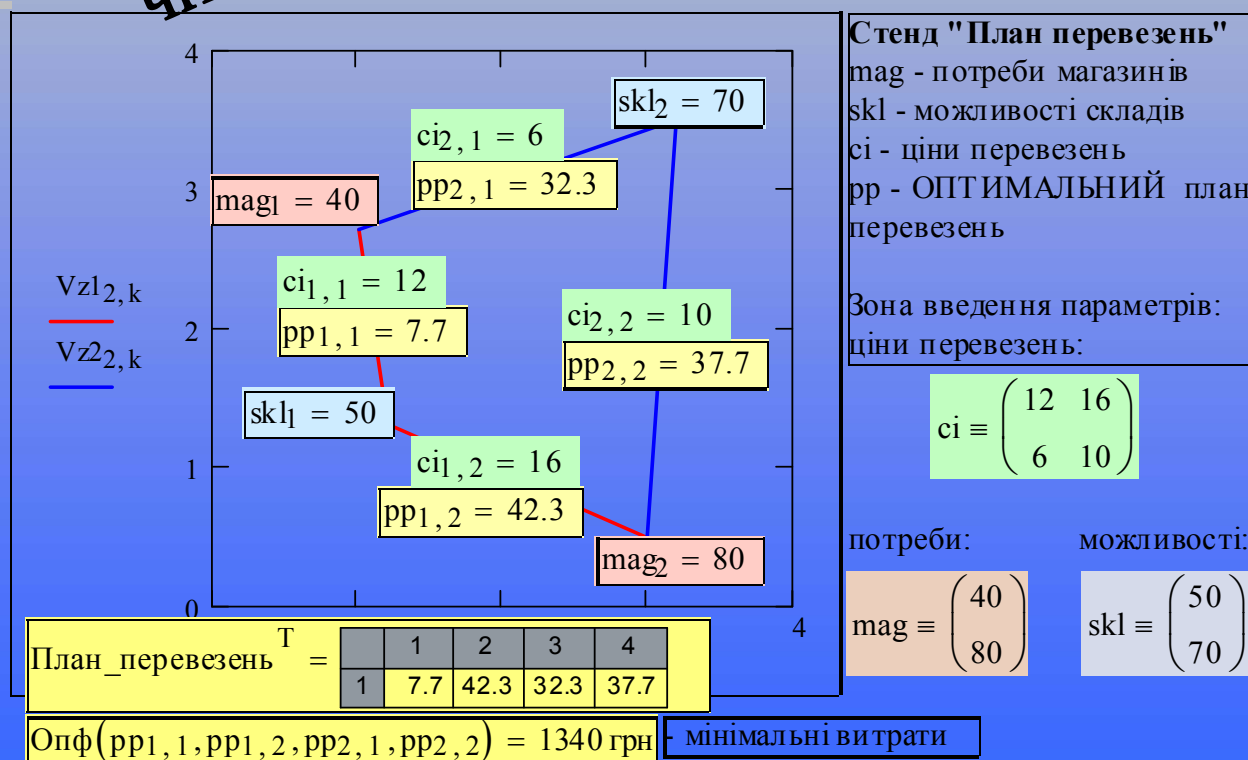


Т. М. Боровська, І. С. Колесник, В. А. Северілов

Основи теорії управління та дослідження операцій

РОЗРОБКА і ДОСЛІДЖЕННЯ:
систем масового обслуговування,
систем управління об'єктами і процесами,
ультранадійних і просто надійних систем,
чітка і нечітка ОПТИМІЗАЦІЯ



Вінницький національний технічний університет

Т. М. Боровська, І. С. Колесник, В. А. Северілов

**ОСНОВИ ТЕОРІЇ УПРАВЛІННЯ
ТА ДОСЛІДЖЕННЯ ОПЕРАЦІЙ**

Рекомендовано Міністерством освіти і науки
України як навчальний посібник для студентів
вищих навчальних закладів

УНІВЕРСУМ-Вінниця 2008

**УДК 681.5(681.3)
Б 31**

Рецензенти:

В. Т. Сусіденко, доктор економічних наук, професор (КНТЕУ ВТЕІ)

Р. Н. Квєтний, доктор технічних наук, професор (ВНТУ)

І. І. Хаїмзон, доктор технічних наук, професор (ВНМУ)

Рекомендовано до видання Міністерством освіти і науки України.
Лист № 14/182 – 1922 від 17.11.03

Боровська Т. М., Колесник І. С., Северілов В. А.

Б 31 Основи теорії управління та дослідження операцій.

Навчальний посібник. – Вінниця: УНІВЕРСУМ-Вінниця, 2008. – 242 с.

ISBN

В посібнику розглянуто ряд задач, що складають основу стандартних курсів теорії управління та дослідження операцій. Це – задачі оптимального управління та стійкості складних систем, імовірнісних процесів – марківських процесів, процесів з розгалуженнями, задачі лінійного і нелінійного програмування, задачі масового обслуговування та управління запасами, задачі надійності складних систем, задачі оптимізації розвитку систем та прийняття рішень на базі нечіткої логіки. Посібник є електронною книгою – в комп'ютерній версії це система діалогових програм, що дозволяють вести серйозні дослідження і є ресурсом для виконання курсових і дипломних проектів. Посібник призначено для широкого кола інженерних та інженерно-економічних спеціальностей.

УДК 681.5

ISBN

© Т. Боровська, І. Колесник, В. Северілов, 2008

ЗМІСТ

Вступ	3
1. Елементи математичної статистики	6
1.1. Міри розташування та розсіювання даних	7
1.2. Міри статистичного зв'язку даних.....	15
1.3. Міри групування даних - частотні розподіли	25
1.4. Ідентифікація частотних розподілів	43
2. Елементи теорії масового обслуговування	56
2.1. Елементи теорії масового обслуговування. Розрахунок та моделювання однофазної СМО	57
2.2. Розрахунок та моделювання багатofазної одноканальної СМО (конвеєрне виробництво)	71
3. Задачі управління запасами	84
3.1. Управління запасами. Базові математичні моделі управління запасами	85
3.2. Задача управління запасами при розподілених поставках і штрафах	94
3.3. Задача управління запасами зі штрафами	99
3.4. Розробка узагальнених математичних моделей управління запасами	103
4. Оптимізація планів виробництва та перевезень	117
4.1. Розробка базової математичної моделі для оптимізації транспортної задачі	118
4.2. Розробка інтерфейсу користувача з підсистемою оптимізації планів перевезень	127
4.3. Масштабування програми оптимізації планів перевезень. Що буде якщо аналіз	132
4.4. Розробка системи для розв'язання задач лінійного програмування альтернативними методами	140
5. Оптимізація розвитку систем	155
5.1. Одновимірна задача Марковиця. Варіаційне обчислення.....	156
5.2. Двовимірна задача Марковиця	167
5.3. Узагальнення задачі Марковиця для випадку нелінійних інвестиційних функцій	181
6. Задачі оптимального управління динамічними системами при обмеженнях	192
6.1. Метод принципу максимуму Понтрягіна. Оптимальне управління інерційним об'єктом	193

6.2. Метод принципу максимуму Понтрягіна. Оптимальне управління об'єктом з демпфуванням	202
6.3. Метод принципу максимуму Понтрягіна. Оптимальне управління об'єктом-маятником	211
7. Оптимізація систем на базі нечіткої логіки	219
7.1. Приклад нечіткої оптимізації на базі експертних оцінок	220
7.2. Оптимізація надійності системи альтернативними методами	226
Література	239

Вступ

В даному посібнику розглянуто класичні і нові задачі оптимізації, моделювання, розрахунку того, що називають “кібернетичними системами”. Є досить вітчизняних і іноземних підручників і посібників з такого класу дисциплін. Але – з’являються нові задачі, нові методи для старих задач, нові інструменти для розв’язання задач – вчора аналогові обчислювальні машини, сьогодні – сучасні ЦОМ, мережі та ін.

У посібнику кожній задачі відповідає документ, що існує у двох формах – **паперовий документ і електронний документ** – фактично діалогова програма з відкритим для вивчення та змін текстом.

Відповідь на питання: що таке електронна книга, може бути надійним і вичерпним тестом для діагностики психічних відхилень у спеціалістів. Справді, якщо людина намагається визначити це через розмір шрифту, інтервали, місце для номерів сторінок... Усім відомо, що звичайна книга, це “корабель думки” (згідно Френсісу Бекону). Є підозра, що і електронна книга – теж корабель, ну, з новим типом двигуна.

Електронна книга – це процес, це певним чином повернення до традицій усної творчості докнижкових і ранньокнижкових епох. Кожен може її скопіювати, змінити на свій розсуд – як пісню? Результатом більшості документів є не отримання певного числового результату, а „інструменту” для розв’язання певного класу задач, „заготовки” для розробки більш досконалих „інструментів”. Щоб підкреслити, що результати є тільки елементами процесу, використовується замість номінального терміну „розв’язок” термін „розв’язання”. В посібнику розглядаються задачі саме управління в довільних системах – технічних, економічних, екологічних, тому використовується термін „управління”, а не „керування”.

Усі розділи починаються з постановки задачі і завдань і далі даються приклади відповідей на усі питання завдання. На питання з економіки, безпеки даються дуже стислі відповіді, бо ця книга присвячена аспектам моделювання, синтезу та дослідження в режимі обчислювального експерименту.

Те, що подано в документах, з одного боку, - закінчені, робочі програми, з другого – це тільки “центри кристалізації” – кожна розробка орієнтована на проведення широких досліджень, подальшу модифікацію і вдосконалення математичних моделей. 10-15 сторінок розділу книги можуть бути розгорнені у 40-70 сторінок дипломного чи курсового проекту.

У формі електронної книги документи посібника пов’язані через стандартні гіперзв’язки (hyperlink) із центральним документом – ЗМІСТОМ, між собою і з різними довідками. Матеріали посібника можна копіювати і переносити у власні розробки. В цьому плані він – склад модулів, переважно таких, що відсутні в фірмових електронних книгах.

1

Елементи математичної статистики і теорії імовірностей

У цьому розділі розглядаються:

- ✓ Міри розташування та розсіювання даних
- ✓ Міри статистичного зв'язку даних
- ✓ Міри групування даних - частотні розподіли
- ✓ Ідентифікація частотних розподілів

Після вивчення цього розділу ви **зможете:**

1. Визначати середні і відхилення від середніх, а головне – вибирати їх відповідно конкретній задачі.
 2. Визначати параметри регресії і кореляції для заданих масивів даних, а головне – критично оцінювати отримані результати. Виконувати аналіз, фільтрацію і розшарування даних.
 4. Виконувати групування даних – будувати емпіричні частотні розподіли - гістограми, визначати параметри гістограм.
 5. Визначати параметри теоретичних розподілів імовірностей.
 6. Визначати імовірності знаходження даних в певних діапазонах.
 7. Визначити інтервали, в яких з певною імовірністю знаходиться випадкова величина.
 8. Знаходити згортки розподілів.
 9. Виконувати ідентифікацію емпіричних частотних розподілів, всебічно перевіряти гіпотези про характер теоретичного розподілу, „схованого” в емпіричному, а точніше – ідентифікувати „механізми”, що породжують статистичні дані.
10. Дати визначення термінам:
- | | |
|--|---|
| 1. Середнє: арифметичне, квадратичне, геометричне, гармонічне та ін. | 5. Мода, медіана, математичне очікування. |
| 2. Середнє відхилення від середнього. | 6. Гістограма, розподіл імовірностей. |
| 3. Дисперсія, стандартне відхилення. | 7. Згортка розподілів. |
| 4. Регресія, кореляція, коваріація. | 8. Статистичний критерій. |

1. Елементи математичної статистики

1.1. Міри розташування та розсіювання даних

1.1.1. Міри розташування даних

Важко знайти таку царину знань, де використовуються методи математичної статистики і теорії ймовірності. Кібернетика і дослідження операцій в значній мірі базуються на математичній статистиці використовують математичну статистику і теорію ймовірності. Природно, що приходить час глянути на все це зверху, узагальнити, ув'язати і просто повторити. Почнемо з найпростішого і немовби очевидного.

