

УДК 514.17, 515.1

Рубльов Б.В.¹, д.ф.-м.н., проф.
Танцюра Б.В., студент
Молодцов О.І., к.ф.-м.н., доц.

Визначення метрики на деревах спеціального вигляду

У цій статті розглядаються деякі попередні підходи до розв'язання задачі розрізнення топологічних форм функцій, їх переваги та недоліки, а також ідея нового підходу

Ключові слова: функція, топологічна форма, метрика

¹ Київський національний університет імені Тараса Шевченка, 03127, Київ-127, просп. акад. Глушкова, 6, факультет кібернетики, e-mail: rublyovbv@gmail.com

Статтю представив д.т.н., с.н.с. Кудін В.І.

Вступ

Порівняння функцій є важливим елементом у сучасній прикладній математиці. Якщо ці функції задані своїми формулами чи графіками, то методів їх порівняння багато, майже будь-яка відома метрика дає змогу тим чи іншим чином оцінити близькість функцій. На практиці функції можуть визначатися самими різноманітними засобами, тому і порівнювати доводиться об'єкти досить складної природи. В цій статті ми плануємо досліджувати топологічну близькість функцій.

Позначимо через $S^{(0)}$ сукупність усіх неперервних функцій, що визначені на $[a; b]$, та мають скінченну кількість екстремумів. Крім того ми будемо розглядати кусково-лінійні функції, оскільки для вивчення топологічних властивостей, які нас цікавлять, цього цілком достатньо. Розглянемо довільну функцію $f \in S^{(0)}$, і зробимо з цією функцією такі перетворення.

Якщо $f(a) \neq f(b)$, то продовжимо функцію f до функції f_1 , що визначена на проміжку $[a; b] \cup [b; 2b - a]$ „парним чином”, тобто $\forall t \in [0; b - a]$ виконується рівність: $f_1(b - t) = f_1(b + t)$ (рис. 1). В подальшому будемо вважати, що $f \in S^{(0)}$ та $f(a) = f(b)$.

B. V. Rublov¹, PhD
B. V. Tantsiura, student
O. I. Molodcov, PhD

Definition of metrics on trees of special shape

In this article some previous approaches for solving the problem of distinguishing topological shapes of functions, their advantages and disadvantages, and the idea of a new approach are considered

Key words: function, topological shape, metrics

¹ Taras Shevchenko National University of Kyiv, 03127, Kyiv, Glushkova st., 6, faculty of cybernetics
e-mail: rublyovbv@gmail.com

Якщо $f(a) = f(b)$ не є глобальним мінімумом функції, то нехай f має глобальний мінімум у точці $c \in (a; b)$. Тоді розглянемо замість функції f таку функцію f_2 , яка

визначена на проміжку $[c; c + b - a]$ шляхом перенесення відповідного шматка функції з проміжку $[a; c]$ на

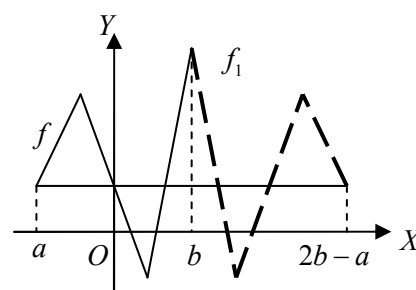


Рис. 1

проміжок $[b; c + b - a]$ (рис. 2).

Паралельним переносом зсунемо графік функції

таким чином, щоб вона стала визначеною на проміжку $[0; b - a]$, а множина значень дорівнювала

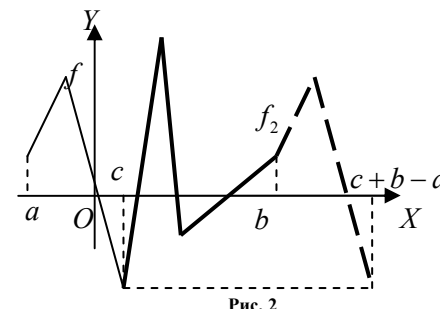


Рис. 2

$[0; d]$. Тепер стисканням (чи розтягуванням) вздовж координатних осей досягнемо того, щоб

$D(f) = E(f) = [0; 1]$. Множину усіх таких функцій назовемо класом S .

Опишемо більш точно представника класу функцій S : тобто

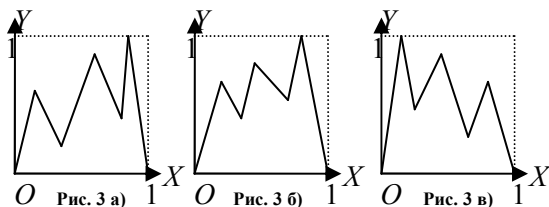
$$f \in S \Leftrightarrow f \in C[0; 1], \quad f(0) = f(1) = 0, \\ E(f) = [0; 1],$$

f має скінченну кількість екстремумів на $[0; 1]$.

Легко зрозуміти, що ми тепер можемо порівнювати неперервні функції, що визначені на сегменті, а також періодичні неперервні функції.

Через S_1 позначимо такий підклас функцій класу S , у яких кожен два екстремуми мають різну амплітуду, окрім двох глобальних мінімумів. Ці мінімуми однакові за побудовою функцій класу S , для періодичної функції це фактично одне й те ж значення.

Розрізнення топологічної форми функцій є важливим не тільки з теоретичної точки зору, а й з прикладної і має важливе застосування в таких науках як медицина та біологія. Це питання вивчали багато математиків (див. [1]). Метою даної роботи є розробка нового дескриптора топологічних форм одновимірних функцій. Дескриптором в даному випадку будемо вважати деяку метрику, яка має бути близькою до нуля для топологічно схожих функцій та неблизькою до нуля для несхожих. Топологічно схожими будемо вважати функції із схожими послідовностями локальних мінімумів та максимумів з певним урахуванням їх амплітуд. Наведемо приклади схожих та несхожих функцій. На рис. 3 а) та рис. 3 б) функції топологічно схожі, а на рис. 3 а) та рис. 3 в) – несхожі.



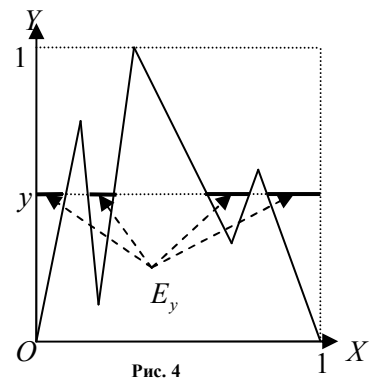
Два методи інтерпретації функцій

Найбільш поширеним підходом для цих цілей зараз використовується метод «діаграм стійкості» («Persistence diagrams») (див. [1]), опишемо його більш детально. По-перше, це буде корисно читачам, які мало обізнані поки що в цій тематиці, по-друге, ми будемо показувати недоліки запропонованого методу та порівнювати його з новим підходом. Для зручності, опишемо паралельно також і новий підхід.

Нехай є деяка функція $f \in S_1$. Для кожного y із області значень функції f розглянемо множину

$$E_y = \{x \mid f(x) \leq y\}$$

. На рис. 4 E_y позначена суцільною лінією. Зазвичай, E_y є

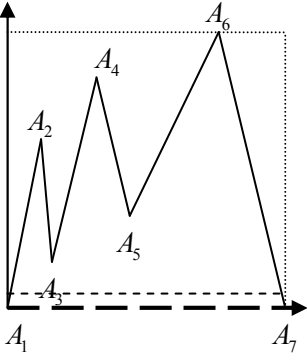
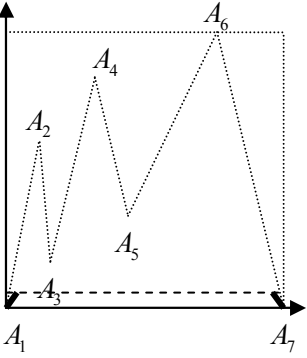
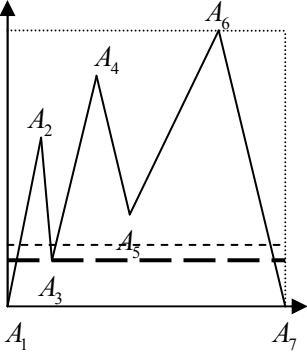
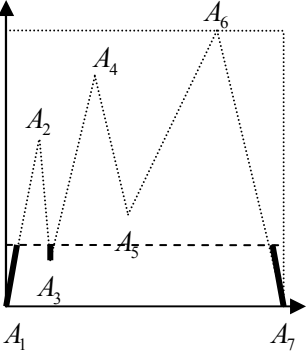
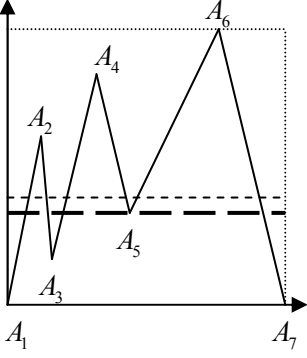
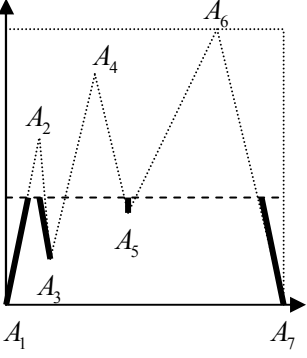
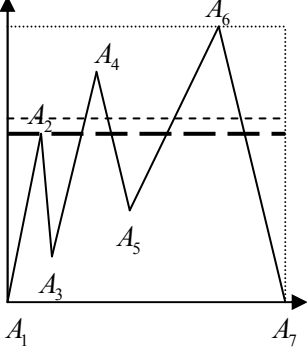
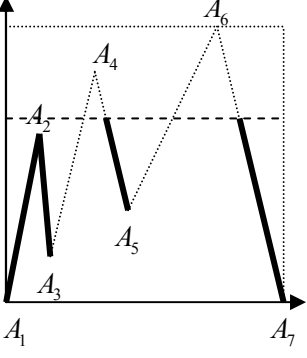


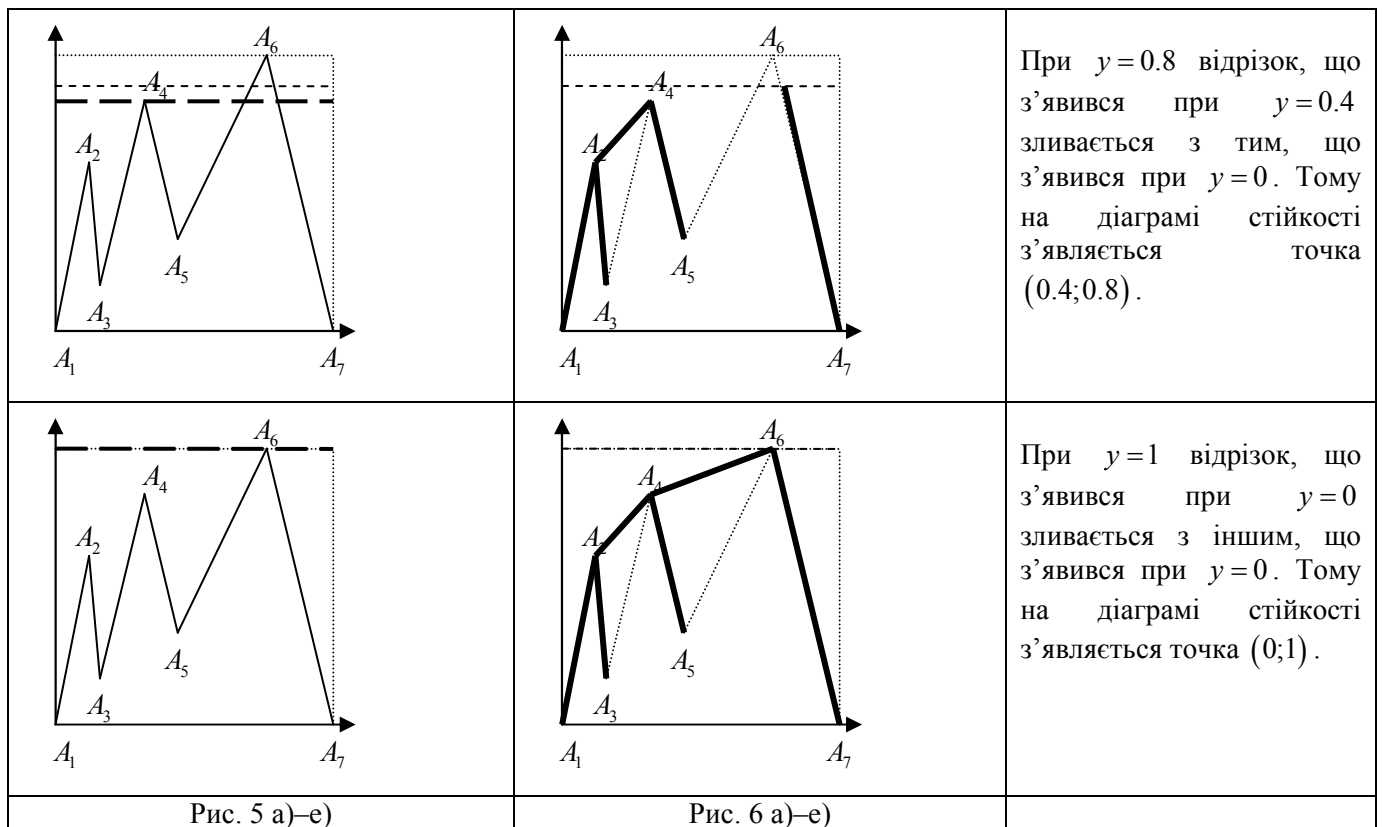
об'єднанням

деякої кількості відрізків. Розглянемо як буде змінюватись E_y при збільшенні y від найменшого до найбільшого. В кожній точці локального мінімуму буде з'являтися новий відрізок, а в кожній точці локального максимуму два відрізки будуть «зливатись». При «злитті» відрізків будемо вважати, що відрізок, який з'явився пізніше «помирає», а інший продовжує «жити».

Для нового алгоритму, який ми назовемо «двійкове дерево», у кожний момент нова компонента зв'язності множини E_y (тобто прохід через локальний мінімум функції) породжує нове ребро графу, а зникнення компоненти зв'язності (прохід через локальний максимум функції) – з'єднання двох ребер. При цьому будемо зберігати абсциси та ординати відповідних точок.

Розглянемо виконання алгоритмів, що реалізують обидва методи, на прикладі.

Графік функції	Двійкове дерево	Діаграма стійкості
		<p>При $y=0$ з'являються два відрізки.</p>
		<p>При $y=0.2$ з'являється ще один відрізок.</p>
		<p>При $y=0.4$ з'являється ще один відрізок.</p>
		<p>При $y=0.6$ відрізок, що з'явився при $y=0.2$ зливається з тим, що з'явився при $y=0$. Тому на діаграмі стійкості з'являється точка $(0.2; 0.6)$.</p>



Як легко бачити з реалізації двійкового алгоритму, отриманий граф, який ми будемо називати графом з класу T_1 , має вузли у точках локальних екстремумів функції f , при цьому ми в точності зберегли при побудові графа їх координати. Таким чином в подальшому ми можемо стверджувати, що один вузол графа лежить вище за інший та навпаки.

Розглянемо декілька прикладів побудови діаграм стійкості. Для функції з рис. 3а отримуємо діаграму з трьох точок: $(0.2; 0.6)$, $(0.4; 0.8)$ та $(0; 1)$. Абсолютно таку ж діаграму отримаємо для функції з рис. 3в. Для функції з рис. 3б отримаємо: $(0.4; 0.7)$, $(0.55; 0.85)$ та $(0; 1)$.

Порівняння функцій методом діаграми стійкості

У відповідність кожному відрізку, який з'явився в множині E_y , будемо ставити точку на площині координатами якої є час «народження» та час «смерті» відрізка, тобто точка буде в першій четверті вище бісектриси кута першої четверті, оскільки точка «помирає» пізніше, ніж «народжується». Тоді функції буде відповідати

діаграма з точок, яку і будемо називати «діаграмою стійкості» («Persistence diagram»). Для двох функцій, які потрібно порівняти, побудуємо діаграми стійкості. Тепер ці діаграми намалюємо на одному малюнку, але різними кольорами (наприклад, точки, що відповідають першій функції, намалюємо червоним, а ті, що другій – зеленим). Також намалюємо бісектрису кута першої четверті. Тепер кожен червоний відрізок з'єднаємо з зеленою, або з бісектрисою, кожен зелений відрізок з червоною або з бісектрисою. Тобто побудуємо бісекцію між червоними та зеленими точками, але дозволимо ставити у відповідність будь-яким точкам бісектрису (це необхідно, наприклад, коли кількість червоних точок не дорівнює кількості зелених). Відстанню між діаграмами стійкості при даному з'єднанні будемо вважати суму довжин всіх відрізків (відстань можна визначати також інакше, наприклад, максимум довжин відрізків). Нам потрібно знайти таке з'єднання, для якого відстань між діаграмами буде якнайменшою, і саме цю відстань і будемо вважати відстанню між діаграмами, а відповідно і між функціями.

Аналіз двійкового дерева типу T_1

Нагадаємо, що ми спочатку вивчаємо функцію з класу S_1 , які відрізняються від функцій класу S тим, що усі максимуми мають різну амплітуду. Тепер опишемо властивості графу класу

Властивість 1. Граф, отриманий за алгоритмом двійкового дерева, є деревом.

Властивість 2. Дерево типу T_1 розташоване на координатній площині у квадраті $[0; 1] \times [0; 1]$.

Властивість 3. Дерево типу T_1 має вузли з відмінними абсцисами та ординатами від інших вузлів.

Властивість 3. Дерево типу T_1 має один вузол степені 1–3:

- вузол степені 2, який ми назвемо **коренем**, має координати $(x_0; 1)$, де $0 < x_0 < 1$;
- з кожного вузла степені 1, які ми будемо називати **листочками**, виходить ребро, що направлене вгору, тобто до вузла з більшою ординатою; до листочків обов'язково належать точки $(0; 0)$ та $(1; 0)$;
- з кожного вузла степені 3, які ми будемо називати **гілками**, виходить одне ребро нагору та 2 донизу.

Назвемо **піддеревом** дерева типу T_1 , що відповідає вузлу A , усі вузли дерева, до яких можна дістатись ребрами дерева з вузла A , рухаючись донизу. Нехай з вузла A виходять два ребра у вузол A_1 , що має меншу від A абсцису, та A_2 , що має більшу від A абсцису. Тоді піддеревом, що відповідає вузлу A_1 (A_2) разом з ребром AA_1 (AA_2) називається **лівим (правим) піддеревом**, що відповідає вузлу A . При цьому вузол A_1 (A_2) називається **лівим (правим) сином** вузла A .

Так на рис. 7 для вузла A_4 правим піддеревом є лише ребро A_4A_5 , а лівим – ребра A_4A_2 , A_2A_1 та A_2A_3 . Разом усі ці 4 ребра

утворюють піддеревом, що відповідає вузлу A_4 . Лівим сином вузла A_4 є вузол A_2 , а правим – A_5 .

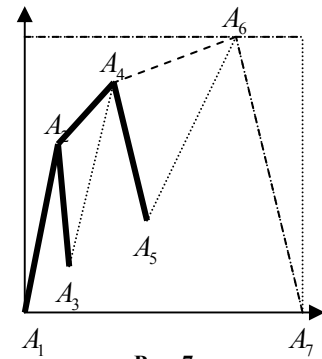


Рис. 7

Властивість 4. Кожне піддеревом будь-якого вузла

A має властивості аналогічні властивостям 1–3 за виключенням координат кореня та вузлів $(0; 0)$ та $(1; 0)$.

Доведення цих властивостей впливає безпосередньо з класу функцій S_1 та алгоритму його побудови. Дерево – це зв'язний граф, який не має циклів. З властивостей функцій класу S_1 , зрозуміло, що це так. З неперервності функції на сегменті випливає, що усі відрізки, які утворюють множину E_y в решті решт при $y \geq 1$ зіллються у один, тобто граф буде зв'язним. Якщо дивитись на утворення ребер графу, тобто розглядати множину E_y при зростанні ординати y , то жодного разу не буде ситуації, щоб з вузла вгору пішло 2 чи більше ребер. А при наявності циклу, повинний бути вузол з найменшою ординатою, тоді з нього вгору повинно йти щонайменше 2 ребра. Аналогічним чином доводяться й решта властивостей.

Теорема 1. Якщо розглянути дерево t , яке має властивості 1–4, то існує функція $f \in S_1$, для якої двійковий алгоритм дає дерево t .

Доведення.

Достатньо просто навести приклад такої функції. Для цього усі вузли дерева t відсортуємо по абсцисах зліва направо і позначимо їх B_1, B_2, \dots, B_n . Тепер з'єднаємо відрізками

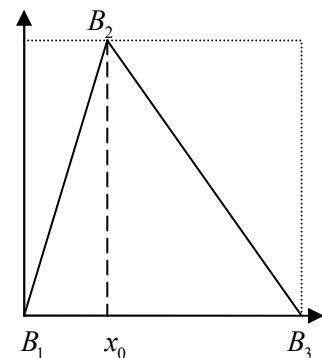


Рис. 8

послідовно ці вершини. Покажемо, що одержана функція f належить класу S_1 та задовольняє потрібні умови.

Будемо доводити твердження методом математичної індукції по кількості вершин. За визначенням класів T_0 та T_1 $n \geq 3$, оскільки точки $(0; 0)$, $(1; 0)$ є вузлами дерев, а крім того існує також вузол $(x_0; 1)$, де $0 < x_0 < 1$, який також повинен бути у деревах кожного типу.

Для $n = 3$ твердження очевидно виконується, бо тут і граф і функція співпадають (рис. 7).

Нехай $n > 3$. Позначимо корінь дерева B_s , а лівого сина B_l . Для піддерева B_l твердження теореми виконується за припущенням індукції, оскільки в ньому вершин менше ніж у всьому дереві. Зазначимо, що вершина B_l має найбільшу ординату серед всіх вершин розташованих лівіше B_s . Тому після проходу прямою (тією, що рухається знизу вгору) вершини B_l від неї буде продовжуватись ребро, яке з'єднається з деяким іншим вже в вершині B_s . Таким чином в дереві, що відповідає функції f , отримуємо ребро, яке було в дереві t . Отже, ліва частина дерева, що відповідає функції f , співпадає з лівою частиною дерева t . Абсолютно аналогічні міркування справедливі і для правої частини дерева.

Теорема доведена.

Перетворення дерева типу T_1 у дерево типу T_2

Тепер зробимо перетворення дерева типу T_1 у подібне дерево, але з певним зсувом по горизонталі його вузлів, яке будемо називати деревом типу T_2 . Корінь $A_0(x_0; 1)$ дерева типу T_1

переміщуємо в точку дерева типу T_2 з координатами $C_0(\frac{1}{2}; 1)$. Далі задамо процес побудови ітераційним чином. Нехай вузол A_k має координати $(x_k; y_k)$ у дереві типу T_1 , а координати відповідного йому вузлу

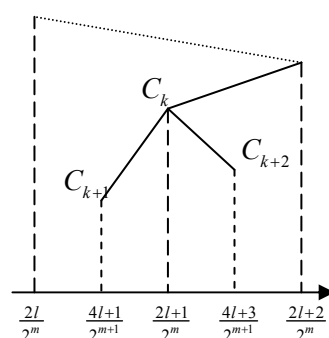


Рис. 9

C_k у дереві типу T_2 дорівнюють $(\frac{2l+1}{2^m}; y_k)$.

Розглянемо вузли $A_{k+1}(x_{k+1}; y_{k+1})$ та $A_{k+2}(x_{k+2}; y_{k+2})$ – лівого та сина цього вузла A_k , тоді абсциси x'_{k+1} та x'_{k+2} відповідних вузлів C_{k+1} та C_{k+2} знаходяться таким чином (рис. 9) $x'_{k+1} = \frac{4l+1}{2^{m+1}}$ та $x'_{k+2} = \frac{4l+3}{2^{m+1}}$. Якщо вузол A_k є листочком, то переходимо до аналогічного переміщення синів вже переміщених вузлів.

Так на рис. 10 показано перетворення дерева типу T_1 з рис. 7 до дерева типу T_2 . Абсциси вершин мають двійково-раціональні координати, а ординати – зберігаються.

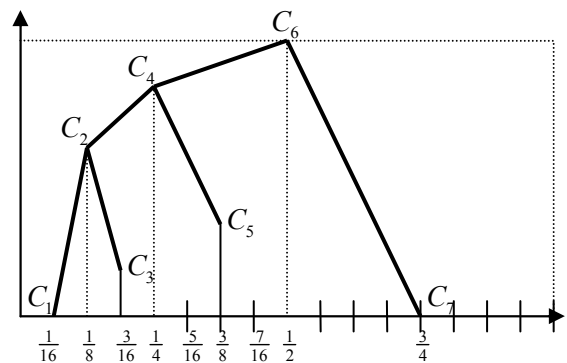


Рис. 10

Список використаних джерел

1. Herbert Edelsbrunner, John Harer, Computational topology. – Departments of Computer Science and Mathematics, Duke University.
2. H. Edelsbrunner, D. Letscher and A. Zomorodian. Topological persistence and simplification. Discrete Comput. Geom.
3. A. Zomorodian and G. Carlsson. Computing persistent homology. Discrete Comput. Geom.

Надійшла до редколегії 18.02.2013