ESAIM: COCV 25 (2019) 62
https://doi.org/10.1051/cocv/2018062

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

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 Seidl [*Phys. Rev. A* **60** (1999) 4387].

Mathematics Subject Classification. 49N15, 49J45, 49K30.

Received May 11, 2018. Accepted November 4, 2018.

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 functions of the type

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

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: $f(\mathrm{d}(x,y))=-\log(\mathrm{d}(x,y))$.

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

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

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

Keywords and phrases: Multi-marginal optimal transport, repulsive costs, Kantorovich duality.

Department of Mathematics and Statistics, University of Jyvaskyla, PO Box 35, 40014 Jyvaskyla, Finland.

** Corresponding author: tapio.m.rajala@jyu.fi

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 [14] 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 [7, 12, 13, 27, 28], as well as the problem (1.1) in a lower space dimensions $X = \mathbb{R}^d, d = 1, 2$ [5, 6, 11, 21, 25]. 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 section 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 [12] 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 Seidl [26]. 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

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

As pointed out in [26] (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, 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] [18, 22, 24]. This was suggested in the physics literature by Gori-Giorgi, Seidl and Vignale [16] and, established rigorously by Cotar, Friesecke and Klüppelberg [8, 9].

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

The same theorem has been proved also by Bindini-De Pascale [1] ($N = 3$ case) and by Lewin [23], 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, 27, 28], the repulsive harmonic cost

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

was studied: Friesecke *et al.* [13] have shown the existence of optimal transport plans supported in $(N-1)d$-dimensional sets; in [12] 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 [15], 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

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

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 [20] under the assumption that there exist $L^1_\rho(X)$-functions $h_1, \ldots, h_N$ and a constant C such that

$$C \leq c(x_1, \ldots, x_N) \leq h(x_1) + \cdots + h(x_N).$$

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 ($1 \leq s \leq d$)

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

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

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

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

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

and

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

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 Sect. 3).

In Section 4, we extend the duality results of [3, 10, 20] for a class of unbounded cost functions (Thm. 4.2) and in Section 5 we obtain regularity results of Kantorovich potentials (Thm. 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

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

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 ith coordinate

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

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

[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.

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

this is the transportation cost related to γ.

We want to study the dual problem, so we set

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

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*

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

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 $x \in X \setminus \mathrm{spt}(\rho)$ 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 [17] for the differentiation result, and for instance [19] 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 $\mathrm{spt}(\rho)$ so that for each k the subcollection $\{Q_{k_i} : i \in N_k\}$ is a partition of $\mathrm{spt}(\rho)$ with $\mathrm{diam}(Q_{k_i}) < 2^{-k}$ and $\rho(Q_{k_i}) > 0$ satisfying $\tilde{u}(x) = u(x)$ for ρ-almost every $x \in \mathrm{spt}(\rho)$, where we have defined for every $x \in \mathrm{spt}(\rho)$

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

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,

$$\sum_{i=1}^N \tilde{u}(x_i) = \sum_{i=1}^N \liminf_{k \to \infty} \frac{1}{\rho(Q_k(x_i))} \int_{Q_k(x_i)} u(y) \, \mathrm{d}\rho(y)$$

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

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 $D_n \xrightarrow{\Gamma^+} D$ if for all $u\in S$ the following two conditions hold:

for any sequence $(u_n)_{n\in\mathbb{N}}$ converging to u we have
$$\limsup_n D_n(u_n) \le D(u) \text{ and} \tag{I+}$$

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

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

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

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], Thm. 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*

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

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

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

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], Lem. 4.3) to find an upper semicontinuous function h such that

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

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

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

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

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

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

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

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)

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

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

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

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

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

Proof. The proof presented in ([3], Thm. 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 [20]. We then fix $\gamma \in \mathcal{P}(X^N)$ and show that

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

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

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 [20]. 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

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

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*

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

where $f\colon [0,+\infty[\to \mathbb{R}\cup\{+\infty\}$ is a function satisfying (F1) and (F2) 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], Thm. 2.4), where it was shown that not only is

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

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 [20] 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*

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

where $f\colon [0,+\infty[\to \mathbb{R}\cup\{+\infty\}$ is a function satisfying $(F1)$ and $(F2)$. 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 $\gamma_L\neq 0$ 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 $\gamma^P_{R/2}$ with the cost c_R and the marginals $\rho_{R/2}$. Let us fix $R_0>0$ such that $\gamma_{R_0/2}\neq 0$, 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 $d(x_1, \overline{x}_j) \geq \frac{\alpha}{2}$ 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 $R \geq 2L$,

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

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 $R > R_1$

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

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:

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

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]

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

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

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

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

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

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

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 $(x_1,\dots,x_N)\in \operatorname{spt}(\gamma)$ the estimate

