ANALYSIS & PDE

Volume 16 No. 9 2023

LUIGI AMBROSIO, AYMERIC BARADAT AND YANN BRENIER

MONGE–AMPÈRE GRAVITATION AS A Γ-LIMIT OF GOOD RATE FUNCTIONS

msp

ANALYSIS AND PDE
Vol. 16 (2023), No. 9, pp. 2005–2040
DOI: 10.2140/apde.2023.16.2005

MONGE–AMPÈRE GRAVITATION AS A Γ-LIMIT OF GOOD RATE FUNCTIONS

LUIGI AMBROSIO, AYMERIC BARADAT AND YANN BRENIER

Monge–Ampère gravitation is a modification of the classical Newtonian gravitation where the linear Poisson equation is replaced by the nonlinear Monge–Ampère equation. This paper is concerned with the rigorous derivation of Monge–Ampère gravitation for a finite number of particles from the stochastic model of a Brownian point cloud, following the formal ideas of a recent work by Brenier (*Bull. Inst. Math. Acad. Sin.* **11**:1(2016), 23–41). This is done in two steps. First, we compute the good rate function corresponding to a large deviation problem related to the Brownian point cloud at fixed positive diffusivity. Second, we study the Γ-convergence of this good rate function, as the diffusivity tends to zero, toward a (nonsmooth) Lagrangian encoding the Monge–Ampère dynamic. Surprisingly, the singularities of the limiting Lagrangian correspond to dissipative phenomena. As an illustration, we show that they lead to sticky collisions in one space dimension.

1. Introduction

Monge–Ampère gravitation. On a periodic domain such as $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$, Newtonian gravitation is commonly described in terms of the density of probability $f(t,x,\xi)$ to find gravitating matter at time t, position $x \in \mathbb{T}^d$ and velocity $\xi \in \mathbb{R}^d$, subject to the Vlasov–Poisson equation

$$\partial_t f(t,x,\xi) + \operatorname{div}_x(\xi f(t,x,\xi)) - \operatorname{div}_\xi(\nabla\varphi(t,x) f(t,x,\xi)) = 0,$$
$$\Delta\varphi(t,x) = \int_{\mathbb{R}^d} f(t,x,\xi)\,\mathrm{d}\xi - 1, \qquad (t,x,\xi) \in \mathbb{R}\times\mathbb{T}^d\times\mathbb{R}^d,$$

where φ is the gravitational potential. Notice that the averaged density, say 1, has been subtracted out from the right-hand side of the Poisson equation, due to the periodicity of the spatial domain. This is a common feature of computational cosmology and it lets the uniform density be a stationary solution. The Vlasov–Poisson system can be seen as an "approximation" to the more nonlinear Vlasov–Monge–Ampère (VMA) system

$$\partial_t f(t,x,\xi) + \operatorname{div}_x(\xi f(t,x,\xi)) - \operatorname{div}_\xi(\nabla\varphi(t,x) f(t,x,\xi)) = 0, \tag{1}$$
$$\det(\mathbb{I} + \mathrm{D}^2\varphi(t,x)) = \int_{\mathbb{R}^d} f(t,x,\xi)\,\mathrm{d}\xi, \qquad (t,x,\xi) \in \mathbb{R}\times\mathbb{T}^d\times\mathbb{R}^d, \tag{2}$$

where the fully nonlinear Monge–Ampère equation substitutes for the linear Poisson equation of Newtonian gravitation. Indeed, for "weak" gravitational potentials, by expanding the determinant around the identity matrix $\mathbb{I}$, we get

$$\det(\mathbb{I} + \mathrm{D}^2\varphi(t,x)) \sim 1 + \operatorname{tr}(\mathrm{D}^2\varphi(t,x)) = 1 + \Delta\varphi(t,x)$$

MSC2020: 49J52, 60F10, 70F40, 70B05.
Keywords: Monge–Ampère gravitation, large deviations, Γ-convergence, Lagrangian mechanics, interacting particle systems.

and recover the Newtonian model approximately (and exactly as $d = 1$). In this paper, we will speak of "Monge–Ampère gravitation" ("MAG" in short). The Vlasov–Monge–Ampère system was introduced and related to the Vlasov–Poisson system in [Brenier and Loeper 2004], and studied as an ODE on the Wasserstein space in [Ambrosio and Gangbo 2008]. It can also be solved numerically thanks to efficient Monge–Ampère solvers recently designed by Mérigot [2011]. It was argued in [Brenier 2011] that MAG may also be seen as an approximation of Newtonian gravitation for which the "Zeldovich approximation" [1970] (see [Frisch et al. 2002; Brenier et al. 2003]), popular in computational cosmology, becomes exact.

Derivation of a discrete model of MAG. In what follows, we will not be directly interested in the aforementioned system, but rather in its discrete version, i.e., when the number of particles is finite. As is well known in optimal transport theory [Brenier 1987; 1991; Villani 2003], the Monge–Ampère equation (2) is solved by the unique function φ such that the map $\mathrm{Id}+\nabla\varphi$ realizes the optimal transport with quadratic cost from the density $\int f \, \mathrm{d}\xi$ to the Lebesgue measure. Then, the kinetic equation (1) is known to be the continuous version of the Newton equations of classical mechanics in a potential given by φ.

In the discrete setting, the stationary Lebesgue measure is replaced by a family $A = (a_1, \dots, a_N) \in (\mathbb{R}^d)^N$ of $N \geq 1$ points in $\mathbb{R}^d$ (here we make the presentation in $\mathbb{R}^d$ instead of $\mathbb{T}^d$ for the sake of simplicity). One can for instance think of a regular lattice approximating in some region a constant density, even though in the sequel the particular choice of $(a_1, \dots, a_N)$ will play no role. We will consider the evolution of a cloud $X = (x_1, \dots, x_N)$ of N particles $x_1, \dots, x_N$ in $\mathbb{R}^d$ whose dynamic is ruled by the $(-1/N)$-convex function induced by the discrete optimal transport problem

$$F(X) := -\min_{\sigma\in\mathfrak{S}_N} \frac{1}{2N}\sum_{i=1}^N |x_i - a_{\sigma(i)}|^2 = -\frac{1}{2} W_2^2\left(\frac{1}{N}\sum_{i=1}^N \delta_{a_i}, \frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right), \tag{3}$$

where W_2 is the so-called Wasserstein distance on $\mathcal{P}_2(\mathbb{R}^d)$, the set of Borel probability measures on $\mathbb{R}^d$ having a finite second-order moment. At least in the case where the optimization problem in (3) admits a unique minimizer $\sigma_{\mathrm{opt}} = \sigma^X_{\mathrm{opt}}$, the analogue of (1), (2) in this framework is easily seen to be formally,

$$\text{for all } i = 1, \dots, N, \quad \frac{\mathrm{d}^2}{\mathrm{d}t^2} x_i(t) = x_i(t) - a_{\sigma_{\mathrm{opt}}(i)}, \tag{4}$$

which can be rewritten as, letting $\mathcal{X}_t := (x_1(t), \dots, x_N(t))$,

$$\frac{1}{N}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{X}_t = -\nabla F(\mathcal{X}_t). \tag{5}$$

Following the ideas of the recent paper [Brenier 2016], we will derive this discrete dynamic from the very elementary stochastic model of a Brownian point cloud. However, in [Brenier 2016], the derivation was obtained by applying two successive large deviation principles (LDP), through a purely formal use of the Freidlin–Wentzell theory [1998]. The main purpose of the present paper is to explain how such a derivation can be made rigorous by substituting for one of the applications of the LDP a PDE method inspired by the famous concept of "onde pilote" introduced by Louis de Broglie [1927] at the early stage of quantum mechanics.

Dealing with the singularities. Due to the lack of uniqueness in the discrete optimal transport problem, solutions of (4) are not always well-defined a priori. Otherwise stated, F is singular, and therefore ∇F in (5) is not everywhere meaningful. A standard choice to give sense to (5) is to restate it as

$$-\frac{1}{N}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{X}_t \in \partial F(\mathcal{X}_t), \tag{6}$$

where $\partial F(\mathcal{X}_t)$ is the subdifferential of F at $\mathcal{X}_t$, or

$$-\frac{1}{N}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{X}_t = \overline{\nabla} F(\mathcal{X}_t),$$

where $\overline{\nabla} F(\mathcal{X}_t)$ is the element of $\partial F(\mathcal{X}_t)$ with minimal Euclidean norm (see Definition 8 below). In these formulations, existence results are available even in the nondiscrete case [Ambrosio and Gangbo 2008].

This is not what we do: our approach selects minimizers of actions appearing as Γ-limits of good rate functions associated with some LDP, under endpoint constraints. These curves do solve (4) in the case where σ_{opt} is unique, but this time, the relaxation is made at the level of the Lagrangian formulation, and not at the level of the Hamiltonian one. In view of (5), we would expect to find the action

$$\int_{t_0}^{t_1} \left\{ \frac{|\dot{\mathcal{X}}_t|^2}{2} - NF(\mathcal{X}_t) \right\} \mathrm{d}t, \tag{7}$$

where t_0, t_1 are some prescribed initial and final times. Instead, our derivation ends up with the smaller action

$$\int_{t_0}^{t_1} \left\{ \frac{|\dot{\mathcal{X}}_t|^2}{2} + \frac{|\mathcal{X}_t - \overline{\nabla} f(\mathcal{X}_t)|^2}{2} \right\} \mathrm{d}t, \tag{8}$$

$$f(X) := \max_{\sigma\in\mathfrak{S}_N} \sum_{i=1}^N x_i a_{\sigma(i)} = \sum_{i=1}^N x_i a_{\sigma_{\mathrm{opt}}^X(i)}, \qquad X = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N.$$

Note that these two actions coincide on curves $\mathcal{X}$ such that, for almost every t, $\sigma_{\mathrm{opt}}^{\mathcal{X}_t}$ is unique (see Section 2.7 for more details). Unexpectedly, this action is *exactly* the one previously suggested by the third author in [Brenier 2011] in order to include dissipative phenomena (such as sticky collisions in one space dimension) in the Monge–Ampère gravitational model!

The classical theory for sticky particles vs. our approach. Systems of particles moving along the line and that stick together when they meet have been studied for a long time, for instance because they were suggested to model the formation of large structures in the universe [Zeldovich 1970]. On the mathematical side, a lot of works have been devoted to studying the limit of this kind of system when the number of particles tends to infinity (see for instance [E et al. 1996; Brenier and Grenier 1998]), and the most recent works generally build on a connection with the theory of optimal transport (see [Natile and Savaré 2009; Brenier et al. 2013; Hynd 2020]). An example illustrating this link, which is one of the main theorems in [Natile and Savaré 2009], is that up to a change of time, the one-dimensional pressureless Euler system with sticky collisions is the gradient flow in the Wasserstein space of $-\frac{1}{2}W_2^2(\mathsf{m}, \cdot)$, where $\mathsf{m} \in \mathcal{P}_2(\mathbb{R})$ is a reference probability measure on the line. In plain English, in these models, particles are only allowed to stick when they meet, and it corresponds to the optimal way of decreasing a certain functional.

Our approach is different. In fact, our model is a least action principle, and therefore is *conservative* and *time-reversible*. In this context, sticky collisions happen due to the presence of an internal energy, corresponding to the discontinuities of the potential energy $X \mapsto -\frac{1}{2}|X - \overline{\nabla} f(X)|^2$ (see formula (51)), and which grows when particles aggregate. Kinetic energy can hence be transferred into internal energy through perfectly inelastic shocks. An output of these considerations is that in our case, particles are not only allowed to stick together; they can also separate.

Outline. In Section 2 we show how to derive MAG starting from a finite number of Brownian particles. This is done in several steps, the main one being the Γ-convergence towards the "effective" singular functional (8) of the good rate functions associated with the large deviations of the solutions of a family of SDEs (up to a change of time). This is stated in Theorem 9, which is our main result. Section 3 is devoted to the proof of Theorem 9. The purpose of Section 4 is to show that in one space dimension, the dissipative phenomena induced by this functional lead to sticky collisions.

Notation. We will work with N particles in $\mathbb{R}^d$, and hence in $(\mathbb{R}^d)^N$. Points of $(\mathbb{R}^d)^N$ will be denoted with capital letters, mainly X, Y or Z. Curves with values in $(\mathbb{R}^d)^N$ will be denoted with calligraphic letters $\mathcal{X}$, $\mathcal{Y}$ or $\mathcal{Z}$. The positions of $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ at time $t \in \mathbb{R}$ will be denoted by $\mathcal{X}_t$, $\mathcal{Y}_t$ and $\mathcal{Z}_t$ respectively.

In order to avoid heavy notation, in most cases, the laws of the processes that we will consider will be continuously parametrized. In these cases, we will use abuses of notation: for instance, we will say that the family of laws $(\mu_\eta)_{\eta>0}$ is tight whenever it is tight for sufficiently small values of η. This is equivalent to $(\mu_{\eta_n})_{n\in\mathbb{N}}$ being tight for all $(\eta_n) \in (\mathbb{R}_+^*)^{\mathbb{N}}$ decreasing to 0.

2. Derivation of the discrete model

2.1. ***The stochastic model of a lattice with Brownian motion.*** Take $A = (a_1, \dots, a_N) \in (\mathbb{R}^d)^N$ to be a family of $N > 1$ points in $\mathbb{R}^d$. We assume each point of this lattice to be subject to Brownian motion for times $t \geq 0$. At time t, the position of point i is

$$a_i + \sqrt{\varepsilon} B_t^i,$$

where $(B^i)_{i=1,\dots,d}$ is a family of N independent normalized Brownian curves and ε monitors the (common) level of noise. As a consequence, at time $t > 0$, the density of probability $\rho_\varepsilon(t, X)$ for the point cloud

$$(a_1 + \sqrt{\varepsilon} B_t^1, \dots, a_N + \sqrt{\varepsilon} B_t^N)$$

to be observed at location $X = (x_1, \dots, x_d) \in (\mathbb{R}^d)^N$, up to a permutation $\sigma \in \mathfrak{S}_N$ of the labels, is easy to compute. We find

$$\rho_\varepsilon(t, X) = \frac{1}{N!\,\sqrt{2\pi\varepsilon t}^{dN}} \sum_{\sigma\in\mathfrak{S}_N} \prod_{i=1}^{N} \exp\left(-\frac{|x_i - a_{\sigma(i)}|^2}{2\varepsilon t}\right),$$

or, in short,

$$\frac{1}{N!\,\sqrt{2\pi\varepsilon t}^{Nd}} \sum_{\sigma\in\mathfrak{S}_N} \exp\left(-\frac{|X - A^\sigma|^2}{2\varepsilon t}\right),$$

where $|\cdot|$ denotes the euclidean norm in $\mathbb{R}^d$ or $(\mathbb{R}^d)^N$ depending on the context, and where, for all $X=(x_1,\dots,x_N)\in(\mathbb{R}^d)^N$, we let

$$X^\sigma=(x_{\sigma(1)},\dots,x_{\sigma(N)}).$$

This was the starting point of the discussion made in [Brenier 2016], using a double large deviation principle.

In the present paper, we rather turn to a PDE viewpoint, where ρ_ε is the solution of the heat equation in $(\mathbb{R}^d)^N$,

$$\frac{\partial\rho_\varepsilon}{\partial t}(t,X)=\frac{\varepsilon}{2}\Delta\rho_\varepsilon(t,X), \tag{9}$$

with, as initial condition, the delta measure located at $A=(a_1,\dots,a_N)\in(\mathbb{R}^d)^N$ and symmetrized with respect to $\sigma\in\mathfrak{S}_N$, namely

$$\rho_\varepsilon(0,X)=\frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N}\delta_{A^\sigma}. \tag{10}$$

In some sense, we have solved the heat equation in the space of "point clouds" $(\mathbb{R}^d)^N/\mathfrak{S}_N$, with initial position A, defined up to a permutation $\sigma\in\mathfrak{S}_N$ of the labels $i=1,\dots,N$.

2.2. ***"Surfing" the "heat wave".*** After solving the heat equation (9)–(10), in the space of "clouds" $(\mathbb{R}^d)^N/\mathfrak{S}_N$, we introduce the companion ODE in the space $(\mathbb{R}^d)^N$:

$$\frac{\mathrm{d}\mathcal{X}_t^\varepsilon}{\mathrm{d}t}=v_\varepsilon(t,\mathcal{X}_t^\varepsilon),\qquad v_\varepsilon(t,X)=-\frac{\varepsilon}{2}\nabla\log\rho_\varepsilon(t,X), \tag{11}$$

or, more explicitly

$$\begin{aligned} v_\varepsilon(t,X)&=\frac{1}{2t}\frac{\sum_{\sigma\in\mathfrak{S}_N}(X-A^\sigma)\exp(-|X-A^\sigma|^2/(2\varepsilon t))}{\sum_{\sigma\in\mathfrak{S}_N}\exp(-|X-A^\sigma|^2/(2\varepsilon t))}\\ &=\frac{1}{2t}\left(X-\frac{\sum_{\sigma\in\mathfrak{S}_N}A^\sigma\exp((X\cdot A^\sigma)/(\varepsilon t))}{\sum_{\sigma\in\mathfrak{S}_N}\exp((X\cdot A^\sigma)/(\varepsilon t))}\right), \end{aligned} \tag{12}$$

where if U and V are in $(\mathbb{R}^d)^N$, then $U\cdot V$ denotes the inner product between U and V. This velocity is chosen so that

$$\frac{\partial\rho_\varepsilon}{\partial t}(t,X)+\operatorname{div}(\rho_\varepsilon(t,X)v_\varepsilon(t,X))=0,$$

i.e., for the density ρ_ε to be transported by the velocity field v_ε. We may solve this ODE for arbitrarily chosen position $\mathcal{X}_{t_0}\in(\mathbb{R}^d)^N$ (up to reordering) and initial time $t_0>0$.

Put another way, we consider the characteristics corresponding to the heat equation (9)–(10), interpreted as a continuity equation, associated to our Brownian point cloud.

Remark 1. By doing that change of perspective, we just mimic the idea of quantum particles driven by the "onde pilote", as imagined by Louis de Broglie [1927; 1959] at the early stage of quantum mechanics.

Indeed, in our case, the velocity $v^\varepsilon = \nabla\varphi^\varepsilon$ is the gradient of the scalar function $\varphi^\varepsilon := (-\varepsilon/2)\log\rho^\varepsilon$, and the pair $(\rho^\varepsilon, \varphi^\varepsilon)$ is easily seen to solve the system

$$\begin{cases} \partial_t\rho^\varepsilon + \operatorname{div}(\rho^\varepsilon\nabla\varphi^\varepsilon) = 0, \\ \partial_t\varphi^\varepsilon + \dfrac{1}{2}|\nabla\varphi^\varepsilon|^2 = -\dfrac{\varepsilon^2}{2}\dfrac{\Delta\sqrt{\rho^\varepsilon}}{\sqrt{\rho^\varepsilon}}, \end{cases} \tag{13}$$

that is, the characteristics follow the trajectories of Newton's law in a potential induced by ρ^ε.

In the quantum case, something very similar can be found with the help of the Madelung transform [1927]. Namely, if the complex function Ψ^ε solves the Schrödinger equation

$$i\partial_t\Psi^\varepsilon + \frac{\varepsilon}{2}\Delta\Psi^\varepsilon = 0,$$

writing $\Psi^\varepsilon = \sqrt{\rho^\varepsilon}e^{i\varphi^\varepsilon/\varepsilon}$ for a pair $(\rho^\varepsilon, \varphi^\varepsilon)$ of real functions, then this pair is shown to formally solve the very similar system

$$\begin{cases} \partial_t\rho^\varepsilon + \operatorname{div}(\rho^\varepsilon\nabla\varphi^\varepsilon) = 0, \\ \partial_t\varphi^\varepsilon + \dfrac{1}{2}|\nabla\varphi^\varepsilon|^2 = \dfrac{\varepsilon^2}{2}\dfrac{\Delta\sqrt{\rho^\varepsilon}}{\sqrt{\rho^\varepsilon}}, \end{cases} \tag{14}$$

and this observation was the starting point of de Broglie's interpretation of quantum mechanics. In this case, the potential in the right-hand side of the second equation is called the Bohm quantum potential. However, the analysis of (14) is substantially more difficult than the one of (13), due to the possible vanishing of the wave function Ψ^ε during the evolution.

This analogy is not a coincidence. Indeed, it is known [von Renesse 2012] that the Schrödinger equation in its Madelung formulation (14) is formally the Hamiltonian flow corresponding to the Lagrangian

$$\mathcal{L}^\varepsilon_{\text{quantum}}(\rho, \nabla\varphi) := \frac{1}{2}\int\left\{|\nabla\varphi|^2 - \left|\frac{\varepsilon}{2}\nabla\log\rho\right|^2\right\}\rho,$$

in the geometry of optimal transport, while system (13), which admits solutions of the heat equations as particular solutions, is *rigorously* the Hamiltonian flow corresponding to the Lagrangian

$$\mathcal{L}^\varepsilon_{\text{heat}}(\rho, \nabla\varphi) := \frac{1}{2}\int\left\{|\nabla\varphi|^2 + \left|\frac{\varepsilon}{2}\nabla\log\rho\right|^2\right\}\rho,$$

in the geometry of optimal transport [Conforti 2019]. The latter Lagrangian appears naturally in the theory of entropic optimal transport; see [Gentil et al. 2017; Gigli and Tamanini 2020].

2.3. ***Large deviations of the "heat wave" ODE.*** Let us now add to the ODE of the previous subsection a noise of the form

$$\mathrm{d}\mathcal{X}^{\varepsilon,\eta}_t = v_\varepsilon(t, \mathcal{X}^{\varepsilon,\eta}_t)\,\mathrm{d}t + \sqrt{\frac{\eta}{t}}\,\mathrm{d}\mathcal{W}_t, \tag{15}$$

where η is a positive number, $(\mathcal{W}_t)$ is a standard Brownian motion in $(\mathbb{R}^d)^N$, and where the scaling prefactor $1/\sqrt{t}$ has been chosen to recover MAG at Section 2.6. That is, we include a second time-dependent level of noise to the model: we perturb the characteristics that were already generated, through the heat equation, by the Brownian motion of our original point cloud.

Our main finding is that when η and ε are small and up to a change of time, the trajectories charged by the solution of this SDE starting from $P \in (\mathbb{R}^d)^N$ at time $t_0 > 0$ and which happen to be close to $Q \in (\mathbb{R}^d)^N$ at time $t_1 > t_0$ are well-approximated by MAG.

The purpose of the rest of this section will be to make this rough statement precise. When we say that some random trajectories are well-approximated by MAG, we mean that they are close in the uniform topology to minimizers of the action (8), with large probability. Justifying this fact will require several steps and intermediate functionals. As the times t_0 and t_1, as well as the endpoints P and Q, will be fixed in what follows, we decided not to refer to them in the notation for the different functionals and laws that will appear.

Since for fixed $\varepsilon > 0$ and $t \geq t_0 > 0$, v_ε is a smooth velocity field, the existence of a strong solution and pathwise uniqueness for (15) is standard once fixed a law for the initial position $\mathcal{X}^{\varepsilon,\eta}_{t_0}$ at some $t_0 > 0$. Since we want to consider indistinguishable particles, a relevant choice of initial law consists in taking $\mathcal{X}^{\varepsilon,\eta}_{t_0} = P^\sigma$ with probability $1/(N!)$, given some $P \in (\mathbb{R}^d)^N$ and $\sigma \in \mathfrak{S}_N$. For convenience, from now on, we denote by $\{P^\sigma\}$ the set $\{P^\sigma : \sigma \in \mathfrak{S}_N\}$. The law just described is nothing but the uniform law on $\{P^\sigma\}$.

Remark 2. Actually, at this stage, it would be possible to reintroduce distinguishability: Theorem 3, Corollary 4, Proposition 7 and Theorem 9 below could easily be written for distinguishable particles, that is, with constraints on the endpoints of the curves, and not on these endpoints up to reordering. We decided to keep on working on clouds of indistinguishable particles in order to avoid crossings of trajectories in Section 4.

The first step consists in using classical Freidlin–Wentzell theory [1998] (see also [Dembo and Zeitouni 1998]) in order to pass to the limit $\eta \to 0$, while $\varepsilon > 0$ is kept fixed, in the sense of large deviations (we omit the proof since it consists in adapting in a straightforward way [Dembo and Zeitouni 1998, Theorem 5.6.3] to time-dependent entries and more general initial law for the SDE).

Theorem 3. *Let us fix two positive times $0 < t_0 < t_1$ and $P \in (\mathbb{R}^d)^N$. For fixed $\varepsilon > 0$ and as $\eta \to 0$, the family of laws $(\mu_{\varepsilon,\eta})$ of the solution of (15) between times t_0 and t_1 and starting from the uniform law on $\{P^\sigma\}$ satisfies the LDP on $C^0([t_0, t_1]; (\mathbb{R}^d)^N)$ with good rate function L^0_ε defined for all $\mathcal{X} = (\mathcal{X}_t)_{t\in[t_0,t_1]}$ by*

$$L^0_\varepsilon(\mathcal{X}) = \begin{cases} \frac{1}{2}\int_{t_0}^{t_1} |\dot{\mathcal{X}}_t - v_\varepsilon(t, \mathcal{X}_t)|^2 \times t \,\mathrm{d}t & \text{if } \mathcal{X} \in H^1([t_0, t_1]; (\mathbb{R}^d)^N) \text{ and } \mathcal{X}_{t_0} \in \{P^\sigma\}, \\ +\infty, & \text{else.} \end{cases} \tag{16}$$

In the rest of the article, we will call these kind of functionals “actions”, instead of the usual terminology “good rate function”.

An outcome of this result is that with large probability, when η is small, $\mathcal{X}^{\eta,\varepsilon}_{t_1}$ is close to the position at time t_1 of the solution of the ODE (11) starting from P, up to reordering. But now, we want to use Theorem 3 in order to describe the behavior of the solutions of the SDE (15) when η is small, under the large deviation assumption that its final position $\mathcal{X}^{\varepsilon,\eta}_{t_1}$ is far from this expected value.

For this, we take $Q \in (\mathbb{R}^d)^N$, and we assume that we observe $\mathcal{X}^{\varepsilon,\eta}_{t_1}$ to be close to Q, up to reordering. To quantify this closeness, we consider a new small parameter $\delta > 0$, and we work with the laws $(\mu_{\varepsilon,\eta})$ from Theorem 3, conditioned to the event $\{\mathcal{X}^{\varepsilon,\eta}_{t_1} \in \bigcup_{\sigma\in\mathfrak{S}_N} \bar{B}(Q^\sigma, \delta)\}$, where for a given $X \in (\mathbb{R}^d)^N$, $\bar{B}(X, \delta)$ stands for the closed ball of center X and radius δ. MAG will be obtained by studying the limit of these conditional laws when $\eta \to 0$, then $\delta \to 0$ and finally $\varepsilon \to 0$. We refer to Remark 12 for a discussion about the order in which we let the different parameters tend to 0.

Concerning the limit $\eta \to 0$, Theorem 3 implies the following.

Corollary 4. *Let us fix $\varepsilon, \delta > 0$, and call $\mathcal{E}^\delta$ the closed subset of $C^0([t_0, t_1]; (\mathbb{R}^d)^N)$ defined by*

$$\mathcal{E}^\delta := \Big\{ \mathcal{X} \in C^0([t_0, t_1]; (\mathbb{R}^d)^N) : \mathcal{X}_{t_1} \in \bigcup_{\sigma \in \mathfrak{S}_N} \bar{B}(Q^\sigma, \delta) \Big\}.$$

The family of conditional laws $(\mu^\delta_{\varepsilon,\eta} := \mu_{\varepsilon,\eta}(\,\cdot : \mathcal{E}^\delta))_{\eta>0}$ is tight. Moreover, its limit points for the topology of narrow convergence as $\eta \to 0$ only charge minimizers of the functional

$$L^\delta_\varepsilon(\mathcal{X}) = \begin{cases} \frac{1}{2}\int_{t_0}^{t_1} |\dot{\mathcal{X}}_t - v_\varepsilon(t, \mathcal{X}_t)|^2 \times t \,\mathrm{d}t & \text{if } \mathcal{X} \in H^1([t_0, t_1]; (\mathbb{R}^d)^N), \mathcal{X}_{t_0} \in \{P^\sigma\} \text{ and } \mathcal{X}_{t_1} \in \bigcup_{\sigma \in \mathfrak{S}_N} \bar{B}(Q^\sigma, \delta), \\ +\infty, & \text{else.} \end{cases} \tag{17}$$

Proof. Let us first prove the tightness property. Let $\mathcal{X}$ be a curve in the interior of $\mathcal{E}^\delta$. As it satisfies an LDP associated with a good rate function in a Polish space, by virtue of [Dembo and Zeitouni 1998, Exercise 4.1.10], for fixed $\varepsilon > 0$, the family of laws $(\mu_{\varepsilon,\eta})_{\eta>0}$ is exponentially tight. Hence, there is a compact K (we call K^c its complement in $C^0([t_0, t_1]; (\mathbb{R}^d)^N)$) such that

$$\limsup_{\eta \to 0} \eta \log \mu_{\varepsilon,\eta}(K^c) \le -L^0_\varepsilon(\mathcal{X}) - 1.$$

Therefore, we find

$$\begin{aligned} \limsup_{\eta \to 0} \eta \log \mu^\delta_{\varepsilon,\eta}(K^c) &= \limsup_{\eta \to 0} \{\eta \log \mu_{\varepsilon,\eta}(K^c \cap \mathcal{E}^\delta) - \eta \log \mu_{\varepsilon,\eta}(\mathcal{E}^\delta)\} \\ &\le \limsup_{\eta \to 0} \eta \log \mu_{\varepsilon,\eta}(K^c) - \liminf_{\eta \to 0} \eta \log \mu_{\varepsilon,\eta}(\mathring{\mathcal{E}}^\delta) \\ &\le -L^0_\varepsilon(\mathcal{X}) - 1 + L^0_\varepsilon(\mathcal{X}) \le -1. \end{aligned}$$

The tightness follows.

Now, let us consider μ a limit point of $(\mu^\delta_{\varepsilon,\eta})$ as $\eta \to 0$, and (η_n) a sequence of positive numbers decreasing to 0, with $\mu^\delta_{\varepsilon,\eta_n} \to \mu$ as $n \to +\infty$. We will argue that whenever $\mathcal{X}$ is not a minimizer of L^δ_ε, then $\mathcal{X}$ is not in the support of μ. First, for all $\eta > 0$, the support of $\mu^\delta_{\varepsilon,\eta}$ is a subset of $\mathcal{E}^\delta$. As the latter is closed, this is also the case for the support of μ. So let us take $\mathcal{X} \in \mathcal{E}^\delta$, which is not a minimizer of L^δ_ε. In particular, $L^0_\varepsilon(\mathcal{X}) > \inf_{\mathcal{E}^\delta} L^0_\varepsilon$. As L^0_ε is lower semicontinuous, there exists an open set U of $C^0([t_0, t_1]; (\mathbb{R}^d)^N)$ containing $\mathcal{X}$ such that $\inf_{\bar{U}} L^0_\varepsilon > \inf_{\mathcal{E}^\delta} L^0_\varepsilon$. Let us show that $\mu(U) = 0$.

By the Portmanteau theorem, we have

$$\mu(U) \le \liminf_{n \to +\infty} \mu^\delta_{\varepsilon,\eta_n}(U).$$

By the definition of $(\mu^\delta_{\varepsilon,\eta})$, we have

$$\eta_n \log \mu^\delta_{\varepsilon,\eta_n}(U) = \eta_n \log \mu_{\varepsilon,\eta_n}(U \cap \mathcal{E}^\delta) - \eta_n \log \mu_{\varepsilon,\eta_n}(\mathcal{E}^\delta) \le \eta_n \log \mu_{\varepsilon,\eta_n}(\bar{U}) - \eta_n \log \mu_{\varepsilon,\eta_n}(\mathring{\mathcal{E}}^\delta).$$

The large deviation principle of Theorem 3 lets us estimate the lim sup of this quantity by

$$\limsup_{n \to +\infty} \eta_n \log \mu^\delta_{\varepsilon,\eta_n}(U) \le \inf_{\mathring{\mathcal{E}}^\delta} L^0_\varepsilon - \inf_{\bar{U}} L^0_\varepsilon.$$

To conclude that this quantity is negative, and therefore that $\mu^{\delta}_{\varepsilon,\eta_n}(U)$ tends to 0 as $n \to +\infty$, it suffices to notice that $\inf_{\mathring{\mathcal{E}}^{\delta}} L^0_\varepsilon = \inf_{\mathcal{E}^{\delta}} L^0_\varepsilon$ (for instance by the easy fact that the infimum of L^δ_ε is continuous with respect to δ), and to use the definition of U. The result follows. □

2.4. *From the Γ-convergence of the actions to the narrow convergence of the laws.* In the previous subsection, we justified why the conditional laws $(\mu^{\delta}_{\varepsilon,\eta})$ from Corollary 4 are well-described by the action L^δ_ε defined by formula (17) as $\eta \to 0$: in this limit, these laws mainly charge small neighborhoods of minimizers of that action. Now, we want to argue that in order to study these laws when not only η is small, but also δ and ε, we have to study the action L^δ_ε in that regime, *in the sense of* Γ*-convergence*.

This assertion relies on the two following lemmas:

Lemma 5. *Let (Ω, d) be a metric space, and $(\mathcal{L}_n)_{n\in\mathbb{N}}$ be a sequence of lower semicontinuous functionals from Ω to $\mathbb{R}_+ \cup \{+\infty\}$ having compact sublevels, uniformly in $n \in \mathbb{N}$. Assume that $(\mathcal{L}_n)$ has a Γ-limit $\mathcal{L}$. Assume furthermore that $\mathcal{L}$ is not uniformly equal to $+\infty$. Finally, consider $(\mu_n) \in \mathcal{P}(\Omega)^{\mathbb{N}}$ a sequence of Borel probability measures on Ω, such that, for all n, μ_n only charges minimizers of $\mathcal{L}_n$. Then, (μ_n) is tight, and any of its limit points in the narrow topology only charges minimizers of $\mathcal{L}$.*

Lemma 6. *The family of actions (L^δ_ε) defined in (17) have compact sublevels in $C^0([t_0, t_1]; (\mathbb{R}^d)^N)$, uniformly in $\varepsilon, \delta > 0$.*

Using these lemmas, we see that if we manage to identify a Γ-limit L for L^δ_ε as $\varepsilon, \delta \to 0$, then in this limit, any family (μ^δ_ε) of limits of $(\mu^\delta_{\varepsilon,\eta})$ as $\eta \to 0$ will mainly charge small neighborhoods of minimizers of the limiting L. Before doing so in the next subsection, let us prove our two lemmas.

Proof of Lemma 5. Let x be a minimizer of $\mathcal{L}$, and (x_n) be an associated recovery sequence, that is, $x_n \to x$ as $n \to +\infty$, and $\limsup_{n\to+\infty} \mathcal{L}_n(x_n) \le \mathcal{L}(x) = \inf \mathcal{L}$. Up to forgetting the first terms, we can assume that $\mathcal{L}_n(x_n)$ is finite for all $n \in \mathbb{N}$. Now, call $M := \sup_{n\in\mathbb{N}} \mathcal{L}_n(x_n)$. By assumption, the set

$$K := \overline{\bigcup_{n\in\mathbb{N}} \{z \in \Omega : \mathcal{L}_n(z) \le M\}}$$

is compact, and by definition of M it contains all the minimizers of all the functionals $\mathcal{L}_n$, $n \in \mathbb{N}$. Therefore, for all $n \in \mathbb{N}$, $\mu_n(K) = 1$, and the tightness follows.

Let μ be a limit point of (μ_n) for the topology of narrow convergence. Up to considering a subsequence, we assume that $\mu_n \to \mu$. Let x be in the support of μ. It is easy to see that there exists a sequence (x_n) such that $x_n \to x$ as $n \to +\infty$, and, for all $n \in \mathbb{N}$, x_n is in the support of μ_n. But then by assumption, for all n, x_n is a minimizer of $\mathcal{L}_n$, and therefore, by standard considerations about Γ-convergence, x is a minimizer of $\mathcal{L}$. □

Proof of Lemma 6. For all $\varepsilon, \delta > 0$, the action L^δ_ε coincides with L^0_ε (defined in (16)) inside of the closed set $\mathcal{E}^\delta$ and is $+\infty$ outside of this closed set. Therefore, we just need to prove that L^0_ε has compact sublevels, uniformly in $\varepsilon > 0$. Actually, precompacity suffices by lower semicontinuity of L^0_ε. To do so, we will use the following bound, which holds as a consequence of (12) for all $\varepsilon > 0$, $t \in [t_0, t_1]$ and $X \in (\mathbb{R}^d)^N$:

$$|v_\varepsilon(t, X)| \le \frac{|A| + |X|}{2t_0}. \tag{18}$$

We will prove that, for all $M > 0$, there exists $M' > 0$ (uniform in ε) such that, for all $\varepsilon > 0$ and $\mathcal{X} \in C^0([t_0, t_1]; (\mathbb{R}^d)^N)$, whenever $L^0_\varepsilon(\mathcal{X}) \le M$, we have

$$\frac{1}{2}\int_{t_0}^{t_1} |\dot{\mathcal{X}}_t|^2 \,\mathrm{d}t \le M'.$$

This is enough to conclude since it is well-known that the set $H^1([t_0, t_1]; (\mathbb{R}^d)^N)$ is compactly embedded in $C^0([t_0, t_1]; (\mathbb{R}^d)^N)$.

So let us consider $M, \varepsilon > 0$, and a curve $\mathcal{X}$ such that $L^0_\varepsilon(\mathcal{X}) \le M$. Note that in particular, $\mathcal{X}_{t_0} \in \{P^\sigma\}$. We have, for all $t \in [t_0, t_1]$,

$$\begin{aligned}
\frac{1}{2}\int_{t_0}^{t} |\dot{\mathcal{X}}_s|^2 \,\mathrm{d}s &= \int_{t_0}^{t} \frac{|\dot{\mathcal{X}}_s - v_\varepsilon(s, \mathcal{X}_s) + v_\varepsilon(s, \mathcal{X}_s)|^2}{2} \,\mathrm{d}s \\
&\le \int_{t_0}^{t_1} |\dot{\mathcal{X}}_s - v_\varepsilon(s, \mathcal{X}_s)|^2 \,\mathrm{d}s + \int_{t_0}^{t} |v_\varepsilon(s, \mathcal{X}_s)|^2 \,\mathrm{d}s \\
&\le \frac{1}{t_0}\int_{t_0}^{t_1} |\dot{\mathcal{X}}_s - v_\varepsilon(s, \mathcal{X}_s)|^2 \times s \,\mathrm{d}s + \frac{1}{4t_0^2}\int_{t_0}^{t} (|A| + |\mathcal{X}_s|)^2 \,\mathrm{d}s \\
&\le \frac{2M}{t_0} + \frac{(t_1 - t_0)|A|^2}{2t_0^2} + \frac{1}{2t_0^2}\int_{t_0}^{t} \left| \mathcal{X}_{t_0} + \int_{t_0}^{s} \dot{\mathcal{X}}_\tau \,\mathrm{d}\tau \right|^2 \mathrm{d}s \\
&\le \frac{2M}{t_0} + \frac{t_1 - t_0}{t_0^2}\left\{ \frac{|A|^2}{2} + |P|^2 + \int_{t_0}^{t}\int_{t_0}^{s} |\dot{\mathcal{X}}_\tau|^2 \,\mathrm{d}\tau \,\mathrm{d}s \right\},
\end{aligned}$$

where we used (18) to get the third line. We deduce our claim from the Grönwall lemma. □

2.5. ***The convergence results.*** As already explained, understanding the behavior of families (μ^δ_ε) of limit points of $(\mu^\delta_{\varepsilon,\eta})$ as $\eta \to 0$ when ε and δ are small amounts to understanding the behavior of the family of actions (L^δ_ε) in the Γ-convergence sense. This is what we propose to do now. More specifically, we will see that (L^δ_ε) has a Γ-limit, when first $\delta \to 0$, and then $\varepsilon \to 0$. Doing so, we ensure that limit points of the family (μ^δ_ε) in the relevant asymptotic only charge minimizers of the corresponding actions; see Corollary 11 below. We discuss the question of swapping these limits in Remark 12.

