CHARACTERIZATION OF OPTIMAL TRANSPORT PLANS FOR THE MONGE-KANTOROVICH-PROBLEM

WALTER SCHACHERMAYER AND JOSEF TEICHMANN

Abstract. We prove that c-cyclically monotone transport plans π optimize the Monge-Kantorovich transportation problem under an additional measurability condition. This measurability condition is always satisfied for finitely valued, lower semi-continuous cost functions. In particular, this yields a positive answer to Problem 2.25 in C. Villani's book. We emphasize that we do not need any regularity conditions as were imposed in the previous literature.

We consider the *Monge-Kantorovich optimal transport problem* (μ, ν, c) for Borel measures μ, ν on polish spaces X, Y and a lower semi-continuous cost function $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ (see C. Villani's beautiful book [**?**] for all necessary details): we define the set of *transport plans* $\Pi(\mu, \nu)$ as the set of probability measures π with marginal μ on X, and marginal ν on Y, respectively. Furthermore, we define $\Phi(\mu, \nu)$ as the set of pairs (ϕ, ψ) of Borel functions $\phi : X \to \mathbb{R} \cup \{-\infty\}$ and $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ with $\phi \in \mathcal{L}^1(\mu)$ and $\psi \in \mathcal{L}^1(\nu)$, such that

$$\phi(x) + \psi(y) \leq c(x, y)$$

for all $(x, y) \in X \times Y$. The Monge-Kantorovich problem is to minimize the cost

$$I(\pi) = \int_{X \times Y} c(x, y)\pi(dx, dy)$$

for transport plans $\pi \in \Pi(\mu, \nu)$. Dually to the Monge-Kantorovich problem, we maximize

$$J(\phi, \psi) = \int_X \phi(x)\mu(dx) + \int_Y \psi(y)\nu(dy)$$

over $(\phi, \psi) \in \Phi(\mu, \nu)$.

It is well-known that under the previous conditions, the equality

$$\inf_{\Pi(\mu,\nu)} I(\pi) = \sup_{\Phi(\mu,\nu)} J(\phi, \psi)$$

holds true, and the infimum is in fact a minimum, if $\inf_{\Pi(\mu,\nu)} I(\pi) < \infty$ (see [**?**], Th.1.3). We know furthermore that optimizers $\widehat{\pi}$ with $I(\widehat{\pi}) < \infty$ are concentrated on c-monotone (see the following definition) Borel sets $\Gamma \subset X \times Y$. These assertions are proved for instance in [**?**, ch.2] or in [**?**, Th. 3.2].

2000 *Mathematics Subject Classification.* 49J45,28A35.

Financial support from the Austrian Science Fund (FWF) under grant P 15889, from the Vienna Science Foundation (WWTF) under grant MA13, from the European Union under grant HPRN-CT-2002-00281 is gratefully acknowledged. Furthermore this work was financially supported by the Christian Doppler Research Association (CDG). The authors gratefully acknowledge a fruitful collaboration and continued support by Bank Austria through CDG.

Definition 1. *A set $\Gamma \subset X \times Y$ is called* c-monotone *if for all $n \geq 1$, for all points $(x_i, y_i) \in \Gamma$, $i = 1, \ldots, n$, and all permutations $\sigma \in \mathfrak{S}_n$*

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{\sigma(i)})$$

holds true. A transport plan $\pi \in \Pi(\mu, \nu)$ is called c -monotone *if there is a c-monotone Borel set $\Gamma \subset X \times Y$, where π is concentrated on, i.e. $\pi(\Gamma) = 1$.*

In [**?**] it was posed as an open problem whether c-monotone transport plans $\pi \in \Pi(\mu, \nu)$ are necessarily optimizers of the Monge-Kantorovich problem (see Problem 2.25 in [**?**] and the references therein). Due to [**?**] it is known that this conjecture holds true under a certain moment conditions on the probability measures μ, ν with respect to the cost function c.

Crucial for these considerations is Rüschendorf's Theorem (see [**?**] and [**?**, ch.2]), which relates c-concave functions and c-monotone sets Γ. Given a function $\phi : X \to \mathbb{R} \cup \{-\infty\}$ and a lower semi-continuous cost function $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$. Then ψ is called c-concave on Y, if there is a function $\phi : X \to \mathbb{R} \cup \{-\infty\}$, such that

$$\psi(y) = \inf_{x \in X} (c(x, y) - \phi(x))$$

for $y \in Y$. Accordingly we define c-concave functions ϕ on X.

Let $\phi : X \to \mathbb{R} \cup \{-\infty\}$ be a function not identical to $-\infty$, then

$$\phi^c(y) = \inf_{x \in X} (c(x, y) - \phi(x)) \tag{1}$$

is called the c-transform of ϕ, which is obviously a c-concave function on Y. We have that ϕ is c-concave on X if and only if $\phi = \phi^{cc}$. The c-superdifferential $\partial_c \phi \subset X \times Y$ of a c-concave function ϕ is defined as the set of all pairs $(x, y) \in X \times Y$ such that for all $z \in X$,

$$\phi(z) \leq \phi(x) + (c(z, y) - c(x, y)).$$

Rüschendorf's Theorem finally states that a set $\Gamma \subset X \times Y$ is c-monotone if and only if there is c-concave function ϕ such that $\Gamma \subset \partial_c \phi$ (see [**?**, ch.2] and [**?**] for further details and references).

We cite the Theorem from [**?**], which inspired our subsequent considerations.

Theorem 1. *Let X, Y be polish spaces and μ, ν be Borel probability measures thereon. Assume that c is a finitely-valued, lower semi-continuous cost function and that π is c-monotone. Furthermore we assume that*

$$\nu\left(\{y, \int_X c(x, y) \mu(dx) < \infty\}\right) > 0,$$
$$\mu\left(\{x, \int_Y c(x, y) \nu(dy) < \infty\}\right) > 0.$$

Then π is an optimizer and there exists a dual optimizer $(\phi, \psi) \in \mathcal{L}^1(\mu) \times \mathcal{L}^1(\nu)$, where ϕ is a c-concave Borel function, ψ is a Borel function, $\psi = \phi^c$ almost surely with respect to ν and

$$\phi(x) + \psi(y) \leq c(x, y)$$

for $(x, y) \in X \times Y$, where equality holds π-almost surely.

We shall prove the same assertion without assuming that c is finitely valued and without any assumptions on the existence of moments of c. Here we do assume a measurability condition related to c-monotone transport plans, namely the existence of a Borel-measurable "hull" for a sequence of dual optimizers. We shall prove (Proposition **??** below) that this measurability condition always holds true for *finitely valued*, lower semi-continuous cost functions c thus covering in particular the setting of problem 2.25 in [**?**].

Definition 2. *Given polish spaces X, Y, a lower semi-continuous cost function $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$, we call a Borel set $\Gamma \subset X \times Y$* strongly c-monotone, *if there exist Borel functions $\phi : X \to \mathbb{R} \cup \{-\infty\}$ and $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ such that*

$$\phi(x) + \psi(y) \leq c(x, y),$$

for all $(x, y) \in X \times Y$ and equality holds if $(x, y) \in \Gamma$. We call a transport plan π strongly c-monotone *if there exists a strongly c-monotone Borel set $\Gamma \subset X \times Y$, on which π is concentrated, i.e. $\pi(\Gamma) = 1$.*

Remark 1. *A seemingly innocent, but important point is the Borel measurability of ϕ and ψ.*

Remark 2. *Notice that we do neither assume the Borel-measurable functions ϕ, ψ to be c-concave nor to be conjugate.*

Remark 3. *A strongly c-monotone set $\Gamma \subset X \times Y$ is c-monotone, since*

$$\begin{aligned}
\sum_{i=1}^{n} c(x_i, y_i) &= \sum_{i=1}^{n} (\phi(x_i) + \psi(y_i)) \\
&= \sum_{i=1}^{n} (\phi(x_i) + \psi(y_{\sigma(i)})) \\
&\leq \sum_{i=1}^{n} c(x_i, y_{\sigma(i)}),
\end{aligned}$$

for all $(x_i, y_i)_{i=1,\dots,n} \subset \Gamma$ and permutations $\sigma \in \mathfrak{S}_n$ for $n \geq 1$.

The following Proposition, which is based on the proof of Theorem 3.2 in [**?**], shows that **all c-monotone transport plans are strongly c-monotone** provided that c is finitely valued.

Proposition 1. *Given polish spaces X, Y, a finitely valued, lower semi-continuous cost function $c : X \times Y \to \mathbb{R}_{\geq 0}$ and a c-monotone transport plan π. Then there exist Borel-measurable functions ϕ, ψ on X, Y, respectively, taking values in $\mathbb{R} \cup \{-\infty\}$, such that ϕ is c-concave, the c-superdifferential $\partial_c \phi$ contains a set where π is concentrated, ψ coincides ν-almost surely with ϕ^c, the inequality*

$$\phi(x) + \psi(y) \leq c(x, y)$$

holds true for all $(x, y) \in X \times Y$, and equality holds true π-almost surely. In particular π is strongly c-monotone.

Proof. In order to apply a construction of [**?**] we have to show that there is a c-monotone Borel set $\Gamma = \cup_{k \geq 1} \Gamma_k$ with Γ_k compact, $c|_{\Gamma_k}$ continuous for $k \geq 1$ and $\pi(\Gamma) = 1$. Take any c-monotone Borel set Γ' with $\pi(\Gamma') = 1$, which exists by assumption. By Egorov's Theorem and the inner regularity on polish spaces of the transport plan π (see [?] for both results), we know that there are compact subsets $L_k \subset X \times Y$, such that $\pi(X \times Y \setminus L_k) < \frac{1}{k}$ for $k \geq 1$ and $c|_{L_k}$ is continuous for $k \geq 1$. Consequently we can define compact subsets $\Gamma_k := \overline{\Gamma' \cap L_k} \subset L_k \subset X \times Y$ for $k \geq 1$. In particular $c|_{\Gamma_k}$ is continuous for $k \geq 1$ and the Borel set $\Gamma := \cup_{k \geq 1} \Gamma_k$ satisfies $\Gamma \supset \Gamma'$ by construction. Since $c|_{\Gamma_k}$ is continuous and $\Gamma' \cap L_k$ is c-monotone by assumption, the c-monotonicity extends to the closure $\Gamma_k := \overline{\Gamma' \cap L_k}$ for $k \geq 1$. A union of c-monotone sets is c-monotone, so Γ is c-monotone. Whence we have the claim and can apply [?], Th. 3.2 (step 1 of the proof) in order to construct – in the spirit of Rüschendorf's Theorem – a Borel-measurable function $\phi : X \to \mathbb{R} \cup \{-\infty\}$ with $\Gamma \subset \partial_c \phi$.

We then define ϕ^c via (??) and obtain by applying again [?], Th. 3.2 (step 2 of the proof), that ϕ^c is ν-measurable. Setting the values of ϕ^c to $-\infty$ on a ν-negligible set we can define a Borel measurable function ψ on Y such that

$$\phi(x) + \psi(y) \leq c(x, y)$$

holds true for all $(x, y) \in X \times Y$. If $(x, y) \in \Gamma \subset \partial_c \phi$, we know from the definition of the superdifferential of ϕ that

$$\phi(z) \leq \phi(x) + (c(z, y) - c(x, y))$$

for all $z \in X$, hence

$$c(x, y) - \phi(x) \leq c(z, y) - \phi(z)$$

for all z, so

$$c(x, y) - \phi(x) \leq \psi(y),$$

for $(x, y) \in \Gamma$. Hence equality in

$$\phi(x) + \psi(y) \leq c(x, y)$$

holds true π-almost surely, since $\pi(\Gamma) = 1$. □

Next we show that **optimal transport plans are strongly c-monotone** provided that c is $\mu \otimes \nu$-almost surely finitely valued. This result suggests to work with strongly c-monotone transport plans instead of c-monotone ones. For the proof we need the following Lemma, which generalizes results in [?] and [?] (see in particular the references therein).

Lemma 1. *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ be probability spaces and let $\phi_n : \Omega_1 \to \mathbb{R}$ and $\psi_n : \Omega_2 \to \mathbb{R}$ be measurable functions for $n \geq 1$. Assume that*

$$\xi_n(\omega_1, \omega_2) = \phi_n(\omega_1) + \psi_n(\omega_2)$$

converges in $\mu_1 \otimes \mu_2$-probability to $\xi : \Omega_1 \times \Omega_2 \to \mathbb{R} \cup \{-\infty\}$. Then there exist real numbers $(r_n)_{n \geq 1}$ and measurable functions $\phi : \Omega_1 \to \mathbb{R} \cup \{-\infty\}$, $\psi : \Omega_2 \to \mathbb{R} \cup \{-\infty\}$, such that – along a fixed subsequence – $(\phi_n + r_n)_{n \geq 0}$ converges in probability to ϕ and $(\psi_n - r_n)_{n \geq 0}$ to ψ. Furthermore,

$$\xi(\omega_1, \omega_2) = \phi(\omega_1) + \psi(\omega_2)$$

$\mu_1 \otimes \mu_2$-almost surely.

Proof. We choose a complete, bounded metric d on $\mathbb{R} \cup \{-\infty\}$. Then $\xi_n \to \xi$ in probability as $n \to \infty$ is equivalent to $E(d(\xi_n, \xi)) \to 0$ as $n \to \infty$. By Fubini's theorem the real-valued function

$$\omega_2 \mapsto \int_{\Omega_1} d(\xi_n(\omega_1, \omega_2), \xi(\omega_1, \omega_2)) \mu_1(d\omega_1)$$

is μ_2-almost surely well-defined and converges in μ_2-probability to 0 as $n \to \infty$ by assumption.

We shall now distinguish two cases, either ξ is finitely valued with positive probability, or $\mu_1 \otimes \mu_2$-almost surely $\xi = -\infty$. First we assume that there is $A \subset \Omega_1 \times \Omega_2$ such that $\xi|_A > -\infty$ and $(\mu_1 \otimes \mu_2)(A) > 0$. Then there is $\widetilde{\omega_2} \in \Omega_2$ such that

$$\xi_n(\omega_1, \widetilde{\omega_2}) = \phi_n(\omega_1) + \psi_n(\widetilde{\omega_2}) \to \phi(\omega_1) := \xi(\omega_1, \widetilde{\omega_2})$$

converges in μ_1-probability on Ω_1 as $n \to \infty$, and the function $\phi : \Omega_1 \to \mathbb{R} \cup \{-\infty\}$ is measurable with $\phi > -\infty$ on a set of positive μ_1-probability. Here we possibly have to pass to a subsequence to guarantee that there is $\widetilde{\omega_2}$ where $\xi_n(., \widetilde{\omega_2})$ converges in probability. We assume from now on that we have already passed to such a subsequence. We then define $r_n := \psi_n(\widetilde{\omega_2})$, for $n \geq 1$. Furthermore we obtain that the sequence

$$\psi_n - r_n = \xi_n - \phi_n - r_n$$

converges in μ_2-probability on Ω_2 as $n \to \infty$ to a function $\psi : \Omega_2 \to \mathbb{R} \cup \{-\infty\}$. Indeed, the right hand side does not depend on the first variable ω_1 by construction. We choose a measurable subset $A_1 \subset \Omega_1$ of positive μ_1-probability with $\phi|_{A_1} > -\infty$. The sequence $(\xi_n)_{n\geq 1}$ converges in probability to ξ and the sequence $(\phi_n + r_n)_{n \geq 1}$ converges in probability to ϕ on A_1, so the difference converges in probability on $A_1 \times \Omega_2$ to a limit ψ, which does not depend on the first variable. Hence in particular

$$\xi(\omega_1, \omega_2) = \phi(\omega_1) + \psi(\omega_2)$$

$\mu_1 \otimes \mu_2$-almost sure.

Now we assume that ξ is almost surely equal to $-\infty$. We know that, for each fixed $0 < \alpha < 1$, the α-quantiles

$$\begin{aligned} q_n^1(\alpha) &:= \inf\{x;\ \mu_1(\phi_n \geq x) \leq \alpha\}, \\ q_n^2(\alpha) &:= \inf\{y;\ \mu_2(\psi_n \geq y) \leq \alpha\} \end{aligned}$$

have the property

$$\lim_{n\to\infty} (q_n^1(\alpha) + q_n^2(\alpha)) = -\infty$$

by assumption. Indeed, otherwise there is a subsequence $(n_k)_{k\geq 1}$ and a lower bound $M > -\infty$ such that $q_{n_k}^1(\alpha) + q_{n_k}^2(\alpha) \geq M$ for all $k \geq 1$. Hence $\phi_{n_k} + \psi_{n_k} \geq M$ with $\mu_1 \otimes \mu_2$-probability greater than α^2 for $k \geq 1$. However, this contradicts the convergence in probability to $-\infty$. By choosing a sequence $(\alpha_n)_{n\geq 1}$ with $\alpha_n \downarrow 0$ slowly enough as $n \to \infty$, we can maintain this property, i.e.

$$\lim_{n\to\infty} (q_n^1(\alpha_n) + q_n^2(\alpha_n)) = -\infty.$$

We define now

$$\begin{aligned} z_n &:= \frac{1}{2}(q_n^1(\alpha_n) + q_n^2(\alpha_n)) \\ r_n &:= \frac{1}{2}(q_n^2(\alpha_n) - q_n^1(\alpha_n)) \\ &= q_n^2(\alpha_n) - z_n = -(q_n^1(\alpha_n) - z_n). \end{aligned}$$

We have $\lim_{n\to\infty} z_n = -\infty$ and hence $\phi_n + r_n \to -\infty$ and $\psi_n - r_n \to -\infty$ converges in the respective probabilities as $n \to \infty$ by construction. □

In the proof of the subsequent proposition we shall apply a Komlos-type result, which we cite here from [**?**, Lemma A1.1] for the sake of completeness.

Lemma 2. *Let $(f_n)_{n\geq 1}$ be a sequence of $\mathbb{R}_{\geq 0}\cup\{+\infty\}$-valued random variables on a probability space $(\Omega,\mathcal{F},P)$. Then there is a sequence $g_n\in\operatorname{conv}(f_n,f_{n+1},\dots)$, the convex hull of $f_n,f_{n+1},\dots$, for $n\geq 1$, which converges almost surely to a random variable g taking values in $\mathbb{R}_{\geq 0}\cup\{+\infty\}$.*

The interesting feature of this result is that we do not need any integrability assumption on the sequence $(f_n)_{n\geq 1}$ to obtain an almost sure convergence of convex combinations. It suffices to assume non-negativity of the functions f_n. This situation is related to the setting of [**?**], where it is shown that a Bipolar-Theorem can be formulated for the non-locally-convex space L^0, provided that one restricts to positive functions.

Proposition 2. *Let $c: X\times Y\to\mathbb{R}_{\geq 0}\cup\{+\infty\}$ be a lower semi-continuous cost function on polish spaces X,Y. Given Borel probability measures μ on X, ν on Y and an optimizer $\widehat{\pi}\in\Pi(\mu,\nu)$ with $I(\widehat{\pi})<\infty$. We assume that c is $\mu\otimes\nu$-almost surely finitely valued. Then there exist Borel functions ϕ,ψ on X,Y, respectively, taking values in $\mathbb{R}\cup\{-\infty\}$, such that*

$$\phi(x)+\psi(y)\leq c(x,y)$$

for $x\in X$ and $y\in Y$, and $\phi(x)+\psi(y)=c(x,y)$ almost surely with respect to $\widehat{\pi}$. In particular the transport plan $\widehat{\pi}$ is strongly c-monotone.

Proof. By [**?**], Th. 3.1, we know that there is a maximizing sequence (ϕ_n,ψ_n) of bounded Borel functions on X,Y, respectively, such that

$$\lim_{n\to\infty}E_{\widehat{\pi}}(\xi_n)=\lim_{n\to\infty}E_{\mu\otimes\nu}(\xi_n)=E_{\widehat{\pi}}(c)=I(\widehat{\pi}),$$

where $\xi_n(x,y)=\phi_n(x)+\psi_n(y)\leq c(x,y)$ for all $(x,y)\in X\times Y$ and the convergence of $(E_{\mu\otimes\nu}(\xi_n))_{n\geq 0}$ is monotonely increasing. By passing to convex combinations (apply the above Lemma **??** with $f_n=c-\xi_n$ for $n\geq 0$) we may assume that ξ_n converges $\mu\otimes\nu$-almost surely to a Borel function ξ taking values in $\mathbb{R}\cup\{-\infty\}$. By Lemma **??** we know that

$$\xi(x,y)=\phi(x)+\psi(y),$$

with functions ϕ and ψ, which are μ- and ν-almost surely defined, respectively. Furthermore – by passing to a subsequence if necessary – there is a sequence of real numbers $(r_n)_{n\geq 1}$ such that $\phi_n+r_n\to\phi$ and $\psi_n-r_n\to\psi$ as $n\to\infty$ almost surely with respect to μ and ν, respectively. Choosing appropriate nullsets in X and Y, we can redefine ϕ,ϕ_n and ψ,ψ_n on these nullsets by $-\infty$, such that $\phi_n+r_n\to\phi$ and $\psi_n-r_n\to\psi$ *surely*, without violating the now *sure inequality*

$$\xi(x,y)=\phi(x)+\psi(y)\leq c(x,y).$$

Obviously $E_{\widehat{\pi}}(|c-\xi_n|)=E_{\widehat{\pi}}(c-\xi_n)\to 0$, as $n\to\infty$, by assumption, hence $\phi(x)+\psi(y)=c(x,y)$ almost surely with respect to $\widehat{\pi}$. □

Our main Theorem below states that strongly c-monotone transport plans are always optimal. For sake of clarity we shall formulate an elementary lemma on monotone convergence of truncations, whose proof is obvious.

Lemma 3. *Let a,b be real numbers. We define*

$$\begin{aligned} a_n &:= (-n\vee a)\wedge n,\\ b_n &:= (-n\vee b)\wedge n,\\ \xi_n &:= a_n+b_n. \end{aligned}$$

Then we have $\xi_0 = 0$ and $(\xi_n)_{n\geq 0}$ converges monotonically to $a+b$, i.e. if $a+b \geq 0$ the sequence $(\xi_n)_{n\geq 0}$ increases to $a+b$ and if $a+b \leq 0$ the sequence decreases to $a+b$.

Theorem 2. *Let $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ be a lower semi-continuous cost function on polish spaces X, Y. Given Borel probability measures μ on X, ν on Y and a strongly c-monotone transport plan $\pi \in \Pi(\mu, \nu)$, then π is an optimizer of the Monge-Kantorovich problem.*

Proof. We assume that there is a transport plan π_0 with finite cost $I(\pi_0) < \infty$. We aim to show that $I(\pi) \leq I(\pi_0)$. By the assumption of strong c-monotonicity of π there exist Borel-measurable functions ϕ, ψ on X, Y, respectively, taking values in $\mathbb{R} \cup \{-\infty\}$, such that for all $(x, y) \in X \times Y$

$$\phi(x) + \psi(y) \leq c(x, y),$$

where equality holds π-almost surely. We define

$$\begin{aligned} \phi_n(x) &:= (-n \vee \phi(x)) \wedge n, \\ \psi_n(y) &:= (-n \vee \psi(y)) \wedge n, \\ \xi_n(x, y) &= \phi_n(x) + \psi_n(y), \\ \xi(x, y) &= \phi(x) + \psi(y), \end{aligned}$$

for $(x, y) \in X \times Y$ and $n \geq 0$. Due to the previous Lemma observe that $\xi_n \uparrow \xi$ on $\{\xi \geq 0\}$ and $\xi_n \downarrow \xi$ on $\{\xi \leq 0\}$, as $n \to \infty$.

Additionally, $E_\pi(\xi) = E_\pi(c)$ exists, taking possibly the value $+\infty$, since equality $\xi = c$ holds π-almost surely and $c \geq 0$. The integral of ξ with respect to π_0 exists, too, as

$$E_{\pi_0}(\xi) \leq E_{\pi_0}(c) < \infty.$$

Note that $E_{\pi_0}(\xi)$ possibly equals $-\infty$. By the assumption on equal marginals of π and π_0 we obtain

$$\begin{aligned} E_\pi(\xi_n) &= E_\pi(\phi_n) + E_\pi(\psi_n) \\ &= E_{\pi_0}(\phi_n) + E_{\pi_0}(\psi_n) \\ &= E_{\pi_0}(\xi_n), \end{aligned}$$

for $n \geq 0$, hence

$$E_\pi(\xi_n 1_{\{\xi \geq 0\}} + \xi_n 1_{\{\xi \leq 0\}}) = E_{\pi_0}(\xi_n 1_{\{\xi \geq 0\}} + \xi_n 1_{\{\xi \leq 0\}}).$$

By our previous considerations we can pass to the limits and obtain $E_\pi(\xi) = E_{\pi_0}(\xi)$. Indeed the limits are monotone on $\{\xi \geq 0\}$ and $\{\xi \leq 0\}$, the expectation with respect to π of ξ_- is finite, namely $E_\pi(\xi_-) = 0$, and the expectation with respect to π_0 of ξ_+ is finite, $E_{\pi_0}(\xi_+) < \infty$. Hence the limits of $E_\pi(\xi_n) = E_{\pi_0}(\xi_n)$ exist as $n \to \infty$ and are equal. Consequently $I(\pi) = E_\pi(\xi) = E_{\pi_0}(\xi) \leq I(\pi_0)$. □

The following Corollary answers Problem 2.25 of [**?**] pertaining to the quadratic cost function affirmatively. In fact, all finite, lower semi-continuous cost functions are covered and no additional assumptions on integrability or measurability are necessary.

Corollary 1. *Let $c : X \times Y \to \mathbb{R}_{\geq 0}$ be a finitely valued, lower semi-continuous cost function on polish spaces X, Y. Given Borel probability measures μ on X, ν*

on Y and a c-monotone transport plan $\pi \in \Pi(\mu,\nu)$, then π is an optimizer of the Monge-Kantorovich problem.

Proof. Due to Proposition **??**, for a finitely valued, lower semi-continuous cost function c, a c-monotone transport plan π is strongly c-monotone. Hence Theorem **??** applies. $\square$

The subsequent example, which is taken from [**?**], Example 3.5, shows that there are c-monotone transport plans, which are not strongly c-monotone.

Example 1. *This example is just a re-interpretation of the illuminating Example 3.5. in [**?**]. Take $X = Y = T^1 = \{\exp(2\pi i r),\ 0 \le r < 1\} \subset \mathbb{C}$ the 1-dimensional Torus with uniform distribution $\mu = \nu$ and choose $\alpha \in T^1$, so that $\{\alpha^n, n \ge 0\}$ is dense in T^1, i.e. $\alpha = \exp(2\pi i r)$ with r an irrational number: we define a lower semicontinuous cost function $c : X \times Y \to \mathbb{R}_{\ge 0} \cup \{+\infty\}$ via*

$$c(x,y) = \begin{cases} 1 \text{ if } x = y \\ 2 \text{ if } x = y\alpha \end{cases}$$

and $+\infty$ elsewhere. Then there are two disjoint c-monotone subsets of $X \times Y$, namely

$$\Gamma_1 = \{(x,x), x \in T^1\},$$
$$\Gamma_2 = \{(x,x\alpha), x \in T^1\}.$$

For the first set strong c-monotonicity is clear with $\phi_1(\exp(2\pi i r)) = r$ for $r \in [0,1[$, and $\psi_1(\exp(ir)) = \inf_x(c(x,\exp(ir)) - \phi_1(x)) = 1 - r$ for $r \in [0,1[$. Then we obtain the result, that $\phi_1(x) + \psi_1(y) \le c(x,y)$ with equality if $x = y$.

*The second set Γ_2 is c-monotone (see [**?**], Example 3.5.), but not strongly c-monotone (this follows from non-optimality, but we shall also show it directly).*

Take $n \ge 1$ and $(x_i, x_i\alpha)$ for $i = 1,\dots,n$. Then for all $\sigma \in \mathfrak{S}_n$,

$$\sum_{i=1}^n c(x_i, x_i\alpha) \le \sum_{i=1}^n c(x_i, x_{\sigma(i)}\alpha).$$

Indeed, otherwise there is $\sigma \in \mathfrak{S}_n$ and at least one $x_i \in T^1$ such that $c(x_i, x_{\sigma(i)}\alpha) = 1$ and therefore

$$x_i = x_{\sigma(i)}\alpha$$

Iteration of this equation yields $\alpha^m = 1$ for some $m \ge 1$, a contradiction showing the c-monotonicity of Γ_2.

We shall now show that Γ_2 fails to be strongly c-monotone. In fact, we shall show the stronger assertion that, for every Borel measure $B \subset T^1$ with $\mu(B) > 0$ the set

$$\Gamma_3 = \{(x,x\alpha), x \in B\}$$

fails to be strongly c-monotone. Indeed, we suppose that there are Borel-measurable ϕ_2 and ψ_2 taking value in $\mathbb{R} \cup \{-\infty\}$, such that

$$\phi_2(x) + \psi_2(y) \le c(x,y) \tag{2}$$

*for all $(x,y) \in T^1 \times T^1$, with equality holding for $x\alpha = y$ and $x \in B$. Using the special form of c, we see that (**??**) implies*

$$\phi_2(z) + \psi_2(z) \le 1, \tag{3}$$

$$\phi_2(x) + \psi_2(x\alpha) = 2, \tag{4}$$

for $x \in B$ and $z \in T^1$. Combining (??) and (??) we obtain

$$\phi_2(x) \geq \phi_2(x\alpha) + 1 \tag{5}$$

*for $x \in B$ or for $x\alpha \in B$. Defining $B_k := B \cap \phi_2^{-1}([k, k+1[)$ for $k \in \mathbb{Z}$, we obtain a partition $(B_k)_{k \in \mathbb{Z}}$ of B. We consider now an equivalence relation on T^1, namely $x \sim_\alpha y$ if $x\alpha^m = y$ for some $m \in \mathbb{Z}$. Fix $k \in \mathbb{Z}$; due to **??** there can be at most one element of each equivalence class $[x] \subset T^1$ in $B \cap \phi_2^{-1}([k, k+1[)$. On the other hand every Borel set $A \subset T^1$, which contains at most one element of each equivalence class $[x]$, has measure $\mu(A) = 0$ (since then $\cup_{m \in \mathbb{Z}}(A\alpha^m) \subset T^1$, $A\alpha^m$ are pairwise disjoint and $\mu(A\alpha^m) = \mu(A)$ for $m \in \mathbb{Z}$ by translation invariance). Consequently we obtain a contradiction to the assumption that the sets B_k are Borel sets.*

We can finally formulate an approximation result for the optimal transport plan of Monge-Kantorovich problems. The result can also be interpreted as discretization result for numerical purposes. In particular, we obtain – without any moment conditions – that the adherence points of candidate approximations are optimizers, provided the cost function c is continuous.

Definition 3. *Let $c : X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ be a lower semi-continuous cost function on polish spaces X, Y. Given Borel probability measures μ on X and ν on Y. Assume that there are sequences $(\mu_n)_{n \geq 1}$ and $(\nu_n)_{n \geq 1}$ of Borel probability measures converging weakly to μ and ν on the respective spaces X and Y. We denote by π_n a solution of the Monge-Kantorovich problem associated to (μ_n, ν_n, c) for $n \geq 1$ and we assume that*

$$\int_{X \times Y} c(x, y)\pi_n(dx, dy) < \infty,$$

for each $n \geq 1$. We call the sequence $(\pi_n)_{n \geq 1}$ an approximating sequence of optimizers.

Theorem 3. *Let $c : X \times Y \to \mathbb{R}_{\geq 0}$ be a finitely valued, continuous cost function on polish spaces X, Y. Let $(\pi_n)_{n \geq 1}$ be an approximating sequence of optimizers associated to weakly converging sequences $\mu_n \to \mu$ and $\nu_n \to \nu$ as $n \to \infty$. Then there is a subsequence $(\pi_{n_k})_{k \geq 0}$ converging weakly to a transport plan π on $X \times Y$, which optimizes the Monge-Kantorovich problem (μ, ν, c). Any other converging subsequence of $(\pi_n)_{n \geq 0}$ also converges to an optimizer of the Monge-Kantorovich problem, i.e. the non-empty set of adherence points of $(\pi_n)_{n \geq 0}$ is a set of optimizers.*

Proof. Fix $\epsilon > 0$. By Prohorov's Theorem (see for instance [**?**], Th. 16.3) there are compact sets $K \subset X$ and $L \subset Y$ such that $\mu_n(X \setminus K) \leq \epsilon$ and $\nu_n(Y \setminus L) \leq \epsilon$ for all n. Hence $\pi_n(X \times Y \setminus K \times L) \leq \pi_n(X \times (Y \setminus L)) + \pi_n((X \setminus K) \times Y) \leq 2\epsilon$. Again by Prohorov's Theorem we know that there is a weakly converging subsequence $(\pi_{n_k})_{k \geq 0}$ with weak limit π. We have to prove that π is c-monotone, since by Proposition **??** the transport plan π is then strongly c-monotone and hence optimal. We know that for $m \geq 1$ and points $(x_i, y_i) \in \operatorname{supp} \pi$ for $i = 1, \ldots, m$, there exist sequences of points $(x_i^k, y_i^k)_{k \geq 0}$ such that $(x_i^k, y_i^k) \in \operatorname{supp} \pi_{n_k}$ and $(x_i^k, y_i^k) \to (x_i, y_i)$ as $k \to \infty$ (see again [**?**], Th. 16.3). Since π_{n_k} is c-monotone by Proposition **??** and Remark **??**, and since c is continuous, $\operatorname{supp} \pi_{n_k}$ is a c-monotone subset of $X \times Y$, hence

$$\sum_{i=1}^m c(x_i^k, y_i^k) \leq \sum_{i=1}^m c(x_{\sigma(i)}^k, y_{\sigma(i)}^k)$$

for permutations $\sigma \in \mathfrak{S}_m$, whence also for $k \to \infty$ by continuity of c. By the same reasoning any converging subsequence of $(\pi_n)_{n\geq 0}$ converges to a c-monotone transport plan, which is by Proposition **??** a strongly c-monotone one and hence optimal by Theorem **??**. □

Remark 4. *In particular the set of optimizers of a Monge-Kantorovich problem is compact in the weak topology on probability measures, since the constant sequences $\mu_n = \mu$ and $\nu_n = \nu$ for $n \geq 0$ are approximating sequences. This also yields that we cannot improve the result to the assertion on convergence of the sequence, but we point out that the convergence of the sequence in Theorem **??** holds true if and only if the optimizer is unique.*

Example 2. *Let c be a continuous, finitely valued cost function on the product $X \times Y$ of two polish spaces X, Y and let μ, ν be probability measures thereon. By the law of large numbers we can always find empirical distributions (see for instance* [**?**]*, Prop. 4.24)*

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i},$$
$$\nu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i},$$

which converge weakly to μ and ν, respectively. Each associated approximating sequence of optimizers $(\pi_n)_{n\geq 1}$ is given through a sequence of permutations $\sigma_n \in \mathfrak{S}_n$, $n \geq 1$, and

$$\pi_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i, y_{\sigma_n(i)})}.$$

*Notice that always $I(\pi_n) = \int_{X\times Y} c(x,y)\pi_n(dx,dy) < \infty$ for each fixed $n \in \mathbb{N}$. We deduce from Theorem **??** that all adherence points of $(\pi_n)_{n\geq 1}$ are optimizers of the Monge-Kantorovich problem and that there is at least one adherence point. Notice that each π_n for $n \geq 1$ is obtained by solving a finite optimization problem.*

The authors would like to thank the participants of the Seminar "Topics in Optimal Transportation" for the fruitful discussions on C. Villani's book. The Seminar is supported by the FWF-Wissenschaftskolleg W 800.

It is a particular pleasure to thank Lugi Ambrosio, Aldo Pratelli and Cedric Villani for their very helpful comments on a previous version of this note. After completion of our work we were kindly informed about the interesting and independent work [**?**] on the same subject in a slightly different framework. We also thank the anonymous referee for her/his pertinent and insightful report, which led to an improvement of the paper.

References

[1] Luigi Ambrosio and Aldo Pratelli, *Existence and Stability Results in the L^1-theory of optimal transportation*, CIME Course, Lecture in Mathematics, Springer Verlag, 2002.

[2] Werner Brannath and Walter Schachermayer, *A Bipolar Theorem for Subsets of $L^0_+(\Omega, \mathcal{F}, P)$*, Séminaire de Probabilités XXXIII, Springer Lecture Notes in Mathematics, Vol. 1709, 349-354 (1999).

[3] Freddy Delbaen and Walter Schachermayer, *A General Version of the Fundamental Theorem of Asset Pricing*, Mathematische Annalen **300**, 463–520, 1994.

[4] Olav Kallenberg, *Foundations of Modern Probability*, Probability and its Applications, 2nd edition, Springer New York, 2001.
[5] D. Ramachandran and Ludger Rüschendorf, *A general duality theorem for marginal problems*, Prob. Theory Rel. Fields **101**, 311-319, 1995.
[6] A. Pratelli, *About the sufficiency of the c-cyclical monotonicity for the optimal transport plans*, personal communication, 2006.
[7] Lars Svensson, *Sums of complemented subspaces in locally convex spaces*, Arkiv för Matematik **25** (1), 147–153, 1987.
[8] Cédric Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics **58**, American Mathematical Society, Providence Rhode Island, 2003.

Technical University Vienna, Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria