JYU DISSERTATIONS 57

Anna Kausamo

On the Structure of Multi-marginal Optimal Mass Transportation in Metric Spaces

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF MATHEMATICS AND SCIENCE

JYU DISSERTATIONS 57

Anna Kausamo

On the Structure of Multi-marginal Optimal Mass Transportation in Metric Spaces

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Historica-rakennuksen salissa H320
helmikuun 9. päivänä 2019 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in building Historica, auditorium H320, on February 9, 2019 at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2019

Editors
Tapio Rajala
Department of Mathematics and Statistics, University of Jyväskylä
Ville Korkiakangas
Open Science Centre, University of Jyväskylä

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-7678-1

ISBN 978-951-39-7678-1 (PDF)
URN:ISBN:978-951-39-7678-1
ISSN 2489-9003

Acknowledgements

First of all, I want to thank my supervisor Tapio Rajala for all his help and patience. I also wish to express my gratitude to Augusto Gerolin for being my co-author and mentor, and for having energetically encouraged me all the way – and also for having pushed me when pushing was needed.

No words can express what I would like to say to my dear 'MaD people', that is, to a certain subset of the class of all PhD students, postdocs, and visitors of the Department of Mathematics and Statistics in the University of Jyväskylä. You made it possible for me to... find the way, not to get lost.

For financial support I am indebted to the Academy of Finland and the Vilho, Yrjö and Kalle Väisälä Foundation.

And finally, I wish to express my appreciation to Codina Cotar and Simone di Marino for having pre-examined my work.

Jyväskylä, 23.1.2019
Anna Kausamo

List of included articles

This dissertation consists of an introductory part and the following four publications:

- **[A]** A. Gerolin, A. Kausamo, and T. Rajala, *Non-existence of optimal transport maps for the multi-marginal repulsive harmonic cost*, preprint.
- **[B]** A. Gerolin, A. Kausamo, and T. Rajala, *Duality theory for multi-marginal optimal transport with repulsive costs in metric spaces*, ESAIM Control Optim. Calc. Var., to appear.
- **[C]** A. Gerolin, A. Kausamo, and T. Rajala, *Multi-marginal Entropy-Transport with repulsive cost*, preprint.
- **[D]** A. Kausamo, *Multi-marginal L^∞-transport*, preprint.

The author of this dissertation has actively taken part in research of the joint papers [A], [B] and [C].

INTRODUCTION

1. Multi-Marginal Optimal Mass Transportation Problem

We denote by $\mathscr{P}(X)$ the set of all Borel probability measures on a space X. Let $N \geq 2$ be fixed, and suppose we are given Polish spaces $X_1, \ldots, X_N$, measures $\mu_1, \ldots, \mu_N$, $\mu_i \in \mathscr{P}(X_i)$ for all $i \in \{1, \ldots, N\}$, and a function $c\colon \prod_{i=1}^N X_i \to \overline{\mathbb{R}}$ which we often call *a cost function*. Let us abbreviate $Y := \prod_{i=1}^N X_i$. In the Multi-Marginal Optimal Mass Transportation (MOT) problem one seeks for minimizing the quantity

$$C(\gamma) = \int_Y c(x_1, \ldots, x_N) \mathrm{d}\gamma \tag{1.1}$$

over all *couplings* $\gamma \in \mathscr{P}(Y)$ of the marginal measures $\mu_1, \ldots, \mu_N$, that is, over the set

$$\Pi_N(\mu_1, \ldots, \mu_N) := \{\gamma \in \mathscr{P}(Y) \mid (\mathrm{pr}_i)_\sharp \gamma = \mu_i \text{ for all } i \in \{1, \ldots, N\}\},$$

where pr_i is the projection on the i-th coordinate:

$$\mathrm{pr}_i(x_1, \ldots, x_N) = x_i \text{ for all } (x_1, \ldots, x_N) \in Y\,.$$

Minimizers for this problem exist under quite mild assumptions on the cost function c, like lower semicontinuity. MOT setup can be applied to very different frameworks, and thus the form of the cost function can change dramatically on going from one application to another. Often, if the spaces X_i are all the same, the cost is a function of pairwise distances between the coordinates. If one thinks that marginal measures $\mu_1, \ldots, \mu_N$ represent different mass or particle densities and interprets the cost function as an interaction between those masses, one finds it natural to say that c is *attractive* if it increases with increasing distances, and *repulsive* if it decreases as a function of distances. If, in the latter case, the cost blows up to $+\infty$ when the distances go to zero, we say that the cost is *singular*. An attentive reader may wonder why the word 'cost' keeps popping up. The reason lies in the two-marginal ($N = 2$) case. If $N = 2$, we may interpret the value $c(x, y)$ as the cost of moving one unit of mass from point x to point y. If, in the following, we don't specify that the context is multi-marginal, the word 'transport' refers to the two-marginal case. However, the word is in standard use also in the multi-marginal framework.

Finding the minimal γ for the problem (1.1) is often called solving the *Monge-Kantorovich* (MK) problem, in honor of the French mathematician Gaspard Monge (1746-1818) and the Russian mathematician Leonid Vitaliyevich Kantorovich (1912–1986), both of whom can be considered founders of the field of optimal mass transportation. Monge formulated in [26] the problem of deterministic transport, or transport given by a map. This problem now carries his name: *The Monge problem.* Kantorovich, on his part, studied in [19] the duality between minimizing the cost and maximizing the benefits of the transport, resulting in what we call today *the Kantorovich Duality.* His perspective was the formulation of the transportation problem as a mimimization in the set of all couplings between marginal measures, rather than in the set of transportation maps. He outlined the connection to Monge's work only after having obtained his main results on duality, see [34]. The multi-marginal versions of these fundamental works are in the heart of this thesis. We will state them in the following two sections and build the bridges between the basic formulations and the articles of this thesis. First however, we will describe the particular case where all marginal measures are the same since this is, for repulsive costs, the context of the greatest applicational interest and will be the setup in most of the articles of the thesis.

Let us denote by $\mathcal{S}_N$ the set of permutations of the set $\{1,\dots,N\}$. Assuming that $X_i = X$ for some Polish space X and $\mu_i = \mu \in \mathscr{P}(X)$ for all $i \in \{1,\dots,N\}$, it is convenient to abbreviate

$$\Pi_N(\mu) := \Pi_N(\mu,\dots,\mu)\,.$$

In this case, the MOT problem has symmetry that comes from the marginal constraints. There can also be symmetry that comes from the form of the cost function. We say that a cost function $c\colon X^N \to \overline{\mathbb{R}}$ is *symmetric* if it satisfies the condition

$$c(x_1,\dots,x_N) = c(x_{\sigma(1)},\dots,x_{\sigma(N)}) \text{ for all } (x_1,\dots,x_N) \in X^N \text{ and } \sigma \in \mathcal{S}_N\,.$$

If these two forms of symmetry are combined, we may restrict ourselves to seeking the minimum for the (MK) problem in the set

$$\Pi_N^{sym}(\mu) := \Pi_N(\mu) \cap \{\gamma \in \mathscr{P}(X^N) \mid (\tau_\sigma)_\sharp\gamma = \gamma \text{ for all } \sigma \in \mathcal{S}_N\}\,,$$

where we have denoted

$$\tau_\sigma : X^N \to X^N,\ (x_1,\dots,x_N) \mapsto (x_{\sigma(1)},\dots,x_{\sigma(N)})\,.$$

To every coupling $\gamma \in \Pi_N(\mu)$ we can associate a symmetrized coupling

$$\gamma^S := \frac{1}{N!}\sum_{\sigma\in\mathcal{S}_N}(\tau_\sigma)_\sharp\gamma \in \Pi_N^{sym}(\mu)\,.$$

If the cost function c is symmetric, then $C(\gamma^S) = C(\gamma)$, since the cost functional C is linear. Therefore, for symmetric cost functions the equality

$$\min_{\gamma\in\Pi_N(\mu)} C(\gamma) = \min_{\gamma\in\Pi_N^{sym}(\mu)} C(\gamma)$$

holds.

2. Monge Problem

Let us assume in this section, unless otherwise stated, that all the spaces X_i are the same: $X_i = X$ for some Polish space X.

In the Monge problem, one seeks for a solution γ_{opt} of the type

$$\gamma_{opt} = (\mathrm{id}, T_1,\dots,T_{N-1})_\sharp\mu_1\,, \tag{2.1}$$

where $T_i\colon X \to X$ are Borel functions such that $(T_i)_\sharp\mu_1 = \mu_{i+1}$ for all $i \in \{1,\dots,N-1\}$. This is equivalent to asking whether the equality

$$\begin{aligned}\min_{\gamma\in\Pi_N(\mu_1,\ldots,\mu_N)}&\int_{\mathbb{R}^{Nd}} c(x_1,\ldots,x_N)\mathrm{d}\gamma(x_1,\ldots,x_N)\\ &= \min\left\{\int_{\mathbb{R}^d} c(x,T_1(x),\ldots,T_{N-1}(x))\mathrm{d}\mu_1(x)\ \middle|\ (T_i)_\sharp\mu_1=\mu_{i+1}\right\}\end{aligned}$$

holds. If the answer is affirmative, we call any minimizing coupling of the type (2.1) a Monge minimizer.

Originally, Monge studied this problem for $N=2$ in a Euclidean space for the Euclidean distance cost $c(x_1,x_2):=|x_1-x_2|$. The rigorous treatment of this cost is, however, difficult due to its lack of strict convexity. Much more studied is the better-behaved squared-distance cost $c(x_1,x_2)=|x_1-x_2|^2$. For $N=2$ this cost function is often called 'the Brenier cost' in honor of Yann Brenier who, in 1991, proved [4] that its solutions are of the Monge type:

Theorem 2.1 (Brenier's Theorem)**.** *Let μ_1 and μ_2 be Borel probability measures with finite second moments. Assume that μ_1 is absolutely continuous with respect to the Lebesgue measure. Then the problem*

$$\min_{\gamma\in\Pi(\mu_1,\mu_2)}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x_1-x_2|^2\mathrm{d}\gamma(x_1,x_2)$$

admits a unique minimizer $\gamma_{opt}=(\mathrm{id},T)_\sharp\mu_1$, where $T(x)=\nabla\phi(x)$, and $\phi\colon\mathbb{R}^d\to\mathbb{R}$ is a convex function.

One of the greatest modern applications of OT has been the synthetic definitions of Ricci curvature in metric measure spaces, introduced by Sturm and Lott-Villani in [32, 33, 24]. These definitions are given in terms of minimizers of the Brenier cost. The assumption that the first marginal measure is absolutely continuous with respect to the Lebesgue measure is essential for the existence part of Brenier's theorem. For instance, if $\mu_1=\delta_0$ and $\mu_2=[-1,1]$, there are no Monge minimizers since the mass at the origin has to split into multiple points. For the uniqueness of the transportation map, the strict convexity of the cost plays a key role: Consider the cost $c(x,y)=|x-y|$, which is convex but not strictly convex. For the marginal measures

$$\mu=\mathcal{L}|_{[0,1]}\text{ and }\nu=\mathcal{L}|_{[\frac12,\frac32]}$$

both plans $\gamma_1=(\mathrm{id},T)_\sharp\mu$ and $\gamma_2=(\mathrm{id},S)_\sharp\mu$ are optimal, where $S,T\colon[0,1]\to[\frac12,\frac32]$ are defined by

$$T(x)=x+\frac12\text{ for all }x\in[0,1]$$

and

$$S(x)=\begin{cases}x+1 & \text{if } x\in[0,\frac12]\\ x & \text{if } x\in(\frac12,1]\end{cases}.$$

However, for the cost $c(x,y)=|x-y|^2$, only the plan γ_1 is optimal.

The proof of the existence and uniqueness of a transport map for the Brenier cost boils down to the differentiability of the cost function and the invertitibility of its gradient. In this two-marginal case, the form $c(x,y)=|x-y|^2$ plays no special role and one can indeed show that any cost of the form $c(x,y)=h(|x-y|)$, where $h\colon [0,\infty)\to[0,\infty)$ is increasing and strictly convex, admits a unique minimizer that is of the Monge type. Actually, whenever the cost function satisfies the so-called *twist condition*, that is, whenever the mapping $y\mapsto D_x c(x,y)$ is injective on its domain, there exists a unique solution of the Monge type. What happens if the twist condition fails? This is the case if, for instance, the cost function is of the form $c(x,y)=|x-y|^\alpha$ for $0<\alpha<1$. In 2015, Pegon, Piazzoli, and Santambrogio completed in [27] the characterization, initiated in 1996 by Gangbo and McCann in [13], of optimal transportation plans on $\mathbb{R}^d$ for cost functions of the form $c(x,y)=l(|x-y|)$ where $l\colon[0,\infty)\to[0,\infty)$ is an increasing, strictly concave function. Intuitively, the concavity of l makes small displacements relatively more expensive than bigger ones. In this case, under the assumption that the first marginal measure does not give mass to $(d-1)$-rectifiable sets (which are also called *small sets*), the optimal plan is unique but it is induced by a single graph only if the marginal measures do not overlap. For general marginal measures $\mu,\nu\in\mathscr{P}(\mathbb{R}^d)$, with μ not giving mass to small sets, the optimal transport plan splits into two parts: the mass that is common to both μ and ν will stay unmoved, and the part of μ that is singular with respect to ν will be transported, by a map, to the part of ν that is singular with respect to μ. Formally, the unique optimal $\gamma\in\Pi_2(\mu,\nu)$ is of the form

$$\gamma=(\mathrm{id},\mathrm{id})_\sharp(\mu\wedge\nu)+(\mathrm{id},T)_\sharp(\mu-\nu)_+\,,$$

where $T\colon\mathbb{R}^d\to\mathbb{R}^d$ is a function.

For the two-marginal case, it is possible to use Brenier's theorem to obtain insight into the so-called *Repulsive harmonic cost* which is defined as $c_h(x,y)=-|x-y|^2$. Here h stands for 'harmonic' though, in the literature, also the notation c_w is frequently seen. This latter name stems from the interpretation of c_w as a weak (repulsive) interaction – weak in the sense that putting the particles in the same point in space costs one 0, not for example $+\infty$ which would be the case if the cost was singular. If one assumes that μ_1 and μ_2 are Borel probability measures on $\mathbb{R}^d$, $d\geq 1$, with finite second moments, and that μ_1 is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, then the problem

$$\min_{\gamma\in\Pi(\mu_1,\mu_2)}\int_{\mathbb{R}^d\times\mathbb{R}^d}-|x_1-x_2|^2\mathrm{d}\gamma(x_1,x_2)$$

admits a unique solution that is of the form $\gamma_{opt}=(\mathrm{id},T)_\sharp\mu_1$ where for the transport map $T\colon\mathbb{R}^d\to\mathbb{R}^d$ we have $T=\nabla\phi$ for a concave function ϕ. This follows from Brenier's theorem since we have

$$\inf_{T_\sharp\mu_1=\mu_2}\int_{\mathbb{R}^d}-|x-T(x)|^2\mathrm{d}\mu_1+\int_{\mathbb{R}^d}|x|^2\mathrm{d}\mu_1+\int_{\mathbb{R}^d}|x|^2\mathrm{d}\mu_2=\inf_{G_\sharp\mu_1=(-id)_\sharp\mu_2}\int_{\mathbb{R}^d}|x-G(x)|^2\mathrm{d}\mu_1\,,$$

where $G=-T$.

There also exists a multi-marginal version of the Brenier cost:

$$c_a(x_1, \ldots, x_N) = \sum_{1 \leq i < j \leq N} d(x_i, x_j)^2 \,.$$

This cost function is often called *the attractive harmonic* cost. The name originates from the multi-marginal interpretation of c_a as the coupling of particles under an attractive, parabolic *i.e.* harmonic interaction. Also the name 'Gangbo-Swiech cost' is frequently used, in honor of Wilfrid Gangbo and Andrzej Swiech. They proved in 1998 [15] that, if $\mu_1, \ldots, \mu_N$ have finite second moments and vanish on $(d-1)$-rectifiable sets, then the Monge-Kantorovich problem

$$\min_{\gamma \in \Pi_N(\mu_1, \ldots, \mu_N)} \int_{\mathbb{R}^{Nd}} c_a(x_1, \ldots, x_N) \mathrm{d}\gamma(x_1, \ldots, x_N)$$

admits a unique minimizer that is of the Monge type.

In general, the Monge problem in the case $N > 2$ is highly nontrivial. Much depends on the number of marginals, dimension of the space, or the form of the cost. Arguments that work for two marginals fail in the most cases if one tries to generalize them to $N > 2$. For instance, one cannot go from the attractive harmonic cost to the repulsive harmonic cost if $N > 2$ since the geometric reasoning used for two marginals does not generalize. Also, one cannot in general build multi-marginal Monge solutions from two-marginal ones. For example, the optimal transportation map for the two-marginals repulsive harmonic cost is the anti-monotone rearrangement, as outlined before. If one increases the number of marginals there is no hope for such a map since all pairwise displacements cannot be anti-monotone simultaneously.

What holds for one specific type of cost may fail miserably if one modifies the cost, even though similar modifications do not ruin the arguments in the $N = 2$ case. An example of this is the comparison of Brenier's Theorem and Gangbo-Swiech Theorem: As pointed out before, a Brenier-type result holds for any cost function of the type $c(x, y) = h(|x - y|)$ where h is a strictly convex function. However, the proof of the Gangbo-Swiech theorem uses explicitly the form of c_h, resulting in arguments that do not work even for costs like

$$c(x_1, \ldots, x_N) = \sum_{1 \leq i < j \leq N} |x_i - x_j|^p, \qquad p \geq 3 \,.$$

Indeed, the Monge problem is open for the above form of cost functions.

This thesis concerns, for the most part, the repulsive costs. Let us now turn our attention more closely on those. Hence, unless otherwise stated, the cost function c is repulsive for the rest of this section. On studying the Monge problem, one factor we haven't emphasized so far is the structure of the underlying space. With toy models such as finite spaces, one can easily construct examples and nonexamples for the multi-marginal Monge problem. For spaces with more structure, results tend to depend strongly on that structure, like on the fact that $\mathbb{R}^d$ is an inner product space. One special case that is worth mentioning is when $d = 1$, since the space $\mathbb{R}$, the set of real numbers, has a characteristic that makes affirmative answer to the Monge problem possible: it has an ordering that is compatible with the topology given by the Euclidean distance. This was done in 2015 by Colombo, De Pascale, and Di Marino in [8]. They considered the cost $c\colon \mathbb{R}^N \to [0,\infty]$,

$$c(x_1,\dots,x_N) = \sum_{1\le i<j\le N} f(|x_i - x_j|) \text{ for all } (x_1,\dots,x_N) \in \mathbb{R}^N\,,$$

where $f\colon \mathbb{R}\to[0,\infty]$ is an even function that is strictly convex and non-increasing on the interval $[0,\infty[$. They proved the following:

Theorem 2.2 (Colombo, De Pascale, Di Marino)**.** *Let ρ be a non-atomic probability measure on $\mathbb{R}$ such that*

$$K := \min_{\gamma\in\Pi_N(\rho)} C(\gamma) < \infty\,.$$

Let $-\infty = d_0 < d_1 < \cdots < d_N = +\infty$ be such that

$$\rho([d_i, d_{i+1}]) = \frac{1}{N} \textit{ for all } i \in \{0,\dots,N-1\}\,.$$

Let $T\colon \mathbb{R}\to\mathbb{R}$ be the unique (up to ρ-null sets) function which is increasing on each interval $[d_i, d_{i+1}]$, $i = 0,\dots,N-1$, and which satisfies

$$T_\sharp 1_{[d_i,d_{i+1}]}\rho = 1_{[d_{i+1},d_{i+2}]}\rho \textit{ for all } i\in\{0,\dots,N-2\}, \textit{ and } T_\sharp 1_{[d_{N-1},d_N]}\rho = 1_{[d_0,d_1]}\rho\,.$$

Then T is an admissible map for the cyclical Monge problem, and

$$K := \int_{\mathbb{R}} c(x, T(x), T^{(2)}(x),\dots,T^{(N-1)}(x))\mathrm{d}\rho(x)$$

where we denote by $T^{(i)}$ the i-fold composition of the map T with itself. Moreover, the only symmetric optimal transport plan is the symmetrization of the plan induced by the map T.

The ordering of $\mathbb{R}$ also gives added insight into the attractive L^∞-cost we are considering in Section 5.

Are there conditions, for general enough spaces, under which we can guarantee an affirmative answer to the multi-marginal Monge problem? Kim and Pass studied [21] this question for continuous and semi-concave costs on Riemannian manifolds. They formulated a condition that is analogous to the two-marginal twist condition: if the function $(x_2,\dots,x_N)\mapsto D_{x_1}c(x_1,x_2,\dots,x_N)$ is injective on 'c-splitting sets' that we define below, then the minimizer for the Monge-Kantorovich problem is unique and deterministic. In the definition and the theorem below, $M_1,\dots,M_N$ are Riemannian manifolds and $\mu_1,\dots,\mu_N$ are Borel probability measures on $M_1,\dots,M_N$, respectively.

Definition 2.3. *We say that a set $\Gamma\subset M_1\times\cdots\times M_N$ is c-splitting if there exist Borel functions $u_i\colon M_i\to\mathbb{R}$ such that*

$$\sum_{i=1}^N u_i(x_i) \le c(x_1,\dots,x_N) \textit{ for all } (x_1,\dots,x_N)\in M_1\times\cdots\times M_N\,,$$

with the equality whenever $(x_1,\dots,x_N)\in\Gamma$.

Theorem 2.4 (Kim and Pass 2014). *Assume that c is continuous and semi-concave. Assume also that μ_1 is absolutely continuous with respect to the local coordinates. If for each $x_1 \in M_1$ and a c-splitting set $\Gamma \subset \{x_1\} \times M_2 \times \cdots \times M_N$ the map*

$$(x_2, \dots, x_M) \mapsto D_{x_1} c(x_1, x_2, \dots, x_N)$$

is injective on the subset of Γ where c is differentiable with respect to x_1. Then every solution γ to the (MK) problem induces a Monge solution and is unique.

The proof extends the two-marginal proof of the now-classical twist condition to the multi-marginal case. It pinpoints exactly the properties of the cost function that are essential for the applicability of this kind of a reasoning, based on the Kantorovich duality that we will discuss in more detail in Section 3. Also the proof of the Gangbo-Swiech theorem, that we stated earlier, has similar characteristics.

Let us take a step back. We are given $N \geq 2$ Borel probability measures $\mu_1, \dots, \mu_N$ on, say, $\mathbb{R}^d$. Let us assume that μ_1 is absolutely continuous with respect to the d-dimensional Lebesgue measure. We are given a continuous cost function $c\colon \mathbb{R}^{Nd} \to \mathbb{R} \cup \{+\infty\}$. As outlined before, the question: 'Is there a unique minimizer to (MK) that is of the Monge-type?' is a tricky one. We can try to understand the situation better by asking some more questions. Can we approximate by Monge-type minimizers? If there is a Monge-type minimizer, is it unique? Is there a unique symmetric minimizer? Are there cases in which we never have Monge-type minimizers? Can the minimizer in this kind of case be unique?

The answer to the first question is affirmative, at least if the cost function does not get the value $-\infty$, is continuous, and does not have atoms, that is, does not give positive mass to single points. In this case we have

$$\min_{\gamma \in \Pi_N(\mu_1, \dots, \mu_N)} \int_{\mathbb{R}^{Nd}} c \mathrm{d}\gamma = \inf \left\{ \int_{\mathbb{R}^d} c(x, T_1(x), \dots, T_{N-1}(x)) \mathrm{d}\mu(x) \;\middle|\; (T_i)_\sharp \mu_1 = \mu_i \right\},$$

where $T_i\colon \mathbb{R}^d \to \mathbb{R}^d$ are Borel functions and $i \in \{1, \dots, N-1\}$. This result follows from a two-marginal theorem of Pratelli, published in 2007 in [28]:

Theorem 2.5 (Pratelli). *Let X and Y be Polish spaces. Fix $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ such that μ does not have atoms. Let $c\colon X \times Y \to (-\infty, \infty]$ be a continuous cost function. Then the infimum of the Monge problem equals the minimum of the Monge-Kantorovich problem.*

The multi-marginal version follows from the theorem of Pratelli by choosing, for a minimizer γ of (1.1), $X = X_1$, $Y = X_2 \times \cdots \times X_N$, $\mu = \mu_1$, and $\nu = (\mathrm{pr}_2, \dots, \mathrm{pr}_N)_\sharp \gamma$.

Our next question was: If there is a Monge-type minimizer, is it unique? If the 'twist on splitting sets' condition is satisfied, then we have uniqueness. However, in general, even the Monge minimizers for multi-marginal OT problems can be nonunique. The demonstration of this leads us to the introduction of one of the important cost functions of this thesis, the multi-marginal repulsive harmonic cost. The two-marginal version was defined already, and for $N \geq 3$ we use the same notation $c_h \colon X^N \to [0, \infty)$,

$$c_h(x_1, \dots, x_N) = - \sum_{1 \leq 1 < j \leq N} d^2(x_i, x_j) \text{ for all } (x_1, \dots, x_N) \in X^N .$$

If $X = \mathbb{R}^d$ with the Euclidean metric, a direct computation shows that, if the marginal measures have finite second moments (*i.e.* $\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\mu_i(x) < \infty$ for all $i \in \{1, \dots, N\}$) we have

$$\operatorname*{argmin}_{\gamma} \int_{\mathbb{R}^{Nd}} c_h(x_1, \dots, x_N) \mathrm{d}\gamma(x_1, \dots, x_N) = \operatorname*{argmin}_{\gamma} \int_{\mathbb{R}^{Nd}} |x_1 + x_2 + \dots + x_N|^2 \mathrm{d}\gamma(x_1, \dots, x_N) ;$$

this observation turns out to be particularly useful. The repulsive harmonic cost was studied in 2017 by Di Marino, Gerolin, and Nenna [12]. It is a remarkably fruitful cost function when one seeks for constructing examples and counterexamples, since its minimizers have a very particular structure, as shown by Di Marino, Gerolin, and Nenna in the following theorem:

Theorem 2.6 (Di Marino, Gerolin, Nenna)**.** *Consider the cost $c_h(x_1, \dots, x_N) = h(x_1 + \dots + x_N)$ where $h \colon \mathbb{R}^d \to \mathbb{R}^+$ is a strictly convex function. Let us assume that there exists $\gamma \in \Pi_N(\mu_1, \dots, \mu_N)$ concentrated on some hyperplane $\{x_1 + x_2 + \cdots + x_N = k\}$, $k \in \mathbb{R}^d$. Then γ is optimal for the cost c_h. In this case, a plan $\tilde{\gamma} \in \Pi_N(x_1, \dots, x_N)$ is optimal for the cost c_h if and only if*

$$\operatorname{spt} \tilde{\gamma} \subset \left\{ (x_1, \dots, x_N) \in \mathbb{R}^{Nd} \;\middle|\; x_1 + x_2 + \dots + x_N = k \right\} .$$

In particular, the constant k can be computed explicitly:

$$k = \sum_{i=1}^{N} \int_X x \mathrm{d}\mu_i(x) .$$

The proof of the above theorem is based on the Jensen's inequality; the strict convexity of the function h is needed in showing that if a coupling $\tilde{\gamma}$ is optimal, then it must satisfy the condition

$$x_1 + x_2 + \dots + x_N = k \text{ for } \tilde{\gamma}\text{-a.e. } (x_1, \dots, x_N) .$$

Theorem 2.6 can now be used to show that a multi-marginal Monge solution is not necessarily unique. A simple example of would be the case of $2N$ particles on the ring $\mathbb{S}^1$. If $\mu_1, \dots, \mu_N$ are uniform probability measures on $\mathbb{S}^1$, then the rotation map $\mathbb{R}_\theta \colon \mathbb{R}^2 \to \mathbb{R}^2$ with angle $\theta = \pi/N$ is optimal, but also the rotation maps $\mathbb{R}_{k\theta}$, $k \in \{2, \dots, N\}$ are. However, we posed our questions for the cases in which the first marginal is absolutely continous with respect to the Lebesgue measure on $\mathbb{R}^d$. Interestingly, even with this assumption there is in general no uniqueness, and in the proof of this, Theore 2.6 plays a crucial role. Di Marino, Gerolin, and Nenna constructed in [12] (Theorem 4.6 and Corollary 4.7) an example of a Monge-minimizer for the case $\mu_i = \mu = \mathcal{L}^d|_{[0,1]}$ for all $i \in \{1, \dots, N\}$,

$N \geq 3$, the cost function being repulsive harmonic, and showed that there exists a Monge minimizer that is not only nonunique but also fractal, that is, not differentiable almost everywhere. This example is important since it shows that, even for MOT problems that have a relatively simple form like the repulsive harmonic, there can exist minimizers that are far from regular.

The question of the uniqueness of the minimizers for multi-marginal OT problems is interesting also in the nondeterministic Monge-Kantorovich formulation. In many cases, the symmetries of the problem guarantee the *non-uniqueness* of the solutions, except under rather special conditions. This was proven in 2017 by Moameni and Pass [25]. For the sake of simplicity, the statement is formulated for $N = 3$. However, it also holds for $N \geq 4$.

Theorem 2.7 (Moameni and Pass)**.** *Suppose that the spaces $X_i = X$ and the marginal measures $\mu_i = \mu$ are the same for each i, and that the cost function $c(x, y, z)$ is symmetric with respect to any permutation of the arguments. Assume that there exists mutually disjoint sets $S_1, S_2, S_2 \subset X$ and an optimal coupling γ that charges the set $S := S_1 \times S_2 \times S_3$, that is, $\gamma(S) > 0$. Then the solution to (1.1) is nonunique.*

In the light of this theorem, if the marginal measures are the same as is the case in most applications, the question should not be: 'Is the optimal coupling unique?' but rather: 'Is the optimal symmetric coupling unique?' If the answer is affirmative, one may look back at the Monge problem and ask: 'Is the unique optimal symmetric coupling given by a map from the first marginal?' This brings us to the first paper [A] of the thesis. We study the repulsive harmonic cost on $\mathbb{R}^d$ in the case $N = 3$. We construct a measure μ, absolutely continous with respect to $\mathcal{L}^d$, so that the multi-marginal OT problem with all marginals equal to μ has no minimizer that is induced by a map from the first coordinate. In addition, we show that the problem has a unique symmetric minimizer. The proofs are heavily based on the fact that, in the case of the repulsive harmonic cost, the structure of the minimizers is quite tractable as Theorem 2.6 shows. Another important ingredient is the existence of a unique minimizer for c_h in the two-marginal case. We did not succeed in generalizing the arguments for $N \geq 4$.

3. Kantorovich Duality

The classical Monge-Kantorovich problem

$$(P) := \min_{\gamma \in \Pi_2(\mu,\nu)} \int_{X \times Y} c(x, y) \mathrm{d}\gamma(x, y)$$

can be considered a linear programming problem; above (P) stands for the *primal* problem which will be shortly connected to a *dual* problem. Due to the linearity of the integral, the functional

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

is linear with respect to the measure γ. The set $\Pi_2(\mu, \nu)$ is compact and convex, so actually we are minimizing a linear functional in a compact and convex set. Like all linear programming problems, also this one admits a dual formulation. In this case, the dual formulation reads [34]

$$(D) := \sup \left\{ \int_X \varphi(x)\mathrm{d}\mu + \int_Y \psi(y)\mathrm{d}\nu \;\middle|\; (\varphi,\psi) \in L^1_\mu(X) \times L^1_\nu(Y) \right.$$
$$\left. \text{such that } \varphi(x) + \psi(y) \le c(x,y) \right\}.$$

The celebrated Kantorovich duality theorem states that, if the cost function $c\colon X \times Y \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and if there exist real-valued upper semicontinuous functions $a \in L^1_\mu(X)$ and $b \in L^1_\nu(Y)$ such that

$$c(x,y) \ge a(x) + b(y) \quad \text{for all } (x,y) \in X \times Y\,,$$

then

$$(P) = (D).$$

We can also (see [34] for details) take the supremum over the space $\mathcal{C}_b(X) \times \mathcal{C}_b(Y)$, that is, over the product space of the space of continuous, bounded functions on X and the space of continuous, bounded functions on Y. We could even assume that the functions φ and ψ are Lipschitz. Or, we could consider the supremum only over functions $\varphi \in L^1_\mu(X)$ under the conjugacy relation

$$\psi(y) = \inf_x [c(x,y) - \varphi(x)] \quad \text{for all } y \in Y\,,$$

or over the functions $\psi \in L^1_\nu(Y)$ under the conjugacy relation

$$\varphi(x) = \inf_y [c(x,y) - \psi(y)] \quad \text{for all } x \in X\,.$$

We know that the primal problem (P) has a minimizer; this is why we wrote it as a minimum, not as an infimum. A natural question to ask is whether the dual problem (D) admits a maximizer. Kantorovich duality theorem guarantees the existence of a maximizer if the following three conditions are satisfied: the cost function is real-valued, $(P) < \infty$, and there exist functions $f \in L^1_\mu(X)$, $g \in L^1_\nu(Y)$ such that

$$c(x,y) \le f(x) + g(y) \text{ for all } (x,y) \in X \times Y\,.$$

If the spaces X and Y are the same Polish space X, equipped with a distance function d, and if $c(x,y) = d(x,y)$, then Kantorovich Duality actually gives the equivalence between notions of distance on the space of Borel probability measures on X: the Wasserstein distance and the Kantorovich-Rubinstein distance.

Let us now turn our attention on the multi-marginal duality theory. We denote again the primal problem by (P), that is,

$$(P) := \min_{\gamma \in \Pi_N(\mu_1,\dots,\mu_N)} \int_{X_1 \times \cdots \times X_N} c\mathrm{d}\gamma\,.$$

The dual problem now takes the form

$$(D) := \sup \left\{ \sum_{i=1}^{N} \int_{X_i} u_i \mathrm{d}\mu_i \mid (u_1, \ldots, u_N) \in \mathcal{F} \right\}$$

where

$$\mathcal{F} := \left\{ (u_1, \ldots, u_N) \in L^1_{\mu_1}(X_1) \times \cdots \times L^1_{\mu_N}(X_N) \;\middle|\; \sum_{i=1}^{N} u_i(x_i) \leq c(x_1, \ldots, x_N) \text{ for all } (x_1, \ldots, x_N) \in X_1 \times \cdots \times X_N \right\}.$$

There is, however, some subtlety in this notation that needs, at least, a comment. The same issue was present already in the two-marginal case yet we ignored it deliberately, for the sake of clarity. The elements of the space $L^1_{\mu_i}$ are, strictly speaking, equivalence classes, the class of a μ_i-integrable function u_i consisting of all μ_i-integrable functions that agree with u_i outside a set of μ_i-measure zero. Therefore, there is little sense in writing the pointwise constraint

$$u_1(x_1) + \cdots + u_N(x_N) \leq c(x_1, \ldots, x_N)$$

and asking it to be satisfied for *classes* $u_i \in L^1_{\mu_i}$. In [B] we ask the constraint to hold for $(\mu_1 \otimes \cdots \otimes \mu_N)$-almost every point $(x_1, \ldots, x_N)$ and show (see Lemma 2.1 in [Article [B]) that there always exists a representative that fulfills the constraint at every point. The formulation of Lemma 2.1 in [B] is a bit different since we consider there only the symmetric case where spaces X_i and marginals μ_i are all the same. However, the reasoning is valid also if the symmetries are removed.

