BACHELORARBEIT / BACHELOR'S THESIS

Titel der Bachelorarbeit / Title of the Bachelor's Thesis

Selected Geometric and Gaussian Inequalities and Optimal Transport

verfasst von / submitted by

Alexander Posch

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Bachelor of Science (BSc.)

Wien, September 2020 / Vienna, September 2020

Studienkennzahl lt. Studienblatt/ degree programme code as it appears on the student record sheet:	A 033621
Studienrichtung lt. Studienblatt / field of study as it appears on the student record sheet:	Mathematik / Mathematics
Betreuer / Supervisor:	Univ.-Prof. Dr. Mathias Beiglböck

Acknowledgements

I wish to thank my supervisor for his support and detailed feedback. Special thanks, for his support with the investigation of the conditional Gaussian correlation inequality and involving also Assoc. Prof. Peter Pivovarov who pointed at the counterexample to inequality (8) for centrally symmetric sets. Many thanks also to may father for spirited discussions and my sister who supported me with proofreading.

Contents

Acknowledgements 1
Abstract 3
1. Introduction 3
2. A brief introduction to optimal transportation 4
3. Brunn-Minkowski inequality 5
4. The Gaussian correlation inequality 10
4.1. Applications of Caffarelli's contraction theorem to a special case of the Gaussian correlation inequality 10
4.2. A note on conditional versions of the Gaussian correlation inequality 11
4.2.1. A conditional Gaussian correlation inequality for centrally symmetric hyperrectangles 13
References 16
Appendix A. Python program for Figure 1 and animations. 18
Appendix B. Python program to compute the counterexample in Section 4 22
Appendix C. Python program for Figure 2 and animations. 22

Abstract

Some results from optimal transport are useful for deriving geometric and Gaussian inequalities. In this thesis we review proofs of the Brunn-Minkowski, isoperimetric and a special case of the Gaussian correlation inequality that use results form optimal transport. We start with a short introduction to optimal transport and introduce the displacement interpolation. As an illustrative example we use it to interpolate between two black and white images. Finally, we derive sufficient conditions under witch a conditional version of the Gaussian correlation inequality for hyperrectangles holds. In particular, this condition is satisfied for any centered bivariate normal distribution.

1. Introduction

The optimal transport problem was first studied by the French geometer Gaspard Monge in the 18th century. While his methods were not rigorous by today's standards he did make some valid observations and is considered the founder of the field [17]. Monge considered the problem of how to move earth from a hole to form structures in the most efficient way. More rigorous results were discovered in the early 20th century by the Russian mathematician and economist Leonid Vitaliyevich Kantorovich [17]. He founded the field of linear programming and later received a Nobel prize in economics, alongside Tjalling Charles Koopmans for their work in the field.

From there on the topic has vastly gained popularity and has found applications in a large number of fields as, for example, image registration [9], meteorology [5], and machine learning [12, 8]. Furthermore, two Fields Medals were awarded to researchers studying optimal transportation, Cedric Villani in 2010 and Alessio Figalli in 2018. Today the standard introductory literature on the topic are the books [16, 17] by Cedric Villani which are also the main references for this thesis.

In this thesis we will discuss a few results from optimal transport and their application to prove geometric and Gaussian type inequalities. In Section 2 we introduce the topic of optimal transport along with some of its central results including Brenier's theorem. In Section 3 we discuss the topic of displacement convexity and present a proof of the Brunn-Minkowski inequality, which is subsequently used to prove the isoperimetric inequality. In Section 4 we revisit Cordero-Erausquin's proof [4] of a special case of the Gaussian correlation inequality using Caffarelli's contraction theorem. [2]. Finally we derive sufficient conditions under witch a conditional version of the Gaussian correlation inequality for hyperrectangles holds. In particular, this condition is satisfied for any centered bivariate normal distribution.

2. A brief introduction to optimal transportation

Definition 2.1. Let $(X, \mathcal{F}_X)$ and $(Y, \mathcal{F}_Y)$ be measure spaces, μ a measure on X and $f : X \to Y$ a measurable function. Then, $f\#\mu$ denotes the pushforward measure of f on Y. This measure is given by $f\#\mu(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{F}_Y$.

Let $(X, \mathcal{F}_X, \mu)$ and $(Y, \mathcal{F}_Y, \nu)$ be two probability spaces. An admissible transference plan is any measure π on $X \times Y$ with marginals μ and ν, here $\pi(A, B)$ denotes the amount of measure transported from A to B for all $A \in \mathcal{F}_X, B \in \mathcal{F}_Y$. Furthermore, we define a cost function $c : X \times Y \to \mathbb{R}$, where $c(x, y)$ denotes the cost of transporting one unit of mass from x to y. The cost of a transference plan π is given by

$$I[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y).$$

Minimizing this transportation cost over the set of all admissible transference plans is known as Kantorovich's optimal transport problem.
A special case of the Kantorovich problem was proposed earlier by Monge. Monge considered transport maps $T : X \to Y$ such that $\nu = T\#\mu$. Here $T(x)$ denotes the location the mass in x is moved to. Thus, the mass in each point is moved to exactly one other point and must not be split up. Naturally every transport map T induces a transference plan by setting $\pi = (Id \times T)\#\mu$, and a transport map is considered optimal if it minimizes the cost of the associated transference plan over the set of all possible transport maps.
In general the solutions of Monge's problem and the Kantorovich problem do not have to coincide. There are even cases where no transportation map exists, but under suitable assumptions on μ, ν and c one can prove the existence of such a transport map and that the solutions of the two problems coincide (see Theorem 1).

Definition 2.2. $P_{2,ac}(\mathbb{R}^n)$ is defined as the set of probability measures on $\mathbb{R}^n$ equivalent to the Lebesgue measure with finite second moment.

Theorem 1 (Brenier's theorem)**.** *[16, Theorem 2.12] Let $\mu, \nu \in P_{2,ac}(\mathbb{R}^n)$ and c denote the quadratic cost function $c(x, y) = \|x - y\|^2$. Then, the Kantorovich problem coincides with Monge's problem and there exists a unique gradient of a convex function $\nabla\varphi$ such that $\nu = \nabla\varphi\#\mu$. This transport map is called the Brenier map. It is the unique solution to Monge's problem and the associated transference plan is the unique solution to the Kantorovich problem.*

For a proof see [16][Chapter 2]. To finish this introduction we will look at Brenier maps between probability measures on $\mathbb{R}$.

