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

ROBERT J. BERMAN, MAGNUS ÖNNHEIM

Abstract. Motivated by a probabilistic approach to Kähler-Einstein metrics we consider a general non-equilibrium statistical mechanics model in Euclidean space consisting of the stochastic gradient flow of a given quasi-convex N-particle interaction energy. We show that a deterministic “macroscopic” evolution equation emerges in the large N-limit of many particles. The proof uses the theory of weak gradient flows on the Wasserstein space and in particular De Giorgi’s notion of “minimizing movements. Applied to the setting of permanental point processes at “negative temperature” the corresponding limiting evolution equation yields a new 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 non-linear version of various extensively studied dissipative evolution equations and conservations laws, including the Keller-Segal equation and Burger’s equation. We also obtain a real probabilistic analog of the complex geometric Yau-Tian-Donaldson conjecture in this setting. In another direction applications to singular pair interactions in 1D are given. Complex geometric aspects of these results will be discussed elsewhere.

Contents

1. Introduction 1
2. General setup and proof of Theorem 1.1 11
3. Permanental processes and toric Kähler-Einstein metrics 28
4. Singular pair interactions 31
References 35

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 [6, 10], formulated in terms of certain β-deformations of determinantal (fermionic) point processes. The approach in [6, 10] 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 dynamic version of the probabilistic approach in [6, 10]. In other words, we are in the realm of non-equilibrium 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 non-linearity of the corresponding interaction energies, we will here focus on the real analog of the complex setting introduced in [9], taking place in $\mathbb{R}^n$ and which corresponds to the case when $X$ is a toric complex algebraic variety. As explained in [9] in this real setting the determinantal (fermionic) processes are replaced by permanental (bosonic) processes and convexity plays the role of positive Ricci curvature/plurisubharmonicity.

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 non-linear real Monge-Ampère operator. It exhibits a phase transition at a certain geometrically determined critical parameter. 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 [38, 17], Burger’s equation [34, 27] in the theory of non-linear waves and scalar conservation laws and the deterministic version of the Kardar–Parisi–Zhang (KPZ) equation describing surface growth [37]. In the higher dimensional real case the equation can be viewed as a dissipative viscous version of the semi-geostrophic equation appearing in dynamic meteorology (see [43, 1] and references therein). For the corresponding static problem we establish a real analog of the Yau-Tian-Donaldson conjecture (Theorem 1.4) which, in particular, yields a probabilistic construction of toric Kähler-Einstein metrics. The relation to complex geometry will be elaborated on elsewhere.

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-Savaree [2], which provides a rigorous framework for the Otto calculus [52]. In particular, as in [2] convexity (or more generally $\lambda$–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 [7] 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 and whose gradient flow with respect to a different metric, the Mabuchi-Donaldson-Semmes metric, is the renowned Calabi flow. The regularity and large time properties of the evolution equation in question will be studied elsewhere [11].

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 final Section 4 of the paper we also give some applications of our approach to singular pair interactions on the real line (including a seemingly new convexity result, Proposition 4.3). This appears to generalize and complement some previous results in the literature concerning the log gas on the real line (including Dyson’s Brownian motion for random matrices).
and models for swarm aggregation and populations dynamics (see [21] and references therein).

1.1. **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 $E^{(N)}(x_1, x_2, ..., x_N)$. At a fixed inverse temperature $\beta$ the distribution of particles at time $t$ is, according to non-equilibrium statistical mechanics, described by the following system of stochastic differential equations (SDEs), under suitable regularity assumptions on $E^{(N)}$:

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

where $B_i$ denotes $N$ independent Brownian motions on $\mathbb{R}^n$ (called the *Langevin equation* in the physics literature).

In other words, this is the Ito 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 [36] 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 $\delta_N$ the empirical measures

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

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 $\mu_0$ then at any later time $t$ the empirical measure $\delta_N(t)$ converges in law to a deterministic curve $\mu_t$ of measures on $\mathbb{R}^n$

$$ \lim_{N \to \infty} \delta_N(t) = \mu_t $$

In the terminology of Kac [36] (see also [55]) this means that *propagating of chaos* holds at any time $t$. It should be stressed that the previous statement admits a pure PDE formulation, not involving any stochastic calculus (see Section 2.5) and it is this analytic point of view that we will adopt here. Moreover, from a differential geometric point of view the SDEs (1.1) correspond, for $E$ smooth, to the heat flow on $X^N$ of the Witten Laplacian of the “Morse function” $E$, appearing in Witten’s super symmetry approach to Morse theory.

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 $E^{(N)}(x_1, x_2, ..., x_N)$ can be as written as

$$ E^{(N)}(x_1, x_2, ..., x_N) = NE(\delta_N) $$

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. By the results in [55] [21] 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 non-linear evolution equation on $\mathcal{P}(X)$ of the following form:

$$ \frac{d\rho_t}{dt} = \frac{1}{\beta} \Delta \rho_t + \nabla \cdot (\rho_t v[\rho_t]) $$

3
where we have identified $\mu = \rho \, dx$ with its density $\rho$ and $v[\mu]$ is a function on $\mathcal{P}(X)$ taking valued in the space of vector fields on $X$ defined as the gradient of the differential $dE[\mu]$ of $E$ on $\mathcal{P}(X)$

\begin{equation}
 v[\mu] = \nabla(dE[\mu])
\end{equation}

where the differential $dE[\mu]$ at $\mu$ is identified with a function on $X$, by standard duality (the alternative suggestive notation $v[\rho] = \nabla \frac{\partial E(\rho)}{\partial \rho}$ is often used in the literature). In the kinetic theory literature drift-diffusion equations of the form (1.6) are often called McKean-Vlasov equations. More generally, the results referred to above hold in the more general setting where the gradient vector field $\frac{\partial}{\partial x_i} E(N)(x_1, x_2, \ldots, x_N)$ on $X$ appearing in equation (1.1) is replaced by $v[\delta_N]$ for a given vector field valued function $v[\mu]$ on $\mathcal{P}(X)$, satisfying appropriate continuity properties.

One of the main aims of the present work is to introduce a new approach to the propagating of chaos result (1.3) for the stochastic dynamics (1.1) which exploits the gradient structure of the equations in question and which appears to apply 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 there is a functional $E(\mu)$ on $\mathcal{P}(\mathbb{R}^n)$ such that

\begin{equation}
 \frac{1}{N} E(N)(x_1, x_2, \ldots, x_N) = E(\delta_N) + o(1),
\end{equation}

for a sequence of functionals $E(N)$ which are uniformly Lipschitz continuous in each variable separately, i.e. there is a constant $C$ such that

\begin{equation}
 |\nabla x_i E(N)| \leq C
\end{equation}

and the error term $o(1)$ tends to zero, as $N \to \infty$ (for $x_i$ uniformly bounded). Moreover, $E(N)$ is $\lambda-$convex on $X^N$ for some real number $\lambda$, which, by symmetry, means that the (distributional) Hessian are uniformly bounded from below for any fixed index $i$:

\begin{equation}
 \nabla^2 x_i E(N) \geq \lambda I,
\end{equation}

where $I$ denotes the identity matrix. This implies that there exists a unique solution to the evolution equation (1.5) in the sense of weak gradient flows on the space $\mathcal{P}_2(X)$ of all probability measures with finite second moments equipped with the Wasserstein $L^2$–metric [2]:

\[
\frac{d\mu_t}{dt} = -\nabla F_\beta(\mu_t)
\]

where $F_\beta$ is the free energy type functional corresponding to the macroscopic energy $E(\mu)$ at inverse temperature $\beta$ :

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

and where $H(\mu)$ is the Boltzmann entropy of $\mu$ (see Section 2.1 for notation).

**Theorem 1.1.** Suppose that $E(N)$ is a sequence of symmetric functions on $(\mathbb{R}^n)^N$ satisfying the Main Assumptions (1.7), (1.8) and (1.9). Then, for any fixed positive time $t$, the empirical measure $\frac{1}{N} \sum \delta_{x_i(t)}$ of the system of SDEs (1.1) with independent
initial data distributed according to $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$ converges in law, as $N \to \infty$, to the deterministic measure $\mu_t$ evolving by the gradient flow on the Wasserstein space of the corresponding free energy functional $F_\beta$ emanating from $\mu_0$.

It should be stressed that the key point of our approach is that we do not need to assume that the drift $v[\mu](x)$ defined by formula (1.6) has any continuity properties with respect to $\mu$ or $x$. Or more precisely, even if the $N-$dependent drift vector field $v^{(N)}$ may very well be smooth for any fixed $N$, we do not assume that it is uniformly bounded in $N$. This will be crucial in the applications to toric Kähler-Einstein metrics below.

We recall that if the drift has suitable continuity assumptions, then the existence of a solution to the drift-diffusion equation (1.5) can be established using fix point type arguments [55]. However, in our case we have, in general, to resort to the weak gradient flow solutions provided by the general theory in [2], where the solution $\rho_t$ can be characterized uniquely by a differential inequality called the Evolutionary Variational Inequality (EVI). As shown in [2] the corresponding solution $\rho_t$ satisfies the drift-diffusion equation (1.5) in a suitable weak sense (as follows formally from the Otto calculus [52]).

Our approach is inspired by the approach of Messer-Spohn [48] concerning the static problem for the SDEs (1.1), i.e., the study of the Boltzmann-Gibbs measure on $X^N$ associated to $E^{(N)}$ at inverse temperature $\beta$:

$$\mu_\beta^{(N)} = \frac{e^{-\beta E^{(N)}}}{Z_N} dx^\otimes N$$

assuming that the normalization function $Z_N$ (the partition function) is finite:

$$Z_N := \int_{X^N} e^{-\beta E^{(N)}} dx < \infty$$

From a statistical mechanical point of view this probability measure describes the microscopic equilibrium distribution at a fixed inverse temperature $\beta$ and it appears as the law of the large time limit (for $N$ fixed) of the empirical measures $\delta_N(t)$.

We thus get a uniform approach which applies both to the dynamic and the static setting and which in the latter case leads to the following generalization of [48]:

**Theorem 1.2.** Suppose that $E^{(N)}$ is uniformly Lipschitz continuous and convex in each fixed variable and satisfies the following uniform properness assumption:

$$E^{(N)}(x) \geq \frac{1}{C} \sum_{i=1}^{N} |x_i| - CN$$

for some positive constant $C$. Then the Boltzmann-Gibbs measures corresponding to $E_N$ are well-defined and

$$\lim_{N \to \infty} -\frac{1}{N\beta} \lim_{N \to \infty} \int e^{-\beta E^{(N)}} dx^\otimes N = \inf_{\mathcal{P}_2(\mathbb{R}^n)} F_\beta(> -\infty)$$

Moreover, the corresponding free energy functional $F_\beta$ on $\mathcal{P}_2(\mathbb{R}^n)$ admits a minimizer $\mu_\beta$ and if it is uniquely determined, then the corresponding empirical measures on $\mathbb{R}^n$ converge in law as $N \to \infty$ to the deterministic measure $\mu_\beta$ (which, as a consequence, is log concave).
It should be stressed that the properness assumption in the previous theorem which will appear naturally in the setting of toric Kähler-Einstein metrics below, corresponds to properness wrt the $L^1$–Wasserstein metric on $\mathcal{P}(\mathbb{R}^n)$, which is thus strictly weaker than demanding properness with respect to the $L^2$–Wasserstein metric. But using the convexity assumption on $E_N$, we will bypass this difficulty using Prekopa’s inequality and Borell’s lemma (see also the appendix of the paper which motivates the properness assumption above).

We briefly point out that the previous theorem is also related to previous work on lattice spin models such as the Kac model [33], as well as lattice models for random growth of surfaces [30], where the large $N$–limit corresponds to the “thermodynamic limit”, where the lattice is approximated by a finite volume lattice.

1.1.1. Generalizations.

- By rescaling $E^{(N)}$ we may as well allow the “inverse temperature” $\beta$ 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 [40].

- The assumption [1.7] in conjunction with the Lipschitz assumption [1.8] may be replaced by the assumption that the limit of the mean energies corresponding to $E^{(N)}$, in the sense of statistical mechanics, exists (i.e. that Proposition [2.18] below holds) and that $E^{(N)}$ has a uniform coercivity property. For example, one can add to the Lipschitz function $E^{(N)}$ any term of the form $N\mathcal{V}(\delta_N)$, for a given coercive $\lambda$–convex $\mathcal{V}$ functional on $\mathcal{P}_2(\mathbb{R}^n)$ (one then replaces $E(\mu)$ with $E(\mu) + \mathcal{V}(\mu)$).

- To illustrate that our method of proof also applies under even more singular situations we consider in Section [4] some applications to pair interactions on the real line (including Dyson’s Brownian motion) where the corresponding functional $E$ is rather singular (for example it may be equal to $\infty$ on the image of $\delta_N$).

1.1.2. Idea of the proof of Theorem [1.1] and comparison with previous results. The starting point of the proof if the basic fact that the SDEs [1.1] on $X^N$ admit a PDE formulation: 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 [2]: the rough idea to show that the any weak limit curve $\Gamma(t)$ of the laws $\Gamma_N(t) := (\delta_N)_* \mu_N(t)$ is of the form $\Gamma(t) := \delta_{\mu_t}$, where the curve $\mu_t$ 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 $\Gamma_N(t_j)$, defined by a variational Euler scheme (a “minimizing movement” in De Georgi’s terminology) as in [35, 2]. We then establish a discretized version of Theorem [1.1] saying that if, at a given discrete time $t_j$ the following convergence
holds in the $L^2$--Wasserstein metric
\[
\lim_{N \to \infty} \Gamma^N_{t_j} = \delta_{\mu_j},
\]
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 $\tau$ tend to zero. This last step uses that the very precise error estimates established in [2], for discretizations schemes as above, only depend on a uniform lower bound $\lambda$ on the convexity of the interaction energies.

Our proof appears to be to rather different from the probabilistic approaches in [55, 24] (and elsewhere) which are based on a study of non-linear martingales and the recent PDE approach in [50]. As pointed out above these approaches seem to require a Lipschitz control on the drift vector field $v^{(N)}$ and hence a two-sided uniform bound on the Hessian of the interaction energy $E^{(N)}$, while we only require a uniform lower bound.

