A DOUBLE LARGE DEVIATION PRINCIPLE FOR MONGE-AMPERE GRAVITATION

Yann Brenier

A DOUBLE LARGE DEVIATION PRINCIPLE FOR MONGE-AMPÈRE GRAVITATION

YANN BRENIER,
CNRS UMR 7640,
ECOLE POLYTECHNIQUE, PALAISEAU, FRANCE.

ABSTRACT. Monge-Ampère gravitation is a nonlinear modification of classical Newtonian gravitation, when the Monge-Ampère equation substitutes for the Poisson equation. We establish, through two applications of the large deviation principle, that the MA gravitation for a finite number of particles can be reduced, through a double application of the large deviation principle, to the simplest possible stochastic model: a collection of independent Brownian motions with vanishing noise.

This paper is dedicated to Professor Tai-Ping Liu for his seventieth birthday.

INTRODUCTION

The purpose of this paper is twofold. We first want to make a short presentation of the Monge-Ampère gravitational (MAG) model, which has been introduced in [13], in close connection with earlier works such as [11, 15, 31, 20, 2] as well as with optimal transport theory (see [38, 3]). This model can be seen as a nonlinear modification of the classical model of Newtonian gravitation, for which we use the fully nonlinear Monge-Ampère equation as a substitute for the linear Poisson equation to derive the gravitational potential from the density field. We also briefly compare the Monge-Ampère and the Newton gravitational models, emphasizing that in the MAG model there is an absolute control of the acceleration of the gravitating particles, no matter how concentrated they can be, in sharp contrast with Newtonian gravitation, for which some particles may runaway to infinity in finite time. However, the main purpose of this paper is about deriving the MAG model from a completely elementary microscopic model in which a finite number of particles just move as independent Brownian trajectories without any interaction. In order to get the MAG model, we need two applications of the large deviation principle (LDP) [25, 21]. (Let us mention the interesting connection between large deviation principles, gradient flows and optimal transport theory recently made in [29, 32], which was certainly influential for us.) Through the first application of the LDP, we get a first order (in time) dynamical system based on the indistinguishability of the gravitating particles. (In a rather paradoxical way, the indistinguishability principle leads to a model where particles do interact.) The second application of the LDP enables us to lift this first order dynamical system to a second order one which is nothing but the discrete version of the MAG model. Let us emphasize that this derivation is purely formal and further investigations are clearly needed to get a complete and rigorous theory. The first section of this paper will be devoted

to a presentation of both the classical Newtonian and the Monge-Ampère models, with a suitable formulation of the MAG model relying on "optimal transport" tools [9, 10, 38]. The second section is devoted to the double application of the LDP in order to recover the discrete version of the MAG model.

1. Monge-Ampère and Newton gravitations

1.1. Classical Newtonian gravitation.
To describe classical Newton gravitation, let us attach to each gravitating particle a label a, which we suppose to belong to an abstract Borel measure space $(\mathcal{A}, \lambda)$, and its position at time t, $X_t(a) \in \mathbb{R}^d$ (classically $d = 3$). For each time t, we denote by ρ_t the image measure of λ by X_t, defined on $\mathbb{R}^d$ by

$$(1) \qquad \int_{\mathbb{R}^d} \gamma(x)\rho_t(dx) = \int_{\mathcal{A}} \gamma(X_t(a))\lambda(da) \quad \forall \gamma \in C_c^0(\mathbb{R}^d).$$

For the sake of simplicity, we further assume $\rho_t(dx)$ to be $\mathbb{Z}^d$ periodic in x and of unit mean on the unit cube $[0,1]^d$, so that we can see $\rho_t(dx)$ as a probability measure on the flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. (This kind of assumption is common in computational Cosmology [26].) We are now ready to write the Newtonian model:

$$(2) \qquad \frac{d^2 X_t(a)}{dt^2} = -\nabla \phi_t(X_t(a)), \quad \triangle \phi_t = \rho_t - 1,$$

where the average density 1 has been subtracted out from the field ρ_t so that the "gravitation potential" ϕ_t is well defined from the Poisson equation as a zero mean $\mathbb{Z}^d$−periodic function. (This is a perfectly meaningful assumption: there should be no resulting force $\nabla \phi_t$ out of a uniform ρ_t .) If, at time t, ρ_t is just a plain probability measure on $\mathbb{T}^d$, then $\nabla \phi_t$, obtained through the Poisson equation $\triangle \phi_t = \rho_t - 1$, is merely an L^p function for $p < (1 - d^{-1})^{-1}$ on $\mathbb{T}^d$. This makes very dubious the meaning of ODE

$$\frac{d^2 X_t(a)}{dt^2} = -\nabla \phi_t(X_t(a))$$

in the general case. This is why Newtonian's gravitation usually splits up into two, rather unrelated theories, according to the choice of its initial conditions:
1) As $\mathcal{A}$ is a discrete set of N particles, then Newtonian gravitation boils down to an "N-body problem", set on the flat torus $\mathbb{T}^d$, which is of form

$$(3) \qquad \frac{d^2 X_t(a)}{dt^2} = R(X_t(a)) + \sum_{b \neq a} G(X_t(a), X_t(b)),$$

where R is a fixed function depending on the torus, and the singular kernel $G(x, y)$, associated with the operator $-\nabla \triangle^{-1}$, behaves as $(y - x)/|y - x|^d$, as x approaches y. This "bad" singularity makes the study of the "N-body problem" very hard, in particular because of the possibility that particles may collide or runaway at infinity, in finite time. (See [18] as a recent reference.)
2) As $\mathcal{A}$ corresponds to a "continuum" of particles, it is fruitful to introduce, at each time t, the nonnegative measure f_t defined on the "phase space" $\mathbb{T}^d \times \mathbb{R}^d$ by

$$(4) \qquad \int_{\mathbb{T}^d \times \mathbb{R}^d} \gamma(x, v) f_t(dx, dv) = \int_{\mathcal{A}} \gamma(X_t(a), \frac{dX_t(a)}{dt})\lambda(da), \quad \forall \gamma \in C_c^0(\mathbb{T}^d \times \mathbb{R}^d).$$

