Optimal measure transportation with respect to non-traditional costs

S. Artstein-Avidan\textsuperscript{1} · S. Sadovsky\textsuperscript{1} · K. Wyczesany\textsuperscript{1}

Received: 27 July 2021 / Accepted: 1 October 2022 / Published online: 6 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
We study optimal mass transport problems between two measures with respect to a non-traditional cost function, i.e. a cost $c$ which can attain the value $+\infty$. We define the notion of $c$-compatibility and strong $c$-compatibility of two measures, and prove that if there is a finite-cost plan between the measures then the measures must be $c$-compatible, and if in addition the two measures are strongly $c$-compatible, then there is an optimal plan concentrated on a $c$-subgradient of a $c$-class function. This function is the so-called potential of the plan. We give two proofs of this theorem, under slightly different assumptions. In the first we utilize the notion of $c$-path-boundedness, showing that strong $c$-compatibility implies a strong connectivity result for a directed graph associated with an optimal map. Strong connectivity of the graph implies that the $c$-cyclic monotonicity of the support set (which follows from classical reasoning) guarantees its $c$-path-boundedness, implying, in turn, the existence of a potential. We also give a constructive proof, in the case when one of the measures is discrete. This approach adopts a new notion of ‘Hall polytopes’, which we introduce and study in depth, to which we apply a version of Brouwer’s fixed point theorem to prove the existence of a potential in this case.

Mathematics Subject Classification 49Q22 · 52A23 · 49K27 · 52A01

\textsuperscript{1} All considered measures are Borel measures on Polish spaces, which are complete, separable metric spaces equipped with their Borel $\sigma$-algebra (see also the beginning of Sect. 2).
1 Introduction and results

The Monge transport problem is concerned with finding a transport map moving mass from one probability measure\(^1\) to another, in a way which is optimal with respect to some cost function. The most widely studied case of this problem is for the quadratic cost \(c(x, y) = \|x - y\|^2/2\), for which the Brenier–Gangbo-McCann theorem \([10, 12]\) implies that under mild conditions on the measures involved, optimal transport maps exist and are given by gradients of convex functions. In this work the main emphasis will be on non-traditional cost functions, i.e. costs that can attain the value \(+\infty\), as this project is motivated by the study of transportation with respect to the so-called polar cost given by

\[
p(x, y) = -\ln \left(\langle x, y \rangle - 1\right),
\]

where \(p(x, y) = +\infty\) if \(\langle x, y \rangle \leq 1\). This cost function is linked with the polarity transform (see \([3, 4]\)), similarly to the strong connection of the quadratic cost with the Legendre transform. Transportation with respect to the polar cost was first considered in \([7]\).

We provide necessary conditions on pairs of measures, together with a cost \(c\), for which finite cost plans exist. To this end, we discuss the class of functions connected with a cost, called its \(c\)-class (see the definition in Sect. 2.2). The optimality of a plan is linked with the possibility of finding a “potential” for the plan, which is a \(c\)-class function such that the plan lies on its \(c\)-subgradient (yet another important notion we discuss in depth, see the definition in Eq. (7)).

We will see shortly that the mere existence of a finite cost plan between two measures \(\mu\) and \(\nu\) implies that the two measures considered are \(c\)-compatible, namely that for any (Borel) measurable set \(A \subseteq X\) one has \(\mu(A) \leq \nu(\{y : \exists x \in A, \; c(x, y) < \infty\})\). This is quite intuitive—all points (up to a set of measure zero) in \(A\) must be mapped to points in the target space with which they have finite cost. This \(c\)-compatibility of two measures is thus a necessary condition (for the formal definition of \(c\)-compatibility see Definition 3.2, and for the statement of the necessity of this condition see Lemma 3.3). As Example 3.5 will show, \(c\)-compatibility is not a sufficient condition for the existence of a finite cost plan. However, if a finite cost plan exists, a slight strengthening of the \(c\)-compatibility condition (in which we demand a strict inequality) is already sufficient to ensure that the optimal plan has a potential. We will also show why the notion of “strong compatibility” is a very natural strengthening of compatibility, and discuss cases where two measures are \(c\)-compatible but not strongly \(c\)-compatible and how this implies that the transport problem is decomposable into sub-problems.

In this note we only consider symmetric cost functions \(c : X \times X \to (-\infty, \infty]\) with \(c(x, y) = c(y, x)\), but to see the difference between the two variables we denote the second copy of \(X\) by \(Y\) and write \(c : X \times Y \to (-\infty, \infty]\). Our results hold for the non-symmetric case as well, with only minor adjustments. We will also add a lower-bound assumption on the cost \(c\) which allows to integrate it without the risk of getting \(-\infty\) total cost (see Example 2.5). We say that \(c\) is essentially bounded from below with respect to \(\mu\) and \(\nu\) if there exist functions \(a(x) \in L^1(\mu), b(x) \in L^1(\nu)\) such that \(c(x, y) \geq a(x) + b(y)\). For the polar cost this condition is satisfied if, for example, both measures have finite second moment.

Our main theorem is the following (here \(\delta^c\varphi\) denotes the \(c\)-subgradient of \(\varphi\), see the definition in Eq. (7), and \(\Pi(\mu, \nu)\) denotes all transport plans between \(\mu\) and \(\nu\), see the beginning of Sect. 2 for all relevant definitions).

**Theorem 1.1** Let \(X = Y\) be a Polish space, let \(c : X \times Y \to (-\infty, \infty]\) be a continuous and symmetric cost function, essentially bounded from below with respect to probability measures...
\( \mu \in \mathcal{P}(X) \) and \( \nu \in \mathcal{P}(Y) \). Assume \( \mu \) and \( \nu \) are strongly c-compatible, namely satisfy that for any measurable \( A \subset X \) we have
\[
\mu(A) + \nu(\{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}) \leq 1,
\]
and for any measurable \( A \subset X \) with \( \mu(A) \neq 0, 1 \) we have
\[
\mu(A) + \nu(\{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}) < 1.
\]

If there exists some finite cost plan transporting \( \mu \) to \( \nu \), then there exists a c-class function \( \phi \) and an optimal transport plan \( \pi \in \Pi(\mu, \nu) \) concentrated on \( \partial^c \phi \).

The proof uses results from [5] on c-path-boundedness, which is a notion that replaces c-cyclic monotonicity from the Rockafellar-Rochet-Rüscheidorf result (see [16–18]) in the case when the cost is non-traditional. The c-path-boundedness is a necessary and sufficient condition for a set to be included in a c-subgradient of a c-class function.

Our initial main interest in developing this theory was transportation with respect to the polar cost, in which case we have a precise form for c-subgradients. We state the relevant theorem, which is almost a direct application of the theorem above, together with some simple analysis of polar-subgradients as performed in [4]. By \( \text{Cvx}_0(\mathbb{R}^n) \) we denote a class of lower semi-continuous convex functions from \( \mathbb{R}^n \) to \([0, \infty]\) which take the value zero at the origin. By \( \mathcal{A} \) we denote the polarity transform on the class \( \text{Cvx}_0(\mathbb{R}^n) \), defined in [3] and given in (16). To obtain Theorem 1.2 from Theorem 1.1, we make use of Lemma 3.12.

**Theorem 1.2** Let \( X = Y = \mathbb{R}^n \) and let \( \mu, \nu \in \mathcal{P}(\mathbb{R}^n) \) be probability measures with finite second moment, which are strongly p-compatible where \( p(x, y) = -\ln(\langle x, y \rangle - 1) \) is the polar cost, that is
\[
\mu(K) + \nu(K^\circ) < 1
\]
for any convex set \( K \) with \( \mu(K) \neq 0, 1 \), and \( \nu(K^\circ) = 0 \) if \( \mu(K) = 1 \). Assume further that \( \mu \) is absolutely continuous. Assume there exists some finite cost plan mapping \( \mu \) to \( \nu \). Then there exists \( \phi \in \text{Cvx}_0(\mathbb{R}^n) \) such that \( \partial^c \phi \) is an optimal transport map between \( \mu \) and \( \nu \), where we define
\[
\partial^c \phi(x) := \{ y \in \mathbb{R}^n : \phi(x)A\phi(y) = \langle x, y \rangle - 1 > 0 \}.
\]
In particular, for \( \mu \)-almost every \( x \), the set \( \partial^c \phi(x) \) is a singleton.

We remark that the existence of a potential function for the cost \( p \) and other non-traditional costs leads naturally to the question regarding regularity of such potentials (as introduced by Caffarelli in [11] and developed, among others, by Trudinger and Wang in [21]). In this work we do not pursue this direction, and instead focus on the analysis of the existence of potentials, leaving the question of regularity for future work.

In the second half of the paper we specialize to the case where \( \nu \) is discrete. In this case we give a constructive proof for the existence of a transport map, where the c-class function is given as a finite infimum of “basic functions” (see (6)) associated with the cost. The advantage of this method is that much of the geometry of the problem is revealed. In the proof, we generalize a method used by Ball [6] for the quadratic cost, where all possible maps are parametrized by a weight vector, and the existence of the required one is shown using Brouwer’s fixed point theorem. However, in contrast with the case of the classical quadratic cost function and other traditional costs, when the cost attains infinite values the set of all discrete measures with a given support, to which a measure \( \mu \) can be mapped with finite cost,
is given by an interesting polytope which we call the Hall polytope of the measure \( \mu \) (see Definition 3.6).

We present a thorough study of the structure and geometry of Hall polytopes (which for traditional costs are just simplices) in Sect. 5, and their connection with transportation to discrete measures. One of our results, presented in Theorem 5.16, has to do with properties of mappings from the simplex to a Hall polytope, which is the main tool in the proof of Theorem 1.3 below. It turns out that the condition of strong \( c \)-compatibility corresponds to measures with weight vectors in the interior of the Hall polytope. An advantage of using the method of Hall polytopes is that we can relax the conditions on the cost function. We do, however, need a condition of \( c \)-regularity for the measure \( \mu \) (given in Definition 5.1), which for the polar cost is satisfied if, say, \( \mu \) is absolutely continuous. These geometric conditions and results are stated in purely measure-theoretic terms as follows.

**Theorem 1.3** Let \( X \) be some Polish space and \( Y = \{u_1\}_{i=1}^m \) a discrete measure space. Assume \( c : X \times Y \rightarrow (-\infty, \infty] \) is a measurable cost function, \( \mu \in \mathcal{P}(X) \) is \( c \)-regular and \( \nu = \sum_{i=1}^m \alpha_i > 1 u_i \in \mathcal{P}(Y) \). Assume furthermore, that there exists a \( c \)-regular probability measure concentrated on each intersection

\[
\{x \in X : c(x, u_i) < \infty\} \cap \{x \in X : c(x, u_j) < \infty\}
\]

for \( i, j \in \{1, \ldots, m\} \) (in particular, each intersection must be non empty). If \( \mu \) and \( \nu \) are strongly \( c \)-compatible then there exists an optimal transport plan \( \pi \in \Pi(\mu, \nu) \) whose graph lies in the c-subgradient \( \partial_c \varphi \) of a \( c \)-class function \( \varphi : X \rightarrow [-\infty, \infty] \).

It turns out that for \( X = \mathbb{R}^n \), which is the space we consider in the framework of the polar cost \( p(x, y) = -\ln(x, y) - 1 \), the assumption on the existence of a \( c \)-regular measure is satisfied under simpler assumptions. We formulate the result for the polar cost, but as the proof will demonstrate it holds for more general costs on \( \mathbb{R}^n \) (more precisely, costs for which the uniform measure on any ball is \( c \)-regular) and a \( c \)-regular measure \( \mu \).

**Theorem 1.4** Let \( X = \mathbb{R}^n \) and \( Y = \{u_i\}_{i=1}^m \subset \mathbb{R}^n \) where \( u_i \neq -tu_j \) for any \( t > 0 \). Let \( \mu \in \mathcal{P}(X) \) and \( \nu = \sum_{i=1}^m \alpha_i > 1 u_i \in \mathcal{P}(Y) \) be probability measures and assume that \( \mu \) is \( p \)-regular. If \( \mu \) and \( \nu \) are strongly \( p \)-compatible then there exists an optimal (with respect to the polar cost) transport plan \( \pi \in \Pi(\mu, \nu) \) whose graph lies in the polar subgradient \( \partial_p \varphi \) of some function \( \varphi \in \text{Cvx}_0(\mathbb{R}^n) \). Moreover, for \( \mu \)-almost every \( x \), the set \( \partial_p \varphi(x) \) is a singleton.

Finally, the case where the measures \( \mu \) and \( \nu \) are \( c \)-compatible but not strongly so, can be analyzed as well. In this case we can write \( \mu = \mu_1 + \mu_2 \) and \( \nu = v_1 + v_2 \) where \( \mu_1(X) = v_1(Y) \) (and so \( \mu_2(X) = v_2(Y) \)), where the measures \( \mu_1 \) and \( \mu_2 \) are concentrated on disjoint sets, as are \( v_1 \) and \( v_2 \), and in such a way that any finite cost transport plan \( \pi \in \Pi(\mu, \nu) \) is given as a sum of \( \pi_1 \in \Pi(\mu_1, v_1) \) and \( \pi_2 \in \Pi(\mu_2, v_2) \). We illustrate this in Sect. 7.

**Structure of the paper**

Section 2 is dedicated to gathering all the required definitions, notions and previous results. In Sect. 3 we discuss the notion of \( c \)-compatibility and strong \( c \)-compatibility together with their geometric interpretation. In Sect. 4 we prove Theorem 1.1. In Sect. 5 we consider transportation to a discrete measure and show how one may treat this special case using some deep structural properties of Hall polytopes which we establish and utilize to prove Theorem 1.3. In Sect. 6 we specialize to the polar cost and give proofs of Theorems 1.2 and 1.4. In Sect. 7
we discuss the case of measures which are $c$-compatible but not strongly $c$-compatible. For completeness we include an appendix A in which we review $c$-subgradients, with detailed examples and geometric intuition.

2 Background and preliminary observations

In what follows, our standing assumption is that the spaces considered are Polish spaces with the Borel $\sigma$-algebra. The definitions make sense for general measure spaces and much of the theory holds with weaker assumptions on the spaces and the $\sigma$-algebra. However, since some of the theorems we use require this assumption, we assume throughout the text that $X$, $Y$ are Polish spaces, simplifying the exposition.

2.1 Transport plans and maps

Given two Polish spaces $X$, $Y$ and probability measures $\mu \in P(X)$ and $\nu \in P(Y)$, and a measurable\(^2\) cost function $c : X \times Y \to (-\infty, \infty]$, we say that there exists a $c$-optimal transport map between them if the following infimum is attained:

$$\inf_T \int_X c(x, T(x))d\mu(x),$$

where $T : X \to Y$ are measurable transport maps, i.e. $\nu(B) = \mu(T^{-1}(B))$ for all measurable sets $B \subset Y$. We say that there exists a $c$-optimal plan between them if the infimum

$$\inf_\pi \int_{X \times Y} c(x, y)d\pi(x, y) \tag{2}$$

is attained, where $\pi \in \Pi(\mu, \nu)$, that is $\pi$ is a probability measure on $X \times Y$ satisfying

$$\pi(A \times Y) = \mu(A), \quad \pi(X \times B) = \nu(B)$$

for all measurable sets $A \subset X$ and $B \subset Y$. Every transport map induces a transport plan supported on its graph, while not every plan is induced by a map. We denote the infimum in (2), also called the “total cost”, by $C(\mu, \nu)$. Due to the Kantorovich Duality Theorem [13, 14], when $c$ is lower semi-continuous, the total cost is equal to

$$\sup_{\phi, \psi} \left\{ \int_X \phi d\mu + \int_Y \psi d\nu : \phi \in L_1(X, \mu), \psi \in L_1(Y, \nu) \text{ admissible} \right\} \tag{3}$$

where $(\phi, \psi)$ is called an admissible pair, if $\phi : X \to [-\infty, \infty]$, $\psi : Y \to [-\infty, \infty]$ satisfy

$$\forall(x, y) \in X \times Y, \quad \phi(x) + \psi(y) \leq c(x, y).$$

For a formulation and proof of the Kantorovich Duality Theorem as stated above, see Theorem 3.1 in [1]. In the case where $\phi = +\infty$ and $\psi = -\infty$ we stipulate $-\infty + \infty = -\infty$, namely in such a case the condition above holds regardless of the value of $c(x, y)$.

\(^2\) When referring to a function on $X \times Y$ as “measurable” we assume it is both measurable with respect to the product $\sigma$-algebra and its fibers $f(\cdot, y)$ and $f(x, \cdot)$ are measurable functions on $X$ and $Y$ respectively, for any $x \in X$ and $y \in Y$. 

\(\text{Springer}\)
2.2 The \(c\)-transform

Motivated by (3), for every function \(\psi : Y \to [-\infty, \infty]\) one may consider the largest function \(\varphi\) for which \((\varphi, \psi)\) is an admissible pair, and vice versa. This gives rise to the \(c\)-transform, defined by

\[
\psi^c(x) = \inf_y (c(x, y) - \psi(y)),
\]

and

\[
\varphi^c(y) = \inf_x (c(x, y) - \varphi(x)).
\]

**Remark 2.1** Here if on the right hand side are infinities of opposite signs, which may occur only if \(\psi(y) = \infty\) (as \(c \neq -\infty\)), we use the opposite convention, namely \(-\infty + \infty = +\infty\), since when the cost \(c(x, y)\) is infinite there is no restriction on the sum \(\varphi(x) + \psi(y)\). In general one must be careful with sums of opposite side infinities, as there is no obvious “rule of thumb” that can be applied everywhere.

Note that for a general cost we may lose the measurability of \(\varphi\) when applying the \(c\)-transform, as well as integrability, even under the assumption that \(c\) is measurable in the strong sense we have postulated. When \(c\) is continuous, however, this is less of a problem. Also, by truncating the functions and taking limits, the issue of integrability can sometimes be resolved. Nevertheless, one should be extra careful when using (3) for a pair \(\varphi, \psi\) when the cost is non-traditional, and in the existing literature it is not always clear for which theorems does the non-traditional case follow without substantially changing the conditions of the proofs.

When \(X = Y\) and \(c(\cdot, \cdot)\) is symmetric in its arguments the transforms in (4) and (5) coincide. Hence, abusing notation, we use the same notation for both. We define the \(c\)-class as the image of the \(c\)-transform \(\{\psi^c : \psi : Y \to [-\infty, \infty]\}\), or equivalently, as all the functions \(\varphi\) such that \(\varphi^c = \varphi\). By definition, any function in the \(c\)-class is an infimum of basic functions, which are functions of the form

\[
c(x) = c(x, y_0) + t
\]

for some \(y_0 \in Y\) and \(t \in \mathbb{R}\). It is useful to notice that the \(c\)-class is always closed under pointwise infimum (this fact is commonly known and used, see e.g. [1, 22], and a simple proof can be found in [24]).

