MONGE-AMPÈRE GRAVITATING FLUIDS. LEAST ACTION PRINCIPLES AND PARTICLE SYSTEMS

CHRISTIAN LÉONARD [1,2] AND ROYA MOHAYAEE [3,4]

ABSTRACT. The Monge-Ampère gravitation theory (MAG) was introduced by Brenier [8] to obtain an approximate solution of the early Universe reconstruction problem. It is a modification of Newtonian gravitation which is based on quadratic optimal transport. Later, Brenier [9], then Ambrosio, Baradat and Brenier [2] discovered a double large deviation principle for Brownian particles whose rate function is precisely MAG's action functional.

In the present article, following Brenier we first recap MAG's theory. Then, we slightly extend it from particles to fluid. This allows us to revisit the Ambrosio-Baradat-Brenier particle system and propose another one which is easier to interpret and whose large deviation rate function is MAG's action functional for fluids.

This model leads to a Gibbs conditioning principle that is an entropy minimization problem close to the Schrödinger problem. While the setting of the Schrödinger problem is a system of noninteracting particles, the particle system we work with is subject to some branching mechanism which regulates the thermal fluctuations and some quantum force which balances them.

This article is dedicated to my long-time friend Patrick Cattiaux, on the occasion of his (official) retirement. CL.

CONTENTS

[1]MODAL'X, UNIV. PARIS NANTERRE, CNRS, F92000 NANTERRE FRANCE

[2]INRIA, PARIS, FRANCE

[3]SORBONNE UNIVERSITÉ, CNRS, INSTITUT D'ASTROPHYSIQUE DE PARIS, FRANCE

[4]RUDOLF PEIERLS CENTRE FOR THEORETICAL PHYSICS, UNIV. OF OXFORD, UNITED KINGDOM

E-mail addresses: christian.leonard@math.cnrs.fr, mohayaee@iap.fr.

Date: February, 2025.

2010 *Mathematics Subject Classification.* 49Q22, 60F10, 70F45, 85A40.

Key words and phrases. Optimal transport, Otto-Wasserstein manifold, large deviations, Schrödinger problem, early Universe reconstruction.

This research has been conducted within the FP2M federation (CNRS FR 2036).

Introduction

The Monge-Ampère gravitation (MAG) theory was introduced by Brenier in [8] to obtain an approximate solution of the early Universe reconstruction (EUR) problem. It is a modification of Newtonian gravitation which is based on quadratic optimal transport. Its action functional is, for any path $\omega = (\omega_s)_{s_0 \le s \le s_1}$ in $(\mathbb{R}^d)^k$,

$$A(\omega) := \int_{s_0}^{s_1} \frac{1}{2} \|\dot{\omega}_s - \mathsf{v}_s(\omega_s)\|^2 \kappa_s \, ds,$$

for some specific velocity vector field $\mathsf{v}_s(x)$, some positive number κ_s whose inverse κ_s^{-1} is a diffusion coefficient and where as usual $\dot{\omega}_s := d\omega_s/ds$ stands for the velocity of ω. Any element of $(\mathbb{R}^d)^k$ should be interpreted as a cloud of k particles living in the configuration space $\mathbb{R}^d$. Let us call any element of $(\mathbb{R}^d)^k$ a k-mapping[1]. See (3.9) for the exact formulation of the action functional.

A short note about action functionals and their link with the equations of motion is proposed at Appendix B.

The coexistence of k particles is essential to convey (semi-discrete) optimal transport into the model.

A double large deviation principle. In two subsequent articles by Brenier, Ambrosio and Baradat [9, 2], MAG's action functional was interpreted as the rate function of some double large deviation principle involving Brownian trajectories. More precisely, it happens that

$$A = \Gamma\text{-}\lim_{\epsilon \to 0} A^\epsilon, \quad \text{where} \quad A^\epsilon(\omega) = \int_{s_0}^{s_1} \frac{1}{2} \|\dot{\omega}_s - \mathsf{v}^\epsilon_s(\omega_s)\|^2 \kappa_s \, ds,$$

and v^ϵ is the current velocity of a rescaled Brownian motion in $(\mathbb{R}^d)^k$

$$\mathsf{X}^\epsilon_s = \mathsf{X}_{s_0} + \sqrt{\epsilon}\, \mathsf{B}_s, \quad s_0 \le s \le s_1,$$

see (3.8). Since the time marginal flow $(r^\epsilon_s)_{s_0 \le s \le s_1}$ of X^ϵ solves the heat equation, we have

$$\partial_s r^\epsilon - \epsilon \Delta r^\epsilon / 2 = \partial_s r^\epsilon + \nabla \cdot (r^\epsilon \mathsf{v}^\epsilon) = 0, \tag{0.1}$$

with $\mathsf{v}^\epsilon_s = -\epsilon \nabla \log \sqrt{r^\epsilon_s}$. In [2], the stochastic differential equation in $(\mathbb{R}^d)^k$,

$$d\mathsf{Z}^{\epsilon,\eta}_s = \mathsf{v}^\epsilon_s(\mathsf{Z}^{\epsilon,\eta}_s)\, ds + \sqrt{\eta \kappa_s^{-1}}\, d\mathsf{W}_s, \quad 0 < s_0 \le s \le s_1, \tag{0.2}$$

is introduced, where W is another $(\mathbb{R}^d)^k$-valued Brownian motion, see (3.7). It is known that this collection of random evolutions obeys the Freidlin-Wentzell large deviation principle when the parameter η tends to zero, ϵ being fixed,

$$\text{Proba}(\mathsf{Z}^{\epsilon,\eta} \in \bullet) \underset{\eta \to 0}{\asymp} \exp\left(-\eta^{-1} \inf_{\mathsf{z} \in \bullet} A^\epsilon(\mathsf{z})\right)$$

[1]For the time being, one can visualize a k-mapping as a "k-cloud". The reason for interpreting this cloud as some mapping will appear later in relation with optimal transport, see Definition 2.24.

with rate function A^ϵ, [17, 14]. Therefore, taking into account the already mentioned limit: $\Gamma\text{-}\lim_{\epsilon\to 0} A^\epsilon = A$, we see that *letting $\eta \to 0$, then $\epsilon \to 0$, the family of Brownian diffusion processes $\mathsf{Z}^{\epsilon,\eta}$ obeys the double large deviation principle*

$$\text{Proba}(\mathsf{Z}^{\epsilon,\eta} \in \bullet) \underset{\eta\to 0,\epsilon\to 0}{\asymp} \exp\left(-(\epsilon\eta)^{-1} \inf_{\mathsf{z}\in\bullet} A(\mathsf{z})\right)$$

with MAG's action functional A as its rate function. This also implies the Gibbs conditioning principle

$$\text{Proba}(\mathsf{Z}^{\epsilon,\eta} \in \bullet \mid \mathsf{Z}^{\epsilon,\eta}_{s_0} = \mathsf{z}_0, \mathsf{Z}^{\epsilon,\eta}_{s_1} = \mathsf{z}_1) \underset{\eta\to 0,\epsilon\to 0}{\asymp} \exp\left(-(\epsilon\eta)^{-1} \inf_{\mathsf{z}\in\bullet, \mathsf{z}(s_0)=\mathsf{z}_0, \mathsf{z}(s_1)=\mathsf{z}_1} A(\mathsf{z})\right),$$

which in turns implies that, conditionally on $\mathsf{z}(s_0) = \mathsf{z}_0$ and $\mathsf{z}(s_1) = \mathsf{z}_1$, the most likely path of $\mathsf{Z}^{\epsilon,\eta}$ as $\eta \to 0$, then $\epsilon \to 0$, solves MAG's least action principle

$$\inf_{\mathsf{z}:\mathsf{z}(s_0)=\mathsf{z}_0, \mathsf{z}(s_1)=\mathsf{z}_1} A(\mathsf{z}).$$

The stochastic evolution of $\mathsf{Z}^{\epsilon,\eta}$ is interpreted in [2] as a cloud of k Brownian particles *surfing the heat wave.*

Aim of the article. The main goal of the present article is to revisit this interpretation. Indeed, the above model does not provide a clear physical picture. In particular, the *forward* velocity vector field v^ϵ of $\mathsf{Z}^{\epsilon,\eta}$ happens to be the *current* velocity vector field of the *other* diffusion process: X^ϵ. This enigmatic substitution is uneasy to interpret.

Another particle system. In the present paper, we drop the particle system $\mathsf{Z}^{\epsilon,\eta}$. But we keep [2]'s key idea of working with X^ϵ, because it establishes a crucial link with optimal transport: a central feature of MAG. We are going to investigate the large deviations of the empirical process of a family of branching Brownian particles, as their number tends to infinity. In absence of an extra force field, each branch is a copy of X^ϵ. But to arrive at MAG, it is necessary that the whole system is immersed in some additional quantum force field.

From clouds to fluids. Passing to the limit as the number of particles tends to infinity, one does not work anymore with k-mappings, but with fluids, i.e. probability measures on $\mathbb{R}^d$. In fact, we shall look at probability measures on $(\mathbb{R}^d)^k$, i.e. fluids of k-mappings. We call these probability measures: k-fluids. Keeping k-mappings is essential to trace the effect of optimal transport. However, the relevant system to be observed is not the k-fluid, which contains the full details of the history of the k particles that are necessary to play with optimal transport, but its one-particle marginal measure on $\mathbb{R}^d$.

We do not perform the limit $\epsilon \to 0$. Neither do we look at the limit $k \to \infty$, nor at the one-particle marginal projection of the k-fluid, leaving these steps to a forthcoming investigation. We only look at the fluid analog of the above action functional A^ϵ, which is

$$(p_s)_{s_0\le s\le s_1} \mapsto \int_{s_0}^{s_1} \frac{1}{2}\|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\, \kappa_s\, ds, \tag{0.3}$$

where $(p_s)_{s_0\le s\le s_1}$ is a k-fluid-valued path, $\dot{\mathsf{p}}$ is its velocity in the Otto-Wasserstein (OW) manifold, $\|\bullet\|_p$ is the norm of the tangent vector space at p and

$$\dot{\mathsf{r}}^\epsilon_s = \mathsf{v}^\epsilon_s = -\epsilon\nabla \log \sqrt{r^\epsilon_s}$$

is the current velocity of the heat flow $(r^\epsilon_s)_{s_0\le s\le s_1}$ – recall (0.1) – in the OW-manifold. We call the corresponding model: ϵ-MAG for k-fluid.

Branching particles. Let us first briefly describe the branching mechanism; the quantum force will come later. We look at an empirical process $(\overline{X}^N_s)_{s_0\le s\le s_1}$ such that for any s, $\overline{X}^N_s$ is the empirical measure of $\lfloor \kappa_s N\rfloor$ particles in $(\mathbb{R}^d)^k$ ($\lfloor a\rfloor$ denotes the integer part of the real number a). The factor κ_s is analogous to κ_s in formula (0.2). The trajectory of each particle is a copy of X^ϵ, but these copies are not independent. As $s\mapsto\kappa_s$ is an increasing function, during any small time interval $[s,s+h]$, a fraction $\kappa'_s\,h+o_{h\to 0}(h)$ of the particles branch: each of them gives birth to a new particle starting at the same place as its genitor and evolving in the future according to the kinematics of X^ϵ and independently of the other particles. Although the number of particles increases with time, $\overline{X}^N$ is normalized so that its total mass remains constant: $\overline{X}^N_s(\mathbb{R}^{dk})=1$ for all s. As a consequence, the random fluctuation of $s\mapsto\overline{X}^N_s$ decreases with time. *This branching mechanism acts as a cooling.*

The large deviation rate function of $(\overline{X}^N)_{N\ge 1}$ as N tends to infinity leads us to the action functional

$$(p_s)_{s_0\le s\le s_1}\mapsto\int_{s_0}^{s_1}\tfrac12\|\dot{\mathsf{p}}_s-\dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\kappa_s\,ds+\epsilon^2\int_{s_0}^{s_1}I(p_s|r^\epsilon_s)\kappa_s\,ds, \tag{0.4}$$

where $I(p|r):=\displaystyle\int_{\mathbb{R}^{dk}}\frac12\Big|\nabla\log\sqrt{\frac{dp}{dr}}\Big|^2\,dp$ is the Fisher information of p with respect to r.

Schrödinger problem. Without any branching, that is if the function κ is equal to 1, $\overline{X}^N$ is the usual empirical process of an iid sample of X^ϵ and the action functional becomes

$$(p_s)_{s_0\le s\le s_1}\mapsto\int_{s_0}^{s_1}\tfrac12\|\dot{\mathsf{p}}_s-\dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\,ds+\epsilon^2\int_{s_0}^{s_1}I(p_s|r^\epsilon_s)\,ds. \tag{0.5}$$

It corresponds to the dynamics of an entropic interpolation, i.e. the time marginal flow $(p_s=P_s)_{s_0\le s\le s_1}$ of the solution P of the Schrödinger problem

$$\inf_{P:P_{s_0}=p_{s_0},P_{s_1}=p_{s_1}}H(P|R^\epsilon),$$

where R^ϵ is the law of X^ϵ, P is a path measure and

$$H(P|R^\epsilon):=\int\log(dP/dR^\epsilon)\,dP$$

is the relative entropy of P with respect to R^ϵ. The "branching modification" that is described above permits to pass from (0.5) to (0.4).

Quantum force. To arrive at MAG's action functional (0.3), it suffices to subtract $\epsilon^2\int_{s_0}^{s_1}I(p|r^\epsilon_s)\kappa_s\,ds$ from (0.5). It is convenient to express this operation by means of the corresponding Newton equation of motion in the OW-manifold. Also, things are clearer after doing some change of time, see the parameter setting 3.10, to arrive at

$$(q_t)_{t_0\le t\le t_1}\mapsto\int_{t_0}^{t_1}\tfrac12\|\dot{\mathsf{q}}_t-\dot{\mathsf{m}}^\epsilon_t\|^2_{q_t}\,dt \tag{0.6}$$

$$(q_t)_{t_0\le t\le t_1}\mapsto\int_{t_0}^{t_1}\tfrac12\|\dot{\mathsf{q}}_t-\dot{\mathsf{m}}^\epsilon_t\|^2_{q_t}\,dt+\epsilon^2\int_{t_0}^{t_1}\sigma_t I(q_t|m^\epsilon_t)\,dt, \tag{0.7}$$

instead of (0.3) and (0.4). Here $\dot{\mathsf{m}}^\epsilon_t$ is the velocity of the time-changed heat flow (m^ϵ_t) in the OW-manifold, see (5.7) and (5.8), and $t\mapsto\sigma_t$ is some positive function, see (9.2).

We show that the Newton equation corresponding to ϵ-MAG's action functional (0.6) is

$$\ddot{\mathsf{q}}_t = \ddot{\mathsf{m}}^\epsilon_t. \tag{0.8}$$

This is the equation of motion of ϵ-MAG for a k-fluid.
In terms of an equation of motion, subtracting the Fisher information term from the action functional (0.7) arising from the large deviations of $\overline{X}^N$, amounts to apply the force field

$$-\epsilon^2 \sigma_t \operatorname{grad}^{\mathrm{OW}}_{q_t} I(\bullet|m^\epsilon_t).$$

It is a quantum force field, because we show (informally) that any solution Ψ (wave function) of the nonlinear Schrödinger equation

$$\big(- i\hbar\partial_t - \hbar^2\Delta/2 - \hbar^2\mathcal{Q}(m^\epsilon_t|\mathrm{Leb})\big)\Psi + (\sigma_t\epsilon^2 - \hbar^2)\mathcal{Q}(|\Psi|^2|m^\epsilon_t)) = 0, \tag{0.9}$$

is such that $q = |\Psi|^2$ solves the Newton equation

$$\ddot{\mathsf{q}}_t = -\epsilon^2 \sigma_t \operatorname{grad}^{\mathrm{OW}}_{q_t} I(\bullet|m^\epsilon_t). \tag{0.10}$$

Here $\hbar$ is Planck's constant and $\mathcal{Q}(\bullet|\mathrm{Leb})$ and $\mathcal{Q}(\bullet|m^\epsilon_t)$ are quantum potentials defined at (9.11). This derivation is inspired from von Renesse's article [34].

Picking up the pieces. We conclude that the equation of motion, as N tends to infinity, of the branching Brownian particle system $\overline{X}^N$, immersed in the above quantum force field, is ϵ-MAG's equation of motion (0.8).

Stepping back to model (0.2), the parameter N replaces η, the branching mechanism replaces the parameter κ in the factor $\sqrt{\eta\kappa_s^{-1}}\, d\mathsf{W}_s$, and the quantum force field supersedes the enigmatic substitution of the forward velocity of $\mathsf{Z}^{\epsilon,\eta}$ by the the current velocity of X^ϵ.

Outline of the article. Section 1 begins with a short description of the physical basis of the Monge-Ampère gravitation theory. We also sketch the early Universe reconstruction problem as the main motivation for this modified theory of gravitation. Section 2 is dedicated to the exposition of the mathematics of MAG, as introduced by Brenier in [8], and Section 3 gives the details about the model (0.2) which is the main object of the article [2] by Ambrosio, Baradat and Brenier. At Section 4, we introduce the analogs for fluids of the action functionals A and A^ϵ, such as (0.3). This is the place where some basic material about the Otto-Wasserstein geometry is gathered; some more is presented at the beginning of Section 6. The action functional (0.3)/(0.6) we mainly work with is introduced at Section 5.

Newton's equation of motion (0.8) is proved at Section 6. A detailed description of the acceleration $\ddot{\mathsf{q}}$ is also obtained at Theorem 6.2. It reveals divergences of the force field at some specific places which are responsible for the concentration of matter.

The last three sections are dedicated to the construction of the particle system that was presented during this introductory section. At Section 7, we recall already known facts about the dynamics of the solutions of the Schrödinger problem. This leads us to the action functional (0.5). Then, we introduce the factor κ_s, at Section 8 to arrive at the action functional (0.4). This is the place where we prove that the branching mechanism that was previously described transforms (0.5) into (0.4). Finally, at Section 9, the quantum force (0.10) is shown to be associated with the nonlinear Schrödinger equation (0.9).

Sections 1, 2, 3 and 7 are expository: most of their content is already known. The rest of the article consists of new material.

This program of investigation is far from being complete: a short list of remaining things to be done is proposed at Appendix A. We also provide at Appendix B a basic reminder about the minimization of action functionals, and at Appendix C some simple analogies are presented to illustrate the concentration of matter that results from MAG's dynamics.

1. MAG. Motivation and definition

Newtonian gravitation. The equation of motion of a test particle in gravitational interaction with a fluid is

$$\ddot{x}_t = -\nabla\varphi_t(x_t), \quad t \geq 0, \tag{1.1}$$

where the scalar potential φ solves the Poisson equation

$$\Delta\varphi_t = \mu_t. \tag{1.2}$$

Here, $x_t \in \mathbb{R}^d$ is the position of the test particle at time t, $\ddot{x}_t$ is its acceleration and $\mu_t(x)$ is the density of the fluid at time t and location x. While Newton's equation (1.1) is physically correct (the mass of the test particle is irrelevant), for simplicity we do not write the gravitational constant in (1.2).

In computational cosmology, the configuration space $\mathbb{R}^d$ is replaced by the flat torus

$$\mathbb{T}^d_L = \mathbb{R}^d/[0,L]^d$$

with size $L > 0$. Assuming that $\mu_t(\mathbb{T}^d_L) := \int_{\mathbb{T}^d_L} \mu_t(x)\,dx < \infty$, without loss of generality one can normalize μ_t as a probability measure.
Periodic boundary conditions imply that for any regular $[0,L]^d$-periodic function φ

$$\begin{aligned}\int_{[0,L]^d} \Delta\varphi\, d\mathrm{Leb} &= \int_{\mathbb{R}^d} \mathbf{1}_{[0,L]^d} \nabla\cdot(\nabla\varphi)\, d\mathrm{Leb} \\ &= -\int_{\mathbb{R}^d} \nabla\mathbf{1}_{[0,L]^d}\cdot\nabla\varphi\, d\mathrm{Leb} = \int_{\partial[0,L]^d} \nabla\varphi\cdot\vec{n}\, d\sigma = 0\end{aligned}$$

where Leb stands for the Lebesgue measure, $\vec{n}$ is the outer unit normal vector on the boundary $\partial[0,L]^d$ of $[0,L]^d$ equipped with the surface measure $\sigma = \mathrm{Leb}_{|\partial[0,L]^d}$. This is Stokes formula and its vanishing means that the mass of any distribution of matter evolving along the integral curves of the vector field $\nabla\varphi$ remains constant since $\mathbb{T}^d_L$ has no physical boundary.
Hence, for a Poisson equation in $\mathbb{T}^d_L$ to admit a solution it is necessary that its right hand term has a zero mass. Let us replace equation (1.2) by $\Delta\varphi_t = \mu_t - \lambda$ where λ is some probability measure on $\mathbb{T}^d_L$. The best natural choice for λ is the uniform unit volume measure on the torus

$$\lambda_L = L^{-d}\,\mathrm{Leb}_{|\mathbb{T}^d_L}.$$

Indeed, doing this we see that, with the balanced Poisson equation

$$\Delta\varphi_t = \mu_t - \lambda_L \quad \text{on } \mathbb{T}^d_L, \tag{1.3}$$

a uniform distribution of matter: $\mu_t = \lambda_L$, does not generate any gravitational force, as expected. Moreover, it successfully passes the "large box"-test

$$\lim_{L\to\infty} (1.3) = (1.2). \tag{1.4}$$

Indeed, letting L tend to infinity, one recovers the standard Poisson equation (1.2) in $\mathbb{R}^d$ because the background source term λ_L vanishes as L tends to infinity. For further physical justification of this model in cosmology, see [21, 5].

Monge-Ampère gravitation (MAG). Brenier [8] introduced a modified theory of gravitation which, although not fundamental, is effective for solving the *early Universe reconstruction* (EUR) problem. The main feature of this modified theory consists in replacing the Poisson equation (1.3) by the Monge-Ampère equation

$$L^{-d}\det(\mathbb{I} + \mathrm{Hess}(L^d\varphi_t)) = \mu_t, \quad \text{on } \mathbb{T}_L^d, \tag{1.5}$$

where $\mathbb{I}$ is the identity matrix. The couple of equations (1.1) and (1.5) is called *Monge-Ampère gravitation*, MAG for short. Note that (1.5) admits (1.3) as its linearization in the limit of weak gravitation, that is

$$\lim_{\|\,\mathrm{Hess}\,\varphi\| \to 0} (1.5) = (1.3),$$

or more precisely $\Delta\varphi_t = \mu_t - \lambda_L + o_{\|\,\mathrm{Hess}\,\varphi_t\| \to 0}(\|\,\mathrm{Hess}\,\varphi_t\|)$, L being fixed. Moreover, it is exactly (1.3) in dimension one. But in dimension $d \geq 2$, it fails the "large box"-test:

$$\lim_{L \to \infty} (1.5) \neq (1.2).$$

For instance, in dimension $d = 3$,

$$\begin{aligned} L^{-3}\det(\mathbb{I} + \mathrm{Hess}(L^3\varphi)) &= L^{-3}\det(\mathbb{I} + L^3\mathrm{diag}(a,b,c)) = L^{-3}(1 + L^3a)(1 + L^3b)(1 + L^3b) \\ &= L^{-3} + \Delta\varphi + L^3(ab + ac + bc) + L^6abc, \end{aligned} \tag{1.6}$$

where a, b and c are the eigenvalues of $\mathrm{Hess}\,\varphi$. We see that the dominating term as L tends to infinity is not $\Delta\varphi$ as desired, but $L^6\det(\mathrm{Hess}\,\varphi)$. This remains true in any dimension where the dominating term is $L^{d^2-d}\det(\mathrm{Hess}\,\varphi)$.
Moreover, when $d \geq 2$, there is no mixed asymptotic regime $(L \to \infty, \|\,\mathrm{Hess}\,\varphi\| \to 0)$, where (1.3)&(1.5) is an approximation of Newtonian gravity which would be valid for *any* φ. Let us show it. Denoting $H := \|\,\mathrm{Hess}\,\varphi\|$, we see that

$$L^{-d}\det(\mathbb{I} + \mathrm{Hess}(L^d\varphi)) = \Delta\varphi + L^{-d} + L^{-d}\sum_{2 \leq n \leq d} O((L^dH)^n).$$

A regime where $\Delta\varphi$ would be the dominating term, must satisfy

$$H^{-1}\Big(L^{-d} + L^{-d}\sum_{2 \leq n \leq d} O((L^dH)^n)\Big) \longrightarrow 0,$$

because $\Delta\varphi = O(H) \to 0$. Since this term is of order $(L^dH)^{-1} + \sum_{1 \leq n \leq d-1}(L^dH)^n$, this would imply that L^dH tends simultaneously to ∞ and 0, a contradiction. □

Nevertheless, despite this negative result, MAG reveals to be effective for approximately solving the early Universe reconstruction problem. This is illustrated by Figure 2 and will be partly explained during a short discussion at page 8.

Early Universe reconstruction (EUR). This problem was addressed by Peebles in the seminal paper [30]. One specific feature with cosmology is that the distribution of matter/energy of the very early Universe was highly uniform as testified by the observation of the cosmic microwave background which exhibits a relative fluctuation from uniformity of order 10^{-4}: typically $3K \pm 300\mu K$ (the unit K is a Kelvin degree).

This very peculiar property of the initial condition is one reason for the Monge-Ampère strategy to be successful when solving EUR. However, the other reasons for its effectiveness are not fully understood at present time. We hope that this article will be a step towards a clearer picture.

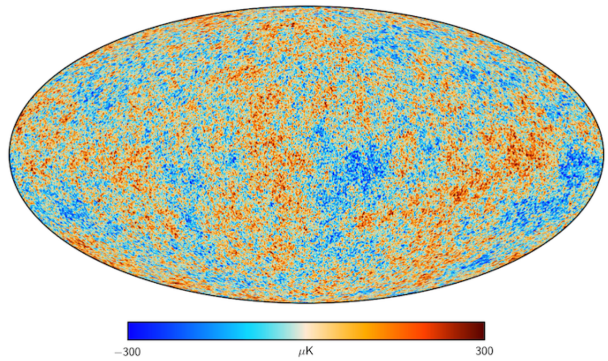


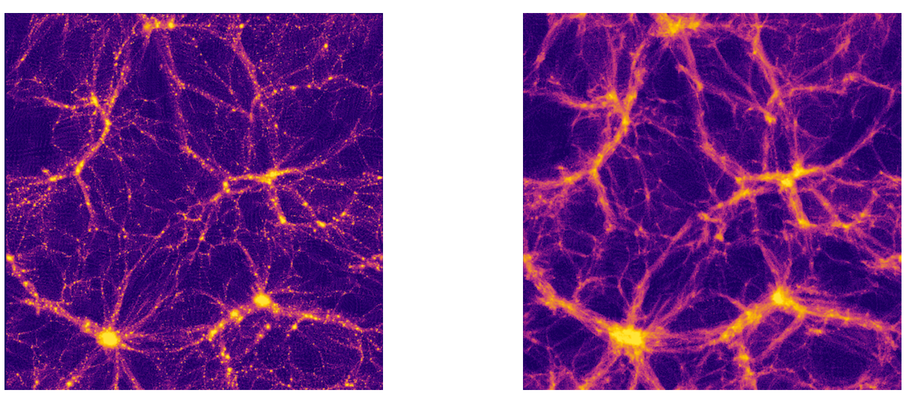
FIGURE 1. Cosmic microwave background

One aim of EUR is to give estimates of the field of very early fluctuations from this uniformity

$$\mu_0'(x) := \lim_{t \to 0^+} \frac{\mu_t(x) - L^{-d}}{t},$$

a crucial information to test the cosmic inflation theory and provide details about the initial quantum fluctuations.

An effective theory in computational cosmology. As an illustration of the good performance of MAG in cosmology, we provide at Figure 2 a couple of images representing typical structures of the actual Universe. Both are obtained by running numerical simulations starting from the same initial condition: a tiny perturbation of the uniform measure on $\mathbb{T}_L^3$. The left-hand side image (A) is obtained using the balanced Poisson equation (1.3), while the right-hand side (B) results from using the Monge Ampère equation (1.5) instead of the physically more realistic law (1.3).



(A) Poisson

(B) Monge-Ampère

FIGURE 2. Typical structures of the actual Universe

Courtesy of Bruno Lévy

MAG works well with EUR. Let us discuss a little bit about MAG being a good approximation of the Newtonian gravity in the special setting of the EUR problem. The symmetric operator $\operatorname{Hess}\varphi$ admits an orthogonal basis of eigenvectors. Let $K := \{x \in \mathbb{R}^3; \operatorname{Hess}\varphi(x) = 0\}$ denotes its kernel, and $K^\perp$ be K's orthogonal subspace in $\mathbb{R}^3$. Suppose that $\operatorname{Hess}\varphi$ admits at least one zero eigenvalue and at least one non-zero one. This means that both K and $K^\perp$ are eigenspaces and that $(\dim(K), \dim(K^\perp))$ is either $(1, 2)$

or $(2,1)$. Since $\operatorname{Hess}\varphi$ is the Jacobian matrix $\operatorname{Jac}(\nabla\varphi)$ of the opposite of the acceleration field $\ddot{x} = -\nabla\varphi$, this implies that the acceleration is locally constant in direction K, so that any concentration of matter in direction $K^\perp$ is pushed by a force field in direction K, possibly fostering the formation of a structure of dimension $\dim(K)$.

(1) At places where $\operatorname{Hess}\varphi$ admits a single zero-eigenvalue, that is $\dim(K) = 1$, then 1D singular structures (filaments) might appear.
(2) At places where $\operatorname{Hess}\varphi$ admits a double zero-eigenvalue, that is $\dim(K) = 2$, then 2D singular structures (sheets) might appear.

