ANALYSIS & PDE

Volume 11 No. 6 2018

ROBERT J. BERMAN AND MAGNUS ÖNNHEIM

PROPAGATION OF CHAOS, WASSERSTEIN GRADIENT FLOWS AND TORIC KÄHLER–EINSTEIN METRICS

msp

ANALYSIS AND PDE
Vol. 11, No. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1343

PROPAGATION OF CHAOS, WASSERSTEIN GRADIENT FLOWS AND TORIC KÄHLER–EINSTEIN METRICS

ROBERT J. BERMAN AND MAGNUS ÖNNHEIM

Motivated by a probabilistic approach to Kähler–Einstein metrics we consider a general nonequilibrium statistical mechanics model in Euclidean space consisting of the stochastic gradient flow of a given (possibly singular) quasiconvex N-particle interaction energy. We show that a deterministic "macroscopic" evolution equation emerges in the large N-limit of many particles. This is a strengthening of previous results which required a uniform two-sided bound on the Hessian of the interaction energy. The proof uses the theory of weak gradient flows on the Wasserstein space. Applied to the setting of permanental point processes at "negative temperature", the corresponding limiting evolution equation yields a drift-diffusion equation, coupled to the Monge–Ampère operator, whose static solutions correspond to toric Kähler–Einstein metrics. This drift-diffusion equation is the gradient flow on the Wasserstein space of probability measures of the K-energy functional in Kähler geometry and it can be seen as a fully nonlinear version of various extensively studied dissipative evolution equations and conservation laws, including the Keller–Segel equation and Burger's equation. In a companion paper, applications to singular pair interactions in one dimension are given.

1. Introduction

The present work is motivated by the probabilistic approach to the construction of canonical metrics, or more precisely Kähler–Einstein metrics, on complex algebraic varieties introduced in [Berman 2013a; 2017], formulated in terms of certain β-deformations of determinantal (fermionic) point processes. The approach in those papers uses ideas from equilibrium statistical mechanics (Boltzmann–Gibbs measures) and the main challenge concerns the existence problem for Kähler–Einstein metrics on a complex manifold X with *positive* Ricci curvature, which is closely related to the seminal Yau–Tian–Donaldson conjecture in complex geometry. In this paper, which is one in a series, we will be concerned with a

MSC2010: 00A05.
Keywords: statistical mechanics, Kähler–Einstein metrics, propagation of chaos, Langvin equation.

dynamic version of the probabilistic approach in [Berman 2013a; 2017]. In other words, we are in the realm of nonequilibrium statistical mechanics, where the relaxation to equilibrium is studied. As the general complex geometric setting appears to be extremely challenging, due to the severe singularities and nonlinearity of the corresponding interaction energies, we will here focus on the real analog of the complex setting introduced in [Berman 2013b], taking place in $\mathbb{R}^n$ and which corresponds to the case when X is a *toric* complex algebraic variety. As explained in that paper in this real setting the determinantal (fermionic) processes are replaced by *permanental* (bosonic) processes and convexity plays the role of positive Ricci curvature/plurisubharmonicity (see Section 5C for some geometric background).

Our main result (Theorem 1.1) shows that a deterministic evolution equation on the space of all probability measures on $\mathbb{R}^n$ emerges from the underlying stochastic dynamics, which as explained below, can be seen as a new "propagation of chaos" result. The evolution equation in question is a drift-diffusion equation coupled to the fully nonlinear real Monge–Ampère operator. It turns out that in the case of the real line (i.e., $n = 1$) this equation is closely related to various extensively studied evolution equations, notably the Keller–Segel equation in chemotaxis [Keller and Segel 1970], Burger's equation [Hopf 1950; Frisch and Bec 2001] in the theory of nonlinear waves and scalar conservation laws and the deterministic version of the Kardar–Parisi–Zhang (KPZ) equation describing surface growth [Kardar et al. 1986]. In the higher-dimensional real case, the equation can be viewed as a dissipative viscous version of the semigeostrophic equation appearing in dynamic meteorology; see [Loeper 2006; Ambrosio et al. 2014]. Moreover, closely related evolution equations appear in cosmology and in particular in Brenier's approach to the Zeldovich model used in the early universe reconstruction problem [Shandarin and Zel'dovich 1989; Frisch et al. 2002; Brenier 2011; 2016].

As we were not able to deduce the type of propagation of chaos result we needed from previous general results and approaches, the main body of the paper establishes the appropriate propagation of chaos result, which, to the best of our knowledge, is new and hopefully the result, as well as the method of proof, is of independent interest. As will be clear below, our approach heavily relies on the theory of weak gradient flows on the Wasserstein L^2-space $\mathcal{P}_2(\mathbb{R}^n)$ of probability measure on $\mathbb{R}^n$ developed in the seminal work of Ambrosio, Gigli and Savaré [Ambrosio et al. 2005], which provides a rigorous framework for the Otto calculus [2001]. In particular, just as in [Ambrosio et al. 2005], convexity (or more generally λ-convexity) plays a prominent role. Our limiting evolution equation will appear as the gradient flow on $\mathcal{P}_2(\mathbb{R}^n)$ of a certain free-energy-type functional F. Interestingly, as observed in [Berman 2013c; Berman and Berndtsson 2013] the functional F may be identified with Mabuchi's K-energy functional on the space of Kähler metrics, which plays a key role in Kähler geometry. The gradient flows of F with respect to other metrics (the Mabuchi–Donaldson–Semmes metric and Calabi's gradient metric) are the renowned Calabi flow and Kähler–Ricci flow respectively [Chen and Zheng 2013]. The regularity and large-time properties of the evolution equation appearing here will be studied elsewhere [Berman and Lu $\geq$ 2018; Berman $\geq$ 2018].

In the remaining part of the introduction we will state our main results: first a general propagation of chaos result, assuming a uniform Lipschitz and convexity assumption on the interaction energy, and then the application to permanental point processes and toric Kähler–Einstein metrics. In the companion paper

[Berman and Önnheim 2016] we give a more general formulation of the propagation of chaos result, by relaxing some of the assumptions (in particular, this yields sharp convergence results for strongly singular repulsive pair interactions when $n = 1$).

1A. ***Propagation of chaos and Wasserstein gradient flows.*** Consider a system of N identical particles diffusing on the n-dimensional Euclidean space

$$X := \mathbb{R}^n$$

and interacting by a symmetric energy function $E^{(N)}(x_1, x_2, \dots, x_N)$, at a fixed inverse temperature β. According to nonequilibrium statistical mechanics, the distribution of particles at time t is described by the following system of stochastic differential equations (SDEs), under suitable regularity assumptions on $E^{(N)}$:

$$dx_i(t) = -\frac{\partial}{\partial x_i} E^{(N)}(x_1, x_2, \dots, x_N)\,dt + \sqrt{\frac{2}{\beta}}\,dB_i(t), \tag{1-1}$$

where B_i denotes N independent Brownian motions on $\mathbb{R}^n$; the equation is called the (overdamped) *Langevin equation* in the physics literature [Schwabl 2002, Section 8.1.2]. In other words, this is the Itô diffusion on $\mathbb{R}^n$ describing the (downward) gradient flow of the function $E^{(N)}$ on the configuration space X^N perturbed by a noise term. A classical problem in mathematical physics going back to Boltzmann and made precise by Kac [1956] is to show that, in the many-particle limit where $N \to \infty$, a *deterministic* macroscopic evolution emerges from the stochastic microscopic dynamics described by (1-1). More precisely, denoting by δ_N the empirical measures

$$\delta_N := \frac{1}{N} \sum \delta_{x_i}, \tag{1-2}$$

the SDEs (1-1) define a curve $\delta_N(t)$ of random measures on X. The problem is to show that, if at the initial time $t = 0$ the random variables x_i are independent with identical distribution μ_0, then the empirical measure $\delta_N(t)$ converges in law to a curve μ_t of measures on $\mathbb{R}^n$,

$$\lim_{N\to\infty} \delta_N(t) = \mu_t \tag{1-3}$$

at any time $t > 0$. In the terminology of [Kac 1956], see also [Sznitman 1991], this means that *propagation of chaos* holds. It should be stressed that the previous statement admits a pure PDE formulation, not involving any stochastic calculus (see Section 2C) and it is this analytic point of view that we will adopt here.[1]

Of course, if propagation of chaos is to hold then some consistency assumptions have to be made on the sequence $E^{(N)}$ of energy functions as N tends to infinity. The standard assumption in the literature ensuring that propagation of chaos does hold is that $E^{(N)}(x_1, x_2, \dots, x_N)$ can be as written as

$$E^{(N)}(x_1, x_2, \dots, x_N) = N E(\delta_N) \tag{1-4}$$

for a fixed functional E on the space of $\mathcal{P}(X)$ of all probability measures on X, where E is assumed to have appropriate regularity properties (to be detailed below). This is sometimes called a mean field model.

[1] From a differential geometric point of view the SDEs (1-1) correspond, under the transformation $\mu \mapsto e^{E/2}\mu$, to the heat flow on X^N of the Witten Laplacian of the "Morse function" E, but we will not elaborate on this point here.

By the results in [Braun and Hepp 1977; Sznitman 1991; Dawson and Gärtner 1987; Mischler et al. 2015], it then follows that the limit $\mu_t(=\rho_t\,dx)$ with initial data $\mu_0(=\rho_0\,dx)$ is uniquely determined and satisfies an explicit nonlinear evolution equation on $\mathcal{P}(X)$ of the form

$$\frac{d\rho_t}{dt}=\frac{1}{\beta}\Delta\rho_t-\nabla\cdot(\rho_t b[\rho_t]), \tag{1-5}$$

where we have identified $\mu(=\rho\,dx)$ with its density ρ and $b[\mu]$ is a function on $\mathcal{P}(X)$ taking values in the space of vector fields on X:

$$b[\mu]=-\nabla(dE_{|\mu}), \tag{1-6}$$

where the differential $dE_{|\mu}$ at μ is identified with a function on X, by standard duality (the alternative suggestive notation $b[\rho]=-\nabla(\partial E(\rho)/\partial\rho)$ is often used in the literature). In the kinetic theory literature drift-diffusion equations of the form (1-6) are usually called *McKean–Vlasov equation*s [McKean 1966; 1967]. More generally, the results referred to above hold in the more general setting where the gradient vector field $-(\partial/\partial x_i)E^{(N)}(x_1,x_2,\dots,x_N)$ on X appearing in (1-1) is replaced by $b[\delta_N]$ for a given function $b[\mu]$ on $\mathcal{P}(X)$, taking values in the space of vector fields and satisfying appropriate continuity properties.

One of the main aims of the present work is to introduce a new approach to the propagation of chaos result (1-3) for the stochastic dynamics (1-1) which exploits the gradient structure of the equations in question and which applies under weaker assumptions than the previous results, referred to above. As indicated above, our main motivation for weakening the assumptions comes from the applications to toric Kähler–Einstein metrics described below. In that case $E^{(N)}$ satisfies the following assumptions, which will be referred to as the *main assumptions*: $E^{(N)}$ is uniformly Lipschitz continuous in each variable separately, i.e., there is a constant C such that

(MA1)
$$|\nabla_{x_i}E^{(N)}|\leq C, \tag{1-7}$$

and there exists a (finite) functional $E(\mu)$ on the Wasserstein space $\mathcal{P}(\mathbb{R}^n)$ such that

(MA2)
$$\frac{1}{N}E^{(N)}(x_1,x_2,\dots,x_N)=E(\delta_N)+o(1), \tag{1-8}$$

where $o(1)$ denotes a sequence of functionals on $\mathcal{P}(\mathbb{R}^n)$ converging pointwise to zero on $\mathcal{P}(\mathbb{R}^n)$ as $N\to\infty$. Moreover, $E^{(N)}$ is λ-*convex* on X^N for some real number λ, which means that the (distributional) Hessians are uniformly bounded from below on $\mathbb{R}^{nN}$,

(MA3)
$$(\nabla^2E^{(N)})\geq\lambda I, \tag{1-9}$$

where I denotes the identity matrix on $\mathbb{R}^{nN}$. This implies, in particular, that there exists a unique solution to the evolution equation (1-5) in the sense of weak gradient flows on the space $\mathcal{P}_2(\mathbb{R}^n)$ of all probability measures with finite second moments equipped with the Wasserstein L^2-metric [Ambrosio et al. 2005]:

$$\frac{d\mu_t}{dt}=-\nabla F_\beta(\mu_t),$$

where F_β is the free-energy-type functional corresponding to the macroscopic energy $E(\mu)$ at inverse temperature β,

$$F_\beta(\mu) = E(\mu) + \frac{1}{\beta} H(\mu),$$

and where $H(\mu)$ is the Boltzmann entropy of μ (see Section 2A for notation).

Theorem 1.1. *Let $E^{(N)}$ be a sequence of symmetric functions on $(\mathbb{R}^n)^N$ satisfying the main assumptions* (1-8), (1-7) *and* (1-9) *and consider the corresponding system of SDEs* (1-1). *If the initial data $x_i(0)$ consists of independent and identically distributed random vectors with law $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$, then, at any fixed positive time t, the corresponding empirical measures converge in law, as $N \to \infty$, to the measure $\mu_t \in \mathcal{P}_2(\mathbb{R}^n)$, where the curve $t \mapsto \mu_t$ is the gradient flow on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$ of the free energy functional F_β, emanating from μ_0.*

In fact, the convergence of the laws will be shown to hold in the L^2-Wasserstein topology. This leads to a strong form of propagation of chaos in the present setting (implying that the correlations between the random vectors $x_i(t)$ and $x_j(t)$ tend to zero as $N \to \infty$, if $i \neq j$; see Section 3B).

It should be stressed that the key point of our approach is that we do not need to assume that the drift $b[\mu](x)$ defined by (1-6) has any continuity properties with respect to μ or x, in contrast to previous work [Braun and Hepp 1977; Sznitman 1991; Dawson and Gärtner 1987; Mischler et al. 2015]. This will be crucial in the applications to toric Kähler–Einstein metrics, where the interaction energies $E^{(N)}$ are smooth and convex, but the norms of corresponding Hessians are not uniformly bounded in N (which is reflected in the fact that the corresponding function $x \mapsto b(\mu)(x)$ is not continuous for a general μ).

We recall that if the drift is assumed to have suitable continuity properties, then the existence of a solution to the drift-diffusion equation (1-5) can be established using fixed-point-type arguments [Sznitman 1991]. However, in our case we have, in general, to resort to the weak gradient flow solutions provided by the general theory in [Ambrosio et al. 2005], where the solution ρ_t can be characterized uniquely by a differential inequality called the *evolutionary variational inequality* (*EVI*). As shown in that work, the corresponding solution ρ_t satisfies the drift-diffusion equation (1-5) in a suitable weak sense, as follows formally from the Otto calculus [Otto 2001].

1A1. *Idea of the proof of Theorem 1.1 and comparison with previous results.* The starting point of the proof is the basic fact that the SDEs (1-1) on X^N admit a PDE formulation. Indeed, as recalled in Section 2C, they correspond to a linear evolution $\mu_N(t)$ of probability measures (or densities) on X^N, given by the corresponding forward Kolmogorov equation (also called the Fokker–Planck equation). Given this fact, our proof of Theorem 1.1 proceeds in a variational manner, building on [Jordan et al. 1998; Ambrosio et al. 2005] (and inspired by the approach introduced in [Messer and Spohn 1982] in the static setting of Gibbs measure): the rough idea is to show that any weak limit curve $\Gamma(t)$ of the laws

$$\Gamma_N(t) := (\delta_N)_* \mu_N(t) \in \mathcal{P}_2(Y), \quad Y = \mathcal{P}_2(\mathbb{R}^n),$$

is of the form $\Gamma(t) := \delta_{\mu_t}$, where the curve μ_t in $\mathcal{P}_2(\mathbb{R}^n)$ is uniquely determined by a "dynamic minimizing property". To this end we first discretize time, by fixing a small time mesh $\tau := t_{j+1} - t_j$, and replace, for any fixed N, the curve $\Gamma_N(t)$ with its discretized version $\Gamma_N^\tau(t_j)$, defined by a variational Euler scheme (a "minimizing movement" in De Giorgi's terminology) as in [Jordan et al. 1998; Ambrosio et al. 2005]. We then establish a discretized version of Theorem 1.1 saying that if, at a given discrete time t_j, the convergence

$$\lim_{N\to\infty} \Gamma_{t_j}^N = \delta_{\mu_{t_j}^\tau}$$

holds in the L^2-Wasserstein metric then the convergence also holds at the next time step t_{j+1} (using a variational argument). In particular, since, by assumption, the convergence above holds at the initial time 0 it "propagates" by induction to hold at any later discrete time. Finally, we prove Theorem 1.1 by letting the mesh τ tend to zero. This last step uses that the error estimates established in [Ambrosio et al. 2005], for discretization schemes as above, only depend on a uniform lower bound λ on the convexity of the interaction energies.

