Math-Net.Ru

Общероссийский математический портал

В. И. Богачев, А. В. Колесников, Задача Монжа–Канторовича: достижения, связи и перспективы, *УМН*, 2012, том 67, выпуск 5, 3–110

DOI: <https://doi.org/10.4213/rm9490>

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

Параметры загрузки:

IP: 18.208.226.222

26 апреля 2023 г., 08:11:36

УДК 519.2+517.9

*Посвящается столетию со дня рождения
Леонида Витальевича Канторовича*

Задача Монжа–Канторовича: достижения, связи и перспективы

В. И. Богачев, А. В. Колесников

В статье дается обзор недавних исследований, связанных с задачей Монжа–Канторовича. Приведены основные результаты о существовании решений и их свойствах как в задаче Монжа об оптимальной транспортировке, так и в задаче Канторовича об оптимальном плане, а также результаты о связях между обеими задачами и случаях их равносильности. Рассказано о многообразных приложениях этих задач в нелинейном анализе, теории вероятностей, дифференциальной геометрии.

Библиография: 193 названия.

Ключевые слова: задача Монжа, задача Канторовича, метрика Канторовича–Рубинштейна, оптимальная транспортировка, транспортное неравенство.

СОДЕРЖАНИЕ

Введение	4
Глава 1. Задача Канторовича	11
1.1. Метрики Канторовича и нормы Канторовича–Рубинштейна	11
1.2. Существование оптимальных планов	19
1.3. Двойственность Канторовича	23
Глава 2. Задача Монжа	26
2.1. Существование и единственность оптимальных отображений	26
2.2. Регулярность решений и априорные оценки	31
2.2.1. Принцип максимума	32
2.2.2. Регулярность решений уравнения Монжа–Ампера	33
2.2.3. Классический подход	34
2.2.4. Регулярность в пространствах Гёльдера и Соболева	36
2.2.5. Глобальные оценки и приложения	37
2.3. Связи с задачей Канторовича и приближенные решения	39

Работа выполнена при поддержке РФФИ (гранты № 10-01-00518, 11-01-90421-Укр-ф-а, 11-01-12104-офи-м, 12-01-33009) и программы SFB 701 при Билефельдском университете. В данной научной работе использованы результаты, полученные в ходе выполнения проекта № 11-01-0175, реализованного в рамках программы “Научный фонд НИУ ВШЭ” в 2012–2013 гг.

Глава 3. Приложения	47
3.1. Изопериметрические неравенства и неравенство Брунна–Минковского	47
3.2. Неравенства Соболева и их обобщения	49
3.2.1. Классические неравенства Соболева	49
3.2.2. Логарифмические неравенства Соболева	50
3.2.3. Равномерно выпуклые потенциалы	54
3.2.4. Ограниченность снизу второй производной и интегрируемость	55
3.3. Транспортные неравенства	56
3.4. Неравенства концентрации и большие отклонения	61
3.5. Иерархия неравенств	62
3.6. Геодезические в пространстве мер и градиентные потоки	63
3.6.1. Геодезические в метрике Канторovichа	63
3.6.2. Формула Бенаму–Бренье	65
3.6.3. Динамический план переноса массы	66
3.6.4. Выпуклые функционалы	67
3.6.5. Дополнительные свойства пространства $\mathcal{P}$	69
3.6.6. Субдифференциалы	70
3.6.7. Градиентные потоки. Формализм Отто	71
3.6.8. Приложения к уравнениям математической физики	73
3.6.9. Риманов объем на $\mathcal{P}^2$	74
3.7. Геометрические приложения	75
3.7.1. Классические задачи дифференциальной геометрии	75
3.7.2. Задача Монжа–Канторovichа на римановом многообразии. Существование и единственность	77
3.7.3. Задача Монжа–Канторovichа на римановом многообразии. Замена переменных, тензор Риччи и геометрические неравенства	79
3.7.4. Пространства неотрицательной кривизны	82
3.7.5. Геометрические потоки: потоки гауссовой кривизны и параболические уравнения	85
3.7.6. Геометрические потоки: потоки Риччи	87
3.7.7. Транспортные сети	89
3.8. Бесконечномерные задачи Монжа–Канторovichа	90
3.8.1. Существование оптимальных отображений на пространстве Винера	90
3.8.2. Формула замены переменной	97
3.9. Некоторые другие виды преобразований мер	98
3.9.1. Треугольные отображения	98
3.9.2. Преобразования Мозера, полугруппы, транспортные уравнения	100
Список литературы	100

Введение

В нынешнем году столетия со дня рождения Леонида Витальевича Канторovichа исполняется и 70 лет с момента выхода одной из его самых коротких, но несомненно выдающихся работ [1] (с еще более кратким продолжением [2]),

идеям которой была суждена долгая жизнь; некоторые из этих идей были развиты в двух совместных статьях Л. В. Канторовича и Г. Ш. Рубинштейна [3], [4]. Как отмечал А. М. Вершик [5], “красота и естественность основной постановки вопроса, фундаментальность основной теоремы (критерия оптимальности) и, наконец, богатство приложений (часть из них реализована, но обнаруживаются все новые, в областях, открывающихся только сейчас) – все это позволяет отнести эти работы к классике математики XX века”.

Цель нашего обзора – рассказать о современном состоянии исследований, идейно связанных как с работой [1], так и со значительно более старой задачей Г. Монжа [6], о которой Л. В. Канторович узнал уже после выхода [1] (см. [2]); интересные материалы биографического характера собраны в [7], [8], а в юбилейный год можно ожидать и новых публикаций.

Хотя этой тематике посвящен целый ряд основательных обзоров и монографий (см. [9]–[17]), представляется разумным вкратце пересказать известную историю ее возникновения и развития. В современных терминах задача Монжа состоит в следующем (в общем виде такая постановка принадлежит А. М. Вершику [18]). Даны два вероятностных пространства $(X, \mathcal{A}, \mu)$ и $(Y, \mathcal{B}, \nu)$ и неотрицательная измеримая функция h на $X \times Y$, называемая функцией стоимости (часто обозначаемая через c). Требуется найти отображение $T: X \rightarrow Y$, измеримое относительно пары $(\mathcal{A}, \mathcal{B})$, т. е. $T^{-1}(B) \in \mathcal{A}$ при всех $B \in \mathcal{B}$, переводящее меру μ в ν и доставляющее минимум выражению

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

среди всех таких отображений. Условие, что μ переходит в ν , означает, что ν совпадает с образом меры μ при отображении T , который задается формулой $\mu \circ T^{-1}(B) := \mu(T^{-1}(B))$, $B \in \mathcal{B}$. Если некоторое $T \in T(\mu, \nu)$ дает минимум, то это T называют *оптимальным отображением* меры μ в ν или *оптимальной транспортной*.

В интерпретации самого Монжа речь шла о переносе грунта для ведения строительных работ, т. е. обе меры были обычными объемами, функция – обычным расстоянием, а минимизировалась выполненная работа (при условии, что работа по переносу элемента массы ΔV на расстояние Δr есть $\Delta V \Delta r$). Внешне задача выглядит как прикладная, но довольно ясно, что она таковой не была, хотя в то время французские академики привлекались и к решению самых настоящих прикладных задач совершенно конкретного практического характера (см. [19]). Конечно, это не единственная возможная “экономическая” формулировка транспортной задачи. Например, речь может идти о перевозке грузов со складов по заданным адресам или вообще распределении каких-либо ресурсов (в некоторых статьях модельный пример – доставка булочек и круассанов из пекарен в парижские кафе), а дискретный вариант задач Монжа и Канторовича возникает в столь многих современных областях, что даже не требует изобретения искусственных приложений. Гаспар Монж был весьма универсальным ученым, внесшим заметный вклад в начертательную геометрию, дифференциальную геометрию, инженерное дело, организацию науки и высшего образования, он был активным деятелем Французской революции (морским министром, подписавшим решение суда о казни Людовика XVI), соратником Наполеона,

проведшим его в члены Академии наук (так что и в те годы, как и сейчас, для избрания в члены академии каких-либо научных достижений могло и не требоваться) и участником его египетского похода. Среди учеников Монжа были такие известные ученые, как О. Коши, С. Д. Пуассон, Ж. Менье, Л. Карно (цикл Карно – изобретение его сына С. Карно), Ж.-В. Понселе, Г.-Г. Кориолис; позже, после падения Наполеона и “французской перестройки”, он был изгнан из Академии (проворно “перестроившимися” коллегами), но за решение поставленной им задачи оптимального переноса масс для конкретной функции $h(x, y) = |x - y|$ много позже Академия наук объявляла призы. Правда, само существование решения подразумевалось, а вопрос состоял в выяснении некоторых его геометрических свойств. Эту проблему решил П. Аппель [20], получивший такой приз. Однако и в его решении не давалось доказательства существования оптимального отображения, а были установлены лишь определенные свойства такого отображения. Сама же проблема существования не только не была решена, но даже и точно сформулирована вплоть до 70-х годов XX века, когда она была явно приведена в современных терминах А. М. Вершиком в упомянутой статье [18] и появилась весьма важная во многих отношениях и вошедшая в классику теории меры работа В. Н. Судакова [21], поставившая, как полагали, точку. Однако спустя 20 лет в ней нашли пробел, который, к счастью, в итоге удалось закрыть (см. [9], [22]–[24] и более точные комментарии ниже). Можно сказать, что лишь к началу нынешнего тысячелетия не только была решена специальная задача Монжа, но и были осознаны многие связанные с ней тонкости, невидимые ранее. В последнее десятилетие исследования в этой области значительно активизировались. Расширение этих исследований было связано как с рассмотрением различных задач, возникающих при том или ином выборе функций стоимости h и классов преобразуемых мер μ и ν , так и с выявлением связей такого рода задач с самыми разными направлениями нелинейного анализа, геометрии, уравнений с частными производными, стохастического анализа. Вот эти открывшиеся связи и привлекают к рассматриваемым вопросам столь много специалистов из столь разных областей.

Однако все это вряд ли можно было предвидеть, когда в тяжелую военную осень 1942 г. майор Л. В. Канторович, заведовавший кафедрой математики в ВИТУ ВМФ и участвовавший в прикладных военных исследованиях, сдал в печать свою заметку [1]. К тому времени он уже был довольно известным математиком, опубликовавшим несколько работ по теории функций вещественного и комплексного переменного, серьезные исследования по дескриптивной теории множеств, полупорядоченным векторным пространствам и линейным задачам функционального анализа, примыкающим к теории меры и интегральным представлениям линейных функционалов. Несомненное влияние на творчество Канторовича, особенно по обсуждаемой теме, оказало то, что он был учеником Г. М. Фихтенгольца, замечательного математика и педагога, крупного специалиста по теории интеграла (о его работах см. [25]; другим учителем Канторовича был В. И. Смирнов). В указанной заметке Канторовича ставится очень близкая к проблеме Монжа задача (даже со сходными примерами “прикладного” характера о переносе грунтов), но с одним принципиально важным нюансом: в задаче Канторовича вместо поиска отображения T (“оптимальной транспортировки”) предлагается нахождение лишь “оптимального плана

транспортировки”, т. е. такой вероятностной меры $\hat{\sigma}$ на $(X \times Y, \mathcal{A} \otimes \mathcal{B})$, что ее проекции на X и Y суть μ и ν соответственно, причем $\hat{\sigma}$ доставляет минимум выражению

$$K(\mu, \nu, \sigma) = K_h(\mu, \nu, \sigma) := \int_{X \times Y} h(x, y) \sigma(dx dy)$$

по всем вероятностным мерам σ из класса $\Pi(\mu, \nu)$ вероятностных мер на произведении $(X \times Y, \mathcal{A} \otimes \mathcal{B})$, дающих μ и ν при проектировании на X и Y . Стоит сказать, что заметка [1] была частью проекта Канторовича, начатого им в 1938 г. и изложенного в брошюре [26], с которой фактически и начинается современное линейное программирование; транспортная задача – одна из намеченных в ней тем. В отличие от сложной нелинейной задачи Монжа задача Канторовича линейна: ищется минимум линейного функционала на выпуклом множестве, причем сам он рассмотрел случай, когда X и Y – метрические компакты, что делает $\Pi(\mu, \nu)$ компактом в слабой топологии, в которой сходимост мер ν_α к мере ν есть сходимость интегралов по ним от каждой ограниченной непрерывной функции. Конечно, и функционал $\sigma \mapsto K(\mu, \nu, \sigma)$ оказывается непрерывным в указанной топологии, если непрерывна функция h . Таким образом, здесь при любой непрерывной функции h есть решения, причем все они достигаются на крайних точках компакта $\Pi(\mu, \nu)$ (т. е. точках, не являющихся внутренними ни для каких отрезков с концами в $\Pi(\mu, \nu)$). В общем случае решением считается мера из $\Pi(\mu, \nu)$, на которой минимум достигается; она называется *оптимальной* или *оптимальным планом*. Для некомпактных пространств ситуации несколько сложнее, но и здесь имеется много достаточных условий существования решений, в то время как задача Монжа может быть неразрешимой даже для довольно простых функций и мер на компактах в $\mathbb{R}^n$. Надо сказать, что Канторович в работе [2] назвал задачу Монжа частным случаем своей задачи и писал, что из решения последней легко извлечь решение первой. В общем случае это неверно, но если задача Монжа для непрерывной функции стоимости имеет решение T , то мера σ на графике T , равная образу меры μ при отображении $x \mapsto (x, T(x))$, будет решением задачи Канторовича. Отметим, что задачу Канторовича можно трактовать как “задачу Монжа с многозначным отображением”, представив план σ в виде $\sigma(dx dy) = \sigma^x(dy) \mu(dx)$ посредством условных мер σ^x на Y (см. ниже), что приводит к поиску оптимального отображения $x \mapsto \sigma^x$ со значениями в пространстве вероятностных мер на Y ; в этой картине существование точечной оптимальной транспортировки означает возможность выбора дираковских условных мер σ^x . Здесь возникают задачи о полиморфизмах (см. [27]). В некоторых важных частных случаях действительно удается по решению задачи Канторовича построить решение задачи Монжа, но в общем это довольно большая редкость. Тем не менее связь задач Монжа и Канторовича оказывается удивительно тесной, так что общепринятым стал термин “задача Монжа–Канторовича” (впервые такое сочетание в названии статьи нам встретилось в работе В.Л. Левина [28]). Для обеих же задач полезно ввести величины

$$M(\mu, \nu) = M_h(\mu, \nu) := \inf\{M_h(\mu, \nu, T) : T \in T(\mu, \nu)\},$$

где $T(\mu, \nu)$ – класс всех измеримых отображений, переводящих μ в ν , и

$$K(\mu, \nu) = K_h(\mu, \nu) := \inf\{K_h(\mu, \nu, \sigma) : \sigma \in \Pi(\mu, \nu)\}.$$

Имея дело с какой-то одной функцией стоимости h , указание на нее в этих обозначениях будем опускать. Однако индекс h будет важен при сравнении функций стоимости. В общем случае эти инфимумы не являются минимумами, причем имеет место оценка

$$K(\mu, \nu) \leq M(\mu, \nu).$$

Наконец, обозначим через Φ_h множество пар функций (φ, ψ) , $\varphi: X \rightarrow \mathbb{R}^1$, $\psi: Y \rightarrow \mathbb{R}^1$, измеримых относительно $\mathcal{A}$ и $\mathcal{B}$ соответственно и удовлетворяющих неравенству

$$\varphi(x) + \psi(y) \leq h(x, y), \quad x \in X, \quad y \in Y.$$

Для этого множества Канторович рассмотрел так называемую двойственную задачу нахождения величины (со значениями в $[0, +\infty]$)

$$J(\mu, \nu) = J_h(\mu, \nu) := \sup\left\{\int_X \varphi d\mu + \int_Y \psi d\nu, (\varphi, \psi) \in \Phi_h\right\}.$$

Эта задача, навеянная его исследованиями в линейном программировании и применением там двойственных задач, играет важную роль в обсуждаемом круге вопросов. Следует оговорить случай, когда при рассмотрении каких-либо инфимумов все возникающие величины бесконечны: естественно, тогда и инфимум считается бесконечным. Тем самым в задаче Канторовича, двойственной задаче и задаче Монжа допускаются бесконечные значения. Например, если интеграл от h бесконечен для всех $\sigma \in \Pi(\mu, \nu)$, то полагаем $K(\mu, \nu) = +\infty$ и считаем любую меру из $\Pi(\mu, \nu)$ оптимальной. Поскольку очевидным образом

$$\int_X \varphi d\mu + \int_Y \psi d\nu \leq \int_{X \times Y} h d\sigma$$

для всех $(\varphi, \psi) \in \Phi_h$ и $\sigma \in \Pi(\mu, \nu)$, то имеет место оценка

$$J(\mu, \nu) \leq K(\mu, \nu) \leq M(\mu, \nu).$$

Оказывается, что при весьма широких предположениях эти неравенства на самом деле являются равенствами, что и будет основным предметом обсуждения в первой главе. Во второй главе мы обратимся к собственно задаче Монжа, изложим основные результаты о существовании ее решений и свойствах этих решений (в частности, дифференциальных). Существенное отличие задачи Монжа проявляется и в том, что от функции h здесь требуется значительно больше, чем в задаче Канторовича, причем эти требования весьма ограничительны даже для хороших мер на пространстве $\mathbb{R}^d$. Пока можно констатировать, что хорошо изучен лишь случай функции $h(x, y) = |x - y|^2$ и некоторых достаточно регулярных функций с условиями типа выпуклости/вогнутости. Последняя глава посвящена разнообразным приложениям

и связям задачи Монжа–Канторовича. В ней речь идет о различных нелинейных неравенствах и вариационных проблемах (изопериметрия, логарифмические неравенства Соболева и т. п.), геометрических потоках на конечномерных римановых многообразиях и градиентных потоках на бесконечномерных многообразиях вероятностных мер, транспортных уравнениях, бесконечномерных вариантах задачи Монжа и многом другом; наконец, кратко упоминается о некоторых других интересных преобразованиях мер (своего рода конкурентах оптимальных отображений). Конечно, объем нашего обзора не позволяет углубиться в детали, поэтому многие направления современных исследований лишь упомянуты со ссылками (впрочем, и во в 10 раз более объемной весьма основательной монографии С. Виллани [17] подробно обсуждаются лишь избранные задачи), но такая зарисовка общей панорамы и есть одна из целей обзора, тесно связанная с другой важной целью: привлечь внимание наших начинающих математиков к этой интересной и многообещающей области, созданной 50–70 лет назад в работах Л. В. Канторовича, а также в оказавшихся тесно с ними связанными работах других наших выдающихся исследователей, таких как А. Н. Колмогоров, А. Д. Александров, В. А. Рохлин, А. В. Погорелов, Р. Л. Добрушин, Ю. В. Прохоров.

Обозначения и терминология. Далее символом $\mathcal{B}(X)$ обозначается σ -алгебра борелевских множеств на топологическом пространстве X , т. е. наименьшая σ -алгебра, содержащая все открытые множества, а символом $C_b(X)$ – множество непрерывных функций на X . Ниже рассматриваются только хаусдорфовы (отделимые) пространства. Напомним, что борелевской мерой называется вещественная счетно-аддитивная функция μ на $\mathcal{B}(X)$; такая функция единственным образом записывается в виде $\mu = \mu^+ - \mu^-$, где μ^+ и μ^- – неотрицательные борелевские меры, сосредоточенные на непересекающихся множествах (называемые положительной и отрицательной частями μ); мера $|\mu| = \mu^+ + \mu^-$ называется полной вариацией меры μ , а величина $\|\mu\| = |\mu|(X)$ – вариацией или вариационной нормой μ . Борелевская мера μ называется радоновской, если для всякого $B \in \mathcal{B}(X)$ и всякого $\varepsilon > 0$ найдется такой компакт K , что $|\mu|(B \setminus K) < \varepsilon$. Множество $B \in \mathcal{B}(X)$ называется атомом меры μ , если $\mu(B) > 0$ и всякое множество $A \in \mathcal{B}(X)$, входящее в множество B , имеет меру либо 0, либо $\mu(B)$. Наименьшее замкнутое множество полной меры называется топологическим носителем меры μ и обозначается через $\text{supp}(\mu)$. Такой носитель существует для всякой меры Радона на топологическом пространстве, а также для всякой борелевской меры на сепарабельном метрическом пространстве.

Положим $a \wedge b = \min(a, b)$. Пусть δ_x – мера Дирака в x .

При обсуждении некоторых вопросов очень полезно понятие суслинского пространства – хаусдорфова пространства, являющегося непрерывным образом полного сепарабельного метрического пространства (см. [29; гл. 6, 7]). На таких пространствах все борелевские меры радоновы. В полном сепарабельном метрическом пространстве все борелевские множества являются суслинскими, но в несчетных пространствах всегда есть неборелевские суслинские множества. Для обсуждаемых вопросов важным преимуществом суслинских множеств перед борелевскими является то, что непрерывный и даже борелевский образ суслинского множества есть суслинское множество (образ борелевского

множества может не быть борелевским даже для бесконечно дифференцируемой функции на прямой).

Многие аспекты задачи Монжа–Канторовича связаны с условными мерами, построение и эффективное использование которых восходит к работам А. Н. Колмогорова, В. А. Рохлина и других классиков (см. гл. 10 в [29]). Здесь укажем лишь, что для суслинских пространств X и Y всякая радоновская вероятностная мера μ на $X \times Y$ допускает условные меры μ^x , $x \in X$, на Y , т. е. такие вероятностные радоновские меры, что для всякого борелевского множества $B \in X \times Y$ функция $y \mapsto \mu^x(B_x)$, где $B_x = \{y \in Y : (x, y) \in B\}$, является борелевской на X , причем

$$\mu(B) = \int_Y \mu^x(B_x) \mu_X(dx),$$

где μ_X – образ μ при проекции на X . Меры $\mu^x(dy)$ обозначают также через $\mu(dy|x)$. В частности, полагая $B = X \times C$, $C \in \mathcal{B}(Y)$, получаем, что функция $x \mapsto \mu^x(C)$ является борелевской. Для более общих пространств само понятие условных мер сохраняет смысл, но существуют они не всегда.

Через $\mathcal{M}_r(X)$ обозначается пространство всех (в том числе знакопеременных) мер Радона на X , через $\mathcal{M}_r^+(X)$ – множество неотрицательных мер Радона, а через $\mathcal{P}_r(X)$ – его подмножество, состоящее из вероятностных мер. Для большинства обсуждаемых далее вопросов можно считать, что речь идет о полных сепарабельных метрических пространствах, а на таких пространствах все борелевские меры радоновы, поэтому будут использоваться символы $\mathcal{M}(X)$ и $\mathcal{P}(X)$ без индекса r . В общем случае бывает полезно рассматривать также бэровскую σ -алгебру $\mathcal{B}_b(X)$, порожденную множествами вида $\{f > 0\}$, где $f \in C_b(X)$; заданные на ней меры называются бэровскими. Ясно, что $\mathcal{B}_b(X) \subset \mathcal{B}(X)$, для метрических пространств имеет место равенство, но в общем случае включение строгое даже для компактов. Пространство всех бэровских мер обозначим через $\mathcal{M}_\sigma(X)$, а его подмножество вероятностных мер – через $\mathcal{P}_\sigma(X)$. На этом пространстве (тем самым и на $\mathcal{M}_r(X)$) вводится слабая топология, порождаемая полунормами

$$p_f(\mu) = \left| \int_X f d\mu \right|, \quad f \in C_b(X).$$

Сходимость последовательности (или направленности) мер в слабой топологии есть сходимость интегралов от каждой фиксированной функции из $C_b(X)$.

Согласно классической теореме Прохорова, если X – полное сепарабельное метрическое пространство, то следующие условия равносильны для последовательности вероятностных мер $\{\mu_n\}$ на X : 1) каждая ее подпоследовательность имеет слабо сходящуюся подпоследовательность; 2) она равномерно плотна, т. е. для всякого $\varepsilon > 0$ найдется такой компакт K_ε , что $\mu_n(X \setminus K_\varepsilon) < \varepsilon$ для всех n .

Затронутые в обзоре вопросы обсуждались со многими коллегами. Мы особенно благодарны Л. Амброзио, Ф. Барту, С. Г. Бобкову, Ф.-Ю. Вангу, А. М. Верпику, А. В. Гасникову, Ф. Гётце, П. Катё, Б. Клартагу, Ф. Клеману, А. Колесанти, В. Л. Левину, М. Леду, Р. Маккэну, Э. Мильману, Ф. Моргану, М. К. фон Ренессе, М. Рёкнеру, А. Н. Соболевскому, Е. О. Степанову, В. Н. Судакову.

Глава 1. Задача Канторовича

1.1. Метрики Канторовича и нормы Канторовича–Рубинштейна

В этом разделе X – метрическое пространство с метрикой ϱ . Таким образом, классы борелевских и бэровских мер совпадают; для части результатов бывает важна полнота и сепарабельность пространства. Известно (см. [29; гл. 8]), что, за исключением случая конечного X , слабая топология на всем пространстве $\mathcal{M}_\sigma(X)$ неметризуема, тем более она не задается нормой. Однако на $\mathcal{M}_\sigma(X)$ можно задать такую норму, что порождаемая ею топология совпадает со слабой топологией на множестве неотрицательных радоновских мер (а значит, и на множестве вероятностных мер). Рассмотрим на пространстве $\mathcal{M}_\sigma(X)$ следующую норму Канторовича–Рубинштейна:

$$\|\mu\|_0 = \sup \left\{ \int_X f d\mu : f \in \text{Lip}_1(X), \sup_{x \in X} |f(x)| \leq 1 \right\},$$

$$\text{Lip}_1(X) := \{f: X \rightarrow \mathbb{R}^1, |f(x) - f(y)| \leq \varrho(x, y) \forall x, y \in X\}.$$

Эта норма порождает метрику Канторовича–Рубинштейна

$$d_0(\mu, \nu) := \|\mu - \nu\|_0.$$

Ясно, что $\|\mu\|_0 \leq \|\mu\|$. Если в пространстве X есть бесконечная сходящаяся последовательность, то норма $\|\cdot\|_0$ строго слабее нормы $\|\cdot\|$ (вариации). Действительно, если $x_n \rightarrow x$, то меры δ_{x_n} сходятся по норме $\|\cdot\|_0$ к мере δ_x , ибо $|f(x_n) - f(x)| \leq \varrho(x_n, x)$ при $f \in \text{Lip}_1(X)$, однако $\|\delta_x - \delta_{x_n}\| = 2$ при $x_n \neq x$. В частности, при указанном условии пространство $\mathcal{M}_\sigma(X)$ не может быть полным по норме $\|\cdot\|_0$, ибо оно полно по вариации, а тогда по теореме Банаха эти две нормы были бы эквивалентны. Если же $\varrho(x, y) \geq \delta > 0$ при $x \neq y$, то нормы $\|\cdot\|_0$ и $\|\cdot\|$ эквивалентны, поскольку в этом случае $f \in \text{Lip}_1(X)$ при $|f| \leq \delta/2$. Согласно приводимой ниже теореме 1.1.2, порожденная нормой $\|\cdot\|_0$ топология совпадает со слабой топологией на множестве неотрицательных τ -аддитивных мер. Ниже вводится модифицированная метрика Канторовича–Рубинштейна. Часто используется также эквивалентная норма

$$\|\mu\|_{\text{BL}}^* := \sup \left\{ \int_X f d\mu : f \in \text{BL}(X), \|f\|_{\text{BL}} \leq 1 \right\},$$

где $\text{BL}(X)$ – пространство ограниченных липшицевых функций на X с нормой

$$\|f\|_{\text{BL}} := \sup_{x \in X} |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{\varrho(x, y)}.$$

Легко проверить, что $\text{BL}(X)$ с этой нормой полно. Ясно, что

$$\|\mu\|_{\text{BL}}^* \leq \|\mu\|_0 \leq 2\|\mu\|_{\text{BL}}^*,$$

ибо $\|f\|_{\text{BL}} \leq 2$, если $f \in \text{Lip}_1(X)$ и $\sup_X |f(x)| \leq 1$.

ЗАМЕЧАНИЕ 1.1.1. Слабая сходимость направленности $\{\mu_\alpha\}$ неотрицательных мер к мере μ равносильна (см. [29; замечание 8.3.1]) равенству

$$\lim_{\alpha} \int_X f(x) \mu_\alpha(dx) = \int_X f(x) \mu(dx)$$

для всех ограниченных липшицевых функций f на X . В частности, из сходимости направленности неотрицательных мер в метрике Канторовича–Рубинштейна следует слабая сходимость. Однако если X некомпактно, то при некотором выборе метрики на X , задающей прежнюю топологию, обязательно найдутся такие последовательность знакопеременных мер μ_n и мера μ , что интегралы по μ_n от каждой ограниченной равномерно непрерывной функции f сходятся к интегралу от f по мере μ , но меры μ_n не сходятся слабо к μ (см. задачу 8.10.89 в [29]). Исходная метрика не всегда обладает таким свойством (пример: случай $X = \mathbb{N}$ с обычной метрикой), но для $X = \mathbb{R}^1$ годится и стандартная метрика: достаточно иметь такие последовательности $\{x_n\}$ и $\{y_n\}$ с $x_n \neq y_n$ без предельных точек, что расстояние между x_n и y_n стремится к нулю.

Для всякого $B \subset X$ положим $B^\varepsilon = \{x: \text{dist}(x, B) < \varepsilon\}$.

В следующей теореме используется еще одно техническое понятие: борелевская мера μ называется τ -аддитивной, если $\mu(Z_\alpha) \rightarrow 0$ для всякой направленности замкнутых множеств, убывающей к пустому множеству. Таким свойством обладают все радоновские меры, а также все борелевские меры на сепарабельном метрическом пространстве.

ТЕОРЕМА 1.1.2. Порожденная нормой $\|\cdot\|_0$ топология совпадает со слабой топологией на множестве $\mathcal{M}_\tau^+(X)$ неотрицательных τ -аддитивных мер. Кроме того, на множестве $\mathcal{P}_\tau(X)$ вероятностных τ -аддитивных мер слабая топология задается следующей метрикой Леви–Прохорова:

$$d_P(\mu, \nu) = \inf\{\varepsilon > 0: \nu(B) \leq \mu(B^\varepsilon) + \varepsilon, \mu(B) \leq \nu(B^\varepsilon) + \varepsilon \forall B \in \mathcal{B}(X)\}.$$

В частности, если пространство X сепарабельно, то слабая топология на множестве $\mathcal{M}_\sigma^+(X)$ порождается метрикой $d_0(\mu, \nu) = \|\mu - \nu\|_0$.

Если же $\mathcal{P}_\sigma(X) \neq \mathcal{P}_\tau(X)$, то слабая топология на $\mathcal{P}_\sigma(X)$ неметризуема.

Надо иметь в виду, что несовпадение $\mathcal{P}_\sigma(X)$ и $\mathcal{P}_\tau(X)$ – довольно экзотическая для приложений ситуация; оно возможно лишь для несепарабельного X , а в случае полного X требует еще и дополнительных теоретико-множественных допущений (типа существования измеримых мощностей, см., например, [29; предложение 7.2.10]). Тем не менее стоит пояснить причину неметризуемости слабой топологии на $\mathcal{P}_\sigma(X)$: ведь все рассмотренные метрики определены в любом случае. Дело в том, что множество мер с конечными носителями всюду плотно в $\mathcal{P}_\sigma(X)$ в слабой топологии (это верно для всех пространств). Значит, существование метрики, задающей слабую топологию на $\mathcal{P}_\sigma(X)$, дает счетные слабо сходящиеся последовательности дискретных мер для каждой меры из $\mathcal{P}_\sigma(X)$, которая, тем самым, оказывается мерой с сепарабельным носителем (замыканием объединения атомов этой счетной последовательности), а это означает τ -аддитивность (но не всегда радоновость, если пространство неполно).

ТЕОРЕМА 1.1.3. *Для всяких двух борелевских вероятностных мер μ и ν на метрическом пространстве X справедливы следующие соотношения между метриками Леви–Прохорова и Канторовича–Рубинштейна:*

$$\frac{2d_P(\mu, \nu)^2}{2 + d_P(\mu, \nu)} \leq \|\mu - \nu\|_{\text{BL}}^* \leq \|\mu - \nu\|_0 \leq 3d_P(\mu, \nu).$$

Кроме того, $\|\mu - \nu\|_{\text{BL}}^* \leq 2d_P(\mu, \nu)$. Если X полно, то полно и пространство $\mathcal{P}_r(X) = \mathcal{P}_\tau(X)$ с любой из указанных метрик.

Тем самым, для полного сепарабельного метрического пространства X пространство вероятностных борелевских мер на X полно относительно этих метрик, причем они задают на нем слабую топологию.

Из теоремы Прохорова следует, что слабо сходящаяся последовательность знакопеременных борелевских мер на полном сепарабельном метрическом пространстве сходится по норме $\|\cdot\|_0$; но обратное неверно: $\|\sqrt{n}(\delta_{1/n} - \delta_0)\|_0 \rightarrow 0$, хотя меры $\sqrt{n}(\delta_{1/n} - \delta_0)$ не сходятся слабо.

Пусть $\mathcal{P}^1(X)$ – множество всех борелевских вероятностных мер на X , для которых функция $x \mapsto \varrho(x, x_0)$ интегрируема при некотором $x_0 \in X$ (а тогда и при всех x_0 ввиду неравенства треугольника). На $\mathcal{P}^1(X)$ определена следующая модифицированная метрика Канторовича–Рубинштейна (введенная также Р. Форте и Э. Мурье [30] для общего сепарабельного метрического пространства):

$$\|\mu - \nu\|_0^* := \sup \left\{ \int_X f d(\mu - \nu) : f \in \text{Lip}_1(X) \right\}.$$

Отметим еще, что в [31] эта же норма была введена на линейной оболочке дираковских мер при изучении изометрических вложений X . Ясно, что

$$\|\mu - \nu\|_0 \leq \|\mu - \nu\|_0^*,$$

ибо при взятии супремума снято ограничение $\sup_x |f(x)| \leq 1$. Кроме того,

$$\|\mu - \nu\|_0^* \leq \int \varrho(x, a) (\mu + \nu)(dx) \quad \forall a \in X,$$

поскольку $f(x)$ можно заменить на $f(x) - f(a)$ в силу равенства $\mu(X) = \nu(X)$ и неравенства $|f(x) - f(a)| \leq \varrho(x, a)$. Если диаметр X не превосходит 1, то $\|\mu - \nu\|_0 = \|\mu - \nu\|_0^*$, так как $|f(x) - f(a)| \leq 1$ при $f \in \text{Lip}_1(X)$.

Величина $\|\mu - \nu\|_0^*$ действительно является нормой меры $\mu - \nu$, если ввести линейное пространство $\mathcal{M}_0(X)$ всех таких знакопеременных борелевских мер σ на X , что $\sigma(X) = 0$ и функция $x \mapsto \varrho(x, x_0)$ интегрируема относительно $|\sigma|$ (равносильное условие: $\text{Lip}_1(X) \in L^1(|\sigma|)$). На пространстве $\mathcal{M}_0(X)$ указанная выше формула задает норму $\sigma \mapsto \|\sigma\|_0^*$. Заметим, что всякая мера $\sigma \in \mathcal{M}_0(X)$ имеет вид $\|\sigma^+\| \mu - \|\sigma^-\| \nu$, где $\mu, \nu \in \mathcal{P}^1(X)$, $\mu = \sigma^+ / \|\sigma^+\|$, $\nu = \sigma^- / \|\sigma^-\|$. Норму $\|\cdot\|_0^*$ можно распространить на линейное пространство всех ограниченных борелевских мер на X , относительно которых интегрируемы все липшицевы функции. Для этого при фиксированном $x_0 \in X$ положим

$$\|\sigma\|_0^* = |\sigma(X)| + \sup \left\{ \int_X f d\sigma : f \in \text{Lip}_1(X), f(x_0) = 0 \right\}.$$

Если X сепарабельно, то сопряженным к пространству мер с нормой $\|\cdot\|_0^*$ служит пространство $\text{Lip}(X)$ липшицевых функций.

ПРЕДЛОЖЕНИЕ 1.1.4. Для всех $\mu, \nu \in \mathcal{P}^1(X)$ верно равенство

$$\|\mu - \nu\|_0^* = \sup \left\{ \int_X f d\mu + \int_X g d\nu : \right. \\ \left. f \in C(X) \cap L^1(\mu), g \in C(X) \cap L^1(\nu), f(x) + g(y) \leq \varrho(x, y) \right\}.$$

Сам Канторович ввел следующую величину $W(\mu, \nu)$, которую теперь называют метрикой Канторовича (доказательство см. в [29; т. 2, с. 298], [32]).

ТЕОРЕМА 1.1.5. Пусть μ и ν – радоновские меры из $\mathcal{P}^1(X)$. Тогда

$$\|\mu - \nu\|_0^* = W(\mu, \nu) := \inf_{\lambda \in \Pi(\mu, \nu)} \int_{X \times X} \varrho(x, y) \lambda(dx, dy), \quad (1.1.1)$$

причем есть мера $\lambda_0 \in \Pi(\mu, \nu)$, на которой достигается значение $W(\mu, \nu)$.

Если X ограничено, то $\mathcal{P}^1(X) = \mathcal{P}_\sigma(X)$. Отметим, что для всякого метрического пространства (X, ϱ) метрика $\varrho_0 := \varrho/(\varrho + 1)$ (или $\varrho_1 := \min(\varrho, 1)$) задает исходную топологию, но ограничена, поэтому построенная по ϱ_0 функция W из (1.1.1) метризует слабую топологию на $\mathcal{P}_r(X)$. Если диаметр X не превосходит 1, то вложение $X \rightarrow \mathcal{P}_r(X)$, $x \mapsto \delta_x$, сохраняет расстояния при наделении $\mathcal{P}_r(X)$ метрикой d_0 , ибо $d_0(\delta_a, \delta_b) = \varrho(a, b)$, что легко получить, взяв $f(x) = \varrho(x, a)$. Некоторое техническое преимущество модифицированной метрики Канторовича–Рубинштейна $\|\mu - \nu\|_0^*$ состоит в том, что вложение $X \rightarrow \mathcal{P}^1(X)$ всегда сохраняет расстояния независимо от диаметра X : равенство

$$\|\delta_a - \delta_b\|_0^* = \varrho(a, b)$$

верно из-за оценки $|f(a) - f(b)| \leq \varrho(a, b)$ для всех $f \in \text{Lip}_1(X)$, причем на функции $f(x) = \varrho(x, a)$ супремум достигается. Тем самым, для полного сепарабельного метрического пространства X мы получаем возрастающую последовательность полных сепарабельных метрических пространств

$$X \rightarrow \mathcal{P}^1(X) \rightarrow \mathcal{P}^1(\mathcal{P}^1(X)) \rightarrow \dots,$$

где элемент V_{n+1} с номером $n + 1$ имеет вид $\mathcal{P}^1(V_n)$, пространства мер наделены модифицированной метрикой Канторовича–Рубинштейна, порожденной нормой $\|\cdot\|_0^*$, причем все вложения сохраняют расстояния. Эту цепочку уместно назвать *башней Вершика*, который ввел ее более 40 лет назад (см. [5], [33], где есть дополнительные ссылки и обсуждения). Если компактно X , то компактны и все этажи башни Вершика. В связи с изометричностью вложения $X \rightarrow \mathcal{P}^1(X)$ отметим, что норма Канторовича–Рубинштейна максимальна на пространстве V_0 конечных линейных комбинаций σ дираковских мер с условием $\sigma(X) = 0$ среди норм q , для которых $q(\delta_x - \delta_y) = \varrho(x, y)$; о связанных с этим интересных задачах и результатах см. [34].

Известно, что общее вполне регулярное хаусдорфово пространство X гомеоморфно подмножеству дираковских мер в пространстве $\mathcal{P}_r(X)$ вероятностных

мер Радона на X со слабой топологией (см. [29; § 8.9]). При этом компактность X равносильна компактности $\mathcal{P}_r(X)$.

Далее, метризуемость X равносильна метризуемости $\mathcal{P}_r(X)$, а метризуемость X полной метрикой равносильна метризуемости $\mathcal{P}_r(X)$ полной метрикой. Разумеется, на метризуемом пространстве неполная метрика может задавать ту же топологию, что некоторая полная метрика (как на интервале $(0, 1)$). В отличие от слабой топологии, не связанной с метриками, все рассмотренные нами метрики на $\mathcal{P}_r(X)$ существенно зависят от исходной метрики на X .

Аналогично вводится L^p -метрика Канторовича (явно это сделано в [35], см. также [36], [37], [15] и имеющиеся там ссылки). Величина

$$W_p(\mu, \nu) = \inf_{\sigma \in \Pi(\mu, \nu)} \left(\int_{X \times X} \varrho(x, y)^p \sigma(dx dy) \right)^{1/p}$$

является метрикой при $p \geq 1$ на пространстве $\mathcal{P}_r^p(X)$ таких радоновских вероятностных мер μ , что функция $x \mapsto \varrho(x, x_0)$ входит в $L^p(\mu)$ при каком-то (а тогда и при всех) $x_0 \in X$ (при $0 < p < 1$ метрикой оказывается W_p^p). Интегрируемость $\varrho(x, y)^p$ относительно $\sigma \in \Pi(\mu, \nu)$ следует из неравенства $\varrho(x, y)^p \leq 2^p \varrho(x, x_0)^p + 2^p \varrho(y, x_0)^p$ и предполагаемой интегрируемости слагаемых в правой части относительно μ и ν соответственно. Метрика W_p называется расстоянием (или p -расстоянием) Канторовича. В зарубежных работах ее часто называют расстоянием Вассерштайна (Wasserstein distance) под влиянием известной работы Р. Л. Добрушина [38], эффективно использовавшего эту метрику и полагавшего, что впервые она была введена в работе Л. Н. Васерштейна [39], действительно применившего ее при $p = 1$ (см. с. 68 цитированной работы). Тем не менее исторически неправильное название прочно вошло в зарубежную литературу вместе с искаженной транскрипцией имени автора работы [39] (кстати, своевременно переведенной на английский с написанием фамилии автора Vassershtein).

Ясно, что $W_p(\mu, \nu) = W_p(\nu, \mu)$; легко проверить, что $W_p(\mu, \nu) = 0$ лишь при $\mu = \nu$ (равенство $W_p(\mu, \mu) = 0$ видно из того, что в качестве σ можно взять образ μ на диагонали при отображении $x \mapsto (x, x)$). В отличие от большинства прикладных метрик проверка неравенства треугольника для W_p не вполне очевидна. Если даны три меры $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_r^p(X)$, то как оценить $W_p(\mu_1, \mu_3)$ через $W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)$? Если $\sigma_{1,2}$ и $\sigma_{2,3}$ – оптимальные планы для пар (μ_1, μ_2) и (μ_2, μ_3) , то достаточно иметь вероятностную меру η на $X \times X \times X$, у которой проекция на произведение первых двух сомножителей есть мера $\sigma_{1,2}$, на произведение двух последних – мера $\sigma_{2,3}$. Тогда проекции η на первый и третий сомножители есть μ_1 и μ_3 , откуда

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq \left(\int \varrho(x_1, x_3)^p \eta(dx_1 dx_2 dx_3) \right)^{1/p} \\ &\leq \left(\int [\varrho(x_1, x_2) + \varrho(x_2, x_3)]^p \eta(dx_1 dx_2 dx_3) \right)^{1/p} \\ &\leq \left(\int \varrho(x_1, x_2)^p \eta(dx_1 dx_2 dx_3) \right)^{1/p} + \left(\int \varrho(x_2, x_3)^p \eta(dx_1 dx_2 dx_3) \right)^{1/p}, \end{aligned}$$

что совпадает с $\|\varrho\|_{L^p(\sigma_{1,2})} + \|\varrho\|_{L^p(\sigma_{2,3})} = W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)$. Однако существование меры η требует обоснования. Это сделано в следующей лемме.

ЛЕММА 1.1.6. Пусть X_1, X_2, X_3 – вполне регулярные пространства, причем на $X_1 \times X_2$ и $X_2 \times X_3$ есть вероятностные меры Радона $\sigma_{1,2}$ и $\sigma_{2,3}$ с равными проекциями на X_2 . Тогда на $X_1 \times X_2 \times X_3$ есть вероятностная мера Радона η с проекциями $\sigma_{1,2}$ и $\sigma_{2,3}$ на $X_1 \times X_2$ и $X_2 \times X_3$ соответственно.

ДОКАЗАТЕЛЬСТВО. Пусть данные пространства компактны. Тогда на линейном подпространстве в $C(X_1 \times X_2 \times X_3)$ функций вида

$$\varphi(x_1, x_2, x_3) = f(x_1, x_2) + g(x_2, x_3),$$

где $f \in C(X_1 \times X_2)$, $g \in C(X_2 \times X_3)$,

можно задать линейный функционал

$$l(\varphi) = \int f d\sigma_{1,2} + \int g d\sigma_{1,3}.$$

Условие равенства проекций на X_2 гарантирует корректность определения: если $f(x_1, x_2) + g(x_2, x_3) = f_0(x_1, x_2) + g(x_2, x_3)$, то $f(x_1, x_2) - f_0(x_1, x_2)$ не зависит от x_1 и имеет вид $\psi(x_2)$, а меры $\sigma_{1,2}$ и $\sigma_{2,3}$ дают один и тот же интеграл от $\psi(x_2)$. Норма l равна 1, так как $l(1) = 1$ и $l(f + g) \leq 1$, если $f + g \leq 1$. Последнее видно из представления

$$f(x_1, x_2) + g(x_2, x_3) = f(x_1, x_2) + \max_{x_3} g(x_2, x_3) + g(x_2, x_3) - \max_{x_3} g(x_2, x_3),$$

где $f(x_1, x_2) + \max_{x_3} g(x_2, x_3) \leq 1$ и $g(x_2, x_3) - \max_{x_3} g(x_2, x_3) \leq 0$. По теореме Хана–Банаха l продолжается до функционала на $C(X_1 \times X_2 \times X_3)$ с единичной нормой, который по теореме Рисса задается как интеграл по некоторой мере Радона ν с $\|\nu\| = 1$; так как $l(1) = 1$, то мера ν оказывается вероятностной, причем она имеет требуемые проекции (достаточно подставить $g = 0$, затем подставить $f = 0$).

В общем случае в силу радоновости рассматриваемых мер можно считать, что все три пространства являются счетными объединениями компактов. Поэтому они оказываются борелевскими множествами в своих компактификациях Стоуна–Чеха $\overline{X}_1, \overline{X}_2, \overline{X}_3$. Применив доказанное к продолжениям мер $\sigma_{1,2}$ и $\sigma_{2,3}$ на $\overline{X}_1 \times \overline{X}_2$ и $\overline{X}_2 \times \overline{X}_3$, получаем вероятностную меру Радона η на $\overline{X}_1 \times \overline{X}_2 \times \overline{X}_3$ с нужными проекциями. Эта мера сосредоточена на $X_1 \times X_2 \times X_3$, ибо $\eta(X_1 \times X_3 \times \overline{X}_3) = \eta(\overline{X}_1 \times X_2 \times X_3) = 1$ в силу условия на проекции.

Отметим, что для многих пространств (скажем, суслинских) искомую меру легко явно выразить через условные меры:

$$\eta(dx_1 dx_2 dx_3) = \sigma_{1,2}(dx_1|x_2) \sigma_{2,3}(dx_3|x_2) \pi(dx_2),$$

где π – общая проекция $\sigma_{1,2}$ и $\sigma_{2,3}$ на X_2 , $\sigma_{1,2}(\cdot|x_2)$ и $\sigma_{2,3}(\cdot|x_2)$ – соответствующие условные меры, т.е. $\sigma_{1,2}(dx_1 dx_2) = \sigma_{1,2}(dx_1|x_2) \pi(dx_2)$, $\sigma_{2,3}(dx_2 dx_3) = \sigma_{2,3}(dx_3|x_2) \pi(dx_2)$. Иначе говоря,

$$\int f d\eta = \int_{X_2} \int_{X_3} \int_{X_1} f(x_1, x_2, x_3) \sigma_{1,2}(dx_1|x_2) \sigma_{2,3}(dx_3|x_2) \pi(dx_2).$$

Например, если f зависит лишь от x_2, x_3 , то первый интеграл дает $f(x_2, x_3)$, что приводит к интегралу от f по мере $\sigma_{2,3}$. Если f не зависит от x_3 , то аналогично получаем интеграл от f по мере $\sigma_{1,2}$, что и означает выполнение нужных условий на проекции (о связи с умножением полиморфизмов см. [27]). Лемма доказана.

Метрика W_p оценивается через вариацию следующим образом (см. [16; предложение 7.10]):

$$W_p(\mu, \nu)^p \leq 2^{p-1} \|\varrho(\cdot, x_0)(\mu - \nu)\|, \quad \mu, \nu \in \mathcal{P}_r^p(X), \quad p \geq 1, \quad x_0 \in X. \quad (1.1.2)$$

Отметим следующий простой факт. Пусть дана последовательность функций $\theta_j \in \text{Lip}_1(X)$, $0 \leq \theta_j \leq 1$, причем $\theta_j(x) = 1$ для всех x из шара $B(x_0, j)$ радиуса j с некоторым общим центром x_0 и $\theta_j(x) = 0$ при $x \notin B(x_0, j+1)$. Ясно, что для всякой меры μ меры $\theta_j \cdot \mu$ слабо сходятся к μ . Аналогичное верно и для метрики W_p . Обозначим через $\xi_j(\mu)$ вероятностную меру $C_{n,j} \theta_j \cdot \mu$, $C_{n,j} = (\theta_j \cdot \mu(X))^{-1}$, если $\theta_j \cdot \mu(X) > 0$, а если $\theta_j \cdot \mu(X) = 0$, то положим $\xi_j(\mu) = \mu$.

ЛЕММА 1.1.7. *Для всякой меры $\mu \in \mathcal{P}_r^p(X)$ имеем $W_p(\xi_j(\mu), \mu) \rightarrow 0$ при $j \rightarrow \infty$. Более того, $\lim_{j \rightarrow \infty} \sup_n W_p(\xi_j(\mu_n), \mu_n) = 0$, если $\mu_n, \mu \in \mathcal{P}_r^p(X)$ таковы, что меры $(1 + \varrho(\cdot, x_0)^p)\mu_n$ слабо сходятся к мере $(1 + \varrho(\cdot, x_0)^p)\mu$.*

ДОКАЗАТЕЛЬСТВО. Первое утверждение сразу следует из (1.1.2). Для доказательства второго надо еще проверить, что $\|\varrho(\cdot, x_0)(\xi_j(\mu_n) - \mu_n)\| \rightarrow 0$ при $j \rightarrow \infty$ равномерно по n . Из слабой сходимости указанных мер следует, что для данного $\delta \in (0, 1/2)$ найдется такое $R > 0$, что интегралы от $1 + \varrho(x, x_0)^p$ по внешности шара $B(R, x_0)$ по всем мерам μ_n и μ меньше δ . Интегралы же по всему пространству ограничены некоторым числом C . Можно рассматривать далее только такие n , что $\mu_n(B(R, x_0)) > 1 - \delta$. При $j > R$ имеем $\theta_j(x) = 1$ при $x \in B(R, x_0)$, поэтому ограничение меры $\xi_j(\mu_n)$ на $B(R, x_0)$ равно ограничению меры $C_{n,j} \mu_n$, где $1 \leq C_{n,j} \leq 1 + \delta$. Кроме того, $\xi_j(\mu_n) \leq 2\mu_n$. Значит, $\|\varrho(\cdot, x_0)(\xi_j(\mu_n) - \mu_n)\|$ оценивается через $C\delta + 3\delta$. Лемма доказана.

СЛЕДСТВИЕ 1.1.8. *В $(\mathcal{P}_r^p(X), W_p)$ всюду плотно множество мер с ограниченными носителями, а тогда и множество мер с конечными носителями.*

Метрика W_p при $p > 1$ следующим образом связана со слабой топологией.

ТЕОРЕМА 1.1.9. *Последовательность мер $\mu_n \in \mathcal{P}_r^p(X)$ сходится к мере $\mu \in \mathcal{P}_r^p(X)$ по метрике W_p в точности тогда, когда $\{\mu_n\}$ сходится к μ слабо и выполнено равенство*

$$\lim_{n \rightarrow \infty} \int_X \varrho(x, x_0)^p \mu_n(dx) = \int_X \varrho(x, x_0)^p \mu(dx) \quad (1.1.3)$$

для некоторого (а тогда и всякого) $x_0 \in X$. Это также равносильно тому, что меры $(1 + \varrho(x, x_0)^p)\mu_n$ слабо сходятся к $(1 + \varrho(x, x_0)^p)\mu$.

ДОКАЗАТЕЛЬСТВО. Слабая сходимость $\{\mu_n\}$ к μ вместе с (1.1.3) следует из слабой сходимости мер $\nu_n = (1 + \varrho(x, x_0)^p)\mu_n$ к мере $\nu = (1 + \varrho(x, x_0)^p)\mu$, а также и влечет ее, ибо при условии $\nu_n(X) \rightarrow \nu(X)$ для слабой сходимости

достаточно иметь сходимость интегралов от функций $f \in C_b(X)$ с ограниченными носителями. Пусть $W_p(\mu_n, \mu) \rightarrow 0$. Тогда $W(\mu_n, \mu) \rightarrow 0$, поэтому есть слабая сходимость. Проверим (1.1.3). Из слабой сходимости и рассмотрения срезок $\min(\varrho(x, x_0)^p, N)$ следует, что правая часть не больше нижнего предела интегралов от $\varrho(x, x_0)^p$ по μ_n . С другой стороны, из неравенства $\varrho(x_0, x) \leq \varrho(x, y) + \varrho(x_0, y)$ следует, что для всякого $q > 1$ найдется такое $C > 0$, что $\varrho(x, x_0)^p \leq C\varrho(x, y)^p + q\varrho(x_0, y)$. Взяв меру $\sigma_n \in \Pi(\mu_n, \mu)$, интеграл от $\varrho(x, y)^p$ по которой равен $W_p(\mu_n, \mu)^p$, и проинтегрировав по ней предыдущее неравенство, получаем оценку

$$\int_X \varrho(x, x_0)^p \mu_n(dx) \leq CW_p(\mu_n, \mu)^p + q \int_X \varrho(x, x_0)^p \mu(dx).$$

Так как $W_p(\mu_n, \mu) \rightarrow 0$, то верхнего предела от левой части этой оценки не больше произведения q и правой части (1.1.3). При $q \rightarrow 1$ получаем оценку, завершающую вывод (1.1.3). Наконец, покажем, что слабая сходимость мер ν_n к ν влечет сходимость $W_p(\mu, \mu_n) \rightarrow 0$. Лемма 1.1.7 сводит все к случаю мер, равных нулю вне некоторого шара, а тем самым к случаю ограниченной метрики. В последнем случае утверждение верно в силу теорем 1.1.2 и 1.1.5. Теорема доказана.

При $p > 1$ метрика W_p на симплексе вероятностных мер не может быть сужением никакой нормы на пространстве мер из-за отсутствия выпуклости на отрезках: уже для $X = \{0, 1\}$ мы имеем $W_p(\delta_0, (1-t)\delta_0 + t\delta_1) = t^{1/p}$. Однако верен следующий факт.

СЛЕДСТВИЕ 1.1.10. *На $\mathcal{P}_r^p(X)$ топология, порожденная метрикой W_p , задается также нормой*

$$K_p(\mu) := \|(1 + \varrho(\cdot, x_0)^p)\mu\|_0, \quad \mu \in \mathcal{M}_r^p(X),$$

где линейное пространство $\mathcal{M}_r^p(X)$ состоит из всех таких мер Радона μ , что функция $\varrho(\cdot, x_0)^p$ интегрируема относительно $|\mu|$.

ДОКАЗАТЕЛЬСТВО. Как мы знаем, сходимость последовательностей в пространстве $\mathcal{P}_r^p(X)$ с указанной нормой равносильна слабой сходимости после умножения на $1 + \varrho(\cdot, x_0)^p$, что по предыдущей теореме равносильно сходимости в метрике W_p . Надо отметить, что введенная норма может не порождать метрику W_p на $\mathcal{P}_r^p(X)$, она порождает лишь ту же топологию.

Еще в работах Р. Л. Добрушина метрика W_1 использовалась для оценки скорости сходимости к инвариантному распределению. Интересная геометрическая интерпретация дана в [40]. Будем говорить, что дискретная марковская цепь M с переходными вероятностями P и метрикой d имеет кривизну Риччи не меньше $k \geq 0$, если $W_1(P_x, P_y) \leq (1-k)d(x, y)$, $x, y \in M$, $P_x = P\delta_x$, $P_y = P\delta_y$. Из этого можно вывести много интересных свойств. Например, для любой вероятностной меры μ на M выполнено неравенство $W_1(\mu P^N, \pi) \leq (1-k)^N W_1(\mu, \pi)$, где π – инвариантное распределение, т. е. имеет место экспоненциальная скорость сходимости к инвариантному распределению. Аналогия с кривизной Риччи для многообразий подтверждается многочисленными примерами (дискретные версии теорем сравнения, неравенств типа Соболева и т. п.).

1.2. Существование оптимальных планов

Известен ряд весьма широких условий существования решения задачи Канторовича, т. е. меры, минимизирующей функционал $\sigma \mapsto K(\mu, \nu, \sigma)$. Приведем основные факты. Сначала отметим, что в случае мер Радона μ и ν на вполне регулярных пространствах X и Y множество $\Pi(\mu, \nu)$ оказывается равномерно плотным: для каждого $\varepsilon > 0$ найдется такой компакт $K \subset X \times Y$, что $\sigma((X \times Y) \setminus K) < \varepsilon$; в качестве K можно взять $K_1 \times K_2$, где компакты K_1 и K_2 таковы, что $\mu(X \setminus K_1) + \nu(Y \setminus K_2) < \varepsilon$. Кроме того, $\Pi(\mu, \nu)$ замкнуто в слабой топологии. Поэтому $\Pi(\mu, \nu)$ оказывается слабо компактным (см. [29; теорема 8.6.7]). Ввиду линейности функционала $K(\mu, \nu, \cdot)$ для доказательства существования его минимума на $\Pi(\mu, \nu)$ достаточно его непрерывности на $\Pi(\mu, \nu)$ в слабой топологии или хотя бы слабой замкнутости множеств $\{\sigma \in \Pi(\mu, \nu) : K(\mu, \nu, \sigma) \leq c\}$, т. е. полунепрерывности снизу. Обратим внимание, что эти условия слабее соответствующих требований на всем пространстве мер. Например, если функция h ограничена и полунепрерывна снизу, то мы получаем полунепрерывность снизу для $K(\mu, \nu, \cdot)$. Действительно, как известно (см. [29; следствие 8.2.5]), для любой слабо сходящейся направленности мер $\sigma_\alpha \rightarrow \sigma$ в $\Pi(\mu, \nu)$ верно неравенство

$$\liminf_{\alpha} \int h d\sigma_{\alpha} \geq \int h d\sigma.$$

Если функция $h \geq 0$ неограничена, но полунепрерывна снизу, мы получаем функционал $K(\mu, \nu, \sigma)$ на $\Pi(\mu, \nu)$ в виде возрастающей последовательности функционалов, задаваемых ограниченными полунепрерывными снизу функциями $h_n = \min(h, n)$, откуда очевидным образом следует полунепрерывность снизу предельного функционала. Это дает следующее утверждение (напомним, что существование решения не означает, что $K(\mu, \nu) < \infty$).

ТЕОРЕМА 1.2.1. *Для мер Радона μ и ν на вполне регулярных пространствах и полунепрерывной снизу функции $h \geq 0$ задача Канторовича имеет решения.*

Понятно, что условия на h можно несколько ослабить. Ведь фактически нам была достаточна полунепрерывность снизу функционалов, заданных ограниченными функциями $\min(h, n)$, а для нее, как легко проверить, достаточно полунепрерывности снизу функции h не на всем $X \times Y$, а на какой-либо последовательности компактов вида $A_j \times B_j$, где $\mu(A_j) \rightarrow 1$, $\nu(B_j) \rightarrow 1$. Тем самым, область действия теоремы значительно расширяется.

Хорошо известно и просто проверяется, что функция h на вполне регулярном пространстве полунепрерывна снизу (т. е. все множества $\{h \leq c\}$ замкнуты) в точности тогда, когда $h = \sup_{\alpha \in A} h_{\alpha}$, где $h_{\alpha} \geq 0$ – непрерывные функции, A – некоторое множество индексов, причем если $h \geq 0$, то можно выбрать $h_{\alpha} \geq 0$; если пространство метризуемо, то можно взять $A = \mathbb{N}$. Эти соображения вместе со слабой компактностью сводят нахождение $K_h(\mu, \nu)$ к случаю ограниченной непрерывной функции стоимости.

ПРЕДЛОЖЕНИЕ 1.2.2. Для мер Радона μ и ν на вполне регулярных пространствах и полунепрерывной снизу функции $h \geq 0$ верно равенство

$$K_h(\mu, \nu) = \sup_{\alpha} K_{h_{\alpha}}(\mu, \nu),$$

где $h_{\alpha} \geq 0$ – такие непрерывные функции, что $h = \sup_{\alpha \in A} h_{\alpha}$. Кроме того,

$$K_h(\mu, \nu) = \sup_N K_{h \wedge N}(\mu, \nu) = \sup_{N, \alpha} K_{h_{\alpha} \wedge N}(\mu, \nu).$$

ДОКАЗАТЕЛЬСТВО. Предположим, что $\sup_{\alpha} K_{h_{\alpha}}(\mu, \nu) < K_h(\mu, \nu) - 2\delta$, где $\delta > 0$. Перейдем к новому набору непрерывных функций h_{α} , индексиремых конечными подмножествами $\alpha \subset A$, частично упорядоченными по включению, положив $h_{\alpha} = \max(h_{\beta_1}, \dots, h_{\beta_k})$, $\alpha = \{\beta_1, \dots, \beta_k\}$. Тогда h есть предел возрастающей направленности $\{h_{\alpha}\}$ (где $h_{\alpha} \geq h_{\beta}$ при $\alpha \geq \beta$). Для всякой меры Радона $\sigma \geq 0$ на $X \times Y$, по которой функция h интегрируема, выполнено равенство

$$\int_{X \times Y} h d\sigma = \lim_{\alpha} \int_{X \times Y} h_{\alpha} d\sigma.$$

Оно проверяется сначала для $h \wedge N$ (см. [29; лемма 7.2.6]). Как показано выше, есть оптимальный план $\hat{\sigma}$ для полунепрерывной функции h , а для каждой функции h_{α} есть оптимальный план $\hat{\sigma}_{\alpha}$. Пользуясь слабой компактностью $\Pi(\mu, \nu)$, перейдем к поднаправленности в $\{\hat{\sigma}_{\alpha}\}$, слабо сходящейся к некоторой мере $\sigma_0 \in \Pi(\mu, \nu)$. Можно считать, что это вся исходная направленность. Заметим, что $K_h(\mu, \nu, \sigma_0) = K_h(\mu, \nu)$. В самом деле, если $K_h(\mu, \nu, \sigma_0) > K_h(\mu, \nu) + C$, где $C > 0$, то в силу сказанного выше найдется β , для которого $K_{h_{\beta}}(\mu, \nu, \sigma_0) > K_h(\mu, \nu) + C$. Тогда из слабой сходимости $\hat{\sigma}_{\alpha}$ к σ_0 и непрерывности h_{β} получаем существование такого $\alpha > 0$, что $K_{h_{\alpha}}(\mu, \nu, \hat{\sigma}_{\alpha}) \geq K_{h_{\beta}}(\mu, \nu, \hat{\sigma}_{\alpha}) > K_h(\mu, \nu) + C$, что невозможно. Таким образом, $K_h(\mu, \nu, \sigma_0) = K_h(\mu, \nu)$, поэтому можно считать, что $\hat{\sigma} = \sigma_0$. Теперь возьмем такое γ , что $K_{h_{\gamma}}(\mu, \nu, \sigma_0) > K_h(\mu, \nu, \sigma_0) - \delta$. Еще раз пользуясь слабой сходимостью, найдем такое $\beta > \gamma$, что $K_{h_{\beta}}(\mu, \nu, \hat{\sigma}_{\beta}) > K_h(\mu, \nu, \sigma_0) - \delta$. Так как $K_{h_{\beta}}(\mu, \nu, \hat{\sigma}_{\beta}) \geq K_{h_{\gamma}}(\mu, \nu, \hat{\sigma}_{\beta})$, то приходим к противоречию, что завершает доказательство первого утверждения. Оставшиеся равенства являются его следствиями. Предложение доказано.

ЗАМЕЧАНИЕ 1.2.3. Для ограниченной непрерывной функции стоимости h совершенно аналогичные соображения позволяют получить $K_h(\mu, \nu)$ в виде предела величин $K_h(\mu_{\alpha}, \nu_{\alpha})$, где меры μ_{α} и ν_{α} имеют конечные носители. Для этого находим направленность таких мер σ_{α} на $X \times Y$, слабо сходящуюся к оптимальному плану $\hat{\sigma}$. Их проекции μ_{α} и ν_{α} на X и Y слабо сходятся к μ и ν соответственно; берем для них оптимальные планы $\hat{\sigma}_{\alpha}$ и переходим к слабо сходящейся поднаправленности $\{\hat{\sigma}_{\alpha}\}$, для слабого предела σ_0 которой получаем оценку $K_h(\mu, \nu, \sigma_0) \leq K(\mu, \nu)$, ибо $K_h(\mu, \nu, \sigma_0) = \lim_{\alpha} K_h(\mu, \nu, \hat{\sigma}_{\alpha}) \leq \lim_{\alpha} K_h(\mu, \nu, \sigma_{\alpha}) = K(\mu, \nu)$. Ясно, что $\sigma_0 \in \Pi(\mu, \nu)$, значит, $K_h(\mu, \nu, \sigma_0) = K(\mu, \nu)$. В случае метризуемых пространств (или пространств с метризуемыми компактами) вместо направленностей можно взять счетные последовательности.

Оптимальные планы описываются через одно интересное свойство их носителей. Пусть X, Y – множества, c – функция на $X \times Y$.

ОПРЕДЕЛЕНИЕ 1.2.4. Множество $S \subset X \times Y$ называют c -циклически монотонным, если $\sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{i+1})$ при $(x_1, y_1), \dots, (x_n, y_n) \in S$, где $y_{n+1} := y_1$. Если же $\sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{\sigma(i)})$ для всякой перестановки σ индексов $\{1, \dots, n\}$, то S называют c -монотонным.

Как несложно проверить, эти понятия равносильны.

Для функций ψ на Y и φ на X со значениями в $[-\infty, +\infty]$ положим

$$\begin{aligned}\psi^{c+}(x) &= \inf_{y \in Y} [c(x, y) - \psi(y)], & \psi^{c-}(x) &= \sup_{y \in Y} [-c(x, y) - \psi(y)], & x &\in X, \\ \varphi^{c+}(y) &= \inf_{x \in X} [c(x, y) - \varphi(x)], & \varphi^{c-}(y) &= \sup_{x \in X} [-c(x, y) - \varphi(x)], & y &\in Y.\end{aligned}$$

Функция $\varphi: X \rightarrow [-\infty, +\infty)$ называется c -вогнутой, если существует такая функция $\psi: Y \rightarrow [-\infty, +\infty)$, что $\varphi = \psi^{c+}$. Это равносильно тому, что $\varphi = \varphi^{c+c+}$. Функция φ называется c -выпуклой, если c -вогнута функция $-\varphi$. Аналогично определяют c -вогнутость и c -выпуклость функций на Y .

Для c -вогнутой функции φ на X определим c -супердифференциал $\partial^{c+}\varphi \subset X \times Y$ как множество $\partial^{c+}\varphi = \{(x, y): \varphi(x) + \varphi^{c+}(y) = c(x, y)\}$. Положим $\partial^{c+}\varphi(x) := \{y \in Y: (x, y) \in \partial^{c+}\varphi\}$. Аналогично, c -субдифференциал выпуклой функции φ на X есть множество $\partial^{c-}\varphi = \{(x, y): \varphi(x) + \varphi^{c-}(y) = -c(x, y)\}$.

Включение $y \in \partial^{c+}\varphi(x)$ равносильно оценке $\varphi(x) - c(x, y) \geq \varphi(z) - c(z, y)$ при всех $z \in X$. Непосредственно проверяется, что c -супердифференциал c -вогнутой функции c -монотонен. В [41] получен важный обратный результат.

ТЕОРЕМА 1.2.5. *Всякое c -циклически монотонное множество S лежит в c -супердифференциале некоторой c -вогнутой функции.*

В случае функции $c(x, y) = \langle x, y \rangle$ на $\mathbb{R}^d$ множество $C \subset \mathbb{R}^d \times \mathbb{R}^d$ называется циклически монотонным (без указания функции), если для всякого набора $(x_1, y_1), \dots, (x_n, y_n) \in C$ верно неравенство $\sum_{i=1}^n \langle x_i, y_{i+1} - y_i \rangle \leq 0$, где $y_{n+1} = x_1$.

Субдифференциалом выпуклой функции φ на $\mathbb{R}^d$ в точке x называют множество $\partial\varphi(x) = \{y: \forall z \in \mathbb{R}^d, \varphi(z) \geq \varphi(x) + \langle y, z - x \rangle\}$. Общее определение выше происходит из этого случая. Выпуклая функция φ дифференцируема в x в точности тогда, когда субдифференциал $\partial\varphi(x)$ состоит из одного элемента. В этом случае $\partial\varphi(x) = \{\nabla\varphi(x)\}$. Следующая теорема Рокафеллара – классический результат выпуклого анализа, который обобщает предыдущая теорема.

ТЕОРЕМА 1.2.6. *Всякое циклически монотонное множество C в $\mathbb{R}^d \times \mathbb{R}^d$ лежит в графике субдифференциала некоторой выпуклой функции φ .*

Функцию φ можно явно задать так. Зафиксируем $(x_0, y_0) \in C$ и положим $\varphi(x) = \sup\{\langle y_m, x - x_m \rangle + \dots + \langle y_0, x_1 - x_0 \rangle; (x_0, y_0), \dots, (x_m, y_m) \in C\}$.

Следующий результат восходит к [42], доказательство можно прочесть в [10], [17] (для польских пространств, общий случай аналогичен).

ТЕОРЕМА 1.2.7. Пусть X, Y вполне регулярны, $\mu \in \mathcal{P}_r(X)$, $\nu \in \mathcal{P}_r(Y)$, функция $c \geq 0$ непрерывна на $X \times Y$ и существуют такие конечные функции $a \in L^1(\mu)$ и $b \in L^1(\nu)$, что $c(x, y) \leq a(x) + b(y)$. Тогда для $\pi \in \Pi(\mu, \nu)$ следующие условия равносильны: (i) план π оптимален, (ii) топологический носитель π является c -циклически монотонным, (iii) существует такая c -вогнутая функция φ на X , что $\varphi^+ \in L^1(\mu)$ и топологический носитель π лежит в $\partial^{c+}\varphi$.

СЛЕДСТВИЕ 1.2.8. Пусть в ситуации предыдущей теоремы $\hat{\sigma}$ – оптимальный для μ и ν , а носитель меры $\sigma_0 \in \mathcal{P}_r(X \times Y)$ лежит в носителе $\hat{\sigma}$. Тогда мера σ_0 является оптимальным планом для своих проекций.

Упомянем еще результат работы [43], где ослаблены условия на c .

ТЕОРЕМА 1.2.9. Пусть X и Y – полные сепарабельные метрические пространства и функция $c: X \times Y \rightarrow [0, +\infty]$ измерима по Борелю. Тогда всякий конечный оптимальный план сосредоточен на c -монотонном множестве. Всякий план с конечной стоимостью, сосредоточенный на c -циклически монотонном множестве, оптимален, если существует замкнутое множество F и $\mu \otimes \nu$ -нулевое множество N такие, что $c^{-1}(+\infty) = F \cup N$.

Условием теоремы 1.2.9 охватываются как конечные, так и полунепрерывные снизу c , но без каких-либо условий на c обойтись нельзя (см. [43]).

ЗАМЕЧАНИЕ 1.2.10. Если X и Y – полные сепарабельные метрические пространства, конечная функция $c \geq 0$ полунепрерывна снизу и $K_c(\mu, \nu) < \infty$, то имеется борелевское c -циклически монотонное множество $\Gamma \subset X \times Y$ с таким свойством: для всякой меры $\pi \in \Pi(\mu, \nu)$ оптимальность π равносильна тому, что $\pi(\Gamma) = 1$.

Отметим одно свойство выпуклости $K(\mu, \nu)$ (см. [17; теорема 4.8]). Пусть X, Y – полные сепарабельные метрические пространства, $(T, \mathcal{B}, P)$ – вероятностное пространство, $t \mapsto \mu_t$ и $t \mapsto \nu_t$ – измеримые отображения в пространства $\mathcal{P}(X)$ и $\mathcal{P}(Y)$, $\mu = \int_T \mu_t P(dt)$, $\nu = \int_T \nu_t P(dt)$.

ТЕОРЕМА 1.2.11. Пусть функция c полунепрерывна снизу, причем $c(x, y) \leq a(x) + b(y)$, где $a \in L^1(\mu)$, $b \in L^1(\nu)$. Тогда $K_c(\mu, \nu) \leq \int_T K_c(\mu_t, \nu_t) P(dt)$.

Стоит отметить, что еще сам Канторович в случае компакта и функции стоимости, равной метрике ϱ , показал, что оптимальный план π характеризуется существованием липшицевой функции U , для которой $U(x) - U(y) = \varrho(x, y)$ п. в. относительно π .

Явных решений задач Монжа и Канторовича обычно нет (см. [44] об одномерном случае, а также [45] про некоторые примеры на плоскости).

1.3. Двойственность Канторовича

Для произвольной функции стоимости h и ассоциированной с ней задачи Канторовича при широких условиях верно следующее замечательное соотношение, называемое двойственностью Канторовича. Для заданных вероятностных мер μ и ν на пространствах $(X, \mathcal{A})$ и $(Y, \mathcal{B})$ соответственно рассматриваются так называемые двойственный функционал и двойственная задача

$$J(\mu, \nu, \varphi, \psi) = \int_X \varphi d\mu + \int_Y \psi d\nu, \quad J(\mu, \nu) = \sup\{J(\mu, \nu, \varphi, \psi) : (\varphi, \psi) \in \Phi_h\}.$$

Теорема о двойственности утверждает, что при определенных условиях выполнено равенство

$$J(\mu, \nu) = K(\mu, \nu).$$

Первый случай такого равенства восходит к работе Канторовича [1] и совместным работам Канторовича и Рубинштейна [3], [4] для метрических компактов. Затем круг пространств, для которых оно верно, был значительно расширен. Прежде чем переходить к точным формулировкам, отметим, что еще в [3], [4] была рассмотрена близкая (но не равносильная) задача нахождения инфимума (или минимума) $K_0(\varrho_0)$ функционала

$$K_0(\varrho_0, \eta) = \int_{X \times X} h(x, y) \eta(dx dy)$$

по всем неотрицательным мерам η на квадрате $(X^2, \mathcal{A} \otimes \mathcal{A})$ измеримого пространства $(X, \mathcal{A})$, для которых проекции η_1 и η_2 на первый и второй сомножители удовлетворяют равенству $\eta_1 - \eta_2 = \varrho_0$ для заданной (знакопеременной) меры ϱ_0 на X с $\varrho_0(X) = 0$. Например, если $\varrho_0 = 0$, то нулевая мера дает нулевое решение, но если даны две различные вероятностные меры μ и ν и $\varrho_0 = \mu - \nu$, то задача становится нетривиальной. Известно [46], что если h удовлетворяет неравенству треугольника (и только в этом случае), мы имеем равенство $K_0(\mu - \nu) = K(\mu, \nu)$ при довольно широких условиях, но в общем случае такой связи нет. В некоторых отношениях эта задача сложнее, но методы исследования обеих задач имеют много общего. Мы остановимся на технически более простой задаче для $K(\mu, \nu)$.

Приведем основные результаты, связанные с равенством $J(\mu, \nu) = K(\mu, \nu)$. В [47] доказана следующая теорема весьма общего характера, в которой используется понятие совершенной вероятностной меры P на измеримом пространстве $(\Omega, \mathcal{F})$, т. е. такой меры, что для всякой $\mathcal{F}$ -измеримой функции f в множестве $f(\Omega)$ есть борелевское множество B , для которого $P(f^{-1}(B)) = 1$, иначе говоря, множество $f(\Omega)$ измеримо относительно $P \circ f^{-1}$; о совершенных мерах см. [29; § 7.5]. Хотя существуют меры, не являющиеся совершенными, все радоновские меры совершенны.

ТЕОРЕМА 1.3.1. *Если хотя бы одна из мер μ и ν совершенна, то равенство $K(\mu, \nu) = J(\mu, \nu)$ верно для всех ограниченных $\mathcal{A} \otimes \mathcal{B}$ -измеримых функций h или, более общим образом, $\mathcal{A} \otimes \mathcal{B}$ -измеримых функций, допускающих поточечную оценку $h(x, y) \leq a(x) + b(y)$ с некоторыми функциями $a \in L^1(\mu)$, $b \in L^1(\nu)$. Тем самым, это равенство верно, если хотя бы одна из двух мер радонова,*

а последнее выполнено, если какая-то из этих мер является борелевской на полном сепарабельном метрическом пространстве.

В [48], [49] вероятностное пространство $(X, \mathcal{A}, \mu)$ названо пространством со свойством двойственности, если $K(\mu, \nu) = J(\mu, \nu)$ для всякого вероятностного пространства $(Y, \mathcal{B}, \nu)$ и всякой ограниченной измеримой функции h на $X \times Y$. По-видимому, остается неизвестным, существуют ли вероятностные пространства без этого свойства; как указано авторами работы [49], их контрпримеры, описанные в [47], [48], ошибочны. Поэтому из доказанного в этих работах следует, что совершенность μ необходима для более сильного свойства сильной двойственности, которое в дополнение к свойству двойственности требует еще, чтобы для каждого вероятностного пространства $(Y, \mathcal{B}, \nu)$ и всякой под- σ -алгебры $\mathcal{B}_0 \subset \mathcal{B}$ для каждой ограниченной $\mathcal{A} \otimes \mathcal{B}_0$ -измеримой функции h величина $J(\mu, \nu, h)$ была одинаковой при рассмотрении h как на $\mathcal{A} \times \mathcal{B}$, так и на $\mathcal{A} \times \mathcal{B}_0$. Неясно, является ли это техническое свойство в самом деле более сильным.

В некоторых важных частных случаях доказательство может быть получено из общих результатов выпуклого анализа (двойственность Фенхеля–Рокафеллара). Его можно найти в [16] вместе с весьма интересными эвристическими выводами этого соотношения, основанными на принципе минимакса (см. также [12] о связи двойственности Канторовича с линейным программированием). Однако даже для простых пространств с более сложными функциями стоимости требуются более специальные рассуждения. Из такого рода результатов необходимо упомянуть следующее недавнее достижение из [50].

ТЕОРЕМА 1.3.2. Пусть X и Y – полные сепарабельные метрические пространства с борелевскими вероятностными мерами μ и ν соответственно и функция $h: X \times Y \rightarrow [0, +\infty]$ измерима по Борелю и конечна $\mu \otimes \nu$ -почти всюду. Предположим также, что существует мера в $\Pi(\mu, \nu)$, относительно которой функция h интегрируема. Тогда $J(\mu, \nu) = K(\mu, \nu)$.

Из формулировки понятно, что вместо борелевских мер на полных сепарабельных метрических пространствах можно брать меры на абстрактных измеримых пространствах, изоморфные (в смысле существования измеримых в обе стороны и сохраняющих меру взаимно однозначных отображений, определенных почти всюду) таким измеримым пространствам.

Какие-то ограничивающие условия на h , конечно, нужны. Скажем, в [50; пример 4.1] рассмотрена такая ситуация: $X = Y = [0, 1]$, $\mu = \nu$ – мера Лебега, $h(x, y) = +\infty$ при $x < y$, $h(x, x) = 1$, $h(x, y) = 0$ при $x > y$. Тогда $K(\mu, \nu) = 1$ (транспортный план – нормированная мера Лебега на диагонали), но $J(\mu, \nu) = 0$. В [51] доказан следующий результат.

ТЕОРЕМА 1.3.3. Пусть μ и ν – меры Радона. Равенство $J(\mu, \nu) = K(\mu, \nu)$ верно, либо если функция h полунепрерывна снизу, либо если h измерима по Борелю и $h(x, y) \leq h_1(x) + h_2(y)$ для некоторых $h_1 \in L^1(\mu)$, $h_2 \in L^1(\nu)$. В последнем случае величина $J(\mu, \nu)$ достигается на некоторой паре $(\hat{h}_1, \hat{h}_2)$ с $\hat{h}_1 \in L^1(\mu)$, $\hat{h}_2 \in L^1(\nu)$.

В работе [52] (где, в нашей терминологии, берется функция $-h$) дано более короткое доказательство равенства $J(\mu, \nu) = K(\mu, \nu)$ в случае мер Радона на

вполне регулярных пространствах и неотрицательных функций h , полунепрерывных снизу (из-за перехода к $-h$ там речь идет о полунепрерывности сверху, которой в нашем случае недостаточно, как показывает пример 2.5 в [51]).

ЗАМЕЧАНИЕ 1.3.4. Даже при наличии равенства $J(\mu, \nu) = K(\mu, \nu)$ инфимум $J(\mu, \nu)$ может не достигаться (не быть минимумом), причем такое случается как для полунепрерывных снизу h на $[0, 1]$ и $\mu = \nu$, равных мере Лебега, так и для абсолютно непрерывных мер μ и ν на $\mathbb{R}$ и $h(x, y) = (x - y)^2$ (см. [50], где $J(\mu, \nu)$ даже равно интегралу от $\varphi(x) + \psi(y)$ по оптимальному плану для некоторой пары (φ, ψ) , но эти функции не лежат в $L^1(\mu)$ и $L^1(\nu)$ соответственно).

ЗАМЕЧАНИЕ 1.3.5. Пусть X и Y – сепарабельные метрические пространства, функция $h \geq 0$ полунепрерывна снизу и $K_h(\mu, \nu) < \infty$. Для существования решений прямой и двойственной задач Канторовича и равенства их значений достаточно (см. [53; теорема 3.2]), чтобы интеграл от $h(x, y)$ по μ не был бесконечен при ν -почти всех y , а интеграл от $h(x, y)$ по ν не был бесконечен при μ -почти всех x .

Из формулы двойственности видно, что для функций h общего вида решение $\hat{\sigma}$ задачи Канторовича сконцентрировано на небольшой части $X \times Y$. Скажем, если (φ, ψ) – решение двойственной задачи, то интеграл от $\varphi(x) - \psi(y)$ по мере $\hat{\sigma}$ совпадает с интегралом от $h(x, y)$, что дает равенство $\varphi(x) - \psi(y) = h(x, y)$ почти всюду относительно $\hat{\sigma}$. Это дает иное объяснение указанного выше факта, что оптимальный план сосредоточен на циклически монотонном множестве (см. предыдущий раздел). В ситуации замечания 1.2.10 указанное там множество $\Gamma \subset X \times Y$ можно выбрать так, что оптимальность $\pi \in \Pi(\mu, \nu)$ равносильна существованию такой h -выпуклой функции ψ , что $h(x, y) = \psi^h(y) - \psi(x)$ π -п.в., а также равносильно существованию таких функций $\varphi: X \rightarrow [-\infty, +\infty)$, $\psi: Y \rightarrow [-\infty, +\infty)$, что $\varphi(x) + \psi(y) \leq h(x, y)$ и $\varphi(x) + \psi(y) = h(x, y)$ π -п.в.

ЗАМЕЧАНИЕ 1.3.6. Существует иная формулировка двойственной задачи Канторовича в случае вполне регулярных топологических пространств X и Y . А именно, вместо класса Φ_h пар измеримых функций φ и ψ можно рассмотреть более узкий класс Φ_h^C пар ограниченных непрерывных функций φ и ψ , для которых $\varphi(x) - \psi(y) \leq h(x, y)$ при всех $x \in X$, $y \in Y$. Затем вместо $J(\mu, \nu)$ можно взять величину $J^C(\mu, \nu)$, беря супремум $J(\mu, \nu, \varphi, \psi)$ по всем парам из Φ_h^C . Ясно, что $\Phi_h^C \leq \Phi_h$. В общем случае неравенство строгое, причем в работе [46] показано, что равенство $\Phi_h^C = \Phi_h$ равносильно полунепрерывности снизу функции h (см. также [15; теорема 4.6.8]).

Число работ по общей задаче Канторовича в топологической постановке и в постановке общей теории меры слишком велико, чтобы быть полноценно представленным в библиографии, но особо отметим многолетний цикл работ В. Л. Левина [28], [46], [54]–[58], где есть дополнительные ссылки.

Обратимся к случаю $X = Y = \mathbb{R}^n$, $h(x, y) = |x - y|^2/2$. Перейдем к функциям $\varphi(x) = |x|^2/2 - u(x)$, $\psi(y) = |y|^2/2 - v(y)$. Тогда задача, двойственная задаче

Канторовича, принимает вид минимизации функционала

$$\tilde{J}(\mu, \nu, \varphi, \psi) = \int_X \varphi d\mu + \int_Y \psi d\nu, \quad \varphi(x) + \psi(y) \geq \langle x, y \rangle.$$

Саму задачу Канторовича можно переформулировать как поиск меры m с заданными проекциями, на которой достигается максимум функционала

$$\tilde{K}(m) \mapsto \int_{X \times Y} \langle x, y \rangle m(dx dy).$$

Здесь видна связь задачи Канторовича с преобразованием Лежандра, ибо из постановки задачи ясно, что с самого начала можно ограничиться функциями, удовлетворяющими равенству $\psi = \varphi^*$, где $\varphi^*(y) = \sup_x (\langle x, y \rangle - \varphi(x))$. Аналогично, $\varphi = \psi^* = \varphi^{**}$. Следовательно, функционал $\tilde{J}$ можно сузить на множество пар (φ, φ^*) с выпуклой функцией φ . В этом случае имеет место такой факт.

ТЕОРЕМА 1.3.7. *Найдется решение двойственной задачи (φ, ψ) , где φ выпукла, $\psi = \varphi^*$ и $\varphi(x) + \varphi^*(y) \geq \langle x, y \rangle$, причем равенство достигается на топологическом носителе меры m , дающей решение задачи Канторовича.*

Ниже в теореме 1.3.7 мы приведем основные факты о двойственной задаче в этом случае при дополнительном предположении, что мера μ равна нулю на множествах хаусдорфовой размерности не выше $d - 1$ (скажем, абсолютно непрерывна). Тогда окажется, что отображение $T = \nabla \varphi$ – оптимальная транспортировка μ в ν , т.е. решение задачи Канторовича дает и решение задачи Монжа. Как видно, здесь мы подошли к соприкосновению обеих задач.

Глава 2. Задача Монжа

2.1. Существование и единственность оптимальных отображений

Обратимся к задаче Монжа. Итак, задана пара вероятностных мер μ и ν на измеримых пространствах $(X, \mathcal{A})$ и $(Y, \mathcal{B})$ соответственно. Решением задачи Монжа называется отображение $T \in T(\mu, \nu)$, на котором достигается минимум $M(\mu, \nu, T)$. В отличие от задачи Канторовича, решения задачи Монжа может не быть в самых простых случаях.

ПРИМЕР 2.1.1. (i) Пусть $X = Y = [-1, 1]$, $\mu = \delta_0$, $\nu = 2^{-1}(\delta_{-1} + \delta_1)$, $h(x, y) = |x - y|^2$. Тогда μ вообще нельзя перевести в ν , а единственным решением задачи Канторовича служит полусумма мер Дирака в точках $(0, -1)$ и $(0, 1)$.

(ii) Пусть $X = Y = [-1, 1]^2$, $\mu = \lambda \otimes \delta_0$, $\nu = 2^{-1}(\lambda \otimes \delta_{-1} + \lambda \otimes \delta_1)$, λ – нормированная мера Лебега на $[-1, 1]$, $h(x, y) = |x - y|^2$, где $|\cdot|$ – норма в $\mathbb{R}^2$. Тогда обе меры не имеют атомов, задача Канторовича имеет решение и $K(\mu, \nu) = 1$, но задача Монжа не имеет решений (хотя имеет приближенные решения, см. теорему 2.3.1).

ДОКАЗАТЕЛЬСТВО. (i) Если мера σ на $[-1, 1]^2$ имеет проекции μ и ν , то она сосредоточена на пересечении отрезка $\{0\} \times [-1, 1]$ с объединением отрезков $[-1, 1] \times \{-1\}$ и $[-1, 1] \times \{1\}$, т.е. является комбинацией мер Дирака

в $(0, -1)$ и $(0, 1)$, откуда видно, что она должна быть их полусуммой. Тем самым, здесь $\Pi(\mu, \nu)$ состоит из одного элемента.

(ii) Здесь $\Pi(\mu, \nu)$ содержит много мер, но величина $K(\mu, \nu)$ находится легко, ибо при минимизации интеграла от функции $|x - y|^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2$ по мерам $\sigma \in \Pi(\mu, \nu)$ мы получаем, что интеграл от первого слагаемого равен 1 из-за того, что $x_1 = 0$ σ -п.в. и $y_1^2 = 1$ σ -п.в., так что минимум, равный 1, достигается в точности тогда, когда $x_2 = y_2$ σ -п.в. Оптимальным планом служит распределение случайного вектора $(0, \eta, \xi, \eta)$, где ξ и η – независимые случайные величины, причем ξ принимает значения -1 и 1 с вероятностями $1/2$, а η имеет равномерное распределение в $[-1, 1]$. Если предположить, что задача Монжа имеет решение $T = (T_1, T_2)$, то получаем, что $|T_1(x)| = 1$ μ -п.в., так что интеграл от $|x - T(x)|^2 = (x_1 - T_1(x))^2 + (x_2 - T_2(x))^2$ по мере μ равен интегралу от $1 + (x_2 - T_2(x))^2$ относительно μ . С учетом равенства $K(\mu, \nu) = 1$ и теоремы 2.3.1 это означает, что $T_2(x) = x_2$ μ -п.в., так что $T_1(x) = \xi(x_2)$ при почти всех x_2 есть 1 или -1 . Тогда мера $\mu \circ T^{-1}$ будет сосредоточена на объединении множеств $A \times \{-1\}$ и $B \times \{1\}$, где A и B – дизъюнктные измеримые множества с $A \cup B = [-1, 1]$, и, следовательно, не может совпадать с мерой ν .

Приведем один общий результат об однозначной разрешимости задачи Монжа (см. [17; теорема 5.28]).

ТЕОРЕМА 2.1.2. Пусть X, Y – полные сепарабельные метрические пространства, функция $h \geq 0$ полунепрерывна снизу и $K_h(\mu, \nu) < \infty$. Если для μ -почти всех x для каждой h -выпуклой функции φ на X множество $\partial^h \varphi(x)$ содержит не более одного элемента, то задача Монжа имеет единственное решение T , причем $T(x) = \partial^h \psi(x)$ μ -п.в. для некоторой h -выпуклой функции ψ , а образ μ при отображении $x \mapsto (x, T(x))$ дает решение задачи Канторовича.

В оставшейся части этого раздела мы обсудим важный частный случай, когда $X = Y = \mathbb{R}^d$, а h имеет специальный вид $h = |x - y|^p$, $p \geq 1$, где $|x|$ – обычная евклидова норма. Будем предполагать, что $\mu, \nu \in \mathcal{P}^p(\mathbb{R}^d)$. Отсюда следует, что существует решение m_p соответствующей задачи Канторовича. Напомним, что решение задачи Канторовича дает более слабый результат, чем решение исходной задачи Монжа. Тем не менее оказывается, что при довольно общих условиях эти задачи эквивалентны. Для случая $p = 1$ первый результат существования появился в работе В. Н. Судакова [21], согласно которой для любой нормы $\|\cdot\|$ на $\mathbb{R}^d$ (не обязательно порожденной скалярным произведением) и всякой пары вероятностных мер вида $\mu = \varrho_\mu dx$, $\nu = \varrho_\nu dx$ на $\mathbb{R}^d$ существует такое отображение $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$, что $\nu = \mu \circ T^{-1}$ и образ меры μ при отображении $x \mapsto (x, T(x)) \in \mathbb{R}^d \times \mathbb{R}^d$ дает решение соответствующей задачи Канторовича для $h(x, y) = \|x - y\|$. Другими словами, мера m_1 на $\mathbb{R}^d \times \mathbb{R}^d$ сосредоточена на графике отображения T : $m_1 = \mu \circ (x, T(x))^{-1}$. Мы будем называть T оптимальным отображением для функции стоимости $h(x, y) = \|x - y\|$ или L^1 -оптимальным отображением. Позднее в доказательстве Судакова был найден пробел, связанный с весьма тонкими вопросами теории условных мер и нелинейными версиями теоремы Фубини, который к настоящему времени удалось полностью устранить (хотя при этом к одному из промежуточных технических утверждений построили контрпример). Однако

впечатляет, что на это ушло три десятилетия, включая немалый срок после обнаружения пробела. Имеется целый ряд работ, в которых разными методами доказывалось существование L^1 -оптимального отображения для норм разных типов, см. [53], [59], [23], [60]–[62] и, наконец, совсем недавнюю работу [24], в которой на норму $\|\cdot\|$ нет ограничений. Во всех них применяются весьма сложные конструкции. Само решение T имеет нетривиальное описание.

Более просто доказывалось существование решения в случае так называемого квадратичного оптимального отображения, или L^2 -оптимального отображения, т. е. отображения $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$, преобразующего μ в ν при функции стоимости $|x - y|^2$. Для такого T справедливо равенство

$$W_2(\mu, \nu) = \left(\int_{\mathbb{R}^d} |T(x) - x|^2 \mu(dx) \right)^{1/2}.$$

Оказывается, что такое T существует и имеет вид $T(x) = \nabla \varphi(x)$, где φ – некоторая выпуклая функция.

Удивительно, но первый общий результат о существовании квадратичных оптимальных отображений появился значительно позже работы Судакова об L^1 -оптимальных отображениях, хотя квадратичные оптимальные отображения существенно проще устроены и имеют обширные приложения в различных областях математики. Более того, квадратичные оптимальные отображения фактически давно возникли и изучались в геометрии как решения некоторого нелинейного уравнения в частных производных, так называемого уравнения Монжа–Ампера (см. [63], [64]). Уравнению Монжа–Ампера и связанным с ним геометрическим задачам были посвящены работы А. Д. Александрова, И. Я. Бакельмана, А. В. Погорелова, Э. Калаби, Ш.-Т. Яу, Л. Ниренберга и других исследователей. Это уравнение возникает, например, в классической задаче Минковского. Возможно, причина упомянутого запаздывания в том, что в квадратичной задаче Монжа–Канторовича отсутствовал внятный физический смысл, в то время как функционал Канторовича с функцией стоимости, равной норме (а не квадрату нормы), допускал физическую интерпретацию “работы”. Как бы то ни было, впервые существование оптимальных отображений при достаточно общих предположениях было доказано в работе Я. Бренье [65], мотивированной анализом уравнений механики сплошной среды. Основные результаты этой работы таковы: 1) доказаны существование и единственность оптимальных отображений для широкого класса мер, 2) найдена связь транспортной задачи Монжа–Канторовича с уравнением Монжа–Ампера, 3) получена теорема о полярной факторизации, согласно которой отображение $h: \Omega \rightarrow \mathbb{R}^d$ класса L^2 на области $\Omega \subset \mathbb{R}^d$, переводящее меру Лебега в абсолютно непрерывную меру, имеет вид $h = \nabla \psi \circ U$, где $U: \Omega \rightarrow \Omega$ сохраняет меру Лебега, ψ – выпуклая функция. Приведем более общую версию результата Бренье, доказанную Р. Маккэном [66].

ТЕОРЕМА 2.1.3. Пусть μ и ν – вероятностные меры на $\mathbb{R}^d$ и $\mu(A) = 0$ для всякого множества A хаусдорфовой размерности не выше $d - 1$. Тогда существует такое борелевское отображение $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$, что $\nu = \mu \circ T^{-1}$, причем $T = \nabla \varphi$ для некоторой выпуклой функции φ . Это отображение единственно с точностью до переопределения на множестве μ -меры нуля, т. е. для любых двух таких отображений T_1, T_2 выполнено равенство $T_1 = T_2$ μ -п.в.

Если также $\nu(A) = 0$ для всякого множества A хаусдорфовой размерности не выше $d-1$, то существует аналогичное отображение S , для которого $\mu = \nu \circ S^{-1}$, причем $S = \nabla \varphi^*$, где $\varphi^*(y) = \sup_{x \in \mathbb{R}^d} \{\langle x, y \rangle - \varphi(x)\}$. Для μ -почти всех x и ν -почти всех y выполнены равенства $\nabla \varphi^* \circ \nabla \varphi(x) = x$, $\nabla \varphi \circ \nabla \varphi^*(y) = y$.

Доказательство этой теоремы опирается на теорию двойственности Канторовича и технику выпуклого анализа. Мы укажем основные шаги упрощенного варианта доказательства из [66]. Задача минимизации функционала

$$m \mapsto \int |x - y|^2 m(dx dy)$$

равносильна задаче максимизации функционала $m \mapsto \int \langle x, y \rangle m(dx dy)$, ибо

$$\int |x - y|^2 m(dx dy) = \int |x|^2 \mu(dx) + \int |y|^2 \nu(dy) - 2 \int \langle x, y \rangle m(dx dy).$$

ПРИМЕР 2.1.4. Пусть $\mu = N^{-1} \sum_{i=1}^N \delta_{x_i}$ и $\nu = N^{-1} \sum_{i=1}^N \delta_{y_i}$ сосредоточены в различных точках $x_1, \dots, x_N$ и $y_1, \dots, y_N$ из $\mathbb{R}^d$. Транспортировка меры μ в ν задается перестановкой $\sigma \in S_N$: $\left(N^{-1} \sum_{i=1}^N \delta_{x_i}\right) \circ T_\sigma^{-1} = N^{-1} \sum_{i=1}^N \delta_{y_{\sigma(i)}}$. Несложно показать, что T_σ оптимально (минимизирует $\sum_{i=1}^N |x_i - y_{\sigma(i)}|^2$) в точности тогда, когда $\sum_{k=1}^m \langle y_{\sigma(i_k)}, x_{i_{k+1}} - x_{i_k} \rangle \leq 0$, где $i_{m+1} = i_1$, при всех $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$, т. е. график T_σ циклически монотонен.

СЛЕДСТВИЕ 2.1.5. Существует решение задачи Канторовича π_n для пары дискретных мер $\mu = N^{-1} \sum_{i=1}^N \delta_{x_i}$, $\nu = N^{-1} \sum_{i=1}^N \delta_{y_i}$, сосредоточенное на циклически монотонном дискретном множестве вида $\{(x_i, y_{\sigma(i)})\}$.

Пусть μ, ν – вероятностные меры на $\mathbb{R}^d$. Очевидным образом существуют слабо сходящиеся последовательности $\mu_n \rightarrow \mu$, $\nu_n \rightarrow \nu$ дискретных мер. Более того, без ограничения общности можно считать, что каждая из мер μ_n и ν_n равномерно распределена на $N(n)$ атомах. Из сказанного выше следует, что для всякого n существуют меры π_n с соответствующими проекциями μ_n и ν_n , сосредоточенные на дискретных циклически монотонных множествах. Заметим, что семейство $\{\pi_n\}$ слабо компактно по теореме Прохорова. Из свойств слабой сходимости нетрудно вывести, что существует мера π с проекциями μ и ν , топологический носитель которой является циклически монотонным множеством. По теореме 1.2.6 он лежит в графике субдифференциала некоторой выпуклой функции φ . Хорошо известно, что выпуклая функция дифференцируема почти всюду по мере Лебега. Менее известно, что множество точек недифференцируемости выпуклой функции имеет хаусдорфову размерность не

более $d - 1$. Предположим, что $\mu(A) = 0$, если хаусдорфова размерность A не больше $d - 1$. Тогда из сказанного следует, что для μ -почти всех x пересечение $\text{supp}(\pi) \cap \{(x, y) : y \in \mathbb{R}^d\} \subset \mathbb{R}^d \times \mathbb{R}^d$ состоит из единственной точки $\nabla\varphi(x)$. Нетрудно видеть, что $\nabla\varphi$ и есть искомое отображение. Итак, существование оптимального отображения следует из теоремы 1.2.6 и примера 2.1.4.

ЗАМЕЧАНИЕ 2.1.6. Единственность отображения $\nabla\varphi$ вытекает из следующего наблюдения А. Д. Александрова: пусть φ_1, φ_2 – две выпуклые функции, причем $\varphi_1(x_0) = \varphi_2(x_0)$, но $\nabla\varphi_1(x_0) \neq \nabla\varphi_2(x_0)$. Тогда $\nabla\varphi_1(D)$ строго содержится в $\nabla\varphi_2(D)$, где $D = \{\varphi_1 > \varphi_2\}$ и $\nabla\varphi_i(D) = \bigcup_{x \in D} \partial\varphi_i(x)$. Существование двух оптимальных отображений $\nabla\varphi_1$ и $\nabla\varphi_2$ противоречит наблюдению Александрова, ибо $\nabla\varphi_1(D)$ и $\nabla\varphi_2(D)$ имеют одинаковую ν -меру.

Наконец, отметим, что если отображение $\nabla\varphi$ гладкое и $\mu = \varrho_\mu dx$, $\nu = \varrho_\nu dx$, то должна быть выполнена следующая формула замены переменных:

$$\varrho_\nu(\nabla\varphi) \det D^2\varphi = \varrho_\mu.$$

Это равенство можно рассматривать как уравнение на φ . Уравнения такого вида называются *уравнениями Монжа–Ампера*. Удивительным образом оказывается, что формула замены переменных для мер с плотностями верна в некотором смысле всегда. Как мы уже отмечали выше, первые производные выпуклой функции существует почти всюду. Вторые производные выпуклых функций можно понимать в обобщенном смысле, т. е. по определению $\partial_{e_i}\partial_{e_j}\varphi$ есть обобщенная функция, удовлетворяющая равенству

$$\langle \partial_{e_i}\partial_{e_j}\varphi, \psi \rangle = - \int_{\mathbb{R}^d} \partial_{e_i}\psi \partial_{e_j}\varphi dx$$

для всех гладких финитных функций ψ . В силу выпуклости φ функционал $\partial_{e_i}\partial_{e_j}\varphi$ неотрицателен, т. е. задается мерой. Пользуясь выпуклостью, несложно показать также, что обобщенная частная производная $\partial_{e_i}\partial_{e_j}\varphi$ тоже является мерой (возможно, знакопеременной). Как известно, существует разложение меры $\partial_{e_i}\partial_{e_j}\varphi = (\partial_{e_i}\partial_{e_j}\varphi)_a dx + (\partial_{e_i}\partial_{e_j}\varphi)_{\text{sing}}$ на абсолютно непрерывную и сингулярную части. Согласно теореме А. Д. Александрова (см. [67; теорема 2.3.2]), функция $(\partial_{e_i}\partial_{e_j}\varphi)_a$ является пределом разностного отношения

$$\begin{aligned} & \frac{1}{2t^2} [\varphi(x + te_i + te_j) + \varphi(x - te_i - te_j) - \varphi(x + te_i) - \varphi(x - te_i) \\ & - \varphi(x + te_j) - \varphi(x - te_j) + 2\varphi(x)] \end{aligned}$$

для почти всех x по мере Лебега. Эта теорема имеет следующую упрощенную формулировку: “каждая выпуклая функция почти всюду дважды дифференцируема”. Всюду далее под $D_a^2\varphi$ подразумевается матрица, составленная из функций $(\partial_{e_i}\partial_{e_j}\varphi)_a$. Оказывается, что только $D_a^2\varphi$ влияет на формулу замены переменных. Следующая теорема была доказана в [68].

ТЕОРЕМА 2.1.7. Равенство $\varrho_\nu(\nabla\varphi) \det D_a^2\varphi = \varrho_\mu$ верно μ -п.в.

Метод множителей Лагранжа. Нестрогое, но поучительное доказательство того факта, что T является градиентом, можно получить путем вывода уравнения Эйлера–Лагранжа (см. [12]). Пусть T – произвольное отображение, переводящее μ в ν . Решение задачи Монжа можно искать как условный экстремум функционала $\int_{\mathbb{R}^d} \langle T(x), x \rangle \mu(dx)$ при условии $\varrho_\nu(T) \det DT = \varrho_\mu$. Составим функцию Лагранжа:

$$\int_{\mathbb{R}^d} \left(\langle T(x), x \rangle \varrho_\mu(x) + \lambda(x) (\varrho_\nu(T(x)) \det DT(x) - \varrho_\mu(x)) \right) dx.$$

Функция λ играет роль множителя Лагранжа (заметим, что по формуле замены переменных $\varrho_\nu(T) \det DT = \varrho_\mu$). Так как

$$\det(A + \varepsilon B) = \det A \det(I + \varepsilon A^{-1}B) \sim \det A \cdot (1 + \varepsilon \operatorname{Tr}(A^{-1}B)),$$

то первая вариация функции Лагранжа равна

$$\int_{\mathbb{R}^d} \left(\langle \omega(x), x \rangle \varrho_\mu + \lambda \cdot \varrho_\mu \operatorname{Tr}[DT^{-1}D\omega] + \lambda \langle \nabla \varrho_\nu(T), \omega \rangle \frac{\varrho_\mu}{\varrho_\nu(T)} \right) dx,$$

где ω – гладкое векторное поле с компактным носителем. Интегрируя по частям и применяя замену переменных, несложно убедиться в том, что

$$\begin{aligned} \int_{\mathbb{R}^d} \lambda \cdot \operatorname{Tr}[DT^{-1}D\omega] \varrho_\mu dx &= \int_{\mathbb{R}^d} \lambda(T^{-1}) \operatorname{div}(\omega(T^{-1})) \varrho_\nu dx \\ &= - \int_{\mathbb{R}^d} \langle \nabla[\lambda(T^{-1})], \omega(T^{-1}) \rangle \varrho_\nu dx - \int_{\mathbb{R}^d} \lambda(T^{-1}) \left\langle \omega(T^{-1}), \frac{\nabla \varrho_\nu}{\varrho_\nu} \right\rangle \varrho_\nu dx. \end{aligned}$$

Пусть $\lambda = u(T)$, где u – некоторая функция. Применяя опять формулу замены переменных, получаем, что для любого гладкого поля ω выполнено равенство

$$\int_{\mathbb{R}^d} (\langle \omega(x), x \rangle - \langle \nabla u(T(x)), \omega(x) \rangle) \varrho_\mu(x) dx = 0.$$

Следовательно, $\nabla u(T(x)) = x$ и $T^{-1} = \nabla u$. В силу симметричности задачи относительно μ и ν аналогичное верно для самого отображения T .

Неквадратичные функции стоимости. Интересный класс отображений мер составляют оптимальные отображения для неквадратичной функции стоимости. Для функции стоимости $h(x, y) = h(x - y)$, где h строго выпукла, они имеют вид $T(x) = x - \nabla h^*(\nabla \varphi(x))$, где φ – некоторая h -выпуклая функция. Последнее означает, что φ имеет вид $\varphi(x) = \inf_{y \in \mathbb{R}^d} (h(x, y) - \psi(y))$ для некоторой функции ψ . Подробнее об этом см. [16], [15]. О единственности без условий выпуклости см. [69].

2.2. Регулярность решений и априорные оценки

Здесь мы обсудим свойства регулярности оптимальных транспортировок. Под этим понимаются свойства типа непрерывности, гёльдеровости, дифференцируемости (обычной и соболевской), оценки роста. Наиболее полная картина имеется пока для квадратичной функции стоимости $|x - y|^2$ на $\mathbb{R}^d$, к которой и относятся все приводимые в этом разделе результаты. Лишь в конце будет сделано несколько замечаний о других функциях стоимости и о многообразиях.

2.2.1. Принцип максимума. Сначала мы обсудим нелинейный принцип максимума, который был впервые получен в работах А. Д. Александрова по геометрии выпуклых поверхностей. Он обобщает классический принцип максимума и является важным инструментом в теории регулярности линейных эллиптических уравнений (в недивергентной форме), а также широкого класса нелинейных эллиптических уравнений. Ниже приведено классическое доказательство, которое, как нетрудно видеть, основано на идее транспортировки мер.

ОПРЕДЕЛЕНИЕ 2.2.1. Выпуклой оболочкой непрерывной функции f называется функция $f_* = \sup\{u \leq f, u \text{ выпукла}\}$. Множеством точек контакта называется множество $C_*(f) = \{x: f(x) = f_*(x)\}$.

Аналогично определяется вогнутая оболочка.

ОПРЕДЕЛЕНИЕ 2.2.2. Мерой Монжа–Ампера, ассоциированной с выпуклой функцией φ на выпуклом множестве в $\mathbb{R}^d$, называют меру

$$\mu_\varphi(A) = \lambda_d\left(\bigcup_{x \in A} \partial\varphi(x)\right).$$

Если функция φ дважды непрерывно дифференцируема, то

$$\mu_\varphi(A) = \int_A \det D^2\varphi(x) dx.$$

Если $\det D^2\varphi > 0$, то μ_φ – прообраз меры Лебега при отображении $\nabla\varphi$.

ТЕОРЕМА 2.2.3 (принцип максимума). Пусть f – непрерывная функция на выпуклом множестве A . Тогда для некоторой постоянной $C(d)$, зависящей только от размерности, выполнено неравенство

$$\sup_{x \in A} f(x) \leq \sup_{x \in \partial A} f(x) + C(d) \operatorname{diam}(A) [\mu_{(-f)_*}(C_*(-f))]^{1/d}.$$

В частности, если f дважды непрерывно дифференцируема, то

$$\sup_A f \leq \sup_{\partial A} f + C(d) \operatorname{diam}(A) \left[\int_{\{x: D^2 f(x) \leq 0\}} |\det D^2 f| dx \right]^{1/d}.$$

ДОКАЗАТЕЛЬСТВО. Переходя к функции $g(x) = \sup_{y \in \partial A} f(y) - f(x)$, получаем равносильное неравенство $\inf_{x \in A} g(x) \geq -C(d) \operatorname{diam}(A) [\mu_{g_*}(C_*(g))]^{1/d}$ для g с условием $\inf_{x \in \partial A} g(x) \geq 0$. Если $m = \inf_{x \in A} g(x) \geq 0$, неравенство тривиально. В противном случае зафиксируем точку минимума x_0 функции g . Рассмотрим конус в пространстве $\mathbb{R}^{d+1}$ с вершиной в точке (x_0, m) и основанием $\partial A \subset \{x: x_{d+1} = 0\}$, заданный уравнением $x_{d+1} = K(x)$. Пусть $v \in \partial K(x)$ для некоторой точки $x' \in A$. Рассмотрим гиперплоскость L , заданную уравнением $x_{d+1} = K(x') + \langle v, x - x' \rangle$ и касающуюся K в точке x' . Так как график g лежит выше L на ∂A и совпадает с L в точке x' , то, сдвигая непрерывно график L вниз по оси x_{d+1} , находим точку x_v , лежащую в множестве точек контакта $C_*(g)$, причем $v \in \partial g_*(x_v)$. Следовательно, $\partial K(A) \subset \partial g_*(C_*(g))$ и

$\mu_{\partial K}(A) \leq \mu_{g_*}(C_*(g))$. Поскольку функция K однородна, то $\partial K(A) = \partial K(x_0)$. Мера $\partial K(x_0)$ легко оценивается снизу через объем d -мерного шара радиуса $-m/\text{diam}(A)$. Следовательно, $\mu_{g_*}(A) \geq C(d)(-m/\text{diam}(A))^d$, откуда вытекает доказываемое.

Используя параболическую транспортировку мер на пространстве $\mathbb{R}^{d+1}$ вида $(t, x) \mapsto (\langle x, \nabla_x u \rangle - u, \nabla_x u)$, где $u(\cdot, \cdot): \Omega = [0, T] \times Q \rightarrow \mathbb{R}$, $Q \subset \mathbb{R}^d$ – компактное выпуклое множество, можно доказать параболический принцип максимума:

$$\sup_{\Omega} u \leq C(d, \text{diam}(Q)) \left(\int_{\Gamma_u} |\partial_t u \cdot \det D_x^2 u| dt dx \right)^{1/d},$$

где $\Gamma_u = \{\partial_t u \leq 0, D^2 u \leq 0\} \subset \Omega$, $u = 0$ на $([0, T] \times \partial Q) \cup (T \times Q)$.

2.2.2. Регулярность решений уравнения Монжа–Ампера. Уравнение Монжа–Ампера, возникающее в задаче транспортировки мер, является частным случаем уравнений вида

$$\det D^2 \varphi(x) = f(x, \nabla \varphi(x)) \quad (2.2.1)$$

из класса так называемых полностью нелинейных эллиптических уравнений (см. обзоры [70], [71]). В [64] описан интересующий нас частный случай (уравнения Монжа–Ампера для транспортной задачи). Задача регулярности для уравнения Монжа–Ампера весьма непростая. Относительно полная картина складывалась более 60 лет в работах А. Д. Александрова, Э. Калаби, А. В. Погорелова, Н. В. Крылова, Дж. Спрука, Л. Каффарелли, Л. Ниренберга и многих других. В геометрии параллельно развивалось близкое направление – комплексное уравнение Монжа–Ампера. Этим занимались Ш.-Т. Яу, Э. Калаби, Т. Обэн. Классический подход к теории регулярности нелинейных уравнений основан на дифференцировании нелинейного уравнения и применении линейной теории к получившемуся уравнению, которое оказывается линейным относительно старших производных. Пример подобной техники можно найти в п. 2.2.3, где будут получены некоторые равномерные оценки на вторые производные потенциала транспортной задачи. Процитируем Н. В. Крылова (см. [70]): “Доказать существование решений уравнения (2.2.1) методами, известными до 1981 г., было нелегкой задачей. Для этого необходимо было найти априорные оценки для решений и их производных до *третьего* порядка. Большая часть этой работы основана на дифференцировании (2.2.1) три раза и некоторых чрезвычайно тонко продуманных приемах Калаби. После 1981 г. подход к полностью нелинейным уравнениям драматически изменился”. В 1980-х годах в работах Н. В. Крылова и М. В. Сафонова, а также (независимо) Л. Эванса были доказаны общие теоремы регулярности для эллиптических уравнений, задаваемых нелинейным оператором вида $F(D^2 \varphi)$, где F – равномерно эллиптический оператор. К сожалению, оператор Монжа–Ампера $\varphi \mapsto \det D^2 \varphi$ не является равномерно эллиптическим даже на пространстве выпуклых функций, и теория Крылова–Сафонова–Эванса не применима напрямую.

Ниже используется понятие решения уравнения Монжа–Ампера в смысле А. Д. Александрова. Решением уравнения

$$\det D^2 \varphi = w$$

по Александрову называется функция φ , у которой мера Монжа–Ампера удовлетворяет равенству $\mu_\varphi = w dx$. Заметим, что решение в смысле Александрова с абсолютно непрерывной правой частью автоматически подразумевает отсутствие сингулярной компоненты у μ_φ , что, вообще говоря, может не выполняться, если речь идет об оптимальной транспортировке $w dx$ в меру Лебега. В случае гладкой φ оба понятия решения совпадают.

2.2.3. Классический подход. Пусть $\nabla\Phi$ – оптимальная транспортировка меры $e^{-V} dx$ в меру $e^{-W} dx$. По формуле замены переменных

$$V = W(\nabla\Phi) - \ln \det D^2\Phi.$$

Зафиксируем единичный вектор e и продифференцируем это равенство вдоль e . Для этого применим соотношение

$$\partial_e \ln \det D^2\Phi = \frac{\partial_e \det D^2\Phi}{\det D^2\Phi} = \text{Tr}[(D^2\Phi)^{-1} \partial_e D^2\Phi].$$

Дифференцируя это соотношение вдоль еще одного вектора v и используя равенство $\partial_v(D^2\Phi)(D^2\Phi)^{-1} + D^2\Phi \partial_v[(D^2\Phi)^{-1}] = 0$, получаем

$$\partial_v \partial_e \ln \det D^2\Phi = \text{Tr}[(D^2\Phi)^{-1} \partial_v \partial_e D^2\Phi] - \text{Tr}[(D^2\Phi)^{-1} \partial_e (D^2\Phi) (D^2\Phi)^{-1} \partial_v D^2\Phi].$$

Используя опять формулу замены переменных, находим

$$\begin{aligned} \partial_e V &= \langle \nabla W(\nabla\Phi), D^2\Phi \cdot e \rangle - \text{Tr}[(D^2\Phi)^{-1} \partial_e D^2\Phi], \\ \partial_e^2 V &= \langle D^2 W(\nabla\Phi) D^2\Phi \cdot e, D^2\Phi \cdot e \rangle + \langle \nabla W(\nabla\Phi), \nabla \partial_e^2 \Phi \rangle \\ &\quad - \text{Tr}[(D^2\Phi)^{-1} \partial_e^2 D^2\Phi] + \text{Tr}[(D^2\Phi)^{-1} \partial_e D^2\Phi]^2. \end{aligned} \quad (2.2.2)$$

Представим эти соотношения в более удобном виде. Определим диффузионный оператор L_Φ формулой $L_\Phi f = \text{Tr}[D^2 f \cdot (D^2\Phi)^{-1}] - \langle \nabla f, \nabla W(\nabla\Phi) \rangle$.

ЗАМЕЧАНИЕ 2.2.4. Если меры μ и ν имеют локально соболевские плотности, то нетрудно проверить, что

$$- \int \langle (D^2\Phi)^{-1} \nabla f, \nabla \eta \rangle d\mu = \int f \cdot L_\Phi \eta d\mu = \int \eta \cdot L_\Phi f d\mu, \quad \eta \in C_0^\infty(\mathbb{R}^d).$$

Таким образом, $\partial_e V = -L_\Phi \partial_e \Phi$, где L_Φ – генератор формы Дирихле

$$\mathcal{E}_\Phi(f, \eta) = \int \langle (D^2\Phi)^{-1} \nabla f, \nabla \eta \rangle d\mu.$$

Из (2.2.2) получаем

$$\partial_e^2 V = \langle D^2 W(\nabla\Phi) D^2\Phi \cdot e, D^2\Phi \cdot e \rangle - L_\Phi \partial_e^2 \Phi + \text{Tr}[(D^2\Phi)^{-1} \partial_e D^2\Phi]^2. \quad (2.2.3)$$

Используя (2.2.3), с помощью принципа максимума можно получить априорные оценки на супремум $\partial_e^2 \Phi$.

ПРИМЕР 2.2.5. Пусть функция $\Phi \in C^4$ выпукла и $\nabla\Phi$ – оптимальная транспортная мера $e^{-V} dx$ с носителем в выпуклом множестве A в меру Лебега на множестве B , e – единичный вектор. Тогда

$$\sup_{x \in A} \partial_e^2 \Phi(x) \leq \sup_{x \in \partial A} \partial_e^2 \Phi(x) + C(d) \operatorname{diam}(A) \left(\int_A (\partial_e^2 V)_+^d e^{-V} dx \right)^{1/d}.$$

ДОКАЗАТЕЛЬСТВО. По формуле замены переменных $\det D^2\Phi = 1/\lambda(B)$. Из равенства (2.2.3) получаем $\partial_e^2 V = -\operatorname{Tr}[(D^2\Phi)^{-1}\partial_e^2 D^2\Phi] + \operatorname{Tr}[(D^2\Phi)^{-1}\partial_e D^2\Phi]^2$. Если $\partial_e^2 D^2\Phi(x) \leq 0$, то

$$\partial_e^2 V \geq \operatorname{Tr}[(D^2\Phi)^{-1}(-\partial_e^2 D^2\Phi)] \geq d \cdot (\det((D^2\Phi)^{-1}(-\partial_e^2 D^2\Phi)))^{1/d}.$$

В силу нелинейного принципа максимума

$$\begin{aligned} \sup_{x \in A} \partial_e^2 \Phi(x) - \sup_{x \in \partial A} \partial_e^2 \Phi(x) &\leq C(d) \operatorname{diam}(A) \left(\int_{\{\partial_e^2 D^2\Phi \leq 0\}} \det(-\partial_e^2 D^2\Phi) dx \right)^{1/d} \\ &\leq C(d) \operatorname{diam}(A) \left(\int_{\{\partial_e^2 D^2\Phi \leq 0\}} \det D^2\Phi \cdot \det[(-\partial_e^2 D^2\Phi) \circ (D^2\Phi)^{-1}] dx \right)^{1/d} \\ &\leq C(d) \operatorname{diam}(A) \left(\int_A (\partial_e^2 V)_+^d e^{-V} dx \right)^{1/d}, \end{aligned}$$

что завершает доказательство.

С помощью классического принципа максимума для эллиптических уравнений доказывается известная лемма Погорелова, в которой принцип максимума применяется к функции $(x, e) \rightarrow (C - \Phi)\partial_e^2 \Phi(x)$ на множестве $S^1 \times \{\Phi \leq C\}$. В результате получаем оценку величины $\sup_x \|D^2\Phi(x)\|$ на множестве подуровня $\Omega = \{\Phi \leq C\}$ величиной, зависящей от V и ее производных до второго порядка. Следующая формулировка леммы Погорелова взята из [72].

ЛЕММА 2.2.6. Пусть $\Phi \in C^4(\Omega)$, где Ω – ограниченная область, $\Phi = 0$ на $\partial\Omega$ и $\det D^2\Phi = e^{-V}$. Тогда $-\Phi(x)\|D^2\Phi(x)\| \leq C(1 + \sup_{y \in \Omega} |\nabla\Phi(y)|^2)$, где C зависит от d , $\sup \Phi$, $\|V\|_{C^2(\Omega)}$.

Таким образом, контролируя L^∞ -норму $\|D^2\Phi\|$, можно применить классическую технику теории регулярности к дифференциальному оператору L_Φ , который будет (локально) равномерно эллиптическим. Обсудим вкратце идею Калаби, позволившую получить глубокие результаты о гладкости решений уравнения Монжа–Ампера. Для произвольной гладкой выпуклой функции Φ введем на $\mathbb{R}^d$ риманову метрику $g_{ij} = \Phi_{ij}$, где положено $\Phi_{ij} = \partial_{x_i}\partial_{x_j}\Phi$ (далее аналогично понимается Φ_{ijk}). Полученное многообразие обозначим через M . Оно принадлежит классу так называемых гессиановых многообразий, являющихся действительными аналогами кэлеровых многообразий, весьма популярных в дифференциальной геометрии и математической физике. Для вывода априорных оценок для уравнения Монжа–Ампера $\det D^2\Phi = 1$ полезно рассмотреть тензор Риччи этого многообразия. Непосредственные вычисления дают следующий результат: $4\operatorname{Ric}_{ik} = g^{jl}g^{ms}(\Phi_{mil}\Phi_{sjk} - \Phi_{mik}\Phi_{sjl})$, где

(g^{ij}) – обратная матрица к (g_{ij}) . С учетом уравнения Монжа–Ампера получаем $4 \operatorname{Ric}_{ik} = g^{jl} g^{ms} \Phi_{mil} \Phi_{sjk}$, откуда следует, что тензор Риччи многообразия M неотрицателен. Обозначим через R скалярную кривизну (свертку тензора Риччи), а через Δ_M – оператор Лапласа–Бельтрами. Основная оценка, полученная Калаби, выглядит так.

ТЕОРЕМА 2.2.7. *Выполнено неравенство $\Delta_M R \geq C(d)R^2$.*

Используя это неравенство, с помощью теорем сравнения на римановом многообразии можно получить оценки на рост R (т.е. на производные Φ третьего порядка). Некоторые обобщения этих результатов на уравнения вида $e^{-V} = e^{-W(\nabla\Phi)} \det D^2\Phi$ получены в готовящейся к печати работе второго автора, где показано, что форма Дирихле $\mathcal{E}_\Phi$, рассмотренная выше, обладает неотрицательным оператором “*carré du champ*”, если V и W выпуклы.

2.2.4. Регулярность в пространствах Гёльдера и Соболева. Как уже отмечалось, теория регулярности для уравнения Монжа–Ампера имеет долгую историю и весьма внушительный список публикаций. Первые результаты в этом направлении были получены А. Д. Александровым, который ввел понятие обобщенного по Александрову решения, и А. В. Погореловым, которому принадлежит хорошо известный сейчас пример (см. [16]) негладкого решения уравнения Монжа–Ампера с бесконечно дифференцируемой правой частью. Причина этого явления лежит в отсутствии равномерной выпуклости у решения. Однако при равномерной выпуклости удается доказать регулярность решения. Дальнейшее развитие связано с работами Н. В. Крылова, Л. Ниренберга, Дж. Спрука, Н. М. Ивочкиной, Дж. Урбаса, Н. Трудингера, Ш.-Т. Яу и др. Относительную завершенность эта теория получила в 90-х годах XX века в работах Л. Каффарелли. В обзоре [72] дано простое изложение многих результатов, ранее имевших длинные и громоздкие доказательства. Приведем два важных результата Каффарелли о регулярности (см. [72], [16], [64]).

ТЕОРЕМА 2.2.8. *Пусть $f dx$ и $g dy$ – вероятностные меры на ограниченных связных открытых множествах X и Y , $\nabla\varphi$ – соответствующая оптимальная транспортировка. Если Y выпукло, а f, g ограничены и отделены от нуля, то $\nabla\varphi$ гёльдерово некоторого порядка. Если f и g еще и гёльдеровы некоторого порядка, то $\nabla\varphi$ имеет производную, гёльдеровскую некоторого порядка.*

В [73] получены конструктивные оценки гёльдеровских норм. Недавно в работе [74] показано, что $\nabla\varphi \in W^{1,1}(X)$ в первом случае из этой теоремы. Выпуклость Y важна: известны примеры разрывных оптимальных транспортировок меры Лебега на шаре в меру Лебега на невыпуклом связном множестве с разрывным $\nabla\varphi$ (см. пример Погорелова в [16]). Если Y несвязно, то $\nabla\varphi$ обязано быть разрывным. В работе [75] непрерывность оптимальных отображений доказана для функции стоимости $|x - y|$ в предположении, что меры μ и ν заданы строго положительными непрерывными плотностями на дизъюнктивных выпуклых компактах в $\mathbb{R}^2$.

ТЕОРЕМА 2.2.9. *Пусть Ω – выпуклое множество, φ – решение уравнения $\det D^2\varphi = f$, $\varphi|_{\partial\Omega} = 0$, в смысле Александрова, причем функция f ограничена и отделена от нуля. Тогда для всякого $p > 1$ существует такое число $\varepsilon > 0$, что если $|f - 1| < \varepsilon$, то $\varphi \in W_{\text{loc}}^{2,p}(\Omega)$.*

В [76] показано, что при общей оценке $\lambda \leq f \leq \Lambda$ показатель p не может быть выбран сколь угодно большим.

2.2.5. Глобальные оценки и приложения. Конечномерная техника теории регулярности редко дает априорные оценки, не зависящие от размерности, или глобальные оценки. Оказывается, что для этих целей эффективен классический подход, основанный на дифференцировании уравнений. Приведем теорему Каффарелли о сжатии.

ТЕОРЕМА 2.2.10. Пусть $T = \nabla \Phi$ есть оптимальная транспортировка вероятностной меры $\mu = e^{-V} dx$ на $\mathbb{R}^d$ в вероятностную меру $\nu = e^{-W} dx$. Пусть V и W дважды непрерывно дифференцируемы и $D^2W \geq K$. Тогда для любого единичного вектора e имеем

$$\sup_{x \in \mathbb{R}^d} \Phi_{ee}^2(x) \leq \frac{1}{K} \sup_{x \in \mathbb{R}^d} V_{ee}(x).$$

В частности, если μ – стандартная гауссовская мера и $K \geq 1$, то T – 1-липшицево отображение.

Идея доказательства такова: пусть x_0 – точка максимума Φ_{ee} ; тогда имеем $L_\Phi \Phi_{ee}(x_0) \leq 0$ (L_Φ – эллиптический оператор); из (2.2.3) получаем $K\Phi_{ee}^2(x_0) \leq \sup_{x \in \mathbb{R}^d} V_{ee}(x)$. Заметим, что теорема из [77] несколько отличается от сформулированной. Ниже приведен оригинальный результат Каффарелли.

ТЕОРЕМА 2.2.11. Пусть $\mu = e^{-Q} dx$ – произвольная гауссовская мера. Тогда для любой меры $\nu = e^{-Q-P} dx$, где P – выпуклая функция, соответствующая оптимальная транспортировка T является 1-липшицевой.

Из этой теоремы и гауссовского изопериметрического неравенства (см. главу 3) следует такой факт.

СЛЕДСТВИЕ 2.2.12 (теорема сравнения Бакри–Леду [78]). Вероятностная мера $\mu = e^{-W} dx$, где $D^2W \geq K$, удовлетворяет изопериметрическому неравенству гауссовского типа $\mu(A^h) \geq \Phi(\Phi^{-1}(\mu(A)) + Kh)$.

Теорема сравнения Бакри–Леду утверждает, что изопериметрические свойства мер вида $\mu = e^{-W} dx$, где $D^2W \geq K$, не хуже изопериметрических свойств гауссовской меры $C_{d,K} e^{-K|x|^2/2}$. Доказательство немедленно следует из существования 1-липшицевой транспортировки меры μ в меру $C_{d,K} e^{-K|x|^2/2}$ и гауссовского изопериметрического неравенства.

ПРИМЕР 2.2.13. В [79] указано, как из результата о сжатии можно вывести известный частный случай так называемого корреляционного неравенства $\gamma(A \cap B) \geq \gamma(A)\gamma(B)$, где γ – стандартная гауссовская мера, если A – абсолютно выпуклое множество и B – эллипсоид с центром в нуле. Неизвестно, верно ли это неравенство для общего абсолютно выпуклого B .

Доказательство теоремы Каффарелли можно также получить, применяя принцип максимума не ко вторым производным, а к разности

$$\Phi(x + te) + \Phi(x - te) - 2\Phi(x).$$

Это позволяет иметь дело с негладкими потенциалами. Следующий результат доказан в [80].

ТЕОРЕМА 2.2.14. Пусть $\mu = \gamma = (2\pi)^{-d/2} e^{-|x|^2/2} dx$, $\nu = e^{-W} dx$, причем $W(x+y) + W(x-y) - W(x) \geq \delta(|y|)$ для некоторой возрастающей неотрицательной функции δ . Тогда $|\nabla\Phi(x) - \nabla\Phi(y)| \leq 8\delta^{-1}(4|x-y|^2)$.

Из этих оценок нетрудно извлечь доказательство неравенства концентрации для мер с равномерно выпуклым потенциалом для стандартной нормы (см. раздел 3.4). Несомненный интерес представляет вопрос: можно ли обобщить теорему Каффарелли на многообразия?

Заметим в связи с этой проблемой, что теорема сравнения Бакри–Леду представляет собой “плоский” аналог теоремы сравнения Леви–Громова из римановой геометрии. Более того, обе они представляют собой частные случаи теоремы сравнения для многообразий с мерой. Однако, в отличие от “плоского” случая, для теоремы Леви–Громова на настоящий момент не существует транспортного доказательства. Другая открытая проблема такова: как оценить константу сжатия, если мера-образ не является равномерно выпуклой? В частности, пусть $\nabla\Phi$ – оптимальная транспортировка стандартной гауссовской меры γ в нормированную меру Лебега на выпуклом множестве K . Имеет ли место оценка

$$\int \|D^2\Phi(x)\| d\gamma \leq C,$$

где C – константа, не зависящая от размерности (например, $C = c \operatorname{diam}(K)$ или C – универсальная константа для изотропных выпуклых множеств)?

Задача навеяна известной гипотезой Каннана–Ловаса–Шимоновича (КЛС-гипотеза). Напомним, что константой Чигера $C_{\text{Chig}}(K)$ выпуклого тела K называется наименьшая константа, для которой выполнено неравенство

$$\int_K \left| f(x) - \frac{1}{\lambda(K)} \int_K f(y) dy \right| dx \leq C \int_K |\nabla f(x)| dx \quad \forall f \in C_0^\infty(\mathbb{R}^d).$$

КЛС-ГИПОТЕЗА. Существует такая универсальная константа c , что неравенство $C_{\text{Chig}}(K) \leq c$ верно для всякого выпуклого $K \subset \mathbb{R}^d$, удовлетворяющего соотношениям

$$\int_K x_i dx = 0, \quad \frac{1}{\lambda(K)} \int_K x_i x_j dx = \delta_i^j;$$

такие тела называются изотропными.

В [80] доказано, что $\sup_x \|D^2\Phi(x)\| \leq c\sqrt{d} \operatorname{diam}(K)$, но это не дает даже оценки $C_{\text{Chig}} \leq c \operatorname{diam}(K)$. С другой стороны, известно, что для оценки константы Чигера достаточно оценить L^1 -норму $\Lambda(x) = \|D^2\Phi(x)\|$. В готовящейся к печати работе второго автора с помощью техники Калаби доказано, что

$$\int \Lambda d\gamma - \left(\int \sqrt{\Lambda} d\gamma \right)^2 \leq c \operatorname{diam}(A).$$

Приложения метрики Калаби и техники кэлеровых многообразий к так называемым оценкам тонких оболочек для выпуклых множеств получены в [81].

Перейдем к глобальным соболевским оценкам. Проинтегрировав соотношение (2.2.3), можно получить глобальные оценки на вторую и третью производные Φ . Действительно, учитывая, что

$$\int V_{ee} d\mu = \int V_e^2 d\mu, \quad \int L_\Phi \Phi_{ee} d\mu = 0,$$

получаем

$$\int V_e^2 d\mu = \int \langle D^2 W(\nabla \Phi) D^2 \Phi \cdot e, D^2 \Phi \cdot e \rangle d\mu + \int \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_e]^2 d\mu.$$

Второе слагаемое в правой части неотрицательно, поэтому при $D^2 W \geq K \cdot \text{Id}$

$$\int |\partial_e V|^2 d\mu \geq K \int |\nabla \partial_e \Phi|^2 d\mu.$$

Более общий результат получен в [82].

ТЕОРЕМА 2.2.15. *Предположим, что $D^2 W \geq K \cdot \text{Id}$, $K > 0$. Тогда для всякого единичного вектора e при $p \geq 1$ выполнены оценки*

$$K \|\Phi_{ee}^2\|_{L^p(\mu)} \leq \|(V_{ee})_+\|_{L^p(\mu)}, \quad K \|\Phi_{ee}^2\|_{L^p(\mu)} \leq \frac{p+1}{2} \|V_e^2\|_{L^p(\mu)}.$$

Более того, при $r \geq 1$ для операторной нормы $\|D^2 \Phi\|$ верно неравенство

$$K^r \int \|D^2 \Phi\|^{2r} d\mu \leq \int \|(D^2 V)_+\|^r d\mu,$$

где $(D^2 V)_+$ – положительная часть оператора $D^2 V$.

При $p \rightarrow \infty$ снова получаем оценку $K \|\Phi_{ee}\|_{L^\infty(\mu)}^2 \leq \|(V_{ee})_+\|_{L^\infty(\mu)}$ из теоремы Каффарелли. Оценки этой теоремы допускают обобщения на пространство Винера, см. раздел 3.8 и работы [83], [84], [82].

Про интегральные оценки см. [85]. Для общих функций стоимости (даже гладких) известно меньше. В [86] дан пример гладкого связного компактного многообразия в $\mathbb{R}^3$ и вероятностных мер с гладкими положительными плотностями, для которых оптимальное отображение (для обычной квадратичной функции стоимости на $\mathbb{R}^3$) разрывно (см. также [10; пример 1.3.7]). Не спасает положение даже неотрицательность секционной кривизны многообразия; оказывается, здесь играет роль так называемый тензор Ма–Трудингера–Ванга. По этим вопросам см. [87], [88], [72]. В работах [89]–[91] исследована задача переноса части массы и приведена обширная библиография.

2.3. Связи с задачей Канторовича и приближенные решения

Как уже отмечалось, задача Монжа не всегда имеет решение даже для очень простых пространств, мер и функций. Тем более удивительно, что при довольно широких предположениях эта задача имеет приближенные решения, т. е. такие преобразования $T \in T(\mu, \nu)$, для которых $M(\mu, \nu, T)$ сколь угодно мало отличается от $K(\mu, \nu)$. Иначе говоря, инфимум Монжа совпадает с минимумом Канторовича. Приведем точную формулировку. А. А. Липчюсом [92] получено усиление результата А. Прателли [93] (в свою очередь усилившего результаты У. Гангбо и Л. Амброзио) и установлен следующий факт.

ТЕОРЕМА 2.3.1. Пусть X и Y – вполне регулярные топологические пространства, в которых все компакты метризуемы, h – непрерывная функция на $X \times Y$, а μ и ν – вероятностные радоновские меры на X и Y соответственно. Пусть μ не имеет атомов. Тогда существует борелевское отображение $T: X \rightarrow Y$, переводящее μ в ν , для которого $M(\mu, \nu, T) \leq K(\mu, \nu, \mu \otimes \nu)$. Более того, справедливо равенство

$$\min_{\sigma \in \Pi(\mu, \nu)} K(\mu, \nu, \sigma) = \inf_{T \in T(\mu, \nu)} M(\mu, \nu, T).$$

Неизвестно, можно ли отказаться от метризуемости компактов в этой теореме, если рассматривать такие пары мер μ и ν , что μ можно перевести в ν . Конечно, для общих пространств отсутствие атомов у меры μ этого уже не гарантирует. Поскольку в [92] были опущены некоторые детали доказательства, мы приведем подробное обоснование (взятое из кандидатской диссертации А. А. Липчюса). Введем вспомогательные обозначения. Если μ и ν – неотрицательные радоновские меры на топологических пространствах X и Y , причем $\mu(X) = \nu(Y)$, то положим

$$\mu \boxtimes \nu := \frac{\mu \otimes \nu}{\mu(X)}.$$

Через $\pi_X \gamma$ и $\pi_Y \gamma$ будем обозначать проекции меры γ на $X \times Y$ на сомножители X и Y соответственно. Замыкание множества A в топологическом пространстве будем обозначать через $\bar{A}$. Пусть I – конечная или бесконечная двоичная (т. е. состоящая из нулей и единиц) последовательность. Тогда через I_k будем обозначать первые k символов в I , если I имеет не менее k символов. Через $l(I)$ будем обозначать число символов в конечной числовой последовательности.

Следующая лемма распространяет доказанное в [93] в случае польских пространств на вполне регулярные пространства с метризуемыми компактами.

ЛЕММА 2.3.2. Пусть μ – вероятностная радоновская мера на вполне регулярном пространстве X , в котором компакты метризуемы. Тогда для каждой двоичной последовательности I найдутся борелевские множества X_I в X и меры Радона μ_I на X со следующими свойствами:

- 1) $\|\mu_I\| = 2^{-k}$ для любой последовательности I длины k ;
- 2) $\mu_I = \mu_{(I,0)} + \mu_{(I,1)}$, причем $\mu = \sum_{l(I)=k} \mu_I$;
- 3) X_I имеет полную μ_I -меру;
- 4) если $I \leq J$, то $X_J \subseteq X_I$;
- 5) множество счетных двоичных последовательностей, для которых пересечение $\bigcap_{k \in \mathbb{N}} \bar{X}_{I_k}$ состоит более чем из одной точки, не более чем счетно;
- 6) если у μ нет атомов, то для любого натурального k множества X_I с $l(I) = k$ открыты и дизъюнкты; при этом $\mu_I = \mu|_{X_I}$.

ДОКАЗАТЕЛЬСТВО. Радоновская вероятностная мера μ на вполне регулярном пространстве X с метризуемыми компактами сосредоточена на суслинском множестве (счетном объединении метризуемых компактов). Поэтому без

потери общности можно считать, что X является вполне регулярным суслинским. Воспользуемся тем фактом, что на вполне регулярном суслинском пространстве существует счетное семейство непрерывных функций, разделяющих точки. Это означает, что всякое вполне регулярное суслинское пространство можно считать подмножеством $\mathbb{R}^\infty$ с более сильной топологией, т.е. можно считать, что на X задана метрика ϱ , причем ϱ -открытые множества являются открытыми в X . При этом граница любого множества в X -топологии лежит в границе относительно ϱ -топологии. Так как $\mathbb{R}^\infty$ сепарабельно, то в X существует счетное семейство таких точек x_n , что шарами сколь угодно малого радиуса по метрике ϱ с центрами в этих точках можно покрыть X . Построим по индукции открытые множества B_I , где I – конечная последовательность натуральных чисел. На первом шаге построим множества B_n , $n \in \mathbb{N}$. Для каждого x_n выберем $r_n \in [1/2, 1]$ так, чтобы шар с центром в x_n радиуса r_n имел ϱ - (а значит, и X -) границу μ -меры нуль. Положим

$$B_1 = B(z_1, r_1), \quad B_n = B(z_n, r_n) \setminus (\overline{B}_1 \cup \dots \cup \overline{B}_{n-1}), \quad n \geq 2.$$

Тогда $\mu(\overline{B}_n \setminus B_n) = 0$, $\mu(\bigcup B_n) = 1$. Далее действуем по индукции: если определены B_I , $l(I) \leq n-1$, то продолжим построение, заменив в первом шаге X на B_I , μ на $\mu|_{B_I}$, $\{x_n\}$ на аналогичную последовательность в B_I , $1/2$ и 1 на 2^{-n} и 2^{1-n} соответственно. Теперь построим по индукции X_I и μ_I . Индукцию будем вести по длине двоичной последовательности J . Положим $B_I^J := B_I \cap X_J$. Считаем, что $X_\emptyset = X$, $\mu_\emptyset = \mu$. Пусть построены все X_J и μ_J с $l(J) < n$. Пусть $l(J) = n$. Найдем такое j_1 , что

$$\mu_J(B_1^J \cup \dots \cup B_{j_1-1}^J) \leq 2^{-n-1} < \mu_J(B_1^J \cup \dots \cup B_{j_1}^J).$$

Положим $X_0^1 := B_1^J \cup \dots \cup B_{j_1-1}^J$, $X_1^1 := B_{j_1+1}^J \cup B_{j_1+2}^J \cup \dots$. При этом считаем, что $B_1^J \cup \dots \cup B_0^J = \emptyset$. Тогда X_0^1 и X_1^1 – открытые непересекающиеся множества с мерами, не большими 2^{-n-1} . Найдем такое j_2 , что

$$\mu_J(X_0^1 \cup B_{j_1,1}^J \cup \dots \cup B_{j_1,j_2-1}^J) \leq 2^{-n-1} < \mu_J(X_0^1 \cup B_{j_1,1}^J \cup \dots \cup B_{j_1,j_2}^J).$$

Положим

$$X_0^2 := X_0^1 \cup B_{j_1,1}^J \cup \dots \cup B_{j_1,j_2-1}^J, \quad X_1^2 := X_1^1 \cup B_{j_1,j_2+1}^J \cup B_{j_1,j_2+2}^J \cup \dots.$$

Тогда X_0^2 и X_1^2 – открытые непересекающиеся множества с мерами, не большими 2^{-n-1} . Продолжая построение, получим множества X_0^i , X_1^i с теми же свойствами. Положим $\widehat{X}_0 = \bigcup_i X_0^i$, $\widehat{X}_1 = \bigcup_i X_1^i$. Заметим, что множество

$X_J \setminus (\widehat{X}_0 \cup \widehat{X}_1) = \bigcap_{j_1, \dots, j_k} B_{j_1, \dots, j_k}^J$ состоит не более чем из одной точки. При этом $\mu(\widehat{X}_0) = 2^{-n-1}$, $\mu(\widehat{X}_1) = 2^{-n-1}$. Если μ не имеет атомов или пересечение пусто, то положим $X_{(J,0)} := \widehat{X}_0$, $X_{(J,1)} := \widehat{X}_1$ и получим открытые непересекающиеся множества меры 2^{-n-1} . При этом положим $\mu_{(J,0)} := \mu_J|_{X_{(J,0)}}$, $\mu_{(J,1)} := \mu_J|_{X_{(J,1)}}$. В противном случае положим

$$X_{(J,i)} := \widehat{X}_i \cup \{z\}, \quad \mu_{(J,i)} := \mu_J|_{\widehat{X}_i} + (2^{-n-1} - \mu_J(\widehat{X}_i))\delta_z, \quad i = 0, 1.$$

В обоих случаях $\mu_J = \mu_{(J,0)} + \mu_{(J,1)}$. Так как новые меры получаются сужением старых мер на измеримые множества и, возможно, дроблением атомов старых мер, то множества B_I сохраняют свои свойства и для построенных мер. Шаг индукции сделан. Из построения видно, что полученные множества и меры обладают нужными свойствами. Дополнительного доказательства требует лишь свойство 5).

По построению для двух конечных двоичных последовательностей J и L одинаковой длины существует не более одной счетной двоичной последовательности I с тем свойством, что $X_{I_k} \cap B_J \neq \emptyset$, $X_{I_k} \cap B_L \neq \emptyset$ для всех натуральных k . Будем обозначать эту последовательность через $I(J, L)$. Пусть $A = \{I(J, L)\}$, J и L – конечные последовательности одинаковой длины. Тогда A не более чем счетно. Покажем, что для всех остальных $I \in \{0, 1\}^{\mathbb{N}}$ множество $\bigcap_{k \in \mathbb{N}} \bar{X}_{I_k}$ состоит не более чем из одной точки. Для этого достаточно показать, что $\bigcap_{k \in \mathbb{N}} X_{I_k}$ состоит не более чем из одной точки, так как $\bar{X}_{I_k} \subset X_{I_{k-1}}$.

Для этого в свою очередь достаточно показать, что $\bigcap_{k \in \mathbb{N}} X_{I_k} \subset B_{J^n}$, где J^n – последовательность длины n , $n \in \mathbb{N}$. Докажем это по индукции. Последовательность множеств $T_k := \{m \in \mathbb{N} : X_{I_k} \cap B_m \neq \emptyset\}$ убывает. При этом пересечение не может содержать более одной точки, так как иначе имело бы место включение $I = I(j, l) \in A$, где j и l – элементы пересечения. Так как $\bigcap_k T_k = \left\{ m \in \mathbb{N} : \left(\bigcap_k X_{I_k} \right) \cap B_m \neq \emptyset \right\} = j$, то мы получили наше утверждение для $n = 1$. Переходя к B_j и заменяя в определении T_k множества B_m на $B_{j,m}$, делаем шаг индукции. Лемма доказана.

ЛЕММА 2.3.3. Пусть X и Y – вполне регулярные топологические пространства с метризуемыми компактными, h – непрерывная функция на $X \times Y$, μ и ν – вероятностные радоновские меры на X и Y соответственно. Пусть γ – вероятностная радоновская мера на $X \times Y$ с $\pi_X \gamma = \mu$, $\pi_Y \gamma = \nu$. Если μ не имеет атомов, то для всякого $\varepsilon > 0$ существуют последовательности измеримых множеств $A_n \subset X$, $B_n \subset Y$ и $X_n \subset A_n$ таких, что

- 1) $\sup_{(x_1, y_1), (x_2, y_2) \in A_n \times B_n} |h(x_1, y_1) - h(x_2, y_2)| < \varepsilon$,
- 2) $(A_n \times B_n) \cap (A_k \times B_k) = \emptyset$ при $n \neq k$,
- 3) $\gamma\left(\bigcup (A_n \times B_n)\right) = 1$,
- 4) $\mu(X_n \cap X_k) = 0$ при $n \neq k$,
- 5) $\mu(X_n) = \gamma(A_n \times B_n)$.

ДОКАЗАТЕЛЬСТВО. Так как μ и ν – радоновские меры, то существуют такие компакты $K_n^1 \subset X$, $K_n^2 \subset Y$, что $\mu(X \setminus K_n^1) < n^{-1}$, $\nu(Y \setminus K_n^2) < n^{-1}$. Можно считать, что $K_n^1 \subset K_{n+1}^1$, $K_n^2 \subset K_{n+1}^2$. Компакты K_n^1 и K_n^2 метризуемы. Обозначим их метрики через ϱ_n^1 и ϱ_n^2 . Компакт $K_n^1 \times K_n^2$ метризуем при помощи метрики

$$\varrho_n(x_1 \times y_1, x_2 \times y_2) := \max(\varrho_n^1(x_1, x_2), \varrho_n^2(y_1, y_2)).$$

Непрерывная функция h на компакте $K_n^1 \times K_n^2$ равномерно непрерывна. Пусть $\delta_n > 0$ таково, что $|h(x_1, y_1) - h(x_2, y_2)| < \varepsilon$, как только $\varrho_n(x_1 \times y_1, x_2 \times y_2) < \delta_n$

и $x_1, x_2 \in K_n^1$, $y^1, y^2 \in K_n^2$. Покроем K_n^1 конечным числом непересекающихся измеримых множеств $A_k^n \subset K_n^1$ диаметра менее δ_n . При этом, переходя к пересечению, можно считать, что если $A_k^n \cap A_l^{n-1} \neq \emptyset$, то $A_k^n \subset A_l^{n-1}$. Покроем K_n^2 конечным числом непересекающихся измеримых множеств $B_k^n \subset K_n^2$ диаметра менее δ_n . Тогда $A_k^n \times B_l^n$ – непересекающиеся прямоугольники (при фиксированном n), покрывающие $K_n^1 \times K_n^2$. При этом разброс значений h на каждом из них меньше ε . Положим

$$\begin{aligned}\tilde{A}_k^n &:= (A_k^n \cap K_n^1) \setminus K_{n-1}^1, & \tilde{B}_k^n &:= (B_k^n \cap K_n^2) \setminus K_{n-1}^2, \\ \hat{A}_k^n &:= A_k^n \cap K_{n-1}^1, & \hat{B}_k^n &:= B_k^n \cap K_{n-1}^2.\end{aligned}$$

Тогда непересекающиеся прямоугольники $\tilde{A}_k^n \times B_l^n$, $\hat{A}_k^n \times \tilde{B}_l^n$ покрывают объединение $\bigcup (K_n^1 \times K_n^2)$. При этом разброс значений h на каждом из них меньше ε . Полученный счетный набор прямоугольников занумеруем по порядку и обозначим через $A_n \times B_n$. Эти прямоугольники обладают свойствами 1), 2), 3). При этом по построению либо $A_n \cap A_k = \emptyset$, либо $A_n \subset A_k$ и $k \leq n$.