2.3 The \(c\)-subgradient

Given a function \(\varphi\) in the \(c\)-class, its \(c\)-subgradient is the subset of \(X \times Y\) given by

\[
\partial^c \varphi = \{(x, y) : \varphi(x) + \varphi^c(y) = c(x, y) \text{ and } c(x, y) < \infty\}.
\]

To illustrate the relevance of \(c\)-subgradients to the study of optimal transport, let us present a folklore argument, which can be made precise for traditional costs, and which we only use as motivation but do not claim it holds in general.

In the Kantorovich Duality Theorem, recalled as (3) above, one is inclined to replace \(\psi\) with the largest admissible partner of \(\varphi\) (at least so long as it is measurable and in \(L_1(\nu)\)), and then replace \(\varphi\) by \(\varphi^c\). In this sense, one may think of (3) applied only to admissible pairs...
$\text{(ϕ, } \varphi^c\text{), where } \varphi = \varphi^{cc}\text{ is in the } c\text{-class. However, for any } \pi \in \Pi(\mu, \nu) \text{ and } \varphi \text{ in the } c\text{-class,}

\int_X \varphi d\mu(x) + \int_Y \varphi^c(y) d\nu(y) = \int_{X \times Y} (\varphi(x) + \varphi^c(y)) d\pi(x, y) \leq \int_{X \times Y} c(x, y) d\pi(x, y)

\text{So for equality between the left and right hand side to be obtained for some (potential) } \varphi \text{ and (optimal plan) } \pi, \text{ we see that } \pi \text{ must be concentrated on the set } \partial^c \varphi. \text{ In other words, finding optimal plans admitting a potential is equivalent to finding some plan concentrated on a } c\text{-subgradient. While this argument is not precise (in particular, we ignored measurability and integrability assumptions, applying (3) to a pair } (\varphi, \varphi^c), \text{ it constitutes the motivation behind searching for potentials in optimal transport problems.}

\text{The above observation shows the importance of the notion of the } c\text{-subgradient mapping. The name } c\text{-subgradient is connected to the fact that for the classical cost } c(x, y) = -(x, y), \text{ the } c\text{-class consists of upper semi-continuous concave functions, the } c\text{-transform of } -(\varphi) \text{ is } -L(\varphi) \text{ (where } L \text{ is the well known Legendre transform), and the } c\text{-subgradient of } -\varphi \text{ at } x \text{ is the usual subgradient } \partial \varphi(x). \text{ So as not to disturb the flow of the paper, we gathered some basic facts about the } c\text{-subgradient, including the geometric intuition behind it, in Appendix A.}

2.4 c-\text{cyclic monotonicity and } c\text{-path-boundedness}

\text{The connection between optimality of a plan and some geometric information on its support is quite intuitive: if a plan is optimal, then we should not gain any profit by interchanging several portions of it. This is the idea behind the well known notion of } c\text{-cyclic monotonicity. Given a cost } c : X \times Y \to (-\infty, \infty], \text{ a subset } G \subset X \times Y \text{ is called } c\text{-cyclically monotone if } c(x, y) < \infty \text{ for all } (x, y) \in G, \text{ and for any } m, \text{ any } (x_i, y_i)_{i=1}^m \subset G, \text{ and any permutation } \sigma \text{ of } [m] = \{1, \ldots, m\} \text{ it holds that}

\sum_{i=1}^m c(x_i, y_i) \leq \sum_{i=1}^m c(x_i, y_{\sigma(i)}). \tag{8}

\text{This definition seems to have been first introduced by Knott and Smith [19], as a generalization of cyclic monotonicity considered by Rockafellar [17] in the case of quadratic cost. It is easy to check that if } \varphi \text{ is a } c\text{-class function then any set } G \subset \partial^c \varphi \text{ is } c\text{-cyclically monotone. The theorems of Rockafellar, Rochet and Rüschendorf give the reverse implication, in the case of a traditional cost. Namely, when } c : X \times Y \to \mathbb{R}, \text{ a set } G \subset X \times Y \text{ is } c\text{-cyclically monotone if and only if there exists a } c\text{-class function such that } G \subset \partial^c \varphi. \text{ For non-traditional costs, this is no longer the case, and one may construct } c\text{-cyclically monotone sets which admit no potential. In [5], the corresponding result for non-traditional costs is provided. Cyclic monotonicity has to be replaced by a stronger notion, which we called } c\text{-path-boundedness.}

\text{Definition 2.2 Fix sets } X, Y \text{ and } c : X \times Y \to (-\infty, \infty]. \text{ A subset } G \subset X \times Y \text{ will be called } c\text{-\textit{path-bounded} if } c(x, y) < \infty \text{ for any } (x, y) \in G, \text{ and for any } (x, y) \in G \text{ and } (z, w) \in G, \text{ there exists a constant } M = M((x, y), (z, w)) \in \mathbb{R}, \text{ such that the following holds: For any } m \in \mathbb{N} \text{ and any } (x_i, y_i)_{i=2}^{m-1} \subset G, \text{ denoting } (x_1, y_1) = (x, y) \text{ and } (x_m, y_m) = (z, w), \text{ we have}

\sum_{i=1}^{m-1} (c(x_i, y_i) - c(x_{i+1}, y_{i+1})) \leq M.
The fact that a $c$-path-bounded set is also $c$-cyclically monotone is easy to establish (see [5]). With this definition the main theorem of [5] can be stated.

**Theorem 2.3** Let $X$, $Y$ be sets and let $c : X \times Y \to (-\infty, \infty]$ be given. A set $G \subset X \times Y$ is $c$-path-bounded if and only if there exists a $c$-class function $\varphi$ such that $G \subset \partial^c \varphi$.

It was also demonstrated in [5] that under certain conditions, the notions of $c$-cyclic monotonicity and $c$-path-boundedness do coincide. One such instance, which will be used in this paper, is explained and formulated in Proposition 4.1 in Sect. 4.

### 2.5 Some known results about existence of optimal plans and potentials

Having fixed a cost, the discussion about the structure of an optimal plan naturally splits into several components. The first, which is relevant only when the cost is non-traditional, is the existence of some finite cost plan (necessary conditions will be discussed in the next section). Further, one can ask whether an optimal plan exists. This is the object of the next theorem, which is quoted from Villani [23].

Recall that $\Pi(\mu, v)$ denotes the set of all probability measures on $X \times Y$ whose marginals are $\mu \in \mathcal{P}(X)$ and $v \in \mathcal{P}(Y)$, and that $c : X \times Y \to (-\infty, \infty]$ is essentially bounded with respect to $\mu$ and $v$ if there exist upper semi-continuous function $a : X \to (-\infty, \infty]$, $a \in L_1(\mu)$ and $b : X \to (-\infty, \infty]$, $b \in L_1(\nu)$ such that $c(x, y) \geq a(x) + b(y)$ for all $x \in X$, $y \in Y$.

**Theorem 2.4** Let $X$, $Y$ be two Polish spaces, let $\mu \in \mathcal{P}(X)$ and $v \in \mathcal{P}(Y)$. Let $c : X \times Y \to (-\infty, \infty]$ be a lower semi-continuous cost function which is essentially bounded with respect to $\mu$ and $v$. Then there exists a $c$-optimal plan $\pi \in \Pi(\mu, v)$.

Let us note that, in the above theorem, the existence of a plan with finite total cost is not assumed as when no finite cost plan exists, any plan (say, $\mu \otimes v$) is optimal in a trivial sense. Further, a simple example demonstrates that without some kind of assumption on boundedness from below of the cost, the total cost may be $-\infty$, and in this case optimal measures can be concentrated on sets which are far from being $c$-cyclically monotone.

**Example 2.5** Let $p(x, y) = -\ln(xy - 1)_+$ be the polar cost on $\mathbb{R}_+ \times \mathbb{R}_+$. Let $\mu$ be a discrete probability measure on $\mathbb{R}_+$ given by $\mu = \sum_{n=2}^{\infty} \alpha_n \delta_{1/n}$, where $\alpha_n > 0$ are such that $\sum_{n=2}^{\infty} \alpha_n = 1$ and $\sum_{n=2}^{\infty} \ln(n^2 - 1)\alpha_n = \infty$ (for example one can take $\alpha_n = \frac{c}{n\ln(n^2)}$ with an appropriate constant $c$). Consider transport plans of $\mu$ to itself, namely $\Pi(\mu, \mu)$.

We claim that in this case, the identity map $x \mapsto x$ is a transport plan whose total cost is $-\infty$ (in particular, it is optimal) but it is not concentrated on a $p$-cyclically monotone set. Indeed, consider the measure $\pi_\mu$ on the diagonal whose projection is $\mu$. Its total cost is $\sum_{n=2}^{\infty} -\ln(n^2 - 1)\alpha_n = -\infty$.

Clearly even for two points $(x_1, y_1) = (2, 2)$ and $(x_2, y_2) = (3, 3)$ it holds that

$$-\ln(2 \cdot 2 - 1) - \ln(3 \cdot 3 - 1) = -\ln(24) > -\ln(2 \cdot 3 - 1) - \ln(3 \cdot 2 - 1) = -\ln(25).$$

We thus see that an optimal plan (albeit with negative infinity cost) may have support which is not $c$-cyclically monotone.

Analysing the geometric structure of an optimal plan, after showing its existence, is a problem which has a long history. After Brenier [10], following Rüschendorf [18] determined the classical structure of cyclic monotonicity of optimal plans, Gangbo and McCann [12]...
extended the result to lower semi-continuous cost functions bounded from below. They showed that every finite optimal plan with respect to such costs lies on a \(c\)-cyclically monotone set. Beiglböck, Goldstern, Maresch, and Schachermayer [8] generalised the result further by removing regularity assumptions on the cost:

**Theorem 2.6** (See [8, Theorem 1.a]) Let \(X, Y\) be two Polish spaces, let \(\mu \in \mathcal{P}(X)\) and \(\nu \in \mathcal{P}(Y)\). Let \(c: X \times Y \rightarrow [0, \infty]\) be a measurable cost function. Then every finite optimal transport plan is \(c\)-cyclically monotone.

The reverse implication, that \(c\)-cyclic monotonicity implies optimality, is not true in general as shown in Example 3.1 in [1]. In [8] Theorem 1.b, it was shown that for a measurable cost function \(c\) the assumption that the “infinity” set \(\{(x, y) : c(x, y) = \infty\}\) is a union of a closed set and a \(\mu \otimes \nu\)-null set, implies that every finite \(c\)-cyclically monotone plan is optimal.

Finally, the question of the existence of a potential for the optimal plan remains. A result in this direction was presented in [8]; it states that, with assumptions as in Theorem 2.6, a finite cost plan admits a potential if and only if it is “robustly optimal” (see Definition 1.6. in [8]). In particular, their result implies that a plan which admits a potential is optimal. In this note, our main goal is to find conditions on the pairs of measures that guarantee the existence of a potential for the optimal transport plan between them, thus guaranteeing robust optimality by Theorem 2 in [8].

### 3 Compatibility

Given two probability measures, before trying to find an element of \(\Pi(\mu, \nu)\) with some good structure (say, a potential), or an optimal element with respect to the cost, one must figure out whether any element \(\pi \in \Pi(\mu, \nu)\) has a finite cost. Clearly, if the cost function is bounded, we may find a finite cost plan between any pair of measures. However, if the cost admits the value \(+\infty\), an obvious necessary condition for the existence of a finite cost plan is that every set in \(X\) has “enough” points in \(Y\) to which it can be mapped for a finite cost.

In the case of two discrete measures, this necessary condition is also sufficient, which is the subject of Hall’s marriage theorem. We start with this simple case as it gives some intuition for our next steps.

#### 3.1 Starting point: Hall’s marriage theorem

In the following motivating example, for some \((x_i)_{i=1}^m \subset X\) let \(\mu = \sum_{i=1}^m \frac{1}{m} > 1_{x_i}\) be a probability measure on \(X\), and for \((y_i)_{i=1}^m \subset Y\) let \(\nu = \sum_{i=1}^m \frac{1}{m} > 1_{y_i}\) be a probability measure on \(Y\). Let \(c: X \times Y \rightarrow (-\infty, \infty]\) be an arbitrary cost. A finite cost map is a given by a bijection \(T: (x_i)_{i=1}^m \rightarrow (y_i)_{i=1}^m\), such that \(c(x_i, T(x_i)) < \infty\) for all \(i = 1, \ldots, m\). The bijection \(T\) corresponds, of course, to a permutation \(\sigma: [m] \rightarrow [m]\). By Birkhoff’s theorem on the extremal points of bi-stochastic matrices, every transport plan \(\pi \in \Pi(\mu, \nu)\) is a convex combination of permutation maps \(T\).

The condition for the existence of a finite cost map/plan can be thus reformulated in a graph-theoretic way: Let \(G\) be a bipartite graph with a vertex set \(V = (x_i)_{i=1}^m \cup (y_i)_{i=1}^m\) and edges \(E = \{(x_i, y_j) : c(x_i, y_j) < \infty\}\). A finite cost map \(T\) corresponds to a matching in this graph. Hall’s Marriage Theorem gives the necessary and sufficient conditions for such a matching to exist.
Theorem 3.1 (Hall’s Marriage Theorem) A bipartite graph $G$ with a vertex set $V_1 \cup V_2$, such that $|V_1| = |V_2|$, contains a complete matching if and only if $G$ satisfies Hall’s condition

$$|N_G(S)| \geq |S| \text{ for every } S \subset V_1,$$

where $N_G(S) \subset V_2$ is the set of all neighbors of vertices in $S$.

The condition can be reformulated in terms of the measures, as

$$\mu(A) \leq v (\{ y \in Y : \exists x \in A, \ c(x, y) < \infty \})$$

for any $A \subset X$, or, equivalently,

$$\mu(A) + v (\{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}) \leq 1.$$

In fact, in this discrete and finite case, once we have determined the existence of a finite cost map, we may consider, among the finite number of possible matchings, the one with minimal cost (there may, of course, be more than one). It is then not hard to show (and will follow from our results as well) that this resulting optimal plan must lie on a $c$-subgradient of a $c$-class function. (This fact follows from a variation of a theorem of Rüschendorf [18], see also [5].)

3.2 The $c$-compatibility condition

The continuous counterpart for Hall’s condition is an obvious necessary condition for the existence of a finite cost plan.

Definition 3.2 Let $X, Y$ be Polish spaces and $c : X \times Y \to (-\infty, \infty]$ be a measurable cost function. We say that two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ are $c$-compatible if for any measurable $A \subset X$ it holds that

$$\mu(A) + v (\{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}) \leq 1.$$

It is not hard to check that $c$-compatibility is in fact a symmetric notion, and the above condition holds if and only if for any $B \subset Y$ we have

$$v(B) + \mu (\{ x \in X : \forall y \in B, \ c(x, y) = \infty \}) \leq 1.$$

Indeed, to get the latter we let $A = \{ x \in X : \forall y \in B, \ c(x, y) = \infty \}$, in which case $B \subset \{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}$. Applying the assumed inequality, we get

$$v(B) + \mu(A) \leq v (\{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}) + \mu(A) \leq 1.$$

The fact that any plan $\pi \in \Pi(\mu, \nu)$ which has finite cost must be concentrated on the finiteness set

$$S = \{ (x, y) : c(x, y) < \infty \} \subset X \times Y$$

implies the necessity of the condition, as is given in the following lemma.

Lemma 3.3 Let $X, Y$ be Polish spaces and $c : X \times Y \to (-\infty, \infty]$ be a measurable cost function. Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, assume there exists $\pi \in \Pi(\mu, \nu)$ which is concentrated on $S = \{ (x, y) \in X \times Y : c(x, y) < \infty \}$. Then $\mu$ and $\nu$ are $c$-compatible.
Example 3.5 Consider once more the polar cost \( c \) used in the previous example. The \( c \)-compatibility condition is sufficient for the existence of a finite-cost transport plan. In some cases, one may use this together with the theorems stated in Sect. 2.5 and the results from \([5]\) to show that a minimizing plan exists and is concentrated on the graph of a \( c \)-subgradient. An example of such reasoning for some explicit cost functions will appear in the forthcoming \([2]\).

**Theorem 3.4** (Strassen) Let \( X, Y \) be Polish spaces and let \( S \) be a non-empty closed subset of \( X \times Y \). Given \( \mu \in \mathcal{P}(X) \) and \( \nu \in \mathcal{P}(Y) \), there exists \( \pi \in \Pi(\mu, \nu) \) which is supported on \( S \) if and only if for all open \( B \subset Y \)

\[
\nu(B) \leq \mu(P_X(S \cap (X \times B))),
\]

(9)

where \( P_X \) is a projection onto \( X \).

In the case of a non-traditional cost \( c \), the relevant set \( S \) considered in Lemma 3.3 is not necessarily closed. If \( S \) is closed, and \( c \) is bounded on it, then the condition in Strassen’s Theorem is sufficient for the existence of a finite-cost transport plan. In some cases, one may use this together with the theorems stated in Sect. 2.5 and the results from \([5]\) to show that a minimizing plan exists and is concentrated on the graph of a \( c \)-subgradient. An example of such reasoning for some explicit cost functions will appear in the forthcoming \([2]\).

However, for certain important costs, and in particular for the polar cost \( p \) defined in (1) which serves as a motivating example for this study, the set \( S \) of finite-cost pairs is not closed.

To illustrate the problem, let us give an example of two measures on intervals which are \( c \)-compatible (we will use the one dimensional polar cost) but do not admit any plan concentrated on the finiteness set \( S \).

**Example 3.5** Consider once more the polar cost \( p(x, y) = -\ln(xy - 1)_+ \) on \( \mathbb{R}^+ \times \mathbb{R}^+ \). Its finiteness set is \( S = \{(x, y) : xy > 1\} \). Let \( \gamma \) be the uniform measure on the set \( S_1 = \{(x, 1/x) \in \mathbb{R}^2 : x \in [1/2, 2]\} \) and let \( \mu \) be its marginal on the first coordinate and \( \nu \) its marginal on the second coordinate.