Our proof appears to be rather different from the probabilistic approaches in [Sznitman 1991; Dawson and Gärtner 1987], which are based on a study of nonlinear martingales, and the recent PDE approach in [Mischler et al. 2015]. The latter approaches require a two-sided uniform bound on the Hessian of the interaction energy $E^{(N)}$, while we only require a uniform lower bound.

It may also be illuminating to think about the convergence of $\Gamma_N(t)$ towards $\Gamma(t)$ as a kind of stability result for the sequence of weak gradient flows on $\mathcal{P}_2(Y)$, associated to the corresponding mean free energies, viewed as functionals on $\mathcal{P}_2(Y)$. This situation is somewhat similar to the stability result for gradient flows on $\mathcal{P}_2(H)$ in [Ambrosio et al. 2005; 2009], where H is a Hilbert space, but the main difference here is that the underlying space Y is not a Hilbert space, as opposed to the setting in those works, which prevents one from directly applying the error estimates in [Ambrosio et al. 2005] on the space $\mathcal{P}_2(Y)$ itself; this analog is expanded on in the companion paper [Berman and Önnheim 2016].

1B. ***Applications to permanental point processes at negative temperature and toric Kähler–Einstein metrics.*** Let P be a convex body in $\mathbb{R}^n$ containing zero in its interior and denote by $P_\mathbb{Z}$ the lattice points in P, i.e., the intersection of the convex body P with the integer lattice $\mathbb{Z}^n$. We fix an auxiliary ordering $p_1, \dots, p_N$ of the N elements of $P_\mathbb{Z}$. Given a configuration $(x_1, \dots, x_N)$ of N points on X, we denote by $\operatorname{Per}(x_1, \dots, x_N)$ the number defined as the permanent of the rank-N matrix with entries $A_{ij} := e^{x_i \cdot p_j}$:

$$\operatorname{Per}(x_1, \dots, x_N) := \operatorname{Per}(e^{x_i \cdot p_j}) = \sum_{\sigma \in S_N} e^{x_1 \cdot p_{\sigma(1)} + \dots + x_N \cdot p_{\sigma(N)}}, \tag{1-10}$$

where S_N denotes the symmetric group on N letters. This defines a symmetric function on $\mathbb{R}^{nN}$ which is canonically attached to P (i.e., it is independent of the choice of ordering of $P_\mathbb{Z}$).[2] We will consider the large-N limit which appears when P is replaced by the sequence kP of scaled convex bodies for any positive integer k. In particular, N depends on k as

$$N_k = \frac{k^n V(P)}{n!} + o(k^n),$$

[2] In many body quantum mechanics $\operatorname{Per}(x_1, \dots, x_N)$ appears as the N-particle wave function for a bosonic system of N particles represented by the N wave functions $e^{x \cdot p_j}$, i.e., N planar waves with imaginary momenta proportional to p_j.

where $V(P)$ denotes the Euclidean volume of P. In this setting the interaction energy is defined by

$$E^{(N_k)}(x_1, \dots, x_{N_k}) = \frac{1}{k} \log \operatorname{Per}(x_1, \dots, x_{N_k}). \tag{1-11}$$

To simplify the notation we will often drop the explicit dependence of N on k.

By the results in [Berman 2013b], the assumptions in Theorem 1.1 hold with

$$E(\mu) := -C(\mu),$$

where $C(\mu)$ is the Monge–Kantorovich optimal cost for transporting μ to the uniform probability measure ν_P on the convex body P, with respect to the standard symmetric quadratic cost function $c(x, p) = -x \cdot p$. Hence, the corresponding free-energy functional may be written as

$$F_\beta(\mu) = -C(\mu) + \frac{1}{\beta} H(\mu). \tag{1-12}$$

Theorem 1.2. *Assume that $\beta > 0$ and consider the system of SDEs* (1-1), *defined by the interaction energy* (1-11). *If the initial data $x_i(0)$ consists of independent and identically distributed random vectors with law $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$, then at any fixed positive time t, the corresponding empirical measures converge in law, as $N \to \infty$, to the measure $\mu_t \in \mathcal{P}_2(\mathbb{R}^n)$, where the curve $t \mapsto \mu_t$ is the gradient flow on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$ of the free energy functional F_β* (1-12) *emanating from μ_0. The corresponding densities ρ_t on $\mathbb{R}^n$ satisfy the following evolution PDE in the distributional sense*:

$$\frac{\partial \rho_t}{\partial t} = \frac{1}{\beta} \Delta \rho_t + \nabla \cdot (\rho_t \nabla \phi_t), \tag{1-13}$$

where $\phi_t(x)$ is the "convex potential" of ρ_t, i.e., the convex function on $\mathbb{R}^n$ solving the Monge–Ampère equation

$$\frac{1}{V(P)} \det(\partial^2 \phi_t) = \rho_t \tag{1-14}$$

(*in the weak sense of Alexandrov*) *normalized so that $\phi(0) = 0$ and satisfying the growth condition*

$$\phi(x) \le \sup_{p \in P} p \cdot x \tag{1-15}$$

(*equivalently*, *the closure of the gradient image of ϕ is equal to P*).

Let us briefly explain how Theorem 1.2 provides a stochastic dynamic approach for constructing Kähler–Einstein metrics on toric varieties; details will appear in a separate publication [Berman $\geq$ 2018]. We first recall that the Kähler potential of such a metric can be identified with a convex function ϕ satisfying the Monge–Ampère equation on $\mathbb{R}^n$

$$\det(\partial^2 \phi) = e^{-\phi}, \tag{1-16}$$

subject to the growth condition (1-15), with P being the moment polytope of the toric variety; see [Wang and Zhu 2004; Berman and Berndtsson 2013]. Now, a direct computation reveals that the corresponding density $\rho := \det(\partial^2 \phi)$ (which may be identified with the volume form of the Kähler–Einstein metric) is a stationary solution of the evolution equation appearing in Theorem 1.2. This is consistent, as it must be, with the fact that the free-energy functional F_β (1-12) may be identified with Mabuchi's K-energy functional on the space of Kähler metrics (when $\beta = 1$), whose minimizers are precisely the Kähler–Einstein metrics [Berman 2013c; Berman and Berndtsson 2013]. As shown in [Berman ≥ 2018], this fact can be used to show that the solution ρ_t of the evolution equation appearing in Theorem 1.2 converges, when $t \to \infty$, to the volume form of a Kähler–Einstein metric on X_P, if such a metric exists, which, in turn, is equivalent to the vanishing of the barycenter of the polytope P [Wang and Zhu 2004; Berman and Berndtsson 2013]. As discussed in Section 5A this can be viewed as a generalization of well-known stability properties for scalar conservation laws. The upshot of all this is that letting first N and then t tend to infinity in the SDEs (1-1), corresponding to the interaction energy (1-11), produces a toric Kähler–Einstein metric, when such a metric exists.

As we point out in Sections 4D, 5C1 our results also apply to the *tropical* analog of the permanental setting above, which can be viewed as the tropicalization of the complex geometric setting on the corresponding toric variety. In the corresponding deterministic setting (i.e., $\beta_N = \infty$) the particles then perform zigzag paths in $\mathbb{R}^n$ generalizing the extensively studied sticky particle system on $\mathbb{R}$ [Weinan et al. 1996; Brenier and Grenier 1998; Natile and Savaré 2009]. This is closely related to the Zeldovich model for the formation of large-scale structures in cosmology; see [Frisch et al. 2002; Brenier 2011; 2016] (compare to the discussion in Section 5B).

There is also a static analog of Theorem 1.2 (formulated in terms of Gibbs measures), which yields a probabilistic tropical analog of the Yau–Tian–Donaldson conjecture on toric Fano varieties linking the existence problem for toric Kähler–Einstein metrics to a notion of stability. This result first appeared in a previous preprint version of the present paper on the arXiv, but it has been deferred to a separate publication to shorten the present paper.

1C. ***Generalizations of Theorem 1.1.*** Let us conclude this introduction by pointing out that Theorem 1.1 admits various generalizations, obtained by weakening the assumptions, which are developed in the companion paper [Berman and Önnheim 2016]:

- By rescaling $E^{(N)}$ we may as well allow the "inverse temperature" β appearing in the SDEs (1-1) to depend on N as long as

$$\beta_N \to \beta \in [0, \infty],$$

as $N \to \infty$. In particular, Theorem 1.1 also applies to $\beta = \infty$ where the evolution equation (1-5) becomes a pure transport equation (i.e., with no diffusion). However, the precise relation to weak solutions becomes much more subtle and is closely related to the notions of entropy solutions and viscosity solutions studied in the PDE literature [Lax 1973], as detailed in [Berman and Önnheim 2016]. In fact, one may even allow that $\beta_N = \infty$, where the corresponding convergence results yields a deterministic mean field particle approximation.

- The assumptions (MA1) and (MA2), formulas (1-7), (1-8), may be replaced by a uniform coercivity assumption on $E^{(N)}$ together with the assumption that the *mean energies* corresponding to $E^{(N)}$ converge as functionals on $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}))$, in a suitable sense (which is closely related to the notion of Γ-convergence). This ensures that a weaker form of Proposition 3.6 holds.

• The convexity assumption (MA3) on $E^{(N)}$ may be replaced by a generalized convexity property of the corresponding mean energy functional on $\mathcal{P}_2(\mathbb{R}^{nN})^{S_N}$.

1D. ***Organization.*** In Section 2 we start by recalling the general setup that we will need from probability, the theory of Wasserstein spaces and weak gradient flows and then turn to the proof of Theorem 1.1 in Section 3B (starting with the discretized situation). In Section 4 we go on to apply the previous general results to the permanental setting and its tropical analog. In the final section we provide on outlook on some relations to conservation laws, sticky-particle-type systems and complex geometry. The Appendix recalls the basics of the formal Otto calculus and is included to serve as a motivation for the material on Wasserstein gradient flows. The rather lengthy setup and preparatory material in Section 2 are due to our effort to make the paper readable to a rather general audience.

2. General setup and preliminaries

2A. ***Notation.*** Given a topological (Polish) space Y we will denote the integration pairing between measures μ on Y (always assumed to be Borel measures) and bounded continuous functions f by

$$\langle f, \mu\rangle := \int f\mu$$

(we will avoid the use of the symbol $d\mu$ since d will usually refer to a distance function on Y). In the case $Y = \mathbb{R}^D$ we will say that a measure μ *has a density,* denoted by ρ, if μ is absolutely continuous with respect to Lebesgue measure dx and $\mu = \rho\, dx$. We will denote by $\mathcal{P}(\mathbb{R}^D)$ the space of all probability measures and by $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^D)$ the subspace containing those with a density. The *Boltzmann entropy* $H(\rho)$ and *Fisher information* $I(\rho)$ (taking values in $]-\infty, \infty]$) are defined by

$$H(\rho) := \int_{\mathbb{R}^D} (\log\rho)\rho\, dx, \qquad I(\rho) = \int_{\mathbb{R}^D} \frac{|\nabla\rho|^2}{\rho}\, dx \tag{2-1}$$

(assuming that $\nabla\rho \in L^1(dx)$ and $\rho^{-1}\nabla\rho \in L^2(\rho dx)$). More generally, given a reference measure μ_0 on Y the entropy of a measure μ relative to μ_0 is defined by

$$H_{\mu_0}(\mu) = \int_{X^N} \left(\log\frac{\mu}{\mu_0}\right)\mu \tag{2-2}$$

if the probability measure μ on X is absolutely continuous with respect to μ_0 and otherwise $H(\mu) := \infty$. The relative Fisher information is defined similarly, by replacing ρ with the density μ/μ_0 in formula (2-1).

Given a lower semicontinuous (*lsc*, for short) function V on Y and $\beta \in]0, \infty]$ (the "inverse temperature") we will denote by F^V_β the corresponding (*Gibbs*) *free-energy functional with potential* V:

$$F^V_\beta(\mu) := \int_X V\mu + \frac{1}{\beta} H_{\mu_0}(\mu), \tag{2-3}$$

which coincides with $1/\beta$ times the entropy of μ relative to $e^{-V}\mu_0$.

2B. ***Wasserstein spaces and metrics.*** We start with the following very general setup. Let (X, d) be a given metric space, which is Polish, i.e., separable and complete, and denote by $\mathcal{P}(X)$ the space of all probability measures on X endowed with the *weak topology,* i.e., $\mu_j \to \mu$ weakly in $\mathcal{P}(X)$ if and only if $\int_X \mu_j f \to \int_X \mu f$ for any bounded continuous function f on X (this is also called the *narrow topology* in the probability literature). The metric d on X induces l^p-type metrics on the N-fold product X^N for any given $p \in [1, \infty[$:

$$d_p(x_1, \dots, x_N; y_1, \dots, y_N) := \left(\sum_{i=1}^N d(x_i, y_i)^p \right)^{1/p}.$$

The permutation group S_N on N letters has a standard action on X^N, defined by $(\sigma, (x_1, \dots, x_N)) \mapsto (x_{\sigma(1)}, \dots, x_{\sigma(N)})$ and we will denote by $X^{(N)}$ and π the corresponding quotient and quotient projection, respectively:

$$X^{(N)} := X^N / S^N, \quad \pi : X^N \to X^{(N)}. \tag{2-4}$$

The quotient $X^{(N)}$ may be naturally identified with the space of all configurations of N points on X. We will denote by $d_{(p)}$ the induced distance function on $X^{(N)}$, suitably normalized:

$$d_{(p)}(x_1, \dots, x_N; y_1, \dots, y_N) := \inf_{\sigma \in S_N} \left(\frac{1}{N} \sum_{i=1}^N d(x_i, y_{\sigma(i)})^p \right)^{1/p}.$$

The normalization factor $1/N^{1/p}$ ensures that the standard embedding of $X^{(N)}$ into the space $\mathcal{P}(X)$ of all probability measures on X,

$$X^{(N)} \hookrightarrow \mathcal{P}(X), \quad (x_1, \dots, x_N) \mapsto \delta_N := \frac{1}{N} \sum \delta_{x_i} \tag{2-5}$$

(where we will call δ_N the *empirical measure*), is an isometry when $\mathcal{P}(X)$ is equipped with the *L^p-Wasserstein metric d_{W^p}* induced by d (for simplicity we will also write $d_{W_p} = d_p$),

$$d_{W_p}^p(\mu, \nu) := \inf_\gamma \int_{X \times X} d(x, y)^p \gamma, \tag{2-6}$$

where γ ranges over all couplings between μ and ν; i.e., γ is a probability measure on $X \times X$ whose first and second marginals are equal to μ and ν respectively (see Lemma 2.3 below). We will denote by $W^p(X, d)$ the corresponding *L^p-Wasserstein space*, i.e., the subspace of $\mathcal{P}(X)$ consisting of all μ with finite p-th moments: for some (and hence any) $x_0 \in X$

$$\int_X d(x, x_0)^p \mu < \infty.$$

We will also write $W^p(X, d) = \mathcal{P}_p(X)$ when it is clear from the context which distance d on X is used.

Remark 2.1. In the terms of the Monge–Kantorovich theory of optimal transport [Villani 2003], $d_{W_p}^p(\mu, \nu)$ is the optimal cost for transporting μ to ν with respect to the cost functional $c(x, p) := d(x, y)^p$. Accordingly a coupling γ as above is often called a *transport plan* between μ and ν and it is said to be a *transport map* T if $\gamma = (I \times T)_* \mu$, where $T_* \mu = \nu$. In particular, if $X = \mathbb{R}^n$, $p = 2$ and μ and ν are absolutely continuous with respect to Lebesgue measure, then, by Brenier's theorem [1991], the optimal transport plan γ is always defined by a transport map $T(:= T_\mu^\nu)$ of the form $T_\mu^\nu = \nabla\phi$, where ϕ is a convex function on $\mathbb{R}^n$ (optimizing the dual Kantorovich functional).

We recall the following standard proposition:

Proposition 2.2. *A sequence μ_j converges to μ in the distance topology in $W^p(X,d)$ if and only if μ_j converges to μ weakly in $\mathcal{P}(X)$ and the p-th moments converge. As a consequence, if μ_j converges to μ weakly in $\mathcal{P}(X)$ and the p-th moments are uniformly bounded, i.e., for some $x_0 \in X$ there exists a constant C_0 such that*

$$\int_X d(x,x_0)^p \mu_j \leq C_0,$$

then μ_j converges to μ in the distance topology in $W^{p'}(X,d)$ for any $p' < p$.