Thanks to the smoothness of v^ε, the first Γ-limit, as $\delta \to 0$, is very simple and we omit the proof.

Proposition 7. *Let $\varepsilon > 0$. As δ tends to zero, the family of actions (L^δ_ε) Γ-converges to*

$$L_\varepsilon(\mathcal{X}) = \begin{cases} \frac{1}{2}\int_{t_0}^{t_1} |\dot{\mathcal{X}}_t - v_\varepsilon(t, \mathcal{X}_t)|^2 \times t \,\mathrm{d}t & \text{if } \mathcal{X} \in H^1([t_0, t_1]; (\mathbb{R}^d)^N),\ \mathcal{X}_{t_0} \in \{P^\sigma\} \text{ and } \mathcal{X}_{t_1} \in \{Q^\sigma\}, \\ +\infty, & \text{else}. \end{cases}$$

The second Γ-convergence, as $\varepsilon \to 0$, is more intricate and can be seen as the main result of this paper, because it involves the singular limit of the vector fields (v^ε) as $\varepsilon \to 0$. Before stating it, we need to introduce a few objects.

Define the following smooth functions, which are convex in X:

$$\text{for all } \varepsilon > 0,\ t > 0,\ X \in (\mathbb{R}^d)^N, \qquad f_\varepsilon(t, X) := \varepsilon t \log\left[\frac{1}{N!}\sum_{\sigma \in \mathfrak{S}_N} \exp\left(\frac{X \cdot A^\sigma}{t\varepsilon}\right)\right]. \tag{19}$$

It has the property that, for all $\varepsilon > 0$, $t > 0$, and $X \in (\mathbb{R}^d)^N$,

$$v_\varepsilon(t, X) = \frac{X - \nabla f_\varepsilon(t, X)}{2t}.$$

As a consequence, we can rewrite L_ε for all $\varepsilon > 0$ as

$$L_\varepsilon(\mathcal{X}) = \begin{cases} \frac{1}{2}\int_{t_0}^{t_1} |\dot{\mathcal{X}}_t - (\mathcal{X}_t - \nabla f_\varepsilon(t, \mathcal{X}_t))/(2t)|^2 \times t \, \mathrm{d}t & \text{if } \mathcal{X} \in H^1([t_0, t_1]; (\mathbb{R}^d)^N), \mathcal{X}_{t_0} \in \{P^\sigma\} \text{ and } \mathcal{X}_{t_1} \in \{Q^\sigma\}, \\ +\infty & \text{else.} \end{cases}$$

When ε tends to zero, by virtue of the so-called Laplace's principle, we have the pointwise convergence

$$\lim_{\varepsilon \to 0} f_\varepsilon(t, X) = \max_{\sigma \in \mathfrak{S}_N} X \cdot A^\sigma =: f(X). \tag{20}$$

Notice that f is linked to the function F defined in (3) by the formula,

$$\text{for all } X \in (\mathbb{R}^d)^N, \quad f(X) = \frac{|A|^2 + |X|^2}{2} + NF(X). \tag{21}$$

The function f no longer depends on the time variable, and it is a convex function with finite values. As a consequence, for each $X \in (\mathbb{R}^d)^N$, the subdifferential $\partial f(X)$ of f at X is nonempty. We will consider the extended gradient $\overline{\nabla} f(X)$ of f at X defined as:

Definition 8 (extended gradient). We define the *extended gradient* of a real-valued convex function h at X, denoted by $\overline{\nabla} h(X)$, to be the element of $\partial h(X)$ with minimal Euclidean norm.

We are now ready to state our result concerning the limit $\varepsilon \to 0$.

Theorem 9. *As ε tends to 0, the family of actions $(L_\varepsilon)_{\varepsilon>0}$ Γ-converges to*

$$L(\mathcal{X}) = \begin{cases} \frac{1}{2}\int_{t_0}^{t_1} |\dot{\mathcal{X}}_t - (\mathcal{X}_t - \overline{\nabla} f(\mathcal{X}_t))/(2t)|^2 \times t \, \mathrm{d}t & \text{if } \mathcal{X} \in H^1([t_0, t_1]; (\mathbb{R}^d)^N), \mathcal{X}_{t_0} \in \{P^\sigma\} \text{ and } \mathcal{X}_{t_1} \in \{Q^\sigma\}, \\ +\infty & \text{else} \end{cases} \tag{22}$$

for the topology of uniform convergence of $C^0([t_0, t_1]; (\mathbb{R}^d)^N)$.

Remark 10. It is relevant to wonder what exactly in the convergence $f_\varepsilon \to f$ implies Theorem 9. It is not so simple to answer due to the dependence in t of f_ε and because the proof involves several manipulations of formula (22). However, the main step of the proof is Lemma 15 below. Now, at least in the autonomous case, several works that are posterior to the first version of the present paper study results similar to Lemma 15 in greater generality, namely in Hilbert spaces [Ambrosio et al. 2021] or in measured metric spaces [Monsaingeon et al. 2023]. In [Ambrosio et al. 2021], the good notion of convergence for $f_\varepsilon \to f$ is *Mosco convergence*. We give more details on this in Remark 16.

As a consequence of Lemmas 5 and 6, this theorem clearly implies the following.

Corollary 11. *Consider the family of laws $(\mu^\delta_{\varepsilon,\eta})$ defined in Corollary 4, and three sequences $(\eta_n)_{n\in\mathbb{N}}$, $(\delta_m)_{m\in\mathbb{N}}$ and $(\varepsilon_p)_{p\in\mathbb{N}}$ decreasing to 0. Then, there exist subsequences $(\eta'_n)_{n\in\mathbb{N}}$, $(\delta'_m)_{m\in\mathbb{N}}$ and $(\varepsilon'_p)_{p\in\mathbb{N}}$ such that the triple limit*

$$\lim_{p\to+\infty} \lim_{m\to+\infty} \lim_{n\to+\infty} \mu^{\delta_m}_{\varepsilon_p,\eta_n}$$

exists in the topology of narrow convergence and only charge minimizers of L as defined by (22).

In particular, if L admits a unique minimizer $\mathcal{X}$, the whole family converges:

$$\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\lim_{\eta\to 0}\mu^\delta_{\varepsilon,\eta} = \delta_{\mathcal{X}}.$$

Let us now comment on the order in which these limits are taken.

Remark 12. Up to potentially considering subsequences, we are studying the behavior of the conditioned laws $(\mu^\delta_{\varepsilon,\eta})$ in the limit $\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\lim_{\eta\to 0}$, and one could wonder whether these limits could be swapped. We recall that ε stands for the level of noise of the original point cloud, that η stands for the level of perturbation of the companion ODE, and that δ is the precision of the observation at the final time.

• Swapping $\lim_{\varepsilon\to 0}$ and $\lim_{\delta\to 0}$ is easy: it amounts to studying the dependence of the limiting action (22) when Q varies. Essentially, this swap would be a consequence of the fact that v^ε is bounded on compact sets, uniformly in time and ε.

• Swapping $\lim_{\delta\to 0}$ and $\lim_{\eta\to 0}$ would be more delicate, but doable as well. We would first need to prove that the family $(\mu^\delta_{\varepsilon,\eta})$ from Corollary 4 converges when $\delta\to 0$, with fixed ε and η, as classically done in the theory of bridges of processes, and then write a large deviation principle for these bridges in place of Theorem 3.

• Finally, not taking into consideration the limit in δ because of the two previous points, the question of how to swap $\lim_{\varepsilon\to 0}$ with $\lim_{\eta\to 0}$ relates to the question of building solutions to SDEs with singular coefficients, and lies beyond the scope of this article. A related question that we also do not want to address is the question of quantifying how small η needs to be with respect to ε to be able to take a simultaneous limit in ε and η. To answer it, we would need to study the dependence in ε of the rates of convergence in the large deviation principle, which is probably a very delicate question, once again because of the singularities of v^ε appearing as $\varepsilon\to 0$.

We will prove Theorem 9 in Section 3 below, but before doing so, let us show that up to changing time, we recover MAG. Notice L has compact sublevels as a consequence of the Γ-convergence and Lemma 6. Hence, the existence of global minimizers for L (and hence for all the forthcoming functionals) follows from the direct method of calculus of variations.

2.6. ***A change of time leading to Monge–Ampère gravitation.*** Through the change of variable

$$t = \exp(2\theta), \quad \mathcal{Z}_\theta = \mathcal{X}_{\exp(2\theta)}, \quad \theta_0 = \tfrac{1}{2}\log t_0, \quad \theta_1 = \tfrac{1}{2}\log t_1,$$

we observe that, for all $\mathcal{X}\in C^0([t_0,t_1];(\mathbb{R}^d)^N)$, $L(\mathcal{X}) = \frac{1}{2}\Lambda(\mathcal{Z})$, with

$$\Lambda(\mathcal{Z}) = \begin{cases} \frac{1}{2}\int_{\theta_0}^{\theta_1} |\dot{\mathcal{Z}}_\theta - (\mathcal{Z}_\theta - \overline{\nabla} f(\mathcal{Z}_\theta))|^2\,\mathrm{d}\theta & \text{if } \mathcal{Z}\in H^1([\theta_0,\theta_1];(\mathbb{R}^d)^N),\ \mathcal{Z}_{\theta_0}\in\{P^\sigma\} \text{ and } \mathcal{Z}_{\theta_1}\in\{Q^\sigma\},\\ +\infty & \text{else.}\end{cases}$$

(Recall the definition (20) of f.)

It turns out to be equivalent to the following one (in which we recognize (8)):

$$\Lambda'(\mathcal{Z}) = \begin{cases} \int_{\theta_0}^{\theta_1} \big\{\frac{1}{2}|\dot{\mathcal{Z}}_\theta|^2 + \frac{1}{2}|\mathcal{Z}_\theta - \overline{\nabla} f(\mathcal{Z}_\theta)|^2\big\}\,\mathrm{d}\theta & \text{if } \mathcal{Z}\in H^1([\theta_0,\theta_1];(\mathbb{R}^d)^N),\ \mathcal{Z}_{\theta_0}\in\{P^\sigma\} \text{ and } \mathcal{Z}_{\theta_1}\in\{Q^\sigma\},\\ +\infty & \text{else.}\end{cases}$$

To see this, it suffices to expand the square and to remark that the mixed product is an exact time derivative, so that its integral only involves the endpoints P and Q. This is done in a slightly different context in the proof of Lemma 14 below.

2.7. ***Application of the least action principle.*** We observe that the points Z where f is differentiable are those for which the maximum in the definition (20) of f is reached by a unique permutation σ_{opt} so that $\nabla f(Z)$ is nothing but $A^{\sigma_{\text{opt}}}$. For such points Z, we get

$$\tfrac{1}{2}|Z-\nabla f(Z)|^2 = \tfrac{1}{2}|Z-A^{\sigma_{\text{opt}}}|^2 = -NF(Z)$$

(by definition (3) of F), while, on the set $\mathcal{N}$ of nondifferentiability of f, we rather have

$$\tfrac{1}{2}|Z-\overline{\nabla} f(Z)|^2 < -NF(Z);$$

see for instance Proposition 27 below in the case of dimension 1. So the action we have obtained in the previous section, namely Λ', bounds from below

$$\Lambda^+(\mathcal{Z}) = \begin{cases} \int_{\theta_0}^{\theta_1}\left\{\frac{1}{2}|\dot{\mathcal{Z}}_\theta|^2 - NF(\mathcal{Z}_\theta)\right\}\mathrm{d}\theta & \text{if } \mathcal{Z}\in H^1([\theta_0,\theta_1];(\mathbb{R}^d)^N),\ \mathcal{Z}_{\theta_0}\in\{P^\sigma\} \text{ and } \mathcal{Z}_{\theta_1}\in\{Q^\sigma\}, \\ +\infty & \text{else.}\end{cases}$$

This second action, already announced in (7), is definitely strictly larger than the first one for those curves $\theta \to \mathcal{Z}_\theta$ which take values in $\mathcal{N}$ (where f and F are not differentiable) on a set of times $\theta \in [\theta_0,\theta_1]$ with positive Lebesgue measure. So the least action principle may provide different optimal curves, depending on the action we choose. However, if a curve is optimal for Λ' and almost surely takes value outside of $\mathcal{N}$, then it must also be optimal for Λ^+. Clearly, it is much easier to get the optimality equation for such a curve, by working with Λ^+ rather than with Λ'. By varying action Λ^+, we get (6) as optimality equation. Therefore, the optimal curves of our functional Λ' taking value in $\mathcal{N}$ for a negligible set of times solve (4) (in a distributional sense), which is the MAG discrete model announced in the Introduction.

Of course, these equations have to be suitably modified for those curves which are optimal for the action Λ' but not for Λ^+ because they take values in $\mathcal{N}$ for a nonnegligible amount of time. At this stage, we do not know how to do it. However, at least in the one-dimensional case $d=1$, such modifications are tractable and correspond to sticky collisions as $x_i(t)=x_j(t)$ occurs for different "particles" of labels $i\neq j$ and during intervals of times of strictly positive Lebesgue measure; see Section 4.

3. Proof of the Γ-convergence

The purpose of this section is to prove Theorem 9.

3.1. ***The proof as a consequence of three lemmas.*** As we will see, Theorem 9 will be a consequence of three lemmas that we state below. Lemmas 14 and 15 both involve a family of smooth functions $(g_\varepsilon)_{\varepsilon>0}$ on $[\theta_0,\theta_1]\times\mathbb{R}^p$ for some $\theta_0<\theta_1$ and $p\in\mathbb{N}$, pointwise converging to a L^1_{loc} function g. On these functions, we will assume the following:

Assumptions 13. (H1) For all $\varepsilon>0$ and $\theta\in[\theta_0,\theta_1]$, $g_\varepsilon(\theta,0)=0$.

(H2) For all $\varepsilon>0$ and $\theta\in[\theta_0,\theta_1]$, $g_\varepsilon(\theta,\cdot)$ is convex. Therefore, $g(\theta,\cdot)$ is convex as well.

(H3) The maps ∇g_ε are uniformly bounded, that is,

$$L := \sup_{\varepsilon>0} \sup_{\theta\in[\theta_0,\theta_1]} \sup_{Y\in\mathbb{R}^p} |\nabla g_\varepsilon(\theta, Y)| < +\infty. \tag{23}$$

Therefore, we also have

$$\sup_{\theta\in[\theta_0,\theta_1]} \sup_{Y\in\mathbb{R}^p} |\overline{\nabla} g(\theta, Y)| \le L.$$

(H4) The distributional derivative $\partial_\theta g$ is $L^2([\theta_0, \theta_1]; L^\infty_{\mathrm{loc}}(\mathbb{R}^d)^N)$, and, for all $\mathcal{Y} \in H^1([\theta_0, \theta_1]; (\mathbb{R}^d)^N)$, the map $\theta \mapsto g(\theta, \mathcal{Y}_\theta)$ is also H^1, with, for almost all $\theta \in [\theta_0, \theta_1]$,

$$\frac{\mathrm{d}}{\mathrm{d}\theta} g(\theta, \mathcal{Y}_\theta) = \partial_\theta g(\theta, \mathcal{Y}_\theta) + \overline{\nabla} g(\theta, \mathcal{Y}_\theta) \cdot \dot{\mathcal{Y}}_\theta. \tag{24}$$

(H5) The maps $\partial_\theta \nabla g_\varepsilon$ are uniformly bounded, that is,

$$M := \sup_{\varepsilon>0} \sup_{\theta\in[\theta_0,\theta_1]} \sup_{Y\in\mathbb{R}^p} |\partial_\theta \nabla g_\varepsilon(\theta, Y)| < +\infty. \tag{25}$$

In order to keep the proofs simple, we did not try to optimize these assumptions for Lemmas 14 and 15, which are probably true in a far more general context (see Remark 16 in the case of Lemma 15). However, as we will see in the proof of Theorem 9, it suffices to check these assumptions for the family $(f_\varepsilon)_{\varepsilon>0}$ after suitable change of temporal and spatial scale. This is done in Lemma 17.

Lemma 14. *Let us consider $\theta_0 < \theta_1 \in \mathbb{R}$, $\eta \in C^\infty([\theta_0, \theta_1]; \mathbb{R}^*_+)$ and a family $(g_\varepsilon)_{\varepsilon>0}$ of smooth functions from $[\theta_0, \theta_1] \times \mathbb{R}^p$ to $\mathbb{R}$ pointwise converging to a function g, which satisfy (H1), (H3), (H4) and (H5) from Assumptions 13. If a family of curves $(\mathcal{Y}^\varepsilon)_{\varepsilon>0}$ in $H^1([\theta_0, \theta_1]; \mathbb{R}^p)$ uniformly converges to a curve $\mathcal{Y} \in H^1([\theta_0, \theta_1]; \mathbb{R}^p)$, then*

$$\int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}^\varepsilon_\theta \cdot \nabla g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta)\eta(\theta)\,\mathrm{d}\theta \xrightarrow[\varepsilon\to 0]{} \int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}_\theta \cdot \overline{\nabla} g(\theta, \mathcal{Y}_\theta)\eta(\theta)\,\mathrm{d}\theta.$$

Lemma 15. *Let us consider $\theta_0 < \theta_1 \in \mathbb{R}$, $\eta \in C^\infty([\theta_0, \theta_1]; \mathbb{R}^*_+)$ and a family $(g_\varepsilon)_{\varepsilon>0}$ of smooth functions from $[\theta_0, \theta_1] \times \mathbb{R}^p$ to $\mathbb{R}$ pointwise converging to a function g, and satisfying (H2), (H3)and (H5) from Assumptions 13. Let us fix $R, S \in \mathbb{R}^p$ and define for $\varepsilon > 0$ and $\mathcal{Y} \in C^0([\theta_0, \theta_1]; \mathbb{R}^p)$*

$$K_\varepsilon(\mathcal{Y}) := \begin{cases} \frac{1}{2}\int_{\theta_0}^{\theta_1} \big\{|\dot{\mathcal{Y}}_\theta|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Y}_\theta)|^2\big\}\eta(\theta)\,\mathrm{d}\theta & \text{if } \mathcal{Y} \in H^1([\theta_1, \theta_1]; \mathbb{R}^p), \mathcal{Y}_{\theta_0} = R \text{ and } \mathcal{Y}_{\theta_1} = S, \\ +\infty & \text{else,} \end{cases}$$

$$K(\mathcal{Y}) := \begin{cases} \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta|^2 + |\overline{\nabla} g(\theta, \mathcal{Y}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta & \text{if } \mathcal{Y} \in H^1([\theta_1, \theta_1]; \mathbb{R}^p), \mathcal{Y}_{\theta_0} = R \text{ and } \mathcal{Y}_{\theta_1} = S, \\ +\infty, & \text{else.} \end{cases}$$

Then $(K_\varepsilon)_{\varepsilon>0}$ Γ-converges to K for the topology of uniform convergence of $C^0([\theta_0, \theta_1]; \mathbb{R}^p)$.

