

# ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE Mathématiques

JAKOB HULTGREN

*Permanental Point Processes on Real Tori, Theta Functions and  
Monge–Ampère Equations*

Tome XXVIII, n° 1 (2019), p. 11-65.

[http://afst.cedram.org/item?id=AFST\\_2019\\_6\\_28\\_1\\_11\\_0](http://afst.cedram.org/item?id=AFST_2019_6_28_1_11_0)

© Université Paul Sabatier, Toulouse, 2019, tous droits réservés.

L'accès aux articles de la revue « Annales de la faculté des sciences de Toulouse Mathématiques » (<http://afst.cedram.org/>), implique l'accord avec les conditions générales d'utilisation (<http://afst.cedram.org/legal/>). Toute reproduction en tout ou partie de cet article sous quelque forme que ce soit pour tout usage autre que l'utilisation à fin strictement personnelle du copiste est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

cedram

Article mis en ligne dans le cadre du  
Centre de diffusion des revues académiques de mathématiques  
<http://www.cedram.org/>

# Permanental Point Processes on Real Tori, Theta Functions and Monge–Ampère Equations <sup>(\*)</sup>

JAKOB HULTGREN <sup>(1)</sup>

---

**ABSTRACT.** — Inspired by constructions in complex geometry we introduce a thermodynamic framework for Monge–Ampère equations on real tori. We show convergence in law of the associated point processes and explain connections to complex Monge–Ampère equations on abelian varieties and optimal transport.

**RÉSUMÉ.** — On introduit un cadre thermodynamique pour des équations Monge–Ampères sur des tores réelles, qui est inspiré par des constructions en géométrie complexe. On démontre la convergence en loi pour les processus aléatoires correspondants, expliquant les liens avec des équations de Monge–Ampères complexes sur des variétés abéliennes et le transport optimal.

---

## 1. Introduction

In a series of papers culminating in [4] Berman introduced a thermodynamic framework for complex Monge–Ampère equations. In particular, he showed how the Monge–Ampère measures of solutions to complex Monge–Ampère equations can be seen as limits of canonically defined (so called temperature-deformations of) determinantal point processes. Inspired by this we will introduce a thermodynamic framework for real Monge–Ampère equations on the real torus  $X = \mathbb{R}^n/\mathbb{Z}^n$ . Using certain families of functions analogous to theta functions on Abelian varieties we construct permanental point processes on  $X$ . Our first result (Theorem 1.2) is that, as long as the Monge–Ampère equation admits a unique solution, the point processes defined by the statistic mechanical framework converges in law towards the Monge–Ampère measure of this unique solution. Equivalently, and in the

---

(\*) Reçu le 19 mai 2016, accepté le 19 février 2017.

(1) Chalmers University of Technology, Chalmersplatsen 4, 412 96 Gothenburg (Sweden) — *Current address:* University of Oslo, P.O. Box 1072 Blindern, 0316 Oslo (Norway) — jakobhu@math.uio.no

Article proposé par Vincent Guedj.

language of thermodynamics, under absence of first order phase transitions the microscopic setting admits a macroscopic limit that is determined by the Monge–Ampère equation. (See [28] for a general reference on the relations between the theory of large deviations used here and statistical mechanics. See also [2] for relations between statistical mechanics and complex geometry).

The real torus should be seen as one of several settings where strong connections between complex geometry, real Monge–Ampère equations and optimal transport are manifested (the related case of toric manifolds is treated in [3]). We will exploit these connections to produce semi-explicit approximations of optimal transport maps on  $X$  (see Corollary 6.6). As such, this work ties in with the works by McCann [24] and Cordero–Erasquin [12] on optimal transport on Riemannian manifolds.

Moreover, motivated by the difficult problem of singular Kähler–Einstein metrics of (almost everywhere) positive curvature on complex varieties we propose a corresponding real Monge–Ampère equation on  $X$  (see equation (1.3) below). The assumption of no first order phase transition 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 of *negative* temperature. As a second result, by proving a uniqueness theorem for Monge–Ampère equations of independent interest (see Theorem 1.4), we rule out first order phase transitions down to the critical temperature of  $-1$ . In a future paper we hope to address the question of uniqueness for temperatures smaller than  $-1$ , which might be seen as the analog of the problem studied in [23] (in particular Theorem 1.2 in [23]). We stress that the point processes and real Monge–Ampère equations treated here are only *analogous* to the related complex geometric objects. In particular, Theorem 1.2 and Theorem 1.4 are independent of corresponding statements about convergence, existence and uniqueness in the complex geometric setting (see Section 6 for more on this).

## 1.1. Setup

Let  $dx$  be the standard volume measure on  $X$  induced from  $\mathbb{R}^n$ . Let  $\beta$  be a real constant and  $\mu_0$  a probability measure on  $X$ , absolutely continuous and with smooth, strictly positive density with respect to  $dx$ . Given the data  $(\mu_0, \beta)$  we will consider the real Monge–Ampère equation on  $X$  given by

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

Here MA is the Monge–Ampère operator defined by

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

where  $(\phi_{ij})$  is the Hessian of  $\phi$  with respect to the coordinates on  $X$  induced from  $\mathbb{R}^n$  and  $\delta_{ij}$  is the Kronecker delta. As usual we demand of a solution  $\phi : X \rightarrow \mathbb{R}$  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  $\mu_0$  is chosen as the measure

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

We get the equation

$$\text{MA}(\phi) = e^{\beta\phi}\gamma. \quad (1.3)$$

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 "temperature - deformations" of certain *permanental point processes* (see [19] 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  $N$  functions on  $X$

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

This defines a matrix valued function on  $X^N$

$$(x_1, \dots, x_N) \rightarrow (\Psi_i^{(N)}(x_j)).$$

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  $\mu_0$  this defines a symmetric probability measure on  $X^N$

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

where  $Z_N$  is a constant ensuring the total mass is one. This is a permanental point process.

Now, for each positive integer  $k$ , let  $N = k^n$  and  $\mathcal{P}_N$  be the set of points in  $X$  given by the image under the quotient map of the lattice  $\frac{1}{k}\mathbb{Z}^n \subset \mathbb{R}^n$ . Note that the number of points in  $\mathcal{P}_N$  is  $N$ . Then let

$$S^{(N)} = \{\Psi_p^{(N)} : p \in \mathcal{P}_N\}$$

where

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

We make the following definition:

DEFINITION 1.1. — Let  $k \in \mathbb{N}$ ,  $N = k^n$ ,  $\beta$  be a real number,  $\mu_0$  a probability measure on  $X$  and  $S^{(N)} = \{\Psi_1, \dots, \Psi_N\}$  be defined as above. We define  $\mu_\beta^{(N)}$  to be the symmetric probability measure on  $X^N$  given by

$$\mu_\beta^{(N)} = \left( \text{perm}(\Psi_i^{(N)}(x_j)) \right)^{\beta/k} d\mu_0^{\otimes N} / Z_{\beta, N} \quad (1.5)$$

where  $Z_{\beta, N}$  is a constant ensuring the total mass is one.

Note that  $\mu_\beta^{(N)}$  is not formally a permanent point process. It differs from (1.4) by an exponent  $\beta/k$  in the density. We will call point processes of this form *temperature-deformation* of permanent point processes. In the case when  $\beta = k$ , and hence (1.5) reduces to (1.4) we will use the term *true* permanent point process to emphasize that  $\mu_\beta^{(N)}$  is really a permanent point process.

Finally, we make two remarks on the definitions. In [25] permanent 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  $\Psi_1^{(N)}, \dots, \Psi_N^{(N)}$ . Secondly, we will explain in Section 6.2 how the functions in  $S^{(N)}$  arise as the "push forward" of  $\theta$ -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  $\delta^{(N)} : X^N \rightarrow \mathcal{M}_1(X)$

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

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

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

Our results concern the weak\* limit (the limit in the topology induced by weak convergence of measures) of  $\Gamma_\beta^{(N)}$  as  $N \rightarrow \infty$ . In particular we will show,

in some cases, that the limit is a Dirac measure concentrated at a certain  $\mu_\vee \in \mathcal{M}_1(X)$  related to (1.1) or (1.3). Loosely speaking, this means  $\mu_\vee$  can be approximated by sampling larger and larger point sets on  $X$  according to  $\mu_\beta^{(N)}$ .

**THEOREM 1.2.** — *Let  $\mu_0 \in \mathcal{M}_1(X)$  be absolutely continuous and have smooth, strictly positive density with respect to  $dx$ . Let  $\Gamma_\beta^{(N)}$  be defined as above and let  $\beta \in \mathbb{R}$ . Assume also that (1.1) admits a unique solution,  $\phi_\vee$ . Then*

$$\Gamma_\beta^{(N)} \rightarrow \delta_{\mu_\vee} \tag{1.7}$$

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

*Remark 1.3.* — 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  $\beta$  as well.

**THEOREM 1.4.** — *Assume  $\mu_0 = \gamma$  and  $\beta \in [-1, 0)$ . Then equation (1.3) admits a unique solution.*

*Remark 1.5.* — The complex Monge–Ampère equation analogous to Equation (1.3) was considered by Lin and Wang in [23]. Theorem 1.4 can be seen as the real analog of Theorem 1.2 in [23]. However, solutions to Equation (1.3) do not induce solutions to the corresponding complex Monge–Ampère equation (see Section 6.1). Hence Theorem 5.6 does not follow from the results in [23].