Суслинское множество $S := \bigcup_n K_n^1$ имеет полную μ -меру. Так как μ не имеет атомов, то по теореме об изоморфизме пространство с мерой (S, μ) изоморфно единичному отрезку с мерой Лебега (см. [29; гл. 9]). Поэтому во всяком измеримом множестве $A \subset S$ для всякого $t \leq \mu(A)$ можно найти измеримое подмножество B с $\mu(B) = t$.

Построим $X_n \subset A_n$ со свойствами 4), 5). Пусть X_1^1 – такое множество в A_1 , что $\mu(X_1^1) = \gamma(A_1 \times B_1)$. Найдем наименьшее k_1 такое, что $\mu(A_{k_1} \cap A_1) > 0$. Пусть $X_2^1, \dots, X_{k_1-1}^1$ – какие-либо множества в $A_2, \dots, A_{k_1-1}$ соответственно, для которых $\mu(X_2^1) = \gamma(A_2 \times B_2), \dots, \mu(X_{k_1-1}^1) = \gamma(A_{k_1-1} \times B_{k_1-1})$. Найдем в A_{k_1} множество $X_{k_1}^1$ с $\mu(X_{k_1}^1) = \gamma(A_{k_1} \times B_{k_1})$. Положим $R_1^1 = A_1 \setminus (X_1^1 \cup X_{k_1}^1)$. Пусть $X_1^2 = (X_1^1 \setminus X_{k_1}^1) \cup D_1^1$, где $D_1^1 \subset R_1^1$ и $\mu(D_1^1) = \mu(X_1^1 \cap X_{k_1}^1)$. Положим

$$X_i^2 = X_i^1, \quad i = 2, \dots, k_1.$$

Найдем наименьшее $k_2 > k_1$ такое, что $\mu(A_{k_2} \cap A_i) > 0$ для некоторого i . Пусть $X_{k_1+1}^2, \dots, X_{k_2-1}^2$ – какие-либо множества в $A_{k_1+1}, \dots, A_{k_2-1}$ соответственно, для которых $\mu(X_{k_1+1}^2) = \gamma(A_{k_1+1} \times B_{k_1+1}), \dots, \mu(X_{k_2-1}^2) = \gamma(A_{k_2-1} \times B_{k_2-1})$. Найдем в A_{k_2} множество $X_{k_2}^2$ с $\mu(X_{k_2}^2) = \gamma(A_{k_2} \times B_{k_2})$. Для $j = k_1, k_1 - 1, \dots, 1$ положим

$$R_j^2 = A_j \setminus \left(\bigcup_{l: A_l \subset A_j} X_l^2 \cup \bigcup_{l>j} X_l^3 \cup X_{k_2}^2 \right),$$

$$D_j^2 \subset R_j^2, \quad \mu(D_j^2) = \mu(X_j^2 \cap X_{k_2}^2), \quad X_j^3 = (X_j^2 \setminus X_{k_2}^2) \cup D_j^2.$$

Пусть $C_j^2 = X_j^2 \cap X_{k_2}^2$. Положим $X_i^3 = X_i^2$, $i = k_1 + 1, \dots, k_2$. Продолжая построение, получим множества X_j^i , R_j^i , D_j^i , C_j^i . При этом $R_j^k \subset R_j^l$ при $k > l$, D_j^i попарно не пересекаются. Множества C_j^i попарно не пересекаются при фиксированном j . По построению для всякого j имеем $\sum_i \mu(D_j^i) = \sum_i \mu(C_j^i)$.

Для каждого n найдем наименьшее m , для которого $n \leq k_m$. Положим

$$X_n = \left(X_n^m \cup \bigcup_{l>m} D_n^l \right) \setminus \left(\bigcup_{l>m} C_n^l \right).$$

Тогда

$$\mu(X_n) = \mu(X_n^m) + \sum_i \mu(D_n^i) - \sum_i \mu(C_n^i) = \mu(X_n^m) = \gamma(A_n \times B_n),$$

так как $\bigcup_{l>m} C_n^l \subset X_n^m \cup \bigcup_{l>m} D_n^l$. При этом $\mu(X_n \cap X_k) = 0$ при $n \neq k$. Лемма доказана.

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 2.3.1. Как отмечено при доказательстве леммы 2.3.2, без потери общности можно считать, что X и Y являются вполне регулярными суслинскими пространствами. Для доказательства первого утверждения, как и в [93], с помощью указанной леммы найдем соответствующие X_I , Y_I , μ_I и ν_I . Построим последовательность

$$\gamma_k = \sum_{I \in \{0,1\}^k} \mu_I \boxtimes \nu_{\varphi_k(I)},$$

где $\varphi_k: \{0,1\}^k \rightarrow \{0,1\}^k$ – биекция со следующими свойствами:

- 1) последовательность $K(\gamma_k)$ не возрастает,
- 2) φ_{k+1} продолжает φ_k ; более точно: при $0 \leq r \leq s$ и $I \in \{0,1\}^s$ выполнено равенство $\varphi_r(I_r) = (\varphi_s(I_s))_r$.

Последовательность φ_k построим по индукции. На первом шаге определим φ_1 так, чтобы $K(\gamma_1) \leq K(\mu \otimes \nu)$. Это можно сделать, ибо

$$2\mu \boxtimes \nu = (\mu_0 \boxtimes \nu_0 + \mu_1 \boxtimes \nu_1) + (\mu_0 \boxtimes \nu_1 + \mu_1 \boxtimes \nu_0).$$

Следовательно, $K(\mu \boxtimes \nu) = 2^{-1}(K(\mu_0 \boxtimes \nu_0 + \mu_1 \boxtimes \nu_1) + K(\mu_0 \boxtimes \nu_1 + \mu_1 \boxtimes \nu_0))$. В последней формуле либо первое, либо второе слагаемое не превосходит величину $K(\mu \boxtimes \nu)$. Если это первое слагаемое, то положим $\varphi_1(0) = 0$, $\varphi_1(1) = 1$. Если это второе слагаемое, то положим $\varphi_1(0) = 1$, $\varphi_1(1) = 0$. Пусть $\varphi_1, \dots, \varphi_{k-1}$ определены. Пусть $l(I) = k$. Тогда $\varphi_k(I) = (\varphi_{k-1}(I_{k-1}), x)$, где x выбирается так, чтобы выполнялось свойство 1). Это всегда можно сделать. Для доказательства достаточно на первом шаге заменить μ и ν на $\mu_{I_{k-1}}$ и $\nu_{\varphi_{k-1}(I_{k-1})}$ и воспользоваться разложением

$$\mu_{I_{k-1}} = \mu_{(I_{k-1},0)} + \mu_{(I_{k-1},1)}, \quad \nu_{\varphi_{k-1}(I_{k-1})} = \nu_{(\varphi_{k-1}(I_{k-1}),0)} + \nu_{(\varphi_{k-1}(I_{k-1}),1)}.$$

Таким образом, определив φ_k , мы получили последовательность радоновских мер γ_k , проекции которых на X и Y суть μ и ν . Следовательно, эти меры равномерно плотны. По усиленной теореме Прохорова семейство $\{\gamma_k\}$ секвенциально слабо компактно (см. теорему 8.6.7 в [29]). Значит, можно выбрать слабо сходящуюся подпоследовательность. Ее предел обозначим через γ . Проекция γ на X и Y суть μ и ν , причем $K(\gamma) \leq K(\mu \otimes \nu)$. Теперь докажем, что носитель γ – график некоторого отображения. Рассмотрим множество

$$Z := \bigcap_{k \in \mathbb{N}} \bigcup_{l(I)=k} X_I$$

и функцию $\psi: Z \rightarrow \{0,1\}^{\mathbb{N}}$, которая ставит в соответствие каждому x из Z единственную счетную двоичную последовательность I со свойством $x = \bigcap_{k \in \mathbb{N}} X_{I_k}$.

Заметим, что Z имеет полную μ -меру. Обозначим через A множество таких счетных последовательностей I , что $\bigcap_{k \in \mathbb{N}} \bar{Y}_{I_k}$ содержит более одной точки. То-

гда по свойству 5) из леммы 2.3.3 получаем, что A не более чем счетно. Пусть $\varphi: \{0, 1\}^{\mathbb{N}} \rightarrow \{0, 1\}^{\mathbb{N}}$ есть продолжение всех φ_k . Из определения φ_k получаем, что φ – биекция. Так как μ не имеет атомов, то множество $\tilde{X} = \{x \in Z: \varphi(\psi(x)) \notin A\}$ имеет полную μ -меру. Множество $S_k = \bigcup \{X_I \times \bar{Y}_{\varphi_k(I)}: I \in \{0, 1\}^k\}$ имеет полную γ_k -меру в силу того, что $\mu(X_I) = \mu(\bar{X}_I)$. При этом S_k имеет полные γ_n -меры для всех $n \geq k$. Покажем, что S_k имеет полную γ -меру. Действительно,

$$\begin{aligned} \gamma(X_I \times \bar{Y}_{\varphi_k(I)}) &= \gamma(\bar{X}_I \times \bar{Y}_{\varphi_k(I)}) \geq \limsup_{n \rightarrow \infty} \gamma_n(\bar{X}_I \times \bar{Y}_{\varphi_k(I)}) \\ &= \limsup_{n \rightarrow \infty} \gamma_n(\bar{X}_I \times Y) = \mu(\bar{X}_I) = \mu(X_I) \\ &= \gamma(X_I \times Y) \geq \gamma(X_I \times \bar{Y}_{\varphi_k(I)}). \end{aligned}$$

Положим $S = \bigcap_k S_k$. Тогда S является борелевским множеством в $\tilde{X} \times Y$ и $\gamma(S) = 1$. Для всякого элемента $x \in \tilde{X}$ множество $\{y: (x, y) \in S\} = \bar{Y}_{\varphi(\psi(x))}$ содержит не более одной точки. Множество $(\tilde{X} \times Y) \cap S$ – борелевское. Его проекция Π на $\tilde{X}$ является суслинским множеством, причем

$$\mu(X \setminus \Pi) = \gamma((X \setminus \Pi) \times Y) = 0,$$

так как $\gamma(S) = 1$. Для всякого $x \in \Pi$ существует единственный элемент $y \in Y$ с $(x, y) \in S$, т. е. суслинское множество $S \cap (\Pi \times Y)$ представляет собой график некоторого отображения $T: \Pi \rightarrow Y$. Как известно (см. лемму 6.7.1 в [29]), T является борелевским отображением из Π в Y . Доопределяя его постоянным значением на дополнении к Π , получим борелевское отображение из X в Y с требуемыми свойствами. Первое утверждение теоремы доказано.

Перейдем ко второму утверждению. Пусть γ – мера на $X \times Y$, на которой достигается минимум. Зафиксируем $\varepsilon > 0$. Воспользуемся леммой 2.3.3 и возьмем указанные там множества A_n , B_n и X_n . По свойству 3) прямоугольники $A_n \times B_n$ покрывают γ -почти все $X \times Y$. На каждом из них разброс значений h меньше ε по свойству 1). По свойству 4) имеем $\mu(X_n) = \gamma(A_n \times B_n) =: m_n$. Введем меры $\gamma_n := \gamma|_{A_n \times B_n}$, $\tilde{\gamma}_n := \mu|_{X_n} \boxtimes \pi_Y \gamma_n$. Проекция меры $\tilde{\gamma}_n$ суть $\mu|_{X_n}$ и $\pi_Y \gamma_n$. Применим к ним первое утверждение теоремы. Получим отображение $t_n: X_n \rightarrow B_n$ со свойством $M(t_n) \leq K(\tilde{\gamma}_n)$. Пусть $T(x) = t_n(x)$ при $x \in X_n$. Поскольку $\bigcup_n X_n$ имеет полную μ -меру (это следует из свойств 3) и 5)), то

отображение T переводит μ в меру $\sum_n \pi_Y \gamma_n = \pi_Y \sum_n \gamma_n = \pi_Y \gamma = \nu$. Так как $|K(\tilde{\gamma}_n) - K(\gamma_n)| \leq m_n \varepsilon$, то окончательно получаем

$$M(T) \leq \sum_n M(t_n) \leq \sum_n K(\tilde{\gamma}_n) \leq \sum_n (K(\gamma_n) + m_n \varepsilon) \leq K(\gamma) + \varepsilon \|\gamma\|.$$

Теорема полностью доказана.

Непрерывность функции стоимости существенна в доказанной теореме и не может быть заменена даже на полунепрерывность снизу. Приведем пример, заимствованный из [93].

ПРЕДЛОЖЕНИЕ 2.3.4. (i) Пусть $X = Y = [0, 1] \times [-1, 1]$, μ – линейная мера Лебега на отрезке $I_0 = [0, 1] \times \{0\}$, ν – умноженная на $1/2$ линейная мера Лебега на объединении отрезков $I_1 = [0, 1] \times \{1\}$ и $I_2 = [0, 1] \times \{-1\}$, $h(x, y) = 0$, если $\|x - y\| = 1$, $h(x, y) = 1$ в остальных случаях, т. е. h – индикатор открытого множества $U = \{(x, y) \in X \times Y : \|x - y\| \neq 1\}$ – дополнения замкнутого множества $Z = \{(x, y) \in X \times Y : \|x - y\| = 1\}$. Тогда задача Канторовича и задача Монжа имеют решения, но они различны и $K(\mu, \nu) = 0$, $M(\mu, \nu) = 1$. При этом задача Канторовича имеет единственное решение.

(ii) Пусть $T(x_1, x_2) = (2x_1, 1)$ при $0 \leq x_1 \leq 1/2$, $T(x_1, x_2) = (2x_1 - 1, -1)$ при $1/2 < x_1 \leq 1$. Переопределим h на графике T , положив на нем $h = 1/2$. Тогда сохраняют силу утверждения из (i), но при этом и задача Монжа будет иметь единственное решение.

ДОКАЗАТЕЛЬСТВО. (i) Решением задачи Канторовича является мера $\hat{\sigma}$, равная умноженной на $1/2$ линейной мере Лебега на объединении диагональных отрезков

$$D_1 = \{(x_1, 0, x_1, 1) : 0 \leq x_1 \leq 1\}, \quad D_2 = \{(x_1, 0, x_1, -1) : 0 \leq x_1 \leq 1\}.$$

Ее проекции суть μ и ν , $K(\mu, \nu, \sigma) = 0$, так как D_1, D_2 лежат в Z . Других мер σ в $\Pi(\mu, \nu)$ с таким же значением нет, ибо должно выполняться равенство $\sigma(Z) = 1$, а также равенства $\sigma(X \times (I_1 \cup I_2)) = 1$, $\sigma(I_0 \times Y) = 1$. Иначе говоря, мера σ должна быть сосредоточена на множестве точек вида $(x_1, 0, y_1, y_2)$, где $|y_2| = 1$ и $(x_1 - y_1)^2 + y_2^2 = 1$, т. е. $x_1 = y_1$. Тем самым, σ сосредоточена на $D_1 \cup D_2$, что с учетом совпадения проекции σ на первый сомножитель с мерой μ делает неизбежным равенство $\sigma = \hat{\sigma}$.

Обратимся к задаче Монжа. Одним из ее решений (но не единственным) оказывается отображение T , указанное в (ii), для которого $M(\mu, \nu, T) = 1$, ибо ни для какой транспортировки S меры μ в ν нельзя получить меньшего значения. Чтобы это увидеть, заметим, что если $\mu \circ S^{-1} = \nu$, то $\|x - S(x)\| \neq 1$ при μ -почти всех x . В самом деле, если $x = (x_1, 0) \in I_0$ таково, что $S(x) \in I_1 \cup I_2$ (таковы μ -почти все x), то равенство $\|x - S(x)\| = 1$ возможно лишь при $S(x) = (x_1, y_2)$, где $|y_2| = 1$. Таким образом, если множество $E = \{x : \|x - S(x)\| = 1\}$ имеет положительную меру, то μ -почти каждая его точка имеет вид $(x_1, 0)$ и сдвигается либо на 1 вверх, либо на 1 вниз. Поэтому можно считать, что $E \subset I_0$ и каждая точка E сдвигается под действием S на 1 вверх или вниз. Это ведет к противоречию с тем, что ν – образ μ : скажем, если положительную меру имеет множество E_1 точек из E , сдвигающихся вверх, то $\nu(E_1 \times \{1\}) = \mu(E_1)/2$, хотя $\nu(E_1) = \mu(S^{-1}(E_1 \times \{1\})) \geq \mu(E_1)$.

(ii) Предыдущие рассуждения сохраняют силу и для новой функции h , но теперь и у задачи Монжа решение единственно. В самом деле, из этих рассуждений видно, что если $M(\mu, \nu, S) \leq M(\mu, \nu, T)$, то $S(x) = T(x)$ при μ -почти всех x . Предложение доказано.

Используя изоморфизм пространств с безатомическими борелевскими мерами, из этого примера легко получить пример функции стоимости h , являющейся индикатором борелевского множества в $[0, 1] \times [0, 1]$, для которой задачи Канторовича и Монжа на $[0, 1]$ с мерами μ и ν , равными мере Лебега, имеют несовпадающие единственные решения.

Приведем также пример, поясняющий разницу в вопросе единственности между функциями стоимости $|x - y|$ и $|x - y|^2$.

ПРИМЕР 2.3.5. Пусть $X = Y = \mathbb{R}$, $h(x, y) = |x - y|$, μ – вероятностная мера на $(0, +\infty)$ с первым моментом, ν – симметричная ей мера на $(-\infty, 0)$. Тогда каждая мера $\sigma \in \Pi(\mu, \nu)$ дает решение задачи Канторовича, поскольку

$$\int |x - y| \sigma(dx) = \int (x - y) \sigma(dx) = 2 \int_{[0, +\infty)} x \mu(dx),$$

ибо σ сосредоточена на нижнем правом квадранте. Такое же значение дает и всякое отображение из μ в ν , которых также много в общем случае (одно из них – отражение).

О связи задач Монжа и Канторовича см. также [5], где есть замечание о сведении задачи Монжа к случаю сохраняющих меру отображений (что делает ее вариационной задачей на группе автоморфизмов). Отметим, что еще в [18] А. М. Вершик ставил задачу минимизации интеграла по мере μ от длины кривой $\{T_t(x)\}_{0 \leq t \leq 1}$ в подходящем классе отображений T_t , для которых $T_0(x) = x$ и T_1 переводит μ в ν .

Глава 3. Приложения

Теперь обсудим приложения транспортной задачи к классическим неравенствам анализа, теории вероятностей и дифференциальной геометрии, а также одну из важнейших идей, на которой основаны многие приложения, – связь геодезических в пространствах мер с задачей Монжа–Канторовича. Далее используются следующие обозначения. Если μ и ν – вероятностные меры на $(X, \mathcal{A})$ и ν имеет плотность ϱ относительно μ , т. е. $\nu = \varrho \cdot \mu$, то энтропию ν (или ϱ) относительно μ зададим формулой

$$\text{Ent}_\mu \nu = \text{Ent}_\mu \varrho = \int_X \varrho \ln \varrho d\mu,$$

если $\varrho \ln \varrho \in L^1(\mu)$; в остальных случаях положим $\text{Ent}_\mu \nu := +\infty$. По неравенству Йенсена энтропия неотрицательна.

3.1. Изопериметрические неравенства и неравенство Брунна–Минковского

Здесь изложено доказательство классического изопериметрического неравенства, принадлежащее М. Грому (см. [94; Appendix]). Он нашел это доказательство, используя треугольные отображения, обсуждаемые в последнем разделе статьи. Те же рассуждения остаются в силе в случае оптимальных

отображений. Известно, впрочем, что изопериметрическое неравенство следует из неравенства Брунна–Минковского, о котором говорится ниже.

ТЕОРЕМА 3.1.1. *Пусть $A \subset \mathbb{R}^d$ – борелевское множество. Тогда выполнено следующее изопериметрическое неравенство:*

$$\lambda^{1-1/d}(A) \leq \kappa_d \mathcal{H}^{d-1}(\partial A), \quad \kappa_d = \frac{\Gamma(1+d/2)^{1/d}}{d\sqrt{\pi}}.$$

ДОКАЗАТЕЛЬСТВО. Рассмотрим шар $B_r = \{x: |x| \leq r\}$, удовлетворяющий условию $\lambda(A) = \lambda(B_r)$. Пусть $T = \nabla W$ – оптимальная транспортировка, отображающая $\lambda|_A$ в $\lambda|_{B_r}$. Применим формулу замены переменных $\det D_a^2 W = 1$ на A . В силу выпуклости W имеем $\Delta_a W dx \leq \Delta W$, где под ΔW понимается лапласиан в смысле обобщенных функций, т. е. мера. Пусть λ_i – собственные числа $D_a^2 W$. В силу неравенства между средним арифметическим и средним геометрическим имеем $\Delta_a W = \sum_{i=1}^d \lambda_i \leq d(\lambda_1 \cdots \lambda_d)^{1/d} = d$. Таким образом,

$$d\lambda(A) \leq \int_A \Delta_a W dx \leq \Delta W(A).$$

Считая, что A имеет достаточно регулярную границу и интегрируя по частям, получаем, что последняя величина равна

$$\int_{\partial A} \langle \nabla W, n_A \rangle d\mathcal{H}^{d-1} \leq r \mathcal{H}^{d-1}(\partial A).$$

Здесь n_A – единичная нормаль к ∂A . Искомое неравенство следует из равенства $\lambda(A) = \lambda(B_r) = \pi^{d/2} r^d / \Gamma(1+d/2)$. Теорема доказана.

Из доказательства видно, что неравенства становятся равенствами, когда A есть сдвиг шара B_r . Итак, шар имеет наименьшую поверхность среди множеств фиксированной меры Лебега.

ЗАМЕЧАНИЕ 3.1.2. Пусть H_d – d -мерное пространство Лобачевского. Рассмотрим модель Пуанкаре $H_d = \mathbb{R}^{d-1} \times \mathbb{R}^+$ с метрикой $g = y_d^{-2} dy_1^2 \cdots dy_d^2$. Используя оптимальную транспортировку (в стандартном “евклидовом” смысле), можно доказать изопериметрическое неравенство

$$\nu^+(\partial A) \geq \max \left[\frac{\nu^{1-1/d}(A)}{\kappa_d}, (d-1)\nu(A) \right],$$

где $\nu = y_d^{-d} I_{\{y_d > 0\}} dy_1 \cdots dy_d$ – риманов объем, $\nu^+ = y_d^{-d+1} \cdot \mathcal{H}^{d-1}$ – соответствующая поверхностная мера.

Обратимся к неравенству Брунна–Минковского. Относительно недавно была обнаружена следующая изящная форма этого классического неравенства.

ТЕОРЕМА 3.1.3. *Пусть f, g, h – такие неотрицательные функции из класса $L^1(\mathbb{R}^d)$, что $h(\lambda x + (1-\lambda)y) \geq f^\lambda(x)g^{1-\lambda}(y)$ при всех $x, y \in \mathbb{R}^d$ и некотором $\lambda \in [0, 1]$. Тогда*

$$\int_{\mathbb{R}^d} h dx \geq \left(\int_{\mathbb{R}^d} f dx \right)^\lambda \left(\int_{\mathbb{R}^d} g dx \right)^{1-\lambda}.$$

ДОКАЗАТЕЛЬСТВО. Мы можем считать, что $\|f\|_{L^1(\mathbb{R}^d)} = \|g\|_{L^1(\mathbb{R}^d)} = 1$. Пусть $\nabla\varphi_f$ и $\nabla\varphi_g$ – оптимальные транспортировки меры Лебега на $[0, 1]^d$ в меры $f dx$ и $g dx$ соответственно. Формула замены переменных принимает вид

$$f(\nabla\varphi_f) \det D_a^2\varphi_f = 1, \quad g(\nabla\varphi_g) \det D_a^2\varphi_g = 1,$$

Положим $\varphi = (1 - \lambda)\varphi_f + \lambda\varphi_g$. По формуле замены переменных и известной оценке для определителя $\det((1 - \lambda)M_1 + \lambda M_2) \geq (\det M_1)^{1-\lambda}(\det M_2)^\lambda$, верной для неотрицательных симметричных матриц M_1, M_2 , имеем

$$\begin{aligned} \int_{\mathbb{R}^d} h dx &= \int_{[0,1]^d} h(\nabla\varphi) \det D_a^2\varphi dx \\ &\geq \int_{[0,1]^d} h((1 - \lambda)\varphi_f + \lambda\varphi_g) (\det D_a^2\varphi_f)^{1-\lambda} (\det D_a^2\varphi_g)^\lambda dx \\ &\geq \int_{[0,1]^d} f^{(1-\lambda)}(\nabla\varphi_f) g^\lambda(\nabla\varphi_g) (\det D_a^2\varphi_f)^{1-\lambda} (\det D_a^2\varphi_g)^\lambda dx = 1. \end{aligned}$$

Теорема доказана.

Применив неравенство к индикаторам множеств A и B , получаем такую форму неравенства Брунна–Минковского:

$$\mathcal{H}^d((1 - \lambda)A + \lambda B) \geq [\mathcal{H}^d(A)]^\lambda [\mathcal{H}^d(B)]^{1-\lambda}.$$

Из нее подбором λ выводится классическое неравенство Брунна–Минковского

$$[\mathcal{H}^d(A + B)]^{1/d} \geq [\mathcal{H}^d(A)]^{1/d} + [\mathcal{H}^d(B)]^{1/d}.$$

3.2. Неравенства Соболева и их обобщения

3.2.1. Классические неравенства Соболева. Хорошо известно, что из изопериметрического неравенства следуют неравенства Соболева. Этот факт давно был замечен независимо в работах В. Г. Мазыи, а также Г. Федерера и В. Флеминга. Если $f = \sum_{i=1}^n \lambda_i I_{A_i}$ – линейная комбинация индикаторов борелевских множеств ограниченного периметра с дизъюнктными границами, то $\|f\|_{L^{d/(d-1)}} \leq \sum_{i=1}^n \lambda_i \|I_{A_i}\|_{L^{d/(d-1)}} \leq \kappa_d \sum_{i=1}^n \lambda_i \mathcal{H}^{d-1}(\partial A_i) = \kappa_d \|\nabla f\|_{L^1}$ по изопериметрическому неравенству. На произвольную функцию $f \in W^{1,1}(\mathbb{R}^d)$ это неравенство распространяется приближением простыми функциями по норме из пространства BV функций ограниченной вариации. Из него следуют вложения Соболева $W^{p,1} \subset L^{dp/(d-p)}$. В работе [95] было найдено непосредственное транспортное доказательство классических неравенств Соболева. Из доказательства автоматически следуют формулы для минимизирующих функций (тех функций, для которых эти неравенства превращаются в точные равенства). При этом были получены интересные соотношения двойственности.

ТЕОРЕМА 3.2.1. Пусть $f \in W^{p,1}(\mathbb{R}^d)$, $p^* = dp/(d - p)$. Тогда

$$\|\nabla f\|_{L^p} \geq \|f\|_{L^{p^*}} \|\nabla h_p\|_{L^p}, \quad \text{где } h_p(x) = (\sigma_p + |x|^{p/(p-1)})^{1-d/p},$$

причем σ_p удовлетворяет условию нормировки $\|h_p\|_{L^{p^*}} = 1$. Равенство выполнено только для функций вида $f(x) = Ch_p(x - x_0)$.

Доказательство можно найти в [95] и [16]. В [96] транспортным методом доказаны неравенства для следов на полупространствах

$$\|f\|_{L^{p^+}(\partial H)} \leq T_p(d) \|\nabla f\|_{L^p(H)}.$$

В [97] получены оптимальные неравенства для следов соболевских функций, в частности неравенства вида

$$\|f\|_{L^{p^*}(\Omega)} \leq S_p(d) \|\nabla f\|_{L^p(\Omega)} + C(p, \Omega) \|f\|_{L^{p^+}(\partial\Omega)},$$

где f – соболевская функция в некоторой области $\Omega \subset \mathbb{R}^d$ с липшицевой границей, $p^+ = (d - 1)p/(d - p)$. При этом наибольшее значения константы $C(p, \Omega)$ достигается на шарах. Обобщения ряда классических неравенств (Гальярдо–Ниренберга, Фабера–Крана, Мозера–Трудингера, евклидова логарифмического неравенства Соболева) на случай следов получены в [97]. В работе [98] с помощью метода Громова найдены усиления классических версий изопериметрического неравенства и неравенства Брунна–Минковского. Надо отметить, что транспортные соображения далеко не всегда хорошо работают на многообразиях. Тем не менее в [99]–[101] получены интересные обобщения неравенства Брунна–Минковского, а также изопериметрических и соболевских неравенств для многообразий с помощью транспортного метода.

Есть интересная связь между расстоянием Канторовича и двойственной соболевской нормой $H^{-1}(\mu)$ относительно меры μ

$$\|u\|_{H^{-1}(\mu)} = \sup \left\{ \int u \varphi d\mu; \varphi \in C_0^\infty(\mathbb{R}^n); \int_{\mathbb{R}^n} |\nabla \varphi|^2 d\mu \leq 1 \right\}.$$

Пусть u – ограниченная борелевская функция с нулевым интегралом по мере μ и $\mu_\varepsilon = \mu + \varepsilon u dx$. Тогда

$$\|u\|_{H^{-1}(\mu)} = \liminf_{\varepsilon \rightarrow 0} \varepsilon^{-1} W_2(\mu, \mu_\varepsilon).$$

Интересные приложения этого факта к выпуклой геометрии см. в [81].

3.2.2. Логарифмические неравенства Соболева. Для вероятностных мер есть аналоги неравенств Соболева. Наиболее известным из них является логарифмическое неравенство Соболева, доказанное Л. Гроссом в [102] для стандартной гауссовской меры на $\mathbb{R}^d$ с плотностью $(2\pi)^{-d/2} e^{-|x|^2/2}$. А именно, для гладкой функции f выполнено следующее неравенство:

$$\text{Ent}_\gamma f^2 := \int_{\mathbb{R}^d} f^2 \ln f^2 d\gamma - \int_{\mathbb{R}^d} f^2 d\gamma \ln \int_{\mathbb{R}^d} f^2 d\gamma \leq 2 \int_{\mathbb{R}^d} |\nabla f|^2 d\gamma. \quad (3.2.1)$$

Логарифмическое неравенство Соболева является естественным аналогом классических неравенств Соболева. Как и тесно связанное с ним свойство гиперсжимаемости диффузионных полугрупп, оно имеет много приложений в математической физике, геометрии, теории вероятностей. Имеется обширная литература по этому вопросу, см. [103], [16], [17], [104]. Полное описание мер,

удовлетворяющих логарифмическому неравенству Соболева на прямой, дано в [105]. В работе Г. Я. Перельмана [106] с помощью логарифмического неравенства Соболева был доказан один из ключевых моментов работы – монотонность так называемого $\mathcal{W}$ -функционала. На первый взгляд неравенство (3.2.1) кажется значительно более слабым, чем классические неравенства Соболева, так как в левой части стоит значительно медленнее растущая функция, чем в аналогичном неравенстве Соболева. Однако его важным преимуществом является то, что (3.2.1) не зависит от размерности, что позволяет применять его в бесконечномерном случае.

Транспортное доказательство (3.2.1) было впервые получено в работе [79]. Предварительно заметим, что неравенство не меняется при умножении f на константу. Поэтому достаточно доказать (3.2.1) для функции f с единичной нормой в $L^2(\gamma)$. Рассмотрим оптимальную транспортировку $T = \nabla\varphi$ меры $f^2 \cdot \gamma$ в меру γ . Будем считать, что функция f достаточно гладкая, ограничена и отделена от нуля, что влечет гладкость T . По формуле замены переменных $f^2(x)e^{-|x|^2/2} = e^{-|\nabla\varphi(x)|^2/2} \det D^2\varphi(x)$, т.е. $\ln f^2(x) - |x|^2/2 = -|\nabla\varphi(x)|^2/2 + \ln \det D^2\varphi(x)$. Перепишем полученное равенство в виде

$$\ln f^2(x) = -\frac{|x - \nabla\varphi(x)|^2}{2} - \langle x, \nabla\varphi(x) - x \rangle + \ln \det D^2\varphi(x)$$

и проинтегрируем его по мере $f^2 \cdot \gamma$, что даст соотношение

$$\begin{aligned} \int_{\mathbb{R}^d} f^2 \ln f^2 d\gamma &= - \int_{\mathbb{R}^d} \frac{|x - \nabla\varphi(x)|^2}{2} f^2(x) \gamma(dx) \\ &\quad - \int_{\mathbb{R}^d} \langle x, \nabla\varphi(x) - x \rangle f^2(x) \gamma(dx) + \int_{\mathbb{R}^d} \ln \det D^2\varphi f^2 d\gamma. \end{aligned}$$

Применим интегрирование по частям:

$$\begin{aligned} &- \int_{\mathbb{R}^d} \langle x, \nabla\varphi(x) - x \rangle f^2(x) \gamma(dx) \\ &= 2 \int_{\mathbb{R}^d} \langle \nabla f(x), \nabla\varphi(x) - x \rangle f(x) \gamma(dx) - \int_{\mathbb{R}^d} (\Delta\varphi - d) f^2 d\gamma. \end{aligned}$$

Для симметричной $(d \times d)$ -матрицы $A \geq 0$ имеем $\text{Tr } A - d - \ln \det A \geq 0$ (это неравенство доказывается приведением матрицы к диагональному виду и применением неравенства $x \ln x \geq x - 1$). Таким образом,

$$\int_{\mathbb{R}^d} f^2 \ln f^2 d\gamma \leq \int_{\mathbb{R}^d} \left[-\frac{|x - \nabla\varphi(x)|^2}{2} f^2(x) + 2\langle \nabla f(x), \nabla\varphi(x) - x \rangle f(x) \right] \gamma(dx).$$

Неравенство (3.2.1) следует теперь из очевидной оценки

$$-\frac{|x - \nabla\varphi(x)|^2}{2} f^2(x) + 2\langle \nabla f(x), \nabla\varphi(x) - x \rangle f(x) \leq 2|\nabla f(x)|^2.$$

Как и в случае неравенства Соболева, существует более сильное изопериметрическое неравенство для гауссовских мер

$$\gamma(A^h) \geq \Phi(\Phi^{-1}(\gamma(A)) + h), \quad A^h = \{x \in \mathbb{R}^d \mid \exists a \in A : |a - x| < h\},$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

В дифференциальной форме оно выглядит следующим образом:

$$\gamma^+(\partial A) := \lim_{h \rightarrow 0} \frac{\gamma(A^h) - \gamma(A)}{h} \geq \frac{1}{[\Phi^{-1}]'(\gamma(A))},$$

где $\gamma^+(\partial A)$ есть поверхностная мера A для γ . Несложно убедиться, что это неравенство качественно эквивалентно следующему неравенству:

$$\gamma^+(\partial A) \geq C \tilde{\gamma}(A) \sqrt{-\ln \tilde{\gamma}(A)}, \quad \tilde{\gamma}(A) = \min(\gamma(A), 1 - \gamma(A)),$$

где C – универсальная константа. Изопериметрическое неравенство было получено В. Н. Судаковым и Б. С. Цирельсоном [107] и позже К. Бореллем [108]. Существует несколько доказательств этого неравенства. Доказательство в [107] было получено как следствие изопериметрического неравенства Леви–Громова на сфере. Кроме того, гауссовские меры – один из редких примеров, в которых известно точное решение изопериметрической задачи. Множествами наименьшей поверхностной меры являются полупространства $\{x: \langle h, x \rangle \leq a\}$. Неизвестно, возможно ли транспортное доказательство изопериметрического гауссовского неравенства.

Гауссовские меры входят в важный класс вероятностных мер с интересными аналитическими и геометрическими свойствами, образованный выпуклыми (называемыми также логарифмически вогнутыми) мерами. Их роль в анализе аналогична роли, которую играют в геометрии многообразия положительной кривизны Риччи. Подробнее о выпуклых мерах написано в книге [67], здесь лишь кратко описаны их основные свойства, важные для дальнейшего. Выпуклой (или логарифмически вогнутой) мерой на $\mathbb{R}^d$ называется вероятностная мера μ , удовлетворяющая неравенству

$$\mu(\lambda A + (1 - \lambda)B) \geq \mu^\lambda(A) \mu(B)^{1-\lambda}$$

для всех непустых борелевских множеств A и B и чисел $\lambda \in [0, 1]$. Согласно известному результату К. Борелля, выпуклая мера сосредоточена на некотором аффинном подпространстве, на котором она обладает плотностью ϱ (относительно соответствующей меры Лебега) вида $\varrho = e^{-V}$, где V – выпуклая функция. Класс выпуклых мер инвариантен относительно аффинных отображений. Как мы увидим ниже, значительное число известных результатов, связанных с оптимальными отображениями и функциональными неравенствами, относится именно к выпуклым мерам.

Существуют различные достаточные условия для того, чтобы вероятностная мера μ удовлетворяла логарифмическому неравенству Соболева. Необходимым условием является выполнение неравенства Пуанкаре.

ОПРЕДЕЛЕНИЕ 3.2.2. Говорят, что вероятностная мера μ удовлетворяет неравенству Пуанкаре, если существует такое $C_p \geq 0$, что для всех гладких функций f верна оценка

$$\mathrm{Var}_\mu f := \int_{\mathbb{R}^d} \left(f - \int_{\mathbb{R}^d} f d\mu \right)^2 d\mu \leq \frac{1}{C_p} \int_{\mathbb{R}^d} |\nabla f|^2 d\mu.$$

Можно проверить, что логарифмическое неравенство Соболева влечет неравенство Пуанкаре, применив (3.2.1) к $1 + \varepsilon f$ и перейдя к пределу при $\varepsilon \rightarrow 0$.

Другим (куда менее очевидным) необходимым условием является существование такого числа $\varepsilon > 0$, что

$$\int_{\mathbb{R}^d} \exp(\varepsilon x^2) \mu(dx) < +\infty. \quad (3.2.2)$$

Следующее обобщение гауссовского неравенства Соболева получено в [109], а транспортное доказательство предложено в [110].

ТЕОРЕМА 3.2.3. Пусть $\mu = e^{-V} dx$ – вероятностная мера на $\mathbb{R}^d$ с дважды непрерывно дифференцируемой функцией V . Предположим, что для некоторого $\lambda \geq 0$ мы имеем $D^2V \geq -\lambda \text{Id}$, при некотором $\varepsilon > 0$ верно неравенство

$$\int_{\mathbb{R}^{2d}} \exp\left(\frac{\lambda + \varepsilon}{2} |y - x|^2\right) \mu(dx) \mu(dy) < \infty$$

и μ удовлетворяет неравенству Пуанкаре. Тогда найдется такое $C > 0$, что

$$\text{Ent}_\mu f^2 \leq C \int_{\mathbb{R}^d} |\nabla f|^2 d\mu.$$

ДОКАЗАТЕЛЬСТВО. Докажем сначала, что есть такие $C_1, C_2 > 0$, что

$$\text{Ent}_\mu f^2 \leq C_1 \int_{\mathbb{R}^d} |\nabla f|^2 d\mu + C_2 \int_{\mathbb{R}^d} f^2 d\mu. \quad (3.2.3)$$

Без ограничения общности можно считать, что f имеет единичную норму в $L^2(\mu)$. Пусть $\nabla\varphi$ – оптимальная транспортировка меры $f^2 \cdot \mu$ в меру μ . Как и в случае гауссовских мер, получаем $\ln f^2 = V - V(\nabla\varphi) + \ln \det D_a^2\varphi$. Из неравенства $V(x) - V(y) - \langle \nabla V(x), x - y \rangle \leq \lambda |x - y|^2 / 2$ следует оценка

$$\ln |f(x)|^2 \leq \langle \nabla V(x), x - \nabla\varphi(x) \rangle + \ln \det D_a^2\varphi(x) + \frac{\lambda |x - \nabla\varphi(x)|^2}{2}.$$

Проинтегрируем полученную оценку по мере $f^2 \cdot \mu = f^2 \cdot e^{-V} dx$ и применим интегрирование по частям. Это дает

$$\begin{aligned} \int_{\mathbb{R}^d} f^2 \ln f^2 d\mu &\leq \int_{\mathbb{R}^d} (d - \text{Tr } D_a^2\varphi + \ln \det D_a^2\varphi) f^2 d\mu \\ &\quad + \frac{\lambda}{2} \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 f(x)^2 \mu(dx) + 2 \int_{\mathbb{R}^d} \langle \nabla f(x), x - \nabla\varphi(x) \rangle f(x) \gamma(dx). \end{aligned}$$

Применяя использованное выше неравенство для определителей и следов, а также неравенство Коши, получаем

$$\int_{\mathbb{R}^d} f^2 \ln f^2 d\mu \leq \frac{\lambda + \varepsilon/2}{2} \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 f(x)^2 \mu(dx) + 2C_\varepsilon \int_{\mathbb{R}^d} |\nabla f|^2 d\gamma. \quad (3.2.4)$$

Далее, из определения оптимальных отображений следует неравенство

$$\frac{\lambda + \varepsilon/2}{2} \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 f(x)^2 \mu(dx) \leq \frac{\lambda + \varepsilon/2}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 f(x)^2 \pi(dx, dy),$$

где π – любая вероятностная мера на $\mathbb{R}^d \times \mathbb{R}^d$ с проекциями, равными μ на обоих сомножителях. Возьмем меру $\pi = \mu \otimes \mu$. Из элементарного неравенства $ab \leq e^a - b + b \ln b$ при $a, b > 0$ получаем

$$\frac{\lambda + \varepsilon/2}{2} |x - y|^2 f(x)^2 \leq \exp\left(\frac{\lambda + \varepsilon}{2} |x - y|^2\right) + \delta f(x)^2 \ln |f(x)^2| + C' f(x)^2,$$

где $\delta = \delta(\lambda, \varepsilon) < 1$. Из условия теоремы вытекает оценка

$$\frac{\lambda + \varepsilon/2}{2} \int_{\mathbb{R}^d} |x - \nabla \varphi(x)|^2 f(x)^2 \mu(dx) \leq \delta \text{Ent}_\mu f^2 + C_0$$

для некоторого $C_0 = C_0(\varepsilon, \lambda)$. Тогда из (3.2.4) следует, что (3.2.3) выполнено. Согласно известному неравенству Ротхауса [111],

$$\text{Ent}_\mu f^2 \leq \text{Ent}_\mu \left(f - \int_{\mathbb{R}^d} f d\mu \right)^2 + 2 \int_{\mathbb{R}^d} \left(f - \int_{\mathbb{R}^d} f d\mu \right)^2 d\mu.$$

Применяя (3.2.3) к $f - \int_{\mathbb{R}^d} f d\mu$ и используя неравенство Ротхауса и неравенство Пуанкаре, получаем утверждение теоремы.

ЗАМЕЧАНИЕ 3.2.4. (i) В формулировке теоремы можно опустить условие, что μ удовлетворяет неравенству Пуанкаре, ибо оно следует из прочих условий. Доказательство этого факта не очень сложно, но довольно длинно, см. [110].

(ii) Если $D^2V \geq \lambda > 0$, то логарифмическое неравенство Соболева выполнено автоматически с константой λ .

СЛЕДСТВИЕ 3.2.5. Если мера μ выпукла, то необходимым и достаточным условием для (3.2.1) является существование такого $\varepsilon > 0$, что верно (3.2.2).

Ниже мы рассмотрим еще несколько ситуаций, в которых верны аналоги логарифмического неравенства Соболева; см. также [112].

3.2.3. Равномерно выпуклые потенциалы. Предположим, что потенциал V меры $\mu = e^{-V}$ является равномерно выпуклым в следующем смысле:

$$V(x + y) + V(x) - \langle \nabla V(x), y \rangle \geq c(\|y\|), \quad (3.2.5)$$

где $\|\cdot\|$ – некоторая норма (вообще говоря, неевклидова), а c – некоторая возрастающая функция. Заметим, что это условие выполнено, например, для функций вида $V = |x|^\alpha$, где $\alpha > 2$. Используя рассуждения из доказательства логарифмического неравенства Соболева и неравенство Юнга, можно доказать, что мера μ удовлетворяет следующему неравенству:

$$\text{Ent}_\mu f^2 \leq \int_{\mathbb{R}^d} c^* \left(\left\| 2 \frac{\nabla f}{f} \right\|_* \right) f^2 d\mu, \quad (3.2.6)$$

где $\|\cdot\|_*$ – сопряженная норма, а c^* – соответствующая выпуклая сопряженная функция:

$$c^*(t) = \sup_{s>0} \{ts - c(s)\}.$$

Неравенства вида (3.2.6) называются модифицированными логарифмическими неравенствами Соболева. Они обобщают классическое логарифмическое неравенство Соболева на случай неквадратичных норм. Важно отметить следующие особенности этого результата: неравенство не зависит от размерности и сохраняется для произведений мер (тензоризация). Последнее означает следующее. Пусть μ_i – набор вероятностных мер на $\mathbb{R}^d$, причем

$$\text{Ent}_{\mu_i} g \leq \int c_i^* \left(\left\| \frac{\nabla g}{g} \right\|_* \right) g d\mu_i.$$

Положим $\mu = \bigotimes_{i=1}^n \mu_i$. Известно (см., например, [103; предложение 5.6]), что энтропия имеет следующее замечательное свойство:

$$\text{Ent}_{\mu} g = \sum_{i=1}^n \int \text{Ent}_{\mu_i} g_i(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots) \mu(dx),$$

где g_i – функции, полученные из g фиксированием всех координат кроме x_i . Следовательно,

$$\text{Ent}_{\mu} g \leq \int \sum_{i=1}^n c_i^* \left(\left\| \frac{\nabla_{x_i} g_i}{g_i} \right\|_* \right) g d\mu.$$

Свойства такого рода важны в бесконечномерном анализе.

В работе [113] модифицированное логарифмическое неравенство Соболева распространено на общие функции стоимости с помощью транспортных методов. В [114] доказаны соответствующие изопериметрические неравенства для вероятностных мер и выпуклых тел с равномерными модулями выпуклости.

Некоторые свойства оптимальных транспортировок мер, удовлетворяющих условию (3.2.5), описаны в п. 2.2.5.

3.2.4. Ограниченность снизу второй производной и интегрируемость. Условие равномерной выпуклости потенциала является довольно ограничительным и может быть ослаблено за счет потери независимости константы от размерности.

ОПРЕДЕЛЕНИЕ 3.2.6. Пусть μ – вероятностная мера на метрическом пространстве (X, ϱ) . Изопериметрической функцией меры μ называется

$$\mathcal{I}_{\mu}(t) = \inf \{ \mu^+(\partial A); \mu(A) = t \},$$

где $\mu^+(\partial A) = \liminf_{h \rightarrow 0^+} \mu(\{x \in X \setminus A; \varrho(x, A) \leq h\})/h$.

ЗАМЕЧАНИЕ 3.2.7. С помощью транспортировки мер в [110] показано, что для выпуклых мер μ верно изопериметрическое неравенство Бобкова [115]:

$$\mu(A) \ln \frac{1}{\mu(A)} + \mu(A^c) \ln \frac{1}{\mu(A^c)} + \ln \mu(B_r) \leq 2r \mu^+(\partial A),$$

где $A \subset \mathbb{R}^d$, $B_r = \{x: |x| \leq r\}$ и $\mu(A) = \mu(B_r)$. Из этого можно вывести, что если для меры выполнено условие

$$\int_{\mathbb{R}^d} \exp(\varepsilon |x|^\alpha) \mu(dx) < +\infty, \quad \text{где } \alpha > 1, \quad (3.2.7)$$

то для множеств малой меры выполнено изопериметрическое неравенство

$$C\mu(A)|\ln \mu(A)|^{1-1/\alpha} \leq \mu^+(\partial A).$$

Для этого надо рассмотреть транспортировки мер $\mu|_A/\mu(A)$ и $\mu|_{A^c}/\mu(A^c)$ в меру $\mu|_{B_r}/\mu(B_r)$.

Как следует из гауссовского изопериметрического неравенства, для стандартной гауссовской меры γ и стандартной нормы имеет место оценка

$$\mathcal{I}_\gamma(t) \geq Ct|\ln t|^{1/2}, \quad t \in \left(0, \frac{1}{2}\right].$$

При этом C не зависит от размерности.

Из замечания 3.2.7 следует, что для изопериметрической функции выпуклой меры μ , удовлетворяющей условию (3.2.7), выполнено неравенство

$$\mathcal{I}_\mu(t) \geq Ct|\ln t|^{1-1/\alpha} \quad \text{при} \quad t \leq \frac{1}{2}.$$

Точнее говоря, $\mathcal{I}_\mu(t) \geq C \min(t, 1-t)|\ln \min(t, 1-t)|^{1-1/\alpha}$, $t \in (0, 1)$. Константа C в общем случае зависит от размерности.

Довольно общий вид соотношений между изопериметрической функцией и неравенствами типа Соболева получен в работах [116], [110], где рассмотрены неравенства вида

$$\int_{\mathbb{R}^d} f^2 F\left(\frac{f^2}{\|f\|_{L^2(\mu)}^2}\right) d\mu \leq C \int_{\mathbb{R}^d} |\nabla f|^2 d\mu,$$

которые называются F -неравенствами и наряду с модифицированными логарифмическими неравенствами Соболева обобщают классическое логарифмическое неравенство Соболева. Из них следуют свойства типа гиперсжимаемости. Есть и другие функциональные неравенства, естественным образом обобщающие логарифмическое соболевское, например неравенства Бекнера–Латалы–Олешкевича и супернеравенства Пуанкаре, введенные Ф.-Ю. Вангом. Обилие неравенств объясняется тем, что разные их виды по-разному адаптированы к различным желаемым свойствам, которыми в полной мере обладает логарифмическое неравенство Соболева.