1.2. 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 $p_1, ..., p_N$ of the $N$ elements of $P_\mathbb{Z}$. Given a configuration $(x_1, ..., x_N)$ of $N$ points on $X$ we denote by $\text{Per}(x_1, ..., x_N)$ the number defined as the permanent of the rank $N$ matrix with entries $A_{ij} := e^{x_i \cdot p_j}$:

(1.11) \[
\text{Per}(x_1, ..., x_N) := \text{Per}(e^{x_1 \cdot p}) = \sum_{\sigma \in S_N} e^{x_1 \cdot p_{\sigma(1)} + \cdots + x_N \cdot p_{\sigma(N)}},
\]

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 $P_\mathbb{Z}$). 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

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

where $V(P)$ denotes the Euclidean volume of $P$. To simplify the notation we will often drop the explicit dependence of $N$ on $k$. Definition the “permanental interaction energy”

(1.12) \[
E^{(N_k)}(x_1, ..., x_{N_k}) = \frac{1}{k} \log \text{Per}(x_1, ..., x_{N_k})
\]

the corresponding SDEs (1.1) may be explicitly written as

(1.13) \[
dx_i = -\frac{1}{k} \sum_{\sigma \in S_N} p_{\sigma(i)} f_\sigma(x_1, ..., x_N) dt + \frac{2}{\beta^{1/2}} dB_i(t),
\]

where the drift is a convex combination of the elements in $P_\mathbb{Z}$ with weights $f_\sigma(x_1, ..., x_N)$ given by

\[
f_\sigma(x_1, ..., x_N) := \frac{e^{x_1 \cdot p_{\sigma(1)} + \cdots + x_N \cdot p_{\sigma(N)}}}{\sum_{\sigma \in S_N} e^{x_1 \cdot p_{\sigma(1)} + \cdots + x_N \cdot p_{\sigma(N)}}}
\]

By the results in [9] the assumptions in Theorem 1.1 hold with

\[
E(\mu) := -C(\mu),
\]
where $C(\mu)$ is the Monge-Kantorovich optimal cost for transporting $\mu$ to the uniform probability measure $\nu_P$ 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

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

**Theorem 1.3.** Assume that $\beta > 0$. Then, for any fixed positive time $t$, the empirical measure $\frac{1}{N} \sum \delta_{x_i}$ of the stochastic process \[1.13\] with initial independent data distributed according to a $\mu_0 \in P_2(\mathbb{R}^n)$ converges in law to the deterministic measure $\mu_t = \rho_t \, dx$ evolving by the gradient flow on the Wasserstein space, defined by the functional $F_\beta$ (formula \[1.14\]) and satisfying the evolution PDE in the distributional sense:

$$\partial_\rho \rho_t = -\frac{1}{\beta} \Delta \rho_t + \nabla \cdot (\rho_t \nabla \phi_t)$$

where $\phi_t(x)$ is the unique convex function on $\mathbb{R}^n$ solving the Monge-Ampère equation

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

(in the weak sense of Alexandrov) normalized so that $\phi(0) = 0$ and satisfying the growth condition $\phi(x) \leq \phi_P(x)$, where $\phi_P(x) := \sup_{p \in P} p \cdot x$.

Integrating twice reveals that the stationary (i.e. time independent) equation corresponding to the evolution PDE in the previous equation may be written as follows in terms of the convex “potential” $\phi$:

$$\det(\partial^2 \phi) = e^{-\beta \phi}$$

where $\rho_t \, dx = \rho \, dx := MA(\phi)$. As shown in [8] (generalizing the seminal result in [58]) there is a solution $\phi := \phi_\beta$ to the previous equation iff the origin is the barycenter $b_P$ of $P$, i.e. iff $b_P = 0$ and then the solution is smooth (see also [23] for generalizations). Moreover, the additive group $\mathbb{R}^n$ acts faithfully by translations on the solution space. Note that up to replacing $P$ with $\beta^{-1} P$ we may as well assume that $\beta = 1$ and the corresponding static equation

$$\det(\partial^2 \phi) = e^{-\phi}$$

is precisely the Kähler-Einstein equation for a toric Kähler potential $\phi$ of a Kähler-Einstein metric with positive Ricci curvature on the toric variety $X_P$ corresponding to $P$, in the case when $P$ is a rational polytope. More precisely, $X_P$ is a toric log Fano variety and $\phi$ corresponds to the Kähler potential of a Kähler-Einstein metric with conical singularities along the divisor $X_P - \mathbb{C}^n$ “at infinity” - the ordinary smooth Fano case appears when the polytope $P$ is a reflexive Delzant polytope [58, 8].

Given the relation to Kähler-Einstein metrics it is natural to ask if Theorem 1.3 has a static analog (as in Theorem 1.2)? However, it follows from symmetry considerations involving the action by translations of the additive group $\mathbb{R}^n$, that the corresponding Boltzmann-Gibbs measure

$$\mu^{(N_k)} := \frac{1}{Z_{N_k}} (\text{Per}(x_1, \ldots, x_N))^{-\beta/k} \, dx^\otimes N_k,$$

describing a permanental point processes at negative temperature, is not even well-defined, i.e. the partition function $Z_{N_k}$ diverges! This is a reflection of the fact
that the static equation (1.17) has a multitude of solutions (due to the translation symmetry) which from a statistical mechanical point of view is a sign of a first order phase transition. However, as we will show, this issue can be bypassed through a symmetry breaking mechanism where one introduces a “background potential” \( V(x) \) with appropriate growth at infinity, dictated by \( P \), i.e.

\[
|V(x) - \phi_P(x)| \leq C, \quad \phi_P(x) := \sup_{p \in P} p \cdot x
\]

and replace the interaction energy \( E^{(N_k)}(x_1, ..., x_{N_k}) \) defined by formula (1.12) by the convex combination

\[
E_{V,\gamma}^{(N_k)}(x_1, ..., x_{N_k}) := \gamma \frac{1}{k} \log \text{Per}(x_1, ..., x_{N_k}) + (1 - \gamma) (V(x_1) + \cdots + V(x_{N_k}))
\]

for a given parameter \( \gamma \in [0, 1] \) (which from the point of view of permanental point processes plays the role of minus the inverse temperature). Then the convergence in Theorem 1.3 still holds with \( F \) replaced by

\[
F_{V,\gamma}(\mu) = -\gamma C(\mu) + (1 - \gamma) \int V d\mu + H(\mu)
\]

and the corresponding static equation now becomes

\[
\det(\partial^2 \phi) = e^{-(\gamma \phi + (1 - \gamma)V)} dx,
\]

which has at most one solution for any given \( \gamma \in [0, 1] \) and convex body \( P \). Moreover, if \( P \) satisfies the barycenter condition \( b_P = 0 \) then there exists a solution \( \phi_\gamma \) to the equation (1.21) for any \( \gamma \in [0, 1] \) and as \( \gamma \to 1 \) it follows from the results in [58, 8] that the solutions \( \phi_\gamma \) converges to a particular solution to the Kähler-Einstein equation (1.18) singled out by the potential \( V \).

**Theorem 1.4.** Let \( P \) be a convex body in \( \mathbb{R}^n \) containing 0 in its interior and denote by \( b_P \) the barycenter of \( P \). For any potential \( V \) in \( \mathbb{R}^n \) satisfying the growth condition (1.19) we have

- If \( b_P = 0 \), then, for any \( \gamma \in [0, 1] \) the Gibbs measure \( \mu^{(N)}_{V,\gamma} \) corresponding to the energy function \( E_{V,\gamma}^{(N_k)}(x_1, ..., x_{N_k}) \), i.e.

\[
\mu^{(N)}_{\phi_\gamma} := \frac{1}{Z_{N_k,\phi_\gamma}} (\text{Per}(x_1, ..., x_N))^{-\gamma/k} (e^{-(1-\gamma)V})^{\otimes N_k}
\]

is well-defined and equal to the weak limit, as \( t \to \infty \) of the law of the empirical measures for the corresponding SDEs (1.1). Moreover, as \( N \to \infty \) the corresponding empirical measures converge in law to the deterministic measure \( \mu_\beta \) defined as \( \mu_\gamma = MA(\phi_\gamma) \) for the unique (mod \( \mathbb{R} \)) solution \( \phi_\gamma \) of the equation (1.21)

- More generally, the Gibbs measure above is well-defined precisely for \( \gamma < R_P \), where \( R_P \in [0, 1] \) is the following invariant of \( P \) :

\[
R_P := \frac{\|q\|}{\|q - b_P\|},
\]

where \( q \) is the point in \( \partial P \) where the line segment starting at \( b_P \) and passing through 0 meets \( \partial P \). Moreover, for any such parameter \( \gamma \) the corresponding convergence results still hold.
As indicated in the introduction this result can be viewed as a probabilistic analog of the seminal Yau-Tian-Donaldson conjecture saying that a Fano manifold $X$ admits a Kähler-Einstein metric if and only if $X$ is K-stable. The latter notion is of algebro-geometric nature, but in the toric setting it is equivalent to the corresponding polytope $P$ having zero as its barycenter. Moreover, from the complex geometric point of view the equation 1.21 is precisely the one appearing in Aubin’s continuity method for Kähler-Einstein metrics and, as shown in [42], the invariant $R_P$ of $P$ coincides with the differential geometric invariant $R(X)$ of $X$ defined as the greatest lower bound on the Ricci curvature. In this complex geometric context the last point in the previous theorem implies, as will be explained elsewhere, one of the inequalities in the conjectural equality relating $R(X)$ to the algebro-geometric invariant $\gamma(X)$ introduced in [10]:

$$R(X) \geq \min\{\gamma(X), 1\},$$

which generalizes a very recent purely algebro-geometric result of Fujita [28] (concerning the case $\gamma(X) = 1$) in the case of a toric Fano variety.

From a statistical mechanical point of view the critical value $R_P$ appearing above can be seen as a real analog of the well-known critical value appearing in the study of the Keller-Segel equation as well as in the study of the 2D log gas in [19, 39]. This connection will be further expanded on elsewhere [11], but the main point is that the invariant $R_P$ may also be characterized as the sup over all $\gamma \in [0, \infty]$ such that the free energy type functional $F_\gamma$ is bounded from below (compare [8]).

1.2.1. Relation to other evolution equations and traveling waves. In the one-dimensional case when $P := [-a_-, -a_+]$ integrating 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 [34] 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 non-linear wave equation and a scalar conservation law (which we used, among many other things, as a toy model for turbulence in the Navier-Stokes equations [27]). Interestingly, the barycenter $b_P$ of $P$ coincides with minus the speed $s := (a_+ + a_-)/2$ of the time-dependent solution $u$, in the terminology of scalar conservation laws [40]. Hence, the vanishing condition $b_P = 0$, which as recalled above is tantamount to the existence of a stationary solution, simply means, from the point of view of non-linear 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 deterministic KPZ-equation [37, 31]:

$$\partial_t \phi = \kappa \partial_x^2 \phi + \frac{1}{2} (\partial_x \phi)^2.$$ (1.23)

As is well-known, for $\kappa > 0$ this equation is integrable in the sense that it is linearized by the Cole-Hopf transformation $f := e^{2\kappa \phi}$, which transforms the equation 1.23 to the ordinary linear heat equation.

In the general higher dimensional case the evolution equation 1.15 can be seen as a dissipative viscous/diffusive version of the semi-geostrophic equation appearing
in dynamic meteorology (see \cite{43} and references therein). 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 1.15 can also be seen as a quadratic perturbation (with diffusion) of the “geodesic flow” on the Wasserstein $L^2$-space (compare \cite{2} Example 11.2.10), which in the one dimensional case appears in the Sticky Particle System \cite{51}. As will be shown in a separate publication the large time asymptotics of the fully non-linear evolution equation 1.15 for the probability density $\rho_t$ in $\mathbb{R}^n$ are governed by traveling wave solutions in $\mathbb{R}^n$ whose speed coincide with minus 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 $\rho$ is uniquely determined from a variant of the Monge-Ampère equation 1.18 together with a moment condition determined by the initial data (breaking the translation symmetry). In complex geometric terms $\rho$ corresponds to a certain canonical Kähler-Einstein metric $\omega$ on $X$ with conical singularities “at infinity”, playing the role of Calabi’s extremal metrics in this context. More generally, the results above apply in a more general setting where the measure $\nu_P$ is multiplied by a density $g$, which amount to replacing the Monge-Ampère equation $MA(\phi)$ with $g(\nabla \phi)MA(\phi)$ and which from the point of view of scalar conservation laws corresponds to a general convex flux function $f$ (when $n = 1$).

1.3. 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 \cite{12} drew our attention to the Otto calculus and the theory of gradient flows on the Wasserstein space.

1.4. 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 2.7 (starting with the discretized situation). Then the proof of the corresponding static result, Theorem 1.2 is given. In Section 3 we go on to apply the previous general results to the permanental setting, as introduced in Section 3. Finally, in Section 4 we illustrate that our approach also applies to more singular situations, by considering the case of singular pair interactions.

The rather lengthy setup and preparatory material in Section 2.1-2.6 is due to our effort to make the paper readable to a rather general audience.