Remark 16. This lemma is the keystone of the proof, and one may wonder how it can be generalized and what is really necessary among our assumptions. In [Ambrosio et al. 2021], we show that at least when (g_ε) and g have no dependence on θ and $\eta \equiv 1$, the result holds true, even in Hilbert spaces, whenever (g_ε) is a family of proper lower semicontinuous uniformly λ-convex functions Mosco converging towards g, plus some uniform Lipschitz conditions at the extreme points.

Lemma 17. *With the notation of Theorem 9, let us define* $\theta_0 := \log t_0/2$, $\theta_1 := \log t_1/2$, $p = dN$, *and for* $\theta \in [\theta_0, \theta_1]$, $\varepsilon > 0$ *and* $Y \in (\mathbb{R}^d)^N$,

$$g_\varepsilon(\theta, Y) := \frac{f_\varepsilon(\exp(2\theta), \exp(\theta)Y)}{\exp(2\theta)} \quad \textit{and} \quad g(\theta, Y) := \frac{f(\exp(\theta)Y)}{\exp(2\theta)}. \tag{26}$$

Then $(g_\varepsilon)_{\varepsilon>0}$ *converges pointwise to* g, *and satisfies* (H1), (H2), (H3), (H4) *and* (H5) *from Assumptions 13.*

In the next subsections, we will prove these three lemmas one by one. The most involved one is undoubtedly Lemma 15, which can be seen as the main step in the proof of Theorem 9. Let us start by proving Theorem 9 using Lemmas 14, 15 and 17.

Proof of Theorem 9. In this proof, the notation $\mathcal{X} = \mathcal{X}_t$ will stand for a generic curve from $[t_0, t_1]$ to $(\mathbb{R}^d)^N$. Associated with $\mathcal{X}$, we define by $\mathcal{Y} = \mathcal{Y}_\theta$ the curve from $[\theta_0, \theta_1]$ to $(\mathbb{R}^d)^N$, where $\theta_0 := \log t_0/2$, $\theta_1 := \log t_1/2$, and, for all $\theta \in [\theta_0, \theta_1]$, $\mathcal{Y}_\theta := \mathcal{X}_{\exp(2\theta)}/\exp(\theta)$. Note that $\mathcal{X}$ is H^1 if and only if $\mathcal{Y}$ is H^1. If $(\mathcal{X}^\varepsilon)_{\varepsilon>0}$ is a family of curves from $[t_0, t_1]$ to $(\mathbb{R}^d)^N$, we define in the same way the family of corresponding curves $(\mathcal{Y}^\varepsilon)_{\varepsilon>0}$ from $[\theta_0, \theta_1]$ to $(\mathbb{R}^d)^N$.

A quick computation shows that, for all $\mathcal{X} \in H^1([\theta_0, \theta_1]; (\mathbb{R}^d)^N)$, considering $\eta(\theta) := \exp(2\theta)$ and $(g_\varepsilon)_{\varepsilon>0}$, g as defined in Lemma 17, we have

$$L_\varepsilon(\mathcal{X}) = \frac{1}{2}\int_{t_0}^{t_1} \left| \dot{\mathcal{X}}_t - \frac{\mathcal{X}_t - \nabla f_\varepsilon(t, \mathcal{X}_t)}{2t} \right|^2 \frac{\mathrm{d}t}{t} = \frac{1}{4}\int_{\theta_0}^{\theta_1} |\dot{\mathcal{Y}}_\theta + \nabla g_\varepsilon(\theta, \mathcal{Y}_\theta)|^2 \eta(\theta)\,\mathrm{d}\theta \tag{27}$$

$$= \frac{1}{4}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Y}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta + \frac{1}{2}\int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}_\theta \cdot \nabla g_\varepsilon(\theta, \mathcal{Y}_\theta)\eta(\theta)\,\mathrm{d}\theta \tag{28}$$

and

$$L(\mathcal{X}) = \frac{1}{2}\int_{t_0}^{t_1} \left| \dot{\mathcal{X}}_t - \frac{\mathcal{X}_t - \overline{\nabla} f(\mathcal{X}_t)}{2t} \right|^2 \frac{\mathrm{d}t}{t} = \frac{1}{4}\int_{\theta_0}^{\theta_1} |\dot{\mathcal{Y}}_\theta + \overline{\nabla} g(\theta, \mathcal{Y}_\theta)|^2 \eta(\theta)\,\mathrm{d}\theta$$

$$= \frac{1}{4}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta|^2 + |\overline{\nabla} g(\theta, \mathcal{Y}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta + \frac{1}{2}\int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}_\theta \cdot \overline{\nabla} g(\theta, \mathcal{Y}_\theta)\eta(\theta)\,\mathrm{d}\theta. \tag{29}$$

(Note that due to Lemma 17, g is convex with respect to the space variable, and so $\overline{\nabla} g$ is well-defined.)

<u>Proof of the Γ-lim inf</u>: Let $\mathcal{X}^\varepsilon \xrightarrow[\varepsilon\to 0]{} \mathcal{X}$ for the topology of uniform convergence. Of course, we also have $\mathcal{Y}^\varepsilon \xrightarrow[\varepsilon\to 0]{} \mathcal{Y}$. Without loss of generality, we can suppose

$$\liminf_{\varepsilon\to 0} L_\varepsilon(\mathcal{X}^\varepsilon) < +\infty.$$

Indeed, if it is not the case, there is nothing to prove. Let us take $(\varepsilon_n)_{n\in\mathbb{N}}$ to be a sequence tending to 0 along which the lim inf is achieved.

As $\nabla g_\varepsilon(\theta, Y)$ is bounded uniformly in ε, θ, Y (this is (H3)), we easily deduce with (27)

$$\limsup_{n\to+\infty} \int_{\theta_0}^{\theta_1} |\dot{\mathcal{Y}}_\theta^{\varepsilon_n}|^2\,\mathrm{d}\theta < +\infty.$$

In particular, by the lower semicontinuity of this H^1 seminorm with respect to uniform convergence, $\mathcal{Y}$ is in $H^1([\theta_0, \theta_1]; (\mathbb{R}^d)^N)$. Applying Lemma 14, thanks to Lemma 17, we have

$$\int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}_\theta^{\varepsilon_n} \cdot \nabla g_{\varepsilon_n}(\theta, \mathcal{Y}_\theta^{\varepsilon_n})\eta(\theta)\,\mathrm{d}\theta \xrightarrow[n\to+\infty]{} \int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}_\theta \cdot \overline{\nabla} g(\theta, \mathcal{Y}_\theta)\eta(\theta)\,\mathrm{d}\theta. \tag{30}$$

On the other hand, for n sufficiently large, $L_{\varepsilon_n}(\mathcal{X}^{\varepsilon_n}) < +\infty$. So the endpoints of $\mathcal{X}^{\varepsilon_n}$ belong to a finite set, and because of the convergence $\mathcal{X}^{\varepsilon_n} \to \mathcal{X}$, for even larger n the endpoints of $\mathcal{X}^{\varepsilon_n}$ are independent of n. In other terms, $\mathcal{X}^{\varepsilon_n}_{t_0} = P^{\sigma_0}$ and $\mathcal{X}^{\varepsilon_n}_{t_1} = Q^{\sigma_1}$ with σ_0, σ_1 independent of n. Hence, for such n, $\mathcal{Y}^{\varepsilon_n}$ satisfies the endpoint constraint for K_{ε_n} with $R := P^{\sigma_0}/\sqrt{t_0}$ and $S := Q^{\sigma_1}/\sqrt{t_1}$. Hence, applying Lemma 15 thanks to Lemma 17, we have

$$\begin{aligned}\frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta|^2 + |\overline{\nabla} g(\theta, \mathcal{Y}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta = K(\mathcal{Y}) \le \liminf_{n\to+\infty} K_{\varepsilon_n}(\mathcal{Y}^{\varepsilon_n}) \\ = \liminf_{n\to+\infty} \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}^{\varepsilon_n}_\theta|^2 + |\nabla g_{\varepsilon_n}(\theta, \mathcal{Y}^{\varepsilon_n}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta. \qquad (31)\end{aligned}$$

The result follows easily by gathering (28), (30), (31) and (29).

Proof of the Γ-lim sup: Let $\mathcal{X} \in C^0([t_0, t_1]; (\mathbb{R}^d)^N)$. Without loss of generality, we can suppose that $\mathcal{X} \in H^1([t_0, t_1]; (\mathbb{R}^d)^N)$ and that it satisfies the endpoint constraint for L. In particular, $\mathcal{Y}$ belongs to $H^1([\theta_0, \theta_1]; (\mathbb{R}^d)^N)$ and satisfies the endpoint constraint for K with $R := \mathcal{X}_{t_0}/\sqrt{t_0}$ and $S := \mathcal{X}_{t_0}/\sqrt{t_1}$. Lemmas 15 and 17 let us find a family $(\mathcal{Y}^\varepsilon)_{\varepsilon>0}$ converging to the corresponding $\mathcal{Y}$ such that

$$\limsup_{\varepsilon\to 0} K_\varepsilon(\mathcal{Y}^\varepsilon) \le K(\mathcal{Y}). \qquad (32)$$

In particular $\mathcal{Y}^\varepsilon$ is in H^1 for sufficiently small ε, and by Lemmas 14 and 17,

$$\int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}^\varepsilon_\theta \cdot \nabla g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta)\eta(\theta)\,\mathrm{d}\theta \xrightarrow[\varepsilon\to 0]{} \int_{\theta_0}^{\theta_1} \dot{\mathcal{Y}}_\theta \cdot \overline{\nabla} g(\theta, \mathcal{Y}_\theta)\eta(\theta)\,\mathrm{d}\theta. \qquad (33)$$

The result follows easily from (28), (32), (33) and (29), by noticing, that because of (32), $\mathcal{Y}^\varepsilon$ satisfies the endpoint constraint for K_ε. Hence, for such ε, $\mathcal{X}^\varepsilon$ satisfies the endpoint constraint for L_ε. □

3.2. *Proof of Lemma 14.* The proof of Lemma 14 just consists in integrating by parts and using the convergence properties of $(g_\varepsilon)_{\varepsilon>0}$.

Proof of Lemma 14. Integration by parts: First, notice that as soon as $\mathcal{Y} \in H^1([\theta_0, \theta_1]; \mathbb{R}^p)$ and $\varepsilon > 0$, then $\theta \mapsto g_\varepsilon(\theta, \mathcal{Y}_\theta)$ and $\theta \mapsto g(\theta, \mathcal{Y}_\theta)$ are also in H^1, with, for almost every θ,

$$\frac{\mathrm{d}}{\mathrm{d}\theta} g_\varepsilon(\theta, \mathcal{Y}_\theta) = \partial_\theta g_\varepsilon(\theta, \mathcal{Y}_\theta) + \nabla g_\varepsilon(\theta, \mathcal{Y}_\theta)\cdot\dot{\mathcal{Y}}_\theta \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}\theta} g(\theta, \mathcal{Y}_\theta) = \partial_\theta g(\theta, \mathcal{Y}_\theta) + \overline{\nabla} g(\theta, \mathcal{Y}_\theta)\cdot\dot{\mathcal{Y}}_\theta.$$

It is clear in the case of g_ε because g_ε is smooth, and it is the assumption (H4) in the case of g. As a consequence, by an integration by parts, it suffices to prove that whenever $(\mathcal{Y}^\varepsilon)_{\varepsilon>0}$ converges to $\mathcal{Y}$ as $\varepsilon \to 0$ for the topology of uniform convergence,

$$\begin{aligned}g_\varepsilon(\theta_1, \mathcal{Y}^\varepsilon_{\theta_1})\eta(\theta_1) - g_\varepsilon(\theta_0, \mathcal{Y}^\varepsilon_{\theta_0})\eta(\theta_0) - \int_{\theta_0}^{\theta_1} g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta)\eta'(\theta)\,\mathrm{d}\theta - \int_{\theta_0}^{\theta_1} \partial_\theta g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta)\eta(\theta)\,\mathrm{d}\theta \\ \xrightarrow[\varepsilon\to 0]{} g(\theta_1, \mathcal{Y}_{\theta_1})\eta(\theta_1) - g(\theta_0, \mathcal{Y}_{\theta_0})\eta(\theta_0) - \int_{\theta_0}^{\theta_1} g(\theta, \mathcal{Y}_\theta)\eta'(\theta)\,\mathrm{d}\theta - \int_{\theta_0}^{\theta_1} \partial_\theta g(\theta, \mathcal{Y}_\theta)\eta(\theta)\,\mathrm{d}\theta.\end{aligned}$$

Convergence term by term: The convergence

$$g_\varepsilon(\theta_1, \mathcal{Y}^\varepsilon_{\theta_1})\eta(\theta_1) - g_\varepsilon(\theta_0, \mathcal{Y}^\varepsilon_{\theta_0})\eta(\theta_0) \xrightarrow[\varepsilon\to 0]{} g(\theta_1, \mathcal{Y}_{\theta_1})\eta(\theta_1) - g(\theta_0, \mathcal{Y}_{\theta_0})\eta(\theta_0)$$

is an easy consequence of the pointwise convergence and of the uniform Lipschitz bound (H3).

For the same reason, we have, for all $\theta \in [\theta_0, \theta_1]$, $g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta) \xrightarrow[\varepsilon\to 0]{} g(\theta, \mathcal{Y}_\theta)$. But on the other hand, because of (H1) and (H3), g_ε is locally bounded, uniformly in ε. Hence,

$$\int_{\theta_0}^{\theta_1} g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta)\eta'(\theta)\,\mathrm{d}\theta \xrightarrow[\varepsilon\to 0]{} \int_{\theta_0}^{\theta_1} g(\theta, \mathcal{Y}_\theta)\eta'(\theta)\,\mathrm{d}\theta$$

is a consequence of the dominated convergence theorem.

Because of (H1) and (H5), for all θ, $(\partial_\theta g_\varepsilon(\theta, \cdot))_{\varepsilon>0}$ is compact for the topology of local uniform convergence. But its only possible limit point is the distributional derivative $\partial_\theta g$. As a consequence, $(\partial_\theta g_\varepsilon)_{\varepsilon>0}$ converges pointwise to $\partial_\theta g$, and because of the uniform bound (H5), for all θ, $\partial_\theta g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta) \xrightarrow[\varepsilon\to 0]{} \partial_\theta g(\theta, \mathcal{Y}_\theta)$. Because of (H1) and (H5), $\partial_\theta g_\varepsilon$ is locally bounded, uniformly in ε, and so

$$\int_{\theta_0}^{\theta_1} \partial_\theta g_\varepsilon(\theta, \mathcal{Y}^\varepsilon_\theta)\eta(\theta)\,\mathrm{d}\theta \xrightarrow[\varepsilon\to 0]{} \int_{\theta_0}^{\theta_1} \partial_\theta g(\theta, \mathcal{Y}_\theta)\eta(\theta)\,\mathrm{d}\theta$$

is also a consequence of the dominated convergence theorem. □

3.3. *Proof of Lemma 15.* Before entering the proof of Lemma 15, we need to state a few standard results concerning the extended gradient $\overline{\nabla}$ as defined in Definition 8, and its links with the so-called *resolvent map*. These tools could even be set in the infinite-dimensional setting, that is, in Hilbert spaces [Strömberg 1996], or in metric spaces [Ambrosio et al. 2005].

The following proposition is a lower semicontinuity property of the slope with respect to both convergence of the function and of the evaluation point.

Proposition 18. *Consider $h : \mathbb{R}^p \to \mathbb{R}$ a convex function with finite values. Let $(h_\varepsilon)_{\varepsilon>0}$ be a family of convex functions on $\mathbb{R}^p$ pointwise converging to h, and let $(X^\varepsilon)_{\varepsilon>0}$ be a family of points in $\mathbb{R}^p$ converging to X. Then*

$$|\overline{\nabla} h(X)| \le \liminf_{\varepsilon\to 0} |\overline{\nabla} h_\varepsilon(X^\varepsilon)|.$$

Proof. As all these functions are convex and h has finite values, standard arguments show that the convergence of $h_\varepsilon \to h$ is also locally uniform. First of all, if

$$\liminf_{\varepsilon\to 0} |\overline{\nabla} h_\varepsilon(X^\varepsilon)| = +\infty,$$

there is nothing to prove. Else, up to considering a subsequence, there exists $D \in \mathbb{R}^p$ such that

$$\lim_{\varepsilon\to 0} \overline{\nabla} h_\varepsilon(X^\varepsilon) = D.$$

But sending $\varepsilon \to 0$ in the inequality,

$$\text{for all } Y \in \mathbb{R}^p, \quad h_\varepsilon(Y) \ge h_\varepsilon(X^\varepsilon) + \langle \overline{\nabla} h_\varepsilon(X^\varepsilon), Y - X^\varepsilon\rangle,$$

and using the local uniformity of the convergence, we see that $D \in \partial h(X)$ (that is, the subdifferential is upper semicontinuous). So $|D| \ge |\overline{\nabla} h(X)|$, and the result follows. □

For $\tau > 0$ and $X \in \mathbb{R}^p$, define the resolvent operator by

$$J_{\tau,h}(X) := \operatorname*{argmin}_{Y \in \mathbb{R}^p} h(Y) + \frac{|Y-X|^2}{2\tau}.$$

Once again, the following proposition is standard. It is an application in the very simple case of convex functions in finite dimension of the so-called maximal monotone operators theory in Hilbert spaces, for which we refer for instance to [Brézis 1973] (see in particular Section 2.4 for the properties of the resolvent in a general setting).

Proposition 19. (1) *We have for all $X \in \mathbb{R}^p$ and $\tau > 0$,*

$$|\overline{\nabla} h(J_{\tau,h}(X))| \le \left| \frac{X - J_{\tau,h}(X)}{\tau} \right| \le |\overline{\nabla} h(X)|. \tag{34}$$

(2) *If h is differentiable at $J_{\tau,h}(X)$ for some $X \in \mathbb{R}^p$, then the following first-order condition holds:*

$$\frac{X - J_{\tau,h}(X)}{\tau} = \nabla h(J_{\tau,h}(X)).$$

(3) *If $(h_\varepsilon)_{\varepsilon>0}$ is a family of convex functions on $\mathbb{R}^p$ pointwise converging to h, then, for all $\tau > 0$ and $X \in \mathbb{R}^p$,*

$$J_{\tau,h_\varepsilon}(X) \xrightarrow[\varepsilon \to 0]{} J_{\tau,h}(X). \tag{35}$$

Proof. By [Brézis 1973, Lemma 2.1], we have

$$\frac{X - J_{\tau,h}(X)}{\tau} \in \partial h(J_{\tau,h}(X)). \tag{36}$$

The first inequality in (34) and the second point of the statement follow.

To get the second inequality in (34), apply the monotone inequality of [Brézis 1973, Definition 2.1] to the maximal monotone operator ∂h (see [Brézis 1973, Example 2.1.4]), with $x_1 = X$, $x_2 = J_{\tau,h}(X)$, $y_1 = \overline{\nabla} h(X) \in \partial h(X)$ and $(X - J_{\tau,h}(X))/\tau \in \partial h(J_{\tau,h}(X))$, thanks to (36). We find

$$\left\langle \overline{\nabla} h(X) - \frac{X - J_{\tau,h}(X)}{\tau}, X - J_{\tau,h}(X) \right\rangle \ge 0,$$

which can be rewritten as

$$\left| \frac{X - J_{\tau,h}(X)}{\tau} \right|^2 \le \left\langle \frac{X - J_{\tau,h}(X)}{\tau}, \overline{\nabla} h(X) \right\rangle.$$

Therefore, the result follows from the Cauchy–Schwarz inequality.