Proof. For the first statement see for example [Villani 2003, Theorem 7.12]. The second statement is certainly also well known, but for completeness we include a simple proof. Fix $x_0 \in X$ and take the decomposition

$$\int_X d(x,x_0)^{p'} \mu_j = \int_{\{d(x,x_0)\leq R\}} d(x,x_0)^{p'} \mu_j + \int_{\{d(x,x_0)> R\}} d(x,x_0)^{p'} \mu_j.$$

Since $d(x,x_0)^{p'} \leq d(x,x_0)^p / R^{(p-p')}$ when $d(x,x_0) \geq R$, the second integral above is bounded from above by $C_0/R^{(p-p')}$. Moreover, by the assumption of weak convergence, the first term above converges to $\int_{\{d(x,x_0)\leq R\}} d(x,x_0)^{p'}\mu$, as $j \to \infty$. Finally, letting R tend to infinity concludes the proof. □

Since $Y_p := (W_p(X), d_{W_p})(:= \mathcal{P}_p(X))$ is also a Polish space we can iterate the previous construction and consider the Wasserstein space $W_q(Y) \subset \mathcal{P}(\mathcal{P}(X))$ that we will write as $W_q(\mathcal{P}_p(X))$, which is thus the space of all probability measures Γ on $\mathcal{P}(X)$ such that, for some $\mu_0 \in W_p(X)$,

$$\int_{\mathcal{P}(X)} d_p(\mu,\mu_0)^q \, \Gamma < \infty.$$

Lemma 2.3 (three isometries).

- *The empirical measure δ_N defines an isometric embedding $(X^{(N)}, d_{(p)}) \to \mathcal{P}_p(X)$.*
- *The corresponding push-forward map $(\delta_N)_*$ from $\mathcal{P}(X^{(N)})$ to $\mathcal{P}(\mathcal{P}(X))$ induces an isometric embedding between the corresponding Wasserstein spaces $W_q(X^{(N)}, d_{(p)})$ and $W_q(\mathcal{P}_p(X))$.*
- *The push-forward π_* of the quotient projection $\pi : X^N \to X^{(N)}$ induces an isometry between the subspace of symmetric measures in $W_q(X^N, (1/N^{1/p})d_p)$ and the space $W_q(X^{(N)}, d_{(p)})$.*

Proof. The first statement is a well-known consequence of the Birkhoff–Von Neumann theorem which gives that for any symmetric function $c(x,y)$ on $X \times X$ we have that if $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and $\nu = \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$ for given $(x_1,\dots,x_N), (y_1,\dots,y_N) \in X^N$, then

$$\inf_{\Gamma(\mu,\nu)} \int c(x,y)\,d\Gamma = \inf_{\Gamma_N(\mu,\nu)} \int c(x,y)\,d\Gamma,$$

where $\Gamma_N(\mu,\nu)\subset\Gamma(\mu,\nu)$ consists of couplings of the form $\Gamma_\sigma := \frac{1}{N}\sum \delta_{x_i}\otimes\delta_{y_{\sigma(i)}}$ for $\sigma\in S_N$, where S_N is the symmetric group on N letters. The second statement then follows from the following general fact: if $f:(Y_1,d_1)\to(Y_2,d_2)$ is an isometry between two metric spaces, then f_* gives an isometry between $W_q(Y_1,d_1)$ and $W_q(Y_2,d_2)$. This follows immediately from the definitions once one observes that one may assume that the coupling γ_2 between $f_*\mu$ and $f_*\nu$ is of the form $f_*\gamma_1$ for some coupling γ_1 between μ and ν. The point is that γ can be taken to be concentrated on $f(Y_1)\times f(Y_2)$ (since this set contains the product of the supports of μ and ν) and hence one can take $\gamma_1 := (f^{-1}\otimes f^{-1})_*\gamma_2$, where $(f^{-1}\otimes f^{-1})(f(y),f(y')) := (y,y')$ is well-defined, since f induces a bijection between Y_1 and $f(Y_1)$. Finally, the last statement follows immediately from the following general claim applied to $Y=X^N$ with $d=(1/N^{1/p})d_{X^N,l^p}$ and $G=S_N$. Let G be a compact group acting by isometries on a metric space (Y,d) and consider the natural projection $\pi: Y\to Y/G$. We denote by d_G the induced quotient metric on Y/G. The push-forward π_* gives a bijection between the space $\mathcal{P}(X)^G$ or all G-invariant probability measures on X and $\mathcal{P}(X/G)$. The claim is that π_* induces an isometry between the corresponding Wasserstein spaces $\mathcal{P}_q(X)^G$ and $\mathcal{P}_q(X/G)$; i.e., $d_{W_q}(\mu,\nu)=d_{W_q}(\pi_*\mu,\pi_*\nu)$ if μ and ν are G-invariant; see [Lott and Villani 2009, Lemma 5.36]. □

Let us also recall the following classical result

Lemma 2.4. *Let μ_0 be a probability measure on X. Then $(\delta_N)_*\mu_0^{\otimes N}\to\delta_{\mu_0}$ in $\mathcal{P}(\mathcal{P}(X))$ weakly as $N\to\infty$.*

In fact, according to Sanov's classical theorem the previous convergence result even holds in the sense of large deviations at speed N with rate functional given by the relative entropy functional $H_{\mu_0}(\,\cdot\,)$ [Dembo and Zeitouni 1993, Theorem 6.2.10].

2B1. *The present setting.* We will apply the previous setup to $X=\mathbb{R}^n$ endowed with the Euclidean metric d. Moreover, we will mainly use the case $p=2$. Then the corresponding metric d_2 on X^N is the Euclidean metric on $X^N=\mathbb{R}^{nN}$. Identifying a symmetric (i.e., S_N-invariant) probability measure μ_N on X^N with a probability measures on the quotient $X^{(N)}$ (as in Lemma 2.3) the second and third points in Lemma 2.3 may (with $q=2$) be summarized by the following chain of equalities that will be used repeatedly below:

$$\frac{1}{N}d_2(\mu_N,\mu'_N)^2 = d_{(2)}(\mu_N,\mu'_N)^2 = d_{W_2(\mathcal{P}_2(\mathbb{R}^n))}(\Gamma_N,\Gamma'_N)^2, \tag{2-7}$$

where Γ_N and Γ'_N denote the push-forwards under δ_N of μ_N and μ'_N respectively. To simplify the notation we will often simply write

$$d := d_{W_2(\mathcal{P}_2(\mathbb{R}^n))}$$

for the metric on $W_2(\mathcal{P}_2(\mathbb{R}^n))$ (or sometimes $d=d_2$).

2C. ***The forward Kolmogorov equation for the SDEs and the mean free energy $F_{N,\beta}$.*** Fix a positive integer N and $\beta>0$ (which may depend on N when we will later on let $N\to\infty$). Let (X,g) be a Riemannian manifold and denote by dV the volume form defined by g. In our case (X,g) will be the Euclidean space $\mathbb{R}^n$.

Consider the SDEs (1-1) on X^N with the initial condition that $x_i(0)$ are independent random variables with identical distribution $\mu_0 \in \mathcal{P}(\mathbb{R}^D)$. As is well known, under suitable regularity assumptions, this defines, for any fixed T, a probability measure η_T on the space of all continuous curves ("sample paths") in X^N, i.e., the space of continuous maps $[0, T] \to X^N$ [Stroock and Varadhan 1997, Chapter 5]. For t fixed we can thus view $x^{(N)}(t)$ as an X^N-valued random variable on the latter probability space. Then its law

$$\mu_t^{(N)} := (x^{(N)}(t))_* \eta_t$$

gives a curve of probability measures on $X^{(N)}$ of the form $\mu_t^{(N)} = \rho_t^{(N)} dV^{\otimes N}$, where the density $\rho_t^{(N)}$ satisfies the corresponding forward Kolmogorov equation

$$\frac{\partial \rho_t^{(N)}}{\partial t} = \frac{1}{\beta} \Delta \rho_t^{(N)} + \nabla \cdot (\rho_t^{(N)} \nabla E^{(N)}), \tag{2-8}$$

which thus coincides with the linear Fokker–Planck equation (A-6) on X^N with potential $V := E^{(N)}$. In this formula the initial conditions for the SDEs translates to

$$\mu_{|t=0}^{(N)} = \mu_0^{\otimes N}. \tag{2-9}$$

In particular, the law of the empirical measures $\delta_N(t)$ for the SDEs (1-1) can be written as the following probability measure on $\mathcal{P}(X)$:

$$\Gamma_N(t) := (\delta_N)_* \mu_t^{(N)},$$

where δ_N is the empirical measure defined by (2-5).

Anyway, for our purposes we may as well forget about the SDEs (1-1) and take the forward Kolmogorov equation (2-8) on X^N as our the starting point, together with the initial condition (2-9). We will exploit the well-known fact, going back to [Jordan et al. 1998] (see Theorem 2.14 below) that the latter evolution equation can be interpreted as the gradient flow on the Wasserstein space $\mathcal{P}_2(X^N)$ of the functional

$$F_\beta^{(N)}(\mu_N) = \int_{X^N} E^{(N)} \mu_N + \frac{1}{\beta} H(\mu_N),$$

where $H(\cdot)$ is the entropy relative to $\mu_0 := dV^{\otimes N}$, formula (2-2); occasionally we will omit the subscript β in the notation $F_\beta^{(N)}$.

Following standard terminology in statistical mechanics we will call the scaled functional $F_{N,\beta} := F_\beta^{(N)}/N$ the *mean free energy*, which is thus a sum of the *mean energy* $E_N (:= F_{N,\infty})$ and the *mean entropy* $H_N(\mu_N)$:

$$F_{N,\beta} = E_N + \frac{1}{\beta} H_N;$$

i.e.,

$$F_{N,\beta}(\mu_N) := \frac{1}{N} F_\beta^{(N)}(\mu_N) = \frac{1}{N} \int_{X^N} E^{(N)} \mu_N + \frac{1}{\beta N} H(\mu_N), \tag{2-10}$$

Note that it follows immediately from the definition that the mean entropy is additive: for any $\mu \in \mathcal{P}(X)$

$$H_N(\mu^{\otimes N}) = H(\mu).$$

In the case dV is a probability measure, it follows immediately from Jensen's inequality that $H(\mu) \geq 0$. In our Euclidean setting this is not the case but using that $\int e^{-\epsilon|x|^2}\,dx < \infty$ for any given $\epsilon > 0$ one then gets

$$H(\mu) \geq -\epsilon \int |x|^2 \mu - C_\epsilon. \tag{2-11}$$

As a consequence we have the following:

Lemma 2.5. *If the mean energy satisfies the uniform coercivity property*

$$\frac{1}{N}\int_{X^N} E^{(N)}(\mu_N) \geq -\frac{1}{2\tau_*} d_2(\mu_N, \Gamma_*)^2 - C \tag{2-12}$$

for some fixed $\tau_* > 0$ *and* $\Gamma_* \in W_2(\mathcal{P}(X))$ *and positive constant* C, *then so does* $F^{(N)}/N$.

Remark 2.6. The linear forward Kolmogorov equation (2-8) can also be viewed as the gradient flow of the *mean* free energy $\frac{1}{N}F^{(N)}$ if one instead uses the scaled metric $g_N := \frac{1}{N} g^{\otimes N}$ on X^N. Moreover, in our case, $E^{(N)}$ will be symmetric, i.e., S_N-invariant, and hence the flow defined with respect to (X^N, g_N) descends to the flow defined with respect to $X^{(N)} := X^N/S_N$ equipped with the distance function $d_{X^{(N)}}$ defined in Section 2B. Using the isometric embedding defined by the empirical measure (Lemma 2.3) we can thus view the sequence of flows on the sequence of spaces $\mathcal{P}(X^N)$ as a sequence of flows on the same (infinite-dimensional) space $W_2(\mathcal{P}(X))$ and this is the geometric motivation for the proof of Theorem 1.1.

2D. *Propagation of chaos and the L^2-Wasserstein topology.* First recall [Sznitman 1991] that a sequence $\mu^{(N)}$ of symmetric probability measures on X^N is said to be *chaotic* if there exists a probability measure μ on X such that, for any given finite number of functions $f_1, \dots, f_k$ in $C_b(X)$,

$$\lim_{N\to\infty} \int_{X^N} f_1(x_1)\cdots f_1(x_k)\mu^{(N)} = \int_X f_1\mu \cdots \int_X f_k\mu \tag{2-13}$$

(more precisely, then $\mu^{(N)}$ is called μ-*chaotic*).

Equivalently [Sznitman 1991, Proposition 2.2], this means that the empirical measure δ_N on the probability space $(X^N, \mu^{(N)})$ converges in law towards μ, i.e., the following convergence holds with respect to the weak topology in $\mathcal{P}(\mathcal{P}(X))$:

$$\lim_{N\to\infty} (\delta_N)_*\mu^{(N)} = \delta_\mu.$$

Now consider the system of SDEs (1-1) and assume that the initial random variables $x_1(0), \dots, x_N(0)$ are independent with identical law μ_0. This means that the corresponding curve of probability measures $\mu^{(N)}(t)$ on X^N (evolving by the forward Kolmogorov equation corresponding to the SDEs) is given by $\mu_0^{\otimes N}$ when $t = 0$ (i.e., the initial condition (2-9) holds). In particular, $\mu^{(N)}(t)$ is μ_0-chaotic when $t = 0$ (by Lemma 2.4). In the terminology introduced by Kac, *propagation of chaos* is said to hold if the sequence $\mu^{(N)}(t)$ remains chaotic for any positive time t, i.e., if there exists a curve $\mu(t)$ in $\mathcal{P}(X)$ emanating from μ_0 such that the sequence $\mu^{(N)}(t)$ is $\mu(t)$-chaotic for any $t \geq 0$.

In the present setting of Theorem 1.1 we will establish propagation of chaos in a stronger sense. Namely, we will show that if $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$, then $(\delta_N)_*\mu^{(N)}(t)$ converges to $\delta_{\mu(t)}$ in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^n))$, with respect to the topology defined by the Wasserstein L^2-metric. This is stronger than propagation of chaos since it also implies that the correlations between the random variables x_1 and x_2 on $((\mathbb{R}^n)^N, \mu_N)$ tend to zero as $N \to \infty$ (by symmetry this equivalently means that the correlations between x_i and x_j tend to zero, if $i \neq j$). This is made precise by the following lemma, where $(x)_\alpha$ denotes the α-th component of a vector $x \in \mathbb{R}^n$:

Lemma 2.7. *Let $\mu^{(N)}$ be a sequence of symmetric probability measures on $(\mathbb{R}^n)^N$ such that $(\delta_N)_*\mu^{(N)}(t)$ converges to δ_μ in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^n))$, with respect to the topology defined by the Wasserstein L^2-metric. Then $\mu^{(N)}$ is μ-chaotic and moreover, for any given $(\alpha_1, \alpha_2) \in \{1, \dots, n\}^2$*

$$\lim_{N\to\infty} \big(\mathbb{E}_N((x_1)_{\alpha_1}(x_2)_{\alpha_2}) - \mathbb{E}_N((x_1)_{\alpha_1})\mathbb{E}_N((x_2)_{\alpha_2})\big) = 0,$$

where $\mathbb{E}_N$ denotes the expectation with respect to $\mu^{(N)}$.

Proof. This follows readily from the definitions, but for completeness we provide a proof. By assumption the probability measures $\Gamma_N := (\delta_N)_*\mu^{(N)}$ converge to δ_μ in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^n))$ in the Wasserstein L^2-metric. Using Proposition 2.2 this convergence is equivalent to having

$$\lim_{N\to\infty} \int \Gamma_N \Phi = \Phi(\mu) \tag{2-14}$$

for any continuous function Φ on $\mathcal{P}_2(\mathbb{R}^n)$ of subquadratic growth, i.e., $\Phi(\nu) \leq C_0 d^2_{W_2}(\nu, \nu_0) + C_0$ for a fixed element $\nu_0 \in \mathcal{P}_2(\mathbb{R}^n)$. Taking $\nu_0 = \delta_0$, the latter growth condition means that

$$\Phi(\nu) \leq C \int |x|^2 \nu + C \tag{2-15}$$

for some constant C. In particular, setting $\Phi(\nu) := \int f_1 \nu \cdots \int f_k \nu$ for given bounded continuous functions $f_1, \dots, f_k$ and expanding reveals that (2-13) holds, showing that propagation of chaos holds. At this point we have only used the convergence of Γ_N towards δ_μ in the weak topology, just as in the proof of one direction of [Sznitman 1991, Proposition 2.2]. But taking $\Phi(\nu) = \int (x)_{\alpha_1} \nu \int (x)_{\alpha_2} \nu$ (which satisfies (2-15), using Hölder's inequality) gives

$$\begin{aligned}\int \Gamma_N \Phi &= N^{-2} \sum_{i,j \leq N} \int (x_i)_{\alpha_1}(x_j)_{\alpha_2} \mu^{(N)} \\ &= N^{-2}(N^2 - N) \int (x_1)_{\alpha_1}(x_2)_{\alpha_2} \mu^{(N)} + N^{-1} \int N^{-1} \sum_{i \leq N} (x_i)_{\alpha_1}(x_i)_{\alpha_2} \mu^{(N)}.\end{aligned}$$

Hence, letting $N \to \infty$ and using the convergence in (2-14) gives

$$\lim_{N\to\infty} \int (x_1)_{\alpha_1}(x_2)_{\alpha_2} \mu^{(N)} + 0 = \int (x)_{\alpha_1} \mu \int (x)_{\alpha_2} \mu.$$

Finally, applying the convergence in (2-14) to $\Phi(\nu) = \int (x)_\alpha \nu$ and using symmetry reveals that the right-hand side above is equal to the limit of $\mathbb{E}_N((x_1)_{\alpha_1})$ times $\mathbb{E}_N((x_2)_{\alpha_2})$ as $N \to \infty$. □

2E. *Gradient flows on the L^2-Wasserstein space and variational discretizations.* In this section we will recall the fundamental results from [Ambrosio et al. 2005] that we will rely on. Let G be a lower semicontinuous function on a complete metric space (M, d). In this generality there are, as explained in that work, various notions of weak gradient flows u_t for G (or "steepest descents") emanating from an initial point u_0 in M, symbolically written as

$$\frac{du_t}{dt} = -\nabla G(u_t), \quad \lim_{t\to 0} u(t) = u_0. \tag{2-16}$$

The strongest forms of weak gradient flows on metric spaces discussed in [Ambrosio et al. 2005] concern λ-convex functionals G and are defined by the property that u_t satisfies the *evolution variational inequalities* (*EVI*)

$$\begin{gathered}\frac{1}{2}\frac{d}{dt}d^2(u_t, v) + G(u(t)) + \frac{\lambda}{2}d^2(\mu_t, \nu)^2 \leq G(v) \quad \text{a.e. } t > 0, \\ \text{for all } v \in M, \quad G(v) < \infty,\end{gathered} \tag{2-17}$$

together with the initial condition $\lim_{t\to 0} u(t) = u_0$ in (M, d). Then u_t is uniquely determined by u_0, as shown in [Ambrosio et al. 2005, Corollary 4.3.3], and we shall say that u_t is the *EVI-gradient flow* of G emanating from u_0. We recall that λ-convexity on a metric space essentially means that the distributional second derivatives are bounded from below by λ along any geodesic segment in M (compare to below). When M has nonpositive curvature, NPC, (in the sense of Alexandrov) the existence of a solution u_t satisfying the EVI was shown by Mayer [1998] for any lower-semicontinuous λ-convex functional, by mimicking the Crandall–Liggett technique in the Hilbert-space setting.