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

Thus, since γ is a probability measure, we have

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

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

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

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*

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

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

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], Lem. 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* $\tilde{u} \geq u$ *which satisfies the representation*

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

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 $x \in X$,

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

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 $h \colon X \to \mathbb{R} \cup \{-\infty, +\infty\}$,

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

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

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], Thm. 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 $C \leq C_R$ 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 $x_1, x_2 \in X$

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

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

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

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 Prop. 3.1). Now, from (5.4) we see that for $d(x_1, x_2) \geq R > 1$, we have

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

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 one 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. □

Acknowledgements. 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) 909–934.
[2] A. Braides, Gamma-Convergence for Beginners, Vol. 22. Clarendon Press, Oxford (2002).
[3] G. Buttazzo, T. Champion and L. De Pascale, Continuity and estimates for multimarginal optimal transportation problems with singular costs. *Appl. Math. Optim.* **78** (2018) 185–200.
[4] G. Buttazzo, L. De Pascale and P. Gori-Giorgi, Optimal-transport formulation of electronic density-functional theory. *Phys. Rev. A* **85** (2012) 062502.
[5] H. Chen, G. Friesecke and C. B. Mendl, Numerical methods for a kohn–sham density functional model based on optimal transport. *J. Chem. Theory Comput.* **10** (2014) 4360–4368.
[6] M. Colombo and F. Stra, Counterexamples to multimarginal optimal transport maps with coulomb cost and radial measures. *Math. Models Methods Appl. Sci.* **26** (2016) 1025–1049.
[7] L. Cort, D. Karlsson, G. Lani and R. van Leeuwen, Time-dependent density-functional theory for strongly interacting electrons. *Phys. Rev. A* **95** (2017) 042505.
[8] C. Cotar, G. Friesecke and C. Klüppelberg, Density functional theory and optimal transportation with coulomb cost. *Comm. Pure Appl. Math.* **66** (2013) 548–599.
[9] C. Cotar, G. Friesecke and C. Klüppelberg, Smoothing of transport plans with fixed marginals and rigorous semiclassical limit of the Hohenberg-Kohn functional. *Arch. Ration. Mech. Anal.* **228** (2018) 891–922.
[10] L. De Pascale, Optimal transport with coulomb cost. approximation and duality. *ESAIM: M2AN* **49** (2015) 1643–1657.
[11] S. Di Marino, L. De Pascale and M. Colombo, Multimarginal optimal transport maps for 1-dimensional repulsive costs. *Can. J. Math.* **67** (2015) 350–368.
[12] S. Di Marino, A. Gerolin and L. Nenna, Topological Optimization and Optimal Transport In the Applied Sciences. De Gruyter (2017).

[13] 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. *J. Chem. Phys.* **139** (2013) 164109.

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

[15] A. Gerolin, A. Kausamo and T. Rajala, Non-existence of optimal transport maps for the multi-marginal repulsive harmonic cost. Preprint arXiv:1805.00417 (2018).

[16] P. Gori-Giorgi, M. Seidl and G. Vignale, Density-functional theory for strongly interacting electrons. *Phys. Rev. Lett.* **103** (2009) 166402.

[17] W. E. Hartnett and A. H. Kruse, Differentiation of set functions using Vitali coverings. *Trans. Amer. Math. Soc.* **96** (1960) 185–209.

[18] P. Hohenberg and W. Kohn, Inhomogeneous electron gas. *Phys. Rev.* **136** (1964) B864.

[19] A. Käenmäki, T. Rajala and V. Suomala, Existence of doubling measures via generalised nested cubes. *Proc. Am. Math. Soc.* **140** (2012) 3275–3281.

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

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

[22] M. Levy, 2016 Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Nati. Acad. Sci.* **12** (76) 6062–6065.

[23] M. Lewin, Semi-classical limit of the Levy-Lieb functional in density functional theory. *C. R. Math. Acad. Sci. Paris* **356** (2018) 449–455.

[24] E. H. Lieb, Density functionals for coulomb systems, in Inequalities. Springer, Berlin, Heidelberg (2002) 269–303.

[25] F. Malet and P. Gori-Giorgi, *Strong* correlation in Kohn-Sham density functional theory. *Phys. Rev. Lett.* **109** (2012) 246402.

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

[27] M. Seidl, S. Di Marino, A. Gerolin, L. Nenna, K. Giesbertz and P. Gori-Giorgi, The strictly-correlated electron functional for spherically symmetric systems revisited. Preprint arXiv:1702.05022 (2017).

[28] M. Seidl, P. Gori-Giorgi and A. Savin, Strictly correlated electrons in density-functional theory: a general formulation with applications to spherical densities. *Phys. Rev. A* **75** (2007) 042511.

[29] C. Villani. Optimal Transport: old and New, Vol. 338. Springer Science & Business Media, Berlin, Heidelberg (2008).

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