



Math-Net.Ru

Общероссийский математический портал

V. I. Bogachev, Kantorovich problem of optimal transportation of measures: new directions of research, *Russian Mathematical Surveys*, 2022, Volume 77, Issue 5, 769–817

DOI: 10.4213/rm10074e

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением <http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 3.147.86.30

5 января 2025 г., 06:47:18



To the 110th anniversary of the birth of
Leonid Vital'evich Kantorovich

Kantorovich problem of optimal transportation of measures: new directions of research

V. I. Bogachev

Abstract. This paper gives a survey of investigations in the last decade and new results on various recent modifications of the classical Kantorovich problem of the optimal transportation of measures. We discuss in detail nonlinear Kantorovich problems, problems with constraints on the densities of transport plans, and optimal transportation problems with a parameter. In addition, we consider some questions relating to the geometry and topology of spaces of measures connected with these new formulations.

Bibliography: 134 items.

Keywords: Kantorovich problem, nonlinear Kantorovich problem, Monge problem, Kantorovich metric, optimal transportation, conditional measure.

Contents

1. Introduction	769
2. Nonlinear Kantorovich problems	774
3. Problems with fixed barycentres	779
4. Problems with conditional measures	783
5. Optimal transportation with density constraints	789
6. Multimarginal and multistochastic problems	793
7. Optimal transportation with parameters	797
8. Metrics and topologies of Kantorovich type	802
9. Other directions of research	807
Bibliography	810

1. Introduction

The goal of this survey is to inform the reader about some new directions of investigation of the Kantorovich problem of the optimal transportation of measures that arose in the last decade. In addition to the well-known monographs [118] and [128], in which a detailed exposition covers the main advances made in the 20th century, now there is a whole series of more recent thorough monographic

This research was supported by the Russian Science Foundation grant no. 22-11-00015 at the Lomonosov Moscow State University.

AMS 2020 Mathematics Subject Classification. Primary 49Q22.

presentations of this circle of problems; see [5], [7], [66], [69], [120], [129], [131], and also the survey [33], dedicated to the centennial of Kantorovich's birth, where the principal results obtained after the publication of his foundational short note [84] were discussed. Nevertheless, the intensive development of this area outpaces its exposition in books. During the last decade several interesting new modifications of the classical Kantorovich problem of optimal transportation of measures appeared. Among many new settings of problems, ideas, and methods connected with Kantorovich problems we can single out nonlinear problems of Kantorovich type, in which one deals with the minimization of integrals of functions which also depend on measures with respect to which the integration is performed, some versions of the classical problem with constraints on the densities of transport plans (which do not fit yet another interesting kind of Kantorovich problem: optimal plans with additional constraints), and also Kantorovich problems with a parameter. The aim of this article is to give a systematic presentation of Kantorovich problems with new formulations. Many results presented below have rather complicated and long proofs, so we state them with references to the original works. However, in some important cases the proofs are also included; this concerns especially assertions that appeared in original works in some special situations, for example, for metric spaces, but that are also valid for general completely regular spaces. Some information about Kantorovich's life and work can be found in the materials published for the centennial of his birth (see [85], [124]–[126], and also [9]).

In order to formulate new versions of the Kantorovich problem we recall its classical version (in the contemporary form, since Kantorovich considered it in a more special case). Suppose we are given probability spaces (X, \mathcal{B}_X, μ) and (Y, \mathcal{B}_Y, ν) and a non-negative $\mathcal{B}_X \otimes \mathcal{B}_Y$ -measurable function h (called a cost function) on the product $X \times Y$. In Kantorovich's foundational paper X and Y were metric compacta with Borel measures, and the cost function (whose value at a pair of points x, y was interpreted as the work on transportation of a unit mass from x to y) was continuous; in the most important examples, in particular, in the joint papers [87] and [88] with Rubinshtein, and also in [86], Chap. VIII, § 4, it equals the distance. However, it was already noted in [84] that "some of the definitions and results presented can be stated for spaces of more general nature". Let $\Pi(\mu, \nu)$ denote the set of all probability measures on the space $(X \times Y, \mathcal{B}_X \otimes \mathcal{B}_Y)$ having projections μ and ν onto the factors, that is, measures σ for which

$$\sigma(A \times Y) = \mu(A), \quad A \in \mathcal{B}_X, \quad \text{and} \quad \sigma(X \times B) = \nu(B), \quad B \in \mathcal{B}_Y.$$

Measures in the set $\Pi(\mu, \nu)$ are called transport plans or Kantorovich plans. The fixed measures μ and ν are called the marginal distributions. Denoting the projections of a measure σ given on $(X \times Y, \mathcal{B}_X \otimes \mathcal{B}_Y)$ onto the factors by σ_X and σ_Y , the above equalities can be written in the form

$$\sigma_X = \mu \quad \text{and} \quad \sigma_Y = \nu.$$

The Kantorovich problem consists in minimizing the integral

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

over the measures $\sigma \in \Pi(\mu, \nu)$. Under broad conditions, this Kantorovich problem has a solution, that is, there exists a measure in $\Pi(\mu, \nu)$ at which the minimum is attained. Such a measure (which is not always unique) is called an optimal measure or an optimal Kantorovich plan. For example, a solution exists if the cost function on a product of completely regular spaces with Radon measures is lower semicontinuous and bounded (see [33]). In place of boundedness it suffices to have a measure in $\Pi(\mu, \nu)$ with respect to which the cost function is integrable. Of course, if the integrals of the cost function over all plans are infinite, then one can also assume that the minimum exists and is infinite. In the general case there is a (possibly infinite) infimum $K_h(\mu, \nu)$ of the indicated integrals:

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

A multi-marginal Kantorovich problem can be stated similarly; in this problem there are n (or even infinitely many) marginals and the cost function is defined on the product of the corresponding spaces.

The Kantorovich problem is closely connected with the Monge problem stated in the 18th century for the same triple (μ, ν, h) as in the Kantorovich problem and consisting in minimizing the integral

$$\int_X h(x, T(x)) \mu(dx)$$

over all measurable maps $T: X \rightarrow Y$ taking the measure μ to ν , that is, satisfying the equality $\nu = \mu \circ T^{-1}$, where the measure $\mu \circ T^{-1}$ is defined by the equality

$$(\mu \circ T^{-1})(B) = \mu(T^{-1}(B))$$

and is called the image of μ under the map T . As in the Kantorovich problem, in the general case there is only an infimum

$$M_h(\mu, \nu) = \inf_T \int_X h(x, T(x)) \mu(dx),$$

where inf is taken over the maps T with the indicated property. If the minimum is attained at some map T , then it is called an optimal Monge map. Unlike the Kantorovich problem, a minimum in the Monge problem is attained much more rarely, even for nice cost functions on an interval. Sufficient conditions for the existence of a minimum have here a rather special character; moreover, some restrictions are needed both on the marginals and the cost function. For example, if $X = Y = \mathbb{R}^n$ and $h(x, y) = |x - y|$, then it suffices to assume the absolute continuity of both marginals. Thus, the Kantorovich problem turns out to be more flexible. Of course, this is not surprising, since it deals with minimizing a linear functional on a convex compact set of measures, while the Monge problem is essentially nonlinear. Moreover, as we will see below, even a nonlinear version of the Kantorovich problem turns out to be closer to its linear version with regard to its properties than to the Monge problem. Although the conditions for the existence of minima in the

two problems differ substantially, for a continuous cost function h the infima coincide under quite general conditions, covering the cases that are most important for applications: the equality

$$K_h(\mu, \nu) = M_h(\mu, \nu)$$

was established in [117] (see also [8]) for atomless measures μ and ν on complete separable metric spaces, extended to Souslin spaces in [102], and was proved for Radon measures on completely regular spaces in [32] under the additional condition of the separability of the measures μ and ν (that is, the separability of their L^1 -spaces); moreover, the condition of separability cannot be omitted as shown in [31]. In the most general case the inequality

$$K_h(\mu, \nu) \leq M_h(\mu, \nu)$$

is true. This is obvious from the fact that for every map T taking the measure μ to the measure ν , the measure σ on the graph of T in the space $X \times Y$, obtained as the image of μ under the map $x \mapsto (x, T(x))$, belongs to $\Pi(\mu, \nu)$, and the integral of h with respect to it is the integral of $h(x, T(x))$ against μ . However, transport plans can exist that are not generated by any maps of the measure μ to a measure ν . For example, if $X = Y = [0, 1]$, then every measure on $[0, 1]^2$ that has a density with respect to Lebesgue measure vanishes on the graph of every Borel map (which is obvious from the Fubini theorem). The set of general transport plans is not just larger than the set of plans generated by maps, but the former is compact. This makes one of the most important distinctions between these two problems, but an even more important feature of the Kantorovich problem is the linearity of the functional minimized in this problem, which makes this more general problem simpler. It is possible that for a long period of time this was the reason why there were no investigations of the problem, quite a natural one from the point of view of applications, of minimizing the integral of a cost function which depends also on the transport plan, that is, of minimizing the nonlinear functional

$$J_h(\sigma) = \int_{X \times Y} h(x, y, \sigma) \sigma(dx dy) \quad (1.1)$$

with cost function h defined on $X \times Y \times \mathcal{P}(X \times Y)$, where $\mathcal{P}(X \times Y)$ is the space of probability measures on $X \times Y$. In the classical problem $h(x, y)$ has the meaning of the cost of transportation of a unit mass from x to y and does not depend on the way of transportation σ . It is clear that in practice it is quite natural to expect that such a dependence can exist. This complicates substantially the search for optimal transportations but, surprisingly, does not complicate the proof of the existence of a minimum under the standard assumptions about h . In the first papers [80], [4], [15], [2], and [16] on the described nonlinear problem the function h had a more special form:

$$h(x, y, \sigma) = h(x, \sigma^x),$$

where the σ^x are the conditional measures on Y representing the plan σ in the form

$$\sigma(dx dy) = \sigma^x(dy) \mu(dx),$$

that is,

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

for all bounded measurable functions f on $X \times Y$. Then the functional takes the form

$$J_h(\sigma) = \int_X h(x, \sigma^x) \mu(dx). \quad (1.2)$$

Formally, this is a particular case of (1.1) but, actually, the functional becomes more singular because of a possible discontinuity in x of the conditional measure.

A nonlinear problem was also considered on the set of plans $\Pi(\mu, \nu)$ with fixed projections, but it led to one more interesting setting of the Kantorovich problem, which turned out to be also new in the linear case. This setting arose in the case where the cost function is not defined on $X \times Y$, but rather on $X \times \mathcal{P}(Y)$, where $\mathcal{P}(Y)$ is the space of probability measures on Y . Of course, in the usual problem one can take an arbitrary space, including $\mathcal{P}(Y)$, for the second space, but the novelty of the problem is that, in place of the projection of the plan onto the second factor, now we are given the barycentre of this projection. The plan itself is a measure on $X \times \mathcal{P}(Y)$, its projection is a measure on $\mathcal{P}(Y)$, that is, a measure on the space of measures, and its barycentre is a measure on Y . Remark 3.5 describes a more general formulation of the transportation problem which covers the case of fixed projections, as well as the case of fixed barycentres. In this formulation the restriction on the plan is that we are given the images of the plan under some maps $\Psi_1: X \times Y \rightarrow E_1$ and $\Psi_2: X \times Y \rightarrow E_2$. The classical problem corresponds to the projections onto factors.

Thus, so far we have mentioned some modifications caused by a more complicated form of the cost function and the replacement of conditions on projections by other conditions on plans. However, relatively recently McCann with coauthors [90]–[93] proposed a problem, a very interesting and natural one from the point of view of applications, in which in the classical linear situation an additional restriction is imposed on transport plans, namely, that only those plans are admissible that are absolutely continuous with respect to a fixed measure λ on $X \times Y$ (in their first papers this was Lebesgue measure on \mathbb{R}^n or a Riemannian manifold) such that the corresponding Radon–Nikodym density does not exceed a given function Φ on $X \times Y$. In this modification the most suitable topology on the space of measures turns out to be the weak topology from the space $L^1(\lambda)$. Investigations of this problem were continued in [64], [30], [37], and [44], and a survey on this was given in [29], so here we sum up briefly the results obtained by taking into account the recent results in [37], where the problem with density constraints was combined with the modifications mentioned above. Thus, we consider nonlinear transportation problems of three types: with fixed marginals, with one fixed marginal and a fixed barycentre of the second marginal, and with constraints on the densities of transport plans. Some subtypes can also be distinguished here, when a nonlinear cost function depends on the plans through their conditional measures. In all these types of Kantorovich problem it is useful to consider parametric problems in which cost functions and marginals (or other objects) depend on a parameter. Parametric problems are discussed in a separate section, but a more detailed exposition can be found in [34]–[36], and [29]. Finally, we touch upon briefly some questions

connected with the topology of spaces of measures, since they have a direct relation to all types of problem we discuss. We present results on these questions from the recent papers [36] and [3], including estimates for the Hausdorff distances between sets of transport plans.

In § 2 we introduce our principal definitions and notation and also discuss general nonlinear Kantorovich problems of optimal transportation, in § 3 we consider linear Kantorovich problems of the classical form, in which the second factor is the space of probability measures on some space Y and, in place of the second fixed marginal, that is, in place of a measure on the space of measures $\mathcal{P}(Y)$, we are given the barycentre of the projection of the plan onto the second factor, which is a measure on Y . The subject of § 4 is the problem with conditional measures. In § 5 we give a brief overview on problems with density constraints (this subject was already presented in the paper [29], also dedicated to an anniversary of Kantorovich's birth). In § 6 new problems with many marginals with additional projections are discussed. Parametric problems are considered in § 7 (to avoid overlaps with [29] this topic is also presented very briefly), and § 8 contains some information about metrics and topologies on spaces of measures connected with Kantorovich problems.

2. Nonlinear Kantorovich problems

Principal versions of various problems of optimal transportation involve measures on topological spaces (although, as we will see below, there is also a version in terms of general spaces with measures). So we recall here the basic concepts and introduce the notation used below. A thorough exposition of these questions can be found in the books [25] and [26].

Let X be a topological space (throughout, we deal with completely regular or metric spaces). We let $\mathcal{B}(X)$ denote its Borel σ -algebra, which is the smallest σ -algebra containing all open sets. A non-negative Borel measure μ on X (that is, a measure on $\mathcal{B}(X)$) is said to be Radon if for every Borel set B in X and every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset B$ such that $\mu(B \setminus K_\varepsilon) < \varepsilon$. A signed Borel measure μ is said to be Radon if so is its total variation $|\mu|$, defined by $|\mu| = \mu^+ + \mu^-$, where μ^+ and μ^- are the positive and negative parts of the measure μ in the Jordan–Hahn decomposition $\mu = \mu^+ - \mu^-$. The total variation norm is defined by

$$\|\mu\| = |\mu|(X).$$

A family M of Borel measures on X is called uniformly tight if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that

$$|\mu|(X \setminus K_\varepsilon) < \varepsilon \quad \forall \mu \in M.$$

The space of all Radon signed measures on the space X is denoted by $\mathcal{M}(X)$, the subset of non-negative measures is denoted by $\mathcal{M}^+(X)$, and the subset of probability measures is denoted by $\mathcal{P}(X)$.

If X is a complete separable metric space, then all Borel measures on it are Radon. The same is true for Souslin spaces, that is, the images of complete separable metric spaces under continuous maps.

The image of a Borel measure μ under a Borel map f from the topological space X to another topological space Y (that is, a map with Borel preimages of

Borel sets) is defined as the Borel measure $\mu \circ f^{-1}$ on Y given by the equality

$$(\mu \circ f^{-1})(B) = \mu(f^{-1}(B)), \quad B \in \mathcal{B}(Y).$$

The space of measures $\mathcal{M}(X)$ is equipped with the weak topology by means of seminorms of the form

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

where f is a bounded continuous function on X . Weak convergence of measures is convergence of the integrals of such functions against these measures.

In the circle of questions under consideration an important role is played by Prohorov’s theorem, according to which a bounded in variation, uniformly tight set of measures in $\mathcal{M}(X)$ is contained in a weakly compact set; moreover, in the case of a complete separable metric space X the converse is also true (see [25] and [26]). A typical example of a weakly compact set is the set of plans $\Pi(\mu, \nu)$ with Radon marginals μ and ν . Here uniform tightness is obvious from the estimate

$$\begin{aligned} \sigma((X \times Y) \setminus (K \times S)) &\leq \sigma((X \times Y) \setminus (K \times Y)) + \sigma((X \times Y) \setminus (X \times S)) \\ &= \sigma((X \setminus K) \times Y) + \sigma(X \times (Y \setminus S)) \\ &= \mu(X \setminus K) + \nu(Y \setminus S) \end{aligned}$$

for all $\sigma \in \Pi(\mu, \nu)$. The right-hand side is estimated by ε if we take the compact sets $K \subset X$ and $S \subset Y$ such that $\mu(X \setminus K) + \nu(Y \setminus S) \leq \varepsilon$.

If (X, d) is a metric space, then we denote by $\text{Lip}_1(d)$ the set of 1-Lipschitz functions, which are functions f on X such that

$$|f(x) - f(y)| \leq d(x, y) \quad \forall x, y \in X.$$

The Kantorovich–Rubinshtein norm on the space $\mathcal{M}(X)$ is defined by the formula

$$\|\mu\|_{\text{KR}} = \sup \left\{ \int_X f d\mu : f \in \text{Lip}_1(d), |f| \leq 1 \right\}.$$

This norm gives rise to the Kantorovich–Rubinshtein metric

$$d_{\text{KR}}(\mu, \nu) = \|\mu - \nu\|_{\text{KR}}.$$

The Kantorovich–Rubinshtein metric generates the weak topology on the set of non-negative measures. However, on the whole space of measures the topology generated by the Kantorovich–Rubinshtein norm differs from the weak topology in non-trivial cases; moreover, these two topologies are uncomparable. Indeed, suppose that X contains an infinite Cauchy sequence $\{x_n\}$. Then the Kantorovich–Rubinshtein norm cannot be continuous in the weak topology, since in that case it would be estimated in terms of the sum of several seminorms of the form p_f , that is, of the sum of the absolute values of several linear functionals on the space $\mathcal{M}(X)$. In our situation this space is infinite-dimensional, hence the intersection of the kernels of a finite system of linear functionals is non-trivial, but the sum of the seminorms under consideration vanishes on it. Thus, it is not true that the weak topology is

stronger than the topology of the Kantorovich–Rubinshtein norm. On the other hand, the latter is not stronger than the weak topology. This is seen from the fact that the sequence of measures $d(x_n, x_k)^{-1/2}(\delta_{x_n} - \delta_{x_k})$, where δ_x is the Dirac measure at the point x , converges to zero in the Kantorovich–Rubinshtein norm in view of the readily verified equality

$$d_{\text{KR}}(\delta_a, \delta_b) = d(a, b)$$

for $d(a, b) \leq 1$. However, this sequence of measures cannot converge weakly, since it is not bounded in variation, while any weakly convergent sequence of measures must be bounded in variation, which follows from the Banach–Steinhaus theorem and the fact that the norm $\|\mu\|$ coincides with the supremum of the integrals against the measure μ of the continuous functions not exceeding 1 in absolute value.

On the subspace $\mathcal{M}^1(X)$ of all measures μ such that for some (and therefore every) $x_0 \in X$ the function $d(x, x_0)$ is integrable with respect to the total variation of μ , we can define the Kantorovich norm

$$\|\mu\|_{\text{K}} = \sup \left\{ \int_X f d\mu : f \in \text{Lip}_1(d), f(x_0) = 0 \right\} + |\mu(X)|,$$

which generates the Kantorovich metric

$$d_{\text{K}}(\mu, \nu) = \|\mu - \nu\|_{\text{K}}.$$

If X is bounded, then these norms are equivalent, and if the diameter of X is not greater than 1, then the Kantorovich–Rubinshtein and Kantorovich metrics coincide on the set of probability measures. Analogues of the Kantorovich–Rubinshtein and Kantorovich norms and metrics on spaces of measures on general completely regular spaces are discussed in § 8.

Given $p \geq 1$, the set $\mathcal{P}^p(X) \subset \mathcal{P}(X)$ of all measures with respect to which the function $x \mapsto d(x, x_0)^p$ is integrable is equipped with the p -Kantorovich metric W_p defined by

$$W_p^p(\mu, \nu) = \inf_{\sigma \in \Pi(\mu, \nu)} \int_{X^2} d(x, y)^p \sigma(dx dy).$$

It was an important observation of Kantorovich that the Kantorovich distance between two probability measures μ and ν coincides with the infimum in the transport problem with marginals μ and ν and cost function equal to the metric, that is, with the minimum of the integrals of the metric against the measures in $\Pi(\mu, \nu)$. Subsequently, this equality, called Kantorovich’s duality formula, was extended to a very general situation of lower semicontinuous cost functions h on the product of completely regular spaces X and Y . Here for measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ the quantity $K_h(\mu, \nu)$ coincides with the supremum of the sums

$$\int_X f d\mu + \int_Y g d\nu,$$

taken over the bounded continuous functions $f: X \rightarrow \mathbb{R}$ and $g: Y \rightarrow \mathbb{R}$ satisfying the condition

$$f(x) + g(y) \leq h(x, y) \quad \forall x \in X, y \in Y.$$

Of course, in place of the sum of f and g we can take their difference, which better shows the connection with the case of the metric $h = d$, where $X = Y$ and $f = g$, and the estimate on the function becomes the condition that f is 1-Lipschitz: $f(x) - f(y) \leq d(x, y)$. Some analogues of the duality formula also appear in the modifications of the Kantorovich problem that we discuss.

The existence of a minimum in the nonlinear Kantorovich problem with fixed marginals or with one fixed marginal and a fixed barycentre is proved very similarly to the case of the linear problem, on the basis of the following readily verifiable fact. However, the problem with conditional measures is an exception: it is not covered by this approach and is considered separately.

Proposition 2.1. *Let X be a completely regular space, Π be a uniformly tight compact subset of the space $\mathcal{P}(X)$ with weak topology, and let $h: X \times \Pi \rightarrow [0, +\infty)$ be a lower semicontinuous function on all sets of the form $K \times \Pi$, where K is compact in X . Then the function*

$$J_h: \Pi \rightarrow [0, +\infty], \quad J_h(\sigma) = \int_X h(x, \sigma) \sigma(dx)$$

is lower semicontinuous. If h is bounded and continuous on the whole of $X \times \mathcal{P}(X)$, then J_h is continuous on $\mathcal{P}(X)$, and if h is lower semicontinuous on $X \times \mathcal{P}(X)$, then J_h is also lower semicontinuous.

Proof. The quantities $J_{\min(h, n)}(\sigma)$ increase to $J_h(\sigma)$ as $n \rightarrow \infty$. Hence the assertion reduces to the case of a bounded function h . We can assume that $h \leq 1$.

First assume that the function h is lower semicontinuous on the whole of $X \times \Pi$. Suppose that a net of measures 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 . Hence the products $\sigma_\alpha \otimes \delta_{\sigma_\alpha}$ on $\Pi \times \mathcal{P}(\Pi)$ converge weakly to the product $\sigma \otimes \delta_\sigma$ (see [26], Theorem 4.3.18). Therefore, by the lower semicontinuity of h we have (see [25], Corollary 8.2.5, or [26], 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 J_h .

Now we turn to 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 [65], 1.7.15(c)) that one can find a family of continuous functions $h_\alpha \geq 0$ on $K \times \Pi$ for which

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

Each 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 whole of $X \times \Pi$ and coincides with h

on $K \times \Pi$; moreover, the corresponding function J_g is also lower semicontinuous as shown above. It remains to observe that

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

because $g = h$ on $K \times \Pi$ and the integrals of the functions $h(x, \sigma)$ and $g(x, \sigma)$ over the complement of K against every measure $\sigma \in \Pi$ are not greater than ε . Thus, the function J_h is uniformly approximated by lower semicontinuous functions, and therefore it also possesses this property.

The last assertion of the proposition is clear from the reasoning above. \square

The additional condition of uniform tightness of the weakly compact set Π is automatically fulfilled for complete separable metrizable spaces, but need not hold for Souslin spaces (for instance, it can be violated even for the set of rational numbers: see the result of Preiss in [26], Theorem 4.8.6). So it is interesting to clarify whether or not it can be dropped in this proposition. One can consider a bounded continuous cost function h , because any lower semicontinuous bounded function h is the limit of an increasing net of bounded continuous functions h_α , hence

$$\int_X h_\alpha(x, \sigma) \sigma(dx) \uparrow \int_X h(x, \sigma) \sigma(dx)$$

for every measure $\sigma \in \Pi$ (see [25], Lemma 7.2.6). For an unbounded function h a further step is needed, involving the truncated functions $\min(h, n)$. The additional condition of uniform tightness is not needed if the function h is continuous and the continuity in the second variable at every point σ_0 is uniform with respect to the first variable, that is, for every $\varepsilon > 0$ there exists a neighborhood U of σ_0 such that

$$|h(x, \sigma) - h(x, \sigma_0)| < \varepsilon \quad \forall \sigma \in U, x \in X.$$

Under this condition, for every net of measures $\sigma_\alpha \in \Pi$ that converges weakly to a measure σ_0 there exists an index α_1 such that

$$\int_X |h(x, \sigma_\alpha) - h(x, \sigma_0)| \sigma_\alpha(dx) \leq \varepsilon \quad \forall \alpha \geq \alpha_1.$$

The same estimate also holds for σ_0 . Since the integrals of $h(x, \sigma_0)$ against the measures σ_α tend to its integral against σ_0 , we obtain that J_h is continuous at the point σ_0 .

Theorem 2.2. *Assume that the cost function h is 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 Proposition 2.1 the function J_h is lower semicontinuous on $\Pi(\mu, \nu)$. Now the existence of an optimal plan follows from the fact that any lower semicontinuous function on a compact set attains its minimum. \square

3. Problems with fixed barycentres

There is a general concept of the barycentre or mean of a Radon measure μ on a locally convex space X such that every continuous linear functional on X is integrable with respect to μ : this is a vector $b \in X$ such that

$$f(b) = \int_X f d\mu \quad \forall f \in X^*.$$

A barycentre exists if X is complete (or at least quasi-complete) and all continuous seminorms are integrable with respect to μ , that is, this measure has a strong first moment (see [41], Corollary 5.6.8). However, in problems of optimal transportation a more special situation arises, when we consider a Radon probability measure Q on the space $\mathcal{P}(E)$ of Radon probability measures on a completely regular topological space E , where the space of measures is equipped with the weak topology. Here the barycentre of Q is the Borel measure β_Q on E given by

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

where the vector integral is understood in the sense of the equality

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

for all Borel sets $A \subset E$. It is known that the function $p \mapsto p(A)$ is Borel on $\mathcal{P}(E)$ and the measure obtained is τ -additive (see [25], Proposition 8.9.8 and Corollary 8.9.9). However, we are interested in Radon barycentres, so the question about conditions ensuring that the measure β_Q is Radon on E arises here.

Proposition 3.1. *The measure β_Q is Radon precisely when the measure Q is concentrated on a countable union of uniformly tight compact sets in $\mathcal{P}(E)$. In particular, this is true if E is a Souslin completely regular space.*

Proof. Assume that there are increasing uniformly tight compact sets S_n in $\mathcal{P}(E)$ for which $Q(E \setminus S_n) \rightarrow 0$. Let $\varepsilon > 0$. Fix n such that

$$Q(E \setminus S_n) \leq \varepsilon.$$

By uniform tightness there exists a compact set $K \subset E$ for which

$$p(K) \geq 1 - \varepsilon \quad \forall p \in S_n.$$

Hence

$$\beta_Q(K) = \int_{\mathcal{P}(E)} p(K) Q(dp) \geq \int_{S_n} p(K) Q(dp) \geq (1 - \varepsilon)^2,$$

so the measure β_Q is tight. Since it is τ -additive, it is Radon (see [25], Proposition 7.2.2).

Conversely, suppose that the measure β_Q is Radon. Then for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset E$ such that $\beta_Q(K_\varepsilon) \geq 1 - \varepsilon^2$, that is,

$$\int_{\mathcal{P}(E)} p(K_\varepsilon) Q(dp) \geq 1 - \varepsilon^2.$$

The set of measures

$$S_\varepsilon := \{p \in \mathcal{P}(E) : p(K_\varepsilon) \geq 1 - \varepsilon\} \quad (3.1)$$

is closed in the weak topology on $\mathcal{P}(E)$. Indeed, if a net of measures p_α in S_ε converges weakly to a measure $p \in \mathcal{P}(E)$, then by the criterion of weak convergence due to A. D. Aleksandrov (see [25], Theorem 8.2.3) the following inequality holds:

$$p(K_\varepsilon) \geq \limsup_\alpha p_\alpha(K_\varepsilon) \geq 1 - \varepsilon.$$

For this set we obtain the estimate

$$Q(S_\varepsilon) \geq 1 - \varepsilon,$$

since by Chebyshev's inequality

$$Q(p : 1 - p(K_\varepsilon) \geq \varepsilon) \leq \varepsilon^{-1} \int_{\mathcal{P}(E)} [1 - p(K_\varepsilon)] Q(dp) \leq \varepsilon^{-1} \varepsilon^2 = \varepsilon.$$

Now, for a fixed number $\delta \in (0, 1)$ we can take sets $S_{\delta 2^{-n}}$ such that their intersection

$$\Pi_\delta = \bigcap_{n=1}^{\infty} S_{\delta 2^{-n}}$$

is also closed in $\mathcal{P}(E)$. For this intersection the inequality

$$Q(\Pi_\delta) \geq 1 - \delta$$

is true, because $Q(\mathcal{P}(E) \setminus S_{\delta 2^{-n}}) \leq \delta 2^{-n}$ for all n . By construction and (3.1) the set Π_δ is uniformly tight. By Prohorov's theorem (see [25], Theorem 8.6.7) it is weakly compact. Thus, the measure Q is concentrated on the union of the uniformly tight weakly compact sets $\Pi_{1/n}$. \square

If the space E is Souslin, then the space of measures $\mathcal{P}(E)$ with weak topology is also Souslin, hence every Borel measure on $\mathcal{P}(E)$ is automatically Radon and is concentrated on a countable union of compact sets. Moreover, these compacta are metrizable (even if E itself is not). However, for rather simple spaces E (for example, the set of rational numbers) compacta in $\mathcal{P}(E)$ need not be uniformly tight. Nevertheless, every measure in $\mathcal{P}(\mathcal{P}(E))$ is concentrated on a countable union of uniformly tight compact sets (see [25], Theorem 8.10.6). This is also true in the more general case of a completely regular space E such that all τ -additive measures on E are Radon. It would be interesting to find an example of a Radon measure on the space of Radon probability measures that vanishes on all uniformly tight compact sets.

Since the sets Π_δ in the above proof were constructed on the basis of the sets K_ε selected in accordance with the values of the barycentre at them, the completely analogous reasoning proves the following assertion.

Proposition 3.2. *Suppose that a set of measures $M \subset \mathcal{P}(\mathcal{P}(Y))$ possesses uniformly tight barycentres in $\mathcal{P}(Y)$. Then this set is uniformly tight in $\mathcal{P}(\mathcal{P}(Y))$ and concentrated on a countable union of uniformly tight weakly compact sets in $\mathcal{P}(Y)$.*

Corollary 3.3. *Suppose that a set of measures $M \subset \mathcal{P}(X \times \mathcal{P}(Y))$ possesses uniformly tight projections onto X and uniformly tight barycentres of the projections onto $\mathcal{P}(Y)$. Then this set is uniformly tight and concentrated on a countable union of sets of the form $K \times S$, where K is a compact set in X and the set S in $\mathcal{P}(Y)$ is weakly compact and uniformly tight.*

In particular, this is true if these measures have equal projections onto X and their projections onto $\mathcal{P}(Y)$ have equal barycentres.

Proof. Set $Z = X \times \mathcal{P}(Y)$. We observe that for every measure P in $\mathcal{P}(Z)$ with projections P_1 and P_2 onto the factors and for any Borel sets $A \subset X$ and $B \subset \mathcal{P}(Y)$ the inequality

$$P(Z \setminus (A \times B)) \leq P(Z \setminus (A \times \mathcal{P}(Y))) + P(Z \setminus (X \times B)) = P_1(X \setminus A) + P_2(\mathcal{P}(Y) \setminus B)$$

holds. Hence it suffices to consider the projections onto $\mathcal{P}(Y)$ and apply the previous proposition. \square

Note that if P is a Radon measure on the product $X \times \mathcal{P}(Y)$, where X and Y are completely regular spaces, μ is its projection onto X , and there exist conditional measures P^x on $\mathcal{P}(Y)$ with respect to μ , then the barycentre $P_{\mathcal{P}}$ of the projection of the measure P onto $\mathcal{P}(Y)$ is given by the formula

$$\beta_{P_{\mathcal{P}}}(B) = \int_X \int_{\mathcal{P}(Y)} p(B) P^x(dp) \mu(dx).$$

Indeed, for every Borel set $B \subset Y$ we have

$$\begin{aligned} \int_X \int_{\mathcal{P}(Y)} p(B) P^x(dp) \mu(dx) &= \int_{X \times \mathcal{P}(Y)} p(B) P(dx dp) \\ &= \int_{\mathcal{P}(Y)} p(B) P_{\mathcal{P}}(dp) = \beta_{P_{\mathcal{P}}}(B). \end{aligned}$$

We go over to setting the nonlinear Kantorovich transportation problem with fixed barycentre. As the space we take here the product $X \times \mathcal{P}(Y)$, where X and Y are completely regular spaces. On this product we are given a lower semicontinuous cost function

$$h: X \times \mathcal{P}(Y) \rightarrow [0, +\infty).$$

In addition, we are given a marginal $\mu \in \mathcal{P}(X)$, but in place of the second marginal we are given a barycentre $\beta \in \mathcal{P}(Y)$ of the projections of admissible plans onto $\mathcal{P}(Y)$; these projections are elements of $\mathcal{P}(\mathcal{P}(Y))$, so that the barycentre is understood in the sense explained above. Thus, on the set of plans

$$\Pi^\beta(\mu) := \{\pi \in \mathcal{P}(X \times \mathcal{P}(Y)): \pi_X = \mu, \beta_{\pi_{\mathcal{P}}} = \beta\},$$

where π_X is the projection of the measure π onto X , we consider the problem

$$\int_{X \times \mathcal{P}(Y)} h(x, p) \pi(dx dp) \rightarrow \min, \quad \pi \in \Pi^\beta(\mu). \tag{3.2}$$

The difference from the usual nonlinear Kantorovich problem is that the second marginal is not prescribed. Instead, we are given the barycentre of the projection onto the second factor.

We recall that in the general situation, given a function h on the product $X \times Z$, a set $\Gamma \subset X \times Z$ is called h -cyclically monotone if for all n the inequality

$$\sum_{i=1}^n h(x_i, z_i) \leq \sum_{i=1}^n h(x_{i+1}, z_i)$$

is true for all pairs $(x_1, z_1), \dots, (x_n, z_n) \in \Gamma$, where $x_{n+1} := x_1$.

It is known (see [20]) that in the classical problem, where X and Z are Souslin spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Z)$, and the Borel cost function h is such that there exists an optimal measure $\sigma \in \Pi(\mu, \nu)$, this measure is concentrated on some Borel h -cyclically monotone set. Applying this reasoning to $Z = \mathcal{P}(Y)$ we prove the following result.

Proposition 3.4. *Let h be a bounded lower semicontinuous function on $X \times \mathcal{P}(Y)$. Then for any measures $\mu \in \mathcal{P}(X)$ and $\beta \in \mathcal{P}(Y)$ Kantorovich problem (3.2) with a prescribed barycentre is solvable.*

Any optimal measure P for this problem is also optimal for the classical linear problem with the same cost function and marginals μ and $P_{\mathcal{P}}$, where $P_{\mathcal{P}}$ is the projection of P onto $\mathcal{P}(Y)$.

Finally, if X and Y are Souslin spaces, then the measure P is concentrated on an h -cyclically monotone set.

The proof can be found in [37]; it coincides with the standard proof for the linear problem.

In the Kantorovich problem for a triple $(\mu, P_{\mathcal{P}}, h)$ we can use the duality theorem, which shows that the minimum in problem (3.2) is equal to the quantity

$$\sup \left(\int_X f \, d\mu + \int_{\mathcal{P}(Y)} g \, dP_{\mathcal{P}} \right),$$

where the supremum is taken over the bounded continuous functions f on X and g on $\mathcal{P}(Y)$ satisfying the inequality

$$f(x) + g(p) \leq h(x, p), \quad x \in X, \quad p \in \mathcal{P}(Y).$$

Remark 3.5. The problem with a fixed marginal is a particular case of the following more general problem with fixed images of plans. Let

$$\Psi_1: \mathcal{P}(X \times Y) \rightarrow E_1 \quad \text{and} \quad \Psi_2: \mathcal{P}(X \times Y) \rightarrow E_2$$

be measurable maps to measurable spaces (E_1, \mathcal{E}_1) and (E_2, \mathcal{E}_2) . In addition, let two probability measures η_1 and η_2 on \mathcal{E}_1 and \mathcal{E}_2 , respectively, be given. For example, assume that we have completely regular spaces X, Y, E_1, E_2 and Borel maps Ψ_1 and Ψ_2 . Consider the set

$$\Pi_{\eta_1, \eta_2} = \{ \sigma \in \mathcal{P}(X \times Y) : \Psi_i(\sigma) = \eta_i, \quad i = 1, 2 \}.$$

Then we can formulate the problem of minimizing the functional J_h , that is, the integral of h , over the set Π_{η_1, η_2} . Here h can be a function on $X \times Y \times \mathcal{P}(X \times Y)$, or a function on $X \times Y \times \mathcal{P}(Y)$ when we consider the problem with conditional measures. If the maps Ψ_1 and Ψ_2 are continuous and the preimages of points under the map (Ψ_1, Ψ_2) are compact, while the cost function h is lower semicontinuous, then the set Π_{η_1, η_2} is compact in the weak topology. Hence the standard reasoning yields the existence of a minimum of the lower semicontinuous functional J_h on Π_{η_1, η_2} . The ordinary problem corresponds to the projections of measures onto X and Y . A problem with a fixed barycentre is obtained if as Ψ_1 we take the projecting of measures onto X , that is, $\Psi_1(\sigma) = \sigma_X$, in place of Y we take $\mathcal{P}(Y)$, and as Ψ_2 we take the map

$$\Psi_2(\sigma) = \beta_{\sigma_{\mathcal{P}(Y)}},$$

where $\sigma_{\mathcal{P}(Y)}$ is the projection of the measure σ onto $\mathcal{P}(Y)$. In place of fixing the first marginal we can state the transportation problem for a cost function on $\mathcal{P}(X) \times \mathcal{P}(Y)$ with given barycentres of the projections onto both factors. Of course, these settings are also meaningful in the case of many marginals. A problem with additional constraints on the densities of plans in Π_{η_1, η_2} also arises here. In this case, in place of the weak topology on the space of measures it is natural to consider the weak topology on L^1 with respect to the corresponding measure λ . Of course, the condition of lower semicontinuity should also refer to this topology. However, now some other conditions on the maps Ψ_i are required if we wish the set Π_{η_1, η_2} to be compact. Clearly, the transportation problem with fixed images of plans is also meaningful in the case of a larger number of maps Ψ_i for which the corresponding class of plans is not empty.

Note also that in the problem with a fixed barycentre the specific features of the space of measures, which is taken as the second factor, play an important role. If we take an abstract locally convex space E for the second factor, then, given a cost function h on $X \times E$, a fixed measure μ on X , and a fixed vector $\beta \in E$, one can introduce the set of Radon probability measures on $X \times E$ with projection μ onto X and with barycentre of the projection onto E equal to β . It is possible to minimize the integral of h over this set. In the particular case $E = \mathcal{P}(Y)$ under consideration this set is compact. However, in the general case there is no compactness because the set of Radon probability measures on E with barycentre β is not necessarily compact. For example, if $\beta = 0$, then the indicated set contains all measures $(\delta_a + \delta_{-a})/2$, where $a \in E$.

The Monge problem also has a modification with a fixed barycentre. We mention it in the next section in connection with nonlinear problems involving conditional measures.

4. Problems with conditional measures

In [80], [4], [15], [2], and [16] the existence of a minimum in a problem with conditional measures was proved under the additional assumption of the convexity of the cost function with respect to the measure-valued argument. The following generalization of these results to completely regular spaces was obtained in [38]. In place of Borel σ -algebras it employs the Baire σ -algebras $\mathcal{B}a(X)$ and $\mathcal{B}a(\mathcal{P}(Y))$,

which are generated by all continuous functions on the corresponding spaces. For a general completely regular space the Baire σ -algebra is smaller than the Borel one, but for Souslin completely regular spaces they coincide. In particular, if Y is a Souslin completely regular space, then $\mathcal{B}a(\mathcal{P}(Y)) = \mathcal{B}(\mathcal{P}(Y))$.

Theorem 4.1. *Assume that the cost function $H: X \times \mathcal{P}(Y) \rightarrow [0, +\infty)$ is measurable with respect to $\mathcal{B}a(X) \otimes \mathcal{B}a(\mathcal{P}(Y))$, lower semicontinuous on all sets of the form $K \times S$, where K is compact in X and $S \subset \mathcal{P}(Y)$ is uniformly tight, and convex in the second argument. Then the infimum*

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

is attained, that is, an optimal plan exists.

The proof reduces to the verification of the lower semicontinuity of the integral functional to be minimized. Since the sum of lower semicontinuous functions is also lower semicontinuous, we can combine the assertions of Theorem 2.2 and Theorem 4.1.

Corollary 4.2. *Consider a cost function of the form*

$$H(x, y, \sigma) = H_1(x, y, \sigma) + H_2(x, \sigma^x),$$

where the function $H_1: X \times Y \times \mathcal{P}(X \times Y) \rightarrow [0, +\infty)$ satisfies the hypotheses of Theorem 2.2 and the function $H_2: X \times \mathcal{P}(Y) \rightarrow [0, +\infty)$ satisfies the hypotheses of Theorem 4.1. Then the minimum is attained in the nonlinear Kantorovich problem with function H , that is, an optimal plan exists.

Note that according to [2], Theorem 3.9, for a continuous bounded cost function $H(x, p)$ on $X \times \mathcal{P}(Y)$, where X and Y are complete separable metric space, the infimum in the nonlinear problem with conditional measures and an atomless marginal $\mu \in \mathcal{P}(X)$ equals the minimum in the same problem with cost function $H^{**}(x, p)$ defined as the maximum function majorized by $H(x, p)$, among the functions that are convex in the second argument and lower semicontinuous.

The case of a cost function of the form $H(x, y, \sigma^x)$, defining the functional

$$\int_{X \times Y} H(x, y, \sigma^x) \sigma(dx dy),$$

has not been studied yet.

There are examples (see [4], Examples 3.2 and 3.3) when there exists no minimum in a problem with conditional measures. In [37] examples of this kind have some additional properties, in particular, both marginal distributions coincide with Lebesgue measure on an interval and the cost function is Lipschitz. Let us describe here these examples, referring to [37] for justifications, which are not very short.

Example 4.3. Let $X = Y = [0, 1]$ and let $\mu = \nu = \lambda$ be Lebesgue measure on $[0, 1]$. Then there exists a bounded Lipschitz function h on $X \times \mathcal{P}(Y)$ (where $\mathcal{P}(Y)$ is equipped with the Kantorovich metric) for which the nonlinear problem with conditional measures

$$\int_X h(x, \sigma^x) \mu(dx) \rightarrow \inf, \quad \sigma \in \Pi(\mu, \nu), \quad \sigma(dx dy) = \sigma^x(dy) \mu(dx)$$

has no minimum. The function h is given by the formula

$$h(x, p) = \min(\|p - \nu_x^1\|_{\mathbb{K}}, \|p - \nu_x^2\|_{\mathbb{K}}),$$

where for every $x \in [0, 1]$ two probability measures ν_x^1 and ν_x^2 on $[0, 1]$ are defined by

$$\nu_x^1(dy) = 2I_{[0, (1+x)/4] \cup [(3+x)/4, 1]} dy, \quad \nu_x^2(dy) = 2I_{[(1+x)/4, (3+x)/4]} dy.$$

This function is 1-Lipschitz in every variable separately, hence it is Lipschitz on the product space. Here the set $\Pi(\mu, \nu)$ consists of all probability measures on the square whose projections are equal to Lebesgue measure. It includes the set of measures given by biprobability densities with respect to Lebesgue measure on $[0, 1]^2$ (that is, densities that are probability densities in every variable separately).

The next example from [37] is interesting in that the cost function splits into a product of functions of one variable.

Example 4.4. As above, let $X = Y = [0, 1]$, and let $\mu = \nu = \lambda$ be Lebesgue measure on the interval $[0, 1]$. There is a bounded continuous function $g: \mathcal{P}(Y) \rightarrow \mathbb{R}$, where $\mathcal{P}(Y)$ is equipped with the weak topology, such that there is no minimum in the nonlinear Kantorovich problem

$$J(\sigma) = \int_0^1 \sqrt{1 + 2x} g(\sigma^x) dx \rightarrow \inf, \quad \sigma \in \Pi(\mu, \nu).$$

Setting $f(x) = \sqrt{1 + 2x}/2$, one can take the following function for g :

$$g(p) = \min(\min\{g_0(t) + M\|p - \nu_t^1\|_{\mathbb{K}} : t \in [f(0), f(1)]\}, \min\{g_0(t) + M\|p - \nu_t^2\|_{\mathbb{K}} : t \in [f(0), f(1)]\}),$$

where

$$\begin{aligned} g_0(t) &= 1 - t, \quad \nu_t^1 = \zeta_t + \eta_{f^{-1}(t)}^1, \quad \nu_t^2 = \zeta_t + \eta_{f^{-1}(t)}^2, \\ \zeta_t &= t^2 \cdot 2I_{[1/2, 3/4]} dy + (1 - t^2) \cdot 2I_{[3/4, 1]} dy, \\ \eta_s^1 &= 2I_{[0, (1+s)/8] \cup [(3+s)/8, 1/2]} dy, \quad \eta_s^2 = 2I_{[(1+s)/8, (3+s)/8]} dy, \quad s \in [0, 1], \end{aligned}$$

and the number M is sufficiently large.

Now we discuss an interesting modification of the Monge problem with a fixed barycentre. This modification arises in the case where as the second space we take the space of Radon probability measures $\mathcal{P}(Y)$ on a Souslin space Y and the cost function h is defined on $X \times \mathcal{P}(Y)$, where X is also a Souslin space. Also let a Radon probability measure β be given on Y . Then for the triple (μ, β, h) the Monge problem on $X \times \mathcal{P}(Y)$ with fixed barycentre β is stated as follows:

$$\int_X h(x, T(x)) \mu(dx) \rightarrow \inf, \quad T: X \rightarrow \mathcal{P}(Y), \quad \int_{\mathcal{P}(Y)} p \mu \circ T^{-1}(dp) = \beta. \quad (4.1)$$

Thus, we minimize the same integral as in the classical Monge problem, but now the minimum is taken over the measurable maps T from X to $\mathcal{P}(Y)$ such that the

barycentre of the image measure $\mu \circ T^{-1}$ is β , but the image itself is not fixed. The last equality can be written as

$$\int_X T(x) \mu(dx) = \beta.$$

The Monge problem with fixed barycentre is connected in an interesting way with the nonlinear Kantorovich problem with conditional measures. As we saw in the previous section, the above Kantorovich problem can fail to have solutions. Nevertheless, it turns out that if the nonlinear Kantorovich problem

$$\int_X h(x, \sigma^x) \mu(dx) \rightarrow \inf, \quad \sigma \in \Pi(\mu, \beta), \quad \sigma(dx dy) = \sigma^x(dy) \mu(dx), \quad (4.2)$$

with conditional measures and fixed marginals μ and β has a solution, then there also exists a solution to the Monge problem for the triple (μ, β, h) with fixed barycentre β . In order to obtain a solution of this problem from a solution σ of the Kantorovich problem, we make the following observation. Assuming first that σ is an arbitrary plan in $\Pi(\mu, \beta)$ with conditional measures σ^x on Y , we set

$$T: X \rightarrow \mathcal{P}(Y), \quad T(x) = \sigma^x.$$

The barycentre of the image measure $\mu \circ T^{-1}$ on $\mathcal{P}(Y)$ equals β , since for every set $B \in \mathcal{B}(Y)$ we have

$$\int_{\mathcal{P}(Y)} p(B) \mu \circ T^{-1}(dp) = \int_X \sigma^x(B) \mu(dx) = \sigma(X \times B) = \beta(B).$$

In addition,

$$\int_X h(x, T(x)) \mu(dx) = \int_X h(x, \sigma^x) \mu(dx) = \int_{X \times Y} h(x, \sigma^x) \sigma(dx dy) \geq K_h(\mu, \beta),$$

and for an optimal plan (if it exists) one has equality. It follows from this equality that there are no measurable transformations $F: X \rightarrow \mathcal{P}(Y)$ of the measure μ into a measure in $\mathcal{P}(\mathcal{P}(Y))$ with barycentre β and a smaller integral of $h(x, F(x))$ with respect to μ than that for the map generated by an optimal plan. Indeed, given such a transformation F , we can take the measure

$$\eta(dx dy) := \eta^x(dy) \mu(dx), \quad \eta^x = F(x).$$

Then for every set $B \in \mathcal{B}(Y)$ we obtain

$$\int_X \eta^x(B) \mu(dx) = \int_X F(x)(B) \mu(dx) = \int_{\mathcal{P}(Y)} p(B) \mu \circ F^{-1}(dp) = \beta(B).$$

This means that the projection of the measure η onto Y equals β : the value $\eta(X \times B)$ equals the left-hand side of the previous equality. Therefore,

$$\int_X h(x, F(x)) \mu(dx) = \int_X h(x, \eta^x) \mu(dx) = J_h(\eta) \geq J_h(\sigma).$$

Thus, any transformation F of the measure μ into the measure with barycentre β generates a plan η in $\Pi(\mu, \beta)$ for which the integral of the function $h(x, \eta^x)$ against μ equals the integral of $h(x, F(x))$.

Hence, if there is a minimizing map $T: X \rightarrow \mathcal{P}(Y)$ in our modified Monge problem, then the measure

$$\eta(dx dy) = \eta^x(dy) \mu(dx), \quad \eta^x = T(x),$$

belongs to $\Pi(\mu, \beta)$ and is minimizing in the Kantorovich problem.

Thus we arrive at the following assertion.

Theorem 4.5. *Let X and Y be Souslin spaces and let two measures $\mu \in \mathcal{P}(X)$ and $\beta \in \mathcal{P}(Y)$ and a Borel function*

$$h: X \times \mathcal{P}(Y) \rightarrow [0, +\infty)$$

be fixed. Then for every plan $\sigma \in \Pi(\mu, \beta)$ there exists a Borel map

$$T: X \rightarrow \mathcal{P}(Y)$$

for which the measure $\mu \circ T^{-1}$ has barycentre β and

$$\int_X h(x, \sigma^x) \mu(dx) = \int_X h(x, T(x)) \mu(dx).$$

Conversely, for every Borel map $T: X \rightarrow \mathcal{P}(Y)$ with $\beta_{\mu \circ T^{-1}} = \beta$ there exists a plan $\sigma \in \Pi(\mu, \beta)$ satisfying the above equality.

Thus, the Kantorovich infimum $K_h(\mu, \beta)$ equals the infimum in the Monge problem with fixed barycentre for the triple (μ, β, h) , and the existence of a solution in one of these two problems is equivalent to the solvability of the other.

We see from the above reasoning that this assertion also remains valid in the more general case where every measure σ in $\Pi(\mu, \beta)$ possesses conditional measures on Y which depend Borel measurably on x .

Now assume that we consider the usual Monge problem with measures μ and ν on completely regular Souslin spaces X and Y and with Borel cost function h on $X \times Y$. The space Y is canonically embedded in the space of probability measures $\mathcal{P}(Y)$ by the map $y \mapsto \delta_y$, where δ_y is the Dirac measure at the point y . The image of Y under this embedding is closed in $\mathcal{P}(Y)$: see [25], Lemma 8.9.2. Let us extend the function h from $X \times Y$ to a Borel function on $X \times \mathcal{P}(Y)$ such that if the original function is continuous or lower semicontinuous, then the extension has the same property. This is possible, since completely regular Souslin spaces are perfectly normal (see [25], Theorem 6.7.7) and the space of measures $\mathcal{P}(Y)$ is also Souslin. To extend a continuous function h we can use Urysohn's theorem and in the case of a lower semicontinuous function h we employ the fact that there exist continuous functions h_α on $X \times Y$ such that $h(x, y) = \sup_\alpha h_\alpha(x, y)$ for all $x \in X, y \in Y$, which enables us to find continuous extensions H_α of these functions to $X \times \mathcal{P}(Y)$ and obtain the lower semicontinuous function $H = \sup_\alpha H_\alpha$ on the space $X \times \mathcal{P}(Y)$. This function coincides with the original function on $X \times Y$. If Monge problem (4.1) on $X \times \mathcal{P}(Y)$ with fixed barycentre ν possesses a minimizing map $T: X \rightarrow \mathcal{P}(Y)$, then

we can call it a relaxed solution of the original Monge problem, and the new problem can be called a relaxation of the original one. Even when the original problem has a solution, the minimum in it need not be the minimum in the relaxed problem. This occurs if $h(x, y) = 1$, but ν is not a Dirac measure, and the extension H of h satisfies $H(x, \nu) = 0$. Then the minimum in the relaxed problem is zero, and the optimal map is identically equal to ν . Hence some restrictions on extensions are required in order to make the relaxed problem meaningful. The infimum in it is not greater than the infimum in the original problem, since if a map $S: X \rightarrow Y$ takes μ to ν , then for the map $T: x \mapsto \delta_{S(x)}$ the barycentre of the measure $\mu \circ T^{-1}$ equals ν , because for every Borel set $B \subset Y$ the relation $\nu = \mu \circ S^{-1}$ yields the equality

$$\int_{\mathcal{P}(Y)} p(B) \mu \circ T^{-1}(dp) = \int_X \delta_{S(x)}(B) \mu(dx) = \int_Y \delta_y(B) \nu(dy) = \nu(B).$$

In addition, $H(x, \delta_{S(x)}) = h(x, S(x))$. Hence the integral of $H(x, T(x))$ against μ coincides with the integral of $h(x, S(x))$. As we know from the above discussion, Monge problem (4.1) need not have a solution even for a continuous cost function (the corresponding Kantorovich problem with conditional measures can fail to have a minimum). So it is of interest to have additional conditions under which there exists a relaxed solution. One such condition is the convexity on $X \times \mathcal{P}(Y)$ of the function H obtained as an extension of the original cost function from $X \times Y$ to $X \times \mathcal{P}(Y)$. There exists an extension that is linear in the second argument. It is given by the explicit formula

$$H(x, p) = \int_Y h(x, y) p(dy).$$

For this extension the minimum in the relaxed Monge problem equals the minimum in the classical Kantorovich problem with function h and marginals μ and ν . If σ is an optimal Kantorovich plan and $\sigma(dx dy) = \sigma^x(dy) \mu(dx)$, then the map $T(x) = \sigma^x$ is optimal in the relaxed Monge problem, since the barycentre of the measure $\mu \circ T^{-1}$ equals ν , which is readily verified, and the integral of $H(x, T(x)) = H(x, \sigma^x)$ against μ equals

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

On the other hand, if $T: X \rightarrow \mathcal{P}(Y)$ is an optimal map in the relaxed Monge problem, then the measure $\sigma(dx dy) = \sigma^x(dy) \mu(dx)$, $\sigma^x = T(x)$, is optimal in the Kantorovich problem for the function h and marginals μ and ν . Indeed, its projection onto Y equals ν and the integral of h against this measure is as above equal to the integral of $H(x, T(x))$ against μ . Other continuous functions H can exist on $X \times \mathcal{P}(Y)$ that are convex in the second argument and for which

$$H(x, \delta_y) = h(x, y) \quad \forall x \in X, y \in Y.$$

The question about the connection between the minimum in the relaxed Monge problem and the infimum in the original one arises for such functions.

Nonlinear problems were also considered in the recent paper [17].

5. Optimal transportation with density constraints

A further new modification of the Kantorovich problem mentioned above was introduced in [90]–[93]; it is connected with constraints on the densities of transport plans. First we formulate it in the case of two marginals. Assume that, as in the classical Kantorovich problem, we are given probability spaces (X, \mathcal{B}_X, μ) and (Y, \mathcal{B}_Y, ν) and also a $\mathcal{B}_X \otimes \mathcal{B}_Y$ -measurable cost function $h \geq 0$. Assume in addition that we are given a probability measure λ on the σ -algebra $\mathcal{B}_X \otimes \mathcal{B}_Y$ of the space $X \times Y$ and a function Φ that is integrable with respect to λ . For example, we can take the product $\mu \otimes \nu$ as λ .

Consider the class $\Pi_\Phi(\mu, \nu)$ of all probability measures σ on $\mathcal{B}_X \otimes \mathcal{B}_Y$ that belong to $\Pi(\mu, \nu)$, are absolutely continuous with respect to λ , and have a Radon–Nikodym density satisfying the estimate

$$\frac{d\sigma}{d\lambda} \leq \Phi.$$

We assume that $\Pi_\Phi(\mu, \nu)$ is not empty. A necessary condition for this is the absolute continuity of the measures μ and ν with respect to the projections of the measure λ onto X and Y , respectively. For example, if $\lambda = \mu \otimes \nu$, then as β we can take a constant $C \geq 1$, but not $C < 1$. In the case when $C = 1$ the set $\Pi_\Phi(\mu, \nu)$ consists of a unique measure λ .

Consider the problem

$$\int_{X \times Y} h d\sigma \rightarrow \min, \quad \sigma \in \Pi_\Phi(\mu, \nu).$$

A sufficient condition for the existence of a minimum in this problem turns out to be substantially simpler and broader than in the classical problems of Kantorovich and Monge.

Theorem 5.1. *If there exists a measure $\sigma \in \Pi_\Phi(\mu, \nu)$ with finite integral of the function h , then a minimum is attained in the above problem.*

Proof. By our assumptions, for some $N > 0$ we have a nonempty set $\Pi_{\Phi, N}(\mu, \nu)$ of measures in $\Pi_\Phi(\mu, \nu)$ such that the integral of h against them does not exceed N . This set of measures can be identified with the set of their densities with respect to the measure λ . Such densities p satisfy $tp \leq \Phi$, and the integral of hp against λ does not exceed N . Hence the set of densities under consideration is weakly compact in $L^1(\lambda)$; see [25], Theorem 4.7.18. The functional

$$p \mapsto \int_{X \times Y} hp d\lambda$$

is lower semicontinuous on $\Pi_{\Phi, N}(\mu, \nu)$ with weak topology, since in the case of a bounded function h it is continuous and in the general case it is the pointwise limit of the sequence of functionals increasing on $\Pi_{\Phi, N}(\mu, \nu)$ and generated by the bounded functions $\min(h, n)$. Therefore, this functional attains its minimum on our compact set. \square

Now we can introduce a nonlinear version of the problem with constraints on the densities of plans. Here we need some additional properties of topological nature for the cost function.

Let \mathcal{P}_λ be the set of all probability densities in $L^1(\lambda)$. It has two natural topologies, the norm topology and the weak topology of the Banach space $L^1(\lambda)$. We denote by $\mathcal{B}(\mathcal{P}_\lambda)$ the Borel σ -algebra with respect to the norm. If $L^1(\lambda)$ is separable (which holds for Borel measures on Souslin spaces), then $\mathcal{B}(\mathcal{P}_\lambda)$ coincides with the Borel σ -algebra with respect to the weak topology.

For problems with density constraints it is useful to note straight away that if a set $M \subset \mathcal{P}(X)$ consists of measures which are absolutely continuous with respect to some probability measure λ_0 , then (after the identification of measures with their densities with respect to λ_0) the weak topology of the space $L^1(\lambda_0)$ arises on M , competing with the weak topology of the space of measures. The former is usually strictly stronger, since it is generated by the duality with the space of all bounded Borel functions on X in place of the space of all bounded continuous functions. However, the Borel structures induced on M by these two different topologies, and by the even stronger norm topology, coincide in the case where X is a Souslin space. This follows from the fact that any sequence of continuous functions separating points on a Souslin space generates its Borel σ -algebra (see [25], Theorem 6.8.9), while the set of probability densities in $L^1(\lambda_0)$ is a Souslin subset in the norm topology since it is closed and $L^1(\lambda_0)$ is separable. Note also that if the set of densities of measures in M is uniformly integrable (this is equivalent to the property that its closure in the weak topology in $L^1(\lambda_0)$ is weakly compact and is also equivalent to its uniform countable additivity; see [25], Chap. 4), then the weak topology of the space of measures on M coincides with the weak topology inherited from $L^1(\lambda_0)$. Indeed, every functional of the form

$$\mu \mapsto \int_X f d\mu,$$

where f is a bounded Borel function on X , is continuous on M with the weak topology of the space of measures, because for every $\varepsilon > 0$ one can find $\delta > 0$ such that $\mu(B) < \varepsilon$ for all $\mu \in M$ whenever $\lambda_0(B) < \delta$. Then by Luzin's theorem we can take a continuous function g on X such that

$$\sup_x |g(x)| \leq \sup_x |f(x)| \quad \text{and} \quad \lambda_0(x: f(x) \neq g(x)) < \delta.$$

Then for all $\mu \in M$ we obtain

$$\left| \int_X (f - g) d\mu \right| \leq 2\varepsilon \sup_x |f(x)|.$$

Let us present a result from the recent paper [37]. Suppose that we are given a function

$$h: X \times Y \times \mathcal{P}_\lambda \rightarrow [0, +\infty)$$

that is measurable with respect to $\mathcal{B}_X \otimes \mathcal{B}_Y \otimes \mathcal{B}(\mathcal{P}_\lambda)$ and such that for λ -almost all $(x, y) \in X \times Y$ the function

$$p \mapsto h(x, y, p)$$

is lower semicontinuous on $\Pi_\Phi(\mu, \nu)$ with respect to the norm of $L^1(\lambda)$.

On the set $\Pi_{\Phi}(\mu_1, \mu_2)$, which is embedded in $L^1(\lambda)$ by means of the identification of measures with their densities with respect to λ , we consider the functional

$$J_h(p) = \int_{X \times Y} h(x, y, p)p(x, y) \lambda(dx dy)$$

with values in $[0, +\infty]$.

Theorem 5.2. *Under the above assumptions, if the functional J_h is convex, then it attains its minimum on the set $\Pi_{\Phi}(\mu_1, \mu_2)$.*

The condition of lower semicontinuity of the function h in p with respect to the norm, which is used in this theorem, is much weaker than a similar condition with respect to the weak topology. However, if the function h itself (rather than the integral) is convex in p , then these conditions are equivalent. Nevertheless, the convexity of h in the last argument does not imply the convexity of J_h , so the hypotheses of the previous theorem can hardly be regarded as constructive. If the cost function has the form

$$h(x, y, p) = h(x, p^x)$$

for some nonnegative function h on $X \times \mathcal{P}(Y)$ and the conditional measures p^x for p with respect to μ , then the convexity of h in the last argument obviously implies the convexity of J_h .

In what follows we deal with Souslin completely regular spaces X and Y . Then, as above, every measure p on $X \times Y$ with projection μ onto X possesses conditional measures p^x on Y with respect to μ , that is,

$$p(dx dy) = p^x(dy) \mu(dx).$$

If $\lambda = \mu \otimes \nu$ and measures $p \in \Pi_{\Phi}(\mu, \nu)$ are identified with their densities $p(x, y)$ with respect to the measure λ , then

$$p^x = p(x, \cdot) \nu.$$

In other words, the conditional density corresponding to a fixed point x is merely $y \mapsto p(x, y)$.

Since we deal with conditional measures on $\mathcal{B}(Y)$, it is reasonable to equip the space $\mathcal{M}(Y)$ of all bounded measures on $\mathcal{B}(Y)$ with the σ -algebra $\mathcal{E}(\mathcal{M}(Y))$ generated by all functions $\nu \mapsto \nu(B)$, $B \in \mathcal{B}(Y)$. Note that, since Y is a completely regular Souslin space, this σ -algebra is countably generated. Indeed, there is a countable family of Borel sets separating points of Y and generating the Borel σ -algebra (see [25], Theorems 6.7.7 and 6.8.9). Hence there is also a countable algebra of sets \mathcal{A} with this property. Consider the σ -algebra \mathcal{E}_0 generated by the countable family of functions $\nu \mapsto \nu(A)$, $A \in \mathcal{A}$. Let \mathcal{B}_0 denote the class of all Borel sets $B \subset Y$ for which the function $\nu \mapsto \nu(B)$ is measurable with respect to \mathcal{E}_0 . This class contains the algebra \mathcal{A} and is monotone, that is, it contains the unions of increasing sequences and the intersections of decreasing sequences of its elements. By the classical monotone class theorem it contains the σ -algebra generated by

the algebra \mathcal{A} (see [25], Theorem 1.9.3), hence it coincides with the whole Borel σ -algebra.

In the next theorem we assume that the function h is measurable with respect to the σ -algebra $\mathcal{B}(X) \otimes \mathcal{E}(\mathcal{M}(Y))$. Then the function

$$x \mapsto h(x, p^x)$$

is Borel measurable if the map $x \mapsto p^x$ from X to $\mathcal{P}(Y)$ is $(\mathcal{B}(X), \mathcal{E}(\mathcal{M}(Y)))$ -measurable, because the map $x \mapsto (x, p^x)$ is measurable with respect to the pair of σ -algebras $\mathcal{B}(X)$ and $\mathcal{B}(X) \otimes \mathcal{E}(\mathcal{M}(Y))$.

Transport problems of this kind can be written in the form

$$\int_X h(x, p^x) \mu(dx) \rightarrow \min, \quad p \in \Pi_{\Phi}(\mu, \nu), \quad p(dx dy) = p^x(dy) \mu(dx). \quad (5.1)$$

The following theorem from [37] is analogous to the previous one, but is not a corollary to it, because the cost function depends on conditional measures rather than on the whole plan.

Theorem 5.3. *If for μ -almost all x the function $p \mapsto h(x, p)$ is lower semicontinuous with respect to the total variation norm on $\mathcal{M}(Y)$ and the function*

$$J_h(p) = \int_X h(x, p^x) \mu(dx)$$

is convex, then it attains its minimum on $\Pi_{\Phi}(\mu, \nu)$. In particular, this is true if the function h is convex in the last argument.

The following example of a nonlinear Kantorovich problem with conditional measures and constraints on the densities of plans, in which there is no solution, was constructed in [37].

Example 5.4. As in the above examples, let $X = Y = [0, 1]$ and let $\mu = \nu = \lambda$ be Lebesgue measure on $[0, 1]$. Then there exists a bounded continuous cost function $h: X \times L^1[0, 1] \rightarrow \mathbb{R}$ (the space $L^1[0, 1]$ is equipped with the weak topology) for which the nonlinear problem with constraints on the densities of plans

$$J_h(\varrho) = \int h(x, \varrho(x, \cdot)) dx \rightarrow \inf,$$

$$\varrho(x, y) \leq 4 \quad \forall x, y, \quad \int_0^1 \varrho(x, y) dy = 1 \quad \forall x, \quad \int_0^1 \varrho(x, y) dx = 1 \quad \forall y,$$

has no minimum. Let $\{q_n\}$ be the set of rational numbers in $[0, 1]$. Set

$$h(x, \varrho) = \min(h_1(x, \varrho), h_2(x, \varrho)),$$

where

$$h_i(x, p) = \sum_{n=1}^{\infty} \min\left(\left|\int_0^{q_n} (p(y) - \varrho_i(x, y)) dy\right|, 1\right) 2^{-n}, \quad i \in \{1, 2\},$$

$$\varrho_1(x, y) = 2I_{[0, (1+x)/4] \cup [(3+x)/4, 1]}(y) \quad \text{and} \quad \varrho_2(x, y) = 2I_{[(1+x)/4, (3+x)/4]}(y).$$

The continuity of the function h follows from the continuity of the functions

$$p \mapsto \int_0^{q_n} p(y) dy \quad \text{and} \quad x \mapsto \int_0^{q_n} \varrho_i(x, y) dy.$$

It was verified in [37] that there exists no biprobability density $\varrho(x, y)$ with the property

$$\int_0^1 h(x, \varrho(x, \cdot)) dx = 0,$$

but the infimum in this problem is zero.

6. Multimarginal and multistochastic problems

Transport problems with many marginals differ from the classical ones in that transport plans are given on a product of more than two (possibly, infinitely many) spaces the projections onto which are fixed. Multimarginal problems were studied by many authors; see the recent papers [19], [22], [51], [62], [67], [71], [74], [77], [81], [110], and [113]–[115], where the reader can find additional references. All the new problems mentioned above can be also set in this situation. However, in the last years so-called multistochastic transport problems have gained popularity, in which plans are given on products of many factors, but there are additional constraints that not only the projections onto separate factors are fixed, but also the ones onto the finite products of some of the factors. For example, for measures on three-dimensional space we are given not only the projections onto coordinate axes, but also the projections onto two-dimensional coordinate subspaces. Of course, in this case the set of admissible plans can be empty, so that the question about conditions ensuring that this set is non-empty arises.

Assume that we are given n completely regular spaces X_1, \dots, X_n and a natural number $k < n$. Let p and q be non-negative integers such that $q \leq p$. Denote by \mathcal{I}_{pq} the family of all subsets of the set $\{1, 2, \dots, p\}$ of size q , and let $\mathcal{I}_p = \bigcup_{q=0}^p \mathcal{I}_{pq}$ denote the family of all subsets of the collection $\{1, 2, \dots, p\}$.

For any $\alpha \in \mathcal{I}_n$ set $X_\alpha = \prod_{i \in \alpha} X_i$. Let $X = \prod_{i=1}^n X_i$. For every $\alpha \in \mathcal{I}_n$ we denote the projection map onto X_α by Pr_α . Assume that for every $\alpha \in \mathcal{I}_{nk}$ we have a Radon probability measure μ_α on X_α . Denote by $\Pi(\{\mu_\alpha\})$ the (possibly empty) set of Radon probability measures with the property $\mu \circ \text{Pr}_\alpha^{-1} = \mu_\alpha$ for all $\alpha \in \mathcal{I}_{nk}$. Measures in $\Pi(\{\mu_\alpha\})$ are called joining. The direct (n, k) -Monge–Kantorovich problem is stated as follows.

Definition 6.1. Fix a Borel cost function $h: X \rightarrow \mathbb{R}$. Then the (n, k) -Monge–Kantorovich problem consists in finding the quantity

$$\inf_{\pi \in \Pi(\{\mu_\alpha\})} \int_X h d\pi.$$

This (n, k) -Monge–Kantorovich problem (also called the multistochastic transport problem) was studied in [75] and [76]. The interest in it is motivated by the following particular case, which turns out to be rather typical for this problem. Recall that bitwise addition of two numbers $x, y \in [0, 1]$ is the following operation: if

$$x = \overline{0.x_1x_2\dots} \quad \text{and} \quad y = \overline{0.y_1y_2\dots}$$

are their binary representations, then the number $z = x \oplus y$ has the form

$$z = \overline{0.z_1z_2\dots}, \quad \text{where } z_i = x_i + y_i \in \mathbb{Z}_2.$$

The set

$$S = \{(x, y, z) : x \oplus y \oplus z = 0\}$$

is a self-similar fractal of dimension 2, known as the Sierpinski tetrahedron.

Theorem 6.2. *Let $n = 3, k = 2$, and $X_i = [0, 1], i = 1, 2, 3$, let $\mu_{xy} = \lambda_{xy}, \mu_{xz} = \lambda_{xz}$, and $\mu_{yz} = \lambda_{yz}$ be copies of two-dimensional Lebesgue measure on $[0, 1]^2$, and let $h(x, y, z) = xyz$. Let the space $\Pi(\{\mu_\alpha\})$ consist of the probability measures on $[0, 1]^3$ whose projections onto the coordinate hyperplanes are Lebesgue measures on $[0, 1]^2$. Then there exists a unique solution of the $(3, 2)$ -problem*

$$\inf_{\pi \in \Pi(\{\mu_\alpha\})} \int_{[0,1]^3} xyz \pi(dx dy dz).$$

It is concentrated on the Sierpinski tetrahedron S .

The idea of the proof of the fact that a measure on S is a solution indeed comes from the following observation. Let $T_1(x, y, z) = (1 - x, y, z)$. We define T_2 and T_3 similarly. For every measure $\pi \in \Pi(\lambda_{xy}, \lambda_{xz}, \lambda_{yz})$ we have the equality

$$\begin{aligned} K(\pi \circ T_1^{-1}) &= \int_{\mathbb{R}^3} xyz \pi \circ T_1^{-1}(dx dy dz) = \int_{\mathbb{R}^3} (1 - x)yz \pi(dx dy dz) \\ &= \int_{\mathbb{R}^3} yz d\lambda_{yz} - \int_{\mathbb{R}^3} xyz \pi(dx dy dz) = \frac{1}{4} - K(\pi). \end{aligned}$$

Therefore, the maps $T_1 \circ T_2, T_1 \circ T_3$, and $T_2 \circ T_3$ preserve the value of the functional $K(\pi)$. Since they preserve the set $\Pi(\lambda_{xy}, \lambda_{xz}, \lambda_{yz})$, there exists a solution π that is invariant with respect to the maps $T_1 \circ T_2, T_1 \circ T_3$, and $T_2 \circ T_3$. We partition $[0, 1]^3$ into two sets S_1 and S_2 , where

$$S_1 = \left[0, \frac{1}{2}\right]^3 \cup \left[\frac{1}{2}, 1\right]^2 \times \left[0, \frac{1}{2}\right] \cup \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]^2 \cup \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$$

and

$$S_2 = [0, 1]^3 \setminus S_1.$$

Observe that S_1 and S_2 are invariant with respect to the operators $T_1 \circ T_2, T_1 \circ T_3$, and $T_2 \circ T_3$, and the equality $S_2 = T_1(S_1) = T_2(S_1) = T_3(S_1)$ is fulfilled. Next, consider the measures $\hat{\pi} = \pi|_{S_2}$ and $\tilde{\pi} = \hat{\pi} \circ T_1^{-1}$. It follows from the symmetry of π that these measures have equal projections onto the coordinate planes O_{xy}, O_{xz} and O_{yz} . Now it is readily proved that

$$\int_{S_2} xyz \hat{\pi}(dx dy dz) \geq \int_{S_1} xyz \tilde{\pi}(dx dy dz).$$

This yields that the support of π is contained in S_1 (otherwise the functional K would attain a smaller value at the measure $\pi - \hat{\pi} + \tilde{\pi}$). Applying this reasoning

to every cube of the four constituting S_1 , we obtain that there exists a solution with support in the set $S(k)$ (where $S(1) = S_1$) that is the union of 4^k cubes of volume $1/8^k$. Passing to the limit as $k \rightarrow \infty$ we obtain that there exists a solution with support $S = \bigcap_{k=1}^{\infty} S(k)$. The set S is exactly the Sierpinski tetrahedron. Modifying the above reasoning we can prove the uniqueness of a solution.

The following duality theorem was proved in [75] (see a more general result in [76]).

Theorem 6.3. *Let X_1, \dots, X_n be compact metric spaces, and let $h \geq 0$ be a continuous cost function on X . Assume that the set $\Pi(\{\mu_\alpha\})$ is not empty. Then*

$$\min_{\pi \in \Pi(\{\mu_\alpha\})} \int h \, d\pi = \sup_{f \leq h} \sum_{\alpha \in \mathcal{I}_{nk}} \int_{X_\alpha} f_\alpha \, d\mu_\alpha,$$

where the supremum is taken over all functions $f_\alpha \in L^1(\mu_\alpha)$ and

$$f(x) = \sum_{\alpha \in \mathcal{I}_{nk}} f_\alpha(x_\alpha).$$

A solution to the problem dual to the problem in Theorem 6.2 is described in the next theorem. The uniqueness of this solution is an open question.

Theorem 6.4. *Let $\mu_{xy} = \lambda_{xy}$, $\mu_{xz} = \lambda_{xz}$, and $\mu_{yz} = \lambda_{yz}$ be copies of two-dimensional Lebesgue measure on $[0, 1]^2$ and let $h(x, y, z) = xyz$. Then the triple of functions $f(x, y)$, $f(x, z)$, $f(y, z)$, where*

$$f(x, y) = \int_0^x \int_0^y t \oplus s \, dt \, ds - \frac{1}{4} \int_0^x \int_0^x t \oplus s \, dt \, ds - \frac{1}{4} \int_0^y \int_0^y t \oplus s \, dt \, ds,$$

is a solution of the dual problem.

The solution of the dual problem given in Theorem 6.4 is connected with the solution π of the linear problem in the following way: the measure π is concentrated on the graph of the map $(x, y) \mapsto f_{xy}(x, y)$, that is, π -almost everywhere

$$z = f_{xy}(x, y); \tag{6.1}$$

moreover, f possesses a non-negative mixed derivative f_{xy} almost everywhere, but the derivatives f_{xx} and f_{yy} do not exist (in the classical sense).

Note that in the multistochastic problem the question of whether the set $\Pi(\{\mu_\alpha\})$ is non-empty is non-trivial. It is clear that for this the system of measures μ_α must have the obvious consistency property: for all $\alpha, \beta \in \mathcal{I}_{nk}$ the equality

$$\mu_\alpha \circ \text{Pr}_{\alpha \cap \beta}^{-1} = \mu_\beta \circ \text{Pr}_{\alpha \cap \beta}^{-1}$$

must be valid. This property is not sufficient for $\Pi(\{\mu_\alpha\})$ to be non-empty, but it is sufficient for the existence of a signed measure P with the property $P \circ \text{Pr}_\alpha^{-1} = \mu_\alpha$. There are ‘dual’ sufficient conditions for the non-emptiness of this set, which are difficult to verify explicitly. One constructive sufficient condition is given in the following theorem.

Theorem 6.5. *Let $\nu_i \in \mathcal{P}(X_i)$, $1 \leq i \leq n$. For any $\alpha \in \mathcal{I}_{nk}$ set $\nu_\alpha = \prod_{i \in \alpha} \nu_i$. Let μ_α be a consistent collection of probability measures on the spaces X_α . Assume that the measure μ_α is absolutely continuous with respect to ν_α for all $\alpha \in \mathcal{I}_{nk}$. Let p_α be the density of μ_α with respect to ν_α . Assume that there exist positive constants m and $M \geq m$ such that $m \leq p_\alpha \leq M$ ν_α -almost everywhere for all $\alpha \in \mathcal{I}_{nk}$. Then there exists a constant $\lambda_{nk} > 1$ such that if $M/m \leq \lambda_{nk}$, then the set $\Pi(\{\mu_\alpha\})$ is not empty.*

Example 6.6. Let $\nu_x, \nu_y,$ and ν_z be some probability measures on the one-dimensional axes $O_x, O_y,$ and O_z and let

$$\nu_{xy} = \nu_x \otimes \nu_y, \quad \nu_{xz} = \nu_x \otimes \nu_z, \quad \text{and} \quad \nu_{yz} = \nu_y \otimes \nu_z.$$

Assume that the measures $\mu_{xy}, \mu_{xz},$ and μ_{yz} satisfy the consistency condition, the conditions

$$\mu_{xy} = p_{xy}\nu_{xy}, \quad \mu_{xz} = p_{xz}\nu_{xz}, \quad \text{and} \quad \mu_{yz} = p_{yz}\nu_{yz}$$

and the condition

$$1 \leq p_{xy}, p_{xz}, p_{yz} \leq c.$$

Then the set $\Pi(\mu_{xy}, \mu_{yz}, \mu_{xz})$ is not empty. In particular, if $c = 3/2$, then it contains the measure

$$\begin{aligned} \mu &= \frac{4}{M^2} \mu_x \otimes \mu_y \otimes \mu_z - \frac{2}{M} (\nu_x \otimes \mu_y \otimes \mu_z + \mu_x \otimes \nu_y \otimes \mu_z + \mu_x \otimes \mu_y \otimes \nu_z) \\ &\quad + 2(\mu_{xy} \otimes \nu_z + \mu_{xz} \otimes \nu_y + \mu_{yz} \otimes \nu_x) \\ &\quad - \frac{1}{M} (\mu_{xy} \otimes \mu_z + \mu_{xz} \otimes \mu_y + \mu_{yz} \otimes \mu_x), \end{aligned}$$

where

$$M = \mu_{xy}(X \times Y) = \mu_{xz}(X \times Z) = \mu_{yz}(Y \times Z).$$

For $c > 2$ the set $\Pi(\mu_{xy}, \mu_{yz}, \mu_{xz})$ can be empty.

Definition 6.7. A consistent collection of probability measures $\mu_\alpha, \alpha \in \mathcal{I}_{nk}$, is called decomposable if there exists a collection of probability measures ν_i on the spaces X_i and a joining measure $\mu \in \Pi(\{\mu_\alpha\})$ given by a bounded and separated from zero density with respect to the measure $\nu_1 \otimes \nu_2 \otimes \dots \otimes \nu_n$.

Let us give simple sufficient conditions for the existence of a dual solution.

Theorem 6.8. *Let X_i be compact metric spaces and $h \geq 0$ be a continuous cost function on X . Let $\{\mu_\alpha\}$ be a decomposable collection of probability measures on the spaces $X_\alpha, \alpha \in \mathcal{I}_{nk}$. Then*

$$\min_{\pi \in \Pi(\{\mu_\alpha\})} \int_X h d\pi = \max_{\alpha \in \mathcal{I}_{nk}} \sum \int_{X_\alpha} f_\alpha d\mu_\alpha,$$

where the maximum is taken over all the systems of functions $f_\alpha \in L^1(\mu_\alpha)$ with values in the set $[-\infty, +\infty)$ for which $\sum_{\alpha \in \mathcal{I}_{nk}} f_\alpha(x_\alpha) \leq h(x)$ for all $x \in X$.

Note that without the condition of decomposability the dual problem can fail to have a solution even in the discrete case.

Finally, for a bounded cost function, in some situations it becomes possible to prove that solutions are bounded.

Theorem 6.9. *Let $X = Y = Z = \mathbb{N}$ and let $\mu_x, \mu_y,$ and μ_z be probability measures on $X, Y,$ and $Z,$ respectively. Consider the $(3, 2)$ -problem for*

$$\mu_{xy} = \mu_x \otimes \mu_y, \quad \mu_{xz} = \mu_x \otimes \mu_z, \quad \mu_{yz} = \mu_y \otimes \mu_z,$$

and a cost function h with values in $[0, 1]$. Assume that the function

$$F(x, y, z) = f(x, y) + g(x, z) + h(y, z)$$

is a solution of the dual problem. Then $F \geq -12$.

7. Optimal transportation with parameters

It is of interest to consider transport problems with a parameter on which the marginals and cost function, as well as other data in the problem, can depend. Such problems were studied in [63], [129], [134], [96], and [34]–[36]. The first questions arising here are the measurability and continuity of solutions and minima with respect to the parameter. Measurability holds under very general conditions. For a simplification of technical details we consider the case where X and Y are completely regular Souslin spaces, for example, complete separable metric spaces.

In problems without constraints on the densities of plans the spaces of measures are equipped with their weak topologies and the corresponding Borel σ -algebras $\mathcal{B}(\mathcal{P}(X)), \mathcal{B}(\mathcal{P}(Y)),$ and $\mathcal{B}(\mathcal{P}(X \times Y))$. In the case of density constraints it is natural, as mentioned above, to use topologies connected with the space L^1 .

Assume that (T, \mathcal{T}) is a measurable space, the map $t \mapsto \mu_t, T \rightarrow \mathcal{P}(X),$ is $(\mathcal{T}, \mathcal{B}(\mathcal{P}(X)))$ -measurable, and the map $t \mapsto \nu_t, T \rightarrow \mathcal{P}(Y),$ is $(\mathcal{T}, \mathcal{B}(\mathcal{P}(Y)))$ -measurable. In our situation such measurability reduces to the \mathcal{T} -measurability of real functions

$$t \mapsto \mu_t(A), \quad t \mapsto \nu_t(B)$$

for all Borel sets $A \subset X, B \subset Y$.

In most important examples (T, \mathcal{T}) is a metric space with its Borel σ -algebra, so we are speaking of the Borel measurability of the indicated maps.

The cost function $h \geq 0$ on $T \times X \times Y$ is assumed to be measurable with respect to $\mathcal{B}(T) \otimes \mathcal{B}(X) \otimes \mathcal{B}(Y)$.

First we consider the case where X and Y are complete separable metric spaces, but (T, \mathcal{T}) is a general measurable space.

Theorem 7.1. *Assume that X and Y are complete separable metric spaces, the cost functions $h_t: (x, y) \mapsto h(t, x, y)$ are continuous for all $t \in T,$ and the quantities $K(t) := K_{h_t}(\mu_t, \nu_t)$ are finite. Then the function K is \mathcal{T} -measurable.*

In addition, there exist optimal measures $\sigma_t \in \Pi(\mu_t, \nu_t)$ for h_t such that the map $t \mapsto \sigma_t$ is measurable with respect to \mathcal{T} and $\mathcal{B}(\mathcal{P}(X \times Y))$.

Note that the formulation of Theorem 4.2 in [4], where this result was proved, contains a typo, which only concerns the last assertion of the theorem: in place of the \mathcal{T} -measurability of K its $(\mathcal{T}, \mathcal{B}(\mathcal{P}(X \times Y)))$ -measurability is mentioned. Note also that, in view of continuity in (x, y) in this theorem, joint measurability in all variables follows from measurability in t for all fixed x and y .

In the next theorem the assumption of continuity of the cost function is relaxed, but T must be a Souslin space.

Theorem 7.2. *Assume that X and Y are complete separable metric spaces, T is a Souslin space, $t \mapsto \mu_t$ and $t \mapsto \nu_t$ are Borel maps with values in the spaces $\mathcal{P}(X)$ and $\mathcal{P}(Y)$, respectively, the function $h \geq 0$ is $\mathcal{B}(T) \otimes \mathcal{B}(X) \otimes \mathcal{B}(Y)$ -measurable, the functions h_t are lower semicontinuous, and the corresponding values $K_{h_t}(\mu_t, \nu_t)$ are finite. Then the function $t \mapsto K_{h_t}(\mu_t, \nu_t)$ is Borel measurable and there exist optimal measures $\sigma_t \in \Pi(\mu_t, \nu_t)$ for h_t such that the map $t \mapsto \sigma_t$ is measurable with respect to $\mathcal{B}(T)$ and $\mathcal{B}(\mathcal{P}(X \times Y))$.*

Theorem 7.3. *Assume that X and Y are completely regular Souslin spaces, T is a Souslin space, and $h: T \times X \times Y \rightarrow [0, +\infty)$ is a Borel function such that the function h_t is continuous for every t . Let $t \mapsto \mu_t$ and $t \mapsto \nu_t$ be Borel maps with values in $\mathcal{P}(X)$ and $\mathcal{P}(Y)$, respectively, such that $K_{h_t}(\mu_t, \nu_t) < \infty$ for all t . Then the function $t \mapsto K(t)$ is Borel, and there exist optimal measures $\sigma_t \in \Pi(\mu_t, \nu_t)$ for h_t such that the map $t \mapsto \sigma_t$ is Borel measurable.*

Moreover, there exists a sequence of Borel maps $\Phi_n: T \rightarrow \mathcal{P}(X \times Y)$ such that for every $t \in T$ the sequence $\{\Phi_n(t)\}$ is dense in the convex compact set M_t of all h_t -optimal measures in the set $\Pi(\mu_t, \nu_t)$.

Although these theorems cover the spaces necessary for applications, it would be interesting to know whether or not Theorem 7.1 is true for Souslin spaces X and Y and lower semicontinuous cost functions (such a result would imply all the three theorems as particular cases). In [34], in the case of Souslin spaces and lower semicontinuous cost functions some results were obtained about the weaker measurability with respect to the σ -algebra generated by Souslin sets. the

The next theorems about the dependence of solutions of the nonlinear problem on a parameter were proved in [35]. Recall that Luzin spaces are the images of complete separable metric spaces under continuous injective maps (they form a subclass of Souslin spaces).

Theorem 7.4. *Let X and Y be Luzin completely regular spaces (for example, complete separable metric spaces). Assume that Borel maps*

$$t \mapsto \mu_t, \quad T \rightarrow \mathcal{P}(X),$$

and

$$t \mapsto \nu_t, \quad T \rightarrow \mathcal{P}(Y),$$

and a Borel function

$$h: T \times X \times Y \times \mathcal{P}(X \times Y) \rightarrow [0, +\infty)$$

are given such that for every $t \in T$ the function

$$h_t : (x, y, \sigma) \mapsto h(t, x, y, \sigma)$$

is lower semicontinuous on all sets of the form $K \times \Pi(\mu_t, \nu_t)$, where $K \subset X \times Y$ is compact, and the quantities $K_{h_t}(\mu_t, \nu_t)$ are finite for all $t \in T$. Then the function $t \mapsto K_{h_t}(\mu_t, \nu_t)$ is Borel and there exists a Borel map $t \mapsto \sigma_t$ from T to $\mathcal{P}(X \times Y)$ such that the measure σ_t is optimal for the triple (μ_t, ν_t, h_t) for every t .

Moreover, there exists a sequence of Borel maps ξ_n from T to $\mathcal{P}(X \times Y)$ such that for every t the sequence of measures $\xi_n(t)$ is everywhere dense in the set of optimal plans for the triple (μ_t, ν_t, h_t) .

Theorem 7.5. Assume that in the previous theorem the cost function h has the form

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

where the function H is defined on $T \times X \times \mathcal{P}(Y)$, the functions

$$H_t : (x, p) \mapsto H(t, x, p)$$

are lower semicontinuous, and the functions $p \mapsto H(t, x, p)$ are convex for all t and x . Then the conclusion of the previous theorem remains valid.

Finally, a parametric version of the transport problem with constraints on the densities of plans was considered in [30], where the following result was obtained.

Given a sequence of measurable spaces (X_n, \mathcal{B}_n) and a probability measure λ on the product $X = \prod_{n=1}^\infty X_n$ equipped with the product \mathcal{B} of the σ -algebras \mathcal{B}_n , denote the projection of λ onto X_n by λ_n . Assume that the measure λ is separable, that is, the space $L^1(\lambda)$ is separable. Then the measures λ_n are also separable. For every n fix a sequence of bounded \mathcal{B}_n -measurable functions $\varphi_{n,j}$ that is everywhere dense in $L^1(\lambda_n)$. In addition, fix a countable system of sets $A_j \in \mathcal{B}$ with the following property: for any $\varepsilon > 0$ any set in \mathcal{B} coincides with some A_j up to a set of λ -measure less than ε . Such a system exists because λ is separable.

Assume that (T, \mathcal{T}) is a measurable space and for every $t \in T$ we are given probability measures $\mu_{n,t}$ on \mathcal{B}_n which are absolutely continuous with respect to λ_n , a non-negative \mathcal{B} -measurable function h_t on X , and a non-negative λ -integrable function Φ_t on X such that the functions

$$(x, t) \mapsto h_t(x) \quad \text{and} \quad (x, t) \mapsto \Phi_t(x)$$

are $\mathcal{B} \otimes \mathcal{T}$ -measurable and the maps $t \mapsto \mu_{n,t}$ are measurable in the following sense: the functions

$$t \mapsto \int_{X_n} \varphi_{n,j}(x) \mu_{n,t}(dx)$$

are \mathcal{T} -measurable for all n and j . If X_n and T are separable metric spaces and $\mathcal{T} = \mathcal{B}(T)$, then it suffices to assume that the measure $\mu_{n,t}$ depends Borel measurably on t , that is, the functions $t \mapsto \mu_{n,t}(B)$ are Borel for Borel sets B .

The Radon–Nikodym density of the measure $\mu_{n,t}$ with respect to the measure λ_n is denoted by $\varrho_{n,t}$, so that

$$\mu_{n,t} = \varrho_{n,t} \cdot \lambda_n.$$

For any fixed $t \in T$ we denote by \mathcal{L}_t the set of probability densities $\psi \in L^1(\lambda)$ such that $\psi \leq \Phi_t$ almost everywhere with respect to λ and the projection of the measure $\psi \cdot \lambda$ onto X_n equals $\mu_{n,t}$ for all n . This condition on projections can be expressed as

$$\varrho_{n,t} = E(\psi \mid \mathcal{B}_n),$$

where $E(\psi \mid \mathcal{B}_n)$ is the conditional expectation of the function ψ with respect to the measure λ and the σ -algebra \mathcal{B}_n , that is, a \mathcal{B}_n -measurable function that is integrable with respect to λ and such that

$$\int_{E \times Y_n} \psi \, d\lambda = \int_{E \times Y_n} E(\psi \mid \mathcal{B}_n) \, d\lambda$$

for all $E \in \mathcal{B}_n$, where $Y_n = \prod_{i \neq n} X_i$. Under our assumptions this is equivalent to the equalities

$$\int_X \varphi_{n,j} \psi \, d\lambda = \int_X \varphi_{n,j} \varrho_{n,t} \, d\lambda \quad \forall j, n \in \mathbb{N}. \tag{7.1}$$

The estimate $\psi \leq \Phi_t$ in $L^1(\lambda)$ is equivalent to the countable system of scalar inequalities

$$\int_{A_j} \psi \, d\lambda \leq \int_{A_j} \Phi_t \, d\lambda \quad \forall j \in \mathbb{N}. \tag{7.2}$$

Thus, we are concerned with the existence of probability densities f_t satisfying the countable system of linear constraints (7.1) and (7.2) and minimizing the integrals of the functions $f_t h_t$ against the measure λ for every t ; moreover, the joint measurability of $f_t(x)$ on $X \times T$ is also required.

Each set \mathcal{L}_t is compact in the weak topology since it is obviously weakly closed and lies in the set of non-negative functions not exceeding Φ_t , while the latter is weakly compact because of uniform integrability (see [25], § 4.7(iv)).

If the densities $\varrho_{n,t}$ are bounded by a number C , then in order that \mathcal{L}_t be non-empty it suffices to have the estimate $\Phi_t \geq C$. If there are only two factors and $\lambda = \lambda_1 \otimes \lambda_2$, then it suffices to have $\varrho_{1,t}(x_1)\varrho_{2,t}(x_2) \leq \beta_t(x_1, x_2)$. If \mathcal{L}_t is not empty and there is a function $v_t \in \mathcal{L}_t$ such that $h_t v_t \in L^1(\lambda)$, then the functional

$$v \mapsto \int_X h_t v \, d\lambda$$

has a finite minimum $M(h_t, \Phi_t)$ on the set

$$\mathcal{K}_t = \{v \in \mathcal{L}_t : v h_t \in L^1(\lambda)\},$$

which is convex and compact in the weak topology. Hence we obtain the non-empty convex weakly compact sets of optimal measures

$$\mathcal{M}_t = \left\{ v \in \mathcal{K}_t : M(h_t, \Phi_t) = \int_X v h_t \, d\lambda \right\}.$$

For a σ -algebra \mathcal{T} we denote by $\mathcal{S}(\mathcal{T})$ the class of sets obtained from \mathcal{T} by the Souslin operation (see [25]), and we denote the σ -algebra generated by this class

by $\sigma(\mathcal{S}(\mathcal{T}))$. Let $\widehat{\mathcal{T}}$ be the class of universally measurable sets generated by \mathcal{T} , that is, the intersection of the completions of \mathcal{T} with respect to all probability measures on \mathcal{T} . It is known (see [25], Theorem 1.10.5) that $\sigma(\mathcal{S}(\mathcal{T})) \subset \widehat{\mathcal{T}}$.

Theorem 7.6. *Let T be a non-empty Souslin space with Borel σ -algebra $\mathcal{T} = \mathcal{B}(T)$. Assume that the set \mathcal{K}_t is not empty for any t . Then there exist functions $f_t \in \mathcal{K}_t$ such that the measures $f_t \cdot \lambda$ are optimal for the functions h_t , that is, $f_t \in \mathcal{M}_t$, and the function $(x, t) \mapsto f_t(x)$ is measurable with respect to $\mathcal{B} \otimes \mathcal{T}$. Moreover, there exists a sequence of functions $f_{n,t}$ with the same properties that is dense in \mathcal{M}_t for every t . In addition, the function*

$$t \mapsto M(h_t, \beta_t) = \int_X h_t f_t d\lambda$$

is \mathcal{T} -measurable.

In the case of a general measurable space (T, \mathcal{T}) the functions $f_t \in \mathcal{M}_t$ can be chosen so that the function $(x, t) \mapsto f_t(x)$ is $\mathcal{B} \otimes \sigma(\mathcal{S}(\widehat{\mathcal{T}}))$ -measurable and the function $t \mapsto M(h_t, \beta_t)$ is $\widehat{\mathcal{T}}$ -measurable.

Results concerning continuity with respect to the parameter in optimal transportation problems were obtained in [23], [72], [121], [79], and [36]. The last of these papers contains the most general results. However, for simpler formulations we present them in a less general form.

Let X and Y be complete separable metric spaces and let T be a metric space. Assume that for every $t \in T$ we have measures $\mu_t \in \mathcal{P}(X)$ and $\nu_t \in \mathcal{P}(Y)$ such that the maps $t \mapsto \mu_t$ and $t \mapsto \nu_t$ are continuous in the weak topology. In addition, let $h: T \times X \times Y \rightarrow [0, +\infty)$ be a continuous function. Set

$$h_t(x, y) := h(t, x, y).$$

Theorem 7.7. *If the function h is bounded, then the function $t \mapsto K_{h_t}(\mu_t, \nu_t)$ is continuous on T .*

Corollary 7.8. *If, in the situation of the previous theorem, there is a unique optimal plan σ_t for every t , then this plan is continuous in t .*

However, without uniqueness it can be impossible to choose optimal plans depending continuously on the parameter t . Simple examples of this sort were constructed in [36]. In particular, one can take all measures μ_t and ν_t to be equal to Lebesgue measure on $[0, 1]$ (so that they do not depend on t) and can take the cost function

$$h_t(x, y) = \begin{cases} \min(|x - y|, |x + y - 1| + t), & t \geq 0, \\ \min(|x - y| - t, |x + y - 1|), & t < 0. \end{cases}$$

Thus, we can say that the situation with the continuity of the optimal cost is similar to the case of measurability, but the continuous choice of an optimal plan brings in some difference. Some compensation is provided by the use of approximate optimal plans. Given $\varepsilon > 0$, a measure $\sigma \in \Pi(\mu, \nu)$ will be called ε -optimal for the cost function h if

$$\int_{X \times Y} h d\sigma \leq K_h(\mu, \nu) + \varepsilon.$$

Theorem 7.9. *For every fixed $\varepsilon > 0$ there exist ε -optimal measures $\sigma_t^\varepsilon \in \Pi(\mu_t, \nu_t)$ for the cost functions h_t which depend continuously on t in the weak topology.*

For Monge optimal maps one can also pose the question about measurable or continuous dependence on a parameter. For example, the following result from [36] claims the continuity of the Monge map with respect to the parameter in the metric of convergence in probability.

Proposition 7.10. *Let X be a metric space, let $\mu_n \in \mathcal{P}(X)$ for $n \in \mathbb{Z}_+$, and let $\mu_n \rightarrow \mu_0$ in variation. Let $T_n: X \rightarrow X$ be Borel maps such that the measures σ_n that are the images of the measures μ_n under the maps $x \mapsto (x, T_n(x))$ converge weakly to a measure σ_0 . Then the maps T_n converge to a map T_0 in measure μ_0 .*

Corollary 7.11. *Assume that $X = Y$ is a complete separable metric space, the measures $\mu_n \in \mathcal{P}(X)$ converge to a measure μ_0 in variation, the measures $\nu_n \in \mathcal{P}(X)$ converge weakly to a measure ν_0 , and the continuous cost functions $h_n \geq 0$ on X^2 are uniformly bounded and converge uniformly on compact subsets to a function h_0 . Also assume that for all $n \geq 0$ optimal Kantorovich plans for the triples (μ_n, ν_n, h_n) are unique and are generated by the unique optimal Monge maps T_n . Then the maps T_n converge to T_0 in the measure μ_0 .*

In [8], [61], and [105] close results were previously obtained for some special cases connected with the investigation of conditions for the existence and uniqueness of Monge maps.

8. Metrics and topologies of Kantorovich type

Recall (see [65]) that the topology of any completely regular topological space X is generated by a family of pseudometrics Π (a pseudometric differs from a metric in that it can vanish on a pair of different elements). Given a pseudometric d on X , we denote by $\text{Lip}_1(d)$ the set of 1-Lipschitz functions with respect to d , that is, of functions f on X such that

$$|f(x) - f(y)| \leq d(x, y) \quad \forall x, y \in X.$$

A straightforward analogue of the Kantorovich–Rubinshtein norm is provided by the Kantorovich–Rubinshtein seminorm on the space of Radon measures $\mathcal{M}(X)$ on X , which are defined by

$$\|\mu\|_{\text{KR},d} = \sup \left\{ \int f d\mu : f \in \text{Lip}_1(d), |f| \leq 1 \right\}.$$

On the subspace $\mathcal{M}_d^1(X)$ of measures μ such that for some (hence all) $x_0 \in X$ the function $d(x, x_0)$ is integrable with respect to the total variation of μ we introduce the Kantorovich seminorm

$$\|\mu\|_{\text{K},d} = \sup \left\{ \int f d\mu : f \in \text{Lip}_1(d), f(x_0) = 0 \right\} + |\mu(X)|,$$

which is completely analogous to the Kantorovich norm in the case of a metric space. The topologies generated by these families of seminorms will be called the Kantorovich–Rubinshtein and Kantorovich topologies and denoted by τ_{KR} and τ_{K} , respectively.

Theorem 8.1. *Assume that the topology in X is generated by a family of pseudometrics Π . Then the weak topology on the set of non-negative measures $\mathcal{M}^+(X)$ is generated by the family of seminorms $\| \cdot \|_{\text{KR},p}$, $p \in \Pi$.*

In addition, these seminorms also generate the weak topology on every uniformly tight set in $\mathcal{M}(X)$ that is bounded in variation.

Theorem 8.2. *Assume that a completely regular space X is separable or possesses a countable system of continuous functions separating points. Then the weak topology coincides with the topology τ_{KR} on weakly compact sets in $\mathcal{M}(X)$.*

Note a simple sufficient condition for convergence in the topology τ_{K} . Given a family of pseudometrics Π generating the topology of the space X , we denote by $\mathcal{M}^\Pi(X)$ the class of measures $\mu \in \mathcal{M}(X)$ for which the function $x \mapsto p(x, x_0)$ with $p \in \Pi$ is integrable with respect to $|\mu|$ for a fixed point $x_0 \in X$ (the choice of x_0 does not influence the definition of this class).

Theorem 8.3. *Assume that a net $\{\mu_\alpha\} \subset \mathcal{M}^\Pi(X)$ converges to a measure $\mu \in \mathcal{M}^\Pi(X)$ in the topology τ_{KR} (for non-negative measures or measures in a bounded uniformly tight family this is equivalent to weak convergence). If every pseudometric p from Π satisfies the condition of uniform integrability*

$$\lim_{R \rightarrow \infty} \sup_{\alpha} \int_{\{p \geq R\}} p(x, x_0) |\mu_\alpha|(dx) = 0,$$

then $\{\mu_\alpha\}$ converges in the topology τ_{K} . In the case of probability measures this condition is also necessary.

Finally, in the case of a countable sequence of measures, in place of convergence in the topology τ_{KR} it suffices to have weak convergence.

In Example 8.6 below we will see that for nets of signed measures it is not enough to have weak convergence in place of convergence in the topology τ_{KR} .

An analogous result is true for the topology $\tau_{\text{K},q}$ with $q \geq 1$, which is introduced on the subclass $\mathcal{M}^{\Pi,q}(X)$ of $\mathcal{M}^\Pi(X)$ consisting of the measures μ for which all functions $x \mapsto p(x, x_0)^q$, where $p \in \Pi$, are integrable with respect to $|\mu|$. This topology is generated by all seminorms

$$K_{p,q}(\mu) = \|(1 + p(\cdot, x_0)^q)\mu\|_{\text{KR},p},$$

where $p \in \Pi$ and x_0 is a fixed point.

Theorem 8.4. *Assume that a net of measures $\mu_\alpha \in \mathcal{M}^{\Pi,q}(X)$, where $q \geq 1$, converges to a measure $\mu \in \mathcal{M}(X)$ in the topology τ_{KR} (for non-negative measures or measures in a bounded uniformly tight family this is equivalent to weak convergence). If the equality*

$$\lim_{R \rightarrow \infty} \sup_{\alpha} \int_{\{p \geq R\}} p(x, x_0)^q |\mu_\alpha|(dx) = 0$$

holds for every pseudometric p in Π , then $\mu \in \mathcal{M}^{\Pi,q}(X)$ and $\{\mu_\alpha\}$ converges to μ in the topology $\tau_{\text{K},q}$.

If some measures $\mu_\alpha \in \mathcal{M}^1$ on a locally convex space X have barycentres b_α and converge in the topology τ_K to a measure $\mu \in \mathcal{M}^1$ with barycentre b , then we have the convergence of barycentres $b_\alpha \rightarrow b$. Indeed, for every continuous seminorm p on X we have the estimate

$$p(b_\alpha - b) \leq \|\mu_\alpha - \mu\|_{K,p},$$

since for every continuous linear functional l on X such that $l \leq p$ we have

$$l(b_\alpha - b) = \int_X l d(\mu_\alpha - \mu) \leq \|\mu_\alpha - \mu\|_{K,p},$$

because $l \in \text{Lip}_1(p)$. By the Hahn–Banach theorem the supremum of $l(b_\alpha - b)$ over the functionals satisfying the estimate $l \leq p$ is $p(b_\alpha - b)$.

Applying to the convergence of barycentres we obtain the following.

Corollary 8.5. *If a sequence of Radon measures μ_n on a locally convex space X converges weakly to a Radon measure μ , each measure μ_n has a barycentre b_n , the measure μ has a barycentre b , and every continuous seminorm is uniformly integrable with respect to the sequence $\{\mu_n\}$ in the sense described in Theorem 8.3, then $b_n \rightarrow b$.*

In the case of probability measures the same is true for nets.

The next example shows that the last assertion of the corollary can fail to hold for nets of signed measures, which, according to what we said above, also gives a counterexample to the last assertion of Theorem 8.3 in the case of nets.

Example 8.6. In the Banach space $X = l^1$ there exists a bounded in variation net of signed discrete measures on the unit ball that converge weakly to the zero measure and have barycentres of unit norm.

For the construction we fix a finite set of bounded continuous functions f_1, \dots, f_n on X . Consider the following vectors in \mathbb{R}^n :

$$v_j = (f_1(e_j), \dots, f_n(e_j)), \quad j = 1, \dots, n+1,$$

where $\{e_j\}$ is the standard basis in l^1 . The vectors v_j are linearly dependent, hence there exist numbers c_1, \dots, c_{n+1} not all equal to zero such that

$$\sum_{j=1}^{n+1} c_j v_j = 0;$$

in other words,

$$\sum_{j=1}^{n+1} c_j f_i(e_j) = 0, \quad i = 1, \dots, n.$$

We can assume that $\sum_j |c_j| = 1$. For every basic neighborhood of zero in the weak topology

$$U = U_{f_1, \dots, f_n, \varepsilon} = \left\{ \mu \in \mathcal{M}(X) : \left| \int_X f_i d\mu \right| < \varepsilon, i = 1, \dots, n \right\}$$

consider the discrete measure

$$\mu_U := \sum_{j=1}^{n+1} c_j \delta_{e_j}.$$

By construction $\mu_U \in U$ and this measure is concentrated on the unit sphere. The set of basic neighborhoods is directed by reverse inclusion: a neighborhood V is greater than a neighborhood U if $V \subset U$. By definition the net of measures μ_U constructed in his way converges weakly to zero. The barycentre of a measure μ_U is $\sum_j c_j e_j$, hence $\|\mu_U\| = \sum_{j=1}^{n+1} |c_j| = 1$.

There are simple sufficient conditions for the compactness of sets in $\mathcal{M}(X)$ in the topology τ_{KR} and sets in $\mathcal{M}^\Pi(X)$ in the topology τ_K .

Theorem 8.7. *Assume that a set $S \subset \mathcal{M}(X)$ is bounded in variation and uniformly tight. Then S has a compact closure in the topology τ_{KR} .*

If $S \subset \mathcal{M}^\Pi(X)$ and every pseudometric p in Π satisfies the condition of uniform integrability

$$\lim_{R \rightarrow \infty} \sup_{\mu \in S} \int_{\{p \geq R\}} p(x, x_0) |\mu|(dx) = 0$$

for some $x_0 \in X$, then S is contained in a compact set in the topology τ_K .

The next result from [3] shows that a uniformly tight set of Radon measures on a Banach space with uniformly integrable norm remains uniformly tight with respect to some stronger norm, which is also uniformly integrable (so that, under the condition of boundedness in variation, this family is contained in a compact set with respect to the Kantorovich norm). Moreover, this family is uniformly tight in some compactly embedded separable reflexive Banach space with uniformly integrable norm.

Theorem 8.8. *Let X be a Fréchet space, and let \mathcal{M} be a uniformly tight family of Radon measures on X such that all seminorms p_n in some sequence generating the topology in X are uniformly integrable with respect to the measures in \mathcal{M} , that is,*

$$\lim_{m \rightarrow \infty} \sup_{\mu \in \mathcal{M}} \int_{\{x : p_n(x) > m\}} p_n(x) |\mu|(dx) = 0, \quad n \in \mathbb{N}.$$

Then there exists a linear subspace $E \subset X$ with the following properties:

- (i) *the space E with some norm $\|\cdot\|_E$ is a separable reflexive Banach space with closed unit ball which is compact in the original space X ;*
- (ii) *the family \mathcal{M} is concentrated on E and uniformly tight on E with norm $\|\cdot\|_E$; moreover, this norm is uniformly integrable with respect to the measures in \mathcal{M} .*

Kantorovich and Kantorovich–Rubinshtein type metrics enable us to introduce convenient Hausdorff distances on sets of measures.

We recall that the Hausdorff distance between two non-empty bounded closed subsets A and B of a metric space (M, d) is defined by the formula

$$H(A, B) = \max \left(\sup_{x \in A} d(x, B), \sup_{y \in B} d(y, A) \right).$$

For our purposes this distance is of interest when we consider the space of probability measures on a metric space (X, d) and the set of transport plans in the space $\mathcal{P}(X \times Y)$ with Kantorovich–Rubinshtein metric d_{KR} generated by the natural metric $d_X(x_1, x_2) + d_Y(y_1, y_2)$ on $X \times Y$, where d_X is a metric on X and d_Y is a metric on Y . The Hausdorff distances generated by the Kantorovich–Rubinshtein metrics on spaces of measures will be denoted by H_{KR} . Similarly, H_K will denote the Hausdorff distances generated by the Kantorovich metrics on spaces of measures with finite first moment. From the topological point of view there is no principal difference between two such distances, since any metric can be replaced by a bounded metric generating the same topology. The topology on the space of measures does not change either.

In the case of the set $\mathcal{P}^p(X \times Y)$ with metric W_p the Hausdorff distance H_K^p is defined on the space of closed subsets of $\mathcal{P}^p(X \times Y)$.

For general completely regular spaces X and Y similar constructions arise. The topologies in these spaces can be generated by families of pseudometrics Ψ_X and Ψ_Y . The topology in the product $X \times Y$ is generated by the pseudometrics

$$((x_1, y_1), (x_2, y_2)) \mapsto d_1 \oplus d_2((x_1, y_1), (x_2, y_2)) = d_1(x_1, x_2) + d_2(y_1, y_2),$$

$$d_1 \in \Psi_X, \quad d_2 \in \Psi_Y.$$

On the spaces of probability measures on X, Y , and $X \times Y$, in the way described above we obtain the Kantorovich–Rubinshtein pseudometrics d_{KR, d_1}, d_{KR, d_2} , and $d_{KR, d_1 \oplus d_2}$, the Kantorovich pseudometrics d_{K, d_1} and so on. For example, for a completely regular space X with a fixed family of pseudometrics Ψ_X generating the topology, the set $\mathcal{P}^\Psi(X)$ of Radon probability measures on X with respect to which the functions $x \mapsto d(x, x_0)$ are integrable for all $d \in \Psi_X$ is equipped with the Kantorovich pseudometrics $d_{K, d}$. On the space of closed subsets of the space of measures $\mathcal{P}(X \times Y)$, in the way described above we obtain the Hausdorff pseudometrics of the form $H_{K, d_1 \oplus d_2}$ generated by the pseudometrics d_1 on X and d_2 on Y .

Theorem 8.9. *Let $\mu_1, \mu_2 \in \mathcal{P}(X)$ and $\nu_1, \nu_2 \in \mathcal{P}(Y)$, and let α and β be continuous pseudometrics on X and Y , respectively. Then, for every measure $\sigma_1 \in \Pi(\mu_1, \nu_1)$ there exists a measure $\sigma_2 \in \Pi(\mu_2, \nu_2)$ such that*

$$d_{K, \alpha \oplus \beta}(\sigma_1, \sigma_2) \leq d_{K, \alpha}(\mu_1, \mu_2) + d_{K, \beta}(\nu_1, \nu_2). \tag{8.1}$$

Hence for the corresponding Kantorovich and Hausdorff pseudometrics the inequality

$$H_{K, \alpha \oplus \beta}(\Pi(\mu_1, \nu_1), \Pi(\mu_2, \nu_2)) \leq d_{K, \alpha}(\mu_1, \mu_2) + d_{K, \beta}(\nu_1, \nu_2) \tag{8.2}$$

holds. The analogous assertion is valid for n marginals: if $\mu_i, \nu_i \in \mathcal{P}(X_i)$, $i = 1, \dots, n$, α_i are continuous pseudometrics on the spaces X_i , then for each measure $\sigma \in \Pi(\mu_1, \dots, \mu_n)$ there exists a measure $\pi \in \Pi(\nu_1, \dots, \nu_n)$ such that

$$d_{K, \alpha_1 \oplus \dots \oplus \alpha_n}(\sigma, \pi) \leq d_{K, \alpha_1}(\mu_1, \nu_1) + \dots + d_{K, \alpha_n}(\mu_n, \nu_n).$$

Estimates connecting the Kantorovich distance with Sobolev norms were studied in [39], [42], [43], and [40].

In [27] some properties of sequential continuity were considered for the space of measures $\mathcal{M}(X)$ with weak topology. This space is not metrizable if X is infinite, but in the case where X is a complete separable metric space, every linear functional l on $\mathcal{M}(X)$ that is sequentially continuous in the weak topology (so that $l(m_n) \rightarrow 0$ if a sequence m_n converges weakly to zero) is continuous in the usual topological sense. Theorem 1 in [27] contains a more general result. However, for nonlinear functions sequential continuity does not imply usual continuity.

Results on approximation of measures on infinite-dimensional Banach or locally convex spaces by finite-dimensional images of these measures, that is, images under continuous linear maps with finite-dimensional ranges, can be useful in problems of optimal transportation. Such approximations are obviously possible in spaces with Schauder bases, but in the general case the question is open (see the discussion in [28]).

The smoothing of Kantorovich metrics (erroneously called ‘Wasserstein distances’) was studied in [53].

Various problems connected with transport distances, the topological properties of spaces of measures, and related questions in the theory of metric measure spaces were discussed in [6], [10]–[13], [18], [24], [52], [78], [83], [89], [101], [104], [116], [122], and [123].

9. Other directions of research

Let us mention some other lines of investigation pursued in the area of optimal transportation in recent years.

Martingale optimal transportation has been developing actively: see [17], [21], [47], [58], [73], [82], [95], [103], [112], and [130]. In its simplest form this problem is stated for n Borel probability measures μ_1, \dots, μ_n on the real line and a bounded Borel cost function h on \mathbb{R}^n . It deals with minimizing the integral

$$\int_{\mathbb{R}^n} h d\mu$$

over the Borel probability measures μ on \mathbb{R}^n with the following restrictions: the projection of μ onto the k th factor is μ_k and the coordinate functions x_1, \dots, x_n form a martingale with respect to the measure μ and the σ -algebras $\sigma_1, \dots, \sigma_n$, where σ_k is generated by the coordinate functions x_1, \dots, x_k . The problem is similarly formulated for infinitely many coordinates. In [17], [47], and [130] the reader can find interesting recent results on the continuity of solutions to the martingale transport problem. Questions close to martingale optimal transportation problems were studied in [1], [14], and [97].

This modification of the Kantorovich problem fits the general framework of problems of Kantorovich type with additional linear constraints, connected, for instance, with various symmetries of solutions, was considered in [132] and [133]. However, it possesses a number of important special features. Let us present a precise formulation of a problem with linear constraints. Given completely regular spaces X_i with Radon probability measures μ_i , $i = 1, \dots, n$, let $\Pi(\mu_1, \dots, \mu_n)$ denote the set of measures in $\mathcal{P}(X_1 \times \dots \times X_n)$ with projections μ_i onto the factors. The function spaces

$$C_L^i = \{f \in L^1(\mu_i) \cap C(X_i)\}$$

of continuous integrable functions are equipped with the standard norms from $L^1(\mu_i)$ (more precisely, these are seminorms if the measures μ_i do not have full support), and the space

$$C_L = \left\{ h \in C(X) : |h(x)| \leq \sum_{i=1}^n f_i(x_i), \text{ where } f_i \in L^1(\mu_i) \right\}$$

is equipped with the seminorm

$$\|h\|_L = \sup_{\pi \in \Pi(\mu_1, \dots, \mu_n)} \int_X |h| d\pi.$$

Set

$$F = \bigoplus_{i=1}^n C_L^i \subset C_L.$$

Given a linear subspace $W \subset C_L$ and a cost function $h \in C_L$, the modification of the Kantorovich problem under consideration consists in finding the quantity

$$\inf_{\pi \in \Pi_W} \int_X h d\pi, \quad \Pi_W = \left\{ \pi \in \Pi(\mu_1, \dots, \mu_n) : \int_X w d\pi = 0 \ \forall w \in W \right\}.$$

It was shown in [132] that a minimum is attained in this problem if the set Π_W is not empty. In addition,

$$\inf_{\pi \in \Pi_W} \int_X h d\pi = \sup_{f \leq h} \sum_{k=1}^n \int_{X_k} f_k(x_k) \mu_k(dx_k), \quad f = \sum_{i=1}^n f_i, \quad f_i \in C_b(X_i),$$

where $C_b(X_i)$ is the space of bounded continuous functions on X_i .

Let us also mention the Schrödinger problem, which was posed by him in connection with some questions in statistical physics. It turned out that a particular case of the Monge–Kantorovich problem can be obtained as the limit of a sequence of Schrödinger problems after a suitable normalization. The works [55], [56], [70], [99], and [98] are concerned with this line of research. In the Schrödinger problem one considers the measure R on some space of continuous trajectories on the interval $[0, 1]$ (for instance, on $C([0, 1], \mathbb{R}^n)$) that is the distribution of the Brownian motion for which the distribution of the initial point is given by Lebesgue measure (R can be an unbounded measure). The problem is to minimize the entropy

$$H(P | R) = \int \log \left(\frac{dP}{dR} \right) dP$$

on the set of measures P absolutely continuous with respect to R for which the distributions $\mu_0 = P_0$ and $\mu_1 = P_1$ on X at the initial and final points of the interval are given. The corresponding Kantorovich problem has the form

$$\int C(\omega) P(d\omega) \rightarrow \min, \quad P_0 = \mu_0, \quad P_1 = \mu_1,$$

where $C(\omega) = \|\dot{\omega}_t\|_{L^2}^2/2$ for absolutely continuous trajectories and $C(\omega) = +\infty$ otherwise. The recent monograph [108] is concerned with related questions.

Transport problems connected with Gaussian measures and their nonlinear transformations were studied in [48]–[50].

Optimal transportation of vector (for instance, matrix) measures was considered in [45], [54], [57], [60], and [106]. Another kind of vector optimal transportation was studied in [131], where for non-negative measures μ_1, \dots, μ_d on a space X , non-negative measures ν_1, \dots, ν_d on a space Y , and a cost function h on $X \times Y$ the author considered the problem of minimizing the integral

$$\int_{X \times Y} h \, d\sigma$$

over the non-negative measures σ on $X \times Y$ satisfying the conditions

$$\int_{X \times B} \frac{d\mu_j}{d\mu}(x) \sigma(dx \, dy) = \nu_j(B), \quad j = 1, \dots, d,$$

for all measurable sets $B \subset Y$, where $\mu = \frac{1}{d} \sum_{j=1}^d \mu_j$. Note that one can also consider the following vector analogue of the Monge problem: given atomless Borel probability measures μ_1, \dots, μ_d on a Souslin space X , a Borel probability measure ν on a Souslin space Y , and sufficiently nice cost functions h_1, \dots, h_d on $X \times Y$, minimize the quantity

$$\sum_{i=1}^d \int_X h_i(x, T(x)) \mu_i(dx)$$

over the Borel maps $T: X \rightarrow Y$ that take all measures μ_i simultaneously to the measure ν . The existence of such maps is ensured by Lyapunov's theorem (see [25], Corollary 9.12.37). The indicated sum can be written in the form

$$\int_X h(x, T(x)) \mu(dx), \quad h(x, y) = \sum_{i=1}^d \frac{d\mu_i}{d\mu}(x) h_i(x, y).$$

However, the difference from the usual Monge problem is that not only the measure μ is transformed into ν , but also each measure μ_i . The book [131] contains results for the 'semidiscrete case', where the measure ν is discrete. It would be interesting to study the general case.

For metric barycentres generated by metrics of Kantorovich type, see [68].

Applications of the Kantorovich problem to the Plateau problem of minimal surfaces with prescribed boundary were discussed in [46].

Connections of optimal transportation with the problem of small divisors were considered in [94].

The regularization of transport problems was studied in [61] and [105].

Dynamical aspects of optimal transportation were considered in the papers [59] and [107].

In the results presented above continuous or lower semicontinuous cost functions were considered. However, in many problems, in particular, in ones connected with duality, discontinuous cost functions are considered: see, for example, [20] and [100]. The paper [127] deals with an interesting concept of a virtually continuous function

in the spirit of a certain refinement of the Luzin property. In a number of problems such a property can replace the usual continuity of cost functions.

For characterizations of optimal plans, uniqueness problems, and duality, see [109], [111], and [119].

I thank K. A. Afonin, A. V. Kolesnikov, E. D. Kosov, S. N. Popova, and A. V. Rezbav for useful discussions.

Bibliography

- [1] B. Acciaio, J. Backhoff-Veraguas, and A. Zalashko, “Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization”, *Stochastic Process. Appl.* **130**:5 (2020), 2918–2953.
- [2] 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 (2021), 1423–1453.
- [3] K. A. Afonin and V. I. Bogachev, “Kantorovich type topologies on spaces of measures and convergence of barycenters”, *Commun. Pure Appl. Anal.* **22**:2 (2023), 597–612.
- [4] J.-J. Alibert, G. Bouchitté, and T. Champion, “A new class of costs for optimal transport planning”, *European J. Appl. Math.* **30**:6 (2019), 1229–1263.
- [5] L. Ambrosio, E. Brué, and D. Semola, *Lectures on optimal transport*, Unitext, vol. 130, Springer, Cham 2021, ix+250 pp.
- [6] L. Ambrosio, M. Erbar, and G. Savaré, “Optimal transport, Cheeger energies and contractivity of dynamic transport distances in extended spaces”, *Nonlinear Anal.* **137** (2016), 77–134.
- [7] L. Ambrosio and N. Gigli, “A user’s guide to optimal transport”, *Modelling and optimisation of flows on networks*, Lecture Notes in Math., vol. 2062, Fond. CIME/CIME Found. Subser., Springer, Heidelberg 2013, pp. 1–155.
- [8] L. Ambrosio and A. Pratelli, “Existence and stability results in the L^1 theory of optimal transportation”, *Optimal transportation and applications* (Martina Franca 2001), Lecture Notes in Math., vol. 1813, Springer, Berlin 2003, pp. 123–160.
- [9] A. L. Andrianov, “The development of linear programming in L. V. Kantorovich’s papers of the 1930s–1950s”, *Istor. Mat. Issled. Ser. 2*, vol. 15(50), Yanus-K, Moscow 2014, pp. 25–40. (Russian)
- [10] S. Athreya, W. Löhr, and A. Winter, “The gap between Gromov-vague and Gromov–Hausdorff-vague topology”, *Stochastic Process. Appl.* **126**:9 (2016), 2527–2553.
- [11] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder, “Adapted Wasserstein distances and stability in mathematical finance”, *Finance Stoch.* **24**:3 (2020), 601–632.
- [12] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder, “All adapted topologies are equal”, *Probab. Theory Related Fields* **178**:3-4 (2020), 1125–1172.
- [13] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and J. Wiesel, “Estimating processes in adapted Wasserstein distance”, *Ann. Appl. Probab.* **32**:1 (2022), 529–550.
- [14] J. Backhoff-Veraguas, M. Beiglböck, Yiqing Lin, and A. Zalashko, “Causal transport in discrete time and applications”, *SIAM J. Optim.* **27**:4 (2017), 2528–2562.

- [15] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer, “Existence, duality, and cyclical monotonicity for weak transport costs”, *Calc. Var. Partial Differential Equations* **58**:6 (2019), 203, 28 pp.
- [16] J. Backhoff-Veraguas and G. Pammer, “Applications of weak transport theory”, *Bernoulli* **28**:1 (2022), 370–394.
- [17] J. Backhoff-Veraguas and G. Pammer, “Stability of martingale optimal transport and weak optimal transport”, *Ann. Appl. Probab.* **32**:1 (2022), 721–752.
- [18] M. Barbie and A. Gupta, “The topology of information on the space of probability measures over Polish spaces”, *J. Math. Econom.* **52** (2014), 98–111.
- [19] D. Bartl, P. Cheridito, M. Kupper, and L. Tangpi, “Duality for increasing convex functionals with countably many marginal constraints”, *Banach J. Math. Anal.* **11**:1 (2017), 72–89.
- [20] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer, “Optimal and better transport plans”, *J. Funct. Anal.* **256**:6 (2009), 1907–1927.
- [21] M. Beiglböck and N. Juillet, “On a problem of optimal transport under marginal martingale constraints”, *Ann. Probab.* **44**:1 (2016), 42–106.
- [22] J.-D. Benamou, G. Carlier, and L. Nenna, “Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm”, *Numer. Math.* **142**:1 (2019), 33–54.
- [23] J. Bergin, “On the continuity of correspondences on sets of measures with restricted marginals”, *Econom. Theory* **13**:2 (1999), 471–481.
- [24] S. Bobkov and M. Ledoux, *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, Mem. Amer. Math. Soc., vol. 261, no. 1259, Amer. Math. Soc., Providence, RI 2019, v+126 pp.
- [25] V. I. Bogachev, *Measure theory*, *Regulyarnaya i Khaoticheskaya Dinamika*, Moscow–Izhevsk 2003, 544 pp., 576 pp.; English transl., vols. I, II, Springer-Verlag, Berlin 2007, xviii+500 pp., xiv+575 pp.
- [26] V. I. Bogachev, *Weak convergence of measures*, Math. Surveys Monogr., vol. 234, Amer. Math. Soc., Providence, RI 2018, xii+286 pp.
- [27] V. I. Bogachev, “On sequential properties of spaces of measures”, *Mat. Notes* **110**:3 (2021), 459–464; English transl. in *Math. Notes* **110**:3 (2021), 449–453.
- [28] V. I. Bogachev, “On approximation of measures by their finite-dimensional images”, *Finktsional. Anal. Prilozhen.* **55**:3 (2021), 75–81; English transl. in *Funct. Anal. Appl.* **55**:3 (2021), 236–241.
- [29] V. I. Bogachev, “Kantorovich problems with a parameter and density constraints”, *Sibirsk. Mat. Zh.* **63**:1 (2022), 42–57; English transl. in *Siberian Math. J.* **63**:1 (2022), 34–47.
- [30] V. I. Bogachev, A. N. Doledenok, and I. I. Malofeev, “The Kantorovich problem with a parameter and density constraints”, *Mat. Zametki* **110**:6 (2021), 922–926; English transl. in *Math. Notes* **110**:6 (2021), 952–955.
- [31] V. I. Bogachev and A. N. Kalinin, “A continuous cost function for which the minima in the Monge and Kantorovich problems are not equal”, *Dokl. Ross. Akad. Nauk* **463**:4 (2015), 383–386; English transl. in *Dokl. Math.* **92**:1 (2015), 452–455.
- [32] V. I. Bogachev, A. N. Kalinin, and S. N. Popova, “On the equality of values in the Monge and Kantorovich problems”, *Probability and Statistics*. 25, Zap. Nauchn. Sem. POMI, vol. 457, St. Petersburg Department of Steklov Mathematical Institute, St. Petersburg 2017, pp. 53–73; English transl. in *J. Math. Sci. (N. Y.)* **238**:4 (2019), 377–389.
- [33] V. I. Bogachev and A. V. Kolesnikov, “The Monge–Kantorovich problem: achievements, connections, and perspectives”, *Uspekhi Mat. Nauk* **67**:5(407) (2012), 3–110; English transl. in *Russian Math. Surveys* **67**:5 (2012), 785–890.

- [34] V. I. Bogachev and I. I. Malofeev, “Kantorovich problems and conditional measures depending on a parameter”, *J. Math. Anal. Appl.* **486**:1 (2020), 123883, 30 pp.
- [35] V. I. Bogachev and I. I. Malofeev, “Nonlinear Kantorovich problems with a parameter”, *Izv. Irkutsk. Gos. univ.* **41** (2022), 96–106. (Russian)
- [36] V. Bogachev and S. Popova, *Optimal transportation of measures with a parameter*, 2021, 14 pp., [arXiv:2111.13014](https://arxiv.org/abs/2111.13014).
- [37] V. I. Bogachev, S. N. Popova, and A. V. Rezbaev, “On nonlinear Kantorovich problems with density constraints”, *Moscow Math. J.* (to appear).
- [38] V. I. Bogachev and A. V. Rezbayev, “Existence of solutions to the nonlinear Kantorovich transportation problem”, *Mat. Zametki* **112**:3 (2022), 360–370; English transl. in *Math. Notes* **112**:3 (2022), 369–377.
- [39] V. I. Bogachev and A. V. Shaposhnikov, “Lower bounds for the Kantorovich distance”, *Dokl. Ross. Akad. Nauk* **460**:6 (2015), 631–633; English transl. in *Dokl. Math.* **91**:1 (2015), 91–93.
- [40] V. I. Bogachev, A. V. Shaposhnikov, and F.-Y. Wang, “Sobolev–Kantorovich inequalities under $CD(0, \infty)$ condition”, *Commun. Contemp. Math.* **24**:5 (2022), 2150027, 27 pp.
- [41] V. I. Bogachev, O. G. Smolyanov, and V. I. Sobolev, *Topological vector spaces and their applications*, Regulyarnaya i Khaotichaskaya Dinamika, Moscow–Izhavsk 2012, 584 pp.; English version, V. I. Bogachev and O. G. Smolyanov, *Topological vector spaces and their applications*, Springer Monogr. Math., Springer, Cham 2017, x+456 pp.
- [42] V. I. Bogachev, F.-Y. Wang, and A. V. Shaposhnikov, “Estimates of the Kantorovich norm on manifolds”, *Dokl. Ross. Akad. Nauk* **463**:6 (2015), 633–638; English transl. in *Dokl. Math.* **92**:1 (2015), 494–499.
- [43] V. I. Bogachev, F.-Y. Wang, and A. V. Shaposhnikov, “On inequalities relating the Sobolev and Kantorovich norms”, *Dokl. Ross. Akad. Nauk* **468**:2 (2016), 131–133; English transl. in *Dokl. Math.* **93**:3 (2016), 256–258.
- [44] W. Boyer, B. Brown, A. Loving, and S. Tammen, “Optimal transportation with constant constraint”, *Involve* **12**:1 (2019), 1–12.
- [45] Y. Brenier and D. Vorotnikov, “On optimal transport of matrix-valued measures”, *SIAM J. Math. Anal.* **52**:3 (2020), 2849–2873.
- [46] H. Brezis and P. Mironescu, “The Plateau problem from the perspective of optimal transport”, *C. R. Math. Acad. Sci. Paris* **357**:7 (2019), 597–612.
- [47] M. Brücknerhoff and N. Juillet, “Instability of martingale optimal transport in dimension $d \geq 2$ ”, *Electron. Commun. Probab.* **27** (2022), 24, 10 pp.
- [48] D. B. Bukin, “On the Monge and Kantorovich problems for distributions of diffusion processes”, *Math. Notes* **96**:5 (2014), 864–870.
- [49] D. B. Bukin, “On the Kantorovich problem for nonlinear images of the Wiener measure”, *Mat. Zametki* **100**:5 (2016), 682–688; English transl. in *Math. Notes* **100**:5 (2016), 660–665.
- [50] D. B. Bukin and E. P. Krugova, “Transportation costs for optimal and triangular transformations of Gaussian measures”, *Theory Stoch. Process.* **23**:2 (2018), 21–32.
- [51] G. Buttazzo, T. Champion, and L. De Pascale, “Continuity and estimates for multimarginal optimal transportation problems with singular costs”, *Appl. Math. Optim.* **78**:1 (2018), 185–200.
- [52] C. Castaing, P. Raynaud de Fitte, and M. Valadier, *Young measures on topological spaces. With applications in control theory and probability theory*, Math. Appl., vol. 571, Kluwer Acad. Publ., Dordrecht 2004, xii+320 pp.

- [53] H.-B. Chen and J. Niles-Weed, “Asymptotics of smoothed Wasserstein distances”, *Potential Anal.* **56**:4 (2022), 571–595.
- [54] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum, “On the matrix Monge–Kantorovich problem”, *European J. Appl. Math.* **31**:4 (2020), 574–600.
- [55] Y. Chen, T. T. Georgiou, and M. Pavon, “On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint”, *J. Optim. Theory Appl.* **169**:2 (2016), 671–691.
- [56] Y. Chen, T. T. Georgiou, and M. Pavon, “Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge”, *SIAM Rev.* **63**:2 (2021), 249–313.
- [57] Y. Chen, T. T. Georgiou, and A. Tannenbaum, “Vector-valued optimal mass transport”, *SIAM J. Appl. Math.* **78**:3 (2018), 1682–1696.
- [58] P. Cheridito, M. Kiiski, D. J. Prömel, and H. M. Soner, “Martingale optimal transport duality”, *Math. Ann.* **379**:3–4 (2021), 1685–1712.
- [59] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, “Unbalanced optimal transport: dynamic and Kantorovich formulations”, *J. Funct. Anal.* **274**:11 (2018), 3090–3123.
- [60] K. J. Ciosmak, “Optimal transport of vector measures”, *Calc. Var. Partial Differential Equations* **60**:6 (2021), 230, 22 pp.
- [61] C. Clason, D. A. Lorenz, H. Mahler, and B. Wirth, “Entropic regularization of continuous optimal transport problems”, *J. Math. Anal. Appl.* **494**:1 (2021), 124432, 22 pp.
- [62] M. Colombo and S. Di Marino, “Equality between Monge and Kantorovich multimarginal problems with Coulomb cost”, *Ann. Mat. Pura Appl.* (4) **194**:2 (2015), 307–320.
- [63] J. Dedecker, C. Prieur, and P. Raynaud De Fitte, “Parametrized Kantorovich–Rubinštein theorem and application to the coupling of random variables”, *Dependence in probability and statistics*, Lect. Notes Stat., vol. 187, Springer, New York 2006, pp. 105–121.
- [64] A. N. Doledenok, “On a Kantorovich problem with a density constraint”, *Mat. Zametki* **104**:1 (2018), 45–55; English transl. in *Math. Notes* **104**:1 (2018), 39–47.
- [65] R. Engelking, *General topology*, Transl. from the Polish, 2nd ed., Sigma Ser. Pure Math., vol. 6, Heldermann Verlag, Berlin 1989, viii+529 pp.
- [66] A. Figalli and F. Glaudo, *An invitation to optimal transport, Wasserstein distances, and gradient flows*, EMS Textbk. Math., EMS Press, Berlin 2021, vi+136 pp.
- [67] G. Friesecke, “A simple counterexample to the Monge ansatz in multimarginal optimal transport, convex geometry of the set of Kantorovich plans, and the Frenkel–Kontorova model”, *SIAM J. Math. Anal.* **51**:6 (2019), 4332–4355.
- [68] G. Friesecke, D. Matthes, and B. Schmitzer, “Barycenters for the Hellinger–Kantorovich distance over \mathbb{R}^d ”, *SIAM J. Math. Anal.* **53**:1 (2021), 62–110.
- [69] A. Galichon, *Optimal transport methods in economics*, Princeton Univ. Press, Princeton, NJ 2016, xii+170 pp.
- [70] I. Gentil, C. Léonard, and L. Ripani, “Dynamical aspects of the generalized Schrödinger problem via Otto calculus – a heuristic point of view”, *Rev. Mat. Iberoam.* **36**:4 (2020), 1071–1112.
- [71] A. Gerolin, A. Kausamo, and T. Rajala, “Nonexistence of optimal transport maps for the multimarginal repulsive harmonic cost”, *SIAM J. Math. Anal.* **51**:3 (2019), 2359–2371.

- [72] M. Ghossoub and D. Saunders, “On the continuity of the feasible set mapping in optimal transport”, *Econ. Theory Bull.* **9**:1 (2021), 113–117.
- [73] N. Ghossoub, Y.-H. Kim, and T. Lim, “Structure of optimal martingale transport plans in general dimensions”, *Ann. Probab.* **47**:1 (2019), 109–164.
- [74] N. Ghossoub and B. Maurey, “Remarks on multi-marginal symmetric Monge–Kantorovich problems”, *Discrete Contin. Dyn. Syst.* **34**:4 (2014), 1465–1480.
- [75] N. A. Gladkov, A. V. Kolesnikov, and A. P. Zimin, “On multistochastic Monge–Kantorovich problem, bitwise operations, and fractals”, *Calc. Var. Partial Differential Equations* **58**:5 (2019), 173, 33 pp.
- [76] N. A. Gladkov, A. V. Kolesnikov, and A. P. Zimin, “The multistochastic Monge–Kantorovich problem”, *J. Math. Anal. Appl.* **506**:2 (2022), 125666, 82 pp.
- [77] N. A. Gladkov and A. P. Zimin, “An explicit solution for a multimarginal mass transportation problem”, *SIAM J. Math. Anal.* **52**:4 (2020), 3666–3696.
- [78] J. Goubault-Larrecq, “Kantorovich–Rubinstein quasi-metrics I: Spaces of measures and of continuous valuations”, *Topology Appl.* **295** (2021), 107673, 37 pp.
- [79] F. de Gournay, J. Kahn, and L. Lebrat, “Differentiation and regularity of semi-discrete optimal transport with respect to the parameters of the discrete measure”, *Numer. Math.* **141**:2 (2019), 429–453.
- [80] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, “Kantorovich duality for general transport costs and applications”, *J. Funct. Anal.* **273**:11 (2017), 3327–3405.
- [81] C. Griessler, “ C -cyclical monotonicity as a sufficient criterion for optimality in the multimarginal Monge–Kantorovich problem”, *Proc. Amer. Math. Soc.* **146**:11 (2018), 4735–4740.
- [82] M. Huesmann and D. Trevisan, “A Benamou–Brenier formulation of martingale optimal transport”, *Bernoulli* **25**:4A (2019), 2729–2757.
- [83] M. Iacobelli, “A new perspective on Wasserstein distances for kinetic problems”, *Arch. Ration. Mech. Anal.* **244**:1 (2022), 27–50.
- [84] L. Kantorovitch, “On the translocation of masses”, *Dokl. Akad. Nauk SSSR* **37**:7-8 (1942), 227–229; English transl. in *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **37** (1942), 199–201.
- [85] L. V. Kantorovich, *Works on mathematics and economics*, Selected works, Nauka, Novosibirsk 2011, 760 pp. (Russian)
- [86] L. V. Kantorovich and G. P. Akilov, *Functional analysis*, 2nd ed., Nauka, Moscow 1977, 742 pp.; English transl., Pergamon Press, Oxford–Elmsford, N. Y. 1982, xiv+589 pp.
- [87] L. V. Kantorovich and G. Sh. Rubinshtein, “On a function space and some extremum problems”, *Dokl. Akad. Nauk SSSR* **115**:6 (1957), 1058–1061. (Russian)
- [88] L. V. Kantorovich and G. Sh. Rubinshtein, “On a space of completely additive functions”, *Vestn. Leningr. Univ.* **13**:7 (1958), 52–59. (Russian)
- [89] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov, “A new optimal transport distance on the space of finite Radon measures”, *Adv. Differential Equations* **21**:11-12 (2016), 1117–1164.
- [90] J. Korman and R. J. McCann, “Insights into capacity-constrained optimal transport”, *Proc. Natl. Acad. Sci. USA* **110**:25 (2013), 10064–10067.
- [91] J. Korman and R. J. McCann, “Optimal transportation with capacity constraints”, *Trans. Amer. Math. Soc.* **367**:3 (2015), 1501–1521.

- [92] J. Korman, R. J. McCann, and C. Seis, “Dual potentials for capacity constrained optimal transport”, *Calc. Var. Partial Differential Equations* **54**:1 (2015), 573–584.
- [93] J. Korman, R. J. McCann, and C. Seis, “An elementary approach to linear programming duality with application to capacity constrained transport”, *J. Convex Anal.* **22**:3 (2015), 797–808.
- [94] V. V. Kozlov, “The Monge problem of ‘piles and holes’ on the torus and the problem of small denominators”, *Sibirsk. Mat. Zh.* **59**:6 (2018), 1370–1374; English transl. in *Siberian Math. J.* **59**:6 (2018), 1090–1093.
- [95] D. Kramkov and Y. Xu, “An optimal transport problem with backward martingale constraints motivated by insider trading”, *Ann. Appl. Probab.* **32**:1 (2022), 294–326.
- [96] S. Kuksin, V. Nersesyan, and A. Shirikyan, “Exponential mixing for a class of dissipative PDEs with bounded degenerate noise”, *Geom. Funct. Anal.* **30**:1 (2020), 126–187.
- [97] R. Lassalle, “Causal transport plans and their Monge–Kantorovich problems”, *Stoch. Anal. Appl.* **36**:3 (2018), 452–484.
- [98] C. Léonard, “From the Schrödinger problem to the Monge–Kantorovich problem”, *J. Funct. Anal.* **262**:4 (2012), 1879–1920.
- [99] C. Léonard, “A survey of the Schrödinger problem and some of its connections with optimal transport”, *Discrete Contin. Dyn. Syst.* **34**:4 (2014), 1533–1574.
- [100] V. L. Levin and A. A. Milyutin, “The problem of mass transfer with a discontinuous cost function and a mass statement of the duality problem for convex extremal problems”, *Uspekhi Mat. Nauk* **34**:3(207) (1979), 3–68; English transl. in *Russian Math. Surveys* **34**:3 (1979), 1–78.
- [101] M. Liero, A. Mielke, and G. Savaré, “Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures”, *Invent. Math.* **211**:3 (2018), 969–1117.
- [102] A. A. Lipchius, “A note on the equality in the Monge and Kantorovich problems”, *Teor. Veroyatn. Primenen.* **50**:4 (2005), 779–782; English transl. in *Theory Probab. Appl.* **50**:4 (2006), 689–693.
- [103] C. Liu and A. Neufeld, “Compactness criterion for semimartingale laws and semimartingale optimal transport”, *Trans. Amer. Math. Soc.* **372**:1 (2019), 187–231.
- [104] W. Löhner, “Equivalence of Gromov–Prohorov- and Gromov’s \square_λ -metric on the space of metric measure spaces”, *Electron. Commun. Probab.* **18** (2013), 17, 10 pp.
- [105] D. A. Lorenz, P. Manns, and C. Meyer, “Quadratically regularized optimal transport”, *Appl. Math. Optim.* **83**:3 (2021), 1919–1949.
- [106] A. Marchese, A. Massaccesi, S. Stuard, and R. Tione, “A multi-material transport problem with arbitrary marginals”, *Calc. Var. Partial Differential Equations* **60**:3 (2021), 88, 49 pp.
- [107] R. J. McCann and L. Rifford, “The intrinsic dynamics of optimal transport”, *J. Éc. Polytech. Math.* **3** (2016), 67–98.
- [108] T. Mikami, *Stochastic optimal transportation. Stochastic control with fixed marginals*, SpringerBriefs Math., Springer, Singapore 2021, xi+121 pp.
- [109] A. Moameni, “A characterization for solutions of the Monge–Kantorovich mass transport problem”, *Math. Ann.* **365**:3-4 (2016), 1279–1304.
- [110] A. Moameni and B. Pass, “Solutions to multi-marginal optimal transport problems concentrated on several graphs”, *ESAIM Control Optim. Calc. Var.* **23**:2 (2017), 551–567.

- [111] A. Moameni and L. Rifford, “Uniquely minimizing costs for the Kantorovitch problem”, *Ann. Fac. Sci. Toulouse Math.* (6) **29**:3 (2020), 507–563.
- [112] A. Neufeld and J. Sester, “On the stability of the martingale optimal transport problem: a set-valued map approach”, *Statist. Probab. Lett.* **176** (2021), 109131, 7 pp.
- [113] B. Pass, “Multi-marginal optimal transport: theory and applications”, *ESAIM Math. Model. Numer. Anal.* **49**:6 (2015), 1771–1790.
- [114] B. W. Pass and A. Vargas-Jiménez, “Multi-marginal optimal transportation problem for cyclic costs”, *SIAM J. Math. Anal.* **53**:4 (2021), 4386–4400.
- [115] M. Petrache, “Cyclically monotone non-optimal N -marginal transport plans and Smirnov-type decompositions for N -flows”, *ESAIM Control Optim. Calc. Var.* **26** (2020), 120, 11 pp.
- [116] G. Ch. Pflug and A. Pichler, *Multistage stochastic optimization*, Springer Ser. Oper. Res. Financ. Eng., Springer, Cham 2014, xiv+301 pp.
- [117] A. Pratelli, “On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation”, *Ann. Inst. H. Poincaré Probab. Statist.* **43**:1 (2007), 1–13.
- [118] S. T. Rachev and L. Rüschendorf, *Mass transportation problems*, vol. I: *Theory*, Probab. Appl. (N. Y.), Springer-Verlag, New York 1998, xxvi+508 pp.; vol. II: *Applications*, xxvi+430 pp.
- [119] P. Rigo, “A note on duality theorems in mass transportation”, *J. Theoret. Probab.* **33**:4 (2020), 2337–2350.
- [120] F. Santambrogio, *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*, Progr. Nonlinear Differential Equations Appl., vol. 87, Birkhäuser/Springer, Cham 2015, xxvii+353 pp.
- [121] A. Savchenko and M. Zarichnyi, “Correspondences of probability measures with restricted marginals”, *Proc. Intern. Geom. Center* **7**:4 (2014), 34–39.
- [122] B. Schmitzer and B. Wirth, “A framework for Wasserstein-1-type metrics”, *J. Convex Anal.* **26**:2 (2019), 353–396.
- [123] T. Shioya, *Metric measure geometry. Gromov’s theory of convergence and concentration of metrics and measures*, IRMA Lect. Math. Theor. Phys., vol. 25, EMS Publishing House, Zürich 2016, xi+182 pp.
- [124] V. M. Tikhomirov, “Leonid Vital’evich Kantorovich (to the centenary of his birth)”, *Istor. Mat. Issled. Ser. 1*, vol. 15(50), Publishing house ‘Janus-K’, M. 2014, pp. 16–24. (Russian)
- [125] A. M. Vershik, “Long history of the Monge–Kantorovich transportation problem”, *Math. Intelligencer* **35**:4 (2013), 1–9.
- [126] A. M. Vershik, S. S. Kutateladze, and S. P. Novikov, “Leonid Vital’evich Kantorovich (on the 100th anniversary of his birth)”, *Uspekhi Mat. Nauk* **67**:3(405) (2012), 185–191; English transl. in *Russian Math. Surveys* **67**:3 (2012), 589–597.
- [127] A. M. Vershik, P. B. Zatitskii, and F. V. Petrov, “Virtual continuity of measurable functions and its applications”, *Uspekhi Mat. Nauk* **69**:6(420) (2014), 81–114; English transl. in *Russian Math. Surveys* **69**:6 (2014), 1031–1063.
- [128] C. Villani, *Topics in optimal transportation*, Grad. Stud. Math., vol. 58, Amer. Math. Soc., Providence, RI 2003, xvi+370 pp.
- [129] C. Villani, *Optimal transport. Old and new*, Grundlehren Math. Wiss., vol. 338, Springer-Verlag, Berlin 2009, xxii+973 pp.
- [130] J. Wiesel, *Continuity of the martingale optimal transport problem on the real line*, 2022 (v1 – 2019), 46 pp., arXiv: 1905.04574.

- [131] G. Wolansky, *Optimal transport. A semi-discrete approach*, De Gruyter Ser. Nonlinear Anal. Appl., vol. 37, De Gruyter, Berlin 2021, xii+208 pp.
- [132] D. A. Zaev, “On the Monge–Kantorovich problem with additional linear constraints”, *Mat. Zametki* **98**:5 (2015), 664–683; English transl. in *Math. Notes* **98**:5 (2015), 725–741.
- [133] D. A. Zaev, “On ergodic decompositions related to the Kantorovich problem”, *Representation theory, dynamical systems, combinatorial methods. XXVI*, Zap. Nauchn. Sem. POMI, vol. 437, St.-Petersburg Department of Steklov Mathematical Institute, St. Petersburg 2015, pp. 100–130; English transl. in *J. Math. Sci. (N. Y.)* **216**:1 (2016), 65–83.
- [134] Xicheng Zhang, “Stochastic Monge–Kantorovich problem and its duality”, *Stochastics* **85**:1 (2013), 71–84.

Vladimir I. Bogachev

Lomonosov Moscow State University;

National Research University

“Higher School of Economics”

E-mail: vibogach@mail.ru

Received 27/JUL/22

Translated by THE AUTHOR