It is not hard to check that the measures \( \mu \) and \( \nu \) (which are the same measure) are \( p \)-compatible. Indeed, let \( A \subset \mathbb{R}^+ \) be open, note that

\[
P_X((\mathbb{R}^+ \times A) \cap S) = \bigcup_{y \in A} (1/y, \infty) = (1/\sup(A), \infty).
\]

Additionally, for any number \( \alpha \in [1/2, 2] \) we have, by the definition, that \( \nu([1/2, \alpha]) = \mu([1/\alpha, 2]) \). Combining these observations with the continuity of \( \mu \) and \( \nu \) we see that the measures are polar compatible

\[
\nu(A) \leq \nu([1/2, \sup(A)]) = \mu([1/\sup(A), 2]) = \mu([1/\sup(A), \infty]) = \mu(P_X((A \times X) \cap S)).
\]

We turn to show that there is no transport plan \( \pi \in \Pi(\mu, \nu) \) concentrated on \( S \). Assume towards a contradiction that there exists such a transport plan \( \pi \). In particular, this implies that there exists some open set \( B \subset [x_1, x_2] \times [y_1, y_2] \subset S \) of positive measure (for example, take \( B \) to be a small enough neighborhood of a point in \( \text{supp}(\pi) \)). By the definition of \( S \), we have that \( x_1 y_1 > 1 \). As \( \pi \) is concentrated on \( S \) we see that

\[
\mu([1/2, x_1]) = \pi([1/2, x_1] \times [x_1^{-1}, 2]) \leq \nu([x_1^{-1}, 2]) = \mu([1/2, x_1])
\]
where the last equality follows from the definition of $\mu$ and $\nu$. We thus have equalities all along. Similarly,

$$
\nu([1/2, x_1^{-1}]) = \pi([x_1, 2] \times [1/2, x_1^{-1}]) \leq \mu([x_1, 2]) = \nu([1/2, x_1^{-1}]).
$$

So we conclude that $\pi([1/2, x_1] \times [x_1^{-1}, 2]) + \pi([x_1, 2] \times [1/2, x_1^{-1}]) = \mu([1/2, x_1]) + 
\mu([x_1, 2]) = 1$, that is, $\pi$ is concentrated on $[1/2, x_1] \times [x_1^{-1}, 2] \cup [x_1, 2] \times [1/2, x_1^{-1}])$, which is a contradiction to the fact that $\pi(B) > 0$.

### 3.3 The Hall polytope

Let us consider a special case, which will be the focus of Sect. 5, when one of the measures is discrete and the other one arbitrary. In such a case, the compatibility condition can be realized geometrically by a polytope, which we call the Hall polytope. We use $\Delta_m = \{\alpha \in \mathbb{R}^m : \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1\}$ to denote the $(m-1)$-dimensional simplex.

**Definition 3.6** Let $X$ be a Polish space, and $Y = \{u_i\}_{i=1}^m$ a finite set. Assume $c : X \times Y \to (-\infty, \infty]$ is a measurable cost function, and let $\mu$ be a probability measure concentrated on $\{x \in X : \exists i \in [m], c(x, u_i) < \infty\}$. Define the **Hall polytope** associated with $(u_i)_{i=1}^m$ and $\mu$ by

$$
P = P((u_i)_{i=1}^m, \mu) = \bigcap_{I\subseteq [m]} \left\{ \alpha \in \Delta_m : \sum_{i \in I} \alpha_i \leq \mu(A_I) \right\},
$$

where

$$
A_I := \{ x \in X : \exists i \in I, c(x, u_i) < \infty \}.
$$

Note that the definition implies that $\mu$ and $\nu = \sum_{i=1}^m \alpha_i \mathbb{1}_{u_i}$ are $c$-compatible if and only if $\alpha \in P((u_i)_{i=1}^m, \mu)$.

We get back to this definition, and present a careful study of the resulting polytopes, in Sect. 5.

### 3.4 Strong $c$-compatibility

We saw in Example 3.5 that $c$-compatibility is not a sufficient condition for the existence of a finite cost plan. In fact, we will see in Example 7.2 that there exist $c$-compatible measures
which do admit a finite cost plan but not a potential. Therefore, we consider a slight strengthening of \(c\)-compatibility, which will ensure that the existence of a finite cost plan implies the existence of a potential. We call this condition strong \(c\)-compatibility, and it amounts to asking for a strict inequality in the defining inequalities.

**Definition 3.7** Let \(X, Y\) be Polish spaces and \(c : X \times Y \to (-\infty, \infty]\) be a measurable cost function. We say that two probability measures \(\mu \in \mathcal{P}(X)\) and \(\nu \in \mathcal{P}(Y)\) are **strongly \(c\)-compatible** if they are \(c\)-compatible and for any measurable \(A \subset X\) with \(0 < \mu(A) < 1\) it holds that

\[
\mu(A) + \nu(\{y \in Y : \forall x \in A, \ c(x, y) = \infty\}) < 1.
\]

The motivation for this specific strengthening of the condition of \(c\)-compatibility is twofold: First, if two measures are \(c\)-compatible and not strongly \(c\)-compatible, this means that there exists a decomposition of the transport problem into two sub-problems (see Sect. 7). Indeed, this is quite clear from the definition: if some set \(A\) of measure \(\mu(A) \in (0, 1)\) satisfies the equality

\[
\mu(A) + \nu(\{y \in Y : \forall x \in A, \ c(x, y) = \infty\}) = 1,
\]

then letting \(B = \{y \in Y : \forall x \in A, \ c(x, y) = \infty\}\) we see that \(A\) must be mapped to \(Y \setminus B\) (and they have the same measure) and the preimage of \(B\) must be \(X \setminus A\). That is, the original transport problem is in fact decomposed into two disjoint transport problems.

Second, in the discrete setting of Sect. 3.3, strong \(c\)-compatibility corresponds to the weight vector \(\alpha\) residing in the interior of the Hall polytope, which makes for an elegant assumption.

We stress that strong \(c\)-compatibility is not a necessary condition, only \(c\)-compatibility is. Even if one of the measures is discrete, it could be that the Hall polytope has an empty interior, but good transport maps, admitting a potential, exist.

### 3.5 The geometric meaning of strong \(c\)-compatibility

It will be very useful to rephrase the condition of strong \(c\)-compatibility in terms that are more geometric. In fact, looking back at the proof of the symmetry of the notion of \(c\)-compatibility, it seems evident that we do not need to assume an inequality \(\mu(A) + \nu(\{y \in Y : \forall x \in A, \ c(x, y) = \infty\}) \leq 1\) (or a strict inequality, in the strong \(c\)-compatibility assumption) for all sets \(A\), and it suffices to consider sets of the form \(\{x \in X : \forall y \in B, \ c(x, y) = \infty\}\). To make this observation more precise, we introduce the notion of the \(c\)-dual of a set.

**Definition 3.8** (**\(c\)-duality**) Let \(X, Y\) be two sets and let \(c : X \times Y \to (-\infty, \infty]\). Fix \(t \in (-\infty, \infty]\) (which will be omitted in the notation as it is a fixed parameter). For \(K \subset X\) define the **\(c\)-dual set** of \(K\) as

\[
K^c = \bigcap_{x \in K} \{y \in Y : c(x, y) \geq t\} = \left\{y \in Y : \inf_{x \in K} c(x, y) \geq t\right\}.
\]

It will be convenient to assume \(X = Y\) and that the cost is symmetric, and as this is the case relevant for this note, we restrict to this case. However, the reader will find it easy to generalize to the case where \(X \neq Y\), in which case there are two different “\(c\)-duality” operations, one mapping sets in \(X\) to sets in \(Y\), and one mapping sets in \(Y\) to sets in \(X\), similarly to the \(c\)-transform.

Let us point out that for the polar cost \(p(x, y) = -\ln((x, y) - 1)_+\) and \(t = \infty\), the set \(K^p\) is the well known polar set \(K^\circ\). Indeed, we have that \(\inf_{x \in K} p(x, y) = \infty\) if and only
if \( \sup_{x \in K} \langle x, y \rangle \leq 1 \). For the classical cost \( c(x, y) = -\langle x, y \rangle \) and \( t = -1 \), we also get the polarity map.

**Remark 3.9** If in Definition 3.8 one adds the assumptions that \( X \) and \( Y \) are Polish spaces and that the cost is upper semi-continuous, it follows that for any fixed \( x \) the set \( \{ y \in Y : c(x, y) \geq t \} \) is closed, and hence so is \( K^c \).

Having defined an operation on sets, let us notice some basic properties.

**Lemma 3.10** For every \( K, L \subseteq X \), the following hold

(i) \( K \subseteq (K^c)^c = K^{ccc} \),

(ii) if \( L \subseteq K \) then \( K^c \subseteq L^c \),

(iii) \( K^c = K^{ccc} \).

**Proof** (i) This follows directly from the definition. If \( x \in K \) and \( y \in K^c \) then \( c(x, y) \geq t \) so that \( x \in K^{cc} \).

(ii) Assume that \( L \subseteq K \), and \( y \in K^c \), then \( c(x, y) \geq t \) for all \( x \in K \) and in particular for all \( x \in L \), so \( y \in L^c \).

(iii) From (i) we know that \( K \subseteq K^{ce} \), so from (ii) we get \( K^c \supseteq K^{ce} \). On the other hand, applying (i) directly to \( K^c \) we get \( K^{cc} \supseteq K^c \), and equality is obtained.

The similarity of \( c \)-duality to the \( c \)-transform is apparent. We are thus motivated to define the \( c \)-class of sets, on which the \( c \)-duality is an order reversing bijection. In order to avoid confusion, as we suppressed \( t \) in the notation, we restrict the next definition to \( t = \infty \), the case relevant for this note.

**Definition 3.11** (\( c \)-class and \( c \)-envelope) Fix \( t = \infty \). The \( c \)-class of sets consists of all closed sets \( K \subseteq X \) such that there exists some \( L \subseteq X \) with \( K = L^c \). For any set \( K \subseteq X \) we define its \( c \)-envelope as the set \( K^{cc} \), which is the smallest \( c \)-class set containing \( K \).

Let us note again that for the polar cost and \( t = \infty \), the \( p \)-class consists of closed convex sets containing the origin, and the \( p \)-envelope is the polar convexification operation \( K \mapsto K^{oo} = \text{conv}[0, K] \).

Our first observation is that in Definitions 3.2 and 3.7 it is sufficient to consider \( c \)-class sets, for \( t = \infty \), instead of all measurable sets.

**Lemma 3.12** Let \( X, Y \) be Polish spaces. Let \( c : X \times Y \to (-\infty, \infty] \) be an upper semi-continuous symmetric cost function. Two probability measures \( \mu \in \mathcal{P}(X) \) and \( v \in \mathcal{P}(Y) \) are \( c \)-compatible if and only if for every set \( K = K^{cc} \subseteq X \) in the \( c \)-class we have

\[

v(K^c) \leq 1 - \mu(K).

\]

They are strongly \( c \)-compatible if and only if in addition when \( v(K^c) \neq 0, 1 \) we have

\[

v(K^c) < 1 - \mu(K).

\]

**Proof** If \( \mu \) and \( v \) are \( c \)-compatible then in particular \( v(K^c) \leq \mu(\{ x : \inf_{y \in K^c} c(x, y) < \infty \}) \), which can be rewritten as \( v(K^c) \leq \mu(X \setminus K^{cc}) = \mu(X \setminus K) \).

For the other direction let \( A \subseteq X \) be a measurable set, and consider the set \( K = A^{cc} \). Then

\[

\{ y \in Y : \inf_{x \in A} c(x, y) = \infty \} = \{ y \in Y : \forall x \in A \ c(x, y) = \infty \} = A^c = K^c.

\]
The last equality holds due to Lemma 3.10 (iii). Thus, using the condition on $c$-class sets and Lemma 3.10 (i), we get

$$
\nu(A^c) = \nu(K^c) \leq 1 - \mu(K) = 1 - \mu(A^{cc}) \leq 1 - \mu(A),
$$

so that $\mu$ and $\nu$ are $c$-compatible.

We now turn to show that two probability measures $\mu$ and $\nu$ are strongly $c$-compatible if and only if they are $c$-compatible and for all $c$-class sets $K \subset X$ such that $\nu(K^c) \neq 0, 1$ we have

$$
\nu(K^c) < 1 - \mu(K).
$$

One direction is again immediate, as in the case of $c$-compatibility.

For the other direction, let $A \subset X$ with $\mu(A) \neq 0, 1$. If $\nu(A^c) = 0$ then indeed $\mu(A) + \nu(A^c) < 1$. Otherwise, $\nu(A^c) > 0$ and by $c$-compatibility, $\nu(A^c) \leq 1 - \mu(A) < 1$. Therefore, $\nu(A^c) \neq 0, 1$ and we denote $K = A^{cc}$. The set $K$ is in the $c$-class and satisfies $\nu(K^c) = \nu(A^c) \neq 0, 1$. Hence the strong $c$-compatibility assumption on $K$, together with the fact that $A \subset A^{cc}$, gives

$$
\nu(A^c) = \nu(K^c) < 1 - \mu(K) = 1 - \mu(A^{cc}) \leq 1 - \mu(A).
$$

\[\square\]

In the next lemma we show that the strong $c$-compatibility of two measures implies a vital condition on the distribution of the transport plan between them.

**Lemma 3.13** Let $X = Y$ be a Polish space, let $c : X \times Y \to (-\infty, \infty]$ be a symmetric and upper semi-continuous cost function. Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and let $\pi \in \Pi(\mu, \nu)$ be a finite cost plan. Then $\mu$ and $\nu$ are strongly $c$-compatible if and only if for every $c$-class set $K$ such that $\nu(K^c) \neq 0, 1$, we have that

$$
\pi((X \setminus K) \times (Y \setminus K^c)) > 0.
$$

**Proof** First we note that the existence of a finite cost plan $\pi \in \Pi(\mu, \nu)$ implies $c$-compatibility (see Lemma 3.3). Thus, under our assumptions, strong $c$-compatibility is equivalent, by Lemma 3.12, to the fact that for every $c$-class $K$ with $\nu(K^c) \neq 0, 1$ we have that $\nu(K^c) < 1 - \mu(K) = \mu(X \setminus K)$. Since $\pi \in \Pi(\mu, \nu)$ this can be rewritten as, for $\nu(K^c) \neq 0, 1$,

$$
\pi(X \times K^c) = \pi((X \setminus K) \times Y),
$$

and if $\nu(K^c) = 1$ then $\mu(K) = 0$ (due to $c$-compatibility). Note that as $\pi$ has finite cost, it is concentrated on $S = \{(x, y) : c(x, y) < \infty\}$, and for $(x, y) \in S$, if $y \in K^c$ then we must have $x \notin K$. In particular, from the point of view of the measure $\pi$, the set on the left hand side is contained in the set on the right hand side. We can thus rewrite the first inequality as

$$
0 < \pi(((X \setminus K) \times Y) \setminus (X \times K^c)) = \pi((X \setminus K) \times (Y \setminus K^c)).
$$

completing the proof of the statement claimed. \[\square\]

### 4 Transportation of measure

Let us recall our main theorem, to be proved in this section.
Theorem 1.1 Let $X = Y$ be a Polish space, and $c : X \times Y \to (-\infty, \infty]$ be a continuous and symmetric cost function, essentially bounded from below with respect to $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Assume $\mu$ and $\nu$ are strongly $c$-compatible, and $C(\mu, \nu) < \infty$. Then there exists a $c$-class function $\varphi$ and an optimal transport plan $\pi \in \Pi(\mu, \nu)$ concentrated on $\partial^c \varphi$.

In order to prove the theorem we will use a combination of Theorems 2.3, 2.4, and 2.6. We will show that once we have an optimal transport plan supported on a $c$-cyclically monotone set then it must be $c$-path-bounded. This will follow from an observation presented in [5] which states that indeed in some special cases $c$-cyclic monotonicity implies $c$-path-boundedness. In order to formulate the condition let us introduce some notation.

We consider a directed graph, associated with a cost function $c : X \times Y \to (-\infty, \infty]$ and a set $G \subset S = \{(x, y) : c(x, y) < \infty\} \subset X \times Y$, in which the vertices are elements of $G$ and there is a directed edge from $(x, y)$ to $(z, w)$ if $c(z, y) < \infty$. Since $G \subset S$ we may say that for every point in $G$ there is an edge (loop) with this point as a start and end vertex.

The directed graph induces a (transitive) relation on points in $G$, namely $(x, y) \prec (z, w)$ if there is a directed path from $(x, y)$ to $(z, w)$. We then define an equivalence relation $\sim$ on elements of $G$ where we say that $(x, y) \sim (z, w)$ if $(x, y) \prec (z, w)$ and $(z, w) \prec (x, y)$, i.e., there is a directed cycle passing through both points. To the best of our knowledge, this equivalence relation was first mentioned in [23, Chapter 5, p.75] and studied in [5, 8, 9]. The following proposition was proved (with a different formulation) in [8] and then in [5].

Proposition 4.1 Let $X, Y$ be sets, let $c : X \times Y \to (-\infty, \infty]$ be some cost function and let $G \subset X \times Y$ be a $c$-cyclically monotone set. Assume that all points in $G$ belong to one equivalence class of the equivalence relation $\sim$ defined above. Then $G$ is $c$-path-bounded.

With this proposition in hand, our goal is to show that if $\pi$ is a finite cost plan between two strongly $c$-compatible measures, then we can find a set $G$, on which $\pi$ is concentrated, such that all of points in $G$ are in one equivalence class of $\sim$.

Proposition 4.2 Let $X = Y$ be a Polish space, let $c : X \times Y \to (-\infty, \infty]$ be a continuous and symmetric cost function, and let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ be strongly $c$-compatible. Let $\pi \in \Pi(\mu, \nu)$ be a finite cost transport plan from $\mu$ to $\nu$. Then there exists a set $G$ on which $\pi$ is concentrated such that all the points in $G$ are in one equivalence class of $\sim$.

Proof Let $G_1$ denote the support of $\pi$, and let $G_0$ denote the set $G_1 \cap \{(x, y) \in X \times Y : c(x, y) < \infty\}$. Since $\pi$ has finite total cost, $\pi(G_0) = 1$. Fix a point $(x, y) \in G_0$. We shall show that the set of points $(z, w) \prec (x, y)$ in $G_0$ is of $\pi$-measure one, as is the set of points $(z, w)$ such that $(x, y) \prec (z, w)$. The intersection of these two sets will also be of measure one, and we denote it by $G$. We will then explain why this $G$ fulfills the requirements of the proposition.

Consider $H \subset G_0$ consisting of all points $(a, b) \prec (x, y)$. Assume towards a contradiction that $\pi(H) < 1$. Note that $\pi(H) > 0$ since $(x, y) \in G_0$ is in the support of $\pi$, so that for any neighborhood $U$ of $(x, y)$ we have $\pi(U) > 0$. Picking a small enough neighborhood $U$, we know that if $(z, w) \in U \cap G_0$ then $c(z, y) < \infty$ and so $(z, w) \in H$ (it may be that $\pi(\{(x, y)\}) > 0$ and that $U$ consists of this one point alone).

Since $\pi(H) < 1$, $\pi(G_0 \setminus H) > 0$, so there is some $(z, w) \in G_0 \setminus H$ which is in particular in the support of $\pi$, that is, for any neighborhood $V$ of $(z, w)$ one has $\pi(V) > 0$.

If $z \in (P_X H)^cc$ then $(P_X H)^cc = (P_X H)^ccc \subset \{z\}^c$ by Lemma 3.10(ii) and (iii). Therefore $Y \setminus \{z\}^c \subset Y \setminus (P_X H)^cc = \{v : \exists w \in P_X H, c(u, v) < \infty\}$. Further, since $(z, w) \in G_0$ we have that $c(z, w) < \infty$ and hence $w \in Y \setminus \{z\}^c$. But this means that $w \in Y \setminus (P_X H)^cc$ and
there exists some \( a \in PXH \) (and \( b \) such that \((a, b) \in G_0\)) with \( c(a, w) < \infty \). Therefore \((z, w) < (a, b)\), and by transitivity \((z, w) < (x, y)\), a contradiction.

We may therefore assume that \((z, w)\) is such that \(z \notin (PXH)^c\); since \(K\) is a closed set in \(X\), there is a neighborhood of \(z\) which does not intersect \(K\), and therefore we can find a neighborhood \(V\) of \((z, w)\) which is of positive \(\pi\) measure (as \((z, w)\) is a density point) and such that its projection onto \(X\) does not intersect \(K\). Note that this implies in particular that \(0 < \mu(K) = \pi(K \times Y) < 1\), since \(H \subseteq K \times Y\) and \(V \cap (K \times Y) = \emptyset\). We may therefore use Lemma 3.13 to deduce that

\[
\pi((X \setminus K) \times (Y \setminus K^c)) > 0.
\]

In particular, there exists some point \((e, f) \in G_0\) Such that \(e \notin K\) and \(f \notin K^c\). The fact that \(e \notin K\) means in particular that \((e, f) \notin H\). The fact that \(f \notin K^c = (PXH)^{ccc} = (PXH)^c\) implies that \(f \in \{v : \exists u \in PXH \ c(u,v) < \infty\}\). Hence there is some point \(a \in PXH\) (and \(b\) such that \((a, b) \in H\)) such that \(c(a, f) < \infty\), which means that \((e, f) < (a, b) < (x, y)\), thus contradicting the fact that \((e, f) \notin H\). We conclude that the set \(H\) satisfies \(\pi(H) = 1\).

Similarly we consider \(F \subseteq G_0\) consisting of all points \((a, b)\) such that \((x, y) < (a, b)\). Using the same argument as above we get that \(\pi(F) = 1\).

Hence, we found sets \(F\) and \(H\) of \(\pi\)-measure one. Let \(G = F \cap H\), every point \((z, w) \in G\) satisfies that there is a directed path, going through points in \(G_0\), between it and \((x, y)\). We now claim that these directed paths only go through points in \(G\) itself. Indeed, consider a cycle (in \(G_0\)) which includes \((x, y)\) and \((z, w) \in G\). The existence of this cycle implies that every point on it belongs to both \(H\) and \(F\), by the definition of the relation \(\sim\), so that the whole cycle consists of points in \(G\). The proof is now complete. \(\square\)

**Proof of Theorem 1.1** By assumption, \(C(\mu, v) < \infty\), and we may use Theorem 2.4, the assumptions of which are satisfied, to find a \(c\)-optimal plan \(\pi \in \Pi(\mu, v)\). By Theorem 2.6, the plan \(\pi\) is concentrated on some \(c\)-cyclically monotone set \(G_1\). Proposition 4.2 implies that \(\pi\) is also concentrated on some set \(G_2\) such that all points in \(G_2\) are in one equivalence class of the relation \(\sim\) defined above. Let \(G = G_1 \cap G_2\), then \(G\) is a \(c\)-cyclically monotone set such that all of its elements lie in one equivalence class, therefore, by Proposition 4.1 the set \(G\) is \(c\)-path-bounded. Finally, we use Theorem 2.3 which implies that a \(c\)-path-bounded set admits a potential, to find some \(c\)-class function \(\varphi\) such that \(G \subseteq \partial c\varphi\).

We have thus determined that there exists a \(c\)-optimal plan \(\pi\) which is concentrated on \(\partial c\varphi\) for some \(c\)-class \(\varphi\), as needed. \(\square\)

## 5 Transportation to a discrete measure

In this section we present a different approach to the problem of finding transport maps which lie on \(c\)-subgradients of functions. We consider the case where one measure is arbitrary (we will add some mild assumptions on it, related with the cost, later on) and the second measure is discrete. As explained in Sect. 3.3, fixing the support of \(v\) to be the set \(\{y_i\}_{i=1}^m\), a necessary condition for the existence of a finite cost plan \(\pi \in \Pi(\mu, v)\) is that the weight vector \(\alpha \in \Delta_m\) associated with the probability measure \(v = \sum_{i=1}^m \alpha_i \delta_{y_i}\) lies in the Hall polytope

\[
P = P((u_i)_{i=1}^m, \mu) = \cap_{I \subseteq [m]} \left\{ \alpha \in \Delta_m : \sum_{i \in I} \alpha_i \leq \mu(A_I) \right\},
\]
where
\[ A_I := \left\{ x \in X : \min_{i \in I} c(x, u_i) < \infty \right\}. \]

So, our main objective is to show that indeed, for a measure \( \nu \) corresponding to a weight vector in the polytope, a finite cost transport plan exists, and further, it is supported on the \( c \)-subgradient of some \( c \)-class function. We are able to do this under very general assumptions on the measure \( \mu \), and provided that \( \alpha \) lies in the interior of the polytope (this is Theorem 1.3). Let us introduce the notion of \( c \)-regularity of a measure, which will be important for the construction given in this section. Roughly speaking, a measure is \( c \)-regular if it gives 0-measure to sets where two different basic functions \( c(\cdot, y_1) + a_1 \) and \( c(\cdot, y_0) + a_0 \), coincide and equal some finite number.

**Definition 5.1** Let \( X, Y \) be Polish spaces and let \( c : X \times Y \to (-\infty, \infty] \) be a measurable cost function, and \( \mu \) a probability measure on \( X \). If for any \( y_1 \neq y_0 \in Y \) and \( t \in \mathbb{R} \)
\[ \mu \left( \{ z \in X : c(z, y_1) - c(z, y_0) = t \} \right) = 0, \]
then we say that \( \mu \in \mathcal{P}(X) \) is a \( c \)-regular measure.

For example, when the cost \( c \) on \( \mathbb{R}^n \times \mathbb{R}^n \) is such that \( \{ z : c(z, y_1) - c(z, y_0) = t \} \) is of lower dimension, and the measure is absolutely continuous, the \( c \)-regularity property is satisfied.

### 5.1 Building transport maps

The idea of the proof is to manually construct functions whose \( c \)-subgradient is a transport map of a \( c \)-regular measure \( \mu \) to a certain discrete measure \( \nu \). We will consider basic functions and use the fact that the \( c \)-class is closed under the pointwise infimum. Formally, we have the following lemma.

**Lemma 5.2** Let \( X, Y \) be Polish spaces and let \( c : X \times Y \to (-\infty, \infty] \) be a measurable cost function. Fix a set of vectors \( (u_i)_{i=1}^m \subset Y \) and let \( \mu \) be a \( c \)-regular probability measure on \( X \), which is concentrated on the set \( \{ x \in X : \exists i \in [m] \text{ with } c(x, u_i) < \infty \} \). Given numbers \( (t_i)_{i=1}^m \subset \mathbb{R} \) let
\[ \varphi(x) = \varphi(u_\cdot, t_\cdot)(x) = \min_{1 \leq i \leq m} \left( c(x, u_i) + t_i \right) \]
be a function in the \( c \)-class and denote \( U_i = \{ x \in X : \arg\min_{1 \leq j \leq m} \left( c(x, u_j) + t_j \right) = i \} \). Then the mapping \( T \), defined to be equal to \( u_i \) on the set \( U_i \), is well defined \( \mu \)-almost everywhere and satisfies that \( (x, T(x)) \in \partial^c \varphi \) for \( \mu \)-almost every \( x \). Moreover, it transports \( \mu \) to the measure \( \nu = \sum_{i=1}^m \alpha_i \cdot 1_{U_i} \) on \( Y \), where \( \alpha_i = \mu(U_i) \).

**Proof** Let \( \varphi(x) \) be the function defined in the statement and note that it induces a partition of \( X \) into \( m \) sets \( \{U_i\}_{i=1}^m \), where
\[ U_i = \left\{ x \in X : \arg\min_{1 \leq j \leq m} \left( c(x, u_j) + t_j \right) = i \right\}. \]
By the definition of the \( c \)-subgradient given in (7), \( (x, u_i) \in \partial^c \varphi \) for all \( x \in U_i \). Equivalently, we write \( u_i \in \partial^c \varphi(x) \) (namely \( \partial^c \varphi(x) = \{ y \in Y : (x, y) \in \partial^c \varphi \} \)). Let \( T : X \to Y \) be the map given by \( T(x) = u_i \) for all \( x \in U_i \), so indeed \( T(x) \in \partial^c \varphi(x) \). For \( \mu \) which is
c-regular, the intersections of the sets $U_i$ are of zero measure and thus $T$ is well defined $\mu$ almost everywhere.

Clearly, the map $T$ transports the measure $\mu$ on $X$ to the measure $\sum_{i=1}^{m} \mu(U_i) \geq 1_{U_i}$. \hfill $\square$

**Remarks 5.3** (i) In general, the partition to sets $\{U_i\}$ as above is not disjoint, so without the additional assumption of $c$-regularity of $\mu$ the map $T$ is not well-defined.

(ii) Since we may add a constant to all $(t_i)_{i=1}^{m}$ without changing the $c$-subgradient, we will assume that $t_i \geq 0$. Thus, given a finite set $(u_i)_{i=1}^{m} \subset Y$, it will be convenient for us to consider the family of functions

$$
\varphi(x) = \min_{1 \leq i \leq m} (c(x, u_i) - \ln(t_i)),
$$

with $t = (t_i)_{i=1}^{m}$ in the $m$-dimensional simplex $\Delta_m$.

Lemma 5.2 guarantees that given a $c$-regular measure $\mu$ and points $(u_i)_{i=1}^{m} \subset Y$, the map $\varphi(u_i), (t_i)$ induces a transport map $T : X \to Y$ mapping $\mu$ to $\sum \alpha_i \geq 1_{U_i}$. This simple idea will be very important in proving Theorem 1.3, and the bulk of the proof lies in analyzing which $\alpha$ cones are spanned by $\{(u_i)_{i=1}^{m}\}$. Fixing a measure $\mu$ and an $m$-tuple $(u_i)_{i=1}^{m}$, the construction in Lemma 5.2 gives rise to a mapping $H$ from the $(m-1)$-dimensional simplex $\Delta_m$ onto the set of ‘weight vectors’ $\alpha = (\alpha_i)_{i=1}^{m}$ of the measure $\nu$ to which $\mu$ can be transported. We will show that $H$ is a surjection from the interior of the simplex onto the interior of the relevant Hall polytope. To this end we define and analyze Hall polytopes, and in particular construct, under some assumptions, a continuous map $R$ from the boundary of the polytope to the boundary of the simplex, which respects certain constraints connected with the face structure of the polytope. We use a variant of Brouwer’s fixed point theorem for the composition $R \circ p \circ H$, where $p$ is a radial projection from some point in the polytope, to obtain the surjectivity.

### 5.2 Structure of the Hall polytope

We introduce the following notation: For $I \subset [m]$, $A \subset \mathbb{R}^{\lvert I \rvert}$ and $B \subset \mathbb{R}^{m-\lvert I \rvert}$, we denote by $A \times_{I} B$ points in $\mathbb{R}^{m}$ with $I$-coordinates in $A$ and $I^{c}$-coordinates in $B$. For a measure $\mu$ and a set $A \subset X$ we denote by $\mu_{\mid A}$ the measure that is equal to $\mu$ on $A$ and zero on $A^{c}$.

Hall polytopes have faces only in specific pre-determined directions. (Their faces’ normal cones are spanned by $[0, 1)$-vectors in $\mathbb{R}^{m}$, projected onto the span of the polytope which is $(m-1)$-dimensional.) As we shall see in Proposition 5.4, each of these faces has a product structure, of which each component is a Hall polytope itself.

Throughout this section $X$ and $Y$ will be Polish spaces and $c : X \times Y \to (-\infty, \infty]$ a measurable cost function.

**Proposition 5.4** Let $P = P((u_i)_{i=1}^{m}, \mu)$ be the Hall polytope associated with some $m$-tuple $(u_i)_{i=1}^{m} \subset Y$ and a probability measure $\mu \in \mathcal{P}(X)$ concentrated on $\{x \in X : \exists i \in$
Then for each $I \subset [m]$, the face of $P$ given by

$$F_I = \left\{ \alpha \in P : \sum_{i} \alpha_i = \mu(A_I) \right\}, \quad (10)$$

admits a splitting $F_I = \mu(A_I) P_I \times_I \mu(A^c_I) \hat{P}_I$ where $P_I$ is the Hall polytope associated with the measure $\frac{1}{\mu(A_I)} \mu|_{A_I}$ and the vectors $(u_i)_{i \in I}$, and $\hat{P}_I$ is the Hall polytope associated with the measure $\frac{1}{\mu(A^c_I)} \mu|_{A^c_I}$ and the vectors $(u_i)_{i \in [m]\setminus I}$. In particular, in case $\mu(A_I) = 0$, we have $P_I = \{0\}_I$, and in case $\mu(A^c_I) = 0$, $\hat{P}_I = \{0\}_{[m]\setminus I}$.

**Proof** Let $\alpha \in F_I$, then by the definition of $F_I$ we have $\sum_{i \in I} \alpha_i = \mu(A_I)$ and thus $\alpha|_I \in \mu(A_I) \Delta_{|I|}$. Furthermore, for every $J \subset I$ it still holds that $\sum_{i \in J} \alpha_i \leq \mu(A_J)$ and as $A_J \subset A_I$ we also get that $\sum_{i \in J} \alpha_i \leq \mu(A_I \cap A_J)$. Recall that $P_I$ is the Hall polytope associated with $(u_i)_{i \in I}$ and $\mu(A_I)^{-1} \mu|_{A_I}$, so re-normalizing the previous inequalities by $\mu(A_I)$ we see that the vector $\alpha|_I \in \mu(A_I) P_I$, as claimed.

Similarly in the $I^c$ coordinates, $\alpha \in F_I$ satisfies $\sum_{i \in [m]\setminus I} \alpha_i = \mu(A^c_I)$. To show $\alpha|_{[m]\setminus I} \in \mu(A^c_I) \hat{P}_I$ we need to check that for every $K \subset [m] \setminus I$ we have $\sum_{i \in K} \alpha_i \leq \mu(A_K \cap A^c_I)$. To this end consider the new subset of $[m]$ given by $J = I \cup K$. By the assumptions,

$$\sum_{i \in J} \alpha_i \leq \mu(A_J) = \mu \left( \bigcup_{i \in I} \{x \in X : c(x, u_i) < \infty\} \cup \bigcup_{i \in K} \{x \in X : c(x, u_i) < \infty\} \right).$$

Since the first of these unions is in fact all of $A_I$, we may rewrite the inequality as

$$\sum_{i \in J} \alpha_i \leq \mu(A_I) + \mu \left( A^c_I \cap \bigcup_{i \in K} \{x \in X : c(x, u_i) < \infty\} \right).$$

The sum on the left hand side is simply $\sum_{i \in I} \alpha_i + \sum_{i \in K} \alpha_i = \mu(A_I) + \sum_{i \in K} \alpha_i$, since we have assumed $\alpha \in F_I$. Plugging into the inequality and canceling, we see

$$\sum_{i \in K} \alpha_i \leq \mu \left( A^c_I \cap \bigcup_{i \in K} \{x \in X : c(x, u_i) < \infty\} \right),$$

as claimed.

We have thus shown, so far, that $F_I \subset \mu(A_I) P_I \times_I \mu(A^c_I) \hat{P}_I$. For the opposite direction, assume we are given some point $\alpha \in \mu(A_I) P_I \times_I \mu(A^c_I) \hat{P}_I$, and we want to show that it belongs to $F_I$. Clearly, using that if $K \subset J$ then $A_K \subset A_J$, we have for any $J \subset [m]$ that

$$\sum_{i \in J} \alpha_i = \sum_{i \in J \cap I} \alpha_i + \sum_{i \in J \cap ([m]\setminus I)} \alpha_i \leq \mu|_{A_J \cap I} + \mu|_{A^c_I} (A_J \cap ([m]\setminus I)) \leq \mu|_{A_J} (A_I) + \mu|_{A^c_I} (A_J) = \mu(A_J).$$

This completes the second part of the proof. \(\square\)

**Remark 5.5** The above proposition also follows from the more general Proposition 7.1, using that if $\alpha \in F_I$ one can decompose both measures in the way explained in Sect. 7.

We will discuss the facial structure of the polytope, and make use of the following simple observation.
Lemma 5.6 Under the conditions and notations of Lemma 5.4, for any $I \subset [m]$, the part of the boundary of $F_I$ given by $\mu(A_I)\partial P_I \times_1 \mu(A^c_I) \hat{P}_I$ is a subset of $\bigcup_{J \subset I} F_J$.

Proof The boundary of $P_I$ consists of points whose $I^{th}$ coordinates add up to one, and for some $J \subset I$ one of the inequalities defining the Hall polytope associated with $\mu$ and $(u_i)_{i \in I}$ is an equality. In other words, if $\alpha \in \mu(A_I)\partial P_I \times_1 \mu(A^c_I) \hat{P}_I$ there is some $J \subset I$ such that $\alpha_i = \mu(A_I)$, which means $\alpha \in F_J$, as claimed. □

5.3 Non-degenerate polytopes

In this subsection we continue analyzing properties of Hall polytopes, under an additional assumption on $\mu$ and $(\alpha_i)_{i=1}^m$ which will imply that all of the Hall polytopes’ faces $F_I$ (defined in (10)) are ‘full dimensional’ in the $I$ coordinates, i.e. that in the splitting described in Proposition 5.4, the polytope $P_I$ is $|I| - 1$ dimensional (Fig. 2).

Definition 5.7 Let $X$, $Y$ be Polish spaces and let $c : X \times Y \to (\infty, \infty]$ be a measurable cost function. Given $(u_i)_{i=1}^m \subset Y$, and a probability measure $\mu$ which is concentrated on $\{x \in X : \exists i \in [m], c(x, u_i) < \infty\}$, we say that $\mu$ is non-degenerate with respect to $(u_i)_{i=1}^m$ if for every $1 \leq i < j \leq m$ it holds that

$$
\mu\left(\{x \in X : c(x, u_i) < \infty\} \cap \{x \in X : c(x, u_j) < \infty\}\right) > 0.
$$

Proposition 5.8 Given a probability measure $\mu \in \mathcal{P}(X)$, which is non-degenerate with respect to $(u_i)_{i=1}^m \subset Y$, the Hall polytope $P = P((u_i)_{i=1}^m, \mu)$ satisfies that its dimension (meaning the dimension of its affine hull) is $\dim(P) = m - 1$.

Proof We shall prove this fact using induction on $m$. For $m = 1$ this is clearly true since the polytope $P$ consists of one point $a_1 = 1$, that is, has dimension 0. Assume that the claim is true for $(m-1)$-tuples. Then, for $m$ and a given set of vectors $(u_i)_{i=1}^m \subset Y$, we know by Proposition 5.4 that $P$ has faces $F_I$ where $I \subset [m]$, each of the form $F_I = \mu(A_I)P_I \times_1 \mu(A^c_I) \hat{P}_I$. Let $F_1 = F_{[m-1]}$ and $F_2 = F_{[m]}$. The set $[m-1]$ still satisfies, along with $\mu|_{A_{[m-1]}}$, the conditions of the proposition, so by the inductive assumption $P_{[m-1]}$ is a polytope of full dimension, that is, of dimension $m - 2$.

It remains to show that $F_2$ does not lie within the affine hull of $F_1$, and hence $P$ has dimension at least $m - 1$ (and of course it cannot have a higher dimension, as it is a subset of $\Delta_m$). Note that the affine hull of $F_1$ is characterized by the equality $\sum_{i=1}^{m-1} \alpha_i = \mu(A_{[m-1]})$, which equivalently can be written as $\alpha_m = 1 - \mu(A_{[m-1]})$. The face $F_2$ satisfies $\alpha_m =
Corollary 5.9 Given a probability measure $\mu \in \mathcal{P}(X)$ which is non-degenerate with respect to $(u_i)_{i=1}^m \subset Y$, the Hall polytope $P = P((u_i)_{i=1}^m, \mu)$ satisfies that each face $F_I$ admits a splitting $F_I = \mu(A_I)\, P_I \times_I \mu(A_J^c)\, \hat{P}_I$ such that $\dim(P_I) = |I| - 1$.

Proof The fact that $F_I$ has such a splitting was proven already in Proposition 5.4, with $P_I$ being the Hall polytope of the normalized restriction of the measure $\mu$ to $A_I$. $\mu_I$ satisfies, together with the subset $(u_i)_{i \in I}$, conditions of Proposition 5.8, namely that it is non-degenerate with respect to $(u_i)_{i \in I}$ as one may easily check that $\mu|_{A_J}$ is non-degenerate with respect to $(u_i)_{i \in I}$. Therefore, $P_I$ is full dimensional, as claimed.

Furthermore, in this case the associated polytope satisfies a “good” face-structure situation, explained in the next two propositions.

Proposition 5.10 Given a probability measure $\mu \in \mathcal{P}(X)$ concentrated on $\{x \in X : \exists i \in [m], c(x, u_i) < \infty \}$ which is non-degenerate with respect to $(u_i)_{i=1}^m \subset Y$, let $P = P((u_i)_{i=1}^m, \mu)$ be the associated Hall polytope. Given $I, J \subset [m]$, the intersection $F_I \cap F_J$ is a subset of $F_{I \cap J}$ (we let $F_{\emptyset} = P$, so that if $I \cap J = \emptyset$ the claim is trivial).

Proof Let $\alpha \in F_I \cap F_J$, so that $\sum_{i \in I} \alpha_i = \mu(A_I)$ and $\sum_{i \notin [m] \setminus I} \alpha_i = \mu(A_J^c)$, as well as $\sum_{i \in J} \alpha_i = \mu(A_J)$ and $\sum_{i \notin [m] \setminus J} \alpha_i = \mu(A_I^c)$. Consider the following equation

$$\sum_{i \in I \cap J} \alpha_i + \sum_{i \in I \cup J} \alpha_i = \sum_{i \in I} \alpha_i + \sum_{i \in J} \alpha_i = \mu(A_I) + \mu(A_J)$$

$$= \mu(A_I \cap A_J) + \mu(A_I \cup A_J) \geq \mu(A_{I \cap J}) + \mu(A_{I \cup J}),$$

where the final inequality follows from the inclusion $A_{I \cap J} \subset A_I \cap A_J$.

Pairing this with the fact that each of the extreme terms satisfies that

$$\sum_{i \in I \cap J} \alpha_i \leq \mu(A_{I \cap J}) \quad \text{and} \quad \sum_{i \in I \cup J} \alpha_i \leq \mu(A_{I \cup J}),$$

we conclude that both of these inequalities are in fact equalities, which implies that

$$\sum_{i \in I \cap J} \alpha_i = \mu(A_{I \cap J}),$$

so that $\alpha$ belongs to the face $F_{I \cap J}$.

In fact, if $\mu$ is non-degenerate and $I \cap J = \emptyset$, we know much more.

Proposition 5.11 Given a probability measure $\mu \in \mathcal{P}(X)$ concentrated on $\{x \in X : \exists i \in [m], c(x, u_i) < \infty \}$ which is non-degenerate with respect to $(u_i)_{i=1}^m \subset Y$, let $P = P((u_i)_{i=1}^m, \mu)$ be the associated Hall polytope. Then given $I, J \subset [m]$, which are disjoint, the faces $F_I$ and $F_J$ do not intersect.
Proof By Corollary 5.9, we know that for any \( I \), the face \( F_I = \mu(A_I)P_I \times_1 \mu(A'_I)\hat{P}_I \) satisfies that \( \dim(P_I) = |I| - 1 \). Assume \( |I| = k_1, |J| = k_2 \), and, towards a contradiction, that the intersection \( F_I \cap F_J \) is non-empty. Denoting \( \beta_I = \mu(A_I) \) and \( \beta_J = \mu(A_J) \), every point \( \alpha \) in the intersection must satisfy \( \sum_{i\in I} \alpha_i = \beta_I \) and \( \sum_{i\in J} \alpha_i = \beta_J \). Letting \( K = I \cup J \), \( \beta_K = \mu(A_K) \), by the fact that \( I \) and \( J \) are disjoint, we see that \( \sum_{i\in K} \alpha_i = \beta_I + \beta_J \), and since \( \alpha \) is a point in \( P \), \( \beta_I + \beta_J \leq \beta_K \). However, using again that \( I \cap J = \emptyset \), we know that all points \( \alpha \in P \) also satisfy \( \sum_{i\in K} \alpha_i = \beta_I + \beta_J \) and since by Proposition 5.8 \( F_K \) is non-empty (it is full dimensional in its \( I \) coordinates), there exists some \( \alpha \in F_K \) such that the equality \( \sum_{i\in K} \alpha_i = \beta_K \) is satisfied. This implies \( \beta_K = \beta_I + \beta_J \). We conclude that \( F_I \cap F_J = F_K \). Indeed, \( F_K \subset F_I \cap F_J \), since each such \( \alpha \in P \) satisfies \( \sum_{i\in I} \alpha_i \leq \beta_I \) and \( \sum_{i\in J} \alpha_i \leq \beta_J \), and for points in \( F_K \) an equality must be attained in both inequalities. The reverse inclusion \( F_I \cap F_J \subset F_K \) is clear.

Next, we use the decomposition of faces in Proposition 5.4 to say that

\[
F_K = \mu(A_K)P_K \times_K \mu(A'_K)\hat{P}_K = \mu(A_I)P_I \times_1 \mu(A'_I)\hat{P}_I \cap \mu(A_J)P_J \times_1 \mu(A'_J)\hat{P}_J.
\]

In particular, using only the coordinates in \( K \), we have

\[
\mu(A_K)P_K \subset \mu(A_I)P_I \times \mu(A_J)P_J.
\]

(Here the product is with respect to the coordinates \( I \) considered as a subset of \( K \)). This implies that the dimension of \( P_K \) is at most \( k_1 - 1 + k_2 - 1 < k_1 + k_2 - 1 = |K| - 1 \), which contradicts the non-degeneracy assumption on \( P \). We conclude that if \( I \) and \( J \) are disjoint, the intersection \( F_I \cap F_J \) must be empty. \( \square \)

### 5.4 Mapping the Hall polytope to the simplex

In this subsection we make one final preparation, and show that for any Hall polytope \( P \), associated with a non-degenerate measure and some \( m \)-tuple, there exists a special mapping \( R \) from \( \partial P \) to \( \partial \Delta_m \) such that for any \( I \subset [m] \) the set \( F_I \) is mapped to \( \partial_I \Delta_m \), and on \( F_I \) the map only depends on the \( I \) coordinates of a point.

Let us explain the notation. The relative boundary of the simplex (its boundary in the affine space \( \{\alpha \in \mathbb{R}^m : \sum_{i=1}^m \alpha_i = 1\} \)) will be denoted by \( \partial \Delta_m \), and the \( I \)th component of this boundary is the lower dimensional simplex defined by

\[
\partial_I \Delta_m = \{\alpha \in \Delta_m : \sum_{i\in I} \alpha_i = 1\}.
\]

Additionally, for \( I \subset [m] \) we say that ‘a point \( x \in \mathbb{R}^m \) has \( I \)th coordinates \( y \in \mathbb{R}^{|I|} \) if the restriction of \( x \) to its coordinates indexed by \( I \) is equal to \( y \).

**Proposition 5.12** Given a probability measure \( \mu \in \mathcal{P}(X) \) concentrated on \( \{x \in X : \exists i \in [m], c(x,u_i) < \infty \} \) which is non-degenerate with respect to \( (u_i)_{i=1}^m \subset Y \), let \( P = \mathcal{P}((u_i)_{i=1}^m, \mu) \) be the associated Hall polytope. Then there exists a continuous mapping \( R : \partial P \to \partial \Delta_m \) such that for any \( I \subset [m] \) it holds that \( \partial_I P := F_I = \mu(A_I)P_I \times_1 \mu(A'_I)\hat{P}_I \) is mapped to \( \partial_I \Delta_m \) with

\[
R(y,z) = R(y,z') \tag{11}
\]

for \( y \in \mu(A_I)P_I, z, z' \in \mu(A'_I)\hat{P}_I \), that is \( R(x) = R(x') \) if \( x|_I = x'|_I \).
Proof The construction of $R$ is recursive. We define the map first only on faces $F_I$ with $|I| = 1$. We then assume it has been defined on faces $F_I$ with $|I| < k$ and define it on $F_I$ with $|I| = k$. At each step we make sure the map we construct is well defined and continuous on its domain.

We denote the center of mass of the face $\partial_I \Delta_m$ by $q_I$ and the center of mass of the polytope $\mu(A_I)P_I$ (the $I$-th component of $F_I$) by $p_I$. We will ensure, within the proof, that all points in $F_I$ with $I$th coordinates $p_I$ are mapped to $q_I$, and that in general the map on a face $F_I$ depends only on the $I$th coordinates of the point at hand.

The basis for the construction are thus faces $F_I$ of $P$ with $|I| = 1$. We let $R(F_{|I|}) = e_i$ (note that $\partial_{|I|}\Delta_m = e_i$). As these faces are disjoint by Proposition 5.10, and the map $R$ is constant on each face, we conclude that it is continuous.

For the induction step, assume we have defined $R$ on all faces $F_J$ with $|J| < k$. Let $F_I$ be a face of $P$ with $|I| = k$. Since $F_I = \mu(A_I)P_I \times \mu(A_I')\hat{P}_I$ by Proposition 5.4, we have that

$$\partial F_I = \left(\mu(A_I)\text{relint}(P_I) \times \mu(A_I')\hat{P}_I\right) \cup \left(\mu(A_I)\partial P_I \times \mu(A_I')\hat{P}_I\right).$$

Since $R$ is already defined, by assumption, on all faces $F_J$ for $J \subsetneq I$, and since $\mu(A_I)\partial P_I \times \mu(A_I')\hat{P}_I \subset \bigcup_{J \subsetneq I} F_J$ by Lemma 5.6, the map $R$ is already defined on this second component of the boundary. Furthermore, again by assumption, it is defined in such a way that the image of $F_J$ is the simplex $\partial_J \Delta_m$, and that on $F_J$ the map $R$ only depends on the $J$th coordinates of the point at hand. Note that $\bigcup_{J \subsetneq I} \partial_J \Delta_m$ is precisely the boundary of $\partial_I \Delta_m$. So, we essentially are given a continuous mapping $\tilde{R}$ from the boundary of $P_I$ to the boundary of $\partial_I \Delta_m$. We extend it by first imposing that $\tilde{R}(p_I) = q_I$ for the specified points $p_I$ and $q_I$ (note that $p_I \in \mu(A_I)\text{relint}(P_I)$, as $P_I$ is $|I|-1$ dimensional). Next we extend $\tilde{R}$ radially for points with $I$th coordinates in $\mu(A_I)\text{relint}(P_I)$. More formally, given $p \in \mu(A_I)\text{relint}(P_I)$ we let $t^*$ be the maximal $t > 0$ such that $p_I + t(p - p_I) \in \mu(A_I)P_I$. We then define

$$\tilde{R}(p) = q_I + \frac{1}{t^*}\left(\tilde{R}(p_I + t^*(p - p_I)) - q_I\right),$$

where the right hand side is well defined since the point to which we apply $\tilde{R}$ is on the boundary of $\mu(A_I)P_I$. Finally, for a point $(p, z) \in F_I$ we let $R(p, z) = \tilde{R}(p)$. The resulting map is well defined, since by Propositions 5.10 and 5.11 the intersections of the faces $\{F_I\}_{|I|=k}$ are included in faces $F_J$ with $|J| < k$. By construction $R$ is a continuous mapping that sends $F_I$ to $\partial_I \Delta_m$ and, on $F_I$, depends only on the $I$th coordinates. $\blacksquare$

As we described in Subsect. 5.1, the main idea of the proof of Theorem 1.3 is to show surjectivity of a map taking a potential function $\varphi$ (indexed by some variables $(t_i) \in \Delta_m$) to the weight vector $\alpha$ of the measure $\nu$ to which the $c$-subgradient $\partial^c \varphi$ maps $\mu$. We present this formally in the next subsection, where we define and analyse this map.

### 5.5 Mapping the simplex to the Hall polytope

Having fixed some $m$-tuple $(u_i)_{i=1}^m \subset Y$ and a probability measure $\mu \in \mathcal{P}(X)$ concentrated on $\{x \in X : \exists i \in [m] \ c(x, u_i) < \infty\}$ and $c$-regular, we define a map on the interior of $\Delta_m$, and then extend it (using converging subsequences) to a set-valued map on the boundary.

More precisely, for $t = (t_i)_{i=1}^m \in \text{int}(\Delta_m)$, define

$$H_{(u_i)_{i=1}^m}(t) = \alpha$$

(12)
with $\alpha \in \Delta_m$ given by

$$\alpha_i = \mu \left( \left\{ x \in X : \arg\min_{1 \leq j \leq m} (c(x, u_j) - \ln(t_j)) = i \right\} \right).$$

For $t \in \partial \Delta_m$, we let $H^\mu_{\{u_i\}_{i=1}^m}(t)$ be the closure of the function in the usual sense, namely the set of all limit points $\lim H^\mu_{\{u_i\}_{i=1}^m}(t(k))$ as $t(k) \to t$ and $t(k) \in \text{int}(\Delta_m)$. When $\mu$ and $(u_i)_{i=1}^m$ are fixed in advance, we denote $H = H^\mu_{\{u_i\}_{i=1}^m}$. By Lemma 5.2 there is a transport map from $\mu$ to $v = \sum_i \alpha_i \delta_{u_i}$ when $\alpha = H(t)$ for $t \in \text{int}(\Delta_m)$, and moreover the transport map’s graph is included in the $c$-subgradient of the function $\min_{1 \leq j \leq m} (c(x, u_j) - \ln(t_j))$. In particular, the image of $H$ is inside $P = P((u_i)_{i=1}^m, \mu)$, the associated Hall polytope.

Our first claim regards the continuity of $H$.

**Proposition 5.13** Let $X, Y$ be Polish spaces, $c : X \times Y \to (-\infty, \infty]$ measurable, let $(u_i)_{i=1}^m \subset Y$, and let $\mu \in \mathcal{P}(X)$ be $c$-regular and concentrated on $\{ x \in X : \exists i \in [m] \ c(x, u_i) < \infty \}$. Then, the function $H = H^\mu_{\{u_i\}_{i=1}^m} : \Delta_m \to P((u_i)_{i=1}^m, \mu)$ is well defined and continuous on $\text{int}(\Delta_m)$.

**Proof** First note that the function $H$ is well defined as $\mu$ is $c$-regular, and the subsets

$$U_i := \left\{ x \in X : \arg\min_{1 \leq j \leq m} (c(x, u_j) - \ln(t_j)) = i \right\}$$

form a measurable partition of $X$ (the intersections are of measure 0, as well as the set where the minimum is $+\infty$), as in Lemma 5.2.

To show that $H$ is continuous on $\text{int}(\Delta_m)$, let $t \in \text{int}(\Delta_m)$, $\varepsilon > 0$ be fixed. We will show that then there exists $\delta > 0$ such that for all $t' \in \Delta_m$ with $\|t' - t\|_2 < \delta$, we have $\|H(t') - H(t)\|_2 < \varepsilon$. To see this, note that the $i^{th}$ coordinate of the difference is given by $\|\mu(U_i) - \mu(V_i)\|$, where

$$V_i := \left\{ x \in X : \arg\min_{1 \leq j \leq m} (c(x, u_j) - \ln(t'_j)) = i \right\}.$$

Clearly, this difference is bounded (in absolute value) by $\|\mu(U_i \Delta V_i)\|$, where $\Delta$ denotes the symmetric difference of the two sets. To estimate the measure of the symmetric difference, when $t$ and $t'$ are close, we use the following sets, which converge to measure 0 sets as $k \to \infty$.

Define for $i \in [m], k \in \mathbb{N}$

$$U_i^k = \left\{ x \in U_i : \exists j \neq i, c(x, u_j) - \ln(t_j) - c(x, u_i) + \ln(t_i) \leq \frac{1}{k} \right\}.$$

Note that $U_i^{k+1} \subset U_i^k$, and as $\mu$ is finite $\mu(U_i^k) \leq \mu(\lim_k U_i^k)$. Moreover, $\mu(\lim_k U_i^k) = 0$ since by $c$-regularity of $\mu$ the limit set $\bigcup_j \{ x \in X : c(x, u_j) - \ln(t_j) - c(x, u_i) + \ln(t_i) \leq 0 \}$ has zero measure. In particular, for every $i \in [m]$ there exists some $k_i$ such that for all $k > k_i$, we have $\mu(U_i^k) < \varepsilon/m$. Denote $k_0 = \max_i k_i$, and note that for any $k \geq k_0$ we have $\mu(\bigcup_{j=1}^m U_j^k) < \varepsilon$ (and in particular for $k = k_0$).