Let us now focus on the third point. Let us fix $\tau > 0$ and $X \in \mathbb{R}^p$, and set,

$$\text{for all } \varepsilon > 0,\ Y \in \mathbb{R}^p, \quad f_\varepsilon(Y) := h_\varepsilon(Y) + \frac{|Y-X|^2}{2\tau} \quad \text{and} \quad f(Y) := h(Y) + \frac{|Y-X|^2}{2\tau}.$$

The family $(f_\varepsilon)_{\varepsilon>0}$ converges pointwise to f, but by convexity and finiteness of the limit, as before, this convergence is also locally uniform. As a consequence, the only thing to prove is that for sufficiently small

$\varepsilon_0 > 0$, the set $\{J_{\tau,h_\varepsilon}(X) : 0 < \varepsilon \le \varepsilon_0\}$ is bounded. Indeed, if it is the case, by local uniform convergence, any limit point Z of $J_{\tau,h_\varepsilon}(X)$ as ε tends to 0 would satisfy

$$f(Z) \le \limsup_{\varepsilon\to 0} f_\varepsilon(J_{\tau,h_\varepsilon}(X)) \le \lim_{\varepsilon\to 0} f_\varepsilon(J_{\tau,h}(X)) = f(J_{\tau,h}(X)),$$

so that, by the definition of $J_{\tau,h}(X)$, $Z = J_{\tau,h}(X)$, which lets us conclude.

Call B the open ball of center $J_{\tau,h}(X)$ and radius 1. We have by the strict convexity of f and minimality of $J_{\tau,h}(X)$

$$f(J_{\tau,h}(X)) < \inf_{Y\in\partial B} f(Y),$$

and this property is open for the topology of local uniform convergence. Hence, we can find ε_0 sufficiently small so that for all $\varepsilon \le \varepsilon_0$

$$f_\varepsilon(J_{\tau,h}(X)) < \inf_{Y\in\partial B} f_\varepsilon(Y).$$

Then, if $Y \notin B$, we call $\overline{Y}$ the projection of Y on ∂B and $\lambda := 1/|Y - J_{\tau,h}(X)| \le 1$, so that $\overline{Y} = (1-\lambda)J_{\tau,h}(X) + \lambda Y$. As soon as $\varepsilon \le \varepsilon_0$, $f_\varepsilon(\overline{Y}) > f_\varepsilon(J_{\tau,h}(X))$. By using the convexity inequality

$$f_\varepsilon(\overline{Y}) \le (1-\lambda) f_\varepsilon(J_{\tau,h}(X)) + \lambda f_\varepsilon(Y),$$

we find $f_\varepsilon(Y) > f_\varepsilon(J_{\tau,h}(X))$. As a consequence, $\{J_{\tau,h_\varepsilon}(X) : 0 < \varepsilon \le \varepsilon_0\} \subset B$ and the result follows. $\square$

We are now ready for the proof of Lemma 15.

Proof of Lemma 15. Proof of the Γ-$\liminf$: It is straightforward using Fatou's lemma, Proposition 18 and the lower semicontinuity of $\mathcal{Y} \mapsto \int_{\theta_0}^{\theta_1} |\dot{\mathcal{Y}}_\theta|^2 \,\mathrm{d}\theta$ with respect to the topology of uniform convergence.

Proof of the Γ-$\limsup$: Let us consider a curve $\mathcal{Y} \in H^1([\theta_0, \theta_1]; \mathbb{R}^p)$ with $\mathcal{Y}_{\theta_0} = R$ and $\mathcal{Y}_{\theta_1} = S$ (else there is nothing to prove). For all $\varepsilon > 0$ and $\tau > 0$, we define

$$\mathcal{Y}^{\tau,\varepsilon} : \theta \mapsto J_{\tau, g_\varepsilon(\theta,\cdot)}(\mathcal{Y}_\theta),$$

and correspondingly

$$\mathcal{Y}^{\tau} : \theta \mapsto J_{\tau, g(\theta,\cdot)}(\mathcal{Y}_\theta).$$

First, we prove

$$\limsup_{\tau\to 0}\limsup_{\varepsilon\to 0} \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}^{\tau,\varepsilon}_\theta|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Y}^{\tau,\varepsilon}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta \le \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta|^2 + |\overline{\nabla} g(\theta, \mathcal{Y}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta. \quad (37)$$

We will then choose τ as a function of ε and show how to fix the endpoints.

Proof of (37): By the second point of Proposition 19, for all $\varepsilon, \tau, \theta$, we have

$$\mathcal{Y}_\theta = \mathcal{Y}^{\tau,\varepsilon}_\theta + \tau\nabla g_\varepsilon(\theta, \mathcal{Y}^{\tau,\varepsilon}_\theta).$$

For all θ, $g_\varepsilon(\theta,\cdot)$ is convex, so $Y \mapsto Y + \nabla g_\varepsilon(\theta, Y)$ is invertible and its inverse is 1-Lipschitz. In addition, the smoothness of $g_\varepsilon = g_\varepsilon(\theta, Y)$ with respect to θ lets us deduce from $\mathcal{Y} \in H^1$ that $\mathcal{Y}^{\tau,\varepsilon}$ is in H^1, and that, for almost all θ,

$$\dot{\mathcal{Y}}_\theta = (\mathbb{I} + \tau \mathrm{D}^2 g_\varepsilon(\theta, \mathcal{Y}^{\tau,\varepsilon}_\theta)) \cdot \dot{\mathcal{Y}}^{\tau,\varepsilon}_\theta + \tau\partial_\theta \nabla g_\varepsilon(\theta, \mathcal{Y}^{\tau,\varepsilon}_\theta).$$

By the convexity of g_ε, we have $\mathbb{I} \le \mathbb{I} + \tau \mathrm{D}^2 g_\varepsilon$ in the sense of symmetric matrices, and hence

$$|\dot{\mathcal{Y}}_\theta^{\tau,\varepsilon}| \le |\dot{\mathcal{Y}}_\theta - \tau \partial_\theta \nabla g_\varepsilon(\theta, \mathcal{Y}_\theta^{\tau,\varepsilon})| \le |\dot{\mathcal{Y}}_\theta| + \tau M. \tag{38}$$

Recall that M was defined in the uniform integrability assumption (25) on $\partial_\theta \nabla g_\varepsilon$. (In the case when $\partial_\theta \nabla g_\varepsilon = 0$, we recover the known fact that for h independent of time, $J_{\tau,h}$ is contractive.) Then, we deduce

$$\begin{aligned}\limsup_{\varepsilon\to 0} \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta^{\tau,\varepsilon}|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Y}_\theta^{\tau,\varepsilon})|^2\}\eta(\theta)\,\mathrm{d}\theta &\overset{(34),(38)}{\le} \limsup_{\varepsilon\to 0} \frac{1}{2}\int_{\theta_0}^{\theta_1} \left\{(|\dot{\mathcal{Y}}_\theta| + \tau M)^2 + \left|\frac{\mathcal{Y}_\theta - \mathcal{Y}_\theta^{\tau,\varepsilon}}{\tau}\right|^2\right\}\eta(\theta)\,\mathrm{d}\theta \\ &\overset{(35)}{\le} \frac{1}{2}\int_{\theta_0}^{\theta_1} \left\{(|\dot{\mathcal{Y}}_\theta| + \tau M)^2 + \left|\frac{\mathcal{Y}_\theta - \mathcal{Y}_\theta^{\tau}}{\tau}\right|^2\right\}\eta(\theta)\,\mathrm{d}\theta \\ &\overset{(34)}{\le} \frac{1}{2}\int_{\theta_0}^{\theta_1} \{(|\dot{\mathcal{Y}}_\theta| + \tau M)^2 + |\overline{\nabla} g(\theta, \mathcal{Y}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta.\end{aligned}$$

Formula (37) follows.

Choice of $\tau = \tau(\varepsilon)$: Because of (37), and because,

$$\text{for all } \varepsilon > 0, \quad \mathcal{Y}_{\theta_0}^{\tau,\varepsilon} \xrightarrow[\tau\to 0]{} R \quad \text{and} \quad \mathcal{Y}_{\theta_1}^{\tau,\varepsilon} \xrightarrow[\tau\to 0]{} S,$$

it is possible to find a nonincreasing function $\tau = \tau(\varepsilon)$ converging sufficiently slowly to 0 so that

$$\limsup_{\varepsilon\to 0} \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta^{\tau(\varepsilon),\varepsilon}|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Y}_\theta^{\tau(\varepsilon),\varepsilon})|^2\}\eta(\theta)\,\mathrm{d}\theta \le \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta|^2 + |\overline{\nabla} g(\theta, \mathcal{Y}_\theta)|^2\}\eta(\theta)\,\mathrm{d}\theta, \tag{39}$$

$$\mathcal{Y}_{\theta_0}^{\tau(\varepsilon),\varepsilon} \xrightarrow[\varepsilon\to 0]{} R \quad \text{and} \quad \mathcal{Y}_{\theta_1}^{\tau(\varepsilon),\varepsilon} \xrightarrow[\tau\to 0]{} S. \tag{40}$$

Fixing the endpoints: For fixed ε and small $\delta > 0$, we will define $\mathcal{Z}^{\delta,\varepsilon}$ as a slight modification of the curve $\mathcal{Y}^{\varepsilon,\tau(\varepsilon)}$ in such a way that $\mathcal{Z}^{\delta,\varepsilon}$ joins R to S. For this, we just set for $\theta \in [\theta_0, \theta_1]$

$$\mathcal{Z}_\theta^{\delta,\varepsilon} = \begin{cases} R + ((\theta - \theta_0)/\delta)(\mathcal{Y}_{\delta_0+\delta}^{\tau(\varepsilon),\varepsilon} - R) & \text{if } \theta \in [\theta_0, \theta_0 + \delta], \\ \mathcal{Y}_\theta^{\tau(\varepsilon),\varepsilon} & \text{if } \theta \in [\theta_0 + \delta, \theta_1 - \delta], \\ S + ((\theta_1 - \theta)/\delta)(\mathcal{Y}_{\delta_1-\delta}^{\tau(\varepsilon),\varepsilon} - S) & \text{if } \theta \in [\theta_1 - \delta, \theta_1]. \end{cases}$$

A quick computation shows

$$\begin{aligned}&\frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Z}}_\theta^{\delta,\varepsilon}|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Z}_\theta^{\delta,\varepsilon})|^2\}\eta(\theta)\,\mathrm{d}\theta \\ &\le \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta^{\tau(\varepsilon),\varepsilon}|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Y}_\theta^{\tau(\varepsilon),\varepsilon})|^2\}\eta(\theta)\,\mathrm{d}\theta + \|\eta\|_\infty \left(\frac{|\mathcal{Y}_{\theta_0+\delta}^{\tau(\varepsilon),\varepsilon} - R|^2}{2\delta} + \frac{|\mathcal{Y}_{\theta_1-\delta}^{\tau(\varepsilon),\varepsilon} - S|^2}{2\delta} + \delta L^2\right),\end{aligned} \tag{41}$$

where L is defined in the uniform Lipschitz assumption (23) for g_ε.

Let us estimate $|\mathcal{Y}_{\theta_0+\delta}^{\tau(\varepsilon),\varepsilon} - R|^2/2\delta$. We have

$$\frac{|\mathcal{Y}_{\theta_0+\delta}^{\tau(\varepsilon),\varepsilon} - R|^2}{2\delta} \le \frac{|\mathcal{Y}_{\theta_0}^{\tau(\varepsilon),\varepsilon} - R|^2}{\delta} + \frac{|\mathcal{Y}_{\theta_0+\delta}^{\tau(\varepsilon),\varepsilon} - \mathcal{Y}_{\theta_0}^{\tau(\varepsilon),\varepsilon}|^2}{\delta} \le \frac{|\mathcal{Y}_{\theta_0}^{\tau(\varepsilon),\varepsilon} - R|^2}{\delta} + \int_{\theta_0}^{\theta_0+\delta} |\dot{\mathcal{Y}}_\theta^{\tau(\varepsilon),\varepsilon}|^2\,\mathrm{d}\theta.$$

Because of (38), (25) and $\mathcal{Y} \in H^1$, the integral $\int_{\theta_0}^{\theta_0+\delta} |\dot{\mathcal{Y}}_\theta^{\tau(\varepsilon),\varepsilon}|^2 \,\mathrm{d}\theta$ tends to 0 as $\delta \to 0$, uniformly in ε: we bound it by a function $v_i = v_i(\delta)$ tending to 0 as $\delta \to 0$. In the same way,

$$\frac{|\mathcal{Y}_{\theta_1-\delta}^{\tau(\varepsilon),\varepsilon} - S|^2}{2\delta} \le \frac{|\mathcal{Y}_{\theta_1}^{\tau(\varepsilon),\varepsilon} - S|^2}{\delta} + v_f(\delta),$$

where $v_f(\delta) \to 0$ as $\delta \to 0$.

Plugging these bounds into (41), we get

$$\frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Z}}_\theta^{\delta,\varepsilon}|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Z}_\theta^{\delta,\varepsilon})|^2\}\eta(\theta)\,\mathrm{d}\theta \le \frac{1}{2}\int_{\theta_0}^{\theta_1} \{|\dot{\mathcal{Y}}_\theta^{\tau(\varepsilon),\varepsilon}|^2 + |\nabla g_\varepsilon(\theta, \mathcal{Y}_\theta^{\tau(\varepsilon),\varepsilon})|^2\}\eta(\theta)\,\mathrm{d}\theta + \|\eta\|_\infty\left(\frac{u(\varepsilon)}{\delta} + v(\delta)\right),$$

where $u(\varepsilon) := |\mathcal{Y}_{\theta_0}^{\tau(\varepsilon),\varepsilon} - R|^2 + |\mathcal{Y}_{\theta_1}^{\tau(\varepsilon),\varepsilon} - S|^2 \to 0$ as $\varepsilon \to 0$ by (40), and $v(\delta) := v_i(\delta) + v_f(\delta) + \delta L^2 \to 0$ as $\delta \to 0$. Hence, choosing $\delta(\varepsilon) := \sqrt{u(\varepsilon)}$, we find with the help of (39) that $\mathcal{Z}^{\delta(\varepsilon),\varepsilon}$ is a recovery sequence for the $\Gamma - \limsup$ of K_ε towards K. □

3.4. *Proof of Lemma 17.*

The proof is straightforward, and relies on explicit computations.

Proof of Lemma 17. Let us define for $X \in (\mathbb{R}^d)^N$

$$h(X) := \log\left[\frac{1}{N!}\sum_{\sigma \in \mathfrak{S}_N} \exp(X \cdot A^\sigma)\right]. \tag{42}$$

For $\varepsilon > 0$, $\theta \in [\theta_0, \theta_1]$ and $Y \in (\mathbb{R}^d)^N$, we have by the definition of f_ε and g_ε (formulas (19) and (26) respectively)

$$g_\varepsilon(\theta, Y) = \varepsilon h\left(\frac{Y}{\varepsilon \exp(\theta)}\right). \tag{43}$$

Proof of (H1): It is obvious.

Proof of (H2): By (43), it suffices to check that h is convex. Differentiating (42) twice, we get for all $X \in (\mathbb{R}^d)^N$

$$\mathrm{D}^2 h(X) = \langle A^\sigma \otimes A^\sigma\rangle_X - \langle A^\sigma\rangle_X \otimes \langle A^\sigma\rangle_X = \langle A^\sigma - \langle A^\sigma\rangle_X\rangle_X \otimes \langle A^\sigma - \langle A^\sigma\rangle_X\rangle_X, \tag{44}$$

where if a is a function of σ, then $\langle a(\sigma)\rangle_X$ stands for

$$\langle a(\sigma)\rangle_X := \frac{\sum_{\sigma\in\mathfrak{S}_N} a(\sigma)\exp(X \cdot A^\sigma)}{\sum_{\sigma\in\mathfrak{S}_N} \exp(X \cdot A^\sigma)}.$$

It follows that $\mathrm{D}^2 h(X)$ is a nonnegative symmetric matrix.

Proof of (H3): In view of (43) and as $\theta_0 > -\infty$, it suffices to check that ∇h is bounded. Differentiating (42) at $X \in (\mathbb{R}^d)^N$ leads to

$$\nabla h(X) = \langle A^\sigma\rangle_X,$$

which is clearly bounded by $|A|$.

Proof of (H4): By the definitions (20) of f and (26) of g, we have for all $\theta \in [\theta_0, \theta_1]$ and $Y \in (\mathbb{R}^d)^N$

$$g(\theta, Y) = \frac{f(Y)}{\exp(\theta)}.$$

The integrability property of $\partial_\theta g$ is clear; let us check (24). Let us consider $\mathcal{Y} \in H^1([\theta_0, \theta_1]; (\mathbb{R}^d)^N)$. The function g is locally Lipschitz both in θ and Y. As a consequence, the map $G : \theta \mapsto g(\theta, \mathcal{Y}_\theta)$ is also H^1.

Now, instead of proving (24), we will prove that for all curves $\mathcal{D} = \mathcal{D}_\theta$ such that, for almost all $\theta \in [\theta_0, \theta_1]$, $\mathcal{D}_\theta$ belongs to the subdifferential of $g(\theta, \cdot)$ at $Y = \mathcal{Y}_\theta$, we have for almost all $\theta \in [\theta_0, \theta_1]$

$$\frac{\mathrm{d}}{\mathrm{d}\theta} g(\theta, \mathcal{Y}_\theta) = \partial_\theta g(\theta, \mathcal{Y}_\theta) + \mathcal{D}_\theta \cdot \dot{\mathcal{Y}}_\theta,$$

so that (24) is an application of this property to $\mathcal{D}_\theta := \overline{\nabla} g(\theta, \mathcal{Y}_\theta)$. Notice that this property implies that up to negligible sets, $\mathcal{D}_\theta \cdot \dot{\mathcal{Y}}_\theta$ does not depend on the choice of $\mathcal{D}$. Let us give ourselves such a curve $\mathcal{D}$.

Let us take a point $\theta \in (\theta_0, \theta_1)$ where both $\mathcal{Y}$ and G are differentiable (this happens for almost every θ). We have

$$\begin{aligned} G'(\theta) &= \lim_{\delta \downarrow 0} \frac{1}{\delta} \left\{ \frac{f(\mathcal{Y}_{\theta+\delta})}{\exp(\theta+\delta)} - \frac{f(\mathcal{Y}_\theta)}{\exp(\theta)} \right\} = -\frac{f(\mathcal{Y}_\theta)}{\exp(\theta)} + \lim_{\delta \downarrow 0} \frac{g(\theta, \mathcal{Y}_{\theta+\delta}) - g(\theta, \mathcal{Y}_\theta)}{\delta} \\ &\geq -\frac{f(\mathcal{Y}_\theta)}{\exp(\theta)} + \limsup_{\delta \downarrow 0} \mathcal{D}_\theta \cdot \frac{\mathcal{Y}_{\theta+\delta} - \mathcal{Y}_\theta}{\delta} = \partial_\theta g(\theta, \mathcal{Y}_\theta) + \mathcal{D}_\theta \cdot \dot{\mathcal{Y}}_\theta, \end{aligned}$$

where we used $g(\theta, \mathcal{Y}_{\theta+\delta}) \geq g(\theta, \mathcal{Y}_\theta) + \mathcal{D}_\theta \cdot (\mathcal{Y}_{\theta+\delta} - \mathcal{Y}_\theta)$ to get the second line.

In the same way, we have

$$G'(\theta) = \lim_{\delta \downarrow 0} \frac{1}{\delta} \left\{ \frac{f(\mathcal{Y}_\theta)}{\exp(\theta)} - \frac{f(\mathcal{Y}_{\theta-\delta})}{\exp(\theta-\delta)} \right\} \leq \partial_\theta g(\theta, \mathcal{Y}_\theta) + \mathcal{D}_\theta \cdot \dot{\mathcal{Y}}_\theta.$$

The result follows from gathering these two inequalities.