## 2. General setup and proof of Theorem 1.1

2.1. Notation. Given a topological (Polish) space $Y$ we will denote the integration pairing between measures $\mu$ 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 case $Y = \mathbb{R}^d$ we will say that a measure $\mu$ has a density, denoted by $\rho$, if $\mu$ is absolutely continuous wrt Lebesgue measure $dx$ and $\mu = \rho dx$.\footnote{In this case the density $\rho$ is a function $\rho : Y \to (0, \infty)$ and $\mu = \int_{\mathbb{R}^d} \rho(x) \delta_\mu(x) dx$.}
We will denote by \( \mathcal{P}(\mathbb{R}^d) \) the space of all probability measures and by \( \mathcal{P}_{ac}(\mathbb{R}^d) \) the subspace containing those with a density. The Boltzmann entropy \( H(\mu) \) and Fisher information \( I(\mu) \) (taking values in \( [-\infty, \infty] \)) are defined by

\[
H(\mu) := \int_{\mathbb{R}^d} (\log \rho) \rho \, dx, \quad I(\mu) = \int_{\mathbb{R}^d} \frac{\nabla \rho}{\rho} \, dx
\]

(assuming that \( \nabla \rho \in L^1(dx) \) and \( \rho^{-1} \nabla \rho \in L^2(dx) \)). More generally, given a reference measure \( \mu_0 \) on \( Y \) the entropy of a measure \( \mu \) relative to \( \mu_0 \) is defined by

\[
H_{\mu_0}(\mu) = \int_{X} \left( \log \frac{\mu}{\mu_0} \right) \mu
\]

if the probability measure \( \mu \) on \( X \) is absolutely continuous with respect to \( \mu \) and otherwise \( H(\mu) := \infty \). The relative Fisher information is defined similarly. Given a lower semi-continuous (lsc, for short) function \( V \) on \( Y \) and \( \beta \in [0, \infty] \) (the “inverse temperature”) we will denote by \( F_\beta^V \) the corresponding (Gibbs) free energy functional with potential \( V \):

\[
F_\beta^V(\mu) := \int_X V \mu + \frac{1}{\beta} H_{\mu_0}(\mu),
\]

which coincides with \( \frac{1}{\beta} \) times the entropy of \( \mu \) relative to \( e^{-V} \mu_0 \). In particular, since by Jensen’s inequality, \( H_{\mu_0}(\mu) \geq 0 \), when \( \mu \) and \( \mu_0 \) are probability measures, with equality iff \( \mu = \mu_0 \), the Boltzmann-Gibbs measure

\[
\frac{e^{-\beta V}}{Z} \mu_0, \quad Z := \int e^{-V} \mu_0
\]

of \( V \), at inverse temperature \( \beta \) is the unique minimizer of \( F_\beta^V \) on the space \( \mathcal{P}(Y) \) of probability measures, under the integrability assumption that \( Z < \infty \) (this is usually called Gibbs’ variational principle).

2.2. 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. \( \mu_j \rightharpoonup \mu \) weakly in \( \mathcal{P}(X) \) iff \( \int_X \mu_j f \to \int_X \mu f \) 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 \( p \in [1, \infty] \):

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

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

\[
X^{(N)} := X^N / S_N, \quad \pi : X^N \to X^{(N)}
\]

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

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

(2.5)

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

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

(2.6)

where $\gamma$ ranges over all couplings between $\mu$ and $\nu$, i.e. $\gamma$ is a probability measure on $X \times X$ whose first and second marginals are equal to $\mu$ and $\nu$, respectively (see Lemma [2.3] below). We will denote $W^p(X, d)$ the corresponding $L^p$-Wasserstein space, i.e. the subspace of $P(X)$ consisting of all $\mu$ 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 $W^p(X, d) = P_p(X)$ 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 [57] $d_{W_p}^p(\mu, \nu)$ is the optimal cost to for transporting $\mu$ to $\nu$ with respect to the cost functional $c(x, p) := d(x, y)^p$. Accordingly a coupling $\gamma$ as above is often called a transport plan between $\mu$ and $\nu$ and it said to be defined be a transport map $T$ if $\gamma = (I \times T)_\ast \mu$ where $T_\ast \mu = \nu$. In particular, if $X = \mathbb{R}^n$, $p = 2$ and $\mu$ and $\nu$ are absolutely continuous with respect to Lebesgue measure, then, by Brenier’s theorem [18], the optimal transport plan $\gamma$ is always defined by a transport map $T(= T^\mu_\nu)$ of the form $T^\mu_\nu = \nabla \phi$, where $\phi$ is a convex function on $\mathbb{R}^n$ (optimizing the dual Kantorovich functional).

We recall the following standard

**Proposition 2.2.** A sequence $\mu_j$ converges to $\mu$ in the distance topology in $W^p(X, d)$ iff $\mu_j$ converges to $\mu$ in the weakly in $P(X)$ and the $p$th moments converge (the latter assumption is automatic if $X$ is compact). In particular, if $\mu_j$ converges to $\mu$ in weakly in $P(X)$ and the $p$th moments are uniformly bounded, then $\mu_j$ converges to $\mu$ in the distance topology in $W^{p'}(X, d)$ for any $p' < p$.

**Proof.** For the first statement see for example [57] Theorem 7.12]. The second statement is certainly also well-known, but for completeness we include a simple proof. Decompose

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

By the assumption and Chebishev’s inequality the second terms may be estimated from above by $C/R^{(p-p')}$ and by the assumption of weak convergence the first
term converges to \( \int_{\{d(x,x_0) \leq R\}} d(x,x_0)^p \mu \) as \( j \to \infty \) (by taking \( f \) to be a suitable regularization of \( 1_{\{d(x,x_0) \leq R\}}d(x,x_0)^p \)). Finally, letting \( R \) tend to infinity concludes the proof. \( \square \)

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 \( \Gamma \) 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 \( \delta_N \) 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 \( \pi_* \) of the quotient projection \( \pi : X^N \to X^{(N)} \) induces an isometry between the subspace of symmetric measures in \( (W_q(X^N),\frac{1}{Nq^q}d_{p}) \) 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 \( (x_1,...,x_N),(y_1,...,y_N) \in X^N \), then

\[
\inf_{\Gamma \in \mathcal{P}(\mathcal{P}(X))} \int c(x,y) d\Gamma = \inf_{\Gamma_N(\mu,\nu)} \int c(x,y) d\Gamma
\]

where \( \Gamma_N(\mu,\nu) \subset \Gamma(\mu,\nu) \) consists of couplings of the form \( \Gamma_\sigma := \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \delta_{y_{\sigma(i)}} \), for \( \sigma \in \Sigma_N \), where \( \Sigma_N \) 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 \( \gamma_2 \) between \( f_*\mu \) and \( f_*\nu \) is of the form \( f_*\gamma_1 \) for some coupling \( \gamma_1 \) between \( \mu \) and \( \nu \). The point is that \( \gamma \) can be taken to be concentrated on \( f(Y_1) \times f(Y_2) \) (since this set contains the product of the supports of \( \mu \) and \( \nu \)) 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 \( d = \frac{1}{Nq^q}d_{X^N,p} \) 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 \( \pi_* \) 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 \( \pi_* \) 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 \( \mu \) and \( \nu \) are \( G \)-invariant (see [44] Lemma 5.36) \( \square \)

Let us also recall the following classical result

**Lemma 2.4.** Let \( \mu_0 \) 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 results even holds in the sense of large deviations at speed $N$ with a rate functional the relative entropy functional $H_{\mu_0}(\cdot)$ [25, Theorem 6.2.10]. We note that (a non-standard) proof of this classical result can be obtained using the argument in the proof of Theorem 1.2 applied to $E = 0$.

2.2.1. The present setting. We will apply the previous setup to $X = \mathbb{R}^n$ endowed with the Euclidean metric $d$. Moreover, we will always take $p = 2$. 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 $\mu_N$ on $X^N$ with a probability measures on the quotient $X^{(N)}$ (as in Lemma 2.3) the second and third point in Lemma 2.3 may (with $q = 2$) be summarized by the following chain of equalities that will be used repeatedly below:

\[
\frac{1}{N} d_2(\mu_N, \mu'_N)^2 = d_2(\mu_N, \mu'_N)^2 = d_2(\Gamma_N, \Gamma'_N)^2,
\]

where $\Gamma_N$ and $\Gamma'_N$ denote the push-forwards under $\delta_N$ of $\mu_N$ and $\mu'_N$, respectively.

2.3. Intermezzo: Otto’s formal Riemannian structure on the $L^2$–Wasserstein space.

2.3.1. The Otto metric. Let us make a brief digression to recall Otto’s [52] beautiful (formal) Riemannian interpretation of the Wasserstein $L^2$–metric $d_2$ on $\mathcal{P}^2(\mathbb{R}^n)$, which motivates the material on gradient flows on $\mathcal{P}^2(\mathbb{R}^n)$ recalled in Section 2.6.

For simplicity we will consider probability measures of the form $\mu = \rho dx$ where $\rho$ is smooth positive everywhere (in order to make the arguments below rigorous one should also specify the rate of decay of $\rho$ at $\infty$ 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 $\rho_t$ in $\mathcal{P}$ at $\rho_0 := \rho$, when $\rho_t$ 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

\[
\dot{\rho}(x) := \frac{d\rho_t(x)}{dt} \bigg|_{t=0}
\]

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 (non-affine) “tangent vector” of $\rho_t$ at $\rho$, i.e. $v_{\dot{\rho}} \in T \rho \mathcal{P}$. First, since the total mass of $\rho_t$ is preserved we have $\int \dot{\rho} dx = 0$ and hence there is a vector field $v$ on $\mathbb{R}^n$ solving the following continuity equation:

\[
\dot{\rho} = -\nabla \cdot (\rho v)
\]

In geometric terms this means that

\[
\rho_t = (F_t^V)_* \rho_0 + o(t),
\]

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

\[
g(v_{\dot{\rho}}, v_{\dot{\rho}}) = \inf_V \int \rho |v|^2 dx = \int \rho |v_{\dot{\rho}}|^2 dx,
\]
which can be seen as the linearized version of the defining formula 2.10 for the Wasserstein $L^2$-metric. By duality the optimal vector field $v_\rho$ may be written as $v_\rho = \nabla \phi$, for unique normalized function $\phi$ on $\mathbb{R}^n$.

2.3.2. 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 $\delta_N$ (Lemma 2.23). Indeed, given a curve $(x_1(t), \ldots, x_N(t))$ in the Riemannian product $(X^N, \frac{1}{N} g^{\otimes N})$ with tangent vector $(\frac{dx_1(t)}{dt}, \ldots, \frac{dx_N(t)}{dt})$ at $t = 0$ we can write its squared Riemannian norm at $(x_1(0), \ldots, x_N(0))$ as

$$\left\| \left( \frac{dx_1(t)}{dt}, \ldots, \frac{dx_N(t)}{dt} \right) \right\|^2 = \int |v|^2 \delta_N(0) \tag{2.11}$$

where $\delta_N(t) := \frac{1}{N} \sum x_{i(t)}$ and $V$ is any vector field on $X = \mathbb{R}^n$ such that $V(x_i) = \frac{dx_i(t)}{dt} \big|_{t=0}$. Note that setting $\rho_t := \delta_N(t)$ the vector field $v$ satisfies the push-forward relation 2.10 (with vanishing error term). Moreover, since passing to the quotient $X^N/S_N$ does not effect the corresponding curve $\rho_t$, minimizing with respect to the action of the permutation group $S_N$ in formula 2.11 corresponds to the infimum defining the Otto metric in formula 2.10.

2.3.3. 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 flow $\nabla G$ satisfies the equation

$$\frac{\partial \rho_t(x)}{\partial t} = \nabla_x \cdot (\rho_v_t(x)), \quad v_t(x) = \nabla_x \frac{\partial G(\rho_t)}{\partial \rho} \big|_{\rho = \rho_t} \tag{2.12}$$

In other words, the gradient flow of $G(\rho)$ may be written as $v_t(x) = \nabla_x \frac{\partial G(\rho_t)}{\partial \rho} \big|_{\rho = \rho_t}$.

In particular, for the Boltzmann entropy $H(\rho)$ (formula 2.11) one gets, since $\frac{\partial G(\rho)}{\partial \rho} = \log \rho$ (using that the mass is preserved) that the corresponding gradient flow is the heat (diffusion) equation and the gradient flow structure then reveals 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 non-negative 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 $F^\beta$ is given by the diffusion equation with linear drift $\nabla_x V$:

$$\frac{\partial \rho_t}{\partial t} = \frac{1}{\beta} \Delta_x \rho_t + \nabla_x \cdot (\rho_t \nabla_x V), \tag{2.13}$$

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 [35] (compare Section 2.3).

2.4. 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:

(1) The functional $E^{(N)}$ is uniformly Lipschitz continuous $(X,d)$ in each variable, or equivalently, under the isometric embedding of $X^{(N)}$ in the $\mathcal{P}(X)$
by the empirical measure $\delta_N$ the sequence $E^{(N)}/N$ extends to define a sequence of functionals on $\mathcal{P}_2(X)$ which are uniformly Lipschitz continuous.

(2) The sequence $E^{(N)}/N$ of functions on $\mathcal{P}_2(X)$ has a unique point-wise limit $E(\mu)$.

(3) The sequence $E^{(N)}$ is $\lambda$-convex on $(X,d)$ in each variable separately (where $\lambda$ is independent of $N$).

Note that since $E(\mu)$ above is assumed Lipschitz continuous on $\mathcal{P}_2(X)$ it is uniquely determined by its restriction to the subspace $\mathcal{P}(X)_c$ consisting of all $\mu$ with compact support.

**Lemma 2.5.** Assume given a sequence $E^{(N)}$ satisfying the uniform Lipschitz assumption (1).

- Then the second point is (2) is equivalent to point-wise convergence of $E^{(N)}/N$ towards $E(\mu)$ for any $\mu$ in $\mathcal{P}(X)$ with compact support.

- The second point (2) implies that

\begin{equation}
\lim_{N \to \infty} \frac{1}{N} \int E^{(N)} \mu^{\otimes N} = E(\mu)
\end{equation}

for any $\mu \in \mathcal{P}_2(X)$.

- Conversely, if $[2.12]$ holds for any $\mu \in \mathcal{P}(X)$ with compact support then $E^{(N)}/N$ converges towards $E(\mu)$ for any $\mu$ in $\mathcal{P}_2(X)$.