Going back to (1.6), it appears that (1.5) is exactly (1.3) if two eigenvalues of $\operatorname{Hess}\varphi$ vanish. In practice, for MAG being close to Newtonian gravitation, one eigenvalue should be very close to zero to kill the leading term L^6abc and another one should be small enough to control the intermediate term $L^3(ab + ac + bc)$.
As can be seen at Figure 2a, it happens that N-body simulations based on the Newtonian gravitation reveal that, after some time, most of the matter is concentrated in singular structures: sheets and filaments (Figure 2 depicts 2D-slices of 3D-cubes). This is a clue in favor of the good fit between these two theories in this special case. However, estimating the accuracy of MAG for solving EUR still remains a mathematical open problem.

Remark that the above classification of singular structures in terms of $\operatorname{Hess}\varphi$ differs from the standard approach by Zeldovich [36] which is based on the Jacobian of a velocity field rather than an acceleration field. The way our classification complements Zeldovich's one will be explored elsewhere.

Pro and cons. From a practical point of view, one advantage of MAG is that it provides us with faster computations than the standard Poisson-based algorithms, because of its connection with optimal transport (this will be made precise later). More about cosmological simulations using MAG and their comparison with standard N-body simulations can be found in the recent article [25] by Lévy, Brenier and the second author. From a theoretical point of view, another advantage is that its tight relation with optimal transport provides us with easy geometrical interpretations. Moreover, it is shown in the article [6] by Bonnefous, Brenier and one of us (RM) that MAG can describe a scalar field, often evoked in modified theories of gravity such as Galileons, opening a promising domain to further explore.
However, the main drawback of MAG is that it is not a fundamental theory of gravitation; it is only effective for solving EUR.

From now on we drop the size L by setting $L = 1$, and denote $\mathbb{T}^d = \mathbb{T}^d_{L=1}$.

Zeldovich approximation. In [8] Brenier transposed Peebles' EUR problem in the framework of the semi-Newtonian gravitational model of the early Universe where the trajectory $t \mapsto x_t \in \mathbb{T}^3$ of each particle (typically a cluster of galaxies!) satisfies the Newton-like equation of motion

$$\frac{2t}{3}\ddot{x}_t + \dot{x}_t = -\nabla\varphi_t(x_t), \quad t \geq 0, \tag{1.7}$$

where the potential φ solves

$$1 + t\Delta\varphi_t = \mu_t \tag{1.8}$$

and μ_t is the matter density. This describes classical Newtonian interactions taking place in an Einstein-de Sitter space corresponding to a Big Bang scenario. The couple of equations (1.7)&(1.8) is called the semi-Newtonian system (SNS).

At time $t = 0$, we see that

$$\mu_0 \equiv 1, \quad \dot{x}_0 = -\nabla\varphi_0(x_0), \quad \Delta\varphi_0(x) = \lim_{t\to 0^+} \frac{\mu_t(x) - 1}{t} =: \mu_0'(x).$$

Remarks 1.9.
(a) One is far from the classical N-body problem.
(b) Equation (1.8) is simply (1.3) with $L = 1$ and $t\varphi_t$ instead of φ_t.
(c) The physical assumption: $\mu_0 \equiv 1$, is experimentally verified up to a very high precision level, see Figure 1.

Zeldovich approximation [36] is simply

$$\widetilde{x}_t = x_0 - t\nabla\varphi_0(x_0), \qquad 0 \le t \le t_*(x_0),$$

where $0 < t_*(x_0) \le \infty$ is the first time where a collision with another particle occurs. In this scenario, the initial fluctuation of the field generates a path with constant velocity until its first collision.

Brenier proved in [8] that keeping (1.7), but replacing the balanced Poisson equation (1.8) by the Monge-Ampère equation

$$\det(\mathbb{I} + t\,\mathrm{Hess}\,\varphi_t) = \mu_t, \tag{1.10}$$

one obtains a least action principle admitting the Zeldovich approximations as its *exact* solutions. A similar reasoning will be exposed in a while to arrive at (2.19) for the MAG problem (1.1)&(1.13), see Definition 1.12 below. Note that (1.8) and (1.10) are exactly (1.3) and (1.5) with $L = 1$ and different notations for the potential.

Optimal transport versus N-body simulation. Frisch, Matarrese, Sobolevskii and the second author of the present article have shown in [18], more than twenty years ago, that with the simplified dynamics of the Zeldovich approximation, EUR is exactly the Monge quadratic optimal transport problem between the initial uniform distribution of matter $\mu_0 \equiv 1$ and the distribution of matter of the present epoch μ_T, provided that the Zeldovich map $x_0 \mapsto \widetilde{x}_T(x_0)$ is the gradient of a convex potential. See also [10] for more mathematical details. Numerical simulations in [18] also demonstrate that the solution of this quadratic optimal transport problem is close to the result of N-body simulations as performed following [21, 5] for instance. Figure 3 illustrates the comparison between a standard N-body simulation and a construction using optimal transport. More precisely, one compares the joint distributions of the couples of initial and final positions of the N-body simulation with the optimal transport plan between the initial and final marginal distributions of the N-body simulation. The yellow points of the left-hand side picture correspond to significant errors of matching. The dots near the diagonal on the right-hand side graphic are a scatter plot of reconstructed (using optimal transport) initial points vs. simulation initial points. The upper left inset is a histogram (by percentage) of distances between such points; 62% are assigned exactly (up to the grid precision). The lower right inset is a similar histogram for reconstruction on a finer grid where 34% are assigned exactly.

MAG's definition. The first definition of MAG that appeared in the literature is the following

Definition 1.11 (MAG's approximation of SNS, [8])**.** *On the torus $\mathbb{T}^d = \mathbb{R}^d/[0,1]^d$, the dynamics defined by the couple of equations* (1.7)*&*(1.10) *is the* Monge-Ampère gravitation *(MAG) system approximating the semi-Newtonian system (SNS)* (1.7)*&*(1.8).

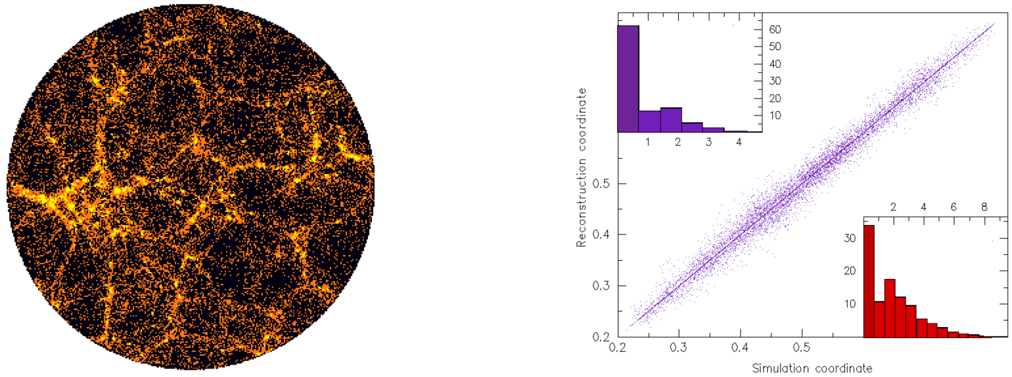


FIGURE 3. Optimal transport vs. N-body simulation, [18].

MAG's approximation of SNS is effective in the regime of short time and weak gravity.

From now on, following Brenier and his co-authors [9, 1, 2] we drop the relativistic dynamics (1.7) and go back to the usual Newton equation (1.1). With this choice the underlying physics is less realistic, but the mathematics are clearer and easier to handle. This simplification seems to us to be a reasonable choice for a preliminary approach of a more complete theory to be developed later.

Definition 1.12 (MAG on $\mathbb{T}^d$, absolutely continuous fluid, [8])**.** *The dynamics defined on the torus $\mathbb{T}^d = \mathbb{R}^d/[0,1]^d$ by the couple of equations (1.1) and*

$$\det(\mathbb{I} + \mathrm{Hess}\,\varphi_t) = \mu_t, \quad \text{on } \mathbb{T}^d, \tag{1.13}$$

is called Monge-Ampère gravitation*, MAG for short.*

Furthermore, to simplify the mathematics and the presentation, we keep Brenier's choice in [8, 2] to consider MAG in the whole space $\mathbb{R}^d$ rather than in the torus $\mathbb{T}^d$, see Definition 2.6 below. However, a severe critic of this choice is expressed at Remark 2.9-(i).

2. MAG. Action functional

This section is aimed at giving a short presentation of the Monge-Ampère gravitation theory. It describes step by step the way for justifying Brenier's Definition 2.26 of MAG's action functional.

Optimal transport. At first sight, it seems to be useless to replace the linear equation (1.3) by a nonlinear one. But, be aware that solving the gravitational problem (1.1)&(1.3), or its standard counterpart (1.1)&(1.2), remains inaccessible in most situations. On the other hand, as will be seen in a moment, the connection of the Monge-Ampère equation (1.13) with quadratic optimal transport permits a rather simple geometric interpretation of MAG which leads to a fast numerical algorithm in [8].

Following Brenier [8], let us make precise the connection of (1.13) with quadratic optimal transport. Take a probability measure λ on $\mathbb{R}^d$ and transport it by the measurable map $\overrightarrow{T} : \mathbb{R}^d \to \mathbb{R}^d$ to obtain its image

$$\mu := \overrightarrow{T}_{\#}\lambda,$$

defined by $\mu(dy) = \lambda\big((\overrightarrow{T})^{-1}(dy)\big)$. We have in mind that μ plays the role of the actual distribution of matter. The measure λ will be specified at (2.5) so that some analogue of (1.13) is satisfied, see (2.7) below.

If μ is absolutely continuous and $\overrightarrow{T}$ is injective and differentiable almost everywhere, then λ is also absolutely continuous, the inverse map $(\overrightarrow{T})^{-1} =: \overleftarrow{T}$ is differentiable μ-a.e., $\overleftarrow{T}_{\#}\mu = \lambda$ and the Monge-Ampère equation

$$\mu(y) = \lambda(\overleftarrow{T}(y)) \, |\det(\nabla \overleftarrow{T})(y)|, \quad y \in \mathbb{R}^d, \text{ a.e.} \tag{2.1}$$

is satisfied. Here and in the remainder of this article, any absolutely continuous measure and its density with respect to Lebesgue measure are denoted by the same letter.
The Monge optimal transport problem which is relevant for our purpose is

$$\inf_{\overleftarrow{T}:\overleftarrow{T}_{\#}\mu=\lambda} \int_{\mathbb{R}^d} |y - \overleftarrow{T}(y)|^2 \, \mu(dy), \tag{2.2}$$

where, as for equation (2.1), the unknown is the map $\overleftarrow{T} : \mathbb{R}^d \to \mathbb{R}^d$, while the probability measures λ and μ are prescribed. Brenier's theorem [7] tells us that when $\int_{\mathbb{R}^d} |x|^2 \, \lambda(dx) < \infty$, $\int_{\mathbb{R}^d} |y|^2 \, \mu(dy) < \infty$ and μ is absolutely continuous, (2.2) admits a unique solution $\overleftarrow{T}$. Moreover,

$$\overleftarrow{T} = \nabla\theta, \quad \text{a.e.}, \tag{2.3}$$

where $\theta : \mathbb{R}^d \to \mathbb{R}$ is a convex function. As Aleksandrov's theorem states that any convex function on $\mathbb{R}^d$ is twice differentiable almost everywhere, the Jacobian matrix $\nabla\overleftarrow{T} = \operatorname{Hess}\theta$ is well-defined a.e. and positive semi-definite. Introducing the function

$$\varphi(y) := \theta(y) - |y|^2/2, \tag{2.4}$$

with (2.3) equation (2.1) reads as $\mu(y) = \lambda(\overleftarrow{T}(y)) \det(\mathbb{I} + \operatorname{Hess}\varphi)(y)$, $y \in \mathbb{R}^d$, a.e. Finally, we observe that in the special case where the source distribution λ is uniform, that is:

$$\lambda(dx) = \mathbf{1}_D(x)\, dx \tag{2.5}$$

for some measurable subset $D \subset \mathbb{R}^d$ with a unit volume $\operatorname{Leb}(D) = 1$, (2.1) simplifies as $\mu(y) = \mathbf{1}_{\{\overleftarrow{T}(y)\in D\}} \det(\mathbb{I} + \operatorname{Hess}\varphi)(y)$, $y \in \mathbb{R}^d$, a.e. With (2.3) and (2.4), one can recast this equation for the unknown φ :

$$\mathbf{1}_{\{y+\nabla\varphi(y)\in D\}} \det(\mathbb{I} + \operatorname{Hess}\varphi)(y) = \mu(y), \quad y \in \mathbb{R}^d, \text{ a.e.}$$

It is an actualization of the balanced Poisson equation (1.13), once the configuration space $\mathbb{T}^d$ is replaced by $\mathbb{R}^d$.

Definition 2.6. (Dynamics of MAG pushed by D in $\mathbb{R}^d$. Absolutely continuous distribution of matter). *Let $D \subset \mathbb{R}^d$ satisfy* $\operatorname{Leb}(D) = 1$. *The dynamics of the* Monge-Ampère gravitation pushed by the source set D *is defined by the following system of Newton's equations* (1.1)*:*

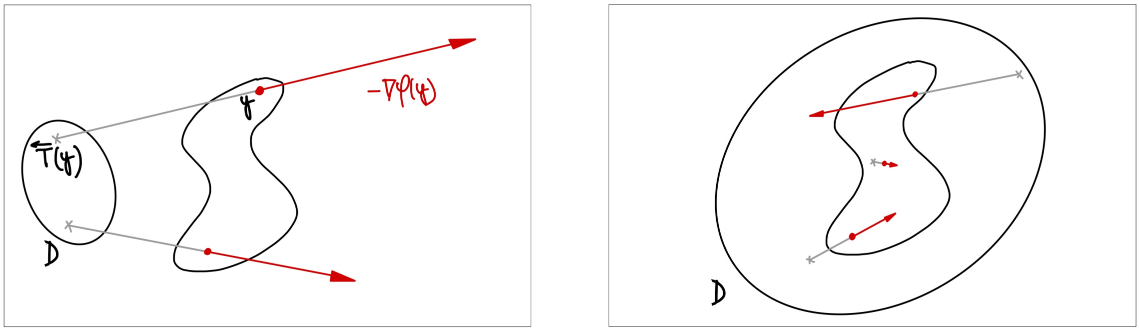
$$\begin{cases} \ddot{X}_t(\xi) = -\nabla\varphi_t(X_t(\xi)), & 0 \le t < t_*, \\ X_0(\xi) = \xi, & t = 0, \end{cases} \qquad \xi \in \operatorname{supp}(\mu_0),$$

where the initial probability measure μ_0 is assumed to be absolutely continuous, the distribution of matter

$$\mu_t = (X_t)_{\#}\mu_0, \qquad 0 \le t < t_*$$

where $t_* \in [0, \infty]$ is the infimum of all times t such that μ_t fails to be absolutely continuous, and the Monge-Ampère equation

$$\mathbf{1}_{\{y+\nabla\varphi_t(y)\in D\}} \det(\mathbb{I} + \operatorname{Hess}\varphi_t)(y) = \mu_t(y), \quad y \in \mathbb{R}^d, \ \text{a.e.}, \quad 0 \le t < t_*. \tag{2.7}$$



(A) Matter outside D (B) Matter inside D

FIGURE 4. The force field depends on D

MAG's force field. Replacing MAG's equation (1.13) by (2.7), the right hand side of Newton's equation (1.1) is

$$-\nabla\varphi(y) = y - \nabla\theta(y) = y - \overleftarrow{T}(y), \quad y \in \overrightarrow{T}(D), \text{ a.e.} \tag{2.8}$$

This is the explicit connection of MAG with quadratic optimal transport. It holds both on $\mathbb{T}^d$ and $\mathbb{R}^d$, since on $\mathbb{T}^d$ one chooses

$$D = \mathbb{T}^d.$$

Remarks 2.9.

(i) Since the force field $-\nabla\varphi$ depends on the choice of the source set D, it cannot be considered as a proxy for a physical gravitational field in $\mathbb{R}^d$. This is illustrated at Figure 4. The relative position of D and the cloud of matter plays an essential role. In particular, one sees that Figure 4a illustrates anything but a self-attractive gravitation!
However, when the configuration space is $\mathbb{T}^d$, as is usual in computational cosmology, the source measure λ is chosen to be the uniform probability measure on the whole set $\mathbb{T}^d$. Since (2.7) becomes (1.13) in this special case, this leads to MAG gravitation as defined at Definition 1.12.
Still, we go on working in $\mathbb{R}^d$ for simplicity.

(ii) Note also that $-\nabla\varphi$ is only defined (almost everywhere) on the support of the target measure μ. This is not an issue because we are only concerned with the evolution of μ_t.

(iii) When considering such an evolution, it is not necessary to assume that the initial matter density μ_0 is equal to the source measure λ specified at (2.5). Indeed, λ *should be regarded as an artefact.*

Polar factorization. The last ingredient to be introduced for a geometric interpretation of MAG is Brenier's polar factorization theorem [7]. During this subsection, λ is any *absolutely continuous* probability measure on $\mathbb{R}^d$ satisfying $\int_{\mathbb{R}^d} |x|^2\,\lambda(dx) < \infty$. The polar factorization theorem states that for any measurable mapping $\mathsf{y} : D \subset \mathbb{R}^d \to \mathbb{R}^d$ in

$$H := L^2_{\mathbb{R}^d}(D, \lambda),$$

i.e. verifying $\|\mathsf{y}\|_H^2 := \int_D |\mathsf{y}(x)|^2\,\lambda(dx) < \infty$, and such that the probability measure

$$\mu := \mathsf{y}_\#\lambda$$

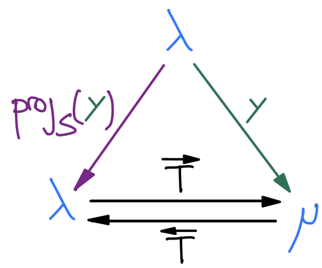
is also absolutely continuous, we have

$$\overleftarrow{T} \circ \mathsf{y} = \mathrm{proj}_S(\mathsf{y}), \tag{2.10}$$

where $\overleftarrow{T}$ is the optimal transport map from μ to λ, and $\mathrm{proj}_S : H \to H$ is the orthogonal projection in the Hilbert space H onto the subset

$$S := \{\mathsf{x} \in H : \mathsf{x}_\# \lambda = \lambda\}$$

of all λ-preserving maps.



From now on, elements in H will be written with bold letters. Note that

$$\|\mathsf{x}\|_H^2 = \int_D |\mathsf{x}(x)|^2\, \lambda(dx) = \int_D |x|^2\, \mathsf{x}_\# \lambda(dx) = \int_D |x|^2\, \lambda(dx) =: r^2 < \infty, \quad \forall \mathsf{x} \in S. \tag{2.11}$$

Hence, S is a subset of a sphere; in particular it is not convex. The uniqueness of the projection of y on S, which is implied by the assumed absolute continuity of μ, is part the statement. This theorem is often expressed the reverse way:

$$\mathsf{y} = \overrightarrow{T} \circ \mathrm{proj}_S(\mathsf{y}),$$

where $\overrightarrow{T} = (\overleftarrow{T})^{-1}$ is the optimal transport map from λ to μ. It is worth seeing $\mathrm{proj}_S(\mathsf{y})$ as a type of permutation preserving λ primer to the least cost mapping $\overrightarrow{T}$.

Geometric expression of MAG. In view of (1.1), (2.8) and (2.10), the collection of mappings $(\mathsf{y}_t; t \geq 0)$ in H is driven by MAG if

$$\ddot{\mathsf{y}}_t = \mathsf{y}_t - \mathrm{proj}_S(\mathsf{y}_t), \quad 0 \leq t < t_*. \tag{2.12}$$

It is a detailed description of the evolution of the infinite particle system $(\mathsf{y}_t(x); x \in D)_{0 \leq t < t_*}$. In particular it gives the evolution of the fluid density $\mu_t := (\mathsf{y}_t)_\# \lambda$.

Concentration of matter. However, for (2.12) to be valid it is necessary that μ_t is absolutely continuous to ensure that y_t admits a *unique* projection on S. Considering the converse of this implication, one sees that *as soon as y_t has several projections on S, the matter distribution μ_t becomes singular with respect to Lebesgue measure.* This appearance of singularities should be interpreted as concentration of matter. As S is non-convex, this happens as a rule.
Choosing *one* candidate $\widehat{\mathrm{proj}}_S(\mathsf{y}_t)$ as a function of the whole set $\mathrm{Proj}_S(\mathsf{y}_t)$ of all the orthogonal projections of y_t on S, determines the dynamics when concentration of matter occurs. A natural one is

$$\widehat{\mathrm{proj}}_S(\mathsf{y}) := \mathrm{proj}_{S(\mathsf{y})}(\mathsf{y}) \tag{2.13}$$

where $S(\mathsf{y}) := \mathrm{cl\,cv}(\mathrm{Proj}_S(\mathsf{y}))$ is the closed convex hull of $\mathrm{Proj}_S(\mathsf{y})$. For a justification of this choice, see the appendix section.

Definition 2.14 (MAG dynamics allowing for matter concentration, [8])**.** *MAG dynamics described at Definition 2.6 is only valid for an absolutely continuous distribution of matter μ_t. In view of* (2.12)*, an extension allowing for matter concentration is*

$$\ddot{\mathsf{y}}_t = \mathsf{y}_t - \widehat{\mathrm{proj}}_S(\mathsf{y}_t) \in H, \quad t \geq 0. \tag{2.15}$$

Remarks 2.16.

(a) Because of the introduction of the extension $\widehat{\mathrm{proj}}_S$ of proj_S when Proj_S contains several points, this definition allows for mass concentration: time t goes beyond t_*.
(b) It depends on the choice of the extension $\widehat{\mathrm{proj}}_S$ of proj_S.
(c) Later, it will be convienient to work with action functionals, rather than with their Euler-Lagrange solutions. One must be aware that the connection between (2.15) and the natural candidate (2.21) below for a corresponding action functional is not fully established, see Remark 2.22-(b).

Action functional. A short reminder about action functionals and their connection with the equations of motion is proposed at Appendix B.

In [8], it is emphasized that the function

$$\Phi(\mathsf{y}) := \inf_{\mathsf{x}\in S} \|\mathsf{y} - \mathsf{x}\|_H^2/2 = \|\mathsf{y}\|_H^2/2 - \Pi_S(\mathsf{y}) + r^2/2, \quad \mathsf{y} \in H,$$

where

$$\mathsf{y} \mapsto \Pi_S(\mathsf{y}) := \sup_{\mathsf{x}\in S} \langle \mathsf{x}, \mathsf{y}\rangle_H\,,$$

is differentiable at any y which admits a unique projection on S. Moreover, for any such "nice" y,

$$\nabla\Phi(\mathsf{y}) = \mathsf{y} - \nabla\Pi_S(\mathsf{y}) = \mathsf{y} - \mathrm{proj}_S(\mathsf{y}). \tag{2.17}$$

This implies

$$\Phi(\mathsf{y}) = \|\nabla\Phi(\mathsf{y})\|_H^2/2,$$

and also that a Lagrangian associated to Newton's equation (2.12) is

$$\|\dot{\mathsf{y}}\|_H^2/2 + \Phi(\mathsf{y}). \tag{2.18}$$

Plugging the last but one identity into the last one, the Lagrangian becomes $\|\dot{\mathsf{y}}\|_H^2/2 + \Phi(\mathsf{y}) = \|\dot{\mathsf{y}}\|_H^2/2 + \|\nabla\Phi(\mathsf{y})\|_H^2/2 = \|\dot{\mathsf{y}} - \nabla\Phi(\mathsf{y})\|_H^2/2 + \langle\nabla\Phi(\mathsf{y}), \dot{\mathsf{y}}\rangle_H$. As $\langle\nabla\Phi(\mathsf{y}), \dot{\mathsf{y}}\rangle_H$ is a null Lagrangian (because $\langle\nabla\Phi(\mathsf{y}_t), \dot{\mathsf{y}}_t\rangle_H = \frac{d}{dt}\Phi(\mathsf{y}_t)$ is a total derivative), one can choose the alternate Lagrangian $\|\dot{\mathsf{y}} - \nabla\Phi(\mathsf{y})\|_H^2/2$ without modifying the dynamics. Its leads to the action functional

$$\int_{t_0}^{t_1} \tfrac{1}{2}\|\dot{\mathsf{y}}_t - \nabla\Phi(\mathsf{y}_t)\|_H^2\, dt = \int_{t_0}^{t_1} \tfrac{1}{2}\|\dot{\mathsf{y}}_t - \mathsf{y}_t + \mathrm{proj}_S(\mathsf{y}_t)\|_H^2\, dt, \tag{2.19}$$

which becomes meaningless as soon as t_1 is larger than the first time t^* when the set $\mathrm{Proj}_S(\mathsf{y}_{t^*})$ contains at least two elements. Finally, Brenier proposes the following

Definition 2.20 (Action functional of MAG pushed by D, [8])**.** *It is*

$$\int_{t_0}^{t_1} \tfrac{1}{2}\|\dot{\mathsf{y}}_t - \widehat{\nabla}\Phi(\mathsf{y}_t)\|_H^2\, dt = \int_{t_0}^{t_1} \tfrac{1}{2}\|\dot{\mathsf{y}}_t - \mathsf{y}_t + \widehat{\mathrm{proj}}_S(\mathsf{y}_t)\|_H^2\, dt, \tag{2.21}$$

the extended gradient $\widehat{\nabla}\Phi(\mathsf{y})$ being defined by

$$\widehat{\nabla}\Phi(\mathsf{y}) := \mathsf{y} - \widehat{\nabla}\Pi_S(\mathsf{y}) = \mathsf{y} - \widehat{\mathrm{proj}}_S(\mathsf{y})$$

where $\widehat{\nabla}\Pi_S(\mathsf{y})$ is the (unique) element with minimal norm in the subdifferential $\partial\Pi_S(\mathsf{y})$ of the convex function Π_S.

Remarks 2.22.

(a) In view of (2.18), another candidate for an action functional of MAG is

$$\|\dot{\mathsf{y}}\|_H^2/2 + \widehat{\Phi}(\mathsf{y}).$$

Unlike this action, the action (2.21) looks like a Freidlin-Wentzell large deviation rate function. This will be exploited at Section 3, see (3.7).

(b) It is noticed in [8] that it is not clear that these action functionals are equivalent and that one of them admits (2.15) as its Euler-Lagrange equation.

(c) It happens that $\widehat{\nabla}\Pi_S(\mathsf{y}) = \widehat{\mathrm{proj}}_S(\mathsf{y})$, as defined at (2.13). See Proposition C.6.

Extension to a discrete source measure. Let us replace the normalized volume measure λ of some set $D \in \mathbb{R}^d$ with a strictly positive Lebesgue volume, see (2.5), by the normalized counting measure

$$\lambda^{(k)} := \frac{1}{k} \sum_{1\le j\le k} \delta_{x_j} \tag{2.23}$$

on some finite set $D^{(k)} := \{x_1, \dots, x_k\} \subset \mathbb{R}^d$. Any mapping $\mathsf{y} : D^{(k)} \to \mathbb{R}^d$ is encoded by the vector $\mathsf{y} := (\mathsf{y}(x_1), \dots, \mathsf{y}(x_k)) \in (\mathbb{R}^d)^k$ whose squared norm in the Hilbert space $H := L^2_{\mathbb{R}^d}(\lambda^{(k)})$ is

$$\|\mathsf{y}\|_H^2 = k^{-1} \sum_{1\le j\le k} |\mathsf{y}(x_j)|^2 = k^{-1}\|\mathsf{y}\|^2$$

where $|\bullet|$ and $\|\bullet\|$ are respectively the Euclidean norms on $\mathbb{R}^d$ and $(\mathbb{R}^d)^k$. Hence,

$$H \simeq (\mathbb{R}^d)^k.$$

Definition 2.24 (k-mapping)**.** *When the source measure is the discrete measure $\lambda^{(k)}$, we call any application $\mathsf{y} : D^{(k)} \to \mathbb{R}^d$ in H a k-mapping.*

For the ease of notation, we shall write $\mathbb{R}^{dk}$ instead of $(\mathbb{R}^d)^k$, but one should keep in mind the structure of $(\mathbb{R}^d)^k$. Also, keeping our convention of writing elements of H with bold letters, any k-mapping will also be written this way.
The set S of all k-mappings preserving $\lambda^{(k)}$ is

$$S = \{\mathsf{x}^\sigma; \sigma \in \mathcal{S}\} \tag{2.25}$$

where $\mathcal{S}$ is the set of all permutations of k elements and

$$\mathsf{x}^\sigma := (x_{\sigma(1)}, \dots, x_{\sigma(k)}) \in H, \qquad \sigma \in \mathcal{S}.$$

Brenier proposed in [9] to extend Definition 2.20 to this semi-discrete setting which is natural to implement numerical simulations [25].

Definition 2.26 (Action functional of MAG pushed by $\{x_1, \dots, x_k\}$, [9])**.** *It is still (2.21) as in Definition 2.20, but $H \simeq (\mathbb{R}^d)^k$ and S is given by (2.25).*

This definition is the basis for the definitions of new action functionals we will work with in the rest of the article, see (4.18) and Definitions 4.7, 5.9 and 5.13. At present time, it is only justified by analogy with Definition 2.20. Its effectiveness in the numerical simulations suggests that one should recover (2.21) by letting k tend to infinity and preparing $\lambda^{(k)}$ such that $\lim_{k\to\infty} \lambda^{(k)} = \mathbf{1}_D\,\mathrm{Leb}$.

3. MAG. Particle system

This section is aimed at giving a presentation of a random particle system, see (3.7) below, introduced by Ambrosio, Baradat and Brenier in [2] whose dynamics is related to the action functional of MAG pushed by $\{x_1, \dots, x_k\}$ introduced at Definition 2.26, see (3.12) below.

Let us introduce the following stochastic differential equation in the set $\mathbb{R}^{dk}$ of k-mappings

$$\mathsf{X}^\epsilon_s = \mathsf{X}_0 + \sqrt{\epsilon}\, \mathsf{B}_s, \quad s \geq 0, \tag{3.1}$$

where B is a standard Brownian motion in $\mathbb{R}^{dk}$ starting from zero, $\epsilon > 0$ is a fluctuation parameter which is intended to tend to zero, and the law of the initial position X_0 in $\mathbb{R}^{dk}$ is

$$r_0 := \mathrm{Law}(\mathsf{X}_0) = \frac{1}{k!} \sum_{\sigma \in \mathcal{S}} \delta_{\mathsf{x}^\sigma}. \tag{3.2}$$

The process $\mathsf{X}^\epsilon_s = (\mathsf{X}^{\epsilon,1}_s, \dots, \mathsf{X}^{\epsilon,k}_s)$ describes a cloud of k *indistinguishable* Brownian particles in $\mathbb{R}^d$ starting from $\{x_1, \dots, x_k\}$; indistinguishability being a direct consequence of the choice (3.2) of the initial law. It is a random *path of k-mappings*. The j-th coordinate $\mathsf{X}^{\epsilon,j}$ is the path of the j-th particle starting from the j-th random draw without replacement from the set $\{x_1, \dots, x_k\}$. Remark that although the coordinates $\mathsf{X}^{\epsilon,j}$, $1 \leq j \leq k$, are correlated, they share the same law with initial distribution $\lambda^{(k)}$, see (2.23).

For any $s > 0$, the law of X^ϵ_s is the following mixture of Gaussian measures in $\mathbb{R}^{dk}$ with means x^σ and covariance matrices $\epsilon s\, \mathbb{I}$:

$$r^\epsilon_s(d\mathsf{x}) = \frac{1}{k!} \sum_{\sigma \in \mathcal{S}} (2\pi\epsilon s)^{-dk/2} \exp\left(-\frac{\|\mathsf{x} - \mathsf{x}^\sigma\|^2}{2\epsilon s} \right) d\mathsf{z}. \tag{3.3}$$

It solves the heat equation

$$\partial_s r^\epsilon = \epsilon \Delta r^\epsilon / 2 \tag{3.4}$$

which rewrites as the continuity equation

$$\partial_s r^\epsilon + \nabla\cdot(r^\epsilon \dot{\mathsf{r}}^\epsilon) = 0$$

with the current velocity field $\dot{\mathsf{r}}^\epsilon$ on $\mathbb{R}^{dk}$ of the diffusion process X^ϵ defined for any $s > 0$ and any $\mathsf{x} \in \mathbb{R}^{dk}$ by

$$\begin{aligned} \dot{\mathsf{r}}^\epsilon_s(\mathsf{x}) &= -\epsilon \nabla \log \sqrt{r^\epsilon_s}(\mathsf{x}) \\ &= -\frac{\epsilon \nabla r^\epsilon_s}{2 r^\epsilon_s}(\mathsf{x}) = \frac{\sum_{\sigma \in \mathcal{S}} (\mathsf{x} - \mathsf{x}^\sigma) \exp\big(- \|\mathsf{x} - \mathsf{x}^\sigma\|^2/(2\epsilon s)\big)}{2s \sum_{\sigma \in \mathcal{S}} \exp\big(- \|\mathsf{x} - \mathsf{x}^\sigma\|^2/(2\epsilon s)\big)} \\ &= \frac{1}{2s} \left(\mathsf{x} - \frac{\sum_{\sigma \in \mathcal{S}} \mathsf{x}^\sigma \exp\big(- \|\mathsf{x} - \mathsf{x}^\sigma\|^2/(2\epsilon s)\big)}{\sum_{\sigma \in \mathcal{S}} \exp\big(- \|\mathsf{x} - \mathsf{x}^\sigma\|^2/(2\epsilon s)\big)} \right). \end{aligned} \tag{3.5}$$

Letting ϵ tend to zero, we see with the Laplace principle that for any $s > 0$, $\lim_{\epsilon \to 0} \dot{\mathsf{r}}^\epsilon_s(\mathsf{x}) = (2s)^{-1}(\mathsf{x} - \mathsf{x}^{\sigma(\mathsf{x})})$ where $\mathsf{x}^{\sigma(\mathsf{x})}$ is the closest point from x among all the x^σ in S, *provided that this closest point is unique*. In view of (2.25), under this uniqueness assumption this means that

$$\lim_{\epsilon \to 0} \dot{\mathsf{r}}^\epsilon_s(\mathsf{x}) = \frac{\mathsf{x} - \mathrm{proj}_S(\mathsf{x})}{2s}, \tag{3.6}$$

a formula similar to the right hand side of equation (2.12). Getting rid of the factor $(2s)^{-1}$ will be a matter of change of time, see the parameter setting 3.10 below.

Since the action functional (2.19) can be read as some large deviation rate function, it is proposed in [2] to consider the stochastic differential equation in $\mathbb{R}^{dk}$

$$d\mathsf{Z}_s^{\epsilon,\eta} = \dot{\mathsf{r}}_s^{\epsilon}(\mathsf{Z}_s^{\epsilon,\eta})\,ds + \sqrt{\eta k^2 \kappa_s^{-1}}\,d\mathsf{W}_s, \quad 0 < s_0 \le s \le s_1, \tag{3.7}$$

where $\dot{\mathsf{r}}^\epsilon$ is the current velocity (3.5), $s \mapsto \kappa_s$ is a positive function, $\eta > 0$ is a parameter intended to decrease to zero and W is a standard Brownian motion on $\mathbb{R}^{dk}$. The Freidlin-Wentzell large deviation principle roughly states that, for any fixed $\epsilon > 0$ and any initial state x_o in $\mathbb{R}^{dk}$, when η tends to zero

$$\mathrm{Proba}(\mathsf{Z}^{\epsilon,\eta} \in \bullet \mid \mathsf{Z}_{s_0}^{\epsilon,\eta} = \mathsf{z}_o) \underset{\eta\to 0}{\asymp} \exp\left(-\eta^{-1} \inf_{\mathsf{z}\in\bullet,\mathsf{z}(s_0)=\mathsf{z}_o} \widetilde{A}^\epsilon(\mathsf{z})\right)$$

with the large deviation rate function

$$\widetilde{A}^\epsilon[(\mathsf{z}_s)_{s_0\le s\le s_1}] = \int_{s_0}^{s_1} \frac{1}{2}\|\dot{\mathsf{z}}_s - \dot{\mathsf{r}}_s^\epsilon(\mathsf{z}_s)\|_H^2\,\kappa_s\,ds \tag{3.8}$$

where $(\mathsf{z}_s)_{s_0\le s\le s_1}$ stands for a generic absolutely continuous path taking its values in $\mathbb{R}^{dk}$. Remember that $\|\bullet\|_H = k^{-1}\|\bullet\|_{\mathbb{R}^{dk}}$ and also that the factor k is part of the diffusion coeffective in (3.7). It is proved in [2], see also [1], that

$$\Gamma\text{-}\lim_{\epsilon\to 0} \widetilde{A}^\epsilon\big[(\mathsf{z}_s)_{s_0\le s\le s_1})\big] = \int_{s_0}^{s_1} \frac{1}{2}\left\|\dot{\mathsf{z}}_s - \frac{\mathsf{z}_s - \widehat{\mathrm{proj}}_S(\mathsf{z}_s)}{2s}\right\|_H^2 \kappa_s\,ds, \tag{3.9}$$

with $\widehat{\mathrm{proj}}_S$ defined at (2.13). Compare with (3.6).
It is easily seen (see (5.8) below for details) that applying the following

Parameter setting 3.10.

- *choose* $\kappa_s = 2s$,
- *change time:* $s = e^{2t}$,

and denoting $\mathsf{y}_t := \mathsf{z}_{e^{2t}}$, the right-hand side of (3.8) becomes

$$A^\epsilon\big[(\mathsf{y}_t)_{t_0\le t\le t_1})\big] := \int_{t_0}^{t_1} \frac{1}{2}\left\|\dot{\mathsf{y}}_t - \mathsf{y}_t + \frac{\sum_{\sigma\in\mathcal{S}} \mathsf{x}^\sigma \exp\big(-\|\mathsf{y}_t - \mathsf{x}^\sigma\|^2/(2\epsilon e^{2t})\big)}{\sum_{\sigma\in\mathcal{S}} \exp\big(-\|\mathsf{y}_t - \mathsf{x}^\sigma\|^2/(2\epsilon e^{2t})\big)}\right\|_H^2 dt, \tag{3.11}$$

where $t_0 = \log\sqrt{s_0}$ and $t_1 = \log\sqrt{s_1}$ and the Γ-limit (3.9) becomes

$$\Gamma\text{-}\lim_{\epsilon\to 0} A^\epsilon\big[(\mathsf{y}_t)_{t_0\le t\le t_1})\big] = \int_{t_0}^{t_1} \frac{1}{2}\|\dot{\mathsf{y}}_t - \mathsf{y}_t + \widehat{\mathrm{proj}}_S(\mathsf{y}_t)\|_H^2\,dt. \tag{3.12}$$

As desired, it is MAG's action (2.21). Remark that the appearance of the hat upon proj_S is a consequence of the Γ-limit.

Definition 3.13 (Action of ϵ-MAG pushed by $\{x_1, \dots, x_k\}$)**.**
The functional A^ϵ defined at (3.11) is the action of ϵ-MAG pushed by $\{x_1, \dots, x_k\}$.
Its s-version is $\widetilde{A}^\epsilon$ defined by (3.8).

Remarks.

(a) In [2], the k-mapping-valued stochastic process (3.7) is interpreted as *surfing the heat wave*, pointing out some analogy with de Broglie's pilot wave theory in quantum physics.

(b) Let us quote a sentence from [2]: *"Unexpectedly, the action* (3.12) *is exactly the one previously suggested by the third author* (Brenier) *in* [8] *to include dissipative phenomena (such as sticky collisions in one space dimension) in the Monge-Ampère gravitational model!"* The physical interpretation of the *intriguing* particle system (3.7) is also unclear to us at first sight, while its connection with MAG seems tight enough to think that it might not be incidental.

The significant feature of the stochastic differential equation (3.7) is that the *forward velocity* of $\mathsf{Z}^{\epsilon,\eta}$: the drift field $\dot{\mathsf{r}}^\epsilon$, is the *current velocity* of someone else, namely Z^ϵ. *How to give a meaning to this substitution?* One purpose of this article is to propose at Section 9 alternate Brownian particle systems with a clearer physical meaning.

In order to proceed in this direction, we need to extend MAG from mappings y to fluids.

4. MAG for a fluid

We look at the evolution of a self-gravitating *fluid* in $\mathbb{R}^d$ governed by a MAG force field. This section only deals with MAG: we drop ϵ-MAG for a while.

Sections 2 and 3 were dedicated to the flow of mappings $t \mapsto \mathsf{y}_t = (\mathsf{y}_t(x))_{x\in D}$, keeping track of the source element $x \in D$. This was necessary to obtain a representation in terms of optimal transport. However, fluid particles being indistinguishable, we do not observe the detail of the mapping y_t, but only the profile

$$\mu_t := (\mathsf{y}_t)_\# \lambda \in \mathrm{P}(\mathbb{R}^d) \tag{4.1}$$

of positions at time t. The following term is part of the integrand of the action (2.21):

$$\|\dot{\mathsf{y}}_t - \{\mathsf{y}_t - \widehat{\mathrm{proj}}_S(\mathsf{y}_t)\}\|_H^2 = \int_D |\dot{\mathsf{y}}_t(x) - \{\mathsf{y}_t(x) - [\widehat{\mathrm{proj}}_S(\mathsf{y}_t)](x)\}|^2 \, \lambda(dx).$$

Properties 4.2. *Suppose that there exist a vector field $v_t(y)$ and a map $\widehat{T}(\mu, y)$ such that for any x in D,*

- $\dot{\mathsf{y}}_t(x) = v_t\big(\mathsf{y}_t(x)\big) \in \mathbb{R}^d,$ (4.3)
- $[\widehat{\mathrm{proj}}_S(\mathsf{y}_t)](x) = \widehat{T}\big(\mu_t, \mathsf{y}_t(x)\big) \in \mathbb{R}^d$ *where* $\mu_t := (\mathsf{y}_t)_\#\lambda.$ (4.4)

Then, (2.21) writes as

$$\begin{aligned}
\int_{t_0}^{t_1} \frac{1}{2} \|\dot{\mathsf{y}}_t &- \{\mathsf{y}_t - \widehat{\mathrm{proj}}_S(\mathsf{y}_t)\}\|_H^2 \, dt \\
&= \int_{D\times[t_0,t_1]} \frac{1}{2} |v_t\big(\mathsf{y}_t(x)\big) - \{\mathsf{y}_t(x) - \widehat{T}\big(\mu_t, \mathsf{y}_t(x)\big)\}|^2 \, \lambda(dx)dt \\
&= \int_{\mathbb{R}^d\times[t_0,t_1]} \frac{1}{2} |v_t(y) - [y - \widehat{T}(\mu_t, y)]|^2 \, \mu_t(dy)dt \\
&= \int_{t_0}^{t_1} \frac{1}{2} \left\langle |v_t - (\mathrm{Id} - \widehat{T}_{\mu_t})|^2, \mu_t \right\rangle dt = \int_{t_0}^{t_1} \frac{1}{2} \|v_t - (\mathrm{Id} - \widehat{T}_{\mu_t})\|_{\mu_t}^2 \, dt,
\end{aligned} \tag{4.5}$$

where last but one equality is simply a change of notation: $\langle f, \mu \rangle := \int_{\mathbb{R}^d} f \, d\mu$, and $\|\bullet\|_\mu$ is a shorthand for $\|\bullet\|_{L^2_{\mathbb{R}^d}(\mu)}$. By (4.3), v_t is the velocity field of the fluid with density μ_t. Hence, the continuity equation

$$\partial_t \mu_t + \nabla\cdot(\mu_t v_t) = 0 \tag{4.6}$$

is satisfied in a weak sense. Although this notation suggests that μ_t should be absolutely continuous, no such hypothesis is done at Proposition 4.11 below: the weak formulation (4.12) is valid for any probability measure μ_t.

Dropping the requirement (4.1) that the density profile is represented by $\mu_t = (\mathsf{y}_t)_\# \lambda$ for some flow of mappings $(\mathsf{y}_t)_{t_0 \le t \le t_1}$, the identity (4.5) suggests the following extension to a fluid of the definition of Monge-Ampère gravitation.

Definition 4.7 (MAG's action for a fluid pushed by λ)**.** *Let λ be any probability measure on $D \subset \mathbb{R}^d$. A least action principle for a MAG self-gravitating fluid pushed by λ is*

$$\inf_{(\mu,v)} \int_{t_0}^{t_1} \tfrac{1}{2} \| v_t - (\mathrm{Id} - \widehat{T}_{\mu_t}) \|^2_{\mu_t} \, dt \tag{4.8}$$

where

(i) the infimum runs through all (μ, v) satisfying the continuity equation (4.6) in the weak sense (4.12), and such that the endpoint marginals μ_{t_0} and μ_{t_1} are prescribed,

(ii) $\widehat{T}_\mu$ is the optimal map $\overleftarrow{T}_\mu$ transporting μ to λ, if the Monge transport problem admits a unique solution, or some extension of it if uniqueness fails, see Remark 4.9-(c) below.

This model depends on the choices of the measure λ and the extension $\widehat{T}$ of $\overleftarrow{T}$.

Remarks 4.9. Clearly, its physical adequateness depends on the fulfillment of Properties 4.2. Let us comment on them.

(a) As far as one is interested in the evolution of the density μ_t forgetting the details of the mapping y_t, property (4.3) is justified by Proposition 4.11 below.

(b) Let us have a look at property (4.4). As long as μ_t remains absolutely continuous, for any $x \in D$, the x-th component $[\widehat{\mathrm{proj}}_S(\mathsf{y}_t)](x) \in \mathbb{R}^d$ of $\widehat{\mathrm{proj}}_S(\mathsf{y}_t)$ only depends on μ_t and the position $\mathsf{y}_t(x)$ of the x-th particle in $\mathbb{R}^d$. Indeed, in this situation $\mathrm{Proj}_S(\mathsf{y}_t)$ contains the single element $\mathrm{proj}_S(\mathsf{y}_t) = \overleftarrow{T}_{\mu_t} \circ \mathsf{y}_t$ where by (2.3)

$$\overleftarrow{T}_{\mu_t} = \nabla \theta_t \tag{4.10}$$

is the optimal transport map from the absolutely continuous measure μ_t to the measure λ, in which case $[\widehat{\mathrm{proj}}_S(\mathsf{y}_t)](x) = [\mathrm{proj}_S(\mathsf{y}_t)](x) = \overleftarrow{T}_{\mu_t}(\mathsf{y}_t(x))$.

(c) We think that it is *physically reasonable* to assume that property (4.4) still holds when μ_t fails to be absolutely continuous. We think that a privileged model should consist of replacing $\overleftarrow{T}_{\mu_t}(y)$ by the orthogonal projection in $\mathbb{R}^d$ of y on the closed convex hull of the subset $\cup_{\pi_{mu_t}} \mathrm{supp}\, \pi_{\mu_t}(\bullet \mid y)$, where the union runs through the collection $\{\pi_{\mu_t}\}$ of all solutions of the Monge-Kantorovich optimal transport problem *from* μ_t *to* λ, and $\mathrm{supp}\, \pi_{\mu_t}(\bullet \mid y)$ stands for the support of the optimal plan π_{μ_t} conditioned by the knowledge of the source location y.

Let us recall the following standard result justifying Remark 4.9-(a).

Proposition 4.11. *Suppose that the (normalized) density $\mu_t = (Y_t)_\# P$ is the time marginal at time t of some path measure $P \in \mathrm{P}(\Omega)$ which is supported by absolutely continuous sample paths, i.e.*

$$dY_t = \dot{Y}_t \, dt, \quad P\text{-a.s.},$$

where $(Y_t)_{t_0 \le t \le t_1}$ stands for the canonical process on the path space

$$\Omega := C([t_0, t_1], \mathbb{R}^d)$$

and $\dot{Y}_t$ is some random vector, possibly depending (a priori) on the whole history of the path and satisfying

$$E_P \int_{t_0}^{t_1} |\dot{Y}_t|^2\, dt < \infty.$$

Then, there exists some measurable vector field $v_t(y)$ such that for almost all t, v_t belongs to the closure $H^{-1}_{\mathbb{R}^d}(\mu_t)$ in $L^2_{\mathbb{R}^d}(\mu_t)$ of the space $\{\nabla u; u \in C^1_c(\mathbb{R}^d)\}$ of regular gradient vector fields,

$$\int_{\mathbb{R}^d \times [t_0,t_1]} |v_t(y)|^2\, \mu_t(dy)dt < \infty,$$

and the continuity equation (4.6) holds in the following weak sense: For any function f in $C^1_c(\mathbb{R}^d)$ and any $t_0 \le t_* \le t_1$,

$$\int_{\mathbb{R}^d} f\, d\mu_{t_*} - \int_{\mathbb{R}^d} f\, d\mu_{t_0} = \int_{\mathbb{R}^d \times [t_0,t_*]} \nabla f(y)\cdot v_t(y)\, \mu_t(dy)dt. \tag{4.12}$$

Proof. Take any function g in $C^{1,1}_c(\mathbb{R}^d \times [t_0, t_*])$ with $t_0 \le t_* \le t_1$. Then,

$$\begin{aligned}
\int_{\mathbb{R}^d} g_{t_*}\, d\mu_{t_*} - \int_{\mathbb{R}^d} g_{t_0}\, d\mu_{t_0} &= E_P \int_{t_0}^{t_*} [\partial_t g_t(Y_t) + \nabla g_t(Y_t)\cdot \dot{Y}_t]\, dt \\
&= E_{\bar{P}}(\partial_t g(\tau, Y_\tau) + \nabla g_\tau(Y_\tau)\cdot \dot{Y}_\tau) = E_{\bar{P}} E_{\bar{P}}\big[\partial_t g(\tau, Y_\tau) + \nabla g_\tau(Y_\tau)\cdot \dot{Y}_\tau \mid (Y_\tau, \tau)\big] \\
&= E_{\bar{P}}\big[\partial_t g(\tau, Y_\tau) + \nabla g_\tau(Y_\tau)\cdot \tilde{v}_\tau(Y_\tau)\big] = \int_{\mathbb{R}^d \times [t_0,t_*]} \big[\partial_t g_t(y) + \nabla g_t(y)\cdot \tilde{v}_t(y)\big]\, \mu_t(dy)dt \\
&= \int_{\mathbb{R}^d \times [t_0,t_*]} \big[\partial_t g_t(y) + \nabla g_t(y)\cdot v_t(y)\big]\, \mu_t(dy)dt,
\end{aligned}$$

where $\bar{P}(d\omega dt) := P(d\omega)dt$, the canonical time is τ, $\tilde{v}_t(y) := E_{\bar{P}}\Big(\dot{Y}_\tau \mid Y_\tau = y, \tau = t\Big)$, and v is the orthogonal projection in $L^2_{\mathbb{R}^d}(\mu_t(dy)dt)$ of $\tilde{v}$ on the closure of the space $\{\nabla g; g \in C^{1,1}_c(\mathbb{R}^d \times [t_0, t_*])\}$. This implies (4.12) and the joint measurability of $(t, y) \mapsto v_t(y)$. □

From $\inf_{(\mu,u)}$ **to** $\inf_{(\mu)}$**, Otto calculus.** The basic insight of Otto calculus [29, 3, 33], is to interpret the velocity field $v_t \in H^{-1}_{\mathbb{R}^d}(\mu_t)$ appearing at Proposition 4.11 as a tangent vector at μ_t in some Riemannian-like manifold $\mathrm{P}_2 \subset \mathrm{P}(\mathbb{R}^d)$, called the Otto-Wasserstein manifold. We denote this velocity by

$$\dot{\mu}_t := v_t \in \mathrm{T}_{\mu_t}\mathrm{P}_2 \subset H^{-1}_{\mathbb{R}^d}(\mu_t) \tag{4.13}$$

where $\mathrm{T}_\mu \mathrm{P}_2$ stands for the tangent space of P_2 at μ. In particular, the continuity equation (4.6) writes as

$$\partial_t \mu + \nabla\cdot(\mu\dot{\mu}) = 0.$$

We emphasize that the vertical variation $\partial_t \mu_t$ differs form the horizontal variation $\dot{\mu}_t$.
The relation between quadratic optimal transport and Otto calculus is best illustrated by the Benamou-Brenier formula

$$\begin{aligned}
W_2^2(\alpha, \beta) &\overset{\text{(i)}}{=} \inf_{(\mu,u)} \int_0^1 \Big(\int_{\mathbb{R}^d} |u_t(y)|^2\, \mu_t(dy)\Big)\, dt \\
&\overset{\text{(ii)}}{=} \inf_{(\mu)} \int_0^1 \Big(\int_{\mathbb{R}^d} |\dot{\mu}_t(y)|^2\, \mu_t(dy)\Big)\, dt \overset{\text{(iii)}}{=} \inf_{(\mu)} \int_0^1 \|\dot{\mu}_t\|^2_{\mu_t}\, dt,
\end{aligned}$$

where $\alpha, \beta \in \mathrm{P}(\mathbb{R}^d)$ and

$$W_2^2(\alpha, \beta) := \inf_{\pi} \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^2 \, \pi(dxdy)$$

is the optimal quadratic transport cost from α to β.
Let us comment on these infimums:

- in the expression $\inf_\pi$, $\pi \in \mathrm{P}(\mathbb{R}^d \times \mathbb{R}^d)$ runs through all the couplings of α and β, that is: $\pi(dx \times \mathbb{R}^d) = \alpha(dx)$ and $\pi(\mathbb{R}^d \times dy) = \beta(dy)$;
- in the expression $\inf_{(\mu)}$, the infimum runs through all the paths $(\mu) = (\mu_t)_{0 \le t \le 1}$ in $\mathrm{P}(\mathbb{R}^d)$ such that $\mu_0 = \alpha$ and $\mu_1 = \beta$;
- in the expression $\inf_{(\mu,u)}$, the infimum runs through all the paths $(\mu, u) = (\mu_t, u_t)_{0 \le t \le 1}$ where $\mu_0 = \alpha$, $\mu_1 = \beta$ and the u_t's are vector fields such that the continuity equation $\partial_t \mu + \nabla \cdot (\mu u) = 0$ holds in the weak sense.

Identity (i) is the Benamou-Brenier formula. Identity (ii) relies on Proposition 4.11 which states that one can replace u_t in the continuity equation $\partial_t \mu + \nabla \cdot (\mu u) = 0$ by $\dot{\mu}_t$ which, as an element of $H^{-1}_{\mathbb{R}^d}(\mu_t)$, minimizes $\int_{\mathbb{R}^d} |u_t|^2 \, d\mu_t$ (Hilbertian projection onto $H^{-1}_{\mathbb{R}^d}(\mu_t)$). The last equality (iii) directly follows from the notation

$$\|\dot{\mu}_t\|^2_{\mu_t} := \int_{\mathbb{R}^d} |\dot{\mu}_t(y)|^2 \, \mu_t(dy) \tag{4.14}$$

where $\|\dot{\mu}_t\|^2_{\mu_t}$ should be interpreted as the analogue of the squared Riemannian norm of a tangent vector at μ_t.
Consequently, for a standard Lagrangian $(t, \mu, \dot{\mu}) \mapsto \alpha_t \|\dot{\mu}\|^2_\mu / 2 - \mathcal{U}_t(\mu)$, where $\alpha : [t_0, t_1] \to [0, \infty)$ is a nonnegative function,

$$\inf_{(\mu,u)} \int_{t_0}^{t_1} \Big(\int_{\mathbb{R}^d} \tfrac{1}{2} |u_t|^2 \alpha_t \, d\mu_t - \mathcal{U}_t(\mu_t) \Big) \, dt = \inf_{(\mu)} \int_{t_0}^{t_1} \big(\tfrac{1}{2} \|\dot{\mu}_t\|^2_{\mu_t} \alpha_t - \mathcal{U}_t(\mu_t) \big) \, dt.$$

We shall also be concerned by Lagrangian of type $(t, \mu, \dot{\mu}) \mapsto \frac{1}{2} \|\dot{\mu} - \nabla w_t\|^2_\mu \alpha_t$ where, again, α is a nonnegative function.

Lemma 4.15. *Suppose that $w : [t_0, t_1] \times \mathbb{R}^d \to \mathbb{R}$ is a regular function, then*

$$\begin{aligned} \int_{t_0}^{t_1} \tfrac{1}{2} \|\dot{\mu}_t - \nabla w_t\|^2_{\mu_t} \alpha_t \, dt & \\ = \int_{t_0}^{t_1} \Big\{ \tfrac{1}{2} \|\dot{\mu}_t\|^2_{\mu_t} \alpha_t + \tfrac{1}{2} \|\nabla w_t\|^2_{\mu_t} \alpha_t &+ \int_{\mathbb{R}^d} \partial_t (\alpha_t w_t) \, d\mu_t \Big\} \, dt \\ &- \int_{\mathbb{R}^d} \alpha_{t_1} w_{t_1} \, d\mu_{t_1} + \int_{\mathbb{R}^d} \alpha_{t_0} w_{t_0} \, d\mu_{t_0}. \end{aligned}$$

Proof. Expanding the square gives us

$$\|\dot{\mu} - \nabla w_t\|^2_\mu = \|\dot{\mu}\|^2_\mu + \|\nabla w_t\|^2_\mu - 2(\nabla w_t, \dot{\mu})_\mu = \|\dot{\mu}\|^2_\mu + \|\nabla w_t\|^2_\mu - 2(\mathrm{grad}^{\mathrm{OW}}_\mu \mathcal{W}_t, \dot{\mu})_\mu$$

where $\mathrm{grad}^{\mathrm{OW}}$ is the gradient with respect to the Otto-Wasserstein metric derived from (4.14), and $\mathcal{W}_t(\mu) := \int_{\mathbb{R}^d} w_t \, d\mu$. The result follows from

$$\frac{d}{dt} (\alpha_t \mathcal{W}_t(\mu_t)) = \alpha_t (\mathrm{grad}^{\mathrm{OW}}_{\mu_t} \mathcal{W}_t, \dot{\mu}_t)_{\mu_t} + \int_{\mathbb{R}^d} \partial_t (\alpha_t w_t) \, d\mu_t.$$

See [3] for a proof of this chain rule. □

Therefore, the Lagrangian $\frac{1}{2}\|\dot\mu - \nabla w_t\|^2_\mu \alpha_t$ is equivalent to the modified Lagrangian $\alpha_t\|\dot\mu\|^2_\mu/2 + \alpha_t\|\nabla w_t\|^2_\mu/2 + \int_{\mathbb{R}^d} \partial_t(\alpha_t w_t)\, d\mu$ which has the form $\alpha_t\|\dot\mu\|^2_\mu/2 - \mathcal{U}_t(\mu)$. Hence,

$$\inf_{(\mu,u)} \int_{t_0}^{t_1} \Big(\int_{\mathbb{R}^d} \frac{1}{2}|u_t - \nabla w_t|^2\, d\mu_t \Big) \alpha_t\, dt = \inf_{(\mu)} \int_{t_0}^{t_1} \frac{1}{2}\|\dot\mu_t - \nabla w_t\|^2_{\mu_t}\, \alpha_t\, dt. \tag{4.16}$$

In particular, with some minor additional work, we obtain the following

Proposition 4.17 (MAG's action for a fluid pushed by λ)**.** *If* (4.10) *extends to*

$$\widehat{T}_{\mu_t} \in \partial\theta_t,$$

meaning that the extension $\widehat{T}_t$ *of* $\overleftarrow{T}$ *is still a subgradient of a convex function* θ_t *for almost all* t, *then the least action principle* (4.8) *of Definition 4.7 reads as*

$$\inf_{(\mu)} \int_{t_0}^{t_1} \frac{1}{2}\|\dot\mu_t - (\mathrm{Id} - \widehat{T}_{\mu_t})\|^2_{\mu_t}\, dt, \tag{4.18}$$

where the infimum runs through all (μ) *such that the endpoint marginals* μ_{t_0} *and* μ_{t_1} *are prescribed.*

5. ϵ-MAG for a k-fluid. Action functional

The main difficulty of the least action principle (4.18) is to handle the optimal transport term $\widehat{T}_\mu$. To do so, we keep the idea of [2] of working with the k-mapping-valued stochastic process X^ϵ, see (3.1)-(3.2), because the symmetrization operator $(k!)^{-1}\sum_\sigma$ together with the Laplace principle when ϵ tends to zero, is a good way to recover proj_S and therefore optimal transport, see (3.9). But, this is at the price of replacing the diffuse source measure λ defined at (2.5), by its discrete analogue $\lambda^{(k)}$ defined at (2.23).

