

На правах рукописи

Коротин Александр Андреевич

Параметрические методы вычисления оптимальных транспортных отображений, расстояний и барицентров

Специальность 1.2.2 —
«Математическое моделирование, численные методы и комплексы
программ»

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2022

Работа выполнена в Автономной некоммерческой образовательной организации высшего образования «СКОЛКОВСКИЙ ИНСТИТУТ НАУКИ И ТЕХНОЛОГИЙ».

Научный руководитель:

Бурнаев Евгений Владимирович

доктор физико-математических наук, доцент,
АНОО ВО «Сколковский институт науки и
технологий», профессор,
рук. центра прикладного ИИ Сколтеха

Официальные оппоненты:

Орлов Юрий Николаевич

доктор физико-математических наук, доцент,
ФИЦ Институт прикладной математики им.
М. В. Келдыша Российской академии наук,
главный научный сотрудник отдела №6

Ватолин Дмитрий Сергеевич

кандидат физико-математических наук,
ФГБОУ ВО «Московский государственный
университет им. М.В. Ломоносова»,
старший научный сотрудник, зав. лаб.
«Компьютерная графика и мультимедиа»

Ведущая организация:

ФГАОУ ВО «Московский физико-технический
институт (национальный исследовательский
университет)»

Защита состоится «20» февраля 2023 г. в 12 часов 30 минут на заседании диссертационного совета 24.1.224.01 на базе федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН) по адресу: 17312, г. Москва, проспект 60-летия Октября, 9.

С диссертацией можно ознакомиться в библиотеке ФИЦ ИУ РАН по адресу:
г. Москва, ул. Вавилова, д. 40 и на официальном сайте: <http://www.frccsc.ru>.

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, просьба направлять по адресу: 17312, г. Москва, проспект 60-летия Октября, 9, ученому секретарю диссертационного совета 24.1.224.01.

Автореферат разослан «___» _____ 2022 года.

Телефон для справок: +7 (499) 135-51-64.

Ученый секретарь
диссертационного совета
24.1.224.01,
канд. физ.-мат. наук, доцент

Смирнов Иван Валентинович

Общая характеристика работы

Актуальность темы. Основной целью генеративного моделирования является построение математических моделей, описывающих сложную структуру многомерных данных, и дальнейшее использование этих моделей для симуляции новых данных со схожими статистиками или для изменения характеристик исходных данных [1]. За последние десять лет в области генеративного моделирования широкое распространение получили численные методы на основе математической теории **оптимального транспорта** [1; 2], одним из основоположников которой является советский математик и экономист Леонид Витальевич Канторович.

Оптимальный транспорт (OT) — это общее название обширного класса задач нахождения эффективного способа перемещения массы между вероятностными распределениями. В наши дни численные методы OT преимущественно используются для вычисления функций потерь при обучении генеративных моделей на основе глубоких нейросетей для синтеза искусственных данных [1], например, изображений. Другое важное применение численных методов OT — доменная адаптация [3], то есть адаптация модели (например, классификационной) для применения к данным с нового источника. Эта задача крайне важна, например, в медицине, где типичный размер выборок данных не превышает сотни объектов и необходимо комбинировать знания, полученные с нескольких выборок из разных источников. Для решения таких задач используются OT отображения, а также центры масс (OT барицентры) для агрегации выборок [4].

Наиболее популярна задача OT с квадратичной функцией стоимости, определяющая известную Васерштейн-2 метрику. Градиентные потоки функционалов на пространстве распределений, оснащенном этой метрикой, позволяют моделировать решения ряда дифференциальных уравнений, например, уравнения Фоккера-Планка, используемого для описания процессов в физике, финансах, демографии. Моделирование градиентных потоков можно осуществлять, используя методы OT [5].

Несмотря на значительные достижения в вычислительном OT за последнее десятилетие, существующие вычислительные методы решения задач OT по-прежнему обладают рядом критичных недостатков (описываемых далее), ограничивающих их эффективное применение в ряде вышеупомянутых задач моделирования распределений данных.

Большинство методов [6] в области вычислительного OT предназначены для данных, описываемых **дискретными** вероятностными распределениями. Такие методы способны аппроксимировать решения задачи OT для распределений с дискретным носителем из $10^5 - 10^6$ точек. При больших количествах точек проблема неразрешима из-за большой вычислительной сложности. Таким образом, дискретные методы не могут обеспечить точную аппроксимацию решения задачи OT для данных

высокой размерности, имеющих непрерывные распределения, поскольку требуются большие размеры выборок.

В то же время машинное обучение все чаще применяется для решения задач ОТ [2; 7]. Построение математических моделей на основе глубоких нейронных сетей для параметризации транспортных отображений позволяет масштабировать применения ОТ на высокоразмерные пространства и практические задачи, в которых дискретные методы работают неудовлетворительно [8]. В частности, численные методы ОТ на основе глубоких нейронных сетей позволяют получать решения задач ОТ для **непрерывных** распределений. Такие методы обычно называются непрерывными или параметрическими, а процесс вычисления ОТ – обучением модели.

Численные непрерывные методы ОТ условно делятся на две группы: методы на основе энтропийной регуляризации [2] и минимаксные методы [7]. Энтропийные методы из-за регуляризации находят решение задачи ОТ, которое смещено относительно истинного решения, что ведет к большой ошибке решения задачи ОТ и усложняет использование методов на практике. Минимаксные непрерывные методы не имеют этой проблемы, но требуют решения задачи поиска седловой точки некоторого функционала. На практике такая оптимизация может быть неустойчива, что усложняет процесс обучения с использованием этих методов.

Важно отметить, что хотя в области вычислительного ОТ активно ведется разработка новых непрерывных методов, вопрос оценки качества их функционирования все еще остается открытым. Связано это с тем, что существует лишь ограниченное число непрерывных задач ОТ, в которых аналитически известно истинное решение задачи, с которым можно сравнить решение вычисленное непрерывным методом.

Таким образом, существует необходимость разработки новых численных методов, которые позволили бы находить решения непрерывных задач ОТ в больших размерностях без смещения и без решения минимаксной задачи оптимизации. Для оценки точности таких методов требуется разработка количественной методики тестирования непрерывных методов вычисления ОТ.

Основными математическими задачами является аналитический вывод двойственных формулировок задач ОТ, позволяющих осуществлять вычисление транспортных отображений с помощью нейросетей без необходимости использовать энтропийную регуляризацию или решать минимаксную задачу. Важной подзадачей является получение теоретических результатов, связывающих точность нахождения оптимума двойственного функционала (в смысле его значения) с ошибкой восстановленного решения прямой задачи ОТ. Построение надежных и теоретически обоснованных нейросетевых методов вычисления ОТ позволит усовершенствовать применения ОТ к ранее упомянутым практическим задачам.

Целью диссертационной работы является разработка математических моделей на основе нейросетей для решения задач непрерывного ОТ с квадратичной функцией стоимости (называемой Васерштейн-2 метрикой): вычисление ОТ отображения и расстояния между непрерывными вероятностными распределениями, нахождение барицентров семейств вероятностных распределений и вычисление градиентных потоков функционалов на пространстве вероятностных распределений. Для достижения цели в работе были поставлены и решены следующие **задачи**:

1. Разработать численный метод на основе двойственной формулировки ОТ для поиска Васерштейн-2 оптимальных отображений между непрерывными распределениями с помощью нейросетей; получить теоретические оценки того, как связаны ошибки аппроксимации решений предложенной двойственной и исходной задач.
2. Разработать численный метод на основе двойственной формулировки ОТ для поиска Васерштейн-2 барицентров семейства непрерывных распределений с помощью нейронных сетей; получить теоретические оценки того, как связаны ошибки аппроксимации решений предложенной двойственной и исходной задач.
3. Разработать численный метод на основе дискретизации времени и нейронных сетей для моделирования Васерштейн-2 градиентных потоков функционалов на пространстве распределений.
4. Разработать методологию для количественного сравнения непрерывных методов ОТ; предложить и математически обосновать метод создания эталонных пар непрерывных распределений с аналитически известным Васерштейн-2 ОТ отображением.

Научная новизна:

1. Впервые предложен масштабируемый метод вычисления Васерштейн-2 оптимальных отображений между непрерывными распределениями в пространствах большой размерности, который состоит из одного этапа обучения и не требует минимаксной оптимизации. В основе метода лежит новый подход на основе введения нового циклического регуляризатора в двойственную постановку задачи ОТ и использования нейросетей с выпуклой по входу архитектурой. В отличие от существующих регуляризаторов циклический регуляризатор не вносит смещение в решение задачи.
2. Впервые предложен масштабируемый метод вычисления Васерштейн-2 барицентров непрерывных вероятностных распределений, который состоит из одного этапа обучения и не требует минимаксной оптимизации. В основе подхода лежит комбинация циклического регуляризатора (см. выше), нового регуляризатора на конгруэнтность решений двойственной задачи и выпуклых по входу нейронных сетей. В отличие от существующих регуляризаторов, предложенная комбинация не вносит смещение в решение.

3. Впервые предложен масштабируемый численный метод для моделирования Васерштейн-2 градиентных потоков на пространстве высокоразмерных вероятностных распределений. Метод является практической реализацией теоретической схемыJKO [5], для которой ранее не были известны эффективные численные методы для её реализации в пространствах высокой размерности.
4. Предложена оригинальная методология синтеза пар многомерных непрерывных вероятностных распределений, для которых аналитически известно эталонное Васерштейн-2 транспортное отображение между ними. Разработанная методология и предложенные пары распределений позволяют сравнивать непрерывные численные методы ОТ в больших размерностях и устраниют недостатки существующих методик, связанные с узостью классов используемых тестовых распределений и не позволяющие объективно оценивать качество численных методов ОТ.

Теоретическая значимость работы заключается в том, что предложенные новые подходы на основе нейронных сетей и численных методов ОТ позволяют эффективно решать ряд задач математического моделирования, избавляясь от недостатков существующих подходов таких как использование численно нестабильной минимаксной оптимизации или наличие систематических ошибок в решении. Принципиальной особенностью предложенных подходов является то, что разработанные в работе решения для не требуют минимаксной оптимизации и не смешают оптимальное решение задачи, что приводит к более вычислительно эффективному и точному нахождению решения по сравнению с существующими численными методами ОТ на основе нейронных сетей. В частности, эти улучшения значительно упрощают теоретический анализ полученных решений, и, как следствие, позволяют выводить оценки ошибки оптимизации по ограниченному классу функций, например, представленному нейросетями.

Практическая значимость. Предложенные подходы математического моделирования на основе нейронных сетей и численных методов ОТ имеют потенциал для решения практически важных научно-исследовательских и промышленных задач. Например, разработанный метод, протестированный в задачах доменной адаптации и переноса стиля на искусственных и модельных данных, имеет ряд потенциальных приложений в задачах компьютерного зрения, таких как обучение моделей для улучшения разрешения изображений, устранения шума и пр. Другим возможным применением разработанных подходов является агрегация данных из нескольких источников, например, медицинских сканеров разных производителей, что позволит улучшить качество моделей классификации и сегментации медицинских изображений, построенных на них.

Методология и методы исследования. Для достижения поставленных в диссертационной работе целей используются:

1. Подходы теории оптимального транспорта для аналитического вывода двойственных формулировок задач ОТ и соотношений между прямыми/двойственными переменными, а также для построения методов регуляризации двойственных задач, позволяющих эффективно находить их решение.
2. Численные методы стохастической оптимизации для построения практически эффективных алгоритмов нахождения решений рассматриваемых задач ОТ с помощью нейронных сетей.
3. Подходы глубинного обучения для построения нейронных сетей с выпуклой по входу архитектурой для аппроксимации минимизаторов разработанных целевых функций.
4. Методы выпуклого анализа, линейной алгебры, математического и функционального анализа, теории вероятностей и теории меры для теоретического обоснования предложенных целевых функций оптимизации, анализа соотношений решений прямых и предложенных регуляризованных двойственных задач.
5. Способы количественной и качественной оценки результатов экспериментов, используемых в области вычислительного оптимального транспорта и генеративного машинного обучения.
6. Подходы объектно-ориентированного программирования с использованием языка программирования Python и фреймворка Pytorch глубинного обучения и численной оптимизации для практической реализации и тестирования предложенных методов.

Основные положения, выносимые на защиту:

1. Предложен алгоритм для вычисления Васерштейн-2 оптимального отображения и расстояния между непрерывными распределениями, базирующийся на нейросетях с выпуклой по входу архитектурой и применении разработанной циклической регуляризации в двойственной задаче. Получены теоретические оценки, связывающие ошибку минимизации разработанного целевого функционала с ошибкой решения исходной задачи нахождения ОТ отображения.
2. Предложен алгоритм для вычисления Васерштейн-2 барицентра семейства непрерывных распределений, базирующийся на нейросетях с выпуклой по входу архитектурой и использовании разработанных циклической регуляризации и регуляризации на конгруэнтность двойственных переменных задачи поиска барицентра. Получены теоретические оценки, связывающие ошибку минимизации разработанного целевого функционала с ошибкой решения исходной задачи нахождения барицентра.
3. Предложен численный метод моделирования Васерштейн-2 градиентных потоков функционалов на пространствах вероятностных распределений, основанный на дискретизации градиентного потока по времени и на использовании выпуклых по входу нейросетей.

4. Разработана методология на основе нейронных сетей с выпуклой по входу архитектурой для синтеза пар непрерывных распределений с аналитически известным ОТ отображением между ними. Данная методология позволяет устраниить имеющийся пробел в тестировании непрерывных методов решения задачи ОТ и исследования математических моделей, на которых они основаны.

Полученные результаты соответствуют следующим пунктам паспорта специальности 1.2.2 (математическое моделирование, численные методы и комплексы программ):

1. Результаты 1-3 соответствуют **п.1** «Разработка новых математических методов моделирования объектов и явлений», **п.3** «Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий» и **п.4** «Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента».
2. Результат 4 соотв. **п.2** «Развитие качественных и приближенных аналитических методов исследования математических моделей».

Достоверность полученных результатов обеспечивается корректностью применения апробированного в научной практике математического аппарата теории оптимального транспорта и других смежных теоретических областей, а также экспериментальной проверкой разработанных численных методов на большом количестве модельных и практических задач оптимального транспорта. Полученные теоретические результаты обосновываются математически строгими доказательствами, а для проведенных вычислительных экспериментов даются детальные описания, обеспечивающие их воспроизводимость. Результаты опубликованы в трудах ведущих рецензируемых международных конференций по машинному обучению и искусственному интеллекту.

Публикации. Список публикаций приведен в конце авторефера. Результаты диссертации изложены в 5 работах. Все работы опубликованы в трудах **Core rank A*** ведущих международных конференций по машинному обучению.

Апробация работы. Результаты работы доложены на **3** ведущих международных научных конференциях ранга **A*** (суммарно **5** докладов) по машинному обучению и искусственному интеллекту:

1. The 9th International Conference on Learning Representations (ICLR 2021, **Core A***, онлайн) – **2** презентации по [[A3](#), [A4](#)];
2. The 35th Conference on Neural Information Processing Systems (NeurIPS 2021, **Core A***, онлайн) – **2** презентации по [[A1](#), [A2](#)];
3. The 10th International Conference on Learning Representations (ICLR 2022, **Core A***, онлайн) – **1** презентация по [[A5](#)];

Результаты работы [A3] были представлены на 5 следующих научных семинарах по машинному обучению и искусственному интеллекту:

- Huawei Machine Learning Workshop (2019, Сочи, Россия);
- Семинар по байесовским методам (2020, НИУ ВШЭ, онлайн);
- Geometry Data Processing Group Seminar (2020, MIT, онлайн);
- SMILES Machine Learning Summer School (2020, онлайн);
- Math of Machine Learning Summer School (2020, Сочи, Россия).

За научные результаты, включая результаты диссертации, автор в 2019 и 2021 годах становился лауреатом премии им. Ильи Сегаловича для поддержки молодых исследователей в области машинного обучения и искусственного интеллекта от компании Яндекс.

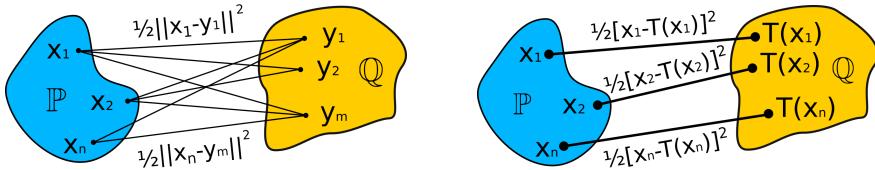
Личный вклад. Все положения, выносимые на защиту, получены автором лично. Все алгоритмы и теоретические результаты, приведенные в диссертации и работах [A1-A5], сформулированы и обоснованы непосредственно автором диссертации. В диссертации использованы результаты совместных исследований: экспериментальное тестирование метода для нахождения ОТ отображения [A3] проведено совместно с соавторами В. Егиазаряном, А. Асадулаевым и А. Сафирным; тестирование метода для вычисления барицентров [A4] проведено совместно с соавтором Л. Ли; автор диссертации осуществил предварительную реализацию и тестирование разработанного метода для вычисления градиентных потоков [A2], адаптация кода для рассматриваемых в экспериментальной части практических задач сравнения с альтернативными подходами осуществлена соавтором П. Мокровым; все вычислительные эксперименты в работе [A1] осуществлены непосредственно автором диссертации, а в работе [A5] – соавтором Л. Роут.

Объем и структура работы. Диссертация состоит из введения, 5 глав и заключения. Полный объём диссертации составляет 148 страниц, включая 30 рисунков и 17 таблиц. Список литературы содержит 138 наименований.

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, формулируется цель, ставятся задачи работы, излагается научная новизна, теоретическая и практическая значимость представляемой работы.

В **первой главе** приводится общее описание 4 основных задач, решаемых в работе: вычисление Васерштейн-2 оптимальных транспортных отображений и расстояний (**задача 1**), вычисление Васерштейн-2 барицентров (**задача 2**), вычисление Васерштейн-2 градиентных потоков функционалов (**задача 3**) и построение эталонных пар вероятностных распределений для Васерштейн-2 оптимального транспорта (**задача 4**). Ниже мы приводим математические постановки вышеописанных задач, а также кратко приводим недостатки существующих подходов к их решению.



а) Постановка Канторовича.

б) Постановка Монже.

Рис. 1 — Схематически изображенная задача 1 оптимального транспорта между распределениями \mathbb{P}, \mathbb{Q} с квадратичной функцией стоимости.

Через $(\mathbb{R}^D, \|\cdot\|)$ обозначим D -мерное евклидово пространство с евклидовой метрикой $\|\cdot\|$. Через $\mathcal{P}_2(\mathbb{R}^D)$ обозначим множество всех борелевских вероятностных распределений на \mathbb{R}^D с конечным вторым моментом. Также через $\mathcal{P}_{2,\text{ac}}(\mathbb{R}^D) \subset \mathcal{P}_2(\mathbb{R}^D)$ обозначим подмножество всех абсолютно непрерывных распределений (относительно меры Лебега). Образ распределения $\mathbb{P} \in \mathcal{P}_2(\mathbb{R}^D)$ под действием измеримого отображения $T : \mathbb{R}^D \rightarrow \mathbb{R}^D$ обозначим через $T\sharp\mathbb{P}$.

Задача 1 (Вычисление оптимального транспорта с квадратичной ценой). Для пары вероятностных распределений $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^D)$ задача оптимального транспорта с квадратичной ставится следующим образом:

$$\mathbb{W}_2^2(\mathbb{P}, \mathbb{Q}) \stackrel{\text{def}}{=} \min_{\pi \in \Pi(\mathbb{P}, \mathbb{Q})} \int_{\mathbb{R}^D \times \mathbb{R}^D} \frac{\|x - y\|^2}{2} d\pi(x, y), \quad (1)$$

где $\Pi(\mathbb{P}, \mathbb{Q})$ – множество транспортных планов, то есть вероятностных распределений на $\mathbb{R}^D \times \mathbb{R}^D$, чьи маргиналы суть \mathbb{P}, \mathbb{Q} соответственно.

Решение задачи (1) предполагает нахождение оптимального плана $\pi^* \in \Pi(\mathbb{P}, \mathbb{Q})$, доставляющего минимум. Соответствующее минимальное значение называется оптимальной транспортной стоимостью, а её корень называется расстоянием Васерштейн-2 и обозначается через \mathbb{W}_2 .

Формулировка (1) задачи оптимального транспорта часто называется формулировкой **Канторовича** задачи ОТ (Рис. 1a) и представляет собой оптимизацию линейного функционала от π (интеграла по π) на пространстве вероятностных мер на $\mathbb{R}^D \times \mathbb{R}^D$ в линейных ограничениях $\pi \in \Pi(\mathbb{P}, \mathbb{Q})$. Однако аналитические решения (1) для пар \mathbb{P}, \mathbb{Q} общего вида неизвестны, а решать данную задачу оптимизации на практике вычислительно сложно.

В случае, когда \mathbb{P}, \mathbb{Q} – **дискретные** вероятностные распределения с конечным носителем, задача обычно решается численными методами линейного программирования или же используются дополнительные регуляризаторы, приводящие к итеративным матричным методам [6]. К сожалению, когда \mathbb{P} и \mathbb{Q} **непрерывные** распределения (доступные через

конечные или бесконечные случайные выборки), вышеупомянутые дискретные методы не дают достаточного качества, так как необходимые точные дискретизации распределений вычислительно недостижимы.

Далее мы кратко рассмотрим основные существующие методы решения задачи (1) в случае непрерывных \mathbb{P}, \mathbb{Q} и обозначим их ключевые недостатки, которые устраняются в данной диссертации. Для начала отметим, что если $\mathbb{P} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$, то ОТ план π единственным образом определяется в виде $\pi^* = [\text{id}_{\mathbb{R}^D}, T^*] \sharp \mathbb{P}$, где $T^* : \mathbb{R}^D \rightarrow \mathbb{R}^D$ – единственный минимизатор задачи ОТ в форме Монжа (Рис. 1б) [9]:

$$\mathbb{W}_2^2(\mathbb{P}, \mathbb{Q}) = \min_{T \sharp \mathbb{P} = \mathbb{Q}} \int_{\mathbb{R}^D} \frac{\|x - T(x)\|^2}{2} d\mathbb{P}(x), \quad (2)$$

где минимум берется по измеримым отображениям $T : \mathbb{R}^D \rightarrow \mathbb{R}^D$, удовлетворяющим условию переноса распределения $T \sharp \mathbb{P} = \mathbb{Q}$. Таким образом, в непрерывном случае задача (1) сводится к *нахождению оптимального транспортного отображения* $T^* : \mathbb{R}^D \rightarrow \mathbb{R}^D$.

Существующие непрерывные методы основываются на решении двойственной задачи к (1) и (2) и дальнейшем восстановлении оптимального транспортного отображения T^* (решения прямой задачи) из найденных двойственных переменных. Для $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^D)$ **двойственная формулировка** задачи квадратичного ОТ имеет следующий вид (см. [9]):

$$\mathbb{W}_2^2(\mathbb{P}, \mathbb{Q}) = \max_{f \oplus g \leq \frac{1}{2}\|\cdot\|^2} \left[\int_{\mathbb{R}^D} f(x) d\mathbb{P}(x) + \int_{\mathbb{R}^D} g(y) d\mathbb{Q}(y) \right], \quad (3)$$

где максимум берется по всем \mathbb{P} -интегрируемым функциям $f : \mathbb{R}^D \rightarrow \mathbb{R} \cup \{\infty\}$ и \mathbb{Q} -интегрируемым $g : \mathbb{R}^D \rightarrow \mathbb{R} \cup \{\infty\}$ (потенциалам), для которых выполнено $f(x) + g(y) \leq \frac{1}{2}\|x - y\|^2$ при всех $x, y \in \mathbb{R}^D$. Существуют альтернативные к (3) двойственные формы, например,

$$\mathbb{W}_2^2(\mathbb{P}, \mathbb{Q}) = \text{Const}(\mathbb{P}, \mathbb{Q}) - \min_{\psi \in \text{Convex}} \overbrace{\left[\int_{\mathbb{R}^D} \psi(x) d\mathbb{P}(x) + \int_{\mathbb{R}^D} \bar{\psi}(y) d\mathbb{Q}(y) \right]}^{\text{Corr}(\mathbb{P}, \mathbb{Q} | \psi)}, \quad (4)$$

где минимум берется по выпуклым потенциалам ψ . При этом $\bar{\psi}(y) \stackrel{\text{def}}{=} \max_{x \in \mathbb{R}^D} [\langle x, y \rangle - \psi(x)]$ – сопряженная функция к ψ , которая тоже является выпуклой функцией. Второе слагаемое $\text{Corr}(\mathbb{P}, \mathbb{Q} | \psi)$ в (4) называется корреляцией, а функции $\psi, \bar{\psi}$ – потенциалами Бренье.

В случае, когда $\mathbb{P} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$, \mathbb{P} -почти всюду выполнено $T^*(x) = \nabla \psi^*(x)$, то есть градиент оптимального потенциала Бренье является оптимальным транспортным отображением из \mathbb{P} в \mathbb{Q} . Этот факт позволяет для решения задач ОТ (1), (2) фокусироваться лишь на решении двойственной задачи, которая представляется более простой, так в задаче устраняется

нетривиальное граниченое условие $\pi \in \Pi(\mathbb{P}, \mathbb{Q})$ в (1) или $T \# \mathbb{P} = \mathbb{Q}$ в (2). Однако решение (4), то есть нахождение двойственных потенциалов, все еще нетривиально, т.к. задача представляет собой оптимизацию по пространству выпуклых функций ψ . При этом вычисление сопряженной функции $\bar{\psi}$ представляет собой внутреннюю задачу оптимизации.

В работе [7] авторы предлагают аппроксимировать ψ^* , $\bar{\psi}^*$ с помощью выпуклых по входу нейронных сетей (англ. input-convex neural networks, сокращенно – ICNNs) и свести задачу к оптимизации параметров нейронных сетей. Однако для использования предложенных методов предполагается решение минимаксной задачи поиска седловой точки некоторого функционала. Полученная задача представляется сложной на практике из-за численной нестабильности минимаксной оптимизации.

В работе [2] авторы рассматривают классическую двойственную форму (3) вместо (4) и ослабляют условие $f \oplus g \leq c$ с помощью наложения регуляризации (энтропийной и пр.), которая мягко штрафует потенциалы f , g за нарушение этого условия. Из-за регуляризации решение задачи (оптимальный f^*) смещается относительно истинного решения, что не позволяет из него точно находить решение исходной задачи ОТ (2).

В данной диссертации **предлагается численный метод** (глава 2 работы), который устраняет недостатки вышеописанных непрерывных методов, а именно: (1) оптимизируется неманиакский критерий, (2) решением которого является истинный несмешенный потенциал ψ^* , градиент $\nabla \psi^*$ которого является искомым ОТ отображением T^* .

Задача 2 (Задача поиска Васерштейн-2 барицентра). *Рассмотрим распределения $\mathbb{P}_1, \dots, \mathbb{P}_N \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$. Тогда их барицентр $\bar{\mathbb{P}}$ относительно весов $\alpha_1, \dots, \alpha_N$ (при $\alpha_n > 0$ и $\sum_{n=1}^N \alpha_n = 1$) определяется через (рис. 2а)*

$$\bar{\mathbb{P}} \stackrel{\text{def}}{=} \arg \min_{\mathbb{P} \in \mathcal{P}_2(\mathbb{R}^D)} \mathcal{F}_{\text{Bar}}(\mathbb{P}) \stackrel{\text{def}}{=} \arg \min_{\mathbb{P} \in \mathcal{P}_2(\mathbb{R}^D)} \sum_{n=1}^N \alpha_n \mathbb{W}_2^2(\mathbb{P}_n, \mathbb{P}). \quad (5)$$

Здесь \mathcal{F}_{Bar} называется барицентрическим функционалом, и его минимальное значение обозначается через $\mathcal{F}_{\text{Bar}}^* \stackrel{\text{def}}{=} \mathcal{F}_{\text{Bar}}(\bar{\mathbb{P}})$.

Проблемы существующих методов вычисления барицентров аналогичны проблемам методов решения задачи ОТ (2). В случае, когда $\mathbb{P}_1, \dots, \mathbb{P}_n$ являются многомерными непрерывными распределениями, дискретные методы не дают достаточного качества, так как необходимые точные дискретизации распределений вычислительно недостижимы. Непрерывные же методы для вычисления барицентров либо требуют решения сложной $\min \max \min$ задачи оптимизации [10], либо используют энтропийную регуляризацию [8], ведущую к смещению решения.

В данной диссертации **предлагается численный метод** (глава 3 работы), который лишен недостатков вышеописанных методов, а именно:

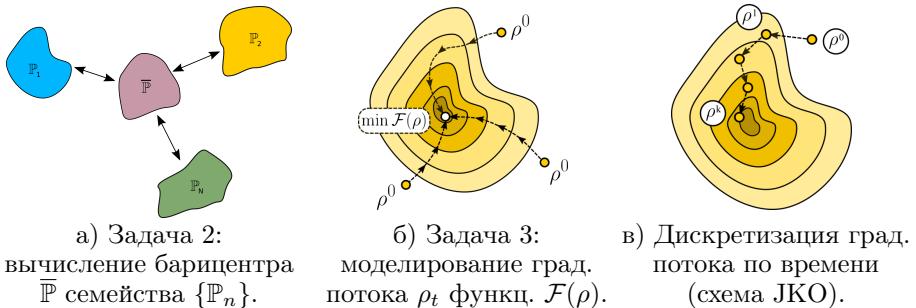


Рис. 2 – Схематически изображенные задачи 2, 3 и схема JKO [5].

(1) оптимизируется неманиакский критерий, (2) решение которого позволяет восстановить истинный несмещенный барицентр.

Задача 3 (Вычисление Васерштейн-2 градиентного потока функционала на пространстве распределений). Рассмотрим функционал на пространстве вероятностных распределений $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^D) \rightarrow \mathbb{R}$. Кривая, состоящая из распределений $\{\rho_t\}_{t \in \mathbb{R}_+}$ и являющаяся решением уравнения

$$\frac{\partial \rho_t}{\partial t} = \operatorname{div}(\rho_t \nabla_x \mathcal{F}'(\rho_t)), \quad s.t. \quad \rho_0 = \rho^0, \quad (6)$$

называется Васерштейн-2 градиентным потоком функционала \mathcal{F} , стартующим из точки $\rho^0 \in \mathcal{P}_2(\mathbb{R}^D)$, см. рис. 2б. Здесь $\mathcal{F}'(\rho_t) : \mathbb{R}^D \rightarrow \mathbb{R}$ обозначает первую вариацию функционала \mathcal{F} в точке ρ_t [9].

Научно-практический интерес в вычислении градиентных потоков заключается в том, что многие диффузионные процессы, возникающие на практике, являются Васерштейн-2 градиентными потоками определенных функционалов. В данной диссертационной работе основной объект исследования представляют диффузии, являющиеся градиентными потоками функционала \mathcal{F}_{FP} свободной энергии Фоккера-Планка ($\beta > 0$)

$$\mathcal{F}_{\text{FP}}(\rho) \stackrel{def}{=} \int_{\mathbb{R}^D} \Phi(x) d\rho(x) + \beta^{-1} \int_{\mathbb{R}^D} \log \frac{d\rho}{dx}(x) d\rho(x), \quad (7)$$

являющегося разностью потенциальной энергии и энтропии. Соответствующий градиентный поток удовлетворяет $\frac{\partial \rho_t}{\partial t} = \operatorname{div}(\nabla \Phi(x) \rho_t) + \beta^{-1} \Delta \rho_t$ при $\rho_0 = \rho^0$. Отметим, что при $\beta = \infty$ выражение является классическим уравнением переноса, а в случае $\Phi(x) \equiv 0$ – уравнением теплопроводности.

В общем случае задача вычисления градиентных потоков Васерштейна-2 является сложной. Обычно аналитическое решение получить не удается, и поэтому приходится применять методы численной аппроксимации. Джордан, Киндерлерер и Отто предложили подход (позже

получивший сокращенное название “*JKO-интегрирование*”) для аппроксимации динамики ρ_t в (6), см. [5]. В этом методе осуществляется дискретное по времени итеративное уточнение непрерывного потока (рис. 2в):

$$\rho^{(k)} \leftarrow \arg \min_{\rho \in \mathcal{P}_2(\mathbb{R}^n)} \left[\mathcal{F}(\rho) + \frac{1}{h} \mathbb{W}_2^2(\rho^{(k-1)}, \rho) \right], \quad (8)$$

где $\rho^{(0)} = \rho^0$ – начальное условие и $h > 0$ – величина шага дискретизации по времени. Дискретный временной градиентный поток сходится к непрерывному потоку при $h \rightarrow 0$, т.е. $\rho^{(k)} \approx \rho_{kh}$.

Стоит отметить, что выполнение JKO-итераций является сложной задачей из-за наличия минимизации по \mathbb{W}_2^2 . Типичный подход к выполнению JKO-шага основан на дискретизации пространственного домена. При размере носителя $\lesssim 10^6$ задача (8) может быть решена стандартными дискретными алгоритмами ОТ [6], упомянутыми в предыдущих разделах. В размерностях $D \geq 3$ дискретные распределения обычно неточно аппроксимируют непрерывные распределения и, как следствие, динамику градиентных потоков. Чтобы обойти эту трудность, в [11] предложен параметрический метод для аппроксимации ρ_t на основе энтропийного ОТ, который приводит к смещению найденного решения.

В данной диссертации **предлагается численный метод** (глава 4 работы) для моделирования несмещенной динамики JKO с помощью замены оптимизации (8) по вероятностным распределениям на эквивалентную, но практически более обозримую оптимизацию по выпуклым функциям.

Задача 4 (Построение эталонных пар непрерывных распределений для Васерштейн-2 оптимального транспорта). *Построить пары непрерывных вероятностных распределений (\mathbb{P}, \mathbb{Q}) , для которых оптимальное решение T^* задачи 1 (ОТ отображение) аналитически известно. Подразумевается, что построенные распределения \mathbb{P}, \mathbb{Q} на практике должны быть доступны через случайные независимые выборки.*

Эталонные пары могут быть использованы для количественного тестирования численных методов для решения задачи 1 поиска ОТ плана или отображения. На данный момент возможности такого тестирования значительно ограничены. Этот пробел в научной области вычислительного ОТ не позволяет объективно оценить качество существующих численных методов для решения задач ОТ. В литературе для оценки качества существующих непрерывных методов используются два типа метрик.

В явных метриках вычисленное транспортное отображение \hat{T} сравнивается с истинным T^* . В настоящее время имеется сравнительно немного явных метрик, поскольку число примеров \mathbb{P}, \mathbb{Q} с известным эталоном T^* (т.е. решений задачи 4) достаточно мало: T^* может быть получено в дискретном, 1-мерном и линейно-разбросанном случаях.

Неявные метрики применяют численный метод ОТ как компоненту в многоэтапной схеме обучения. При этом качество конечного решения задачи используется как косвенная оценка качества самого численного метода. Например, в генеративном моделировании, при котором ОТ используется как функция потери нейронной сети генератора, качество генератора оценивают через популярные метрики генеративного моделирования [1]. Такие метрики не дают четкого понимания о качестве самого численного метода, так как зависят от компонент модели, не связанных с ОТ.

Таким образом, возникает необходимость разработки нетривиальных пар непрерывных вероятностных распределений (\mathbb{P}, \mathbb{Q}) , которые могли бы быть использованы для **явной** оценки качества численных методов ОТ, то есть имели бы аналитически известное ОТ отображение. В данной диссертационной работе **предлагается метод** (глава 5 работы) построения эталонных непрерывных пар \mathbb{P}, \mathbb{Q} с известным T^* .

Далее обозреваются основные результаты глав 2-5. Каждая глава посвящена решению конкретной задачи ОТ 1, 2, 3, 4.

Во второй главе решается задача поиска Васерштейн-2 ОТ отображения между двумя непрерывными вероятностными распределениями (задача 1). Предлагается и теоретически обосновывается численный метод для решения рассматриваемой задачи на основе выпуклых по входу нейронных сетей и добавления циклической регуляризации к двойственной задаче. Ниже детализируются результаты главы.

Для нахождения оптимального выпуклого потенциала ψ^* в (4), градиент $\nabla \psi^*$ которого является ОТ отображением, предлагается рассматривать следующую неминимаксную задачу оптимизации:

$$\min_{\phi, \psi \in \text{Convex}} \text{Corr}(\mathbb{P}, \mathbb{Q} | \phi, \psi; \lambda) \stackrel{\text{def}}{=} \min_{\phi, \psi \in \text{Convex}} \left[\frac{\lambda}{2} \mathcal{R}_2^{\mathbb{Q}}(\psi, \phi) + \left(\int_{\mathbb{R}^D} \psi(x) d\mathbb{P}(x) + \int_{\mathbb{R}^D} [\langle \nabla \phi(y), y \rangle - \psi(\nabla \phi(y))] d\mathbb{Q}(y) \right) \right], \quad (9)$$

где оптимизация ведется по выпуклым функциям $\psi, \phi : \mathbb{R}^D \rightarrow \mathbb{R}$, а $\mathcal{R}_2^{\mathbb{Q}}(\psi, \phi)$ – специально разработанный циклический регуляризатор (с весом $\lambda > 0$):

$$\mathcal{R}_2^{\mathbb{Q}}(\psi, \phi) \stackrel{\text{def}}{=} \int_{\mathbb{R}^D} \|\nabla \psi \circ \nabla \phi(y) - y\|^2 d\mathbb{Q}(y). \quad (10)$$

Регуляризатор штрафует потенциалы ϕ, ψ за нарушение условия циклической состоятельности (взаимнообратности) их градиентов $\nabla \phi, \nabla \psi$. Отметим, что при определенных предположениях о выпуклых функциях ϕ, ψ взаимнообратность их градиентов $\nabla \phi, \nabla \psi$ соответствует сопряженности с точностью до добавочной константы.

Ключевой особенностью предложенной задачи оптимизации и регуляризатора $\mathcal{R}_2^{\mathbb{Q}}$ является то, что при правильном подборе параметра $\lambda > 0$

оптимальным решением задачи является действительная оптимальная пара потенциалов Бренье $(\psi^*, \phi^*) = (\psi^*, \bar{\psi}^*)$, которая решает задачу (4). В диссертационной работе доказывается теорема 12 на связь ошибки решения задачи (9) с помощью данных выпуклых потенциалов $(\hat{\phi}, \hat{\psi})$ и ошибки восстановленного решения прямой задачи нахождения ОТ отображения (2) через $\hat{\psi}$ с использованием отображения $\hat{T} = \nabla \hat{\psi}$.

Теорема 1 (транспортное свойство для приближения регуляризованных корреляций). Пусть $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$. Пусть $\psi^* : \mathbb{R}^D \rightarrow \mathbb{R}$ – оптимальный выпуклый потенциал Бренье:

$$\psi^* = \arg \min_{\psi \in \text{Convex}} \text{Corr}(\mathbb{P}, \mathbb{Q} | \psi) = \arg \min_{\psi \in \text{Convex}} \left[\int_{\mathbb{R}^D} \psi(x) d\mathbb{P}(x) + \int_{\mathbb{R}^D} \bar{\psi}(y) d\mathbb{Q}(y) \right]. \quad (11)$$

Пусть дифференцируемые выпуклые функции $\hat{\psi} : \mathbb{R}^D \rightarrow \mathbb{R}$ и $\hat{\phi} : \mathbb{R}^D \rightarrow \mathbb{R}$ таковы, что при некотором $\epsilon \in \mathbb{R}$ выполнено

$$\text{Corr}(\mathbb{P}, \mathbb{Q} | \hat{\psi}, \hat{\phi}; \lambda) \leq \text{Corr}(\mathbb{P}, \mathbb{Q} | \psi^*) + \epsilon. \quad (12)$$

Предположим, что $\hat{\psi}$ является β -сильно выпуклой ($\beta > \frac{1}{\lambda} > 0$) и \mathcal{B} -гладкой ($\mathcal{B} \geq \beta$). Предположим также, что $\hat{\phi}$ имеет биективный градиент $\nabla \hat{\phi}$. Тогда имеют место следующие неравенства:

1. Оценка сверху для корреляции:

$$\text{Corr}(\mathbb{P}, \mathbb{Q} | \hat{\psi}, \hat{\phi}; \lambda) \geq \text{Corr}(\mathbb{P}, \mathbb{Q} | \psi^*) \quad (\text{m.e. } \epsilon \geq 0);$$

2. Прямое транспортное свойство:

$$\frac{1}{2} \int_{\mathbb{R}^D} \|\nabla \hat{\psi}(x) - \nabla \psi^*(x)\|^2 d\mathbb{P}(x) \leq \frac{(\mathcal{B})^2 \cdot \epsilon}{\lambda \beta - 1} \cdot \left[\frac{1}{\sqrt{\beta}} + \sqrt{\lambda} \right]^2 = O(\epsilon);$$

3. Обратное транспортное свойство:

$$\frac{1}{2} \int_{\mathbb{R}^D} \|\nabla \hat{\phi}(y) - \nabla \bar{\psi}^*(y)\|^2 d\mathbb{Q}(y) \leq \frac{\epsilon}{\beta - \frac{1}{\lambda}} = O(\epsilon).$$

Теорема утверждает, что чем лучше мы аппроксимируем корреляции (9) при помощи потенциалов $\hat{\psi}, \hat{\phi}$, тем ближе градиенты этих потенциалов к прямому $\nabla \psi^*$ (и обратному $\nabla \bar{\psi}^* = (\nabla \psi^*)^{-1}$) транспортному отображению. При этом несложно видеть, что оптимальные потенциалы Бренье $(\psi^*, \phi^*) = (\psi^*, \bar{\psi}^*)$, решающие (4), доставляют минимум и в задаче (9). В следующей теореме демонстрируется, что в действительности корреляции можно аппроксимировать требуемым образом, если аппроксимирующий класс функций для потенциалов является достаточно большим.

Теорема 2 (аппроксимируемость корреляций). Пусть $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$, а $\psi^* : \mathbb{R}^D \rightarrow \mathbb{R}$ – оптимальный выпуклый потенциал. Пусть Ψ, Φ – классы дифференцируемых выпуклых функций $\mathbb{R}^D \rightarrow \mathbb{R}$. Предположим, что

1. $\exists \psi_1 \in \Psi$ с ϵ_1 -близким градиентом прямого отображения $\nabla\psi^*$, то есть $\int_{\mathbb{R}^D} \|\nabla\psi_1(x) - \nabla\psi^*(x)\|^2 d\mathbb{P}(x) \leq \epsilon_1$, где ψ_1 является \mathcal{B} -гладкой;
2. $\exists \phi_2 \in \Phi$ с ϵ_2 -близким градиентом к обратному отображению $\nabla\psi^*$, то есть $\int_{\mathbb{R}^D} \|\nabla\phi_2(y) - \nabla\psi^*(y)\|^2 d\mathbb{Q}(y) \leq \epsilon_2$.

Пусть $(\hat{\psi}, \hat{\phi}) = \arg \min_{\psi \in \Psi, \phi \in \Phi} \text{Corr}(\mathbb{P}, \mathbb{Q} | \psi, \phi; \lambda)$ – минимизаторы регуляризованных корреляций в $\Psi \times \Phi$. Тогда выполняется неравенство:

$$\begin{aligned} & \text{Corr}(\mathbb{P}, \mathbb{Q} | \hat{\psi}, \hat{\phi}; \lambda) \leq \\ & \text{Corr}(\mathbb{P}, \mathbb{Q} | \psi^*) + \left[\frac{\lambda}{2} (\mathcal{B}\sqrt{\epsilon_2} + \sqrt{\epsilon_1})^2 + (\mathcal{B}\sqrt{\epsilon_2} + \sqrt{\epsilon_1}) \cdot \sqrt{\epsilon_2} + \frac{\mathcal{B}}{2} \epsilon_2 \right], \end{aligned} \quad (13)$$

т.е. регуляризованные корреляции не большие истинных плюс $O(\epsilon_1 + \epsilon_2)$.

Применяя теоремы 2 и 1, мы заключаем, что градиенты $\nabla\hat{\psi}, \nabla\hat{\phi}$ приближенных решений $\hat{\psi}, \hat{\phi}$ являются $O(\epsilon_1 + \epsilon_2)$ -близкими (в смысле \mathcal{L}^2) к $\nabla\psi^*$ и $\nabla\overline{\psi^*}$ соответственно. Фактически, полученные теоретические оценки показывают, что точность решения предложенной неминимаксной задачи (9) с точки зрения значения целевого функционала пропорциональна точности восстановления решения основной прямой задачи Монжа (2), то есть нахождения прямого $\nabla\psi^*$ и обратного $\nabla\overline{\psi^*}$ ОТ отображений.

В работе предлагается **вычислительный алгоритм** для минимизации разработанного функционала (2) на основе методов глубинного обучения. В качестве классов выпуклых функций Ψ, Φ , по которым ведется минимизация, используются выпуклые по входу нейронные сети ICNN ψ_θ, ϕ_ω [12] с параметрами (весами) θ, ω . Для оптимизации их параметров θ, ω используется метод стохастического градиентного спуска с использованием случайных выборок из входных распределений \mathbb{P}, \mathbb{Q} .

Для эмпирической проверки работоспособности предложенного алгоритма оптимизации функционала (9) с помощью нейронных сетей проводится ряд вычислительных экспериментов. Приводятся как качественная, так и количественная оценки оптимальных транспортных отображений, вычисленных алгоритмом. Для этого рассматривается широкий спектр задач и приложений как с синтетическими, так и с реальными данными. Обучаются ОТ отображения (а) между 2D распределениями; (б) между многомерными гауссианами; (с) для задачи транспорта масс в латентном пространстве автокодировщиков данных; (д) для задачи доменной адаптации в задаче классификации изображений рукописных цифр; (е) в задаче переноса стиля между непарными датасетами изображений.

Пример 1 (ОТ между гауссовскими распределениями). Ниже приведены количественные экспериментальные результаты для восстановления ОТ отображения T^* между случайно инициализированными гауссианами \mathbb{P}, \mathbb{Q} в размерностях $D = 2, 4, 8, \dots, 2^{12}$ с помощью предложенного метода. В качестве метрики для оценки обученного отображения $\hat{T} = \nabla\hat{\psi}$ используется \mathcal{L}^2 -UVP(\hat{T}) $\stackrel{\text{def}}{=} 100 \cdot \int_{\mathbb{R}^D} \|\hat{T}(x) - T^*(x)\|_2^2 d\mathbb{P}(x) / \text{Var}(\mathbb{Q})\%$.

<i>Dim</i>	2	4	8	16	32	64	128	256	512	1024	2048	4096
LSOT	< 1	3.7	7.5	14.3	23	34.7	46.9					> 50
MM-1	< 1	< 1	< 1	< 1	< 1	1.2	1.4	1.3	1.5	1.6	1.8	2.7
MM-2	< 1	< 1	< 1	< 1	< 1	< 1	1	1.1	1.2	1.3	1.5	2.1
W2GN	< 1	< 1	< 1	< 1	< 1	< 1	1	1.1	1.3	1.3	1.8	1.5

Таблица 1 — Сравнение \mathcal{L}^2 -UVP (%) (меньше=лучше) для методов LSOT [2], MM-1, MM-2 [7] и предложенного метода.

Разработанный метод значительно превосходит по качеству метод [2] на основе энтропийной регуляризации (Таблица 1) и эмпирически сходится быстрее, чем минимаксные методы [7] (Рис. 3).

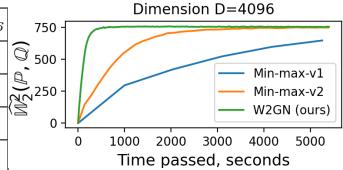


Рис. 3 — Сравнение сходимости для методов MM-1, MM-2 и предложенного метода.

В третьей главе решается задача поиска Васерштейн-2 барицентра набора непрерывных вероятностных распределений (задача 2). Предлагается и теоретически обосновывается численный метод для решения задачи на основе выпуклых по входу нейронных сетей, добавления циклической регуляризации и регуляризации на конгруэнтность в двойственной задаче.

В основе предлагаемого алгоритма лежит реформулировка двойственной задачи поиска Васерштейн-2 барицентра. Для $\mathbb{P}_1, \dots, \mathbb{P}_N \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ и весов $\alpha_1, \dots, \alpha_N$, удовлетворяющих $\alpha_n > 0$ и $\sum_{n=1}^N \alpha_n = 1$, **двойственная задача** барицентра [4] имеет следующий вид:

$$\mathcal{F}_{\text{Bar}}^* = - \min_{\{\psi_n\} \text{ конгруэнтны}} \left[\underbrace{\sum_{n=1}^N \alpha_n \int_{\mathbb{R}^D} \psi_n(y) d\mathbb{P}_n(y)}_{\text{MultiCorr}(\{\alpha_n, \mathbb{P}_n\} | \{\psi_n\})} \right] + \text{Const}(\{\alpha_n\}, \{\mathbb{P}_n\}), \quad (14)$$

где условие конгруэнтности обозначает, что функции $\psi_n : \mathbb{R}^D \rightarrow \mathbb{R}$ выпуклые и их сопряженные функции $\bar{\psi}_n$ удовлетворяют равенству $\sum_{n=1}^N \alpha_n \bar{\psi}_n(x) = \frac{\|x\|^2}{2} + c$ для всех $x \in \mathbb{R}$. Первое слагаемое в (14) называется *множественными корреляциями*.

Аналогично двойственной постановке задачи ОТ (4), выпуклые функции ψ_n называются потенциалами Бренье. Оптимальные потенциалы ψ_n^* позволяют немедленно получить решение $\bar{\mathbb{P}}$ прямой задачи (5). Для всех $n = 1, 2, \dots, N$ выполняется $\nabla \psi_n^* \sharp \mathbb{P}_n = \bar{\mathbb{P}}$, то есть из входных распределений с помощью $\nabla \psi_n^*$ можно получить барицентр $\bar{\mathbb{P}}$. Однако оптимизация (14) затруднительна из-за граничного условия на конгруэнтность.

В данной диссертационной работе предлагается рассматривать следующую неминимаксную задачу оптимизации ($\lambda > 0, \tau \geq 1$) по набору N

пар выпуклых функций $\{\psi_n, \phi_n\}_{n=1}^N$:

$$\min_{\{\psi_n, \phi_n\}} \text{MultiCorr}(\{\mathbb{P}_n\} \mid \{\psi_n\}, \{\phi_n\}; \tau, \hat{\mathbb{P}}, \lambda) \stackrel{\text{def}}{=} \min_{\{\psi_n, \phi_n\}} \left\{ \tau \cdot \mathcal{R}_1^{\hat{\mathbb{P}}}(\{\phi_n\}) + \right. \\ \left. \lambda \sum_{n=1}^N \alpha_n \mathcal{R}_2^{\mathbb{P}_n}(\psi_n, \phi_n) + \sum_{n=1}^N \left[\alpha_n \int_{\mathbb{R}^D} [\langle x, \nabla \psi_n(x) \rangle - \phi_n(\nabla \psi_n(x))] d\mathbb{P}_n(x) \right] \right\}, \quad (15)$$

где \mathcal{R}_2 является циклическим регуляризатором (10), предложенным в первой главе работы, а $\mathcal{R}_1^{\hat{\mathbb{P}}}(\{\phi_n\}) \stackrel{\text{def}}{=} \int_{\mathbb{R}^D} \left[\sum_{n=1}^N \alpha_n \phi_n(y) - \frac{\|y\|^2}{2} \right]_+ d\hat{\mathbb{P}}$ – новый предлагаемый регуляризатор для конгруэнтности потенциалов. В нем вводится вспомогательное распределение $\hat{\mathbb{P}} \in \mathcal{P}_2(\mathbb{R}^D)$.

Ключевой особенностью предложенной задачи оптимизации и регуляризатора $\mathcal{R}_1^{\hat{\mathbb{P}}}$ является то, что при правильном подборе параметров $\lambda, \tau, \hat{\mathbb{P}}$ оптимальным решением задачи являются оптимальные действительные пары потенциалов Бренье $(\psi_n^*, \phi_n^*) = (\psi_n^*, \overline{\psi_n^*})$, решающие (14). В диссертации доказывается теорема 3, связывающая ошибку решения задачи (15) с помощью некоторых потенциалов $(\hat{\phi}_n, \hat{\psi}_n)$ и ошибку восстановленного решения прямой задачи нахождения ОТ отображения (5) через $\nabla \hat{\psi}_n$.

Теорема 3. Пусть $\bar{\mathbb{P}} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ – барицентр для $\mathbb{P}_1, \dots, \mathbb{P}_N \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ с весами $\alpha_1, \dots, \alpha_N$ и пусть $\{\psi_n^*\}$ – оптимальные конгруэнтные потенциалы для задачи о барицентре. Предположим, что $\tau, \hat{\mathbb{P}}$ таковы, что $\tau \geq 1$ и $\tau \cdot \hat{\mathbb{P}} \geq \bar{\mathbb{P}}$. Пусть нам также даны выпуклые потенциалы $\{\hat{\psi}_n\}$ и β -сильно выпуклые и \mathcal{B} -гладкие потенциалы $\{\hat{\phi}_n\}$ с $0 < \beta \leq \mathcal{B} < \infty$ и $\lambda > \frac{\mathcal{B}}{2(\beta)^2}$. Положим $\epsilon = \text{MultiCorr}(\{\alpha_n, \mathbb{P}_n\} \mid \{\hat{\psi}_n\}, \{\hat{\phi}_n\}; \tau, \hat{\mathbb{P}}, \lambda) - \text{MultiCorr}(\{\alpha_n, \mathbb{P}_n\} \mid \{\psi_n^*\})$. Тогда $\epsilon \geq 0$ и при всех $n \in \{1, \dots, N\}$ имеем

$$\int_{\mathbb{R}^D} \|\nabla \hat{\psi}_n(y) - \nabla \psi^*(y)\|^2 d\mathbb{P}_n(y) \leq \frac{2\epsilon}{\alpha_n} \cdot \left(\sqrt{\frac{1}{\beta}} + \sqrt{\frac{1}{\lambda(\beta)^2 - \frac{\mathcal{B}}{2}}} \right)^2 = O(\epsilon), \quad (16)$$

что, в частности, означает $\mathbb{W}_2^2(\nabla \hat{\psi}_n \# \mathbb{P}_n, \bar{\mathbb{P}}) \leq O(\epsilon)$.

Полученная теоретическая оценка показывает, что точность решения предложенной неминимаксной задачи (15) с помощью $(\hat{\phi}, \hat{\psi})$ с точки зрения значения целевого функционала мажорирует точность восстановления барицентра $\bar{\mathbb{P}}$ с помощью $\{\nabla \hat{\psi}_n\}$ из входных распределений \mathbb{P}_n .

В работе предлагается **вычислительный алгоритм** для минимизации разработанного функционала (15) на основе методов глубинного обучения. В качестве $2N$ выпуклых потенциалов $\{\psi_n, \phi_n\}$ используются выпуклые по входу нейронные сети ICNN $\{\psi_{\theta_n}, \phi_{\omega_n}\}$ с параметрами (весами) θ_n, ω_n . Для оптимизации их параметров используется метод стохастического градиентного спуска с использованием случайных выборок из входных распределений $\{\mathbb{P}_n\}$ и априорного распределения $\hat{\mathbb{P}}$.



Рис. 4 — Барицентр линейно-разбросанных “швейцарских рулетов” с весами $(\alpha_1, \dots, \alpha_4) = (\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10})$, вычисленный алгоритмом.

Для эмпирической проверки работоспособности предложенного алгоритма в работе приводятся как качественная, так и количественная оценки ОТ барицентров, обученных алгоритмом в задачах: (а) вычисление барицентров 2D распределений; (б) вычисление барицентров линейно-разбросанных распределений; (с) применение барицентров к байесовской задаче агрегации апостериорных распределений, полученных по подвыборкам данных; (д) применение барицентров к задаче агрегации цветовых палитр RGB изображений. Результаты подтверждают превосходство алгоритма над существующими численными методами для барицентров.

Пример 2 (Барицентр локально-разбросанных распред.). См. рис. 4.

В четвертой главе решается задача вычисления Васерштейн-2 градиентного потока функционала на пространстве вероятностных распределений (задача 3). Предлагается численный метод для её решения на основе выпуклых по входу нейронных сетей и схемы JKO, рассмотренной ранее. Ниже детализируются результаты главы.

Согласно теореме Бренье [13] для любых $\rho \in \mathcal{P}_2(\mathbb{R}^D)$, $\rho^{(k-1)} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ существует $\rho^{(k-1)}$ -единственный градиент $\nabla \psi : \mathbb{R}^D \rightarrow \mathbb{R}^D$ выпуклой функции ψ , удовлетворяющий $\rho = \nabla \psi \sharp \rho^{(k-1)}$. Положим $\rho = \nabla \psi \sharp \rho^{(k-1)}$ и запишем шаг JKO (8) как оптимизационную задачу с выпуклой ψ :

$$\psi^{(k)} \leftarrow \arg \min_{\text{Convex } \psi} \left[\mathcal{F}(\nabla \psi \sharp \rho^{(k-1)}) + \frac{1}{h} \int_{\mathbb{R}^D} \frac{1}{2} \|x - \nabla \psi(x)\|_2^2 d\rho^{(k-1)}(x) \right]. \quad (17)$$

Так как ρ является образом $\rho^{(k-1)}$ под действием градиента выпуклой функции $\nabla \psi$, то член \mathbb{W}_2^2 в (17) может быть вычислен явно, что приводит к явному представлению расстояния Васерштейна-2. Для перехода на следующий шаг JKO определим $\rho^{(k)} \stackrel{\text{def}}{=} \nabla \psi^{(k)} \sharp \rho^{(k-1)}$. При такой переформулировке мы избегаем сложностей, связанных с вычислением расстояния Васерштейна-2. При этом для получения выборки из $\rho^{(k)} = [\nabla \psi^{(k)} \circ \dots \circ \nabla \psi^{(1)}] \sharp \rho^0$ можно выбрать выборку $x_0 \sim \rho^{(0)}$, и тогда $\nabla \psi^{(k)} \circ \dots \circ \nabla \psi^{(1)}(x_0)$ дает выборку из $\rho^{(k)}$.

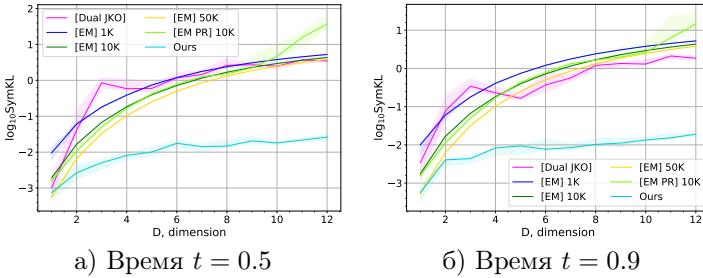


Рис. 5 — Значения симметричной дивергенции Кульбака–Лейблера между вычисленным и истинным распределением ρ_t в моменты времени $t = 0.5$ (слева) и $t = 0.9$ (справа) для размерностей $D = 1, 2, \dots, 12$.

В диссертации предлагается **вычислительный алгоритм** для осуществления переформулированного JKO интегрирования (17) на основе методов глубинного обучения. Для вычисления шага JKO (17), предлагается аппроксимировать $\psi^{(k)}$ с помощью выпуклой по входу нейронной сети ICNN ψ_{θ_k} с параметрами (весами) θ_k и для оптимизации параметров использовать алгоритм стохастического градиентного спуска с использованием выборок из полученного на предыдущем шаге распределения $\rho^{(k-1)}$.

Для эмпирической проверки предложенного метода проводится ряд вычислительных экспериментов: **(a)** моделирование процессов Орнштейна–Уленбека; **(b)** нахождение минимумов функционалов на пространстве распределений с помощью Васерштейн-2 градиентного потока; **(c)** применение градиентных потоков к вычислению апостериорного распределения в задаче байесовской логистической регрессии.

Пример 3 (Вычисление динамики процесса Орнштейн–Уленбека). На рис. 5 приводится количественный анализ восстановленной динамики процесса Орнштейн–Уленбека, т.е., градиентного потока функционала \mathcal{F}_{FP} (7) с квадратичным потенциалом $\Phi(x) = \frac{1}{2}(x-b)^T A(x-b)$, $A \succeq 0$ и гауссовским ρ^0 . В больших размерностях результаты экспериментов показывают значительное превосходство разработанной методологии над существующими методами вычисления градиентных потоков.

В **пятой главе** разрабатывается методология на основе нейросетей с выпуклой по входу архитектурой для синтеза пар непрерывных вероятностных распределений с аналитически известным ОТ отображением между ними (задача 4). Ниже описываются полученные результаты главы.

Ключевая идея предлагаемого метода построения эталонных пар заключается в том, что для дифференцируемой выпуклой $\psi : \mathbb{R}^D \rightarrow \mathbb{R}$ ее градиент $\nabla \psi$ является ОТ отображением между произвольным $\mathbb{P} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ и его образом $\nabla \psi \sharp \mathbb{P}$ под действием $\nabla \psi : \mathbb{R}^D \rightarrow \mathbb{R}^D$. Это следует из теоремы Бренье [13]. Таким образом, для непрерывного распределения \mathbb{P} с выборочным доступом и известным выпуклым ψ в качестве эталонной пары можно

использовать $(\mathbb{P}, \nabla \psi \sharp \mathbb{P})$. Получение выборки из $\nabla \psi \sharp \mathbb{P}$ можно осуществлять путем взятия выборки из \mathbb{P} и применением $\nabla \psi$.

В диссертационной работе разрабатываются пары $\mathbb{P}, \nabla \psi \sharp \mathbb{P}$ с аналитически известным (представленным градиентом выпуклой по входу нейронной сети) ОТ отображением: **(а)** многомерные пары непрерывных вероятностных распределений в размерностях $D = 2, 4, 8, \dots, 256$; **(б)** три пары вероятностных распределений на высокоразмерном пространстве $64 \times 64 \times 3$ изображений. Далее на разработанных парах осуществляется количественное тестирование существующих численных методов.

Результаты главы – пары непрерывных распределений с известным по построению эталонным ОТ отображением. Данная разработка позволяет устраниТЬ имеющийся пробел в тестировании непрерывных численных методов для задачи 1. Как результат ожидается, что разработанная методика тестирования может улучшить качество и надежность методов ОТ.

В заключении подводятся итоги результатов работы, а также обсуждаются потенциальные дальнейшие направления исследований.

Публикации автора по теме диссертации

В изданиях, индексируемых Scopus:

- A1 Korotin, A., Li, L., Genevay, A., Solomon, J. M., Filippov, A., & Burnaev, E. Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark // Advances in Neural Information Processing Systems. — 2021. — Vol. 34. — P. 14593–14605.
- A2 Mokrov, P., Korotin, A., Li, L., Genevay, A., Solomon, J. M., & Burnaev, E. Large-scale wasserstein gradient flows // Advances in Neural Information Processing Systems. — 2021. — Vol. 34. — P. 15243–15256.

В трудах конференций:

- A3 Korotin, A., Egiazarian, V., Asadulaev, A., Safin, A., & Burnaev, E. Wasserstein-2 Generative Networks // International Conference on Learning Representations. — 2021. [Электронный ресурс] URL: https://openreview.net/pdf?id=bEoxzW_EXsa (дата обращения 29.11.2022)
- A4 Korotin, A., Li, L., Solomon, J., & Burnaev, E. Continuous Wasserstein-2 Barycenter Estimation without Minimax Optimization // International Conference on Learning Representations. — 2021. [Электронный ресурс] URL: <https://openreview.net/pdf?id=3tFAs5E-Pe> (дата обращения 29.11.2022)
- A5 Rout, L., Korotin, A., & Burnaev, E. Generative Modeling with Optimal Transport Maps // International Conference on Learning Representations. — 2022. [Электронный ресурс] URL: <https://openreview.net/pdf?id=5JdLZg346Lw> (дата обращения 29.11.2022)

Список литературы

1. *Arjovsky M., Chintala S., Bottou L.* Wasserstein generative adversarial networks // International conference on machine learning. — PMLR. 2017. — P. 214—223.
2. Large Scale Optimal Transport and Mapping Estimation / V. Seguy [et al.] // International Conference on Learning Representations. — 2018.
3. Optimal transport for domain adaptation / N. Courty [et al.] // IEEE transactions on pattern analysis and machine intelligence. — 2016. — Vol. 39, no. 9. — P. 1853—1865.
4. *Aguech M., Carlier G.* Barycenters in the Wasserstein space // SIAM Journal on Mathematical Analysis. — 2011. — Vol. 43, no. 2. — P. 904—924.
5. *Jordan R., Kinderlehrer D., Otto F.* The variational formulation of the Fokker–Planck equation // SIAM journal on mathematical analysis. — 1998. — Vol. 29, no. 1. — P. 1—17.
6. Computational optimal transport / G. Peyré, M. Cuturi, [et al.] // Foundations and Trends® in Machine Learning. — 2019. — Vol. 11, no. 5/6. — P. 355—607.
7. Optimal transport mapping via input convex neural networks / A. Makkonen [et al.] // International Conference on Machine Learning. — PMLR. 2020. — P. 6672—6681.
8. Continuous regularized wasserstein barycenters / L. Li [et al.] // Advances in Neural Information Processing Systems. — 2020. — Vol. 33. — P. 17755—17765.
9. *Santambrogio F.* Optimal transport for applied mathematicians // Birkhäuser, NY. — 2015. — Vol. 55, no. 58—63. — P. 94.
10. *Fan J., Taghvaei A., Chen Y.* Scalable Computations of Wasserstein Barycenter via Input Convex Neural Networks // International Conference on Machine Learning. — PMLR. 2021. — P. 1571—1581.
11. *Frogner C., Poggio T.* Approximate inference with wasserstein gradient flows // International Conference on Artificial Intelligence and Statistics. — PMLR. 2020. — P. 2581—2590.
12. *Amos B., Xu L., Kolter J. Z.* Input convex neural networks // Proceedings of the 34th International Conference on Machine Learning-Volume 70. — JMLR. org. 2017. — P. 146—155.
13. *Brenier Y.* Polar factorization and monotone rearrangement of vector-valued functions // Communications on pure and applied mathematics. — 1991. — Vol. 44, no. 4. — P. 375—417.