**Proof.** Given $\mu \in \mathcal{P}_2(X)$ we define the truncation $\mu_R := \frac{1}{1 \{d(\omega,x) \leq R\}} \int 1 \{d(\omega,x) \leq R\} \mathcal{P}(X)$. By the Lipschitz assumption $|E^{(N)}/N(\mu) - E^{(N)}/N(\mu_R)| \leq Cd_2(\mu,\mu_R)$. In particular, $a_N := (E^{(N)}/N)(\mu)$ is a uniformly bounded sequence in $\mathbb{R}$. Next, letting $N \to \infty$ gives $|a - E(\mu_R)| \leq d_2(\mu,\mu_R)$ for any limit point $a \in \mathbb{R}$ of the sequence $a_N$. Finally letting $R \to \infty$ and using the Lip continuity forces $a = E(\mu)$, as desired and hence $(E^{(N)}/N)(\mu)$ converges towards $E(\mu)$, as desired. To prove formula $2.14$ we first remark that it follows from the general convergence in Proposition $2.18$ below that

\[\lim_{N \to \infty} \frac{1}{N} \int E^{(N)} \mu^{\otimes N} = \int_{\mathcal{P}(X)} E(\nu) \Gamma(\nu),\]

where $\Gamma$ is a weak limit point of $(\delta_N)_{i=1}^\infty$. But by Lemma $2.4$ the limit point is unique and given by $\Gamma = \delta_{\mu}$. Hence, the rhs above is equal to $E(\mu)$, as desired. \( \square \)

### 2.5. The forward Kolmogorov equation for the SDEs and the mean free energy $F_N$

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 $\mathbb{R}^n$. As is well-known, under suitable regularity assumptions the SDEs $\mathbb{R}^n$ on $X^N$ defines, for any fixed $T$, a probability measure $\eta_T$ on the space of all continuous curves (“sample paths”) in $X^N$, i.e. continuous maps $[0,T] \to X^N$ (see for example [55] and reference therein). For $t$ fixed we can thus view $x^{(N)}(t)$ as a $X^N$-valued random variable on the latter probability space. Then its law $\mu^{(N)}_t := (x^{(N)}(t)_{i=0}^\infty \eta_t$
gives a curve of probability measures on $X^{(N)}$ of the form $\mu^{(N)}_t = \rho^{(N)}_t dV$, where the density $\rho^{(N)}_t$ satisfies the corresponding forward Kolmogorov equation:

$$(2.15) \quad \frac{\partial \rho^{(N)}_t}{\partial t} = \frac{1}{\beta} \Delta \rho^{(N)}_t + \nabla \cdot (\rho^{(N)}_t \nabla E^{(N)}),$$

which thus coincides with the linear Fokker-Planck equation $2.13$ on $X^N$ with potential $V := E^{(N)}$. 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 $\mathcal{P}(X^N)$:

$$
\Gamma^{(N)}_t := (\delta^{(N)}_t)^* \mu^{(N)}_t,
$$

where $\delta^{(N)}_t$ is the empirical measure defined by formula $2.5$.

Anyway, for our purposes we may as well forget about the SDEs $1.1$ and take the forward Kolmogorov equation $2.15$ as our the starting point. We will exploit the well-known fact, going back to [35] (see Prop $2.12$ below) that the latter evolution equation can be interpreted as the gradient-flow on the Wasserstein space $W_2(Y)$, of the functional

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

where $H(\cdot)$ is the entropy relative to $\mu_0 := dV^{\otimes N}$ (formula $2.2$); occasionally we will omit the subscript $\beta$ in the notation $F^{(N)}_{\beta}$.

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

$$F_N = E_N + \frac{1}{\beta} H_N,$$

i.e.

$$(2.16) \quad F_N(\mu_N) := \frac{1}{N} F^{(N)}(\mu_N) = \frac{1}{N} \int_{X^N} E^{(N)} \mu_N + \frac{1}{3N} H(\mu_N),$$

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 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 $\int e^{-\epsilon|x|} dx < \infty$ for any given $\epsilon > 0$ one then gets

$$(2.17) \quad H(\mu) \geq -\epsilon \int |x| dx - C\epsilon$$

As a consequence we have the following

**Lemma 2.6.** If the mean energy satisfies the uniform coercivity property

$$(2.18) \quad \frac{1}{N} \int_{X^N} E^{(N)}(\mu_N) \geq -\frac{1}{2\tau_*} d_2(\mu_N, \Gamma_*)^2 - C$$

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

For example, this is trivially the case under a uniform Lipschitz assumption on $E^{(N)}$ (as in the Main Assumptions).
Remark 2.7. The linear forward Kolmogorov equation \[2.15\] 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 \( g_N := \frac{1}{N^2} g \otimes g \) on \( X^N \). Moreover, in our case \( E(x) \) will be symmetric, i.e. \( S_N \)-invariant and hence the flow defined wrt \( (X^N, g_N) \) 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 2.2. Using the isometric embedding defined by the empirical measure \( E \) we can thus view the sequence of flows on \( P(X^N) \) as a sequence of flows on the same (infinite dimensional) space \( W_2(P(X)) \) and this is the geometric motivation for the proof of Theorem 1.1.

2.6. Gradient flows on the \( L^2 \)-Wasserstein space and variational discretizations. In this section we will recall the fundamental results from \[2\] that we will rely on. Let \( G \) a lower semi-continuous function on a complete metric space \((M, d)\). In this generality there are, as explained in \[2\], 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

\[
\frac{d u_t}{dt} = -\nabla G(u_t), \quad \lim_{t \to 0} u(t) = u_0
\]

The strongest form of weak gradient flows on metric spaces discussed in \[2\] concern \( \lambda \)-convex functionals \( G \) and are defined by the property that \( u_t \) satisfies the following Evolution Variational Inequality (EVI)

\[
\frac{1}{2} \frac{d}{dt} d^2(u_t, v) \leq G(v) - G(u(t)) - \frac{\lambda}{2} d^2(\mu_t, \nu)^2 \quad \text{a.e.} \quad t > 0, \quad \forall v : G(v) < \infty
\]

among all locally absolutely continuous curves such that \( \lim_{t \to 0} u(t) = u_0 \) in \((M, d)\).

Then \( u_t \) is uniquely determined, as shown in \[2\] and we shall then say that \( u_t \) is the EVI-gradient flow of \( G \) emanating from \( u_0 \). We recall that \( \lambda \)-convexity on a metric space essentially means that the distributional second derivatives are bounded from below by \( \lambda \) along any geodesic segment in \( M \) (compare below). When \( M \) has non-negative curvature, NC (in the sense of Alexandrov) the existence of a solution \( u_t \) satisfying the EVI was shown by Meyer for any lower-semicontinuous \( \lambda \)-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 \( \mu \) on \( \mathbb{R}^d \) which does not have non-negative curvature (when \( d > 1 \)). Still, as shown in \[2\], the analog of Meyer’s result does hold under the stronger assumption that \( G \) be \( \lambda \)-convex along any generalized geodesic \( \mu_s \) in \( \mathcal{P}_2(\mathbb{R}^d) \). For our purposes it will be enough to consider lsc \( \lambda \)-convex functionals with the property that \( \mathcal{P}_{2,ac}(\mathbb{R}^d) \) is weakly dense in \( \{ G < \infty \} \). Then the \( \lambda \)-convexity of \( G \) means (compare \[2\] Proposition 9.210) that for any generalized geodesic \( \mu_s = \rho_s dx \) in \( \mathcal{P}_{2,ac}(\mathbb{R}^d) \) the function \( G(\rho_s) \) is continuous on \([0, 1]\) and the distributional second derivatives on \([0, 1]\) satisfy

\[
\frac{d^2 G(\rho_s)}{ds^2} \geq \lambda,
\]

where the generalized geodesic \( \mu_s \) connecting \( \mu_0 \) and \( \mu_1 \) in \( \mathcal{P}_{2,ac}(\mathbb{R}^d) \) (with base \( \nu \in \mathcal{P}_{2,ac}(\mathbb{R}^d) \)) is defined as the following family of push-forwards:

\[
\mu_s = ((1 - s) T_0 + s T_1)_* \nu
\]
where $T_s$ is the optimal transport map (defined with respect to the cost function $|x-y|^2/2$) pushing forward $\nu$ to $\mu_t$ (compare Remark 2.1). The bona fide Wasserstein geodesics in $P_{2,ac}(\mathbb{R}^d)$ are obtained by taking $\nu = \mu_0$ (the study of convexity along such geodesics was introduced by McCann [15], who called it displacement convexity).

We will be relying on the following version of Theorem 4.0.4 and Theorem 11.2.1 in [2]:

**Theorem 2.8.** Suppose that $G$ is a lsc real-valued functional on $P_2(\mathbb{R}^d)$ which is $\lambda$-convex along generalized geodesics and satisfies the following coercivity property: there exist constants $\tau, C > 0$ and $\mu_0 \in P_2(\mathbb{R}^d)$ such that

$$G(\cdot) \geq -\frac{1}{\tau} d_2(\cdot, \mu_0)^2 - C$$

Then there is a unique solution $\mu_t$ to the EVI-gradient flow of $G$, emanating from any given $\mu_0 \in \{G < \infty\}$. The flow has the following regularizing effect: $\mu_t \in \{|\partial G| < \infty\} \subset \{G < \infty\}$. Moreover, if $\lambda \geq 0$, then, for any $t > 0$ and $\nu \in P_2(\mathbb{R}^d)$

$$|\partial G|^2(\mu_t) \leq |\partial G|^2(\nu) + d_2(\mu_t, \nu)/t^2,$$

Here $|\partial G|$ denotes the metric slope of $G$, i.e.

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

In fact, many more properties of the EVI-gradient flow $\mu_t$ are established in [2], for example $\mu_t$ defines an absolutely continuous curve $\mathbb{R} \to P_2(\mathbb{R}^n)$ (in the sense of metric spaces) which is locally Lipschitz continuous and has the usual semi-group property. Moreover, under suitably regularity assumptions it shown in [2] that the EVI-gradient flow $\mu_t = \rho_t dx$ furnished by the previous theorem satisfies Otto’s evolution equation in the weak sense:

**Proposition 2.9.** Suppose in addition to the assumptions in the previous theorem that $\mu_t$ has a density $\rho_t$ for $t > 0$. Then $\rho_t$ satisfies the the continuity equation (2.12) 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 $|\partial G|^2 < \infty \subset P_{2,ac}(\mathbb{R}^n)$) the (many-valued) subdifferential $\partial G$ on the subspace $P_{2,ac}(\mathbb{R}^n)$ is a metric generalization of the (Frechet) subdifferential Hilbert space theory; by definition, it satisfies a “slope inequality along geodesics”:

$$(\partial G)(\mu) := \left\{ \xi \in L^2(\mu) : \forall \nu : G(\nu) \geq G(\mu) + \langle \xi, \nu \rangle_{L^2(\mu)} + \frac{\lambda}{2} d_2(\nu, \mu)^2, \quad v(x) := T^v_\mu(x) - x \right\}$$

where $T^v_\mu$ denotes the optimal transport map between $\mu$ and $\nu$, as in Remark 2.1. (note that $v$ is the tangent vector field at 0 of the geodesic $\mu_s$ from $\mu$ to $\nu$). The **minimal subdifferential** $\partial^0 G$ on $P_{2,ac}(\mathbb{R}^n)$ at $\mu$ is defined as the unique element in the subdifferential $\partial G$ at $\mu$ minimizing the $L^2-$norm in $L^2(\mu)$; in fact, its norm coincides with the metric slope of $G$ at $\mu$ (in [2] there is also a more general notion of extended subdifferential which, however, will not be needed for our purposes).
Example 2.10. In the case when $G = H$ is the Boltzmann entropy and $\mu$ satisfies $H(\mu) < \infty$, so that $\mu$ has a density $\rho$, we have $(\partial^2 H)(\mu) = \rho^{-1} \nabla \rho \in L^2(\mu)$ and hence

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

is the Fisher information of $\rho$ (formula (2.21); see [2] Theorem 10.4.17)

The following result goes back to McCann [45] (see also [2] for various elaborations):

Lemma 2.11. The following functionals are lsc and $\lambda-$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 $\lambda-$convex and lsc function $V$ on $\mathbb{R}^d$ (and the converse also holds)
- The functional $\mu \mapsto \int V_N \otimes N$ defined by a given $\lambda-$convex function $V_N$ on $\mathbb{R}^{dN}$ and in particular the “interaction energy” functional

$$\mathcal{W}(\mu) := \int W(x - y)\mu(x) \otimes \mu(x)$$

defined by a given lsc $\lambda-$convex function $W$ on $\mathbb{R}^d$.
- The Boltzmann entropy $H(\mu)$ (relative to $dx$) is lsc and convex along any generalized geodesics.

In particular, for any $\lambda-$convex function $V$ on $\mathbb{R}^d$ the corresponding free energy functional $F_\beta^V$ (formula (2.25) is $\lambda-$convex along generalized geodesics, if $\beta \in ]0, \infty]$.

Combining the results above we arrive at the following

Theorem 2.12. Assume given $\beta \in ]0, \infty]$. Let be $E(\mu)$ a lsc functional on $\mathcal{P}_2(\mathbb{R}^d)$ which is $\lambda-$convex along generalized geodesics and satisfies the coercivity condition [2.27]. Then the EVI-gradient flow $\mu_t$ of the corresponding free energy functional $F_\beta := E + H/\beta$ exists. Moreover, if $\beta < \infty$, then $\mu_t = \rho_t dx$, where $\rho_t$ has finite Boltzmann entropy. In particular,

- If $V$ is a lsc finite $\lambda-$convex function on $\mathbb{R}^d$, then the gradient flow of $F_\beta^V$ exists (defining a weak solution of the corresponding forward Kolmogorov equation/Fokker-Planck equation)
- If $E(\mu)$ is a Lipschitz continuous functional on $\mathcal{P}_2(\mathbb{R}^d)$ which is $\lambda-$convex along generalized geodesics, then the gradient flow exists for any initial data $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$ and if $\beta < \infty$, then $\mu_t = \rho_t dx$, where $\rho_t$ has finite Boltzmann entropy and Fisher information and the following continuity equation holds in the distributional sense on $\mathbb{R}^n \times \mathbb{R}$

$$(2.22) \quad \frac{\partial \rho_t}{\partial t} = \frac{1}{\beta} \Delta \rho_t + \nabla (\rho_t v_t),$$

where $v_t = \partial^0 E$ is the minimal subdifferential of $E$ at $\mu_t = \rho_t dx$.

