

Existence of Solutions to the Nonlinear Kantorovich Transportation Problem

V. I. Bogachev^{1,2,3,4*} and A. V. Rezbaev^{5**}

¹ Lomonosov Moscow State University, Moscow, 119991 Russia

² National Research University Higher School of Economics, Moscow, 101000 Russia

³ St. Tikhon's Orthodox University, Moscow, 115184 Russia

⁴ Moscow Center for Fundamental and Applied Mathematics, Moscow, 119991 Russia

⁵ National Research University Higher School of Economics, Moscow, 101000 Russia

Received April 14, 2022; in final form, April 14, 2022; accepted April 20, 2022

Abstract—We study the existence of solutions to the Kantorovich optimal transportation problem with a nonlinear cost functional generated by a cost function depending on the transport plan. We also consider the case of a cost function depending on the conditional measures of the transport plan. Broad sufficient conditions are obtained for the existence of optimal plans for Radon marginal distributions on completely regular spaces and a lower semicontinuous cost function.

DOI: 10.1134/S0001434622090048

Keywords: *Kantorovich problem, conditional measure, nonlinear cost functional, cost function.*

1. INTRODUCTION

It is well known that in the classical Kantorovich problem of minimizing the integral of a nonnegative Borel cost function h on the product of completely regular topological spaces X and Y with Radon probability measures μ and ν with respect to measures σ from the set $\Pi(\mu, \nu)$ of Radon probability measures on $X \times Y$ with projections μ and ν on the factors (such measures are called plans), a solution exists if the function h is lower semicontinuous (or at least lower semicontinuous on compact sets). A minimizing measure is called an optimal measure or an optimal Kantorovich plan, and the measures μ and ν are called marginals or marginal distributions. About this problem see [1]–[5]. In the general case, there exists an infimum

$$K_h(\mu, \nu) = \inf_{\sigma \in \Pi(\mu, \nu)} \int_{X \times Y} h d\sigma.$$

In the recent works [6]–[10] the Kantorovich problem has been considered with a more general, nonlinear in σ , cost functional given by a nonlocal cost function that depends on the plan. The cost function considered in these papers has the form

$$h(x, y, \sigma) = H(x, \sigma^x), \quad (1.1)$$

where σ^x are the conditional measures on Y with respect to the projection of σ (which equals μ) on X . The existence of a solution is proved in [8] in the case of Polish spaces and a lower semicontinuous function h on $X \times \mathcal{P}(Y)$, where the function $\sigma \mapsto h(x, \sigma)$ is convex, and $\mathcal{P}(Y)$ is the space of Borel probability measures on Y equipped with the weak topology. The goal of this paper is to extend this

*E-mail: vibogach@mail.ru

**E-mail: aratyo@yandex.ru

existence theorem to a wider class of functionals and spaces. We consider nonlinear functionals of the form

$$J_h(\sigma) = \int_{X \times Y} h(x, y, \sigma) \sigma(dx dy), \quad (1.2)$$

where the cost function $h: X \times Y \times \mathcal{P}(X \times Y) \rightarrow [0, +\infty)$ is lower semicontinuous. For Radon marginal distributions, we prove the existence of an optimal plan minimizing this functional. However, the very natural requirement of lower semicontinuity does not enable one to cover cost functions of the form $h(x, \sigma^x)$, since conditional measures can depend on σ only in a Borel way. Therefore, functionals of type (1.2) are considered separately and under the more restrictive condition that the cost function is convex with respect to the measure argument.

2. EXISTENCE THEOREM FOR NONLINEAR FUNCTIONALS

For a completely regular space X we denote by $\mathcal{P}_r(X)$ the space of Radon probability measures on X , i.e., Borel measures μ such that, for every Borel set B and every $\varepsilon > 0$, there exists a compact set $K \subset B$ such that $\mu(B \setminus K) \leq \varepsilon$. We equip $\mathcal{P}_r(X)$ with the weak topology, which on the whole space of signed measures is generated by the family of seminorms

$$p_f(\mu) = \left| \int_X f d\mu \right|,$$

where f is a continuous bounded function on X . If X is a complete metrizable space, then so is $\mathcal{P}_r(X)$, and if X is a Souslin space, i.e., the image of a complete separable metric space under a continuous mapping, then $\mathcal{P}_r(X)$ is also a Souslin space.