From the form of the constraint of (D) one readily sees that the inequality $(D) \leq (P)$ always holds. Hence, the first question is: do we have the equality $(D) = (P)$. The second question should then be: does there exist a maximizer for (D). In his comprehensive and much-cited study [20], published in 1984, Kellerer considered the problem in topological spaces for very general cost functions. On Polish spaces that are of our interest, his main results are the following:

(a) **Theorem 2.6**
If $c\colon X_1 \times \cdots \times X_N \to \overline{R}$ is lower-semicontinuous, then the duality $(D) = (P)$ holds.

(b) **Theorem 2.21**
If $(D) > -\infty$ and if there exist μ_i-integrable functions h_i such that

$$c(x_1, \ldots, x_N) \leq h(x_1) + \cdots + h(x_N) \text{ for all } (x_1, \ldots, x_N) \in X_1 \times \cdots \times X_N\,, \tag{3.1}$$

then the dual problem (D) admits a maximizer.

The starting point in Kellerer's proof of the duality $(D) = (P)$ is the case when all spaces X_i are finite. For finite spaces, the duality is a well-known theorem in linear programming and, actually, a consequence of the Hahn-Banach theorem in finite-dimensional spaces.

From finite spaces and the corresponding finite cost functions one can go, by suitable embeddings, to the case where the cost is a product of finite elementary functions; a finite elementary function is a Borel measurable function whose image is a finite set. Finally, the lower-semicontinuous cost function h is approximated by those products of finite elementary functions, completing the proof. In the proof of the existence of a maximizer, the assumption of boundedness of the cost from above by a sum of intergrable functions is needed to guarantee the suitable compactness properties in the spaces $L^1_{\mu_i}(X_i)$. An attentive reader, who seeks for going back to the work of Kellerer to study his proofs in more detail, should be ready to a very different notation to what is used today. Also, Kellerer formulated his problem as a negative of our $(P) = (D)$ setup. Therefore, when we take infimums, Kellerer takes supremums and vice versa, and his upper semicontinous is lower-semicontinuous in our framework. One can run across the sign convention of Kellerer also when studying, for instance, problems in linear programming.

Now we move from general multi-marginal duality theory to the context of this thesis. The rest of this section considers the case where the cost function is singular, the spaces X_i are all equal to some Polish space X, and all the marginals are the same, equal to a measure $\mu \in \mathscr{P}(X)$. These setup is highly relevant in the application of MOT theory on Density Functional Theory where the now-famous Coulombic cost

$$c(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} \frac{1}{d(x_i, x_j)} \text{ for all } (x_1, \dots, x_N)$$

can be used to represent the so-called *strong interaction limit* [30]. Here the marginal measure represents the electron density and thus the assumption of all marginals being the same is natural: electrons are indistinguishable. The optimal couplings represent equilibrium electronic configurations under the electrostatic Coulombic interaction.

Even though we state most of the multi-marginal theorems and definitions for the case $X_i = X$, $\mu_i = \mu$ for all i, with mostly minor modifications everything works also if the spaces and the marginal measures are different. The duality theorem now reads:

$$(P) := \min_{\gamma \in \Pi_N(\mu)} \int_{X^N} c \mathrm{d}\gamma = \sup \sum_{i=1}^{N} \int_X u_i \mathrm{d}\mu \,,$$

where the supremum on the right-hand side is taken over the set $\mathcal{F}$ all N-tuples $(u_1, \dots, u_N)$ of μ-integrable functions on X such that

$$u_1(x_1) + \dots + u_N(x_N) \le c(x_1, \dots, x_N) \text{ for } \mu^{\otimes N}\text{-a. e. } (x_1, \dots, x_N) \in X^N \,.$$

The dual formulation can be simplified: For each N-tuple

$$(u_1, \dots, u_N) \in \mathcal{F}_1$$

we can associate a function $u = \frac{1}{N}\sum_{i=1}^{N} u_i$. For this function, we have

$$N\sum_{i=1}^{N}\int_X u(x)\mathrm{d}\mu = N\cdot\frac{1}{N}\sum_{i=1}^{N}\int_X u_i\mathrm{d}\mu = \sum_{i=1}^{N}\int_X u_i\mathrm{d}\mu$$

and for all $(x_1,\ldots,x_N)$

$$u(x_1)+\cdots+u_(x_N) = \sum_{i=1}^{N} u_i(x_i) \leq c(x_1,\ldots,x_N)\,.$$

Hence it is equivalent to take define the dual problem as

$$(D) = \sup\left\{N\int_X u\mathrm{d}\mu \;\middle|\; u\in L^1_\mu(X) \text{ and } \sum_{i=1}^{N} u(x_i)\leq c(x_1,\ldots,x_N) \right.$$
$$\left. \text{for } \mu^{\otimes N}\text{-a.e. } (x_1,\ldots,x_N)\right\}.$$

For singular costs, the equality $(P)=(D)$ follows from the Theorem (a) of Kellerer stated earlier. However, we cannot use Theorem (b) to prove the existence of a maximizer. This theorem is not applicable to singular costs. The problem is in the key assumption: there cannot exist μ-integrable functions h_i, $i=1,\ldots,N$ on the space X such that

$$c(x_1,x_2,\ldots,x_N)\leq h(x_1)+h(x_2)+\cdots+h(x_N) \text{ for all } (x_1,\ldots,x_N)\in X^N\,,$$

since the right-hand side allows the choice $x_1=x_2=\cdots x_N$ in sets of positive μ-measure, whereas the left-hand-side blows up whenever $x_i=x_j$.

The problem of generalizing the duality theory to cover also singular costs was first studied by Luigi De Pascale for the Coulomb cost in [11]. He used Γ-convergence tools for establishing the existence of a maximizer for (D) by approximating the Coulomb cost with cost functions which are truncated close to the origin and which, therefore, satisfy Theorem (b) of Kellerer. The work of De Pascale was a bit later generalized by Buttazzo, Champion, and De Pascale in [5] to cover all cost functions of the form

$$c\colon\mathbb{R}^{Nd}\to[0,\infty],\ c(x_1,\ldots,x_N)=\sum_{1\leq i<j\leq N}\phi(|x_i-x_j|) \text{ for all } (x_1,\ldots,x_N)\,,$$

where $\phi\colon(0,\infty)\to[0,\infty)$ is a continuous, strictly decreasing function satisfying $\lim_{t\to 0+}\phi(t)=+\infty$. The authors also comment that the assumption of ϕ being strictly decreasing can be replaced by the assumptions that ϕ is bounded at $+\infty$. This class of cost functions covers, for instance, all cost functions of the type

$$c(x_1,\ldots,x_N)=\sum_{1\leq i<j\leq N}\frac{1}{|x_i-x_j|^s}\,,\ s\geq 1\,, \tag{3.2}$$

also called as Riesz costs. Those include, for the choice $s=1$, the Coulomb cost. In his proof for the Coulomb cost, De Pascale assumed that the marginal measure is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. The work of Buttazzo, Champion, and De Pascale not only widened the class of cost functions but also relaxed the assumption on the marginal measure. They assumed that the marginal measure μ has the *small concentration*:

$$\lim_{r\to 0} \sup_{x\in\mathbb{R}^d} \mu(B(x,r)) < \frac{1}{N(N-1)^2}\,. \tag{A}$$

This assumption means, roughly speaking, that the marginal measure should not have too heavy atoms. The upper bound $\frac{1}{N(N-1)^2}$ has been further slackened in a work in progress by Colombo, Di Marino, and Stra.

The key ingredient of the proof of Buttazzo, Champion, and De Pascale is the observation that, for small enough $\alpha > 0$, the support of an optimal coupling stays uniformly out of the diagonal

$$D_\alpha := \{(x_1,\dots,x_N) \in \mathbb{R}^{Nd} \mid x_i = x_j \text{ for some } i \neq j\}\,.$$

The authors also proof regularity properties of the maximizers of the dual problem and continuity results of the cost functional C.

What remained open after the previous works was the case when a singular cost function is not bounded from below. An example of such a cost if the repulsive logarithmic cost

$$c(x_1,\dots,x_N) = -\sum_{1\le i<j\le N} \frac{1}{\log(d(x_i,x_j))} \quad \text{for all } (x_1,\dots,x_N) \in X\,.$$

ths cost function occurs in the case '$N = 2$, $X = \mathbb{S}^d$' in the so-called *reflector problem* studied by Wang in [35], and Gangbo and Oliker in [14]. In the multi-marginal case, logarithmic cost function is the natural form of the electrostatic potential if the particles are confined to the plane $\mathbb{R}^2$.

The second paper of thesis [B] extends the multi-marginal duality theory to cover also the singular cost functions that are not bounded from below. We state and prove our theorems in general Polish spaces, since the arguments do not need the structure of $\mathbb{R}^d$ and since some applications (like the reflector problem) are not formulated on a Euclidean space.

To conclude this section, let us remark that one can also consider the multi-marginal optimal transportation problem as a function of the number N of marginal measures μ and seek for expressing the minimal cost as a power series expansion of the variable N. In 2017, the expansions up to the second order have been obtained for the Coulomb cost on $\mathbb{R}^3$ by Lewin, Lieb, and Seiringer in [23]. A more general work appeared also in 2017 by Cotar and Petrache [9], covering Riesz costs (3.2) for exponents $0 < s < d$ where $d \geq 1$ is the dimension of the underlying space $\mathbb{R}^d$.

4. Entropy-regularized multi-marginal optimal transport with repulsive costs

The third paper of this thesis [C] is a preliminary version of our work in progress on the topic of entropy-regularized multi-marginal optimal transport for singular cost functions. We work on general Polish measure space $(X, d, \mathfrak{m})$. Let us denote by $\mathscr{P}^{ac}(X)$ the set of Borel probability measures on X which are absolutely continuous with respect to the reference measure $\mathfrak{m}$. For $N \geq 2$ and for a fixed marginal measure $\rho\mathfrak{m} \in \mathscr{P}^{ac}(X)$ we consider the following minimization problem

$$\inf_{\gamma \in \Pi_N^{sym}(\rho)} C_\epsilon(\gamma) \text{ for some } \epsilon \geq 0\,.$$

Here C_ϵ is the entropy-regularized cost functional

$$C_\epsilon(\gamma) = C_0(\gamma) + \epsilon E(\gamma)\,,$$

where $C(\gamma) = \int_{X^N} c \mathrm{d}\gamma$ is a non-regularized MOT cost and E is the entropy functional

$$E(\gamma) = \begin{cases} \int_{X^N} \rho_\gamma \log \rho_\gamma \mathrm{d}\mathfrak{m}_N, & \text{if } \gamma \ll \mathfrak{m}_N \\ +\infty, & \text{otherwise} \end{cases}.$$

Above we have denoted by $\mathfrak{m}_N$ the N-fold product of the measure $\mathfrak{m}$, and by ρ_γ the Radon-Nikodym derivative of the measure γ with respect to the reference measure $\mathfrak{m}_N$. The notation $\gamma \ll \mathfrak{m}_N$ means that γ is absolutely continuous with respect to $\mathfrak{m}_N$.

The motivation for studying entropy-regularized OT problems comes from both theory and numerics. The theoretical applications include second order calculus on RCD spaces[16], the Schrödinger problem introduced by Schrödinger in [29] and revisited from the entropy-transport point-of-view by Léonard in [22], and bounding from below the Hohemberg-Kohn functional in Density Functional Theory [31]. In the numerical implementations of optimal transportation problems, the entropy functional serves as a regularizer that makes computational solution of the Monge-Kantorovich problem more feasible [10, 2]. Numerical experiments suggest that, when the regularization parameter ϵ goes to 0, the minimizers γ_ϵ for the entropy-regularized MOT problems converge to a minimizer of the unregularized problem whose entropy is minimal in the class of minimizers of C_0. This convergence was stated and proved rigorously for the case $N = 2$, $c(x, y) = |x-y|^2$ on $\mathbb{R}^d$ by Carlier, Duval, Peyré, and Schmitzer in [6]. The notion of convergence used is the Γ-convergence of functionals on topological spaces; this is the most useful form of convergence if one wants to study optimization (minimization and maximization) problems since it gives information of the convergence and existence of optimizers under suitable compactness assumptions of the underlying spaces [3]. In [C] we prove the Γ-convergence of the entropy-regularized functionals to the unregularized cost functional for multi-marginal optimal transportation problems with singular costs in Polish measure spaces.

5. Multi-marginal L^∞ transport

The last paper [D] of this thesis has a different nature to the previous three ones. The difference is two-fold. First, the cost function is, in general, not singular – not even repulsive. Second, even the form of the minimization is different to the standard integral Monge-Kantorovich problem. In the multi-marginal L^∞ transportation problem formulated in [D] one aims at mimimizing the quantity

$$C_\infty(\lambda) := \lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N)\in X^N} c(x_1,\ldots,x_N)$$

over all couplings λ of the marginals $\mu_1,\ldots,\mu_N$, that is, over the set $\Pi_N(\mu_1,\ldots,\mu_N)$. The two-marginal version of this problem was first extensively studied on $\mathbb{R}^d$ for the cost $c(x,y) = |x-y|$ by Champion, De Pascale, and Juutinen in [7] and then extended to general Polish spaces and more general (though mostly continuous) cost functions by Jylhä in [17]. The functional C_∞ is much more cumbersome than the Monge-Kantorovich functional C, since it is neither linear nor convex. Therefore, it also lacks simple dual interpretation, though a form of duality was established by Barron, Bocea, and Jensen in [1]. The name L^∞-transport originates from the fact, proven for $N = 2$ and $c(x,y) = |x-y|$ in [7], that if the cost is given by the distance function, then the transportation problem

$$\inf_{\lambda\in\Pi_2(\mu,\nu)} \lambda - \operatorname{ess\,sup} d(x,y)$$

is actually the $p\to\infty$ limit of the L^p-transportation problems

$$\inf_{\lambda\in\Pi_2(\mu,\nu)} \int_{X^2} d(x,y)\mathrm{d}\lambda\,.$$

In [D], we generalize these results to the multi-marginal language. We also generalize the study of Jylhä and Rajala [18] of the question when and how can optimal transportation costs be estimated from below by the L^∞ cost.

References

[1] E. N. Barron, M. Bocea, and R. R. Jensen, *Duality for the L^∞ optimal transport problem*, Trans. Amer. Math. Soc. **369** (2017), 3289–3323.

[2] J. D. Benamou, G. Carlier, and L. Nenna, *A numerical method to solve optimal transport problems with Coulomb cost*, to appear as a chapter in 'Splitting Methods in Communication and Imaging, Science and Engineering', Springer, (2015).

[3] A. Braides, *Gamma-convergence for Beginners*, vol. 22, Clarendon Press, (2002).

[4] Y. Brenier, *Polar Factorization and Monotone Rearrangement of Vector-Valued Functions*, Commu. Pure Applied Math., Vol. XLIV, 375–417 (1991).

[5] G. Buttazzo, T. Champion, and L. De Pascale, *Continuity and Estimates for Multimarginal Optimal Transportation Problem with Singular costs*, Appl. Math. Optim. **78** (2018), 185–200.

[6] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM J. Math. Anal., **49** (2017), 1385–1418.

[7] T. Champion, L. De Pascale, and P. Juutinen, *The ∞-Wasserstein distance: local solutions and existence of optimal transport maps*, SIAM J. Math. Anal. **40** (2008), 1-20.

[8] M. Colombo, S. Di Marino, and L. De Pascale, *Multimarginal Optimal Transport Maps for one-dimensional Repulsive Costs*, Canad. J. Math. **67** (2015), 350–368.

[9] C. Cotar and M. Petrache, *Next-order asymptotic expansion for N-marginal optimal transport with Coulomb and Riesz costs*, Adv. Math. **344** (2019), 137–233.

[10] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, Advances in neural information processing systems, 2292–2300, (2013)

[11] L. De Pascale, *Optimal transport with Coulomb cost, approximation and duality*, ESAIM Math. Model. Numer. Anal. **49** (2015), 1643–1657.

[12] S. Di Marino, A. Gerolin, and L. Nenna, *Optimal Transport Theory for repulsive costs*, Topological Optimization and Optimal Transport: In the Applied Sciences, **17**, (2017).

[13] W. Gangbo and R. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), 113–161.

[14] W. Gangbo and V. Oliker, *Existence of optimal maps in the reflector-type problems*, ESAIM Control Optim. Calc. Var. **13** (2007), 93–106.

[15] W. Gangbo and A. Swiech, *Optimal maps for the multidimensional Monge-Kantorovich problem*, Comm. Pure Appl. Math. **51** (1998), 23–45.

[16] N. Gigli and L. Tamanini, *Second order differentiation formula on compact $RCD^*(K,N)$ spaces*, arXiv:1701.03932, (2017).

[17] H. Jylhä, *The L^∞ Optimal Transport: infinite cyclical monotonicity and the existence of optimal transport maps*, Calc. Var. Partial Differential Equations **52** (2015), 303–326.

[18] H. Jylhä and T. Rajala, *L^∞ estimates in optimal mass transportation*, J. Funct. Anal. **270** (2016), 4297–4321.

[19] L.V. Kantorovich, *On the translocation of masses*, C.R. (Dokl.) Acad. Sci. URSS, **37** (1942), 199–201.

[20] H. Kellerer, *Duality theorems for marginal problems*, Z. Wahrsch. Verw. Gebiete **67** (1984), 399–432.

[21] Y. Kim and B. Pass, *A general condition for Monge solutions in the multi-marginal optimal transport problem*, SIAM J. Math. Anal., **46** (2014), 1538–1550.

[22] C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst. **34** (2014), 1533–1574.

[23] M. Lewin, E. H. Lieb, and R. Seiringer, *Statistical Mechanics of the Uniform Electron Gas*, J. Éc. polytech. Math. **5** (2018), 79–116.

[24] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport.* Ann. of Math. **169** (2009), 903–991.

[25] A. Moameni and B. Pass, *Solutions to multi-marginal optimal transport problems concentrated on several graphs*, ESAIM Control Optim. Calc. Var. **23** (2017), 551-567.

[26] G. Monge, *Mémoire sur la théorie des Deblais et des Rémblais*, Histoire de l'Academie des Sciences de Paris, 1781.

[27] P. Pegon, F. Santambrogio, and D. Piazzoli, *Full characterization of optimal transport plans for concave costs*, Discrete Contin. Dyn. Syst. **35** (2015), 6113–6132.

[28] A. Pratelli, *On the equality between Monge's infimum and Kantorovich's infimum in optimal mass transportation*, Ann. Inst. H. Poincaré Probab. Statist. **43** (2007), 1–13.

[29] E. Schrödinger, *Über die umkehrung der naturgesetze*, Sonderausgabe a. d. Sitz.-Ber. d. Preuss. Akad. d. Wiss., Phys.-math. Klasse, (1931).

[30] M. Seidl, *Strong-interaction limit of density-functional theory*, Phys. Rev. A., **60** (1999), 4387.

[31] M. Seidl, S. Di Marino, A. Gerolin, K. Giesbertz, L. Nenna, and P. Gori-Giorgi, *The Strictly-Correclated Electron functional for spherically symmetric systems revisited*, accepted in Phys. Rev. A. (arXiv:1702.05022) (2016).

[32] K.-T. Sturm, *On the geometry of metric measure spaces. I.* Acta Math. **196** (2006), 65–131.

[33] K.-T. Sturm, *On the geometry of metric measure spaces. II.* Acta Math. **196** (2006), 133–177.

[34] C. Villani, *Optimal transport. Old and new.* vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin (2009).

[35] X.-J. Wang, *On the design of a reflector antenna ii*, Calc. Var. Partial Differential Equations **20** (2004), 329–341.

Included articles

[**A**]

Non-existence of optimal transport maps for the multi-marginal repulsive harmonic cost

A. Gerolin, A. Kausamo, and T. Rajala

Preprint

NON-EXISTENCE OF OPTIMAL TRANSPORT MAPS FOR THE MULTI-MARGINAL REPULSIVE HARMONIC COST

AUGUSTO GEROLIN, ANNA KAUSAMO, AND TAPIO RAJALA

ABSTRACT. We give an example of an absolutely continuous measure μ on $\mathbb{R}^d$, for any $d \geq 1$, such that no minimizer of the 3-marginal harmonic repulsive cost with all marginals equal to μ is supported on a graph over the first variable.

1. INTRODUCTION

Let $\mu_1, \ldots, \mu_N$ be probability measures in $\mathbb{R}^d$, $c\colon (\mathbb{R}^d)^N \to \mathbb{R}$ be a cost function and denote by $\Gamma(\mu_1, \ldots, \mu_N)$ the set of probability measures $\gamma \in \mathscr{P}(\mathbb{R}^{dN})$ having marginals $\mu_1, \mu_2, \ldots, \mu_N$. We are interested in understanding the structure of the minimizers of the multi-marginal Monge-Kantorovich problem

$$\inf_{\gamma \in \Gamma(\mu_1, \ldots, \mu_N)} \int_{\mathbb{R}^{dN}} c(x_1, \ldots, x_N) \, \mathrm{d}\gamma(x_1, \ldots, x_N). \tag{1.1}$$

The problem has been studied recently in the literature for different cost functions c motivated by problems in economics [2, 3, 4], physics [7, 8, 11, 13, 18] and mathematics [1, 10, 14, 15]. We refer to [16] for an extensive discussion on multi-marginal optimal transport and to [9] for an introduction to the topic and for references in the context of optimal transport and density functional theory.

Among the cost functions that are considered in the literature, we concentrate in this paper to the attractive harmonic or Gangbo-Swiech cost c_a [1, 10] and to the repulsive harmonic cost c_w [9, 19],

$$c_a(x_1, \ldots, x_N) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} |x_j - x_i|^2, \quad c_w(x_1, \ldots, x_N) = -\sum_{i=1}^{N} \sum_{j=i+1}^{N} |x_j - x_i|^2.$$

Although the case $N = 2$ is well-understood for a wide class of cost functions depending on the distance, thanks to the celebrated Brenier's theorem [5], only partial results are currently known for the general case $N \geq 3$. In fact, when $N = 2$, if $h\colon \mathbb{R}^d \to \mathbb{R}^+$ is an increasing strictly convex function, $c(x_1, x_2) = h(|x_1 - x_2|)$, and μ_1 is absolutely continuous, then the problem (1.1) admits a unique solution $\gamma = (\mathrm{id}, T)_\sharp \mu_1$, which is concentrated on the graph of a map $T : \mathbb{R}^d \to \mathbb{R}^d$.

The main theorem of this paper is stated below. For simplicity, we denote by $\Gamma_N(\mu)$ the set of probability measures in $\mathbb{R}^{dN}$ with all the N marginals equal to μ.

Date: January 24, 2019.

The authors acknowledge the Academy of Finland projects no. 274372, 284511 and 312488.

Theorem 1.1. *Given $d \in \mathbb{N}$ there exists an absolutely continuous measure μ in $\mathbb{R}^d$ with bounded support such that there is no minimizer for the quantity*

$$\min_{\gamma \in \Gamma_3(\mu)} \int_{\mathbb{R}^{3d}} -|x_1 - x_2|^2 - |x_1 - x_3|^2 - |x_2 - x_3|^2 \, \mathrm{d}\gamma(x_1, x_2, x_3)$$

that is induced by a map from one of the marginals. Moreover, there is a unique symmetric minimizer.

We remark that when μ has finite second moments (or bounded support) and $N = 2$ the problem (1.1) corresponding to the cost function $c_w(x_1, x_2) = -|x_1 - x_2|^2$ admits a unique solution $\gamma = (\mathrm{id}, G)_\sharp\mu$, where G is the gradient of a concave function. This is a simple consequence of Brenier's theorem [5].

In [14], A. Moameni and B. Pass gave two general conditions on the cost functions in (1.1) which ensure that any solution must concentrate on either finitely many or countably many graphs. Moreover, the authors exhibit two examples of cost functions and marginals μ_1, μ_2, μ_3 such that there exists a unique minimizer which is concentrated in two graphs [14, Example 4.2] and [14, Example 4.4].

A few words about the proof of Theorem 1.1: We will first construct a measure μ on $\mathbb{R}$ and after that show the general case $d \geq 1$. The measure μ will be defined as a sum $\mu = \mu^1 + \mu^2$, where the measures μ^1 and μ^2 overlap so that this forces the optimal couplings to be non-graphical. The support of each measure μ^k consists of three connected components that are separated so that any optimal coupling is forced to have one marginal in each of the components. We use mainly two ingredients in the proof: (i) the fact that the support of an optimal transport γ for the repulsive harmonic cost must be contained in the set $\{x_1 + \cdots + x_N = k\}$, where $k \in \mathbb{R}^d$; (ii) The classical 2-marginal Brenier's theorem [5]. The main ideas are given in the proofs of Theorem 3.5 and Lemma 3.6.

Structure of the paper: In Section 2, we recall the main results regarding the attractive and repulsive harmonic cost. In particular, we show that in the 2-marginal case the Monge-Kantorovich problem (1.1) for both costs coincide under the hypothesis that the measure μ has finite second moments. Finally, in the Section 3, we present the main construction of the paper, namely prove the counterexample for the existence of Monge-type solutions in the repulsive harmonic case.

2. Multi-marginal optimal transport for the attractive and repulsive harmonic costs

2.1. **The attractive harmonic cost.** In [10], W. Gangbo and A. Swiech introduced the multi-marginal optimal transport problem for the attractive harmonic cost (also known as the Gangbo-Swiech cost)

$$c_a(x_1, \ldots, x_N) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} |x_i - x_j|^2, \tag{2.1}$$

and showed that the Monge-Kantorovich problem

$$\min_{\gamma \in \Gamma(\mu_1, \ldots, \mu_N)} \int_{\mathbb{R}^{dN}} c_a(x_1, \ldots, x_N) \, \mathrm{d}\gamma(x_1, \ldots, x_N), \tag{2.2}$$

admits a unique minimizer γ_{opt} provided that the marginals $\mu_1, \dots, \mu_N$ have finite second moments and are not concentrated in *small sets*. Moreover, γ_{opt} is of Monge-type, that is, $\gamma_{opt} = (\mathrm{id}, T_1 \dots, T_{N-1})_\sharp \mu_1$, where $T_i : \mathbb{R}^d \to \mathbb{R}^d$ satisfy $T_{i\sharp}\mu_1 = \mu_i$, for all $i \in \{1, \dots, N-1\}$.

For the sake of completeness, we briefly present some heuristics on this result. First, by opening the squares of $c_a(x_1, \dots, x_N)$, we notice that studying the minimizers of (2.2) is equivalent to studying the problem

$$\max_{\gamma \in \Gamma(\mu_1, \dots, \mu_N)} \int_{\mathbb{R}^{dN}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i \cdot x_j \,\mathrm{d}\gamma(x_1, \dots, x_N). \tag{2.3}$$

The main idea in [10] is to write the Kantorovich formulation of (2.3) as

$$\sup \left\{ \sum_{i=1}^{N} \int_{\mathbb{R}^d} u_i(x_i) \,\mathrm{d}\mu_i : \sum_{i=1}^{N} u_i(x_i) \leq -\sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i \cdot x_j \right\}, \tag{2.4}$$

and to show that the above supremum is attained by some functions $u_i \in L^1(\mu_i)$. Then, letting γ be an optimal transport plan for (2.3), and using the optimality conditions in (2.4), we see that the Kantorovich potentials u_i must satisfy the equality

$$u_1(x_1) + \dots + u_N(x_N) = -\sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i \cdot x_j, \text{ on } \mathrm{spt}(\gamma). \tag{2.5}$$

Now, applying the operators D_{x_j}, $j \in \{1, \dots, N\}$, on both sides of (2.5), one aims at solving the following system of equations

$$\begin{cases} D_{x_1} u_1(x_1) = -\sum_{i=2}^{N} x_i \\ \dots \\ D_{x_j} u_j(x_j) = -\sum_{i=1, i\neq j}^{N} x_i \\ \dots \\ D_{x_N} u_N(x_N) = -\sum_{i=1}^{N-1} x_i \end{cases} \quad \text{or, equivalently,} \quad \begin{cases} -x_1 + D_{x_1} u_1(x_1) = -\sum_{i=1}^{N} x_i \\ \dots \\ -x_j + D_{x_j} u_j(x_j) = -\sum_{i=1}^{N} x_i \\ \dots \\ -x_N + D_{x_N} u_N(x_N) = -\sum_{i=1}^{N} x_i \end{cases},$$

which is handily solvable for x_1 provided that one can, for each $j \in \{2, \dots, N\}$, invert the function $f_j \colon \mathbb{R}^d \to \mathbb{R}^d$,

$$f_j(x) = D_{x_j}(\tfrac{1}{2}|x|^2 - u_j(x)) \text{ for all } x \in \mathbb{R}^d.$$

Indeed, we have by the above system of equalities that $D_{x_1}(\frac{1}{2}|x_1|^2 - u_1(x_1)) = D_{x_j}(\frac{1}{2}|x_j|^2 - u_j(x_j))$ for all $j \in \{2, \dots, N\}$. Therefore, $x_j = f_j^{-1}(D_{x_1}(\frac{1}{2}|x_1|^2 - u_1(x_1)) =: T_j(x_1)$. The proof of the invertibility of functions f_j relies on the fact that those functions turn out to be strictly convex and to admit inverses which are their respective convex conjugates f_j^* (see [10], for details).

The above approach leads to the following result by Gangbo-Swiech [10].

Theorem 2.1 (Gangbo-Swiech, [10])**.** *Let $\mu_1, \dots, \mu_N$ be non-negative Borel probability measures in $\mathbb{R}^d$ vanishing on $(d-1)-$rectifiable sets and having finite second moments, and let c_a be the attractive harmonic cost* (2.1)*. Then*

(1) *the problem* (2.4) *admits maximizers u_i which are μ_i-differentiable.*

(2) *There exists a unique minimizer γ_{opt} in (2.2) such that $\gamma_{opt} = (id, T_1, \dots, T_{N-1})_{\sharp\mu_1}$, where $T_i\colon \mathbb{R}^d \to \mathbb{R}^d$ are defined by $T_i(x_1) = f_i^*(D_{x_1}(\frac{1}{2}|x_1|^2 + u_1(x_1))$ for $i = 2, \dots, N$ where $f_j : \mathbb{R}^d \to \mathbb{R}^d$,*

$$f_j(x) = D_{x_j}(\tfrac{1}{2}|x|^2 - u_j(x)) \text{ for all } x \in \mathbb{R}^d.$$

Theorem 2.1 was extended by H. Heinrich [12], allowing the inclusion of more general cost functions $c(x_1, \dots, x_N) = l(x_1 + \dots + x_N)$ with $l\colon \mathbb{R}^d \to \mathbb{R}^+$ concave. M. Agueh and G. Carlier [1] remarked that the attractive harmonic cost is equivalent to the so-called *barycenter problem* in multi-marginal optimal transport, which has many applications in inverse problems in imaging sciences. We refer to the survey [16] and the references therein for further applications and extensions of the attractive harmonic cost.

2.2. **The repulsive harmonic cost.** In this section, we consider cost functions of the type

$$c_w(x_1, \dots, x_N) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} -|x_i - x_j|^2,$$

or, more generally,

$$c_h(x_1, \dots, x_N) = h(x_1 + x_2 + \dots + x_N),$$

where $h\colon \mathbb{R}^d \to \mathbb{R}^+$ is a convex function. Costs, such as c_w, are said to be *repulsive* because optimal transport plans must place points x_i as far as possible from each other. We show in Subsection 2.3, that in the case where the number of marginals N is equal to 2, the attractive and repulsive costs are equivalent via Brenier's theorem. In particular, the Monge-Kantorovich problem associated with the repulsive harmonic cost admits a unique minimizer which is of Monge-type.

This situation, however, changes drastically in the multi-marginal case where generally there is no hope of uniqueness of optimal transport plans for the Monge-Kantorovich problem associated with the cost c_w. Moreover, examples of diffuse-like, fractal-like, and Monge-like solutions can be constructed [9].

Although the most interesting cost in the class of repulsive costs is the Coulomb one, since it plays an important role in density functional theory [11, 19], the repulsive harmonic cost has been used as a toy-model to study problems in that context. As observed by physicists [19, 6], it has several advantages compared to the Coulomb one, since easy Monge-type solutions can be constructed in any dimension d for specific densities. Additionally, the repulsive harmonic cost admits a very simple characterization for optimality (see Proposition 2.3).

We now recall a few results and central examples in the repulsive harmonic cost case. First, we notice that the Monge-Kantorovich problem associated with c_w is equivalent to the one associated with $c_h(x_1, \dots, x_N) = h(x_1 + \dots + x_N)$ when $h(z) = |z|^2$.

Proposition 2.2. *Assume that $\mu_1, \dots, \mu_N$ are probability measures in $\mathbb{R}^d$ with finite second moments. Then,*

$$\underset{\gamma\in\Gamma(\mu_1,\dots,\mu_N)}{\operatorname{argmin}} \int_{\mathbb{R}^{dN}} c_w(x_1, \dots, x_N)\,\mathrm{d}\gamma = \underset{\gamma\in\Gamma(\mu_1,\dots,\mu_N)}{\operatorname{argmin}} \int_{\mathbb{R}^{dN}} |x_1 + \dots + x_N|^2\,\mathrm{d}\gamma.$$

As noticed in [9], the Monge-Kantorovich problem for cost functions c_h have a simple characterization.

Proposition 2.3 ([9])**.** *Assume that $\mu_1, \dots, \mu_N$ are probability measures in $\mathbb{R}^d$ with finite second moments and $c_h(x_1, \dots, x_N) = h(x_1 + \cdots + x_N)$, where $h\colon \mathbb{R}^d \to \mathbb{R}^+$ is a strictly convex function. Then, if there exists a plan γ concentrated in some hyperplane $\{x_1 + \cdots + x_N = k\}$, $k \in \mathbb{R}^d$, then γ is an optimal transport plan for the problem*

$$\min_{\gamma \in \Gamma(\mu_1,\dots,\mu_N)} \int_{\mathbb{R}^{dN}} h(x_1 + \cdots + x_N)\, \mathrm{d}\gamma(x_1, \dots, x_N). \tag{2.6}$$

In this case, $\tilde{\gamma} \in \Gamma(\mu_1, \dots, \mu_N)$ is optimal in (2.6), if and only if,

$$\operatorname{spt}(\tilde{\gamma}) \subset \{(x_1, \dots, x_N) \in \mathbb{R}^{dN} \mid x_1 + \cdots + x_N = k\}. \tag{2.7}$$

In particular, the constant vector k can be computed explicitly as

$$k = \sum_{j=1}^{N} \int x \, \mathrm{d}\mu_j(x).$$

In the multi-marginal setting, the optimality condition

$$\operatorname{spt}(\gamma) \subset \Big\{(x_1, \dots, x_N) \in \mathbb{R}^{dN} \mid x_1 + \cdots + x_N = k\Big\}$$

is much easier to handle than the c-cyclical monotonicity and the optimality conditions given by the Kantorovich problem.

