OPTIMAL TRANSPORTATION OF PROCESSES WITH INFINITE KANTOROVICH DISTANCE. INDEPENDENCE AND SYMMETRY.

ALEXANDER V. KOLESNIKOV AND DANILA A. ZAEV

Abstract. We consider probability measures on $\mathbb{R}^\infty$ and study optimal transportation mappings for the case of infinite Kantorovich distance. Our examples include 1) quasi-product measures, 2) measures with certain symmetric properties, in particular, exchangeable and stationary measures. We show in the latter case that existence problem for optimal transportation is closely related to ergodicity of the target measure. In particular, we prove existence of the symmetric optimal transportation for a certain class of stationary Gibbs measures.

MSC 49J27, 28C20

1. Introduction

Let us consider two Borel probability measures μ, ν on $\mathbb{R}^d$. The central result (Brenier theorem) of the finite-dimensional optimal transportation theory establishes under fairy general assumptions existence of the corresponding optimal transportation mapping T, which can be characterized by the following properties:

1) $T = \nabla \varphi$, where φ is a convex function

2) ν is the image of μ under T: $\nu = \mu \circ T^{-1}$.

The mapping T exists, in particular, when both measures are absolutely continuous and have finite second moments. The second assumption can be replaced by the weaker assumption of the finiteness of the corresponding Kantorovich distance $W_2(\mu, \nu)$ but it does not make much difference for the finite-dimensional problems. However, this difference becomes essential in the infinite-dimensional case.

It is well-known that the optimal transportation mapping T solves the so-called Monge problem, meaning that T gives minimum to the functional

$$\int_{\mathbb{R}^d} \|r(x) - x\|^2 d\mu(x)$$

among of the mappings $r\colon \mathbb{R}^d \mapsto \mathbb{R}^d$ pushing forward μ onto ν; here $\|\cdot\|$ is the standard Euclidean norm. The corresponding minimal value coincides with the squared Kantorovich distance $W_2^2(\mu, \nu)$.

Now let us consider a couple of measures on an infinite-dimensional linear space X; to avoid unessential technicalities, we will assume everywhere that $X = \mathbb{R}^\infty$.

Key words and phrases. Monge–Kantorovich problem, optimal transportation, Kantorovich duality, Gaussian measures, Gibbs measures, log-concave measures, exchangeability, stationarity, ergodicity, transportation inequalities, entropy, and Kullback-Leibler distance;

The first named author was supported by RFBR project 14-01-00237 and the DFG project CRC 701. This study (research grant No 14-01-0056) was supported by The National Research University-Higher School of Economics' Academic Fund Program in 2014/2015. The second author was partially supported by AG Laboratory NRU-HSE, RF government grant, ag. 11.G34.31.0023.

We deal throughout with the standard Hilbert norm

$$\|x\|^2 := \|x\|^2_{l^2} = \sum_{i=1}^{\infty} x_i^2,$$

which takes infinite value almost everywhere with respect to most of the measures we are interested in.

What is a natural analog of the Brenier theorem in this setting? To understand the situation better let us consider the Gaussian model.

Example 1.1. Let $\gamma = \prod_{i=1}^{\infty} \gamma_i = \prod_{i=1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \ dx_i$ be the standard Gaussian product measure on $\mathbb{R}^\infty$ and $H = l^2$ be the corresponding Cameron–Martin space. More generally, one can consider any abstract Wiener space.

The optimal transportation problem is well-understood for the case of measures μ and ν which are absolutely continuous with respect to γ. The most general results were obtained in [12] (another approach has been developed in [15]). In particular, for a broad class of probability measures $f \cdot \gamma$ absolutely continuous w.r.t. γ there exists a transportation mapping $T(x) = x + \nabla\varphi(x)$ minimizing the cost

$$\int \|T(x) - x\|^2_{l^2} \ d\gamma$$

and pushing forward γ onto $f \cdot \gamma$. Analogously, there exists a transportation mapping pushing forward $f \cdot \gamma$ onto γ. The gradient operator ∇ is understood with respect to $\langle \cdot, \cdot \rangle_{l^2}$-scalar product.

It is known (this follows from the so-called Talagrand transportation inequality) that under assumption $\int f \log f \ d\gamma < \infty$ the Kantorovich distance between γ and $f \cdot \gamma$ is finite

$$W_2^2(\gamma, f \cdot \gamma) = \int \|T(x) - x\|^2_{l^2} \ d\gamma < \infty.$$

In particular, $\nabla\varphi(x) \in l^2$ for γ-almost all x. More on optimal transportation on the Wiener space, the corresponding Monge–Ampére equation, regularity issues, and transportation on other infinite-dimensional spaces see in [5], [6], [8], [11], and [10].

In this paper we study situation when the Kantorovich distance between measures is a priori **infinite**. This makes impossible in general to understand T as a solution to a certain minimization problem. Nevertheless, we have many good candidates to be called "optimal transportation" in many particular cases. The following example motivates our study.

Example 1.2. 1) Let $\mu = \prod_{i=1}^{\infty} \mu_i(dx_i)$, $\nu = \prod_{i=1}^{\infty} \nu_i(dx_i)$ be product probability measures. Assume that all μ_i have densities. Then there exists a mass transportation mapping T pushing forward μ onto ν which has the form

$$T(x) = (T_1(x_1), \cdots, T_i(x_i), \cdots),$$

where $T_i(x_i)$ is the one-dimensional optimal transportation pushing forward μ_i onto ν_i.

2) Let us consider the Gaussian measure μ which is a push-forward image of the standard Gaussian measure γ under a linear mapping $T(x) = Ax$ with A symmetric and positive. It is well-known (and can be obtained from the law of large numbers) that γ and μ are mutually singular even in the simplest case $A = 2 \cdot \mathrm{Id}$.

T is "optimal" because it is linear and given by a positive symmetric operator. Heuristically,

$$T(x) = \frac{1}{2}\nabla\langle Ax, x\rangle.$$

It is clear that in both cases T cannot be obtained as a minimizer of a functional of the type $\int \|T(x) - x\|^2_{l^2}\ d\mu$.

We state now the central problem of this paper.

Problem 1.3. Let μ and ν be two probability measures on $\mathbb{R}^\infty$. When does exist a transportation mapping T pushing forward μ onto ν which is "optimal" for the cost function $c(x,y) = \|x - y\|^2_{l^2}$?

In this paper we deal with two model situations.

Quasi-product measures.

We assume that both measures have densities with respect to product probability measures

$$\mu = f \cdot \mu_0,\ \ \nu = g \cdot \nu_0,$$

$$\mu_0 = \prod_{i=1}^{\infty} \mu_i(dx_i), \nu_0 = \prod_{i=1}^{\infty} \nu_i(dx_i).$$

Then the corresponding "optimal tranportation" is a small perturbation of the diagonal mapping, considered in Example 1.2.

Symmetric measures.

It is possible to give a meaning to the Monge–Kantorovich optimization problem if we restrict ourselves to a certain class of symmetric measures. In this paper we consider two types of symmetry: exchangeable measures (invariant with respect to finite permutations of coordinates) and stationary measures on $\mathbb{R}^\infty$ (invariant with respect to shifts of coordinates). Note that $\|x - y\|^2_{l^2}$ is symmetric with respect to both types of symmetry. More generally, let G be a group of linear operators which acts on $X = Y = \mathbb{R}^\infty$ and $X \times Y$: $x \to gx$, $(x,y) \to (gx, gy)$, $g \in G$ and preserves the cost function $c(x,y)$. We assume that every basic vector e_j can be obtained from any other e_i by action of this group: there exists $g \in G$ such that $e_i = ge_j$. Note that under these assumptions all the coordinates are identically distributed. This leads us to the following definition: given G-invariant marginals μ and ν we call π an optimal (symmetric, invariant) solution to the Monge–Kantorovich problem if π solves the Monge–Kantorovich problem

$$\int (x_1 - y_1)^2\ d\pi \to \min$$

among all of the measures which are invariant with respect to G. If there exists a mapping T such that its graph $\Gamma = \{x, T(x)\}$ satisfies $m(\Gamma) = 1$, we say that T is an optimal transportation mapping pushing forward μ onto ν.

The following counter-example, however, demonstrates that the optimal transportation may fail to exist by a quite simple reason.

Example 1.4. Let $\mu = \gamma$ be the standard Gaussian measure on $\mathbb{R}^\infty$ and

$$\nu = \frac{1}{2}(\gamma + \gamma_2)$$

be the average of γ and its homothetic image $\gamma_2 = \gamma \circ S^{-1}$, where $S(x) = 2x$. There is no any mass transportation T of μ to ν which commutes with any cylindrical rotation. Indeed, any mapping of such a type must have the form $T(x) =$

$g(x)(x_1, x_2, \cdots) = g(x) \cdot x$, where g is invariant with respect to any "rotation", in particular, with respect to any coordinate permutation. But any function g of this type is constant γ-a.e. This is a corollary of the Hewitt–Savage $0-1$ law. It is clear that there is no any mass transportation of this type for the given target measure.

There is a general principle behind of this simple example. Recall that a measure μ is called ergodic with respect to a group action G, if for every G-invariant set A one has either $\mu(A) = 1$ or $\mu(A) = 0$. It follows directly from the definition that *there does not exists a bijective mass transportation T pushing forwarg μ onto ν, such that $T \circ g = g \circ T$ for every $g \in G$, provided μ is G-ergodic but ν is not.*

This observation leads to the following problem.

Problem. Let G be a group of linear operators acting on $\mathbb{R}^\infty$ and preserving l_2-distance (model example: group of shifts). Let μ, ν be **ergodic** G-invariant measures. When does exist a transportation $T\colon \mathbb{R}^\infty \mapsto \mathbb{R}^\infty$ pushing forward μ onto ν, which commutes with G and gives minimum to the Monge functional $T \mapsto \int_{\mathbb{R}^\infty} (T_1(x) - x_1)^2 \ d\mu$?

Trivially, the ergodicity by itself is not sufficient for the affirmative answer to this problem. In addition to it, we need to have certain infinite-dimesional analogs of "absolute continuity" for the source measure μ.

We believe that the symmetric transportation problem must have deep and very interesting relation with the ergodic theory. The second named author studied the interplay between ergodic decompositions and transportation theory in [26]. Another interesting connection has been established in [3]. It was shown that the Birkhoff ergodic theorem implies equivalence between optimality and the so-called cyclical monotonicity property. The related problems on optimal transportation in symmetric settings have been considered in [22] (stationary processes), in [23] (symmetric measures on graphs), and in [19], [20], [9] (ergodic theory). Transportation problems with symmetries have been studied in [13], [21]. Further development of the duality theory for transportation problem with linear restriction has been obtained in [25].

The paper is organized as follows: in Section 2 we give preliminaries in transportation theory, ergodic theory, and recall some important results on log-concave measures. In Section 3 we establish sufficient conditions for existence of optimal transportation mappings which are obtained as a.e.-limits of finite-dimensional approximations. The applications of this result are obtained in Section 4. Here we prove existence of optimal transportation for a couple of measures having densities with respect to product measures. In Section 5 we discuss the invariant optimal transportation problem, consider examples and prove some basic facts. In Section 6 we briefly discuss Kantorovich duality for problem which is invariant with respect to the action of a group. In Section 7 we construct a non-trivial example of a symmetric optimal transportation T. Namely, we establish sufficient conditions for existence of T pushing forward a stationary measure into the standard Gaussian measure. Finally, we apply this result to a certain class of Gibbs measures.

2. Preliminaries

2.1. **Optimal transportation problem.**

Kantorovich problem. Given two probability measures μ and ν on the spaces X and Y respectively, and a cost function $c : X \times Y \mapsto \mathbb{R} \cup \{+\infty\}$ we are looking for the minimum of the functional

$$W_2^2(\mu, \nu) = \inf\Bigl\{ \int \|x - y\|^2 \ dm \colon m \in P(\mu, \nu)\Bigr\},$$

on the space $P(\mu, \nu)$ of probability measures with fixed projections: $Pr_X m = \mu, Pr_Y m = \nu$.

In the classical setup $X = Y = \mathbb{R}^n$, $c = |x - y|^2$ the solution m is supported on the graph of a mapping $T : \mathbb{R}^n \mapsto \mathbb{R}^n$:

$$m(\Gamma) = 1, \ \ \text{where} \ \ \Gamma = \{(x, T(x)), \ x \in \mathbb{R}^d\}.$$

(see [1], [7], [24].). The functional $W_2(\mu, \nu)$ is a distance in the space of probability measures. In what follows we call it the Kantorovich distance. The mapping T is called optimal transportation of μ onto ν.

Another well-known fact which will be used throughout the paper is the following relation called the Kantorovich duality:

$$W_2(\mu, \nu) = -\frac{1}{2} J(\varphi, \psi),$$

where

$$J(\varphi, \psi) = \inf_{\varphi,\psi}\Bigl\{ \int \Bigl(\varphi(x) - \frac{|x|^2}{2}\Bigr) \ d\mu + \int \Bigl(\psi(y) - \frac{|y|^2}{2}\Bigr) \ d\nu, \ \ \varphi(x) + \psi(y) \geq \langle x, y\rangle \Bigr\},$$

where the infimum is taken over couples of integrable Borel functions $\varphi(x), \psi(y)$. The function φ in the dual problem coincides with the potential generating the transportation mapping

$$T = \nabla \varphi.$$

2.2. **Ergodic decomposition.** Given a Borel transformation $S : X \mapsto X$ of the space X we call a Borel probability measure μ ergodic if any S-invariant measurable set A has the property $\mu(A) = 1$ or $\mu(A) = 0$. A similar terminology is used if instead of a single mapping S we deal with a family G of transformations.

The ergodic G-invariant measures are extreme points of the set of all G-invariant measures, hence any G-invariant measure can be represented as the average of G-invariant ergodic measures. The famous de Finetti theorem establishes decomposition of this type for a class of exchangeable measures, i.e. measures, invariant with respect to a permutation of a finite number of coordinates.

Theorem 2.1. *Let $\mathcal{P}$ be the space of Borel probability measures on $\mathbb{R}$ equipped with the weak topology. Then for every Borel exchangeable μ on $\mathbb{R}^\infty$ there exists a Borel probability measure Π on $\mathcal{P}$ such that*

$$\mu(B) = \int m^\infty(B) \Pi(dm),$$

for every Borel $B \subset \mathbb{R}^\infty$.

Yet another example of the ergodic decomposition where a precise description is possible is given by rotationally invariant measures (see Example 5.9).

2.3. **Log-concave measures and functional inequalities.** We recall that a probability measure μ on $\mathbb{R}^n$ is called log-concave if it has the form $e^{-V} \cdot \mathcal{H}^k|_L$, where $\mathcal{H}^k$ is the k-dimensional Hausdorff measure, $k \in \{0, 1, \cdots, n\}$, L is an affine subspace, and V is a convex function.

In what follows we consider uniformly log-concave measures. Roughly speaking, these are the measures with potential V satisfying

$$V(x) - V(y) - \langle \nabla V(y), x - y \rangle \geq \frac{K}{2}|x-y|^2,$$

which is equivalent to $D^2 V \geq K \cdot \mathrm{Id}$ in the smooth (finite-dimensional) case. Here K is a positive constant.

More precisely, we say that a probability measure μ is K-uniformly log-concave ($K > 0$) if for any $\varepsilon > 0$ the measure $\hat{\mu} = \frac{1}{Z} e^{\frac{K-\varepsilon}{2}|x|^2} \cdot \mu$ is log-concave for a suitable renormalization factor Z. It is well-known (C. Borell) that the projections of log-concave measures are log-concave (this is in fact a corollary of the Brunn-Minkowski theorem). It can be easily checked that the uniform log-concavity is preserved by projections as well. We can extend this notion to the infinite-dimensional case. Namely, we call a probability measure μ on a locally convex space X log-concave (K-uniformly log-concave with $K > 0$) if its images $\mu \circ l^{-1}$, $l \in X^*$ under linear continuous functionals are all log-concave (K-uniformly log-concave with $K > 0$).

Throughout the paper we apply the following estimate (see [15], [16]), which generalizes the famous Talagrand transportation inequality.

Theorem 2.2. *(***Generalized Talagrand inequality.***) Let m be a K-uniformly log-concave probability measure with some $K > 0$. Then for any couple of probability measures $\mu = e^{-V}\, dx$, $\nu = e^{-W}\, dx$ and the corresponding optimal mappings $\nabla \varphi_\mu$, $\nabla \varphi_\nu$, pushing forward μ, ν onto m respectively, one has the following estimate*

$$\mathrm{Ent}_\nu\Bigl(\frac{\mu}{\nu}\Bigr) = \int \log \frac{d\mu}{d\nu}\, d\mu = \int (W - V)\, d\mu \geq \frac{K}{2} \int |\nabla \varphi_\mu - \nabla \varphi_\nu|^2\, d\mu.$$

Another result used in the paper is the Cafarelli's contraction theorem. Here is the version from [16] (see also [17]).

Theorem 2.3. **(Caffarelli contraction theorem).** *Let $\nabla \Phi$ be the optimal transportation of the probability measure $\mu = e^{-V} dx$ into $\nu = e^{-W} dx$. Assume that for some positive c, C one has $D^2 V \leq C \cdot \mathrm{Id}$, $D^2 W \geq c \cdot \mathrm{Id}$. Then $\nabla \Phi$ is Lipschitz with $\|\nabla \Phi\|_{Lip} \leq \sqrt{\frac{C}{c}}$.*

The quantity $\mathrm{Ent}_\nu\Bigl(\frac{\mu}{\nu}\Bigr)$ is called the relative entropy or the Kullback-Leibler distance between μ and ν.

3. Sufficient condition for existence of limits of finite-dimensional optimal mappings

3.1. **Preliminary finite-dimensional estimates.** Let μ and ν be probability measures on $\mathbb{R}^d$ and $T(x) = \nabla \varphi(x)$ be the optimal transportation mapping pushing forward μ onto ν. Let us denote by μ_v the images of μ under the shifts $x \mapsto x + v$, $v \in \mathbb{R}^d$.

It will be assumed throughout that μ_v have densities with respect to μ:

$$\frac{d\mu_v}{d\mu} = e^{\beta_v}.$$

Lemma 3.1. *For every $p, q \ge 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, $\varepsilon \ge 0$, and $e \in \mathbb{R}^d$*

$$\int |\varphi(x+te) - \varphi(x)|^{1+\varepsilon} \ \ d\mu \le t^{1+\varepsilon} \| \ |\langle x, e\rangle|^{1+\varepsilon}\|_{L^p(\nu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}}\|_{L^q(\mu)}.$$

$$\int \bigl(\varphi(x+te) - \varphi(x) - t\partial_e \varphi(x)\bigr) \ d\mu \le t \|\langle x, e\rangle\|_{L^p(\nu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}} - 1\|_{L^q(\mu)}.$$

Proof. One has $\varphi(x+te) - \varphi(x) = \int_0^t \partial_e \varphi(x+se) \ ds$. Hence

$$\begin{aligned}
&\int |\varphi(x+te) - \varphi(x)|^{1+\varepsilon} \ d\mu \le t^{\varepsilon} \int \int_0^t |\partial_e \varphi|^{1+\varepsilon}(x+se) \ ds \ d\mu \\
&\quad = t^{\varepsilon} \int_0^t \Bigl[\int |\partial_e \varphi|^{1+\varepsilon} e^{\beta_{se}} \ \ d\mu\Bigr] \ ds \le t^{1+\varepsilon} \| |\partial_e \varphi|^{1+\varepsilon}\|_{L^p(\mu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}}\|_{L^q(\mu)} \\
&\quad = t^{1+\varepsilon} \| \ |\langle x, e\rangle|^{1+\varepsilon}\|_{L^p(\nu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}}\|_{L^q(\mu)}.
\end{aligned}$$

Applying the same arguments one gets

$$\begin{aligned}
&\int \bigl(\varphi(x+te) - \varphi(x) - t\partial_e \varphi(x)\bigr) \ d\mu = \int \int_0^t (\partial_e \varphi(x+se) - \partial_e \varphi(x)) \ ds \ d\mu \\
&\quad = \int \Bigl[\int_0^t (e^{\beta_{se}} - 1) \ ds\Bigr] \partial_e \varphi(x) \ d\mu \le t^{\frac{1}{p}} \|\partial_e \varphi\|_{L^p(\mu)} \Bigl[\int \int_0^t |e^{\beta_{se}} - 1|^q \ ds \ d\mu\Bigr]^{\frac{1}{q}}.
\end{aligned}$$

The desired estimate follows from the change of variables formula and trivial uniform bounds. □

In addition, we will apply the following elementary Lemma.

Lemma 3.2. *Assume that a sequence $\{T_n\}$ of measurable mappings $T_n \colon \mathbb{R}^\infty \to \mathbb{R}^\infty$ converges to a mapping T in the following sense: for every e_i, $\lim_n \langle T_n, e_i\rangle = \langle T, e_i\rangle$ in measure with respect to μ. Then the measures $\{\mu \circ T_n^{-1}\}$ converge weakly to $\mu \circ T^{-1}$.*

3.2. **Existence theorem.** We consider a couple of Borel probability measures μ and ν on $\mathbb{R}^\infty$, where $\mathbb{R}^\infty$ is the space of all real sequences: $\mathbb{R}^\infty = \prod_{i=1}^\infty \mathbb{R}_i$. We deal with the standard coordinate system $x = (x_1, x_2, \cdots, x_n, \cdots)$ and the standard basis vectors $e_i = (\delta_{ij})$. The projection on the first n coordinates will be denoted by P_n: $P_n(x) = (x_1, \cdots, x_n)$. We use notations $\|x\|$, $\langle x, y\rangle$ for the Hilbert space norm and inner product: $\|x\| = \sum_{i=1}^\infty x_i^2$, $\langle x, y\rangle = \sum_{i=1}^\infty x_i y_i$. We use notation $\mathbb{E}_\mu^n$ for the conditional expectation with respect to μ and the σ-algebra generated by $x_1, \cdots, x_n$. For any product measure $P = \prod_{i=1}^\infty p_i(x_i) \ dx_i$ its projection $P_n = P \circ P_n^{-1}$ has the form $\prod_{i=1}^n p_i(x_i) \ dx_i$ and the projection $(f \cdot P) \circ P_n^{-1} = f_n \cdot P_n$ of the measure $f \cdot P$ satisfies $f_n = \mathbb{E}_P^n f$. Everywhere below we agree that every cylindrical function $f = f(x_1, \cdots, x_n)$ can be extended to $\mathbb{R}^\infty$ by the formula $x \to f_n(P_n x)$.

It will be assumed throughout the paper that the shifts of μ along any vector $v = te_i$ are absolutely continuous with respect to μ:

$$\frac{d\mu_v}{d\mu} = e^{\beta_v}.$$

In Section 3, moreover, the following assumption holds.

Assumption (A). For every basic vector $e = e_i$ there exist $p \geq 1$, $q \geq 1$, satisfying $\frac{1}{p} + \frac{1}{q} = 1$, and $\varepsilon > 0$ such that

$$\int |\langle x, e\rangle|^{(1+\varepsilon)p} \, d\nu < \infty$$

and

$$p(t) = \sup_{0 \leq s \leq t} \int |e^{\beta_{se}} - 1|^q \, d\mu$$

satisfies $\lim_{t\to 0} p(t) = 0$.

Let $\mu_n = \mu \circ P_n^{-1}(x)$, $\nu_n = \nu \circ P_n^{-1}(y)$ be the projections of μ, ν. For every $v = te_i$ let us set

$$\frac{d(\mu_n)_v}{d\mu_n} = e^{\beta_v^{(n)}}.$$

It is easy to check that the projections of μ, ν satisfy Assumption (A).

Lemma 3.3. *For every $n \in \mathbb{N}$ and every $e = e_i$ one has*

$$\int |\langle P_n(x), e\rangle|^p \, d\nu_n \leq \int |\langle x, e\rangle|^p \, d\nu, \quad \int |e^{\beta_e^{(n)}} - 1|^q \, d\mu_n \leq \int |e^{\beta_e} - 1|^q \, d\mu.$$

Proof. The first estimate is trivial. To prove the second one, let us note that $e^{\beta_v^{(n)}} = \mathbb{E}_\mu^n e^{\beta_v}$. The claim follows from the Jensen inequality and convexity of the function $t \to |t-1|^q$. □

We denote by π_n the optimal transportation plan for the couple (μ_n, ν_n). Let $\varphi_n(x)$ and $\psi_n(y)$ solve the dual Kantorovich problem. Let us recall that $\nabla\varphi_n$ ($\nabla\psi_n$) is the optimal transportation mapping sending μ_n to ν_n (ν_n to μ_n). One has

$$\varphi_n(x) + \psi_n(y) \geq \langle P_n x, P_n y\rangle$$

for every x, y. The equality is attained on the support of π_n. In particular,

$$\varphi_n(x) + \psi_n(\nabla\varphi_n(x)) = \langle P_n x, \nabla\varphi_n(x)\rangle.$$

It is easy to check that $\{\pi_n\}$ is a tight sequence. By the Prokhorov theorem one can extract a weakly convergent subsequence $\pi_{n_k} \to \pi$. Note that π_n **is not** the projection of π.

The main result if the section is the following theorem.

Theorem 3.4. *Assume that* **(A)** *is fulfilled and, in addition,*

$$F_n(x, y, 0, 0) = \varphi_n(x) + \psi_n(y) - \langle P_n x, P_n y\rangle \to 0$$

in measure with respect to π. Then there exists a mapping $T\colon \mathbb{R}^\infty \mapsto \mathbb{R}^\infty$ such that

$$T(x) = y$$

for π-almost all (x, y).

In what follows we will pass several time to subsequences and use for the new subsequences the same index n again, with the agreement that n takes values in another infinite set $\mathbb{N}' \subset \mathbb{N}$. Let us fix unit vectors e_i, e_j for some $i, j \in \mathbb{N}$ and consider the following sequence of non-negative functions:

$$F_n(x, y, t, s) = \varphi_n(x + te_i) + \psi_n(y + se_j) - \langle P_n(x + te_i), P_n(y + se_j)\rangle$$

with $n > i, n > j$.

Lemma 3.5. *There exists a $L^{1+\varepsilon}(\pi)$-weakly convergent subsequence*

$$\varphi_{n_k}(x+te_i) - \varphi_{n_k}(x) \to U(x).$$

The following relation holds for the limiting function $U(x)$:

$$\Big| \int U(x)\ d\mu - t \int \langle y, e_i\rangle\ d\nu \Big| \le Ctp(t).$$

Proof. Taking into account that $\int F_n(x,y,0,0)\ d\pi_n = 0$, one obtains

$$\int F_n(x,y,t,0)\ d\pi_n = \int F_n(x,y,t,0)\ d\pi_n - \int F_n(x,y,0,0)\ d\pi_n \ge 0.$$

Note that the right-hand side equals

$$\int (F_n(x,y,t,0) - F_n(x,y,0,0))\ d\pi_n = \int \big[\varphi_n(x+te_i) - \varphi_n(x) - t\langle y, e_i\rangle\big]\ d\pi_n.$$

Taking into account that the projection of π_n onto X coincides with μ_n and φ_n depends on the first n coordinates, one finally obtains that for $n > i$ the latter is equal to

$$\int \big[\varphi_n(x+te_i) - \varphi_n(x)\big]\ d\mu - t\int \langle y, e_i\rangle\ d\nu = \int \big[\varphi_n(x+te_i) - \varphi_n(x) - t\partial_{e_i}\varphi_n(x)\big]\ d\mu.$$

It follows from Lemma 3.1, Lemma 3.3 and Assumption (A) that

$$\Big| \int F_n(x,y,t,0)\ d\pi_n \Big| \le Ctp(t). \tag{1}$$

Since φ_n depends on a finite number of coordinates ($\le n$), one has

$$\int |\varphi_n(x+te_i) - \varphi_n(x)|^{1+\varepsilon}\ d\mu = \int |\varphi_n(x+te_i) - \varphi_n(x)|^{1+\varepsilon}\ d\mu_n.$$

Hence by Lemma 3.1

$$U_n(x) = \varphi_n(x+te_i) - \varphi_n(x) \in L^{1+\varepsilon}(\mu)$$

and, moreover, $\sup_n \|U_n\|_{L^{1+\varepsilon}(\mu)} < \infty$. Thus there exists function $U \in L^{1+\varepsilon}(\mu)$ such that for some subsequence n_k

$$\varphi_{n_k}(x+te_i) - \varphi_{n_k}(x) \to U(x)$$

weakly in $L^{1+\varepsilon}(\mu)$. Passing to the limit we obtain from (1) that

$$\Big| \int U(x)\ d\mu - t \int \langle y, e_i\rangle\ d\nu \Big| \le Ctp(t).$$

□

Lemma 3.6. *Assume that $F_n(x,y,0,0) \to 0$ in measure with respect to π. Then*

$$U(x) - t\langle y, e_i\rangle \ge 0$$

for π-almost all (x,y).

Proof. Note that

$$\big[\varphi_n(x+te_i) - \varphi_n(x) - t\langle y, e_i\rangle\big] + F_n(x,y,0,0) = \varphi_n(x+te_i) + \psi_n(y) - \langle P_n y, P_n(x+te_i)\rangle$$

is a non-negative function for every n. Since $F_n(x,y,0,0) \to 0$ in measure, there exists a subsequence (denoted again by F_n) which converges to zero π-almost everywhere. Since $f_n = \varphi_n(x+te_i) - \varphi_n(x) - t\langle y, e_i\rangle$ converges to $f = U(x) -$

$t\langle y, e_i\rangle$ weakly in $L^{1+\varepsilon}(\pi)$, one can assume (passing again to a subsequence) that $\frac{1}{N}\sum_{n=1}^N f_n \to f$ π-a.e. Since $f_n + F_n \geq 0$, this implies that $f \geq 0$ π-a.e. □

Proposition 3.7. *Assume that there exists a sequence of continuous functions*

$$f_n(x_1, \cdots, x_n), g_n(y_1, \cdots, y_n) \in L^1(\pi_n)$$

such that $G_n = f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i$ *has the following properties:*

1) $G_n \geq 0$,
2) $G_n \leq G_m, \quad \forall\ n \leq m, x, y \in \mathbb{R}^m$,
3) $\sup_n \int G_n \ d\pi_n < \infty$.

Then $F_n(x, y, 0, 0) \to 0$ *in* $L^1(\pi)$.

Proof. We start with the identity $\int F_n(x, y, 0, 0)\ d\pi_n = 0$ and rewrite it in the following way:

$$(2) \qquad 0 = \int (\varphi_n - f_n)\ d\mu + \int (\psi_n - g_n)\ d\nu + \int \big(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\big)\ d\pi_n.$$

Since φ_n, ψ_n are defined up to a constant, one can assume that $\int (\psi_n - g_n)\ d\nu = 0$. Thus $-\int (\varphi_n - f_n)\ d\mu = \int \big(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\big)\ d\pi_n$. It follows from 1) and 3) that the right-hand side is a bounded sequence of non-negative numbers. Passing to a subsequence we may assume that the right-hand side has a limit. It follows from the weak convergence $\pi_n \to \pi$ and the monotonicity property 2) that for every k

$$\begin{aligned}\underline{\lim}_n \int \big(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\big)\ d\pi_n &\geq \underline{\lim}_n \int \big(f_k(x) + g_k(y) - \sum_{i=1}^k x_i y_i\big)\ d\pi_n \\ &= \int \big(f_k(x) + g_k(y) - \sum_{i=1}^k x_i y_i\big)\ d\pi.\end{aligned}$$

Hence

$$\underline{\lim}_n \int \big(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\big)\ d\pi_n \geq \lim_k \int \big(f_k(x) + g_k(y) - \sum_{i=1}^k x_i y_i\big)\ d\pi,$$

where the limit in the right-hand side exists, because the sequence is monotone. Hence we get from (2)

$$0 \geq \lim_n \int (\varphi_n - f_n)\ d\mu + \lim_n \int \big(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\big) d\pi.$$

Taking into account that $\int g_n\ d\pi = \int g_n\ d\nu = \int \psi_n\ d\nu = \int \psi_n\ d\pi$, we obtain

$$\begin{aligned}0 &\geq \lim_n \int (\varphi_n - f_n)(x)\ d\mu + \lim_n \int \big(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\big) d\pi \\ &= \lim_n \Big(\int (\varphi_n(x) + \psi_n(y) - \sum_{i=1}^n x_i y_i)\ d\pi\Big) \geq 0.\end{aligned}$$

The proof is complete. □

Finally, we obtain a sufficient condition for the existence of an optimal mapping in the infinite-dimensional case.

Proof. (**Theorem 3.4)** Let us fix e_i and choose a sequence of numbers $t_n \to 0$. We get from Lemma 3.5 and Lemma 3.6 that there exist π-a.e. nonnegative functions $U_{t_n}(x) - t_n\langle y, e_i\rangle$ with $\int \bigl(U_{t_n}(x) - t_n\langle y, e_i\rangle\bigr)\, d\pi = o(t_n)$. Hence, $\lim_{t_n \to 0} \int \bigl(\frac{U_{t_n}(x)}{t_n} - \langle y, e_i\rangle\bigr)\, d\pi = 0$. Taking into account that $\frac{U_{t_n}(x)}{t_n} - \langle y, e_i\rangle \ge 0$ for π-almost all (x, y), we conclude that $\frac{U_{t_n}(x)}{t_n}$ converges μ-a.e. and in $L^1(\mu)$ to a function $u_i(x)$ satisfying $u_i(x) - \langle y, e_i\rangle \ge 0$, π-a.e. and $\int (u_i(x) - \langle y, e_i\rangle)\, d\pi = 0$. Clearly, $u(x) = \langle y, e_i\rangle$ for π-almost all (x, y). Repeating these arguments for every $i \in \mathbb{N}$, we get the claim. □

4. Application: quasi-product case

The main result of this section is a generalization of the optimal transport existence theorem for Gaussian measures. Recall that by results from [12], [15] that for the standard Gaussian measure $\gamma = \prod_{i=1}^{\infty} \gamma_i(dx_i)$, $\gamma_i \sim \mathcal{N}(0, 1)$ the existence of the optimal transportation mapping pushing forward $f \cdot \gamma$ into $g \cdot \gamma$ is established, for instance, under assumption $\int f \log f\, d\gamma < \infty, \int g \log g\, d\gamma < \infty$. We give in this section a generalization of this result for a wide class of quasi-product measures.

Let us consider two product reference measures

$$P = \prod_{i=1}^{\infty} p_i(x_i)\ dx_i, \quad Q = \prod_{i=1}^{\infty} q_i(x_i) dx_i$$

and fix the diagonal infinite transportation mapping

$$T(x) = (T_1(x_1), \cdots, T_n(x_n), \cdots)$$

where $T_i(x_i)$ pushes forward $p_i(x_i)dx_i$ onto $q_i(x_i)dx_i$. Clearly, T takes P onto Q. The inverse mapping $S = T^{-1}$ has the same diagonal structure:

$$S(x) = (S_1(x_1), \cdots, S_n(x_n), \cdots).$$

Theorem 4.1. *Let $\mu = f \cdot P$ and $\nu = g \cdot Q$ be probability measures satisfying the Assumption (A) of the previous section. Assume, in addition, that*

1) *there exists $K > 0$ such that every q_i is K-uniformly log-concave;*
2) *there exists $M > 0$ such that*

$$S_i'(x_i) \le M$$

for all i, x_i;

3) *Assume that either a) or b) holds for some constants $C > c > 0$*
   a) $g \log^2 g \in L^1(Q)$, $\frac{1}{f} \in L^1(P)$, $f \le C$,
   b) $f \log f \in L^1(P)$, $c \le g \le C$.

Then there exists a transportation mapping T pushing forward μ onto ν which is μ-a.e. limit of finite-dimensional optimal transportation mappings T_n.

Remark 4.2. It follows from Caffarelli's contraction theorem (see Section 2) that assumption 2) is satisfied if $(-\log p_i(x_i))'' \ge C_0$, $(-\log q_i(x_i))'' \le C_1$ for some $C_0, C_1 > 0$ and every i. Of course, there exist many other examples when this assumption is satisfied.

Proof. Consider the finite-dimensional projections $\mu_n = f_n \cdot P_n$, $\nu_n = g_n \cdot Q_n$, where $P_n = \prod_{i=1}^{n} p_i(x_i)\ dx_i$, $Q_n = \prod_{i=1}^{n} q_i(x_i)\ dx_i$. Here f_n and g_n are the conditional expectations of f, g with respect to P, Q and the σ-algebra $\mathcal{F}_n$, generated by the first n coordinates. Recall that $\nabla\varphi_n$ is the optimal transportation of μ_n to ν_n. Let

$$u_i(x_i),\ v_i(y_i) = u_i^*$$

be the one-dimensional convex potentials associated to the mappings T_i, S_i, respectively:

$$T_i = u_i',\ S_i = v_i'.$$

Note that $\tilde{T}_n = (T_1, \cdots, T_n)$ pushes forward P_n onto Q_n and $\nabla\varphi_n$ pushes forward $\frac{f_n}{g_n(\nabla\varphi_n)} \cdot P_n$ onto Q_n.

According to Proposition 2.2 one has the following estimate:

$$\frac{K}{2}\int |\tilde{T}_n - \nabla\varphi_n|^2 dP_n \le \int \log\Bigl(\frac{g_n(\nabla\varphi_n)}{f_n}\Bigr) dP_n. \tag{3}$$

To see that the right-hand side is finite, let us estimate

$$\begin{aligned}\int \log\Bigl(\frac{g_n(\nabla\varphi_n)}{f_n}\Bigr) dP_n &\le \int \log\frac{1}{f_n} dP_n + \frac{1}{2}\int \log^2 g_n(\nabla\varphi_n) f_n dP_n + \frac{1}{2}\int \frac{dP_n}{f_n} \\ &= \int \log\frac{1}{f_n} dP_n + \frac{1}{2}\int g_n \log^2 g_n dQ_n + \frac{1}{2}\int \frac{dP_n}{f_n}.\end{aligned}$$

Applying Assumption 3a of the Theorem and the Jensen inequality one can easily get that the right-hand side is uniformly bounded.

We complete the proof by applying Theorem 3.4 and Proposition 3.7. For application of Proposition 3.7 set

$$f_n = \sum_{i=1}^n u_i(x_i),\ g_n = \sum_{i=1}^n v_i(y_i).$$

We need to estimate $\sum_{i=1}^n \int (u_i(x_i) + v_i(y_i) - x_i y_i)\ d\pi_n$. Taking into account that π_n is supported on the graph of $\nabla\varphi_n$, and the relation $u_i(x_i) + v_i(T_i(x)) = x_i T_i(x)$ we obtain that the latter equals to

$$\begin{aligned}&\int \bigl(u_i(x_i) + v_i(\partial_{x_i}\varphi_n) - x_i \partial_{x_i}\varphi_n(x)\bigr)\ d\mu_n \\ &= \int \Bigl[v_i(\partial_{x_i}\varphi_n(x)) - v_i(T_i(x)) - x_i(\partial_{x_i}\varphi_n(x) - T_i(x))\Bigr]\ d\mu_n \\ &= \int \Bigl[v_i(\partial_{x_i}\varphi_n(x)) - v_i(T_i(x)) - v_i'(T_i(x))(\partial_{x_i}\varphi_n(x) - T_i(x))\Bigr]\ d\mu_n \\ &\le M \int (\partial_{x_i}\varphi_n(x) - T_i)^2\ d\mu_n.\end{aligned}$$

Here we use the uniform bound $v_i'' = S_i' \le M$. Finally, using the uniform bound $f \le C$ and the Jensen inequality we obtain that

$$\sum_{i=1}^n \int (u_i(x_i) + v_i(y_i) - x_i y_i)\ d\pi_n \le MC \int |\nabla\varphi_n - \tilde{T}_n|^2\ dP_n.$$

We have already shown that the right-hand side is bounded. The result now follows from Proposition 3.7.

The proof follows the same line under Assumption 3b, but we use another corollary of Proposition 2.2:

$$\frac{K}{2}\int |\tilde{T}_n - \nabla\varphi_n|^2 \frac{f_n}{g_n(\nabla\varphi_n)} dP_n \le \int \log\Bigl(\frac{f_n}{g_n(\nabla\varphi_n)}\Bigr)\frac{f_n}{g_n(\nabla\varphi_n)} dP_n.$$

The detailes are left to the reader. □

5. Symmetric transportation problem and ergodic decomposition of optimal transportation plans

5.1. **Symmetric transportation problem.** In this section we discuss the mass transportation of symmetric (mainly exchangeable) measures, where the word "symmetric" means "invariant under action of a group Γ".

Recall that a probability measure is exchangeable if it is invariant with respect to any permutation of finite number of coordinates. Before we consider $\mathbb{R}^\infty$, let us make some remarks on the finite-dimensional case.

Consider the group S_d of all permutations of $\{1, \cdots, d\}$ acting on $\mathbb{R}^d$ as follows:

$$L_\sigma(x) = (x_{\sigma(1)}, x_{\sigma(2)}, \cdots, x_{\sigma(d)}), \quad \sigma \in S_d.$$

Let $\Gamma \subset S_d$ be any subgroup with the property that for every couple i, j there exists $\sigma \in \Gamma$ such that $\sigma(i) = j$.

Assume that the source and target measures are both invariant with respect to Γ. Under additional assumption that the cost function c is Γ-invariant (for instance, $c = |x - y|^2$) one can easily check that the Kantorivich potential φ is Γ-invariant as well: $\varphi = \varphi \circ L_\sigma$ for any $\sigma \in \Gamma$ see [21], [25]. Consequently, the optimal transportation $T = \nabla \varphi$ has the following commutation property:

$$T = L^*_\sigma (T \circ L_\sigma) = L^{-1}_\sigma \circ T \circ L_\sigma.$$

Equivalently,

$$L_\sigma \circ T = T \circ L_\sigma.$$

The optimal transportation plan $\pi(dx, dy)$ is also Γ-invariant under the following extension of the action of Γ to $\mathbb{R}^d \times \mathbb{R}^d$:

$$L_\sigma(x, y) = (L_\sigma x, L_\sigma y).$$

Now let $\sigma(i) = j$. One has

$$\begin{aligned} \int x_i y_i \ d\pi &= \int \langle e_i, x\rangle \langle e_i, y\rangle \ d\pi = \int \langle L_\sigma e_i, L_\sigma x\rangle \langle L_\sigma e_i, L_\sigma y\rangle \ d\pi \\ &= \int \langle e_j, L_\sigma x\rangle \langle e_j, L_\sigma y\rangle \ d\pi = \int x_j y_j \ d\pi. \end{aligned}$$

Consequently,

$$W_2^2(\mu, \nu) = \int \|x - y\|^2 \ d\pi = \sum_{i=1}^{d} \int (x_i - y_i)^2 \ d\pi = d \int (x_i - y_i)^2 \ d\pi, \quad \forall i. \tag{4}$$

Lemma 5.1. *The standard quadratic Kantorovich problem on $\mathbb{R}^d$ with Γ-invariant marginals is equivalent to the transportation problem for the cost $|x_1 - y_1|^2$ with additional constraint that the solution is a Γ-invariant probability measure*

Proof. Let π be the solution to the quadratic Kantorovich problem for the marginals μ, ν and $\tilde{\pi}$ be a measure giving the minimum to the functional $m \mapsto \int |x_1 - y_1|^2 dm$ among of the Γ-invariant measures with the same marginals. By optimality of π

$$\int \|x - y\|^2 d\pi \le \int \|x - y\|^2 d\tilde{\pi}.$$

Since π and $\tilde{\pi}$ are both Γ-invariant, (4) implies that $\int |x_1 - y_1|^2 d\pi \le \int |x_1 - y_1|^2 d\tilde{\pi}$. By optimality of $\tilde{\pi}$ one gets $\int |x_1 - y_1|^2 d\pi = \int |x_1 - y_1|^2 d\tilde{\pi}$, and, finally

$\int \|x-y\|^2 d\pi = \int \|x-y\|^2 d\tilde{\pi}$. This means that $\tilde{\pi}$ solves the quadratic Kantorovich problem as well and, vice versa, π solves the Kantorovich problem with symmetric constraints. □

The conclusion made above helps us to give a variational meaning to the transportation problem in the infinite-dimensional case.

Definition 5.2. Symmetric Kantorovich problem. Let Γ be a group of linear operators acting on $\mathbb{R}^\infty$ and μ, ν be Γ-invariant probability measures. Assume in addition that

- For every $i, j \in \mathbb{N}$ there exists $g \in \Gamma$ such that
$$g(e_i) = e_j.$$
- The space of probability measures $\Pi^\Gamma(\mu,\nu)$ on $\mathbb{R}^\infty \times \mathbb{R}^\infty$ which are invariant with respect to the action $(x, y) \mapsto (g(x), g(y))$, $g \in \Gamma$ of Γ and have marginals μ, ν, is non-empty and closed in the weak topology.

We say that a measure $\pi \in \Pi^\Gamma(\mu,\nu)$ is a solution to the Γ-symmetric (quadratic) Kantorovich problem if it gives the minimum to the functional

$$\Pi^\Gamma(\mu,\nu) \ni m \mapsto \int (x_1 - y_1)^2 \, dm. \tag{5}$$

Definition 5.3. Symmetric optimal transportation. Let m be a solution to the symmetric Kantorovich problem. A measurable mapping $T \colon \mathbb{R}^\infty \mapsto \mathbb{R}^\infty$ is called optimal transportation mapping of μ onto ν if

$$m(\{(x, T(x))\}) = 1.$$

The standard compactness arguments imply that a solution to the Kantorovich problem (5) exists provided $\int x_1^2 \, d\mu < \infty, \int y_1^2 \, d\nu < \infty$. If, in addition, there exists an optimal transportation mapping T, it commutes with any $g \in \Gamma$. This means that for μ-almost all x and every $g \in \Gamma$

$$(T \circ g)(x) = (g \circ T)(x). \tag{6}$$

Example 5.4. **Exchangeable measures.** We denote by S_∞ the group of permutation of $\mathbb{N}$ which change only a finite number of coordinates. We consider its natural action on $\mathbb{R}^\infty$ defined by

$$\sigma(x) = (x_{\sigma(i)}), \quad x = (x_i) \in \mathbb{R}^\infty, \quad \sigma \in S_\infty.$$

Consider measures μ and ν which are invariant with respect to any $\sigma \in S_\infty$:

$$\mu = \mu \circ \sigma^{-1}, \quad \nu = \nu \circ \sigma^{-1}.$$

The measures of this type are called exchangeable. The basic example is given by the countable power m^∞ of some Borel measure m on $\mathbb{R}$. The structure of mappings satisfying (6) in the case $\mu = m^\infty$ is very easy to describe. Consider the function $T_1(x) = \langle T(x), e_1 \rangle$ and fix the first coordinate x_1. Then the function $F \colon (x_2, x_3, \cdots) \to T_1(x)$ is invariant with respect to S_∞ (acting on $(x_2, x_3, \cdots)$). Hence F is constant according by the Hewitt–Sawage $0-1$ law applied to the measure μ. Thus $T_1(x) = T_1(x_1)$ depends on x_1 only (up to a set of measure zero). The same arguments applied to other coordinates imply that T is diagonal: $(T_1(x_1), T_2(x_2), \cdots)$. Moreover, $T_i(x) = T_1(x)$ because T commutes with every permutation of coordinates.

Example 5.5. **Optimal transportation not always exists.** Let μ_1, μ_2 be countable powers of two different one-dimensional measures. By the Kakutani dichotomy theorem they are mutually singular. There is no any mass transportation T of $\mu = \mu_1$ onto $\nu = \frac{1}{2}(\mu_1 + \mu_2)$ satisfying (6). Indeed, according to Example 5.4 any T satisfying (6) must be diagonal, hence the measure $\mu \circ T^{-1}$ must be a product measure.

Thus, we see that the optimal transportation does not always exist. This example can be easily generalized to many other linear groups Γ and Γ-invariant measures. It can be easily understood that T does not exist provided the source measure is ergodic, but the target measure is not.

5.2. **Ergodic decomposition of optimal transportation plans.** The connection between Kantorovich problem and ergodic decomposition has been established under fairy general assumptions by the second-named author in [26]. A particular case of this result is given in the following theorem.

Let Γ be an amenable group acting by continuous one-to-one mappings on a Polish space X. Let Π^{Γ} be the set of all Borel probability Γ-invariant measures and $\mu, \nu \in \Pi^{\Gamma}$. The set of Γ-invariant transportation plans with marginals μ, ν will be denoted by $\Pi^{\Gamma}(\mu, \nu)$. Assume that the cost function c is lower semicontinuous and $\Pi^{\Gamma}(\mu, \nu)$ is non-empty and closed in the weak topology.

Let us fix a solution π to the Γ-invariant Kantorovich problem with marginals μ, ν. Denote by $\Delta(X)$ the set all Γ-invariant ergodic measures on X. Assume we are given ergodic decompositions

$$\mu = \int_{\Delta(X)} \mu^x \ d\sigma_\mu, \ \ \nu = \int_{\Delta(Y)} \nu^y \ d\sigma_\nu \tag{7}$$

of μ, ν, where $X = Y$, σ_μ, σ_ν are probability measures on $\Delta(X), \Delta(Y)$ and, similarly, the ergodic decomposition of π:

$$\pi = \int_{\Delta(X \times Y)} \pi^{x,y} d\delta \tag{8}$$

(recall that the Γ-invariance for π means the invariance with respect to the action $(x, y) \mapsto (g(x), g(y))$). We stress that in (7) the integrals are taken not with respect to variables x, y, but with respect to variables μ^x, ν^y (x, y indicate the spaces where the measures are defined), the same holds for (8). It is straightforward that δ-almost all $\pi^{x,y}$ have ergodic marginals and taking the projections of the both sides of (8) we obtain decompositions (7). Moreover, the following statement holds:

Theorem 5.6. *Under δ almost every measure $\pi^{x,y}$ solves the Γ-symmetric Kantorovich problem with marginals μ^x, ν^y:*

$$K_c^{\Gamma}(\mu^x, \nu^y) = \inf_{m \in \Pi^{\Gamma}(\mu^x, \nu^y)} \int c dm = \int c d\pi^{x,y}$$

and the following representation formula holds:

$$\inf_{\pi \in \Pi^{\Gamma}(\mu, \nu)} \int c d\pi = \inf_{\delta \in \Pi(\sigma_\mu, \sigma_\nu)} \int K_c^{\Gamma}(\mu^x, \nu^y) \ d\delta.$$

Remark 5.7. In the situation of Theorem 5.6 one can decompose the optimal transportation plan for ergodic marginals μ, ν: $\pi = \int_{\Delta(X \times Y)} \pi^{x,y} d\delta$. Ergodicity of the marginals implies immediately that δ-almost all $\pi^{x,y}$ have the same marginals μ

and ν. The optimality of $\pi^{x,y}$ for the cost c follows from Theorem 5.6. Thus we get that any solvable symmetric Kantorovich problem with ergodic marginals admits, in particular, an ergodic solution.

Thus the symmetric transportation problem can be reduced to the following steps:

Q1) Construct a solution to the symmetric Kantorovich problem for ergodic measures.

Q2) Given two non-ergodic measures μ, ν and the corresponding ergodic decompositions (7) construct a solution to the Kantorovich problem to measures σ_μ, σ_ν on $\Delta(X)$ with the cost function K_c^Γ.

Consider application of Theorem 5.6 to several classical groups.

Example 5.8. **Exchangeable measures revisited.** Consider invariant transportation problem for exchangeable measures and $c = (x_1 - y_1)^2$. The answer to Q1) is trivial, because ergodic measures are countable powers and the structure of the corresponding solution is trivial. As for Q2), by the de Finetti theorem the space of ergodic measures is isomorphic to the space $\mathcal{P}(\mathbb{R})$ of probability measures on $\mathbb{R}$. Thus to resolve an optimal transportation problem for exchangeable measures, we need to study the optimal transportation problem for a couple of measures μ_0, ν_0 on $\mathcal{P}(\mathbb{R})$ arising from the de Finetti decomposition. It is clear that the cost function c on $\mathcal{P}(\mathbb{R})$ satisfies

$$c(p_1, p_2) = W_2^2(p_1, p_2),$$

where W_2 is the standard Kantorovich distance on $\mathbb{R}$.

Example 5.9. **Rotationally invariant measures.** Consider invariant transportation problem for measures invariant with respect to operators of the type $U \times Id$, where U is a rotation of $\mathbb{R}^n = Pr_n(\mathbb{R}^\infty)$ and Id is the identical operator on the orthogonal complement to $\mathbb{R}^n$ As usual $c = (x_1 - y_1)^2$. This is an example where the optimal transportation problem admits a precise solution. By a well known result (see [14]) every rotationally invariant measure μ on $\mathbb{R}^\infty$ admits a representation

$$\mu = \int \gamma_t dp_\mu(t),$$

where γ_t is the distribution of the Gaussian i.i.d. with zero mean and variance t and p_μ is a measure on $\mathbb{R}_+$. The optimal transportation problem is reduced obviously to the one-dimensional optimal transportation between p_μ and p_ν.

Example 5.10. **Stationary measures.** These are the measures which are invariant with respect to the shift:

$$T\colon x = (x_1, x_2, \cdots) \mapsto (x_2, x_3, \cdots).$$

Note that the powers of T generates the semigroup $\{0\} \cup \mathbb{N}$, but not the group. However, it makes no difference for our analysis, we are still able to consider the corresponding ergodic decompositions. In this case the description of ergodic measures is nontrivial and we do not know any general sufficient conditions for existence even in the case when both measures are ergodic. Some sufficient conditions are given in Section 7.

We conclude the section with the remark that existence of a transportation mapping for (not necessary optimal) symmetric plan π with ergodic X-marginal implies ergodicity of π.

Proposition 5.11. *Let $X = Y$ be Polish space and Γ be a group of Borel one-to-one transformations acting on X. Assume that π and μ are Γ-invariant Borel probability measures on $X \times Y$ and X respectively. Assume, in addition, that $Pr_X \pi = \mu$, μ is ergodic, and $\pi(\{x, T(x)\}) = 1$ for some Borel mapping T. Then π is ergodic.*

Proof. Assuming the contrary we represent π as a convex combinations of two Γ-invariant measures

$$\pi = \lambda \pi_1 + (1 - \lambda)\pi_2,$$

$\pi_1 \neq \pi_2$, $0 < \lambda < 1$. Clearly, this implies a similar decomposition for the projections $\mu = \lambda Pr_X \pi_1 + (1-\lambda) Pr_X \pi_2$. If we show that μ_1, μ_2 are Γ-invariant and distinct, we will get a contradiction. The Γ-invariance of both measures follows immediately from the Γ-invariance of π_i. Let us show that $\mu_1 \neq \mu_2$. Assume the contrary and take a Borel set $B \subset X \times Y$. We get that $\pi_i(B)$ equals to $\mu_i(A)$, where $A = \mathrm{Pr}_X(B \cap \mathrm{Graph}(T)))$ (note that A is universally measurable as a projection of a Borel set). Then it follows that π_i coincide because μ_i do coincide. ☐

6. Kantorovich duality

In this section we start to study measures which are invariant under actions of some group. The results of this section will not be used in this paper, but they are of independent interest.

Let X, Y be Polish spaces, Γ be a locally-compact amenable group with *continuous* actions L^X_Γ, L^Y_Γ on X, Y respectively. The action L_Γ on the product space $X \times Y$ is defined as follows:

$$L_g(x, y) = (L_g(x), L_g(y)).$$

where L_g is an element of L_Γ corresponding to $g \in \Gamma$.

Let us define the space $W_\Gamma \subset C_b(X \times Y)$ as the closure of linear span of the following set:

$$\{f - f \circ L_g : \ f \in C_b(X \times Y), \ g \in \Gamma\}.$$

It can be checked that the property

$$\int \omega d\pi = 0, \ \forall \omega \in W_\Gamma \tag{9}$$

of a probability measure $\pi \in \mathcal{P}(X \times Y)$ is equivalent to its invariance w.r.t. L_Γ.

Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ be invariant under the actions L^X_Γ, L^Y_Γ respectively. Then a transport plan $\pi \in \Pi(\mu, \nu)$ is invariant iff the property (9) is satisfied. We denote the set of all invariant transport plans by $\Pi^\Gamma(\mu, \nu)$.

The following Theorem is a refinement of the duality result, which was proved in [25] (Theorem 2.5). In there we considered only $C_b(X \times Y)$ cost functions (we warn the reader that the classical duality statement from Section 2 is formulated in a slightly different but equivalent way: in notations of this section $\Phi = \frac{x^2}{2} - \varphi, \Psi = \frac{y^2}{2} - \psi$).

Theorem 6.1. *Let $c \in C(X \times Y)$ be a nonnegative function such that there exist $f \in L^1(X, \mu)$, $g \in L^1(Y, \nu)$, and*

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

Then, in the setting described above,

$$\inf_{\pi\in\Pi^\Gamma}\int c d\pi = \sup_{\Phi+\Psi+\omega\le c}\int_X \Phi(x)d\mu + \int_Y \Psi(y)d\nu,$$

where $\Phi \in L^1(X)$, $\Psi \in L^1(Y)$, $\omega \in W_\Gamma$.

Proof. The inequality

$$\inf_{\pi\in\Pi^\Gamma}\int c d\pi \ge \sup_{\Phi+\Psi+\omega\le c}\int \Phi d\mu + \int \Psi d\nu$$

can be easily obtained:

$$\inf_{\pi\in\Pi^\Gamma}\int c d\pi \ge \inf_{\pi\in\Pi^\Gamma}\left(\sup_{\Phi+\Psi+\omega\le c}\int(\Phi+\Psi+\omega)d\pi\right) =$$
$$= \inf_{\pi\in\Pi^\Gamma}\left(\sup_{\Phi+\Psi+\omega\le c}\int \Phi d\mu + \int \Psi d\nu\right) = \sup_{\Phi+\Psi+\omega\le c}\int \Phi d\mu + \int \Psi d\nu.$$

To obtain the opposite inequality we use the following statement from Theorem 2.5 of [25].

$$\inf_{\pi\in\Pi^\Gamma}\int c_b d\pi = \sup_{\Phi+\Psi+\omega\le c_b}\int_X \Phi(x)d\mu + \int_Y \Psi(y)d\nu$$

for $c_b \in C_b(X\times Y)$, $\Phi \in C_b(X)$, $\Psi \in C_b(Y)$, $\omega \in W_\Gamma$. Let $c_n(x,y) := \min\{c(x,y), n\}$ for each $n \in N$. The inequality

$$\sup_{\Phi+\Psi+\omega\le c_n}\int_X \Phi(x)d\mu + \int_Y \Psi(y)d\nu \le \sup_{\Phi+\Psi+\omega\le c}\int_X \Phi(x)d\mu + \int_Y \Psi(y)d\nu$$

is obvious for any natural n. Thus it remains to prove that

$$\lim_{n\to\infty}\inf_{\pi\in\Pi^\Gamma}\int c_n d\pi = \inf_{\pi\in\Pi^\Gamma}\int c d\pi.$$

Recall that the functional $\pi \to \int c_b d\pi$ is weakly continuous for every $c_b \in C_b(X\times Y)$. It follows from the characterization (9) of invariant measures, that $\Pi^\Gamma(\mu,\nu)$ is a closed subset of $\Pi(\mu,\nu)$, which is known to be compact. Thus $\Pi^\Gamma(\mu,\nu)$ is compact in the topology of weak convergence. If π_n is the solution for

$$\inf_{\pi\in\Pi^\Gamma}\int c_n d\pi,$$

the sequence (π_n) has to have a subsequence converging to some element $\pi^* \in \Pi^\Gamma$. Since for any fixed $m \in \mathbb{N}$ the inequality: $\lim_{n\to\infty}\int c_n d\pi^* \ge \int c_m d\pi^*$ is satisfied, and, by monotone convergence theorem, $\lim_{m\to\infty}\int c_m d\pi^* = \int c d\pi^* \le \int (f(x)+g(y))d\pi^* < \infty$, we obtain

$$\lim_{n\to\infty}\int c_n d\pi_n \ge \lim_{m\to\infty}\int c_m d\pi^* = \int c d\pi^* \ge \inf_{\pi\in\Pi^\Gamma}\int c d\pi.$$

This fact concludes the proof of the theorem. □

As one can see, the form of the duality theorem is similar to the well-known classic result, but the difference is substantial: dual functionals are related to each other in a more complicated way. Moreover, there is no existence result for the dual problem without any additional assumptions.

It was shown in [25] (Theorem 5.7) that in case of compact group Γ and under the assumptions of Theorem 6.1,

$$\inf_{\pi\in\Pi^\Gamma}\int c d\pi = \sup_{\Phi+\Psi\le\bar{c}}\int_X \Phi(x)d\mu + \int_Y \Psi(y)d\nu.$$

where $\bar{c} := \int_\Gamma (c\circ g)d\chi(g)$ and $\chi(g)$ is the probability Haar measure. It is clear that if cost function is Γ-invariant, the invariant dual problem coincides with the usual one.

Moameni ([21]) proved that for $\Gamma = \mathbb{Z}$ and an invariant cost function c, the corresponding invariant dual problem coincides with the usual one, and, moreover, both prime and dual Kantorovich problems have an invariant solution.

7. Existence of invariant optimal mapping for stationary measures

Recall that the measures on $\mathbb{R}^\infty$ which are invariant with respect to the shift

$$\sigma(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$$

are called stationary measures. Unlike exchangeable measures, the projections of stationary measures are in general not invariant with respect to some reasonable family of linear transformation.

As usual we assume that $\mathbb{R}^\infty$ is approximated by the sequence of finite-dimensional spaces $\mathbb{R}^n$ in the following sense: we identify $\mathbb{R}^n$ with the subset

$$P_n(\mathbb{R}^\infty) = \{x = (x_1, x_2, \cdots, x_n, 0, 0, \cdots)\} \subset \mathbb{R}^\infty.$$

On every finite-dimensional space $\mathbb{R}^n$ we will apply the following operator of cyclical shift:

$$\sigma_n(x_1, x_2, \cdots, x_n) = (x_2, x_3, \cdots, x_n, x_1).$$

Let us associate with every stationary measure μ the cyclical average of its projections:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} (\mu\circ P_n^{-1})\circ\sigma_n^{-(i-1)}.$$

In addition, let us denote by $\mathbb{R}_{m,n}$ the orthogonal complement of $\mathbb{R}^m \subset \mathbb{R}^n$:

$$\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}_{m,n},\ m < n.$$

The marginal measures are always assumed to satify the following property:

Assumption B. The measures μ, ν are stationary Borel probability measures such that their projections on every $\mathbb{R}^n$

$$\mu\circ Pr_n^{-1},\ \nu\circ Pr_n^{-1}$$

have Lebesgue densities and bounded second moments.

We consider symmetric Monge-Kantorovich problem

$$\int (x_1 - y_1)^2\ d\pi \to \min \tag{10}$$

where the infimum is taken among of all stationary measures $\Pi^\Gamma(\mu,\nu)$ with marginals μ, ν.

Remark 7.1. Minimizing $\int (x_1-y_1)^2\ d\pi$ is equivalent to maximizing of $\int x_1 y_1\ d\pi$, because $\int x_1^2\ d\pi = \int x_1^2\ d\mu,\ \int y_1^2\ d\pi = \int y_1^2 d\nu$ are fixed.

Theorem 7.2. *Let μ be a stationary measure which satisfies the following assumptions:*

1) *μ is a weak limit of a sequence of σ_n-invariant measures μ_n on $\mathbb{R}^n$.*
2) *For every $m < n$ there exists a probability measure $\mu_{m,n}$ on $\mathbb{R}_{m,n}$ such that the relative entropy (the Kullback-Leibler distance) between $\mu_m \times \mu_{m,n}$ and μ_n is uniformly bounded in n:*
$$\int \log\Bigl(\frac{d\mu_n}{d(\mu_m \times \mu_{m,n})}\Bigr) d\mu_n < C_m$$
with C_m satisfying
$$\lim_m \frac{C_m}{m} = 0;$$
3) *The cyclical average $\hat{\mu}_n$ of the n-dimensional projection $\mu \circ P_n^{-1}$ has finite second moments and admits a density ρ_n with respect to μ satisfying*
$$\sup_n \int \rho_n^{-\varepsilon} d\mu < \infty$$
for some $\varepsilon > 0$.

Then there exists a mapping T with the properties

- *T pushes forward μ onto the standard Gaussian measure on $\mathbb{R}^\infty$:*
$$\nu = \gamma.$$
- *T a μ-a.e. limit of finite dimensional mappings $T_n : \mathbb{R}^n \mapsto \mathbb{R}^n$ such that every T_n is a solution to an optimal transportation problem on $\mathbb{R}^n$.*

Proof. We consider the sequence of n-dimensional optimal transportation mappings T_n with cost function $\sum_{i=1}^n (x_i - y_i)^2$ pushing forward μ_n onto γ_n. It follows from the σ_n-invariance of μ_n and γ_n that the mapping T_n is cyclically invariant:
$$\langle T_n \circ \sigma_n, e_i \rangle = \langle T_n, e_{i-1} \rangle, \quad \mu_n - \text{a.e.}$$

Fix a couple of numbers m, n with $n > m$. Let $T_{m,n}$ be the optimal transportation mapping for the cost function $\sum_{i=n+1}^m (x_i - y_i)^2$ pushing foward $\mu_{m,n}$ onto the standard Gaussian measure on $\mathbb{R}_{m,n}$. We stress that T_m and $T_{m,n}$ depend on different collections of coordinates.

We extend T_m onto $\mathbb{R}^n$ in the following way:
$$T_m(x) = T_m(P_m x) + T_{m,n}(P_{m,n} x).$$
Clearly, T_m pushes forward $\mu_m \times \mu_{m,n}$ onto the standard Gaussian measure on $\mathbb{R}_n$. Applying Proposition 2.2 to the couple of mappings T_m, T_n, we get
$$\frac{1}{2} \int \|T_n - T_m\|^2 d\mu_n \le \int \log\Bigl(\frac{d\mu_n}{d(\mu_m \times \mu_{m,n})}\Bigr) d\mu_n. \tag{11}$$
This implies
$$\sum_{i=1}^m \int \langle T_n - T_m, e_i \rangle^2 \, d\mu_n \le \int \|T_n - T_m\|^2 d\mu_n \le 2C_m \tag{12}$$
for every m, n, $m < n$.

Let us note that for every i one can extract a weakly convergent subsequence from a sequence of (signed) measures $\{\langle T_n, e_i \rangle \cdot \mu_n\}$. Indeed, for any compact set K
$$\Bigl(\int_{K^c} |\langle T_n, e_i \rangle| d\mu_n\Bigr)^2 \le \int |\langle T_n, e_i \rangle|^2 d\mu_n \cdot \mu_n(K^c) = \int x_i^2 \, d\gamma \cdot \mu_n(K^c).$$

Using the tightness of $\{\mu_n\}$ we get that $\{|\langle T_n, e_i\rangle| \cdot \mu_n\}$ is a tight sequence. In addition, note that for every continuous f

$$\lim_n \Bigl(\int f |\langle T_n, e_i\rangle| d\mu_n\Bigr)^2 \le \int x_i^2 \ d\gamma \cdot \int f^2 d\mu.$$

This implies that any limiting point of $\{\langle T_n, e_i\rangle \cdot \mu_n\}$ is absolutely continuous with respect to μ. Applying the diagonal method and passing to a subsequence one can assume that convergence takes place for all i simultaneously. Consequently, there exists a subsequence $\{n_k\}$ and a measurable mapping T with values in $\mathbb{R}^\infty$ such that

$$\langle T_{n_k}, e_i\rangle \cdot \mu_{n_k} \to \langle T, e_i\rangle \cdot \mu$$

weakly in the sense of measures for every i. It is easy to check that the standard property of L^2-weak convergence holds also in this case:

$$\int \langle T, e_i\rangle^2 d\mu \le \underline{\lim}_k \int \langle T_{n_k}, e_i\rangle^2 d\mu_n = \int x_i^2 \ d\gamma = 1. \tag{13}$$

Finally, we pass to the limit in (12) and get

$$\sum_{i=1}^m \int \langle T - T_m, e_i\rangle^2 \ d\mu \le 2C_m. \tag{14}$$

The claim follows from (13) and the fact that $\lim_n \int \varphi \ d\mu_n = \int \varphi \ d\mu$ for every $\varphi \in L^2(\mu)$. Indeed, if φ is bounded and continuous, this follows from the weak convergence $\langle T_n, e_i\rangle \cdot \mu_n \to \langle T, e_i\rangle \cdot \mu$. For arbitrary $\varphi \in L^2(\mu)$ we find continuous bounded cylindrical function $\tilde{\varphi}$ such that $\|\varphi - \tilde{\varphi}\|_{L^2(\mu)} < \varepsilon$. One has $\lim_n \int \varphi \ d\mu_n = \lim_n \int (\varphi - \tilde{\varphi}) \ d\mu_n + \int \tilde{\varphi} \ d\mu$. The claim follows from the estimate

$$\Bigl(\int |\varphi - \tilde{\varphi}| \ d\mu_n\Bigr)^2 \le \int (\varphi - \tilde{\varphi})^2 \ d\mu \cdot \int \rho_n^2 \ d\mu \le (\sup_n \int \rho_n^2 \ d\mu)\varepsilon^2.$$

Note that T commutes with the shift σ: $\langle T \circ \sigma, e_i\rangle = \langle T, e_{i-1}\rangle$. Indeed, for every bounded cylindrical φ one has

$$\int \varphi \langle T_n, e_{i-1}\rangle d\mu_n = \int \varphi \langle T_n(\sigma_n), e_i\rangle d\mu_n = \int \varphi(\sigma_n^{-1}) \langle T_n, e_i\rangle d\mu_n = \int \varphi(\sigma^{-1}) \langle T_n, e_i\rangle d\mu_n.$$

Here we use that $\varphi(\sigma_n^{-1}) = \varphi(\sigma^{-1})$ for sufficiently large values of n and the cyclical invariance of T_n. Passing to the limit in the n_k-subsequence one gets

$$\int \varphi \langle T, e_{i-1}\rangle d\mu = \int \varphi(\sigma^{-1}) \langle T, e_i\rangle d\mu = \int \varphi \langle T \circ \sigma, e_i\rangle d\mu.$$

Hence $T \circ \sigma = \sigma \circ T$.

Hence by assumptions of the theorem and (14) we get

$$\limsup_m \frac{1}{m} \sum_{i=1}^m \int \langle T - T_m, e_i\rangle^2 \ d\mu = 0.$$

To prove that T pushes forward μ into γ it is sufficient to show that that $\langle T_m, e_i\rangle \to \langle T, e_i\rangle$ in measure (see Lemma 3.2). To this end let us approximate T_1 by a bounded function $\xi_1(x_1, \ldots, x_k)$ depending on finite number of coordinates in $L^2(\mu)$: $\int \|T_1 - \xi_1\|^2 d\mu < \varepsilon$, where ε is chosen sufficiently small. Set: $\xi_i = \xi \circ \sigma^{i-1}$. Clearly, we get by the shift invariance

$$\frac{1}{m} \int \sum_{i=1}^{m} (T_i - \xi_i)^2 d\mu = \int (T_1 - \xi_1)^2 d\mu < \varepsilon.$$

Hence

$$\limsup_m \frac{1}{m} \int \|T_m - \xi\|^2 d\mu \le \varepsilon, \ \xi = (\xi_1, \xi_2, \ldots).$$

Let make the change of variables under the cyclical shift σ_n. One has

$$\langle T_m, e_i \rangle \circ \sigma_m^{-(i-1)} = T_1$$

for all $1 \le i \le m$ and

$$\xi_i \circ \sigma_m^{-(i-1)} = \xi_1$$

as soon as $i - 1 + k \le m$. Hence for the latter values of i one has

$$\int \langle \xi - T, e_i \rangle^2 d\mu = \int \langle \xi - T, e_1 \rangle^2 \ d\mu \circ \sigma_n^i.$$

The number of indices which do not satify this property is limited by k. Clearly, it doses not affect the limit of averages. Finally we obtain

$$\varepsilon \ge \limsup_m \frac{1}{m} \int \sum_{i=1}^{n} \langle \xi - T_m, e_i \rangle^2 d\mu = \limsup_m \int \langle \xi - T_m, e_1 \rangle^2 d\hat{\mu}_m.$$

Recall that $\int (T_1 - \xi_1)^2 d\mu \le \varepsilon$. Finally

$$\begin{aligned} \limsup_m \int \langle T - T_m, e_1 \rangle^2 d\hat{\mu}_m &\le 2 \limsup_m \int \langle \xi - T_m, e_1 \rangle^2 d\hat{\mu}_m \\ &+ 2 \limsup_m \int (T_1 - \xi_1)^2 d\hat{\mu}_m \le 4\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, one gets $\int \langle T - T_m, e_1 \rangle^2 d\hat{\mu}_m \to 0$. By the Hölder inequality

$$\int \langle T - T_m, e_1 \rangle^{\frac{2}{p}} d\mu \le \Bigl(\int \langle T - T_m, e_1 \rangle^2 d\hat{\mu}_m \Bigr)^{\frac{1}{p}} \Bigl(\int \rho_m^{-\frac{1}{p-1}} d\mu \Bigr)^{\frac{1}{q}}.$$

Take $p = 1 + \frac{1}{\varepsilon}$ we get by the assumption of the theorem that the latter tends to zero. The proof is complete. □

Remark 7.3. In Theorem 7.2 the Gaussian measure γ can be replaced by any countable power of an uniformly log-concave one-dimensional measure.

In the following proposition we prove that the transportation mapping T is indeed optimal under additional assumptions.

Proposition 7.4. *Let the assumptions of Theorem 7.2 hold. Assume in addition that*

$$\lim_{n \to \infty} \frac{1}{n} W_2^2(\hat{\mu}_n, \mu_n) = 0.$$

Then there exists a solution π of problem (10) in the class of stationary measures such that $\pi\{(x, T(x)), x \in \mathbb{R}^\infty\} = 1$.

Proof. We show that the measure $\pi = \mu \circ (x, T(x))^{-1}$, which is the weak limit of measures π_n is optimal. Recall that π_n gives minimum to $m \to \int \sum_{i=1}^{n} (x_i - y_i)^2 dm$ and has marginals μ_n, γ_n, hence measure π has marginals μ, γ. Indeed,

$$\int (x_1 - y_1)^2 d\pi = \lim_n \int (x_1 - y_1)^2 d\pi_n = \lim_n \frac{1}{n} \int \sum_{i=1}^{n} (x_i - y_i)^2 d\pi_n.$$

If π is not optimal, when there exists a stationary measure π_0 with projections μ, ν such that

$$\int (x_1 - y_1)^2 d\pi_0 + \varepsilon < \frac{1}{n} \int \sum_{i=1}^{N} (x_i - y_i)^2 d\pi_n$$

for some $\varepsilon > 0$ and all sufficiently big values of n. Taking into account stationarity of π_0 we get $\int x_i y_i d\pi_0 = \int x_j y_j \pi_0$ for every i, j, thus

$$\int \sum_{i=1}^{n} (x_i - y_i)^2 d\hat{\pi}_0 + n\varepsilon = \int \sum_{i=1}^{n} (x_i - y_i)^2 d\pi_0 + n\varepsilon < \int \sum_{i=1}^{n} (x_i - y_i)^2 d\pi_n,$$

where $\hat{\pi}_0 = \frac{1}{n} \sum_{i=1}^{n} (\pi_0 \circ Pr_n^{-1}) \circ \sigma_n^{-(i-1)}$. The latter inequality implies

$$W_2^2(\hat{\mu}_n, \gamma_n) + n\varepsilon \le W_2^2(\mu_n, \gamma_n).$$

By the triangle inequality

$$\begin{aligned} W_2^2(\hat{\mu}_n, \gamma_n) + n\varepsilon &\le (W_2(\mu_n, \tilde{\mu}_n) + W_2(\hat{\mu}_n, \gamma_n))^2 \\ &\le W_2^2(\mu_n, \hat{\mu}_n) + 2W_2(\hat{\mu}_n, \gamma_n) W_2(\mu_n, \hat{\mu}_n) + W_2^2(\hat{\mu}_n, \gamma_n). \end{aligned}$$

Hence

$$\varepsilon \le \frac{1}{n} (2W_2(\hat{\mu}_n, \gamma_n) W_2(\mu_n, \hat{\mu}_n) + W_2^2(\hat{\mu}_n, \mu_n)). \tag{15}$$

The quantity $W_2^2(\hat{\mu}_n, \gamma_n)$ can be trivially estimated by $2 \sum_{i=1}^{n} (\int x_i^2 d\hat{\mu}_n + \int y_i^2 d\gamma_n) \le Cn$. Then the using the assumption of the theorem we get that the right-hand side of (15) tends to zero, which contradicts to positivity of ε. $\square$

We finish this section with a concrete application of Theorem 7.2. We study a transportation of a Gibbs measure μ which can be formally written in the form

$$\mu = e^{-H(x)} dx,$$

where the potential H admits the following heuristic representation:

$$H(x) = \sum_{i=1}^{\infty} V(x_i) + \sum_{i=1}^{\infty} W(x_i, x_{i+1}).$$

Here V and W are smooth functions and $W(x, y)$ is symmetric: $W(x, y) = W(y, x)$. The existence of such measures was proved in [2].

Let us specify the assumptions about V and W. These are a particular case of assumptions A1-A3 from [2].

1)
$$W(x, y) = W(y, x);$$
2) There exist numbers $J > 0$, $L \ge 1$, $N \ge 2$, $\sigma > 0$, and $A, B, C > 0$ such that
$$|W(x, y)| \le J(1 + |x| + |y|)^{N-1}, \ |\partial_x W(x, y)| \le J(1 + |x| + |y|)^{N-1}$$

3)

$$|V(x)| \leq C(1+|x|)^L, \quad |V'(x)| \leq C(1+|x|)^{L-1};$$

4) (coercivity assumption)

$$V'(x)\cdot x \geq A|x|^{N+\sigma} - B.$$

Let us define the following probability measure on E_n:

$$\mu_n = \frac{1}{Z_n}\exp\Bigl(-\sum_{i=1}^{n}\bigl(V(x_i)+W(x_i,x_{i+1})\bigr)\Bigr),$$

with the convention $x_{n+1} := x_1$. Here Z_n is the normalizing constant.

Proposition 7.5. *The sequence μ_n admits a weakly convergent subsequence $\mu_{n_k} \to \mu$ satisfying the assumptions of Theorem 7.2.*

Proof. It was proved in Theorem 3.1 of [2] that any sequence of probability measures

$$\tilde{\mu}_n = c_n e^{-H_n} dx_{-n}\cdots dx_n,$$

where H_n is obtained from H by fixing a boundary condition $\tilde{x}$

$$H_n = \sum_{i=1}^{n} V(x_i) + \sum_{i=1}^{n-1} W(x_i,x_{i+1}) + W(x_n,\tilde{x}_1),$$

has a weakly convergent subsequence $\tilde{\mu}_{n_k} \to \tilde{\mu}$. In addition (see [2]), μ satisfies the following a priori estimate: for every $\lambda > 0$

$$\sup_{k\in\mathbb{N}} \int \exp(\lambda|x_k|^N)\ d\tilde{\mu} < \infty.$$

The same estimate holds for $\tilde{\mu}_n$ uniformly in n.

Following the reasoning from [2] it is easy to show that the sequence $\{\mu_n\}$ is tight and satisfies the same a priori estimate. Thus, we can pass to a subsequence $\{\mu_{n'}\}$ which weakly converges to a measure μ. For the sake of simplicity this subsequence will be denoted by $\{\mu_n\}$ again. The limiting measure μ satisfies

$$\sup_{k\in\mathbb{N}} \int \exp(\lambda|x_k|^N)\ d\mu < \infty, \tag{16}$$

moreover,

$$\sup_{n}\sup_{k\in\mathbb{N}} \int \exp(\lambda|x_k|^N)\ d\mu_n < \infty. \tag{17}$$

Let us estimate the relative entropy. We note that μ_n and μ_m $(n > m)$ are related in the following way:

$$\frac{e^Z\mu_n}{\int e^Z d\mu_n} = \mu_m \times \nu_{m,n},$$

where $Z = -W(x_m,x_1) + W(x_m,x_{m+1}) + W(x_n,x_1)$, and $\nu_{m,n}$ is a probability measure on $E_{m,n}$. Set: $\mu_{m,n} = \nu_{m,n}$. Then

$$\int \log\Bigl(\frac{d\mu_n}{d(\mu_m\times\mu_{m,n})}\Bigr)d\mu_n = \int (Z - \log\int e^Z\ d\mu_n)\ d\mu_n.$$

The desired bound follows immediately from (17) and the assumptions about W.

In order to prove assumption 3) we note that

$$\frac{\left[e^{W(x_n,x_{n+1})+W(x_1,x_n)}\cdot\mu\right]\circ P_n^{-1}}{\int e^{W(x_n,x_1)+W(x_1,x_n)}d\mu}=\frac{e^{W(x_1,x_n)}\cdot\mu_n}{\int e^{W(x_1,x_n)}\ d\mu_n}.$$

The normalizing constants can be easily estimated with the help of a priori bounds for μ and μ_n. Applying assumptions on W one can easily get that

$$Ae^{-B(|x_n|^{N-1}+|x_q|^{N-1})}\le\frac{d\mu_n}{d\mu\circ P_n^{-1}}\le Ae^{B(|x_n|^{N-1}+|x_1|^{N-1})}$$

where $A, B > 0$ do not depend on n. Hence, Assumption 3) follows immediately from (17), the Jensen inequality and convexity if the function $x^{-\varepsilon}$. □

Remark 7.6. Finally, let us briefly discuss when the transportation mapping obtained in Proposition 7.5 by Theorem 7.2 solves the corresponding optimal transportation problem. To this end we apply Proposition 7.4.

Following the estimates obtained in Proposition 7.5 and applying Jensen inequality one can easily show that the sequence of the entropies

$$\int\log\Big(\frac{d\hat{\mu}_n}{d\mu_n}\Big)\ d\hat{\mu}_n$$

is bounded. Then the assumption of Proposition 7.4 holds, for instance, if every μ_n satisfies the Talagrand inequality

$$W_2^2(\mu_n,\rho\cdot\mu_n)\le C\int\rho\log\rho d\mu_n$$

with constant which does not depends on n. We don't investigate here sufficient condition for measures μ_n to satisfy this inequality, we just mention that this clearly holds in many natural situations (e.g. under assumption of uniform log-concavity or finiteness of the log-Sobolev constant).

In addition, we emphasize, that in many applications the measures do indeed satisfy the Talagrand inequality, but Proposition 7.4 should actually work under much milder assumptions.

References

[1] Ambrosio L., Gigli N., Savaré G., Gradient flows in metric spaces and in the Wasserstein spaces of probability measures, Birkhäuser, 2008.

[2] Albeverio S., Kondratiev Yu.G., Röckner M., Tsikalenko T.V., A priori estimates for symmetrizing measures and their applications to Gibbs states, Journ. of Func. Anal., 171, 366–400, 2000.

[3] Beiglböck M., Cyclical monotonicity and the ergodic theorem,. Ergod. Theory and Dynam. Syst., 35(3), 710–713, 2015.

[4] Bogachev V.I., Measure theory. V. 1,2. Springer, Berlin – New York, 2007.

[5] Bogachev V.I., Kolesnikov A.V., On the Monge–Ampère equation in infinite dimensions, Infin. Dimen. Anal. Quantum Probab. Related Topics, 8(4), 547–572, 2005.

[6] Bogachev V.I., Kolesnikov A.V., Sobolev regularity for the Monge–Ampere equation in the Wiener space. Kyoto Jour. Math., 53(4), 713–738, 2013.

[7] Bogachev V.I., Kolesnikov A.V., The Monge–Kantorovich problem: achievements, connections, and perspectives, Russian Mathematical Surveys, 67(5), 785–890, 2012.

[8] Cavalletti F., The Monge problem in Wiener Space. Calcul. of Var.and PDE's, 45(1-2), 101–124, 2012.

[9] Contreras G., Lopes A.O., Oliveira E.R., Ergodic Transport Theory, periodic maximizing probabilities and the twist condition. In: Modeling, Dynamics, Optimization and Bioeconomics I, vol. 73 of the series Springer Proceedings in Mathematics and Statistics, 183–219, 2014.
[10] Fang S., Nolot V., Sobolev estimates for optimal transport maps on Gaussian spaces, Jour. Func. Anal., 266(8), 5045–5084, 2014.
[11] Fang S., Shao J. Optimal transport maps for Monge–Kantorovich problem on loop groups, Journ. of Func. Anal., 248(1), 225–257, 2007.
[12] Feyel D., Üstünel A.S., Monge–Kantorovich measure transportation and Monge–Ampère equation on Wiener space, Prob. Theory and Related Fields, 128, 347–385, 2004.
[13] Ghoussoub N., Moameni A., Symmetric Monge-Kantorovich problems and polar decompositions of vector fields, Geom. Func. Anal., 24(4), 1129–1166, 2014.
[14] Kallenberg O., Probabilistic symmetries and invariance principles, Springer-Verlag New York, 2005.
[15] Kolesnikov A.V., Convexity inequalities and optimal transport of infinite-dimensional measures, J. Math. Pures Appl. (9), 83(11), 1373–1404, 2004.
[16] Kolesnikov A.V., On Sobolev regularity of mass transport and transportation inequalities, Theory of Probability and its Applications, to appear. Translated from Teor. Veroyatnost. i Primenen., 57(2), 296–321, 2012.
[17] Kolesnikov A.V., Mass transportation and contractions, MIPT Proc. 2(4), 90–99, 2010.
[18] Kolesnikov A.V., Röckner M., On transport equation in infinite dimensions. Jour. Func. Anal. 266(7), 4490–4537, 2014.
[19] Lopes A.O., Mengue J.K., Duality Theorems in Ergodic Transport, Journ Stat. Phys., 149(5), 921–942, 2012.
[20] Lopes A.O., Oliveira E.O.,Thieullen P., The Dual Potential, the involution kernel and Transport in Ergodic Optimization, Dynamics, Games and Science Vol. 1 of the series CIM Series in Mathematical Science, 357–398.
[21] Moameni A., Invariance properties of the Monge-Kantorovich mass transport problem, arXiv: 1311.7051.
[22] Rüschendorf L., Sei T., On optimal stationary couplings between stationary processes Electr. Journ. Probab., 17, article 17, 2012.
[23] Vershik A.M., The problem of describing central measures on the path spaces of graded graphs., Func.l Analysis and Its Appl., 48(4), 256–271, 2014.
[24] Villani C., Topics in optimal transportation, Amer. Math. Soc. Providence, Rhode Island, 2003.
[25] Zaev D.A., On the Monge-Kantorovich problem with additional linear constraints, Mat. Zametki, 98(5), 664–683, 2015.
[26] Zaev D.A., On ergodic decompositions related to the Kantorovich problem, Zapiski POMI, 437, 100–130, 2015.

Higher School of Economics, Moscow, Russia
E-mail address: Sascha77@mail.ru

Higher School of Economics, Moscow, Russia
E-mail address: zaev.da@gmail.com