3.3. Транспортные неравенства

Транспортное неравенство для стандартной гауссовской меры было доказано М. Талаграном [117]. Оно выглядит следующим образом:

$$\frac{1}{2}W_2^2(\gamma, g \cdot \gamma) \leq \text{Ent}_\gamma g, \quad (3.3.1)$$

где $g \cdot \gamma$ – произвольная вероятностная мера, абсолютно непрерывная относительно γ , для которой правая часть конечна. Это неравенство равносильно оценке

$$\frac{1}{2} \int_{\mathbb{R}^d} |x - \nabla \varphi(x)|^2 \gamma(dx) \leq \int_{\mathbb{R}^d} g \ln g d\gamma,$$

где $\nabla\varphi$ – оптимальная транспортировка меры γ в $g \cdot \gamma$. Следствием (3.3.1) и неравенства треугольника для W_2 является симметричное неравенство

$$W_2(f \cdot \gamma, g \cdot \gamma) \leq \sqrt{2 \operatorname{Ent}_\gamma f} + \sqrt{2 \operatorname{Ent}_\gamma g}, \quad (3.3.2)$$

где f и g – вероятностные плотности относительно γ .

Для обоснования (3.3.1) рассмотрим формулу замены переменных

$$-\frac{x^2}{2} = \ln g(\nabla\varphi(x)) - \frac{|\nabla\varphi(x)|^2}{2} + \ln \det D_a^2\varphi(x).$$

Записав ее как $|x - \nabla\varphi(x)|^2/2 = \langle x, x - \nabla\varphi(x) \rangle + \ln g(\nabla\varphi(x)) + \ln \det D_a^2\varphi(x)$, проинтегрировав по мере γ и применив уже знакомые рассуждения, получим

$$\frac{1}{2} \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 \gamma(dx) \leq \int_{\mathbb{R}^d} \ln g(\nabla\varphi) d\gamma = \int_{\mathbb{R}^d} g \ln g d\gamma.$$

Неравенство (3.3.1) доказано.

Можно доказать (см. [118]), что если $g \in L^p(\gamma)$, $p > 1$, то существует такое число $\varepsilon = \varepsilon(p) > 0$, что

$$\int_{\mathbb{R}^d} \exp(\varepsilon |x - \nabla\varphi(x)|^2) \gamma(dx) \leq C(p) \int_{\mathbb{R}^d} g^p d\gamma.$$

Ниже нам понадобится более общая версия неравенства Талагранна.

Определитель Фредгольма–Карлемана оператора A в $\mathbb{R}^n$ задается формулой $\det_2 A := e^{\operatorname{trace}(I-A)} \det A$.

ТЕОРЕМА 3.3.1. *Для любых двух вероятностных мер вида $\mu = e^{-V} dx$ и $\nu = e^{-W} dx$ на $\mathbb{R}^d$ и соответствующих оптимальных отображений $\nabla\Phi_\mu$ и $\nabla\Phi_\nu$, переводящих μ и ν в меру $m = e^{-P} dx$, удовлетворяющую с некоторым числом $K > 0$ условию $D^2P \geq K \cdot \operatorname{Id}$, выполнено неравенство*

$$\operatorname{Ent}_\nu\left(\frac{\mu}{\nu}\right) \geq \frac{K}{2} \int |\nabla\Phi_\mu - \nabla\Phi_\nu|^2 d\mu - \int \ln \det_2[D^2\Phi_\nu \cdot (D^2\Phi_\mu)^{-1}] d\mu,$$

если все интегралы в неравенстве существуют.

Последнее слагаемое в правой части неотрицательно. Это следует из соотношения

$$-\ln \det_2[D^2\Phi_\nu \cdot (D^2\Phi_\mu)^{-1}] = -\ln \det_2[(D^2\Phi_\mu)^{-1/2} \cdot D^2\Phi_\nu \cdot (D^2\Phi_\mu)^{-1/2}]$$

и неотрицательности матрицы $(D^2\Phi_\mu)^{-1/2} \cdot D^2\Phi_\nu \cdot (D^2\Phi_\mu)^{-1/2}$.

Приведем идею доказательства. По формуле замены переменных

$$e^{-V} = \det D^2\Phi_\mu \cdot e^{-P(\nabla\Phi_\mu)} \quad \mu\text{-п.в.}$$

Следовательно, $-V((\nabla\Phi_\mu)^{-1}) = -P + \ln \det D^2\Phi_\mu \circ ((\nabla\Phi_\mu)^{-1})$. Аналогично, применив формулу замены переменных для ν , получаем

$$-W((\nabla\Phi_\mu)^{-1}) = -P(\nabla\Phi_\nu \circ (\nabla\Phi_\mu)^{-1}) + \ln \det D^2\Phi_\nu \circ ((\nabla\Phi_\nu)^{-1}).$$

Положим $S = (\nabla \Phi_\mu)^{-1}$ и вычтем это равенство из предыдущего, что даст

$$(-V + W) \circ S = P(\nabla \Phi_\nu \circ S) - P - \ln \det[D^2 \Phi_\nu \cdot (D^2 \Phi_\mu)^{-1}] \circ S.$$

Заметим, что

$$P(\nabla \Phi_\nu \circ S(x)) - P(x) \geq \frac{K}{2} |\nabla \Phi_\nu(S(x)) - x|^2 + \langle \nabla P, \nabla \Phi_\nu(S(x)) - x \rangle.$$

Следовательно,

$$\begin{aligned} (-V + W) \circ S(x) &\geq K |\nabla \Phi_\nu(S(x)) - x|^2 \\ &\quad + \langle \nabla P, \nabla \Phi_\nu(S(x)) - x \rangle - \ln \det[D^2 \Phi_\nu \cdot (D^2 \Phi_\mu)^{-1}] \circ S(x). \end{aligned}$$

Проинтегрируем полученное соотношение по мере m . Из формулы интегрирования по частям и соотношения $m = \mu \circ (\nabla \Phi_\mu)^{-1}$ получим

$$\begin{aligned} &\int \langle \nabla P(x), \nabla \Phi_\nu(S(x)) - x \rangle m(dx) \\ &\geq \int \text{Tr}[D^2 \Phi_\nu \circ S \cdot (D^2 \Phi_\mu)^{-1} \circ S] dm - d = \int \text{Tr}[D^2 \Phi_\nu \cdot (D^2 \Phi_\mu)^{-1}] d\mu - d. \end{aligned}$$

Нетрудно видеть, что доказательство транспортного неравенства для гауссовских мер сходно с доказательством логарифмического неравенства Соболева. В некотором смысле транспортное неравенство является “двойственным” к логарифмическому неравенству Соболева. Будем говорить, что вероятностная мера μ удовлетворяет транспортному неравенству, если

$$\frac{1}{2} W_2^2(\mu, g \cdot \mu) \leq C \text{Ent}_\mu g. \quad (3.3.3)$$

Логарифмическое неравенство Соболева (ЛНС) и транспортное неравенство (ТН) связаны нетривиальным образом. Согласно работе [119], (ЛНС) влечет (ТН). Приведем короткое доказательство из [120].

ТЕОРЕМА 3.3.2. *Если вероятностная мера μ удовлетворяет (ЛНС), т. е.*

$$\text{Ent}_\mu f^2 \leq 2C \int_{\mathbb{R}^d} |\nabla f|^2 d\mu, \quad f \in C_0^\infty(\mathbb{R}^d),$$

то для всякой вероятностной меры $g \cdot \mu$ имеем $W_2^2(\mu, g \cdot \mu) \leq 2C \text{Ent}_\mu g$.

ДОКАЗАТЕЛЬСТВО. Для простоты предположим, что $C = 1$. Доказательство основано на свойствах полугруппы Хопфа–Лакса

$$v(t, x) = Q_t f(x) = \inf_{y \in \mathbb{R}^d} \left[f(y) + \frac{1}{2t} |x - y|^2 \right].$$

Известно, что для v выполнено уравнение Гамильтона–Якоби

$$\frac{\partial v}{\partial t} + \frac{1}{2} |\nabla v|^2 = 0.$$

Кроме того, $v(0, x) = f(x)$. Положим $g(x, \lambda) = Q_1(\lambda f)(x)$. Из уравнения Гамильтона–Якоби следует равенство $g = \lambda \frac{\partial g}{\partial \lambda} + \frac{|\nabla_x g|^2}{2}$. Положим

$$M(\lambda) = \int_{\mathbb{R}^d} e^{\lambda g} d\mu.$$

Применяя (ЛНС), получаем $\lambda M'(\lambda) \leq M(\lambda) \ln M(\lambda)$. Отсюда следует оценка $M(1) \leq e^{M'(0)}$. Заметим, что $M'(0) = \int_{\mathbb{R}^d} f d\mu$. Таким образом,

$$\int_{\mathbb{R}^d} \exp(Q_1 f) d\mu \leq \exp\left(\int_{\mathbb{R}^d} f d\mu\right). \quad (3.3.4)$$

Пусть теперь $T = \nabla \varphi$ – оптимальная транспортировка меры μ в меру $g \cdot \mu$. Тогда

$$W_2^2(\mu, g \cdot \mu) = \int_{\mathbb{R}^d} |x - \nabla \varphi(x)|^2 \mu(dx).$$

Как мы отмечали выше, $\varphi(x) + \varphi^*(\nabla \varphi(x)) = \langle x, \nabla \varphi(x) \rangle$ для μ -почти всех x . Отсюда получаем $|x - \nabla \varphi(x)|^2/2 = x^2/2 - \varphi(x) + (x^2/2 - \varphi^*(x)) \circ \nabla \varphi(x)$. По формуле замены переменных

$$\frac{1}{2} W_2^2(\mu, g \cdot \mu) = \int_{\mathbb{R}^d} \left(\frac{1}{2} x^2 - \varphi(x)\right) \mu(dx) + \int_{\mathbb{R}^d} \left(\frac{1}{2} x^2 - \varphi^*(x)\right) g(x) \mu(dx).$$

Без ограничения общности φ можно выбрать таким образом, что

$$\int_{\mathbb{R}^d} \varphi d\mu = \frac{1}{2} \int_{\mathbb{R}^d} x^2 \mu(dx).$$

Заметим, что $Q_1(x^2/2 - \varphi^*)(x) = \varphi(x) - x^2/2$. Используя известное неравенство $ab \leq e^a - b + b \ln b$ и (3.3.4), получаем

$$\begin{aligned} \frac{1}{2} W_2^2(\mu, g \cdot \mu) &\leq \exp\left(\int_{\mathbb{R}^d} \left(\varphi(x) - \frac{x^2}{2}\right) \mu(dx)\right) + \int_{\mathbb{R}^d} g \ln g d\mu - \int_{\mathbb{R}^d} g d\mu \\ &= \int_{\mathbb{R}^d} g \ln g d\mu. \end{aligned}$$

Теорема доказана.

Обратное утверждение (транспортное неравенство влечет неравенство Соболева) в общем случае неверно, что было некоторое время открытой проблемой. Для некоторых классов мер (например, выпуклых) эти неравенства равносильны. Первый контрпример был построен в [121]. Пример меры $e^{-V} dx$ на прямой, удовлетворяющей транспортному неравенству, но не удовлетворяющей логарифмическому неравенству Соболева, дается потенциалом $V(x) = |x|^3 + 3x^2 \sin^2 x + |x|^\beta$, $2 < \beta < 5/2$. Доказательство основано на некоторых критериях для обоих неравенств для мер на прямой. Н. Гозлан в работах [122], [123] дал красивое описание широкого класса мер, удовлетворяющих (ТН). Необходимые и достаточные условия для логарифмического неравенства Соболева известны только в одномерном случае (см. [105]).

Важным свойством гауссовского транспортного неравенства (3.3.1) (как и гауссовского логарифмического неравенства Соболева) является его независимость от размерности. Естественным обобщением на общие выпуклые функции стоимости c является неравенство

$$K_c(\mu, f \cdot \mu) \leq C \operatorname{Ent}_\mu f. \quad (3.3.5)$$

Неравенства такого типа следуют, например, из модифицированных неравенств Соболева (см. [103], [120], [124]).

Можно показать, что для равномерно выпуклого потенциала V , т. е. при

$$V(x+y) + V(x) - \langle \nabla V(x), y \rangle \geq c(\|y\|),$$

для меры $\mu = e^{-V} dx$ и $c(x, y) = c(\|x - y\|)$ выполнено (3.3.5).

В работе [125] найдено необходимое и достаточное условие для следующего аналога неравенства Талаграна для функции стоимости $|x - y|$:

$$W_1(\mu, g \cdot \mu) \leq \sqrt{C \operatorname{Ent}_\mu g}.$$

Оказывается, что здесь достаточно наличия такого $\varepsilon > 0$, что

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{\varepsilon|x-y|^2} \mu(dx) \mu(dy) < \infty.$$

Транспортное неравенство близко по своей форме другому классическому неравенству, называемому неравенством Пинскера–Кульбака–Чизара. Пусть μ и $\nu = f \cdot \mu$ – вероятностные меры и $f > 0$. Тогда

$$\frac{1}{2} \|\mu - \nu\|^2 := \frac{1}{2} \left(\int |f - 1| d\mu \right)^2 \leq \operatorname{Ent}_\mu f.$$

Здесь $\|\mu - \nu\|$ – расстояние по вариации между μ и ν . Оказывается, это неравенство можно рассматривать, как транспортное неравенство для подходящей функции стоимости (дискретной). Подробнее об этом см. [126].

Ниже будет обсуждаться результат Н. Гозлана об эквивалентности транспортного неравенства неравенству безразмерной концентрации и связи с теорией больших уклонений. Недавно в [127] получено неравенство типа Соболева, эквивалентное неравенству концентрации. С помощью этого доказано, что, аналогично логарифмическому неравенству Соболева, транспортное неравенство сохраняется при ограниченных возмущениях потенциала. Там же показано, что для широкого класса функций стоимости c неравенство

$$K_c(\mu, \nu) \leq C \operatorname{Ent}_\mu \nu = C \int \ln \frac{\mu}{\nu} d\mu$$

равносильно неравенству

$$\operatorname{Ent}_\mu(e^f) \leq \frac{f - Q_\lambda f}{1 - \lambda C} e^f d\mu, \quad \lambda \in \left(0, \frac{1}{C}\right),$$

где $Q_\lambda f = \inf_y (f(y) + \lambda c(x - y))$.

О связи транспортных неравенств с другими функциональными неравенствами, теорией больших уклонений, информационными неравенствами, функциями Ляпунова см. [128]–[130] и обзор [131]. Динамические транспортные неравенства рассмотрены в [132].

3.4. Неравенства концентрации и большие отклонения

Следующее важное наблюдение сделала К. Мартон. Пусть мера μ удовлетворяет (3.3.3), $A \subset \mathbb{R}^d$ и $\mu(A) \geq 1/2$. Рассмотрим оптимальное отображение $\nabla\varphi$ меры $\mu_1 = \frac{1}{\mu(A)}I_A$ в $\mu_2 = \frac{1}{\mu((A^h)^c)}I_{(A^h)^c}$. Тогда $W_2(\mu_1, \mu_2) \geq h$. Кроме того,

$$W_2(\mu_1, \mu_2) \leq W_2(\mu_1, \mu) + W_2(\mu_2, \mu) \leq \sqrt{2C \operatorname{Ent}_\mu \mu_1} + \sqrt{2C \operatorname{Ent}_\mu \mu_2}.$$

Так как $\operatorname{Ent}_\mu \mu_1 = \ln \frac{1}{\mu(A)}$, $\operatorname{Ent}_\mu \mu_2 = \ln \frac{1}{\mu((A^h)^c)}$, то из полученного неравенства следует, что существуют такие $a, b > 0$, что

$$\mu(A^h) \geq 1 - ae^{-bh^2}. \quad (3.4.1)$$

Это рассуждение переносится на более общие функции стоимости. Можно проверить, что (3.4.1) эквивалентно следующему неравенству для 1-липшицевых функций: $\mu(x: f(x) - m_f \geq h) \leq ae^{-bh^2}$, $h > 0$, m_f – медиана f .

Неравенство (3.4.1) называется неравенством концентрации. Неравенства концентрации возникли в асимптотической теории выпуклых тел как инструмент изучения не зависящих от размерности свойств выпуклых тел. Особенно активно это направление развивалось в работах В. Мильмана и его учеников; интересные применения в теории вероятностей см. в [103].

Следующий факт легко следует из неравенства Чебышёва: если мера μ удовлетворяет (3.4.1), то для некоторого $\varepsilon > 0$ имеем

$$\int_{\mathbb{R}^d} e^{\varepsilon x^2} \mu(dx) < +\infty.$$

В частности, получаем необходимое условие для (ЛНС) и (ТН).

В работах [133], [129], [123] исследована связь оптимальных отображений с теорией больших отклонений. Поясним ее на примере одного красивого результата из [123].

ТЕОРЕМА 3.4.1. *Вероятностная мера μ на $\mathbb{R}^d$ удовлетворяет транспортному неравенству тогда и только тогда, когда она обладает безразмерным свойством гауссовской концентрации, т. е. существуют такие константы a, b, r_0 , что для всякой степени μ^n меры μ выполнено неравенство*

$$\mu^n(A^r) \geq 1 - be^{-a(r-r_0)^2}, \quad \text{где } r \geq r_0, \quad A \subset (\mathbb{R}^d)^n, \quad \mu^n(A) \geq \frac{1}{2}.$$

Доказательство в одну сторону тривиально и следует из свойства транспортного неравенства сохраняться при произведениях (тензоризация) и того факта, что транспортное неравенство влечет концентрацию. Предположим теперь, что выполнено свойство концентрации. Для любого n рассмотрим эмпириче-

ское среднее $L_n = n^{-1} \sum_{i=1}^n \delta_{x_i}$, где $\{x_i\}$ – независимые случайные величины

с законом распределения μ . Известно (теорема Варадарайана), что меры L_n сходятся слабо к μ . Кроме того, известно (см. раздел 1.1), что сходимость

в метрике Канторовича W_p равносильна слабой сходимости мер с конечными моментами порядка p и сходимости этих моментов. По теореме Санова величина $P(L_n \in \tilde{A})$ для множества $\tilde{A}$ в пространстве мер ведет себя как $e^{-nH(\tilde{A}|\mu)}$ с ростом n , где $H(\tilde{A}|\mu) = \inf\{\text{Ent}_\mu(d\nu/d\mu), \nu \in \tilde{A}\}$. Более точно, используя технику теории больших отклонений, можно показать, что

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \ln P(W_p(L_n, \mu) > t) \geq -\inf\left\{\text{Ent}_\mu\left(\frac{d\nu}{d\mu}\right) : W_p(\nu, \mu) > t\right\}$$

(см. подробности в [123]). Для $x \in (\mathbb{R}^d)^n$ положим $L_n^x = n^{-1} \sum_{i=1}^n \delta_{x_i}$. По теореме 1.2.11 имеем

$$|W_2(L_n^x, \mu) - W_2(L_n^y, \mu)| \leq W_2(L_n^x, L_n^y) \leq \left(\frac{1}{n} \sum_{i=1}^n |x^i - y^i|^2\right)^{1/2} = \frac{1}{\sqrt{n}} |x - y|.$$

Положим $A = \{x : W_2(L_n^x, \mu) \leq m_n\}$, где m_n — медиана $W_2(L_n, \mu)$. Из доказанной оценки следует неравенство $A^r \subset \{x : W_2(L_n^x, \mu) \leq m_n + r/\sqrt{n}\}$. Для случайного вектора $x \in (\mathbb{R}^d)^n$ с независимыми компонентами с распределением μ предполагаемое свойство концентрации дает

$$P\left(W_2(L_n^x, \mu) > m_n + \frac{r}{\sqrt{n}}\right) \leq be^{-a(r-r_0)^2}, \quad r \geq r_0.$$

Это равносильно оценке $P(W_2(L_n^x, \mu) > u) \leq b \exp(-a(\sqrt{n}(u - m_n) - r_0)^2)$ при $\sqrt{n}(u - m_n) \geq r_0$. Из сходимости $W_2(L_n, \mu)$ к нулю по вероятности следует сходимость m_n к нулю. Следовательно, если $u > 0$, то

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln P(W_2(L_n^x, \mu) > u) \leq -au^2.$$

Тогда по полученной выше оценке $\inf\{\text{Ent}_\mu(d\nu/d\mu) : W_2(\nu, \mu) > u\} \geq au^2$. Таким образом, $aW_2^2(\mu, \nu) \leq \text{Ent}_\mu(d\nu/d\mu)$. Теорема доказана.

3.5. Иерархия неравенств

Рассмотренные примеры выявляют некоторую закономерность. В классе вероятностных мер имеется следующая иерархия функциональных неравенств:

- 1) изопериметрические неравенства $\mathcal{I}_\mu(t) \geq t\varphi(t)$, $t \leq 1/2$
 $\Downarrow$
- 2) неравенства типа Соболева $\text{Ent}_\mu f^2 \leq \lambda_S \int c^* \left(\left\| \frac{\nabla f}{f} \right\| \right) f^2 d\mu$
 $\Downarrow$
- 3) транспортные неравенства $W_c(\mu, f \cdot \mu) \leq \lambda_T \text{Ent}_\mu f$
 $\Downarrow$
- 4) неравенства концентрации $\mu(A^r) \geq 1 - e^{-\lambda_{CC}(r)}$
 $\Downarrow$
- 5) экспоненциальная интегрируемость $\exists \varepsilon > 0 : \int e^{\varepsilon c(|x|)} \mu(dx) < \infty$.

В некоторых ситуациях верны также обратные импликации. Мы уже видели, например, что для выпуклых мер логарифмическое неравенство Соболева и соответствующее изопериметрическое неравенство вытекают из условия экспоненциальной интегрируемости 5) для $c(x) = |x|^2$. Оказывается, в выпуклом случае все эти неравенства эквивалентны. Наиболее общий результат такого вида был получен в работе [134] (см. также [135]). При этом константы в неравенствах не зависят от размерности.

ТЕОРЕМА 3.5.1. Пусть для выпуклой меры $\mu = e^{-V} dx$ выполнены следующие условия: (i) мера μ удовлетворяет неравенству концентрации

$$1 - \mu(A_r) \leq e^{-\mathcal{K}(r)}, \quad \text{где} \quad \mu(A) \geq \frac{1}{2}, \quad \mathcal{K} - \text{некоторая функция};$$

(ii) для некоторой возрастающей функции $\alpha(r): \mathbb{R}^+ \rightarrow \mathbb{R}^+$ с $\lim_{t \rightarrow \infty} \alpha(t) = \infty$ верна оценка $\mathcal{K}(r) \geq \alpha(r)$.

Положим $\widetilde{\mathcal{I}}_\mu(t) = \min(\mathcal{I}_\mu(t), \mathcal{I}_\mu(1-t))$. Тогда выполнено изопериметрическое неравенство $\widetilde{\mathcal{I}}_\mu(t) \geq \min(ct\gamma(\ln(1/t)), c_\alpha)$, $\gamma(t) = t/\alpha^{-1}(t)$, где c – универсальная константа, а c_α зависит только от α .

Теорема верна и для многообразия с мерой и неотрицательным тензором Бакри–Эмери (п. 3.7.4).

3.6. Геодезические в пространстве мер и градиентные потоки

3.6.1. Геодезические в метрике Канторовича. Рассмотрим оптимальную транспортировку T вероятностной меры μ_0 в μ_1 на метрическом пространстве (X, ϱ) с функцией стоимости $h = \varrho^p(x, y)$, $p > 1$ (будем для краткости называть такие отображения p -оптимальными). Пространство радоновских вероятностных мер с конечным p -моментом и метрикой W_p на них будем обозначать символом $\mathcal{P}_r^p(X)$.

Как мы увидим, физически транспортировку массы можно представить себе как движение частиц по прямым с постоянной скоростью $T(x) - x$:

$$T_t: x \mapsto (1-t)x + tT(x), \quad x \in \mathbb{R}^d.$$

Таким образом, $T_0(x) = x$, $T_1 = T$. Надо сказать, что еще сам Монж мыслил свои транспортировки как процессы во времени и пространстве, в частности привлекал к рассмотрению пути перемещения масс; этот аспект пока совсем не присутствовал в наших обсуждениях, поскольку все преобразования мер ассоциируются с мгновенным действием без каких-либо “путей перемещения”; между тем он тоже может иметь значение. Несмотря на свою простоту, семейство интерполирующих отображений $\{T_t\}_{t \in [0,1]}$ обладает весьма глубокими свойствами. Введем интерполирующее семейство мер $\mu_t = \mu_0 \circ T_t^{-1}$. Плодотворным и важным для приложений (в первую очередь для уравнений в частных производных) является представление о семействе $\{\mu_t\}$ как о *геодезической* в пространстве мер, наделенном метрикой Канторовича. Близкая идея была высказана В.И. Арнольдом в работах по гидромеханике, в которых уравнения Эйлера представлены как уравнения геодезических на (бесконечномерной) группе диффеоморфизмов с некоторой римановой метрикой.

Мы определили интерполирующее семейство мер, используя понятие оптимального отображения (которого может и не быть). В общем случае интерполирующее семейство можно определить через линейную интерполяцию оптимального плана, а также посредством $\mu_t = \Pi \circ e_t^{-1}$, где Π – динамический оптимальный план (см. определение 3.6.6). Интерполирующая кривая, вообще говоря, неединственна. В произвольном метрическом пространстве длину параметрической кривой $\gamma(t)$, $t \in [0, 1]$, определяют как

$$L(\gamma) = \int_0^1 |\dot{\gamma}(t)| dt, \quad \dot{\gamma}(t) = \limsup_{\varepsilon \rightarrow 0} \frac{\varrho(\gamma(t+\varepsilon), \gamma(t))}{\varepsilon}.$$

Другое (эквивалентное для абсолютно непрерывных кривых) определение:

$$L(\gamma) = \sup_n \sup_{0=t_0 < t_1 < \dots < t_n=1} \sum_{i=0}^{n-1} \varrho(\gamma(t_i), \gamma(t_{i+1})).$$

В дальнейшем в пп. 3.6.1–3.6.3 разделе речь будет идти о специальном классе (LS) метрических пространств (называемых пространствами Александрова–Буземана; иное название: length space), в котором предполагается выполненным следующее условие: $\varrho(x, y) = \inf_{\Gamma^{x,y}} L(\gamma)$, где $\Gamma^{x,y}$ – пространство непрерывных кривых с $\gamma(0) = x$, $\gamma(1) = y$. Класс (LS)-пространств содержит все гладкие римановы многообразия и нормированные пространства. Этот класс пространств оказывается очень естественным в тех задачах, где возникают геодезические.

ОПРЕДЕЛЕНИЕ 3.6.1. Кривая $\gamma: [0, 1] \rightarrow X$ в (LS)-пространстве называется геодезической постоянной скорости, если

$$\varrho(\gamma(t), \gamma(s)) = |t - s| \varrho(\gamma(0), \gamma(1)).$$

Нетрудно видеть, что геодезическая постоянной скорости удовлетворяет равенству $L(\gamma) = \varrho(\gamma(0), \gamma(1))$ и является кратчайшей кривой, соединяющей γ_0 и γ_1 . Оказывается, что интерполирующее семейство $\{\mu_t\}$ является геодезической постоянной скорости в пространстве $(\mathcal{P}_r^p(X), W^p)$.

ТЕОРЕМА 3.6.2. Пусть μ_0, μ_1 – вероятностные меры на $X = \mathbb{R}^d$, T – p -оптимальное отображение μ_0 в μ_1 , $\mu_t = \mu_0 \circ T_t^{-1}$, $T_t(x) = (1-t)x + tT(x)$. Тогда семейство мер $\{\mu_t\}$ удовлетворяет соотношению

$$W_p(\mu_t, \mu_s) = (t - s)W_p(\mu_0, \mu_1).$$

ДОКАЗАТЕЛЬСТВО. Оценка $W_p(\mu_t, \mu_s) \leq (t - s)W_p(\mu_0, \mu_1)$ следует из оптимальности T и неравенства (для любых меры μ и отображений r_1, r_2)

$$W_p^p(\mu \circ r_1, \mu \circ r_2) \leq \int |r_1 - r_2|^p d\mu.$$

Для доказательства последнего заметим, что $\mu \circ (r_1, r_2)^{-1}$ является мерой на $X \times X$ с маргиналами $\mu \circ r_1^{-1}, \mu \circ r_2^{-1}$. Следовательно,

$$W_p^p(\mu \circ r_1, \mu \circ r_2) \leq \int_{X^2} |x_1 - x_2|^p d(\mu \circ (r_1, r_2)^{-1}) = \int_X |r_1 - r_2|^p d\mu.$$

Пусть теперь $W_p(\mu_t, \mu_s) < (t-s)W_p(\mu_0, \mu_1)$ для некоторой пары точек s, t . Так как $W_p(\mu_0, \mu_s) \leq sW_p(\mu_0, \mu_1)$, $W_p(\mu_t, \mu_1) \leq (1-t)W_p(\mu_0, \mu_1)$, то неравенство треугольника приводит к противоречию. Теорема доказана.

ЗАМЕЧАНИЕ 3.6.3. Если X – (LS)-пространство, то и пространство вероятностных мер $\mathcal{P}_r^p(X)$ оказывается (LS)-пространством (см. [136], [11]).

ЗАМЕЧАНИЕ 3.6.4. Также верно и обратное – любая геодезическая постоянной скорости соответствует некоторому оптимальному плану π для пары мер (см. теорему 7.22 в [11] и теорему 3.6.7 ниже).

По этой теме см. также [137], [138], [139]. Новый класс расстояний рассмотрен в [140].

3.6.2. Формула Бенаму–Бренье. <http://www.ams.org/mathscinet-getitem?mr=086>

Формула Бенаму–Бренье интерпретирует метрику $\mathcal{W}_p(\mathbb{R}^d)$ как риманову длину на (бесконечномерном) многообразии вероятностных мер $\mathcal{P}^p(\mathbb{R}^d)$. Пусть $v_t: \mathbb{R}^d \rightarrow \mathbb{R}^d$, $t \in [0, 1]$, – семейство гладких векторных полей (поле скоростей) и $T_t: \mathbb{R}^d \rightarrow \mathbb{R}^d$ – порожденное им семейство преобразований, заданное уравнением $dT_t/dt = v_t(T_t)$, $T_0(x) = x$. Рассмотрим семейство мер, заданное соотношением

$$\mu_t = \mu_0 \circ T_t^{-1} = \varrho_t dx, \quad \mu_0 = \varrho_0 dx.$$

Нетрудно проверить, что μ_t удовлетворяет классическому уравнению неразрывности (транспортному уравнению)

$$\frac{\partial \varrho_t}{\partial t} + \operatorname{div}(\varrho_t \cdot v_t) = 0. \quad (3.6.1)$$

Следующая формула получена в работах Бенаму–Бренье (см. [141]) для $p = 2$:

$$W_p^p(\mu_0, \mu_1) = \inf_v \int_0^1 \int |v_t|^p d\mu_t dt,$$

где $\mu_t = \varrho_t dx$ удовлетворяет (3.6.1) и инфимум берется по векторным полям из некоторого подходящего класса. Полученное выражение напоминает формулу для длины геодезической на римановом многообразии. Заметим, что неравенство

$$W_p^p(\mu_0, \mu_1) \leq \int_0^1 \int |v_t|^p d\mu_t dt$$

следует из соотношений

$$\begin{aligned} W_p^p(\mu_0, \mu_1) &\leq \int |T_t(x) - x|^p \varrho_0 dx \leq \int \int_0^1 \left| \frac{d}{dt} T_t(x) \right|^p dt \varrho_0 dx \\ &= \int \int_0^1 v_t^p(T_t) dt d\mu_0 = \int_0^1 \int v_t^p d\mu_t dt. \end{aligned}$$

Равенство формально достигается при $T_t(x) = (1-t)x + tT_0$ (случай геодезических постоянной скорости), $v_t = (T_0 - x) \circ T_t^{-1}$. Строгое обоснование (и формулировку) можно найти в [16; теорема 8.1] и в [11; п. 8.3].

3.6.3. Динамический план переноса массы. Теорема о геодезических имеет широкое обобщение на случай абстрактных “лагранжевых действий”. Подробно об этом рассказано в [17; гл. 7], ниже приведена краткая формулировка. Лагранжевым действием называется семейство полунепрерывных снизу функционалов $\mathcal{A}^{s,t}: C \rightarrow \mathbb{R}$ на некотором классе C непрерывных кривых $\gamma: [0, 1] \rightarrow X$ и семейство функций стоимости $c^{s,t}: X \times X \rightarrow \mathbb{R}$, удовлетворяющих следующим условиям:

- 1) $\mathcal{A}^{t_1, t_2} + \mathcal{A}^{t_2, t_3} = \mathcal{A}^{t_1, t_3}$, $0 \leq t_1 < t_2 < t_3 \leq 1$;
- 2) $c^{s,t}(x, y) = \inf \{ \mathcal{A}^{s,t}(\gamma), \gamma \in C, \gamma(s) = x, \gamma(t) = y \}$ для всех $x, y \in X$;
- 3) для всех $\gamma \in C$ выполнено равенство

$$\mathcal{A}^{s,t}(\gamma) = \sup_{N \in \mathbb{N}} \sup_{0=t_0 \leq t_1 \leq t_N=1} \sum_{i=0}^{N-1} c^{t_i, t_{i+1}}(\gamma_{t_i}, \gamma_{t_{i+1}}).$$

Класс кривых C и топология на нем зависят от конкретной задачи. Типичным примером является функционал

$$\mathcal{A}^{s,t}(\gamma) = \int_s^t L(\gamma_\tau, \dot{\gamma}_\tau, \tau) d\tau, \quad \text{где } L(x, v, t) \text{ — некоторая функция Лагранжа.}$$

ПРИМЕР 3.6.5. Пусть X — (LS)-пространство, c — строго выпуклая функция. Положим

$$\mathcal{A}(\gamma) = \int_0^1 c(|\dot{\gamma}_t|) dt.$$

Из неравенства Йенсена следует оценка

$$c(d(\gamma_0, \gamma_1)) \leq \int_0^1 c(|\dot{\gamma}_t|) dt.$$

Следовательно, минимум $\mathcal{A}$ достигается на геодезических постоянной скорости и равен $c(d(x, y))$.

Оказывается, теорема о геодезических имеет широкое обобщение на случай коэрцитивных действий. Лагранжево действие называется коэрцитивным, если

- 1) $\inf_{s < t} \inf_{\gamma} \mathcal{A}^{s,t} > -\infty$,
- 2) для всех $t < s$ и всяких двух компактных множеств K_s, K_t множество минимизирующих кривых $\Gamma_{K_s \rightarrow K_t}^{s,t}$ с началом в K_s в момент s и концом в K_t в момент t компактно и непусто.

ОПРЕДЕЛЕНИЕ 3.6.6. Пусть Γ — множество кривых, минимизирующих действие $\mathcal{A}$. Динамическим оптимальным планом называется такая вероятностная мера Π на Γ , что мера $\Pi \circ (e_0, e_1)^{-1}$, где $e_t: \Gamma \rightarrow X$, $e_t(\gamma) = \gamma(t)$, является оптимальным планом для пары μ_0, μ_1 .

Следующий результат [17; теорема 7.19] дает широкое обобщение теоремы 3.6.2. Как показано в [17], теорема 3.6.2 следует из этого результата в случае польского локально компактного (LS)-пространства.

ТЕОРЕМА 3.6.7. Пусть X – польское пространство, $\mathcal{A}^{0,1}$ – коэрцитивное лагранжево действие. Предположим, что функции стоимости $c^{s,t}$ непрерывны. Обозначим через $C^{s,t}(\mu, \nu)$ стоимость оптимальной транспортировки для пары μ, ν . Тогда для кривой $\{\mu_t\}_{0 \leq t \leq 1}$ следующие свойства равносильны:

(i) при $0 \leq t \leq 1$ мера μ_t равна распределению $\gamma(t)$, где $t \mapsto \gamma(t)$ – случайная кривая, минимизирующая действие $\mathcal{A}$, причем (γ_0, γ_1) – оптимальный план для пары μ_0, μ_1 ;

(ii) при $t_1 < t_2 < t_3$ выполнено равенство

$$C^{t_1, t_2}(\mu_{t_1}, \mu_{t_2}) + C^{t_2, t_3}(\mu_{t_2}, \mu_{t_3}) = C^{t_1, t_3}(\mu_{t_1}, \mu_{t_3});$$

(iii) кривая $\{\mu_t\}$ в пространстве мер минимизирует действие следующего коэрцитивного функционала на пространстве мер $\mathcal{P}(X)$:

$$\mathcal{A}^{s,t}(\gamma) = \sup_{N \in \mathbb{N}} \sup_{0=t_0 \leq t_1 \leq \dots \leq t_N=1} \sum_{i=0}^{N-1} C^{t_i, t_{i+1}}(\mu_{t_i}, \mu_{t_{i+1}}) = \inf_{\gamma} \mathbb{E} \mathcal{A}^{s,t}(\gamma).$$

Последний инфимум взят в пространстве случайных кривых $\gamma: [s, t] \rightarrow X$, для которых распределение γ_τ совпадает с μ_τ , $s \leq \tau \leq t$.

СЛЕДСТВИЕ 3.6.8. Пусть X – польское (LS)-пространство и Π – динамический оптимальный план переноса массы для функции стоимости $d^p(x, y)$, $p > 1$. Тогда кривая мер $\mu_t = \Pi \circ e_t^{-1}$ является геодезической постоянной скорости между μ_0 и μ_1 . Более того, любая геодезическая постоянной скорости $\{\mu_t\}$ имеет вид $\mu_t = \Pi \circ e_t^{-1}$ для некоторого динамического оптимального плана переноса массы.

3.6.4. Выпуклые функционалы. Функционалы, выпуклые вдоль геодезических на пространстве мер, были рассмотрены в работе Р. Маккэна [68] в связи с вопросами единственности решения некоторых вариационных задач. Выпуклость такого рода получила название “переносной выпуклости” (“displacement convexity”). Эта техника нашла широкое применение как в вариационных, так и в вероятностных задачах (см. пример 3.7.14). Функцию f на $\mathbb{R}^n$ называют λ -выпуклой, если выпукла функция $f(x) - \lambda|x|^2/2$.

ОПРЕДЕЛЕНИЕ 3.6.9. Функционал $\mathcal{F}: \mathcal{P}_r^p(X) \rightarrow (-\infty, \infty]$ называют λ -выпуклым, если для всякой пары мер $\mu_0, \mu_1 \in W_p(X)$ функция $t \mapsto \mathcal{F}(\mu_t)$ является λ -выпуклой, где $\{\mu_t\}$ – произвольное интерполирующее семейство мер в $\mathcal{P}_r^p(X)$.

Приведем примеры выпуклых функционалов на пространстве $\mathcal{P}_r^p(X)$.

ПРИМЕР 3.6.10. Обобщенный функционал энтропии: для заданной функции F положим

$$\mathcal{F}(\mu) = \int F(\varrho) dx,$$

если $\mu = \varrho dx$ и $\mathcal{F}(\mu) = \infty$ в противном случае. Если функция $t \mapsto t^d F(t^{-d})$ выпукла и убывает, то $\mathcal{F}$ – выпуклый функционал на $\mathcal{P}_r^p(\mathbb{R}^d)$.

Идея доказательства состоит в применении формулы замены переменной. Опуская вопросы регулярности, считаем, что выполнена формула замены переменной $\varrho_t(T_t) = \varrho_0 / \det DT_t$ для интерполирующего отображения T_t . Тогда

$$\begin{aligned} \mathcal{F}(\mu_t) &= \int F(\varrho_t) d = \int F(\varrho_t(T_t)) \det DT_t dx \\ &= \int F\left(\frac{\varrho_0}{\det DT_t}\right) \det DT_t dx. \end{aligned} \quad (3.6.2)$$

Заметим теперь, что матрица $(1-t)I + tDT_t$ диагонализируема и имеет неотрицательные значения на диагонали. Отсюда несложно извлечь, что функция

$$t \mapsto \det^{1/d}((1-t)I + tDT_t)$$

вогнута на $[0, 1]$. Тогда в силу условий, наложенных на F , выпуклой оказывается функция $t \mapsto F(\varrho_0 / \det DT_t) \det DT_t$. Следовательно,

$$F\left(\frac{\varrho_0}{\det DT_t}\right) \det DT_t \leq (1-t)F(\varrho_0) + tF(\varrho_1).$$

Интегрирование полученного неравенства дает искомый результат.

Следующие функции удовлетворяют условию примера 3.6.10:

$$F(t) = t \ln t, \quad F(t) = \frac{1}{m-1} s^m, \quad m \geq 1 - \frac{1}{d}.$$

ПРИМЕР 3.6.11 (потенциальная энергия). Для функции V , удовлетворяющей условию $V(x) \geq -a - b|x|^p$, возьмем функционал

$$\mathcal{V}(\mu) = \int V d\mu.$$

Если V – λ -выпуклая функция, то $\mathcal{V}$ – λ -выпуклый функционал на $\mathcal{P}^2(\mathbb{R}^d)$. Это следует из соотношений

$$\begin{aligned} \mathcal{V}(\mu_t) &= \int V((1-t)x + tT_t(x)) \mu_0(dx) \\ &\leq \int \left((1-t)V(x) + tV(T_t(x)) - \frac{\lambda}{2}t(1-t)|x - T_t(x)|^2 \right) \mu_0(dx) \\ &\leq (1-t) \int V d\mu_0 + t \int V d\mu_1 - \frac{\lambda}{2}t(1-t)W_2^2(\mu_0, \mu_1) \\ &= (1-t)\mathcal{V}(\mu_0) + t\mathcal{V}(\mu_1) - \frac{\lambda}{2}t(1-t)W_2^2(\mu_0, \mu_1). \end{aligned}$$

Результат верен при $p \leq 2$, $\lambda \geq 0$ и при $p \geq 2$, $\lambda \leq 0$ (см. [11; п. 9.3]).

СЛЕДСТВИЕ 3.6.12. Пусть $\nu = e^{-V} dx$ – выпуклая мера. Тогда относительная энтропия $\text{Ent}_\nu(\mu)$ – выпуклый функционал на $\mathcal{P}_r^p(\mathbb{R}^d)$.

ДОКАЗАТЕЛЬСТВО. Это следует из представления относительной энтропии $\text{Ent}_\nu(\varrho dx)$ в виде суммы двух выпуклых функционалов: интеграла от $\varrho \ln \varrho$ по мере Лебега и интеграла от V по мере μ .

ПРИМЕР 3.6.13 (энергия взаимодействия). Полунепрерывная снизу функция $W(x_1, \dots, x_k): (\mathbb{R}^d)^k \rightarrow \mathbb{R}$ определяет следующий функционал на $\mathcal{P}_r^p(\mathbb{R}^d)$:

$$\mathcal{W}(\mu) = \int_{(\mathbb{R}^d)^k} W d\mu^k.$$

Если функция W выпукла, то $\mathcal{W}$ – выпуклый функционал на $\mathcal{P}_r^p(\mathbb{R}^d)$. Это видно из того, что $\mathcal{W}$ – сужение функционала потенциального взаимодействия

$$\tilde{\mu} \mapsto \int W d\tilde{\mu}, \quad \tilde{\mu} \in \mathcal{P}_r^p((\mathbb{R}^d)^k),$$

на подмножестве произведений мер (т. е. мер-произведений).

Следующий результат доказан в работе [68].

ТЕОРЕМА 3.6.14. *Для строго выпуклых V и W функционал*

$$F(\mu) = \int U(\varrho) dx + \int V d\mu + \frac{1}{2} \int_{(\mathbb{R}^d)^2} W(x-y) \mu(dx) \mu(dy)$$

обладает единственной точкой минимума.

Естественно определяется “переносная выпуклость” множеств; в [142] недавно показано, что этим свойством не обладает множество мер в $\mathcal{P}^2(\mathbb{R}^d)$, равных нулю на множествах недифференцируемости выпуклых функций на $\mathbb{R}^d$.

3.6.5. Дополнительные свойства пространства $\mathcal{P}^p$. Укажем ряд важных для дальнейшего свойств пространства $\mathcal{P}_r^p(X)$ со ссылками на работы, где они обсуждаются.

1) Формула для дифференцирования W_2 (см. [11; теорема 8.4.7]):

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) = \int_{X^2} \langle v_t(x_2), x_1 - x_2 \rangle \gamma_t(dx_1 dx_2),$$

где σ – фиксированная мера, μ_t – абсолютно непрерывная кривая в $\mathcal{P}_2(X)$,

$$\frac{\partial}{\partial t} \mu_t + \operatorname{div}(v \cdot \mu_t) = 0,$$

γ_t – оптимальный транспортный план между μ_t и σ . При этом функционал $W_2^2(X)$ не является выпуклым на $\mathcal{P}^2(X)$.

2) $\mathcal{P}_r^p(X)$ – (LS)-пространство, если таково X (теорема 3.6.7).

3) “Неравенство параллелограмма” (см. [11]) для $\mathcal{P}^2(\mathbb{R}^n)$:

$$W_2(\mu_t^{1 \rightarrow 2}, \mu^3) \geq (1-t)W_2^2(\mu^1, \mu^3) + tW_2^2(\mu^2, \mu^3) - t(1-t)W_2^2(\mu^1, \mu^2),$$

где $\mu_t^{1 \rightarrow 2}$ – интерполяция мер μ_1 и μ_2 . Из этого неравенства следует, что $\mathcal{P}^2(\mathbb{R}^n)$ – пространство неотрицательной кривизны. Более общим образом, пространство $\mathcal{P}_r^p(M)$, где M – гладкое многообразие, имеет неотрицательную кривизну Александрова, если (и только если) M имеет неотрицательную секционную кривизну (см. [143], [136]).

4) В работе [144] вычислены некоторые геометрические величины (связность, тензор Римана) на пространстве $\mathcal{P}_r^p$.

3.6.6. Субдифференциалы. Субдифференциалы для выпуклых функций $\Phi: \mathcal{P}^2(X) \rightarrow \mathbb{R}$ изучены в книге [11]. Ниже мы приведем основные понятия и результаты. Через $\mathcal{D}(\Phi)$ обозначим множество точек x , где $\Phi(x) < \infty$. Для $v \in \mathcal{D}(\Phi)$ определим модуль метрического градиента как

$$|\partial\Phi|(v) = \limsup_{\omega \rightarrow v} \frac{(\Phi(\omega) - \Phi(v))_+}{W_2(\omega, v)}.$$

Если мера μ регулярна (например, имеет плотность), то для любой меры ν существует оптимальная транспортировка T из μ в ν . Следующее определение субдифференциала для регулярной меры хорошо согласуется с классическим понятием субдифференциала для функций на гильбертовом пространстве.

ОПРЕДЕЛЕНИЕ 3.6.15. Пусть $\mu \in \mathcal{D}(|\partial\Phi|)$. Векторное поле $\xi \in L^2(\mu; \mathbb{R}^d)$ принадлежит субдифференциалу $\partial\Phi$, если

$$\Phi(\nu) - \Phi(\mu) \geq \int_{\mathbb{R}^d} \langle \xi(x), T(x) - x \rangle \mu(dx) + o(W_2(\mu, \nu))$$

для оптимальной транспортировки T из μ в ν .

Минимальной селекцией субдифференциала $\partial\Phi$ называется вектор $\xi \in \partial\Phi$ с минимальной нормой $\|\cdot\|_{L^2(\mu; \mathbb{R}^d)}$. Понятие субдифференциала можно распространить на случай нерегулярной меры, если вместо оптимальной транспортировки использовать оптимальные транспортные планы (см. [11; п. 10.3]).

ПРИМЕР 3.6.16. Для формального вычисления субдифференциала воспользуемся формулой (3.6.2), предполагая, что все вычисляемые объекты гладкие:

$$\mathcal{F}(\mu_t) = \int F\left(\frac{\varrho_0}{\det((1-t)I + tD^2\varphi)}\right) \det((1-t)I + tD^2\varphi) dx.$$

Принимая во внимание, что $\frac{d}{dt} \det((1-t)I + tD^2\varphi)|_{t=0} = \Delta\varphi - d$, легко получаем

$$\left. \frac{d}{dt} \mathcal{F}(\mu_t) \right|_{t=0} = \int (F(\varrho_0) - \varrho_0 F'(\varrho_0)) (\Delta\varphi - d) dx.$$

Положим $L_F(\varrho_0) = \varrho_0 F'(\varrho_0) - F(\varrho_0)$. Интегрируя по частям, получаем

$$\left. \frac{d}{dt} \mathcal{F}(\mu_t) \right|_{t=0} = \int \left\langle \frac{\nabla L_F(\varrho(x))}{\varrho(x)}, \nabla\varphi(x) - x \right\rangle \varrho(x) dx.$$

Точно так же получаются формулы для субдифференциалов функционалов потенциальной энергии и энергии взаимодействия:

$$\begin{aligned} \left. \frac{d}{dt} \mathcal{V}(\mu) \right|_{t=0} &= \int \langle \nabla V(x), \nabla\varphi(x) - x \rangle \varrho_0(x) dx, \\ \left. \frac{d}{dt} \mathcal{W}(\mu) \right|_{t=0} &= \int \langle (\nabla W * \mu(x)), \nabla\varphi(x) - x \rangle \varrho_0(x) dx. \end{aligned}$$

Строгое обоснование полученных формул технически довольно сложно. Следующий результат доказан в [11; теорема 10.4.13].