We list some examples of minimizers in (2.6) that can be constructed via Proposition 2.3. An interesting case for repulsive costs is when all the N marginals are the same and given by an absolutely continuous measure $\mu = \rho\mathcal{L}^d$.

- *Diffuse optimal plan or "fat" plan:*
  Suppose $d = 1$, $N = 3$ and the measure μ is defined by $\mu = \frac{1}{2}\mathcal{L}|_{[-1,1]}$. Let $\gamma = \frac{1}{2}\mathcal{H}^2|_H g(\max\{|x_1|, |x_2|, |x_3|\})$, where H is defined as $H = \{x_1 + x_2 + x_3 = 0\} \cap \{|x| \leq 1, |y| \leq 1, |z| \leq 1\}$, $g(x) = \frac{\sqrt{3}}{6}x$ and $\mathcal{H}^2$ denotes the 2-dimensional Hausdorff measure. Then, by Proposition 2.3, γ is an optimal plan, since $\gamma \in \Pi_3(\mu)$ and $k = 3\int x d\mu = 0$. Moreover, this plan is not concentrated on a graph of a map [9].

- *Fractal-like optimal plan concentrated on a graph of a map* [9]*:*
  Let $\mu = \mathcal{L}^d|_{[0,1]^d}$ be the uniform measure in the d-dimensional unit cube and $N \geq 3$. Every point $z \in [0, 1]$ can be represented in its N-th base. Namely, $z = \sum_{k=1}^{\infty} a_k/N^k$ with $a_k \in \{0, \dots, N-1\}$. We define a map T by permuting the N symbols of a_k.

  $$T : z \in [0, 1] \mapsto T(z) = \sum_{k=1}^{\infty} \mathcal{S}(a_k)/N^K \in [0, 1],$$
  $$\text{where } \mathcal{S}(i) = i + 1 \text{ and } \mathcal{S}(N-1) = 0.$$

  One can show that [9], if $x_1 = (z_1, \dots, z_N)$ the map $\tilde{T}(x_1) = (T(z_1), T(z_2), \dots, T(z_N))$ preserves the uniform measure in the d-dimensional cube and has the property that

  $$x_1 + \tilde{T}(x_1) + \cdots + \tilde{T}^{(N-1)}(x_N) = N/2 = N\left(\int_{\mathbb{R}^d} x_1 \, \mathrm{d}\mu(x_1)\right)^2.$$

  Therefore, by Proposition 2.3 the plan $\gamma = (\mathrm{id}, \tilde{T}, \dots, \tilde{T}^{(N)})_\sharp\mu$ is optimal. The map $\tilde{T}$ is not continuous at any point [9].

- *Optimal plan concentrated on the graph of a regular cyclical map (N even):* Suppose that, for all $i = 1, \dots, 2k$, $\mu_i = \mu = \rho\mathcal{L}^d$ is such that $\rho(|x|) = \rho(x)$ for all $x \in \mathbb{R}^d$. Then $\gamma(x_1, \dots, x_{2k}) = (\mathrm{id}, T, T^{(2)}, \dots, T^{(2k-1)})_\sharp\mu$ where $T\colon x \in \mathbb{R}^d \mapsto -x \in \mathbb{R}^d$ is an optimal plan for (2.6). In fact,
$$N \int_{\mathbb{R}^d} x\rho(x)\,\mathrm{d}x = 0$$
and
$$x + T(x) + T^{(2)}(x) + \cdots + T^{(N-1)}(x) = 0.$$

2.3. **Discussion on the $N = 2$ case and Brenier's Theorem.** We start by recalling the classical Brenier's theorem in optimal transport. Notice that the attractive harmonic cost is a natural generalization of the distance squared cost when $N > 2$.

Theorem 2.4 (Brenier, [5])**.** *Let μ_1 and μ_2 be Borel probability measures in $\mathbb{R}^d$ with finite second moments. Assume that μ_1 is absolutely continuous with respect to the Lebesgue measure. Then the problem*
$$\min_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2\,\mathrm{d}\gamma(x_1, x_2),$$
admits a unique minimizer $\gamma_{opt} = (id, T)_\sharp\mu_1$, where $T(x_1) = \nabla\phi(x_1)$, and $\phi\colon \mathbb{R}^d \to \mathbb{R}$ is a convex function. As a consequence, the Monge problem with squared distance cost
$$\min\left\{ \int_{\mathbb{R}^d} |x_1 - T(x_1)|^2\,\mathrm{d}\mu_1(x_1) \,:\, T\colon \mathbb{R}^d \to \mathbb{R}^d, T_\sharp\mu_1 = \mu_2 \right\},$$
admits a unique minimizer.

We call the optimal map T the *Brenier's map.* In the two marginal case, a solution of the Monge-Kantorovich problem for the repulsive harmonic cost is an immediate consequence of Brenier's theorem.

Corollary 2.5. *Let μ_1 and μ_2 be Borel probability measures in $\mathbb{R}^d$ having finite second moments. Assume that μ_1 is absolutely continuous with respect to the Lebesgue measure. Then the problem*
$$\min_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} -|x_1 - x_2|^2\,\mathrm{d}\gamma(x_1, x_2),$$
admits a unique minimizer $\gamma_{opt} = (id, T)_\sharp\mu_1$, where $T(x_1) = \nabla\psi(x_1)$, and $\psi : \mathbb{R}^d \to \mathbb{R}$ is a concave function.

Proof. First notice that, since μ_1 and μ_2 have finite second moments,
$$\begin{aligned} 2\int_{\mathbb{R}^d} |x_1|^2\,\mathrm{d}\mu_1 + 2\int_{\mathbb{R}^d} |x_2|^2\,\mathrm{d}\mu_2 &+ \min_{T_\sharp\mu_1 = \mu_2} \int_{\mathbb{R}^d} -|x_1 - T(x_1)|^2\,\mathrm{d}\mu_1 \\ &= \min_{G_\sharp\mu_1 = \tilde{\mu_2}} \int_{\mathbb{R}^d} |x_1 - G(x_1)|^2\,\mathrm{d}\mu_1, \end{aligned}$$
where $G = -T$ and $\tilde{\mu_2} = (-\mathrm{id})_\sharp\mu_2$. By Brenier's theorem the problem on the right-hand side has a unique solution $G = \nabla\phi$, with $\phi\colon \mathbb{R}^d \to \mathbb{R}$ a convex function. Then, $T = \nabla\psi$ is an optimal map for the repulsive harmonic cost, which is the gradient of a concave function $\psi = -\phi$. $\square$

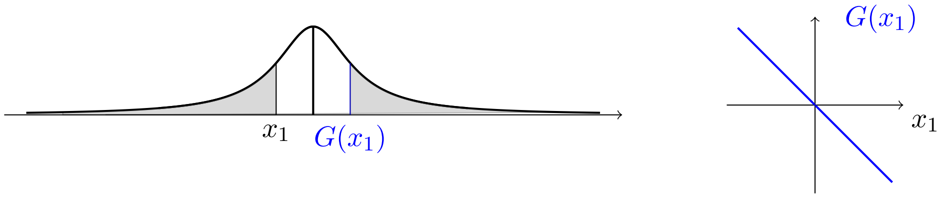

FIGURE 1. Example of an optimal map given by Corollary 2.5 in the one-dimensional case $d = 1$: $\mu = \mu_1 = \mu_2$ is the Gaussian-like shape density in the left-side picture. The map $G\colon \mathbb{R} \to \mathbb{R}$ is the anti-monotone map $G(x) = -x$ (right-hand picture).

3. Repulsive Harmonic cost: Monge is equal to Monge-Kantorovich?

Let μ be an absolutely continuous probability measure with a finite second moment. It is natural to inquire if there exists a Monge-type minimizer for the Monge-Kantorovich problem ($N \geq 3$)

$$\min_{\gamma\in\Gamma_N(\mu)} \int_{\mathbb{R}^{dN}} c_w(x_1,\dots,x_N)\,\mathrm{d}\gamma(x_1,\dots,x_N),$$

or, for the cost $c_h(x_1,\dots,x_N) = h(x_1+\cdots+x_N)$, with $h\colon \mathbb{R}^d \to \mathbb{R}^+$ strictly convex.

A particular case of a theorem of A. Pratelli [17] states that the minimum of the Monge-Kantorovich problem coincides with the infimum of the Monge problem for a large class of cost functions, including Coulomb and c_h cost functions.

Theorem 3.1 (A. Pratelli, [17])**.** *Assume that μ is a Borel probability measure without atoms and $c\colon (\mathbb{R}^d)^N \to \mathbb{R}\cup\{+\infty\}$ is a continuous and bounded from below cost function. Then,*

$$\min_{\gamma\in\Gamma_N(\mu)} \int_{\mathbb{R}^{dN}} c\,\mathrm{d}\gamma = \inf\left\{ \int_{\mathbb{R}^d} c(x,T_1(x),\dots,T_{N-1}(x))\,\mathrm{d}\mu(x) :\ T_{i\sharp}\mu = \mu \right\}, \tag{3.1}$$

where $T_i\colon \mathbb{R}^d \to \mathbb{R}^d$ are Borel functions and $i \in \{1,\dots,N-1\}$.

In other words, we are interested in knowing if the above infimum in (3.1) is achieved. The following counterexample proves that this is generally not the case for the repulsive harmonic cost c_w or, more generally, for cost functions $c_h(x_1,\dots,x_N) = h(\sum x_i)$ depending on a convex function $h\colon \mathbb{R}^d \to \mathbb{R}$.

Let us recall the notion of a symmetric transport plan.

Definition 3.2 (Symmetric measures)**.** *A measure $\gamma \in \mathcal{P}(\mathbb{R}^{dN})$ is symmetric if*

$$\int_{\mathbb{R}^{dN}} g(x_1,\dots,x_N)\,\mathrm{d}\gamma = \int_{\mathbb{R}^{dN}} g(\sigma(x_1,\dots,x_N))\,\mathrm{d}\gamma, \text{ for all } g \in \mathcal{C}(\mathbb{R}^{dN})$$

and for all permutations of N symbols $\sigma \in \mathfrak{S}_N$. We denote by $\Gamma_N^{sym}(\mu)$, the space of all $\gamma \in \Gamma_N(\mu)$ which are symmetric.

The following result follows by observing that

(i) $\Gamma_N^{sym}(\mu) \subset \Gamma_N(\mu)$;
(ii) if $\gamma \in \Gamma_N(\mu)$ then $\gamma_{sym} = \frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N} \sigma_\sharp\gamma \in \Gamma_N^{sym}(\mu)$.

Proposition 3.3. *If μ is an absolutely continuous measure in $\mathbb{R}^d$ with finite second moments, then*

$$\min_{\gamma\in\Gamma_N(\mu)} \int_{\mathbb{R}^{dN}} c_w(x_1,\dots,x_N)\,\mathrm{d}\gamma = \min_{\gamma\in\Gamma_N^{sym}(\mu)} \int_{\mathbb{R}^{dN}} c_w(x_1,\dots,x_N)\,\mathrm{d}\gamma. \tag{3.2}$$

3.1. **Construction of the counterexample.** The main goal of this section is to exhibit an absolutely continuous measure μ such that, for the repulsive harmonic problem (2.6), there is no Monge-type minimizer. We recall our main theorem.

Theorem 3.4. *Given $d \in \mathbb{N}$ there exists an absolutely continuous measure μ in $\mathbb{R}^d$ with bounded support such that there is no minimizer for the quantity*

$$\min_{\gamma\in\Gamma_3(\mu)} \int_{\mathbb{R}^{3d}} -|x_1-x_2|^2 - |x_1-x_3|^2 - |x_2-x_3|^2\,\mathrm{d}\gamma(x_1,x_2,x_3) \tag{3.3}$$

that is induced by a map from one of the marginals. Moreover, there is a unique symmetric minimizer.

Let us first define the absolutely continuous measure μ in $\mathbb{R}$, since all the ideas of the construction are present already in the one-dimensional case. Then, in Subsection 3.4 we give the necessary modifications for the higher dimensions.

The measure μ can be understood as a sum $\mu = \mu^1 + \mu^2$, where the support of both measures μ^k consists of three connected components that are separated so that any optimal coupling is forced to have one marginal in each of the components. Let us now define the connected components. Let

$$C = [0, 1/2]\,,$$

and for $k \in \{1, 2\}$, let

$$R^k = \left[3^k, 3^k + 1/2\right] \qquad \text{and} \qquad L^k = \left[-3^k - 1, -3^k\right].$$

Using these we define $R = R^1 \cup R^2$ and $L = L^1 \cup L^2$.

In order to study the structure of the minimizers, let us consider μ as the a sum

$$\mu = \frac{1}{3}\left(\mu_L + \mu_C + \mu_R\right), \tag{3.4}$$

where different components correspond to the measure on the left, center and right, respectively. We define these parts as

$$\mu_L = \frac{1}{2}\mathcal{L}|_L, \qquad \mu_C = 2\mathcal{L}|_C, \qquad \mu_R = \mathcal{L}|_R. \tag{3.5}$$

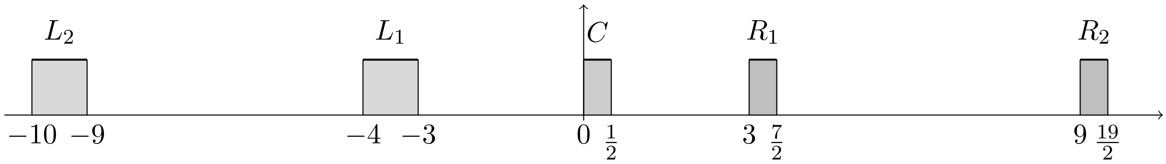

FIGURE 2. Sketch of the support of μ.

3.2. **Unique non-graphical minimizer in $\Gamma(\mu_C, \mu_R, \mu_L)$.** Let us first state a version of Theorem 3.4 where the marginals are given by μ_C, μ_R, μ_L. Theorem 3.5 contains the main ideas in the proof of Theorem 3.4; the existence of a unique non-graphical minimizer.

Theorem 3.5. *Let μ_L, μ_C, μ_R be the absolutely continuous measures defined in* (3.5)*. Then there exists a unique minimizer γ_0 of*

$$\min_{\gamma\in\Gamma(\mu_C,\mu_R,\mu_L)} \int_{\mathbb{R}^3} -|x_1-x_2|^2 - |x_1-x_3|^2 - |x_2-x_3|^2 \,\mathrm{d}\gamma(x_1,x_2,x_3) \tag{3.6}$$

which is given by

$$\gamma_0 = \frac{1}{2}\sum_{k=1}^{2}\left((H_k)_\sharp \mathcal{L}_{|C}\right),$$

where

$$H_k(x) = (x, x+3^k, -2x-3^k).$$

In particular, this minimizer is not induced by a map from the first coordinate.

Proof. By construction, the projections of γ_0 are μ_C, μ_R and μ_L. Moreover, since for all $k \in \{1,2\}$ and $x \in \mathbb{R}$, writing $H_k(x) = (x_1, x_2, x_3)$, we have

$$x_1 + x_2 + x_3 = x + x + 3^k - 2x - 3^k = 0, \tag{3.7}$$

we get from Proposition 2.3 that the measure γ_0 is a minimizer. Since the disintegration of γ_0 with respect to the projection to the first coordinate is for μ_C-almost every $x \in \mathbb{R}$ unique and equal to

$$\frac{1}{2}\delta_{(x,x+3,-2x-3)} + \frac{1}{2}\delta_{(x,x+9,-2x-9)},$$

we conlculde that γ_0 is not induced by a map from the first coordinate.

What remains to show is that γ_0 is the unique minimizer in (3.6). Let γ be a minimizer of (3.6). Since for γ_0 we have by (3.7) that

$$\int_{\mathbb{R}^3} |x_1+x_2+x_3|^2 \,\mathrm{d}\gamma_0(x_1,x_2,x_3) = 0,$$

we conclude that the same must hold for γ. Consequently, for any $(x_1,x_2,x_3) \in \operatorname{spt}(\gamma)$ we have

$$x_1 + x_2 + x_3 = 0. \tag{3.8}$$

The idea of the proof is now to look at the variational problem

$$\min\left\{ \int |x_2+x_3|^2 \,\mathrm{d}\omega(x_2,x_3) \,\middle|\, \omega \in \Gamma(\mu_R,\mu_L) \right\}.$$

By Corollary 2.5, there exists a unique minimizer $\omega_a = (\mathrm{id}, G)_\sharp\mu_R$, which is given by the anti-monotone map G. Let $T(x_2,x_3) = -x_2 - x_3$. Since $\mathcal{L}_{|C} = T_\sharp(\omega_a)$ and

$$\int |x_1|^2 \,\mathrm{d}T_\sharp(\omega)(x_1) = \int |x_2+x_3|^2 \,\mathrm{d}\omega(x_2,x_3),$$

we have that ω_a is the unique coupling of μ_R and μ_L that is pushed to $\mathcal{L}_{|C}$ by T. Recall that by Proposition 2.3, we have $x_1 = -x_2 - x_3$ for any $(x_1,x_2,x_3) \in \operatorname{spt}(\gamma)$. Therefore, $\gamma = S_\sharp(\omega_a) = \gamma_0$, with $S(x_2,x_3) = (-x_2-x_3, x_2, x_3)$. We conclude that γ_0 is the unique minimizer of (3.6). □

3.3. **Structure of the minimizers in in $\Gamma_3(\mu)$.** Let us now study the minimizers in the case where all the three marginals are μ. We start with the following lemma.

Lemma 3.6. *Let γ be a minimizer in (3.3) and $(x_1, x_2, x_3) \in \operatorname{spt}(\gamma)$. Then there exist $k \in \{1, 2\}$ and a permutation σ of $\{1, 2, 3\}$ so that*

$$x_{\sigma(1)} \in L^k, \qquad x_{\sigma(2)} \in C, \qquad x_{\sigma(3)} \in R^k. \tag{3.9}$$

Proof. We prove the claim by ruling out the other possibilities. By applying a permutation, if necessary, we may assume that $x_1 \leq x_2 \leq x_3$.

If $x_1, x_2, x_3 \in C \cup R^1 \cup R^2$, then (3.8) forces $x_1 = x_2 = x_3 = 0$. Then by the definition of the support and the fact that $\mu(\{0\}) = 0$, there exist points $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) \in \operatorname{spt}(\gamma) \setminus \{(x_1, x_2, x_3)\}$ arbitrarily close to (x_1, x_2, x_3). For close enough points, the form of μ then gives that $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) \in C^3$ and thus $\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 > 0$, contradicting (3.8). Thus $x_1 \in L^1 \cup L^2$.

If now $x_2 \in L^1 \cup L^2$, then by (3.8),

$$x_3 \in [6, 8] \cup [12, 14] \cup [18, 20].$$

But then, $x_3 \notin \operatorname{spt}(\mu)$, giving a contradiction. Therefore, $x_2, x_3 \in C \cup R^1 \cup R^2$.

If $x_2, x_3 \in C$, then $x_1 + x_2 + x_3 < 0$. Thus, $x_3 \in R^1 \cup R^2$. In order to $x_2 \in \operatorname{spt}(\mu)$, the only possibility is then to have $x_1 \in L^1$ and $x_3 \in R^1$, or $x_1 \in L^2$ and $x_3 \in R^2$. In both cases, one must have $x_2 \in C$. This proves the claim. $\square$

Proof of Theorem 3.4. Let γ be a minimizer in (3.3). Let us translate the problem to the setting of Theorem 3.5 using the following mapping $F \colon \mathbb{R}^3 \to \mathbb{R}^3$ defined as

$$F((x_1, x_2, x_3)) = \begin{cases} (x_3, x_2, x_1), & \text{if } (x_1, x_2, x_3) \in L \times R \times C, \\ (x_3, x_1, x_2), & \text{if } (x_1, x_2, x_3) \in R \times L \times C, \\ (x_2, x_3, x_1), & \text{if } (x_1, x_2, x_3) \in L \times C \times R, \\ (x_2, x_1, x_3), & \text{if } (x_1, x_2, x_3) \in R \times C \times L, \\ (x_1, x_3, x_2), & \text{if } (x_1, x_2, x_3) \in C \times L \times R, \\ (x_1, x_2, x_3), & \text{otherwise.} \end{cases}$$

The map F permutes the coordinates differently in different parts of the space so that γ-almost every point will be mapped into $C \times R \times L$, due to Lemma 3.6. Since F is pointwise just a permutation, we have that γ-almost every $(x_1, x_2, x_2) \in \mathbb{R}^3$ the point $(y_1, y_2, y_3) = F((x_1, x_2, x_3))$ satisfies

$$y_1 + y_2 + y_3 = 0.$$

Therefore, by Proposition 2.3, the measure $F_\sharp\gamma$ is an optimal coupling (for its marginal measures). We claim that $F_\sharp\gamma \in \Pi(\mu_C, \mu_R, \mu_L)$. Let us check this for the first marginal. Let $A \subset \mathbb{R}$. Then by Lemma 3.6 and the definition of F, we get

$$\begin{aligned} F_\sharp\gamma(A \times \mathbb{R} \times \mathbb{R}) &= F_\sharp\gamma((A \cap C) \times \mathbb{R} \times \mathbb{R}) \\ &= \gamma((A \cap C) \times \mathbb{R} \times \mathbb{R}) + \gamma(\mathbb{R} \times (A \cap C) \times \mathbb{R}) + \gamma(\mathbb{R} \times \mathbb{R} \times (A \cap C)) \\ &= 3\mu(A \cap C) = \mu_C(A). \end{aligned}$$

The claim for the other marginals follows similarly.

Thus, due to the optimality of $F_\sharp\gamma$, from Theorem 3.5 we conclude that $F_\sharp\gamma = \gamma_0$ with the measure γ_0 defined in Theorem 3.5. Now, the form of $\operatorname{spt}(\gamma_0)$ forces

$$(\mathrm{p}_2)_\sharp(\gamma|_{R \times \mathbb{R} \times \mathbb{R}}) + (\mathrm{p}_3)_\sharp(\gamma|_{R \times \mathbb{R} \times \mathbb{R}}) = \frac{1}{2}\mu|_{C \cup L}, \tag{3.10}$$

where $\mathtt{p}_i$ is the projection to the i:th coordinate. Similarly,

$$(\mathtt{p}_2)_\sharp(\gamma|_{L\times\mathbb{R}\times\mathbb{R}}) + (\mathtt{p}_3)_\sharp(\gamma|_{L\times\mathbb{R}\times\mathbb{R}}) = \frac{1}{2}\mu|_{C\cup R}. \tag{3.11}$$

Together (3.10) and (3.11) imply

$$(\mathtt{p}_2)_\sharp(\gamma|_{C\times\mathbb{R}\times\mathbb{R}}) + (\mathtt{p}_3)_\sharp(\gamma|_{C\times\mathbb{R}\times\mathbb{R}}) = \frac{1}{2}\mu|_{L\cup R}.$$

Therefore, again from the form of γ_0, we conclude that for μ-almost every $x \in C$ the disintegration of

$$(\mathtt{p}_1,\mathtt{p}_2)_\sharp(\gamma|_{C\times\mathbb{R}\times\mathbb{R}}) + (\mathtt{p}_1,\mathtt{p}_3)_\sharp(\gamma|_{C\times\mathbb{R}\times\mathbb{R}})$$

with respect to the projection on the first coordinate gives positive measure for at least 4 points and thus $\gamma|_C$ is not induced by a map from the first coordinate. Since the argument is symmetric with respect to the marginals, we conclude that γ is not induced by a map from any of the coordinates.

Let us then show the uniqueness of the symmetric minimizer in (3.3). First of all, by Proposition 3.3 we already know that there exists a symmetric minimizer. We will show that the unique symmetric minimizer is the plan $\gamma_1 \in \Gamma_3^{sym}(\mu)$ defined by

$$\gamma_1 = \frac{1}{6}\sum_\sigma \sigma_\sharp\gamma_0,$$

where the sum is taken over all permutations of the marginals $\sigma\colon \mathbb{R}^3 \to \mathbb{R}^3$. Since $x_1+x_2+x_3 = 0$ for γ_1-almost every (x_1, x_2, x_3), by Proposition 2.3 the plan γ_1 is a minimizer. Now, let $\gamma \in \Gamma_3^{sym}(\mu)$ be a minimizer of (3.3). By symmetry, we know that

$$\sigma_\sharp(\gamma|_{\sigma^{-1}(C\times R\times L)}) = \gamma|_{C\times R\times L}$$

for all permutations σ of the marginals. Thus by Lemma 3.6 and the fact that $F_\sharp\gamma = \gamma_0$, we get

$$\gamma_0 = \sum_\sigma \sigma_\sharp(\gamma|_{\sigma^{-1}(C\times R\times L)}) = 6\gamma|_{C\times R\times L},$$

implying $\gamma = \gamma_1$. □

3.4. **The higher-dimensional case.** Let us generalize the previous example to the higher dimensions. Using the same notation C, L^k, R^k for $k \in \{1, 2\}$ as in the one dimensional case, we write

$$C_d = C^d,\ L_d^k = (L^k)^d,\ \text{and}\ R_d^k = (R^k)^d$$

and

$$L_d = L_d^1 \cup L_d^2,\ \text{and}\ R_d = R_d^1 \cup R_d^2.$$

Let us denote by $\mathcal{L}$ the Lebesgue measure on $\mathbb{R}^d$. The marginal measures are now

$$\mu_L = \frac{1}{2}\mathcal{L}|_{L_d}, \qquad \mu_C = 2^d\mathcal{L}|_{C_d}, \qquad \mu_R = 2^{d-1}\mathcal{L}|_{R_d}.$$

The optimal coupling γ_0 in $\Pi(\mu_C, \mu_R, \mu_L)$ now reads

$$\gamma_0 = \sum_{k=1}^{2}\left((H_k)_\sharp 2^{d-1}\mathcal{L}|_{C_d}\right),$$

where for $k \in \{1,2\}$, the maps $H_k \colon \mathbb{R}^d \to \mathbb{R}^{3d}$ are given by

$$T_k(x) = (x, x + (3^k, \dots, 3^k), -2x - (3^k, \dots, 3^k))\,.$$

We again have the optimality conditition (2.7). This gives us that for each minimizer γ of c we must have, for each point $(x_1, x_2, x_3) \in \mathrm{spt}(\gamma)$ the equality

$$x_1 + x_2 + x_3 = 0.$$

The uniqueness of optimal $\gamma_0 \in \Pi(\mu_C, \mu_R, \mu_L)$ follows as in the proof of Theorem 3.5 using the functional $F \colon \Gamma(\mu_L, \mu_R) \to \mathbb{R}$

$$F(\omega) = \int_{\mathbb{R}^{2d}} \left(|x_1^1 + x_3^1|^2 + \cdots + |x_1^d + x_3^d|^2 \right) \mathrm{d}\omega(x_1, x_3)\,,$$

and the component-wise anti-monotone transport ω_a of μ_L to μ_R.

The analysis of the minimizers with marginals $\mu = \frac{1}{3}(\mu_C + \mu_L + \mu_R)$, Lemma 3.6 and the proof of Theorem 3.4 go as in the one dimensional case, by component-wise considerations.

References

[1] M. Agueh and G. Carlier, *Barycenters in the Wasserstein space*, SIAM J. on Mathematical Analysis, 43 (2011), pp. 904–924.

[2] M. Beiglböck, P. Henry-Labordère, and F. Penkner, *Model-independent bounds for option prices—a mass transport approach*, Finance and Stochastics, 17 (2013), pp. 477–501.

[3] M. Beiglboeck, A. Cox, and M. Huesmann, *Optimal transport and Skorokhod embedding*, Inventiones mathematicae, 208 (2017).

[4] M. Beiglboeck and N. Juillet, *On a problem of optimal transport under marginal martingale constraints*, to appear Ann. Probab. ArXiv:1208.1509, (2012).

[5] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, CPAM, 44 (1991), pp. 375–417.

[6] L. Cort, D. Karlsson, G. Lani, and R. van Leeuwen, *Time-dependent Density-Functional Theory for Strongly Interacting Electrons*, Physical Review A, 95 (2017), p. 042505.

[7] C. Cotar, G. Friesecke, and C. Klüppelberg, *Smoothing of transport plans with fixed marginals and rigorous semiclassical limit of the Hohenberg-Kohn functional*, arXiv:1706.05676, (2017).

[8] S. Di Marino, L. De Pascale, and M. Colombo, *Multimarginal optimal transport maps for 1-dimensional repulsive costs*, Canadian Journal of Mathematics – Journal Canadien des Mathématiques, 67 (2015), pp. 350–368.

[9] S. Di Marino, A. Gerolin, and L. Nenna, *Optimal transport theory for repulsive costs*, Topological Optimization and Optimal Transport: In the Applied Sciences, 17 (2017).

[10] W. Gangbo and A. Swiech, *Optimal maps for the multidimensional Monge-Kantorovich problem*, Communications on pure and applied mathematics, 51 (1998), pp. 23–45.

[11] P. Gori-Giorgi, M. Seidl, and G. Vignale, *Density Functional Theory for strongly interacting electrons*, Physical review letters, 103 (2009), p. 166402.

[12] H. Heinich, *Problème de Monge pour n probabilités*, Comptes Rendus Mathematique, 334 (2002), pp. 793–795.

[13] G. Lani, S. Di Marino, A. Gerolin, R. van Leeuwen, and P. Gori-Giorgi, *The adiabatic strictly-correlated-electrons functional: kernel and exact properties*, Physical Chemistry Chemical Physics, 18 (2016), pp. 21092–21101.

[14] A. Moameni and B. Pass, *Solutions to multi-marginal optimal transport problems concentrated on several graphs*, ESAIM: Control Optim. Calc. Var., (2017), pp. 551–567.

[15] B. Pass, *Multi-marginal optimal transport and multi-agent matching problems: uniqueness and structure of solutions*, Discrete Contin. Dyn. Syst., 34:1623-1639, (2014).

[16] ———, *Multi-marginal optimal transport: theory and applications*, ESAIM: Math. Model. Numer. Anal. (Special issue on "Optimal transport in applied mathematics"), 49 (2015), pp. 1771–1790.

[17] A. Pratelli, *On the equality between Monge's infimum and kantorovich's minimum in optimal mass transportation*, in Annales de l'Institut Henri Poincare (B) Probability and Statistics, vol. 43, Elsevier, 2007, pp. 1–13.
[18] M. Seidl, S. Di Marino, A. Gerolin, K. Giesbertz, L. Nenna, and P. Gori-Giorgi, *The Strictly-Correlated Electron functional for spherically symmetric systems revisited*, Accepted in Physical Review A (available in arXiv:1702.05022), (2016).
[19] M. Seidl, P. Gori-Giorgi, and A. Savin, *Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities*, Physical Review A, 75 (2007), p. 042511.

University of Jyvaskyla, Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of Jyvaskyla, Finland
E-mail address: augustogerolin@gmail.com
E-mail address: anna.m.kausamo@jyu.fi
E-mail address: tapio.m.rajala@jyu.fi

[**B**]

Duality theory for multimarginal optimal transport with repulsive costs in metric spaces

A. Gerolin, A. Kausamo, and T. Rajala

To appear in ESAIM Control Optim. Calc. Var.

DUALITY THEORY FOR MULTI-MARGINAL OPTIMAL TRANSPORT WITH REPULSIVE COSTS IN METRIC SPACES

AUGUSTO GEROLIN, ANNA KAUSAMO, AND TAPIO RAJALA

Abstract. In this paper we extend the duality theory of the multi-marginal optimal transport problem for cost functions depending on a decreasing function of the distance (not necessarily bounded). This class of cost functions appears in the context of SCE Density Functional Theory introduced in *Strong-interaction limit of density-functional theory* by M. Seidl [27].

Contents

1. Introduction

We consider the following multi-marginal optimal transport (MOT) problem

$$\inf_{\gamma\in\Gamma(\rho)} \int_{X^N} c(x_1,\dots,x_N)\,\mathrm{d}\gamma(x_1,\dots,x_N), \tag{1.1}$$

where (X,d) is a Polish space and $\Gamma(\rho)$ denotes the set of Borel probability measures in X^N having all N marginals equal to a Borel probability measure ρ. We are interested in cost

Date: October 29, 2018.
The authors acknowledge the Academy of Finland projects no. 274372, 284511, 312488, and 314789.

functions of the type

$$c(x_1,\dots,x_N) = \sum_{1\leq i<j\leq N} f(d(x_i,x_j)),$$

where $f\colon [0,+\infty[\to \mathbb{R}\cup\{+\infty\}$ is a continuous, decreasing function, not necessarily bounded from above or below. An interesting example of such a cost is given by minus the logarithm: $f(d(x,y)) = -\log(d(x,y))$.

Our aim is to study properties of the so-called *Kantorovich formulation* of (1.1) for such costs

$$\sup\left\{ N\int_X u\,\mathrm{d}\rho \;\middle|\; u\in L^1_\rho(X), \sum_{i=1}^N u(x_i)\leq c(x_1,\dots,x_N) \text{ for } \rho^{\otimes(N)}\text{-a.e. } (x_1,\dots,x_N)\right\}, \quad (1.2)$$

where $\rho^{\otimes(N)}$ denotes the product of N measures ρ. Optimal Transport problems with logarithmic-type costs were first considered in the literature by W. Wang [30] and W. Gangbo and V. Oliker [15] motivated by the *reflector problem*. In this case, $X = \mathrm{S}^d$, $N=2$ and the authors show the existence of optimal transport plans $\gamma = (\mathrm{Id},T)_\sharp\rho$ in (1.1) concentrated on the graph of a map $T\colon \mathrm{S}^d\to\mathrm{S}^d$. Generally, in the reflector problem, the marginals are not necessarily equal.

In the multi-marginal case, logarithmic-type costs appear in Density Functional Theory (DFT), in the so-called *strictly correlated limit* (SCE). In SCE-DFT, the multi-marginal optimal transport problem is interpreted as the equilibrium configuration of a distribution of N charges in $(x_1,\dots,x_N)\in(\mathbb{R}^d)^N$ subject to the (minus) logarithmic electrostatic interaction depending on the distance between each two of the particles. Due to the indistinguishability of the particles, the charge density $\rho(x_i)$ is the same for all the particles $x_i, i=1,\dots,N$.

Although the interesting case in chemistry is when the system of N electrons are in the physical space $X=\mathbb{R}^3$ subject to a Coulomb electronic-electronic interaction cost, in physics and mathematics 2-body interactions other than the Coulombian one have been considered [12, 13, 28, 14, 7], as well as the problem (1.1) in a lower space dimensions $X=\mathbb{R}^d, d=1,2$ [11, 26, 5, 6, 22]. In particular, when the particles are confined in the plane $\mathbb{R}^2$, the natural model of electrostatic potential between two charges x_i and x_j is given by the logarithmic interaction. We present in subsection 1.2 a pedagogical example of a charged wire, where the logarithmic electrostatic potential appears naturally.

In the following, we give a brief overview on DFT-OT. For a complete presentation on the topic, we refer the reader to [13] and the references therein.