Proposition 1. *Let $\mu, \nu \in P_{2,ac}(\mathbb{R})$ and let Γ_μ, Γ_ν be their respective cumulative distribution functions. Then, the Brenier map from μ to ν is given by $\Gamma_\nu^{(-1)} \circ \Gamma_\mu$ where $\Gamma_\nu^{(-1)}(x) = \inf(r \in \mathbb{R} | \Gamma_\nu(r) \geq x)$.*

Proof. Since $\Gamma_\nu^{(-1)} \circ \Gamma_\mu$ is increasing and continuous in all but countably many points it is almost everywhere the gradient of a convex function. Thus according to Brenier's theorem, it only remains to verify that $\nu = \Gamma_\nu^{(-1)} \circ \Gamma_\mu \# \mu$. Let $b \in \mathbb{R}$ then $\Gamma_\nu^{(-1)} \circ \Gamma_\mu \# \mu((-\infty, b]) = \mu(\Gamma_\mu^{-1} \circ (\Gamma_\nu^{(-1)})^{-1}((-\infty, b]) = \mu(\Gamma_\mu^{-1}([0, \Gamma_\nu(b)]) = \mu((-\infty, sup\{r \in \mathbb{R} | \Gamma_\mu(r) = \Gamma_\nu(b)\}]) = \nu((-\infty, b])$. ☐

3. Brunn-Minkowski inequality

In this section we will revisit the proof of the Brunn-Minkowski inequality, presented in [16, Chapter 6] and [11]. This proof for the Brunn-Minkowski inequality relies on displacement-convexity, so we will start by proving the necessary results.

Definition 3.1. [11, 16, p. 148] Let μ and $\nu = T\#\mu$ be two probability measures with T the optimal transport map between the two measures. Then, $[\mu, \nu]_t := (1-t)id + t\nabla\varphi\#\mu$ denotes the displacement interpolation between μ and ν.

Brenier's theorem implies that the displacement interpolation is well defined for the quadratic cost function and any two probability measures with finite second moments that are absolutely continuous with respect to the Lebesgue measure. Note that the measure $[\mu, \nu]_t$ is a probability measure for all $t \in [0, 1]$ and is absolutely continuous if either μ or ν is absolutely continuous. [16, Proposition 5.9]

Lemma 1. *[16, Lemma 6.3] Let μ and ν be two probability measures on $\mathbb{R}^n$ absolutely continuous with compact supports X and Y. Then, supp$([\mu, \nu]_t) \subseteq (1-t)X + tY$, where $(1-t)X + tY = \{(1-t)x + ty | x \in x, y \in Y\}$ is the Minkowski sum.*

Proof. Since $[\mu, \nu]_t = (1-t)id + t\nabla\varphi\#\mu$, $s \in \operatorname{supp}[\mu, \nu]_t \Rightarrow ((1-t)id + t\nabla\varphi)^{-1}(s) = x \in X$ and therefore $s = (1-t)x + t\nabla\varphi(x)$. The result now follows since $\nabla\varphi(x) \in Y$ ☐

The next proposition holds true for any two probability measures on $\mathbb{R}^n$, that are absolutely continuous with respect to the Lebesgue measure. But since the proof is quite technical we will only examine the case where $\nu = T\#\mu$ and T is a diffeomorphism. For a proof of the general case, see [16, Theorem 4.8].

Proposition 2. *Let μ be a probability measure on $\mathbb{R}^n$ with density p. Let $\nu = \nabla\varphi\#\mu$ where φ is a convex function and $\nabla\varphi$ is a diffeomorphism. Let $\gamma_t = (1-t)id + t\nabla\varphi$. Then the density of $[\mu,\nu]_t$ is given by $\frac{p\circ\gamma_t^{-1}}{|\det((\nabla\gamma_t)\circ\gamma_t^{-1})|}$ in $\gamma_t(\mathbb{R}^n)$ and by 0 in $\mathbb{R}^n\backslash\gamma_t(\mathbb{R}^n)$.*

Proof. Let $A \subseteq \gamma_t(\mathbb{R}^n)$. Then,

$$[\mu,\nu]_t(A) = \mu(\gamma^{-1}(A)) = \int_{\gamma_t^{-1}(A)} p(x)dx = \int_A p(\gamma_t^{-1}(x))\left|\det\nabla\gamma_t^{-1}(x)\right|dx$$
$$= \int_A \frac{p(\gamma_t^{-1}(x))}{|\det\nabla\gamma_t(\gamma_t^{-1}(x))|}dx.$$

□

As an example consider $\mu(x) = 2x\chi_{[0,1]}$ and $\nu = 0.5\chi_{[0,2]}$. Proposition 1 tells us that $T(x) = 2x^2$ is the Brenier map from μ to ν. Thus, the displacement interpolation is given by $[\mu,\nu]_t = t2id^2 + (1-t)id\#\mu$. From Proposition 2 we know that the density is given by $[\mu,\nu]_t(x) = \frac{\chi_{[0,1]}(x)\sqrt{2x}}{2t\sqrt{2x}+(1-t)}$. As illustrated in Figure 1 this results in a smooth transition between the two measures.

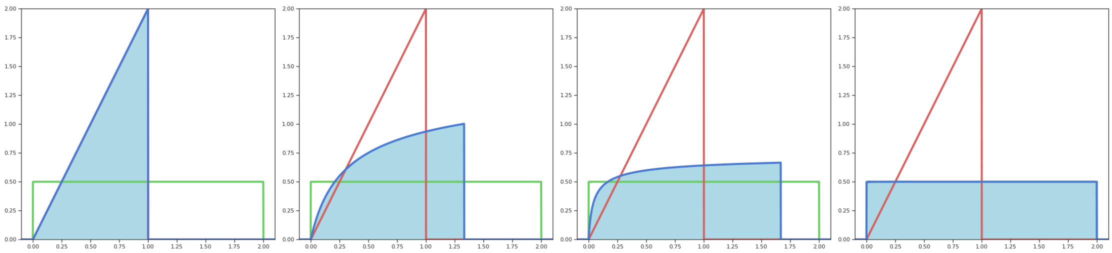

FIGURE 1. Displacement interpolation between the measures $\mu(x) = 2x\chi_{[0,1]}$ and $\nu(x) = 0.5\chi_{[0,2]}$. The figure shows $[\mu,\nu]_t$ for $t = 0, \frac{1}{3}, \frac{2}{3}, 1$. An animation for this example can be found at https://youtu.be/xX4XFLP5ijc. The graphs and the animation were created with the Python program in Appendix A. The program can numerically calculate this interpolation for any two measures on $\mathbb{R}$.