Нагадаємо, що міри розташування - це середні, а міри відхилення - це середні відхилення від середніх. Дехто з вас, інтенсивно використовуючи середнє арифметичне та середнє квадратичне, не підозрює що існують інші середні. Розглянемо ці міри на простих прикладах, і заодно ознайомимось з вбудованими статистичними функціями пакету. Припустимо, що вам дано масив експериментальних, емпіричних даних, наприклад:

$$\begin{pmatrix} \text{"Зріст"} & 130 & 145 & 176 & 220 \\ \text{"Вага"} & 110 & 100 & 60 & 55 \end{pmatrix}, \text{ або } \begin{pmatrix} \text{"Середній бал"} & 3.3 & 4.5 & 5.0 & 2.2 \\ \text{"Вага"} & 110 & 100 & 60 & 55 \end{pmatrix}, \text{ або}$$
$$M := \begin{pmatrix} \text{"Дата"} & 13.1 & 15.1 & 17.1 & 33.1 & 1.2 & 12.2 \\ \text{"Кільк покупців"} & 90 & 110 & 55 & 45 & 50 & 150 \\ \text{"Індекс цін"} & 1.00 & 1.03 & 1.3 & 1.1 & 1.0 & 0.9 \end{pmatrix}. \quad \text{ORIGIN} := 1$$

Виділимо з матриці M вектор даних - для наскрізного прикладу. Це можна зробити за допомогою вбудованої функції `submatrix(A, ir, jr, ic, jc)`. Нагадаємо, що **вбудовані статистичні функції пакета звичайно працюють з даними у формі вектора-стовпця**.

$$Vr := \text{submatrix}(M, 2, 2, 2, 7); \quad V := Vr^T; \quad Vr = (90 \ 110 \ 55 \ 45 \ 50 \ 150) .$$

Незалежно від специфіки задачі для контролю та управління завжди потрібна така найпростіша первинна обробка даних:

- виділення максимуму і мінімуму,
- визначення розмірів масивів, сортування,
- визначення середнього, відхилень від середнього,

а вже потім визначаються більш складні характеристики - регресії і кореляцій, частотні розподіли та ін. Сьогодні функції статистичної обробки вбудовані в усі пакети і системи програмування.

Наприклад, в пакеті Mathcad є такі функції

максимум $\max(V) = 150$; мінімум $\min(V) = 45$; кількість рядків $\text{rows}(M) = 3$ і стовпців (колонок) масиву $\text{cols}(M) = 7$. Функція сортування:

$$\text{sort}(V)^T = (45 \ 50 \ 55 \ 90 \ 110 \ 150) ; \quad -\text{sort}(-V)^T = (150 \ 110 \ 90 \ 55 \ 50 \ 45) .$$

Що таке середнє - майже очевидно: середня ціна, середня швидкість, ... але з ймовірністю = 1 можна стверджувати, що ви не знаєте, скільки існує видів середніх і яке середнє - "найкраще"? Згадайте, придумайте можливі середні.

Середнє - важлива характеристика набору кількісних даних. Вона вказує на "центр" даних. Точно так же, як люди можуть мати різні думки відносно "центру міста" (торговий центр, діловий центр, культурний центр...), так і в статистиці, в залежності від обставин використовують різні середні:

- 1) середнє діапазону,
- 2) середнє арифметичне,
- 3) середнє геометричне,
- 4) середнє гармонічне,
- 5) середнє квадратичне,
- 6) медіана (=медіанне середнє),
- 7) мода.

Розглянемо стисло визначення і формули цих середніх. Зробимо тестовий масив випадкових z : $N := 100$ величин; $i := 1..N$; $V_i := (rnd(10) + rnd(10)) \cdot By$;
 $rows(V) = 100$; $cols(V) = 1$; $max(V) = 18.319$; $min(V) = 0.838$

1. Напівсума крайніх значень (=середнє діапазону)

Подаємо формулу визначення цього середнього у вигляді функції користувача. *Самостійно, не дивлячись у відповідь, знайдіть його переваги і вади цього середнього. (Легко обчислювати, але коли дані сконцентровані на якійсь границі діапазону - буде велика різниця з середнім арифметичним).*

$$polysym(V) := (max(V) + min(V)) \cdot 0.5; \quad polysym(V) = 9.578.$$

2. Середнє арифметичне

Це найбільш розповсюджене середнє, воно здається таким природним, і для цього є деякі підстави - ми потім розглянемо фундаментальну властивість середнього арифметичного і середньоквадратичного. Подаємо формулу і вбудовану функцію:

$$Sar(V) := \frac{1}{N} \cdot \sum_i V_i; \quad Sar(V) = 9.842; \quad mean(V) = 9.842$$

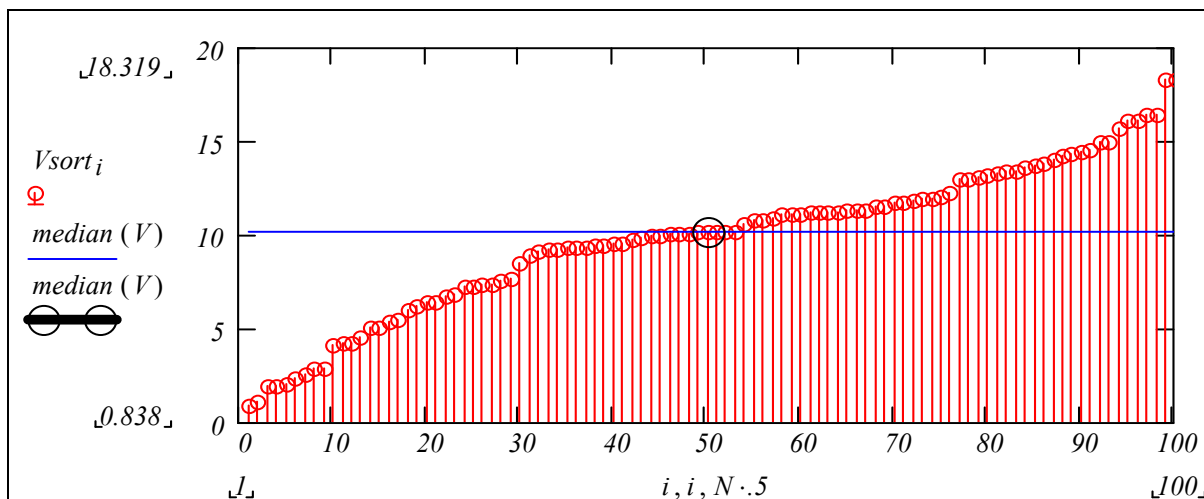
3. Медіанне середнє, або просто медіана

Медіана - величина, що знаходиться в середині упорядкованого набору даних. Звичайно медіану визначають на частотному розподілі (див. підрозділ 1.3).

Подаємо функцію користувача: і вбудовану функцію обчислення медіани:

$$Smed(v) := sort(v)_{ceil(length(v) \cdot 0.5)}; \quad Smed(V) = 10.13; \quad median(V) = 10.138.$$

На графіку сортованих даних медіана проходить через **середню** (25-ту) точку масиву. Якщо побудувати гістограму (= емпіричний частотний розподіл), то медіана ділить площу частотного розподілу навпіл. $Vsort := sort(V)$

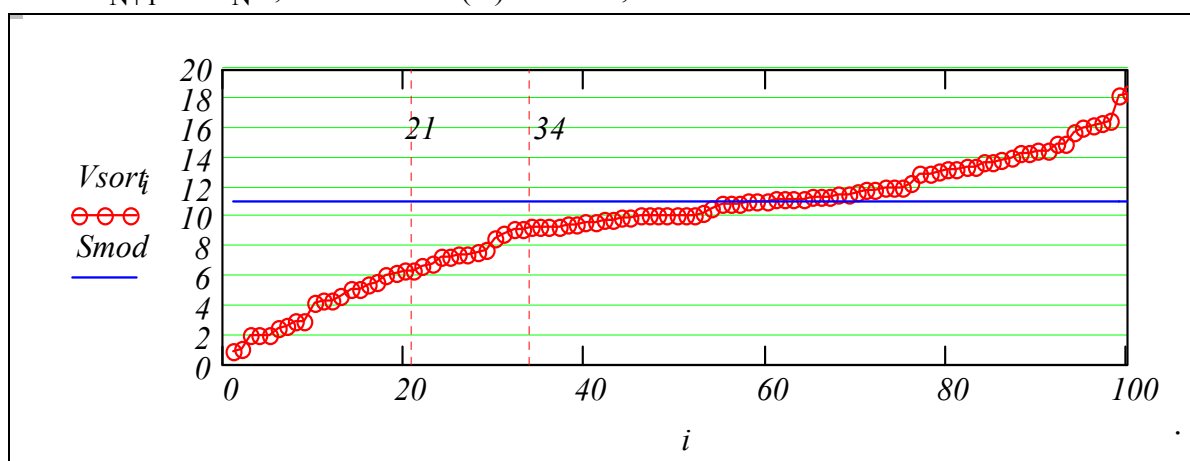


4. Мода

Мода є значенням, що найбільш часто зустрічається в наборі даних. Інакше можна сказати, що це координата, де дані розташовуються найбільш щільно. **Звичайно, моду визначають на частотному розподілі**, і мод може бути декілька.

Визначати моду є смисл на великих вибірках, $N > 100$..., в даному випадку можна знайти її приблизно за графіком: бачимо, що в діапазоні 10-12 більше всього даних = $34 - 21 = 13$. На частотному розподілі - мода - це екстремум. Вам зрозуміло тепер, що таке бімодальні, полімодальні розподіли. В пакеті є відповідна вбудована функція `mode()`.

$$V_{N+1} := V_N ; \quad \text{mode}(V) = 4.58 ; \quad \text{Smod} := 11 .$$



5. Середнє геометричне

Середнє геометричне можна визначити й так: **"антилогарифм від середнього арифметичного від логарифмів величин набору даних"**. Використовується це середнє для усереднення послідовності дробів і, між іншим, в так званому "геометричному програмуванні" - ефективному методі оптимізації. Далі подано пару формул обчислення середнього геометричного і відповідну вбудовану функцію.

$$S_{geom} := \left(\prod_i V_i \right)^{\frac{1}{N}} ; \quad S_{geo} := \exp \left[\frac{1}{N} \cdot \left(\sum_{i=1}^N \ln(V_i) \right) \right] ;$$

$$S_{geom} = 8.671 ; \quad S_{geo} = 8.671 ; \quad gmean(V) = 8.617 .$$

6. Середнє гармонічне

Середнє гармонічне можна визначити отак: **"зворотна величина від середнього арифметичного від зворотних величин набору даних"**. Далі подано формулу обчислення середнього гармонічного, що точно втілює словесну формулу.

Використовується це середнє для усереднення послідовності "швидкостей на однакових дистанціях" - придумайте змістовну задачу, ну, хоч для цеха автоматів: "задано випуск деталей за 8 годин для кожного наладчика (усього їх $=N$)...Чому дорівнює середній випуск усього цеха за годину?"

$$Sharm := \frac{N}{\sum_i \frac{1}{V_i}} ; \quad Sharm = 6.763 ; \quad hmean(V) = 6.731 .$$

7. Середнє квадратичне

Корінь із середнього квадратичного дорівнює кореню із середнього арифметичного квадратів величин набору даних. (Є легка плутанина - часто середнім квадратичним називають саме корінь із середньоквадратичного). Використовується середнє квадратичне для моментів інерції та ін.

$$Sqvad := \sqrt{\left(\frac{1}{N}\right) \cdot \sum_i (V_i)^2} ; \quad Sqvad = 10.597 .$$

Збираємо усі середні для нашого прикладу разом і бачимо, що вони досить різні, як і наші визначення центру міста.

$$\begin{aligned} Sar(V) &= 9.842 ; & polysym(V) &= 9.578 ; & Smod &= 11 ; & Sqvad &= 10.597 \\ Smed(V) &= 10.13 ; & Sgeom &= 8.671 ; & Sharm &= 6.763 . \end{aligned}$$

Контрольні питання:

1. Яким повинен бути набір даних, щоб усі середні були однаковими?
2. Назвіть переваги, або "екологічну нішу" для кожного з середніх.

1.1.2. Міри розсіювання (= Середні відхилення від середніх)

Почнемо розгляд з такої актуально-практичної задачі. Уявіть собі, що Ви виробили 5 підшипників, **середній** розмір яких 50мм при номіналі 50мм, а ваш колега теж 5 і з **середнім** розміром 49.99. Хто працює краще?...

