Состоятельность барицентров Фреше

Алексей Крошнин
Институт проблем передачи информации им. А. А. Харкевича
Московский физико-технический институт
kroshnin@phystech.edu

Аннотация

В статье рассматривается пространство $\mathscr{P}(X)$ вероятностных мер на произвольном пространстве Радона X, снабженное транспортным функционалом $J(\cdot,\cdot)$. Исследуется топология, индуцированная транспортным расстоянием, в частности, для случая $X = \mathbb{R}^d$ с выпуклой ценовой функцией. Вводится понятие обобщенного среднего для набора мер и для вероятностного распределения на $\mathscr{P}(X)$, называемое здесь барицентром Фреше. Для барицентров распределений доказываются достаточные условия сходимости, в частности, показывается состоятельность эмпирических барицентров.

Ключевые слова: *задача Монжа-Канторовича, транспортное расстояние, выпуклая ценовая функция, барицентр, состоятельность.*

1. Введение

Меры, в том числе вероятностные, являются естественными объектами во многих областях. Существуют различные формализации понятия близости распределений, например: дивергенция Кульбака-Лейблера, расстояние полной вариации или слабая сходимость. В то же время, первые две из них не позволяют учитывать геометрическую структуру пространства, на котором рассматриваются распределения, а топология слабой сходимости хоть и метризуема, но метрика, ее порождающая, вряд ли применима на практике. В этом плане оказывается весьма полезной задача Монжа-Канторовича и определяемое с ее помощью транспортное расстояние. Задача Монжа-Канторовича, или задача оптимального транспорта, состоит в нахождении отображения одной вероятностной меры в другую с минимальной стоимостью, если задана ценовая функция $c(x,y)$ — стоимость переноса единицы массы из точки x в точку y (см. [7], гл. 1). Соответственно, транспортным расстоянием между двумя мерами будем называть эту минимальную стоимость преобразования. Очевидно, данное расстояние позволяет учитывать структуру пространства в терминах подходящим образом определенной функции $c(\cdot,\cdot)$, поэтому задача Монжа-Канторовича нашла широкое применение во многих областях, например, в анализе изображений (в том числе многомерных) и других сложных объектов.

Однако интерес представляет не только задача Монжа-Канторовича между двумя конкретными мерами, но и структура пространства $\mathscr{P}(X)$ вероятностных мер на X, снабженного транспортным расстоянием. Как правило, рассматриваются меры на $\mathbb{R}^d$ и ценовая функция вида $c(x,y) = |x-y|^p$, где $p \geq 1$. Пространство вероятностных мер с конечным p-м моментом, оснащенное соответствующим транспортным расстоянием $W_p(\cdot,\cdot)$, называется пространством p-Вассерштейна ([7], гл. 5). Оказывается, пространство Вассерштейна обладает многими хорошими свойствами, в частности, $\sqrt[p]{W_p(\cdot,\cdot)}$ является метрикой, более того, строго внутренней (см. [2], гл. 2), т. е. это геодезическое пространство (Теорема 5.24, [7]).

Возможность измерять расстояние между мерами позволяет также поставить другую важную задачу — о нахождении среднего в пространстве мер. Очевидно, меры можно усреднять как элементы векторного пространства, но такое усреднение, опять же, не позволяет никак учитывать геометрию исходного пространства. В статье Аге и Карлье 2011 года [3] вводится понятие *барицентра Вассерштейна* (для пространства 2-Вассерштейна) как среднего в смысле Фреше, т. е. меры, которая минимизирует сумму стоимостей своего оптимального переноса в каждую из заданного набора мер. С точки зрения статистических приложений, естественно рассматривать случайные величины в пространстве мер. Например, если дана последовательность i. i. d. мер или мерозначный случайный процесс, то возникает вопрос о состоятельности эмпирических барицентров, т. е. барицентров первых n мер, в смысле транспортного расстояния. Определение барицентра очевидным образом можно обобщить на распределение на мерах, как в статьях [6, 5], где также доказывается сходимость эмпирических барицентров к барицентру распределения (но все так же для пространств Вассерштейна).

В данной работе рассматривается наиболее общая постановка транспортной задачи: для мер над произвольным пространством Радона X и широкого класса ценовых функций. Оказывается, что даже при достаточно слабых ограничениях пространство мер, снабженное транспортным расстоянием $J(\cdot,\cdot)$, обладает хорошими свойствами. В частности, транспортное расстояние порождает топологию τ_J, и само пространство мер $\mathscr{P}(X)$ (вернее, каждый класс эквивалентности, на котором расстояние Монжа-Канторовича между мерами конечно) является радоновским. Также в работе будет показано, что при тех же самых ограничениях барицентр распределения на мерах всегда существует. Кроме того, будет доказан результат, аналогичный Теореме 2 из [6] — состоятельность барицентров. В частности, отсюда следует сходимость барицентров по i. i. d. выборке мер к барицентру распределения, аналогично Теореме 6.1 из [5] (для пространства 2-Вассерштейна). Помимо общего случая, будет рассмотрено пространство $\mathbb{R}^d$ с выпуклой ценовой функцией, как наиболее часто возникающее в приложениях.

Статья построена следующим образом. В разделе 2 дается строгая формулировка транспортной задачи и рассматриваются свойства расстояния Монжа-Канторовича. Также, при дополнительных предположениях на ценовую функцию, исследуются свойства топологии, порожденной транспортным расстоянием. В следующем разделе 3 отдельно рассматривается случай евклидова пространства с выпуклой ценовой функцией, что наиболее часто встречается на практике. В разделе 4 вводится определение барицентра распределения на пространстве мер и доказывается его существование. Также в этом разделе устанавливается состоятельность барицентров, в том числе непрерывность барицентра от конечного набора мер и функциональный закон больших чисел для эмпирических барицентров.

2. Пространство Монжа-Канторовича

2.1. Используемые обозначения

Введем некоторые обозначения, которые будут использоваться в статье.

Через $\mathscr{P}(X)$ будем обозначать пространство вероятностных мер на измеримом пространстве X. Если не оговорено обратное, то подразумевается, что X — топологическое пространство с борелевской сигма-алгеброй $\mathscr{B}(X)$.

Пусть даны измеримые пространства X, Y и измеримое отображение $T\colon X \to Y$. Пусть μ — мера на пространстве X. Через $T_{\#}\mu$ будем обозначать образ меры (the image measure) μ при отображении T, т. е. меру на Y, определяемую следующим образом:

$$(T_{\#}\mu)(A) := \mu\big(T^{-1}(A)\big),$$

где $A \subset Y$ — любое измеримое множество.

Пусть дана мера μ на X и интегрируемая относительно нее функция $f(\cdot)$. Определим меру $f\lfloor\mu$:

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

для всех измеримых множеств $A \subset X$. В частности, для любого измеримого множества B определим меру $B\lfloor\mu := \chi_B\lfloor\mu$, где $\chi_B(\cdot)$ — характеристическая функция множества B.

При интегрировании обозначения аргумента функции и пространства иногда будут опускаться, если это не может вызвать неоднозначности.

Максимум из двух чисел будем обозначать $a \vee b := \max\{a,b\}$.

2.2. Транспортное расстояние

Определение 2.1. *Пространством Радона* будем называть сепарабельное метризуемое топологическое пространство, на котором любая борелевская вероятностная мера является радоновской, т. е. допускает внутреннюю аппроксимацию компактами (опр. 5.1.4, [4]).

Замечание 2.1. Как известно (см. [1], С. 84-85), на метрическом пространстве борелевская вероятностная мера является радоновской тогда и только тогда, когда она плотна. Отсюда следует, в частности, что любое польское пространство и любое сепарабельное локально-компактное пространство являются радоновскими.

Пусть X — пространство Радона, и задана непрерывная *ценовая функция* (cost function) $c\colon X\times X \to [0,\infty)$ такая, что $c(x,y)=0 \Leftrightarrow x=y$. Введем в рассмотрение “шары” относительно ценовой функции $c(\cdot,\cdot)$:

$$B^c_r(x) := \big\{y\in X : c(x,y) < r\big\},$$
$$\overline{B}^c_r(x) := \big\{y\in X : c(x,y) \le r\big\},$$

для произвольных $x\in X$, $r>0$. В силу непрерывности ценовой функции, “открытый шар” $B^c_r(x)$ является открытым множеством, а “замкнутый шар” $\overline{B}^c_r(x)$ — замкнутым. Более того, потребуем условия *согласованности* ценовой функции с топологией: пусть для любой точки $x\in X$ и ее открытой окрестности $U(x)$ найдется шар $B^c_r(x)\subset U(x)$. Заметим, что вместе с непрерывностью это эквивалентно тому, чтобы $c(x,x_n)\to 0$ было равносильно $x_n\to x$ для произвольных x, $\{x_n\}_{n\in\mathbb{N}}\subset X$.