1.1. **A brief review on the literature in DFT-OT.** The problem (1.1) when $X=\mathbb{R}^3$ and c is the Coulomb cost ($f(|x-y|) = 1/|x-y|$) was introduced in 1999 by M. Seidl [27]. By using arguments from physics, Seidl suggested that, at least in the case when ρ is radially symmetric, a minimizer γ in (1.1) exists and is concentrated on the graph of a map $T:\mathbb{R}^3\to\mathbb{R}^3$, $T_\sharp\rho=\rho$, and its iterates, i.e.

$$\gamma = (\mathrm{Id},T,T^{(2)},\dots,T^{(N-1)})_\sharp\rho,$$

where $T^{(N)} = \mathrm{Id}$ and $T^{(i)}$ is the i-times composition of the map T with itself. In particular, via the map T, the optimality condition in the Kantorovich formulation of (1.2) with Coulomb cost reads

$$\nabla u(x) = -\sum_{i=1}^N \frac{x-T^{(i)}(x)}{|x-T^{(i)}(x)|^3}. \quad (1.3)$$

As pointed out in [27] (see also [4]), the constraint in (1.2),

$$\sum_{i=1}^{N} u(x_1) \leq \sum_{1\leq i<j\leq N} \frac{1}{|x_i - x_j|},$$

has a simple physical meaning: it is required that, at optimality, the allowed manifold of the full $3D$ configuration space is the minimum of the classical potential energy given by the Coulomb interaction. Also, the equation (1.3) means that if such an optimal map T exists, the Kantorovich potential $u(x)$ must compensate the net force acting on the electron in x, resulting from the repulsion of the other $N-1$ electrons at positions $T^{(i)}(x)$ [28].

In Density Functional Theory (DFT), the problem (1.1) can be seen as a sort of a semi-classical limit (dilute limit of DFT) of the Hohenberg-Kohn functional[1] [19, 23, 25]. This was suggested in the physics literature by Gori-Giorgi, Seidl and Vignale [17] and, established rigorously by Cotar, Friesecke and Klüppelberg [8, 9]. The same theorem has been proved also by Bindini-De Pascale [1] ($N = 3$ case) and by Lewin [24], the latter allowing also mixed states.

For the Coulomb cost in the 2-marginal case ($N = 2$), the existence of a unique optimal transport plan in (1.1) of type $\gamma = (\mathrm{Id}, T)_\sharp\rho$ ($N = 2$) was obtained, independently, by Cotar, Friesecke and Klüppelberg [8] and by Buttazzo, De Pascale and Gori-Giorgi [4]. In the multi-marginal case ($N > 2$) on the real line ($d = 1$), Colombo, De Pascale and Di Marino [11] proved the existence of optimal transport plans $\gamma = (\mathrm{Id}, T, \dots, T^{(N-1)})_\sharp\rho$ in (1.1) for Coulomb costs. In [12, 13, 28], the repulsive harmonic cost

$$c_w(x_1, \dots, x_N) = -\sum_{1\leq i,j\leq N} |x_i - x_j|^2$$

was studied: Friesecke *et al* [14] have shown the existence of optimal transport plans supported in $(N-1)d$-dimensional sets; in [13] explicit examples of such higher dimensional optimal transport plans as well as an example of an optimal transport plan γ concentrated on the graphs of $\mathrm{Id}, T, \dots, T^{(N-1)}$ for a nowhere continuous map $T\colon [0,1]^d \to [0,1]^d$ are presented. In [16], we gave an example of a three-marginal harmonic repulsion case with absolutely continuous marginals in $\mathbb{R}^n$ for which there is a unique optimal transport plan which is not induced by a map.

1.2. **Logarithmic Eletrostatic potential: Charged wire.** Consider a uniformly charged (infinitely thin) wire on the z-axis:

$$\mathcal{W} := \{\mathbf{x} = (x, y, z) \in \mathbb{R}^3 \ : \ |(x, y)| < \delta\}, \qquad 0 < \delta \ll 1.$$

Suppose that the wire has a charge density $\rho(\mathbf{x})$. The resulting electric field is defined by

$$E(\mathbf{x}) = \frac{1}{4\pi\epsilon_0} \int_{\mathbb{R}^3} \frac{\mathbf{x} - s}{|\mathbf{x} - s|^3} \rho(s)\,\mathrm{d}s,$$

where $\epsilon_0 > 0$ is a constant (permittivity of the free space). Due to *Maxwell's first equation* (or *Gauss' law* of eletrostatics) the scalar field $\rho\colon \mathbb{R}^3 \to \mathbb{R}$ and the vector field $E(\mathbf{x})$ are related by

$$\nabla \cdot E(\mathbf{x}) = \frac{1}{\epsilon_0}\rho(\mathbf{x}).$$

[1] Also known as the Levy-Lieb functional.

We define the total amount of charge Q_Ω in a cylinder $\Omega = \Omega_{R,H} \subset \mathbb{R}^3$ of radius $R > 0$ and height H, which has the wire as its axis of symmetry:

$$Q_\Omega = \int_\Omega \rho(s)\,\mathrm{d}s = \epsilon_0 \int_\Omega \nabla \cdot E(x)\,\mathrm{d}x = \epsilon_0 \oint_{\partial\Omega} E(a) \cdot \mathrm{d}a, \tag{1.4}$$

where the second equality is obtained using the Gauss' theorem. Due to symmetry, the magnitude $|E(\mathbf{x})|$ of the electric field depends only on the Euclidean distance $s = d(\mathbf{x}, \mathbf{w}) = d(\mathbf{x}, \mathcal{W})$ of a point $\mathbf{x}$ from the wire, $|E(\mathbf{x})| = E(s)$, i.e $E(\mathbf{x}) = (E(s)\cos\theta, E(s)\sin\theta, 0)$. Moreover, at each point $\mathbf{w}$ on the lateral surface of this cylinder, the vector $E(\mathbf{w})$ is normal to the surface and has everywhere the same magnitude $|E(\mathbf{w})| = E(R)$.

Therefore, if $\rho(\mathbf{x}) = \overline{\rho} > 0$ is constant inside the cylinder, the flux integral and the total amount of charge in the cylinder $\Omega_{R,H}$ in (1.4) read

$$\frac{1}{\epsilon_0}\overline{\rho} H = (2\pi R)H \cdot E(R), \quad \text{and therefore,} \quad E(R) = \frac{1}{2\pi\epsilon_0}\frac{1}{R}.$$

Let us write $E(s) = 1/(2\pi\epsilon_0 s)$. Since $E(s) = -V'(s)$, the corresponding electrostatic potential $V(s)$ is of logarithmic form

$$V(s) = -\frac{1}{2\pi\epsilon_0}\log\frac{s}{s_0}, \quad s_0 > 0.$$

1.3. **Kantorovich duality.** The duality between the Monge-Kantorovich (1.1) and the Kantorovich problem (1.2) as well as the existence of a maximizer in (1.2) was shown by Kellerer [21] under the assumption that there exist $L^1_\rho(X)$-functions $h_1, \dots, h_N$ and a constant C such that

$$C \le c(x_1, \dots, x_N) \le h(x_1) + \dots + h(x_N).$$

More recently, De Pascale [10] and Buttazzo, Champion and De Pascale [3] extended the duality theory for a class of repulsive cost functions $c\colon \mathbb{R}^{dN} \to \mathbb{R} \cup \{+\infty\}$ which are bounded from below, allowing, for instance, the inclusion of the Coulomb ($s = 1$) and Riesz cost functions ($1 \le s \le d$)

$$c(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|^s}.$$

The main contribution of this paper is to extend the duality theory for logarithmic costs. Some of our proofs are based on arguments present in [3]. One ingredient to tackle the problem of costs that are not bounded from below is to consider, for $R \in]0, \infty[$, the truncated cost functions

$$c_R(x_1, \dots, x_N) := \sum_{1 \le i < j \le N} \max\{f(R), f(d(x_i, x_j))\}, \text{ for all } (x_1, \dots, x_N) \in X^N, \tag{1.5}$$

and related total cost C_R, and collection $\mathcal{F}_R$ of functions for the dual problem:

$$C_R(\gamma) := \int_{X^N} c_R(x_1, \dots, x_N)\,\mathrm{d}\gamma(x_1, \dots, x_N), \text{ for each } \gamma \in \Gamma(\rho),$$

and

$$\mathcal{F}_R := \left\{ u \in L^1_\rho(X) \;\middle|\; u(x_1) + \dots + u(x_N) \le c_R(x_1, \dots, x_N) \text{ for } \rho^{\otimes(N)}\text{-a.e. } (x_1, \dots, x_N) \right\}.$$

In this paper, we will deal with the unbounded costs via the Γ-limit of their truncations.

1.4. **Organization of the paper.** This paper is divided as follows: in Section 2 we present the general setting and introduce briefly some properties of Γ-convergence. In Section 3, we discuss the existence of a minimizer in (1.1) by assuming that the marginals ρ satisfy, with respect to the function f that appears in our cost c, a condition analogous to the common assumption of the marginal measures having finite second moments (see condition (B) in Section 3).

In Section 4, we extend the duality results of [21, 10, 3] for a class of unbounded cost functions (Theorem 4.2) and in Section 5 we obtain regularity results of Kantorovich potentials (Theorem 5.2) as well as continuity of the cost functional as a function of the marginal ρ.

Finally, in Section 6 we give some applications of our results: we note the existence of optimal plans in (1.1), for log-type costs, which are concentrated on maps when $X = \mathbb{R}$, and we prove the existence of an optimal transport map for the logarithmic cost when $N = 2$.

2. Preliminaries

2.1. **General assumptions.** Let (X, d) be a Polish space and $N > 1$ be an integer. We consider a Borel probability measure $\rho \in \mathcal{P}(X)$ having *small concentration*[2], meaning

$$\lim_{r\to 0} \sup_{x\in X} \rho(B(x,r)) < \frac{1}{N(N-1)^2}. \tag{A}$$

We denote by $(x_1, \dots, x_N)$ points in X^N, so $x_i \in X$ for each i. If we do not otherwise specify, each quantification with respect to i or i, j is from 1 to N. For a fixed $N \geq 1$, we assume that the cost $c\colon X^N \to \mathbb{R} \cup \{+\infty\}$ is of the form

$$c(x_1, \dots, x_N) = \sum_{1\leq i<j\leq N} f(d(x_i, x_j)), \qquad \text{for all } (x_1, \dots, x_N) \in X^N, \tag{2.1}$$

where $f\colon [0, \infty[\to \mathbb{R} \cup \{+\infty\}$ satisfies the following conditions

$$f|_{[0,\infty[} \text{ is continuous and decreasing, and} \tag{F1}$$

$$\lim_{t\to 0+} f(t) = +\infty. \tag{F2}$$

Let us denote for a fixed $R > 0$, for all $t > 0$

$$f_R(t) = \begin{cases} f(t) & \text{if } t < R \\ f(R) & \text{otherwise} \end{cases} \quad \text{and}$$

$$f_R^{-1}(t) = \inf\{s \mid f_R(s) = t\};$$

of course, if f is not strictly decreasing, the inverse function f^{-1} is not well defined, but still the *left-inverse* of f can be defined as above.

We denote the set of couplings or transport plans having N marginals equal to ρ by

$$\Gamma(\rho) = \left\{\gamma \in \mathcal{P}(X^N) \;\middle|\; \mathrm{pr}^i_\sharp\gamma = \rho \text{ for all } i\right\},$$

where pr^i is the projection on the i-th coordinate

$$\mathrm{pr}^i(x_1, \dots, x_i, \dots, x_N) = x_i, \quad \text{for all } (x_1, \dots, x_i, \dots, x_N) \in X^N.$$

[2] The upper bound $\frac{1}{N(N-1)^2}$ in (A) has been weakened to $\frac{1}{N}$ in a work in progress by Colombo, Di Marino and Stra. However, we will assume here the previously used upper bound.

In addition, we set for each $\gamma \in \Gamma(\rho)$,

$$C(\gamma) = \int_{X^N} c(x_1, \dots, x_N) \, \mathrm{d}\gamma(x_1, \dots, x_N);$$

this is the transportation cost related to γ.

We want to study the dual problem, so we set

$$\mathcal{F} := \left\{ u \in L^1_\rho(X) \;\middle|\; u(x_1) + \cdots + u(x_N) \le c(x_1, \dots, x_N) \text{ for } \rho^{\otimes(N)}\text{-a.e. } (x_1, \dots, x_N) \right\}$$

and

$$D\colon L^1_\rho(X) \to \mathbb{R}, \quad D(u) = N \int_X u \, \mathrm{d}\rho \text{ for all } u \in L^1_\rho(X).$$

Here one should note that, in the definition of $\mathcal{F}$ and also in future considerations, we identify the elements of $\mathcal{F}$ with a representative given by the following Lemma 2.1 (unless otherwise stated). The lemma also implies that for every $u \in \mathcal{F}$ there is a set $E \subset X$ of ρ-measure zero and a real-valued representative $\tilde{u}$ of u so that for all $x_1, \dots, x_N \in X \setminus E$ the constraint (2.2) below holds.

Lemma 2.1. *Let $u \in \mathcal{F}$. Then there exists a representative $\tilde{u}$ of u (possibly attaining value $-\infty$ in a set of measure zero) such that*

$$\tilde{u}(x_1) + \cdots + \tilde{u}(x_N) \le c(x_1, \dots, x_N) \tag{2.2}$$

holds for all $x_1, \dots, x_N \in X$.

Proof. Let us start by fixing any representative of u (still denoting it by u). First of all, for $x \in X \setminus \operatorname{spt}(\rho)$ we may define $\tilde{u}(x) = -\infty$. Similarly, if our procedure below would give the value $+\infty$ for $\tilde{u}$ at some point, we may freely change the value at that point to 0.

We also note that, if we were in a setting where the standard Lebesgue differentiation theorem holds, we could define

$$\tilde{u}(x) = \liminf_{r \to 0} \frac{1}{\rho(B(x,r))} \int_{B(x,r)} u(y) \, \mathrm{d}\rho(y).$$

However, since the standard Lebesgue differentiation theorem using balls might fail in our setting, we take a more general version using Vitali coverings, see [18] for the differentiation result, and for instance [20] for a generalized nested cubes construction (that can be made in any separable metric space with a trivial modification) which serves as the Vitali covering. We only recall here the properties needed for our proof: There exists a collection $\{Q_{k_i} : k \in \mathbb{N}, i \in N_k \subset \mathbb{N}\}$ of Borel subsets of $\operatorname{spt}(\rho)$ so that for each k the subcollection $\{Q_{k_i} : i \in N_k\}$ is a partition of $\operatorname{spt}(\rho)$ with $\operatorname{diam}(Q_{k_i}) < 2^{-k}$ and $\rho(Q_{k_i}) > 0$ satisfying $\tilde{u}(x) = u(x)$ for ρ-almost every $x \in \operatorname{spt}(\rho)$, where we have defined for every $x \in \operatorname{spt}(\rho)$

$$\tilde{u}(x) := \liminf_{k \to \infty} \frac{1}{\rho(Q_k(x))} \int_{Q_k(x)} u(y) \, \mathrm{d}\rho(y)$$

with $Q_k(x)$ defined as the unique $Q_{k,i}$ in the level k partition for which $x \in Q_{k,i}$.

Now, for every $x_1, \dots, x_N \in \mathrm{spt}(\rho)$ we get by using the $\rho^{\otimes N}$-almost everywhere inequality and the continuity of the cost c,

$$\begin{aligned}
\sum_{i=1}^N \tilde{u}(x_i) &= \sum_{i=1}^N \liminf_{k\to\infty} \frac{1}{\rho(Q_k(x_i))} \int_{Q_k(x_i)} u(y)\,\mathrm{d}\rho(y) \\
&= \sum_{i=1}^N \liminf_{k\to\infty} \frac{1}{\rho^{\otimes N}(\prod_{j=1}^N Q_k(x_j))} \int_{\prod_{j=1}^N Q_k(x_j)} u(y_i)\,\mathrm{d}\rho^{\otimes N}(y_1,\dots,y_N) \\
&\le \liminf_{k\to\infty} \frac{1}{\rho^{\otimes N}(\prod_{j=1}^N Q_k(x_j))} \int_{\prod_{j=1}^N Q_k(x_j)} \sum_{i=1}^N u(y_i)\,\mathrm{d}\rho^{\otimes N}(y_1,\dots,y_N) \\
&\le \liminf_{k\to\infty} \frac{1}{\rho^{\otimes N}(\prod_{j=1}^N Q_k(x_j))} \int_{\prod_{j=1}^N Q_k(x_j)} c(y_1,\dots,y_N)\,\mathrm{d}\rho^{\otimes N}(y_1,\dots,y_N) \\
&= c(x_1,\dots,x_N).
\end{aligned}$$

This concludes the proof. ☐

We aim at showing that

$$\min_{\gamma\in\Gamma(\rho)} C(\gamma) = \max_{u\in\mathcal{F}} D(u). \tag{2.3}$$

In order to guarantee the existence of a minimizer on the left-hand side of (2.3), we also assume that there exist a point $o \in X$ and a radius $r_0 > 0$ such that

$$\int_{X\setminus B(o,r_0)} f\left(2d(x,o)\right)\,\mathrm{d}\rho(x) > -\infty. \tag{B}$$

This is a similar assumption than requiring, in the case of quadratic cost, that the marginal measures have finite second moments.

Notice that even when $X = \mathbb{R}^d$ the cost function c in (2.1) does not fall in the class of functions considered by Buttazzo, Champion and de Pascale [3], since it may not be bounded from below. However, by suitably truncating the cost c, the truncated functions c_R are bounded from below for each R and, modulo translation, fall into the category of functions considered in [3].

2.2. **Γ-convergence.** We briefly outline the relevant definitions and properties of Γ and Γ^+-convergences. The former is a type of convergence of functionals adjusted to minimal value problems and the latter to maximal value problems. For a thorough presentation of Γ-convergence, we refer the reader to Braides' book [2].

Definition 2.2 (Γ-convergence and Γ^+-convergence)**.** *Let (S,d) be a metric space. We say that a sequence $(F_n)_{n\in\mathbb{N}}$ of functions $F_n\colon S\to\overline{\mathbb{R}}$ Γ-converges to a function $F\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$ and denote $F_n \xrightarrow{\Gamma} F$ if for all $y\in S$ the following two conditions hold:*

For all sequences $(y_n)_{n\in\mathbb{N}}$ that converge to y we have

$$\liminf_n F_n(y_n) \ge F(y), \text{ and} \tag{I}$$

there exists a sequence $(y_n)_{n\in\mathbb{N}}$ converging to y such that

$$\limsup_n F_n(y_n) \le F(y). \tag{II}$$

Correspondingly, we say that a sequence $(D_n)_{n\in\mathbb{N}}$ *of functions* $D_n\colon S\to\overline{\mathbb{R}}$, Γ^+*-convergence to a function* $D\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$ *and denote* $D_n\overset{\Gamma^+}{\to}D$ *if for all* $u\in S$ *the following two conditions hold:*

$$\begin{aligned}&\textit{For any sequence } (u_n)_{n\in\mathbb{N}} \textit{ converging to } u \textit{ we have}\\ &\limsup_n D_n(u_n)\le D(u), \textit{ and}\end{aligned}\tag{I+}$$

$$\begin{aligned}&\textit{there exists a sequence } (u_n)_{n\in\mathbb{N}} \textit{ converging to } u \textit{ such that}\\ &\limsup_n D_n(u_n)\le D(u).\end{aligned}\tag{II+}$$

In order to be able to take advantage of these notions, the sub-levels of the functionals must satisfy some compactness properties. The following definition takes care of this.

Definition 2.3. *Let* (S,d) *be a metric space. We say that a sequence* $(F_n)_{n\in\mathbb{N}}$ *of functions* $F_n\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$ *is equi-mildly coercive on* S *if there exists a compact and non-empty subset* K *of* S *such that for all* $n\in\mathbb{N}$ *we have*

$$\inf_{y\in S}F_n(x)=\inf_{y\in K}F_n(y).$$

Analogously, we say that a sequence $(D_n)_{n\in\mathbb{N}}$ *of functions* $D_n\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$ *is equi-mildly* $^+$*-coercive on* S *if there exists a compact and non-empty subset* K *of* S *such that for all* $n\in\mathbb{N}$ *we have*

$$\sup_{u\in S}D_n(u)=\sup_{u\in K}D_n(u).$$

Theorem 2.4. [2, Theorem 1.21] *Let* (S,d) *be a metric space. Let* $(F_n)_{n\in\mathbb{N}}$ *be an equi-mildly coercive sequence of functions* $F_n\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$ *that* Γ*-converges to some function* $F\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$*. Then there exists a minimum* $y\in S$ *of* F *and the sequence* $(\inf_{y\in S}F_n(y))_{n\in\mathbb{N}}$ *converges to* $\min_{y\in S}F(y)$*. In addition, if* $(y_n)_{n\in\mathbb{N}}$ *is a sequence of elements of* S *such that*

$$\lim_n F_n(y_n)=\lim_n\inf_{y\in S}F_n(y),$$

then every limit of a subsequence of $(y_n)_{n\in\mathbb{N}}$ *is a minimizer of* F*.*

Similarly, let $(D_n)_{n\in\mathbb{N}}$ *be an equi-mildly* Γ^+*-coercive sequence of functions* $D_n\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$ *that* Γ^+*-converges to some function* $D\colon S\to\mathbb{R}\cup\{-\infty,+\infty\}$*. Then there exists a maximum* $u\in S$ *of* D *and the sequence* $(\sup_{u\in S}D_n(u))_n$ *converges to* $\max_{u\in S}D(u)$*. In addition, if* $(u_n)_{n\in\mathbb{N}}$ *is a sequence of elements of* S *such that*

$$\lim_n D_n(u_n)=\lim_n\sup_{u\in S}D_n(u),$$

then every limit of a subsequence of $(u_n)_{n\in\mathbb{N}}$ *is a maximizer of* D*.*

3. Monge-Kantorovich problem

First, we prove the existence of a minimizer for the Monge-Kantorovich problem (1.1) in our framework. Notice that the conditions (A) and (B) guarantee that the cost has a finite value.

Proposition 3.1. *Let (X, d) be a Polish space. Suppose that $\rho \in \mathcal{P}(X)$ satisfies (A) and (B), and $c\colon X^N \to \mathbb{R} \cup \{+\infty\}$ is a cost function*

$$c(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} f(d(x_i, x_j)), \quad \textit{for all } (x_1, \dots, x_N) \in X^N,$$

where $f\colon [0, \infty[\to \mathbb{R}$ satisfies $(F1)$ and $(F2)$. Then, the following minimum is achieved

$$\min_{\gamma \in \Gamma(\rho)} \int_{X^N} c(x_1, \dots, x_N) \, \mathrm{d}\gamma(x_1, \dots, x_N).$$

Proof. The proof follows standard arguments. From Prokhorov theorem we know that $\Gamma(\rho)$ is weakly compact. Therefore, it suffices to prove the lower semicontinuity of the cost $C(\gamma)$. For this, it suffices (see [29, Lemma 4.3]) to find an upper semicontinuous function h such that

$$h \in L^1_\gamma(X^N) \text{ for all } \gamma \in \Gamma(\rho), \tag{3.1}$$

$$c \ge h, \text{ and} \tag{3.2}$$

$$\int_{X^N} h \, \mathrm{d}\gamma' = \int_{X^N} h \, \mathrm{d}\gamma \text{ for all } \gamma, \gamma' \in \Gamma(\rho). \tag{3.3}$$

Using the point $o \in X$ and radius $r_0 > 0$ from the condition (B), we define $g\colon [0, \infty[\to \mathbb{R}$ by

$$g(r) = \begin{cases} \min(0, f(2r_0)) & \text{if } r < 2r_0 \\ \min(0, f(r)) & \text{if } r \ge 2r_0 \end{cases},$$

and set $h\colon X^N \to \mathbb{R}$

$$h(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} (g(2d(x_i, o)) + g(2d(x_j, o))).$$

As a finite sum of continuous functions, h is continuous and thus trivially upper semicontinuous. In addition, for any $\gamma \in \Gamma(\rho)$ we have by Assumption (B)

$$\begin{aligned}
\int_{X^N} h \, \mathrm{d}\gamma &= \sum_{1 \le i < j \le N} \int_{X^N} (g(2d(x_i, o)) + g(2d(x_j, o))) \, \mathrm{d}\gamma \\
&= N(N-1) \int_X g(2d(x, o)) \, \mathrm{d}\rho(x) \\
&\ge N(N-1) \left(\int_{B(o, r_0)} \min(0, f(2r_0)) \, \mathrm{d}\rho(x) + \int_{X \setminus B(o, r_0)} \min(0, f(2d(x, o))) \, \mathrm{d}\rho(x) \right) \\
&> -\infty.
\end{aligned}$$

Since h is also nonpositive, condition (3.1) holds. Similarly, condition (3.3) follows by

$$\int_{X^N} h \, \mathrm{d}\gamma' = N(N-1) \int_X g(2d(x, o)) \, \mathrm{d}\rho(x) = \int_{X^N} h \, \mathrm{d}\gamma.$$

Finally, to prove condition (3.2), we fix $(x_1,\dots,x_N)\in X^N$. By (F1) and the nonpositivity of g we have that

$$\begin{aligned} c(x_1,\dots,x_N) &= \sum_{1\le i<j\le N} f(d(x_i,x_j)) \ge \sum_{1\le i<j\le N} g(d(x_i,x_j)) \\ &\ge \sum_{1\le i<j\le N} g(d(x_i,o)+d(x_j,o)) \\ &\ge \sum_{1\le i<j\le N} g(2\max\{d(x_i,o),d(x_j,o)\}) \\ &\ge \sum_{1\le i<j\le N} (g(2d(x_i,o))+g(2d(x_j,o))) = h(x_1,\dots,x_N). \end{aligned}$$

This concludes the proof. □

For $\alpha>0$ we define the set D_α as

$$D_\alpha := \left\{(x_1,\dots,x_N)\in X^N \mid \text{there exist } i,j \text{ such that } d(x_i,x_j)<\alpha\right\}.$$

The next theorem from [3] states that under the previous hypotheses there exists $\overline{\alpha}>0$ for which the support of any optimal plan is concentrated away from the set $D_{\overline{\alpha}}$.

Theorem 3.2. *Let (X,d), ρ, f, c as in the Proposition 3.1 and let γ be a minimizer of*

$$C(\rho)=\min_{\gamma\in\Gamma(\rho)}\int_{X^N} c(x_1,\dots,x_N)\,\mathrm{d}\gamma(x_1,\dots,x_N).$$

Let us fix $0<\beta<1$ such that

$$\sup_{x\in X}\rho(B(x,\beta))<\frac{1}{N(N-1)^2}.$$

Then, we have for all

$$\alpha < f^{-1}\left(\frac{N^2(N-1)}{2}f(\beta)\right) \tag{3.4}$$

the inclusion

$$\mathrm{spt}(\gamma)\subset X^N\setminus D_\alpha. \tag{3.5}$$

Proof. The proof presented in [3, Theorem 2.4] also works here. The fact that optimal plans stay out of the diagonal reflect the properties of the cost close to the singularity, not to the tail. □

We recall that for all $R>0$, the truncated costs c_R and C_R

$$c_R(x_1,\dots,x_N)=\sum_{1\le i<j\le N}\max\{f(R),f(d(x_i,x_j))\} \text{ for all } (x_1,\dots,x_N)\in X^N,$$

$$C_R(\gamma)=\int_{X^N} c_R(x_1,\dots,x_N)\,\mathrm{d}\gamma(x_1,\dots,x_N) \text{ for each } \gamma\in\Gamma(\rho).$$

Using these we define the functionals $K_R, K\colon \mathcal{P}(X^N)\to\mathbb{R}\cup\{+\infty\}$,

$$K_R(\gamma) := \begin{cases} C_R(\gamma) & \text{if } \gamma\in\Gamma(\rho) \\ +\infty & \text{otherwise} \end{cases},$$

$$K(\gamma) := \begin{cases} C(\gamma) & \text{if } \gamma \in \Gamma(\rho) \\ +\infty & \text{otherwise} \end{cases}.$$

An approximation result of convergence of minimizers of the truncated costs $(K_R)_{R\in\mathbb{N}}$ is given by the following proposition.

Proposition 3.3. *The sequence of functionals $(K_R)_{R\in\mathbb{N}}$ is equicoercive and Γ-converges to K with respect to the weak convergence of measures.*

Proof. First we notice that the equicoerciviness of $(K_R)_{R\in\mathbb{N}}$ follows from the fact that $\Gamma(\rho)$ is weakly compact [21]. We then fix $\gamma \in \mathcal{P}(X^N)$ and show that

$$\begin{aligned} &\text{for all sequences } (\gamma_R)_{R\in\mathbb{N}} \text{ such that } \gamma_R \rightharpoonup \gamma \text{ we have} \\ &\liminf_{R\to\infty} K_R(\gamma_R) \geq K(\gamma), \text{ and} \end{aligned} \tag{3.6}$$

$$\begin{aligned} &\text{there exists a sequence } (\gamma_R)_{R\in\mathbb{N}} \text{ such that } \gamma_R \rightharpoonup \gamma \text{ and} \\ &\limsup_{R\to\infty} K_R(\gamma_R) \leq K(\gamma). \end{aligned} \tag{3.7}$$

Fix a sequence $(\gamma_R)_{R\in\mathbb{N}}$ in $\mathcal{P}(X^N)$ such that $\gamma_R \rightharpoonup \gamma$. By going to a subsequence we may assume that $\liminf_{R\to\infty} K_R(\gamma_R) = \lim_{R\to\infty} K_R(\gamma_R)$. Thus, we may also suppose that $K_R(\gamma_R) < \infty$ for all $R \in \mathbb{N}$, since otherwise (3.6) would trivially hold. Consequently, we have that $\gamma_R \in \Gamma(\rho)$ for all $R \in \mathbb{N}$ and thus also $\gamma \in \Gamma(\rho)$ by compactness of $\Gamma(\rho)$, see [21]. Now, by monotonicity of the integral and lower semi-continuity of $K(\gamma)$ we get

$$\liminf_{R\to\infty} K_R(\gamma_R) \geq \liminf_{R\to\infty} K(\gamma_R) \geq K(\gamma),$$

so (3.6) is satisfied. Finally, the condition (3.7) is satisfied by the constant sequence $\gamma_R = \gamma$ for all $R \in \mathbb{N}$. □

3.1. **Symmetric probability measures.** We remark that the Monge-Kantorovich problem (1.1) can be restricted to symmetric transport plans.

Definition 3.4 (Symmetric measures)**.** *A measure $\gamma \in \mathcal{P}(X^N)$ is symmetric if*

$$\int_{X^N} \phi(x_1,\dots,x_N)\,\mathrm{d}\gamma = \int_{X^N} \phi(\overline{\sigma}(x_1,\dots,x_N))\,\mathrm{d}\gamma, \ \textit{for all } \phi \in \mathcal{C}(X^N)$$

and for all permutations $\overline{\sigma}$ of N symbols. We denote by $\Gamma^{sym}(\rho)$, the space of all $\gamma \in \Gamma(\rho)$ which are symmetric.

Proposition 3.5. *Let (X,d) be a Polish space. Suppose $\rho \in \mathcal{P}(X)$ such that (A) and (B) hold and $c\colon X^N \to \mathbb{R}\cup\{+\infty\}$ is a continuous cost function. Then,*

$$\min_{\gamma\in\Gamma(\rho)} \int_{X^N} c(x_1,\dots,x_N)\,\mathrm{d}\gamma = \min_{\gamma\in\Gamma^{sym}(\rho)} \int_{X^N} c(x_1,\dots,x_N)\,\mathrm{d}\gamma. \tag{3.8}$$

Proof. The minimum on the left-hand side in (3.8) is surely smaller than or equal to the minimum on the right-hand side, since $\Gamma^{sym}(\rho) \subset \Gamma(\rho)$. Suppose $\gamma \in \Gamma(\rho)$, we can define a symmetric plan

$$\gamma_{sym} = \frac{1}{N!}\sum_{\sigma} \sigma_\sharp \gamma,$$

where the sum is taken over all permutations σ of N-symbols. Thanks to the linearity of the cost function $C(\gamma)$, γ_{sym} and γ have the same cost and, therefore, (3.8) holds. □

4. Duality Theory for log-type cost functions

Let us start by recalling a version of Kellerer's theorem when the cost is bounded from below.

Theorem 4.1. *Let (X, d) be a Polish space. Suppose $\rho \in \mathcal{P}(X)$ such that (A) and (B) hold and $c\colon X^N \to \mathbb{R} \cup \{+\infty\}$ is a cost function*

$$c(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} f(d(x_i, x_j)), \qquad \text{for all } (x_1, \dots, x_N) \in X^N,$$

where $f\colon [0, +\infty[\to \mathbb{R} \cup \{+\infty\}$ is a function satisfying (F1) and (F2) and f is bounded from below. Then for any $\gamma \in \Gamma(\rho)$ minimizing $\int_{X^N} c \, \mathrm{d}\gamma$ there exists a Kantorovich potential $u \in \mathcal{F}$ such that the following hold:

$$\int_{X^N} c \, \mathrm{d}\gamma = N \int_X u(x) \, \mathrm{d}\rho(x),$$

for ρ-almost every $x_1 \in X$ we have

$$u(x_1) = \inf \left\{ c(x_1, x_2, \dots, x_N) - \sum_{j=2}^{N} u(x_j) \;\middle|\; (x_2, \dots, x_N) \in X^{N-1} \right\}, \tag{4.1}$$

and for γ-almost every $(x_1, \dots, x_N) \in X^N$ we have

$$\sum_{j=1}^{N} u(x_j) = c(x_1, x_2, \dots, x_N). \tag{4.2}$$

Proof. As for the proof of Theorem 3.2 we refer to [3, Theorem 2.4], where it was shown that not only is

$$\operatorname{spt}(\gamma) \subset X^N \setminus D_\alpha$$

for $\alpha > 0$ small enough, but also the same holds if we consider the cost

$$c^{f(\alpha)}(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} \min(f(d(x_i, x_j)), f(\alpha)).$$

Consequently, γ is also a minimizer for the cost $c^{f(\alpha)}$. Therefore, we can invoke Kellerer's Theorem [21] for the bounded cost $c^{f(\alpha)}$. According to it, there exists a Kantorovich potential $u \in L^1_\rho(X)$ solving the dual problem for $c^{f(\alpha)}$ so that for any $x_1 \in X$

$$u(x_1) = \inf \left\{ c^{f(\alpha)}(x_1, x_2, \dots, x_N) - \sum_{j=2}^{N} u(x_j) \;\middle|\; (x_2, \dots, x_N) \in X^{N-1} \right\}$$

and for γ-almost every $(x_1, \dots, x_N) \in X^N$

$$\sum_{j=1}^{N} u(x_j) = c^{f(\alpha)}(x_1, x_2, \dots, x_N).$$

Now, observe that since $c^{f(\alpha)} \le c$, we have $u \in \mathcal{F}$. Also, since for γ-almost every $(x_1, \dots, x_N) \in X$ we have

$$c^{f(\alpha)}(x_1, x_2, \dots, x_N) = c(x_1, x_2, \dots, x_N),$$

the equation (4.2), and consequently also (4.1), hold. ☐

The following theorem then extends Kantorovich duality for our larger class of cost functions.