However, in our case (M, d) will be the L^2-Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$ for the space of all probability measures μ on $\mathbb{R}^d$, which does not have nonpositive curvature (when $d > 1$). Still, as shown in [Ambrosio et al. 2005], the analog of Meyer's result does hold under the stronger assumption that G be λ-convex along any *generalized* geodesic μ_s in $\mathcal{P}_2(\mathbb{R}^d)$. For our purposes it will be enough to consider lsc λ-convex functionals G with the property that $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ is weakly dense in $\{G < \infty\}$. Then the λ-convexity of G means, compare to [Ambrosio et al. 2005, Proposition 9.210], that for any generalized geodesic $\mu_s = \rho_s\, dx$ in $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ the function $G(\rho_s)$ is continuous on $[0, 1]$ and the distributional second derivatives on $]0, 1[$ satisfy

$$\frac{d^2 G(\rho_s)}{d^2 s} \geq \lambda.$$

We recall that a *generalized geodesic* μ_s connecting μ_0 and μ_1 in $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ is determined by specifying a "base measure" $\nu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$. Then μ_s is defined as the following family of push-forwards:

$$\mu_s = ((1-s)T_0 + sT_1)_*\nu,$$

where T_i is the optimal transport map (defined with respect to the cost function $|x - y|^2/2$) pushing forward ν to μ_i (compare to Remark 2.1).

Remark 2.8. The bona fide Wasserstein geodesics in $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ are obtained by taking $\nu = \mu_0$ (the study of convexity along such geodesics was introduced by McCann [1997], who called it displacement convexity). But as shown in [Ambrosio et al. 2005], the point of working with general base measures ν is that they can be adapted to the discrete variational scheme for constructing EVI-gradient flows by taking $\nu = \mu_{t_j}$ at the j-th time step (compare to Section 2E1).

We will be relying on the following version of Theorems 4.0.4 and 11.2.1 in [Ambrosio et al. 2005]:

Theorem 2.9. *Suppose that G is an lsc real-valued functional on $\mathcal{P}_2(\mathbb{R}^d)$ which is λ-convex along generalized geodesics and satisfies the following coercivity property: there exist constants $\tau_*, C > 0$ and $\mu_* \in \mathcal{P}_2(\mathbb{R}^d)$ such that*

$$G(\cdot) \geq -\frac{1}{\tau_*} d_2(\cdot, \mu_*)^2 - C. \tag{2-18}$$

Then there is a unique solution μ_t to the EVI-gradient flow of G, emanating from any given $\mu_0 \in \overline{\{G < \infty\}}$. The flow has the following regularizing effect: $\mu_t \in \{|\partial G| < \infty\} \subset \{G < \infty\}$. Moreover, $G(\mu_t)$ and $e^{\lambda t}|\partial G|^2(\mu_t)$ are decreasing, where $|\partial G|$ denotes the metric slope of G:

$$|\partial G|(\mu) := \limsup_{\nu \to \mu} \frac{(G(\nu) - G(\mu))^+}{d(\mu, \nu)}.$$

Remark 2.10. Many more properties of the EVI-gradient flow μ_t are established in [Ambrosio et al. 2005]. For example, μ_t defines an absolutely continuous curve $\mathbb{R} \to \mathcal{P}_2(\mathbb{R}^n)$ (in the sense of metric spaces) which is locally Lipschitz continuous on $]0, \infty[$, which is a λ-contracting semigroup. Moreover, the flows are stable under suitable approximation of the initial data and the functional G.

Under suitably regularity assumptions it shown in [Ambrosio et al. 2005] that the EVI-gradient flow $\mu_t = \rho_t \, dx$ furnished by the previous theorem satisfies Otto's evolution equation (recalled in the Appendix) in the weak sense:

Proposition 2.11. *Suppose in addition to the assumptions in the previous theorem that μ_t has a density ρ_t for $t > 0$. Then ρ_t satisfies the continuity equation* (A-5) *in the sense of distributions on $\mathbb{R}^d \times \mathbb{R}$ with*

$$v_t = -(\partial^0 G)(\rho_t \, dx),$$

where $\partial^0 G$ denotes the minimal subdifferential of G.

We recall that under the assumptions in the previous theorem (and assuming $\{|\partial G|^2 < \infty\} \subset \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$) the many-valued *subdifferential* ∂G on the subspace $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$ is a metric generalization of the (Fréchet) subdifferential Hilbert space theory; by definition, it satisfies a "slope inequality along geodesics":

$$(\partial G)(\mu) := \left\{ \xi \in L^2(\mu) : G(\nu) \geq G(\mu) + \langle \xi, T_\mu^\nu(x) - x \rangle_{L^2(\mu)} + \tfrac{1}{2}\lambda d_2(\nu, \mu)^2 \text{ for all } \nu \right\},$$

where T_μ^ν denotes the optimal transport map between μ and ν, as in Remark 2.1. The *minimal subdifferential* $\partial^0 G$ on $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$ at μ is defined as the unique element in the subdifferential ∂G at μ minimizing the L^2-norm in $L^2(\mu)$; in fact, its norm coincides with the metric slope of G at μ. In [Ambrosio et al. 2005] there is also a more general notion of extended subdifferential which, however, will not be needed for our purposes.

Example 2.12. In the case when $G = H$ is the Boltzmann entropy and μ satisfies $H(\mu) < \infty$, so that μ has a density ρ, we have $(\partial^0 H)(\mu) = \rho^{-1}\nabla\rho \in L^2(\mu)$ and hence

$$|\partial H|^2(\mu) = I(\rho)$$

is the Fisher information of ρ (2-1); see [Ambrosio et al. 2005, Theorem 10.4.17].

The following result goes back to [McCann 1997]; see also [Ambrosio et al. 2005] for various elaborations:

Lemma 2.13. *The following functionals are lsc and λ-convex along any generalized geodesics in $\mathcal{P}_2(\mathbb{R}^d)$:*

- *The "potential energy" functional $\mathcal{V}(\mu) := \int V\mu$, defined by a given lsc λ-convex and lsc function V on $\mathbb{R}^d$ (and the converse also holds).*
- *The functional $\mu \mapsto \int V_N \mu^{\otimes N}$, defined by a given λ-convex function V_N on $\mathbb{R}^{dN}$.*
- *The Boltzmann entropy $H(\mu)$ (relative to dx).*

In particular, for any λ-convex function V on $\mathbb{R}^d$ the corresponding free energy functional F_β^V (2-3), is lsc and λ-convex along generalized geodesics if $\beta \in \,]0, \infty]$.

Combining the results above we arrive at the following

Theorem 2.14. *Assume given $\beta \in \,]0, \infty]$ and $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$. Let $E(\mu)$ be a lsc functional on $\mathcal{P}_2(\mathbb{R}^d)$ which is λ-convex along generalized geodesics and satisfies the coercivity condition* (2-18). *Denote by F_β the corresponding free energy functional, $F_\beta := E + H/\beta$. Then the EVI-gradient flow μ_t on $\mathcal{P}_2(\mathbb{R}^d)$, emanating from μ_0, of the functional F_β exists. Moreover, if $\beta < \infty$, then $\mu_t = \rho_t\, dx$, where ρ_t has finite Boltzmann entropy. In particular:*

- *If V is an lsc finite λ-convex function on $\mathbb{R}^d$, then the gradient flow of F_β^V exists, defining a weak solution of the corresponding forward Kolmogorov equation/Fokker–Planck equation* (2-8) *with initial condition* (2-9).
- *If moreover $E(\mu)$ is Lipschitz continuous on $\mathcal{P}_2(\mathbb{R}^d)$ and $\beta > 0$ then ρ_t has finite Boltzmann entropy and Fisher information and the following continuity equation holds in the distributional sense on $\mathbb{R}^n \times \mathbb{R}$:*
$$\frac{\partial \rho_t}{\partial t} = \frac{1}{\beta}\Delta\rho_t + \nabla(\rho_t v_t), \tag{2-19}$$
where $v_t = \partial^0 E$ is the minimal subdifferential of E at $\mu_t = \rho_t\, dx$.

Proof. By the previous lemma, F_β is also lsc and λ-convex and by Lemma 2.5 it also satisfies the coercivity condition. Hence, the EVI-gradient flow exists according to Theorem 2.9. Moreover, by the general results in [Ambrosio et al. 2005] F_β is decreasing along the flow and, in particular, locally uniformly bounded from above on $]0, \infty[$. But, by the coercivity assumption $E > -\infty$ on $\mathcal{P}_2(\mathbb{R}^d)$ and hence it follows that $H(\mu_t) < \infty$. The second statement then follows by the previous lemma and the fact that the coercivity condition holds: by λ-convexity $f(x) := v(x) + \lambda|x|^2$ is convex and hence $f(x) \geq -C|x|$ for some constant C, proving coercivity of v. To prove the last point first observe that

$E(\mu) \geq -A - Bd(\mu, \mu_0)^2 < \infty$ on $\mathcal{P}_2(\mathbb{R}^n)$ by the Lipschitz assumption. Since $F_\beta(\mu_t) \leq C$ it follows that $H(\mu_t) < \infty$, which in particular implies that μ_t has a density ρ_t. Moreover, by Theorem 2.9 $|\partial F_\beta(\mu_t)| < \infty$ for $t > 0$. But since E is assumed Lipschitz continuous we have $|\partial F_\beta(\mu_t)| < \infty$ if and only if $|\partial H(\mu_t)| < \infty$, which means that $I(\mu_t)$ has finite Fisher information (see Example 2.12). Finally, the distributional equation follows from Proposition 2.11. □

2E1. *The variational discretization scheme* (*minimizing movements*). Recall that the proof of Theorem 2.9 in [Ambrosio et al. 2005] uses a discrete approximation scheme introduced by De Giorgi, called the *minimizing movement* scheme. It can be seen as a variational formulation of the (backward) Euler scheme. Consider the fixed time interval $[0, T]$ and fix a (small) positive number τ (the "time step"). In order to define the "discrete flow" u_j^τ corresponding to the sequence of discrete times $t_j := j\tau$, where $t_j \leq T$ with initial data u_0, one proceeds by iteration: given $u_j \in M := \mathcal{P}_2(\mathbb{R}^d)$ the next step u_{j+1} is obtained by minimizing the functional

$$u \mapsto \frac{1}{2\tau} d(u, u_j)^2 + G(u)$$

on M. The minimizer exists and is unique as long as $\tau \leq \tau_0$, where τ_0 only depends on λ and the constant τ_* appearing in the inequality (2-18). Next, one defines $u^\tau(t)$ for any $t \in [0, T]$ by setting $u^\tau(t_j) := u_j^\tau$ and demanding that $u^\tau(t)$ be constant on $]t_j, t_{j+1}[$ and right continuous; we are using a slightly different notation than the one in [Ambrosio et al. 2005, Chapter 2].

The curve u_t is then defined as the large-m limit of $u_t^{(m)}$ in (M, d); as shown in [Ambrosio et al. 2005] the limit indeed exists and satisfies the EVI (2-17) and is thus uniquely determined. More precisely, the following quantitative convergence result holds; see Theorem 4.07, formula 4.024, and Theorem 4.09 of that same work:

Theorem 2.15. *Let G be a functional on $\mathcal{P}_2(\mathbb{R}^n)$ satisfying the assumptions in Theorem 2.9 with $\lambda \geq 0$. Then*

$$d^2(u^\tau(t), u(t)) \leq \tfrac{1}{2}|\tau|^2 |\partial G|^2(u_0),$$

where $|\partial G|(u_0)$ denotes the metric slope of G at u_0. If G is only assumed to be λ-convex for some, possibly negative, λ then

$$d(u^\tau(t), u(t)) \leq C|\tau|(G(u_0) - \inf G)$$

for some constant C only depending on λ and T.

Remark 2.16. By the last paragraph on page 79 in [Ambrosio et al. 2005] even if $\lambda < 0$, one does not need a lower bound on $\inf G$ if one replaces $|\tau|$ with $|\tau|^{1/2}$, as long as u_0 is assumed to satisfy $G(u_0) < \infty$.

3. Proof of Theorem 1.1

3A. ***The main assumptions on the interaction energy $E^{(N)}$.*** Set $X = \mathbb{R}^n$ and denote by d the Euclidean distance function on X. Throughout the paper $E^{(N)}$ will denote a symmetric, i.e., S_N-invariant, sequence of functions on X^N and we will make the following *main assumptions:*

(MA1) The functional $E^{(N)}$ is Lipschitz continuous in each variable on (X, d), uniformly in N.

(MA2) There exists a finite functional $E(\mu)$ on $\mathcal{P}_2(X)$ such that

$$\left|N^{-1}E^{(N)}(x_1,\dots,x_N)-E(\delta_N)\right|\le\epsilon_N(\delta_N)$$

for a sequence of functionals ϵ_N on $\mathcal{P}_2(X)$, converging pointwise to zero.

(MA3) The sequence $E^{(N)}$ is λ-convex on (X^N,d), uniformly in N.

Lemma 3.1. *Assume that* (MA1) *holds. Then, under the embedding*

$$\delta_N : X^{(N)}\to\mathcal{P}_1(X)$$

the sequence $E^{(N)}/N$ admits an extension which is uniformly Lipschitz continuous on $(\mathcal{P}_1(X),d_1)$ (*and hence on* $(\mathcal{P}_2(X),d_2)$, *by Hölder's inequality*). *If moreover*, (MA2) *holds then the extended functionals converge pointwise on $\mathcal{P}(X)$ to the functional E, which thus defines a Lipschitz continuous functional on* $(\mathcal{P}_1(X),d_1)$ (*and hence on* $(\mathcal{P}_2(X),d_2)$).

Proof. If $E^{(N)}$ is Lipschitz continuous in each variable with Lipschitz constant L, then taking the decomposition

$$\begin{aligned}&E^{(N)}(x_1,\dots,x_N)-E^{(N)}(y_1,\dots,y_N)\\&=\bigl(E^{(N)}(x_1,x_2,\dots,x_N)-E^{(N)}(y_1,x_2,\dots,x_N)\bigr)+\dots+\bigl(E^{(N)}(y_1,\dots,y_{N-1},x_N)-E^{(N)}(y_1,\dots,y_N)\bigr),\end{aligned}$$

where the right-hand side consists of N terms, gives

$$N^{-1}\left|E^{(N)}(x_1,\dots,x_N)-E^{(N)}(y_1,\dots,y_N)\right|\le LN^{-1}\sum_{i=1}^{N}d(x_i,y_i).$$

Since, $E^{(N)}$ is assumed S_N-invariant we deduce that, for any given $\sigma\in S_N$,

$$N^{-1}\left|E^{(N)}(x_1,\dots,x_N)-E^{(N)}(y_1,\dots,y_N)\right|\le LN^{-1}\sum_{i=1}^{N}d(x_i,y_{\sigma(i)}).$$

Hence, taking the infimum over all $\sigma\in S_N$ shows that $E^{(N)}/N$ is Lipschitz continuous on $(X^{(N)},d_{(1)})$. By the isometry property in Lemma 2.3, this means that we can identify $E^{(N)}/N$ with a Lipschitz continuous function f_N on a subset F_N of $\mathcal{P}_1(X)$. The desired extension property now follows from the general fact that any Lipschitz continuous function f defined on a subset of a metric space Y admits a Lipschitz continuous extension to all of Y. For example, the extension (that we still denote by f) can be taken as an infimal convolution [Hiriart-Urruty 1980].