A family of measures $M \subset \mathcal{P}_r(X)$ is said to be *uniformly tight* if, for every $\varepsilon > 0$, there exists a compact set $K \subset B$ such that $\mu(B \setminus K) \leq \varepsilon$ for all $\mu \in M$. By Prohorov's theorem (see [11, Chap. 8]) such a set has compact closure in the weak topology. For Polish spaces, the converse is true, but in the general case this is false (even for the set of rational numbers, see [12, Sec. 4.8]).

Lemma 1. *Let X be a completely regular space, let Π be a uniformly tight compact subset in $\mathcal{P}_r(X)$, and let a function $h: X \times \Pi \rightarrow [0, +\infty)$ be lower semicontinuous on all sets of the form $K \times \Pi$, where K is compact in X . Then, the following function is lower semicontinuous:*

$$J_h(\sigma) = \int_X h d\sigma, \quad \Pi \rightarrow [0, +\infty].$$

Proof. The values $J_{\min(h, n)}(\sigma)$ are increasing to $J_h(\sigma)$. Therefore, we can consider only bounded functions h . We can assume that $h \leq 1$.

Assume first that the function h is lower semicontinuous on the whole product $X \times \Pi$. Suppose that a net σ_α converges weakly in Π to a measure σ . Then the Dirac measures δ_{σ_α} on Π converge weakly to the Dirac measure δ_σ . Therefore, the products $\sigma_\alpha \otimes \delta_{\sigma_\alpha}$ on $\Pi \times \mathcal{P}_r(\Pi)$ converge weakly to the product $\sigma \otimes \delta_\sigma$, see [12, Theorem 4.3.18]. Hence, due to the lower semicontinuity of h we have (see [11, Corollary 8.2.5] or [12, Corollary 4.3.5])

$$\liminf_\alpha \int_\Pi \int_X h(x, p) \sigma_\alpha(dx) \delta_{\sigma_\alpha}(dp) \geq \int_\Pi \int_X h(x, p) \sigma(dx) \delta_\sigma(dp);$$

in other words,

$$\liminf_\alpha \int_X h(x, \sigma_\alpha) \sigma_\alpha(dx) \geq \int_X h(x, \sigma) \sigma(dx),$$

which is equivalent to the lower semicontinuity of the function J_h .

We now consider the general case, still assuming that $h \leq 1$. Fix $\varepsilon > 0$. By assumption, there exists a compact set $K \subset X$ such that $\sigma(K) > 1 - \varepsilon$ for all $\sigma \in \Pi$. It is known (see [13, 1.7.15(c)]) that one can find a family of continuous functions $h_\alpha \geq 0$, on $K \times \Pi$ for which

$$h(x, \sigma) = \sup_\alpha h_\alpha(x, \sigma) \quad \forall x \in K, \quad \sigma \in \Pi.$$

Each function h_α extends to a continuous function $g_\alpha: X \times \Pi \rightarrow [0, 1]$. The function $g(x, \sigma) = \sup_\alpha g_\alpha(x, \sigma)$ is lower semicontinuous on the entire product $X \times \Pi$ and coincides with h on $K \times \Pi$ and the corresponding function J_g , according to what has been proved above, is also lower semicontinuous. It remains to note that

$$|J_g(\sigma) - J_h(\sigma)| \leq 2\varepsilon \quad \forall \sigma \in \Pi,$$

since $g = h$ on $K \times \Pi$, and the integrals of $h(x, \sigma)$ and $g(x, \sigma)$ over the complement of K with respect to any measure $\sigma \in \Pi$ do not exceed ε . Thus, the function J_h is uniformly approximated by lower semicontinuous functions and therefore it is also lower semicontinuous. \square

Remark 1. It is seen from the proof that if the function h is bounded and continuous on the whole product $X \times \mathcal{P}_r(X)$, then the function J_h is continuous on the space $\mathcal{P}_r(X)$.

Theorem 1. *Let h be lower semicontinuous on all sets of the form $K \times \Pi(\mu, \nu)$, where K is compact in $X \times Y$. Then there exists an optimal plan.*

Proof. Since the set of plans $\Pi(\mu, \nu)$ is uniformly tight and weakly compact, by Lemma 1, the function J_h is lower semicontinuous on $\Pi(\mu, \nu)$. Now the existence of an optimal plan follows from the fact that a lower semicontinuous function on a compact set attains its minimum on this compact set. \square

3. COST FUNCTIONS WITH CONDITIONAL MEASURES

In this section we assume that X and Y are completely regular spaces, $\mu \in \mathcal{P}_r(X)$ and $\nu \in \mathcal{P}_r(Y)$. Let $\mathcal{B}(T)$ denote the Borel σ -algebra of a topological space T and let $\mathcal{Ba}(T)$ denote the Baire σ -algebra that is generated by all continuous functions on T . In the case of a completely regular Souslin space, the equality $\mathcal{Ba}(T) = \mathcal{B}(T)$ holds, see [11, Theorem 6.7.7]. If $T = \mathcal{P}_r(Y)$ with the weak topology, then $\mathcal{Ba}(T)$ is generated by all functions on T of the form

$$p \mapsto \int_Y \varphi(y) p(dy),$$

where φ is a continuous bounded function on Y , see [11, Theorem 6.10.6].

We additionally require from the set $\Pi(\mu, \nu)$ that, for every measure $\sigma \in \Pi(\mu, \nu)$, there exist unique μ -a.e. conditional measures $\sigma^x \in \mathcal{P}_r(Y)$ with respect to the projection of σ on X (that equals μ). This means that σ has the form $\sigma(dx dy) = \sigma^x(dy)\mu(dx)$, the function $x \mapsto \sigma^x(B)$ is measurable with respect to μ for all $B \in \mathcal{B}(Y)$ and, for every bounded function f on $X \times Y$, measurable with respect to $\mathcal{B}(X) \otimes \mathcal{B}(Y)$, the equality

$$\int_{X \times Y} f d\sigma = \int_X \int_Y f(x, y) \sigma^x(dy) \mu(dx)$$

holds. It is clear that it suffices to have this equality for all functions of the form $I_A(x)I_B(y)$, where $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(Y)$.

We assume in addition that the mapping $x \mapsto \sigma^x$, $X \mapsto \mathcal{P}_r(Y)$ is measurable with respect to the σ -algebras $\mathcal{B}(X)$ and $\mathcal{Ba}(\mathcal{P}_r(Y))$. Then the mapping

$$x \mapsto (x, \sigma^x), \quad X \rightarrow X \times \mathcal{P}_r(Y)$$

is measurable with respect to the σ -algebras $\mathcal{B}(X)$ and $\mathcal{B}(X) \otimes \mathcal{Ba}(\mathcal{P}_r(Y))$. Hence, for any function H on $X \times \mathcal{P}_r(Y)$ that is measurable with respect to $\mathcal{B}(X) \otimes \mathcal{Ba}(\mathcal{P}_r(Y))$, the function

$$x \mapsto H(x, \sigma^x)$$

is Borel on X .

Finally, we assume that, for every measure $P \in \mathcal{P}_r(X \times \mathcal{P}_r(Y))$ with the projection μ on X , there exist conditional measures $P^x \in \mathcal{P}_r(\mathcal{P}_r(Y))$. This is automatically fulfilled if the set $\{(x, p, x): x \in X, p \in \mathcal{P}_r(Y)\}$, which is closed, belongs to $\mathcal{B}(X \times \mathcal{P}_r(Y)) \otimes \mathcal{B}(X)$, see [11, Corollary 10.5.7].

These requirements are natural for a correct setting of the nonlinear Kantorovich problem and are satisfied, for example, when the measures μ and ν are concentrated on Souslin sets (see [11, Sec. 10.4]). Moreover, it is sufficient that only the space Y be Souslin, since for every topological space T and every completely regular Souslin space E , one has

$$\mathcal{B}(T \times E) = \mathcal{B}(T) \otimes \mathcal{B}(E)$$

(see [14], [15, Sec. 4A3X]), and the Borel σ -algebra of any Souslin space is countably generated (see [11, Corollary 6.7.5]).

For a Radon probability measure Q on the space of measures $\mathcal{P}_r(Y)$, the barycenter is given by

$$\beta_Q := \int_{\mathcal{P}_r(Y)} p Q(dp),$$

where the vector integral with values in the space of measures is understood as the equality

$$\beta_Q(A) = \int_{\mathcal{P}_r(Y)} p(A) Q(dp)$$

for all Borel sets $A \subset Y$. It is a well known fact that the measure β_Q is τ -additive (see [11, Corollary 8.9.9]), but we are interested in Radon barycenters. The Radon property holds if all τ -additive measures on Y are Radon (say, in the case of a Souslin space Y), and also if the measure Q is concentrated on the countable union of some uniformly tight compact set $S_n \subset \mathcal{P}_r(Y)$. Indeed, the barycenters of the measures $I_{S_n} \cdot Q$ converge in variation to β_Q , and for each $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset Y$ such that $p(K_\varepsilon) \geq 1 - \varepsilon$ for all $p \in S_n$,

$$\beta_{I_{S_n} \cdot Q}(K_\varepsilon) \geq (1 - \varepsilon)Q(S_n),$$

i.e., the measures $\beta_{I_{S_n} \cdot Q}$ are tight, hence they are Radon. Below, barycenters will be considered for measures concentrated on countable unions of uniformly tight compact sets and, therefore, they will be Radon measures. It is worth noting that if all τ -additive measures are Radon on Y , then every Radon measure on $\mathcal{P}_r(Y)$ is concentrated on a countable union of uniformly tight compact sets, see [11, Theorem 8.10.6].

Lemma 2. (i) *Suppose that E is a completely regular space, a net $P_\alpha \in \mathcal{P}_r(E)$ is uniformly tight and converges weakly to a measure $P \in \mathcal{P}_r(E)$ and $H: E \rightarrow [0, 1]$ is a function such that, for each $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset E$ for which $P_\alpha(E \setminus K_\varepsilon) < \varepsilon$ for all α and the restriction of H to K_ε is lower semicontinuous. Then*

$$\int_E H dP \leq \liminf_\alpha \int_E H dP_\alpha. \quad (3.1)$$

(ii) *Suppose that Y is a completely regular space, a measure $Q \in \mathcal{P}_r(\mathcal{P}_r(Y))$ is concentrated on a countable union of uniformly tight sets and a bounded function H on $\mathcal{P}_r(Y)$ is convex and lower semicontinuous on uniformly tight sets. Then*

$$H\left(\int_{\mathcal{P}_r(Y)} p Q(dp)\right) \leq \int_{\mathcal{P}_r(Y)} H(p) Q(dp). \quad (3.2)$$

Proof. (i) Fix $\varepsilon > 0$. By assumption, there exists a compact set $K \subset E$, such that $P_\alpha(K) \geq 1 - \varepsilon$ for all α and the restriction of H to K is lower semicontinuous. By weak convergence, we have $P(K) \geq 1 - \varepsilon$. As above, one can find a family of continuous functions $H_\alpha \geq 0$ on E , such that $H(x) = \sup_\alpha H_\alpha(x)$ for $x \in K$. We can assume that $H_\alpha \leq 1$ for all α passing to $\min(H_\alpha, 1)$. The function $G(x) = \sup_\alpha H_\alpha(x)$ is lower semicontinuous on the whole space E and equal to H on K . By weak convergence of the measures P_α and the lower semicontinuity of G , we have (see [11, Corollary 8.2.5])

$$\liminf_\alpha \int_E G dP_\alpha \geq \int_E G dP.$$

Using the inequalities

$$\int_E H dP_\alpha \geq \int_K H dP_\alpha = \int_K G dP_\alpha \geq \int_E G dP_\alpha - \varepsilon,$$

we obtain the estimates

$$\liminf_\alpha \int_E H dP_\alpha \geq \int_E G dP - \varepsilon \geq \int_E H dP - 3\varepsilon.$$

Since ε is arbitrary, we arrive at (3.1).

(ii) By the convexity of H the indicated inequality holds for any convex combinations of Dirac measures. Let us extend it to measures Q with compact uniformly tight supports. Since the convex hull of a uniformly tight set is uniformly tight, we can assume that the measure Q is concentrated on a set $C \subset \mathcal{P}_r(Y)$ that is convex, compact, and uniformly tight. Let us find a net of discrete probability measures Q_t on C that converges weakly to Q and has the property

$$\int_{\mathcal{P}_r(Y)} H dQ_t \rightarrow \int_{\mathcal{P}_r(Y)} H dQ.$$

This can be done similarly to the standard proof of the density of discrete measures in the space of Baire measures (see [11, Example 8.1.6]). Let us recall the construction. The indices t of the directed set indexing the selected measures are taken in the form of collections f_1, \dots, f_k, r , where f_i are bounded continuous functions on $\mathcal{P}_r(Y)$ and $r \in \mathbb{N}$, and the partial order on collections is introduced as follows:

$$(f_1, \dots, f_k, r) \leq (g_1, \dots, g_m, s),$$

if $\{f_1, \dots, f_k\} \subset \{g_1, \dots, g_m\}$ and $r \leq s$. A discrete measure $Q_{f_1, \dots, f_k, r}$ with support in C is found in such a way that the difference between the integrals of f_1, \dots, f_k, H with respect to $Q_{f_1, \dots, f_k, r}$ and Q is not greater than $1/r$. The barycenters of the measures Q_t belong to C . The following equality holds:

$$\int_C p Q(dp) = \lim_t \int_C p Q_t(dp),$$

because by the definition of the weak topology this equality is equivalent to the fact that, for every bounded continuous function f on Y , we have

$$\int_C \int_Y f(y) p(dy) Q(dp) = \lim_t \int_C \int_Y f(y) p(dy) Q_t(dp),$$

but the latter is true by weak convergence of Q_t to Q and the continuity of the inner integral in p . Thus, by the lower semicontinuity of H on C , we have

$$\begin{aligned} H\left(\int_C p Q(dp)\right) &\leq \liminf_t H\left(\int_C p Q_t(dp)\right) \\ &\leq \liminf_t \int_C H(p) Q_t(dp) = \int_C H(p) Q(dp). \end{aligned}$$

In the general case, the measure Q is concentrated on the union of increasing sets C_n of the considered form. This measure is the limit in variation of the normalized restrictions $Q_n = Q(C_n)^{-1}Q|_{C_n}$. The barycenters β_{Q_n} of these measures converge in variation to the barycenter β_Q of the measure Q , because

$$\int_{\mathcal{P}_r(Y) \setminus C_n} p(A) Q(dp) \leq Q(\mathcal{P}_r(Y) \setminus C_n)$$

for any Borel set $A \subset \mathcal{P}_r(Y)$. Therefore, the sequence β_{Q_n} is uniformly tight, and due to the lower semicontinuity of H on uniformly tight sets we have the estimate

$$H(\beta_Q) \leq \liminf_{n \rightarrow \infty} H(\beta_{Q_n}).$$

For the measure Q_n the required inequality has already been established, so it holds for the measure Q , since by convergence in variation the integral of H with respect to the measure Q_n tends to the integral with respect to the measure Q .

Note that this step of justification is omitted in [8], but it is also necessary in the case of lower semicontinuity on the whole space considered there. \square

Lemma 3. *Suppose that a function $H: X \times \mathcal{P}_r(Y) \rightarrow [0, +\infty)$ is measurable with respect to $\mathcal{B}(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y))$ and lower semicontinuous on the sets of the form $K \times S$, where K is a compact set in X and $S \subset \mathcal{P}_r(Y)$ is uniformly tight and convex in the second argument. Then the function*

$$J_H(\sigma) = \int_X H(x, \sigma^x) \mu(dx) \quad (3.3)$$

is lower semicontinuous on $\Pi(\mu, \nu)$.

Proof. 1. We use slightly modified and partially simplified arguments from [8], in which we consider nets instead of sequences. For every measure $\sigma \in \Pi(\mu, \nu)$ we introduce the mapping

$$F_\sigma(x) = (x, \sigma^x), \quad X \rightarrow X \times \mathcal{P}_r(Y).$$

By our assumption, this mapping is μ -measurable if the space $X \times \mathcal{P}_r(Y)$ is equipped with the product of the σ -algebras $\mathcal{B}(X)$ and $\mathcal{B}a(\mathcal{P}_r(Y))$ on the factors. Therefore, on $\mathcal{B}a(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y)) \subset \mathcal{B}(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y))$ the measure

$$P_\sigma = \mu \circ F_\sigma^{-1}$$

is defined, i.e., the image of the measure μ under the indicated mapping given by the formula

$$(\mu \circ F_\sigma^{-1})(B) = \mu(x : (x, \sigma^x) \in B).$$

In the case of Souslin spaces, the measures P_α are automatically Radon. But in the general case, the existence of their Radon extensions requires verification. We will keep the same notation for these extensions. It is known (see [11, Corollary 7.3.5]) that a Radon extension exists if each measure P_σ is tight on $\mathcal{B}a(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y))$, i.e., for every $\varepsilon > 0$ here exists a compact set C in $X \times \mathcal{P}_r(Y)$ such that $P_\sigma(D) \leq \varepsilon$, whenever a set $D \in \mathcal{B}a(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y))$ is disjoint C . Since the projection of the measure P_σ on X equals μ on Baire sets and hence is tight, it suffices to verify that the projection of P_σ on $\mathcal{P}_r(Y)$ is tight. We will show that in fact the set of measures $\{P_\sigma : \sigma \in \Pi(\mu, \nu)\}$ on $X \times \mathcal{P}_r(Y)$ is uniformly tight. Moreover, the set Λ of their projections on $\mathcal{P}_r(Y)$ not only is uniformly tight, but is concentrated on a countable union of compact uniformly tight sets from $\mathcal{P}_r(Y)$ in the sense that there exists a sequence of uniformly tight compact sets $M_n \subset \mathcal{P}_r(Y)$, for which $\sup_{\lambda \in \Lambda} \lambda^*(M_n) \geq 1 - 1/n$, where the outer measure is taken with respect to $\mathcal{B}a(\mathcal{P}_r(Y))$ (recall that compact sets in $\mathcal{P}_r(Y)$ are not necessarily uniformly tight). It suffices to verify the latter. Indeed, let $\varepsilon > 0$. For every $\delta > 0$ there exists a compact set $C_\delta \subset Y$ such that $\nu(C_\delta) > 1 - \delta^2$. By the Chebyshev inequality, for every measure $\sigma \in \Pi(\mu, \nu)$, we obtain

$$\begin{aligned} \mu(x : \sigma^x(C_\delta) \leq 1 - \delta) &= \mu(x : 1 - \sigma^x(C_\delta) \geq \delta) \leq \delta^{-1} \int_X (1 - \sigma^x(C_\delta)) \mu(dx) \\ &= \delta^{-1}(1 - \sigma(X \times C_\delta)) = \delta^{-1}(1 - \nu(C_\delta)) \leq \delta. \end{aligned}$$

For δ we take the numbers $\varepsilon 2^{-n}$. Let

$$S_\varepsilon = \{p \in \mathcal{P}_r(Y) : p(C_{2^{-n}\varepsilon}) \geq 1 - 2^{-n}\varepsilon \forall n \geq 1\}.$$

The set S_ε is closed in the weak topology and uniformly tight by definition. Hence it is compact in $\mathcal{P}_r(Y)$. Moreover, for every measure $\sigma \in \Pi(\mu, \nu)$, we have

$$\mu(x : \sigma^x \in S_\varepsilon) \geq 1 - \varepsilon,$$

since

$$\mu(x : \sigma^x(C_{2^{-n}\varepsilon}) < 1 - 2^{-n}\varepsilon) \leq 2^{-n}\varepsilon.$$

Now for the required sets M_n we can take $S_{1/n}$. Indeed, suppose that a set $E \in \mathcal{B}a(\mathcal{P}_r(Y))$ is disjoint with $S_{1/n}$. Let π_σ be the projection of the measure P_σ on $\mathcal{P}_r(Y)$. Then by definition

$$\pi_\sigma(E) = \mu(x : \sigma^x \in E) \leq \mu(x : \sigma^x \in \mathcal{P}_r(Y) \setminus S_{1/n}) \leq 1/n.$$

Hence $\pi_\sigma^*(S_{1/n}) \geq 1 - 1/n$. The necessity to use the outer measure is explained by the fact that the measure P_σ is initially defined not on the whole Borel σ -algebra, but only on $\mathcal{B}a(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y))$. By the established tightness, it has an extension, but its values on compact sets need not be calculated according to the formula by which they are defined on $\mathcal{B}a(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y))$.

2. Note that so far we have not used the function H . Since the quantities $J_{\min(H,n)}(\sigma)$ increase to $J_H(\sigma)$, the assertion reduces to the case of a bounded function H , so one can assume that $H \leq 1$. Suppose that a net of measures σ_α converges weakly to a measure $\Pi(\mu, \nu)$. We prove the inequality

$$\liminf_\alpha \int_X H(x, \sigma_\alpha^x) \mu(dx) \geq \int_X H(x, \sigma^x) \mu(dx), \tag{3.4}$$

which implies the lower semicontinuity of the cost functional in (3.3).

We set $P_\alpha := \mu \circ F_{\sigma_\alpha}^{-1}$, where the Radon extensions mentioned above are used. According to what has been proved above, the family of measures P_α is uniformly tight, therefore, by the Prokhorov theorem, it has compact closure (see [11, Theorem 8.6.7]). Passing to a subnet, we can assume that the measures P_α converge weakly to a measure $P \in \mathcal{P}_r(X \times \mathcal{P}_r(Y))$. Moreover, one can assume that

$$\liminf_\alpha \int_X H(x, \sigma_\alpha^x) \mu(dx) = \lim_\alpha \int_X H(x, \sigma_\alpha^x) \mu(dx).$$

Note that by the definition of the image of the measure and due to the measurability of H with respect to $\mathcal{B}a(X) \otimes \mathcal{B}a(\mathcal{P}_r(Y))$, the following equality holds:

$$\int_X H(x, \sigma_\alpha^x) \mu(dx) = \int_{X \times \mathcal{P}_r(Y)} H(x, p) P_\alpha(dx dp).$$

According to the previous lemma

$$\liminf_\alpha \int_{X \times \mathcal{P}_r(Y)} H(x, p) P_\alpha(dx dp) \geq \int_{X \times \mathcal{P}_r(Y)} H(x, p) P(dx dp).$$

3. The projection of the measure P on X equals the limit of the projections of the measures P_α which are equal to μ , hence it also coincides with μ . The conditional measures P^x with respect to this projection belong to $\mathcal{P}_r(\mathcal{P}_r(Y))$. The barycenter of the measure P^x (as explained above, it is μ -almost everywhere included in $\mathcal{P}_r(Y)$ namely for those x for which the measure P^x is concentrated on the countable union of compact uniformly tight sets) is defined by the equality

$$\beta_x := \int_{\mathcal{P}_r(Y)} p P^x(dp).$$

It is easily seen that

$$\beta_x = \sigma^x \quad \mu\text{-a.e.}$$

Indeed, for any bounded continuous functions f on X and g on Y , the following equalities hold:

$$\begin{aligned} \int_X \int_Y f(x)g(y) \beta_x(dy) \mu(dx) &= \int_X \int_{\mathcal{P}_r(Y)} \int_Y f(x)g(y) p(dy) P^x(dp) \mu(dx) \\ &= \int_{X \times \mathcal{P}_r(Y)} \int_Y f(x)g(y) p(dy) P(dx dp) \\ &= \lim_\alpha \int_{X \times \mathcal{P}_r(Y)} \int_Y f(x)g(y) p(dy) P_\alpha(dx dp) = \lim_\alpha \int_X \int_Y f(x)g(y) \sigma_\alpha^x(dy) \mu(dx) \\ &= \lim_\alpha \int_{X \times Y} f(x)g(y) \sigma_\alpha(dx dy) = \int_{X \times Y} f(x)g(y) \sigma(dx dy). \end{aligned}$$

By virtue of the supposed uniqueness of conditional measures for $\sigma \in \Pi(\mu, \nu)$ this implies that β_x coincides with σ^x .

4. According to the previous lemma, due to the convexity of the function H in the second argument, for each x for which there also is the lower semicontinuity in the second argument on uniformly tight sets from $\mathcal{P}_r(Y)$, and the measure P^x is concentrated on a countable union of compact uniformly tight sets, the following inequality holds:

$$H\left(x, \int_{\mathcal{P}_r(Y)} p P^x(dp)\right) \leq \int_{\mathcal{P}_r(Y)} H(x, p) P^x(dp).$$

5. Finally, we find a compact set $K \subset X$ such that $\mu(K) > 1 - \varepsilon$ and for each $x \in K$ the functions $p \mapsto H(x, p)$ are lower semicontinuous on uniformly tight sets from $\mathcal{P}_r(Y)$, while the measure P^x is concentrated on a countable union of uniformly tight sets. The latter can be achieved according to what has been proved at the first step.

Applying the above inequality, we obtain

$$\begin{aligned} \int_{X \times \mathcal{P}_r(Y)} H(x, p) P(dx, dp) &= \int_X \int_{\mathcal{P}_r(Y)} H(x, p) P^x(dp) \mu(dx) \\ &\geq \int_K H\left(x, \int_{\mathcal{P}_r(Y)} p P^x(dp)\right) \mu(dx) \geq \int_X H\left(x, \int_{\mathcal{P}_r(Y)} p P^x(dp)\right) \mu(dx) - \varepsilon. \end{aligned}$$

Since ε is arbitrary, we have

$$\int_{X \times \mathcal{P}_r(Y)} H(x, p) P(dx, dp) \geq \int_X H\left(x, \int_{\mathcal{P}_r(Y)} p P^x(dp)\right) \mu(dx),$$

which completes the proof. \square

The following fact follows from what has been proved and the weak compactness of the set of plans $\Pi(\mu, \nu)$.

Theorem 2. *Suppose that the cost function $H: X \times \mathcal{P}_r(Y) \rightarrow [0, +\infty)$ is measurable with respect to $\mathcal{B}_a(X) \otimes \mathcal{B}_a(\mathcal{P}_r(Y))$, lower semicontinuous on all sets of the form $K \times S$, where K is a compact set in X and $S \subset \mathcal{P}_r(Y)$ is uniformly tight, and convex in the second argument. Then*

$$\inf_{\sigma \in \Pi(\mu, \nu)} \int_X H(x, \sigma^x) \mu(dx)$$

is attained, that is, an optimal plan exists.

Remark 2. (i) The statements obtained above remain valid in the situation where the functions H and h take values in $[0, +\infty]$. It suffices to apply the established facts to the functions $\min(h, N)$ and $\min(H, N)$.

(ii) It is clear from the proof that it suffices to impose the measurability condition on the function H on the sets of the form $K \times S$ with compact factors. For a broad class of spaces, the lower semicontinuity on such sets implies the $\mathcal{B}(K) \otimes \mathcal{B}_a(S)$ -measurability. For example, this is true if Y is Souslin and the Borel and Baire σ -algebras coincide on compact sets in X . In the case of general spaces (even Souslin spaces), the condition of lower semicontinuity on compact sets is weaker than the global lower semicontinuity (in the theorem, the condition is even slightly weaker, since we are speaking of uniformly tight compact sets). In Souslin spaces, compact sets are metrizable, therefore, this condition can be verified by using countable sequences. Moreover, for such spaces, the lower semicontinuity on compact sets implies the Borel measurability on compact sets, which coincides with the $\mathcal{B}_a(X) \otimes \mathcal{B}_a(\mathcal{P}_r(Y))$ -measurability, therefore, it need not be required additionally. On some sequential properties of spaces of measures, see [16].

Nonlinear cost functionals can be also considered for the transport problem with constraints on the densities of optimal plans in the spirit of the papers [17], [18]. This will be done in a separate paper. In addition, the role of convexity of the cost function depending on conditional measures will be studied.

FUNDING

This work was supported by the Russian Foundation for Basic Research under grant 20-01-00432, the project “Kantorovich’s parametric problem of optimal transportation”, with the support of PSTGU and the “Living Tradition” foundation, and also supported by the Ministry of Education and Science of the Russian Federation as part of the program of the Moscow Center for Fundamental and Applied Mathematics under the agreement No. 075-15-2022-284 (results in Sec. 2).

REFERENCES

1. L. Ambrosio and N. Gigli, “A user’s guide to optimal transport,” in *Lecture Notes in Math.*, Vol. 2062: *Modelling and Optimisation of Flows on Networks* (Springer, Heidelberg, 2013), pp. 1–155.
2. V. I. Bogachev and A. V. Kolesnikov, “Monge–Kantorovich problem: achievements, connections, and prospects,” *Russian Math. Surveys* **67** (5), 785–890 (2012).
3. S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems. I* (Springer, New York, 1998); *Mass Transportation Problems. II* (Springer, New York, 1998).
4. F. Santambrogio, *Optimal Transport for Applied Mathematicians* (Birkhäuser, New York, 2015).
5. C. Villani, *Optimal Transport, Old and New* (Springer, New York, 2009), pp. xxii+973.
6. N. Gozlan, C. Roberto, P. M. Samson, and P. Tetali, “Kantorovich duality for general transport costs and applications,” *J. Funct. Anal.* **273** (11), 3327–3405 (2017).
7. J. J. Alibert, G. Bouchitté, and T. Champion, “A new class of costs for optimal transport planning,” *European J. Appl. Math.* **30** (6), 1229–1263 (2019).
8. J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer, “Existence, duality, and cyclical monotonicity for weak transport costs,” *Calc. Var. Partial Differ. Equ.* **58** (203), 1–28 (2019).
9. B. Acciaio, M. Beiglböck, and G. Pammer, “Weak transport for non-convex costs and model-independence in a fixed-income market,” *Math. Finance* **31** (4), 1423–1453 (2021).
10. J. Backhoff-Veraguas and G. Pammer, “Applications of weak transport theory,” *Bernoulli J.* **28** (1), 370–394 (2022).
11. V. I. Bogachev, *Measure Theory. I, II* (Springer, Berlin, 2007).
12. V. I. Bogachev, *Weak Convergence of Measures* (Amer. Math. Soc., Providence, RI, 2018).
13. P. Engelking, *General Topology* (Polish Sci. Publ., Warszawa, 1977), pp. 626.
14. D. H. Fremlin, R. A. Johnson, and E. Wajch, “Countable network weight and multiplication of Borel sets,” *Proc. Amer. Math. Soc.* **124** (9), 2897–2903 (1996).
15. D. H. Fremlin, *Measure Theory. Vol. 4: Topological Measure Spaces. Part I, II* (Colchester, 2006).
16. V. I. Bogachev, “On sequential properties of spaces of measures,” *Math. Notes* **110** (3), 449–453 (2021).
17. J. Korman and R. J. McCann, “Optimal transportation with capacity constraints,” *Trans. Amer. Math. Soc.* **367** (3), 1501–1521 (2015).
18. V. I. Bogachev, A. N. Doledenok, and I. I. Malofeev, “The Kantorovich problem with a parameter and density constraints,” *Nath. Notes* **110** (6), 952–955 (2021).