On the other hand, we leave apart the enigmatic process $\mathsf{Z}^{\epsilon,\eta}$ defined at (3.7). It will be replaced at Section 9 by some empirical process built on Z^ϵ, see (7.3).

From now on, we only consider ϵ-approximations of MAG, in the sense that we do not let ϵ tend to zero, leaving open the problem of the fulfillment of property (4.4).

Definitions 5.1 (k-mapping and k-fluid)**.**

- *Any element* z *of* $\mathbb{R}^{dk}$ *is interpreted as a* k-mapping, *i.e.* $\mathsf{z} : \{x_1, \dots, x_k\} \to \mathbb{R}^d$.
- *Any probability measure* $p \in \mathrm{P}(\mathbb{R}^{dk})$ *on* $\mathbb{R}^{dk}$ *is interpreted as a fluid of* k*-mappings, and it is called a* k-fluid, *for short.*

An ϵ-approximation of MAG for k-fluids. Arguing as in Section 4, the ϵ-MAG action functional (3.8): $\widetilde{A}^\epsilon[\mathsf{z}] = \int_{s_0}^{s_1} \frac{1}{2}\|\dot{\mathsf{z}}_s - \dot{\mathsf{r}}^\epsilon_s(\mathsf{z}_s)\|^2_H\, \kappa_s\, ds$, admits the k-fluid analogue:

$$\int_{s_0}^{s_1} \frac{1}{2}\|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\, \kappa_s\, ds, \tag{5.2}$$

where $\dot{\mathsf{r}}^\epsilon$ is the current velocity (3.5) of X^ϵ, $s \mapsto p_s \in \mathrm{P}(\mathbb{R}^{dk})$ is the path of density distributions of a k-fluid, $\dot{\mathsf{p}}$ is its *gradient* velocity field, meaning that the continuity equation

$$\partial_s p + \nabla\cdot(p\, \dot{\mathsf{p}}) = 0 \tag{5.3}$$

is satisfied in the weak sense and $\dot{\mathsf{p}}_s \in H^{-1}_{\mathbb{R}^{dk}}(p_s)$. The action (5.2) is valid because $\dot{\mathsf{r}}^\epsilon$ is a gradient field, so that one can apply Lemma 4.15, leading to (4.16).

Similarly, at the limit $\epsilon \to 0$, reasoning as during the derivation of Proposition 4.17, the k-fluid analogue of the k-mapping action (3.9) is

$$\int_{s_0}^{s_1} \frac{1}{2}\left\| \dot{\mathsf{p}}_s - \frac{\mathrm{Id} - \widehat{T}_{p_s}}{2s} \right\|_{p_s}^2 \kappa_s \, ds \tag{5.4}$$

provided that the extension $\widehat{T}$ of $\overleftarrow{T}$ is still a subgradient of some convex function.
In view of the Γ-limit (3.9), we see that (5.2) is a reasonable ϵ-approximation of (5.4).

Applying to (5.4) the parameter setting 3.10: $\kappa_s = 2s$, $s = e^{2t}$, as was done in order to go from (3.9) to (3.12), and setting

$$q_t := p_s = p_{e^{2t}}, \tag{5.5}$$

we obtain its t-analogue

$$\int_{t_0}^{t_1} \frac{1}{2}\|\dot{\mathsf{q}}_t - (\mathrm{Id} - \widehat{T}_{q_t})\|_{q_t}^2 \, dt, \tag{5.6}$$

which is the MAG action for k-fluids, analogous to the MAG action functional for k-mappings obtained at Proposition 4.17. Defining

$$m_t^\epsilon(d\mathsf{y}) := r_s^\epsilon(d\mathsf{y}) = r_{e^{2t}}^\epsilon(d\mathsf{y}) \overset{(3.3)}{=} \frac{1}{k!}\sum_{\sigma\in\mathcal{S}} (2\pi\epsilon e^{2t})^{-dk/2} \exp\left(-\frac{\|\mathsf{y}-\mathsf{x}^\sigma\|^2}{2\epsilon e^{2t}}\right) d\mathsf{y}, \tag{5.7}$$

we obtain

$$\dot{\mathsf{m}}_t^\epsilon(\mathsf{y}) \overset{\checkmark}{=} 2e^{2t}\dot{\mathsf{r}}_{e^{2t}}^\epsilon(\mathsf{y}) \overset{(3.5)}{=} \mathsf{y} - \frac{\sum_{\sigma\in\mathcal{S}} \mathsf{x}^\sigma \exp\big(-\|\mathsf{y}-\mathsf{x}^\sigma\|^2/(2\epsilon e^{2t})\big)}{\sum_{\sigma\in\mathcal{S}} \exp\big(-\|\mathsf{y}-\mathsf{x}^\sigma\|^2/(2\epsilon e^{2t})\big)}, \tag{5.8}$$

where the marked equality follows from

$$\partial_t m_t = \partial_t r_{e^{2t}} = 2e^{2t}\partial_s r_s = -2e^{2t}\nabla\cdot(r_s\dot{\mathsf{r}}_s) = -2e^{2t}\nabla\cdot(m_t\dot{\mathsf{r}}_{e^{2t}})$$

with $s = e^{2t}$, that is: $\partial_t m_t + \nabla\cdot\big(m_t(2e^{2t}\dot{\mathsf{r}}_{e^{2t}})\big) = 0$. It is natural to propose the following

Definition 5.9 (ϵ-MAG for a k-fluid). *We consider flows $(q_t)_{t_0\le t\le t_1}$ of k-fluids. A least action principle for an ϵ-MAG self-gravitating k-fluid is*

$$\inf_{(q)} \int_{t_0}^{t_1} \frac{1}{2}\|\dot{\mathsf{q}}_t - \dot{\mathsf{m}}_t^\epsilon\|_{q_t}^2 \, dt \tag{5.10}$$

The infimum runs through all (q) with prescribed endpoint marginals q_{t_0} and q_{t_1}.

The action (5.10) is a reasonable ϵ-approximation of (5.6).

From k-fluids to fluids. The detail of the evolution of the flow of k-mappings $p_s \in \mathrm{P}(\mathbb{R}^{dk})$ is necessary for optimal transport to enter the game. However, particles are not coloured by any rank of trial (the k slots in $\mathbb{R}^d$ of a vector in $H \simeq (\mathbb{R}^d)^k$). We only see a monochrome cloud. This means that instead of a probability measure p on $(\mathbb{R}^d)^k$, we have to consider its projection

$$\mathrm{proj}_{dk\to d}(p) := k^{-1} \sum_{1\le j\le k} p_j \in \mathrm{P}(\mathbb{R}^d), \qquad p \in \mathrm{P}(\mathbb{R}^{dk}),$$

on $\mathrm{P}(\mathbb{R}^d)$, where for any $1 \le j \le k$, the j-th marginal $p_j \in \mathrm{P}(\mathbb{R}^d)$ of p is defined by: $p_j(dy) := p(\mathbb{R}^d \times \cdots \mathbb{R}^d \times dy \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d)$ where $dy \subset \mathbb{R}^d$ occupies the j-th slot. The weights k^{-1} are those of $\lambda^{(k)}$ because

$$\lambda^{(k)} \overset{(2.23)}{=} k^{-1} \sum_{1\le j\le k} \delta_{x_j} = \mathrm{proj}_{dk\to d}\Big((k!)^{-1}\sum_{\sigma\in\mathcal{S}} \delta_{\mathsf{x}^\sigma}\Big) \overset{(3.2)}{=} \mathrm{proj}_{dk\to d}(\mathrm{Law}(\mathsf{Z}_0)).$$

Introducing the j-th projection from $(\mathbb{R}^d)^k$ to $\mathbb{R}^d$ defined for any $(y_1, \dots, y_k) \in (\mathbb{R}^d)^k$ by

$$\mathrm{proj}_j(y_1, \dots, y_k) = y_j \in \mathbb{R}^d,$$

we see that $p_j = (\mathrm{proj}_j)_\# p$. On the other hand, the continuity equation $\partial_s p + \nabla \cdot (p\mathsf{v}) = 0$ in $\mathbb{R}^{dk}$, implies the continuity equation

$$\partial_s \mu + \nabla\cdot(\mu v) = 0 \tag{5.11}$$

in $\mathbb{R}^d$, via the transformation

$$\mathrm{proj}_{dk\to d}(p, \mathsf{v}) = (\mu, v) \quad \text{where} \quad \begin{cases} \mu = \mathrm{proj}_{dk\to d}(p) = k^{-1} \sum_{1\le j\le k} (\mathrm{proj}_j)_\# p, \\ v(y) = k^{-1} \sum_{1\le j\le k} E_p(\mathrm{proj}_j(\mathsf{v}) \mid \mathrm{proj}_j = y), \ y \in \mathbb{R}^d. \end{cases}$$

Indeed, we see with (4.12) that $\partial_s p + \nabla \cdot (p\mathsf{v}) = 0$ means that for any $s_0 \le s_* \le s_1$ and any function f in $C^1_c(\mathbb{R}^{dk})$, we have: $\int_{\mathbb{R}^{dk}} f\, d(p_{s_*} - p_{s_0}) = \int_{\mathbb{R}^{dk}\times[s_0,s_*]} \nabla f(\mathsf{y}) \cdot \mathsf{v}_s(\mathsf{y})\, p_s(d\mathsf{y}) ds$. Applying this identity with $f(y_1, \dots, y_k) = k^{-1} \sum_{1\le j\le k} g(y_j)$ for any function g in $C^1(\mathbb{R}^d)$, gives the announced result.

Remark that if $\mathrm{proj}_{dk\to d}(p) = \mu$, then $\mathrm{proj}_{dk\to d}(p, \dot{\mathsf{p}}) = (\mu, \dot\mu)$. This holds because $k^{-1} \sum_{1\le j\le k} E_p(\mathrm{proj}_j(\dot{\mathsf{p}}) \mid \mathrm{proj}_j = \bullet)$ is a gradient field on $\mathbb{R}^d$ since $\dot{\mathsf{p}}$ is a gradient field on $\mathbb{R}^{dk}$. As a consequence, the least action principle in $\mathrm{P}(\mathbb{R}^{dk})$ based on (5.2):

$$\inf_{(p)} \int_{s_0}^{s_1} \frac{1}{2} \|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\, \kappa_s\, ds \tag{5.12}$$

where the infimum runs through all the (p) with prescribed marginal measures p_{s_0} and p_{s_1}, leads to the following main definition of this article.
Let us denote

$$\Omega_{\mathrm{P}} := C([t_0, t_1], \mathrm{P}(\mathbb{R}^d)) \quad \text{or} \quad \Omega_{\mathrm{P}} := C([s_0, s_1], \mathrm{P}(\mathbb{R}^d))$$

(depending on the s or t context), the set of $\mathrm{P}(\mathbb{R}^d)$-valued trajectories, and similarly

$$\Omega^{(k)}_{\mathrm{P}} := C([t_0, t_1], \mathrm{P}(\mathbb{R}^{dk})) \quad \text{or} \quad \Omega^{(k)}_{\mathrm{P}} := C([s_0, s_1], \mathrm{P}(\mathbb{R}^{dk})),$$

the set of $\mathrm{P}(\mathbb{R}^{dk})$-valued trajectories.

Definition 5.13 (Least action principle for a fluid driven by ϵ-MAG pushed by $\{x_1, \dots, x_k\}$)**.**

It is the least action principle in $\mathrm{P}(\mathbb{R}^d)$

$$\inf_{(\nu)} \inf \left\{ \int_{t_0}^{t_1} \frac{1}{2} \|\dot{\mathsf{q}}_t - \dot{\mathsf{m}}^\epsilon_t\|^2_{q_t}\, dt; (q) \in \Omega^{(k)}_{\mathrm{P}} : \mathrm{proj}_{kd\to d}(q_t) = \nu_t, \forall t \right\} \tag{5.14}$$

where the leftmost infimum runs through all the $(\nu) \in \Omega_{\mathrm{P}}$ *such that* ν_{t_0} *and* ν_{t_1} *are prescribed. Its s-version is obtained replacing* (5.14) *by*

$$\inf_{(\mu)} \inf \left\{ \int_{s_0}^{s_1} \frac{1}{2} \|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\, \kappa_s\, ds; (p) \in \Omega^{(k)}_{\mathrm{P}} : \mathrm{proj}_{kd\to d}(p_s) = \mu_s, \forall s \right\}. \tag{5.15}$$

6. ϵ-MAG for a k-fluid. Newton equation

In this section, we partly stay at an informal level, applying Otto's heuristics, but we also prove rigorous results. Otto's heuristics means that, while investigating the Euler-Lagrange equation of a least action principle in the Otto-Wasserstein manifold, we only consider a finite dimensional analogy. This type of equation is usually referred to as a Newton equation.

Otto's heuristics. In the Otto-Wasserstein manifold, the velocity at time t of a moving profile of matter (q_t) is the vector field $\dot{\mathsf{q}}_t \in H^{-1}_{\mathbb{R}^{dk}}(q_t)$ of the continuity equation (5.3): $\partial_t q + \nabla\cdot(q\dot{\mathsf{q}}) = 0$. Staying at a heuristic level mainly consists of replacing $\dot{\mathsf{q}}_t \in H^{-1}_{\mathbb{R}^{dk}}(q_t)$ by

$$\dot{\mathsf{q}}_t = \nabla\theta_t$$

for some $C^{1,2}$-regular function $(t,x) \mapsto \theta(t,x)$. If this wishful thinking is realized, the acceleration is given by

$$\nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t}\dot{\mathsf{q}}_t = (\partial_t + \dot{\mathsf{q}}_t\cdot\nabla)\dot{\mathsf{q}}_t = \nabla\big(\partial_t\theta_t + \tfrac{1}{2}|\nabla\theta_t|^2\big).$$

We see that, at least informally, $\nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t}$ is identified with the convective derivative $\partial_t + \dot{\mathsf{q}}_t\cdot\nabla$. In fact, this works fine when calculating an acceleration, but in the general case $\nabla^{\mathrm{OW}}_{\nabla\alpha}\dot{\mathsf{q}}_t$ should be identified with $\mathrm{proj}_{H^{-1}_{\mathbb{R}^{dk}}(q_t)}\big[(\partial_t + \nabla\alpha\cdot\nabla)\dot{\mathsf{q}}_t\big]$ where $\mathrm{proj}_{H^{-1}_{\mathbb{R}^{dk}}(q_t)}$ is the orthogonal projection onto the space of *gradient* vector fields $H^{-1}_{\mathbb{R}^{dk}}(q_t)$.
For a presentation of Otto's heuristics, see the chapter entitled *Otto calculus* in [33]. Rigorous material for proving (6.11) below is presented in [26, 4, 20].

Let us introduce the field on $\mathbb{R}^{dk}$ of probability measures on $S \overset{(2.25)}{=} \{\mathsf{x}^\sigma; \sigma \in \mathcal{S}\} \subset \mathbb{R}^{dk}$,

$$\pi^\epsilon_t(\mathsf{y}) := \sum_\sigma \pi^{\epsilon,\sigma}_t(\mathsf{y})\,\delta_{\mathsf{x}^\sigma} \in \mathrm{P}(S), \qquad \mathsf{y} \in \mathbb{R}^{dk},$$

where

$$\pi^{\epsilon,\sigma}_t(\mathsf{y}) := \frac{w^{\epsilon,\sigma}_t(\mathsf{y})}{\sum_{\sigma'} w^{\epsilon,\sigma'}_t(\mathsf{y})} \qquad \text{with} \qquad w^{\epsilon,\sigma}_t(\mathsf{y}) := \exp\left(-\frac{\|\mathsf{y}-\mathsf{x}^\sigma\|^2}{2\epsilon e^{2t}}\right).$$

The main reason for introducing π^ϵ_t is the following expression of (5.8)

$$\dot{\mathsf{m}}^\epsilon_t(\mathsf{y}) = \mathsf{y} - \sum_\sigma \pi^{\epsilon,\sigma}_t(\mathsf{y})\,\mathsf{x}^\sigma = \mathsf{y} - \int_S \mathsf{x}\,[\pi^\epsilon_t(\mathsf{y})](d\mathsf{x}). \tag{6.1}$$

For any $\mathsf{y} \in \mathbb{R}^{dk}$, $\epsilon > 0$ and t, we write

$$\tilde{\mathsf{x}}^\epsilon_t(\mathsf{x},\mathsf{y}) := \mathsf{x} - \int_S \mathsf{x}'\;[\pi^\epsilon_t(\mathsf{y})](d\mathsf{x}'), \quad \mathsf{x} \in S$$
$$F^\epsilon_t(\mathsf{y}) := (\epsilon e^{2t})^{-1}\int_S \big((\mathsf{y}-\mathsf{x})\cdot\tilde{\mathsf{x}}^\epsilon_t(\mathsf{x},\mathsf{y})\big)\,\tilde{\mathsf{x}}^\epsilon_t(\mathsf{x},\mathsf{y})\,\pi^\epsilon_t(\mathsf{y})(d\mathsf{x}).$$

Let

$$S_*(\mathsf{y}) := \big\{\mathsf{x}^1_*(\mathsf{y}), \dots, \mathsf{x}^{n_*(\mathsf{y})}_*(\mathsf{y})\big\} = \mathrm{argmin}_S\,\|\mathsf{y} - \bullet\| \subset S$$

be the subset of all the $n_*(\mathsf{y})$ closest points to y in S, i.e. the orthogonal projection in $\mathbb{R}^{dk}$ of y on S. We also introduce

$$\pi_*(\mathsf{y}) := n_*(\mathsf{y})^{-1}\sum_{1\le n\le n_*(\mathsf{y})} \delta_{\mathsf{x}^n_*(\mathsf{y})}$$

the uniform probability measure on $S_*(\mathsf{y})$, and

$$A_*(\mathsf{y}) := [\mathrm{Cov}(\mathsf{x}_*^1 \dots, \mathsf{x}^{n_*})\, \bar{\mathsf{x}}_*](\mathsf{y}) = n_*^{-1} \sum_{1\le n\le n_*} [\bar{\mathsf{x}}_* \cdot (\mathsf{x}_*^n - \bar{\mathsf{x}}_*)\ (\mathsf{x}_*^n - \bar{\mathsf{x}}_*)](\mathsf{y})$$

where $\bar{\mathsf{x}}_*(\mathsf{y}) = n_*(\mathsf{y})^{-1} \sum_{1\le n\le n_*(\mathsf{y})} \mathsf{x}_*^n(\mathsf{y})$.

Theorem 6.2 (Newton equation for a k-fluid driven by ϵ-MAG)**.** *Any solution (q) of the least action principle (5.10) solves the Newton equation with the acceleration field $\nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t} \dot{\mathsf{q}}_t$ given, for any $\mathsf{y} \in \mathbb{R}^{dk}$, by*

$$\nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t} \dot{\mathsf{q}}_t(\mathsf{y}) = \dot{\mathsf{m}}_t^\epsilon(\mathsf{y}) + F_t^\epsilon(\mathsf{y}). \tag{6.3}$$

There exist $C, a > 0$ such that for all t and all $\mathsf{y} \in \mathbb{R}^{dk}$,

$$\left|F_t^\epsilon(\mathsf{y}) - \epsilon^{-1} e^{-2t} A_*(\mathsf{y})\right| \le C(\|\mathsf{y}\| + 1) e^{-2t} \epsilon^{-1} \exp(-a e^{-2t} \|\mathsf{y}\|/\epsilon) \underset{\epsilon\to 0}{\longrightarrow} 0. \tag{6.4}$$

Moreover, there exists a negligible subset $\mathcal{N}$ of $\mathbb{R}^{dk}$ which is a finite union of vector subspaces with codimension at least 2, such that

$$A_*(\mathsf{y}) = 0, \quad \textit{for all } \mathsf{y} \notin \mathcal{N}. \tag{6.5}$$

Therefore,

$$\textit{if } \mathsf{y} \notin \mathcal{N}, \qquad \nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t} \dot{\mathsf{q}}_t(\mathsf{y}) = \dot{\mathsf{m}}_t^\epsilon(\mathsf{y}) \qquad + o_{\epsilon\to 0}(1), \tag{6.6}$$

$$\textit{if } \mathsf{y} \in \mathcal{N}, \qquad \nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t} \dot{\mathsf{q}}_t(\mathsf{y}) = \dot{\mathsf{m}}_t^\epsilon(\mathsf{y}) + \epsilon^{-1} e^{-2t} A_*(\mathsf{y}) \ + o_{\epsilon\to 0}(1) \tag{6.7}$$

with $\sup_{\mathsf{y}\in\mathbb{R}^{dk}} \|A_*(\mathsf{y})\| \le 2r^3 < \infty$, where r is the common norm of the elements of S. Furthermore,

$$\nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t} \dot{\mathsf{q}}_t = \nabla^{\mathrm{OW}}_{\dot{\mathsf{m}}_t^\epsilon} \dot{\mathsf{m}}_t^\epsilon. \tag{6.8}$$

Proof. Its is a direct consequence of Lemmas 6.10, 6.13, 6.14 and 6.18 below. □

Remarks 6.9.

(a) With the notation of Lemma 6.18, the thin set is $\mathcal{N} = \{\mathsf{y} : n_*(\mathsf{y}) \ge 3\}$.
(b) Statement (6.6) with (6.1) is reminiscent to (2.15).
(c) Statement (6.7) expresses the divergence as ϵ tends to zero of the force field on $\mathcal{N}$. This strong force field on $\mathcal{N}$ is responsible for *concentration of matter*.
(d) Identity (6.8) is the precise meaning of [2]'s expression: *surfing the heat wave*.

Let us prove these lemmas.

Lemma 6.10. *Any solution (q) of the least action principle (5.10) solves the Newton equation*

$$\nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t} \dot{\mathsf{q}}_t = 2\dot{\mathsf{m}}_t^\epsilon + 4\epsilon^2 e^{4t} \nabla[\mathcal{Q}(m_t^\epsilon|\mathrm{Leb})], \tag{6.11}$$

where we introduced the quantum potential

$$\mathcal{Q}(m|\mathrm{Leb}) := -\frac{\Delta\sqrt{m}}{2\sqrt{m}}. \tag{6.12}$$

Heuristics for a proof. Defining for any $q \in \mathrm{P}(\mathbb{R}^{dk})$

$$\mathcal{F}_t(q) = -2\epsilon e^{2t} \int_{\mathbb{R}^{dk}} \log \sqrt{m_t^\epsilon}\, dq,$$

we see that $\dot{\mathsf{m}}_t^\epsilon = \mathrm{grad}_q^{\mathrm{OW}} \mathcal{F}_t$. Hence the Lagrangian $\frac{1}{2}\|\dot{\mathsf{q}} - \dot{\mathsf{m}}_t^\epsilon\|_q^2$ of (5.10) writes as

$$\frac{1}{2}\|\dot{\mathsf{q}} - \mathrm{grad}^{\mathrm{OW}} \mathcal{F}_t(q)\|_q^2,$$

the analog of which in an n-dimensional Riemannian manifold is

$$L(t,q,v)=\tfrac{1}{2}|v-\operatorname{grad} f_t(q)|_q^2$$

for some function f. More precisely, in a coordinate system

$$L(t,q,v)=\tfrac{1}{2}g_{ij}(q)\big(v^i-g^{ik}(q)\partial_k f_t(q)\big)\big(v^j-g^{jl}(q)\partial_l f_t(q)\big)$$

where the metric tensor is $g=(g_{ij})$, its inverse is $g^{-1}=(g^{ij})$ and we use Einstein's summation convention. Expending the square

$$L(t,q,v)=\tfrac{1}{2}|v|_q^2+\tfrac{1}{2}|\operatorname{grad} f_t(q)|_q^2-(\operatorname{grad} f_t(q),v)_q.$$

As

$$\frac{d}{dt}[f_t(\omega_t)]=\partial_t f_t(\omega_t)+df_t(\omega_t)\cdot\dot\omega_t=\partial_t f_t(\omega_t)+(\operatorname{grad} f_t(\omega_t),\dot\omega_t)_{\omega_t},$$

the Lagrangian L is equivalent to

$$\tfrac{1}{2}|v|_q^2+\tfrac{1}{2}|\operatorname{grad} f_t(q)|_q^2+\partial_t f_t(q),$$

in the sense that both least action principles attached to these Lagrangians and sharing the same prescribed endpoints admit the same solution. Let us introduce the scalar potential

$$U_t(q):=-\tfrac{1}{2}|\operatorname{grad} f_t(q)|_q^2-\partial_t f_t(q),$$

so that this Lagrangian has the standard form: $\tfrac{1}{2}|v|_q^2-U_t(q)$. Therefore, the Euler-Lagrange equation $\frac{d}{dt}[\partial_v L(t,\omega_t,\dot\omega_t)]-\partial_q L(t,\omega_t,\dot\omega_t)=0$ writes as the Newton equation

$$\nabla_{\dot\omega_t}\dot\omega_t=-\operatorname{grad} U_t(\omega_t),$$

where ∇ is the Levi-Civita connection of the metric tensor g, see [22].
By analogy, the Newton equation in the Otto-Wasserstein manifold is

$$\nabla^{\mathrm{OW}}_{\dot{\mathsf q}_t}\dot{\mathsf q}_t=-\operatorname{grad}^{\mathrm{OW}}_{\mathsf q_t}\mathcal{U}_t$$

where the analog of U_t is

$$\mathcal{U}_t(q)=-\frac{1}{2}\int_{\mathbb{R}^{dk}}|\operatorname{grad}^{\mathrm{OW}}_q\mathcal{F}_t|^2\,dq-\partial_t\mathcal{F}_t(q),\qquad q\in\mathrm{P}(\mathbb{R}^{dk}).$$

Let us compute $\mathcal{U}_t(q)$. As $\operatorname{grad}^{\mathrm{OW}}_q\mathcal{F}_t=-2\epsilon e^{2t}\nabla\log\sqrt{m_t^\epsilon}$, we obtain

$$\frac{1}{2}\int_{\mathbb{R}^{dk}}|\operatorname{grad}^{\mathrm{OW}}_q\mathcal{F}_t|^2\,dq=2\epsilon^2e^{4t}\int_{\mathbb{R}^{dk}}|\nabla\log\sqrt{m_t^\epsilon}|^2\,dq.$$

Let us look at $\partial_t\mathcal{F}_t(q)$. We have $m_t^\epsilon=r^\epsilon_{e^{2t}}$ and $\partial_s r_s^\epsilon=\epsilon\Delta r_s^\epsilon/2$. Hence $\partial_t m_t^\epsilon=2e^{2t}\partial_s r^\epsilon_{s|s=e^{2t}}=\epsilon e^{2t}\Delta r^\epsilon_{s|s=e^{2t}}=\epsilon e^{2t}\Delta m_t^\epsilon$ and

$$-\partial_t\mathcal{F}_t(q)=4\epsilon e^{2t}\int_{\mathbb{R}^{dk}}\log\sqrt{m_t^\epsilon}\,dq+\epsilon e^{2t}\int_{\mathbb{R}^{dk}}\partial_t\log m_t^\epsilon\,dq.$$

But

$$\begin{aligned}\partial_t\log m_t^\epsilon&=\frac{\partial_t m_t^\epsilon}{m_t^\epsilon}=\epsilon e^{2t}\frac{\Delta m_t^\epsilon}{m_t^\epsilon}=\epsilon e^{2t}\big(\Delta\log m_t^\epsilon+|\nabla\log m_t^\epsilon|^2\big)\\&=\epsilon e^{2t}\big(2\Delta\log\sqrt{m_t^\epsilon}+4|\nabla\log\sqrt{m_t^\epsilon}|^2\big),\end{aligned}$$

where we used $\Delta u/u = \Delta \log u + |\nabla \log u|^2$. This implies that

$$\begin{aligned}\mathcal{U}_t(q) &= 4\epsilon e^{2t} \int_{\mathbb{R}^{dk}} \log \sqrt{m_t^\epsilon}\, dq + 2\epsilon^2 e^{4t} \int_{\mathbb{R}^{dk}} \left(\Delta \log \sqrt{m_t^\epsilon} + |\nabla \log \sqrt{m_t^\epsilon}|^2\right) dq \\ &= 4\epsilon e^{2t} \int_{\mathbb{R}^{dk}} \log \sqrt{m_t^\epsilon}\, dq + 2\epsilon^2 e^{4t} \int_{\mathbb{R}^{dk}} \frac{\Delta \sqrt{m_t^\epsilon}}{\sqrt{m_t^\epsilon}}\, dq.\end{aligned}$$

Its gradient is the *constant* vector field

$$\mathrm{grad}_q^{\mathrm{OW}} \mathcal{U}_t = 4\epsilon e^{2t} \nabla \log \sqrt{m_t^\epsilon} + 4\epsilon^2 e^{4t} \nabla \left(\frac{\Delta \sqrt{m_t^\epsilon}}{2\sqrt{m_t^\epsilon}} \right) = -2\dot{\mathsf{m}}_t^\epsilon - 4\epsilon^2 e^{4t} \nabla [\mathcal{Q}(m_t^\epsilon|\mathrm{Leb})]$$

Finally, the Newton equation we are after is (6.11). □

Lemma 6.13. *Identity* (6.8) *holds.*

Proof. We also obtain (6.8) because the trajectory (m_t^ϵ) trivially minimizes the Lagrangian $\frac{1}{2}\|\dot{\mathsf{q}} - \dot{\mathsf{m}}_t^\epsilon\|_q^2$, implying that it solves the least action principle (with well chosen endpoints). With (6.11), this leads us to $\nabla^{\mathrm{OW}}_{\dot{\mathsf{m}}_t^\epsilon} \dot{\mathsf{m}}_t^\epsilon = 2\dot{\mathsf{m}}_t^\epsilon + 4\epsilon^2 e^{4t} \nabla [\mathcal{Q}(m_t^\epsilon)|\mathrm{Leb}]$. □

Notation.

- For any $\mathsf{y} \in \mathbb{R}^{dk}$, $\quad \mathsf{y} = (y_i^l)_{1\le i\le k, 1\le l\le d} = (\underbrace{y_1^1, \dots, y_1^d}_{\mathsf{y}_1 \in \mathbb{R}^d}; \dots; \underbrace{y_k^1, \dots, y_k^d}_{\mathsf{y}_k \in \mathbb{R}^d}) \in (\mathbb{R}^d)^k$.
- For any $1 \le i \le k$,
  - for any regular $f : \mathbb{R}^{dk} \to \mathbb{R}$, $\quad \partial_i f = \partial_{\mathsf{y}_i} f = (\partial_{y_i^l} f)_{1\le l\le d} \in \mathbb{R}^d$;
  - for any regular $u : \mathbb{R}^{dk} \to \mathbb{R}^d$, $\quad \partial_i u = \partial_{\mathsf{y}_i} u = (\partial_{y_i^l} u^n)_{1\le l,n\le d} \in \mathrm{M}_{d\times d}$.
- For any $\pi \in \mathrm{P}(S)$,
  - for any vector valued function u on S (as S is a finite set, u is a vector), $\langle u, \pi\rangle := \int_S u(\mathsf{x})\, \pi(d\mathsf{x}) =: \langle u(\mathsf{x}), \pi\rangle$ where last identity is a practical abuse of notation which permits us to write for instance $\langle \mathsf{x}, \pi\rangle = \int_S \mathsf{x}\, \pi(d\mathsf{x})$;
  - once π is clear from the context, for any $\mathsf{x} \in S$, we write $\tilde{\mathsf{x}} := \mathsf{x} - \langle \mathsf{x}, \pi\rangle = \mathsf{x} - \int_S \mathsf{x}'\, \pi(d\mathsf{x}')$ or more specifically $\tilde{\mathsf{x}}(\mathsf{y}) = \tilde{\mathsf{x}}(\mathsf{x}, \mathsf{y}) := \mathsf{x} - \langle \mathsf{x}', \pi(\mathsf{y})\rangle$.
- For any $a = (a^l)_{1\le l\le d}, b = (b^n)_{1\le n\le d} \in \mathbb{R}^d$, $a \otimes b$ is the $d \times d$-matrix defined by $a \otimes b := (a^l b^n)_{1\le l,n\le d}$.

Lemma 6.14. $\quad 4\epsilon^2 e^{4t} \nabla [\mathcal{Q}(m_t^\epsilon|\mathrm{Leb})](\mathsf{y}) = -\dot{\mathsf{m}}_t^\epsilon(\mathsf{y}) + F_t^\epsilon(\mathsf{y})$.

Proof. Since ϵ and t are fixed, we do not write them as indices. The leftmost equality in (5.8) is

$$\dot{\mathsf{m}} = -2\epsilon e^{2t} \nabla \log \sqrt{m}.$$

It implies

$$\begin{aligned}\mathcal{Q}(m|\mathrm{Leb}) &= -\tfrac{1}{2}\Delta \log \sqrt{m} - \tfrac{1}{2}|\nabla \log \sqrt{m}|^2 = -\tfrac{1}{2}\nabla\cdot\nabla \log \sqrt{m} - \tfrac{1}{2}|\nabla \log \sqrt{m}|^2 \\ &= \frac{1}{4\epsilon e^{2t}} \nabla\cdot\dot{\mathsf{m}} - \frac{1}{8\epsilon^2 e^{4t}} |\dot{\mathsf{m}}|^2,\end{aligned}$$

leading to

$$4\epsilon^2 e^{4t} \nabla [\mathcal{Q}(m_t^\epsilon|\mathrm{Leb})] = \epsilon e^{2t} \nabla(\nabla\cdot\dot{\mathsf{m}}) - \nabla(\|\dot{\mathsf{m}}\|^2)/2. \tag{6.15}$$

This shows that $\mathsf{y} \mapsto \dot{\mathsf{m}}(\mathsf{y})$ is an ingredient that we have to work with. It will be convenient to use its representation (6.1):

$$\dot{\mathsf{m}}(\mathsf{y}) = \mathsf{y} - \langle \mathsf{x}, \pi(\mathsf{y})\rangle.$$

The basic block of our calculation is, for any $1 \le i \le k$ and any $\sigma \in \mathcal{S}$,

$$\partial_i \pi^\sigma(\mathsf{y}) = (\epsilon e^{2t})^{-1}\pi^\sigma(\mathsf{y})\tilde{\mathsf{x}}_i^\sigma(\mathsf{y}), \quad \forall \mathsf{y} \in \mathbb{R}^{dk}. \tag{6.16}$$

Let us show it. First of all

$$\partial_i w^\sigma(\mathsf{y}) = -(\epsilon e^{2t})^{-1}(\mathsf{y}_i - \mathsf{x}_i^\sigma)w^\sigma(\mathsf{y}).$$

Hence,

$$\begin{aligned}\partial_i \pi^\sigma &= \frac{\partial_i w^\sigma}{\sum_{\sigma'} w^{\sigma'}} - \frac{w^\sigma \sum_{\sigma'} \partial_i w^{\sigma'}}{(\sum_{\sigma'} w^{\sigma'})^2} = -(\epsilon e^{2t})^{-1}\pi^\sigma \Big[\mathsf{y}_i - \mathsf{x}_i^\sigma - \sum_{\sigma'}(\mathsf{y}_i - \mathsf{x}_i^{\sigma'})\pi^{\sigma'}\Big] \\ &= (\epsilon e^{2t})^{-1}\pi^\sigma(\mathsf{x}_i^\sigma - \langle \mathsf{x}_i, \pi\rangle) = (\epsilon e^{2t})^{-1}\pi^\sigma \tilde{\mathsf{x}}_i^\sigma,\end{aligned}$$

which is (6.16). This implies that for any $1 \le i, j \le k$,

$$\partial_j \dot{\mathsf{m}}_i = \delta_{ij}\mathrm{Id}_{\mathbb{R}^d} - (\epsilon e^{2t})^{-1}\langle \tilde{\mathsf{x}}_i \otimes \tilde{\mathsf{x}}_j, \pi\rangle, \tag{6.17}$$

because

$$\begin{aligned}\partial_j \dot{\mathsf{m}}_i(\mathsf{y}) &= \partial_j(\mathsf{y}_i - \langle \mathsf{x}_i, \pi(\mathsf{y})\rangle) = \delta_{ij}\mathrm{Id}_{\mathbb{R}^d} - \langle \mathsf{x}_i \otimes \partial_j \pi(\mathsf{y})\rangle \\ &\overset{(6.16)}{=} \delta_{ij}\mathrm{Id}_{\mathbb{R}^d} - (\epsilon e^{2t})^{-1}\langle \mathsf{x}_i \otimes \tilde{\mathsf{x}}_j(\mathsf{y}), \pi(\mathsf{y})\rangle = \delta_{ij}\mathrm{Id}_{\mathbb{R}^d} - (\epsilon e^{2t})^{-1}\langle \tilde{\mathsf{x}}_i(\mathsf{y}) \otimes \tilde{\mathsf{x}}_j(\mathsf{y}), \pi(\mathsf{y})\rangle\end{aligned}$$

since $\langle \tilde{\mathsf{x}}_j, \pi\rangle = 0$. Consequently,

$$\nabla\cdot \dot{\mathsf{m}} = \sum_{i,l} \partial_{y_i^l} \dot{m}_i^l = \sum_{i,l} \Big(1 - (\epsilon e^{2t})^{-1}\langle (\tilde{x}_i^l)^2, \pi\rangle\Big) = kd - (\epsilon e^{2t})^{-1}\langle \|\tilde{\mathsf{x}}\|^2, \pi\rangle$$

and for any $1 \le i \le k$,

$$\begin{aligned}\epsilon e^{2t}\partial_i(\nabla\cdot \dot{\mathsf{m}}) &= -\partial_i\langle \|\tilde{\mathsf{x}}\|^2, \pi\rangle = -\langle \partial_i \|\tilde{\mathsf{x}}\|^2, \pi\rangle - \langle \|\tilde{\mathsf{x}}\|^2, \partial_i \pi\rangle \\ &\overset{(6.16)}{=} -\langle \partial_i \|\tilde{\mathsf{x}}\|^2 + (\epsilon e^{2t})^{-1}\tilde{\mathsf{x}}_i, \pi\rangle.\end{aligned}$$

As

$$\partial_i \|\tilde{\mathsf{x}}\|^2 = \partial_i \sum_j |\tilde{\mathsf{x}}_j|^2 = 2\sum_j [\partial_i \tilde{\mathsf{x}}_j]\,\tilde{\mathsf{x}}_j \overset{(6.16)}{=} -2(\epsilon e^{2t})^{-1}\sum_j \langle \tilde{\mathsf{x}}_i \otimes \tilde{\mathsf{x}}_j, \pi\rangle\,\tilde{\mathsf{x}}_j,$$

we have

$$\langle \partial_i \|\tilde{\mathsf{x}}\|^2, \pi\rangle = 0,$$

because $\langle \tilde{\mathsf{x}}_j, \pi\rangle = 0$. We finally obtain

$$\epsilon e^{2t}\partial_i(\nabla\cdot \dot{\mathsf{m}}) = -(\epsilon e^{2t})^{-1}\langle \|\tilde{\mathsf{x}}\|^2\,\tilde{\mathsf{x}}_i, \pi\rangle.$$

On the other hand,

$$\begin{aligned}\partial_i(\|\dot{\mathsf{m}}\|^2/2) &= \partial_i \sum_j |\dot{\mathsf{m}}_j|^2/2 = \sum_j (\partial_i \dot{\mathsf{m}}_j)\,\dot{\mathsf{m}}_j \overset{(6.17)}{=} \sum_j \delta_{ij}\dot{\mathsf{m}}_j - (\epsilon e^{2t})^{-1}\sum_j \langle \tilde{\mathsf{x}}_i \otimes \tilde{\mathsf{x}}_j, \pi\rangle\,\dot{\mathsf{m}}_j \\ &= \dot{\mathsf{m}}_i - (\epsilon e^{2t})^{-1}\sum_j \langle \tilde{\mathsf{x}}_i \otimes \tilde{\mathsf{x}}_j, \pi\rangle\,\dot{\mathsf{m}}_j.\end{aligned}$$

These last two identities, together with (6.15), lead us to

$$4\epsilon^2 e^{4t}\nabla[\mathcal{Q}(m|\mathrm{Leb})] = -\dot{\mathsf{m}}_t^\epsilon + (\epsilon e^{2t})^{-1}\Big(\langle \tilde{\mathsf{x}} \otimes \tilde{\mathsf{x}}, \pi\rangle\,\dot{\mathsf{m}} - \langle \|\tilde{\mathsf{x}}\|^2\,\tilde{\mathsf{x}}, \pi\rangle\Big).$$

As $\dot{\mathsf{m}}(\mathsf{y}) = \mathsf{y} - \langle \mathsf{x}, \pi\rangle$ does not depend on the variable x, which is integrated, we see that

$$\begin{aligned}\langle \tilde{\mathsf{x}} \otimes \tilde{\mathsf{x}}, \pi\rangle\,\dot{\mathsf{m}} - \langle \|\tilde{\mathsf{x}}\|^2\,\tilde{\mathsf{x}}, \pi\rangle &= \langle [\tilde{\mathsf{x}} \otimes \tilde{\mathsf{x}}]\,\dot{\mathsf{m}} - \|\tilde{\mathsf{x}}\|^2\,\tilde{\mathsf{x}}, \pi\rangle = \langle [\tilde{\mathsf{x}} \otimes \tilde{\mathsf{x}}]\,(\dot{\mathsf{m}} - \tilde{\mathsf{x}}), \pi\rangle \\ &= \langle [\tilde{\mathsf{x}} \otimes \tilde{\mathsf{x}}]\,(\mathsf{y} - \mathsf{x}), \pi\rangle = \Big\langle \big((\mathsf{y} - \mathsf{x})\cdot\tilde{\mathsf{x}}\big)\,\tilde{\mathsf{x}}, \pi\Big\rangle.\end{aligned}$$

This completes the proof of the lemma. □

Since $(\epsilon e^{2t})^{-1}\big\langle\{(\mathsf{y}-\mathsf{x})\cdot\tilde{\mathsf{x}}\}\,\tilde{\mathsf{x}},\pi^\epsilon\big\rangle$ might diverge as ϵ tends to zero, we have a closer look at it. Let us denote the energy gap

$$c(\mathsf{y}) := \min_{\mathsf{x}\in S\setminus S_*(\mathsf{y})}\|\mathsf{y}-\mathsf{x}\|^2/2 - \min_{\mathsf{x}\in S}\|\mathsf{y}-\mathsf{x}\|^2/2 > 0.$$

Lemma 6.18. *Let $r := \|\mathsf{x}\|$, $\mathsf{x}\in S$, be the common norm of the elements of S. There is some $a>0$ such that for any nonzero $\mathsf{y}\in\mathbb{R}^{dk}$,*

(a) $0 < a\|\mathsf{y}\| \le c(\mathsf{y}) \le 2r\|\mathsf{y}\|$ *and*

(b) $|F_t^\epsilon(\mathsf{y}) - \epsilon^{-1}e^{-2t}A_*(\mathsf{y})| \le 8k!r^2(\|\mathsf{y}\|+r)e^{-2t}\epsilon^{-1}\exp(-e^{-2t}a\|\mathsf{y}\|/\epsilon)$.

(c) $\sup_{\mathsf{y}}|A_*(\mathsf{y})| \le 2r^3$.

(d) *Furthermore, if $n_*(\mathsf{y})=1$ or $n_*(\mathsf{y})=2$, then $A_*(\mathsf{y})=0$.*
But this generally fails when $n_*(\mathsf{y})\ge 3$.

Proof. Let us prove (a). For $\mathsf{y}=0$, $S_*(0)=S$. Hence, $c(0)=+\infty$ because a minimum on an empty set can be set as infinite (this is coherent with the bounds to appear below). For any nonzero y, denote $\mathsf{x}_*(\mathsf{y})$ an element of $S_*(\mathsf{y})$ and $\hat{\mathsf{x}}(\mathsf{y})$ a minimizer of $\min_{\mathsf{x}\in S\setminus S_*(\mathsf{y})}\|\mathsf{y}-\mathsf{x}\|^2/2$, so that

$$\begin{aligned}c(\mathsf{y}) &= \|\mathsf{y}-\hat{\mathsf{x}}(\mathsf{y})\|^2/2 - \|\mathsf{y}-\mathsf{x}_*(\mathsf{y})\|^2/2 \\ &= (\mathsf{x}_*-\hat{\mathsf{x}})(\mathsf{y})\cdot\mathsf{y} + \|\hat{\mathsf{x}}(\mathsf{y})\|^2/2 - \|\mathsf{x}_*(\mathsf{y})\|^2/2 = (\mathsf{x}_*-\hat{\mathsf{x}})(\mathsf{y})\cdot\mathsf{y}\end{aligned}$$

because $\|\hat{\mathsf{x}}(\mathsf{y})\| = \|\mathsf{x}_*(\mathsf{y})\| = r$. The same computation shows that x_* is an orthogonal projection of y on S if and only if $\mathsf{x}_*\cdot\mathsf{y}\ge\mathsf{x}\cdot\mathsf{y}$, for all $\mathsf{x}\in S$. It follows that $\mathsf{y}\mapsto\mathsf{x}_*(\mathsf{y})$ and $\mathsf{y}\mapsto\hat{\mathsf{x}}(\mathsf{y})$ are functions of the unit vector $\mathsf{u}_\mathsf{y} := \mathsf{y}/\|\mathsf{y}\|$. Therefore, for any $\mathsf{y}\neq 0$, $0 < c(\mathsf{y}) = \|\mathsf{y}\|\,\mathsf{u}_\mathsf{y}\cdot(\hat{\mathsf{x}}-\mathsf{x}_*)(\mathsf{u}_\mathsf{y})$. As S is a finite set, $a := \inf_{\mathsf{u}:\|\mathsf{u}\|=1}\mathsf{u}\cdot(\hat{\mathsf{x}}-\mathsf{x}_*)(\mathsf{u}) > 0$. Finally,

$$0 < a\|\mathsf{y}\| \le c(\mathsf{y}) = \mathsf{u}_\mathsf{y}\cdot(\hat{\mathsf{x}}-\mathsf{x}_*)(\mathsf{u}_\mathsf{y})\,\|\mathsf{y}\| \le 2r\|\mathsf{y}\|.$$

Let us prove (b). We easily see that the total variation between $\pi_t^\epsilon(\mathsf{y})$ and $\pi_*(\mathsf{y})$ is upper bounded by

$$\|\pi_t^\epsilon(\mathsf{y}) - \pi_*(\mathsf{y})\|_{\mathrm{TV}} \le 2k!n_*(\mathsf{y})^{-1}\exp(-e^{-2t}c(\mathsf{y})/\epsilon) \le 2k!\exp(-e^{-2t}c(\mathsf{y})/\epsilon).$$

This implies that

$$\big|\big\langle\{(\mathsf{y}-\mathsf{x})\cdot\tilde{\mathsf{x}}\}\,\tilde{\mathsf{x}},\pi_t^\epsilon(\mathsf{y})-\pi_*(\mathsf{y})\big\rangle\big| \le 8k!(\|\mathsf{y}\|+r)r^2\,\exp(-e^{-2t}c(\mathsf{y})/\epsilon).$$

It remains to evaluate and bound

$$A := \big\langle\{(\mathsf{y}-\mathsf{x})\cdot\tilde{\mathsf{x}}\}\,\tilde{\mathsf{x}},\pi_*(\mathsf{y})\big\rangle = n_*^{-1}\sum_{1\le n\le n_*}\{(\mathsf{y}-\mathsf{x}_*^n)\cdot(\mathsf{x}_*^n-\bar{\mathsf{x}}_*)\}\,(\mathsf{x}_*^n-\bar{\mathsf{x}}_*),$$

where we drop the explicit dependence on y. We are going to take advantage of both invariances: $\|\mathsf{x}_*^n\| = r$ and $\|\mathsf{y}-\mathsf{x}_*^n\| = \ell$ for all $1\le n\le n_*$. As for any n, $\ell^2 := \|\mathsf{y}-\mathsf{x}_*^n\|^2 = \|\mathsf{y}\|^2+\|\mathsf{x}_*^n\|^2-2\mathsf{y}\cdot\mathsf{x}_*^n = \|\mathsf{y}\|^2+r^2-2\mathsf{y}\cdot\mathsf{x}_*^n$, we see that $\mathsf{y}\cdot\mathsf{x}_*^n$ does not depend on n. Hence, $\mathsf{y}\cdot(\mathsf{x}_*^n-\bar{\mathsf{x}}_*)=0$ for all n, and

$$\begin{aligned}A &= -n_*^{-1}\sum_{1\le n\le n_*}\{\mathsf{x}_*^n\cdot(\mathsf{x}_*^n-\bar{\mathsf{x}}_*)\}\,(\mathsf{x}_*^n-\bar{\mathsf{x}}_*) \\ &= -n_*^{-1}\sum_{1\le n\le n_*}\{\|\mathsf{x}_*^n\|^2-\bar{\mathsf{x}}_*\cdot\mathsf{x}_*^n\}\,(\mathsf{x}_*^n-\bar{\mathsf{x}}_*) = -n_*^{-1}\sum_{1\le n\le n_*}\{r^2-\bar{\mathsf{x}}_*\cdot\mathsf{x}_*^n\}\,(\mathsf{x}_*^n-\bar{\mathsf{x}}_*) \\ &= n_*^{-1}\sum_{1\le n\le n_*}\{\bar{\mathsf{x}}_*\cdot\mathsf{x}_*^n\}\,(\mathsf{x}_*^n-\bar{\mathsf{x}}_*) = \mathrm{Cov}(\mathsf{x}_*^1\ldots,\mathsf{x}^{n_*})\,\bar{\mathsf{x}}_*\end{aligned}$$

where the last two equalities follow from $\sum_n (\mathsf{x}_*^n - \bar{\mathsf{x}}_*) = 0$.

Let us prove (c): $A = n_*^{-1} \sum_{1\le n\le n_*} \{\bar{\mathsf{x}}_* \cdot \mathsf{x}_*^n\} (\mathsf{x}_*^n - \bar{\mathsf{x}}_*)$ implies $\sup_{\mathsf{y}} |A_*(\mathsf{y})| \le 2r^3$.

Let us prove (d).

(i) If $n_* = 1$, then $A = 0$ because $\mathsf{x}_* = \bar{\mathsf{x}}_*$.

(ii) If $n_* = 2$, we also have $A = 0$ because, denoting $\mathsf{x}_*^1 = a$ and $\mathsf{x}_*^2 = b$,

$$A = \frac{1}{2}\Big(\frac{a+b}{2}\cdot a\Big)\frac{a-b}{2} + \frac{1}{2}\Big(\frac{a+b}{2}\cdot b\Big)\frac{b-a}{2}$$
$$= \Big((a+b)\cdot(b-a)\Big)(b-a)/8 = (\|b\|^2 - \|a\|^2)\,(b-a)/8 = 0.$$

Last equality holds because $\|a\| = \|b\|$, since a and b belong to S.

(iii) If $n_* \ge 3$, this is not true anymore: A does not vanish in general. As an example, take the three vectors $a = (1,0,0,\dots,0)$, $b = (0,1,0,\dots,0)$ and $c = (-1,0,0,\dots,0) = -a$, lying on a sphere centered at zero. Their barycenter $\bar{\mathsf{x}}_*$ is $b/3$, and

$$A = 1/3\ [(a\cdot b/3)\,(a - b/3) + (b\cdot b/3)\,(b - b/3) + (c\cdot b/3)\,(c - b/3)]$$
$$= 1/3\ [(b\cdot b/3)\,(b - b/3)] = 2b/27 \ne 0,$$

because $a\cdot b = c\cdot b = 0$ and $\|b\|^2 = 1$.

This completes the proof of the lemma. □

7. Schrödinger problem

This section is dedicated to already well known results about the large deviation principle for the empirical process of a collection of independent copies of diffusion processes by Dawson and Gärtner [13] and Föllmer [16], its connection with the Schrödinger problem [31, 32, 16, 24] and an expression of its large deviation rate function as a Lagrangian action in the Otto-Wasserstein manifold derived in [11]. This is a preliminary step for the construction at Section 9 of an interacting Brownian particle system whose empirical process satisfies the Gibbs conditioning principle of Statement 7.1 below.

Our goal. To provide some physical representation differing from the enigmatic model (3.7), we are in search for some collection $(\widetilde{X}^N)_{N\ge1}$ of random elements in the set $\Omega_{\mathrm{P}}^{(k)} := C([s_0,s_1], \mathrm{P}(\mathbb{R}^{dk}))$ of all continuous paths on the set $\mathrm{P}(\mathbb{R}^{dk})$ of k-fluids which satisfies the following

Statement 7.1. (Gibbs conditioning principle). *For any probability measures α and β on $\mathbb{R}^{dk}$, conditionally on $\widetilde{X}^N(s_0) \simeq \alpha$ and $\widetilde{X}^N(s_1) \simeq \beta$, the most likely trajectory $p \in \Omega_{\mathrm{P}}^{(k)}$ of $\widetilde{X}^N$ as N tends to infinity solves the least action principle* (5.12).

Remarks 7.2.

(a) The fluctuation parameter $\epsilon > 0$ is fixed once for all. We are only concerned by limits as N tends to infinity.

(b) This statement is fuzzy: a rigorous one should consider a Γ-limit along decreasing neighborhoods of α and β.

(c) The desired collection $(\widetilde{X}^N)_{N\ge1}$ of random processes will be built upon a sequence of independent copies the stochastic process X^ϵ already encountered at Section 3.

Time s versus time t. In the present section, we shall stick to time s and consider the least action principle (5.12) rather than its t-version (5.10). Once a particle system corresponds to the least action principle (5.12), it simply remains to apply the parameter setting 3.10 to arrive at (5.10).

A Brownian cloud related to ϵ-MAG for a k-fluid. As a first step, we recall the large deviation principle satisfied by the empirical process

$$X^N : s \in [s_0, s_1] \mapsto \frac{1}{N} \sum_{1 \le i \le N} \delta_{\mathsf{X}_i^\epsilon(s)} \in \mathrm{P}\big(\mathbb{R}^{dk}\big) \tag{7.3}$$

of a sequence $(\mathsf{X}_i^\epsilon)_{i\ge 1}$ of independent copies of X^ϵ, see (3.1) and (3.2), that is

$$\mathrm{Law}(\mathsf{X}_i^\epsilon)_{i\ge1} = (R^\epsilon)^{\otimes\infty}$$

where R^ϵ is the law of the process X^ϵ.
The random process X^N describes a Brownian cloud of N particles evolving in $\mathbb{R}^{dk}$. Its large deviations in $\Omega_{\mathrm{P}}^{(k)}$:

$$\mathrm{Proba}(X^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- N \inf_{p\in\bullet} J(p)\Big),$$

are well known since the pioneering article [13] by Dawson and Gärtner, and its presentation by Föllmer in [16]. The rate function $p \in \Omega_{\mathrm{P}}^{(k)} \mapsto J(p) \in [0,\infty]$ is expressed below at Proposition 7.11. This implies that, for any two prescribed time marginals α and β in $\mathrm{P}(\mathbb{R}^{dk})$,

$$\begin{aligned}\mathrm{Proba}(X^N \in \bullet \mid X^N(s_0) \simeq \alpha, X^N(s_1) \simeq \beta)&\\ \underset{N\to\infty}{\asymp} \exp\Big(- N \Big[\inf_{p\in\bullet, p_{s_0}=\alpha, p_{s_1}=\beta} J(p) - &\inf_{p: p_{s_0}=\alpha, p_{s_1}=\beta} J(p)\Big]\Big),\end{aligned}$$

which in turns implies the following

Statement 7.4. (Gibbs conditioning principle). *For any probability measures α and β on $\mathbb{R}^{dk}$, conditionally on $X^N(s_0) \simeq \alpha$ and $X^N(s_1) \simeq \beta$, the most likely trajectory $p \in \Omega_{\mathrm{P}}^{(k)}$ of X^N as N tends to infinity solves the least action principle*

$$\inf J(p), \quad p \in \Omega_{\mathrm{P}}^{(k)} : p_{s_0} = \alpha, p_{s_1} = \beta. \tag{7.5}$$

We focus on X^N because the minimization problem (7.5) happens to be close to the least action principle (5.12) we are after, see Proposition 7.13 below. Let us give some indications about the computation of J. The random paths Z_i^ϵ take their values in the space

$$\Omega^{(k)} := C([s_0, s_1], \mathbb{R}^{dk})$$

of all continuous paths on $\mathbb{R}^{dk}$. The large deviation principle as N tends to infinity of their empirical measures

$$\widehat{X}^N := \frac{1}{N} \sum_{1\le i\le N} \delta_{\mathsf{X}_i^\epsilon} \in \mathrm{P}(\Omega^{(k)}),$$

which take their values in the set $\mathrm{P}(\Omega^{(k)})$ of all probability measures on $\Omega^{(k)}$, is given by Sanov's theorem, see [14] for instance, which roughly states that

$$\mathrm{Proba}(\widehat{X}^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- N \inf_{P\in\bullet} H(P|R^\epsilon)\Big)$$

where

$$H(P|R^\epsilon) := E_P \log(dP/dR^\epsilon) \in [0,\infty], \quad P \in \mathrm{P}(\Omega^{(k)}),$$

is the relative entropy of P with respect to R^ϵ. By the contraction principle, the large deviation rate function for X^N is

$$J(p) = \inf\{H(P|R^\epsilon); P \in \mathrm{P}(\Omega^{(k)}) : P_s = p_s, \forall s \in [s_0, s_1]\}, \quad p = (p_s)_{s_0\le s\le s_1} \in \Omega_{\mathrm{P}}^{(k)} \tag{7.6}$$

where $P_s \in \mathrm{P}(\mathbb{R}^{dk})$ is the s-th marginal of P.