To prove the last statement in the lemma we assume that (MA2) holds. Taking $\mu_0:=\delta_{x_0}$ it follows that $f_N(\mu_0)\to E(\mu_0)$. By the Arzelà–Ascoli theorem this implies that there exists a Lipschitz continuous function f on $(\mathcal{P}_1(X),d_1)$ such that, after perhaps passing to a subsequence, $f_N\to f$ uniformly on compacts of $\mathcal{P}(X)$. By the assumption (MA2) we must have $f=E$ and hence the whole sequence f_N has to convergence to E, which is thus Lipschitz continuous. As a consequence, the sequence $\epsilon_N:=f-f_N$ is also uniformly Lipschitz continuous on $\mathcal{P}(X)$. Finally, fix $\mu\in\mathcal{P}_1(X)$ and take some sequence $\boldsymbol{x}_N$ in X^N such that $\delta_N(\boldsymbol{x}_N)\to\mu$ in $\mathcal{P}_1(X)$. Then, using the triangle inequality three times together with the uniform Lipschitz continuity of $E^{(N)}$, $\epsilon^{(N)}$ and E we have

$$|E(\mu) - N^{-1}E^{(N)}(\mu)| \leq |\epsilon_N(\mu)| + 3Ld_1(\delta_N(\boldsymbol{x}_N), \mu),$$

which, by the assumption (MA2) converges to zero, as desired (we have used the same notation E^N/N for the extended functional f_N). □

The next lemma verifies that the mean free energy functional (2-10) and the free energy functional $F_\beta(:= E + H/\beta)$, corresponding to the sequence $E^{(N)}$, satisfy the assumptions in Theorem 2.9:

Lemma 3.2. *If the main assumptions hold, then the following hold for any given $\beta \in]0, \infty]$:*

- *The mean free energy functional $N^{-1}F_\beta^{(N)}$ is λ-convex along generalized geodesics in $\mathcal{P}_2(X^{(N)}, d_{(2)})$ and satisfies the following uniform coercivity property: there exist constants $\tau_*, C > 0$*

$$N^{-1}F_\beta^{(N)} \geq -\frac{1}{\tau_*}, d_{(2)}(\,\cdot\,, \delta_{(0,\ldots,0)})^2 - C. \tag{3-1}$$

- *The free energy functional F_β is λ-convex along generalized geodesics in $\mathcal{P}_2(X)$ and satisfies*

$$F_\beta \geq -\frac{1}{\tau_*} d_2(\,\cdot\,, \delta_0)^2 - C. \tag{3-2}$$

Proof. The λ-convexity of $F_\beta^{(N)}$ follows directly from the assumption (MA3) combined with first and third points in Lemma 2.13. Moreover, (MA1) together with (MA2) implies (using Lemma 3.1) that there exists a constant C such that

$$N^{-1}E^{(N)}(x_1, \ldots, x_N) \geq -LN^{-1}\sum_{i=1}^{N}|x_i| - C.$$

Using Hölder's inequality and integrating over X^N gives the uniform coercivity property (3-1) when $\beta = \infty$. The general case then follows from Lemma 2.5. Next, since E is Lipschitz continuous (by the previous lemma) the inequality (3-2) also follows in a similar manner. All that remains is thus to check that $E(\mu)$ is λ-convex along generalized geodesics in $\mathcal{P}_2(X)$. To this end we note that $E(\mu)$ is the pointwise limit on $\mathcal{P}_2(X)$ of the functionals

$$\mu \mapsto \int N^{-1}E^{(N)}\mu^{\otimes N},$$

as follows from Proposition 3.6 below (applied to $\mu_N = \mu^{\otimes N}$). For any fixed N the functional above is λ-convex along generalized geodesics (by (MA3) combined with the second point in Lemma 2.13). Letting $N \to \infty$ thus reveals that E is indeed λ-convex along generalized geodesics. As a consequence, so is F_β for any $\beta \in]0, \infty]$ (by the third point in Lemma 2.13). □

3B. ***Propagation of chaos in the time-discretized setting.*** In this section we will formulate and prove a discretized version of Theorem 1.1, assuming that the main assumptions hold. Let $\mu_0^{(N)}$ be a given sequence of symmetric elements in $\mathcal{P}_2(X^N)$ and $\mu_0 \in \mathcal{P}(X)$ be an element. Given a (small) "time step" τ we denote by $\mu_{t_j}^{(N)}$ the discretized minimizing movement of the mean free energy functional $N^{-1}F_\beta^{(N)}$ on $\mathcal{P}_2(X^{(N)}, d_{(2)})$ (2-10) emanating from $\mu_0^{(N)}$ and by μ_{t_j} the discretized minimizing movement of the free energy functional $F_\beta(:= E + H/\beta)$ on $\mathcal{P}_2(X)$ emanating from μ_0. We recall that this means (see Section 2E1) that, given $\mu_{t_j}^{(N)} \in \mathcal{P}_2(X^N)$, the next measure $\mu_{t_{j+1}}^{(N)}$ is defined as the minimizer of the following functional on $\mathcal{P}(X^N)$:

$$\frac{1}{N} J_{j+1}^{(N)}(\,\cdot\,) := \frac{1}{2\tau}\frac{1}{N} d(\,\cdot\,, \mu_{t_j}^{(N)})^2 + \frac{1}{N} F_\beta^{(N)}(\,\cdot\,). \tag{3-3}$$

Similarly, given $\mu_{t_j} \in \mathcal{P}(X)$, the next measure $\mu_{t_{j+1}}$ is defined as the minimizer of the following functional on $\mathcal{P}(X)$:

$$J_{j+1}(\,\cdot\,) = \frac{1}{2\tau}\frac{1}{N} d(\,\cdot\,, \mu_{t_j})^2 + \frac{1}{N} F_\beta(\,\cdot\,)$$

The sequences $\mu_{t_j}^{(N)}$ and μ_{t_j} are well-defined according to Lemma 3.2 and Theorem 2.9 (or rather its proof using minimizing movements, recalled in Section 2E1). We note that the sequence $\mu_{t_j}^{(N)}$ may (by the third isometry property in Lemma 2.3) be identified with the minimizing movement of the mean free energy functional $F^{(N)}/N$ on $\mathcal{P}_2(X^{(N)}, d_{(2)})$, which in turn embeds isometrically to give a discrete flow $\Gamma_{t_j}^{(N)}$ in $W_2(\mathcal{P}_2(X), d_2)$.

Theorem 3.3. *Assume that at time t_j*

$$\lim_{N\to\infty} (\delta_N)_*\mu_{t_j}^{(N)} = \delta_{\mu_{t_j}}$$

in $W_2(\mathcal{P}_2(X), d_2)$. Then, at the next time step t_{j+1}

$$\lim_{N\to\infty} (\delta_N)_*\mu_{t_{j+1}}^{(N)} = \delta_{\mu_{t_{j+1}}}$$

in $W_2(\mathcal{P}_2(X), d_2)$. As a consequence, if $\mu_{t_j}^{(N)}$ is of the form $\mu_{t_j}^{(N)} = \mu_0^{\otimes N}$ when $t_j = 0$, then $(\delta_N)_*\mu_{t_j}^{(N)}$ converges to $\delta_{\mu_{t_j}}$ in $W_2(\mathcal{P}_2(X), d_2)$ for any t_j.

The last statement follows directly from induction using the first statement and the following basic observation:

$$\mu_0 \in \mathcal{P}_2(X) \quad \Longrightarrow \quad (\delta_N)_*\mu_0^{\otimes N} \to \delta_{\mu_0} \text{ in } W_2(\mathcal{P}_2(X), d_2). \tag{3-4}$$

Indeed, by Lemma 2.4 the convergence holds in $\mathcal{P}(\mathcal{P}(X))$. Moreover, setting $\Gamma_0 := \delta_{\delta_0}$ gives

$$\int_{\mathcal{P}(X)} d^2(\Gamma, \Gamma_0)(\delta_N)_*\mu_0^{\otimes N} = \int_X |x|^2\mu_0 = \int d^2(\Gamma_{\delta_{\mu_0}}, \Gamma_0)(\delta_N)_*\mu_0^{\otimes N}$$

and hence (3-4) follows from Proposition 2.2. Thus it will be enough to prove the first statement in the previous theorem.

3C. *Proof of Theorem 3.3.* We start with the following direct consequence of Proposition 2.2 combined with Lemma 2.3:

Lemma 3.4. *Let μ_N be a sequence of symmetric probability measures on X^N and denote by $\Gamma_N := (\delta_N)_*\mu_N$ the corresponding probability measures on $\mathcal{P}(X)$. Assume that the d_2-distance of Γ_N to a fixed element in the Wasserstein space $W_q(\mathcal{P}_2(X))$ is uniformly bounded from above. Then, after perhaps passing to a subsequence, there is a probability measure Γ in $W_2(\mathcal{P}_2(X))$ such that*

$$\lim_{N\to\infty} (\delta_N)_*\mu_N = \Gamma$$

in $W_1(\mathcal{P}_2(X)$.

We next recall the following well-known result about the asymptotics of the mean entropy, proved in [Robinson and Ruelle 1967]; see also Theorem 5.5 in [Hauray and Mischler 2014] for generalizations. The proof is based on the subadditivity properties of the entropy.

Proposition 3.5. *Let μ_N be a sequence of probability measures on X^N such that $(\delta_N)_*\mu_N$ converges weakly to $\Gamma \in \mathcal{P}(\mathcal{P}(X))$. Then*

$$\liminf_{N\to\infty} H^{(N)}(\mu_N) \geq \int_{\mathcal{P}(X)} H(\mu)\Gamma.$$

We will also use the following result, which generalizes a result in [Messer and Spohn 1982] concerning the case when E_N is quadratic:

Proposition 3.6. *Let $\mu^{(N)}$ be a sequence of probability measures on X^N such that $\Gamma_N := (\delta_N)_*\mu_N$ converges to Γ in $W_1(\mathcal{P}_2(X))$. If $E^{(N)}$ satisfies the assumptions* (MA1) *and* (MA2) *in the main assumptions* (*Section 3A*), *then*

$$\lim_{N\to\infty} \frac{1}{N}\int_{X^N} E^{(N)}\mu^{(N)} = \int_{\mathcal{P}(X)} E(\mu)\Gamma.$$

Proof. Recall that the L^1-Wasserstein distance d_{W_1} on $W_1(Y,d)$ admits the following dual representation (the Kantorovich–Rubinstein theorem [Villani 2003, page 34]):

$$d_{W_1}(\mu,\nu) = \sup_{u\in \mathrm{Lip}_1} \int u(\mu-\nu),$$

where u ranges over all Lipschitz continuous functions on Y with Lipschitz constant 1. In the present setting we take (Y,d) as $\mathcal{P}_2(X)$ endowed with the d_2-distance. By assumption

$$d_{W_1}(\Gamma_N,\Gamma) \to 0. \tag{3-5}$$

Using the empirical measure δ_N we can identify $N^{-1}E^{(N)}$ with a uniformly Lipschitz continuous sequence of functions on $(\mathcal{P}_2(X), d_2)$, which by the main assumptions converges pointwise to the Lipschitz continuous functional $E(\mu)$ (using Lemma 3.1 applied to $p=2$). Since $N^{-1}E^{(N)}$ is uniformly Lipschitz continuous we have

$$\lim_{N\to\infty}\int_{\mathcal{P}_2(X)} N^{-1}E^{(N)}(\Gamma_N-\Gamma) = 0$$

using the (3-5) combined with the dual representation of the L^1-Wasserstein distance. Hence, all that remains is to verify that

$$\lim_{N\to\infty}\int_{\mathcal{P}_2(X)} N^{-1}E^{(N)}\Gamma = \int_{\mathcal{P}_2(X)} E(\mu)\Gamma.$$

But this follows from the dominated convergence theorem. Indeed, $N^{-1}E^{(N)}$ converges pointwise to E on $\mathcal{P}_2(X)$ and (by the uniform Lipschitz property) is uniformly dominated by the function $A + Bd_2$, which is in $L^1(\Gamma)$, since $\Gamma \in W_1(\mathcal{P}_2(X))$. □

Next we turn to the asymptotics of the distances, establishing the following key property:

Proposition 3.7. *Assume that a sequence ν_N of symmetric probability measures on X^N satisfies*

$$\lim_{N\to\infty}(\delta_N)_*\nu_N=\delta_\nu$$

in the distance topology in $W_2(\mathcal{P}_2(X))$. Then any sequence μ_N such that $(\delta_N)_*\mu_N$ converges weakly to $\Gamma\in\mathcal{P}(\mathcal{P}(X))$ satisfies

$$\liminf_{N\to\infty}\frac{1}{N}d(\mu_N,\nu_N)^2\geq\int_{\mathcal{P}(X)}d(\mu,\nu)^2\Gamma(\mu)$$

and equality holds if and only if $(\delta_N)_*\mu_N$ converges to Γ in the distance topology in $W_2(\mathcal{P}_2(X))$.

Proof. Consider the isometry

$$\delta_N:(X^{(N)},d_{X^{(N)}})\hookrightarrow(\mathcal{P}(X),d_W),\quad (x_1,\dots,x_N)\mapsto\delta_N:=\frac{1}{N}\sum\delta_{x_i},$$

defined in terms of the L^2-distances. We equip the space $\mathcal{P}(\mathcal{P}(X))$ with the L^2-Wasserstein (pre)metric d induced from distance d_W on $\mathcal{P}(X)$; i.e., we consider the subspace $W_2(\mathcal{P}(X))$. By Lemma 2.3

$$\frac{1}{N}d(\mu_N,\nu_N)^2=d((\delta_N)_*\mu_N,(\delta_N)_*\nu_N)^2.$$

We now first assume that $(\delta_N)_*\mu_N$ converges to Γ in the d-distance topology in $W_2(\mathcal{P}_2(X))$. Then the "triangle inequality" for d immediately gives

$$\lim_{N\to\infty}d((\delta_N)_*\mu_N,(\delta_N)_*\nu_N)^2=d(\Gamma,\delta_\nu)^2.$$

Next we will use the following simple general fact for the Wasserstein distance on $\mathcal{P}(Y,d)$:

$$d(\mu,\delta_{y_0})^2=\int d(y,y_0)^2\mu(y),$$

which follows from the fact that the only coupling between μ and δ_{y_0} is the product $\mu\otimes\delta_{y_0}$. Applied to $Y=\mathcal{P}(X)$ this gives

$$d((\delta_N)_*\mu_N,\delta_\nu)^2=\int_{\mathcal{P}(X)}d(\mu,\nu)^2\,\Gamma(\mu)$$

which concludes the proof using that $d(\delta_\mu,\delta_\nu)=d(\mu,\nu)$ by the general fact above. More generally, if $(\delta_N)_*\mu_N$ is only assumed to converge to Γ weakly in $\mathcal{P}(\mathcal{P}(X))$, then the lower semicontinuity of the Wasserstein distance function with respect to the weak topology instead gives

$$\liminf_{N\to\infty}\frac{1}{N}d(\mu_N,\nu_N)^2\geq\int_{\mathcal{P}(X)}d(\mu,\nu)^2\,\Gamma(\mu).$$

Finally, if equality holds above, then, by the previous arguments,

$$\lim_{N\to\infty}\int_{\mu\in\mathcal{P}(X)}d(\mu,\nu)^2(\delta_N)_*\mu_N=\int_{\mathcal{P}(X)}d(\mu,\nu)^2\,\Gamma(\mu)$$

(i.e., the "second moments of $(\delta_N)_*\mu_N$ converge to the second moments of Γ) and then it follows from Proposition 2.2 that $(\delta_N)_*\mu_N$ converges to Γ in the distance topology in $W_2(\mathcal{P}(X))$. □

3C1. *Conclusion of the proof of Theorem 3.3.* Without loss of generality we may set $\beta = 1$ and we will thus drop the subindex β from the notations. We start by observing that for any fixed μ in $\mathcal{P}(X)$ we have, by the defining property of $\mu^{(N)}_{t_{j+1}}$, that

$$\frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) \leq \frac{1}{N} J^{(N)}_{j+1}(\mu^{\otimes N}),$$

where the right-hand side converges, by the propositions above, to $J_{j+1}(\mu)$ as $N \to \infty$, where $J_{j+1}(\mu) = \frac{1}{2\tau} d(\mu, \mu_j)^2 + F(\mu)$. In particular, taking $\mu = \mu_{j+1}$ gives

$$\limsup_{N\to\infty} \frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) \leq J_{j+1}(\mu_{j+1}), \tag{3-6}$$

where μ_{j+1} is the unique minimizer of J_{j+1}.