A straightforward calculation shows that f must satisfy, at least formally, the "Vlasov-Poisson" system [35]

$$(5) \quad \partial_t f_t + \nabla_x \cdot (v f_t) - \nabla_v \cdot (\nabla \phi_t f_t) = 0, \quad \triangle \phi_t = \rho_t - 1 = \int_{\mathbb{R}^d} f_t(dv) - 1.$$

The meaning of this equation is unclear in the general case, because of the dubious product of $\nabla \phi_t$ by f_t. However, it is not hard to see that $0 \le f_t(dx, dv) \le C dx dv$ for some positive constant C is a consistent property for equation (5). It turns out that this property is sufficient to justify the multiplication of f_t by $\nabla \phi_t$ at least when the initial condition f_0 vanishes for large values of v, say for $|v| \ge C'$ for some other constant C'. Then, the Newtonian model gets well defined and, at least for $d \le 3$, global weak solutions can be shown to exist globally in time, as f_0 is given, whatever are constants C and C'. In addition, such solutions are unique and smooth as long as f_0 has a smooth density with respect to the Lebesgue measure. (See [34, 30, 8] for such results, stated on the whole space $\mathbb{R}^d$ rather than on $\mathbb{T}^d$. See also [35], as well as [22] for the closely related Euler-Poisson system.)

1.2. **Monge-Ampère gravitation.** We now introduce the Monge-Ampère gravitation (MAG) model which differs from (2) just by the substitution of the Monge-Ampère equation for the Poisson equation:

$$(6) \qquad \frac{d^2 X_t(a)}{dt^2} = -\nabla \phi_t(X_t(a)), \quad \det(\mathbb{I} + D^2 \phi_t) = \rho_t,$$

where $\mathbb{I}$ denotes the $d \times d$ identity matrix and ϕ_t is solution to the Monge-Ampère equation (in a suitable sense), which is $\mathbb{Z}^d-$periodic with zero mean and satisfies the (weak) ellipticity condition $\mathbb{I} + D^2 \phi_t(x) \ge 0$, in the sense of symmetric matrices, for every x. [Notice that ϕ_t is unique and Lipschitz continuous (resp. smooth) as soon as $\rho_t(dx)$ has an integrable (resp. smooth and positive) density with respect to the Lebesgue measure dx, see [19] for example.] We see that Newtonian gravitation can be formally recovered from the MAG model just by expanding the determinant about $\mathbb{I}$ and retaining only the linear part:

$$\det(\mathbb{I} + D^2 \phi_t) \sim 1 + \text{Trace} D^2 \phi_t = 1 + \triangle \phi_t.$$

Notice that, as $d = 1$ (which is a case of limited interest, describing gravitating "parallel pancakes"), the MAG model coincide with Newtonian gravitation. (As a consequence, our derivation of the MAG model from a double application of the large deviation principle, obtained in the second section of the present paper, is also valid for the Newtonian gravitational model in one space dimension.) However, let us emphasize that, to the best of our knowledge, MAG has never been considered by any physicist in dimension larger than one. It has been so far only considered by mathematicians (see [13] and [15, 31, 20, 2] for closely related topics), mostly because of its close connection with optimal transport theory (as discussed in the next subsection). It is easy to describe, as we did for the Newtonian model, the MAG model through a "kinetic equation", the so-called "Vlasov-Monge-Ampère" (VMA) system [15, 31]:

$$(7) \quad \partial_t f_t + \nabla_x \cdot (v f_t) - \nabla_v \cdot (\nabla \phi_t f_t) = 0, \quad \det(\mathbb{I} + D^2 \phi_t) = \rho_t = \int_{\mathbb{R}^d} f_t(dv).$$

As for the Vlasov-Poisson system (5), the existence of global weak solutions can be shown as soon as

$$0 \leq f_0(dx, dv) \leq C dx dv, \quad f_0(\mathbb{T}^d \times \{v \in \mathbb{R}^d, \ |v| \leq C'\}) = 0,$$

for some positive constants C, C'. Existence of a unique smooth solution, but only for a short time interval, has been proven by Loeper [31] provided $f_0(dx, dv)$ (resp. $\rho_0(dx)$) has a smooth density with respect to $dxdv$ (resp. dx) and is uniformly compactly supported in $v \in \mathbb{R}^d$ (resp. is strictly positive). These are still limited results, due to the full non-linearity of the Monge-Ampère equation that leads to analytic difficulties, in particular when seeking for smooth solutions. However, the Monge-Ampère equation enjoys remarkable properties, closely related to the theory of optimal transportation [10, 38]. This is why we are going to introduce a related formulation of the MAG model, with interesting geometric features.

