Thesis for the Degree of Licentiate of Engineering

Permanental Point Processes on Real Tori

Jakob Hultgren

CHALMERS | UNIVERSITY OF GOTHENBURG

Division of Mathematics
Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden, 2016

Permanental Point Processes on Real Tori
Jakob Hultgren

Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg
SE–412 96 Gothenburg
Sweden
Phone: +46 (0)31-772 10 00

Typeset with LaTeX.
Printed in Gothenburg, Sweden, 2016.

Abstract

The main motivation for this thesis is to study *real Monge-Ampère equations*. These are fully nonlinear differential equations that arise in differential geometry. They lie at the heart of optimal transport and, as such, are related to probability theory, statistics, geometrical inequalities, fluid dynamics and diffusion equations. In this thesis we setup and study a thermodynamic formalism for a certain type of Monge-Ampère equations on real tori. We define a family of permanental point processes and show that their asymptotic behavior (when the number of particles tends infinity) is governed by Monge-Ampère equations.

Acknowledgments

There is a tremendous amount of work going in to a thesis like this. It goes without saying that it is not the achievement of only one person, and that completing it requires great amounts of support from people around you.

First I want to thank my advisor Robert Berman. For introducing me to this topic (and countless other topics for that matter) and for supporting me and guiding me through this incomprehensible jungle which is the life of new PhD students. Thanks also for being an inspiring person and sharing your enthusiasm for mathematics. Secondly I would like to thank my co-advisor Bo Berndtsson whose interest in and comments on the thesis made the writing process a really fun and stimulating experience.

Thanks to my fellow PhD students. Both old and new. Oskar, Hossein, Dawan, Peter, Malin, Christoffer, Ivar. Thanks Elizabeth and David for friendship and support which I value a lot. Thanks Magnus for disturbing me with fascinating question. And for teaching me how to drive. Thanks to my family for being the most amazing hosts (in Kalmar, Malmö and Gothenburg) and always welcoming me when I need to escape from Chalmers and my job. Finally, thanks Elina for putting up with me and always taking me seriously, even at times when I don't deserve it.

Jakob Hultgren
Gothenburg, January 2016

Contents

Introduction

We will start this introduction with an example from the Feynman Lectures on Physics [3]. The example treats a mono-atomic gas. A mono-atomic gas is a gas where each particle consists of a single atom. Some examples of gases that are mono-atomic at room temperature are helium, neon, argon, krypton, xenon and radon. Now, two quantities that are easy to measure in a gas are its volume, V, and its pressure, P. Imagine we have a box containing a mono-atomic gas. Now, by moving one of the walls of the box we may alter the volume of the box without letting any gas in to or out of the box. We would expect a decrease in pressure if we enlarge the box and we would expect a rise in pressure if we make the box smaller. By making experiments it is possible to verify the following law:

$$PV^{5/3} = C \tag{0.1}$$

where C is a constant.

On the other hand, we also know that a gas consists of a large number of small particles. Moreover, these particles are assumed to, to a large extent, satisfy Newtons laws of motion. From this point of view, the pressure the gas exerts on the walls of the box is caused by the numerous collisions between the particles and the walls. Moreover, the assumption that the gas is mono-atomic means that there is no motion inside the particles, in other words all the kinetic energy of the system is tied up in motion of the particles themselves.

This means we have one *large scale* perspective, given by (0.1), and one *small scale* perspective, given by the particle model above. How can we unify these two perspectives? Feynman proceeds to explain how to do this. He uses probabilistic arguments to derive (0.1) from the assumptions in the small scale description. He starts by noting that the number of collisions per second involving a certain wall should be proportional to the density of particles and the average velocity in the direction normal to the wall. Moreover, the impact on the wall in each collision is proportional to the the velocity and the mass of the particle. If v_n^2 denotes the average squared velocity of the particles in the direction normal to the wall, m denotes the total mass of the particles in the box and P denotes the pressure the gas exerts on the wall, he arrives at

$$P = \frac{mv_n^2}{V},$$

or equivalently

$$PV = mv_n^2.$$

Moreover, let v^2 denote the average squared velocity of the particles and v_x^2, v_y^2 and v_z^2 denote the average squared velocity in the directions of an orthonormal basis. Then we have

$$v^2 = v_x^2 + v_y^2 + v_z^2.$$

and, by symmetry,

$$v_x^2 = v_y^2 = v_z^2 = v_n^2 = \frac{v^2}{3}.$$

This means

$$PV = \frac{mv^2}{3}$$

and if we let $U = mv^2/2$ be the kinetic energy of the system then

$$U = \frac{3}{2}PV.$$

As the work done when moving the wall is given by PdV and this should be transformed into kinetic energy we get

$$-PdV = dU = \frac{3}{2}(PdV + VdP).$$

Rearranging the terms in this differential equation gives

$$0 = \frac{5}{3}\frac{dV}{V} + \frac{dP}{P}.$$

Integrating we get

$$c = \frac{5}{3}\log V + \log P$$

for some constant c. This implies equation (0.1) and we can conclude that it is a consequence of the particle model.

Now, this thesis is not about mono-atomic gases. However, similarly as the example above, this thesis connects a small scale description of something with a large scale description of the same thing. The system we are studying is given by a space X and a type of particles on X. The small scale description of the system is given by a rule for how the particles on X move and interact. The rule is probabilistic in the sense that given a certain number of particles on X, the rule tells us the probability that we will find them in a certain configuration. Mathematically, this is specified by a sequence of probability measures $\{\mu^{(N)}\}$, each defined on the configuration space X^N, where N denotes the number of particles. These measures will be symmetric, in other words, they are invariant under permutations of the coordinates on X^N. This reflects the fact that the particles are identical.

The large scale description of the system is given by a certain differential equation. The solution to this differential equation defines a probability measure μ_*. In its simplest form, the main result of the thesis says that if we sample the position of a large number of particles on X and restrict our attention to the first particle, then the result is very close to sampling a point according to μ_*. However, the full result is stronger than that. In general, there will be interaction between the particles. In other words, the probability of finding one particle in a certain position depend on the position of the other $N-1$ particles. The result says that if we sample a large number of particles on X and restrict our attention to the first d particles, where d is a number that is much smaller than the number of particles, then the result will be close to sampling d points independently and according to μ_*.

The differential equation that provides the large scale description of our system is a so called *real Monge-Ampère equation*. It is a fully non-linear partial differential equation of second order. The particular class of Monge-Ampère equations we are studying are interesting for their connections to complex geometry.

In the following two sections we will state this result in a more precise way. Then, in Section 3 we will explain a very important tool used in the proof, namely so called *large deviation principles*. In Section 4 we will give some background to real Monge-Ampère equations. In Section 5 we will explain what the variational approach to differential equations is and how it ties in with the large deviation principles. In Section 6 we will explain a tool used to prove large deviation principles. In Section 7 we will explain some of the background for the small scale description we are using. It is closely related to certain point processes in complex geometry modeled on so called Fekete points (see Section 7.1).

1. A More Precise Description of the Setting

As mentioned above the setting we are in is given by a space X. This will be the real n-dimensional torus

$$X = \mathbb{R}^n/\mathbb{Z}^n.$$

Let $C(X)$ be the space of continuous, real valued functions on X, $\mathcal{M}(X)$ be the space of finite signed measures on X and $\mathcal{M}_1(X)$ be the space of probability measures on X. When working with X we will always use the standard coordinates inherited from $\mathbb{R}^n$.

Given a twice differentiable function ϕ on X we can form its Hessian matrix (ϕ_{ij}) where the second derivatives are taken with respect to the standard coordinates on X. This will allow us to define the Monge-Ampère operator. The original Monge-Ampère operator, which is defined for convex functions on $\mathbb{R}^n$, is given by the determinant of the Hessian of a function. Now, since X is compact we get, since a convex function can not have a maximum in the interior of its domain, that all convex functions on X are constant. It is thus necessary to relax the assumption of convexity. The domain of the Monge-Ampère operator on X is the space of so called *pseudo convex* functions, satisfying

$$(\phi_{ij} + \delta_{ij}) > 0$$

where δ_{ij} is the Kronecker delta. If we let dx be the standard volume form on X then the Monge-Ampère operator is defined as

$$\mathrm{MA}(\phi) = \det(\phi_{ij} + \delta_{ij})dx.$$

The differential equation that constitutes the large scale description of our system is, given a background measure μ_0 and a constant $\beta \in \mathbb{R}$, the following Monge-Amère equation

$$\mathrm{MA}(\phi) = e^{\beta\phi}d\mu_0. \tag{1.1}$$

It turns out that for certain data μ_0 and β, this equation admits a unique solution ϕ_*. In those cases we define the measure μ_* to be the Monge-Ampère measure of ϕ_*, in other words

$$\mu_* = \mathrm{MA}(\phi_*).$$

This measure is the subject of the main theorem of the thesis. For large N we expect the particles to distribute according to μ_*.

We will now define the small scale description of our system. It is given by a sequence of so called *Permanental Point Processes.* Originally, these were introduced to model many particle systems of bosons in quantum mechanics. We will not go into this here. For now we will only present the special instance of permanental point processes we will use in this thesis. In the last chapter of this introduction we will explain how they relate to Fekete Points and certain point processes in complex geometry.

For each positive integer k, $\frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n$ defines a set of $N = k^n$ points in X. Let $S^{(N)}$ be the set of functions

$$S^{(N)} = \{\Psi_p^{(N)} : p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n\}$$

where

$$\Psi_p(x) = \sum_{m \in \mathbb{Z}^n + p} e^{-|x-m|^2/2}.$$

Enumerating the points in $\frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n$ we get for each k a matrix valued function on X^N

$$(x_1, \ldots, x_N) \mapsto (\Psi_{p_i}(x_j)).$$

The permanent of a matrix (a_{ij}) is given by the expression

$$\operatorname{perm}(a_{ij}) = \sum_\sigma \prod_i a_{i\sigma(j)}$$

where the sum is taken over all permutations, σ, of N elements. This gives us, for each k, a symmetric function on X^N

$$(x_1, \ldots, x_N) \mapsto \operatorname{perm}(\Psi_{p_i}(x_j)).$$

Now, introducing a background measure μ_0 of mass one and a constant $\beta \in \mathbb{R}$ (which in a thermodynamic interpretation is the inverse temperature of the system) we get a sequence of symmetric probability measures on X^N defined as

$$\mu^{(N)} = \operatorname{perm}(\Psi_{p_i}(x_j))^{\beta/k} \, d\mu_0^{\otimes N}/Z_N.$$

where Z_N is a constant making sure $\mu^{(N)}$ is a probability measure. These probability measures define the point processes that constitute the small scale description of our system. If the number of particles on X is N, then the probability of finding these particles in a certain configuration is determined by $\mu^{(N)}$.

2. Main Result

The main result of the thesis is a statement about what happens as the number of particles becomes very large. Since each probability measure, $\mu^{(N)}$, is defined on a separate space, it is not immediately clear how to speak about the limit of $\mu^{(N)}$ as $N \to \infty$. One way to deal with this is to consider the *mariginals* of $\mu^{(N)}$. Let d be a positive integer. The d'th marginal of $\mu^{(N)}$ is a probability measure on X^d. It is denoted $(\mu^{(N)})_d$ and is defined by

$$(\mu^{(N)})_d(A) = \mu^{(N)}(A \times X^{N-d})$$

for any measurable $A \subset X^d$. The point is that $(\mu^{(N)})_d \in \mathcal{M}_1(X^d)$ for every N. Hence, given d there is a natural question to ask, namely

- Does $(\mu^{(N)})_d \to \mu_*^{\otimes d}$?

The convergence here is in terms of the weak* topology on $\mathcal{M}_1(X)$. If it holds then sampling a large number of particles and restricting our attention to the first d particles is close to sampling d points independently and according to μ_*.

The main result of the thesis states that, under certain conditions on μ_0 and β, the answer to the this question is yes for any d. However, the theorem is not stated in this way. An alternative way to deal with the problem that $\mu^{(N)}$ are defined on different spaces is to map the spaces X^N into $\mathcal{M}(X)$. We think of an element $x = (x_1, \ldots, x_N)$ in X^N as representing the position of N particles. Since we don't care about the order of the particles we might just as well represent them by the measure

$$\delta^{(N)}(x) = \frac{1}{N} \sum_{i}^{N} \delta_{x_i}.$$

We get a map $\delta^{(N)} : X^N \to \mathcal{M}(X)$. Each probability measure $\mu^{(N)}$ defines a random variable $x^{(N)}$ on X^N and we can think of $\delta^{(N)}(x^{(N)})$ as a random measure on X. This is in fact the usual way to represent a point process. It is called the *empirical measure*. The law of $\delta^{(N)}(x^{(N)})$ is given by the push-forward measure

$$\Gamma_{(N)} = \left(\delta^{(N)}\right)_* \mu^{(N)} \in \mathcal{M}_1(\mathcal{M}(X)).$$

Since $\Gamma_{(N)} \in \mathcal{M}_1(\mathcal{M}(X))$ for all N, we can ask the question if $\Gamma_{(N)}$ converges to something as $N \to \infty$. The main theorem is formulated in terms of a convergence of this type.

Theorem 2.1 (Main Theorem). *Let $\mu_0 \in \mathcal{M}_1(X)$ be absolutely continuous and have smooth, strictly positive density with respect to dx. Let $\Gamma^{(N)}$ be defined as above and let $\beta \in \mathbb{R}$. Assume also that* (1.1) *admits a unique solution, ϕ_*. Then*

$$\Gamma_\beta^{(N)} \to \delta_{\mu_*} \tag{2.1}$$

in the weak topology of $\mathcal{M}_1(\mathcal{M}(X))$, where $\mu_* = MA(\phi_*)$.*

The following basic proposition connects this with (2).

Proposition 2.2 (See Proposition 2.2 in [23]).

$$\Gamma_N \to \delta_{\mu_*} \tag{2.2}$$

in the weak topology of $\mathcal{M}_1(\mathcal{M}(X))$ if and only if*

$$(\mu_N)_d \to \mu_*^{\otimes d} \tag{2.3}$$

for all $d \in \mathbb{N}$.

We will not prove this in full here but we will provide an argument for one of the implications.

Proof of (2.2) $\Rightarrow$ (2.3). Assume (2.2) holds. We wish to prove that

$$(\mu_N)_d \to \mu_*^{\otimes d}$$

in the weak* topology on $\mathcal{M}_1(X)$, in other words that for all continuous functions, g, on X^d we have

$$\int_{X^d} gd(\mu_N)_d \to \int_{X^d} gd\mu_*^{\otimes d}. \tag{2.4}$$

Now, for a continuous function g on X^d, consider the continuous function on $\mathcal{M}_1(X)$ given by

$$\mu \mapsto A_g(\mu) = \int_{X^d} gd\mu^{\otimes d}.$$

On the one hand, by (2.2) we have

$$\int_{\mathcal{M}_1(X)} A_g(\mu)d\Gamma(\mu) \to A_g(\mu_*) = \int_{X^N} gd\mu_*^{\otimes d}. \tag{2.5}$$

On the other hand, since $\Gamma = (\delta^{(N)})_* \mu_N$,

$$\begin{aligned}
\int_{\mathcal{M}_1(X)} A_g(\mu) d\Gamma(\mu) &= \int_{X^N} A_g(\frac{1}{N}\sum_i \delta_{x_i}) d\mu_N \\
&= \int_{X^N} \left(\int_{X^N} g d(\frac{1}{N}\sum_i \delta_{x_i})^{\otimes d} \right) d\mu_N \\
&= \int_{X^N} \frac{1}{N^d} \sum_{I\in\mathbb{I}} g(x_{I(1)}, \ldots, x_{I(d)}) d\mu_N \quad (2.6)
\end{aligned}$$

where $\mathbb{I}$ is the set of functions $I : \{1, \ldots, d\} \to \{1, \ldots, N\}$. Let $\mathbb{I}_d$ be the set of injective $I \in \mathbb{I}$. Then, by the symmetry of μ_N

$$\int_{X^N} g(x_{I(1)}, \ldots, x_{I(d)}) d\mu_N = \int_{X^N} g(x_1, \ldots, x_d) d\mu_N$$

for any $I \in \mathbb{I}_d$. As there are $N!/(N-d)!$ elements in $\mathbb{I}_d$ we get that

$$\begin{aligned}
(2.6) &= \int_{X^N} \frac{N!}{N^d(N-d)!} g(x_1, \ldots, x_d) d\mu_N \\
&+ \int_{X^N} \frac{1}{N^d} \sum_{I\in\mathbb{I}\setminus\mathbb{I}_d} g(x_{I(1)}, \ldots, x_{I(d)}) d\mu_N.
\end{aligned}$$

Since $N!/(N-d)! \to 1$ as $N \to \infty$ we get that the first term of this converges to

$$\int_{X^N} g(x_1, \ldots, x_d) d\mu_N = \int_{X^d} g d(\mu_N)_d.$$

Since

$$\int_{X^N} g(x_{I(1)}, \ldots, x_{I(d)}) d\mu_N \leq \sup_{X^d} g$$

for any $I \in \mathbb{I}$ and the number of elements in $\mathbb{I} \setminus \mathbb{I}_d$ is $N^d - N!/(N-d)!$ we may bound the second term by

$$\frac{N^d - N!/(N-d)!}{N^d} \sup_{X^d} g$$

which vanishes as $N \to \infty$. This means (2.4) holds. $\square$

3. Large Deviation Principles

To prove Theorem 2.1 we prove a so called *large deviation principle*. Essentially, a large deviation principle for a sequence of of probability measures consists of bounds on how fast the probability of "unlikely events", or events that deviates a lot from what is expected, tends to zero. These bounds are encoded in a *rate* and a *rate function*. The rate is a sequence of real numbers that tend to ∞ and the rate function is a non-negative real valued function on the sample space. A good but somewhat imprecise way of thinking of a large deviation principle is that, for large N, the probability measures behave roughly as the densities

$$\Gamma_N \sim e^{-r_N G}$$

where r_N is the rate and G is the rate function. This captures the important fact that Γ_N is, for large N, concentrated around the minimizers of G.

The precise definition is as follows:

Definition 3.1. *Let χ be a topological space, $\{\Gamma_N\}$ a sequence of probability measures on χ, G a lower semi continuous function on χ and r_N a sequence of numbers such that $r_N \to \infty$. Then $\{\Gamma_N\}$ satisfies a* large deviation principle *with rate function G and rate r_N if, for all measurable $E \subset \chi$,*

$$-\inf_{E^\circ} G \leq \liminf_{N\to\infty} \frac{1}{r_N} \log \Gamma_N(E) \leq \limsup_{N\to\infty} \frac{1}{r_N} \log \Gamma_N(E) \leq -\inf_{\bar{E}} G \quad (3.1)$$

where E° and $\bar{E}$ are the interior and the closure of E.

Note that putting $E = \chi$ in (3.1) gives

$$-\inf_{\chi} G \leq 0 \leq -\inf_{\chi} G$$

hence $\inf_\chi G = 0$. Assume G admits a unique minimizer, μ_*. In other words, μ_* is the unique point where $G = 0$. Then an "unlikely event" is a subset E of χ such that $\mu_* \notin \overline{E}$. By the lower semi-continuity of G we get $\inf_{\overline{E}} G > 0$. A large deviation principle then provides a bound on the probability of the event E in the sense that for any $\delta < \inf_{\overline{E}} G$ we get

$$\frac{1}{r_N} \log \Gamma_N(E) \leq -\delta,$$

or equivalently

$$\Gamma_N(E) \leq e^{-r_N \delta},$$

for large enough N.

We will now give an example of a large deviation principle.

Example 1. Let $\{x_i\}_{i=1}^{\infty}$ be a sequence of independent, normally distributed random variables with mean 0 and variance 1. Then the average of the first N variables

$$\hat{x}_N = \frac{1}{N}\sum_{i}^{N} x_i$$

is a normally distributed random variable with mean 0 and variance $1/\sqrt{N}$. If we let $\Gamma_N \in \mathcal{M}_1(\mathbb{R})$ denote the law of $\hat{x}_N$, we have

$$\Gamma_N = e^{-Nx^2/2}dx.$$

Then $\{\Gamma_N\}$ satisfy a large deviation principle with rate $r_N = N$ and rate function $G(x) = x^2/2$.

Proof of the statement in example. We will only prove the upper bound in (3.1). Let $E \subset \mathbb{R}$ be a Borel measurable set. Assume δ is a number such that $\delta < \inf_{\overline{E}} x^2/2$ and let $\epsilon > 0$ be a number such that

$$\delta < (1-\epsilon)\inf_{\overline{E}} x^2/2.$$

We get

$$\Gamma_N(E) = \int_E e^{-Nx^2/2}dx \leq e^{-N\delta}\int_E e^{-\epsilon N x^2/2}dx$$

Hence

$$\limsup_{N\to\infty}\frac{1}{N}\log\Gamma_N(E) \leq -\delta + \limsup_{N\to\infty}\frac{\log\int e^{-\epsilon N x^2/2}}{N}dx.$$

The limit in the right hand side of this is clearly 0 and since the argument can be repeated for any $\delta < \inf_{\overline{E}} x^2/2$ we get

$$\limsup_{N\to\infty}\frac{1}{N}\log\Gamma_N(E) \leq -\inf_{\overline{E}} x^2/2. \quad \Box$$

To conclude this section and explain how Theorem 2.1 can be deduced from a suitable Large Deviation Principle we have the following lemma. As we explained above, a large deviation principle guaranteees that for large N the probability measures are concentrated around the minimizers of G. When G admits a unique minimizer we have the following

Lemma 3.2 (See for example the proof of Theorem 1 in the paper). *Let χ be a topological vector space and Γ_N be a sequence of probability measures on χ that satisfies a large deviation principle with rate r_N and rate function G. Assume G admits a unique minimizer μ_*. Then*

$$\Gamma_N \to \delta_{\mu_*}$$

in the weak topology on $\mathcal{M}_1(\chi)$.*

This means proving Theorem 2.1 comes down to proving a large deviation principle with a certain rate function. We want this rate function to admit as its unique minimizer the Monge-Ampère measure of the unique solution to (1.1). In the next section we will give a brief background to Monge-Ampère equations. Then we will explain the variational approach to Monge-Ampère equations and how this relates to a rate function with this property.

