Journal of Convex Analysis
Volume 25 (2018), No. 4, 1371–1395

Fréchet Barycenters in the Monge-Kantorovich Spaces*

Alexey Kroshnin
Moscow Institute of Physics and Technology, Institute for Information Transmission Problems, 9 Institutskiy per., Dolgoprudny – Moscow Region 141701, Russia
kroshnin@phystech.edu

Received: June 19, 2017
Accepted: October 22, 2017

We consider the space $\mathcal{P}(X)$ of probability measures on arbitrary Radon space X endowed with a transportation cost $J(\mu, \nu)$ generated by a nonnegative continuous cost function. For a probability distribution on $\mathcal{P}(X)$ we formulate a notion of average with respect to this transportation cost, called here the *Fréchet barycenter*, prove a version of the law of large numbers for Fréchet barycenters, and discuss the structure of $\mathcal{P}(X)$ related to the transportation cost J.

Keywords: Optimal transport, Wasserstein space, Wasserstein barycenter, law of large numbers.

2010 Mathematics Subject Classification: 60D05, 28C99, 54E40.

1. Introduction

In this paper we consider averaging in the space $\mathcal{P}(X)$ of probability measures over a Radon space X, using a transport optimization procedure to define a suitable concept of a "typical element", which extends the notion of *Fréchet mean.* For the first time a construction of this kind was introduced by M. Agueh and G. Carlier in [1]: a *Wasserstein barycenter* of a family of measures on the Euclidean space $\mathbb{R}^d$ is defined as the Fréchet mean using the 2-Wasserstein distance W_2 on $\mathcal{P}(\mathbb{R}^d)$, which is given by minimization of the mean-square displacement. In [1], the authors establish existence, uniqueness, and regularity results for the Wasserstein barycenter and, when $d = 1$, provide an explicit formula for the Wasserstein barycenter in terms of quantile functions of the measures involved.

*The research was performed at IITP RAS with support of a Russian Science Foundation grant, project no. 14-50-00150.

ISSN 0944-6532 / $ 2.50

Here we take a general Radon space X (e.g. Polish space) and a general transportation cost

$$J(\mu, \nu) = \inf \left\{ \int c(x, y)\, d\gamma(x, y) : \gamma \in \mathcal{P}(X \times X),\ \pi^x_{\#}\gamma = \mu,\ \pi^y_{\#}\gamma = \nu \right\} \geq 0,$$

where $c(\cdot, \cdot) \geq 0$ is a continuous *cost function* that satisfies $c(x, y) = 0$ iff $x = y$. Since $J(\mu, \nu) = 0$ iff $\mu = \nu$, this cost quantifies separation between measures μ and ν in $\mathcal{P}(X)$ but does not necessarily satisfy the triangle inequality. Although $J(\cdot, \cdot)$ is not a metric, it generates a *transportation topology* on X, and the space X endowed with this topology is divided into equivalence classes each of which is a Radon space (in particular, the classes are separable and metrizable).

Let a measure ν be fixed and $\boldsymbol{\mu}$ be a random element of $\mathcal{P}(X)$ with distribution P_μ. We introduce a notion of Fréchet typical element of P_μ with respect to $J(\cdot, \cdot)$, which we propose to call the *Fréchet barycenter* of P_μ. It is defined as any measure ν for which the expected cost

$$\mathbb{E}\, J(\boldsymbol{\mu}, \nu) = \int_{\mathcal{P}(X)} J(\mu, \nu)\, dP_\mu$$

attains its minimum over $\mathcal{P}(X)$. Rigorous definitions of such a distribution and an integral are formulated in Section 5. Suppose $\mathbb{E}\, J(\boldsymbol{\mu}, \cdot)$ is not identically equal to $+\infty$ on $\mathcal{P}(X)$. Then there exists a Fréchet barycenter of P_μ. This averaging appears to be quite reasonable. Namely, if distributions P_n converge to P with respect to the transportation cost in $\mathcal{P}\big(\mathcal{P}(X)\big)$ corresponding to $J(\cdot, \cdot)$ as cost function, then barycenters of P_n also converge in some sense to the barycenter of P. For instance, this result implies a law of large numbers for Fréchet barycenters.

The paper is organized as follows. In Section 2 we introduce some standard definitions and notations and recall properties of the Monge–Kantorovich distance. In Section 3 properties of the topology on $\mathcal{P}(X)$ induced by the Monge–Kantorovich transportation cost $J(\cdot, \cdot)$ are considered. In Section 4 we consider the particular case of $X = \mathbb{R}^d$ and $c(x, y) = g(x - y)$, where $g(\cdot) \geq 0$ is a convex function. Then we define in Section 5 the generalized barycenter of distribution on $\mathcal{P}(X)$ for general X. The central result of this paper is proved in Subsection 5.2: the convergence of barycenters of distributions P_n is established provided P_n themselves converge to some distribution P.

2. The Monge–Kantorovich distance

2.1. Notations

For a measurable space X denote the space of probability measures on X by $\mathcal{P}(X)$. In particular, if the space X is topological, we assume it is endowed with the standard Borel σ-algebra $\mathcal{B}(X)$.

For two measurable spaces X, Y, a measurable map $T\colon X \to Y$ induces a map $T_\#\colon \mathcal{P}(X) \to \mathcal{P}(Y)$ given by $T_\#\mu(A) := \mu\big(T^{-1}(A)\big)$ for any measurable $A \subset Y$. Recall that for any integrable function f

$$\int_Y f(y)\, d(T_\#\mu) = \int_X f\big(T(x)\big)\, d\mu.$$

For a measure μ on X and an integrable function $f\colon X \to \mathbb{R}$ define the measure $f\lfloor\mu$ as

$$(f\lfloor\mu)(A) := \int_A f(x)\, d\mu$$

for any measurable $A \subset Y$. Moreover, for any measurable set B define $B\lfloor\mu := \chi_B\lfloor\mu$, where $\chi_B(\cdot)$ is the characteristic function of B.

We will often drop the argument of function and the symbol of domain of integration if there is no risk of confusion.

2.2. The transport functional

Let (X, ρ) be a *Radon space*, i.e. a separable metric space such that for every Borel probability measure $\mu \in \mathcal{P}(X)$ and $\epsilon > 0$ there exists a compact set D^μ_ϵ for which $\mu(X \setminus D^\mu_\epsilon) < \epsilon$ [2, def. 5.1.4]. E.g., any Polish space is a Radon space.

Fix a measurable nonnegative function $c\colon X \times X \to [0, \infty)$ and call it the *cost function*. Assume $c(\cdot, \cdot)$ is continuous and *consistent* in the sense that $c(x, x_n) \to 0$ iff $c(x_n, x) \to 0$ iff $x_n \to x$ for any $x \in X$, $\{x_n\}_{n\in\mathbb{N}} \subset X$ (cf. discussion after Assumption 3.13). In particular, $c(x, y) = 0$ iff $x = y$.

For two measures $\mu, \nu \in \mathcal{P}(X)$ define the set of *transport plans* taking μ to ν as

$$\Pi(\mu, \nu) := \big\{\gamma \in \mathcal{P}(X \times X) : \pi^x_\#\gamma = \mu,\ \pi^y_\#\gamma = \nu\big\},$$

where π^x and π^y are projections of $X \times X$ to the first and second factor respectively. Observe that $\Pi(\mu, \nu)$ is always nonempty because it contains the direct product measure $\mu \otimes \nu$.

The *transportation cost* of a transport plan γ is defined as

$$K(\gamma) := \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\, d\gamma.$$

It is easy to obtain that $K(\cdot)$ is lower semicontinuous with respect to the weak convergence of measures. The *Monge–Kantorovich problem* for given $\mu, \nu \in \mathcal{P}(X)$ consists in minimizing the transportation cost $K(\gamma)$ over all $\gamma \in \Pi(\mu, \nu)$. Accordingly, the *Monge–Kantorovich distance* (or *transportation functional*) between μ and ν is the infimum of transportation costs:

$$J(\mu, \nu) := \inf_{\gamma\in\Pi(\mu,\nu)} K(\gamma).$$

A transport plan γ^* is called *optimal* if $K(\cdot)$ attains its minimum over $\Pi(\mu,\nu)$ at γ^*.

Theorem 2.1. *For any $\mu, \nu \in \mathcal{P}(X)$ there exists an optimal transport plan.*

Proof. Let $J(\mu,\nu) < \infty$ and $\{\gamma_n\}_{n\in\mathbb{N}} \subset \Pi(\mu,\nu)$ be a minimizing sequence for $K(\cdot)$. Observe that μ and ν are tight measures as X is a Radon space, hence the set of transport plans $\Pi(\mu,\nu)$ is also tight. Indeed, for any $\epsilon > 0$ there exist such compact sets D^ϵ_μ, D^ϵ_ν that $\mu\left(X \setminus D^\epsilon_\mu\right) < \epsilon/2$ and $\nu\left(X \setminus D^\epsilon_\nu\right) < \epsilon/2$; then $D^\epsilon := D^\epsilon_\mu \times D^\epsilon_\nu \subset X \times X$ is compact and for any $\gamma \in \Pi(\mu,\nu)$ it holds

$$\begin{aligned}\gamma\left(X \times X \setminus D^\epsilon\right) \leq \gamma\big(X \times (X \setminus D^\epsilon_\nu)\big) + \gamma\big(\big(X \setminus D^\epsilon_\mu\big) \times X\big) = \\ = \nu\left(X \setminus D^\epsilon_\nu\right) + \mu\left(X \setminus D^\epsilon_\mu\right) < \epsilon.\end{aligned}$$

Then by Prokhorov's theorem one can choose a weakly convergent subsequence $\gamma_{n_k} \rightharpoonup \gamma^* \in \Pi(\mu,\nu)$. Since $K(\cdot)$ is lower semicontinuous with respect to the weak convergence,

$$K(\gamma^*) \leq \liminf K(\gamma_{n_k}) = \inf_{\gamma\in\Pi(\mu,\nu)} K(\gamma) := J(\mu,\nu),$$

i.e. γ^* is an optimal transport plan from μ to ν. $\square$

Notice, however, that optimal transport plan from μ to ν may be not unique.

Corollary 2.2. $J(\mu,\nu) = 0$ *iff* $\mu = \nu$.

Let us recall some properties of the transportation functional $J(\cdot,\cdot)$ (see e.g. [2, 9, 10]).

Lemma 2.3. *The functional $J(\cdot,\cdot)$ is convex, i.e. for any measures μ_0, ν_0, μ_1, $\nu_1 \in \mathcal{P}(X)$ and $t \in [0,1]$ it holds*

$$J(\mu_t,\nu_t) \leq (1-t)J(\mu_0,\nu_0) + tJ(\mu_1,\nu_1),$$

where $\mu_t := (1-t)\mu_0 + t\mu_1$, $\nu_t := (1-t)\nu_0 + t\nu_1$.

Proof. Let γ_s be an optimal transport plan from μ_s to ν_s, $s \in \{0,1\}$. Then $\gamma_t := (1-t)\gamma_0 + t\gamma_1 \in \Pi(\mu_t,\nu_t)$, and hence

$$\begin{aligned}J(\mu_t,\nu_t) \leq K(\gamma_t) = (1-t)K(\gamma_0) + tK(\gamma_1) = \\ = (1-t)J(\mu_0,\nu_0) + tJ(\mu_1,\nu_1).\end{aligned}$$ $\square$

Corollary 2.4. *Fix $\mu_0, \mu_1 \in \mathcal{P}(X)$. Then for any $0 \leq t < t' \leq 1$*

$$J(\mu_t,\mu_{t'}) \leq (t'-t)J(\mu_0,\mu_1).$$

Proof. From the convexity of $J(\cdot,\cdot)$ it follows

$$J(\mu_t,\mu_{t'}) \le \frac{1-t'}{1-t} J(\mu_t,\mu_t) + \frac{t'-t}{1-t} J(\mu_t,\mu_1) \le \\ \le \frac{t'-t}{1-t}\big((1-t)J(\mu_0,\mu_1) + tJ(\mu_1,\mu_1)\big) = (t'-t)J(\mu_0,\mu_1). \qquad \square$$

Lemma 2.5. *The functional $J(\cdot,\cdot)$ is lower semicontinuous with respect to the weak convergence of measures.*

Proof. Let $\mu_n \rightharpoonup \mu$, $\nu_n \rightharpoonup \nu$ and $\gamma_n \in \Pi(\mu_n,\nu_n)$ be an optimal transport plan from μ_n to ν_n. Notice that $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\nu_n\}_{n\in\mathbb{N}}$ weakly converge so they are tight [2, p. 108] and $\{\gamma_n\}_{n\in\mathbb{N}}$ is the same. Consider subsequence whose lower limit $\liminf J(\mu_n,\nu_n) \in [0,\infty]$ is attained. It contains a weakly convergent subsequence $\gamma_{n_k} \rightharpoonup \tilde{\gamma} \in \Pi(\mu,\nu)$. So, due to the lower semicontinuity of $K(\cdot)$ one can obtain

$$J(\mu,\nu) \le K(\tilde{\gamma}) \le \liminf K(\gamma_{n_k}) = \lim J(\mu_{n_k},\nu_{n_k}) = \liminf J(\mu_n,\nu_n). \qquad \square$$

Corollary 2.6. *$J(\cdot,\cdot)$ is measurable with respect to the product of Borel σ-algebras $\mathcal{B}_w\big(\mathcal{P}(X)\big) \times \mathcal{B}_w\big(\mathcal{P}(X)\big)$ induced by the topology of weak convergence.*

Theorem 2.7. *Let $\{\nu_n\}_{n\in\mathbb{N}}$ be a sequence of measures from $\mathcal{P}(X)$ such that $J(\nu^*,\nu_n) \to 0$ (or $J(\nu_n,\nu^*) \to 0$) for some $\nu^* \in \mathcal{P}(X)$; then $\nu_n \rightharpoonup \nu^*$.*

Proof. Assume $\{\nu_n\}_{n\in\mathbb{N}}$ fails to converge to ν^*. Then there exists a closed set $F \subset X$ such that $\limsup \nu_n(F) > \nu^*(F)$. Let $3\epsilon := \lim \nu_n(F) - \nu^*(F) > 0$ without relabelling. Consider the following open neighbourhood of F:

$$F_r := \left\{ x \in X : \inf_{y\in F} c(x,y) < r \right\} \supset F,\ r > 0.$$

For any $r > 0$ set F_r is open due to continuity of $c(\cdot,\cdot)$. One can obtain that for any $x \notin F$ there exist an "open ball" $B_r^c(x) := \big\{y \in X : c(x,y) < r\big\}$ such that $B_r^c(x) \cap F = \emptyset$. Therefore $\bigcap_{r>0} F_r = F$ and $\nu^*(F) = \lim_{r\to 0} \nu^*(F_r)$. Let $r_0 > 0$ be such that $\nu^*(F_{r_0}) < \nu^*(F) + \epsilon$. As $\nu_n(F) > \nu^*(F) + 2\epsilon$ starting from some n,

$$\gamma_n\big((X \setminus F_{r_0}) \times F\big) = \nu_n(F) - \gamma_n\left(F_{r_0} \times F\right) \ge \nu_n(F) - \nu^*\left(F_{r_0}\right) > \epsilon,$$

where $\gamma_n \in \Pi(\nu^*,\nu_n)$ is an optimal transport plan. Consequently,

$$J(\nu^*,\nu_n) = K(\gamma_n) \ge \gamma_n\big((X \setminus F_{r_0}) \times F\big) \inf\big\{c(x,y) : x \notin F_{r_0},\, y \in F\big\} \ge \epsilon r_0$$

and $\liminf J(\nu^*,\nu_n) \ge \epsilon r_0 > 0$, which contradicts the assumption. $\square$

As we have seen, convergence with respect to the transportation functional implies the weak convergence. Actually, the converse also holds under some additional assumptions.

Theorem 2.8. *Let $\nu_n \rightharpoonup \nu^*$ and $\operatorname{supp}\nu_n \subset F$ for all n, where $F \subset X$ is a closed set such that $\sup\limits_{x,y\in F} c(x,y) < \infty$. Then $\lim J(\nu_n,\nu^*) = \lim J(\nu^*,\nu_n) = 0$.*

Proof. Fix $\epsilon > 0$. Due to separability of X and continuity of $c(\cdot,\cdot)$ one can cover X with a countable union of closed balls $\{\overline{B}_{r_i}(x_i)\}_{i\in\mathbb{N}}$ such that $c(x,y) < \epsilon$ whenever $x,y \in \overline{B}_{2r_i}(x_i)$ for all i. Fix $m \in \mathbb{N}$ such that $\nu^*\big(X\setminus\bigcup_{i=1}^m B_{r_i}(x_i)\big) < \epsilon$ and consider continuous functions $f_i(\cdot)\colon X \to [0,1]$, $0 \le i \le m$ which satisfy $\sum_{i=0}^m f_i(x) \equiv 1$, $f_0(x) = 0$ for $x \in \overline{B}_{r_i}(x_i)$, and $f_i(x) = 0$ for $x \notin \overline{B}_{2r_i}(x_i)$, $1 \le i \le m$.

Without loss of generality $\int f_i\,d\nu^* > 0$ for all i. Define measures

$$\lambda_n^i := \frac{(f_i\lfloor\nu_n)\otimes(f_i\lfloor\nu^*)}{\max\left\{\int f_i\,d\nu_n, \int f_i\,d\nu^*\right\}},\ 1 \le i \le m,\ n \in \mathbb{N};$$

$$\hat{\nu}_n := \nu_n - \sum_{i=1}^m \pi^x_\#\lambda_n^i = \left(1 - \sum_{i=1}^m \frac{\int f_i\,d\nu^*}{\max\left\{\int f_i\,d\nu_n, \int f_i\,d\nu^*\right\}} f_i\right)\lfloor\nu_n \ge f_0\lfloor\nu_n \ge 0;$$

$$\hat{\nu}_n^* := \nu^* - \sum_{i=1}^m \pi^y_\#\lambda_n^i = \left(1 - \sum_{i=1}^m \frac{\int f_i\,d\nu_n}{\max\left\{\int f_i\,d\nu_n, \int f_i\,d\nu^*\right\}} f_i\right)\lfloor\nu^* \ge f_0\lfloor\nu^* \ge 0.$$

Since $\nu_n \rightharpoonup \nu^*$

$$\hat{\nu}_n(X) = \hat{\nu}_n^*(X) = \int\left(1 - \sum_{i=1}^m \frac{\int f_i\,d\nu_n}{\max\left\{\int f_i\,d\nu_n, \int f_i\,d\nu^*\right\}} f_i\right) d\nu^*$$
$$\to \int\left(1 - \sum_{i=1}^m f_i\right) d\nu^* = \int f_0\,d\nu^* \le \nu^*\left(X \setminus \bigcup_{i=1}^m B_{r_i}(x_i)\right) < \epsilon.$$

Consider the following transport plans:

$$\gamma_n := \frac{\hat{\nu}_n \otimes \hat{\nu}_n^*}{\hat{\nu}_n(X)} + \sum_{i=1}^m \lambda_n^i \in \Pi(\nu_n,\nu^*),\ n \in \mathbb{N}.$$

From $\operatorname{supp}\nu_n \subset F$ it follows $\operatorname{supp}\nu^* \subset F$ and $\operatorname{supp}\gamma_n \subset F \times F$. Define $M := \sup_{x,y\in F} c(x,y) < \infty$. $\operatorname{supp}\lambda_n^i \subset \overline{B}_{2r_i}(x_i) \times \overline{B}_{2r_i}(x_i)$ by the definition of functions $f_i(\cdot)$. Now one can obtain that

$$\begin{aligned}\limsup J(\nu_n,\nu^*) &\le \limsup K(\gamma_n)\\ &\le \limsup\left(M\frac{\hat{\nu}_n(X)\hat{\nu}_n^*(X)}{\hat{\nu}_n(X)} + \epsilon\sum_{i=1}^m \lambda_n^i(X\times X)\right)\\ &\le M\limsup \hat{\nu}_n^*(X) + \epsilon \le (1+M)\,\epsilon.\end{aligned}$$

It proves that $J(\nu_n, \nu^*) \to 0$ because of the arbitrary choice of ϵ. In the same way one can show that $J(\nu^*, \nu_n) \to 0$. $\square$

3. Transportation topology

Assume that the cost function $c(\cdot,\cdot)$ satisfies the following weak triangle inequalities.

Assumption 3.1. There exist constants $A, B \geq 0$ such that the following set of inequalities holds for all $x, y, z \in X$:

$$\begin{aligned} c(x,y) &\leq A + B\big(c(x,z) + c(y,z)\big), \quad c(x,y) \leq A + B\big(c(x,z) + c(z,y)\big), \\ c(x,y) &\leq A + B\big(c(z,x) + c(y,z)\big), \quad c(x,y) \leq A + B\big(c(z,x) + c(z,y)\big). \end{aligned}$$

This is a quite natural assumption which holds for a wide class of functions, e.g. for $c(x,y) = \rho^p(x,y)$, where $\rho(\cdot,\cdot)$ is a metric on X and $p > 0$ (the case of p-Wasserstein spaces).

Let us now show that the Monge-Kantorovich distance “inherits” the inequalities for the cost function.

Lemma 3.2. *For all $\mu, \nu, \lambda \in \mathcal{P}(X)$*

$$\begin{aligned} J(\mu,\nu) &\leq A + B\big(J(\mu,\lambda) + J(\nu,\lambda)\big), \quad J(\mu,\nu) \leq A + B\big(J(\mu,\lambda) + J(\lambda,\nu)\big), \\ J(\mu,\nu) &\leq A + B\big(J(\lambda,\mu) + J(\nu,\lambda)\big), \quad J(\mu,\nu) \leq A + B\big(J(\lambda,\mu) + J(\lambda,\nu)\big). \end{aligned}$$

Proof. Consider measures $\mu, \nu, \lambda \in \mathcal{P}(X)$ and optimal plans $\gamma_1 \in \Pi(\mu, \lambda)$, $\gamma_2 \in \Pi(\nu, \lambda)$. By disintegration theorem [2, Theorem 5.3.1] there exists a measure $\sigma \in \Pi(\mu, \nu, \lambda)$ such that $\pi^{x,z}_{\#}\sigma = \gamma_1$ and $\pi^{y,z}_{\#}\sigma = \gamma_2$ [2, Lemma 5.3.2]. Now, applying the weak triangle inequality, one can obtain

$$\begin{aligned} J(\mu,\nu) \leq K(\pi^{x,y}_{\#}\sigma) &= \int c(x,y)\, d\sigma \leq \int \Big(A + B\big(c(x,z) + c(y,z)\big)\Big)\, d\sigma \\ &= A + B\big(K(\gamma_1) + K(\gamma_2)\big) = A + B\big(J(\mu,\lambda) + J(\nu,\lambda)\big). \end{aligned}$$

The other inequalities might be proved similarly. $\square$

Notice that if $c(\cdot,\cdot)$ is a metric on X, then $J(\cdot,\cdot)$ is a metric on $\mathcal{P}(X)$ (which may make the value $+\infty$). Moreover, it is an inner metric, even if X is a disconnected space. Indeed, if $J(\mu,\nu) < \infty$, then the curve given by $[0,1] \ni t \mapsto (1-t)\mu + t\nu$ is a minimizing geodesic connecting μ to ν due to Corollary 2.4.

Lemma 3.3. *Let $D \subset X$ be a compact set. Then for any $\epsilon > 0$ there exist an open neighbourhood $U_\epsilon(D) \supset D$ and a constant B^D_ϵ such that*

$$c(x,y) \leq \epsilon + c(x,z) + B^D_\epsilon c(y,z) \quad \textit{and} \quad c(x,y) \leq \epsilon + c(z,y) + B^D_\epsilon c(z,x)$$

for all $x, y, z \in U_\epsilon(D)$.

Proof. Fix $\epsilon > 0$. As $c(\cdot,\cdot)$ is continuous, it is uniformly continuous on $D \times D$, hence there exists a relatively open subset $V \subset D \times D$ such that $(y, y) \in V$ for all $y \in D$ and $c(x, y) < c(x, z) + \epsilon/2$ for all $x \in X$, $(y, z) \in V$. Define

$$M := \max_{x,y \in D} c(x, y) < \infty \quad \text{and} \quad \delta := \min_{(y,z) \in D^2 \setminus V} c(y, z) > 0$$

which is positive due to compactness of D. If $(y, z) \in D^2 \setminus V$ then $c(x, y) \leq M \leq \frac{M}{\delta} c(y, z)$ for all $x \in D$. Consequently,

$$c(x, y) \leq \epsilon/2 + c(x, z) + \frac{M}{\delta} c(y, z) \tag{1}$$

for all $x, y, z \in D$. Denote $\frac{M}{\delta}$ by B^D_ϵ.

Due to the continuity of $c(\cdot,\cdot)$ one can choose an open neighbourhood W of D^3 such that for all $x, y, z \in W$ there exist $x', y', z' \in D$ for which $|c(x, y) - c(x', y')| < \gamma$, $|c(x, z) - c(x', z')| < \gamma$ and $|c(y, z) - c(y', z')| < \gamma$ where $\gamma := \epsilon / \big(6(1 + B^D_\epsilon)\big)$. As D^3 is compact, $\rho(D^3, X^3 \setminus W) > 0$ and there exists such neighbourhood $U_\epsilon(D)$ that $\big(U_\epsilon(D)\big)^3 \subset W$. Now, applying inequality (1) and the definition of W one can obtain that

$$\begin{aligned} c(x, y) &\leq \gamma + c(x', y') \leq \gamma + \epsilon/2 + c(x', z') + B^D_\epsilon c(y', z') \\ &\leq \gamma \left(2 + B^D_\epsilon\right) + \epsilon/2 + c(x, z) + B^D_\epsilon c(y, z) \\ &\leq \epsilon + c(x, z) + B^D_\epsilon c(y, z) \end{aligned}$$

for all $x, y, z \in U_\epsilon(D)$. The last inequality can be treated in the same way. $\square$

Lemma 3.4. (Continuity) *Take two sequences $\{\mu_n\}_{n\in\mathbb{N}}$, $\{\nu_n\}_{n\in\mathbb{N}}$ such that $J(\mu^*, \mu_n) \to 0$ and $J(\nu^*, \nu_n) \to 0$ for some measures $\mu^*, \nu^* \in \mathcal{P}(X)$. Then $J(\mu_n, \nu_n) \to J(\mu^*, \nu^*)$.*

Proof. Let $\gamma^1_n \in \Pi(\mu^*, \mu_n)$, $\gamma^2 \in \Pi(\mu^*, \nu^*)$, $\gamma^3_n \in \Pi(\nu^*, \nu_n)$ be optimal transport plans. Consider measures $\sigma_n \in \mathcal{P}(X^4)$ such that $(\pi^{x_2} \times \pi^{x_1})_\# \sigma_n = \gamma^1_n$, $(\pi^{x_2} \times \pi^{x_3})_\# \sigma_n = \gamma^2$ and $(\pi^{x_3} \times \pi^{x_4})_\# \sigma_n = \gamma^3_n$. Since the sequences are tight one can fix $\epsilon > 0$ and a compact set D such that $\mu_n(X \setminus D)$, $\mu^*(X \setminus D)$, $\nu_n(X \setminus D)$, $\nu^*(X \setminus D)$ and $\int_{X^2 \setminus D^2} c(x_2, x_3)\, d\gamma^2$ are less than ϵ. Obviously

$$J(\mu_n, \nu_n) \leq K\big((\pi^{x_1} \times \pi^{x_4})_\# \sigma_n\big) = \int c(x_1, x_4)\, d\sigma_n.$$

Consider the set $Y := U_\epsilon(D) \times D^2 \times U_\epsilon(D)$. Now one can obtain due to Lemma 3.3 that

$$
\begin{aligned}
&\int_Y c(x_1,x_4)\,d\sigma_n \le \int_Y \big(\epsilon + B_\epsilon^D c(x_2,x_1) + (1+\epsilon)c(x_2,x_4)\big)\,d\sigma_n \le \\
&\le \int_Y \big(\epsilon + B_\epsilon^D c(x_2,x_1) + (1+\epsilon)\epsilon + (1+\epsilon)B_\epsilon^D c(x_3,x_4) + (1+\epsilon)^2 c(x_2,x_3)\big)\,d\sigma_n \\
&\le \epsilon + B_\epsilon^D K(\gamma_n^1) + (1+\epsilon)\epsilon + (1+\epsilon)B_\epsilon^D K(\gamma_n^3) + (1+\epsilon)^2 K(\gamma^2) = \\
&= \epsilon + B_\epsilon^D J(\mu^*,\mu_n) + (1+\epsilon)\epsilon + (1+\epsilon)B_\epsilon^D J(\nu^*,\nu_n) + (1+\epsilon)^2 J(\mu^*,\nu^*) \\
&\xrightarrow[n\to\infty]{} \epsilon + (1+\epsilon)\epsilon + (1+\epsilon)^2 J(\mu^*,\nu^*) \to J(\mu^*,\nu^*) \text{ as } \epsilon \to 0.
\end{aligned}
$$

The remaining term may be bounded by Assumption 3.1 in the following way:

$$
\begin{aligned}
&\int_{X^4\setminus Y} c(x_1,x_4)\,d\sigma_n \le \int_{X^4\setminus Y} \big(A + Bc(x_2,x_1) + Bc(x_2,x_4)\big)\,d\sigma_n \le \\
&\le \int_{X^4\setminus Y} \big(A + Bc(x_2,x_1) + AB + B^2 c(x_2,x_3) + B^2 c(x_3,x_4)\big)\,d\sigma_n \le \\
&\le (A+AB)\sigma_n(X^4\setminus Y) + BJ(\mu^*,\mu_n) + B^2\int_{X^4\setminus Y} c(x_2,x_3)\,d\sigma_n + B^2 J(\nu^*,\nu_n) \\
&\le 4(A+AB)\epsilon + BJ(\mu^*,\mu_n) + B^2 J(\nu^*,\nu_n) + B^2\int_{X^4\setminus Y} c(x_2,x_3)\,d\sigma_n.
\end{aligned}
$$

Notice that

$$
X^4\setminus Y = \Big[X\times(X^2\setminus D^2)\times X\Big]\cup\Big[\big(X\setminus U_\epsilon(D)\big)\times D^2\times X\Big]\cup\Big[X\times D^2\times\big(X\setminus U_\epsilon(D)\big)\Big]
$$

and

$$
\int_{X\times(X^2\setminus D^2)\times X} c(x_2,x_3)\,d\sigma_n = \int_{X^2\setminus D^2} c(x_2,x_3)\,d\gamma^2 < \epsilon.
$$

Moreover, since $J(\mu^*,\mu_n)\to 0$ and $J(\nu^*,\nu_n)\to 0$, $\gamma_n^1 \rightharpoonup (x\times x)_\#\mu^*$ and $\gamma_n^3 \rightharpoonup (x\times x)_\#\nu^*$, so $\sigma_n \rightharpoonup (\pi^x\times\pi^x\times\pi^y\times\pi^y)_\#\gamma^2$. Since $\big(X\setminus U_\epsilon(D)\big)\times D^2\times X$ is a closed set and $c(x_2,x_3)$ is continuous and bounded on it we have

$$
\limsup \int_{(X\setminus U_\epsilon(D))\times D^2\times X} c(x_2,x_3)\,d\sigma_n \le \int_{(D\cap X\setminus U_\epsilon(D))\times D} c(x,y)\,d\gamma^2 = 0
$$

as $D\cap X\setminus U_\epsilon(D) = \emptyset$. In the same way one can obtain that

$$
\limsup \int_{X\times D^2\times(X\setminus U_\epsilon(D))} c(x_2,x_3)\,d\sigma_n = 0.
$$

Thus $J(\mu_n,\nu_n)\to J(\mu^*,\nu^*)$. $\square$

Corollary 3.5. (Topology) *The balls* $B_r^J(\mu) := \{\nu \in \mathcal{P}(X) : J(\mu,\nu) < r\}$ *form a basis of a* topology τ_J *on* $\mathcal{P}(X)$*, and* $J(\cdot,\cdot)$ *is continuous with respect to this topology.*

Let us denote the convergence of a sequence $\{\nu_n\}_{n\in\mathbb{N}}$ to ν^* in the topology τ_J (*the transportation convergence*) by $\nu_n \xrightarrow{J} \nu^*$. It is equivalent to $J(\nu_n,\nu^*) \to 0$ and $J(\nu^*,\nu_n) \to 0$. If X is a compact space, it follows from Theorems 2.7 and 2.8 that weak convergence of measures is equivalent to the transportation convergence. But in this case the space $\mathcal{P}(X)$ with the topology of weak convergence is compact itself, and so is $(\mathcal{P}(X),\tau_J)$. Notice that if X is not compact, $\mathcal{P}(X)$ is neither compact nor locally compact.

Lemma 3.6. *Let* $\{\mu_n\}_{n\in\mathbb{N}}$*,* $\{\nu_n\}_{n\in\mathbb{N}}$*,* $\{\lambda_n\}_{n\in\mathbb{N}}$ *be tight sequences such that* $J(\mu_n,\nu_n) \to 0$ *and* $J(\nu_n,\lambda_n) \to 0$*; then* $J(\nu_n,\mu_n) \to 0$ *and* $J(\mu_n,\lambda_n) \to 0$*.*

Proof. Fix $\epsilon > 0$ and a compact set $D \subset X$ such that $\mu_n(D)$, $\nu_n(D)$ and $\lambda_n(D)$ are greater than $1-\epsilon$. Consider again measures σ_n such that $\pi^{x,y}_{\#}\sigma_n = \gamma^1_n$ and $\pi^{y,z}_{\#}\sigma_n = \gamma^2_n$ where γ^1_n and γ^2_n are optimal transport plans from μ_n to ν_n and from ν_n to λ_n, respectively. Due to Assumption 3.1 and Lemma 3.3 one can obtain that

$$\begin{aligned}
&J(\mu_n,\lambda_n) \le K\big((\pi^x\times\pi^z)_{\#}\sigma_n\big) = \int c(x,z)\,d\sigma_n \\
&\le \int\limits_{D^3}\big(\epsilon + (1+\epsilon)c(x,y) + B^D_\epsilon c(y,z)\big)\,d\sigma_n + \int\limits_{X^3\setminus D^3}\big(A + Bc(x,y) + Bc(y,z)\big)\,d\sigma_n \\
&\le \epsilon + (1+\epsilon)J(\mu_n,\nu_n) + B^D_\epsilon J(\nu_n,\lambda_n) + 3\epsilon A + BJ(\mu_n,\nu_n) + BJ(\nu_n,\lambda_n) \\
&\to \epsilon + 3\epsilon A \to 0 \quad \text{as} \quad \epsilon \to 0.
\end{aligned}$$

Thus $J(\mu_n,\lambda_n) \to 0$. Similarly,

$$\begin{aligned}
J(\nu_n,\mu_n) &\le K\big((\pi^y\times\pi^x)_{\#}\gamma^1_n\big) = \int c(y,x)\,d\gamma^1_n \\
&\le \int_{D^3}\big(\epsilon + B^D_\epsilon c(x,y)\big)\,d\gamma^1_n + \int_{X^3\setminus D^3}\big(A + Bc(x,y)\big)\,d\gamma^1_n \\
&\le \epsilon + B^D_\epsilon J(\mu_n,\nu_n) + 2\epsilon A + BJ(\mu_n,\nu_n) \to \epsilon + 2\epsilon A,
\end{aligned}$$

and therefore $J(\nu_n,\mu_n) \to 0$. □

Let the relation $\mu \sim \nu$ be defined as $J(\mu,\nu) < \infty$. Then it is an equivalence on $\mathcal{P}(X)$ and splits the space into equivalence classes $E(\mu) := \{\nu \in \mathcal{P}(X) : J(\mu,\nu) < \infty\}$. Notice that every equivalence class is path-connected, even if X is disconnected, since curve $[0,1] \ni t \mapsto (1-t)\mu + t\nu$ is continuous by Corollary 2.4, whenever $J(\mu,\nu) < \infty$. Let us denote by E_0 the class containing delta-measures, i.e.

$$E_0 := E(\delta_{x_0}) = \left\{ \nu \in \mathcal{P}(X) : \int c(x, x_0)\, d\nu(x) < \infty \right\}$$

for an arbitrary $x_0 \in X$ (obviously, it does not depend on the choice of x_0).

Consider the following useful construction: fix some point $x_0 \in X$ and for given $R > 0$ take a continuous function $f_R \colon X \times X \to [0, 1]$ such that $f_R(x, y) = 1$ for $x, y \in B^c_R(x_0)$ and $f_R(x, y) = 0$ if $x \notin B^c_{R+1}(x_0)$ or $y \notin B^c_{R+1}(x_0)$. Let us take measures μ, ν, $\gamma \in \Pi(\mu, \nu)$ and consider $\lambda := f_R \lfloor \gamma$. Define

$$\tilde{\gamma} := \gamma - \lambda + (\pi^y \times \pi^y)_{\#}\lambda$$

and $\tilde{\nu} := (\pi^x)_{\#}\tilde{\gamma}$. So $\tilde{\gamma} \in \Pi(\tilde{\nu}, \nu)$ and $J(\tilde{\nu}, \nu) \le K(\tilde{\gamma}) = K(\gamma) - K(\lambda) = K(\gamma) - K'(\gamma)$ where $K'(\gamma) = K'_R(\gamma) := K(f_R \lfloor \gamma)$.

Now consider a weakly convergent sequence of plans $\Pi(\mu, \nu_n) \ni \gamma_n \rightharpoonup \gamma^* \in \Pi(\mu, \nu^*)$. One has $\tilde{\gamma}_n \rightharpoonup \tilde{\gamma}^*$ hence $\tilde{\nu}_n \rightharpoonup \tilde{\nu}^*$. But on the complement of the ball $B^c_{R+1}(x_0)$ all the measures $\tilde{\nu}$ coincide with μ, consequently, $\tilde{\nu}_n \xrightarrow{J} \tilde{\nu}^*$ by Theorem 2.8.

Theorems 2.7 and 2.8 state some results about the relation between transportation convergence and weak convergence. The following theorem clarifies this relation and gives the necessary and sufficient condition of convergence in the topology τ_J.

Theorem 3.7. (Criterion of the transportation convergence) *Take measures ν^* and $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(X)$. Then the following conditions are equivalent:*

(1) $\nu_n \xrightarrow{J} \nu^*$;

(2) $\nu_n \rightharpoonup \nu^*$ *and* $J(\mu, \nu_n) \to J(\mu, \nu^*)$ *for any* $\mu \in \mathcal{P}(X)$;

(3) $\nu_n \rightharpoonup \nu^*$ *and* $J(\mu, \nu_n) \to J(\mu, \nu^*) < \infty$ *for some* $\mu \in E(\nu^*)$.

Proof. Obviously, (1) $\Rightarrow$ (2) $\Rightarrow$ (3). Let us show that (3) $\Rightarrow$ (1). Without loss of generality assume that $J(\mu, \nu_n) < \infty$ for all n. Let $\gamma_n \in \Pi(\mu, \nu_n)$ be an optimal transport plan. Since the sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ is tight, one can extract a subsequence $\gamma_n \rightharpoonup \hat{\gamma} \in \Pi(\mu, \nu^*)$ (without relabelling). Fix $\epsilon > 0$ and an $R > 0$ such that $K'(\hat{\gamma}) = K'_R(\hat{\gamma}) > K(\hat{\gamma}) - \epsilon$. Using the construction above get transport plans $\Pi(\tilde{\nu}_n, \nu_n) \ni \tilde{\gamma}_n \rightharpoonup \tilde{\gamma} \in \Pi(\tilde{\nu}^*, \nu^*)$. Therefore $J(\tilde{\nu}_n, \tilde{\nu}^*) \to 0$, $J(\tilde{\nu}^*, \nu^*) \le K(\tilde{\gamma}) = K(\hat{\gamma}) - K'(\hat{\gamma}) < \epsilon$ and

$$\begin{aligned} J(\tilde{\nu}_n, \nu_n) &\le K(\tilde{\gamma}_n) = K(\gamma_n) - K'(\gamma_n) = J(\mu, \nu_n) - K'(\gamma_n) \\ &\to J(\mu, \nu^*) - K'(\hat{\gamma}) \le K(\hat{\gamma}) - K'(\hat{\gamma}) < \epsilon. \end{aligned}$$

So, one can construct sequences $\{\tilde{\nu}_n\}_{n\in\mathbb{N}}$, $\{\tilde{\nu}^*_n\}_{n\in\mathbb{N}}$ such that

$$\lim J(\tilde{\nu}_n, \nu_n) = \lim J(\tilde{\nu}_n, \tilde{\nu}^*_n) = \lim J(\tilde{\nu}^*_n, \nu^*) = 0.$$

All the sequences are tight hence $J(\nu_n, \nu^*) \to 0$ by Lemma 3.6. $\square$

Remark 3.8. Obviously, the arguments of $J(\cdot,\cdot)$ can be swapped simultaneously in each of the conditions (2)–(3) without violating the theorem.

Moreover, for the class E_0 there also exists the following dual formulation of transportation convergence which is extremely similar to the notion of weak convergence of measures.

Theorem 3.9. *Take measures $\nu^* \in E_0$ and $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(X)$. Then $\nu_n \xrightarrow{J} \nu^*$ iff $\int f\, d\nu_n \to \int f\, d\nu^*$ for any continuous function $f(\cdot)$ such that $|f(x)| \le \alpha + \beta c(x, x_0)$ for any $x \in X$ and for some constants α, β.*

Proof. (1) Let $\int f\, d\nu_n \to \int f\, d\nu^*$ for any continuous function $f(\cdot)$ such that $|f(x)| \le \alpha + \beta c(x, x_0)$. Then $\nu_n \rightharpoonup \nu^*$ and for any $x_0 \in X$

$$J(\nu_n, \delta_{x_0}) = \int c(x, x_0)\, d\nu_n \to \int c(x, x_0)\, d\nu^* = J(\nu^*, \delta_{x_0}).$$

By Theorem 3.7 that implies $\nu_n \xrightarrow{J} \nu^*$.

(2) Now let $\nu_n \xrightarrow{J} \nu^*$. Then $\nu_n \rightharpoonup \nu^*$ and $\int c(x, x_0)\, d\nu_n \to \int c(x, x_0)\, d\nu^*$. Obviously, it is enough to consider the case when $f(\cdot)$ is nonnegative and $f(x) \le 1 + c(x, x_0)$. Consider functions $f_h(\cdot) := \min\{h, f(\cdot)\} \in C_b(X)$, $h \ge 0$. From weak convergence of ν_n it follows that $\int f_h\, d\nu_n \to \int f_h\, d\nu^*$ for any $h \ge 0$. On the other hand,

$$\begin{aligned}
0 &\le \int (f - f_h)\, d\nu_n = \int (f - h)_+\, d\nu_n \le \int \big(1 + c(x, x_0) - h\big)_+\, d\nu_n \\
&= \int \big(c(x, x_0) - c_{h-1}(x, x_0)\big)\, d\nu_n \\
&\xrightarrow[n\to\infty]{} \int \big(c(x, x_0) - c_{h-1}(x, x_0)\big)\, d\nu^* \to 0 \quad \text{as} \quad h \to +\infty.
\end{aligned}$$

Consequently, $\int f\, d\nu_n \to \int f\, d\nu^*$. □

Corollary 3.10. *Let X be a Polish space. Then the space (E_0, τ_J) is also Polish.*

Proof. Consider the following embedding of E_0 in the space $\mathcal{M}_+(X)$ of all finite nonnegative Borel measures on X: $\nu \mapsto F(\nu) := \big(1 + c(x, x_0)\big)\lfloor\nu$. From Theorem 3.9 it follows that $\nu_n \xrightarrow{J} \nu^*$ iff $F(\nu_n) \rightharpoonup F(\nu^*)$. Moreover, the image $F(E_0)$ is weakly closed in $\mathcal{M}_+(X)$: indeed, if $\mu_n := F(\nu_n) \rightharpoonup \mu$, then

$$\int \frac{1}{1 + c(x, x_0)}\, d\mu = \lim \int \frac{1}{1 + c(x, x_0)}\, d\mu_n = \int d\nu_n = 1,$$

thus $\nu := \frac{1}{1+c(x,x_0)} \lfloor \mu \in E_0$. Space $\mathcal{M}_+(X)$ endowed with the topology of weak convergence is Polish [5, Theorem 8.9.3] and (E_0, τ_J) is isomorphic to its closed subspace, consequently, it also is Polish. $\square$

As we have seen, the function $\nu \mapsto \int f \, d\nu$ for any $f(\cdot) \in \bigl(1 + c(\cdot, x_0)\bigr) C_b(X)$ is continuous w.r.t. the topology τ_J. The next theorem shows that in some cases it is possible to quantify the modulus of continuity for this function.

Theorem 3.11. *Take function $f(\cdot)$ such that $\bigl|f(x) - f(y)\bigr| \le g\bigl(c(x,y)\bigr)$ for all $x, y \in X$ and for some concave function $g(\cdot)$. Then for any $\mu, \nu \in E_0$ the following inequation holds:*

$$\left| \int f \, d\mu - \int f \, d\nu \right| \le g\bigl(J(\mu, \nu)\bigr).$$

Proof. Let γ be an optimal transport plan from μ to ν. Then by Jensen's inequality

$$\begin{aligned} \left| \int f \, d\mu - \int f \, d\nu \right| &= \left| \int f(x) \, d\gamma - \int f(y) \, d\gamma \right| \le \int \bigl|f(x) - f(y)\bigr| \, d\gamma \\ &\le \int g\bigl(c(x,y)\bigr) \, d\gamma \le g\left(\int c \, d\gamma \right) = g\bigl(J(\mu, \nu)\bigr). \end{aligned}$$ $\square$

For example, if $c(x,y) = \rho^p(x,y)$ where $p \ge 1$, and $f(\cdot)$ is Lipschitz continuous with a constant L then $\bigl|\int f \, d\mu - \int f \, d\nu\bigr| \le L J(\mu,\nu)^{1/p}$. Moreover, in this case $J(\mu,\nu)^{1/p}$ itself is a metric on $\mathcal{P}(X)$ (see Section 4), thus the function $\mu \mapsto \int f \, d\mu$ is also L-Lipschitz.

Let us now show that, under Assumption 3.1 and local compactness of X, any class $E(\mu_0)$ endowed with the transportation topology is a Radon space. In order to prove this, we show that $\bigl(E(\mu_0), \tau_J\bigr)$ is separable, metrizable, and that any probability measure on this space is tight.

Lemma 3.12. *Take an arbitrary measure $\mu_0 \in \mathcal{P}(X)$. The equivalence class $E(\mu_0)$ endowed with the topology τ_J is separable and metrizable.*

Proof. (1) Let $\rho_w(\cdot, \cdot)$ be a metric on $\mathcal{P}(X)$ that induces weak convergence. Then

$$\rho_J(\mu, \nu) := \rho_w(\mu, \nu) + \bigl|J(\mu, \mu_0) - J(\nu, \mu_0)\bigr|$$

is also a metric and, obviously, $\mu_n \xrightarrow{J} \mu^* \in E(\mu_0)$ iff $\rho_J(\mu_n, \mu^*) \to 0$ by Theorem 3.7.

(2) Let $\mathcal{S}_{\mu_0}$ be a countable family of measures of type $\nu := \bigl(X \setminus B_m(x_0)\bigr) \lfloor \mu_0 + \alpha \sum_{i=1}^n p_i \delta_{x_i}$ where $m, n \in \mathbb{N}$, $p_i \in \mathbb{Q}_+$, x_i belong to some countable dense subset of X and α is a normalizing constant. Fix measure $\mu \in C(\mu_0)$, $\epsilon > 0$

and such $R > 0$ that $\int_{B_R(x_0)\times B_R(x_0)} c(x,y)\,d\gamma > K(\gamma)-\epsilon$, where γ is an optimal transport plan from μ_0 to μ. Thus $K'(\gamma) := K(f_R\lfloor\gamma) > K(\gamma)-\epsilon$ and $J(\tilde\mu,\mu) \le K(\gamma) - K'(\gamma) < \epsilon$. But $\tilde\mu$ obviously lie in the weak closure of $\mathcal{S}_{\mu_0}$, so there exists a sequence $\mathcal{S}_{\mu_0} \ni \nu_n \xrightarrow{J} \tilde\mu$ and $\lim J(\nu_n,\mu) = J(\tilde\mu,\nu) < \epsilon$. Consequently, $\mathcal{S}_{\mu_0}$ is a dense set in $E(\mu_0)$. $\square$

The following assumption about local compactness of the space X allows us to obtain weak local compactness of $\mathcal{P}(X)$.

Assumption 3.13. For some (and therefore for any) $x_0 \in X$ any "closed ball" $\overline{B}^c_r(x_0) := \left\{y \in X : c(x_0,y) \le r\right\}$ is compact.

Notice that under this assumption from $c(x,y) = 0$ iff $x = y$ it follows that $c(x,x_n) \to 0$ iff $c(x_n,x) \to 0$ iff $x_n \to x$. Moreover, X is locally-compact and therefore separability of the space implies that it is a Radon space without any additional assumption.

Lemma 3.14. *Let a sequence $\{\nu_n\}_{n\in\mathbb{N}}$ be such that $\limsup J(\mu,\nu_n) < \infty$ for some $\mu \in \mathcal{P}(X)$; then the sequence is tight.*

Proof. Fix $\epsilon > 0$ and a ball $\overline{B}^c_r(x_0)$ such that $\mu\left(X \setminus \overline{B}^c_r(x_0)\right) < \epsilon/2$. Consider $R > r$ and any measure λ_R such that $\lambda_R\big(X \setminus B^c_R(x_0)\big) \ge \epsilon$. Then

$$J(\mu,\lambda_R) \ge \frac{\epsilon}{2}\inf\left\{c(x,y) : x \in \overline{B}^c_r(x_0),\, y \notin B^c_R(x_0)\right\} \to \infty \text{ as } R \to \infty.$$

But $\limsup J(\mu,\nu_n) < \infty$ and therefore one can find such a compact ball $\overline{B}^c_{R_\epsilon}(x_0)$ that $\nu_n\left(X \setminus \overline{B}^c_{R_\epsilon}(x_0)\right) < \epsilon$ for all n, i.e. $\{\nu_n\}_{n\in\mathbb{N}}$ is a tight sequence. $\square$

Corollary 3.15. *Any "closed ball" $\overline{B}^J_r(\mu) := \left\{\nu \in \mathcal{P}(X) : J(\mu,\nu) \le r\right\}$ is weakly compact.*

Proof. Consider any sequence $\{\nu_n\}_{n\in\mathbb{N}} \subset \overline{B}^J_r(\mu)$. It is tight by the Lemma 3.14 and hence there is weakly convergent subsequence $\nu_{n_k} \rightharpoonup \nu^*$. By lemma 2.5 $J(\mu,\nu^*) \le \liminf J(\mu,\nu_{n_k}) \le r$ thus $\nu^* \in \overline{B}^J_r(\mu)$. $\square$

Now let us show that space $\mathcal{P}(X)$ is "complete" with respect to the Monge–Kantorovich distance.

Lemma 3.16. *Let $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ be a sequence such that $J(\nu_n,\nu_m) \to 0$ as $n \ge m \to \infty$; then $\nu_n \xrightarrow{J} \nu^*$ for some $\nu^* \in \mathcal{P}(X)$.*

Proof. Since $\{\nu_n\}_{n\in\mathbb{N}}$ is a tight sequence by Lemma 3.14 there exists a subsequence $\nu_{n_k} \rightharpoonup \nu^* \in \mathcal{P}(X)$. Assume ν_{n_k} fails to converge to ν^*; then one can choose such $\epsilon > 0$, $N \in \mathbb{N}$ that $J(\nu^*,\nu_N) > \epsilon$ and $J(\nu_n,\nu_N) < \epsilon$ for $n \ge N$.

But from weak lower semicontinuity of $J(\cdot,\cdot)$ it holds $\epsilon \geq \liminf J(\nu_{n_k}, \nu_N) \geq J(\nu^*, \nu_N) > \epsilon$ what leads to a contradiction. Thereby $\nu_n \xrightarrow{J} \nu^*$. $\square$

Now it is enough to show that any Borel probability measure over $E(\mu_0)$ is tight in order to prove that $\big(E(\mu_0), \tau_J\big)$ is a Radon space. Let us adapt to our case the proof for Polish space from [5, Theorem 7.1.7].

Theorem 3.17. *For any $\mu_0 \in \mathcal{P}(X)$ the class $E(\mu_0)$ endowed with the transportation topology τ_J is a Radon space.*

Lemma 3.18. (Compactness in τ_J) *Let a closed set $\mathcal{H} \subset E(\mu_0)$ be totally bounded with respect to the transportation cost $J(\cdot,\cdot)$, i.e. for any $\epsilon > 0$ let there exist measures $\mu_1^\epsilon, \dots, \mu_n^\epsilon$ such that $\mathcal{H} \subset \bigcup_{i=1}^n B_\epsilon^J(\mu_i^\epsilon)$; then $\mathcal{H}$ is compact in the transportation topology.*

Proof. Consider an arbitrary sequence $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$. Using Cantor's diagonal argument and conditions of the lemma one can find a sequence of measures $\{\mu_n\}_{n\in\mathbb{N}}$ and a subsequence of $\{\nu_n\}_{n\in\mathbb{N}}$ such that (without relabelling) $J(\mu_n, \nu_k) < 1/n$ for all $k \geq n$. Then $\{\nu_n\}_{n\in\mathbb{N}}$ and $\{\mu_n\}_{n\in\mathbb{N}}$ are tight by Lemma 3.14 and therefore $J(\nu_n, \nu_m) \to 0$ as $m, n \to \infty$ by Lemma 3.6. Now it follows from Lemma 3.16 that there exists $\nu^* \in \mathcal{H}$ such that $\nu_n \xrightarrow{J} \nu^*$. $\square$

Now one can prove Theorem 3.17 in the same way as [5, Theorem 7.1.7]. The only necessary change regards the criterion of compactness in τ_J, which is considered in Lemma 3.18.

4. The case of $\mathbb{R}^d$

Now consider the locally compact Polish space $X = \mathbb{R}^d$ with Euclidean metric and take $c(x, y) = g(x - y)$, where the function $g(\cdot)$ is convex, $g(0) = 0$ and $g(x) > 0$ whenever $x \neq 0$. Obviously, this cost function is continuous and *consistent* in the sense that $c(x, x_n) \to 0$ iff $c(x_n, x) \to 0$ iff $x_n \to x$. Moreover, the space and the cost function satisfy Assumption 3.13.

Let us assume the following inequality holds:

$$B := \sup_{x,y} \frac{g(x+y)}{g(x) + g(y)} < \infty. \tag{2}$$

Notice that $B \geq 1$ due to convexity of the function $g(\cdot)$.

Theorem 4.1. *Let inequality* (2) *hold. Then there exists $q \geq 1$ such that $g^{1/q}(\cdot)$ satisfies the triangle inequality:*

$$g^{1/q}(x+y) \leq g^{1/q}(x) + g^{1/q}(y)$$

for any $x, y \in \mathbb{R}^d$.

Proof. Consider points $x, y \in \mathbb{R}^d$ such that $g(y) = \xi g(x)$, $\xi \le 1$. Due to convexity of $g(\cdot)$ one can obtain that for any $n \ge 1$ it holds

$$g(x+y) \le \frac{n-1}{n} g(x) + \frac{1}{n} g(x+ny) \le g(x) + \frac{B}{n}\bigl(g(x) + g(ny)\bigr).$$

Consider $n = 2^k$; it follows from inequality 2 that $g(2^k y) \le (2B)^k g(y)$ and therefore

$$g(x+y) \le g(x) + 2^{-k} B g(x) + B^{k+1} g(y) = g(x)\left(1 + B\left(2^{-k} + \xi B^k\right)\right).$$

Take $k := \left\lfloor -\frac{\ln \xi}{\ln 2B} \right\rfloor$; then

$$2^{-k} + \xi B^k \le 2^{\frac{\ln \xi}{\ln 2 + \ln B} + 1} + \xi B^{-\frac{\ln \xi}{\ln 2 + \ln B}} = 3\xi^{1/q_0},$$

where $q_0 := \frac{\ln 2B}{\ln 2} \ge 1$. Thus $g(x+y) \le g(x)\left(1 + 3B\xi^{1/q_0}\right)$. Since $\xi \le 1$ one can obtain that for $q = \max\{3B, q_0\}$ it holds

$$\begin{aligned} g^{1/q}(x+y) &\le g^{1/q}(x)\left(1 + 3B\xi^{1/q_0}\right)^{1/q} \le g^{1/q}(x)\left(1 + \frac{3B}{q}\xi^{1/q_0}\right) \\ &\le g^{1/q}(x)\left(1 + \xi^{1/q}\right) = g^{1/q}(x) + g^{1/q}(y) \end{aligned}$$

for all $x, y \in \mathbb{R}^d$. □

Corollary 4.2. *For all $\mu, \nu, \lambda \in \mathcal{P}(\mathbb{R}^d)$: $J^{1/q}(\mu, \nu) \le J^{1/q}(\mu, \lambda) + J^{1/q}(\lambda, \nu)$.*

Proof.
Let us take measures $\mu, \nu, \lambda \in \mathcal{P}(\mathbb{R}^d)$ and optimal transport plans $\gamma_1 \in \Pi(\mu, \lambda)$, $\gamma_2 \in \Pi(\lambda, \nu)$. Similarly to the proof of Lemma 3.2 consider a measure $\sigma \in \Pi(\mu, \lambda, \nu)$ such that $\pi^{x,y}_{\#}\sigma = \gamma_1$, $\pi^{y,z}_{\#}\sigma = \gamma_2$. Applying Theorem 4.1 and the Minkowski inequality one obtains

$$\begin{aligned} J^{1/q}(\mu, \nu) &\le K^{1/q}(\pi^{x,z}_{\#}\sigma) = \left(\int \left(g^{1/q}(x-z)\right)^q d\sigma\right)^{1/q} \\ &\le \left(\int \left(g^{1/q}(x-y) + g^{1/q}(y-z)\right)^q d\sigma\right)^{1/q} \\ &\le \left(\int g(x-y)\, d\sigma\right)^{1/q} + \left(\int g(y-z)\, d\sigma\right)^{1/q} \\ &= K^{1/q}(\gamma_1) + K^{1/q}(\gamma_2) = J^{1/q}(\mu, \lambda) + J^{1/q}(\lambda, \nu). \end{aligned}$$ □

Corollary 4.3. *The function $\rho(\mu, \nu) := \max\left\{J^{1/q}(\mu, \nu), J^{1/q}(\nu, \mu)\right\} \in [0, \infty]$ is a metric on $\mathcal{P}(\mathbb{R}^d)$ (which may make the value $+\infty$).*

As we have seen, under assumption (2) $\big(\mathcal{P}(X), J\big)$ is similar to a q-Wasserstein space for some degree q.

Now consider Assumption 3.1 in the Euclidean case. Obviously, one can rewrite it as

$$g(\pm x \pm y) \le A + B\big(g(x) + g(y)\big).$$

Theorem 4.4. *Under Assumption* 3.1 *for any* $\epsilon > 0$ *there exist* $q_\epsilon \ge 1$, $C_\epsilon > 0$ *such that for all* $x, y \in \mathbb{R}^d$ *it holds*

$$\big(g(\pm x \pm y) + \epsilon\big)^{1/q_\epsilon} \le \big(g(x) + \epsilon\big)^{1/q_\epsilon} + \big(g(y) + \epsilon\big)^{1/q_\epsilon},$$
$$g(x \pm y) \le \epsilon + (1+\epsilon) g(x) + C_\epsilon g(y).$$

Proof. Fix $\epsilon > 0$. Consider $r > 0$ such that $g(x) \le \epsilon$, $\|x\| \le r$. Since $g(x) = 0$ iff $x = 0$ define $a_r := \inf_{\|x\| \ge r} g(x) > 0$. If $\|x + y\| > r$ then $\|x\| > r/2$ or $\|y\| > r/2$ therefore

$$\begin{aligned} g(x+y) &\le A + B\big(g(x) + g(y)\big) \le \frac{A}{a_{r/2}}\big(g(x) + g(y)\big) + B\big(g(x) + g(y)\big) \\ &= \left(\frac{A}{a_{r/2}} + B\right)\big(g(x) + g(y)\big). \end{aligned}$$

Consequently, there exists $D_r = \frac{A}{a_{r/2}} + B \ge 1$ such that

$$g(\pm x \pm y) \le \max\big\{\epsilon, D_r\big(g(x) + g(y)\big)\big\}$$

for all $x, y \in \mathbb{R}^d$. In particular,

$$g(\pm x \pm y) + \epsilon \le D_r\big(g(x) + \epsilon + g(y) + \epsilon\big)$$

and one can prove the first inequality in the same way as in Theorem 4.1.

In order to prove the second inequality, let us choose $k \in \mathbb{N}$, $r > 0$ such that

$$2^{-k} B < \epsilon, \quad 2^{-k} A < \frac{\epsilon}{2}, \quad \text{and} \quad 2^{-k} B g(x) < \frac{\epsilon}{2} \text{ as } \|x\| \le 2^k r.$$

Then similarly to the proof of Theorem 4.1 one can show that

$$\begin{aligned} g(x+y) &\le \left(1 + 2^{-k} B\right) g(x) + 2^{-k} B g\left(2^k y\right) + 2^{-k} A \\ &\le (1+\epsilon) g(x) + \frac{\epsilon}{2} + \max\left\{\frac{\epsilon}{2}, D_r^k B g(y)\right\} \le \epsilon + (1+\epsilon) g(x) + C_\epsilon g(y), \end{aligned}$$

where $C_\epsilon := D_r^k B$. □

Corollary 4.5. *For any* $\epsilon > 0$ *and measures* $\mu, \nu, \lambda \in \mathcal{P}(\mathbb{R}^d)$ *the following inequalities hold:*

$$\big(J(\mu,\nu) + \epsilon\big)^{1/q_\epsilon} \le \big(J(\mu,\lambda) + \epsilon\big)^{1/q_\epsilon} + \big(J(\lambda,\nu) + \epsilon\big)^{1/q_\epsilon},$$
$$J(\mu,\nu) \le \epsilon + (1+\epsilon) J(\mu,\lambda) + C_\epsilon J(\lambda,\nu),$$
$$J(\mu,\nu) \le \epsilon + (1+\epsilon) J(\lambda,\nu) + C_\epsilon J(\mu,\lambda).$$

The proof of Corollary 4.5 is completely similar to the proofs of Lemma 3.2 and Corollary 4.2.

5. Fréchet barycenters

As we have obtained in Sec. 3, the space of probability measures endowed with the transportation topology has some good properties. In this section the *barycenter* of measures will be defined, i.e. some kind of averaging w.r.t. the transportation structure of the space. It generalizes the construction from [1], where the 2-Wasserstein space is considered. Other related works on barycenters are [8], where p-Wasserstein barycenters are introduced and studied, and [6] devoted to barycenters of measures on Riemannian manifolds. Regularized barycenters are also considered in the recent work [3].

In the section, the Fréchet barycenter will be shown to be "upper semicontinuous" in some sense and statistically consistent. Analogous results for measures over $\mathbb{R}$ and a convex cost function were proved in [7].

5.1. Generalized averaging in $\mathcal{P}(X)$

Let the space $\mathcal{P}(X)$ be endowed with the Borel σ-algebra $\mathcal{B}(\tau_w)$ induced by the topology of weak convergence τ_w. This σ-algebra is weaker than $\mathcal{B}(\tau_J)$, induced by the transportation topology. However, as we will see later, they are equivalent for defining an averaging in $\mathcal{P}(X)$.

Definition 5.1. Take the functional $G \colon \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ (regularizer), consider a finite set $\mu_1, \mu_2, \dots, \mu_n$ of measures in $\mathcal{P}(X)$ and positive *weights* $\lambda_1 > 0$, $\lambda_2 > 0$, …, $\lambda_n > 0$. The *G-regularized Fréchet barycenter* $\mathrm{bar}_G(\mu_i, \lambda_i)_{1 \le i \le n} \in \mathcal{P}(X)$ (or just the Fréchet barycenter if $G \equiv$ *const*) with respect to the transportation functional $J(\cdot, \cdot)$ is a measure that minimizes

$$\sum_{1 \le i \le n} \lambda_i J(\mu_i, \nu) + G(\nu) \tag{3}$$

over $\nu \in \mathcal{P}(X)$. By $\mathrm{Bar}_G(\mu_i, \lambda_i)_{1 \le i \le n} \in \mathcal{P}(X)$ we denote the set of all Fréchet barycenters of $(\mu_i, \lambda_i)_{1 \le i \le n}$.

Definition 5.2. Take the functional $G \colon \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ and the distribution $P \in \mathcal{P}\big(\mathcal{P}(X)\big)$. Consider the problem of minimizing

$$\mathbb{E}_{\boldsymbol{\mu} \sim P}\, J(\boldsymbol{\mu}, \nu) + G(\nu) \to \min_{\nu \in \mathcal{P}(X)}$$

and denote its solution by $\mathrm{bar}_G(P)$. We call the measure $\mathrm{bar}_G(P)$ the *G-regularized Fréchet barycenter* of the distribution P. Respectively, $\mathrm{Bar}_G(P)$ is the set of all G-regularized Fréchet barycenters of P.

Obviously, Definition 5.1 is a particular case of Definition 5.2 hence one can consider distribution $P_n := \sum_{i=1}^n \lambda_i \delta_{\mu_i}$ such that $\mathrm{Bar}_G(P_n) = \mathrm{Bar}_G(\mu_i, \lambda_i)_{1 \le i \le n}$. Under Assumptions 3.1 and 3.13 the G-regularized Fréchet barycenter with lower semicontinuous and bounded below regularizer G always exists.

Theorem 5.3. *Let $G\colon \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be bounded below and lower semi-continuous with respect to weak convergence, and distribution $P \in \mathcal{P}\big(\mathcal{P}(X)\big)$ be such that*

$$\inf_{\nu \in \mathcal{P}(X)} \left(\int J(\mu, \nu)\, dP(\mu) + G(\nu) \right) < \infty;$$

then there exists the G-regularized Fréchet barycenter of P. Moreover, any minimizing sequence $\{\nu_n\}_{n \in \mathbb{N}}$, i.e. such that

$$\int J(\mu, \nu_n)\, dP(\mu) + G(\nu_n) \to \inf_{\nu \in \mathcal{P}(X)} \left(\int J(\mu, \nu)\, dP(\mu) + G(\nu) \right),$$

is precompact in the topology τ_J and every its partial limit is the G-regularized barycenter of the distribution. In particular, $\mathrm{Bar}_G(P)$ is compact.

Proof. Fix $\epsilon > 0$ and a ball $B = \overline{B}_r^c(x_0)$ such that $\mu(X \setminus B) < \epsilon/2$ for all measures from some set $\mathcal{U} \subset \mathcal{P}(X)$, $P(\mathcal{U}) > 1/2$. Consider any $R > r$ and measure λ_R such that $\lambda_R\big(X \setminus B_R^c(x_0)\big) \ge \epsilon$. One can obtain that

$$\int J(\mu, \lambda_R)\, dP(\mu) \ge \frac{1}{2}\frac{\epsilon}{2} \inf\big\{c(x,y) : x \in B,\, y \notin B_R^c(x_0)\big\} \to \infty \text{ as } R \to \infty.$$

Let m be a lower bound for $G(\cdot)$, then

$$\begin{aligned}\limsup \int J(\mu, \nu_n)\, dP(\mu) &\le \limsup \left(\int J(\mu, \nu_n)\, dP(\mu) + G(\nu_n) - m \right) \\ &= \inf_{\nu \in \mathcal{P}(X)} \left(\int J(\mu, \nu_n)\, dP(\mu) + G(\nu_n) \right) - m < +\infty,\end{aligned}$$

consequently, $\{\nu_n\}_{n \in \mathbb{N}}$ is a tight sequence (similarly to Lemma 3.14) and there exists weakly convergent subsequence $\nu_{n_k} \rightharpoonup \nu^*$.

By Fatou's lemma and lower semicontinuity of $J(\cdot,\cdot)$ and $G(\cdot)$

$$\int J(\mu, \nu^*)\, dP(\mu) + G(\nu^*) \le \int \liminf J(\mu, \nu_{n_k})\, dP(\mu) + \liminf G(\nu_{n_k}) \le$$
$$\le \liminf \left(\int J(\mu, \nu_{n_k})\, dP(\mu) + G(\nu_{n_k}) \right) = \inf_{\nu \in \mathcal{P}(X)} \left(\int J(\mu, \nu)\, dP(\mu) + G(\nu) \right).$$

Thus $\nu^* \in \mathrm{Bar}_G(P)$. Moreover, $J(\mu, \nu^*) = \liminf J(\mu, \nu_{n_k})$ for P-a.e. μ, so by Theorem 3.7 there is a subsequence $\nu_{n_k} \xrightarrow{J} \nu^*$ (without relabelling). $\square$

In particular, notice that $\mathrm{bar}_G(\mu_i, \lambda_i)_{1 \le i \le n} \in \mathcal{P}(X)$ exists iff all the μ_i belong to the same equivalence class and $G(\cdot) \not\equiv +\infty$ on this class. For $\mathrm{bar}_G(P)$ to exist it is necessary but not sufficient that $\operatorname{supp} P \subset E(\mu)$ for some μ, i.e. $P\big(E(\mu)\big) = 1$. Notice that since $E(\mu)$ and every ball $B_r^J(\nu)$ are measurable w.r.t. $\mathcal{B}(\tau_w)$, the restriction of $\mathcal{B}(\tau_w)$ to $E(\mu)$ coincides with $\mathcal{B}(\tau_J)$, as the restriction of τ_J on $E(\mu)$ has a countable basis of balls. Therefore, it is enough to consider the space $\mathcal{P}(X)$ endowed with $\mathcal{B}(\tau_w)$ instead of stronger σ-algebra $\mathcal{B}(\tau_J)$.

As example of bounded below and lower semicontinuous regularizer one can consider a characteristic function of some weakly closed subset $\mathcal{G} \subset \mathcal{P}(X)$, or entropy-type functionals: $G(\nu) = \int g\left(\frac{d\nu}{d\nu_0}\right) d\nu_0$, where $\nu \ll \nu_0$ and $g(\cdot)$ is a convex nonnegative function. See [3] for more details on regularized barycenters in the 2-Wasserstein space.

Consider the case when $G(\cdot)$ is convex. Due to convexity of $J(\cdot,\cdot)$ by Lemma 2.3, $\mathrm{Bar}_G(P)$ also is a convex set. Moreover, if $X = \mathbb{R}^d$ and $c(x,y) = g(x-y)$, where $g(\cdot)$ is *strictly* convex, then $J(\mu,\cdot)$ is also strictly convex, whenever μ is absolutely continuous w.r.t. the Lebesgue measure $\mathcal{L}$. It follows from the fact that in this case for any $\nu \sim \mu$ there exists a unique optimal transport plan from μ to ν of form $\gamma = (id \times T)_{\#}\mu$, where T is an optimal *transport map* (see [9], Section 1.3). Therefore, there exists a *unique* barycenter of P, whenever $P\big(\{\mu : \mu \ll \mathcal{L}\}\big) > 0$. However, even without any assumption about measures one can take a strictly convex regularizer and ensure the uniqueness of the barycenter.

5.2. Consistency of barycenters

Let us consider distributions on $E_0 = E(\delta_{x_0})$ where x_0 is some fixed point in X. One can define the Monge–Kantorovich distance between them with $J(\cdot,\cdot)$ as a cost function:

$$\mathcal{J}(P, P') := \inf_{F \in \Pi(P,P')} \int J(\mu,\nu)\, dF(\mu,\nu), \;\; P, P' \in \mathcal{P}(E_0).$$

As E_0 endowed with the topology τ_J is a Polish space, and $J(\cdot,\cdot)$ as a cost function satisfies Assumption 3.1 and other conditions, all the results from Section 3 hold except for those which need Assumption 3.13.

Now let us show that convergence of distributions with respect to $\mathcal{J}(\cdot,\cdot)$ implies the transportation convergence of its barycenters. This result is similar to [8, Theorem 2] in case of the p-Wasserstein spaces. Also, we will obtain the law of large numbers for empirical barycenters proved in [4, Theorem 6.1] for the 2-Wasserstein space and measures with compact support.

Firstly, one can show that for measures from E_0 the following analogue of Prokhorov's theorem holds.

Lemma 5.4. *A family of measures $\mathcal{H} \subset E_0$ is precompact in the transportation topology iff for any $\epsilon > 0$ there exists a compact set $D_\epsilon \subset X$ such that $\int_{X\setminus D_\epsilon} c(x, x_0)\, d\nu \le \epsilon$ for all $\nu \in \mathcal{H}$.*

Proof. Let the condition of the lemma holds. Consider the image of $\mathcal{H}$ under the isomorphism $F(\nu) := \big(1 + c(x, x_0)\big)\lfloor\nu$ introduced in the proof of Corollary 3.10. The set $F(\mathcal{H})$ is tight and uniformly bounded in $\mathcal{M}_+(X)$; indeed, for any $\nu \in \mathcal{H}$ one has

$$\begin{aligned}\int d\big(F(\nu)\big) &:= \int \big(1 + c(x, x_0)\big)\, d\nu = 1 + \int\limits_{X\setminus D_1} c(x, x_0)\, d\nu + \int\limits_{D_1} c(x, x_0)\, d\nu \\ &\le 2 + \max_{x\in D_1} c(x, x_0) < \infty.\end{aligned}$$

Therefore, $F(\mathcal{H})$ is weakly precompact by the variant of Prokhorov's theorem [5, Theorem 8.6.7]. Since $F(E_0)$ is weakly closed in $\mathcal{M}_+(X)$, it follows that $\mathcal{H}$ is precompact in E_0 with transportation topology by Theorem 3.9.

On the other hand, if $\mathcal{H}$ is precompact then $F(\mathcal{H})$ is weakly precompact. Hence $F(\mathcal{H})$ is tight what implies the statement of the lemma. □

Theorem 5.5. *Take a regularizer G which is bounded below and lower semocontinuous with respect to τ_w. Let the sequence $\{P_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(E_0)$ be such that $P_n \xrightarrow{\mathcal{J}} P$ for some distribution P and let there exist the G-regularized Fréchet barycenter of P. Then there exist barycenters $\nu_n \in \mathrm{Bar}_G(P_n)$ beginning from some n_0, the sequence $\{\nu_n\}_{n\ge n_0}$ is precompact and every its partial limit is the barycenter of P. In particular, if $\nu^* := \mathrm{bar}_G(P)$ is unique, then $\nu_n \xrightarrow{J} \nu^*$.*

Remark 5.6. One can rewrite the statement of the theorem as follows: let the distribution P have a G-regularized Fréchet barycenter; then for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\mathrm{Bar}_G(P') \subset U_\epsilon\big(\mathrm{Bar}_G(P)\big) \text{ as } \mathcal{J}(P, P') < \delta,$$

where $U_\delta\big(\mathrm{Bar}_G(P)\big) := \bigcup_{\mu\in\mathrm{Bar}_G(P)} B^J_\epsilon(\mu)$ is an open neighbourhood of $\mathrm{Bar}_G(P)$. One can say that set-valued map $P \mapsto \mathrm{Bar}_G(P)$ is upper-semicontinuous with respect to Hausdorf-like distance.

Proof. Obviously, for any $\mu_0 \in E_0$ it holds $\limsup J(\nu_n, \mu_0) < \infty$ therefore $\{\nu_n\}_{n\in\mathbb{N}}$ is tight. Let $\nu_n \rightharpoonup \nu^*$ without relabelling. Assume that there is no subsequence convergent to ν^* in τ_J. Then by Lemma 5.4 one can assume without loss of generality that for some $\epsilon_0 > 0$ and any $R > 0$

$$\liminf \int_{X\setminus B_R} c(x, x_0)\, d\nu_n > \epsilon_0 B.$$

For any $R > 0$ fix continuous function $f_R \colon X \to [0,1]$ such that $f_R(x) = 1$, $x \in B_R$ and $f_R(x) = 0$, $x \notin B_{R+1}$. For given R and measure ν let us define measure $\tilde{\nu} = \tilde{\nu}_R := f_R \lfloor \nu + \big(1 - (f_R \lfloor \nu)(X)\big)\delta_{x_0}$ similarly to one in Section 3. Notice that $\tilde{\nu}_n \xrightarrow{J} \tilde{\nu}^*$ for any R and $\tilde{\nu}^* \xrightarrow{J} \nu^*$ as $R \to \infty$.

Consider an arbitrary measure $\mu \in E_0$. Due to Lemma 5.4,

$$\sup_{\mu' \in B^J_\delta(\mu)} \int_{X \setminus B_r} c(x, x_0)\, d\mu' \to 0, \text{ as } r \to \infty,\ \delta \to 0.$$

Now one can choose $r, R, \delta > 0$ and $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ and for any $\mu' \in B^J_\delta(\mu)$ the following inequalities hold:

$$J(\mu', \nu^*) < J(\mu', \tilde{\nu}_n) + \frac{\epsilon_0}{4}, \quad \int_{X \setminus B_{R+1}} c(y, x_0)\, d\nu_n(y) > \epsilon_0 B,$$

$$\int_{X \setminus B_r} c(x, x_0)\, d\mu'(x) < \frac{\epsilon_0}{8}, \quad r\nu_n\big(X \setminus B_R\big) < \frac{\epsilon_0}{8}, \quad \text{and} \quad \nu_n\big(X \setminus B_{R+1}\big) < \epsilon_0 \frac{B}{8A}.$$

Let us denote by $\gamma^{\nu_n}_{\mu'}$ an optimal transport plan from μ' to ν_n. Then

$$J(\mu', \tilde{\nu}_n) \leq \int_{X \times B_{R+1}} c\, d\gamma^{\nu_n}_{\mu'} + \int_{X \times (X \setminus B_R)} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'}.$$

One can obtain that

$$\begin{aligned} \int_{X \times (X \setminus B_R)} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} &\leq \int_{B_r \times (X \setminus B_R)} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} + \int_{(X \setminus B_r) \times X} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} \\ &\leq r\nu_n(X \setminus B_R) + \int_{X \setminus B_r} c(x, x_0)\, d\mu' \leq \frac{\epsilon_0}{4}, \end{aligned}$$

hence $\quad J(\mu', \tilde{\nu}_n) \leq \displaystyle\int_{X \times B_{R+1}} c(x, y)\, d\gamma^{\nu_n}_{\mu'} + \frac{\epsilon_0}{4}.$ On the other hand,

$$\begin{aligned} J(\mu', \nu_n) &= \int_{X \times B_{R+1}} c(x, y)\, d\gamma^{\nu_n}_{\mu'} + \int_{X \times (X \setminus B_{R+1})} c(x, y)\, d\gamma^{\nu_n}_{\mu'} \geq \\ &\geq \int_{X \times B_{R+1}} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} + \int_{X \times (X \setminus B_{R+1})} \big(\frac{1}{B} c(y, x_0) - c(x, x_0) - \frac{A}{B}\big)\, d\gamma^{\nu_n}_{\mu'} \\ &= \int_{X \times B_{R+1}} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} + \frac{1}{B} \int_{X \setminus B_{R+1}} c(y, x_0)\, d\nu_n - \\ &\qquad - \int_{X \times (X \setminus B_{R+1})} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} - \frac{A}{B} \nu_n(X \setminus B_{R+1}) \\ &\geq \int_{X \times B_{R+1}} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} + \epsilon_0 - \frac{\epsilon_0}{4} - \frac{\epsilon_0}{8} = \int_{X \times B_{R+1}} c(x, x_0)\, d\gamma^{\nu_n}_{\mu'} + \frac{5\epsilon_0}{8}. \end{aligned}$$

Therefore $J(\mu', \tilde{\nu}_n) \le J(\mu', \nu_n) - 3\epsilon_0/8$ and

$$J(\mu', \nu^*) < J(\mu', \tilde{\nu}_n) + \frac{\epsilon_0}{4} \le J(\mu', \nu_n) - \frac{\epsilon_0}{8}$$

for all $\mu' \in B^J_\delta(\mu)$. Thus for any $k \in \mathbb{N}$ there exists an open set $U_k \subset B^J_k(\nu^*)$ such that for all $\mu \in U_k$

$$J(\mu, \nu_n) \ge J(\mu, \nu^*) + \frac{\epsilon_0}{8}, \; n \ge k,$$

and $\bigcup_{k=1}^\infty U_k = E_0$. Now one can obtain that for any k

$$\begin{aligned}\liminf \left(\int J(\mu, \nu_n)\, dP_n + G(\nu_n) \right) &\ge \liminf \left(\int_{U_k} J(\mu, \nu_n)\, dP_n + G(\nu_n) \right) \\ &\ge \liminf \left(\int_{U_k} J(\mu, \nu^*)\, dP_n + \frac{\epsilon_0}{8} P_n(U_k) + G(\nu_n) \right) \\ &\ge \int_{U_k} J(\mu, \nu^*)\, dP + \frac{\epsilon_0}{8} P(U_k) + G(\nu^*).\end{aligned}$$

But since $\bigcup_{k=1}^\infty U_k = E_0$

$$\liminf \left(\int J(\mu, \nu_n)\, dP_n + G(\nu_n) \right) \ge \int J(\mu, \nu^*)\, dP + G(\nu^*) + \frac{\epsilon_0}{8},$$

what contradicts the fact that

$$\int J(\mu, \nu_n)\, dP_n + G(\nu_n) \le \int J(\mu, \nu^*)\, dP_n + G(\nu^*) \to \int J(\mu, \nu^*)\, dP + G(\nu^*).$$

Consequently, there exists a convergent subsequence of barycenters $\nu_{n_k} \xrightarrow{J} \nu^*$. Let us show that ν^* is the G-regularized barycenter of P. Indeed, consider any $\nu \in \mathcal{P}(X)$. Continuity of the transportation distance implies

$$\begin{aligned}\int J(\mu, \nu^*)\, dP + G(\nu^*) &\le \lim \int J(\mu, \nu_{n_k})\, dP_{n_k} + \liminf G(\nu_{n_k}) \\ &\le \liminf \left(\int J(\mu, \nu)\, dP_{n_k} + G(\nu) \right) = \int J(\mu, \nu)\, dP + G(\nu),\end{aligned}$$

thus, $\nu^* \in \mathrm{Bar}_G(P)$. □

Remark 5.7. Although set-valued map $P \mapsto \mathrm{Bar}_G(P)$ is in some sense "upper semicontinuous", in general case there no exists a *continuous* function $P \mapsto \mathrm{bar}_G(P)$, even for $G(\cdot) \equiv$ *const*. However, if $G(\cdot)$ is strictly convex then $P \mapsto \mathrm{bar}_G(P)$ is actually continuous.

Corollary 5.8. (Upper semicontinuity of empirical barycenters) *Take the sequences of measures* $\{\mu_i^n\}_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ *and weights* $\{\alpha_i^n\}_{n\in\mathbb{N}} \subset [0,+\infty)$ *such that* $\mu_i^n \xrightarrow{J} \mu_i$, $\alpha_i^n \to \alpha_i$ *for* $1 \le i \le m$. *Then, if all* $\mu_i \in E_0$, *the sequence of Fréchet barycenters* $\{\mathrm{bar}_G(\mu_i^n, \alpha_i^n)_{1\le i\le m}\}_{n\in\mathbb{N}}$ *is precompact and every its partial limit belongs to* $\mathrm{Bar}_G(\mu_i, \alpha_i)_{1\le i\le m}$.

Proof. Let us consider $P_n := \sum_i \alpha_i^n \delta_{\mu_i^n}$, $P := \sum_i \alpha_i \delta_{\mu_i}$, so $\mathrm{Bar}_G(\mu_i^n, \alpha_i^n)_{1\le i\le m} = \mathrm{Bar}_G(P_n)$ and $\mathrm{Bar}_G(\mu_i, \alpha_i)_{1\le i\le m} = \mathrm{Bar}_G(P)$. Notice that $J(\mu_i^n, \mu_j) \to J(\mu_i, \mu_j)$ and $\max_{i,j} J(\mu_i, \mu_j) < \infty$ hence

$$\mathcal{J}(P_n, P) \le \sum_{i=1}^m \min\{\alpha_i^n, \alpha_i\}\, J(\mu_i^n, \mu_i) + \max_{i,j} J(\mu_i^n, \mu_j) \sum_{i=1}^m |\alpha_i^n - \alpha_i| \to 0.$$

This shows that the conditions of Theorem 5.5 hold. □

Corollary 5.9. (Law of large numbers) *Let* $\{\boldsymbol{\mu}_n\}_{n\in\mathbb{N}} \subset E_0$ *be a sequence of i.i.d. random elements with distribution* P_μ *such that there exists* $\mathrm{bar}_G(P_\mu)$, *and* $\overline{\boldsymbol{\mu}}_n \in \mathrm{Bar}_G(\boldsymbol{\mu}_i, 1/n)_{1\le i\le n}$ *be a measurable choice of empirical Fréchet barycenters. Then the sequence* $\{\overline{\boldsymbol{\mu}}_n\}_{n\in\mathbb{N}}$ *is precompact a.s. and every its partial limit is the barycenter of the distribution* P_μ.

Proof. Let us consider empirical measures $\boldsymbol{P}_n := \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{\mu}_i}$. Obviously, the empirical barycenter $\boldsymbol{\nu}_n := \mathrm{bar}_G(\boldsymbol{\mu}_i, 1/n)_{1\le i\le n} = \mathrm{bar}_G(\boldsymbol{P}_n)$. By the law of large numbers

$$\mathcal{J}(\boldsymbol{P}_n, \delta_{\nu^*}) = \frac{1}{n}\sum_i J(\boldsymbol{\mu}_i, \nu^*) \to \mathbb{E}\, J(\boldsymbol{\mu}, \nu^*) = \mathcal{J}(P_\mu, \nu^*) < \infty \text{ a.s.},$$

and $\boldsymbol{P}_n \rightharpoonup P_\mu$ since the topology of weak convergence in $\mathcal{P}(E_0)$ has a countable basis due to separability of E_0. Then by Theorem 3.7 $\boldsymbol{P}_n \xrightarrow{\mathcal{J}} P_\mu$ almost surely, i.e. the conditions of the theorem hold. □

Actually, the results just proved also hold for an arbitrary equivalence class under some additional assumption: let for any $\epsilon > 0$ there exist constants $A_\epsilon, C_\epsilon > 0$ such that

$$\begin{aligned} c(x,y) &\le A_\epsilon + (1+\epsilon)c(x,z) + C_\epsilon c(y,z), \\ c(x,y) &\le A_\epsilon + (1+\epsilon)c(z,y) + C_\epsilon c(z,x), \end{aligned}$$

for all $x, y, z \in X$. This condition is stronger than that in Assumption 3.1, but they coincide i.e. in Euclidean case with convex cost function considered in Section 4 (Corollary 4.5). Under such an assumption Theorem 5.5 holds for any equivalence class $E(\mu_0)$.

Notice that all the statements in this section also hold for the space $\mathcal{P}(X)$ instead of $\mathcal{P}\big(\mathcal{P}(X)\big)$ because one can identify a point $x \in X$ with a Dirac measure $\delta_x \in \mathcal{P}(X)$ so that $J(\delta_x, \delta_y) = c(x,y)$ for all $x, y \in X$.

Acknowledgement. The author thanks Andrei Sobolevski for many fruitful discussions and precious advice.

References

[1] M. Agueh, G. Carlier: Barycenters in the Wasserstein space, SIAM J. Math. Analysis 43(2) (2011) 904–924.

[2] L. Ambrosio, N. Gigli, G. Savaré: Gradient Flows in Metric Spaces and in the Space of Probability Measures, Birkhäuser, Basel (2008).

[3] J. Bigot, E. Cazelles, N. Papadakis: Regularization of barycenters in the Wasserstein space, ArXiv e-prints (2016) http:// arxiv.org/ abs/ 1606.01025v1.

[4] J. Bigot, T. Klein: Consistent estimation of a population barycenter in the Wasserstein space, ArXiv e-prints (2015) http:// arxiv.org/ abs/ 1212.2562v4.

[5] V. Bogachev: Measure Theory, volume 2, Springer, Berlin (2007).

[6] Y.-H. Kim, B. Pass: Wasserstein barycenters over Riemannian manifolds, Adv. Math. 307 (2017) 640–683.

[7] A. Kroshnin, A. Sobolevski: Fréchet barycenters and a law of large numbers for measures on the real line, ArXiv e-prints (2015) http:// arxiv.org/ abs/ 1512.08421.

[8] T. Le Gouic, J.-M. Loubes: Existence and consistency of Wasserstein barycenters, Probab. Theory Rel. Fields 168(3-4) (2017) 901–917.

[9] F. Santambrogio: Optimal Transport for Applied Mathematicians, Birkhäuser, Basel (2015).

[10] C. Villani: Optimal Transport, Old and New, Springer, Berlin (2009).