Theorem 4.2. *Let (X,d) be a Polish space. Suppose $\rho \in \mathcal{P}(X)$ such that (A) and (B) hold and $c\colon X^N \to \mathbb{R}\cup\{+\infty\}$ is a cost function*

$$c(x_1,\dots,x_N) = \sum_{1\le i<j\le N} f(d(x_i,x_j)), \quad \textit{for all } (x_1,\dots,x_N)\in X^N,$$

where $f\colon [0,+\infty[\to \mathbb{R}\cup\{+\infty\}$ is a function satisfying $(F1)$ and $(F2)$. Then, the duality holds:

$$\min_{\gamma\in\Gamma(\rho)} \int_{X^N} c\,\mathrm{d}\gamma = \max_{u\in L^1_\rho(X)} \left\{ N\int_X u(x)\,\mathrm{d}\rho(x) \ :\ \sum_{i=1}^N u(x_i) \le c(x_1,\dots,x_N)\ \rho^{\otimes(N)}\text{-}a.e.\right\}. \tag{4.3}$$

Proof. Due to Proposition 3.1 the minimum on the left-hand side is realized. By using the monotonicity of integral and the fact that $\gamma\in\Gamma(\rho)$, we easily get

$$\min_{\gamma\in\Gamma(\rho)} C(\gamma) \ge \sup_{u\in\mathcal{F}} D(u).$$

Hence, we need to show that

$$\min_{\gamma\in\Gamma(\gamma)} C(\gamma) \le \sup_{u\in\mathcal{F}} D(u) \tag{4.4}$$

and that a maximizer for $\max_{u\in\mathcal{F}} D(u)$ exists.

Towards this goal, let us fix a minimizer $\gamma\in\Gamma^{sym}(\rho)$ of C. It now suffices to show that there exists a function $u\in\mathcal{F}$ such that

$$C(\gamma)\le D(u)\,.$$

For each $L>0$, let us denote $\gamma_L = \gamma|_{B(o,L)^N}$, where o is the point in the condition (B). Let us further denote $\gamma_L^P = \frac{1}{\gamma_L(X^N)}\gamma_L$. Notice that $\gamma_L\ne 0$ for large enough $L>0$. Let us denote the marginals of γ_L^P by ρ_L.

Now, γ_L^P is optimal also for all C_R with $R\ge 2L$, since $C=C_R$ for all couplings of ρ_L. Let (u_R) be a sequence of Kantorovich potentials given by Theorem 4.1, each corresponding to $\gamma_{R/2}^P$ with the cost c_R and the marginals $\rho_{R/2}$. Let us fix $R_0>0$ such that $\gamma_{R_0/2}\ne 0$, and a point $(\overline{x}_1,\dots,\overline{x}_N)\in \operatorname{spt}(\gamma_{R_0/2})$.

We may then assume that for all $R\ge R_0$, we have

$$u_R(\overline{x}_i) = \frac{1}{N}c_R(\overline{x}_1,\dots,\overline{x}_N) = \frac{1}{N}c(\overline{x}_1,\dots,\overline{x}_N) \text{ for all } i,$$

since $(\overline{x}_1,\dots,\overline{x}_N)\in\operatorname{spt}(\gamma_{R_0/2})\subset\operatorname{spt}(\gamma_{R/2})$.

Now we have, for all $R\ge R_0$ and for $\rho_{R/2}$-almost every $x_1\in X$, by (4.1), for some $\alpha>0$ coming from Theorem 3.2 which depends on γ, but not on R, the estimate

$$\begin{aligned} u_R(x_1) &\le c_R(x_1,\overline{x}_2,\dots,\overline{x}_N) - \frac{N-1}{N}c(\overline{x}_1,\dots,\overline{x}_N)\\ &\le \frac{N(N-1)}{2}f(\frac{\alpha}{2}) - \frac{N-1}{N}c(\overline{x}_1,\dots,\overline{x}_N) =: M, \end{aligned}$$

since by the fact that $(\overline{x}_1,\dots,\overline{x}_N)\in X^N\setminus D_\alpha$, we may assume (by changing $\overline{x}_1$ with some other $\overline{x}_i$), that $d(x_1,\overline{x}_j)\ge\frac{\alpha}{2}$ for all $j\in\{2,\dots,N\}$.

For the lower bound, we use (4.2) for u_R, since it was taken to be the potential given by Theorem 4.1, together with the upper bound that we just obtained. For γ_L^P-almost every $(x_1, \dots, x_N) \in X^N$, and thus ρ_L-almost every $x_1 \in X$, when $R \geq 2L$, we have

$$\begin{aligned} u_R(x_1) &= \sum_{1 \leq i < j \leq N} f(d(x_i, x_j)) - \sum_{j=2}^{N} u_R(x_j) \\ &\geq \frac{N(N-1)}{2} f(2L) - (N-1)M. \end{aligned}$$

Let us now consider a sequence $(\bar{u}_R)$ built from (u_R) by setting $\bar{u}_R := \chi_{A_R} u_R$, where

$$A_R := \left\{ x \in X \, : \, \frac{\mathrm{d}\rho_R}{\mathrm{d}\rho} \geq \frac{1}{2\gamma_R(X^N)} \right\}.$$

We aim at showing that $(\bar{u}_R)$ has a weakly converging subsequence in $L^1_\rho(X)$. This follows from Dunford-Pettis theorem if we show that $(\bar{u}_R)$ is equibounded and equi-integrale in $L^1_\rho(X)$.

Let us first prove the equiboundedness. Using the definition of A_R and $\bar{u}_R$, the upper bound M for u_R, and the fact that u_R is a Kantorovich potential, we get

$$\begin{aligned} \int_X |\bar{u}_R(x)| \, \mathrm{d}\rho(x) &\leq 2\gamma_R(X^N) \int_X |\bar{u}_R(x)| \, \mathrm{d}\rho_R(x) \leq 2\gamma_R(X^N) \int_X |u_R(x)| \, \mathrm{d}\rho_R(x) \\ &\leq 2\gamma_R(X^N) M + 2\gamma_R(X^N) \left| \int_X u_R(x) \, \mathrm{d}\rho_R(x) \right| \\ &= 2\gamma_R(X^N) M + \frac{2\gamma_R(X^N)}{N} |C(\gamma_R^P)|. \end{aligned}$$

Since $C(\gamma_R^P) \to C(\gamma)$ as $R \to \infty$, we see that $\bar{u}_R$ is equibounded in $L^1_\rho(X)$.

Let us then show the equi-integrability of $\bar{u}_R$. For this, let

$$B_L := \left\{ x \in X \, : \, \frac{\mathrm{d}\rho_L}{\mathrm{d}\rho} > 0 \right\}.$$

Since $R \geq 2L$ the functions $\bar{u}_R$ are uniformly bounded on B_L, and since for the finitely many $R \in \{R_0, \dots, 2L-1\}$ we may rely on the absolute continuity of the integral, we only need a uniform estimate outside B_L. We start by similar estimates as before to obtain

$$\begin{aligned} \int_{X \setminus B_L} |\bar{u}_R(x)| \, \mathrm{d}\rho(x) &\leq 2\gamma_R(X^N) \int_{X \setminus B_L} |u_R(x)| \, \mathrm{d}\rho_R(x) \\ &\leq 2\gamma_R(X^N) \int_{X^N \setminus B(o,L)^N} |u_R(x_1)| \, \mathrm{d}\gamma_R^P(x) \\ &\leq 2\gamma_R(X^N) M \gamma_R^P(X^N \setminus B(o,L)^N) \\ &\quad + 2\gamma_R(X^N) \left| \int_{X^N \setminus B(o,L)^N} u_R(x_1) \, \mathrm{d}\gamma_R^P(x) \right|. \end{aligned}$$

Since $\sup_{R>0} \gamma_R^P(X^N \setminus B(o,L)^N) \to 0$ as $L \to \infty$, we only need to control the second term in the sum above. To control it we use (4.2) to obtain, for $R \geq 2L$,

$$\begin{aligned}\left|\int_{X^N \setminus B(o,L)^N} u_R(x_1)\,\mathrm{d}\gamma_R^P(x)\right| &\leq \left|\int_{X^N} u_R(x_1)\,\mathrm{d}\gamma_R^P(x) - \int_{B(o,L)^N} u_R(x_1)\,\mathrm{d}\gamma_R^P(x)\right| \\ &= \frac{1}{N}\left|C(\gamma_R^P) - \frac{\gamma_R(X^N)}{\gamma_L(X^N)} C(\gamma_L^P)\right| \to 0\end{aligned}$$

as $L \to \infty$. Thus the equi-integrability of $\bar{u}_R$ with respect to ρ is established.

By the Dunford-Pettis theorem there exists a weak limit u in $L^1_\rho(X)$ of $(\bar{u}_R)$ along some subsequence. What still needs to be shown is that $u \in \mathcal{F}$ and $C(\gamma) = D(u)$. Let us start by showing that $u \in \mathcal{F}$. We already know that $u \in L^1_\rho(X)$. Towards showing the validity of the constraint, we assume for the contrary that there exists a Borel set $A \subseteq X^N$ such that $\rho^{\otimes(N)}(A) > 0$ and

$$u(x_1) + \cdots + u(x_N) > c(x_1, \ldots, x_N) \text{ for all } (x_1, \ldots, x_N) \in A. \tag{4.5}$$

Notice that A_L is an increasing sequence of sets and $\rho(A_L) \to 1$ as $L \to \infty$. Therefore, by going into a subset of A if necessary, we may assume that $A \subset (A_L)^N$ for some $L > 0$.

Now, by Mazur's lemma, there is a sequence $(\tilde{u}_R)$ of convex combinations of $(\bar{u}_R)_{R\geq 2L}$ strongly converging to u in $L^1_\rho(X)$. Since $c_R = c$ on A for all $R \geq 2L$, we have

$$\tilde{u}_R(x_1) + \cdots + \tilde{u}_R(x_N) \leq c(x_1, \ldots, x_N) \text{ for } \rho\text{-almost all } (x_1, \ldots, x_N) \in A, \tag{4.6}$$

for all $R \geq 2L$, as the inequality is preserved under convex combinations.

Let us denote

$$l := \int_A (u(x_1) + \cdots + u(x_N) - c(x_1, \ldots, x_N))\,\mathrm{d}\rho^{\otimes(N)}.$$

Due to (4.5) we have $l > 0$. Because $\tilde{u}_R \to u$ strongly, there exists $R_1 \geq 2L$ such that

$$\int_A \sum_{i=1}^N |\tilde{u}_R(x_i) - u(x_i)|\,\mathrm{d}\rho^{\otimes(N)} < \frac{l}{2} \text{ for all } R \geq R_1. \tag{4.7}$$

Then we have for all $R > R_1$

$$\begin{aligned}&\int_A \left(\sum_{i=1}^N \tilde{u}_R(x_i) - c(x_1, \ldots, x_N)\right)\mathrm{d}\rho^{\otimes(N)} \\ &= \int_A \sum_{i=1}^N (\tilde{u}_R(x_i) - u(x_i))\,\mathrm{d}\rho^{\otimes(N)} + \int_A \sum_{i=1}^N u(x_i) - c(x_1, \ldots, x_N)\,\mathrm{d}\rho^{\otimes(N)} \\ &> l - \frac{l}{2} = \frac{l}{2} > 0,\end{aligned}$$

contradicting (4.6).

Finally, we show that $C(\gamma) = D(u)$. We get this by using $\rho(A_R) \to 1$ as $R \to \infty$, the definition of γ_R^P, the equality (4.2), and the $L^1_\rho(X)$-convergence:

$$\begin{aligned} C(\gamma) &= \lim_{R\to\infty} \int_{A_R^N} c \,\mathrm{d}\gamma = \lim_{R\to\infty} \int_{A_R^N} c_R \,\mathrm{d}\gamma_R^P \\ &= \lim_{R\to\infty} N \int_X \bar{u}_R(x) \,\mathrm{d}\rho_R(x) = \lim_{R\to\infty} N \int_X \bar{u}_R(x) \,\mathrm{d}\rho(x) = D(u). \end{aligned}$$

This concludes the proof. □

5. Properties of the Kantorovich potentials

Let $C(\gamma)$ be as before

$$C(\gamma) = \int_X \sum_{1\le i<j\le N} f(d(x_i, x_j)) \,\mathrm{d}\gamma.$$

We denote by $C^R(\gamma)$ the truncation of a cost $C(\gamma)$ from above[3],

$$C^R(\gamma) = \int_{X^N} c^R(x_1, \ldots, x_N) \,\mathrm{d}\gamma, \text{ for all } \gamma \in \mathcal{P}(X^N),$$

where we have denoted by c^R the corresponding truncation of c,

$$c^R(x_1, \ldots, x_N) = \sum_{1\le i<j\le N} \min\{R, f(d(x_i, x_j))\}.$$

Proposition 5.1. *Let $\rho \in \mathcal{P}(X)$ satisfy the assumptions (A) and (B). Fix $\beta > 0$ such that*

$$\sup_{x\in X} \rho(B(x, \beta)) < \frac{1}{N(N-1)^2}.$$

Then, for any $\alpha < f^{-1}\left(\frac{N^2(N-1)}{2} f(\beta)\right)$ and for all optimal $\gamma \in \Gamma(\rho)$ associated to $C(\gamma)$, we have

$$C(\gamma) \le \frac{N^3(N-1)^2}{4} f(\beta) \quad \text{and} \quad C(\gamma) = C^{f(\alpha)}(\gamma). \tag{5.1}$$

Moreover, for the same α, any Kantorovich potential u_α for $C^{f(\alpha)}$ is also a Kantorovich potential for C.

Proof. For each

$$\alpha < f^{-1}\left(\frac{N^2(N-1)}{2} f(\beta)\right),$$

we know by Theorem 3.2 that the support of γ can intersect at most the boundary of D_α. Therefore, since f is decreasing, we have for all $(x_1, \ldots, x_N) \in \operatorname{spt}(\gamma)$ the estimate

$$c(x_1, \ldots, x_N) = \sum_{1\le i<j\le N} f(d(x_i, x_j)) \le \frac{N(N-1)}{2} f(\alpha).$$

Thus, since γ is a probability measure, we have

$$C(\gamma) \le \int_{X^N} \frac{N(N-1)}{2} f(\alpha) \,\mathrm{d}\gamma = \frac{N(N-1)}{2} f(\alpha).$$

[3]Notice that we have used the notation C_R to correspond to the cost truncated from below.

Taking $\alpha \to f^{-1}\left(\frac{N^2(N-1)}{2}f(\beta)\right)$, we then get

$$\begin{aligned} C(\gamma) &\le \frac{N(N-1)}{2} f\left(f^{-1}\left(\frac{N^2(N-1)}{2}f(\beta)\right)\right) \\ &= \frac{N(N-1)}{2}\cdot\frac{N^2(N-1)}{2}f(\beta) = \frac{N^3(N-1)^2}{4}f(\beta), \end{aligned}$$

which gives the left-hand side in (5.1). Let us then fix an optimal plan γ_α for the cost $C^{f(\alpha)}$. Then $\operatorname{spt}(\gamma_\alpha) \subset X^N \setminus D_\alpha$, so $c = c^{f(\alpha)}$ on $\operatorname{spt}(\gamma_\alpha)$. Thus,

$$C(\gamma) \le \int_{X^N} c \,\mathrm{d}\gamma_\alpha = \int_{X^N} c^{f(\alpha)}\,\mathrm{d}\gamma_\alpha = C^{f(\alpha)}(\gamma_\alpha).$$

The opposite inequality is simply due to the monotonicity of the integral. It remains to prove the last part of the statement. We fix a Kantorovich potential u_α for $C^{f(\alpha)}$. It satisfies, for $\rho^{\otimes(N)}$-almost every $(x_1,\dots,x_N) \in X^N$ the estimate

$$u_\alpha(x_1) + \cdots + u_\alpha(x_N) \le c^{f(\alpha)}(x_1,\dots,x_N) \le c(x_1,\dots,x_N).$$

Hence, u_α is also a Kantorovich potential for the cost function c and, moreover,

$$\int_X u(x)\,\mathrm{d}\rho(x) = \min_{\gamma\in\Gamma(\rho)} C(\gamma) = \min_{\gamma\in\Gamma(\rho)} C^{f(\alpha)}(\gamma) = N\int_X u_\alpha(x)\,\mathrm{d}\rho(x).$$

This concludes the proof. □

Theorem 5.2. *Let (X,d) be a Polish space. Suppose $\rho \in \mathcal{P}(X)$ such that (A) and (B) hold and $c\colon X^N \to \mathbb{R}\cup\{+\infty\}$ is a cost function*

$$c(x_1,\dots,x_N) = \sum_{1\le i<j\le N} f(d(x_i,x_j)), \quad \text{for all } (x_1,\dots,x_N)\in X^N,$$

where $f\colon [0,+\infty[\to \mathbb{R}\cup\{+\infty\}$ is a function satisfying (F1) and (F2).

Let $\beta > 0$ be such that

$$\sup_{x\in X} \rho(B(x,\beta)) < \frac{1}{N(N-1)^2}.$$

Assume additionally that, for some $\alpha < f^{-1}\left(\frac{N^2(N-1)}{2}f(\beta)\right)$, the restriction $f|_{[\alpha,\infty[}$ is Lipschitz. Then, there exists a Kantorovich potential w in (4.3) that is Lipschitz.

The following lemma is useful for proving Theorem 5.2. The proof follows in the same way as the proof of [3, Lemma 3.3].

Lemma 5.3. *Let u be a Kantorovich potential for the problem (1.2), i.e. a maximizer of the problem (1.2). Then there exists a Kantorovich potential $\tilde u$ such that $\tilde u \ge u$ which satisfies the representation*

$$\tilde u(x) = \inf\left\{ c(x,x_2,\dots,x_N) - \sum_{i\ge 2}\tilde u(x_i) \;:\; x_j\in X \text{ for all } j \right\}. \tag{5.2}$$

Proof of the Theorem 5.2. According to Lemma 5.3, we may choose a Kantorovich potential u_α for the truncated cost $C^{f(\alpha)}$ satisfying, for all $x \in X$,

$$u_\alpha(x) = \inf \left\{ c^{f(\alpha)}(x, x_2, \dots, x_N) - \sum_{j=1}^{N} u_\alpha(x_j) \mid x_j \in X \right\}.$$

By Proposition 5.1, due to the choice of α, u_α is also a Kantorovich potential for C. So, it suffices to show that u_α is Lipschitz. Since $f|_{[\alpha,\infty[}$ and d are Lipschitz, the function $h\colon X \to \mathbb{R} \cup \{-\infty, +\infty\}$,

$$h(x) = \sum_{1 \le i < j \le N} c^{f(\alpha)}(x, x_2, \dots, x_N) - \sum_{j=2}^{N} u_\alpha(x_j) \text{ for all } x \in X,$$

is Lipschitz with a Lipschitz constant that does not depend on $(x_2, \dots, x_N)$. Since the infimum of a family of uniformly Lipschitz functions is Lipschitz, we have that u_α is Lipschitz. $\square$

Finally, we can move on to the continuity properties of the cost functional $C(\rho)$ with respect to the marginal ρ.

Proposition 5.4. *Under the same assumptions as in Theorem 5.2, let (ρ_n) be a sequence in $\mathcal{P}(X^N)$, weakly converging to some $\rho_\infty \in \mathcal{P}(X^N)$ that satisfies (A). If*

$$\int_{X \setminus B(o,r)} f\left(2d(x,o)\right) \mathrm{d}\rho_n(x) \to 0 \quad \textit{uniformly when } r \to 0, \tag{5.3}$$

then

$$\lim_{n \to \infty} C(\rho_n) = C(\rho).$$

Proof. By [3, Theorem 3.9], the above result holds for the singular costs C_R which are bounded from below. Therefore, it suffices to show that for each $\varepsilon > 0$ there exists $R \in \mathbb{N}$ such that

$$|C(\rho_n) - C_R(\rho_n)| < \varepsilon$$

for all $n \in \mathbb{N} \cup \{\infty\}$. Since the inequality $C \le C_R$ always holds, it suffices to show that $C_R(\rho_n) - C(\rho_n) < \varepsilon$ for R large enough. Let us assume $R > 1$. Then, by the triangle inequality, for $x_1, x_2 \in X$

$$\max\{2d(x_1,o), 2d(x_2,o)\} \ge d(x_1,x_2). \tag{5.4}$$

Since f is decreasing, we have from (5.4) for every $x_1, x_2 \in X$ the estimate

$$f(d(x_1,x_2)) \ge f(\max\{2d(x_1,o), 2d(x_2,o)\}). \tag{5.5}$$

In order to obtain $C_R(\rho_n) - C(\rho_n) < \varepsilon$, we take a minimizer γ_n for C with marginals ρ_n (given by Proposition 3.1). Now, from (5.4) we see that for $d(x_1,x_2) \ge R > 1$, we have

$$f(1) - f(\max\{2d(x_1,o), 2d(x_2,o)\}) \ge 0$$

Using this together with the estimate (5.5), and assuming $\gamma_n \in \Gamma^{sym}(\rho_n)$ by Proposition 3.5, we get

$$\begin{aligned} C_R(\rho_n) - C(\rho_n) &\leq \int_{X^N} (C_R - C)\,\mathrm{d}\gamma_n = \int_{X^N} \sum_{1\leq i<j\leq N} \max\{f(R) - f(d(x_i,x_j)), 0\}\,\mathrm{d}\gamma_n \\ &\leq N(N-1)\int_{d(x_1,x_2)\geq R} (f(1) - f(d(x_1,x_2)))\,\mathrm{d}\gamma_n \\ &\leq N(N-1)\int_{d(x_1,x_2)\geq R} (f(1) - f(\max\{2d(x_1,o), 2d(x_2,o)\}))\,\mathrm{d}\gamma_n \\ &\leq 2N(N-1)\int_{d(x,o)\geq \frac{R}{2}} (f(1) - f(2d(x,o)))\,\mathrm{d}\rho_n < \varepsilon, \end{aligned}$$

for large enough R by assumption (5.3). □

6. Monge Problem for log-type costs

Regarding the existence of Monge-type minimizers in (1.1), the first positive result for repulsive type costs is shown in [11] where, in dimension $d = 1$, $X = \mathbb{R}$, M. Colombo, L. De Pascale and S. Di Marino prove that, for an absolutely continuous measure, a symmetric optimal plan γ is always induced by a cyclical optimal map T.

Theorem 6.1 (Colombo, De Pascale and Di Marino, [11])**.** *Let $\mu \in \mathcal{P}(\mathbb{R})$ be an absolutely continuous probability measure and $f\colon \mathbb{R} \to \mathbb{R}$ strictly convex, bounded from below and non-increasing function. Then there exists a unique optimal symmetric plan $\gamma \in \Gamma^{sym}(\mu)$ that solves*

$$\min_{\gamma\in\Gamma^{sym}(\mu)} \int_{\mathbb{R}^N} \sum_{1\leq i<j\leq N} f(|x_j - x_i|)\,\mathrm{d}\gamma.$$

Moreover, this plan is induced by an optimal cyclical map T, that is, $\gamma_{sym} = \frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N} \sigma_\sharp\gamma_T$, where $\gamma_T = (Id, T, T^{(2)}, \dots, T^{(N-1)})_\sharp\mu$. An explicit optimal cyclical map is

$$T(x) = \begin{cases} F_\mu^{-1}(F_\mu(x) + 1/N) & \text{if } F_\mu(x) \leq (N-1)/N \\ F_\mu^{-1}(F_\mu(x) + 1 - 1/N) & \text{otherwise.} \end{cases}$$

Here $F_\mu(x) = \mu(-\infty, x]$ is the distribution function of μ, and F_μ^{-1} is its lower semicontinuous left inverse.

We remark that, due to Theorem 2.4, the above Theorem 6.1 also holds for unbounded cost functions satisfying $(F1)$ and $(F2)$ and under the additional assumption (B) on the absolutely continuous measure μ. This can be seen for instance by taking a minimizer for the unbounded cost and observing that its restriction to a bounded set is also a minimizer of a truncated for and thus of the form given by Theorem 6.1.

6.1. Log-type cost ($N = 2$).

Here we consider $X = \mathbb{R}^d$ with $d \geq 1$.

Theorem 6.2. *Let $\rho \in \mathcal{P}(\mathbb{R}^d)$ be a probability measure such that (A) and (B) hold. Then there exists a unique optimal plan $\gamma_O \in \Gamma(\rho,\rho)$ for the problem*

$$\min_{\gamma\in\Gamma(\rho,\rho)} \int_{\mathbb{R}^d\times\mathbb{R}^d} -\log(|x_1 - x_2|)\,\mathrm{d}\gamma(x_1,x_2). \tag{6.1}$$

Moreover, this plan is induced by an optimal map T, that is, $\gamma = (Id, T)_\sharp \rho$, and $T(x) = x - \frac{\nabla u}{|\nabla u|^2}$ ρ-almost everywhere, where u is a Lipschitz maximizer for the dual problem (1.2).

Proof. Let us consider γ a minimizer for the problem (6.1) and u a maximizer of the dual problem, which is Lipschitz by Theorem 5.2. Then,

$$F(x_1, x_2) = u(x_1) + u(x_2) + \log(|x_1 - x_2|) \leq 0,$$

for $\rho \otimes \rho$-almost every $(x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$. Moreover, $F = 0$ γ-almost everywhere. But then F has a maximum on the support of γ and so $\nabla F = 0$ in this set; in particular we have that $\nabla u(x_1) = \frac{(x_1 - x_2)}{|x_1 - x_2|^2}$ on the support of γ. By solving this equation for x_2, we have

$$x_2 = x_1 - \frac{\nabla u(x_1)}{|\nabla u(x_1)|^2}, \quad \gamma - \text{almost everywhere},$$

which implies $\gamma = (Id, T)_\sharp \mu$ as we wanted to show. □

Acknowledgments

The authors thank Michael Seidl for fruitful discussions and for suggesting the charged wire model presented in the introduction of this paper. We also thank the anonymous referee for carefully reading the paper and for his many suggestions.

References

[1] U. Bindini and L. De Pascale, *Optimal transport with Coulomb cost and the semiclassical limit of density functional theory*, J. Éc. polytech. Math., 4 (2017), pp. 909–934.
[2] A. Braides, *Gamma-convergence for Beginners*, vol. 22, Clarendon Press, 2002.
[3] G. Buttazzo, T. Champion, and L. De Pascale, *Continuity and estimates for multimarginal optimal transportation problems with singular costs*, Applied Mathematics & Optimization, 78 (2018), pp. 185–200.
[4] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi, *Optimal-transport formulation of electronic density-functional theory*, Physical Review A, 85 (2012), p. 062502.
[5] H. Chen, G. Friesecke, and C. B. Mendl, *Numerical methods for a kohn–sham density functional model based on optimal transport*, Journal of Chemical Theory and Computation, 10 (2014), pp. 4360–4368.
[6] M. Colombo and F. Stra, *Counterexamples to multimarginal optimal transport maps with coulomb cost and radial measures*, Mathematical Models and Methods in Applied Sciences, 26 (2016), pp. 1025–1049.
[7] L. Cort, D. Karlsson, G. Lani, and R. van Leeuwen, *Time-dependent density-functional theory for strongly interacting electrons*, Physical Review A, 95 (2017), p. 042505.
[8] C. Cotar, G. Friesecke, and C. Klüppelberg, *Density functional theory and optimal transportation with coulomb cost*, Comm. Pure Appl. Math., 66 (2013), pp. 548–599.
[9] ———, *Smoothing of transport plans with fixed marginals and rigorous semiclassical limit of the Hohenberg-Kohn functional*, Arch. Ration. Mech. Anal., 228 (2018), pp. 891–922.
[10] L. De Pascale, *Optimal transport with coulomb cost. approximation and duality*, ESAIM: Math. Model. Numer. Anal. (Special issue on "Optimal transport in applied mathematics"), 49 (2015), pp. 1643–1657.
[11] S. Di Marino, L. De Pascale, and M. Colombo, *Multimarginal optimal transport maps for* 1*-dimensional repulsive costs*, Canadian Journal of Mathematics – Journal Canadien des Mathématiques, 67 (2015), pp. 350–368.
[12] S. Di Marino, A. Gerolin, K. Giesbertz, L. Nenna, M. Seidl, and P. Gori-Giorgi, *The strictly-correlated electron functional for spherically symmetric systems revisited.*, in preparation, (2016).
[13] S. Di Marino, A. Gerolin, and L. Nenna, *Optimal transport for repulsive costs*, Topological Optimization and Optimal Transport In the Applied Sciences, (2017).
[14] G. Friesecke, C. B. Mendl, B. Pass, C. Cotar, and C. Klüppelberg, *N-density representability and the optimal transport limit of the hohenberg-kohn functional*, The Journal of chemical physics, 139 (2013), p. 164109.

[15] W. Gangbo and V. Oliker, *Existence of optimal maps in the reflector-type problems*, ESAIM: Control, Optimisation and Calculus of Variations, 13 (2007), pp. 93–106.
[16] A. Gerolin, A. Kausamo, and T. Rajala, *Non-existence of optimal transport maps for the multi-marginal repulsive harmonic cost*, arXiv preprint arXiv:1805.00417, (2018).
[17] P. Gori-Giorgi, M. Seidl, and G. Vignale, *Density-functional theory for strongly interacting electrons*, Physical review letters, 103 (2009), p. 166402.
[18] W. E. Hartnett and A. H. Kruse, *Differentiation of set functions using Vitali coverings*, Trans. Amer. Math. Soc., 96 (1960), pp. 185–209.
[19] P. Hohenberg and W. Kohn, *Inhomogeneous electron gas*, Phys. Rev., 136 (1964).
[20] A. Käenmäki, T. Rajala, and V. Suomala, *Existence of doubling measures via generalised nested cubes*, Proc. Amer. Math. Soc., 140 (2012), pp. 3275–3281.
[21] H. G. Kellerer, *Duality theorems for marginal problems*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 67 (1984), pp. 399–432.
[22] G. Lani, S. Di Marino, A. Gerolin, R. van Leeuwen, and P. Gori-Giorgi, *The adiabatic strictly-correlated-electrons functional: kernel and exact properties*, Accepted to PCCP for the Baerends special issue, (2016).
[23] M. Levy, *Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem*, Proceedings of the National Academy of Sciences, 12 (76), pp. 6062–6065.
[24] M. Lewin, *Semi-classical limit of the Levy-Lieb functional in density functional theory*, C. R. Math. Acad. Sci. Paris, 356 (2018), pp. 449–455.
[25] E. H. Lieb, *Density functionals for coulomb systems*, in Inequalities, Springer, 2002, pp. 269–303.
[26] F. Malet and P. Gori-Giorgi, *Strong Correlation in Kohn-Sham Density Functional Theory*, Physical Review Letters, 109 (2012), p. 246402.
[27] M. Seidl, *Strong-interaction limit of density-functional theory*, Physical Review A, 60 (1999), p. 4387.
[28] M. Seidl, P. Gori-Giorgi, and A. Savin, *Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities*, Physical Review A, 75 (2007), p. 042511.
[29] C. Villani, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008.
[30] X.-J. Wang, *On the design of a reflector antenna ii*, Calculus of Variations and Partial Differential Equations, 20 (2004), pp. 329–341.

Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of Jyväskylä, Finland
E-mail address: augusto.gerolin@jyu.fi
E-mail address: anna.m.kausamo@jyu.fi
E-mail address: tapio.m.rajala@jyu.fi

[C]

Multi-marginal Entropy-Transport with repulsive cost

A. Gerolin, A. Kausamo, and T. Rajala

Preprint

MULTI-MARGINAL ENTROPY-TRANSPORT WITH REPULSIVE COST

AUGUSTO GEROLIN, ANNA KAUSAMO, AND TAPIO RAJALA

Abstract. In this paper we study theoretical properties of the entropy-transport functional with repulsive cost functions. In this preliminary version, we provide sufficient conditions for the existence of a minimizer in a class of metric spaces and prove the Γ-convergence of the entropy-transport functional to a multi-marginal optimal transport problem with a repulsive cost.

1. Introduction

In this paper we consider the following multi-marginal entropy-transport problem

$$I_\varepsilon[\rho] = \inf_{\gamma \in \Pi_N^{\mathrm{sym}}(\rho)} (C_0[\gamma] + \varepsilon E[\gamma]), \tag{1.1}$$

where $C_0[\gamma] = \int_{X^N} c d\gamma$ is the transportation cost related to a cost function c, $E[\gamma]$ is the entropy, and $\varepsilon \geq 0$ is a parameter, see Section 2 for details. We consider the setting where $(X, d, \mathfrak{m})$ is a Polish measure space, and $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ is an absolutely continuous probability measure with respect to the reference measure $\mathfrak{m}$. An element $\gamma \in \Pi_N^{\mathrm{sym}}(\rho)$ is called a symmetric coupling (or transport plan), i.e. a symmetric probability measure in X^N having all marginals equals to $\rho\mathfrak{m}$.

We are interested in a class of repulsive cost functions $c\colon X^N \to \mathbb{R} \cup \{+\infty\}$ of the form

$$c(x_1, \dots, x_N) = \sum_{1 \leq i < j \leq N} f(d(x_i, x_j)), \quad \text{for all } (x_1, \dots, x_N) \in X^N.$$

We assume $f\colon]0, \infty[\to \mathbb{R}$ to be a continuous and decreasing function (eventually singular). Among the examples of such cost functions we have the Coulomb cost $f(z) = 1/|z|$, the Riesz cost $f(z) = 1/|z|^s, d \geq s \geq \max\{d-2, 0\}$ and the logarithmic cost $f(z) = -\log(|z|)$. We observe that when $\varepsilon = 0$, the entropy-transport problem above reduces to the classical multi-marginal optimal transport problem with repulsive costs [7, 9, 10, 14].

The motivation of this paper comes from both theory and numerics. For repulsive cost functions, the entropy term in (1.1) plays a role of a *regularizer* to compute numerically a solution γ of the multi-marginal optimal transport problem $I_0[\rho]$, see [4]. Numerical experiments suggest that when the regularization parameter ε goes to 0, the minimizer γ_ε converges to a minimizer of $I_0[\rho]$ having maximal entropy among the minimizer of $I_0[\rho]$.

From a theoretical viewpoint, this type of functional has direct relevance in Density Functional Theory. By choosing carefully the parameter ε, the functional (1.1) provides a lower-bound for the Hohenberg-Kohn functional in Density Functional Theory [22, 25]. This is a immediate consequence of the Log-Sobolev Inequality.

Date: November 22, 2018.

The authors acknowledge the support of the Academy of Finland, projects no. 274372, 284511, 312488, and 314789.

The entropy-transport problem has appeared previously in the literature in the attractive case, in particular when $c(x_1, x_2) = d(x_1, x_2)^2$. We mention briefly below some of connections of the entropy-transport with other fields and point out the relevance in the Coulomb case.

