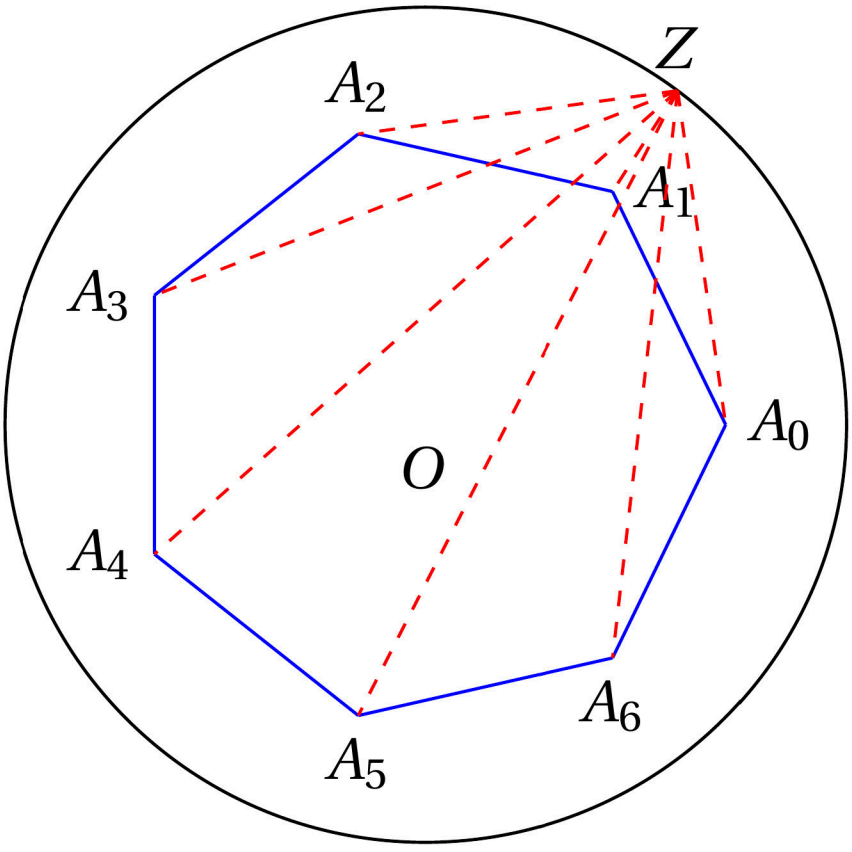
ISSN 2524-2407 (online), ISSN 1029-4171 (print)
Київський національний університет імені Тараса Шевченка

У світі математики

2021 р., 1(25)



Зміст

Горіхова задача. *О. К. Толпиго* 3

Обрії геометрії

Про деякі властивості чотирикутників. *В. М. Петечук, К. М. Петечук, Ю. В. Петечук* 9

Із яких точок простору Землю і Місяць видно під однаковими кутами. *О. В. Вишенська* 19

Студентська сторінка

Парадокси статистики. *Р. Є. Майборода* 25

Уявні числа та нове життя евклідової геометрії. *С. Є. Гладун* 36

Професія: математик

Робота актуарієм у Великій Британії. *Т. О. Андрощук* 48

Внутрішні хвилі — відсутній елемент глобального кліматичного пазла. *К. В. Терлецька* 57

В океані великих даних. *О. М. Корольов* 64

Видатні постаті

Леонардо да Вінчі. *Я. О. Жук* 78

Математика на шахівниці

Рекурсія в етюдах *Е. Г. Ейлазян* 86

Горіхова задача

Толпиго Олексій Кирилович,
кандидат фізико-математичних наук

У цій статті проаналізуємо «задачу про горіхи», яка декілька років тому пропонувалася на олімпіаді Турніру міст у такому формулюванні.

Задача 1. За столом сидять (за годинниковою стрілкою) Аня, Боря і Вітя. Вони їдять горіхи. Горіхів багато (більше трьох). Спочатку всі горіхи в Ані. Вона ділить їх порівну між Борею і Вітею, а той горіх, що залишився (якщо він є), з'їдає. Потім Боря робить так само: усі свої горіхи він ділить порівну між двома друзями та з'їдає той, що залишився (якщо він є). Потім горіхи ділить Вітя, знову Аня і т. д. Так триває досить довго.

а) Доведіть, що хоча б один горіх буде з'їдений.

б) Доведіть, що не всі горіхи будуть з'їдені.

Виявляється, що із цієї простенької задачі можна витягти досить багатий зміст і що вона пов'язана з найрізноманітнішими математичними темами. Ми розглянемо низку більш загальних питань: для довільної кількості горіхів N постараємося знайти кількість $E(N)$ горіхів, які будуть з'їдені в результаті, наскільки «чесно» розподіляться з'їдені горіхи, наскільки великою може бути частка з'їдених горіхів за великих значень N тощо.

Спочатку розберемо конкретний приклад. Нехай $N = 19$, тоді крок за кроком одержимо такі розподіли: $(19, 0, 0) \to (0, 9, 9)* \to (4, 0, 13)* \to (10, 6, 0)* \to (0, 11, 5) \to (5, 0, 10)* \to (10, 5, 0) \to (0, 10, 5) \to \ldots$ (зірочкою тут позначено з'їдені горіхи).

Очевидно, далі продовжувати немає сенсу. Горіхи будуть циклічно переходити від Ані до Борі й Віті, але завжди будуть розподілені за законом $(10, 5, 0)$, і жоден горіх більше не буде з'їдений. Про таку ситуацію ми будемо говорити, що «процес вийшов на цикл». Бачимо, що $E(19) = 4$, причому Борі дісталося два горіхи (на другому і п'ятому кроках), а Ані й Віті — по одному.

Проаналізуємо те, що ми дізналися. Перш за все, упадає в очі, що на кожному кроці хтось (той, хто щойно ділив горіхи) не має жодного горіха (має нуль горіхів). Більш того, легко помітити, що горіхи завжди, починаючи з другого ходу, розподілені за законом

$$(a, b, 0),\ a > b > 0, \qquad (1)$$

причому людина з найбільшою кількістю горіхів щокроку зсувається в напрямку руху годинникової стрілки.

Це легко доводиться. Справді, завжди після першого кроку маємо розподіл $(0, d, d)$, $d > 0$. Але тоді на наступному кроці у другого не буде горіхів, а у третього більше горіхів, ніж у першого; оскільки зроблено два кроки, ми робимо зсув на двох осіб і отримуємо якраз розподіл (1).

Очевидно, те ж саме буде виконуватися і на всіх наступних кроках. Можна сказати й більше: якщо після якогось ходу ми мали розподіл $(a, b, 0)$, то після наступного ми

отримаємо (з поправкою на зсув) розподіл

$$(b + [a/2], [a/2], 0). \qquad (2)$$

Наслідок 1. Якщо спочатку було більше трьох горіхів, то, яка б не була їхня початкова кількість, принаймні три горіхи не будуть з'їдені.

Дійсно, горіхи їдять по одному. Припустимо, що їхня кількість весь час зменшується, тоді в деякий момент їх залишиться рівно три. Але з (1) видно, що для трьох горіхів можливий тільки розподіл $(2, 1, 0)$, і в цей момент, аналогічно до того, що ми вже бачили, процес вийшов на цикл, тобто поїдати горіхи більше не будуть.

Зауваження 1. Це неправильно, якщо із самого початку було лише три горіхи, оскільки в початковий момент формула (1) не виконується. При трьох або менше горіхах усі горіхи будуть з'їдені.

Перш ніж рухатися далі, домовимося про позначення.

По-перше, надалі перехід до наступного розподілу ми будемо позначати штрихом: якщо на m-му кроці ми мали розподіл K, то відповідний розподіл на $(m+1)$-му кроці будемо позначати K'. Відповідно, розподіл на $(m + r)$-му кроці ми позначимо через $K^{(r)}$.

Крім того, оскільки ми поки що займаємося тільки кількістю з'їдених горіхів, а хто їх з'їсть, нас не цікавить, то зручніше буде записувати розміщення горіхів на m-му кроці просто у формі $(a, b \mid m)$, де $a \geq b$. Інакше кажучи, ми, по-перше, завжди починаємо рахунок із того, у кого найбільше горіхів (він або вона наступного разу буде їх розподіляти), по-друге, записуємо номер кроку, і нарешті, відкидаємо третього учасника, в якого немає горіхів. Те, що на кожному кроці цей «третій зайвий» змінюється, для нас несуттєво.

Отже, формула (2) набуває тепер такого вигляду. Якщо на m-му кроці ми мали розподіл $K = (a, b \mid m)$, то на наступному кроці він буде таким:

$$K' = (b + [a/2], [a/2], m + 1). \qquad (3)$$

Зокрема, з формули (3) випливає, що якщо $K = (2k, k \mid m)$, то $K' = (k + 2k/2, 2k/2 \mid m + 1) = (2k, k \mid m + 1)$, тобто всі наступні розподілу будуть такими ж, ми вийшли на цикл, і процес фактично завершився.

Доведемо тепер, що якою б не була початкова кількість горіхів, через деякий час процес неодмінно вийде на цикл $(2k, k)$.

Для доведення для довільного розподілу горіхів $K = (a, b \mid m)$ розглянемо величину $r(m) = a - 2b$. На наступному кроці розподіл буде, залежно від парності числа a, або

$$K' = (b + a/2, a/2, m + 1), \qquad (4)$$

або

$$K' = (b + (a - 1)/2, (a - 1)/2, m + 1). \qquad (5)$$

У першому випадку ми отримуємо

$$r(m + 1) = b + \frac{a}{2} - a = b - \frac{a}{2} = \frac{-r(m)}{2}, \qquad (6)$$

а у другому —

$$r(m + 1) = b + \frac{a - 1}{2} - (a - 1) = \\ = b - \frac{a}{2} + \frac{1}{2} = \frac{-r(m)}{2} + \frac{1}{2}. \qquad (7)$$

Із цих формул видно, що $|r(m + 1)| < |r(m)|$, якщо $|r(m)| > 1$. Оскільки $r(m)$ ціле, то врешті-решт воно дорівнюватиме 1 або -1. Зазначимо, що в цей момент a, безумовно, непарне, і отже, ми маємо справу з другим випадком.

Якщо $r(m) = 1$, то згідно з формулою (7) маємо $r(m + 1) = 0$, тобто $a = 2b$ і процес вийшов на цикл. Якщо ж $r(m) = -1$, то з тієї ж формули випливає, що $r(m + 1) = 1$, а отже, він вийде на цикл на наступному кроці.

Приклад 1. Нехай в якийсь момент має місце розподіл $(8,3\mid m)$. Це означає, що $r(m)=2$. Тоді на наступних кроках маємо

$$\begin{aligned} K' &= (7,4\mid m+1), \quad & r(m+1) &= -1,\\ K'' &= (7,3\mid m+2), \quad & r(m+2) &= 1,\\ K''' &= (6,3\mid m+3), \quad & r(m+3) &= 0, \end{aligned}$$

тобто ми вийшли на цикл.

З наших міркувань видно також, що $r(m)$ кожен раз зменшується приблизно вдвічі (з неточністю за рахунок з'їдених горіхів). Тому, якщо спочатку було N горіхів, то вихід на цикл займе приблизно $\log_2 N$ кроків.

Приклад 2. Доведіть, що коли спочатку є 4^n горіхів, то вихід на цикл відбудеться за $2n+1$ крок.

Звідки береться цикл $(2k,k)$? Чому він саме такий?

Щоб у цьому розібратися, змінимо формулювання задачі. Припустимо, по-перше, що йдеться не про горіхи, а про дійсні числа, які (на відміну від горіхів) можна ділити необмежено, і, відповідно, відкинемо умову про «з'їдання». Відкинемо також третю людину, яка все одно завжди має нуль. Задача набуде такого вигляду.

Задача 2. Грають двоє: A і B. Перед кожним лежить папірець, на якому написано якесь число (у A — число a, у B — число b). Один хід полягає в тому, що вони закреслюють свої числа і замість них пишуть: A — число $b+a/2$, B — число $a/2$. Що станеться після досить великої кількості кроків?

Якщо спочатку $a=2b$, то числа просто-напросто не змінюватимуться («вийшли на цикл»). В іншому разі в цій грі (на відміну від гри з горіхами) розподіл ніколи не вийде на цикл, але прямуватиме до нього. Справді, розглянемо, як вище, величину $r(m)=a-2b$. Вона на кожному кроці буде зменшуватися за модулем рівно вдвічі, тобто прямуватиме до нуля, не досягаючи його. (У цьому відмінність від «горіхової» задачі, у якій $r(m)$ обов'язково ціле)

Приклад 3. Нехай початковий розподіл є $(5,7)$. Як було сказано, процес прямуватиме до циклу. Зрозуміло, що сума наших чисел завжди буде однакова. У цьому разі вона дорівнює 12, тобто граничним циклом буде $(8,4)$.

Наш початковий розподіл можна записати у вигляді $(8,4)+(-3,3)$. Легко бачити, що другий доданок на кожному ході зменшуватиметься вдвічі і мінятиме знак. Отже, після n ходів ми отримаємо розподіл

$$\left(8\pm\frac{3}{2^n},4\mp\frac{3}{2^n}\right)\to(8,4).$$

Зауваження 2. Є один випадок, коли розподіл не прямує до циклу, а саме, якщо спочатку сума чисел дорівнює 0, тобто ми мали розподіл $(s,-s)$. У цьому разі числа будуть щоразу множитися на $-1/2$, тобто прямувати до нуля.

Читач, трохи знайомий із матричним численням, легко помітить, що проведене нами щойно мінідослідження є нічим іншим як пошуком власних чисел нашого перетворення, і що ці власні числа дорівнюють $-1/2$ і 1, а відповідні їм власні вектори — це $(s,-s)$ і $(2k,k)$, тобто наш цикл.

Повернемося до задачі про горіхи і введемо два означення: *суми* і *нерівності* двох розподілів.

Нехай дано два розподіли на одному й тому ж ході: $K=(a,b\mid m)$ і $L=(c,d\mid m)$. Домовимося писати $K\le L$, якщо одночасно виконуються нерівності $a\le b$ і $c\le d$.

Сума також визначається природно: якщо розглядаються дві задачі, і в одній після m-го кроку маємо розподіл $K=(a,b\mid m)$, а в іншій, також після m-го кроку, $L=(c,d\mid m)$, то *сумою* цих задач ми назвемо розподіл $K+L=(a+c,b+d\mid m)$.

Як пов'язані між собою три поняття: крок гри (перехід від K до K'), сума і

нерівність? Відповідь дає таке основне твердження.

Теорема 1. *Якщо* $M \geq K + L$, *то* $M' \geq (K + L)' \geq K' + L'$.

Доведення. Перша нерівність легко випливає з формули (2). Доведемо другу. Нехай $K = (a, b)$, $L = (c, d)$. Якщо обидва числа a і c парні, то у формулі (2) усюди можна замість цілих частин $[a/2]$ і $[c/2]$ ставити просто дроби, звідки видно, що в цьому разі $(K + L)' = K' + L'$.

Ця рівність залишається справедливою і в разі, коли тільки одне з чисел a або c парне, а інше непарне (тоді треба відповідні дроби зменшити на $1/2$). Якщо ж обидва ці числа непарні, то, очевидно, рівність перетворюється на строгу нерівність:

$$(K + L)' = (b + d + (a + c)/2, (a + c)/2),$$

тоді як

$$K' + L' = (b + d + (a + c)/2 - 1, (a + c)/2 - 1).$$

Теорему доведено. □

Зокрема, ми бачимо, що справедливим є таке твердження.

Твердження 1. *При переході на наступний крок сума зберігається в тому і тільки в тому випадку, коли хоча б одне із чисел a або c є парним.*

Застосовуючи теорему 1 декілька разів, приходимо до такого твердження.

Твердження 2. *Якщо* $M \geq K+L$, *то також* $M^{(r)} \geq (K+L)^{(r)} \geq K^{(r)}+L^{(r)}$ *для будь-якого* r.

Зробимо ще два важливі зауваження. По-перше, строга нерівність, узагалі кажучи, не зберігається: вона може перетворитися на рівність. Наприклад, $(21, 7) > (20, 6)$, але, як легко переконатися, через два ходи ці розподіли перетворяться на однакові.

По-друге (дуже важливий для нас факт), написані нерівності справедливі саме для кількостей нез'їдених горіхів. Для кількості $E(N)$ з'їдених горіхів вони можуть не виконуватися. Більш того, можна говорити про те, що для цієї кількості справедлива обернена нерівність, а саме, маємо таку теорему.

Теорема 2. *Нехай* $N = M + S$. *Тоді* $E(N) \leq E(M) + E(S)$.

Доведення. Позначимо через $Q(N)$ кількість горіхів, які не будуть з'їдені, тоді, очевидно, $N = E(N) + Q(N)$. Якщо даний розподіл є сумою двох розподілів, то спочатку кількість горіхів суми дорівнює сумі кількості горіхів доданків. Але для $Q(N)$, як ми бачили, має місце нерівність $Q(N) \geq Q(M) + Q(S)$. Отже, для $E(N)$ справедлива нерівність у зворотний бік, тобто $E(N) \leq E(M) + E(S)$. □

Далі буде корисний наступний окремий випадок твердження 1.

Твердження 3. *Якщо в деякий момент розподіл горіхів є* $(2k + a, k + b \mid m)$, *то розподіл після будь-якої кількості кроків матиме вигляд* $(2k + a^{(r)}, k + b^{(r)} \mid m + n)$, *де* $(a^{(r)}, b^{(r)})$ *— розподіл, який виходить за r кроків з* (a, b).

Доведення. Застосовуємо твердження 1 і враховуємо, що перший доданок завжди буде парним, тому його можна застосовувати необмежену кількість разів. Отже, якщо з даного розподілу можна «виділити» цикл $(2k, k)$, то його можна не тільки виділити, але і викинути, не змінюючи наслідків. □

Враховуючи це, ми надалі писатимемо

$$(2k + a, k + b) \sim (a, b),$$

маючи на увазі, що обидва початкові набори завжди призводять до еквівалентної ситуації: кількість з'їдених горіхів буде однаковою, вони розподіляться між Анею, Борею й Вітею однаково і т. д. Наприклад,

$$(17, 13) = (16, 8) + (1, 5) \sim (1, 5).$$

Тепер підготовчу роботу нарешті завершено, і залишається лише сформулювати деякі остаточні результати: їх доведення вже практично очевидні.

Перш за все, сформулюємо окремий випадок твердження 3. Нехай спочатку кількість горіхів ділиться на 4, тоді маємо

$$(4k, 0) \to (2k, 2k) \to (3k, k) = \\ = (2k, k) + (k, 0) \sim (k, 0).$$

Отже, справедливе таке твердження.

Твердження 4. *Якщо не цікавитися тим, кому саме дістануться горіхи, а стежити тільки за кількістю з'їдених, то результат для чисел k і $4k$ однаковий.*

Однак, якщо стежити за тим, кому саме дістануться горіхи, то це вже не так. Перевірте самі випадок $k = 4$ і $4k = 16$.

Твердження 5. *Якщо $N = a \cdot 4^k$, $a \le 3$, то буде з'їдено a горіхів.*

Справді, як зазначалося вище, якщо кількість горіхів не більше трьох, то всі вони будуть з'їдені. Наш випадок відразу зводиться до цього, якщо k разів застосувати твердження 4.

Твердження 6. *а) Якщо $N = 4k + r$, $R < 2$, то $E(N) = E(k) + r$.*

б) Якщо $N = 16k + r$, $R < 4$, то $E(N) = E(k) + r$.

Доведення. а) За два кроки число $4k$, як ми вже бачили, перетвориться на k, а той єдиний горіх, який ще є (якщо він є), буде з'їдений. б) Доведення повністю аналогічне, з поправкою на те, що треба зробити не два, а чотири кроки, за які всі r горіхів, $r < 4$, будуть з'їдені. □

Звідси видно, що нам буде зручно розкласти дане число N за степенями четвірки або, що те саме, подати його в четвірковій системі числення.

Суму цифр у четвірковому записі ми будемо позначати $S(N)$. Наприклад, $97 = 1201_4$, оскільки це число можна представити у вигляді $97 = 1 \cdot 64 + 2 \cdot 16 + 0 \cdot 4 + 1$. Відповідно, $S(97) = 4$.

Скільки ж горіхів буде з'їдено? Згідно з твердженням 6, $E(97) = E(24) + 1 = E(6) + 1$. Останнє число можна знайти безпосередньо, але можна також скористатися такими міркуваннями:

- залишиться принаймні три горіхи;
- кількість горіхів, які залишилися, має ділитися на 3;
- хоча б один із шести горіхів буде з'їдений.

З урахуванням усього цього, виходить єдина можливість: $E(6) = 3$. Відповідно, $E(97) = 4$.

Отже, у цьому разі ми бачимо, що для $N = 97$ числа $S(N)$ і $E(N)$ рівні між собою. Чи завжди це так?

Ні, не завжди. Так, наприклад, якщо $N = 47 = 233_4$, то сума цифр дорівнює 8, проте з'їдено буде, як неважко переконатися, тільки 5 горіхів із виходом на цикл (28, 14).

Однак із тверджень 5 і 6 легко випливає, що рівність $S(N) = E(N)$ справедлива принаймні у двох випадках: якщо всі цифри у четвірковому записі числа рівні 0 або 1, а також, якщо перед кожною цифрою 2 або 3 стоїть нуль.

Справедливе також таке твердження.

Твердження 7. *Різниця $S(N) - E(N)$ невід'ємна та кратна трьом:*

$$S(N) - E(N) = 3m, \quad m \ge 0.$$

Доведення. Щоб довести першу частину твердження, зауважимо, розподіл (N) можна подати у вигляді суми декількох розподілів вигляду $(a \cdot 4^k)$, $a \le 3$, та скористаємося теоремою 2 та твердженням 5.

Доведемо другу частину. Для цього, по-перше, зауважимо, що різниця $N - E(N)$ є кількістю горіхів у циклі, а вона ділиться на 3. По-друге, помітимо, що числа N і $S(N)$ дають однакову остачу від ділення на

3. Це доводиться так само, як аналогічна властивість для десяткової системи та остач від ділення на 3 та на 9: замість того, що $10 - 1$ ділиться на 9, о отже, й на 3, використовується те, що $4 - 1$ ділиться на 3.

Отже, числа $E(N)$, N і $S(N)$ дають однакову остачу від ділення на 3. □

Нарешті, розглянемо питання, наскільки «чесно» розподіляться горіхи, і як дізнатися, кому вони дістануться.

Для цього перш за все слід дізнатися, кому вони дістануться, якщо кількість горіхів є одноцифровим числом (у четвірковій системі), тобто якщо $N \leq 3$.

Пряма перевірка показує, що

- якщо $N = 0$, то ніхто нічого не отримає;
- якщо $N = 1$, то єдиний горіх отримує Аня;
- якщо $N = 2$, то по одному горіху отримають Боря і Вітя;
- якщо $N = 3$, то кожен отримає по одному горіху.

Чи зміниться щось, якщо помножити кількість горіхів на 4? Вище було показано, що кількість з'їдених горіхів буде такою самою. Але це неправильно, якщо говорити про їхній розподіл: як ми бачили, якщо $N = 4k$, то на перших двох кроках горіхи з'їдені не будуть, а після другого кроку можна вважати, що їх залишилося k, але вони вже не в Ані: вони зрушили на 2 кроки.

Звідси вже легко зрозуміти, як саме розподіляться горіхи. Припустимо, наприклад, що їх спочатку було $N = 20301_4$. Тоді горіх, що відповідає останній цифрі числа, дістанеться Ані, три горіхи, що відповідає третій цифрі — по одному кожному, і нарешті, два горіхи від п'ятої цифри дістануться Віті й Ані.

З усього доведеного вже легко випливають, наприклад, такі факти, довести які ми пропонуємо читачам:

- Для довільного $\alpha > 0$ існує лише скінченна кількість таких N, для яких частка з'їдених горіхів $E(N)/N > \alpha$.
- Для довільного натурального числа E існує нескінченно багато таких N, для яких кількість з'їдених горіхів $E(N) = E$. Більш того, для будь-яких цілих невід'ємних p, q, r існує таке N, при якому Аня, Боря й Вітя отримають відповідно p, q і r горіхів.

Залишилося, зрозуміло, ще питання: скільки саме горіхів буде з'їдено в загальному випадку? Використовуючи проведені вище міркування, в цьому неважко розібратися в окремих випадках, але написати загальну формулу не так легко. Можливо, хтось із читачів захоче розібрати цю задачу глибше і знайти таку формулу.

Про деякі властивості чотирикутників

Петечук Василь Михайлович, Петечук Катерина Михайлівна,
Закарпатський інститут післядипломної педагогічної освіти
Петечук Юлія Василівна,
Закарпатський угорський інститут імені Ференца Ракоці II

У цій статті розглядаються багатокутники, базові означення площі і кутів багатокутників. На прикладі чотирикутників показано виникнення формул й ідей, які вплинули на розвиток планіметрії. Деякі твердження про трикутники випливають як окремі випадки геометрії чотирикутників, а деякі твердження про трикутник мають узагальнення або аналоги в геометрії чотирикутника. Зокрема, формула Герона обчислення площі трикутника (насправді це формула, яка була відкрита Архімедом за 200 років до Герона) випливає з формули площі чотирикутника як окремий випадок. Водночас формула Ейлера, що виражає відстань між центрами вписаного у трикутник й описаного навколо нього кіл через їхні радіуси, не виконується для чотирикутника, проте для нього справедлива формула Фусса, як формулювання, так і в доведення якої мають певну аналогію з формулою Ейлера.

У цій статті для $n \geq 3$ багатокутником (n-кутником) $A_1 \dots A_n$ ми називатимемо геометричну фігуру, яка складається з точок $A_1, \dots, A_n$, жодні три з яких не лежать на одній прямій, і відрізків A_1A_2, A_2A_3, $\dots$, $A_{n-1}A_n$, A_nA_1, що попарно не перетинаються. Точки $A_1, \dots, A_n$ прийнято називати вершинами багатокутника, а відрізки A_1A_2, $A_2A_3, \dots, A_nA_1$ — його сторонами. Для зручності запису загальних формул часто ототожнюють вершину A_1 з вершиною A_{n+1}, а вершину A_n з вершиною A_0.

Багатокутник називають опуклим, якщо він лежить по один бік від будь-якої прямої, що містить сторону багатокутника; інакше багатокутник називається неопуклим. Неопуклі багатокутники бувають без самоперетинів сторін (прості), а також із самоперетинами сторін. Зрозуміло, що трикутники є опуклими багатокутниками.

Площею S простого багатокутника $A_1 \dots A_n$ називається модуль суми площ трикутників $A_1A_{i-1}A_i$, $2 < i \leq n$, узятих зі знаком плюс, якщо обхід трикутника здійснюється за годинниковою стрілкою, і зі знаком мінус у протилежному разі.

Кутом багатокутника $A_1 \dots A_n$ при вершині A_i, $1 \leq i \leq n$, називається кут, на який потрібно повернути проти годинникової стрілки навколо вершини A_i напівпряму $A_{i-1}A_i$, щоб вона перейшла в напівпряму A_iA_{i+1}; він позначається $\angle A_i$.

Зокрема, якщо $n = 3$, то вищезгадані означення площі й кутів багатокутника збігаються з означеннями площі й кутів трикутника. Якщо $n = 4$, то в чотирикутнику $ABCD$ вони мають такий вигляд:

- для опуклого чотирикутника

$$S = S_{\triangle ABC} + S_{\triangle ACD},$$
$$\angle A = \angle DAC, \angle B = \angle ABC,$$
$$\angle C = \angle BCD, \angle D = \angle ADC;$$

- для неопуклого простого чотирикутника з вершиною D всередині трикутника ABC

$$S = |S_{\triangle ABC} - S_{\triangle ACD}| = S_{\triangle BCD} + S_{\triangle BDA},$$
$$\angle A = \angle DAC, \angle B = \angle ABC,$$
$$\angle C = \angle BCD, \angle D = 360^\circ - \angle ADC.$$

Формула суми квадратів діагоналей паралелограма і трапеції. Формула Ейлера

Нехай $ABCD$ — паралелограм, B_1 і C_1 — основи перпендикулярів із вершин B і C на пряму AD (рис. 1).

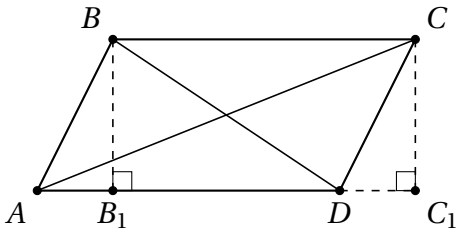


Рис. 1

Скористаємось тим, що різниця квадратів похилих дорівнює різниці квадратів їх проєкцій:

$$AC^2 - CD^2 = AC_1^2 - C_1D^2$$

і

$$BD^2 - AB^2 = B_1D^2 - AB_1^2.$$

Додавши ці рівності, знаходимо, що

$$AC^2 + BD^2 = CD^2 + AB^2 +$$
$$+AC_1^2 - C_1D^2 + B_1D^2 - AB_1^2.$$

Використовуючи рівності

$$B_1D^2 - AB_1^2 = AD(B_1D - AB_1),$$
$$AC_1^2 - C_1D^2 = AD(AC_1 + C_1D),$$

одержимо

$$AC_1^2 - C_1D^2 + B_1D^2 - AB_1^2 =$$
$$= AD(AC_1 + C_1D + B_1D - AB_1) =$$
$$= 2AD \cdot BC.$$

Тим самим ми довели, що

$$AC^2 + BD^2 = CD^2 + AB^2 + 2AD \cdot BC =$$
$$= 2(AB^2 + BC^2).$$

Якщо позначити $a = AB$, $b = BC$, $c = CD$, $d = DA$, $d_1 = AC$, $d_2 = BD$, то отримана формула має вигляд

$$d_1^2 + d_2^2 = 2(a^2 + b^2).$$

Вона означає, що сума квадратів діагоналей паралелограма дорівнює сумі квадратів його сторін.

З неї випливає, що квадрат подвоєної медіани трикутника є різницею подвоєної суми квадратів прилеглих сторін і квадрата сторони, до якої її проведено:

$$4m_a^2 = 2(b^2 + c^2) - a^2.$$

Використаємо формулу довжини медіани трикутника для відшукання відстані між серединами діагоналей AC і BD чотирикутника $ABCD$, які позначимо M і N відповідно (рис. 2).

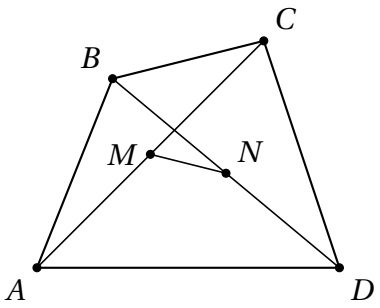


Рис. 2

З формули квадрата медіани маємо, що

$$4AN^2 = 2(AB^2 + AD^2) - BD^2,$$
$$4CN^2 = 2(BC^2 + CD^2) - BD^2,$$
$$4MN^2 = 2(AN^2 + CN^2) - AC^2 =$$
$$= AB^2 + BC^2 + CD^2 +$$
$$+AD^2 - AC^2 - BD^2.$$

Отримана формула для відшукання відстані між точками M і N називається *формулою Ейлера*. У коротких позначеннях вона має вигляд

$$4MN^2 = a^2 + b^2 + c^2 + d^2 - d_1^2 - d_2^2.$$

Зокрема, якщо $ABCD$ — трапеція з основами BC і AD, то пряма MN паралельна основам трапеції і

$$MN = \frac{|AD - BC|}{2}.$$

У такому разі

$$4MN^2 = AD^2 + BC^2 - 2AD \cdot BC.$$

Зіставляючи отриману рівність із формулою Ейлера, знаходимо, що

$$AC^2 + BD^2 = AB^2 + CD^2 + 2AD \cdot BC.$$

Іншими словами, сума квадратів діагоналей трапеції дорівнює сумі квадратів бічних сторін і подвоєного добутку основ:

$$d_1^2 + d_2^2 = a^2 + c^2 + 2bd.$$

Часто буває корисним таке формулювання: подвоєний добуток основ трапеції дорівнює різниці суми квадратів діагоналей і суми квадратів бічних сторін.

Формули Ейлера і Фусса

Формула довжини медіани трикутника може бути використана для знаходження відстані між центром вписаного кола I і центром описаного кола O опуклого чотирикутника, в який можна вписати і описати навколо нього кола радіусів r і R відповідно (рис. 3). Це відома формула Фусса

$$\frac{1}{(R-OI)^2} + \frac{1}{(R+OI)^2} = \frac{1}{r^2}.$$

Адже якщо провести бісектриси з вершин A і C до перетину з описаним колом у точках M і N відповідно, то MN стягує дугу, яка складається з дуг, на які спираються вписані кути α і β.

Оскільки

$$\alpha + \beta = \frac{\angle A + \angle C}{2} = 90^\circ,$$

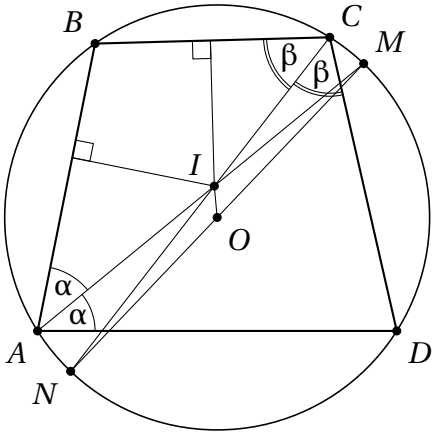


Рис. 3

то MN є діаметром кола, а точка O — його серединою. За формулою довжини медіани

$$4OI^2 = 2(IM^2 + IN^2) - MN^2.$$

Нехай $m = OI^2 - R^2$ — степінь точки I відносно кола, описаного навколо трикутника ABC.

Справедлива рівність

$$AI \cdot IM = CI \cdot IN = (R - OI)(R + OI),$$

з якої випливає, що

$$IM^2 + IN^2 = \frac{m^2}{AI^2} + \frac{m^2}{CI^2}.$$

Оскільки

$$AI = \frac{r}{\sin\alpha} \quad \text{і} \quad CI = \frac{r}{\sin\beta} = \frac{r}{\cos\alpha},$$

то

$$IM^2 + IN^2 = \frac{m^2}{r^2}.$$

Отже,

$$2OI^2 = \frac{m^2}{r^2} - 2R^2.$$

Після перетворення знаходимо, що

$$\frac{1}{r^2} = \frac{2(OI^2 + R^2)}{m^2} = \frac{2(OI^2 + R^2)}{(R^2 - OI^2)^2} = \\ = \frac{1}{(R-OI)^2} + \frac{1}{(R+OI)^2}.$$

Подібні міркування можна застосувати до трикутника ABC (рис. 4).

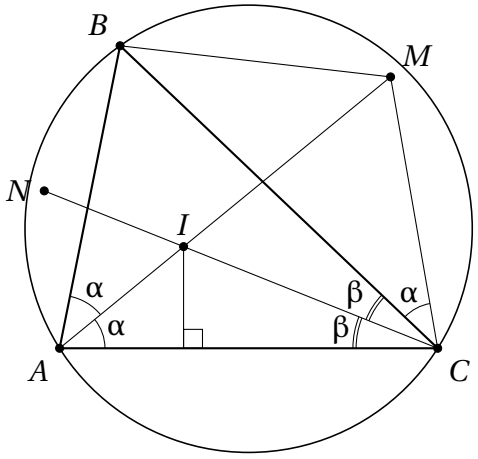


Рис. 4

Стверджувати, що MN є діаметром кола й застосовувати формулу довжини медіани не можна. Однак рівності

$$AI \cdot IM = CI \cdot IN = (R - OI)(R + OI)$$

і

$$AI = \frac{r}{\sin \alpha}$$

виконуються. Оскільки

$$\alpha + \beta = \frac{\angle A + \angle C}{2} = \angle MIC = \angle MCI,$$

то

$$IM = MC = 2R \sin \alpha.$$

Звідси маємо

$$AI \cdot IM = 2rR.$$

Тим самим доведено, що

$$R^2 - OI^2 = 2rR.$$

Отже, має місце формула Ейлера

$$OI^2 = R^2 - 2rR,$$

яка може бути записана у вигляді, аналогічному до формули Фусса для чотирикутника

$$\frac{1}{R - OI} + \frac{1}{R + OI} = \frac{1}{r}.$$

Нерівність Птолемея – Гіппарха. Теорема Бретшнейдера

Нехай $ABCD$ — чотирикутник, $a = AB$, $b = BC$, $c = CD$, $d = DA$, $d_1 = AC$, $d_2 = BD$. Повернемо трикутник ABC навколо вершини C проти годинникової стрілки таким чином, щоб із точністю до подібності з коефіцієнтом k, трикутник ABC перейшов у трикутник DMC. У такому разі вершина A перейде у вершину D, а вершина B — у вершину M (рис. 5–7). Крім того, $CD = kd_1 = c$, $MD = ka$, $CM = kb$, $\angle B = \angle ABC = \angle DMC$.

Оскільки $\angle BCM = \angle ACD$ і

$$\frac{BC}{AC} = \frac{b}{d_1} = \frac{CM}{CD},$$

то трикутники BCM і ACD подібні. Тому

$$\frac{BM}{AD} = \frac{BC}{AC} = \frac{CM}{CD} = \frac{b}{d_1},$$

і $\angle BCM = \angle ADC$.

Це означає, що $BM = \frac{db}{d_1}$ і

$$\angle BMD = |\angle BMC \pm \angle DMC|,$$

де плюс береться у разі опуклого чотирикутника (рис. 5), а мінус — у разі неопуклого чотирикутника (рис. 6 і 7). За нерівністю трикутника й теоремою косинусів, застосованою до трикутника BMD, отримуємо відповідно нерівність

$$BD \leq BM + MD$$

(нерівність Птолемея – Гіппарха), а також рівність

$$BD^2 = BM^2 + MD^2 - 2BM \cdot MD \cos \angle BMD$$

(теорема Бретшнейдера). Теорему Бретшнейдера ще називають другою теоремою косинусів.

Зауважимо, що нерівність Птолемея – Гіппарха випливає з теореми Бретшнейдера. Рівність у нерівності Птолемея – Гіппарха досягається тільки тоді, коли то-

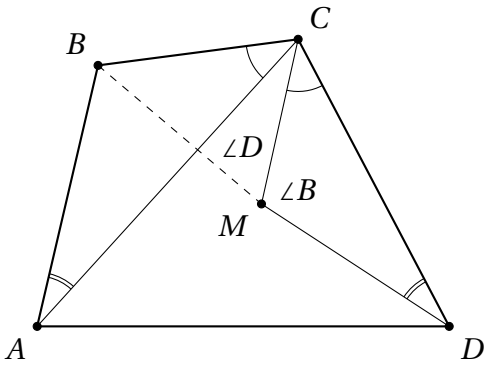


Рис. 5

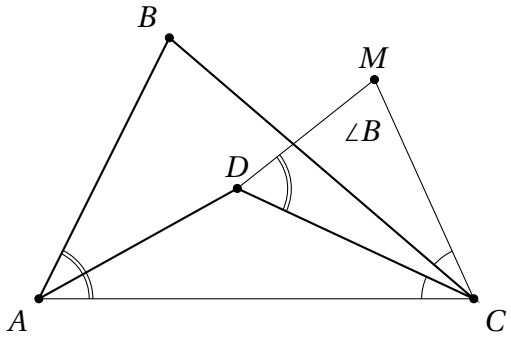


Рис. 6

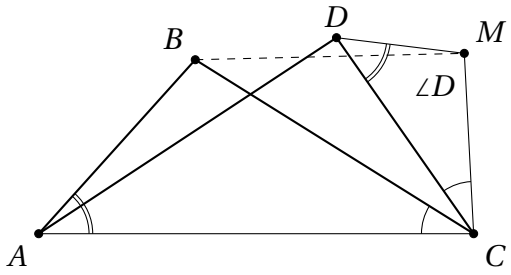


Рис. 7

чка M належить відрізку BD. Оскільки

$$\angle BMC = \angle ADC = \angle D,$$
$$\angle BMD = \angle B + \angle D \text{ (рис. 5)},$$
$$\angle BMC = \angle ADC = 360^\circ - \angle D \text{ (рис. 6, 7)},$$
$$\angle BMD = \angle BMC - \angle DMC =$$
$$= 360^\circ - \angle D - \angle B \text{ (рис. 6)},$$
$$\angle BMD = \angle DMC - \angle BMC =$$
$$= \angle B - (360^\circ - \angle D) =$$
$$= \angle B + \angle D - 360^\circ \text{ (рис. 7)},$$

то в усіх випадках

$$\cos \angle BMD = \cos(\angle B + \angle D).$$

Очевидно, що

$$BD \cdot d_1 = d_1 d_2, \quad BM \cdot d_1 = bd,$$
$$MD \cdot d_1 = kad_1 = ac.$$

Перепишемо нерівність Птолемея – Гіппарха й теорему Бретшнейдера в коротких позначеннях, попередньо домноживши їх на d_1 і d_1^2 відповідно:

$$d_1 d_2 \le ac + bd,$$
$$d_1^2 d_2^2 = (ac)^2 + (bd)^2 -$$
$$- 2abcd \cos(\angle B + \angle D).$$

З'ясуємо, коли досягається рівність у нерівності Птолемея – Гіппарха. Зрозуміло, що це можливо тільки тоді, коли точка M належить відрізку BD.

В опуклому чотирикутнику $ABCD$ (рис. 8) рівність у нерівності Птолемея – Гіппарха досягається тільки тоді, коли точка M належить діагоналі BD, тобто коли навколо опуклого чотирикутника $ABCD$ можна описати коло.

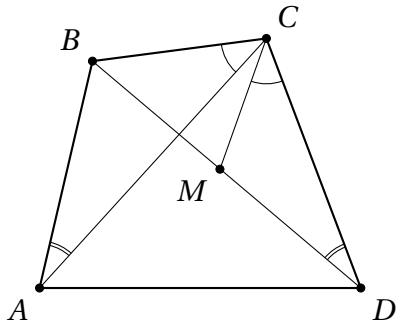


Рис. 8

Якщо чотирикутник $ABCD$ неопуклий, то точка M не може належати відрізку BD. Тому нерівність Птолемея – Гіппарха для неопуклого чотирикутника $ABCD$ строга.

Водночас нерівності $MD \le MB + BD$ і $MB \le MD + DB$ в опуклому чотирикутнику $ABCD$ строгі (рис. 8), але в неопуклому чотирикутнику $ABCD$ із самоперетинами сторін може досягатися рівність $MB = MD + DB$ (рис. 9).

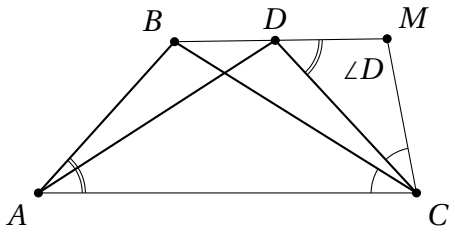


Рис. 9

Найпростіші формули площі чотирикутника

Нехай $ABCD$ — чотирикутник, $a = AB$, $b = BC$, $c = CD$, $d = DA$, $d_1 = AC$, $d_2 = BD$, φ — кут між діагоналями або їх продовженнями, B_1 і D_1 — основи перпендикулярів, опущених із вершин B і D на пряму AC відповідно, B_2 — основа перпендикуляра, опущеного з вершини D на пряму BB_1, S — площа чотирикутника $ABCD$.

Тоді

$$S = \left|\frac{1}{2}\,BB_1 \cdot AC \pm \frac{1}{2}\,DD_1 \cdot AC\right| = \frac{1}{2}\,AC \cdot BD \sin\varphi,$$

де знак плюс береться в разі опуклого чотирикутника (рис. 10) і мінус — у разі неопуклого чотирикутника (рис. 11).

У коротких позначеннях формула площі чотирикутника має вигляд

$$S = \frac{1}{2}\,d_1 d_2 \sin\varphi.$$

Перетворимо отриману формулу

$$S = \frac{1}{2}\,AC \cdot BB_2,$$

скориставшись теоремою Піфагора

$$4S^2 = AC^2 \cdot BB_2^2 = AC^2(BD^2 - DB_2^2) = AC^2 \cdot BD^2 - (AC \cdot BD_2)^2.$$

Оскільки в трапеції подвоєний добуток основ дорівнює різниці суми квадратів

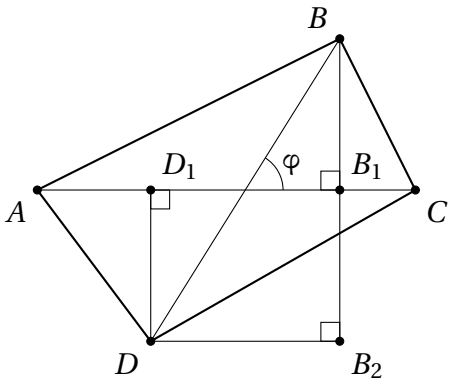


Рис. 10

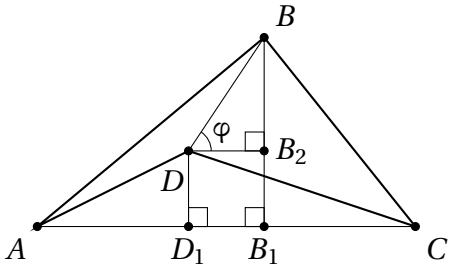


Рис. 11

бічних сторін і суми квадратів діагоналей, то

$$16S^2 = 4AC^2 \cdot BD^2 - (AD^2 + CB_2^2 - AB_2^2 - CD^2)^2.$$

За теоремою Піфагора різниця квадратів похилих дорівнює різниці квадратів їхніх проєкцій. Тому

$$CB_2^2 - AB_2^2 = CB_1^2 - AB_1^2 = CB^2 - AB^2.$$

Отже, справедлива формула

$$16S^2 = 4AC^2 \cdot BD^2 - (AD^2 + CB^2 - AB^2 - CD^2)^2.$$

У коротких позначеннях формула площі чотирикутника має вигляд

$$16S^2 = 4d_1^2 d_2^2 - (a^2 + c^2 - b^2 - d^2)^2.$$

Нехай

$$p = \frac{a + b + c + d}{2}.$$

За формулою різниці квадратів

$$16(p-a)(p-b)(p-c)(p-d) =$$
$$= (a+b+c-d)(a+b-c+d)\times$$
$$\times(a-b+c+d)(-a+b+c+d) =$$
$$= ((a+c)^2-(b-d)^2)\times$$
$$\times((b+d)^2-(a-c)^2) =$$
$$= \big(2(ac+bd)+a^2+c^2-b^2-d^2\big)\times$$
$$\times\big(2(ac+bd)-(a^2+c^2-b^2-d^2)\big) =$$
$$= 4(ac+bd)^2-(a^2+c^2-b^2-d^2)^2.$$

Тому формула площі чотирикутника задовольняє рівність

$$16S^2 = 4d_1^2d_2^2 + 16(p-a)(p-b)\times$$
$$\times(p-c)(p-d) - 4(ac+bd)^2.$$

Це означає, що справедлива рівність

$$4S^2 = 4(p-a)(p-b)(p-c)(p-d)+$$
$$+d_1^2d_2^2-(ac+bd)^2.$$

Узагальнення формули Герона площі чотирикутника

Перепишемо знайдену формулу площі чотирикутника. Для цього скористаємось теоремою Бретшнейдера

$$d_1^2d_2^2 = (ac)^2+(bd)^2-$$
$$-2abcd\cos(\angle B+\angle D) =$$
$$= (ac+bd)^2-$$
$$-2abcd\big(1+\cos(\angle B+\angle D)\big) =$$
$$= (ac+bd)^2 - 4abcd\cos^2\frac{\angle B+\angle D}{2}.$$

Тому має місце ще одна формула площі чотирикутника, яка є узагальненням формули Герона площі трикутника:

$$S^2 = (p-a)(p-b)(p-c)(p-d)-$$
$$-abcd\cos^2\frac{\angle B+\angle D}{2}.$$

Зрозуміло, що формула Герона – Архімеда випливає із цієї формули при $d=0$.

Якщо чотирикутник $ABCD$ вписаний, то $\angle B+\angle D = 180^\circ$ і виконується формула Брахмагупти:

$$S^2 = (p-a)(p-b)(p-c)(p-d).$$

Якщо чотирикутник $ABCD$ описаний, то $a+c=b+d=p$ і

$$S^2 = abcd\sin^2\frac{\angle B+\angle D}{2}.$$

Якщо ж чотирикутник $ABCD$ і вписаний, і описаний, то

$$S^2 = abcd.$$

Векторно-координатна формула площі чотирикутника

Розглянемо $\triangle ABC$, який побудований на ненульових векторах $\vec{a}=\overrightarrow{AB}$ і $\vec{b}=\overrightarrow{AC}$, які виходять з однієї точки. Нехай S — площа $\triangle ABC$, a і b — абсолютні величини векторів $\vec{a}$ і $\vec{b}$, а γ — кут між ними.

За формулою площі трикутника

$$S = \frac{1}{2}ab\sin\gamma$$

і

$$4S^2 = a^2b^2(1-\cos^2\gamma).$$

Тому справедлива векторна формула площі трикутника ABC

$$4S^2 = \vec{a}^2\vec{b}^2 - \big(\vec{a}\vec{b}\big)^2.$$

Зокрема, якщо $\vec{a}(a_1;a_2)$ і $\vec{b}(b_1;b_2)$, то виконується координатна рівність

$$4S^2 = (a_1^2+a_2^2)(b_1^2+b_2^2) -$$
$$-(a_1b_1+a_2b_2)^2 = (a_1b_2-a_2b_1)^2.$$

Тому площа трикутника ABC обчислюється за координатною формулою

$$2S = |a_1b_2 - a_2b_1|.$$

Виявляється, що якщо при обході вздовж вектора $\vec{a}$ трикутник ABC, побудований на векторах $\vec{a}$ і $\vec{b}$, що виходять з однієї точки, залишається зліва, то площа S трикутника ABC обчислюється за формулою

$$2S = a_1b_2 - a_2b_1,$$

і за формулою

$$2S = a_2b_1 - a_1b_2$$

у протилежному разі. Це випливає з такої леми.

Лема 1. *Площа S трикутника ABC, побудованого на ненульових векторах $\vec{m}(a;b)$ і $\vec{n}(x;y)$, що виходять з однієї точки, обчислюється за формулою*

$$2S = ay - bx,$$

якщо вектор $\vec{n}$ міститься в лівій півплощині відносно вектора $\vec{m}$, і за формулою

$$2S = bx - ay$$

у протилежному разі.

Доведення. Для доведення леми розглянемо функцію

$$f(x,y) = ay - bx.$$

Функція f набуває нульових значень тільки тоді, коли вектор $\vec{n}$ колінеарний вектору $\vec{m}$, тобто коли точки з координатами $(x;y)$ належать прямій, яка містить вектор $\vec{m}$.

Функція $f(x,y)$ є неперервною, а тому зберігає знак у півплощинах, на які ця пряма ділить числову площину. Визначимо знак функції $f(x,y)$ у цих півплощинах.

Вектор $\vec{m}_1(-b;a)$ утворюється поворотом вектора $\vec{m}(a;b)$ на 90° проти годинникової стрілки відносно фіксованого початку вектора $\vec{m}$.

У такому разі $f(-b,a) = a^2 + b^2 > 0$. Це означає, що площа трикутника, побудованого на векторах $\vec{m}(a;b)$ і $\vec{n}(x;y)$ з фіксованим початком обчислюється за формулою

$$2S = ay - bx,$$

якщо при русі вздовж вектора $\vec{m}$ вектор $\vec{n}$ міститься у півплощині, що є зліва від вектора $\vec{m}$, і за формулою

$$2S = bx - ay$$

у протилежному разі. □

Це означає, що площа трикутника, побудованого на векторах $\vec{a}(a_1;a_2)$ і $\vec{b}(b_1;b_2)$ з фіксованим початком, обчислюється за формулою

$$2S = a_1b_2 - a_2b_1,$$

якщо при обході трикутника вздовж вектора $\vec{a}$ трикутник залишається зліва, і за формулою

$$2S = a_2b_1 - a_1b_2$$

у протилежному разі.

Зокрема, якщо $A(x_1;y_1)$, $B(x_2;y_2)$, $C(x_3;y_3)$ — вершини трикутника ABC такого, що при обході від вершини A до вершини B і від вершини B до вершини C трикутник ABC залишається зліва, то $\vec{m} = \overrightarrow{AB}$, $\vec{n} = \overrightarrow{AC}$ і площа S трикутника ABC знаходиться за формулою

$$\begin{gathered}2S = (x_2 - x_1)(y_3 - y_1) - \\ -(x_3 - x_1)(y_2 - y_1) = \\ = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2).\end{gathered}$$

Скориставшись цією формулою, знайдемо площу S опуклого чотирикутника $ABCD$ з вершинами $A(x_1;y_1)$, $B(x_2;y_2)$, $C(x_3;y_3)$, $D(x_4;y_4)$. Очевидно, що

$$S = S_{\triangle ABC} + S_{\triangle ACD}.$$

Вважаємо, що при обході чотирикутника $ABCD$ він залишається зліва. Згідно з

формулами площ трикутників ABC і ACD знаходимо

$$2S_{\triangle ABC} = (x_2 - x_1)(y_3 - y_1) - \\ -(x_3 - x_1)(y_2 - y_1) = \\ = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2),$$

$$2S_{\triangle ACD} = (x_3 - x_1)(y_4 - y_1) - \\ -(x_4 - x_1)(y_3 - y_1) = \\ = x_1(y_3 - y_4) + x_3(y_4 - y_1) + x_4(y_1 - y_3).$$

Тому

$$2S = x_1(y_1 - y_3 + y_2 - y_1 + y_1 - y_4 + \\ + y_3 - y_1) + x_2(y_3 - y_1) + \\ + x_3(y_4 - y_1) + x_4(y_3 - y_1) = \\ = x_1(y_2 - y_4) + x_2(y_3 - y_1) + \\ + x_3(y_4 - y_2) + x_4(y_1 - y_3).$$

Аналогічно до того, як була знайдена формула площі опуклого чотирикутника, знаходиться формула площі довільного опуклого багатокутника.

Площа опуклого багатокутника

Нехай $A_1(x_1; y_1)$, …, $A_n(x_n; y_n)$ — послідовні вершини опуклого багатокутника $A_1 \dots A_n$, $n \geq 3$. Вважаємо, що при обході багатокутника $A_1 \dots A_n$ у напрямку від A_1 до A_2, …, від A_{n-1} до A_n, від A_n до A_1 багатокутник $A_1 \dots A_n$ залишається зліва. Випадки $n = 3$ і $n = 4$ розглянуто вище. Нехай $n > 4$.

Неважко бачити, що багатокутник $A_1 \dots A_n$ складається з $(n-2)$-х трикутників $\triangle A_1A_2A_3$, $\triangle A_1A_3A_4$,…, $\triangle A_1A_{n-1}A_n$. Тому площа S багатокутника $A_1 \dots A_n$ є сумою площ цих трикутників. Згідно з формулами площ трикутників справедлива рівність

$$2S = (x_2 - x_1)(y_3 - y_1) - \\ -(x_3 - x_1)(y_2 - y_1) + \dots + \\ + (x_{n-1} - x_1)(y_n - y_1) - \\ - (x_n - x_1)(y_{n-1} - y_1).$$

Отриману формулу можна записати інакше:

$$2S = x_1(y_1 - y_3 + y_2 - y_1 + y_1 - y_4 + \dots + \\ + y_1 - y_n + y_{n-1} - y_1) + \\ + x_2(y_3 - y_1) + \dots + x_n(y_1 - y_{n-1}).$$

Після перетворень вона набуває вигляду

$$2S = x_1(y_2 - y_n) + x_2(y_3 - y_1) + \dots + \\ + x_n(y_1 - y_{n-1}).$$

Якщо скористатися ототожненням вершини $A_{n+1}(x_n + 1; y_n + 1)$ з вершиною $A_1(x_1; y_1)$, а вершини $A_n(x_n; y_n)$ з вершиною $A_0(x_0; y_0)$, при якому x_n ототожнюється з x_0, а x_1 ототожнюються з x_{n+1}, а також y_n ототожнюється з y_0, а y_1 ототожнюється з y_{n+1}, то формула площі багатокутника $A_1 \dots A_n$ набуває вигляду

$$2S = \sum_{i=1}^{n} x_i(y_{i+1} - y_{i-1}),$$

або

$$2S = \sum_{i=1}^{n} (x_{i-1} - x_{i+1})y_i.$$

Література

[1] Петечук В. М. Геометрія для 8-го класу з поглибленим вивченням математики / В. М. Петечук. — Ужгород : Карпати, 1992. — 128 с.

[2] Бевз Г. П. Геометрія чотирикутника / Г. П. Бевз. — Х. : Видав. гр. «Основа», 2003. — 80 с. — (Серія «Бібліотека журналу "Математика в школах України"»; Вип. 5).

[3] Ясінський В. А. Секрети підготовки школярів до Всеукраїнських та Міжнародних олімпіад. Геометрія : навчально-методичний посібник / В. А. Ясінський, О. Б. Панасенко. — Вінниця : ТОВ «Нілан-ЛТД», 2014. — 224 с.

[4] Кушнир И. А. Школа формульной геометрии / И. А. Кушнир — Днепр : Середняк Т. К., 2018. — 360 с.

[5] Збірник задач республіканських математичних олімпіад / В. І. Михайловський та ін. — К. : Вища школа, 1979.

[6] Математичні олімпіади школярів України: 1991–2000 / В. М. Лейфура та ін. — К. : Техніка, 2003. — 541 с.

[7] Математичні олімпіади школярів України: 2001–2006 / В. М. Лейфура та ін. — Львів : Каменяр, 2008. — 348 с.

[8] Математичні олімпіади школярів України 2011–2012 / І. М. Мітельман та ін. — Чернівці : 2013. — 134 с.

[9] Математичні олімпіадні змагання школярів України 2006–2007 / А. В. Анікушін та ін. — К. : Літера ЛТД, 2008. — 135 с.

[10] Математичні олімпіадні змагання школярів України 2007–2008 та 2008–2009 / А. В. Анікушін та ін.; за ред. Б. В. Рубльова — Львів : Каменяр, 2010. — 549 с.

Із яких точок простору Землю і Місяць видно під однаковими кутами

Вишенська Оксана Володимирівна,
Національний транспортний університет

У цій статті ми спробуємо дати відповідь на поставлене в заголовку питання, а саме, визначити геометричне місце точок тривимірного простору, з яких дві дані кулі видно від однаковими кутами. Спершу ми маємо виконати певну підготовчу роботу.

Поділ відрізка в даному відношенні

Нехай маємо пряму l і на ній відрізок AB (рис. 1). Запитаємо: чи є на прямій l така точка C, щоб відношення відстаней від неї до точок A і B дорівнювало деякому числу $\frac{m}{n}$, тобто, щоб $\frac{CA}{CB} = \frac{m}{n}$?

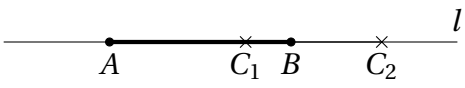


Рис. 1

Скільки є таких точок? Передусім, така точка є на самому відрізку AB (на рис. 1 її позначено C_1). Справді, нехай $AB = a$. Позначивши відстань від A до шуканої точки через x, знаходимо її місце на відрізку AB з рівняння

$$\frac{x}{a-x} = \frac{m}{n}.$$

Отримуємо

$$AC_1 = \frac{am}{m+n}, \quad BC_1 = \frac{an}{m+n}.$$

Про точку C_1 кажуть, що вона ділить відрізок AB у відношенні $\frac{m}{n}$ внутрішнім чином.

Якщо $\frac{m}{n} = 1$, то точка C_1 лежить посередині відрізка AB і на прямій l більше немає точок, рівновіддалених від A і B .

Інакше буде тоді, коли $\frac{m}{n} \neq 1$. Від точок C, що лежать на прямій l праворуч від B, до точки B ближче, ніж до A, а тому для кожної з них

$$\frac{CA}{CB} > 1.$$

Але чи є серед них точка C, для якої це відношення дорівнює даному заздалегідь числу $\frac{m}{n}$ за умови, що $m > n$?

Позначимо через y відстань від точки C до B. Тоді відстань CA становитиме $y + a$. Беремо будь-яке число $\frac{m}{n} > 1$. Згадана точка існуватиме, якщо рівняння

$$\frac{y+a}{y} = \frac{m}{n}$$

має розв'язок. Це справді так. Його розв'язок

$$y = \frac{an}{m-n}.$$

Отже, праворуч від точки B (за межами відрізка AB) є одна точка (на рис. 1 її позначено C_2), відстані від якої до точок A і B відносяться як m до n. Відстані від неї до точок A і B становлять відповідно

$$C_2A = \frac{am}{m-n}, \quad C_2B = \frac{an}{m-n}.$$

Про точку C_2 кажуть, що вона ділить відрізок AB у відношенні $\frac{m}{n}$ зовнішнім чином.

Якщо $\frac{m}{n} < 1$, то ролі точок A і B поміняються. Точка C, що ділить відрізок AB у відношенні $\frac{m}{n}$ зовнішнім чином, лежатиме ліворуч від A.

Підсумуємо отримані результати. Є тільки одна точка, що ділить відрізок AB у відношенні 1:1. Це середина відрізка AB. Якщо $m \neq n$, то у відношенні $m : n$ відрізок AB ділять дві точки. Одна з них лежить на відрізку AB, а друга на його продовженні за точку A, якщо $m < n$, і за точку B, якщо $m > n$. Перша з них ділить відрізок внутрішнім чином, а друга зовнішнім.

Нам знадобляться властивості внутрішньої та зовнішньої бісектрис трикутника. Нехай ABC — довільний трикутник, CK — його внутрішня бісектриса, a і b — відповідно довжини сторін BC і AC.

Тоді

$$\frac{AK}{KB} = \frac{b}{a}.$$

Проведемо зовнішню бісектрису трикутника через вершину C. Оскільки вона перпендикулярна внутрішній бісектрисі CK, то вона паралельна стороні AB тоді й тільки тоді, коли CK є висотою, тобто коли трикутник є рівнобедреним з основою AB, або $a = b$.

Таким чином, коли $a \neq b$, то зовнішня бісектриса перетинає сторону AB. Позначимо цю точку L та припустимо, що вона лежить на продовженні AB за точку B (рис. 2).

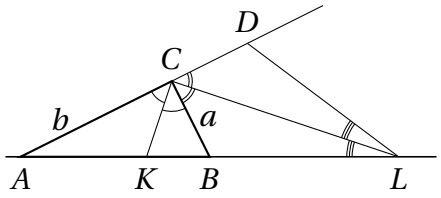


Рис. 2

На продовженні сторони AC трикутника ABC за точку C відкладаємо відрізок CD, що дорівнює CB. Сполучаємо точку D з L. Придивимося пильно до трикутників BCL і DCL. Вони рівні, бо $BC = DC$, CL — їхня спільна сторона і $\angle BCL = \angle DCL$ (адже CL — бісектриса кута BCD). Із рівності цих двох трикутників висновуємо, що $\angle BLC = \angle DLC$, тобто, що LC — бісектриса кута BLD. Кут BLD є внутрішнім кутом трикутника ALD, із чого випливає

$$\frac{LA}{LD} = \frac{AC}{CD} = \frac{b}{a}.$$

Ця рівність означає, що точка L ділить зовнішнім чином відрізок AB у відношенні $b : a$. Наше припущення про те, що точка L лежить праворуч від B, рівнозначне умові $a > b$.

Коло Аполлонія

Візьмімо на площині дві точки A і B, два числа m і n, та запитаймо: для яких точок M площини відстані AM і BM відносяться так, як числа m і n:

$$AM : BM = m : n?$$

Із шкільних підручників із геометрії ми знаємо відповіді на це запитання в тому випадку, коли $m = n$.

На однакових відстанях від A і B перебувають ті й тільки ті точки площини, які лежать на серединному перпендикулярі, проведеному до відрізка AB. Іншими словами, точки M, для яких $AM = BM$, утворюють пряму, що перпендикулярна до відрізка AB і проходить через його середину (рис. 3).

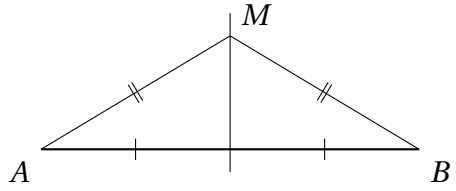


Рис. 3

Нехай тепер $m \neq n$, для визначеності, $m > n$. На прямій l, що містить відрізок AB,

є дві точки, кожна з яких ділить відрізок AB у відношенні $m : n$: перша — внутрішнім чином, друга — зовнішнім (див. рис. 1). Позначимо ці точки відповідно через K та L.

Нехай M — будь-яка інша точка площини, для якої

$$\frac{MA}{MB} = \frac{m}{n}.$$

Тоді, як ми вже знаємо, MK — бісектриса кута AMB, а ML — бісектриса кута BMP. Оскільки кут KML прямий, то точка M лежить на колі, діаметром якого є відрізок KL. Висновок: усі шукані точки неодмінно лежать на колі, діаметром якого є відрізок KL. Лишається з'ясувати, чи всі точки цього кола мають потрібну властивість.

Нехай масштаб для вимірювання відрізків обрано так, що відрізок AB має довжину $m+n$, а отже, $AK = m$ і $KB = n$. Якщо точка M має потрібну властивість, то $AM = cm$, а $MB = cn$, де c — додатне число, від якого вимагається лише, щоб точка M існувала. Оскільки ми вважаємо, що $m > n$, то це буде тоді, коли для чисел cm, cn і $m + n$ справджуються дві нерівності

$$cm + cn \geq m + n \quad \text{і} \quad cm \leq cn + m + n,$$

тобто коли $1 \leq c \leq \frac{m+n}{m-n}$.

При $c = 1$ отримуємо точку K, при $c = \frac{m+n}{m-n}$ — точку L. Решта шуканих точок лежать на перетинах кіл радіусів $r_a = cm$ і $r_b = cn$.

Кожна така точка лежить також на колі S, що має діаметр KL. Оскільки $m \leq r_a \leq m \cdot \frac{m+n}{m-n}$, $AK = m$ і $AL = m \cdot \frac{m+n}{m-n}$, то через кожну точку кола S проходить одне з кіл радіуса r_a. Тому для кожної точки M кола S справджується рівність $MA : MB = m : n$.

Підсумовуємо. Нехай A і B — дві дані точки площини, а m і n — дані додатні числа. Ми з'ясували, яким є геометричне місце тих точок M площини, відстані від яких до точок A і B відносяться так, як числа m і n:

$$MA : MB = m : n.$$

Якщо $m = n$, то таким геометричним місцем точок є серединний перпендикуляр до відрізка AB, тобто пряма лінія, щодо якої точки A і B дзеркально симетричні одна одній.

Якщо ж $m \neq n$, то шукане геометричне місце — коло S, двома діаметрально протилежними точками якого є K і L, які ділять відрізок AB внутрішнім і зовнішнім чином у відношенні $m : n$.

Коло S називають *колом Аполлонія* на честь знаменитого давньогрецького математика, що жив у другому сторіччі до нової ери. Саме йому належить перший розв'язок наведеної вище задачі.

Обертаючи коло Аполлонія навколо його діаметра KL (тобто прямої AB), отримаємо сферу, центр і радіус якої збігаються із центром і радіусом цього кола. Відстані від будь-якої точки N цієї сфери до точок A і B також відносяться як $m : n$, адже при обертанні кола S навкруг прямої AB через точку N проходить траєкторія однієї з його точок M. Тому $NA = MA$ і $NB = MB$.

Звідки два кола видно під однаковими кутами

Розглянемо таку задачу. У площині проведено два кола, що лежать зовні одне одного та не мають спільних точок (іншими словами, відстань між центрами кіл більша, ніж сума їхніх радіусів). Із яких точок площини ці кола видно під однаковими кутами? Треба знайти всі такі точки, тобто визначити геометричне місце цих точок.

Звісно, братися до розв'язання задачі можна тільки після того, як домовлено, що вважати величиною кута, під яким із даної точки видно дане коло. Годі й казати, що це величина кута між двома дотичними до кола, який містить це коло всередині (рис. 4).

Це означення придатне лише для того випадку, коли дана точка M лежить за

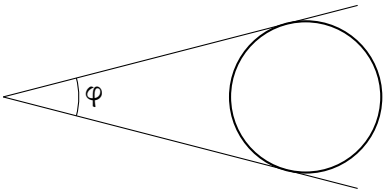


Рис. 4

межами кола, проте для інших точок M обговорюване поняття не цікаве.

Можемо переходити до розв'язання задачі. Нехай є два кола радіусів R_1 і R_2, які не мають спільних точок. Позначимо їхні центри O_1 і O_2. Якщо із точки M обидва кола видно під однаковими кутами, то $\angle AMO_1 = \angle BMO_2$ (рис. 5).

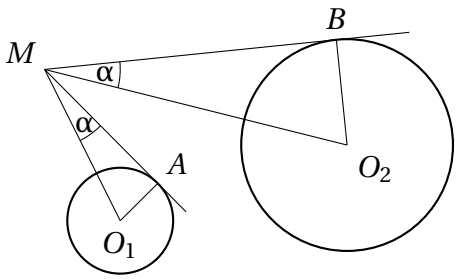


Рис. 5

Оскільки кути MAO_1 і MBO_2 прямі, то трикутники MAO_1 і MBO_2 подібні. Тому $MO_1 : MO_2 = R_1 : R_2$.

Ця пропорція засвідчує, що точка M лежить на колі Аполлонія, яке визначають точки O_1 і O_2 та числа R_1 і R_2. Навпаки, якщо M – точка цього кола, тобто виконується остання пропорція, то трикутники AMO_1 і BMO_2 подібні, а їхні кути AMO_1 і BMO_2 рівні. Виходить, що дані два кола видно під однаковими кутами з усіх точок кола Аполлонія, побудованого для точок O_1 і O_2 та відношення $R_1 : R_2$, і тільки із цих точок.

Звичайно ж, усе сказане стосується того випадку, коли радіуси кіл різні. Якщо ж вони однакові, то для шуканих точок M справджується рівність

$$MO_1 : MO_2 = 1 : 1.$$

Такі точки утворюють серединний перпендикуляр, проведений до відрізка O_1O_2.

Повернімося знову до найцікавішого випадку, коли $R_1 \neq R_2$. На прямій O_1O_2 є дві точки, з яких обидва кола видно під однаковими кутами, й ці точки легко знайти (побудувати). Одна з них — точка перетину двох спільних внутрішніх дотичних до даних кіл, друга — точка перетину спільних зовнішніх дотичних (рис. 6).

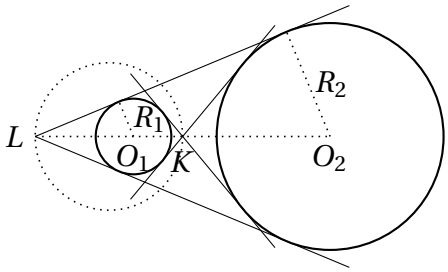


Рис. 6

Ці дві точки ділять відрізок O_1O_2 у відношенні $R_1 : R_2$ внутрішнім та зовнішнім чином. Побудоване на відрізку LK як на діаметрі коло є водночас колом Аполлонія для точок O_1 і O_2 та відношення $R_1 : R_2$, і тим колом, з будь-якої точки якого дані кола видно під однаковими кутами.

Якщо всі три кола, зображені на рис. 6, обертати навколо прямої LO_2, що проходить через їхні центри, то вони опишуть три сфери: дві із центрами O_1 і O_2 й радіусами R_1 і R_2, а третю з діаметром LK. Ця просторова конфігурація є повним аналогом плоскої конфігурації, поданої на рис. 6. Найменуємо три згадані сфери так: A — сфера із діаметром LK; Φ_1 — сфера з центром O_1 і радіусом R_1; Φ_2 — сфера із центром O_2 і радіусом R_2. Візьмімо на колі з діаметром LK (тобто коло Аполлонія для точок O_1 і O_2 та відношення $R_1 : R_2$) довільну точку M. Якщо її обертати навколо прямої LO_2, то вона опише коло σ, що лежить на сфері A у перпендикулярній до осі обертання площині. Відстань від будь-якої точки цього кола до точки O_1 така сама, як від точки M. Іншими словами, усі точки цього кола (траєкторії точки M при її обертанні навкруг LO_2) рівновіддалені від O_1. Так са-

мо всі його точки однаково віддалені від O_2. Оскільки

$$MO_1 : MO_2 = R_1 : R_2,$$

то ця рівність лишиться правильною для всіх точок кола σ. Узявши тепер до уваги, що попередні міркування стосуються довільної точки M, що лежить на колі з діаметром LK, доходимо висновку, що рівність

$$TO_1 : TO_2 = R_1 : R_2$$

справджується для всіх точок T сфери A. Її, цю сферу, можна назвати сферою Аполлонія для точок O_1 і O_2 та відношення $R_1 : R_2$.

Коли ми для двох даних точок O_1 і O_2 і даного відношення $R_1 : R_2$ ставимо запитання, яким є геометричне місце тих точок, для яких відстані до O_1 і O_2 відносяться так, як R_1 до R_2, то відповідь на нього залежить від того, де ми шукаємо такі точки. Якщо у фіксованій площині, що містить точки O_1 і O_2, то матимемо коло Аполлонія. Якщо ж у всьому просторі, то отримаємо сферу Аполлонія. Ця сфера є результатом обертання кола Аполлонія навкруг прямої O_1O_2. Інакше кажучи, вона зіткана з кіл Аполлонія для всіх площин, які містять пряму O_1O_2.

Як ми з'ясували, коло Аполлонія щонайтісніше пов'язане з відповіддю на таке запитання: з яких точок площини два ізольовані одне від одного кола, що лежать у цій площині, видно під однаковими кутами?

Чи зберігається цей зв'язок, коли від площини переходимо до простору, замінивши кола сферами? Відповідь позитивна. Слід лише умовитися, що маємо на увазі, коли кажемо, що з даної точки дану сферу видно під тим чи іншим кутом.

Отож, нехай є сфера Ф і точка M, що лежить поза нею (рис. 7).

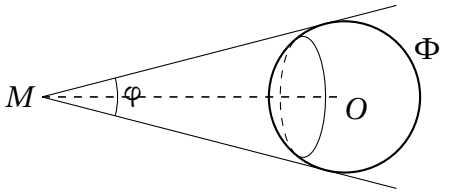


Рис. 7

Провівши з точки M усі можливі дотичні до сфери Ф, отримаємо конічну поверхню з вершиною M. Кулю Ф вписано в цю поверхню. Просторовий кут при вершині конічної поверхні і є тим кутом, під яким сферу Ф видно з точки M. Ми не торкатимемося тут методів вимірювання довільних просторових кутів (таких, зокрема, як багатогранні кути при вершинах багатогранників). Натомість зазначимо, що «розхил» кругової конічної поверхні цілком визначає плоский кут φ при вершині її осьового перерізу. Величину цього кута вважатимемо також величиною просторового кута при вершині кругової конічної поверхні.

Варто зауважити, що ідея вимірювання складніших кутів пов'язаними з ними певним чином плоскими кутами застосовується не тільки у випадку кутів при вершинах кругових конічних поверхонь. Пригадаємо, наприклад, як упроваджується міра двогранного кута. Величиною такого кута називають величину кута його плоского перерізу, перпендикулярного до його ребра.

Беремося тепер до задачі про дві сфери. Отже, дано дві ізольовані одна від одної сфери. Позначимо їхні центри O_1 і O_2, а радіуси R_1 і R_2. Слід знайти всі ті точки простору, з яких обидві дані сфери видно під однаковими кутами. Щоб уникати непорозумінь, підкреслимо, що ми не зважатимемо на те, чи затуляє одна зі сфер частину іншої, чи ні, коли споглядати їх з якоїсь точки. Вимірюючи кут, під яким видно одну сферу, ми повністю ігноруємо другу. Це зауваження стосується також двох кіл, про що йшлося раніше.

Нехай із якоїсь точки M простору кожну з двох наших сфер видно під кутом γ. Проведімо площину П, яку визначають точки M, O_1 і O_2 (якщо M належить прямій O_1O_2, то проводимо будь-яку площину,

яка містить пряму O_1O_2). Площина П перетинає наші сфери по великих колах (радіусів R_1 і R_2). Із точки M обидва ці кола видно під кутом γ. Тому точка M належить колу Аполлонія для точок O_1 і O_2 та відношення $R_1 : R_2$ (якщо $R_1 = R_2$, то замість кола Аполлонія матимемо пряму — серединний перпендикуляр до відрізка O_1O_2). Із цього випливає таке:

а) точка M лежить на сфері Аполлонія для точок O_1 і O_2 та відношення $R_1 : R_2$, якщо $R_1 \neq R_2$, або ж

б) вона лежить у площині, щодо якої точки O_1 і O_2 дзеркально симетричні.

Повторивши міркування попереднього абзацу у зворотному порядку, впевнимося, що кожна точка сфери Аполлонія (якщо $R_1 \neq R_2$) або ж кожна точка згаданої вище площини (якщо $R_1 = R_2$) має потрібну властивість.

Висновки

Якщо $R_1 = R_2$, то точки, з яких обидві сфери видно під однаковими кутами, складають «серединну» площину, тобто площину, яка перпендикулярна до відрізка O_1O_2 і проходить через його середину.

Якщо ж $R_1 \neq R_2$, то ті точки, з яких дані сфери видно під однаковими кутами, утворюють сферу. Ця сфера не що інше, як сфера Аполлонія для точок O_1 і O_2 та відношення $R_1 : R_2$. Як ми показали вище, радіус цієї сфери становить

$$\frac{R_1 \cdot R_2 \cdot d}{\left|R_1^2 - R_2^2\right|}.$$

Літерою d тут позначено відстань між центрами куль O_1 і O_2.

Якщо Землю і Місяць вважати кулями, що мають радіуси R_1 і R_2, а відстань d між їхніми центрами незмінною, то останній вираз означатиме радіус сфери Ω, із кожної точки якої Місяць і Землю видно під однаковими кутами. Центр цієї сфери лежить на лінії центрів Землі і Місяця поза відрізком O_1O_2 на відстані

$$\frac{d \cdot R_2^2}{R_1^2 - R_2^2}.$$

від центра Місяця (R_2 — радіус Місяця). Гігантська уявна сфера Ω, звичайно ж, переміщується у просторі, супроводжуючи два реальні космічні тіла: Землю та Місяць.

За бажанням читач легко визначить числове значення її радіуса, скориставшись середніми значеннями радіусів Землі й Місяця та відстані між їхніми центрами: $R_З = 6400$ км, $R_М = 1700$ км, $d = 384000$ км.

Парадокси статистики

Майборода Ростислав Євгенович,
Київський національний університет імені Тараса Шевченка

У статистиці ми часто маємо справу з великими наборами даних, які потрібно описати одним-двома числами. Найпростіший спосіб такого узагальненого опису — усереднення.

Наприклад, потрібно порівняти ефективність двох методів лікування однієї хвороби. Для цього можна провести експеримент: взяти дві групи пацієнтів, скажімо, по 10 осіб у кожній, і пацієнтів з першої групи лікувати першим методом, а з другої — другим. Якщо для кожного пацієнта записати час від початку лікування до моменту одужання, отримаємо 20 чисел. Зрозуміло, що чим менші ці числа, тим ефективнішим було лікування. Якщо, наприклад, усі тривалості лікування з першого набору менші, ніж будь-яка тривалість із другого, то перший метод однозначно ефективніший ніж другий. Але таке трапляється досить рідко: лікування пацієнта інколи може затягнутись і за ефективного підходу. А хтось може швидко одужати сам майже без лікування.

Тому медичні статистики рахують середні тривалості для кожному набору та порівнюють їх. Якщо середня тривалість лікування для першого методу на 30 % менша, ніж для другого, це вже помітний ефект, навіть коли не всі пацієнти першої групи одужали швидше.

Коли хочуть продемонструвати неадекватність такого підходу, згадують «середню температуру по палаті»: у п'яти хворих температура тіла підвищена — 38 °C, у п'яти знижена — 35 °C. Середня для всіх температура — 36,5 °C, нормальна. Всіх виписуємо!

Але й повна відмова від усереднення та виключно індивідуальний підхід до кожного окремого випадку теж, вочевидь, неможливі. Якби ми щоразу шукали індивідуальні ліки для кожного окремого хворого, ігноруючи дані про аналогічні випадки, ми ніколи не досягли б такого ефективного лікування багатьох хвороб, яке маємо нині.

Здоровий глузд вимагає спочатку ділити об'єкти спостереження по однорідних групах, а вже потім описувати такі групи узагальненими характеристиками і робити висновки. Це розумний підхід, але і він не без вад. Далі у статті я трохи розповім про деякі камені спотикання, пов'язані з такою технікою обробки й аналізу даних. Коли в них тицьнути пальцем, вони виглядають цілком очевидними. Однак, не здогадуючись про їхнє існування, обов'язково перечепишся.

Такі прості, але часом неочікувані ефекти статистичної обробки, що легко можуть привести до хибних висновків, називають парадоксами статистики. Насправді якоїсь логічної парадоксальності у них немає, це лише результат розбіжності між статистичним підходом до аналізу і звичайним здоровим глуздом.

Ефект Вілла Роджерса

У 30-ті роки XX століття США переживали Велику депресію. Це змушувало жителів одних штатів переселятись у інші. Зокрема, мешканці Оклахоми (їх назива-

ли *окі*) мігрували до Каліфорнії. Місцеві не дуже полюбляли приїжджих, це викликало непорозуміння і було приводом багатьох жартів. Один із них приписують комічному акторові Віллу Роджерсу, втім, цей жарт приписують і Дж. Стейнбеку та іншим відомим американцям тих часів: «Коли окі переїхали з Оклахоми до Каліфорнії, вони підвищили середній рівень інтелекту в обох штатах».

На перший погляд це виглядає нісенітницею. Коли геній залишає Оклахому, то середній рівень інтелекту там має знизитись, а у Каліфорнії, куди він переїхав, — зрости. З дурнем все навпаки. А з людиною середніх здібностей?

Давайте порахуємо. Нехай у Каліфорнії живуть n осіб з рівнями інтелекту $x_1, x_2, \ldots, x_n$. Їхній середній інтелект[1]

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

До них переїхав новий, $(n+1)$-й житель з інтелектом x_{n+1}. Середній рівень інтелекту став:

$$\bar{x}^{+} = \frac{1}{n+1}\sum_{i=1}^{n+1} x_i = \frac{1}{n+1}(n\bar{x} + x_{n+1}) = \\ = \bar{x} + \frac{1}{n+1}(x_{n+1} - \bar{x}).$$

Другий доданок у останньому виразі — це добавка, на яку змінюється середнє з появою нового жителя. Вона додатна, коли $x_{n+1} > \bar{x}$, і від'ємна, коли $x_{n+1} < \bar{x}$. Можна сказати, що кожен новий об'єкт (житель) притягує до себе середнє значення: воно зміщується в той бік, де цей новий об'єкт з'явився.

При вибутті n-го об'єкта з набору даних середнє «відштовхується» від його значення:

$$\bar{x}^{-} = \frac{1}{n-1}\sum_{i=1}^{n-1} x_i = \frac{1}{n-1}(n\bar{x} - x_n) = \\ = \bar{x} + \frac{1}{n-1}(\bar{x} - x_n).$$

Таким чином, якщо середній інтелект у Каліфорнії був $\bar{x}_K$, а в Оклахомі — $\bar{x}_O$, то при виїзді з Оклахоми жителя з інтелектом x середнє $\bar{x}_O$ збільшиться, якщо $x < \bar{x}_O$, а з приїздом його у Каліфорнію збільшиться, якщо $x > \bar{x}_K$.

Остаточно, для виконання твердження Вілла Роджерса досить, щоб

$$\bar{x}_K < x < \bar{x}_O,$$

тобто щоб середній інтелект каліфорнійців був нижчим, ніж в оклахомців, а інтелект мігрантів містився між цими середніми.

Звичайно, це лише жарт, і, можливо, не дуже вдалий. Але, якщо не враховувати можливості такого ефекту, легко прийти до хибних висновків. Наприклад, розглянемо знову медичних статистиків, що збирають інформацію про тривалість лікування пацієнтів у госпіталі. Щоб не міряти середню температуру по палаті, пацієнтів розділили на дві групи — легко- і важкохворих. Легкохворими вважають тих, хто перебував у госпіталі на лікуванні менше п'яти діб, важкохворими — решту. По кожній групі реєструється середня тривалість лікування. За цими даними оцінюється ефективність роботи госпіталю.[2]

Аж раптом приймається рішення вважати легкохворими тих, хто лікувався лише три доби або менше, а решту — важкохворими. Внаслідок цього група легкохворих втрачає тих, хто провів у госпіталі рівно 4 дні. Тривалість їхнього лікування була найбільшою у групі, тому середнє по групі зменшиться.

[1] У статистиці середні величини прийнято позначати рискою вгорі.

[2] Так страхові компанії на основі статистичних даних можуть піддати сумніву необхідність занадто довгого лікування та відмовитись покривати госпіталю відповідні витрати.

Табл. 1. *Дані про вступ до UCB*

	Чоловіки			Жінки		
Факультет	Прийнято	Відхилено	%	Прийнято	Відхилено	%
A	512	313	62	89	19	**82**
B	353	207	63	17	8	**68**
C	120	205	**37**	202	391	34
D	138	279	33	131	244	**35**
E	53	138	**28**	94	299	24
F	22	351	6	24	317	**7**
Усього	1198	1493	**45**	557	1278	30

Ці пацієнти поповнюють групу важкохворих. Але тут вони — найлегші. Тому середнє по важкохворим теж зменшиться. Менеджмент госпіталю може відрапортувати про помітне підвищення ефективності роботи: тривалість лікування зменшилася в обох групах: і у важко-, і у легкохворих. Хоча насправді ефект чисто уявний: жоден пацієнт у реальності не вилікувався швидше. Ми лише переклали картки з одного стосу до іншого.

Це звуть *ефектом перегрупування*. У цій ситуації осмисленою характеристикою даних є саме загальна середня тривалість лікування. З неї може виходити менеджер госпіталю, прогнозуючи завантаження та витрати, плануючи кількість пацієнтів, яких можна обслужити, тощо. Окремі середні по групах принципового значення не мають.

Парадокс Симпсона

Але так буває не завжди.

У 1973 році помітного розголосу у США набула історія з дискримінацією за статтю під час вступу до магістратури університету Берклі[3]. Тоді на 6 основних факультетів цього університету надійшло 4526 заяв, зокрема, 2691 від чоловіків і 1835 — від жінок. З тих чоловіків, що подали заяви, прийнято було 45 % (1198) а з жінок — лише 30 % (557) жінок[4].

Така велика різниця між відсотками прийнятих чоловіків і жінок викликала підозру, що приймальні комісії недостатньо об'єктивно проводили відбір, надаючи необґрунтованих переваг представникам однієї статі. Дивлячись лише на ці дані, важко не погодитися з такою підозрою.

Однак розглянемо дані окремо по кожному з шести факультетів (табл. 1).

Кожному факультету відповідає один рядок таблиці. Окремо для чоловіків та жінок вказано кількість прийнятих і відхилених заяв і відсоток прийнятих по відношенню до кількості всіх заяв, поданих на цей факультет особами даної статі. Ми бачимо, що для чотирьох факультетів із шести (*A*, *B*, *D*, *F*) відсоток прийнятих серед жінок переважає такий відсоток для чоловіків. Для факультетів *C* і *E* відсотки прийнятих більші у чоловіків, але різниця між цими відсотками складає 3 і 4, а не 15, як по загальних даних.

[3] UCB — University of California, Berkeley.

[4] Це дещо спрощений опис. Насправді в UCB тоді був 101 департамент (department), і кількості студентів і викладачів на них коливалися дуже сильно. За українською номенклатурою деякі з департаментів скоріше слід назвати кафедрами, інші — факультетами. Я використовую дані лише по 6 найбільших департаментах, узяті зі стандартного набору системи `R` — об'єкт `UCBAdmissions`. Вони вміщені також у статті Simpson's paradox з англійської вікіпедії. Справжні назви департаментів/факультетів замінені умовними літерами. Сама історія розібрана у статті [1].

Табл. 2. *Як студенти вивчали шведську мову*

	Студент *A*			Студент *B*		
День	Завдань	Правильно	%	Завдань	Правильно	%
Субота	20	10	50	5	3	**60**
Неділя	5	1	20	20	5	**25**
Усього	25	11	**44**	25	8	32

За такими даними навряд чи варто підіймати питання про дискримінацію осіб якоїсь статі. Хіба що приймальну комісію факультету *A* можна запідозрити у наданні більшого сприяння жінкам.

Тут ми зустрічаємось із проявом *парадоксу Симпсона*. Здоровий глузд підказує, що якби жінки вигравали у чоловіків по відсотках на кожному факультеті, то вони виграли б і по всьому університету разом. Але це не так.

Щоб зрозуміти, чому це може статись, розглянемо простіший приклад. Нехай два студенти *A* і *B* вирішили вивчати шведську мову і поставили на своїх смартфонах відповідну програму. Ця програма задає їм завдання і перевіряє правильність виконання. Щоб було цікавіше, студенти змагаються один з одним — хто набере більший відсоток правильних відповідей.

Почали у суботу — студент *A* виконав 20 завдань, з них 10 правильно. Студент *B* зробив лише 5 завдань, правильно — 3. У неділю студент *B* вирішив надолужити і зробив 20 завдань. Та правильними виявились тільки 5. Студент *A* полінувався — зробив 5 завдань, з них вгадав тільки в одному. Результати маємо у табл. 2.

Бачимо, що і в суботу і у неділю переміг студент *B*. Але по сумі за обидва дні — студент *A*. Чому? Тому, що кількості виконаних завдань (знаменники тих пропорцій, за якими ми визначаємо перемогу) були дуже різними у різні дні та для різних студентів. Суботня перемога *B* була «нечесна» — він зробив правильно значно менше завдань, ніж *A*. У неділю *B* переміг «чесно», якщо порівнювати з результатами *A* теж у неділю. Але якщо подивитись на суботні результати *A*, коли той виконав стільки ж завдань, скільки *B* у неділю, то можна помітити, що тут *A* встиг набути собі хорошої фори.

У цій задачі розбиття результатів по днях тижня не виглядає принциповим. Зрозуміло, що краще впорався з роботою той, хто правильно виконав більшу частину всіх (і суботніх, і недільних) завдань. Думаю, читач погодиться, що по суті переміг *A*[5].

У попередньому прикладі з можливою дискримінацією за статтю, навпаки, розбиття даних по факультетах має принциповий характер. Кожен факультет має свою приймальну комісію і своїх претендентів на вступ. Ті, хто хоче вивчати юриспруденцію, не конкурують із тими, хто подав заявку на геологію. Звичайно, можна обговорювати питання про те, чому 1973 року факультет *A* користувався більшим попитом у чоловіків, а факультет *C* — у жінок. Але це інше явище, не пов'язане безпосередньо з дискримінаційним підходом приймальних комісій до своєї роботи.

Таким чином, вибір, групувати чи не групувати, статистик має робити з урахуванням не математичних міркувань, з логіки прикладної проблеми, яку він розв'язує.

[5]Звичайно, якщо друзі наперед домовились про формальні умови свого змагання, за якими перемагає той, хто переміг у більшості днів, то треба погодитись із перемогою *B*. Це дещо нагадує історію з американськими виборами, де Х. Клінтон набрала на 2 868 686 більше голосів виборців, а Д. Трамп переміг за результатами голосувань у більшості штатів. Проте виборча система в США значно складніша за приклади, які ми тут розбираємо.

Регресія до середнього

1877 року Френсіс Гальтон[6] зробив доповідь, присвячену статистичному дослідженню законів спадковості у живих істот. Вона була опублікована у [3]. Для дослідів Гальтон використовував духмяний горошок. Він вимірював вагу горошини (x), а потім вирощував із неї рослину, зважував горошини-нащадки та рахував середню вагу всіх нащадків даної горошини-предка (y). При цьому виявилась така закономірність: якщо x був великим, то y теж, але, як правило, менше, ніж x і ближче до середньої ваги всіх горошин даного виду. Якщо ж x був малим, то y, як правило, був більшим за x. Тобто середня вага нащадків не в точності успадковувалась від предка, а зміщувалась у бік загального середнього.

Схожі дослідження Гальтон проводив також, вивчаючи успадкування зросту людей та інших характеристик живих об'єктів. І всюди були помітні аналогічні явища — зміщення значення характеристики у ході успадкування до середнього по всій популяції.

Гальтон назвав цей ефект «поверненням до середини» — «regression toward the mean» і намагався з'ясувати механізм передачі спадковості, який його забезпечує. Але вже 1885 року він писав: «Я був тоді сліпим і не бачив того, що приймаю тепер як просте пояснення цього явища».

У чому полягає це просте пояснення? Припустимо спочатку, що між вагою горошини-предка і горошини-нащадка взагалі немає зв'язку — вони коливаються навколо середнього значення ваги незалежно одна від одної. Тоді, якщо вибрати горошину-предка навмання, для нащадка однаково можливою буде і більша вага, ніж у предка, і менша. Чим більшою є вага предка, тим менше шансів у нащадка перевершити її. Скажімо, у найважчої можливої горошини всі предки обов'язково будуть легшими, ніж вона. Аналогічно і з легкими горошинами — їхні нащадки в середньому важчі за свого предка не тому, що до цього їх спонукає спеціальний механізм, а тому, що простіше бути важчим за легкого, ніж легшим.

Зрозуміло, що у реальному житті вага нащадка залежить від ваги предка. У важких горошин частіше народжуються важкі нащадки, ніж у легких. Саме це ми і називаємо успадкуванням ознак. Воно, як ми тепер знаємо, забезпечується складними механізмами молекулярної генетики, які практично неможливо зрозуміти, просто вимірюючи вагу горошин[7]. Але крім спадкової інформації на вагу горошини впливає купа інших випадкових причин: освітлення, температура під час вирощування, наявність поживних речовин у ґрунті тощо. Чим більше вплив цих випадкових факторів, що діють незалежно на предка і нащадка, тим сильніше буде виражений ефект «регресії до середнього». Цей ефект не пов'язаний із біологічними механізмами спадковості, а пояснюється взаємодією випадковостей, яка досить просто описується теорією ймовірностей.

Гальтон багато зробив для розвитку застосування статистики в біології. В результаті велика група методів виявлення реальних статистичних залежностей між різними змінними отримала назву «регресійний аналіз» — наче на честь цього помилкового висновку Гальтона, який він сам потім викрив. А термін «регресія до середнього»

[6] Англійський філософ, біолог, статистик. Відомий, зокрема, «системою Гальтона» для класифікації відбитків пальців, яка широко застосовувалась у криміналістиці.

[7] Цікаво, що піонерська робота Менделя, яка започаткувала сучасну генетику, була теж виконана на духмяному горошку. Мендель опублікував її у 1866 році (за десять років до Гальтона!), але вона пройшла практично непоміченою і була перевідкрита лише на початку XX століття. Нагадаю, що Мендель працював не з неперервними характеристиками, подібними до ваги, а з бінарними, на зразок білого/фіолетового кольору квіток, гладенької/зморшкуватої поверхні горошин тощо.

залишився як назва самого статистичного ефекту, що легко може ввести в оману, якщо не звертати увагу на його існування.

Гарний приклад наводить психолог Даніел Канеман у своїй книжці [5]. Працюючи у школі військових пілотів, він пояснював інструкторам, що у процесі навчання позитивне підкріплення працює краще ніж негативне, тому учнів треба не стільки сварити за недоліки, скільки хвалити за успіхи. Практично всі інструктори відповідали, що ця загальноприйнята психологічна теорія може бути правильною для пацюків і голубів, але для курсантів вона неправильна.

Дійсно, казали інструктори, ми неодноразово помічали, що після похвали курсант, зазвичай, розслаблюється і робить вправу не так вдало, як раніше. А от після догани людина, як правило, наступну спробу виконує краще. Канеман каже, що в цьому якраз і виявляється регресія до середнього. Справді, догани, як правило, роблять після особливо невдалого виконання завдання. Отже і перевершити такий результат не надто важко. А хвалять за особливі успіхи. Для курсанта часто такий успіх є не наслідком надзвичайної вправності, а результатом щасливого збігу обставин. Тому не дивно, що наступні спроби дають гірший результат.

Ще один приклад проблеми, пов'язаної з регресією до середнього, я візьму зі статті *regression to the mean* у Енциклопедії статистичних наук [2]. Коли міська влада вирішує, які перехрестя треба наново обладнати світлофорами, вибирають ті, на яких за статистикою фіксується найбільше дорожніх інцидентів. Після встановлення світлофорів кількість зафіксованих інцидентів на перехресті, як правило, знижується.

Це може бути наслідком ефективної роботи світлофора. А може бути ефектом регресії до середнього: адже дорожні інциденти відбуваються випадково і, якщо ми вибрали перехрестя, де їх було надзвичайно багато, малоймовірно, що у наступний період часу їх стане ще більше. Скоріше, можна сподіватись на випадкове зменшення кількості інцидентів. Хотілося б уміти оцінювати, яку саме частку ефекту зниження можна записати на користь роботи світлофора, а яку — списати на регресію до середнього. Однак, для цього потрібно спочатку побудувати ймовірнісну модель розподілу кількості інцидентів, а це вже виходить за межі нашої статті.

Мабуть, найбільш ефектний сюжет, пов'язаний із регресією до середнього, ми розглянемо в наступному розділі.

Тріумф посередності в бізнесі

1933 року американський статистик Орас Секріст опублікував монографію під назвою «Тріумф посередності у бізнесі»[8]. Секріст був професором у Північно-Східному університеті у Чикаго і головою Бюро економічних досліджень цього університету. Книга підводила підсумки десятирічної праці Секріста та його колег з вивчення роботи американських торговельних, фінансових, промислових і транспортних підприємств. У ній було 140 таблиць і 103 діаграми, що описували розвиток американської економіки у період 1920–1930 років. Результатом його праці було економічне відкриття, яке претендувало на ту ж роль, яку теорія Дарвіна відіграє у біології.

Дарвін у «Походженні видів» продемонстрував, як біологічна боротьба за існування приводить до зростання різноманітності живого світу. Секріст показував, як економічна боротьба на вільних ринках зменшує різноманітність підприємств, змушує їх орієнтуватись на посередні показники. Для цього він застосовував єдину техніку

[8]Horace Secrist «The Triumph of Mediocrity in Business». Зміст цієї книжки та всю історію з Секрістом і Хотелінгом я викладаю, переважно, за [6].

аналізу всіх досліджуваних галузей економіки.

Так, у книзі розглянуто 50 великих американських універмагів. Секріст дослідив, як змінювались протягом 1920–1930 років такі характеристики їхньої діяльності, як відношення чистого прибутку до виручки. Для того, щоб відслідкувати загальну тенденцію, Секріст розбив усі розглядувані підприємства на 4 групи залежно від того, яким був їхній прибуток на початку дослідження — у 1920 році. У першу групу ввійшли 25 % підприємств із найбільшим прибутком, у другу — наступні 25 %, …, у четверту — 25 % найменш прибуткових. Далі Секріст підрахував середній прибуток у кожній із цих груп по кожному року досліджень від 1920 до 1930 і відобразив їх у вигляді таблиць і графіків. Приклад того, як виглядав типовий такий графік, можна побачити на рис. 1 ліворуч[9].

Як бачимо, дійсно, середні по групах виявляють стабільну тенденцію до зближення. Якщо 1920 року різниця між середнім найкращої і найгіршої груп становила 5,5, то у 1930 році вона була вже приблизно 2,5.

Аналогічний ефект Секріст виявив, аналізуючи роботу продуктових магазинів, залізниць, банків та інших підприємств. Ту ж закономірність можна було побачити, вибравши для початку дослідження не тільки 1920, а й інші роки. Завжди групи, які спочатку займали крайні позиції, з плином часу виявляли тенденцію прямувати до середини. Всюди спостерігався «тріумф посередності».

Щоб продемонструвати, що виявлений ефект має саме економічний характер, Секріст розглянув набір даних про природні явища: середні температури червня у 191 американському місті протягом тих же 11 років. Діаграма, побудована за методом Секріста, не виявила сходження групових середніх до загального середнього!

Робота Секріста викликала схвальні відгуки у провідних статистичних журналах світу. Однак у журналі американської статистичної асоціації з'явилась стаття Гарольда Хотеллінга — провідного американського спеціаліста з математичної статистики. Уважно дочитавши до цього місця, ви вже знаєте, що там було написано: виявлений Секрістом «ефект» — це регресія до середнього, яка виникає не внаслідок реальних економічних закономірностей, а через специфічну обробку даних із виділенням «крайніх» (найбільших, найменших) спостережень. Якби Секріст для групування використав не дані 1920 року, а дані 1930 року (останній рік спостережень) то його діаграми виглядали б так, як на рис. 1 праворуч. За логікою секрістових пояснень це мало б свідчити про наростання різноманітності у роботі універмагів!

Але чому ж регресія до середнього не спостерігалася для рядів температури червня? Справа в тому, що ступінь вираженості цього ефекту пов'язаний із тим, наскільки залежними (корельованими) є повторні вимірювання. Зрозуміло, що коли повторні вимірювання через рік дають в точності ті ж самі значення (максимальна залежність) — лінії на діаграмі Секріста будуть горизонтальними, жодного прямування до середнього побачити буде неможливо. Якщо залежності між повторними вимірюваннями немає взагалі, значення на діаграмі впадуть одразу на другому ж році спостереження і далі коливатимуться навколо середнього цілком випадково. Ті діаграми для економічних даних, які Секріст навів у своїй книжці, відповідають проміжній ситуації: є залежність між послідовними значеннями спостережуваної характеристики, але на цю залежність накладаються випадкові коливання. При цьому з плином часу

[9] Справжні графіки Секріста можна подивитись у статті [6]. Графік на рис. 1 побудований за секрістовою методикою на основі модельованих даних.

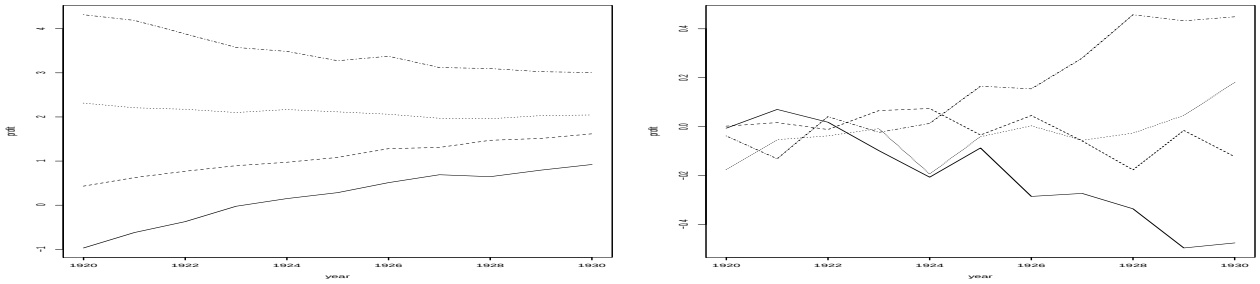


Рис. 1. *Діаграми Секріста. Ліворуч — оригінальна, праворуч — згрупована по 1930 року*

залежність від початкових значень зменшується. Саме такій, природній для економічних характеристик підприємства картині і відповідає та поведінка діаграм, яку помітив Секріст.

Ряди температур для різних міст виявились дуже корельованими. Це і не дивно. Секріст взяв міста, розташовані у географічно віддалених регіонах, тому їх середні температури червня більше відрізнялись внаслідок різних географічних умов, ніж у результаті випадкових коливань.

На відміну від Гальтона, Секріст так і не визнав помилки і до кінця обстоював правильність своїх висновків. У [6] висловлена гіпотеза, що саме завдяки впливам Секріста, Хотеллінг утратив можливість друкувати свої роботи у журналах американської статистичної асоціації і був змушений узяти участь у створенні нового журналу «Annals of Mathematical Statistics». Це і досі один із найвпливовіших журналів у галузі математичної статистики[10]. А Секріста нині згадують практично тільки у зв'язку із цією історією.

Біоритми

В одній простенькій історії з використанням групування й усереднення мені самому довелось узяти участь. Почалась вона іще у XIX столітті, коли друг та послідовник Зігмунда Фрейда отоларинголог Вільгельм Флісс (Wilhelm Fliess) висунув теорію біоритмів. Теорію Флейса потім розвивали та модифікували багато авторів. Найбільш поширену її версію можна описати так.

У кожної людини є три різні «потенціали» — фізичний, емоційний та інтелектуальний. Їхні значення коливаються періодично по синусоїдах із періодами, відповідно, 22, 28 і 33 дні. Стартують усі синусоїди з 0 у момент народження[11] і коливаються далі без змін — аж до смерті людини. Чим вищий інтелектуальний потенціал людини у даний момент, тим краще в неї йде розумова робота, й аналогічно — для інших потенціалів. Усе це разом і зветься біоритмами[12].

Прибічники теорії біоритмів вважають, що особливо небезпечними для людини є *критичні дні*, коли відповідний потенціал змінюється якнайшвидше. Оскільки швидкість зміни синуса найбільша, коли він дорівнює 0, отримуємо три послідовності критичних днів із періодами 11, 14 і 16,5 відповідно[13]. Усі три послідовності починаються у день народження. У ці дні на людину чекає небезпека, особливо, якщо вона займатиметься діяльністю, для якої день є критичним.

[10] Нині видається під назвою «Annals of Statistics».

[11] За іншою версією — у момент зачаття.

[12] «Теорію біоритмів» прийнято вважати псевдонауковою, разом із «теоріями» Бермудського трикутника, палеоконтакту, телепатії і т. п. Наука, що вивчає реальні циклічні біологічні процеси (такі як чергування сну та неспання) зветься хронобіологією.

[13] Наприклад, кожен чотирнадцятий день є емоційно критичним.

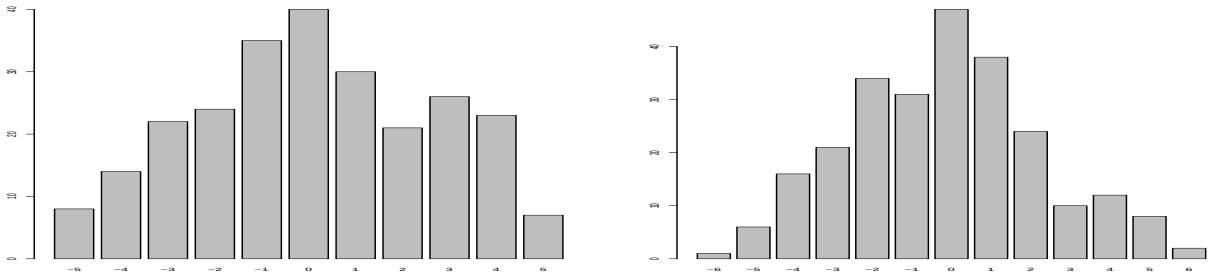


Рис. 2. *Діаграми частот для критичних днів. Ліворуч — для фізичного й емоційного циклів, праворуч — для циклів із періодами 13 і 15*

У 80-ті роки XX ст. ця теорія набула масового поширення, значною мірою, через появу персональних комп'ютерів із кольоровими дисплеями і принтерами. Пам'ятаю, що якийсь час на залізничному вокзалі у Києві стояв такий комп'ютер біоритмів, у якого недорого можна було дізнатись свої найближчі критичні дні, а за додаткову плату — отримати роздруківку власних біоритмів на кілька місяців уперед. Нині для цього можна встановити собі відповідну програму на смартфоні, погугливши її за словами «biorythm caculator».

Безпосередня експериментальна перевірка гармонійного коливання біопотенціалів неможлива, оскільки не зрозуміло, як вимірювати ці потенціали. З критичними днями легше: досить перевірити, наскільки часто певні негативні події відбуваються саме у критичні дні порівняно з відповідними частотами цих подій у дні некритичні. Наприклад, на роль негативної події можна взяти попадання у аварію на дорозі з вини водія. Чи пов'язана частота цієї події з критичними днями водія? Такі дослідження проводились неодноразово.

У коректно проведених дослідженнях результати виходили негативними. У роботі Т. Гінеса [4] проведено аналіз 134 експериментальних робіт із виявлення біоритмів. З них у 99 підтверджень цієї теорії не виявлено. У 35 роботах зроблено висновок про виявлення ефектів, пов'язаних із біоритмами, але ці роботи містять методичні та статистичні помилки.

Десь приблизно тоді ж, коли Гінес опублікував свою роботу, до мене звернувся по консультацію колега з іншого факультету. Назву його NN. Він мав цікаве пояснення того, чому статистичні пошуки ефекту критичних точок не принесли успіху. На його думку, це пов'язано з неточністю визначення дня народження людини: немовля може народитись у понеділок увечері, а запишуть його у вівторок. Або навпаки. Зрозуміло, що в результаті при підрахунку кількості подій, що припали на критичні дні, дані такої людини будуть зараховані неправильно.

Для усунення цього гіпотетичного недоліку NN розробив новий підхід. Він увів поняття *індексу дня* щодо набору критичних днів. Критичні дні отримують індекс 0. Дні, які йдуть одразу після критичних (якщо вони самі не є критичними) — індекс 1. Дні напередодні критичних — індекс −1. Наступний після дня з індексом 1 буде мати індекс 2, якщо йому не присвоєний індекс −1, і т. д. — для кожного некритичного дня ми визначаємо його відстань до найближчого критичного і беремо її зі знаком +, якщо цей найближчий — у минулому, та зі знаком −, якщо він — у майбутньому. Це і буде індекс нашого дня.

Зрозуміло, що під час вибору знаку індексу може виникнути невизначеність. Наприклад, некритичний день, що міститься між двома критичними, повинен одночасно мати індекси 1 і −1. У таких випадках будемо вибирати знак випадково, скажімо,

підкидаючи монетку: якщо випав герб, пишемо +, випала цифра — пишемо −.

NN запропонував підраховувати кількості подій, що припали на дні з певним значенням індексу. Якщо навіть день народження людини записали зі зсувом, подія, що трапилась у справжній критичний день, буде зарахована на день із невеликим індексом. Отже, якщо теорія біоритмів є правильною, слід чекати скупчення неприємних подій у дні з індексами малими за абсолютною величиною. А у дні з великими індексами таких подій повинно бути мало.

NN використав дані київських станцій швидкої допомоги про виклики до пацієнтів, що мали інфаркти або інсульти. У даних вказувалися дата народження пацієнта та дата, коли був виклик швидкої допомоги. Це дозволяє розрахувати індекс дня, у який для даного пацієнта стався інфаркт/інсульт, а потім за всіма наявними даними підрахувати, скажімо, кількість інфарктів, що стались у дні з індексом −2. Результати можна відобразити у вигляді діаграм, складених зі стовпців, як на рис. 2.

Кожен стовпець відповідає одному значенню індексу, а його висота пропорційна кількості подій, що припали на дні із цим індексом.

На жаль, у мене немає даних, що зібрав NN[14]. Тому я взяв 13-й том Великої радянської енциклопедії і виписав звідти підряд дані про дати народження і смерті для 250 осіб, для яких вони там вказані. Звичайно, це не так гарно, як дані про інфаркти, але можна припустити, що смерті теж частіше мають відбуватись у критичні дні. На рис. 2 ліворуч діаграма побудована за цими даними для випадку, коли критичними вважаються кожен 11-й і кожен 14-й дні. Це відповідає фізичному й емоційному біоритмам (інтелектуальний не враховується). Як бачимо, рисунок відповідає нашим припущенням про те, що повинно бути за виконання теорії біоритмів: найбільша кількість смертей відбувається у критичні дні. Мірою віддалення від цих небезпечних днів кількість смертей зменшується. Звичайно, є невеликі відхилення від цієї закономірності, які природно пояснити випадковими коливаннями.

NN побудував багато таких діаграм для різних наборів біоритмів і для подій різних типів. І всі вони мали той же характер, підтверджуючи теорію біоритмів. Після ознайомлення з ними мені знадобився цілий день, щоб вийти з-під впливу їхньої магії. Аж тоді я зрозумів, що нічого несподіваного у цих трикутних діаграмах немає. Такі самі картинки будуть утворюватись не тільки для «теоретичних» послідовностей із періодами 11, 14 або 16,5 днів, а і для будь-якої іншої пари періодичних послідовностей критичних днів[15], якщо у ній довший період не кратний короткому.

Так, на рис. 2 зображена діаграма побудована за тими ж даними, що і ліворуч, але для послідовностей із періодами 13 і 15. Чому це так?

Пояснення просте: для таких наборів критичних точок на будь-якому достатньо довгому часовому інтервалі критичних точок буде більше, ніж точок із будь-яким іншим індексом. І чим більше абсолютне значення індексу, тим менше буде зустрічатись відповідних точок. Дійсно, якщо на часовій осі є точка з додатним індексом i, то ліворуч від неї має бути точка з індексом $i - 1$ — це безпосередньо випливає з нашого означення індексу. А от праворуч може бути точка або з індексом $i + 1$, або з індексом $-i$. Тому $i + 1$ у послідовності індексів повинно зустрічатись рідше, ніж i.

Відповідно, якщо події, які ми досліджуємо, ніяк не пов'язані з біоритмами та критичними днями, а можуть відбутись у будь-який день з однаковою ймовірністю, то частіше вони будуть потрапляти на точки з

[14] А якби й були, я не став би їх друкувати без його дозволу.

[15] або трійки, четвірки…

меншим (за абсолютною величиною) індексом — тому, що цих точок просто більше.

Єдиний випадок, коли кількість точок із різними індексами приблизно однакова — коли розглядається одна періодична послідовність критичних точок із фіксованим періодом. Тому, якщо вже пробувати застосовувати таку техніку до виявлення справжніх критичних днів, то починати треба з одного біоритму. За відсутності залежності подій від критичних днів стовпці на відповідній діаграмі будуть мати приблизно однакову висоту. А наявність спадання кількості подій при зростанні абсолютної величини індексу свідчила б про наявність реальної залежності між подіями і критичними днями.

Коли я став пояснювати це NN і дійшов до цього місця, він сказав: «Ні, з однією послідовністю там нічого хорошого не виходить». Більше ми з ним не спілкувалися.

Висновки

Можливо, читач, дійшовши до цього місця, вирішить, що кваліфікований статистик здатний будь-які дані подати так, щоб спантеличити слухачів і нав'язати їм те пояснення, яке йому потрібно. Якщо так — моя розповідь не досягла мети.

Мені хотілось показати які простенькі ефекти можуть збити з пантелику людину, що береться за статистичну обробку даних без належної підготовки. А завдання кваліфікованого статистика — допомагати неспеціалістам уникати таких пасток. Та інших, значно більш хитрих і неприємних несподіванок.

Література

[1] Bickel P. J. Sex bias in graduate admissions: Data From Berkeley / P. J. Bickel, E. A. Hammel, J. W. O'Connell // Science. — 1975. — Vol. 187 (4175). — P. 398–404.

[2] Kotz S. Encyclopedia of Statistical Sciences. Vol. 11 / S. Kotz, N. L. Johnson, C. B. Read. — Wiley, 2006. — 620 p.

[3] Galton F. Typical laws of heredity / F. Galton // Nature. — 1877. — Vol. 15. —P. 492–495, 512–514, 532–533.

[4] Hines T. Comprehensive review of biorhythm theory / T. Hines // Psychological Reports. — 1998. — Vol. 83 (1). — P. 19–64.

[5] Kahneman D. Thinking, Fast and Slow / D. Kanneman — Farrar, Straus and Giroux, 2011. — 499 p.

[6] Stigler S. M. The history of statistics in 1933 / S. M. Stigler // Statistical Science. — 1996. — Vol. 11 (3). — P. 244–252.

Уявні числа та нове життя евклідової геометрії

Гладун Сергій Євгенійович,
магістр прикладної математики

Відомо, що комплексні числа допомагають розв'язувати складні геометричні задачі. Яскравим прикладом застосування комплексних чисел у планіметрії служить розв'язок німецьким математиком Карлом Фрідріхом Гауссом у 1796 році стародавньої проблеми побудови за допомогою циркуля та лінійки правильних багатокутників. Сформулюємо і доведемо, використовуючи апарат уявних чисел, ряд нових теорем елементарної геометрії. Базою для досліджень автору послугували дивовижна теорема французького геометра Віктора Тебо середини минулого сторіччя [1, с. 55, задача 262], а також відома теорема про прямі Гаусса та Обера [2, с. 68]. У статті сформульовано цікаві властивості точок Фейєрбаха, Монжа та Мікеля у вигляді теорем, а також наведено їхні короткі доведення, що обґрунтовує застосування методу уявних чисел у даних конкретних випадках. Наостанок наведено нове доведення древньої японської теореми.

Теорема 1. *Нехай у трикутнику ABC проведено висоти AA_1, BB_1 і CC_1. Тоді прямі Ейлера e_1, e_2, e_3 трикутників AB_1C_1, A_1BC_1 та A_1B_1C перетинаються в ортополюсі T прямої Ейлера e $\triangle ABC$. Нехай далі e_1, e_2, e_3 повернуто на кут θ навколо відповідних точок Ейлера $\triangle ABC$, а пряма Ейлера самого $\triangle ABC$ на кут $-\theta$ навколо центра описаного кола. Тоді образи p_1, p_2, p_3 прямих e_1, e_2, e_3 перетинаються в ортополюсі T' прямої p (образі прямої e) відносно $\triangle ABC$.*

Доведення. Доведемо другу частину теореми, з якої випливає перша (при $\theta = 0$). Припустимо, що коло, описане навколо $\triangle ABC$ є одиничним із центром у точці O з $z = 0$ комплексної площини (рис. 1). Нехай афікси точок A, B та C дорівнюють a, b і c відповідно. Комплексна координата ортоцентра H $\triangle ABC$ за теоремою Гамільтона дорівнює $a + b + c$. Уведемо симетричні функції від трьох комплексних чисел a, b та c:

$$\sigma_1 := a+b+c, \quad \sigma_2 := ab+bc+ac, \quad \sigma_3 := abc.$$

Комплексний кутовий коефіцієнт прямої Ейлера e запишеться у вигляді

$$\kappa = \frac{\sigma_1\sigma_3}{\sigma_2},$$

а прямої p

$$\kappa' = \frac{\sigma_1\sigma_3}{\sigma_2}e^{-2i\theta}. \qquad (1)$$

Точка перетину медіан $\triangle A_1B_1C$ має комплексну координату

$$m_3 = \frac{1}{3}\left\{a + b + 2c - \frac{1}{2}\left(\frac{bc}{a} + \frac{ac}{b}\right)\right\},$$

а центр описаного кола $\triangle A_1B_1C$ — афікс c_2, який дорівнює

$$c_2 = \frac{1}{2}(a + b + 2c). \qquad (2)$$

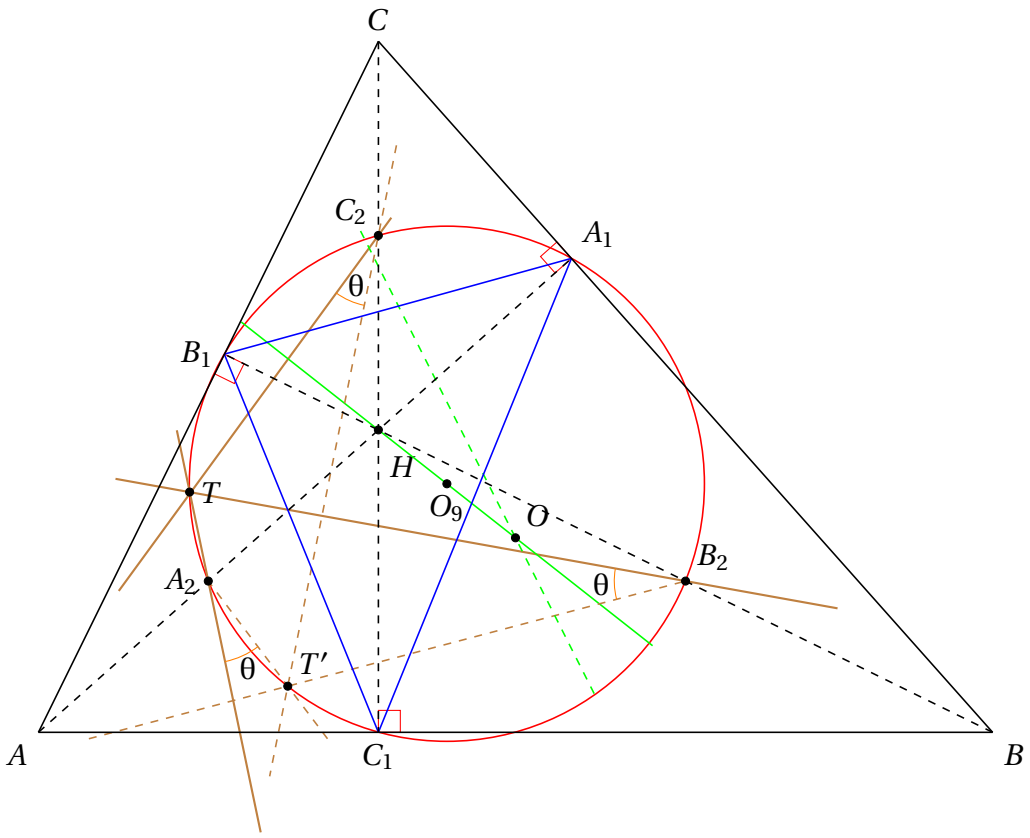


Рис. 1. *Загальна теорема про ортополюс*

Тому комплексний кутовий коефіцієнт прямої Ейлера e_3 запишеться у вигляді

$$\lambda = \frac{c_2 - m_3}{\overline{c_2} - \overline{m_3}} = \frac{\sigma_2}{\sigma_1} c,$$

а комплексний кутовий коефіцієнт прямої p_3

$$\lambda' = \lambda e^{2i\theta}. \qquad (3)$$

Афікс ортополюса T' прямої p відносно $\triangle ABC$ з урахуванням (1) набуває значення [3, с. 85]:

$$\omega' = \frac{1}{2}\left(\sigma_1 - \frac{\sigma_2}{\sigma_1} e^{2i\theta}\right). \qquad (4)$$

Для доведення загального твердження теореми достатньо перевірити справедливість рівності

$$\frac{\omega' - c_2}{\overline{\omega'} - \overline{c_2}} = \lambda'. \qquad (5)$$

Рівність (5) випливає з (2), (3) та (4). Доведення завершено. □

Із загального твердження попередньої теореми можна отримати цікаві властивості точок Фейєрбаха. Теорема Фейєрбаха є однією з дивовижних теорем геометрії трикутника й обов'язково повинна бути представлена у курсі шкільної математики.

Тут приділена увага одному з найзагальніших методів доведення геометричних фактів, котрий іноді помилково вважають нераціональним і громіздким. Насправді, **метод комплексних чисел** буває просто необхідний і дозволяє отримати набагато більше інформації про геометричні об'єкти, що вивчаються, ніж інші методи порівняно коротким шляхом. Приведені у цій статті теореми обґрунтовують цю точку зору. Автор вважає, що метод комплексних чисел (у широкому сенсі) при доведенні планіметричних і фундаментальних алгебраїчних теорем (наприклад, основної теореми алгебри за допомогою теореми Коші – Ліувілля) також необхідно включити у про-

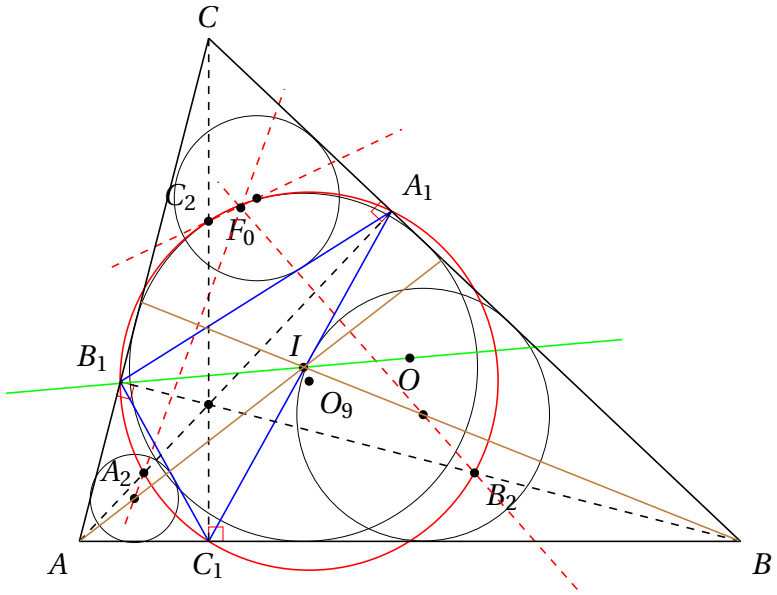


Рис. 2. *Властивості точок Фейєрбаха*

граму середньої школи. Доведення знаменитої теореми Руффіні – Абеля про те, що алгебраїчне рівняння вище четвертого степеня не розв'язується в радикалах у загальному випадку, можна подати схематично на заняттях математичного гуртка, де вказати основні ідеї, що привели до бурхливого розвитку всієї математики та суміжних наук. Для цього необхідні знання основ теорії функцій комплексної змінної та теорії груп. І, взагалі, теорію функцій комплексної змінної непогано викладати разом з ілюстрацією застосування методу комплексних чисел у планіметрії.

Теорема 2. *Нехай у трикутнику ABC проведено висоти AA_1, BB_1 і CC_1. Тоді прямі, які сполучають центри вписаних та описаних кіл трикутників AB_1C_1, A_1BC_1 і A_1B_1C перетинаються у точці Фейєрбаха F_0 — точці дотику кола дев'яти точок Ейлера та вписаного кола $\triangle ABC$ (рис. 2).*

Доведення. Твердження теореми випливає з теореми 1 та [3, с. 206, приклад 40]. Аналогічні твердження справедливі для трьох інших точок Фейєрбаха $\triangle ABC$. □

Теорема 3. *Нехай s — пряма Сімсона трикутника ABC та точки M його описаного кола, а W — ортополюс s відносно $\triangle ABC$. Тоді пряма HW паралельна прямій s, де H — ортоцентр $\triangle ABC$ (рис. 3).*

Доведення. Нехай коло, описане навколо $\triangle ABC$, є одиничним із центром у точці O з $z = 0$ комплексної площини. Нехай афікси точок A, B та C дорівнюють a, b і c відповідно. Симетричні функції трьох комплексних чисел a, b та c були введені при доведенні теореми 1.

Афікс ортополюса прямої Сімсона для точки $M(z_0)$ у цих позначеннях запишеться у вигляді [3, с. 85]

$$w = \frac{3\sigma_1 {z_0}^2 - {z_0}^3 - \sigma_2 z_0 - \sigma_3}{4{z_0}^2}$$

або

$$w = \sigma_1 - \frac{(z_0 + a)(z_0 + b)(z_0 + c)}{4{z_0}^2}. \qquad (6)$$

Формулу (6) можно переписати так:

$$WH = \frac{1}{4R^2} MA_1 \cdot MB_1 \cdot MC_1, \qquad (7)$$

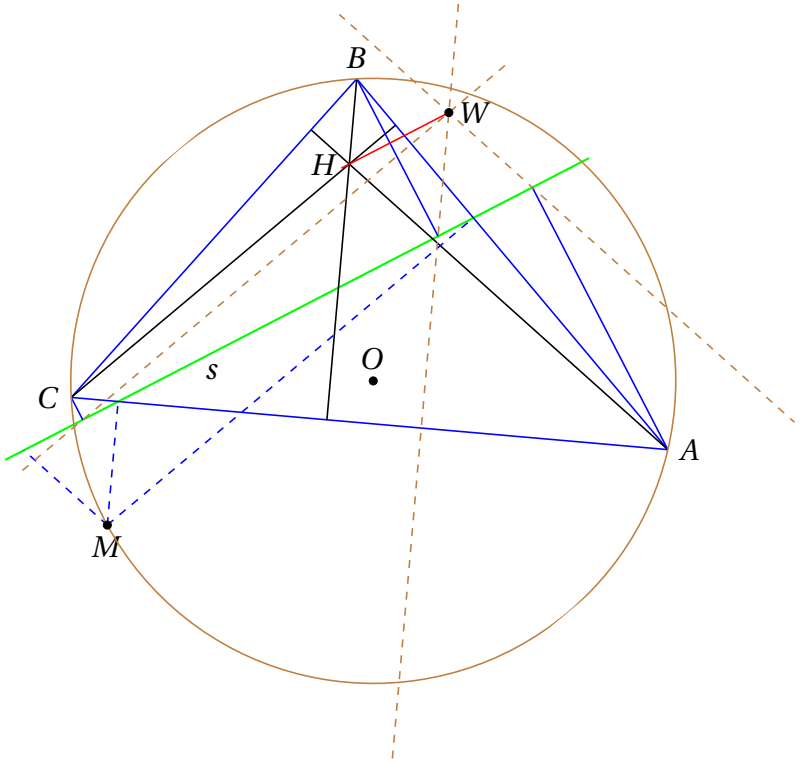


Рис. 3. *Ортополюс прямої Сімсона s точки M*

де R, у загальному випадку, є радіусом описаного кола $\triangle ABC$, а A_1, B_1 и C_1 — точки, діаметрально протилежні вершинам $\triangle ABC$.

З урахуванням (6) ми можемо отримати, що

$$\frac{w-\sigma_1}{\overline{w}-\overline{\sigma_1}}=\frac{\sigma_3}{z_0}. \tag{8}$$

Рівність (8) доводить твердження теореми [3, с. 77]. □

Зауваження 1. Ортополюс W належить прямій Ейлера $\triangle ABC$, коли

$$\frac{h}{\bar{h}}=\frac{\sigma_3}{z_0},$$

звідки

$$z_0=\frac{\sigma_2}{\sigma_1}. \tag{9}$$

Зауваження 2. Нехай $\angle B_1A_1C_1=\alpha$, $\angle B_1C_1A_1=\gamma$, і точка M належить дузі A_1B_1, яка є доповненням до дуги $A_1C_1B_1$, а $\angle MA_1B_1=x$, де $x\in(0,\gamma)$. Тоді WH досягає максимуму, коли x є коренем рівняння

$$\operatorname{ctg}(x)+\operatorname{ctg}(\alpha+x)=\operatorname{ctg}(\gamma-x)$$

на вказаному інтервалі. Оскільки функція

$$K(x):=\operatorname{ctg}(x)+\operatorname{ctg}(\alpha+x)+\operatorname{ctg}(x-\gamma)$$

монотонно спадає на $(0,\gamma)$, а

$$\lim_{x\to 0+}K(x)=+\infty,\ \lim_{x\to\gamma-}K(x)=-\infty,$$

то, як випливає з першої теореми Больцано – Коші, вона має єдиний нуль на інтервалі $(0,\gamma)$ — точку максимуму функції (див. рівність (7))

$$f(x):=WH=2R\sin(x)\sin(\alpha+x)\sin(\gamma-x).$$

Нехай тепер точка M пробігає все описане коло $\triangle ABC$. Тоді максимуму WH відповідає комплексна координата точки M, яка задовольняє рівняння

$$\frac{z-a}{z+a}+\frac{z-b}{z+b}+\frac{z-c}{z+c}=0$$

або

$$3z^3 + \sigma_1 z^2 - \sigma_2 z - 3\sigma_3 = 0. \qquad (10)$$

Нехай G_1 — центроїд $\triangle A_1B_1C_1$, а G_0 — центроїд трикутника, вершини якого відповідають трьом кореням рівняння (10). Тоді

$$\overrightarrow{OG_1} = 3\overrightarrow{OG_0}.$$

Зауваження 3. Ортополюс прямої Сімсона збігається з ортоцентром $\triangle ABC$, якщо точка $M(z_0)$ діаметрально протилежна одній із вершин $\triangle ABC$.

Твердження 1. *Нехай сторони трикутника $A'B'C'$ дотикаються описаного кола $\triangle ABC$ у точках A, B і C. Розглянемо точку Феєрбаха F_0 $\triangle A'B'C'$, яка є точкою дотику кола дев'яти точок Ейлера та вписаного кола $\triangle A'B'C'$, а також її пряму Сімсона s для $\triangle ABC$. Тоді ортополюс W прямої s належить прямій Ейлера $\triangle ABC$. Таким чином, для точки F_0 прямі Ейлера та Сімсона $\triangle ABC$ паралельні. Для точки F_0', симетричній F_0 відносно центра описаного кола $\triangle ABC$, прямі Ейлера та Сімсона $\triangle ABC$ будуть перпендикулярними.*

Доведення. Твердження випливає з рівності (9) зауваження 1 до теореми 3 та [3, с. 185]. Друге твердження дає ще один спосіб розв'язку задачі [2, с. 191]. □

Твердження 2. *Якщо точка $M(z_0)$ збігається з точками, симетричними точкам Бутена $\triangle ABC$ відносно центра описаного кола, то пряма Сімсона для $M(z_0)$ дотикається до кола дев'яти точок $\triangle ABC$.*

Доведення. Середина відрізка MH — точка K, розташована на колі дев'яти точок із центром O_9 та на прямій Сімсона і має комплексну координату

$$k = \frac{\sigma_1 + z_0}{2}.$$

Тому комплексний кутовий коефіцієнт прямої O_9K такий:

$$\kappa_1 = \frac{o_9 - k}{\overline{o_9} - \overline{k}} = z_0^2,$$

оскільки $o_9 = \sigma_1/2$. Умова дотику прямої Сімсона та кола дев'яти точок еквівалентна

$$\frac{\sigma_3}{z_0} + \kappa_1 = \frac{\sigma_3}{z_0} + z_0^2 = 0,$$

звідки і випливає твердження. □

Ще одним цікавим геометричним об'єктом є повний чотирибічник. Теореми, пов'язані з повним чотирибічником, носять, як правило, чисто афінний характер, а доводити їх можна і метричними засобами (векторами або комплексними числами). Наведемо декілька прикладів так званих змішаних теорем.

Теорема 4. *Нехай прямі α, β, γ і δ утворюють повний чотирибічник із вершинами $\alpha \cap \beta = C$, $\gamma \cap \delta = C_1$, $\alpha \cap \gamma = B$, $\beta \cap \delta = B_1$, $\beta \cap \gamma = A$, $\alpha \cap \delta = A_1$. Якщо точки A, B, A_1 та B_1 належать одному колу, то пряма Обера повного чотирибічника проходить через точку Монжа M вписаного чотирикутника ABA_1B_1 та через точку D перетину його діагоналей, а точки D, M, C і C_1 розташовані на одному колі (рис. 4).*

Доведення. Нехай коло, яке проходить через точки A, B, A_1 та B_1 є одиничним із центром у точці $z = 0$ комплексної площини. Розглянемо $\triangle ABC$, котрий орієнтований додатно. Виразимо тангенси кутів цього трикутника через афікси a, b, a_1 та b_1 точок A, B, A_1 і B_1:

$$\operatorname{tg}\hat{A} = i\frac{b - b_1}{b + b_1},$$
$$\operatorname{tg}\hat{B} = i\frac{a_1 - a}{a + a_1},$$
$$\operatorname{tg}\hat{C} = i\frac{ab_1 - ba_1}{ba_1 + ab_1}.$$

Афікс вершини C знаходимо, використовуючи формулу XII [2, с. 181]:

$$c = \frac{\dfrac{1}{a} + \dfrac{1}{b_1} - \dfrac{1}{b} - \dfrac{1}{a_1}}{\dfrac{1}{ab_1} - \dfrac{1}{ba_1}}.$$

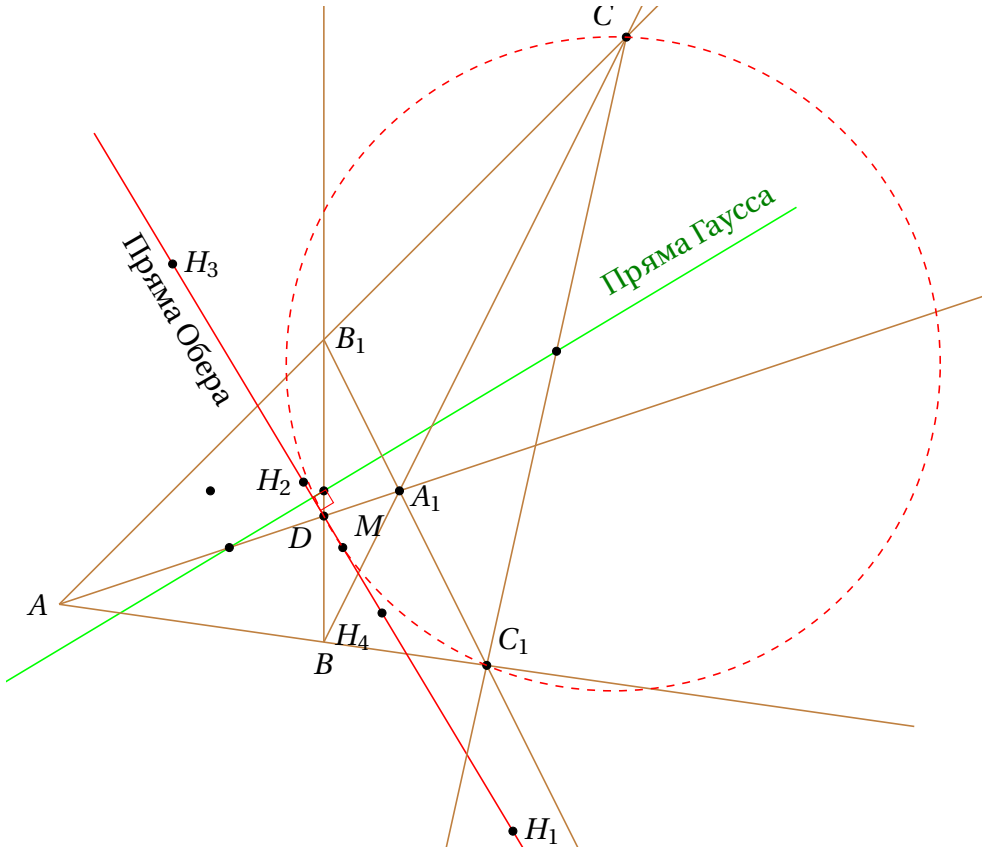


Рис. 4. *Повний чотирибічник, пряма Обера o і точка Монжа*

Далі, згідно з [2, с. 64], комплексна координата ортоцентра H_1 $\triangle ABC$ така:

$$h_1 = \frac{-aa_1b_1 + ba_1b_1 + a_1b^2 - a^2b_1}{ba_1 - ab_1}. \quad (11)$$

Згідно з [2, с. 161], афікс точки Монжа M вписаного чотирикутника ABA_1B_1

$$m = \frac{1}{2}(a + a_1 + b + b_1). \quad (12)$$

Із (11) та (12) отримуємо вираз для комплексного кутового коефіцієнта прямої H_1M:

$$\kappa = \frac{(a + a_1 - b - b_1)aa_1bb_1}{aa_1b - bb_1a_1 + b_1aa_1 - abb_1}.$$

Комплексний кутовий коефіцієнт прямої Гаусса запишемо у вигляді

$$\kappa_G = \frac{a + a_1 - b - b_1}{\frac{1}{a} + \frac{1}{a_1} - \frac{1}{b} - \frac{1}{b_1}},$$

оскільки ця пряма проходить через середини відрізків AA_1 та BB_1.

Легко перевірити, що

$$\kappa_G + \kappa = 0.$$

Ця рівність доводить перше твердження теореми, оскільки відомо, що прямі Гаусса та Обера перпендикулярні [2, с. 68].

Афікс точки D перетину діагоналей вписаного чотирикутника ABA_1B_1 знаходимо, використовуючи формулу XII [2, с. 181]:

$$d = \frac{\frac{1}{a} + \frac{1}{a_1} - \frac{1}{b} - \frac{1}{b_1}}{\frac{1}{aa_1} - \frac{1}{bb_1}}.$$

Тепер легко довести, що точки H_1, M та D колінеарні, використовуючи критерій II [2, с. 180].

Твердження про те, що точки D, M, C та C_1 розміщені конциклічно можна

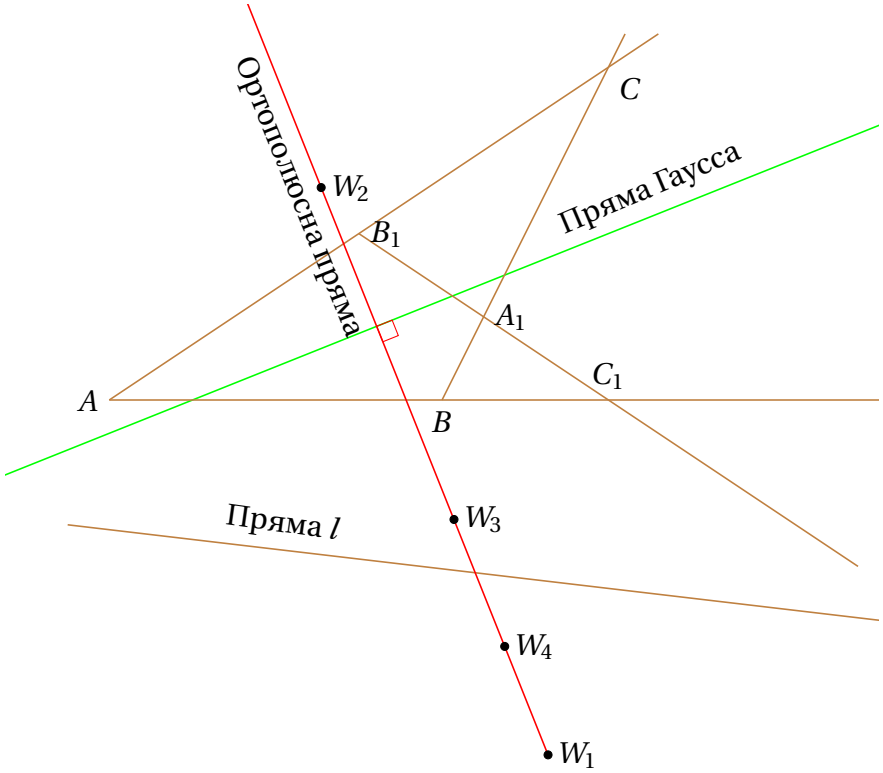


Рис. 5. *Узагальнення теореми про прямі Гаусса та Обера*

перевірити, використовуючи критерій III [2, с. 180], оскільки подвійне відношення цієї четвірки точок

$$(d,c;m,c_1)=\frac{(b-b_1)(a-a_1)}{(b-a_1)(b_1-a)}\times$$
$$\times\frac{-abb_1-a_1bb_1+aba_1+aa_1b_1}{a_1bb_1+aba_1-aa_1b_1-abb_1}\times$$
$$\times\frac{(a+a_1-b-b_1)(aa_1+bb_1)}{(a+b_1-b-a_1)(ab_1+ba_1)}$$

збігається зі спряженим. □

Далі ми доведемо теорему, що узагальнює твердження про взаємне розташування прямих Гаусса та Обера [2, с. 68]. Хоча першу частину теореми (про те, що чотири ортополюси належать одній прямій) можна довести за допомогою засобів проєктивної геометрії, ми доведемо обидві частини теореми методом комплексних чисел.

Теорема 5. *Нехай на площині дано повний чотирибічник (див. формулювання теореми 4) та пряма l. Тоді ортополюси цієї прямої відносно чотирьох трикутників, утворених трійками прямих повного чотирибічника, належать одній прямій, яка перпендикулярна прямій Гаусса (рис. 5).*

Схема доведення. Нехай коло, яке описане навколо $\triangle ABC$ є одиничним із центром у початку координат комплексної площини. Умови колінеарності точок A, B_1, C та A_1, B, C матимуть вигляд

$$a(\overline{b_1}-\overline{c})+b_1(\overline{c}-\overline{a})+c(\overline{a}-\overline{b_1})=0$$

та

$$c(\overline{a_1}-\overline{b})+a_1(\overline{b}-\overline{c})+b(\overline{c}-\overline{a_1})=0,$$

звідки отримуємо вирази для спряжених величин

$$\overline{a_1}=\frac{b+c}{bc}-\frac{a_1}{bc},\qquad \overline{b_1}=\frac{a+c}{ac}-\frac{b_1}{ac}.$$

Нехай W_4 — ортополюс прямої l відносно $\triangle A_1B_1C$, а W_1 — ортополюс прямої l

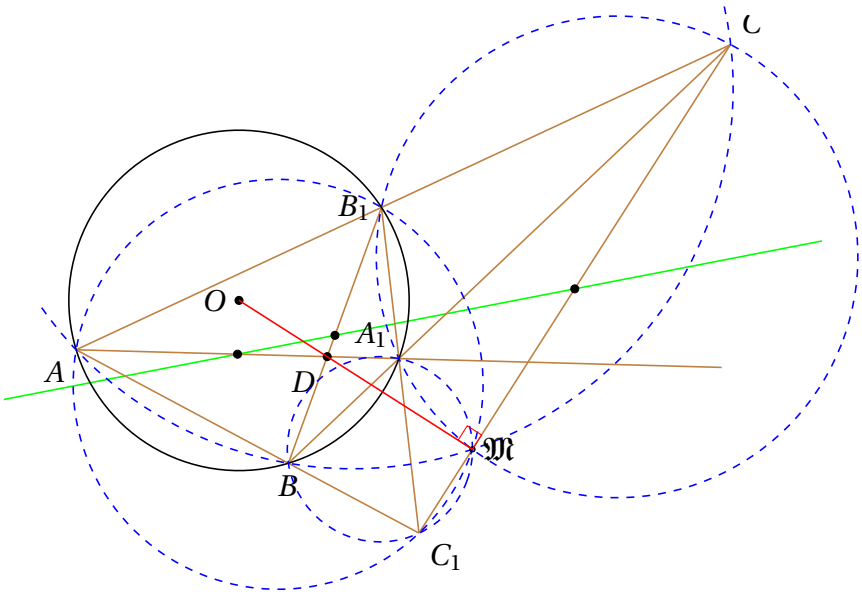


Рис. 6. *Повний чотирибічник і точка Мікеля*

відносно $\triangle ABC$. Пряма l задана рівнянням $z - z_0 = \kappa(\overline{z} - \overline{z_0})$. Тоді афікси ортополюсів W_4 і W_1 запишуться у вигляді

$$w_4 = \frac{abc}{2\kappa} \cdot \frac{b_1 - a_1}{b - a} + \frac{aa_1 - bb_1}{2(b - a)} + \frac{c}{2} -$$
$$-\frac{1}{2}\left\{\frac{ab_1 - ba_1}{b - a} - z_0 - \frac{\kappa}{c}\left(1 + \frac{b_1 - a_1}{b - a}\right) + \kappa\overline{z_0}\right\},$$
$$w_1 = \frac{1}{2}\left\{a + b + c - \frac{abc}{\kappa} + z_0 - \kappa\overline{z_0}\right\}.$$

Тому

$$w_4 - w_1 = \frac{1}{2} \cdot \frac{(b_1 - a_1 + b - a)(bc - \kappa)(ac - \kappa)}{(b - a)\kappa c},$$

а отже,

$$\overline{w_4} - \overline{w_1} = \frac{1}{2} \cdot \frac{(aa_1 - bb_1)(\kappa - bc)(\kappa - ac)}{(a - b)\kappa abc^2}.$$

З останніх двох рівностей випливає, що комплексний кутовий коефіцієнт прямої W_1W_4 буде такий:

$$\kappa_{W_1W_4} = \frac{a + a_1 - b - b_1}{aa_1 - bb_1} abc.$$

Тому

$$\kappa_{W_1W_4} + \kappa_g = 0,$$

де κ_g — комплексний кутовий коефіцієнт прямої Гаусса g. Аналогічно доводиться, що $W_4W_2 \perp g$ і $W_4W_3 \perp g$. Звідси випливають обидва твердження теореми.

Тепер, якщо пряма l збігається з однією з прямих, які утворюють даний повний чотирибічник, то **ортополюсна пряма** переходить у пряму Обера. При цьому три ортополюси переходять у відповідні ортоцентри. Таким чином, на прямій Обера містяться чотири ортоцентри і ще чотири додаткові ортополюси відносно трикутників, сторони яких не належать вибраним (у якості прямої l) прямим даного повного чотирибічника. □

Доведемо тепер властивість точки Мікеля [4, с. 40] для одного частинного випадку розташування чотирьох прямих на площині.

Теорема 6. *Нехай прямі* α, β, γ *і* δ *утворюють повний чотирибічник із вершинами* $\alpha \cap \beta = C$, $\gamma \cap \delta = C_1$, $\alpha \cap \gamma = B$, $\beta \cap \delta = B_1$, $\beta \cap \gamma = A$, $\alpha \cap \delta = A_1$. *Нехай точки* A, B, A_1 *та* B_1 *належать одному колу* ω *із центром у точці* O, *а* $AA_1 \cap BB_1 = D$. *Тоді прямі* OD *та* CC_1 *перетинаються у точці Мікеля* $\mathfrak{M}$ *повного чотирибічника, причому* $OD \perp CC_1$, *а точка* D *переходить у точку* $\mathfrak{M}$ *при інверсії відносно* ω *(рис. 6).*

Доведення. Нехай коло, яке проходить через A, B, A_1 та B_1 є одиничним із центром у точці $z = 0$ комплексної площини. Нехай дано афікси a, b, a_1 та b_1 точок A, B, A_1 и B_1. Афікси точок C і C_1 знаходимо, використовуючи формулу XII [2, с. 181]:

$$c = \frac{\frac{1}{a} + \frac{1}{b_1} - \frac{1}{b} - \frac{1}{a_1}}{\frac{1}{ab_1} - \frac{1}{ba_1}},$$

$$c_1 = \frac{\frac{1}{a} + \frac{1}{b} - \frac{1}{a_1} - \frac{1}{b_1}}{\frac{1}{ab} - \frac{1}{a_1 b_1}}.$$

Записавши рівняння кіл, які проходять через вершини трикутників A_1B_1C і ABC [2, с. 180, пункт V], та, спростивши вирази, отримуємо систему рівнянь для знаходження афіксу точки Мікеля $\mathfrak{M}$:

$$\begin{cases} \dfrac{z - b_1}{z - a_1} \cdot \dfrac{b}{a} = \dfrac{\overline{z}b_1 - 1}{\overline{z}a_1 - 1} \\ \dfrac{z - b}{z - a} \cdot \dfrac{b_1}{a_1} = \dfrac{\overline{z}b - 1}{\overline{z}a - 1}. \end{cases}$$

Таким чином, точка Мікеля визначається комплексною координатою

$$z_{\mathfrak{M}} = \frac{bb_1 - aa_1}{b + b_1 - a - a_1}.$$

Колінеарність трійок $(\mathfrak{M}, O, D)$ і $(\mathfrak{M}, C, C_1)$ доводиться застосуванням критерію II [2, с. 180].

Зауважимо, що точка D перетину діагоналей вписаного в одиничне коло ω чотирикутника ABA_1B_1 переходить у точку Мікеля $\mathfrak{M}$ повного чотирибічника при інверсії відносно даного кола. Це твердження випливає з виразів для афіксів $z_{\mathfrak{M}}$ та d (див. доведення теореми 4), оскільки в цьому випадку виконується рівність

$$d \cdot \overline{z_{\mathfrak{M}}} = 1.$$

Таким самим чином можна довести, що $OD \perp CC_1$. □

Застосування комплексних чисел особливо доцільне, коли мова йде про правильні багатокутники. Доведемо одну з теорем такого роду.

Теорема 7. *Нехай на площині дано правильний n-кутник $A_0A_1 \ldots A_{n-1}$, коло Ω довільного радіуса із центром, який збігається із центром багатокутника, та довільну точку Z кола Ω. Тоді для натурального числа m, такого, що $m < n$, виконано*

$$\sum_{l=0}^{n-1} ZA_l{}^{2m} = \text{const}$$

(рис. 7).

Доведення. Нехай вершини даного правильного n-кутника належать одиничному колу із центром у початку координат, і точка Z має комплексну координату z_0, а вершина A_0 міститься на дійсній осі у правій півплощині. Покладемо $\omega = e^{2\pi i/n}$. Тоді

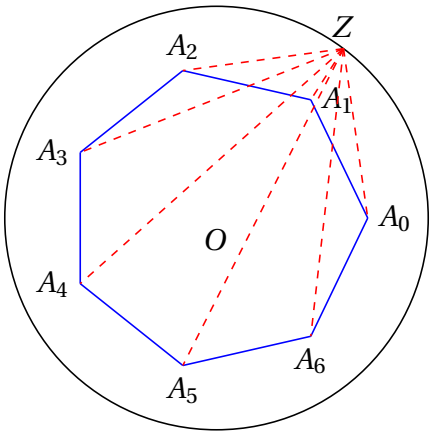


Рис. 7. *Властивість правильного багатокутника*

$$\mathfrak{S} := \sum_{l=0}^{n-1} ZA_l{}^{2m} = \sum_{l=0}^{n-1} \left(z_0 - \omega^l\right)^m \left(\overline{z_0} - \omega^{-l}\right)^m =$$

$$= \sum_{l=0}^{n-1} \left\{ \left(|z_0|^2 + 1\right) - \left(\overline{z_0}\omega^l + z_0\omega^{-l}\right) \right\}^m.$$

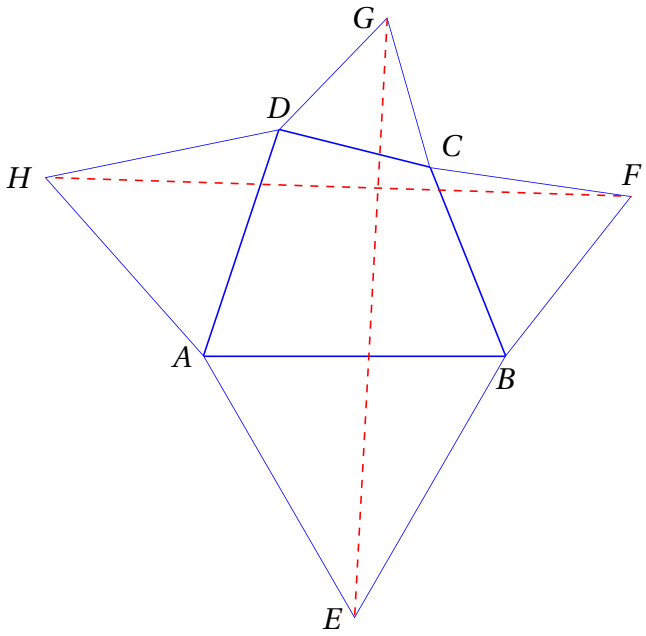


Рис. 8. *Чотирикутник і правильні трикутники, які побудовані на його сторонах*

Застосувавши двічі формулу бінома Ньютона, отримаємо

$$\mathfrak{S} = \sum_{l=0}^{n-1}\sum_{j=0}^{m}(-1)^j\binom{m}{j}\left(|z_0|^2+1\right)^{m-j}\times$$
$$\times\left(\overline{z_0}\omega^l + z_0\omega^{-l}\right)^j =$$
$$= \sum_{l=0}^{n-1}\sum_{j=0}^{m}\sum_{t=0}^{j}(-1)^j\binom{m}{j}\binom{j}{t}\left(|z_0|^2+1\right)^{m-j}\times$$
$$\times\left(\overline{z_0}\omega^l\right)^t\left(z_0\omega^{-l}\right)^{j-t}.$$

Спростивши праву частину останньої рівності, прийдемо до шуканого виразу для $\mathfrak{S}$:

$$\mathfrak{S} = n\sum_{k=0}^{[m/2]}\binom{m}{2k}\binom{2k}{k}\left(|z_0|^2+1\right)^{m-2k}|z_0|^{2k},$$

з якого випливає твердження теореми. □

Доведемо цікавий планіметричний факт, пов'язаний із правильними трикутниками. У цьому прикладі ми побачимо силу методу уявних чисел.

Теорема 8. *Нехай у чотирикутнику $ABCD$ на сторонах AB, BC, CD і AD назовні побудовано правильні трикутники AEB, CBF, DCG і ADH (рис. 8). У такому разі:*

1) відрізки EG та FH перпендикулярні тоді і тільки тоді, коли діагоналі AC та BD рівні;

2) відрізки EG та FH рівні тоді і тільки тоді, коли діагоналі AC та BD перпендикулярні.

Доведення. Нехай точка A збігається з початком координат у комплексній площині, а точка B належить дійсній осі. Введемо позначення $\omega := e^{2\pi i/3}$.

Доведемо перше твердження. Комплексні кутові коефіцієнти прямих HF та EG записуються у вигляді

$$\varkappa_{HF} = \frac{d-b+\omega(d+c-b)}{\overline{d}-b+\omega^2(\overline{d}+\overline{c}-b)},$$
$$\varkappa_{EG} = \frac{c+\omega(b-d+c)}{\overline{c}+\omega^2(b-\overline{d}+\overline{c})}.$$

Неважко бачити, що

$$\varkappa_{HF}+\varkappa_{EG}=0 \Leftrightarrow (d-b)(\overline{d}-b)=c\overline{c},$$

тобто, прямі EG и FH перпендикулярні тоді і тільки тоді, коли $AC = BD$.

Доведемо твердження (б). Умову рівності відрізків EG та FH можна записати так:

$$\begin{aligned}\big(c+\omega(b+c-d)\big)\big(\overline{c}+\omega^2(b+\overline{c}-\overline{d})\big) &= \\ = \big(d-b+\omega(d+c-b)\big)&\times \\ \times\big(\overline{d}-b+\omega^2(\overline{d}+\overline{c}-b\big),&\end{aligned}$$

звідки

$$\overline{c}(b-d)+c(b-\overline{d})=0,$$

тобто $EG=FH \iff AC\perp BD$. □

Наостанок наших досліджень доведемо за допомогою уявних чисел древню японську теорему.

Теорема 9. *Нехай на площині даний n-кутник $A_1A_2\dots A_n$, вписаний у коло Ω із центром у початку координат комплексної площини радіуса R. Вибрана довільна вершина A_p цього багатокутника, з якої проведено діагоналі, ща розбивають наш багатокутник на $n-2$ трикутники. Нехай у кожен із цих трикутників вписано коло. Тоді сума радіусів цих кіл не залежить від вибору вершини A_p:*

$$\sum_{k=1}^{n-2} r_{p,k}=\text{const}.$$

Доведення. Нехай вершини даного n-кутника мають комплексні координати Rz_k, де $|z_k|=1$. Нехай $z_k=e^{i\phi_k}$, де $0\le\phi_k<2\pi$, а точки з афіксами $\sqrt{Rz_k}=\sqrt{R}e^{i\phi_k/2}$ містяться у верхній півплощині. Тому всі трикутники з вершинами в точках із вибраними значеннями $\sqrt{z_k}$ тупокутні, а значить, якщо кут при вершині $\sqrt{Rz_{p+k}}$ тупий, то в цьому випадку виконується потрібна нам нерівність

$$|\sqrt{z_p}-\sqrt{z_{p+k}}+\sqrt{z_{p+k+1}}|<1.$$

Позначимо через $\Delta_{p,k}$ трикутник із вершинами, які мають комплексні координати Rz_p, Rz_{p+k} и Rz_{p+k+1}.

Відповідно до формули Ейлера, квадрат відстані між центрами вписаного й описаного кіл кожного трикутника $\Delta_{p,k}$ обчислюється за формулою:

$$d_{p,k}^2=R^2-2Rr_{p,k}.$$

Тому

$$\sum_{k=1}^{n-2} r_{p,k}=\frac{1}{2R}\sum_{k=1}^{n-2}\{R^2-d_{p,k}^2\},$$

звідки

$$\sum_{k=1}^{n-2} r_{p,k}=\frac{R(n-2)}{2}-\frac{1}{2R}\sum_{k=1}^{n-2} d_{p,k}^2. \qquad (13)$$

Запишемо вирази для $d_{p,k}^2$, спираючись на логіку обчислень у [3, с. 197]:

$$\begin{aligned}d_{p,1}^2&=R^2\left(\sqrt{z_pz_{p+1}}+\sqrt{z_{p+1}z_{p+2}}-\sqrt{z_pz_{p+2}}\right)\times\\ &\times\left(\frac{1}{\sqrt{z_pz_{p+1}}}+\frac{1}{\sqrt{z_{p+1}z_{p+2}}}-\frac{1}{\sqrt{z_pz_{p+2}}}\right),\\ d_{p,2}^2&=R^2\left(\sqrt{z_pz_{p+2}}+\sqrt{z_{p+2}z_{p+3}}-\sqrt{z_pz_{p+3}}\right)\times\\ &\times\left(\frac{1}{\sqrt{z_pz_{p+2}}}+\frac{1}{\sqrt{z_{p+2}z_{p+3}}}-\frac{1}{\sqrt{z_pz_{p+3}}}\right),\\ &\dots\\ d_{p,n-p}^2&=R^2\left(\sqrt{z_pz_1}+\sqrt{z_nz_1}-\sqrt{z_pz_n}\right)\times\\ &\times\left(\frac{1}{\sqrt{z_pz_1}}+\frac{1}{\sqrt{z_nz_1}}-\frac{1}{\sqrt{z_pz_n}}\right),\\ d_{p,n-p+1}^2&=R^2\left(\sqrt{z_pz_2}+\sqrt{z_1z_2}-\sqrt{z_pz_1}\right)\times\\ &\times\left(\frac{1}{\sqrt{z_pz_2}}+\frac{1}{\sqrt{z_1z_2}}-\frac{1}{\sqrt{z_pz_1}}\right),\\ &\dots\\ d_{p,n-2}^2&=R^2\left(\sqrt{z_pz_{p-1}}+\sqrt{z_{p-2}z_{p-1}}-\sqrt{z_pz_{p-2}}\right)\times\\ &\times\left(\frac{1}{\sqrt{z_pz_{p-1}}}+\frac{1}{\sqrt{z_{p-2}z_{p-1}}}-\frac{1}{\sqrt{z_pz_{p-2}}}\right).\end{aligned}$$

Розкривши дужки та просумувавши отримані рівності, прийдемо до формули

$$\begin{aligned}\sum_{k=1}^{n-2} d_{p,k}^2&=3R^2(n-2)-\\ -R^2&\sum_{k=-1}^{n-2}\left\{\sqrt{\frac{z_{p+k}}{z_{p+k+1}}}+\sqrt{\frac{z_{p+k+1}}{z_{p+k}}}\right\}.\end{aligned} \qquad (14)$$

Із рівностей (13) та (14) отримаємо

$$\sum_{k=1}^{n-2} r_{p,k} = -R(n-2)+$$
$$+\frac{R}{2}\sum_{k=-1}^{n-2}\left\{\sqrt{\frac{z_{p+k}}{z_{p+k+1}}}+\sqrt{\frac{z_{p+k+1}}{z_{p+k}}}\right\}=$$
$$=-R(n-2)+\frac{R}{2}\mathfrak{G},$$

де форма $\mathfrak{G}$ не залежить від p (тут позначено $z_{n+m}=z_m$). Доведення завершене. □

Наведене доведення, звичайно, не є оптимальним, однак його перевага в тому, що ми отримуємо точний вираз для суми радіусів через комплексні координати вершин, а пошук оптимального шляху міг би зайняти значно більше часу. Зауважимо також, що остання формула вказує на оптимальний шлях геометричного (класичного) доведення з мінімальним числом формул.

Автор сподівається, що геометричні теореми цієї статті разом із доведенням древньої японської теореми є новими та необхідними і що вони ще раз підкреслюють важливість методу комплексних чисел у планіметрії.

Автор дуже вдячний редакції журналу «У світі математики» за увагу, приділену цій роботі, за допомогу та цінні поради при підготовці матеріалу до друку, а також бажає всім любителям математики творчих успіхів.

Література

[1] Шарыгин И. Ф. Задачи по геометрии. Планиметрия / И. Ф. Шарыгин — М. : Наука, 1982. — 160 с.

[2] Скопец З. А. Геометрические миниатюры / З. А. Скопец — М. : Просвещение, 1990. — 224 с.

[3] Моденов П. С. Задачи по геометрии / П. С. Моденов — М. : Наука, 1979. — 368 с.

[4] Прасолов В. В. Задачи по планиметрии. Ч. 1 / В. В. Прасолов — М. : Наука, 1991. — 320 с.

Робота актуарієм у Великій Британії

Андрощук Тарас Олександрович,
кандидат фізико-математичних наук

У попередньому номері журналу було опубліковано початок моєї розповіді про актуарну професію. У другій частині своєї розповіді я зупинюся більш детально на питаннях щодо навчання та роботи актуарієм у Великій Британії, навівши всесторонній практичний огляд, що базується на власному досвіді.

Як і де навчаються британські актуарії?

У Великій Британії все актуарне навчання сконцентроване навколо кваліфікаційних вимог до актуарія, які задає професійна організація актуаріїв, що називається «Інститут та факультет актуаріїв» (Institute and Faculty of Actuaries, www.actuaries.org.uk). Така дивна подвійна назва пояснюється тим, що раніше в Британії існувало дві актуарні професійні організації — англійська, «Інститут актуаріїв», та шотландська, «Факультет актуаріїв». 2010 року ці організації об'єднались в одну і вирішили зберегти частинки своїх минулих назв у новій назві. Досі офіційне актуарне звання позначається трьома літерами у двох варіантах, залежно від того, з англійською чи шотландською організацією кандидат себе більш асоціює: «FIA» (Fellow of the Institute of Actuaries) чи «FFA» (Fellow of the Faculty of Actuaries). Ці три літери є предметом чималої гордості для кваліфікованих актуаріїв і об'єктом потаємних мрій і бажань для актуарних студентів.

Узагалі, актуарна професія є порівняно давньою — вона з'явилась саме тут, у Британії, майже 170 років тому. За цей час професія стала досить консервативною та зарегульованою. Так, для того, щоб отримати право називатися кваліфікованим актуарієм у Великобританії, потрібно здати десь близько 15 (важких або дуже важких) іспитів і виконати ряд інших вимог. Весь цей процес (від першого іспиту до кваліфікації) може зайняти від 5 до 15 років життя, або взагалі може ніколи не завершитись повною кваліфікацією, що теж трапляється.

Але професія актуарія вважається однією з найважчих не лише для отримання повної офіційної кваліфікації, а і для її утримання: окрім сплати щорічно значної суми членських внесків, повні (кваліфіковані) члени Інституту також повинні пройти й офіційно задекларувати принаймні 15 годин професійного розвитку, так зване «CPD» (Continious Professional Development), які можна отримати або відвідуючи відповідні конференції та семінари, або навчаючись на онлайн-курсах на сайті самого Інституту — і одне, і друге є досить затратним як з погляду часу, так і у грошах.

Оскільки шлях до актуарної кваліфікації є довгим, у реальному житті різні люди його проходять досить різними стежками. Звичайно, варіанти тут значно залежать від бекграунду студентів, зокрема від країни походження.

Шлях більшості британських студентів такий. Багато місцевих університетів мають спеціалізовані актуарні бакалаврські та магістерські курси. Програми таких курсів узгоджено з Інститутом актуаріїв: теми і, що важливіше, іспити за ними є дуже подібними до кваліфікаційних вимог Інституту. Університети часто мають прямі домовленості з Інститутом про автоматичне зарахування результатів іспитів за відповідними предметами в результати екзаменів Інституту. Це зручно — студент після вищої школи поступає на бакалаврський курс університету за актуарним напрямком, одержує диплом, і, якщо гарно вчився, вже має 4 – 7 зданих професійних іспитів. Якщо він потім успішно вчиться на магістерській програмі, то має ще 3 – 6 зданих іспитів, тільки вже складнішого рівня.

Таким чином, деякі студенти до закінчення університету проходять більше половини кваліфікаційного шляху. Завершують кваліфікацію вони вже працюючи; про це йтиметься далі. Зауважу, що не у всіх британських студентів такий прискорений шлях. Є багато молодих людей, які обирають фах актуарія, уже завершивши повне університетське (а іноді і наукове, тобто PhD) навчання за іншою спеціальністю: математика, фізика, інженерія, економіка тощо. Такі студенти не мають жодних переваг порівняно з більшістю українських студентів.

Цікавою є категорія студентів із країн, де існують актуарні професійні і навчальні організації, афільовані з британським Інститутом актуаріїв. У силу історичних чи інших причин, їхні місцеві «Інститути актуаріїв» мають системи навчання й іспитів, майже ідентичні до британської. Здані в таких країнах іспити визнаються британським інститутом нарівні зі своїми, що надає певну перевагу актуарним студентам, які вирішують переїхати до Великобританії, щоб продовжувати вчитись і працювати актуаріями. Я не знаю повний список таких країн, але, для прикладу, до нього входять ПАР та Індія.

Щодо інших країн у світі — більшість розвинених країн мають свої професійні та/або навчальні актуарні організації (Україна теж має свою актуарну організацію та систему кваліфікування, про що поговоримо далі). Стандарти і системи навчання та кваліфікації є дуже різними, від просто проходження спеціалізованого актуарного курсу в університеті до вимоги здавати незалежні системи іспитів, як у Британії.

Британський інститут актуаріїв визнає деякі національні кваліфікації та дозволяє актуаріям, кваліфікованим за правилами тих країн, перезарахувати свою кваліфікацію у Британії. У таких випадках все одно потрібно здати один чи декілька додаткових іспитів (наприклад, на Британське страхове законодавство), виконати певні умови щодо досвіду роботи у Британії, але це все є мінімальними вимогами порівняно із проходженням повного шляху кваліфікації. Я мав колег із декількох західноєвропейських країн та Австралії, які кваліфікувалися таким шляхом. США та Канада, здається, теж належать до цього списку. На жаль, українська актуарна кваліфікація британцями для перезарахування не приймається.

Навчання українських актуаріїв в Україні та Великій Британії

Повернімося до України. Хоча українська актуарна кваліфікація поки що не є еквівалентною британській (щодо перезарахування), у нашій країні процес актуарного навчання вже запущено. Теперішні актуарні студенти мають значно більше можливостей вчитися, ніж це було, наприклад, ще 10 років, а тим більше, 20 років тому. Зазначимо, що за бажанням, можна

користуватися українською, а не лише англійською мовою, і не обов'язково «заочно» навчатися у Британії чи США (як це робили мої однолітки) або виїжджати з країни.

Наскільки мені відомо, головними досягненнями останнього часу стали офіційна реєстрація предметів з актуарної математики як можливих дисциплін навчання у реєстрі МОН України; відкриття актуарної спеціалізації для бакалаврської та магістерської програм навчання на механіко-математичному факультеті Київського національного університету імені Тараса Шевченка, а також спрощення й орієнтація на «вітчизняне» актуарне навчання вимог до актуарної кваліфікації Державною комісією з регулювання ринків фінансових послуг.

Я можу не знати всіх деталей цих змін, тому відсилаю читача до сайтів першоджерел:

1) `http://mechmat.univ.kiev.ua` та `http://probability.univ.kiev.ua/` — сайти мехмату та спеціальності кафедри теорії ймовірностей, статистики та актуарної математики;

2) `http://zakon2.rada.gov.ua/laws/show/z0265-05` — посилання на нові кваліфікаційні вимоги;

3) `http://actuary.in.ua/` — сайт ТАУ.

Окрім ситуації з актуарним навчанням у самій Україні, розкажу трохи більше зі свого досвіду — які є реальні шляхи для українських студентів кваліфікуватися за вимогами британської актуарної професії та працювати на британському страховому ринку, у разі, якщо таким буде їхній вибір. Очевидно, потрібно володіти англійською мовою достатньою мірою для можливості навчатися та складати нею іспити. Наступний шлях уже був пройдений багатьма випускниками мехмату:

1) *Самостійно реєструєтеся як студент у Британському інституті актуаріїв (*`www.actuaries.org.uk`*).* Це можна зробити ще під час навчання в українському виші. Я особисто та мої однокурсники робили це вже після завершення навчання на мехматі, але тому весь процес у нас і зайняв довший час, ніж міг би бути. Студенти Інституту повинні сплачувати щорічний членський внесок до Інституту, але для категорії «студентів» він є відносно невисоким. Крім того, для студентів із деяких країн, включаючи Україну, є суттєва знижка на членський внесок (за умови, якщо щорічний дохід студента не перевищує певний установлений Британським інститутом рівень).

2) *Починаєте помалу здавати іспити, необхідні для кваліфікації.* Навчання в Інституті побудовано так. Ви самі обираєте, які предмети і коли здавати. Є певний логічний порядок і залежність предметів один від іншого, тому варто дотримуватися рекомендацій щодо порядку здачі іспитів, які дають на сайті Інституту. Сесії іспитів проходять двічі на рік — навесні та восени. Ви завчасно обираєте предмет(и), який(які) здаватимете у наступну сесію, і починаєте готуватися. Матеріали для підготовки до кожного іспиту можна придбати на сайті Інституту — це рекомендоване джерело підготовки, оскільки вони «заточені» під сам іспит, тобто мають «необхідну і достатню» кількість інформації та рівень деталей для здачі іспиту. Також можливо пройти курси підготовки до конкретних іспитів (у мій час такі курси епізодично діяли у Львові та Києві).

Рекомендовано здавати один предмет за сесію (хоча декому вдавалося здавати два або більше предметів за раз). Ближче до дати іспиту ви офіційно реєструєтесь на цей іспит (це коштує деяких грошей, знову ж таки, зі знижкою для України). Іспити в Україні проходять у спеціальних екзаменаційних центрах, які вказані на сайті Інституту під час реєстрації (ще коли я навчався, такий центр спочатку був у Харкові, а потім у Києві — на мехматі КНУ імені Тараса Шевченка).

На самому іспиті дуже строгі умови: завдання відкривають із запечатаного конвер-

та безпосередньо на початку іспиту; роздають і збирають завдання за годинником у точно визначений час (оскільки іспити проходять синхронно й одночасно по всьому світу); дозволяється приносити на іспит лише ручку і калькулятор; при будь-якому порушенні «правил чесної гри», таких як списування, розмови тощо — «фейл, бан», та, можливо, кінець усій майбутній кар'єрі актуарія — наполегливо не рекомендується. Результати іспитів публікуються на сайті Інституту через декілька місяців після складання.

3) *Деякі предмети теоретично можна перезарахувати без складання іспитів.* Якщо основна спеціалізація, за якою ви навчаєтесь в українському виші, має предмети, програми яких покривають програми іспитів Інституту, то є шанс перезарахувати оцінки за деякими предметами у результати актуарних іспитів. Цей шлях дає перевагу, як у британських студентів, які навчаються на актуарних програмах і мають право перезарахувати свої університетські курси в іспити Інституту, про що я писав вище. Єдина відмінність полягає у тому, що в процесі перезарахування для наших студентів потрібно підготувати набагато більше документів, і при цьому немає ніяких гарантій щодо результату. Тим не менш, мені та деяким моїм друзям, наприклад, вдалося перезарахувати два предмети: CT3 «Імовірність та математична статистика» та CT7 «Бізнес-економіка».

Для цього ми робили таку послідовність кроків: брали офіційні програми за відповідними курсами на мехматі (відповідність була неоднозначною — декілька мехматівських курсів покривали один актуарний предмет); перекладали програми англійською; робили таблицю покриття тем програми актуарного іспиту програмами наших курсів; робили копії та переклади наших дипломів із відповідними оцінками за цими предметами і відсилали це все в Інститут. Результат бував різний — декому відмовляли, декому перезараховували. Але, враховуючи те, що підготовка до складання кожного іспиту в середньому займає пів року, спробувати варто. Можливо зараз, коли в Україні вже є навчання за офіційною актуарною спеціалізацією, процедура перезарахування стала простішою, і, можливо, список предметів, які можна перезарахувати, теж розширився; щодо цього вже доведеться дізнаватися самостійно.

4) *Проблема складання останніх, спеціалізованих, іспитів.* Іспити, які потрібно складати для повної кваліфікації, поділяються на 4 ступені: Базові технічні (серія «CT»), Базові практичні (серія «CA»), Спеціалізовані технічні (серія «ST») та Спеціалізовані практичні (серія «SA»). Я знаю багатьох людей, які перебуваючи ще в Україні, змогли скласти багато або всі іспити серії «CT». Також дехто міг здати якісь іспити серії «CA» (особливо, враховуючи, що деякі з них тепер можна складати онлайн). Але мені не відомі випадки, щоб хтось заочно, з України, здав іспити серії «ST» чи «SA». Я, звичайно, можу просто не знати про такі випадки — вітаю, якщо вони є! — але беззаперечним є те, що здати ці спеціалізовані іспити, перебуваючи поза британським «актуарним середовищем», дуже важко. Причин на те є декілька.

По-перше, банально, англійська мова. Якщо іспити базових рівнів є більш технічно орієнтованими, то й вимоги до англійської у них простіші. Зрозумівши умову задачі, записуєш відповідну математику і користуєшся англійською у більшості випадків лише для простих пояснень і логічних переходів між своїми висновками. Спеціалізовані ж іспити, за своєю суттю, потребують набагато менше математичних висновків, зате прямо вимагають від студента дати визначення або описати, пояснити, аргументувати, продискутувати, оцінити «за» та «проти» і так далі. Не перебуваючи у мовному середовищі і не «проговоривши» усі ці вимоги багато разів і в деталях безпосередньо під час навчання англійською мовою,

важко знайти потрібний лексикон, «перекладаючи» в голові з української.

По-друге, спеціалізовані іспити написані в контексті страхового ринку Британії. Відповідно, тут будуть використовуватись (і вимагатися від студента) приклади та посилання на британські страхові продукти, практики, закони, оподаткування, певні статистичні дані і так далі. Зважаючи на те, що частини елементів цього списку (наприклад, деяких продуктів, практик) в Україні просто не існує, а інша частина списку (наприклад, закони, оподаткування) є іншими та нерелевантними, то використовувати практичні знання з українського страхового ринку у відповідях на іспиті не можна. Отже, бажано перебувати у середовищі британського страхового ринку і при навчанні чи роботі напрацьовувати досвід саме цього ринку — це допоможе скласти відповідні іспити.

5) *Очне магістерське навчання у Британії та робота.* Як же вирішити цю проблему відсутності практики мови та контексту (ринку) для складання практичних іспитів українському студенту? Одне практичне вирішення, яке мені відомо — і я знаю чимало людей, які успішно ним скористались, включаючи мене — це приїхати до Великої Британії навчатись на очній актуарній магістерській програмі, а потім почати тут працювати актуарієм.

Магістерська програма («Master», «MSc») зазвичай триває один рік. При успішному навчанні, за цей рік вдається:

1) покращити мову (особливо спеціалізовану професійну);

2) здати/перезарахувати декілька необхідних для кваліфікації іспитів, включаючи деякі з вищих рівнів (CT, CA чи ST, але зазвичай не SA);

3) отримати британський магістерський диплом від хорошого університету чи бізнес-школи.

Усі три здобутки допомагають після завершення навчання знайти роботу. Робота, у свою чергу, допомагає завершити кваліфікацію, спонсоруючи навчання та складання останніх іспитів, що залишились (включаючи SA), та надаючи необхідний досвід роботи («work based skills»), який також є частиною кваліфікаційних вимог.

Цей шлях вимагає певних грошових інвестицій: навчання в магістратурі платне, прожиття в Британії в перший рік без роботи теж недешеве. Тому весь цей проєкт є чималим ризиком, і необхідно реалістично зважити всі «за» та «проти» перед тим, як наважитись на нього (застосувати своє найкраще «актуарне судження»).

Приклади того, коли їхати, теж є різні: дехто приїхав уже після здобуття наукового ступеня та/або отримання суттєвого досвіду роботи в Україні, а дехто приїхав одразу після завершення магістерської, або навіть бакалаврської програми в Україні. У кожного свій шлях, який залежить від індивідуальних обставин і оцінювань своїх власних ризиків і потенційних переваг. Для того, щоб краще зрозуміти, якими є ці потенційні переваги та ризики, пропоную ознайомитись із наступним, останнім, розділом цієї статті.

Робота і кар'єра актуарія у Великій Британії

Ми так багато говорили про саму актуарну професію, про актуарне навчання та кваліфікацію, навіть про те, як «податися на чужину в найми», але поки що, напевно, не дуже зрозуміло, навіщо та заради чого. Тож я спробую тут коротко описати, як складається «серединна» кар'єра актуарія у Великій Британії і які її можливі плюси та мінуси.

Перша робота

Першу роботу актуарієм у Великій Британії можна знайти одразу після завершення

навчання в університеті. Починати, швидше за все, доведеться з однієї з найнижчих позицій «кар'єрної драбини» — тобто аналітик («analyst») або помічник («associate»).

Для цих стартових позицій, ще під час навчання, деякі роботодавці пропонують конкурсні «випускні» програми («analyst/graduate program»), за якими можна отримати пропозицію (оффер, «offer») першого місця роботи ще до завершення навчання. Для цього необхідно пройти низку тестів і співбесіду та виграти чималий конкурс, але спроба може бути вартою того, оскільки, в разі успіху, надає гарантію працевлаштування, а також можливість знайти інші позиції та вибрати найкращу.

Для українського студента (студентки), який (яка) приїхав (приїхала) на британську актуарну магістерську програму одразу після завершення українського університету, почати кар'єру з найнижчої сходинки не повинно бути проблемою — так починають тут усі, і ти будеш дертися по цій «драбині» угору разом з усіма у відносно рівних умовах — найвище та найшвидше це зробить найсильніший.

Водночас, українському «студенту», який перед тим, як піти на цю магістерську програму, мав певний досвід актуарної роботи в Україні, або може навіть мав якісь високі посади в українських страхових компаніях, такий «даун-шифтинг» може здатись не дуже привабливим. Що ж, для таких «досвідчених» («experienced») кандидатів, це може бути першим ризиком і компромісом. Можна спробувати «продати» свій минулий досвід і отримати вищу аніж стартову посаду, але на великий «підйом» розраховувати не варто, максимум один-два щаблі.

Причина в тому, що український актуарний досвід не є таким, що його легко можна «конвертувати» у британський. Такий досвід у вашому резюме не те, щоб зовсім не брався до уваги при пошуку роботи, але на практиці, як мінімум, дуже сильно «дисконтується». Буває навіть, що деякі записи у вашому резюме, які в Україні наче, здавалось, виглядали солідно та поважно, тут можуть ускладнити процес співбесіди, оскільки викликатимуть здивування і потребуватимуть додаткових незручних пояснень. Наприклад, на співбесіді можуть спитати: «Як це можливо, що одразу після випуску з університету, не будучи кваліфікованим актуарієм, ви були єдиним актуарієм на всю компанію, а через рік уже мали посаду начальника відділу актуарних розрахунків?»

Правильною тактикою в таких випадках є не акцентувати увагу на посадах та назвах (хоча, навмисне видаляти їх із резюме теж, звичайно, не потрібно), а робити акцент на тому, які саме обов'язки ви виконували і який практичний досвід актуарних розрахунків із того часу маєте. Досвідчені кандидати, так само, як і недосвідчені, які не гарантували собі місце роботи через випускні програми, роботу шукають через агентства.

Наостанок додам лише, що актуарних вакансій на ринку є багато, конкуренція між роботодавцями й агенціями, що їх представляють, теж велика; тому, якщо маєте бажання, енергію, гарну технічну підготовку, пристойний рівень англійської і певний рівень загальних комунікаційних навичок, першу роботу знайти цілком реально. Але зважте на всі перераховані умови — якщо якась із них не виконується, є ризик довго шукати першу роботу, або не знайти її взагалі. Це є суттєвим ризиком для людини, яка приїхала в іншу країну, має обмежену за часом візу і потребує якнайшвидшого регулярного доходу, щоб мати змогу оплачувати своє прожиття та навчання.

Наводити більше деталей, ніж уже сказано, вважаю, немає сенсу, бо спеціалізованої інформації є занадто багато. Крім того, у студента, який навчатиметься тут на магістерській програмі, велика частина його останнього року навчання повинна бути присвячена саме тому, щоб гарно розібратись в усіх пропозиціях на ринку пра-

ці і сформувати тактику найкращого і найшвидшого самостійного входу в цей ринок.

Професійний розвиток і кар'єрне зростання

Після отримання першої роботи кар'єрне зростання актуарія поділяється на два етапи: ДО кваліфікації, та ПІСЛЯ кваліфікації. З точки зору ієрархії, цей поділ відбувається приблизно на рівні помічника чи менеджера (друге радше як виняток). Зате після кваліфікації актуарії дуже швидко підіймаються до менеджера, а також їм відкривається шлях рости вище (що неможливо без кваліфікації).

Більшість роботодавців мають різні програми підтримки для ще не кваліфікованих працівників. Це може включати в себе оплату різних витрат на навчання, як ось: членські внески в Інститут актуаріїв, реєстрацію на іспити, матеріали для підготовки до іспитів, очні курси чи участь у пробних тренувальних іспитах тощо. Також деякі роботодавці виділяють якусь кількість оплачуваних робочих днів як час для підготовки до іспитів.

Але всі ці оплати і преференції зазвичай проводяться за умови достатнього прогресу у складанні іспитів. Якщо ж студент недостатньо швидко просувається у кваліфікації, преференції можуть відмінити, і всі витрати доведеться нести самому. Також без швидкого прогресу в навчанні розраховувати на кар'єрне просування не доводиться. Очікування роботодавців для іспитів вищих рівнів — щонайменше один іспит у сесію (тобто два на рік). Для іспитів початкового рівня (CT), ці очікування можуть бути ще вищими.

Отже, зростання від аналітика до менеджера (середня ланка в ієрархії) дуже сильно пов'язано із кваліфікацією та може тривати у середньому від двох-трьох до п'яти років. Після кваліфікації кар'єрне зростання пов'язано, як і у всіх інших професіях, із багатьма факторами, які включають у себе технічні знання й експертизу; розуміння бізнесу своєї компанії і ринкового середовища навколо неї; так звані «м'які навички» («soft skills»): професійне спілкування, пояснення технічних деталей простою бізнесовою мовою, просування своїх продуктів і послуг, заведення та підтримання бізнес-зв'язків («networking») тощо. Тут у кожного вже будуть свої особисті обставини та прогрес.

Скажу лише, що на дуже стрімке зростання я б не розраховував (принаймні, не на такі карколомні кар'єри, які бувають в Україні). Зростання на один рівень позиції за два-три роки тут вважається дуже гарним прогресом. І далеко не всі просуваються навіть із такою швидкістю. Тому зважайте на цю реальність і пов'язані з нею ризики при побудові власних фінансових планів на майбутнє.

Потенційний дохід

У Великій Британії актуарії є відносно високооплачуваною професією. Навіть якщо починаєш працювати з найнижчої сходинки актуарної кар'єрної ієрархії (тобто, аналітик, «analyst»), отримуваний дохід має бути достатнім для прожиття, і, найшвидше, вже буде вищим і за медіанну, і за середню зарплату у Великій Британії.

За просування далі по кар'єрі дохід буде рости суттєво, і досить швидко перетне і 75 %, і 90 % процентилі у загальному розподілі доходу в цій країні, а за успішної кар'єри може перетнути і 95 %, і 99 % процентилі. Вище цього теж можливо рости, але це вже буде категорія «дуже успішної» кар'єри — рівень головних актуаріїв у великих страхових компаніях і партнерів у великих консалтингових компаніях — але для нас, простих українських актуарних гастарбайтерів, я б обмежив свої реалістичні очікуванні на рівні попередньої категорії

(хоча поганий той козак, який не хоче бути отаманом).

Якщо читача цікавлять конкретніші суми у реальних фунтах стерлінгів, я раджу попрацювати з Гугл, сайтом «Скляні двері» (www.glassdoor.co.uk) та розділом «Пропоную роботу» журналу «Актуарій» Британського інституту актуаріїв («The Actuary», www.theactuary.com).

Водночас, хочу зауважити декілька важливих нюансів для формування правильних очікувань і оцінювання ризиків. Починати доведеться з нижчих рівнів зарплати; до вищих рівнів доростають не одразу і не всі; над своїм зростанням потрібно працювати, і воно не відбувається просто так.

У Великій Британії діє сильно прогресивна система оподаткування — це означає, що за зростання номінальної зарплатні відповідне зростання реального доходу після оподаткування є значно менш суттєвим. Життя у Великій Британії, а особливо в Лондоні, є дорогим. У іноземців виникає також багато додаткових витрат, пов'язаних із візами, поїздками додому тощо. Життя разом зі своєю сім'єю теж значно збільшує рівень витрат. Тому розраховувати на швидкі великі накопичення від роботи актуарієм за кордоном я б не став.

Заключне слово

Чи корисними будуть відомості, викладені в цій статті, українським школярам та студентам? Маю надію, що так. Ми живемо у глобальному світі, який швидко змінюється, розвивається, а також є раціональним, відкритим та ефективним. Це означає, що у світі завжди існує попит на адекватних людей зі світлими головами та належними знаннями, і цей світ готовий включати таких людей у процеси глобальної економіки, незалежно від походження людей.

Патріотично це чи ні — включатись у такі глобальні економічні процеси? Відповідь на це питання є поза темою цієї статті, але моя думка така, що якщо країна стоїть на периферії глобальної економіки, це означатиме, що вона дорого купуватиме послуги з високою доданою вартістю (у тих же ж фінансах чи технологіях), а розраховуватиметься за них відносно дешевими товарами з низькою доданою вартістю (ті ж пшениця, сталь і дешева робоча сила). Щоб змінити ситуацію, людям такої країни якраз і потрібно починати розбиратись, включатись, а потім і переносити «додому» найбільш сучасні й ефективні практики, процеси, технології, ринки, професії.

Здається, першими в Україні, хто зрозумів нові широкі можливості глобалізованого світу, були програмісти. Вони швидко знайшли свої ніші у глобальному технологічному секторі: починаючи від індивідуальних робіт розробниками в різноманітних індустріях тут, на Заході (у мене, наприклад, є чимало знайомих українських програмістів, які працюють у тому ж фінансовому секторі у Британії чи у США), до масштабного корпоративного аутсорсингу великих проєктів безпосередньо з України.

Але те ж саме можливе і для спеціалістів фінансового ринку, зокрема, актуаріїв та інших математиків у фінансах (наприклад, «квонтів»). Те, що раніше сприймалось нами[1] «іншим світом» по ту сторону вебсторінки агенції Bloomberg чи газети «Financial Times», насправді, перебуває від нас лише у декількох (рішучих і потужних, але досяжних) кроках.

Якщо вжити економічні терміни, на талановитих професійних людей у глобальному світі є великий «попит» і непогана «ціна». Щоправда, також існує і достатня «пропозиція» таланту з найрізноманітніших куточків світу, що створює здорову конкуренцію. За час роботи в Лондоні, моїми колега-

[1] Можливо вже не «вами» — сучасними школярами та студентами з «інтернетами» та відповідними широкими обріями, але точно ще «нами» — поколінням, яке бачило народження пошуковика Гугл на останніх курсах університету.

ми були чудові люди із десятків країн з усіх континентів світу (Європа, Африка, Азія, обидві Америки, Австралія) — усі ці люди є хорошими спеціалістами, і всі свого часу теж робили свій вибір і свої відповідні «рішучі кроки», щоб приїхати в Лондон і працювати в одному із центрів глобальних світових фінансів. Як мені здається, всі ці люди, незалежно від країни походження, задля своєї мети проходили більш-менш однаковий за важкістю та ризиками шлях — тож українські молоді математики мають абсолютно ті ж самі можливості і ризики, що й їхні ровесники деінде у світі.

Я маю надію, що ця стаття допоможе комусь з українських школярів чи студентів поглянути на практичне застосування математики з нової точки зору, побачити нові перспективи, а маючи інформацію, можливо, зробити власний вибір та ті «рішучі і потужні» кроки до того бажаного майбутнього. Успіхів!

Внутрішні хвилі — відсутній елемент глобального кліматичного пазла

Терлецька Катерина Валеріївна,
Інститут проблем математичних машин та систем НАНУ

Ми всі чули про руйнівні поверхневі хвилі цунамі. Вони зароджуються у відкритому океані, де мають амплітуди у декілька метрів і розповсюджуються на тисячі кілометрів у вигляді усамітнених водяних пагорбів — солітонів, які були описані Джоном Расселом ще 1832 року [1]. Коли ж цунамі виходять на мілководдя, їхня амплітуда значно зростає і може сягати десятків метрів. Такі екстремальні поверхневі хвилі час від часу завдають удару по прибережних зонах морів і океанів та несуть цим ділянкам руйнування та лихо.

Але, мабуть, несподіваним буде факт, що кожен день хвиля з амплітудою до двох сотень метрів [2] та з довжиною сотні кілометрів формується у водах між островом Тайвань і Філіпінами і за декілька днів проходить через Південнокитайське море, хоча на поверхні вона ледь помітна. Хвиля подорожує від Лусонської протоки, де вона народжується, до шельфової зони Південнокитайського моря, де руйнується та гине. Крім того, вплив від життєвого циклу таких хвиль є надзвичайно важливим у формуванні глобального клімату.

Ця величезна хвиля заввишки з хмарочос називається внутрішньою гравітаційною хвилею і своїм існуванням вона зобов'язана різній густині води у водоймі біля поверхні та у її глибині.

Така різниця в густині виникає або за рахунок різниці солоності (наприклад, біля гирл річок, де на поверхні утворюється шар прісної води), або внаслідок різних температур (за рахунок прогрівання верхнього шару) та називається стратифікацією. Через те, що різниця в густині між двома шарами невелика (порівняно з різницею в густині на межі вода – повітря), амплітуди внутрішніх гравітаційних хвиль у рази більші за амплітуди поверхневих хвиль і можуть сягати сотень метрів. На рис. 1 зображено локації й амплітуди спостережуваних внутрішніх хвиль великих амплітуд і профіль найбільшої зафіксованої внутрішньої хвилі.

Побачити внутрішні хвилі у природних водоймах неможливо. Але їх можна створити і спостерігати у склянці. Так перше в історії задокументоване спостереження внутрішніх хвиль було зроблено в олійній лампі Бенджаміном Франкліном під час подорожі на кораблі. Він спостерігав хвилі на межі олії та води в італійській олійній лампі [3]. І хоча цьому документу близько 250 років, знайомство людства із цим феноменом має довгу історію з містичним забарвленням.

Містичні внутрішні хвилі

З давніх часів морякам було відоме загадкове явище «мертвої води». Воно проявлялось у тому, що судно, яке йшло малим ходом, раптово зупинялося, наче чиясь рука тримала його знизу. Вітрильні судна під

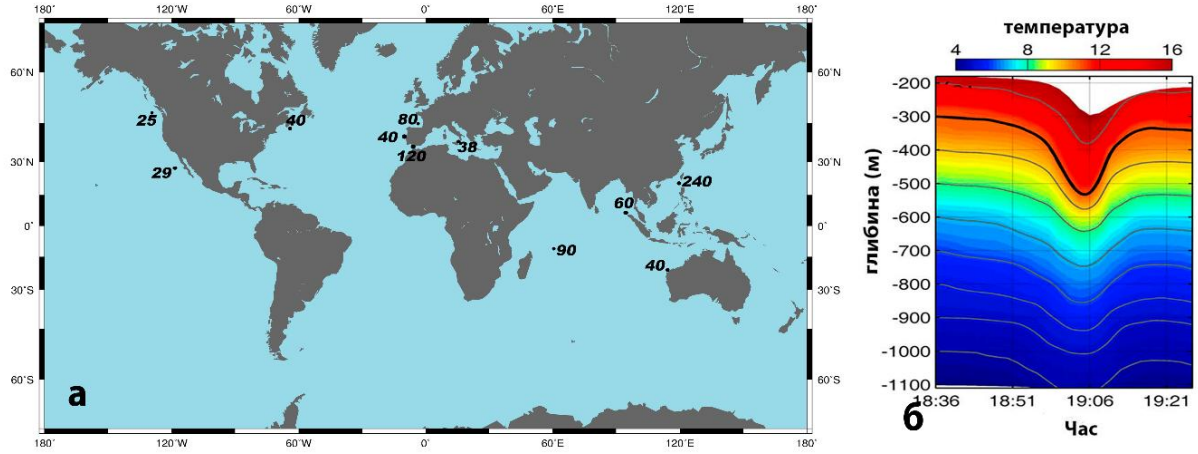


Рис. 1. *Найбільші внутрішні хвилі, що спостерігалися у Світовому океані: місця спостережень та амплітуди в метрах (а); профіль найбільшої зафіксованої внутрішньої хвилі (б) [2]*

дією «мертвої води» збивалися з курсу й переставали слухатися керма. Цей ефект приписували або містичним чудовиськам, або особливому типу риб, які прикріплювалися до кіля. Протягом століть «мертва вода» засмучувала досвідчених капітанів, а простих моряків взагалі жахала, аж поки не з'явилось фізичне пояснення.

Після плавання знаменитої полярної експедиції 1893–1896 рр. Фрітьйоф Нансен виразно описав картину збурень на поверхні води у разі потрапляння судна в «мертву воду» в норвезьких фіордах та біля берегів Таймиру, коли швидкість шхуни «Фрам» падала з 6 до 1,5 вузлів. Він ініціював постановку циклу експериментальних досліджень внутрішніх хвиль, що забирають енергію руху судна, які виконав океанолог Ваґн Вальфрід Екман на початку минулого століття.

Екман писав: «Якщо шар прісної води лежить над солоною, то корабель спричиняє не тільки звичайні видимі хвилі на межі між водою та повітрям, але і буде породжувати невидимі хвилі на межі солоної та несолоної води, я припускаю, що величезний опір, який відчуває корабель, зумовлений роботою, що витрачається на утворення цих невидимих хвиль» [3].

Ці невидимі хвилі можуть з'являтися у вузьких протоках, де змішуються водні маси з різними характеристиками. Наприклад, у Мессінській протоці, що міститься між східним берегом Сицилії та південним берегом Калабрії. Тут змішуються важкі солоні води Іонічного моря, що протікають у нижньому шарі, із легкою водою Тірренського моря, що тече на поверхні. При посиленні поверхневого потоку, може утворитися внутрішня прибійна хвиля заввишки до 60 метрів [4].

У північній частині протоки внаслідок її сильного звуження обидві водні маси розташовуються поруч, тому вихровий рух внутрішньої прибійної хвилі сягає поверхні й викликає поверхневе збурення, що відоме з давніх часів з «Одіссеї» Гомера в легендах про Сціллу і Харібду.

Інша містична історія пов'язана з існуванням так званих «озерних монстрів», що мають за домівки довгі, вузькі озера з крутими схилами у середніх широтах. Таким озерам притаманний чітко виражений сезонний термоклин (перемішаний верхній шар), що призводить до виникнення інтенсивних внутрішніх хвиль. Зважаючи на те, що внутрішні хвилі мають прояви на поверхні, є припущення [5], що ці явища можуть пояснити факти спостережень раптового хвилювання поверхні озера, які приписують загадковій Нессі та її родичам: чорту з озера Лабинкир (Росія), змію Огопо

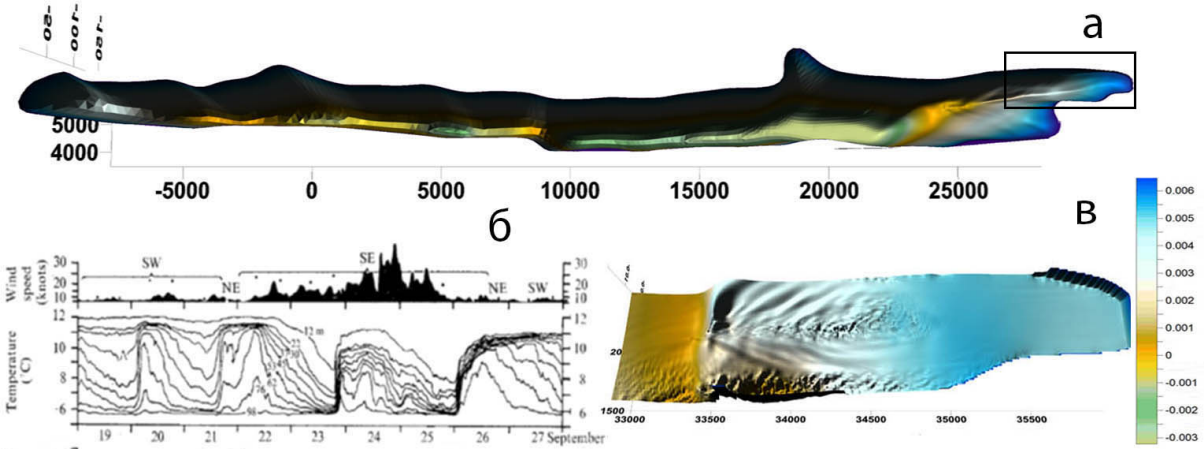


Рис. 2. *Карта глибин озера Лох-Несс (а); спостереження внутрішнього бору з [2] (б); чисельне моделювання вільної поверхні (в)*

з озера Оканаган (Канада), монстру Сельма з озера Сельйорд (Норвегія), монстру Лорі з озера Лох'Рі (Ірландія), монстру з озера Лоен (Норвегія) та іншим. Для того, щоб зрозуміти, яким же чином і за рахунок чого відбувається це раптове хвилювання поверхні озера, було проведене чисельне моделювання [6–8].

Об'єктом моделювання обрано відоме озеро Лох-Несс, яке є природною лабораторією з дослідження внутрішніх хвиль. В озері Лох-Несс були зроблені перші в історії спостереження внутрішніх хвиль в озерах, більше ніж 100 років тому [9]. Саме озеро має довжину близько 38 км, ширину — у середньому 1,6 км і глибину до 230 м (рис. 2, а) і міститься в западині Грейт-Глен — низині у межах тектонічного рову із напрямком, що збігається з домінуючим напрямком дії сезонних вітрів.

У жовтні поверхневий перемішаний шар (термоклин) тут має глибину 40 м і розділяє шари води з температурою 12 та 6 градусів, а період коливань внутрішньої сейші (стоячої внутрішньої хвилі масштабу довжини озера) дорівнює 52 годинам. В озері був зафіксований внутрішній бор (сильнонелінійний фронт), схожий на припливний вал (рис. 2, б, [10]), що формується після дії інтенсивних вітрів, а потім подорожує озером із швидкістю 35 см/с, відбивається від країв озера та поступово затухає. За допомогою чисельного моделювання було відтворено взаємодію інтенсивних внутрішніх хвиль зі східним краєм озера.

У результаті моделювання виявлено цікавий феномен: існування надкритичних (зі швидкостями, що більші за швидкість розповсюдження лінійних хвиль) підводних струменів, які виникають унаслідок кумулятивного ефекту, що пов'язаний із взаємодією внутрішніх хвиль із «ложкоподібною» формою східного торця озера. Надкритичний струмінь формує вихрову пару, що рухається в термоклині з надкритичною швидкістю та формує за собою слід, що можна побачити на поверхні (рис. 2, в). Саме поява цього сліду на поверхні озера і може бути джерелом містичних історій і казкового прибутку у десятки мільйонів фунтів стерлінгів на рік.

Тут доречно згадати фразу, що приписують Альберту Ейнштейну «Процес наукових відкриттів — це, по суті, безперервна втеча від чудес». Убивати монстрів, це, звичайно, справа корисна. Але цим необхідність вивчення внутрішніх хвиль не обмежується.

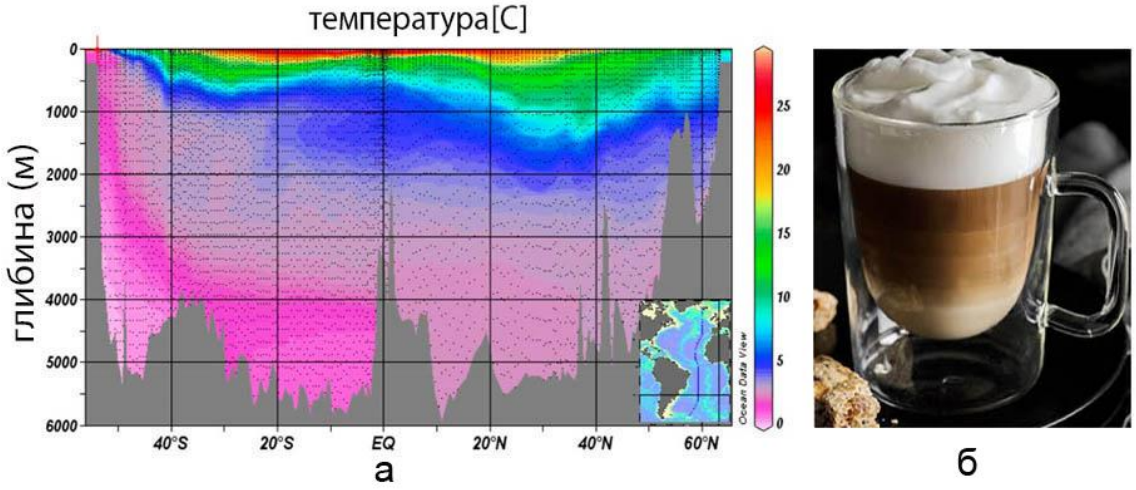


Рис. 3. *Поле температури у вертикальному перерізі у меридіональному напрямку вздовж Атлантичного океану (а); філіжанка кави з вершками (б)*

Внутрішні хвилі — океанічний міксер

Виклик сьогодення — зміни клімату. Наслідки яких — збільшення глобальної температури, зменшення площі льодовиків, підвищення рівня моря, збільшення частоти проявів екстремальних природних явищ, таких як сильні шторми, торнадо та рясні повені. Ключова роль океану у змінах клімату визначається його сильною інерційністю, в силу величезної теплоємності океан є одним із найповільніших компонентів кліматичної системи. Масштаби мінливості океану — роки, десятки і сотні років — значно перевершують характерні масштаби змін в атмосфері. Іншими словами, океан може впливати на клімат у тому діапазоні масштабів, де відбувається антропогенний вплив на клімат.

Якщо ми подивимось на вертикальний переріз температури в океані, то побачимо, що океан схожий на шарувату філіжанку кави з вершками (рис. 3). Результати спостережень показують, що верхній та нижній шари в океані перемішуються дуже слабо, та не відбувається перенесення теплоти й імпульсу по всій глибині океану. Як показують останні спостереження [11], перемішування у відкритому океані розподілено по площі нерівномірно, а локалізовано у вигляді окремих плям поблизу великомасштабних нерівностей морського дна типу континентальних схилів і серединних океанських хребтів. Нині вважається, що ймовірною причиною такої плямистої структури може бути життєвий цикл внутрішніх хвиль, а саме їх генерація і подальша дисипація та руйнування. Іншими словами, передбачається, що внутрішні хвилі є реальним кандидатом на роль дисипаторів припливної енергії, що забезпечує нерівномірне перемішування в глибинному шарі відкритого океану.

Згадаємо про хвилю-хмарочос, що зароджується та помирає у Південнокитайському морі. Як і багато інших внутрішніх хвиль вона народжується завдяки дії припливних баротропних течій, що виникають під час взаємодії Землі в основному з Місяцем. Довга припливна хвиля виникає у неоднорідно стратифікованому океані, рухається та трансформується за взаємодії з неоднорідностями дна, де вона частково руйнується, втрачає частину енергії та перетворюється на низку внутрішніх хвиль із меншими довжинами хвиль, але з великими амплітудами.

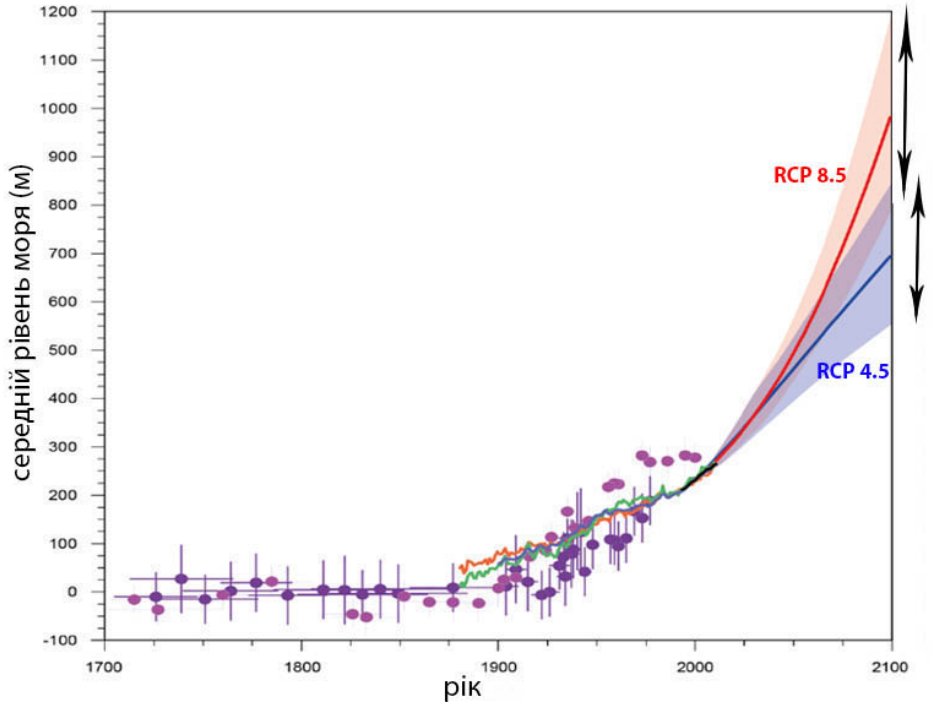


Рис. 4. *Зміни глобального рівня Світового океану за спостереженнями в період 1700–2000 рр. та проєкції змін рівня під впливом змін клімату на 2010–2100 рр. [12]: верхня крива — сценарій RCP8.5, нижня крива — сценарій RCP4.5*

Ці хвилі рухаються зі швидкостями декілька метрів за секунду та проходять тисячі кілометрів, практично не змінюючи своєї форми, аж поки не руйнуються у шельфових зонах, спричиняючи інтенсивне перемішування. Таким чином, енергія припливів перетворюється на енергію внутрішніх хвиль. Енергія припливів передається по каскаду від великих масштабів та, за посередництвом внутрішніх хвиль, переноситься до найменших масштабів і є головним чинником діапікнічного перемішування. Без внутрішніх хвиль океан був би зовсім іншим: поверхневий шар більш теплим, а абісальні води холоднішими. Дія внутрішніх хвиль схожа на перемішування вершків у каві, вони наче уявна «океанічна ложка».

Глобальні кліматичні моделі

Поступово глобальні моделі, що використовуються для прогнозування змін клімату, доповнюються різноманітними уточненнями, такими як урахування впливу хмар, льодового покрову та перемішування від внутрішніх хвиль.

Як бачимо з прогнозів чисельних моделей (рис. 4), рівень океану до кінця сторіччя може збільшитись від 30 до 50 см для сценарію RCP4.5, що є сценарієм середнього рівня викидів парникових газів та від 0,5 до 1 м, для сценарію RCP8.5, що є сценарієм високого рівня викидів. Невизначеність у цих оцінках (смуги навколо кривих прогнозу на рис. 4) спричиняється ще недостатнім розумінням багатьох важливих процесів, таких як урахування впливу льодового покрову, хмар, перемішування від внутрішніх хвиль та інших процесів.

Яким же чином описати та включити ці процеси у глобальні кліматичні моделі, для того щоб адекватно спрогнозувати температуру та зміну рівня океану у майбутньому? Для того, щоб описати процеси втрати енергії та перемішування, що спричинені внутрішніми хвилями в межах глобальних чисельних кліматичних моделей явно, необхідно, щоб роздільна здатність розрахункової сітки була не менше 1 м по вертикалі та в межах 1 км по горизонталі. Це надзвичайно висока роздільна здатність, яка робить розрахунки неможливими для глобальних кліматичних моделей навіть при сучасному рівні обчислювальних потужностей. Таким чином, перемішування в океані, що виникає внаслідок народження, поширення та руйнування дії внутрішніх хвиль, має бути параметризовано та застосовано у глобальних океанічних моделях для розуміння наслідків у глобальній циркуляції та глобального клімату.

Ця задача-виклик є настільки актуальною, що під неї був створений проєкт, який мав об'єднати всі здобутки у розумінні процесів перемішування від внутрішніх хвиль. Він виконувався п'ять років (2010–2015) командою CPT (Climate Process Team), в яку входили вчені з провідних інститутів США та за підтримки таких організацій і програм, як US CLIVAR, NSF та NOAA. Ключовим висновком цього проєкту є те, що результати моделювання глобальними кліматичними чисельними моделями надзвичайно чутливі до просторового розподілу характеристик перемішування та дисипації енергії внутрішніх хвиль [13].

Це загалом означає, що й досі проблема параметризації внутрішніх хвиль, яка б адекватно відображала фізичні процеси в океані, залишається актуальною, а такий елемент, як «внутрішні хвилі», відсутній у пазлі моделювання глобального клімату та потребує подальшого дослідження.

Література

[1] Russell J. S. Report of the committee on waves, Report of the 7th Meeting of British Association for the Advancement of Science / J. S. Russell — John Murray : London, 1838. — P. 417–496.

[2] An extreme internal solitary wave event observed in the northern South China Sea / X. Huang et al. // Scientific Reports, 2016. — Article No. 30041.

[3] Гилл А. Динамика атмосферы и океана: В 2-х т. Т. 2 / А. Гилл; пер. с англ. — М. : Мир, 1986. — 415 с.

[4] Alpers W. Internal waves generated in the straits of Gibraltar and Messina: observations from space / P. Brandt, A. Rubino. In: V. Barale, M. Gade M. (Eds.) // Remote Sensing of the European Seas. — Springer : Dordrecht, 2008.

[5] Thorpe S. Near-resonant forcing in a shallow two-layer fluid: a model for the internal surge in Loch New? / S. Thorpe // Journal of Fluid Mechanics. — 1974. — Vol. 63 (3). — P. 509–527.

[6] Терлецкая Е. Сильно-нелинейные внутренние сейши в удлиненных стратифицированных озерах и феномен «озерных монстров» / Е. Терлецкая, В. Мадерич, И. Бровченко // Прикладна гідромеханіка. — 2011. — Т. 13, № 1. — С. 51–55.

[7] Internal seiches of large amplitude in the elongated lakes and phenomenon of «Loch Ness monster» / K. V. Terletska, V. S. Maderich, I. O. Brovchenko, K. T. Jung // Proceedings of the 23rd International Congress of Theoretical and Applied Mechanics (ICTAM), August 19–24 2012, Beijing, China, FM16-010. — 2 p.

[8] Numerical simulations of the nonhydrostatic transformation of basin-scale internal gravity waves and wave-enhanced meromixis in

lakes / K. V. Terletska, V. S. Maderich, I. O. Brovchenko, K. Hutter. In: K. Hutter (Ed.) // Nonlinear internal waves in lakes. — Springer, Series: Advances in Geophysical and Environmental Mechanics, 2012. — P. 192–276.

[9] Watson E. R. Movements of the waters of Loch Ness, as indicated by temperature observations / E. R. Watson // The Geographical Journal. —1904. — Vol. 24. — P. 430–437.

[10] Mortimer C. H. Some effects of the Earth's rotation on water movements in stratified lakes / C. H. Mortimer // Proceedings of the International Association of Theoretical and Applied Limnology. — 1955. — Vol. 12. — P. 66–77.

[11] Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate / A. F. Waterhouse et al. // Journal of Physical Oceanography. — 2014. — Vol. 44 (7). — P. 1854–1872.

[12] IPCC Fifth Assessment Report (AR5), 2013: https://www.ipcc.ch/assessment-report/ar5/

[13] Melet A. Climatic impacts of parameterized local and remote tidal mixing / A. Melet, S. Legg, R. Hallberg // Journal of Climate. — 2016. — Vol. 29 (10). — P. 3473–3500.

В океані великих даних

Корольов Олексій Миколайович,
кандидат технічних наук

Подивіться на рис. 1. На перший погляд, на ньому зображено звичайну вітрильну гоночну яхту-катамаран чорного кольору. Проте, якщо придивитися до вітрил, то можна побачити, що її грот — не просто полотнище, а жорстка конструкція, схожа на крило літака (насправді вона так і називається — wingsail, парус-крило). Ось яхта виходить в океан і починає прискорюватися, злітаючи на своїх підводних крилах... Крила — і підводні, і поставлене вертикально крило вітрила — дозволяють яхті буквально летіти швидше за вітер: на гонках яхти цього класу можуть досягати швидкостей, які в 2–2,5 рази перевищують швидкість істинного вітру. Рекорд — 47,2 вузла (або 87,4 км/год) — це швидкість скоріше автомобіля, ніж парусної яхти.



Рис. 1. *Тренувальний вихід у море яхти Team Oracle USA*

Що ж дозволяє чорному гоночному катамарану з червоними логотипами Oracle розвивати таку швидкість? Нові сучасні матеріали? Безумовно — яхта практично повністю зроблена з вуглепластика (а точніше — надрукована вуглепластиком на 3D-принтері новозеландської верфі Core Builders Composites), що робить її легкою і міцною одночасно. Вивірена форма крил? І це теж — недарма розробкою і виробництвом окремих вузлів займався Airbus, а вітрило-крило катамарана — це і є поставлене вертикально крило літака з усіма його елементами управління і гідравлічними підсилювачами.

Але чи є ще щось крім ультрасучасних матеріалів, вивіреної форми і відточеної тренуваннями майстерності команди, що дозволяє цій гоночної яхті команди Team Oracle USA ставити свої рекорди? Так. Це «щось» якраз і є внеском Oracle Corporation в успіхи її команди. Що це? Кошти. І — набагато важливіше — сучасні інформаційні технології.

Гоночні яхти-катамарани класу AC50 обов'язково мають поплавці на підводних крилах і вітрило-крило. Гідравлічні елементи управління дозволяють членам еки-

пажу, керуючим яхтою, розвивати такі самі керуючі зусилля, як і гідравлічні системи авіалайнера Airbus A350 XWB. Єдине, що вбило б спортивний дух змагань і чого не може бути на яхті — це двигун. А ще — бортовий комп'ютер.

Так, комп'ютера на борту гоночного катамарана Oracle немає. Все програмне забезпечення, яке вкладає компанія в успіх своєї команди, працює на березі. Із собою в море команда бере тільки датчики. Датчики, які відстежують параметри курсу. Датчики, що вимірюють швидкість і напрямок вітру (тільки на вітрилі-крилі їх більше 300). Датчики, що вимірюють напруги і деформації, що виникають у вітрилі і в корпусі. Датчики серцебиття (і не тільки) членів екіпажу. Датчики, датчики, датчики... Сотні (як правило, більше тисячі одночасно) датчиків пов'язані оптоволоконними кабелями і бездротовими радіоканалами Bluetooth із накопичувачами даних дуже великого об'єму. Кожен вихід яхти в море приносить від 200 до 500 гігабайт даних, які необхідно обробляти для того, щоб на льоту вдосконалювати конструкцію яхти і виробляти оптимальну стратегію екіпажу для тренувань і змагань. Під час перегонів ті самі дані передаються на берег по радіо й аналізуються в реальному часі.

З Атлантичного океану — в океан великих даних

Ще недавно робота з великими обсягами даних вважалася неможливою (або, у кращому випадку, дуже складною). Уже 1944 року Фремонт Райдер, бібліотекар із Весліанського університету в Коннектикуті, опублікував дослідження «The Scholar and the Future of the Research Library», в якому показав, що об'єм книг і періодики, що зберігаються в університетській бібліотеці, подвоюється кожні шістнадцять років, тож до 2040 року бібліотека Єльського університету міститиме приблизно 200 мільйонів томів, які займатимуть 6 тисяч миль полиць і вимагатимуть штат у 6000 бібліотекарів для каталогізації й обслуговування. Очевидно, що робота з такою бібліотекою була б практично неможливою, а зібрані в ній сотні мільйонів томів було б простіше спалити.

Хоча сам термін «великі дані» і не пролунав в статті Райдера, в ній виник інший — «інформаційний вибух». 1961 року Дерек Прайс у своїй книзі «Наука після Вавилона» проаналізував статистику публікацій у наукових журналах і дійшов висновку, що зростання кількості нових журналів (НЕ публікацій!) також носить не лінійний, а експоненційний характер — кількість періодичних видань із будь-якої тематики подвоюється кожні п'ятнадцять років і зростає в десять разів кожні пів століття. Це справді був вибух.

Усвідомлення інформаційного вибуху привело насамперед до появи нових методів стиснення інформації, які дозволяють зберігати більше даних і/або прискорювати їхню передачу. На жаль, інформаційні технології того часу не були готові до такого вибуху, і 1975 року японське Міністерство пошти і телекомунікацій у своєму дослідженні «Перепис інформаційних потоків» дійшло невтішного висновку: інформація вже виробляється в кількостях, у яких її неможливо спожити (проаналізувати й обробити).

Ішли роки. Удосконалювалися пристрої зберігання інформації й алгоритми її стиснення, але інформація чітко слідувала Законам Паркінсона, неминуче займаючи весь наданий їй об'єм. Розвиток обчислювальної техніки (а вона з 1960-х рр. змінилася кардинально) не так допоміг вирішити вже наявні проблеми «інформаційного вибуху», як додав до них нову: тепер уже самі комп'ютери почали породжувати нову інформацію.

1997 року Джон Меші працював провідним ученим компанії Silicon Graphics. Подо-

рожуючи країною, він проводив безліч зустрічей, популяризуючи як програмні продукти його компанії, так і нову революційну концепцію — «великі дані» (Big Data). Напевно, саме він і використав цей термін першим.

Що таке великі дані?

Усупереч поширеній помилці, словосполучення «Big Data» не позначає просто «великі обсяги даних» — термін «великі дані» було введено за аналогією з термінами «велика нафта», «велика руда» тощо. Розмір «великих даних», звичайно ж, теж має значення, однак основну роль тут відіграє не він, а ті нові можливості, які отримує будь-яка людина, уряд або бізнес, які виявились здатними з ними впоратися.

Так що ж таке «великі дані»? Оксфордський словник англійської мови визначає термін «Big Data» як «Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions» — тобто йдеться про надзвичайно великі об'єми даних, які можуть бути піддані кількісному аналізу для виявлення прихованих шаблонів, тенденцій і асоціацій, що належать, у першу чергу, до поведінки і взаємодії людей. У ширшому сенсі, термін «великі дані» характеризує набори даних або занадто великі, або занадто неструктуровані для аналізу традиційними методами.

Який об'єм мають «великі» дані? Очевидно, що це не мегабайти і навіть не гігабайти. Перші оцінки варіювалися від декількох терабайт (смішно, правда?) до десятків, а то й сотень, петабайт (1 Пбайт = = 10^{15} байт). Досить скоро стало зрозуміло, що одного об'єму недостатньо — для того, щоб бути «великими», великі дані повинні задовольняти цілий ряд формальних ознак. Перший набір таких ознак назвали «три V», оскільки він включав у себе Volume (об'єм), Velocity (швидкість — як надходження даних, так і їхньої обробки) і Variety (різноманіття). Пізніше робилися неодноразові спроби розширити цей список іншими англійськими словами, що починаються на букву V — Value (цінність), Veracity (достовірність), Viability (життєздатність), Variability (мінливість), Visualization (візуалізація, візуалізовність) тощо.

Перше V — Volume (обсяг)

Великі дані насамперед є ВЕЛИКИМИ. Їх настільки багато, що їх неможливо вивчати безпосередньо або через статистичні вибірки — можна тільки спостерігати їх як єдине ціле, аналізувати їх і виявляти приховані тенденції, залежності і т. п. Можна порівняти набір великих даних із картинкою з високою роздільною здатністю на екрані монітора: здалеку ми бачимо зображення цілком, але, наблизивши око до поверхні екрана, починаємо бачити, що воно складається з окремих пікселів, проте перестаємо бачити картину в цілому.

Очевидно, що великі обсяги даних вимагають і більших обсягів пам'яті та пристроїв зберігання інформації. Один диск або сервер не здатний вирішити цю задачу: нині у проєкті, що займається збиранням і аналізуванням великих даних, ви швидше за все побачите кластер Hadoop або аналогічну йому систему розподіленого зберігання даних (рис. 2).

Що таке кластер? Це кілька комп'ютерів, об'єднаних між собою високошвидкісними каналами зв'язку, які виглядають для зовнішнього спостерігача як єдиний комп'ютер. Кластери потрібні для підвищення надійності, рівномірного перерозподілу навантаження між окремими комп'ютерами і пришвидшення. Проєкт Apache Hadoop — це вільно поширюваний набір утиліт, бібліотек і програмних платформ для розробки та виконання роз-



Рис. 2. *Фрагмент кластера Wikimedia Foundation*

поділених програм, що працюють на кластерах із сотень–тисяч вузлів. Саме він забезпечує пошук у Yahoo!, Facebook і безлічі інших систем. Безкоштовний продукт із «несерйозною» назвою (засновник проєкту назвав його на честь іграшкового слоненяти) обслуговує інтереси більшої частини компаній-лідерів зі списку журналу Fortune.

Якщо в минулому столітті основною проблемою під час зберігання даних була висока вартість дисків, мікросхем пам'яті та процесорів, то з їх здешевленням найкритичнішим місцем у кластері стали канали зв'язку між комп'ютерами й окремими мікросхемами. Обмежені швидкістю світла, дослідники змушені винаходити нові методи роздільної паралельної обробки даних окремими вузлами обчислювальної мережі.

Ще одна новинка в роботі з даними, продиктована їхнім обсягом, — це застосування розвідувального аналізу даних для підготовки й очищення масиву даних. Американський математик Джон Тьюкі, який запропонував цю концепцію в однойменній книзі 1977 року, протиставляв розвідувальний аналіз даних поширеному в той час методу перевірки статистичних гіпотез і визначав його цілі як аналіз основних властивостей даних, знаходження в них загальних закономірностей, розподілів і аномалій, побудову початкових моделей.

Друге V — Velocity (швидкість)

У XVII столітті британський галантерейник Джон Граунт вивчав парафіяльні списки померлих. Аналізуючи їх, він намагався створити своєрідну систему попередження поширення бубонної чуми, водночас заклавши основи науки, яку ми називаємо демографією. Спостереження, зведені Граунтом у книзі «Природні та політичні спостереження над списками померлих», що витримала чотири видання, принесли йому визнання; за сприяння Карла II він став членом Лондонського королівського товариства. На жаль, своє основне завдання він так і не розв'язав: оновлення даних про смертність в обраному ним джерелі була занадто повільним і відставало від розповсюдження чуми.

Розмірковуючи про швидкість стосовно великих даних, ми маємо на увазі в першу чергу швидкість, з якою вони надходять в систему. Як правило, йдеться про надходження (і, відповідно, оброблення) даних у реальному часі. Хороший приклад тут — та ж яхта Team Oracle USA під час перегонів: дані з її сотень датчиків одночасно надходять у берегову систему на аналіз, і рекомендації екіпажеві на їхній основі повинні вироблятися негайно.

У поєднанні з обсягом, швидкість нерідко породжує серйозні виклики для команд, що працюють із великими даними. Автор цієї статті мав задоволення працювати в стартапі, який займався аналізом англомовної текстової інформації в Інтернеті. Повний дамп усього англомовного сегмента Інтернет купувався за попередньою оплатою у компанії, що займалася його викачуванням, і у вигляді коробки з компакт-дисками з максимально можливою дозволеною на хайвеї швидкістю їхав на машині в датацентр — це виходило швидше, ніж пересилання тих самих даних по самому широкосмуговому з доступних тоді каналів зв'язку. (У 50-х рр. таким способом переда-

валися навіть дуже невеликі об'єми даних, рис. 3.)



Рис. 3. *Дані вирушають у подорож... на вантажівці (диск IBM об'ємом 5 мегабайт)*

Не завжди великі дані настільки повільні: в деяких випадках вони надходять безперервно і з високою швидкістю, і так само вони повинні оброблятися. Це вимагає застосування методів статистичного послідовного аналізу, стійких до короткострокових надходжень нерелевантних даних. Крім цього, в окремих областях (наприклад, у фізиці елементарних частинок) експериментальні дані можуть надходити «сплесками» такого гігантського обсягу і за такий короткий проміжок часу, що не тільки обробка, але і збереження їх можуть бути фізично неможливими. Такі завдання також потрібно вирішувати, працюючи з великими даними.

Третє V — Variety (різноманіття)

Великі дані можуть бути дуже різними: чисельними, текстовими, мультимедійними. Для великих даних у цьому полягає їхня перевага, а не недолік. Якщо різноманіття вхідних великих даних велике, то їхня аналітична обробка нерідко включає в себе злиття даних із різнорідних джерел. При цьому різні дані, що належать до одного об'єкта, географічної позиції тощо, об'єднуються, усувається їхнє дублювання і підвищується достовірність.

Різноманіття великих даних веде до якісного підвищення складності вирішуваних завдань. Класичні задачі статистики припускають вибірки з великої кількості об'єктів n, кожен з яких окремо характеризується відносно малою кількістю змінних p, формуючи тим самим вузьку і високу матрицю структури даних розмірності $n \times p$.

У великих даних об'єкт може бути чим завгодно — дво- або тривимірним зображенням, повідомленням із соціальних мереж, аудіоданими, відеороликом... У багатьох випадках різні об'єкти вибірки характеризуються різною кількістю змінних, змушуючи тим самим або розпрощатися з «прямокутно-матричним» поданням даних, або ж перетворювати їх на матрицю штучно (наприклад, заповнюючи «відсутні» змінні нулями) — і тим самим втрачати оригінальну структуру даних.

Навіть якщо кількість змінних, що характеризують об'єкт, стала, вона може виявитися занадто великою. Наприклад, матриця, що описує комунікації в мережі з n вузлів, кожен з яких пов'язаний з усіма, швидше за все виявиться як мінімум квадратною, розмірності $n \times n$. У проєкті «тисячі геномів» кількість описаних індивідуумів n незначна (1092), проте геном кожного з них описується дуже великою (близько 38 мільйонів однонуклеотидних поліморфізмів) кількістю змінних p.

А ще великі дані завжди неповні. Як би багато ми не зібрали б великих даних, ще більше їх залишається незібраними — тому для роботи з ними замість дескриптивних статистик традиційного аналізу даних необхідно використовувати індуктивні статистики та нелінійну ідентифікацію систем.

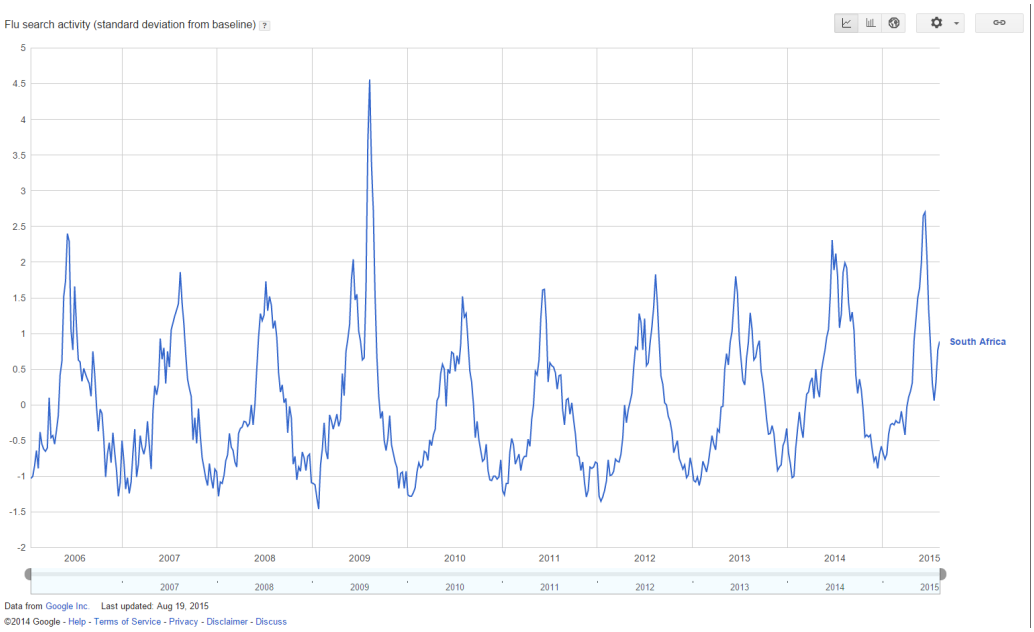


Рис. 4. *Динаміка «грипозних» пошукових запитів для Південної Африки*

І ще одне V наостанок — Veracity (достовірність)

1936 року американський журнал «The Literary Digest» провів опитування серед своїх десяти мільйонів передплатників. Темою цього опитування були президентські вибори, фаворитами яких стали губернатор-республіканець штату Канзас Альфред Лендон і Франклін Делано Рузвельт, кандидат від демократів і діючий президент Сполучених Штатів Америки. Всі попередні такі опитування, що проводилися журналом в 1916, 1920, 1924, 1928 і 1932 роках, правильно передбачали результати президентських виборів. Тож багато хто серйозно поставився до того, що близько двох з половиною мільйонів передплатників «The Literary Digest» переконливо «віддали» перемогу на прийдешніх виборах Лендону. Але переміг, як ми знаємо, Рузвельт. Що ж сталося? Чому великих за тодішніми мірками (2,4 мільйона опитаних — це дуже багато) даних виявилося недостатньо, і чому в той же час значно скромніше за обсягом вибірки (50 тисяч опитаних) опитування Американського інституту громадської думки Джорджа Геллапа досить точно передбачило перемогу Рузвельта? Причина — у джерелі даних опитування і їхній достовірності: якщо Геллап складав статистично репрезентативну вибірку, то журнал просто опитував своїх передплатників. Але хто міг продовжувати підписуватися на не самий дешевий журнал під час Великої Депресії? Як виявилося — виборці Лендона.

Майже через століття, 2008 року корпорація Google запустила сервіс «Google Flu Trends», який мав передбачати епідемії грипу на основі пошукових запитів — тобто вже справжніх великих даних (рис. 4). В основі проєкту лежала цілком правдоподібна ідея: люди, які захворюють на грип, шукають інформацію про грип, і їхні пошукові запити повинні були це відображати. Систему навчили на розподіленій за регіонами інформації про епідемії грипу та хворих, а також на пошукових запитах із США з 2003 по 2007 роки, яких щотижня було близько 50 мільйонів. Протягом навчання виділили 45 «найбільш грипозних» пошукових запитів. Початкові оцінки припускали точність передбачення в 97 %. Очікувалося, що за 7–10 днів до початку епідемії вона вже буде передбачена, що дозволить вжити всіх необхідних профілактичних за-

ходів. Насправді ж модель почала «передбачати» епідемії грипу або високий рівень епідемічної небезпеки навіть тоді, коли в реальності нічого такого не відбувалося. Як і у випадку опитування «The Literary Digest», слабкою ланкою для Google виявилися його власні користувачі, які були нездатними діагностувати у себе правильно грип і тому шукали інформацію про його симптоми навіть тоді, коли в ній не було необхідності. Після серії гарячкових дій з «уточнення» моделі, які увінчалися відчайдушною спробою підвищити точність за рахунок включення до вибірки крім пошукових запитів ще й повідомлення агентств новин, Google 2013 року закрив доступ до результатів роботи «Google Flu Trends».

Проблема достовірності великих даних має рішення — той же Google продовжує працювати над епідемічними передбачувальними моделями, хоча і без сенсаційних заяв. Усе, що потрібно — пам'ятати, що великі дані можуть бути зашумлені або частково відсутні, і відповідно коригувати методики роботи з ними.

Решта V і зовсім не V

Хоча додаткові V-критерії і складалися перш за все в маркетингових цілях, вони також можуть бути корисні. Наприклад, критерій Value (цінність) передбачає, що обробка великих даних повинна бути економічно доцільною, а Visualization (візуалізація) відображає той факт, що великі дані зручно аналізувати у графічному поданні.

У той же час, наявні і досить важливі властивості великих даних, єдиним недоліком яких є відсутність в англійській мові відповідного слова на букву V. Слід особливо виокремити два з них: орієнтованість великих даних на завдання машинного навчання і їх породження «цифровим слідом».

Цифровий слід (також цифрова тінь або цифровий відбиток) — це сукупність інформації про дії людини, що здійснюються в Інтернеті, яка охоплює все від спілкування в соціальних мережах до епізодичних відвідувань сайтів. Як результат, велика частина населення земної кулі сама безперервно і безкоштовно формує великі дані про самих себе. Аналітична обробка таких великих даних дозволяє вирішувати найрізноманітніші завдання — від маркетингових досліджень і збору демографічної інформації до шпигунства і соціальної інженерії. Пів століття тому для отримання подібних результатів досліднику або бізнесу треба було б проводити тривалі дорогі масові опитування з анкетуванням, спецслужбі — закликати на службу половину населення країни для того, щоб стежити за іншою половиною.

З машинним навчанням ситуація цікавіша і заслуговує на окрему розповідь. Машинним навчанням називають технології штучного інтелекту, завдяки застосуванню яких замість прямого програмування розв'язання задачі комп'ютером алгоритмічною мовою дослідник може навчити комп'ютер розв'язуванню певного класу задач на основі вибірки відомих розв'язків задач із цього класу (наприклад, інтелектуальний фільтр спаму в електронній пошті може брати як навчальну вибірку всі листи, помічені або НЕ помічені користувачем як спам, і на основі навчання будувати гіпотези про те, з якою ймовірністю наступний лист виявиться спамом із точки зору одержувача).

Що ж робити з великими даними?

Новачок, що вирішує зайнятися роботою з великими даними, нерідко чує: «В ідеалі для цього треба закінчити сильний технічний виш, відмінно знати матаналіз, лінійну алгебру (матриці і т. д.), дискретну математику, математичну статистику, теорію ймовірностей, теорію алгоритмів то-

що.» Чи дійсно це так? Безумовно. Але чому це так? Для відповіді на поставлене питання спробуємо подивитися, що саме відбувається з великими даними після їх збору.

Отже, великі дані можуть надходити з різних джерел, з різною швидкістю і в різній формі, мати різний ступінь достовірності. Їх збір може відбуватися сам собою («цифровий слід») або ж вимагати додаткових зусиль (дані фізичних експериментів і астрономічних спостережень). Так чи інакше, зібрані дані потрапляють у пристрій зберігання, і постає питання: що зробити з ними далі, щоб одержати з них користь?

Випас даних

Чи можна пасти великі дані? Виявляється — так: терміном «випас даних» (data wrangling) в англомовній літературі називають процес зведення «сирих» даних до вигляду, більш придатного для аналізу. Сенс аналогії очевидний: у «стаді» великих даних жоден «цифровий ковбой» не зможе зайнятися окремо кожною «коровою», але може пасти все «стадо», поки воно не нагуляє цифровий жирок.

Кожне «стадо» даних вимагає своєї стратегії випасу. Не можна переглянути і відредагувати вручну мільярди записів даних, тому застосовуються різноманітні автоматичні фільтри й обробники. Текстові дані проходять через лінгвістичні процесори, які виділяють ключові слова, граматичні ознаки і навіть смисловий зміст. Графічні об'єкти проходять через системи розпізнавання образів. Різні способи представлення географічної інформації (через поштову адресу, через широту і довготу, через IP-адресу комп'ютера тощо) зводяться в єдину універсальну форму. Занадто точні дані перетворюються в нечіткі (fuzzy). Зайві дані виключаються з вибірки, відсутні — додаються шляхом екстраполяції.

Візуалізація

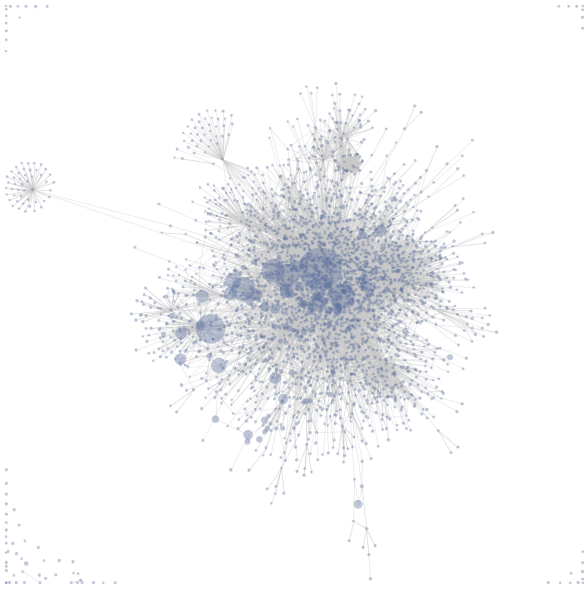


Рис. 5. *Візуалізація структури даних Wikimedia за допомогою пакету Prefuse*

Часто візуалізацію великих даних називають першим етапом їхнього аналізу. Як ми вже говорили вище, неможливо переглянути всі великі дані окремо, проте можна подивитися на них, як на єдине ціле, і побачити цілісну картину із прихованими закономірностями. Або не побачити. Вибір правильного методу і параметрів візуалізації сам по собі може бути нетривіальним творчим завданням, розв'язання якого дозволить значно прискорити наступні етапи аналізу.

Маючи справу з результатами візуалізації, дослідник може не відразу виявити шукані закономірності — тому візуалізація повинна бути інтерактивною, забезпечувати можливість масштабування і зміну кута зору (як, наприклад, візуалізація за допомогою пакету Prefuse на рис. 5).

Зниження розмірності

Великі дані багатовимірні, тому зниження їх розмірності є досить поширеною частиною роботи з даними. Зниження розмірності очевидно потрібне для візуаліза-

ції і може бути корисним для ефективного зберігання великих даних, потреба в ньому виникає і в задачах класифікації.

Класичним прикладом способу зниження розмірності даних за мінімальних втрат інформації є метод головних компонент, запропонований Карлом Пірсоном 1901 року. Нерідко перші дві-три головні компоненти візуалізують для того, щоб ідентифікувати структуру даних.

Розрідженість і регуляризація

Великі дані можуть бути розрідженими, тобто значна частина координат у багатовимірних векторах, що представляють об'єкти, виявляється заповненою нулями. Хороший приклад таких розріджених даних — подання текстів у вигляді векторів, в яких окремим координатам відповідає кількість входжень у текст документа відповідних ключових слів. Хоча в цілому в мові можуть бути десятки, якщо не сотні тисяч слів, у будь-якому конкретному документі (якщо тільки це не тлумачний словник) велика частина цих слів не буде використовуватися, а відповідні їм координати у векторах будуть нульовими.

Багатовимірність таких розріджених великих даних виявляється таким чином багатовимірністю надлишковою, і завдання зниження їхньої розмірності напрошується само собою як під час зберіганні, так і під час аналізу. Класичними шляхами розв'язання цього завдання є різноманітні форми регуляризації. Типовими представниками методів регуляризації є LASSO (least absolute shrinkage and selection operator) і регуляризація Тихонова, задача ж у цілому відкриває широкий простір для математичних досліджень.

Оптимізація

Задачі оптимізації (нерідко — досить складні) виникають під час розв'язання різноманітних статистичних задач над великими даними — зокрема, при їх регуляризації, про яку ми говорили вище. Складні моделі з великою кількістю прихованих шарів, що виникають, наприклад, у ході побудови нейронних мереж і випадкових марковських полів (машин Больцмана), зазвичай потребують розв'язання задач неопуклої оптимізації.

Вимірювання відстаней

Здійснюючи аналіз багатовимірних даних, необхідно вміти знаходити відстань між окремими «точками» в багатовимірному метричному просторі об'єктів. У найпростішому випадку така відстань може обчислюватись за декартовою формулою для лінійного багатовимірного простору.

Наступна відстань має багато назв: «Відстань міських кварталів», «Манхеттенська відстань», «метрика прямокутного міста», «метрика таксі»... Манхеттенська відстань між двома точками дорівнює сумі модулів різниць їх координат — так, наче обидві точки містяться на перехрестях вулиць, що перетинаються під прямими кутами і розбивають усе місто на прямокутну сітку (як у Нью-Йорку на Манхеттені), а водій таксі шукає найкоротшу відстань між ними (рис. 6).

Але у світі транспортних засобів існує не тільки таксі. Ще одним способом визначити відстань між багатовимірними об'єктами є так звана «французька залізнична метрика». Така назва пов'язана з тим, що в минулому всі залізниці у Франції виходили наче промені із загального центра — Парижа. Знаходячи найкоротший шлях між двома станціями в такій залізничній мережі, ми повинні переконатися, чи не перебувають вони на одній дорозі — якщо так, то ми зможемо визначити відстань безпосередньо між ними. Якщо ж станції належать до різних залізничних гілок, то відстань між ними стане сумою відстаней від

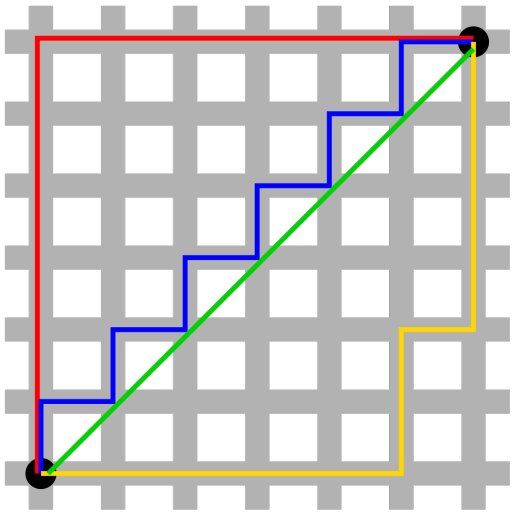


Рис. 6. *Манхеттенська та евклідова відстані*

кожної з них до Парижа. До речі, якщо ми забороняємо нашим «поїздам» ходити безпосередньо між станціями, вимагаючи натомість кожен раз спочатку заїжджати в «Париж», то це вже буде «метрика британської залізниці», вона ж «метрика пошти» (Post Office metric).

Але повернемося з екзотичних курйозів у реальний світ і закінчимо наш короткий і неповний екскурс у світ різноманітних відстаней на відстані Хеммінга. Що це таке? Уявімо собі, що наші n-вимірні вектори — це слова довжиною n букв, і ми граємо у просту гру: беремо перше слово (координати першої точки) і дивимося, скільки букв у ньому потрібно замінити на інші, щоб отримати друге слово. Відстань Хеммінга між «КИТ» і «КІТ» дорівнює 1, а між «КІТ» і «ПЕС» — 3. Що робити, якщо одне слово коротше? Перейти від відстані Хеммінга до відстані Левенштейна (відомою також, як «редакційна відстань» або «дистанція редагування»), в якій відстань між двома словами — це мінімальна кількість операцій із видалення літери, вставлення нової літери і заміни однієї літери на іншу, які потрібно виконати над першим словом, щоб отримати друге. До речі, якщо додати ще одну «редакторську» операцію — поміняти місцями дві сусідні букви — то ми отримаємо вже відстань Дамерау – Левенштейна, яка використовується в біоінформатиці.

У деяких випадках замість відстаней зручніше виявляється використовувати міри подібності — безрозмірні величини в діапазоні від 0 (подібність відсутня) до 1 (повна подібність). Прикладом такої міри може бути косинус кута між двома векторами.

Глибоке навчання

Метою глибокого навчання (deep learning) є пошук прихованих і складних структур. Нерідко воно має багаторівневу структуру для виявлення ієрархічного представлення даних. Модель для такої багатошарової структури може перевершувати за розмірністю вихідні дані — в цьому випадку глибоке навчання доводиться поєднувати з різними способами зниження розмірності даних.

Найбільш складною проблемою у глибокому навчанні на великих даних є неможливість вручну розмітити всі дані (наприклад, якщо у нас є трильйон електронних листів, то ми не можемо вручну позначити кожне як спам або не спам). Для її вирішення застосовуються як різноманітні способи автоматичної розмітки вхідних даних, так і перехід від машинного навчання з учителем до машинного навчання без учителя.

Послідовне навчання

Як ми вже говорили вище, однією з характеристик великих даних є швидкість. Якщо ця швидкість висока, то ми не можемо працювати зі статичним сформованим сховищем даних і змушені перейти від звичайного навчання до послідовного (sequential learning, incremental learning, online learning).

Трапляється й так, що занадто великою виявляється не швидкість, а об'єм. Хоча при цьому всі дані доступні одночасно, по-

слідовне навчання виявляється вигідним уже з інших причин (наприклад, у силу неможливості помістити в пам'ять відразу всі дані).

Навіщо потрібні великі дані?

Світом правлять об'єктивні закони природи — такі, наприклад, як принцип Ле Шательє – Брауна («Система, виведена зі стійкої рівноваги, прагне до неї повернутися»), закон Мерфі («Якщо якась неприємність може статися, то вона неодмінно станеться») або «бритва Хенлона» («Не потрібно пояснювати таємною змовою те, що можна пояснити явною некомпетентністю»). З них великих даних стосується в першу чергу так зване «узагальнене правило номер 34»: «Великі дані застосовні скрізь. Жодних винятків».

І справді, великі дані працюють усюди. Вони дозволяють проводити адресні рекламні компанії (заздалегідь вивчивши інтереси їхнього адресата за його «цифровим слідом») і об'єктивно оцінювати їхню ефективність (без нудних опитувань, у яких ніхто не хоче брати участь).

Великі дані допомагають інтернет-магазину як порекомендувати покупцеві те, що йому швидше за все потрібно (перевірено на мільйонах інших покупок!), так і заздалегідь замовити поповнення полиць на складі тими товарами, які ще не продані, але точно будуть (причому саме на складі, найближчому до потенційних покупців).

Великі дані шукають екзопланети у далеких зірок і розкривають таємниці того, як улаштований Всесвіт на мікрорівні, у прискорювачах елементарних частинок.

Великі дані передбачають епідемії й оцінюють ефективність повсякденної діяльності закладів охорони здоров'я.

Великі дані розкривають причини злочинності і передбачають масові заворушення. Великі дані знаходять те, що загрожує навколишньому середовищу, і одночасно дозволяють відрізнити реальні загрози від уявних, пов'язаних із природними явищами.

Великі дані вибирають колір і дизайн кузова нового автомобіля, який всім захочеться купити, і оптимізують дорожній рух так, щоб цей автомобіль їхав, а не стояв у пробках.

Великі дані сидять за кермом автомобілів з автопілотом і займаються обслуговуванням клієнтів у чатах.

Великі дані роблять усе. Кожним своїм кроком, кожною дією сучасна людина породжує нові великі дані, і кожен із нас може знайти нові способи їх використання на благо суспільства.

Подивимося, наприклад, на кілька проєктів, пов'язаних з великими даними, які вже працюють в Україні.

Великі дані на замовлення: «Київстар» і «Vodafone Україна»

Коли Джордж Оруелл створював свою антиутопію «1984», єдиним технічним рішенням для тотального стеження тоталітарної держави Великого Брата за своїми громадянами, яке він зміг придумати, стали відеокамери «телекранів» (аналога телевізорів), через які співробітники тамтешньої держбезпеки могли випадковим чином вибірково підглядати за громадянами. Втілення такої системи в реальному житті вимагало б від держави закликати на службу цілу армію шпигунів, єдиним завданням яких було б сидіти перед екранами і виглядати крамолу.

Тоді нікому не могло прийти в голову, що в майбутньому громадяни будуть носити з собою в кишені або сумці своїх власних персональних шпигунів і донощиків — мобільні телефони. У будь-який момент часу, залишаючись на зв'язку, ми одночасно по-

відомляємо стільниковому оператору, де саме ми перебуваємо. Навіть якщо телефон не має GPS-чипа (а такі ще існують?), він все одно може бути точно локований по тому, у трикутнику між якими трьома найближчими базовими станціями він перебуває, а по різниці рівнів сигналів можна доволі точно визначити і його позицію всередині цього трикутника.

Стеження за конкретним абонентом без санкції суду незаконне — але немає нічого незаконного у зборі всієї інформації про переміщення всіх абонентів, якщо вона анонімізована (тобто, не має прив'язки до конкретних людей). Оператори стільникового зв'язку використовують ці дані у своїй роботі (перш за все, для вибору місць установки нових базових станцій, відкриття нових магазинів тощо), а також надають усім бажаючим, готовим заплатити. І таких бажаючих є безліч.

Великі дані України починаються насамперед саме тут. І дані це дійсно великі: одні тільки абоненти «Vodafone Україна» генерують за добу близько 400 терабайт великих даних про себе, сумарні ж обсяги даних, за словами представників компанії, обчислюються петабайтами.

Поряд із платним доступом «Vodafone Україна» підтримує проєкт «Big Data Lab», в межах якого відкриває дослідникам масив із частиною своїх анонімізованих великих даних, а потім підтримує найвдаліші проєкти. Аналогічний проєкт «Київстару» називається «Big Data School». Крім проєктів, ці компанії підтримують і окремих дослідників, пропонують їм роботу.

Дихання великого міста: громадський транспорт і дорожній рух

Однією з основних проблем управління великим містом є те, що в місті живуть люди, і керувати ним треба так, щоб у першу чергу від цього ставало кращим життя людей. Скільки людей живе в Києві і де? Скільки в ньому туристів, і звідки вони? Скільки людей приїжджають кожен день до Києва з міст-супутників? Скільки переміщається щодня між лівим і правим берегами? Жодні переписи й опитування не можуть дати вичерпних відповідей на ці питання — їх можна знайти тільки у великих даних, пройшовши по «цифровому сліду» кожного киянина. КМДА вдалася до допомоги компанії «Київстар».

Першою метою для аналізу цих великих даних став київський громадський транспорт. Із січня по вересень 2015 року відбувалась побудова транспортної моделі міста, яка дозволила розрахувати, з якою швидкістю і з якою кількістю пересадок пасажир може дістатися в різні частини міста з найбільш завантажених районів. Результат у більшості випадків виявився незадовільним. На основі моделі були створені нові маршрути, що скорочують граничний внутрішньоміський маршрут до 60 хвилин і тільки з однією пересадкою, розмічені виділені смуги для громадського транспорту, установлено нові світлофори.

Аналіз переміщень туристів по місту дозволив (на основі того, що їм найцікавіше) в реальному часі коректувати інформацію на офіційному сайті *Kyiv.Travel* і в путівниках по Києву, формувати персональні програми лояльності по картках туристів. У найближчому майбутньому міська влада планує використовувати великі дані для управління світлофорами: побудувати і підтримувати динамічну транспортну модель, яка сама буде інтелектуально керувати всіма міськими світлофорами, щоб зменшити пробки і прискорити переміщення містом.

Гостинна Одеса: народження туристичної перлини

Загальновідомо, що в останні роки істотно зросла кількість туристів в Одесі й Одеській області. Але «загальновідомо» — не

завжди означає, що відомо щось конкретне. Точної офіційної статистики не було: дані, що формуються туристичними компаніями й інформація від прикордонних пунктів про вхідний потік закордонних туристів іноді могли відрізнятися на кілька порядків. Рішенням, як і у випадку Києва, стало застосування великих даних від мобільного оператора. Основною відмінністю Одеської області від Києва стала менша географічна точність даних: поза містом стільники мобільного зв'язку мають більший розмір. Через це й оперативність отримання інформації стала менш важливою. Для обласної влади двічі на рік — у жовтні (після закінчення сезону) і січні-лютому (за підсумками минулого року) — «Київстар» готує дані по місяцях і регіонах, а також надає розбивку на внутрішніх та іноземних туристів, де «турист» — це той, хто провів у межах Одеської області більше доби.

Великі дані одеського туризму публікуються у відкритому доступі. Їхній аналіз дозволив перемаршрутизувати транспортні потоки — зокрема, домовитися з «Укрзалізницею» про запуск поїзда Київ – Ізмаїл на південь Одеської області в напрямку популярних курортів Затока і Сергіївка, куди раніше можна було потрапити тільки на електричці або по автомобільній дорозі.

Galactic Exchange: лопата для великих даних

12 травня 1848 року вулицями Сан-Франциско йшов чоловік. Він ніс над головою пляшку, наповнену золотим пилом, і кричав: «Золото! Золото! Золото з річки Американ-Рівер!» Цього чоловіка звали Семюель Бреннан (Samuel Brannan), він був видавцем газети і торговцем. Перш ніж почати свою екстравагантну рекламну кампанію, він скупив усі кирки і лопати в окрузі. Парадоксально, але саме такі люди, а не шукачі золота, й заробили більшу частину грошей за часів каліфорнійської «золотої лихоманки» (рис 7).



Рис. 7. *Золотошукач з киркою на річці Трініті-Рівер, Каліфорнія*

Що є аналогом кирки або лопати золотошукача для «великих даних»? Комп'ютери? Частково. Висловлюючись фігурально, у людей, що займаються глибоким навчанням на великих даних, спільна лопата із сучасними золотошукачами, що майнять криптовалюти: і ті, й інші використовують цілі ферми з відеокарт, на яких крутяться їхні складні паралельні багатопотокові обчислення. Але крім кирки GPU в розпорядженні людей, що займаються великими даними, є і лопата — Hadoop, або інший кластер для зберігання даних. І ось тут вступає у гру наш герой — харківський стартап Galactic Exchange.

«Лопата», яку пропонує Galactic Exchange, називається ClusterGX. Це безкоштовний у використанні (при обсязі даних до 1 ТБ) продукт із відкритим кодом, який дозволяє зібрати Hadoop-кластер за п'ять хвилин і працює у хмарі на Windows, Linux або macOS. Харківський офіс стартапу працює в тісному контакті з Харківським національним університетом та Харківським національним університетом радіоелектроніки.

Leantegra: Маленький Великий Брат

Якщо стільникові оператори збирають великі дані із «цифрових слідів» своїх або-

нентів як відходи від своєї основної діяльності, то київський стартап Leantegra займається тим самим більш активно, використовуючи популярні зараз технології «Інтернету речей» (IoT, Internet of Things). Крім інформації від смартфонів (Bluetooth, Wi-Fi, радіосигнали), Leantegra використовує і свої власні бездротові пристрої: «Маяк» PowerBeacon і радіопозначки PowerMote. Бездротова інформація збирається для великого комерційного об'єкта (наприклад, супермаркету або торговельно-розважального центру) і формує великі дані, на основі яких розв'язуються найрізноманітніші задачі: контекстний маркетинг, прив'язаний до координат конкретного користувача, управління торговим залом, індивідуальна навігація...

Знання — сила. Клієнти Leantegra (а серед них, наприклад, київський «ArtMall») отримують такі детальні великі дані про своїх клієнтів, наче ті переміщуються та купують товари не на місцевості, а в інтернеті. Виграють від цього і клієнти: так, наприклад, покупці в «Ашані» на Петрівці, якщо на їхньому смартфоні встановлено додаток гіпермаркету, можуть отримувати інформацію про діючі акції та знижки в той момент, коли проходять повз полиці з відповідними товарами — майже як у фільмі «Minority Report».

Відкриті великі дані: від розумних міст до розумної країни

Не всі знають, але в Україні вже роки зо два як діє закон, згідно з яким громадяни України повинні мати відкритий доступ до великих даних своєї держави у формі «відкритих баз даних, придатних для автоматизованої обробки без попереднього втручання людини». Розпорядники інформації — тобто, державні органи — зобов'язані публікувати точну і повну інформацію і постійно її оновлювати. Дані публікуються на Єдиному порталі відкритих даних (`http://data.gov.ua/`). Усього в законі вказано 290 масивів даних, які повинні зробити загальнодоступними міністерства.

Відкриті дані можна вільно копіювати, публікувати, поширювати і доповнювати, зокрема — в комерційних цілях, а доступ до них, крім прозорості роботи влади, служить для боротьби з корупцією і надання допомоги українським компаніям і бізнесу. Проте основне досягнення — переведення всіх даних в єдиний формат і форму, придатні для машинного оброблювання. Тепер не треба буде збирати по крихтах в Інтернеті розклади електричок або видряпувати з вордового файлу з помилками статистичну таблицю.

Плавання тільки починається...

Вище ми зупинилися тільки на двох українських стартапах, які працюють із великими даними — насправді їх десятки, якщо не сотні. Інвестиції, залучені успішним стартапом, вимірюються сотнями тисяч або мільйонами доларів, прибутки у цій галузі ще вище. І, що важливо не менше грошових потоків, постійно з'являються розв'язки для нових задач, які раніше або вважали важковирішуваними, або взагалі не намагалися розв'язувати. Майбутнє приходить до нас. Чим більше відкритих великих даних буде доступно (а їх з'являтиметься в Інтернеті все більше і більше), тим вища буде можливість у кожного бажаючого спробувати розв'язати якусь свою проблему з великими даними. В океані великих даних відкриті всі шляхи. Так, іноді плавання ними вимагає знань і досвіду, іноді на них підстерігають підводні камені й айсберги, але результат того вартий.

Леонардо да Вінчі

Жук Ярослав Олександрович,
Київський національний університет імені Тараса Шевченка

2 травня 2019 року виповнилося 500 років із дня смерті одного з найвидатніших людей епохи Середньовіччя — Леонардо да Вінчі. За пів тисячоліття, що минуло відтоді, світ людей дуже сильно змінився як в матеріальному, так і в моральному розумінні. Але можна стверджувати абсолютно точно, що без Леонардо він був би іншим.

За одним із висловів, знайденим серед давньоєгипетських текстів, «… існує дещо, перед чим відступають і байдужість зірок і довічний шепіт хвиль — це діяння людини, що відбирають у смерті її здобич». Ці слова

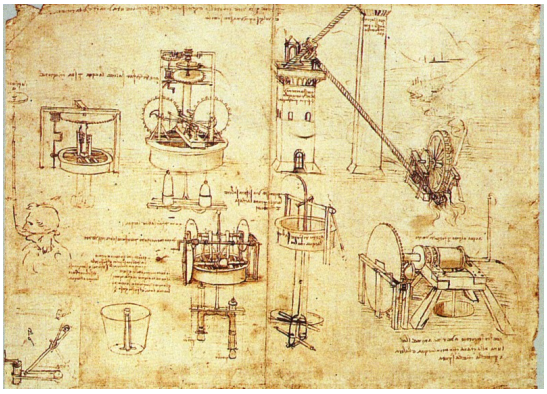


Рис. 1. *Відцентровий насос*

можуть бути повністю віднесені до спадку Леонардо да Вінчі, який був типовим представником Ренесансу, людиною багатьох талантів. Здебільшого він відомий широкому загалу як митець, хоча був також науковцем, інженером, винахідником, анатомістом, механіком і, певною мірою, математиком. В усякому разі, математика і механіка виразно виступають практично у всіх його роботах. Серед них, перш за все, слід зазначити такі твори, як «Таємна вечеря», «Мона Ліза», «Вітрувіанська людина» і «Поклоніння волхвів». Віддамо ж належне багатогранному італійському генію, коротко наголосивши на його внеску в математику і механіку.

Інженер-механік

Інженерний геній Леонардо добре відомий у світі. Йому присвячені численні дослідження, фільми, виставки і телепрограми, кількість яких поступається хіба що дослідженням його художнього спадку. Нам важко уявити глибину його інтелекту, особливо враховуючи той факт, що Леонардо жив 500 років тому, коли не було електрики, надійних засобів вимірювання часу і навіть мостів через водні перешкоди.

Як і багато творчих людей того часу, хто не міг заробити достатньо грошей для власного проживання, да Вінчі змушений був шукати інші способи підтримати себе. Він це робив, використовуючи власний художній геній у таких царинах, як військова інженерна справа, розробка зброї, архітектура і навіть будування каналів. Деякий час він отримував платню, працюючи на посаді військового інженера у правителя Мілана. Проте найбільш цікавим є те, що цю роботу, як і багато інших, він робив талановито! Леонардо часто називають «генієм», хоча як би його не називали, да Вінчі мав хист до застосування своїх художніх здібностей у

надзвичайний спосіб. Він хотів створювати «нові машини» для «нового світу».



Рис. 2. *Багатоствольний кулемет*

Добре відомо, що Леонардо називав війну «тваринним божевіллям» і виступав проти неї. Проте він не міг залишатись осторонь, коли його рідна Італія часів Ренесансу, здавалось, перебувала у стані постійної війни. Да Вінчі вважають розробником багатьох видів тогочасної зброї: гранат, мортир, ракет і багатоствольних "кулеметів" (рис. 2). Протягом років своєї роботи для правителя Мілана Людовіко Сфорца, що тривала з 1482 до 1499, Леонардо займався багатьма проєктами від споруд і устаткування до вивчення механіки, геометрії та створення літальних апаратів (рис. 3). Зокрема у період 1485–1490 років да Вінчі присвятив свої зусилля розробці підводного човна, танка й іншого складного озброєння. Серед багатьох інших слід зокрема згадати проєкт 1495 року — детальний план триколісного заводного самохідного «автомобіля», який також можна вважати першим «роботом».

Протягом сімнадцяти років свого «інженерного» міланського періоду Леонардо створив лише шість мистецьких творів, серед них — «Пані з горностаєм» (рис. 4) і «Таємна вечеря» (рис. 5), оскільки через властивості свого характеру й особливості завдань весь час «перестрибував» з одного предмета на інший. Художній талант Леонардо дозволяв йому «записувати» свої механічні ідеї з небувалою для тих часів зрозумілістю, що навіть у наш час дозволяє створювати на основі його малюнків готові креслення, за якими можна будувати працюючі моделі. Цей талант дозволяє людству і нині оцінити його геній. Тільки недосконалість середньовічних технологій і рівень тогочасного матеріалознавства стали на заваді втіленню цих проєктів у робочі прототипи.

Отже, чи можна назвати да Вінчі механіком або інженером? Для цього, напевно, корисно звернутись до самого означення механіки: механіка — від грецького μηχανη, μηχανικη — винахідливість, мистецтво побудови, виготовлення знарядь, споруд. Ньютон розділяє цю науку на раціональну (умоглядну) і практичну: «До практичної механіки належать усі ремесла та виробництва, що називають механічними, від яких взяла назву і сама наука... Раціональна механіка — це вчення про рухи, що викликаються будь-якими силами, і сили, які є необхідними для спричинення будь-яких рухів, точно, послідовно викладене і доведене».

Таким чином, Леонардо да Вінчі, який не займався розробкою теоретичних основ механіки, можна назвати геніальним винахідником та інженером-механіком. Він був, напевно, першим інженером-проєктувальником, що створив підхід до розуміння як працюють машини.

Да Вінчі любив механіку, яку називав «раєм математичних наук». Глибокі знання того, як функціонують окремі частини, дозволяли Леонардо модифікувати та комбінувати ці частини у такий спосіб, щоб покращити існуючі машини, або винайти такі, що доти не існували. Він дав перший систематичний опис того, як працюють машини

і як слід комбінувати їхні елементи. Тому, на наш погляд, коректніше віддати да Вінчі шану як інженеру-прикладнику, на відміну від механіків-науковців.

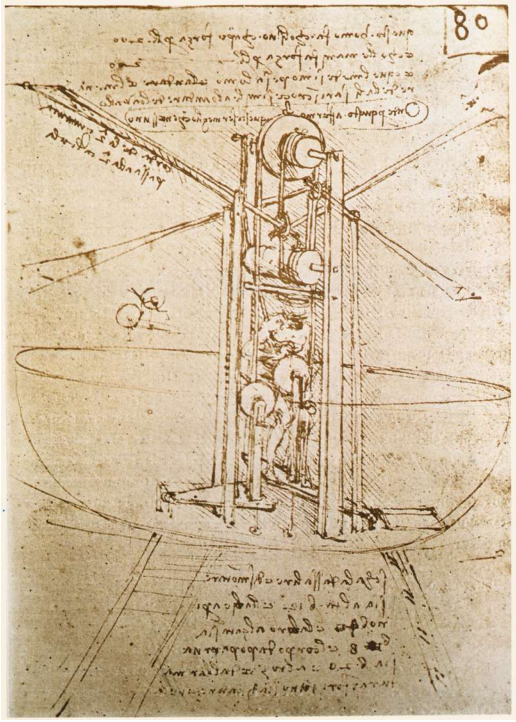


Рис. 3. *Літальний апарат*

Математика і мистецтво

А як же математика? Чи можемо ми вважати Леонардо да Вінчі видатним математиком? Як охарактеризувати його внесок у «царицю наук»? Якщо підходити з точки зору розвитку математичної теорії, то напевно ні. Із сучасних позицій тогочасними математиками називають декількох учених. Фібоначчі (близько 1170 — близько 1250) задовго до Леонардо та Сципіон Даль Феро (1465–1526), Нікколо Тарталья (1499–1557) і Джироламо Кардано (1501–1576) у його часи цікавилися питаннями, зовсім відмінними від тих, яким присвятив свої зусилля Леонардо. Внесок да Вінчі у математику, на відміну від названих алгебраїстів, полягає у прикладному розвитку геометрії. Так, саме художники доби Ренесансу надали поштовх у прогресі геометрії.

Що стосується геометрії, то її зв'язок із мистецтвом сягає сивої давнини. Такі математичні ідеї, як лінійна перспектива, симетрія, золотий перетин і геометрична форма мали безпосередній вплив на мистецтво. Ці геометричні поняття сформували розуміння того, як ми бачимо предмети і світ загалом, та що треба зробити, щоб покращити твори образотворчого мистецтва в сенсі реалістичнішого відображення навколишнього світу. Якраз за доби Ренесансу увагу Леонардо да Вінчі та інших митців привернули ідеї залучення математики у світ мистецтва.



Рис. 4. *Пані з горностаєм*

Знання Леонардо в царині математики були більш ніж скромними аж до того часу, поки йому не виповнилось сорок років. Загалом він як позашлюбна дитина не здобув систематичної освіти у навчальному закладі. Усі знання митець отримав самотужки, займаючись самоосвітою протягом усього свого життя. До математики справа дійшла, коли Леонардо вже став зрілою людиною, вивчати її він почав під керівни-



Рис. 5. *Таємна вечеря*

цтвом монаха-францисканця Луки Пачолі (Luca Pacioli). Отримані математичні знання да Вінчі використав, коли перейшов до вивчення інженерної справи й анатомії.

Золотий перетин. Божественна пропорція

Золотий перетин у мистецтві є добре відомою математичною концепцією. Це особливе число (пропорція), відоме ще із часів Стародавнього Єгипту та Греції. Воно особливе, бо є ірраціональним числом як, наприклад, і добре відоме число π. Золотий перетин, природно, має своє власне позначення ф і дорівнює 1,61803398874989484820... Слід зазначити, що протягом тривалого часу «Золотий перетин » був також відомий під іншими назвами: «Золота пропорція», «Золоте середнє» і, серед інших, «Божественна пропорція». «Золотий перетин» достатньо часто зустрічається у природі, науці й, без сумніву, у мистецтві. Отже, «науковий» зв'язок «Золотого перетину» й мистецтва був остаточно оформлений приблизно 1400 року Леонардо да Вінчі та згадуваним вище францисканцем Лукою Пачолі.

Пачолі був власне математиком у тому розумінні, як ми зараз собі уявляємо математика на рубежі XV століття. Зокрема він написав трактат «Божественна пропорція» (De divina proportione). Принагідно зазначимо, що не варто вважати тогочасних церковників (як і їхніх наступників — інквізиторів, що судили Галілео Галілея сто років потому) збіговиськом твердолобих невігласів. Це були найосвіченіші люди своєї епохи (а серед згаданих інквізиторів були друзі Галілео та його університетські товариші). Отже, у тексті книги обговорювались питання перспективи і застосування математики в архітектурі. Значну увагу також було приділено дискусії про п'ять тіл, які стародавні греки вважали ідеальними (як же ж без стародавніх греків, це ж доба Ренесансу!). Монах-францисканець дружив із багатьма відомими митцями того часу. І одним з них був Леонардо, який на той час став відомим художником, да Вінчі був одним з ілюстраторів трактату Пачолі. Загалом Леонардо зробив близько 60 малюнків для книги. Деякі з них містили геометричні форми та фігури для пояснення математичних понять і концепцій.

Підкреслимо одну важливу рису характеру великого майстра. Він вмів захоплюватись! Якщо якась тема викликала інтерес

у Леонардо, то він легко жертвував усіма іншими справами (разом із замовленнями) на користь вивчення нового питання. Це, напевно, приводило до зриву термінів здачі багатьох робіт і незадоволенню замовників. Власне, через таку свою особливість Леонардо да Вінчі як митець не був дуже продуктивним порівняно з іншими сучасними йому художниками. Так сталось і з математикою. Проте, захопившись геометрією і продовжуючи навчання з Пачолі, да Вінчі використовував золотий перетин у своїх малюнках і картинах у подальшому. Його присутність можна побачити у відомих творах «Таємна вечеря» (рис. 5) і «Мона Ліза» (рис. 6).

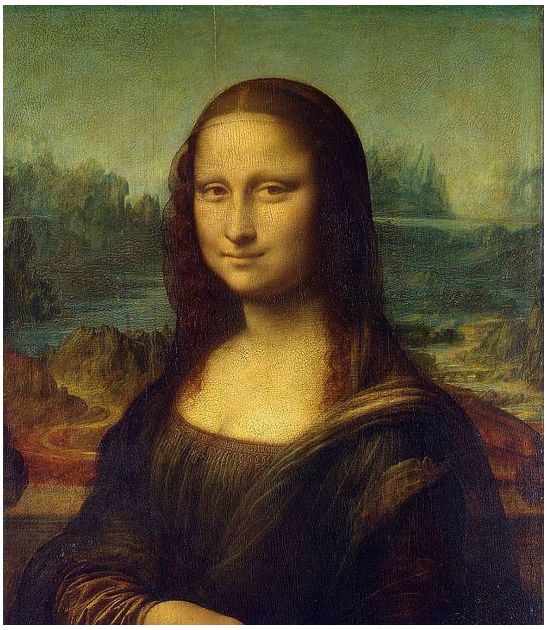


Рис. 6. *Мона Ліза*

Перспектива

Що стосується математики в мистецтві, то Леонардо да Вінчі використовував принципи лінійної перспективи у своїх роботах. Для неї необхідні три основні елементи: паралельні лінії, лінія горизонту і нескінченно віддалена точка. Вміле оперування ними дозволяє художнику реалістично зображати об'єкти, що розміщені далі від спостерігача. Використовуючи лінійну перспективу, да Вінчі створював ілюзію глибини на пласкій поверхні картини або малюнка.

Вивчивши принципи лінійної перспективи, Леонардо знайшов їм вдале використання у деяких ранніх роботах. У своїй картині «Благовіщення» (рис. 7) він використав перспективу для виділення і підкреслення декількох деталей, що суттєво покращило роботу. За допомогою лінійної перспективи художник виділив ріг будинку, огороджений муром садок і доріжку. В ансамблі ці елементи створюють контраст між обмеженістю двору і відкритістю простору за ним.

Серед інших робіт Леонардо відмітимо незакінчену «Поклоніння волхвів», у якій спостерігаємо найбільш виразно майстерність володіння перспективою при зображенні фігур людей. У процесі роботи над картиною да Вінчі зробив велику кількість начерків, зокрема і по достатньо глибокому вивченню перспективи. Один із них містить ліву частину картини, де митець планував зобразити складні руїни класичних споруд. Зрозуміло, що це вимагало глибоких і ґрунтовних знань геометрії (зокрема законів уже згаданої перспективи).

Визнано, що при використанні перспективи, Леонардо був першим, хто застосував анаморфоз — техніку, в якій приховане зображення можна побачити лише у викривленому дзеркалі або дивлячись на картину із певного положення.

Математика і роботи да Вінчі

Двома найвідомішими творами Леонардо да Вінчі, в яких математика використовується найбільш широко, є, безсумнівно, «Таємна вечеря» і «Мона Ліза». У композиціях обох застосовано принципи перспективи, золотий перетин і пропорції. В ро-



Рис. 7. *Благовіщення*

боті «Таємна вечеря» він використовує золотий перетин для розташування різних елементів композиції, зокрема для розміщення зображень людей за столом пропорційно до стін і вікон на задньому плані. У такий спосіб математичний підхід дозволяє да Вінчі узгодити різні елементи на фресці.

У композиції картини Леонардо да Вінчі «Мона Ліза» також використано золотий перетин. Для всіх центральних елементів композиції застосовано золотий прямокутник. І хоча він завжди залишається прямокутником, золота пропорція проявляється у розмірах форм. Звичайно, золотий перетин спостерігаємо перш за все у зображенні обличчя Мони Лізи. Якщо вписати обличчя у прямокутник і провести лінію через очі, то утвориться новий золотий прямокутник. Як результат, відношення довжини голови Джоконди до відстані між очима є золотим перетином; це співвідношення можна відшукати і для інших частин тіла на картині.

Ідеальні пропорції людини

В епоху раннього Ренесансу побутувала думка, що людина створена за подобою божою, а, отже, є ідеальним створінням. Якщо так, то закони математики повинні описувати таку ідеальність. За твердим переконанням тодішніх мислителів і митців математичні принципи лежать в основі всіх форм, які, у свою чергу, підкоряються божественній пропорції. Багато хто щиро дотримувався думки, що людина є найідеальнішим творінням бога, тому людські форми повинні відповідати певним пропорціям.

Використовуючи свої математичні знання, Леонардо да Вінчі якраз і був одним з тих, хто створив дуже багато малюнків, що демонстрували, як вважалось, ідеальні пропорції людини. Однією з найвідоміших робіт, присвячених цьому питанню, є «Вітрувіанська людина» (рис. 8). Це візуальне зображення ідеального тіла людини, виконане на основі математики. Математика для таких людей, як да Вінчі, була засобом побачити пропорційність в усьому, що їх оточувало.

На малюнку «Вітрувіанська людина», зробленому да Вінчі пером і чорнилом на папері, бачимо людину з розпростертими руками і розставленими ногами. Зображення людини є вписаним у круг і у квадрат. Леонардо використовує це зображення, щоб показати різні пропорції тіла люди-

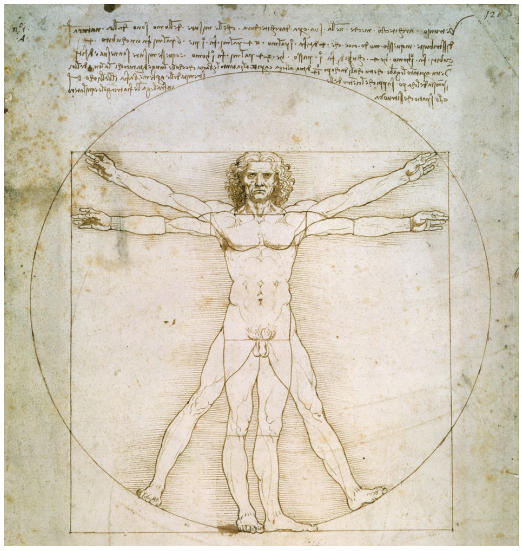


Рис. 8. *Вітрувіанська людина*

ни. У такий спосіб із використанням математики він описує ідеальні фізичні пропорції людини. Для ілюстрації своєї думки та ідеї ідеальних пропорцій важливими є не тільки сам малюнок, а й текст, що його супроводжує. Власне текст надає детальний опис різноманітних пропорцій тіла людини, сформульований відомим римлянином на ім'я Вітрувій. Сам да Вінчі використовує свій малюнок, власне, як ілюстрацію до цього опису.

На додаток до «Вітрувіанської людини» Леонардо зробив декілька інших малюнків, які ілюструють пропорції людського тіла. Захоплений бажанням покращити свої мистецькі творіння, да Вінчі завжди робив велику кількість ескізів і начерків людей з намаганням найточніше відобразити їхні пропорції.

Геометричні форми

Вважають, що, саме робота над ілюстраціями до трактату Луки Пачолі «Божественна пропорція» стала поштовхом до великого зацікавлення Леонардо геометрією. Відтоді протягом усього свого життя великий італієць присвячував дуже багато часу вивченню цього предмета і практичному його застосуванню у своїй творчості. Зацікавлення да Вінчі такими геометричними формами як, зокрема, багатогранники, є помітним у багатьох ескізах, які ми зустрічаємо на всіх малюнках Леонардо.

Для книги «Божественна пропорція» да Вінчі створив шістдесят ілюстрацій геометричних форм, ґрунтуючись на описах і моделях Пачолі. Деякі з них є першими ілюстраціями багатогранників, що зображують фігури із суцільними ребрами, тобто наче зробленими з балок (рис. 9). Такий підхід дозволяв бачити як ребра, розташовані на передньому плані, так і ті, що утворювались задніми гранями. Це запобігало очевидним проблемам сприйняття під час виконання малюнка простими лініями (з ребрами у вигляді дротинок). Відсутність граней у цієї форми означало, що читач міг бачити «крізь фігуру» передні і задні ребра та грані. Геометричні ілюстрації надавали багато корисної геометричної інформації. Окрім багатогранників, да Вінчі створив подібні малюнки й для інших геометричних форм. Причому ілюстрації йшли парами: одна показувала модель із суцільними гранями, тоді як для другої використовувалась техніка суцільних ребер. Отже, Леонардо да Вінчі був скоріше видатним геометром-прикладником, аніж геніальним математиком.

Висновки

Новаторський підхід до мистецтва, розроблений на основі глибоких геометричних знань, дозволив Леонардо створити шедеври, які суттєвим чином виокремлювалися поміж робіт сучасників. Звичайно, да Вінчі запропонував декілька цікавих математичних ідей, проте не вивчив, не дослідив і не довів жодну з них. Він висловив деякі цікаві припущення у своїх щоденниках, проте так їх зашифрував, що це унеможливило дістатись їх точного змісту і породило

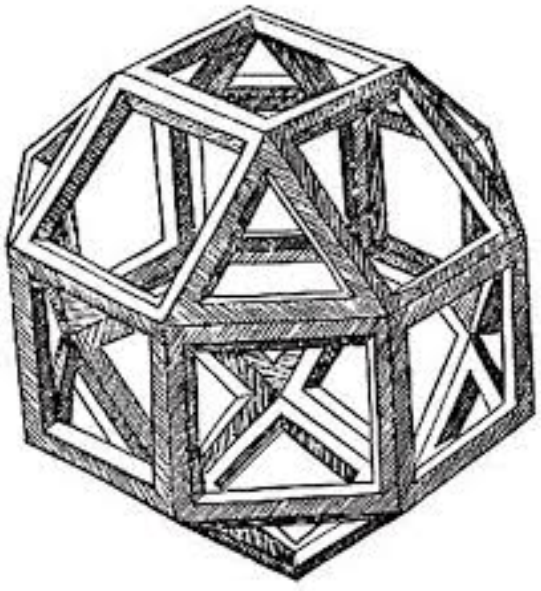


Рис. 9. *Ромбокубоктаедр*

безліч інтерпретацій і «грандіозних» теорій. Але наука загалом і математика зокрема є корисними лише тоді, коли вони побудовані на поняттях, які є однозначними і не містять суперечностей, а також використовують філософію тієї чи іншої області знань. Як з'ясувалося, серед «математичних ідей» да Вінчі дуже мало було правильних, а, можливо, їх і зовсім не було [8]. У науковій літературі неможливо знайти жодного посилання на найменший оригінальний внесок Леонардо у математику.

У результатах, доведеннях і розв'язках, які дають математики, немає місця здогадкам або неоднозначним трактуванням. Вони вивчають різні методи та підходи, на це йдуть цілі роки. Математики вчаться коректно застосовувати численні математичні моделі, щоб знайти розв'язок. Вони порівнюють результати, отримані в межах різних альтернативних підходів, узагальнюють задачі, щоб отримати розв'язки, застосовні до цілих класів задач, і перевіряють свої відповіді, щоб знайти недоліки чи протиріччя. Конструкція математичного результату передбачає доведення, яке може бути перевіреним іншими математиками. Тільки після такої довгої й важкої роботи народжується іскра осяяння, що дозволяє створити щось настільки вагоме, що можна назвати геніальним.

Отже, да Вінчі був далеко не математичним генієм. Проте, талановито застосовуючи тогочасні математичні знання і помноживши їх на свій художній талант, він зміг піднятись до рівня генія мистецького, за що його вважають чи не найгеніальнішим митцем усіх часів і народів.

Література

[1] Ільченко О. Про Леонардо да Вінчі, Карла Ліннея, Жюля Верна, Джона Рокфеллера, Лесю Українку, Вінстона Черчілля / О. Ільченко. — К. : Грані-Т, 2007. — 92 с.

[2] Леонардо та винахід колісного замка вогнепальної зброї // Світ науки. — Травень 1998. — С. 72–77.

[3] Зубов В. П. Леонардо да Винчи. 1452–1519 / В. П. Зубов; отв. ред. М. В. Зубов. — М. : Наука, 2008. — 350 с.

[4] Гуковский М. А. Механика Леонардо да Винчи // М. А. Гуковский — М. : Едиториал УРСС, 2016. — 832 с.

[5] https://en.wikipedia.org/wiki/Leonardo_da_Vinci

[6] http://www-history.mcs.st-andrews.ac.uk/Biographies/Leonardo.html

[7] http://www.lairweb.org.nz/leonardo/index.html#stages

[8] https://www.asme.org/engineering-topics/articles/history-of-mechanical-engineering/leonardo-da-vinci

[9] https://www.engineering.com/Library/ArticlesPage/tabid/85/ArticleID/34/Leonardo-da-Vinci.aspx

Рекурсія в етюдах

Ейлазян Едуард Гайкович,
майстер спорту із шахової композиції

Термін «рекурсія» широко вживається в математиці та програмуванні. Зазвичай його використання пов'язано з методом задання функцій і числових послідовностей, з організацією обчислювальних процедур і циклічних процесів. Рекурсія зустрічається не лише в науці, але й в усіх видах мистецтва, у природних явищах і об'єктах, навіть у життєвих ситуаціях.

Як і будь-який вид творчості, шахова композиція своїми засобами відображає деякі суттєві риси реального світу. Природно виникає питання: чи можна у жанрі етюдної композиції складати твори, основний зміст яких достатньо точно виражається терміном *рекурсія*?

Дивно, що в етюдній композиції тема рекурсії практично не розроблялася, незважаючи на глибокі традиції і широке розповсюдження цього способу структурної організації творів в інших видах мистецтва. Водночас, *тема рекурсії* в етюдах видається цікавою у творчому плані та досить перспективною для розробки. Проте для викладення нової теми однієї назви явно недостатньо. Необхідно надати чітке формулювання, описання й приклади реалізації теми.

Розкрити істинний сенс поняття рекурсії в усьому його різноманітті, не вдаючися до строгих математичних і формально-логічних побудов, доволі складно. Навіть у самій математиці цей термін на початковому етапі вживався без точного означення. Приблизний інтуїтивний сенс рекурсії розкривається на конкретних прикладах. Ось зразок рекурсивного означення: *матрьошка — це лялька, всередині якої знаходиться матрьошка.*

У літературі можна зустріти багато різних означень рекурсії. *Рекурсивним* називається об'єкт (структура, система, процес), що містить у собі ряд подібних вкладених об'єктів як складових. Або таке означення: *рекурсія* — це процес повторення самоподібних об'єктів. Властивість самоподібності полягає у тому, що частина об'єкта володіє деякими (але не усіма!) властивостями цілого об'єкта, причому у свою чергу, є самоподібною за тими самими властивостями.

У рекурсивному об'єкті прийнято виділяти рівень вкладеності. Кількість таких рівнів визначає глибину рекурсії. Якщо у рекурсивній структурі останній вкладений об'єкт є ідентичним першому, то така структура називається *кільцевою* (див., наприклад, [1]). Кільцеву рекурсію можна виявити у поліфонічній музиці Й. С. Баха та у гравюрах М. Ешера.

Спираючись на ці досить загальні означення, сформулюємо тему рекурсії в етюді.

В етюді *виконано тему рекурсії*, якщо розв'язок етюду має рекурсивну структуру, тобто містить не менше трьох послідовно вкладених один в одний самоподібних етюдів.

Як бачимо з означення, елементарною структурною одиницею рекурсії є *етюд в етюді*. Рекурсивну структуру розв'язання етюду утворює послідовність вкладених варіантів розв'язку та/або хибних слідів (спроб), кожен із яких являє собою етюд,

подібний початковому етюду. Подібність етюдів забезпечується наявністю якої-небудь спільної характерної особливості, притаманної кожній із них. Цю особливість називатимемо *формотворчим фактором*.

Класифікуючи типи рекурсії в етюді, будемо спиратися в основному на класифікацію, прийняту в інформатиці. Рекурсія називається *лінійною*, якщо рекурсивна структура не містить розгалужень, тобто кожен етюд, що входить до рекурсивної структури, містить не більше одного вкладеного етюду. В іншому випадку рекурсія називається *каскадною*.

Можна виокремити *два різновиди лінійної рекурсії*:

а) Рекурсивну структуру утворюють послідовно вкладені варіанти розв'язку етюдів. У цьому випадку рекурсивна структурна схема має деяку подібність до форми запису ланцюгового дробу (рис. 1). Тому рекурсію такого типу називатимемо *ланцюговою* (рис. 2).

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}}$$

Рис. 1. *Ланцюговий дріб*

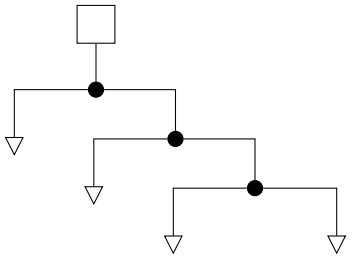


Рис. 2. *Ланцюгова рекурсія*

б) Рекурсивну структуру утворюють послідовно вкладені тематичні хибні сліди (ТХС), тобто тематичні спроби. У цьому випадку кожний перехід на наступний рівень рекурсії супроводжується зміною (кольору) провідної сторони. Тому рекурсію такого типу називатимемо *хамелеонною* (рис. 3).

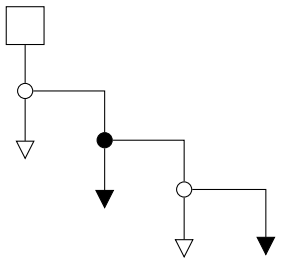
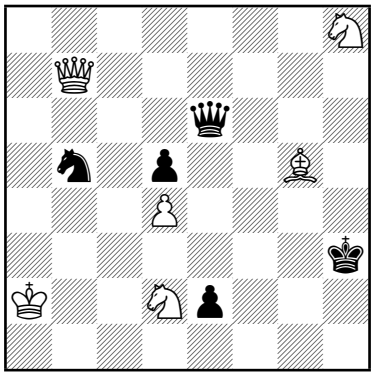


Рис. 3. *Хамелеонна рекурсія*

Можливості реалізації теми рекурсії розглянемо на конкретних прикладах.

№ 1. Е. Ейлазян
«Schach», 2009, третій схвальний відгук



Виграш 6+5

№ 1. Етюд-1. 1. ♘g6! ♕×g6 2. ♕d7+! Перше розгалуження:

I. 2...♔h2 (a) **3. ♕×b5 ♕c2+ 4.♔a3!** (A!) **♕c3+ 5. ♕b3! ♕a1+ 6. ♕a2! ♕c3+ 7. ♘b3!** та виграш.

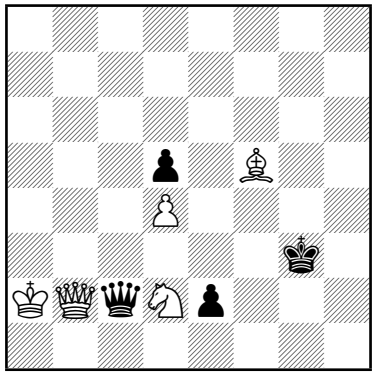
ТХС-1: 4.♕b2? (B?) ♕a4+! (4...♕×b2+? 5. ♔×b2 з виграшем) 5. ♕a3 ♕c2+ 6. ♕b2 ♕a4+... нічия.

II. 2...♔g3 (b) **3. ♕×b5 ♕c2+ 4.♕b2!** (B!).

ТХС-2: 4.♔a3? (A?) e1♕! (4...♕c3+? 5. ♕b3!) 5. ♗h4+ ♔×h4 нічия (6. ♘f3+ ♔g4 7. ♘×e1 ♕c3+ 8. ♔a2 ♕×d4).

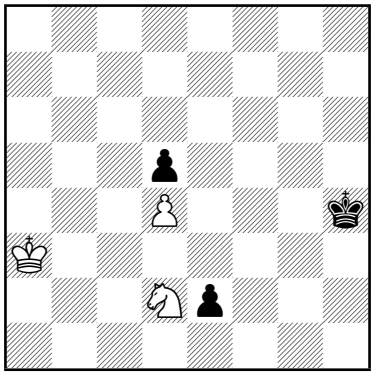
Зміна хибного сліду у варіантах **I–II**.

Етюд-2 (етюд в етюді-1)



Чорні починають, білі виграють

Етюд-3 (етюд в етюді-2)



Білі починають і виграють

Друге розгалуження:

III. 4...♕×b2+ (c) **5.♔×b2 ♔g4 6.♘f3!** (C!) **♔×f3 7.♗d2! ♔e4 8.♔c3** (цугцванг 5+), і білі виграють.

ТХС-3: 6.♗h4? (D?) ♔×h4 7.♘f3+ ♔g3 8.♘e1 ♔f2! Нічия. (Проте не 8...♔f4? 9.♘c2! ♔f3(f5) 10.♔b3! ♔f4 11.♔b4! ♔e4 12.♔c3! (цугцванг 1+), або 11...♔f3 12.♔c5! ♔e4 13.♘e1 (цугцванг 2+) з виграшем).

ТХС: 7.♗h4? ♔e3(f4)! (7...♔e4? 8.♔c3 (цугцванг 4+)) 8.♔c3 ♔e4! (цугцванг 4–) 9.♗g3 ♔f3! 10.♗e1 ♔e3! 11.♗d2+ ♔e4! (цугцванг 5–) 12.♗e1 ♔e3 13.♗h4 ♔e4 14.♗f2 ♔f3 15.♗e1 ♔e3! 16.♗g3 ♔f3!... позиційна нічия.

IV. 4...♕a4+ (d) **5.♕a3+ ♕×a3+ 6.♔×a3 ♔g4 7.♗h4!** (D!).

ТХС-4: 7.♘f3? (C?) ♔×f3 8.♗d2 ♔e4 з нічиєю.

Зміна хибного сліду у варіантах **III–IV**. **7...♔×h4 8.♘f3+ ♔g3 9.♘e1 ♔f4! 10.♔b4!!**

І третє розгалуження:

V. 10...♔e4 (e) **11.♔c5!** (E!) (цугцванг 2+) **♔e3 12.♔×d5** і білі виграють.

ТХС-5: 11.♔c3? (F?) ♔e3! (цугцванг 3–) 12.♘c2+ ♔e4 (цугцванг 1–) 13.♔d2 e1♕+! 14.♔×e1 ♔d3! 15.♔d1 ♔c3 16.♔c1 ♔d3 17.♔b1 ♔c3! 18.♔c1 ♔d3... позиційна нічия.

VI. 10...♔e3 (f) **11.♔c3!** (F!) (цугцванг 3+) **♔e4 12.♘c2!** (цугцванг 1+) (12.♘g2? ♔f3!) **12...♔f3 13.♔d3(d2)** з виграшем.

ТХС-6: 11.♔c5? (E?) ♔e4 (цугцванг 2–) 12.♘c2 ♔d3 13.♘e1+ ♔e4 14.♘g2 ♔f3 15.♘e1+ ♔e4 позиційна нічия.

Зміна хибного сліду у варіантах **V–VI**.

Три послідовно вкладені один в інший тематично подібні етюди утворюють ланцюгову рекурсію. В кожному з етюдів реалізовано *тему зміни хибного сліду* в паралельній формі (формотворчий фактор):

- Етюд-1: X – (a – A!(B?))V(b – B!(A?)).
- Етюд-2: X – (c – C!(D?))V(d – D!(C?)).
- Етюд-3: X – (e – E!(F?))V(f – F!(E?)).

Загальна формула етюду наочно відображає рекурсивний характер структури його розв'язку:

X – (a – A!(B?))V(b – B!(A?) – (c – C!(D?))
V(d – D!(C?) – (e – E!(F?))V(f – F!(E?)))).

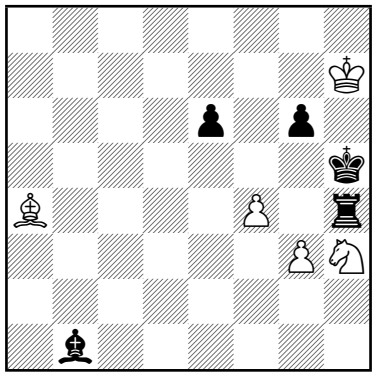
Тут перехід від одного рівня рекурсії до наступного двічі здійснюється шляхом розгалуження основного розв'язку етюду. Крім того, в етюді органічно синтезовано п'ять різних позицій взаємного цугцвангу.

№ 2. Етюд-1. 1.♗d1+ ♖g4 2.♘f2!

[ТХС-1: 2.♔g7? — **Етюд-2** (етюд в етюді-1)]

2...g5+ 3. f5! e×f5! 4.♔g7 ♗d3! Жертву приймати не можна: 5.♘×d3? f4 6.♘f2 f3! 7.♘e4 ♖×g3! 8.♘×g3+ ♔g4 9.♘f1 ♔f4 10.♔g6 g4 11.♔h5 g3 нічия. Чорного слона на d3 можно взяти тільки при білому слоні на f3! **5.♗f3!** Білі проводять *логічний*

№ 2. Е. Ейлазян
«Argentina Bicentennial», 2011,
шостий схвальний відгук



Виграш 5+5

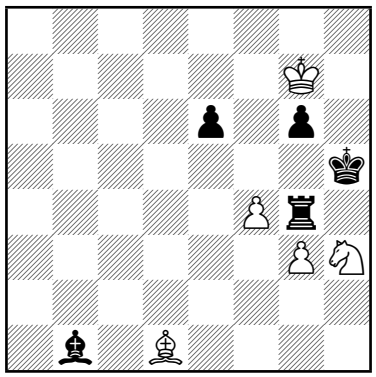
маневр для передачі ходу чорним. **5...♗c2 6. ♗e2 ♗d3 7. ♗d1! ♗c2 8. ♗f3!** Чорні в цугцвангу! (Перша дуель слонів. Рахунок 1:0 на користь білого слона!) Якщо тепер 8...♗d3, то 9. ♘×d3! f4 10. ♘f2 f×g3 11. ♘×g4 ♔h4 12. ♘e3 з перемогою. Два тематичні варіанти:

A) 8...♗b1 9.♘×g4! Але не 9.♗×g4+? f×g4 нічия, наприклад, 10. ♘d1 ♗h7 11. ♘c3 ♗e4 12. ♘×e4 пат. **9...f×g4 10. ♗c6! ♗e4 11. ♗e8+ ♗g6 12. ♗×g6** мат.

Або **B) 8...♗a4 9.♗×g4+!** Але не 9.♘×g4? f×g4 10. ♗d5 ♗b3! 11. ♗c6 ♗a4! 12. ♗e4 ♗c2! 13. ♗×c2 пат. **9...f×g4 10. ♘e4 B~ 11...♘f6** мат.

Два однорідні варіанти **A)** та **B)** пов'язані між собою *зміною хибного сліду*.

Етюд-2 (етюд в етюді-1)



Чорні починають і роблять нічию

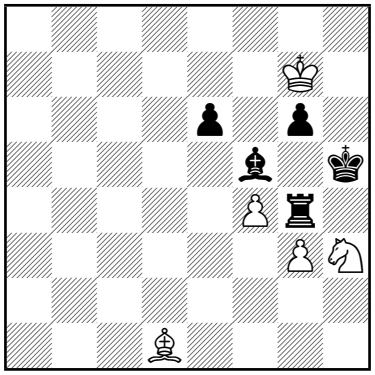
[**ТХС-2:** 2...♗f5? — **Етюд-3** (етюд в етюді-2)]

2...♗c2! Спочатку *логічний маневр* — переведення білого слона на e2. **3. ♗f3 ♗e4! 4. ♗e2** і лише тепер **4...♗f5!** — цугцванг! (Друга дуель слонів. Рахунок вирівнявся, 1:1) **5. ♗d1 ♗c2 6. ♗e2 ♗f5!** і варіанти:

a) 7. ♘g5 e5! 8. ♘h7 e×f4 9. ♘f6+ ♔g5 10. ♘h7+ ♔h5 — позиційна нічия; або

b) 7. ♘f2 g5! (але не 7...e5? 8. f×e5 з виграшем). *Зміна захисту.* **8. f×g5 ♔×g5 9. ♘h3+ ♔h5+ 10. ♔f6 ♔h6!** нічия.

Етюд-3 (етюд в етюді-2)



Білі починають і виграють

3. ♗e2! — цугцванг! І два варіанти:

a) 3...♗d3 4.♗×g4+! ♔×g4 5. ♘f2+ ♔×g3 6. ♘×d3 ♔f3 7. ♔×g6... виграш. Але не 4.♗f3? ♗e4!(...g5? 5. f×g5!) 5. ♗e2 ♗f5! нічия (див. *Розв'язок етюду-2*). Чорні проводять *логічний маневр*, зсуваючи пішака f4 на g5.

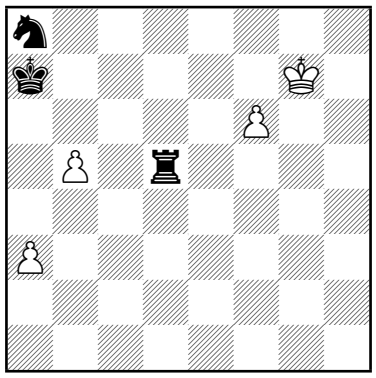
b) 3...g5! 4. f×g5 і тепер **4...♗d3** у розрахунку на 5.♗×g4+? ♔×g4 6. ♘f2+ ♔×g5 7. ♘×d3 ♔g4 нічия. Але **5.♗f3!** — у відповідь *логічний маневр* із римським переведенням чорного слона із d3 на e4. **5...♗e4 6. ♗×g4+!** (Третя дуель слонів. Рахунок 2:1 на користь білого слона!) **6...♔×g4 7. ♘f2+ ♔×g5 8. ♘×e4+** виграш.

Зміна хибного сліду у варіантах **a)** та **b)**.

Три послідовно вкладені один в інший логічні етюди утворюють лінійну хамелеонну рекурсію. Ідейно-тематична однорі-

дність етюдів забезпечується тим, що на кожному рівні рекурсії здійснюються *логічні маневри, що базуються на дуелі слонів.* Крім того, *тема зміни* входить до основного змісту дійсного розв'язку кожного із вкладених етюдів. Тож рекурсивну структуру етюду породжують два різні формотворчі фактори! Перехід з одного рівня рекурсії до наступного двічі здійснюється за допомогою хибного сліду, та двічі відбувається зміна кольору провідної сторони.

№ 3. Е. Ейлазян
Ювілейний конкурс «В. Н. Долгов–75», 2001, Приз



Нічия 4+3

№ 3. Етюд-1. 1. f7 ♘**c7** [**ТХС-1:** 2…♔f6? — **Етюд-2**] **2. b6+! ♔×b6 3. ♔f6 ♖d6+ 4. ♔e7 ♖e6+ 5. ♔d7 ♖f6 6. ♔e7** ♘**d5+ 7. ♔e8 ♖e6+ 8. ♔d8 ♖d6+ 9. ♔e8** ♘**c7+ 10. ♔e7 ♖e6+ 11. ♔d7** — *тематична позиційна нічия* (ТПН) — формотворчий фактор.

Етюд-2 (етюд в етюді-1) *Чорні починають і виграють.*

2…♖d6+ 3. ♔e7 ♖e6+ 4. ♔d7 ♖f6 і розгалуження:

А) 5. ♔e7 ♘**d5+ 6. ♔e8 ♖e6+ 7. ♔d8 ♖b6!** (подолання ТПН — перша варіація!) **8. f8♕ ♖b8+ 9. ♔d7 ♖×f8** виграш.

В) 5. b6+ Етюд-3 (етюд в етюді-2) *Чорні починають і виграють.* **5…♔b7!** [**ТХС-2:** 5…♔a6? — **Етюд-4**; **ТХС-3:** 5…♔×b6? — **Етюд-5**.] **6. ♔e7** ♘**d5+ 7. ♔e8 ♖e6+ 8. ♔d8 ♖c6!** (подолання ТПН — друга варіація!) **9. f8♕ ♖c8+ 10. ♔d7 ♖×f8** виграш.

Етюд-4 (етюд в етюді-3) *Білі починають і роблять нічию.*

6. b×c7! ♖×f7+ 7. ♔c6 ♖f6+ 8. ♔d7 ♖f7+ 9. ♔c6 ♖f8 10. a4 ♖h8 11. a5 ♔a7 12. ♔d7! ♔b7 13. a6+! нічия.

Тематична спроба: 6. ♔e7? ♘d5+ 7. ♔e8 ♖e6+ 8. ♔d8 ♖×b6! (подолання ТПН – третя варіація!) 9. f8♕ ♖b8+ 10. ♔d7 ♖×f8 виграш.

Етюд-5 (етюд в етюді-3) *Білі починають і роблять нічию.*

6. ♔e7 ♘**d5+ 7. ♔e8 ♖e6+ 8. ♔d8 ♖d6+ 9. ♔e8** ♘**c7+ 10. ♔e7 ♖e6+ 11. ♔d7 ♖f6 12. ♔e7** ♘**d5+ 13. ♔e8** — *тематична позиційна нічия.*

Каскадну рекурсію утворюють п'ять вкладених етюдів. Формотворчим фактором, що породжує рекурсивну структуру розв'язку етюду, є ТПН. Рекурсія має чотири рівні вкладення. На нульовому рівні (початковий етюд) ТПН реалізується у дійсному розв'язку. На першому рівні (етюд-2) ТПН долається в першому варіанті розв'язку, на другому рівні (етюд-3) ТПН долається у другому варіанті розв'язку. На третьому рівні відбувається *розгалуження рекурсивної структури*: у ТХС-2 (етюд-4) ТПН долається у тематичній спробі, а в ТХС-3 (етюд-5) реалізується ТПН. Розв'язок етюду-5 ідентичний розв'язку початкового етюду. Тож ця гілка рекурсії утворює *кільцеву* структуру. Перехід з одного рівня рекурсії до наступного здійснюється як шляхом хибного сліду, так і шляхом розгалуження розв'язку.

Наведені приклади не вичерпують усіх можливостей теми рекурсії. Так, наприклад, різноманітні комбінації формотворчих факторів дозволяють отримувати досить складні рекурсивні конструкції. Можна навіть уявити таку фантастичну ситуацію, коли функцію формотворчого фактора трирівневої рекурсії виконує тема ре-

курсії. Рекурсивна форма будови етюду визначає не лише структуру розв'язку, але й пов'язує цю структуру з конкретним шаховим змістом, і у цьому проявляється основний закон естетики — єдиність форми і змісту.

Завершити свій короткий опис теми рекурсії в етюдній композиції мені хочеться цитатою [2]: «Останнім часом найбільш явно проявляється тяжіння літератури до використання рекурсії. Багато авторів інтуїтивно застосовували цей засіб як відображення реального розвитку світу. І це було визнано читачами. Тепер, коли рекурсія є впізнаною, можливе свідоме використання складних рекурсивних конструкцій, широко відомих у теорії алгоритмів і програмуванні... Це інтелектуальні ігри недалекого майбутнього». У літературі це «недалеке майбутнє» вже настало...

Література

[1] Кретов А. Сказки рекурсивной структуры / А. Кретов; ред. П. Рейфман // Труды по русской и славянской филологии. Литературоведение. I. (Новая серия). — Тарту, 1994. — С. 204–214; http://sites.utoronto.ca/tsq/15/kretov15.shtml

[2] Анисимов А. В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. / А. В. Анисимов. — К. : Наук. думка, 1991.