ТЕОРЕМА 3.6.17. Рассмотрим на $\mathcal{P}_r^p(\mathbb{R}^d)$ функционал

$$\Phi(\mu) = \int F(\varrho) dx + \int V d\mu + \frac{1}{2} \int W d\mu \times d\mu, \quad \mu = \varrho dx,$$

где $\Phi(\mu) = \infty$, если μ не имеет плотности. Предположим, что функция $F: [0, \infty) \rightarrow \mathbb{R}$ выпукла, дифференцируема и суперлинейна, причем функция $t \mapsto t^d F(t^{-d})$ выпукла и убывает и существует такое $\alpha > d/(d+p)$, что

$$F(0) = 0, \quad \liminf_{s \rightarrow 0} F(s)s^{-\alpha} > -\infty,$$

а также выполнено условие удвоения: существует такое число $C > 0$, что

$$F(a+b) \leq C(1 + F(a) + F(b)).$$

Кроме того, пусть $V: \mathbb{R}^d \rightarrow (-\infty, +\infty)$ – полунепрерывная снизу λ -выпуклая функция, $\mathcal{D}(V)$ имеет непустую внутренность Ω и $W: \mathbb{R}^d \rightarrow [0, +\infty)$ – выпуклая дифференцируемая четная функция, удовлетворяющая условию удвоения. При этих условиях мера $\mu = \varrho dx$ входит в $\mathcal{D}(|\partial\varphi|)$ в точности тогда, когда $L_F(\varrho) \in W_{\text{loc}}^{1,1}(\Omega)$, $L_F = sF'(s) - F(s)$ и

$$\varrho\omega = \nabla L_F(\varrho) + \varrho\nabla V + \varrho(\nabla W) * \varrho$$

для некоторого $\omega \in L^q(\mu; \mathbb{R}^d)$. В этом случае вектор ω является минимальной селекцией в $\partial\Phi(\mu)$.

3.6.7. Градиентные потоки. Формализм Отто. Градиентным потоком в $\mathbb{R}^d$ называется решение $X(t)$ уравнения $dX(t)/dt = -\nabla E(X(t))$, где E – некоторый потенциал. Заметим, что $dE(X(t))/dt = -|\nabla E(X(t))|^2$. Существует понятие градиентного потока на метрическом пространстве. Подробному изложению этой концепции посвящена первая часть книги [11]. Градиентный поток можно определить в некотором смысле как “кривую наименьшего спуска”. Градиентный поток на метрическом пространстве X удовлетворяет уравнению

$$-\frac{d}{dt}E(X(t)) = \frac{1}{2}|X'(t)|^2 + \frac{1}{2}|\partial E(X(t))|^2,$$

где $|X'(t)|$ – метрическая производная кривой $X(t)$. Оказывается, что многие уравнения математической физики допускают описание в виде градиентных потоков некоторых функционалов на пространстве $\mathcal{P}^2(X)$.

ПРИМЕР 3.6.18. Рассмотрим уравнение $\partial\varrho/\partial t - \operatorname{div}(\nabla L_F(\varrho)) = 0$ относительно функции $\varrho: (t, x) \mapsto \varrho_t(x)$; для функции $F(s) = s \ln s$ получаем уравнение теплопроводности $\partial\varrho/\partial t = \Delta\varrho$. Если $v = -\nabla L_F(\varrho)/\varrho$, то получаем транспортное уравнение $\partial\varrho/\partial t + \operatorname{div}(\varrho \cdot v) = 0$. Естественно интерпретировать v как скорость кривой $t \mapsto \varrho_t(\cdot)$. По теореме 3.6.17 функция $\nabla L_F(\varrho)/\varrho$ служит (суб)дифференциалом функционала $\mathcal{F}$. Следовательно, исходное уравнение является градиентным потоком и (формально) $v = -\nabla \mathcal{F}$ или $v \in -\partial\mathcal{F}$.

Градиентные потоки, связанные с функционалами на пространстве $\mathcal{P}^2(X)$, допускают интересную геометрическую интерпретацию, предложенную Ф. Отто [145] (см. также [146]). Им была введена формальная риманова структура

на $\mathcal{P}^2(X)$, в которой W_2 оказывается римановой метрикой на “многообразии” мер. Отметим, однако, что “формализм Отто” пока что остается во многих случаях эвристическим подходом, нуждающимся в строгом обосновании.

Напомним, что по формуле Бенаму–Бренье

$$W_2(\mu_0, \mu_1) = \inf \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \varrho_t(x) dx dt,$$

где инфимум взят среди всех ϱ_t, v_t с $\varrho_t|_{t=0} = \varrho_0, \varrho_t|_{t=1} = \varrho_1$, удовлетворяющих транспортному уравнению $\partial \varrho_t / \partial t + \operatorname{div}(\varrho_t v_t) = 0$. Минимум достигается на кривой $\{\varrho_t\}$, удовлетворяющей уравнениям

$$\varrho_t(x) = \varrho_t((1-t)x + t\nabla\varphi(x)) \det((1-t)I + tD^2\varphi(x)), \quad \frac{\partial \varrho_t}{\partial t} + \operatorname{div}(\varrho_t v_t) = 0,$$

где $v_t(x) = \nabla\varphi(S_t(x)) - S_t(x)$, $S_t(x) = ((1-t)I + t\nabla\varphi)^{-1}(x)$ и $\nabla\varphi$ – оптимальная транспортировка μ_0 в μ_1 . Ф. Отто ввел следующую дифференциальную структуру на $\mathcal{P}^2(X)$, где $X = \mathbb{R}^d$. Касательное пространство $T\mathcal{P}^2(X)$ в точке ϱdx состоит из всех мер с плотностью вида $-\operatorname{div}(\varrho v)$, где v – векторное поле. Определим скалярное произведение (риманову метрику) на $T\mathcal{P}^2(X)$ так:

$$\int_X \langle v_1, v_2 \rangle \varrho dx.$$

Рассмотрим с этой точки зрения функционал энтропии $\operatorname{Ent}: \varrho dx \mapsto \int_{\mathbb{R}^d} \varrho \ln \varrho dx$. Дифференцирование вдоль некоторой геодезической $t \mapsto \mu_0 \circ T_t^{-1}$ дает

$$\left. \frac{\partial \operatorname{Ent} \varrho_t}{\partial t} \right|_{t=0} = - \int_X \varrho (\Delta\varphi - d) dx = \int_X \langle \nabla \varrho, \nabla \psi \rangle dx,$$

где $\psi(x) = \varphi(x) - x^2/2$. Заметим, что $\partial \varrho_t / \partial t|_{t=0} + \operatorname{div}(\nabla \psi \cdot \varrho_0) = 0$. Следовательно, “градиент” функционала Ent удовлетворяет равенствам

$$\left\langle \nabla \operatorname{Ent}(\varrho), \frac{\partial \varrho_t}{\partial t} \right\rangle \Big|_{t=0} := \left. \frac{\partial \operatorname{Ent} \varrho_t}{\partial t} \right|_{t=0} = \int_X \langle \nabla \varrho, \nabla \psi \rangle dx = \int_X \left\langle \frac{\nabla \varrho}{\varrho}, \nabla \psi \right\rangle \varrho dx,$$

а $\nabla \operatorname{Ent}(\varrho)$ отождествляется с функцией $-\operatorname{div}\left(\varrho \frac{\nabla \varrho}{\varrho}\right) = -\Delta \varrho$. Соответствующий *градиентный поток* задается уравнением теплопроводности $\partial \varrho / \partial t = \Delta \varrho$. Зафиксируем вероятностную меру $\nu = e^{-V} dx$. Положим $\mu = f \cdot \nu = \varrho dx$. Рассмотрев таким же образом относительную энтропию

$$\operatorname{Ent}_\nu \mu = \int_X f \ln f d\nu = \int_X \varrho \ln \varrho dx + \int_X V \varrho dx,$$

получим, что соответствующий градиентный поток удовлетворяет уравнению

$$\frac{\partial \varrho}{\partial t} = \Delta \varrho + \operatorname{div}(\varrho \cdot \nabla V).$$

Ниже приведено несколько других примеров градиентных потоков.

ПРИМЕР 3.6.19. Уравнение пористых сред:

$$E(\mu) = \frac{1}{m-1} \int \varrho^m dx, \quad \frac{\partial \varrho}{\partial t} = \Delta \varrho^m.$$

Уравнение нелинейной диффузии: $\frac{\partial \varrho}{\partial t} - \Delta L_F(\varrho) = 0$.

Уравнение Маккина–Власова:

$$E(\mu) = \frac{1}{2} \int W(x-y) \varrho(x) \varrho(y) dx dy, \quad \frac{\partial \varrho}{\partial t} = \operatorname{div}(\varrho \nabla(\varrho * \nabla W)).$$

Уравнение квантовой диффузии со сносом (см. [147]):

$$E(\mu) = \int \frac{|\nabla \varrho|^2}{\varrho} dx, \quad \frac{\partial \varrho}{\partial t} + 4 \operatorname{div} \left(\varrho \nabla \frac{\Delta \sqrt{\varrho}}{\sqrt{\varrho}} \right) = 0.$$

Метрика Канторовича (под именем метрики Васерштейна) была применена к таким уравнениям еще Р. Л. Добрушиным [148]; из более новых работ отметим [149], [132], [150]. Связанные с метрикой W_p оценки решений уравнений теплопроводности см. в [151].

3.6.8. Приложения к уравнениям математической физики. Строгое обоснование приведенных выше эвристических выкладок можно делать разными способами. Один из возможных путей – формализация понятия римановой структуры и дифференциального исчисления Отто. Ситуация сильно осложняется тем, что рассматриваемые объекты не являются гладкими. Мы изложим ниже некоторые результаты книги [11], где основной акцент делался на работе с кривыми в пространстве мер и формализации понятия субдифференциала. Оказывается, что во многих случаях (включая λ -выпуклый) для обоснования существования/единственности/регулярности градиентного потока вместо равенства $v = -\nabla \mathcal{F}$ достаточно рассматривать включение $v \in -\partial \mathcal{F}$.

ТЕОРЕМА 3.6.20 [11; теорема 8.3.1]. *Абсолютно непрерывная кривая $\{\mu_t\}$ обладает касательным вектором v_t : $\partial \mu_t / \partial t + \operatorname{div}(v_t \cdot \mu_t) = 0$, где равенство понимается в слабом смысле.*

ОПРЕДЕЛЕНИЕ 3.6.21. Кривая $\{\mu_t\} \in AC_{\text{loc}}^p(\mathbb{R}^+; \mathcal{P}_r^p(X))$ называется градиентным потоком для функционала Φ , если $j_p(v_t) \in -\partial \Phi(\mu_t)$, $t > 0$, где $j_p(v) = |v|^{p-2}v$ – двойственное отображение.

Следующая теорема имеет большое значение, ибо позволяет применить “метрическую” теорию градиентных потоков к градиентным потокам в $\mathcal{P}_r^p(X)$.

ТЕОРЕМА 3.6.22 [11]. *Для λ -выпуклых функционалов понятие градиентного потока в $\mathcal{P}_r^p(X)$ совпадает с понятием градиентного потока в метрическом пространстве $\mathcal{P}_r^p(X)$.*

Для доказательства существования градиентного потока применяется аппроксимация минимизирующими движениями. В основе лежит следующее простое наблюдение. Предположим для простоты, что ψ – полунепрерывный снизу функционал на гильбертовом пространстве и v минимизирует функционал

$$\omega \mapsto \psi(\omega) + \frac{1}{2t} |\omega - a|^2.$$

Тогда $-(v - a)/t \in \partial\psi(v)$. Аналогом этого для $\mathcal{P}^2$ является следующий факт.

ЛЕММА 3.6.23. *Если ν минимизирует функционал $\nu \mapsto \Phi(\nu) + W_2^2(\nu, \mu)/(2t)$ и T – оптимальная транспортировка ν в μ , то $(T - x)/\tau \in \partial\Phi(\nu)$.*

Разделим теперь числовую прямую на интервалы длины $\tau > 0$. Положим $M_\tau^0 = \mu_0$, а M_τ^{n+1} определим как точку минимума функционала

$$\mu \mapsto \Phi(\mu) + \frac{1}{2\tau} W_2^2(\mu, M_\tau^{n-1}).$$

Если M_τ^{n+1} – достаточно хорошая мера, то для оптимального отображения T_τ^n меры M_τ^{n+1} в M_τ^n поле $-(x - T_\tau^n(x))/\tau$ входит в $\partial\Phi(M_\tau^{n+1})$. Интуитивно это означает, что $(x - T_\tau^n(x))/\tau$ дает аппроксимацию скорости v предельной ($\tau \rightarrow 0$) кривой μ_t , к которой должна сойтись дискретная ломаная $\mu_t^\tau: t \mapsto M_\tau^n$, где $n\tau \leq t < (n+1)\tau$. При этом вектор скорости v должен удовлетворять включению $v \in -\partial\Phi$. Описанная схема действительно работает в случае λ -выпуклых функционалов. Следующая теорема – следствие теоремы 11.2.1 из [11], в которой помимо всего прочего доказаны важные оценки для самих аппроксимаций и даны различные вариационные формулировки рассматриваемой задачи.

ТЕОРЕМА 3.6.24. *Если X – гильбертово пространство, Φ – λ -выпуклый функционал, то при $\tau \rightarrow 0$ дискретные ломаные $\{\mu_t^\tau\}$ сходятся локально равномерно в метрике W_2 к единственному градиентному потоку $\{\mu_t\}$ функционала Φ .*

Единственность градиентного потока $\{\mu_t\}$ доказывается следующим образом. Так как $-v_t \in \partial\Phi(\mu_t)$, то по определению дифференциала

$$\Phi(\sigma) \geq \Phi(\mu_t) + \int_{X^2} \langle v_t(x_2), x_1 - x_2 \rangle \gamma_t(dx_1 dx_2) + \frac{\lambda}{2} W_2^2(\mu_t, \sigma),$$

где γ_t – оптимальный план для пары μ_t, σ . Используя соотношение

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) = \int_{X^2} \langle v_t(x_2), x_1 - x_2 \rangle \gamma_t(dx_1 dx_2),$$

получаем

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) + \frac{\lambda}{2} W_2^2(\mu_t, \sigma) \leq \Phi(\sigma) - \Phi(\mu_t).$$

В частности, для двух различных градиентных потоков μ_t^1, μ_t^2 получаем

$$\frac{d}{dt} W_2^2(\mu_t^1, \mu_t^2) \leq -\lambda W_2^2(\mu_t^1, \mu_t^2).$$

Следовательно, $W_2^2(\mu_t^1, \mu_t^2) \leq e^{-\lambda t} W_2^2(\mu_0^1, \mu_0^2)$, откуда следует единственность.

3.6.9. Риманов объем на $\mathcal{P}^2$. В работе М. К. фон Ренессе и К.-Т. Штурма [152] было введено понятие риманова объема на $\mathcal{P}^2(X)$, где $X = [0, 1]$. Эвристически была определена мера

$$P_\beta = e^{-\beta \text{Ent}_\nu \mu} \cdot P_0,$$

где ν – некоторая фиксированная мера, а P_0 – “равномерное распределение” на $\mathcal{P}^2(X)$. Используем естественный изоморфизм между мерами и монотонными функциями на $[0, 1]$. Для каждой вероятностной меры μ на $[0, 1]$ найдем неубывающую функцию ξ_μ , отображающую меру Лебега на $[0, 1]$ в μ . Тогда

$$\text{Ent}_\lambda \mu = \int_0^1 \varrho \ln \varrho \, dx = \int_0^1 \ln \varrho(\xi_\mu(t)) \, dt = - \int_0^1 \ln \xi'_\mu(t) \, dt.$$

Используя это соотношение, можно построить меру Q_β на пространстве монотонных функций. Мера Q_β определяется своими конечномерными проекциями

$$\begin{aligned} Q_\beta(\xi_1 \in dx_1, \dots, \xi_N \in dx_N) \\ = \frac{1}{Z_{\beta, N}} \exp \left(\beta \sum_{i=1}^{N+1} \ln \frac{x_i - x_{i-1}}{t_i - t_{i-1}} \cdot (t_i - t_{i-1}) \right) q_N(x_1, \dots, x_N). \end{aligned} \quad (3.6.3)$$

Здесь $x_0 = 0$, $x_{N+1} = 1$ и $q_N(x_1, \dots, x_N)$ – некоторое распределение в $\Sigma_N = \{(x_1, \dots, x_N) \subset [0, 1]^N : 0 < x_1 < \dots < x_N < 1\}$, $x_0 = 0$, $x_N = 1$. При требовании некоторых свойств инвариантности и согласованности условных распределений мера q_N единственным образом определена формулой

$$q_N(x_1, \dots, x_N) = C^N \frac{dx_1 \cdots dx_N}{x_1(x_2 - x_1) \cdots (x_N - x_{N-1})(1 - x_N)}.$$

Заметим, что q_N определяет σ -конечную меру. Тем не менее при подходящем выборе константы равенство (3.6.3) задает вероятностную меру. Существование Q_β следует из теоремы Колмогорова. Учитывая соответствие между мерами и функциями, получаем меру P_β на $\mathcal{P}^2(X)$. Полученная мера обладает следующими свойствами: 1) ее носитель состоит из чисто сингулярных мер, 2) она квазиинвариантна относительно преобразований следующего вида; пусть h – C^2 -диффеоморфизм отрезка $[0, 1]$ с $h(0) = 0$; определим его действие на мерах посредством $\mu \mapsto \mu \circ h^{-1}$; тогда верна следующая формула для плотности Радона–Никодима:

$$\frac{dP_\beta(\mu \circ h^{-1})}{dP_\beta(\mu)} = \exp \left(\beta \int_0^1 \ln h'(s) \, d\mu \right) \frac{1}{\sqrt{h'(0) \cdot h'(1)}} \prod_{I \in \text{gaps}(\mu)} \frac{\sqrt{h'(I_-)h'(I_+)}}{|h(I)|/|I|},$$

где $\text{gaps}(\mu)$ – множество сегментов максимальной длины $I = [I_-, I_+]$ с $\mu(I) = 0$.

3.7. Геометрические приложения

3.7.1. Классические задачи дифференциальной геометрии. Уравнение Монжа–Ампера активно изучалось в связи с рядом задач дифференциальной геометрии. К ним относится задача о существовании поверхности заданной гауссовой кривизны. Надо найти такую функцию $\varphi(x)$, что гауссова кривизна ее графика в точке x совпадает с заданной функцией $K(x)$. Задача сводится к решению уравнению Монжа–Ампера

$$\det D^2\varphi(x) = K(x)(1 + |\nabla\varphi(x)|^2)^{(d+2)/2}.$$

Классическая задача Минковского состоит в построении выпуклого многогранника, имеющего заданные нормали $n_1, \dots, n_k$ и заданные площади соответствующих граней $S_1, \dots, S_k$. В непрерывном случае надо найти выпуклую поверхность, кривизна которой в точке с нормалью n равна заданной функции $K(n)$. Оказывается, что решение существует при условии

$$\int_{S^{d-1}} \frac{n}{K(n)} d\mathcal{H}^{d-1} = 0.$$

В решение задачи Минковского весьма существенный вклад внесли Г. Минковский, А. Д. Александров, Г. Леви, Э. Калаби, Л. Ниренберг, А. В. Погорелов, Ш.-Ю. Ченг и Ш.-Т. Яу (см. [72]). В качестве примера приложения транспортной техники обсудим кратко задачу А. Д. Александрова и ее транспортное решение. Для произвольной выпуклой поверхности $F \subset \mathbb{R}^d$ рассмотрим ее обобщенное нормальное отображение в сферу S^{d-1} : $F \ni x \mapsto N(x)$, где под $N(x)$ подразумевается множество всех нормалей к опорным плоскостям к F в точке x . Пусть начало координат находится внутри F . Тогда F можно параметризовать с помощью радиальной функции: $F \ni r(x) = \varrho(x)x$, $x \in S^{d-1}$. Определим многозначное отображение $\alpha_F: S^{d-1} \rightarrow S^{d-1}$, $\alpha_F(x) = N(r(x))$. Определим меру μ_F формулой

$$\mu_F(A) = \mathcal{H}^{d-1} \left(\bigcup_{x \in A} \alpha_F(x) \right).$$

Мера μ_F называется интегральной гауссовой кривизной F . Задача Александрова (поставленная и решенная А. Д. Александровым [153]) состоит в следующем: для заданной меры μ найти такую выпуклую поверхность F , что $\mu = \mu_F$. Для решения задачи Александрова необходимо и достаточно выполнение следующих условий (необходимость легко проверяется):

- 1) $\mu(S^{d-1}) = \mathcal{H}^{d-1}(S^{d-1})$,
- 2) $\mu(S^{d-1} \setminus A) \geq \mathcal{H}^{d-1}(A^*)$ для всякого сферически выпуклого множества $A \subset S^{d-1}$, $A^* = \{y \in S^{d-1} : \langle x, y \rangle \leq 0, x \in A\}$.

В [154] найдено интересное транспортное решение этой задачи. Определим опорную функцию h формулой

$$h(y) = \sup_{x \in S^{d-1}} \varrho(x) \langle x, y \rangle, \quad y \in S^{d-1}.$$

Радиальная и опорная функции связаны соотношениями двойственности через преобразование Лежандра (в этой форме также носящее название преобразования Юнга): $1/\varrho(x) = \sup_{y \in S^{d-1}} \langle x, y \rangle / h(y)$, $y \in S^{d-1}$. Таким образом, $h(y)/\varrho(x) \geq \langle x, y \rangle$ или

$$\ln h(y) - \ln \varrho(x) \geq \ln \langle x, y \rangle. \quad (3.7.1)$$

Точное равенство достигается при $y \in N(r(x))$. Теорема о двойственности наводит на мысль, что пара $(\ln h, -\ln \varrho)$ является решением следующей оптимальной задачи (двойственной задаче Монжа–Канторовича): найти минимум функционала

$$\int_{S^{d-1}} \ln h d\mathcal{H}^{d-1} - \int_{S^{d-1}} \ln \varrho d\mu$$

среди функций $(\ln h, -\ln \varrho)$, удовлетворяющих (3.7.1). Действительно, в [154] было доказано, что решение этой задачи дает решение задачи Александра (искомая поверхность восстанавливается однозначно по любой из данных функций h, ϱ). При этом соответствующая задача Монжа–Канторовича имеет такой вид: для $c(x, y) = \ln \langle x, y \rangle$ найти максимум функционала

$$\int_{S^{d-1} \times S^{d-1}} c(x, y) dm(x, y) \quad \text{при} \quad m_x = \mathcal{H}^{d-1}, \quad m_y = \mu,$$

где $c(x, y) = -\infty$, если $\langle x, y \rangle < 0$. Аналитически задача сводится к уравнению типа Монжа–Ампера на сфере.

3.7.2. Задача Монжа–Канторовича на римановом многообразии. Существование и единственность. Задача Монжа–Канторовича естественно обобщается на римановы многообразия. Решение этой задачи было получено Р. Маккэнном в работе [155]. Ниже мы изложим основные элементы его конструкции. По этой теме см. [10], [156]–[158], [86], [88], [72]. Пусть (M, g) – гладкое (класса C^3) связное компактное риманово многообразие. Через $T_x M$ обозначим касательное пространство в точке $x \in M$. Расстояние $d(x, y)$ между точками $x, y \in M$ определяется как

$$d(x, y) = \inf_{\gamma_{x,y}} \int \sqrt{g_{ij}(\gamma) \dot{\gamma}^i \dot{\gamma}^j} dt,$$

где инфимум берется по всевозможным кривым γ , соединяющим x и y . Геодезической называется кривая $t \mapsto \gamma(t)$, удовлетворяющая условию $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$, где $\dot{\gamma} \in TM_{\gamma(t)}$ – вектор скорости кривой, а ∇ – ковариантная производная. Экспоненциальное отображение $\exp_x: TM_x \rightarrow M$ определяется как отображение $TM_x \ni v \mapsto y$, где y – конечная точка геодезической с началом в x и начальным касательным вектором v . Задача Монжа для квадратичного функционала имеет следующую естественную постановку. Для двух вероятностных мер μ и ν , имеющих плотности ϱ_μ и ϱ_ν относительно риманова объема dx , найти минимум функционала

$$M(\mu, \nu, T) = \int_M d^2(x, T(x)) \mu(dx), \quad T \in T(\mu, \nu).$$

Далее в этом пункте $c(x, y) = d^2(x, y)$. Аналогично плоскому случаю рассмотрим операцию выпуклого сопряжения $\varphi^{c+}(y) = \inf_{x \in M} (c(x, y) - \varphi(x))$. Естественно ожидать, что оптимальная транспортировка примет следующий вид: $T(x) = \exp_x(-\nabla \varphi(x))$, где $\exp_x(v)$ – экспоненциальное отображение в точке x , а потенциал φ удовлетворяет условию $\varphi^{cc} = \varphi$. Отметим, что для $\mathbb{R}^d$ это дало бы $T(x) = x - \nabla \varphi(x)$, а не $T(x) = x + \nabla \varphi(x)$, как у нас было выше. Как и в евклидовом случае, соответствующая задача Канторовича на пространстве мер имеет двойственную формулировку $J(\varphi, \psi) \rightarrow \sup$, где функционал J рассматривается на парах непрерывных функций φ, ψ , удовлетворяющих условию $\varphi(x) + \psi(y) \leq d^2(x, y)$. Результаты теории двойственности полностью переносятся на случай риманова многообразия. В частности, мы получаем существование потенциала φ . Затем доказывается, что $\varphi^{cc} = \varphi$, $\psi = \varphi^c$. Основная задача теперь – доказать, что соответствующее отображение задает искомый

перенос масс. Следующая лемма (теорема Радемахера на многообразиях) доказана в работе [155].

ЛЕММА 3.7.1. *Всякая липшицева функция φ на многообразии M дифференцируема во всех точках M за исключением множества меры нуль (относительно риманова объема).*

ЗАМЕЧАНИЕ 3.7.2. Представляя потенциал φ как результат операции сопряжения, можно доказать, что φ – липшицева функция.

Основная сложность – недифференцируемость римановой метрики $d(x, y)$ (например, в сопряженных точках, где теряется единственность соединяющей геодезической). Интуитивно ясно (хотя требует непростого обоснования), что $d(x, y)$ – гладкая функция для y из малой окрестности x . В общем случае (см. детали в [155]) квадрат расстояния $f(y) = d^2(x, y)$ супердифференцируем, т. е. существует такой вектор $p \in TM_x$, что для всех $v \in TM_x$ имеем

$$f(\exp_x(v)) \leq f(x) + g^{ij}(x)p^i v^j + o(|v|).$$

Аналогично определяется субдифференцируемость. Из локальной дифференцируемости $d(x, y)$ и неравенства треугольника можно вывести супердифференцируемость функции $y \mapsto d(x, y)$ и супердифференцируемость $d^2(x, y)$. Из замечания 3.7.2 и версии теоремы Радемахера следует, что $\nabla\varphi$ существует μ -почти всюду. Более того, верен следующий факт, вытекающий из определения операции сопряжения и супердифференцируемости функции $c(x, y)$ (см. [155]).

ЛЕММА 3.7.3. *Для всех $x, y \in M$ верно неравенство $\varphi(x) + \varphi^c(y) \leq c(x, y)$. Если x – точка дифференцируемости функции φ , то равенство*

$$\varphi(x) + \varphi^c(y) = c(x, y)$$

верно ровно тогда, когда $y = \exp_x(-\nabla\varphi)$. При этом $|\nabla\varphi(x)| = d(x, y)$.

ДОКАЗАТЕЛЬСТВО. Неравенство следует непосредственно из определения. Пусть x – точка дифференцируемости функции φ . Если $\varphi(x) + \varphi^c(y) = c(x, y)$ для некоторой точки $y \in M$, то для всех $z \in M$

$$c(z, y) - \varphi(z) - \varphi^c(y) \geq c(x, y) - \varphi(x) - \varphi^c(y).$$

Положим $z = \exp_x(v)$, $v \in TM_x$ и $f(z) = c(z, y)$. Тогда

$$f(z) \geq f(x) - \varphi(x) + \varphi(z) = f(x) - \varphi(x) + \varphi(x) + g(\nabla\varphi(x), v) + o(|v|).$$

Следовательно, $\nabla\varphi(x)$ принадлежит субградиенту f в точке x . С другой стороны, в силу классических результатов дифференциальной геометрии (теорема Хопфа–Ринова) существует минимальная геодезическая, соединяющая y и x с начальным касательным вектором v . Из этого можно вывести, что f супердифференцируема в x , а v принадлежит ее супердифференциалу в x . Отсюда следует, что f дифференцируема в x и $y = \exp_x(-\nabla\varphi(x))$, $v = \nabla\varphi(x)$. Первая часть утверждения доказана.

Теперь для доказательства второй части достаточно доказать, что равенство $\varphi(x) + \varphi^c(y) = c(x, y)$ достигается хотя бы в одной точке. В силу того, что

$\varphi = \varphi^{cc}$, замечание 3.7.2 влечет, что функция φ^c липшицева. Поэтому в соотношении $\varphi^{cc}(x) = \inf_{y \in M} (c(x, y) - \varphi^c(y))$ инфимум достигается в некоторой точке (так как M – компактно и f непрерывна). Лемма доказана.

ТЕОРЕМА 3.7.4. *Образ μ относительно $T(x) = \exp_x(-\nabla\varphi(x))$ равен ν .*

ДОКАЗАТЕЛЬСТВО. Пусть $f \in C_b(M)$. Докажем, что интеграл от f по мере ν равен интегралу от $f \circ T$ по мере μ . Для $\varepsilon \in (-1, 1)$ рассмотрим возмущения $\psi_\varepsilon(y) := \psi(y) + \varepsilon h(y)$ и $\varphi_\varepsilon(x) = (\psi_\varepsilon)^c(x) = \inf_{y \in M} (c(x, y) - \psi(y) - \varepsilon h(y))$. Пусть функция φ дифференцируема в точке x . Тогда минимум $c(x, y) - \psi(y)$ достигается в единственной точке $t(x)$ в силу леммы 3.7.3. Для малых ε минимум $c(x, y) - \psi(y) - \varepsilon h(y)$ достигается в близкой точке $y_\varepsilon = t(x) + o(1)$. Таким образом, для всех $y \in M$ имеем

$$c(x, t(x)) - \varphi(t(x)) - \varepsilon f(t_\varepsilon(x)) \leq \varphi_\varepsilon(x) \leq c(x, y) - \varphi(y) - \varepsilon f(y).$$

Возьмем $y = t(x)$. Тогда $\varphi_\varepsilon(x) = \varphi(x) - \varepsilon f(t(x)) + o(\varepsilon)$, причем выполнена оценка $o(\varepsilon) \leq \varepsilon \sup_{x \in M} |f(x)|$. Так как $\varepsilon \mapsto J(\varphi_\varepsilon, \psi_\varepsilon)$ имеет максимум при $\varepsilon = 0$, то

$$\lim_{\varepsilon \rightarrow 0} \frac{J(\varphi_\varepsilon, \psi_\varepsilon) - J(\varphi, \psi)}{\varepsilon} = \int f d\nu + \lim_{\varepsilon \rightarrow 0} \int \frac{\varphi_\varepsilon - \varphi}{\varepsilon} d\mu = \int f d\nu + \int f(t(x)) d\mu = 0$$

по теореме о мажорируемой сходимости. Теорема доказана.

Имеет место следующая теорема единственности [155].

ТЕОРЕМА 3.7.5. *Отображение $T = \exp_x(\nabla\varphi)$ минимизирует функционал $M(\mu, \nu, S)$ среди отображений $S \in T(\mu, \nu)$, причем с точностью до эквивалентности оно единственное минимизирующее отображение.*

3.7.3. Задача Монжа–Канторовича на римановом многообразии. Замена переменных, тензор Риччи и геометрические неравенства. Как и в плоском случае, оптимальное отображение T на римановом многообразии почти всюду дифференцируемо. Здесь ситуация осложняется тем, что T является композицией двух отображений: градиента функции и экспоненциального отображения $TM_x \ni v \mapsto \exp_x(v)$. Последнее дифференцируемо, вообще говоря, не во всех $v \in TM_x$. Для $x \in M$ определим множество $\text{cut}(x) \subset M$ (cut locus), как множество точек y , для которых существует кратчайшая геодезическая $t \mapsto \exp(tv)$, $0 \leq t \leq 1$, соединяющая x и y , но любая такая геодезическая не является кратчайшей при $t \in [0, 1 + \varepsilon)$. Хорошо известно, что экспоненциальное отображение $v \mapsto \exp_x(v)$ дифференцируемо для всех v , удовлетворяющих условию $\exp(v) \notin \text{cut}(x)$. Очевидно, что $|v| = d(x, \exp_x(v))$ для всех v , для которых $\exp(tv)$ не пересекает $\text{cut}(x)$, $0 \leq t \leq 1$. Если $x \notin \text{cut}(y)$, то $y = \exp_x(-\nabla_x[d^2(y, x)]/2)$. В [99] получена следующая формула замены переменных. Как и в плоском случае, под гессианом понимается гессиан в смысле Александрова (точнее, его обобщение на многообразия).

ТЕОРЕМА 3.7.6. *Пусть $\mu = f dx$ и $\nu = g dx$ – вероятностные меры с компактными носителями, dx – риманов объем. Пусть $T(x) = \exp_x(-\nabla\varphi(x))$ – оптимальное отображение μ в ν для “квадратичной” функции стоимости*

$d^2(x, y)$. Тогда существует борелевское множество K полной меры μ , для которого (i) функция φ обладает гессианом $\text{Hess}_x \varphi$ в каждой точке $x \in K$, причем $T(x) \notin \text{cut}(x)$, (ii) для $x \in K$, положив $Y = d(\exp_x)_{-\nabla \varphi(x)}$ и $H = (1/2) \text{Hess } d^2(x, y)|_{y=T(x)}$, имеем $f(x) = g(T(x)) \det[Y(H - \text{Hess}_x \varphi)]$.

Для приложений оптимальной транспортировки большое значение имеет интерполирующее семейство отображений T_t , в плоском случае определяемое как $T_t(x) = (1 - t)x + tT(x)$. Аналогичным образом, для многообразий положим

$$T_t(x) = \exp_x(-t\nabla \varphi(x)).$$

Это семейство отображений оказывается тесно связанным с важным понятием полей Якоби. Напомним, что полем Якоби $J(t)$ называется любое поле, полученное путем вариации семейства геодезических: $J(t) := \frac{d}{ds} \gamma_s(t) \Big|_{s=0}$, где $[0, 1] \ni t \mapsto \gamma_s(t)$ – геодезическая для любого s . Как известно из римановой геометрии, произвольное поле Якоби удовлетворяет уравнению

$$\nabla_{\dot{\gamma}(t)}^2 J(t) + R_{\gamma(t)}(\dot{\gamma}(t), J(t))\dot{\gamma}(t) = 0. \quad (3.7.2)$$

Здесь $R_x: TM_x \times TM_x \times TM_x \mapsto TM_x$ – тензор Римана в точке x , $\dot{\gamma}(t) = \frac{d\gamma}{dt}(t)$, $\nabla_{\dot{\gamma}(t)}^2$ – вторая ковариантная производная вдоль γ .

Дифференциальные уравнения вдоль геодезических удобно записывать в ортогональном движущемся базисе. Зафиксируем в $T_{\gamma(0)}M$ ортогональный базис $\{e_1(0), \dots, e_n(0)\}$ и рассмотрим его параллельный перенос вдоль γ . Пусть $Y(t)$ – координаты $J(t)$ в базисе $\{e_1(t), \dots, e_n(t)\}$. Уравнение (3.7.2) перепишется в виде $Y'' + RY = 0$, где $R = R(t)$ – некоторая симметричная матрица, след которой дает значение тензора Риччи на векторе $\dot{\gamma}$. Заметим, что для любого $v_0 \in TM_{x_0}$ поле $t \mapsto A(t)(x_0)v_0 := d_x(T_t)|_{x=x_0} v_0 = d_x(\exp_x(-t\varphi(x)))|_{x=x_0} v_0$, где $d_x(T_t)|_{x=x_0}$ – дифференциал отображения $x \mapsto T_t(x)$ в x_0 , представляет собой поле Якоби. При этом формула замены переменных для семейства интерполирующих мер принимает вид $\varrho_0(x) = \varrho_t(T_t x) \det A(t)$, где $\varrho_0 = f$, $\varrho_1 = g$, A_t – единственное тензорное поле Якоби вдоль геодезической $\exp(-t\nabla \varphi)$, заданное соотношением $A(0) = I$, $A'(0) = -\text{Hess } \varphi$.

Посмотрим, как с помощью этой техники можно получить обобщение неравенства Брунна–Минковского на многообразия. Для двух точек x и y положим

$$Z_t(x, y) = \{z \in M; d(x, z) = td(x, y), d(z, y) = (1 - t)d(x, y)\}.$$

Если две точки соединяет единственная геодезическая, то $Z_t(x, y)$ – единственная точка.

Тензор кривизны Риччи будем обозначать символом Ric .

ТЕОРЕМА 3.7.7. Пусть $\mu = e^{-V} dx$ – вероятностная мера на M . Предположим, что для некоторого $\lambda \in \mathbb{R}$ тензор Бакри–Эмери многообразия M удовлетворяют соотношению $\text{Ric} + \text{Hess } V \geq \lambda$. Пусть f, g, h – неотрицательные функции, удовлетворяющие неравенству

$$h(z) \geq e^{-\lambda s(1-s)d^2(x,y)/2} f^{1-s}(x) g^s(y)$$

для всех $x, y \in M$, $s \in [0, 1]$, $z \in Z_s(x, y)$. Тогда

$$\int_M h d\mu \geq \left(\int_M f d\mu \right)^{1-s} \left(\int_M g d\mu \right)^s.$$

Поясним идею доказательства. Пусть ϱ_s – плотность меры $\mu \circ T_s^{-1}$, где T_t определено перед формулой (3.7.2). Пусть $\gamma(s)$ – геодезическая, выходящая из точки x (указание на которую мы опускаем) со скоростью $-\nabla\varphi(x)$. Как и в теореме 3.1.3, достаточно доказать неравенство $h(\gamma(s)) \exp(-V(\gamma(s))) \geq \varrho_s(\gamma(s))$ (потому что интеграл от правой части по формуле замены переменных равен 1). Положим $\psi(t) = -\ln \det d(T_t)$. В силу формулы замены переменных

$$f(\gamma(0))e^{-V(\gamma(0))} = \varrho_s(\gamma(s))e^{-\psi(s)} = g(\gamma(1))e^{-V(\gamma(1))-\psi(1)}.$$

В силу условия на функцию h имеем

$$h(\gamma(s))e^{-V(\gamma(s))} \geq f^{1-s}(\gamma(0))g^s(\gamma(1)) \exp\left(-V(\gamma(s)) - \frac{1}{2}\lambda s(1-s)d^2(\gamma(0), \gamma(1))\right).$$

В силу уравнений выше логарифм правой части, деленной на $\varrho_s(\gamma(s))$, равен

$$(1-s)V(\gamma(0)) + sV(\gamma(1)) - V(\gamma(s)) + s\psi(1) - \psi(s) - \frac{1}{2}\lambda s(1-s)d^2(\gamma(0), \gamma(1)).$$

Положим $\alpha(t) := V(\gamma(t)) + \psi(t)$. Учитывая, что $\psi(0) = 0$, получаем, что достаточно доказать оценку $(1-s)\alpha(0) + s\alpha(1) - \alpha(s) \geq \lambda s(1-s)d^2(\gamma(0), \gamma(1))/2$. Заметим, что $\alpha''(t) = \text{Hess } V(\gamma(t))(\dot{\gamma}(t), \dot{\gamma}(t)) + \psi''(t)$. Если мы покажем, что выполнено неравенство $\psi''(t) \geq \text{Ric}_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))$, то из оценки на тензор Бакри–Эмери мы получим неравенство $\alpha''(t) \geq \lambda|\dot{\gamma}(t)|^2$, что влечет искомое соотношение. Нужная нам оценка действительно имеет место. Ее доказательство основано на методах, развитых в римановой геометрии для доказательств теорем сравнения. Перейдем в движущийся ортогональный базис и воспользуемся тем, что $A(t) = d_x(T_t)$ является решением уравнения $A''(t) + R(t)A(t) = 0$. В силу симметричности начальных условий симметричность A сохраняется. Положим $B = A'A^{-1}$. Дифференцируя определитель, находим $\psi' = -\text{Tr } B$, $\psi'' = -\text{Tr } B'$. Дифференцируя $B^* - B$ и используя симметричность R , нетрудно показать, что $(B^* - B)' = 0$, следовательно, B симметрично. Поэтому

$$\begin{aligned} \psi'' &= -\text{Tr } B' = \text{Tr } B^2 + \text{Tr } R \geq \frac{(\text{Tr } B)^2}{d} + \text{Tr } R \geq \frac{(\text{Tr } B)^2}{d} + \text{Tr } R \\ &= \frac{(\psi')^2}{d} + \text{Ric}(\gamma(t))(\dot{\gamma}(t), \dot{\gamma}(t)), \end{aligned}$$

что и требовалось.

Приведем еще результаты работы [99]. Коэффициент искажения объема определяется следующим образом: $v_t(x, y) = \lim_{r \rightarrow 0} Z_t(x, B_r(y)) / \text{vol}(B_{tr}(y))$. Коэффициент искажения объема связан со степенью “искривленности” многообразия. Для евклидова пространства $v_t = 0$. Положим теперь

$$S_k(t) = \frac{\sin(\sqrt{k}t)}{\sqrt{k}t}, \quad \text{если } k > 0, \quad S_k(t) = \frac{\sinh(\sqrt{-k}t)}{\sqrt{-k}t}, \quad \text{если } k < 0.$$

ТЕОРЕМА 3.7.8. Для d -мерного многообразия M , удовлетворяющего условию $\text{Ric} \geq (d-1)k$, выполнено неравенство

$$v_t(x, y) \geq \left(\frac{S_k(td(x, y))}{S_k(d(x, y))} \right)^{d-1}.$$

Точное равенство достигается на модельных пространствах постоянной секционной кривизны (сфера S^d , $\mathbb{R}^d$ и пространство Лобачевского H^d).

По этой теме см. также [159].

3.7.4. Пространства неотрицательной кривизны. Многие аналитические и геометрические свойства многообразий выражаются в терминах тензора Риччи. Особенно много интересных результатов известно о многообразиях, для которых тензор Риччи ограничен снизу постоянным тензором: $\text{Ric} \geq K$ (другая форма записи: $\text{Ric} \geq K \cdot g$, g – метрический тензор на M). Упомянем наиболее известные.

1. *Теорема сравнения Бишоп–Громова.* Если $x \in M$, то объем шара с центром x радиуса r растет не быстрее объема шара на модельном пространстве постоянной секционной кривизны K . К модельным пространствам относятся сфера кривизны K , если $K > 0$, $\mathbb{R}^d$, если $K = 0$, пространство Лобачевского кривизны K , если $K < 0$.

2. *Теорема Бонне–Майерса о компактности.* Если $K > 0$, то диаметр M не превосходит $\pi\sqrt{(d-1)/K}$.

3. *Изопериметрическое неравенство Леви–Громова.* Если $K > 0$, то

$$\frac{\nu^+(\partial A)}{\nu(A)^{(d-1)/d}} \geq \frac{\mu^+(\partial B)}{\mu(B)^{(d-1)/d}}.$$

Здесь $A \subset M$, ν – риманов объем на M , ν^+ – поверхностная мера на M ; S – модельная сфера, $B \subset S$ – шар в римановой метрике на сфере S и μ , μ^+ – аналогичные меры на сфере. Предполагается, что $\nu(A)/\nu(M) = \mu(B)/\mu(S)$. Неравенство Леви–Громова тесно связано с гауссовским изопериметрическим неравенством, которое можно считать “бесконечномерной версией” последнего.

Это далеко не полный список интересных результатов в этой области. Можно было бы упомянуть еще точные неравенства Соболева, неравенства типа Ли–Яу, неравенства Харнака, градиентные оценки и оценки ядер теплопроводности, различные топологические результаты для пространств неотрицательной кривизны. Сюда же примыкает и теорема Бакри–Эмери, которая влечет логарифмическое неравенство Соболева. Ключевой аналитический объект в этом результате – тензор Бакри–Эмери $\text{Hess } V + \text{Ric}$. Оказывается, именно тензор Бакри–Эмери отвечает за многие аналитические свойства многообразий. Идея изучения геометрии многообразий путем наделения последних вероятностными мерами и рассмотрения сохраняющих меру изометрий предложена М. Громовым [160], назвавшим этот объект “метрическими пространствами с мерами” (“metric-measure space”); для набора из пространства, метрики и меры на нем А. М. Вершик [161], [162] предложил термин “метрическая тройка”. Этот материал подробно изложен в книге [17] (см. также [163]–[165], [40]), поэтому мы ограничимся кратким изложением основных концепций и формулировками основных результатов. В основе этих результатов лежит понятие

выпуклости подходящих энергетических функционалов на пространстве мер с метрикой Канторовича. В литературе распространен термин “переносная выпуклость” (“displacement convexity”). Памятуя о том, что интерполяция мер с помощью оптимальных переносов задает геодезические в пространстве мер, можно также говорить о геодезической выпуклости. В качестве формальной техники для работы с выпуклыми функционалами применяется формализм Отто. Мы уже обсуждали вычисление градиентов от таких функционалов. Следующая формула (см. [17]) понимается эвристически (в рамках того же формализма Отто). Пусть dx – риманов объем и функционал $\mathcal{F}$ задан формулой

$$\mathcal{F}(\mu) = \int_M F(\varrho) d\nu, \quad \mu = \varrho d\nu, \quad \nu = e^{-V} dx,$$

$v = \nabla f$ – векторное поле, задающее кривую $\dot{\varrho} + \operatorname{div}(\nabla f \cdot \varrho) = 0$. Тогда непосредственным вычислением проверяется следующая формула.

ТЕОРЕМА 3.7.9. *Гессиан на пространстве $\mathcal{P}^2(\mathbb{R}^d)$ имеет вид*

$$\begin{aligned} \operatorname{Hess}_\mu \mathcal{F}(\nabla f, \nabla f) &= \int [\|D^2 f\|_{\mathcal{H}, \mathcal{F}}^2 + (\operatorname{Hess} V + \operatorname{Ric})(\nabla f, \nabla f)] p(\varrho) d\nu \\ &+ \int (Lf)^2 p_2(\varrho) d\nu = \int \Gamma_2(f) p(\varrho) d\nu + \int (Lf)^2 p_2(\varrho) d\nu, \\ Lf &= \Delta f - \langle \nabla V, \nabla f \rangle, \quad p(\varrho) = \varrho F'(\varrho) - F(\varrho), \quad p_2(\varrho) = \varrho p'(\varrho) - p(\varrho). \end{aligned}$$

Выражение (итерация диффузионного оператора L)

$$\Gamma_2(f) = L\left(\frac{|\nabla f|^2}{2}\right) - \langle \nabla f, \nabla Lf \rangle$$

носит название оператора “carré du champ itéré”. Нетривиальное и весьма полезное равенство

$$\Gamma_2(f) = \|D^2 f\|_{\mathcal{H}, \mathcal{F}}^2 + (\operatorname{Hess} V + \operatorname{Ric})(\nabla f, \nabla f)$$

носит название (обобщенной) формулы Бохнера. Для $V = 0$ оно принимает вид

$$\frac{1}{2} \Delta |\nabla f|^2 = \langle \nabla f, \nabla \Delta f \rangle + \|D^2 f\|_{\mathcal{H}, \mathcal{F}}^2 + \operatorname{Ric}(\nabla f, \nabla f). \quad (3.7.3)$$

Тензор Бакри–Эмери определяется посредством $R_{\infty, \mu} = \operatorname{Ric} + \operatorname{Hess} V$. Рассмотрим модифицированный (зависящий от размерности) тензор Бакри–Эмери

$$R_{N, \mu} = \operatorname{Ric} + \operatorname{Hess} V - \frac{\nabla V \oplus \nabla V}{N - d}, \quad N > d.$$

Непосредственно проверяется неравенство

$$\Gamma_2(f) \geq \frac{(Lf)^2}{N} + \operatorname{Ric}_{N, \mu}(\nabla f, \nabla f).$$

Наконец, отсюда и из формулы для гессиана вытекает важная оценка

$$\operatorname{Hess}_\mu \mathcal{F} \geq \int \operatorname{Ric}_{N, \mu}(\nabla f, \nabla f) p(\varrho) d\nu + \int (Lf)^2 \left[p_2(\varrho) + \frac{p(\varrho)}{N} \right] d\nu.$$

В частности, из этого соотношения можно формально вывести свойство выпуклости соответствующих функционалов.

ПРИМЕР 3.7.10. Если $\text{Ric}_{\infty, \mu} \geq 0$, то $\text{Hess}_{\mu} \mathcal{F}(\nabla f, \nabla f) \geq 0$ для $F(\varrho) = \varrho \ln \varrho$. Если $\text{Ric}_{N, \mu} \geq 0$, то $\text{Hess}_{\mu} \mathcal{F}(\nabla f, \nabla f) \geq 0$ для $F(\varrho) = -N(\varrho^{1-1/N} - \varrho)$.

ОПРЕДЕЛЕНИЕ 3.7.11. Пусть V – дважды непрерывно дифференцируемая функция на гладком римановом многообразии M с метрикой ϱ и $\nu = e^{-V} dx$. Будем говорить, что пространство (M, ϱ, ν) принадлежит классу $CD(K, N)$, где $K \in \mathbb{R}$, $N \in [n, \infty]$, если $\text{Ric}_{N, \nu} \geq K$.

Случай $N = n$ имеет смысл при постоянном V . В этом случае полагаем по определению $\text{Ric}_{N, \mu} = \text{Ric}$.

ПРИМЕР 3.7.12. Следующие примеры мер на $\mathbb{R}$ со стандартной метрикой дают модельные примеры пространств из $CD(K, N)$:

- 1) $M = (-\pi\sqrt{N-1}/(2K), \pi\sqrt{N-1}/(2K))$, $K > 0$, $1 < N < \infty$, $\nu = \cos^{N-1}(\sqrt{K/(N-1)}x) dx$,
- 2) $M = \mathbb{R}$, $K < 0$, $1 < N < \infty$, $\nu = \cosh^{N-1}(\sqrt{-K/(N-1)}x) dx$,
- 3) $M = (0, +\infty)$, $K = 0$, $1 < N < \infty$, $\nu = x^{N-1} dx$,
- 4) $K > 0$, $N = \infty$, $\nu = e^{-Kx^2/2} dx$.

Принцип, лежащий в основе аналитических теорем для пространств с мерой и метрикой, а также теорем сравнения римановой геометрии, гласит: аналитические свойства пространств $CD(K, N)$ не хуже, чем у соответствующих модельных пространств.