Proof. By the previous Lemma $F_\beta$ is also lsc and $\lambda-$convex and by Lemma [2.10] it also satisfies the coercivity condition. Hence, the EVI-gradient flow exists according to Theorem [2.8]. Moreover, by the general results in [2] $F_\beta$ is decreasing along the flow and in particular uniformly bounded from above. 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 $\lambda-$convexity $f(x) := v(x) + \lambda|x|^2$ is convex and hence $f(x) \geq$
fine the “discrete flow” $u$ produced by De Georgi, called the minimizing movement that the proof of Theorem 2.8 in [2] uses a discrete approximation scheme introduced by De Georgi, called the minimizing movement scheme. It can be seen as a variational discretization scheme (“minimizing movements”). □

2.6.1. The variational discretization scheme (“minimizing movements”). We recall that the proof of Theorem 2.8 in [2] uses a discrete approximation scheme introduced by De Georgi, 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 $\tau$ (the “time step”). In order to define the “discrete flow” $u^\tau_j$ 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 $u_j \in M$ the next step $u_{j+1}$ is obtained by minimizing the following functional on $(M,d) := W_2(\mathbb{R}^d)$:

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

Next, one defines $u^\tau(t)$ for $t \in [0,T]$ by demanding that $u^\tau(t_j) = u^\tau_j$ and demanding that $u^\tau(t)$ be piece-wise constant and right continuous (we are using a slightly different notion than the one in [2, Chapter 2]).

The curve $u_\tau$ is then defined as the large $m$ limit of $u^{(m)}_t$ in $(M,d)$; as shown in [2] the limit indeed exists and satisfies the EVI 2.20 and is thus uniquely determined. More precisely, the following quantitative convergence result holds (see Theorem 4.07, formuka 4.024] and [2, Theorem 4.09]:

**Theorem 2.13.** Let $G$ be a functional on $\mathcal{P}_2(\mathbb{R}^n)$ satisfying the assumptions in Theorem 2.8 with $\lambda \geq 0$. Then

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

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

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

for some constant $C$ only depending on $\lambda$ and $T$.

**Remark 2.14.** By the last paragraph on page 79 in [2] 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 $G(u_0) < \infty$.

2.7. Proof of propagation of chaos in the discretized setting. In this section we fix once and for all the time interval $[0,T]$ and the time step $\tau > 0$. We denote by $\mu^{(N)}_{ij}$ the corresponding discretized minimizing movement of the free energy functional $F^{(N)}$ on $\mathcal{P}_2(X^N,d_2)$ with given initial data $\mu^{(N)}_{ij}$. The sequence $\mu^{(N)}_{ij}$ is well-defined according to Theorem 2.8 and the Main Assumptions. Moreover, by the third isometry property in Lemma 2.3 $\mu^{(N)}_{ij}$ may 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 $\Gamma^{(N)}_{ij}$ in $W_2(\mathcal{P}_2(X),d_2)$. 

---

$-C|x|$ 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 Lip assumption. Since $F^{(0)}(\mu_0) \leq C$ it follows that $H(\mu_0) < \infty$, which in particular implies that $\mu_t$ has a density $\rho_t$. Moreover, by Theorem 2.8 $|\partial F^{(0)}(\mu_t)| < \infty$ for $t > 0$. But since $E$ is assumed Lip continuous we have $|\partial F^{(0)}(\mu_t)| < \infty$ iff $|\partial H(\mu_t)| < \infty$, which means that $I(\mu_t)$ has finite Fisher information (see Example 2.10). Finally, the distributional equation follows from Proposition 2.39
Theorem 2.15. Assume that at time $t_j$

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

in the $L^2$-Wasserstein metric. 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 the $L^2$-Wasserstein metric.

We recall that given $\mu_{t_j}^{(N)}$ the next measure $\mu_{t_{j+1}}^{(N)}$ is defined as the minimizer of the following functional on $\mathcal{P}(X^N)$:

$$\frac{1}{N} J_{t_{j+1}}^{(N)}(\cdot) := \frac{1}{2\tau_N} \frac{1}{N} d(\cdot, \mu_{t_{j+1}}^{(N)})^2 + \frac{1}{N} P^{(N)}(\cdot)$$

2.7.1. Proof of Theorem 2.15. We start with the following direct consequence of Proposition 2.2 combined with Lemma 2.3:

Lemma 2.16. Let $\mu_N$ 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 $d_q$-distance of $\Gamma_N$ to a fixed element in the Wasserstein space $W_q(\mathcal{P}_2(X))$ is uniformly bounded from above, for some fixed $q \in [1, \infty]$. Then, after perhaps passing to a subsequence, there is a probability measure $\Gamma$ in $W_q(\mathcal{P}_2(X))$ such that

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

weakly in $\mathcal{P}(X)$ or more precisely in $W_{q'}(\mathcal{P}_2(X))$ if $1 \leq q' < q$.

We next recall the following well-known result about the asymptotics of the mean entropy (proved in [53]; see also Theorem 5.5 in [32] for generalizations). The proof is based on the sub-additivity properties of the entropy.

Proposition 2.17. Let $\mu_N^{(N)}$ 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

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

We will also use the following result, which generalizes a result in [48] concerning the case when $E_N$ is quadratic:

Proposition 2.18. Let $\mu_N^{(N)}$ be a sequence of probability measures on $X^N$ such that $\Gamma_N := (\delta_N)_* \mu_N$ converges to $\Gamma$ in $W_1(\mathcal{P}_2(X))$. Then

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

Proof. Recall that the $L^1$-Wasserstein distance $d_1$ on $\mathcal{P}(Y)$, where $Y = \mathcal{P}_2(X)$, admits the following dual representation:

$$d(\mu, \nu) = \sup_{u \in \text{Lip}_1} \int u(\mu - \nu)$$

where $u$ ranges over all Lip-functions on $Y$ with Lip-constant one. By assumption (2.23) $d_1(\Gamma_N, \Gamma) \to 0$. 

23
Using the empirical measure $\delta_N$ we identify $N^{-1}E^{(N)}$ with a uniformly Lipschitz continuous sequence of functions on $\mathcal{P}(X)$ which by the Main Assumptions pointwise to to $E(\mu)$. First observe that since $N^{-1}E^{(N)}$ is uniformly Lipschitz continuous we have that
\[
\lim_{N \to 0} \int_{\mathcal{P}(X)} N^{-1}E^{(N)}(\Gamma_N - \Gamma) = 0
\]
using the dominated convergence theorem in the last step, which applies thanks to the bound $|N^{-1}E^{(N)}| \leq A + Bd_2$ resulting from the Main Assumptions.

Next we turn to the asymptotics of the distances:

**Proposition 2.19.** Assume that a sequence $\nu_N$ of symmetric probability measures on $X_N$ satisfies
\[
\lim_{N \to \infty} (\delta_N)_{\ast} \nu_N = \delta_{\nu}
\]
in the distance topology in $W_2(\mathcal{P}_2(X))$. Then any sequence $\mu_N$ such that $(\delta_N)_{\ast} \mu_N$ 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 iff $(\delta_N)_{\ast} \mu_N$ converges to $\Gamma$ in the distance topology in $W_2(\mathcal{P}_2(X))$.

**Proof.** Consider the isometry
\[
\delta_N : (X^{(N)}, d_{X^{(N)}}) \to (\mathcal{P}(X), d_W) \quad (x_1, \ldots, x_N) \mapsto \delta_N := \frac{1}{N} \sum \delta_{x_i}
\]
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 2.23
\[
\frac{1}{N} d(\mu_N, \nu_N)^2 = d(\delta_N)_{\ast} \mu_N, (\delta_N)_{\ast} \nu_N)^2.
\]
We now first assume that $(\delta_N)_{\ast} \mu_N$ converges to $\Gamma$ in the $d$-distance topology in $W_2(\mathcal{P}_2(X))$. Then the “triangle inequality” for $d$ immediately gives
\[
\lim_{N \to \infty} d((\delta_N)_{\ast} \mu_N, (\delta_N)_{\ast} \nu_N)^2 = d(\Gamma, \delta_{\nu})^2.
\]
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 $\mu$ and $\delta_{y_0}$ is the product $\mu \otimes \delta_{y_0}$. Applied to $Y = \mathcal{P}(X)$ this gives
\[
d((\delta_N)_{\ast} \mu_N, \delta_{\nu})^2 = \int_{\mathcal{P}(X)} d(\mu, \nu)^2 \Gamma(\mu)
\]
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 \( \Gamma \) weakly in \( \mathcal{P}(\mathcal{P}(X)) \), then the lower semi-continuity of the Wasserstein distance function wrt 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 \( \Gamma \)) and then it follows from Proposition 2.22 that \((\delta_N)_* \mu_N\) converges to \( \Gamma \) in the distance topology in \( W_2(\mathcal{P}(X)) \). \( \square \)

### 2.7.2. Conclusion of the proof of Theorem 2.15

First observe that for any fixed \( \mu \) in \( \mathcal{P}(X) \) we have, by the defining property of \( \mu_{(N)} \), that

\[
J_{j+1}(\mu_{(N)}^{(j)})/N \leq J_{j+1}(\mu^{(j)})/N
\]

where the rhs converges, by the propositions above, to \( J_{j+1}(\mu_{j+1}) \) as \( N \to \infty \), where \( J_{j+1}(\mu) = \frac{1}{2N} d(\mu_{j+1}, \mu)^2 + F(\mu_{j+1}) \). In particular,

\[
\limsup_{N \to \infty} J_{j+1}(\mu_{(N)}^{(j)})/N \leq J_{j+1}(\mu_{j+1})
\]

where \( \mu_{j+1} \) is the unique minimizer of \( J_{j+1} \).

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

\[
d_2((\delta_N)_* \mu_{(N)}^{(j+1)}, (\delta_N)_* \mu_{(j)})^2 = \frac{1}{N} d_2(\mu_{(N)}^{(j+1)}, \mu_{(j)})^2 \leq C
\]

Indeed, the minimizing property together with the previous bound gives

\[
\frac{1}{N} d_2((\delta_N)_* \mu_{(N)}^{(j+1)}, (\delta_N)_* \mu_{(j)})^2 \leq C - \frac{1}{N} F(\mu_{(N)}^{(j+1)})
\]

Hence, it is enough to verify that the uniform coercivity property 2.18 holds. But this follows the uniform Lipschitz assumption on \( E^{(N)} \).

Now, it follows from the induction assumption and the triangle inequality for \( d \) that \( \mu_{(N)}^{(j+1)} \) satisfies the assumptions of Lemma 2.22. Accordingly, we may, after passing to a subsequence, assume that \( \mu_N := \mu_{(N)}^{(j+1)} \) converges as in Lemma 2.16 or more precisely that \((\delta_N)_* \mu_{(N)}^{(j+1)} \to \Gamma \) in \( W_1(\mathcal{P}(X)) \), where \( \Gamma \in W_2(\mathcal{P}(X)) \). It then follows from Propositions 2.17, 2.18 and 2.19 that

\[
\liminf_{N \to \infty} J_{j+1}(\mu_{(N)}^{(j+1)})/N \geq \int d(\mu) J_{j+1}(\mu)
\]

Combining the lower and upper bound above and using that \( \mu_{j+1} \) is the unique minimizer of \( J_{j+1} \) then forces \( \Gamma = \delta_{\mu_{j+1}} \) and

\[
\lim_{N \to \infty} J_{j+1}(\mu_{(N)}^{(j+1)})/N = J_{j+1}(\mu)
\]

But this means

\[
\lim_{N \to \infty} (\delta_N)_* \mu_{(N)}^{(j+1)} = \delta_{\mu_{j+1}}
\]
weakly in \( P(X) \) and by the equality \( 2.27 \) that
\[
\lim_{N \to \infty} d(\delta_N, \mu_{t_{i+1}}(N), \delta_{t_{i+1}}) = d(\delta_{t_{i+1}}, \delta_{t_i}).
\]
But then it follows from Proposition \( 2.19 \) (applied to \( \nu = \delta_{t_i} \)) that \( (\delta_N, \mu_N \) converges to \( \Gamma \) in the distance topology in \( W_2(P(X)) \), as desired.

2.8. Convergence in the non-discrete setting: proof of Theorem \( 1.1 \). We first assume that \( \lambda \geq 0 \). By the Main Assumptions the limiting free energy functional \( F(\mu) := E(\mu) + H(\mu) \) is also bounded from below, Lipschitz continuous and \( \lambda \)–convex along generalized geodesics in \( P_2(X) \). Indeed, by \( 2.11 \) \( E(\mu) \) is the limit of the mean energy functionals \( \mu \mapsto \int_X E(\mu) / N \mu \otimes N \) which are \( \lambda \)–convex along generalized geodesics, since \( E(\mu) / N \) is assumed \( \lambda \)–convex (see Lemma \( 2.11 \)). In particular, by Theorem \( 2.12 \) the gradient flow \( \mu_t \) of \( F \) emanating from a given \( \mu_0 \in P(X) \) exists and is uniquely determined in the sense of Theorem \( 2.8 \). We let \( \Gamma_t := \delta_{\mu_t} \) be the corresponding flow on \( P_2(P(X)) \).

Consider the fixed time interval \([0, T]\) and fix a small time step \( \tau > 0 \). Denote by \( \mu^\tau(t) \) the discretized minimized movement of \( F(\mu) \) with time step \( \tau \) and set \( \Gamma^\tau_t := \delta_{\mu^\tau(t)}. \) For any fixed \( t \in [0, T] \) we then have, by the triangle inequality,
\[
d(\Gamma_N(t), \Gamma(t)) \leq d(\Gamma_N(t), \Gamma^\tau_N(t)) + d(\Gamma(t), \Gamma^\tau(t)) + d(\Gamma^\tau_N(t), \Gamma^\tau(t))
\]
By the isometry property in Lemma \( 2.3 \) and the assumed convexity properties we have, by Theorem \( 2.8 \) that \( d(\Gamma_N(t), \Gamma^\tau_N(t)) \leq C \tau \) (uniformly in \( N \)) and \( d(\Gamma(t), \Gamma^\tau(t)) \leq \tau C \). Moreover, by Theorem \( 2.15 \) \( \lim_{N \to \infty} d(\Gamma^\tau_N(t), \Gamma^\tau(t)) = 0 \) for any fixed \( \tau \). Hence, letting first \( N \to \infty \) and then \( \tau \to 0 \) gives \( \lim_{N \to \infty} d(\Gamma_N(t), \Gamma(t)) = 0 \), which concludes the proof.

In the case when \( \lambda \leq 0 \) the previous argument still applies (with the error \( O(\tau) \) replaced by \( O(\tau^{1/2}) \) according to Remark \( 2.14 \)).

2.9. Proof of Theorem \( 1.2 \) (the static case). By Gibbs variational principle (which follows immediately from Jensen’s lemma) \( \mu^{(N)}(\cdot) \) is a minimizer of the mean free energy \( F^{(N)}(\cdot) \). In particular, for any fixed \( \mu \in P(\mathbb{R}^n) \)
\[
\frac{1}{N} F^{(N)}(\mu^{(N)}) \leq \frac{1}{N} F^{(N)}(\mu) = H(\mu) + \int \frac{1}{N} E^{(N)}(\mu) \mu \otimes N \leq C,
\]
where the last inequality is obtained by taking \( \mu = \mu^{(N)} \) to be any measure with compact support. Next observe that by the properness assumption on \( F^{(N)}(\cdot) \) this gives
\[
\int_{\mathbb{R}^n} \frac{1}{N} \sum_{i=1}^N |x_i| \mu^{(N)} \leq A \frac{1}{N} F^{(N)}(\mu^{(N)}) - B \leq AC - B.
\]
However, on order to apply Proposition \( 2.18 \) we would rather need a bound on the \( p \)--moments for some \( p > 1 \):
\[
\int_{\mathbb{R}^n} \frac{1}{N} \sum_{i=1}^N |x_i|^p \mu^{(N)} \leq C_p.
\]
But using the convexity assumption this follows automatically from the bound on the first moments using the following well-consequence of Borell’s lemma \( 14 \) (see \( 49 \), Appendix III), which gives a “Reversed Hölder’s inequality”:
Lemma 2.20. Let $\mu$ be a log concave measure, i.e. $\mu = e^{-\phi}dx$ for some convex functions $\phi$. Then, for any $q > p$:

$$\left( \int_{\mathbb{R}^n} |x|^q \mu \right)^{1/q} \leq C_{p,q,n} \left( \int_{\mathbb{R}^n} |x|^p \mu \right)^{1/p},$$

where the constant $C_{p,q,n}$ is independent of $\mu$.

To prove the bound $2.29$ we first observe that for any $p$

$$\int_{(\mathbb{R}^n)^N} \frac{1}{N} \sum_{i=1}^N |x_i|^p \mu^{(N)} = \int_{\mathbb{R}^n} |x|^p (\mu^{(N)})_1,$$

where $(\mu^{(N)})_1$ denotes the “first marginal” of $\mu^{(N)}$, i.e. its push-forward under the natural projection $(\mathbb{R}^n)^N \to \mathbb{R}^n$ onto the first factor. But, by assumption $\mu^{(N)}$ is log concave and hence, by the Prekopa inequality [57], so is its first marginal $(\mu^{(N)})_1$. Applying the previous lemma thus gives

$$\int_{\mathbb{R}^n} |x|^p (\mu^{(N)})_1 \leq C_{p,n}$$

for any $p \geq 1$ where $C_{p,n}$ is independent of $N$. In particular, by Proposition 2.2 combined with the isometric embedding in Lemma 2.3 we get that $(\delta_N)_* \mu^{(N)} := \Gamma^{(N)}$ converges to some $\Gamma$ in the distance topology in $W_p(\mathcal{P}(X))$ for any $p \in [1, \infty]$. Applying this to $p = 1$ and invoking Proposition 2.17 thus gives the convergence of the mean energy:

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

We then deduce, using the asymptotics of the entropy in Prop 2.17 precisely as in the proof of Theorem 2.15 that

$$\int_{\mathcal{P}(X)} F(\mu) \Gamma(\mu) \leq \liminf_{N \to \infty} \frac{1}{N} F^{(N)}(\mu^{(N)}) \leq \frac{1}{N} F^{(N)}(\mu)$$

Next we claim that there exists a unique minimizer $\mu_*$ on $W_2(X)$ of the functional $F(\mu)$. Accepting this for the moment we get, using the upper bound $2.28$ and Proposition 2.17 applied to $\mu_N = \mu^{N}_*$ that

$$\int_{\mathcal{P}(X)} F(\mu) \Gamma(\mu) \leq \liminf_{N \to \infty} \frac{1}{N} F^{(N)}(\mu^{(N)}) = F(\mu_*) = \inf_{W_2(X)} F(\mu)$$

Hence, it must be that $\Gamma = \delta_{\mu_*}$ which concludes the proof of the convergence assuming the existence and uniqueness of the minimizer $\mu_*$. In fact, the existence of $\mu_*$ also follows from the previous argument: indeed, since, by well-known properties of the entropy [45] we have $H^{(N)}(\mu)/N \geq H(\mu^{(N)})$ the previous argument gives that any limit point $\mu_*$ of the sequence $\mu_1^{(N)}$ in $\mathcal{P}(X)$ (which exists by tightness and as explained above is in $W_p(X)$ for any $p$) satisfies

$$F(\mu_*) \leq \liminf_{N \to \infty} \frac{1}{N} F^{(N)}(\mu^{(N)}) = F(\mu_*) = \inf_{W_2(X)} F(\mu)$$

and hence minimizes $F$ on $W_2(X)$. Finally, since $\mu_1^{(N)}$ is log concave so is the limit $\mu_*$. As for the convergence cr it follows immediately from the formula $-\frac{1}{N} \log Z_N =
and the proof of cr above does not use the existence or uniqueness of a minimizer of $F$.

3. Permanental processes and toric Kähler-Einstein metrics

3.1. Permanental processes: setup. Let $P$ be a convex body in $\mathbb{R}^n$ containing zero in its interior and denote by $\nu_P$ the corresponding uniform probability measure on $P$, i.e. $P = 1_P d\lambda/V(P)$, 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

$$E^{(N_k)}(x_1, \ldots, x_{N_k}) := \frac{1}{k} \log \sum_{\sigma \in \Sigma_{N_k}} e^{k(x_1, p_{\sigma(1)} + \cdots + x_{N_k} p_{\sigma(N_k)})},$$

which, as explained in the introduction of the paper, 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 $N \to \infty$ iff $k \to \infty$.

**Proposition 3.1.** The Main Assumptions for $E^{(N)}$ are satisfied with $\lambda = 0$ and $E = -C(\mu)$, where $C(\mu)$ is the Monge-Kantorovich optimal cost for transporting $\mu$ to the uniform probability measure $\nu_P$ on the convex body $P$, with respect to the standard symmetric quadratic cost function $c(x, p) = -x \cdot p$. Equivalently, formulated in terms of the Wasserstein $L^2$-distance

$$C(\mu) = \frac{1}{2} d_{W_2}(\mu, \nu_P)^2 - \frac{1}{2} \int x^2 d\mu - c_P, \quad c_P := \frac{1}{2} \int p^2 d\nu_P$$

In particular, $C(\mu)$ is convex along generalized geodesics.

**Proof.** This follows from the results in [9]. In fact, the first and second point follows immediately from basic fact that if $\phi_s$ is a family of smooth convex functions on $\mathbb{R}^n$ and $\nu$ a probability measure on the parameter space, then $f := \log \int d\nu(s) e^{\phi_s}$ is also convex and $\nabla \phi$ is contained in the convex support of $\{\nabla \phi_s\}$, which in the present case if contained in $kP$. Hence, $\nabla x, E^{(N)} \in P$ which is uniformly bounded, since $P$ is a convex body and in particular bounded. Finally, the convergence of $E^{(N)}$ was shown in [9] for $\mu$ with compact support (which is enough by Lemma 2.5). The convexity of $C(\mu)$ then follows from Lemma 2.11. Equivalently, this means that $\frac{1}{2} d_{W_2}(\mu, \nu_P)^2$ is $-1$ convex. In fact, as shown in [2] using a different argument $\frac{1}{2} d_{W_2}(\cdot, \nu)^2$ is $-1$ convex for any fixed $\nu \in P_2(\mathbb{R}^n)$. \hfill \Box

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

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

where $d\lambda$ denotes Lebesgue measure and $\partial \phi$ denotes the subgradient of $\phi$ (which defines a multivalued map from $\mathbb{R}^n$ to $\mathbb{R}^n$). In particular, if $\phi \in C^2_{\text{loc}}$, then

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

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

$$(\partial \phi)(\mathbb{R}^n) \subset P$$
Proof. Given formula 3.1 this follows immediately from Theorem 10.4.12 in [2] and the fact that if \( R \) plan (coupling) from \( C \) normalized solution in \( L \) the complex also [8] for a direct variational proof of Brenier’s theorem which can be seen as the map gives back the transport map (see [2, Th 12.4.4] this concludes the proof. See appearing in Theorem 10.4.12 in [2] for the transport plan defined by a transport isfies \( P \) all probability measures in a given point \( \rho dx \), 3.2. □.

Lemma 3.2. The minimal subdifferential of \(-C(\mu)\) on the subspace \( P_{2,ac}(\mathbb{R}^n) \) of all probability measures in \( P_2(\mathbb{R}^n) \) which are absolutely continuous wrt dx, may, at a given point \( p dx \), be represented by the \( L^\infty \) vector field \( \nabla \phi \), where \( \phi \) is the unique normalized solution in \( C_P \) to the equation \( 3.2 \).

Proof. Given formula [3.1] this follows immediately from Theorem 10.4.12 in [2] and the fact that if \( P_{2,ac}(\mathbb{R}^n) \), then Brenier’s theorem gives that the optimal transport plan (coupling) from \( \mathbb{R}^n \) to \( P \) realizing the infimum defining \( dW_2(\mu, \nu_P)^2 \) is given by the \( L^\infty \) map \( \nabla \phi \), where \( \phi \) solves the equation \( 3.2 \). Since the barycentric projection appearing in Theorem 10.4.12 in [2] for the transport plan defined by a transport map gives back the transport map (see [2, Th. 12.4.4]) this concludes the proof. See also [5] for a direct variational proof of Brenier’s theorem which can be seen as the real analogue of the variational approach to complex Monge-Ampère equations in [5]. □.

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

Proposition 3.3. The gradient flow \( \mu_t \) of \( F_\beta \) on \( P_2(\mathbb{R}^n) \) emanating from a given \( \mu_0 \) exists for any \( \beta \in [0, \infty] \). Moreover, for \( \beta < \infty \) we have that \( \mu_t = \rho_t(x) dx \) where \( \rho_t \) has finite Boltzmann entropy and Fisher information and \( \rho(x, t) := \rho_t(x) \) satisfies the following equation in the sense of distributions on \( \mathbb{R}^n \times [0, \infty[ \)

\[
\frac{d\rho_t}{dt} = \frac{1}{\beta} \Delta \rho_t + \nabla \cdot (\rho_t \nabla \phi_t),
\]

where \( \phi_t \) is the unique normalized solution in \( C_P \) to the equation \( 3.2 \) and \( \nabla \phi_t \) defines a vector field with coefficients in \( L^\infty \).

Proof. This follows from Thm 8.3.1 and Cor 11.1.8 in [2]. □.

More generally, as explained in the introduction it is natural to introduce a parameter \( \gamma \in [0, 1] \) and a back-ground potential \( V(x) \), i.e. a convex function on \( \mathbb{R}^n \) satisfying the growth condition \([1.19]\). The one replaces \( E^{(N)} \) with its weighted generalization \( E^{(N)}_{\gamma V} \) defined by formula \([1.20]\). Then the previous proposition still holds with \( F_\beta \) replaced by the corresponding functional \( F_{\gamma V} \) and \( \phi_t \) in the evolution equation \( 3.3 \) is replaced by \( \gamma \phi_t + (1 - \gamma)V \) and with \( \beta = 1 \) (up to rescaling time \( t \) and the potential \( V \) this is equivalent to taking \( \beta = \gamma \)).

3.3. The dynamic setting: Proof of Theorem 1.3. By Proposition 3.1 the Main Assumptions are satisfied.
3.4. The static setting: Proof of Theorem 1.4. Let us first verify the properness assumption in Theorem 1.2 holds for \( \gamma < R_P \). To this end we first assume that \( b_P = 0 \) and observe that

\[
E(N)(x_1, \ldots, x_N) \geq -o(1) \sum_{i=1}^{N} |x_i| - o(1) \tag{3.4}
\]

Indeed, applying Jensen’s inequality to the concave function \( \log \) gives

\[
E(N_k)(x_1, \ldots, x_{N_k}) = \frac{1}{k} \log \frac{1}{N_k!} \sum_{\sigma \in S_{N_k}} e^{k(x_1 p_{\sigma(1)} + \cdots + x_{N_k} p_{\sigma(N_k)})} + \frac{1}{k} \log \frac{1}{N_k!} \geq
\]

\[
\geq \frac{1}{N_k!} \sum_{j=1}^{N} p_{\sigma(j)} - o(1),
\]

But for any fixed \( i \) we have that \( \sum_{\sigma} p_{\sigma(i)} = (N-1)! \sum_{p_j \in \mathbb{P} \cap \mathbb{Z}_k} p_j \) and hence we get a Riemann sum:

\[
\frac{1}{N_k!} (\sum_{\sigma} p_{\sigma(i)}) = \frac{1}{N_k!} \sum_{p_j \in \mathbb{P} \cap \mathbb{Z}_k} p_j := b_p^{(k)} = b_P + o(1),
\]

where \( b_P := \int \mu_N \nu_P \), which is assumed to vanish and hence the inequality 3.3 follows. But then the properness for \( E_{\gamma,V}^{(N)} \) (defined by formula 1.20) in the case \( b_P = 0 \) follows immediately from the definition of \( E_{\gamma,V}^{(N)} \) and the growth assumption 1.19 on \( V \) ensuring that that \( V(x) \geq |x|/C - C \) since 0 is assumed to be an interior point of \( P \). Finally, the case then \( b_P \neq 0 \) can be reduced to the previous case by translating \( P \). More precisely, by the previous argument

\[
E_{\gamma,V}^{(N)} + o(1) \geq \sum_{i=1}^{N} (\gamma + o(1)) x_i \cdot b_P + (1 - \gamma) V(x_i)
\]

But, as shown in the proof of Theorem 2.18 in \( \mathbb{R}^d \) (defined by formula 1.22) is the sup of all \( r \in [0,1] \) such that \( r x \cdot b_P + (1 - r) \phi_P(x) \geq 0 \) and since \( |\phi_P(x) - V(x)| \leq C \) and \( \gamma < R_P \) this gives the desired properness. The convergence of the Boltzmann-Gibbs measures, as \( N \to \infty \), now follows from Theorem 1.2. Moreover, the convergence as \( t \to \infty \) for \( N \) fixed follows from well-known results about the linear Fokker-Planck equation with a convex potential \( E \) such that \( \int e^{-E} dx < \infty \) (see for example [13] and reference therein).

Finally, to prove the divergence of the partition function for \( \gamma = R_P \) we first recall that if \( \psi \) is a convex function on \( \mathbb{R}^d \) then a necessary condition for \( \int e^{-\psi} dx < \infty \) is that \( \psi(x) \to \infty \) as \( |x| \to \infty \) (as follows, for example from Borell’s lemma [14] Lemma 3.1) which gives that the integrability also holds for \( \psi - \epsilon |x| \), for \( \epsilon \) any sufficiently small number [13] Theorem 3.1], and hence, by Jensen’s inequality, \( \psi(x) - \epsilon |x| \geq -C \) := \( -\log \int e^{-\psi(x)} dx \). To violate the previous condition it is clearly enough to find a vector \( a \in \mathbb{R}^d \) such that \( t \mapsto \psi(ta) \) is an affine function on \( \mathbb{R} \). We now consider the convex function \( \psi := E_{R_P,\phi_P} \) on \( \mathbb{R}^{N_k} \) and observe that for any fixed \( a \in \mathbb{R}^N \) we can write, with \( a := (a, a, \ldots, a) \),

\[
(3.6) \quad \psi(a) - \psi(0) = ((1 - R_P) \phi_P(a) + R_P a \cdot b_P) + a \delta_k, \quad \delta_k := R_P(b_p^{(k)} - b_P) \in \mathbb{R}^n
\]

(compare formula 3.5). Moreover, as shown in the proof of Theorem 2.18 in [8] when \( a \) is taken as a normal vector to a facet of \( P \) containing the point \( q \) appearing
in formula \[1.22\] the bracket in formula \[3.6\] vanishes. But this means that \(t \mapsto \psi(ta)\) is an affine function on \(\mathbb{R}\) and hence \(\int e^{-\psi} dx = \infty\). Since \(|\phi_p(x) - V(x)| \leq C\) the divergence also holds when \(\phi_p\) is replaced by \(V\), which concludes the proof.

4. Singular pair interactions

4.1. Setup. Let \(w(s)\) be a lsc and \(\lambda\)–convex real-valued function on \([0, \infty]\) such that there exists positive constants \(A\) and \(B\) with

\[
\liminf_{s \to 0} w(s) \geq -A, \quad \liminf_{s \to \infty} w(s)/s^2 \geq -B
\]

Extend \(w\) to a lsc function \(\mathbb{R} \to \infty\) by demanding that \(w(-s) = w(s)\) for \(s \neq 0\) and \(w(0) := \liminf_{s \to 0} w(s)\). We define the corresponding two-point interaction function by

\[
W(x, y) := w(x - y),
\]

which is called repulsive (attractive) if \(w(s)\) is decreasing (increasing) on \([0, \infty]\).

Given a lsc \(\lambda\)–convex function \(V(x)\) we set

\[
E_{W,V}(x_1, x_2, \ldots, x_N) := \frac{1}{N-1} \sum_{i \neq j} w(x_i - x_j) + V(x_i)
\]

We will consider the general setting of an \(N\)–dependent inverse temperature

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

Then the corresponding SDEs can be formally written as

\[
(4.1) \quad dx_i(t) = -\sum_{j \neq i} (\nabla w)(x_i - x_j)dt - (\nabla V)(x_i)dt + \frac{2}{\beta_N^{1/2}} dB_i(t),
\]

but some care has to be taken when dealing with the singularities which appear when \(x_i = x_j\). Anyway, for our purposes it will be enough to use the EVI gradient flow formulation of the corresponding forward Kolmogorov equations, as in Section 2.5.

We fix a sequence of bounded continuous functions \(w_R\) and \(V_R\) increasing to to \(w\) and \(V\), respectively (which exist by the assumption on lower semi-continuity), where \(R\) will be referred to as the “truncation parameter”. Note however that the corresponding bound lsc function \(W_R(x, y) := w_R(|x-y|)\) may not always be taken to be convex.

**Example 4.1.** (power-laws and the logarithmic case) Our setup applies in particular to the repulsive power-laws

\[
w(s) \sim s^\alpha, \quad \alpha \in ]-1, 0[
\]

whose role in the case \(\alpha = 0\) it played by the repulsive logarithmic potential \(w(s) = -\log s\) and for \(\alpha > 0\) by

\[
w(s) \sim -s^\alpha, \quad \alpha \in ]0, 1[
\]

(the strict lower bound on \(\alpha\) ensures that the corresponding interaction energy \(E_{W,V}^{(N)}\) is in \(L^1_{\text{loc}}\)). The results also apply in the case of the attractive potentials

\[
w(s) \sim s^\alpha, \quad \alpha \in [1, \infty[
\]

(where \(w\) is convex on all of \(\mathbb{R}\)). In particular, in the linear case \(\alpha = 1\) (which gives the 1D Newton potential) our setting applies both to the repulsive case \(-s\) and the attractive case \(s\). In fact, in the repulsive case \(w(s) = -s\) it can be shown by direct
calculation that the corresponding macroscopic energy is \( E_W(\mu) \) is equal to \( C(\mu) \) for \( P = [-1, 1] \), where \( C(\mu) \) is the cost functional appearing in Proposition 3.1 which was shown to be convex using the convexity of the corresponding \( N \)–point interaction energy \( E^{(N)} \). The convexity will also be obtained as a special case of Proposition 4.3 below.

4.2. Propagation of chaos in the large \( N \)–limit and convexity.

**Proposition 4.2.** The functional

\[
E_{W,V}(\mu) := \int_{\mathbb{R} \times \mathbb{R}} W(\mu) \otimes \mu + \int \mathcal{V}(\mu) := W(\mu) + \mathcal{V}(\mu) \in [\infty, \infty]
\]

is well-defined, and lsc on \( \mathcal{P}_2(\mathbb{R}) \) and satisfies the coercivity property 2.21. Moreover,

\[
\frac{1}{N} \int E_{W,V}^{(N)}(x_1, x_2, \ldots, x_N) \mu^{\otimes N} = E_{W,V}(\mu)
\]

and

\[
E_{W,V}(\mu) = \lim_{R \to \infty} E_{W_{R,V}}(\mu)
\]

where \( E_{W_{R,V}}(\mu) \) defines a bounded continuous functional on \( \mathcal{P}(\mathbb{R}) \) (wrt the weak topology). In particular, taking \( \mu = \delta_N(x_1, \ldots, x_N) \) we have

\[
E_{W_{R,V}}(\delta_N(x_1, \ldots, x_N)) + O\left(\frac{C_R}{N}\right) = \frac{1}{N} E_{W_{R,V}}^{(N)}(x_1, x_2, \ldots, x_N)
\]

**Proof.** To simplify the notation we assume that \( V = 0 \), but the general case is similar. First note that the fact that \( E_W(\mu) \) is well-defined is trivial in case \( \mu \) has compact support since then \( W \geq -C \) on the support of \( \mu \). Formula 4.2 then follows immediately from the definition. In the general case we note that fixing \( \delta > 0 \) and setting \( U_\delta := \{(x, y) : |x - y| > \delta\} \) gives \( \int_{U_\delta} W(\mu) \otimes \mu \geq -C_\delta \int |x - y|^2 \mu \otimes \mu \geq 2C_\delta \int (|x|^2 + |y|^2) \mu \otimes \mu \geq C_\delta > \infty \) since \( \mu \in \mathcal{P}_2(\mathbb{R}) \). Hence, \( E_W(\mu) := \int_{\mathbb{R} \times \mathbb{R}} W(\mu) \otimes \mu := \int_{U_\delta} W(\mu) \otimes \mu + \int_{U_\delta^c} W(\mu) \otimes \mu \) is well-defined, since \( W \geq A_\delta \) on \( U_\delta^c \).

The convergence 4.3 as \( R \to \infty \) then follows from the monotone and dominated convergence theorems. Finally, since \( E_W(\mu) \) is an increasing sequence of continuous functionals \( E_{W_R}(\mu) \) it follows that \( E_W(\mu) \) is lower semi-continuous on \( \mathcal{P}_2(\mathbb{R}) \). To prove the last statement we note that, by definition, the error term in question comes from the missing diagonal terms in the definition of \( E_{W_R}^{(N)}(x_1, x_2, \ldots, x_N) \) corresponding to \( i = j \) i.e. from

\[
\frac{1}{N} \frac{1}{N - 1} \sum_{i=1}^{N} w_R(x_i - x_i) = \frac{1}{N} \frac{N}{N - 1} w_R(0) = O\left(\frac{C_R}{N}\right),
\]

which concludes the proof. \( \square \)

**Proposition 4.3.** (convexity). The mean energy functional

\[
\mu^N \mapsto \frac{1}{N} \int_{\mathbb{R}^N} E_{W,V}^{(N)}(\mu^N)
\]

on the subspace of symmetric probability measures in \( \mathcal{P}_{2,ac}(\mathbb{R}^N) \) is \( \lambda \)–convex along generalized geodesics with symmetric base \( \nu_N \). In particular, the functional \( E_{W,V}(\mu) \) is lsc and \( \lambda \)–convex along generalized geodesics in \( \mathcal{P}_{2,ac}(\mathbb{R}^N) \) and satisfies the coercivity condition 2.21.
Proof. We will write $x := (x_1, ..., x_N)$ etc. Let $\mu_0^{(N)}$, $\mu_1^{(N)}$ and $\nu^{(N)}$ be three given symmetric measures in $P_{2,ac}(\mathbb{R}^N)$ and denote by $T_0$ and $T_1$ the optimal maps pushing forward $\nu^{(N)}$ to $\mu_0^{(N)}$ and $\mu_1^{(N)}$, respectively. Let $T_t := (1-t)T_0 + tT$ so that $\mu_t := T_t\nu^{(N)}$ is the corresponding generalized geodesic. The key point of the proof is the following

Claim: (a) $T_t$ commutes with the $S_N$–action and (b) $T_t$ preserves order, i.e. $x_i < x_j$ iff $T(x)_i < T(x)_j$.

The first claim (b) follows directly from Kantorovich duality [35, 57]. Indeed, $T_i$ (for $i \in \{0, 1\}$) is an optimal transport map iff $T_i = \nabla \phi_i$ where the convex function $\phi_i$ on $\mathbb{R}^N$ minimizes the Kantorovich functional $J_i$ corresponding to the two $S_N$–invariant measures $\mu_i^{(N)}$ and $\nu^{(N)}$. But then it follows from general principles that the minimizer can also be taken $S_N$–invariant. To prove the claim (b) we will use the well-known fact that any optimal map $T$ is cyclical monotone and in particular for any $x$ and $x'$ in $\mathbb{R}^N$

$$|x - T(x)|^2 + |x' - T(x')|^2 \leq |x - T(x')|^2 + |x' - T(x)|^2$$

(as follows from the fact that $T$ is the gradient of a convex function). In particular, denoting by $\sigma (= (ij)) \in S_N$ the map on $\mathbb{R}^N$ permuting $x_i$ and $x_j$ we get,

$$|x - T(x)|^2 + |\sigma x - T(\sigma x)|^2 \leq |x - T(\sigma x)|^2 + |\sigma x - T(x)|^2$$

But since (by (a)) $T \sigma = \sigma T$ and $\sigma$ acts as an isometry on $\mathbb{R}^N$ the left hand side above is equal to $2|\sigma x - T(x)|^2$ and similarly, since $\sigma^{-1} = \sigma$ the right hand side is equal to $2|\sigma x - T(x)|^2$. Hence setting $y := T(x)$ gives

$$|x - y|^2 \leq |\sigma x - y|^2$$

Finally, expanding the squares above and using that $\sigma = (ij)$ gives $-2(x_i y_i + x_j y_j) \leq -2(x_j y_i + x_i y_j)$ or equivalently: $(x_i - x_j)(y_i - y_j) \geq 0$, which means that $x_i < x_j$ iff $y_i < y_j$ and that concludes the proof of (b).

Now, by the previous claim the map $T_1$ preserves the fundamental domain $\Lambda := \{x : x_1 < x_2 < ... < x_N\}$ for the $S_N$–action on $\mathbb{R}^N$. But, by assumption, on the subset $\Lambda$ the function $E^{(N)}$ is convex and this is enough to run the usual argument to get convexity of the mean energy on the subspace of symmetric measures. Indeed, we can decompose

$$\int_{X^N} E^{(N)}(\mu_t^{(N)}) = \sum_{\sigma} \int_{\sigma(\Lambda)} E^{(N)}(T_t)_* \nu^{(N)}$$

For any fixed $\sigma$ the integral above is equal to $\int_{\sigma(\Lambda)} T_t^* E^{(N)}(\nu^{(N)})$ (since $T_t$ preserves $\sigma(\Lambda)$) which, by the $S_N$–invariants of $\nu^{(N)}$ and $E^{(N)}$ in turn is equal to $\int_{\Lambda} T_t^* E^{(N)}(\nu^{(N)})$. But since $E^{(N)}$ is convex on $\Lambda$ and $T_t$ preserves $\Lambda$ the function $T_t^* E^{(N)}$ is convex in $t$ for any fixed $x \in \Lambda$ and hence, by the decomposition [4.4]

$$\int_{X^N} E^{(N)}(\mu_t^{(N)})$$

is convex wrt $t$, as desired.

Finally, the convexity of $E(\mu)$ follows immediately by taking $\mu^{(N)}$ to be a product measure $\mu^{\otimes N}$ and using formula [4.2].