Note that if  $\beta \neq 0$  and  $\mu_\vee = MA(\phi_\vee)dx$  where  $\phi_\vee$  is a solution to (1.1), then  $\phi_\vee$  can be recovered from  $\mu_\vee$  as  $\phi_\vee = \frac{1}{\beta} \log \rho$  where  $\rho$  is the density of  $\mu_\vee$  with respect  $\mu_0$ . In fact we get the following corollary of Theorem 1.2.

**COROLLARY 1.6.** — *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,  $\phi_\vee$ . Let  $\phi_N : X \rightarrow \mathbb{R}$  be the function defined by*

$$\phi_N(x_1) = \frac{1}{\beta} \log \int_{X^{N-1}} \left( \text{perm}(\Psi_{p_i}^{(N)}(x_j)) \right)^{\beta/k} d\mu_0^{\otimes(N-1)}(x_2, \dots, x_n) / Z_{\beta, N}.$$

*Then  $\phi_N$  converges uniformly to  $\phi_\vee$ .*

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 1.6 does not 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 semi-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.

*Remark 1.7.* — It might be worthwhile to clarify the relationship between Theorem 1.2 in this paper and Theorem 1.5 in [4]. In Section 6.2 we explain that permanental point processes arise as push forwards of determinantal point processes. This is an indication that convergence in the real setting would be closely connected to convergence in the complex setting. However, the theorems above are independent of each other. In other words, Theorem 1.2 does not follow from Theorem 1.5 in [4] and Theorem 1.2 does not imply a special case of Theorem 1.5 in [4]. The reason for this is that the push forward relation only holds for true permanental and determinantal point processes. Theorem 1, as well as Theorem 1.5 in [4], concerns cases when the densities of the point processes are taken to the power of  $\beta/k$  (i.e. the point processes are temperature-deformed). As temperature-deforming and taking push forward are two operations that does not commute, the convergence in Theorem 1 does not imply, and does not follow from, the convergence in Theorem 1.5 in [4].

## 1.4. Outline

**Convergence in Theorem 1.2 and a Large Deviation Principle.** Theorem 1.2 will follow from a *large deviation principle* for the sequence  $\Gamma_\beta^{(N)}$  (see Theorem 3.2). This large deviation principle provides a quantitative description of the convergence in Theorem 1.2, recording the speed of convergence in a *rate function*  $G : \mathcal{M}_1(X) \rightarrow [0, \infty)$ , satisfying  $\inf G = 0$  and a *rate*  $\{r_N\} \subset \mathbb{R}$  such that  $r_N \rightarrow \infty$  as  $N \rightarrow \infty$ . 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  $\Gamma_\beta^{(N)}(U)$  behaves as

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

as  $N \rightarrow \infty$ . This means  $\Gamma_\beta^{(N)}$ , for large  $N$ , is concentrated where  $G$  is small. In particular, if  $G$  admits a unique minimizer,  $\mu_\nu$ , (where  $G = 0$ ) then it follows that (see Section 4.3)  $\Gamma_\beta^{(N)}$  converges in the weak\* topology to  $\delta_{\mu_\nu}$ .

**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 the Large Deviation Principle (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  $\beta_N$  such that  $\beta_N \rightarrow \infty$  and study the family  $\{\Gamma_{\beta_N}^{(N)}\}$ . 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  $\Gamma_{\beta_N}^{(N)}$  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 we show how the large deviation principle in Theorem 3.2 can be deduced from this. This is based on essentially well known arguments. However, for completeness we give a proof of this in Section 3.2. It turns out that the crucial point is the equicontinuity and uniform boundedness of the (normalized) energy functions

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

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

**Connection to the Monge–Ampère Equation.** The final ingredients in the proof of Theorem 1.2 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.

**Existence and Uniqueness of Solutions (Theorem 1.4).** 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 Caffarelli explained in Lemma 5.5. These type of existence results for Monge–Ampère equations on affine manifolds was originally proven by Caffarelli and Viaclovsky [10] on the one hand and



Cheng and Yau [11] on the other. However, we will provide an alternative proof based on the variational principle above. Uniqueness, which is one of the main new contributions in this paper is proved in Section 5.3. Here we look at the space of quasi-convex functions equipped with a geodesic structure different from the standard one. The geodesic structure we are using is natural from a convex geometric perspective and it coincides with the geodesic structure induced by Mabuchi geodesics in the space of Kähler-Metrics on Abelian varieties. It will then follow from the Prekopa inequality that the energy functional associated to (1.3) is strictly convex with respect to this geodesic structure, hence admits no more than one minimizer. This is an extension of an argument used in [5] to prove uniqueness of Kähler–Einstein metrics on toric Fano manifolds. Curiously, there does not seem to be any direct argument for this using the Prekopa theorem on Riemannian manifolds (see [13]). Instead, we need to lift the problem to the covering space  $\mathbb{R}^n$  and use that  $\gamma$  is the push forward of a measure on  $\mathbb{R}^n$  with strong log-concavity properties.

**Geometric Motivation.** In Section 6 we explain the connections to the point processes on compact Kähler manifolds introduced by Berman in [4]. 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 functions in  $S^{(N)}$  and permanent 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 used to get semi-explicit approximations of optimal transport maps on  $X$ .

We end this section with two comments. While some parts of Section 3 might be well known to readers with a probabilistic background and, likewise, some parts of Section 4 might be familiar to readers with a background in geometry or optimal transport we nevertheless want to provide a paper that is accessible to readers from all three of these fields. This should (at least partly) explain the length of the paper.

Finally, comparing this paper to [3] where a similar probabilistic framework was set up for the toric setting, we will include a paragraph explaining some of the differences between the two papers. The setup in [3] is based on the same idea (namely the push-forward formula in Lemma 6.5) and Theorem 3.2 and Theorem 3.6 has clear analogs in [3] (Theorem 1.1 and Proposition 5.3). However, the setup in terms of the functions in  $S^{(N)}$  as analogs of theta functions on Abelian varieties and the part fitting these into the theory of optimal transport had to be developed specifically for this paper. This should be contrasted with the toric setting where real analogs

of metrics on line bundles, sections and their connection to optimal transport is more well-known. Moreover, the process of reducing Theorem 3.2 to Theorem 3.6, i.e. Theorem 3.12 (compare Theorem 1.4 in [3]) is done in a more direct way in this paper (based on an argument which is actually outlined in Section 4.6 [3]). There are also important differences related to the case of negative  $\beta$ . From the perspective of complex geometry this is of high interest. One crucial point in the case of negative  $\beta$  is the finiteness of the partition functions

$$Z_{N,\beta} = \int_{X^N} \text{perm}(\Psi_i(x_j))^{\beta/k} \mu_0^{\otimes N}$$

which, in our setting, follows from compactness. Another, more difficult, obstacle in the negative setting for us is that in order for the convergence in Theorem 1.2 to hold we need to establish uniqueness of solutions to (1.1). In the toric setting this follows from known results. In the setting of this paper, this is a subtle question which is partly settled in Theorem 1.4 where we establish this uniqueness in a family of cases. However, we conjecture that uniqueness does not hold in general for negative  $\beta$  and  $\mu_0$  satisfying the assumptions of Theorem 1.2.

**Acknowledgements.** The author would like to thank Robert Berman and Yanir Rubinstein for many interesting discussions on the subject of the paper (see for example Remark 3.10). Moreover, the author is grateful to Bo Berndtsson who read and commented on an earlier version of the paper as well as both referees for their many comments and suggestions, greatly improving the exposition and accuracy of the paper.

## 2. Preliminaries: Optimal Transport on Real Tori

In this section we will recall some basic theory of optimal transport. The content of the section is well known. Early contributors to the theory are Cordera-Erasquin [12] who studied the theory of optimal transport on real tori and McCann [24] generalized it to 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 gives a variational approach to real Monge–Ampère equations and an explicit formula for the Legendre transform of the functional  $\mu \rightarrow W^2(\mu, dx)$ , 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  $\Gamma_\beta^{(N)}$  as  $N \rightarrow \infty$ .

## 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*  $c : Y \times Y \rightarrow [0, \infty)$ , 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*,  $\Pi(\mu, \nu)$ , consisting of measures  $\gamma \in \mathcal{M}_1(Y \times Y)$  such that the first and second marginals of  $\gamma$  equal  $\mu$  and  $\nu$  respectively. The optimal transport distance between  $\mu$  and  $\nu$  is the quantity

$$\inf_{\gamma \in \Pi(\mu, \nu)} C(\gamma). \quad (2.1)$$

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  $\pi : \mathbb{R}^n \rightarrow X$  is the quotient map, then

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

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

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

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

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

Let  $C(\mathbb{R}^n)$  be the space of continuous functions on  $\mathbb{R}^n$  and if  $\Phi \in C(\mathbb{R}^n)$  let  $\Phi^*$  denote the Legendre transform of  $\Phi$ . 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

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

Then

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

*Proof.* — Note that

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

This means

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

which proves the lemma.  $\square$

It follows that  $\phi \in C(X)$  satisfies  $(\phi^c)^c = \phi$  if and only if  $\Phi$  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  $P(X)$ . In other words

$$\begin{aligned} P(X) &= \{ \phi \in C(X) : (\phi^c)^c = \phi \} \\ &= \{ \phi \in C(X) : \Phi = \phi \circ \pi + \frac{|x|^2}{2} \text{ convex} \}. \end{aligned}$$

Since  $\Phi^*$  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  $\Phi$  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  $\Phi$  be the image of  $\phi$  in  $P_{\mathbb{Z}^n}(\mathbb{R}^n)$ . Then  $\Phi$  is differentiable at a point  $x \in \mathbb{R}^n$  if and only if  $\phi$  is differentiable at  $\pi(x)$ . 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  $\Phi$  is differentiable at  $x$  and  $\nabla\Phi(x) = y$  if and only if  $\Phi$  is differentiable at  $x+m$  and  $\nabla\Phi(x+m) = y+m$ . This means the map  $\nabla\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , where it is defined, factors through to a map  $X \rightarrow X$ . 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,  $\Phi$  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

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

We conclude that  $\phi$  is differentiable and  $\nabla^c\phi(\pi(x)) = \pi(y)$  if and only if  $\pi(y)$  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  $\phi$  is differentiable at  $\pi(x)$  and  $\phi^c$  is differentiable at  $\pi(y) = \nabla^c\phi(\pi(x))$ , then  $\nabla^c\phi^c(\pi(y)) = \pi(x)$ .