якщо підшипники мали такі розміри:

Ваші - 45.00, 55.00, 46.00, 48.00, 56.00 Його - 49.98, 50.00, 49.98, 50.00, 49.99

Бачимо, що є якась різниця, але як її вимірити? Для кількісної оцінки саме того, чим відрізняються ці два набори даних використовуються **міри розсіювання**. Як і для середніх можна запропонувати декілька альтернативних мір розсіювання. Ознайомимось з цими мірами та проведемо порівняльний аналіз.

1. Розмах

Розмах - простіша міра розсіювання, дорівнює різниці максимального і мінімального значень набору даних. Цю величину можна зробити відносною, безрозмірною, поділивши її на середнє:

$$\frac{(\max(V) - \min(V))}{\text{mean}(V)} = 1.786 .$$

Задача. Якщо середнє від частки можна визначити як частку від середніх, то з мірами розкиду це не так. А як саме?

2. Середньоарифметичне відхилення від середньоарифметичного

Ця міра має унікальну властивість, через яку ніколи не використовується. Доведіть цю властивість. Це відсіюючий тест - завжди пам'ятайте такий факт:

$$OO := \left(\frac{1}{N}\right) \cdot \sum_i (V_i - \text{mean}(V)) ; \quad \text{mean}(V - \text{mean}(V)) = 0 ; \quad OO = 0.1$$

3. Середнє абсолютне відхилення від середньоарифметичного

На практиці часто і додатні і від'ємні відхилення є однаково не бажаними. Ну от, для мисливця, яка різниця - вище чи нижче пролетіла куля? Ця міра визначається як середнє арифметичне модулів відхилень від середнього. Визначаємо її двома альтернативними методами. Між іншим, риска над виразом - оператор векторизації. Означає вона "виконати операцію над кожним компонентом вектора чи матриці". Векторизація прискорює обчислення.

$$\text{dmod} := \left(\frac{1}{N-1} \right) \cdot \sum_i |V_i - \text{mean}(V)| ; \quad \text{Dmod} := \text{mean} \left(\overrightarrow{|V - \text{mean}(V)|} \right)$$

$$\text{dmod} = 3.1 ; \quad \text{Dmod} = 3.09 .$$

4. Середнє квадратичне відхилення від середньоарифметичного

Для аналітичного розв'язання оптимізаційних задач статистики, коли треба брати похідну від міри розсіювання, легше взяти похідну від x^2 - це буде $2 \cdot x$ - лінійна функція. А похідна від модуля? От тому масова наука так полюбить квадратичні критерії та помилки. Ця міра визначається як **середньоарифметичне від квадратів відхилень від середнього**. Визначаємо її 4-ма альтернативними методами - за формулами і за вбудованою функцією $\text{var}()$ - variance (варіація, дисперсія), $\text{stdev}()$ - standard deviation (стандартне відхилення).

$$\text{dqva} := \sqrt{\left(\frac{1}{N} \right) \cdot \sum_i (V_i - \text{mean}(V))^2} ; \quad \text{Dqva} := \sqrt{\text{mean} \left(\overrightarrow{(V - \text{mean}(V))^2} \right)} ;$$

$$\sqrt{\text{dqva}} = 3.929 ; \quad \sqrt{\text{Dqva}} = 3.944 ; \quad \sqrt{\text{var}(V)} = 3.944 ; \quad \sqrt{\text{stdev}(V)} = 3.944 .$$

Фактично із множини мір відхилення в практиці використовується тільки ця остання міра. На то є підстави.

Властивості середньоквадратичного відхилення від середньоарифметичного

1. Обчислення цієї міри (назви: variance, дисперсія) полегшується, **якщо врахувати властивість (не дивлячись у відповідь доведіть це самостійно!?)**:

$$\sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (V_i - \text{mean}(V))^2} = 3.93 ; \quad \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (V_i)^2 - \text{mean}(V)^2} = 4.06$$

Доведення (майже без коментарів):

$$\sum_i (V_i - \text{mean}(V))^2 ; \quad \sum_i (V_i)^2 - 2 \cdot N \cdot \text{mean}(V) \cdot \sum_i V_i + \sum_i (\text{mean}(V))^2 ;$$

$$\text{враховуємо, що } \sum_{i=1}^N V_i = N \cdot \text{mean}(V) , \text{ тому } \sum_{i=1}^N \text{mean}(V) = N \cdot \text{mean}(V) .$$

В результаті отримуємо:

$$\sum_i (V_i)^2 - 2 \cdot \text{mean}(V) \cdot (N \cdot \text{mean}(V)) + N \cdot (\text{mean}(V))^2 = \sum_i (V_i)^2 - N \cdot (\text{mean}(V))^2 .$$

- тобто твердження доведено.

2. Якщо обчислити середньоквадратичні відхилення від **різних середніх**, то виявиться, що найменше з них - **середньоквадратичне від середньоарифметичного**.

Тобто, якщо за міру розсіювання вибрати середньоквадратичне, то найкращим центром, (серединою) даних виявляється середнє арифметичне. Тому основними в статистиці стали саме ці міри і саме вони є вбудованими функціями багатьох мов і середовищ програмування.

Це легко перевірити експериментально і досить важко довести. Спробуйте? Виконаємо експериментальну перевірку - запишемо вирази для середньоквадратичних відхилень від різних середніх:

$$dqsgm := \sqrt{\left(\frac{1}{N}\right) \cdot \sum_i (V_i - S_{geom})^2}; \quad dqsgm = 4.1 \quad ;$$

$$dqsqv := \sqrt{\left(\frac{1}{N}\right) \cdot \sum_i (V_i - S_{qvad})^2}; \quad dqsqv = 4.001 \quad ;$$

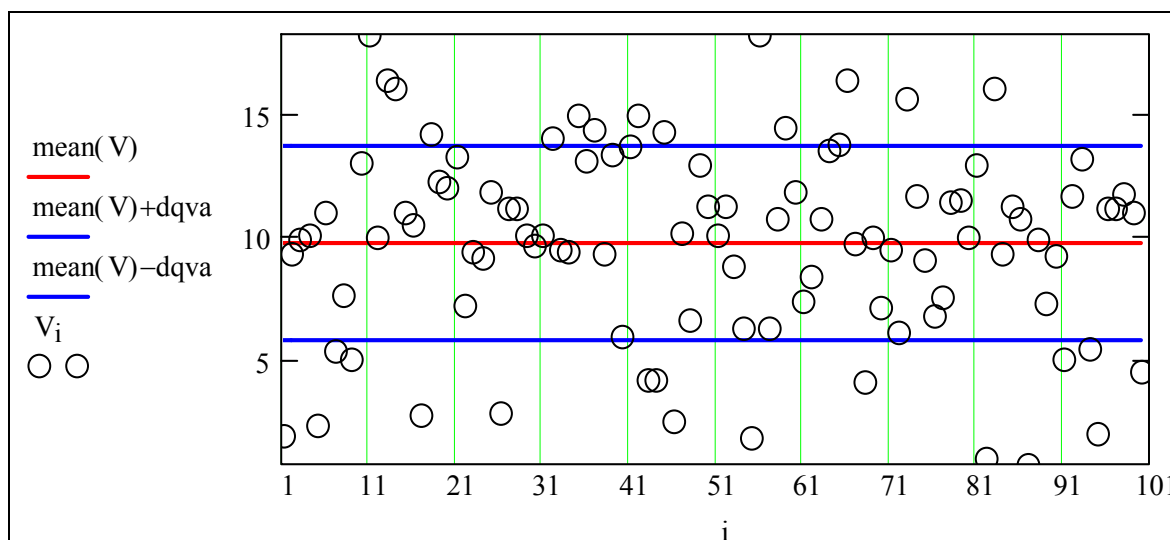
$$dqva = 3.929$$

$$dqshar := \sqrt{\left(\frac{1}{N}\right) \cdot \sum_i (V_i - S_{harm})^2}; \quad dqshar = 4.992 \quad ;$$

Бачимо, що середньоквадратичне відхилення від середньоарифметичного є найменшим серед середньоквадратичних відхилень від інших середніх.

Подіємо варіант програми - відповідь на наступне самостійне завдання. Прогнозуйте його, якщо не хочете втратити потенційний власний варіант.

$kk_i := \text{if}(|V_i - \text{mean}(V)| - dqva > 0, 0, 1);$ Част_даних := mean(kk); Част_даних = 0.67



На цьому графіку подано "статистичні" дані, середнє арифметичне і середні квадратичні відхилення.

Завдання. Напишіть САМОСТІЙНО програму для обчислення частки даних, що попадають в діапазон між стандартними відхиленнями. Модифікуйте програму для обробки масиву у 130 даних, 1300 даних. Подаємо завдання у графічному вигляді.

Завдання для самостійного виконання. Вбудовані функції дуже зручні, але вони:
 1) не завжди вдалі; 2) є чорними ящиками; 3) розтренировують, відучують мислити і працювати.

Тому перевірте себе - не дивлячись на зразки напишіть банальні програми для знаходження максимуму і мінімуму.

brainpowder:=	<pre> mi ← 100000001 ma ← -999999999 for i ∈ 1..N mi ← V_i if V_i < mi mi otherwise ma ← V_i if V_i > ma ma otherwise Otv ← (mi (ma) Otv</pre>
---------------	---

Бачимо, наша програма працює як фірмова:

$$\text{brainpowder} = \begin{pmatrix} 0.838 \\ 18.319 \end{pmatrix}; \quad \begin{pmatrix} \min(V) \\ \max(V) \end{pmatrix} = \begin{pmatrix} 0.838 \\ 18.319 \end{pmatrix}.$$

Тест 1. Перевірте експериментально, який відсоток даних попадає у діапазон $(x - \sigma, x + \sigma)$ для 1) рівномірного розподілу 2) нормального розподілу.

Який відсоток даних попадає у діапазон $(x - 2\sigma, x + 2\sigma)$ для 1) рівномірного Розподілу; 2) нормального розподілу.

Підказка – вбудована функція $\text{rnd}(a)$ видає випадкове число x з діапазону $(0 \leq x \leq a)$, а сума трьох таких чисел буде мати розподіл близький до...

$$V_{2i} := \text{rnd}(10) + \text{rnd}(10); \quad dqva := \sqrt{\left(\frac{1}{N}\right) \cdot \sum_i (V_{2i} - \text{mean}(V_2))^2}.$$

$$kk_i := \text{if}(|V_{2i} - \text{mean}(V_2)| - dqva > 0, 0, 1); \quad \text{частка_даних} := \text{mean}(kk).$$

Тест 2. Прокоментуйте програму подану далі. Побудуйте графік видачі цієї програми як функцію N та M . Зробіть висновки відносно розкиду даних.

$$kk_i := \text{if}(|V_{2i} - \text{mean}(V_2)| - dqva > 0, 0, 1); \quad \text{частка_даних} := \text{mean}(kk)$$

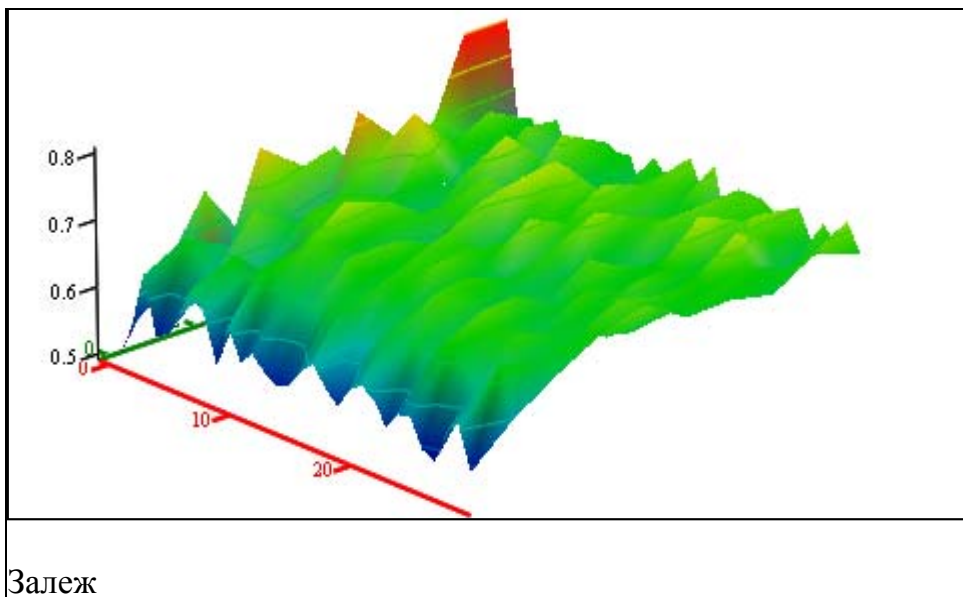
Доля_Д(К, М) :=	for i ∈ 1 .. К $V3_i \leftarrow \sum_{q=1}^M \text{rnd}\left(\frac{10}{M}\right)$ "Енд" очікування ← mean(V3) $\text{відх_серед_ква} \leftarrow \text{mean}\left[(V3 - \text{mean}(V3))^2\right]^{0.5}$ Верх ← очікування + відх_серед_ква Низ ← очікування - відх_серед_ква for i ∈ 1 .. К $\text{попало}_i \leftarrow \begin{cases} 1 & \text{if Низ} < V3_i < \text{Верх} \\ 0 & \text{otherwise} \end{cases}$ "Енд" Відсоток ← mean(попало)
-----------------	---

$$\overline{\text{частка_даних}} = 0.63$$

Доля_Д(50, 2) = 0.76 Доля_Д(50, 2) = 0.64 Доля_Д(50, 2) = 0.6

Останнє питання: Чому для одних і тих же вхідних параметрів програма Доля_Д(50, 2) видає різні значення?

