КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет комп'ютерних наук та кібернетики

Випускна кваліфікаційна робота бакалавра

Оптимальний транспорт та алгоритми пошуку барицентрів Вассерштейна

Студентки 4 курсу
кафедри обчислювальної математики
Коваленко Олександри Юріївни

Науковий керівник
доктор фізико-математичних наук, професор
Семенов Володимир Вікторович

____

"____"____________2021р.

Робота заслухана на засіданні кафедри обчислювальної математики та рекомендована до захисту в ДЕК, протокол

Завідувач кафедри обчислювальної математики проф. Ляшко С.І.

Київ

2021

1 РЕФЕРАТ

Обсяг роботи 42 сторінки, 12 джерел посилання, 3 ілюстрації, 2 таблиці. *Ключові слова:* ВАРІАЦІЙНА НЕРІВНІСТЬ, БАРИЦЕНТР ВАССЕРШТЕЙНА, МОНОТОННИЙ ОПЕРАТОР, ОПТИМАЛЬНИЙ ТРАНСПОРТ.

Мета роботи написати програмну бібліотеку, яка допоможе у розв'язанні варіаційної нерівності для задачі пошуку барицентра Вассерштейна, використовуючи алгоритми Попова, Корпелевич, Ценга, операторної екстраполяції та алгоритмів Маліцького (відбиваючого проєкційного алгоритму та методу дзеркального спуску) та порівняти результати на основі проведеного тестування для поставленої задачі.

Для досягнення зазначеної мети, необхідно виконати наступні *завдання*: охарактеризувати сучасні методи для алгоритмів пошуку барицентрів Вассерштейна, виділити алгоритми транспортих задач, визначити способи оптимізації алгоритма пошуку барицентра Вассерштейна за допомогою алгоритмів екстраполяції з минулого, дзеркального спуску, методу Корпелевич порівняти існуючі алгоритми, а також виділити найбільш оптимальний з них.

Об'єктом дослідження є алгоритм для визначення відстанні Вассерштейна. *Предмет* дослідження - оптимальний алгоритм визначенні відстані Вассерштейна, отриманий на основі розробленої програмної бібліотека для знаходження розв'язку варіаційної нерівності.

Методи розробки: порівняльний аналіз, комп'ютерне моделювання, розробка програмного продукту. Інструменти написання програми: мова програмування Python, середовище розробки Spyder та Jupyter Notebook.

Результати роботи: розроблено набір проекційних методів для розв'язування задачі варіаційної нерівності, з використанням алгоритмів Попова, Корпелевич, Ценга, операторної екстраполяції та алгоритмів Маліцького, проведено

тестування бібліотеки на модельних задачах, з подальшою оцінкою і аналізом результатів.

*Апробація результатів*Результати роботи було представлено на міжнародній науковій конференції молодих вчених "Шевченківська весна – 2021"(15 квітня 2021 р.) на та міжнародній конференції "Problems of decision making under uncertainties (PDMU-2021 14 травня 2021 р.).

Зміст

2 ВСТУП

Оптимальний транспорт та відстань Вассерштейна - це тісно пов'язані поняття, які мають довгу історію в математиці. Останнім часом активізувався інтерес до цієї теми, особливо в контексті різних наукових галузей, включаючи візуалізацію, статистику та машинне навчання. Завдяки різкому збільшенню обсягу даних та великому попиту на обчислення відстані Вассерштейна, збільшився інтерес до розробки більш ефективних алгоритмів в області оптимального транспорту та способів їх оптимізації.

Ця дипломна робота присвячена алгоритмам розв'язання оптимальної транспортної задачі та обчислення відстаней Вассерштейна. У роботі буде розглянуто як відомі методи, так і нові сучасні підходи обчислення дискретних задач оптимального транспорту.

Історія теорії оптимального транспортування сягає XVIII століття, коли французький математик Гаспард Монж опублікував свою роботу про постановку задачі на межі між аналізом та теорією ймовірностей, яку тепер називають формулюванням Монжа оптимального транспорту. Через два століття Леонід Канторович переглянув цю тему і доповнив її більш загальним викладом проблеми, що заклало основу для теорії оптимального транспортування. Сьогодні підхід Л.В. Канторовича широко застосовується у сучасних розробках.

У дискретному випадку обидві постановки задач (Монжа і Канторовича) дуже тісно пов'язані з добре вивченими задачами оптимізації. Наприклад, проблема присвоєння є окремим випадком формулювання Монжа, а проблему Канторовича можна розглядати як проблему мінімальних витрат на повній двосторонній мережі.

Концептуально ця проблема обертається навколо ідеї ефективної перебудови та діє на ймовірнісних вимірах. Окремі дослідники застосовують цю теорію в різних сферах, зокрема, для пошуку оптимального маршруту перевезення фізичних товарів, людей, перенесення частинок, а також в більш абстрактних поняттях, наприклад, знаходження відтінків сірого пікселів на зображеннях, класифікації нейронної мережі тощо.

Сьогодні розроблюються нові теоретичні підходи, сучасні методи вирішення оптимальних транспортних проблем. Однак застосування оптимальних транспортних задач або відстаней Вассерштейна стримується великим обчислювальним навантаженням. Незважаючи на можливість описувати та подавати оптимальний транспорт у формі простої лінійної програми, розміри вхідних даних часто значно перевищують межі здатності до їх обробки новітніми методами лінійного програмування. Ця проблема була частково вирішена завдяки досягненню регульованого оптимального транспорту та впровадженню алгоритму масштабування Сінкхорна до оптимального транспорту.

За допомогою теорії оптимального транспорту можна визначити, наприклад, зміну відстані об'єктів між різними зображеннями. Для цього необхідно перенести інтенсивність кожного пікселя в будь-яке інше місце, при цьому, якщо пересунити їх занадто далеко, можна отримати більшу похибку, ніж, якщо здійснювати менше переміщення, або взагалі не пересувати піксель.

Крім знаходження відстані, є можливість обчислювати зважений центр мас (барицентр) декількох зображень, шляхом усереднення загального значення та положення зображення. Наприклад, можна багато разів написати на аркуші паперу певне слово, і, якщо знайти усереднене зображення, то можна отримати такий вигляд цього слова, що узагальнює всі попередні. Або, взявши два набори різних зображень, і, знайшовши їх усереднене значення, можна додати одне зображення й віднести його до однієї з груп зображень.

Одним із найпоширеніших способів застосування оптимального транспорту є їх використання у нейродослідженнях, зокрема, при дослідженні активності головного мозку, для пошуку аномалій чи відхилень від норми. Якщо серія зображень велика, то обробити їх традиційним способом буде важко. В таких випадках застосування алгоритмів оптимального транспортування є вирішенням ситуації. Вони можуть замість відстежування змін порівнювати сусідні кадри, але їм можуть заважати шуми і те, що зміни відбуваються не за один кадр. Іншими словами, алгоритм оптимального транспортування допомагає усереднити серію знімків, а потім порівнює знайдені "барицентри" і їх відмінності між собою.

Тому вивчення проблеми розробки алгоритму оптимального транспорту та визначення барицентра Вассерштейна є актуальним питанням сучасної математики, яке має абсолютно прикладне застосування в різних сферах.

Мета роботи - написати програмну бібліотеку, яка допоможе у розв'язанні варіаційної нерівності для задачі пошуку барицентра Вассерштейна, використовуючи алгоритми Попова, Корпелевич, Ценга, операторної екстраполяції та алгоритмів Маліцького (відбиваючого проєкційного алгоритму та методу дзеркального спуску) та порівняти результати на основі проведеного тестування для поставленої задачі.

Для досягнення зазначеної мети, необхідно виконати наступні *завдання*: охарактеризувати сучасні методи для алгоритмів пошуку барицентрів Вассерштейна, виділити алгоритми транспортих задач, визначити способи оптимізації алгоритма пошуку барицентра Вассерштейна за допомогою алгоритмів екстраполяції з минулого, дзеркального спуску, методу Корпелевич порівняти існуючі алгоритми, а також виділити найбільш оптимальний з них.

Об'єктом дослідження роботи буде знаходження "оптимального"алгоритму для визначення відстанні Вассерштейна. А *предмет* дослідження - оптимальний алгоритм визначенні відстані Вассерштейна, отриманий на основі розробленої програмної бібліотека для знаходження розв'язку варіаційної нерівності.

Результати роботи було представлено на міжнародній науковій конференції молодих вчених "Шевченківська весна – 2021"(15 квітня 2021 р.) на та міжнародній конференції "Problems of decision making under uncertainties-(PDMU-2021 (14 травня 2021 р.)).

2.1 Опис проблеми алгоритмів пошуку барицентра Вассерштейна

Пошук стохастичної рівноваги в багатостадійних моделях транспортних потоків призводить до вирішення наступної сідлової задачі з опукло-увігнутою структурою: де - $g(y)$ і $c_{ij} \geq 0$ – увігнуті гладкі функції, c_{ij} - випукла обмежена множина простої структури.

$$\min_{\sum_{j=1} z_{ij}=L_i, \sum_{i=1} z_{ij}=W_j} \max_{x \in Q} \left\{ \sum_{i,j=1}^{n} z_{ij} \ln z_{ij} + \sum_{i,j=1}^{n} c_{ij}(x) z_{ij} + g(x) \right\} =$$

$$= \max_{x \in Q} \max_{\lambda, \mu \in \mathbb{R}^n} \left\{ \langle \lambda, L \rangle + \langle \mu, W \rangle - \sum_{i,j=1}^{n} e^{(-c_{ij}(x) z_{ij} + \lambda_i + \mu_j)} + g(x) \right\}.$$

Зрозуміло, що система балансових обмежень несумісна $\sum_{i=1}^{n} L_i \neq \sum_{j=1}^{n} W_j$, або вироджена. В останньому випадку це призводить до того, що двоїсті змінні λ, μ визначені з точністю до довільної сталої $\tilde{\alpha}$:

$$(\lambda + \tilde{\alpha} e, \mu - \tilde{\alpha} e), e = (1, ..., 1) \in \mathbb{R}^n.$$

Без обмеження загальності покладемо $\sum_{i=1}^{n} L_i = \sum_{j=1}^{n} W_j = 1$. Тоді задачу можна переписати таким чином:

$$\min_{\sum_{j=1} z_{ij}=L_i, \sum_{i=1} z_{ij}=W_j} \max_{x\in Q} \left\{ \sum_{i,j=1}^{n} z_{ij} \ln z_{ij} + \sum_{i,j=1}^{n} c_{ij}(x) z_{ij} + g(x) \right\} =$$

$$= \max_{x\in Q} \max_{\tilde{\lambda},\tilde{\mu}\in\mathbb{R}^n} \left\{ \langle \tilde{\lambda}, L\rangle + \langle \tilde{\mu}, W\rangle - \sum_{i,j=1}^{n} e^{(-c_{ij}(x)z_{ij}+\tilde{\lambda}_i+\tilde{\mu}_j)} + g(x) \right\} =$$

$$= \max_{x\in Q} \max_{\lambda,\mu\in\mathbb{R}^n} \left\{ \langle \lambda, L\rangle + \langle \mu, W\rangle + 1 - \sum_{i,j=1}^{n} e^{(-c_{ij}(x)z_{ij}+\lambda_i+\mu_j)} + g(x) \right\} =$$

$$= -\min_{x\in Q}(f(x) - g(x)),$$

де опуклу функцію $f(x)$ можна визначити як:

$$f(x) = \min_{\sum_{j=1} z_{ij}=L_i, \sum_{i=1} z_{ij}=W_j} \sum_{i,j=1}^{n} z_{ij} \ln z_{ij} + \sum_{i,j=1}^{n} c_{ij}(x) z_{ij} =$$

$$= \max_{\lambda,\mu\in\mathbb{R}^n} \left\{ \langle \lambda, L\rangle + \langle \mu, W\rangle - \sum_{i,j=1}^{n} e^{(-c_{ij}(x)z_{ij}+\lambda_i+\mu_j)} \right\}.$$

Окрім цього, основну задачу можна переписати у вигляді:

$$\min_{\sum_{j=1} z_{ij}=L_i, \sum_{i=1} z_{ij}=W_j} \max_{x\in Q} \left\{ \sum_{i,j=1}^{n} z_{ij} \ln z_{ij} + \sum_{i,j=1}^{n} c_{ij}(x) z_{ij} + g(x) \right\} =$$

$$= \max_{x\in Q} \max_{\lambda,\mu\in\mathbb{R}^n} \left\{ \langle \lambda, L\rangle + \langle \mu, W\rangle - \ln(\sum_{i,j=1}^{n} e^{(-c_{ij}(x)z_{ij}+\lambda_i+\mu_j)}) + g(x) \right\} =$$

$$= -\min_{x\in Q}(f(x) - g(x)),$$

І функцію $\hat{f}(x)$ можна позначити через:

$$\hat{f}(x) = \min_{\sum_{j=1} z_{ij}=L_i, \sum_{i=1} z_{ij}=W_j} \sum_{i,j=1}^{n} z_{ij} \ln z_{ij} + \sum_{i,j=1}^{n} c_{ij}(x) z_{ij} =$$

$$= \max_{\lambda,\mu\in\mathbb{R}^n}\left\{\langle\lambda, L\rangle + \langle\mu, W\rangle - \ln(\sum_{i,j=1}^{n} e^{(-c_{ij}(x)z_{ij}+\lambda_i+\mu_j)})\right\}.$$

Оскільки, додана умова $\sum_{i,j=1}^{n,n} z_{ij} = 1$ є наслідком балансових рівнянь, то це призводить до того, що дуальні (*) змінні (λ, μ) визначені з точністю до двох довільних постійних.

3 ПОСТАНОВКА ЗАДАЧІ

Варіаційна нерівність забезпечує універсальну структуру для вирішення питань оптимізації чи рівнянь із нерухомою точкою.

Основною задачею, буде знаходження розв'язку варіаційних нерівностей, тобто нерівностей що мають наступний вигляд:

$$\langle G(x), x' - x \rangle \geq 0, \forall x' \in X \tag{3.1}$$

X - підмножина гільбертового простору (будемо вважати, що вона опукла та замкнена), G - оператор, що діє з X у гільбертів простір (надалі припускаємо, що він монотонний та L-ліпшиців), тобто такий, що задовольняє умові:

$$||G(x_1) - G(x_2)|| \leq L||x_1 - x_2|| \ \ \forall x_1, x_2 \in X \tag{3.2}$$

Така задача (3.1) дуже тісно пов'язана, з знаходженням критерію оптимальності в наступній задачі,

$$f(x) \to \min_{x \in X} \tag{3.3}$$

Інакше кажучи

$$x^* - \text{розв'язок задачі} \Leftrightarrow \langle \nabla f(x^*), x - x^*) \rangle \geq 0 \ \ \forall x \in X \tag{3.4}$$

Варіаційні нерівності, що задовольняють (3.2), часто називають гладкими. Крім того, припускаємо, що G задовольняє узагальнену умову монотонності

$$\langle G(x), x' - x \rangle \geq \mu||x' - x||^2, \forall x' \in X \tag{3.5}$$

для деякого $\mu \geq 0$. У цій роботі ми припускаємо існування розв'язку x' задачі (3.1) -(3.5). Очевидно, умова (3.5) виконується, якщо G монотонна, тобто існує $\mu \geq 0$, так що

$$\langle G(x_1) - G(x_2), x_1 - x_2 \rangle \geq \mu||x_1 - x_2||^2, \forall x_1, x_2 \in X \tag{3.6}$$

Зокрема, G є сильно монотонним, якщо $\mu > 0$ в (3.6). Однак умова (3.5) не обов'язково означає монотонність G. Наприклад, псевдомонотонні варіаційні нерівності задовольняють (3.5), але не (3.6). Поняттям, що стосується сильної монотонності, є наступна умова для деякого $\mu > 0$ так, що

$$\langle G(x'), x - x\rangle \geq \mu||x - x'||^2, \forall x \in X \tag{3.7}$$

Зауважимо, що якщо G є монотонним і задовольняє (3.7), то (3.5) повинно виконуватися.

Тому сильно монотонні варіаційні нерівності охоплюють багато проблем варіаційних нерівностей, які були вивчені в літературі. Як окремий випадок монотонних варіаційніих нерівностей можна виокремити задачу (3.1) - (3.5) монотонної варіаційної нерівності, при $\mu > 0$ в (3.5).

3.1 Сучасні підходи до визначення відстані Вассерштейна

Барицентри Вассерштейна забезпечують геометрично значущий спосіб агрегування розподілів ймовірностей, побудований на теорії оптимального транспорту.

Визначення 3.1. *Нехай (M, d)- метричний простір. Для $p \geq 1, \exists x_0$*

$$\int_M d(x, x_0)^p \,\mathrm{d}\mu(x) < +\infty \tag{3.8}$$

Тоді p-ю метрику Васерштейну $W_{(\mu,\nu)}$ для двох ймовірносних мір μ та ν можна визначити як:

$$W_p(\mu, \nu) := \left(\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{M\times M} d(x, y)^p \,\mathrm{d}\gamma(x, y)\right)^{1/p}, \tag{3.9}$$

де $\Gamma(\mu, \nu)$ визначається як сукупність усіх мір по $M \times M$ з розподілами μ та ν.

Визначення 3.2. *Нехай* (M, d) *- метричний простір. Для* $p \geq 1, \exists x_0$

$$\sum_i d^p(x_i, x_0)\pi_{i,0} < +\infty \tag{3.10}$$

відстань Вассерштейна можна розглядати як:

$$W_p(\mu, \nu) = \left(\min_{\pi \in \Pi(\mu,\nu)} \sum_{ij} d^p(x_i, y_j)\ pi_{i,j}\right)^{\dfrac{1}{p}}.$$

Визначення 3.3. *Барицентром Вассерштейна називається метрика* N $\{\nu_1, ..., \nu_N\}$ *у* $\mathbf{P} \subset P(\Omega)$, *що мінімізує функціонал* f *над* $\mathbf{P}$, *де*

$$f(\mu) = \frac{1}{N}\sum_{i=1}^{n} W_p(\mu, \nu_i).$$

В теоретичному розділі дипломної роботи розглянуто алгоритми, які будуть протестовані в практичній частині.

Окремі методи вирішення оптимальних транспортних проблем розроблювалися давно. Однак багато алгоритмів, що використовуються в сучасних додатках, були розроблені протягом останнього десятиліття. У 2018 році Габріель Пейре та Марко Кутурі випустили книгу „Обчислювальний оптимальний транспорт“, яка містить більшість відповідних алгоритмів оптимального транспорту, яка надає вичерпний огляд різних типів методів та додатків.

Нехай маємо 2 вектори r, c з $n-$вимірного ймовірнісного простору, що представдений симплексом Δ^n та матрицею витрат $C \in \mathbb{R}_+^{n\times n}$, тоді задача оптимального транспотру виглядатиме так:

$$\min_{X \in \mathcal{U}_{r,c}} \langle C, X \rangle\,,\ \text{де}\ \mathcal{U}_{r,c} = \{X \in \mathbb{R}_+^{n\times n}, \mathbf{1}X = r, X^\top \mathbf{1} = c\} \tag{3.11}$$

Ця задача виникає внаслідок визначення відстані Вассерштайна або Earth‘s mover distance між дискретними r і c, як "найдешевший"зв’язок між розподілами, де вартість зв’язку $X \in \mathcal{U}_{r,c}$ дорівнює $\langle C, X\rangle$. Якщо r і c розглядати як розподіл ваги, розміщених на n точках в деякому просторі (зазвичай, метричному), відстань Вассерштайна - це найдешевший спосіб переміщення з r у c. У (3.11) X представляє переміщення (X_{ij} - сума, переміщена з r_i в c_j) і C являє собою вартість руху (C_{ij} - вартість переміщення з r_i до c_j).

Назвемо $\hat{X} \in \mathcal{U}_{r,c}$ ε – наближеним планом перенесення, якщо

$$\left\langle C, \hat{X} \right\rangle \leq \min_{X \in \mathcal{U}_{r,c}} \langle C, X \rangle + \varepsilon.$$

Нехай d – векторизована матриця витрат C, а Δ^{n^2} - n^2– вимірний симплекс. Знаючи, що r, c задані суми рядків і стовпців, так що $1^\top r = 1^\top c = 1$. Тоді задачу оптимального транспорту можна сформулювати як:

$$\min_{x \in \Delta^n, Ax=b} d^\top x \tag{3.12}$$

де $A \in \{0,1\}^{2n \times m}$ $n^2 = m$- матриця інцидентності $b \in \mathbb{R}_0^{2n}$, де $\mathbb{R}_0^{2n}$ - множина дісних невід'ємних чисел, а $\mathbf{1}$ – одиничний вектор.

Визначення 3.4. *Симплекс - опукла оболонка з $n+1$ точки, які не лежать в одній $n-1$ -вимірній гіперплощині.* $\Delta = \left\{ \sum_{i=0}^{n} x_i A_i : \left(\sum_{i=0}^{n} x_i = 1 \right) (\forall i \ x_i \geq 0) \right\}$. *Вважатимемо, ці точки x_i вершинами симплексу.*

Ймовірносний симплекс $\Delta^d = \{v | v \in \mathbb{R}_+^d \geq, 1^\top = 1\}$.

Для $x \in \Delta^d, h(x) = \sum_{i \in d} x_i \ln x_i$ є ентропією, припустивши, що $0 \ln_0 = 0$. Визначимо поняття дивергенції Брегмана та прокс-оператора дивергенції.

Визначення 3.5. *Дивергенцію Брегмана можна визначити як:*

$$V_g(x,z) := V(x,z) = g(x) - g(z) - \langle \nabla g(z), x - z \rangle \quad \forall x, z \in X \tag{3.13}$$

де g - сильно опукла на $X : \mathbb{R}$, диференційована на X

А також для сильно опуклої g справедливо:

$$V_g(x, y) = \frac{1}{2}||x - y||^2. \tag{3.14}$$

Визначивши відстань Брегмана, можна замінити узагальнене припущення про сильну монотонність в (3.5) за

$$\langle \nabla G(x), x' - x \rangle \geq 2\mu V(x, x') \ \ \forall x \in X \tag{3.15}$$

Визначення 3.6. *Представимо проксимальний оператор у вигляді*

$$Prox(x, y, \gamma) = \operatorname*{argmin}_{z \in X} \{ \langle \nabla y, z \rangle + \frac{1}{\gamma} V(z, x) \},$$

при $g = \frac{1}{2}||x||^2$ *проксимальний оператор дорівнює оператору проектування* X_i

$$Prox(x, y, \gamma) = P_X(x - y\gamma)$$

Деякі змінні мають спеціалізоване значення. Всі графіки, що розглядаються, матимуть $2n$ вершини з m ребрами, тобто $m = n^2$. $A \in R^{2n \times m}$ - матриця. d - матриця векторизованих витрат C. b - вектор обмеження, що об'єднує обмеження r, c. В алгоритмах для знаходження розв'язку, x та y є прямими (у симплексі) та двоїстими змінними відповідно, де змінна $x \in \Delta m$ є вектор. $\mathcal{U}_{r,c}$ є можливий багатокутник: коли область визначення є векторами, $\mathcal{U}_{r,c}$ дорівнює $x | Ax = b$, а коли це матриці, $\mathcal{U}_{r,c}$ дорівнює $X | X\mathbf{1} = r, X^\top \mathbf{1} = c$ (при згладжуванні X це узгоджується).

Для зручності запишемо $\min\limits_{X \in \mathcal{U}_{r,c}} \langle C, X \rangle$ як

$$\min_{x \in \Delta^m} d^\top x + 2||d||_\infty ||Ax - b|| \ \ [1] \tag{3.16}$$

Задачу (3.16) можна звести до мінімаксної, враховуючи властивості з [11]. Для $F(x, y)$, опуклої в x і увігнутої в y, стандартний спосіб вимірювання розриву двоїстості полягає у визначенні оператора градієнта $g(x, y) = (\nabla_x F(x, y), -\nabla_y F(x, y))$, і доведенні того, що для $z = (x, y)$ та будь-якого u на просторі, залишок, $\langle g(z), z - u\rangle$ невеликий. Відповідно, визначаємо

$$g(x, y) = (d + 2||d||_\infty A^\top y, 2||d||_\infty (b - Ax)) \tag{3.17}$$

4 ТЕОРЕТИЧНІ ВІДОМОСТІ

4.0.1 Необхідні визначення

Для нас будуть важливі наступні визначення:

Визначення 4.1. *Функція f називаєтьcя L-ліпшицевою з множини S щодо норми $||\cdot||$ тоді, коли для будь-яких точок $u, w \in S$ виконується $|f(u) - f(w)| \leq L||u - w||$.*

Визначення 4.2. *Функція $f = f(x)$ назвемо μ-*сильно опуклою *для $\exists \mu > 0$ якщо*

$$f(\alpha \cdot x + (1-\alpha) \cdot y) \leq \alpha \cdot f(x) + (1-\alpha) \cdot f(y) - \mu \cdot \alpha \cdot (1-\alpha) \cdot ||x-y||^2. \quad (4.1)$$

Або:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \mu \cdot ||x - y||^2. \quad (4.2)$$

Визначення 4.3. *Кажемо, що оператор G: $H \to H$* монотонний *якщо*

$$\langle G(x) - G(y), x - y \rangle \geq 0 \quad \forall x \, \forall y. \quad (4.3)$$

Аналогічно розгядається поняття μ-сильно опуклого оператора

Визначення 4.4. *Оператор G: $H \to H$ називаєтьcя μ-сильно монотонним зі сталою $\mu > 0$ якщо*

$$\langle G(x) - G(y), x - y \rangle \geq \mu \cdot ||x - y||^2 \quad \forall x \, \forall y. \quad (4.4)$$

Подивимося на задачу знаходження розв'язку варіаційної нерівності як на задачу знаходження нерухомої точки оператора

$$T : x \mapsto P_C\left(x - \rho G(x)\right), \tag{4.5}$$

Або

$$x_{k+1} = P_C\left(x_k - \rho_k G(x_k)\right). \tag{4.6}$$

Для оператора G – μ-сильно монотонний та L-ліпшиців можна отримати наступну оцінку:

$$\begin{aligned}
||T(x) - T(x)||^2 &\leq ||x - y - \rho(G(x) - G(y))||^2 \leq \\
&\leq ||x-y||^2 - 2\rho\left\langle G(x) - G(y), x - y\right\rangle + \rho^2||G(x) - G(y)||^2 \leq \\
&\leq ||x-y||^2 - 2\rho\mu||x-y||^2 + \rho^2||G(x) - G(y)||^2 \leq \\
&\leq ||x-y||^2 - 2\rho\mu||x-y||^2 + \rho^2L^2||x-y||^2 = \\
&= (1 - 2\rho\mu + \rho^2L^2)||x-y||^2 = \eta \cdot ||x-y||^2.
\end{aligned} \tag{4.7}$$

де $\eta(\rho) \in (0, 1)$.

Регуляризація - процес заміни довільної опуклої функції f на ε-сильно опуклу функцію $f_\varepsilon = f + \varepsilon||x||^2$.

Отже, й монотонний оператор G можна подати, як ε-сильно опуклий оператор $G_\varepsilon = G + \varepsilon\mathbf{1}$, де $\mathbf{1}x \mapsto x$ – одиничний оператор. Тоді варіаційну нерівність, можна записати для $\langle G(x) - G(y), x - y\rangle \geq 0$ у такій формі

$$\langle G_\varepsilon(x) - G_\varepsilon(y), x - y\rangle = \langle G(x) - G(y), x - y\rangle + \varepsilon\cdot||x-y||^2 \geq \varepsilon\cdot||x-y||^2, \tag{4.8}$$

У цій роботі ми спробуємо побудувати алгоритми з регуляризацією, для яких ітерація матиме наступний вигляд:

$$x_{k+1} = P_C\left(x_k - \alpha_k G_{\varepsilon_k} x_k\right), \tag{4.9}$$

Для регуляризованих варіантів алгоритмів, нам також буде необхідно визначити послідовність α_n, таку що $\alpha_n \in (0, 1)$, $\sum\limits_{n=0}^{\infty} \alpha_n = +\infty$, $\lim\limits_{n\to\infty} \alpha_n = 0$

4.1 Алгоритми розв'язання транспортної задачі

4.1.1 Лінійне програмування

Оскільки формулювання Канторовича в дискретному середовищі є скінченною лінійною програмою, тому для розв'язання транспортної задачі застосовуються стандартні алгоритми, такі як симплекс-метод. Однак це формулювання має дуже особливу форму як випадок проблеми мережі з мінімальними витратами, що дозволяє використовувати більш спеціалізовані алгоритми для її вирішення.

Переозначимо: $\min\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\pi_{ij}$ замінивши $\sum_{j=1}^{n}\pi_{ij} = \mu = 1, \sum_{i=1}^{m} = \nu \ \forall i = \overline{1,n}, j = \overline{1,m}$.

Де c_{ij} - елементи матриці витрат $C \in \mathbb{R}_+^{n\times m}$ та міри μ і ν задовольняють $\sum_i \mu_i = \sum_j \nu_j$, щоб задача була здійсненною. Двоїсте формулювання подано наступним чином: $\max\sum_{i=1}^{n} u_i\mu_i + \sum_{j=1}^{m} v_j\mu_j$ за умови $u_i + v_j \leq c_{ij} \ \forall i, j$. Двоїсті змінні u_i та v_j також називаються потенціалами. Це формулювання використовується у більшості методів лінійного програмування. Додаткова умова нежорсткості $(c_{ij} - u_i - v_j) \cdot \pi_{ij} = 0 \ \forall i, j$. Це означає, що коли знаходиться первинне можливе рішення π та двоїсте можливе рішення (u, v), що задовольняє наведеній вище умові, розв'язок повинен бути оптимальним.

Розглянемо такі методи, як угорський метод для завдань присвоєння та транспортний симплекс, а також більш сучасні модифікації, такий як метод короткого списку.

Угорський метод розробили два математика Кун та Манкрес в 1955 р. Цей алгоритм є частинним випадком транспортної задачі з $n = m$ та $\mu_i = \nu_j = 1 \forall i, j$. Тут одне джерело призначається саме одному стоку таким чином, щоб загальна вартість задачі була мінімізована. Алгоритм починається з визначення двоїстого можливого рішення $(\hat{u}, \hat{v})$ через $\hat{u}_i := \min_{j=\overline{1,n}} c_{ij}; \overline{v}_j := \min_{i=\overline{1,n}} c_{ij} - \hat{u}_i \forall i, j = \overline{1, n}$.

Матриця приведених витрат $\hat{C}$ визначається за $\hat{c}_{ij} := c_{ij} - \hat{u}_i - \hat{v}_j$. Тепер, якщо можна знайти таке присвоєння, що кожне ненульове введення β відповідає нульовому входу $\hat{C}$, завдяки додатковій нежорсткості це призначення є оптимальним.

Якщо ні, розглядається мінімальна кількість рядків і стовпців, необхідна для покриття всіх нульових записів у $\hat{C}$, фіксуючи одну таку оболонку та визначаючи набори непокритих рядків і стовпців як S_r та S_c, відповідно. Далі знаходиться мінімальний непокритий запис в $\hat{C}$ ($c_0 := \min_{i \in S_r, j \in S_c} \hat{c}_{ij} > 0$). Отримується нове можливе двоїсте рішення $(\overline{u}, \overline{v})$, встановлюючи $\overline{u}_i = \hat{u}i + c_0$ для $i \in S_r, \overline{u}_i = \hat{u}_i$, $\overline{u}_i = \hat{u}_i$ для $i \notin S_r, \overline{v}_j = \hat{v}_j$ для $j \in S_c$ та $\overline{v}_j = \hat{v}_j - c_0$ для $j \notin S_c$ і, в свою чергу, нова матриця знижених витрат $\overline{C}$. Цей процес повторюється, поки не будуть виконані додаткові умови нежорсткості, і отримане оптимальне призначення через нульові записи матриці знижених витрат.

Транспортний симплекс - це спеціальна версія симплекс мережі. Як і інші симплекс-методи, транспортний симплекс має дві фази: одну фазу для побудови початкового базового можливого рішення π та другого етапу для вдосконалення цього рішення до оптимальності. Як правило, більшість часу витрачається на другу фазу, оскільки початкове рішення оптимального транспорту легко отримати.

Транспортний симплекс також розглядається як модифіковане правило мінімуму рядків, що має універсальну продуктивність як за часом виконання, так і за якістю побудованого рішення. Перебираючи всі місця розташування джерел (рядки) $x_i \in X$, які все ще мають масу, для кожного джерела вибирається доступна ціль c_j з найменшими витратами та включається в рішення, встановлюючи π_{ij} на найбільше можливе значення. Цей процес повторюється, поки всі джерела не будуть вичерпані. Отримане таким чином рішення автоматично стає базовим.

Якщо інтерпретувати вихідне та цільове розташування як вузли в графіку та малювати ребра для кожного можливого транспорту, можна отримати повний двосторонній графік. Кожне невироджене базове можливе рішення тепер може бути представлене охопленням дерева на цьому графіку, вибравши всі дуги, що належать до активних транспортів. З огляду на базове можливе рішення, симплексний крок виконується наступним чином:

1. Для введення базису вибирається нова змінна (π_{ij})
2. Створюється цикл у попередньому дереві, який потім ідентифікується.
3. Максимальна кількість маси зміщується по цьому циклу, тобто, по черзі додається і віднімається від послідовних перевезень.
4. Змінна, яка в процесі стала нульовою, вилучається з базису.

Для пошуку нової базової змінної є багато варіантів, але мінімальна стратегія рядка є найбільш ефективною.

Використавши поточне рішення для обчислення значень двоїстих змінних u_i та v_j через додаткові умови нежорсткості, що $u_i + v_j = c_{ij}$, коли завгодно $\pi_{ij} > 0$. Це лінійна система рівнянь $n + m - 1$ із $n + m$ змінними, яку можна вирішити за допомогою підстановки назад після встановлення $u_1 = 0$. Тоді, розглянувши неосновні змінні π_{ij}, рядок за рядком можна обрахувати зменшені витрати $r_{ij} = c_{ij} - u_i - v_j$. У випадку отримання змінної з негативними зменшеними витратами, небхідно зупитинися в кінці цього рядка і вибрати змінну з найменшим зменшенням вартості, що встановлюється як нова базова змінна. Якщо кандидати не знайдені серед усіх рядків (поточне рішення є оптимальним), то відбувається зупинка роботи алгоритму.

Друга фаза транспортного симплексу дуже схожа на мережевий симплекс. Різниця лише в тому, що ми завжди маємо повну двосторонню мережу в оптимальній транспортній проблемі. На першому етапі ця конструкція є дуже вигідною і дозволяє легко знайти рішення, тоді як у більш загальному випадку для введення штучного базису часто потрібні додаткові змінні.

Варіацією транспортного симплекс-методу є метод короткого списку, який містить три параметри: s - довжина короткого переліку, p, k - кількість змінних, які визначаються, щоб знайти нову базову змінну.

Перед початком оптимізації створюється короткий список для кожного джерела, що складається з цілей з найменшими транспортними витратами, упорядкованими за вартістю. Основне можливе рішення за аналогією знаходиться за модифікованим мінімальним правилом рядка, де списки мають пріоритет. Після цього отримане рішення вдосконалюється за допомогою симплексних кроків, схожих на транспортний симплекс, але пошук здійснюється за обмеженими списками.

Списки перевіряються до тих пір, поки не з'являться будь-які k змінні з негативно зменшеними витратами, або не знайдено p відсотків короткого списку. Потім для введення в базис обирається кандидат з мінімально зниженою вартістю. Якщо в короткому списку досягти покращення вже неможливо, то останнє рішення визначається як оптимальне.

4.1.2 Ентропійно-регуляризаційна транспортна задача

Ідея регулювати оптимальні транспортні проблеми, додавши ентропійний множник, була запропонована в 1969 р., але лише в 2013 році завдяки розробкам Кутурі така можливість набула значної переваги в теорії обчислювального оптимального транспорту. Він запропонував розглядати ентропічно-регульовані оптимальні транспортні проблеми за допомогою алгоритму матричного масштабування. Цей алгоритм був запропонований в 1967 році і називається масштабуванням Сінкхорна або Сінкхорна-Ноппа. Саме через це ентропія використовується у регуляризації і регуляризований оптимальний транспорт розглядається як алгоритм.

Розглянемо дискретний випадок проблеми Канторовича. Для зв'язки, π розглядається як ентропійний член регуляризації

$$H(\pi) := -\sum_{ij} \pi_{ij}(\ln(\pi_{ij} - 1)).$$

Для параметра регуляризації $\lambda > 0$, ентропічно-регуляризована оптимальна транспортна задача буде мати вигляд

$$\min_{\pi \in \Pi(\mu,\nu)} \sum_{ij} c_{ij}\pi_{ij} - \lambda H(\pi).$$

Оскільки ентропія є строго увігнутою функцією, це призводить до того, що задача регуляризованого транспорту є опуклою. Отже, допускається єдиний розв'язок $\pi_\lambda^* \ \forall \lambda > 0$. Це рішення завжди має повний носій, на відміну від основних рішень вихідної проблеми, отриманих за допомогою лінійного програмування, які є рідкісними.

Якщо λ велике, тоді опуклість сильніша (цільова функція λ -сильно опукла). Це відображається на розв'язку π_λ^*, коли $\lim\limits_{\lambda\to\infty} \pi_\lambda^* = \mu \times \nu$, оскільки

$\lim\limits_{\lambda \to 0} \pi^*_\lambda = \pi^*$, є оптимальним рішенням вихідної транспортної проблеми з найвищою ентропією.

У випадку Вассерштейна $c = d^p$, можна розглядати результати регуляризованої задачі як дещо іншу оптимальну транспортну відстань

$$W_{p,\lambda}(\mu, \nu) = \left(\min_{\pi \in \Pi(\mu,\nu)} \sum_{ij} d^p(x_i, y_j)\pi_{i,j} - \lambda H(\pi) \right)^{\dfrac{1}{p}},$$

що можна вважати як наближення до відстані Вассерштейна, особливо для малих λ, оскільки $W_{p,\lambda}(\mu, \nu) \to W_p(\mu, \nu)$, при $\lambda \to 0$.

Алгоритм масштабування Сікхорна.

Для фіксованого значення $\lambda > 0$, ядро Гіббса, пов'язане з матрицею витрат C, становить $K \in \mathbb{R}^{n \times m}$, де,

$$K_{ij} = e^{-\dfrac{c_{ij}}{\lambda}},\ i = \overline{1, n}; j = \overline{1, m}.$$

Основний факт, який використовується алгоритмом, полягає в тому, що єдиний розв'язок π^*_λ має вигляд $\pi^*_{\lambda ij} = u_i K_{ij} v_j$ з векторами $u \in \mathbb{R}^n, v \in \mathbb{R}^m$, подібними до двоїстих векторів у лінійних програмах. Оскільки π^*_λ має бути поєднано, потрібно знайти вектори u і v, такі, щоб $\pi^*_\lambda \in \Pi(\mu, \nu)$, тобто u і v, задовольняли обмеженням $diag(u)Kv = \mu$ і $diag(v)K^t u = \nu$, де μ і ν інтерпретуються як масові вектори, а $diag(u)$ - це матриця n на n, яка має записи u на діагоналі та нуль в будь-якому іншому місці.

Обмеження в u і v - нелінійні. Загальна стратегія виконання цих обмежень полягає в тому, щоб починати обчислення з довільного додатного вектора (наприклад, $v^{(0)} = (1, ..., 1)^t$) та по черзі оновлювати u та v.

Це досягається за допомогою ітерацій

$$u^{k+1} := \frac{\mu}{Kv^k};\ v^{k+1} := \frac{\nu}{K^t u^{k+1}},$$

де розподіл виконується почергово. По суті, це метод змінних проекцій - поточне рішення проектується по черзі на множини рішень (u, v), що задовольняють обмеженням джерела $diag(u)Kv = \mu$ та обмеженням стоку $diag(v)K^t u = \nu$.

Цей алгоритм широко застосовується на практиці, де необхідні широкомасштабні обчислення або наближення оптимальних транспортних відстаней, наприклад, при обчисленні барицентрів Васерштайна. Однак, проблемою алгоритму є, зокрема, числова стабільність. Як згадувалося раніше, апроксимація відстані Вассерштайна через регуляризовану версію краща, при малому λ.

У цьому алгоритмі є дві задачі з вибором λ:

1. Маленьке λ призводить до задачі зі слабкою опуклістю. Отже, збіжність $\pi_\lambda^{(k)} \to \pi_\lambda^*$, повільніша.

2. Ядра Гіббса швидко наближаються до нуля при малих λ.

Це призводить до чисельно нестабільних ітерацій масштабування матриці, що особливо проблематично, коли K наближається або падає нижче порогу точності машини. Це означає, що λ не можна обирати довільно малим.

4.1.3 Двоїста екстраполяція

Алгоритм пошуку пари (x, y) з невеликим розривом двоїстості щодо цілі описується таким чином:

$$\min_{x \in X} \max_{y \in Y} d^\top x + 2||d||_\infty (y^\top Ax - b^\top y), X = \Delta^m; Y = [-1; 1]^{2n}.$$

Основною ідеєю є розгляд алгоритму, який пов'язує його з аналізом двоїстої екстраполяції, тобто алгоритмом наведення приблизних сідлових точок з більш стандартним аналізом.

Для $F(x, y)$, опуклої в x і увігнутої в y, стандартний спосіб вимірювання розриву двоїстості полягає у визначенні оператора градієнта $g(x, y) = (\nabla_x F(x, y), -\nabla_y F(x, y))$, і доведенні того, що для $z = (x, y)$ та будь-якого u на просторі, залишок, $\langle g(z), z - u\rangle$ невеликий. Відповідно, визначаємо

$$g(x, y) = (d + 2||d||_\infty A^\top y, 2||d||_\infty (b - Ax))$$

Система двоїстої екстраполяції вимагає регуляризатора простору. Реалізація алгоритму полягає у тому, що для кожної ітерації потрібно два „дзеркально“ схожих кроки, що забезпечують стан s_t у спряженому просторі. Типовою установкою є оператор градієнта Ліпшица та регуляризатор, який є сумою канонічних сильно опуклих регуляторів у нормах, що відповідають добутку множин X, Y.

За визначенням, регулятор r k-вимірно-опуклий щодо оператора g, якщо для будь-яких точок a, b, c у своїй області,

$$k\left(r(a) + r(b) + r(c) - 3r\left(\frac{a + b + c}{3}\right)\right) \geq \langle g(b) - g(a), b - c\rangle$$

Вимірна-опуклість названа так, оскільки $\langle g(b) - g(a), b - c\rangle$ можна розглядати як вимірювання „площі“ трикутника з вершинами a, b, c відносно деякої матриці Якобіана. У випадку білінійних фунцкій ліва сторона у визначенні вимірної опуклості є незмінною до перестановки a, b, c, тоді як знак правої сторони можна змінити, помінявши місцями a, c, тому і передбачається опуклість.

Однак це не означає, що регуляризатор r сильно-опуклий, що є типовим припущенням про збіжність методу дзеркального спуску. Сформулюємо алгоритм для T. Єдина відмінність від попередньо створених алгоритмів – коефіцієнт 2 при s_{t+1}, тобто відбувається додавання кратного $1/2k$ замість $1/k$.

Алгоритм двоїстої екстраполяції (k, r, g, T).

Ініціалізуємо $s_0 = 0$, нехай $\bar{z}$ - мінімізація r.

for $t < T$ **do**

$z_t \leftarrow \mathrm{Prox}_z^r(s_t)$

$w_t \leftarrow \mathrm{Prox}_z^r(s_t + \frac{1}{k} g(z_t))$

$s_{t+1} \leftarrow s_t + \frac{1}{2k} g(w_t)$

$t + 1 \leftarrow t$

end for

return $\bar{w} = \frac{1}{T} \sum_{t \in [T]} w_t$

Лема (Про збіжність двоїстої екстраполяції). Нехай r є k-вимірною опуклістю щодо g. Нехай для деяких $u, \Theta \geq r(u) - r(\bar{z})$. Тоді, $\bar{w}$ до алгоритму задовольняє

$$\langle g(\bar{w}, \bar{w} - w) \rangle \leq \frac{2k\Theta}{T}$$

На першому кроці, необхідно довести наступну нерівність:

$$\frac{1}{2k} \langle g(w_t), w_t - \bar{z} \rangle \leq \langle s_{t+1}, z_{t+1} - \bar{z} \rangle + V_x^r(z_{t+1}) - \langle s_t, z_t - \bar{z} \rangle - V_z^r(z_t)$$

Припустимо, що $c_t = \dfrac{z_t + w_t + z_t + 1}{3}$. Доведення випливає з мінімальності z_t відносно c_t, мінімальності w_t відносно z_{t+1} та опуклості площі щодо z_t, w_t та z_{t+1}. Відповідно,

$$\langle s_t, z_t\rangle + r(z_t) \le \langle s_t, c_t\rangle + r(c_t)$$

$$\langle s_t, w_t\rangle + \frac{1}{k}\langle g(z_t), w_t\rangle + r(w_t) \le \langle s_t, z_{t+1}\rangle + \frac{1}{k}\langle g(z_t), z_{t+1}\rangle + r(z_{t+1})$$

$$\frac{1}{k}\langle g(w_t) - g(z_t), w_t - z_{t+1}\rangle \le r(z_t) + r(w_t) + r(z_{t+1}) - 3r(c_t)$$

Підставивши перше рівняння у третє і використовуючи визначення c_t, маємо

$$\frac{1}{k}\langle g(w_t) - g(z_t), w_t - z_{t+1}\rangle \le r(w_t) + r(z_{t+1}) - 2r(z_t) + \langle s_t, w_t + z_{t+1} - 2z_t\rangle$$

Переставляючи друге рівняння, маємо

$$\frac{1}{k}\langle g(z_t), w_t - z_{t+1}\rangle \le r(z_{t+1}) - r(z_t) + \langle s_t, 2z_{t+1} - 2z_t\rangle$$

Додавши ці два рівняння, маємо

$$\frac{1}{k}\langle g(w_t), w_t - z_{t+1}\rangle \le 2r(z_{t+1}) - 2r(z_t) + \langle s_t, 2z_{t+1} - 2z_t\rangle$$

Ділимо на 2 і додаємо $\frac{1}{2k}\langle g(w_t), w_{t+1} - \bar{z}\rangle$ з обох боків, отримуємо шукану функцію. Тепер, визначимо потенційну функцію

$$\Phi_k = \frac{1}{2k}\sum_{t=0}^{k-1}\langle g(w_t), w_t - \bar{z}\rangle - \langle s_k, z_k - \bar{z}\rangle - V_z^r(z_k)$$

Тоді, Φ_k не зростає в k. Тому для будь-якого u, за визначенням Θ,

$$\frac{1}{T}\sum_{t=0}^{T-1}\langle g(w_t), w_t - u\rangle \le \frac{1}{T}\sum_{t=0}^{T-1}\langle g(w_t), w_t - \bar{z}\rangle + \frac{1}{T}\sum_{t=0}^{T-1}\langle g(w_t), \bar{z} - u\rangle + (\frac{2k\Theta}{T} - \frac{2kV_z(u)}{T}) =$$

$$= \frac{1}{T}\sum_{t=0}^{T-1}\langle g(w_t), w_t - \bar{z}\rangle + \frac{1}{T}\sum_{t=0}^{T-1}\langle g(w_t), \bar{z} - z_T\rangle + (\frac{2k\Theta}{T} - \frac{2kV_z(x_T)}{T})$$

$$= \frac{2k}{T} + \frac{2k\Theta}{T} \le \frac{2k\Phi_0}{T} + \frac{2k\Theta}{T} = \frac{2k\Theta}{T}$$

Нерівність у другому рядку використала визначення $z_T = \text{Prox}_z^T(\frac{1}{2k} \sum_{t \in [T-1]} g(w_t))$, а остання нерівність – $\Phi_T \leq \Phi_0$. Висновок про збіжність двоїстої екстраполяції випливає із визначення g (оскільки вона є лінійною).

4.1.4 Метод дзеркального спуску

Попередні аналізи градієнтного спуску припускають, що функція f є ліпшицевою щодо норми Евкліда. Якщо f є $(\sqrt{n}L)$ - ліпшицевою відносно $||\cdot||_2$, то коефіцієнт $\sqrt{n}$ може бути як завгодно великий.

Градієнтний спуск чітко не розмежовує основний та спряжений векторні простори, тобто ітерації x_i є векторами в $\mathbb{R}^n$, тоді як градієнти $\nabla f(x_i)$ є лінійними функціоналами в $\mathbb{R}^n$.

Отже, $\nabla f(x_i)$ знаходиться в спряженому просторі. Можна пов'язати $\mathbb{R}^n$ та спряжений до нього простір, оскільки вони ізоморфні (за допомогою операції транспонування). Тим не менше, градієнтний спуск обчислює лінійну комбінацію ітерації x_i та градієнта $\nabla f(x_i)$, не звертаючи уваги на те, що ці об'єкти лежать у різних векторних просторах.

Основна ідея методу дзеркального спуску полягає у чіткому розмежуванні основного та спряженого до нього просторів та у визначенні корисної бієкції між ними двома.

Бієкція визначається дзеркальною функцією $\Phi : \mathbb{R}^n \to \mathbb{R}$. Канонічним прикладом Φ є від'ємна функція ентропії.

Бієкція між основним простором і спряженим простором полягає в наступному: основна точка x відображається в двоїсту точку $\nabla\Phi(x)$. Для відображення зворотного з спряженого простору в простір ми будемо використовувати обернену до відображення $x \mapsto \nabla\Phi(x)$. (Насправді це зворотне відображення просто $y \mapsto \nabla\Phi^*(y)$.

Метод дзеркального спуску грунтується на двох основних ідеях:

Головна ідея. Замість того, щоб робити кроки градієнта в основному просторі, дзеркальний спуск робить кроки градієнта в спряженому просторі. Бієкція $\nabla\Phi$ та її оберенена $\nabla\Phi^*$ використовуються для відображення вперед і назад між точками.

Забезпечення доцільності. Метою є обмежена оптимізація над опуклою множиною X в основному просторі, тому проблема полягає в тому, що крок градієнта, можливо, створює точку поза X. Наступна основна ідея - спроектувати цю точку на X під дивергенцією Брегмана D_Φ.

Алгоритм методу дзеркального спуску..

Початкова точка - будь-яка $x_1 \in \mathcal{X} \cap \mathcal{D}$

for $i \in$ **N do**

$\hat{x}_i \leftarrow \nabla\Phi(x_i)$– відображення точки до спряженої

$\hat{y}_{i+1} \leftarrow \hat{x}_i - \eta g_i$ – градієнтний крок у спряженому просторі

$y_{i+1} \leftarrow \nabla\Phi^*(\hat{y}_{i+1})$ – спряжена точка до точки з $\mathcal{D}$

$x_{i+1} \leftarrow \Pi^{\Phi}_{\mathcal{X}\cap\mathcal{D}}(y_{i+1})$ – проекція нової точки на можливий область

end for

Теорема 4.1 (Про дзеркальний спуск)**.** *Для будь-якого $\eta > 0$ та дзеркального відображення $\Phi \to \mathcal{D} : \mathbb{R}$, такого, що Φ - строго опукла та диференційована на $\mathcal{D}$; спряжений простір це весь $\mathbb{R}^n$; а градієнт Φ розбігається на межі $\mathcal{D}$, а також $\mathcal{D}$- відкрита підмножина з $\mathbb{R}^n$.*

Область $X \subseteq \mathbb{R}^n$ задовольняє таким умовам: X- замкнена і опукла; $X \subseteq \hat{\mathcal{D}}$ замикання $\mathcal{D}$; $X \cap \mathcal{D} \neq \varnothing$.

$f_1, f_2, ...f_n \to \mathbb{R}$ - *опуклі фінкції.*

Нехай строго $\rho-$опукла функція з нормою $||\cdot||$. Тоді для алгоритма дзер-

кального спуску характерна оцінка виду

$$\sum_{i=1}^{t}(f_i(x_i)) \le \frac{D_\Phi(x^*, x_1)}{\eta} + \sum(\langle g_i, x_i - y_{i+1}\rangle) - \frac{\rho}{2\eta}||x_i - y_{i+1}||^2 \le$$

$$\le \frac{D_\Phi(x^*, x_1)}{\eta} + \frac{\eta}{2\rho}\sum_{i=1}^{t}||g_i||_*^2$$

Для довільного $z \in X$

$$f_i(x_i) - f_i(z) \le g_i \top (x_i - z)$$

З кроком градієнта $g_i = \dfrac{\hat{x}_i - \hat{y}_{i+1}}{\eta}$, так що

$$f_i(x_i) = \frac{1}{\eta}(\nabla\Phi(x_i) - \nabla\Phi(y_i))^\top (x_i - z) =$$

$$= \frac{1}{\eta}(\nabla\Phi(x_i, y_{i+1}) + D_\Phi(z, x_i) - D_\Phi(z, y_{i+1}))$$

Застосувавши узагальнену тотожність Піфагора, можна отримати наступне:

$$\le \frac{1}{\eta}(\nabla\Phi(x_i, y_{i+1}) + D_\Phi(z, x_i) - D_\Phi(z, x_{i+1}))$$

Просумувавши по i, отримаємо

$$\sum_{i=1}^{t}(f_i(x_i) - f_i(z)) \le \frac{1}{\eta}(D_\Phi(z, x_1) + \sum_{i=1}^{t}(x_i, y_{i+1}))$$

Суми з правої сторони обмежені сильною опуклістю

$$D_\Phi = \Phi(x_i) - \Phi(y_{i+1}) - \langle\nabla\Phi(y_{i+1}), x_i - y_{i+1}\rangle =$$

$$= \Phi(x_i) - \Phi(y_{i+1}) + \langle\nabla\Phi(x_i), y_{i+1} - x_i\rangle + \langle\nabla\Phi(x_i) - \nabla\Phi(y_{i+1}), x_i - y_{i+1}\rangle$$

|з сильної опуклості випливає| $\le -\dfrac{\rho}{2}||x_i - y_{i+1}||^2 + \eta\,\langle g_i, x_i - y_{i+1}\rangle$.

Поєднавши з виразом вище та припустивши, що $z = x^*$.

Далі, оскільки функція є L-ліпшицевою, то

$$D_\Phi(x_i, y_{i+1}) \le \eta||g_i||_*||x_i - y_{i+1}||^2 - \frac{\rho}{2}||x_i - y_{i+1}||^2 \le \frac{\eta^2||g_i||_*^2}{2\rho}$$

що й треба було довести.

4.2 Способи оптимізації алгоритма пошуку барицентра Вассерштейна

У цьому розділі ми розглянемо алгоритми для розв'язування варіаційної нерівності, що забезпечують слабку збіжність x_n, а також спробуємо, побудувати сильно збіжні алгоритми з використанням ітеративної регуляризації Бакушинського [10].

Для регуляризованих варіантів алгоритмів, нам також буде необхідно визначити послідовність α_n, таку що $\alpha_n \in (0,1)$, $\sum_{n=0}^{\infty} \alpha_n = +\infty$, $\lim_{n\to\infty} \alpha_n = 0$

4.2.1 Простий проекційний метод

Перевагою цього методу є його лінійна збіжність, а недоліком, сильна монотонність G. Нехай, α - стала оберненої сильної монотонності оператора, тоді можна записати алгоритм:

Initialization: $x_0 \in X; \lambda \in (0, \frac{2\alpha}{L^2})$
for $i \in \overline{0, t-1}$ **do**
$x_{i+1} = P_X(x_i - \lambda G(x_i))$
end for
return x_t

Вимагатимемо, щоб $\lim_{n\to+\infty} \frac{\alpha_{n+1} - \alpha_n}{\alpha_{n+1}} = 0$, для аналогічного методу з регуляризацією. Його алгоритм матиме наступний вигляд:

Initialization: $x_0 \in X; \lambda \in (0, \frac{2\alpha}{L^2})$
for $i \in \overline{0, t-1}$ **do**
$x_{i+1} = P_X(1 - \alpha_n)P_X(x_i - \lambda G(x_i))$
end for
return x_t

4.2.2 Відбиваючий проекційний метод

Цей метод було запропоновано Маліцьким у 2018 році[5]. Головною особливістю методу, є те, що знаходити значення оператора та проекції потрібно лише один раз.

Initialization: $x_0 \in X; \lambda \in (0, \frac{\sqrt{2}-1}{L})$
for $i \in \overline{0, t-1}$ **do**
$x_{i+1} = P_X(x_i - \lambda G(2x_i - x_{i-1}))$
end for
return $\overline{x}_t = \frac{1}{t}\sum_{s=1}^{t}(2x_s - x_{s-1})$

А його алгоритм з регуляризацією можна записати, як:

Initialization: $x_0 \in X; \lambda \in (0, \frac{\sqrt{2}-1}{L})$
for $i \in \overline{0, t-1}$ **do**
$x_{i+1} = P_X(x_i - \lambda G(2x_i - x_{i-1})(1 - \alpha_n))$
end for
return $\overline{x}_t = \frac{1}{t}\sum_{s=1}^{t}(2x_s - x_{s-1})$

4.2.3 Екстраградієнтний метод Ценга

Для методу Ценга [7] нам необхідно стверждувати, що оператор буде монотонний і $L-$ліпщиців.

Основною перевагою методу, є необхідність рахувати проєкцію один раз, серед недоліків виділимо необхідність знаходити значення оператора у двох точках.

Initialization: $x_0 \in X; \lambda \in (0, \frac{1}{L})$
for $i \in \overline{0, t-1}$ **do**
$y_i = P_X(x_i - \lambda G(x_i))$
$x_{i+1} = y_i - \lambda(G(y_i) - G(x_i))$
end for
return $\overline{x}_t = \frac{1}{t}\sum_{s=1}^{t}(x_s)$

Задаємо, $a \in \mathcal{H}$, тоді метод Ценга з регуляризацією виглядатиме так:

Initialization: $x_0 \in X; \lambda \in (0, \frac{1}{L})$
for $i \in \overline{0, t-1}$ **do**
$y_i = P_X(x_i - \lambda G(x_i))$
$x_{i+1} = \alpha_n a + (1 - \alpha_n)(y_i - \lambda(G(y_i) - G(x_i)))$
end for
return $\overline{x}_t = \frac{1}{t}\sum_{s=1}^{t}(x_s)$

4.2.4 Операторна екстраполяція

Одним з нових методів для варіаційних нерівностей є метод операторної екстраполяції[4]:

Initialization: $x_0 \in X; \lambda \in (0, \frac{\sqrt{2}-1}{L})$
for $i \in \overline{0, t-1}$ **do**
$x_{i+1} = P_X(G(x_i) - \lambda G(x_i - G(x_{i-1})))$
end for
return $\overline{x}_t = \frac{1}{t}\sum_{s=1}^{t}(x_s)$

Серед зауважень щодо відмінностей між методом OE та кількома іншими існуючими методами для варіаційних нерівностей є наступне. Алгоритм методу вимагає лише однієї оцінки оператора та однієї проекції над множиною X.

По-друге, Нестеров та Скімалі показують, що метод операторної екстраполяції може бути використаний для оптимального вирішення сильно монотонних задач. Крім того, цей метод вимагає сильної монотонності.

Аналог цього методу, буде мати вигляд:

Initialization: $x_0 \in X; \lambda \in (0, \frac{\sqrt{2}-1}{L})$

for $i \in \overline{0, t-1}$ **do**

$x_{i+1} = P_X(G(x_i) - \lambda G(x_i - G(x_{i-1})))(1 - \alpha_n)$

end for

return $\overline{x}_t = \frac{1}{t}\sum_{s=1}^{t}(x_s)$

4.2.5 Екстраградієнтний метод Корпелевич

Метод Корпелевич [8] є одним з найвідоміших методів для знаходження розв'язку задачі варіаційних нерівностей. Алгоритм екстраградієнтного методу виглядає наступним чином:

Initialization: $x_0 \in X; \lambda \in (0, \frac{1}{L})$

for $i \in \overline{0, t-1}$ **do**

$x_i = P_X(x_i - \lambda G(x_i))$

$x_{i+1} = P_X(x_i - \lambda G(y_i))$

end for

return $\overline{x}_t = \frac{1}{t}\sum_{s=1}^{t}(x_s)$

Серед переваг методу можна відзначити, те, що його можна застосовувати для монотонного оператора G. Серед недоліків, можна виділити необхідність знаходити двічі проекцію на множину X, що знижує швидкість роботи алгоритму.

Його аналог має вигляд:

Initialization: $x_0 \in X; \lambda \in (0, \dfrac{1}{L})$

for $i \in \overline{0, t-1}$ **do**

$x_i = P_X(x_i - \lambda G(x_i))$

$x_{i+1} = P_X(x_i - \lambda G(y_i))(1 - \alpha_n)$

end for

return $\overline{x}_t = \dfrac{1}{t}\sum_{s=1}^{t}(x_s)$

4.2.6 Метод екстраполяції з минулого

Метод Попова [9], або Extrapolation from the past[6], є більш зручною варіацією метода Корпелевич. Головною ідеєю є використання обчисленого значення монотонного оператора G, та двічі знаходити проєкцію на множину:

Initialization: $x_0 \in X; \lambda \in (0, \dfrac{1}{2L})$

for $i \in \overline{0, t-1}$ **do**

$y_i = P_X(x_i - \lambda G(y_{i-1}))$

$x_{i+1} = P_X(x_i - \lambda G(y_i))$

end for

return $\overline{x}_t = \dfrac{1}{t}\sum_{s=1}^{t}(x_s)$

Можна записати, аналогічний йому метод, як:

Initialization: $x_0 \in X; \lambda \in (0, \dfrac{1}{2L})$

for $i \in \overline{0, t-1}$ **do**

$y_i = P_X(x_i - \lambda G(y_{i-1}))$

$x_{i+1} = P_X(x_i - \lambda G(y_i))(1 - \alpha_n)$

end for

return $\overline{x}_t = \dfrac{1}{t}\sum_{s=1}^{t}(x_s)$

4.2.7 Алгоритм дзеркального спуску

Цей алгоритм[2] можна інтерпретувати як модифікацію двоступеневого алгоритму Попова із застосуванням проекції на здійсненну множину у сенсі відстані Брегмана. Як і інші схеми дзеркального спуску, метод дозволяє ефективно враховувати структуру здійсненної сукупності проблеми. Основним теоретичним результатом є доведена теорема про збіжність методу. Очевидним недоліком алгоритму є припущення, що константа Ліпшица оператора відома або допускає просту оцінку. Більше того, в певних задачах оператори можуть не задовольняти глобальну умову Ліпшіца.

Initialization: $x_0 \in X; \lambda \in (0, \frac{1}{L})$

for $i \in \overline{0, t-1}$ **do**

$y_i = P_X(x_i - \lambda G(y_{i-1}))$

$x_{i+1} = P_X(x_i - \lambda G(y_i))$

end for

return $\overline{x}_t = \frac{1}{t} \sum_{s=1}^{t} (x_s)$

Алгоритм дзеркального спуску з регуляризацією, матиме вигляд:

Initialization: $x_0 \in X; \lambda \in (0, \frac{1}{L})$

for $i \in \overline{0, t-1}$ **do**

$y_i = P_X(x_i - \lambda G(y_{i-1}))$

$x_{i+1} = P_X(x_i - \lambda G(y_i))(1 - \alpha_n)$

end for

return $\overline{x}_t = \frac{1}{t} \sum_{s=1}^{t} (x_s)$

5 ПРАКТИЧНА ЧАСТИНА

Теорія нерухомої точки пропонує важливі інструменти для аналізу при вивченні існування та наближення рішень різних задач (задачі оптимізації, варіаційні проблеми нерівності, проблеми включення, задачі рівноваги тощо). Шуканий розв'язок такої нелінійної задачі виражається як нерухома точка відповідного оператора, тобто як розв'язок еквівалентної задачі з нерухомою точкою:

$$x = Tx, \tag{5.1}$$

Де T визначено на просторі X. Тепер дана задача вирішується застосуванням теореми про нерухому точку, і таким чином ми можемо відразу побудувати ітераційну послідовність, визначену як:

$$x_{n+1} = Tx_n, \;\; n \geq 0, \text{ де } x_0 \text{ - початкове наближення.} \tag{5.2}$$

Нагадаємо, що задача знаходження розв'язку варіаційної нерівності для відображення $G : H \to H$ визначається наступним чином:

Знайти $x^* \in X$ таким, що

$$\langle G(x^*), x - x^* \rangle \geq 0, \; \forall x \in X, \tag{5.3}$$

де X - непорожня, опукла та замкнута підмножина дійсного гільбертового простору H з нормою $|| \cdot ||$. Відомо, що задача варіаційної нерівності, еквівалентна задачі нерухомої точки:

$$x = P_X(x - \lambda G(x)), x \in X, \tag{5.4}$$

де P_X - найближча точкова проекція на X.

5.1 Тестування алгоритмів без регуляризації

Для початку, можемо ознайомитися з приблизною швидкодією алгоритмів для невеликих початкових розмірностей.

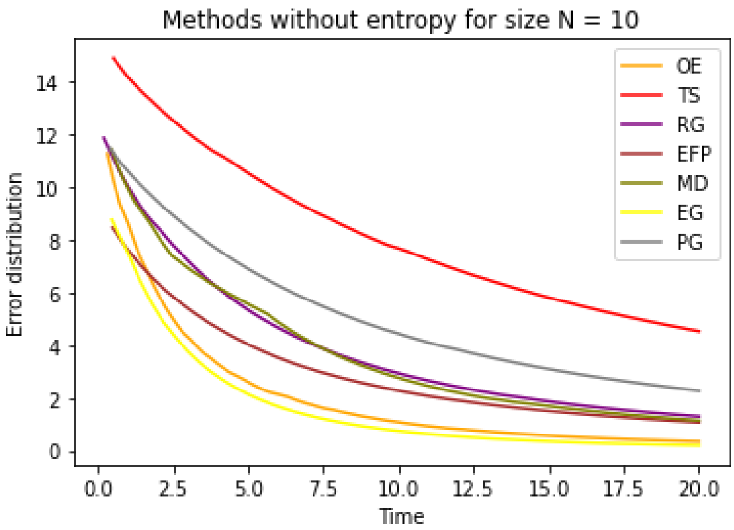

Рис. 1: Алгоритми без регуляризації для множини з 10 блоків

	Розмірність = 25		Розмірність = 49	
Назва методу	Кількість ітерацій	Значення	Кількість ітерацій	Значення
Проєкційний	908	0.41	972	0.234
Відбиваючий проекц.	834	0.52	956	0.247
Операторної екстр.	774	0.65	895	0.267
Екстраполяції з мин	1790	0.1	1356	0.116
Ценга	697	0.49	719	0.249
Корпелевич	925	0.26	1088	0.177
Дзеркальний спуск	825	0.44	959	0.210

Табл. 1: Алгоритми без регуляризації для множин з 25, 49 блоків

Можна побачити, що алгоритм Ценга не є достатньо точним для цієї задачі. Алгоритми операторної екстраполяції та метод Корпелевич переважають за швидкістю методи екстраполяції з минулого та метод дзеркального спуску, проте потребують більшої кількості ітерацій для обчислення наступного кроку. Також маємо непрямий доказ подібності алгоритмів методів дзеркального спуску та методу відбиваючої проєкції, що можна побачити, з близького значення кількості ітерації та похибки обчислення. Найменші похибки було знайдено для методів екстраполяції з минулого, відбиваючого проекційного та методу дзеркального спуску. Проте найменше число ітерацій було у методу Ценга, операторної екстраполяції та дзеркального спуску.

5.2 Тестування алгоритмів з регуляризацією

Аналогічно до алгоритмів без регуляризації знайдемо відстань Вассерштейна для визначнених методів.

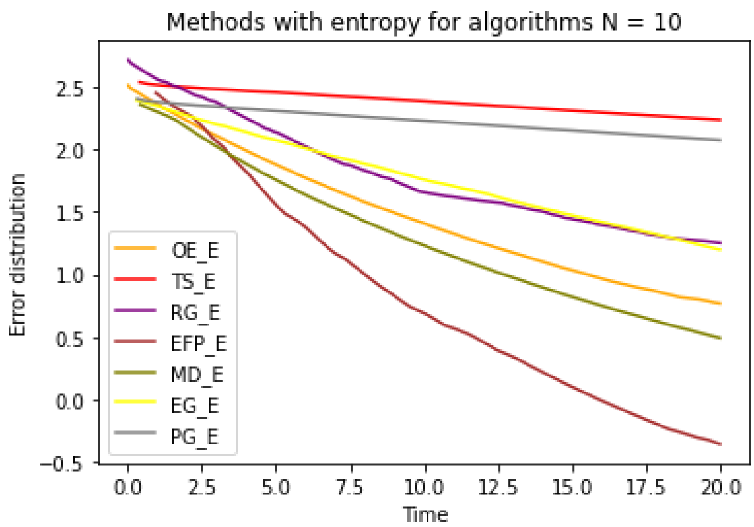

Рис. 2: Алгоритми з регуляризацією для множини з 10 блоків

	Розмірність = 25		Розмірність = 49	
Назва методу	Кількість ітерацій	Значення	Кількість ітерацій	Значення
Проєкційний	1338	0.65	1332	0.276
Відбиваючий проекц.	1935	0.37	934	0.368
Операторної екстр.	1245	0.65	3425	0.234
Ценга	1295	0.64	1302	0.278
Екстраполяції з мин.	1541	0.30	1356	0.198
Корпелевич	1205	0.77	1027	0.291
Дзеркальний спуск	1909	0.35	1026	0.350

Табл. 2: Алгоритми з регуляризацією для множин з 25, 49 блоків

Можна побачити, що алгоритм Ценга для цієї задачі наближений до простого проєкційного методу для цієї задачі. Алгоритми операторної екстраполяції та метод дзеркального спуску переважають за швидкістю методи екстраполяції з минулого та метод дзеркального спуску, проте потребують більшої кількості ітерацій для обчислення наступного кроку. Також маємо непрямий доказ подібності алгоритмів методів для великих розмірностей дзеркального спуску та методу відбиваючої проєкції, що можна побачити, з близького значення кількості ітерації та похибки обчислення. Найменші похибки було знайдено для методів екстраполяції з минулого та операторної екстраполяції. Проте найменше число ітерацій було у методу Ценга, операторної екстраполяції та алгоритму Корпелевич.

5.3 Тестування алгоритму для знаходження барицентру Вассерштейна

Знаходження та аналіз отриманих даних з досліджень відстані Вассерштейна для задачі варіаційної нерівності дало можливість визначити, принаймні з розглянутих алгоритмів, кращий для знаходження барицентра Вассерштейна. Таким оптимальним алгоритмом визначено метод Корпелевич.

Застосувавши алгоритм відповідно до методу Корпелевич. Отримали такі результати при кількості градієнтих кроків 20 та 50.

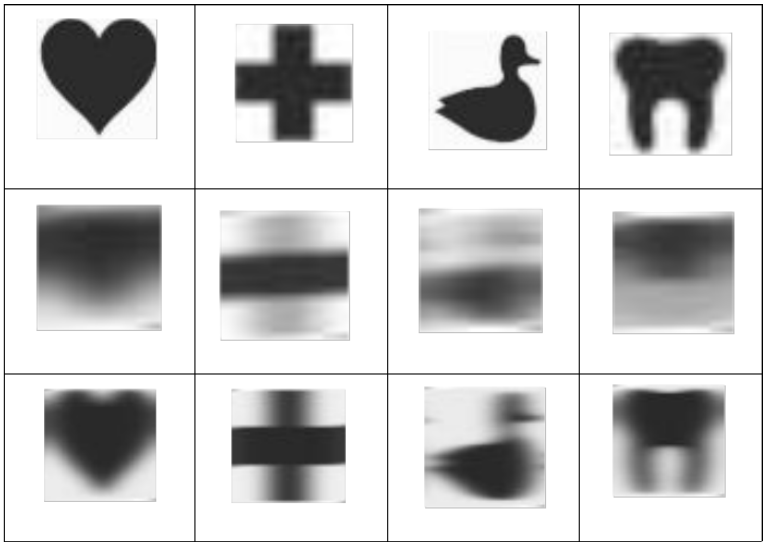

Рис. 3: Приклад знаходження барицентрів Вассерштейна для 20 та 50 кроків градієнта

6 ВИСНОВКИ

Розв'язання поставленої задачі уможливило сформулювати наступні висновки.

1. Хоча вирішенням проблеми знаходження барицентра Вассерштейна почали займатися ще у XVIII-столітті, сьогодні ці методи продовжують активно розвиватися. Вони знаходять серйозне практичне застосування в різних галузях, зокрема, медицині, обробці зображень, логістичних задачах та інших мінімаксних задачах. Якщо раніше універсальним методом розв'язання такого роду задач були методи лінійного програмування, то сьогодні задача ускладнюється тим, що через велику кількість елементів, велику кількість даних стандартні алгоритми працюють повільно. Отже є необхідність в отриманні максимально точного результату за найменшу кількість ітерацій, з найменшими часовими витратами.

2. Розглянуті алгоритми розв'язання транспортних задач показали, що не існує універсального підходу до для пошуку барицентра Вассерштейна монотонним оператором за допомогою алгоритмів Попова, Корпелевич, Ценга, операторної екстраполяції та алгоритмів Маліцького (відбиваючого проєкційного та методу дзеркального спуску).

3. За допомогою розробленого алгоритму та бібліотеки проєкційних методів показано, що для малої розмірності вхідних даних класичні алгоритми операторної екстраполяції та метод Корпелевич переважають за швидкістю методи екстраполяції з минулого та метод дзеркального спуску. Проте зазначені алгоритми потребують більшої кількості ітерацій для обчислення наступного кроку. Водночас серед алгоритмів з регуляризацією методи екстраполяції з минулого та операторної екстраполяції мають

найменшу похибку.

4. Розроблена програмна бібліотека для розв'язування варіаційної нерівності із використанням алгоритми Попова, Корпелевич, Ценга, операторної екстраполяції та алгоритми Маліцького, дозволила запропонувати аналоги вирішення задачі з регуляризацією для класичних алгоритмів. Можемо стверджувати, щоб побачити істотну перевагу оптимальності алгоритмів необхідно задати велику вхідну розмірність для початкових даних.

ДЖЕРЕЛА

[1] Jambulapati, A., Sidford, A., Tian, K. (2019). A Direct $\tilde{O}(1/\varepsilon)$ Iteration Parallel Algorithm for Optimal Transport, arXiv https://arxiv.org/abs/1906.00618

[2] Semenov, V. V.(2017). A version of the mirror descent method to solve variational inequalities. Cybernetics and Systems Analysis, Vol. 53, No. 2.

[3] Agueh, M., Carlier, G. (2011). Barycenters in the Wasserstein space. SIAM J. Math. Anal. Vol. 43, No. 2, pp.904–924.

[4] Kotalis, G., Lan, G., Li, T. (2020). Simple optimal optimal methods for stochastic variational inequalities, I: operator extrapolation, arXiv https://arxiv.org/abs/2011.02987

[5] Malitsky, Yu. (2015). Projected Reflected Gradient Methods for Monotone Variational Inequalities. SIAM Journal on Optimization. Vol. 25. pp. 502–520.

[6] Gidel, G., Berard, H., Vincent, P., Lacoste–Julien S. (2018). A Variational Inequality Perspective on Generative Adversarial Networks, arXiv https://arxiv.org/abs/1802.10551.

[7] Tseng, P. (2000). A modified forward-backward splitting method for maximal monotone mappings. SIAM Journal on Optimization. Vol. 38. pp. 431–446.

[8] Korpelevich, G. (1976). Extragradient method for finding saddle points and other problems. Economics and Mathematical Methods. No 4, pp. 747–756.

[9] Popov, L. (1980). Modification of the Arrow-Hurwitz method for finding saddle points. Mathematical notes. Vol. 28, No 5, pp. 777–784.

[10] Bakushinsky, A., Goncharsky, A. (1989). Ill-posed problems. Numerical methods and applications. Moscow: MSU Publishing House

[11] Sherman, J. (2017 June). Area-convexity, l_∞ regularization, and undirected multicommodity flow. In Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, pp. 452–460.

[12] Cuturi, M. (2013). Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In: Proc. NIPS, pp. 2292–2300