We next claim that there exists $\delta$ such that if $t'$ is such that $|t' - t_i| < \delta$ for all $i$, we have that $U_i \Delta V_i \subset \bigcup_{j=1}^m U_j^{k_0}$, which completes the proof. Indeed, we will choose $\delta$ such that if $|t - t'| < \delta$ then $t_i e^{-1/2k_0} \leq t'_i \leq t_i e^{1/2k_0}$ for every $i$. 

\[ \square \]
First consider the case \( x \in U_l \setminus V_l \), and note that then there exists \( 1 \leq l \leq m \) such that \( x \in V_l \) (since the sets \((V_i)_{i=1}^m\) are a partition of \( X \)) and hence for all \( 1 \leq j \leq m \) we have

\[
c(x, u_i) - \ln(t'_j) - c(x, u_j) + \ln(t'_j) \leq 0.
\]

By taking \( j = i \) and by the choice of \( \delta \),

\[
c(x, u_i) - \ln(t_i) - c(x, u_i) + \ln(t_i) \leq \frac{1}{k_0},
\]

which yields that \( x \in U_i^{k_0} \). Similarly, in the case where \( x \in V_l \setminus U_l \), there exists \( 1 \leq l \leq m \) such that \( x \in U_l \) and since \( x \in V_l \) we have that for all \( 1 \leq j \leq m \)

\[
c(x, u_i) - \ln(t'_j) - c(x, u_j) + \ln(t'_j) \leq 0.
\]

By taking \( j = l \) and using the assumption on \( \delta \), this yields

\[
c(x, u_i) - \ln(t_i) - c(x, u_i) + \ln(t_i) \leq \frac{1}{k_0},
\]

which implies \( x \in U_i^{k_0} \), and in particular, in both cases, \( x \in \bigcup_{j=1}^m U_j^{k_0} \). Since the parameters were chosen so that the measure of this set is at most \( \varepsilon \), we conclude that \( \mu(U_l \Delta V_l) < \varepsilon \), as long as \( |t' - t| < \delta \), which completes the proof.

\[\square\]

A main feature of the map \( H_{(u_i)_{i=1}^m}^\mu \) is that it “respects the product structure” on the faces of \( \Delta_m \). More precisely, when applied to a point on a face \( \partial_I \Delta_m \), the map is usually set-valued. The set which such a point is mapped to, however, has a specified \( I^{th} \)-coordinate (given by another map of such form, associated with a different measure), and the \( I^c \)-coordinates of points in the image span a full Hall polytope of another associated measure – exactly the one given in the face splitting discussed in Proposition 5.4. This is formally described in the next proposition.

**Proposition 5.14** Under the assumptions of Proposition 5.13, consider some subset \( I \subset [m] \) and let \( t = t_I \times 1 \delta_I \) be a vector with positive \( I^{th} \)-coordinates. Let

\[
\mu_I = \frac{1}{\mu(A_I)} \mu|_{A_I}, \quad \mu_I^c = \frac{1}{\mu(A_I^c)} \mu|_{A_I^c}.
\]

Then,

\[
H_{(u_i)_{i=1}^m}^\mu (t_1, \ldots, t_m) = \mu(A_I) H_{(u_i)_{i \in I}}^\mu_I (t_i)_{i \in I} \times 1 \mu(A_I^c) H_{(u_j)_{j \notin I}}^\mu_I (s_j)_{j \notin I} (\Delta_m - |I|).
\]

In particular, \( H_{(u_i)_{i=1}^m}^\mu \) maps the face \( \partial_I \Delta_m \) to the face \( F_I \) of the Hall polytope \( P((u_i)_{i=1}^m, \mu) \).

**Proof of Proposition 5.14** To prove equality we show two inclusions.

For the direction \( \supseteq \) take a point \((\alpha_1, \ldots, \alpha_m)\) in the right hand side, namely assume that

\[
\alpha \in \mu(A_I) H_{(u_i)_{i \in I}}^\mu_I (t_i)_{i \in I} \times 1 \mu(A_I^c) H_{(u_j)_{j \notin I}}^\mu_I (s_j)_{j \notin I}
\]

where \( \sum_{i \in I} t_i = 1 \) and \( \sum_{j \notin I} s_j = 1 \). For \( \delta \in (0, 1) \) define \( t^\delta_i \in \Delta_m \) in the following way

\[
t^\delta_i = \begin{cases} 
(1 - \delta)t_i & \text{if } i \in I \\
\delta s_i & \text{if } i \notin I
\end{cases}
\]
Clearly, \( t^\delta \to t \) as \( \delta \to 0 \), thus, by the definition of \( H \) it suffices to show that

\[
(\alpha_1, \ldots, \alpha_m) = \lim_{\delta \to 0} H^{\mu}_{(u_i)_{i=1}^m} (t^\delta_{i=1}^m).
\]

We will show that for every \( \varepsilon > 0 \) there exists some \( \delta_0 \) such that for every \( \delta < \delta_0 \) we have

\[
\| (\alpha_1, \ldots, \alpha_m) - H^{\mu}_{(u_i)_{i=1}^m} (t^\delta) \|_\infty \leq \varepsilon.
\] (13)

Denote \((\beta_1, \ldots, \beta_m) = H^{\mu}_{(u_i)_{i=1}^m} (t^\delta)\). Let us reinterpret \( \beta_i \),

\[
\beta_i = \mu \left( \left\{ x \in X : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(t^\delta_k)) = i \right\} \right)
\]

\[
= \mu \left( \left\{ x \in A_I^c \cap (c(x, u_k) - \ln(t^\delta_k)) = i \right\} \right)
\]

\[
+ \mu \left( \left\{ x \in A_I : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(t^\delta_k)) = i \right\} \right).
\]

On \( A_I^c \) the minimum is attained for \( k \notin I \), hence

\[
\beta_i = \mu(A_I^c) \mu_I^c \left( \left\{ x \in X : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(s_k)) = i \right\} \right)
\]

\[
+ \mu \left( \left\{ x \in A_I : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(t^\delta_k)) = i \right\} \right)
\]

\[
= \mu(A_I^c) \mu_I^c \left( \left\{ x \in X : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(s_k)) = i \right\} \right)
\]

\[
+ \mu \left( \left\{ x \in A_I : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(t^\delta_k)) = i \right\} \right).
\]

Observe that the first summand is by definition equal to

\[
\mu(A_I^c) \mu_I^c (\{ x \in X : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(s_k)) = i \}) = \begin{cases} 0 & i \in I \\ \alpha_i & i \notin I \end{cases}.
\] (14)

We first deal with the case \( i \notin I \), in which (14) gives that

\[
\beta_i = \alpha_i + \mu(\{ x \in A_I : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(t^\delta_k)) = i \}).
\]

For \( x \in A_I \), \( \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(t^\delta_k)) = i \) means, in particular, that

\[
\min_{j \in I} (c(x, u_j) - \ln((1 - \delta) t_j)) \geq c(x, u_i) - \ln(\delta s_i)
\]

for all \( j \in I \).

It remains to find \( \delta_0 > 0 \) such that for any \( \delta < \delta_0 \) we have

\[
\mu(\{ x \in A_I : \min_{j \in I} (c(x, u_j) - \ln(t_j)) - c(x, u_i) \geq - \ln(s_i) + \ln \left( \frac{1 - \delta}{\delta} \right) \}) < \varepsilon/m.
\]

Such \( \delta \) exists by the regularity of the measure (as \( \delta \to 0 \) the set considered saturates \( A_I \); note that we have used the fact that \( t_i > 0 \)). For this choice of \( \delta_0 \) we have \( \alpha_i \leq \beta_i \leq \alpha_i + \varepsilon/m \) for all \( i \notin I \).
For the case $i \in I$, 

$$
\beta_i = \mu \left( \left\{ x \in A_I : \arg \min_{1 \leq k \leq m} (c(x, u_k) - \ln(t_k^j)) = i \right\} \right)
$$

$$
\leq \mu \left( \left\{ x \in A_I : \arg \min_{k \in I} (c(x, u_k) - \ln(t_k^j)) = i \right\} \right)
$$

$$
= \mu \left( \left\{ x \in A_I : \arg \min_{k \in I} (c(x, u_k) - \ln(1 - \delta) - \ln(t_k)) = i \right\} \right)
$$

$$
= \mu(A_I) \mu_I \left( \left\{ x \in A_I : \arg \min_{k \in I} (c(x, u_k) - \ln(t_k)) = i \right\} \right) = \alpha_i.
$$

Thus we have $\beta_i \leq \alpha_i$. Since $\sum_{j=1}^m \beta_j = 1 = \sum_{j=1}^m \alpha_j$, and we have already shown that $\alpha_j \leq \beta_j \leq \alpha_j + \varepsilon/m$ when $j \notin I$, we see that

$$
\sum_{i \in I} \alpha_i - \varepsilon \leq 1 - \sum_{j \notin I} (\alpha_j + \varepsilon/m) \leq \sum_{i \in I} \beta_i \leq \sum_{i \in I} \alpha_i
$$

and we conclude $\beta_i \geq \alpha_i - \varepsilon$ for $i \in I$. This completes the proof of (13) and the first inclusion.

We proceed to show the second inclusion $\subseteq$. Let $\alpha \in H_{(u_t)}^\mu(t)$ for $t = t_1 \times_I \{0, \ell\}$. Denote $\alpha = (\alpha_I, \alpha_{I'})$. We claim $\alpha_I \in H_{(u_t)}^\mu(t_I)$ and $\alpha_{I'} \in H_{(u_t)}^\mu(t_{I'})$ for some $s \in \mathbb{R}^{|I'|}$.

By the definition of $H$ on the boundary of the simplex, there exists a sequence

$$
(t_1^{(k)}, \ldots, t_m^{(k)}) = t^{(k)} \rightarrow t = (t_1, \ldots, t_m)
$$

with $t^{(k)} \in \text{int}(\Delta_m)$, and

$$
H_{(u_t)}^\mu_{(u_t)}(t^{(k)}) \rightarrow \alpha.
$$

In particular $\sum_{i \in I} t_i^{(k)} \rightarrow 1$ and $\sum_{j \notin I} t_j^{(k)} \rightarrow 0$. Note that for $x \in A_I$ and large enough $k$, $\arg \min (c(x, u_j) - \ln(t_j^{(k)})) \in I$, and if $x \notin A_I$ the minimum will not be attained on $i \in I$. Therefore

$$
\sum_{j \in I} \alpha_j \leftarrow \sum_{j \in I} (H_{(u_t)}^\mu_{(u_t)}(t_i^{(k)}))_j = \mu(x \in X : \arg \min_{1 \leq i \leq m} (c(x, u_i) - \ln(t_i^{(k)})) \in I) \rightarrow \mu(A_I)
$$

(15)

(limits with respect to $k \rightarrow \infty$). In particular, this implies $\sum_{j \in I} \alpha_j = \mu(A_I)$ and thus $\sum_{j \notin I} \alpha_j = \mu(A_{I'})$.

To show the first assertion, namely that $\alpha_I \in H_{(u_t)}^\mu(t_I)$, define $t'^{(k)} \in R^I$ by $t'^{(k)} = t^{(k)} \frac{t_i^{(k)}}{\sum_{i \in I} t_i^{(k)}}$. It is well defined as $t^{(k)} \in \text{int}(\Delta_m)$ and clearly $t'^{(k)} \rightarrow t_I$. It remains to check that $\mu(A_I) H_{(u_t)}^\mu(t'^{(k)}) \rightarrow \alpha_I$. Indeed,
\[ 
\mu(A_I) H_{(u_i)_{i \leq I}}^{\mu_{i}}(t'(k)) = \left( \mu \left( \left\{ x \in A_I : \arg \min_{j \in I} (c(x, u_j) - \ln(t_j^{(k)})) = i \right\} \right) \right)_{i \leq I} \\
= \left( \mu \left( \left\{ x \in A_I : \arg \min_{j \in I} (c(x, u_j) - \ln(t_j^{(k)})) = i \right\} \right) \right)_{i \leq I}, 
\]

which by assumption converges to \(\alpha_I\) as \(k \to \infty\).

For the second assertion, define \(t''(k) \in R^{I^c}\) by \(t''(k) = \frac{t'(k)}{\sum_{j \in I} t_j^{(k)}}\). The sequence \(t''(k)\) has a converging subsequence in \(\Delta_{I^c}\), so without loss of generality assume \(t''(k) \to t''\). As above

\[ 
\mu(A_I^c) H_{(u_i)_{i \in I^c}}^{\mu_{i}}(t''(k)) = \left( \mu \left( \left\{ x \in A_I^c : \arg \min_{j \in I^c} (c(x, u_j) - \ln(t_j^{(k)})) = i \right\} \right) \right)_{i \in I^c} \\
= \left( \mu \left( \left\{ x \in A_I^c : \arg \min_{j \in I^c} (c(x, u_j) - \ln(t_j^{(k)})) = i \right\} \right) \right)_{i \in I^c}. 
\]

Note that our assumption implies that

\[ 
\left( \mu \left( \left\{ x \in X : \arg \min_{j \in [m]} (c(x, u_j) - \ln(t_j^{(k)})) = i \right\} \right) \right)_{i \leq I} \to \alpha_{I^c}, 
\]

which is a similar statement. However, we can either reason as in (15), or simply use the fact that \(\sum \alpha_i = 1\), to conclude that as \(k \to \infty\), \(\mu(\{x \in A_I : \arg \min_{j \in [m]} (c(x, u_j) - \ln(t_j^{(k)})) = i\}) \to 0\) for any \(i \notin I\). Therefore, the limit of \(\mu(A_I^c) H_{(u_i)_{i \in I^c}}^{\mu_{i}}(t''(k))\) is \(\alpha_{I^c}\) as well. Concluding that \(\alpha_{I^c} \in H_{(u_i)_{i \in I^c}}^{\mu_{i}}(t'')\) as needed. \(\square\)

### 5.6 Transporting a non-degenerate measure to a discrete measure

We proceed to the proof of the following theorem, which is a version of Theorem 1.3, with a stronger non-degeneracy assumption (recall Definition 5.7).

**Theorem 5.15** Let \(X\) be some Polish space and \(Y = \{u_i\}_{i=1}^m\) a discrete measure space. Assume \(c : X \times Y \to (-\infty, \infty]\) is a measurable cost function, \(\mu \in \mathcal{P}(X)\) is \(c\)-regular and \(\nu = \sum_{i=1}^m \alpha_i > 1, u_i \in \mathcal{P}(Y)\). Assume furthermore, that \(\mu\) is non-degenerate with respect to the vectors \((u_i)_{i=1}^m\), that is, for every \(i, j \in \{1, \ldots, m\}\) we have

\[ 
\mu(\{x \in X : c(x, u_i) < \infty\} \cap \{x \in X : c(x, u_j) < \infty\}) > 0. 
\]

If \(\mu\) and \(\nu\) are strongly \(c\)-compatible then there exists an optimal transport plan \(\pi \in \Pi(\mu, \nu)\) whose graph lies in the \(c\)-subgradient \(\partial^c \varphi\) of a \(c\)-class function \(\varphi : X \to [-\infty, \infty]\).

We proceed to the proof of the following theorem, which together with Lemma 5.2 implies Theorem 5.15 as we will show below.

**Theorem 5.16** Let \(X, Y\) be Polish spaces, \(c : X \times Y \to (-\infty, \infty]\) measurable, fix \((u_i)_{i=1}^m \subset Y\) and let \(\mu \in \mathcal{P}(X)\) be \(c\)-regular and concentrated on \(\{x \in X : \exists i \in [m], c(x, u_i) < \infty\}\). Assume, in addition, that \(\mu\) is non-degenerate with respect to \((u_i)_{i=1}^m\). Then, the set \(\text{int}(P((u_i)_{i=1}^m, \mu))\) is included in the image of the mapping \(H_{(u_i)_{i=1}^m}^{\mu} : \text{int}(\Delta_m) \to (P((u_i)_{i=1}^m, \mu))\), that is, for any \(\alpha \in \text{int}(P((u_i)_{i=1}^m, \mu))\) there exists some \(t \in \text{int}(\Delta_m)\) such that \(H_{(u_i)_{i=1}^m}^{\mu}(t) = \alpha\).\(\square\)
Proof Denote $H = H_{(u_i)_{i=1}^m}^\mu$ and $P = P((u_i)_{i=1}^m, \mu)$. By the non-degeneracy assumption, $P$ is full dimensional, and in particular has non-empty interior. If the image of $H$ did not cover the interior of $P$, there would be some $\alpha \in \text{int}(P)$ such that $H(t) \neq \alpha$ for all $t \in \text{int}(\Delta_m)$. We use $\alpha$ to define the radial projection $\nu$ of $P \setminus \{\alpha\}$ to its boundary. It follows that $\nu \circ H$ is well defined and continuous on $\text{int}(\Delta_m)$. Then we use the function $\nu$ as given in Proposition 5.12 to map the boundary of $P$ to the boundary of the simplex. Since $H(\partial_t \Delta_m) \subset F_1$ we see that $F \circ \nu \circ H$ is a mapping (a priori, set-valued) from the simplex to its boundary which maps the $I^{th}$ face to itself.

Note that this composition map is a well defined function, i.e. a point-valued map: It is clearly point-valued on $\text{int}(\Delta_m)$. Let $t \in \partial_t \Delta_m$, and take the minimal $I$ (with respect to inclusion) such that $t \in \partial_I \Delta_m$. Then, the coordinates $t_I$ are all non-zero, and by Proposition 5.14 points in the set $H(t)$ diver only on their $I^c$ coordinates. Again by Proposition 5.14, $H(t) \in F_I \subset \partial P$, so $p(H(t)) = H(t)$. Further, since the map $\nu$ depends only on the $I$ coordinates of $H(t)$ (as $H(t) \in F_I$), we conclude that the set $H(t)$ is mapped to a single point, and thus $F \circ \nu \circ H$ is point-valued.

Next we claim that the composition $F \circ \nu \circ H$ is a continuous function on $\Delta_m$. For points in the interior of $\Delta_m$ this follows from the fact that all three maps are continuous (see Propositions 5.13 and 5.12). We proceed to explain why the composition is continuous on the boundary. Let $t = t_I \times_{I^c} 0_{I^c}$ be some boundary point, with $t_i > 0$ for $i \in I$ (so, $t \in \text{relint}(\partial_I \Delta_m)$). Consider a sequence $t^{(k)} \to t$ with $(F \circ \nu \circ H)(t^{(k)})$ converging to some vector $s$ on the boundary of the simplex. We need to show that $s = (F \circ \nu \circ H)(t)$. By the definition of $H$ on boundary points, and the continuity of $\nu$ and $P$, we may without loss of generality assume $t^{(k)} \in \text{int}(\Delta_m)$. Indeed, for any $t' \in \partial \Delta_m$, $y \in H(t')$ and any $\epsilon > 0$, there is some $t'_k \in \text{int}(\Delta_m)$ with $\|y - H(t'_k)\| < \epsilon$, so given any sequence $t^{(k)} \to t$ with $(F \circ \nu \circ H)(t^{(k)})$ converging to $s$ we can construct a sequence in the interior, converging to $t$, whose image under $F \circ \nu \circ H$ converges to the same $s$. By definition of $H$, all accumulation points of the sequence $H(t^{(k)})$ belong to $H(t)$. By continuity of $F \circ \nu$, we conclude that all accumulation points of $F \circ \nu \circ H(t^{(k)})$ (which we have assumed converge to the point $s$) belong to $(F \circ \nu)(H(t))$. However, as we have already seen, $(F \circ \nu)(H(t))$ is a point, and we get that $s = (F \circ \nu \circ H)(t)$.

However, there does not exist a continuous mapping from the simplex to its boundary which preserves the faces. Indeed, this can be shown, for example, using Brouwer’s fixed point theorem — as such a map could then be composed with a permutation of coordinates, arriving at a continuous mapping from the simplex to itself with no fixed point. Hence, $H$ covers the interior of $P$, and for every $\alpha \in \text{int}(P)$ there is some preimage $t$. Moreover, this $t$ satisfies $t \in \text{int}(\Delta_m)$, otherwise, if $t_i = 0$ for some set of indices $i \in I$, then by Proposition 5.14, $H(t) \in F_{I^c}$, which does not contain $\alpha$ (as $F_{I^c}$ is not in the interior of $P$).