Proof of (H5): Using (43), we get for all $\varepsilon > 0$, $\theta \in [\theta_0, \theta_1]$ and $Y \in (\mathbb{R}^d)^N$,

$$\partial_\theta \nabla g_\varepsilon(\theta, Y) = -\frac{1}{\exp(\theta)} \left(\nabla h\left(\frac{Y}{\varepsilon \exp(\theta)} \right) + \mathrm{D}^2 h\left(\frac{Y}{\varepsilon \exp(\theta)} \right) \cdot \frac{Y}{\varepsilon \exp(\theta)} \right).$$

As we already saw in (H3) that ∇h is bounded, it suffices to prove that $X \mapsto \mathrm{D}^2 h(X) \cdot X$ is bounded. Let us expand everything in (44) and apply X to the right. We get

$$\mathrm{D}^2 h(X) \cdot X = \frac{\sum_{\sigma, \eta \in \mathfrak{S}_N} X \cdot (A^\sigma - A^\eta) A^\sigma \exp(X \cdot (A^\sigma + A^\eta))}{\sum_{\sigma, \eta \in \mathfrak{S}_N} \exp(X \cdot (A^\sigma + A^\eta))}.$$

As a consequence, it suffices to show that, for each $\sigma, \eta \in \mathfrak{S}_N$,

$$T(\sigma, \eta, X) := \frac{X \cdot (A^\sigma - A^\eta) \exp(X \cdot (A^\sigma + A^\eta))}{\sum_{\sigma', \eta' \in \mathfrak{S}_N} \exp(X \cdot (A^{\sigma'} + A^{\eta'}))}$$

is bounded, uniformly in X. First, if $\eta = \sigma$, then $T(\sigma, \sigma, X) = 0$. Else, let us use the bound

$$\sum_{\sigma', \eta' \in \mathfrak{S}_N} \exp(X \cdot (A^{\sigma'} + A^{\eta'})) \leq \exp(2X \cdot A^\sigma) + \exp(2X \cdot A^\eta),$$

obtained by only keeping the terms corresponding to $\sigma' = \eta' = \sigma$ and $\sigma' = \eta' = \eta$ in the sum. This leads to

$$|T(\sigma, \eta, X)| \leq \frac{|X \cdot (A^\sigma - A^\eta)| \exp(X \cdot (A^\sigma + A^\eta))}{\exp(2X \cdot A^\sigma) + \exp(2X \cdot A^\eta)} = \frac{|X \cdot (A^\sigma - A^\eta)|}{\exp(-|X \cdot (A^\sigma - A^\eta)|) + \exp(|X \cdot (A^\sigma - A^\eta)|)},$$

which is clearly bounded uniformly in X. The result follows. $\square$

4. The case of dimension 1: sticky collisions

In this section, we will study the global minimizers of the functional Λ' obtained in Section 2.6, in dimension $d = 1$. If we call t the time variable and if we replace θ_0 and θ_1 by 0 and T respectively, due to the invariance of the functional through translation in time, Λ' reads

$$\Lambda'(\mathcal{Z}) = \begin{cases} \int_0^T \left\{\frac{1}{2}|\dot{\mathcal{Z}}_t|^2 + \frac{1}{2}|\mathcal{Z}_t - \overline{\nabla} f(\mathcal{Z}_t)|^2\right\} \mathrm{d}t & \text{if } \mathcal{Z} \in H^1([0,T]; \mathbb{R}^N),\, \mathcal{Z}_0 \in \{P^\sigma\} \text{ and } \mathcal{Z}_T \in \{Q^\sigma\}, \\ +\infty & \text{else,} \end{cases} \tag{45}$$

where

$$f(X) = \max_{\sigma \in \mathfrak{S}_N} X \cdot A^\sigma, \qquad X \in \mathbb{R}^N. \tag{46}$$

Here, we chose a *strictly ordered* $A = (a_1, \ldots, a_N)$, that is, such that $a_1 < \cdots < a_N$, $P, Q \in \mathbb{R}^N$ and $T > 0$. Once again, when $X = (x_1, \ldots, x_N) \in \mathbb{R}^N$ and $\sigma \in \mathfrak{S}_N$, we let $X^\sigma := (x_{\sigma(1)}, \ldots, x_{\sigma(N)})$, and $\{P^\sigma\}$ and $\{Q^\sigma\}$ refer to $\{P^\sigma : \sigma \in \mathfrak{S}_N\}$ and $\{Q^\sigma : \sigma \in \mathfrak{S}_N\}$ respectively. Of course $P = (p_1, \ldots, p_N)$ and $Q = (q_1, \ldots, q_N)$ can be supposed to be ordered, that is, $p_1 \le \cdots \le p_N$ and $q_1 \le \cdots \le q_N$. We recall that we defined the *extended gradient* $\overline{\nabla} f$ in Definition 8. As already noticed in Section 2.5, the existence of global minimizers for Λ' follows from the direct method of calculus of variations. Uniqueness does not hold in general, even up to permutations.

The purpose of the section is two-fold. On the one hand, we will show that the model has nice regularity properties: any global minimizer of Λ' is smooth except on a finite number of "sticking" or "separation" times.[1] On the other hand, we will justify as claimed in Section 2 that Λ' describes a model with sticky collisions in the sense that a minimizer $\mathcal{Z} = (z_1(t), \ldots, z_N(t))$ of Λ' will typically exhibit some sticking effects as $z_i(t) = z_j(t)$ for $i \neq j$ on nontrivial intervals.

To describe the sticking effect, it is convenient to introduce the following definition:

Definition 20 (partition of $[\![1, N]\!]$). Let $X \in \mathbb{R}^N$. We say that X is divided according to $\pi(X)$ when $\pi(X)$ is the partition of $[\![1, N]\!]$ induced by the relation,

$$\text{for all } (i, j) \in [\![1, N]\!]^2, \quad i \sim j \iff x_i = x_j.$$

We call $C(X, i)$ the class of $i \in [\![1, N]\!]$ in $\pi(X)$, namely, $C(X, i) = \{j : x_j = x_i\}$.

The main result of the section is the following:

Theorem 21 (regularity of the optimal trajectories). *For given $A, P, Q \in \mathbb{R}^N$ and $T > 0$ as before, let $\mathcal{Z}$ be a global minimizer of Λ' defined in (45). Then $\mathcal{Z}$ is continuous and there exist*

$$0 = t_0 < t_1 < \cdots < t_p = T,$$

a family of times such that, for each $i = 1, \ldots, p$, $\mathcal{Z}$ is smooth on $[t_{i-1}, t_i]$, and $\pi(\mathcal{Z})$ is constant on (t_{i-1}, t_i).

It will be quite clear from the proof that sticking effects do occur. This exactly means that there exist trajectories $\mathcal{Z}$ for which, with the notation of Section 2.7, $\Lambda'(\mathcal{Z}) < \Lambda^+(\mathcal{Z})$. For such trajectories, $\mathcal{Z}_t$ is

[1] Notice that Λ' is invariant under time inversion, so that if particles are allowed to stick, they are also allowed to separate.

located on the set where f is not differentiable for a set of times of positive Lebesgue measure. But in dimension 1, this set is exactly the set where at least two particles are located at the same place. That is, the set of times when $\pi(\mathcal{Z}) \neq \{\{1\}, \ldots, \{N\}\}$ is typically of positive Lebesgue measure. As a consequence of Theorem 21, it is even a finite union of intervals.

Still it might be convenient to illustrate the sticking effects included in the model by the following easy proposition. It asserts that the set of times when all the particles are stuck is an interval: if all the particles are stuck at two different times, the cheapest behavior between these two times is to remain stuck. It also shows that this phenomenon occurs: if all the particles are sufficiently close at the initial and final time, then they necessarily stick together during a nontrivial interval along the evolution.

Proposition 22 (intervals of full degeneration). (1) *For given $A, P, Q \in \mathbb{R}^N$ and $T > 0$ as before, let $\mathcal{Z} = (z_1(t), \ldots, z_N(t))$ be a global minimizer of Λ'. Suppose there exist two times $0 \leq t_1 < t_2 \leq T$ such that*

$$z_1(t_1) = \cdots = z_N(t_1) \quad \text{and} \quad z_1(t_2) = \cdots = z_N(t_2).$$

Then, for all $t \in [t_1, t_2]$, we have $z_1(t) = \cdots = z_N(t)$.

(2) *For given $A \in \mathbb{R}^N$ and $T > 0$ as before, the set $\mathcal{U}$ of endpoints $P, Q \in \mathbb{R}^N$ with the property that, for all minimizers $\mathcal{Z} = (z_1(t), \ldots, z_N(t))$ of Λ', the set of times*

$$\{t \in [0, T] : z_1(t) = \cdots = z_N(t)\}$$

is a nontrivial interval, is a neighborhood of $\{P, Q \in \mathbb{R}^N : p_1 = \cdots = p_N$ and $q_1 = \cdots = q_N\}$.

The proof of Proposition 22 uses almost nothing and is given in Section 4.2. Except for that, the whole section is dedicated to the proof of Theorem 21. For this we take once for all $A, P, Q \in \mathbb{R}^N$ and $T > 0$, A being strictly ordered and P, Q being ordered.

Even if all the arguments are elementary, we will need a certain number of steps, including:

- The explicit computation of the potential $|X - \overline{\nabla} f(X)|^2$ (Section 4.1 and 4.4).
- The justification of a priori knowledge on the optimal trajectories: they can be supposed to be ordered at all times (Section 4.3).
- The conservation of energy and momentum holds during shocks[2] (Section 4.5).

Then, the main ingredient in the proof of Theorem 21 is an estimate given in Section 4.6: during a nonpathological shock (pathological shocks are excluded a posteriori), at least one particle has a lower-bounded jump in its velocity (Proposition 31). We finally provide the proof of Theorem 21 in Section 4.7.

Throughout the section, we will work with several types of finite sets: the partitions of type $\pi(X)$ and the class of particles of type $C(X, i)$. Some of the arguments or computations will deal with their cardinality. Thus, if $\mathcal{F}$ is a finite set, we will denote by $\#\mathcal{F}$ its cardinality.

[2] We say that $\mathcal{Z}$ presents a shock at time t if t is a discontinuity point of $\pi(\mathcal{Z})$; see Definition 30.

4.1. ***Properties of the extended gradient.*** In Lemma 24, we gather easy properties of $\overline{\nabla} f$ that will be needed in the following. Before doing so, let us introduce some notation.

Definition 23. Let π be a partition of $[\![1, N]\!]$. We call E_π the linear subspace of $\mathbb{R}^N$ of all X such that π is a refinement of $\pi(X)$, that is,

$$E_\pi := \bigcap_{C\in\pi}\bigcap_{i,j\in C}\{X = (x_1, \dots, x_N) \in \mathbb{R}^N : x_i = x_j\}.$$

Lemma 24 (properties of $\overline{\nabla} f$). (1) *The extended gradient $\overline{\nabla} f$ has the following symmetry*:

$$\textit{for all } X \in \mathbb{R}^N,\ \sigma \in \mathfrak{S}_N, \quad \overline{\nabla} f(X^\sigma) = (\overline{\nabla} f(X))^\sigma. \tag{47}$$

(2) *The function $X \mapsto |X - \overline{\nabla} f(X)|$ is symmetric*:

$$\textit{for all } X \in \mathbb{R}^N,\ \sigma \in \mathfrak{S}_N, \quad |X^\sigma - \overline{\nabla} f(X^\sigma)|^2 = |X - \overline{\nabla} f(X)|^2. \tag{48}$$

(3) *If X is ordered, then $\overline{\nabla} f(X)$ is the orthogonal projection of A on $E_{\pi(X)}$.*

(4) *If X is ordered and $i \in \{1, \dots, N\}$,*

$$(\overline{\nabla} f(X))_i = \frac{1}{\#C(X,i)}\sum_{j\in C(X,i)} a_j. \tag{49}$$

(*Recall that $C(X, i)$ is defined in Definition 20.*)

Remark 25. The extended gradient $\overline{\nabla} f$ is completely characterized by points (1) and (3) (or (4)) of Lemma 24.

Proof. (1) Let $\sigma \in \mathfrak{S}_N$. By the definition (46) of f, for all $X \in \mathbb{R}^N$, $f(X^\sigma) = f(X)$. Letting $I^\sigma : X \mapsto X^\sigma$, we easily deduce that at the level of subdifferentials: $\partial f(X^\sigma) = I^\sigma(\partial f(X))$. We conclude by the fact that I^σ is orthogonal.

(2) It is a direct consequence of point (1).

(3) Let $X = (x_1, \dots x_N) \in \mathbb{R}^N$ be an ordered vector. Considering the definition (46) of f and noticing that the maximum is achieved exactly for those σ such that $X^\sigma = X$, it appears that $\overline{\nabla} f(X)$ belongs to the convex hull:

$$\mathrm{Conv}(\{A^\sigma : \sigma \in \mathfrak{S}_N \text{ such that } X^\sigma = X\}).$$

For a given $i \in \{1, \dots, N\}$, we call $V^i \in \mathbb{R}^N$ the vector whose j-th coordinate is 1 if $j \in C(X, i)$ and 0 otherwise. On the one hand, we have $E_{\pi(X)} = \mathrm{Span}\{V^i : i = 1, \dots, N\}$, and on the other hand, for all i, the scalar product $V^i \cdot Y$ is constant on the above-mentioned convex hull. So we deduce

$$A - \overline{\nabla} f(X) \in (E_{\pi(X)})^\perp.$$

Hence, we just have to prove that $\overline{\nabla} f(X) \in E_{\pi(X)}$. If $i, j \in \{1, \dots, N\}$ are such that $x_i = x_j$, let us apply formula (47) to the permutation $\sigma := (i, j)$:

$$(\overline{\nabla} f(X))_i = ((\overline{\nabla} f(X))^\sigma)_j = (\overline{\nabla} f(X^\sigma))_j = (\overline{\nabla} f(X))_j.$$

The result follows.

(4) Let X be ordered and $i \in \{1, \dots, N\}$. As $\overline{\nabla} f(X) \in E_{\pi(X)}$, with the notation of the proof of (3),

$$
\begin{aligned}
(\overline{\nabla} f(X))_i &= \frac{1}{\#C(X,i)} \sum_{j \in C(X,i)} (\overline{\nabla} f(X))_j = \frac{1}{\#C(X,i)} \overline{\nabla} f(X) \cdot V^i \\
&= \frac{1}{\#C(X,i)} A \cdot V^i = \frac{1}{\#C(X,i)} \sum_{j \in C(X,i)} a_j,
\end{aligned}
$$

where we used $A - \overline{\nabla} f(X) \perp V^i$ to get the first identity in the second line. □

The three next subsections will be dedicated to consequences of this lemma:

- A proof of Proposition 22.
- When proving Theorem 21, it is enough to consider ordered trajectories (Proposition 26).
- For ordered trajectories, the potential in Λ' can be decomposed as sum of a smooth "external" potential and an "internal" energy only depending on $\pi(X)$ (Proposition 27).

4.2. *Proof of Proposition 22.* With the help of Lemma 24, we are ready to prove Proposition 22.

Proof of Proposition 22. (1) Without loss of generality, we can suppose $t_1 = 0$ and $t_2 = T$, that is, $P = (p_1, \dots, p_N)$ and $Q = (q_1, \dots, q_N)$ are such that $p_1 = \dots = p_N$ and $q_1 = \dots = q_N$.

Call Ψ the orthogonal projection on the line $E_{[\![1,N]\!]} := \{X = (x_, \dots, x_N) \in \mathbb{R}^N \mid x_1 = \dots = x_N\}$. It suffices to prove that when $\mathcal{Z}$ is a continuous trajectory joining P to Q, then $\Lambda'(\Psi(\mathcal{Z})) \le \Lambda'(\mathcal{Z})$, and with equality if and only if $\mathcal{Z} = \Psi(\mathcal{Z})$. As Ψ is 1-Lipschitz, it reduces the kinetic part of Λ'. For the potential part, we remark that, for all $X \in \mathbb{R}^N$, $E_{\pi(\Psi(X))} = E_{[\![1,N]\!]} \subset E_{\pi(X)}$. As a consequence, by point (3) of Lemma 24, we have as soon as X is ordered $\overline{\nabla} f(\Psi(X)) = \Psi(\overline{\nabla} f(X))$. Hence

$$
|\Psi(X) - \overline{\nabla} f(\Psi(X))|^2 = |\Psi(X - \overline{\nabla} f(X))|^2 \le |X - \overline{\nabla} f(X)|^2,
$$

with equality if and only if $X \in E_{[\![1,N]\!]}$, i.e., if and only if $\Psi(X) = X$. This property is extended to nonordered X using (48), and the result follows.

(2) The function $\overline{\Lambda'} = \overline{\Lambda'}(P, Q)$, defined for all $P, Q \in \mathbb{R}^N$ as the minimal value of Λ', is continuous. Indeed, if $P, P', Q, Q' \in \mathbb{R}^N$ are chosen so that $|P' - P| + |Q' - Q| \ll 1$ and if $\mathcal{Z}$ is a trajectory joining P to Q, we can find a trajectory $\widetilde{\mathcal{Z}}$ joining P' to Q' with[3]

$$
\Lambda'(\widetilde{\mathcal{Z}}) \le \Lambda'(\mathcal{Z}) + \underset{(P',Q') \to (P,Q)}{o}(1). \tag{50}
$$

To do so, it suffices to choose $\tau \sim |P' - P| + |Q' - Q|$, and to define $\widetilde{\mathcal{Z}}$ as the trajectory joining P' to P in straight line between times 0 and τ, joining P to Q between times τ and $T - \tau$ by following $\mathcal{Z}$ with a proper affine change of time, and finally joining Q to Q' in straight line between times $T - \tau$ and T. This shows that $\overline{\Lambda'}$ is lower semicontinuous, but the continuity is obtained by noticing that the o in (50) is locally uniform on $P, Q \in \mathbb{R}^N$. The argument is easily adapted to show that $\widetilde{\Lambda'} = \widetilde{\Lambda'}(P, Q)$, defined

[3] With a slight abuse of notation, we do not refer explicitly to the dependence of Λ' on P, Q.

for $P, Q \in \mathbb{R}^N$ by

$$\widetilde{\Lambda}'(P, Q) := \inf\{\Lambda'(\mathcal{Z}) : \mathcal{Z} \text{ whose set of } t \text{ such that } \mathcal{Z}_t \in E_{[\![1,N]\!]} \text{ is negligible}\},$$

is also continuous. Additionally, the set $\mathcal{U}$ defined in the statement clearly satisfies

$$\mathcal{V} := \{P, Q \in \mathbb{R}^N : \overline{\Lambda}'(P, Q) < \widetilde{\Lambda}'(P, Q)\} \subset \mathcal{U}.$$

By the continuity of $\overline{\Lambda}'$ and $\widetilde{\Lambda}'$, $\mathcal{V}$ is an open set. Hence it remains to prove that

$$\{P, Q \in \mathbb{R}^N : p_1 = \cdots = p_N \text{ and } q_1 = \cdots = q_N\} = E_{[\![1,N]\!]} \times E_{[\![1,N]\!]} \subset \mathcal{V}.$$

To do so, we take $P, Q \in E_{[\![1,N]\!]}$ and $\mathcal{Z}$ a curve joining P to Q such that $\{t : \mathcal{Z}_t \in E_{[\![1,N]\!]}\}$ is negligible, we still call Ψ the orthogonal projection on $E_{[\![1,N]\!]}$, and we prove that

$$\Lambda'(\mathcal{Z}) \geq \Lambda'(\Psi(\mathcal{Z})) + a,$$

where $a > 0$ does not depend on $\mathcal{Z}$. Let us call $\Phi := \mathrm{Id} - \Psi$ the orthogonal projection on the orthogonal of $E_{[\![1,N]\!]}$. As in the proof of the first point, $\overline{\nabla} f \circ \Psi = \Psi \circ \overline{\nabla} f$. As a consequence

$$\begin{aligned}\Lambda'(\mathcal{Z}) &= \int_0^T \{|\Psi(\dot{\mathcal{Z}}_t)|^2 + |\Psi(\mathcal{Z}_t) - \Psi(\overline{\nabla} f(\mathcal{Z}_t))|^2\}\,\mathrm{d}t + \int_0^T \{|\Phi(\dot{\mathcal{Z}}_t)|^2 + |\Phi(\mathcal{Z}_t) - \Phi(\overline{\nabla} f(\mathcal{Z}_t))|^2\}\,\mathrm{d}t \\ &= \Lambda'(\Psi(\mathcal{Z})) + \int_0^T \{|\dot{\mathcal{Z}}_t^\perp|^2 + |\mathcal{Z}_t^\perp - \Phi(\overline{\nabla} f(\mathcal{Z}_t))|^2\}\,\mathrm{d}t,\end{aligned}$$

where $\mathcal{Z}^\perp = \mathcal{Z}_t^\perp := \Phi(\mathcal{Z}_t)$ is a curve joining 0 to 0. But for almost all t, $\mathcal{Z}_t \notin E_{[\![1,N]\!]}$, so as we saw in the proof of the first point, $\overline{\nabla} f(\mathcal{Z}_t) \notin E_{[\![1,N]\!]}$. As $\overline{\nabla} f$ only takes a finite number of values (see Lemma 24), for almost all t, $\Phi(\overline{\nabla} f(\mathcal{Z}_t))$ belongs to some finite set, say $\mathcal{G}$, which does not contain 0. Hence,

$$\int_0^T \{|\dot{\mathcal{Z}}_t^\perp|^2 + |\mathcal{Z}_t^\perp - \Phi(\overline{\nabla} f(\mathcal{Z}_t))|^2\}\,\mathrm{d}t \geq \int_0^T \{|\dot{\mathcal{Z}}_t^\perp|^2 + \mathrm{dist}(\mathcal{Z}_t^\perp, \mathcal{G})^2\}\,\mathrm{d}t,$$

where $\mathrm{dist}(Z, \mathcal{G})$ denotes the distance from Z to $\mathcal{G}$. Because $\mathcal{Z}^\perp$ joins 0 to 0 and $\mathcal{G}$ does not contain 0, this last integral is easily seen to be bounded below away from 0 independently of $\mathcal{Z}$, and the result follows. □

4.3. ***Ordering of the particles.*** The purpose of this subsection is to show that when proving Theorem 21, we can restrict ourselves to study trajectories that remain ordered (see Figure 1). This is due to the following proposition.