Рассмотрим произвольные меры $\mu,\nu \in \mathscr{P}(X)$ и определим множество *транспортных планов* (transport plan)

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

где π^x, π^y — операторы проекции на первую и вторую компоненту, соответственно.

Определение 2.2. *Расстоянием Монжа-Канторовича*, или *транспортным функционалом*, будем называть

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

где $K(\gamma) := \int c(x,y)\,d\gamma(x,y)$.

Замечание 2.2. Вообще говоря, функционал $J(\cdot,\cdot)$ не является расстоянием в обычном смысле — так, он может быть не симметричным и не удовлетворять неравенству треугольника.

Определение 2.3. Оптимальным транспортным планом из μ в ν называется элемент $\gamma^* \in \Pi(\mu,\nu)$, на котором достигается минимум $K(\cdot)$, т. е. $K(\gamma^*) = J(\mu,\nu)$.

***Теорема* 2.3.** *Для любых мер $\mu,\nu \in \mathscr{P}(X)$ существует оптимальный транспортный план.*

Доказательство. Рассмотрим минимизирующую $K(\cdot)$ последовательность $\{\gamma_n\}_{n\in\mathbb{N}} \subset \Pi(\mu,\nu)$. Меры μ и ν являются плотными, поэтому и множество транспортных планов $\Pi(\mu,\nu)$ — тоже плотное. Действительно, для любого $\varepsilon > 0$ найдутся такие компакты K^ε_μ, K^ε_ν, что $\mu(X \setminus K^\varepsilon_\mu) < \varepsilon/2$ и $\nu(X \setminus K^\varepsilon_\nu) < \varepsilon/2$. Тогда для компакта $K^\varepsilon := K^\varepsilon_\mu \times K^\varepsilon_\nu \subset X \times X$ и любого транспортного плана $\gamma \in \Pi(\mu,\nu)$ выполняется

$$\gamma(X\times X\setminus K^\varepsilon) \le \gamma\big(X\times (X\setminus K^\varepsilon_\nu)\big) + \gamma\big((X\setminus K^\varepsilon_\mu)\times X\big) = \\ = \nu(X\setminus K^\varepsilon_\nu) + \mu(X\setminus K^\varepsilon_\mu) < \varepsilon.$$

Следовательно, по теореме Прохорова найдется слабо сходящаяся подпоследовательность $\gamma_{n_k} \rightharpoonup \gamma^* \in \Pi(\mu,\nu)$. Как несложно видеть, $K(\cdot)$ — полунепрерывен снизу относительно слабой сходимости, поэтому

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

т. е. γ^* — оптимальный транспортный план. $\square$

***Следствие* 2.4.** $J(\mu,\nu) = 0$ *тогда и только тогда, когда* $\mu = \nu$.

Напомним еще некоторые свойства функционала $J(\cdot,\cdot)$, многие из которых хорошо известны (см., например, [7, 8, 4]).

***Лемма* 2.5.** *Функционал $J(\cdot,\cdot)$ полунепрерывен снизу относительно слабой сходимости.*

Доказательство. Пусть $\mu_n \rightharpoonup \mu$, $\nu_n \rightharpoonup \nu$, $\gamma_n \in \Pi(\mu_n,\nu_n)$ — оптимальный транспортный план из μ_n в ν_n. Так как $\{\mu_n\}_{n\in\mathbb{N}}$ и $\{\nu_n\}_{n\in\mathbb{N}}$ слабо сходятся, то это плотные последовательности ([4], с. 108), поэтому $\{\gamma_n\}_{n\in\mathbb{N}}$ — также плотная. Рассмотрим подпоследовательность, на которой достигается нижний предел $\liminf J(\mu_n,\nu_n) \in [0,\infty]$. По теореме Прохорова, у нее существует слабо сходящаяся подпоследовательность $\gamma_{n_k} \rightharpoonup \tilde\gamma$. Очевидно, $\tilde\gamma \in \Pi(\mu,\nu)$. Т.к. $K(\cdot)$ — полунепрерывен снизу относительно слабой сходимости, то

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

Следовательно, $J(\mu,\nu) \le \liminf J(\mu_n,\nu_n)$. $\square$

***Следствие* 2.6.** *Функционал $J(\cdot,\cdot)$ измерим относительно произведения борелевских сигма-алгебр на X, порожденных топологией слабой сходимости.*

***Теорема* 2.7.** *Пусть дана мера ν^*, последовательность $\{\nu_n\}_{n\in\mathbb{N}}$, и $J(\nu^*,\nu_n)\to 0$. Тогда $\nu_n \rightharpoonup \nu^*$.*

Доказательство. Допустим, $\nu_n \not\rightharpoonup \nu^*$. Тогда существует замкнутое множество $C \subset X$, для которого $\limsup \nu_n(C) > \nu^*(C)$. Без о. о. (ограничения общности) будем считать, что $3\varepsilon := \lim \nu_n(C) - \nu^*(C) > 0$. Рассмотрим следующие окрестности множества C:

$$C_r := \big\{x\in X : \inf_{y\in C} c(x,y) < r\big\} \supset C,\ r>0.$$

Они, очевидно, открыты в силу непрерывности $c(\cdot,\cdot)$. В силу замкнутости C и согласованности ценовой функции, для любой точки $x\notin C$ найдется шар $B^c_r(x)\cap C = \emptyset$. Следовательно, $\bigcap_{r>0} C_r = C$, и $\nu^*(C) = \lim_{r\to0}\nu^*(C_r)$. Выберем такое $r_0>0$, что $\nu^*(C_{r_0}) < \nu^*(C)+\varepsilon$. Начиная с некоторого n, $\nu_n(C) > \nu^*(C)+2\varepsilon$. Тогда для оптимального транспортного плана $\gamma_n \in \Pi(\nu^*,\nu_n)$ выполняется

$$\gamma_n\big((X\setminus C_{r_0})\times C\big) = \nu_n(C) - \gamma_n(C_{r_0}\times C) \ge \\ \ge \nu_n(C) - \nu^*(C_{r_0}) > \varepsilon,$$

следовательно,

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

Но тогда $\liminf J(\nu^*,\nu_n) \ge \varepsilon r_0 > 0$ — противоречие. $\square$

Как только что было показано, из сходимости относительно транспортного функционала следует слабая сходимость. На самом деле, обратное тоже почти верно с некоторыми дополнительными ограничениями.

***Теорема* 2.8.** *Пусть даны меры $\nu_n \rightharpoonup \nu^*$ такие, что $\operatorname{supp}\nu_n \subset A$ для всех n, где $A\subset X$ — замкнутое множество, на котором $\sup_{x,y\in A} c(x,y) < \infty$. Тогда $\lim J(\nu_n,\nu^*) = \lim J(\nu^*,\nu_n) = 0$.*

Доказательство. Выберем какую-нибудь согласованную метрику $\rho(\cdot,\cdot)$ на X (это можно сделать, т. к. X — метризуемо). Зафиксируем $\varepsilon>0$. В силу сепарабельности X и непрерывности $c(\cdot,\cdot)$, можно покрыть X счетным объединением замкнутых шаров $\{\overline{B}_{r_i}(x_i)\}_{i\in\mathbb{N}}$ таких, что $c(x,y)<\varepsilon$ при $x,y\in\overline{B}_{2r_i}(x_i)$ для всех i. Выберем такое $m\in\mathbb{N}$, чтобы $\nu^*\big(X\setminus\bigcup_{i=1}^m B_{r_i}(x_i)\big) < \varepsilon$. Рассмотрим непрерывные функции $f_i(\cdot)\colon X\to[0,1]$, $0\le i\le m$, для которых выполняется

$$\sum_{i=0}^m f_i(x) \equiv 1;$$
$$f_0(x) = 0,\ x\in\overline{B}_{r_i}(x_i),\ 1\le i\le m;$$
$$f_i(x) = 0,\ x\notin\overline{B}_{2r_i}(x_i),\ 1\le i\le m.$$

Без о. о. можно считать, что $\int f_i\,d\nu^* > 0$ для всех i. Определим меры

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

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

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

Из слабой сходимости следует, что

$$\hat\nu_n(X)=\hat\nu_n^*(X)=\int\Big(1-\sum_{i=1}^m\frac{\int f_i\,d\nu_n}{\int f_i\,d\nu_n\vee\int f_i\,d\nu^*}f_i\Big)\,d\nu^*\to \to\int f_0\,d\nu^*\le\nu^*\big(X\setminus\bigcup_{i=1}^m B_{r_i}(x_i)\big)<\varepsilon.$$

