DIPLOMARBEIT

Optimal Transport and Geometric Inequalities

Ausgeführt am Institut für
Diskrete Mathematik und Geometrie
der Technischen Universität Wien

unter der Leitung von
Assoc. Prof. Dr. Franz Schuster

eingereicht an der
Technischen Universität Wien durch

Robert Bruno Schlichtner
Vivenotgasse 1/3/9
1120 Wien

Wien, am 18. Dezember 2012

Contents

Introduction

In this thesis it will be shown that many well known geometric and analytic inequalities, namely the

Brunn-Minkowski inequality,

Prèkopa-Leindler inequality,

Minkowski inequality for mixed volumes,

Brascamp-Lieb inequality and the reverse Brascamp-Lieb inequality,

Gagliardo-Nirenberg-Sobolev inequality

can be proven in a similar way using one tool:

The Brenier map. This is a map from $\mathbb{R}^n$ to $\mathbb{R}^n$ derived from a convex potential pushing forward one probability measure to another. It was shown by Brenier that such a map exists always if the pair of measures satisfies some rather weak regularity conditions.

The questions that led to the discovery of this map however, are not of the kind that we are going to present here. Its origin lies in fact in a question by Gaspard Monge in 1781. He asked for the shortest way to displace an amount of soil from one place of the Euclidean space to a heap of soil at another place.

A modern reformulation of this question known as **Monge's optimal transport problem** can be stated as follows:
Given two probability measures μ and ν, find a map T pushing forward μ to ν that minimises the integral

$$\int c(x, T(x)) d\mu(x)$$

for some cost function c.

It turns out that the most simple case is the one where c is the square of the Euclidean distance: In this case there is always a minimising map and it is the Brenier map.

In this thesis optimal transport will mostly be seen as a vehicle to construct the Brenier map. We will not care that this monotone map minimises the cost of transportation but will rather make use of its analytic properties.

This work is structured as follows: In the first part, i.e. the Chapters 1 and 2, we will introduce some basic notions about optimal transport and develop the main tools we will need later.

In the second part we will give short introductions to the diverse geometric and analytic inequalities we want to discuss, and we will give proofs for them that are related to optimal transport.

To be more precise, Chapter 1 will introduce some basic notions about convex analysis that we will need in later chapters.

Chapter 2 will detail Monge's optimal transport problem and introduce a relaxed version of it, called Kantorovich's optimal transport problem. Limiting ourselves to the case of a quadratic cost function we will sketch the proof of the Knott-Smith optimality criterion and obtain Brenier's theorem as a corollary.

For absolutely continuous measures we will deduce that the very useful Monge-Ampère equation, a partial differential equation, linking the densities of these measures and the Brenier map, holds almost everywhere.

This introduction will mostly follow Cedric Villani [17][especially Chapters 2 and 4].

In Chapter 3 we consider the Brunn-Minkowski inequality and its functional version, the Prèkopa-Leindler inequality. We will give two direct proofs (from McCann [13], Barthe [3] and Ball [2] respectively) using optimal transport for each inequality. This chapter will also introduce the notions of McCann interpolation and displacement convexity.

In Chapter 4 we will further study the Minkowski sum of convex bodies, introducing the notion of mixed volumes and giving a proof of the Minkowski inequality, using a method introduced by S. Alesker, S. Dar and V. Milman [1].

In Chapter 5 the Brascamp-Lieb inequalities and their duals, the reverse Brascamp-Lieb inequalities, will be considered and proved via the Brenier map. This was originally done by Barthe [4]. A classification of optimisers by S. I. Valdimarsson [16] will be stated for the Brascamp-Lieb inequalities.

In Chapter 6 it will be shown how optimal transport does not only allow us to proof certain Sobolev inequalities, but also to obtain a classification of all their optimisers in a rather simple way. These proofs closely follow an article by D. Cordero-Erausquin, B. Nazaret and C. Villani [7].

1 Background

In this first chapter we will introduce some results, which, although they are not directly linked to optimal transportation, will be important for later proofs.

First we will state the well known arithmetic/geometric mean inequality and some consequences of it for determinants of symmetric nonnegative matrices. These results can also be found in [17][pages 156-157]. Nonnegativity of a symmetric matrix M here means that all eigenvalues of M are greater or equal than zero or equivalently that $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$.

Theorem 1.1. The arithmetic/geometric mean inequality

(i) Let $(x_i)_{1 \leq i \leq n}$ and $(t_i)_{1 \leq i \leq n}$ be nonnegative real numbers satisfying $\sum\limits_{i=1}^{n} t_i = 1$.

Then (with the convention that $0^0 = 1$),

$$\sum_{i=1}^{n} t_i \cdot x_i \geq \prod_{i=1}^{n} x_i^{t_i}.$$

(ii) Let A and B be two nonnegative symmetric $n \times n$ matrices and $t \in [0, 1]$. Then,

$$\begin{array}{rrcl} & \det(t \cdot A + (1-t) \cdot B)^{\frac{1}{n}} & \geq & t \cdot \det(A)^{\frac{1}{n}} + (1-t) \cdot (\det(B))^{\frac{1}{n}} \\ \textit{and} & \det(t \cdot A + (1-t) \cdot B) & \geq & (\det(A))^t (\det(B))^{1-t}. \end{array}$$

(iii) Let S be a symmetric matrix with all its eigenvalues less or equal than 1. Then the function

$$t \mapsto \det(I_n - tS)^{\frac{1}{n}}$$

is concave for $t \in [0, 1]$.

Remark 1.1. In the proof of Theorem 1.1 some well known facts about symmetric nonnegative matrices will be needed:

Let A and B be symmetric nonnegative matrices. Then the following statements hold:

(i) There exists a symmetric nonnegative matrix $\sqrt{(A)}$ with $A = \sqrt{(A)}^2$.

(ii) The matrix ABA is itself symmetric nonnegative.

Proof. As A is symmetric it can be diagonalised by an orthogonal matrix O and a diagonal matrix D with nonnegative entries:

$$A = O^T DO = O^T \sqrt{(D)} O \cdot O^T \sqrt{(D)} O,$$

where $\sqrt{D} = \sqrt{(\operatorname{diag}(t_1, ..., t_n))} = \operatorname{diag}(\sqrt{(t_1)}, ..., \sqrt{(t_n)})$. Defining $\sqrt{(A)}$ as $O^T \sqrt{(D)} O$ finishes the proof of the first part.

Obviously ABA is symmetric as $(ABA)^T = A^T B^T A^T = ABA$. To prove nonnegativity it suffices to show that $x^T ABAx = (Ax)^T B(Ax) \geq 0$ for all $x \in \mathbb{R}^n$. This is obvious as B is nonnegative. $\square$

Proof of the arithmetic/geometric mean inequality. (i) is a simple consequence of the concavity of the logarithm.

To prove (ii) it will be sufficient to prove

$$(\det(A + B))^{\frac{1}{n}} \geq (\det A)^{\frac{1}{n}} + (\det B)^{\frac{1}{n}}$$

because of the n-homogeneity of the determinant.
As a first step we will prove this for regular matrices A. As A is symmetric nonnegative it has an invertible symmetric nonnegative squareroot $\sqrt{(A)}$.
By multiplying the inequality with $\det(\sqrt{(A)})^{-\frac{1}{n}}$ from both sides we obtain

$$(\det(I_n + D))^{\frac{1}{n}} \geq (\det I_n)^{\frac{1}{n}} + (\det D)^{\frac{1}{n}} \tag{1.1}$$

with $D = \sqrt{(A)}^{-1} \cdot B \cdot \sqrt{(A)}^{-1}$. By Remark 1.1, D is symmetric nonnegative itself. As D can be written as $\operatorname{diag}(t_1, ..., t_n)$ with $t_i > 0$ with respect to an appropriate orthonormal base, we can write (1.1) as

$$\left(\prod_{i=1}^{n} (1 + t_i)\right)^{\frac{1}{n}} \geq 1 + \left(\prod_{i=1}^{n} t_i\right)^{\frac{1}{n}}.$$

It follows from (i) that

$$1 = \frac{1}{n} \sum \frac{1}{1 + t_i} + \frac{1}{n} \sum \frac{t_i}{1 + t_i} \geq \left(\prod \frac{1}{1 + t_i}\right)^{\frac{1}{n}} + \left(\prod \frac{t_i}{1 + t_i}\right)^{\frac{1}{n}}.$$

Multiplying this by $(\prod(1 + t_i))^{\frac{1}{n}}$ finishes the proof for regular A.

If A is a not invertible, we can approximate it by regular matrices. Since det is continuous, (ii) holds.

The proof of (iii) is done by the following simple calculation using (ii):

$$\begin{aligned}\det(I_n - \lambda t_1 S - (1-\lambda)t_2 S)^{\frac{1}{n}} &= \det(\lambda(I_n - t_1 S) + (1-\lambda)(I_n - t_2 S))^{\frac{1}{n}} \\ &\geq \lambda \det(I_n - t_1 S)^{\frac{1}{n}} + (1-\lambda)\det(I_n - t_2 S)^{\frac{1}{n}}.\end{aligned}$$

Here the nonnegativity of $I_n - t_i S$ is a consequence of the assumption that all eigenvalues of S are less than 1. □

1.1 Lebesgue Density Theorem

In this section we will introduce derivatives of measures and proof a useful density theorem by Lebesgue. This proof can be found in [15][Chapter 8].

To define the notion of a derivative of a measure we will introduce **substantial families** of open sets:

Definition 1.1. *A collection Ξ of open sets in $\mathbb{R}^n$ will be called a substantial family if it satisfies the following conditions:*

(i) There is some positive constant $C < \infty$ such that for all $X \in \Xi$ there exists an open ball B containing X with $\mathrm{vol_n}[B] < C\,\mathrm{vol_n}[X]$.

(ii) For every ball $B \subset \mathbb{R}^n$ there is an $X \in \Xi$ with $X \subset B$.

The first condition means essentially that sets that are small in volume also have to be small in diameter. The second condition guarantees that there are enough sets.
The collections of all balls or all cubes are simple examples of substantial families.

Now we can define the derivative of a measure:

Definition 1.2. *Let μ be a Borel measure on $\mathbb{R}^n$. For every $r > 0$ and $x \in \mathbb{R}^n$ define*

$$\begin{aligned}\underline{\Delta}_r(x) &= \inf\left\{\frac{\mu[X]}{\mathrm{vol_n}[X]} : x \in X \subset B(x,r), X \in \Xi\right\},\\ \bar{\Delta}_r(x) &= \sup\left\{\frac{\mu[X]}{\mathrm{vol_n}[X]} : x \in X \subset B(x,r), X \in \Xi\right\},\end{aligned}$$

where Ξ is a given substantial family, and define the lower and upper derivative of μ by

$$(\underline{D}\mu)(x) = \lim_{r\to 0} \underline{\Delta}_r(x) \quad \textit{and} \quad (\bar{D}\mu)(x) = \lim_{r\to 0} \bar{\Delta}_r(x)$$

respectively.

We say that μ is differentiable at a point $x \in \mathbb{R}^n$ if the lower is equal to the upper derivative and than we define its derivative as

$$(D\mu)(x) = (\underline{D}\mu)(x) = (\bar{D}\mu)(x).$$

Now our aim in this section is to prove the following theorem:

Theorem 1.2. *Let $\mu = fdx + \sigma$ be a Borel measure on $\mathbb{R}^n$, where fdx is the absolute continuous and σ the singular part of μ. We assume further that μ is locally finite, meaning that for every $x \in \mathbb{R}^n$ there is an open neighborhood N_x of x with $\mu[N_x] < \infty$.*

Then μ is differentiable almost everywhere and the equation

$$f(x) = (D\mu)(x),$$

where $(D\mu)$ is defined with respect to some substantial family Ξ, holds for almost all x.

The points satisfying this are called Lebesgue points of μ (or f).

For the proof we will need some lemmata. We begin with a covering lemma for substantial families:

Lemma 1.3. *Let Ξ be a substantial family in $\mathbb{R}^n$ and let Φ be a finite subcollection of Ξ. Then there exists a subcollection Φ' of Φ consisting of pairwise disjoint sets satisfying*

$$\mathrm{vol_n}\left[\bigcup \Phi\right] \leq C \cdot 3^n \cdot \mathrm{vol_n}\left[\bigcup \Phi'\right].$$

Here C is the constant from the definition of substantial families.

Proof. We order the elements $X_1, X_2, X_3, ..., X_k$ of the collection Φ by their diameter in decreasing order.
Let $i_1 = 1$, let i_2 be the smallest integer such that X_{i_2} is disjoint from X_{i_1} and then let generally i_m be the smallest index such that X_{i_m} is disjoint from all X_{i_p} with $1 \leq p < m$. Do this as long as it is possible to find such indices and let $\Phi' = \{X_{i_1}, X_{i_2}, ...\}$.

By definition, each X_{i_p} lies in some ball B_p satisfying

$$\mathrm{vol_n}[B_p] < C\,\mathrm{vol_n}[X_{i_p}].$$

Now we observe that for each X_k there is some index $i_p \leq k$ such that X_{i_p} intersects X_k. As the diameter of X_k is smaller as that of X_{i_p} by our construction, X_k has to be included in the ball $3B_p$.
Thus we can calculate

$$\mathrm{vol_n}\left[\bigcup \Phi\right] \leq \mathrm{vol_n}\left[\bigcup 3B_p\right] \leq 3^n \cdot \mathrm{vol_n}\left[\bigcup B_p\right] \leq C \cdot 3^n \cdot \mathrm{vol_n}\left[\bigcup \Phi'\right]$$

which completes the proof. □

Lemma 1.4. *If μ is a Borel measure, then $\bar{D}\mu$ is measurable.*

Proof. We consider the functions $\bar{\Delta}_r$ as defined in Definition 1.2. As $\bar{D}\mu$ is defined as a limit of such functions it will be sufficient to show that these are measurable.

To do this we will prove that the sets $\{x : \bar{\Delta}(x)_r > \alpha\}$ are measurable for all $C > 0$. If α is such a number and x is some point with $\bar{\Delta}_r(x) > \alpha$, then there exists an $X \in \Xi$ such that $x \in X \subset B$ where B is some ball with radius r and $\mu[X] > \alpha \,\mathrm{vol_n}[X]$. It follows that $\bar{\Delta}_r(y) > \alpha$ for all $y \in X$ and so $\{x : \bar{\Delta}(x)_r > \alpha\}$ can be written as a union of open sets and is as such open and therefore measurable itself. $\square$

Lemma 1.5. *Let Ξ be a substantial family. Let μ be a positive Borel measure which is finite on compact sets. Let A be a measurable set with $\mu[A] = 0$. Then $(D\mu)(x) = 0$ for almost all $x \in A$.*

Proof. Let P be the set of all x for which $(\bar{D}\mu)(x) > 0$. By Lemma 1.4, P is measurable and so is $A \cap P$. So it is sufficient to show that $\mathrm{vol_n}[A \cap P] = 0$.

We assume that this is false. Then there is a constant $\alpha > 0$ and a Borel set $X_\alpha \subset A \cap P$ with $\mathrm{vol_n}[X_\alpha] > 0$ and $(\bar{D}\mu)(x) > \alpha$. By the regularity of the volume we see that X_α contains a compact subset K with $\mathrm{vol_n}[K] > 0$.
Fix $\delta > 0$. For each $x \in K$ there is a set $S \in \Xi$ such that $\mathrm{diam}(S) < \delta$ and $\mu[S] > \alpha \,\mathrm{vol_n}[S]$. Since K is compact, there is a finite subcollection of these sets S covering K and, by Lemma 1.3, there is a subcollection $\{S_1, S_2, ..., S_k\}$ with the following properties:

1. All the S_i are pairwise disjoint.
2. $\mu[S_i] > \alpha \,\mathrm{vol_n}[S_i]$.
3. $\sum_{i=1}^{k} \mathrm{vol_n}[S_i] \geq C^{-1} \cdot 3^{-n} \cdot \mathrm{vol_n}[K]$.

Let K_δ be the set of all points whose distance from K is less than δ. As $S_i \subset K_\delta$ for all i, we can calculate

$$\mu\left[K_\delta\right] \geq \mu\left[\bigcup_{i=1}^{k} S_i\right] = \sum_{i=1}^{k} \mu[S_i] > \alpha \sum_{i=1}^{k} \mathrm{vol_n}[S_i] \geq \alpha \cdot C^{-1} \cdot 3^{-n} \cdot \mathrm{vol_n}[K]. \tag{1.2}$$

Taking δ to be $\frac{1}{m}$ for $n \geq 1$ we obtain $\mu[K] = \lim_{m\to\infty} K_{\frac{1}{m}}$ as K is the intersection of the decreasing sequence $\{K_1, K_{\frac{1}{2}}, K_{\frac{1}{3}}, ...\}$ and $\mu[K_1]$ is finite.
Therefore, (1.2) implies that

$$0 = \mu[A] \geq \mu[K] \geq \alpha \cdot C^{-1} \cdot 3^{-n} \cdot \mathrm{vol_n}[K] > 0.$$

This contradiction shows that $(\bar{D}\mu)(x) \leq 0$ almost everywhere and, as μ was positive, $(\bar{D}\mu)(x) = 0$ almost everywhere, finishing the proof. $\square$

Lemma 1.6. *If μ is a singular Borel measure then $(D\mu)(x) = 0$ almost everywhere.*

Proof. Without loss of generality we can assume that μ is positive. Since $\mu \perp \mathrm{vol_n}$ there is a Borel set A such that $\mu[A] = 0$ and $\mathrm{vol_n}[A^c] = 0$. Applying Lemma 1.5 finishes the proof. $\square$

Now we are in a position to prove Lebesgue's density theorem.