Proposition 26. *Let $\mathcal{Z} = \mathcal{Z}_t$ be a global minimizer of Λ'. We call $\widetilde{\mathcal{Z}} = \widetilde{\mathcal{Z}}_t$ the trajectory obtained by reordering the coordinates of $\mathcal{Z}$ in increasing order. Then $\widetilde{\mathcal{Z}}$ is also a global minimizer of Λ'.*

Moreover, $\mathcal{Z}$ has the regularity stated in Theorem 21 if and only if $\widetilde{\mathcal{Z}}$ does.

In particular, Λ' always admits an ordered minimizer, and it is enough to prove Theorem 21 for such minimizers.

Thanks to this proposition, from now on, we only work with ordered minimizers of Λ'. These minimizers $\mathcal{Z} = \mathcal{Z}_t$ satisfy in particular $\mathcal{Z}_0 = P$ and $\mathcal{Z}_T = Q$ (as we chose them to be ordered in the first place).

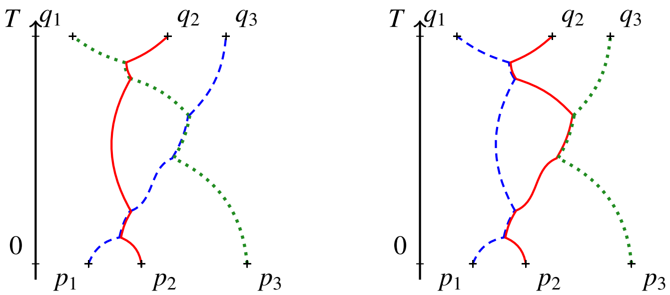

Figure 1. These two trajectories share their initial and final positions up to ordering and their actions. But to the right, the order is preserved, while to the left, this is not the case.

Proof. Let $\mathcal{Z}$ and $\widetilde{\mathcal{Z}}$ be as in the statement of the proposition. Point (2) of Lemma 24 implies

$$\int_0^T |\widetilde{\mathcal{Z}}_t - \overline{\nabla} f(\widetilde{\mathcal{Z}}_t)|^2 \, \mathrm{d}t = \int_0^T |\mathcal{Z}_t - \overline{\nabla} f(\mathcal{Z}_t)|^2 \, \mathrm{d}t.$$

We call $\Psi : \mathbb{R}^N \to \mathbb{R}^N$ the operator that reorders the coordinates of a vector in increasing order, so that in particular, for all t, $\widetilde{\mathcal{Z}}_t = \Psi(\mathcal{Z}_t)$. A simple application of the rearrangement inequality shows that Ψ is 1-Lipschitz. In particular, it reduces the action of curves

$$\int_0^T |\dot{\widetilde{\mathcal{Z}}}_t|^2 \, \mathrm{d}t \le \int_0^T |\dot{\mathcal{Z}}_t|^2 \, \mathrm{d}t.$$

By adding the two last formulas, and by noticing that the endpoint constraint is fulfilled, we get $\Lambda'(\widetilde{\mathcal{Z}}) \le \Lambda'(\mathcal{Z})$. As $\mathcal{Z}$ is a minimizer, this inequality is in fact an equality, and $\widetilde{\mathcal{Z}}$ is also a minimizer.

Note that both $\mathcal{Z}$ and $\widetilde{\mathcal{Z}}$ are continuous because they have finite action. Hence, the second claim of the proposition is a consequence of the two following facts:

- For all $t \in [0, T]$, $\#\pi(\widetilde{\mathcal{Z}}_t) = \#\pi(\mathcal{Z}_t)$.
- For any continuous trajectory $t \in I \mapsto \mathcal{X}_t \in \mathbb{R}^N$, where I is an interval, $t \mapsto \pi(\mathcal{X}_t)$ is constant if and only if $t \mapsto \#\pi(\mathcal{X}_t)$ is constant.

Indeed in that case, $t \mapsto \pi(\mathcal{Z}_t)$ and $t \mapsto \pi(\widetilde{\mathcal{Z}}_t)$ are constant on the same intervals, and the result follows.

The first point and the "only if" part of the second point are trivial.

For the "if" part of the second one, we reason by contraposition. Suppose $s \mapsto \pi(\mathcal{X}_s)$ has a discontinuity at time t and we prove that $s \mapsto \#\pi(\mathcal{X}_s)$ also does. If $s \mapsto \pi(\mathcal{X}_s)$ has a discontinuity at time t, we can find two distinct accumulation points π_1 and π_2 of $s \mapsto \pi(\mathcal{X}_s)$ at time t. As the set E_π is closed for all π, we find that $\mathcal{X}_t$ belongs to $E_{\pi_1} \cap E_{\pi_2}$. But this set is nothing but $E_{\bar{\pi}}$, where $\bar{\pi}$ is the finest partition of which π_1 and π_2 are refinements, that is, the partition corresponding to the relation

$$i \sim j \quad \Longleftrightarrow \quad \text{there exists } C \in \pi_1 \cup \pi_2 \text{ such that } \{i, j\} \subset C.$$

In particular, $\pi(\mathcal{X}_t)$ is a refinement of $\bar{\pi}$ and as $\pi_1 \ne \pi_2$,

$$\#\pi(\mathcal{X}_t) \le \#\bar{\pi} < \max(\#\pi_1, \#\pi_2).$$

So $s \mapsto \#\pi(\mathcal{X}_s)$ has a discontinuity at time t, and the result follows. □

4.4. ***Decomposition of the potential.*** Here, we compute explicitly the values of the potential $X \mapsto |X - \overline{\nabla} f(X)|^2$ on ordered vectors $X \in \mathbb{R}^N$. Notice that, for such vectors X, $\pi(X)$ has an additional structure: if $C \in \pi(X)$, then C is an interval of integers. We say that such partitions are *ordered*. We prove the following:

Proposition 27. *For all ordered $X \in \mathbb{R}^N$,*

$$|X - \overline{\nabla} f(X)|^2 = |X - A|^2 + h(\pi(X)) - |A|^2, \tag{51}$$

where h is defined on a partition π of $[\![1, N]\!]$ by

$$h(\pi) := \sum_{C \in \pi} \frac{1}{\#C} \left| \sum_{j \in C} a_j \right|^2. \tag{52}$$

In particular, h has the following monotonicity property: if π and π' are two ordered partitions and if π' is a strict refinement of π, then $h(\pi) < h(\pi')$.

The more particles are stuck together, the lower h is. This is the reason for which Λ' favors the sticking of particles. The function $-h/2$ can be understood as the internal energy of the system.

Dropping the constant term $|A|^2/2$ in (51) and defining Λ'' on a trajectory $\mathcal{Z}$ by

$$\Lambda''(\mathcal{Z}) = \begin{cases} \frac{1}{2}\int_0^T \{|\dot{\mathcal{Z}}_t|^2 + |\mathcal{Z}_t - A|^2 + h(\pi(\mathcal{Z}_t))\}\,\mathrm{d}t & \text{if } \mathcal{Z} \in H^1([0,T];\mathbb{R}^N),\ \mathcal{Z}_0 = P \text{ and } \mathcal{Z}_T = Q, \\ +\infty & \text{else,} \end{cases} \tag{53}$$

it is clear that Λ' and Λ'' have the same minimizers in the class of ordered trajectories. Hence, as a consequence of Proposition 26, it suffices to prove the conclusion of Theorem 21 for the minimizers of Λ'' in the class of ordered trajectories.

Proof of Proposition 27. Let $X \in \mathbb{R}^N$ be an ordered vector. By point (3) of Lemma 24, we have $A - \overline{\nabla} f(X) \in (E_{\pi(X)})^\perp$ and both X and $\overline{\nabla} f(X) \in E_{\pi(X)}$. So using the Pythagorean theorem twice, we get

$$|X - \overline{\nabla} f(X)|^2 = |X - A|^2 - |A - \overline{\nabla} f(X)|^2 = |X - A|^2 + |\overline{\nabla} f(X)|^2 - |A|^2.$$

The identities (51) and (52) are obtained by computing $|\overline{\nabla} f(X)|^2$ using (49).

If we recap, $h(\pi)$ is the squared norm of the orthogonal projection of A on E_π. But if π' is a refinement of π, $E_\pi \subset E_{\pi'}$, and hence $h(\pi) \le h(\pi')$. The strict inequality is obtained by noticing with the help of (49) and using the strict ordering of A that if in addition π and π' are ordered and $\pi' \neq \pi$, then the projection of A on $E_{\pi'}$ does not belong to E_π. □

4.5. ***Conserved quantities.*** In this subsection, we discuss two simple and yet structural properties of the dynamic prescribed by the functionals Λ', Λ'': the Hamiltonian of the system is conserved (Proposition 28), and its center of mass draws a smooth curve (Proposition 29). In particular, the momentum of the system is conserved during shocks.

Proposition 28. *Let $\mathcal{Z}$ be an ordered minimizer of Λ''. Then*

$$\mathcal{E} = \mathcal{E}(t) := \tfrac{1}{2}\{|\dot{\mathcal{Z}}_t|^2 - |\mathcal{Z}_t - A|^2 - h(\pi(\mathcal{Z}_t))\} \tag{54}$$

is constant in the sense of distributions.

Proof. The proof is completely standard and is done by comparing the value of Λ'' on $\mathcal{Z}$ and $t \mapsto \mathcal{Z}_{t+\varepsilon\varphi(t)}$ for small ε and functions φ that are smooth and compactly supported in $(0, T)$. □

Proposition 29. *Let $\mathcal{Z} = (z_1(t), \dots, z_N(t))$ be an ordered minimizer of Λ''. Call $a := (a_1 + \cdots + a_N)/N$ and for $t \in [0, T]$*

$$\mathcal{M}(t) := \frac{1}{N}\sum_{i=1}^N z_i(t) \quad \text{and} \quad \mathcal{P}(t) := \frac{1}{N}\sum_{i=1}^N \dot{z}_i(t).$$

($\mathcal{M}$ is well-defined for all t, and $\mathcal{P}$ for almost all t.) Then $\mathcal{M}$, $\mathcal{P}$ solve distributionally

$$\dot{\mathcal{M}}(t) = \mathcal{P}(t), \qquad \dot{\mathcal{P}}(t) = \mathcal{M}(t) - a.$$

In particular, $\mathcal{M}$ is smooth and $\mathcal{P}$ coincide almost surely with a smooth function.

Proof. Here the proof consists in comparing the value of Λ'' on $\mathcal{Z}$ and $t \mapsto \mathcal{Z}_t + \varepsilon\varphi(t)V$ for small ε, smooth and compactly supported φ, and where we call $V = (1, \dots, 1)$. The only somehow unusual thing to remark is that π and hence $h \circ \pi$ are invariant under translations in the direction of V. □

4.6. *Shocks, isolated shocks and minimal deviation.* This subsection contains the main estimate that allows us to prove Theorem 21. Roughly speaking, if at time t some of the particles stick or separate, there is a lower bound on the change of the velocity of at least one particle. The proof of Theorem 21 will then consist in showing that this cannot happen an infinite number of times.

Let us first define as "shocks" these sticking and separating behaviors:

Definition 30 (shocks). Let $\mathcal{X} = \mathcal{X}_t = (x_1(t), \dots, x_N(t))$ be a continuous trajectory on $\mathbb{R}^N$.

(1) We call a shock of $\mathcal{X}$ a triplet (t, q, C) with $t \in [0, T]$, $q \in \mathbb{R}$ and $C \subset [\![1, N]\!]$ such that
- $C \in \pi(\mathcal{X}_t)$,
- for all $i \in C$, $x_i(t) = q$,
- for all $\tau > 0$, there exists $s \in (t - \tau, t + \tau)$ such that $C \notin \pi(\mathcal{X}_s)$.

(2) If (t, q, C) is a shock of $\mathcal{X}$, we say that it is isolated if (t, q) is isolated in

$$\{(t', q') : \text{there exists } C' \subset [\![1, N]\!] \text{ such that } (t', q', C') \text{ is a shock}\},$$

i.e., if there is no other shock than (t, q, C) in the neighborhood of $(t, q) \in [0, T] \times \mathbb{R}$.

We provide in Figure 2 a picture of a shock which does not seem to be isolated. The following result is the main step in the proof of Theorem 21.

Proposition 31. *Let $\mathcal{Z} = (z_1(t), \dots, z_N(t))$ be an ordered minimizer of Λ' (or equivalently a minimizer of Λ'' in the class of ordered trajectories), and let $t \in [0, T]$.*

(1) *If particle i is not involved in a shock at time t, then for s in the neighborhood of t, $C := C(\mathcal{Z}_s, i)$ is constant and z_i is a smooth solution of*

$$\ddot{z}_i(s) = z_i(s) - \frac{1}{\#C}\sum_{j \in C} a_j. \tag{55}$$

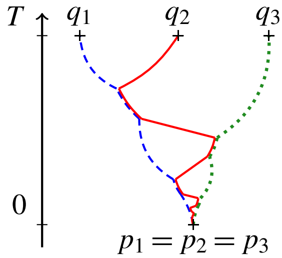

Figure 2. A shock with three particles which does not seem to be isolated. We will see later on that this kind of shock cannot occur in our model.

In particular, if i is involved in an isolated shock at time t, then z_i admits left and right derivatives at time t, denoted by $\dot{z}_i(t-)$ and $\dot{z}_i(t+)$ respectively.

(2) *There is $\alpha = \alpha(N, A) > 0$ such that for any isolated shock (t, q, C), calling $i := \min C$,*

$$\dot{z}_i(t-) - \dot{z}_i(t+) \geq \alpha. \tag{56}$$

(*Note that the quantity $\dot{z}_i(t-) - \dot{z}_i(t+)$ is not affected by time inversion. In particular, this lower bound is coherent with the invariance of the Lagrangian through time inversion.*)

Proof. (1) If particle i is not involved in a shock at time t, by the definition of a shock, it means that $C := C(\mathcal{Z}_t, i) \in \pi(\mathcal{Z}_s)$ for all s in a neighborhood of t. In particular, for all $j \in C$ and s sufficiently close to t, by (49),

$$(\overline{\nabla} f(\mathcal{Z}_s))_j = \frac{1}{\#C} \sum_{k \in C} a_k.$$

On the other hand, it is easy to find a neighborhood U of $(t, z_i(t))$ in $[0, T] \times \mathbb{R}$ such that, for all $j \in \{1, \dots, N\}$ and all $s \in [0, T]$, $(s, z_j(s)) \in U$ implies $j \in C$.

As a consequence, if $\xi : [0, T] \to \mathbb{R}$ is smooth and compactly supported in a sufficiently small neighborhood of t, and if ε is sufficiently small, by defining $\widetilde{\mathcal{Z}} = (\tilde{z}_1(s), \dots, \tilde{z}_N(s))$ for any $j \in \{1, \dots, N\}$ and $s \in [0, T]$ by

$$\tilde{z}_j(s) := \begin{cases} z_j(s) + \varepsilon\xi(s) & \text{if } j \in C, \\ z_j(s) & \text{else,} \end{cases}$$

then $\pi(\mathcal{Z})$ and $\pi(\widetilde{\mathcal{Z}})$ (and hence $\overline{\nabla} f(\mathcal{Z})$ and $\overline{\nabla} f(\widetilde{\mathcal{Z}})$) coincide at all time. The ODE follows from comparing the values of Λ' on $\mathcal{Z}$ and trajectories of type $\widetilde{\mathcal{Z}}$.

In particular, by boundedness of $\mathcal{Z}$, if particle i is not involved in a shock at time t, $|\ddot{z}_i|$ is bounded by a constant that is not depending on t. The existence of $\dot{z}_i(t-)$ and $\dot{z}_i(t+)$ at the times of isolated shocks follows easily.

(2) This is the heart of our study of the dynamical system, and maybe the less standard part of Section 4. But still the idea is very easy: with the notation of the statement, if $\dot{z}_i(t-) - \dot{z}_i(t+)$ is too small, then it is cheaper to stick particle i with other particles, as shown in Figure 3. The proof goes as follows.

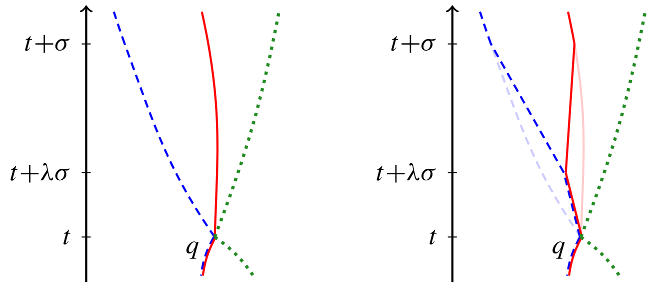

Figure 3. To the left, a piece of the trajectory $\mathcal{Z}$, and to the right, the competitor $\mathcal{Z}^{\sigma,\lambda}$ that we describe in the proof.