Построим транспортные планы

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

Т.к. $\operatorname{supp}\nu_n\subset A$, то и $\operatorname{supp}\nu^*\subset A$, поэтому $\operatorname{supp}\gamma_n\subset A\times A$. Обозначим $M:=\sup_{x,y\in A}c(x,y)<\infty$. $\operatorname{supp}\lambda_n^i\subset\overline{B}_{2r_i}(x_i)\times\overline{B}_{2r_i}(x_i)$ по построению функций $f_i(\cdot)$. Тогда

$$J(\nu_n,\nu^*)\le K(\gamma_n)\le M\frac{\hat\nu_n(X)\hat\nu_n^*(X)}{\hat\nu_n(X)}+\varepsilon\sum_{i=1}^m\lambda_n^i(X\times X)\le \le M\hat\nu_n^*(X)+\varepsilon\to M\int f_0\,d\nu^*+\varepsilon\le(1+M)\varepsilon.$$

Таким образом, $J(\nu_n,\nu^*)\to 0$. Аналогично доказывается, что $J(\nu^*,\nu_n)\to 0$. □

Следующее предположение позволит получить слабую локальную компактность $\mathscr{P}(X)$, что будет использовано в дальнейшем при доказательстве свойств барицентров Фреше.

Предположение 2.1. Пусть существует такая точка $x_0\in X$, что любой замкнутый шар $\overline{B}^c_r(x_0)$ компактен. Пусть также для любого фиксированного $r>0$ выполняется

$$\inf\{c(x,y): x\in\overline{B}^c_r(x_0),\ y\notin B^c_R(x_0)\}\to\infty \text{ при } R\to\infty.$$

Замечание 2.9. Вообще говоря, если это предположение верно для какой-то x_0, то оно выполняется и для любой другой точки $x\in X$.

Заметим, что отсюда следует условие согласованности ценовой функции. Кроме того, при этом предположении X является локально-компактным. Следовательно, для того, чтобы оно было радоновским, достаточно сепарабельности. Отметим также, что если X — компактно, то данное предположение всегда выполняется.

***Лемма* 2.10.** *Пусть дана последовательность мер $\{\nu_n\}_{n\in\mathbb{N}}$ такая, что для некоторой меры μ выполняется $\limsup J(\mu,\nu_n)<\infty$. Если выполнено Предположение 2.1, то эта последовательность плотная.*

Доказательство. Зафиксируем $\varepsilon>0$ и шар $B=\overline{B}^c_r(x_0)$ такой, что $\mu(X\setminus B)<\varepsilon/2$. Рассмотрим $R>r$ и произвольную меру λ_R, для которой $\lambda_R\big(X\setminus B^c_R(x_0)\big)\ge\varepsilon$. Тогда

$$J(\mu,\lambda_R)\ge\frac{\varepsilon}{2}\inf\{c(x,y): x\in B,\ y\notin B^c_R(x_0)\}\to\infty$$

при $R\to\infty$. Учитывая, что $\limsup J(\mu,\nu_n)<\infty$, любой замкнутый шар $\overline{B}^c_R(x_0)$ компактен, и каждая отдельная мера является плотной, получаем, что найдется компакт K_ε, для которого $\nu_n(X\setminus K_\varepsilon)<\varepsilon$ при всех n. □

***Следствие* 2.11.** *При Предположении 2.1 любой “замкнутый шар” $\overline{B}^J_r(\mu):=\{\nu\in\mathscr{P}(X): J(\mu,\nu)\le r\}$ является слабым компактом.*

Доказательство. Рассмотрим произвольную последовательность $\{\nu_n\}_{n\in\mathbb{N}}\subset\overline{B}^J_r(\mu)$. Она является плотной, и по теореме Прохорова у нее есть слабо сходящаяся подпоследовательность $\nu_{n_k}\rightharpoonup\nu^*$. По Лемме 2.5, $J(\mu,\nu^*)\le\liminf J(\mu,\nu_{n_k})\le r$, т. е. $\nu^*\in\overline{B}^J_r(\mu)$. □

2.3. Топология, индуцированная транспортным функционалом

В этом разделе будем считать, что для ценовой функции выполняется следующее ослабленное неравенство треугольника.

Предположение 2.2. Пусть для любого $\varepsilon>0$ существует такая константа $C_\varepsilon>0$, что для всех $x,y,z\in X$ выполняется

$$c(x,y)\le\varepsilon+(1+\varepsilon)c(x,z)+C_\varepsilon c(y,z),$$
$$c(x,y)\le\varepsilon+(1+\varepsilon)c(z,y)+C_\varepsilon c(z,x).$$

Замечание 2.12. Пусть $\rho(\cdot,\cdot)$ — некоторая согласованная метрика на X. Тогда Предположение 2.2 выполняется, например, для функций вида $c(x,y):=f\big(\rho(x,y)\big)$, где $f\colon[0,\infty)\to[0,\infty)$ — строго возрастающая выпуклая функция, $f(0)=0$ и существуют такие константы A,B, что $f(u+v)\le A+B\big(f(u)+f(v)\big)$ для всех $u,v\ge 0$. Это следует из неравенства треугольника для метрики и леммы, которая будет доказана ниже.

Также, условие выполнено, если ценовая функция является степенью расстояния $c(x,y):=\rho^p(x,y)$, для любого $p>0$.

***Лемма* 2.13.** *Если пространство X компактно, то Предположение 2.2 всегда выполняется.*

Доказательство. Зафиксируем $\varepsilon>0$. В силу равномерной непрерывности ценовой функции, можно выбрать открытое множество $U\in X\times X$ такое, что $(y,y)\in U$ для всех $y \in X$ и $c(x,y) < c(x,z)+\varepsilon$ для любого $x \in X$, если $(y,z) \in U$. В силу компактности, $\delta := \min_{(y,z)\notin U} c(y,z) > 0$. Пусть $M := \max_{x,y\in X} c(x,y)$. Следовательно, если $(y,z) \notin U$, то $c(x,y) \le M \le \frac{M}{\delta} c(y,z)$ при любом x. Таким образом,

$$c(x,y) \le \varepsilon + c(x,z) + \frac{M}{\delta} c(y,z).$$

Второе неравенство доказывается аналогично. □

Покажем, что расстояние Монжа-Канторовича "наследует" неравенства на ценовую функцию.

***Теорема* 2.14.** *Для любого $\varepsilon > 0$ и всех мер $\mu, \nu, \lambda \in \mathscr{P}(X)$ выполняются неравенства*

$$J(\mu,\nu) \le \varepsilon + (1+\varepsilon)J(\mu,\lambda) + C_\varepsilon J(\nu,\lambda),$$
$$J(\mu,\nu) \le \varepsilon + (1+\varepsilon)J(\lambda,\nu) + C_\varepsilon J(\lambda,\mu).$$

Доказательство. Рассмотрим меры $\mu, \nu, \lambda \in \mathscr{P}(X)$. Пусть $\gamma_1 \in \Pi(\mu,\lambda)$, $\gamma_2 \in \Pi(\nu,\lambda)$ — оптимальные транспортные планы. Из теоремы о дезинтеграции меры (Теорема 5.3.1, [4]) следует, что существует мера $\sigma \in \Pi(\mu,\nu,\lambda)$, для которой $\pi^{x,z}_\# \sigma = \gamma_1$, $\pi^{y,z}_\# \sigma = \gamma_2$ (Лемма 5.3.2., [4]). Тогда непосредственно из Предположения 2.2 следует, что

$$\begin{aligned} J(\mu,\nu) &\le K(\pi^{x,y}_\# \sigma) = \int c(x,y)\, d\sigma \le \\ &\le \int \big(\varepsilon + (1+\varepsilon)c(x,z) + C_\varepsilon c(y,z)\big)\, d\sigma = \\ &= \varepsilon + (1+\varepsilon)K(\gamma_1) + C_\varepsilon K(\gamma_2) = \\ &= \varepsilon + (1+\varepsilon)J(\mu,\lambda) + C_\varepsilon J(\nu,\lambda). \end{aligned}$$

Второе неравенство доказывается аналогично. □

***Следствие* 2.15** (топология)**.** *Семейство шаров $B^J_r(\mu) := \{\nu \in \mathscr{P}(X) : J(\mu,\nu) < r\}$, где $\mu \in \mathscr{P}(X)$, $r > 0$, образует базу топологии τ_J в $\mathscr{P}(X)$, и функционал $J(\cdot,\cdot)$ непрерывен относительно этой топологии.*

Сходимость последовательности $\{\nu_n\}_{n\in\mathbb{N}}$ к мере ν^* в топологии τ_J будем обозначать $\nu_n \xrightarrow{J} \nu^*$, что эквивалентно $J(\nu_n,\nu^*) \to 0$, а также $J(\nu^*,\nu_n) \to 0$ (так как $J(\mu,\nu) \le \varepsilon + C_\varepsilon J(\nu,\mu)$). Как следует из Теоремы 2.7, τ_J не слабее, чем топология слабой сходимости, поэтому будем называть сходимость относительно транспортного функционала сильной сходимостью.

***Теорема* 2.16.** *Если пространство X компактно, то $\big(\mathscr{P}(X),\tau_J\big)$ тоже компактно, и слабая сходимость мер совпадает с сильной.*

Доказательство. Как известно, в этом случае $\mathscr{P}(X)$ с топологией слабой сходимости является метризуемым компактным пространством. Как было показано выше, для компактного пространства X всегда выполняются Предположения 2.1 и 2.2. Следовательно, в силу Теорем 2.7 и 2.8, слабая сходимость в $\mathscr{P}(X)$ совпадает с сильной, откуда следует компактность $\big(\mathscr{P}(X),\tau_J\big)$. □

Замечание 2.17. Если X не компактно, то пространство $\mathscr{P}(X)$ не является не только компактным, но даже локально-компактным, в отличие от X.

Введем отношение эквивалентности на мерах: $\mu \sim \nu$, если $J(\mu,\nu) < \infty$. Это отношение задает разбиение пространства $\mathscr{P}(X)$ на классы эквивалентности $C(\mu) := \{\nu \in \mathscr{P}(X) : J(\mu,\nu) < \infty\}$.

Рассмотрим следующую конструкцию, которая поможет установить соотношение между сильной и слабой сходимостью: зафиксируем некоторую точку $x_0 \in X$ и для заданного $R > 0$ выберем непрерывную функцию $f_R\colon X\times X \to [0,1]$ такую, что $f_R(x,y) = 1$ при $x,y \in B^c_R(x_0)$, $f_R(x,y) = 0$ при $x \notin B^c_{R+1}(x_0)$ или $y \notin B_{R+1}(x_0)$. Пусть даны меры μ, ν и $\gamma \in \Pi(\mu,\nu)$. Рассмотрим $\lambda := f_R\lfloor\gamma$. Определим меру

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

Также введем $\tilde\nu := (\pi^x)_\# \tilde\gamma$. Очевидно, $\tilde\gamma \in \Pi(\tilde\nu,\nu)$ и $J(\tilde\nu,\nu) \le K(\tilde\gamma) = K(\gamma) - K(\lambda) = K(\gamma) - K'(\gamma)$, где $K'(\gamma) = K'_R(\gamma) := K(f_R\lfloor\gamma)$.

Если мы имеем слабо сходящуюся последовательность $\Pi(\mu,\nu_n) \ni \gamma_n \rightharpoonup \gamma^* \in \Pi(\mu,\nu^*)$, то и $\tilde\gamma_n \rightharpoonup \tilde\gamma^*$, $\tilde\nu_n \rightharpoonup \tilde\nu^*$. Заметим, что на дополнении шара $B^c_{R+1}(x_0)$ все меры $\tilde\nu$ совпадают с μ, поэтому по Теореме 2.8 из слабой сходимости следует сильная: $\tilde\nu_n \xrightarrow{J} \tilde\nu^*$.

Как было показано в Теореме 2.7, из сильной сходимости следует слабая. В то же время, согласно Теореме 2.8, для мер с ограниченным носителем имеет место обратная импликация. Следующая теорема дает необходимые и достаточные условия сходимости в сильной топологии τ_J.

***Теорема* 2.18** (критерий сильной сходимости)**.** *Пусть дана мера ν^* и последовательность $\{\nu_n\}_{n\in\mathbb{N}}$. Следующие условия эквивалентны:*

1. $\nu_n \xrightarrow{J} \nu^*$.
2. $\nu_n \rightharpoonup \nu^*$ *и* $J(\mu,\nu_n) \to J(\mu,\nu^*)$ *для любой меры* $\mu \in \mathscr{P}(X)$.
3. $\nu_n \rightharpoonup \nu^*$ *и* $J(\mu,\nu_n) \to J(\mu,\nu^*) < \infty$ *для какой-то меры* $\mu \in C(\nu^*)$.

Доказательство. Очевидно, 1) ⇒ 2) ⇒ 3). Покажем, что 3) ⇒ 1). Без о. о. можно считать, что все $J(\mu,\nu_n) < \infty$. Пусть $\gamma_n \in \Pi(\mu,\nu_n)$ — оптимальный транспортный план. Последовательность $\{\gamma_n\}_{n\in\mathbb{N}}$ — плотная, поэтому можно выделить слабо сходящуюся подпоследовательность $\gamma_n \rightharpoonup \hat\gamma \in \Pi(\mu,\nu^*)$ (без переобозначения). Зафиксируем $\varepsilon > 0$ и выберем такое $R > 0$, чтобы $K'(\hat\gamma) = K'_R(\hat\gamma) > K(\hat\gamma) - \varepsilon$. Построим меры $\Pi(\tilde\nu_n,\nu_n) \ni \tilde\gamma_n \rightharpoonup \tilde\gamma \in \Pi(\tilde\nu^*,\nu^*)$, как описано выше. Получаем, что $J(\tilde\nu_n,\tilde\nu^*) \to 0$, $J(\tilde\nu^*,\nu^*) \le K(\tilde\gamma) = K(\hat\gamma) - K'(\hat\gamma) < \varepsilon$ и

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

Следовательно, $J(\nu^*,\nu_n) \to 0$. □

Замечание 2.19. Очевидно, аргументы можно поменять местами: $J(\nu_n,\mu)\to J(\nu^*,\mu)$.

Теперь мы хотим показать, что при Предположениях 2.1 и 2.2 для любой меры $\mu_0\in\mathscr{P}(X)$ пространство $\big(C(\mu_0),\tau_J\big)$ является радоновским. Для этого необходимо доказать его метризуемость, сепарабельность и плотность произвольной вероятностной борелевской меры на этом пространстве.

***Лемма* 2.20.** *Пусть дана мера $\mu_0\in\mathscr{P}(X)$. Тогда класс эквивалентности $C(\mu_0)$ сепарабельный.*

Доказательство. Рассмотрим счетное семейство мер $\mathscr{S}_{\mu_0}$, состоящее из мер вида $\nu:=\big(X\setminus B_m(x_0)\big)\lfloor\mu_0+\alpha\sum_{i=1}^n p_i\delta_{x_i}$, где $m,n\in\mathbb{N}$, $p_i\in\mathbb{Q}_+$, x_i из некоторого счетного всюду плотного подмножества X, α — нормирующий множитель. Зафиксируем меру $\mu\in C(\mu_0)$, число $\varepsilon>0$ и выберем $R>0$ таким, что $\int_{B_R(x_0)\times B_R(x_0)}c(x,y)\,d\gamma>K(\gamma)-\varepsilon$, где γ — оптимальный транспортный план из μ_0 в μ. Следовательно, $K'(\gamma):=K(f_R\lfloor\gamma)>K(\gamma)-\varepsilon$, и $J(\tilde\mu,\mu)\le K(\gamma)-K'(\gamma)<\varepsilon$. Очевидно, $\tilde\mu$ содержится в слабом замыкании $\mathscr{S}_{\mu_0}$, поэтому найдется мера $\nu\in\mathscr{S}_{\mu_0}$, для которой $J(\tilde\mu,\nu)<\varepsilon$. Тогда из Теоремы 2.14 следует, что $\mathscr{S}_{\mu_0}$ — всюду плотное множество в $C(\mu_0)$ □

***Лемма* 2.21.** *Для любой меры $\mu_0\in\mathscr{P}(X)$ пространство $\big(C(\mu_0),\tau_J\big)$ метризуемо.*

Доказательство. Для доказательства метризуемости воспользуемся леммой Урысона (в форме Тихонова). Для этого нужно показать, что пространство регулярно и обладает счетной базой топологии.