$i := 1 \dots 30$ $j := 1 \dots 10$ Залеж_{i, j} := Доля_Д(i·10, j)



1.2. Міри статистичного зв'язку даних (регресія, кореляція)

Ми розглянули міри групування та розсіювання даних. Тепер переходимо до мір статистичного зв'язку між даними. Де, коли, для чого використовуються ці міри, як вони визначаються? Розглянемо це з позицій інженера.

Розробка моделей усталених процесів "вхід-вихід" для підприємств і техпроцесів, перевірка теоретичних моделей, побудованих на базі певних фізичних законів, звичайно базується на експериментально виявлених залежностях між цими входами і виходами. Експерименти з вимірюванням реальних входів і виходів також виконуються і в процесі функціонування системи - для покращення та уточнення моделей систем та процесів, тобто для оптимізації, настроювання системи управління.

Регресійний аналіз - це метод побудови моделі, яка **найкраще** відповідає набору експериментальних даних. Апріорно вважається, що дані вимірювань містять в собі "істинні" дані, засмічені **випадковими** помилками та шумами. Термін "найкраще відповідає" не означає, що модель точно збігається з наявною вибіркою даних. Нам треба сконструювати якусь числову міру розходження моделі з експериментальними даними, яку ми і будемо мінімізувати. Ви вже морально підготовлені до того, що формально нема єдиної найкращої міри, але в певному сенсі вона є - це сума квадратів відхилень - аналог дисперсії, середньоквадратичного відхилення - ну дуже зручна міра. Згадаємо (див підрозділ 1.1.2), що вона є найменшою серед інших (правда, невідомо, куди цю властивість застосувати).

1.2.1. Побудова лінійної регресійної моделі

На рисунку 1 подано N-мірний набір пар даних (наприклад, кров'яний тиск у пацієнта - кров'яний тиск у лікаря після бесіди з пацієнтом). Прийнято попередню гіпотезу про статистичну **лінійну залежність** між цими величинами (якщо серйозно, то існує таке явище як **емпатія**). На графіку подано і цю модель і "помилки" даних відносно цієї моделі.

Вводимо дані для побудови графіків регресії. $ORIGIN := 1$ $N := 100$; $i := 1..N$.

Модель статистичних даних: $A0d := 2$; $A1d := 0.5$; ; $S = 5$;

$$X_i := .1 \cdot i + \text{rnd}(.5) ; \quad Y_i := A1d \cdot X_i + A0d + \left(\text{rnd}(S) - \frac{S}{2} \right). \quad (1)$$

$$Y_{o_i} = A1d \cdot X_{o_i} + A0d$$

"Істинні" величини X_o та Y_o "засмічені" помилками та шумами вимірювань. Ми беремо слово "істинний" в лапки, тому, що істинних значень в багатьох випадках не існує. Зараз не обговорюємо цю проблему.

Зауважимо, що фізика елементарних часток якось пристосувалась до того, що точні значення імпульсу і координати електрона принципово не існують (принцип невизначеності). Це фізика, економіку характеризують як вимірювання резинових речей резиновим метром. В соціології, взагалі, результати вимірювання, саме створюються в процесі вимірювання. Але віковий досвід підказує, що в кожному хаосі можна знайти гармонію і порядок.

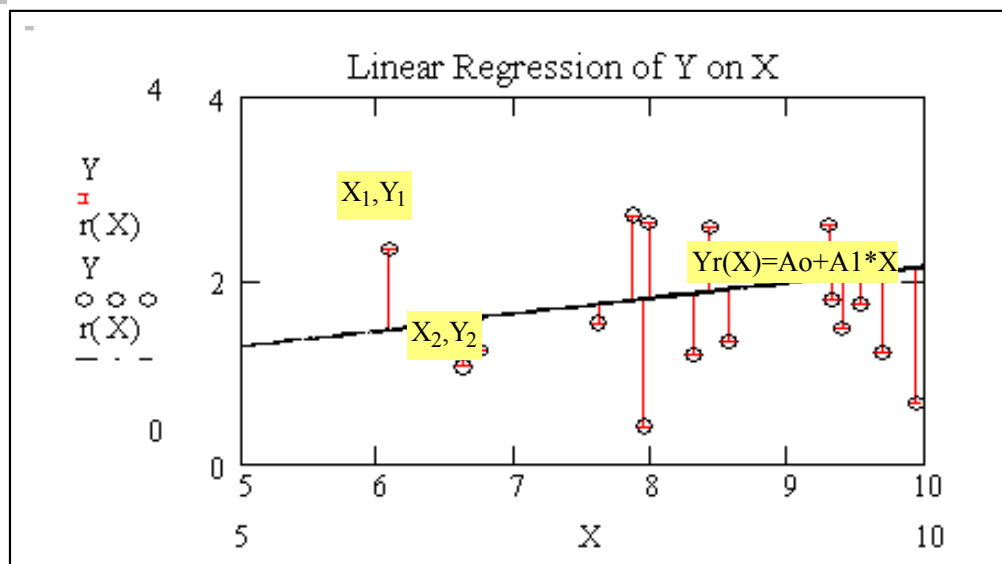


Рис.1. Схема лінійної регресії

Постановка задачі побудови лінійної регресії. Маємо набір пар даних (між якими підозрюється залежність) і лінійну модель залежності

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \quad y_i = A_0 + A_1 \cdot X_i$$

де y_i - величина, що прогнозується регресійною моделлю. **Необхідно отримати такі значення коефіцієнтів A_0, A_1 , для яких сума квадратів помилок є мінімальною.**

Розв'язання задачі

Для кожної точки "незалежної" (такої, що ми прийняли за причину) експериментальної змінної X можна записати рівняння прогнозу :

$$\left. \begin{array}{l} y_1 = A_0 + A_1 \cdot X_1 \\ \dots\dots\dots \\ y_N = A_0 + A_1 \cdot X_N \end{array} \right\}$$

Для кожної точки можна записати рівняння помилки прогнозування:

$$\left. \begin{array}{l} E_1 = y_1 - Y_1 = A_0 + A_1 \cdot X_1 - Y_1 \\ \dots\dots\dots \\ E_N = y_N - Y_N = A_0 + A_1 \cdot X_N - Y_N \end{array} \right\}$$

Складаючи усі ці помилки, отримуємо вираз для критерію оптимізації - **суми квадратів помилок** :

$$F = (E_1)^2 + \dots + (E_i)^2 + \dots + (E_N)^2 \quad ; \quad F = \sum_{i=1}^N (E_i)^2 = \sum_{i=1}^N (A_0 + A_1 \cdot X_i - Y_i)^2 .$$

Як розв'язати цю оптимізаційну задачу? (поставте собі 7 балів, якщо вам одразу прийшло в голову те, що написано далі). Шукаємо екстремум функції двох змінних A_0 та A_1 (а не X_i, Y_i , як Ви могли подумати). Завдяки квадратичності критерію, отримуємо лінійні рівняння для визначення оптимальних значень A_0 і A_1 .

Записуємо необхідні умови мінімуму:

$$\frac{d}{dA_0} F(A_0, A_1) = 0 ; \quad \frac{d}{dA_1} F(A_0, A_1) = 0 .$$

Увага! Цей посібник є електронною книгою, створеною в середовищі пакета Mathcad. Певні математичні операції тут виконуються символьним процесором. Для того, щоб отримувати чисто символьні результати, усі змінні символьних перетворень НЕ ПОВИННІ МАТИ ЧИСЛОВІ ЗНАЧЕННЯ. Тому ми трохи змінили назви змінних. Записуємо вираз для цільової функції

$$F(A_0, A_1) := \sum_{k=1}^K (A_0 + A_1 \cdot X_k - Y_k)^2 .$$

Знаходимо вирази для похідних від цільової функції за допомогою символьного процесора

$$\begin{aligned} \frac{d}{dA_0} F(A_0, A_1) &\rightarrow 2 \cdot K \cdot (A_0 + A_1 \cdot X_k - Y_k) ; \\ \frac{d}{dA_1} F(A_0, A_1) &\rightarrow 2 \cdot K \cdot (A_0 + A_1 \cdot X_k - Y_k) \cdot X_k . \end{aligned}$$

Прирівнюємо вирази похідних до нуля. Виконуємо алгебраїчні перетворення, множник "2" в обох рівняннях виносимо і скорочуємо.

$$N \cdot A_0 + \sum_{i=1}^N A_1 \cdot X_i = \sum_{i=1}^N Y_i ; \quad A_0 \cdot \sum_{i=1}^N X_i + A_1 \cdot \sum_{i=1}^N (X_i)^2 = \sum_{i=1}^N X_i \cdot Y_i .$$

Отримали систему двох рівнянь з двома невідомими - A_0 , A_1 . Записуємо її у матрично-векторному вигляді (використовуємо скорочені означення суми):

$$\begin{bmatrix} N & \sum_i X_i \\ \sum_i X_i & \sum_i (X_i)^2 \end{bmatrix} \cdot \begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} \sum_i Y_i \\ \sum_i X_i \cdot Y_i \end{pmatrix} .$$

Розв'язуємо цю систему рівнянь, отримуємо формули для параметрів регресії

$$A_0 := \left[\frac{\sum_i Y_i \cdot \left[\sum_i (X_i)^2 \right] - \sum_i X_i \cdot \sum_i X_i \cdot Y_i}{N \cdot \left[\sum_i (X_i)^2 \right] - \left(\sum_i X_i \right)^2} \right] ; \quad A_1 := \left[\frac{N \cdot \sum_i X_i \cdot Y_i - \sum_i X_i \cdot \sum_i Y_i}{N \cdot \left[\sum_i (X_i)^2 \right] - \left(\sum_i X_i \right)^2} \right]$$

Нижче подано "кухню" - копію блока розв'язання системи рівнянь у символьному вигляді. Отримані нами вирази є досить складними, в них дуже легко допустити помилки. Нагадуємо - **сучасні стандарти програмування та роботи з математикою вимагають виключати людину із складних, але рутинних процедур** . В багатьох випадках забороняється виконувати алгебраїчні перетворення вручну.

Given

$$N \cdot A_0 + \sum_{i=1}^N A_1 \cdot X_i = \sum_{i=1}^N Y_i \quad A_0 \cdot \sum_{i=1}^N X_i + A_1 \cdot \sum_{i=1}^N (X_i)^2 = \sum_{i=1}^N X_i \cdot Y_i$$

$$\text{Find}(A_0, A_1) \rightarrow \begin{bmatrix} \frac{-\sum_{i=1}^N X_i \cdot \sum_{i=1}^N X_i \cdot Y_i + \sum_{i=1}^N Y_i \cdot \sum_{i=1}^N (X_i)^2}{\left[-\left(\sum_{i=1}^N X_i \right)^2 + \sum_{i=1}^N (X_i)^2 \cdot N \right]} \\ \frac{\left(-\sum_{i=1}^N X_i \cdot \sum_{i=1}^N Y_i + \sum_{i=1}^N X_i \cdot Y_i \cdot N \right)}{\left[-\left(\sum_{i=1}^N X_i \right)^2 + \sum_{i=1}^N (X_i)^2 \cdot N \right]} \end{bmatrix}$$

В пакеті є вбудовані функції для обчислення параметрів регресії. Виводимо їх, порівнюємо з обчисленими за отриманими формулами:

$$a_0 := \text{intercep}(X, Y); \quad a_1 := \text{slope}(X, Y); \quad (3)$$

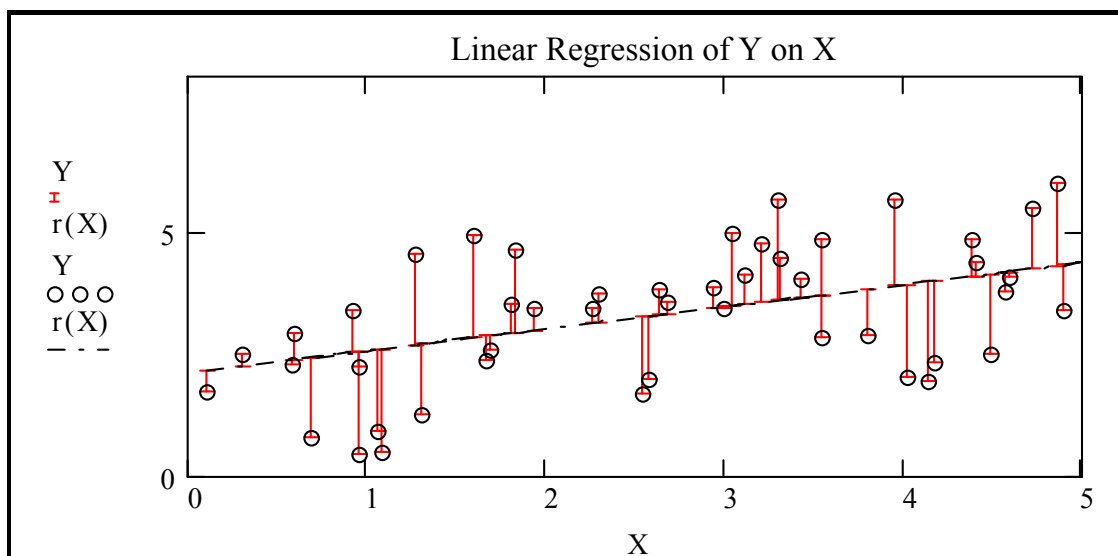
Обчислено за вбудованими функціями: $a_0 = 2.139$; $a_1 = 0.453$;

за нашими формулами: $A_0 = 2.139$; $A_1 = 0.453$;

точні значення, що ми "сховали" у модель: $A_0d = 2$; $A_1d = 0.5$.

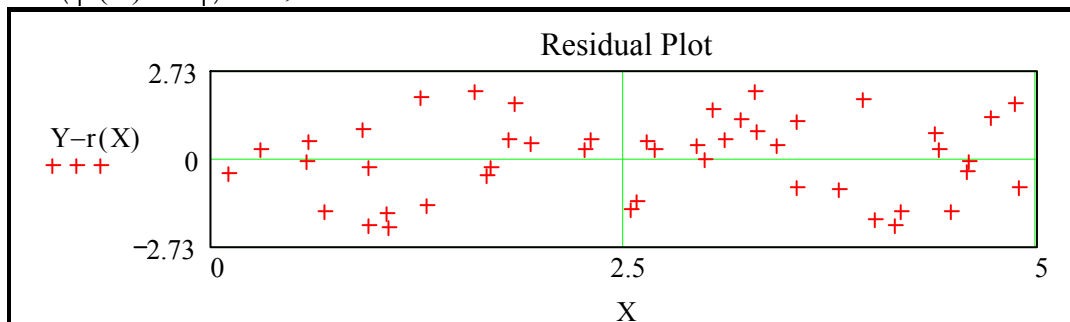
Введення розкиду $S \equiv 5$ Рівняння регресії: $r(x) := (A_0 + A_1 \cdot x)$.

Якщо ви вивчаєте цей матеріал на комп'ютері, то можете проекспериментувати: натисніть декілька раз розкид S в моделі даних (1). Так створюються нові реалізації випадкових даних. Можемо бачити, що значення параметрів регресії кожен раз децю змінюються. Можете змінювати величину розкиду - при бажанні можна буде виявити, що значення параметрів регресії стануть менш точними. Порівняйте ці результати з A_0d і A_1d - параметрами "істинної" залежності.



Наша модель даних (1) - лінійна залежність "засмічена" випадковими числами, що імітують шуми вимірювання та інші випадкові фактори. Будуємо графік регресії і вхідних даних. Увага! Графіки регресії взято з електронної книги - "Ресурсного центру" пакету Mathcad і залишено в оригінальному вигляді. Користуйтеся переважно фірмовими засобами допомоги і навчання. Масштаб для графіка помилки регресії:

$$\text{scale} := \max(|r(X) - Y|) \cdot 1.1 ; \text{scale} = 2.732$$



Вище подано графік так званих "залишків" - відхилень даних Y від середнього, що не пояснюються регресією, тобто впливом параметра X . Ці залишки можуть бути спричинені: помилками вимірювання, впливом ще якихось факторів і, взагалі, властивістю самого процесу.

1.2.2. Міра статистичного зв'язку між даними - кореляція

Задамо собі таке питання - а що, як статдані X і Y незалежні, не мають ніякого зв'язку. Формули регресії завжди видадуть якісь параметри. Інакше кажучи, як оцінити ступінь зв'язку між даними, як оцінити точність регресії? Введемо ще одну міру статистичного зв'язку між даними - кореляцію, точніше - **коваріацію (Covariance)**. Запишемо вираз для коваріації і відповідну вбудовану функцію:

$$s_{xy} := \frac{1}{N} \cdot \left[\sum_i (X_i - \text{mean}(X)) \cdot (Y_i - \text{mean}(Y)) \right] ; \quad s_{xy} = 3.808$$

Формально це **середнє арифметичне від добутку центрованих парних випадкових X та Y** . Ця величина не дуже зручна для оцінки лінійного зв'язку бо залежить ще від дисперсій X та Y . Нагадаємо їх для повторення:

$$s_x := \sqrt{\frac{1}{N} \cdot \left[\sum_i (X_i - \text{mean}(X))^2 \right]} ; \quad s_y := \sqrt{\frac{1}{N} \cdot \left[\sum_i (Y_i - \text{mean}(Y))^2 \right]}$$

Порівнюємо значення отримані за цими формулами та за вбудованими функціями.
 $\sigma_x := \text{stdev}(X) ; \sigma_x = 2.898 ; s_x = 2.898 ; \sigma_y := \text{stdev}(Y) ; \sigma_y = 1.895 ; s_y = 1.895$

Звичайно використовують нормовану величину - **коефіцієнт кореляції (Correlation coefficient)**, що змінюється в діапазоні $-1 < rk < 1$ (-1 - від'ємна, $+1$ - додатна кореляція, 0 - нема кореляції).

$$rk := \frac{s_{xy}}{s_x \cdot s_y} ; \quad rk = 0.693 ; \quad R_{xy} := \text{corr}(X, Y) ; \quad R_{xy} = 0.693$$

Тепер можемо сформулювати практичне правило для побудови регресії: обчисліть спочатку коефіцієнт кореляції даних, якщо він близький до нуля, то дані або незалежні або дуже засмічені помилками вимірювань чи збуреннями.

Зв'язок між регресією, кореляцією та дисперсіями легко вивести самостійно, подивившись на отримані вище вирази для A_0 , A_1 , або знайти в підручниках з статистики. *Перевірємо, бачимо, що результати обчислень збігаються.*

$$A_{k1} := \frac{s_{xy}}{s_x^2} ; \quad A_{k0} := \text{mean}(Y) - \frac{s_{xy}}{s_x^2} \cdot \text{mean}(X) ; \quad (4)$$

за формулами (4): $A_{k1} = 0.453$; $A_{k0} = 2.139$;

за вбудованими функціями (3) $a_1 = 0.453$; $a_0 = 2.139$;

за формулами (2): $A_1 = 0.453$; $A_0 = 2.139$.

1.2.3. Як і коли використовувати статистичні характеристики

У світовій науковій і не тільки науковій літературі багато писалося і пишеться відносно зловживання або некоректного вживання статметодів. У совковій науці, особливо у економічних і гуманітарних напрямках статистичні методи були засобом прикриття пустих робіт. Зрозуміло чому у радянські часи у нас не могло бути сильних статистиків - для цього треба мати **досить статданих** і бути **досить незалежним у висновках**.

Беллман у бібліографії до розділу книги "Процеси управління з адаптацією" [3] подав поему "Гайавата ставить експеримент" написану **Кендалом**. Як інструкцію, (з перших рук!), з застосування статистичних методів наведемо уривки з цієї поеми. Багато прочитати повністю не тільки цю поему, але й роботи Беллмана, Фішера, Кендала. Це - **першоджерела**.

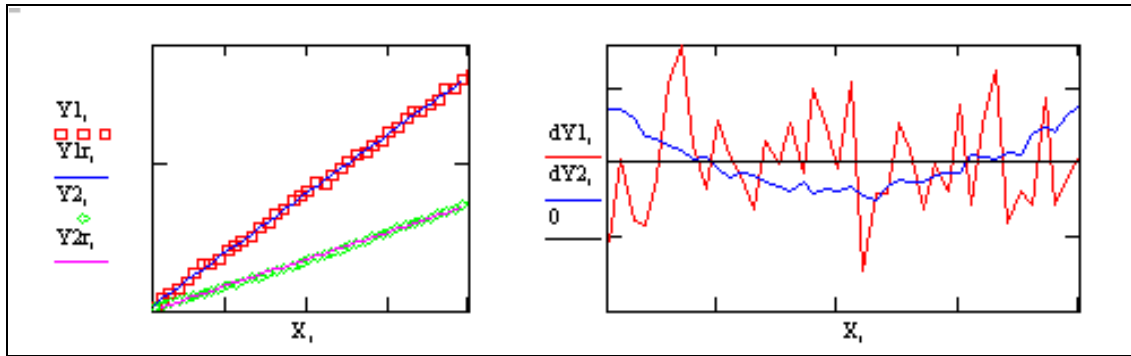
4. Но упрямые индейцы
Возразили Гайавате,
Что они не понимают
Столь туманных рассуждений.
Им совсем неинтересен
Результат его попыток.
И они предполагают,
Что охотник должен метко
В цель стрелять. А если будет
Он впустую тратить стрелы-
Должен сам за них платить.

5. Раздражённый Гайавата
Стал цитировать обильно
Р.А. Фишера и Итса,
Приводит работы Финни,
Книги Кемпторна Оскара,
Главы Кокрана и Кокса,
Андерсена и Банкрофта.
Он взывал к авторитетам,
Убеждая несогласных,
Что в стрельбе всего важнее
Не прямое попадание,
А научно безупречный
Статистический подход.

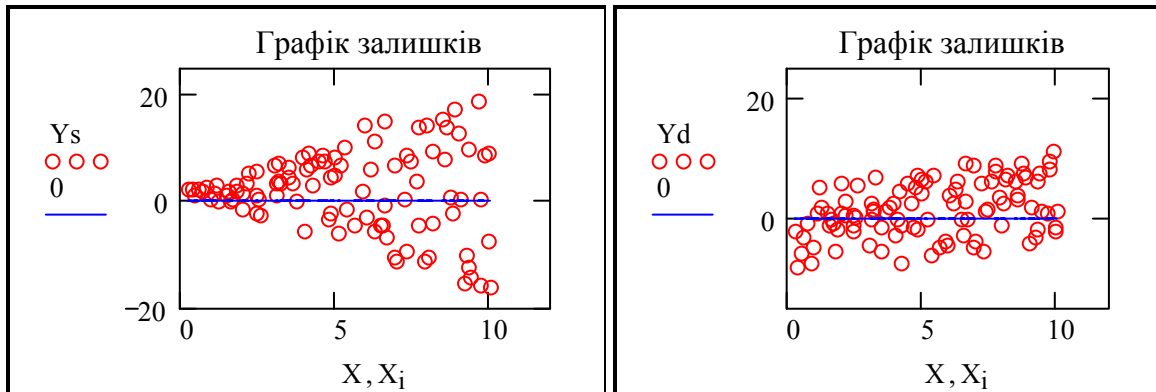
12. Тем не менее индейцы,
Не поверившие цифрам,
Отобрали у героя
Лёгкий лук его и стрелы
И сказали, что, возможно,
Гайавата в самом деле
Выдающийся статистик,
Но при этом совершенно
Бесполезен как стрелок.
Что ж касается дисперсий,
То какой то грубый неуч
Произнёс **такое** слово,
Что его сказать по чести,
В статистическом издании
Я не смею повторить.