Proof of Theorem 1.2. By Lemma 6.7 and the Lebesgue decomposition theorem we need only consider absolutely continuous measures μ. By the Radon-Nikodym theorem there is a density $f \in L_1$ with

$$\mu[E] = \int_E f(x)dx$$

for all measurable sets E. It is therefore sufficient to prove that the equality

$$(D\mu)(x) = f(x)$$

holds almost everywhere. Let r be a rational number and define sets A and B by

$$A = \{x : f(x) < r\} \qquad \text{and} \qquad B = \{x : f(x) \geq r\}.$$

Let $\lambda[E] = \int_{E\cap B} (f(x) - r)dx$ for all Borel sets E. For every $E \in \Xi$ we obtain

$$\mu[E] - r\,\mathrm{vol_n}[E] = \int_E (f(x) - r)dx \leq \lambda[E]$$

or, calculating the upper derivatives of both sides of these inequalities,

$$(\bar{D}\mu)(x) - r \leq (\bar{D}\lambda)(x).$$

Since $(D\lambda)(x) = 0$ almost everywhere on A, we can conclude (using Lemma 1.5) that

$$(\bar{D}\mu)(x) \leq r$$

almost everywhere.
Therefore, if $E_r = \{x : f(x) < r < (\bar{D}\mu)(x)\}$, we have shown that E_r is a null set. Since $\{x : f(x) < (\bar{D}\mu)(x)\} = \bigcup_{r\in\mathbb{Q}} E_r$ we obtain

$$f(x) \geq (\bar{D}\mu)(x)$$

almost everywhere.
On the other hand, if we replace μ with $-\mu$ we can in the same way obtain that $f(x) \leq (\underline{D}\mu)(x)$ almost everywhere. Combining these results finishes the proof. $\square$

1.2 Some Facts About Convex Functions

In this section we will state without proof some results about convex functions, most of them concerning (twice) differentiability of convex functions. These results can be found in [14].

Convex Functions

Definition 1.3. *A convex function ϕ on $\mathbb{R}^n$ is a function $\phi : \mathbb{R}^n \to \mathbb{R}$ satisfying*

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y)$$

for all $x, y \in \mathbb{R}^n$ and $t \in [0,1]$. If $\psi : \mathbb{R}^n \to \mathbb{R}$ is a function, such that

$$\psi(tx + (1-t)y) \geq t\psi(x) + (1-t)\psi(y)$$

for all $x, y \in \mathbb{R}^n$ and $t \in [0,1]$, then it is called concave.

A convex function ϕ is automatically continuous and locally Lipschitz. Therefore, by Rademacher's theorem, its gradient $\nabla\phi$ exists almost everywhere.

In order to deal with nondifferentiable points one defines the subdifferential $\partial\phi : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n) = \{X \subset \mathbb{R}^n\}$ of a convex function ϕ by

$$y \in \partial\phi(x) :\Leftrightarrow \phi(z) \geq \phi(x) + y \cdot (z - x) \qquad \forall z \in \mathbb{R}^n .$$

If ϕ is differentiable in x, the subdifferential coincides with the gradient, $\partial\phi(x) = \{\nabla\phi(x)\}$.

Legendre Duality

Definition 1.4. *For a convex function ϕ one defines its convex conjugate function, or Legendre transformation ϕ^* by*

$$\phi^*(y) = \sup_{x \in \mathbb{R}^n} x \cdot y - \phi(x).$$

Obviously, for all $x, y \in \mathbb{R}^n$,

$$x \cdot y \leq \phi(x) + \phi^*(y). \tag{1.3}$$

If ϕ is lower semicontinuous then ϕ^* is also lower semicontinuous and $\phi^{**} = \phi$. (Legendre duality).

Theorem 1.7. *Convex conjugate functions give us a characterisation of subdifferentials:*

$$y \in \partial\phi(x) \Leftrightarrow x \in \partial\phi^*(y).$$

Proof. The subgradient inequality $\phi(z) \geq \phi(x) + y \cdot (z - x)$, which defines $y \in \partial\phi(x)$, can be rewritten as

$$x \cdot y - \phi(x) \geq z \cdot y - \phi(z) \qquad \forall z \in \mathbb{R}^n .$$

Taking the supremum in z we get

$$x \cdot y - \phi(x) \geq \phi^*(y) \text{ and thus by (1.3), } x \cdot y - \phi(x) = \phi^*(y).$$

Using the Legendere duality this is the same as $x \cdot y - \phi^*(y) = \phi^{**}(x)$ and therefore we obtain

$$x \cdot y - \phi^*(y) \geq x \cdot z - \phi^*(z) \qquad \forall z \in \mathbb{R}^n,$$

which can be rewritten as $x \in \partial\phi^*(y)$. $\square$

Second Differentiability

A convex function is twice differentiable almost everywhere in the interior of its domain. This fact is known as Alexandrov's theorem.

If x_0 is a point where ϕ is twice differentiable then $D^2_A(\phi(x_0))$ is invertible if and only if ϕ^* is twice differentiable in $\nabla\phi(x_0)$ and the following formula holds

$$D^2_A\phi^*(\nabla\phi(x)) = (D^2_A\phi(x))^{-1}.$$

The second derivative $D^2_A\phi$ of the convex function ϕ can be interpreted as a rate of volume distortion. More precisely:

Theorem 1.8. *Let ϕ be a convex function which is twice differentiable at x_o. Then*

$$\lim_{r\to 0} \frac{\mathrm{vol_n}[\partial\phi(B_r(x_0))]}{\mathrm{vol_n}[B_r(x_0)]} = \det D^2_A\phi(x_0).$$

If $D^2_A\phi(x_0)$ is invertible, then there is a sequence $r_k \to 0$ and there exist sequences of balls $(B_k)_{k\in\mathbb{N}}$ and $(B'_k)_{k\in\mathbb{N}}$ and a fixed constant $C > 0$ with

$$B_k \subset \partial\phi(B_{r_k}) \subset B'_k \quad \textit{and} \quad \frac{\mathrm{vol_n}(B_k)}{\mathrm{vol_n}(B'_k)} > C > 0.$$

2 Optimal Transport

2.1 Monge's Optimal Transportation Problem

This introduction to Monge's problem will mostly follow [17].

Assume that we are given some goods we want to redistribute in a certain way. For example we may have a pile of sand which we want to use to fill a hole.

Obviously such a process is not going to change the mass of our goods. Without loss of generality we will normalize this mass to 1.

We shall model both the original configuration and the distribution we want to achieve with probability measures on $\mathbb{R}^n$. For any measurable subset A of $\mathbb{R}^n$, $\mu[A]$ will give a measure of how much of our goods is located in A, while $\nu[A]$ will measure the amount that shall be in A.

For any given instance of this problem there may be many ways to achieve such a redistribution. Each of these needs some effort, which will be modeled using a measurable cost function c.

The question is now the following:

How to realize the redistribution with minimal effort.

In more technical terms, given two probability measures μ and ν and a measurable cost function $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ we want to minimize

$$I[T] = \int\limits_X c(x, T(x))d\mu$$

over all $T : \mathbb{R}^n \to \mathbb{R}^n$ with $\nu = T\#\mu$, where $T\#\mu$ denotes the push-forward measure defined by $T\#\mu[X] = \mu[T^{-1}(X)]$.

This question can be ill-posed as there may be no maps T satisfying $\nu = T\#\mu$.
Consider for example the case that μ is a Dirac measure, while ν is not.

On the other hand, even if such maps T do exist, there may not be a unique minimizer for $I[T]$, as shown by the following example.
Let $\mu = \frac{1}{2}(\delta_1 + \delta_2)$ and $\nu = \frac{1}{2}(\delta_2 + \delta_3)$ where δ_a denotes the Dirac measure concentrated

in a.
Obviously, there are two T satisfying the conditions of our problem.

$$T_1(x) = \begin{cases} 2 & x = 1, \\ 3 & x = 2, \end{cases} \qquad T_2(x) = \begin{cases} 3 & x = 1, \\ 2 & x = 2. \end{cases}$$

Let the cost function c be defined by $c(x, y) = |x - y|$. Then a simple calculation shows that $I[T_1] = I[T_2] = 1$.

In the following we will only consider measures μ and ν that do not give measure to small sets, where "small sets" shall be sets with Hausdorff measure zero, and the quadratic cost function $c(x, y) = |x - y|^2$.

As we will see in the following sections, these assumptions will guarantee the existence of a unique solution to Monge's problem.

2.2 Kantorovich's Reformulation of Monge's Problem

As mentioned before, Monge's problem may not always have a solution, for example there may not even be a map T with $\nabla T \# \mu = \nu$.

Therefore we will loosen the restrictions on solutions of our optimal transport problem: We will allow mass from a point $x \in \mathrm{Supp}(\mu)$ to be split up while being transported to $\mathrm{Supp}(\nu)$.

To achieve this a way of transportation will now be modelled by a probability measure π, a so called admissible transference plan, on the product space $\mathbb{R}^n \times \mathbb{R}^n$. Informally, $\pi(x, y)$ measures the amount of mass that is transported from x to y. Obviously for this to make sense, the mass taken from a location x must coincide with $\mu(x)$ and the mass transported to a location y must coincide with $\nu(y)$.

More precisely:

Definition 2.1. *Given two probability measures μ and ν on $\mathbb{R}^n$ an* **admissible transference plan** *is a probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ satisfying*

$$\pi[A \times \mathbb{R}^n] = \mu[A] \text{ and } \pi[\mathbb{R}^n \times B] = \nu[B]$$

for all measurable subsets A and B.
Let $\Pi(\mu, \nu)$ be the set of all such probability measures.

The set $\Pi(\mu, \nu)$ is always nonempty since it contains at least the product measure $\mu \times \nu$.

Given two probability measures μ and ν and a measurable cost function c, Kantorovich's optimal transportation problem is now to minimize

$$I[\pi] = \int\limits_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y) d\pi(x,y)$$

for $\pi \in \Pi(\mu, \nu)$.

Remark 2.1. Let T be a map pushing μ forward to ν. If we define $\pi = (\mathrm{Id} \times T)\#\mu$, then π is an admissible transference plan as

$$\pi[A \times \mathbb{R}^n] = \mu[(\mathrm{Id} \times T)^{-1}(A \times \mathbb{R}^n)] = \mu[A]$$

and

$$\pi[\mathbb{R}^n \times B] = \mu[(\mathrm{Id} \times T)^{-1}(\mathbb{R}^n \times B)] = \mu[T^{-1}(B)] = T\#\mu[B] = \nu[B].$$

Using the push-forward formula we can also calculate

$$I[T] = \int\limits_{\mathbb{R}^n} c(x, T(x)) d\mu(x) = \int\limits_{\mathbb{R}^n} [c \circ (\mathrm{Id} \times T)](x) d\mu(x) = \int\limits_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y) d(\mathrm{Id} \times T)\#\mu = I[\pi].$$

Thus, Kantorovich's problem is indeed a relaxed version of Monge's problem.

For this more general problem a solution always exists in the case of the quadratic cost function. To prove this we will need Prokhorov's theorem:

Theorem 2.1. Prokhorov *Let $(X, \mathcal{T})$ be a topological space and let M be a set of probability measures defined on the Borel σ-algebra of X. Then the following statements are equivalent:*

1. *M is tight, that means for every $\epsilon > 0$ there is a compact set K such that $\pi[X \backslash K] < \epsilon$ for all $\pi \in M$.*

2. *M is sequentially compact in the space of probability measures equipped with the topology of weak convergence.*

A proof of this theorem can be found in [8][Chapter 8].

Theorem 2.2. *For $c = |x-y|^2$ the Kantorovich optimal transportation problem admits a minimizer.*

Proof. As a first step we will show that $\Pi(\mu, \nu)$ is sequentially compact. Let $\epsilon > 0$ be given, and let $K, L \subset \mathbb{R}^n$ be such that

$$\mu[\mathbb{R}^n \backslash K] < \epsilon \quad \text{and} \quad \mu[\mathbb{R}^n \backslash L] < \epsilon.$$

Then

$$\pi[(\mathbb{R}^n \times \mathbb{R}^n) \backslash (K \times L)] \leq \pi[\mathbb{R}^n \times (\mathbb{R}^n \backslash L)] + \pi[(\mathbb{R}^n \backslash K) \times \mathbb{R}^n] = \nu[\mathbb{R}^n \backslash L] + \mu[\mathbb{R}^n \backslash K] < 2\epsilon.$$

So $\Pi(\mu, \nu)$ is tight and therefore sequentially compact.

Let $(\pi_k)_{k\in\mathbb{N}}$ be a minimizing sequence and let π be a cluster point of this sequence. Then π is also an element of $\Pi(\mu, \nu)$. We write the cost function $c = |x-y|^2$ as the supremum of a nondecreasing sequence of continuous bounded functions c_l. Then we obtain, using the monotone convergence theorem,

$$\begin{aligned}\int |x-y|^2 d\pi(x,y) &= \lim_{l\to\infty} \int c_l(x,y) d\pi(x,y) \\ &\leq \lim_{l\to\infty} \limsup_{k\to\infty} \int c_l(x,y) d\pi_k(x,y) \\ &\leq \limsup_{k\to\infty} \int |x-y|^2 d\pi_k(x,y) = \inf_{p\in\Pi(\mu,\nu)} I[p].\end{aligned}$$

So π is a minimizer. □

2.3 The Brenier Map

As we have seen above there is always a solution to Kantorovich's optimal transportation problem for a quadratic cost function.
In this section our aim will be to show that such a solution π has to have a special kind of structure, namely its support has to be included in the graph of the subdifferential of a convex function ϕ on $\mathbb{R}^n$.
As such a convex function is in fact differentiable outside a small set, we will in fact be able to show that such a solution is also a solution in the stronger sense of Monge's transportation problem if the measure μ does not give mass to such sets.

The tool we will use to connect optimal transport with convexity is the notion of cyclical monotonicity:

Definition 2.2. Cyclical monotonicity. *A subset $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone if it satisfies the following condition: For all $m \in \mathbb{N}$ and for all $(x_1, y_1), ..., (x_m, y_m) \in \Gamma$,*

$$\sum_{i=1}^{m} |x_i - y_i|^2 \leq \sum_{i=1}^{m} |x_i - y_{i-1}|^2,$$

where $y_0 = y_m$.

In the following we will prove that the support of an optimal transference plan has to be cyclically monotone and that a cyclically monotone set has to be included in the graph of the subdifferential of a convex function.

Theorem 2.3. *Let μ and ν be two probability measures on $\mathbb{R}^n$ and let $\pi \in \Pi(\mu, \nu)$ be optimal in the Kantorovich problem of mass transference between μ and ν with quadratic cost function $c(x,y) = |x-y|^2$. Then the support of π is cyclically monotone.*

Sketch of proof. Let π be optimal. Assume that the support of π is not cyclically monotone, that is, there are points $(x_1, y_1), ..., (x_m, y_m)$ in the support of π with

$$\sum_{i=1}^{m} |x_i - y_i|^2 > \sum_{i=1}^{m} |x_i - y_{i-1}|^2.$$

Consider balls B_i centered at (x_i, y_i) respectively, each carrying a small mass ϵ of the measure π. Let the translated balls $\tilde{B}_i$ be defined by $B_i + (0, y_{i-1} - y_i)$. Define a measure $\tilde{\pi}$ in the following way:

$$\tilde{\pi}[X] = \pi[X] - \sum \pi[(X \cap B_i)] + \sum \pi[(X \cap \tilde{B}_i) - (0, y_{i-1} - y_i)].$$

This means we shifted the mass carried by the B_i along the second axis.
Let $\tilde{\mu}$ and $\tilde{\nu}$ be the marginals of $\tilde{\pi}$. Obviously $\tilde{\mu}$ is equal to μ as there was no change concerning the first axis. $\tilde{\nu}$ on the other hand is approximately equal to ν, because the measure removed at each y_i by the removal of the measure carried by B_i is compensated by the measure of B_{i+1}. On the other hand calculating

$$\begin{aligned}
\int\limits_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^2 d(\tilde{\pi} - \pi) &= -\sum_{i=1}^{m} \int\limits_{B_i} |x-y|^2 d\pi + \sum_{i=1}^{m} \int\limits_{B_i} |x - (y - y_i + y_{i-1})|^2 d\pi \\
&\approx -\sum_{i=1}^{m} \int\limits_{B_i} |x_i - y_i|^2 d\pi + \sum_{i=1}^{m} \int\limits_{B_i} |x_i - (y_i - y_i + y_{i-1})|^2 d\pi \\
&= \epsilon \left(-\sum_{i=1}^{m} |x_i - y_i|^2 i + \sum_{i=1}^{m} |x_i - y_{i-1}|^2 \right) < 0
\end{aligned}$$

shows that $\tilde{\pi}$ is in fact a better transference plan than π, which we assumed to be optimal.
Of course this proof is not rigorous, because $\tilde{\pi}$ is only nearly a member of $\Pi[\mu, \nu]$. □

Theorem 2.4. Rockafellar's theorem *A nonempty subset $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone if and only if it is included in the (graph of the) subdifferential of a proper convex lower semicontinuous function ϕ on $\mathbb{R}^n$. Maximal cyclically monotone subsets with respect to inclusion are exactly the subdifferentials of such functions.*

Proof. Let ϕ be a convex function and let $(x_i, y_i) \in \text{Graph}(\nabla\phi)$ for $1 \leq i \leq m$, where $m \in \mathbb{N}$. By definition, this means that

$$\phi(z) \geq \phi(x_i) + y_i \cdot (z - x_i)$$

for all $z \in \mathbb{R}^n$, and therefore

$$\begin{aligned}
\phi(x_2) &\geq \phi(x_1) + y_1 \cdot (x_2 - x_1) \\
\phi(x_3) &\geq \phi(x_2) + y_2 \cdot (x_3 - x_2) \\
&\vdots \\
\phi(x_1) &\geq \phi(x_m) + y_m \cdot (x_1 - x_m).
\end{aligned}$$

Summing up these inequalities, we find that

$$\sum \phi(x_i) \geq \sum \phi(x_i) + \sum y_i \cdot (x_{i+1} - x_i)$$

with the convention that $x_{m+1} = x_1$. This is of course equivalent to

$$\begin{aligned} 0 &\geq 2\sum y_i \cdot x_{i+1} - 2\sum y_i \cdot x_i = -\Big(\sum x_{i+1} \cdot x_{i+1} - 2\sum y_i \cdot x_{i+1} + \sum y_i \cdot y_i\Big) \\ &+ \Big(\sum x_i \cdot x_i - 2\sum y_i \cdot x_i + \sum y_i \cdot y_i\Big) = -\sum_{i=1}^{m} |x_i - y_{i-1}|^2 + \sum_{i=1}^{m} |x_i - y_i|^2, \end{aligned}$$

and so $\mathrm{Graph}(\nabla\phi)$ and all of its subsets are cyclically monotone.