Next we consider the lower bound. By the minimizing property of $\mu^{(N)}_{t_{j+1}}$ we have a uniform control on the d_2-distance:

$$d_2\big((\delta_N)_*\mu^{(N)}_{t_{j+1}}, (\delta_N)_*\mu^{(N)}_{t_j}\big)^2 = \frac{1}{N} d_2(\mu^{(N)}_{t_{j+1}}, \mu^{(N)}_{t_j})^2 \leq C. \tag{3-7}$$

Indeed, the minimizing property together with the previous bound gives

$$\frac{1}{\tau} d_2\big((\delta_N)_*\mu^{(N)}_{t_{j+1}}, (\delta_N)_*\mu^{(N)}_{t_j}\big)^2 \leq C - \frac{1}{N} F^{(N)}(\mu^{(N)}_{t_{j+1}}).$$

Hence, the inequality (3-7) follows from the uniform coercivity property of $\frac{1}{N} F^{(N)}$, formula (3-1).

Now, it follows from the induction assumption and the triangle inequality for d that $\mu^{(N)}_{t_{j+1}}$ satisfies the assumptions of Proposition 2.2. Accordingly, we may, after passing to a subsequence, assume that $\mu_N := \mu^{(N)}_{t_{j+1}}$ converges as in Lemma 3.4, or more precisely that

$$(\delta_N)_*\mu^{(N)}_{t_{j+1}} \to \Gamma$$

in $W_1(\mathcal{P}_2(X))$ for some $\Gamma \in W_2(\mathcal{P}_2(X))$. It then follows from Propositions 3.5, 3.6 and 3.7 that

$$\liminf_{N\to\infty} \frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) \geq \int d\Gamma(\mu) J_{j+1}(\mu). \tag{3-8}$$

Combining the previous lower bound with the upper bound (3-6) and using that μ_{j+1} is the unique minimizer of J_{j+1} then forces $\Gamma = \delta_{\mu_{j+1}}$ and

$$\lim_{N\to\infty} \frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) = J_{j+1}(\mu). \tag{3-9}$$

But this means that

$$\lim_{N\to\infty} (\delta_N)_*\mu^{(N)}_{t_{j+1}} = \delta_{\mu_{t_{j+1}}}$$

weakly in $\mathcal{P}(X)$ and by the equality (3-9) that

$$\lim_{N\to\infty} d((\delta_N)_*\mu^{(N)}_{t_{j+1}}, \delta_{\mu_{t_{j+1}}}) = d(\delta_{\mu_{t_{j+1}}}, \delta_{\mu_j}).$$

But then it follows from Proposition 3.7 (applied to $\nu = \delta_{\mu_j}$) that $(\delta_N)_*\mu_N$ converges to Γ in the distance topology in $W_2(\mathcal{P}_2(X))$, as desired.

3D. ***Convergence in the nondiscrete setting: proof of Theorem 1.1.*** We recall that in the previous section we had fixed a time step τ. In this section we will emphasize the dependence on τ by setting

$$\Gamma_N^\tau(t) := (\delta_N)_*(\mu_{t_j}^{(N)}), \quad \Gamma^\tau(t) := \delta_{\mu_{t_j}}.$$

The assumptions in Theorem 1.1 imply, by the last statement in Theorem 3.3, that

$$\lim_{N\to\infty} d(\Gamma_N^\tau(t), \Gamma^\tau(t)) = 0 \tag{3-10}$$

in $W_2(\mathcal{P}(X))$. Next, set

$$\Gamma_N(t) := (\delta_N)_*(\mu_t^{(N)}), \quad \Gamma(t) := \delta_{\mu_t},$$

where $\mu_t^{(N)}$ and μ_t denote the EVI-gradient flows of $F_\beta^{(N)}$ and F_β, respectively (whose existence is a consequence of Theorem 2.9 combined with Lemma 3.2). Consider now a fixed time interval $[0, T]$. For any fixed $t \in]0, T[$ we then have, by the triangle inequality,

$$d(\Gamma_N(t), \Gamma(t)) \leq d(\Gamma_N(t), \Gamma_N^\tau(t)) + d(\Gamma(t), \Gamma^\tau(t)) + d(\Gamma_N^\tau(t), \Gamma^\tau(t)).$$

First assume, for simplicity, that the assumption (MA3) (Section 3A) holds with $\lambda \geq 0$. By the convexity properties in Lemma 3.2 and the isometry property in Lemma 2.3 we have, using Theorem 2.15, that $d(\Gamma_N(t), \Gamma_N^\tau(t)) \leq C\tau$ (uniformly in N) and $d(\Gamma(t), \Gamma^\tau(t)) \leq \tau C$. Hence, combining the previous two inequalities with the convergence (3-10) and letting first $N \to \infty$ and then $\tau \to 0$, gives $\lim_{N\to\infty} d(\Gamma_N(t), \Gamma(t)) = 0$, which proves Theorem 1.1 when $\lambda \geq 0$. Finally, in the case when $\lambda \leq 0$ the previous argument still applies, with the error $O(\tau)$ replaced by $\mathcal{O}(\tau^{1/2})$ according to Remark 2.16.

4. Permanental processes and toric Kähler–Einstein metrics

In this section we will deduce Theorem 1.2, stated in the Introduction, from Theorem 1.1, proved in the previous section.

4A. ***Permanental processes: setup.*** Let P be a convex body in $\mathbb{R}^n$ containing zero in its interior and denote by ν_P the corresponding uniform probability measure on P; i.e.,

$$\nu_P = \frac{1_P d\lambda}{V(P)},$$

where $d\lambda$ denotes Lebesgue measure and $V(P)$ is the Euclidean volume of P. Setting $P_k := P \cap (\mathbb{Z}/k)^n$, we let N_k be the number of points in P_k and fix an auxiliary ordering $p_1, \ldots, p_{N_k}$ of the N_k elements of P_k. Given a configuration $(x_1, \ldots, x_{N_k})$ of points on $X := \mathbb{R}^n$ we set

$$E^{(N_k)}(x_1, \ldots, x_{N_k}) := \frac{1}{k} \log \sum_{\sigma \in S_{N_k}} e^{k(x_1\cdot p_{\sigma(1)} + \cdots + x_N \cdot p_{\sigma(N_k)})}, \tag{4-1}$$

which, as explained in Section 1B, can be written as the scaled logarithm of a permanent. To simplify the notation we will often drop the subscript k and simply write $N_k = N$, since anyway $N \to \infty$ if and only if $k \to \infty$. We will denote by $C(\mu, \nu)$ the Monge–Kantorovich optimal cost for transport between the probability measures μ and ν, with respect to the standard symmetric quadratic cost function $c(x, p) = -x \cdot p$:

$$C(\mu, \nu) := \inf_{\gamma} - \int_{X\times X} x \cdot p\gamma, \tag{4-2}$$

where the γ ranges over all couplings (transport plans) between μ and ν (see Remark 2.1).

Proposition 4.1. *The main assumptions are satisfied for $E^{(N)}$ with $\lambda = 0$ and $E(\mu) = -C(\mu, \nu_P)$. Equivalently, formulated in terms of the Wasserstein L^2-distance*

$$E(\mu) = -\frac{1}{2} d_{W_2}(\mu, \nu_P)^2 + \frac{1}{2}\int x^2\, d\mu + c_P, \quad c_P := \frac{1}{2}\int p^2 \nu_P. \tag{4-3}$$

In particular, $-C(\cdot, \nu_P)$ is convex along generalized geodesics.

Proof. This follows essentially from the results in [Berman 2013b]. But for completeness we give a direct proof here:

Step 1: (MA2) holds. First observe that

$$N^{-1}|E^{(N)} - E^{(N)}_{\text{trop}}| \leq C\frac{N}{\log N}, \tag{4-4}$$

where $E^{(N)}_{\text{trop}}$ denotes the tropical analog of $E^{(N)}$ (see formula (4-8) below). Indeed, fixing $(x_1, \dots, x_N)$ and denoting by σ_0 the element in S_N maximizing $\sigma \mapsto e^{x_1 p_{\sigma(1)}+\cdots+x_1 p_{\sigma(1)}}$ we have

$$e^{kx_1\cdot p_{\sigma_0(1)}+\cdots+x_1\cdot p_{\sigma_0(1)}}(1+0+\cdots+0) \leq E^{(N)}(x_1, \dots, x_N) \leq N!\, e^{kx_1\cdot p_{\sigma_0(1)}+\cdots+kx_1\cdot p_{\sigma_0(1)}}.$$

Hence, taking the log and dividing by k proves the inequality (4-4), using that $k^{-1}N^{-1}\log N! \to \infty$ (by Stirling, since $N \sim k^n$). Next observe that

$$-N^{-1}E^{(N)}_{\text{trop}}(\boldsymbol{x}) = \frac{1}{2}d^2(\delta_N(\boldsymbol{x}), \delta_N(\boldsymbol{p})) - \frac{1}{2}\int_{\mathbb{R}^n} |x|^2 \delta_N(\boldsymbol{x}) - \frac{1}{2}\int_{\mathbb{R}^n} |p|^2 \delta_N(p). \tag{4-5}$$

Indeed, rewriting $x \cdot p = |x-p|^2/2 - |p|^2/2 - |p|^2/2$ reveals that $2N^{-1}E^{(N)}_{\text{trop}}(\boldsymbol{x})$ is equal to the distance between $\boldsymbol{x}$ and $\boldsymbol{p}$ in $(X^{(N)}, d_{(2)})$ minus two quadratic terms. Since $\delta_N^* d = d_{(2)}$, this proves (4-5). All in all this means that assumption (MA2) is satisfied with

$$2\epsilon_N(\mu) := \left|d^2(\mu, \delta_N(\boldsymbol{p})) - d^2(\mu, \nu_P)\right| + \left|\int_{\mathbb{R}^n} |p|^2(\nu_P - \delta_N(p))\right|.$$

Step 2: (MA1) and (MA3) hold. First recall the basic fact that if ϕ_σ is a family of smooth convex functions on $\mathbb{R}^m$ and γ is a probability measure on the parameter space S then $\phi := k^{-1}\log\int e^{k\phi_\sigma}\, d\gamma(\sigma)$ is also convex, for any given positive number k, and $\nabla\phi$ is contained in the convex hull of $\{\nabla\phi_\sigma\}$. In the present setting we take $S := S_N$ endowed with the counting measure γ and $\phi_\sigma(\boldsymbol{x}) := \boldsymbol{x} \cdot \boldsymbol{p}_\sigma$, which is clearly convex and satisfies $\nabla_{x_i}\phi_\sigma \in P$. Since P is convex and uniformly bounded, this concludes the proof of Step 2. The convexity of $-C(\cdot, \nu_P)$ then follows from Lemma 2.13. Equivalently, this means that $-\frac{1}{2}d_{W_2}(\mu, \nu_P)^2$ is -1-convex. In fact, as shown in [Ambrosio et al. 2005] using a different argument $-\frac{1}{2}d_{W_2}(\cdot, \nu)^2$ is -1-convex for any fixed $\nu \in \mathcal{P}_2(\mathbb{R}^n)$. □

Next, we recall that the *Monge–Ampère measure* $\mathrm{MA}(\phi)$ of a convex function ϕ on $\mathbb{R}^n$ (4-4) is defined by the property that, for a given Borel set E,

$$\int_E \mathrm{MA}(\phi) := \int_{(\partial\phi)(E)} d\lambda,$$

where $d\lambda$ denotes Lebesgue measure and $\partial\phi$ denotes the subgradient of ϕ (which defines a multivalued map from $\mathbb{R}^n$ to $\mathbb{R}^n$). This is also called the Hessian measure; see [Villani 2003, Section 4.1.4]. In particular, if $\phi \in C^2$, then

$$\mathrm{MA}(\phi) = \det(\partial^2\phi)\, dx,$$

where $\partial^2\phi$ denotes the Hessian matrix of ϕ. We will denote by $\mathcal{C}_P$ the space of all convex functions ϕ on $\mathbb{R}^n$ whose subgradient $\partial\phi$ satisfies

$$(\partial\phi)(\mathbb{R}^n) \subset P$$

and we will say that ϕ is *normalized* if $\phi(0) = 0$. By the convexity of ϕ the gradient condition above equivalently means that ϕ is bounded from above by the support function ϕ_P of P, where $\phi_P(x) := \sup_{p\in P} p \cdot x$.

By Brenier's theorem [1991], given $\mu = \rho\, dx$ in $\mathcal{P}_2(\mathbb{R}^n)$ there exists a unique normalized $\phi \in \mathcal{C}_P$ such that

$$\mathrm{MA}(\phi) = \rho\, dx, \tag{4-6}$$

which equivalently means that the corresponding L^∞-map $\nabla\phi$ from $\mathbb{R}^n$ to P satisfies

$$(\nabla\phi)_*\mu = \nu_P.$$

Given the previous proposition we can use the differentiability result in [Ambrosio et al. 2005] for the Wasserstein L^2-distance to get the following result (see Proposition 2.11 and the subsequent discussion for the definition of the minimal subdifferential):

Lemma 4.2. *The minimal subdifferential of $-C(\cdot, \nu_P)$ on the subspace $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$ of all probability measures in $\mathcal{P}_2(\mathbb{R}^n)$ which are absolutely continuous with respect to dx, may, at a given point $\rho\, dx$, be represented by the L^∞-vector field $\nabla\phi$, where ϕ is the unique normalized solution in $\mathcal{C}_P$ to* (4-6).

Proof. Given formula (4-3) this follows immediately from Theorem 10.4.12 in [Ambrosio et al. 2005] and the fact that if $\mu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$, then Brenier's theorem gives that the optimal transport plan (coupling) from $\mathbb{R}^n$ to P realizing the infimum defining $d_{W_2}(\mu, \nu_P)^2$ is given by the L^∞-map $\nabla\phi$, where ϕ solves (4-6). Since the barycentric projection appearing in Theorem 10.4.12 in [Ambrosio et al. 2005] for the transport plan defined by a transport map gives back the transport map, see Theorem 12.4.4 of the same work, this concludes the proof. □

4B. ***Existence of the gradient flow for $F_\beta(\mu)$.*** Given $\beta \in]0, \infty]$ we set $F_\beta(\mu) := -C(\mu, \nu_P) + H(\mu)/\beta$.

Proposition 4.3. *The gradient flow μ_t of F_β on $\mathcal{P}_2(\mathbb{R}^n)$ emanating from a given μ_0 exists for any $\beta \in]0, \infty]$. Moreover, for $\beta < \infty$ we have that $\mu_t = \rho_t(x)\, dx$, where ρ_t has finite Boltzmann entropy* *and Fisher information and* $\rho(x,t) := \rho_t(x)$ *satisfies the following equation in the sense of distributions on* $\mathbb{R}^n \times]0,\infty[$:

$$\frac{d\rho_t}{dt} = \frac{1}{\beta}\Delta\rho_t + \nabla\cdot(\rho_t\nabla\phi_t), \tag{4-7}$$

where ϕ_t *is the unique normalized solution in* $\mathcal{C}_P$ *to* (4-6) *and* $\nabla\phi_t$ *defines a vector field with coefficients in* L^∞_{loc}.

Proof. Given the previous lemma this follows immediately from Theorem 8.3.1 and Corollary 11.1.8 in [Ambrosio et al. 2005] (the case $\beta = \infty$ has previously been considered by Brenier [2010; 2011; 2016] by lifting the problem to the space of L^2-maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ where Hilbert space techniques can be applied). □

4C. ***Conclusion of the proof of Theorem 1.2.*** By Proposition 4.1 the main assumptions are satisfied. Hence, Theorem 1.1 implies that the corresponding empirical measures converge in law, as $N\to\infty$, to the measure $\mu_t \in \mathcal{P}_2(\mathbb{R}^n)$, where the curve $t \mapsto \mu_t$ is the gradient flow on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$ of the corresponding free energy functional F_β, emanating from μ_0. Finally, Proposition 4.3 says that the gradient flow in question satisfies the evolution equation appearing in Theorem 1.2.

4D. ***The tropical setting.*** The results above are also valid when the permanental interaction energy $E^{(N_k)}(x_1,\dots,x_{N_k})$ is replaced by its tropical analog, i.e., the convex piecewise affine convex function

$$E^{(N_k)}_{\mathrm{trop}}(x_1,\dots,x_{N_k}) := \max \sum_{\sigma\in S_{N_k}} x_1\cdot p_{\sigma(1)} + \cdots + x_N\cdot p_{\sigma(N_k)}. \tag{4-8}$$

In other words this is a *tropical permanent,* i.e., the permanent of the rank-N matrix $(x_i\cdot p_j)$ in the tropical semiring over $\mathbb{R}$, i.e., the set $\mathbb{R}\cup\{-\infty\}$ where the plus and multiplication operations are defined by $\max\{a,b\}$ and $a+b$, respectively [Itenberg and Mikhalkin 2012]. Equivalently, in terms of discrete transport theory this means that

$$E^{(N_k)}_{\mathrm{trop}}(x_1,\dots,x_{N_k}) := -C((\delta_N(x),\delta_N(p))).$$

Passing to the tropical setting has, in particular, computational advantages. Indeed, while all known methods for evaluating (general) permanents take exponential time, the tropical permanent above is, by its very definition, the optimal value of a linear assignment problem and can be computed using an algorithm of cubic-time complexity; see the discussion in [Brenier et al. 2003].

5. Outlook

In this final section we point out some relations between the limiting evolution equation appearing in Theorem 1.2 (whose static solutions correspond to toric Kähler–Einstein metrics) and other well-known evolution equations. We also indicate some relations to sticky particle systems appearing at the microscopic level (i.e., for finite N) and the complex geometric picture. These relations will be elaborated on in a sequel to the present paper [Berman $\geq$ 2018].

5A. ***Relation to other evolution equations and traveling waves.*** In the one-dimensional case when $P := [-a_-, -a_+]$, integrating the evolution equation for ρ_t in Theorem 1.2 once reveals that the bounded decreasing function $u(x,t) := -\partial_x \phi_t$ (physically playing the role of a velocity field) satisfies *Burger's equation* [Hopf 1950] with positive *viscosity* $\kappa := \beta^{-1}$:

$$\partial_t u = \kappa \partial_x^2 u - u \partial_x u$$

with the left and right space asymptotics $\lim_{x\to\pm\infty} u(x,t) = a_\pm$. We recall that Burger's equation is the prototype of a nonlinear wave equation and a scalar conservation law, which is used, among many other things, as a toy model for turbulence in the Navier–Stokes equations [Frisch and Bec 2001]. Interestingly, the barycenter b_P of the polytope P coincides, in this one-dimensional situation, with the negative of the *speed* $s := (a_+ + a_-)/2$ of the time-dependent solution u in the terminology of scalar conservation laws [Lax 1973]. Hence, the vanishing condition $b_P = 0$, which in general is tantamount to the existence of a stationary solution $\rho_t = \rho$ (as discussed in connection to Theorem 1.2) simply means, from the point of view of nonlinear wave theory, that the speed s vanishes.

Similarly, the function $\phi(x,t) := \phi_t(x)$, which in complex-geometric terms is a Kähler potential, satisfies (after the appropriate normalization) the following viscous Hamilton–Jacobi equation, known as the deterministic KPZ equation in the literature on growth of random surfaces [Kardar et al. 1986; Hairer 2013]:

$$\partial_t \phi = \kappa \partial_x^2 \phi + \tfrac{1}{2}(\partial_x \phi)^2. \tag{5-1}$$

In the general higher-dimensional case, the evolution equation (1-13) (which is different than the higher-dimensional version of Burger's equation) can be seen as a dissipative viscous/diffusive version of the semigeostrophic equation appearing in dynamic meteorology; see [Loeper 2006; Ambrosio et al. 2014; Brenier 2011] for a similar situation in cosmology. Moreover, since

$$E(\mu) = -\frac{1}{2} d^2(\mu, \nu_P) + \frac{1}{2}\int |x|^2 \mu + C,$$

where d denotes the Wasserstein L^2-distance, the evolution equation (1-13) can also be seen as a quadratic perturbation (with diffusion) of the "geodesic flow" on the Wasserstein L^2-space, compare to [Ambrosio et al. 2005, Example 11.2.10], which in the one-dimensional case appears in connection to the sticky particle system [Natile and Savaré 2009]. As will be shown in [Berman $\geq$ 2018], the large-time asymptotics of the fully nonlinear evolution equation (1-13) for the probability density ρ_t in $\mathbb{R}^n$ are governed by *traveling wave solutions* in $\mathbb{R}^n$ whose speeds coincide with the negative of the barycenter b_P of the convex body P:

$$\rho_t(x) = \rho(x - b_P t) + o(t), \quad t \to \infty,$$

where the error terms $o(t)$ tends to zero in $L^1(\mathbb{R}^n)$ (and even in relative entropy) and where the limiting profile ρ is uniquely determined from a variant of the Monge–Ampère equation (1-16) together with the condition that its barycenter coincides with the barycenter of the initial data (thus breaking the translation symmetry). In complex-geometric terms, ρ corresponds to a certain canonical Kähler–Einstein metric ω on X with conical singularities "at infinity", playing the role of Calabi's extremal metrics in this context. More generally, as will be elaborated on in [Berman $\geq$ 2018], the results above apply in a more general setting where the measure ν_P is multiplied by a density g, which amounts to replacing the Monge–Ampère equation $\mathrm{MA}(\phi)$ with $g(\nabla\phi)\,\mathrm{MA}(\phi)$ and which from the point of view of scalar conservation laws corresponds to a general concave flux function f (when $n = 1$).

5B. ***The microscopic picture: sticky particles in*** $\mathbb{R}^n$**.** It can be shown that the attractive Newtonian interaction energy in $\mathbb{R}$ is the one-dimensional version of the tropical permanental energy $E^{(N)}_{\mathrm{trop}}(x_1, \dots, x_N)$ appearing in Section 4D. In the general higher-dimensional setting it turns out that a very concrete interpretation of the corresponding EVI gradient flow of $E^{(N)}_{\mathrm{trop}}(x_1, \dots, x_N)$ on $\mathbb{R}^{nN}$ can be given; in particular the particles perform zigzag paths with velocity vectors contained in the polytope $-P$ generalizing the sticky N-particle system on the real line.[3] Moreover, there is a static solution to the corresponding deterministic N-particle system if and only if the "discrete" barycenter of P vanishes,

$$\frac{1}{N}(p_1 + \cdots + p_N) = 0,$$

which is consistent with the fact that the discrete barycenter can be interpreted as the mean velocity of the particles. In general, any initial configuration of points $(x_1, \dots, x_N)(0)$ is assembled, in a finite time, into a single particle x_*, namely the barycenter of $\{x_1, \dots, x_N\}$, which moves at the mean velocity above. The results in the present paper can also be used to study the large N-limit of this deterministic system (which can be seen as a dissipative version of the Hamiltonian particle system introduced in [Cullen et al. 2007] as a discretization of the semigeostrophic equations). But the key point of our approach is that it allows noise to be added to the particle system. Then the role of the large N-limit of x_* is played by the volume form μ_* of a Kähler–Einstein metric on the toric variety determined by the polytope P (compare the discussion in Section 5A).

Interestingly, a similar particle system on $\mathbb{R}^n$ appears in Brenier's approach [2011; 2016] to the early universe reconstruction problem in cosmology [Frisch and Bec 2001] (in connection to the so-called Zeldovich approximation). In fact, our results can be used to validate the formal large N-limit of the N-particle system with noise introduced in [Brenier 2016, Section 2.3].[4]

5C. ***The complex geometric picture.*** In this final section we provide some complex-geometric motivation for the present paper; a more detailed account, including the relations to the Yau–Tian–Donaldson conjecture and tropicalization, will appear elsewhere

Let X be an n-dimensional compact complex manifold. A metric g on X is said to be Kähler–Einstein if g has constant Ricci curvature and g is Kähler; i.e., in local holomorphic coordinates g can be represented as the real part of the positive definite complex Hessian $\partial\phi(z)/(\partial z_i \partial \bar{z}_j)$ of a local function $\phi(z)$ called the Kähler potential of g. If such a metric g exists with positive Ricci curvature, then X is necessarily a projective algebraic variety which is Fano; i.e., the holomorphic (anticanonical) line bundle $L := \det(TX)$ over X is positive.

[3] When $n = 1$ the dynamics is determined by the property that total mass and momentum is conserved in collisions and that the particles stick together when they collide; see [Brenier and Grenier 1998].

[4] As pointed out in [Brenier 2016, Section 2.3], the formal argument used there, which is based on the classical Freidlin–Wentzel theory, as in [Dawson and Gärtner 1987], would require a Lipschitz bound on the drift.

As shown in [Berman 2013a], a Fano manifold comes with a sequence of canonical N-particle random point processes. The number of particles N arises as the pluriantigenera of X:

$$N = N_k := \dim H^0(X, L^{\otimes k}), \quad k = 1, 2, 3, \dots,$$

where $H^0(X, L^{\otimes k})$ denotes the complex vector space consisting of the global holomorphic sections of the k-th tensor power of L. The Fano condition ensures that $N_k \to \infty$ as $k \to \infty$. The local density of the corresponding canonical symmetric probability measure $\mu^{(N_k)}$ on X^{N_k} is defined by

$$\rho^{(N_k)}(z_1, \dots, z_{N_k}) := \frac{1}{Z_{N_k}} \left|\det(z_1, \dots, z_{N_k})\right|^{-2/k}, \quad \det(z_1, \dots, z_{N_k}) := \det(s_i(z_j)), \tag{5-2}$$

where $\det(z_1, \dots, z_{N_k}) \in H^0(X^{N_k}, L^{\otimes k})$ is the Vandermonde-type determinant formed from a given base $s_1, \dots, s_{N_k}$ in $H^0(X, L^{\otimes k})$ and Z_{N_k} is the corresponding normalization constant ensuring that the probability measure has unit mass (by homogeneity $\rho^{(N_k)}$ is independent of the choice of base). However, since the local density $\rho^{(N_k)}(z_1, \dots, z_{N_k})$ has singularities (for example when two points on X merge), the normalization constant Z_{N_k} may be infinite, which means that the random point processes are only well-defined if $Z_{N_k} < \infty$. Such a Fano manifold X was called Gibbs stable in [Berman 2013a], where it was shown that the condition can be rephrased in purely algebrogeometric terms (see also [Fujita 2016] for further developments). It was conjectured in [Berman 2013a] that this condition is equivalent to X admitting a (unique) Kähler–Einstein metric (which necessarily has positive Ricci curvature) whose volume form may be recovered as the deterministic large N-limit of the empirical measures of the corresponding random point processes.[5]

The motivation for the present paper comes from a dynamic approach to the latter conjecture where one introduces the interaction energy

$$E^{(N_k)}(z_1, \dots, z_N) := \frac{1}{k} \log|\det(z_1, \dots, z_{N_k})|^2,$$

which is attractive, in the sense that it tends to $-\infty$ as two particles merge. Locally, this object is represented by a plurisubharmonic function, but in order to get a globally well-defined function on X^{N_k} one also has to fix a background Kähler metric g on X (representing the first Chern class of X) whose volume form dV_g then induces a metric $\|\cdot\|$ on L which is used to replace the absolute values above. The point is that, if X is Gibbs stable, the canonical probability measure $\mu^{(N_k)}$ on X^{N_k} can then be represented globally as the corresponding Gibbs measure at inverse temperature $\beta = 1$ (which is independent of the choice of metric g),

$$\mu^{(N_k)} = \frac{1}{Z_{N_k}} e^{-E^{(N_k)}} dV_g^{\otimes N_k} \left(= \frac{1}{Z_{N_k}} \|\det\|^{-2/k} dV_g^{\otimes N_k} \right),$$

i.e., as a determinantal point process on X at negative temperature. The different zero-temperature case was studied in [Berman et al. 2011].

[5] The convergence of the processes in the opposite case when the dual $\det(T^*X)$ of $\det(TX)$ is positive was settled in [Berman 2013a] (the limit is then the volume form of the unique Kähler–Einstein metric on X with negative Ricci curvature, whose existence was first established in the seminal works of Aubin and Yau).

At any rate, even if X is not Gibbs stable one can still look at the stochastic gradient flow of $E^{(N_k)}$ on the N_k-fold product of the Riemannian manifold (X, g). From this dynamic perspective Gibbs stability simply means that the corresponding stochastic process has an invariant measure, to wit, $\mu^{(N_k)}$. Accordingly, the natural dynamic generalization of the conjecture referred to above is that a (unique) Kähler–Einstein metric g_{KE} exists precisely when the stochastic gradient flow of $E^{(N_k)}$ admits a stationary measure and then its volume form $dV_{g_{\mathrm{KE}}}$ can be recovered from the joint large N- and large t-limit of the flow. More precisely, conjecturally the large N-limit of the corresponding stochastic gradient flows is described by the complex version of the evolution equation (4-7), obtained by replacing the real Monge–Ampère operator with its complex counterpart. The latter flow is, at least formally, the Wasserstein gradient flow of a free-energy-type functional $F(\mu)$ on $\mathcal{P}_2(X, g)$ and F can be identified with the K-energy functional on the space of Kähler metrics in $c_1(X)$ (using the Calabi–Yau isomorphism) [Berman 2013c]. Unfortunately, the study of the latter flows is plagued by various analytical difficulties stemming from the singularities of $E^{(N_k)}$ and the lack of convexity. For example, even in the simplest case when X is the Riemann sphere, i.e., the one-point compactification of the complex plane $\mathbb{C}$, so that $E^{(N_k)}$ is simply the attractive logarithmic pair interaction between N_k equal charges on $\mathbb{C}$, the convergence of the large N-limit, for a fixed time, is a long-standing open problem (however, see [Fournier and Jourdain 2017] for very recent partial results).

5C1. *The toric setting and its tropicalization.* The complex geometric setting which is relevant to the present paper appears when X is a toric Fano manifold, i.e., X admits a holomorphic action of the real n-torus T such that (X, T) can be realized as an equivariant compactification of the complex torus $\mathbb{C}^{*n}$ (with its standard T-action) [Donaldson 2008]. Such a compactification X is determined by a convex polytope P, which has the property that under the dense embedding of $\mathbb{C}^{*n}$ into X, the complex vector space $H^0(X, L^{\otimes k})$ may be identified with the space of all holomorphic Laurent polynomials $f(z)$ on $\mathbb{C}^{*n}$ of the form

$$f(z) = \sum_{m \in kP \cap \mathbb{Z}^n} a_m z^m$$

(using multi-index notation). In particular, introducing an ordering $m_1, \dots, m_{N_k}$ on the integer points of $kP \cap \mathbb{Z}^n$ gives a basis $s_{m_1}(z), \dots, s_{m_{N_k}}$ of multinomials in $H^0(X, L^{\otimes k})$, which can be used to represent

$$\det(z_1, \dots, z_{N_k}) = \sum_{\sigma \in S_N} (-1)^{\mathrm{sign}(\sigma)} z_1^{m_{\sigma(1)}} \cdots z_{N_k}^{m_{\sigma(N)}}. \tag{5-3}$$

Now, the real vector space $\mathbb{R}^n$ makes its appearance when introducing logarithmic coordinates on $\mathbb{C}^{*n}$, i.e., as the image of the Log map

$$\mathrm{Log} : \mathbb{C}^{*n} \to \mathbb{R}^n, \quad z \mapsto x := (\log |z_1|^2, \dots, \log |z_n|^2),$$

whose fibers are the orbits of the action of T. Using this map, T-invariant metrics on $L \to X$ with positive curvature may be identified with convex functions $\phi(x)$ on $\mathbb{R}^n$ such that $(\partial\phi)(\mathbb{R}^n) \subset P$. In this picture the permanental density $\mathrm{Per}(x_1, \dots, x_{N_k})$ arises as the push-forward to $\mathbb{R}^n$, under the Log map, of the determinant density (5-3). In other words, the smooth convex permanental energy $E_{\mathrm{per}}^{(N)}(x_1, \dots, x_N)$, formula (1-11), on $\mathbb{R}^n$ is an averaged version of the singular plurisubharmonic interaction energy $E^{(N_k)}$ on $\mathbb{C}^{*n}$:

$$E^{(N)}_{\mathrm{per}}(x_1, \dots, x_N) = \frac{1}{k} \log \int_{T^{N_k}} e^{kE^{(N_k)}}\, d\theta^{\otimes N_k}. \tag{5-4}$$

Similarly, its tropical version $E^{(N)}_{\mathrm{trop}}(x_1, \dots, x_N)$ is the piecewise affine convex function on $\mathbb{R}^{nN}$ obtained as the tropicalization of the Laurent polynomial $\det(z_1, \dots, z_{N_k})$ on $\mathbb{C}^{*nN_k}$.[6] Accordingly, Theorem 1.2 should be seen in the light of the well-known philosophy of replacing an elusive complex-geometric problem by a more tractable convex-geometric one, by the process of tropicalization; see, for example, [Itenberg and Mikhalkin 2012].

Appendix: The Otto calculus

In this appendix we briefly recall Otto's [2001] beautiful (formal) Riemannian interpretation of the Wasserstein L^2-metric d_2 on $\mathcal{P}_2(\mathbb{R}^n)$. The material is included with the nonexpert in mind as a motivation for the material on gradient flows on $\mathcal{P}^2(\mathbb{R}^n)$ recalled in Section 2E.

The Otto metric. For simplicity we will consider probability measures of the form $\mu = \rho\, dx$, where ρ is smooth positive everywhere (in order to make the arguments below rigorous one should also specify the rate of decay of ρ at ∞ in $\mathbb{R}^n$). The corresponding subspace of probability measures in $\mathcal{P}_2(\mathbb{R}^n)$ will be denoted by $\mathcal{P}$. First recall that the ordinary "affine tangent vector" of a curve ρ_t in $\mathcal{P}$ at $\rho := \rho_0$, when ρ_t is viewed as a curve in the affine space $L^1(\mathbb{R}^n)$), is the function $\dot\rho$ on $\mathbb{R}^n$ defined by

$$\dot\rho(x) := \frac{d\rho_t(x)}{dt}_{|t=0}.$$

Next, let us show how to identify $\dot\rho$ with a vector field $v_{\dot\rho}$ in $L^2(\rho\, dx, \mathbb{R}^n)$, which, by definition, is the (nonaffine) "tangent vector" of ρ_t at ρ; i.e., $v_{\dot\rho} \in T_\rho \mathcal{P}$. First, since the total mass of ρ_t is preserved, we have $\int \dot\rho\, dx = 0$ and hence there is a vector field v on $\mathbb{R}^n$ solving the continuity equation

$$\dot\rho = -\nabla \cdot (\rho v). \tag{A-1}$$

In geometric terms this means that

$$\rho_t\, dx = (F_t^V)_*(\rho_0\, dx) + o(t), \tag{A-2}$$

where F_t^V is the family of maps defined by the flow of V. Now, under suitable regularity assumptions, $v_{\dot\rho}$ may be defined as the "optimal" vector field v solving the previous equation, in the sense that it minimizes the L^2-norm in $L^2(\rho\, dx, \mathbb{R}^n)$. The Otto metric is then defined by

$$g(v_{\dot\rho}, v_{\dot\rho}) = \inf_v \int \rho |v|^2\, dx = \int \rho |v_{\dot\rho}|^2\, dx, \tag{A-3}$$

[6] Incidentally, tropicalization may be interpreted as a zero-temperature limit by writing the tropical sum $\max\{a, b\}$ as the limit of $T^{-1} \log(e^{(1/T)a} + e^{(1/T)b})$ as $T \to 0$; compare to the discussion in [Itenberg and Mikhalkin 2012].

which can be seen as the linearized version of the defining formula (2-6) for the Wasserstein L^2-metric. By Hodge theory, the optimal vector field $v_{\dot{\rho}}$ may be written as $v_{\dot{\rho}} = \nabla\phi$ for a unique normalized function ϕ on $\mathbb{R}^n$ (under suitable assumptions).

The microscopic point of view. Let us remark that a simple heuristic "microscopic" derivation of the Otto metric can be given using the isometry defined by the empirical measure δ_N (Lemma 2.3). Indeed, given a curve $(x_1(t), \dots, x_N(t))$ in the Riemannian product $\left(X^N, \frac{1}{N}g^{\otimes N}\right)$ with tangent vector $(dx_1(t)/dt, \dots, dx_1(t)/dt)$ at $t = 0$ we can write its squared Riemannian norm at $(x_1(0), \dots, x_N(0))$ as

$$\left\| \left(\frac{dx_1(t)}{dt}, \dots, \frac{dx_1(t)}{dt} \right) \right\|^2 = \int |v|^2 \delta_N(0), \tag{A-4}$$

where $\delta_N(t) := \frac{1}{N}\sum \delta_{x_i(t)}$ and v is any vector field on $X = \mathbb{R}^n$ such that $v(x_i) = dx_i(t)/dt_{|t=0}$. Note that setting $\rho_t := \delta_N(t)$, the vector field v satisfies the push-forward relation (A-2) (with vanishing error term). Moreover, since passing to the quotient X^N/S_N does not effect the corresponding curve ρ_t, minimizing with respect to the action of the permutation group S_N in formula (A-4) corresponds to the infimum defining the Otto metric in formula (A-3).

Relation to gradient flows and drift-diffusion equations. If G is a smooth functional on $\mathcal{P}$ then a direct computations reveals that its (formal) gradient with respect to the Otto metric at ρ corresponds to the vector field $v(x) = \nabla_x(\partial G(\rho)/\partial\rho)$. In other words, the gradient flow of $G(\rho)$ may be written as

$$\frac{\partial\rho_t(x)}{\partial t} = \nabla_x \cdot (\rho v_t(x)), \quad v_t(x) = \nabla_x \frac{\partial G(\rho)}{\partial\rho}_{|\rho=\rho_t} \tag{A-5}$$

In particular, for the Boltzmann entropy $H(\rho)$, formula (2-1), one gets, since $\partial G(\rho)/\partial\rho = \log\rho$ (using that the mass is preserved), that the corresponding gradient flow is the heat (diffusion) equation and the gradient flow structure then implies that $H(\rho_t)$ is decreasing along the heat equation. Moreover, a direct calculation reveals that H is *convex* on $\mathcal{P}$ in sense that the Hessian of H is nonnegative and hence it also follows from general principles that the squared Riemannian norm $|\nabla H|^2(\rho_t)$ is decreasing. In fact, by definition $|\nabla H|^2(\rho)$ coincides with the Fisher information functional $I(\rho)$, formula (2-1). More generally, the gradient flow of the Gibbs free energy F_β^V is given by the diffusion equation with linear drift $\nabla_x V$,

$$\frac{\partial\rho_t}{\partial t} = \frac{1}{\beta}\Delta_x\rho_t + \nabla_x \cdot (\rho_t \nabla_x V), \tag{A-6}$$

often called the linear Fokker–Planck equation in the mathematical physics literature. The study of the previous flow using a variational discretization scheme on $\mathcal{P}^2(\mathbb{R}^n)$ was introduced in [Jordan et al. 1998] (compare to Section 5C1).

Acknowledgments

It is a pleasure to thank Eric Carlen for several stimulating discussions and whose inspiring lecture in the Kinetic Theory seminar at Chalmers concerning [Blanchet et al. 2012] drew our attention to the Otto calculus and the theory of gradient flows on the Wasserstein space. Thanks also to Luigi Ambrosio for helpful comments on the paper and to Yann Brenier, José Carrillo, Maxime Hauray and Bernt Wennberg for providing us with references and to the referee for many comments which helped to improve the exposition of the paper. This work was supported by grants from the Swedish Research Council, the Knut and Alice Wallenberg Foundation and the European Research Council.

References

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

[Ambrosio et al. 2009] L. Ambrosio, G. Savaré, and L. Zambotti, "Existence and stability for Fokker–Planck equations with log-concave reference measure", *Probab. Theory Related Fields* **145**:3-4 (2009), 517–564. MR Zbl

[Ambrosio et al. 2014] L. Ambrosio, M. Colombo, G. De Philippis, and A. Figalli, "A global existence result for the semi-geostrophic equations in three dimensional convex domains", *Discrete Contin. Dyn. Syst.* **34**:4 (2014), 1251–1268. MR Zbl

[Berman 2013a] R. J. Berman, "Kähler–Einstein metrics, canonical random point processes and birational geometry", preprint, 2013. To appear in *AMS Proceedings of the 2015 Summer Research Institute on Algebraic Geometry*. arXiv

[Berman 2013b] R. J. Berman, "Statistical mechanics of permanents, real-Monge–Ampère equations and optimal transport", preprint, 2013. arXiv

[Berman 2013c] R. J. Berman, "A thermodynamical formalism for Monge–Ampère equations, Moser–Trudinger inequalities and Kähler–Einstein metrics", *Adv. Math.* **248** (2013), 1254–1297. MR Zbl

[Berman 2017] R. J. Berman, "Large deviations for Gibbs measures with singular Hamiltonians and emergence of Kähler–Einstein metrics", *Comm. Math. Phys.* **354**:3 (2017), 1133–1172. MR Zbl

[Berman $\geq$ 2018] R. J. Berman, "Stability of some fully-nonlinear shock waves and toric Kähler–Einstein metrics", in preparation.

[Berman and Berndtsson 2013] R. J. Berman and B. Berndtsson, "Real Monge–Ampère equations and Kähler–Ricci solitons on toric log Fano varieties", *Ann. Fac. Sci. Toulouse Math.* (6) **22**:4 (2013), 649–711. MR Zbl

[Berman and Lu $\geq$ 2018] R. J. Berman and C. H. Lu, "A drift-diffusion equation in complex geometry and convergence towards Kähler–Einstein metrics", in preparation.

[Berman and Önnheim 2016] R. J. Berman and M. Önnheim, "Propagation of chaos for a class of first order models with singular mean field interactions", preprint, 2016. arXiv

[Berman et al. 2011] R. Berman, S. Boucksom, and D. Witt Nyström, "Fekete points and convergence towards equilibrium measures on complex manifolds", *Acta Math.* **207**:1 (2011), 1–27. MR Zbl

[Blanchet et al. 2012] A. Blanchet, E. A. Carlen, and J. A. Carrillo, "Functional inequalities, thick tails and asymptotics for the critical mass Patlak–Keller–Segel model", *J. Funct. Anal.* **262**:5 (2012), 2142–2230. MR Zbl

[Braun and Hepp 1977] W. Braun and K. Hepp, "The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles", *Comm. Math. Phys.* **56**:2 (1977), 101–113. MR Zbl

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

[Brenier 2010] Y. Brenier, "Hilbertian approaches to some non-linear conservation laws", pp. 19–35 in *Nonlinear partial differential equations and hyperbolic wave phenomena*, edited by H. Holden and K. H. Karlsen, Contemp. Math. **526**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

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

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

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

[Brenier et al. 2003] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Moyhayaee, and A. Sobolevskiĭ, "Reconstruction of the early universe as a convex optimization problem", *Mon. Not. R. Astron. Soc.* **346**:2 (2003), 501–524.

[Chen and Zheng 2013] X. Chen and K. Zheng, "The pseudo-Calabi flow", *J. Reine Angew. Math.* **674** (2013), 195–251. MR Zbl

[Cullen et al. 2007] M. Cullen, W. Gangbo, and G. Pisante, "The semigeostrophic equations discretized in reference and dual variables", *Arch. Ration. Mech. Anal.* **185**:2 (2007), 341–363. MR Zbl

[Dawson and Gärtner 1987] D. A. Dawson and J. Gärtner, "Large deviations from the McKean–Vlasov limit for weakly interacting diffusions", *Stochastics* **20**:4 (1987), 247–308. MR Zbl

[Dembo and Zeitouni 1993] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Jones and Bartlett, Boston, 1993. MR Zbl

[Donaldson 2008] S. K. Donaldson, "Kähler geometry on toric manifolds, and some other manifolds with large symmetry", pp. 29–75 in *Handbook of geometric analysis, I*, edited by L. Ji et al., Adv. Lect. Math. **7**, International Press, Somerville, MA, 2008. MR Zbl

[Fournier and Jourdain 2017] N. Fournier and B. Jourdain, "Stochastic particle approximation of the Keller–Segel equation and two-dimensional generalization of Bessel processes", *Ann. Appl. Probab.* **27**:5 (2017), 2807–2861. MR Zbl

[Frisch and Bec 2001] U. Frisch and J. Bec, "Burgulence", pp. 341–383 in *New trends in turbulence* (Les Houches, 2000), edited by M. Lesieur et al., EDP Sci., Les Ulis, 2001. MR Zbl

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

[Fujita 2016] K. Fujita, "On Berman–Gibbs stability and K-stability of $\mathbb{Q}$-Fano varieties", *Compos. Math.* **152**:2 (2016), 288–298. MR Zbl

[Hairer 2013] M. Hairer, "Solving the KPZ equation", *Ann. of Math.* (2) **178**:2 (2013), 559–664. MR Zbl

[Hauray and Mischler 2014] M. Hauray and S. Mischler, "On Kac's chaos and related problems", *J. Funct. Anal.* **266**:10 (2014), 6055–6157. MR Zbl

[Hiriart-Urruty 1980] J.-B. Hiriart-Urruty, "Extension of Lipschitz functions", *J. Math. Anal. Appl.* **77**:2 (1980), 539–554. MR Zbl

[Hopf 1950] E. Hopf, "The partial differential equation $u_t + uu_x = \mu u_{xx}$", *Comm. Pure Appl. Math.* **3** (1950), 201–230. MR Zbl

[Itenberg and Mikhalkin 2012] I. Itenberg and G. Mikhalkin, "Geometry in the tropical limit", *Math. Semesterber.* **59**:1 (2012), 57–73. MR Zbl

[Jordan et al. 1998] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation", *SIAM J. Math. Anal.* **29**:1 (1998), 1–17. MR Zbl

[Kac 1956] M. Kac, "Foundations of kinetic theory", pp. 171–197 in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, III*, University of California Press, Berkeley, 1956. MR Zbl

[Kardar et al. 1986] M. Kardar, G. Parisi, and Y.-C. Zhang, "Dynamic scaling of growing interfaces", *Phys. Rev. Lett.* **56**:9 (1986), 889–892. Zbl

[Keller and Segel 1970] E. F. Keller and L. A. Segel, "Initiation of slime mold aggregation viewed as instability", *J. Theo. Biol.* **26**:3 (1970), 399–415. Zbl

[Lax 1973] P. D. Lax, *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, CBMS-NSF Regional Conference Series in Applied Mathematics **11**, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1973. MR Zbl

[Loeper 2006] G. Loeper, "A fully nonlinear version of the incompressible Euler equations: the semigeostrophic system", *SIAM J. Math. Anal.* **38**:3 (2006), 795–823. MR Zbl

[Lott and Villani 2009] J. Lott and C. Villani, "Ricci curvature for metric-measure spaces via optimal transport", *Ann. of Math.* (2) **169**:3 (2009), 903–991. MR Zbl

[Mayer 1998] U. F. Mayer, "Gradient flows on nonpositively curved metric spaces and harmonic maps", *Comm. Anal. Geom.* **6**:2 (1998), 199–253. MR Zbl

[McCann 1997] R. J. McCann, "A convexity principle for interacting gases", *Adv. Math.* **128**:1 (1997), 153–179. MR Zbl

[McKean 1966] H. P. McKean, Jr., "A class of Markov processes associated with nonlinear parabolic equations", *Proc. Nat. Acad. Sci. U.S.A.* **56** (1966), 1907–1911. MR Zbl

[McKean 1967] H. P. McKean, Jr., "Propagation of chaos for a class of non-linear parabolic equations", pp. 41–57 in *Stochastic differential equations* (*Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967*), Air Force Office Sci. Res., Arlington, VA, 1967. MR

[Messer and Spohn 1982] J. Messer and H. Spohn, "Statistical mechanics of the isothermal Lane–Emden equation", *J. Statist. Phys.* **29**:3 (1982), 561–578. MR

[Mischler et al. 2015] S. Mischler, C. Mouhot, and B. Wennberg, "A new approach to quantitative propagation of chaos for drift, diffusion and jump processes", *Probab. Theory Related Fields* **161**:1-2 (2015), 1–59. MR Zbl

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

[Otto 2001] F. Otto, "The geometry of dissipative evolution equations: the porous medium equation", *Comm. Partial Differential Equations* **26**:1-2 (2001), 101–174. MR Zbl

[Robinson and Ruelle 1967] D. W. Robinson and D. Ruelle, "Mean entropy of states in classical statistical mechanics", *Comm. Math. Phys.* **5** (1967), 288–300. MR

[Schwabl 2002] F. Schwabl, *Statistical mechanics*, Springer, 2002. MR Zbl

[Shandarin and Zel'dovich 1989] S. F. Shandarin and Y. B. Zel'dovich, "The large-scale structure of the universe: turbulence, intermittency, structures in a self-gravitating medium", *Rev. Modern Phys.* **61**:2 (1989), 185–220. MR

[Stroock and Varadhan 1997] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Grundlehren der mathematischen Wissenschaften **233**, Springer, 1997. Zbl

[Sznitman 1991] A.-S. Sznitman, "Topics in propagation of chaos", pp. 165–251 in *École d'Été de Probabilités de Saint-Flour XIX—1989*, edited by P. L. Hennequin, Lecture Notes in Math. **1464**, Springer, 1991. MR Zbl

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

[Wang and Zhu 2004] X.-J. Wang and X. Zhu, "Kähler–Ricci solitons on toric manifolds with positive first Chern class", *Adv. Math.* **188**:1 (2004), 87–103. MR Zbl

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

Received 28 Jun 2016. Revised 9 Oct 2017. Accepted 12 Jan 2018.

ROBERT J. BERMAN: robertb@chalmers.se
Department of Mathematical Sciences, Chalmers University of Technology, Göteborg, Sweden

MAGNUS ÖNNHEIM: onnheimm@chalmers.se
Department of Mathematical Sciences, Chalmers University of Technology and University of Götenburg, Göteborg, Sweden

Analysis & PDE

msp.org/apde

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

The subscription price for 2018 is US $275/year for the electronic version, and $480/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS & PDE

Volume 11 No. 6 2018

2157-5045(2018)11:6;1-O