Step 1: Definition of a competitor. Let us consider (t, q, C) an isolated shock. Because it is isolated, we can find $\tau > 0$ such that the particles of C are not involved in another shock between times $t-\tau$ and $t+\tau$. By the definition of a shock, we cannot have $C \in \pi(\mathcal{Z}_s)$ for all $s \in (t-\tau, t+\tau)$, so either, for all $s \in (t-\tau, t)$, $C \notin \pi(\mathcal{Z}_s)$ or, for all $s \in (t, t+\tau)$, $C \notin \pi(\mathcal{Z}_s)$. Without loss of generality, we suppose that the second one holds: the particles of C are not all stuck right after the shock. Moreover, by our choice of τ, for all $C' \subset C$, the assertion $C' \in \pi(\mathcal{Z}_s)$ is either true or false independently of $s \in (t, t+\tau)$. Then, for $s \in (t, t+\tau)$, the following definitions of $C_1, C_2 \in \pi(\mathcal{Z}_s)$ do not depend on s:

$$C_1 := C(\mathcal{Z}_s, i) \quad \text{for } i = \min C \qquad \text{and} \qquad C_2 := C(\mathcal{Z}_s, i) \quad \text{for } i = \min C \backslash C_1.$$

(The classes C_1 and C_2 are the two leftmost packs of particles of C right after the shock.) Let us define for $j = 1, 2$

$$k_j := \# C_j, \qquad v_j := \dot{z}_i(t+) \quad \text{for } i \in C_j, \qquad \text{and} \qquad p := \frac{k_1 v_1 + k_2 v_2}{k_1 + k_2}. \tag{57}$$

For $0 \leq \sigma < \tau$ and $\lambda \in [0, 1)$, we define a competitor $\mathcal{Z}^{\sigma,\lambda} = (z_1^{\sigma,\lambda}(s), \dots, z_N^{\sigma,\lambda}(s))$ by setting for all $i = \{1, \dots, N\}$ and $s \in [0, T]$

$$z_i^{\sigma,\lambda}(s) = \begin{cases} z_i(s) & \text{if } i \notin C_1 \cup C_2 \text{ or } s \notin (t, t+\sigma), \\ q + (s-t)p & \text{if } i \in C_1 \cup C_2 \text{ and } s \in (t, t+\lambda\sigma), \\ \dfrac{t+\sigma-s}{(1-\lambda)\sigma}(q + \lambda\sigma p) + \dfrac{s-(t+\lambda\sigma)}{(1-\lambda)\sigma} z_i(t+\sigma) & \text{if } i \in C_1 \cup C_2 \text{ and } s \in (t+\lambda\sigma, t+\sigma). \end{cases}$$

(See Figure 3 for an illustration of this competitor.) We will get a lower bound on $v_2 - v_1$ by comparing the value of Λ'' on $\mathcal{Z}$ and $\mathcal{Z}^{\sigma,\lambda}$, and by differentiating the corresponding inequality first with respect to σ at $\sigma = 0$ (we zoom so that the particles of $\mathcal{Z}$ only travel along straight lines), and then with respect to λ at $\lambda = 0$ (we compute the first variation of the action when we let the particles stick together).

Step 2: A lower bound on $v_2 - v_1$. The partitions $\pi(\mathcal{Z}_s^{\sigma,\lambda})$ and $\pi(\mathcal{Z}_s)$ coincide at all times except between t and $t + \lambda\sigma$, when $\pi(\mathcal{Z}_s)$ is a strict refinement of $\pi(\mathcal{Z}_s^{\sigma,\lambda})$. Hence, letting

$$\delta = \delta(N, A) := \min\{h(\pi) - h(\pi') : (\pi, \pi') \text{ ordered partition of } [\![1, N]\!],\ \pi \text{ strict refinement of } \pi'\} > 0,$$

we have, for all $s \in (t, t+\lambda\sigma)$,

$$h(\pi(\mathcal{Z}_s^{\lambda,\sigma})) + \delta \le h(\pi(\mathcal{Z}_s)). \tag{58}$$

As $\mathcal{Z}^\sigma$ coincide with $\mathcal{Z}$ for times outside $(t, t+\sigma)$ and for coordinates that are not in $C_1 \cup C_2$, by definition (53) of Λ'', we have

$$\begin{aligned}\Lambda''(\mathcal{Z}^{\sigma,\lambda}) - \Lambda''(\mathcal{Z}) &= \sum_{i\in C_1\cup C_2} \int_t^{t+\sigma} \{|\dot z_i^{\sigma,\lambda}(s)|^2 + |z_i^{\sigma,\lambda}(s) - a_i|^2 - |\dot z_i(s)|^2 - |z_i(s)-a_i|^2\}\,\mathrm{d}s \\ &\qquad + \int_t^{t+\lambda\sigma} \{h(\pi(\mathcal{Z}_s^{\sigma,\lambda})) - h(\pi(\mathcal{Z}_s))\}\,\mathrm{d}s \\ &\le \sum_{i\in C_1\cup C_2} \int_t^{t+\sigma} \{|\dot z_i^{\tau,\sigma}(s)|^2 - |\dot z_i(s)|^2\}\,\mathrm{d}s - \delta\lambda\sigma + \underset{\sigma\to 0}{o}(\sigma),\end{aligned} \tag{59}$$

where to obtain the second line, we used (58) and the fact that between times t and $t+\sigma$, both z_i and $z_i^{\sigma,\lambda}$ remain at a distance of order σ of q.

Let us consider $i \in C_j$ for $j = 1, 2$. On one hand, as z_i admits v_j as a right derivative at time t, we have

$$\int_t^{t+\sigma} |\dot z_i(s)|^2\,\mathrm{d}s = v_j^2\sigma + \underset{\sigma\to 0}{o}(\sigma). \tag{60}$$

On the other hand, we can compute explicitly

$$\begin{aligned}\int_t^{t+\sigma} |\dot z_i^{\sigma,\lambda}(s)|^2\,\mathrm{d}s &= \lambda p^2\sigma + (1-\lambda)\sigma\left(\frac{z_i(t+\sigma) - (q+\lambda p\sigma)}{(1-\lambda)\sigma}\right)^2 \\ &= \lambda p^2\sigma + \frac{1}{(1-\lambda)\sigma}\Big(q + v_j\sigma + \underset{\sigma\to 0}{o}(\sigma) - q - \lambda p\sigma\Big)^2 \\ &= \lambda p^2\sigma + (v_j - \lambda p)^2\frac{\sigma}{1-\lambda} + \underset{\sigma\to 0}{o}(\sigma).\end{aligned} \tag{61}$$

By plugging (60) and (61) into (59) and by using the definition (57) of k_1, k_2 and p, we get

$$\begin{aligned}\Lambda''(\mathcal{Z}^{\sigma,\lambda}) - \Lambda''(\mathcal{Z}) &\le \left\{(k_1+k_2)\lambda p^2 + \frac{k_1(v_1-\lambda p)^2 + k_2(v_2-\lambda p)^2}{1-\lambda} - k_1v_1^2 - k_2v_2^2 - \delta\lambda\right\}\sigma + \underset{\sigma\to 0}{o}(\sigma) \\ &= \{(k_1+k_2)p^2 + k_1v_1^2 + k_2v_2^2 - 2p(k_1v_1 + k_2v_2) - \delta(1-\lambda)\}\frac{\lambda}{1-\lambda}\sigma + \underset{\sigma\to 0}{o}(\sigma) \\ &= \left\{k_1v_1^2 + k_2v_2^2 - \frac{(k_1v_1+k_2v_2)^2}{k_1+k_2} - \delta(1-\lambda)\right\}\frac{\lambda}{1-\lambda}\sigma + \underset{\sigma\to 0}{o}(\sigma) \\ &= \left\{\frac{k_1k_2}{k_1+k_2}(v_2-v_1)^2 - \delta(1-\lambda)\right\}\frac{\lambda}{1-\lambda}\sigma + \underset{\sigma\to 0}{o}(\sigma).\end{aligned}$$

By the minimality of $\Lambda''(\mathcal{Z})$, this quantity must be nonnegative. If we divide it by $\lambda\sigma$, and if we let σ and then λ go to zero, we end up with

$$\frac{k_1k_2}{k_1+k_2}(v_2-v_1)^2 \ge \delta. \tag{62}$$

Step 3: Conservation of momentum during an isolated shock and conclusion. Because (t, q, C) is isolated, it is easy to justify that we can replace V by the vector V^C whose j-th coordinate is 1 if $j \in C$ and 0 otherwise in the proof of Proposition 29. Doing so, we obtain the "local" conservation of momentum

$$\frac{1}{\#C}\sum_{i\in C}\dot{z}_i(t-) = \frac{1}{\#C}\sum_{i\in C}\dot{z}_i(t+) =: \mathcal{P}^C(t).$$

By ordering of the particles, we have, for $i = \min C$,

$$\dot{z}_i(t-) \geq \mathcal{P}^C(t) = \frac{1}{\#C}\sum_{i\in C}\dot{z}_i(t+) \geq \frac{k_1}{\#C}v_1 + \frac{\#C - k_1}{\#C}v_2.$$

(Indeed, $j \in C \mapsto \dot{z}_j(t-)$ and $j \in C \mapsto \dot{z}_j(t+)$ are clearly nonincreasing and nondecreasing respectively.) By recalling that $v_1 = \dot{z}_i(t+)$ and using (62), we get

$$\dot{z}_i(t-) - \dot{z}_i(t+) \geq \frac{\#C - k_1}{\#C}(v_2 - v_1) \geq \frac{\#C - k_1}{\#C}\sqrt{\frac{k_1 + k_2}{k_1 k_2}\delta}.$$

The minimal right-hand side's value is $\sqrt{\delta/(\#C^2 - \#C)}$, obtained for $k_1 = \#C - 1$ and $k_2 = 1$. Hence, we get the result by choosing $\alpha = \sqrt{\delta/(N^2 - N)}$. □

4.7. *Conclusion: proof of Theorem 21.* We are now ready to give the proof of Theorem 21. We give ourselves $\mathcal{Z}$ a global minimizer of Λ'. Thanks to Proposition 26, we can suppose that $\mathcal{Z}$ is ordered, and thanks to Proposition 27, we can consider Λ'' instead of Λ'.

Because of Proposition 31, it suffices to prove that there is a finite number of shocks. Indeed, in that case one can take for $0 = t_0 < t_1 < \cdots < t_p = T$ the moments of these shocks (and the endpoints of $[0, T]$). The smoothness of $\mathcal{Z}$ on each $[t_{i-1}, t_i]$, $i = 1, \ldots, p$, follows directly from Proposition 31. Then $\pi(\mathcal{Z})$ is constant on each (t_{i-1}, t_i), $i = 1, \ldots, p$, because by Definition 30 of a shock, at each time of discontinuity of $\pi(\mathcal{Z})$, there is at least one shock.

The set

$$\{(t', q') : \text{there exists } C' \subset [\![1, N]\!] \text{ such that } (t', q', C') \text{ is a shock}\}$$

is easily seen to be compact. So if it is not finite, it admits at least one accumulation point. That is, if there is an infinite number of shocks, then there is at least one shock which is not isolated. Let us consider such a shock (t, q, C) with minimal number of particles involved, i.e., with minimal $\#C$. The rest of the proof is dedicated to showing that the existence of (t, q, C) leads to a contradiction.

Step 1: The velocities are bounded. As $\mathcal{Z}$ is continuous on $[0, T]$, it is bounded. On the other hand, by definition, $h \leq |A|^2$. Now if $i \in \{1, \ldots, N\}$ and $t \in [0, T]$ is such that $\mathcal{Z}$ is differentiable at t (which is true for almost any t), recalling the definition (54) of $\mathcal{E}$,

$$\dot{z}_i(t)^2 \leq |\dot{\mathcal{Z}}_t|^2 \leq 2\mathcal{E} + |\mathcal{Z}_t - A|^2 + h(\pi(\mathcal{Z}_t)),$$

which is bounded uniformly in t.

Step 2: All the shocks in the neighborhood of (t, q) that are distinct from (t, q) are isolated. Let U be a neighborhood of (t, q) in $[0, T] \times \mathbb{R}$ such that, for all $s \in [0, T]$ and $i \in \{1, \ldots, N\}$, $(s, z_i(s)) \in U$ implies $i \in C$. This is possible since $\mathcal{Z}$ is continuous and, for all $j \notin C$, $z_j(t) \neq q$ by Definition 30 of a shock. Let us consider (t', q', C') a shock with $(t', q') \in U$. If $\#C' < \#C$, then (t', q', C') is isolated by the minimality of $\#C$. If $\#C' = \#C$, then $C' = C$ by the definition of U. But then it is easy to adapt the proof of point (1) of Proposition 22 to prove that $C \in \pi(\mathcal{Z}_s)$ for all s between t and t', and so there is no shock in U between t and t'. Hence there exists at most one such shock in U: either one before t or one after t, but not both because else (t, q, C) would contradict the third point of the definition of a shock. Up to reducing U, we can then exclude (t', q', C').

Step 3: Conclusion using Proposition 31. As (t, q, C) is not isolated, there is an infinite number of (isolated) shocks in U. Without loss of generality, we can assume that there is an infinite number of shocks in U after time t. Call $i \in C$ the smallest index such that particle i is involved in an infinite number of shocks in U after time t. When $i \neq \min C$, up to reducing U and by the minimality of i, we can assume that no particle $j \in C$ with $j < i$ is involved in a shock in U after time t.

As the shocks in U involving i after time t are isolated (Step 2), we can enumerate their times in decreasing order $(t_p)_{p\in\mathbb{N}}$. The boundedness of $\mathcal{Z}$ together with (55) allows us to take M as an upper bound for $\ddot{z}_i$ between the times of shocks. For all $p \in \mathbb{N}$ and $s \in (t_{p+1}, t_p)$, taking α as in (56), we have

$$\begin{aligned}\dot{z}_i(s) &= \dot{z}_i(t_0-) + \sum_{k=1}^{p} \left\{ \dot{z}_i(t_k-) - \dot{z}_i(t_k+) - \int_{t_{k-1}}^{t_k} \ddot{z}_i(\tau)\,\mathrm{d}\tau \right\} - \int_{t_p}^{s} \ddot{z}_i(\tau)\,\mathrm{d}\tau \\ &\geq \dot{z}_i(t_0-) + p\alpha - M(t_0 - t),\end{aligned}$$

which contradicts Step 1 as soon as p is sufficiently large. □

Acknowledgment

Ambrosio acknowledges the support of the PRIN 2017 project "Gradient Flows, Optimal Transport and Metric Measure Structures".

References

[Ambrosio and Gangbo 2008] L. Ambrosio and W. Gangbo, "Hamiltonian ODEs in the Wasserstein space of probability measures", *Comm. Pure Appl. Math.* **61**:1 (2008), 18–53. MR Zbl

[Ambrosio et al. 2005] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Birkhäuser, Basel, 2005. MR Zbl

[Ambrosio et al. 2021] L. Ambrosio, A. Baradat, and Y. Brenier, "Γ-convergence for a class of action functionals induced by gradients of convex functions", *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* **32**:1 (2021), 97–108. MR Zbl

[Brenier 1987] Y. Brenier, "Décomposition polaire et réarrangement monotone des champs de vecteurs", *C. R. Acad. Sci. Paris Sér. I Math.* **305**:19 (1987), 805–808. MR Zbl

[Brenier 1991] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions", *Comm. Pure Appl. Math.* **44**:4 (1991), 375–417. MR Zbl

[Brenier 2011] Y. Brenier, "A modified least action principle allowing mass concentrations for the early universe reconstruction problem", *Confluentes Math.* **3**:3 (2011), 361–385. MR Zbl

[Brenier 2016] Y. Brenier, "A double large deviation principle for Monge–Ampère gravitation", *Bull. Inst. Math. Acad. Sin. (N.S.)* **11**:1 (2016), 23–41. MR Zbl

[Brenier and Grenier 1998] Y. Brenier and E. Grenier, "Sticky particles and scalar conservation laws", *SIAM J. Numer. Anal.* **35**:6 (1998), 2317–2328. MR Zbl

[Brenier and Loeper 2004] Y. Brenier and G. Loeper, "A geometric approximation to the Euler equations: the Vlasov–Monge–Ampère system", *Geom. Funct. Anal.* **14**:6 (2004), 1182–1218. MR Zbl

[Brenier et al. 2003] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskiĭ, "Reconstruction of the early universe as a convex optimization problem", *Monthly Notices Roy. Astronom. Soc.* **346**:2 (2003), 501–524.

[Brenier et al. 2013] Y. Brenier, W. Gangbo, G. Savaré, and M. Westdickenberg, "Sticky particle dynamics with interactions", *J. Math. Pures Appl.* (9) **99**:5 (2013), 577–617. MR Zbl

[Brézis 1973] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Math. Stud. **5**, North-Holland, Amsterdam, 1973. MR Zbl

[de Broglie 1927] L. de Broglie, "La mécanique ondulatoire et la structure atomique de la matière et du rayonnement", *J. Phys. Radium* **8** (1927), 225–241. Zbl

[de Broglie 1959] L. de Broglie, "L'interprétation de la mécanique ondulatoire", *J. Phys. Radium* (8) **20** (1959), 963–979. MR Zbl

[Conforti 2019] G. Conforti, "A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost", *Probab. Theory Related Fields* **174**:1-2 (2019), 1–47. MR Zbl

[Dembo and Zeitouni 1998] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd ed., Appl. Math. (New York) **38**, Springer, 1998. MR Zbl

[E et al. 1996] W. E, Y. G. Rykov, and Y. G. Sinai, "Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics", *Comm. Math. Phys.* **177**:2 (1996), 349–380. MR Zbl

[Freidlin and Wentzell 1998] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, 2nd ed., Grundlehren der Math. Wissenschaften **260**, Springer, 1998. MR Zbl

[Frisch et al. 2002] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski, "A reconstruction of the initial conditions of the universe by optimal mass transportation", *Nature* **417** (2002), 260–262.

[Gentil et al. 2017] I. Gentil, C. Léonard, and L. Ripani, "About the analogy between optimal transport and minimal entropy", *Ann. Fac. Sci. Toulouse Math.* (6) **26**:3 (2017), 569–601. MR Zbl

[Gigli and Tamanini 2020] N. Gigli and L. Tamanini, "Benamou–Brenier and duality formulas for the entropic cost on $\mathrm{RCD}^*(K, N)$ spaces", *Probab. Theory Related Fields* **176**:1-2 (2020), 1–34. MR Zbl

[Hynd 2020] R. Hynd, "A trajectory map for the pressureless Euler equations", *Trans. Amer. Math. Soc.* **373**:10 (2020), 6777–6815. MR Zbl

[Madelung 1927] E. Madelung, "Quantentheorie in hydrodynamischer Form", *Z. Phys.* **40** (1927), 322–326. Zbl

[Mérigot 2011] Q. Mérigot, "A multiscale approach to optimal transport", *Comput. Graphics Forum* **30**:5 (2011), 1583–1592.

[Monsaingeon et al. 2023] L. Monsaingeon, L. Tamanini, and D. Vorotnikov, "The dynamical Schrödinger problem in abstract metric spaces", *Adv. Math.* **426** (2023), art. id. 109100. MR Zbl

[Natile and Savaré 2009] L. Natile and G. Savaré, "A Wasserstein approach to the one-dimensional sticky particle system", *SIAM J. Math. Anal.* **41**:4 (2009), 1340–1365. MR Zbl

[von Renesse 2012] M.-K. von Renesse, "An optimal transport view of Schrödinger's equation", *Canad. Math. Bull.* **55**:4 (2012), 858–869. MR Zbl

[Strömberg 1996] T. Strömberg, *The operation of infimal convolution*, Dissertationes Math. **352**, Polish Acad. Sci., Warsaw, 1996. MR Zbl

[Villani 2003] C. Villani, *Topics in optimal transportation*, Grad. Stud. in Math. **58**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl

[Zeldovich 1970] Y. B. Zeldovich, "Gravitational instability: an approximate theory for large density perturbations", *Astron. Astrophys.* **5** (1970), 84–89.

Received 8 Sep 2020. Revised 17 Jan 2022. Accepted 9 Apr 2022.

LUIGI AMBROSIO: luigi.ambrosio@sns.it
Scuola Normale Superiore, Pisa, Italy

AYMERIC BARADAT: baradat@math.univ-lyon1.fr
Université Lyon, CNRS, Université Claude Bernard Lyon 1, UMR5208, Institut Camille Jordan, Villeurbanne, France

YANN BRENIER: brenier@dma.ens.fr
École Normale Supérieure, Paris, France

Analysis & PDE

msp.org/apde

See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2023 is US $405/year for the electronic version, and $630/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Analysis & PDE

Volume 16 No. 9 2023