1.3. **The MAG model written in optimal transportation terms.** For the description of the MAG model in optimal transport terms, it is convenient to discuss the model in a (slightly) different and more abstract framework. We consider a metric measure space $(\mathcal{A}, \lambda)$, made of a compact subset $\mathcal{A}$ of $\mathbb{R}^d$ equipped with a Borel probability measure λ. Two typical examples are, on one side, the unit cube with the Lebesgue measure, and, on the other side, any set of N points equipped with the (normalized) counting measure. In the first case, we will speak of the "continuous" case, while, in the second case, we will speak of the "discrete" case. (Of course many others situations could be also considered, in particular the flat torus $\mathbb{T}^d$ as we did in the previous subsections, but we will focus on these two cases only.) We introduce the separable Hilbert space H of all $\lambda-$square-integrable maps from $\mathcal{A}$ to $\mathbb{R}^d$, $H = L^2(\mathcal{A}, \lambda; \mathbb{R}^d)$, with norm and inner product respectively denoted by $|| \cdot ||$ and $((\cdot, \cdot))$. (Notice that, when $\mathcal{A}$ is a finite set of N points in $\mathbb{R}^d$, then $H \sim \mathbb{R}^{Nd}$ is of finite dimension.) We crucially consider the subset S of all measure-preserving maps s of A:

$$(8) \qquad S = \{s \in H, \ \int_{\mathcal{A}} \gamma(s(a))\lambda(da) = \int_{\mathcal{A}} \gamma(a)\lambda(da), \ \forall \gamma \in C^0(\mathbb{R}^d)\}.$$

(Observe, in the discrete case, when $\mathcal{A}$ is made of N distinct points $A(a) \in \mathbb{R}^d$, for $a = 1, \cdots, N$, S just reads $S = \{(A(\sigma(1)), \cdots, A(\sigma(N)) \in H, \ \sigma \in \mathcal{S}_N\}$ with $N!$ elements, where $\mathcal{S}_N$ denotes the group of all permutations of the first N integers.)

According to Edelstein's theorem [23, 4], given a separable Hilbert space H and a closed bounded subset S, there is, in the sense of Baire, a generic set (i.e. containing a countable intersection of dense open subsets of H) of points X for which there exists a unique closest point $\pi(X)$ on S. In addition $\pi(X)$ is nothing but the gradient, at point X, of the Lipschitz convex function Π defined on the Hilbert space H by

$$(9) \qquad \Pi(X) = \sup_{s \in S}((X, s)) - \frac{||s||^2}{2}, \quad \pi(X) = \nabla\Pi(X) = \text{Arg} \inf_{s \in S} \frac{||X - s||^2}{2}.$$

In our particular case, we can say much more in the continuous case, when $\mathcal{A}$ is the unit cube with λ as the Lebesgue measure, thanks to "optimal transport theory"

Theorem 1.1. *([10]) Let $X \in H$ be a non degenerate map, in the sense that the image measure ρ of λ by X is absolutely continuous with respect to the Lebesgue*

measure on $\mathbb{R}^d$. Then, X has a unique closest point $\pi(X)$ on S. We also have

$$\pi(X) = T \circ X, \tag{10}$$

where T is the unique map (in the ρ a.e. sense) $y \in \mathbb{R}^d \to T(y) \in \mathcal{A}$ such that:
i) there is a convex Lipschitz function $\psi : \mathbb{R}^d \to \mathbb{R}$, $\rho-$a.e. differentiable, with:

$$T(y) = \nabla \psi(y), \quad \rho - a.e. \ y \in \mathbb{R}^d \tag{11}$$

ii) λ is the image of ρ by T:

$$\int_{\mathbb{R}^\Gamma} \gamma(T(y))\rho(dy) = \int_{\mathcal{A}} \gamma(a)\lambda(da), \quad \forall \gamma \in C^0(\mathbb{R}^d). \tag{12}$$

As a matter of fact, equations (11,12) form a generalized formulation of the Monge-Ampère problem on $\mathbb{R}^d$

$$\rho = \det(D_x^2 \psi), \ \ (\nabla \psi)(\text{support}(\rho)) = \mathcal{A}, \ \ D^2\psi \geq 0, \tag{13}$$

with a unique solution $\nabla \psi$ (in the $\rho-$a.e. sense). (See [9, 10, 38] for more details.) This generalized formulation of the Monge-Ampère equation allows us to write the MAG model (6) in a much more geometric way. Indeed, (6) just reads

$$\frac{d^2 X_t(a)}{dt^2} = X_t(a) - \nabla \psi_t(X_t(a)), \ \ \det(D^2\psi_t) = \rho_t,$$

after setting $\psi_t(x) = |x|^2/2 + \phi_t(x)$, that we complete with the weak ellipticity condition $D^2\psi_t \geq 0$ and the range condition $\nabla \psi_t(\text{support}(\rho_t)) = \mathcal{A}$ (which substitutes for the $\mathbb{Z}^d$-periodicity condition we used in writing (6)). Assume that, at time t, X_t is non degenerate (or, in other words, ρ_t is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$). Then, using Theorem 1.1, we may write

$$\nabla \psi_t \circ X_t = \pi(X_t) = \nabla \Pi(X_t)$$

and finally obtain

$$\frac{d^2 X_t}{dt^2} = X_t - \pi \circ X_t = X_t - \nabla \Pi \circ X_t, \tag{14}$$

where Π is the Lipschitz convex function defined by (9), π its gradient and S is the set of measure preserving maps defined by (8). In the rest of this paper, we will retain (8,9,14) as our definition of the MAG model.