Ось як починає книгу "Часові ряди" Кендел: "Час, можливо, є найбільшою таємницею в таємничому всесвіті. Але його таємнича природа в цій книзі нас, на щастя, не цікавить", а так продовжує: "Але, якщо маленький хлопчик обстрілює маятник го-рохом нерегулярним чином, то його рух буде збурено".

Розглянемо стисло деякі особливості і правила аналізу даних. Почнемо з аналізу статистичних залежностей. Аналіз залишків - важлива неформальна частина аналізу даних. Нижче подано дві досить близькі до детермінованих залежностей. "На око" вони лінійні, але, якщо побудувати збільшені графіки, то можна бачити, що залишки в одному випадку - випадкові, в іншому - досить детерміновані. Висновок - у другому випадку **залежність нелінійна**, потрібні інші моделі регресії.



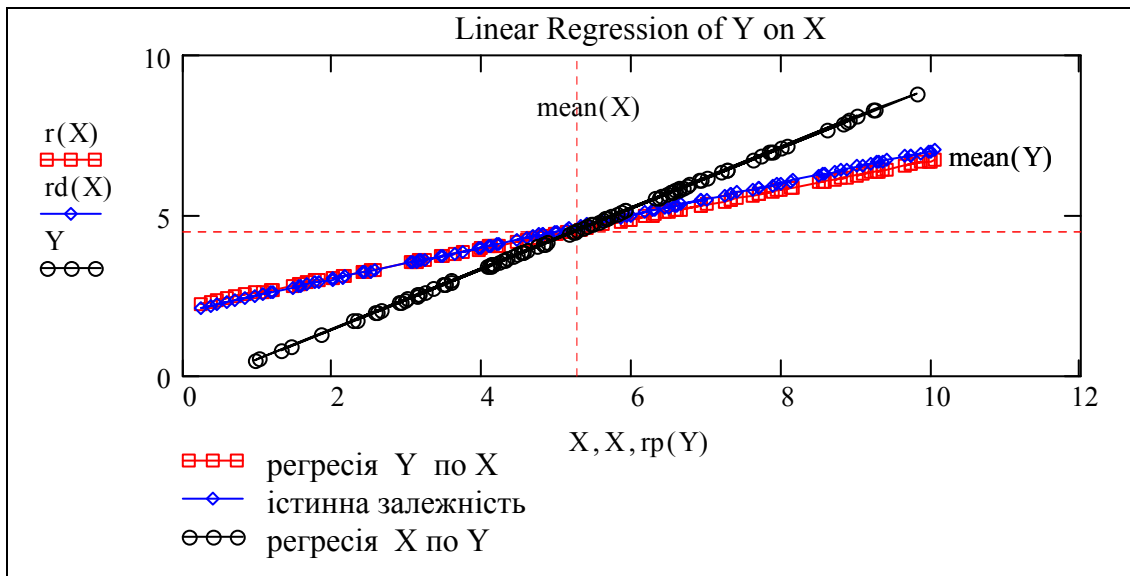
Розглянемо ще два випадки - 2) розкид залишків збільшується з ростом X , це називається гетероскедастичність - треба враховувати це в моделі регресії; 3) розкид має лінійний тренд по X - це ознака наявності ще одної (крім X) незалежної змінної.



Розглянемо ще одна особливість регресії. В деяких книгах стверджується, що *лінія регресії наближує істинну функціональну залежність між y та x* , але це **зовсім не так**. Обчислимо і побудуємо регресії Y по X , X по Y та істинну залежність.

$$ao := \text{intercept}(X, Y) ; a1 := \text{slope}(X, Y) ; bo := \text{intercept}(Y, X) ; b1 := \text{slope}(Y, X) ;$$

$$r(x) := ao + a1 \cdot x ; \quad rp(y) := bo + b1 \cdot y ; \quad rd(x) := A0d + A1d \cdot x .$$



В загальному випадку "пряма" і "зворотна" регресії не збігаються. Однак вони обов'язково повинні перетинатись в точці $(X_{\text{середнє}}, Y_{\text{середнє}})$. Якщо шуми вимірю-

вання центровані, то істинна залежність теж проходить через цю точку. Але для кореляції завжди має місце: $\text{corr}(Y, X) = 0.695$; $\text{corr}(X, Y) = 0.695$.

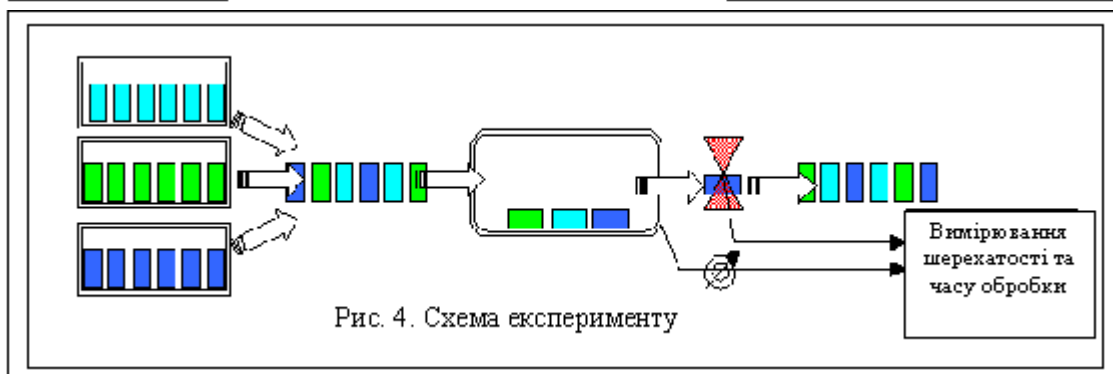
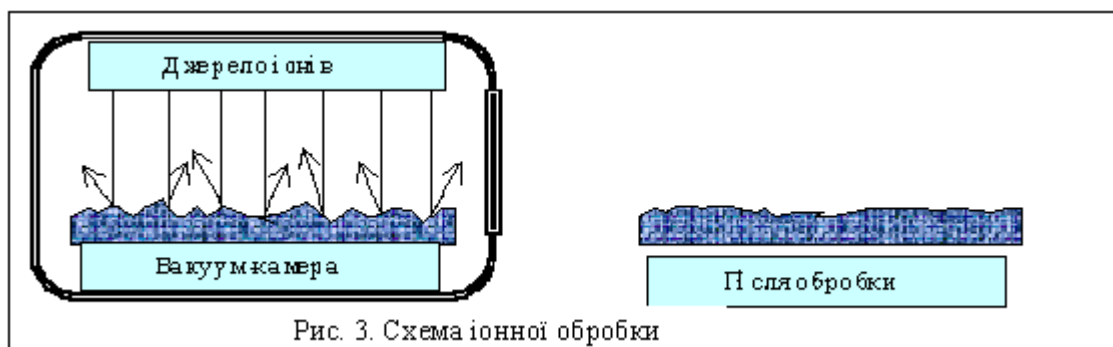
Завдання. Покажіть самостійно, чому в загальному випадку ці регресії розрізняються, і при виконанні яких умов вони збігаються.

Дивлячись на ці графіки запишемо собі в пам'ять: **побудова лінії регресії Y по X має за мету мінімізацію квадратів залишків "по напрямку Y"**, а тим самим оптимізацію прогнозування Y по X. Навпаки ж, лінія регресії X по Y мінімізує квадрати залишків "по напрямку X" і оптимальна для прогнозування значень X по Y.

Ще одне останнє зауваження: перш ніж обробляти дані, треба ретельно перевірити процедури їх отримання, вимірювання. За результатами такого аналізу дані слід **розширювати за типами об'єктів, коректувати, вносити поправки на час, сезонність (тренди), відкидати явно некоректні дані (робастне оцінювання)**. Останній пункт відноситься і до анекдотичних випадків, наприклад, в даних про зріст замість 1,82 м записано 18,2м. Розглянемо конкретний приклад.

1.2.4. Приклад розширення даних

Постановка задачі. В неформальному плані це, наприклад, така задача. В цеху 234 на ділянці **іонної шліфовки** пластин проведено дослідження залежності шорсткості (шерехатості, шерхливості, шершавості, або по-простому - мікронерівностей). Згідно з гіпотезою здорового глузду шорсткість зменшується з ростом часу обробки. Але, як саме залежить якість поверхні від часу обробки, аналітично визначити неможливо через відсутність точних математичних моделей процесу обробки. Експериментальне визначення висоти нерівностей ускладнюється їх малими розмірами і впливом збурень (рис.3).



Проведено **експеримент**, в якому тривалість часу обробки $T_{об}$ змінювалась (точність реєстрації часу обробки - 1%). **Нерівності вимірювались з середньою випадковою помилкою $\pm 25\%$ і систематичною - 0%.**

Було проведено 90 експериментів, для чого **систематично-випадково** було відібрано 90 деталей-заготовок: по 30 штук від кожного з трьох постачальників заготовок, ці

30 штук вибирались з упаковки (на 100 штук) випадковим чином.

Час обробки деталі задавався з дискретного ряду значень 5, 10, 15, 20, 25, 30 хвилин так, щоб мати по 15 деталей на кожне значення часу обробки і реалізувався з вимірюваними відхиленнями (рис.4).

Проблема. За результатами статистичної обробки даних виявилось, що регресія (статистична залежність між парними статистичними даними) "шорсткість - час обробки" має антиінтуїтивний нахил - **чим більший час обробки, - тим більша шорсткість?**

Задача. Для заданого варіанта вхідних даних обчислити статистичні міри розташування (середні), розсіювання (дисперсії), зв'язку (параметри регресії і кореляції). Треба провести аналіз для виявлення дійсних залежностей між шорсткістю і часом обробки.

Подаємо дані для такого аналізу - три реалізації (комп'ютерні імітації) випробування, наприклад, в перший, другий, третій день при абсолютно однакових умовах.

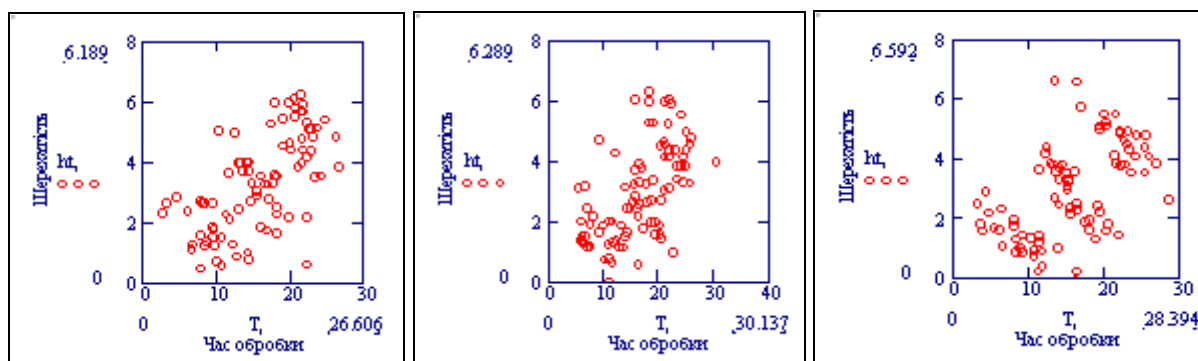


Рис. 5. Залежності "час обробки-шорсткість" (три серії по 90 випробувань)

Статистики високого рангу дозволяють собі вживання терміну "окомірний аналіз". Не завжди він допомагає, але іноді без критеріїв довіри Фішера і Колмогорова можна побачити дійсну залежність. Перш ніж перейти до регресії, спробуйте і Ви подивитись неупереджено на рис. 5.