4. Monge-Ampère Equations

Monge-Ampère equations arise in several different areas of mathematics. We will be interested in their connections to mass transport on the one hand and their connections to complex geometry on the other. The connection to mass transport is easiest to grasp in its original setting on $\mathbb{R}^n$. As mentioned in Section 1 the Monge-Ampère operator on $\mathbb{R}^n$ is defined on the space of twice differentiable convex functions. It takes the form

$$f \mapsto \det(f_{ij})dx$$

where (f_{ij}) is the Hessian of f. Assume μ is a measure on $\mathbb{R}^n$ and f satisfies the Monge-Ampère equation

$$\det(f_{ij}) = \mu.$$

The gradient of f, ∇f, defines a map from $\mathbb{R}^n$ to $\mathbb{R}^n$. As $\det(f_{ij})$ is the Jacobian determinant of this map we get the following change of variables formula

$$\int_{\nabla f(\mathbb{R}^n)} h d\mu = \int_{\nabla f(\mathbb{R}^n)} h \det(f_{ij})dx = \int_{\mathbb{R}^n} h \circ \nabla f dx$$

for any continuous function, h, on $\mathbb{R}^n$.

This property is generally referred to as a *push forward* property. We say that the measure dx is the push forward of μ under the map ∇f and we write

$$(\nabla f)_* \mu = dx.$$

In fact, solutions to Monge-Ampère equations define maps that are optimal with respect to a certain optimization problem over all maps $T : \mathbb{R}^n \to \mathbb{R}^n$ that satisfies $T_*\mu = dx$. For more on this we refer the reader to Chapter 2 in the paper where we give an introduction to optimal transport.

4.1. The Complex Monge-Ampère operator. Before explaining the complex Monge-Ampère operator we will make an observation about the Monge-Ampère operator in X. Let ϕ be a twice differentiable function on X such that the matrix $(\phi_{ij} + \delta_{ij})$ is strictly positive definite. Then the tensor

$$(\phi_{ij} + \delta_{ij}) dx_i \otimes dx_j$$

defines a Riemannian metric on X. The volume form associated to this metric is given by

$$\sqrt{\det(\phi_{ij} + \delta_{ij})} dx_1 \wedge \ldots \wedge dx_N,$$

in other words, it is closely related to the Monge-Ampère measure of ϕ. There is a similar correspondence in complex geometry. There, given a certain type of reference object, ϕ_0, (a positive metric on a line bundle) a Kähler metric on a complex manifold is represented by a real valued function, f. The complex Monge-Ampère measure of f is given by the volume form associated to the metric. In coordinates $(z_1, \ldots, z_n)$ this is

$$MA^{\mathbb{C}}_{\phi_0}(f) = \det\left(\frac{\partial^2 f}{\partial z_i \partial \bar{z}_j} + \omega_{ij}\right) dz_1 \wedge d\bar{z}_1 \wedge \ldots \wedge dz_n \wedge d\bar{z}_n,$$

where $(\frac{\partial f}{\partial z_i \partial \bar{z}_j})$ is the complex Hessian of f and (ω_{ij}) is the matrix representing the curvature tensor of the reference object ϕ_0.

Representing a metric as a function in this manner has proven exceptionally useful in complex geometry. It allows certain properties of the metric to be stated in terms of the function. Among other things, it transforms the equations defining so called *Kähler-Einstein metrics* from a tensor equation into a scalar equation involving the complex Monge-Ampère operator.

The connection to the real Monge-Ampère operator is manifested when the complex manifold is a certain type of fibration over a real manifold. For example, if we let $Y = \mathbb{C}^n/(\mathbb{Z}^n + i\mathbb{Z}^n)$ we get a fibration over X with projection map $\pi : Y \to X$ given by

$$z \to x$$

where $z = x + iy$. Here we may choose ϕ_0 so that $\omega_{ij} = \delta_{ij}$. If ϕ is a real valued function $\mathbb{C}^n$ such that $\phi(z) = \phi(x+iy)$ is independent of y, then, abusing notation slightly we may think of ϕ as a function on X and the following push forward formula holds

$$\mathrm{MA}(\phi) = \pi_* \, \mathrm{MA}^{\mathbb{C}}_{\phi_0}(\phi).$$

In other words, the real Monge-Ampère measure arises as the push forward of the complex Monge-Ampère measure on a certain "larger" (twice the dimension) complex manifold.

5. The Variational Approach to Differential Equations

It is quite common that differential equations in mathematics and physics have an alternative formulation in terms of a minimization problem. The variational approach to a differential equation is to study it in terms of this minimization problem. For example, if $x : [0,1] \to \mathbb{R}^3$ describes the trajectory of a single particle in a force field with potential V, then, given boundary conditions $x(0) = x_0$ and $\dot{x}(0) = v_0$, x is determined by Newton's laws of motion

$$\ddot{x} = -\nabla V(x).$$

On the other hand, x is also determined by the fact that it minimizes the action

$$S(x) = \int_{[0,1]} \frac{\dot{x}(t)^2}{2} - V(x(t))dt$$

over all possible trajectories $x : [0,1] \to \mathbb{R}^3$ satisfying $x(0) = z_0$ and $\dot{x}(0) = v_0$.

Another example, which is closer to our setting, is that solutions to the homogeneous Laplace equation

$$\Delta\phi = 0$$

in a domain $\Omega \subset \mathbb{R}^n$ with boundary conditions

$$\phi = \Phi$$

on $\delta\Omega$ arise as minimizers of the *Dirichlet Energy*

$$E(\phi) = \frac{1}{2}\int_\Omega \|\nabla\phi\|^2 dx.$$

To prove these things it is useful to introduce a way of differentiating the functionals above.

Definition 5.1. *Let V be a topological vector space, V^* its topological dual and $\langle\cdot,\cdot\rangle$ the pairing of V and V^*. A function $F : V \to \mathbb{R}$ is said to be Gateaux differentiable at a point $\phi \in V$ if there exist $\mu_\phi \in V^*$ such that for every $v \in V$*

$$\frac{\partial}{\partial t} F(\phi + tv) = \langle v, \mu_\phi \rangle .$$

If that is the case then we say that μ_ϕ is the Gateaux derivative of F at ϕ and we write

$$dF|_\phi = \mu_\phi.$$

Now, if a function is Gateaux differentiable in its domain then its Gateaux derivative vanishes at interior extremal points. The key point in proving the two statements above is to verify that, on suitable domains for S and E

$$dS|_x = \ddot{x} - \nabla V(x)$$

and

$$dE|_\phi = -\Delta\phi dx.$$

In the variational approach to (1.1) it is convenient to first normalize the equation. Instead of studying (1.1) we study

$$\mathrm{MA}(\phi) = \frac{e^{\beta\phi}dx}{\int_X e^{\beta\phi}dx}. \tag{5.1}$$

Questions of existence and uniqueness of solutions to (1.1) can be translated into corresponding questions about (5.1). For example, if ϕ is a solution to (1.1) then ϕ also a solution to (5.1). This follows from the striking fact that the left hand side of (1.1) will, for any ϕ, be a probability measure. Conversely, if ϕ is a solution to (5.1) then $\phi + C$ for a suitable constant C will be a solution to (1.1). The functional on $C(X)$ that provide a variational approach to (5.1) is a sum of two terms, each corresponding to one of the two sides of (5.1). The first term, corresponding to the left hand side in (5.1), is given by

$$\xi(\phi) = \int_X \phi^c dx$$

where

$$\phi^c(y) = \sup_{x\in X} -d(x,y)^2/2 - \phi(x)$$

and d is the standard distance function on X

$$d(x,y)^2 = \inf_{m\in\mathbb{Z}^n} |x - y - m|^2.$$

Now, ξ plays the same role for the Monge-Ampère operator as the Dirichlet Energy plays for the Laplace operator. It is Gateaux Differentiable and satisfies (in a weak sense)

$$d\xi|_\phi = -\operatorname{MA}(\phi).$$

The second term, corresponding to the right hand side of (5.1) is given by

$$I_\mu(\phi) = \log \int_X e^\phi d\mu.$$

It is Gateaux differentiable as well and satisfies

$$dI_{\mu_0}|_\phi = \log \int_X e^\phi d\mu_0.$$

Defining

$$F_\beta(\phi) = \xi(\phi) + \frac{1}{\beta} I_{\mu_0}(\beta\phi)$$

we have that F_β is Gateaux differentiable and

$$dF_\beta|_\phi = -\operatorname{MA}(\phi) + \frac{e^{\beta\phi} d\mu_0}{\int_X e^{\beta\phi} d\mu_0}.$$

In particular, any minimizer of F is a weak solution to (5.1). Hence F provides a variational approach to (5.1).

5.1. Legendre Transform and Dual Variation Approach. An important tool in this thesis is *Legendre transform*. As above, let V be a topological vector space, V^* the dual vector space and $\langle \cdot, \cdot \rangle$ the pairing. Given a function, F, on V there is a way of associating a dual function, F^*, defined on V^*. The function F^* is called the Legendre Transform of F and is defined by the following formula:

$$F^*(y) = \sup_{\phi \in V} \langle x, y \rangle - F(x).$$

Now, since F^* is defined as the pointwise supremum of affine functions it will be a convex function.

Using Legendre transform we will be able to get a dual variational approach for (5.1), given by a function on $\mathcal{M}(X)$, the topological dual of $C(X)$. The principle that gives rise to this become rather technical when applied to infinite dimensional, non-reflexive vector spaces so we will present it for finite dimensional vector spaces here.

Assume that V is finite dimensional. Then it turns out that F^* is strictly convex and differentiable if and only if F is strictly convex and differentiable. Assume F is strictly convex and differentiable. Then dF defines an injective map from V to V^*. The key property of Legendre transform we will use here is that $dF^* : V^* \to (V^*)^* = V$ is the inverse of this map. This is used in

Lemma 5.2. *Assume F and G are strictly convex and differentiable functions on a finite dimensional vector space, V, then $x \in V$ is a minimizer of*

$$x \mapsto F(x) + G(x) \tag{5.2}$$

if and only if $y = dF|_x$ is a minimizer of

$$y \mapsto F^*(y) + G^*(-y). \tag{5.3}$$

Proof. Since both F and G are strictly convex and differentiable we have that x is the unique minimizer of (5.2) if and only if $y = dF|_x = -dG|_x$. Since the same is true for F^* and G^* we have that y is the unique minimizer of (5.3) if and only if $dF^*|_y = dG^*|_{-y}$. Now, if $dF|_x = -dG|_x$ then, putting $y = dF|_x$, gives $dF^*|_y = x = dG^*|_{-y}$, hence that y is the unique minimizer of (5.3). Conversely, if $dF^*_y = dG^*_{-y}$, then, if $y = dF|_x$ we have $dG^*|_{-y} = x$ and hence $dG_x = -y = -dF_x$. In other words x is the unique minimizer of (5.2). $\square$

In fact, even though the space of continuous functions on X, $C(X)$, does not satisfy the assumptions in Lemma 5.2, the same principle can be used to, from the variational approach given by F_β, get a dual variational approach defined as a functional on $\mathcal{M}(X)$. This is proved in Lemma 13 in the paper. The functional on $\mathcal{M}(X)$ is referred to as the *Gibbs Free Energy* and it is the the rate function of the large deviation principle used to prove Theorem 2.1.

6. The Gärtner-Ellis theorem

A useful tool when proving large deviation principles is the Gärtner-Ellis theorem. It provides a way to deduce large deviation principles for a sequence of probability measures from the behavior of their moment generating functions. Recall that if Γ is a probability measure on a topological vector space χ then the moment generating function is the function on the topological dual χ^* defined by

$$Z_\Gamma(\phi) = \int_\chi e^{\langle \phi, \mu \rangle} \Gamma.$$

The Gärtner-Ellis Theorem is stated for exponentially tight sequences of probability measures. A sequence,$\{\Gamma_N\}$, is *exponentially tight* if for each $\epsilon \in \mathbb{R}$ there is a compact $K_\epsilon \subseteq \chi$ such that for all N

$$\limsup_{\mathbb{N}\to\infty} \frac{1}{N} \log \Gamma_N(\chi \setminus K_\epsilon) \leq \epsilon. \tag{6.1}$$

We are now ready to state the Gärtner-Ellis Theorem.

THEOREM 6.1 (The Gärtner-Ellis Theorem. See for example Corollary 4.5.27 in [13]). *Let χ be a locally convex topological vector space, $\{\Gamma_N\}$ an exponentially tight sequence of probability measures on χ and r_N a sequence such that $r_N \to \infty$. Let Z_{Γ_N} be the moment generating function of Γ_N and assume*

$$F(\phi) = \lim_{N\to\infty} \frac{1}{r_N} \log Z_{\Gamma_N}(r_N \phi)$$

exist, is finite valued, lower semi continuous and Gateaux differentiable. Then Γ_N satisfies a large deviation principle with rate r_N and rate function given by the Legendre transform of F.

Before moving on we will give a nice application of this theorem. We will prove the classical Sanov's Theorem. Let μ_0 be a probability measure on $\mathbb{R}^n$ with compact support and consider the point process defined by

$$\mu_0^{\otimes N}.$$

This describes a process where N points are chosen independently and according to μ_0. When N is large we expect the empirical measure to approximate μ_0. This is made precise by Sanov's theorem. It connects this point process to the *relative entropy function.*

DEFINITION 6.2. *Assume $\mu, \mu_0 \in \mathcal{M}(X)$ and, if μ is absolutely continuous with respect to μ, let μ/μ_0 denote the density of μ with respect to μ_0. The* relative entropy *of μ with respect to μ_0 is*

$$Ent_{\mu_0}(\mu) = \begin{cases} \int_X \mu \log \frac{\mu}{\mu_0} & \text{if } \mu \text{ is a probability measure and absolutely} \\ & \text{continuous with respect to } \mu_0 \\ +\infty & \text{otherwise,} \end{cases}$$

A key fact, which we will use below, is that the relative entropy function arises as the Legendre transform of the functional I_{μ_0}.

THEOREM 6.3 (Sanov's theorem, see for example 6.2.10 in [13]). *Let $\mu_0 \in \mathcal{M}_1(X)$. Then the family*

$$\left\{ \left(\delta^{(N)}\right)_* \mu_0^{\otimes N} \right\}$$

satisfies a large deviation principle with rate $r_N = N$ and rate function Ent_{μ_0}.

PROOF. Let $\Gamma_N = \left\{ \left(\delta^{(N)}\right)_* \mu_0^{\otimes N} \right\}$. We want to apply the the Gärtner-Ellis Theorem. Note that if F is a function on $\mathcal{M}_1(X)$, then,

$$\int_{\mathcal{M}_1(X)} F(\mu)\Gamma_N = \int_{X^N} F\left(\delta^{(N)}(x)\right) d\mu_0^{\otimes N}.$$

Moreover,

$$\left\langle N\phi, \delta^{(N)}(x) \right\rangle = N \int_X \phi \frac{1}{N} \sum \delta_{x_i} = \sum \phi(x_i).$$

This means the moment generating function of Γ_N are

$$\begin{aligned} Z_{\Gamma_N}(N\phi) &= \int_{\mathcal{M}_1(X)} e^{\langle N\phi, \mu \rangle} \Gamma_N = \int_{X^N} e^{\sum \phi(x_i)} d\mu_0^{\otimes N}. \\ &= \left(\int_X e^{\phi} d\mu_0 \right)^N \end{aligned}$$

and

$$\frac{1}{N} \log Z_{\Gamma_N}(N\phi) = I(\phi).$$

Exponential tightness is a consequence of the compact support of μ_0 and, as mentioned in the previous section, I_{μ_0} is Gateaux differentiable. Hence we may apply the the Gärtner-Ellis Theorem. Since Ent_{μ_0} is the Legendre transform of I_{μ_0} this finishes the proof. ☐

7. Probabilistic Background

7.1. Fekete Points. Consider a compact subset, K, of the complex plane. Let $z_1, \ldots, z_n \in K$ denote the positions of N electrons on the surface represented by K. Two electrons repel each other with forces inversely proportional to the distance between them. This means that, in a suitable choice of units, the potential energy of the systems of electrons is given by

$$-\sum_{1 \le i < j \le N} \log |z_i - z_j|. \tag{7.1}$$

A set of points $\{z_1, \ldots, z_N\}$ representing a configuration of electrons that minimizes the potential energy is called a set of Fekete points. There is a classical result which describes the distribution of Fekete points as $N \to \infty$. In potential theory there is a certain measure associated to K called the *equilibrium measure of* K which is defined in terms of the Laplace operator and which we will denote by μ_K.

THEOREM 7.1. *Let $\{P^{(N)}\}$ be a sequence of finite sets such that for each $N \in \mathbb{N}$, $P^{(N)}$ is a set of Fekete points and $|P^{(N)}| = N$. Then*

$$\delta(P^{(N)}) \to \mu_K$$

in the weak topology on the space of measures on $\mathbb{C}$.*

There are at least two natural ways to formulate higher dimensional generalization of Fekete points. One way, which is perhaps the one that would first come to mind, is to, given a compact subset K of $\mathbb{C}^n$, consider sets $\{z_1, \ldots, z_N\}$ that minimize the functional

$$-\sum_{1 \leq i,j \leq N} \log\|z_i - z_j\|, \tag{7.2}$$

where $\|\cdot\|$ is the standard norm on $\mathbb{C}^n$. This would take us to linear higher dimensional potential theory and would preserve the relation to the Laplace operator. However, we will rewrite the expression in (7.1) in a way that allows for another, more algebraic, higher dimensional generalization. This generalization is related to optimal sampling of higher dimensional polynomials (see [2]). It is part of pluripotential theory rather than linear higher dimensional potential theory and related to the Monge-Ampère operator instead of the Laplace operator.

Given N points $\{z_0, \ldots, z_{N-1}\} \subset \mathbb{C}$ we may form the associated *Vandermonde matrix*

$$(z_i^j) = \begin{pmatrix} 1 & z_0 & z_0^2 & \ldots & z_0^{(n-1)} \\ \vdots & & & \ddots & \vdots \\ 1 & z_{N-1} & z_{N-1}^2 & \ldots & z_{N-1}^{(n-1)} \end{pmatrix}.$$

and the associated Vandermonde determinant, given by $\det(z_i^j)$. We have

LEMMA 7.2. *Let $z_0, \ldots, z_{N-1} \in \mathbb{C}$. Then*

$$\prod_{0 \leq i < j \leq N-1} |z_i - z_j| = |\det(z_i^j)|^2. \tag{7.3}$$

Proof. If we fix the variables $z_0, \ldots, z_{N-2}$ then we may consider the right hand side of (7.3) as a polynomial in z_{N-1} of degree $N-1$. Moreover, it is zero whenever $z_{N-1} - z_j = 0$ for some $0 \leq j \leq N-2$. If we assume that $z_i \neq z_j$ for all $i, j \in \{0, \ldots, N-2\}$ then this implies

$$\det(z_i^j) = C \prod_{0 \leq j \leq N-2} (z_{N-1} - z_j). \tag{7.4}$$

where C is the leading coefficient of the polynomial. By continuity (7.4) holds for all (not necessearliy distinct) $z_0, \ldots, z_{N-2} \in \mathbb{C}$. Further, by expanding the determinant we get that

$$\det(z_i^j) = \det(z_i^j)_{0 \leq i,j \leq N-2} z_{N-1}^{N-1} + c_{N-2} z^{N-2} + \ldots + c_0,$$

for some $c_0, \ldots, c_{N-2}$. This means $C = \det(z_i^j)_{0 \leq i,j \leq N-2}$. Proceeding by induction gives the statement of the lemma. $\square$

This means we can think of Fekete points as maximizers of the function

$$\{z_0, \ldots, z_{N-1}\} \mapsto |\det(z_i^j)|^2. \tag{7.5}$$

Note that $\{1, z, \ldots, z^{N-1}\}$ is a basis of the space of polynomials in one variable of degree at most $N-1$. Moreover, replacing this basis by another basis for the same vector space only changes (7.5) by a constant. In particular this doesn't affect its minimizers. This means Fekete points is an invariant of finite dimensional vector spaces of functions. Now, complex algebraic geometry abound with interesting finite dimensional vector spaces of functions. Typically, they are given by the set of meromorphic functions on a manifold that share a certain polar set. These spaces are called the *total linear systems* associated to a *divisor*. The space of polynomials in one variable of degree at most $N-1$ can be identified with a space of this type, namely the space of meromorphic functions on the Riemann sphere $\mathbb{C} \cup \{\infty\}$ that have a pole of order at most $N-1$ at ∞. This indicates that there might be a natural generalisation of Fekete points to this setting. The way to state this generalization is in terms of sections of line bundles over a complex manifold Y. We have the following

Definition 7.3. *Let Y be a compact complex manifold, L a line bundle over Y and ϕ_0 a positive metric on L. Moreover, let $k \in \mathbb{N}$, K be a closed subset of Y and $\{s_1, \ldots, s_N\}$ be a basis of the space of holomorphic sections in kL. Then a set of points $\{z_1, \ldots, z_N\} \subset K$ that maximizes* *the quantity*

$$\|\det(s_i(z_j))\|^2_{kN\phi_0} \tag{7.6}$$

over all sets of N points in K is a set of Fekete points of K.

This generalization is due to Leja on the one hand, who generalized Fekete points to $\mathbb{C}^n$, and Berman, Boucksom and Witt Nyström [2] on the other hand, who took the step to the geometric setting. The last group of authors presented this definition together with a description of the asymptotic behavior of Fekete points in the same manner as Theorem 7.1. However, in this non-linear higher dimensional case the equilibrium measure is defined in terms of the Monge-Ampère operator instead of the Laplace operator.

7.2. Determinantal Point Processes in Complex Geometry. There is no randomness involved in the definition of Fekete points. In Theorem 7.1 the sets of points, $P^{(N)}$, are deterministically chosen. We will now introduce some randomness into the procedure. Doing this we will arrive at a special case of so called *β-deformed determinantal point processes.* Similarly as the point processes we defined in section 1, these will be defined in terms of symmetric probability measures on X^N. They were introduced by Berman. In [2] he established their connection to complex Monge-Ampère equations.

As before, let Y be a compact complex manifold, L a line bundle over Y and ϕ_0 a positively curved metric on L. Instead of, like in the case of Fekete points, fixing a compact set K, fix a probability measure μ_0 on Y. We also fix a constant $\beta \geq 0$. We define

$$\mu^{(N)} = \|\det(s_i(x_j))\|^{2\beta/k}_{kN\phi_0} d\mu_0^{\otimes N}/Z$$

where Z is a constant ensuring the total mass of the measure is one. As above, $\{s_1, \ldots, s_N\}$ is a basis of the space of holomorphic sections in kL. This means that, instead of choosing maximizers of (7.6) we pick points according to a probability measure whose density is proportional to (the β/k'th power of) (7.6).

Now, the connection between determinantal point processes and complex Monge-Ampère equations discovered by Berman can be thought of as the complex analog of Theorem 2.1. We will state it in a slightly less general setting than it was presented by the author

Theorem 7.4. *[2] Let Y be a compact complex manifold, L a line bundle over Y and ϕ_0 a positively curved metric on L. Fix a constant*

$\beta > 0$ *and a volume form* μ_0*. Then the laws of the associated point processes*

$$\Gamma_N = (\delta^{(N)})_* \mu^{(N)}$$

satisfies

$$\Gamma_N \to \delta_{\mu_*}$$

where $\mu_* = \mathrm{MA}^{\mathbb{C}}_{\phi_0}(\phi_*)$ *and* ϕ_* *is the unique solution to the equation*

$$\mathrm{MA}^{\mathbb{C}}_{\phi_0}(\phi) = e^{\beta\phi}\mu_0.$$

As explained in Section 4.1 the real Monge-Ampère operator arise as the push forward of the complex Monge-Ampère operator on a certain complex manifold. In fact there is a similar relation between permanents and determinants. This relation is explained in the last section of the paper.

References

[1] R. Berman *Kähler-Einstein metrics, Canonical Random Point Processes and Birational Geometry* arXiv:1307.3634v1 [Math.DG] 2013

[2] R. Berman. S Boucksom. D Witt Nyström. *Fekete Points and Convergence Towards Equilibrium Measures on Complex Manifolds* Acta Math. Vol 207, Issue 1 (2011), 1-27.

[3] R. Feynman *The Feynman Lectures on Physics* Addison Wesley Longman. 1970.

[4] A-S Sznitman *Topics in Propagation of Chaos* École d'Été de Probabilités de Saint-Flour XIX-1989. 165-251. Lecture notes in Math. 1464. Springer. 1991.

Paper

Permanental Point Processes on Real Tori, Theta Functions and Optimal Transport

Jakob Hultgren

Permanental Point Processes on Real Tori, Theta Functions and Optimal Transport

Jakob Hultgren

Abstract. Inspired by objects in complex geometry we introduce a family of permanental point processes on real tori and show that the empirical measures converge in law towards the real Monge-Ampère measures of solutions to certain real Monge-Ampère equations.

1. Introduction

We introduce a thermodynamic approach for producing solutions to certain real Monge-Ampère equations on the real torus $X = \mathbb{R}^n/\mathbb{Z}^n$. Motivated by the problem of singular Kähler-Einstein metrics of (almost everywhere) positive curvature on the Abelian variety $\mathbb{C}^n/(4\pi\mathbb{Z}^n + i\mathbb{Z}^n)$ we propose a corresponding real Monge-Ampère equation on X (see equation (1.3)). Inspired by Berman's construction in [1] we set up a statistic mechanical framework that applies to a wide range of real Monge-Ampère equations on X (see equation (1.1)). Our first result is that, as long as (1.1) admits a unique solution, the point process defined by the statistic mechanical framework converges to a measure on X related to equation (1.1). More precisely, and in the language of thermodynamics, under absence of first order phase transitions the microscopic setting admits a macroscopic limit determined by equation (1.1). The assumption of no first order phase transition, in other words that (1.1) admits a unique solution, always holds for positive temperature. However, a reflection of the fact that the related complex geometric problem is one of *positive* curvature is that the statistical mechanic setting for (1.3) is one of *negative* temperature. As a second result we prove the absence of first order phase transitions down to the critical temperature of -1.

1.1. Setup.

Let dx be the standard volume measure on X induced from $\mathbb{R}^n$. Let β be a real constant and μ_0 a probability measure on X, absolutely continuous and with smooth, strictly positive density with respect to dx. Given the data (μ_0, β) we will consider the real Monge-Ampère equation on X given by

$$\mathrm{MA}(\phi) = e^{\beta\phi} d\mu_0. \tag{1.1}$$

Here MA is the Monge-Ampère operator defined by

$$\phi \mapsto \det(\phi_{ij} + \delta_{ij})dx. \tag{1.2}$$

where (ϕ_{ij}) is the Hessian of ϕ with respect to the coordinates on X induced from $\mathbb{R}^n$ and δ_{ij} is the Kronecker delta. As usual we demand of a solution $\phi : X \to \mathbb{R}$ that it is twice differentiable and quasi-convex in the sense that $(\phi_{ij} + \delta_{ij})$ is a positive definite matrix.

We will pay specific attention to the case when μ_0 is choosen as the measure

$$\gamma = \sum_{m \in \mathbb{Z}^n} e^{-|x-m|^2/2} dx.$$

We get the equation

$$\mathrm{MA}(\phi) = e^{\beta\phi}\gamma. \tag{1.3}$$

As mentioned above this equation has an interpretation in terms of complex geometry. For $\beta = -1$, (1.3) arises as the "push forward" of a twisted Kähler-Einstein equation on the Abelian variety $\mathbb{C}^n/4\pi\mathbb{Z}^n + i\mathbb{Z}^n$. A more detailed exposition of this relation will follow in Section 6.1.

1.2. Construction of the Point Processes. The point processes we will study arise as the so called "β-deformations" of certain *permanental point processes* (see [17] for a survey). Let's first recall the general setup of a permanental point process with N particles. We begin by fixing a set of N *wave functions* on X

$$S^{(N)} = \{\Psi_1^{(N)}, \dots, \Psi_N^{(N)}\}.$$

This defines a matrix valued function on X^N

$$(x_1, \dots, x_N) \to (\Psi_i(x_j)).$$

Recall that the permanent of a matrix $A = a_{ij}$ is the quantity

$$\sum_\sigma \prod_i a_{i,\sigma(i)}$$

where the sum is taken over all permutations of the set $\{1 \dots, N\}$. Together with the background measure μ_0 this defines a symmetric probability measure on X^N

$$\mathrm{perm}(\Psi_i^{(N)}(x_j)) d\mu_0^{\otimes N}/Z_N, \tag{1.4}$$

where Z_N is a constant ensuring the total mass is one. This is a *pure* permanental point process. We will define, for each $k \in \mathbb{N}$, a set of

$N = N_k$ wave functions and, for a given $\beta \in \mathbb{R}$, study the so called β-deformations of (1.4)

$$\mu_\beta^{(N)} = (\mathrm{perm}(\Psi_i(x_j))^{\beta/k} \, d\mu_0^{\otimes N}/Z_{\beta,N} \tag{1.5}$$

where, as above, $Z_{\beta,N}$ is a constant ensuring the total mass is one. We will now define the sets of wave functions. Note that $\mu_\beta^{(N)}$ does not depend on the order of the element in $S^{(N)}$. For each positive integer k, let

$$S^{(N)} = \{\Psi_p^{(N)} : p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n\}$$

where

$$\Psi_p^{(N)}(x) = \sum_{m \in \mathbb{Z}^n + p} e^{-k|x-m|^2/2} dx.$$

Before we move on we should make a comment on the notation. We get $N = N_k = k^n$. Throughout the text, in formulas where both N and k occur, the relation $N = k^n$ will always be assumed.

Finally, we will make two remarks on the definitions. In [21] permanental point processes are used to model a bosonian many particle system in quantum mechanics. In that interpretation $\Psi_i^{(N)}$ defines a 1-particle wave function and the permanent above is the corresponding N-particle wave function defined by $\Psi_1^{(N)}, \ldots, \Psi_N^{(N)}$. Secondly, we will explain in Section 6.2 how the wave functions arises as the "push forward" of θ-functions on $\mathbb{C}^n/(4\pi\mathbb{Z}^n + i\mathbb{Z}^n)$.

1.3. Main Results. Denote the space of probability measures on X by $\mathcal{M}_1(X)$ and consider the map $\delta^{(N)} : X^N \to \mathcal{M}_1(X)$

$$\delta^{(N)}(x) = \delta^{(N)}(x_1, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}.$$

Let $x = (x_1, \ldots, x_N) \in X^N$ be the random variable with law $\mu_\beta^{(N)}$. Its image under $\delta^{(N)}$, $\delta^{(N)}(x)$, is the *empirical measure*. This is a random measure with law given by the push-forward measure

$$\Gamma_\beta^{(N)} = \left(\delta^{(N)}\right)_* \mu_\beta^{(N)} \in \mathcal{M}_1(\mathcal{M}_1(X)) \tag{1.6}$$

Our results concern the weak* limit of $\Gamma_\beta^{(N)}$ as $N \to \infty$. In particular we will show, in some cases, that the limit is a dirac measure concentrated at a certain $\mu_* \in \mathcal{M}_1(X)$ related to (1.1) or (1.3). Loosely speaking, this means μ_* can be approximated by sampling larger and larger point sets on X according to $\mu_\beta^{(N)}$.

Theorem 1.1. *Let $\mu_0 \in \mathcal{M}_1(X)$ be absolutely continuous and have smooth, strictly positive density with respect to dx. Let $\Gamma^{(N)}$ be defined as above and let $\beta \in \mathbb{R}$. Assume also that* (1.1) *admits a unique solution, ϕ_*. Then*

$$\Gamma_\beta^{(N)} \to \delta_{\mu_*} \tag{1.7}$$

in the weak topology of $\mathcal{M}_1(\mathcal{M}_1(X))$, where $\mu_* = MA(\phi_*)$.*

Remark 1.2. The assumption that (1.1) admits a unique solutions is always satisfied when $\beta > 0$. This follows from standard arguments (see Theorem 5.6). However, the case $\beta < 0$ is a lot more subtle. In our second result we show that, in the special case $\mu_0 = \gamma$, the assumption holds for certain negative values of β as well.

Theorem 1.3. *Assume $\mu_0 = \gamma$ and $\beta \in [-1, 0)$. Then equation* (1.3) *admits a unique solution.*

Note that if $\beta \neq 0$ and $\mu_* = MA(\phi_*)dx$ where ϕ_* is a solution to (1.1), then ϕ_* can be recovered from μ_* as $\phi_* = \frac{1}{\beta}\log\rho$ where ρ is the density of μ_* with respect μ_0. In fact we get the following corollary of Theorem 1.1.

Corollary 1.4. *Let $\mu_0 \in \mathcal{M}_1(X)$ be absolutely continuous and have smooth, strictly positive density with respect to dx. Let $\beta \neq 0$. Assume also that* (1.1) *admits a unique solution, ϕ_*. Let $\phi_N : X \to \mathbb{R}$ be the function defined by*

$$\phi_N(x_1) = \frac{1}{\beta}\log\int_{X^{N-1}}\left(\mathrm{perm}(\Psi_{p_i}^{(N)}(x_j)\right)^{\beta/k} d\mu^{\otimes(N-1)}(x_2,\ldots,x_n)/Z_{\beta,N}.$$

Then ϕ_N converges uniformly to ϕ^*.

If we put $\beta = 0$ in (1.1) we get the inhomogenous Monge-Ampère equation. Solutions then determine Optimal Transport maps on X. Now, although Corollary 1.4 doesn't cover the case $\beta = 0$, by considering $\mu_{\beta_N}^{(N)}$ for the sequence of constants $\beta_N = 1/N$ we will be able to produce explicit approximations of optimal transport maps. However, when working with optimal transport it is natural to consider a more general setting than the one proposed for equation (1.1). Because of this we will not state this corollary here but postpone it to Section 6.3.

1.4. Outline.

1.4.1. *Convergence in Theorem 1.1 and a Large Deviation Principle.* Theorem 1.1 will follow from a *large deviation principle* for the sequence $\Gamma^{(N)}$ (see Theorem 3.2). This large deviation principle provides a quantitative description of the convergence in Theorem 1.1, recording the speed of convergence in a *rate function* $G : \mathcal{M}_1(X) \to [0, \infty)$, satisfying $\inf G = 0$ and a *rate* $\{r_N\} \subset \mathbb{R}$ such that $r_N \to \infty$ as $N \to \infty$. We will give a formal definition of large deviation principles in Section 3. Roughly speaking, a large deviation principle with rate function G and rate r_N holds if, for $U \subset \mathcal{M}_1(X)$, the probability $\Gamma(U)$ behaves as

$$e^{-r_N \inf_U G}$$

as $N \to \infty$. This means $\Gamma^{(N)}$, for large N, is concentrated where G is small. In particular, if G admits a unique minimizer, μ_*, (where $G = 0$) then it follows that $\Gamma^{(N)}$ converges in the weak* topology to δ_{μ_*}.

1.4.2. *Proof of the Large Deviation Principle.* It turns out that the rate function above is related to the Wasserstein metric of optimal transport. In Section 2 we will recall some basic facts about optimal transport. In particular, we explain how Kantorovich' duality principle gives an explicit formula for the Legendre transform of the squared Wasserstein distance from a fixed measure. The proof of Theorem 3.2 is given in Section 3 and it is divided into two parts of which the first part uses this explicit formula. In the first part, given in Section 3.1, we take a sequence of constants β_N such that $\beta_N \to \infty$ and study the family $\{\Gamma^{(N)}_{\beta_N}\}$. In the thermodynamic interpretation this means we are studying the zero temperature limit of the system. Using the formula given by Kantorovich duality and the Gärtner-Ellis theorem, relating the moment generating functions of $\Gamma^{(N)}_{\beta_N}$ to the Legendre transform of a rate function, we prove a large deviation principle for this family (see Theorem 3.6). In the second part of the proof, given in Section 3.2, we show how the large deviation principle in Theorem 3.2 can be deduced from this. This second part is originally explained in [1] and it turns out that the crucial point is the equicontinuity and uniform boundedness of the (normalized) energy functions

$$-\frac{1}{kN} \log \operatorname{perm}(\Psi_i^{(N)}(x_j)).$$

These properties will follow from equicontinuity properties and bounds on the wave functions $\Psi_i^{(N)}$ and we give a proof of these properties in Section 3.3.

1.4.3. *Connection to the Monge-Ampère Equation.* The final ingredients in the proof of Theorem 1.1 are given in Section 4.1 and Section 4.2 (essentially by Lemma 4.1 and Theorem 4.3). These sections connect the large deviation principle above with the Monge-Ampère equation (1.1). Note that, as $\inf G = 0$, G admits a unique point where $G = 0$ if and only if G admits a unique minimizer. We apply a variational approach to (1.1). Uniqueness and existence of solutions is studied by means of a certain energy functional on $C(X)$ whose stationary points corresponds to weak solutions of (1.1). The rate function above, G, is closely related to this energy functional. This relation encodes the fact that minimizers of G arise as the Monge-Ampère measures of solutions to (1.1). Moreover, it follows from this relation that G admits a unique minimizer if the energy functional does, which is true if and only if (1.1) admits a unique solution.

1.4.4. *Theorem 1.3.* Existence of weak solutions will follow from the variational approach and compactness properties of the space of quasi convex functions on X (see Section 5.1) and regularity will follow from results by Cafarelli explained in Lemma 5.5. These type of existence properties for solutions to Monge-Ampère equations on affine manifolds was studied by Caffarelli and Viaclovsky [9] on the one hand and Cheng and Yau [10] on the other. It is interesting to note however, that while their result would primarily apply to the case $\beta \geq 0$, the variational approach which we apply gives a simple proof for existence for any $\beta \in \mathbb{R}$. Uniqueness is proved in Section 5.3. Here we look at the space of quasi-convex functions equipped with an affine structure different from the standard one. It will then follow from the Prekopa inequality that the energy functional associated to (1.3) is strictly convex with respect to this affine structure, hence admits no more than one minimizer. This is an extension of an argument used in [3] to prove uniqueness of Kähler-Einstein metrics on toric Fano manifolds. Curiously, there doesn't seem to be any direct argument for this using the Prekopa theorem on Riemannian manifolds (see [12]). Instead, we need to lift the problem to the covering space $\mathbb{R}^n$ and use that γ is the push forward of a measure on $\mathbb{R}^n$ with strong log-concavity properties.

1.4.5. *Geometric Motivation.* In Section 6 we explain the connections to the point processes on compact Kähler manifolds introduced by Berman in [2]. More precisely, we explain the connection with a complex Monge-Ampère equations on $\mathbb{C}^n/4\pi\mathbb{Z}^n + i\mathbb{Z}^n$ and how the wave functions and permanental point processes defined here are connected to theta-functions and determinantal point processes on $\mathbb{C}^n/4\pi\mathbb{Z}^n + i\mathbb{Z}^n$. Finally, in Section 6.3 we show how the connection to optimal transport can be exploited to get explicit approximations of optimal transport maps on X.

2. Preliminaries: Optimal Transport on Real Tori

In this section we will recall some basic theory of optimal transport. The content of the chapter is well known. Early contributors to the theory are Cordera-Erasquin [11] who established a a theory of optimal transport on real tori and McCann [20] who took it to the very general setting of Riemannian manifolds. The reason for this is the close relation between optimal transport and real Monge-Ampère equations. The most important part is Corollary 2.7. There we explain how Kantorovich' duality theorem give a variational approach to real Monge-Ampère equations and an explicit formula for the Legendre transform of the functional $\mu \to W^2(\mu, dx)$, where $W^2(\cdot,\cdot)$ is the Wasserstein metric, a distance function on $\mathcal{M}_1(X)$ defined in terms of optimal transport and which turn up in the rate function describing the behaviour of the point process $\Gamma^{(N)}$ as $N \to \infty$.

2.1. Kantorovich' Problem of Optimal Transport. We will use Kantorovich' formulation (as opposed to Monge's formulation) of the optimal transport problem. The given data is a smooth manifold Y, a *cost function* $c : Y \times Y \to [0, \infty)$, a *source measure*, $\mu \in \mathcal{M}_1(Y)$ and a *target measure*, $\nu \in \mathcal{M}_1(Y)$. Kantorovich problem of optimal transport is the problem of minimizing the functional

$$C(\gamma) = \int_{Y\times Y} c(x,y)d\gamma(x,y)$$

over the set of *transport plans*, $\Pi(\mu, dx)$, consisting of measures $\gamma \in \mathcal{M}_1(Y \times Y)$ such that the first and second marginals of γ equal μ and ν respectively. The optimal transport distance between μ and ν is the quantity

$$\inf_{\gamma\in\Pi(\mu,dx)} C(\gamma). \tag{2.1}$$

In our case $Y = X$, $\nu = dx$ and $c = d(\cdot,\cdot)^2/2$ where d is the distance function on X induced from $\mathbb{R}^n$. In other words, if $x, y \in \mathbb{R}^n$ and $\pi : \mathbb{R}^n \to X$ is the quotient map, then

$$c(\pi x, \pi y) = \frac{d(\pi x, \pi y)^2}{2} = \frac{\inf_{m \in Z^n} |x - y - m|^2}{2}.$$

With this choice of cost function, (2.1) is often referred to as the (squared) Wasserstein distance, $W^2(\mu, dx)$, between μ and dx.

2.2. The c-Transform and c-Convex Functions. A cost function in optimal transport defines a c-transform, closely related to Legendre transform on $\mathbb{R}^n$. Let $C(X)$ be the space of continuous functions on X. For $\phi \in C(X)$ the c-transform of ϕ is

$$\phi^c(y) = \sup_{x \in X} -c(x, y) - \phi(x) = \sup_{x \in X} -\frac{d(x, y)^2}{2} - \phi(x) \tag{2.2}$$

Note that if ϕ is a smooth function on X such that $(\phi_{ij} + \delta_{ij})$ is positive definite, then there is a natural way of associating to ϕ a convex function on $\mathbb{R}^n$, namely

$$\Phi(x) = \phi(\pi x) + \frac{x^2}{2}. \tag{2.3}$$

Let $C(\mathbb{R}^n)$ be the space of continuous functions on $\mathbb{R}^n$ and if $\Phi \in C(X)$ let Φ^* denote the Legendre transform of Φ. The map from $C(X)$ to $C(\mathbb{R}^n)$ given by $\phi \mapsto \Phi$, relates c-transform on X to Legendre transform on $\mathbb{R}^n$ in the sense that

Lemma 2.1. *Let $\phi \in C(X)$ and*

$$\Phi(x) = \phi(\pi x) + \frac{x^2}{2}.$$

Then

$$\Phi^*(y) = \phi^c(\pi y) + \frac{y^2}{2}.$$

PROOF. Note that

$$\begin{aligned}
\sup_{x\in\mathbb{R}^n} -\frac{|x-y|^2}{2} - \phi(\pi x) &= \sup_{x\in[0,1]^n, m\in\mathbb{Z}^n} -\frac{|x-y-m|^2}{2} - \phi(\pi x) \\
&= \sup_{x\in[0,1]^n} - \inf_{m\in\mathbb{Z}^n} \frac{|x-y-m|^2}{2} - \phi(\pi x) \\
&= \sup_{x\in X} -\frac{d(x,\pi y)^2}{2} - \phi(x) \\
&= \phi^c(\pi y).
\end{aligned}$$

This means

$$\begin{aligned}
\Phi^*(y) &= \sup_{x\in\mathbb{R}^n} \langle x, y\rangle - \Phi(x) \\
&= \sup_{x\in\mathbb{R}^n} -\frac{|x-y|^2}{2} - \phi(\pi x) + \frac{y^2}{2} \\
&= \phi^c(\pi y) + \frac{y^2}{2}.
\end{aligned}$$

which proves the lemma. □

It follows that $\phi \in C(X)$ satisfies $(\phi^c)^c = \phi$ if and only if Φ is convex. The property $(\phi^c)^c = \phi$ is often referred to as c-convexity and we will denote the set of functions in $C(X)$ that satisfy this $P(X)$. Since Φ^* is convex for any $\Phi \in C(\mathbb{R}^n)$ we get that $\phi^c \in P(X)$, for any $\phi \in C(X)$. Moreover, also from the theory of convex functions on $\mathbb{R}^n$, we get that the projection $\phi \mapsto (\phi^c)^c$ of $C(X)$ onto $P(X)$ is monotone in the sense that $(\phi^c)^c(x) \leq \phi(x)$ for all $x \in X$.