.

For a visually appealing example, we look at case of a discrete distribution. Consider two black and white images with the same number of black pixels. For a picture with K black pixels, those can be interpreted as distributions on the grid where $P(x) = 0$ if x is a white pixel and $P(x) = 1/K$ if x is a black pixel. Using

POT's (Pythons Optimal transport module[7]) emd (earth movers distance) function we can find the discrete optimal transport map from one picture to the other where we are transporting the black pixels with regard to a quadratic cost function. The algorithm used by the emd function is explained in [1]. As illustrated in Figure 2 the displacement interpolation gives a smooth interpolation between the two images. The Python code can be found in Appendix C.

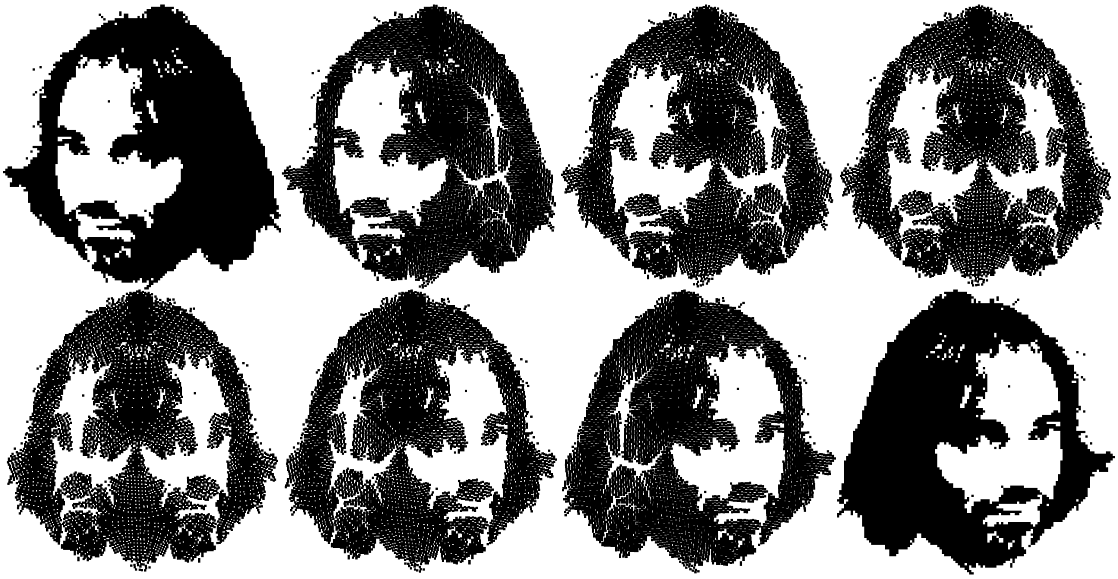

FIGURE 2. The figure shows the optimal transport between a picture of the mathematician Cédric Villani and it's reflection. The black pixels are transported in accordance to minimize transportation cost with regard to a quadratic cost function. The picture was adapted from the photograph [3] using the program Gimp. For an animation see https://youtu.be/-esQj8i3-ZQ. Further examples can be found at https://www.youtube.com/playlist?list=PLH81iF4KWa50-6_xGO3gFPI8dALhoF6f_.

.

Definition 3.2. [11, 16, Definition 5.10 ii] A functional F is called displacement convex if the function $t \to F([\mu,\nu]_t)$ is convex for any probability measures μ and ν absolutely continuous with respect to the Lebesgue measure.

Lemma 2. *[16, Lemma 5.23 (ii)] Let $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix then $\psi(t) := |\det(1-t)I + tA|^{1/n}$ is concave for $t \in [0,1]$.*

Proof. To prove concavity we show that for $a, b \in [0,1]$ we get $(1-t)\psi(a)+t\psi(b) \geq \psi((1-t)a+tb)$ for all $t \in [0,1]$.
We first note that for all $s \in [0,1]$ $(1-s)I + sA$ is a positive definite matrix, since the sum of two positive definite matrices is again positive definite. Furthermore, for all $t \in [0,1]$

$$[1-((1-t)a+tb)]I + [(1-t)a+tb]A = (1-t)[(1-a)I + aA] + t[(1-b)I + bA].$$

Therefore it is sufficient to prove that for all positive definite matrices A, B,

$$|\det A + B|^{1/n} \geq |\det A|^{1/n} + |\det B|^{1/n}. \tag{1}$$

Because of continuity it suffices to prove the assertion for A invertible. In this case we can multiply both sides of the equation (1) by $|\det(A^{-1})|^{1/n}$. If we now set $C := A^{-1}B$ then C is positive definite and equation (1) becomes

$$|\det I + C|^{1/n} \geq 1 + |\det C|^{1/n}. \tag{2}$$

Now let $c_1, \ldots, c_n \geq 0$ denote the eigenvalues of C then the eigenvalues of $I+C$ are $c_1+1, c_2+1, \ldots, c_n+1$. Since the determinant is the product of the eigenvalues, equation (2) simplifies to

$$\prod_{i=0}^{n}(c_i+1)^{1/n} \geq 1 + \prod_{i=0}^{n} c_i^{1/n} \tag{3}$$

and dividing by the left hand side yields

$$1 \geq \prod_{i=0}^{n}\left(\frac{1}{c_i+1}\right)^{1/n} + \prod_{i=0}^{n}\left(\frac{c_i}{c_i+1}\right)^{1/n}. \tag{4}$$

This follows from the arithmetic geometric mean inequality since

$$\prod_{i=0}^{n}\left(\frac{1}{c_i+1}\right)^{1/n} + \prod_{i=0}^{n}\left(\frac{c_i}{c_i+1}\right)^{1/n} \leq \frac{1}{n}\sum_{i=0}^{n}\frac{1}{c_i+1} + \frac{1}{n}\sum_{i=0}^{n}\frac{c_i}{c_i+1} = 1.$$

$\square$

Remark. Note that ϕ_s is convex for all $s \geq 0$ if ϕ_1 is convex, where $\phi_s : \mathbb{R}^+ \to \mathbb{R}$ with $\phi_s(t) = t^n U(st^{-n})$ and $U(0) = 0$. For $s > 0$, this is a direct consequence of the convexity of $s\,\phi_s$.