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,  $\text{MA}(\phi)$ , of  $\phi$  as*

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

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

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

*as weak solutions to*

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

The definition above is a priori different from the usual viscosity definition. The definition is natural from the perspective of optimal transport (see Theorem 2.4). 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) (which is equivalent to viscosity solutions). In fact, we will see in Lemma 5.4 that the map  $\phi \mapsto \Phi$  where  $\Phi$  is given by (2.3) gives a direct link between these two types of solutions. First of all, however, we include a lemma that shows that smooth weak solutions are classical solutions.

LEMMA 2.3. — *Assume  $\phi$  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  $\nabla^c \phi^c : X \rightarrow X$  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  $\Phi^*$  be the image of  $\phi^*$  in  $P_{\mathbb{Z}^n}(\mathbb{R}^n)$ . We get

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

for some  $m \in \mathbb{Z}^n$ . By (2.4) we get  $\nabla \Phi^*(\tilde{x}_1) = \nabla \Phi^*(\tilde{x}_2 + m)$ . But since  $\phi$ , and hence  $\Phi$ , is smooth  $\Phi^*$  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  $\pi$  maps  $\nabla \Phi^*([0, 1]^n)$  diffeomorphically to  $X$ . Further,

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

and the denominator of the right hand side of (2.8) is the Jacobian determinant of the map  $\nabla \Phi^* : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Note that the last equality in (2.8) follows from standard properties of convex functions on  $\mathbb{R}^n$ . Let  $h \in C(X)$ . Then

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

which proves the lemma.  $\square$

### 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  $\xi$  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** ([21],[22],[6]). — *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*

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

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

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

Moreover,

$$\text{MA}(\phi_\mu) = \mu. \quad (2.12)$$

*Remark 2.5.* — Equation 2.10 is called Kantorovich’ duality [21] and property (2.12) is the Knott-Smith criterion which, in the context of Monge’s problem of optimal transport in  $\mathbb{R}^n$ , was discovered independently by Knott and Smith in 1984 [22] and by Brenier in 1987 [6]. It was established for general Riemannian manifolds by McCann [24]. In this case (2.12) involves a curvature term which, since the real torus with the metric induced by  $\mathbb{R}^n$  is flat, vanishes.

*Proof of Theorem 2.4.* — The theorem is essentially given by Theorem 5.10 in [30]. 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

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

hence the assumptions in 5.10.i, 5.10.ii and 5.10.iii in [30] 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  $\gamma$  is concentrated on the set

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

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

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

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  $\text{dom } \nabla^c \phi_\mu^c \subset X$  be the set where  $\phi_\mu^c$  is differentiable. We have

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

As  $\gamma$  is concentrated on (2.13) we get that  $\gamma$  is concentrated on the set

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

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).  $\square$

## 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 Gâteaux differentiability of the functional  $\xi$ . 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

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

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

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

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



LEMMA 2.6. — *The functional  $\xi$  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  $\phi_t^c$  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.  $\square$

This allows 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  $\xi$ . Moreover, for any  $\mu \in \mathcal{M}_1(X)$  there is  $\phi$  such that*

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

*Finally,  $\xi$  is Gâteaux differentiable on  $C(X)$  and*

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

*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  $C \rightarrow \infty$  if  $\mu(X) < 1$  and  $C \rightarrow -\infty$  if  $\mu(X) > 1$  gives

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

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

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

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

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

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

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

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

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

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 weakly to some  $\mu \in \mathcal{M}_1(X)$ . By the lower semi-continuity of  $W^2(\cdot, dx)$  we get

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

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

$$(\nabla^c \phi^c)_* dx = \mu. \quad (2.18)$$

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

Now, to see that (2.18) holds, note that (2.16) implies  $W^2(\mu, dx) < \infty$  and hence  $\mu \in \mathcal{M}_1(X)$ . By Theorem 2.4 there is a function  $\phi_\mu \in P(X)$  such that  $\text{MA}(\phi_\mu) = \mu$  and

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

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

### 3. A Large Deviation Principle

This section is devoted to Theorem 3.2 which will be the key part in the proof of Theorem 1.2. 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  $\mu$  is absolutely continuous with respect to  $\mu_0$ , let  $\mu/\mu_0$  denote the density of  $\mu$  with respect to  $\mu_0$ . The relative entropy of  $\mu$  with respect to  $\mu_0$  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  $\Gamma_\beta^{(N)}$  is defined as in Section 1.2. Then*

$$\left\{ \Gamma_\beta^{(N)} \right\}$$

*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  $\mu_0$  (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  $\chi$  be a topological space,  $\{\Gamma_N\}$  a sequence of probability measures on  $\chi$ ,  $G$  a lower semi continuous function on  $\chi$  and  $r_N$  a sequence of numbers such that  $r_N \rightarrow \infty$ . Then  $\{\Gamma_N\}$  satisfies a large deviation principle with rate function  $G$  and rate  $r_N$  if, for all measurable  $E \subset \chi$ ,*

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

*where  $E^\circ$  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  $\chi$ , 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, and well known, criteria for when a large deviation principle exist.

LEMMA 3.4. — *Let  $\chi$  be a compact metric space,  $\{\Gamma_N\}$  a sequence of probability measures on  $\chi$ ,  $G$  a function on  $\chi$  and  $r_N$  a sequence of numbers such that  $r_N \rightarrow \infty$ . Let  $B_d(\mu)$  denote the open ball in  $\chi$  with center  $\mu$  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  $\mu \in \chi$*

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

*Proof.* — Let  $\mathcal{B}$  be the basis of the topology on  $\chi$  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  $\chi$  is compact by assumption) in [15],  $\{\Gamma_N\}$  satisfies a large deviation principle with rate function  $G$  and rate  $r_N$  if and only if

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

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

$$\lim_{d \rightarrow 0} \left( \limsup_{N \rightarrow \infty} -\frac{1}{r_N} \log \Gamma_N(B_\delta(\mu)) \right) \quad (3.1)$$

is increasing as  $d \rightarrow 0$ , that

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

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  $\limsup$  replaced by  $\liminf$ . This proves the lemma.  $\square$

Finally we recall the well known

THEOREM 3.5 (Sanov’s theorem, see for example 6.2.10 in [15]). — *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  $Ent_{\mu_0}$ .

### 3.1. The Zero Temperature Case and the Gärtner–Ellis 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 1.2 and Theorem 3.2 are both concerned with the behavior of these families. In this section we will consider the family  $\{\Gamma_k^{(N)}\}_{k \in \mathbb{N}}$  (in other words, we will let  $\beta = k$ ). 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  $\Gamma$  is a probability measure on a topological vector space  $\chi$ , then the moment generating function of  $\Gamma$  is the functional on the dual vector space  $\chi^*$  given by

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

where  $\langle \cdot, \cdot \rangle$  is the pairing of  $\chi$  and  $\chi^*$ . 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  $\chi$  is *exponentially tight* if for each  $\epsilon \in \mathbb{R}$  there is a compact  $K_\epsilon \subseteq \chi$  such that for all  $N$

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \Gamma_N(\chi \setminus K_\epsilon) \leq \epsilon. \quad (3.3)$$

In our case, when  $\chi$  is compact, this is automatically satisfied since choosing  $K_\epsilon = \chi$  for any  $\epsilon$  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 [15]). — *Let  $\chi$  be a locally convex topological vector space,  $\{\Gamma_N\}$  an exponentially tight sequence of probability measures on  $\chi$  and  $r_N$  a sequence such that  $r_N \rightarrow \infty$ . Let  $Z_{\Gamma_N}$  be the moment generating function of  $\Gamma_N$  and assume*

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

exist, is finite valued, lower semi continuous and Gâteaux differentiable. Then  $\Gamma_N$  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 strictly positive density with respect to  $dx$ . Assume  $\Gamma_\beta^{(N)}$  is defined as in Section 1.2. Then*

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

*Proof.* — Note that if  $\mu_N$  is a measure on  $X^N$  and  $F$  is a function on  $\mathcal{M}_1(X)$ , then, since  $\Gamma_k^{(N)} = (\delta^{(N)})_* \mu_k^{(N)}$ ,

$$\int_{\mathcal{M}_1(X)} F(\mu) \Gamma_k^{(N)} = \int_{X^N} F(\delta^{(N)}(x)) d\mu_k^{(N)}.$$

Moreover,

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

This means

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

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

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

Introducing the notation

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

we get

$$Z_{\Gamma_k^{(N)}}(kN\phi) = N! \prod_{p \in \mathcal{P}_N} \int_X e^{k(-c_p^{(N)} + \phi)} d\mu_0 / Z_{k,N}. \quad (3.5)$$

Now, we claim that

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

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

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

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

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

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  $p, x_* \in X$

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

as long as  $x \in B_d(x_*)$ . Further, as  $\mu_0$  has full support, is absolutely continuous and has smooth density with respect to  $dx$  there is a large constant  $C$  such that

$$C\mu_0(B_d(x_*)) \geq 1 \quad (3.9)$$

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

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

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

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

Using (3.8) and (3.9) gives

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

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

$$\begin{aligned} \frac{1}{kN} \log Z_{\Gamma_k^{(N)}}(kN\phi) &= \frac{1}{kN} \log N! \prod_{p \in \mathcal{P}_N} \int_X e^{k(-c_p^{(N)} + \phi)} d\mu_0 / Z_{k,N} \\ &= \frac{\log N!}{kN} + \frac{1}{N} \sum_{p \in \mathcal{P}_N} \frac{1}{k} \log \int_X e^{k(-c_p^{(N)} + \phi)} d\mu_0 - \log Z_{k,N} \end{aligned} \quad (3.12)$$

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  $k \rightarrow \infty$ . Moreover, using (3.7) we get, since  $\frac{1}{N} \sum_{p \in \mathcal{P}_n} \delta_p \rightarrow dx$  in the weak\* topology, that the second term converges to

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

Finally, applying this to the third term

$$Z_{k,N} = \int_{X^N} \text{perm}(\Psi_{p_i}(x_j)) \mu_0^{\otimes N}$$

gives that  $Z_{k,N}$  converges to  $\xi(0)$ . Since the zero-function is mapped to itself under  $c$ -transform we get that  $Z_{k,N}$  vanishes as  $N \rightarrow \infty$ . This proves the lemma.  $\square$

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



LEMMA 3.9. — *The functional  $\xi$  is continuous on  $C(X)$ .*

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

$$\sup_X |\phi_0^c - \phi_1^c| \leq \sup_X |\phi_1 - \phi_0|. \quad (3.13)$$

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  $\phi_0$  and  $\phi_1$  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  $\chi = \mathcal{M}_1(X)$  is compact, tightness of  $\Gamma^{(N)}$  holds automatically. By Lemma 3.8

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

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

*Remark 3.10.* — As pointed out by Yanir Rubinstein, there is an alternative proof of Theorem 3.6 bypassing the use of the Gärtner–Ellis theorem. We give a brief sketch of it here because it provides a good intuitive picture. By Lemma 3.4 it suffices to verify that

$$\lim_{\delta \rightarrow 0} \limsup_{N \rightarrow \infty} -\frac{1}{kN} \log \int_{B_\delta^{(N)}(\mu)} \text{perm}(\Psi_{p_i}(x_j)) d\mu_0^{\otimes N} = W^2(\mu, dx) \quad (3.14)$$

and similarly with  $\limsup$  is replaced by  $\liminf$ . Here,  $B_d(\mu)$  is the ball in  $\mathcal{M}_1(X)$  of (Wasserstein-1) radius  $d$  centered at  $\mu$  and  $B_d^{(N)}(\mu)$  is its pre-image in  $X^N$  under  $\delta^{(N)}$ . For large  $N$ , the logarithm of the permanent

in (3.14) can be approximated by its tropical counterpart

$$\sup_{\sigma} \sum_i \log \Psi_{p_i}^{(N)}(x_{\sigma(i)}).$$

It follows from (3.6) that this is close to  $kN$  times the Wasserstein-2 distance between  $\frac{1}{N} \sum_i \delta_{p_i}$  and  $\frac{1}{N} \sum_i \delta_{x_i}$ . By general theory, this is approximated by the right hand side of (3.14) as long as  $\frac{1}{N} \sum_i \delta_{x_i}$  is close to  $\mu$  in Wasserstein-1 distance. Approximating the integrand in (3.14) like this and applying Sanov’s theorem to get

$$\limsup -\frac{1}{kN} \log \int_{B_{\delta}^{(N)}(\mu)} d\mu_0^{\otimes N} = 0$$

proves Theorem 3.6.

### 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 (see Theorem 3.12 below). This result is closely related to the main result in [17]. Before we go deeper into this we explain how  $\{\mu_{\beta}^{(N)}\}$  can be seen as the Gibbs measures of certain thermodynamic systems. If we introduce the  $N$ -particle Hamiltonian

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

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

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

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  $\mu_0$ . In this interpretation  $\mu_{\beta}^{(N)}$  is the equilibrium state of the system when the temperature is assumed fixed at  $Temp = 1/\beta$ . In Theorem 3.6,  $Temp = 1/\beta = 1/k \rightarrow 0$ , hence Theorem 3.6 describes the zero-temperature limit. Theorem 3.2 will follow from Theorem 3.6 and a theorem on equi-continuous and uniformly bounded Hamiltonians. To state that theorem we need to define what it means for the family  $\{\frac{H^{(N)}}{N}\}$  to be equi-continuous. 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

$$d^{(N)}(x, y) = d^{(N)}(x_1, \dots, x_N, y_1, \dots, y_N) = \frac{1}{N} \inf_{\sigma} \sum_i d(x_i, y_{\sigma(i)}) \quad (3.15)$$

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

$$\left| \frac{1}{N} H^{(N)}(x) - \frac{1}{N} H^{(N)}(y) \right| \leq \epsilon \quad (3.16)$$

whenever  $d^{(N)}(x, y) \leq d$ . Before we move on to state Theorem 3.12 we prove the following well known lemma.

LEMMA 3.11. — *Let  $x = (x_1, \dots, x_N) \in X^N$  and  $y = (y_1, \dots, y_N) \in X^N$ . Then (3.15) 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

$$(3.15) = \inf_{\gamma} \int_{X \times X} d(x, y) \gamma \quad (3.17)$$

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

$$\left\{ \sum_{i,j} a_{ij} \delta_{(x_i, y_j)} : a_{ij} \geq 0, \sum_i a_{ij} = 1/N, \sum_j a_{ij} = 1/N \right\}, \quad (3.18)$$

in other words a polytope in  $\mathcal{M}_1(X \times X)$ . It follows that the infimum in (3.17) is attained on one or more of the vertices of (3.18). Moreover, any permutation,  $\sigma$ , 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 (3.18) occur as  $\gamma_{\sigma}$  for some permutation  $\sigma$ . This proves the lemma.  $\square$

Note that this lemma implies that if we equip  $\mathcal{M}_1(X)$  with the Wasserstein 1-metric, which metrizes the weak\* topology on  $\mathcal{M}_1(X)$ , then the distance function defined in (3.15) makes the embeddings

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

isometric embeddings.

**THEOREM 3.12.** — *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  $\beta_N$  is a sequence of numbers tending to infinity. Assume also that*

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

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

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

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

For completeness, we will include a proof of Theorem 3.12 here. The key point is the following result from [17].

**PROPOSITION 3.13** ([17] Theorem 3.2, page 1028). — *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 lower semi-continuous functional  $E$  on  $\mathcal{M}_1(X)$  satisfying*

$$\sup_{X^N} \left| \frac{H^{(N)}}{N} - E \circ \delta^{(N)} \right| \rightarrow 0 \quad (3.19)$$

*as  $N \rightarrow \infty$ . Then*

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

*satisfies a Large Deviation Principle with rate  $N$  and rate function*

$$\beta E + Ent_{\mu_0} + C_{E, \mu_0, \beta} \quad (3.20)$$

*where  $C_{E, \mu_0, \beta}$  is a constant making sure the infimum of (3.20) is 0.*

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

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

Using (3.19) we get

$$\begin{aligned}
 & \lim_{d \rightarrow 0} \liminf_{N \rightarrow \infty} -\frac{1}{N} (\delta^{(N)})_* e^{-\beta H^{(N)}} \mu_0^{\otimes N} (B_d(\mu)) / Z_{\beta, N} \\
 &= \lim_{d \rightarrow 0} \liminf_{N \rightarrow \infty} -\frac{1}{N} \log \int_{B_d^{(N)}(\mu)} e^{-\beta H^N(x)} d\mu_0^{\otimes N} / Z_{\beta, N} \\
 &= \lim_{d \rightarrow 0} \liminf_{N \rightarrow \infty} -\frac{1}{N} \log \int_{B_d^{(N)}(\mu)} e^{-\beta N(E \circ \delta^{(N)}(x) + o(1))} d\mu_0^{\otimes N} / Z_{\beta, N} \\
 &= \beta E(\mu) + \lim_{d \rightarrow 0} \liminf_{N \rightarrow \infty} \left( -\frac{1}{N} \log \int_{B_d^{(N)}(\mu)} d\mu_0^{\otimes N} + \frac{1}{N} \log Z_{\beta, N} \right). \quad (3.21)
 \end{aligned}$$

and similarly with  $\liminf$  replaced by  $\limsup$  (here  $o(1) \rightarrow 0$  uniformly in  $x$  as  $N \rightarrow \infty$ ). By Sanov's theorem  $(\delta^{(N)})_* \mu_0^{\otimes N}$  satisfies a large deviation principle with rate  $N$  and rate function  $Ent_{\mu_0}$ . Hence, by Lemma 3.4, the first term in the parenthesis of (3.21) converges to  $Ent_{\mu_0}(\mu)$ . To handle the second term in the parenthesis, note that

$$\begin{aligned}
 \liminf_{N \rightarrow \infty} \frac{1}{N} \log Z_{\beta, N} &= \liminf_{N \rightarrow \infty} \frac{1}{N} \log \int_{X^N} e^{-\beta H^N(x)} d\mu_0^{\otimes N} \\
 &\geq \liminf_{N \rightarrow \infty} \frac{1}{N} \log \int_{B_d^{(N)}(\mu)} e^{-\beta H^N(x)} d\mu_0^{\otimes N}
 \end{aligned}$$

for all  $d > 0$ . Letting  $d \rightarrow 0$ , the argument above gives

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log Z_{\beta, N} \geq -\beta E(\mu) - Ent_{\mu_0}(\mu). \quad (3.22)$$

for all  $\mu$ . On the other hand, fixing  $d > 0$  we may take a finite subset  $\{\mu_1, \dots, \mu_m\}$  in  $\mathcal{M}_1(X)$  such that  $\{B_d(\mu_i)\}$  covers  $\mathcal{M}_1(X)$ . Then

$$\begin{aligned}
 \frac{1}{N} \log Z_{\beta, N} &\leq \frac{1}{N} \log \sum_i \int_{B_d^{(N)}(\mu_i)} e^{-\beta H^N(x)} \mu_0^{\otimes N} \\
 &= \frac{1}{N} \log \sum_i e^{-N a_{d, i}(N)}
 \end{aligned}$$

where

$$a_{d, i}(N) = -\frac{1}{N} \log \int_{B_d^{(N)}(\mu_i)} e^{-\beta H^N(x)} \mu_0^{\otimes N}.$$

