# Обзор минимаксных параметрических методов вычисления оптимальных транспортных отображений

М.Д. Газдиева[I,II], А.А. Коротин[I,II]

[I] Сколковский институт науки и технологий, г. Москва, Россия
[II] Институт AIRI, г. Москва, Россия

**Аннотация.** В последние годы наблюдается новая волна интереса к теории оптимального транспорта (ОТ), обусловленная развитием масштабируемых и вычислительно эффективных методов решения задач ОТ и их активным использованием в задачах машинного обучения и генеративного моделирования. Кроме того, в настоящее время активно развиваются обобщенные постановки задачи ОТ (в частности, несбалансированные) и численные методы их решения, направленные на повышение практический применимости методов ОТ. В статье представлен обзор классических и обобщенных постановок задачи оптимального транспорта, а также современных минимаксных параметрических подходов к решению таких задач (включая нейросетевые модели), которые активно используются для решения задач машинного обучения.

**Ключевые слова:** *оптимальный транспорт, численные методы, машинное обучение, нейросетевые алгоритмы.*

**DOI:** 10.14357/20790279260303 **EDN:** AAELCW

## Введение

Под задачами оптимального транспорта понимается совокупность различных прикладных задач, которые могут быть сформулированы в виде математических принципов поиска способа переноса одного вероятностного распределения в другое, обладающего определенными свойствами [1]. В теории ОТ такой способ переноса называется ОТ-отображением или ОТ-планом в зависимости от постановки математической задачи.

Изначально задача ОТ, восходящая к работам Л.В. Канторовича [2], была сформулирована в контексте решения прикладных задач поиска оптимального плана перевозки грузов и распределения ресурсов. Предполагается, что груз распределен по $N$-поставщикам $x_1, \dots, x_N$ и его необходимо доставить $M$-потребителям $y_1, \dots, y_M$. Для каждого поставщика $x_i$ и каждого потребителя $y_j$ заданы следующие величины: $p(x_i)$ – объем груза у поставщика $x_i$, $q(y_j)$ – объем потребления для потребителя $y_j$, а также стоимость перевозки единицы груза от предприятия $x_i$ до потребителя $y_j$, обозначаемая как $c(x_i,y_j)$ и называемая *транспортной стоимостью*. Предполагается, что суммарное производство равно суммарному потреблению $\sum_{i=1}^{N} p(x_i) = \sum_{j=1}^{M} q(y_j) = Const$. Без ограничения общности можно полагать, что *Const*=1, и трактовать $p$, $q$ как дискретные *вероятностные* распределения грузов по поставщикам и необходимых объемов товара по потребителям, соответственно. Целью является нахождение оптимального транспортного плана перевозок $\pi(x_i,y_j)$, показывающего, какую долю груза от поставщика $x_i$ необходимо доставить потребителю $y_j$ для оптимизации стоимости перевозок (для всех $i,j$). Данный план представляет собой матрицу размера $N \times M$ и является решением следующей математической задачи линейного программирования (ЛП), известной как *транспортная задача Монжа-Канторовича*:

$$\sum_{j=1}^{M} \pi(x_i, y_j) = p(x_i), i = 1, \dots, N;$$
$$\sum_{i=1}^{N} \pi(x_i, y_j) = q(y_j), j = 1, \dots, M;$$
$$\sum_{i=1}^{N}\sum_{j=1}^{M} c(x_i, y_j)\pi(x_i, y_j) \to \min_{\pi}. \qquad (1)$$

Будем называть систему (1) – задачей дискретного ОТ. Она может быть аналогичным образом сформулирована для непрерывных распределений $p$ и $q$, заданных на непрерывных пространствах $X$, $Y$ (как правило, подмножествах Евклидова пространства $\mathbb{R}^D$). Здесь для заданной непрерывной функции транспортной стоимости $c(x,y)$ требуется найти транспортный план (вероятностное распределение) $\pi(x,y)$, такой что:

$$\pi_x = \int_Y d\pi(x,y) = p,$$

$$\pi_y = \int_X d\pi(x,y) = q,$$

$$\int_{X\times Y} c(x,y)d\pi(x,y) = \int_X [\int_Y c(x,y)d\pi(y|x)]dp(x) \to \min_{\pi}. \quad (2)$$

Данную задачу будем называть задачей непрерывного (вероятностного) ОТ. Представление (2) показывает, что условное распределение $\pi(y|x)$ должно быть сосредоточено на тех значениях переменной $y$, для которых функция $c(x,y)$ мала (в частном случае, распределение $\pi(y|x)$ может быть сосредоточено в одной точке). К примеру, такая задача может возникать при моделировании переноса массы в непрерывной среде, такой как жидкость или газ, где распределение вещества описывается плотностью [3]. Строгая математическая постановка задачи вероятностного ОТ дана в Разделе 1.

Кроме того, возможна *статистическая* постановка задачи ОТ, которая отличается от непрерывной (2) тем, что вероятностные распределения $p$ и $q$ неизвестны, а заданы лишь выборки данных из этих распределений $(x_1, x_2, \dots, x_N) \sim p$ и $(y_1, y_2, \dots, y_M) \sim q$. Детали постановки описаны в Разделе 2.

Имеется небольшое количество публикаций, посвященных решению непрерывной задачи ОТ для конкретных распределений $p$ и $q$ и функции $c$ [4].

В последние годы интерес к ОТ существенно возрос в связи с тем, что многие прикладные задачи генеративного моделирования (такие как перенос стиля [5], сверхразрешение изображений [6]) могут быть сформулированы как математические задачи ОТ. Эти прикладные задачи в общем виде выглядят следующим образом. Рассматривается два множества объектов $X$ и $Y$ (например, изображений), которые представлены конечными наборами элементов из этих множеств (выборками) $X_N = (x_1, x_2, \dots, x_N)$ и $Y_M = (y_1, y_2, \dots, y_M)$, в предположении, что данные выборки взяты из «генеральных совокупностей» $X$ и $Y$ в соответствии с распределениями $p$ и $q$. Рассматривается некоторое свойство, которым обладают элементы из множеств $X$ и $Y$, и выбранная функция $c(x,y)$ определяет некоторую степень различия элементов $x\in X$ и $y\in Y$ с точки зрения выбранного свойства.

Тогда решение статистической задачи ОТ, определяющее условное распределение $\pi(y|x)$, позволяет решить следующую прикладную задачу, известную как задача условного генеративного моделирования: по заданному элементу $x\in X$ сгенерировать элемент $y\in Y$, который максимально сохраняет свойство, имеющееся у элемента $x$. Известным примером такой прикладной задачи является проблема переноса стиля изображений. В данном примере множества $X$ и $Y$ обозначают два класса изображений (два стиля), а выбранное свойство может отражать содержание (контент) изображения. Мы хотим для элемента $x\in X$ одного стиля создать (сгенерировать) элемент $y\in Y$ другого стиля, который максимально сохраняет контент изображения.

Отметим, что рассматриваемая задача отличается от проблемы поиска: для заданного $x\in X$ *найти* в множестве $Y_M$ элемент $y$, минимизирующий функцию $c(x,y)$. В нашей задаче мы хотим создать (сгенерировать) новый такой элемент $y\in Y$, которого ранее не было в выборке $Y_M$. Этим отличается задача поиска от задачи генеративного моделирования.

Данная статья посвящена обзору методов статистического ОТ. В Разделе 1 систематизируются различные постановки задач ОТ между непрерывными распределениями. В Разделе 2 представлен обзор современных методов для решения задач статистического ОТ.

## 1. Математические постановки задач оптимального транспорта

В данном разделе кратко изложены базовые постановки и ключевые понятия теории оптимального транспорта.

В начале изложим основные обозначения раздела. Далее в разделе $X$ и $Y$ обозначают компактные подмножества $D$-мерного евклидова пространства $\mathbb{R}^D$. Множество вероятностных мер на пространстве $X$ обозначается как $P(X)$, множество произвольных неотрицательных мер – как $M_+(X)$. Для измеримой функции $T{:}X\to Y$ оператор переноса меры, порожденный $T$, обозначается как $T_\#$. Мы используем $I_A$ для обозначения выпуклого индикатора множества $A{:}\ I_A(x) = 0$, если $x\in A$, иначе $I_A(x) = +\infty$. Мы используем $Id_X$ для обозначения тождественного отображения на пространстве $X$, такого что $Id_X(x) = x$ для любого $x\in X$.

Для положительных мер $\mu_1, \mu_2 \in M_+(X)$ и полунепрерывной снизу функции $f: \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ $f$-дивергенция между $\mu_1, \mu_2$ определяется так

$$D_f(\mu_1||\mu_2) =$$

$$\begin{cases} \int_X f(\frac{\mu_1(x)}{\mu_2(x)}) d\mu_2(x), & \text{если } \mu_1(x) \ll \mu_2(x), \\ \{+\infty\}, & \text{иначе.} \end{cases}$$

Известным примером $f$-дивергенции является дивергенция Кульбака-Лейблера.

Начнем с изложения математических фактов, касающихся постановки непрерывной задачи ОТ (2). План $\pi^*$, доставляющий минимум задаче (2) называется оптимальным транспортным планом, а значение формулы (2) при оптимальном плане – оптимальной транспортной стоимостью $OT_c(p,q) := \int_{X\times Y} c(x,y) d\pi^*(x,y)$. Заметим, что при мягких условиях на функцию стоимости и распределения $p$ и $q$, ОТ план всегда существует, хотя его единственность не гарантирована [7]. Классическим примером ОТ стоимости является расстояние Васерштейна ($W_1(P,Q)$), которое задается формулой (2) при $c(x,y) = ||x-y||$.

В некоторых частных случаях ОТ-план является детерминированным, то есть представим его в виде $\pi^* = [Id_X, T^*]_{\#}p$, где $T^*$ – некоторое измеримое отображение $T^*: X \to Y$. Тогда $T^*$ называется ОТ отображением из $p$ и $q$, а ОТ задача в формулировке Канторовича (2) сводится к так называемой ОТ задаче Монжа:

$$OT_c(p,q) := inf_{T:T\#p=q} \int_X c(x,T(x)) dp(x). \quad (3)$$

Здесь инфимум берется по всем измеримым отображениям $T: X \to Y$, которые переносят $p$ и $q$, то есть удовлетворяют условию $T_{\#}p=q$. Заметим, что в случае произвольных распределений $p$ и $q$ аналитическое решение ОТ задачи (3) зачастую неизвестно.

Имеет место следующая *двойственная* формулировка задачи ОТ (2):

$$OT_c(p,q) = \\ = sup_v \{\int_X v^c(x) dp(x) + \int_Y v(y) dq(y)\}. \quad (4)$$

где $v: Y \to \mathbb{R}$ – измеримая функция, называемая *потенциалом*, $v^c(x)$ обозначает $c$-преобразование потенциала $v^c(x) = inf_{y\in Y} \{c(x,y) - v(y)\}$. Функция $v^*$, доставляющая максимум задаче (4), называется оптимальным транспортным потенциалом. При этом известно [7], что если ОТ-план является детерминированным, то есть существует ОТ отображение $T^*$, такое что $\pi^* = [Id_X, T^*]_{\#}p$, то выполнено:

$$T^*(x) \in arg\ \underset{y\in Y}{min}[c(x,y) - v^*(y)] \Rightarrow \\ (v^*)^c(x) = c(x,T^*(x)) - v^*(T^*(x)). \quad (5)$$

Данное свойство показывает, что ОТ-потенциал $v^*$ связан с решением прямой задачи (3) – ОТ отображением $T^*$, задающим детерминистический ОТ план $\pi^* = [Id_X, T^*]_{\#}p$. При некоторых предположениях на меры $p$, $q$, функцию стоимости $c(x,y)$ и потенциал $v^*$ [8] можно гарантировать, что $arg\ \underset{y\in Y}{min}[c(x,y) - v^*(y)]$ содержит лишь одну точку и формула (5) преобразуется в равенство $T^*(x) = arg\ \underset{y\in Y}{min}[c(x,y) - v^*(y)]$, тем самым позволяя восстанавливать решение прямой задачи по двойственному потенциалу $v^*$.

В последние годы стали рассматриваться обобщенные постановки задачи оптимального транспорта, такие как несбалансированный (unbalanced OT, UOT) ОТ [9], в котором допускается ослабление маргинальных ограничений.

Для двух распределений $p$ и $q$ задача UOT состоит в поиске плана $\lambda \in M_+(X\times Y)$ (совместного распределения на $X$ и $Y$), доставляющего минимум задаче:

$$UOT_c(p,q) = inf_{\gamma \in M_+(X\times Y)}\{\int_{X\times Y} c(x,y) d\gamma(x,y) \\ + D_{f_1}(\gamma_x||\, p) + D_{f_2}(\gamma_y||\, q)\}, \quad (6)$$

где $D_{f_1}$, $D_{f_2}$ – $f$-дивергенции на пространстве неотрицательных мер $M_+(X)$ и $M_+(Y)$, $\gamma_x$, $\gamma_y$ – маргиналы плана. Заметим, что классическая задача оптимального транспорта является частным случаем постановки UOT при $f_1 = f_2 = I_{x=1}$.

UOT-задача также имеет двойственную формулировку:

$$UOT_c(p,q) = sup_v \{-\int_X \bar{f}_1(-v^c(x)) dp(x) \\ - \int_Y \bar{f}_2(-v(y)) dq(y)\}, \quad (7)$$

где $\bar{f}$ – выпуклая сопряженная к $f$ функция $\bar{f}(y) = sup_{x\in\mathbb{R}} \{< x,y > - f(x)\}$ для $f: \mathbb{R} \to \mathbb{R} \cup \{\infty\}$.

Постановка задачи статистического ОТ и обзор статистических методов ОТ представлены в следующем разделе.

## 2. Статистические постановки ОТ задач и методы их решения

Недавний прогресс в области вычислительного ОТ и его активное использование в задачах машинного обучения связаны с развитием методов решения статистической постановки задачи ОТ, называемых методами *статистического* ОТ.

В статистической задаче ОТ необходимо найти решение непрерывной задачи ОТ (3) или (4), то есть оптимальный транспортный план $\pi^*$ между мерами $p$ и $q$, в предположении, что распределения $p$, $q$ неизвестны, а доступны лишь конечные

выборки данных из них $X_N = (x_1, x_2, \ldots, x_N) \sim p$ и $Y_M = (y_1, y_2, \ldots, y_M) \sim q$. Целью современных методов статистического ОТ является восстановление (приближение) оптимального транспортного плана $\pi$, который может использоваться для генерации данных $y \sim \pi(y|x_{new})$, где $x_{new}$ – новый экземпляр, ранее не встречавшийся в обучающей выборке $X_N$.

Из постановки ОТ задачи (2) следует, что для оптимального транспортного плана выполнены условия: $\pi_x = p$ и $\pi_y = q$. Дополнительно поясним значение данных условий в применении к задаче генерации. Планируется, что модель будет применяться к новым данным $x_{new}$, выбранным из «генеральной совокупности» $X$ в соответствии с распределением $p$. Кроме того, новые сгенерированные объекты должны быть «схожи» с элементами, выбранными из «генеральной совокупности» $Y$ в соответствии с распределением $q$, что объясняет практическое значение маргинальных условий.

Прямая задача ОТ представляет собой проблему с ограничениями вида $\pi_x = p$ и $\pi_y = q$ , которые сложно проверить, имея только выборки из распределений $p$ и $q$. По этой причине большинство методов статистического ОТ восстанавливают оптимальный транспортный план, решая двойственную задачу ОТ (4), в которой переменной оптимизации является функция (потенциал) $v: Y \to \mathbb{R}$. Существующие методы работают в предположениях существования детерминистического плана (что верно, например, если $p$ абсолютно непрерывна, а функция стоимости – квадратична), когда восстановление ОТ плана $\pi^*$ сводится к задаче восстановления ОТ отображения $T^*$.

В основе таких подходов лежит взаимосвязь решений прямой (2) и двойственной (4) задач ОТ, (см. формулу (5)), что позволяет восстанавливать ОТ-план на основе восстановленного потенциала. Поскольку в статистической задаче ОТ доступны лишь выборки данных из распределений $p$ и $q$, рассматривается приближение задачи (4) по выборкам $X_N$, $Y_M$:

$$sup_{v:Y\to\mathbb{R}} \{\tfrac{1}{N}\textstyle\sum_{i=1}^{N} v^c(x_i) + \tfrac{1}{M}\sum_{j=1}^{M} v(y_j)\}. \quad (8)$$

Заметим, что решение задачи (8) требует вычисления $c$-преобразования потенциала $v^c(x) = inf_{y\in Y}\{c(x,y) - v(y)\}$, что является вычислительно затратной процедурой, поскольку требует решения отдельной оптимизационной задачи для каждой точки $x \in X$.

Стандартный подход, который был предложен в статьях [5,6,8,10–14], состоит в том, что в двойственной задаче (4) рассматривается эквивалентная форма интеграла от $c$-преобразовании потенциала, в которой оптимизация по $y \in Y$ заменяется на эквивалентную оптимизацию по функциям $T: X \to Y$:

$$\int_X v^c(x)dp(x) = \int_X inf_{y\in Y}\{c(x,y) - v(y)\}dp(x)$$

$$= inf_{T:X\to Y}\int_X \{c(x,T(x)) - v(T(x))\}dp(x), \quad (9)$$

где перестановка интеграла и инфимума осуществляется с использованием теоремы Рокафеллара [15]. Тогда с использованием формулы (9) двойственная задача ОТ (4) может быть представлена как задача поиска седловой точки вида:

$$sup_v inf_T \{\int_X [c(x,T(x)) - v\big(T(x)\big)]dp(x) + \textstyle\int_Y v(y)dq(y)\} = sup_v inf_T\, L(v,T). \quad (10)$$

Заметим, что на практике решение задачи (10) ищется в некоторых параметрических семействах функций $\{v_\omega\ (\omega \in \Omega)\}$, $T_\theta\ (\theta \in \Theta)$, например, нейросетевых, по доступным выборкам из распределений. Тогда в статистической формулировке задача (10) сводится к задаче поиска седловой точки (оптимизации типа max-min) по $\theta$ и $\omega$:

$$sup_\omega inf_\theta \{\frac{1}{N}\sum_{i=1}^{N}\big[c\big(x_i, T_\theta(x_i)\big) - v_\omega\big(T_\theta(x_i)\big)\big] + \tfrac{1}{M}\textstyle\sum_{j=1}^{M} v_\omega\big(y_j\big)\}, \quad (11)$$

где параметры $\theta$ и $\omega$ оптимизируются с помощью алгоритма градиентного подъема-спуска с использованием случайных выборок из распределений $p$, $q$ аналогично известным алгоритмам генеративного моделирования [16]. Решением задачи (11) называется пара потенциала и отображения $(v_{\omega^*}, T_{\theta^*})$ для оптимальных параметров $(\omega^*, \theta^*)$.

В статье [17] проводится теоретический анализ решений минимаксных методов. А именно, авторы выводят статистическую оценку на разницу между истинным ОТ отображением $T^*$ и восстановленным отображением $T_{\theta^*}$: $\| T^* - T_{\theta^*} \|^2$, называемую *ошибкой обобщения*. В статье вводятся следующие предположения: ошибка оптимизации при решении задачи (11) равняется нулю и находится оптимальное решение этой задачи $(v_{\omega^*}, T_{\theta^*})$, оптимальный двойственный потенциал $v^*$ является $\beta$-сильно выпуклым, а $v_{\omega^*}$ принадлежит параметрическому классу липшицевых $\beta$-сильно выпуклых функций (нейросетей), вполне ограниченному относительно нормы Липшица. В данных предположениях показывается, что ошибка обобщения может быть сделана угодно малой при

достаточно большом размере выборок $N,M$ и выразительном классе нейросетей, то есть $T_{\theta^*} \approx T^*$. Хотя указанные предположения довольно ограничивающие, на практике такие минимаксные алгоритмы позволяют достичь хорошего качества восстановления ОТ отображения $T^*$ и без этих предположений [18].

В недавнее время стали также активно развиваться методы решения несбалансированных (обобщенных) постановок ОТ (6) в статистической формулировке. Напомним, что ключевое отличие несбалансированной постановки ОТ (6) от классической (3) заключается в том, что строгое условие на маргиналы ОТ-плана $\pi_x=p$ и $\pi_y=q$ в задаче (6) заменяется штрафами. Такая задача рассматривается по двум причинам. Во-первых, на практике имеющиеся выборки данных $X_N$ и $Y_M$ могут содержать выбросы и некорректно характеризовать генеральные совокупности $X$ и $Y$. Во-вторых, такие постановки позволяют разрабатывать более эффективные практические алгоритмы [19].

Аналогично классическим методам статистического ОТ стандартные методы решения задачи UOT в статистической формулировке [19,20] также восстанавливают UOT-план, решая двойственную задачу UOT (7), используя доступные выборки $X_N$, $Y_M$ из распределений $p$, $q$:

$$sup_{v:Y\to\mathbb{R}} \{-\tfrac{1}{N}\textstyle\sum_{i=1}^{N} \bar{f}_1(-v^c(x_i)) - \tfrac{1}{M}\sum_{j=1}^{M} \bar{f}_2(-v(y_j))\}. \quad (12)$$

Затем с использованием эквивалентной формы $c$-преобразования потенциала $v^c$ (9) задача (12) сводится к проблеме поиска седловой точки:

$$sup_{\omega} inf_{\theta} \{-\frac{1}{N}\sum_{i=1}^{N} \bar{f}_1(-c(x_i, T_\theta(x_i)) + v_\omega(T_\theta(x_i))) - \frac{1}{M}\sum_{j=1}^{M} \bar{f}_2(-v_\omega(y_j))\}. \quad (13)$$

где параметры $\theta$ и $\omega$ оптимизируются с помощью алгоритма градиентного подъема-спуска с использованием случайных выборок из распределений $p$, $q$.

Заметим, что в UOT-задаче транспортный план $\pi^*$ имеет вид $\pi^* = [Id_X, T^*]_{\#}\pi_x{}^*$, где маргинал $\pi_x{}^*$ не обязан совпадать с распределением $p$. По этой причине помимо задачи восстановления ОТ отображения $T^*$, возникает проблема восстановления маргинала $\pi_x{}^*$. Для восстановления $T^*$ используется стандартный подход, описанный выше, а маргинал $\pi_x{}^*$ может быть восстановлен на основе выученного потенциала $v_{\omega^*}$ и функции стоимости $c(x,y)$ [19]. Однако на практике маргинал плана обычно не используется, требуется восстановить лишь отображение $T^*$.

В заключение отметим, что существуют некоторые методы статистического ОТ, не основанные на состязательном обучении, однако они обладают рядом недостатков относительно состязательных подходов [18]. Также отметим прочие существующие подходы для решения статистической UOT-задачи. В работе [21] двойственный потенциал параметризуется при помощи смеси Гауссиан, что позволяет свести задачу к несостязательной постановке и вывести оценку на ошибку обобщаемости для восстановленного плана. В работе [22] предложено обобщение описанного состязательного подхода для задачи поиска «средних» распределений данных. Есть большое число методов, предлагающих схемы восстановления UOT-планов на основе решения дискретных UOT-задач [23–25]. Однако такие методы зачастую основаны на эвристических принципах и не имеют теоретических гарантий.

## Заключение

В работе представлен обзор современных минимаксных методов статистического оптимального транспорта с акцентом на их применение в задачах машинного обучения и генеративного моделирования. Отдельное внимание уделено современным параметрическим подходам, в которых транспортные отображения параметризуются (в том числе с использованием нейросетевых моделей) и настраиваются по выборочным данным, что обеспечивает возможность применения полученных отображений к новым тестовым данным.

Вместе с тем реальные данные в прикладных сценариях часто содержат неточности, такие как выбросы или шумы, что ограничивает применимость методов на основе классических ОТ постановок. Это мотивирует использование моделей на основе обобщенного (несбалансированного, частичного) ОТ, которые позволяют повысить устойчивость к указанным неточностям в данных. В работе систематизированы обобщенные (несбалансированные) постановки ОТ, а также представлен обзор соответствующих параметрических методов их решения.

Проведенный обзор позволяет выделить ключевые направления развития вычислительного ОТ в контексте генеративного моделирования и подчеркивает значимость дальнейших исследований, направленных на повышение масштабируемости и устойчивости современных параметрических методов ОТ.

## Литература

1. Santambrogio F. Optimal transport for applied mathematicians: calculus of variations, PDEs, and

modeling. Cham: Birkhäuser, 2015. 353 p.
2. Канторович Л.В. О перемещении масс // Доклады Академии наук СССР. 1942. Т. 37. № 7-8. С. 227-229.
3. Benamou J.D., Brenier Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem // Numerische Mathematik. 2000. Vol. 84. No. 3. P. 375-393.
4. Knott M., Smith C.S. On the optimal mapping of distributions // Journal of Optimization Theory and Applications. 1984. Vol. 43. No. 1. P. 39-49.
5. Rout L., Korotin A., Burnaev E. Generative modeling with optimal transport maps // International Conference on Learning Representations, 2022.
6. Gazdieva M., Mokrov P., Rout L., Korotin A., Kravchenko A., Filippov A., Burnaev E. An optimal transport perspective on unpaired image super-resolution // Journal of Optimization Theory and Applications. 2025. Vol. 207. No. 2. Art. 40.
7. Villani C. Optimal transport: old and new. Berlin; Heidelberg: Springer, 2009. 976 p.
8. Fan J., Liu S., Ma S., Zhou H.-M., Chen Y. Neural Monge map estimation and its applications // Transactions on Machine Learning Research, 2023.
9. Séjourné T., Peyré G., Vialard F.-X. Unbalanced optimal transport, from theory to numerics // Numerical Control: Part B. Elsevier, 2023. P. 407-471.
10. Korotin A., Selikhanovych D., Burnaev E. Neural optimal transport // International Conference on Learning Representations, 2023.
11. Korotin A., Selikhanovych D., Burnaev E. Kernel neural optimal transport // International Conference on Learning Representations, 2023.
12. Asadulaev A., Korotin A., Egiazarian V., Mokrov P., Burnaev E. Neural optimal transport with general cost functionals // The Twelfth International Conference on Learning Representations, 2024.
13. Buzun N., Bobrin M., Dylov D.V. Expectile regularization for fast and accurate training of neural optimal transport // Advances in Neural Information Processing Systems. 2024. Vol. 37. P. 119811-119837.
14. Nguyen B., Nguyen B., Nguyen H.T., Nguyen V.A. Generative conditional distributions by neural (entropic) optimal transport // Proceedings of the 41st International Conference on Machine Learning. 2024. Vol. 235. P. 37761-37775.
15. Rockafellar R.T. Integral functionals, normal integrands and measurable selections // Nonlinear Operators and the Calculus of Variations. Berlin; Heidelberg: Springer, 1976. P. 157-207.
16. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial networks // Communications of the ACM. 2020. Vol. 63. No. 11. P. 139-144.
17. Tarasov R., Mokrov P., Gazdieva M., Burnaev E., Korotin A. A statistical learning perspective on semi-dual adversarial neural optimal transport solvers // International Conference on Learning Representations, 2026.
18. 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.
19. Choi J., Choi J., Kang M. Generative modeling through the semi-dual formulation of unbalanced optimal transport // Advances in Neural Information Processing Systems. 2023. Vol. 36. P. 42433-42455.
20. Gazdieva M., Korotin A., Selikhanovych D., Burnaev E. Extremal domain translation with neural optimal transport // Advances in Neural Information Processing Systems. 2023. Vol. 36. P. 40381-40413.
21. Gazdieva M., Asadulaev A., Burnaev E., Korotin A. Light unbalanced optimal transport // Advances in Neural Information Processing Systems. 2024. Vol. 37. P. 93907-93938.
22. Gazdieva M., Choi J., Kolesov A., Choi J., Mokrov P., Korotin A. Robust barycenter estimation using semi-unbalanced neural optimal transport // International Conference on Learning Representations, 2025.
23. Lübeck F., Bunne C., Gut G., del Castillo J.S., Pelkmans L., Alvarez-Melis D. Neural unbalanced optimal transport via cycle-consistent semi-couplings // NeurIPS 2022 Workshop on Learning Meaningful Representations of Life. 2022.
24. Eyring L., Klein D., Uscidda T., Palla G., Kilbertus N., Akata Z., Theis F.J. Unbalancedness in neural Monge maps improves unpaired domain translation // The Twelfth International Conference on Learning Representations, 2024.
25. Klein D., Uscidda T., Theis F., Cuturi M. GENOT: entropic (Gromov) Wasserstein flow matching with applications to single-cell genomics // Advances in Neural Information Processing Systems. 2024. Vol. 37. P. 103897-103944.

**Газдиева Милена Джамбулатовна.** Сколковский институт науки и технологий, г. Москва, Россия. Инженер-исследователь. Институт AIRI, г. Москва, Россия. Научный сотрудник. Область научных интересов: машинное обучение, генеративное моделирование. E-mail: milena.gazdieva@skoltech.ru (Ответственный за переписку).

**Коротин Александр Андреевич.** Сколковский институт науки и технологий, г. Москва, Россия. Кандидат физико-математических наук, доцент. Институт AIRI, г. Москва, Россия. Ведущий научный сотрудник. Область научных интересов: машинное обучение, генеративное моделирование. E-mail: a.korotin@skoltech.ru

## A survey of minimax parametric methods for computing optimal transport maps

M.D. Gazdieva[I,II], A.A. Korotin[I,II]

[I] Skolkovo Institute of Science and Technology, Moscow, Russia

[II] AIRI Institute, Moscow, Russia

**Abstract.** Optimal transport (OT) theory, whose foundations were laid in the works of the Soviet mathematician and economist L. V. Kantorovich in the mid-20th century, was originally developed to address problems of optimal cargo transportation and resource allocation. In recent years, OT has experienced a new wave of interest, driven by the development of scalable and computationally efficient methods for solving OT problems, as well as by their widespread use in machine learning and generative modeling. In addition, generalized formulations of the OT problem, in particular unbalanced ones, and numerical methods for their solution are currently being actively developed to improve the practical applicability of OT-based approaches. This paper presents a survey of classical and generalized formulations of the optimal transport problem, as well as modern minimax parametric approaches to solving such problems, including neural network-based models, which are actively used in machine learning applications.

**Keywords:** *optimal transport, numerical methods, machine learning, neural network algorithms*.
**DOI:** 10.14357/20790279260303 **EDN:** AAELCW

### References

1. Santambrogio F. *Optimal transport for applied mathematicians: calculus of variations, PDEs, and modeling.* Cham: Birkhäuser; 2015. 353 p.
2. Kantorovich L.V. On the translocation of masses. *Doklady Akademii Nauk SSSR*. 1942;37(7-8):227-229. (In Russ.).
3. Benamou J.D., Brenier Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*. 2000;84(3):375-393.
4. Knott M., Smith C.S. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications.* 1984;43(1):39-49.
5. Rout L., Korotin A., Burnaev E. Generative modeling with optimal transport maps. In: *International Conference on Learning Representations*, 2022.
6. Gazdieva M., Mokrov P., Rout L., Korotin A., Kravchenko A., Filippov A., Burnaev E. An optimal transport perspective on unpaired image super-resolution. *Journal of optimization theory and applications.* 2025;207(2):40.
7. Villani C. *Optimal transport: old and new.* Berlin; Heidelberg: Springer; 2009. 976 p.
8. Fan J., Liu S., Ma S., Zhou H.-M., Chen Y. Neural Monge map estimation and its applications. *Transactions on machine learning research,* 2023.
9. Séjourné T., Peyré G., Vialard F.-X. Unbalanced optimal transport, from theory to numerics. In: *Numerical control: part B.* Elsevier. 2023:407-471.
10. Korotin A., Selikhanovych D., Burnaev E. Neural optimal transport. In: *International Conference on Learning Representations*, 2023.
11. Korotin A., Selikhanovych D., Burnaev E. Kernel neural optimal transport. In: *International Conference on Learning Representations*, 2023.
12. Asadulaev A., Korotin A., Egiazarian V., Mokrov P., Burnaev E. Neural optimal transport with general cost functionals. In: *The Twelfth International Conference on Learning Representations*, 2024.
13. Buzun N., Bobrin M., Dylov D.V. Expectile regularization for fast and accurate training of neural optimal transport. In: *Advances in Neural Information Processing Systems.* 2024;37:119811-119837.
14. Nguyen B., Nguyen B., Nguyen H.T., Nguyen V.A. Generative conditional distributions by neural

(entropic) optimal transport. In: *Proceedings of the 41st International Conference on Machine Learning.* 2024;235:37761-37775.
15. Rockafellar R.T. Integral functionals, normal integrands and measurable selections. In: *Nonlinear operators and the calculus of variations.* Berlin; Heidelberg: Springer, 1976:157-207.
16. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial networks. *Communications of the ACM.* 2020;63(11):139-144.
17. Tarasov R., Mokrov P., Gazdieva M., Burnaev E., Korotin A. A statistical learning perspective on semi-dual adversarial neural optimal transport solvers. In: *International Conference on Learning Representations*, 2026.
18. Korotin A., Li L., Genevay A., Solomon J.M., Filippov A., Burnaev E. Do neural optimal transport solvers work? A continuous Wasserstein-2 benchmark. In: *Advances in Neural Information Processing Systems.* 2021;34:14593-14605.
19. Choi J., Choi J., Kang M. Generative modeling through the semi-dual formulation of unbalanced optimal transport. In: *Advances in Neural Information Processing Systems.* 2023;36:42433-42455.
20. Gazdieva M., Korotin A., Selikhanovych D., Burnaev E. Extremal domain translation with neural optimal transport. In: *Advances in Neural Information Processing Systems.* 2023;36:40381-40413.
21. Gazdieva M., Asadulaev A., Burnaev E., Korotin A. Light unbalanced optimal transport. In: *Advances in Neural Information Processing Systems.* 2024;37:93907-93938.
22. Gazdieva M., Choi J., Kolesov A., Choi J., Mokrov P., Korotin A. Robust barycenter estimation using semi-unbalanced neural optimal transport. In: *International Conference on Learning Representations*, 2025.
23. Lübeck F., Bunne C., Gut G., del Castillo J.S., Pelkmans L., Alvarez-Melis D. Neural unbalanced optimal transport via cycle-consistent semi-couplings. In: *NeurIPS 2022 Workshop on Learning Meaningful Representations of Life*, 2022.
24. Eyring L., Klein D., Uscidda T., Palla G., Kilbertus N., Akata Z., Theis F.J. Unbalancedness in neural Monge maps improves unpaired domain translation. In: *The Twelfth International Conference on Learning Representations*, 2024.
25. Klein D., Uscidda T., Theis F., Cuturi M. GENOT: entropic (Gromov) Wasserstein flow matching with applications to single-cell genomics. In: *Advances in Neural Information Processing Systems.* 2024;37:103897-103944.

**Gazdieva Milena D**. Skolkovo Institute of Science and Technology, 30 Bolshoy Boulevard, bld. 1, Moscow, Russia. AIRI Institute, Moscow, Russia. Research interests: machine learning, generative modeling. E-mail: milena.gazdieva@skoltech.ru

**Korotin Alexander A**. Candidate of Physical and Mathematical Sciences. Skolkovo Institute of Science and Technology, 30 Bolshoy Boulevard, bld. 1, Moscow, Russia. AIRI Institute, Moscow, Russia. Research interests: machine learning, generative modeling. E-mail: a.korotin@skoltech.ru