Commun. Math. Phys. 214, 547 – 563 (2000)

Communications in
**Mathematical
Physics**

Monotonicity Properties of Optimal Transportation and the FKG and Related Inequalities*

Luis A. Caffarelli

Department of Mathematics, The University of Texas at Austin, Austin, TX 78712-1082, USA.
E-mail: caffarel@math.utexas.edu

Received: 18 October 1999 / Accepted: 24 March 2000

Abstract: Optimal transportation between densities $f(X)$, $g(Y)$ can be interpreted as a joint probability distribution with marginally $f(X)$, and $g(Y)$. We prove monotonicity and concavity properties of optimal transportation $(Y(X))$ under suitable assumptions on f and g. As an application we obtain the Fortuin, Kasteleyn, Ginibre correlation inequalities as well as some generalizations of the Brascamp–Lieb momentum inequalities.

0. Introduction

We start this introduction by giving some background on optimal transportation and the FKG inequalities.

0.1. The problem of optimal transportation. We are given two probability densities $f(X)$, $g(Y)$, and we want to transport the (variable X with) density f onto the (variable Y with) density g in a way that minimizes transportation costs, say for simplicity, $C(Y-X)$. Let us first say what we mean by transporting f to g.

(Pre) Definition. *A smooth map $Y(X)$ transports f to g if*

$$g(Y(X)) \det D_X Y = f(X).$$

That is, a small differential of volume

$$g(Y)\,dy$$

is pulled back to

$$f(X)\,dx$$

by the map $Y(X)$.

* Research was supported in part by the National Science Foundation, DMS-9714758.

A weak formulation is the following:

Definition 1. *A (weak) transport is a measurable map $Y(X)$, such that for any C_0 function $h(Y)$ the following ("change of variable") formula is valid:*

$$\int h(Y)g(Y)\,dY = \int h(Y(X))f(X)\,dX.$$

Now, given the cost function $C(X)$, we define

Optimal transportation. The (weak) transportation $Y(X)$ is optimal if it minimizes

$$J(Y) = \int C(Y(X) - X)f(X)\,dx$$

among all weak transportation.

Existence and regularity of such an optimal transportation has been studied in detail. (See for instance [B,C2,C3] and [G-M].) We will discuss (and use) in this paper the particular case where

$$C(X - Y) = \tfrac{1}{2}|X - Y|^2.$$

The correlation inequalities part of the paper holds true for more general cost functions, still convex and with the appropriate symmetries, but the proofs are technically involved and we will present it elsewhere.

The second derivative estimates for the Monge-Ampere like equations corresponding to non-quadratic cost functions, is a completely open matter. In the quadratic case, there is a rather complete existence and regularity theory ([B,C2,C3]). We will be interested in the following results.

Theorem 1 (Existence and stability, [B])). *Let Ω_1, Ω_2 be two open domains in $\mathbb{R}^n$, $f(X)$, $g(Y)$ two strictly positive bounded, measurable functions in $\overline{\Omega_i}$, with*

$$\int_{\Omega_1} f(X)\,dX = \int_{\Omega_2} g(Y)\,dY = 1.$$

Then,

a) *There exists a unique optimal transportation map $Y(X)$.*
b) *The optimal transportation $Y(X)$ (and its inverse $X(Y)$) are obtained from the following minimization process:*
   b_1) *Among all pairs of continuous functions $\varphi(X)$, $\psi(Y)$ satisfying the constraint*

$$\varphi(X) + \psi(Y) \geq \langle X, Y\rangle$$

minimize

$$J(\varphi, \psi) = \int_{\Omega_1} \varphi(X)f(X)\,dX + \int_{\Omega_2} \psi(Y)g(Y)\,dY.$$

   b_2) *φ and ψ are unique and convex and $Y(X)$ is defined as the (possibly multiple valued) map $Y \in Y(X)$ if*

$$\varphi(X) + \psi(Y) = \langle Y, X\rangle.$$

Theorem 2 (Regularity, [C2,C3]). *Hypothesis as before, assume further that* Ω_1, Ω_2 *are convex. Then*

a) *If* $0 < \lambda \le f, g \le \Lambda$, *the map* $Y(X)$ *and its inverse* $X(Y)$ *are single valued, of class* C^α *in* Ω_i *for some* α.
b) *If* f, g *are Hölder continuous, with exponent* β *for some* β *then* $Y(X)$, $X(Y)$ *are of class* $C^{1,\beta}$.
c) *In both cases, (a) and b)), there exists a pair of convex potentials* $\varphi(X)$, $\psi(Y)$ *such that*

$$Y(X) = \nabla\varphi(X), X(Y) = \nabla\psi(Y).$$

d) φ *satisfies the Monge–Ampére equation*

$$\det D^2\varphi(X) = \frac{f(X)}{g(\nabla\varphi(X))}$$

in case a) in the Alexandrov weak sense, in case b) in the classical sense.

(Note that $\varphi \in C^{2,\beta}$*.) By approximation, we will develop all our discussion for* f, g *of class* C^α*, so we will always talk of "classical" solutions.*

From the variational construction of Y, we also have a stability theorem.

Theorem 3 (Stability). *Let* f_j, g_j *be uniformly bounded, measurable and supported in a bounded domain* B_R. *Assume that* $f_j \to f$ *in* L^1, $g_j \to g$ *in* L^1. *Then* $\varphi_j \to \varphi$, $\psi_j \to \psi$ *uniformly in* B_R. *In particular if* φ_j, ψ_j *are uniformly* $C^{1,\alpha}$, *then* $\nabla\varphi_j$, $\nabla\psi_j$ *also converge uniformly to* $\nabla\varphi$, $\nabla\psi$.

We complete the discussion with the following interpretation (see [B]).

If we think of $f(X)$, $g(Y)$ as probability densities, we may think of the map $Y(X)$ as a joint probability distribution: $\nu_0(X, Y)$ in $\Omega_1 \times \Omega_2$, sitting on the graph $X, Y(X)$ with the property that the marginals $\mu_1(X)$, $\mu_2(Y)$ of ν_0 are exactly $f(X)\,dx$ and $g(Y)\,dy$.

In fact ν_0 has the following minimizing property:

Theorem ([B]). *Among all probability measures* $\nu(X, Y)$ *with marginals* $f(X)\,dX$ *and* $g(Y)\,dY$, $Y(X)$ *minimizes*

$$E(\nu) = \int |X - Y|^2\,d\nu(X, Y).$$

0.2. The FKG inequalities. The FKG inequalities (see [FKG,H,P]) play a fundamental role in statistical mechanics.

In this paper, we are interested in a theorem of Holley [H] from which the inequalities follow. Holley's Theorem establishes a monotonicity condition for probability measures μ_1, μ_2 defined on a finite lattice, Λ.

Let us discuss briefly his two main theorems. We consider a finite lattice Λ (that we will think of as embedded in the set P of vertices of the unit cube of $\mathbb{R}^N$ for some N (i.e., the set of all N-tuples, $X = (x_1, \ldots, x_N)$ with $x_i = 0$ or 1. On Λ, we have two non-vanishing probability measures $\mu_1(X)$, $\mu_2(X)$ with the "monotonicity property":

Given X, Y in Λ,

$$\mu_2(X \vee Y)\mu_1(X \wedge Y) \ge \mu_2(X)\mu_1(Y).$$

(As usual $\vee$ denotes taking max in each entry, $\wedge$ min.) Then

Theorem 4 ([H]**).** *There exists a joint measure*

$$\nu(X, Y)$$

with marginals $\mu_1(X)$, $\mu_2(Y)$ *such that*

$$\nu(X, Y) \neq 0 \implies X \leq Y.$$

As a corollary, he obtains

Corollary 1. *If* h *is an increasing function of* X*, then*

$$\int_\Lambda h(X)\, d\mu_1(X) \leq \int h(X)\, d\mu_2(X)$$

(that is μ_2 *is "concentrated more to the right" than* μ_1*).*

The purpose of this paper is to study the relation between optimal transportation and the FKG inequalities, in particular to show:

a) In the continuous case, the optimal transportation from the unit cube of $\mathbb{R}^n$ into itself ($\mu_1 = f(X)$, $\mu_2 = g(Y)$) has the proper monotonicity properties ($Y(X) \geq X$) of Holley's joint probability density provided that f, g do).
b) If we "spread" the measures μ_i from the vertices of the unit cube to half cubes, the densities f, g so obtained satisfy these properties, recuperating from this approach Holley's theorem, for the lattice formed by all vertices of the cube.
c) For a general sublattice, one can extend the "spread" measure to all of the half cubes recuperating in full the theorem of Holley.
d) In fact the discrete optimal transportation satisfies $Y(X) \geq X$.

Our proof is based on the fact that first derivatives of solutions of the Monge–Ampére equation satisfy an equation themselves. But it is also known that second derivatives are subsolutions of an elliptic equation.

In the last section we explore what the implications of that fact are in terms of correlation inequalities.

In closing this introduction we want to stress that in the continuous case the optimal transport map $Y(X)$ interpreted as a joint probability measure

$$\nu(X, Y) = \delta_{X,Y(X)}(X, Y) f(X)\, dX = \delta_{X,Y(X)}(X, Y) g(Y)\, dY$$

is not just a joint distribution but a "change of variables", i.e., a one to one map that carries one density to the other, and it is further the gradient of a convex potential, giving the map (or the measure $\nu(X, Y)$) a lot of stability.

1. Optimal Transportation from the Unit Cube to the Unit Cube and Periodic Monge–Ampére

We start this section with a reflection property of optimal transportation maps. Given $X \in \mathbb{R}^n$ we denote by $\bar{X}$ its reflection with respect to x_1, i.e., if $X = (x_1, x_2, \ldots, x_n)$ then $\bar{X} = (-x_1, x_2, \ldots, x_n)$.

Lemma 1. *Assume that*

a) Ω_1, Ω_2 *are symmetric with respect to* x_1*, i.e.,* $X \in \Omega_i \Leftrightarrow \bar{X} \in \Omega_i$*,*

b) *f, g are also symmetric, i.e.,*

$$f(X) = f(\bar{X}), \quad g(X) = g(\bar{X}).$$

Then the optimal transportation is also symmetric, i.e.,

a) $\varphi(X) = \varphi(\bar{X}), \quad \psi(Y) = \psi(\bar{Y}),$
b) $Y(\bar{X}) = \bar{Y}(X).$

Proof. By Brenier [B] φ, ψ are the unique minimizing pair of

$$\int \varphi(X) f(X)\, dX + \int \psi(Y) f(Y)\, dY$$

under the constraint

$$\varphi(X) + \psi(Y) \geq \langle X, Y \rangle.$$

By uniqueness, then,

$$\varphi(X) = \varphi(\bar{X}), \quad \psi(Y) = \psi(\bar{Y})$$

since $\varphi(\bar{X}), \psi(\bar{Y})$ are a competing pair with the same energy. □

Remark. The lemma is valid for a general cost function $C(X)$ symmetric in x_1.

Corollary 2. *Under the hypothesis and with the notation of the lemma, if Y^+ is the optimal transportation from Ω_1^+ to Ω_2^+ then $Y^+ = Y|_{\Omega_1^+}$, where Y is again $\varphi(X), \psi(Y)$ restricted to X, Y in $(\mathbb{R}^n)^+ = \{X : x_1 > 0\}$ must be the minimizing pair.*

We apply the previous lemma and corollary to densities $f(X)$ and $g(Y)$ in the unit cube of $\mathbb{R}^n$. Let f, g be densities in the unit cube of $\mathbb{R}^n$, $Q_1 = \{X : 0 \leq x_i \leq 1\}$ and Y be the optimal transportation.

Let us write $\boldsymbol{Y} = X + \boldsymbol{V}$ and respectively

$$\varphi(X) = \frac{1}{2}|X|^2 + u(X)$$

(that is $\boldsymbol{V} = \nabla u$). Then

Theorem 5. *If we extend f, g to f^*, g^* on a larger cube Q by even reflections, then $u(X)$ also extends periodically to u^*, to the same cube Q^* by even reflection and $Y(X)$ to the optimal transportation map*

$$Y^* = X + \nabla u^*(X)$$

from Q^* to Q^*.

Corollary 3. *If f, g are strictly positive and C^α in the unit cube Q_1, then $Y(X)$ maps each face of the cube to itself and both $Y(X)$, $X(Y)$ have a $C^{1,\alpha}$ extension across ∂Q.*

Proof. It follows from the interior regularity theory (the above theorem) since each face of Q becomes interior after a reflection.

Remark. The problem of finding "periodic" solutions to the Monge–Ampére equation was solved by Yanyan Li [L] by a different method.

2. Monotonicity Properties of $Y(X)$

We start with a heuristic discussion. Recall that the Holley condition on μ_2, μ_1 was that

$$\mu_2(A \vee B)\mu_1(A \wedge B) \geq \mu_2(A)\mu_1(B).$$

Logarithmically

$$\log \mu_2(A \vee B) - \log \mu_2(A) \geq \log \mu_1(B) - \log \mu_1(A \wedge B).$$

Let us now think on smooth densities $f(X)$, $g(Y)$ on the unit cube, and assume we are trying to prove, by a continuity argument that $Y(X)$ is monotone, that is $Y(X) \geq X$. So we are looking at a continuous family of densities f^t, g^t for which $Y(X) > X$ and we find a first time t_0 and a point X_0, for which $Y(X_0) \not> X_0$, that is some coordinate, say $y_1(X_0) = x_1(X_0)$. That means that $y_1(X) - x_1(X)$ has a local minimum, zero, at X_0.

But it is well known that $y_1 = D_1\varphi$, satisfies an elliptic equation, obtained by differentiating the equation for φ. From

$$\log \det D^2\varphi = \log f(X) - \log g(\nabla\varphi)$$

we get

$$M_{ij} D_{ij}(D_1\varphi) = (\log f(X))_1 - (\log g(\nabla\varphi))_i D_{i1}\varphi.$$

Since $\varphi_1 - x_1$ has a minimum, zero, at X_0,

$$D_{i1}\varphi = \delta_{i1},$$

and we get at X_0, $Y(X_0)$,

$$M_{ij} D_{ij}[y_1 - x_1] = (\log f)_1(X) - (\log g)_1(Y).$$

Since M_{ij} is a strictly positive matrix for φ strictly convex and $y_1 - x_1$ has a minimum, the left-hand side must be non-negative.

If we impose the right-hand to be non-positive we have a contradiction.

About the right-hand side, we know that $Y > X$ and that

$$\langle Y - X, \mathrm{e}_1 \rangle = 0,$$

so the natural hypothesis we want to impose on f, g is that

Monotonicity hypothesis. *If* $Y \geq X$ *and* $\langle Y - X, \mathrm{e}_i \rangle = 0$, *then*

$$D_i(\log g)(Y) \geq D_i(\log f)(X).$$

Note. If we think of $A = Y$ and $B = X + t\mathrm{e}_i$ we can argue that heuristically $B \vee A = Y + t\mathrm{e}_i$ and $B \wedge A = X$, so

$$\log g(Y + t\mathrm{e}_i) - \log g(Y) \geq \log f(X + t\mathrm{e}_i) - \log f(X)$$

becomes Holley's condition. We will show in fact later how to associate to a discrete "Holley" pair a continuous one satisfying our hypothesis.

But first we prove our main comparison theorem.

Theorem 6. *Let f, g be $C^{1,\alpha}$, strictly positive probability densities in the unit cube Q of $\mathbb{R}^n$. Assume that given any X, Y, e_j with $X \le Y$, and $\langle X - Y, \mathrm{e}_j\rangle = 0$ (i.e., $y_j - x_j = 0$)*

$$(D_j \log f)(X) \le (D_j \log g)(Y),$$

and let $Y(X)$ be the optimal transportation map. Then for any X in Q,

$$Y(X) \ge X.$$

Proof. As we pointed out before, we know that the potentials $\varphi(X)$, $\psi(Y)$ are of class $C^{2,\alpha}$ across ∂Q_j and the $C^{1,\alpha}$ optimal transportations $Y(X)$, $X(Y)$ map each face of the cube into itself in a $C^{1,\alpha}$ fashion.

In particular, classical regularity theory for fully non linear equations applies to φ in the interior of the cube. More precisely, φ satisfies

$$\det D_{ij}\varphi = \frac{f(X)}{g(\nabla\varphi)}$$

(see [G-T]) and f, g being $C^{1,\alpha}$ (this is not kept by reflection along the faces), we have that:

$$\varphi \text{ is of class } C^{3,\alpha}(Q).$$

We now study directional derivatives along the boundary of Q_j.

Consider $D_1\varphi$ outside the faces $x_1 = 0, x_1 = 1$. Then, across the remaining boundary of Q_1, $y_1(X) = D_1\varphi$ satisfies

$$M_{ij} D_{ij}(D_1\varphi) = D_1 \log f(X) - D_\ell(\log g) D_{\ell 1}\varphi.$$

Both M_{ij} and the right-hand side are of class C^α (since $D_1 \log f$ is tangential to the face). Hence $y_1(X)$ is of class $C^{2,\alpha}$ across that part of the boundary and the equation is satisfied in the classical sense.

In order to make the f, g relation strict we change g to g_ε by defining

$$\log g_\varepsilon(Y) = \log g + \sum \varepsilon y_i + C_\varepsilon,$$

where the constant C_ε is chosen so that

$$\int g_\varepsilon(Y) = 1.$$

Then from the condition

$$D_j \log f(X) \le D_j \log g(Y)$$

for $y_j - x_j = 0$, we now have for $0 < \gamma < \delta(\varepsilon)$ small enough:

$$D_j \log f(X) \le D_j \log g_\varepsilon(Y) - \delta$$

if $|y_{j_0} - x_{j_0}| < \gamma$ for some j_0 and $y_j - x_j > -\gamma$ for the remaining j. □

We now look at the continuous family of densities f_t, g_t defined by

$$\log f_t = t \log f + C(t),$$
$$\log g_t = t \log g_\varepsilon + D(t),$$

where $C(t)$, $D(t)$ are chosen to keep $\int f_t = \int g_t = 1$ and we show

Lemma 2. *For any* $0 < t < 1$ *the corresponding (continuous in* t*) family of optimal transports* $Y_t(X)$*, satisfies*

$$y_j^t \geq x_j^t - \tfrac{1}{2}\gamma.$$

Proof. For $t = 0$, $Y(X)$ is the identity map, and thus the inequality is satisfied for t small. As usual, suppose there exists a first value $t_0 > 0$, for which the inequality is not satisfied. Thus, there exists X_0 and a j (say $j = 1$) such that

$$y_1(X_0) = x_1(X_0) - \tfrac{1}{2}\gamma$$

and still $y_1(X) \geq x_1(X) - \frac{1}{2}\gamma$ everywhere else.

We first note that $x_1(X_0) \neq 0, 1$ because, if not

$$y_1(X_0) = x_1(X_0).$$

But everywhere else we have

$$0 \leq M_{ij} D_{ij} y_1(X_0)$$

(since $y_1 - x_1$ has a minimum at X_0) and

$$D_1 \log f(X_0) \leq D_1 \log g(Y(X_0)) - t\delta$$

(since $|y_1 - x_1| = \gamma/2$ and $y_j \geq x_j - \gamma/2$ for the remaining j).

This is a contradiction that completes the proof of the lemma and the theorem. □

Corollary 4. *Let* $0 < \lambda \leq f, g \leq \Lambda$ *be measurable. Suppose that* $\log f$, $\log g$ *satisfies the hypothesis of the theorem in the sense of distributions. Then, the theorem still holds, i.e.,*

$$Y(X) \geq X.$$

Proof. Mollify $\log f$, $\log g$ to $\log f_\varepsilon$, $\log g_\varepsilon$ with a standard (radially symmetric, non-negative, compactly supported) mollifier φ_ε. Then the hypothesis of Theorem 6 is satisfied as long as X, Y stay at distance ε from ∂Q_1.

Take as center of coordinates the center of the cube: $X = (\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$ and make a 2ε-dilation. The new f_ε, g_ε satisfy the hypothesis of Theorem 6 when restricted to the unit cube. Thus Theorem 1 holds for them. By passing to the limit on the maps, the theorem holds for f, g. □

3. Holley's Theorem when the Lattice is all of the Vertices of the Unit Cube

Given a vertex $X \in P$, we will denote by Q_X the subcube of Q_j, of side $1/2$ that has X as a vertex

$$Q_X = \{Z : |Z - X|_{L^\infty} \leq 1/2\}.$$

We prove the following theorem.

Theorem 7. *Let f, g be step functions*

$$f = \sum_{X \in P} \mu_1(X) \chi_{Q_X},$$

$$g = \sum_{X \in P} \mu_2(X) \chi_{Q_X}.$$

Assume that given vertices $X, Y, X + \mathrm{e}_j, Y + \mathrm{e}_j$ with $Y \geq X$ and $\langle Y, \mathrm{e}_j \rangle = \langle X, \mathrm{e}_j \rangle = 0$ we have

$$\log \mu_2(Y + \mathrm{e}_j) - \log \mu_2(Y) \geq \log \mu_1(X + \mathrm{e}_j) - \log \mu_1(X).$$

Then $Y(X) \geq X$.

Proof. As a distribution $D_i \log f$ (resp. $D_i \log g$) is the jump function

$$\log \mu_i(X + \mathrm{e}_j) - \log \mu_1(X)$$

supported on the face of Q_X laying in the plane $x_j = 1/2$. $\square$

Corollary 5. *Let $Z_1, Z_2 \in P$. Define*

$$\begin{aligned}\nu(Z_1, Z_2) &= \mu_1(Z_1)/|Q_{1/2}| \, |\{X \in Q_{Z_1} / Y(X) \in Q_{Z_2}\}| \\ &= \mu_2(Z_2)/|Q_{1/2}| \, |\{Y \in Q_{Z_2} / X(Y) \in Q_{Z_1}\}|.\end{aligned}$$

Then

a) *ν is a probability measure with marginals $\mu_1(Z_1), \mu_2(Z_2)$,*
b) $\nu(Z_1, Z_2) \neq 0 \implies Z_2 \geq Z_1$.

4. Holley's Theorem for General Lattices

Given a lattice $\Lambda \subset P$, and two measures μ_1, μ_2 satisfying the Holley condition we want to extend μ_1, μ_2 to small perturbations μ_1^*, μ_2^* in all of P keeping the inequalities. Usually, μ_1, μ_2 are extended by zero. We need to be a little more careful.

We state the following presentation of Λ.

Lemma 3. *There is a partition of*

$$\mathbb{R}^N = \mathbb{R}^{k_1} \otimes \mathbb{R}^{k_2} \otimes \cdots \otimes \mathbb{R}^{k_\ell}$$

and a family of elements w_i^j $(1 \leq j \leq \ell, 1 \leq i \leq k_j)$ such that any non zero element $X \in \Lambda$ is the max of w_i^j,

$$x = \bigvee_{i,j \in I_X} w_i^j$$

and

$$w_i^j = \mathrm{e}_i^j + \boldsymbol{v}$$

with the coordinates $v_i^s = 0 \; \forall \; s \geq j$. (More precisely $w_i^1 = \mathrm{e}_i^1$, $w_i^2 = \mathrm{e}_i^2 + \boldsymbol{v}$, with $v \in \mathbb{R}^{k_1}$, $w_i^3 = \mathrm{e}_i^3 + \boldsymbol{v}$ with $v \in \mathbb{R}^{k_1+k_2}$ and so on.

Proof. The decomposition is by first choosing the minimal elements $\bar{e}_1, \bar{e}_2, \dots, \bar{e}_{k_1}$ and contracting the ones in them to only one position. Next we choose minimal elements among those not in $\mathbb{R}^{k_1}$ and so on.

We now extend the lattice and the measure. Let $\bar{\Lambda}$ be the following extension of Λ:

$$\bar{\Lambda} = \Lambda \cup \Lambda_0, \qquad \text{where } w \in \Lambda_0 \Leftrightarrow \max(w, e_1) \in \Lambda$$

(that is, we add to all those elements with a 1 as first coordinates, those with a zero). Given w in $\bar{\Lambda}$ define

$$w^+ = w \vee e_1,$$
$$w^- = w^+ - e_1 \ \text{(i.e., } w \text{ with a zero in the position } e_1).$$

Define

$$\mu^*(w) = \begin{cases} \mu(w) & \text{if } w \in \Lambda \\ \mu(w^+)/M & \text{otherwise } (M \text{ large}) \end{cases}. \qquad \square$$

Theorem 8. $\bar{\Lambda}$ *is a lattice and* μ_1^*, μ_2^* *still satisfy*

$$\log \mu_2^*(v_1 \vee v_2) - \log \mu_2^*(v_2) \geq \log \mu_1^*(v_1) - \log \mu_1^*(v_1 \wedge v_2).$$

Proof. Elements in $\bar{\Lambda}$ are w^+ and w^- of elements in Λ (w^+ is always in Λ since $e_1 \in \lambda$). Then

$$v_1 \wedge v_2 = w_1^\pm \wedge w_2^\pm$$

for $w \in \Lambda$.

If one of the signs is a $-$,

$$v_1 \wedge v_2 = (w_1 \wedge w_2)^-.$$

If not

$$v_1 \wedge v_2 = w_1 \wedge w_2.$$

Also

$$v_1 \vee v_2 = w_1^\pm \vee w_2^\pm.$$

If one of the signs is a $+$ (since $w^+ \in \Lambda$),

$$v_1 \vee v_2 = w_1 \vee w_2.$$

If not

$$v_1 \vee v_2 = (w_1 \vee w_2)^-.$$

About the measures μ_1^*, μ_2^*, let us verify the proper inequalities. For that purpose we choose $M \gg \mu_i(X)$ for any X. There are several cases to consider

a) $w_1, w_2 \in \Lambda$, then $w_1 \wedge w_2$, $w_1 \vee w_2 \in \Lambda$ and everything is as before.

b) $w_1 \in \Lambda$, $w_2 \notin \Lambda$ (thus $w_2 = w_2^-$).

- b$_1$) If $w_1 = w_1^-$, we have that $w_1 \wedge w_2 \in \Lambda$ and $w_1 \vee w_2 \notin \Lambda$ and the factor $\log M$ cancels in the μ_2^* expression.
- b$_2$) If $w_1 = w_1^+$, $w_1 \vee w_2 \in \Lambda$. If $w_1 \wedge w_2 \in \Lambda$, the extra factor $\log M$ in the μ_2^* expression controls everything else (we choose $\log M \gg \sup |\log \mu_i|$. If $w_1 \wedge w_2 \notin \Lambda$, $\mu_1^*(w_1 \wedge w_2) = \mu_1(w_1 \wedge w_2^+)/M$, and $\mu^*(w_2) = \mu(w_2^+)/M$, thus each term has an extra $\log M$ factor that cancels.

c) $w_2 \in \Lambda$, $w_1 \notin \Lambda$.

c$_1$) If $w_2 = w_2^+$, then $w_1 \vee w_2 \in \Lambda$. If $w_1 \wedge w_2 \in \Lambda$ the extra term $-\log M$ in the μ_1 expression controls everything. If $w_1 \wedge w_2 \notin \Lambda$ then

$$\mu_1^*(w_1 \wedge w_2) = \mu(w_1^+ \wedge w_2)/M,$$
$$\mu_1^*(w_1) = \mu(w_1^+)/M,$$

and we have $\log M$ cancellation.

c$_2$) If $w_2 = w_2^-$, then $w_1 \wedge w_2 \in \Lambda$, but $w_1 \vee w_2 \notin \Lambda$ and we have

$$\mu_2^*(w_1 \vee w_2) = \mu_2(w_1^+ \vee w_2)/M,$$
$$\mu_1^*(w_1)* = \mu_1(w_1^+)/M,$$

and there is a $\log M$ factor cancellation.

d) If $w_1 \notin \Lambda$, $w_2 \notin \Lambda$, then $w_1 \vee w_2 \notin \Lambda$. If $w_1 \wedge w_2 \notin \Lambda$, the factors $\log M$ cancel. If not, the extra factor $\log M$ in the μ_1^* expression controls everything else.

The proof of the theorem is complete. $\square$

Theorem 9. *We are given $\Lambda \subset P$ and μ_1, μ_2. As before, let f, g be the step functions*

$$f = \sum_{w_i \in \Lambda} \mu_1(w_i)\chi_{Q_{w_i}},$$
$$g = \sum_{w_i \in \Lambda} \mu_2(w_i)\chi_{Q_{w_i}}.$$

Then, the optimal transportation map $Y(X)$ is monotone.

Proof. If we start with $M = M_0$ and we repeat the extension process ($M_1 \gg M_0$, $M_2 \geq M_1$ and so on) we exhaust P. Note that once we have extended through $\mathrm{e}_1^1, \dots, \mathrm{e}_{k_1}^1$, the elements $\mathrm{e}_1^2, \dots, \mathrm{e}_{k_2}^2$ belong now to the lattice and are minimal, so we can keep extending. As M_0 goes to infinity the measures μ_i^* converge to μ_i. $\square$

We complete this work by showing that, actually, the discrete optimal transportation map is monotone. In this case the map is in general multi-valued. That is the mass $\mu_1(w)$ may have to be spread through several points v. Still, for all those v's, $v(w) \geq w$.

Theorem 10. *Let Λ be a sublattice of P, the set of vertices of the unit cube on $\mathbb{R}^n$, and let μ_1, μ_2 be positive measures in Λ satisfying the usual monotonicity condition. Let $\nu(X, Y)$ be the (discrete) optimal transportation. Then $\nu(X, Y) \neq 0 \implies Y \geq X$.*

Proof. From the previous theorem we may assume that μ_i is defined and positive in all of P. We will approximate it by bounded densities f, g that satisfy the hypothesis of Theorem 6. We define them as follows.

Let 1 be the vector $1 = (1, 1, \dots, 1)$. In the strip $S_\omega^\varepsilon = \{\varepsilon\omega < X \leq \omega + \varepsilon 1\}$, let $N(X, \omega)$ be the number of coordinates, j, for which $w_j - x_j > \varepsilon$ and we define there, for $\delta \ll \varepsilon$,

$$f(X) = \mu_1(\omega)\delta^N.$$

Note that S_ω^ε cover Q_1 disjointly (given X we determine w by those coordinates $x_j > \varepsilon$). Same definition for g.

Of course, we have to multiply as usual by a normalization constant to make $\int f = \int g = 1$, but this does not affect the logarithmic inequality. Also if δ goes to zero much faster than ε, (say like ε^{2N}) f and g converge to μ_1 and μ_2, since most of the mass concentrates in the cube $Q_\varepsilon(\omega) = \{|x_i - \omega_i| < \varepsilon\}$.

About $D_i \log f$, $D_i \log g$, they are jump functions concentrated on the planes $x_j = \varepsilon$ or $1 - \varepsilon$ so we have to check that the jump inequalities are satisfied. We also may disregard plane intersections since they will not affect $D_i f$ in the distributional sense.

So we check that

a) For $X \leq Y$ and $x_i = y_i = \varepsilon$ we have $\mathrm{Jump}(\log g) \geq \mathrm{Jump}(\log f)$. Indeed when x_i, y_i go through ε we change from evaluating the measures at w_1, (resp. w_2) to $w_1 + \mathrm{e}_i$, $w_2 + \mathrm{e}_i$, and both $N(X)$, $N(Y)$ increase by one, so the jump relation holds (they are the lattice relations plus a factor $\log \delta$.
b) When x_i, y_i go through $(1-\varepsilon)$, w_1 and w_2 remain unchanged and $N(X)$, $N(Y)$ both decrease by one.

Also here the jump relation holds (both jumps are just $\log \delta$).

This completes the proof. □

5. Second Derivative Estimates

In this section we explore what the implications are of the fact that second derivatives of solutions to Monge–Ampére equations are subsolutions of an elliptic equation.

First an heuristic discussion: Let us take a second pure derivative of the equation

$$\log \det D_{ij}\varphi = \log f(x) - \log g(\nabla\varphi).$$

We get

$$M_{ij} D_{ij} \varphi_{\alpha\alpha} + M_{ij,k\ell} D_{ij\alpha}\varphi D_{ij\beta}\varphi = D_{\alpha\alpha} \log f - (\log g)_{ij} \varphi_{i\alpha}\varphi_{j\alpha} - (\log g)_i \varphi_{\alpha\alpha i}.$$

From the concavity of $\log \det$ the second term on the left is negative. If $\varphi_{\alpha\alpha}$ reaches at X_0 the maximum value among all pure second derivatives, then the right-hand side must be negative. Let us look at the explicit case in which up to a constant, $f = \mathrm{e}^{-Q(X)}$ and $g = \mathrm{e}^{-(Q(Y)+F(Y))}$, where Q is a nonnegative quadratic polynomial, $a_{ij}x_i x_j$ (for instance, near neighborhood or other “Dirichlet Integral” like interactions in field theory).

We may assume that $\alpha = \mathrm{e}_1$. Then, we must compute

$$D_{11}(-Q(X) + Q(\nabla\varphi) + F(\nabla\varphi),$$

we have

$$\begin{aligned} D_{11}(-Q)(X) &= -a_{11}, \\ D_{11} Q(\nabla\varphi) &= a_{ij}\varphi_{i1}\varphi_{j1} + a_{ij}\varphi_{i11}\varphi_j. \end{aligned}$$

But since $\varphi_{11}(X_0)$ is the maximum among all pure second derivatives, $\varphi_{11i} = 0$ for all i, and $\varphi_{1i} = 0$ for $i \neq 1$. So $D_{11} Q(\nabla\varphi(X_0)) = a_{11}(\varphi_{11})^2$. Finally, if F is convex

$$D_{11} F(\nabla\varphi) = F_{ij}\varphi_{i1}\varphi_{j1} + F_i \varphi_{i11}$$

is non-negative.

Therefore D_{11}(R.H.S.) $\geq a_{11}((\varphi_{11})^2 - 1)$. We get a contradiction if $\varphi_{11} > 1$. That is

Theorem 11. *Let, up to a multiplicative constant,*

$$f(X) = \mathrm{e}^{-Q(X)},$$
$$g(Y) = \mathrm{e}^{-(Q(Y)+F(Y))}$$

with F convex. Then the potential φ of the optimal transportation satisfies

$$0 \leq \varphi_{\alpha\alpha} \leq 1.$$

In particular

$$Y = X + \nabla u(X),$$

where

$$u = \varphi - \tfrac{1}{2}|X|^2$$

is concave and

$$-1 \leq u_{\alpha\alpha} \leq 0$$

(independently of dimension).

Proof. To make the previous theorem valid we have to take care of what happens when X goes to infinity.

Again by approximation we may assume that the convex function $F(X)$ is $+\infty$ outside the ball B_R (that is g is supported in the ball of radius R, and smooth bounded away from zero and infinity inside it.

We will replace the second derivative by an incremental quotient, and show that it still satisfies a maximum principle and goes to zero at infinity. Let

$$(\delta\varphi_{\mathrm{e}})(X) = \varphi(X + h\mathrm{e}) + \varphi(X - h\mathrm{e}) - 2\varphi(X).$$

We fix h, and study what happens if $\delta\varphi = \delta\varphi_{\mathrm{e}_1}$ attains a local maximum at X_0, for all possible e. From the concavity of log det, we still have that, for the linearization coefficients M_{ij}, of log det at X_0,

$$M_{ij}\delta\varphi(X_0) \leq \delta(\log f - \log g) = \delta(-Q(X_)) + Q(\nabla\varphi) + F(\nabla\varphi).$$

From the fact that $\delta\varphi_{\mathrm{e}_1}$ realizes a maximum among X and e, we obtain

a) $\nabla\delta\varphi = \nabla\varphi(X_0 + h\mathrm{e}_1) + \nabla\varphi(X_0 - h\mathrm{e}_1) - 2\nabla\varphi(X_0) = 0$
and
b) for any $\tau \perp \mathrm{e}_1$,

$$D_\tau\delta\varphi = \tau \cdot (\nabla\varphi(X_0 + h\mathrm{e}_1) - \nabla\varphi(X_0 - h\mathrm{e}_1) = 0.$$

Therefore

$$\nabla\varphi(X \pm h\mathrm{e}_1) = \nabla\varphi(X) \pm \lambda\mathrm{e}_1$$

and $\delta\varphi = 2\lambda$ (λ positive). Then, from the convexity of F,

$$\delta F(\nabla\varphi(X_0)) \geq 0.$$

If we write $Q(X)$ as a bilinear form $Q(X) = B(X, X)$,

$$\begin{aligned}\delta Q(\nabla\varphi) = \; & B(\nabla\varphi(X_0) + \lambda e_1, \nabla\varphi(X_0) + \lambda e_1)\\ & + B(\nabla\varphi(X_0) - \lambda e_1, \nabla\varphi(X_0) - \lambda e_1)\\ & - 2B(\nabla\varphi(X_0), \nabla\varphi(X_0))\\ = \; & \lambda^2 B(e_1, e_1).\end{aligned}$$

Similarly $\delta Q(X) = h^2 B(e_1, e_1)$ so we get: If $\delta\varphi$ has an interior maximum at X_0, then it must hold:

$$\nabla\varphi(X_0 \pm h e_1) = \nabla\varphi(X_0) \pm \lambda e_1$$

with $\lambda < h$.

But, since φ is convex

$$\varphi(X_0 \pm h e_1) - \varphi(X_0) \le \langle \nabla\varphi(X_0 \pm h e_1) - \nabla\varphi(X_0), \pm h e_1\rangle = \lambda h \le h^2.$$

Thus,

$$\delta\varphi \le 2h^2,$$

the desired inequality. □

To complete the proof of the theorem it would be enough to show that $\delta\varphi$ goes to zero (for fixed δ) when X goes to infinity.

We show that:

Lemma 4. *As X goes to infinity Y converges uniformly to $R\frac{X}{|X|}$.*

Proof. Let $X_0 = \lambda e_1$ for λ large and Y_0 its image. Let ν be a unit vector with

$$\text{angle}\,(\nu, e_1) \le \frac{\pi}{2} - \varepsilon.$$

From the monotonicity of the map, any point on B_R of the form

$$Y' = Y_0 + t\nu$$

must come from a vector

$$X' = X_0 + s\mu,$$

with $\langle \mu, \nu\rangle \ge 0$.

In particular, we must have

$$\text{angle}\,(\mu, e_1) \le (\pi - \varepsilon).$$

In other words if in Y space we consider the cone,

$$\Gamma = \{Y' = Y_0 + t\nu, \;\; \text{with} \;\; t > 0, \;\; \text{angle}\,(\nu, e_1) \ge \frac{\pi}{2} - \varepsilon,$$

its intersection with B_R must be covered by the image of the (concave) cone

$$\overline{\Gamma} = \{X' = X_0 + s\mu, \;\; \text{with } s > 0 \text{ and angle}\,(\mu, e_1) \le \pi - \varepsilon\}.$$

But $\overline{\Gamma}$ has very small f measure

$$\mu_f(\overline{\Gamma}) \le (\varepsilon\lambda)^n e^{-(\varepsilon\lambda)^2}, \qquad \varepsilon\lambda > \lambda^{1/2},$$

since the ball of radius $\varepsilon\lambda$ is not contained in $\overline{\Gamma}$.

On the other hand, g is strictly positive in B_R, so

$$\mu_g(\Gamma \cap B_R) \sim |\Gamma \cap B_R| \le \mu_f(\overline{\Gamma}).$$

This forces the exponential convergence of Y to $R\mathrm{e}_1$.

This completes the proof of the lemma and the theorem, since the uniform convergence of $\nabla\varphi$ to $\frac{X}{|X|}$, makes $\delta\varphi$ go to zero (for any fixed, positive h). □

We state three corollaries of this last inequality. The first two are a generalization of the classic Brascamp–Lieb moment inequality and the third an eigenvalue inequality.

Corollary 6. *Let $f(X) = e^{-Q(X)}$, $g(H) = e^{-[Q(Y)+F(Y)]}$ with Q quadratic and F convex, and let Γ be a convex function of one variable ($|x_1|^\alpha$ in [B-L]). Then*

$$E_g(\Gamma(y_1 - E_g(y_1)) \le E_f(\Gamma(x_1)).$$

Proof. It follows from [B-L] that it is enough to prove it in the one dimensional case (see Theorem 5.1 of [B-L]). We can also assume by a translation that $E_g(y_1) = 0$.

By the change of variable formula that means

$$\int y(x) f(x)\, dx = 0.$$

Also

$$E_g(\Gamma(y_1)) = \int \Gamma(y_1(x) f(x)\, dx.$$

But $y(x) = x + u(x)$, where $y = \varphi'(x)$, φ convex and $u = \psi'(x)$, ψ concave. Thus y is increasing, and u is decreasing and changes sign, since

$$\int u(x) f(x)\, dx = \int y(x) f(x)\, dx = 0.$$

Say $u(x_0) = 0$. Then, we write

$$\int \Gamma(y(x)) f(x) \le \int [\Gamma(x) + \Gamma'(y(x))(y - x)] f(x).$$

Since Γ is convex,

$$\le E_f(\Gamma(x)) + \int [\Gamma'(y(x)) - \Gamma'(x_0)](y - x) f(x).$$

But at x_0, $\Gamma'(y(x_0)) = \Gamma'(x_0)$ and $y(x_0) = x_0$, and further Γ' is increasing, while $y - x = u$ is decreasing, thus the last integrand is negative, and this completes the proof.

If we want to repeat the argument above for functions Γ that depend on more than one variable, and we want to prove that

$$E_g(\Gamma(Y - E_g(Y)) \le E_f(\Gamma(X)),$$

we may as before assume that $E_g(Y) = 0$.

That means, with $Y = X + U$, that $U(X_0) = 0$ for some X_0 (i.e., the concave function $-\psi$ has a maximum). The same computation then gives us

$$E_g(\Gamma(Y)) \le E_f(\Gamma(X)) + \int (\nabla\Gamma(Y) - \nabla(\Gamma(X_0))(-\nabla\psi(Y) f(X)\, dx,$$

where ψ and $\Gamma - \langle \nabla\Gamma(X_0), X - X_0 \rangle$ are both convex with a minimum at X_0, so there is some hope that the integrand be negative. □

For instance, if we are looking at statistics of k-variables we have the following corollary.

Corollary 7. *Assume that $Q(X)$, $F(X)$ in the definition of $f(X)$, $g(Y)$ are symmetric with respect to $(x_1, \dots, x_k)$ and that $\Gamma(x_1, \dots, x_k)$ is convex and symmetric. Then*

$$E_g(\Gamma(Y)) \le E_f(\Gamma(X)).$$

Proof. As before we may assume the problem is k-dimensional ([B-L], Theorem 4.3). Since Q and F are symmetric, the potentials $\varphi(X)$, $\psi(X)$ are symmetric. Therefore $\nabla\varphi, \nabla\psi, \nabla\Gamma = 0$ for $X = 0$ and further,

$$\text{sign } \varphi_i(X) = \text{sign } \psi_i(X) = \text{sign } \Gamma_i(X) = \text{sign } x_i = \text{sign } y_i.$$

From the computation above it suffices to show that for all Y,

$$\nabla\Gamma \cdot \nabla\psi \ge 0.$$

That follows since $\Gamma_i \cdot \psi_i \ge 0$ for all i. □

A final consequence of the estimate $\varphi_{\alpha\alpha} \le 1$ for log concave perturbations of the Gaussian is that any Raleigh-like quotient (log Sobolev inequality, isoperimetric inequality, Poincaré inequality) that involves a quotient between first derivatives and the function themselves is smaller for the perturbation than for the Gaussian.

For instance, let $F(t)$, $G(t)$, $H(t)$, $K(t)$ be non-negative, non-decreasing functions of $t \in [0, \infty)$, then we have the

Corollary 8. *Let f, g be densities as in Theorem 11 (i.e., a Gaussian and its log concave perturbation) then consider the "Raleigh" quotient*

$$\lambda_f = \inf \frac{F(\int G(|\nabla u|) f(X)\, dX)}{H(\int K(|u|) f(X)\, dX)}.$$

Then $\lambda_g \ge \lambda_f$.

Proof. If we apply the change of variable formula to any function $u(Y)$, we get

$$\int K(|u(Y)|) g(Y)\, dY = \int K(|u(X)|) f(X)\, dX,$$

while $\nabla_X u(Y(X)) = D_X(Y) \nabla_Y u(X)$. But $D_X Y$ is a symmetric matrix with all eigenvalues less than one, so $|\nabla_X u(Y(X))| \le |\nabla_Y u(Y)|$ which proves the corollary. □

Remark. The monotonicity for the log Sobolev inequality under log concave perturbations of the Gaussian follows from the Bakry–Emery theorem ([B-E]).

References

[B-E] Bakry, D. and Emery, M.: Diffusions hypercontractives. In: *Sém. Prob. XIX*, LNM 1123 Berlin–Heidelberg–New York: Springer, 1985, pp. 177–206
[B-L] Brascamp, H. and Lieb, E.: On extentions of the Brunn-Minkowski and Prékopa-Leindler Theorems, Including Inequality for Log Concave functions, and with an Application to the Diffusion Equation. J. Funct. Anal. **22**, 366–389 (1976)
[B] Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Comm. Pure Appl. Math. **XLIV**, 375–417 (1991)
[C1] Caffarelli, L.A.: Interior $W^{2,p}$ estimates for solutions of the Monge–Ampére equation. Ann. of Math. **131**, 135–150 (1989)
[C2] Caffarelli, L.A.: The regularity of mappings with a convex potential. J.A.M.S. **5**, 99–104 (1992)
[C3] Caffarelli, L.A.: Boundary regularity of maps with convex potential I. Comm. Pure Appl. Math. **45**, 1141–1151 (1992)
[FKG] Fortiun, C.M., Kasteleyn, P.W., Ginibre, J.: Correlation inequalities on some partially ordered sets. Commun. Math. Phys. **22**, 89–103 (1971)
[G-M] Ganzbo, W., McCann, R.J.: The geometry of optimal transport. Acta Math. **177**, 2, 113–161 (1996)
[G-T] Gilbarg, P., Trudinger, N.: *Elliptic partial differential equations of second order*. Second edition, Berlin–Heidelberg–New York: Springer, 1983
[H] Holley, R.: Remarks on the FKG inequalities. Commun. Math. Phys. **36**, 227–231 (1974)
[L] Li, Yanyan: Some existence results of fully non-linear elliptic equations of Monge-Ampere type. Comm. Pure Appl. Math. **43**, 233–271 (1990
[P] Preston, C.J.: A generalization of the FKG inequalities. Commun. Math. Phys. **36**, 232–241 (1974)

Communicated by J. L. Lebowitz