Proposition 3. *[16, Theorem 5.15 (i)] Let $U : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$ measurable with $U(0) = 0$ and $\phi_1(t) = t^n U(t^{-n})$ be convex and non-increasing. Then, the functional*

$\mathcal{U}(p) = \int_{\mathbb{R}^n} U(p(x))dx$, *where the probability measure p is identified with its density, is displacement convex.*

Proof. Let μ, ν be two absolutely continuous probability measures. Then, we have to show that $t \to \mathcal{U}([\mu,\nu]_t)$ is convex. Since $\nabla\gamma_t = (1-t)I + t\nabla^2\varphi$, with $\nabla^2\varphi$ positive definite (because φ is convex), it follows from Lemma 2 that, $\psi_x(t) = |\det \nabla\gamma_t(x)|^{1/n}$ is concave in t. It now follows from Proposition 2 that

$$\begin{aligned}\mathcal{U}([\mu,\nu]_t) &= \int_{\mathbb{R}^n} U\left(\frac{p(\gamma_t^{-1}(x))}{|\det \nabla\gamma_t(\gamma_t^{-1}(x))|}\right)dx = \int_{\mathbb{R}^n} U\left(\frac{p(x)}{|\det \nabla\gamma_t(x)|}\right)|\det \nabla\gamma_t(x)|dx \\ &= \int_{\mathbb{R}^n} U\left(\frac{p(x)}{\psi_x(t)^n}\right)\psi(x,t)^n dx = \int_{\mathbb{R}^n} \phi_{p(x)} \circ \psi_x(t)dx.\end{aligned}$$

The result now follows from $\phi_a \circ \psi_b$ being convex for all $a, b > 0$, since ϕ_a is convex non-increasing and ψ_b is concave. $\square$

We now have all the tools needed to prove the Brunn-Minkowski inequality

Theorem 2 (Brunn-Minkowski inequality)**.** *[16, p. 184] Let $X, Y \subset \mathbb{R}^n$ compact, and let $|X|$ denote the Lebesgue measure of X. Then, the following inequality holds for the Minkowski sum:*

$$|X+Y|^{1/n} \geq |X|^{1/n} + |Y|^{1/n}$$

Proof. Let $\mu = \frac{\chi_X}{|X|}$ and $\nu = \frac{\chi_Y}{|Y|}$ be defined via their densities. Let $U(x) = -|x|^{1-1/n}$. Then $U(0) = 0$ and $t \to t^n U(t^{-n}s) = -t^n t^{-n+1}|s|^{1-1/n} = -t|s|^{1-1/n}$ is convex and non-increasing in t. Therefore $\mathcal{U}(p) = \int_{\mathbb{R}^n} U(p(x))dx$ is displacement convex according to Proposition 3. Furthermore, note that $\mathcal{U}(\frac{\chi_X}{|X|}) = -|X|^{1/n}$. Therefore

$$(1-t)|X|^{1/n} + t|Y|^{1/n} = -(1-t)\mathcal{U}(\mu) - t\mathcal{U}(\nu) \leq -\mathcal{U}([\mu,\nu]_t).$$

It therefore just remains to show that $-\mathcal{U}([\mu,\nu]_t) \leq |(1-t)X + tY|^{1/n}$. This is a consequence of Jensen's inequality. Set $S_t = (1-t)X + tY$ then

$$\mathcal{U}([\mu,\nu]_t) = \int_{S_t} U([\mu,\nu]_t(x))dx \geq |S_t|\, U\left(\frac{1}{|S_t|}\int_{S_t} d[\mu,\nu]_t\right) = |S_t|\, U(\frac{1}{|S_t|}) = -|S_t|^{1/n}.$$

$\square$

The Brunn-Minkowski inequality can now be used to prove the well known isoperimetric inequality.

Theorem 3 (Isoperimetric inequality)**.** *[16, p.185] Let A be a compact set in $\mathbb{R}^n$ and $per(A)$ the surface area of A then the isoperimetric inequality states that*

$$per(A) \geq n\,|A|^{\frac{n-1}{n}}\,|B_1|^{\frac{1}{n}}\,. \tag{5}$$

Proof. Since $per(A) = \liminf_{\varepsilon\to 0}\frac{|A+B_e|-|A|}{\varepsilon}$ and since for all $\varepsilon > 0$

$$\frac{|A+B_e|-|A|}{\varepsilon} \geq \frac{(|A|^{\frac{1}{n}}+|B_\varepsilon|^{\frac{1}{n}})^n-|A|}{\varepsilon} \geq \frac{n\,|A|^{\frac{n-1}{n}}\,|B_\varepsilon|^{\frac{1}{n}}}{\varepsilon} = n\,|A|^{\frac{n-1}{n}}\,|B_1|^{\frac{1}{n}}$$

the inequality also holds for the infimum and the proof is complete. □

Remark. Since $|B_1| = per(B_1)n$ we get

$$\frac{per(A)}{per(B_1)} \geq \left(\frac{|A|}{|B_1|}\right)^{\frac{n-1}{n}}.$$

In this notation it can be easily seen that the unit sphere has the lowest perimeter to volume ratio of any compact set with its volume.

4. The Gaussian correlation inequality

The Gaussian correlation inequality is a famous inequality in probability and convex geometry. It was first conjectured in 1955 [6] but it was proven only recently in 2014 by T. Royen [13]. The inequality states that for any two convex centrally symmetric sets A, B

$$\gamma(A\cap B) \geq \gamma(A)\gamma(B) \tag{6}$$

where γ is a centred multivariate normal distribution. While the proof by T. Royen is relatively short and proves an even more general statement, it mainly relies on an argument using the Laplace transform and uses hardly any optimal transport. Therefore we will consider the case where one of the sets is an ellipsoid. In this case Cordero-Erausquin [4] found that the result follows directly from Caffarelli's contraction theorem.

4.1. **Applications of Caffarelli's contraction theorem to a special case of the Gaussian correlation inequality.** In this subsection we state an important result by L. Caffarelli [2]. We will then discuss its applications to the Gaussian correlation inequality [4].

Theorem 4 (Caffarelli's Contraction Theorem)**.** *[2] Let $\mu = e^{-Q(x)}$ and $\nu = e^{-(Q(x)+F(x))}$ be two probability measures on $\mathbb{R}^n$, such that Q is a non negative* *quadratic polynomial and F is convex. Then, the Brenier map pushing μ to ν is a contraction*

Definition 4.1. A set A is said to be centrally symmetric if $A = -A := \{-a | a \in A\}$. A measure μ is said to be centrally symmetric if $\mu = -id\#\mu$.

Theorem 5. *Let γ denote the multivariate Gaussian with the covariance matrix C. Given two centrally symmetric sets in $\mathbb{R}^n$, A convex and B an ellipsoid, then the Gaussian correlation inequality holds:*

$$\gamma_A(B) = \frac{\gamma(A \cap B)}{\gamma(A)} \geq \gamma(B) \tag{7}$$

Proof. Since B is an ellipsoid, we can apply a linear transformation L such that $L(B) = B_1$ the unit sphere. The measure on the transformed space is still a multivariate Gaussian but with the covariance matrix LCL^T. Furthermore, $L(A)$ is still a convex set. Therefore it suffices to examine the case where B is the unit sphere B_1. If we now define $F(x) = \log \gamma(A)$ for $x \in A$ and $F(x) = +\infty$ for $x \notin A$, then $\gamma_A = e^{-(Q+F)}$. Since F is convex the Brenier map T from γ to γ_A is a contraction. Since A is centrally symmetric it follows from Lemma 3 that $T(0) = 0$ and therefore $T(B_1) \subseteq B_1$. It now follows that $\gamma_A(B_1) = \gamma(T^{-1}(B_1)) \geq \gamma(B_1)$ $\square$

Lemma 3. *[4] Let μ and $\nu = T\#\mu$ be any two centrally symmetric measures, where T is the Brenier map. Then, $T(0) = 0$.*

Proof. Let r be any linear isometry. We will show that $T_r(x) := r^{-1}(Tr((x)))$ is the Brenier map pushing $r^{-1}\#\mu$ towards $r^{-1}\#\nu$. Since $r^{-1}Tr\#(r^{-1}\#\mu) = r^{-1}T\#\mu = r^{-1}\#(T\#\mu) = r^{-1}\#\nu$ and by the uniqueness of the Brenier map it suffices to show that T_r is the gradient of a convex function. Since T is a Brenier map it is the gradient of a convex function ϕ. If we define $\phi_r(x) = \phi(r(x))$ then we get $\nabla\phi_r(x) = r^*\nabla\phi(r(x)) = r^{-1}T(r(x)) = T_r(x)$. Because of the uniqueness of the Brenier map and $\mu = -id\#\mu$ and $\nu = -id\#\nu$, it follows that $T = T_{-id}$ and therefore $T(0) = -T(-0)$ and we get the desired result. $\square$

4.2. **A note on conditional versions of the Gaussian correlation inequality.** When reading through the proof one could ask oneself if, or under what assumptions a conditional version of the inequality holds.

$$\gamma(A \cap B | C) \geq \gamma(A|C)\gamma(B|C) \tag{8}$$

Note that for $A, B \subseteq C$ this can be rewritten as

$$\gamma(A \cap B)\gamma(C) \geq \gamma(A)\gamma(B) \tag{9}$$

For example consider the case when $A, B, C \subset \mathbb{R}^n$ are convex centrally symmetric sets and γ is any Gaussian measure. In this case the inequality (8) does not hold. A counterexample is given for $n = 2$ and γ the standard multivariate normal distribution. For an illustration see Figure 3. Let A, B be rectangles with lengths a_1, a_2 and widths b_1, b_2, such that $a_1 < b1$ and $b_2 < a_2$. If we choose C as the smallest rectangle containing $A \cup B$ than C has length b_1 and width a_2 while $A \cap B$ has length a_1 and width b_2. Inserting these values into the equation (9) we get $\gamma(|x| \leq \frac{a_1}{2}, |y| \leq \frac{b_2}{2})\gamma(|x| \leq \frac{a_2}{2}, |y| \leq \frac{b_2}{2}) \geq \gamma(|x| \leq \frac{a_1}{2}, |y| \leq \frac{a_2}{2})\gamma(|x| \leq \frac{b_1}{2}, |y| \leq \frac{b_2}{2})$. But since γ is the standard multivariate normal distribution x and y are independent. Separating the terms shows that the proposed inequality is an equality for our choice of A, B, C. Equation (9) now implies that for any convex set $\widetilde{C} \subset C$ such that $A \cup B \subseteq \widetilde{C}$ and $\gamma(\widetilde{C}) < \gamma(C)$ the inequality does not hold. Therefore choosing $\widetilde{C} = \mathrm{Conv}(A \cup B)$ the convex hull of $A \cup B$ gives an counterexample to the inequality (8).

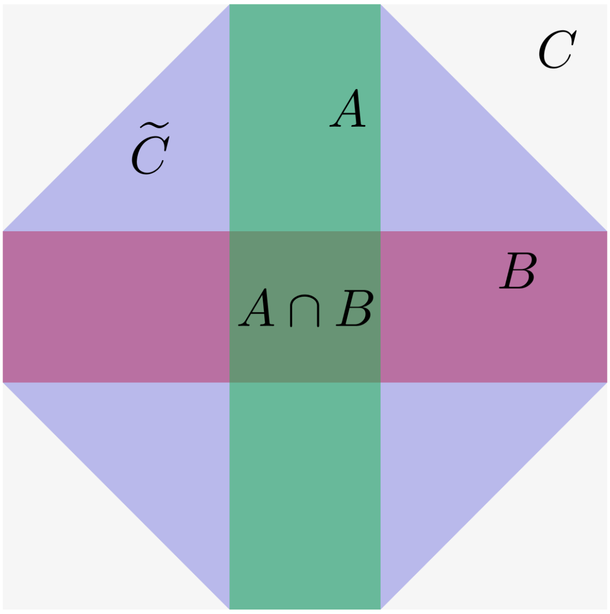

FIGURE 3. This figure illustrates the counterexample to the conditional Gaussian correlation inequality in (8), where A, B, C are centrally symmetric sets in $\mathbb{R}^n$. The picture was created with the program Geogebra.