ЗАМЕЧАНИЕ 3.7.13 (касательный формализм). Техника вывода аналитических соотношений, основанная на понятии геодезической выпуклости функционалов, дает далеко идущие обобщения многих упомянутых выше результатов для $\mathbb{R}^d$ на римановы многообразия. При этом используется следующие элементарное соотношение: если $D^2F \geq K$, то

$$\begin{aligned} F(tx + (1-t)y) &\leq tF(x) + (1-t)F(y) - \frac{Kt(1-t)}{2}|x-y|^2, \quad 0 \leq t \leq 1, \\ F(y) &\geq F(x) + \langle F(x), y-x \rangle - \frac{K}{2}|y-x|^2, \end{aligned} \quad (3.7.4)$$

обобщенное должным образом на функционалы на $\mathcal{P}^2$.

ПРИМЕР 3.7.14 (неравенства Соболева и касательный формализм). Пусть $K > 0$, $N = \infty$, $T = I + \nabla \varphi$ – оптимальная транспортировка μ в ν . Рассмотрим стандартную интерполяцию с помощью геодезической в пространстве мер между ν и $\mu = \varrho d\nu$. Воспользуемся выпуклостью этого функционала. Применим соотношение (3.7.4) и формулу

$$\nabla \mathcal{F}(\nabla \varphi) = \int \left\langle \frac{\nabla p(\varrho(x))}{\varrho(x)}, \nabla \varphi(x) \right\rangle \varrho(x) dx.$$

Отсюда вытекает хорошо известное нам в плоском случае соотношение

$$\int \varrho \ln \varrho d\nu \leq - \int \langle \nabla \varrho, \nabla \varphi \rangle d\nu - \frac{K}{2} W_2^2(\mu, \nu),$$

следствием которого является логарифмическое неравенство Соболева. Точно так же стандартные неравенства Соболева на $\mathbb{R}^d$ следуют из выпуклости функционала, задаваемого интегралом от $-\varrho^{1-1/d}$.

В частности, мы получаем теорему Бакри–Эмери [166].

ТЕОРЕМА 3.7.15. Пусть M – риманово многообразие с вероятностной мерой $\mu = e^{-V} dx$, где dx – риманов объем. Если $R_{\infty, \mu} \geq K$, $K > 0$, то выполнено логарифмическое неравенство Соболева $\text{Ent}_\mu f^2 \leq \frac{2}{K} \int_M |\nabla f|^2 d\mu$.

Есть различные характеристики $CD(K, N)$ -пространств (см. [143], [136]). Приведем основные результаты, следуя [17].

ТЕОРЕМА 3.7.16. Гладкое многообразие M с мерой $\nu = e^{-V} dx$, где $V \in C^2(M)$, принадлежит классу $CD(K, N)$ в точности тогда, когда функционал

$$\mathcal{F}(\mu) = \int_M F(\varrho) d\nu$$

является выпуклым для $F(\varrho) = -N(\varrho^{1-1/N} - \varrho)$, $F(\varrho) = \varrho \ln \varrho$ для $N = +\infty$.

Эта теорема позволяет определить класс $CD(K, N)$ -пространств в слабом смысле как пределов $CD(K, N)$ -пространств в топологии сходимости Громова–Хаусдорфа. Такой предел является метрическим пространством с мерой, но уже может не быть гладким многообразием. В частности, можно определить пространства с ограниченной снизу кривизной Риччи как LS-пространства, для которых соответствующий функционал является выпуклым в пространстве мер с метрикой Канторовича. Существуют также различные “локальные” характеристики $CD(K, N)$ -пространств. Мы уже видели в теоремах 3.7.7 и 3.7.8, что при работе с объемами, геометрическими неравенствами и формулой замены переменной естественным образом возникают коэффициенты искажения для сдвигов вдоль геодезических. Эти коэффициенты в силу своей вариационной природы являются решениями уравнения Якоби. Его решения оцениваются стандартными методами. Доказательство следующего факта см. в [17].

ТЕОРЕМА 3.7.17. Пусть $K > 0$, $N < \infty$. Слабое $CD(K, N)$ -пространство удовлетворяет неравенству $\text{diam}(\text{supp } \nu) \leq \pi \sqrt{(N-1)/K}$.

3.7.5. Геометрические потоки: потоки гауссовой кривизны и параболические уравнения. В этом пункте мы рассмотрим так называемую гауссову транспортировку мер – класс отображений мер, близких к оптимальным. Общий вид гауссовой транспортировки: $T = \varphi \nabla \varphi / |\nabla \varphi|$, где φ – функция (потенциал) с выпуклыми множествами $A_t = \{\varphi \leq t\}$. Для широкого класса мер существование таких отображений установлено в [167]. Название “гауссовы” появилось потому, что данный объект тесно связан с гауссовыми (сферическими или нормальными) отображениями поверхностей, а также потоками гауссовой кривизны. Рассмотрим простой пример. Пусть $\gamma: S^1 \rightarrow \mathbb{R}^2$ – диффеоморфное вложение окружности в плоскость, т.е. $\gamma(S^1)$ – гладкий, замкнутый контур без самопересечений, ограничивающий односвязную область. Пусть $\gamma_t: S^1 \rightarrow \mathbb{R}^2$ – семейство кривых, заданных уравнениями

$$\gamma_0 = \gamma, \quad \frac{d}{dt} \gamma_t(s) = -K(\gamma_t(s)) \cdot n(\gamma_t(s)),$$

где $K(\gamma_t(s))$ и $n(\gamma_t(s))$ – кривизна и внешняя нормаль кривой γ_t в соответствующей точке. Другими словами, γ_t – поток кривых, стартующих из γ и движущихся со скоростью K по направлению нормали (т. е. внутрь, если $K > 0$, и наружу, если $K < 0$). В англоязычной литературе этот объект носит название “curve shortening flow” и является простейшим примером геометрических потоков, к которым относятся потоки средней и гауссовой кривизны, а также потоки Риччи. Легко видеть, что площадь области A_t , которую охватывает γ_t , меняется с постоянной скоростью

$$\frac{d}{dt} \mathcal{H}^2(A_t) = - \int_{\gamma_t} K d\mathcal{H}^1 = -2\pi.$$

Последнее равенство выполнено по теореме Гаусса–Бонне. Нетрудно доказать, что длина контура ∂A_t меняется по формуле

$$\frac{d}{dt} \mathcal{H}^1(\partial A_t) = - \int_{\gamma_t} K^2 d\mathcal{H}^1.$$

Формула для вариации площади показывает, что со временем кривые должны сомкнуться к некоторому множеству меры нуль. Это действительно так. Более того, известно, что если начальная кривая γ была невыпуклой (т. е. множество A_0 не было выпукло), то за конечное время они становятся выпуклыми. Далее выпуклость сохраняется и множества A_t оказываются вложенными друг в друга: $A_{t_2} \subset A_{t_1}$, если $t_2 \geq t_1$. Со временем A_t становится все “круглее”, что означает, что кривизна γ_t распределяется все более равномерно на кривой. Наконец, кривые сжимаются в точку. Естественным обобщением этой конструкции являются потоки многообразий. Пусть $M \subset \mathbb{R}^d$ есть $(d-1)$ -мерная поверхность в $\mathbb{R}^d$. Геометрическим потоком называется семейство вложений $M_t \subset \mathbb{R}^d$, $M_0 = M$, движущихся по направлению нормали n с некоторой скоростью F : $\dot{x}(s) = -F(x) \cdot n(x)$. Если F равна средней кривизне H , то говорят о потоке средней кривизны. Если F равна гауссовой кривизне K , то говорят о потоке гауссовой кривизны. В отличие от одномерного случая и те, и другие потоки гораздо менее регулярны. В многомерном случае за конечное время может произойти разрыв и связное многообразие распадется на части. Впрочем, если начальная поверхность была выпуклой, то оба типа потоков сохраняют выпуклость и поверхность сожмется в точку за конечное время. В случае потоков средней кривизны это было доказано в [168], а в случае потоков гауссовой кривизны – в [169]. Поясним связь потоков кривизны с транспортной мерой. Пусть $\{\gamma_t\}$ – поток одномерных кривых на плоскости и $\gamma = \gamma_0$ – гладкая выпуклая кривая. Как мы знаем, все остальные кривые γ_t остаются выпуклыми, а поток существует некоторое конечное время T . Пусть φ – функция, равная $T - t$ на γ_t . Каков образ меры Лебега при отображении $T = \varphi \nabla \varphi / |\nabla \varphi|$? Заметим, что $T = \varphi \cdot n$, где n – нормаль к γ_t . Пусть v – касательный вектор к γ_t , так что (n, v) – ортонормированный базис. Продифференцируем T в этом базисе. Учитывая, что φ постоянна на γ_t , получаем $\partial_n T = \partial_n \varphi \cdot n + \varphi \cdot \partial_n n$, $\partial_v T = \varphi \cdot \partial_v n$. Заметим, что $\partial_v n = Kv$ (формула Френе), $\partial_n \varphi = |\nabla \varphi|$. Следовательно, мы получаем $\det DT = K\varphi |\nabla \varphi|$. По формуле замены переменных под действием отображения T мера Лебега переходит в меру ϱdx , где

$\varrho(\varphi \nabla \varphi / |\nabla \varphi|) K \varphi |\nabla \varphi| = 1$. Линии уровня $\{x: \varphi(x) = T - t\}$ функции φ меняются со скоростью $1/|\nabla \varphi|$. Так как мы имеем дело с потоком кривизны, то $K|\nabla \varphi| = 1$. Следовательно, $\varrho(\varphi \nabla \varphi / |\nabla \varphi|) = 1/\varphi$, т.е. образ меры Лебега относительно отображения T есть мера $\nu = dx/|x|$. В общем случае верна следующая теорема.

ТЕОРЕМА 3.7.18. Пусть $A \subset \mathbb{R}^d$ – компактное выпуклое множество и $\mu = \varrho_0 dx$ – вероятностная мера на A , эквивалентная ограничению меры Лебега. Пусть $\nu = \varrho_1 dx$ – вероятностная мера на $B_r = \{x: |x| \leq r\}$, эквивалентная ограничению меры Лебега. Тогда существуют борелевское отображение $T: A \rightarrow B_r$ и непрерывная функция $\varphi: A \rightarrow [0, r]$ с выпуклыми множествами $A_s = \{\varphi \leq s\}$ такие, что $\nu = \mu \circ T^{-1}$ и $T = \varphi \cdot n$ почти всюду относительно $\mathcal{H}^d$, где $n = n(x)$ – единичный внешний нормальный вектор к границе множества подуровня $\{y: \varphi(y) \leq \varphi(x)\}$.

Если функция φ является гладкой, то линии уровня φ движутся согласно следующему уравнению потока гауссовой кривизны:

$$\dot{x}(s) = -s^{d-1} \frac{\varrho_1(\text{sn})}{\varrho_0(x)} K(x) \cdot n(x),$$

$x(0) \in \partial A$ – произвольная начальная точка. Если множество A строго выпукло, то $\varphi|_{\partial A} = r$.

Доказательство основано на технике оптимальной транспортировки. В качестве следствия получаем транспортное доказательство существования гауссового потока для строго выпуклых поверхностей (см. детали в [167]). Отметим, что путем замены координат уравнение потока гауссовой кривизны сводится (хотя бы локально) к параболическому уравнению Монжа–Ампера

$$\partial_t u(t, x) \det D_x^2 u(t, x) = f(t, x), \quad u(0, x) = u_0, \quad (t, x) \in \mathbb{R}^+ \times \mathbb{R}^{d-1}.$$

Функция u предполагается выпуклой по x и возрастающей по t . Это уравнение переписывается в “транспортном” виде. В качестве “параболической” транспортировки выступает отображение $T: (t, x) \mapsto (\langle x, \nabla_x u \rangle - u, \nabla_x u)$. При этом $\det DT = \partial_t u \det D_x^2 u$. Развитие этой темы (формула замены переменных, связь с параболическим уравнением Монжа–Ампера, параболический принцип максимума) см. в [170]. О параболических оптимальных уравнениях см. также [171].

3.7.6. Геометрические потоки: потоки Риччи. Пусть M – гладкое n -мерное риманово многообразие с метрикой g . Эволюция римановой метрики, задаваемая уравнением $\partial g / \partial t = -2 \text{Ric}(g)$, называется потоком Риччи. Потоки Риччи стали основным инструментом в доказательстве Г. Я. Перельмана гипотезы Пуанкаре. В первом приближении поток Риччи является нелинейным аналогом полугруппы теплопроводности и действует на многообразие “усредняюще”. Тем не менее надежда на то, что многообразие под действием потока Риччи сожмется к пространству постоянной кривизны, не оправдывается. Оказывается, у многообразия могут возникнуть разрывы. Чтобы этого не происходило, приходится усложнять процедуру и контролировать разные

аналитические характеристики, например константы в неравенствах типа Соболева на многообразии. Делается это в том числе методом введения различных функционалов типа энтропии или энергетического функционала, которые оказываются монотонными под воздействием потока Риччи. Как мы видели, для двух градиентных потоков μ_t^1, μ_t^2 выпуклого функционала расстояние Канторовича $t \mapsto W_2(\mu_t^1, \mu_t^2)$ является невозрастающей функцией. В частности, монотонность имеет место для диффузии с генератором $\Delta - \nabla V \cdot \nabla$ и выпуклой функцией V . Оказывается, подходящее обобщение этого свойства на потоки Риччи влечет монотонность некоторых функционалов, доказанную Перельманом. Рассмотрим пример из [172]. Введем “обратное время” $\tau = b - t$. Рассмотрим нелинейное “сопряженное” уравнение теплопроводности

$$\frac{\partial u}{\partial \tau} = \Delta_{g(\tau)} u - \frac{1}{2} R u,$$

где $R = \text{Tr}(\text{Ric})$ – скалярная кривизна, $\Delta_{g(\tau)}$ – оператор Лапласа–Бельтрами в метрике $g(\tau)$, $u(x, 0)$ – неотрицательная функция с $\int_M u(x, 0) d\text{vol}_0 = 1$, где vol_t – риманов объем, соответствующий метрике $g(t)$. Продифференцировав соответствующие величины по t и используя уравнения выше, можно убедиться, что $\nu_t = u(\tau, x) \text{vol}_\tau$ – вероятностная мера при всех τ ; здесь используется формула для эволюции объема $d\text{vol}_\tau/d\tau = 2^{-1} \text{Tr}(dg(\tau)/d\tau) \text{vol}_{g(\tau)}$. Поток мер, построенный таким способом, будем называть “диффузией”; более того, постоянен интеграл

$$\int f(x, t) \nu_t(dx),$$

если f – решение уравнения теплопроводности $\partial_t f(x, t) = \Delta_{g(t)} f(x, t)$.

Рассмотрим теперь две диффузии: ν_τ^1, ν_τ^2 . В работе [172] (см. также [173]) было доказано, что функции $\tau \mapsto W_1^{g(\tau)}(\nu_\tau^1, \nu_\tau^2)$, $\tau \mapsto W_2^{g(\tau)}(\nu_\tau^1, \nu_\tau^2)$ являются убывающими (это дает обобщение свойства монотонности стандартного расстояния Канторовича относительно броуновской диффузии в плоском случае). Доказательство для W_1 основано на следующих наблюдениях.

1) Константа Липшица функции f , удовлетворяющей уравнению теплопроводности $\partial_t f = \Delta_{g(t)} f$, не возрастает вдоль t . В этом можно убедиться, применив формулу Бохнера (3.7.3) к метрике $g(t)$ и функции f и найдя $\partial|\nabla f|_{g(t)}^2/\partial t$. Учítывая, что для дифференцируемой матричной функции имеет место равенство $\frac{d}{dt} A(t)^{-1} = -A(t)^{-1} \frac{d}{dt} A(t) A(t)^{-1}$, получаем

$$\begin{aligned} \frac{\partial}{\partial t} |\nabla f|_g^2 &= \frac{\partial}{\partial t} (g^{ij} f_{x_i} f_{x_j}) = -|\nabla f|_{\partial_t g}^2 + 2\langle \nabla f, \nabla \Delta f \rangle \\ &= |\nabla f|_{\partial_\tau g}^2 + \Delta |\nabla f|^2 - 2|\text{Hess}(f)|^2 - 2\text{Ric}(\nabla f, \nabla f). \end{aligned}$$

Следовательно, (в метрике $g(t)$) $\partial|\nabla f|^2/\partial t = \Delta|\nabla f|^2 - 2|\text{Hess}(f)|^2 \leq \Delta|\nabla f|^2$. Искомое свойство следует из принципа максимума.

2) Для метрики W_1 верно соотношение двойственности

$$W_1^{g(\tau)}(\nu_\tau^1, \nu_\tau^2) = \max \left\{ \int \varphi d\nu_\tau^1 - \int \varphi d\nu_\tau^2 : \varphi \text{ есть 1-липшицева функция} \right\}.$$

Если теперь φ – решение двойственной задачи Монжа–Канторовича для метрики g_{τ_0} , то, положив $\varphi_\tau = -\Delta_{g(\tau)}\varphi$, мы получим семейство 1-липшицевых функций, для которых сохраняется величина

$$\int \varphi d\nu_\tau^1 - \int \varphi d\nu_\tau^2.$$

Отсюда немедленно вытекает монотонность W_1 . Более общий результат был получен в [174], [173]. Естественным объектом, связанным с потоками Риччи, является так называемая $\mathcal{L}$ -длина Перельмана

$$\mathcal{L}(\gamma) = \int_{\tau_1}^{\tau_2} \sqrt{\tau} (R(\gamma(\tau), \tau) + |\gamma'(\tau)|_{g(\tau)}^2) d\tau,$$

где γ – некоторая кривая на M (в каждый момент τ задействована своя риманова метрика); см. [175]. Указанный $\mathcal{L}$ -функционал определяет риманову “метрику” на $M \times [\tau_1, \tau_2]$ (последнее не вполне соответствует классическому понятию длины, ибо $\mathcal{L}$ может быть отрицательным). П. Топшинг рассмотрел $\mathcal{L}$ -оптимальную транспортировку вероятностных мер и соответствующее “расстояние” Канторовича. Оказывается, на пространстве мер существует функционал Θ , построенный как некоторое “ренормализованное” $\mathcal{L}$ -расстояние Канторовича, обладающий монотонностью вдоль потока диффузий. При этом из монотонности этого функционала можно извлечь (при подходящем предельном переходе) как результат Маккэна и Топшинга из [172], так и монотонность так называемой $\mathcal{W}$ -энтропии, введенной Перельманом в доказательстве гипотезы Пуанкаре.

3.7.7. Транспортные сети. Теория оптимальных транспортных сетей – интересное и в некотором смысле альтернативное ответвление теории оптимальной транспортировки. В отличие от знакомых нам регулярных решений задачи Монжа–Канторовича, решения задач теории транспортных сетей устроены совсем иначе. Визуально эти решения можно представить как дискретные наборы ветвящихся одномерных объектов, подобных деревьям или кровеносным сосудам. Стоимость транспортировки довольно легко определяется на дискретных объектах. Например, для дискретного множества вершин (терминалов) заданы потоки между φ терминалами. Общая стоимость перевозки равна $\sum_i \varphi_i^\alpha \lambda_i$ для некоторой константы $0 < \alpha < 1$ (λ_i – длина

ребра). Для постановки задачи в непрерывном случае напомним, что задача Монжа–Канторовича допускает формулировку на языке динамических оптимальных планов, т.е. мер на пространстве кривых. Аналогично оптимальная транспортная сеть определяется как мера P на пространстве Ω липшицевых кривых $\gamma: [0, +\infty) \rightarrow \mathbb{R}^d$ с заданными проекциями $P \circ \pi_0^{-1}$ (орошающая мера) и $P \circ \pi_\infty^{-1}$ (орошаемая мера), минимизирующая функционал Джилльберта

$$\mathcal{E}^\alpha(P) = \int_\Omega \int_{\mathbb{R}_+} |\gamma(\omega, t)|_P^{\alpha-1} |\dot{\gamma}(t)| dt dP(\gamma).$$

Здесь $0 \leq \alpha \leq 1$, а величина $|\gamma(\omega, t)|_P$ называется кратностью и определяется как $P(\{\gamma: \text{есть } t, \text{ для которого } \gamma(t) = x\})$. При $\alpha = 1$ мы получаем классическую задачу Монжа. Подробнее об этом см. [176], [177].

3.8. Бесконечномерные задачи Монжа–Канторовича

3.8.1. Существование оптимальных отображений на пространстве Винера. В этом разделе речь пойдет об анализе на бесконечномерном локально выпуклом пространстве X , наделенном радоновской гауссовской мерой γ .

ОПРЕДЕЛЕНИЕ 3.8.1. Радоновская вероятностная мера γ на локально выпуклом пространстве X называется гауссовской, если ее одномерный образ $\gamma \circ f^{-1}$ является гауссовской мерой для всех $f \in X^*$. Если все $\gamma \circ f^{-1}$ симметричны, то γ называется симметричной или центрированной.

В дальнейшем для простоты изложения мы будем считать (если специально не оговорено иное), что $X = \mathbb{R}^\infty$ (счетная степень прямой), а под бесконечномерной гауссовской мерой будет подразумеваться счетное произведение $\gamma = \bigotimes_{i=1}^{\infty} \gamma_i$ стандартных гауссовских мер на прямой. Согласно известной теореме Цирельсона, всякая центрированная радоновская гауссовская мера, не сосредоточенная на конечномерном подпространстве, изоморфна γ посредством линейного измеримого отображения, поэтому обсуждаемые ниже результаты верны для всего этого класса мер. Общая теория гауссовских мер подробно изложена в книге [104]. Ниже мы кратко опишем особенности анализа на пространстве с гауссовской мерой. В отличие от конечномерного случая мера γ_h , получаемая из меры γ сдвигом на некоторый вектор $h \in H$, т. е. $\gamma_h(A) = \gamma(A - h)$, не обязана быть абсолютно непрерывной относительно γ . Рассмотрим этот вопрос для счетного произведения $\gamma = \bigotimes_{i=1}^{\infty} \gamma_i$. Пусть $h = (h_i) \in \mathbb{R}^\infty$. Формальное вычисление плотности Радона–Никодима дает выражение

$$\frac{d\gamma_h}{d\gamma} = \mathbb{E} \left(\sum_{i=1}^{\infty} h_i x_i - \frac{1}{2} \sum_{i=1}^{\infty} h_i^2 \right). \quad (3.8.1)$$

Оно имеет смысл, если $\sum_{i=1}^{\infty} h_i^2 < \infty$, т. е. $h \in l^2$. В самом деле, координатные функции x_i – независимые стандартные гауссовские случайные величины на пространстве $\mathbb{R}^\infty$ с мерой γ . Так как $\|S_{m+n} - S_m\|_{L^2(\gamma)}^2 = \sum_{i=m+1}^{m+n} h_i^2$, где $S_n = \sum_{i=1}^n \langle x, h_i \rangle$, то ряд $\sum_{i=1}^{\infty} h_i x_i$ сходится в $L^2(\gamma)$ при $h \in l^2$. Более того, имеет место сходимост γ -п.в., так как $\{S_n\}$ – мартингал с $\sup_n \mathbb{E} S_n^2 < \infty$. Из этого следует эквивалентность мер γ_h и γ при $h \in l^2$. Если же $h \notin l^2$, то меры γ и γ_h взаимно сингулярны, ибо найдется элемент $k \in l^2$, для которого $\sum_{n=1}^{\infty} k_n h_n = +\infty$, тогда множество K последовательностей (x_n) , для которых ряд $\sum k_n x_n$ сходится, имеет γ -меру 1, но $(K + h) \cap K = \emptyset$.

ОПРЕДЕЛЕНИЕ 3.8.2. Пространство $H = \{h \in X : \gamma_h \sim \gamma\}$ называется пространством Камерона–Мартина меры γ .

Из сказанного явствует, что $\gamma_h \sim \gamma$ (т.е. меры эквивалентны – каждая абсолютно непрерывна относительно другой) для всех $h \in H$, а для прочих h меры γ и γ_h взаимно сингулярны. Пространство H линейно и имеет меру нуль относительно γ , если бесконечномерно.

Напомним, что винеровский процесс $\{w_t\}_{t \in [0,1]}$, $w_0 = 0$, задает меру на пространстве траекторий на $[0, 1]$. Ее можно специальным образом сузить на пространство непрерывных траекторий, что дает меру Винера P^W на пространстве $C[0, 1]$ непрерывных функций.

ПРИМЕР 3.8.3. Для счетного произведения стандартных гауссовских мер γ имеем $H = l^2$. Для P^W пространство Камерона–Мартина есть класс Соболева

$$W_0^{2,1} = \{h: h \text{ абсолютно непрерывна, } h(0) = 0, h' \in L^2[0, 1]\}.$$

Всякий вектор h из пространства Камерона–Мартина задает так называемый измеримый линейный функционал $\hat{h}$ на X . Для $\gamma = \prod_{i=1}^{\infty} \gamma_i$ это $\hat{h}(x) = \sum_{i=1}^{\infty} h_i x_i$. В случае меры P^W это стохастический интеграл

$$w \mapsto \int_0^1 h'(t) dw_t$$

от детерминированной функции h' . Таким образом, вектору h соответствует случайная величина (измеримый линейный функционал) $\hat{h}$ на пространстве с гауссовской мерой. На H вводится естественная норма с помощью равенства

$$|h|_H^2 = \|\hat{h}\|_{L^2(\gamma)}^2 = \int_X \hat{h}^2 d\gamma,$$

с которой H превращается в сепарабельное гильбертово пространство.

Анализ на пространстве X с мерой γ тесно связан с пространством Камерона–Мартина. В некотором смысле естественным образом дифференцируемые функции дифференцируемы только вдоль пространства Камерона–Мартина. Понимая дифференцирование таким образом, можно построить аналог соболевских пространств на пространстве Винера. Для определения соболевских пространств введем некоторые обозначения. По-прежнему удобно считать, что $X = \mathbb{R}^\infty$, $x = (x_i)$. Пусть $L^2(\gamma, H)$ обозначает гильбертово пространство H -значных γ -измеримых отображений v с конечной нормой

$$\|v\|_{L^2(\gamma, H)} = \left(\int_X |v(x)|_H^2 \gamma(dx) \right)^{1/2}.$$

Норма Гильберта–Шмидта симметричного оператора A на H определяется по формуле $\|A\|_{\mathcal{H}} = \left(\sum_{i=1}^{\infty} (Ae_i, Ae_i)_H \right)^{1/2}$, где $\{e_i\}$ – ортогональный базис в H .

Сигма-алгебра, порожденная координатными функциями $x_1, \dots, x_n$, обозначается через $\mathcal{F}_n$. Пространство гладких цилиндрических функций $\mathcal{FC}_b^\infty$ состоит из всех функций вида $\zeta(x_1, \dots, x_n)$, где $\zeta \in C_b^\infty(\mathbb{R}^n)$ для некоторого n . Соболевский класс $W^{2,1}(\gamma)$ состоит из всех функций $f \in L^2(\gamma)$, имеющих обобщенный

градиент $\nabla_H f \in L^2(\gamma, H)$ вдоль H , определяемый равенством

$$\int_X \partial_h \varphi f d\gamma = - \int_X \varphi \langle \nabla_H f, h \rangle_H d\gamma + \int_X \varphi f \widehat{h} d\gamma$$

для всех $h \in H$ и $\varphi \in \mathcal{FC}_b^\infty$, где $\partial_h \varphi(x) = \lim_{t \rightarrow 0} (\varphi(x + th) - \varphi(x))/t$. Соболевский класс $W^{2,2}(\gamma)$ состоит из дважды обобщенно H -дифференцируемых функций с конечной нормой

$$\|f\|_{W^{2,2}(\gamma)} = \left(\int_X f^2 d\gamma + \int_X |\nabla_H f|_H^2 d\gamma + \int_X \|D_H^2 f\|_{\mathcal{H}}^2 d\gamma \right)^{1/2}.$$

Матричные элементы $\partial_{e_i} \partial_{e_j} f$ оператора $D_H^2 f$ определяются через формулу интегрирования по частям.

Не зависящие от размерности конечномерные результаты (логарифмическое неравенство Соболева и т. д.) автоматически оказываются верными на бесконечномерном пространстве, если вместо обычной евклидовой нормы понимать норму на пространстве Камерона–Мартина. Например, логарифмическое неравенство Соболева принимает вид

$$\int f^2 \ln \frac{f^2}{\|f\|_{L^2(\gamma)}^2} d\gamma \leq 2 \int |\nabla_H f|_H^2 d\gamma.$$

Плотность Радона–Никодима $d\gamma_h/d\gamma$ в случае абстрактной гауссовской меры имеет вид

$$\frac{d\gamma_h}{d\gamma} = \mathbb{E} \left(\widehat{h}(x) - \frac{1}{2} |h|_H^2 \right).$$

В случае $\gamma = \bigotimes_{i=1}^{\infty} \gamma_i$ это есть формула (3.8.1). В случае P^W приходим к классической формуле Камерона–Мартина

$$\frac{dP_h^W}{dP^W} = \mathbb{E} \left(\int_0^1 h'(t) dw_t - \frac{1}{2} \int_0^1 |h'(t)|^2 dt \right).$$

Для гауссовских мер полезно рассматривать нелинейные сдвиги вдоль пространства Камерона–Мартина вида

$$T(x) = x + F(x), \quad (3.8.2)$$

где F – отображение, принимающее значения в H . Для отображений такого вида при довольно широких предположениях удается доказать, что мера $\gamma \circ T^{-1}$ абсолютно непрерывна относительно γ , и явно написать формулу замены переменных. Например, такие отображение могут задаваться решениями стохастических дифференциальных уравнений, а соответствующая “формула замены переменных” носит название формулы Гирсанова. Довольно долго был открыт вопрос (обсуждавшийся в частности в [178]): какие меры вида $\nu = g \cdot \gamma$ являются образами γ при отображениях вида (3.8.2)? Ответ был получен в работе [179].

ТЕОРЕМА 3.8.4. *Всякая вероятностная мера вида $\nu = g \cdot \gamma$ представляется в виде $\nu = \gamma \circ T^{-1}$, где $T(x) = x + F(x)$, $F: X \rightarrow H$.*

Отображение T было построено с помощью треугольных отображений, причем существенно использовался аналог транспортного неравенства для треугольных отображений; такие отображения обсуждаются в последнем разделе. Однако открыт вопрос, всегда ли можно само T вида (3.8.2) выбрать треугольным. Как мы увидим, нелинейные сдвиги вдоль пространства Камерона–Мартина возникают и в задаче Монжа–Канторовича для $|\cdot|_H^2$, изученную впервые Д. Фейелем и А. С. Устюнелем [180], [181]. В рассматриваемой ситуации ставится задача минимизации функционала

$$K_H(\mu, \nu, m) = \int_{X \times X} |x - y|_H^2 dm, \quad m \in \Pi(\mu, \nu).$$

Пусть

$$W_2(\mu, \nu) = \sqrt{K_H(\mu, \nu)}.$$

Положив $K_H(\mu, \nu) = \inf\{K_H(\mu, \nu, m) : m \in \Pi(\mu, \nu)\}$ и введя конечномерные непрерывные функции $h_n(x, y) = \sum_{i=1}^n |x_i - y_i|^2$, замечаем, что

$$K_H(\mu, \nu) = \sup_n K_{H,n}(\mu_n, \nu_n), \quad (3.8.3)$$

где $K_{H,n}$ соответствует конечномерной функции стоимости h_n и μ_n, ν_n – проекции μ и ν на $\mathbb{R}^n$. Это следует из теоремы о монотонной сходимости с учетом того, что $h_n \uparrow h$ и интеграл от функции h_n по мере σ равен интегралу по проекции σ на $\mathbb{R}^n \times \mathbb{R}^n$.

Существенным новшеством в данной ситуации оказывается то, что здесь гораздо более узок класс мер m , на которых $K_H(\mu, \nu, m)$ принимает конечные значения. Отчасти это связано с тем, что $h(x, y) = +\infty$ почти всюду относительно $\gamma \otimes \gamma$. С таким явлением мы не сталкивались в предыдущих разделах. Конечно, если иметь дело с мерами, сосредоточенными на H , то все эти неприятности исчезнут, но вместе с ними уйдет и вся содержательная часть, связанная с привязкой к мере γ . Подчеркнем, что это касается обеих мер, а не только μ . Например, если $\mu = \gamma$, а мера ν сосредоточена на H (даже является дираковской мерой в нуле), то $K_H(\mu, \nu, m) = +\infty$ при всех $m \in \Pi(\mu, \nu)$, так как в противном случае, записав m в виде $m(dx dy) = m^y(dx) \nu(dy)$, для ν -почти всех $y \in H$ получаем, что мера m^y тоже должна быть сосредоточена на H (если $|x - y|_H < \infty$, то $x \in H$), поэтому $\gamma(H) = 1$, что неверно. Если величина $K_H(\mu, \nu)$ конечна, то общие соображения о существовании соответствующего решения задачи Канторовича остаются в силе из-за сохраняющейся здесь полунепрерывности снизу функционала $m \mapsto K_H(\mu, \nu, m)$.

Рассмотрим другой пример: ν есть образ γ при растяжении $x \mapsto 2x$. Тогда $K_{H,n}(\gamma_n, \nu_n) = n$, ибо отображение $Tx = 2x$, переводящее γ_n в ν_n , является градиентом, следовательно, по теореме единственности оно и служит оптимальной транспортировкой меры γ_n в ν_n для функции стоимости h_n , а тогда $K_{H,n}(\gamma_n, \nu_n) = M(\gamma_n, \nu_n, T)$ есть интеграл от $|x - 2x|_H^2 = |x|_H^2$ по мере γ_n , равный n .

Таким образом, на меру ν приходится накладывать какие-то дополнительные условия. Распространенная (но не единственно возможная) ситуация возникает, когда обе меры μ и ν абсолютно непрерывны относительно γ (но и это еще не гарантирует, что $K_H(\mu, \nu) < \infty$). Эффективно проверяемое достаточное (но отнюдь не необходимое) условие состоит в следующем.

ПРЕДЛОЖЕНИЕ 3.8.5. Пусть $\mu = \varrho_\mu \cdot \gamma$, $\nu = \varrho_\nu \cdot \gamma$ – такие вероятностные меры, что $\text{Ent}_\gamma \mu < \infty$, $\text{Ent}_\gamma \nu < \infty$. Тогда

$$K_H(\mu, \nu) = W_2^2(\mu, \nu) \leq 4 \text{Ent}_\gamma \mu + 4 \text{Ent}_\gamma \nu. \quad (3.8.4)$$

ДОКАЗАТЕЛЬСТВО. Это вытекает из (3.8.3) и конечномерного неравенства Талаграна (3.3.2) с учетом того факта, что $\text{Ent}_\gamma \mu = \sup_n \text{Ent}_{\gamma_n} \mu_n$ и аналогично для ν . Последний факт следует из того, что плотности Радона–Никодима $\varrho_n = d\mu_n/d\gamma_n$, рассматриваемые на пространстве $\mathbb{R}^\infty$ с мерой γ , образуют мартингал относительно σ -алгебр, порожденных проекциями на $\mathbb{R}^n$, сходящийся γ -п.в. к $\varrho = d\mu/d\gamma$.

Отметим, что в силу неравенства Йенсена для условных математических ожиданий величины $\text{Ent}_{\gamma_n} \mu_n$ возрастают и не превосходят $\text{Ent}_\gamma \mu$.

Без неравенства Талаграна неочевидно даже, что $K_H(\gamma, \nu) < \infty$ для меры $\nu = \gamma/\gamma(A)$, где A – множество с $\gamma(A) > 0$. Таким образом, здесь редко удастся найти явно хотя бы какую-то меру $\sigma \in \Pi(\mu, \nu)$ с $K(\mu, \nu, \sigma) < \infty$.

Разумеется, кроме аналога задачи Канторовича здесь появляется и аналог задачи Монжа минимизации функционала

$$\int_X |x - T(x)|_H^2 \mu(dx)$$

по отображениям T , переводящим меру μ в меру ν . Здесь T заведомо должно иметь вид (3.8.2). Рассмотрим пару вероятностных мер $\mu = \varrho_\mu \cdot \gamma$, $\nu = \varrho_\nu \cdot \gamma$. Нас интересует, можно ли отобразить μ в ν отображением вида (3.8.2), минимизирующим указанный функционал. Оказывается, что это возможно, причем при широких условиях F является в некотором смысле градиентом вдоль H . Отметим, что даже для тождественного отображения $T(x) = x$ равенство $x = \nabla_H |x|_H^2/2$ не имеет смысла в бесконечномерном пространстве, ибо пространство Камерона–Мартина имеет меру нуль. Поэтому здесь не приходится рассчитывать на то, что само T будет градиентом вдоль H . Возникает ситуация, схожая с ситуацией на многообразиях. Решение надо искать в виде

$$T(x) = x + \nabla_H f(x),$$

где f – так называемая 1-выпуклая функция (см. [182]). Функция f называется 1-выпуклой, если отображение $h \mapsto F_h(x) := f(x + h) + |h|^2/2$ из H в $L^0(\gamma)$ выпукло, т. е. при фиксированных $h, k \in H$ и $\alpha \in [0, 1]$ имеем

$$F_x(\alpha h + (1 - \alpha)k) \leq \alpha F_x(h) + (1 - \alpha)F_x(k) \quad \text{для } \gamma\text{-почти всех } x,$$

где соответствующее множество меры нуль может зависеть от h, k, α . Можно показать, что для выбранного ортонормированного базиса $\{e_i\}$ в H при

фиксированном i есть такая версия f , что функции $t \mapsto f(x + te_i) + t^2/2$ выпуклы. Поэтому почти всюду существуют частные производные $\partial_{e_i} f$. Определим $\nabla_H f(x)$ как $\nabla_H f = \sum_{i=1}^{\infty} \partial_{e_i} f \cdot e_i$, если этот ряд сходится в H . Конечномерные 1-выпуклые функции – с точностью до модификации функции вида $\varphi(x) - |x|^2/2$, где φ выпукла. Следует отметить, что в таких вопросах частные производные определяют в смысле Соболева (см. [104]).

Приведем теперь результат Д. Фейеля и А. С. Устюнеля [180].

ТЕОРЕМА 3.8.6. Пусть $\mu = \varrho_\mu \cdot \gamma$, $\nu = \varrho_\nu \cdot \gamma$ – вероятностные меры, причем $\text{Ent}_\gamma \mu < \infty$, $\text{Ent}_\gamma \nu < \infty$. Предположим дополнительно, что μ удовлетворяет неравенству Пуанкаре (например, $\mu = \gamma$), т. е. существует такое $C > 0$, что

$$\int \left(f - \int f d\mu \right)^2 d\mu \leq C^{-1} \int |\nabla_H f|^2 d\mu \quad \text{для всех } f \in \mathcal{FC}_b^\infty.$$

Тогда существует такое отображение $T(x) = x + F(x)$, что $\nu = \mu \circ T^{-1}$, $F = \nabla_H f \in H$, где f – 1-выпуклая функция. Отображение T единственно с точностью до μ -эквивалентности, причем оно дает единственное решение задачи Монжа и порождает единственное решение задачи Канторовича.

ДОКАЗАТЕЛЬСТВО. Как указано выше, описанная задача Канторовича для функционала K_H и пары μ, ν имеет решение. Кроме того, $W_2(\mu, \nu) < \infty$ в силу (3.8.4). Далее считаем, что $X = \mathbb{R}^\infty$, γ – счетная степень стандартной гауссовской меры. Тогда $H = l^2$, поэтому в обозначении нормы и градиента опускаем индекс H . Обозначим через π и π_n некоторые решения задачи Канторовича для пар μ, ν и μ_n, ν_n соответственно. В силу того, что проекция меры π на $\mathbb{R}^n \times \mathbb{R}^n$ обладает маргиналами μ_n и ν_n , имеем очевидное неравенство

$$W_2(\mu_n, \nu_n) \leq \int \sum_{i=1}^n |x_i - y_i|^2 \pi(dx dy) \leq W_2(\mu, \nu).$$

Аналогично $W_2(\mu_m, \nu_m) \leq W_2(\mu_n, \nu_n)$ при $m \leq n$. Отсюда и из неравенств выше следует, что $\lim_{n \rightarrow \infty} W_2(\mu_n, \nu_n) = W_2(\mu, \nu)$. Пусть $\Phi_n(x)$, $\Psi_n(y)$ – решение двойственной задачи Канторовича для μ_n, ν_n . Положим $\varphi_n(x) = \Phi_n(x) - |x|^2/2$, $\psi_n(y) = \Psi_n(y) - |y|^2/2$. Заметим, что

$$F_n(x, y) = \varphi_n(x) + \psi_n(y) + \frac{1}{2}|x - y|^2 \geq 0,$$

причем точное равенство достигается π_n -п.в. Ясно, что

$$\int F_n(x, y) \pi(dx dy) = W_2(\mu, \nu) - W_2(\mu_n, \nu_n) \rightarrow 0.$$

Следовательно, $F_n \rightarrow 0$ в $L^1(\pi)$. Поэтому последовательность $\{F_n\}$ равномерно интегрируема. Так как μ удовлетворяет неравенству Пуанкаре, то этому же неравенству удовлетворяют ее проекции μ_n (с той же константой). В силу неравенства Пуанкаре имеем

$$W_2(\mu, \gamma) \geq W_2(\mu_n, \gamma_n) = \frac{1}{2} \int |\nabla \varphi_n|^2 d\mu_n \geq \frac{C}{2} \int \left(\varphi_n - \int \varphi_n d\mu_n \right)^2 d\mu_n.$$

Вычитанием постоянных переходим к случаю $\int \varphi_n d\mu_n = 0$. Полученная оценка дает равномерную интегрируемость последовательности $\{\varphi_n\}$ относительно меры μ . Значит, последовательность $\{\psi_n\}$ равномерно интегрируема относительно меры ν . Таким образом, можно считать, что последовательности $\{\varphi_n\}$ и $\{\psi_n\}$ равномерно интегрируемы относительно меры π . В силу известного результата для некоторой подпоследовательности индексов $\{n_k\}$ последовательности выпуклых комбинаций $\{\varphi'_n\}, \{\psi'_n\}$ функций φ_n и ψ_n соответственно сходятся π -почти всюду и в $L^1(\pi)$ к некоторым функциям φ, ψ . Положим

$$\varphi = \limsup \varphi'_n, \quad \psi = \limsup \psi'_n.$$

Имеем $\varphi(x) + \psi(y) + |x - y|^2/2 \geq 0$, причем π -п.в. выполнено равенство. Зафиксируем $h \in H$. Из соотношения $|x + h - y|^2 = |x - y|^2 + 2\langle h, x - y \rangle + |h|^2$ следует, что π -п.в. $\varphi(x + h) - \varphi(x) \geq -\langle h, x - y \rangle - |h|^2/2$. Последнее означает, что $y = x + \nabla\varphi(x)$ для μ -почти всех x и $T(x) = x + \nabla\varphi(x)$ – искомое отображение. Заметим теперь, что любой $L_1(\pi)$ -предел $\tilde{\varphi}$ выпуклых комбинаций φ_n будет π -п.в. удовлетворять равенству $\nabla\tilde{\varphi}(x) = y - x$. Отсюда следует, что $\nabla\varphi(x) = \nabla\tilde{\varphi}(x)$ для μ -почти всех x . Если теперь $\tilde{\gamma}$ – другое решение задачи Канторовича, то оно также сосредоточено на графике $\nabla\varphi$, где φ – любая предельная точка конечномерных приближений. Отсюда следует, что получено единственное решение задач как Канторовича, так и Монжа.

ЗАМЕЧАНИЕ 3.8.7. Нетрудно доказать, что отображение $T(x) = x + \nabla\varphi(x)$ обратимо и $T^{-1}(y) = S(y)$ для ν -почти всех y и $S(y) = y + \nabla\psi(x)$ также является оптимальным отображением.

В работе [180] есть формулировка (с не вполне нам ясным доказательством) более общего результата о существовании оптимального отображения между парой мерой с конечным расстоянием Канторовича. В [183] было предложено альтернативное доказательство существования транспортировки меры $f \cdot \gamma$ в меру γ , основанное на использовании неравенства (3.3.1). Применяя это неравенство к случаю $m = \gamma$, $\mu = f \cdot \gamma$, $\nu = g \cdot \gamma$, мы получаем неравенство

$$\text{Ent}_\mu \nu = \int g \ln \frac{f}{g} d\gamma \geq \frac{1}{2} \int |\nabla\Phi_f - \nabla\Phi_g|^2 f d\gamma. \quad (3.8.5)$$

Из (3.8.5) несложно вывести существование оптимального отображения T меры $f \cdot \gamma$ с $\text{Ent}_\gamma f < \infty$ в меру γ . Действительно, конечномерные оптимальные транспортировки T_n , отображающие $f_n \cdot \gamma$ в γ , где $f_n = E_\gamma^{\mathcal{F}_n} f$, удовлетворяют неравенству

$$\int f_n \ln \frac{f_m}{f_n} d\gamma \geq \frac{1}{2} \int |\nabla\Phi_{f_m} - \nabla\Phi_{f_n}|^2 f_m d\gamma.$$

Используя, что

$$\lim_{n,m \rightarrow \infty} \int f_n \ln \frac{f_m}{f_n} d\gamma = 0,$$

нетрудно показать, что последовательность $\{\nabla\Phi_{f_n}\}$ сходится $f \cdot \gamma$ -п.в. Отсюда получаем существование искомого отображения. Более того, доказательство можно распространить на более широкие классы мер, в частности, построить

примеры оптимальных транспортировок взаимно сингулярных бесконечномерных мер (см. теорему 3.8.8).

В недавней работе [184] установлено существование оптимального отображения между всякими двумя вероятностными мерами, абсолютно непрерывными относительно γ .

Несомненно, вопрос о том, какие пары мер (вообще говоря, взаимно сингулярных) обладают оптимальной транспортировкой, относится к центральным в бесконечномерной теории Монжа–Канторовича. Более слабый вариант решения этой задачи дан в следующей теореме, которую мы приведем без доказательства (в его основе лежит оценка (3.8.5)); в ней T можно получить как предел (почти всюду) конечномерных оптимальных отображений.

ТЕОРЕМА 3.8.8. Пусть μ – вероятностная мера на $\mathbb{R}^\infty$. Предположим, что конечномерные проекции $\mu_n = \mu \circ P_n^{-1} = g_n dx$ имеют то свойство, что $\lim_{m \rightarrow \infty} \sup_{n > m} e(n, m) = 0$, где $e(n, m) = \inf_{\tau \in M_{(n, m)}} \text{Ent}_{\mu_m \times \tau}(d\mu_n/d(\mu_m \times \tau))$, $M_{(n, m)}$ – пространство вероятностных мер на подпространстве $L_{m, n}$, порожденном $\{e_{m+1}, \dots, e_n\}$. Тогда существует отображение T из μ в γ , почти всюду равное пределу конечномерных оптимальных отображений.

Условие теоремы выполнено для меры μ , имеющей конечную энтропию относительно некоторой продукт-меры.

3.8.2. Формула замены переменной. При некоторых ограничениях на меры для оптимальных отображений оказывается возможным доказать бесконечномерную формулу замены переменных. Пусть γ – стандартная гауссовская мера на $\mathbb{R}^\infty$, как и выше. Пусть $T(x) = x + \nabla\varphi(x)$ – оптимальная транспортировка меры $g \cdot \gamma$ в γ , $S(x) = T^{-1}(x) = x + \nabla\psi(x)$ – обратное отображение. Формально выражение для замены переменных принимает вид

$$g = \det_2(I + D^2\varphi) \mathbb{E} \left(\mathcal{L}\varphi - \frac{1}{2} |\nabla\varphi|^2 \right),$$

где D^2 – вторая производная,

$$\mathcal{L}\varphi(x) = \Delta\varphi(x) - \langle x, \nabla\varphi(x) \rangle = \text{div}_\gamma(\nabla\varphi)(x)$$

есть оператор Орнштейна–Уленбека, а $\det_2$ – определитель Фредгольма–Карлемана, определяемый формулой $\det_2(I + K) = \prod_{i=1}^{\infty} (1 + k_i) e^{-k_i}$, где K – симметричный оператор Гильберта–Шмидта с собственными значениями k_i . В классе $W^{p,2}(\gamma)$ оператор $\mathcal{L}$ можно ввести как генератор полугруппы Орнштейна–Уленбека в $L^p(\gamma)$ или задать $\mathcal{L}f$ как предел в $L^p(\gamma)$ функций $\mathcal{L}f_n$ для функций $f_n \in \mathcal{FC}_b^\infty$, сходящихся к f в $W^{p,2}(\gamma)$.

Для обратного отображения $T^{-1}(x) = x + \nabla\psi(x)$ формула замены переменных принимает вид

$$g(x + \nabla\psi(x)) \det_2(I + D^2\psi(x)) \mathbb{E} \left(\mathcal{L}\psi - \frac{1}{2} |\nabla\psi|^2 \right) = 1. \quad (3.8.6)$$

В общем случае вопрос о том, для каких отображений выполнена формула замены переменной, представляет собой нетривиальную проблему (см. [104],

[67], [178]). В нашем случае можно воспользоваться специальным видом T . Соотношение (3.8.6) получено в [180] в предположении равномерной выпуклости $g \cdot \gamma$, т. е. $-D^2 \ln g + I \geq \varepsilon \text{Id}$, $\varepsilon > 0$. Как показано в [185], в предположении $\text{Ent}_\gamma g < \infty$ и $g > c > 0$ выполнено соотношение

$$g = \det_2(I + D_a^2 \varphi) \mathbb{E} \left(\mathcal{L}_a \varphi - \frac{1}{2} |\nabla \varphi|^2 \right),$$

где $D_a^2 \varphi$ и $\mathcal{L}_a \varphi$ – абсолютно непрерывные части $D^2 \varphi$ и $\mathcal{L} \varphi$ соответственно. Аналогично в предположении $\text{Ent}_{g \cdot \gamma} \frac{1}{g} < \infty$, $g < c$ выполнено соотношение

$$g(x + \nabla \psi(x)) \det_2(I + D_a^2 \psi(x)) \mathbb{E} \left(\mathcal{L}_a \psi - \frac{|\nabla \psi|^2}{2} \right) = 1.$$

Следующая теорема доказана в [84].

ТЕОРЕМА 3.8.9. *Предположим, что $I_\gamma g := \int \frac{|\nabla g|^2}{g} d\gamma < \infty$. Тогда $D^2 \varphi$ входит в $\mathcal{HS}$ $g \cdot \gamma$ -н.в., $\int \|D^2 \varphi\|_{\mathcal{HS}}^2 g d\gamma < \infty$, $\mathcal{L} \varphi \in L^1(g \cdot \gamma)$ и $g \cdot \gamma$ -н.в. выполнена формула замены переменных $g = \det(I + D^2 \varphi) \mathbb{E}(\mathcal{L} \varphi - |\nabla \varphi|^2/2)$.*

В этой теореме $\mathcal{L} \varphi$ понимается в смысле интегрирования по частям, но при дополнительном условии, что $1/g \in L^r(\gamma)$ с некоторым $r > 1$, имеет место включение $\varphi \in W^{2r/(1+r), 2}(\gamma)$, так что $\mathcal{L} \varphi$ существует в смысле $W^{2r/(1+r), 2}(\gamma)$. В [84] получены и некоторые другие результаты о регулярности φ (см. также п. 2.2.5). Естественная проблема в этой области – регулярность оптимального отображения (частично решенная в теореме 3.8.9). Пусть γ – гауссовская мера на бесконечномерном пространстве, $T(x) = x + \nabla \varphi(x)$ – оптимальная транспортировка вероятностной меры $f \cdot \gamma$ в вероятностную меру $g \cdot \gamma$. Интересно найти в общем случае соболевские априорные оценки для φ .

Связанная с винеровским пространством метрика Канторовича изучается также в [186]. В работе [187] рассмотрена оптимальная транспортировка на пространстве конфигураций.

3.9. Некоторые другие виды преобразований мер

3.9.1. Треугольные отображения. Пусть даны пространства X_k в конечном или счетном числе и $X = \prod_k X_k$. Отображение $T = (T_1, T_2, \dots): X \rightarrow X$ называется *треугольным*, если $T_k(x) = T_k(x_1, \dots, x_k)$ при всех k . Если же $X_k = \mathbb{R}^1$ при всех k , то треугольное отображение называется *возрастающим*, когда все функции $x_k \mapsto T_k(x_1, \dots, x_k)$ являются возрастающими. Аналогично определяются возрастающие треугольные отображения, заданные на подмножествах $\mathbb{R}^\infty$. Отметим, что монотонности по другим переменным не требуется. Треугольные отображения появились в 50-х годах (М. Rosenblatt, Н. Knothe, см. ссылки в [17]). Позже М. Громов применил треугольные отображения для вывода классического изопериметрического неравенства (см. п. 3.1). Основные

свойства треугольных отображений изложены в [29; §10.10(vii)], [67; §10.7]. Слово “треугольное” в названии объясняется тем, что производная дифференцируемого треугольного отображения в $\mathbb{R}^n$ задается треугольной матрицей. Несмотря на довольно специальный вид, треугольные отображения обладают богатыми возможностями в плане преобразования мер. В одномерном случае меру Лебега λ на $(0, 1)$ можно перевести в произвольную вероятностную меру ν на $(0, 1)$ возрастающей функцией T , используя функцию распределения F_ν меры ν (если F_ν строго возрастает, то $T = F_\nu^{-1}$). Из этого следует, что всякую безатомическую вероятностную меру μ на прямой можно перевести в любую вероятностную меру на прямой (если μ задана положительной плотностью, то преобразование можно сделать неубывающим). Покажем, как построить треугольное отображение на плоскости, преобразующее абсолютно непрерывную вероятностную меру μ в вероятностную меру ν . Обозначим через μ_1 и ν_1 проекции мер μ и ν на первую координатную прямую и возьмем одномерное отображение T_1 , переводящее μ_1 в ν_1 . Через μ^{x_1} и ν^{x_1} обозначим условные вероятностные меры на второй координатной прямой, $x_1 \in \mathbb{R}^1$. При μ_1 -почти всяком x_1 мера μ^{x_1} имеет плотность. Поэтому ее можно перевести в меру $\nu^{T_1(x_1)}$ функцией $x_2 \mapsto T_2(x_1, x_2)$, которая может быть сделана возрастающей в случае положительной плотности. Ясно, что $T := (T_1, T_2)$ оказывается треугольным отображением, причем оно возрастающее, если плотность μ положительна. Непосредственно проверяется, что оно переводит μ в ν . Построение продолжается индуктивно с использованием одномерных условных мер на последней координатной прямой. В результате для любой пары абсолютно непрерывных вероятностных мер на $\mathbb{R}^n$ можно построить треугольное отображение T , преобразующее одну меру в другую. Совершенно аналогично доказывается более общий факт.

ТЕОРЕМА 3.9.1. Пусть $X = \prod_{n=1}^{\infty} X_n$, где X_n – суслинские пространства. Пусть μ – такая борелевская вероятностная мера на X , что ее проекции на $\prod_{j=1}^n X_j$ и условные меры на X_n не имеют атомов для всех n . Тогда для каждой борелевской вероятностной меры ν на X найдется такое треугольное борелевское отображение $T: X \rightarrow X$, что выполнено равенство $\mu \circ T^{-1} = \nu$.

Если $X_n = \mathbb{R}^1$ при всех n и проекции μ на все $\mathbb{R}^n$ имеют положительные плотности, то это треугольное отображение можно сделать возрастающим. Последним свойством оно определяется однозначно с точностью до μ -эквивалентности и называется каноническим треугольным отображением.

Треугольные отображения почти никогда не являются оптимальными (кроме одномерного случая); это видно уже из того, что матрица Якоби у треугольного отображения треугольна, а у оптимального симметрична, что одновременно возможно лишь для диагональных матриц. Тем не менее многие свойства возрастающих треугольных отображений близки к свойствам оптимальных транспортировок. Например, в [188] доказано следующее транспортное неравенство. Меру $e^{-V} dx$ на $\mathbb{R}^n$ называют выпуклой с константой $C > 0$, если $V'' \geq C \cdot \text{Id}$.

ТЕОРЕМА 3.9.2. Пусть борелевская вероятностная мера μ на $X = \mathbb{R}^\infty$ равномерно выпукла с константой $C > 0$, т. е. все ее проекции на $\mathbb{R}^n$ выпуклы с константой $C > 0$, и пусть $H = l^2$. Если $\nu \ll \mu$ — такая вероятностная мера, что для $f = d\nu/d\mu$ мы имеем $f \ln f \in L^1(\mu)$, то каноническое треугольное отображение $T_{\mu,\nu}$ обладает тем свойством, что

$$\int_X |T_{\mu,\nu}(x) - x|_H^2 \mu(dx) \leq \frac{2}{C} \int_X f \ln f d\mu.$$

О вероятностных применениях треугольных отображений см. [188], [189]. В работе [190] треугольные отображения представлены в виде пределов решений задач Монжа со специально подобранными функциями стоимости.

3.9.2. Преобразования Мозера, полугруппы, транспортные уравнения. Ряд отображений, преобразующих заданную меру-прообраз в заданную меру-образ, можно построить с помощью классической теоремы Лиувилля об изменении объема. Идея таких конструкций восходит к работе Ю. Мозера [191] о диффеоморфизмах многообразий с мерами. Например, в работе [192] было выполнено следующее построение. Рассмотрим полугруппу $P_t = e^{tL}$, порожденную эллиптическим оператором $L = \Delta - \langle \nabla V, \nabla \rangle = e^V \operatorname{div}(e^{-V} \cdot \nabla)$, и поток вероятностных мер $\nu_t = P_t(e^{-W+V}) \cdot \mu$. Очевидно, что μ — инвариантная мера для P_t , $\nu_0 = \nu$ и $\nu_\infty = \mu$. Запишем уравнение для ν_t :

$$\frac{d}{dt} \nu_t = LP_t(e^{-W+V}) \cdot \mu = \operatorname{div}[\nabla P_t(e^{-W+V}) \cdot e^{-V}] = \operatorname{div}[\nabla \ln P_t(e^{-W+V}) \cdot \nu_t].$$

По теореме Лиувилля, если $X(t)$ — решение уравнения $dX(t)/dt = b(X(t))$, $X(0) = I$, где b — гладкое, то образ $\varrho_t dx$ меры Лебега относительно диффеоморфизма $X(t)$ удовлетворяет уравнению $\partial \varrho_t / \partial t = \operatorname{div}(\varrho_t \cdot b)$. Соответствующий поток диффеоморфизмов задан уравнением $dS_t/dt = -\nabla \ln P_t(e^{-W+V}) \circ S_t$, $S_0 = \operatorname{Id}$, где ν_t и S_t связаны соотношением $\nu_t = \nu \circ S_t^{-1}$. В частности, предельное отображение $S_\infty = \lim_{t \rightarrow \infty} S_t$ переводит ν в μ .

Связанные с оптимальными транспортировками результаты об эволюционных уравнениях см. в [193]. О связях с мерами Янга см. [194]. Про задачи Монжа и Канторовича с несколькими мерами см. [195], [15]. Дифференциальные формы на пространстве вероятностных мер рассмотрены в [196]. Поток публикаций по рассмотренным вопросам весьма интенсивен, представленная библиография охватывает заметно менее его половины.

Список литературы

- [1] Л. В. Канторович, “О перемещении масс”, *Докл. АН СССР*, **37** (1942), 227–229.
- [2] Л. В. Канторович, “Об одной проблеме Монжа”, *УМН*, **3:2(24)** (1948), 225–226.
- [3] Л. В. Канторович, Г. Ш. Рубинштейн, “Об одном функциональном пространстве и некоторых экстремальных задачах”, *Докл. АН СССР*, **115:6** (1957), 1058–1061.
- [4] Л. В. Канторович, Г. Ш. Рубинштейн, “Об одном пространстве вполне аддитивных функций множества”, *Вестн. ЛГУ*, **13:7** (1958), 52–59.
- [5] А. М. Вершик, “Метрика Канторовича: начальная история и малоизвестные применения”, *Теория представлений, динамические системы. XI*, Специальный выпуск, Зап. научн. сем. ПОМИ, **312**, ПОМИ, СПб., 2004, 69–85; англ.

- пер.: A. M. Vershik, “The Kantorovich metric: initial history and little-known applications”, *J. Math. Sci. (N. Y.)*, **133**:4 (2006), 1410–1417.
- [6] G. Monge, *Mémoire sur la théorie des déblais et de remblais*, Histoire de l’Académie Royale des Sciences avec les Mémoires de Mathématique & de Physique, Paris, 1781.
- [7] Леонид Витальевич Канторович: человек и ученый, т. 1, 2, Наука Сибири в лицах, ред. В. Л. Канторович, С. С. Кутателадзе, Я. И. Фет, Изд-во СО РАН, филиал “Гео”, Новосибирск, 2002, 543, 614 с.
- [8] А. М. Вершик, С. С. Кутателадзе, С. П. Новиков, “Леонид Витальевич Канторович (к 100-летию со дня рождения)”, *УМН*, **67**:3(405) (2012), 185–191; англ. пер.: A. M. Vershik, S. S. Kutateladze, S. P. Novikov, “Leonid Vital’evich Kantorovich (on the 100th anniversary of his birth)”, *Russian Math. Surveys*, **67**:3 (2012), 589–597.
- [9] L. Ambrosio, “Lecture notes on optimal transport problems”, *Mathematical aspects of evolving interfaces* (Funchal, 2000), Lecture Notes in Math., **1812**, Springer, Berlin, 2003, 1–52.
- [10] L. Ambrosio, N. Gigli, *A user’s guide to optimal transport*, Scuola Normale Superiore di Pisa, Pisa, 2011.
- [11] L. Ambrosio, N. Gigli, G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures Math. ETH Zurich, Birkhäuser Verlag, Basel, 2005, viii+333 pp.
- [12] L. C. Evans, “Partial differential equations and Monge–Kantorovich mass transfer”, *Current developments in mathematics*, 1997 (Cambridge, MA), Int. Press, Boston, MA, 1999, 65–126.
- [13] W. Gangbo, R. J. McCann, “The geometry of optimal transportation”, *Acta Math.*, **177**:2 (1996), 113–161.
- [14] R. J. McCann, N. Guillen, *Five lectures on optimal transportation: geometry, regularity and applications*, 2010, arXiv: 1011.2911v1.
- [15] S. T. Rachev, L. Rüschendorf, *Mass transportation problems*, v. I, II, Probab. Appl. (N. Y.), Springer-Verlag, New York, 1998, xxvi+508, xxvi+430 pp.
- [16] C. Villani, *Topics in optimal transportation*, Grad. Stud. Math., **58**, Amer. Math. Soc., Providence, RI, 2003, xvi+370 pp.
- [17] C. Villani, *Optimal transport. Old and new*, Grundlehren Math. Wiss., **338**, Springer-Verlag, Berlin, 2009, xxii+973 pp.
- [18] А. М. Вершик, “Несколько замечаний о бесконечномерных задачах линейного программирования”, *УМН*, **25**:5(155) (1970), 117–124; англ. пер.: A. M. Vershik, “Some remarks on the infinite-dimensional problems of linear programming”, *Russian Math. Surveys*, **25**:5 (1970), 117–124.
- [19] А. Н. Боголюбов, *Гаспар Монж*, Наука, М., 1978, 184 с.
- [20] P. Appell, “Mémoire sur les déblais et les remblais des systèmes continus ou discontinus”, *Mém. Acad. Sci. Inst. France*, **29**, № 3, Paris, 1887, 208 pp.
- [21] В. Н. Судakov, “Геометрические проблемы теории бесконечномерных вероятностных распределений”, *Тр. МИАН СССР*, **141**, 1976, 3–191; англ. пер.: V. N. Sudakov, “Geometric problems of the theory of infinite-dimensional probability distributions”, *Proc. Steklov Inst. Math.*, **141** (1979), 1–178.
- [22] L. Ambrosio, B. Kirchheim, A. Pratelli, “Existence of optimal transport maps for chrystalline norms”, *Duke Math. J.*, **125**:2 (2004), 207–241.
- [23] L. C. Evans, W. Gangbo, *Differential equations methods for the Monge–Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., **137**, № 653, 1999, viii+66 pp.
- [24] T. Champion, L. De Pascale, “The Monge problem in $\mathbb{R}^d$ ”, *Duke Math. J.*, **157**:3 (2011), 551–572.
- [25] В. И. Богачев, “О работах Г. М. Фихтенгольца по теории интеграла”, *Истор.-матем. исслед. Сер. 2*, 2005, № 9(44), 252–264.

- [26] Л. В. Канторович, *Математические методы организации и планирования производства*, Изд-во ЛГУ, Л., 1939, 67 с.; репринтное изд.: Изд. дом СПбГУ, СПб., 2012.
- [27] А. М. Вершик, “Многозначные отображения с инвариантной мерой (полиморфизмы) и марковские операторы”, *Проблемы теории вероятностных распределений. IV*, Зап. научн. сем. ЛОМИ, **72**, Наука, Л., 1977, 26–61; англ. пер.: A. M. Vershik, “Many-valued measure-preserving mappings (polymorphisms) and Markovian operators”, *J. Soviet Math.*, **23**:3 (1983), 2243–2266.
- [28] В. Л. Левин, “О теоремах двойственности в задаче Монжа–Канторовича”, *УМН*, **32**:3(195) (1977), 171–172.
- [29] В. И. Богачев, *Основы теории меры*, т. 1, 2, 2-е изд., НИЦ “Регулярная и хаотическая динамика”, М.–Ижевск, 2006, 584, 680 с.; англ. пер.: V. I. Bogachev, *Measure theory*, v. I, II, Springer-Verlag, Berlin, 2007, xviii+500, xiv+575 pp.
- [30] R. Fortet, E. Mourier, “Convergence de la répartition empirique vers la répartition théorique”, *Ann. Sci. Ecole Norm. Sup.* (3), **70**:3 (1953), 267–285.
- [31] R. F. Arens, J. Eells, Jr., “On embedding uniform and topological spaces”, *Pacif. J. Math.*, **6** (1956), 397–403.
- [32] D. A. Edwards, “On the Kantorovich–Rubinshtein theorem”, *Expos. Math.*, **29**:4 (2011), 387–398.
- [33] A. M. Vershik, “Dynamics of metrics in measure spaces and their asymptotic invariants”, *Markov Process. Related Fields*, **16**:1 (2010), 169–184.
- [34] J. Melleray, F. V. Petrov, A. M. Vershik, “Linearly rigid metric spaces and the embedding problem”, *Fund. Math.*, **199**:2 (2008), 177–194.
- [35] В. М. Золотарев, “О непрерывности стохастических последовательностей, порождаемых рекуррентными процедурами”, *Теория вероятн. и ее примен.*, **20**:4 (1975), 834–847; англ. пер.: V. M. Zolotarev, “On the continuity of stochastic sequences generated by recurrent processes”, *Theory Probab. Appl.*, **20**:4 (1976), 819–832.
- [36] В. М. Золотарев, “Метрические расстояния в пространствах случайных величин и их распределений”, *Матем. сб.*, **101(143)**:3(11) (1976), 416–454; англ. пер.: V. M. Zolotarev, “Metric distances in spaces of random variables and their distributions”, *Math. USSR-Sb.*, **30**:3 (1976), 373–401.
- [37] С. Т. Рачев, “Задача Монжа–Канторовича о перемещении масс и ее применения в стохастике”, *Теория вероятн. и ее примен.*, **29**:4 (1984), 625–653; англ. пер.: S. T. Rachev, “Monge–Kantorovich problem on mass transition and its applications in stochastics”, *Theory Probab. Appl.*, **29**:4 (1985), 625–653.
- [38] Р. Л. Добрушин, “Задание системы случайных величин при помощи условных распределений”, *Теория вероятн. и ее примен.*, **15**:3 (1970), 469–497; англ. пер.: R. L. Dobrushin, “Prescribing a system of random variables by conditional distributions”, *Theory Probab. Appl.*, **15**:3 (1970), 458–486.
- [39] Л. Н. Васерштейн, “Марковские процессы на счетном произведении пространства, описывающие большие системы автоматов”, *Пробл. передачи информ.*, **5**:3 (1969), 64–72; англ. пер.: L. N. Vasershtein, “Markov processes over denumerable products of spaces describing large system of automata”, *Problems Inform. Transmission*, **5**:3 (1969), 47–52.
- [40] Y. Olivier, “A survey of Ricci curvature for metric spaces and Markov chains”, *Probabilistic approach to geometry*, Adv. Stud. Pure Math., **57**, Math. Soc. Japan, Tokyo, 2010, 343–381.
- [41] L. Rüschendorf, “On c -optimal random variables”, *Statist. Probab. Lett.*, **27**:3 (1996), 267–270.
- [42] M. Knott, C. S. Smith, “On the optimal mapping of distributions”, *J. Optim. Theory Appl.*, **43**:1 (1984), 39–49.

- [43] M. Beiglböck, M. Goldstern, G. Maresch, W. Schachermayer, “Optimal and better transport plans”, *J. Funct. Anal.*, **256**:6 (2009), 1907–1927.
- [44] А. Ю. Плахов, “Точные решения одномерной задачи Монжа–Канторовича”, *Матем. сб.*, **195**:9 (2004), 57–74; англ. пер.: A. Yu. Plakhov, “Precise solutions of the one-dimensional Monge–Kantorovich problem”, *Sb. Math.*, **195**:9 (2004), 1291–1307.
- [45] В. Л. Левин, “Условия оптимальности и точные решения двумерной задачи Монжа–Канторовича”, *Теория представлений, динамические системы. XI*, Специальный выпуск, Зап. научн. сем. ПОМИ, **312**, ПОМИ, СПб., 2004, 150–164; англ. пер.: V. L. Levin, “Optimality conditions and exact solutions to the two-dimensional Monge–Kantorovich problem”, *J. Math. Sci. (N. Y.)*, **133**:4 (2006), 1456–1463.
- [46] V. L. Levin, “General Monge–Kantorovich problem and its applications in measure theory and mathematical economics”, *Functional analysis, optimization, and mathematical economics*, A collection of papers dedicated to the memory of L. V. Kantorovich, ed. L. J. Leifman, Oxford Univ. Press, New York, 1990, 141–176.
- [47] D. Ramachandran, L. Rüschendorf, “A general duality theorem for marginal problems”, *Probab. Theory Relat. Fields*, **101**:3 (1995), 311–319.
- [48] D. Ramachandran, L. Rüschendorf, “Duality and perfect probability spaces”, *Proc. Amer. Math. Soc.*, **124**:7 (1996), 2223–2228.
- [49] D. Ramachandran, L. Rüschendorf, “On the Monge–Kantorovich duality theorem”, *Теория вероятн. и ее примен.*, **45**:2 (2000), 403–409; D. Ramachandran, L. Rüschendorf, “On the Monge–Kantorovich duality theorem”, *Theory Probab. Appl.*, **45**:2 (2001), 350–356.
- [50] M. Beiglböck, W. Schachermayer, “Duality for Borel measurable cost functions”, *Trans. Amer. Math. Soc.*, **363**:8 (2011), 4203–4224.
- [51] H. G. Kellerer, “Duality theorems for marginal problems”, *Z. Wahrscheinlichkeitstheor. verw. Geb.*, **67**:4 (1984), 399–432.
- [52] D. A. Edwards, “A simple proof in Monge–Kantorovich duality theory”, *Studia Math.*, **200**:1 (2010), 67–77.
- [53] L. Ambrosio, A. Pratelli, “Existence and stability results in the L^1 theory of optimal transportation”, *Optimal transportation and applications* (Martina Franca, 2001), Lecture Notes in Math., **1813**, Springer, Berlin, 2003, 123–160.
- [54] В. Л. Левин, “Формула для оптимального значения задачи Монжа–Канторовича с гладкой функцией стоимости и характеристика циклически монотонных отображений”, *Матем. сб.*, **181**:12 (1990), 1694–1709; англ. пер.: V. L. Levin, “A formula for the optimal value in the Monge–Kantorovich problem with a smooth cost function, and a characterization of cyclically monotone mappings”, *Math. USSR-Sb.*, **71**:2 (1992), 533–548.
- [55] В. Л. Левин, “К теории двойственности для нетопологических вариантов задачи о перемещении масс”, *Матем. сб.*, **188**:4 (1997), 95–126; англ. пер.: V. L. Levin, “On duality theory for non-topological variants of the mass transfer problem”, *Sb. Math.*, **188**:4 (1997), 571–602.
- [56] В. Л. Левин, “Двойственность Монжа–Канторовича и ее применение в теории полезности”, *Экономика и матем. методы*, **47**:4 (2011), 143–165.
- [57] V. Levin, “Abstract cyclical monotonicity and Monge solutions for the general Monge–Kantorovich problem”, *Set-Valued Anal.*, **7**:1 (1999), 7–32.
- [58] В. Л. Левин, А. А. Милютин, “Задача о перемещении масс с разрывной функцией стоимости и массовая постановка проблемы двойственности выпуклых экстремальных задач”, *УМН*, **34**:3(207) (1979), 3–68; англ. пер.: V. L. Levin, A. A. Milyutin, “The problem of mass transfer with a discontinuous cost function and a mass statement of the duality problem for convex extremal problems”, *Russian Math. Surveys*, **34**:3 (1979), 1–78.

- [59] L. A. Caffarelli, M. Feldman, R. J. McCann, “Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs”, *J. Amer. Math. Soc.*, **15**:1 (2002), 1–26.
- [60] N. S. Trudinger, X.-J. Wang, “On the Monge mass transfer problem”, *Calc. Var. Partial Differential Equations*, **13**:1 (2001), 19–31.
- [61] T. Champion, L. De Pascale, “The Monge problem for strictly convex norms in $\mathbb{R}^d$ ”, *J. Eur. Math. Soc.*, **12**:6 (2010), 1355–1369.
- [62] L. Caravenna, “A proof of Sudakov theorem with strictly convex norms”, *Math. Z.*, **268**:1-2 (2011), 371–407.
- [63] А. В. Погорелов, *Многомерное уравнение Монжа–Ампера* $\det \|z_{ij}\| = \phi(z_1, \dots, z_n, z, x_1, \dots, x_n)$, Наука, М., 1988, 94 с.
- [64] C. E. Gutiérrez, *The Monge–Ampère equation*, Progr. Nonlinear Differential Equations Appl., **44**, Birkhäuser Boston, Inc., Boston, MA, 2001, xii+127 pp.
- [65] Y. Brenier, “Polar factorization and monotone rearrangement of vector valued functions”, *Comm. Pure Appl. Math.*, **44**:4 (1991), 375–417.
- [66] R. J. McCann, “Existence and uniqueness of monotone measure-preserving maps”, *Duke Math. J.*, **80**:2 (1995), 309–323.
- [67] В. И. Богачев, *Дифференцируемые меры и исчисление Маллявена*, НИЦ “Регулярная и хаотическая динамика”, М.–Ижевск, 2008, 544 с.; англ. пер.: V. I. Bogachev, *Differentiable measures and the Malliavin calculus*, Math. Surveys Monogr., **164**, Amer. Math. Soc., Providence, RI, 2010, xvi+488 pp.
- [68] R. J. McCann, “A convexity principle for interacting gases”, *Adv. Math.*, **128**:1 (1997), 153–179.
- [69] R. van der Putten, “Uniqueness and non uniqueness of optimal maps in mass transport problem with not strictly convex cost”, *Comment. Math. Univ. Carolin.*, **51**:1 (2010), 67–83.
- [70] N. V. Krylov, “Fully nonlinear second order elliptic equations: recent development”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci* (4), **25**:3-4 (1997), 569–595.
- [71] X. Cabré, “Elliptic PDE’s in probability and geometry: symmetry and regularity of solutions”, *Discrete Contin. Dyn. Syst.*, **20**:3 (2008), 425–457.
- [72] N. S. Trudinger, X.-J. Wang, “The Monge–Ampère equation and its geometric applications”, *Handbook of geometric analysis*, № 1, Adv. Lect. Math. (ALM), **7**, Int. Press, Somerville, MA, 2008, 467–524.
- [73] L. Forzani, D. Maldonado, “Properties of the solutions to the Monge–Ampère equation”, *Nonlinear Anal.*, **57**:5-6 (2004), 815–829.
- [74] G. De Philippis, A. Figalli, *$W^{2,1}$ -regularity for solutions of the Monge–Ampère equation*, 2011, arXiv: 1111.7207v2.
- [75] I. Fragalà, M. S. Gelli, A. Pratelli, “Continuity of an optimal transport in Monge problem”, *J. Math. Pures Appl.* (9), **84**:9 (2005), 1261–1294.
- [76] X.-J. Wang, “Some counterexamples to the regularity of Monge–Ampère equations”, *Proc. Amer. Math. Soc.*, **123**:3 (1995), 841–845.
- [77] L. A. Caffarelli, “Monotonicity properties of optimal transportation and the FKG and related inequalities”, *Comm. Math. Phys.*, **214**:3 (2000), 547–563.
- [78] D. Bakry, M. Ledoux, “Lévy–Gromov’s isoperimetric inequality for an infinite-dimensional diffusion generator”, *Invent. Math.*, **123**:2 (1996), 259–281.
- [79] D. Cordero-Erausquin, “Some applications of mass transport to Gaussian-type inequalities”, *Arch. Ration. Mech. Anal.*, **161**:3 (2002), 257–269.
- [80] А. В. Колесников, “Глобальные гёльдеровы оценки для оптимальной транспортировки”, *Матем. заметки*, **88**:5 (2010), 708–728; англ. пер.: A. V. Kolesnikov, “Global Hölder estimates for optimal transportation”, *Math. Notes*, **88**:5-6 (2010), 678–695.

- [81] B. Klartag, *Poincaré inequalities and moment maps*, 2011, arXiv: 1104.2791.
- [82] А. В. Колесников, “Соболевская регулярность транспортировки вероятностных мер и транспортные неравенства”, *Теория вероятн. и ее примен.*, **57**:2 (2012), 296–321.
- [83] В. И. Богачев, А. В. Колесников, “Соболевская регулярность для бесконечномерного уравнения Монжа–Ампера”, *Докл. РАН*, **444**:2 (2012), 131–136.
- [84] V. I. Bogachev, A. V. Kolesnikov, “Sobolev regularity for the Monge–Ampère equation in the Wiener space”, *Kyoto J. Math.* (to appear); 2011, arXiv: 1110.1822.
- [85] L. De Pascale, L. C. Evans, A. Pratelli, “Integral estimates for transport densities”, *Bull. London Math. Soc.*, **36**:3 (2004), 383–395.
- [86] G. Loeper, “On the regularity of solutions of optimal transportation problems”, *Acta Math.*, **202**:2 (2009), 241–283.
- [87] J. Liu, N. S. Trudinger, X.-J. Wang, “Interior $C^{2,\alpha}$ regularity for potential functions in optimal transportation”, *Comm. Partial Differential Equations*, **35**:1 (2010), 165–184.
- [88] G. Loeper, C. Villani, “Regularity of optimal transport in curved geometry: the nonfocal case”, *Duke Math. J.*, **151**:3 (2010), 431–485.
- [89] L. Caffarelli, R. J. McCann, “Free boundaries in optimal transport and Monge–Ampère obstacle problems”, *Ann. of Math.* (2), **171**:2 (2010), 673–730.
- [90] A. Figalli, “The optimal partial transport problem”, *Arch. Ration. Mech. Anal.*, **195**:2 (2010), 533–560.
- [91] A. Figalli, Y.-H. Kim, “Partial regularity of Brenier solutions of the Monge–Ampère equation”, *Discrete Contin. Dyn. Syst.*, **28**:2 (2010), 559–565.
- [92] А. А. Липчюс, “Замечание о равенстве в задачах Монжа и Канторовича”, *Теория вероятн. и ее примен.*, **50**:4 (2005), 779–782; англ. пер.: A. A. Lipchius, “A note on the equality in the Monge and Kantorovich problems”, *Theory Probab. Appl.*, **50**:4 (2006), 689–693.
- [93] A. Pratelli, “On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation”, *Ann. Inst. H. Poincaré Probab. Statist.*, **43**:1 (2007), 1–13.
- [94] V. D. Milman, G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, with an appendix by M. Gromov, Lecture Notes in Math., **1200**, Springer-Verlag, Berlin, 1986, viii+156 pp.
- [95] D. Cordero-Erausquin, B. Nazaret, C. Villani, “A mass-transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities”, *Adv. Math.*, **182**:2 (2004), 307–332.
- [96] B. Nazaret, “Best constant in Sobolev trace inequalities on the half space”, *Nonlinear Anal.*, **65**:10 (2006), 1977–1985.
- [97] F. Maggi, C. Villani, “Balls have the worst best Sobolev inequalities”, *J. Geom. Anal.*, **15**:1 (2005), 83–121; “Balls have the worst best Sobolev inequalities. II: Variants and extensions”, *Calc. Var. Partial Differential Equations*, **31**:1 (2008), 47–74.
- [98] A. Figalli, F. Maggi, A. Pratelli, “A mass transportation approach to quantitative isoperimetric inequalities”, *Invent. Math.*, **182**:1 (2010), 167–211.
- [99] D. Cordero-Erausquin, R. J. McCann, M. Schmuckenschläger, “A Riemannian interpolation inequality à la Borell, Brascamp and Lieb”, *Invent. Math.*, **146**:2 (2001), 219–257.
- [100] D. Cordero-Erausquin, R. J. McCann, M. Schmuckenschläger, “Prékopa–Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport”, *Ann. Fac. Sci. Toulouse Math.* (6), **15**:4 (2006), 613–635.

- [101] P. Castillon, “Submanifolds, isoperimetric inequalities and optimal transportation”, *J. Funct. Anal.*, **259**:1 (2010), 79–103; 2009, arXiv: 0908.2711.
- [102] L. Gross, “Logarithmic Sobolev inequalities”, *Amer. J. Math.*, **97**:4 (1975), 1061–1083.
- [103] M. Ledoux, *The concentration of measure phenomenon*, Math. Surveys Monogr., **89**, Amer. Math. Soc., Providence, RI, 2001, x+181 pp.
- [104] В. И. Богачев, *Гауссовские меры*, Наука, М., 1997, 352 с.
- [105] S. G. Bobkov, F. Götze, “Exponential integrability and transportation cost related to logarithmic Sobolev inequality”, *J. Funct. Anal.*, **163** (1999), 1–28.
- [106] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, 2002, arXiv: math/0211159.
- [107] В. Н. Судаков, Б. С. Цирельсон, “Экстремальные свойства полупространств для сферически инвариантных мер”, *Проблемы теории вероятностных распределений. II*, Зап. научн. сем. ЛОМИ, **41**, Наука, Л., 1974, 14–24.
- [108] C. Borell, “The Brunn–Minkowski inequality in Gauss space”, *Invent. Math.*, **30**:2 (1975), 207–216.
- [109] F.-Y. Wang, “Logarithmic Sobolev inequalities on non-compact Riemannian manifolds”, *Probab. Theory Related Fields*, **109**:3 (1997), 417–424.
- [110] F. Barthe, A. V. Kolesnikov, “Mass transport and variants of the logarithmic Sobolev inequality”, *J. Geom. Anal.*, **18**:4 (2008), 921–979.
- [111] O. S. Rothaus, “Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities”, *J. Funct. Anal.*, **64**:2 (1985), 296–313.
- [112] F. Barthe, P. Cattiaux, C. Roberto, “Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry”, *Rev. Mat. Iberoam.*, **22**:3 (2006), 993–1067.
- [113] D. Cordero-Erausquin, W. Gangbo, C. Houdré, “Inequalities for generalized entropies and optimal transportation”, *Recent advances in the theory and applications of mass transport*, Contemp. Math., **353**, Amer. Math. Soc., Providence, RI, 2004, 73–94.
- [114] E. Milman, S. Sodin, “An isoperimetric inequality for uniformly log-concave measures and uniformly convex bodies”, *J. Funct. Anal.*, **254**:5 (2008), 1235–1268.
- [115] S. G. Bobkov, “Isoperimetric and analytic inequalities for log-concave probability measures”, *Ann. Probab.*, **27**:4 (1999), 1903–1921.
- [116] A. V. Kolesnikov, “Modified log-Sobolev inequalities and isoperimetry”, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, **18**:2 (2007), 179–208.
- [117] M. Talagrand, “Transportation cost for Gaussian and other product measures”, *Geom. Funct. Anal.*, **6**:3 (1996), 587–600.
- [118] В. И. Богачев, А. В. Колесников, “Интегрируемость абсолютно непрерывных преобразований мер и применения к оптимальному переносу масс”, *Теория вероятн. и ее примен.*, **50**:3 (2005), 433–456; англ. пер.: V. I. Bogachev, A. V. Kolesnikov, “Integrability of absolutely continuous transformations of measures and applications to optimal mass transportation”, *Theory Probab. Appl.*, **50**:3 (2006), 367–385.
- [119] F. Otto, C. Villani, “Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality”, *J. Funct. Anal.*, **173**:2 (2000), 361–400.
- [120] S. G. Bobkov, I. Gentil, M. Ledoux, “Hypercontractivity of Hamilton–Jacobi equations”, *J. Math. Pures Appl.* (9), **80**:7 (2001), 669–696.
- [121] P. Cattiaux, A. Guillin, “On quadratic transportation cost inequalities”, *J. Math. Pures Appl.* (9), **86**:4 (2006), 342–361.

- [122] N. Gozlan, “Characterization of Talagrand’s like transportation-cost inequalities on the real line”, *J. Funct. Anal.*, **250**:2 (2007), 400–425.
- [123] N. Gozlan, “A characterization of dimension free concentration in terms of transportation inequalities”, *Ann. Probab.*, **37**:6 (2009), 2480–2498.
- [124] I. Gentil, A. Guillin, L. Miclo, “Modified logarithmic Sobolev inequalities and transportation inequalities”, *Probab. Theory Related Fields*, **133**:3 (2005), 409–436.
- [125] H. Djellout, A. Guillin, L. Wu, “Transportation cost-information inequalities and applications to random dynamical systems and diffusions”, *Ann. Probab.*, **32**:3B (2004), 2702–2732.
- [126] F. Bolley, C. Villani, “Weighted Csiszár–Kullback–Pinsker inequalities and application to transportation inequalities”, *Ann. Fac. Sci. Toulouse Math.* (6), **14**:3 (2005), 331–352.
- [127] N. Gozlan, C. Roberto, P.M. Samson, “A new characterization of Talagrand’s transport-entropy inequalities and applications”, *Ann. Probab.*, **39**:3 (2011), 857–880.
- [128] P. Cattiaux, A. Guillin, L.-M. Wu, “A note on Talagrand’s transportation inequality and logarithmic Sobolev inequality”, *Probab. Theory Related Fields*, **148**:1-2 (2010), 285–304.
- [129] A. Guillin, C. Léonard, L. Wu, N. Yao, “Transportation-information inequalities for Markov processes”, *Probab. Theory Related Fields*, **144**:3-4 (2009), 669–695.
- [130] A. Guillin, C. Leonard, F.-Y. Wang, L. Wu, *Transportation-information inequalities for Markov processes (II): relations with other functional inequalities*, 2009, arXiv: 0902.2101.
- [131] N. Gozlan, C. Léonard, “Transport inequalities. A survey”, *Markov Process. Related Fields*, **16**:4 (2010), 635–736.
- [132] J. Dolbeault, B. Nazaret, G. Savaré, “A new class of transport distances between measures”, *Calc. Var. Partial Differential Equations*, **34**:2 (2009), 193–231.
- [133] L. Gozlan, C. Léonard, “A large deviation approach to some transportation cost inequalities”, *Probab. Theory Related Fields*, **139**:1-2 (2007), 235–283.
- [134] E. Milman, “Isoperimetric and concentration inequalities: equivalence under curvature lower bound”, *Duke Math. J.*, **154**:2 (2010), 207–239.
- [135] E. Milman, “Properties of isoperimetric, functional and transport-entropy inequalities via concentration”, *Probab. Theory Related Fields*, **152**:3-4 (2012), 475–507.
- [136] J. Lott, C. Villani, “Ricci curvature for metric-measure spaces via optimal transport”, *Ann. Math.* (2), **169**:3 (2009), 903–991.
- [137] L. Ambrosio, A. Figalli, “Geodesics in the space of measure-preserving maps and plans”, *Arch. Ration. Mech. Anal.*, **194**:2 (2009), 421–462.
- [138] S. Angenent, S. Haker, A. Tannenbaum, “Minimizing flows for the Monge–Kantorovich problem”, *SIAM J. Math. Anal.*, **35**:1 (2003), 61–97.
- [139] L. De Pascale, M. S. Gelli, L. Granieri, “Minimal measures, one-dimensional currents and the Monge–Kantorovich problem”, *Calc. Var. Partial Differential Equations*, **27**:1 (2006), 1–23.
- [140] A. Figalli, N. Gigli, “A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions”, *J. Math. Pures Appl.* (9), **94**:2 (2010), 107–130.
- [141] J.-D. Benamou, Y. Brenier, “A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem”, *Numer. Math.*, **84**:3 (2000), 375–393.
- [142] N. Juillet, “On displacement interpolation of measures involved in Brenier’s theorem”, *Proc. Amer. Math. Soc.*, **139**:10 (2011), 3623–3632.
- [143] K.-Th. Sturm, “On the geometry of metric measure spaces, I, II”, *Acta Math.*, **196** (2006), 65–131, 133–177.

- [144] J. Lott, “Some geometric calculations on Wasserstein spaces”, *Comm. Math. Phys.*, **277**:2 (2008), 423–437.
- [145] F. Otto, “The geometry of dissipative evolution equations: the porous medium equation”, *Comm. Partial Differential Equations*, **26**:1-2 (2001), 101–174.
- [146] F. Otto, M. Westdickenberg, “Eulerian calculus for the contraction in the Wasserstein distance”, *SIAM J. Math. Anal.*, **37**:4 (2005), 1227–1255 (electronic).
- [147] U. Gianazza, G. Savaré, G. Toscani, “The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation”, *Arch. Ration. Mech. Anal.*, **194**:1 (2009), 133–220.
- [148] Р. Л. Добрушин, “Уравнения Власова”, *Функц. анализ и его прил.*, **13**:2 (1979), 48–58; англ. пер.: R. L. Dobrushin, “Vlasov equations”, *Funct. Anal. Appl.*, **13**:2 (1979), 115–123.
- [149] R. Jordan, D. Kinderlehrer, F. Otto, “The variational formulation of the Fokker–Planck equation”, *SIAM J. Math. Anal.*, **29**:1 (1998), 1–17.
- [150] L. Natile, M. A. Peletier, G. Savaré, “Contraction of general transportation costs along solutions to Fokker–Planck equations with monotone drifts”, *J. Math. Pures Appl.* (9), **95**:1 (2011), 18–35.
- [151] K. Kuwada, “Duality on gradient estimates and Wasserstein controls”, *J. Funct. Anal.*, **258**:11 (2010), 3758–3774.
- [152] M.-K. von Renesse, K.-Th. Sturm, “Entropic measure and Wasserstein diffusion”, *Ann. Probab.*, **37**:3 (2009), 1114–1191.
- [153] А. Д. Александров, *Избранные труды*, т. 1, 2, Наука, Новосибирск, 2006, 2007, lii+748, iv+491 с.
- [154] V. Ollier, “Embeddings of S^n into $\mathbb{R}^{n+1}$ with given integral Gauss curvature and optimal mass transport on S^n ”, *Adv. Math.*, **213**:2 (2007), 600–620.
- [155] R. J. McCann, “Polar factorization of maps on Riemannian manifolds”, *Geom. Funct. Anal.*, **11**:3 (2001), 589–608.
- [156] A. Fathi, A. Figalli, “Optimal transportation on non-compact manifolds”, *Israel J. Math.*, **175** (2010), 1–59.
- [157] A. Figalli, L. Rifford, “Mass transportation on sub-Riemannian manifolds”, *Geom. Funct. Anal.*, **20**:1 (2010), 124–159.
- [158] Y.-H. Kim, R. J. McCann, “Continuity, curvature, and the general covariance of optimal transportation”, *J. Eur. Math. Soc.*, **12**:4 (2010), 1009–1040.
- [159] J.-P. Bourguignon, “Ricci curvature and measures”, *Japan. J. Math.*, **4**:1 (2009), 27–45.
- [160] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progr. Math., **152**, Birkhäuser Boston, Inc., Boston, MA, 1999, xx+585 pp.; пер. с франц.: M. Gromov, *Structures métriques pour les variétés riemanniennes*, Textes Math., **1**, eds. J. Lafontaine, P. Pansu, CEDIC, Paris, 1981, iv+152 pp.
- [161] А. М. Вершик, “Универсальное пространство Урысона, метрические тройки Громова и случайные метрики на натуральном ряде”, *УМН*, **53**:5(323) (1998), 57–64; англ. пер.: A. M. Vershik, “The universal Urysohn space, Gromov metric triples and random metrics on the natural numbers”, *Russian Math. Surveys*, **53**:5 (1998), 921–928.
- [162] А. М. Вершик, “Случайные метрические пространства и универсальность”, *УМН*, **59**:2(356) (2004), 65–104; англ. пер.: A. M. Vershik, “Random metric spaces and universality”, *Russian Math. Surveys*, **59**:2 (2004), 259–295.
- [163] K. Bacher, “On Borell–Brascamp–Lieb inequalities on metric measure spaces”, *Potential Anal.*, **33**:1 (2010), 1–15.
- [164] K. Bacher, K.-Th. Sturm, “Localization and tensorization properties of the curvature-dimension condition for metric measure spaces”, *J. Funct. Anal.*, **259**:1 (2010), 28–56.

- [165] Q. Deng, K.-Th. Sturm, “Localization and tensorization properties of the curvature-dimension condition for metric measure spaces, II”, *J. Funct. Anal.*, **260**:12 (2011), 3718–3725.
- [166] D. Bakry, M. Émery, “Diffusions hypercontractives”, *Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math.*, **1123**, Springer, Berlin, 1985, 177–206.
- [167] V. I. Bogachev, A. V. Kolesnikov, “Mass transport generated by a flow of Gauss maps”, *J. Funct. Anal.*, **256**:3 (2009), 940–957.
- [168] G. Huisken, “Flow by mean curvature of convex surfaces into spheres”, *J. Differential Geom.*, **20**:1 (1984), 237–266.
- [169] K. Tso, “Deforming a hypersurfaces by its Gauss–Kronecker curvature”, *Comm. Pure Appl. Math.*, **38**:6 (1985), 867–882.
- [170] A. V. Kolesnikov, *Weak regularity of Gauss mass transport*, 2009, arXiv:0904.1852.
- [171] Y.-H. Kim, J. Streets, M. Warren, *Parabolic optimal transport equations on manifolds*, 2010, arXiv:1008.3824v1.
- [172] R. J. McCann, P. M. Topping, “Ricci flow, entropy and optimal transportation”, *Amer. J. Math.*, **132**:3 (2010), 711–730.
- [173] P. Topping, *Ricci flows: the foundations via optimal transportation*, <http://homepages.warwick.ac.uk/~maseq/grenoble.20100324.pdf>.
- [174] P. Topping, “ $\mathcal{L}$ -optimal transportation for Ricci flows”, *J. Reine Angew. Math.*, **636** (2009), 93–122.
- [175] J. Lott, “Optimal transport and Perelman’s reduced volume”, *Calc. Var. Partial Differential Equations*, **36**:1 (2009), 49–84.
- [176] G. Buttazzo, A. Pratelli, S. Solimini, E. Stepanov, “Optimal urban networks via mass transportation”, *Lecture Notes in Math.*, **1961**, Springer-Verlag, Berlin, 2009, x+150 pp.
- [177] M. Bernot, V. Caselles, J.-M. Morel, *Optimal transportation networks. Models and theory*, *Lecture Notes in Math.*, **1955**, Springer-Verlag, Berlin, 2009, x+200 pp.
- [178] A. S. Üstünel, M. Zakai, *Transformation of measure on Wiener space*, Springer Monogr. Math., Springer-Verlag, Berlin, 2000, xiv+296 pp.
- [179] В. И. Богачев, А. В. Колесников, К. В. Медведев, “Треугольные преобразования мер”, *Матем. сб.*, **196**:3 (2005), 3–30; англ. пер.: V. I. Bogachev, A. V. Kolesnikov, K. V. Medvedev, “Triangular transformations of measures”, *Sb. Math.*, **196**:3-4 (2005), 309–335.
- [180] D. Feyel, A. S. Üstünel, “Monge–Kantorovitch measure transportation and Monge–Ampère equation on Wiener space”, *Probab. Theory Related Fields*, **128**:3 (2004), 347–385.
- [181] D. Feyel, A. S. Üstünel, “Solution of the Monge–Ampère equation on Wiener space for general log-concave measures”, *J. Funct. Anal.*, **232**:1 (2006), 29–55.
- [182] D. Feyel, A. S. Üstünel, “The notions of convexity and concavity on Wiener space”, *J. Funct. Anal.*, **176**:2 (2000), 400–428.
- [183] A. V. Kolesnikov, “Convexity inequalities and optimal transport of infinite-dimensional measures”, *J. Math. Pures Appl.* (9), **83**:11 (2004), 1373–1404.
- [184] F. Cavalletti, “The Monge problem in Wiener space”, *Calc. Var. Partial Differential Equations*, **45**:1-2 (2012), 101–124; 2011, arXiv:1103.2798v2.
- [185] V. I. Bogachev, A. V. Kolesnikov, “On the Monge–Ampère equation in infinite dimensions”, *Infin. Dimen. Anal. Quantum Probab. Related Top.*, **8**:4 (2005), 547–572.
- [186] S. Fang, J. Shao, K.-Th. Sturm, “Wasserstein space over the Wiener space”, *Probab. Theory Related Fields*, **146**:3-4 (2010), 535–565.
- [187] L. Decreusefond, “Wasserstein distance on configuration space”, *Potential Anal.*, **28**:3 (2008), 283–300.

- [188] В. И. Богачев, А. В. Колесников, “Нелинейные преобразования выпуклых мер”, *Теория вероятн. и ее примен.*, **50**:1 (2005), 27–51; англ. пер.: V. I. Bogachev, A. V. Kolesnikov, “Nonlinear transformations of convex measures”, *Theory Probab. Appl.*, **50**:1 (2006), 34–52.
- [189] S. G. Bobkov, “Large deviations via transference plans”, *Advances in mathematics research*, v. 2, Adv. Math. Res., **2**, Nova Sci. Publ., Hauppauge, NY, 2003, 151–175.
- [190] G. Carlier, A. Galichon, F. Santambrogio, “From Knothe’s transport to Brenier’s map and a continuation method for optimal transport”, *SIAM J. Math. Anal.*, **41**:6 (2009/10), 2554–2576.
- [191] J. Moser, “On the volume elements on a manifold”, *Trans. Amer. Math. Soc.*, **120** (1965), 286–294.
- [192] Y.-H. Kim, E. Milman, “A generalization of Caffarelli’s contraction theorem via (reverse) heat flow”, *Math. Ann.* (to appear); 2010, arXiv:1002.0373.
- [193] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, D. Slepčev, “Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations”, *Duke Math. J.*, **156**:2 (2011), 229–271.
- [194] P. Bernard, “Young measures, superposition and transport”, *Indiana Univ. Math. J.*, **57**:1 (2008), 247–275.
- [195] W. Gangbo, A. Świąch, “Optimal maps for the multidimensional Monge–Kantorovich problem”, *Comm. Pure Appl. Math.*, **51**:1 (1998), 23–45.
- [196] W. Gangbo, H. K. Kim, T. Pacini, *Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems*, Mem. Amer. Math. Soc., **211**, № 993, 2011, vi+77 pp.

В. И. Богачев (V. I. Bogachev)

Московский государственный университет
им. М. В. Ломоносова;
Православный Свято-Тихоновский
государственный университет
E-mail: vibogach@mail.ru

Поступила в редакцию
20.06.2012

А. В. Колесников (A. V. Kolesnikov)

Высшая школа экономики
E-mail: sascha77@mail.ru