Brief comments on some applications of the entropy-transport
Optimal Transport and Sinkhorn algorithm: The entropy-transport (1.1) was introduced by M. Cuturi [12] in order to compute numerically a solution of the optimal transport plan for the distance squared cost in the 2-marginals case via the Sinkorn algorithm. Due to its reasonable computational cost[1], it has been applied to a wide range of problems in various research areas, including Information Theory, Computer Graphics, Statistical Inference, Machine Learning, and Mean-Field Games. The entropic regularization method was also considered in the (attractive) multi-marginal case in the so-called *barycenter problem* introduced by M. Agueh and G. Carlier [1] (see also [8]) and in numerical methods in the time discretization of Brenier's relaxed formulation of the incompressible Euler equation [5]. For a thorough presentation of the computational aspects we refer to M. Cuturi and G. Peyré's book [23].

Second-order Calculus on RCD spaces: N. Gigli and L. Tamanini [17] studied the entropic-transport problem on a class of metric spaces with (Riemannian) Ricci curvature bounded from below (2-marginals case, $c(x_1, x_2) = d(x_1, x_2)^2$). The entropic regularization procedure was crucial to establish a second-order differential structure in that setting.

Schrödinger Problem: In 1926, E. Schrödinger introduced the (linear) Schrödinger equations describing the non-relativistic evolution of a single particle in an electric field with potential energy and also established an equivalence between such equations and a system of diffusion equations [24]. Roughly speaking, the variational problem (see (1.1) with $X = C([0,1], \mathbb{R}^d)$ and $N = 2$) arises in the Schrödinger manuscript while studying the limit $k \to \infty$ ($N = 2$) of the empirical measures associated to the evolution of k i.i.d. Brownian motions. We refer the reader to C. Léonard survey [20] for technical details and historical notes.

Lower bound of the Hohenberg-Kohn functional in Density Functional Theory: This is the particular case where the Entropy-Transport problem with Coulomb cost comes into play. It has been shown in [22, 25] that the functional (1.1) provides a lower bound for computing the ground state energy of the Hohenberg-Kohn functional [7, 9, 10, 14, 21]. Below we give a brief description of the result. Notice that in this context $X = \mathbb{R}^d$ and $\mathfrak{m}$ is the Lebesgue measure $\mathcal{L}^d$.

Assume that $\gamma \in \Pi_N(\rho)$ such that $\sqrt{\gamma} \in H^1(\mathbb{R}^{dN})$. This is the case, for example, when $\gamma(x_1, \dots, x_N) = |\psi(x_1, \dots, x_N)|^2$, where $\psi \in H^1(\mathbb{R}^{dN})$ is a ground-state wave function solving the N-electron Schrödinger Equation (see [9, 10, 14, 25] for details). Then, we can define the Hohenberg-Kohn functional by

$$\tilde{F}_\hbar^{HK}[\rho] = \inf_{\gamma \in \Pi_N(\rho), \sqrt{\gamma} \in H^1(\mathbb{R}^{dN})} \left\{ \frac{\hbar^2}{2} \int_{\mathbb{R}^{dN}} |\nabla \sqrt{\gamma}|^2 dx_1 \dots dx_N + \int_{\mathbb{R}^{dN}} \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|} d\gamma \right\}.$$

Therefore, as a consequence of the logarithmic Sobolev inequality for the Lebesgue measure [18], the following result holds: if $\rho\mathcal{L}^d \in \mathcal{P}(\mathbb{R}^d)$ and $\sqrt{\gamma} \in H^1(\mathbb{R}^{dN})$ then

$$C_0[\rho] \le C_\varepsilon[\rho] \le \tilde{F}_\hbar^{HK}[\rho], \quad \text{with } \varepsilon = \pi\hbar^2/2.$$

[1] In the 2-marginal case, the discreate optimal coupling γ can be computed in $O(N\log(N))$ operations (FFT, IIR approximation), see [3].

Example of optimal entropy couplings. Let us present some computational examples of minimizers of $I_\varepsilon[\rho]$ illustrating the role of the parameter ε. Before this, we recall a result on the characterization of minimizers in the one-dimensional case [13]. In particular, according to it the minimizer of $I_0[\rho]$ is concentrated on finitely many graphs and thus singular with respect to the product reference measure.

Theorem 1.1. *Let $\mu \in \mathcal{P}(\mathbb{R})$ be an absolutely continuous probability measure and $f\colon \mathbb{R} \to \mathbb{R}$ strictly convex, bounded from below and non-increasing function. Then there exists a unique optimal symmetric plan $\gamma \in \Gamma^{\mathrm{sym}}(\mu)$ that solves*

$$\min_{\gamma \in \Pi_N^{\mathrm{sym}}(\mu)} \int_{\mathbb{R}^N} \sum_{1 \le i < j \le N} f(|x_j - x_i|)\, \mathrm{d}\gamma.$$

Moreover, this plan is induced by an optimal cyclical map T, that is, $\gamma_{\mathrm{sym}} = (\gamma_T)^S$, where $\gamma_T = (\mathrm{id}, T, T^{(2)}, \dots, T^{(N-1)})_\sharp \mu$. An explicit optimal cyclical map is

$$T(x) = \begin{cases} F_\mu^{-1}(F_\mu(x) + 1/N) & \text{if } F_\mu(x) \le (N-1)/N \\ F_\mu^{-1}(F_\mu(x) + 1 - 1/N) & \text{otherwise.} \end{cases}$$

Here $F_\mu(x) = \mu(-\infty, x]$ is the distribution function of μ, and F_μ^{-1} is its lower semicontinuous left inverse.

One-dimensional Entropic-Transport with Coulomb cost and a Gaussian measure. Let ρ be a normal distribution on the real line with zero mean and standard deviation $\sigma = 5$. We compute numerically the solution of the entropic-transport problem with Coulomb cost in the real line using the Sinkorn algorithm [11]. Notice that by Theorem 1.1, we know that the minimizer of $I_0[\rho]$ is concentrated on a graph. See Figure 1 for an illustration of the computational results. Our code is based on the Python implementation available at POT library [15].

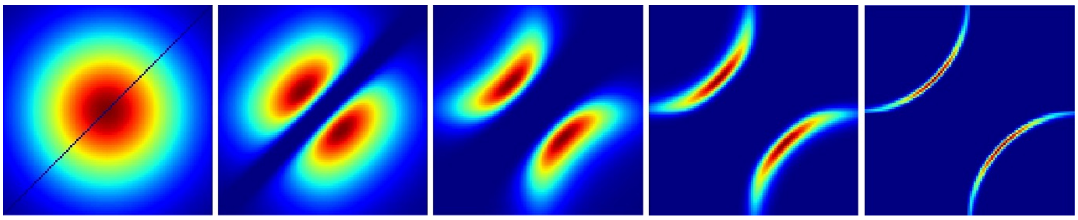

Figure 1. The dependence of the minimizer of the entropic-transport problem (1.1) on the entropic parameter ε for the one-dimensional Coulomb cost, $N = 2$ and $\rho \sim N(0, 5)$. The pictures show part of the support of the optimal coupling γ_ε around the origin. From the left to right: $\varepsilon = 10^4, 1/e^2, 1/e^3, 1/e^4, 1/e^5$.

Organization of the paper. In Section 2 we introduce the setting and study sufficient conditions for the existence of minimizers for the entropy-transport problem (1.1). Section 3 is devoted to the development of the main tools for the $\Gamma-$convergence proof of the entropic-transport functional $C_\varepsilon[\gamma]$ to the multi-marginal optimal transport with repulsive costs $C_0[\gamma]$.

Strategy of the main proof and some technical remarks. The main result of this paper is Theorem 3.1, in which we prove the Γ-convergence of the entropic-regularized functional $C_\varepsilon[\gamma]$ to $C_0[\gamma]$. The technical difficulty on dealing with the Γ-convergence comes from the fact that while for the entropic part $E[\gamma]$ the minimizer γ tends to be *as spread as possible* with respect to $\mathfrak{m}$, for the cost $C_0[\gamma]$ a minimizer can be very singular and have infinite entropy.

We divide the proof in two parts. The part (I), the $\liminf$ −inequality, follows basically from the lower-semicontinuity of the costs $C_0[\gamma]$ and $C_\varepsilon[\gamma]$ - which are obtained from the assumption $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$ on the marginal measure $\rho\mathfrak{m}$, giving a lower bound on the entropy. The part (II), the $\limsup$ −inequality, is more involved. In Section 3.2, we construct a block approximation γ'_n for a coupling γ with $C_0[\gamma] < +\infty$. Such construction is done in several steps, since we need to construct a competitor γ'_n such that $E[\gamma'_n] < \infty$ and $\gamma'_n \in \Pi^{\mathrm{sym}}_N(\rho)$.

Finally, we point out that our construction can deal with the case when the space X is a domain in $\mathbb{R}^d$, answering a question raised in [4]. In [8], G. Carlier, V. Duval, G. Peyré and B. Schmitzer carried out a similar block approximation procedure for the two-marginal squared distance cost in the Euclidean space.

2. The entropy-regularized repulsive costs

Let (X, d) be a Polish space and $\mathfrak{m}$ be a reference measure on X. We denote by $\mathcal{P}(X)$ the set of Borel probability measures on X, and $\mathcal{P}^{ac}(X)$ the set of Borel probability measures on X that are absolutely continuous with respect to $\mathfrak{m}$. We denote by $\mathfrak{m}_N$ the product measure $\mathfrak{m} \otimes \mathfrak{m} \otimes \cdots \otimes \mathfrak{m}$. This is the reference measure we use on the product space X^N. On X^N we use the sup-metric, which we denote by d_N.

The class of cost functions $c\colon X^N \to \mathbb{R} \cup \{+\infty\}$ of our interests is given by functions of the form

$$c(x_1, \ldots, x_N) = \sum_{1 \le i < j \le N} f(d(x_i, x_j)), \qquad \text{for all } (x_1, \ldots, x_N) \in X^N,$$

where $f : [0, \infty[\to \mathbb{R} \cup \{+\infty\}$ satisfies the following conditions

$$f|_{]0,\infty[} \text{ is continuous, decreasing and} \tag{F1}$$

$$\lim_{t \to 0+} f(t) = +\infty\,. \tag{F2}$$

Above and from now on, we denote by $(x_1, \ldots, x_N)$ points in X^N, so $x_i \in X$ for each i.

We denote by

$$\Pi(\rho) = \left\{\gamma \in \mathcal{P}(X^N) \mid \mathtt{pr}^i_\sharp \gamma = \rho \text{ for all } i \in \{1, \ldots, N\}\right\}$$

the set of *couplings* or *transport plans*, where $\mathtt{pr}^i$ is the projection

$$\mathtt{pr}^i(x_1, \ldots, x_i, \ldots, x_N) = x_i \text{ for all } (x_1, \ldots, x_i, \ldots, x_N) \in X^N\,.$$

For the definition of the of symmetric couplings $\Pi^{\mathrm{sym}}_N(\rho)$, see Definition 2.7. We define the functional $C_0[\gamma]$ to be the cost related to the coupling γ

$$C_0[\gamma] = \int_{X^N} c(x_1, \ldots, x_N)\, \mathrm{d}\gamma(x_1, \ldots, x_N)\,.$$

Given $\varepsilon \ge 0$, we denote by $C_\varepsilon[\gamma]$ the entropy-regularized cost

$$C_\varepsilon[\gamma] = C_0[\gamma] + \varepsilon E[\gamma], \qquad \text{for all } \gamma \in \Pi^{\mathrm{sym}}_N(\rho), \tag{2.1}$$

where the entropy $E : \mathcal{P}(X^N) \to \mathbb{R} \cup \{-\infty, +\infty\}$ is defined as

$$E[\gamma] = \begin{cases} \int_{X^N} \rho_\gamma \log \rho_\gamma \, \mathrm{d}\mathfrak{m}_N & \text{if } \gamma \ll \mathfrak{m}_N \\ +\infty & \text{otherwise} \end{cases} . \tag{2.2}$$

The notation ρ_γ stands for the Radon-Nikodym derivative of γ with respect to the reference measure $\mathfrak{m}_N$ and $\gamma \ll \mathfrak{m}_N$ means that γ is an absolutely continuous with respect to the reference measure $\mathfrak{m}_N$. Let $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$. In this paper we are interested in the following infimum

$$I_\varepsilon[\rho] := \inf_{\gamma \in \Pi_N^{\mathrm{sym}}(\rho)} C_\varepsilon[\gamma] \, . \tag{2.3}$$

In order to guarantee the lower semicontinuity for C_ε, we will assume $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$. This will take care of the entropy part $E[\cdot]$. In order to have lower semicontinuity for the functional $F_0[\cdot]$, we assume that ρ satisfies

$$\lim_{r \to 0} \sup_{x \in X} \rho(B(x,r)) < \frac{1}{N(N-1)^2} \tag{A}$$

and

$$\int_{X \setminus B(o,r_0)} f\left(2d(x,o)\right) \mathrm{d}\rho(x) > -\infty. \tag{B}$$

The condition (B) is a similar assumption than requiring, in the case of quadratic cost, that the marginal measures have finite second moments; while the condition (A) guarantees that the cost is finite.

If we endow the spaces $\mathcal{P}(X^N)$ and $\mathcal{P}(X)$ with w^*-topology, then, by Prokhorov's theorem, any subset of $\mathcal{P}(X)$ (or $\mathcal{P}(X^N)$) is tight if and only if it is relatively compact.

Proposition 2.1. *Let $(X, d, \mathfrak{m})$ be a Polish metric measure space. Assume that $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ satisfies $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$ and the conditions (A) and (B). Assume that $c\colon X^N \to \mathbb{R} \cup \{+\infty\}$ satisfies the conditions $(F1)$ and $(F2)$. Then, for each $\varepsilon \geq 0$, there exists a minimizer $\gamma \in \Pi_N^{\mathrm{sym}}(\rho)$ for the entropic-regularized cost $C_\varepsilon[\gamma]$. In addition, if $\varepsilon > 0$ the minimizer is unique.*

Proof. We notice that the set $\Pi(\rho)$ is a compact set in the w^*-topology [19]. The functional E is lower semicontinuous by Proposition 2.4, whereas in our setting the lower semicontinuity of C_0 is proven as part of the proof of [16, Proposition 3.1]. Moreover, due to the hypothesis $\rho \log \rho \in L^1_{\mathfrak{m}}(\rho)$, we have that $E[\gamma] < +\infty$, since the product measure has finite entropy. For each $\varepsilon \geq 0$, the functional C_ε is convex, and for $\varepsilon > 0$, the convexity is strict. Thus, we have the existence of the minimizers, and uniqueness in the case $\varepsilon > 0$. □

Remark 2.2 (Entropy-transport seen as a Kullback-Leibler divergence)**.** *We recall (see for instance* [20, 23]*) that the cost function $C_\varepsilon[\gamma]$ can be alternatively written as the Kullback-Leibler divergence between γ and a kernel $\tilde{\kappa}$ defined below*

$$C_\varepsilon[\gamma] = \mathrm{KL}(\gamma|\tilde{\kappa}) = \int_{X^N} \rho_\gamma \ln\left(\frac{\rho_\gamma}{\rho_\kappa}\right) \mathrm{d}\mathfrak{m}_N,$$

where $\tilde{\kappa} \in \mathcal{P}(X^N)$ and its Radon-Nikodym derivative with respect to $\mathfrak{m}_N$ is $\rho_\kappa = e^{-c/\varepsilon}$.

2.1. **Some properties of the entropy functional.** Let us start by noting that the minimum of the entropy is attained by the product measure and that its value is not $-\infty$.

Proposition 2.3. *Let $(X, d, \mathfrak{m})$ be a Polish metric measure space, and let $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ with $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$. Then*

$$\min_{\gamma \in \Pi_N^{\mathrm{sym}}(\rho)} E[\gamma] = \int_{X^N} \left(\otimes_{i=1}^N \rho \right) \log \left(\otimes_{i=1}^N \rho \right) \mathrm{d}\mathfrak{m}_N = N \int_X \rho \log \rho \, \mathrm{d}\mathfrak{m} > -\infty.$$

Proof. As we will see, the minimality is an immediate consequence of Jensen's inequality. Let $\gamma \in \Pi(\rho)$. Then

$$\begin{aligned} E[\gamma] &= \int_{X^N} \rho_\gamma \log(\rho_\gamma) \, \mathrm{d}\mathfrak{m}_N = \int_{X^N} \frac{\rho_\gamma}{\otimes_{i=1}^N \rho} \left(\log \left(\frac{\rho_\gamma}{\otimes_{i=1}^N \rho} \right) + \log \left(\otimes_{i=1}^N \rho \right) \right) \otimes_{i=1}^N \rho \, \mathrm{d}\mathfrak{m}_N \\ &\geq \left(\int_{X^N} \rho_\gamma \, \mathrm{d}\mathfrak{m}_N \right) \log \left(\int_{X^N} \rho_\gamma \, \mathrm{d}\mathfrak{m}_N \right) + \int_{X^N} \rho_\gamma \log \left(\otimes_{i=1}^N \rho \right) \, \mathrm{d}\mathfrak{m}_N \\ &= 0 + E[\otimes_{i=1}^N \rho]. \end{aligned}$$

□

Using Proposition 2.3 we immediately get the lower semicontinuity of the entropy functional by representing the entropy as relative entropy against the probability measure $\otimes_{i=1}^N (\rho\mathfrak{m})$.

Proposition 2.4. *Let $(X, d, \mathfrak{m})$ be a Polish metric measure space, and let $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ with $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$. Then $E[\cdot]$ is lower semicontinuous in $\Pi_N^{\mathrm{sym}}(\rho)$.*

2.2. **Some properties of the coupling cost $C_0[\gamma]$.** We first state the existence of a minimizer in (2.3) when $\varepsilon = 0$. Notice that in this section $\Pi_N(\rho)$ denotes the set of couplings in X^N (not necessarily symmetric). Moreover, we need to assume extra hypothesis on the probability measure ρ in order to guaranteee that $C_0[\gamma]$ is bounded from below for a $\gamma \in \Pi_N(\rho)$ (e.g. $f(z) = -\log(|z|)$).

Proposition 2.5. [16, Proposition 3.1] *Let (X, d) be a Polish space. Suppose that $\rho \in \mathcal{P}(X)$ satisfies (A) and (B), and $c\colon X^N \to \mathbb{R} \cup \{+\infty\}$ is a cost function*

$$c(x_1, \dots, x_N) = \sum_{1 \leq i < j \leq N} f(d(x_i, x_j)), \quad \text{for all } (x_1, \dots, x_N) \in X^N,$$

where $f\colon [0, \infty[\to \mathbb{R}$ satisfies $(F1)$ and $(F2)$. Then, the following minimum is achieved

$$I_0[\rho] = \min_{\gamma \in \Pi(\rho)} \int_{X^N} c(x_1, \dots, x_N) \, \mathrm{d}\gamma(x_1, \dots, x_N).$$

For $\alpha > 0$, we define the set D_α as

$$D_\alpha := \left\{ (x_1, \dots, x_N) \in X^N \mid \text{there exist } i, j \text{ such that } d(x_i, x_j) < \alpha \right\}.$$

The next theorem from [6] (see also [16, Theorem 3.2]) states that for measures ρ as in Proposition 2.5 there exists $\overline{\alpha} > 0$ for which the support of any optimal plan is concentrated away from the set $D_{\overline{\alpha}}$.

Theorem 2.6. [6, Theorem 3.2] *Let (X, d), ρ, f, c as in the Proposition 2.5 and let γ be a minimizer of*

$$I_0[\rho] = \min_{\gamma \in \Pi(\rho)} \int_{X^N} c(x_1, \dots, x_N) \, \mathrm{d}\gamma(x_1, \dots, x_N).$$

Let us fix $0 < \beta < 1$ *such that*

$$\sup_{x \in X} \rho(B(x,\beta)) < \frac{1}{N(N-1)^2}.$$

Then, we have for all

$$\alpha < f^{-1}\left(\frac{N^2(N-1)}{2} f(\beta)\right) \tag{2.4}$$

the inclusion

$$\mathrm{spt}(\gamma) \subset X^N \setminus D_\alpha, \tag{2.5}$$

where f^{-1} *stands for the left-inverse of* f.

As a last remark, we observe that one can restrict the problem $\min_{\gamma \in \Pi(\rho)} C_0[\gamma]$ to the class of symmetric couplings in X^N having all the marginals equal to ρ.

Definition 2.7 (Symmetric measures)**.** *A measure* $\gamma \in \mathcal{P}(X^N)$ *is symmetric if*

$$\int_{X^N} \phi(x_1, \dots, x_N)\, \mathrm{d}\gamma = \int_{X^N} \phi(\overline{\sigma}(x_1, \dots, x_N))\, \mathrm{d}\gamma, \ \textit{for all } \phi \in \mathcal{C}(X^N)$$

and for all permutations $\overline{\sigma}$ *of* N *symbols. We denote by* $\mathcal{P}^{\mathrm{sym}}(X^N)$ *the set of symmetric probability measures in* X^N*, and notice that* $\Pi_N^{\mathrm{sym}}(\rho) := \Pi_N(\rho) \cap \mathcal{P}^{\mathrm{sym}}(X^N)$.

Let us also introduce the notation for symmetrized measures. If γ is a Borel measure on X^N, we denote by γ^S the symmetrized measure

$$\gamma^S := \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \sigma_\sharp \gamma\,,$$

where $\mathcal{S}_N$ is the set of permutations of $\{1, \dots, N\}$. The following result follows immediately.

Proposition 2.8. *Under the hypothesis of Proposition 2.5, we have that*

$$\min_{\gamma \in \Pi(\rho)} \int_{X^N} c(x_1, \dots, x_N)\, \mathrm{d}\gamma = \min_{\gamma \in \Pi^{\mathrm{sym}}(\rho)} \int_{X^N} c(x_1, \dots, x_N)\, \mathrm{d}\gamma. \tag{2.6}$$

3. The Γ-convergence of Entropic-regularized cost

Now let us turn to the Γ-convergence. From now on, $(\tau_n)_{n \in \mathbb{N}}$ is any sequence of positive real numbers decreasing to zero. Let us introduce the following functionals: for each $n \in \mathbb{N}$

$$\mathcal{C}_n : \mathcal{P}^{\mathrm{sym}}(X^N) \to \mathbb{R} \cup \{+\infty\}, \ \mathcal{C}_n(\gamma) = \begin{cases} C_{\tau_n}[\gamma] & \text{if } \gamma \in \Pi(\rho) \\ +\infty & \text{otherwise} \end{cases}$$

and

$$\mathcal{C} : \mathcal{P}^{\mathrm{sym}}(X^N) \to \mathbb{R} \cup \{+\infty\}, \ \mathcal{C}(\gamma) = \begin{cases} C[\gamma] & \text{if } \gamma \in \Pi(\rho) \\ +\infty & \text{otherwise} \end{cases}.$$

The goal of this section is to prove that the sequence $(\mathcal{C}_{n \in \mathbb{N}})$ Γ-converges to $\mathcal{C}$ with respect to γ.

Theorem 3.1. *Let* $(X, d, \mathfrak{m})$ *be a Polish metric measure space. Let* $\rho \in \mathcal{P}^{ac}(X)$ *with* $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$ *satisfying (A) and (B). Then the sequence* $(\mathcal{C}_n)$ Γ*-converges to* $\mathcal{C}$ *with respect to* γ.

Let us fix $\gamma \in \mathcal{P}^{\mathrm{sym}}(X^N)$. We need to show that

$$\text{For each sequence } (\gamma_n)_{n\in\mathbb{N}} \text{ that converges to } \gamma \text{ we have } \liminf_{n\to\infty} \mathcal{C}_n[\gamma_n] \geq C[\gamma], \text{ and} \tag{I}$$

$$\text{There exists a sequence } (\gamma_n)_{n\in\mathbb{N}} \text{ that converges to } \gamma \text{ and } \limsup_{n\to\infty} \mathcal{C}_n[\gamma_n] \leq C[\gamma]\,. \tag{II}$$

The proof of Theorem 3.1 is divided into two parts. The proof of the first part, the liminf-inequality (I), is short and is established in the next subsection. The remainder of this section is then divided into subsections in which the second part, the limsup-inequality (II) is proven.

3.1. **Proof of condition (I).** We fix a sequence $(\gamma_n)_{n\in\mathbb{N}}$ that converges to γ. If $\gamma \notin \Pi(\rho)$, then since $\Pi(\rho)$ is compact, for large indices also $\gamma_n \notin \Pi(\rho)$, so both sides of inequality (I) are $+\infty$, and we are done. So, we may assume that γ and γ_n's are elements of $\Pi(\rho)$. Since now $\gamma_n \in \Pi(\rho)$, (I) follows from the lower-semicontinuity of $\gamma \mapsto \int c\mathrm{d}\gamma$ and from the entropy lower bound shown in Proposition 2.3.

3.2. **Constructing an approximation of the coupling γ.** First of all, we need to construct an approximation of γ only in the case where $C[\gamma] < \infty$. The idea of the construction is to redefine a large part of γ to be a product measure on finitely many Borel sets with small diameter. In order not to increase the cost by too much, the Borel sets we use have to be far away from the diagonal compared to the diameter of the sets. We call the part of the measure defined in this way *the core part* of the approximation. For the rest of the measure, we take another finite combination of product measures. However, this time the sets do not have small (or even bounded) diameter, but just small measure. This part will be called *the remainder part* of the approximation.

We start the construction by taking out a small part of γ that will later be used to deal with the remainder part of the approximation. For this we take a sequence of radii defined as $r_n = 1/n$. Since $C[\gamma] < \infty$, there exists a point $x = (x_1, \ldots, x_N) \in \mathrm{spt}(\gamma)$ with

$$x_i \neq x_j \quad \text{if } 1 \leq i < j \leq N.$$

Moreover, since $\gamma \in \Pi_N(\rho)$ and ρ satisfies (A), we have

$$\begin{aligned}
\gamma(\{(y_1, \ldots, &y_N) \in X^N \mid y_i \neq x_j \text{ for all } i,j\}) \\
&\geq 1 - \sum_{i\neq j} \gamma(\{(y_1, \ldots, y_N) \in X^N \mid y_i = x_j\}) \\
&= 1 - \sum_{i\neq j} \rho(X \setminus \{x_j\}) \geq 1 - N(N-1)\frac{1}{N(N-1)^2} > 0.
\end{aligned}$$

Thus, using again $C[\gamma] < \infty$, there exists another point $x' = (x_{N+1}, \ldots, x_{2N}) \in \mathrm{spt}(\gamma)$, so that

$$x_i \neq x_j \quad \text{if } 1 \leq i < j \leq 2N.$$

From now on, we consider x, x' fixed. Therefore, for $n \in \mathbb{N}$ sufficiently large we have

$$d(x_i, x_j) > r_n \quad \text{if } 1 \leq i < j \leq 2N. \tag{3.1}$$

Let us denote by

$$B_n = B(x, \tfrac{r_n}{10}) \quad \text{and} \quad B_n' := B(x', \tfrac{r_n}{10})$$

the balls around x and x' with radii $r_n/10$ in the sup-metric of the product space. So,

$$y = (y_1, \dots, y_N) \in B_n$$

if and only if

$$d(x_i, y_i) < \tfrac{r_n}{10} \text{ for all } i \in \{1, \dots, N\},$$

and analogously for B'_n with the relevant index modifications.

Let us now define

$$\gamma_{B_n} = \left(\frac{\gamma|_{B_n}}{\gamma(B_n)}\right)^S \quad \text{and} \quad \gamma_{B'_n} = \left(\frac{\gamma|_{B'_n}}{\gamma(B'_n)}\right)^S .$$

Observe that γ_{B_n} and $\gamma_{B'_n}$ are symmetric probability measures. Since the marginals of a symmetric measure are the same, we may denote by ρ_{B_n} the marginal of γ_{B_n} and similarly by $\rho_{B'_n}$ the marginal of $\gamma_{B'_n}$. Let us further denote $\tilde{B}_n := \operatorname{spt} \gamma_{B_n}$, $\tilde{B}'_n := \operatorname{spt} \gamma_{B'_n}$ and

$$\varepsilon_n := \frac{1}{N} \min \left\{ \gamma(\tilde{B}_n), \gamma(\tilde{B}'_n), r_n, \frac{r_n}{f(2r_n/5)} \right\} . \tag{3.2}$$

We then define a measure

$$\gamma_{0,n} := \gamma|_{X^n \setminus (\tilde{B}_n \cup \tilde{B}'_n)} + \frac{\gamma(\tilde{B}_n) - \varepsilon_n}{\gamma(\tilde{B}_n)} \gamma|_{\tilde{B}_n} + \frac{\gamma(\tilde{B}'_n) - \varepsilon_n}{\gamma(\tilde{B}'_n)} \gamma|_{\tilde{B}'_n} .$$

The idea behind the measure $\gamma_{0,n}$ is that we have chopped of a small part of the measure around the points x and x' (symmetrically) for later use. Since we are working with a singular cost, we still need to take out a small neighbourhood of the diagonals before approximating by product measures. We do this now.

We fix a compact $K_n \subset X$ such that

$$\gamma_{0,n}(X^N \setminus K_n^N) < \frac{\varepsilon_n}{2} \tag{3.3}$$

and take a small enough $\delta_n \in (0, r_n)$ so that

$$\gamma_{0,n}(D_{\delta_n}) < \frac{\varepsilon_n}{2}. \tag{3.4}$$

Using K_n and δ_n we then define

$$\gamma_{1,n} := \gamma_{0,n}|_{K_n^N \setminus D_{\delta_n}} . \tag{3.5}$$

The measure $\gamma_{1,n}$ is now the core part of the measure that we approximate. We denote by $\rho_{1,n}$ the marginals of the symmetric measure $\gamma_{1,n}$.

Let us then approximate the measure $\gamma_{1,n}$. We take $\lambda_n \in (0, \delta_n/n)$ so that

$$|f(r) - f(s)| < \varepsilon_n \qquad \text{for all } r, s \in [\delta_n/2, 2\operatorname{diam}(K_n)] \text{ with } |r - s| \le 2\lambda_n. \tag{3.6}$$

Such λ_n exists by the uniform continuity of f on $[\delta_n/2, 2\operatorname{diam}(K_n)]$. Since the set K_n is compact, we may fix a finite Borel partition $\{B_n^i\}_{i=1}^{M_n}$ of $\operatorname{spt}(\rho_{1,n})$ such that

$$\operatorname{diam}(B_n^i) < \lambda_n \quad \text{and} \quad 0 < \rho_{1,n}(B_n^i) < \varepsilon_n \text{ for all } i \in \{1, \dots, M_n\}.$$

We are now ready to define the core part approximants $\gamma_{1,n}^a$ as

$$\gamma_{1,n}^a = \sum_{(k_1, \dots, k_N) \in M_n^N} \frac{\gamma_{1,n}(B_n^{k_1} \times \cdots \times B_n^{k_N})}{\rho_{1,n}(B_n^{k_1}) \cdots \rho_{1,n}(B_n^{k_N})} \rho_{1,n}|_{B_n^{k_1}} \otimes \cdots \otimes \rho_{1,n}|_{B_n^{k_N}} . \tag{3.7}$$

Now let us handle the main part of the remainder of the measure, namely the measure

$$\gamma_{2,n} := \gamma_{0,n}|_{D_{4\varepsilon_n} \cup (X^N \setminus K_n^N)} .$$

Because $\gamma_{0,n}$ and the set where we restrict it is symmetric, also $\gamma_{2,n}$ is. We may thus denote its marginals by $\rho_{2,n}$.

In order to determine which part of the remaining marginal measure should be coupled where, we define a partition $\{A_{i,n}\}_{i=1}^N$ of the space X by setting, for all $i \in \{1, \dots, N-1\}$

$$A_{i,n} := \{y \in X \mid d(x_i, y) \le \tfrac{r_n}{2}\} ,$$

and

$$A_{N,n} := X \setminus \bigcup_{i=1}^{N-1} A_{i,n} .$$

Condition (3.1) guarantees that the sets $A_{i,n}$ are pairwise disjoint.

Now we approximate $\gamma_{2,n}$ by the measure

$$\gamma_{2,n}^a := N \left(\sum_{i=1}^N \eta_{n,i} \right)^S ,$$

where for all i the measure $\eta_{n,i}$ is the product

$$\eta_{n,i} := \left(\bigotimes_{k=1}^{i-1} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right) \otimes \rho_{\gamma_{2,n}}|_{A_{i,n}} \otimes \left(\bigotimes_{k=i+1}^{N} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right) .$$

By the definition of the sets $A_{i,n}$, for every $(y_1, \dots, y_N) \in \mathrm{spt}(\gamma_{2,n}^a)$ we have for each $i \ne j$

$$d(y_i, y_j) \ge |d(y_i, x_j) - d(x_j, y_j)| > \frac{r_n}{2} - \frac{r_n}{10} = \frac{2r_n}{5}, \tag{3.8}$$

where we have assumed (which we can do without loss of generality) that $y_j \in \overline{B}(x_j, r_n/10)$.

What we have done using the measure $\gamma_{2,n}^a$ is that we have coupled the marginals of the measure $\gamma_{2,n}$ with some suitable parts of the marginals of the reserved measure that was taken out around the point x. In this way we have used unevenly the marginals of this reserved part. To handle the rest of the reserved part of the measure around the point x, we now use the reserved measure around the point x'. So, we need to redefine the coupling for the part of the marginal given by

$$\rho_{3,n} := (\mathtt{pr}^1)_\sharp \frac{\varepsilon_n}{\gamma(\tilde{B}_n)} \gamma|_{\tilde{B}_n} + \rho_{2,n} - (\mathtt{pr}^1)_\sharp \gamma_{2,n}^a.$$

We define it as

$$\gamma_{3,n}^a := \left(\sum_{i=1}^N \phi_{n,i} \right)^S ,$$

where each $\phi_{n,i}$ is defined as

$$\phi_{n,i} := \left(\bigotimes_{k=N+1}^{N+i-1} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right) \otimes \rho_{3,n} \otimes \left(\bigotimes_{k=N+i+1}^{2N} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right) .$$

Since $\mathrm{spt}(\rho_{3,n}) \subset \mathrm{spt}(\rho_{B_n})$, we have that for every $(y_1, \dots, y_N) \in \mathrm{spt}(\gamma_{3,n}^a)$ and each $i \ne j$

$$d(y_i, y_j) \ge |d(x_{k(i)}, x_{k(j)}) - d(y_i, x_{k(i)} - d(x_{k(j)}, y_j)| > r_n - \frac{r_n}{10} - \frac{r_n}{10} = \frac{4r_n}{5}, \tag{3.9}$$

where $k(i) \neq k(j)$ are the indices for which $y_j \in \overline{B}(x_{k(j)}, r_n/10)$ and $y_i \in \overline{B}(x_{k(i)}, r_n/10)$.

What remains is the part of the measure around x' that was not used for $\gamma^a_{3,n}$. Since $\gamma^a_{3,n}$ used the marginals from this part of the reserved measure evenly, we may simply couple the rest by a measure

$$\gamma^a_{4,n} := b\left(\bigotimes_{k=N+1}^{2N} \frac{\rho_{B_n}|_{B(x_k,r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))}\right)^S,$$

with b being the correct scaling constant. Similarly as for the previous remainder part, we have that for every $(y_1, \ldots, y_N) \in \mathrm{spt}(\gamma^a_{4,n})$ and each $i \neq j$ the inequality (3.9) holds.

Now we are ready to define the full approximation as

$$\gamma'_n = \gamma^a_{1,n} + \gamma^a_{2,n} + \gamma^a_{3,n} + \gamma^a_{4,n}.$$

By construction $\gamma'_n \in \Pi^{\mathrm{sym}}(\rho)$.

3.3. **Narrow convergence of the approximations.** Let us now prove that the sequence $(\gamma'_n)_n$ narrowly converges to γ. We could argue this by using the Wasserstein distance. However, let us do it here directly using the definition of narrow convergence.

Lemma 3.2. *The sequence $(\gamma'_n)_n$ narrowly converges to γ.*

Proof. Let $\varphi \in C_b(X^N)$ and $\varepsilon > 0$. Since φ is continuous, there exist $\delta > 0$ and $A \subset X^N$ Borel so that $|f(x), f(y)| < \varepsilon/2$ for all $x, y \in A$ with $d_N(x, y) < \delta$, and $\gamma(A) > 1 - \varepsilon$. Now, let $n \in \mathbb{N}$ be so large that

$$\lambda_n < \delta \qquad \text{and} \qquad 12\varepsilon_n \sup_{x\in X^N} |\varphi(x)| < \varepsilon.$$

Then,

$$\left|\int_{X^N} \varphi \,\mathrm{d}\gamma_{1,n} - \int_{X^N} \varphi \,\mathrm{d}\gamma^a_{1,n}\right| < \varepsilon/2.$$

For the remainder part we have

$$(\gamma^a_{2,n} + \gamma^a_{3,n} + \gamma^a_{4,n})(X^N) < 3\varepsilon_n.$$

Thus, combining the above two estimates we get

$$\left|\int_{X^N} \varphi \,\mathrm{d}\gamma - \int_{X^N} \varphi \,\mathrm{d}\gamma'_n\right| < \varepsilon,$$

which concludes the proof. □

3.4. **Convergence of the cost.** In order to obtain (II), we need the cost $C_0[\cdot]$ to converge along the sequence. We prove this in the following lemma.

Lemma 3.3. *We have $C_0[\gamma'_n] \to C_0[\gamma]$ as $n \to \infty$.*

Proof. Let us first consider the remainder part. Recall that for $n \in \mathbb{N}$ we have

$$(\gamma'_n - \gamma^a_{1,n})(X^N) = (\gamma^a_{2,n} + \gamma^a_{3,n} + \gamma^a_{4,n})(X^N) < 3\varepsilon_n.$$

Thus, using the lower bounds (3.8) and (3.9) for distances in the support of the remainder part, and the definition (3.2) of ε_n, we get

$$\int_{X^N} c \,\mathrm{d}(\gamma'_n - \gamma^a_{1,n}) \leq \frac{N(N-1)}{2} f\left(\frac{2r_n}{5}\right) 3\varepsilon_n \leq \frac{3N(N-1)}{2} r_n \to 0 \tag{3.10}$$

as $n \to \infty$. By continuity of the integral, we get

$$\int_{X^N} c \,\mathrm{d}(\gamma - \gamma_{1,n}) \to 0 \tag{3.11}$$

as $n \to \infty$.

Let us now estimate the core part of the approximation. By the construction (3.7) of $\gamma_{1,n}^a$ and the choice (3.6) of λ_n, we have

$$\left| \int_{X^N} c \,\mathrm{d}\gamma_{1,n}^a - \int_{X^N} c \,\mathrm{d}\gamma_{1,n} \right| \le \int_{X^N} \frac{N(N-1)}{2} \varepsilon_n \,\mathrm{d}\gamma_{1,n} < \frac{N(N-1)}{2} \varepsilon_n. \tag{3.12}$$

Combining the above estimate with (3.10), (3.11) and (3.12) we get

$$\begin{aligned} \left| C_0[\gamma_n'] - C_0[\gamma] \right| \le & \left| \int_{X^N} c \,\mathrm{d}\gamma_{1,n}^a - \int_{X^N} c \,\mathrm{d}\gamma_{1,n} \right| + \left| \int_{X^N} c \,\mathrm{d}(\gamma_n' - \gamma_{1,n}^a) \right| \\ & + \left| \int_{X^N} c \,\mathrm{d}(\gamma - \gamma_{1,n}) \right| \to 0 \end{aligned}$$

as $n \to \infty$. $\square$

3.5. **Finiteness of the entropy for the approximations.** Next we show that the entropy is finite for the approximating sequence. Notice that, in order to prove (II), we do not need better estimate on the entropy.

Lemma 3.4. *For each $n \in \mathbb{N}$ we have $E[\gamma_n'] < \infty$.*

Proof. In order to see the finiteness of the entropy, it suffices to notice that each γ_n' is the sum of finitely many measures $(\tilde{\gamma}_{n,k})_{k=1}^{N_n}$ each of which is of the form $\tilde{\gamma}_{n,k} = \tilde{\rho}_1^k \mathfrak{m} \otimes \cdots \otimes \tilde{\rho}_N^k \mathfrak{m}$ with $\tilde{\rho}_i^k \ll \rho$ and $\frac{\mathrm{d}\tilde{\rho}_i^k}{\mathrm{d}\rho} \le 1$. Indeed, by Proposition 2.3, the entropy is always bounded from below, and so we can make a crude estimate:

$$\begin{aligned} E[\gamma_n'] &= \int_{X^N} \log \left(\sum_{k=1}^{N_n} \frac{\mathrm{d}\tilde{\gamma}_{n,k}}{\mathrm{d}\mathfrak{m}} \right) \mathrm{d} \sum_{k=1}^{N_n} \tilde{\gamma}_{n,k} \\ &\le \log(N_n) + \sum_{k=1}^{N_n} \int_{X^N} \log \left(\frac{\mathrm{d}\tilde{\gamma}_{n,k}}{\mathrm{d}\mathfrak{m}} \right) \mathrm{d}\tilde{\gamma}_{n,k} < \infty. \end{aligned}$$ $\square$

3.6. **Proof of condition (II).** We are now ready to prove the lim sup-inequality (II). By Lemma 3.2 we already know that $(\gamma_n')_n$ converges to γ. However, $\mathcal{C}_n[\gamma_n']$ need not converge to $\mathcal{C}[\gamma]$. This can be solved by making the convergence of $(\gamma_n')_n$ slower by repeating always the same measure for sufficiently (but finitely) many times before moving to the next one. We define $k(n)$ for every $n \in \mathbb{N}$ as

$$k(n) = \min \left(n, \max \left(1, \sup \left\{ k \in \mathbb{N} \,|\, \sqrt{\tau_n} E[\gamma_j'] < 1 \text{ for all } j \le k \right\} \right) \right).$$

By definition, $1 \le k(n) \le n$. Moreover, since for every $j \in \mathbb{N}$ we have $E[\gamma_j'] < \infty$ by Lemma 3.4 and $\tau_n \to 0$ by definition, we have that $k(n) \to \infty$ as $n \to \infty$. Thus, defining $\gamma_n = \gamma_{k(n)}'$, for large enough $n \in \mathbb{N}$ we have

$$\mathcal{C}_n[\gamma_n] = C_0[\gamma_{k(n)}'] + \tau_n E[\gamma_{k(n)}'] < C_0[\gamma_{k(n)}'] + \sqrt{\tau_n}.$$

Recalling that by Lemma 3.3 we have $C_0[\gamma_{k(n)}'] \to C_0[\gamma]$, we conclude the proof.

References

[1] M. Agueh and G. Carlier, *Barycenters in the Wasserstein space*, SIAM J. on Mathematical Analysis, 43 (2011), pp. 904–924.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, vol. 254, Clarendon Press Oxford, 2000.

[3] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, *Iterative Bregman projections for regularized transportation problems*, SIAM J. Sci. Comput., 37 (2015), pp. A1111–A1138.

[4] J.-D. Benamou, G. Carlier, and L. Nenna, *A numerical method to solve optimal transport problems with Coulomb cost*, to appear as a chapter in "Splitting Methods in Communication and Imaging, Science and Engineering", Springer, (2015).

[5] ———, *Generalized incompressible flows, multi-marginal transport and sinkhorn algorithm*, arXiv preprint arXiv:1710.08234, (2017).

[6] G. Buttazzo, T. Champion, and L. De Pascale, *Continuity and estimates for multimarginal optimal transportation problems with singular costs*, Applied Mathematics & Optimization, (2016), pp. 1–16.

[7] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi, *Optimal-transport formulation of electronic Density Functional Theory*, Physical Review A, 85 (2012), p. 062502.

[8] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM J. Math. Anal., 49 (2017), pp. 1385–1418.

[9] C. Cotar, G. Friesecke, and C. Klüppelberg, *Density Functional Theory and optimal transportation with Coulomb cost*, Comm. Pure Appl. Math., 66 (2013), pp. 548–599.

[10] ———, *Smoothing of transport plans with fixed marginals and rigorous semiclassical limit of the Hohenberg-Kohn functional*, arXiv:1706.05676, (2017).

[11] M. Cuturi, in Adv. in Neural Information Processing Sys., 2013, pp. 2292–2300.

[12] ———, *Sinkhorn distances: Lightspeed computation of optimal transport*, in Advances in neural information processing systems, 2013, pp. 2292–2300.

[13] S. Di Marino, L. De Pascale, and M. Colombo, *Multimarginal optimal transport maps for* 1-*dimensional repulsive costs*, Canadian Journal of Mathematics – Journal Canadien des Mathématiques, 67 (2015), pp. 350–368.

[14] S. Di Marino, A. Gerolin, and L. Nenna, *Optimal transport theory for repulsive costs*, Topological Optimization and Optimal Transport: In the Applied Sciences, 17 (2017).

[15] R. Flamary and N. Courty, *POT Python Optimal Transport library*, 2017.

[16] A. Gerolin, A. Kausamo, and T. Rajala, *Duality theory for multimarginal optimal transport with repulsive costs in metric spaces*, ESAIM Control Optim. Calc. Var., (to appear), (2018).

[17] N. Gigli and L. Tamanini, *Second order differentiation formula on compact $RCD^*(K,N)$ spaces*, arXiv:1701.03932, (2017).

[18] N. Gozlan and C. Léonard, *Transport inequalities. a survey*, Markov Processes and Related Fields, 16 (2010), pp. 635–736.

[19] H. G. Kellerer, *Duality theorems for marginal problems*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 67 (1984).

[20] C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst., (2014), pp. 1533–1574.

[21] E. H. Lieb, *Density functionals for Coulomb systems*, in Inequalities, Springer, 2002, pp. 269–303.

[22] L. Nenna, *Numerical Methods for Multi-Marginal Opimal Transportation*, PhD thesis, Université Paris-Dauphine, 2016.

[23] G. Peyré and M. Cuturi, *Computational Optimal Transport*, 2018.

[24] E. Schrödinger, *Über die umkehrung der naturgesetze*, Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u. Company, 1931.

[25] M. Seidl, S. Di Marino, A. Gerolin, K. Giesbertz, L. Nenna, and P. Gori-Giorgi, *The Strictly-Correlated Electron functional for spherically symmetric systems revisited*, Accepted in Physical Review A (available in arXiv:1702.05022), (2016).

University of Jyvaskyla, Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of Jyvaskyla, Finland
E-mail address: augustogerolin@gmail.com
E-mail address: anna.m.kausamo@jyu.fi
E-mail address: tapio.m.rajala@jyu.fi

[**D**]

Multi-marginal L^∞-transport

A. Kausamo

Preprint

MULTI-MARGINAL L^∞-TRANSPORT

ANNA KAUSAMO

Abstract. In this paper, we introduce the concept of multi-marginal optimal L^∞-transport on Polish spaces, and compare it with the two-marginal case. We state and prove Kantorovich duality in our setting, and establish L^∞ lower bounds on multi-marginal optimal transportation costs. We also state a version of cyclical monotonicity and restrictability in the multi-marginal L^∞ case, and study how and when optimal multi-marginal transportation costs can be evaluated from below by L^∞-distances.

1. Introduction

Let a Polish space (X, d), a cost function $c : X^2 \to \mathbb{R}$, and two Borel probability measures μ, ν be given. In the Optimal L^∞ Transport Problem one seeks for minimizing the quantity

$$C_\infty(\lambda) := \lambda - \operatorname*{ess\,sup}_{(x,y)\in X\times Y} c(x, y)$$

among all couplings λ of the measures μ and ν, that is, over the set of all probability measures on $X \times X$ the first projection of which is μ and the second projection is ν. The minimizer for this problem exists under mild assumptions on c such as lower-semicontinuity. In [6], Champion, De Pascale, and Juutinen carried out a comprehensive study on the L^∞ transport in the case where the cost of transporting a point x to a point y is given by their distance: $c(x, y) = d(x, y)$. They studied the problem on compact sets of $\mathbb{R}^n$, $n \geq 1$, and their work was generalized in 2015 by Jylhä [5] to Polish spaces with more general costs c. Champion, De Pascale, and Juutinen introduced the concept of L^∞-cyclical monotonicity, which carries the well-known notion of cyclical monotonicity of optimal transportation plans to the L^∞ case. In general, optimal L^∞-transportation plans are not as well-behaved as the minimizers of the standard integral Monge-Kantorovich problem. For instance, their restrictions are not always optimal with respect to their marginals. To tackle this, Champion, De Pascale, and Juutinen introduced the concept of *restrictability*: an optimal L^∞ transportation plan is restrictable if, loosely speaking, its restrictions are also optimal. Moreover, they showed that restrictability and L^∞ cyclical monotonicity are equivalent. This equivalence also holds for more general cost functions at least if they are continuous, as was proven by Jylhä in [5].

If $c = d$, the minimal value $C_\infty(\lambda)$ is called ∞-Wasserstein distance of the measures μ and ν and denoted by $W_\infty(\mu, \nu)$. There is a reason behind this notation: for $c = d$, the quantity $W_\infty(\mu, \nu)$ is actually the $p \to \infty$ limit of the p-Wasserstein distances $W_p(\mu, \nu)$ of the measures μ and ν. Bouchitté, Jimenez, and Rajesh [2], and Jylhä and Rajala [7] studied this ∞-Wasserstein distance and its connections to the p-Wasserstein convergenge of Borel probability measures. Jylhä and Rajala also established necessary and succifient conditions on the existence of W_∞ lower bounds for optimal transportation costs.

Date: November 26, 2018.

The author acknowledges the support of the Academy of Finland, projects no. 312488 and 314789.

The question on whether there exists a dual formulation for the L^∞-transport, similar to the now-standard Kantorovich duality, remained open until Barron, Bocea, and Jensen stated and proved a duality theorem in 2017 [1].

In this paper we generalize the concept of L^∞ transport to multi-marginal case. This paper is constructed as follows: In Section 2 we state basic definitions and properties. In Section 3 the L^∞ duality is considered. Section 4 studies multimarginal L^∞-transportation plans as limits of L^p-transportation plans, which leads us to the existence of L^∞-cyclically monotone plans. In Section 5 we consider L^∞-lower bounds on optimal transportation cost in the multi-marginal case.

2. Preliminaries

Let (X, d) be a complete, separable metric space, $N \geq 2$, and $\mu_1, \ldots, \mu_N$ Borel probability measures on X. Let $c : X^N \to [0, \infty]$ be a lower-semicontinuous cost function. In the multi-marginal L^∞ transportation problem one aims at mimimizing the quantity

$$C_\infty(\lambda) := \lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N)\in X^N} c(x_1, \ldots, x_N)$$

over all couplings λ of the marginals $\mu_1, \ldots, \mu_N$, that is, over the set

$$\Pi_N(\mu_1, \ldots, \mu_N) := \left\{\lambda \in \mathscr{P}(X^N) \mid (\mathrm{pr}_i)_\sharp \lambda = \mu_i \text{ for all } i \in \{1, \ldots, N\}\right\} ,$$

where pr_i is the projection on the i-th coordinate

$$\mathrm{pr}_i(x_1, \ldots, x_N) = x_i \quad \text{for all } (x_1, \ldots, x_N) \in X^N .$$

We call this minimization problem (P_∞) and the cost functional C_∞ if there is no confusion about what cost function we are referring to. If different cost functions are considered, we refer to the minimization with respect to the cost c as $(P_\infty(c))$ and to the cost functional as $C_{\infty,c}$. The marginal measures should be clear from the context. The existence of minimizers in L^∞ transportation problem for $N = 2$ was proven in [5]. The proof also works for $N \geq 2$, and thus we omit the details of the proof.

Proposition 2.1. *The cost C_∞ has a minimizer in the set $\Pi_N(\mu_1, \ldots, \mu_N)$.*

Proof. The claim follows from the lower-semicontinuity of the functional C_∞ and the compactness of the set $\Pi_N(\mu_1, \ldots, \mu_N)$. □

Some of the cost functions we are going to consider in this paper are functions of pairwise distances $d(x_i, x_j)$ between points $x_i, x_j \in X$. To this class belongs a cost that we are going to consider in more detail in Section 5. We denote this cost by $\max_d : X^N \to [0, \infty[$ and define

$$\max_d(x_1, \ldots, x_N) = \max\left\{d(x_i, x_j) \mid i, j \in \{1, \ldots, N\}\right\} \quad \text{for all } (x_1, \ldots, x_N) \in X^N .$$

Remark 2.2. *Instead of using $\max_d$ as the cost function, we could take the essential supremum over sums $\sum_{1\leq i<j\leq N} d(x_i, x_j)$ and consider the cost*

$$C_{\infty,s}(\lambda) := \lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N)\in X^N} \sum_{1\leq i<j\leq N} d(x_i, x_j) \quad \textit{for all } \lambda \in \Pi_N(\mu_1, \ldots, \mu_N) .$$

This form of a cost function often appears in multi-marginal settings. The functionals $C_{\infty,S}$ and C_∞ are related by

$$C_\infty(\gamma) \le C_{\infty,S}(\gamma) \le \frac{N(N-1)}{2} C_\infty(\gamma) \text{ for all } \gamma \in \mathscr{P}(X^N)\,.$$

For $N = 2$, the costs coincide.

Since the function $\max_d$ is continuous, the existence of at least one minimizer for the functional $C_{\infty,\max_d}$ follows from Proposition 2.1.

Let us denote, mimicking the two-marginal case, the minimal $C_{\infty,\max_d}(\lambda)$ by $W_\infty(\mu_1,\dots,\mu_N)$.

2.1. **Cyclical monotonicity.** In the two-marginal case, we know by Theorem 5.10 in [10] that optimality of a transportation plan is equivalent to its support being c-cyclically monotone. In the multi-marginal case, there exists a definition, analogous to the c-cyclical monotonicity, here taken from [4]:

Definition 2.3. *Let us consider the multi-marginal OT problem with respect to a cost function c. We say that a set $\Gamma \subset X^N$ is c-cyclically monotone if for any $k \in \mathbb{N}$ and for any k-tuple of points*

$$(x_1^{(1)},\dots,x_N^{(1)}),\ \dots,\ (x_1^{(k)},\dots,x_N^{(k)}) \in \Gamma$$

and permutations $\sigma_2,\dots,\sigma_N : \{1,\dots,k\} \to \{1,\dots,k\}$ one has

$$\sum_{i=1}^k c(x_1^{(i)},\dots,x_N^{(i)}) \le \sum_{i=1}^k c(x_1^{(i)}, x_2^{(\sigma_2(i))},\dots,x_N^{(\sigma_N(i))})\,.$$

Kim and Pass proved in [9] that, in the multi-marginal case, c-cyclical monotonicity is a necessary condition for optimality. Their proof was stated for the underlying spaces being Riemannian manifolds, but the argument does not use Riemannian structure so we may assume that the space(s) are Polish. Griessler showed in [4] that c-cyclical monotonicity is also a sufficient condition for optimality, under the assumption that the cost function is bounded from above by a sum of integrable functions defined on the marginals. Therefore, we get as a corollary

Corollary 2.4. *Let $c : X^N \to [0,\infty[$ be a continuous cost function satisfying, for some $u_i \in L^1_{\mu_i}(X)$, $i \in 1,\dots,N$, the condition*

$$c(x_1,\dots,x_N) \le \sum_{i=1}^N u_i(x_i) \quad \text{for } \mu_1\times\cdots\times\mu_N\text{-a.e. } (x_1,\dots,x_N) \in X^N\,. \tag{2.1}$$

Let $\lambda \in \Pi_N(\mu_1,\dots,\mu_N)$. Then λ is a minimizer of the multi-marginal OT problem

$$\inf_{\lambda' \in \Pi_N(\mu_1,\dots,\mu_N)} \int_{X^N} c\,\mathrm{d}\lambda'\,,$$

if and only if $\operatorname{supp}\lambda$ *is c-cyclically monotone in the sense of Definition 2.3.*

Remark 2.5. *The integrability condition* (2.1) *is needed only for the proof of the fact that c-cyclical monotonicity is sufficient for optimality.*

3. Kantorovich Duality

In this section we formulate and prove the Kantorovich Duality formula for L^∞-MOT problem. The duality was proven in the two-marginal case by Barron, Bocea, and Jensen in [1]. Here we assume that $c : X^N \to [0, +\infty]$ is a lower-semicontinous cost function, and consider the infimum

$$I(\mu_1, \dots, \mu_N) := \inf_{\lambda \in \Pi_N(\mu_1, \dots, \mu_N)} \lambda - \operatorname*{ess\,sup}_{(x_1, \dots, x_N) \in X^N} c(x_1, \dots, x_N)\,.$$

The infimum on the left is sometimes abbreviated as I if there is no danger of confusion. Before stating the duality, we introduce some notation. We denote, for all $a \geq 0$,

$$\begin{aligned}
S_{c,a} &:= \left\{(x_1, \dots, x_N) \in X^N \mid c(x_1, \dots, x_N) \leq a\right\}\,,\\
\mathcal{F}_c &:= \left\{ (u_1, \dots, u_N) \in (C_b(X))^N \;\middle|\; \sum_{i=1}^N u_i(x_i) \leq c \text{ for all } (x_1, \dots, x_N) \in X^N \right\} \text{ and}\\
\mathcal{F}_{c,a} &:= \left\{ (u_1, \dots, u_N) \in (C_b(X))^N \;\middle|\; \sum_{i=1}^N u_i(x_i) \leq a \text{ for all } (x_1, \dots, x_N) \in S_{c,a} \right\}.
\end{aligned}$$

The duality statement is formulated and proven in the following theorem.

Theorem 3.1.

$$I(\mu_1, \dots, \mu_N) = \inf_{a \geq 0} \sup \left\{ \sum_{i=1}^N \int_X u_i \mathrm{d}\mu_i \;\middle|\; (u_1, \dots, u_N) \in \mathcal{F}_{c,a} \right\}. \tag{3.1}$$

Proof. The argument of Barron, Bocea, and Jensen works also here. We repeat it for completeness.

For each $a \geq 0$ we define a function $c_a : X^N \to [0, \infty]$ by setting

$$c_a(x_1, \dots, x_N) = \begin{cases} a & \text{if } c(x, y) \leq a \\ +\infty & \text{otherwise} \end{cases}.$$

Now c_a is lower semi-continuous and we know from Theorem 2.6 of [8] that the duality

$$\inf \left\{ \int_{X^N} c_a \mathrm{d}\lambda \;\middle|\; \lambda \in \Pi_N(\mu_1, \dots, \mu_N) \right\} = \sup \left\{ \sum_{i=1}^N \int_X u_i \mathrm{d}\mu_i \;\middle|\; (u_1, \dots, u_N) \in \mathcal{F}_{c_a} \right\} \tag{3.2}$$

holds. Now we have $(u_1, \dots, u_N) \in \mathcal{F}_{c_a}$ if and only if

$$\sum_i u_i(x_i) \leq c_a(x_1, \dots, x_N) \text{ for all } (x_1, \dots, x_N) \in X^N$$

if and only if

$$\sum_i u_i(x_i) \leq a \text{ for all } (x_1, \dots, x_N) \in S_{c,a}\,.$$

Hence the equation (3.2) can be written in the form

$$\inf \left\{ \int_{X^N} c_a \mathrm{d}\lambda \;\middle|\; \lambda \in \Pi_N(\mu_1, \dots, \mu_N) \right\} = \sup \left\{ \sum_{i=1}^N \int_X u_i \mathrm{d}\mu_i \;\middle|\; (u_1, \dots, u_N) \in \mathcal{F}_{c,a} \right\}$$

and it suffices to show that

$$I = \inf_{a \geq 0} I_a \,, \tag{3.3}$$

where

$$I_a := \inf \left\{ \int_{X^N} c_a \mathrm{d}\lambda \;\middle|\; \lambda \in \Pi_N(\mu_1, \ldots, \mu_N) \right\} .$$

Let us first show that

$$I \geq \inf_{a \geq 0} I_a \,. \tag{3.4}$$

It suffices to show that, for all $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ we have

$$\lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N) \in X^2} c(x_1, \ldots, x_N) \geq \inf_{a \geq 0} I_a \,.$$

So we fix a coupling λ of the marginals $\mu_1, \ldots, \mu_N$. Let us consider the choice

$$a := \lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N) \in X^N} c(x_1, \ldots, x_N) \,.$$

Now $a \geq 0$ and we may assume that it is finite; if $a = +\infty$, inequality (3.4) trivially holds. Now we have

$$\lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N) \in X^N} c(x_1, \ldots, x_N) = a \overset{a)}{=} \int_{X^N} c_a \mathrm{d}\lambda \geq \inf_a I_a \,,$$

where equality a) follows from the choice of a and the fact that λ is a probability measure. So we are done with the proof of inequality (3.4).

It remains to show that

$$I \leq \inf_{a \geq 0} I_a \,. \tag{3.5}$$

We assume, on the contrary, that

$$I > \inf_{a \geq 0} I_a \,.$$

In this case there exist $a > 0$ and $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ such that

$$\int_{X^N} c_a \mathrm{d}\lambda < I < \infty \,.$$

This means that $c_a \in L^1_\lambda$ and thus, by the definition of c_a we have

$$\lambda(\{(x_1, \ldots, x_N) \in X^N \mid c(x_1, \ldots, x_N) > a\}) = 0 \,.$$

Thus

$$\lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N) \in X^N} c(x_1, \ldots, x_N) \leq a = \int_{X^N} c_a \mathrm{d}\lambda < I \,.$$

This is impossible since $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$. Thus condition (3.5) holds and the proof is complete. $\square$

Remark 3.2. *Similarly to the two-marginal case* [1]*, we have that the choice $a = I$ realizes the infimum in* (3.1) *and thus the duality can be written in the form*

$$I(\mu_1, \ldots, \mu_N) = \sup \left\{ \sum_{i=1}^N \int_X u_i \mathrm{d}\mu_i \;\middle|\; (u_1, \ldots, u_N) \in \mathcal{F}_{c, I(\mu_1, \ldots, \mu_N)} \right\} . \tag{3.6}$$

Furthermore, if $I < \infty$, the functions

$$u_1 = u_2 = \cdots = u_N \equiv \frac{1}{N} I(\mu_1, \ldots, \mu_N)$$

trivially maximize the right-hand side of (3.6) *so the L^∞-duality can be written in the form*

$$I(\mu_1, \ldots, \mu_N) = \max \left\{ \sum_{i=1}^{N} \int_X u_i \mathrm{d}\mu_i \;\middle|\; (u_1, \ldots, u_N) \in \mathcal{F}_{c, I(\mu_1, \ldots, \mu_N)} \right\}. \tag{3.7}$$

Let us call any N-tuple of functions $(u_1, \ldots, u_N)$ that realizes the maximum above ***a Kantorovich solution*** *to problem P_∞.*

Remark 3.3. *It is equivalent to, instead of maximizing over the class $\mathcal{F}_{c,I}$, maximize over the class of functions $(u_1, \ldots, u_N) \in L^1_{\mu_1} \times \cdots \times L^1_{\mu_N}$ for which the condition*

$$\sum_{i=1}^{N} u_i(x_i) \leq I(\mu_1, \ldots, \mu_N) \text{ whenever } c(x_1, \ldots, x_N) \leq I(\mu_1, \ldots, \mu_N)$$

holds for $\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_N$ almost every $(x_1, \ldots, x_N) \in X^N$.

We can now prove results that are similar to standard optimal transportation results concerning Kantorovich potentials.

Proposition 3.4. *Assume that $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ is an optimal coupling. Then the following conditions are equivalent:*

(i) *$(u_1, \ldots, u_N)$ is a Kantorovich solution.*
(ii) *It holds*

$$\sum_{i=1}^{N} u_i(x_i) \leq c(x_1, \ldots, x_N) \text{ for all } (x_1, \ldots, x_N) \in S_{I(\mu_1, \ldots, \mu_N)} \text{ and}$$
$$\sum_{i=1}^{N} u_i(x_i) = I(\mu_1, \ldots, \mu_N) \text{ for } \lambda - a.e. \ (x_1, \ldots, x_N) \in X^N.$$

Proof. In proving that (i) implies (ii) the two-marginal proof works also here; see Proposition 2.8 of [1].

The reverse direction follows directly from the definitions and the fact that μ_i:s are probability measures. □

It is difficult to find non-trivial Kantorovich solutions. Luigi De Pascale and Jean Louet studied in [3] the problem in the two-marginal case on $\mathbb{R}^d$ (inparticular in $\mathbb{R}$) for the cost $c(x, y) = |x - y|$. They proved that the Kantorovich potentials are actually locally constant, when restricted on the support of their respective marginal measures, in points that are moved, in an optimal transportation plan, a distance less than the maximum distance. The same holds in the multi-marginal case, and also for more general cost functions, provided that they are continuous.

Proposition 3.5. *Assume that $c : X^N \to [0, \infty[$ is a continuous cost function. Let $(u_1, \ldots, u_N) \in C_b(X)^N$ be a Kantorovich solution and $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ be an optimal coupling. Then the* *measure μ_1 is concentrated on a set L such that :*

If for a point $\overline{x}_1 \in L$ we have

$$\sup_{\substack{(x_2,\ldots,x_N)\in X^{N-1}\\ (x_1,x_2,\ldots,x_N)\in \operatorname{supp}\lambda}} c(\overline{x}_1, x_2, \ldots, x_N) = C < I \text{ for some } C \in \mathbb{R}\,, \tag{3.8}$$

then there exists a constant $\epsilon > 0$ such that u_1 is constant on the set $B(\overline{x}_1, \epsilon) \cap L$.

Proof. Since $(u_1, \ldots, u_N)$ is a Kantorovich solution, we may fix a set $\Gamma \subset X^N$ such that λ is concentrated on Γ and

$$u_1(x_1) + u_2(x_2) + \cdots + u_N(x_N) = I \quad \text{for all } (x_1, \ldots, x_N) \in \Gamma\,.$$

We set $L := \mathrm{pr}_1(\Gamma)$, and show that this set satisfies the property (3.8). Towards this end we fix $\overline{x}_1 \in L$ such that, for some $C \in \mathbb{R}$,

$$\sup_{\substack{(x_2,\ldots,x_N)\in X^{N-1}\\ (x_1,x_2,\ldots,x_N)\in \operatorname{supp}\lambda}} c(\overline{x}_1, x_2, \ldots, x_N) = C < I\,.$$

Because of the continuity of c and the fact that $\operatorname{supp}\lambda$ is a closed set we may fix an $\epsilon > 0$ such that

$$\sup_{\substack{(x_2,\ldots,x_N)\in X^{N-1}\\ (x_1,x_2,\ldots,x_N)\in \operatorname{supp}\lambda}} c(x_1, \ldots, x_N) < C + \frac{I - C}{2} \text{ for all } x_1 \in B(\overline{x}_1, \epsilon)\,. \tag{3.9}$$

We fix $x_1 \in B(\overline{x}_1, \epsilon) \cap L$. It suffices to show that

$$u_1(x_1) \le u_1(\overline{x}_1) \text{ and} \tag{3.10}$$

$$u_1(\overline{x}_1) \le u_1(x_1)\,. \tag{3.11}$$

Let us start with the first of these claims. Since $\overline{x}_1 \in L = \mathrm{pr}_1(\Gamma)$, there exist points $\overline{x}_2, \ldots, \overline{x}_N$ such that

$$(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_N) \in \Gamma$$

and thus

$$u_1(\overline{x}_1) + u_2(\overline{x}_2) + \cdots + u_N(\overline{x}_N) = I\,. \tag{3.12}$$

Since $x_1 \in B(\overline{x}_1, \epsilon)$ we have, by Condition (3.9) that

$$c(x_1, \overline{x}_2, \ldots, \overline{x}_N) < I\,,$$

and thus, since $(u_1, \ldots, u_N)$ is a Kantorovich solution,

$$u_1(x_1) + u_2(\overline{x}_2) + \cdots + u_N(\overline{x}_N) \le I\,,$$

that is,

$$u_1(x_1) \le I - \sum_{j=2}^{N} u_i(\overline{x}_i) = u_1(\overline{x}_1)\,,$$

where in the last equality we have used Condition (3.12). This proves the inequality (3.10).

The proof of the inequality (3.11) is similar: Because $x_1 \in B(\overline{x}_1, \epsilon) \cap L$, we in particular have $x_1 \in L$. Thus there exist points $x_2, \ldots, x_N$ such that

$$(x_1, \ldots, x_N) \in \Gamma\,.$$

Now we have

$$c(\overline{x}_1, x_2, \ldots, x_N) \le I$$

and then

$$u_1(\overline{x}_1) + u_2(x_2) + \cdots u_N(x_N) \leq I$$

and thus

$$u_1(\overline{x}_1) \leq I - \sum_{j=2}^{N} u_j(x_j) = u_1(x_1)\,. \qquad \square$$

In proving that c-cyclical monotonicity of a transport plan is sufficient for its optimality, Griessler [4] used the concept of splitting sets. The concept can be brought to the multi-marginal L^∞ context:

Definition 3.6. *We say that a set $\Gamma \subset X^N$ is ∞-splitting if there exist functions $(u_1, \ldots, u_N) \in L^2_{\mu_1} \times \cdots \times L^1_{\mu_N}$ such that*

$$u_1(x_1) + \cdots + u_N(x_N) \leq I(\mu_1, \ldots, \mu_N) \text{ for all } (x_1, \ldots, x_N) \in S_{c,I(\mu_1,\ldots,\mu_N)} \text{ and}$$
$$u_1(x_1) + \cdots + u_N(x_N) = I(\mu_1, \ldots, \mu_N) \text{ for all } (x_1, \ldots, x_N) \in \Gamma\,.$$

We call such $(u_1, \ldots, u_N)$ a $(\Gamma, c)_\infty$-splitting tuple.

We immediately get as a corollary to the Kantorovich duality

Corollary 3.7. *If a coupling $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ is optimal then it is concentrated on a ∞-splitting set Γ*

In the L^∞-case, the concept of splitting sets may not be not very fruitful since we cannot use it to prove optimality results: for *any* coupling λ we can use the trivial solution to the Kantorovich problem to form a splitting tuple. So in the L^∞-case, optimality does not follow from 'being concentrated on a splitting set'.

4. Optimal L^∞ plans as limits of L^p plans

In this section, we formulate and proof results that describe L^∞ transport plans as limits of L^p plans when $p \to \infty$.

Here again $c : X^N \to [0, \infty]$ is a lower-semicontinuous cost function. Let us denote, for each $p \geq 1$ and $\lambda \in \mathscr{P}(X^N)$,

$$C_p(\lambda) := \left(\int_{X^N} c(x_1, \ldots, x_N)^p \mathrm{d}\lambda \right)^{\frac{1}{p}} \text{ and}$$
$$C_\infty(\lambda) = \lambda - \operatorname*{ess\,sup}_{(x_1,\ldots,x_N)} c(x_1, \ldots, x_N)\,.$$

Now we have

Proposition 4.1. *Let us assume that $(\lambda_p)_{p=1}^\infty$ is sequence of minimzers of the costs C_p in $\Pi_N(\mu_1, \ldots, \mu_N)$. Then this sequence converges, up to passing to a subsequence, to a measure λ that is a minimizer of C_∞ in $\Pi_N(\mu_1, \ldots, \mu_N)$.*

Proof. The existence of a converging subsequence follows from the fact that the set $\Pi_N(\mu_1, \ldots, \mu_N)$ is compact in the weak* topology.[8] Without change of notation, let us assume that $\lambda_p \rightharpoonup \lambda$ for some $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$.

Let us show that λ minimizes C_∞. For this, we fix $\lambda' \in \Pi_N(\mu_1, \ldots, \mu_N)$. We have to show that

$$C_\infty(\lambda) \leq C_\infty(\lambda')\,.$$

Since $C_\infty(\lambda) = \lim_{q\to\infty} C_q(\lambda)$, it suffices to show that, for every $q \geq 1$ we have

$$C_q(\lambda) \leq C_\infty(\lambda') \,. \tag{4.1}$$

For this, we fix $q \geq 1$. Since $\lambda_p \rightharpoonup \lambda$ we have, as C_q is lower semicontinuous, the inequality

$$C_q(\lambda) \leq \liminf_{p\to\infty} C_q(\lambda_p) \,.$$

Thus, to prove condition (4.1) it suffices to show that for all $p \geq q$ we have

$$C_q(\lambda_p) \leq C_\infty(\lambda') \,. \tag{4.2}$$

We fix $p \geq q$. Using Hölder's inequality one gets

$$\int_{X^N} c^q \mathrm{d}\lambda_p \leq \left(\int_{X^N} (c^q)^{\frac{p}{q}} \mathrm{d}\lambda_p\right)^{\frac{q}{p}} = \left(\int_{X^N} c^p \mathrm{d}\lambda_p\right)^{\frac{q}{p}} ,$$

and therefore

$$C_q(\lambda_p) = \left(\int_{X^N} c^q \mathrm{d}\lambda_p\right)^{\frac{1}{q}} \leq \left(\int_{X^N} c^p \mathrm{d}\lambda_p\right)^{\frac{1}{p}} = C_p(\lambda_p) \,.$$

Since λ_p is optimal for C_p, we also have

$$C_p(\lambda_p) \leq C_p(\lambda') \,.$$

Combining these two estimates gives

$$C_q(\lambda_p) \leq C_p(\lambda') \,.$$

Yet we also have (by Hölder's inequality and passing to the limit $p \to \infty$)

$$C_p(\lambda') \leq C_\infty(\lambda') \,,$$

so condition (4.2) follows, and we are done. □

In [5], it was proven that the weak*-limits of mimizers of C_p for continuous cost functions c satisfy the so called infinite c-cyclical monotonicity property. A similar result also holds in the multi-marginal case. First we introduce the generalized definition

Definition 4.2. *We say that a set $\Gamma \subset X^N$ is infinitely c-cyclically monotone (ICM), if for every k-tuple of points $(x_1^i, \dots, x_N^i)_{i=1}^k$ and every $(N-1)$-tuple of permuations $(\sigma_2, \dots, \sigma_N)_{i=1}^k$ of the set $\{1, \dots, k\}$ we have*

$$\max\{c(x_1^i, x_2, \dots, x_N^i) \mid i \in \{1, \dots, k\}\} \leq \max\{c(x_1^i, x_2^{\sigma_2(i)}), \dots, x_N^{\sigma_N(i)}) \mid i \in \{1, \dots, k\}\} \,.$$

We also say that a coupling $\lambda \in \Pi_N(\mu_1, \dots, \mu_N)$ is infinitely c-cyclically monotone if it is concentrated on an ICM set.

We know by [9] that the mimizers of C_p are c-cyclically monotone in the sense of Definition 2.3 if the cost function c is continous. Now, as in the two-marginal case, the weak*-limits of these minimizers are ICM:

Proposition 4.3. *Let us assume that $c : X^N \to [0, \infty]$ is continuous. Let $(\lambda_p)_{p=1}^\infty$ be a sequence of minimizers of the cost C_p, and let us assume that $C_p(\lambda_p)$ is finite for every p. Then their weak* limit $\lambda := \lim_{p\to\infty} \lambda_p$ is ICM in the sense of Definition 4.2.*

Proof. We fix $k \in \mathbb{N}$, points $\{(x_1^i, \dots, x_N^i)\}_{i=1}^K \subset \operatorname{supp} \lambda$ and permutations $\sigma_2, \dots, \sigma_N$. We have to show that

$$\max\{c(x_1^i, x_2^i, \dots, x_N^i) \mid i \in \{1, \dots, k\}\} \leq \max\{c(x_1^i, x_2^{\sigma_2(i)}, \dots, x_N^{\sigma_N(i)}) \mid i \in \{1, \dots, k\}\}.$$

As minimizers of the costs C_p, the couplings λ_p are c^p-cyclically monotone, and thus we have

$$\sum_{i=1}^n c^p(x_1^{(i)}, \dots, x_N^{(i)}) \leq \sum_{i=1}^n c^p(x_1^{(i)}, x_2^{(\sigma_2(i))}, \dots, x_N^{(\sigma_N(i))}).$$

The claim follows by taking the p:th root from both sides and passing to the limit $p \to \infty$. □

With similar considerations to those in [5], one can prove the existence of ICM couplings for continuous cost functions. Here is the statement, we do not repeat the proof here.

Proposition 4.4. *If c is continuous, then the cost C_∞ admits a ICM minimizer.*

This leads us to the question on whether all ICM plans are optimal. The answer is affirmative at least if c is continuous:

Proposition 4.5. *Assume that c is a continuous cost function. And $\gamma \in \Pi_N(\mu_1, \dots, \mu_N)$ is an ICM coupling. Then γ solves (P_∞).*

Proof. We adapt the argument in [5] to our setting. Assume that γ is ICM. Claim γ is optimal. Assume on the contrary that there exists λ such that

$$\gamma - \operatorname{ess\,sup} c > \lambda - \operatorname{ess\,sup} c + \epsilon.$$

We fix σ-compact sets $\Gamma_\gamma \subset X^N$ and $\Gamma_\lambda \subset X^N$ such that

$$\begin{aligned} &\gamma(X^N \setminus \Gamma_\gamma) = 0, \\ &\lambda(X^N \setminus \Gamma_\lambda) = 0 \text{ and} \\ &\operatorname{pr}_i(\Gamma_\lambda) = \operatorname{pr}_i(\Gamma_\gamma) \text{ for all } i \in \{1, \dots, N\}. \end{aligned}$$

We define

$$\widetilde{A}_1 := \{z \in \Gamma_\gamma \mid c(z) > \lambda - \operatorname{ess\,sup} c + \epsilon\}.$$

This is σ-compact and, because of our antithesis, has positive γ-measure. Then we project $\widetilde{A}_1$ on Γ_λ by defining

$$\widetilde{A}_0 := \{(x_1, \dots, x_N) \in \Gamma_\lambda \mid x \in \operatorname{pr}_1(\widetilde{A}_1)\} = \left(\operatorname{pr}_1(\widetilde{A}_1) \times X^{N-1}\right) \cap \Gamma_\lambda.$$

This is a σ-compact set since Γ_λ is. It also has a positive λ-measure since, due to the fact that γ and λ have the same marginals, we have

$$\lambda(\operatorname{pr}_1(\widetilde{A}_1) \times X^{N-1}) = \mu_1(\operatorname{pr}_1(\widetilde{A}_1) = \gamma(\operatorname{pr}_1(\widetilde{A}_1) \times X^{N-1}) > 0.$$

Since

$$\left(\operatorname{pr}_1(\widetilde{A}_0) \times X^{N-1}\right) \cap \widetilde{A}_1 = \widetilde{A}_1 \cap \left(\operatorname{pr}_1(\Gamma_\lambda) \times X^{N-1}\right) = \widetilde{A}_1 \cap \left(\operatorname{pr}_1(\Gamma_\gamma) \times X^{N-1}\right),$$

we also have

$$\gamma((\operatorname{pr}_1(\widetilde{A}_0) \times X^{N-1}) \cap \widetilde{A}_1 > 0. \tag{4.3}$$

We may assume, by possibly removing a set of λ-measure zero, that

$$\lambda(B(z, r)) > 0 \quad \text{for all } z \in \widetilde{A}_0 \text{ and } r > 0. \tag{4.4}$$

Because of the continuity of c we can, for each $w \in \overline{A}_0$, fix a 'radius' $\delta_w > 0$ such that

$$c(z) < c(w) + \epsilon \leq \lambda - \operatorname{ess\,sup} c + \epsilon \text{ for all } z \in B_w \,, \tag{4.5}$$

where we have denoted

$$B_w := B(x_1, \delta_w) \times B(x_2, \delta_w) \times \cdots \times B(x_N, \delta_w)\,.$$

Since the set $\widetilde{A}_0$ is σ-compact, we can fix a finite cover $\{B_i\}_{i=1}^N$ for it, consisting of sets B_w, $w \in \widetilde{A}_0$. Since the sets B_i form a cover of $\widetilde{A}_0$, we have

$$\bigcup_{i \in \mathbb{N}} \Big(\mathrm{pr}_1(B_i \cap \widetilde{A}_0) \times X^{N-1}) \cap \widetilde{A}_1 \Big) = \Big(\mathrm{pr}_1(\widetilde{A}_0) \times X^{N-1}) \Big) \cap \widetilde{A}_1\,.$$

Hence by Condition (4.3) there exists $i_0 \in \mathbb{N}$ such that

$$\gamma(\mathrm{pr}_1(B_{i_0} \cap \widetilde{A}_0)) \times X^{N-1}) > 0\,.$$

Now we define $A_0 = B_{i_0} \cap \widetilde{A_0}$. Due to Condition (4.4) we have $\lambda(A_0) > 0$. We continue by defining

$$A_1 := \{(x_1, \ldots, x_N) \in \widetilde{A}_1 \mid x_1 \in \mathrm{pr}_1(A_0)\} = (\mathrm{pr}_1(A_0) \times X^{N-1}) \cap \widetilde{A}_1$$

This again is a σ-compact set and has, due to the choice of i_0, a positive γ-measure. We define for all $j \geq 1$

$$\begin{aligned} A_{2j} &:= \big(X \times (\mathrm{pr}_2(A_{2j-1}) \times X^{N-2} \big) \cap \Gamma_\lambda \text{ and} \\ A_{2j+1} &:= \big(\mathrm{pr}_1(A_{2j}) \times X^{N-1} \big) \cap \Gamma_\gamma\,. \end{aligned}$$

and

$$D := \bigcup_{j \geq 1} \mathrm{pr}_1(A_{2j+1}) \text{ and } E := \bigcup_{j \geq 1} \mathrm{pr}_2(A_{2j})\,.$$

Let us show that

$$\mathrm{pr}_2(A_0) \cap E \neq \emptyset\,. \tag{4.6}$$

By the definitions of the sets Γ_γ, Γ_λ, and A_j we have

$$\text{If } x_1 \in D \text{ and } (x_1, x_2, \ldots, x_N) \in \Gamma_\gamma \text{ then } x_2 \in E \text{ and} \tag{4.7}$$

$$\text{If } x_2 \in E \text{ and } (x_1, x_2, \ldots, x_N) \in \Gamma_\lambda \text{ then } x_1 \in E\,. \tag{4.8}$$

Indeed, if $x_1 \in D$ and

$$(x_1, x_2, \ldots, x_N) \in \Gamma_\gamma\,, \tag{4.9}$$

then there exists $j \geq 1$ such that $x_1 \in \mathrm{pr}_1(A_{2j+1})$ This means, by the definition of the set A_{2j+1}, in particular, that

$$x_1 \in \mathrm{pr}_1(A_{2j})\,. \tag{4.10}$$

Now we have $x_2 \in \mathrm{pr}_2(A_{2(j+1)}) \subset E$. To see this, we have to show that for some $x_1', x_3', \ldots, x_N'$ we have

$$(x_1', x_2, x_3', \ldots, x_N') \in \Gamma\lambda \text{ and} \tag{4.11}$$

$$x_2 \in \mathrm{pr}_2(A_{2j+1})\,. \tag{4.12}$$

Claim (4.11) holds since, by the assumption (4.9) we have

$$x_2 \in \mathrm{pr}_2(\Gamma_\gamma) = \mathrm{pr}_2(\Gamma_\lambda)\,,$$

so there exist points $x_1', x_3', \dots, x_N' \in X$ such that Claim (4.11) is satisfied. The second claim follows from the fact that, in fact

$$(x_1, x_2, \dots, x_N) \in A_{2j+1}\,.$$

This is a direct consequence of the Assumption (4.9) and Condition (4.10). Thus Property (4.7) holds. The proof of (4.8) is similar. The claim (4.6) can be proved by using Conditions (4.8) and (4.7) to show that for the choice $A := \mathrm{pr}_1(A_0) \cap D$ we have

$$\mu_1(A) > 0 \text{ and } \lambda(A \times (\mathrm{pr}_2(A_0) \cap E)) > 0\,.$$

The argument is identical to the two-marginal one in [5].

Now we are ready to construct the point cycle that will give the desired contradiction. Since

$$\mathrm{pr}_2(A_0) \cap E = \mathrm{pr}_2(B_{i_0} \cap \widetilde{A}_0) \cap A_0 \neq \emptyset\,,$$

we can fix a point

$$x_2^k \in \mathrm{pr}_2(A_0) \cap \mathrm{pr}_2(A_{2k}) \;\text{ for some } k \geq 1\,.$$

By the definition of the sets A_j we can then fix points

$$\begin{aligned}&(x_1^{j+1}, x_2^{j+1}, x_3^{j+1}, \dots, x_N^{j+1}) \in A_{2j+1} \text{ and}\\ &(x_1^{j+1}, x_2^{j}, x_3^{j+1}, \dots, x_N^{j+1}) \in A_{2j} \text{ for all } j \in \{1, \dots, k-1\}\,.\end{aligned}$$

Finally there exists a point

$$(x_1^1, x_2^1, \dots, x_N^1) \in A_1\,,$$

which can be considered as the completing point of the cycle since by the choice of the point x_2^k we have

$$(x_1^1, x_2^k, x_3^1, \dots, x_N^1) \in B_{i_0}\,. \tag{4.13}$$

By the definition of the set A_1 we have

$$\max\{c(x_1^j, x_2^j, \dots, x_N^j) \mid 1 \leq j \leq k\} > \lambda - \operatorname{ess\,sup} c + \epsilon\,. \tag{4.14}$$

From Conditions (4.13) and (4.5) we get that

$$c(x_1^1, x_2^k, x_3^1, \dots, x_N^1) < \lambda - \operatorname{ess\,sup} c + \epsilon\,.$$

We also have, since $A_{2j} \subset \Gamma_\lambda$, that

$$c(x_1^j, x_2^{j-1}, x_3^j, \dots, x_N^j) < \lambda - \operatorname{ess\,sup} c + \epsilon \text{ for all } j \in \{2, \dots, k\}\,.$$

Hence we get (by denoting $x_2^k := x_2^0$),

$$\max\{c(x_1^j, x_2^{j-1}, x_3^j, \dots, x_N^j) \mid 1 \leq j \leq k\} < \lambda - \operatorname{ess\,sup} c + \epsilon\,. \tag{4.15}$$

Combining Conditions (4.14) and (4.15) we get

$$\max\{c(x_1^j, x_2^{j-1}, x_3^j, \dots, x_N^j) \mid 1 \leq j \leq k\} < \max\{c(x_1^j, x_2^j, \dots, x_N^j) \mid 1 \leq j \leq k\}\,.$$

This contradicts the assumption that γ is ICM because, by the definitions of the sets A_1 and A_{2j+1}, we have

$$(x_1^j, x_2^j, x_3^j, \dots, x_N^j) \in \operatorname{supp} \gamma \;\text{ for all } j \in \{1, \dots, k\}\,. \qquad \square$$

4.1. **Restrictable solutions.** Like in the two-marginal cases, ICM solutions can be described by a property called *restrictability.* Loosely speaking, a restrictable solution to the problem (P_∞) remains optimal with respect to its marginals when restricted to a smaller set. In general, the problem (P_∞) can be considered quite crude since its solutions can present locally behavior that is far from what is intuitively considered optimal. In this sense, restrictable solutions are better-behaved.

Definition 4.6. *We say that a coupling $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ is a restrictable solution of the problem (P_∞) if the following condition holds: For all Borel measures $\bar{\lambda}$ that are absolutely continuous with respect to λ and satisfy $\bar{\lambda}(X^N) > 0$ we have*

$$C_\infty\left(\frac{\bar{\lambda}}{\bar{\lambda}(x^N)}\right) = \min\left\{C_\infty(\hat{\lambda}) \;\middle|\; \hat{\lambda} \in \Pi_N((\mathrm{pr}_1)_\sharp\bar{\lambda}, \ldots, (\mathrm{pr}_N)_\sharp\bar{\lambda}\right\} .$$

Here the measure $\bar{\lambda}$ is normalized only because we have defined C_∞ for elements of $\mathscr{P}(X^N)$.

Like in the two-marginal case, restrictability and ICM property are equivalent, at least if the cost function is continuous:

Proposition 4.7. *Assume that $c : X^N \to [0, \infty]$ is continuous and that $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$. Then λ is ICM if and only if it is restrictable.*

Proof. The proof is, for the most part, direct multi-marginalization of the proof of Theorem 2.19 in [5]. The only significant difference is in the proof of fact that restrictability implies infinite c-cyclical monotonicity. There one fixes points $\{(x_1^i, x_2^i \ldots, x_N^i)\}_{i=1}^k$ and $N-1$ permutations $\sigma_2, \ldots, \sigma_N$ of the set $\{1, \ldots, k\}$ to show that the ICM condition holds. In the proof of Theorem 2.19 in [5] it is assumed that, (using our notation)

$$x_1^i \neq x_1^j \text{ and } x_2^i \neq x_2^j \quad \text{for } i \neq j\,.$$

In the multi-marginal case, we cannot make this assumption. This leads to differences in normalization to handle the possible double-counting of measures of the some sets. □

5. Multi-marginal W_∞-bounds

In [7] Heikki Jylhä and Tapio Rajala answered, in the two-marginal case, the question: when can we evaluate optimal transportation costs from below by the W_∞-distance of the marginal measures. In this section, we state and prove the multi-marginal analogue of their Theorem 1.1:

Theorem 5.1. *Let $h : [0, \infty) \to [0, \infty)$ be a nondecreasing function with $h(t) > 0$ if $t > 0$. let us fix $\mu_1 \in \mathscr{P}(X)$. Then there exists a function $\omega : [0, \infty) \to [0, \infty)$ with $\omega(t) > 0$ for all $t > 0$ such that*

$$\inf_{\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)} \int_{X^N} \sum_{1 \le i < j \le N} h(d(x_i, x_j))\mathrm{d}\lambda \ge \omega(W_\infty(\mu_1, \ldots, \mu_N)) \tag{5.1}$$
$$\text{for all } \mu_2, \ldots, \mu_N \in \mathscr{P}(\operatorname{supp}\mu_1)\,.$$

if and only if $\operatorname{supp}\mu_1$ *is compact and connected. Moreover, in the inequality* (5.1), *we can take $\omega(t) = \frac{1}{2}m(\frac{t}{68})h(\frac{t}{68})$ where*

$$m(t) = \sup_{x \in X} \mu_1(B(x, t)) \text{ for all } t > 0\,.$$

For simplicity we denote the product of N copies of any set $A \subset X$ by $\otimes_N A$. We will need the following lemma:

Lemma 5.2. *Let (X,d) be complete metric space and $\mu_1 \in \mathscr{P}(X)$ with a δ-connected support for some $\delta \geq 0$. Assume also that*

$$m(t) := \inf_{x \in X} \mu_1(B(x,r)) > 0 \text{ for all } r > 0\,. \tag{5.2}$$

Let $\epsilon > 0$ and assume that there exist measures $\mu_2, \ldots, \mu_N \in \mathscr{P}(B(\operatorname{supp}\mu_1, \epsilon))$, a number $r > 0$, and a coupling $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ such that

$$\lambda(\{(x_1, \ldots, x_N) \in X^N \mid d(x_i, x_j) \geq r \text{ for some } i \neq j\}) < \frac{m(r)}{2}\,. \tag{5.3}$$

Then we have $W_\infty(\mu_1, \ldots, \mu_N) \leq 2(17r + 4\epsilon + \delta)$.

Proof. Let us also denote for all $i \in \{2, \ldots, N\}$ $\lambda_{1,i} := (\mathrm{pr}_1, \mathrm{pr}_i)_\sharp \lambda$. Now we have for all $i \in \{2, \ldots, N\}$ that $\lambda_{1,i} \in \Pi_2(\mu_1, \mu_i)$, and because of Assumption (5.3) and the fact that $\operatorname{supp}\mu_i \subset \operatorname{supp}\mu_1$ for $i \in \{2, \ldots, N\}$ we can apply Lemma 2.2 in [7] to get

$$W_\infty(\mu_1, \mu_i) \leq 17r + 4\epsilon + \delta \text{ for all } i \in \{2, \ldots, N\}\,.$$

Let us build a new measure $\lambda' \in \Pi_N(\mu_1, \ldots, \mu_N)$ as follows: Let us denote

$$T : X^2 \to X^2\,,\ T(x_1, x_2) = (x_2, x_1) \text{ for all } (x_1, x_2) \in X^2\,.$$

We start by gluing the measures $T_\sharp \lambda_{1,2} \in \Pi_2(\mu_2, \mu_1)$ and $\lambda_{1,3} \in \Pi_2(\mu_1, \mu_3)$ along their common marginal μ_1, resulting in a measure $\lambda_{2,1,3} \in \Pi_3(\mu_2, \mu_1, \mu_3)$. Then we form from $\lambda_{2,1,3}$ a measure $\lambda_{2,3,1} \in \Pi_3(\mu_2, \mu_3, \mu_1)$ by permuting the coordinates $(x_1, x_2, x_3) \mapsto (x_2, x_3, x_3)$. The resulting measure is glued to $\lambda_{1,4}$ along with the common marginal μ_1. This continues until we have glued the last measure $\lambda_{1,N}$. Then we take the final pushforward under the permutation of the coordinates that gives us a measure $\lambda' \in \Pi_N(\mu_1, \ldots, \mu_N)$ (the permutation is $(x_1, \ldots, x_{N-1}, x_N) \mapsto (x_{N-1}, x_1, \ldots, x_{N-2}, x_N)$).

For all $(x_1, \ldots, x_N) \in \operatorname{supp}\lambda'$ we have for $i, j \in \{1, \ldots, N\}$

$$d(x_i, x_j) \leq 17r + 4\epsilon + \delta$$

and thus for an arbitrary $i < j$, by applying triangle inequality $j - i - 1$ times, the estimate

$$d(x_i, x_j) \leq d(x_i, x_1) + d(x_1, x_j) \leq 2(17r + 4\epsilon + \delta) \leq N(17r + 4\epsilon + \delta)\,.$$

Now

$$W_\infty(\mu_1, \ldots, \mu_N) \leq C_\infty(\lambda') \leq 2(17r + 4\epsilon + \delta)\,. \qquad \square$$

There is another lemma, similar in nature to the two-marginal one in [7], that we need in order to prove the Theorem 5.1.

Lemma 5.3. *Let (X,d) be a complete metric space and $\mu_1 \in \mathscr{P}(X)$. Let us assume that μ_1 satisfies condition* (5.1). *Then there cannot exist nonempty Borel sets $A, B \subset X$ such that*

$$A \cup B = \operatorname{supp}\mu_1 \qquad \textit{and} \qquad \operatorname{dist}(A,B) > 0\,.$$

Proof. The argument used in the two-marginal proof works also here. We repeat the argument for the convenience of the reader.

Let us abbreviate $\mu := \mu_1$. We suppose, contrary to the claim, that such sets A, B exist. We fix $x \in X$ and $R > 0$ large enough to satisfy, for $\tilde{A} := A \cap B(x,R)$ and $\tilde{B} = B \cap B(x,R)$, the conditions

$$\mu(\tilde{A}) > 0 \text{ and } \mu(\tilde{B}) > 0\,.$$

Now we define for all $0 < t < \mu(\tilde{A})$ the measure ν_t by setting

$$\nu_t = \mu|_{X\setminus B(x,R)} + \frac{\mu_1(\tilde{A}) - t}{\mu_1(\tilde{A})}\mu|_{\tilde{A}} + \frac{\mu(\tilde{B}) + t}{\mu(\tilde{B})}\mu|_{\tilde{B}} \in \mathscr{P}(\operatorname{supp}\mu)\,.$$

Let us then define a coupling $\lambda \in \Pi_N(\mu, \nu_t, \mu, \ldots, \mu)$ by

$$\lambda = \left((\mathrm{id},\mathrm{id})_\sharp\mu|_{X\setminus B(x,R)} + \frac{1}{\mu(B(x,R))}\mu|_{B(x,R)} \times \nu_t|_{B(x,R)}\right) \times \underbrace{(\mathrm{id},\ldots,\mathrm{id})}_{N-2 \text{ times}}{}_\sharp\mu$$

We have

$$\begin{aligned}&\inf_{\lambda'\in\Pi_N(\mu,\nu_t,\mu,\ldots,\mu)} \int_{X^N} \sum_{1\le i<j\le N} h(d(x_i,x_j))\mathrm{d}\lambda' \le \int_{X^N} \sum_{1\le i<j\le N} h(d(x_i,x_j))\mathrm{d}\lambda \\ &\le (N-1)th(2R)\,. \end{aligned} \tag{5.4}$$

Since $\nu_t(A) < \mu(A)$, we have $W_\infty(\mu, \nu_t, \mu, \ldots, \mu) \ge \operatorname{dist}(A, B) > 0$. Combining this and (5.4) with (5.1) we get

$$\begin{aligned}(N-1)th(2R) &\ge \inf_{\lambda'\in\Pi_N(\mu,\nu_t,\mu,\ldots,\mu)} \int_{X^N} \sum_{1\le i<j\le N} h(d(x_i,x_j))\mathrm{d}\lambda' \\ &\ge \omega(W_\infty(\mu,\nu_t,\mu,\ldots,\mu)) \ge \omega(\operatorname{dist}(A,B)) > 0\,.\end{aligned}$$

Letting $t \to 0$ gives a contradiction. □

Before proving Theorem 5.1 we repeat a useful lemma, stated and proven in [7], that characterizes measures $\mu \in \mathscr{P}(X)$ that have compact support, using the function m introduced above.

Lemma 5.4. *Let (X, d) be a complete metric space and $\mu \in \mathscr{P}(X)$. Then μ has compact support if and only if*

$$m(t) = \inf_{x\in X} \mu(B(x,t)) > 0 \textit{ for all } t > 0\,.$$

Proof of Theorem 5.1. Let us first assume that $\operatorname{supp}\mu_1$ is compact and connected. Let measures $\mu_2, \ldots, \mu_N \in \mathscr{P}(\operatorname{supp}\mu_1)$ be arbitrary. We may assume that $W := W_\infty(\mu_1, \ldots, \mu_N) > 0$. We show that for all $\lambda \in \Pi_N(\mu_1, \mu_2, \ldots, \mu_N)$ we have

$$\lambda\left(\left\{(x_1,\ldots,x_N) \in X^N \mid d(x_i,x_j) \ge \frac{W}{68} \text{ for some } i \ne j\right\}\right) \ge \frac{1}{2}m(\tfrac{W}{68})\,. \tag{5.5}$$

Let us assume that this is not the case. Then by lemma 5.2, applied with $\delta = \epsilon = 0$, we have

$$W \le 17 \cdot 2 \cdot \frac{W}{68} = \frac{W}{2}\,,$$

which is impossible. Thus the estimate (5.5) holds and we have for all $\lambda \in \Pi_N(\mu_1, \ldots, \mu_N)$ that

$$\int_{X^N} \sum_{1\le i<j\le N} h(d(x_i,x_j))\mathrm{d}\lambda \ge h\left(\frac{W}{64}\right)\cdot\frac{1}{2}m\left(\frac{W}{68}\right)\,,$$

so we can take $\omega(t) = \frac{1}{2}m(\frac{t}{68})h(\frac{t}{68})$; by our assumptions on h, the compactness of $\operatorname{supp}\mu_1$, and Lemma 5.4, the function ω is positive for positive t. It is also inreasing. Hence condition (5.1) is satisfied.

It remains to prove the opposite direction. So we assume that condition (5.1) holds and aim at showing that the support of the first marginal $\mu := \mu_1$ is compact and connected. The connectedness of $\operatorname{supp}\mu$ follows from Lemma 5.3. For compactness, we fix $r > 0$. By Lemma 5.4 it suffices to show that $m(r) > 0$. This follows if we find a lower bound for $\mu(B(x,r))$ that is strictly positive and does not depend on the center $x \in X$.

Let $x \in X$ be arbitrary for a while. We denote

$$B := B(x,r),\ A_1 = (\operatorname{supp}\mu \cap B(x,2r)) \setminus B\,,\ \text{and } A_2 = \operatorname{supp}\mu \setminus B(x,2r)\,.$$

We may assume that $\mu(B) < 1$. This gives, by Lemma 5.3, also that $\mu(A_1) > 0$. We define for all $0 < t < \mu(A_1)$ a measure $\nu_t \in \mathscr{P}(\operatorname{supp}\mu)$ by

$$\nu_t = \mu|_{A_2} + \frac{\mu(A_1) - t}{\mu(A_1)}\mu|_{A_1} + (\mu(B) + t)\delta_x\,.$$

Next we define a coupling $\lambda \in \Pi_N(\mu, \nu_t, \mu, \ldots, \mu)$ by setting

$$\lambda = \left((\mathrm{id},\mathrm{id})_\sharp\left(\mu|_{A_2} + \frac{\mu(A_1)-t}{\mu(A_1)}\mu|_{A_1}\right) + \left(\frac{t}{\mu(A_1)}\mu|_{A_1} + \mu|_B\right) \times \delta_x\right) \times \underbrace{(\mathrm{id},\ldots,\mathrm{id})}_{N-2 \text{ times}}{}_\sharp\mu\,.$$

Using this coupling we can estimate

$$\begin{aligned}\inf_{\lambda' \in \Pi_N(\mu,\nu_t,\mu,\ldots,\mu)} \int_{X^N} \sum_{1\le i<j\le N} h(d(x_i,x_j))\mathrm{d}\lambda' &\le \int_{X^N} \sum_{1\le i<j\le N} h(d(x_i,x_j))\mathrm{d}\lambda \\ &\le (N-1)(th(2r) + \mu(B)h(r))\,.\end{aligned}$$

We also have

$$W_\infty(\mu, \nu_t, \mu, \ldots, \mu) \ge r\,.$$

Since $\nu_t \in \mathscr{P}(\operatorname{supp}\mu)$ we get by condition (5.1) the estimate

$$\omega(W_\infty(\mu,\nu_t,\mu,\ldots,\mu)) \le \inf_{\lambda' \in \Pi_N(\mu,\nu_t,\mu,\ldots,\mu)} \int_{X^N} \sum_{1\le i<j\le N} h(d(x_i,x_j))\mathrm{d}\lambda'\,.$$

Combining these estimates gives

$$\omega(r) \le (N-1)th(2r) + (N-1)h(r)\mu(B)\,,$$

from which we get, by setting $t \to 0$, a lower bound

$$\mu(B) \ge \frac{\omega(r)}{(N-1)h(r)}\,. \qquad \square$$

Let us then study the sharpness of the bound we found.

Example 5.5. *If for a fixed $\mu_1 \in \mathscr{P}(X)$ and $0 < r \le \operatorname{diam}(\operatorname{supp}\mu_1)$ we take $x \in X$ such that*

$$\mu_1(B(x,r)) = m(r)\,,$$

and consider

$$\mu_2 = \cdots = \mu_N = \mu_1|_{X\setminus B(x,r)} + B(x,r)\delta_x\,,$$

we get $W_\infty(\mu_1,\ldots,\mu_N) = r$ and

$$C_{h\circ d} \le (N-1)m(r)h(r)\,,$$

so we are a factor $(N-1)$ away from the sharpness we would like to have.

If we assume that h is convex and thus superadditive (assuming $h(0) = 0$), we get

$$C_{h\circ d} \le (N-1)m(r)h(r) \le m(r)h((N-1)r)\,.$$

References

[1] E. N. Barron, M. Bocea, and R. R. Jensen, *Duality for the L^∞ optimal transport problem*, Trans. Amer. Math. Soc. **369** (2017), 3289–3323.

[2] G. Bouchitté, C. Jimenez, and M. Rajesh, *A new L^∞ estimate in optimal mass transport.* Proc. Amer. Math. Soc. **135** (2007), 3525–3535.

[3] L. De Pascale and J. Louet, *A study of the dual problem of the one-dimensional L^∞ optimal transport problem with applications*, preprint (2017).

[4] C. Griessler, *c-cyclical monotonicity as a sufficient criterion for optimality in the multi-marginal Monge-Kantorovich problem*, Proc. Amer. Math. Soc. **145** (2018), 4735–4740.

[5] H. Jylhä, *The L^∞ Optimal Transport: infinite cyclical monotonicity and the existence of optimal transport maps*, Calc. Var. Partial Differential Equations, **52** (2015), 303–326.

[6] T. Champion, L. De Pascale, and P. Juutinen, *The ∞-Wasserstein distance: local solutions and existence of optimal transport maps*, SIAM J. Math. Anal. **40(1)**, 1-20 (2008).

[7] H. Jylhä and T. Rajala, *L^∞ estimates in optimal mass transportation*, J. Funct. Anal., **270** (2016), 4297–4321.

[8] H. Kellerer, *Duality theorems for marginal problems*, Z. Wahrsch. Verw. Gebiete, **67** (1984), 399–432.

[9] Y. Kim and B. Pass *A general condition for Monge solutions in multi-marginal optimal transport problem*, SIAM J. Math. Anal. **46** (2014), 1538–1550.

[10] C. Villani,*Optimal transport. Old and new.* vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin (2009).

University of Jyvaskyla, Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of Jyvaskyla, Finland

E-mail address: `anna.m.kausamo@jyu.fi`