We are now ready to prove the theorem regarding transportation to a discrete measure, which differs from our main theorem only in the constraint of non-degeneracy of $\mu$.

Proof of Theorem 5.15 Under the conditions of Theorem 5.15 we note that by the $c$-compatibility of $\mu$ and $\nu$, we have that $\mu$ must be concentrated on $\{x \in X : \exists i \in [m], c(x, u_i) < \infty\}$. Thus, all conditions in Theorem 5.16 are satisfied and we conclude that $H = H_{(u_i)_{i=1}^m}^\mu$ is onto the interior of the Hall polytope $P = P((u_i)_{i=1}^m, \mu)$. In particular, for $\alpha$ given in the statement of Theorem 5.15 there exists some $t \in \text{int}(\Delta_m)$ such that $H(t) = \alpha$.

Next we use Lemma 5.2, which ensures that the $c$-subgradient of the function

$$\varphi(x) = \min_{1 \leq i \leq m} (c(x, u_i) - \ln(t_i))$$
supports a transport plan from $\mu$ to $\upsilon$. Moreover, $\mu$-almost everywhere this plan is given by the map which takes the set $\{ x : \arg\min_{1 \leq j \leq m} (c(x, u_j) - \ln(t_j)) = i \}$ to $u_i$ for all $i \in [m]$. This completes the proof. \hfill $\square$

### 5.7 Removing non-degeneracy

The main difference between Theorems 1.3 and 5.15 is that in the former instead of assuming that $\mu$ is non-degenerate with respect to $(u_i)_{i=1}^m$, we assume that there exists a $c$-regular measure concentrated on each of the sets

$$\{ x \in X : c(x, u_i) < \infty \} \cap \{ x \in X : c(x, u_j) < \infty \}.$$

To make the reduction we use a straightforward perturbation argument, similar to constructions used for example, by McCann [15], adding in this case $c$-regular measures concentrated on the intersections, and taking limits. More formally, we make use of the following technical lemma.

**Lemma 5.17** Let $P = P((u_i)_{i=1}^m, \mu)$ be the Hall polytope associated with $(u_i)_{i=1}^m \subset Y$ and the measure $\mu \in \mathcal{P}(X)$ which is $c$-regular and concentrated on $\{ x \in X : \exists i \in [m], c(x, u_i) < \infty \}$, and assume that $P$ is full dimensional. Further, assume that for any $1 \leq i < j \leq m$ there exists a $c$-regular probability measure concentrated on the intersection

$$\{ x \in X : c(x, u_i) < \infty \} \cap \{ x \in X : c(x, u_j) < \infty \}$$

and let $\eta_{i,j}$ be a multiple of this measure so that $\sum_{i < j \in [m]} \eta_{i,j}(X) = 1$. For any $k \in \mathbb{N}$ let

$$\mu_k = \frac{1}{k} \sum_{i < j \in [m]} \eta_{i,j} + \left( 1 - \frac{1}{k} \right) \mu,$$

and $P_k = P((u_i)_{i=1}^m, \mu_k)$ the associated Hall polytope. Then,

(i) $H^\mu_{(u_i)_{i=1}^m} \rightarrow_k \infty H^\mu_{(u_i)_{i=1}^m}$ uniformly on $\text{int}(\Delta_m)$, and

(ii) $P_k \rightarrow P$ as $k \rightarrow \infty$ in the Hausdorff metric.

**Proof** Note first that each $\mu_k$ is non-degenerate, so by Proposition 5.8 $P_k$ is full dimensional. Furthermore, we may apply Theorem 5.16, since $\mu_k$ is $c$-regular as a convex combination of $c$-regular measures. It implies that the set $\text{int}(P_k)$ is included in the image of the mapping $H_k := H^\mu_{(u_i)_{i=1}^m}$. Denote $H := H^\mu_{(u_i)_{i=1}^m}$.

For (i), let $t = (t_1, \ldots, t_m) \in \text{int}(\Delta_m)$. It will be convenient to recall the notation from Lemma 5.2 where we denote $U_i = \{ x \in X : \arg\min_{1 \leq j \leq m} (c(x, u_j) - \ln(t_j)) = i \}$. By definition

$$|(H_k(t) - H(t))| = |\mu_k(U_i) - \mu(U_i)| = \left| \frac{1}{k} \sum_{i < j} \eta_{i,j}(U_i) - \frac{1}{k} \mu(U_i) \right| \leq \frac{2}{k},$$

which clearly implies uniform convergence.

For (ii), let $\varepsilon > 0$, we will show that there exists $k_0 = k(\varepsilon)$ such that for all for all $k \geq k_0$, $P_k \subset P + \varepsilon B_2^n$ and $P \subset P_k + \varepsilon B_2^n$. For the first inclusion, let $\alpha \in \text{int}(P_k)$. Applying Theorem 5.16 again, we get a point $t \in \text{int}(\Delta_m)$ for which $H_k(t) = \alpha$. By (i) $\|\alpha - H(t)\|_2 \leq \frac{2}{\sqrt{m}}$ and as $H(t) \in P$ we get $\text{int}(P_k) \subset P + \frac{2}{\sqrt{m}} B_2^n$ and therefore $P_k$ itself is also included in the same extension of $P$. For the second inclusion, let $\alpha \in P$ and define $\alpha^{(k)}_i := \left( 1 - \frac{1}{k} \right) \alpha_i + \frac{1}{k} \sum_{j \neq i} \eta_{i,j}(X)$. By the definition of $\eta_{i,j}$ it turns out that $\alpha^{(k)} \in P_k$.\hfill $\Box$
Indeed, \( \sum \alpha_i^{(k)} = 1 \), and as the support of \( \eta_{i,j} \) is a subset of \( A_{[i]} \cap A_{[j]} \subset A_I \) for all \( i \in I \) we have
\[
\sum_{i \in I} \alpha_i^{(k)} = \left( 1 - \frac{1}{k} \right) \sum_{i \in I} \alpha_i + \frac{1}{k} \sum_{i \in I \setminus \{j : i < j\}} \eta_{i,j}(X) \\
\leq \left( 1 - \frac{1}{k} \right) \mu(A_I) + \frac{1}{k} \sum_{i < j} \eta_{i,j}(A_I) = \mu_k(A_I).
\]

We compute
\[
\|\alpha - \alpha^{(k)}\|_2 = \left( \sum_{i=1}^m \frac{1}{k^2} (\alpha_i - \sum_{\{j : i < j\}} \eta_{i,j}(X))^2 \right)^{1/2} \leq \frac{2 \sqrt{m}}{k}.
\]
Taking \( k_0 = \frac{2 \sqrt{m}}{\epsilon} \) completes the proof. \( \square \)

We are now set up to prove the existence of a transport map between strongly \( c \)-compatible measures, under the conditions specified in Theorem 1.3.

**Proof of Theorem 1.3** Let \( \mu \in \mathcal{P}(X) \) be a \( c \)-regular measure and \( \nu = \sum_{i=1}^m \alpha_i \Delta_{u_i}, \alpha \in \text{int}(\Delta_m) \). Denote by \( P = P([u_i])_{i=1}^m, \mu \) the associated Hall polytope.

The condition of strong \( c \)-compatibility means precisely that for \( I \neq \emptyset, [m] \) we have \( \sum \alpha_i < \mu(A_I) \), or, in other words, that \( \alpha = (\alpha_i)_{i=1}^m \in \text{int}(P) \). In particular, \( P \) is non-empty and in fact full dimensional.

The conditions of Lemma 5.17 are satisfied so we may use it to define \( P_k \) and \( H_k \), where we have used the same notation as in the proof of the lemma. Since \( \alpha \in \text{int}(P) \) and \( P_k \to P \) in the Hausdorff metric it is easy to check that \( \alpha \in \text{int}(P_k) \) for large enough \( k \). Arguing as in the proof of the lemma, the conditions of Theorem 5.16 are satisfied for the measure \( \mu_k \), so we may find \( t^{(k)} \in \text{int}(\Delta_m) \) such that \( H_k(t^{(k)}) = \alpha \). By the compactness of \( \Delta_m \), there exists a converging subsequence of \( t^{(k)} \) to some \( t \in \Delta_m \), and we assume without loss of generality that \( t^{(k)} \to t \). We claim that \( \alpha \in H(t) \). Indeed,
\[
\|H(t_k) - \alpha\| = \|H(t_k) - H_k(t_k)\| \leq \frac{2 \sqrt{m}}{k} \xrightarrow{k \to \infty} 0.
\]
By definition of \( H \) this means that \( \alpha \in H(t) \). Furthermore, if \( t \in \text{int}(\Delta_m) \) then by continuity of \( H \) on the interior (Proposition 5.13) implies \( \alpha = H(t) \). Finally, we claim that indeed \( t \in \text{int}(\Delta_m) \) since \( \alpha \in \text{int}(P) \) and by Proposition 5.14 it cannot be in the image of points of the boundary of the simplex.

Having established the existence of \( t \in \text{int}(\Delta_m) \) which is mapped to \( \alpha \), we proceed as in the proof of Theorem 5.15 to construct the potential \( \varphi \), which completes the proof. \( \square \)

### 6 Specializing to the polar cost

In this section we will give proofs of Theorems 1.2 and 1.4. It turns out that specializing to the space \( \mathbb{R}^n \) and adding some assumptions on the cost allows to give a cleaner formulation of theorems and stronger results. We present them for the polar cost and the reader may infer the additional conditions which will provide similar theorems for other reasonable costs.

Throughout the paper, excluding Sect. 5, we were careful to discuss transport plans, and not just maps. Indeed, even in the simplest cases of discrete measures, there is no reason for a
transport map to exist, as it may require “atom splitting”, a dangerous endeavor. Nevertheless, in the classical case, for example, when a transport plan from some absolutely continuous measure \( \mu \) to a measure \( \nu \) is concentrated on the usual subgradient of a convex function, \( \varphi \in \text{Cvx}(\mathbb{R}^n) \), it is easy to see that in fact one obtains a map, not just a plan. Indeed, a convex function has a unique subgradient almost everywhere.

For a general cost \( c \) this is no longer the case, but for our main motivating example, the polar cost \( p(x, y) = -\ln((x, y) - 1) \), a similar argument works. Recall that for this cost the \( p \)-class is given by \( -\ln(\varphi) \) where \( \varphi \in \text{Cvx}_0(\mathbb{R}^n) \) is a geometric convex function, that is, a lower semi-continuous non-negative convex function with \( \varphi(0) = 0 \). The \( p \)-subgradient of the function \( -\ln(\varphi) \) coincides with the polar subgradient \( \partial^0 \), introduced in [4], of the function \( \varphi \in \text{Cvx}_0(\mathbb{R}^n) \), and we have that

\[
\partial^p(-\ln(\varphi)) = \partial^0\varphi = \{(x, y) : \varphi(x)A\varphi(y) = (x, y) - 1 > 0\},
\]

where \( A\varphi(y) = \sup_{(x, (x, y) > 1)} \frac{(x, y) - 1}{\varphi(x)} \) is the polarity transform defined in [3]. More details are provided in Appendix A together with the proof of the following lemma, which is a reformulation of a statement that appeared in [4].

**Lemma A.4** Let \( \varphi \in \text{Cvx}_0(\mathbb{R}^n) \) and let \( x \in \text{dom}(\varphi) \backslash Z_\varphi \). Then there is a bijection between the set \( \{z \in \varphi(x) : (x, z) \neq \varphi(x)\} \) and the set \( \partial^0\varphi(x) \) given by \( z \mapsto y = \frac{z}{(x, z) - \varphi(x)} \).

When \( \varphi(x) = 0 \) or \( \varphi(x) = \infty \), then by definition, \( \partial^0\varphi(x) = \emptyset \). When \( \varphi(x) \in (0, \infty) \), the lemma implies that at a differentiability point of \( \varphi \), the set \( \partial^0\varphi(x) \) is either a singleton or is empty, which may happen only if the function \( \varphi \) is linear on \([0, x]\). We are now in the position to prove Theorem 1.2.

**Proof of Theorem 1.2** We shall use Theorem 1.1 for the polar cost \( p \). The assumption of finite second moment implies that the polar cost is essentially bounded from below with respect to \( \mu \) and \( \nu \). Indeed,

\[
-\ln((x, y) - 1) \geq 2 - \|x\|\|y\| \geq 2 - \frac{\|x\|^2}{2} - \frac{\|y\|^2}{2},
\]

so letting \( a(x) = b(x) = 1 - \frac{\|x\|^2}{2} \) establishes the essential boundedness. As the strong \( p \)-compatibility is assumed, together with the existence of a finite cost plan, the theorem implies existence of a potential, namely a function \( \varphi \in \text{Cvx}_0(\mathbb{R}^n) \) such that there is an optimal plan \( \pi \) concentrated on the graph of \( \partial^p(-\log \varphi) = \partial^0\varphi \).

We claim that \( \mu \)-almost everywhere, the set \( \partial^0\varphi(x) \) is a singleton, implying that \( \partial^0\varphi \) is indeed a transport map. Indeed, since \( \pi \) is concentrated on \( \partial^0\varphi \), the measure \( \mu \) is concentrated on the projection of \( \partial^0\varphi \), so in particular on the set of \( x \in \mathbb{R}^n \) with \( \varphi(x) \neq 0, \infty \). We may also restrict to points in the interior of the domain of \( \varphi \), as \( \varphi \) is convex and points on the boundary of its domain have \( \mu \)-measure zero (using absolute continuity of \( \mu \)). Since \( \varphi \) is convex, \( \mu \)-almost every point \( x \) in the interior of its domain is a differentiability point. Using Lemma A.4, \( \mu \)-almost everywhere \( \partial^0\varphi(x) \) is either a singleton or the empty set (in which case \( x \) does not belong to the projection of \( \partial^0\varphi \)). We conclude that indeed \( \partial^0\varphi(x) \) must be a singleton \( \mu \)-almost everywhere, as required.

\( \square \)

Theorem 1.4 follows from Theorem 1.3 and a similar analysis of the polar subgradient.

**Proof of Theorem 1.4** In the case of the polar cost \( p(x, y) = -\ln((x, y) - 1) \), whenever \( u_i \neq -tu_j \) for \( t > 0 \), the intersection

\[
\{x \in \mathbb{R}^n : \langle x, u_i \rangle > 1\} \cap \{x \in \mathbb{R}^n : \langle x, u_j \rangle > 1\}
\]

is a singleton. Therefore, for \( t > 0 \), the intersection

\[
\{x \in \mathbb{R}^n : \langle x, u_i \rangle > 1\} \cap \{x \in \mathbb{R}^n : \langle x, u_j \rangle > 1\}
\]

is a singleton.
contains some open ball $B_{ij}$, so we may define the measure $\eta_{ij} = \mathbb{1}_{B_{ij}}$, i.e. $\eta_{ij}$ is the uniform measure on $B_{ij}$. This measure is $p$-regular, since the polar cost is continuous, and in particular, does not assign any measure to lower dimensional sets of the form
\[
\{z \in X : p(z, y_1) - p(z, y_0) = t\}.
\]
Hence, applying Theorem 1.3 gives a potential $\varphi$. By construction, as in the proof of Theorem 1.2 it follows that $\partial^o \varphi(x)$ is a singleton $\mu$-almost everywhere, completing the proof. \hfill \Box

### 7 Decomposable pairs

We discussed in Sect. 3 that when considering the transport problem of a measure $\mu \in \mathcal{P}(X)$ to $\nu \in \mathcal{P}(Y)$, with respect to a cost function $c : X \times Y \to (-\infty, \infty]$, where $\mu$ and $\nu$ are $c$-compatible but not strongly $c$-compatible, the transport problem splits into two transport problems of disjointly supported measures. Let us make this observation more formal.

**Proposition 7.1** Let $c : X \times Y \to (-\infty, \infty]$, and assume $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ are $c$-compatible measures which are not strongly $c$-compatible. There exists a $c$-class set $A \subset X$, and $B \subset Y$ such that $Y \setminus B$ is $c$-class, with $\mu(A) = \nu(B) \in (0, 1)$, such that, letting $\mu|_A$ and $\nu|_B$ denote the restricted measures, normalized, the pair $\mu|_A$ and $\nu|_B$ is $c$-compatible, as is the pair $\mu|_{X \setminus A}$ and $\nu|_{Y \setminus B}$. Moreover, any $\pi \in \Pi(\mu, \nu)$ which is concentrated in the set $S = \{ (x, y) : c(x, y) < \infty \}$, can be written as $\pi = \mu(A)\pi_1 + (1 - \mu(A))\pi_2$, where $\pi_1 \in \Pi(\mu|_A, \nu|_B)$ and $\pi_2 \in \Pi(\mu|_{X \setminus A}, \nu|_{Y \setminus B})$, and $C(\mu, \nu) = \mu(A)C(\mu|_A, \nu|_B) + (1 - \mu(A))C(\mu|_{X \setminus A}, \nu|_{Y \setminus B})$.

**Proof** Indeed, by Lemma 3.12, the fact that the measures are not strongly $c$-compatible implies that there exists some set $A \subset X$, which is a $c$-class set (this means there is some $D \subset Y$, which can also be assumed to be a $c$-class set, such that $A = \{ x \in X : \forall y \in D, c(x, y) = \infty \}$), and such that $\mu(A) \in (0, 1)$ and
\[
\mu(A) + \nu(\{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}) = 1.
\]
Rearranging, this means that
\[
\mu(A) = \nu(\{ y \in Y : \exists x \in A, \ c(x, y) < \infty \}) \quad \text{and} \quad \mu(X \setminus A) = \nu(\{ y \in Y : \forall x \in A, \ c(x, y) = \infty \}).
\]
Let $B = \{ y \in Y : \exists x \in A, \ c(x, y) < \infty \} = Y \setminus D$. To see that $\mu|_A$ and $\nu|_B$ are $c$-compatible, letting $\mu(A) = a$, say, and fixing some set $A' \subset A$, we see that
\[
\mu(A') + \nu(\{ y \in B : \forall x \in A', \ c(x, y) = \infty \}) = \mu(A') + \nu(\{ y \in Y : \forall x \in A', \ c(x, y) = \infty \}) - \nu(Y \setminus B) \leq 1 - (1 - a) = a,
\]
as required. Similarly for the complementary measures. If a transport plan $\pi \in \Pi(\mu, \nu)$ is concentrated on $S$, then $\pi$ cannot have non-zero measure in $A \times (Y \setminus B)$ or in $(X \setminus A) \times B$.