1.4. **The discrete Monge-Ampère gravitational model.** The geometric formulation (8,9,14) of the MAG model is very convenient to get its discrete version, as $\mathcal{A}$ is just a subset of N points $A(a)$ in $\mathbb{R}^d$, for $a = 1, \cdots, N$, equipped with the counting measure, in which case, we will speak of the "discrete MAG model with N particles". Indeed, at the discrete level, a time-dependent map X_t can be seen as a sequence of N "particles", with positions $X_t(a)$, moving in $\mathbb{R}^d$, for $a = 1, \cdots, N$. (Notice that X_t gets degenerate at time t in case of collisions.) In this discrete setting, the MAG model (8,9,14) reads as the finite dimensional dynamical system:

$$\frac{d^2 X_t(a)}{dt^2} = X_t(a) - A(\sigma_t(a)), \ \ \sigma_t = \text{Arg} \inf_{\sigma \in \mathcal{S}_N} \sum_{a=1}^{N} |X_t(a) - A(\sigma(a))|^2, \tag{15}$$

where $\sigma \in \mathcal{S}_N$ denotes the set of all permutations of the first N integers. So we see that, at the discrete level, the MAG looks both very simple and very different from the classical Newtonian N body problem (3)! (Similar systems have been previously studied in [11, 20].)

The MAG model, as defined by (8,9,14), can be seen as a dynamical system in a Hilbert space H with a force term $\mathcal{F}(x) = x - \nabla\Pi(x)$ with a trivial linear part and the gradient of a Lipschitz convex function Π. To the best of our knowledge, there is no theory for such an ODE in infinite dimension. Indeed, we are very far from the standard Cauchy-Lipschitz setting. This is a challenging open problem. (See the related theory developed by Ambrosio and Gangbo for some infinite dimensional hamiltonian systems [2]. As a matter of fact, their main example is very similar to the MAG model, written in a different way. See also [15, 27].) However, for the discrete MAG model with N particles, $H \sim \mathbb{R}^{Nd}$ is of finite dimension. Thus, equation (14), which also reads (15), can be neatly solved globally in time in the sense of Bouchut and Ambrosio [7, 1], for every initial condition $(X_0, \frac{dX_0}{dt})$, except on a negligible subset of the "phase space" $H \times H$, of zero $2Nd$ dimensional Lebesgue measure. (A more accurate statement can be found in [1].) Indeed, the force term $\mathcal{F}(x) = x - \nabla\Pi(x)$ in the right-hand side of (14) is a smooth perturbation of a "bounded variation" function, since Π is a Lipschitz convex function (which implies that $D^2\Pi$ can be seen as a bounded nonnegative measure valued in the convex cone of nonnegative symmetric matrices, and, therefore, that $\nabla\Pi$ is of bounded variation). Notice that the exceptional set of bad initial conditions in the phase space is not empty, as it follows clearly from formulation (15), where we see that the evolution of the particles becomes ambiguous as different particles depart from the same position with *exactly* the same velocity.

1.5. **Monge-Ampère versus Newtonian gravitational models.** It is now interesting to compare the Monge-Ampère and the Newton gravitational models. We first observe that, according to the MAG model, particles may never runaway to infinity in finite time. Indeed, from the optimal transportation formulation (8,9,14), we immediately get

$$|\frac{d^2 X_t}{dt^2} - X_t| \ \leq \ R = \sup_{a \in \mathcal{A}} \ |a|$$

and deduce that

$$|X_t| + |\frac{dX_t}{dt}| \leq (|X_0| + |\frac{dX_0}{dt}| + 1)C \cosh t, \ \ \forall t \in \mathbb{R},$$

where C depends only on R. Of course, there is nothing similar with Newtonian gravitation. Indeed, the Poisson equation in (2) does not behave well when particles concentrate. For instance, if some particles concentrate as a delta measure at some point y at time t, then $\nabla\phi_t(x)$ has a singularity as bad as $(x-y)/|x-y|^d$.

In this way, the MAG model has a lot of similarity with the Born-Infeld (BI) theory of the electromagnetic field [6, 5, 12, 16], in which any electrostatic field is bounded by a universal constant. (Notice that the BI model, which goes back to 1934, is no longer for use in Electrodynamics, but has enjoyed a remarkable revival in String Theory since the 1990s [33].)

In addition, the MAG model enjoys good properties of approximations by finite sums of Dirac measures (see [20] for closely related results). In sharp contrast, such discrete approximations have never been justified, to the best of our knowledge, in the case of 3D Newtonian gravitation, because of the bad singularities of the Green function for the Poisson equation [28]. As a matter of fact, the treatment of

point particles in classical Electrodynamics remains an outstanding open problem in both Theoretical and Mathematical Physics (see [24, 37] for instance).

2. Formal derivation of the discrete Monge-Ampère gravitation model from a double application of the large deviation principle

In this second section, we provide a *formal* derivation of the discrete MAG model (15) from a double application of the large deviation principle [25, 21].