Schrödinger problem. The minimization problem (7.5) is also called the Schrödinger problem. It was addressed and solved (at least formally) in 1931 by Schrödinger in [31, 32]. In view of (7.6), its solution (p_s) is the time marginal flow of the solution P, i.e.

$$p_s = P_s, \quad s_0 \le s \le s_1, \tag{7.7}$$

of the following entropy minimization problem

$$\inf H(P|R^\epsilon), \quad P \in \mathrm{P}(\Omega^{(k)}) : P_{s_0} = \alpha, P_{s_1} = \beta. \tag{7.8}$$

This formulation in terms of the relative entropy is due to Föllmer [16]. Problem (7.8) admits at most one solution because $H(\bullet|R^\epsilon)$ is strictly convex and the constraint set $\{P \in \mathrm{P}(\Omega^{(k)}) : P_{s_0} = \alpha, P_{s_1} = \beta\}$ is convex. If it exists, it is called the *Schrödinger bridge* with respect to R between α and β. Its marginal flow (7.7) is called the *entropic interpolation* with respect to R between α and β. The entropy minimization problem (7.8) is called the Schrödinger bridge problem.

The function J**.** Finding the minimizer $P \in \mathrm{P}(\Omega^{(k)})$ of (7.6) is called the Dawson-Gärtner problem [13]. Next proposition gives its solution together with an expression of the large deviation rate function J.

Proposition 7.9. *If $J(p)$ is finite, this infimum is uniquely attained at some $P(p) \in \mathrm{P}(\Omega^{(k)})$ which is Markov with a gradient drift field $\overrightarrow{\mathsf{v}_s} \in H^{-1}_{\mathbb{R}^{dk}}(p_s)$ for almost every s, i.e. $P(p)$ solves the martingale problem with Markov generator $\partial_s + \overrightarrow{\mathsf{v}_s}\cdot\nabla + \epsilon\Delta/2$. Moreover,*

$$J(p) = H(P(p)|R^\epsilon) = H(p_{s_0}|r^\epsilon_{s_0}) + \epsilon^{-1}\int_{s_0}^{s_1} \tfrac{1}{2}\|\overrightarrow{\mathsf{v}_s}\|^2_{p_s}\, ds. \tag{7.10}$$

Proof. See [15, 16] for the proof of this result, and also [13] for the original proof leading to an alternate equivalent expression of $J(p)$. □

We prefer departing from the standard representation (7.10) of J which was put forward in [13, 16], to exploit the following alternate representation.

Proposition 7.11. *The large deviation rate function J is given for any $p \in \Omega^{(k)}_{\mathrm{P}}$ by*

$$J(p) = \tfrac{1}{2}H(p_{s_0}|r^\epsilon_{s_0}) + \tfrac{1}{2}H(p_{s_1}|r^\epsilon_{s_1}) + \epsilon^{-1}\int_{s_0}^{s_1}\tfrac{1}{2}\|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\, ds + \epsilon\int_{s_0}^{s_1} I(p_s|r^\epsilon_s)\, ds. \tag{7.12}$$

Proof. For a proof of the identity (7.12) which relies on time reversal, see [11, §6]. □

In formula (7.12),

$$I(p|r) := \tfrac{1}{2}\left\|\nabla \log\sqrt{\frac{dp}{dr}}\right\|^2_p$$

is the Fisher information of p with respect to r, r^ϵ_s is the s-th marginal of R^ϵ, and

$$\dot{\mathsf{p}}_s = \overrightarrow{\mathsf{v}_s} - \epsilon\nabla\log\sqrt{p_s} =: \mathsf{v}^{\mathrm{cu},P(p)}_s$$

is the current velocity of $P(p)$. In particular, $\dot{\mathsf{p}}_s$ belongs to $H^{-1}_{\mathbb{R}^{dk}}(p_s)$ for almost every s and it satisfies the continuity equation (5.3) in the weak sense. Consequently, we obtain

Proposition 7.13. *The least action principle (7.5) is equivalent to*

$$\inf_p \int_{s_0}^{s_1}\left(\tfrac{1}{2}\|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s} + \epsilon^2 I(p_s|r^\epsilon_s)\right) ds \tag{7.14}$$

where the infimum runs through all the $p \in \Omega^{(k)}_{\mathrm{P}}$ satisfying $p_{s_0} = \alpha$ and $p_{s_1} = \beta$.

Hence, a restatement of Statement 7.4 is

Statement 7.15. (Gibbs conditioning principle). *For any probability measures α and β on $\mathbb{R}^{dk}$, conditionally on $X^N(s_0) \simeq \alpha$ and $X^N(s_1) \simeq \beta$, the most likely trajectory $p \in \Omega_{\mathrm{P}}^{(k)}$ of X^N as N tends to infinity solves the least action principle (7.14).*

Comparing (7.14) with (5.12): $\inf_{(p)} \int_{s_0}^{s_1} \frac{1}{2} \|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s} \kappa_s \, ds$, we see that it is necessary

- to introduce the coeffective κ_s into (7.14) and
- to remove from (7.14) the Fisher information term.

8. Plugging κ_s in

Our aim is to introduce the coeffective κ_s into (7.14). The main result of this section is

Theorem 8.1. *Let Z be a continuous Markov process taking its values in $\mathbb{R}^d$ on the time interval $[s_0, s_1]$. Let $(Z''_i; i \geq 1)$ be an iid sequence of copies of Z and denote $Z^N_s := N^{-1} \sum_{i=1}^N \delta_{Z''_i(s)}$, $s_0 \leq s \leq s_1$, the corresponding empirical process. We assume that $(Z^N)_{N\geq 1}$ obeys the large deviation principle:* $\mathrm{Proba}(Z^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(-N \inf_{\zeta\in\bullet} I(\zeta)\Big)$, *in $\mathrm{P}(C([s_0, s_1], \mathbb{R}^d))$ with rate function*

$$I(\zeta) = I_{s_0}(\zeta_{s_0}) + \int_{s_0}^{s_1} L(s, \zeta_s, \dot{\zeta}_s) \, ds, \quad \zeta = (\zeta_s)_{s_0 \leq s \leq s_1}, \tag{8.2}$$

for some Lagrangian function L on the tangent bundle of the Otto-Wasserstein manifold, recall (4.13).
Let $s \in [s_0, s_1] \mapsto \kappa_s \in (0, \infty)$ be a continuously differentiable positive function.
Let $(\overline{Z}^N)_{N\geq 1}$ be the sequence of modified empirical processes described at page 40, see (8.13), (8.15), (8.17) and (8.18).
Then, $(\overline{Z}^N)_{N\geq 1}$ obeys the large deviation principle:

$$\mathrm{Proba}(\overline{Z}^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(-N \inf_{\zeta\in\bullet} I^\kappa(\zeta)\Big),$$

in $\mathrm{P}(C([s_0, s_1], \mathbb{R}^d))$ with rate function

$$I^\kappa(\zeta) = \kappa_{s_0} I_{s_0}(\zeta_{s_0}) + \int_{s_0}^{s_1} \kappa_s L(s, \zeta_s, \dot{\zeta}_s) \, ds, \quad \zeta = (\zeta_s)_{s_0 \leq s \leq s_1}. \tag{8.3}$$

Sketch of proof. Approximate $\overline{Z}^N$ by a sequence $\big((\widetilde{Z}^{R,K,N})_{N\geq 1}\big)_{R,K\geq 1}$ that is constructed as in the first model to be described below at page 37, see (8.6) and (8.7). Then, relying on (8.16), apply the abstract theorem [14, Thm. 4.2.16] on exponentially good approximations. □

We only give a sketch of the proof to save time. However, the main arguments are exposed in the next pages.

Preliminary considerations. Let $(Z_i; i \geq 1)$ be a sequence of independent copies of some Markov process $(Z_s)_{s_0 \leq s \leq s_1}$ taking its values in $\mathbb{R}^d$ and such that the empirical process

$$Z^N_s := N^{-1} \sum_{i=1}^N \delta_{Z_i(s)}, \quad s_0 \leq s \leq s_1,$$

obeys the large deviation principle

$$\text{Proba}(Z^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- N \inf_{\zeta\in\bullet} I(\zeta)\Big)$$

with the rate function I given at (8.2). In the general setting of the present section, we want to find some modification $Z^{\kappa,N}$ of X^N which obeys the large deviation principle

$$\text{Proba}(Z^{\kappa,N} \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- N \inf_{\zeta\in\bullet} I^\kappa(\zeta)\Big)$$

with the modified rate function I^κ given at (8.3).

The key idea for this purpose is the following easy remark. Sanov's theorem states that the empirical measure $\widehat{Z}^N := N^{-1}\sum_{i=1}^N \delta_{Z_i}$ obeys the large deviation principle

$$\text{Proba}(\widehat{X}^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- N \inf_{P\in\bullet} H(P|R)\Big)$$

where $H(P|R) = \int \log(dP/dR)\, dP$ is the relative entropy of P with respect to the law R of the Markov process Z. It immediately follows that for any $\kappa > 0$, the sequence of modified empirical measures

$$\widehat{Z}^{\kappa,N} := \lfloor \kappa N\rfloor^{-1} \sum_{i=1}^{\lfloor \kappa N\rfloor} \delta_{Z_i}$$

where $\lfloor a\rfloor$ is the integer value of a, obeys the large deviation principle

$$\text{Proba}(\widehat{Z}^{\kappa,N} \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- \lfloor \kappa N\rfloor \inf_{P\in\bullet} H(P|R)\Big) \simeq \exp\Big(- N \inf_{P\in\bullet} \kappa\, H(P|R)\Big)$$

with rate function $\kappa H(\bullet|R)$ instead of $H(\bullet|R)$. It also follows with the contraction principle that for any continuous mapping Φ, we obtain

$$\text{Proba}(\Phi(\widehat{X}^N) \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- N \inf_{\zeta\in\bullet} I(\zeta)\Big) \quad \text{where} \quad I(\zeta) = \inf_{P:\Phi(P)=\zeta} H(P|R)$$

and similarly

$$\text{Proba}(\Phi(\widehat{Z}^{\kappa,N}) \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(- N \inf_{\zeta\in\bullet} I^\kappa(\zeta)\Big) \quad \text{where} \quad I^\kappa(\zeta) = \inf_{P:\Phi(P)=\zeta} \kappa H(P|R)$$

with rate function

$$I^\kappa = \kappa I. \tag{8.4}$$

Since $Z^N = \Phi(\widehat{Z}^N)$ where Φ is the continuous application mapping a path measure P to its flow of marginal measures $\Phi(P) = (P_s)_{s_0\le s\le s_1}$, we see that if I denotes the large deviation rate function as N tends to infinity of the empirical process Z^N, then the modified empirical process

$$\lfloor \kappa N\rfloor^{-1} \sum_{i=1}^{\lfloor \kappa N\rfloor} \delta_{Z_i(s)}, \quad s_0 \le s \le s_1,$$

obeys the large deviation principle with rate function $I^\kappa = \kappa I$. Assume in addition that Z is Markov so that the rate function I is additive in time, and more precisely that it has the form of the action functional (8.2). Of course, the rate function $I^\kappa = \kappa I$ of the modified empirical process is associated with the Lagrangian κL instead of L.

Considering a varying coeffective $s \mapsto \kappa_s$, it is tempting to guess that the modified empirical process defined by

$$\lfloor \kappa_s N \rfloor^{-1} \sum_{i=1}^{\lfloor \kappa_s N \rfloor} \delta_{Z_i(s)}, \quad s_0 \le s \le s_1, \tag{8.5}$$

obeys the large deviation principle with the rate function I^κ defined at (8.3).

But this does *not* hold. The reason for this is that during a small time interval $[s, s+h]$, the "new" particles $j \in \{\lfloor \kappa_s N \rfloor + 1, \dots, \lfloor \kappa_{s+h} N \rfloor\}$ to be (a): added if $\kappa_s < \kappa_{s+h}$, or (b): removed if $\kappa_s > \kappa_{s+h}$, are sampled from the random collection (a): $(Z_i)_{i \ge \lfloor \kappa_s N \rfloor}$, or (b): $(Z_i)_{i \le \lfloor \kappa_s N \rfloor}$. This *random* sampling is subject to large deviations so that some additional cost must be added to the Lagrangian cost.
To fix the idea, we assume up to page 41 that $s \mapsto \kappa_s$ *is increasing*. This is satisfied for our model ($\kappa_s = 2s$, see the parameter setting 3.10). The general case will be considered later at (8.17) and (8.18).
This additional cost is the large deviation cost for observing the empirical measure of the newcomers

$$(\lfloor \kappa_{s+h} N \rfloor - \lfloor \kappa_s N \rfloor)^{-1} \sum_{j \in \{\lfloor \kappa_s N \rfloor + 1, \dots, \lfloor \kappa_{s+h} N \rfloor\}} \delta_{Z'_j(s)}$$

close to the actual state $\lfloor \kappa_s N \rfloor^{-1} \sum_{i=1}^{\lfloor \kappa_s N \rfloor} \delta_{Z_i(s)} \simeq \zeta_s$ which might be far from the most likely state. Therefore, the large deviations of the empirical process (8.5) are not governed by the desired rate function (8.3).
This remark indicates the way to a good candidate: one must replace the random sampling of the newcomers by an *almost deterministic* one to control the distance between the actual empirical measure and the empirical measure of the newcomers.

First model. To do so, we approximate the desired empirical process by introducing the newcomers at each time

$$\sigma_k := s_0 + kh, \quad 1 \le k \le K - 1,$$

where $h = (s_1 - s_0)/K$ for some arbitrarily large integer K. Let $\widetilde{Z}^{K,N}$ denote this approximate process. At time σ_k^-, meaning the left limit $\lim_{s \to \sigma_k, s < \sigma_k}$, its value is

$$\widetilde{Z}^{K,N}_{\sigma_k^-} = (N_{k-1})^{-1} \sum_{i=1}^{N_{k-1}} \delta_{Z_i(\sigma_k^-)}, \quad \text{where} \quad N_k := \lfloor \kappa_{\sigma_k} N \rfloor.$$

Beware: although we use the same notation Z_i for our random paths, they are not independent anymore as previously.
The newcomers $\{Z'_j(\sigma_k); N_{k-1} < j \le N_k\}$ at time σ_k are sampled from the support of $Z^{K,N}_{\sigma_k^-}$ in such a way that their empirical measure $(N_k - N_{k-1})^{-1} \sum_{j=1}^{N_k - N_{k-1}} \delta_{Z'_j(\sigma_k)}$ is arbitrarily close, as N tends to infinity, to $Z^{K,N}_{\sigma_k^-}$. This is possible, as explained at next subsection, see (8.10) and (8.12). At time σ_k, we state

$$\widetilde{Z}^{K,N}_{\sigma_k} = N_k^{-1} \Big(N_{k-1} \widetilde{Z}^{K,N}_{\sigma_k^-} + \sum_{j=1}^{N_k - N_{k-1}} \delta_{Z'_j(\sigma_k)} \Big) = N_k^{-1} \sum_{i=1}^{N_k} \delta_{Z_i(\sigma_k)} \tag{8.6}$$

where last equality results from some relabeling. Keeping track of the history of these relabeling procedures, one sees that *at each time σ_k some particles branch in such a way that for each k, $\widetilde{Z}^{K,N}_{\sigma_k}$ is arbitrarily close, as N tends to infinity, to $\widetilde{Z}^{K,N}_{\sigma_k^-}$, while the total*

number of particles jumps from N_{k-1} at time σ_k^- to N_k at time σ_k and the total (unit) mass of the empirical process remains constant.
During the time interval $[\sigma_k, \sigma_{k+1})$, we set

$$\widetilde{Z}_s^{K,N} = N_k^{-1} \sum_{i=1}^{N_k} \delta_{Z_i(s)}, \quad \sigma_k \le s < \sigma_{k+1} \tag{8.7}$$

where the random paths $(Z_i(s); \sigma_k \le s < \sigma_{k+1})_{1\le i\le N_k}$ are independent from each other and each particle $1 \le i \le N_k$ starts from $Z_i(\sigma_k)$ at time $s = \sigma_k$, and follows the Markovian evolution of the reference process Z. Since this evolution specifies the Lagrangian $L(s, \zeta_s, \dot\zeta_s)$, with (8.4) we obtain that the empirical process $\widetilde{Z}^{K,N}$ obeys the large deviation principle with rate function

$$I^{\kappa,K}(\zeta) = I_{\kappa_{s_0}}(\zeta_{\kappa_{s_0}}) + \sum_{k=0}^{K-1} \int_{\sigma_k}^{\sigma_{k+1}} \kappa_{\sigma_k} L(s, \zeta_s, \dot\zeta_s)\, ds, \quad \zeta = (\zeta_s)_{s_0\le s\le s_1}.$$

The sum $\sum_k$ comes from the Markov property of the sample trajectories. It remains to let K tend to infinity and remark that the following Γ-limit

$$\Gamma\text{-}\lim_{K\to\infty} I^{\kappa,K} = I^\kappa,$$

holds as soon as $s \mapsto \kappa_s$ is continuous, with I^κ defined at (8.3). This implies that any limit point as K tends to infinity of the sequence of laws of $(\widetilde{Z}^{K,N})_{K\ge1}$ obeys the large deviation principle as N tends to infinity with rate function I^κ, see [28] for this argument. If the law of the random path Z is the unique solution to its martingale problem, for instance if Z is a diffusion process with Lipschitz drift and diffusion fields as in the present setting, then for each $N \ge 1$, there exists a unique limit point: $\mathrm{Law}(\widetilde{Z}^{\infty,N})$, as K tends to infinity. See [23] for this argument.

Choice of the newcomers. During this sketch of proof, we admitted that one can sample the newcomers from the support of $\widetilde{Z}_{\sigma_k^-}^{K,N}$ in such a way that their empirical measure is arbitrarily close, as N tends to infinity, to $\widetilde{Z}_{\sigma_k^-}^{K,N}$. Let us show this.
We want to pick, in an almost deterministic way, $N_k - N_{k-1}$ distinct points y'_j, $1 \le j \le N_k - N_{k-1}$, from the the set $\{Z_i(\sigma_k^-); 1 \le i \le N_{k-1}\}$.
Let us simplify notation. We wish to choose $1 \le n' \le n$ distinct points $y'_1, \dots, y'_{n'}$ among a subset $\{y_1, \dots, y_n\}$ of $\mathbb{R}^d$ in such a manner that $\tilde y^{n'} := (n')^{-1} \sum_{j=1}^{n'} \delta_{y'_j}$ is arbitrarily close, in the sense of narrow convergence, to $\hat y^n := n^{-1} \sum_{i=1}^n \delta_{y_i}$ as n tends to infinity:

$$\lim_{n,n'\to\infty} W_p(\tilde y^{n'}, \hat y^n) = 0 \tag{8.8}$$

where W_p stands for the Wasserstein distance of order $p \ge 1$. Suppose for the moment that the supports of the empirical measures $\hat y^n$ are included in some fixed bounded box:

$$\cup_n \operatorname{supp}(\hat y^n) \subset \Lambda_R := [-R, R]^d, \tag{8.9}$$

for some large enough $R > 0$. Take an integer $m \ge 1$ (which is intended to tend to infinity as n increases) and cover Λ_R by the $(2m)^d$ cubic boxes $B_{k_1,\dots,k_d} := a_{k_1,\dots,k_d} + [0, R/m]^d$ where $a_{k_1,\dots,k_d} := m^{-1}R\,(k_1, \dots, k_d)$ with $k_1, \dots, k_d \in \{-m, -m+1, \dots, m-2, m-1\}$. In each box B, pick arbitrarily $\lfloor \hat y^n(B)\, n' \rfloor$ distinct points from $B \cap \operatorname{supp}(\hat y^n)$. It is because of this controlled arbitrariness that we describe these trials as *almost deterministic*. The collection of these picked points is the main part of the set of newcomers. Their number $n'' = \sum_{k_1,\dots,k_d} \lfloor \hat y^n(B_{k_1,\dots,k_d})\, n' \rfloor$ statisfies $n' - (2m)^d \le n'' \le n'$ because taking the integer part may induce a lack of at most one point in each box. To complete the set of newcomers, simply add $n'-n''$ not already picked points from $\mathrm{supp}(\hat{y}^n)\cap\Lambda_R$. It is immediate to see that for any $p\geq 1$,

$$W_p(\tilde{y}^{n'},\hat{y}^n)\leq D_R\,(1/m+(2m)^d/n') \tag{8.10}$$

where $D_R=2\sqrt{d}R$ is the diameter of the large box Λ_R and D_R/m is the common diameter of the small boxes $B_{k_1,\dots,k_d}$. At (8.15) below, we shall tune the coeffectives m and n' as functions of n in order to obtain (8.8).

An exponentially good approximation. The assumption (8.9) that the supports of all the measures $\hat{y}^n$ are included in some bounded set is crucial for this construction to be valid. But in general the supports of the random measures $\widetilde{Z}^{K,N}_s$ are not uniformly bounded. This is why it is needed to rely on a sequence $\big((\widetilde{Z}^{R,K,N})_{N\geq 1}\big)_{R\geq 1}$ of approximations of $(\widetilde{Z}^{K,N})_{N\geq 1}$ living on larger and larger bounded sets Λ_R which is an exponentially good approximation of $(\widetilde{Z}^{K,N})_{N\geq 1}$ in the sense of the definition [14, Def. 4.2.14] in order to apply the theorem [14, Thm. 4.2.16]. The proxy $\widetilde{Z}^{R,K,N}$ is defined as $\widetilde{Z}^{K,N}$, but the reference Markov random path $(Z_s)_{s_0\leq s\leq s_1}$ is replaced by its stopped version:

$$Z^R_s:=Z_{s\wedge\tau_R},\quad s_0\leq s\leq s_1, \tag{8.11}$$

where $\tau_R:=\inf\{s;s_0\leq s\leq s_1:Z_s\notin\Lambda_R\}\in[s_0,s_1]\cup\{\infty\}$ is the first exit time from the box Λ_R. Assuming that the sample paths of Z are continuous, we have: $\sup_s|Z^R_s|\leq R$. The main argument for proving that this is indeed an exponentially good approximation is the following estimate

$$\lim_{R\to\infty}\sup_K\limsup_{N\to\infty}N^{-1}\log\mathbb{P}\Big(\widehat{Z}^{K,N}\big(\{\omega=(\omega_s)_{s_0\leq s\leq s_1};\sup_{s_0\leq s\leq s_1}|\omega_s|\geq R\}\big)>\delta\Big)=-\infty \tag{8.12}$$

which holds for any $\delta>0$. This result simply follows from Cramér's theorem applied to an iid sequence of Bernoulli random variables with parameter

$$\epsilon(R):=\mathbb{P}(Z^*\geq R)\leq\mathbb{E}Z^*/R.$$

where we set $Z^*:=\sup_{s_0\leq s\leq s_1}|Z_s|$. Indeed, Cramér's theorem states that, for any $K\geq 1$,

$$\limsup_{N\to\infty}N^{-1}\log\mathbb{P}\Big(\widehat{Z}^{K,N}\big(\{\omega=(\omega_s)_{s_0\leq s\leq s_1};\sup_{s_0\leq s\leq s_1}|\omega_s|\geq R\}\big)>\delta\Big)\leq -h_{\epsilon(R)}(\delta)$$

where

$$\begin{aligned}h_{\epsilon(R)}(\delta)&=\delta\log(\delta/\epsilon(R))+(1-\delta)\log\big((1-\delta)/(1-\epsilon(R))\big)\\&\geq\delta\log(\delta/\epsilon(R))+(1-\delta)\log(1-\delta)\geq\delta\log(\delta R/\mathbb{E}Z^*)+(1-\delta)\log(1-\delta),\end{aligned}$$

which implies (8.12) as soon as

$$\mathbb{E}Z^*:=\mathbb{E}\sup_{s_0\leq s\leq s_1}|Z_s|<\infty.$$

Second model. Previous construction is a little bit frustrating, since it does not provide an explicit representation of the limiting process $\widetilde{Z}^{\infty,N}$, as K and R tend to infinity. Based on these previous considerations, we propose *another* sequence of empirical processes $(\overline{Z}^N)_{N\geq 1}$ and some heuristics for proving that its large deviation rate function as N tends to infinity is also I^κ.

The evolution of the empirical process $(\overline{Z}^N_s; s_0 \le s \le s_1)$ is a concatenation of very short periods (as N tends to infinity) separated by the increasing sequence of times

$$\sigma_0 := s_0, \qquad \sigma_{k+1} := \sigma_k + (\kappa'_{\sigma_k} N^{1-a})^{-1}, \quad k \ge 0, \tag{8.13}$$

where $0 < a < 1$ and $\kappa'_s > 0$ is the derivative of $s \mapsto \kappa_s$ which is assumed to be differentiable. Note that, although σ_k depends on N, for a better readability we do not write explicitly this dependence. At time σ_k, the empirical process is

$$\overline{Z}^N_{\sigma_k} = N_k^{-1} \sum_{i=1}^{N_k} \delta_{Z_i(\sigma_k)}, \quad \text{where} \quad N_k := \lfloor \kappa_{\sigma_k} N \rfloor,$$

and during the time interval $[\sigma_k, \sigma_{k+1})$ all the random paths $(Z_i; 1 \le i \le N_k)$ evolve independently and follow the Markovian evolution of Z^R defined at (8.11), where the large box parameter is chosen such that

$$R = O_{N\to\infty}(N^b)$$

for some constant $b > 0$. At time σ_k^-, the empirical process was

$$\overline{Z}^N_{\sigma_k^-} = (N_{k-1})^{-1} \sum_{i=1}^{N_{k-1}} \delta_{Z_i(\sigma_k^-)}.$$

Beware: although we keep the notation Z_i, these random paths are not the same as in the first model.

Let us describe how to choose the newcomers $(Z'_j(\sigma_k); 1 \le j \le N_k - N_{k-1})$ at time σ_k. Remark that the increment ratio $(N_{k+1} - N_k)/(\sigma_{k+1} - \sigma_k)$ must be close to $N\kappa'_{\sigma_k}$. This implies that $N_{k+1} - N_k \approx N\kappa'_{\sigma_k}(\kappa'_{\sigma_k} N^{1-a})^{-1} = N^a$. Therefore,

$$N_{k+1} - N_k = O_{N\to\infty}(N^a).$$

We choose the newcomers $Z'_j(\sigma_k)$ as in previous subsection, with $R = O_{N\to\infty}(N^b)$ as above and the small box parameter

$$m = O_{N\to\infty}(N^c),$$

for some $c > 0$. As we did at (8.6), we set

$$\overline{Z}^N_{\sigma_k} := N_k^{-1}\Big(N_{k-1}\overline{Z}^N_{\sigma_k^-} + \sum_{j=1}^{N_k - N_{k-1}} \delta_{Z'_j(\sigma_k)}\Big). \tag{8.14}$$

We want that the cumulated error vanishes as N tends to infinity:

$$\sum_k W_p(\overline{Z}^N_{\sigma_k^-}, \overline{Z}^N_{\sigma_k}) \underset{N\to\infty}{\longrightarrow} 0.$$

Considering (8.10) with $n = O_{N\to\infty}(N)$, $n' = O_{N\to\infty}(N^a)$, $R = O_{N\to\infty}(N^b)$ and $m = O_{N\to\infty}(N^c)$, one sees that $W_k := W_p(\overline{Z}^N_{\sigma_k^-}, \overline{Z}^N_{\sigma_k}) = O_{N\to\infty}(N^{b-c} + N^{-a+b+cd})$. On the other hand, (8.13) implies that $\Delta\sigma_k := \sigma_{k+1} - \sigma_k = O_{N\to\infty}(N^{a-1})$. Therefore

$$\sum_k W_p(\overline{Z}^N_{\sigma_k^-}, \overline{Z}^N_{\sigma_k}) \le |s_1 - s_0| \sup_k (W_k/\Delta\sigma_k) = O_{N\to\infty}(N^{\max(1-a+b-c, 1-2a+b+cd)})$$

vanishes as N tends to infinity if one chooses $0 < a < 1$, $b > 0$ and $c > 0$ such that $1 - a + b - c < 0$ and $1 - 2a + b + cd < 0$. One can take for instance:

$$a = \frac{d+2}{d+3}, \quad b = \frac{1}{3d(d+3)} \quad \text{and} \quad c = \frac{2d+1}{2d(d+3)}. \tag{8.15}$$

We have shown that

$$\sum_k W_p(\overline{Z}^N_{\sigma_k^-}, \overline{Z}^N_{\sigma_k}) = O_{N\to\infty}(N^{-\gamma}), \quad \text{for some} \quad \gamma > 0, \quad \text{almost surely.} \tag{8.16}$$

General case for κ. Up to now we decided, to make things easier, to assume that $s \mapsto \kappa_s$ is increasing. The general case where κ is any positive continuously differentiable function is simply obtained by replacing (8.14) by

$$\overline{Z}^N_{\sigma_k} = N_k^{-1}\Big(N_{k-1}\overline{Z}^N_{\sigma_k^-} + \operatorname{sign}(\kappa_{\sigma_k} - \kappa_{\sigma_{k-1}}) \sum_{j=1}^{|N_k - N_{k-1}|} \delta_{Z'_j(\sigma_k)}\Big) = N_k^{-1}\sum_{i=1}^{N_k}\delta_{Z_i(\sigma_k)} \tag{8.17}$$

and replacing (8.13) by

$$\sigma_0 := s_0, \qquad \sigma_{k+1} := \sigma_k + (|\kappa'_{\sigma_k}|N^{1-a})^{-1} + N^{-1}, \quad k \ge 0. \tag{8.18}$$

Last equality in (8.17) results from some relabeling, and last term N^{-1} in (8.18) is added to make sure that $\sigma_{k+1} > \sigma_k$ even when $\kappa'_{\sigma_k} = 0$.

Looking at (8.17), we see that when $\kappa_{\sigma_k} < \kappa_{\sigma_{k-1}}$, the newcomers are removed from the support of $\overline{Z}^N_{\sigma_k^-}$ and when $\kappa_{\sigma_k} > \kappa_{\sigma_{k-1}}$, the newcomers are added to $\overline{Z}^N_{\sigma_k^-}$, while the total (unit) mass of $\overline{Z}^N_s$ is conserved.

9. ϵ-MAG for a k-fluid. Particle system

Let us denote $\overline{X}^N$ the process $\overline{Z}^N$ of previous section where $Z = \mathsf{X}^\epsilon$, as in Section 7.

We see that at each time s, $\overline{X}^N_s$ is the empirical measure of a cloud of $\lfloor \kappa_s N\rfloor$ k-mappings. Each of these k-mappings is a copy of X^ϵ. But these copies are not independent. Assuming that κ is an increasing function, during any small time interval $[s, s+h]$, a fraction $\kappa'_s\, h + o_{h\to 0}(h)$ of the particles branch: each of them gives birth to a new particle starting at the same place as its genitor and evolving in the future according to the kinematics of X^ϵ and independently of the other particles.
Remark that although the number $\lfloor \kappa_s N\rfloor$ of particles in the cloud increases with time, $\overline{X}^N$ is normalized so that its total mass remains constant: $\overline{X}^N_s(\mathbb{R}^{dk}) = 1$ for all s. As a consequence, the random fluctuation of the whole cloud decreases.

Theorem 8.1 and Proposition 7.11 tell us that $(\overline{X}^N)_{N\ge 1}$ obeys the large deviation principle

$$\operatorname{Proba}(\overline{X}^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\Big(-N \inf_{p\in\bullet} J^\kappa(p)\Big),$$

in $\Omega^{(k)}_{\mathrm{P}}$ with rate function given for any $p \in \Omega^{(k)}_{\mathrm{P}}$ by

$$\begin{aligned} J^\kappa(p) = \tfrac12 \kappa_{s_0} H(p_{s_0}|r^\epsilon_{s_0}) + \tfrac12 \kappa_{s_1} H(p_{s_1}|r^\epsilon_{s_1}) \\ + \epsilon^{-1}\int_{s_0}^{s_1} \tfrac12 \|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}^\epsilon_s\|^2_{p_s}\kappa_s\, ds + \epsilon \int_{s_0}^{s_1} I(p_s|r^\epsilon_s)\kappa_s\, ds. \end{aligned} \tag{9.1}$$

Note that, because of the s_1-term, the rate function J of Proposition 7.11 has not exactly the form of I at Theorem 8.1. To obtain the announced result, apply Theorem 8.1 with Proposition 7.9, and also with its time reversed analog which shares the same rate function since time reversal is one-one. Then take the half sum of these two expressions to arrive at (9.1), see [11, § 6] for details. It follows that an action functional attached to $(\overline{X}^N)_{N\geq 1}$ is

$$p \in \Omega_{\mathrm{P}}^{(k)} \mapsto \int_{s_0}^{s_1} \frac{1}{2}\|\dot{\mathsf{p}}_s - \dot{\mathsf{r}}_s^\epsilon\|_{p_s}^2 \kappa_s\, ds + \epsilon^2 \int_{s_0}^{s_1} I(p_s|r_s^\epsilon)\kappa_s\, ds.$$

Let us switch to time t by means of the Parameter setting 3.10: $\kappa_s = 2s$ and $s = e^{2t}$, and (5.5): $q_t := p_s = p_{e^{2t}}$, to obtain the t-action

$$q \in \Omega_{\mathrm{P}}^{(k)} \mapsto \int_{t_0}^{t_1} \frac{1}{2}\|\dot{\mathsf{q}}_t - \dot{\mathsf{m}}_t^\epsilon\|_{q_t}^2\, dt + \epsilon^2 \int_{t_0}^{t_1} 4e^{4t} I(q_t|m_t^\epsilon)\, dt, \tag{9.2}$$

whose first term is (5.10) as desired.

It remains to remove the rightmost term from action (9.2) to arrive at ϵ-MAG's action (5.10). This amounts to subtract the potential energy $\epsilon^2\mathcal{I}_t(q)$ where

$$\mathcal{I}_t(q) := 4e^{4t} I(q|m_t^\epsilon)$$

from the Lagrangian $(t, q, \dot{\mathsf{q}}) \mapsto \frac{1}{2}\|\dot{\mathsf{q}} - \dot{\mathsf{m}}_t^\epsilon\|_q^2 + \epsilon^2\mathcal{I}(t, q)$ to arrive at the desired Lagrangian

$$(t, q, \dot{\mathsf{q}}) \mapsto \frac{1}{2}\|\dot{\mathsf{q}} - \dot{\mathsf{m}}_t^\epsilon\|_q^2.$$

In terms of a Newton equation, this means that an additional force field

$$-\epsilon^2 \operatorname{grad}_q^{\mathrm{OW}} \mathcal{I}_t$$

is applied to the k-fluid that minimizes the action (9.2). Next result tells us that it is a *quantum* force field.

Theorem 9.3. *Any solution Ψ of the nonlinear Schrödinger equation*

$$\big(-i\hbar\partial_t - \hbar^2\Delta/2 - \hbar^2\mathcal{Q}(m_t^\epsilon|\mathrm{Leb})\big)\Psi + (4e^{4t}\epsilon^2 - \hbar^2)\mathcal{Q}(|\Psi|^2|m_t^\epsilon)) = 0,$$

is such that $q = |\Psi|^2$ solves the Newton equation

$$\ddot{\mathsf{q}}_t = -\epsilon^2 \operatorname{grad}_{q_t}^{\mathrm{OW}} \mathcal{I}_t.$$

The measure m_t^ϵ is defined at (5.7) and the quantum potentials $\mathcal{Q}(m|\mathrm{Leb})$ and $\mathcal{Q}(p|m)$ are defined at (9.18) and (9.11) respectively.

Proof. It is Theorem 9.29 applied with $m = m_t^\epsilon$ and $c = 4e^{4t}\epsilon^2/\hbar^2$. □

The remainder of this section is dedicated to the proof Theorem 9.29.

Schrödinger bridges. We recall some results from [35] and [12] about the Schrödinger problem that we already met at (7.8). This will help us to understand the force field $+\epsilon^2 \operatorname{grad}_q^{\mathrm{OW}} \mathcal{I}_t$, with a plus instead of a minus sign, which is easier to justify than its opposite.

A reversible path measure: R^m. Let us consider an abstract setting where the configuration space is $\mathbb{R}^n$, the measure

$$m(dx) := e^{-U(x)/\epsilon}\, dx$$

on $\mathbb{R}^n$ is the equilibrium measure of the Markov process with generator

$$A^m := (-\nabla U\cdot\nabla + \epsilon\Delta)/2,$$

where the scalar function $U : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $\epsilon > 0$. It is assumed that there is a unique path measure R^m that solves the martingale problem with generator A^m and initial measure m. Denoting $(X_t)_{t_0 \le t \le t_1}$ the canonical process, this is equivalent to

$$\begin{aligned} &dX_t = -\tfrac{1}{2}\nabla U(X_t)\,dt + \sqrt{\epsilon}\,dB_t^{R^m}, \quad t_0 \le t \le t_1, \quad R^m\text{-a.e.}, \\ &X_{t_0} \sim m, \end{aligned} \tag{9.4}$$

where B^{R^m} is an R^m-Brownian motion. Not only R^m is m-stationary, but also it is reversible.

The Schrödinger problem and its solution. We choose as a reference path measure

$$R := \exp\left(\epsilon^{-1}\int_{t_0}^{t_1} V_t(X_t)\,dt\right) R^m, \tag{9.5}$$

where V is some scalar potential. It is known that, in a generic situation, the unique solution of the *Schrödinger problem*, recall (7.8),

$$\inf_{Q: Q_{t_0}=q_{t_0}, Q_{t_1}=q_{t_1}} H(Q|R)$$

where Q is a path measure with prescribed initial and final marginals: q_{t_0} and q_{t_1}, writes as

$$Q = f_{t_0}(X_{t_0}) g_{t_1}(X_{t_1})\,R, \tag{9.6}$$

for some nonnegative measurable functions f_{t_0} and g_{t_1} on $\mathbb{R}^n$, see [24] and the references therein for an overview of the Schrödinger problem. This entropic minimizer is called the *Schrödinger bridge* between q_{t_0} and q_{t_1} with respect to R. It is also a Markov measure.

Entropic interpolation. As a definition, the entropic interpolation between q_{t_0} and q_{t_1} with respect to R, is the flow $(q_t := (X_t)_\# Q; t_0 \le t \le t_1)$ of time-marginals of the Schrödinger bridge Q.
Let us denote the average potential by

$$\mathcal{V}(p) := \int_{\mathbb{R}^n} V\,dp.$$

Theorem 9.7 ([12]). *Any entropic interpolation (q_t) with respect to R solves the Newton equation*

$$\ddot{\mathsf{q}}_t = -\operatorname{grad}^{\mathrm{OW}}_{q_t}(\mathcal{V} - \epsilon^2 I(\bullet|m))$$

in the Otto-Wasserstein manifold.

The proof of this theorem is postponed at page 46. It relies on Propositions 9.8, 9.12 and Lemma 9.16 below. We give some details of its proof because several elements of the proof will be utilized later.

Proposition 9.8. *For any t, $q_t = (X_t)_\# Q$ is absolutely continuous with respect to m and*

$$\rho_t := dq_t/dm = f_t g_t, \quad t_0 \le t \le t_1, \tag{9.9}$$

where for any $t_0 \le t \le t_1$ and $x \in \mathbb{R}^n$,

$$\begin{aligned} f_t(x) &:= E_{R^m}\left(f_{t_0}(X_{t_0}) \exp\left(\epsilon^{-1}\int_{t_0}^{t} V_s(X_s)\,ds\right) \mid X_t = x\right), \\ g_t(x) &:= E_{R^m}\left(\exp\left(\epsilon^{-1}\int_{t}^{t_1} V_s(X_s)\,ds\right) g_{t_1}(X_{t_1}) \mid X_t = x\right). \end{aligned} \tag{9.10}$$

Proof. By a general result of integration theory

$$\frac{dq_t}{dm}(x) = \frac{d(X_t)_\# Q}{d(X_t)_\# R^m}(x) = E_{R^m}\Big(\frac{dQ}{dR^m} \mid X_t = x\Big).$$

With (9.5) and (9.6), we see that for any $t_0 \le t \le t_1$,

$$\begin{aligned}
\frac{dq_t}{dm}(X_t) &= E_{R^m}\left(\frac{dQ}{dR}\frac{dR}{dR^m} \mid X_t\right)\\
&= E_{R^m}\Big(f_{t_0}(X_{t_0}) g_{t_1}(X_{t_1}) \exp\Big(\epsilon^{-1}\int_{t_0}^{t_1} V_t(X_t)\,dt\Big) \mid X_t\Big)\\
&= E_{R^m}\Big(f_{t_0}(X_{t_0}) \exp\Big(\epsilon^{-1}\int_{t_0}^{t} V_s(X_s)\,ds\Big)\\
&\qquad\qquad \times \exp\Big(\epsilon^{-1}\int_{t}^{t_1} V_s(X_s)\,ds\Big) g_{t_1}(X_{t_1}) \mid X_t\Big)\\
&= E_{R^m}\Big(f_{t_0}(X_{t_0}) \exp\Big(\epsilon^{-1}\int_{t_0}^{t} V_s(X_s)\,ds\Big) \mid X_t\Big)\\
&\qquad\qquad \times E_{R^m}\Big(\exp\Big(\epsilon^{-1}\int_{t}^{t_1} V_s(X_s)\,ds\Big) g_{t_1}(X_{t_1}) \mid X_t\Big)\\
&=: f_t g_t(X_t),
\end{aligned}$$

where we used the Markov property of R^m at last but one equality. □

The functions $f(t,x)$ and $g(t,x)$ are solutions of the parabolic equations

$$\begin{cases} (-\partial_t + A^m + V/\epsilon) f = 0, & t_0 < t \le t_1,\\ f(t_0) = f_{t_0}, & t = t_0, \end{cases} \qquad \begin{cases} (\partial_t + A^m + V/\epsilon) g = 0, & t_0 \le t < t_1,\\ g(t_1) = g_{t_1}, & t = t_1, \end{cases}$$

and (9.10) are their Feynman-Kac representations. The Born-like formula (9.9) was first established by Schrödinger in [31, 32] in the Brownian case and extended later by Zambrini in [35]. Introducing

$$\theta_t(x) := \epsilon \log \sqrt{g_t(x)/f_t(x)},$$

we see with (9.9) that

$$f = \sqrt{\rho} e^{-\theta/\epsilon}, \qquad g = \sqrt{\rho} e^{\theta/\epsilon}.$$

For any regular enough probability measure p, the quantum potential of p with respect to m is defined by

$$\mathcal{Q}(p|m) := -\epsilon^{-1}\frac{A^m\sqrt{\ell}}{\sqrt{\ell}} = -\nabla \log\sqrt{m}\cdot\nabla\log\sqrt{\ell} - \frac{\Delta\sqrt{\ell}}{2\sqrt{\ell}} \quad \text{where} \quad \ell := dp/dm. \tag{9.11}$$

Proposition 9.12. *Denoting $q_t(x) = dq_t/dx$, we have*

$$\begin{aligned}
&\partial_t q + \nabla\cdot(q\,\nabla\theta) = 0,\\
&\partial_t \theta + |\nabla\theta|^2/2 + V - \epsilon^2 \mathcal{Q}(q|m) = 0.
\end{aligned} \tag{9.13}$$

Proof. Beside θ, let us introduce the functions $\varphi := \epsilon \log f$, $\psi := \epsilon\log g$ and $\eta := (\varphi + \psi)/2$. We also see that $\theta = (\psi - \varphi)/2$. The above parabolic equations are equivalent to $\epsilon e^{-\varphi/\epsilon}(-\partial_t + A^m + V/\epsilon) e^{\varphi/\epsilon} = 0$ and $\epsilon e^{-\psi/\epsilon}(\partial_t + A^m + V/\epsilon)e^{\psi/\epsilon} = 0$, that is

$$\begin{aligned}
&(a) \qquad -\partial_t\varphi + A^m\varphi + |\nabla\varphi|^2/2 + V = 0,\\
&(b) \qquad \quad\ \partial_t\psi + A^m\psi + |\nabla\psi|^2/2 + V = 0.
\end{aligned}$$

Taking the half difference $[(b)-(a)]/2$ and half sum $[(a)+(b)]/2$ of these equations, we obtain

$$\begin{aligned} &(1) \quad \partial_t \eta + A^m \theta + \nabla\theta \cdot \nabla\eta = 0 \\ &(2) \quad \partial_t \theta + |\nabla\theta|^2/2 + V + A^m \eta + |\nabla\eta|^2/2 = 0. \end{aligned}$$

By (9.9): $\rho = q/m$ and $\eta = \frac{1}{2}\epsilon \log q - \epsilon \log \sqrt{m}$, so that (1) writes as :

$$\frac{\epsilon}{2}\frac{\partial_t q}{q} - \frac{\epsilon}{2}\frac{\partial_t m}{m} + \epsilon\nabla \log\sqrt{m}\cdot\nabla\theta + \frac{\epsilon}{2}\Delta\theta + \frac{\epsilon}{2}\frac{\nabla q}{q}\cdot\nabla\theta - \epsilon\nabla\log\sqrt{m}\cdot\nabla\theta = 0.$$

Multiplying by $2q\epsilon^{-1}$, this amounts to

$$0 = \partial_t q + q\Delta\theta + \nabla q\cdot\nabla\theta = \partial_t q + \nabla\cdot(q\nabla\theta),$$

which is the continuity equation in (9.13). The Hamilton-Jacobi equation in (9.13) is simply (2), once one notices that

$$\begin{aligned} A^m\eta + |\nabla\eta|^2/2 &= \nabla\log\sqrt{m}\cdot\nabla\eta + \epsilon\Delta\eta/2 + |\nabla\eta|^2/2 \\ &= \epsilon\nabla\log\sqrt{m}\cdot\nabla\log\sqrt{\rho} + \epsilon^2(\Delta\log\sqrt{\rho} + |\nabla\log\sqrt{\rho}|^2)/2 \\ &= \epsilon\nabla\log\sqrt{m}\cdot\frac{\nabla\sqrt{\rho}}{\sqrt{\rho}} + \epsilon^2\frac{\Delta\sqrt{\rho}}{2\sqrt{\rho}} = \epsilon\frac{A^m\sqrt{\rho}}{\sqrt{\rho}} = -\epsilon^2\mathcal{Q}(q|m), \end{aligned}$$

where we used $\eta = \epsilon\log\sqrt{\rho}$ at second equality. □

Lemma 9.14. *The Fisher information of p with respect to m satisfies*

$$I(p|m) = \int_{\mathbb{R}^n} \mathcal{Q}(p|m)\, dp,$$

meaning that it is the average of the quantum potential, and

$$I(p|m) = \frac{1}{2}\int_{\mathbb{R}^n} |\nabla\sqrt{\ell}|^2\, dm = \frac{1}{2}\int_{\mathbb{R}^n} |\nabla\log\sqrt{\ell}|^2\, dp \quad \textit{where} \quad \ell := dp/dm.$$

Proof. Taking $I(p|m) := \int_{\mathbb{R}^n}\mathcal{Q}(p|m)\,dp$ as a definition, $I(p|m) = \frac{1}{2}\int_{\mathbb{R}^n}|\nabla\sqrt{\ell}|^2\,dm$ follows from the integration by parts formula

$$\epsilon\int_{\mathbb{R}^n}\nabla u\cdot\nabla v\,dm = \int_{\mathbb{R}^n}\Gamma^m(u,v)\,dm = -2\int_{\mathbb{R}^n} uA^m v\,dm \tag{9.15}$$

which holds because R^m is m-reversible. The carré du champ of its generator A^m is denoted by $\Gamma^m(uv) := A^m(uv) - uA^m v - vA^m u$. □

Lemma 9.16 ([34])**.** *The Otto-Wasserstein gradients of $\mathcal{V}$ and $I(\bullet|m)$ are*

$$(a) \quad \mathrm{grad}_p^{\mathrm{OW}}\,\mathcal{V} = \nabla V \qquad \textit{and} \qquad (b) \quad \mathrm{grad}_p^{\mathrm{OW}}\, I(\bullet|m) = \nabla\mathcal{Q}(p|m).$$

Proof. For any regular path (p_t),

$$\frac{d}{dt}\mathcal{F}(p_t) = (\mathrm{grad}_{p_t}^{\mathrm{OW}}\,\mathcal{F}, \dot{\mathsf{p}}_t)_{p_t}^{\mathrm{OW}} = \int_{\mathbb{R}^n}\mathrm{grad}_{p_t}^{\mathrm{OW}}\,\mathcal{F}\cdot\dot{\mathsf{p}}_t\,dp_t \tag{9.17}$$

where $(\nabla u, \nabla v)_p^{\mathrm{OW}} = \int_{\mathbb{R}^n}\nabla u\cdot\nabla v\,dp$ is the inner product of the tangent space at p of the Otto-Wasserstein manifold.

(a) Taking $\mathcal{F}=\mathcal{V}$, we see that

$$\frac{d}{dt}\mathcal{V}(p_t)=\int_{\mathbb{R}^n} V\,\partial_t p_t\,d\mathrm{Leb}=-\int_{\mathbb{R}^n} V\,\nabla\cdot(p_t\dot{\mathsf{p}}_t)\,d\mathrm{Leb}=\int_{\mathbb{R}^n}\nabla V\cdot\dot{\mathsf{p}}_t\,dp_t.$$

We used the continuity equation $\partial_t p+\nabla\cdot(p\dot{\mathsf{p}})=0$ at second identity and a standard integration by parts at last identity. Comparing with (9.17) leads to the announced result.

(b) For any small perturbation h with a small gradient, we have

$$\begin{aligned}|\nabla\sqrt{\ell+h}|^2&=|\nabla(\sqrt{\ell}+h/(2\sqrt{\ell})+o(h))|^2=|\nabla\sqrt{\ell}|^2+\nabla\sqrt{\ell}\cdot\nabla(h/\sqrt{\ell})+o(h,\nabla h)\\&=|\nabla\sqrt{\ell}|^2+\nabla\log\sqrt{\ell}\cdot\nabla h-|\nabla\log\sqrt{\ell}|^2h+o(h,\nabla h).\end{aligned}$$

Hence, denoting $\ell_t=dp_t/dm$ and taking $\mathcal{F}=I(\bullet|m)$ in (9.17), we see that

$$\begin{aligned}\frac{d}{dt}I(p_t|m)&=\frac12\int_{\mathbb{R}^n}\partial_t|\nabla\sqrt{\ell_t}|^2\,dm=\frac12\int_{\mathbb{R}^n}(\nabla\log\sqrt{\ell_t}\cdot\nabla\partial_t\ell_t-|\nabla\log\sqrt{\ell_t}|^2\partial_t\ell_t)\,dm\\&\overset{(i)}{=}-\epsilon^{-1}\int_{\mathbb{R}^n}(A^m\log\sqrt{\ell_t}+\epsilon|\nabla\log\sqrt{\ell_t}|^2/2)\,\partial_t p_t\,d\mathrm{Leb}\\&\overset{(ii)}{=}\int_{\mathbb{R}^n}\mathcal{Q}(p_t|m)\,\partial_t p_t\,d\mathrm{Leb}\overset{(iii)}{=}\int_{\mathbb{R}^n}\nabla\mathcal{Q}(p_t|m)\cdot\dot{\mathsf{p}}_t\,dp_t.\end{aligned}$$

At $\overset{(i)}{=}$, we used $\partial_t\ell_t\,dm=\partial_t p_t\,d\mathrm{Leb}$ and the integration by parts formula (9.15). Equality $\overset{(ii)}{=}$ follows from $\Delta u/u=\Delta\log u+|\nabla\log u|^2$. At $\overset{(iii)}{=}$, we used the continuity equation $\partial_t p+\nabla\cdot(p\dot{\mathsf{p}})=0$ and a standard integration by parts. We conclude comparing with (9.17).

This completes the proof of the lemma. □

Proof of Theorem 9.7. The first identity in (9.13) is a continuity equation: $\nabla\theta$ is the current velocity of (q_t), that is $\dot{\mathsf{q}}_t=\nabla\theta_t$. The second identity in (9.13) is a Hamilton-Jacobi equation. Taking its gradient leads us to Newton's equation: $(\partial_t+\nabla\theta_t\cdot\nabla)\nabla\theta_t+V-\epsilon^2\nabla\mathcal{Q}(q_t|m)=0$. But $(\partial_t+\nabla\theta_t\cdot\nabla)\nabla\theta_t=\nabla^{\mathrm{OW}}_{\nabla\theta_t}\nabla\theta_t=\nabla^{\mathrm{OW}}_{\dot{\mathsf{q}}_t}\dot{\mathsf{q}}_t=:\ddot{\mathsf{q}}_t$. Hence, $\ddot{\mathsf{q}}_t=-\nabla V+\epsilon^2\nabla\mathcal{Q}(q_t|m)$. We conclude with Lemma 9.16. □

From m to Leb. In a moment, we shall establish an analogy between the thermal evolution of the entropic interpolation and some quantum evolution. As Schrödinger equation refers to densities with respect to Lebesgue measure, it is worth switching from R^m to the new reversible path measure R^{Leb} which is defined as the unique solution of the martingale problem (9.4) with $U=0$, that is

$$\begin{aligned}&dX_t=\sqrt{\epsilon}\,dB_t^{R^{\mathrm{Leb}}},\quad t_0\le t\le t_1,\quad R^{\mathrm{Leb}}\text{-a.e.},\\&X_{t_0}\sim\mathrm{Leb},\end{aligned}$$

where $B^{R^{\mathrm{Leb}}}$ is an R^{Leb}-Brownian motion. Its Markov generator is

$$A^{\mathrm{Leb}}=\epsilon\Delta/2.$$

The corresponding quantum potential is (9.11) with $m=\mathrm{Leb}$, that is (6.12):

$$\mathcal{Q}(p|\mathrm{Leb}):=-\epsilon^{-1}\frac{A^{\mathrm{Leb}}\sqrt{p}}{\sqrt{p}}\quad\text{where}\quad p:=dp/\mathrm{Leb},$$

Note that

$$\mathcal{Q}(p|\mathrm{Leb}) = -\frac{\Delta\sqrt{p}}{2\sqrt{p}} = -\frac{1}{2}(\Delta \log \sqrt{p} + |\nabla \log \sqrt{p}|^2). \tag{9.18}$$

Lemma 9.19. *Let $m = \exp(-U/\epsilon)$ denote the density of the equilibrium measure m with respect to Lebesgue measure. Then,*

$$\epsilon^2 \mathcal{Q}(m|\mathrm{Leb}) = \epsilon \Delta U/4 - |\nabla U|^2/8, \tag{9.20}$$

$$\mathcal{Q}(q|\mathrm{Leb}) = \mathcal{Q}(q|m) + \mathcal{Q}(m|\mathrm{Leb}). \tag{9.21}$$

Proof. The first identity is a direct calculation with $\log\sqrt{m} = -U/(2\epsilon)$. For the second one, use: $\Delta(uv) = u\Delta v + v\Delta u + 2\nabla u \cdot \nabla v$ and $q = \rho m$ to obtain

$$\begin{aligned} -2\mathcal{Q}(q|\mathrm{Leb}) &= \frac{\Delta\sqrt{q}}{\sqrt{q}} = \frac{\Delta\sqrt{\rho m}}{\sqrt{\rho m}} = \frac{\sqrt{\rho}\Delta\sqrt{m} + \sqrt{m}\Delta\sqrt{\rho} + 2\nabla\sqrt{\rho}\cdot\nabla\sqrt{m}}{\sqrt{\rho}\sqrt{m}} \\ &= \frac{\Delta\sqrt{m}}{\sqrt{m}} + \frac{\Delta\sqrt{\rho}}{\sqrt{\rho}} + 2\nabla \log\sqrt{m} \cdot \frac{\nabla\sqrt{\rho}}{\sqrt{\rho}} = -2\mathcal{Q}(m|\mathrm{Leb}) - 2\mathcal{Q}(q|m), \end{aligned}$$

as announced. □

Corollary 9.22. *Denoting $q_t(x) = dq_t/dx$, we have*

$$\begin{aligned} &\partial_t q + \nabla \cdot (q\,\nabla\theta) = 0, \\ &\partial_t \theta + |\nabla\theta|^2/2 + V - \epsilon^2[\mathcal{Q}(q|\mathrm{Leb}) - \mathcal{Q}(m|\mathrm{Leb})] = 0. \end{aligned} \tag{9.23}$$

Proof. It is a direct consequence of (9.13) and (9.21). □

This easy analytical proof partly hides what is really at stake with the potential $\mathcal{Q}(m|\mathrm{Leb})$. This is why we present another proof of (9.23) where the role of this potential is more explicit.

Lemma 9.24. *The Radon-Nikodym density of R^m with respect to R^{Leb} is*

$$\frac{dR^m}{dR^{\mathrm{Leb}}} = \sqrt{m}(X_{t_0}) \exp\left(\int_{t_0}^{t_1} \epsilon\mathcal{Q}(m|\mathrm{Leb})(X_t)\,dt\right)\sqrt{m}(X_{t_1}).$$

Proof. This follows from Girsanov's formula which is the first equality below

$$\begin{aligned} \frac{dR^m}{dR^{\mathrm{Leb}}} &= \frac{dR^m_{t_0}}{dR^{\mathrm{Leb}}_{t_0}}(X_{t_0}) \exp\left(\int_{t_0}^{t_1} -\frac{\nabla U}{2\epsilon}(X_t)\cdot dX_t - \frac{1}{2}\int_{t_0}^{t_1} \left|\frac{\nabla U}{2\epsilon}\right|^2 (X_t)\,\epsilon dt\right) \\ &= \frac{dR^m_{t_0}}{dR^{\mathrm{Leb}}_{t_0}}(X_{t_0}) \exp\left(\frac{U(X_{t_0})}{2\epsilon} - \frac{U(X_{t_1})}{2\epsilon} + \int_{t_0}^{t_1} (\frac{\Delta U}{4} - \frac{|\nabla U|^2}{8\epsilon})(X_t)\,dt\right) \\ &= \exp\Big(-\frac{U(X_{t_0}) + U(X_{t_1})}{2\epsilon}\Big) \exp\left(\int_{t_0}^{t_1} (\frac{\Delta U}{4} - \frac{|\nabla U|^2}{8\epsilon})(X_t)\,dt\right) \\ &= \sqrt{m}(X_{t_0})\sqrt{m}(X_{t_1}) \exp\left(\int_{t_0}^{t_1} \epsilon\mathcal{Q}(m|\mathrm{Leb})(X_t)\,dt\right). \end{aligned}$$

These identities hold R^{Leb}-almost everywhere. At second equality, we used Itô's formula: $dU(X_t) = \nabla U(X_t)\cdot dX_t + \epsilon\Delta U(X_t)/2\ dt$, R^{Leb}-a.e., and at third equality we used $\dfrac{dR^m_{t_0}}{dR^{\mathrm{Leb}}_{t_0}}(X_{t_0}) = \dfrac{dm}{d\mathrm{Leb}}(X_{t_0}) = \exp(-U(X_{t_0})/\epsilon)$. □

This result is a probabilistic version of the *ground state transform*.

Another proof of (9.23). With Lemma 9.24, our reference path measure rewrites as

$$\begin{aligned} R &\overset{(9.5)}{=} \exp\left(\epsilon^{-1}\int_{t_0}^{t_1} V_t(X_t)\,dt\right) R^m \\ &= \sqrt{m}(X_{t_0}) \exp\left(\epsilon^{-1}\int_{t_0}^{t_1} [V_t + \epsilon^2 \mathcal{Q}(m|\mathrm{Leb})](X_t)\,dt\right) \sqrt{m}(X_{t_1})\, R^{\mathrm{Leb}}, \end{aligned}$$

and the Schrödinger bridge rewrites as

$$\begin{aligned} Q &\overset{(9.6)}{=} f_{t_0}(X_{t_0}) g_{t_1}(X_{t_1})\, R \\ &= \tilde f_{t_0}(X_{t_0}) \exp\left(\epsilon^{-1}\int_{t_0}^{t_1} \widetilde V_t(X_t)\,dt\right) \tilde g_{t_1}(X_{t_1})\, R^{\mathrm{Leb}} \end{aligned}$$

where

$$\tilde f := \sqrt{m} f, \qquad \tilde g := \sqrt{m} g \quad \text{and} \quad \widetilde V_t = V_t + \epsilon^2 \mathcal{Q}(m|\mathrm{Leb}).$$

Applying Proposition 9.8 with Leb instead of m and $\widetilde V$ instead of V, we see that

$$q_t := dq_t/d\mathrm{Leb} = \tilde f_t \tilde g_t, \quad t_0 \le t \le t_1, \tag{9.25}$$

where for any $t_0 \le t \le t_1$ and $x \in \mathbb{R}^n$,

$$\begin{aligned} \tilde f_t(x) &:= E_{R^{\mathrm{Leb}}}\left(\tilde f_{t_0}(X_{t_0}) \exp\left(\epsilon^{-1}\int_{t_0}^{t} \widetilde V_s(X_s)\,ds\right) \mid X_t = x\right), \\ \tilde g_t(x) &:= E_{R^{\mathrm{Leb}}}\left(\exp\left(\epsilon^{-1}\int_{t}^{t_1} \widetilde V_s(X_s)\,ds\right) \tilde g_{t_1}(X_{t_1}) \mid X_t = x\right). \end{aligned} \tag{9.26}$$

The functions $\tilde f(t,x)$ and $\tilde g(t,x)$ are solutions of the parabolic equations

$$\begin{cases} (-\partial_t + \epsilon\Delta/2 + [\epsilon^{-1} V_t + \epsilon \mathcal{Q}(m|\mathrm{Leb})])\tilde f &= 0 \\ (\partial_t + \epsilon\Delta/2 + [\epsilon^{-1} V_t + \epsilon \mathcal{Q}(m|\mathrm{Leb})])\tilde g &= 0. \end{cases} \tag{9.27}$$

We also see that $\tilde\theta := \epsilon \log\sqrt{\tilde g/\tilde f} = \epsilon\log\sqrt{g/f} =: \theta$ and with (9.25) that

$$\tilde f = \sqrt{q} e^{-\theta/\epsilon}, \qquad \tilde g = \sqrt{q} e^{\theta/\epsilon}. \tag{9.28}$$

Replacing $\mathcal{Q}(q|m)$ by $\mathcal{Q}(q|\mathrm{Leb})$, V by $\widetilde V$ and keeping $\tilde\theta = \theta$, we see that the analog of (9.13) is (9.23). □

An analogy with a quantum evolution. We are now ready to present an informal proof of

Theorem 9.29. *Let c be a real number. Any solution Ψ of the nonlinear Schrödinger equation*

$$\big(-i\hbar\partial_t - \hbar^2\Delta/2 - \hbar^2\mathcal{Q}(m|\mathrm{Leb})\big)\Psi + (c-1)\hbar^2\mathcal{Q}(|\Psi|^2|m)) = 0,$$

is such that $q = |\Psi|^2$ solves the Newton equation

$$\ddot{\mathsf{q}}_t = -\hbar^2 c\, \mathrm{grad}^{\mathrm{OW}}_{q_t} I(\bullet|m).$$

Informal proof. In [31, 32, 35], the standard case $U = 0$, i.e. $m = \mathrm{Leb}$, is considered. In this case (9.27) writes as

$$\begin{cases} -\epsilon\partial_t \tilde f + \epsilon^2\Delta\tilde f/2 + V\tilde f &= 0 \\ +\epsilon\partial_t \tilde g + \epsilon^2\Delta\tilde g/2 + V\tilde g &= 0. \end{cases}$$

Applying the correspondences

$$\epsilon \leftrightarrow i\hbar, \qquad (\tilde{f}, \tilde{g}) \leftrightarrow (\Psi, \overline{\Psi}) \tag{9.30}$$

where Ψ is a complex function and $\overline{\Psi}$ is its complex conjugate, we obtain

$$\big(- i\hbar\partial_t - \hbar^2\Delta/2 + V\big)\Psi = 0,$$

the second equation: $+i\hbar\partial_t\overline{\Psi} - \hbar^2\Delta\overline{\Psi}/2 + V\overline{\Psi} = 0$, being its complex conjugate. This is Schrödinger equation for the wave function Ψ, while (9.25) becomes Born's formula of quantum mechanics:

$$q = \Psi\overline{\Psi} = |\Psi|^2.$$

Formula (9.28) becomes the polar decomposition $\Psi = \sqrt{q}e^{i\theta/\hbar}$ and (9.23) with $m = \mathrm{Leb}$ becomes the couple of Madelung equations [27]

$$\begin{aligned} &\partial_t q + \nabla\cdot(q\,\nabla\theta) = 0, \\ &\partial_t\theta + |\nabla\theta|^2/2 + V + \hbar^2\mathcal{Q}(q|\mathrm{Leb}) = 0, \end{aligned} \tag{9.31}$$

leading, as in the proof of Theorem 9.7, to Newton's equation

$$\ddot{\mathsf{q}}_t = -\operatorname{grad}^{\mathrm{OW}}_{q_t}\Big(\mathcal{V} + \hbar^2 I(\bullet|\mathrm{Leb})\Big).$$

This result was first discovered by von Renesse in the article [34],
Choosing $V(q) = \hbar^2(c\mathcal{Q}(q|m) - \mathcal{Q}(q|\mathrm{Leb})) \overset{(9.21)}{=} -\hbar^2\mathcal{Q}(m|\mathrm{Leb}) + (c-1)\hbar^2\mathcal{Q}(q|m)$ completes the "proof" of the theorem. □

The correspondences (9.30) are far from being mathematically rigorous. They are a quantization rule. Although Schrödinger did not write them explicitly in [31, 32], it is clearly what he had in mind.

Appendix A. What remains to be done

To derive a tractable least action principle for a fluid subject to ϵ-MAG, we need to

(1) give a more explicit formula for the least action principle (5.15);
(2) as (5.15) depends on k, look at its limit as k tends to infinity when the limiting source measure $\lim_{k\to\infty}\lambda^{(k)} = \mathbf{1}_D\,\mathrm{Leb}$ is the normalized volume of some set D, see (2.5).

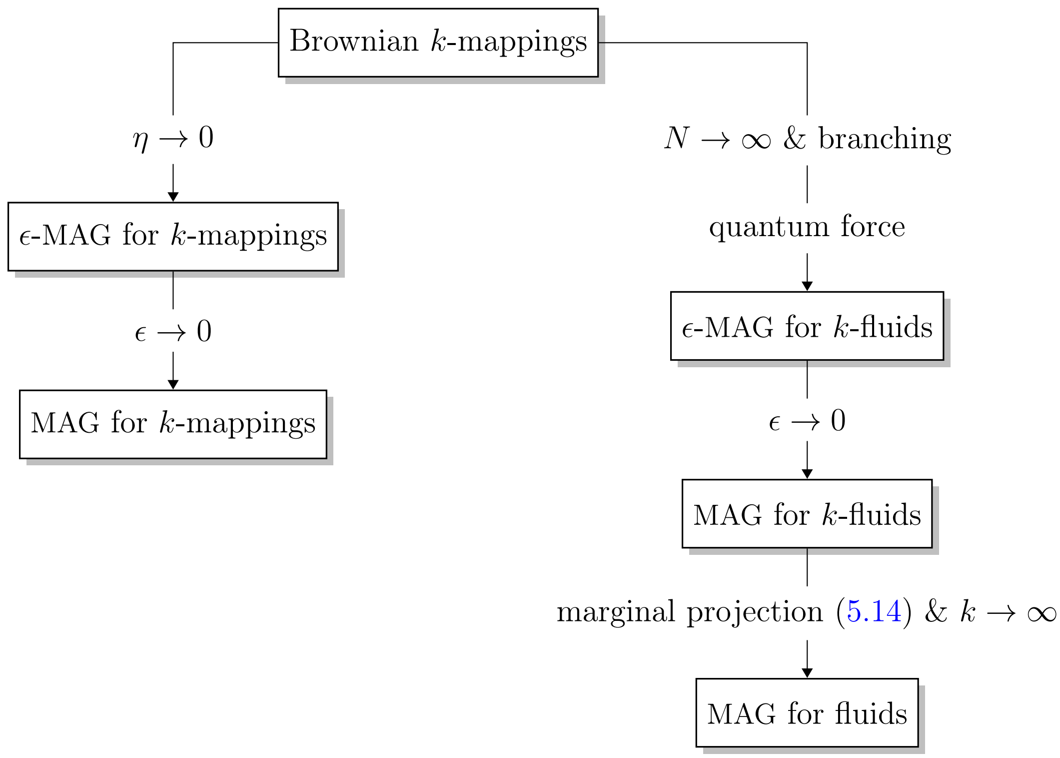
Of course,

(3) let ϵ tend to zero.

Finally, with Remark 2.9-(i) in mind,

(4) translate these results from $\mathbb{R}^d$ to the torus $\mathbb{T}^d$.

This list only pertains to mathematics. Translating these mathematical results – or more likely some modified versions of them – into meaningful physics is an open challenge.



The left-hand side of this flow chart corresponds to the Ambrosio-Baradat-Brenier particle system (3.7)-(3.12). Its right-hand side corresponds to the approach of the present article. We stopped at ϵ-MAG for k-fluids. The two lowest arrows on the right remain to be explored.

Appendix B. Least action principle

This short note is about basic calculus of variations with no emphasis on rigorous derivations. A good reference is the textbook by Gelfand and Fomin [19] which doesn't seek mathematical rigor either. The action functional is defined by

$$\mathcal{A}(\omega) = \int_{t_0}^{t_1} L(t, \omega_t, \dot{\omega}_t)\, dt$$

where $\omega = (\omega_t)_{t_0 \le t \le t_1}$ is a regular $\mathbb{R}^n$-valued path with time derivative $\dot{\omega}$. The function

$$L : (t, q, v) \in [t_0, t_1] \times \mathbb{R}^d \times \mathbb{R}^d \mapsto L(t, q, v) \in \mathbb{R}$$

is called the Lagrangian. It is assumed to be sufficiently differentiable.
The first variation of $\mathcal{A}$ at ω in the direction η is

$$\mathrm{d}\mathcal{A}_\omega(\eta) := \lim_{h \to 0} h^{-1}[\mathcal{A}(\omega + h\eta) - \mathcal{A}(\omega)]$$

and, as a definition, a critical path ω of $\mathcal{A}$ satisfies $\mathrm{d}\mathcal{A}_\omega(\eta) = 0$ for all η.

Theorem B.1. *Any critical path ω of $\mathcal{A}$ solves the Euler-Lagrange equation*

$$\partial_q L(t, \omega_t, \dot{\omega}_t) - \frac{d}{dt}\{\partial_v L(t, \omega_t, \dot{\omega}_t)\} = 0, \quad t_0 \le t \le t_1.$$

Proof. The Taylor expansion of L leads us to

$$\begin{aligned}
\mathrm{d}\mathcal{A}_\omega(\eta) &= \lim_{h\to 0} h^{-1}\int_{t_0}^{t_1}[L(t,\omega_t+h\eta_t,\dot\omega_t+h\dot\eta_t)-L(t,\omega_t,\dot\omega_t)]\,dt\\
&= \int_{t_0}^{t_1}[\partial_q L(t,\omega_t,\dot\omega_t)\cdot\eta_t+\partial_v L(t,\omega_t,\dot\omega_t)\cdot\dot\eta_t]\,dt\\
&= \partial_v L(t_1,\omega_{t_1},\dot\omega_{t_1})\cdot\eta(t_1)-\partial_v L(t_0,\omega_{t_0},\dot\omega_{t_0})\cdot\eta(t_0)\\
&\qquad +\int_{t_0}^{t_1}\left[\partial_q L(t,\omega_t,\dot\omega_t)-\frac{d}{dt}\{\partial_v L(t,\omega_t,\dot\omega_t)\}\right]\cdot\eta_t\,dt
\end{aligned}$$

where last identity is obtained by integrating by parts. Since η is arbitrary, it is necessary that along any critical path of $\mathcal{A}$ the integrand $\partial_q L(t)-\frac{d}{dt}\{\partial_v L(t)\}$ vanishes. □

The least action problem is

$$\inf\{\mathcal{A}(\omega);\omega:\omega_{t_0}=a,\omega_{t_1}=b\}$$

where the endpoint positions a and b are prescribed. Under these constraints, the variation η must verify $\eta_{t_0}=\eta_{t_1}=0$, so that

$$\mathrm{d}\mathcal{A}_\omega(\eta)=\int_{t_0}^{t_1}\left[\partial_q L(t,\omega_t,\dot\omega_t)-\frac{d}{dt}\{\partial_v L(t,\omega_t,\dot\omega_t)\}\right]\cdot\eta_t\,dt.$$

Of course, any minimizer is critical, so that any solution of the least action problem solves the Euler-Lagrange equation.
An important example in classical mechanics is given by the Lagrangian

$$L(t,q,v)=m|v|^2/2-U(t,q),$$

because the corresponding Euler-Lagrange equation is the standard equation of motion

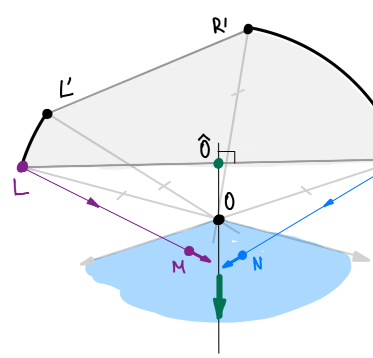
$$m\ddot\omega_t=-\nabla U_t(\omega_t).$$

A least action *principle* is a law of nature which stipulates that the trajectory of the physical system solves some least action *problem*. As is customary, although it is slightly incorrect, we keep saying that this system solves a least action principle.

Appendix C. Concentration of matter

Let us explain informally, using a 2D analogy, why the expression (2.21) at Definition 2.20 of MAG's action functional is reasonable.

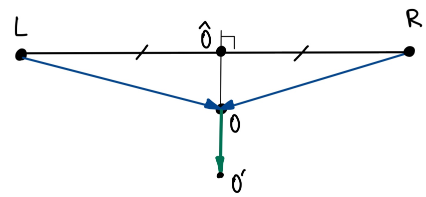
Definition C.1 (proj^o_S)**.** *Let* $\mathrm{Proj}_S(\mathsf{y})$ *be the set of all the closest points to* y *in* S *and* $\mathrm{cl\,cv}(\mathrm{Proj}_S(\mathsf{y}))$ *be its closed convex hull. We define* $\mathrm{proj}^o_S(\mathsf{y})$ *as the (unique) element in* $\mathrm{cl\,cv}(\mathrm{Proj}_S(\mathsf{y}))$ *with minimal norm.*



In the above figure, the point O is the center of the circle passing through the points L, L', R, R', and the set S is the union of the arcs: $\mathrm{arc}_{LL'}$ and $\mathrm{arc}_{R'R}$.

We see that $\mathrm{proj}^o_S(M) = \mathrm{proj}_S(M) = L$, $\mathrm{proj}^o(N) = \mathrm{proj}_S(N) = R$. But $\mathrm{proj}^o(O) = \widehat{O}$ because $\mathrm{Proj}_S(O) = S$ and $\mathrm{cl}\,\mathrm{cv}(S)$ is the gray area.

In order to minimize the kinetic action (2.19) the points M and N must evolve respectively in the direction of the purple and blue arrows.
The situation for O is a little bit different because its projection on S is the set $\mathrm{Proj}_S(O) = S$. This implies that O must evolve in the direction of the blue cone. Suppose it moves on the left of the segment bisector of the points L and R which is the line extending $\widehat{O}O$. Then as happens to M, it is instantaneously pushed back onto the right towards the bisector. It can neither move on the right, for the same symmetric reason. No choice, then: the only option is to move in the direction of the green arrow. It stays on the bisector.



By symmetry, the effective velocity $\overrightarrow{OO'}$ is equal to the average of the velocities generated by L on the left and R on the right. That is something proportional to

$$\overrightarrow{OO'} = (\overrightarrow{LO} + \overrightarrow{RO})/2 = \overrightarrow{\widehat{O}O}.$$

On the other hand, we have $\widehat{O} = \mathrm{proj}^o_S(O)$, and it is proved at Proposition C.6 below that, going back to the notation of Definition 2.20, we have $\widehat{\Phi}(0) = 0 - \widehat{\mathrm{proj}}_S(0)$. $\square$

A convex analytic result. Denoting $\Psi := -\Phi$, the action (2.19) writes as

$$\int_{t_0}^{t_1} \frac{1}{2} \|\dot{\mathsf{y}}_t + \mathrm{grad}_H \Psi(\mathsf{y}_t)\|^2_H \, dt, \tag{C.2}$$

and any solution of the gradient flow equation

$$\dot{\mathsf{y}}_t = -\,\mathrm{grad}_H \Psi(\mathsf{y}_t), \tag{C.3}$$

is a global minimizer of (C.2). For any $\mathsf{y} \in H$,

$$\Psi(\mathsf{y}) = -\inf_{\mathsf{x}\in S} \|\mathsf{y} - \mathsf{x}\|^2_H/2 = \iota^*_S(\mathsf{y}) - \|\mathsf{y}\|^2_H/2 - r^2/2,$$

where $\iota^*_S(\mathsf{y}) := \sup_{\mathsf{x}\in S} \mathsf{x}\cdot\mathsf{y}$, is the convex conjugate of $\iota_S(\mathsf{x}) = \begin{cases} 0, & \text{if } \mathsf{x} \in S \\ +\infty, & \text{otherwise} \end{cases}$, and we used (2.11). Since ι^*_S is convex, Ψ is a α-convex function (with $\alpha = -1$). Following [3, Cor. 1.4.2, Thm. 2.4.15], equation (C.3) admits as a natural extension

$$\dot{\mathsf{y}}_t = -\partial^o \Psi(\mathsf{y}_t), \quad \text{for a.e. } t, \tag{C.4}$$

Definition C.5 ($\partial^o\Psi(\mathsf{y})$)**.** *We denote by $\partial^o\Psi(\mathsf{y})$ the (unique) element with minimal norm of the local subdifferential $\partial\Psi(\mathsf{y})$ defined by*

$$\zeta \in \partial\Psi(\mathsf{y}) \iff \liminf_{\mathsf{y}'\to\mathsf{y}} \frac{\Psi(\mathsf{y}') - [\Psi(\mathsf{y}) + \langle\zeta, \mathsf{y}' - \mathsf{y}\rangle_H]}{\|\mathsf{y}' - \mathsf{y}\|_H} \geq 0, \qquad \zeta \in H.$$

Proposition C.6. *For any $\mathsf{y} \in H$, the element of $\partial\Psi(\mathsf{y})$ with minimal norm is*

$$\partial^o\Psi(\mathsf{y}) = \mathrm{proj}^o_S(\mathsf{y}) - \mathsf{y}.$$

Proof. The definition of $\partial\Psi(\mathsf{y})$ clearly implies that it is a convex set, hence the uniqueness of its element with minimal norm if it is nonempty. But, because Ψ is the sum of the differentiable function $f = -\|\bullet\|_H^2/2$ and the convex function ι_S^*, its local subdifferentials are nonempty. More precisely,

$$\partial\Psi(\mathsf{y}) = f'(\mathsf{y}) + \partial\iota_S^*(\mathsf{y}) = -\mathsf{y} + \partial\iota_S^*(\mathsf{y})$$

where

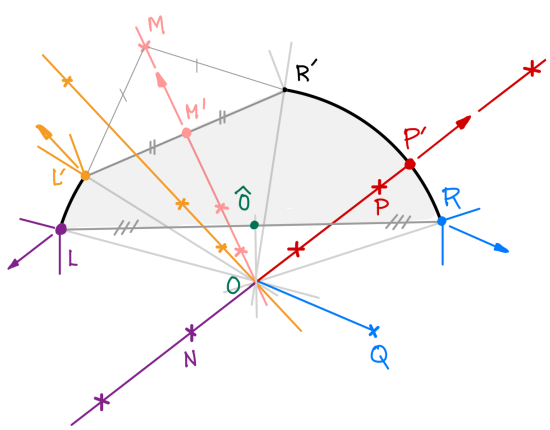
$$\partial\iota_S^*(\mathsf{y}) = \{\zeta \in H : \iota_S^*(\mathsf{y}') - \iota_S^*(\mathsf{y}) - \langle\zeta, \mathsf{y}' - \mathsf{y}\rangle \geq 0,\ \forall \mathsf{y}' \in H\} \neq \emptyset$$

is the (global) subdifferential of ι_S^* at y. The convex conjugate of ι_S^* is the convex indicator $\iota_{\mathrm{cl\,cv}\,S}$ of the closed convex hull $\mathrm{cl\,cv}(S)$ of S. Consequently, $\zeta \in \partial\iota_S^*(\mathsf{y})$ is equivalent to $\mathsf{y} \in \partial\iota_{\mathrm{cl\,cv}(S)}(\zeta)$, that is: $\zeta \in \mathrm{cl\,cv}(S)$ and y is in the cone of outer normals of $\mathrm{cl\,cv}(S)$ at ζ. In the general case, this property does not imply that ζ is the orthogonal projection $\mathrm{proj}_S(\mathsf{y})$ of y on S. But in the present setting where S is a subset of a sphere centered at zero, we obtain

$$\partial\iota_S^*(\mathsf{y}) = \mathrm{cl\,cv}(\mathrm{Proj}_S(\mathsf{y})).$$

In the regular case where $\mathrm{Proj}_S(\mathsf{y})$ is reduced to the single point $\mathrm{proj}_S(\mathsf{y})$, we clearly obtain: $\partial^o\Psi(\mathsf{y}) = \mathrm{proj}_S(\mathsf{y}) - \mathsf{y}$. In the general case, we have to replace $\mathrm{proj}_S(\mathsf{y})$ by $\mathrm{proj}_S^o(\mathsf{y})$. Again, this holds because S is included in a sphere centered at zero. □

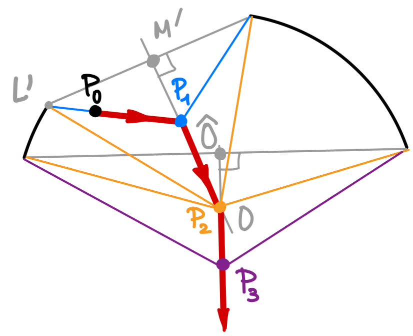
An analogical illustration for the concentration of matter. In the figure below illustrating a 2D analogy where the black arcs are S and the gray area is $\mathrm{cl\,cv}(S)$, we see that $\partial\iota_S^*(P) = \{\mathrm{proj}_S(P)\}$, $\partial\iota_S^*(N) = \{L\}$, $\partial\iota_S^*(M) = [L', R']$. These identities are unchanged when replacing P by any red cross point on the picture, N by any purple cross point, and M by any pink cross point. The orange crosses are mapped to L' and $\partial\iota_S^*(O) = \mathrm{cl\,cv}(S)$.
We also see that $\mathrm{proj}_S^o(P) = \mathrm{proj}_S(P)$, $\mathrm{proj}_S^o(Q) = \mathrm{proj}_S(Q)$, but $\mathrm{proj}_S^o(M) = M'$ and $\mathrm{proj}_S^o(O) = \widehat{O}$.



Keeping our 2D analogy, the figure below illustrates the motion of a particle moving according to the modified gradient flow (C.4). Shocks occur. Successively: $\mathrm{proj}_S^o(P_0) = A$, $\mathrm{proj}_S^o(P_1) = M'$, $\mathrm{proj}_S^o(P_2 = O) = \mathrm{proj}_S^o(P_3) = \widehat{O}$, and we see that:

- $s \mapsto \mathrm{dist}(\mathsf{y}_t, S)$ increases continuously
- with a discontinuity of its derivative when passing through P_1 and P_2.
- Furthermore, we observe that at each shock the modulus of the force decreases. Indeed, (2.17) tells us that it is proportional to $|\mathsf{y}_t - \mathrm{proj}_S^o(\mathsf{y}_t)|$, and this quantity

jumps from $L'P_1$ to $M'P_1 < L'P_1$ at P_1, and from $L'O = r$ to $\widehat{O}O < r$ at $P_2 = O$. Energy dissipates during each shock.



References

[1] 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 Cl. Sci. Fis. Mat. Natur.*, 32(1):97–108, 2021.

[2] L. Ambrosio, A. Baradat, and Y. Brenier. Monge-Ampère gravitation as a Γ-limit of good rate functions. *Anal. PDE*, 16(9):2005–2040, 2023.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, second edition, 2008.

[4] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Mathematical Journal*, 163(7):1405–1490, 2014.

[5] J.S. Bagla and T. Padmanabdan. Cosmological N-body simulations. *Pramana*, 49(2):161–192, 1997. arXiv:astro-ph/ 0411730v1.

[6] A. Bonnefous, Y. Brenier, and R. Mohayaee. Monge-Ampère gravity, optimal transport theory and their link to the Galileons. *Phys. Rev. D*, 110:123513, 2024. Preprint: arXiv 2405.15035.

[7] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[8] Y. Brenier. A modified least action principle allowing mass concentrations for the early Universe reconstruction problem. *Confluentes Mathematici*, 3(3):361–385, 2011.

[9] Y. Brenier. A double large deviation principle for Monge-Ampère gravitation. *Bulletin of the Institute of Mathematics. Academia Sinica (New Series)*, 11(1):23–41, 2016.

[10] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskii. Reconstruction of the early Universe as a convex optimization problem. *Mon. Not. R. Astron. Soc.*, 346(2):501–524, 2003.

[11] P. Cattiaux, G. Conforti, I. Gentil, and C. Léonard. Time reversal of diffusion processes under a finite entropy condition. *Annales de l'Institut Henri Poincaré Probab. Statist.*, 59(4):1844–1881, 2023.

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

[13] D.A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20:247–308, 1987.

[14] A. Dembo and O. Zeitouni. *Large deviations techniques and applications. Second edition.* Applications of Mathematics 38. Springer Verlag, 1998.

[15] H. Föllmer. Time reversal on Wiener space. In *Stochastic Processes - Mathematics and Physics*, volume 1158 of *Lecture Notes in Math.*, pages 119–129. Springer, Berlin, 1986.

[16] H. Föllmer. *Random fields and diffusion processes, in École d'été de probabilités de Saint-Flour XV-XVII-1985-87*, volume 1362 of *Lecture Notes in Mathematics*. Springer, Berlin, 1988.

[17] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems.* Springer, Berlin, New York, 2nd ed. edition, 1998.

[18] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevskii. A reconstruction of the initial conditions of the Universe by optimal mass transportation. *Nature*, 417, 16 May 2002.

[19] I.M. Gelfand and S.V. Fomin. *Calculus of variations.* Dover Publications Inc., 2000.

[20] N. Gigli. Second order analysis on $(\mathcal{P}_2(M), W_2)$. *Mem. Amer. Math. Soc.*, 216(1018):xxi+154, 2012.

[21] L. Hernquist, F. R. Bouchet, and Y. Suto. Application of the Ewald method to cosmological N-body simulations. *The Astrophysical Journal Supplement Series*, 75:231–240, 1991.

[22] J. Jost and X. Li-Jost. *Calculus of variations*. Number 64 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1998.

[23] C. Léonard. Girsanov theory under a finite entropy condition. In *Séminaire de probabilités, vol. 44.*, pages 429–465. Lecture Notes in Mathematics 2046. Springer, 2012.

[24] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst. A*, 34(4):1533–1574, 2014.

[25] B. Lévy, Y. Brenier, and R. Mohayaee. Monge Ampère gravity: from the large deviation principle to cosmological simulations through optimal transport. arXiv:2404.07697.

[26] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, 169(3):903–991, 2009.

[27] E. Madelung. Quantentheorie in hydrodynamischer Form. *Z. Phys.*, 40:322–326, 1926.

[28] M. Mariani. A Gamma-convergence approach to large deviations. *Ann. Sc. Norm. Super. Pisa. Cl. Sci.*, 18(3):951–976, 2018.

[29] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[30] P.J.E. Peebles. Tracing galaxy orbits back in time. *Astrophys. J.*, 344:L53–L56, 1989.

[31] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.

[32] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, 2:269–310, 1932.

[33] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

[34] M. von Renesse. An optimal transport view on Schrödinger's equation. *Canad. Math. Bull.*, 55(4):858–869, 2011.

[35] J.-C. Zambrini. Variational processes and stochastic versions of mechanics. *J. Math. Phys.*, 27:2307–2330, 1986.

[36] Ya. B. Zeldovich. Gravitational instability: an approximate theory for large density perturbations. *Astron. and Astrophys.*, 5:84–89, 1970.