Indeed, as
\[
\mu(A) = \pi((A \times Y) \cap S) = \pi(A \times B) \leq \nu(B) = \mu(A),
\]
we have equalities all along, and $\pi(A \times (Y \setminus B)) = 0$. Similarly, as $Y \setminus B = D$ is a $c$-class set, $D = \{ y : \forall x \in A, \ c(x, y) = \infty \}$, so $D$ must be mapped to $X \setminus A$, and as these sets have the same measure,
\[
\nu(D) = \pi((X \times D) \cap S) = \pi((X \setminus A) \times D) \leq \mu(X \setminus A) = \nu(D),
\]
and by the same reasoning, \( \pi((X \setminus A) \times B) = 0 \). (Fig. 1 is a good illustration of this event.)

In other words, such a transport plan can be split into its components, \( \pi_1 = \pi|_{A \times B} \) and \( \pi_2 = \pi|_{(X \setminus A) \times (Y \setminus B)} \) (where, as in the statement of this Proposition, restriction stands for restricting a measure and renormalizing it to a probability measure). This completes the proof.

Of course, the fact that the problem splits into two sub-problems does not necessarily imply we may solve it in a satisfactory way. Indeed, it may be the case that each sub-problem has an associated potential function, but these two functions cannot be “glued” so as to form a potential for the original problem. This is the case for example for the polar cost in the following example.

**Example 7.2** Consider the set

\[ A = \{ (x, y) : x \in \left( \frac{1}{2}, 1 \right), \ y = 3 - 2x \} \cup \{ (x, y) : x \in (1, 2), \ y = \frac{3}{2} - \frac{1}{2}x \} \subset \mathbb{R}^+ \times \mathbb{R}^+. \]

The set is a \( p \)-cyclically monotone (with respect to the polar cost \( p(x, y) = -\ln(1, y) \)) since for every point \( (x, y) \in A \) we have \( (x, y) > 1 \) and it is a graph of non-increasing function on its domain, which characterize \( p \)-cyclically monotone sets on the ray \( \mathbb{R}^+ \), see [5]. However, the set is not \( p \)-path-bounded, and thus admits no potential.

Next, consider the measure \( \mu = \nu \) on \([1/2, 2] \) with density \( 1 \) on \([1/2, 1] \) and density \( 1/2 \) on \([1, 2] \). This is a probability measure. In fact, \( \mu \) and \( \nu \) are \( p \)-compatible as the normalized uniform Lebesgue measure on the set \( A \) constitutes a plan \( \pi \in \Pi(\mu, \nu) \). However, they are not strongly \( p \)-compatible since the set \( A = [1/2, 1] \) must be mapped to \( B = [1/2, 2] \) and vice versa.

In this case we see the splitting very clearly, and indeed \( A \) is written as the union of two sets, each of which admits a potential (so, in particular, each is \( c \)-path-bounded, and is an optimal plan between the corresponding restricted measures). However, there is no potential for the full set \( A \), as it is not \( c \)-path-bounded, and in particular no “gluing” of the two potentials is possible.

**Acknowledgements** The authors were partially supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 770127). The first named author was partially supported by ISF Grant No. 784/20. The second named author is grateful to the Azrieli foundation for the award of an Azrieli fellowship. We thank the anonymous referee whose comments helped improve and clarify this manuscript.

### Appendix A: \( c \)-subgradients and polar subgradients

Since \( c \)-subgradients play such an important role in this theory, we gather here some relevant information regarding them but which we did not include in the main text so as not to disturb its flow.

Let us recall that given a function \( \varphi \) in the \( c \)-class, its \( c \)-subgradient is defined by

\[ \partial^c \varphi = \left\{ (x, y) : \varphi(x) + \varphi^c(y) = c(x, y) \text{ and } c(x, y) < \infty \right\}. \]

Denoting by \( \partial^c \varphi(x) \) the set of points \( y \in Y \) for which \( (x, y) \in \partial^c \varphi \), we have by the definition that

\[ x \in \partial^c \varphi^c(y) \iff y \in \partial^c \varphi(x). \]

Notice that \( y \in \partial^c \varphi(x) \) if and only if the function \( c(\cdot, y) - \varphi^c(y) \) is above \( \varphi \) and coincides with it at \( x \). This provides the first simple but useful way to think about \( c \)-subgradients,
summarized in Lemma A.1. Given a function $\varphi$ in the $c$-class, it is the image, under the $c$-transform, of another $c$-class function $\psi = \varphi^c$ and therefore, it can be written as an infimum over basic functions as follows:

$$\varphi(x) = \inf_y \left( c(x, y) - \varphi^c(y) \right).$$

All the functions on the right hand side lie above $\varphi$. If any one of the basic functions (indexed by $y$) on the right hand side is equal to $\varphi$ at the point $x$, then the pair $(x, y)$ belongs to $\partial^c \varphi$, and $y \in \partial^c \varphi(x)$.

**Lemma A.1** Let $\varphi$ be a $c$-class function, and $x \in X$ and assume that $\varphi(x) < \infty$. Then $y_0 \in \partial^c \varphi(x)$ if and only if $c(x, y_0) < \infty$ and the function $\ell(z) = c(z, y_0) - c(x, y_0) + \varphi(x)$ satisfies

$$\ell(z) \geq \varphi(z) \text{ for all } z \in X.$$

**Proof** By the definition we have that $y_0 \in \partial^c \varphi(x)$ if and only if

$$\varphi(x) + \varphi^c(y_0) = c(x, y_0) < \infty.$$

Using the definition of the $c$-transform we see that

$$\varphi(x) = c(x, y_0) - \varphi^c(y_0) = \sup_z (c(x, y_0) - c(z, y_0) + \varphi(z)),$$

which holds if and only if for all $z$ we have $c(z, y_0) - c(x, y_0) + \varphi(x) \geq \varphi(z)$.

It is useful to understand the structure of the $c$-subgradient of the basic functions. In parallel to the classical case, where the linear functions have constant subgradient, we show that under mild assumptions the same is true for $c$-subgradients of basic functions. This was, of course, our motivation for using the specific candidates for the potential functions in Sect. 5.

**Lemma A.2** Let $X, Y$ be Polish spaces and let $c : X \times Y \to (-\infty, \infty]$ be a measurable cost function. Consider a basic function $\varphi(x) = c(x, y_0) + t$ for some $y_0 \in Y$. If $c(x, y_0) < \infty$, then $y_0 \in \partial^c \varphi(x)$. If, in addition, for any $y_1 \neq y_0$ we have that $\inf_z (c(z, y_1) - c(z, y_0))$ is not attained at $x$ (for example, if the infimum is $-\infty$, or bounded but not attained at all) then \{ $y_0$ \} = $\partial^c \varphi(x)$.

**Proof** Indeed, let $\varphi$ be as in the statement. From the definition it follows that $y \in \partial^c \varphi(x)$ if and only if $c(x, y) < \infty$ and

$$c(x, y) - \varphi(x) = \varphi^c(y) = \inf_z (c(z, y) - \varphi(z)),$$

which can be reformulated as

$$c(x, y) - \varphi(x) \leq c(z, y_1) - \varphi(z) \text{ for all } z \in X.$$

Plugging in the definition of $\varphi$ we get

$$c(x, y) - c(x, y_0) \leq c(z, y) - c(z, y_0) \text{ for all } z \in X.$$

We see that $y = y_0$ always satisfies the equality, so that $y_0 \in \partial^c \varphi(x)$. Clearly for $y_1 \neq y_0$, such an inequality means precisely that the infimum is attained at $x$. 

An important and motivating first example is the one coming from the classical cost function $c(x, y) = -\langle x, y \rangle$. 

\copyright Springer
Example A.3 For the cost function $c(x, y) = -\langle x, y \rangle$, whose transport plans and maps coincide with those associated to the quadratic cost, the $c$-subgradient coincides, up to a minus sign, with the well known subgradient. More formally, a function $\varphi$ is in the $c$-class if and only if $-\varphi \in \text{Cvx}_0(\mathbb{R}^n)$, namely is convex and lower semi-continuous. Denoting $\psi = -\varphi$ and using the definition of the $c$-transform we see that $(x, y) \in \partial^c \varphi$ if and only if for all $z \in X$ we have

$$\psi(z) - \psi(x) = \varphi(x) - \varphi(z) \geq c(x, y) - c(z, y).$$

Plugging in the cost we indeed get that $y \in \partial^c \varphi(x)$ if for all $z$ it holds that $\psi(x) + \langle z - x, y \rangle \leq \psi(z)$, namely $y \in \partial \psi(x)$.

The second motivating example, which is our main point of interest, is that of the polar cost $p: \mathbb{R}^n \times \mathbb{R}^n \to (-\infty, \infty]$, which we once again recall

$$p(x, y) = -\ln(\langle x, y \rangle - 1) = \begin{cases} -\ln(\langle x, y \rangle - 1), & \text{if } \langle x, y \rangle > 1 \\ +\infty, & \text{otherwise}. \end{cases}$$

It was shown in [7] that for the polar cost the $p$-class consists of all functions of the form $-\ln(\varphi)$, where $\varphi$ is a geometric convex function, that is, a lower semi-continuous non-negative convex function with $\varphi(0) = 0$. We denote this class by $\text{Cvx}_0(\mathbb{R}^n)$. The associated cost transform is linked with the $A$-transform defined in [3] and given by

$$A\varphi(y) = \sup_{\langle x, (x,y) > 1 \rangle} \frac{\langle x, y \rangle - 1}{\varphi(x)}.$$  

(16)

More precisely, one may easily verify that $-\ln(A\varphi) = (-\ln(\varphi))^p$. Further, the $p$-subgradient of the function $-\ln(\varphi)$ can be rewritten as the polar subgradient $\partial^p$, introduced in [4], of the function $\varphi \in \text{Cvx}_0(\mathbb{R}^n)$. Indeed, we have that

$$\partial^p(-\ln(\varphi)) = \partial^p\varphi = \{(x, y) : \varphi(x)A\varphi(y) = \langle x, y \rangle - 1 > 0\}.$$  

(17)

This convenient form is a reason for us to sometimes consider a “multiplicative” setting, where the basic functions are of the form

$$\varphi_{u,t}(x) = t(\langle x, u \rangle - 1)_+ = t \max(\langle x, u \rangle - 1, 0).$$

The next lemma, which is a version of [4, Lemma 3.3], describes the connection between the polar subgradient and the classical subgradient. We will also use the following notation for the zero set $Z_{\varphi} = \{x : \varphi(x) = 0\}$ and $\text{dom}(\varphi) = \{x : \varphi(x) < \infty\}$ for the domain where $\varphi$ is finite.

Lemma A.4 Let $\varphi \in \text{Cvx}_0(\mathbb{R}^n)$ and let $x \in \text{dom}(\varphi)\setminus Z_{\varphi}$. Then there is a bijection between the set $\{z \in \partial \varphi(x) : \langle x, z \rangle \neq \varphi(x)\}$ and the set $\partial^p\varphi(x)$ given by $z \mapsto y = \frac{z}{\langle x, z \rangle - \varphi(x)}$. (Note that these sets might be empty.)

Proof We first show that the mapping has the desired range. Let $z \in \partial \varphi(x)$ with $\langle x, z \rangle \neq \varphi(x)$, which means that for every $w$ we have $\langle w, z \rangle - \varphi(w) \leq \langle x, z \rangle - \varphi(x)$. In particular, $\langle x, z \rangle - \varphi(x) > 0$. Hence, letting $y = \frac{z}{\langle x, z \rangle - \varphi(x)}$ we have that $\langle x, y \rangle > 1$.

To show that $y \in \partial^p\varphi(x)$, it remains to show that $\varphi(x)A\varphi(y) = \langle x, y \rangle - 1$. According to the definition of $A$, this holds if for every $w$ with $\langle w, y \rangle > 1$ and $\varphi(w) > 0$, we have

$$\frac{\langle w, y \rangle - 1}{\varphi(w)} \leq \frac{\langle x, y \rangle - 1}{\varphi(x)}.$$
Plugging in the formula for $y$ and rearranging gives
\[
\frac{(w, \frac{z}{(x,z) - \varphi(x)}) - 1}{\varphi(w)} \leq \frac{(x, \frac{z}{(x,z) - \varphi(x)}) - 1}{\varphi(x)} = \frac{1}{(x,z) - \varphi(x)}.
\]

Using that $(x,z) - \varphi(x) > 0$, the above inequality is equivalent to our initial assumption $(w,z) - \varphi(w) \leq (x,z) - \varphi(x)$.

Having established that the mapping is indeed into $\partial^0 \varphi(x)$, we proceed to show that it is surjective. Given $y \in \partial^0 \varphi(x)$ it follows from the definition that $(x,y) > 1$. Consider
\[
z = \frac{y \varphi(x)}{(y,x) - 1},
\]

(18)

which is well defined. Taking scalar product with $x$ we see that $(z,x) = \frac{(y,x) \varphi(x)}{(y,x) - 1} > 0$, which can be rearranged to $(y,x)((z,x) - \varphi(x)) = (z,x) > 0$ so, in particular, $(z,x) \neq \varphi(x)$. Solving for $(x,y)$ and plugging back into the Eq. (18) we get that $y = \frac{z}{(x,z) - \varphi(x)}$. To prove surjectivity, it remain to show that $z \in \partial^0 \varphi(x)$. To this end, as before, we use that if $y \in \partial^0 \varphi(x)$ then for any $w$ with $(w,y) > 1$ and $\varphi(w) > 0$ we have \(\frac{(w,y) - 1}{\varphi(w)} \leq \frac{(x,y) - 1}{\varphi(x)}\). Plugging in the formula for $y$ and rearranging, we get that
\[
\varphi(x) + (w - x, z) \leq \varphi(w)
\]

holds for any $w$ such that $(w,z) > (x,z) - \varphi(x)$ and $\varphi(w) > 0$. In the case when $w$ is such that $(w,z) \leq (x,z) - \varphi(x)$, this actually means that $\varphi(x) + (w-x,z) \leq 0$ and since the geometric convex functions are non-negative the desired inequality trivially follows. Finally, we consider the case when $w \in Z_\varphi$, i.e. when $\varphi(w) = 0$. Then, plugging in $z$ as defined in (18), we have that the inequality defining the subgradient of $\varphi$ at $x$ becomes simply
\[
(w,y) \leq 1.
\]

That is, we need to show that $\partial^0 \varphi(x)$ is contained in the polar set of $Z_\varphi$. Indeed, $y \in \partial^0 \varphi(x)$ implies, in particular, that $y \in \text{dom } (A\varphi)$ (since the value of $A\varphi(y) = \frac{(x,y) - 1}{\varphi(x)} < \infty$) and it follows from the definition of $A$ that $\text{dom } (A\varphi) \subset Z^0_\varphi$, which completes the proof of surjectivity.

To show injectivity, assume that for $z,w \in \partial\varphi(x)$ we have that $y = \frac{z}{(x,z) - \varphi(x)} = \frac{w}{(x,w) - \varphi(x)}$. This means that these vectors are parallel. Taking scalar product of each of them with $x$ and simplifying we get that $(x,z) = (x,w)$, and plugging this back into the equality we get $z = w$, as claimed. \hfill \Box

We end this appendix with one explicit example of a function and its $p$-subgradient. More examples and applications can be found in [7, 24] and in the forthcoming [2].

**Example A.5** Let $\varphi(x) = \|x\|^2/2$, in which case $A\varphi(y) = \|y\|^2/2$ and the supremum in the definition of $A\varphi$ is satisfied for $x = 2y/\|y\|^2$. Hence, $\partial^0 \varphi(x) = \frac{2x}{\|x\|^2}$. Note that the mapping $x \mapsto \partial^0 \varphi(x)$ in this case is a (rescaled) spherical inversion.

**References**

1. Ambrosio, L., Pratelli, A.: Existence and stability results in the $L^1$ theory of optimal transportation. Optimal transportation and applications, pp. 123–160. Springer, Berlin, Heidelberg (2003)
2. Artstein-Avidan, S., Barel, H., Rubinstein, Y., Sadoovsky, S., Wyczesany, K.: Transportation induced by the polarity transform, in preparation
3. Artstein-Avidan, S., Milman, V.: Hidden structures in the class of convex functions and a new duality transform. J. Eur. Math. Soc. 13(4), 975–1004 (2011)
4. Artstein-Avidan, S., Rubinstein, Y.A.: Differential analysis of polarity: Polar Hamilton-Jacobi, conservation laws, and Monge Ampère equations. J. d’Anal. Math. 132(1), 133–156 (2017)
5. Artstein-Avidan, S., Sadovsky, S., Wyczesany, K.: A Rockafellar-type theorem for non-traditional costs. Adv. Math. 395, 108157 (2022)
6. Ball, K.: An elementary introduction to monotone transportation. In: Geometric aspects of functional analysis, Lecture Notes in Mathematics, pp. 41–52. Springer, Berlin, Heidelberg (2004)
7. Barel, H.: Optimal transportation problem for polar cost, Master’s thesis, Tel Aviv University (2019)
8. Beiglböck, M., Goldstern, M., Maresh, G., Schachermayer, W.: Optimal and better transport plans. J. Funct. Anal. 256(6), 1907–1927 (2009)
9. Bianchini, S., Caravenna, L.: On optimality of c-cyclically monotone transference plans. Comptes Rendus Math. 348(11–12), 613–618 (2010)
10. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. 44(4), 375–417 (1991)
11. Caffarelli, L.A.: The regularity of mappings with a convex potential. J. Am. Math. Soc. 5(1), 99–104 (1992)
12. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. Acta Math. 177(2), 113–161 (1996)
13. Kantorovich, L.V.: On the transfer of masses. Dokl. Acad. Sci. SSSR 37, 227–229 (1942)
14. Kantorovich, L.V.: On a problem of Monge. Uspekhi Mat. Nauk. 3, 225–226 (1948)
15. McCann, R.J.: A convexity theory for interacting gases and equilibrium crystals, Ph.D. thesis, Princeton University (1994)
16. Rochet, J.C.: A necessary and sufficient condition for rationalizability in a quasi-linear context. J. Math. Econ. 16(2), 191–200 (1987)
17. Rockafellar, R.T.: Characterization of the subdifferentials of convex functions. Pac. J. Math. 17(3), 497–510 (1966)
18. Rüschendorf, L.: On c-optimal random variables. Stat. Probab. Lett. 27(3), 267–270 (1996)
19. Smith, C., Knott, M.: On Hoeffding-Fréchet bounds and cyclic monotone relations. J. Multivar. Anal. 40(2), 328–334 (1992)
20. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Stat. 36(2), 423–439 (1965)
21. Trudinger, N.S., Wang, X.: On strict convexity and continuous differentiability of potential functions in optimal transportation. Arch. Ration. Mech. Anal. 192(3), 403–418 (2009)
22. Villani, C.: Topics in optimal transportation, vol. 58. Graduate Studies in Mathematics American Mathematical Society, Ann Arbor (2003)
23. Villani, C.: Optimal transport: old and new, vol. 338. Springer Science & Business Media, Springer, New York, (2008)
24. Wyczesany, K.: Topics in high-dimensional geometry and optimal transport, Ph.D. thesis, University of Cambridge (2020)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.