2.1. **A basic model of N independent Brownian particles.** We fix a positive integer N and a finite set $\mathcal{A}$ of N distinct points $A(a) \in \mathbb{R}^d$, for $a = 1, \cdots, N$. Using the notations of subsection 1.3, we set $H = \mathbb{R}^{Nd}$ and

$$(16) \qquad S = \{(A(\sigma(1)), \cdots, A(\sigma(N))) \in H, \ \sigma \in \mathcal{S}_N\},$$

where $\mathcal{S}_N$ denotes the set of all possible permutations of the first N integers. Given $\sigma_0 \in \mathcal{S}_N$, we consider the motion of N "particles" that move in the Euclidean space $\mathbb{R}^d$, according to

$$(17) \qquad X_t^\varepsilon(a) = s_0(a) + \varepsilon B_t(a), \ s_0(a) = A(\sigma_0(a)), \ \forall a = 1, \cdots, N.$$

where $(t \in \mathbb{R}_+ \to B_t(a) \in \mathbb{R}^d)_{a=1,\cdots,N}$ denote N independent realizations of the "standard" (i.e. normalized) Brownian motion in $\mathbb{R}^d$.
Fixing $t^* > 0$ and

$$Y^*(a) \in \mathbb{R}^d, \ \forall a = 1, \cdots, N,$$

it is easy to compute the probability that, at time t^*, each particle occupies the position given by Y^*:

$$\text{Prob}(X_{t^*}^\varepsilon(a) \approx Y^*(a), \ \forall a = 1, \cdots, N)$$

$$\approx \prod_{a=1}^N [\exp(\frac{-|Y^*(a) - s_0(a)|^2}{2\varepsilon t^*})(2\pi\varepsilon t^*)^{-d/2}],$$

or, in other words

$$\text{Prob}(X_{t^*}^\varepsilon \approx Y^*) \approx \exp(\frac{-||Y^* - s_0||^2}{2\varepsilon t^*})(2\pi\varepsilon t^*)^{-Nd/2}.$$

Here we have denoted by $|\cdot|$ and $||\cdot||$ the Euclidean distance on respectively $\mathbb{R}^d$ and $H = (\mathbb{R}^d)^N$. We have also used symbol $\approx$ just to make notations lighter. [What we precisely mean is: for any Borel $\mathcal{B}$ subset of $(\mathbb{R}^d)^N$, the probability that $X_{t^*}^\varepsilon$ belongs to $Y^* + \mathcal{B}$ is given by

$$\int_{Y \in Y^* + \mathcal{B}} \exp(\frac{-||Y - s_0||^2}{2\varepsilon t^*})(2\pi\varepsilon t^*)^{-Nd/2} dY,$$

but we hope that our simplified notation is acceptable. Indeed, we will use it again, without further notice.]

2.2. **First application of the large deviation principle.** We now want to compute the probability that $X_{t^*}^\varepsilon$ and Y^* coincide, up to a permutation, property that we denote by $(X_{t^*}^\varepsilon) \underset{\text{perm}}{\approx} (Y^*)$. We find

$$(18) \qquad \text{Prob}[(X_{t^*}^\varepsilon) \underset{\text{perm}}{\approx} (Y^*)] \approx \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \exp(\frac{-||Y^* \circ \sigma - s_0||^2}{2\varepsilon t^*})(2\pi\varepsilon t^*)^{-Nd/2}.$$

When the level of noise ε goes to zero, we immediately get

$$-\lim_{\varepsilon \to 0} \varepsilon \log \text{Prob}[(X_{t^*}^\varepsilon) \underset{\text{perm}}{\approx} (Y^*)] \approx \inf_{\sigma \in \mathcal{S}_N} \frac{||Y^* \circ \sigma - s_0||^2}{2t^*},$$

which is a rather trivial illustration of the Laplace method and the large deviation principe (LDP) [25]. There is a more sophisticated aspect of the LDP: as the level of noise goes to zero, the Brownian trajectories of the particles conditioned by $(X_{t^*}^\varepsilon) \underset{\text{perm}}{\approx} (Y^*)$ behave more and more as constant speed minimizing geodesic curves. More precisely, for all $t \in [0, t^*]$,

$$(19) \quad X_t^\varepsilon \sim_{\varepsilon \to 0} \quad X_t = s_0 + \frac{t}{t^*}(Y^* \circ \sigma^* - s_0), \quad \sigma^* = \text{Arg} \inf_{\sigma \in \mathcal{S}_N} \frac{||Y^* \circ \sigma - s_0||^2}{2t^*}.$$

which implies

$$(20) \qquad \frac{dX_t}{dt} = \frac{Y^* \circ \sigma^* - s_0}{t^*} = \frac{X_{t^*} - s_0}{t^*} = \frac{X_t - s_0}{t}, \quad \forall t \in]0, t^*].$$

Now, we are going to translate this large deviation principle into a self-consistent dynamical system for the particles. Let us first denote by $\pi(Y)$ (as in subsection 1.3) the unique closest point on S of a "generic" point $Y \in H = (\mathbb{R}^d)^N$

$$(21) \qquad \pi(Y) = \text{Arg} \inf_{s \in S} \frac{||Y - s||^2}{2} = \nabla \Pi(Y), \quad \Pi(Y) = \sup_{s \in S}((Y, s)) - \frac{||s||^2}{2}$$

(where $((\cdot, \cdot))$ denotes the inner product attached to $|| \cdot ||$ on H). Then we state:

Proposition 2.1. *Equation (20), that we have derived from (17) (through a large deviation principle), can be written as a self-consistent ordinary differential equation for X_t at least for $t > 0$,*

$$(22) \qquad t\frac{dX_t}{dt} = X_t - \pi(X_t).$$

Proof. To get this result, our simple but crucial observation is that, along the geodesic curve defined by (19), namely

$$X_t = s_0 + \frac{t}{t^*}(Y^* \circ \sigma^* - s_0), \quad \sigma^* = \text{Arg} \inf_{\sigma \in \mathcal{S}_N} ||Y^* \circ \sigma - s_0||,$$

s_0 is the closest point in S not only of the end-point $X_{t^*} = Y^* \circ \sigma^*$ but also of all points X_t, $\forall t \in [0, t^*]$. [Then, we can write $s_0 = \pi(X_t)$ in equation (20) which immediately leads to (22) and completes the proof of our Proposition.]
Although this property is geometrically quite obvious, let us provide a comprehensive proof for the sake of completeness.
By definition (21) of the closest point operator π, it is enough to show that

$$\kappa = ||X_t - s_0 \circ \sigma||^2 - ||X_t - s_0||^2 \geq 0$$

for each permutation $\sigma \in \mathcal{S}_N$ and each $t \in [0, t^*]$. By definition (19),

$$\kappa = ||s_0 + \frac{t}{t^*}(X^* - s_0) - s_0 \circ \sigma||^2 - ||\frac{t}{t^*}(X^* - s_0)||^2$$

where we set, still according to (19),

$$(23) \qquad X^* = X_{t^*} = Y^* \circ \sigma^*, \quad \sigma^* = \text{Arg} \inf_{\sigma \in \mathcal{S}_N} ||Y^* \circ \sigma - s_0||.$$

Expanding the squares, we obtain

$$\kappa = ||s_0 - s_0 \circ \sigma||^2 + \frac{2t}{t^*}((X^* - s_0, s_0 - s_0 \circ \sigma)).$$

By definition of $X^* = Y^* \circ \sigma^*$ and σ^*, we get from (23)

$$||X^* - s_0|| = ||Y^* \circ \sigma^* - s_0|| \leq ||Y^* \circ \sigma^* \circ \sigma^{-1} - s_0|| = ||X^* - s_0 \circ \sigma||.$$

Thus $((X^*, s_0 - s_0 \circ \sigma)) \geq 0$. So, we deduce

$$\kappa \geq ||s_0 - s_0 \circ \sigma||^2 - \frac{2t}{t^*}((s_0, s_0 - s_0 \circ \sigma)) = 2(1 - \frac{t}{t^*})[||s_0||^2 - ((s_0, s_0 \circ \sigma))] \geq 0$$

(using $||s_0|| = ||s_0 \circ \sigma||$ and the Cauchy-Schwarz inequality), which is the desired inequality and completes the proof of Proposition 2.1.

Equation (22) is clearly singular at time $t = 0$. We can lift this singularity with an exponential rescaling of time $t = \exp(\theta)$, $\theta \in \mathbb{R}$, and finally obtain:

$$(24) \qquad \frac{dX_\theta}{d\theta} = X_\theta - \pi(X_\theta) = X_\theta - \nabla\Pi(X_\theta).$$

So we have obtained a first order (in time) dynamical system, of "gradient type".

Miscellaneous remarks. i) Since π is the gradient of a Lipschitz convex function (namely Π defined by (9)), equation (24) is uniquely solvable in the framework of "maximal monotone operator theory" [17]. As already discussed in [13], this is a way to introduce a dissipative mechanism in the motion of particles, such as sticky collisions when particles stick to each other while conserving their momentum (but not their kinetic energy which decreases). At the level of the present paper, we do not want to enter such considerations and leave (24) just as a *formal* equation.
ii) Notice that (24) is a gradient flow in the variable X, valued in the Hilbert space H, which is the counterpart of the gradient flow in the variable $\text{Law}(X)$, in the so-called "Wasserstein space" of half the negative squared "Wasserstein distance", as discussed in [3]. For us, it is important to keep a formulation in terms of $X \in H$ and not in terms of $\text{Law}(X)$, because, in the sequel of our discussion, it is crucial to restore the "individuality" of the gravitating particles during their motion, in order to get a second order dynamical system for them.
iii) Quite remarkably, as explained in [13], equation (22) is nothing but the Zeldovich model used in Cosmology [39, 36, 26, 14] as an approximation of semi-Newtonian gravitation in an Einstein-de Sitter space!
iv) Let us finally provide a possible interpretation of equation (24), vaguely related to the so-called "anthropic principle": What is observed at out "present" time t^* is just a random output of the independent Brownian trajectories of a large number N of *indistinguishable* particles initially located on the set $\mathcal{A}$. As the noise vanishes, the motion of these particles, conditioned by what we can observe now, just looks driven by the deterministic law (24). Of course, this is a highly

speculative and questionable interpretation coming from a mathematician and not from an authorized physicist.

2.3. **Second application of the large deviation principle.** Equation (24) is a *first* order dynamical system, a so-called "gradient flow" since (24) can also be written

$$dX_\theta = \nabla\Phi(X_\theta)d\theta,$$

where Φ is half of the squared distance function to S:

$$(25) \qquad \Phi(Y) = \frac{||Y||^2}{2} - \Pi(Y) = \inf_{s\in S}\frac{||Y-s||^2}{2},$$

where Π is defined by (9). Describing a gravitational theory by a gradient flow does not sound reasonable. We would rather like to get a *second* order, conservative, dynamical system. Here again, the large deviation principle turns out to be useful to get such a second order system out of (24). Fixing $\eta > 0$, we introduce the "noisy" version of (25) defined by:

$$(26) \qquad dX_\theta^\eta = \nabla\Phi(X_\theta^\eta)d\theta + \eta\, dB_\theta,$$

where B_θ (again) denotes a Brownian process in $(\mathbb{R}^d)^N$. Given two points Y_0 and Y_1 in $(\mathbb{R}^d)^N$, we expect from the LDP (or, more precisely, the Freidlin-Wentzell theorem [25, 21]) that, as the level of noise η goes to zero,

$$-\eta\log\mathrm{Prob}[X_{\theta_0}^\eta{\approx}Y_0 \text{ and } X_{\theta_1}^\eta{\approx}Y_1] \sim_{\eta\to 0}\quad \mathcal{A}(\theta_0, Y_0, \theta_1, Y_1)$$

$$(27)\quad \mathcal{A}(\theta_0, Y_0, \theta_1, Y_1) = \inf\{\int_{\theta_0}^{\theta_1}\frac{1}{2}||\frac{dX_\theta}{d\theta} - \nabla\Phi(X_\theta)||^2 d\theta,\ \ X_{\theta_0} = Y_0,\ \ X_{\theta_1} = Y_1\}.$$

In addition, as η goes to zero,

$$X_\theta^\eta \sim X_\theta,\ \ \forall\theta\in[\theta_0,\theta_1],$$

where

$$X = \mathrm{Arg}\inf\{\int_{\theta_0}^{\theta_1}\frac{1}{2}||\frac{dX_\theta}{d\theta} - \nabla\Phi(X_\theta)||^2 d\theta,\ \ X_{\theta_0} = Y_0,\ \ X_{\theta_1} = Y_1\}.$$

Strictly speaking, this is correct when $\nabla\Phi$ is Lipschitz continuous, which is not true in our case (where $\nabla\Phi$ is not even continuous), without further assumptions on the data. However, from the pure modeling viewpoint, it is very tempting to find the second order dynamical system linked to the least action principle (27). Since Φ, as defined by (25), is half of a squared distance function, we have

$$(28) \qquad \frac{1}{2}||\nabla\Phi(Y)||^2 = \Phi(Y),$$

for every $Y \in H \setminus \mathcal{N}$, where $\mathcal{N}$ is the set on which Φ is not differentiable, which is a negligible subset of H, both in the Lebesgue almost everywhere sense and in the Baire category sense. Thus, at least for each curve X that stays away from $\mathcal{N}$ for Lebesgue almost every $\theta \in [\theta_0, \theta_1]$, we have

$$\int_{\theta_0}^{\theta_1}\frac{1}{2}||\frac{dX_\theta}{d\theta} - \nabla\Phi(X_\theta)||^2 d\theta$$

$$= \int_{\theta_0}^{\theta_1}(\frac{1}{2}||\frac{dX_\theta}{d\theta}||^2 + ||\nabla\Phi(X_\theta)||^2)d\theta - \int_{\theta_0}^{\theta_1}\nabla\Phi(X_\theta)\cdot\frac{dX_\theta}{d\theta}d\theta$$

$$= \int_{\theta_0}^{\theta_1} (\frac{1}{2}||\frac{dX_\theta}{d\theta}||^2 + \Phi(X_\theta))d\theta - \Phi(X_{\theta_1}) + \Phi(X_{\theta_0}).$$

Therefore, the action principle (27) is equivalent to

$$(29) \quad \tilde{\mathcal{A}}(\theta_0, Y_0, \theta_1, Y_1) = \inf\{\int_{\theta_0}^{\theta_1} (\frac{1}{2}||\frac{dX_\theta}{d\theta}||^2 + \Phi(X_\theta))d\theta, \ \ X_{\theta_0} = Y_0, \ \ X_{\theta_1} = Y_1\}.$$

From this equivalent LAP, we find as optimality equation

$$(30) \qquad\qquad \frac{d^2 X_\theta}{d\theta^2} = \nabla \Phi(X_\theta) = X_\theta - \pi(X_\theta),$$

which is just the second order version of the gradient flow equation (24). More explicitly, we have obtained

$$(31) \qquad\qquad \frac{d^2 X_\theta(a)}{d\theta^2} = X_\theta(a) - A(\sigma_\theta(a)),$$

where

$$\sigma_\theta = \text{Arg} \inf_{\sigma \in \mathcal{S}_N} \sum_{a=1}^{N} |X_\theta(a) - A(\sigma(a))|^2,$$

which is nothing but the discrete version (15) of the Monge-Ampère model of gravitation discussed in the first section.

So we have achieved, at least at a formal level, the main goal of our paper which was, through a double application of the large deviation principle, the derivation of the (discrete) Monge-Ampère gravitational model from one of the simplest thinkable model of particles: N independent Brownian trajectories, with an intriguing interplay between their indistinguishability (for the first application of the LDP) and their individuality (for the second application of the LDP).

REFERENCES

[1] L. Ambrosio, *Transport equation and Cauchy problem for BV vector fields,* Invent. Math. 158 (2004) 227-260.

[2] L. Ambrosio, W. Gangbo, *Hamiltonian ODE in the Wasserstein spaces of probability measures,* Comm. Pure Appl. Math. 61(2008) 18-53.

[3] L. Ambrosio, N.Gigli, G. Savaré *Gradient flows in metric spaces and the Wasserstein spaces of probability measures,* Lectures in Mathematics, ETH Zurich, Birkhäuser, 2005.

[4] J.-P. Aubin, *Mathematical methods of game and economic theory,* Studies in Mathematics and its Applications, 7. North-Holland 1979.

[5] G. Boilliat, C. Dafermos, P.D. Lax, T.P. Liu, *Recent mathematical methods in nonlinear wave propagation,* C.I.M.E. Lecture Notes in Mathematics, 1640, Springer-Verlag, 1996.

[6] M. Born, L. Infeld, *Foundations of the new field theory,* Proc. Roy. Soc. London, A 144 (1934), 425-451.

[7] F. Bouchut, *Renormalized solutions to the Vlasov equation with coefficients of bounded variation,* Arch. Ration. Mech. Anal. 157 (2001) 75-90.

[8] F. Bouchut, F. Golse, M. Pulvirenti, *Kinetic equations and asymptotic theory.* Gauthier-Villars, Elsevier, 2000.

[9] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs,* C. R. Acad. Sci. Paris I Math. 305 (1987) 805-808.

[10] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions,* Comm. Pure Appl. Math. 44 (1991) 375-417.

[11] Y. Brenier, *Derivation of the Euler equations from a caricature of Coulomb interaction,* Comm. Math. Phys. 212 (2000) 93-104.

[12] Y. Brenier, *Hydrodynamic sturcture of the augmented Born-Infeld equations,* Arch. Rational Mech. Anal., 172 (2004), 65-91.

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

[14] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, Mohayaee, Sobolevskii, *Reconstruction of the early universe as a convex optimization problem,* Mon. Not. R. Astron. Soc. 2002.

[15] Y. Brenier, G., Loeper, *A geometric approximation to the Euler equations: The Vlasov-Monge- Ampère equation,* Geom. Funct. Anal. 14(2004) 1182-1218.

[16] Y. Brenier, W. Yong, *Derivation of particle, string, and membrane motions from the Born-Infeld electromagnetism,* J. Math. Phys. 46 (2005) 062305.

[17] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert,* North-Holland Mathematics Studies, No. 5. 1973.

[18] A. Chenciner, R. Montgomery, *A remarkable periodic solution of the three-body problem in the case of equal masses,* Ann. of Math. (2) 152 (2000) 881-901.

[19] D. Cordero-Erausquin, *Sur le transport de mesures périodiques,* C. R. Acad. Sci. Paris Sér. I Math. 329 (1999) 199-202.

[20] M.Cullen, W. Gangbo, G. Pisante, *The semigeostrophic equations discretized in reference and dual variables,* Arch. Ration. Mech. Anal. 185 (2007) 341-363.

[21] A. Dembo, O. Zeitouni, *Large deviations techniques and applications,* Springer-Verlag, 2010.

[22] Y. Deng, T.P. Liu, T. Yang, Z-a Yao, *Solutions of Euler-Poisson equations for gaseous stars,* Arch. Ration. Mech. Anal. 164 (2002) 261-285.

[23] M. Edelstein *On nearest points of sets in uniformly convex Banach spaces,* J. London Math. Soc. 43 (1968) 375-377.

[24] R. Feynman, *The Feynman lectures in Physics, vol.II, chapter 29,* AddisonWesley 1964.

[25] M. Freidlin, A. Wentzell (Ventcel), *Random perturbations of dynamical systems,* Springer 2012.

[26] U. Frisch, S. Matarrese, R. Mohayaee, A. Sobolevski, *A reconstruction of the initial conditions of the Universe by optimal mass transportation,* Nature 417 (2002) 260-262.

[27] W. Gangbo, T. Nguyen, A. Tudorascu, *Hamilton-Jacobi equations in the Wasserstein space,* Methods Appl. Anal. 15 (2008) 155-183.

[28] M. Hauray, P.-E. Jabin, *N-particles approximation of the Vlasov equations with singular potential,* Arch. Ration. Mech. Anal. 183 (2007) 489-524.

[29] Ch. Léonard, *From the Schrödinger problem to the Monge-Kantorovich problem,* J. Funct. Anal. 262 (2012) 1879-1920.

[30] P.-L. Lions, B. Perthame, *Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson System,* Invent. Math. 105 (1991) 415-430.

[31] G. Loeper, *Quasi-neutral limit of the Euler-Poisson and Euler-Monge-Ampère systems,* Comm. Partial Differential Equations 30 (2005) 1141-1167.

[32] A. Mielke, M. Peletier, D. Renger, *On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion,* Potential Anal. 41 (2014) 1293-1327.

[33] J. Polchinski, *String theory, Volume I,* Cambridge University Press, 1998.

[34] K. Pfaffelmoser, *Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data,* J. Differential Equations 95 (1992) 281-303.

[35] G. Rein, *Collisionless kinetic equations from astrophysics-the Vlasov-Poisson system,* Handbook of differential equations: evolutionary equations. Vol. III, 383-476, Handb. Differ. Equ., Elsevier 2007.

[36] S. Shandarin, Y. Zeldovich, *The large-scale structure of the universe: turbulence, intermittency, structures in a self-gravitating medium,* Rev. Modern Phys. 61 (1989) 185-220.

[37] H. Spohn, *Large Scale Dynamics of Interacting Particles,* Springer-Verlag 1991.

[38] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58, AMS, Providence, 2003.

[39] Y. Zeldovich, *Gravitational instability: An approximate theory for large density perturbations,* Astron. Astrophys. 5, 84-89 (1970).