Now let $\Gamma \in \mathbb{R}^n \times \mathbb{R}^n$ be cyclically monotone and let $(x_0, y_0) \in \Gamma$. Define a function ϕ by

$$\begin{aligned} \phi(x) \;=\; & \sup\{y_m \cdot (x - x_m) + y_{m-1} \cdot (x_m - x_{m-1}) + ... + y_0 \cdot (x_1 - x_0) \\ & : m \in \mathbb{N} \text{ and } (x_i, y_i) \in \Gamma\}. \end{aligned}$$

As a supremum of affine functions, ϕ is a convex function. Also, $\phi(x_0) \leq 0$ by cyclical monotonicity and so, ϕ is proper.

Now we only have to prove that γ is indeed included in the subdifferential $\mathrm{Graph}(\partial\phi)$: For $(x, y) \in \Gamma$ we have to check that

$$\phi(z) \geq \phi(x) + y \cdot (z - x)$$

for all $z \in \mathbb{R}^n$. It suffices to check that

$$\phi(z) \geq \alpha + y \cdot (z - x) \tag{2.1}$$

for all $\alpha < \phi(x)$. If $\alpha < \phi(x)$, then there exist m and (x_i, y_i) such that

$$\alpha \leq y_m \cdot (x - x_m) + y_{m-1} \cdot (x_m - x_{m-1}) + ... + y_0 \cdot (x_1 - x_0)$$

or equivalently

$$\alpha + y \cdot (z - x) \leq y \cdot (z - x) + y_m \cdot (x - x_m) + y_{m-1} \cdot (x_m - x_{m-1}) + ... + y_0 \cdot (x_1 - x_0).$$

By setting $x = x_{m+1}$ and $y = y_{m+1}$ and applying the definition of ϕ, we obtain (2.1).

□

Combining these two last results we get as a corollary the **Knott-Smith optimality criterion**: If $\pi \in \Pi(\mu, \nu)$ is optimal, then there exists a convex function ϕ such that

$$\mathrm{Supp}(\pi) \subset \mathrm{Graph}(\partial\phi)$$

or equivalently:

$$y \in \partial\phi(x) \text{ for } \pi\text{-almost all } (x, y)$$

Finally we can now prove Brenier's theorem:

Theorem 2.5. *Let μ and ν be probability measures and consider the optimal transportation problem with the quadratic cost function $c(x, y) = |x - y|^2$. If μ does not give mass to small sets, then there is an optimal π given by*

$$\pi = (\mathrm{Id} \times \nabla\phi)\#\mu,$$

where $\nabla\phi$ is the gradient of a convex function which pushes μ forward to ν: $\nabla\phi\#\mu = \nu$.

We shall refer to this mapping $\nabla\phi$ as the Brenier map pushing μ forward to ν.

Proof. We already know that an optimal π exists and that its support is included in the graph of the subdifferential of a function ϕ, or in other words that $y \in \partial\phi(x)$ for π-almost all (x, y). Outside of a Lebesgue null set this means that $y = \nabla\phi(x)$ and as μ does not give mass to such a set this equation holds μ-almost everywhere and as such also π almost everywhere. □

2.4 The Monge-Ampère Equation

Let μ and ν be two probability measures with densities f and g. As we know from the last section there exists a convex function ϕ with $\nabla\phi\#\mu = \nu$. This means that we can use the push-forward formula to obtain

$$\int \zeta(y)g(y)dy = \int \zeta(\nabla\phi(x))f(x)dx.$$

for all bounded continuous functions ζ.

From now on we will assume that $\nabla\phi$ is a differentiable function that is also one-to-one. Then we can use the change of variable formula to obtain

$$\int \zeta(y)g(y)dy = \int \zeta(\nabla\phi(x))g(\nabla\phi(x)) \det D^2(\phi(x))dx.$$

Combining the last two equations we get

$$\int \zeta(\nabla\phi(x))f(x)dx = \int \zeta(\nabla\phi(x))g(\nabla\phi(x)) \det D^2(\phi(x))dx.$$

Since ζ was arbitrary, we obtain the **Monge-Ampère equation**

$$f(x) = g(\nabla\phi(x)) \det D^2(\phi(x))dx$$

This formulation of measure transportation will be very useful for the geometric applications in the remaining chapters.

Our goal in this section is to obtain this equation without assuming smoothness of the transportation map $\nabla\phi$.
Before stating the main result we need one more definition:

Definition 2.3. Hessian measure *The Hessian measure associated with ϕ, denoted by $\det_H D^2\phi$, is defined by*

$$\det{}_H D^2\phi[E] = \mathrm{vol_n}[\partial\phi(E)] = \mathrm{vol_n}\left[\{\partial\phi(x)|x \in E\}\right]$$

for any Borel measurable set $E \subset \mathbb{R}^n$.

Theorem 2.6. The Monge-Ampère equation *Let μ and ν be absolutely continuous measures with densities f and g respectively, and let ϕ be a convex function with $\nabla\phi\#\mu = \nu$.*
Let $\det D_A^2\phi$ be the determinant of the Hessian of ϕ in the Aleksandrov sense, i.e. defined almost everywhere, and let $\det D_A^2\phi dx$ denote the measure with density $\det D_A^2\phi$.
Let M be the set of points where D_A^2 is defined and invertible.

Then the following statements hold:

(i) *M is of full measure for μ and $\partial\phi(M)$ is of full measure for ν.*

(ii) *The measure with density $\det D_A^2(\phi)$ is the absolute continuous part of the Hessian measure $\det_H D^2(\phi)$. It is concentrated on M and satisfies the push-forward formula*

$$\nabla\phi\#[\det D_A^2\phi dx] = \chi_{\partial\phi(M)}dx.$$

(iii) *For almost all $x \in \mathbb{R}^n$ the Monge-Ampère equation*

$$\det D_A^2\phi(x)g(\nabla\phi(x)) = f(x) \tag{2.2}$$

holds.

(iv) *For all nonnegative measurable functions U on $\mathbb{R}_+$ with $U(0) = 0$,*

$$\int_{\mathbb{R}^n} U(g(y))dy = \int_{\mathbb{R}^n} U\left(\frac{f(x)}{\det D_A^2\phi(x)}\right)\det D_A^2\phi(x)dx. \tag{2.3}$$

The main tool in the proof of this theorem will be the density theorem of Lebesgue.

The next result will allow us to identify the absolute continuous part of the Hessian measure $\det_H D_2\phi(x)$.

Theorem 2.7. *Let ϕ be a convex function. Let $\det_H D^2\phi(x)$ denote the Hessian measure associated with ϕ and let $\det D_A^2\phi(x)$ denote the determinant of the Hessian.*
Then $\det D_A^2\phi(x)dx$ is the absolute continuous part of $\det_H D^2\phi(x)$ and $\det_H D^2\phi(x)$ is locally finite.

Proof. It is a consequence of the convexity of ϕ that for any compact set K,

$$\det{}_H D^2\phi[K] = \mathrm{vol_n}[\partial\phi(K)] < \infty,$$

and, therefore, the Hessian measure is locally finite and we may apply Lebesgue's density theorem. Applying Theorem 1.8 we obtain

$$\lim_{r\to 0} \frac{\det_H D^2\phi[B_r(x)]}{\mathrm{vol_n}[B_r(x)]} = \lim_{r\to 0} \frac{\mathrm{vol_n}[\partial\phi(B_r(x))]}{\mathrm{vol_n}[B_r(x)]} = \det D_A^2\phi(x)$$

for almost all x. An application of Lebesgue's density theorem finishes the proof. $\square$

Another tool in the proof will be the following lemma, which expresses the push-forward measure $\nabla\phi\#\mu$ in terms of the subdifferential $\partial\phi$.

Lemma 2.8. *Let ϕ be a convex function and let μ and $\nu = \nabla\phi\#\mu$ be absolutely continuous probability measures. Denote by f and g the respective densities of μ and ν. Then for all Borel sets $A \subset \mathbb{R}^n$,*

$$\nabla\phi\#\mu[A] = \mu[\partial\phi^*(A)] \qquad \textit{and} \qquad \int\limits_{\partial\phi(A)} g(y)dy = \int\limits_A f(x)dx. \tag{2.4}$$

Proof. Since $\nabla\phi\#\mu[A] = \mu[(\nabla\phi)^{-1}(A)]$ by the definition of the push-forward, the first equation is equivalent to $\mu[(\nabla\phi)^{-1}(A)] = \mu[\partial\phi^*(A)]$.
Its a general property of convex functions that $x \in \partial\phi^*(y)$ is equivalent to $y \in \partial\phi(x)$ $\quad(= \{\nabla\phi(x)\}$ if $\nabla\phi(x)$ exists). Therefore $x \in (\nabla\phi)^{-1}(y) \Rightarrow \nabla\phi(x) = y \Rightarrow x \in \partial\phi^*(y)$ and so $(\nabla\phi)^{-1}(A) \subset \partial\phi^*(A)$.
It will suffice to prove that the difference

$$Z = [\partial\phi^*(A)] - [(\nabla\phi)^{-1}(A)]$$

is a set with zero Lebesgue measure, because then,

$$\mu[\partial\phi^*(A)] = \mu[(\nabla\phi)^{-1}(A)] + \mu[Z] = \mu[(\nabla\phi)^{-1}(A)]$$

as $\mu[Z] = 0$ because μ is absolutely continuous.

Let $x \in \partial\phi^*(A)$. Obviously there is a $y \in A$ with $x \in \partial\phi^*(y)$ or equivalently $y \in \partial\phi(x)$. If x is a differentiability point of ϕ then $\partial\phi(x) = \{\nabla\phi(x)\}$ and so $y = \nabla\phi(x)$. Therefore $x \in (\nabla\phi)^{-1}(y) \subset (\nabla\phi)^{-1}(A)$.
Hence Z is included in the set of points where ϕ is not differentiable. As this set has zero Lebesgue measure, Z has zero Lebesgue measure and the proof of the first equation is complete.

To prove the second equation, recall that $\mu = \nabla\phi^*\#\nu$. Therefore, by applying the first equation on ϕ^* instead of ϕ and using the push-forward formula

$$\int\limits_A 1d\mu = \int\limits_{\nabla(\phi^*)^{-1}(A)} 1d\nu = \int\limits_{\partial\phi(A)} 1d\nu,$$

finishing the proof of the second equation as f and g are the densities of μ and ν. $\square$

With these tools at hand we can finally give the proof of the Monge-Ampère equation.

Proof of Theorem 2.6 . Recall from Section 1.2 that $D^2_A\phi(x)$ is invertible if and only if $D^2_A\phi^*(y := \nabla\phi(x))$ is defined.
This implies that if $D^2_A\phi(x)$ is not invertible, then x is included in $\partial\phi^*(y)$, where y is included in the set of points where $D^2_A\phi^*$ does not exist. We denote this set by C. By Alexandrov's theorem C has zero Lebesgue measure.

Using Lemma 2.8 we see that

$$\mu[M^c] \le \mu[\partial\phi^*(C)] = \nabla\phi\#\mu[C] = \nu[C] = 0$$

as we assumed that ν is absolutely continuous.

Obviously this means that $\mu[M] = 1$ and $\nu[\partial\phi(M)] = \nabla\phi^*\#\nu[M] = \mu[M] = 1$ by another application of Lemma 2.8. This finishes the proof of (i).

The first part of (ii) has already been proven in Theorem 2.7. As $\det D^2_A\phi(x)$ is zero outside of M it is clear that $\det D^2_A\phi(x)dx$ is concentrated on M.

To complete the proof of (ii) we will show first that $\nabla\phi\#(\det D^2_A\phi(x)dx)$ is absolutely continuous. Consider a subset A of $\partial\phi(M)$ with zero Lebesgue measure. Then

$$\begin{aligned}
\nabla\phi\#(\det D^2_A\phi(x)dx)[A] &= \det D^2_A\phi(x)dx[(\nabla\phi)^{-1}(A)] \\
&\le \det{}_H D^2\phi(x)dx[(\nabla\phi)^{-1}(A)] = \mathrm{vol_n}[\partial\phi((\nabla\phi)^{-1}(A)] \\
&= \mathrm{vol_n}[\nabla\phi((\nabla\phi)^{-1}(A)] = \mathrm{vol_n}[A] = 0.
\end{aligned}$$

To finish the proof of (ii) we have to show that the density of $\nabla\phi\#(\det D^2_A\phi(x)dx)$ is equal to one almost everywhere on $\partial\phi(M)$. To do this we use Lebesgue's density theorem again:
Let $y \in \partial\phi(M)$. Then there is an $x \in M$ with $\nabla\phi(x) = y$. Recall from Section 1.2 that

$$\lim_{r\to 0} \frac{\mathrm{vol_n}[\partial\phi^*(B_r(y))]}{\mathrm{vol_n}[B_r(y)]} = \det D^2_A\phi^*(y) = \det D^2_A\phi^*(\nabla\phi(x)) = \frac{1}{\det D^2_A\phi(x)}.$$

Using the second part of Theorem 1.8 we see that there is a sequence $r_k \to 0$ such that we can apply Lebesgue's density theorem on the measure $\det D^2_A\phi(x)dx$ and the sets $C_k = \partial\phi^*(B_r(y))$. Doing this,

$$\lim_{k\to\infty} \frac{\det D^2_A\phi(x)dx[\partial\phi^*(B_r(y))]}{\mathrm{vol_n}[\partial\phi^*(B_r(y))]} = \det D^2_A\phi(x).$$

Multiplying these two limits and using Lemma 2.8 again, we find

$$\lim_{k\to\infty} \frac{\det D^2_A\phi(x)dx[\partial\phi^*(B_r(y))]}{\mathrm{vol_n}[B_r(y)]} = \lim_{k\to\infty} \frac{\nabla\phi\#(\det D^2_A\phi(x)dx)[B_r(y)]}{\mathrm{vol_n}[B_r(y)]} = 1.$$

Using Lebesgue's density theorem finishes the proof.

Let $A \subset \mathbb{R}^n$ be a Borel set. Using (ii) and the push-forward formula, we obtain

$$\int\limits_{\partial\phi(A)} g(y)dy = \int\limits_{\partial\phi(M)} \chi_{\partial\phi(A)}(y)g(y)dy = \int\limits_{M} \chi_{\partial\phi(A)}(\nabla\phi(x))g(\nabla\phi(x)) \det D_A^2\phi(x)dx.$$

For the integrand to be non zero, $\nabla\phi(x)$ has to be an element of $\partial\phi(A)$. This means there is a $y \in A$ with $y \in \partial\phi^*(\nabla\phi(x))$. As x is an element of M, the second derivative $D_A^2\phi(x)$ is invertible, and so ϕ^* is actually twice differentiable at $\nabla\phi(x)$ and therefore

$$y \in \partial\phi^*(\nabla\phi(x)) = \{\nabla\phi^*(\nabla\phi(x))\} = \{x\}.$$

So we find that $x = y$ and $x \in A$. If, on the other hand, we assume that $x \in A$, then clearly

$$\{\nabla\phi(x)\} = \partial\phi(x) \subset \partial\phi(A).$$

Thus, we see that $x \in A$ and $\nabla\phi(x) \in \partial\phi(A)$ are equivalent and, therefore, we can write

$$\int\limits_{\partial\phi(A)} g(y)dy = \int\limits_{M} \chi_A(x)g(\nabla\phi(x)) \det D_A^2\phi(x)dx.$$

Using the fact that $\det D_A^2\phi(x)dx$ is concentrated on M we deduce

$$\int\limits_{\partial\phi(A)} g(y)dy = \int\limits_{A} g(\nabla\phi(x)) \det D_A^2\phi(x)dx.$$

By Lemma 2.8 part (ii), we find that

$$\int\limits_{A} f(x)dx = \int\limits_{A} g(\nabla\phi(x)) \det D_A^2\phi(x)dx.$$

Since A is an arbitrary measurable set, this concludes the proof of (iii).

From (ii) we know that

$$\int\limits_{\partial\phi(M)} U(g(y))dy = \int\limits_{M} U(g(\nabla\phi(x)) \det D_A^2\phi(x)dx$$

and by (iii) we can write

$$g(\nabla\phi(x)) = \frac{f(x)}{\det D_A^2\phi(x)}$$

for almost all $x \in M$. Combining these two statements we find that

$$\int\limits_{\partial\phi(M)} U(g(y))dy = \int\limits_{M} U\left(\frac{f(x)}{\det D_A^2\phi(x)}\right) \det D_A^2\phi(x)dx.$$

Since $\det D_A^2\phi(x)dx$ is concentrated on M, we can extend the integral on the right–hand side to $\mathbb{R}^n$. Since $\nu = g(x)dx$ is concentrated on $\partial\phi(M)$, we know that $g(z) = 0$ for almost all $z \in (\partial\phi(M))^c$. Therefore $U(g(z)) = 0$ almost everywhere outside of $\partial\phi(M)$ and we can also extend the left–hand side integral. $\square$

3 The Brunn-Minkowski and Prèkopa-Leindler Inequality

3.1 The Brunn-Minkowski Inequality

Let A and B be subsets of $\mathbb{R}^n$. Then the Minkowski sum $A+B$ is defined as the set $\{a+b|a \in A, b \in B\}$. The Brunn-Minkowski inequality gives a lower bound for the volume of the Minkowski sum of two sets.

Theorem 3.1. The Brunn-Minkowski Inequality *Let A and B be non-empty compact sets in $\mathbb{R}^n$. Then the following inequality holds*

$$\mathrm{vol_n}[(1-t)\cdot A + t\cdot B]^{\frac{1}{n}} \geq (1-t)\cdot \mathrm{vol_n}[A]^{\frac{1}{n}} + t\cdot \mathrm{vol_n}[B]^{\frac{1}{n}}$$

for all $t \in [0,1]$.

Remark 3.1. The compactness of A and B implies that $A+B$ is also a compact set and therefore a Borel set.

An eloborate introduction to this topic is given in [10].

First we will give an equivalent, dimension free formulation of the Brunn-Minkowski inequality. Although this inequality is formally weaker than the Brunn-Minkowski inequality for fixed A, B and t, it is equivalent to Theorem 3.1 if it holds for all A, B and t.

Theorem 3.2. The Multiplicative Version of the Brunn-Minkowski Inequality *Let A and B be non-empty compact Sets in $\mathbb{R}^n$. Then the following inequality holds*

$$\mathrm{vol_n}[(1-t)\cdot A + t\cdot B] \geq \mathrm{vol_n}[A]^{1-t}\cdot \mathrm{vol_n}[B]^{t} \tag{3.1}$$

for all $t \in [0,1]$.

Lemma 3.3. *The Brunn-Minkowski inequality holds if and only if Theorem 3.2 holds.*

Proof of Lemma 3.3. We can show that the Brunn-Minkowski inequality implies its multiplicative form by a simple application of the arithmetic/geometric mean inequality:

$$\mathrm{vol_n}[(1-t)\cdot A + t\cdot B]^{\frac{1}{n}} \geq (1-t)\cdot \mathrm{vol_n}[A]^{\frac{1}{n}} + t\cdot \mathrm{vol_n}[B]^{\frac{1}{n}} \geq \mathrm{vol_n}[A]^{\frac{1-t}{n}}\cdot \mathrm{vol_n}[B]^{\frac{t}{n}}$$

which is obviously equivalent to inequality (3.1)

To proof the other implication let $A' = \mathrm{vol_n}[A]^{-\frac{1}{n}} \cdot A$ and $B' = \mathrm{vol_n}[B]^{-\frac{1}{n}} \cdot B$. Then by (3.1)

$$\mathrm{vol_n}[(1-t)A' + tB'] \geq \mathrm{vol_n}[A']^{1-t}\,\mathrm{vol_n}[B']^{t} = 1 \tag{3.2}$$

for every $t \in [0,1]$ as the volume of A' and B' is 1. Let $t \in [0,1]$ and let

$$t' = \frac{t \cdot \mathrm{vol_n}[B]^{\frac{1}{n}}}{(1-t)\cdot \mathrm{vol_n}[A]^{\frac{1}{n}} + t \cdot \mathrm{vol_n}[B]^{\frac{1}{n}}}.$$

Then we get

$$(1-t')\cdot A' + t' \cdot B' = \frac{(1-t)\cdot A + t \cdot B}{(1-t)\cdot \mathrm{vol_n}[A]^{\frac{1}{n}} + t \cdot \mathrm{vol_n}[B]^{\frac{1}{n}}}.$$

If we take the volume of both sides of this equation and use (3.2) we see that

$$1 \leq \mathrm{vol_n}\left(\frac{(1-t)\cdot A + t \cdot B}{(1-t)\cdot \mathrm{vol_n}[A]^{\frac{1}{n}} + t \cdot \mathrm{vol_n}[B]^{\frac{1}{n}}}\right)^{\frac{1}{n}}$$

and finally the n-homogeneity of the n-dimensional volume implies the Brunn-Minkowski inequality. □

An important implication of the Brunn-Minkowski inequality is the isoperimetric inequality.

Theorem 3.4. *Let X be a compact set and let B_n denote the n-dimensional unit Ball. Then*

$$\left(\frac{\mathrm{vol_{n\text{-}1}}[\partial X]}{\mathrm{vol_{n\text{-}1}}[\partial B_n]}\right)^{\frac{1}{n-1}} \geq \left(\frac{\mathrm{vol_n}[X]}{\mathrm{vol_n}[B_n]}\right)^{\frac{1}{n}}.$$

In other words, among all compact sets the sphere has the highest volume to surface area ratio.

Proof. Let B_ϵ denote a Ball with radius $\epsilon > 0$ centered at 0. Then we can calculate

$$\mathrm{vol_{n\text{-}1}}[\partial X] = \liminf_{\epsilon \to 0} \frac{\mathrm{vol_n}[X + B_\epsilon] - \mathrm{vol_n}[X]}{\epsilon}. \tag{3.3}$$

Applying the Brunn-Minkowski inequality with $Y = B_n$ we find

$$\mathrm{vol_n}[X + B_\epsilon]^{\frac{1}{n}} - \mathrm{vol_n}[X]^{\frac{1}{n}} \geq \mathrm{vol_n}[B_\epsilon]^{\frac{1}{n}} = \epsilon\, \mathrm{vol_n}[B_n]^{\frac{1}{n}}. \tag{3.4}$$

Dividing this by ϵ and passing to the lim inf on the left hand side we obtain

$$\begin{aligned}\mathrm{vol_n}[B_n]^{\frac{1}{n}} &\leq \liminf_{\epsilon \to 0} \frac{\mathrm{vol_n}[X + B_\epsilon]^{\frac{1}{n}} - \mathrm{vol_n}[X]^{\frac{1}{n}}}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{\mathrm{vol_n}[X + B_\epsilon]^{\frac{1}{n}} - \mathrm{vol_n}[X]^{\frac{1}{n}}}{\mathrm{vol_n}[X + B_\epsilon] - \mathrm{vol_n}[X]} \cdot \liminf_{\epsilon \to 0} \frac{\mathrm{vol_n}[X + B_\epsilon] - \mathrm{vol_n}[X]}{\epsilon}.\end{aligned}$$

Using now the fact that the left factor is a differential quotient and (3.3) we further get

$$\mathrm{vol_n}[B_n]^{\frac{1}{n}} \leq \frac{1}{n} \mathrm{vol_n}[X]^{\frac{1}{n}-1} \mathrm{vol_{n\text{-}1}}[\partial X].$$

With $\mathrm{vol_n}[B_n] = \frac{1}{n} \mathrm{vol_{n\text{-}1}}[\partial B_n]$ we finally see that

$$\mathrm{vol_n}[B_n]^{\frac{1}{n}-1} \mathrm{vol_{n\text{-}1}}[\partial B_n] \leq \mathrm{vol_n}[X]^{\frac{1}{n}-1} \mathrm{vol_{n\text{-}1}}[\partial X].$$

thus finishing the proof. □

There are many known proofs of the Brunn-Minkowski inequality. In the following sections we will give two proofs using optimal transport, the first one can be found in [2], the second one in [17][Chapter 5].

3.1.1 A First Proof Using Optimal Transport

Using the Brenier map we can give a first proof of the Brunn-Minkowski inequality. It is noteworthy that the proof does not depend on the optimality of the Brenier map in the sense of the Monge-Kantorovich problem, but only on its analytic properties.

A first proof of the Brunn-Minkowski inequality. Without loss of generality we may assume that both A and B have finite non-zero measure. Let μ and ν be the uniform probability measures on A and B, respectively. Let T be the Brenier map transporting μ to ν and let T_t be the map given by

$$x \mapsto (1-t)x + tT(x).$$

This map transports the measure μ to a probability measure supported on $(1-t)A+tB$. Let f_t be the density of this measure.

If we could prove that f_t is bounded from above almost everywhere, say by a constant C, than $C \cdot \mathrm{vol_n}((1-t)A+tB)$ would have to be greater or equal than 1 because f_t is a probability measure. Of course we can also write this as

$$\mathrm{vol_n}((1-t)A+tB) \geq \frac{1}{C}. \tag{3.5}$$

In particular, to prove the multiplicative version of the Brunn-Minkowski inequality it is now sufficient to show that we can choose C to be $\frac{1}{\mathrm{vol_n}(A)^{1-t}\,\mathrm{vol_n}(B)^t}$.

By Theorem 2.6, the densities $\frac{1}{\mathrm{vol_n}(A)}, \frac{1}{\mathrm{vol_n}(B)}, f_t$ satisfy the Monge-Ampère equations

$$\begin{aligned} \frac{1}{\mathrm{vol_n}(A)} &= f_t(T_t(x)) \cdot \det(T_t'(x)), && (3.6)\\ \frac{1}{\mathrm{vol_n}(A)} &= \frac{1}{\mathrm{vol_n}(B)} \cdot \det(T'(x)) && (3.7) \end{aligned}$$

almost everywhere.

Using these, the inequality (3.5) with $C = \frac{1}{\mathrm{vol_n}(A)^{1-t}\,\mathrm{vol_n}(B)^t}$ will follow from

$$\det(T_t'(x)) \geq \left(\frac{vol(B)}{vol(A)}\right)^t = (\det(T'(x))^t$$

which, simply by the definition of f_t, is equivalent to

$$\det((1-t)I + tT'(x)) \geq (\det(T'(x))^t. \tag{3.8}$$

As $T'(x)$ is the Hessian of a convex function by Brenier's Theorem 2.5, and therefore symmetric and nonnegative, it is a diagonal matrix with respect to an appropriate orthonormal basis, say $\mathrm{diag}(t_1, t_2, ..., t_n)$ with $t_i > 0$. Therefore we can rewrite the last inequality as

$$\prod_{i=1}^{n}(1 - t + t \cdot t_i) \geq \prod_{i=1}^{n} t_i^t.$$

But this follows directly from the arithmetic/geometric mean inequality, because we can obtain $1 - t + t \cdot t_i \geq t_i^t$ for each t_i. $\square$

3.1.2 A Second Proof by Displacement Convexity

Displacement Convexity

Our formulation of the Monge problem depends only on the starting and the end point of the transportation process. One can also take the whole history of the transportation process into account:
To each point $x \in \mathbb{R}^n$ we will associate a trajectory $(T_t x)$ for $t \in [0,1]$ and a corresponding cost $C(T_t x)$. This leads to the following time dependent reformulation of Monge's problem:

Time Dependent Version of Monge's Problem

We want to minimize

$$I^*[T_t] = \int_X C(T_t x) d\mu,$$

where $T_0 = \mathrm{Id}$ and $T_1 \# \mu = \nu$.

As we considered only the quadratic cost function $c(x,y) = |x-y|^2$ in the original formulation, we will only consider the quadratic cost $C(T_t x) = \int_0^1 |\dot{T}_t x|^2 dt$ in this problem.

These two problems are then equivalent in the sense that a solution T_t for the time dependent problem gives rise to a solution T of the original problem T, via $T = T_1$.

Proof. Jensen's inequality implies that

$$C(T_t x) = \int_0^1 |\dot{T}_t x|^2 dt \geq \left| \int_0^1 \dot{T}_t x dt \right|^2 = |T_0(x) - T_1(x)|^2 = |x - T_1(x)|^2 = c(x, T_1(x))$$

and for $T_t(x) = tx + (1-t)y$ we get $C(T_t(x)) = \int_0^1 |x-y|^2 dx = c(x,y)$.

Therefore an optimizer T_t of the time dependent problem has to be of the form $t\,\mathrm{Id} + (1-t)T$ for some T with $T\#\mu = \nu$ and for such an optimizer $I^*[T_t] = I[T]$ holds.

Assume that $T_1 = T$ is not an optimizer of the original problem. This would imply the existence of a map T' with $I^*[\mathrm{Id} + (1-t)T'] = I[T'] < I[T] = I[T_t]$. As T_t was an optimizer this is a contradiction. $\square$

These arguments together with the existence of the Brenier map imply that McCann's interpolation as defined below is the solution of our time dependent problem.

Definition 3.1. McCann's Interpolation *Let μ and ν be two probability measures that do not give mass to small sets. Let $\nabla\phi$ be the transportation map with $\nabla\phi\#\mu = \nu$. Define*

$$\rho_t = [\mu, \nu]_t = ((1-t)\,\mathrm{Id} + t\nabla\phi)\#\mu.$$

This family of probability measures interpolates between μ and ν and $[\mu,\nu]_0 = \mu$ and $[\mu,\nu]_1 = \nu$.
If F is a functional on the space of probability measures $\mathcal{P}$ then we can study its behavior on $[\mu,\nu]_t$ as t varies in $[0,1]$. Is, for example, $F([\mu,\nu]_t)$ convex as a function of t?

Definition 3.2. *A functional F on $\mathcal{P}$ is called displacement convex if for all ρ_0, ρ_1 in $\mathcal{P}$, $t \to F(\rho_t)$ is convex on $[0,1]$ where $\rho_t = [\rho_0, \rho_1]$.*

The following Lemma will show the displacement convexity of a class of functionals.

Lemma 3.5. *Let $\mathcal{U}(\rho) = \int_{\mathbb{R}^n} U(\rho(x))dx$ for U measurable $\mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$. Let Ψ be defined by $\Psi : r \mapsto r^n \cdot U(r^{-n})$.*

Then $\mathcal{U}$ is displacement convex if $U(0) = 0$ and Ψ is convex nonincreasing on $(0, +\infty)$.

Before we give the proof of this lemma, we will prove a result about the convexity of composite functions.

Lemma 3.6. *Let f be concave function and let g be a convex nonincreasing function. Then $g \circ f$ is also convex.*

Proof. As f is concave the following inequality holds for all $\lambda \in [0,1]$:

$$f(\lambda x + (1-\lambda y)) \geq \lambda f(x) + (1 - \lambda f(y)).$$

Using that g is nonincreasing and convex we obtain that

$$g(f(\lambda x + (1-\lambda y))) \leq g(\lambda f(x) + (1-\lambda f(y))) \leq \lambda g(f(x)) + (1 - \lambda g(f(y)))$$

which completes the proof. $\square$

Proof of Lemma 3.5. Let μ and ν be probability densities. We consider the interpolant $\rho_t = [\mu, \nu] = [(1-t)\operatorname{Id} + t\nabla\phi]\#\mu$.

As a consequence of 2.6(iv), we can write

$$\mathcal{U}(\rho_t) = \int_{\mathbb{R}^n} U\left(\frac{\rho(x)}{\det(I_n - t(I_n - D_A^2\phi(x)))}\right) \det(I_n - t(I_n - D_A^2\phi(x)))dx.$$

The integrand can be written as a composition of the two following maps.

$$t \mapsto r = \det(I_n - tS)^{\frac{1}{n}}$$
$$r \mapsto r^n \cdot U(r^{-n}\rho(x))$$

where $S = I_n - D_A^2\phi$.

As the first mapping is concave by Theorem 1.1(iii) and the second nonincreasing convex by our assumptions, their composition is – by Lemma (3.6) – convex and so $\mathcal{U}(\rho)$ is displacement convex.

$\square$

Remark 3.2. In the following proof of the Brunn-Minkowski Inequality we will use the following functional

$$\mathcal{U}(\rho) = -\int_{\mathbb{R}^n} \rho(x)^{1-\frac{1}{n}} dx.$$

This is of course a special case of the functionals we were looking at in Lemma 3.5 with $U(r) = -r^{1-\frac{1}{n}}$. It is therefore displacement convex as $\Psi(r) = r^n \cdot -r^{-n(1-\frac{1}{n})} = -r$ is obviously convex and nonincreasing.

Another property of this functional which we will use in the proof is the following:

Let μ be the uniform probability measure on X. Then

$$\mathcal{U}(\mu) = -\int\limits_{\mathbb{R}^n} \mu^{1-\frac{1}{n}} dx = -\int\limits_{X} 1/\operatorname{vol_n}(X)^{1-1/n} dx = -\operatorname{vol_n}(X)^{\frac{1}{n}}.$$

Proof of the Brunn-Minkowski inequality. Let $\mu = \rho_0$ and $\nu = \rho_1$ be the uniform probability measures on A and B, respectively. Then the interpolant $\rho_t = [\mu, \nu]_t$ has its support included in the set $(1-t)A + tB$ for all $t \in [0,1]$.

Let $\lambda \in (0,1)$ and let S_t be the support of the interpolant ρ_t. Obviously, $\frac{\chi_{S_t}}{\operatorname{vol_n}(S_t)}$ is the density of a probability measure. Therefore, using Jensen's inequality:

$$\begin{aligned}\mathcal{U}(\rho_t) &= \int_{S_t} U(\rho_t) dx = \operatorname{vol_n}(S_t) \int_{S_t} U(\rho_t) \frac{dx}{\operatorname{vol_n}(S_t)} \geq \operatorname{vol_n}(S_t) U\left(\frac{1}{\operatorname{vol_n}(S_t)} \int_{S_t} \rho_t dx\right) \\ &\geq \operatorname{vol_n}(S_t) U\left(\frac{1}{\operatorname{vol_n}(S_t)}\right) = -\operatorname{vol_n}(S_t)^{\frac{1}{n}}.\end{aligned}$$

As S_t is a subset of $(1-t)A + tB$, the inequality $-\operatorname{vol_n}(S_t)^{\frac{1}{n}} \geq -\operatorname{vol_n}((1-t)A + tB)^{\frac{1}{n}}$ holds.

Finally the displacement convexity of $\mathcal{U}$ and Remark 3.2 imply that

$$-\operatorname{vol_n}(S_t)^{\frac{1}{n}} \leq \mathcal{U}(\rho_t) \leq (1-t) \cdot \mathcal{U}(\rho_0) + t \cdot \mathcal{U}(\rho_1) = -(1-t) \cdot \operatorname{vol_n}(A)^{\frac{1}{n}} - t \cdot \operatorname{vol_n}(B)^{\frac{1}{n}}.$$

If we put the last two statements together we finally get

$$\operatorname{vol_n}((1-t)A + tB)^{\frac{1}{n}} \geq (1-t) \cdot \operatorname{vol_n}(A)^{\frac{1}{n}} + t \cdot \operatorname{vol_n}(B)^{\frac{1}{n}}.$$

□

3.2 The Prèkopa-Leindler Inequality

Another important inequality, functional in nature, that can be proved using the tools optimal transport provides, is the Prèkopa-Leindler inequality. As shown below, it can be seen as a more flexible variant of the Brunn-Minkowski inequality.

Theorem 3.7. *Let f,g, and h be nonnegative integrable functions on $\mathbb{R}^n$ and let $t \in [0,1]$. Assume that for all $x, y \in \mathbb{R}^n$,*

$$h((1-t)x + ty) \geq f(x)^{1-t} \cdot g(y)^t. \tag{3.9}$$

Then the Prèkopa-Leindler inequality holds:

$$\int_{\mathbb{R}^n} h dx \geq \left(\int_{\mathbb{R}^n} f dx\right)^{1-t} \cdot \left(\int_{\mathbb{R}^n} g dx\right)^t. \tag{3.10}$$

Remark 3.3. Of course, given f,g one can choose

$$h(z) = \sup_{z=(1-t)x+ty} f(x)^{1-t} g(y)^t$$

to fullfill the assumption (3.9).

Remark 3.4. The Prèkopa-Leindler inequality is a functional version of the Brunn-Minkowski inequality. In fact, we can regain the Brunn-Minkowski inequality if we use the Prèkopa-Leindler inequality for the characteristic functions of A and B.

Proof. Let f and g be the characteristic functions of A and B, respectively. Let h be the characteristic function of $(1-t)\cdot A + t\cdot B$.
Since f and g are characteristic functions, the right hand side of inequality (3.9) only takes two values, zero and one.
If it is equal to zero the inequality obviously holds.

On the other hand, if both $f(x)$ and $g(y)$ are equal to one, in other words if $x \in A$ and $y \in B$, then, clearly, $(1-t)x + ty$ is an element of their Minkowski sum, and therefore the left hand side of (3.9) is also equal to one.

Therefore we can use the Prèkopa-Leindler inequality to obtain

$$\mathrm{vol_n}[(1-t)\cdot A + t\cdot B] \geq \mathrm{vol_n}[A]^{1-t} \cdot \mathrm{vol_n}[B]^t$$

which is exaxtly the multiplicative Brunn-Minkowski inequality. □

There are many classical proofs of the Prèkopa-Leindler inequality, in the following, two proofs using optimal transport will be given. The first one can be found in [13], the second one in [3].

First proof of the Prèkopa-Leindler inequality. Without loss of generality we may assume that $\int f = \int g = 1$ and then proof that $\int h \geq 1$.

Let μ and ν be the probability measures with densities f and g, respectively. We consider the displacement interpolants $\rho_t = [\mu, \nu]_t$. Since $\int \rho_t = 1$, it will be sufficient to show that $\rho_t \leq h$ almost everywhere for all $t \in [0,1]$.

Let ϕ be a convex function with $\nabla\phi\#\mu = \nu$. Then $\rho_t = [(1-t)\cdot \mathrm{Id} + t\cdot\nabla\phi]\,\#\mu$ and, with Theorem 2.6, the Monge-Ampère equation

$$f(x) = \rho_t((1-t)x + t\nabla\phi(x)) \det((1-t)I_n + tD^2_A\phi(x)) \tag{3.11}$$

holds almost everywhere.
As a special case of (3.11), $f(x) = g(\nabla\phi(x))\det(D^2_A\phi(x))$. Since f is the density of μ,

it is obviously positive μ-almost everywhere and therefore μ-almost everywhere we can write

$$\det(D_A^2\phi(x)) = \frac{f(x)}{g(\nabla\phi(x))}.$$

When we consider the Jacobi determinant in equation (3.11), we see, using the arithmetic/geometric inequality, that

$$\det((1-t)I_n + tD_A^2\phi(x))^{\frac{1}{n}} \geq (1-t)(\det I_n)^{\frac{1}{n}} + t(\det(D_A^2\phi))^{\frac{1}{n}} = (1-t) + t\cdot\left(\frac{f(x)}{g(\nabla\phi(x))}\right)^{\frac{1}{n}}.$$

Using equation (3.11) again, we see that

$$\begin{aligned}
\rho_t((1-t)x + t\nabla\phi(x)) &= \frac{f(x)}{\det((1-t)I_n + tD_A^2\phi(x))} \\
&\leq \frac{f(x)}{\left((1-t) + t\cdot\left(\frac{f(x)}{g(\nabla\phi(x))}\right)^{\frac{1}{n}}\right)^n} \\
&= \left[(1-t)f(x)^{-\frac{1}{n}} + t\cdot g(\nabla\phi(x))^{-\frac{1}{n}}\right]^n \\
&= f(x)^{1-t}\cdot g(\nabla\phi(x))^t
\end{aligned}$$

for μ-almost all x.
As both sides are equal to zero for all x with $f(x) = 0$ this inequality in fact holds even for λ-almost all x.
Since we assumed that $h((1-t)x + ty) \geq f(x)^{1-t}\cdot g(y)^t$ for all x, y, the inequality $\rho_t((1-t)x + t\nabla\phi(x)) \leq h((1-t)x + t\nabla\phi(x))$ obviously holds almost everywhere.

As the displacement interpolants ρ_t are supported in $[(1-t)x + t\nabla\phi(x)]\,(\mathbb{R}^n)$, we have finally shown that $\rho_t \leq h$ almost everywhere. □

Now we will look at yet another proof of this inequality. The main difference is the following: While we were interpolating between the measures defined by f and g in the first proof, we will now introduce a third (arbitrary) measure and interpolate between this new measure, and the two measures defined by f and g.

Second proof of the Prèkopa-Leindler inequality. Without loss of generality we may assume that $\int f = \int g = 1$ and then proof that $\int h \geq 1$.

Let μ and ν be the probabibity measures with densities f and g, respectively.
Let p be the uniform probability measure on the unit cube $[0,1]^n$ and let ϕ_1 and ϕ_2 be the respective Brenier maps transporting p to μ and ν.

Define $\phi = (1-t)\phi_1 + t\phi_2$. Then

$$\begin{aligned}
\int_{\mathbb{R}^n} h &\geq \int_{\nabla\phi([0,1]^n)} h = \int_{[0,1]^n} h(\nabla\phi(x)) \det(D_A^2\phi(x))dx \\
&= \int_{[0,1]^n} h((1-t)\nabla\phi_1(x) + t\nabla\phi_2(x)) \det((1-t)D_A^2\phi_1(x) + tD_A^2\phi_2(x))dx \\
&\geq \int_{[0,1]^n} h((1-t)\nabla\phi_1(x) + t\nabla\phi_2(x)) \det(D_A^2\phi_1(x))^{1-t} \det(D_A^2\phi_2(x))^t dx \\
&\geq \int_{[0,1]^n} f(\nabla\phi_1(x))^{1-t} \det(D_A^2\phi_1(x))^{1-t} g(\nabla\phi_2(x))^t \det(D_A^2\phi_2(x))^t dx.
\end{aligned}$$

Since the following two Monge-Ampère equations hold

$$f(\nabla\phi_1(x)) \det(D_A^2\phi_1(x)) = 1, \qquad g(\nabla\phi_2(x)) \det(D_A^2\phi_2(x)) = 1,$$

we finally see that

$$\int_{\mathbb{R}^n} h \geq \int_{[0,1]^n} 1dx = 1.$$

$\square$

4 The Minkowski Inequality

In this chapter we will further study the volume $\mathrm{vol_n}[K + \lambda T]$ for convex bodies K and T. More precisely, we will show how optimal transport can be used to describe the linear combination $K + \lambda T$ in a different way, using a diffeomorphism which was first introduced by S. Alesker, S. Dar and V. Milman in [1].
We will then use this method to introduce the notion of mixed volumes, and to give a proof, also originally found in [1], for the famous Minkowski inequality, which gives a lower bound for these mixed volumes.

4.1 The Alesker-Dar-Milman Diffeomorphism

When studying the Minkowski sum, the following question arises: Which geometrical and topological properties of X and Y are inherited by their sum $X + Y$?
Simple examples show, that, in fact, this question is very complicated. For example, the sum of two simply connected sets is not necessarily simply connected. (Consider in $\mathbb{R}^2$ the unit ball and the broken annulus $\{(x, y) \in \mathbb{R}^2 \,|9 < x^2 + y^2 < 10, y < 0 \text{ or } |x| > \frac{1}{4}\}$. While both are clearly simply connected, their Minkowski sum is not.)
The following theorem by S. Alesker et al [1] shows that, when X and Y are bounded, convex and open, such problems are avoided, and $X + Y$ is in fact diffeomorph to both X and Y.
More precisely, the following statement holds:

Theorem 4.1 (Alesker-Dar-Milman theorem)**.** *Let K and T be open convex bounded subsets of $\mathbb{R}^n$ of volume 1. Then there exists a C^1-diffeomorphism Ψ, preserving the Lebesgue measure, such that*

$$\{x + \lambda\Psi(x) | x \in K\} = K + \lambda T$$

holds for any $\lambda > 0$.

Before the proof we state – without proof – two results which will be needed.

The first result, by Cafarelli, shows that for a certain class of measures μ and ν the Brenier map $\nabla\phi$ transporting μ to ν is in fact not only differentiable allmost everywhere, but a C^1-smooth function. This and other regularity results can be found in [5] and [6].

Theorem 4.2 (Cafarrelli's regularity theorem for optimal transportation)**.** *Let Y be a bounded convex open set. Let f and g be locally Hölder continuous probability densities defined on $\mathbb{R}^n$ and Y, respectively. Assume that*

1. *f is locally bounded and bounded away from zero on compact sets, i.e. there are constants c_r and C_r with*

$$0 < c_r \leq f(x) \leq C_r \qquad \text{for } |x| < r$$

for each $r > 0$.

2. *g is bounded.*

Let ϕ be the unique Brenier solution of the Monge-Ampère equation

$$\det D^2\phi(x) = \frac{f(x)}{g(\nabla\phi(x))}.$$

Then ϕ is C^2-smooth.

The second tool we will need in the proof is the following result about the images of smooth convex functions, a proof for which can be found in [11].

Theorem 4.3. *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be C^2-smooth convex functions with strictly positive Hessian. Then the following statements hold:*

1. *The image of ∇f is open and convex.*
2. $(\nabla f + \nabla g)(\mathbb{R}^n) = \nabla f(\mathbb{R}^n) + \nabla g(\mathbb{R}^n)$.

Proof of Theorem (4.1). Let ρ be the standard Gaussian measure and denote by μ and ν the uniform probability measures on K and T, respectively. Consider the Brenier maps

$\nabla\phi_1 : \mathbb{R}^n \to K$, transporting ρ to μ, and

$\nabla\phi_2 : \mathbb{R}^n \to T$, transporting ρ to ν.

Note that these are C^1-smooth by Theorem 4.2. By Theorem 4.3

$$(\nabla\phi_1 + \lambda\nabla\phi_2)(\mathbb{R}^n) = K + \lambda T.$$

Setting $\Psi = \nabla\phi_2 \circ (\nabla\phi_1)^{-1}$ finishes the proof. $\square$

4.2 Mixed Volumes

A classical result of convex geometry states that the volume of a linear combination of convex bodies – in the sense of the Minkowski sum– is in fact a polynomial in the coefficients of the linear combination. These coefficients are called mixed volumes. We will prove this fact using the above method. A more elementary proof as well as much more about this topic can be found in [12].

Theorem 4.4 (Minkowski's theorem on mixed volumes)**.** *Let $K_1, ..., K_m$ be convex bodies and let $\lambda_1, ..., \lambda_m$ be nonnegative real numbers. Then there are coefficients $V(K_{i_1}, ..., K_{i_m})$, $1 \leq i_1, ..., i_m \leq m$, symmetric in their arguments, such that*

$$\mathrm{vol_n}\left[\sum_{i=1}^{m} \lambda_i K_i\right] = \sum_{i_1,...,i_d=1}^{m} V(K_{i_1}, ..., K_{i_m})\lambda_{i_1}...\lambda_{i_n}.$$

Proof. Let μ_i be the uniform probability measure on K_i and let ρ denote the standard Gaussian measure. Consider the Brenier maps $\phi_i : \mathbb{R}^n \to K_i$, transporting ρ to μ_i.
By Theorem 4.2 and Theorem 4.3 we obtain

$$\left(\sum_{i=1}^{m} \lambda_i \nabla\phi_i\right)(\mathbb{R}^n) = \sum_{i=1}^{m} \lambda_i K_i,$$

and using the change of variable formula we find:

$$\mathrm{vol_n}\left[\sum_{i=1}^{m} \lambda_i K_i\right] = \int_{\mathbb{R}^n} 1 \det\left(\sum_{i=1}^{m} \lambda_i D^2\phi_i\right) d\,\mathrm{vol_n}\,.$$

Obviously this determinant is a polynomial in the λ_i and, by the linearity of the integral, so is the left hand side. $\square$

Using again methods similiar to those above, we will prove Minkowski's inequality.

Theorem 4.5 (Minkowski's inequality)**.** *Let K_1 and K_2 be compact convex bodies and let $V_k(K_1, K_2)$ denote the mixed volume of K_1 taken k times and K_2 taken $n-k$ times. Then the inequality*

$$V_k(K_1, K_2) \geq \mathrm{vol_n}[K_1]^{\frac{k}{n}} \cdot \mathrm{vol_n}[K_2]^{\frac{k-n}{n}} \tag{4.1}$$

holds.

Before giving the proof we need a lemma concerning simultaneous diagonalisation of symmetric positive matrices:

Lemma 4.6. *Let N and M be symmetric positive matrices. Then there exists a matrix T with determinant 1 such that $T^T MT$ and $T^T NT$ are diagonal matrices.*

Proof. Let $Q^T Q$ be the Cholesky decomposition of N^{-1}. Then QMQ^T is a symmetric matrix and can therefore be diagonalised using an orthogonal matrix O. Hence we can calculate

1. $(Q^T O)^T M (Q^T O) = O^T Q M Q^T O = D$ for some diagonal matrix D.

2. $(Q^T O)^T N (Q^T O) = (Q^T O)^T (Q^T Q)^{-1} (Q^T O) = O^T (Q Q^{-1})(Q^{-T} Q^T) O = I_n$.

Hence, setting $T = \frac{1}{\sqrt[n]{\det(Q^T O)}} Q^T O$ completes the proof. $\square$

Proof of Theorem 4.5. By homogeneity we can normalise the volumes of K_1 and K_2 to 1. Let μ and ν be the uniform probability measures on K_1 and K_2 respectively, and let ρ denote both the standard Gaussian measure on $\mathbb{R}^n$ and its density.
Consider the Brenier maps $\nabla\phi_1$, transporting ρ to μ, and $\nabla\phi_2$, transporting ρ to ν. By (4.2) these maps are C^1-smooth, and, by Theorem 4.3

$$(\lambda_1 \nabla\phi_1 + \lambda_2 \nabla\phi_2)(\mathbb{R}^n) = \lambda_1 K_1 + \lambda_2 K_2$$

holds.

Minkowski's theorem on mixed volumes implies that

$$\mathrm{vol}_\mathrm{n}[\lambda_1 K_1 + \lambda_2 K_2] = \sum_{i=0}^{n} \binom{n}{i} \lambda_1^i \lambda_2^{n-i} V_k(K_1, K_2) \tag{4.2}$$

and on the other hand, using the change of variable formula,

$$\mathrm{vol}_\mathrm{n}\left[\lambda_1 K_1 + \lambda_2 K_2\right] = \int\limits_{\mathbb{R}^n} 1 \det\left(\lambda_1 D^2\phi_1 + \lambda_2 D^2\phi_2\right) d\,\mathrm{vol}_\mathrm{n} \tag{4.3}$$

holds. Hence, to calculate the mixed volumes $V_k(K_1, K_2)$ it is sufficient to determine the coefficient of $\lambda_1^k \lambda_2^{n-k}$ in the polynomial (4.3). Applying Lemma 4.6 to the integrand, we find that

$$\det\left(\lambda_1 D^2\phi_1 + \lambda_2 D^2\phi\right) = \det\left(\lambda_1 A + \lambda_2 B\right) = \prod_{i=1}^{n} (\lambda_1 a_i + \lambda_2 b_i)$$

for two diagonal matrices $A = \mathrm{diag}(a_1, ..., a_n)$ and $B = \mathrm{diag}(b_1, ... b_n)$. Expanding the last term we obtain

$$\prod_{i=1}^{n} (\lambda_1 a_i + \lambda_2 b_i) = \sum_{i=0}^{n} \left(\sum_{\substack{I \subset \{1,\dots,n\} \\ |I| = i}} \left(\prod_{j \in I} a_j \right) \left(\prod_{j \in I^c} b_j \right) \right) \lambda_1^i \lambda_2^{n-i}.$$

Therefore, we can estimate the relevant coefficient using the arithmetic-geometric inequality

$$\binom{n}{k}\binom{n}{k}^{-1}\sum_{\substack{I\subset\{1,\dots,n\}\\|I|=k}}\left(\prod_{j\in I}a_j\right)\left(\prod_{j\in I^c}b_j\right)\geq\binom{n}{k}\prod_{\substack{I\subset\{1,\dots,n\}\\|I|=k}}\left(\left(\prod_{j\in I}a_j\right)\left(\prod_{j\in I^c}b_j\right)\right)^{\binom{n}{k}^{-1}} \tag{4.4}$$

$$= \binom{n}{k}\left(\left(\prod_{j=1}^n a_j\right)^{\binom{n-1}{k-1}}\left(\prod_{j=1}^n b_j\right)^{\binom{n-1}{k}}\right)^{\binom{n}{k}^{-1}} = \binom{n}{k}\left(\prod_{j=1}^n a_j\right)^{\frac{k}{n}}\left(\prod_{j=1}^n b_j\right)^{\frac{n-k}{n}} \tag{4.5}$$

$$= \binom{n}{k}\det(A)^{\frac{k}{n}}\det(B)^{\frac{n-k}{n}}. \tag{4.6}$$

Using (4.2) and (4.3), we find that

$$V_k(K_1,K_2)\geq\int_{\mathbb{R}^n}\det\left(D^2\phi_1\right)\cdot\det\left(D^2\phi_2\right)d\,\mathrm{vol_n},$$

and using the Monge-Ampère equations for $\nabla\phi_1$ and $\nabla\phi_2$, i.e. $D^2\phi_i=\rho$ for $i=1,2$, we obtain

$$\int_{\mathbb{R}^n}\det\left(D^2\phi_1\right)\cdot\det\left(D^2\phi_2\right)d\,\mathrm{vol_n}=\int_{\mathbb{R}^n}\rho d\,\mathrm{vol_n}=1,$$

completing the proof. □

Remark 4.1. Analyzing this proof further, it is also possible to settle the equality cases of the Minkowski inequality: Equality in (4.4) implies that

$$\left(\prod_{j\in I}a_j\right)\left(\prod_{j\in I^c}b_j\right)$$

amounts to the same value for any index set I with $|I|=k$. Equivalently, $a_1/b_1=a_j/b_j$ for all indices j (Consider two index sets $I'\cup 1$ and $I'\cup j$, where $|I'|=k-1$ and I' contains neither 1 nor j.), in other words A is a positive multiple of B. Hence, $\nabla\phi_1$ is a positive multiple of $\nabla\phi_2$ and therefore K_1 and K_2 have to be positive homothetic.

Remark 4.2. Using a similar argument it is also possible to obtain a more general result, namely some of the famous Aleksandrov-Fenchel inequalities for mixed volumes. More precisely, it is possible to proof the following statement: Let $K_1,...,K_n$ be convex bodies. Then the following inequality holds:

$$V(K_1,...,K_n)\geq\left(\prod_{i=1}^n\mathrm{vol_n}[K_i]\right)^{\frac{1}{n}}.$$

This proof can be found in [1].

5 Brascamp-Lieb Inequalities

5.1 The Brascamp-Lieb Inequalities and the Reverse Brascamp-Lieb Inequalities

In this section the multidimensional Brascamp-Lieb inequalities and their duals, the reverse Brascamp-Lieb inequalities, are proved. This proof we present, using optimal transport, is extracted from an article by Barthe [4].
It will be shown that optimal constants in these inequalities can be calculated by considering centered Gaussian functions, that is functions of the form $\gamma(x) = \exp(-x^T Ax)$, where A is a positive symmetric matrix.

Theorem 5.1. *Let N and $(n_i)_{1\leq i\leq m}$ be integers satisfying $n_i \leq N$ and let $(c_i)_{1\leq i\leq m}$ be positive real numbers such that*

$$\sum_{i=1}^{m} c_i n_i = N.$$

Let $B_i : \mathbb{R}^N \to \mathbb{R}^{n_i}$ be surjective linear mappings such that

$$\bigcap_i \ker B_i = \{0\}.$$

For integrable nonnegative functions f_i define the two functions

$$\begin{aligned} I(f_1, f_2, ..., f_m) &= \int_{\mathbb{R}^N} \prod_i f_i^{c_i}(B_i x)dx, \\ J(f_1, f_2, ..., f_m) &= \int_{\mathbb{R}^N} \sup \left\{ \prod_i f_i^{c_i}(x_i) : x = \sum c_i B_i^T(x_i) \right\} dx. \end{aligned}$$

Then the optimal constants E and F in the inequalities

$$I(f_1, f_2, ..., f_m) \leq E \prod_i \left(\int_{\mathbb{R}^{n_i}} f_i \right)^{c_i} \quad \text{and} \quad J(f_1, f_2, ..., f_m) \geq F \prod_i \left(\int_{\mathbb{R}^{n_i}} f_i \right)^{c_i} \tag{5.1}$$

can be computed by considering only centered Gaussian functions, that is:

$$E = \sup\left\{\frac{I(\gamma_1,...,\gamma_m)}{\prod\limits_i(\int_{\mathbb{R}^{n_i}}\gamma_i)^{c_i}} : \gamma_i \text{ centered Gaussian on } \mathbb{R}^{n_i}\right\},$$
$$F = \inf\left\{\frac{J(\gamma_1,...,\gamma_m)}{\prod\limits_i(\int_{\mathbb{R}^{n_i}}\gamma_i)^{c_i}} : \gamma_i \text{ centered Gaussian on } \mathbb{R}^{n_i}\right\}.$$

Moreover, E and F are given by $E = \frac{1}{\sqrt{D}}$ and $F = \sqrt{D}$, where D is defined by

$$D = \inf\left\{\frac{\det(\sum_i c_i B_i^T A_i B_i)}{\prod_i(\det A_i)^{c_i}} : A_i \text{ positive } n\text{-dimensional symmetric matrix}\right\}.$$

The inequalities in (5.1) are called the Brascamp-Lieb and the reverse Brascamp-Lieb inequalities, respectively.

Before we start with the proof, two special cases of these inequalities are given in the next two remarks.

Remark 5.1. **Young's Inequality** Let $m = 3$ and $N = 2n$ and $n_1 = n_2 = n_3 = n$ and let $c_1 = \frac{1}{p}$, $c_2 = \frac{1}{q}$ and $c_3 = \frac{1}{r'} = 2 - \frac{1}{p} - \frac{1}{q}$ in the Brascamp-Lieb inequality. Then with $B_1(x,y) = x$,$B_2(x,y) = x - y$ and $B_3(x,y) = y$ we find Young's inequality

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} f(x)^{\frac{1}{p}} g(x-y)^{\frac{1}{q}} h(y)^{\frac{1}{r'}} dxdy \le E\left(\int_{\mathbb{R}^n} f(x)dx\right)^{\frac{1}{p}}\left(\int_{\mathbb{R}^n} g(x)dx\right)^{\frac{1}{q}}\left(\int_{\mathbb{R}^n} h(x)dx\right)^{\frac{1}{r'}}.$$

Remark 5.2. **Prèkopa Leindler Inequality** The Prèkopa-Leindler inequality is a special case of the reverse Brascamp-Lieb inequality for $m = 2, n_1 = n_2 = N, B_1 = B_2 = I_n$ and $c_1 + c_2 = 1$. Indeed

$$D = \inf\frac{\det(c_1A_1 + c_2A_2)}{(\det A_1)^{c_1}(\det A_2)^{c_2}} = 1$$

as the numerator is always greater or equal the denominator by the arithmetic-geometric inequality.
So the reverse Brascamp-Lieb inequality becomes

$$\int_{\mathbb{R}^n}\left[\sup_{z=c_1x+c_2y} f^{c_1}g^{c_2}\right]dz \ge \left(\int_{\mathbb{R}^n} f\right)^{c_1}\left(\int_{\mathbb{R}^n} g\right)^{c_2}.$$

In the proof of Theorem 5.1 the following short lemma about gaussian integrals is needed:

Lemma 5.2. *Let M be a positive symmetric n-dimensional matrix. Then*

$$\int\limits_{\mathbb{R}^n}\exp(-x^TMx)dx = \sqrt{\frac{\pi^n}{\det M}}.$$

Proof. Recall that a positive symmetric matrix M has a positive symmetric square root $\sqrt{M}$. Therefore, with the change of variable formula,

$$\begin{aligned}\int_{\mathbb{R}^n} \exp(-x^T M x)dx &= \int_{\mathbb{R}^n} \exp(-(\sqrt{M}x)^T(\sqrt{M}x))dx = \sqrt{\frac{1}{\det M}} \int_{\mathbb{R}^n} \exp(-y^T y)dy \\ &= \sqrt{\frac{1}{\det M}} \prod_{i=1}^{n} \int_{-\infty}^{\infty} e^{-y_i^2} dy_i = \sqrt{\frac{\pi^n}{\det M}}.\end{aligned}$$

□

Proof of Theorem 3.1. We set

$$E_g = \sup \left\{ \frac{I(\gamma_1, ..., \gamma_m)}{\prod\limits_i (\int_{\mathbb{R}^{n_i}} \gamma_i)^{c_i}} : \gamma_i \text{ centered Gaussian on } \mathbb{R}^{n_i} \right\},$$

$$F_g = \inf \left\{ \frac{J(\gamma_1, ..., \gamma_m)}{\prod\limits_i (\int_{\mathbb{R}^{n_i}} \gamma_i)^{c_i}} : \gamma_i \text{ centered Gaussian on } \mathbb{R}^{n_i} \right\}.$$

As a first step we will prove that $E_g = \frac{1}{\sqrt{D}}$ and $F_g = \sqrt{D}$. For positive definite n_i-dimensional matrices A_i we define the quadratic form Q on $\mathbb{R}^n$ by

$$Q(y) = y^T \left(\sum_{i=1}^{m} c_i B_i^T A_i B_i \right) y.$$

We also introduce a function R on $\mathbb{R}^N$ by

$$R(x) = \inf \left\{ \sum_{i=1}^{m} c_i x_i^T A_i^{-1} x_i : x = \sum_{i=1}^{m} c_i B_i^T x_i, x_i \in \mathbb{R}^{n_i} \right\}.$$

The first step is to show that R is the dual quadratic Form Q^* of Q, that is, the quadratic form defined by $Q^*(x) = \sup \{ |x \cdot y|^2 : Q(y) \leq 1 \}$.

Indeed, with $x = \sum\limits_{i=1}^{m} c_i B_i^T x_i$,we find that

$$\begin{aligned} |x \cdot y|^2 &= \left| \sum_{i=1}^{m} (c_i B_i^T x_i)^T y \right|^2 = \left| \sum_{i=1}^{m} \sqrt{c_i} x_i^T \sqrt{A_i}^{-1} \sqrt{A_i} B_i \sqrt{c_i} y \right|^2 \\ &= \left| \sum_{i=1}^{m} \left(\sqrt{c_i} \sqrt{A_i}^{-1} x_i \right) \cdot \left(\sqrt{c_i} \sqrt{A_i} B_i y \right) \right|^2 \\ &\leq \left(\sum_{i=1}^{m} |\sqrt{c_i} \sqrt{A_i}^{-1} x_i|^2 \right) \left(\sum_{i=1}^{m} |\sqrt{c_i} \sqrt{A_i} B_i y|^2 \right) \\ &= \left(\sum_{i=1}^{m} c_i x_i^T A_i^{-1} x_i \right) \left(\sum_{i=1}^{m} c_i y^T B_i^T A_i B_i y \right) \end{aligned}$$

using the Cauchy-Schwarz inequality. Taking the infimum over all decompositions of x of the form $x = \sum_{i=1}^{m} c_i B_i^T x_i$ we get

$$|x \cdot y|^2 \leq R(x)Q(y)$$

and so $Q^*(x) \leq R(x)$.
On the other hand, if $y = \left(\sum c_i B_i^T A_i B_i\right)^{-1} x$, then clearly $x_i = A_i B_i y$ constitute such a decomposition of x. With these we easily find equality in the calculations above, and so $|x \cdot y|^2 \geq R(x)Q(y)$ holds also. It follows that $R = Q^*$.

Calculating $I(\gamma(A_1), .., \gamma(A_m))$ and $J(\gamma(A_1^{-1}), .., \gamma(A_m^{-1}))$,where $\gamma(A_i)(x) = \exp(-x^T A_i x)$ with a positive symmetric matrix A_i we get, using Lemma 5.2,

$$\begin{aligned}
I(\gamma(A_1), .., \gamma(A_m)) &= \int_{\mathbb{R}^N} \prod_i \gamma(A_i)^{c_i}(B_i x) dx \\
&= \int_{\mathbb{R}^N} \exp\left[-x^T \left(\sum_i c_i B_i^T A_i B_i\right) x\right] dx = \sqrt{\frac{\pi^N}{\det Q}}
\end{aligned}$$

and

$$\begin{aligned}
J(\gamma(A_1^{-1}), .., \gamma(A_m^{-1})) &= \int_{\mathbb{R}^N} \sup\left\{\prod_i \gamma(A_i)^{-c_i}(x_i) : x = \sum c_i B_i^T x_i\right\} dx \\
&= \int_{\mathbb{R}^N} \sup\left\{\exp\left(-\sum_i c_i x_i^T A_i^{-1} x_i\right) : x = \sum c_i B_i^T x_i\right\} dx \\
&= \int_{\mathbb{R}^N} \exp(-R(x)) dx = \sqrt{\frac{\pi^N}{\det R}}.
\end{aligned}$$

In particular,

$$\begin{aligned}
E_g &= \sup \sqrt{\frac{\prod_{i=1}^m (\det A_i)^{c_i}}{\det Q}} = \frac{1}{\sqrt{D}}, \\
F_g &= \inf \sqrt{\frac{\prod_{i=1}^m (\det A_i)^{-c_i}}{\det R}} = \sqrt{\frac{\det Q}{\prod_{i=1}^m (\det A_i)^{c_i}}} = \sqrt{D}
\end{aligned}$$

by another application of Lemma 5.2 and using the fact that $\det Q^* \det Q = 1$.

By now, we know that $\sqrt{D} = F_g \geq F$ and that $DE \geq DE_g = \sqrt{D}$. So if we knew that $F \geq DE$ the proof would be finished, as we would have $\sqrt{D} = F = DE = \sqrt{D}$.

So, in fact, it will be sufficient to prove that for all nonnegative integrable functions f_i and g_i with $\int f_i = \int g_i = 1$,

$$I(g_1, ..., g_m) \leq \frac{1}{D} \cdot J(f_1, ..., f_m)$$

because taking the infimum over the g_i and the supremum over the f_i will yield that $F \geq DE$.

Without loss of generality we assume that $D > 0$. Let μ_i be the probability measures with densities f_i and let ν_i be those with densities g_i. Let ϕ_i be the Brenier map transporting μ_i to ν_i.
Then for almost all x the following Monge-Ampère equations hold

$$f_i(x) = \det(D_A^2\phi(x))g_i(\nabla\phi(x)). \tag{5.2}$$

We define a function $\Theta : \mathbb{R}^N \to \mathbb{R}^N$ by

$$\Theta(y) = \sum_{i=1}^{m} c_i B_i^T(\nabla\phi(B_i y)).$$

Its differential $\sum_{i=1}^m c_i B_i^T D_A^2(B_i y)B_i$ is obviously symmetric and nonnegative, but in fact even positive on

$$S = \{y \in \mathbb{R}^N : f_i(B_i^{-1}(y)) > 0 \text{ for all } i \in \{1, ..., m\}\}$$

by the following argument.
On S, obviously, $\det D_A^2(B_i y))^{c_i}$ is non zero by the Monge-Ampère equation (5.2). Thus using the definition of D we find

$$\det \sum_{i=1}^{m} c_i B_i^T D_A^2(B_i y)B_i \geq \prod_i (\det D_A^2(B_i y))^{c_i} > 0.$$

So, in particular Θ is injective.

Now we can finally calculate

$$\begin{aligned}
I(f_1, ..., f_m) &= \int_{\mathbb{R}^N} \prod_{i=1}^{m} f_i^{c_i}(B_i y)dy = \int_S \prod_{i=1}^{m} f_i^{c_i}(B_i y)dy \\
&= \int_S \prod_{i=1}^{m} \left[\det(D_A^2\phi(B_i y))g_i(\nabla\phi(B_i y))\right]^{c_i} dy \\
&\leq \frac{1}{D}\int_S \left(\prod_{i=1}^{m} [g_i(\nabla\phi(B_i y))]^{c_i}\right)\left(\det \sum_{i=1}^{m} c_i B_i^T D_A^2(B_i y)B_i\right) dy \\
&\leq \frac{1}{D}\int_S \sup_{\Theta(y)=\sum c_i B_i^T x_i} \left(\prod_{i=1}^{m} g_i(x_i)^{c_i}\right) \det(D\Theta(y))dy \\
&\leq \frac{1}{D}\int_{\mathbb{R}^n} \sup_{x=\sum c_i B_i^T x_i} \left(\prod_{i=1}^{m} g_i(x_i)^{c_i}\right) dx = \frac{1}{D}\cdot J(g_1, ..., g_m)
\end{aligned}$$

which completes the proof. $\square$

5.2 Optimisers for the Brascamp-Lieb Inequality

The goal of this section is to describe functions for which there is equality in

$$\int_{\mathbb{R}^N} \prod_i f_i^{c_i}(B_i x) dx \leq E \prod_i \left(\int_{\mathbb{R}^{n_i}} f_i \right)^{c_i}.$$

The statements made in this section will not be proven. Originally these results, along with proofs, can be found in [16].

Obviously such optimizers will depend on the tuples (B_i, c_i). (For a given Brascamp-Lieb inequality we will refer to the vector (B_i, c_i) as its Brascamp-Lieb datum.) In fact there are such tuples were no extremisers exist. To answer the question which data are extremisable we will need the following definitions.

Definition 5.1. *A Brascamp-Lieb datum is geometric if $B_i B_i^T = \mathrm{Id}_{\mathbb{R}^{n_i}}$ for each i and*

$$\sum_{i=1}^{m} c_i B_i^T B_i = \mathrm{Id}_{\mathbb{R}^N}.$$

Definition 5.2. *We say that two Brascamp-Lieb data (B_i, c_i) and (B_i', c_i) are equivalent if there are invertible matrices C and C_j such that $B_j' = C_j^{-1} B_j C$ for all j.*

With these definitions we can state

Theorem 5.3. *Every extremisable Brascamp-Lieb datum is equivalent to a geometric datum.*

To be able to state the results on the form of optimizers some knowledge about the structure of the Bsascamp-Lieb inequality is needed.
In the following discussion we shall assume that $c_j > 0$ for all j and concentrate on the case of geometric data. We can then generalize our findings using Theorem 5.3.

We will begin by defining some decompositions of the Euclidean space:

Definition 5.3. *A subspace V of $\mathbb{R}^N$ is said to be critical for (B_i, c_i) if V is neither $\{0\}$ nor $\mathbb{R}^N$ and*

$$\dim V = \sum_i c_i \dim(B_i V).$$

It is possible to get a so called maximal critical decomposition, where we write $\mathbb{R}^N$ as a sum of pairwise orthogonal spaces, each of which is critical and has no critical subspaces.

Definition 5.4. *A subspace V of $\mathbb{R}^N$ is said to be independent if it is not $\{0\}$ and can be written as*

$$V = \bigcap_{i=1}^{m} [\mathbb{R}^{n_i}]^{\alpha},$$

where $[\mathbb{R}^{n_i}]^\alpha$ is either $\mathbb{R}^{n_i}$ or $[\mathbb{R}^{n_i}]^\perp$. An independent decomposition of $\mathbb{R}^N$ is a decomposition of the form

$$\mathbb{R}^N = K_{ind} \oplus K_{dep} = \left(\bigoplus_{k=0}^{k_0} K_k \right) \oplus K_{dep},$$

where K_k are independent subspaces of $\mathbb{R}^N$ and K_{dep} is the orthogonal complement of their sum.

With these definitions we can state the main theorem of this section:

Theorem 5.4. *Let (B_i, c_i) be a geometric Brascamp-Lieb datum and let $\left(\bigoplus_{k=0}^{k_0} K_k \right) \oplus K_{dep}$ be the independent decomposition of $\mathbb{R}^N$. Then the tuple (f_i) is an extremiser if and only if there are*

(i) integrable functions $u_k : K_k \to \mathbb{R}$ for $k = 1, ..., k_0$,

(ii) a critical decomposition $K_{k_0+1} \oplus ... \oplus K_{k_1}$ of K_{dep},

(iii) positive constants α_i for $i = 1, ..., m$ and β_k for $k = k_0 + 1, ..., k_1$

(iv) and an element b from K_{dep} such that almost everywhere

$$f_i(x) = \alpha_i \prod_{k=1}^{k_0} u_k(P_{i,k} B_i^T x) \prod_{k=k_0+1}^{k_1} e^{-\beta_k (P_{i,k} B_i^T x) \cdot (P_{i,k} (B_i^T x + b))},$$

where $P_{i,k}$ is the orthogonal projection from $\mathbb{R}^N$ to $\mathbb{R}^{n_i} \cap K_k$.

For example we can use this to characterize cases of equality in Hölder's inequality.

Remark 5.3. **Hölder's inequality.** Let $m = 2$ and $N = n_1 = n_2 = n$ and let $c_1 = \frac{1}{p}$ and $c_2 = 1 - \frac{1}{p} = \frac{1}{q}$ in the Brascamp-Lieb inequality. Then with $B_1 = B_2 = 1$, we find

$$\int_{\mathbb{R}^n} f_1^{\frac{1}{p}}(x) f_2^{\frac{1}{q}}(x) dx \leq E \left(\int_{\mathbb{R}^n} f_1(x) dx \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} f_2(x) dx \right)^{\frac{1}{q}}.$$

Obviously $((\mathrm{Id}, \frac{1}{p}), (\mathrm{Id}, \frac{1}{q}))$ is of geometric type and the independent decomposition is very simple:

$$\mathbb{R}^n = K_{\text{ind}}.$$

Therefore we find that $f_1 = \alpha_1 u$ and $f_2 = \alpha_2 u$ almost everywhere for some integrable u and positive constants c_1 and c_2. By then it is obvious that $E = 1$.

Using Theorem 5.3 we get the following corollary:

Corollary 5.5. *Let (B_i, c_i) be an extremisable Brascamp-Lieb datum. Let (B_i', c_i) be the geometric datum equivalent to (B_i, c_i) and let M and S_i be such that $B_i' = S_i^{\frac{1}{2}} B_i M^{\frac{1}{2}}$. Furthermore let $\left(\bigoplus_{k=0}^{k_0} K_k\right) \oplus K_{dep}$ be the independent decomposition of $\mathbb{R}^N$. Then the tuple (f_i) is an extremiser if and only if there are*

(i) integrable functions $u_k : K_k \to \mathbb{R}$ for $k = 1, ..., k_0$,

(ii) a critical decomposition $K_{k_0+1} \oplus ... \oplus K_{k_1}$ of K_{dep},

(iii) positive constants α_i for $i = 1, ..., m$ and β_k for $k = k_0 + 1, ..., k_1$

(iv) and an element b from K_{dep} such almost everywhere

$$f_i(x) = \alpha_i \prod_{k=1}^{k_0} u_k(P_{i,k} {B'}_i^T S_i^{\frac{1}{2}} x), \prod_{k=k_0+1}^{k_1} e^{-\beta_k (P_{i,k} {B'}_i^T S_i^{\frac{1}{2}} x) \cdot (P_{i,k}({B'}_i^T S_i^{\frac{1}{2}} x + b))}$$

where $P_{i,k}$ is the orthogonal projection from $\mathbb{R}^N$ to $\mathbb{R}^{n_i} \cap K_k$.

6 Sobolev Inequalities

Dealing with partial differential equations one often finds that Sobolev spaces are the right setting to work in. One of the main results concerning these spaces are Sobolev embeddings and inequalities. An introduction to this theory can be found in [9][Chapter 5].

In this chapter we will consider only a certain kind of Sobolev inequality, the Gagliardo-Nirenberg-Sobolev inequality:

$$||f||_{L^{p^\star}} \leq S_n(p)||\nabla f||_{L^p}, \tag{6.1}$$

where $1 \leq p < n$ and $p^\star = \frac{np}{n-p}$.

Using optimal transport, we will be able to explicitly state optimal values for the constants $S_n(p)$ and functions h_p for which equality holds in this inequality. This aproach is originally due to D. Cordero-Erausquin, B. Nazaret, and C. Villani and can be found in [7].

6.1 Sobolev Spaces

In this section some basic definitions and results concerning Sobolev spaces will be recalled. For Sobolev spaces to be a suitable setting for partial differential equations, they have to contain quite non-smooth functions.
In fact they comprise of functions that only have derivatives in a weak sense.

Definition 6.1. (Weak partial derivatives) *Suppose f and g are locally integrable functions on $\mathbb{R}^n$. We say that g is the weak partial derivative of f with respect to x_i if*

$$\int f \frac{\partial \zeta}{\partial x_i} dx = -\int g\zeta dx$$

for all test functions $\zeta \in C_c^\infty(\mathbb{R}^n)$.

If we are given a locally integrable function f and there exists such a function g that satifies the formula above, f is said to have a weak partial derivative with respect to x_i. If there is no such g then f does not.

∇f is defined accordingly.

This definition is of course motivated by the integration by parts formula.

Definition 6.2. *The Sobolev space*

$$W^{1,p}(\mathbb{R}^n)$$

consists of all locally integrable functions f for which all weak partial derivatives exist and both $||f||_p$ and $||\nabla f||_p$ are finite.

As for L_p spaces we will identify functions in $W^{1,p}$ which are equal almost everywhere.

With the norm $||f||_{W^{1,p}} = ||f||_p + ||\nabla f||_p$ these Sobolev spaces are Banach spaces and $W^{1,p}(\mathbb{R}^n) \cap C_c^\infty(\mathbb{R}^n)$ is a dense subset of $W^{1,p}(\mathbb{R}^n)$.

In a very similar way we define the homogeneous Sobolev space

$$\dot{W}^{1,p}(\mathbb{R}^n)$$

as the set of all locally integrable functions f for which all weak partial derivatives exist and both $||f||_{p^*}$ and $||\nabla f||_p$ are finite where $p^* = \frac{np}{n-p}$.

6.2 Sobolev Inequalities

As already mentioned in the introduction we will only consider the Sobolev spaces $W^{1,p}(\mathbb{R}^n)$ for $p \in [1, n)$. For such p we define its Sobolev conjugate by $p^* = \frac{np}{n-p}$.

For notational reasons we will separate the case $p = 1$ from the rest. Let us start with $p > 1$.

Theorem 6.1. *Let $p \in (1, n)$ and let $q = \frac{p}{p-1}$ be the dual exponent. Define the function h_p by*

$$h_p(x) = \frac{1}{(\sigma_p + |x|^q)^{\frac{n-p}{p}}},$$

where σ_p is chosen such that $||h_p||_{L^{p^\star}} = 1$.

Whenever $f, g \in L^{p^\star}(\mathbb{R}^n)$ are two functions satisfying $||f||_{L^{p^\star}} = ||g||_{L^{p^\star}}$, and if $\nabla f \in L^p(\mathbb{R}^n)$, then

$$\frac{\int |g(x)|^{p^\star(1-\frac{1}{n})}dx}{\left(\int |x|^q |g(x)|^{p^\star}\right)^{\frac{1}{q}}} \leq \frac{p(n-1)}{n(n-p)}||\nabla f||_{L^p} \tag{6.2}$$

with equality if $f = g = h_p$.

Before the proof we state a corollary:

Corollary 6.2. *The following statement is an immediate consequence of Theorem 6.1.*

> *Let f and h_p be defined as in Theorem 6.1. If $f \neq 0$ then the Sharp Sobolev inequality*
> $$\frac{||\nabla f||_{L^p}}{||f||_{L^{p^\star}}} \geq ||\nabla h_p||_{L^p}$$
> *holds.*

Proof of Theorem 6.1. As $|\nabla|f|| = |\nabla f|$ holds almost everywhere, it obviously follows that $||\nabla f||_{L^p} = ||\nabla|f|||_{L^p}$ and therefore we only need to consider nonnegative functions f. Moreover, by homogeneity, we can without loss of generality assume that $||f||_{L^{p^\star}} = ||g||_{L^{p^\star}} = 1$.

Because of the density of smooth functions with compact support in Sobolev spaces, we need only consider such functions f and g.

Let F and G be the two probability densities

$$F(x) = f^{p^\star}, \qquad G(x) = g^{p^\star},$$

and let μ and ν be the corresponding probability measures.

Consider the Brenier map $\nabla\phi$ transporting μ to ν. Then the Monge-Ampére equality holds

$$F(x) = G(\nabla\phi(x)) \det D_A^2\phi(x)$$

for μ-almost all x. Using the arithmetic/geometric mean inequality and the fact that the determinant of a matrix is the product of its eigenvalues and its trace their sum, we get

$$G^{-\frac{1}{n}}(\nabla\phi(x)) = F^{-\frac{1}{n}}(x)(\det D_A^2\phi(x))^{\frac{1}{n}} \leq F^{-\frac{1}{n}}(x)\frac{\Delta_A\phi(x)}{n}.$$

By integrating both sides with respect to μ, we further obtain

$$\int G^{-\frac{1}{n}}(\nabla\phi(x))F(x)dx \leq \int F^{1-\frac{1}{n}}(x)\frac{\Delta_A\phi(x)}{n}dx.$$

Since $\int G^{-\frac{1}{n}}(\nabla\phi(x))F(x)dx = \int G^{-\frac{1}{n}}(x)G(x)dx$ by the definition of the pushforward measure we find that

$$\int G^{1-\frac{1}{n}}(x)dx \leq \int F^{1-\frac{1}{n}}(x)\frac{\Delta_A\phi(x)}{n}dx. \tag{6.3}$$

Considering the right hand side of this inequality, and using the fact that $\Delta_A\phi$ is bounded from above by the distributional Laplacian $\Delta_{D'}\phi$ we can write

$$\frac{1}{n}\int F^{1-\frac{1}{n}}\Delta_A\phi \leq \frac{1}{n}\int F^{1-\frac{1}{n}}\Delta_{D'}\phi = -\frac{1}{n}\int \nabla(F^{1-\frac{1}{n}})\cdot\nabla\phi. \tag{6.4}$$

Putting together (6.3) and (6.4) and the definitions of F and G, we get

$$\begin{aligned}\int g^{p^\star(1-\frac{1}{n})} &\leq -\frac{1}{n}\int \nabla(f^{p^\star(1-\frac{1}{n})})\cdot\nabla\phi = -\frac{1}{n}\int\nabla(f^{\frac{p(n-1)}{n-p}})\cdot\nabla\phi \\ &= -\frac{p(n-1)}{n(n-p)}\int f^{\frac{n(p-1)}{n-p}}\nabla f\cdot\nabla\phi = -\frac{p(n-1)}{n(n-p)}\int f^{p^\star/q}\nabla f\cdot\nabla\phi.\end{aligned}$$

Using Hölder's inequality on the last integral, we obtain

$$-\int f^{p^\star/q}\nabla f\cdot\nabla\phi \leq ||\nabla f||_{L^p}\cdot||f^{p^\star/q}\nabla\phi||_{L^q} = ||\nabla f||_{L^p}\left(\int f^{p^\star}|\nabla\phi|^q\right)^{\frac{1}{q}}. \qquad (6.5)$$

Again by the definition of the pushforward measure we obtain

$$\int f^{p^\star}|\nabla\phi|^q = \int|\nabla\phi|^q d\mu = \int|x|^q d\nu = \int|x|^q g^{p^\star}(x)dx.$$

Combining the last three results, we see that

$$\int g^{p^\star(1-\frac{1}{n})} \leq \frac{p(n-1)}{n(n-p)}||\nabla f||_{L^p}\left(\int|x|^q g^{p^\star}dx\right)^{\frac{1}{q}}$$

which completes the proof of inequality (6.2).

To show the equality in (6.2) for $f = g = h_p$ we only have to verify equality in the inequalities (6.3), (6.4), and (6.5):

If $f = g$, then obviously the measures μ and ν are equal, and therefore $\nabla\phi$ is the identity map, and so we have equality in (6.3) as $\Delta\phi = n$.
Equality in (6.4) follows from integration by parts.
Finally, as there is equality in Hölders inequality $||f\cdot g||_{L^1} \leq ||f||_{L^p}\cdot||g||_{L^q}$ if f^p is proportional to g^q almost everywhere, we only have to show that there is a $c\in\mathbb{R}$ with $(\nabla h_p(x))^p = c\cdot\left(xh_p(x)^{\frac{p^\star}{q}}\right)^q$ for almost all x. This is easily done by an explicit calculation. ☐

For $p = 1$ a very similar statement can be proved:

Theorem 6.3. *If $f \neq 0$ is a smooth compactly supported function, then*

$$\frac{||\nabla f||_{L^1}}{||f||_{L^{n/(n-1)}}} \geq n\,\mathrm{vol_n}[B_n]^{\frac{1}{n}}.$$

This inequality extends to functions with bounded variation, with equality if $f = h_1$ where h_1, is defined by

$$h_1 = \frac{\chi_{B_n}(x)}{\mathrm{vol_n}[B_n]^{\frac{n-1}{n}}}.$$

This variant for $p = 1$ can be proved in a way that is very similar, but will depend on the coarea formula for the equality case h_1. Before starting with the proof we will state this important formula.

Theorem 6.4. *Let $u : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous and assume that for almost all $r \in \mathbb{R}$ the level set*

$$\{x \in \mathbb{R}^n : u(x) = r\}$$

is a smooth hypersurface. Suppose also that $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and integrable. Then

$$\int_{\mathbb{R}^n} f|\nabla u|dx = \int_{-\infty}^{\infty} \left(\int_{\{u=r\}} f d\,\mathrm{vol_{n\text{-}1}} \right) dr.$$

Proof of the Sobolev inequality for $p = 1$. Without loss of generality we will only consider nonnegative functions f with $||f||_{L^{n/(n-1)}} = 1$.
Introduce the two probability densities

$$F(x) = f^{n/(n-1)}, \qquad G(x) = h_1^{n/(n-1)} = \frac{\chi_{B_n}(x)}{\mathrm{vol_n}[B_n]},$$

and let μ and ν be the corresponding probability measures.
As in the proof for $p > 1$ we obtain

$$G^{-\frac{1}{n}}(\nabla\phi(x)) = F^{-\frac{1}{n}}(x)(\det D_A^2\phi(x))^{\frac{1}{n}} \leq F^{-\frac{1}{n}}(x)\frac{\Delta_A\phi(x)}{n},$$

where $\nabla\phi$ is the Brenier map transporting μ to ν, and again by integration and application of the push-forward formula we find

$$\int G^{1-\frac{1}{n}} \leq \frac{1}{n}\int F^{1-\frac{1}{n}}\Delta_A\phi.$$

Using the definitions of F and G and the integration by parts formula (justified as in the last proof) we arrive at

$$\mathrm{vol_n}[B_n]^{\frac{1}{n}} \leq \frac{1}{n}\int f\Delta_A\phi \leq -\frac{1}{n}\int \nabla f \cdot \nabla\phi.$$

As h_1 is supported in B_n we see that $|\nabla\phi(x)| \leq 1$ for almost all x in the support of f. Therefore $-\nabla f(x) \cdot \nabla\phi(x) \leq |\nabla f(x)||\nabla\phi(x)| \leq |\nabla f(x)|$ and thus

$$n\,\mathrm{vol_n}[B_n]^{\frac{1}{n}} \leq \int |\nabla f(x)|dx = ||\nabla f||_{L^1}. \tag{6.6}$$

Using the coarea formula we get

$$||\nabla f||_{L^1} = \int_0^{\infty} \mathrm{vol_{n\text{-}1}}[\{x : f(x) = r\}]dr.$$

Thus, approximating h_1 with smooth functions, we find

$$n\,\mathrm{vol_n}[B_n]^{\frac{1}{n}} \leq \mathrm{vol_{n\text{-}1}}[\partial B_n] \cdot \mathrm{vol_n}[B_n]^{\frac{n}{n-1}}.$$

But $n\,\mathrm{vol_n}[B_n]$ is of course equal to $\mathrm{vol_{n\text{-}1}}[\partial B^n]$ and inequality (6.6) has to be sharp. $\square$

Theorem 6.5. *A function $f \in \dot{W}^{1,p}$ is optimal in the Sobolev inequality*

$$\frac{||\nabla f||_{L^p}}{||f||_{L^{p^*}}} \geq ||\nabla h_p||_{L^p}$$

if and only if there exist $C \in \mathbb{R}$, $\lambda \neq 0$ and $x_0 \in \mathbb{R}^n$ such that

$$f(x) = Ch_p(\lambda(x - x_0)),$$

where h_p is defined as in Theorem 6.1.

To prove this theorem we will consider the proof of Theorem 6.1. In this proof we assumed that f and g were smooth functions with compact support, because we were then able to extend our results with a density argument.
Thus we were able to simply apply the integration by parts formula in inequality (6.4). But this kind of reasoning will not be sufficient to proof the characterization of equality cases.

Therefore our first step will be to generalise the proof to all admissible functions f and g without assuming smoothness or compact support.

Lemma 6.6. *Let $f \in \dot{W}^{1,p}(\mathbb{R}^n)$ and $g \in L^{p^*}(\mathbb{R}^n)$ be two nonnegative functions such that $||f||_{L^{p^*}} = ||g||_{L^{p^*}} = 1$ and $\int g^{p^*}(y)|y|^q dy < \infty$. Let $\nabla\phi$ denote the Brenier map pushing the measure with density f^{p^*} forward to the measure with density g^{p^*}. Then $f^{p^*/q}\nabla\phi \in L^q(\mathbb{R}^n)$ and*

$$\frac{1}{n}\int f^{p^*(1-\frac{1}{n})}\Delta_A\phi \leq -\frac{1}{n}\int \nabla(f^{p^*(1-\frac{1}{n})}) \cdot \nabla\phi.$$

Sketch of proof. We have $\int |f^{p^*/q}\nabla\phi|^q = \int g^{p^*}(y)|y|^q dy$ and so $f^{p^*/q}\nabla\phi \in L^q(\mathbb{R}^n)$.

Let Ω be the convex set, where $\phi < \infty$. Without loss of generality we assume that $0 \in \Omega$. For $\epsilon > 0$ we define the function f_ϵ by

$$f_\epsilon(x) = \min\left[f\left(\frac{x}{1-\epsilon}\right), f(x)\chi(\epsilon x)\right],$$

where $\chi : \mathbb{R}^n \to [0,1]$ is a smooth function mapping all x with $|x| < \frac{1}{2}$ to 1 and all x with $|x| \geq 1$ to 0. The support of f_ϵ is compact and contained in Ω since Ω is convex.

Both functions on the right-hand side of the definition of f_ϵ are bounded in $\dot{W}^{1,p}(\mathbb{R}^n)$, uniformly in ϵ. This is obvious for the first one, as it is just a dilation of f. For the second one this can be seen by using the product rule for differentiation and then using the following calculation based on Hölders inequality:

$$\begin{aligned}
\int_{\mathbb{R}^n} f^p(x)|\nabla[\chi(\epsilon x)]|^p dx &= \epsilon^p \int_{\mathbb{R}^n} f^p(x)|\nabla\chi(\epsilon x)|^p dx \\
&\leq \left(\int_{\mathbb{R}^n} [f^p(x)]^{\frac{n}{n-p}}\right)^{\frac{n-p}{n}} \left(\int_{\mathbb{R}^n} [\epsilon^p|\nabla\chi(\epsilon x)|^p]^{\frac{n}{p}}\right)^{\frac{p}{n}} \\
&\leq \left(\int_{\mathbb{R}^n} f^{p^*}(x)\right)^{\frac{n-p}{n}} \left(\int_{\mathbb{R}^n} |\nabla\chi(x)|^n\right)^{\frac{p}{n}}.
\end{aligned}$$

Using the identity $\min(f,g) = \frac{f+g}{2} - \frac{|f-g|}{2}$, we therefore find that f_ϵ lies in $\dot{W}^{1,p}(\mathbb{R}^n)$ and $||\nabla f||_{L^p}$ is bounded as $\epsilon \to 0$.

We now fix $\epsilon > 0$. Let Ω_ϵ be a bounded open set whose closure is contained in Ω and which contains the support of f_ϵ. Let f_ϵ^δ be a sequence of smooth nonnegative functions with compact support satisfying $\mathrm{Supp}(f_\epsilon^\delta) \subset \Omega_\epsilon$ and $f_\epsilon^\delta \to f_\epsilon$ in $\dot{W}^{1,p}(\mathbb{R}^n)$.

Then we can calculate

$$\frac{1}{n}\int (f_\epsilon^\delta)^{p^*(1-\frac{1}{n})}\Delta_A\phi \leq \frac{1}{n}\int (f_\epsilon^\delta)^{p^*(1-\frac{1}{n})}\Delta_{D'}\phi = -\frac{1}{n}\int \nabla((f_\epsilon^\delta)^{p^*(1-\frac{1}{n})}) \cdot \nabla\phi.$$

With some care one can now use the Lemma of Fatou to show that this inequality remains true as $\delta \to 0$ and $\epsilon \to 0$. □

Proof of Theorem 6.5. Without loss of generality we may assume that f is nonnegative as $|f|$ will be also optimal for all optimal functions f and then the conclusion of our theorem will force f to be either $|f|$ or $-|f|$.

Our goal is to prove that $\nabla\phi$ is a dilation-translation map, that is, it satisfies the equation $\nabla\phi = \lambda(\mathrm{Id} - x_0)$ with some positive λ and $x_0 \in \mathbb{R}^n$. To do this we will trace back the equality cases in the proof of Theorem 6.1.

Let Ω be the convex set, where $\phi < \infty$ and let us first show that f is not only nonnegative but that there is a positive constant α_K for each compact subset K of Ω such that $f(x) \geq \alpha_K > 0$ for almost all $x \in K$.

In order to do this, we will consider the use of Hölders inequality in

$$-\int f^{p^\star/q}\nabla f \cdot \nabla\phi \leq ||\nabla f||_{L^p} \cdot ||f^{p^\star/q}\nabla\phi||_{L^q}.$$

For equality to hold it is necessary here that for some positive constant k, $|\nabla f|^p = kf^{p^\star}|\nabla\phi|^q$ for almost every $x \in \Omega$. Define functions f_m by $f_m(x) = \max(f(x), \frac{1}{m})$. Then $\nabla f_m = \nabla f \chi_{\{x: f(x) > 1/m\}}$ and so $\nabla f_m \in L^p$. Thus

$$|\nabla f_m|^p \leq |\nabla f|^p = kf^{p^\star}|\nabla\phi|^q \leq kf_m^{p^\star}|\nabla\phi|^q.$$

As a consequence

$$|\nabla(f_m^{-p/(n-p)})|^p = \left(\frac{p}{n-p}\right)^p f_m^{\frac{-np}{n-p}} |\nabla f_m|^p \leq k \left(\frac{p}{n-p}\right)^p f_m^{\frac{-np}{n-p}} f_m^{p^\star} |\nabla \phi|^q.$$

Since $|\nabla\phi|$ is locally bounded on Ω, it follows that the functions $f_m^{-p/(n-p)}$ are uniformly in m locally Lipschitz and so locally bounded on Ω. So f is locally bounded away from 0 on Ω.

As a second step we prove that $D^2_{D'}\phi$ is absolutely continuous. Since this is a nonnegative matrix valued measure, it will be enough to proof that $\Delta_{D'}\phi$ is absolutely continuous, as the only nonnegative matrix with trace zero is the zero matrix.
Let $\Delta_s\phi$ be the singular part of $\Delta_{D'}\phi$. Since we assume that there is equality in

$$\frac{1}{n}\int F^{1-\frac{1}{n}}\Delta_A\phi \leq \frac{1}{n}\int F^{1-\frac{1}{n}}\Delta_{D'}\phi$$

and $\Delta_{D'}\phi = \Delta_s\phi + \Delta_A\phi$, we see that

$$\int f^{\frac{p(n-1)}{n-p}}\Delta_s\phi = 0.$$

From the proof of (6.6) we deduce that

$$\lim_{\epsilon\to 0}\lim_{\delta\to 0}\int (f_\epsilon^\delta)^{\frac{p(n-1)}{n-p}}\Delta_s\phi = 0, \tag{6.7}$$

where f_ϵ^δ is defined as in the proof above. Without loss of generality we assume that $0 \in \Omega$. Let K be an arbitrary convex compact subset of Ω with 0 in its interior. If we choose ϵ and δ small enough, we can find a positive constant α such that

$$f_\epsilon^\delta \geq \alpha\chi_K$$

using the first part of the proof. As a consequence we see that

$$\int (f_\epsilon^\delta)^{\frac{p(n-1)}{n-p}}\Delta_s\phi \geq \int_K \alpha^{\frac{p(n-1)}{n-p}}\Delta_s\phi = \alpha^{\frac{p(n-1)}{n-p}}\Delta_s\phi[K].$$

Combining this with (6.7) we find that $\Delta_s\phi[K] = 0$ and as K was arbitrary this means that the singular part of $\Delta_{D'}\phi$ vanishes and so $D^2_{D'}\phi$ is absolutely continuous.

Now we can finally prove that $\nabla\phi$ is indeed a dilation-translation map. In order to do this we consider the last inequality we used in the proof of Theorem 6.1. For equality to hold in

$$F^{-\frac{1}{n}}(x)(\det D_A^2\phi(x))^{\frac{1}{n}} \leq F^{-\frac{1}{n}}(x)\frac{\Delta_A\phi(x)}{n}.$$

for almost all $x \in \Omega$, all the eigenvalues of $D_A^2 = D_{D'}^2$ have to be the same for almost all $x \in \Omega$. This obviously means that D_A^2 is a multiple of the identity almost everywhere in

Ω.
To prove that its actually a multiple of the identity we consider a regularizing kernel κ supported on a small ball B_ϵ. As $D_A^2(\phi * \kappa) = D_A^2\phi * \kappa$ we find that the smooth function $D_A^2(\phi * \kappa)$ is proportional to the identity matrix on $\{x \in \Omega : d(x, \partial\Omega)\}$. By making κ go to a dirac mass, we see that $D_A^2(\phi)$ is a multiple of the identity on the whole of Ω and so $\nabla\phi$ and has to be of the form $\lambda(\mathrm{Id} - x_0)$. $\square$

Bibliography

[1] S. Alesker, S. Dar, and V. Milman. A remarkable measure preserving diffeomorphism between two convex bodies in $\mathbb{R}^n$. *Geometriae Dedicata*, 74:201–212, 1999.

[2] Keith Ball. An elementary introduction to monotone transportation. Department of Mathematics, University College London, Gower Street, London.

[3] Franck Barthe. *Inegalites fonctionelles et geometriques obtenues par transport de mesures*. PhD thesis, Univ. Marne-la-Valle, 1994.

[4] Franck Barthe. On a reverse form of the Brascamp-Lieb inequality. *Inventiones mathematicae*, 134:335–361, 1998.

[5] L. Cafarelli. Boundary regularity of maps with a convex potential. *Communications on Pure and Applied Mathematics*, 45:1141–1151, 1992.

[6] L. Cafarelli. Regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5:99104, 1992.

[7] D. Cordero-Erausquin, B. Nazaret, and C. Villani. A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities. *Advances in Mathematics*, 182:307–332, 2004.

[8] Jürgen Elstrodt. *Mass- und Integrationstheorie*. Springer, 2007.

[9] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, 2002.

[10] R. J. Gardner. The Brunn-Minkowski inequality. *Bulletin of the American Mathematical Society*, 39:355–405, 2002.

[11] M. Gromov. Convex sets and Kähler manifolds. In *Advances in differential geometry and topology*, pages 1–38. World Scientific, 1990.

[12] Peter M. Gruber. *Convex and Discrete Geometry*. Springer, 2007.

[13] Robert J. McCann. *A convexity theorie for interacting gases and equilibrium crystals*. PhD thesis, Princeton University, 1994.

[14] R. Tyrrel Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[15] Walter Rudin. *Real and Compex Analysis*. McGraw-Hill, Inc., 1966.

[16] Stefan Ingi Valdimarsson. Optimisers for the brascamp-lieb inequality. *Israel Journal of Mathematics*, 168:253–274, 2008.

[17] Cedric Villan. *Topics in Optimal Transportation.* American Mathematical Society, 2003.