УДК 519.1
DOI: 10.32626/2308-5878.2021-22.75-80

О. В. Зеленський*, канд. фіз.-мат. наук,
В. М. Дармосюк,** канд. фіз.-мат. наук,
Р. В. Лобач***

* Кам'янець-Подільський національний університет
імені Івана Огієнка, м. Кам'янець-Подільський,
** Миколаївський національний університет
імені В. О. Сухомлинського, м. Миколаїв,
*** Кам'янець-Подільський НВК № 14, м. Кам'янець-Подільський

ВІДНОВЛЕННЯ МАТРИЦЬ ВІДСТАНЕЙ ТА ЇХ ЗАСТОСУВАННЯ

Матриці відстаней застосовуються в геометричному моделюванні та в задачах відновлення геометричних об’єктів, економіці, біоінформатиці, програмуванні. Матриці відстаней застосовуються у машинному навчанні, наприклад, створюються програми, пов'язані з дорожнім трафіком, автобусними маршрутами, геолокацією зокрема компанія Yandex створила сервіс, у якому, за допомогою матриць відстаней прогнозується завантаженість доріг, на потрібний час у майбутньому. Таким чином, автомобілісти можуть запобігти потрапляння у затори. Distance Matrix API — це сервіс, який повідомляє відстань і час в дорозі між початком руху та точкою призначення. Сервіс повертає інформацію на основі запропонованого маршруту між початковою та кінцевою точками, обчисленого API картами Google, і складається із значеннь тривалості подорожі та відстані для кожної пари пунктів.

Також матриці відстаней можуть бути застосовані при створенні будь-якої статистики. У біоінформатиці матриці відстаней використовуються для представлення структур білків незалежним від координат чином, або для відновлення відстаней у ланцюгу ДНК.

У [4] автори розглядають фундаментальні властивості EDM, такі як ранг та не визначеність. У статті досліджують, як різні властивості EDM можуть бути використані для розробки алгоритмів для заповнення та зменшення шумів даних про відстані. Попутно автори демонструють застосування матриць відстаней для калібрування положення мікрофона та ультразвукової томографії.

В роботі знайдено критерій можливості відновлення матриці Евклідових відстаней на прямій, та між вершинами опуклого n-кутника на площині. Розроблено алгоритм передачі ключа к шифру з використанням матриць Евклідових відста-

© О. В. Зеленський, В. М. Дармосюк, Р. В. Лобач, 2021

ней на площині. Розроблено швидкий алгоритм відновлення матриці відстаней між об'єктами на прямій.

Ключові слова: *матриці відстаней, матриця Евклідових відстаней, граф, зв'язний граф, граф матриці відстаней, відновлення матриці відстаней.*

Вступ.

Означення 1. Матрицею відстаней називають квадратну матрицю типу «об'єкт-об'єкт» (порядку n), що містить відстані між об'єктами в метричному просторі в якості елементів [1].

В загальному ж вигляді матриця виглядає наступним чином:

$$\begin{bmatrix} 0 & \cdots & d_{1j} & \cdots & d_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ d_{i1} & \cdots & d_{ij} & \cdots & d_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ d_{n1} & \cdots & d_{nj} & \cdots & 0 \end{bmatrix}.$$

Властивості матриці відстаней:

1. Симетричність відносно діагоналі, тобто $d_{ij} = d_{ji}$.
2. Відображення властивості тотожності відстані $d_{ij} = 0 \leftrightarrow i = j$ в матриці відстаней всі елементи головної діагоналі рівні нулю.
3. Значення відстаней в матриці завжди невід'ємні.
4. Нерівність трикутника приймає форму $d_{ij} + d_{jk} \geq d_{ik}$ для всіх i, j та k.

1. Відновлення матриць Евклідових відстаней.

Означення 2. Евклідова метрика (евклідова відстань) — метрика в евклідовому просторі — відстань між двома точками евклідового простору, що обчислюється за теоремою Піфагора [2]. Для точок $p = (p_1, p_2, ..., p_k)$ і $q = (q_1, q_2, ..., q_k)$ Евклідова відстань визначається наступним чином:

$$d(p,q) = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2}.$$

Розглянемо задачу **відновлення матриці евклідових відстаней.**

Нехай задано n точок площині $A_k(x_k, y_k)$ між деякими з яких відстані відомі, а між деякими ні. Нехай $[D] = (d_{ij})$ — матриця квадратів евклідових відстаней, $d_{ij} = \|A_i, A_j\|^2$, або $d_{ij} = -1$, якщо відстань між об'єктами невідома. Матрицю $[D]$ в якій всі відстані відомі називають матрицю квадратів евклідових відстаней (*EDM*). Задача полягає в тому, що потрібно знайти координати всіх об'єктів, та всі невідомі відстані між об'єктами. Виникає питання, коли по деяким елементам матриці відстаней, можна відновити її повністю. Побудуємо по матриці $[D]$ неорієнтований граф $G = (V, E) = G(D)$ наступним чином $(i, j) \in E(G)$, якщо відстань $\rho(A_i, A_j)$ відома та $(i, j) \notin E(G)$ якщо відстань $\rho(A_i, A_j)$ невідома.

Будемо вважати, що точки між якими ми шукаємо відстані утворюють опуклий n-кутник.

З опуклості n-кутника випливає, що жодні три точки не лежать на одній прямій. Тобто виконується строга нерівність трикутника $d_{ij} + d_{jk} > d_{ik}$.

Твердження. Якщо відомо менше ніж $2n - 3$ відстані, то задача відновлення відстаней має безліч розв'язків.

Доведення. Нехай для $i \in [3 \ldots n]$ відомі відстані d_{1i}, та d_{2i} тобто відомі $2n - 4$ відстані серед яких немає зайвих. Позначимо

$$M = \min_{i \in [3 \ldots n]} \{d_{1i} + d_{2i+}\},\ m = \min_{i \in [3 \ldots n]} \{|d_{1i} - d_{2i}|\}.$$

Для довільного $d \in (m, M)$ можна побудувати матрицю відстаней вважаючи $d_{12} = d$.

Оскільки дійсних чисел з інтервалу (m, M) є нескінченно багато, то їй задача відновлення матриці відстаней має безліч розв'язків.

Теорема 1. Якщо в графі $G(D)$ з n вершинами, $n \geq 3$ можна вибрати $2n - 3$ ребра, так, що будь-який підграф з $k \geq 3$ вершинами містить не більше $2k - 3$ вибраних ребер, то матрицю Евклідових відстаней на площині D можна відновити повністю.

Доведення. Застосуємо метод математичної індукції.

1. При $n = 3$, всі три відстані між трьома точками відомі, теорема виконується.
2. Припустимо, що твердження виконується при $n = k$.
3. Доведемо, що теорема виконується при $n = k + 1$.

Спочатку доведемо, що знайдеться точка для якої ми знаємо не більше трьох відстаней. Припустимо, що такої точки не існує, тобто для довільної точки ми знаємо не менше 4 відстаней. Тоді кількість всіх відомих відстаней не менше за $\frac{4n}{2} = 2n$, але $2n \geq 2n - 3$, ми отримали протиріччя, яке доводить, що знайдеться точка, для якої ми знаємо не більше трьох відстаней.

Розглянемо випадок, що існує точка для якої ми знаємо тільки дві відстані. (Очевидно, що якщо для точки ми знаємо тільки одну відстань, то матрицю відстаней відновити неможливо, оскільки точка може рухатися по колу).

Видалимо цю точку разом з двома відстанями. В нас залишається k точок, та $2k - 3$ відстані, тому за припущенням індукції матрицю відстаней з k точками можна відновити. (Тобто знайти координати вершин та всі відстані). Після чого знаючи дві відстані знаходимо координати видаленої точки та всі відстані.

Розглянемо випадок, що існує точка для якої ми знаємо рівно три відстані. Нехай це буде точка A_1 та відстані d_{12}, d_{13}, d_{14}. Відомо, що для 4 точок опуклого чотирикутника знаючи 5 відстаней, ми однозначно можемо знайти 6, тому серед відстаней d_{23}, d_{24}, d_{34} принаймні одну відстань ми точно не знаємо, оскільки серед відстаней немає зайвих. Тобто, наприклад знаючи d_{23}, d_{24}, ми однозначно знаходимо d_{34}. Тому, використовуючи точку A_1, ми можемо знайти одну відстань між точками A_2, A_3, A_4. Видалимо точку A_1 разом з відстанями d_{12}, d_{13}, d_{14}.

Залишається k точок, $2k - 4$ відстані та одну відстань ми можемо знайти між точками. A_2, A_3, A_4. Тобто всього ми знаємо $2k - 3$ відстаней, тому за припущенням індукції матрицю відстаней можна відновити, а потім додати до неї точку A_1. Теорема доведена.

2. Відновлення матриці відстаней між об'єктами на прямій. Розглянемо одновимірний випадок: об'єкти знаходяться на прямій тобто визначаються однією координатою. Будемо вважати, що $A_i(x_i)$, $x_1 = 0$, $x_j > x_i$ для $j >$ i. Тобто об'єкти на прямій знаходяться в порядку зростання їх номерів та початок відліку координатної прямої співпадає з точкою A_1.

Зрозуміло, що для того щоб відновити матрицю відстаней потрібно знайти $n - 1$ невідоме: $x_2, x_3, \ldots, x_n$, а для цього потрібно $n - 1$ лінійно незалежних рівнянь.

Зауважимо, що розв'язувати задачу про відновлення матриці відстаней порядку $n \geq 10000$ методами лінійної алгебри є досить обчислювально складною задачею. Тому в статті створено алгоритм відновлення таких матриць відстаней.

Розв'яжемо задачу про відновлення відстаней D, та знайдемо критерій коли це можливо зробити.

Для цього ми за відомими відстанями побудуємо неорієнтований граф $G(D) = (V(G), E(G))$, де $V(G) = \{1, 2, \ldots, n\}$, $\sigma_{ij} \in E(G) \Leftrightarrow d_{ij} \in D$

Тобто граф побудований за матрицею відстаней складається з n вершин, дві вершини «i» та «j» з'єднані ребром, якщо відома відстань d_{ij}, та дві вершини «i» та «j» не з'єднані ребром, якщо не відома відстань d_{ij}.

Теорема 2. Матрицю відстаней D можна відновити тоді і тільки тоді, коли граф $G(D)$ є зв'язним.

Доведення. Нехай граф $G(D)$ є зв'язним. Доведемо, що можливо відновити всі відстані матриці D. Для цього достатньо знайти координати x_i. Поділимо вершини графа G на декілька рівнів. Вершину «1» будемо вважати вершиною нульового рівня. Вершини суміжні з вершиною «1» — це вершини першого рівня. Якщо $A_i(x_i)$ — це точка, яка відповідає вершині першого рівня, то $x_i = d_{1i}$. До вершин другого рівня віднесемо вершини відстань від яких до вершини «1» в графі G дорівнює два. Тобто вершини другого рівня не є суміжними з вершиною «1», та є суміжними хоча б з однією вершиною першого рівня. Нехай вершині «i» першого рівня відповідає точка $A_i(x_i)$, вершині «j» другого відповідає точка $A_j(x_j)$, та вершина «i» суміжна з вершиною «j». Тоді координати x_j можна знайти за формулою:

$$x_j = \begin{cases} x_j + d_{ij}, \text{ якщо } j > i, \\ x_j - d_{ij}, \text{ якщо } j < i. \end{cases}$$

Продовжуючи алгоритм, знаючи координати точок, що відповідають вершинам другого рівня, однозначно визначаємо координати для третього рівня і так далі. Оскільки граф G зв'язний, то рівень кожної вершини є скінченим. Отже, для зв'язного графа G, можна знайти всі координати точок, тобто повністю відновити матрицю відстаней D.

Нехай граф $G(D)$ є не зв'язним. Доведемо, що неможливо відновити всі відстані матриці D. Розіб'ємо не зв'язний граф на компоненти зв'язності. Серед відстаней, які відомі немає жодної відстані між об'єктами з різних компонент. Нехай вершина «1» належить першій компоненті зв'язності. Збільшимо всі координати об'єктів з другої компоненти зв'язності на число t. При цьому відстані, які вже відомі не зміняться. Отже, число t може бути довільним, тому координати об'єктів з другої компоненти зв'язності та відстані між об'єктами з першої та другої компоненти зв'язності визначити неможливо. Отже, в цьому випадку матрицю D відновити неможливо.

Висновки. Матриці відстаней застосовуються в геометричному моделюванні. Важливою є задача відновлення матриць відстаней. В роботі доведено, що для відновлення матриці відстаней D між об'єктами на прямій необхідно та досить щоб граф $G(D)$ був зв'язним. Для відновлення матриці евклідових відстаней між об'єктами на площині, необхідно та досить знати $2n - 3$ незалежних відстаней. Знайдено застосування матриць відстаней в криптографії для передачі ключа шифру по відкритому каналу зв'язку, та розглянуто застосування в математиці при розв'язуванні задач, пов'язаних з цикломатичним числом площини.

Список використаних джерел:

1. Patwari N., Ash J. N., Kyperountas S., Hero A. O., Moses R. L., Correal N. S. Locating the Nodes: Cooperative Localization in Wireless Sensor Networks. *N IEEE Signal Process. Mag.* 2005. Vol. 22. № 4. P. 54-69.
2. Alfakih A., Khandani A., Wolkowicz H. Solving Euclidean Distance Matrix Completion Problems via Semidefinite Programming. *Comput. Optim. Appl.* 1999. Vol. 12. № 1-3. P. 13-30.
3. Doherty L., Pister K., El Ghaoui L. Convex Position Estimation in Wireless Sensor Networks. *Proc. IEEE INFOCOM.* 2001. Vol. 3. P. 1655-1663.
4. Dokmanic I., Parhizkar R., Ranieri J., Vetterli M. Euclidean Distance Matrices: Essential Theory, Algorithms and Application. *IEEE Signal Process. Mag.* 2015. Vol. 32. № 6. P. 12-30.

RECONSTRUCTION OF DISTANCE MATRIXES AND THEIR APLICATION

Distance matrix are used in geometric modeling and in the restoration of geometric objects, economics, bioinformatics, and programming. Distance matrices are used in machine learning, for example, programs related to road traffic, bus routes, and geolocation. In particular, Yandex has created a service in which distance matrices predict road congestion at the right time in the future. In this way, motorists can prevent traffic jams. Distance matrices can also be used to create any statistics. In bioinformatics, distance matrices are used to represent protein structures in a coordinate-independent manner, or to reconstruct distances in the DNA chain. Distance Matrix API is a service that provides distance and time for the exit and destination matrix. The API returns information based on the suggested route between the start and end points calculated by the Google Maps API and consists of paths that include duration and distance values for each pair.

In [4] authors reviews the fundamental properties of EDMs, such as rank or (non)definiteness. Authors shows how various EDM properties can be used to design algorithms for completing and denoising distance data. Along the way, authors demonstrates applications to microphone position calibration, ultrasound tomography.

The paper finds a criterion for the possibility of restoring the matrix of Euclidean distances on a line and between the vertices of a convex n-gon on a plane. An algorithm for transferring a key to a cipher using Euclidean distance matrices on a plane has been developed. A fast algorithm for reconstructing the distance matrixes between objects on a straight line has been developed.

Key words: *distance matrix, Euclidean distance matrix, graph, connected graph, graph of distance matrix, reconstruction of distance matrix.*

Отримано: 19.10.2021