4.2.1. *A conditional Gaussian correlation inequality for centrally symmetric hyperrectangles.*

Definition 4.2. We will slightly, abuse notation and identify vectors and hyperrectangles. A centrally symmetric hyperrectangle $A = (a_1, \dots, a_n)$ is defined as the set $\{(x_1, \dots, x_n) \in \mathbb{R}^n | \forall i \, a_i \geq |x_i|\}$. $P_\Sigma(A)$ denotes the measure of the set (hyperrectangle) defined by A under a centered Gaussian with covariance matrix Σ.

Definition 4.3. [10] For two vectors $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_n)$ we define $A \wedge B = (min(a_1, b_1), \dots, min(a_n, b_n))$ and $A \vee B = (max(a_1, b_1), \dots, max(a_n, b_n))$.

Remark. Note that for hyperrectangles $A = (a_1, a_2, \dots, a_n)$ and $B = (b_1, b_2, \dots, b_n)$ we have $A \cap B = A \wedge B$ and it suffices to consider $C = A \vee B$ the smallest hyperrectangle containing both A and B.

Even when we assume that A, B, C are centrally symmetric hyperrectangles, the inequality (8) does not hold for all multivariate Gaussian distributions γ. We show this by giving a counterexample in $\mathbb{R}^3$.

Example. *Consider the centered Gaussian on $\mathbb{R}^3$ with the covariance matrix*

$$\Sigma = \begin{pmatrix} 1 & 0.85 & -0.78 \\ 0.85 & 1 & -0.36 \\ -0.78, & -0.36 & 1 \end{pmatrix}$$

and $A = (1.4, 0.8, 2.3), B = (1.4, 2.1, 1.0)$ two cuboids then $A \cap B = (1.4, 0.8, 1.0)$ and $C = (1.4, 2.1, 2.3)$. Using Pythons scipy stats multivariate normal package one can determine $P_\Sigma(A)P_\Sigma(B) = 0.3395$ and $P_\Sigma(A \cap B)P_\Sigma(C) = 0.3546$ (rounded to four decimals). Thus, the inequality in Equation (9) does not hold.

Definition 4.4. [10] A function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is called multivariate totally positive of order 2 (MTP2) on D if $f(X)f(Y) \leq f(X \vee Y)f(X \wedge Y)$ for all $X, Y \in D$.

For a multivariate normaly distributed random variable X we define $\overline{\Phi}$ as the cumulative distribution function of the random variable $|X|$. Then we have $P(A) = \overline{\Phi}(A)$. Thus, (9) simplifies to the MTP2 condition for $\overline{\Phi}$.

$$\overline{\Phi}(A)\overline{\Phi}(B) \leq \overline{\Phi}(A \vee B)\overline{\Phi}(A \wedge B).$$

Theorem 6. *Consider a centered Gaussian with a covariance matrix Σ such that there exists a diagonal Matrix D with diagonal elements in $(-1,1)$ such that all non diagonal entries of $-D\Sigma^{-1}D$ are non negative. (In particular, this condition is satisfied for any centered bivariate normal distribution.) Then the inequality (8) holds for centrally symmetric hyperrecangles $A, B, C = A \vee B$.*

This is a consequence of Theorem 3.1 in [10] by Karlin and Rinott where they prove the condition of this theorem to be a necessary and sufficient condition on the covariance matrix Σ of a centered normal distribution for the density of it's absolute values to be MTP2. Thus it only remains to show that the cumulative distribution function of a MTP2 density is again MTP2. This property is mentioned in [14, p. 7] and [15, p. 6], first of which is a recent preprint from June 2020 by Royen where he proves a more general version of this theorem. I give an independent proof that is elementary and relies on L'Hopitals monotone theorem.

Lemma 4. *Let $f : \mathbb{R}_+^n \to (0,\infty)$ then the following statements are equivalent:*

(1) *f is MTP2.*
(2) *If $x_j \geq y_j$ for all $j \in [1,\dots,n]$ then $\frac{f(x_1,\dots,x_i=s,\dots,x_n)}{f(y_1,\dots,y_i=s,\dots,y_n)}$ is increasing in s for all $i \in [1,\dots,n]$.*

Proof. Let $X, Y \in \mathbb{R}_+^n$ arbitrary and $S := max(Y-X,0)$ then the MTP2 condition rewrites to

$$f(X+S)f(X \wedge Y) \geq f(X)f(X \wedge Y + S)$$

or equivalently

$$\frac{f(X+S)}{f(X \wedge Y + S)} \geq \frac{f(X)}{f(X \wedge Y)} \tag{10}$$

(2)⇒(1): When f satisfies (2) the inequality (10) clearly holds and thus f is MTP2.
(1)⇒(2): This is a direct consequence of the inequality (10) by setting $X = (x_1,\dots,x_n)$ and $Y = (y_1,\dots,x_i+s,\dots,y_n)$ with $x_j \geq y_j$ for $j \neq i$ and $s \geq 0$. □

Proposition 4. *Let $F(x_1,\dots,x_n) = \int_0^{x_i} f(x_1,\dots,x_i=r,\dots,x_n)dr$ be the integral of a strictly positive MTP2 function f. Then, F is also MTP2.*

Proof. W.l.o.g we can assume $i=1$. Let $X = (x_1,\dots,x_n), Y = (y_1,\dots,y_n) \in \mathbb{R}_+^n$ such that $y_j \leq x_j$ for all $j \in [1,\dots,n]$

Since f is MTP2 we know from Lemma 4 that $\frac{f(r,x_2,..x_n)}{f(r,y_2,...,y_n)}$ is increasing in r. L'Hopital's monotone rule implies that also

$$\frac{\int_0^{\overline{r}} f(r, x_2, \ldots, x_n)dr}{\int_0^{\overline{r}} f(r, y_2, \ldots, y_n)dr} \text{ is increasing in } \overline{r}.$$

This simplifies to $\frac{F(\overline{r},x_2,...,x_n)}{F(\overline{r},y_2,...,y_n)}$ being increasing in $\overline{r}$. Thus, by Lemma 4 F is MTP2. □

Proof of Theorem 6. From Theorem 3.1 in [10] by Karlin and Rinott we know that, given the conditions, the density of the absolute values is MTP2. Furthermore, since that density is strictly positive we know from Proposition 4 that also its integral with regard to any variable is MTP2 and clearly strictly positive on $\mathbb{R}_+^n$. Thus, attractively integrating over all variables yields that also the cumulative distribution function of the absolute values is MTP2. Therefore, both the inequality (9) and the equivalent inequality (8) hold. □

References

[1] Nicolas Bonneel et al. "Displacement interpolation using Lagrangian mass transport". In: *Proceedings of the 2011 SIGGRAPH Asia Conference*. 2011, pp. 1–12.

[2] Luis A Caffarelli. "Monotonicity Properties of Optimal Transportation¶ and the FKG and Related Inequalities". In: *Communications in Mathematical Physics* 214.3 (2000), pp. 547–563.

[3] Marie-Lan Nguyen / Wikimedia Commons. *Cedric_Villani_at_his_office_2015_n3.jpg*. 2015 (accessed September 11, 2020). URL: https://de.wikipedia.org/wiki/C%5C%C3%5C%A9dric_Villani#/media/Datei:Cedric_Villani_at_his_office_2015_n3.jpg.

[4] Dario Cordero-Erausquin. "Some applications of mass transport to Gaussian-type inequalities". In: *Archive for rational mechanics and analysis* 161.3 (2002), pp. 257–269.

[5] Mike Cullen and Hamed Maroofi. "The fully compressible semi-geostrophic system from meteorology". In: *Archive for rational mechanics and analysis* 167.4 (2003), pp. 309–336.

[6] C. W. Dunnett and M. Sobel. "Approximations to the Probability Integral and Certain Percentage Points of a Multivariate Analogue of Student's t-Distribution". In: *Biometrika* 42.1/2 (1955), pp. 258–260. ISSN: 00063444. URL: http://www.jstor.org/stable/2333441.

[7] R'emi Flamary and Nicolas Courty. *POT Python Optimal Transport library*. 2017. URL: https://pythonot.github.io/.

[8] Aude Genevay et al. "Stochastic optimization for large-scale optimal transport". In: *Advances in neural information processing systems*. 2016, pp. 3440–3448.

[9] Steven Haker et al. "Optimal mass transport for registration and warping". In: *International Journal of computer vision* 60.3 (2004), pp. 225–240.

[10] Samuel Karlin, Yosef Rinott, et al. "Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities". In: *The Annals of Statistics* 9.5 (1981), pp. 1035–1049.

[11] Robert J McCann. "A convexity principle for interacting gases". In: *Advances in mathematics* 128.1 (1997), pp. 153–179.

[12] Gabriel Peyré, Marco Cuturi, et al. "Computational Optimal Transport: With Applications to Data Science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.

[13] Thomas Royen. "A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions". In: *arXiv preprint arXiv:1408.1028* (2014).

[14] Thomas Royen. "Some improved Gaussian correlation inequalities for symmetrical n-rectangles extended to some multivariate gamma distributions and some further probability inequalities". In: *arXiv preprint arXiv:2006.00769* (2020).

[15] Sanat K Sarkar et al. "On the Simes inequality and its generalization". In: *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*. Institute of Mathematical Statistics, 2008, pp. 231–242.

[16] C. Villani. "Topics in Optimal Transportation Theory". In: 58 (Jan. 2003). DOI: 10.1090/gsm/058.

[17] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media, 2008.

Appendix A. Python program for Figure 1 and animations.

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.animation as animation
import seaborn as sns
from matplotlib.animation import FuncAnimation
from scipy.interpolate import interp1d
from PIL import Image
sns.set(style="ticks", palette="muted")

def nu(x):#density of nu
    return 2*x*(0<=x)*(x<=1)
def cdfnu(x):#cdf of nu
    return (x>=0)*np.minimum(x,1)**2
def icdfnu(t):#inverse of cdf of nu
    return(t**(1/2))

def mu(x):#density of mu
    return 1/2*(0<=x)*(x<=2)
def cdfmu(x):#cdf of mu
    return (x>=0)*np.minimum(x,2)*1/2
def icdfmu(t):#inverse of cdf of mu
    return 2*t


#Coordinates for graphs
left,right,up=-0.001,2.001,2.001,
Precision=10000
Name="nu_to_mu_displacement.mp4"

def T(s):#Transport map from nu to mu
    Ts=icdfmu(cdfnu(s))
    return np.array([list(zip(s,Ts))[i] for i in range(len(s))])
```

```
def Mcint(t,x): #(1-t)*id+t*T
    return (1-t)*x+t*T3(x)

def iMcint(t,x,y):#inverse of Mcint
    return np.array([x[np.where(np.abs(Mcint(t,x)-y[i])==\
        np.amin(np.abs(Mcint(t,x)-y[i])))[0][0]]for i in range(len(y))])

def DMcint(t):#Derivative of Mcint
    g=0.0001#precision
    return lambda x: (Mcint(t,x+g) -Mcint(t,x-g))/(2*g)

def density(t,x,s):#Density of displacement interpolation
    iM=iMcint(t,x,s)
    return nu(iM)/np.abs(DMcint(t)(iM))

x = np.linspace(left-1,right+1,Precision)
Tx=T(x)
xt,Tx=Tx.T
T3= interp1d(list(xt)+[-1,4*np.pi+1], list(Tx)+[0,1])
x = np.linspace(left-0.1,right+0.1,Precision)


fig = plt.figure()
fig.set_size_inches(9,9, True)
ax = plt.axes(xlim=(left-0.1,right+0.1),ylim=(0, up),aspect="equal")
linemu, = ax.plot([], [], "g",lw=4)
linenu, = ax.plot([], [], "r",lw=4)
liner,= ax.plot([], [], "b",lw=4)
line, = ax.fill([], [], "lightblue")#,lw=4,edgecolor="b")
def init():
    linemu.set_data([], [])
    return linemu,
def animate(i):
```

```
    if i%10==0:
        print(i)
    x = np.linspace(left-0.1,right+.1, 2000)
    y = density(min(0.999,i/199),x,x)
    line.set_xy(np.array([x,y]).T)
    liner.set_data(x,y)
    linenu.set_data(x,nu(x))
    linemu.set_data(x,mu(x))
    plt.savefig("/home/alexander/pythonimages/d"+str(i)+".png",\
        bbox_inches='tight')
    return line,linemu,linenu, liner

anim = FuncAnimation(fig, animate, init_func=init,
                               frames=200, interval=20, blit=True)

anim.save(Name)

#Animation comparing linear and displacement interpolation.

fig, axs = plt.subplots(1, 2, figsize=(16, 8), sharey=True)
#axs[0] = plt.axes(xlim=(left-0.1,right+0.1),ylim=(0, up),aspect="equal")
axs[0].set_xlim((left-0.1,right+0.1))
axs[0].set_ylim((0, up))
axs[0].set_aspect("equal")
axs[0].set_title("McCann Displacement Interpolation",fontdict = {'fontsize'
: 20})

linemu, = axs[0].plot([], [], "g",lw=4)
linenu, = axs[0].plot([], [], "r",lw=4)
liner,= axs[0].plot([], [], "b",lw=4)
line, = axs[0].fill([], [], "lightblue")#,lw=4,edgecolor="b")

axs[1].set_xlim((left-0.1,right+0.1))
axs[1].set_ylim((0, up))
```

```
axs[1].set_aspect("equal")
axs[1].set_title("Linear Interpolation",fontdict = {'fontsize' : 20})

llinemu, = axs[1].plot([], [], "g",lw=4)
llinenu, = axs[1].plot([], [], "r",lw=4)
lliner,= axs[1].plot([], [], "b",lw=4)
lline, = axs[1].fill([], [], "lightblue")#,lw=4,edgecolor="b")




def init():
    linemu.set_data([], [])
    return linemu,
def animate(i):
    t=min(0.9999,i/99)
    if i%10==0:
        print(i)
    x = np.linspace(left-0.1,right+.1, 2000)

    y = density(t,x,x)
    yl = (1-t)*nu(x)+t*mu(x)

    line.set_xy(np.array([x,y]).T)
    liner.set_data(x,y)
    linenu.set_data(x,nu(x))
    linemu.set_data(x,mu(x))

    lline.set_xy(np.array([x,yl]).T)
    lliner.set_data(x,yl)
    llinenu.set_data(x,nu(x))
    llinemu.set_data(x,mu(x))
    return line,linemu,linenu, liner,lline,llinemu,llinenu, lliner
```

```
anim = FuncAnimation(fig, animate, init_func=init,
                               frames=100, interval=20, blit=True)

anim.save("comparison_"+Name)
```

Appendix B. Python program to compute the counterexample in Section 4

```
import numpy as np
from scipy.stats import mvn

S=np.array([[ 1. ,  0.85,-0.78],
            [0.85,  1.  ,-0.36],
            [-0.78,-0.36, 1.  ]])
A=np.array([1.4,0.8,2.3])
B=np.array([1.4,2.1,1.0])

def f(v1,v2,S,gg=7):
    MIN=np.min((v1,v2),axis=0)
    MAX=np.max((v1,v2),axis=0)

    pA,ipA= mvn.mvnun(-v1,v1,(0,0,0),S,maxpts=10**gg)
    pB,ipB= mvn.mvnun(-v2,v2,(0,0,0),S,maxpts=10**gg)
    pC ,ipC = mvn.mvnun(-MAX,MAX,(0,0,0),S,maxpts=10**gg)
    pAB,ipAB= mvn.mvnun(-MIN,MIN,(0,0,0),S,maxpts=10**gg)
    return pC*pAB,pA*pB
print(f(A,B,S))
```

Appendix C. Python program for Figure 2 and animations.

```
import numpy as np
import ot
from PIL import Image
import matplotlib.pyplot as plt

#Loading the images
```

```
myImage1 = Image.open("/home/alexander/Documents/villani.png");
myImage2 = Image.open("/home/alexander/Documents/villaniflipped.png");
m1=np.array(myImage1)
m2=np.array(myImage2)

#Location of points at the start
Pointsm1=np.array([(i,j) for i in range(m1.shape[0]) for j in range(m1.shape[1])
if m1[i][j]==1])
#Location of points at the end
Pointsm2=np.array([(i,j) for i in range(m2.shape[0]) for j in range(m2.shape[1])
if m2[i][j]==1])

#Defining the quadratic cost matrix
M=np.add(Pointsm2,-np.array([Pointsm1]).transpose(1,0,2))
M=np.sum(np.square(M),axis=2)

#Calculating the Optimal Transport map
MMM=ot.emd([],[],M,numItermax=10**6)

def f(x): #location the pixel number x is moved to
    return MMM[x].argmax()
def g2(x,t): #position of the pixel number x at the time t
    return Pointsm1[x]*(1-t)+Pointsm2[f(x)]*t

#Creating the images
path="/home/alexander/pythonimages/faces/villani"
cmap = plt.cm.gray
norm = plt.Normalize(vmin=-1, vmax=0)
rr=100 #number of pictures
ff=10 #increases resolution to better show pixel movement

for i in range(rr):
    KK=np.zeros(np.array(m1.shape)*ff,dtype=int)
    t=i/(rr-1)
```

```
    for x in range(len(Pointsm1)):
        gg=np.array(np.round(g2(x,t)*ff),dtype=int)
        KK[gg[0]:gg[0]+ff,gg[1]:gg[1]+ff]=1

    l=cmap(norm(-KK))
    plt.imsave(path+str(i)+'.png', l)

#Creating an animation:
import imageio
imagenumbers=18*[0]+list(range(100))+18*[99]+list(range(99,-1,-1))
images = [imageio.imread(path+str(i)+'.png') for i in imagenumbers]
imageio.mimsave(path+'.mp4', images,fps=20)
```