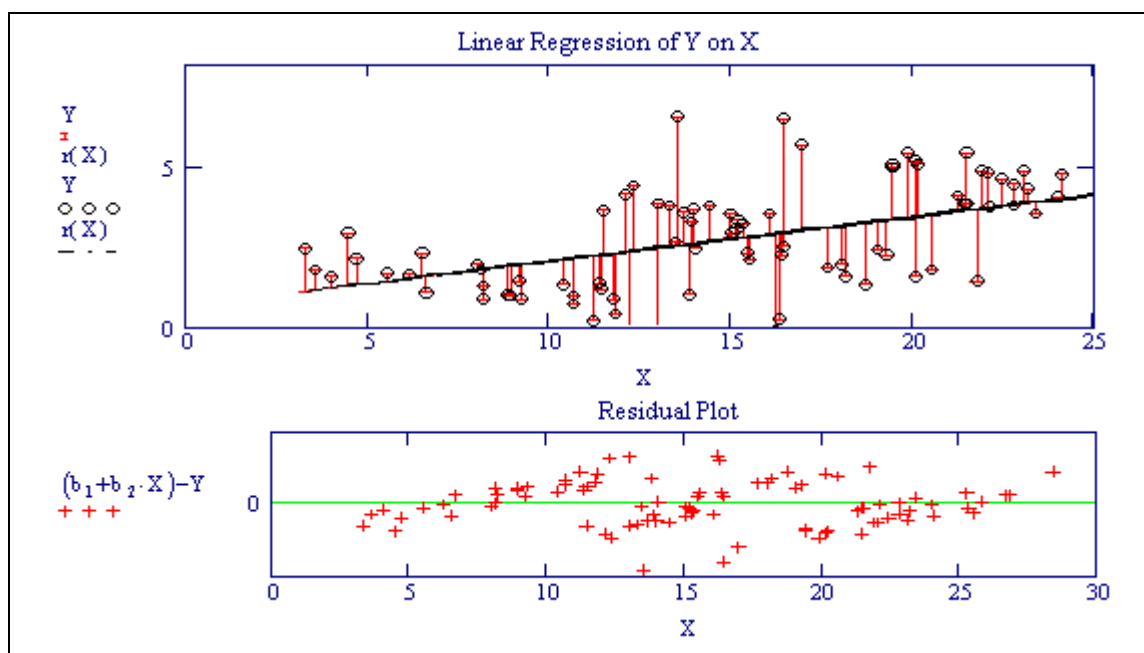


Рис. 6 Регресія та графік "залишків" для нерозшарованих даних

Будуємо графік регресії та залишків на фоні даних (рис. 6, увесь цей приклад повністю зроблено в середовищі пакета і використано як шаблони-фрейми приклади з "Ресурсного центру" математичного пакета. Ми скопіювали їх та підставили в них наші дані).

Бачимо, що результати, як у Гайавати, з точки зору статистики, бездоганні. Все в порядку - і дисперсії і кореляції. Висновок - чим більший час обробки, тим гіршою стає поверхня пластини? Для коментарів саме в таких ситуаціях вживається нестатистична лексика.

Перш ніж перейти до відповіді, подивіться тепер на рис. 4 і 6, спробуйте побачити дещо спільне. В часи Піфагора в доказах теорем з геометрії було тільки одне слово "дивись" та відповідне креслення. Подивіться на рис. 7 - там добавлено відносно рис. 6 тільки три лінії.

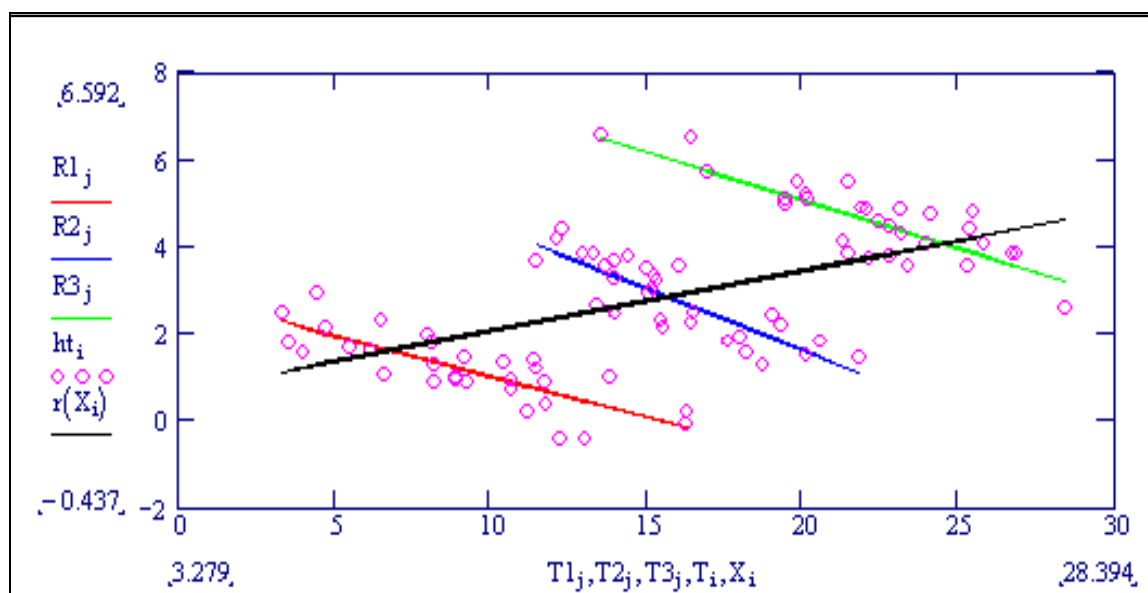


Рис. 7 Регресії для нерозшарованих та розшарованих даних

Дивимось на те, що бачимо, а також на те, що бачили (рис. 3) - схему організації експерименту. Заготовки, які ми вважали ідентичними, бралися з трьох ящиків - від різних постачальників і мали **різний середній рівень** початкової шорсткості.

Якщо будувати регресії для заготовок кожної партії окремо, то отримуємо регресії з приблизно однаковим від'ємним нахилом - **чим більший час обробки, тим менша висота мікронерівностей**. В літературі можна знайти і більш "прокольні" випадки некоректного застосування статистичних методів. Читайте першоджерела.

1.3. Міри групування даних - частотні розподіли

1.3.1. Постановка задачі

Чи знаєте ви, з яких актуальних практичних задач почалася десь 300 років тому сучасна статистика (статистика взагалі існувала завжди)? - Знаєте - поставте собі 7 балів за ерудицію. Ви можете сказати, що людство слідувало за врожаями, цінами, темпера турями, дощами. Так, але першою з актуальних задач для статистики в Англії була задача пошиву одягу для армії. Форма, зрозуміло, шилася масово, не на замовлення. Скільки ж треба було пошити кожного розміру форми, щоб не було дефіцитів та надлишків?

Уявіть собі - у вас на руках первинні дані - списки рекрутів з розмірами одягу та взуття. Що ви з ними зробите (ми вже вивчили середні, дисперсії, кореляції)?

Обчислите середній розмір - пошиєте середню форму? - Не те.

Обчислите ще дисперсію - пошиєте "максимальну" і "мінімальну" форму? - Теж не те. Обчислите кореляцію між розміром голови та ніг? - Зовсім не те.

Для розв'язання нашої задачі первинні дані необхідно **групувати**. Візьмемо репрезентативну (поки не знаємо, що це означає, але неважко домислити) **вибірку** на 1000 чоловік. Визначимо діапазон - максимальний (160) і мінімальний зріст (200). Розіб'ємо цей діапазон на інтервали, наприклад: 160==165==170==175==180....==200. Обчислимо, по скільки рекрутів попадає в кожний інтервал. Нехай ми розбили діапазон на 8 інтервалів (між іншим, скільки існує розмірів одягу для дорослих?).

В результаті замість списку на 1000 чоловік ми отримаємо таку табличку:

	0	1	2
0	"ріст"	"число_ел"	"відсоток"
1	160	40	0.04
2	170	222	0.222
3	180	480	0.48
4	190	218	0.218
5	200	40	0.04

Ми не тільки стиснули дані, але і зробили їх більш інформативними відносно поставлених цілей. Дійсно, якщо чисельність війська 100000 чоловік то форми розміром 170 слід пошити (дивись таблицю): $222 \cdot \left(\frac{100000}{1000} \right) = 22200$ комплекти.

Узагальнимо і формалізуємо задачу. Створимо дані для обчислювальних експериментів. Імітуємо статистичні дані вбудованою функцією

$X := \text{rnorm}(10024, 175, 5 \cdot \text{Натисни} \cdot \text{натисни_F9} \cdot \text{натисни})$,

що означає: створити масив з 10024 чисел, що мають гаусовий (нормальний) розподіл з середнім 175 і дисперсією 5. Змінна "натисни" вводиться далі, вона призначена для переобчислення вибірки.

В порядку повторення матеріалу розділів 1.1 та 1.2 перевіримо цю функцію:

$\text{length}(X) = 10024$; $\text{mean}(X) = 175.036$; $\text{stdev}(X) = 4.983$.i := 1..length(X)

1.3.2. Побудова емпіричного частотного розподілу (гістограми)

Дано: N-мірний набір даних, наприклад, ріст - вектор (масив) чисел. Проведемо попередній аналіз даних - визначимо мінімум, середнє та максимум.

$\text{min}(X) = 156$

$\text{mean}(X) = 175$

$\text{max}(X) = 193$

Задаємо систему (вектор) інтервалів, це трохи мистецтво: змінили число інтервалів на одиницю, трохи зсунули границі інтервалів - і гістограма стає зовсім іншою. Існують такі правила:

1. **Кількість інтервалів** повинна бути такою, щоб в середньому в інтервал попадало не менше 7-ми елементів, тобто:

$$\text{Кільк_інт} = \frac{\text{Розмір_вибірки}}{\text{більше_7}} ; \quad \text{Кільк_інт} := \text{ceil}\left(\frac{\text{length}(X)}{64}\right); \quad \text{Кільк_інт} = 157$$

2. **Границі системи інтервалів** беруться так, щоб до них входили мінімальне і максимальне значення вибірки;

3. **Розміри інтервалів** роблять звичайно однаковими, але, коли вибірка суттєво нерівномірно розташована по діапазону, використовуються системи нерівних інтервалів.

З урахуванням цих вимог вираз для вектора інтервалів буде:

$$q := 1 \dots \text{Кільк_інт} ; \quad \text{In}_q := \frac{(\max(X) - \min(X)) \cdot 1.2}{\text{Кільк_інт}} \cdot (q - 1) + \min(X) \cdot 0.99$$

$\text{In}^T =$	1	2	3	4	5	6	7	8	9
	1	154.017	154.301	154.585	154.869	155.153	155.437	155.721	156.005
									156.289

Необов'язкові множники (1.1 та 0.99) використовуються для точного налаштування системи інтервалів. Для побудови гістограм звичайно використовуються вбудовані функції пакетів.

1. Почнемо з "саморобної" програми побудови гістограми.

$\text{CatWork} :=$ <pre> for j ∈ 1..Кільк_інт n_j ← 0 for i ∈ 1..length(X) j ← 1 while (X_i ≥ In_j) · (X_i < In_{j+1}) = 0 j ← j + 1 "endwhile" n_j ← n_j + 1 n </pre>	Тут n_j - кількість елементів вибірки, що попадають в j-ий інтервал. Спочатку ми обнулюємо їх. Потім для кожного елемента вибірки (цикл for) знаходиться (цикл while) такий інтервал, що: $\text{In}_j \leq X_i < \text{In}_{j+1}$ по виходу з циклу while значення n_j збільшується на 1. Програма повертає вектор n - це власне і є значення гістограми.
--	---

Перевіримо коректність системи інтервалів і програми:

$$\sum_q \text{CatWork}_q = 10024 ; \quad \text{length}(X) = 10024$$

тобто, усі дані попали в систему інтервалів

Тепер отримаємо гістограму за допомогою вбудованої функції

$\text{hist}(\text{вектор_інтервалів}, X_ \text{вектор_даних})$

Ця функція "Returns a vector representing the frequencies with which values in X fall into the intervals represented by intvls. intvls can be a vector of intervals such that the i th element of the returned vector is the number of points in data falling between the i th and $(i+1)$ th element of intvls". В гуманних цілях подаємо: "функція повертає вектор частот, з якими величини із (вектора даних) X попадають в інтервали, що подані вектором інтервалів. Вектор інтервалів може бути таким, що кожен елемент вихідного вектора - це кількість вхідних даних, що попала між i -тим та $(i+1)$ -им елементами вектора інтервалів".