Remark 4.4. The first convexity statement may appear to contradict the second point in Lemma [2.11] which seems to force $E^{(N)}$ to be convex on all of $\mathbb{R}^N$ (which will not be the case in general). But the point is that we are only integrating against symmetric measures. As for the convexity result for $E_{W,V}(\mu)$ is indeed well-known
that it holds precisely when the symmetric function \( w(x) \) is convex on \([0, \infty[\) as can be proved directly by using that in this special case \( T(x) = (f(x_1), f(x_2), \ldots, f(x_N)) \)

Clearly preserves order since \( f \) being the derivative of a convex function is clearly increasing.

Given a sequence \( \beta_N \in [0, \infty[ \) converging to \( \beta \in [0, \infty[ \) we define the corresponding

mean free energy functional \( F_{\beta_N}^{(N)} / N \) on the space of symmetric probability measures on \( \mathbb{R}^N \) by

\[
F_{\beta_N}^{(N)}(\mu_N) := \int_{\mathbb{R}^N} E_{W,V}^{(N)} + \frac{1}{\beta_N} H^{(N)}(\mu_N)
\]

Similarly, the corresponding (macroscopic) free energy functional on \( \mathcal{P}(\mathbb{R}) \) is defined by

\[
F_{\beta}(\mu) := E_{W,V}(\mu) + \frac{1}{\beta} H(\mu)
\]

Combining the previous proposition with Theorem 2.12 shows that the EVI-gradient flows \( \mu_t \) of \( F_{\beta_N}^{(N)} \) and \( F_{\beta}^{(N)} \) on \( \mathcal{P}_2(\mathbb{R}^N) \) and \( \mathcal{P}_2(\mathbb{R}) \), respectively exist for appropriate initial measures \( \mu_0 \).

**Theorem 4.5.** Let \( W \) and \( V \) be an interaction energies and potential as in Section 4.1 and denote by \( \mu_t^{(N)} \) the corresponding probability measures on \( \mathbb{R}^n \) evolving according to the forward Kolmogorov equation associated to the stochastic process \( \mu_t \) Assume that at the initial time \( t = 0 \)

\[
\lim_{N \to \infty} (\delta_N)_* \mu_t^{(N)} = \delta_{\mu_0}
\]

in the \( L^2 \)-Wasserstein metric. Then, at any positive time

\[
\lim_{N \to \infty} (\delta_N)_* \mu_t^{(N)} = \delta_{\mu_t}
\]

in the \( L^2 \)-Wasserstein metric, where \( \mu_t \) is the EVI-gradient flow on \( \mathcal{P}_2(\mathbb{R}) \) of the free energy functional \( F_{\beta} \), emanating from \( \mu_0 \).

**Proof.** The upper bound 2.24 follows precisely as before, using formula 4.2. The only new feature appears when proving the lower bound 2.20 where the convergence of mean energies appearing in Proposition 2.18 has to be verified in the present setting. More precisely, it is enough to prove the lower bound

\[
\lim \inf_{N \to \infty} \frac{1}{N} \int_{\mathbb{R}^N} E^{(N)}(\mu_{t+1}^{(N)}) / N \geq \int \text{d}\Gamma(\mu) E(\mu)
\]

To this end we fix the truncation parameter \( R > 0 \) and observe that, since \( E_{W,V}^{(N)} \geq E_{W,R,V_r}^{(N)} \), formula 4.3 gives

\[
E_{W,V}^{(N)}(\mu_N) / N \geq \int E_{W,R,V_r}(\delta_N(x_1, \ldots, x_N)) \mu_N + C_R / N
\]

But

\[
\int E_{W,R,V_r}(\delta_N(x_1, \ldots, x_N)) \mu_N = \int_{\mathcal{P}(\mathbb{R})} E_{W,R,V_r}(\delta_N)_* \mu_N \to \int_{\mathcal{P}(\mathbb{R})} E_{W,R,V_r}(\mu)_* \Gamma(\mu)
\]

as \( N \to \infty \), since \( E_{W,R,V_r}(\mu) \) is a bounded and continuous. Hence,

\[
\lim \inf_{N \to \infty} \frac{1}{N} \int_{\mathbb{R}^N} E^{(N)}(\mu_{t+1}^{(N)}) / N \geq \int \Gamma(\mu) E_{W,R,V_r}(\mu)
\]
for any $R > 0$. Finally, letting $R \to \infty$ and using the monotone convergence theorem concludes the proof of the inequality.

Under suitable regularity assumptions it follows from [22, 20, 16, 17] that the limiting EVI-gradient flow appearing in the previous theorem solves the drift-diffusion equation in the sense of distributions with drift vector field

$$v[\mu_t](x) = 2^n \int (\nabla w) (x - y) \mu_t(y),$$

where the quotation marks indicate that a suitable regularized version of the integral has to be used to deal with the singularity at $x = y$. For example in the logarithmic case one uses the principal value (so that $v[\rho]$ is the Hilbert transform; see [20]).

**References**

[1] Luigi Ambrosio, Maria Colombo, Guido De Philippis, Alessio Figalli: A global existence result for the semigeostrophic equations in three dimensional convex domains. [http://arxiv.org/pdf/1205.5435.pdf](http://arxiv.org/pdf/1205.5435.pdf)

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savare. Gradient flows in metric spaces and in the space of probability measures Lectures in Mathematics ETH Z"urich. Birkh"auser Verlag, Basel, 2005.

[3] D. Bakry and M. Emery: Diffusions hypercontractives , in S’eminaire de probabilite’es, XIX, 1983/84, vol. 1123, Springer, Berlin, 1985, pp. 177-20

[4] Biane, R. Speicher, Free diffusions, free energy and free Fisher information, Ann. Inst. H. Poincar’ e Probab. Stat. 37 (2001), 581-606

[5] Berman, R.J; Boucksom, S; Guedj,V; Zeriahi: A variational approach to complex Monge-Ampere equations. Publications math. de l’IHES (2012): 1-67, November 14, 2012

[6] Berman, R.J: Kahler-Einstein metrics emerging from free fermions and statistical mechanics. 22 pages, J. of High Energy Phys. (JHEP), Volume 2011, Issue 10 (2011)

[7] R.J.Berman: A thermodynamical formalism for Monge-Ampere equations, Moser-Trudinger inequalities and Kahler-Einstein metrics. Advances in Math. (2013) 1254. Volume: 248, 2013

[8] Berman, R.J: Real Monge-Ampere equations and Kahler-Ricci solitons on toric log Fano varieties arXiv:1207.6128, 2012 .Ann. de Fac. Toulouse (to appear)

[9] Berman, R.J: Statistical mechanics of permanents, real-Monge-Ampère equations and optimal transport. arXiv preprint arXiv:1302.4045

[10] Berman, R.J: Kähler-Einstein metrics, canonical random point processes and birational geometry. [http://arxiv.org/abs/1307.3634](http://arxiv.org/abs/1307.3634)

[11] Berman, R.J: Lu, C.H: Drift-diffusion equations in complex geometry and convergence towards Kähler-Einstein metrics.

[12] A Blanchet, EA Carlen, JA Carrillo- Functional inequalities, thick tails and asymptotics for the critical mass Patlak–Keller–Segel model. Journal of Functional Analysis, 2012

[13] F. Bolley, I Gentil, A Guillin: Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations. Journal of Functional Analysis, 2012 - Elsevier

[14] Borell, C. Convex measures on locally convex spaces. Ark. Mat., 12:239–252, 1974

[15] Bott, R: On a theorem of Lefschetz. Michigan Math. J. Volume 6, Issue 3 (1959), 211-216.

[16] G. A. Bonaschi. Gradient flows driven by a non-smooth repulsive interaction potential. Master’s thesis, Universiy y of Pavia, Italy, 2011. [arXiv:1310.3677](http://arxiv.org/abs/1310.3677)

[17] GA Bonaschi, JA Carrillo, M Di Francesco: Equivalence of gradient flows and entropy solutions for singular nonlocal interaction equations in 1D. [http://arxiv.org/pdf/1310.4110.pdf](http://arxiv.org/pdf/1310.4110.pdf)

[18] Brenier, Y: Polar factorization and monotone rearrangement of vector valued functions. Communications on pure and applied mathematics, 1991

[19] Caglioti,E; Lions, P-L; Marchioro.C; Pulvirenti.M: A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. Communications in Mathematical Physics (1992) Volume 143, Number 3, 501-525

[20] JA Carrillo, LCF Ferreira, JC Precioso: A mass-transportation approach to a one dimensional fluid mechanics model with nonlocal velocity. Advances in Mathematics, 2012
[21] J. A. Carrillo, Y.-P. Choi, M. Hauray: The derivation of Swarming models: Mean-Field Limit and Wasserstein distances. http://arxiv.org/pdf/1304.5776.pdf

[22] J. A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev: Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. Duke Math. J. Volume 156, Number 2 (2011), 229-271

[23] D Cordero-Erausquin, B Klartag: Moment measures. arXiv:1304.0630.

[24] Dawsont, J Gärtner: Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics Volume 20, Issue 4, 1987

[25] Dembo, A; Zeitouni O: Large deviation techniques and applications. Jones and Bartlett Publ. 1993

[26] Donaldson, S.K. Scalar curvature and stability of toric varieties. J. Diff. Geom. 62 (2002), 289-349

[27] U Frisch, J Bec: Burgulence. Les Houches 2000: New Trends in Turbulence. http://arxiv.org/abs/nlin/0012033

[28] Fujita, K: On Berman-Gibbs stability and K-stability of Q-Fano varieties. arXiv:1501.00248

[29] A. Guionnet, Large random matrices: Lectures on macroscopic asymptotics , Springer, 2008

[30] Funaki, T: Stochastic Interface Models. In: Lectures on Probability Theory and Statistics, Ecole d’Eté de Probabilités de Saint-Flour XXIII - 2003 (ed. J. Picard), 103–274, Lect. Notes Math., 1869 (2005), Springer.

[31] M. Hairer: Solving the KPZ equation, Annals of Mathematics, 178 (2013), no. 2, pp. 559–664

[32] Hauray,M; Mischler, S: On Kac’s chaos and related problems. arXiv:1205.3518. Journal of Functional Analysis, 2014

[33] Helffer ,B.and Sjöstrand , J: On the correlation for Kac-like models in the convex case. (1994). J. Statist. Phys. 74 349–409

[34] Hopf E. The partial differential equation $u_t + uu_x = \mu u_{xx}$. Comm. Pure Appl. Math. 1950

[35] Richard Jordan, David Kinderlehrer, and Felix Otto: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal., 29(1):1–17 (electronic), 1998.

[36] Kac, M.: M. Foundations of kinetic theory. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III (Berkeley and Los Angeles, 1956), University of California Press, pp. 171–197

[37] M. Kardar, G. Parisi, and Y.-C. Zhang: Dynamic Scaling of Growing Interfaces, Physical Review Letters, Vol. 56, 889–892 (1986). APS

[38] E. F. Keller and L. A. Segel, Initiation of slime mold aggregation viewed as instability, J. theor. Biol. 26 (1970), 399-415.

[39] Kiessling M.K.H.: Statistical mechanics of classical particles with logarithmic interactions, Comm. Pure Appl. Math. 46 (1993), 27-56.

[40] PD Lax: Hyperbolic systems of conservation laws and the mathematical theory of shock waves. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 11. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1973. v+48 pp.

[41] S Li, XD Li, YX Xie: Generalized Dyson Brownian motion, McKean-Vlasov equation and eigenvalues of random matrices. arXiv:1303.1240, 2013

[42] Li, C: Greatest lower bounds on Ricci curvature for toric Fano manifolds. Adv. Math. 226 (2011), no. 6, 4921–4932

[43] Loeper, G: A fully non-linear version of Euler incompressible equations: the Semi-Geostrophic system. SIAM Journal of Math Analysis (to appear).

[44] J Lott, C Villani: Ricci curvature for metric-measure spaces via optimal transport. Annals of Math, Volume 169 (2009), Issue 3, 903-991

[45] McCann, Robert J: A convexity principle for interacting gases. Adv. Math. 128 (1997), no. 1, 153–179.

[46] HP McKean Jr : A class of Markov processes associated with nonlinear parabolic equations. Proceedings of the National Academy of Sciences. 1966

[47] HP McKean Jr: Propagation of chaos for a class of non-linear parabolic equations. Stochastic Differential Equations (Lecture Series in . . . , 1967

[48] Messer, J; Spohn, H: Statistical mechanics of the isothermal Lane-Emden equation. J. Statist. Phys. 29 (1982), no. 3, 561–578.

[49] V. D. Milman and G. Schechtman, Asymptotic Theory of Finite Dimensional Normed Spaces , Lecture Notes in Math. 1200 (1986), Springer, Berlin.
[50] S Mischler, C Mouhot, B Wennberg: A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. Probability Theory and Related Fields, 2011
[51] Natile, L; Savaré, G: A Wasserstein Approach to the One-Dimensional Sticky Particle System. SIAM J. Math. Anal., 41(4), 1340–1365.
[52] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. Comm. Partial Differential Equations , 26(1-2):101–174, 2001
[53] Robinson, D. W., and Ruelle, D. Mean entropy of states in classical statistical mechanics. Comm. Math. Phys. 5 (1967), 288–300.
[54] L.C.G. Rogers, Z. Shi, Interacting brownian particles and the Wigner law, Probab. Theory Related Fields 95,4(1993), 555-570
[55] Sznitman, A-S: Topics in propagation of chaos. École d'Été de Probabilités de Saint-Flour XIX—1989, 165–251, Lecture Notes in Math., 1464, Springer, Berlin, 1991
[56] S.R.S. Varadhan: Entropy methods in hydrodynamic scaling , in “Proceedings of the International Congress of Mathematicians”, Vol. 1, (Zürich, 1994), 196–208, Birkhäuser, Basel, 1995
[57] Villani, C: Topics in optimal transportation. Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003. xvi+370 pp
[58] Wang, X; Zhu, X: Kähler–Ricci solitons on toric manifolds with positive first Chern class, Advances in Mathematics 188 (2004), 87–103

E-mail address: robertb@chalmers.se, onnheimm@chalmers.se

Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden