A generalized dual maximizer for the Monge–Kantorovich transport problem

Mathias Beiglböck, Christian Léonard, Walter Schachermayer

A GENERALIZED DUAL MAXIMIZER FOR THE MONGE–KANTOROVICH TRANSPORT PROBLEM

MATHIAS BEIGLBÖCK, CHRISTIAN LÉONARD, AND WALTER SCHACHERMAYER

Abstract. The dual attainment of the Monge–Kantorovich transport problem is analyzed in a general setting. The spaces X, Y are assumed to be polish and equipped with Borel probability measures μ and ν. The transport cost function $c : X \times Y \to [0, \infty]$ is assumed to be Borel measurable. We show that a dual optimizer always exists, provided we interpret it as a projective limit of certain finitely additive measures. Our methods are functional analytic and rely on Fenchel's perturbation technique.

1. Introduction

We consider the *Monge-Kantorovich transport problem* for Borel probability measures μ, ν on polish spaces X, Y. See [Vil03, Vil09] for an excellent account of the theory of optimal transportation. The set $\Pi(\mu, \nu)$ consists of all Monge-Kantorovich *transport plans*, that is, Borel probability measures on $X \times Y$ which have X-marginal μ and Y-marginal ν. The *transport costs* associated to a transport plan π are given by

$$\langle c, \pi \rangle = \int_{X \times Y} c(x, y)\, d\pi(x, y). \tag{1}$$

In most applications of the theory of optimal transport, the cost function $c : X \times Y \to [0, \infty]$ is lower semicontinuous and only takes values in $\mathbb{R}_+$. But equation (1) makes perfect sense if the $[0, \infty]$-valued cost function only is Borel measurable. We therefore assume throughout this paper that $c : X \times Y \to [0, \infty]$ is a Borel measurable function which may very well assume the value $+\infty$ for "many" $(x, y) \in X \times Y$. The subset $\{c = \infty\}$ of $X \times Y$ is a set of forbidden transitions.

Optimal transport on the Wiener space [FÜ02, FÜ04a, FÜ04b, FÜ06]) and on configuration spaces [Dec08, DJS08] provide natural infinite dimensional settings where c takes infinite values.

The (primal) Monge-Kantorovich problem is to determine the primal value

$$P := \inf\{\langle c, \pi \rangle : \pi \in \Pi(\mu, \nu)\} \tag{2}$$

and to identify a primal optimizer $\hat{\pi} \in \Pi(\mu, \nu)$ which is also called an *optimal transport plan*. Clearly, without loss of generality this minimization can be performed among the *finite transport plans*, i.e. the infimum is taken over the plans $\pi \in \Pi(\mu, \nu)$ verifying $\langle c, \pi \rangle < \infty$.

The dual Monge-Kantorovich problem consists in determining

$$D := \sup \left\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \right\} \tag{3}$$

Date: September 2010.

Key words and phrases. Monge-Kantorovich problem, dual attainment, Kantorovich potential, optimal transport.

The first author acknowledges financial support from the Austrian Science Fund (FWF) under grant P21209. The third author acknowledges support from the Austrian Science Fund (FWF) under grant P19456, from the Vienna Science and Technology Fund (WWTF) under grant MA13 and by the Christian Doppler Research Association (CDG). All authors thank A. Pratelli for helpful discussions on the topic of this paper. We also thank M. Goldstern and G. Maresch for their advice.

for (φ, ψ) varying over the set of pairs of functions $\varphi : X \to [-\infty, \infty)$ and $\psi : Y \to [-\infty, \infty)$ which are *integrable*, i.e. $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$, and satisfy $\varphi \oplus \psi \leq c$. We have denoted $\varphi \oplus \psi(x, y) := \varphi(x) + \psi(y)$, $x \in X$, $y \in Y$.

We say that there is *no duality gap* if the primal value P of the problem equals the dual value D, there is primal attainment if there exists some optimal plan $\hat{\pi}$ and there is *integrable dual attainment* if the above dual Monge-Kantorovich problem is attained for some $(\hat{\varphi}, \hat{\psi})$. There is a long line of research on these questions, initiated already by Kantorovich ([Kan42]) himself and continued by numerous others (we mention [KR58, Dud76, Dud02, dA82, GR81, Fer81, Szu82, RR95, RR96, Mik06, MT06], see also the bibliographical notes in [Vil09, p 86, 87]). Important progresses were done by Kellerer [Kel84]. We also refer to the seminal paper [GM96] by Gangbo and McCann. Recently the authors of the present article have obtained in [BLS09a] a general duality result which is recalled below at Theorem 1.1.

It is well-known that there is primal attainment under the assumptions that c is lower semicontinuous and the primal value P is finite. On the other hand, it is easy to build examples where c is not lower semicontinuous and no primal minimizer exists.

In this article we focus onto the question of the dual attainment.

The dual optimizers $(\hat{\varphi}, \hat{\psi})$ are sometimes called Kantorovich potentials. In the Euclidean case with a quadratic cost, it is well-known that these potentials are convex conjugate to each other and that any optimal plan is supported by the subdifferential of $\hat{\varphi}$. In the general case, these potentials are c-conjugate to each other, a notion introduced by Rüschendorf [Rüs96].

Kellerer [Kel84, Theorem 2.21] established that integrable dual attainment holds true in the case of bounded c. This was extended by Ambrosio and Pratelli [AP03, Theorem 3.2], who gave appropriate moment conditions on μ and ν which are sufficient to guarantee the existence of integrable dual optimizers. Easy examples show that one cannot expect that the dual problem admits integrable maximizers unless the cost function satisfies certain integrability conditions with respect to μ and ν [BS09, Examples 4.4, 4.5]. In fact [BS09, Example 4.5] takes place in a very "regular" setting, where c is squared Euclidean distance on $\mathbb{R}$. In this case there exist natural candidates $(\hat{\varphi}, \hat{\psi})$ for the dual optimizer which, however, fail to be dual maximizers in the usual sense as they are not integrable.

The following solution was proposed in [BS09, Section 1.1]. If φ and ψ are integrable functions and $\pi \in \Pi(\mu, \nu)$ then

$$\int_X \varphi \, d\mu + \int_Y \psi \, d\nu = \int_{X \times Y} \varphi \oplus \psi \, d\pi. \tag{4}$$

If we drop the integrability condition on φ and ψ, the left hand side need not make sense. But if we require that $\varphi \oplus \psi \leq c$ and if π is a finite cost transport plan, i.e. $\int_{X \times Y} c \, d\pi < \infty$, then the right hand side of (4) still makes good sense, assuming possibly the value $-\infty$, and we set

$$J_c(\varphi, \psi) = \int_{X \times Y} \varphi \oplus \psi \, d\pi.$$

It is not difficult to show (see [BS09, Lemma 1.1]) that this value does not depend on the choice of the finite cost transport plan π and satisfies $J_c(\varphi, \psi) \leq D$. Under the assumption that there exists some finite transport plan, we then say that we have *measurable dual attainment* in the optimization problem (3) if there exist Borel measurable functions $\hat{\varphi} : X \to [-\infty, \infty)$ and $\hat{\psi} : Y \to [-\infty, \infty)$ verifying $\hat{\varphi} \oplus \hat{\psi} \leq c$ such that

$$D = J_c(\hat{\varphi}, \hat{\psi}). \tag{5}$$

In [BS09, Theorem 2] it was shown that, for Borel measurable $c : X \times Y \to [0, \infty]$ such that $c < \infty, \mu \otimes \nu$-almost surely, there is no duality gap and there is measurable dual attainment in the sense of (5).

A necessary and sufficient condition for the measurable dual attainment was proved in [BLS09a, Theorems 1.2 and 3.5]. We need some more notation to state this result below as Theorem 1.1. Fix $0 \leq \varepsilon \leq 1$ and define $\Pi^\varepsilon(\mu,\nu) = \{\pi \in \mathcal{M}^+_{X\times Y}, \|\pi\| \geq 1-\varepsilon, p_X(\pi) \leq \mu, p_Y(\pi) \leq \nu\}$ where $\mathcal{M}^+_{X\times Y}$ denotes the non-negative Borel measures π on $X \times Y$ with norm $\|\pi\| = \pi(X \times Y)$. By $p_X(\pi) \leq \mu$ (resp. $p_Y(\pi) \leq \nu$) we mean that the projection of π onto X (resp. onto Y) is dominated by μ (resp. ν). We denote $P^\varepsilon := \inf\{\langle c,\pi\rangle : \pi \in \Pi^\varepsilon(\mu,\nu)\}$. This partial transport problem has recently been studied by Caffarelli and McCann [CM06] as well as Figalli [Fig09]. In their work the emphasis is on a finer analysis of the Monge problem for the squared Euclidean distance on $\mathbb{R}^n$, and pertains to a fixed $\varepsilon > 0$. In the present paper, we do not deal with these more subtle issues of the Monge problem and always remain in the realm of the Kantorovich problem (2). We call

$$P^{\mathrm{rel}} := \lim_{\varepsilon\to 0} P^\varepsilon \tag{6}$$

the relaxed primal value of the transport plan. Obviously this limit exists (assuming possibly the value $+\infty$) and $P^{\mathrm{rel}} \leq P$.

Theorem 1.1 (Measurable dual attainment [BLS09a])**.** *Let X, Y be polish spaces, equipped with Borel probability measures μ,ν, and let $c : X \times Y \to [0,\infty]$ be Borel measurable.*

(a) There is no duality gap if the primal problem is defined in the relaxed form (6) while the dual problem is formulated in its usual form (3). *In other words, we have $P^{\mathrm{rel}} = D$.*

(b) Assume that in addition there exists a finite transport plan $\pi \in \Pi(\mu,\nu)$. The following statements are equivalent.

  (i) There is measurable dual attainment, i.e. there exist measurable functions $\hat\varphi, \hat\psi$ such that $\hat\varphi \oplus \hat\psi \leq c$ and $P^{\mathrm{rel}} = D = J_c(\hat\varphi,\hat\psi)$.

  (ii) There exists a $\mu\otimes\nu$-a.s. finite function $h : X \times Y \to [0,\infty]$ such that $P^{\mathrm{rel}} = P_{c\wedge h} := \inf\{\langle c \wedge h, \pi\rangle : \pi \in \Pi(\mu,\nu)\}$.

The aim of the present paper is to go beyond the setting of this theorem where the measurable dual attainment is realized. We are going to discuss the existence of an optimizer of an extension of the dual problem (3), without imposing any further conditions on the Borel measurable cost function $c : X \times Y \to [0,\infty]$.

In Theorem 3.1 we take a somewhat unothodox view at the general optimization problem. We start with a transport plan $\pi_0 \in \Pi(\mu,\nu)$ with finite cost, but which is *not* supposed to be optimal. We then optimize over all the transport plans $\pi \in \Pi(\mu,\nu)$ such that the Radon-Nikodym derivative $\frac{d\pi}{d\pi_0}$ is bounded. In this setting we show that there is no duality gap and that there is a dual optimizer. However, this dual optimizer is not given by a pair of functions $(\varphi \oplus \psi) \in L^1(\pi_0)$, but rather as a weak star limit of a sequence $(\varphi_n \oplus \psi_n)_{n=1}^\infty \in L^1(\pi_0)$ in the bidual $L^1(\pi_0)^{**}$. A rather elaborate example in the accompanying paper [BLS09b] shows that this passage to the bidual is indeed necessary, in general.

While Theorem 3.1 depends on the choice of the finite transport plan $\pi_0 \in \Pi(\mu,\nu)$, we formulate in Theorem 4.2 a result which does not depend on this choice. There we pass to a projective limit along a net of finite transport plans. Again we can prove that there is no duality gap and can identify a dual optimizer.

2. Two types of accident

In this section, we point out some difficulties which arise when going one step beyond the measurable dual attainment. We shall face two types of troubles which might be called

- measurability accident;
- singular concentration accident.

Before describing these phenomena, it is worth recalling some results from [BS09] and [Léo09] about optimal plans. The proofs of the present paper and of Theorems 2.1 and 2.2 below rely on three different types of techniques.

About the optimal plans. The following characterization of the optimal plans was proved in [BS09].

Theorem 2.1 ([BS09, Theorem 2]). *Assume that X, Y are polish spaces equipped with Borel probability measures μ, ν, that $c : X \times Y \to [0, \infty]$ is Borel measurable and $\mu \otimes \nu$-a.e. finite and that there exists a finite transport plan.*

(a) Let π be a finite transport plan and assume that there exist measurable functions $\varphi : X \to [-\infty, \infty)$ and $\psi : Y \to [-\infty, \infty)$ which satisfy

$$\begin{cases} \varphi \oplus \psi \le c & \text{everywhere} \\ \varphi \oplus \psi = c & \pi\text{-almost everywhere.} \end{cases} \tag{7}$$

Then $J_c(\varphi, \psi) = \langle c, \pi \rangle$, thus π is an optimal transport plan and φ, ψ are dual maximizers in the sense of (5).

(b) Assume that $\hat{\pi}$ is an optimal transport plan. Then $\hat{\pi}$ verifies (7) *for every pair $(\hat{\varphi}, \hat{\psi})$ of dual maximizers in the sense of* (5).

As a definition which was introduced in [ST09], a transport plan π is said to be *strongly c-cyclically monotone* if there exist measurable functions $\varphi : X \to [-\infty, \infty), \psi : Y \to [-\infty, \infty)$ which satisfy (7).

Denote by $\Pi(\mu, \nu, c)$ the set of finite cost transport plans

$$\Pi(\mu, \nu, c) := \left\{ \pi \in \Pi(\mu, \nu) : \int_{X \times Y} c \, d\pi < \infty \right\},$$

and say that a property holds $\Pi(\mu, \nu, c)$-almost everywhere if it holds true outside a measurable set N such that $\pi(N) = 0$, for all $\pi \in \Pi(\mu, \nu, c)$.

In [Léo09], the assumption that c is $\mu \otimes \nu$-a.e. finite was removed under the extra requirement that c is lower semicontinuous and the following analogous results were obtained.

Theorem 2.2 ([Léo09]). *Assume that X, Y are polish spaces equipped with Borel probability measures μ, ν, that $c : X \times Y \to [0, \infty]$ is lower semicontinuous and that there exists a finite transport plan.*

(a) Let π be a finite plan and assume that there exist measurable functions $\varphi : X \to [-\infty, \infty)$ and $\psi : Y \to [-\infty, \infty)$ which satisfy

$$\begin{cases} \varphi \oplus \psi \le c & \Pi(\mu, \nu, c)\text{-almost everywhere} \\ \varphi \oplus \psi = c & \pi\text{-almost everywhere.} \end{cases} \tag{8}$$

Then $J_c(\varphi, \psi) = \langle c, \pi \rangle$, thus π is an optimal transport plan and φ, ψ are dual maximizers in the sense of (5).

(b) Take any optimal plan $\hat{\pi}$, $\epsilon > 0$ and π_o any probability measure on $X \times Y$ such that $\int_{X \times Y} c \, d\pi_o < \infty$. Then, there exist functions $h \in L^1(\hat{\pi} + \pi_o)$, φ and ψ bounded continuous on X and Y respectively and a measurable subset $Z_\epsilon \subset (X \times Y)$ such that

- *(i) $h = c$, $\hat{\pi}$-almost everywhere on $(X \times Y) \setminus Z_\epsilon$;*
- *(ii) $\int_{Z_\epsilon} (1 + c) \, d\hat{\pi} \le \epsilon$;*
- *(iii) $-c/\epsilon \le h \le c$, $(\hat{\pi} + \pi_o)$-almost everywhere;*
- *(iv) $-c/\epsilon \le \varphi \oplus \psi \le c$, everywhere;*
- *(v) $\|h - \varphi \oplus \psi\|_{L^1(\hat{\pi} + \pi_o)} \le \epsilon$.*

As regards (a), the examples [BGMS09, Example 5.1] and [BS09, Example 4.2] exhibit optimal plans which are not strongly c-cyclically monotone but which satisfy the weaker property (8). As regards (b), let us emphasize the appearance of the probability measure π_o in items (iii) and (v). One can read $(iii\text{-}v)$ as an approximation of $\varphi \oplus \psi \le c$, $(\hat{\pi} + \pi_o)$-a.e. Since it is required that $\int_{X \times Y} c \, d\pi_o < \infty$, one can choose π_o in $\Pi(\mu, \nu, c)$, and the properties $(i\text{-}v)$ are an approximation of (8) where $\Pi(\mu, \nu, c)$-a.e. is replaced by the weaker $(\hat{\pi} + \pi_o)$-a.e.

Note also that for any (φ, ψ) verifying (7) or (8) with $\pi \in \Pi(\mu, \nu, c)$, we have

$$\mu(\varphi = -\infty) = \nu(\psi = -\infty) = 0. \tag{9}$$

As a consequence of this remark and a result of Kellerer [Kel84], see [BLS09a, Lemma A.1], we can replace"$\varphi \oplus \psi \leq c$ everywhere" in (7) by "$\varphi \oplus \psi \leq c$, $\Pi(\mu, \nu)$-almost everywhere." The comparison between (7) and (8) becomes clearer.

Measurability accident. To develop a feeling for what we are after, we consider a specific example.

Example 2.3 (Ambrosio-Pratelli, [AP03, Example 3.2])**.** Let $X = Y = [0, 1)$, equipped with Lebesgue measure $\lambda = \mu = \nu$. Pick $\alpha \in [0, 1)$ irrational. Set

$$\Gamma_0 = \{(x, x) : x \in X\} \quad \Gamma_1 = \{(x, x \oplus \alpha) : x \in X\},$$

where $\oplus$ is addition modulo 1. Define $c : X \times Y \to [0, \infty]$ by

$$c(x, y) = \begin{cases} 1 & \text{for } (x, y) \in \Gamma_0 \\ 2 & \text{for } (x, y) \in \Gamma_1, x \in [0, 1/2) \\ 0 & \text{for } (x, y) \in \Gamma_1, x \in [1/2, 1) \\ \infty & \text{else} \end{cases}.$$

This cost function is a variation on [AP03]'s original example which has been proposed in [BS09, Example 4.3]. For $i = 0, 1$, let π_i be the obvious transport plan supported by Γ_i. Following the arguments of [AP03], it is easy to see that all finite transport plans are given by convex combinations of the form $\rho\pi_0 + (1 - \rho)\pi_1, \rho \in [0, 1]$ and each of these transport plans leads to costs of 1.
Note that since c is lower semicontinuous, there is no duality gap. This was proved in [Kel84] and is an easy consequence of Theorem 1.1-(a). Thus, for each $\varepsilon > 0$, there are integrable functions $\varphi, \psi : [0, 1) \to [-\infty, \infty)$ such that $\varphi \oplus \psi \leq c$ and $0 \leq \int (c - \varphi \oplus \psi) \, d\pi_i \leq \varepsilon$ for $i = 0, 1$.
On the other hand, it is shown in [BS09] that there do not exist *measurable* functions $\varphi, \psi : [0, 1) \to [-\infty, \infty)$ satisfying $\varphi \oplus \psi \leq c$ such that $\varphi \oplus \psi = c$ holds π_0- as well as π_1-almost surely.

Let us have a closer look at the previous example: while it is *not possible* to find Borel measurable limits $\hat{\varphi}, \hat{\psi}$ of an optimizing sequence $(\varphi_n, \psi_n)_{n=1}^{\infty}$, it *is possible* to find a limiting Borel function $\hat{h}(x, y)$ of the sequence of functions $(\varphi_n(x) + \psi_n(y))_{n=1}^{\infty}$ on the set $\{(x, y) \in X \times Y : c(x, y) < \infty\}$. Indeed, on this set, which simply equals $\Gamma_0 \cup \Gamma_1$, any optimizing sequence $(\varphi_n(x) + \psi_n(y))_{n=1}^{\infty}$ for (3) has a subsequence which converges π-a.s. to $\hat{h}(x, y) := c(x, y)$, for any finite cost transport plan π.

Summing up: in the context of the previous example, there is a Borel function $\hat{h}(x, y)$ on $X \times Y$, which equals $c(x, y)$ on $\Gamma_0 \cup \Gamma_1$; it may take any value on $(X \times Y) \setminus (\Gamma_0 \cup \Gamma_1)$, e.g. the value $+\infty$. This function $\hat{h}(x, y)$ may be considered as a kind of dual optimizer: it is, for any finite cost transport plan π, the limit of an optimizing sequence $(\varphi_n(x) + \psi_n(y))_{n=1}^{\infty}$ with respect to the norm $\| \cdot \|_{L^1(\pi)}$.

Singular concentration accident. One can rewrite the sufficient conditions of Theorems 2.1-(a) and 2.2-(a) as follows: $\hat{\pi}$ and $(\hat{\varphi}, \hat{\psi})$ solve the primal and dual problems if $\hat{\pi} \in \Pi(\mu, \nu, c)$, $(\hat{\varphi} \oplus \hat{\psi})\hat{\pi} = c\hat{\pi}$ and $(\hat{\varphi} \oplus \hat{\psi})\pi \leq c\pi$, $\forall \pi \in \Pi(\mu, \nu, c)$, in the space of bounded measures. In view of Example 2.3 and of part (b) of Theorem 2.2, we are aware that $\hat{\varphi} \oplus \hat{\psi}$ should be replaced by a jointly measurable $\hat{h}$ such that for each $\pi \in \Pi(\mu, \nu, c)$, $\hat{h}\pi$ can be approximated in variation norm by a sequence $((\varphi_n \oplus \psi_n)\pi)_{n=1}^{\infty}$ verifying $(\varphi_n \oplus \psi_n)\pi \leq c\pi$ for all $n \geq 1$. But this is not the end of the story.

In the accompanying paper [BLS09b], rather elaborate extensions of the above example are analyzed. By means of examples (which are too long to be recalled here), it is shown that instead of the functions or, equivalently, countably additive measures $\hat{h}\pi$, one has to consider finitely additive measures. This might be seen as a consequence of the limiting behavior of functions $\varphi \oplus \psi$ tending to $-\infty$ somewhere, under the seemingly contradictory requirement (9).

3. Existence of a dual optimizer

The remainder of this article is devoted to developing a theory which makes this circle of ideas precise in the general setting of Borel measurable cost functions $c : X \times Y \to [0, \infty]$. To do so we shall apply Fenchel's perturbation method as in [BLS09a]. In addition, we need some functional analytic machinery, in particular we shall use the space $(L^1)^{**} = (L^\infty)^*$ of finitely additive measures.

Assume $\Pi(\mu, \nu, c) \neq \emptyset$ to avoid the trivial case.

We fix $\pi_0 \in \Pi(\mu, \nu, c)$ and stress that we do *not* assume that π_0 has minimal transport cost. In fact, there is little reason in the present setting (where c is not assumed to be lower semicontinuous) why a primal optimizer $\widehat{\pi}$ should exist. We denote by $\Pi^{(\pi_0)}(\mu, \nu)$ the set of elements $\pi \in \Pi(\mu, \nu)$ such that $\pi \ll \pi_0$ and $\left\| \frac{d\pi}{d\pi_0} \right\|_{L^\infty(\pi_0)} < \infty$. Note that $\Pi^{(\pi_0)}(\mu, \nu) = \Pi(\mu, \nu) \cap L^\infty(\pi_0) \subseteq \Pi(\mu, \nu, c)$.

We shall replace the usual Kantorovich optimization problem over the set $\Pi(\mu, \nu, c)$ by the optimization over the smaller set $\Pi^{(\pi_0)}(\mu, \nu)$ and consider

$$P^{(\pi_0)} = \inf\{\langle c, \pi \rangle = \textstyle\int c \, d\pi : \pi \in \Pi^{(\pi_0)}(\mu, \nu)\}. \tag{10}$$

As regards the dual problem, we define for $\varepsilon > 0$,

$$\begin{aligned} D^{(\pi_0,\varepsilon)} = \sup \Big\{ \int \varphi \, d\mu + \int \psi \, d\nu : \varphi \in L^1(\mu), \psi \in L^1(\nu), \\ \int_{X \times Y} (\varphi \oplus \psi - c)_+ \, d\pi_0 \leq \varepsilon \Big\} \quad \text{and} \\ D^{(\pi_0)} = \lim_{\varepsilon \to 0} D^{(\pi_0,\varepsilon)}. \end{aligned} \tag{11}$$

Define the "summing" map S by

$$\begin{aligned} S : L^1(X, \mu) \times L^1(Y, \nu) &\to L^1(X \times Y, \pi_0) \\ (\varphi, \psi) &\mapsto \varphi \oplus \psi \end{aligned}$$

and denote by $L^1_S(X \times Y, \pi_0)$ the $\|.\|_1$-closed linear subspace of $L^1(X \times Y, \pi_0)$ spanned by $S(L^1(X, \mu) \times L^1(Y, \nu))$. Clearly $L^1_S(X \times Y, \pi_0)$ is a Banach space under the norm $\|.\|_1$ induced by $L^1(X \times Y, \pi_0)$.

We shall also need the bi-dual $L^1_S(X \times Y, \pi_0)^{**}$ which may be identified with a subspace of $L^1(X \times Y, \pi_0)^{**}$. In particular, an element $h \in L^1_S(X \times Y, \pi_0)^{**}$ can be decomposed into $h = h^r + h^s$, where $h^r \in L^1(X \times Y, \pi_0)$ is the regular part of the finitely additive measure h and h^s its purely singular part. Note that it may happen that $h \in L^1_S(X \times Y, \pi_0)^{**}$ while $h^r \notin L^1_S(X \times Y, \pi_0)$, and therefore also $h^s \notin L^1_S(X \times Y, \pi_0)^{**}$.

Theorem 3.1. *Let $c : X \times Y \to [0, \infty]$ be Borel measurable and let $\pi_0 \in \Pi(\mu, \nu, c)$ be a finite transport plan. We have*

$$P^{(\pi_0)} = D^{(\pi_0)}. \tag{12}$$

There is an element $\hat{h} \in L^1_S(X \times Y, \pi_0)^{**}$ which verifies the inequality[1] *$\hat{h} \leq c$ in the Banach lattice $L^1(X \times Y, \pi_0)^{**}$ and*

$$D^{(\pi_0)} = \langle \hat{h}, \pi_0 \rangle.$$

[1]The inequality $\hat{h} \leq c$ pertains to the lattice order of $L^1(X \times Y)^{**}$, where we identify the π_0-integrable function c with an element of $L^1(X \times Y, \pi_0)^{**}$. If $\hat{h}$ decomposes into $\hat{h} = \hat{h}^r + \hat{h}^s$, the inequality $\hat{h} \leq c$ holds true if and only if $\hat{h}^r(x, y) \leq c(x, y)$, π_0-a.s. and $\hat{h}^s \leq 0$ (compare the discussion after (18))

If $\pi \in \Pi^{(\pi_0)}(\mu,\nu)$ (identifying π with $\frac{d\pi}{d\pi_0}$) satisfies $\int c\,d\pi \le P^{(\pi_0)} + \alpha$ for some number $\alpha \ge 0$, then

$$-\alpha \le \langle \hat{h}^s, \pi\rangle \le 0. \tag{13}$$

In particular, if π is an optimizer of (10), then $\hat{h}^s$ vanishes on the set $\{\frac{d\pi}{d\pi_0} > 0\}$.
In addition, we may find a sequence of elements $(\varphi_n, \psi_n) \in L^1(\mu) \times L^1(\nu)$ such that

$$\begin{aligned}
&\varphi_n \oplus \psi_n \to \hat{h}^r, \ \pi_0\text{-a.s.},\\
&\|(\varphi_n \oplus \psi_n - \hat{h}^r)_+\|_{L_1(\pi_0)} \to 0 \quad \text{and}\\
&\lim_{\delta\to 0} \sup_{A\subseteq X\times Y, \pi_0(A)<\delta} \lim_{n\to\infty} -\langle(\varphi_n \oplus \psi_n)\mathbb{1}_A, \pi_0\rangle = \|\hat{h}^s\|_{L_1(\pi_0)^{**}}.
\end{aligned} \tag{14}$$

Proof. It is straightforward to verify the trivial duality relation $D^{(\pi_0)} \le P^{(\pi_0)}$. To show the reverse inequality and to find the dual optimizer $\hat{h} \in L^1(X \times Y, \pi_0)^{**}$, as in [BLS09a] we apply W. Fenchel's perturbation argument. (For an elementary treatment, compare also [BLS09b].) The summing map S factors through $L^1_S(\pi_0)$ as indicated in the subsequent diagram:

$$\begin{tikzcd}
L^1(\mu)\times L^1(\nu) \arrow[rr, "S"] \arrow[dr, "S_1"'] & & L^1(\pi_0) \\
& L^1_S(\pi_0) \arrow[ur, "S_2"'] &
\end{tikzcd}$$

Then S_1 has dense range and S_2 is an isometric embedding. Denote by $\big(L^1_S(\pi_0)^*, \|.\|_{L^1_S(\pi_0)^*}\big)$ the dual of $L^1_S(\pi_0)$ which is a quotient space of $L^\infty(\pi_0)$. Transposing the above diagram we get

$$\begin{tikzcd}
L^\infty(\mu)\times L^\infty(\nu) & & L^\infty(\pi_0) \arrow[ll, "T"'] \arrow[dl, "T_2"] \\
& L^1_S(\pi_0)^* \arrow[ul, "T_1"] &
\end{tikzcd}$$

where T, T_1, T_2 are the transposed maps of S, S_1, resp. S_2. Clearly $T(\gamma) = (p_X(\gamma), p_Y(\gamma))$ for $\gamma \in L^\infty(\pi_0)$, where p_X, p_Y are the projections of a measure γ (identified with the Radon-Nikodym-derivative $\frac{d\gamma}{d\pi_0}$) onto its marginals. By elementary duality relations we have that T_2 is a quotient map and T_1 is injective; the latter fact allows us to identify the space $L^1_S(\pi_0)^*$ with a subspace of $L^\infty(\mu) \times L^\infty(\nu)$.
For example, consider the element $\mathbf{1} \in L^\infty(\pi_0)$, which corresponds to the measure π_0 on $X \times Y$. The element $T_2(\mathbf{1}) \in L^1_S(\pi_0)^*$ may then be identified with the element $(\mathbf{1},\mathbf{1}) = T(\mathbf{1})$ in $L^\infty(\mu) \times L^\infty(\nu)$ which corresponds to the pair (μ,ν). We take the liberty to henceforth denote this element simply by $\mathbf{1}$, independently of whether we consider it as an element of $L^\infty(\pi_0)$, $L^1_S(\pi_0)^*$ or $L^\infty(\mu) \times L^\infty(\nu)$.

We may now rephrase the primal problem (10) as

$$\langle c, \gamma\rangle = \int_{X\times Y} c(x,y)\,d\gamma(x,y) \to \min, \quad \gamma \in L^\infty_+(\pi_0),$$

under the constraint

$$T(\gamma) = \mathbf{1}. \tag{15}$$

The decisive trick is to replace (15) by the trivially equivalent constraint

$$T_2(\gamma) = \mathbf{1},$$

and to perform the Fenchel perturbation argument *not* in the space $L^\infty(\mu) \times L^\infty(\nu)$ but rather in the subspace $L^1_S(\pi_0)^*$ which is endowed with a *stronger norm.* The map Φ: $L^1_S(\pi_0)^* \to [0,\infty]$,

$$\Phi(p) := \inf\{\langle c,\gamma\rangle : \gamma \in L^\infty_+(\pi_0), T_2(\gamma) = p\}, \quad p \in L^1_S(\pi_0)^*,$$

is convex, positively homogeneous and $\Phi(\mathbf{1}) = P^{(\pi_0)}$.

Claim. *There is a neighbourhood V of $\mathbf{1}$ in $L^1_S(\pi_0)^*$ on which Φ is bounded.*
Indeed, let $U = \{\gamma \in L^\infty(\pi_0) \mid \|\gamma - \mathbf{1}\|_{L^\infty(\pi_0)} < \frac{1}{2}\}$. Then U is contained in the positive orthant $L^\infty_+(\pi_0)$ of $L^\infty(\pi_0)$ and

$$\Phi(T_2(\gamma)) \le \langle c, \gamma\rangle \le \tfrac{3}{2}\|c\|_{L^1(\pi_0)} \text{ for all } \gamma \in U.$$

Hence on $T_2(U)$, which simply is the open ball of radius $\frac{1}{2}$ around $\mathbf{1}$ in the Banach space $L^1_S(\pi_0)^*$, we have that Φ is bounded by $\frac{3}{2}\|c\|_{L^1(\pi_0)}$.

It follows from elementary geometric facts that the convex function Φ is continuous on $T_2(U)$ with respect to the norm of $L^1_S(\pi_0)^*$. By Hahn-Banach there exists $f \in L^1_S(\pi_0)^{**}$ such that

$$\langle f, \mathbf{1}\rangle = \Phi(\mathbf{1}),$$
$$\langle f, p\rangle \le \Phi(p) \text{ for all } p \in L^1_S(\pi_0)^*.$$

The adjoint T_2^* of T_2 maps $L^1_S(\pi_0)^{**}$ isometrically onto a subspace E of $L^1(\pi_0)^{**} = L^\infty(\pi_0)^*$. The space E consists of those elements of $L^1(\pi_0)^{**}$ which are σ^*-limits of nets $(\varphi_\alpha \oplus \psi_\alpha)_{\alpha\in I}$ with $\varphi_\alpha \in L^1(\mu)$, $\psi_\alpha \in L^1(\nu)$. Write $\hat h := T_2^*(f)$. Then for all $\gamma \in L^\infty_+(\pi_0)$,

$$\langle \hat h, \gamma\rangle = \langle T_2^*(f), \gamma\rangle = \langle f, T_2(\gamma)\rangle \ \le\ \Phi(T_2(\gamma)) \ \le\ \langle c, \gamma\rangle, \tag{16}$$

and if $\pi \in L^\infty_+(\pi_0), T_2(\pi) = \mathbf{1}$ then

$$\langle \hat h, \pi\rangle = \langle T_2^*(f), \pi\rangle = \langle f, T_2(\pi)\rangle = \langle f, \mathbf{1}\rangle = \Phi(\mathbf{1}) = P^{(\pi_0)}. \tag{17}$$

By (16), the inequality $\hat h \le c$ holds true in the Banach-lattice $L^\infty(\pi_0)^*$. Combining this with (17) we obtain that $\hat h$ is a dual optimizer in the sense of

$$\begin{aligned} D^{(\pi_0)}_{**} := \sup\big\{\langle g, \pi_0\rangle : g \in L^1_S(\pi_0)^{**},\ g \le c \\ \text{in the Banach lattice } L^1(\pi_0)^{**}\big\}\end{aligned} \tag{18}$$

(where we identify π_0 with the element $\mathbf{1}$ of $L^\infty(\pi_0)$) and that there is no duality gap in this sense, i.e. $D^{(\pi_0)}_{**} = P^{(\pi_0)}$.

As mentioned above, every element $g \in L^\infty(\pi_0)^*$ splits in a regular part g^r lying in $L^1(\pi_0)$ and a purely singular part g^s. Given $g_1, g_2 \in L^\infty(\pi_0)^*$, we have $g_1 \le g_2$ if and only if $g_1^r \le g_2^r$ and $g_1^s \le g_2^s$. Since $c \in L^1(\pi_0)$ we have $c^s = 0$. The inequality $\hat h \le c$ implies that $\hat h^s \le c^s = 0$ and $\hat h^r \le c^r = c$. It follows that for each $\pi \in L^\infty_+(\pi_0)$

$$\langle \hat h^r, \pi\rangle \le \langle c, \pi\rangle. \tag{19}$$

Assume additionally that π satisfies $T_2(\pi) = \mathbf{1}$ and choose $\alpha \ge 0$ such that $\langle c, \pi\rangle \le P^{(\pi_0)} + \alpha$. Then $\langle \hat h, \pi\rangle = P^{(\pi_0)}$ and subtracting this quantity from (19) we get

$$\langle -\hat h^s, \pi\rangle = \langle \hat h^r - \hat h, \pi\rangle \le \langle c, \pi\rangle - P^{(\pi_0)} \le \alpha$$

showing (13).

We still have to show the existence of a *sequence* $(\varphi_n, \psi_n)_{n=1}^\infty$ satisfying the above assertions about convergence. So far we know that there is a net $(\varphi_\alpha, \psi_\alpha)_{\alpha\in I}$ such that $\varphi_\alpha \oplus \psi_\alpha$ weak-star converges to $\hat h$. First we claim that there exists a net $(f_\alpha)_{\alpha\in I}$ of elements of $L^1(\pi_0)$, such that $\|f_\alpha\|_1 \le \|\hat h^s\|$, $\hat h^r + f_\alpha \in L^1_S(\pi_0)$ and $\hat h^r + f_\alpha \to \hat h$ in the σ^*-topology. To see this, note that Alaoglu's theorem [RS80, Theorem IV.21] implies that in a Banach space V, the unit ball $B_1(V)$ is σ^*-dense in the unit ball $B_1(V^{**})$ of the bidual. Thus $\hat h^r + \|\hat h^s\| B_1(L^1_S(\pi_0))$ is σ^*-dense in $\hat h^r + \|\hat h^s\| B_1(L^1_S(\pi_0)^{**})$ which yields the existence of a net $(f_\alpha)_{\alpha\in I}$ as required.

As $\hat{h}^s$ is purely singular, we may find a sequence $(\alpha_n)_{n=1}^\infty$ in I such that $\|f_{\alpha_n}\| \leq \|\hat{h}^s\|$ and $\int f_{\alpha_n}\, d\pi_0 = -\|\hat{h}^s\| + 2^{-n}$, and that $\int (|f_{\alpha_n}| \wedge 2^n)\, d\pi_0 \leq 2^{-n}$, which implies that the sequence $(f_{\alpha_n})_{n=1}^\infty$ converges π_0-a.s. to zero.

As $\hat{h}^r + f_{\alpha_n} \in L^1_S(\pi_0)$ we may find $(\varphi_n, \psi_n) \in L^1(\mu) \times L^1(\nu)$ such that

$$\|\varphi_n \oplus \psi_n - (\hat{h}^r + f_{\alpha_n})\|_{L^1(\pi_0)} < 2^{-n}.$$

We then have that $(\varphi_n \oplus \psi_n)_{n=1}^\infty$ converges π_0-a.s. to $\hat{h}^r$ and that $\|(\varphi_n \oplus \psi_n - \hat{h}^r)_+\|_{L^1(\pi_0)} \to 0$.

As regards assertion (14) we note that, for $A_m = \bigcup_{n=m+1}^\infty \{|f_{\alpha_n}| > 2^{-n}\}$ we have $\pi_0(A_m) \leq 2^{-m}$ and

$$\begin{aligned}
\liminf_{n\to\infty} (-\langle (\varphi_n \oplus \psi_n) \mathbb{1}_{A_m}, \pi_0 \rangle) &= -\limsup_{n\to\infty} \langle (\hat{h}^r + f_{\alpha_n}) \mathbb{1}_{A_m}, \pi_0 \rangle \\
&= -\langle \hat{h}^r \mathbb{1}_{A_m}, \pi_0 \rangle - \lim_{n\to\infty} \langle f_{\alpha_n} \mathbb{1}_{A_m}, \pi_0 \rangle \\
&= -\langle \hat{h}^r \mathbb{1}_{A_m}, \pi_0 \rangle + \|\hat{h}^s\|_{L^1(\pi_0)^{**}}.
\end{aligned}$$

Letting m tend to infinity we obtain that the left hand side of (14) is greater than or equal to the right hand side. As regards the reverse inequality it suffices to note that $\|f_{\alpha_n}\|_{L^1(\pi_0)} \leq \|\hat{h}^s\|_{L^1(\pi_0)^{**}}$.

As $\hat{h}^r \leq c$, π_0-a.s., we obtain in particular that $\|(\varphi_n \oplus \psi_n - c)_+\|_{L^1(\pi_0)} \to 0$ showing that $D^{(\pi_0)} \geq P^{(\pi_0)}$ and therefore (12), the reverse inequality being straightforward. ☐

As a by-product of this proof, we have shown in (18) that

$$D^{(\pi_0)}_{**} = D^{(\pi_0)} = P^{(\pi_0)}. \tag{20}$$

Admittedly, Theorem 3.1 is rather abstract. However, we believe that it may be useful in applications to have the possibility to pass to *some kind of limit* $\hat{h}$ of an optimizing sequence $(\varphi_n, \psi_n)_{n=1}^\infty$ in the dual optimization problem, even if this limit is somewhat awkward. To develop some intuition for the message of Theorem 3.1, we shall illustrate the situation at the hand of some examples.

Let us start with Example 2.3. In this case we may apply Theorem 3.1 to the finite transport plan $\pi_{\frac{1}{2}} = \frac{1}{2}(\pi_0 + \pi_1)$, (we apologize for using $\pi_{\frac{1}{2}}$ instead of π_0 in Theorem 3.1 as the notation π_0 is already taken). As we have seen above, there are sequences $(\varphi_n \oplus \psi_n)_{n=1}^\infty$ converging $\pi_{\frac{1}{2}}$-a.s. as well as in the norm of $L^1(\pi_{\frac{1}{2}})$ to $\hat{h} = c$, as defined in Example 2.3 above. In particular we do not have to bother about the singular part $\hat{h}^s$ of $\hat{h}$, as we have $\hat{h} = \hat{h}^r$ in this example. We find again that h represents the limit of $(\varphi_n \oplus \psi_n)_{n=1}^\infty$, considered as a Borel function on $\{c < \infty\}$ which is the support of $\pi_{\frac{1}{2}}$.

We now make the example a bit more interesting and challenging. (See Example 3.2 below.)

Fix in the context of Example 2.3 (where we now write $\tilde{c}$ instead of c to keep the letter c free for a new function to be constructed) a sequence $(\varphi_n, \psi_n)_{n=1}^\infty$ such that $\|\tilde{c} - \varphi_n \oplus \psi_n\|_{L^1(\pi_i)} \to 0$ for $i = 0, 1$. We claim that $(\varphi_n \oplus \psi_n)_{n=1}^\infty$ converges in $\|.\|_{L^1(\pi_k)}$ where, for each $k \in \mathbb{N}$, π_k is the measure which is uniformly distributed on

$$\Gamma_k = \{(x, x \oplus k\alpha) : x \in [0, 1)\}. \tag{21}$$

Let us prove this convergence whose precise statement is given below at (26) and (27). We know that[2]

$$\begin{aligned}
\varphi_n(x)+\psi_n(x) &\to \tilde{c}(x,x) \text{ and} &(22)\\
\varphi_n(x)+\psi_n(x\oplus\alpha) &\to \tilde{c}(x,x\oplus\alpha), \text{whence}\\
\psi_n(x\oplus\alpha)-\psi_n(x) &\to \underbrace{\tilde{c}(x,x\oplus\alpha)-\tilde{c}(x,x)}_{=:g(x)} = \begin{cases} +1 & \text{for } x\in[0,\frac12),\\ -1 & \text{for } x\in[\frac12,1).\end{cases} &(23)
\end{aligned}$$

Replacing x by $x\oplus i\alpha,\ i=1,\dots,k-1$ in (23) this yields

$$\psi_n(x\oplus\alpha)-\psi_n(x)\to\sum_{i=0}^{k-1} g(x\oplus i\alpha).$$

Combined with (22) we have

$$\begin{aligned}
&\lim_{n\to\infty}[\varphi_n(x)+\psi_n(x\oplus k\alpha)] = 1+\sum_{i=0}^{k-1} g(x\oplus i\alpha) &(24)\\
&= \ 1+\ \#\left\{0\le i<k : x\oplus i\alpha\in[0,\tfrac12)\right\} - \ \#\left\{0\le i<k : x\oplus i\alpha\in[\tfrac12,1)\right\}\\
&=: \ \rho_k(x). &(25)
\end{aligned}$$

Define the function h on $X\times Y$

$$h(x,y)=\begin{cases}\rho_k(x) & \text{for } (x,y)\in\Gamma_k, k\in\mathbb{N},\\ \infty & \text{else.}\end{cases} \tag{26}$$

By (24), we have, for each $k\in\mathbb{N}$, $\lim_n \|h-\varphi_n\oplus\psi_n\|_{L^1(\pi_k)}=0$. Somewhat more precisely, one obtains that

$$\|h-\varphi_n\oplus\psi_n\|_{L^1(\pi_k)}\le k\|\tilde{c}-\varphi_n\oplus\psi_n\|_{L^1(\pi_0+\pi_1)}. \tag{27}$$

Now we shall modify the cost function $\tilde{c}$ of Example 2.3 by defining it to be finite not only on $\Gamma_0\cup\Gamma_1$, but rather on $\bigcup_{k\in\mathbb{N}}\Gamma_k$. We then obtain the following situation.

Example 3.2. Using (26) define $c:[0,1)\times[0,1)\to[0,\infty]$ by

$$c(x,y)=h(x,y)_+,$$

so that $\{c<\infty\}=\bigcup_{k\in\mathbb{N}}\Gamma_k$. For the resulting optimal transport problem we then find:

(i) The primal value P of the problem (2) equals zero and $\hat\varphi=\hat\psi=0$ are (trivial) optimizers of the dual problem (3).

(ii) For strictly positive scalars $(a_k)_{k\ge0}$, normalized by $\sum_{k\ge0}a_k=1$ apply Theorem 3.1 to the transport plan $\pi:=\sum_{k\ge0}a_k\pi_k$. (Again we apologize for using the notation π for the measure π_0 in Theorem 3.1, as all the letters π_k are already taken.) If $(a_k)_{\ge0}$ tends sufficiently fast to zero, as $|k|\to\infty$, the following facts are verified.
   - The primal value is
   $$P^{(\pi)}=\inf\left\{\int_{X\times Y}c\,d\bar\pi : \bar\pi\in\Pi(\mu,\nu), \|\tfrac{d\bar\pi}{d\pi}\|_{L^\infty}<\infty\right\}=1.$$
   - The Borel function $h\in L^1(\pi)$ defined in (26) is a dual optimizer in the sense of Theorem 3.1, i.e.
   $$D^{(\pi)}=\int_{X\times Y}h\,d\pi=1.$$
   - There is a sequence $(\varphi_n,\psi_n)_{n=1}^\infty$ in $L^1(\mu)\times L^1(\nu)$ such that $(\varphi_n\oplus\psi_n)_{n=1}^\infty$ converges to h in the norm of $L^1(\pi)$.

[2] The equations (22) to (25) refer to integrable functions on $[0,1)$ and convergence is understood to be with respect to $\|\cdot\|_{L^1(\mu)}$.

Before proving the above assertions let us draw one conclusion: in (ii) we *can not assert* that the functions $(\varphi_n, \psi_n)_{n=1}^{\infty}$ satisfy – in addition to the properties above – the inequality $\varphi_n(x) + \psi_n(y) \le c(x,y)$, for all $(x,y) \in X \times Y$. Indeed, if this were possible then, because of $\lim_{n\to\infty}(\int_X \varphi_n \, d\mu + \int_Y \psi_n \, d\nu) = D^{(\pi)} = 1$, we would have that the dual value D of the original dual problem (3) would equal $D = 1$, in contradiction to (i).

Proof of the assertions of Example 3.2. We start with assertion (ii). Fix an optimizing sequence $(\varphi_n, \psi_n)_{n=1}^{\infty}$ in the context of Example 2.3 such that

$$\|\tilde{c} - \varphi_n \oplus \psi_n\|_{L^1(\pi_0+\pi_1)} \le 1/n^3. \tag{28}$$

Pick a sequence $(a_k)_{k\in\mathbb{N}}$ of positive numbers such that

(a) $a_k \|h\|_{L^1(\pi_k)} \le C 2^{-k}$ for all $k \in \mathbb{N}$,
(b) $a_k(\|\varphi_n\|_1 + \|\psi_n\|_1) \le C 2^{-k}$ for all $k \in \mathbb{N}$ with $n \le k$,

for some real constant C. After re-normalizing, if necessary, we may assume that $\sum_{k=1}^{\infty} a_k = 1$. Set $\pi := \sum_{k=1}^{\infty} a_k \pi_k$. From (a) we obtain $h \in L^1(\pi) \subseteq L^1(\pi)^{**}$ thus h is viable for the problem $D_{**}^{(\pi)}$ and hence $D_{**}^{(\pi)} \ge 1$. Clearly $P^{(\pi)} \le 1$, hence $P^{(\pi)} = D_{**}^{(\pi)} = 1$ and h is a dual maximizer. Combining (28) with (27) we obtain

$$\|h - \varphi_n \oplus \psi_n\|_{L^1(\pi_k)} \le k/n^3.$$

Therefore

$$\begin{aligned} \|h - \varphi_n \oplus \psi_n\|_{L^1(\pi)} &\le \sum_{k\le n} \|h - \varphi_n \oplus \psi_n\|_{L^1(\pi_k)} + \sum_{k>n} a_k(\|h\|_{L^1(\pi_k)} + \|\varphi_n\|_1 + \|\psi_n\|_1) \\ &\le 1/n + 2C \sum_{k>n} 2^{-k}. \end{aligned}$$

Hence $\varphi_n \oplus \psi_n$ converges to h in $\|.\|_{L^1(\pi)}$. This shows assertion (ii) above.

To obtain (i) we construct a transport plan $\pi_\beta \in \Pi(\mu,\nu)$ such that $\int_{X\times Y} c \, d\pi_\beta = 0$. Note in passing that in view of (ii) we must have $\|\frac{d\pi_\beta}{d\pi}\|_{L^\infty(\pi)} = \infty$ for the π constructed above. On the other hand, we must have $\frac{d\pi_\beta}{d\pi} \in L^1(\pi)$, if $a_k > 0$ for all $k \in \mathbb{N}$, as every finite cost transport plan must be absolutely continuous with respect to π.

The idea is to concentrate π_β on the set

$$\begin{aligned} \Gamma &:= \{(x,y) : c(x,y) = 0\} \\ &= \{(x, x \oplus k\alpha) : k \ge 1, \textstyle\sum_{i=0}^{k-1} (\mathbb{1}_{[0,\frac{1}{2})}(x \oplus i\alpha) - \mathbb{1}_{[\frac{1}{2},1)}(x \oplus i\alpha)) \le -1\}. \end{aligned}$$

To prove that this can be done it is sufficient to show that whenever $A \subseteq X$, $B \subseteq Y$, $\mu(A), \nu(B) > 0$, a subset A' of A can be transported to a subset B' of B with $\nu(B') = \mu(A') > 0$ via Γ. Then an exhaustion argument applies.

At this stage we encounter an interesting connection to the theory of measure preserving systems. For $x \in X$ and $m \in \mathbb{N}$ set

$$S(x,m) := \Big(x \oplus \alpha, m + \mathbb{1}_{[0,\frac{1}{2})}(x) - \mathbb{1}_{[\frac{1}{2},1)}(x)\Big).$$

Then S is a measure preserving transformation of the space $([0,1] \times \mathbb{Z}, \lambda \times \#)$. (See [Aar97] for an introduction to infinite ergodic theory and the basic definitions in this field.) It is not hard to see that the ergodic theorem, applied to the rotation by α on the torus, shows that S is non wandering. Much less trivial is the fact that S is also ergodic. This was shown by K. Schmidt [Sch78] for a certain class of irrational numbers $\alpha \in [0,1)$, and in full generality by M. Keane and J.-P. Conze [CK76], see also [AK82].

The relevance of these facts to our situation is that for $k \ge 1$, the pair $(x, x \oplus k\alpha)$ is an element of Γ if and only if $S^k(x,0) \in [0,1) \times \{-1,-2,\ldots\}$. By ergodicity of S, there exists k such that

$$(\lambda \times \#)\big((S^k[A \times \{0\}]) \cap (B \times \{-1,-2,\ldots\})\big) > 0,$$

thus it is possible to shift a positive portion of A to B as required. By exhaustion, there indeed exists a transport π_β such that $\langle c, \pi_\beta \rangle = 0$. □

The above example illustrates some of the subtleties of Theorem 3.1. However, it does not yet provide evidence for the necessity of allowing for the singular part $\hat{h}^s$ of the optimizer $\hat{h}$ in Theorem 3.1. We have constructed yet a more refined – and rather longish – variant of the Ambrosio–Pratelli example above, which shows that, in general, there is no way of avoiding these complications in the statement of Theorem 3.1. We refer to the accompanying paper [BLS09b, Section 3] for a presentation of this example, where it is shown that it can indeed occur that the singular part $\hat{h}^s$ in Theorem 3.1 does not vanish.

4. The Projective Limit Theorem

We again consider the general setting where c is a $[0,\infty]$-valued Borel measurable function. To avoid trivialities we shall always assume that $\Pi(\mu,\nu,c)$ is non-empty.

Theorem 3.1 only pertains to the situation of a *fixed* element $\pi_0 \in \Pi(\mu,\nu,c)$: one then optimizes the transport problem of all $\pi \in \Pi(\mu,\nu)$ with $\|\frac{d\pi}{d\pi_0}\|_{L^\infty(\pi_0)} < \infty$.

The purpose of this section is to find an optimizer h which does work simultaneously, *for all* $\pi_0 \in \Pi(\mu,\nu,c)$. We are not able to provide a result showing that a *function* h – plus possibly some singular part h^s – exists which fulfills this duty, for all $\pi_0 \in \Pi(\mu,\nu,c)$. We have to leave the question whether this is always possible as an open problem. But we can show that a projective limit $\hat{H} = (\hat{h}_\pi)_{\pi \in \Pi(\mu,\nu,c)}$ exists which does the job.

We introduce an order relation on $\Pi(\mu,\nu,c)$: we say that $\pi_1 \preceq \pi_2$ if $\pi_1 \ll \pi_2$ and $\|\frac{d\pi_1}{d\pi_2}\|_{L^\infty(\pi_2)} < \infty$. For $\pi_1 \preceq \pi_2$ there is a natural, continuous projection $P_{\pi_1,\pi_2} : L^1(\pi_2) \to L^1(\pi_1)$ associating to each $h_{\pi_2} \in L^1(\pi_2)$, which is an equivalence class modulo π_2-null functions, the equivalence class modulo π_1-null functions which contains the equivalence class h_{π_2} (and where this inclusion of equivalence classes may be strict, in general). We may define the locally convex vector space E as the projective limit

$$E = \lim_{\longleftarrow \pi \in \Pi(\mu,\nu,c)} L^1(X \times Y, \pi).$$

The elements of E are families $H = (h_\pi)_{\pi \in \Pi(\mu,\nu,c)}$ such that, for $\pi_1 \preceq \pi_2$, we have $P_{\pi_1,\pi_2}(h_{\pi_2}) = h_{\pi_1}$.
A net $(H^\alpha)_{\alpha \in I} \in E$ converges to $H \in E$ if,

$$\lim_{\alpha \in I} \| h^\alpha_\pi - h_\pi \|_{L^1(\pi)} = 0, \quad \text{for each } \pi \in \Pi(\mu,\nu,c).$$

We may also define the projective limit

$$E_S = \lim_{\longleftarrow \pi \in \Pi(\mu,\nu,c)} L^1_S(X \times Y, \pi),$$

which is a closed subspace of E.

We start with an easy result.

Proposition 4.1. *Let X and Y be polish spaces equipped with Borel probability measures μ,ν, and let $c : X \times Y \to [0,\infty]$ be Borel measurable. Assume that $\Pi(\mu,\nu,c)$ is non-empty.*

There is $\pi_0 \in \Pi(\mu,\nu,c)$ such that

$$P^{(\pi_0)} = \inf_{\pi \in \Pi(\mu,\nu,c)} P^{(\pi)}.$$

Proof. Let $(\pi_n)_{n=1}^{\infty}$ be a sequence in $\Pi(\mu,\nu,c)$ such that

$$\lim_{n\to\infty} P^{(\pi_n)} = \inf_{\pi\in\Pi(\mu,\nu,c)} P^{(\pi)}.$$

It suffices to define π_0 as

$$\pi_0 = \sum_{n=1}^{\infty} 2^{-n}\,\pi_n$$

as we then have $\pi_n \preceq \pi_0$, for each $n\in\mathbb{N}$. □

Of course, if the primal problem (2) is attained, we have $P^{(\pi_0)} = P$.

The above proposition allows us to suppose w.l.o.g. in our considerations on the projective limit E that the π appearing in the definition are all bigger than π_0:

$$E = \varprojlim_{\pi\in\Pi(\mu,\nu,c)} L^1(\pi) = \varprojlim_{\pi\in\Pi(\mu,\nu,c),\pi\succeq\pi_0} L^1(\pi).$$

Clearly, we then have that the optimal transport cost $P^{(\pi)}$ is equal to $P^{(\pi_0)}$, for all $\pi\succeq\pi_0$.

Theorem 4.2. *Let X and Y be polish spaces equipped with Borel probability measures μ,ν, and let $c: X\times Y\to[0,\infty]$ be Borel measurable. Assume that $\Pi(\mu,\nu,c)$ is non-empty. Let π_0 be as in Proposition 4.1*

There is an element $\hat H = (\hat h_\pi)_{\pi\in\Pi(\mu,\nu,c),\pi\succeq\pi_0}\in E$ such that, for each $\pi\in\Pi(\mu,\nu,c),\pi\succeq\pi_0$, the element $\hat h_\pi\in L^1_S(\pi)^{**}$ satisfies $\hat h_\pi\le c$ in the order of $L^1(\pi)^{**}$ and $\hat h_\pi$ is an optimizer of the dual problem (18)

$$\langle \hat h_\pi,\pi\rangle = D^{(\pi)}_{**} := \sup\{\langle h,\pi\rangle : h\in L^1_S(\pi)^{**},\, h\le c\}.$$

We then have that, for each $\pi\in\Pi(\mu,\nu,c),\pi\succeq\pi_0$, the decomposition $\hat h_\pi = \hat h^r_\pi + \hat h^s_\pi$ of $\hat h_\pi$ into its regular and singular parts verifies

- *$\hat h^r_\pi\in L^1_S(\pi)$ and $\hat h^r_\pi\le c$ in $L^1(\pi)$;*
- *$\hat h^s_\pi\in L^1_S(\pi)^{**}$ and $\hat h^s_\pi\le 0$ in the space of purely finitely additive measures which are absolutely continuous with respect to π.*

Moreover, for each $\pi\in\Pi(\mu,\nu,c),\pi\succeq\pi_0$, there is no duality gap in the sense that

$$D^{(\pi)}_{**} = D^{(\pi)} = P^{(\pi)} = P^{(\pi_0)} \tag{29}$$

where $D^{(\pi)} := \lim_{\varepsilon\to0}\sup\Big\{\int\varphi\,d\mu + \int\psi\,d\nu : \varphi\in L^1(\mu),\psi\in L^1(\nu),\int(\varphi\oplus\psi-c)_+\,d\pi\le\varepsilon\Big\}$ and $P^{(\pi)} := \inf\{\langle c,\pi'\rangle : \pi'\in\Pi^{(\pi)}(\mu,\nu)\}$. If in addition the primal problem (2) is attained, for instance if c is lower semicontinuous, then $D^{(\pi)}_{**} = D^{(\pi)} = P^{(\pi)} = P$.

Proof. Fix $\pi\in\Pi(\mu,\nu,c),\pi\succeq\pi_0$. We have seen in Theorem 3.1 that the set

$$K_\pi = \{h\in L^1_S(\pi)^{**} : h\le c, \langle h,\pi\rangle = \langle c,\pi\rangle\}$$

is non-empty. In addition K_π is closed and bounded in $L^1(\pi)^{**}$ and hence compact with respect to the $\sigma(L^1_S(\pi)^{**}, L^1_S(\pi)^*)$-topology.

For $\pi,\pi'\in\Pi(\mu,\nu,c)$ with $\pi\preceq\pi'$ the set

$$K_{\pi,\pi'} = P_{\pi,\pi'}(K_{\pi'})$$

is contained in K_π and still a non-empty σ^*-compact convex subset of $L^1(\pi)^{**}$. By compactness the following set is σ^*-compact and non-empty too:

$$K_{\pi,\infty} = \bigcap_{\pi'\succeq\pi} K_{\pi,\pi'}.$$

We have $K_{\pi,\infty} = P_{\pi,\pi'}(K_{\pi',\infty})$ for $\pi\preceq\pi'$. Hence by Tychonoff's theorem the projective limit

$$\varprojlim_{\pi\in\Pi(\mu,\nu,c),\pi\succeq\pi_0} K_{\pi,\infty}$$

of the compact sets $(K_{\pi,\infty})_{\pi\succeq\pi_0}$ is non-empty, which is precisely the main assertion of the present theorem.
Finally, (29) is a restatement of (20) and when the primal problem (2) is attained, the last series of equalities follows from $P^{(\pi_0)} = P$. □

Clearly $P^{\mathrm{rel}} \leq P \leq P^{(\pi_0)}$, hence with Theorem 1.1 and (29) one sees that

$$D = P^{\mathrm{rel}} \leq P \leq P^{(\pi_0)} = P^{(\pi)} = D^{(\pi)}_{**} = D^{(\pi)}$$

for every $\pi \in \Pi(\mu, \nu, c)$ such that $\pi \succeq \pi_0$.

References

[Aar97] J. Aaronson. *An introduction to infinite ergodic theory*, volume 50 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1997.

[AK82] J. Aaronson and M. Keane. The visits to zero of some deterministic random walks. *Proc. London Math. Soc. (3)*, 44(3):535–553, 1982.

[AP03] L. Ambrosio and A. Pratelli. *Existence and stability results in the L^1-theory of optimal transportation. CIME Course*, volume 1813 of *Lecture Notes in Mathematics*, pages 123–160. Springer Verlag, 2003.

[BGMS09] M. Beiglböck, M. Goldstern, G. Maresh, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009.

[BLS09a] M. Beiglböck, C. Léonard, and W. Schachermayer. A general duality theorem for the Monge-Kantorovich transport problem. *submitted*, 2009.

[BLS09b] M. Beiglböck, C. Léonard, and W. Schachermayer. On the duality of the Monge-Kantorovich transport problem. *submitted*, 2009.

[BS09] M. Beiglböck and W. Schachermayer. Duality for Borel measurable cost functions. *Trans. Amer. Math. Soc.*, to appear, 2009.

[CK76] J.-P. Conze and M. Keane. Ergodicité d'un flot cylindrique. In *Séminaire de Probabilités, I (Univ. Rennes, Rennes, 1976), Exp. No. 5*, page 7. Dépt. Math. Informat., Univ. Rennes, Rennes, 1976.

[CM06] L. Cafarelli and R.J. McCann. Free boundaries in optimal transport and Monge-Ampere obstacle problems. *Ann. of Math. (2)*, to appear, 2006.

[dA82] A. de Acosta. Invariance principles in probability for triangular arrays of B-valued random vectors and some applications. *Ann. Probab.*, 10(2):346–373, 1982.

[Dec08] L. Decreusefond. Wasserstein distance on configuration space. *Potential Anal.*, 28(3):283–300, 2008.

[DJS08] L. Decreusefond, A. Joulin, and N. Savy. Rubinstein distances on configuration spaces. Preprint, 2008.

[Dud76] R. M. Dudley. *Probabilities and metrics*. Matematisk Institut, Aarhus Universitet, Aarhus, 1976. Convergence of laws on metric spaces, with a view to statistical testing, Lecture Notes Series, No. 45.

[Dud02] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[Fer81] X. Fernique. Sur le théorème de Kantorovich-Rubinstein dans les espaces polonais. In *Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)*, volume 850 of *Lecture Notes in Math.*, pages 6–10. Springer, Berlin, 1981.

[Fig09] A. Figalli. The optimal partial transport problem. *Arch. Rational Mech. Anal.*, to appear, 2009.

[FÜ02] D. Feyel and A.S. Üstünel. Measure transport on Wiener space and the Girsanov theorem. *C. R. Math. Acad. Sci. Paris*, 334(11):1025–1028, 2002.

[FÜ04a] D. Feyel and A.S. Üstünel. Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004.

[FÜ04b] D. Feyel and A.S. Üstünel. Monge-Kantorovitch measure transportation, Monge-Ampère equation and the Itô calculus. In *Stochastic analysis and related topics in Kyoto*, volume 41 of *Adv. Stud. Pure Math. Math. Soc. Japan*, pages 49–74, Tokyo, 2004.

[FÜ06] D. Feyel and A. S. Üstünel. Solution of the Monge-Ampère equation on Wiener space for general log-concave measures. *J. Funct. Anal.*, 232(1):29–55, 2006.

[GM96] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[GR81] N. Gaffke and L. Rüschendorf. On a class of extremal problems in statistics. *Math. Operationsforsch. Statist. Ser. Optim.*, 12(1):123–135, 1981.

[Kan42] L.V. Kantorovich. On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS*, 37:199–201, 1942.

[Kel84] H. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4):399–432, 1984.

[KR58] L. V. Kantorovič and G. Š. Rubinšteĭn. On a space of completely additive functions. *Vestnik Leningrad. Univ.*, 13(7):52–59, 1958.

[Léo09] C. Léonard. A saddle-point approach to the Monge-Kantorovich transport problem. 2009. To appear in *ESAIM-COCV*.

[Mik06] T. Mikami. A simple proof of duality theorem for Monge-Kantorovich problem. *Kodai Math. J.*, 29(1):1–4, 2006.

[MT06] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic Process. Appl.*, 116(12):1815–1835, 2006.

[RR95] D. Ramachandran and L. Rüschendorf. A general duality theorem for marginal problems. *Probab. Theory Related Fields*, 101(3):311–319, 1995.

[RR96] D. Ramachandran and L. Rüschendorf. Duality and perfect probability spaces. *Proc. Amer. Math. Soc.*, 124(7):2223–2228, 1996.

[RS80] M. Reed and B. Simon. *Methods of Modern Mathematical Physics, I: Functional Analysis*. Academic Press, 1980.

[Rüs96] L. Rüschendorf. On c-optimal random variables. *Statist. Probab. Lett.*, 27(3):267–270, 1996.

[Sch78] K. Schmidt. A cylinder flow arising from irregularity of distribution. *Compositio Math.*, 36(3):225–232, 1978.

[ST09] W. Schachermayer and J. Teichman. Characterization of optimal transport plans for the Monge-Kantorovich problem. *Proc. Amer. Math. Soc.*, 137:519–529, 2009.

[Szu82] A. Szulga. On minimal metrics in the space of random variables. *Teor. Veroyatnost. i Primenen.*, 27(2):401–405, 1982.

[Vil03] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.

[Vil09] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

University of Vienna. Faculty of Mathematics. Nordbergstrasse 15. 1090 Vienna, Austria
E-mail address: mathias.beiglboeck@univie.ac.at

Modal-X, Université Paris Ouest. Bât. G, 200 av. de la République. 92001 Nanterre, France
E-mail address: christian.leonard@u-paris10.fr

University of Vienna. Faculty of Mathematics. Nordbergstrasse 15. 1090 Vienna, Austria
E-mail address: walter.schachermayer@univie.ac.at