By the argument above

$$\liminf_{N \rightarrow \infty} a_{d, i}(N) \rightarrow \beta E(\mu_i) + Ent_{\mu_0}(\mu_i) + o_{d, i}$$

where  $o_{d,i}$  is a number (independent of  $N$ ) that vanishes as  $d \rightarrow 0$ . This implies

$$\begin{aligned}
 \liminf_{N \rightarrow \infty} \frac{1}{N} \log Z_{\beta,N} &= \liminf_{N \rightarrow \infty} \frac{1}{N} \log \sum_i e^{-Na_{d,i}(N)} \\
 &\leq \liminf_{N \rightarrow \infty} \frac{1}{N} \log m \sup_i e^{-Na_{d,i}(N)} \\
 &= \liminf_{N \rightarrow \infty} \left( \frac{\log m}{N} - \inf_i a_{d,i}(N) \right) \\
 &\leq - \inf_i (\beta E(\mu_i) + Ent_{\mu_0}(\mu_i) + o_{d,i}). \tag{3.23}
 \end{aligned}$$

By lower semi-continuity of  $\beta E + Ent_{\mu_0}$ , (3.22) and (3.23) we get

$$\lim_{d \rightarrow 0} \liminf_{N \rightarrow \infty} Z_{\beta_N} = - \inf_{\mathcal{M}_1(X)} \beta E + Ent_{\mu_0}$$

and similarly with  $\liminf$  replaced by  $\limsup$ . 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 3.13 always holds for some functional  $U$  on  $\mathcal{M}_1(X)$ .

LEMMA 3.14. — *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,*

$$\sup_{X^N} \left| \frac{H^{(N)}(x)}{N} - U \circ \delta^{(N)}(x) \right| \rightarrow 0 \tag{3.24}$$

as  $N \rightarrow \infty$ .

*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 (3.24). To define the extensions  $U^{(N)}$ , recall that a real valued function  $f$  on a metric space with distance function  $(\cdot, \cdot)$  is said to satisfy a modulus of continuity  $\omega : [0, \infty) \rightarrow [0, \infty)$  if

$$|f(x) - f(y)| \leq \omega(d(x, y))$$

for all  $x$  and  $y$  in the domain of  $f$ . Note that by assumption the functions  $\frac{H^{(N)}}{N}$  all satisfy the same modulus of continuity,  $\omega$ . We define  $U^{(N)} :$

$\mathcal{M}_1(X) \rightarrow \mathbb{R}$  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  $\omega$  to be sub-additive. It follows that the function  $\omega(d(\mu, \cdot))$  satisfies  $\omega$  as modulus of continuity. This means  $U^{(N)}$ , being an infimum of functions satisfying  $\omega$ , also satisfy  $\omega$ . In particular the family  $\{U^{(N)}\}$  is equi-continuous.  $\square$

We can now prove Theorem 3.12.

*Proof of Theorem 3.12.* — 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  $\mu$  with radius  $d$ . By the assumed Large Deviation Principle and Lemma 3.4, for any  $\mu \in \mathcal{M}_1(X)$ ,

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

On the other hand, by Lemma 3.14 there is a function  $U$  on  $\mathcal{M}_1(X)$  such that, after possibly passing to a subsequence, (3.24) holds. This means

$$\begin{aligned} E(\mu) &= \lim_{d \rightarrow 0} \liminf_{N \rightarrow \infty} -\frac{1}{N\beta_N} \log \int_{B_d^{(N)}(\mu)} e^{-N\beta_N(U \circ \delta^{(N)} + o(1))} d\mu_0^{\otimes N} \\ &= U(\mu) + \lim_{d \rightarrow 0} \liminf_{N \rightarrow \infty} -\frac{1}{N\beta_N} \log \int_{B_d^{(N)}(\mu)} d\mu_0^{\otimes N} \\ &= U(\mu). \end{aligned} \tag{3.25}$$

where the second term in (3.25) is zero by Sanov's theorem. This means  $E = U$  and since  $E$  is the rate function of a large deviation principle and thus lower semi-continuous, the theorem follows from Proposition 3.13.  $\square$

### 3.3. Proof of Theorem 3.2

To use Theorem 3.12 we need to verify that the family  $\{\frac{H^{(N)}}{N}\}$  is equi-continuous. We will use the following two lemmas

LEMMA 3.15. — *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,  $\Phi$ , on  $\mathbb{R}^n$  is  $\lambda$ -convex if  $\Phi - \lambda \frac{|x|^2}{2}$  is convex.

LEMMA 3.16. — Assume  $\Phi_\alpha$  is a family of functions on  $\mathbb{R}^n$  parametrized over some set  $A$ . Assume that for all  $\alpha \in A$ ,  $\Phi_\alpha$  is  $\lambda$ -convex. Let  $\sigma$  be a probability measure on  $A$ . Then

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

is  $\lambda$ -convex.

*Proof.* — Assume first  $\lambda = 0$ . By the convexity of  $\Phi_\alpha$  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,

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

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 3.17. — *The normalized energy functions*

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

is an equi-continuous family (in the sense of (3.16)).

*Proof.* — We claim that

$$c_p^{(N)} = \frac{1}{k} \log \sum_{m \in \mathbb{Z}^n + p} e^{-k|x-m|^2/2} \in P(X) \quad (3.26)$$

for all  $p \in X$  and  $k \in \mathbb{N}$ . To prove the claim it suffices to prove that (3.26) is  $\lambda$ -convex with  $\lambda = -1$ . This follows from Lemma 3.16 as  $-|x - m|^2/2$  is  $\lambda$ -convex with  $\lambda = -1$  for all  $m \in \mathbb{R}^n$ . Further, fixing all but one variable we get a function on  $X$  given by

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

By Lemma 3.16 this function is in  $P(X)$ . By Lemma 3.15 it satisfies the Lipschitz constant 1. This means, if  $x = (x_1, \dots, x_N)$  and  $y = (y_1, \dots, y_N)$  are points in  $X^N$ , that

$$\begin{aligned} &\left| \frac{1}{N} H^{(N)}(x_1, \dots, x_N) - \frac{1}{N} H^{(N)}(y_1, \dots, y_N) \right| \\ &\leq \frac{1}{N} \sum_i \left| H^{(N)}(x_1, \dots, x_{i-1}, y_i, \dots, y_N) - H^{(N)}(x_1, \dots, x_i, y_{i+1}, \dots, y_N) \right| \\ &\leq \sum_i d(x_i, y_i). \end{aligned} \quad (3.27)$$

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 (3.27) equals  $d^{(N)}(x, y)$ . This implies  $H^{(N)}/N$  is equi-continuous in the sense of (3.16).  $\square$

*Proof of Theorem 3.2.* — By Theorem 3.6 and Theorem 3.12 we only need to verify that the family  $\{H^{(N)}/N\}$  is uniformly bounded and equi-continuous. The latter was proved in Corollary 3.17. To see that  $\{H^{(N)}/N\}$  is uniformly bounded recall that in the proof of Theorem 3.6 we proved that  $-\frac{1}{k} \log \Psi_p^{(N)}(x) \rightarrow d(x, p)^2/2$  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  $N$ ,

$$c \leq \frac{1}{k} \log \Psi_p^{(N)}(x) \leq C \quad (3.28)$$

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 (3.28) 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.  $\square$

#### 4. The Rate Function and its relation 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 functional  $G$  are the Monge–Ampère measures of solutions to (1.1) (see Lemma 4.3). This will allow us to finish the proof of Theorem 1.2.

##### 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:

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

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

$$C \mapsto (\phi \mapsto \phi + C). \quad (4.2)$$

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 (the analog

of the Ding functional in complex geometry) whose stationary points correspond to weak solutions of (1.1). For given data  $(\mu_0, \beta)$  this energy functional has the form

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

where  $I_{\mu_0}$  is defined as

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

LEMMA 4.1. — *Let  $\beta \neq 0$ . The functional  $I_{\mu_0}$  is Gâteaux differentiable and*

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

*Consequently,  $F$  is Gâteaux differentiable and  $\phi$  is a stationary point of  $F$  if and only if  $\phi$  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,  $\xi$  is differentiable and  $d\xi|_\phi = -\text{MA}(\phi)$ . This means  $F$  is Gâteaux differentiable and

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

proving the last statements of the lemma. □

## 4.2. The Minimizers of the Gibbs Free Energy

We will use the following well know property of the relative entropy function (see for example Lemma 6.2.13 in [15]).

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

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

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

We can now prove Lemma 4.3.

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

$$\mu_\vee = \text{MA}(\phi_\vee) \tag{4.4}$$

is the unique minimizer of the rate function

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

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

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

This means that, under the assumptions of Lemma 4.3,  $\mu_\vee$  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  $\mu \in \mathcal{M}_1(X)$  and  $\phi \in C(X)$ , the two inequalities

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

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

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  $\phi_\vee$  be the minimizer of  $F$ . Applying (4.5) to the pair  $\mu$  and  $\phi_\vee$  and (4.6) to the pair  $\mu$  and  $\beta\phi_\vee$  we get

$$G(\mu) = \beta W^2(\mu, dx) + \text{Ent}_{\mu_0}(\mu) \tag{4.8}$$

$$\geq -\beta \int \phi_\vee d\mu - \beta \xi(\phi_\vee) + \int \beta \phi_\vee d\mu - I(\beta \phi_\vee) \tag{4.9}$$

$$= -\beta \left( \xi(\phi_\vee) + \frac{1}{\beta} I(\beta \phi_\vee) \right) = -\beta F(\phi_\vee) \tag{4.10}$$

with equality if and only if  $d\xi|_{\phi_\vee} = -MA(\phi_\vee) = -\mu$  and  $\mu = dI|_{\phi_\vee}$  which, since  $d\xi|_{\phi_\vee} + dI|_{\phi_\vee} = 0$ , is true if and only if  $\mu = \text{MA}(\phi_\vee)$ . For the case  $\beta < 0$ , let  $\mu \in \mathcal{M}_1(X)$ . By Corollary 2.7 we may take  $\phi$  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

$$\begin{aligned} G(\mu) &= \beta W^2(\mu, dx) + Ent_{\mu_0}(\mu) \\ &\geq -\beta \int \phi d\mu - \beta \xi(\phi) + \int \beta \phi d\mu - I(\beta \phi) \end{aligned} \quad (4.11)$$

$$= -\beta \left( \xi(\phi) + \frac{1}{\beta} I(\beta \phi) \right) = -\beta F(\phi) \geq -\beta F(\phi_\nu). \quad (4.12)$$

Moreover, equality in (4.12) holds if and only if  $\phi = \phi_\nu$ . But that means  $dI|_\phi = -d\xi|_\phi = \mu$ , hence we have equality in (4.11) as well. This implies  $G(\mu) \geq -\beta F(\phi_\nu)$  with equality if and only if  $\mu = \text{MA}(\phi_\nu)$ .  $\square$

*Remark 4.5.* — At least formally, Lemma 4.3 is a consequence of a general principle for differentiable, strictly convex functions. Namely that if  $A$  and  $B$  are differentiable, strictly convex functions,  $A^*$  and  $B^*$  denotes their respective Legendre transforms and  $x$  is a minimizer of  $A \pm B$ , then  $dA(x)$  is a stationary point of  $A^* \pm B^*(\mp \cdot)$ . This can be verified by differentiating and using  $dA \circ AF^* = dB \circ dB^* = id$ . Moreover, in the case of  $A + B$ , this principle expresses the fact that the point where the difference between a convex and a concave function is minimal defines the element in the dual space such that the difference between the corresponding tangents is maximal.

### 4.3. Proof of Theorem 1.2 and Corollary 1.6

*Proof of Theorem 1.2.* — Let  $\phi_\nu$  be the unique solution to (1.1). It follows that (4.1) admits a unique solution modulo  $\mathbb{R}$  and that  $\phi_\nu$  is a solution to (4.1). Now, we will use two results from the next section. 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  $\phi_\nu$  is a minimizer of  $F$ . Using Lemma 4.3 we get that  $G$  admits the unique minimizer  $\mu_\nu$  satisfying  $\mu_\nu = \text{MA}(\phi_\nu)$ .

We want to prove that  $\Gamma_\beta^{(N)} \rightarrow \delta_{\mu_\nu}$  in the weak\* topology on  $\mathcal{M}_1(\mathcal{M}_1(X))$ . By the Portmanteau Theorem (see for example Theorem D.10 in [15]) it suffices to verify that

$$\limsup_{N \rightarrow \infty} \Gamma_\beta^{(N)}(F) \leq \delta_{\mu_\nu}(F) \quad (4.13)$$

for all closed  $F \subset \mathcal{M}_1(X)$ . If  $\mu_\nu \in F$  then (4.13) holds trivially. Assume  $\mu_\nu \notin F$ . 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  $\mu_\vee \notin F$  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

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

As  $r_N \rightarrow \infty$  we get that  $\limsup \log \Gamma_\beta^{(N)}(F) = -\infty$  and  $\limsup \Gamma_\beta^{(N)}(F) = 0$ . This proves the theorem.  $\square$

*Proof of Corollary 1.6.* — Equation (1.7) implies the first marginals of  $\mu_\beta^{(N)}$ ,

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

converges to  $\mu_\vee$  in the weak\* topology of  $\mathcal{M}_1(X)$  (see Proposition 2.2 in [27]). Now,  $e^{\beta\phi_N}$  is the density with respect to  $\mu_0$  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 3.16,  $\phi^{(N)}$  is  $\lambda$ -convex with  $\lambda = -1$  and hence in  $P(X)$ . By Lemma 3.15 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

$$\phi_N \rightarrow \phi_\infty$$

uniformly. As

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

in the weak\* topology of  $\mathcal{M}_1(X)$  we get that  $\phi_\infty = \phi_\vee$  almost everywhere with respect to  $\mu_0$ . As  $\mu_0$  has full support and  $\phi_\infty, \phi_\vee \in C(X)$ , this means  $\phi_\infty = \phi_\vee$ .  $\square$

## 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  $(\mu_0, \beta)$ . 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 [5], 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 1.3, namely that as long as  $\beta > 0$  equation (1.1) admits at most one solution. Finally we prove Theorem 1.4 regarding  $\beta \in [-1, 0)$  and  $\mu_0 = \gamma$ .

### 5.1. Existence of Weak Solutions

First of all, Lemma 3.15 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,  $\phi_k \rightarrow \phi$  uniformly.*

*Proof.* — By lemma 3.15,  $\{\phi_k\}$  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.  $\square$

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

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

*Then*

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

*Moreover, if  $\mu_0$  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

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

Assume  $\mu_0$  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  $\phi$  are continuous and  $\mu_0$  has full support, strict inequality holds in (5.2) and (5.3). This proves the lemma.  $\square$

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.* — By the Dominated Convergence Theorem  $\frac{1}{\beta} I(\beta\phi)$  is continuous in  $\phi$ . By Lemma 3.9,  $\xi$  is continuous. This means  $F$  is continuous. Let  $\phi_k$  be a sequence such that  $F(\phi_k) \rightarrow \inf F$ . 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  $\phi_k$  satisfies  $\inf \phi_k = 0$  for all  $k$ . By Lemma 5.1, after possibly passing to a subsequence,  $\phi_k \rightarrow \phi$  for some  $\phi \in C(X)$ . By continuity  $F(\phi) = \lim_{k \rightarrow \infty} F(\phi_k) = \inf F$ , hence  $\phi$  is a minimizer of  $F$ .  $\square$

## 5.2. Regularity

In a numbers of papers (see [7], [8], [9]) 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  $\Phi$  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  $\mu_0$  is absolutely continuous with density  $f$  with respect to  $dx$ ,  $\beta \in \mathbb{R}$  and*

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

*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,  $\Phi$  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  $\pi$  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  $\pi$  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  $\text{dom } \nabla\Phi^*$  be the set where  $\nabla\Phi^*$  is defined. We have

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

Since  $\Omega \setminus \text{dom } \nabla\Phi^*$  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  $\Phi$  is proper, note that since  $\phi$  is continuous it is bounded on  $X$ . Let  $C = \inf_X \phi$ . We get

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

LEMMA 5.5. — *Assume  $\mu_0$  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  $\phi$  is smooth.*

*Proof.* — We refer to [5] (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

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

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  $\Phi$ . 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 1.3.

**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  $\phi_0$  and  $\phi_1$  satisfies

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

Let  $\phi_t = t\phi_1 + (1-t)\phi_0$ . Applying Lemma 3.16 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

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

Further, if we let  $\nu_t$  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.  $\square$

We now turn to the proof of Theorem 1.4. We will use

**THEOREM 5.7** (The Prekopa Inequality [26, 16]). — *Let  $\phi : [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}$  be a convex function. Define*

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

*Then  $\hat{\phi}$  is convex. Moreover, if  $\hat{\phi}$  is affine, then there is  $v \in \mathbb{R}^n$  and  $C \in \mathbb{R}$  such that*

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

The first statement above was proved by Prekopa [26]. The second statement, concerning the affine case, was done by Dubuc [16]. Since [16] deals with a more general class of inequalities we will, as a courtesy to the reader, explain how to deduce the second statement of Theorem 5.7 from the statements in [16]. We will use the notation  $\phi_t(x) = \phi(t, x)$ . First of all, note that we may restrict ourselves to the case when

$$\hat{\phi}(t) = 0 \tag{5.9}$$

for all  $t \in [0, 1]$  since adding a term of the form  $A + Ct$  to  $\phi$  will reduce the general case to this. Moreover, it suffices to prove that

$$\phi_1(x + v) = \phi_{1/2}(x + \frac{v}{2}) = \phi_0(x) \tag{5.10}$$

for all  $x \in \mathbb{R}^n$ . To see this, note that if (5.10) holds then by convexity of  $\phi$

$$\phi_t(x + tv) \leq \max\{\phi_0(x), \phi_1(x + v)\} = \phi_0(x)$$

for any  $t \in [0, 1]$ . On the other hand, if  $t > 1/2$  then

$$\phi_t(x + tv) \geq \phi_0(x) + 2t(\phi_{1/2}(x + \frac{v}{2}) - \phi_0(x)) = \phi_0(x)$$

and if  $t < 1/2$  then

$$\phi_t(x + tv) \geq \phi_1(x + v) + 2(t-1)(\phi_1(x + v) - \phi_{1/2}(x + \frac{v}{2})) = \phi_1(x + v) = \phi_0(x).$$

To get (5.10), let  $f = e^{-\phi_0}$ ,  $g = e^{-\phi_1}$  and  $h(\cdot) = e^{-\phi_{1/2}(\cdot/2)}$  and apply Theorem 12 in [16] with  $p(u, v) = \sqrt{uv}$ . Then  $p_n(u, v)$  reduces to  $2^n \sqrt{uv}$  and, since (by (5.9))

$$\int_{\mathbb{R}^n} h(x) dx = 2^n = 2^n \left( \int_{\mathbb{R}^n} f(x) dx \right)^{1/2} \left( \int_{\mathbb{R}^n} g(x) dx \right)^{1/2}$$

we get, for some vector  $v \in \mathbb{R}^n$  and some real number  $m > 0$ , that

$$\begin{aligned} m^n g(mx + v) &= f(x) \\ (m + 1)^n h((m + 1)x + v) &= 2^n f(x). \end{aligned}$$

Using the convexity of  $\phi$  we get

$$\begin{aligned} (m + 1)^{-n} 2^n f(x) &= (m + 1)^{-n} 2^n f\left(\frac{x + mx}{1 + m}\right) = h(x + mx + v) \\ &= e^{-\phi_{1/2}\left(\frac{x + mx + v}{2}\right)} \geq e^{-(\phi_0(x) - \phi_1(mx + v))/2} \\ &= \sqrt{f(x)} \sqrt{g(mx + v)} = m^{-n/2} f(x). \end{aligned}$$

This implies  $(m + 1)^{-1} 2 \geq m^{-1/2}$  and hence  $m = 1$ . We get

$$\phi_1(x + v) = -\log g(x + v) = -\log f(x) = \phi_0(x)$$

and

$$\phi_{1/2}\left(x + \frac{v}{2}\right) = -\log h(2x + v) = -\log f(x) = \phi_0(x)$$

hence (5.10) holds.

We are now ready for

*Proof of Theorem 1.4.* — By Lemma 5.3 and Lemma 5.5 there always exist a solution to (1.3). Similarly 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  $\phi_0$  and  $\phi_1$  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  $\phi_0$  and  $\phi_1$ :

$$\phi_t = (t(\phi_1)^c + (1 - t)(\phi_0)^c)^c. \quad (5.11)$$

Even though it is not needed in the proof we make the following remark about (5.11). If we identify convex functions on  $X$  with  $(S^1)^n$  invariant Kähler metrics on the Abelian variety  $\mathbb{C}^n / (\mathbb{Z}^n + i\mathbb{Z}^n)$  (see Section 6.1), then (5.11) is the Mabuchi geodesic connecting  $\phi_0$  and  $\phi_1$  in the space of Kähler metrics

on  $\mathbb{C}^n/(\mathbb{Z}^n + i\mathbb{Z}^n)$ . We know proceed with the proof. 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

$$F(\phi_t) = \int_X t\phi_1^c + (1-t)\phi_0^c dx + \frac{1}{\beta} \log \int_X e^{\beta\phi_t} d\gamma. \quad (5.12)$$

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

$$\frac{1}{\beta} \log \int_X e^{\beta\phi_t} \sum_{m \in \mathbb{Z}^n} e^{-|x-m|^2/2} dx = \frac{1}{\beta} \log \int_{\mathbb{R}^n} e^{\beta\phi_t \circ \pi - |x|^2/2} dx. \quad (5.13)$$

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

$$\Phi_t(x) = \sup_{y \in \mathbb{R}^n} \left( \langle x, y \rangle - (t\phi_1^c + (1-t)\phi_0^c) \circ \pi(y) - \frac{|y|^2}{2} \right). \quad (5.14)$$

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 (5.14) is convex in  $(t, x)$ . It follows that, as long as  $\beta \in [-1, 0)$ , the exponent in (5.13),

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

is concave in  $(t, x)$ . We may then apply the Prekopa inequality to deduce that (5.13) and hence  $F(\phi_t)$  is convex in  $t$ . In particular, as  $\phi_0$  and  $\phi_1$  are minimizers of  $F$ , this means  $F(\phi_t) = F(\phi_0) = F(\phi_1)$  for all  $t \in [0, 1]$ . This implies (5.13) is affine in  $t$ . By the affine 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 1.4.  $\square$

## 6. Geometric Motivation

The original motivation for this project comes from the paper on statistical mechanics and birational geometry by Berman [4]. Berman introduces a thermodynamic approach to produce solutions to the complex Monge–Ampère equation

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

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

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

where  $n$  is the complex dimension of  $M$  and  $\omega_0$  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,  $\omega_0$  determines, up to a constant, a metric on  $L$ . For each  $k > 0$ , let  $N = N_k = H^0(M, L)$ . By assumption on  $\omega_0$ ,  $L$  is ample and hence  $N_k \rightarrow \infty$  as  $k \rightarrow \infty$ . 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

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

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  $\mu_C$  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  $\omega_0$ ,  $\mu$  and  $\beta$  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 [18] 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  $\theta$  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  $\theta$ -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  $\theta$  defines the *theta divisor*,  $D$ , on  $M$  and, using certain trivializations of the line bundle associated to  $D$ ,  $\theta$  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, for constants  $\alpha$  and  $\eta$ , the twisted Kähler–Einstein equation

$$\text{Ric}(\omega) - \alpha\tau = \eta\omega \tag{6.4}$$

on  $M$ , where  $\text{Ric}(\omega)$  denotes the Ricci curvature of  $\omega$ . This equation describes subtle problems in complex geometry. When  $\alpha \in (0, 2\pi)$  it is related to so called Kähler–Einstein edge metrics of curvature  $\eta$  and an edge singularity of angle  $2\pi - \alpha$  along  $D$  (see [20]). However, we will be interested in the case  $\alpha = -1$  and  $\eta = 1$ , which is treated for  $n = 1$  in [1] and [23]. Then (6.4) define metrics of constant positive Ricci curvature away from  $D$ . There is a standard procedure to rewrite (6.4) into a scalar equation of type (6.1). This process involves choosing a reference form  $\omega_0$  in the cohomology class of  $\tau$  and fixing a Ricci-potential of  $\omega_0$ ,  $F$ , such that

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

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

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

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  $\rho : M \rightarrow X$  be the map  $z \mapsto y$ . If  $\phi$  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,

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

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

LEMMA 6.1. — *Let  $\rho$ ,  $\theta$  and  $\omega_0$  be defined as above. Then*

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

Since  $u$  is rotationally invariant we get that

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

and this is the relation that makes us refer to equation (1.3) as the "push forward" of equation (6.5) and it is because of this we write about the two equations as analogous. However, like we stressed in the introduction, statements about existence and uniqueness of solutions to (1.3) are independent of corresponding statements about solutions to (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 Berman’s 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  $(E, \mu)$  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  $\mu$ , such that, for each  $j$*

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

*if  $k \neq l$ . Then*

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



*Proof.* — Now,

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

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

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

Before we examine its consequences for permanent point processes we illustrate two other applications. The first is given by a quick proof of the following well known formula related to Gram Determinants (see for example [14]):

COROLLARY 6.3. — *Let  $(E, \mu)$  be a measure space and*

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

*Then*

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

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

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

does not affect the formula (6.9). This means we may, by choosing  $A$  so that the matrix

$$\left( \int_E \tilde{f}_j \overline{\tilde{f}_k} d\mu \right) = \overline{A^T} \left( \int_E f_j \overline{f_k} d\mu \right) A$$

is diagonal, 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

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

and, applying Lemma 6.2, that

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

proving the corollary.  $\square$

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*

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

*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

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

which proves the corollary.  $\square$

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 Berman’s 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

$$\{\theta_p^{(k)} : p \in \mathcal{P}_N\} \tag{6.10}$$

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

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

Enumeration the points in  $\mathcal{P}_N$ ,  $\{p_1, \dots, p_N\}$  and using the standard coordinates  $(z_1, \dots, z_N) = (x_1 + iy_1, \dots, x_N + iy_N)$  on  $M^N$  allow us to write the determinant in (6.3) as

$$\left| \det \left( \theta_{p_i}^{(k)}(z_j) e^{-y_j^2/4} \right)_{j,l} \right|^2.$$

Now, recall that the real Monge–Ampère measure on  $X$  may be recovered as the push forward under the projection map,  $\rho : M \rightarrow X$ , 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  $\rho^{\times N} : M^N \rightarrow X^N$ . 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 Berman’s framework for complex Monge–Ampère equations.

LEMMA 6.5. — *Let  $dy$  be the uniform measure on  $X$ . Then*

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

*Proof.* — Let  $y = (y_1, \dots, y_N) \in X^N$ . The point  $y \in X^N$  defines a real torus,  $T_y$ , in  $M^N$

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

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

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

For each  $j, l \in \{1, \dots, N\}$ , let  $F_{jl} : T_y \rightarrow \mathbb{C}$  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

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

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

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

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 semi-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  $\nu_0$  defines a Monge–Ampère operator  $\text{MA}_{\nu_0}$  on  $P(X)$  as

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

Solutions,  $\phi_\nu$ , to the inhomogeneous Monge–Ampère equation

$$\text{MA}_{\nu_0}(\phi) = \mu_0 \quad (6.14)$$

determine optimal transport maps on  $X$  in the sense that  $T = \nabla^c \phi_\nu$  is the optimal *transport map* in the sense of Brenier (see [29]) from the source measure  $\mu_0$  to the target measure  $\nu_0$ .

The fact that the point processes defined in Section 1.2 are related to the standard  $\text{MA} = \text{MA}_{dx}$  is a consequence of the fact that

$$\frac{1}{N} \sum_{p \in \mathcal{P}_N} \delta_p \rightarrow dx$$

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

$$\frac{1}{N} \sum_{p \in P^{(N)}} \delta_p \rightarrow \nu_0.$$

This is possible since the set of atomic measures is dense in  $\mathcal{M}_1(X)$  equipped with the weak\* topology. As in the original definition, associate a 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  $P^{(N)}$

$$P^{(N)} = \{p_1, \dots, p_N\}.$$

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*

$$\phi_N := \frac{1}{N} \log \int_{X^{N-1}} \text{perm} \left( \Psi_{p_i}^{(N)}(x_j) \right) d\mu^{\otimes N}$$

*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  $\mu_0$  to  $\nu_0$ .*

*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 [29]). 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, Gâteaux differentiable and  $dE^{(N)}|_{\mu_0} = \Phi^{(N)}$ . We claim that

$$E^{(N)}(\mu) \rightarrow W^2(\mu, dx) \tag{6.15}$$

for all  $\mu \in \mathcal{M}_1(X)$ . To see this, note that by the proof of Theorem 3.12

$$\sup_{X^N} \left| \frac{1}{N} H^{(N)} - W^2(\cdot, dx) \circ \delta^{(N)} \right| \rightarrow 0$$

as  $N \rightarrow \infty$ . We get, since  $\{\frac{1}{N} H^{(N)}\}$  are uniformly bounded,

$$\begin{aligned} E^{(N)}(\mu) &= \int_{X^N} W^2(\cdot, dx) \circ \delta^{(N)} d\mu^{\otimes N} + o(1) \\ &= \int_{\mathcal{M}_1(X)} W^2(\cdot, dx) \left( \delta^{(N)} \right)_* \mu^{\otimes N} + o(1). \end{aligned} \quad (6.16)$$

where  $o(1) \rightarrow 0$  as  $N \rightarrow \infty$ . Now, it follows from Sanov’s theorem that  $(\delta^{(N)})_* \mu^{\otimes N} \rightarrow \delta_\mu$  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 constant times the distance function. As the Wasserstein 1-metric metrizes 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  $N \rightarrow \infty$ .

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  $\mu_0$ . By standard properties of the Legendre Transform this means

$$\phi = \lim_{N \rightarrow \infty} \phi^{(N)} \quad (6.17)$$

satisfies  $d\xi|_\phi = MA(\phi) = \mu_0$ . This means  $\phi$  is smooth and  $\nabla^c \phi$  defines the optimal transport map transporting  $\mu_0$  to  $\nu_0$ . Now, let  $\Phi_N$  and  $\Phi$  be the images in  $P_{\mathbb{Z}^n}(\mathbb{R}^n)$  of  $\phi_N$  and  $\phi$  respectively. The convergence in (6.17) implies  $\Phi_N \rightarrow \Phi$  and, by standard properties of convex functions,  $\nabla \Phi^{(N)} \rightarrow \nabla \Phi$ . This means  $\nabla^c \phi_N \rightarrow \nabla^c \phi$  which proves the Corollary.  $\square$

## Bibliography

- [1] R. J. BERMAN, “Analytic torsion, vortices and positive Ricci curvature”, <https://arxiv.org/abs/1006.2988v1>, 2010.
- [2] ———, “Kähler-Einstein metrics emerging from free fermions and statistical mechanics”, *J. High Energy Phys.* **10** (2011), no. 106, p. 31.
- [3] ———, “Statistical Mechanics of Permanents, Real Monge–Ampère Equations and Optimal Transport”, <https://arxiv.org/abs/1302.4045v1>, 2013.
- [4] ———, “Kähler-Einstein metrics, canonical random point processes and birational geometry”, in *Algebraic geometry: Salt Lake City 2015*, Proceedings of Symposia in Pure Mathematics, vol. 97, American Mathematical Society, 2018, p. 29-73.
- [5] R. J. BERMAN & B. BERNDTSSON, “Real Monge-Ampère equations and Kähler-Ricci solitons on toric log Fano varieties”, *Ann. Fac. Sci. Toulouse Math. (6)* **22** (2013), no. 4, p. 649-711.

- [6] Y. BRENIER, “Décomposition polaire et réarrangement monotone des champs de vecteurs”, *C. R. Acad. Sci. Paris Sér. I Math.* **305** (1987), no. 19, p. 805-808.
- [7] L. A. CAFFARELLI, “Interior  $W^{2,p}$  estimates for solutions of the Monge-Ampère equation”, *Ann. Math.* **131** (1990), no. 1, p. 135-150.
- [8] ———, “A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity”, *Ann. Math.* **131** (1990), no. 1, p. 129-134.
- [9] ———, “Some regularity properties of solutions of Monge Ampère equation”, *Commun. Pure Appl. Math.* **44** (1991), no. 8-9, p. 965-969.
- [10] L. A. CAFFARELLI & J. A. VIACLOVSKY, “On the regularity of solutions to Monge-Ampère equations on Hessian manifolds”, *Commun. Partial Differ. Equations* **26** (2001), no. 11-12, p. 2339-2351.
- [11] S. Y. CHENG & S. T. YAU, “The real Monge-Ampère equation and affine flat structures”, in *Proceedings of the 1980 Beijing Symposium on Differential Geometry and Differential Equations, Vol. 1, 2, 3 (Beijing, 1980)*, Sci. Press Beijing, Beijing, 1982, p. 339-370.
- [12] D. CORDERO-ERAUSQUIN, “Sur le transport de mesures périodiques”, *C. R. Acad. Sci. Paris Sér. I Math.* **329** (1999), no. 3, p. 199-202.
- [13] D. CORDERO-ERAUSQUIN, R. J. MCCANN & M. SCHMUCKENSCHLÄGER, “Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport”, *Ann. Fac. Sci. Toulouse Math. (6)* **15** (2006), no. 4, p. 613-635.
- [14] P. A. DEIFT, *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, Courant Lecture Notes in Mathematics, vol. 3, New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.
- [15] A. DEMBO & O. ZEITOUNI, *Large deviations techniques and applications*, second ed., Applications of Mathematics (New York), vol. 38, Springer, 1998.
- [16] S. DUBUC, “Critères de convexité et inégalités intégrales”, *Ann. Inst. Fourier* **27** (1977), no. 1, p. 135-165.
- [17] 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** (2000), no. 5-6, p. 999-1064.
- [18] K. FUJITA, “On Berman-Gibbs stability and  $K$ -stability of  $\mathbb{Q}$ -Fano varieties”, *Compos. Math.* **152** (2016), no. 2, p. 288-298.
- [19] J. B. HOUGH, M. KRISHNAPUR, Y. PERES & B. VIRÁG, “Determinantal processes and independence”, *Probab. Surv.* **3** (2006), p. 206-229.
- [20] T. JEFFRES, R. MAZZEO & Y. A. RUBINSTEIN, “Kähler-Einstein metrics with edge singularities”, *Ann. Math.* **183** (2016), no. 1, p. 95-176.
- [21] L. V. KANTOROVITCH, “On the translocation of masses”, *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **37** (1942), p. 199-201.
- [22] M. KNOTT & C. S. SMITH, “On the optimal mapping of distributions”, *J. Optim. Theory Appl.* **43** (1984), no. 1, p. 39-49.
- [23] C. S. LIN & C. L. WANG, “Elliptic functions, Green functions and the mean field equations on tori”, *Ann. Math.* **172** (2010), no. 2, p. 911-954.
- [24] R. J. MCCANN, “Polar factorization of maps on Riemannian manifolds”, *Geom. Funct. Anal.* **11** (2001), no. 3, p. 589-608.
- [25] J. W. NEGELE & H. ORLAND, *Quantum many-particle systems*, Frontiers in Physics, vol. 68, Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1988.
- [26] A. PRÉKOPA, “On logarithmic concave measures and functions”, *Acta Sci. Math.* **34** (1973), p. 335-343.

- [27] A. S. SZNITMAN, “Topics in propagation of chaos”, in *École d’Été de Probabilités de Saint-Flour XIX—1989*, Lecture Notes in Mathematics, vol. 1464, Springer, 1991, p. 165-251.
- [28] H. TOUCHETTE, “The large deviation approach to statistical mechanics”, *Phys. Rep.* **478** (2009), no. 1-3, p. 1-69.
- [29] C. VILLANI, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, 2003.
- [30] ———, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer, 2009, Old and new.