Let $P(\mathbb{R}^n)$ be the set of convex functions on $\mathbb{R}^n$. It is easy to verify that the image of $P(X)$ in $P(\mathbb{R}^n)$ under the map $\phi \mapsto \Phi$ (where Φ is given by (2.3)) is given by the set

$$\begin{aligned}
P_{\mathbb{Z}^n}(\mathbb{R}^n) &= \{\Phi \in P(\mathbb{R}^n) : \Phi(x+m) - \frac{|x+m|^2}{2} = \Phi(x) - \frac{x^2}{2} \forall m \in \mathbb{Z}^n\} \\
&= \{\Phi \in P(\mathbb{R}^n) : \Phi(x+m) = \Phi(x) + \langle x, m\rangle + \frac{m^2}{2} \forall m \in \mathbb{Z}^n\}
\end{aligned} \tag{2.4}$$

Now, let $\phi \in P(X)$ and Φ be the image of ϕ in $P_{\mathbb{Z}^n}(\mathbb{R}^n)$. Then Φ is differentiable at a point $x \in \mathbb{R}^n$ if and only if ϕ is differentiable at πx. Since a convex function on $\mathbb{R}^n$ is differentiable almost everywhere we get that any $\phi \in P(X)$ is differentiable almost everywhere (with respect to dx). Further, it follows from (2.4) that Φ is differentiable at x and $\nabla\Phi(x) = y$ if and only if Φ is differentiable at $x + m$ and $\nabla\Phi(x + m) = y + m$. This means the map $\nabla\Phi : \mathbb{R}^n \to \mathbb{R}^n$, where it is defined, factors through to a map $X \to X$. This map is the so called c-gradient map in optimal transport, denoted $\nabla^c\phi$. It satisfies the formula

$$\nabla^c\phi(\pi x) = \pi\nabla\Phi(x).$$

Further, Φ is differentiable at x and $\Phi(x) = y$ if and only if y is the unique point in $\mathbb{R}^n$ such that

$$\Phi(x) + \Phi^*(y) = \langle x, y\rangle . \tag{2.5}$$

This holds if and only if

$$\phi(\pi x) + \phi^c(\pi y) = -\frac{d(\pi x, \pi y)^2}{2}. \tag{2.6}$$

We conclude that ϕ is differentiable and $\nabla^c\phi(\pi x) = \pi y$ if and only if πy is the unique point in X such that (2.6) holds. In fact, this is the usual definition of the c-gradient and one of its strengths is that it becomes immediately apparent that if ϕ is differentiable at x and ϕ^c is differentiable at $y = \nabla^c\phi(x)$, then $\nabla^c\phi^c(y) = x$.

The definition of the Monge-Ampère operator in (1.2) makes sense for twice differentiable functions. We will now provide an extension of this operator to $P(X)$.

Definition 2.2. *Let $\phi \in P(X)$. We define the Monge-Ampère measure,* $\mathrm{MA}(\phi)$*, of ϕ as*

$$\mathrm{MA}(\phi) = (\nabla^c\phi^c)_*dx.$$

Consequently, we refer to functions in $P(X)$ satisfying

$$(\nabla^c\phi^c)_*dx = \mu$$

as weak solutions to

$$\mathrm{MA}(\phi) = \mu. \tag{2.7}$$

Now, the following lemma will serve as a direct justification of Definition 2.2 and we will see in Theorem 2.4 that it fits nicely into the theory of optimal transport. Moreover, weak solutions to (1.1) in terms of Definition 2.2 is the natural analog of so called Alexandrov solutions to Monge-Ampère equations on $\mathbb{R}^n$ (see Section 5.2). In fact, we will see in Lemma 5.4 that the map $\phi \mapsto \Phi$ where Φ is given by (2.3) gives a direct link between these two types of solutions.

LEMMA 2.3. *Assume ϕ is smooth and $(\phi_{ij}+\delta_{ij})$ is strictly positive definite. Then*

$$\det(\phi_{ij}+\delta_{ij})dx = (\nabla^c\phi^c)_* dx.$$

PROOF. First of all, we claim that $\nabla^c\phi^c : X \to X$ is one-to-one. To see this, assume that $\nabla^c\phi^c(x_1) = \nabla^c\phi^c(x_2)$ for $x_1, x_2 \in X$. Let $\tilde{x}_1, \tilde{x}_2 \in \mathbb{R}^n$ be lifts of x_1 and x_2 respectively and Φ^* be the image of ϕ^* in $P_{\mathbb{Z}^n}(\mathbb{R}^n)$. We get

$$\nabla\Phi^*(\tilde{x}_1) = \nabla\Phi^*(\tilde{x}_2) + m.$$

By (2.4) we get $\nabla\Phi^*(\tilde{x}_1) = \nabla\Phi^*(\tilde{x}_2 + m)$. But since ϕ, and hence Φ, is smooth Φ^* must be strictly convex. This means $\tilde{x}_1 = \tilde{x}_2 + m$ and $x_1 = x_2$, proving the claim.

The previous claim implies, since $\pi \circ \nabla\Phi^* = \nabla^c\phi^c \circ \pi$, that π maps $\nabla\Phi^*([0,1)^n)$ diffeomorphically to X. Further,

$$\det(\phi_{ij}+\delta_{ij}) \circ \pi = \det(\Phi_{ij}) = \frac{1}{\det(\Phi^*_{ij})} \tag{2.8}$$

and the numerator of the right hand side of (2.8) is the Jacobian determinant of the map $\nabla\Phi^* : \mathbb{R}^n \to \mathbb{R}^n$. Let $h \in C(X)$. Then

$$\begin{aligned}\int_X h\det(\phi_{ij}+\delta_{ij})dx &= \int_{\nabla\Phi^*([0,1)^n)} \frac{h\circ\pi}{\det(\Phi^*_{ij})}dx = \int_{[0,1)^n} h\circ\pi\circ\nabla\Phi^* dx \\ &= \int_{[0,1)^n} h\circ\nabla^c\phi^c\circ\pi dx = \int_X h\circ\nabla^c\phi^c dx. \end{aligned} \tag{2.9}$$

which proves the lemma. □

2.3. Kantorovich Duality. We now return to the problem of optimal transport. Although it has very satisfactory solutions providing existence and characterization of minimizers under great generality, we will only give part of that picture here. For us, the important feature of the problem of optimal transport is its dual formulation. Introducing the functional ξ on $C(X)$ defined by

$$\xi(\phi) = \int_X \phi^c dx$$

we get a functional J on $C(X)$

$$J(\phi) = -\int_X \phi d\mu - \xi(\phi).$$

This functional describes the dual formulation of the problem of optimal transport in the sense that $W^2(\mu, dx)$ can be recovered as the supremum of J over $C(X)$. Moreover, the maximizers of J are weak solutions to a certain Monge-Ampère equation. This is recorded in the following theorem.

THEOREM 2.4 ([18],[19],[5]). *Let $\mu \in \mathcal{M}_1(X)$ be absolutely continuous with respect to dx. Let $c = d^2/2$ where d is the distance function on X induced from $\mathbb{R}^n$. Then*

$$W^2(\mu, dx) = \inf_{\gamma\in\Pi(\mu,dx)} I(\gamma) = \sup_{\phi\in C(X)} J(\phi). \tag{2.10}$$

and there is $\phi_\mu \in P(X)$ such that

$$\sup_{\phi\in C(X)} J(\phi) = J(\phi_\mu). \tag{2.11}$$

Moreover,

$$\mathrm{MA}(\phi_\mu) = \mu. \tag{2.12}$$

REMARK 2.5. Equation 2.10 is called Kantorovich' duality [18] and property (2.12) is the Knott-Smith criterion which, in the context of Monge's problem of optimal transport, was discovered independently by Knott and Smith in 1984 [19] and by Brenier in 1987 [5].

PROOF OF THEOREM 2.4. The theorem is essentially given by Theorem 5.10 in [25]. As X is a smooth manifold that can be endowed with a complete metric, X is indeed a Polish space. Further, d is continuous and bounded on X. Putting $\gamma' = \mu \times dx$ gives

$$\inf_{\gamma\in\mathcal{P}(\mu,dx)} I(\gamma) \le I(\gamma') < \infty$$

hence the assumptions in 5.10.i, 5.10.ii and 5.10.iii in [25] holds. In particular we get that (2.10) holds and that there is an optimal transport plan $\gamma \in \Pi(\mu, dx)$ and $\phi_\gamma \in P(X)$ such that γ is concentrated on the set

$$\{(x, y) \in X \times X : \phi_\gamma(x) + \phi^c_\gamma(y) = -c(x, y)\}. \tag{2.13}$$

Let $\phi_\mu = \phi_\gamma$. To see that (2.11) holds, note that, since the first and second marginals of γ are μ and ν respectively,

$$\begin{aligned} W^2(\mu, dx) &= \int_{X\times X} c\gamma = -\int_{X\times X} \left(\phi_\mu(x) + \phi^c_\mu(y)\right)\gamma \\ &= -\int_X \phi_\mu(x)d\mu - \int_X \phi^c_\mu(y)dx. \end{aligned}$$

To see that (2.12) holds note that $\phi_\mu^c \in P(X)$ is differentiable almost everywhere with respect to dx. Let $A \subset X$ be a measurable set and $\operatorname{dom} \nabla^c \phi_\mu^c \subset X$ be the set where ϕ_μ^c is differentiable. We have

$$\gamma(X \times \operatorname{dom} \nabla^c \phi_\mu^c) = dx(\operatorname{dom} \nabla^c \phi_\mu^c) = 1.$$

As γ is concentrated on (2.13) we get that γ is concentrated on the set

$$\{(x,y) : y \in \operatorname{dom} \nabla^c \phi_\mu^c, x = \nabla^c \phi_\mu^c(y)\}.$$

This means

$$\begin{aligned} \int_{(\nabla^c \phi_\mu^c)^{-1}(A)} dx &= \int_{X \times (\nabla^c \phi_\mu^c)^{-1}(A)} d\gamma = \int_{A \times (\nabla^c \phi_\mu^c)^{-1}(A)} d\gamma \\ &= \int_{A \times X} d\gamma = \int_A d\mu, \end{aligned}$$

in other words $(\nabla^c \phi_\mu^c)_* dx = \mu$, which proves (2.12). □

2.4. The Variational Approach to Real Monge-Ampère Equations. We will now reformulate the statement of Theorem 2.4 in terms of the Legendre transform and Gateaux differentiability of the functional ξ. Recall that if A is a functional on $C(X)$, then the *Legendre* transform of A is a functional on the dual vector space of $C(X)$, the space of finite signed measures on X, $\mathcal{M}(X)$. This functional is given by

$$B(\mu) = \sup_{\phi \in C(X)} \int_Y \phi d\mu - A(\phi).$$

Recall also that if A is convex, then A is Gateaux differentiable at a point ϕ and has Gateaux differential μ if μ is the unique point in $\mathcal{M}(X)$ such that

$$B(\mu) = \int_Y \phi d\mu - A(\phi).$$

A priory $W^2(\cdot, dx)$ is defined on $\mathcal{M}_1(X)$. However, we may extend the definition to all of $\mathcal{M}(X)$ by putting $W(\mu, dx) = +\infty$ for any $\mu \notin \mathcal{M}_1(X)$. We begin with the following lemma

Lemma 2.6. *The functional ξ is convex on $C(X)$. Moreover, let $\phi_0, \phi_1 \in C(X)$ and*

$$\phi_t = t\phi_1 + (1-t)\phi_0.$$

Then, if $\xi(\phi_t)$ is affine in t,

$$\nabla^c \phi_0^c = \nabla^c \phi_1^c$$

almost everywhere with respect to dx.

Proof. First of all, for any $y \in X$, the quantity

$$\phi_t^c(y) = \sup_{x \in X} -c(x,y) - \phi_t(x) \tag{2.14}$$

is a supremum of functions that are affine in t, hence it is convex in t. This implies $\xi(\phi_t)$ is convex in t. Now, assume $\xi(\phi_t)$ is affine in t. This implies (2.14) is affine in t for almost all y. Assume y is a point such that $\nabla^c \phi_0^c(y)$, $\nabla^c \phi_{1/2}^c(y)$ and $\nabla^c \phi_1^c(y)$ are defined and (2.14) is affine. Let $x_{1/2} = \nabla^c \phi_{1/2}^c(y)$. This means

$$\phi_{1/2}^c(y) = -c(x_{1/2}, y) - \phi_{1/2}(x_{1/2}).$$

By construction

$$\phi_t^c(y) \geq -c(x_{1/2}, y) - \phi_t(x_{1/2})$$

for any $t \in [0,1]$. As ϕ_t^c and $-c(x_{1/2}, y) - \phi_t(x_{1/2})$ are affine functions (in t) that coincide in one point in the interior of their domains, this inequality implies that they coincide. This means $\nabla^c \phi_0^c(y) = \nabla^c \phi_{1/2}^c(y) = \nabla^c \phi_1^c(y)$. As $\nabla^c \phi_0^c$, $\nabla^c \phi_{1/2}^c$ and $\nabla^c \phi_1^c$ are defined almost everywhere, this proves the lemma. □

This allow us to draw the following conclusions from Theorem 2.4

Corollary 2.7. *The functional on $\mathcal{M}(X)$ defined by $\mu \mapsto W^2(-\mu, dx)$ is the Legendre transform of ξ. Moreover, for any $\mu \in \mathcal{M}_1(X)$ there is ϕ such that*

$$W^2(\mu, dx) + \xi(\phi) = -\int_X \phi d\mu.$$

Finally, ξ is Gateaux differentiable on $C(X)$ and

$$d\xi|_\phi = -\operatorname{MA}(\phi). \tag{2.15}$$

Proof. The first statement is, as long as $\mu \in \mathcal{M}_1(X)$, a direct consequence of (2.10). If $\mu \notin \mathcal{M}_1(X)$ then putting $\phi_C = \phi + C$ for some $\phi \in C(X)$ and $C \in \mathbb{R}$ gives $(\phi_C)^c = \phi^c - C$ and

$$-\int_X \phi_C d\mu - \xi(\phi_C) = -\int_X \phi d\mu - \xi(\phi) + C(1 - \mu(X)).$$

Letting $C \to \infty$ if $\mu(X) < 1$ and $C \to -\infty$ if $\mu(X) > 1$ gives

$$\sup_{\phi \in C(X)} \phi d\mu - \xi(\phi) = +\infty,$$

proving the first statement. The second statement is also a direct consequence of Theorem 2.4. We will now prove that ξ is Gateaux differentiable and that (2.15) holds. Let $\phi \in C(X)$. We claim that there is $\mu \in \mathcal{M}(X)$ such that

$$\xi(\phi) + W^2(\mu, dx) = -\int \phi d\mu, \tag{2.16}$$

in other words μ is a supporting hyperplane of ξ at ϕ. To see this, note that since $W^2(-\cdot, dx)$ is the Legendre transform of ξ we get that $W^2(\cdot, dx)$ is lower semi-continuous and

$$\xi(\phi) + W^2(\mu, dx) \geq -\int \phi d\mu \tag{2.17}$$

for all $\mu \in \mathcal{M}(X)$. By lemma 2.6, ξ is convex on $C(X)$. By the involutive property of Legendre transform

$$\xi(\phi) = \sup_{\mu \in \mathcal{M}(X)} -\int_X \phi d\mu - W^2(\mu, dx).$$

Let $\{\mu_i\} \subset \mathcal{M}(X)$ be a sequence such that

$$-\int_X \phi d\mu_i - W^2(\mu_i, dx) \to \xi(\phi).$$

We may assume, since $W^2(\mu_i, dx) = \infty$ if $\mu_i \notin \mathcal{M}_1(X)$, that $\mu_i \in \mathcal{M}_1(X)$ for all i. Since $\mathcal{M}_1(X)$ is compact we may take a subsequence $\{\mu_{i_k}\}$ converging to some $\mu \in \mathcal{M}_1(X)$. By the lower semi-continuity of $W^2(\cdot, dx)$ we get

$$-\int_X \phi\mu - W^2(\mu, dx) \geq \liminf_{k\to\infty} -\int_X \phi_0 \mu_{i_k} - W^2(\mu_{i_k}, dx) = \xi(\phi_0).$$

which, together with (2.17), proves the claim. We will now prove that this implies

$$(\nabla^c \phi^c)_* dx = \mu. \tag{2.18}$$

As this relation determines μ we get that μ must be the unique supporting hyperplane at ϕ. This implies ξ is Gateaux differentiable at ϕ and $d\xi_\phi = \mu$, proving the second statement in the corollary.

Now, to see that (2.18) holds, note that (2.16) implies $W^2(\mu, dx) < \infty$ and hence $\mu \in M_1(X)$. By Theorem 2.4 there is a function $\phi_\mu \in P(X)$ such that $\mathrm{MA}(\phi_\mu) = \mu$ and

$$W^2(\mu, dx) + \xi(\phi_\mu) = -\int \phi_\mu d\mu.$$

This means μ is a supporting hyperplane of ξ both at ϕ and at ϕ_μ. This implies $\xi(t\phi+(1-t)\phi_\mu)$ is affine. By Lemma 2.6, $\nabla^c\phi^c$ and $\nabla^c\phi^c_\mu$ coincide almost everywhere with respect to dx and hence (2.18) holds. □

3. A Large Deviation Principle

This section is devoted to Theorem 3.2 which will be the key part in the proof of Theorem 1.1. Before we state Theorem 3.2 we will recall the definition of the relative entropy function.

Definition 3.1. *Assume $\mu, \mu_0 \in \mathcal{M}(X)$ and, if μ is absolutely continuous with respect to μ, let μ/μ_0 denote the density of μ with respect to μ_0. The* relative entropy *of μ with respect to μ_0 is*

$$Ent_{\mu_0}(\mu) = \begin{cases} \int_X \mu \log \frac{\mu}{\mu_0} & \text{if } \mu \text{ is a probability measure and absolutely} \\ & \text{continuous with respect to } \mu_0 \\ +\infty & \text{otherwise,} \end{cases}$$

We recall the basic property that $Ent_{\mu_0}(\mu) \geq 0$ with equality if and only if $\mu = \mu_0$.

Theorem 3.2. *Let $\mu_0 \in \mathcal{M}_1(X)$ be absolutely continuous and have positive density with respect to dx. Let $\beta \in \mathbb{R}$. Assume $\Gamma^{(N)}_\beta$ is defined as in section 1.2. Then*

$$\left\{\Gamma^{(N)}_\beta\right\}$$

satisfy a Large Deviation Principle *with rate $r_N = N$ and rate function*

$$G(\mu) = \beta W^2(\mu, dx) + Ent_{\mu_0}(\mu) + C_{\mu_0,\beta}$$

where $W^2(\mu, dx)$ is the squared Wasserstein 2-distance between dx and μ_0 (defined in the previous section) and $C_{\mu_0,\beta}$ is a constant ensuring $\inf_{\mathcal{M}_1(X)} G = 0$.

Before we move on we will recall the definition of a Large Deviation Principle.

Definition 3.3. *Let χ be a topological space, $\{\Gamma_N\}$ a sequence of probability measures on χ, G a lower semi continuous function on χ and r_N a sequence of numbers such that $r_N \to \infty$. Then $\{\Gamma_N\}$ satisfies a* large deviation principle *with rate function G and rate r_N if, for all measurable $E \subset \chi$,*

$$-\inf_{E^\circ} G \leq \liminf_{N\to\infty} \frac{1}{r_N} \log \Gamma_N(E) \leq \limsup_{N\to\infty} \frac{1}{r_N} \log \Gamma_N(E) \leq -\inf_{\bar{E}} G$$

where E° and $\bar{E}$ are the interior and the closure of E.

In our case $\chi = \mathcal{M}_1(X)$. As we may endow $\mathcal{M}_1(X)$ with the Wasserstein 1-metric, metricizing the topology of weak* convergence on χ, we may think of $\mathcal{M}_1(X)$ as a metric space. Further, by Prohorov's Theorem, $\mathcal{M}_1(X)$ is compact. In this setting there is an alternative criteria for when a large deviation principle exist.

LEMMA 3.4. *Let χ be a compact metric space, $\{\Gamma_N\}$ a sequence of probability measures on χ, G a function on χ and r_N a sequence of numbers such that $r_N \to \infty$. Let $B_d(\mu)$ denote the open ball in χ with center μ and radius d. Then $\{\Gamma_N\}$ satisfies a* large deviation principle *with rate function G and rate r_N if and only if, for all $\mu \in \chi$*

$$\begin{aligned} G(\mu) &= \lim_{\delta\to 0} \limsup_{N\to\infty} -\frac{1}{r_N} \log \Gamma_N(B_\delta(\mu)) \\ &= \lim_{\delta\to 0} \liminf_{N\to\infty} -\frac{1}{r_N} \log \Gamma_N(B_\delta(\mu)) \end{aligned}$$

PROOF. Let $\mathcal{B}$ be the basis of the topology on χ given by

$$\mathcal{B} = \{B_d(\mu) : d > 0, \mu \in \chi\}.$$

By Theorem 4.1.11, Theorem 4.1.18 and Lemma 1.2.18 (recall that χ is compact by assumption) in [13], $\{\Gamma_N\}$ satisfies a large deviation principle with rate function G and rate r_N if and only if

$$\begin{aligned} G(\mu) &= \sup_{B\in\mathcal{B}:\mu\in B} \limsup_{N\to\infty} -\frac{1}{r_N} \log \Gamma_N(B_\delta(\mu)) \\ &= \sup_{B\in\mathcal{B}:\mu\in B} \liminf_{N\to\infty} -\frac{1}{r_N} \log \Gamma_N(B_\delta(\mu)). \end{aligned}$$

Now, if $\mu \in B \in \mathcal{B}$ then $B_d(\mu) \subset B$ for d small enough. This means, since

$$\lim_{d\to 0} \limsup_{N\to\infty} -\frac{1}{r_N} \log \Gamma_N(B_\delta(\mu)) \tag{3.1}$$

is increasing as $d \to 0$, that

$$(3.1) \geq \sup_{B\in\mathcal{B}:\mu\in B} \limsup_{N\to\infty} -\frac{1}{r_N} \log \Gamma_N(B_\delta(\mu)). \tag{3.2}$$

Since, for any $d > 0$, $B_d(\mu)$ is a candidate for the supremum in the right hand side of (3.2) we get that equality must hold in (3.2). The same argument goes through with lim sup replaced by lim inf. This proves the lemma. □

Finally we recall the well known

THEOREM 3.5 (Sanov's theorem, see for example 6.2.10 in [13]). *Let $\mu_0 \in \mathcal{M}_1(X)$. Then the family*

$$\left\{ \left(\delta^{(N)} \right)_* \mu_0^{\otimes N} \right\}$$

satisfies a large deviation principle with rate $r_N = N$ and rate function Ent_{μ_0}.

3.1. The Zero Temperature Case and the Gärtner-Theorem. Recall that $N = k^n$. For each $\beta \in \mathbb{R}$ we get a family of probability measures $\{\Gamma_\beta^{(N)}\}_{k \in \mathbb{N}}$. Theorem 1.1 and Theorem 3.2 are both concerned with the behavior of these families. In this section we will consider the family $\{\Gamma_k^{(N)}\}_{k \in \mathbb{N}}$. We will prove a large deviation principle for this family (see Theorem 3.6) which, in Section 3.2, will be used to prove Theorem 3.2.

THEOREM 3.6. *Let $\mu_0 \in \mathcal{M}_1(X)$ be absolutely continuous and have positive density with respect to dx. Assume $\Gamma_\beta^{(N)}$ is defined as in section 1.2. Then*

$$\{\Gamma_k^{(N)}\}$$

satisfies a large deviation principle with rate $r_N = kN$ and rate function

$$G(\mu) = W^2(\cdot, dx).$$

Recall that if Γ is a probability measure on a topological vector space χ, then the moment generating function of Γ is the functional on the dual vector space χ^* given by

$$Z_\Gamma(\phi) = \int_\chi e^{-\langle \phi, \mu \rangle} d\Gamma(\mu)$$

where $\langle \cdot, \cdot \rangle$ is the pairing of χ and χ^*. The significance of this for our purposes lies in the Gärtner-Ellis theorem. Before we state this theorem, recall that a sequence of (Borel) probability measures $\{\Gamma_N\}$ on a space χ is *exponentially tight* if for each $\epsilon \in \mathbb{R}$ there is a compact $K_\epsilon \subseteq \chi$ such that for all N

$$\limsup_{\mathbb{N} \to \infty} \frac{1}{N} \log \Gamma_N(\chi \setminus K_\epsilon) \le \epsilon. \tag{3.3}$$

In our case, when χ is compact, this is automatically satisfied since choosing $K_\epsilon = \chi$ for any ϵ gives that the left hand side of (3.3) is $-\infty$ for all N.

THEOREM 3.7 (The Gärtner-Ellis Theorem. See for example Corollary 4.5.27 in [13]). *Let χ be a locally convex topological vector space, $\{\Gamma_N\}$ an exponentially tight sequence of probability measures on χ and r_N a sequence such that $r_N \to \infty$. Let Z_{Γ_N} be the moment generating function of Γ_N and assume*

$$F(\phi) = \lim_{N\to\infty} \frac{1}{r_N} \log Z_{\Gamma_N}(r_N \phi)$$

exist, is finite valued, lower semi continuous and Gateaux differentiable. Then Γ_N satisfies a large deviation principle with rate r_N and rate function given by the Legendre transform of F.

Theorem 3.6 will follow from the Gärtner-Ellis theorem and the crucial point will be the following lemma.

LEMMA 3.8. *Let $\mu_0 \in \mathcal{M}_1(X)$ be absolutely continuous and have positive density with respect to dx. Assume $\Gamma_\beta^{(N)}$ is defined as in section 1.2. Then*

$$\lim_{N\to\infty} \frac{1}{kN} \log Z_{\Gamma^{(N)}}(kN\phi) = \xi(-\phi).$$

PROOF. Note that if μ_N is a measure on X^N and F is a function on $\mathcal{M}_1(X)$, then, since $\Gamma^{(N)} = (\delta^{(N)})_* \mu_{\beta_N}^{(N)}$,

$$\int_{\mathcal{M}_1(X)} F(\mu)\Gamma^{(N)} = \int_{X^N} F\left(\delta^{(N)}(x)\right) d\mu_{\beta_N}^{(N)}.$$

Moreover,

$$\left\langle kN\phi, \delta^{(N)}(x) \right\rangle = kN \int_X \phi \frac{1}{N} \sum \delta_{x_i} = k \sum \phi(x_i).$$

This means

$$Z_{\Gamma^{(N)}}(kN\phi) = \int_{\mathcal{M}_1(X)} e^{\langle r_N \phi, \mu\rangle} \Gamma^{(N)} = \int_{X^N} e^{k\sum \phi(x_i)} d\mu_{\beta_N}^{(N)}.$$

Using the symmetries in the explicit form of $\mu_{\beta_N}^{(N)}$ we get

$$\begin{aligned} Z_{\Gamma^{(N)}}(kN\phi) &= \int_{X^N} \sum_{\sigma} \prod_i \Psi_{p_i}^{(N)}(x_{\sigma(i)}) e^{k\phi(x_{\sigma(i)})} d\mu_0^{\otimes N} \\ &= \sum_{\sigma} \int_{\sigma^{-1}(X^N)} \prod_i \Psi_{p_i}^{(N)}(x_i) e^{k\phi(x_i)} d\mu_0^{\otimes N} \\ &= N! \int_{X^N} \prod_i \Psi_{p_i}^{(N)}(x_i) e^{k\phi(x_i)} d\mu_0^{\otimes N} \\ &= N! \prod_i \int_X \Psi_{p_i}^{(N)}(x) e^{k\phi(x)} d\mu_0 \end{aligned} \tag{3.4}$$

Introducing the notation

$$c_p^{(N)} = -\frac{1}{k} \log \Psi_p^{(N)}$$

we get

$$Z_{\Gamma^{(N)}}(kN\phi) = N! \prod_{p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n} \int_X e^{k(-c_p^{(N)}+\phi)} d\mu_0. \tag{3.5}$$

Now, we claim that

$$c_p^{(N)} \to d(x,p)^2/2 \tag{3.6}$$

uniformly in p and x. To see this, note first that

$$d(x,p)^2 = \inf_{m \in \mathbb{Z}^n + p} |x-m|^2$$

and

$$\begin{aligned} c_p^{(N)}(x) &= -\frac{1}{k} \log \sum_{m \in \mathbb{Z}^n + p} e^{-k|x-m|^2/2} \\ &\leq -\frac{1}{k} \log \sup_{m \in \mathbb{Z}^n + p} e^{-k|x-m|^2/2} = \inf_{m \in \mathbb{Z}^n + p} |x-m|^2/2. \end{aligned}$$

On the other hand, by the exponential decay of $e^{-|x-m|^2}$ there is a large constant, C, (independent of x and p) such that

$$\sum_{m \in \mathbb{Z}^n + p} e^{-k|x-m|^2/2} \leq C \sup e^{-k|x-m|^2/2}$$

and

$$\begin{aligned} c_p^{(N)}(x) &= -\frac{1}{k}\log \sum_{m\in\mathbb{Z}^n+p} e^{-k|x-m|^2/2} \geq -\frac{1}{k}\log\left(C \sup_{m\in\mathbb{Z}^n+p} e^{-k|x-m|^2/2}\right) \\ &= -\frac{\log C}{k} + \inf_{m\in\mathbb{Z}^n+p} |x-m|^2/2. \end{aligned}$$

This proves the claim. We claim further that

$$\frac{1}{k}\log\int_X e^{k(-c_p^{(N)}+\phi)}d\mu_0 \to (-\phi)^c(p) \tag{3.7}$$

uniformly in p. To see this, note first that (3.6) together with the fact that the family $\{d(\cdot,p)^2/2 : p\in X\}$ is equi-continuous implies that

$$\{c_p^{(N)} : k\in\mathbb{N}, p\in X\}$$

is equi-continuous. This means for any $\epsilon>0$ there is $d>0$ such that for all $k\in\mathbb{N}$ and $p, x_*\in X$

$$|c_p^{(N)}(x)-\phi(x)-(c_p^{(N)}(x_*)-\phi(x_*))|\leq\epsilon \tag{3.8}$$

as long as $x\in B_d(x_*)$. Further, as μ_0 has full support, is absolutely continuous and has smooth density with respect to dx there is a large constant C such that

$$C\mu_0(B_d(x_*))\geq 1 \tag{3.9}$$

for all $x_*\in X$. We get trivially

$$\begin{aligned} \frac{1}{k}\log\int_X e^{k(-c_p^{(N)}+\phi)}d\mu_0 &\leq \frac{1}{k}\log\sup_{x\in X} e^{k(-c_p^{(N)}+\phi)} \\ &= \sup -c_p^{(N)}(x)+\phi(x) \end{aligned} \tag{3.10}$$

For each N, let $x_*^{(N)}$ satisfy

$$-c_p^{(N)}(x_*^{(N)})+\phi(x_*^{(N)}) = \sup_{x\in X} -c_p^{(N)}(x)+\phi(x).$$

Using (3.8) and (3.9) gives

$$\begin{aligned} \frac{1}{k}\log\int_X e^{k(-c_p^{(N)}+\phi)}d\mu_0 &\geq \frac{1}{k}\log\int_{B_\delta(x_*^{(N)})} e^{k(\sup_{x\in X}-c_p^{(N)}+\phi-\epsilon)}d\mu_0 \\ &= \frac{1}{k}\log\int_{B_\delta(x_*^{(N)})} d\mu_0 + \sup_{x\in X} -c_p^{(N)}(x)+\phi(x)-\epsilon \\ &\geq \frac{1}{k}\log\frac{1}{C}\int_X d\mu_0 + \sup_{x\in X} -c_p^{(N)}(x)+\phi(x) -\epsilon \end{aligned} \tag{3.11}$$

Finally, letting $k, N \to \infty$ and $\epsilon \to 0$ in (3.10) and (3.11) proves (3.7). Recalling equation (3.5), we have

$$\begin{aligned} \frac{1}{kN} \log Z_{\Gamma^{(N)}}(kN\phi) &= \frac{1}{kN} \log N! \prod_{p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n} \int_X e^{k(-c_p^{(N)}+\phi)} d\mu_0 \\ &= \frac{\log N!}{kN} + \frac{1}{N} \sum_{p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n} \frac{1}{k} \log \int_X e^{k(-c_p^{(N)}+\phi)} d\mu_0 \end{aligned} \tag{3.12}$$

By Sterling's formula, $\log N! \leq N \log N + O(\log N)$. This means, since $N = k^n$, that the first term in (3.12) is bounded by $(\log k^n)/k + O(\log k^n)/k^{n+1}$ which vanishes as $k \to \infty$. Finally, using (3.7) we get, since $\frac{1}{N}\sum_{p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n} \delta_p \to dx$ in the weak* topology, that the second term converges to

$$\int (-\phi)^c(p)dx = \xi(-\phi).$$

This proves the lemma. $\square$

When proving Theorem 3.6 we will also need the following lemma.

Lemma 3.9. *The functional ξ is continuous on $C(X)$.*

Proof. We will prove that for any $\phi_0, \phi_1 \in C(X)$

$$\sup_X |\phi_0^c - \phi_1^c| \leq \sup_X |\phi_1 - \phi_0|. \tag{3.13}$$

Once this is established the lemma follows from the dominated convergence theorem. To see that (3.13) holds, let $y \in X$. By compactness and continuity there is $x_y \in X$ such that

$$\phi_0^c(y) = \sup_{x \in X} -c(x, y) + \phi_0(x) = -c(x_y, y) - \phi_0(x_y).$$

By construction

$$\phi_1^c(y) = \sup_{x \in X} -c(x, y) + \phi_1(x) \geq -c(x_y, y) - \phi_1(x_y).$$

We get

$$\phi_0^c(y) - \phi_1^c(y) \leq \phi_1(x_y) - \phi_0(x_y) \leq \sup_X |\phi_1 - \phi_0|.$$

By interchanging the roles of ϕ_0 and ϕ_1 we get

$$\phi_1^c(y) - \phi_0^c(y) \leq \sup_X |\phi_1 - \phi_0|$$

and hence that (3.13) holds. $\square$

Proof of Theorem 3.6. We want to apply the Gärtner-Ellis theorem. As $\chi = M_1(X)$ is compact, tightness of $\Gamma^{(N)}$ holds automatically. By Lemma 3.8

$$\lim_{N\to\infty} \frac{1}{r_N} \log \Lambda_{\Gamma_k^{(N)}}(r_N \phi) = \xi(-\phi).$$

Further, ξ is finite valued since ϕ^c is continuous, and hence bounded, for any $\phi \in C(X)$. By Lemma 3.9, ξ is continuous. Finally, by Corollary 2.7, ξ is Gateaux differentiable. As $W^2(-\cdot, dx)$ is the Legendre transform of ξ, and hence $W^2(\cdot, dx)$ is the Legendre transform of $\xi(-\cdot)$, the theorem follows from the Gärtner-Ellis theorem. $\square$

3.2. A Thermodynamic Interpretation and Reduction to the Zero Temperature Case. The proof of Theorem 3.2 is based on a result on large deviation principles for Gibbs measures. Because of this we explain in this section how $\{\mu_\beta^{(N)}\}$ can be seen as the Gibbs measures of certain thermodynamic systems. If we introduce the N-particle Hamiltonian

$$H^{(N)}(x_1, \ldots, x_N) = -\frac{1}{k} \log \operatorname{perm}(\Psi_{p_i}(x_j))$$

we may write $\mu_\beta^{(N)}$ on the form

$$\mu_\beta^{(N)} = e^{-\beta H^{(N)}} d\mu_0^{\otimes N}.$$

This means $\mu_\beta^{(N)}$ admits a thermodynamic interpretation as the *Gibbs measure*, or *canonical ensemble*, of the system determined by the Hamiltonian $H^{(N)}$ and the background measure μ_0. In this interpretation $\mu_\beta^{(N)}$ is the equilibrium state of the system when the temperature is assumed fixed at $Temp = 1/\beta$ and Theorem 3.6 is describing the zero-temperature limit. Theorem 3.2 will follow from Theorem 3.6 and a theorem on equicontinuous and uniformly bounded Hamiltonians. To state that theorem we need to define what it means for the family $\{\frac{H^{(N)}}{N}\}$ to be equicontinuous. Let $d(\cdot,\cdot)$ be the distance function induced by the standard Riemannian metric on X. This defines distance functions, $d^{(N)}(\cdot,\cdot)$, on X^N given by

$$d^{(N)}(x, y) = d^{(N)}(x_1, \ldots, x_N, y_1, \ldots, y_N) = \frac{1}{N} \inf_\sigma \sum_i d(x_i, y_{\sigma(i)}) \quad (3.14)$$

where the infimum is taken over all permutations σ of the set $\{1,\ldots,N\}$. We will say that the family of functions $\frac{H^{(N)}}{N}$ on X^N is (uniformly) equicontinuous if for every $\epsilon>0$ there is $d>0$ such that for all N

$$\left|\frac{1}{N}H^{(N)}(x)-\frac{1}{N}H^{(N)}(y)\right|\leq\epsilon \tag{3.15}$$

whenever $d^{(N)}(x,y)\leq d$. Before we move on to state the Theorem 3.11 we prove the following well known lemma.

LEMMA 3.10. *Let $x=(x_1,\ldots,x_N)\in X^N$ and $y=(y_1,\ldots,y_N)\in X^N$. Then* (3.14) *is the optimal transport cost with respect to the cost function $d(\cdot,\cdot)$, of transporting the measure $\delta^{(N)}(x)=\frac{1}{N}\sum\delta_{x_i}$ to the measure $\delta^{(N)}(y)=\frac{1}{N}\sum\delta_{y_i}$.*

PROOF. We need to prove that

$$(3.14)=\inf_{\gamma}\int_{X\times X}d(x,y)\gamma \tag{3.16}$$

where the infimum is taken over all $\gamma\in\mathcal{M}_1(X\times X)$ with first and second marginal given by $\delta^{(N)}(x)$ and $\delta^{(N)}(y)$ respectively. We will refer to any $\gamma\in\mathcal{M}_1(X\times X)$ satisfying this as a feasible transport plan. The conditions on the marginals imply that any feasible transport plan is supported on the intersection of the sets $\{x_i\}\times X$ and $X\times\{y_i\}$, in other words on the set $\{x_i\}\times\{y_i\}$. We conclude that the set of feasible transport plans is given by

$$\left\{\sum_{i,j}a_{ij}\delta_{(x_i,y_j)}:a_{ij}\geq 0,\sum_i a_{ij}=1/N,\sum_j a_{ij}=1/N\right\}, \tag{3.17}$$

in other words a polytope in $\mathcal{M}_1(X\times X)$. It follows that the infimum in (3.16) is attained on one or more of the vertices of (3.17). Moreover, any permutation, σ, of N elements induce a feasible transport plan

$$\gamma_\sigma=\frac{1}{N}\sum_i\delta_{(x_i,y_{\sigma(i)})}$$

with transport cost

$$\int_{X\times X}d(x,y)\gamma_\sigma=\frac{1}{N}\sum_i d(x_i,y_{\sigma(i)}).$$

It is easy to verify that any vertex of (3.17) occur as γ_σ for some permutation σ. This proves the lemma. $\square$

Note that this lemma implies that if we equip $\mathcal{M}_1(X)$ with the Wasserstein 1-metric, which metricizes the weak* topology on $\mathcal{M}_1(X)$, then the distance function defined in (3.14) makes the embeddings

$$\delta^{(N)} : X^N \hookrightarrow \mathcal{M}_1(X)$$

isometric embeddings.

THEOREM 3.11 ([1]). *Assume X is a compact manifold, $\mu_0 \in \mathcal{M}_1(X)$, $\{\frac{H^{(N)}}{N}\}$ is a uniformly bounded and equi-continuous family of functions on X^N and β_N is a sequence of numbers tending to infinity. Assume also that*

$$\left(\delta^{(N)}\right)_* e^{-\beta_N H^{(N)}} d\mu_0^{\otimes N}$$

satisfies a Large Deviation Principle with rate $N\beta_N$ and rate function E. Then, for any $\beta \in \mathbb{R}$,

$$\left(\delta^{(N)}\right)_* e^{-\beta H^{(N)}} d\mu_0^{\otimes N}$$

satisfies a Large Deviation Principle with rate N and rate function $\beta E + Ent_{\mu_0}$.

The proof is based on the following

PROPOSITION 3.12 ([15]). *Assume X is a compact manifold, $\mu_0 \in \mathcal{M}_1(X)$, $\beta \in \mathbb{R}$, $\{\frac{H^{(N)}}{N}\}$ is a family of functions on X^N. Assume also that there is a functional E on $\mathcal{M}_1(X)$ satisfying*

$$\sup_{X^N}\left|\frac{H^{(N)}}{N} - E \circ \delta^{(N)}\right| \to 0 \tag{3.18}$$

as $N \to \infty$. Then

$$\left(\delta^{(N)}\right)_* e^{-\beta H^{(N)}} \mu_0^{\otimes N}$$

satisfies a Large Deviation Principle with rate N and rate function $\beta E + Ent_{\mu_0}$.

PROOF. Let $\mu \in \mathcal{M}_1(X)$ and $B_d(\mu)$ be the ball of (Wasserstein-1) radius d centred at μ and

$$B_d^{(N)}(\mu) = (\delta^{(N)})^{-1}(B_d(\mu)) \subset X^N.$$

Using (3.18) we get

$$\begin{aligned}
& \lim_{d\to 0} \liminf_{N\to\infty} -\frac{1}{N}(\delta^{(N)})_* e^{-\beta H^{(N)}} \mu_0^{\otimes N}(B_d(\mu)) \\
= & \lim_{d\to 0} \liminf_{N\to\infty} -\frac{1}{N} \log \int_{B_d^{(N)}(\mu)} e^{-\beta H^N(x)} d\mu_0^{\otimes N} \\
= & \lim_{d\to 0} \liminf_{N\to\infty} -\frac{1}{N} \log \int_{B_d^{(N)}(\mu)} e^{-\beta N(E\circ\delta^{(N)}(x)+o(1))} d\mu_0^{\otimes N} \\
= & \beta E(\mu) + \lim_{d\to 0} \liminf_{N\to\infty} -\frac{1}{N} \log \int_{B_d^{(N)}(\mu)} d\mu_0^{\otimes N}. \qquad (3.19)
\end{aligned}$$

and similarily with lim inf replaced by lim sup (here $o(1) \to 0$ uniformly in x as $N \to \infty$). By Sanov's theorem $(\delta^{(N)})_* \mu_0^{\otimes N}$ satisfies a large deviation principle with rate N and rate function Ent_{μ_0}. Hence, by Lemma 3.4, the second term in (3.19) is $Ent_\gamma(\mu)$. Using Lemma 3.4 again, this proves the proposition. $\square$

It turns out that in the compact setting, under the assumptions of uniform boundedness and equi-continuity, the assumption of convergence in Proposition 3.12 always holds for some functional U on $\mathcal{M}_1(X)$.

LEMMA 3.13. *Assume X is a compact manifold, $\mu_0 \in \mathcal{M}_1(X)$ and $\{\frac{H^{(N)}}{N}\}$ is a uniformly bounded and equi-continuous family of functions on X^N. Then there is a function U on $\mathcal{M}_1(X)$ such that, after possibly passing to a subsequence,*

$$\sup_{X^N} \left|\frac{H^{(N)}(x)}{N} - U \circ \delta^{(N)}(x)\right| \to 0 \qquad (3.20)$$

as $N \to \infty$.

PROOF. Using the embeddings $\delta^{(N)} : X^n \hookrightarrow \mathcal{M}_1(X)$ the functions $H^{(N)}$ define a sequence of functionals, $\mathcal{H}^{(N)}$, defined on the subspaces $\delta^{(N)}(X^N) \subset \mathcal{M}_1(X)$. By a standard procedure (we will explain it below) it is possible to define an equi-continuous family of extensions, $\{U^{(N)}\}$, of $\frac{\mathcal{H}^{(N)}}{N}$ on $\mathcal{M}_1(X)$. By Arzelà-Ascoli theorem $U^{(N)}$, after possibly passing to a subsequence, will converge to a functional U satisfying (3.20). We may define the extensions $U^{(N)}$ in the following way: Note that by assumption the functions $\frac{H^{(N)}}{N}$ all satisfy the same modulus of continuity,

ω. We define $U^{(N)} : \mathcal{M}_1(X) \to \mathbb{R}$ as

$$U^{(N)}(\mu) = \inf_{\nu \in \delta^{(N)}(X^N)} \frac{\mathcal{H}^{(N)}(\nu)}{N} + \omega(d(\mu, \nu))$$

where $d(\cdot, \cdot)$ is the Wasserstein 1-distance on $\mathcal{M}_1(X)$. It follows from the definition of moduli of continuity that $U^{(N)} = \frac{\mathcal{H}^{(N)}}{N}$ on $\delta^{(N)}(X^N)$. As $\mathcal{M}_1(X)$ is compact we may take ω to be sub-additive. It follows that the function $\omega(d(\mu, \cdot))$ satisfies ω as modulus of continuity. This means $U^{(N)}$, being a supremum of functions satisfying ω, also satisfy ω. In particular the family $\{U^{(N)}\}$ is equi-continuous. $\square$

We can now prove Theorem 3.11.

Proof of Theorem 3.11. As above, let $B_d^{(N)}(\mu) = (\delta^{(N)})^{-1}(B_d(\mu)) \subset X^N$, where $B_d(\mu)$ is the ball in $\mathcal{M}_1(X)$ centered at μ with radius d. By the assumed Large Deviation Principle and Lemma 3.4, for any $\mu \in \mathcal{M}_1(X)$,

$$E(\mu) = \lim_{d \to 0} \liminf_{N \to \infty} -\frac{1}{N\beta_N} \log \int_{B_d^{(N)}(\mu)} e^{-\beta_N H^N} \mu_0^{\otimes N}.$$

On the other hand, by Lemma 3.13 there is a function U on $\mathcal{M}_1(X)$ such that, after possibly passing to a subsequence, (3.20) holds. This means

$$\begin{aligned} E(\mu) &= \lim_{\delta \to 0} \liminf_{N \to \infty} -\frac{1}{N\beta_N} \log \int_{B_d^{(N)}(\mu)} e^{-N\beta_N (U \circ \delta^{(N)} + o(1))} d\mu_0^{\otimes N} \\ &= U(\mu) + \lim_{\delta \to 0} \liminf_{N \to \infty} -\frac{1}{N\beta_N} \log \int_{B_d^{(N)}(\mu)} d\mu_0^{\otimes N} \qquad (3.21) \\ &= U(\mu). \end{aligned}$$

where the second term in (3.21) is zero by Sanov's theorem. This means $E = U$ and the theorem now follows from Proposition 3.12. $\square$

3.3. Proof of Theorem 3.2. To use Theorem 3.11 we need to verify that the family $\{H^{(N)}\}$ is equi-continuous. We will use the following two lemmas

Lemma 3.14. *The functions in $P(X)$ are Lipschitz with the Lipschitz constant $L = 1$.*

Proof. As the diameter of X is 1 we get that the set

$$\{d(\cdot, y)^2/2 : y \in X\}$$

is Lipschitz with the Lipschitz constant $L = 1$. Now, assume $\phi \in P(X)$ and $x_1, x_2 \in X$. By definition

$$\phi(x) = \sup_{y \in X} -d(x,y)^2/2 - \phi^c(y).$$

for all x. By compactness and continuity there is y_1 such that

$$\phi(x_1) = -d(x_1, y_1)^2/2 - \phi^c(y_1).$$

We have

$$\begin{aligned}\phi(x_2) &\geq -d(x_2,y_1)^2/2 - \phi^c(y_1) = \phi(x_1) - (d(x_2,y_1)^2/2 - d(x_1,y_1)^2/2) \\ &\geq \phi(x_1) - d(x_1,x_2).\end{aligned}$$

By interchanging the roles of x_1 and x_2 we get

$$\phi(x_1) \geq \phi(x_2) - d(x_1, x_2)$$

and hence

$$|\phi(x_1) - \phi(x_2)| \leq d(x_1, x_2). \quad \square$$

We say that a function, Φ, on $\mathbb{R}^n$ is λ-*convex* if $\Phi - \lambda \frac{x^2}{2}$ is convex.

Lemma 3.15. *Assume Φ_α is a family of functions on $\mathbb{R}^n$ parametrized over some set A. Assume that for all $\alpha \in A$, Φ_α is λ-convex. Let σ be a probability measure on A. Then*

$$\log \int e^{\Phi_\alpha} d\sigma(\alpha)$$

is λ-convex.

Proof. Assume first $\lambda = 0$. By the convexity of Φ_α in x and Hölder's inequality we get

$$\begin{aligned}\int_A e^{\Phi_\alpha(tx_1+(1-t)x_0)} d\sigma(\alpha) &\leq \int_A e^{t\Phi_\alpha(x_1)+(1-t)\Phi_\alpha(x_0)} d\sigma(\alpha) \\ &\leq \left(\int_A e^{\Phi_\alpha(x_1)} d\sigma(\alpha)\right)^t \left(\int_A e^{\Phi_\alpha(x_0)} d\sigma(\alpha)\right)^{(1-t)}\end{aligned}$$

and hence, taking the logarithm of both sides of this inequality,

$$\begin{aligned}&\log \textstyle\int_A e^{\Phi_\alpha(tx_1+(1-t)x_0)} d\sigma(\alpha) \\ \leq\ & t \log \textstyle\int_A e^{\Phi_\alpha(x_1)} d\sigma(\alpha) + (1-t) \log \int_X e^{\Phi_\alpha(x_0)} d\sigma(\alpha).\end{aligned}$$

For the general case, note that

$$\log \int_A e^{\Phi_\alpha(x)} d\sigma(\alpha) - \lambda \frac{x^2}{2} = \log \int_A e^{\Phi_\alpha(x) - \lambda x^2/2} d\sigma(\alpha)$$

which is convex by the case considered above. □

We get

COROLLARY 3.16. *The normalized energy functions*

$$\{H^{(N)}/N : k \in \mathbb{N}\}$$

is an equi-continuous family (in the sense of (3.15)*).*

PROOF. We claim that

$$c_p^{(N)} = \frac{1}{k} \log \sum_{m \in \mathbb{Z}^n + p} e^{-k|x-m|^2/2} \in P(X) \tag{3.22}$$

for all $p \in X$ and $k \in \mathbb{N}$. To prove the claim it suffices to prove that (3.22) is -1-convex. This follows from Lemma 3.15 as $-|x-m|^2/2$ is -1-convex for all $m \in \mathbb{R}^n$. Further, fixing all but one variable we get a function on X given by

$$\begin{aligned} x \quad \mapsto \quad & H^{(N)}(x_1, \ldots x_{i-1}, x, x_{i+1}, \ldots, x_n) \\ = \quad & \frac{1}{k} \log \sum_\sigma e^{-kc^{(N)}_{p_{\sigma(i)}}(x)} \prod_{j \neq i} e^{-kc^{(N)}_{p_{\sigma(j)}}(x_j)} \end{aligned}$$

By Lemma 3.15 this function is in $P(X)$. By Lemma 3.14 it satisfies the Lipschitz constant 1. This means, if $x = (x_1, \ldots x_N)$ and $y = (y_1, \ldots, y_N)$ are points in X^N, that

$$\begin{aligned} & |\tfrac{1}{N} H^{(N)}(x_1, \ldots, x_N) - \tfrac{1}{N} H^{(N)}(y_1, \ldots, y_N)| \\ \leq \quad & \tfrac{1}{N} \textstyle\sum_i \left| H^{(N)}(x_1, \ldots, x_{i-1}, y_i, \ldots y_N) - H^{(N)}(x_1, \ldots, x_i, y_{i+1}, \ldots y_N) \right| \\ \leq \quad & \textstyle\sum_i d(x_i, y_i). \end{aligned} \tag{3.23}$$

As $H^{(N)}$ is symmetric we may reorder $\{x_i\}$ so that

$$\sum_i d(x_i, y_i) = \inf_\sigma \sum_i d(x_i, y_{\sigma(i)})$$

and hence the right hand side of (3.23) equals $d^{(N)}(x, y)$. This implies $H^{(N)}/N$ is equi-continuous in the sense of (3.15). □

Proof of Theorem 3.2. By Theorem 3.6 and Theorem 3.11 we only need to verify that the family $\{H^{(N)}/N\}$ is uniformly bounded and equi-continuous. The latter was proved in Corollary 3.16. To see that $\{H^{(N)}/N\}$ is uniformly bounded recall that in the proof of Theorem 3.6 we proved that $-\frac{1}{k}\log\Psi_p^{(N)}(x) \to d(x,p)/2$ uniformly in x and p. Since $d(\cdot,\cdot)$ is bounded on $X \times X$ we get that there is constants $c, C \in \mathbb{R}$ such that, for all but finitely many N,

$$c \leq \frac{1}{k}\log\Psi_p^{(N)}(x) \leq C \tag{3.24}$$

for all x, p. As the functions $\{\frac{1}{k}\log\Psi_p^{(N)}\}$ are bounded on X and there is only finitely many functions for each N, we may choose c and C such that (3.24) holds for all N. We get

$$H^{(N)}(x)/N = \frac{1}{kN}\log\sum_\sigma\prod_i e^{\log\Psi_{p_i}(x)} \leq \frac{1}{kN}\log\sum_\sigma\prod_i e^{kC} = \frac{\log N!}{kN}+C$$

and

$$H^{(N)}(x)/N = \frac{1}{kN}\log\sum_\sigma\prod_i e^{\log\Psi_{p_i}(x)} \geq \frac{1}{kN}\log\prod_i e^{kc} = c$$

for all N and $x \in X^N$. This proves the theorem. □

4. The Rate Function and its Connection to Monge Ampère equations

In this section we will show how the rate function, G, in Theorem 3.2 is related to Monge-Ampère equations. More precisely, we will establish a variational approach to equation (1.1) and then show that, under a certain condition, the minimizers of the G are the Monge-Ampère measures of solutions to (1.1) (see Lemma 4.3). This will allow us to finish the proof of Theorem 1.1.

4.1. The Variational Approach to Equation (1.1)**.** In the variational approach to equation (1.1) it is convenient to consider its normalized version:

$$\mathrm{MA}(\phi) = \frac{e^{\beta\phi}\mu_0}{\int_X e^{\beta\phi}d\mu_0}. \tag{4.1}$$

We see that this equation is invariant under the action of $\mathbb{R}$ on $P(X)$ given by

$$C \mapsto (\phi \mapsto \phi + C). \tag{4.2}$$

Now, we will say that an equation admits a unique solution modulo $\mathbb{R}$ if, for any two solutions $\phi_1, \phi_2 \in C(X)$, $\phi_1 - \phi_2$ is constant. It is easy to verify that (1.1) admits a unique solution if and only if (4.1) admits a unique solution modulo $\mathbb{R}$. We will consider a certain energy functional whose stationary points correspond to weak solutions of (1.1). For given data (μ_0, β) this energy functional has the form

$$F(\phi) = \xi(\phi) + \frac{1}{\beta} I_{\mu_0}(\beta\phi).$$

where I_{μ_0} is defined as

$$I_{\mu_0}(\phi) = \log \int_X e^{\phi}\mu_0.$$

LEMMA 4.1. *Let $\beta \neq 0$. The functional I_{μ_0} is Gateaux differentiable and*

$$dI_{\mu_0}|_{\phi} = \frac{e^{\phi}\mu_0}{\int_X e^{\phi} d\mu_0}.$$

Consequently, F is Gatueux differentiable and ϕ is a stationary point of F if and only if ϕ is a weak solution (in the sense of Section 2.2) to (1.1).

PROOF. Let $v \in C(X)$. As v is bounded an application of the dominated convergence theorem gives

$$\begin{aligned} \frac{d}{dt}|_{t=0} I(\phi + tv) &= \frac{\frac{d}{dt}|_{t=0} \int_X e^{\phi+tv} d\mu_0}{\int_X e^{\phi} d\mu_0} \\ &= \frac{\int_X \frac{d}{dt}|_{t=0} e^{\phi+tv} d\mu_0}{\int_X e^{\phi} d\mu_0} \\ &= \frac{\int_X v e^{\phi} d\mu_0}{\int_X e^{\phi} d\mu_0}, \end{aligned}$$

proving the first two statements of the lemma. By Corollary 2.7, ξ is differentiable and $d\xi|_{\phi} = -\operatorname{MA}(\phi)$. This means F is Gateaux differentiable and

$$dF|_{\phi} = -\operatorname{MA}(\phi) + \frac{e^{\phi}\mu_0}{\int_X e^{\phi} d\mu_0}$$

proving the last statements of the lemma. $\square$

4.2. The Minimizers of the Gibbs Free Energy. We will use the following well know property of the relative entropy function in the proof of Lemma 4.3.

LEMMA 4.2. *Let $\mu \in \mathcal{M}_1(X)$ and $\phi \in C(X)$. Then*

$$I_{\mu_0}(\phi) + Ent_{\mu_0}(\mu) \geq \int_X \phi d\mu \tag{4.3}$$

with equality if and only if $\mu = dI_{\mu_0}|_\phi$.

PROOF. Assume first that μ is absolutely continuous with respect to μ_0 and μ_0 is absolutely continuous with respect to μ. By Jensen's inequality

$$\begin{aligned} I_{\mu_0}(\phi) &= \log \int_X e^\phi \frac{\mu_0}{\mu} d\mu \\ &\geq \int_X \phi d\mu - \int_X \log \frac{\mu}{\mu_0} d\mu \\ &= \int_X \phi d\mu - Ent_{\mu_0}(\mu) \end{aligned}$$

with equality if and only if $e^\phi \frac{\mu_0}{\mu}$ is constant, or, equivalently, μ is proportional to $e^\phi \mu_0$. As μ is a probability measure this means

$$\mu = \frac{e^\phi \mu_0}{\int_X e^\phi d\mu_0} = dI|_\phi$$

proving the lemma in this special case. If μ is not absolutely continuous with respect to μ_0 then $Ent_{\mu_0}(\mu) = +\infty$ and the equality holds trivially. Finally, when μ is absolutely continuous with respect to μ_0 but μ_0 is not absolutely continuous with respect to μ, then replacing μ_0 by $\chi\mu_0$, where χ is the characteristic function of the support of μ doesn't change the right hand side of (4.3). Since

$$I_{\mu_0}(\phi) \geq \log \int e^\phi \chi d\mu_0$$

this reduces this case to the case when μ_0 is absolutely continuous with respect to μ. $\square$

We can now prove Lemma 4.3.

LEMMA 4.3. *Assume $\beta \neq 0$, F admits a unique minimizer modulo $\mathbb{R}$ and ϕ_* is a minimizer of F. Then*

$$\mu_* = \mathrm{MA}(\phi_*) \tag{4.4}$$

is the unique minimizer of the rate function

$$G(\mu) = \beta W^2(\mu, dx) + Ent_{\mu_0}(\mu) + C_{\mu_0,\beta}$$

defined in Theorem 3.2.

REMARK 4.4. Note that $\phi_1 - \phi_2 = C$ implies $\phi_1^c - \phi_2^c = -C$ and hence

$$\mathrm{MA}(\phi_1) = (\nabla^c \phi_1^c)_* dx = (\nabla^c \phi_2^c)_* dx = \mathrm{MA}(\phi_2).$$

This means that, under the assumptions of Lemma 4.3, μ_* is uniquely determined by (4.4).

PROOF OF THEOREM 4.3. Note that by Corollary 2.7 and Lemma 4.2 we have, for all $\mu \in M_1(X)$ and $\phi \in C(X)$, the two inequalities

$$W^2(\mu, dx) + \xi(\phi) \geq -\int \phi d\mu \tag{4.5}$$

$$Ent_{\mu_0}(\mu) + I_{\mu_0}(\phi) \geq \int \phi d\mu \tag{4.6}$$

where equality in (4.5) is characterized by

$$d\xi|_\phi = -MA(\phi) = -\mu \tag{4.7}$$

and equality in (4.6) is characterized by $dI|_\phi = \mu$. We will start with the case $\beta > 0$. Let $\mu \in \mathcal{M}_1(X)$ and ϕ_* be the minimizer of F. Applying (4.5) to the pair μ and ϕ_* and (4.6) to the pair μ and $\beta\phi_*$ we get

$$\begin{aligned} G(\mu) &= \beta W^2(\mu, dx) + Ent(\mu) \\ &\geq -\beta \int \phi_* d\mu - \beta\xi(\phi_*) + \int \beta\phi_* d\mu - I(\beta\phi_*) \\ &= -\beta\left(\xi(\phi_*) + \frac{1}{\beta} I(\beta\phi_*)\right) = -\beta F(\phi_*) \end{aligned}$$

with equality if and only if $d\xi|_{\phi_*} = -MA(\phi_*) = -\mu$ and $\mu = dI|_{\phi_*}$ which, since $d\xi|_{\phi_*} + dI|_{\phi_*} = 0$, is true if and only if $\mu = \mathrm{MA}(\phi_*)$. For the case $\beta < 0$, let $\mu \in \mathcal{M}_1(X)$. By Corollary 2.7 we may take ϕ to satisfy equality in (4.5) and hence (4.7). A similar application of (4.5) and (4.6) as above, keeping in mind that we have equality in (4.5), give

$$\begin{aligned} G(\mu) &= \beta W^2(\mu, dx) + Ent(\mu) \\ &\geq -\beta \int \phi d\mu - \beta\xi(\phi) + \int \beta\phi d\mu - I(\beta\phi) \qquad (4.8) \\ &= -\beta\left(\xi(\phi) + \frac{1}{\beta} I(\beta\phi)\right) = -\beta F(\phi) \geq -\beta F(\phi_*). \qquad (4.9) \end{aligned}$$

Moreover, equality in (4.9) holds if and only if $\phi = \phi_*$. But that means $dI|_\phi = -d\xi|_\phi = \mu$, hence we have equality in (4.8) as well. This implies $G(\mu) \geq -\beta F(\phi_*)$ with equality if and only if $\mu = \mathrm{MA}(\phi_*)$. □

4.3. Proof of Theorem 1.1 and Corollary 1.4.

Proof of Theorem 1.1. Let ϕ_* be the unique solution to (1.1). It follows that (4.1) admits a unique solution modulo $\mathbb{R}$ and that ϕ_* is a solution to (4.1). Now, we will use two results from the next chapter. Namely that any stationary point of F is a *smooth* solution to (4.1) (see Section 5.2) and that F always admit a minimizer (see Section 5.1). Under our assumptions, this implies F admits a unique minimizer modulo $\mathbb{R}$ and that ϕ_* is a minimizer of F. Using Lemma 4.3 we get that G admits the unique minimizer μ_* satisfying $\mu_* = \mathrm{MA}(\phi_*)$.

We want to prove that $\Gamma^{(N)} \to \delta_{\mu_*}$ in the weak* topology on $\mathcal{M}_1(\mathcal{M}_1(X))$. By the Portmanteau Theorem it suffices to verify that

$$\limsup_{N\to\infty} \Gamma^{(N)}(F) \leq \delta_{\mu_*}(F) \qquad (4.10)$$

for all closed $F \subset \mathcal{M}_1(X)$. If $\mu_* \in F$ then (4.10) holds trivially. Assume $\mu_* \notin F$. Recall that $\mathcal{M}_1(X)$ is compact. This means the closed subset F is compact. Since G is lower semi-continuous there is $\mu_F \in F$ such that $\inf_F G = G(\mu_F)$. As $\mu_* \notin F$ is the unique point where $G = \inf G = 0$ we get that $G(\mu_F) = \inf_F G > 0$. By the large deviation principle in Theorem 3.2

$$\limsup_{N\to\infty} \frac{1}{r_N} \log \Gamma^{(N)}(F) \leq -\inf_F G < 0.$$

As $r_N \to \infty$ we get that $\limsup \log \Gamma^{(N)}(F) = -\infty$ and $\limsup \Gamma^{(N)}(F) = 0$. This proves the theorem. □

Proof of Corollary 1.4. Equation (1.7) implies the first marginals of $\mu_\beta^{(N)}$,

$$\int_{X^{N-1}} \mu_\beta^{(N)},$$

converges to μ_* in the weak* topology of $M_1(X)$ (see Proposition 2.2 in [23]). Now, $e^{\beta\phi_N}$ is the density with respect to μ_0 of the first marginal of $\mu_\beta^{(N)}$.

We claim that the collection $\{\phi^{(N)} : k \in \mathbb{N}\}$ is equi-continuous and uniformly bounded. To see this, note that by Lemma 3.15, $\phi^{(N)}$ is -1-convex and hence in $P(X)$. By Lemma 3.14 the functions $\{\phi^{(N)}, k \in \mathbb{N}\}$ satisfy the Lipschitz constant $L = 1$. As

$$\int_X e^{\beta\phi_N}\mu_0 = \int_{X^N} \mu_\beta^{(N)} = 1$$

for all N, this means there are constants $c, C \in \mathbb{R}$, independent of N, such that $c \leq \phi_N \leq C$. This proves the claim. By the Arzelà-Ascoli theorem there is some function $\phi_\infty \in C(X)$ such that

$$\phi_N \to \phi_\infty$$

uniformly. As

$$e^{\beta\phi_N}\mu_0 = \int_{X^{N-1}} \mu_\beta^{(N)} \to \mu_* = e^{\beta\phi_*}\mu_0$$

in the weak* topology of $M_1(X)$ we get that $\phi_\infty = \phi_*$ almost everywhere with respect to μ_0. As μ_0 has full support and $\phi_\infty, \phi \in C(X)$, this means $\phi_\infty = \phi_*$. □

5. Existence and Uniqueness of Solutions

In this section we will treat questions of existence and uniqueness of solutions to (1.1) for different data (μ_0, β). First of all we will prove that, for any data $(\mu_0, \beta \neq 0)$, (1.1) admit a weak solution. We will then explain how to reduce the problem of regularity to the case considered in [3], where the authors use Caffarelli's interior regularity theory for Monge-Ampère equations. In the last part of the section we treat uniqueness. We first prove the claim made in Remark 1.2, namely that as long as $\beta > 0$ equation (1.1) admits at most one solution. Finally we prove Theorem 1.3 regarding $\beta \in [-1, 0)$ and $\mu_0 = \gamma$.

5.1. Existence of Weak Solutions. First of all, Lemma 3.14 implies $P(X)$ satisfies the following (relative) compactness property:

LEMMA 5.1. *Let $\{\phi_k\}$ be a sequence of functions in $P(X)$ such that $\inf_X \phi_k = 0$ for all k, then there is $\phi \in C(X)$ such that, after passing to a subsequence, $\phi_k \to \phi$ uniformly.*

PROOF. By lemma 3.14, $\{\phi_k\}$ are Lipschitz with a uniform Lipschitz constant. As X has finite diameter and $\inf_X \phi_k = 0$ for all k this means $\{\phi_k\}$ is also uniformly bounded, hence the lemma follows from the Arzelà-Ascoli theorem. □

LEMMA 5.2. *Let $\phi \in C(X)$ and*

$$F(\phi) = \xi(\phi) + \frac{1}{\beta} I_{\mu_0}(\beta\phi).$$

Then

$$F\left((\phi^c)^c\right) \leq F(\phi). \tag{5.1}$$

Moreover, if μ_0 has full support, then equality holds in (5.1) *if and only if $\phi \in P(X)$.*

PROOF. Recall that $\phi^c \in P(X)$, and hence $((\phi^c)^c)^c = \phi^c$ for all $\phi \in C(X)$. Also, $(\phi^c)^c \leq \phi$ for all $\phi \in C(X)$. This means $\xi(\phi) = \xi((\phi^c)^c)$ and

$$I_{\mu_0}((\phi^c)^c) = \frac{1}{\beta} \log \int_X e^{\beta(\phi^c)^c} d\mu_0 \leq \frac{1}{\beta} \log \int_X e^{\beta\phi} d\mu_0 = I_{\mu_0}(\phi). \tag{5.2}$$

and hence

$$F\left((\phi^c)^c\right) \leq F(\phi). \tag{5.3}$$

Assume μ_0 has full support. Then, if $\phi \notin P(X)$ and hence $(\phi^c)^c(x) < \phi(x)$ for some $x \in X$, then, as both $(\phi^c)^c$ and ϕ are continuous and μ_0 has full support, strict inequality holds in (5.2) and (5.3). This proves the lemma. □

LEMMA 5.3. *Let $\beta \in \mathbb{R} \setminus \{0\}$. Then F admits a minimizer. In other words,* (1.1) *admits a weak solution.*

PROOF. Recall that

$$F(\phi) = \xi(\phi) + \frac{1}{\beta} I(\beta\phi).$$

By the Dominated Convergence Theorem $\frac{1}{\beta} I(\beta\phi)$ is continuous in ϕ. By Lemma 3.9, ξ is continuous. This means F is continuous. Let ϕ_k be a sequence such that $F(\phi_k) \to \inf F$. By Lemma 5.2 we may assume

$\phi_k \in P(X)$ for all k. As F is invariant under the action of $\mathbb{R}$ given in (4.2) we may assume ϕ_k satisfies $\inf \phi_k = 0$ for all k. By Lemma 5.1, after possibly passing to a subsequence, $\phi_k \to \phi$ for some $\phi \in C(X)$. By continuity $F(\phi) = \lim_{k\to\infty} F(\phi_k) = \inf F$, hence ϕ is a minimizer of F. □

5.2. Regularity. In a numbers of papers (see [6], [7], [8]) Caffarelli developed a regularity theory for various types of weak solutions to Monge-Ampère equations. In particular, Caffarelli's theory applies to so called *Alexandrov solutions.* Recall that if f is a smooth function on $\mathbb{R}^n$, then a convex function Φ on $\mathbb{R}^n$ is an Alexandrov solution to the equation

$$\det(\Phi_{ij}) = f$$

if, for any borel measurable $E \subset \Omega$,

$$\int_E f dx = \int_{\partial\Phi(E)} dx$$

where $\partial\Phi(E)$ is the image of E under the multivalued gradient mapping, in other words

$$\partial\Phi(E) = \{y \in \mathbb{R}^n : \Phi(x) + \Phi^*(y) = \langle x, y\rangle \text{ for some } x \in E\}.$$

We have the following lemma:

Lemma 5.4. *Assume μ_0 is absolutely continuous with density f with respect to dx, $\beta \in \mathbb{R}$ and*

$$\mathrm{MA}(\phi) = e^{\beta\phi}\mu_0. \tag{5.4}$$

in the sense of Definition 2.2. Then $\Phi = \phi \circ \pi + x^2/2$ is an Alexandrov solution to the equation

$$\det(\Phi_{ij}) = e^{\beta(\Phi - x^2/2)} f \circ \pi \tag{5.5}$$

on $\mathbb{R}^n$. Moreover, Φ is proper.

Proof. Assume E is a Borel measurable subset of $\mathbb{R}^n$. To prove the first point in the lemma we need to prove

$$\int_E e^{\beta(\Phi - x^2/2)} f \circ \pi dx = \int_{\partial\Phi(E)} dx.$$

Let $C_0 = [0,1)^n \subset \mathbb{R}^n$ and $\{C_i\}$ be a collection of disjoint translates of C_0 such that $E \subset \cup C_i$. Let $E_i = E \cap C_i$. We have

$$\int_E e^{\beta(\Phi - x^2/2)} f \circ \pi dx = \sum_i \int_{E_i} e^{\beta(\Phi - x^2/2)} f \circ \pi dx = \sum_i \int_{\pi(E_i)} e^{\beta\phi} f dx$$

and by (5.4)

$$\sum_i \int_{\pi(E_i)} e^{\beta\phi} f dx = \sum_i \int_{(\nabla^c \phi^c)^{-1}(\pi(E_i))} dx.$$

Now, we claim that π maps $(\nabla\Phi^*)^{-1}(E_i)$ bijectively onto

$$(\nabla^c \phi^c)^{-1}(\pi(E_i))$$

for all i. To see this note that if $y \in (\nabla\Phi^*)^{-1}(E_i)$, then

$$\nabla^c \phi^c \circ \pi(y) = \pi \circ \nabla\Phi^*(y) \in \pi(E_i),$$

hence $\pi(y) \in (\nabla^c\phi^c)^{-1}(\pi(E_i))$. On the other hand, if $y \in (\nabla^c\phi^c)^{-1}(\pi(E_i))$, let $\tilde{x}$ be the unique lift of $\nabla^c\phi^c(y)$ in E_i. Moreover, let $\tilde{y}$ be a lift of y in $\mathbb{R}^n$. Since $\nabla^c\phi^c(y) = x$ we have $\nabla\Phi^*(\tilde{y}) = \tilde{x} + m_0$ for some $m_0 \in \mathbb{Z}^n$. We have that

$$\pi^{-1}(y) = \{\tilde{y} + m : m \in \mathbb{Z}^n\}$$

and by (2.4)

$$\nabla\Phi^*(\tilde{y} + m) = \nabla\Phi^*(\tilde{y}) + m = \tilde{x} + m_0 + m.$$

We conclude that $\nabla\Phi^*(\tilde{y} + m) \in E_i$ if and only if $m = -m_0$ and then $\nabla\Phi^*(\tilde{y} + m) = \tilde{x}$. This means π maps $(\nabla\Phi^*)^{-1}(E_i)$ bijectively onto $(\nabla^c\phi^c)^{-1}(\pi(E_i))$ as claimed. We get

$$\sum_i \int_{(\nabla^c\phi^c)^{-1}(\pi(E_i))} dx = \sum_i \int_{(\nabla\Phi^*)^{-1}(E_i)} dx = \int_{(\nabla\Phi^*)^{-1}(E)} dx$$

where the second inequality holds since the sets $(\nabla\Phi^*)^{-1}(E_i)$ are disjoint. Now, let $\operatorname{dom}\nabla\Phi^*$ be the set where $\nabla\Phi^*$ is defined. We have

$$\begin{aligned} \operatorname{dom}\nabla\Phi^* \cap \partial\Phi(E) &= \{y \in \mathbb{R}^n : \nabla\Phi^*(y) = x \text{ for some } x \in E\} \\ &= (\nabla\Phi^*)^{-1}(E). \end{aligned}$$

Since $\Omega \setminus \operatorname{dom}\nabla\Phi^*$ is a zero-set with respect to dx we have

$$\int_{(\nabla\Phi^*)^{-1}(E)} dx = \int_{\partial\Phi(E)} dx$$

which proves the first part of the lemma.

To see that Φ is proper, note that since ϕ is continuous it is bounded on X. Let $C = \inf_X \phi$. We get

$$\Phi(x) = \phi(\pi x) + \frac{x^2}{2} \geq C - 1 + |x|. \quad \square$$

LEMMA 5.5. *Assume μ_0 is absolutely continuous with smooth density with respect to dx and $\phi \in P(X)$ satisfies* (1.1) *in the sense of Definition 2.2. Then ϕ is smooth.*

PROOF. We will not refer directly to Caffarelli's papers. Instead we refer to [3] (more precisely, step three in the proof of Theorem 1.1) where the authors explain why, by Caffarelli's regularity theory, proper Alexandrov solutions on $\mathbb{R}^n$ to the equation

$$\det(\Phi_{ij}) = F(\Phi, x), \tag{5.6}$$

where F is smooth, are smooth. Strictly speaking the authors use an additional assumption of "finite energy", but the only way this is used is to guarantee properness of Φ. By Lemma 5.4, $\Phi = \phi \circ \pi + x^2/2$ is proper and satisfies (5.5) in the Alexandrov sense. As (5.5) is indeed a special case of (5.6) this proves the lemma. $\square$

5.3. Uniqueness. We first prove the claim made in Remark 1.2.

THEOREM 5.6. *Let $\mu_0 \in \mathcal{M}_1(X)$ be absolutely continuous with smooth density with respect to dx and $\beta > 0$. Then* (1.1) *admits a unique solution.*

PROOF. By Lemma 5.3 and Lemma 5.5 there always exist a solution to (1.1). To prove uniqueness it suffices to prove that the normalized equation (4.1) admits a unique solution modulo $\mathbb{R}$, in other words that F admits a unique minimizer modulo $\mathbb{R}$. Assume then ϕ_0 and ϕ_1 satisfies

$$F(\phi_0) = F(\phi_1) = \inf_{C(X)} F. \tag{5.7}$$

Let $\phi_t = t\phi_1 + (1-t)\phi_0$. Applying Lemma 3.15 with $A = X$ and $\Phi_\alpha(x) = \phi_x(\alpha)$ gives that

$$I_{\mu_0}(\phi_t) = \log \int_X e^{\phi_t} d\mu_0$$

is convex in t. Now, $\xi(\phi_t)$ is convex in t by Lemma 2.6. This means $F(\phi_t)$ is convex and hence, by (5.7), constant in t. It follows that $I_{\mu_0}(\phi_t)$

is affine in t. However, if we let $v = \frac{d}{dt}\phi_t = \phi_1 - \phi_0$, then

$$\begin{aligned}\frac{d^2}{dt^2} I_{\mu_0}(\phi_t) &= \frac{d}{dt}\left(\frac{\int_X v e^{\phi_t} d\mu_0}{\int_X e^{\phi_t} d\mu_0}\right) \\ &= \frac{\int_X v^2 e^{\phi_t} d\mu_0 \int_X e^{\phi_t} d\mu_0 - \left(\int_X v e^{\phi_t} d\mu_0\right)^2}{\left(\int_X e^{\phi_t} d\mu_0\right)^2} \end{aligned} \tag{5.8}$$

Further, if we let ν_t be the probability measure

$$\nu_t = \frac{e^{\phi_t} d\mu_0}{\int_X e^{\phi_t} d\mu_0}$$

and $\hat{v}$ be the constant

$$\hat{v} = \int_X v \nu_t$$

then

$$(5.8) = \int_X v^2 \nu_t - \hat{v}^2 = \int_X v^2 \nu_t - 2\hat{v}\int_X v\nu_t + \hat{v}^2 = \int_X (v - \hat{v})^2 \nu_t.$$

In particular, since $I_{\mu_0}(\phi_t)$ is affine in t we get that $v = \hat{v}$, hence that $\phi_1 - \phi_0$ is constant. This proves the theorem. □

We now turn to the proof of Theorem 1.3. We will use

Theorem 5.7 (The Prekopa Inequality [4], [14],[22]). *Let $\phi : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ be a convex function. Define*

$$\hat{\phi}(t) = -\log \int_{\mathbb{R}^n} e^{-\phi(t,x)} dx.$$

Then, for all $t \in \mathbb{R}$

$$\hat{\phi}(t) \leq t\hat{\phi}(1) + (1-t)\hat{\phi}(0)$$

with equality if and only if there is $v \in \mathbb{R}^n$ and $C \in \mathbb{R}$ such that

$$\phi(t,x) = \phi(0, x - tv) + tC.$$

Proof of Theorem 1.3. By Lemma 5.3 and Lemma 5.5 there always exist a solution to (1.3). Similarily as in the proof of Theorem 5.6, to prove uniqueness it suffices to prove that F admits a unique minimizer modulo $\mathbb{R}$. Assume ϕ_0 and ϕ_1 satisfies

$$F(\phi_0) = F(\phi_1) = \inf_{C(X)} F.$$

By Lemma 5.2 any minimizer of F is in $P(X)$, hence $(\phi_0^c)^c = \phi_0$ and $(\phi_1^c)^c = \phi_1$. This means the following equation defines a curve in $C(X)$ connecting ϕ_0 and ϕ_1:

$$\phi_t = (t(\phi_1)^c + (1-t)(\phi_0)^c)^c. \tag{5.9}$$

Note that, as $P(X)$ is convex and $\phi_0^c, \phi_1^c \in P(X)$ we get $t\phi_1^c + (1-t)\phi_0^c \in P(X)$ and

$$F(\phi_t) = \int_X t\phi_1^c + (1-t)\phi_0^c dx + \frac{1}{\beta}\log\int_X e^{\beta\phi_t}d\gamma. \tag{5.10}$$

The first term of this is affine in t. The second term is given by

$$\frac{1}{\beta}\log\int_X e^{\beta\phi_t}\sum_{m\in\mathbb{Z}^n} e^{|x-m|^2/2}dx = \frac{1}{\beta}\log\int_{\mathbb{R}^n} e^{\beta\phi_t\circ\pi - x^2/2}dx. \tag{5.11}$$

Let $\Phi_t = \phi_t \circ \pi + x^2/2$. By Lemma 2.1, since ϕ_t is the c-transform $t\phi_1^c + (1-t)\phi_0^c$, we have

$$\Phi_t(x) = \sup_{y\in\mathbb{R}^n} \langle x, y\rangle - (t\phi_1^c + (1-t)\phi_0^c)\circ\pi(y) - \frac{y^2}{2}. \tag{5.12}$$

As

$$\langle x, y\rangle - (t\phi_1^c + (1-t)\phi_0^c)\circ\pi(y) - \frac{y^2}{2}$$

is affine in (t, x) we get that (5.12) is convex in (t, x). It follows that, as long as $\beta \in [-1, 0)$, the exponent in (5.11),

$$\begin{aligned}\beta\phi_t\circ\pi(x) - x^2/2 &= \beta(\phi_t\circ\pi(x) + x^2/2) - (\beta+1)x^2/2 \\ &= \beta\Phi_t(x) - (\beta+1)x^2/2\end{aligned}$$

is concave in (t, x). We may then apply the Prekopa inequality to deduce that (5.11) and hence $F(\phi_t)$ is convex in t. In particular, as ϕ_0 and ϕ_1 are minimizers of F, this means $F(\phi_t) = F(\phi_0) = F(\phi_1)$ for all $t \in [0, 1]$. This imples (5.11) is affine in t. By the equality case in the Prekopa inequality

$$\beta\phi_1\circ\pi(x) - x^2/2 = \beta\phi_0\circ\pi(x-v) - (x-v)^2/2 + C$$

for some $C \in \mathbb{R}$ and $v \in \mathbb{R}^n$. By noting that $\phi_1\circ\pi$ and $\phi_0\circ\pi(\cdot - v)$, and hence

$$\beta\phi_1\circ\pi - \beta\phi_0\circ\pi(\cdot - v) = \langle\cdot, v\rangle + v^2/2 + C,$$

should descend to a function on X (in other words, they should be invariant under the action of $\mathbb{Z}^n$), we get that $v = 0$. This means $\phi_1 = \phi_0 + C$ which proves Theorem 1.3. $\square$

6. Geometric Motivation

The original motivation for this project comes from the paper on statistical mechanics and birational geometry by Berman [2]. Berman introduces a thermodynamic approach to produce solutions to the complex Monge-Ampère equation

$$\mathrm{MA}_{\mathbb{C}}(u) = e^{\beta u}\mu_0 \tag{6.1}$$

on a compact Kähler manifold M. The Monge-Ampère operator in (6.1) is defined as

$$(i\partial\bar{\partial}u + \omega_0)^n \tag{6.2}$$

where n is the complex dimension of M and ω_0 is a fixed Kähler-form on M representing the Chern class of a line bundle L over M. A solution, u, should be a real valued twice differentiable function on M satisfying $i\partial\bar{\partial}u + \omega_0 > 0$. As Berman's thermodynamic approach to this equation has served as an inspiration for us, we outline it here.

The metric, ω_0 determines, up to a constant, a metric on L. For each $k > 0$, let $N = N_k = H^0(M, L)$. By assumption on ω_0, L is ample and hence $N_k \to \infty$ as $k \to \infty$. Let $s_1, \dots s_N$ be a basis of $H^0(M, L)$. Locally we may identify this basis with a collection of functions $f_1, \dots f_N$. The map

$$(x_1, \dots, x_N) \mapsto \det(f_i(x_j))$$

determines a section, $\det(s_1, \dots, s_N)$, of the induced line bundle $L^{\boxtimes N_k}$ over M_N. The metric on L induces a metric, $\|\cdot\|$, on this line bundle and

$$\|\det(s_1, \dots, s_N)\|^{2\beta/k}\mu_0 \tag{6.3}$$

determines a symmetric measure on M^N. Note that changing the basis of $H^0(M, L)$ will give the same result up to a multiplicative constant. As long as this measure has finite volume we may normalize it to get a symmetric probability measure on M^N.

Now, Berman shows that if $\beta > 0$ and the singularities of $\mu_{\mathbb{C}}$ are controlled in a certain way, then the point processes defined by (6.3) converge to the Monge-Ampère measure of a solution to (6.1). However, it should be stressed that when $\beta < 0$ there is no guarantee that (6.3) has finite volume and can be normalized to a probability measure. This turns out to be a subtle property and in one of the most famous versions of equation (6.1), when M is a Fano manifold and ω_0, μ and β are chosen so that solutions to (6.1) define Kähler-Einstein metrics of positive curvature, this reduces to a property of the manifold M which is conjectured to be equivalent to the existence of Kähler-Einstein metrics on M (see [16] for some progress on this). We will explain in Section 6.1 how equation (1.3) can be seen as the "push forward" to a real setting of a complex Monge-Ampère equation whose solution define Kähler-Einstein metrics of almost everywhere positive curvature. In that sense, the present project can be seen as an attempt to study one side of this complex geometric problem.

6.1. Equation (1.3) as the "Push Forward" of a Complex Monge-Ampère Equation. Let $M = \mathbb{C}^n/(4\pi\mathbb{Z}^n + i\mathbb{Z}^n)$ and θ be the function on $\mathbb{C}^n$ defined as

$$\theta(z) = \sum_{m\in\mathbb{Z}^n} e^{-m^2/4+izm/2}.$$

This is the classical θ-function and it satisfies the following transformation properties:

$$\begin{aligned} \theta(z+4\pi) &= \theta(z) \\ \theta(z+i) &= \theta(z)e^{iz/2-1/4}. \end{aligned}$$

In particular, the zero set of θ defines the *theta divisor*, D, on M and, using certain trivializations of the line bundle associated to D, θ descends to a holomorphic section of this line bundle. This means $\tau = i\partial\bar{\partial}\log|\theta|^2$ is a well-defined (1,1)-current on M and we may consider the twisted Kähler-Einstein equation

$$\mathrm{Ric}(\omega) + \tau = \omega \tag{6.4}$$

on M, where $\mathrm{Ric}(\omega)$ denotes the Ricci curvature of ω. The current τ is supported on D so away from D this equation define metrics of constant positive Ricci curvature. Now, there is a standard procedure to rewrite (6.4) into a scalar equation of type (6.1). This process involves choosing a reference form ω_0 in the cohomology class of τ and fixing a Ricci-potential of ω_0, F, such that

$$i\partial\bar{\partial}F = \mathrm{Ric}(\omega_0) + \tau - \omega_0.$$

Choosing $\omega_0 = \sum_i idz_i \wedge d\bar{z}_i$ and $F = -y^2/2 + \log|\theta|^2$ gives the equation

$$\mathrm{MA}_{\mathbb{C}}(u) = e^{-u-y^2/2}|\theta|^2\omega_0^n. \tag{6.5}$$

In other words, we arrive at equation (6.1) with the choices

$$\mu_{\mathbb{C}} = |\theta^2|e^{-y^2/2}\omega_0^n$$

and $\beta = -1$. Now, let $z = x + iy$ be the standard coordinates on M induced from $\mathbb{C}^n$. Let $\rho : M \to X$ be the map $z \mapsto y$. If ϕ is a twice differentiable function on X such that $(\phi_{ij} + \delta_{ij})$ is strictly positive definite, then $u(z) := \phi(y)$ defines a (rotationally invariant) twice differentiable function on M satisfying $i\partial\bar{\partial}u + \omega_0 > 0$. Moreover,

$$\rho_* \mathrm{MA}_{\mathbb{C}}(u) = \mathrm{MA}(\phi) \tag{6.6}$$

where $\mathrm{MA}(u)$ is the complex Monge-Ampère measure on M defined in (6.2) and $\mathrm{MA}(\phi)$ is the real Monge-Ampère measure on X defined in (1.2). Further, at the end of the next sextion we will prove

Lemma 6.1.

$$\rho_*\left(e^{-y^2/2}|\theta|^2\omega_0^n\right) = \gamma \tag{6.7}$$

where dy is the uniform measure on X.

Since u is rotationally invariant we get that

$$\rho_*\left(\mathrm{MA}_{\mathbb{C}}(u) - e^{-u-y^2/2}|\theta|^2\omega_0^n\right) = \mathrm{MA}(\phi) - e^{-\phi}\gamma$$

and this is the relation that makes us refer to equation (1.3) as the "push forward" of equation (6.5).

6.2. Permanental Point Processes as the Push Forward of Determinantal Point Processes. Here we will establish a connection between the permanental point processes defined in Section 1.2 and the determinantal point processes defined in Bermans framework. The connection is a consequence of a certain formula that relates integrals of determinants to permanents. This formula might be of independent interest and is given in the following lemma.

Lemma 6.2. *Let (E, μ) be a measure space. Let $N \in \mathbb{N}$ and*

$$\{F_{jk} : j = 1 \dots N, k = 1 \dots N\}$$

be a collection of complex valued functions on E, square integrable with respect to μ, such that, for each j

$$\int_E F_{jk}\overline{F_{jl}}d\mu = 0$$

if $k \neq l$. Then

$$\operatorname{perm}\left(\int_E |F_{jk}|^2 d\mu\right) = \int_{E^N} |\det(F_{jk}(x_j))|^2 d\mu^{\otimes N}.$$

PROOF. Now,

$$\begin{aligned}
& \int_{E^N} |\det(F_{jk}(x_j))|^2 d\mu^{\otimes N} \\
= & \int_{E^N} \det(F_{jk}(x_j))\overline{\det(F_{jk}(x_j))} d\mu^{\otimes N} \\
= & \int_{E^N} \left(\sum_\sigma (-1)^\sigma \prod_j F_{j\sigma(k)}(x_j)\right) \overline{\left(\sum_{\sigma'} (-1)^{\sigma'} \prod_j F_{j\sigma'(k)}(x_j)\right)} d\mu^{\otimes N} \\
= & \sum_{\sigma,\sigma'} (-1)^{\sigma+\sigma'} \prod_j \int_E F_{j\sigma(k)}\overline{F_{j\sigma'(k)}} d\mu \qquad (6.8)
\end{aligned}$$

By the orthogonality assumption on $\{F_{jk}\}_k$, the only contribution comes from terms where $\sigma = \sigma'$. We get

$$(6.8) = \sum_\sigma \prod_j \int_E |F_{j\sigma(k)}|^2 d\mu = \operatorname{perm}\left(\int_E |F_{jk}|^2 d\mu\right). \quad \square$$

Before we examine its consequences for permanental point processes we illustrate two other applications. The first is given by the following formula related to Gram Determinants [Referens!]:

COROLLARY 6.3. *Let (E, μ) be a measure space and*

$$f_1, \ldots, f_N \in L^2(\mu).$$

Then

$$\det\left(\int_E f_j \overline{f_k} d\mu\right) = \frac{1}{N!} \int_{E^N} \left|\det\left(f_k(x_j)\right)\right|^2 d\mu^{\otimes N}. \qquad (6.9)$$

PROOF. Note that if A is an invertible $N \times N$ matrix with determinant 1, then replacing $\{f_1, \ldots, f_n\}$ by $\{\tilde{f}_1, \ldots, \tilde{f}_N\}$ where $\tilde{f}_i$ is defined by

$$(\tilde{f}_1, \ldots, \tilde{f}_N) = (f_1, \ldots, f_N)A$$

doesn't affect the formula (6.9). This means we may assume $f_1, \dots, f_N$ satisfy

$$\int_E f_j \overline{f_k} d\mu = 0$$

if $j \neq k$. For each $j, k \in \{1, \dots, N\}$, let $F_{jk} = f_k$. We get that

$$\det\left(\int_E f_j \overline{f_k} d\mu\right) = \prod_k \int_E |f_k|^2 d\mu = \frac{1}{N!} \operatorname{perm}\left(\int_E |F_{jk}|^2 d\mu\right)$$

and, applying Lemma 6.2, that

$$\begin{aligned} \det\left(\int_E f_j \overline{f_k} d\mu\right) &= \frac{1}{N!} \operatorname{perm}\left(\int_E |F_{jk}|^2 d\mu\right) = \frac{1}{N!} \int_{E^N} \left|\det\left(F_{jk}(x_j)\right)\right|^2 d\mu^{\otimes N} \\ &= \frac{1}{N!} \int_{E^N} \left|\det\left(f_k(x_j)\right)\right|^2 d\mu^{\otimes N} \end{aligned}$$

proving the corollary. □

The second application of Lemma 6.2 is given by the following formula for the permanent of a matrix of non-negative real numbers.

COROLLARY 6.4. *Let (a_{jk}) be an $N \times N$-matrix of non-negative real numbers. Then*

$$\operatorname{perm}(a_{jk}) = \frac{1}{(2\pi)^N} \int_{[0,2\pi]^N} \left|\det\left(\sqrt{a_{jk}} e^{ikx_j}\right)\right|^2 dx_1 \cdots dx_N.$$

PROOF. Let $F_{jk} = \sqrt{a_{jk}} e^{ikx}$. Then, for each j,

$$\int_{[0,2\pi]} F_{jk} \overline{F_{jl}} dx = \int_{[0,2\pi]} a_{jk} e^{i(k-l)x} dx = \begin{cases} 2\pi a_{jk} & \text{if } l = k \\ 0 & \text{otherwise.} \end{cases}$$

Applying Lemma 6.2 gives

$$\begin{aligned} \operatorname{perm}(a_{jk}) &= \frac{1}{(2\pi)^N} \operatorname{perm} \int_{[0,2\pi]} |F_{jk}|^2 dx \\ &= \frac{1}{(2\pi)^N} \int_{[0,2\pi]^N} \left|\det\left(F_{jk}(x_j)\right)\right|^2 dx_1 \dots dx_N \\ &= \frac{1}{(2\pi)^N} \int_{[0,2\pi]^N} \left|\det\left(\sqrt{a_{jk}} e^{ikx_j}\right)\right|^2 dx_1 \cdots dx_N. \end{aligned}$$

which proves the corollary. □

To see how Lemma 6.2 connects permanental point processes to determinantal point processes, we will now look a bit closer on the point processes defined by Bermans framework when applied to the complex Monge-Ampère equation in Section 6.1. First of all, $\omega_0 = \sum_i idz_i \wedge d\bar{z}_i$ represents the curvature class of the theta divisor D on M. Elements in $H^0(M, kD)$ may be represented by theta functions and a basis at level $k \in \mathbb{N}$ is given by the set

$$\{\theta_p^{(k)} : p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n\} \tag{6.10}$$

where

$$\theta_p^{(k)} = \sum_{m \in \mathbb{Z}^n + p} e^{-km^2/4 + izkm/2}.$$

With respect to these trivializations the norm of $\theta_p^{(k)}$ with respect to the metric on kD with curvature form $k\omega_0$ may be written

$$\|\theta_p^{(k)}\|^2 = |\theta_p^{(k)}|^2 e^{-ky^2/2}.$$

Enumeration the points in $\frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n$, $\{p_1, \ldots, p_N\}$ and using the standard coordinates $(z_1, \ldots, z_N) = (x_1 + iy_1, \ldots, x_N + iy_N)$ on M^N allow us to write the determinant in (6.3) as

$$\left| \det \left(\theta_{p_l}^{(k)}(z_j) e^{-y_j^2/4} \right)_{jl} \right|^2.$$

Now, recall that the real Monge-Ampère measure on X may be recovered as the push forward under the projection map, $\rho : M \to X$, of the complex Monge-Ampère measure on M (see equation (6.6)). Similarly, Lemma 6.2 will allow us to explicitly calculate the push forward of the measure

$$|\det(\theta_{p_i}^{(k)}(z_j)e^{-y_j^2/4})|^2 \omega_0^n$$

on M^N under the map $\rho^{\times N} : M^N \to X^N$. We get the following lemma, which is the key point of this section. It shows that the permanental point processes defined in Section 1.2 are the natural analog of the determinantal point processes defined by Bermans framework for complex Monge-Ampère equations.

Lemma 6.5. *Let dy be the uniform measure on X. Then*

$$\left(\rho^{\times N}\right)_* |\det(\theta_{p_l}^{(k)}(z_j)e^{-y_j^2/4})|^2 \omega_0^n = \operatorname{perm}\left(\Psi_{p_l}^{(N)}(y_j)\right) dy. \tag{6.11}$$

PROOF. Let $y = (y_1, \ldots, y_N) \in X^N$. The point $y \in X^N$ defines a real torus, T_y, in M^N

$$T_y = \left(\rho^{\times N}\right)^{-1}(y) = \left\{x + iy : x \in (\mathbb{R}^n/4\pi\mathbb{Z}^n)^N\right\}.$$

If we let dx be the measure on T_y induced by $(\mathbb{R}^n)^N$, then the density at y of the left hand side of (6.11) with respect to dy is given by the integral

$$\int_{T_y} |\det(\theta_{p_l}^{(k)}(z_j))e^{-y_j^2/4}|^2 dx. \tag{6.12}$$

For each $j, l \in \{1, \ldots, N\}$, let $F_{jl} : T_y \to \mathbb{C}$ be defined by

$$\begin{aligned} F_{jl}(x) &= \theta_{p_l}(x + iy_j)e^{-y_j^2/4} \\ &= \sum_{m\in\mathbb{Z}^n+p_l} e^{-km^2/4+i(x+iy_j)km/2-y_j^2/4} \\ &= \sum_{m\in\mathbb{Z}^n+p_l} e^{-k(m-y_j)^2/4+ikmx/2} \end{aligned}$$

Now, when computing the integral

$$\begin{aligned} &\int_{T_y} F_{jl}\overline{F_{jl'}}dx \\ =& \int_{T_y} \sum_{\substack{m\in\mathbb{Z}^n+p_l \\ m'\in\mathbb{Z}^n+p_{l'}}} e^{-k(m-y_j)^2/4-k(m'-y_j)^2/4+ik(m-m')x/2}dx \end{aligned} \tag{6.13}$$

the only contribution comes from the terms where $m - m' = 0$. If $l \neq l'$, then there are no such terms, in other words $(6.13) = 0$. If $l = l'$ we are left with

$$(6.13) = (4\pi)^N \sum_{m\in\mathbb{Z}+p_l} e^{-k|y_j-m|^2/2} = (4\pi)^N \Psi_{p_l}^{(N)}(y_j).$$

Applying Lemma 6.2 gives

$$(6.12) = \int_{T_y} |\det\left(F_{jl}(x_j)\right)|^2 dx = \operatorname{perm}\left(\int |F_{jl}|^2 dx\right) = \operatorname{perm}\left(\Psi_{p_l}(y_j)\right)$$

proving the lemma. $\square$

Finally, we show that Lemma 6.1 is a special case of this.

PROOF OF LEMMA 6.1. Note that $\theta = \theta_0^{(1)}$ and

$$\gamma = \sum_{m \in \mathbb{Z}^n} e^{-|y-m]^2/2} dy = \Psi_0^{(1)} dy.$$

This means (6.7) is the special case of (6.11) given by $N = k = 1$. Hence the lemma follows from Lemma 6.5. $\square$

6.3. Approximations of Optimal Transport Maps. As mentioned in the introduction the point processes defined here can be used to produce explicit approximations of optimal transport maps. In optimal transport it is natural to consider a larger class of Monge-Ampère operators. Let $\nu_0 \in \mathcal{M}_1(X)$ be absolutely continuous with respect to dx. Then ν_0 defines a Monge-Ampère operator MA_{ν_0} on $P(X)$ as

$$\mathrm{MA}_{\nu_0}(\phi) = (\nabla^c \phi^c)_* \nu_0.$$

Solutions, ϕ_*, to the inhomogenous Monge-Ampère equation

$$\mathrm{MA}_{\nu_0}(\phi) = \mu_0 \tag{6.14}$$

determine optimal transport maps on X in the sense that $T = \nabla^c \phi_*$ is the optimal *transport map* in the sense of Brenier (see [24]) from the source measure μ_0 to the target measure ν_0.

The fact that the point processes defined in Section 1.2 are related to the standard $\mathrm{MA} = \mathrm{MA}_{dx}$ is a consequence of the fact that

$$\frac{1}{N} \sum_{p \in \frac{1}{k}\mathbb{Z}^n/\mathbb{Z}^n} \delta_p \to dx$$

in the weak*-topology. Redefining $S^{(N)}$ in the following way will provide the generalisation we want: Let $P^{(N)}$ be a collection of point sets with the property that $|P^{(N)}| = N$ and

$$\frac{1}{N} \sum_{p \in P^{(N)}} \delta_p \to \nu_0.$$

As in the original definition, associate a wave function, $\Psi_p^{(N)}$, to each point $p \in \cup P^{(N)}$

$$\Psi_{p_i}^{(K)} = \sum_{m \in \mathbb{Z}^n + p_i} e^{-|x-m|^2}$$

and, for each N, enumerate the points in $P^{(N)}$

$$P^{(N)} = \{p_1, \ldots, p_N\}.$$

We get

Corollary 6.6. *Let $\mu_0, \nu_0 \in \mathcal{M}_1(X)$ be absolutely continuous and have smooth, strictly positive densities with respect to dx and $\Psi_{p_i}^{(N)}$ be defined as above. Then*

$$\phi_N := \frac{1}{N} \log \int_{X^{N-1}} \operatorname{perm}\left(\Psi_{p_i}^{(N)}(x_j)\right) d\mu^{\otimes N}$$

converges uniformly to the unique, smooth, strictly convex solution of (6.14)*. Consequently, the associated gradient maps $\nabla^c \phi_N$ converges uniformly to the unique optimal transport map transporting μ_0 to ν_0.*

Proof of Corollary 6.6. First of all, the fact that the optimal transport map is smooth follow from Caffarelli's regularity theory for Monge-Ampère equations. We will not go through the argument as it is similar as in Section 5.2. Uniqueness is a basic result from optimal transport (see for example Theorem 2.4.7 in [24]). Now, to see that the convergence holds, consider the functionals, $\{H^{(N)}\}$, on $\mathcal{M}_1(X)$ defined by

$$E^{(N)}(\mu) = \frac{1}{N} \int_{X^N} H^{(N)} d\mu^{\otimes N}.$$

Direct calculations give that they are continuous, convex, Gateaux differentiable and $dE^{(N)}|_{\mu_0} = \Phi^{(N)}$. We claim that

$$E^{(N)}(\mu) \to W^2(\mu, dx) \tag{6.15}$$

for all $\mu \in \mathcal{M}_1(X)$. To see this, note that by the proof of Theorem 3.11

$$\sup_{X^N} \left| \frac{1}{N} H^{(N)} - W^2(\cdot, dx) \circ \delta^{(N)} \right| \to 0$$

as $N \to \infty$. We get, since $\{\frac{1}{N} H^{(N)}\}$ are uniformly bounded,

$$\begin{aligned} E^{(N)}(\mu) &= \int_{X^N} W^2(\cdot, dx) \circ \delta^{(N)} d\mu^{\otimes N} + o(1) \\ &= \int_{\mathcal{M}_1(X)} W^2(\cdot, dx) \left(\delta^{(N)}\right)_* \mu^{\otimes N} + o(1). \end{aligned} \tag{6.16}$$

where $o(1) \to 0$ as $N \to \infty$. Now, it follows from Sanov's theorem that $(\delta^{(N)})_* \mu^{\otimes N} \to \delta_\mu$ in the weak*-topology on $\mathcal{M}_1(\mathcal{M}_1(X))$. Now, since X has finite diameter we get that the squared distance function on X can be bounded by a a constant times the distance function. As the Wasserstein 1-metric metricizes the weak* topology on $\mathcal{M}_1(X)$ this implies that $W^2(\cdot, dx)$ is continuous on $\mathcal{M}_1(X)$. We get that (6.16) converges to $W^2(\mu, dx)$ as $N \to \infty$.

Further, $W^2(\cdot, dx)$ is convex. By standard properties of convex functions $dE^{(N)}|_{\mu_0}$ converges to a subgradient of $W^2(\cdot, dx)$ at μ_0. By standard properties of the Legendre Transform this means

$$\phi = \lim_{N \to \infty} \phi^{(N)} \tag{6.17}$$

satisfies $d\xi|_\phi = MA(\phi) = \mu_0$. This means ϕ is smooth and $\nabla^c \phi$ defines the optimal transport map transporting μ_0 to ν_0. Now, let Φ_N and Φ be the images in $P_{\mathbb{Z}^n}(\mathbb{R}^n)$ of ϕ_N and ϕ respectively. The convergence in (6.17) implies $\Phi_N \to \Phi$ and, by standard properties of convex functions, $\nabla \Phi^{(N)} \to \nabla \Phi$. This means $\nabla^c \phi_N \to \nabla^c \phi$ which proves the Corollary. $\square$

References

[1] R.J. Berman *Statistical Mechanics of Permanents, Real Monge-Ampère Equations and Optimal Transport* Arxiv:1302.4045v1 [Math.AP] 2013.

[2] R.J. Berman *Kähler-Einstein metrics, Canonical Random Point Processes and Birational Geometry* arXiv:1307.3634v1 [Math.DG] 2013

[3] R.J. Berman, B. Berndtsson *Real Monge-Ampère Equations and Kähler-Ricci Solitons on Toric Log Fano Varieties* Arxiv:1207.6128v1 [Math.DG] 2012.

[4] C. Borell. *Convex set functions in d-space* Period. Math. Hungar., Vol. 6. p 111-136. 1975.

[5] Y. Brenier *Décomposition polaire et réarrangement monotone des champs de vecteurs.* C.R. Acad. Sci. Paris Sér. I Math., 305 (1987), 805.808.

[6] L.A Caffarelli. *Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation.* Ann. of Math. (2) 131 1990, no 1, 135-150.

[7] L.A Caffarelli. *A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity.* Ann. of Math. (2) 131 1990, no 1, 129-134.

[8] L.A. Caffarelli. *Some regularity properties of solutions of Monge-Ampère equations* Comm. Pure Appl. Math. 44 no 8-9, 965-959. 1991.

[9] L.A. Caffarelli. J.A. Viaclovsky. *On the Regularity of Solutions to Monge-Ampère Equations on Hessian Manifolds* Commun. in Partial Differential Equations, 26(11&12, 2339-2351 (2001).

[10] S.-Y. Cheng S.-T. Yau *The Real Monge-Ampère Equation and Affine Flat Structures* Proceedings of the 1980 Beijing Symposium on Differential Geometry and Differential Equations, Vol. 1. 339370, Science Press, Beijing, 1982.

[11] D. Cordero-Erasquin *Sur le transport de mesures périodiques* C.R. Acad. Sci. Paris Sr. I Math. 329 (1999), 199–202.
[12] D. Cordero-Erasquin, R.J. McCann, M. Michael Schmuckenschläger *Prekopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport*
[13] A. Dembo, O. Zeitouni. *Large Deviations Techniques and Applications* Springer. (2nd edition) 1998.
[14] S. Dubuc *Critères de convexit et ingalits intgrales* Annales de l'institut Fourier, tome 27, no 1. p 135-165. 1977.
[15] R.S. Ellis, K. Haven, B. Turkington *Large Deviation Principles and Complete Equivalence and Nonequivalence Results for Pure and Mixed Ensembles* J. Statist. Phys. 101 no 5-6. p 999-1064. 2000.
[16] K. Fujita. *On Berman-Gibbs stability of $\mathbb{Q}$-Fano varieties* arXiv:1501.00248 [math.AG] 2015.
[17] J.B. Hough, M. Krishnapur, Y.I: Peres, B. Virag *Determinantal Point Processes and Independence* Probab. Surv. 3. p 206-229. 2006.
[18] L.V. Kantorovich. *On the Translocation of Masses* C.R. (Dokl) Acad. Sci. URSS 37 (1942), 199-201.
[19] M. Knott, C.S. Smith. *On the Optimal Mapping of Distributions* Journal of Optimization Theory and Applications: VoL 43, No. 1, May 1984
[20] R.J. McCann. *Polar factorization of Maps on Riemannian Manifolds* Geom. Funct. Anal. 11, 3 (2001), 589-608.
[21] J.W. Negele, H. Orland *Quantum Many Particle Systems* Westveiw Press. 1998.
[22] A. Prekopa *On Logarithmic Concave Measures and functions* Acta Sci. Math., 34. p336-343. 1972.
[23] A-S Sznitman *Topics in Propagation of Chaos* École d'Été de Probabilités de Saint-Flour XIX-1989. 165-251. Lecture notes in Math. 1464. Springer. 1991.
[24] C. Villani *Topics in Optimal Transportation* American Mathematical Society. 2003.
[25] C. Villani *Optimal Transport; Old and New* Springer. 2009.