Очевидно, наше пространство удовлетворяет аксиоме отделимости Т-1 (любые две точки отделимы). Также из непрерывности $J(\cdot,\cdot)$ следует, что замыкание шара совпадает с замкнутым шаром $\overline{B}^J_r(\mu)=\big\{\nu\in C(\mu_0):J(\mu,\nu)\le r\big\}$, и для любой меры $\mu\in C(\mu_0)$ и $r>0$ выполняется

$$\mu\in B^J_{r/2}(\mu)\subset\overline{B}^J_{r/2}(\mu)\subset B^J_r(\mu).$$

Следовательно, пространство $\big(C(\mu_0),\tau_J\big)$ регулярное. С помощью Теоремы 2.14 и Леммы 2.20 несложно убедиться, что пространство имеет счетную базу $\big\{B^J_r(\mu):\mu\in\mathscr{S},\,r\in Q_+\big\}$, где $\mathscr{S}$ — счетное всюду плотное множество (например, $\mathscr{S}_{\mu_0}$ из Леммы 2.20). По лемме Урысона получаем, что пространство $\big(C(\mu_0),\tau_J\big)$ метризуемо. □

Покажем, что пространство $\mathscr{P}(X)$ является “полным” относительно расстояния Монжа-Канторовича.

***Лемма* 2.22.** *Пусть выполнено Предположение 2.1. Пусть дана последовательность $\{\nu_n\}_{n\in\mathbb{N}}$ из $\mathscr{P}(X)$, и $J(\nu_n,\nu_m)\to0$ при $n\ge m\to\infty$. Тогда существует такая мера ν^*, что $\nu_n\xrightarrow{J}\nu^*$.*

Доказательство. Последовательность $\{\nu_n\}_{n\in\mathbb{N}}$, очевидно, плотная по Лемме 2.10, поэтому у нее существует слабо сходящаяся подпоследовательность $\nu_{n_k}\rightharpoonup\nu^*\in\mathscr{P}(X)$. Допустим, $J(\nu^*,\nu_n)\nrightarrow0$. Тогда найдутся $\varepsilon>0$, $N\in\mathbb{N}$ такие, что $J(\nu^*,\nu_N)>\varepsilon$ и $J(\nu_n,\nu_N)<\varepsilon$ при $n\ge N$. Но в силу слабой полунепрерывности снизу $\varepsilon\ge\liminf J(\nu_{n_k},\nu_N)\ge J(\nu^*,\nu_N)>\varepsilon$ — противоречие. Следовательно, $J(\nu^*,\nu_n)\to0$. □

Предыдущие леммы позволяют доказать, что каждый класс эквивалентности является пространством Радона. Для этого остается показать, что любая борелевская вероятностная мера является плотной. Адаптируем к нашему случаю доказательство этого факта для польского пространства из [1] (Теорема 7.1.7).

***Теорема* 2.23.** *Если выполнено Предположение 2.1, то для любого $\mu_0\in\mathscr{P}(X)$ пространство $\big(C(\mu_0),\tau_J\big)$ является радоновским.*

Доказательство. Рассмотрим произвольное борелевское распределение P на $C(\mu_0)$. Покажем, что оно плотное. В силу сепарабельности и Теоремы 2.14, для любого $r>0$ пространство $C(\mu_0)$ можно покрыть счетным числом замкнутых шаров $\big\{\overline{B}^J_r(\mu_i)\big\}_{i\in\mathbb{N}}$. Зафиксируем $\varepsilon>0$ и для всех $n\in\mathbb{N}$ выберем такие конечные наборы шаров $\{\overline{B}^J_{1/n}(\mu_i)\}_{1\le i\le m_n}$, что $P\big(\bigcup_{i=1}^{m_n}\overline{B}^J_{1/n}(\mu_i)\big)>1-2^{-n}\varepsilon$. Определим множество $K_\varepsilon:=\bigcap_{n=1}^{\infty}\bigcup_{i=1}^{m_n}\overline{B}^J_{1/n}(\mu_i)$. По построению,

$$P\big(C(\mu_0)\setminus K_\varepsilon\big)\le\sum_{n=1}^{\infty}P\Big(C(\mu_0)\setminus\bigcup_{i=1}^{m_n}\overline{B}^J_{1/n}(\mu_i)\Big)<\sum_{n=1}^{\infty}2^{-n}\varepsilon=\varepsilon.$$

Кроме того, K_ε является замкнутым (как пересечение замкнутых) и “вполне ограниченным” относительно ценовой функции. Покажем, что оно компактно, аналогично критерию компактности для полного метрического пространства. Рассмотрим произвольную последовательность $\{\nu_n\}_{n\in\mathbb{N}}\subset K_\varepsilon$. Используя канторов диагональный процесс и определение K_ε, можно получить такую подпоследовательность $\{\nu_{n_k}\}_{k\in\mathbb{N}}$, что для любого k и всех $m\ge k$ найдется шар $\overline{B}^J_{1/k}(\mu_{i_k})\ni\nu_{n_m}$. Тогда для произвольного $\delta>0$ имеет место

$$\begin{aligned}J(\nu_{n_m},\nu_{n_k})&\le\delta+(1+\delta)J(\mu_{i_k},\nu_{n_k})+C_\delta J(\mu_{i_k},\nu_{n_m})\le\\&\le\delta+\frac{1+\delta+C_\delta}{k}\to\delta\ \text{при}\ k\to\infty.\end{aligned}$$

Следовательно, $J(\nu_{n_m},\nu_{n_k})\to0$ при $m\ge k\to\infty$, и по Лемме 2.22 у $\{\nu_{n_k}\}_{k\in\mathbb{N}}$ есть сильный предел ν^*. В силу замкнутости, $\nu^*\in K_\varepsilon$, откуда следует компактность. Таким образом, любое борелевское распределение P является плотным, поэтому $\big(C(\mu_0),\tau_J\big)$ — пространство Радона. □

Так как при соответствующих предположениях пространство Монжа-Канторовича, точнее, каждый класс эквивалентности $\big(C(\mu_0),\tau_J\big)$, является радоновским, то на нем, в свою очередь, можно ставить задачу Монжа-Канторовича. В качестве ценовой функции при этом естественно выбрать транспортный функционал $J(\cdot,\cdot)$. Такая задача частично будет рассматриваться ниже для исследования барицентров Фреше, т. е. обобщенных средних на пространстве мер.

3. Случай $\mathbb{R}^d$

Пусть теперь $X = \mathbb{R}^d$ — польское локально-компактное пространство, а ценовая функция имеет вид $c(x,y) := g(x-y)$, где $g(\cdot)$ — выпуклая функция (не обязательно строго) и достигает *строгого* минимума в нуле: $g(0)=0$. Очевидно, условие согласованности и Предположение 2.1 выполнены за счет выпуклости и локальной компактности. Наложим дополнительные ограничения на ценовую функцию:

Предположение 3.1. Пусть функция $g(\cdot)$ такова, что

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

***Лемма* 3.1.** *Если выполнено Предположение 3.1, то существует такое $q \geq 1$, что для всех $x,y \in \mathbb{R}^d$ выполняется неравенство треугольника:*

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

Доказательство. Рассмотрим такие точки x,y, что (без о. о.) $g(y) = \xi g(x)$, $\xi \leq 1$. В силу выпуклости функции $g(\cdot)$

$$g(x+y) \leq \frac{n-1}{n} g(x) + \frac{1}{n} g(x+ny) \leq g(x) + \frac{B}{n}\big(g(x)+g(ny)\big).$$

Пусть $n = 2^k$. Очевидно, $g(2^k y) \leq (2B)^k g(y)$. Получаем, что

$$g(x+y) \leq g(x) + 2^{-k} B g(x) + B^{k+1} g(y) = \\ = g(x)\big(1 + B(2^{-k} + \xi B^k)\big).$$

Пусть $0 < \xi \ll 1$. Выберем $k \approx -\frac{\ln \xi}{\ln 2B}$. Тогда $2^{-k} + \xi B^k \approx 2\xi^{1/q_0}$, где $q_0 := \frac{\ln 2B}{\ln 2} > 1$. Таким образом, $g(x+y) \leq g(x)\big(1 + O(\xi^{1/q_0})\big)$. Т.к. $\xi \leq 1$, то несложно видеть, что найдется допустимое $q' \geq 1$, при котором выполняется неравенство треугольника

$$g^{1/q'}(x+y) \leq g^{1/q'}(x)\big(1 + O(\xi^{1/q_0})\big)^{1/q'} \leq \\ \leq g^{1/q'}(x)(1 + \xi^{1/q'}) = g^{1/q'}(x) + g^{1/q'}(y)$$

для всех $x,y \in \mathbb{R}^d$. Величину q можно определить как минимум таких допустимых q' (по непрерывности q тоже будет допустимой). □

Замечание 3.2. Если $g(x) := |x|^p$, то $q = q_0 = p$.

***Теорема* 3.3.** *Для всех мер $\mu,\nu,\lambda \in \mathscr{P}(\mathbb{R}^d)$ выполняются следующие неравенства:*

$$J(\mu,\nu) \leq B\big(J(\mu,\lambda) + J(\lambda,\nu)\big),$$
$$J^{1/q}(\mu,\nu) \leq J^{1/q}(\mu,\lambda) + J^{1/q}(\lambda,\nu).$$

Доказательство. Рассмотрим меры $\mu,\nu,\lambda \in \mathscr{P}(\mathbb{R}^d)$. Пусть $\gamma_1 \in \Pi(\mu,\lambda)$, $\gamma_2 \in \Pi(\lambda,\nu)$ — оптимальные транспортные планы. Аналогично Теореме 2.14, рассмотрим меру $\sigma \in \Pi(\mu,\lambda,\nu)$, для которой $\pi^{x,y}_{\#}\sigma = \gamma_1$, $\pi^{y,z}_{\#}\sigma = \gamma_2$. Тогда непосредственно из Предположения 3.1 следует, что

$$J(\mu,\nu) \leq K(\pi^{x,z}_{\#}\sigma) = \int g(x-z)\,d\sigma \leq \\ \leq \int B\big(g(x-y) + g(y-z)\big)\,d\sigma = \\ = B\,(K(\gamma_1) + K(\gamma_2)) = B\big(J(\mu,\lambda) + J(\lambda,\nu)\big).$$

Используя Лемму 3.1 и неравенство треугольника для q-нормы, получаем

$$J^{1/q}(\mu,\nu) \leq K^{1/q}(\pi^{x,z}_{\#}\sigma) = \left(\int \big(g^{1/q}(x-z)\big)^q\,d\sigma\right)^{1/q} \leq \\ \leq \left(\int \big(g^{1/q}(x-y) + g^{1/q}(y-z)\big)^q\,d\sigma\right)^{1/q} \leq \\ \leq \left(\int g(x-y)\,d\sigma\right)^{1/q} + \left(\int g(y-z)\,d\sigma\right)^{1/q} = \\ = J^{1/q}(\mu,\lambda) + J^{1/q}(\lambda,\nu). \quad \square$$

***Следствие* 3.4.** *Функция $\rho(\mu,\nu) := J^{1/q}(\mu,\nu) \vee J^{1/q}(\nu,\mu) \in [0,\infty]$ является метрикой на $\mathscr{P}(\mathbb{R}^d)$.*

Таким образом, при данных ограничениях на функцию $g(\cdot)$ мы получаем практически пространство Васерштейна подходящей степени q. Заметим, что если Предположение 3.1 не выполнено, то не верна и Теорема 3.3.

***Утверждение* 3.5.** *Пусть $\sup_{x,y} \frac{g(x+y)}{g(x)+g(y)} = \infty$. Тогда для любого $q > 0$ найдутся такие меры $\mu_q,\nu_q,\lambda_q \in \mathscr{P}(\mathbb{R}^d)$, что*

$$J^{1/q}(\mu_q,\nu_q) > J^{1/q}(\mu_q,\lambda_q) + J^{1/q}(\lambda_q,\nu_q).$$

Доказательство. Зафиксируем $q > 0$ и рассмотрим точки $x_q, y_q \in \mathbb{R}^d$, для которых $g(x_q+y_q) > 2^q\big(g(x_q)+g(y_q)\big)$. Тогда

$$J^{1/q}(\delta_{x_q+y_q},\delta_0) = g^{1/q}(x_q+y_q) > \\ > (2^q)^{1/q}\big(g(x_q) \vee g(y_q)\big)^{1/q} \geq g^{1/q}(x_q) + g^{1/q}(y_q) = \\ = J^{1/q}(\delta_{x_q},\delta_0) + J^{1/q}(\delta_{x_q+y_q},\delta_{x_q}). \quad \square$$

Предположение 3.1 контролировало функцию $g(\cdot)$ на бесконечности и в окрестности 0. Теперь несколько ослабим условия на ценовую функцию — будем контролировать ее только на бесконечности.

Предположение 3.2. Пусть существуют константы $A,B \geq 0$ такие, что для всех $x,y \in \mathbb{R}^d$ выполняется

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

***Лемма* 3.6.** *Для любого $\varepsilon > 0$ существуют B_ε, $q_\varepsilon \geq 1$, C_ε такие, что для всех $x, y \in \mathbb{R}^d$ выполняется*

$$g(x+y) \leq \varepsilon \vee B_\varepsilon\big(g(x)+g(y)\big),$$
$$\big(g(x+y)+\varepsilon\big)^{1/q_\varepsilon} \leq \big(g(x)+\varepsilon\big)^{1/q_\varepsilon} + \big(g(y)+\varepsilon\big)^{1/q_\varepsilon},$$
$$g(x+y) \leq \varepsilon + (1+\varepsilon)g(x) + C_\varepsilon g(y).$$

Доказательство. Зафиксируем произвольное $\varepsilon > 0$. В силу выпуклости и единственности минимума $\delta(\varepsilon) := \inf\big\{g(x)+g(y) : g(x+y) \geq \varepsilon\big\} > 0$. Следовательно, если $g(x+y) \geq \varepsilon$, то

$$g(x+y) \leq \left(\frac{A}{\delta(\varepsilon)} + B\right)\big(g(x)+g(y)\big) := B_\varepsilon\big(g(x)+g(y)\big).$$

Таким образом, $g(x+y) \leq \max\{\varepsilon, B_\varepsilon\big(g(x)+g(y)\big)\}$. Заметим, что $B_\varepsilon \geq 1$, поэтому

$$g(x+y)+\varepsilon \leq 2\varepsilon + B_\varepsilon\big(g(x)+g(y)\big) \leq$$
$$\leq B_\varepsilon\big(g(x)+\varepsilon+g(y)+\varepsilon\big).$$

Отсюда, аналогично Лемме 3.1, следует, что найдется $q_\varepsilon \geq 1$, для которого

$$\big(g(x+y)+\varepsilon\big)^{1/q_\varepsilon} \leq \big(g(x)+\varepsilon\big)^{1/q_\varepsilon} + \big(g(y)+\varepsilon\big)^{1/q_\varepsilon}.$$

Рассмотрим произвольное $0 < t \leq 1/2$. В силу выпуклости $g(\cdot)$,

$$g(x+y) \leq (1-t)g(x) + tg\big(x+\frac{y}{t}\big) \leq$$
$$\leq g(x) + t\Big[\varepsilon + B_\varepsilon\big(g(x)+g\big(\frac{y}{t}\big)\big)\Big] \leq$$
$$\leq \frac{\varepsilon}{2} + (1+tB_\varepsilon)g(x) + tB_\varepsilon g\big(\frac{y}{t}\big).$$

Выберем $0 < r \leq 1/2$ так, что $rB_\varepsilon < \varepsilon$ и $rB_\varepsilon \max_{|z|=1} g(z) < \varepsilon/2$. Тогда для $0 < |y| \leq r$ выполняется

$$g(x+y) \leq \frac{\varepsilon}{2} + (1+|y|B_\varepsilon)g(x) + |y|B_\varepsilon g\big(\frac{y}{|y|}\big) \leq$$
$$\leq \varepsilon + (1+\varepsilon)g(x).$$

В то же время, аналогично первому пункту леммы, найдется такое $L > 0$, что для всех $|z| > r$ выполняется $g(2z) \leq Lg(z)$. Выберем $k \in \mathbb{N}$ так, чтобы $2^{-k}B_\varepsilon < \varepsilon$. Тогда

$$g(x+y) \leq \frac{\varepsilon}{2} + (1+2^{-k}B_\varepsilon)g(x) + 2^{-k}B_\varepsilon g(2^k y) \leq$$
$$\leq \varepsilon + (1+\varepsilon)g(x) + 2^{-k}B_\varepsilon L^k g(y) :=$$
$$:= \varepsilon + (1+\varepsilon)g(x) + C_\varepsilon g(y). \quad \square$$

***Теорема* 3.7.** *Для любого $\varepsilon > 0$ и всех мер $\mu, \nu, \lambda \in \mathscr{P}(\mathbb{R}^d)$ выполняются неравенства:*

$$J(\mu,\nu) \leq \max\{\varepsilon, B_\varepsilon\big(J(\mu,\lambda)+J(\lambda,\nu)\big)\},$$
$$\big(J(\mu,\nu)+\varepsilon\big)^{1/q_\varepsilon} \leq \big(J(\mu,\lambda)+\varepsilon\big)^{1/q_\varepsilon} + \big(J(\lambda,\nu)+\varepsilon\big)^{1/q_\varepsilon},$$

$$J(\mu,\nu) \leq \varepsilon + (1+\varepsilon)J(\mu,\lambda) + C_\varepsilon J(\lambda,\nu),$$
$$J(\mu,\nu) \leq \varepsilon + (1+\varepsilon)J(\lambda,\nu) + C_\varepsilon J(\mu,\lambda).$$

Доказательство. Аналогично Теореме 3.3. $\square$

Это условие является естественным в том смысле, что без его выполнения расстояние Монжа-Канторовича является достаточно плохой мерой близости — оно не имеет разумной “аддитивной” оценки сверху.

***Утверждение* 3.8.** *Пусть Предположение 3.2 не выполнено. Тогда существуют такие последовательности мер $\{\mu_n, \nu_n, \lambda_n\}_{n\in\mathbb{N}}$, что $\lim J(\mu_n,\lambda_n) = \lim J(\lambda_n,\nu_n) = 0$, а $J(\mu_n,\nu_n) \equiv \infty$.*

Доказательство. Рассмотрим точки $x_n, y_n \in \mathbb{R}^d$ такие, что $g(x_n+y_n) > n\big(1+g(x_n)+g(y_n)\big)$. Определим точки $z_n \in \mathbb{R}^d$ так, чтобы выполнялось $g(x_n+y_n) < g(z_k + x_k + y_k - z_n)$ и $g(x_n+y_n) < g(z_n + x_n + y_n - z_k)$ для всех $k, n \geq 0$, считая $x_k = y_k = z_k := 0$. Можно подобрать такие веса $\alpha_n \geq 0$, $\sum_{n=1}^\infty \alpha_n = 1$, что $\sum_n \alpha_n g(x_n+y_n) = \infty$, и $\sum_n \alpha_n g(x_n) < \infty$, $\sum_n \alpha_n g(y_n) < \infty$. Определим меры

$$\mu_n := \sum_{k=n}^\infty \alpha_k \delta_{z_k+x_k+y_k} + (1-\sum_{k=1}^{n-1}\alpha_k)\delta_0,$$
$$\lambda_n := \sum_{k=n}^\infty \alpha_k \delta_{z_k+x_k} + (1-\sum_{k=1}^{n-1}\alpha_k)\delta_0,$$
$$\nu_n := \sum_{k=n}^\infty \alpha_k \delta_{z_k} + (1-\sum_{k=1}^{n-1}\alpha_k)\delta_0.$$

Как несложно видеть, из выбора последовательности $\{z_n\}_{n\in\mathbb{N}}$ следует, что $J(\mu_n,\nu_n) = \sum_{k=n}^\infty g(x_k+y_k) \equiv \infty$, при этом $J(\mu_n,\lambda_n) \leq \sum_{k=n}^\infty g(y_k) \to 0$, $J(\lambda_n,\nu_n) \leq \sum_{k=n}^\infty g(x_k) \to 0$. $\square$

Заметим, что пока мы не накладывали никаких ограничений на связь $g(x)$ и $g(-x)$. Чтобы имело место Предположение 2.2, потребуем выполнения следующего условия.

Предположение 3.3. Существуют константы $A, B \geq 0$ такие, что при любых $x, y \in \mathbb{R}^d$

$$g(x-y) \leq A + B\big(g(x)+g(y)\big).$$

Тогда без о. о. можно считать, что с теми же константами для всех $x, y \in \mathbb{R}^d$ выполняется

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

В частности, отсюда следует Предположение 3.2. Также, имеют место результаты, аналогичные Лемме 3.6 и Теореме 3.7 с различными перестановками аргументов. Например, выполняется Предположение 2.2 со всеми опирающимися на него результатами, и следующие неравенства:

$$J(\mu,\nu), J(\nu,\mu) \leq \varepsilon \vee B_\varepsilon\big(J(\mu,\lambda)+J(\nu,\lambda)\big),$$
$$J(\mu,\nu), J(\nu,\mu) \leq \varepsilon + (1+\varepsilon)J(\mu,\lambda) + C_\varepsilon J(\nu,\lambda),$$
$$J(\mu,\nu), J(\nu,\mu) \leq \varepsilon + (1+\varepsilon)J(\lambda,\mu) + C_\varepsilon J(\lambda,\nu).$$

Таким образом, как видим, Предположение 2.2 из предыдущего раздела не является слишком ограничительным и выполняется для достаточно широкого класса функций, имеющих большое практическое значение.

4. Барицентры Фреше

В разд. 2 было показано, что пространство вероятностных мер с топологий, порожденной транспортным функционалом, обладает достаточно хорошими свойствами. В этом разделе будет определен барицентр мер, то есть "среднее" в пространстве $\mathscr{P}(X)$ относительно широкого класса транспортных функционалов, обобщающий конструкцию для пространства 2-Вассерштейна из статьи [3]. Также будет показано, что барицентры обладают статистической состоятельностью и непрерывностью, т.е. что такое усреднение является достаточно разумным.

4.1. Обобщенное среднее в $\mathscr{P}(X)$

Определение 4.1. Пусть даны меры $\mu_1, \ldots, \mu_n \in \mathscr{P}(X)$, и веса $\alpha_1, \ldots, \alpha_n > 0$, $\sum_{i=1}^n \alpha_i = 1$. *Барицентром Фреше набора* $(\mu_i, \alpha_i)_{1 \le i \le n}$ будем называть решение задачи

$$\sum_{i=1}^n \alpha_i J(\mu_i, \nu) \to \inf_{\nu \in \mathscr{P}(X)}$$

(если $\inf_\nu \sum_{i=1}^n \alpha_i J(\mu_i, \nu) < \infty$), обозначая его $\mathrm{bar}(\mu_i, \alpha_i)_{1 \le i \le n}$.

Определение 4.2. Пусть дано борелевское распределение P на пространстве $\mathscr{P}(X)$. *Барицентром Фреше распределения* P будем называть решение задачи

$$\int J(\mu, \nu)\, dP(\mu) \to \inf_{\nu \in \mathscr{P}(X)} \tag{1}$$

(опять же, если инфимум конечен), обозначая его $\mathrm{bar}(P)$.

Покажем, что при выполнении Предположения 2.1 всегда существует барицентр распределения мер.

***Теорема* 4.1.** *Пусть дано распределение* $P \in \mathscr{P}\big(\mathscr{P}(X)\big)$, *и* $\inf_\nu \int J(\mu, \nu)\, dP(\mu) < \infty$. *Если выполнено Предположение 2.1, то существует барицентр распределения* P. *Также, если* $\{\nu_n\}_{n \in \mathbb{N}}$ *— минимизирующая последовательность для задачи* (1), *т. е.* $\int J(\mu, \nu_n)\, dP(\mu) \to \inf_\nu \int J(\mu, \nu)\, dP(\mu)$, *то у нее есть слабо сходящаяся подпоследовательность, предел которой является барицентром* P. *В частности, множество барицентров любого распределения слабо компактно.*

Если дополнительно выполняется Предположение 2.2, то любая минимизирующая последовательность имеет сильный частичный предел, являющийся барицентром, и множество барицентров — сильный компакт.

Доказательство. Зафиксируем $\varepsilon > 0$ и шар $B = \overline{B}_r^c(x_0)$ такой, что $\mu(X \setminus B) < \varepsilon/2$ на некотором множестве $A \in \mathscr{P}(X)$, $P(A) > 1/2$. Рассмотрим $R > r$ и произвольную меру λ_R, для которой $\lambda_R\big(X \setminus B_R^c(x_0)\big) \ge \varepsilon$. Тогда

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

при $R \to \infty$. А так как $\lim \int J(\mu, \nu_n)\, dP(\mu) = \inf_\nu \int J(\mu, \nu)\, dP(\mu) < \infty$, то отсюда следует, что $\{\nu_n\}_{n \in \mathbb{N}}$ — плотная последовательность (аналогично Лемме 2.10), поэтому у нее есть слабо сходящаяся подпоследовательность $\nu_{n_k} \rightharpoonup \nu^*$.

По лемме Фату и полунепрерывности снизу,

$$\int J(\mu, \nu^*)\, dP(\mu) \le \int \liminf J(\mu, \nu_{n_k})\, dP(\mu) \le$$
$$\le \liminf \int J(\mu, \nu_{n_k})\, dP(\mu) = \inf_\nu \int J(\mu, \nu)\, dP(\mu) < \infty.$$

Следовательно, $\nu^* = \mathrm{bar}(P)$ — барицентр распределения. Также, отсюда следует, что $J(\mu, \nu^*) = \liminf J(\mu, \nu_{n_k})$ для P-почти всех μ. Без переобозначения будем считать, что $J(\hat{\mu}, \nu^*) = \lim J(\hat{\mu}, \nu_{n_k}) < \infty$ для некоторой меры $\hat{\mu}$. Если выполнено Предположение 2.2, то по Теореме 2.18 $\nu_{n_k} \xrightarrow{J} \nu^*$. □

4.2. Состоятельность барицентров

Рассмотрим пространство $\mathscr{P}(X)$. Для распределений на нем можно ввести расстояние Монжа-Канторовича с ценовой функцией $J(\cdot,\cdot)$: пусть $P, P' \in \mathscr{P}\big(\mathscr{P}(X)\big)$, тогда

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

В дальнейшем считаем, что выполнены Предположения 2.1 и 2.2. В этом случае барицентр конечного набора мер $(\mu_i, \alpha_i)_{1 \le i \le n}$ существует тогда и только тогда, когда все они лежат в одном классе эквивалентности. Для существования барицентра распределения P необходимо, чтобы вероятность какого-то класса эквивалентности равнялась 1, т. е. $P\big(C(\mu)\big) = 1$ для какой-то меры μ. Поэтому зафиксируем произвольную меру $\mu_0 \in \mathscr{P}(X)$ и будем рассматривать пространство $C(\mu_0)$ с сильной топологией, которое, как было показано в Теореме 2.23, является пространством Радона. Очевидно, ценовая функция $J(\cdot,\cdot)$ является непрерывной и согласованной с топологией τ_J в силу ее определения. Кроме того, для $J(\cdot,\cdot)$ выполняется Предположение 2.2 (Теорема 2.14), поэтому для $\mathscr{P}\big(C(\mu_0)\big)$ имеет место критерий сильной сходимости и прочие результаты, не использующие Предположение 2.1, которое может и не выполняться, если пространство X не компактно.

Теперь мы хотим показать, что из сходимости распределений относительно $\mathscr{J}(\cdot,\cdot)$ следует сходимость их барицентров — результат, аналогичный Теореме 2 из [6] в случае пространства Вассерштейна. Также отсюда мы получим закон больших чисел для эмпирических барицентров, доказанный в статье [5] для квадратичной ценовой функции и мер с компактным носителем (Теорема 6.1).

***Теорема* 4.2** (состоятельность барицентров)**.** *Пусть дана последовательность распределений $\{P_n\}_{n\in\mathbb{N}} \subset \mathscr{P}\big(C(\mu_0)\big)$, $P_n \xrightarrow{\mathscr{J}} P$ для некоторого распределения P, у которого существует барицентр.*

Тогда, начиная с некоторого n, существуют барицентры $\nu_n := \mathrm{bar}(P_n)$. При этом любая подпоследовательность $\{\nu_n\}_{n\in\mathbb{N}}$ имеет частичный предел, который является барицентром распределения P. В частности, если существует единственный барицентр $\nu^* := \mathrm{bar}(P)$, то $\nu_n \xrightarrow{J} \nu^*$.

Доказательство. Т.к. $\int J(\mu,\nu^*)\,dP_n(\mu) \to \int J(\mu,\nu^*)\,dP(\mu) < \infty$ в силу Теоремы 2.18, то, начиная с некоторого n, выполняется $\int J(\mu,\nu^*)\,dP(\mu) < \infty$, поэтому существуют барицентры распределений P_n. Покажем, что $\{\nu_n\}_{n\in\mathbb{N}}$ — минимизирующая последовательность для $\int J(\mu,\cdot)\,dP(\mu)$. Действительно, с помощью Теоремы 2.14 для $\mathscr{J}(\cdot,\cdot)$ получаем

$$\limsup \int J(\mu,\nu_n)\,dP(\mu) = \limsup \mathscr{J}(P,\delta_{\nu_n}) =$$
$$= \limsup \mathscr{J}(P_n,\delta_{\nu_n}) = \limsup \int J(\mu,\nu_n)\,dP_n(\mu) \le$$
$$\le \lim \int J(\mu,\nu^*)\,dP_n(\mu) = \int J(\mu,\nu^*)\,dP(\mu) =$$
$$= \inf_{\nu} \int J(\mu,\nu)\,dP(\mu).$$

В силу Теоремы 4.1, отсюда следует доказываемое утверждение. □

***Следствие* 4.3** (непрерывность)**.** *Пусть даны последовательности мер $\mu_i^n \xrightarrow{J} \mu_i$ и неотрицательные веса $\alpha_i^n \to \alpha_i$, $\sum_i \alpha_i^n = 1$, при $1 \le i \le m$. Пусть все $\mu_i \in C(\mu_1)$. Тогда у $\big\{\mathrm{bar}(\mu_i^n,\alpha_i^n)_{1\le i\le m}\big\}_{n\in\mathbb{N}}$ есть частичный предел, являющийся барицентром набора $(\mu_i,\alpha_i)_{1\le i\le m}$.*

Доказательство. Пусть $P_n := \sum_i \alpha_i^n \delta_{\mu_i^n}$, $P := \sum_i \alpha_i \delta_{\mu_i}$. Заметим, что $J(\mu_i^n,\mu_j) \to J(\mu_i,\mu_j)$ и $\max_{i,j} J(\mu_i,\mu_j) < \infty$. Поэтому

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

Таким образом, выполнены условия теоремы. □

***Следствие* 4.4** (закон больших чисел)**.** *Пусть $\{\boldsymbol{\mu}_n\}_{n\in\mathbb{N}}$ — последовательность i. i. d. случайных величин с распределением P_μ, для которого существует барицентр $\nu^* := \mathrm{bar}(P_\mu)$. Тогда почти наверное любая подпоследовательность эмпирических барицентров $\boldsymbol{\nu}_n := \mathrm{bar}(\boldsymbol{\mu}_i, 1/n)_{1\le i\le n}$ имеет частичный предел, который является барицентром распределения P_μ.*

Доказательство. Рассмотрим эмпирические меры $\boldsymbol{P}_n := \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{\mu}_i}$. Очевидно, эмпирический барицентр $\boldsymbol{\nu}_n := \mathrm{bar}(\boldsymbol{\mu}_i, 1/n)_{1\le i\le n} = \mathrm{bar}(\boldsymbol{P}_n)$. Так как $P_\mu \in \mathscr{P}\big(C(\nu^*)\big)$, то все $\boldsymbol{\mu}_i \in C(\nu^*)$ п. н., поэтому $\boldsymbol{P}_n \in \mathscr{P}\big(C(\nu^*)\big)$. Из усиленного ЗБЧ следует, что почти наверное

$$\mathscr{J}(\boldsymbol{P}_n,\delta_{\nu^*}) = \frac{1}{n}\sum_i J(\boldsymbol{\mu}_i,\nu^*) \to \mathbb{E} J(\boldsymbol{\mu},\nu^*) = \mathscr{J}(P_\mu,\delta_{\nu^*})$$

и $\boldsymbol{P}_n \rightharpoonup P_\mu$, так как топология слабой сходимости в $\mathscr{P}\big(C(\nu^*)\big)$ обладает счетной базой в силу сепарабельности $C(\nu^*)$. Тогда по Теореме 2.18 $\boldsymbol{P}_n \xrightarrow{\mathscr{J}} P_\mu$ п. н., т. е. выполняются условия теоремы. □

Заметим, что все результаты из этого раздела верны и для пространства $\mathscr{P}(X)$ вместо $\mathscr{P}\big(\mathscr{P}(X)\big)$: это следует из того, что любую меру $\nu \in \mathscr{P}(X)$ можно отождествить с распределением $P := (\delta_x)_{\#}\nu \in \mathscr{P}\big(\mathscr{P}(X)\big)$, и $J(\delta_x,\delta_y) = c(x,y)$ для всех $x,y \in X$.

Список литературы

[1] В. И. Богачев, *Основы теории меры*, том 2, НИЦ «РХД», Москва-Ижевск, 2003.

[2] Д. Ю. Бураго, Ю. Д. Бураго и С. В. Иванов, *Курс метрической геометрии*, Институт компьютерных исследований, Москва-Ижевск, 2004.

[3] M. Agueh, G. Carlier, *Barycenters in the Wasserstein space*, SIAM Journal on Mathematical Analysis, 43(2):904–924, 2011.

[4] L. Ambrosio, N. Gigli and G. Savare, *Gradient Flows*, Birkhäuser, Basel, 2008.

[5] J. Bigot, T. Klein, *Consistent Estimation of a Population Barycenter in the Wasserstein space*, 2015, URL: `http://arxiv.org/abs/1212.2562v4`.

[6] T. Le Gouic and J.-M. Loubes, *Existence and Consistency of Wasserstein Barycenters*, 2015, URL: `http://arxiv.org/abs/1506.04153v1`.

[7] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Birkhäuser, Basel, 2015.

[8] C. Villani, *Optimal Transport, Old and New*. Springer-Verlag, Berlin-Heidelberg, 2009.