Підставляємо у цю функцію наші інтервали та дані: $\text{EmpR} := \text{hist}(\text{In}, X)$

Обчислюємо довжину інтервалу (у сантиметрах росту). - Ми множимо вираз для інтервалу на 1.2, щоб із певним запасом охопити діапазон даних. В наших модельних даних "схований" нормальний (=Гаусовий) розподіл з середнім 175 (сантиметрів) і дисперсією 5. Ми його дискретизуємо розподіл розбиваємо на стовпчики, площа яких дорівнює $\text{dnorm}(\text{In}_q, 175, 5) \cdot d\text{In}$. Потім множимо кожний стовпчик на розмір вибірки.

$$d\text{In} := \frac{\max(X) - \min(X)}{\text{Кільк_інт}} \cdot 1.2 ; \quad \text{TeoR}_q := \text{dnorm}(\text{In}_q, 175, 5) \cdot d\text{In} \cdot \text{length}(X).$$

Виводимо таблиці, будуємо графіки. Кнопка "натисни" (курсор на кнопку, натиснути F9) викликає обчислення нової реалізації вектора випадкових даних X. Там деś вище, за полами сторінки є функція-генератор випадкових чисел із заданим розподілом імовірностей). По натискайте її і подивіться як змінюється емпіричний розподіл - гістограма.

qq := 1 .. Кільк_інт - 1 ; Натисни ≡ 1

$$\text{In}^T =$$

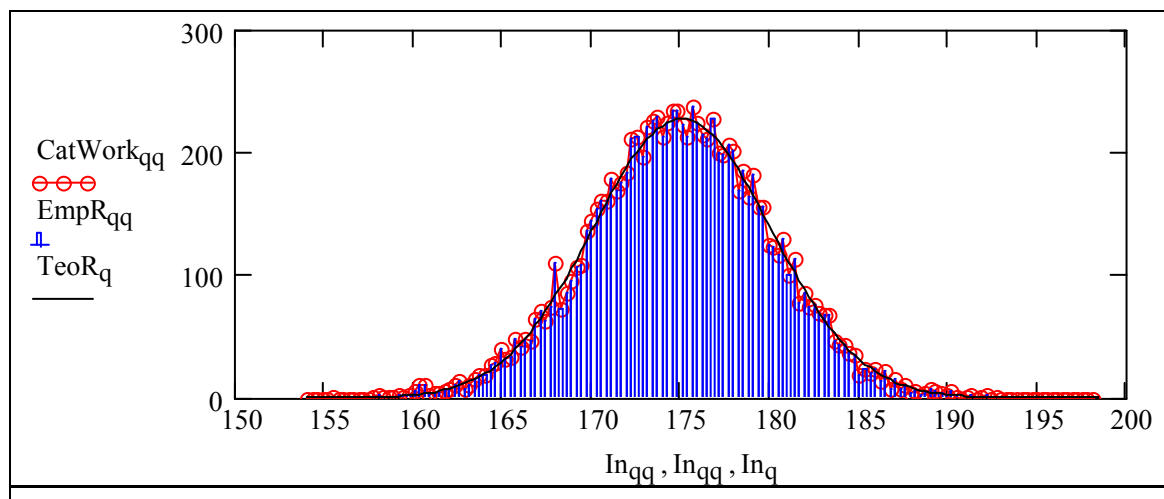
	1	2	3	4	5	6	7	8	9	10
1	154	154	155	155	155	155	156	156	156	157

$$\text{CatWork}^T =$$

	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	1	0	0	0	0

$$\text{EmpR}^T =$$

	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	1	0	0	0	0



$$\min(X) = 155.573 ; \quad \text{mean}(X) = 175.036 ; \quad \max(X) = 192.729$$

$$\text{Контроль: } \text{length}(X) = 10024 ; \quad \sum_q \text{TeoR}_q = 10023.867 ? ; \quad \sum_{qq} \text{EmpR}_{qq} = 10024$$

Бачимо, що:

1. Результати роботи саморобної програми і вбудованої функції - збігаються.
2. Система інтервалів задана коректно - дивись контроль - усі елементи попадають в інтервали. (Якщо не так - наша програма не працюватиме, тому що не знайде, куди покласти елемент, що виходить за межі діапазону).
3. Число стовпців гістограми на одиницю менше числа границь інтервалів. - **Чому?**
4. Теоретичний розподіл зсунутий на один інтервал відносно гістограми. - **Чому?**
3. Нарешті отримуємо гістограму за допомогою вбудованої функції

histogram(n_кількість_інтервалів, вектор_даних)

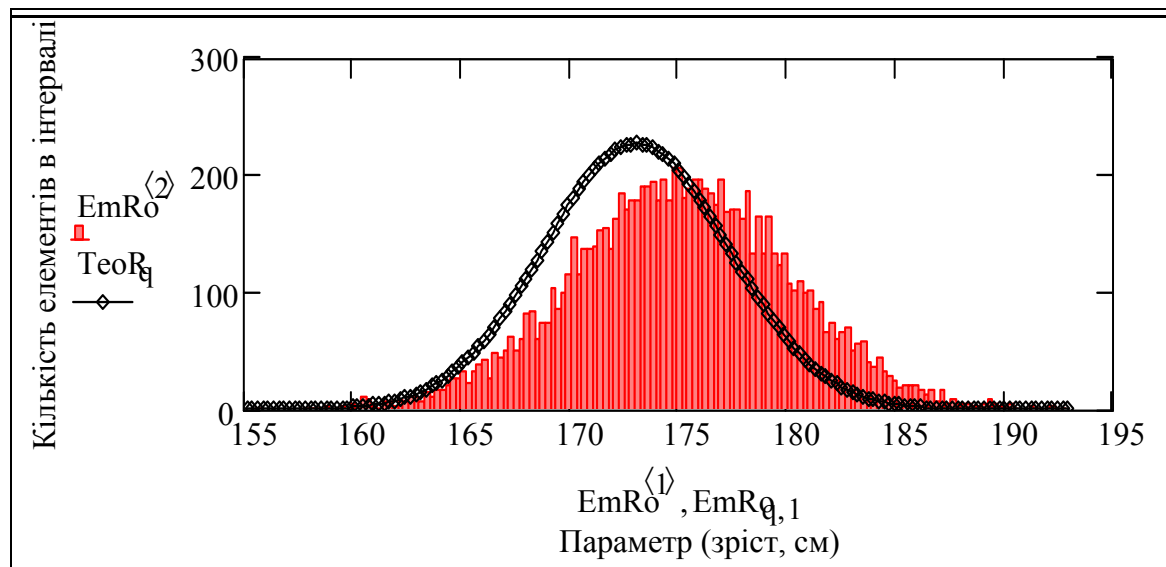
УВАГА! Ця функція є тільки у версіях пакету ≥ 2001 .

Ця функція "Returns a matrix with two columns. The first column contains midpoints of the n subintervals of the range of equal length. The second column is identical to hist(n, A), and hence the resulting matrix has n rows". В гуманних цілях подаємо: "повертає матрицю з двох стовпців. Перший стовпець містить **середини інтервалів** однакової довжини. Другий дає таку ж гістограму, як і функція hist(n, A), однак вихідна матриця має n стовпців".

Що дає ця функція?

1. Не треба мудрувати з інтервалами - систему інтервалів зробить сама функція;
2. Не треба турбуватись про те, що у гістограмі стовпців на один менше, ніж інтервалів;
3. Не треба зсувати гістограму на "півкроку" - система інтервалів подає саме середини інтервалів, а не їх праві чи ліві границі.

Зауваження. Саме на цих трьох пунктах вимогливі викладачі ловили не дуже вимогливих (до себе) студентів. А тепер все це робить комп'ютер. Підставляємо у цю функцію число інтервалів та дані: $EmRo := \text{histogram}(\text{Кільк_інт}, X)$. Поставте курсор на цю "кнопку" **натисни_F9 $\equiv 1$** , по натискайте F9. Ви побачите, що іноді гістограма точно збігається з теоретичним розподілом, іноді - зовсім несхожа.



$EmRo^T =$		1	2	3	4	5	6	7	8	9	10	11	12
1	155.1	155.4	155.6	155.8	156.1	156.3	156.6	156.8	157.1	157.3	157.5	157.8	
2	0	0	1	0	0	0	0	0	0	0	0	0	1

Бачимо, що:

- 1) вбудована функція економно вибрала діапазон вектора інтервалів;
- 2) значення вектора інтервалів - це середини, а не границі інтервалів;
- 3) число стовпчиків гістограми дорівнює числу інтервалів.

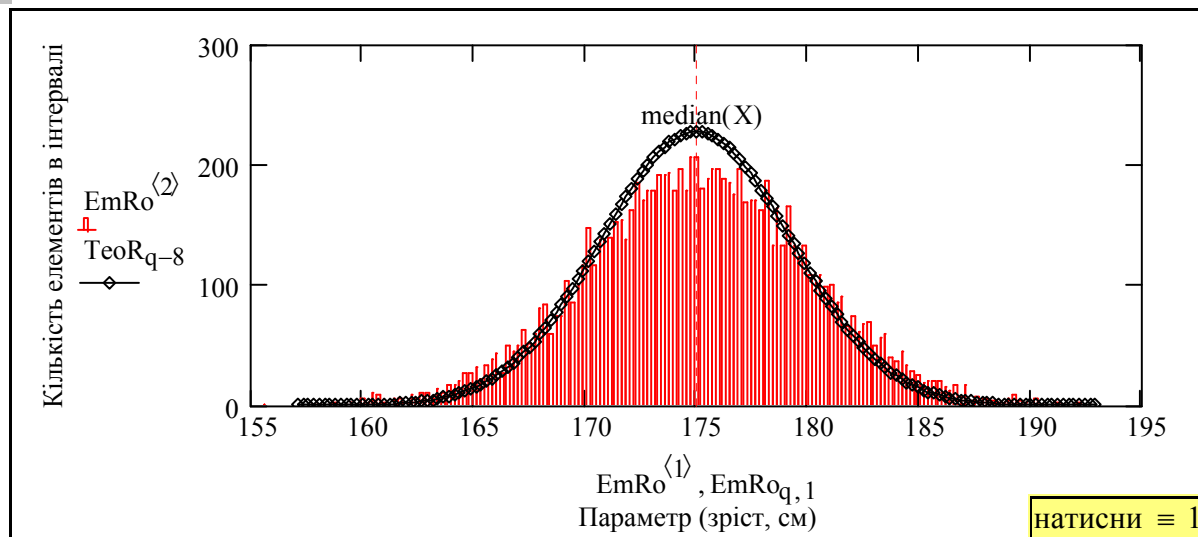
Завдання. Зробіть програму для визначення діапазону, в який попадає задана частина вибірки (масиву даних), наприклад, 80%, 90%, ... Самостійно довизначте задачу.

1.3.3. Визначення медіани та моди емпіричного розподілу

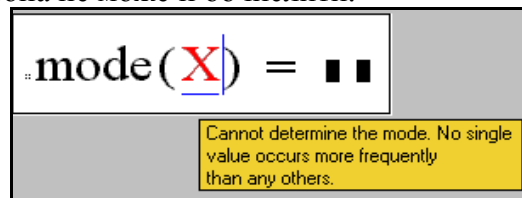
Ми вже визначали моду і медіану на неаггерованих даних (розділ 1.1). Ці види середнього є природними саме для частотного розподілу. Визначимо їх спочатку за допомогою "саморобних" програм, потім - за допомогою вбудованих функцій.

Медіана(Hi, In) :=	<pre> ORIGIN ← 1 N ← length(Hi) sum₁ ← Hi₁ for i ∈ 2..N sum_i ← sum_{i-1} + Hi_i nom ← 1 while sum_{nom} ≤ 0.5·sum_N nom ← nom + 1 In_{nom} </pre>	Мода(Hi, In) :=	<pre> ma ← -10000 for i ∈ 1..length(Hi) if Hi_i > ma ma ← Hi_i nom ← i otherwise ma ← ma nom ← nom In_{nom} </pre>
--------------------	--	-----------------	---

Коментар до програм. В програмах є зайві рядки. Їх можна виключити. Знайдіть їх.
Медіана(EmpR, In) = 174.7 ; median(X) = 175 ; Медіана($EmRo^{(2)}$, $EmRo^{(1)}$) = 175 ;
Мода(EmpR, In) = 175.6 ; mode(floor(X)) = 174 ; Мода($EmRo^{(2)}$, $EmRo^{(1)}$) = 174.7 .



Зауваження. З певних причин вбудовані функції є гіршими від саморобних. Подивіться праворуч на збільшений фрагмент, де вбудована функція замість обчислення медіани поясняє, чому вона не може її обчислити.



1.3.4. Приклад розшарування даних

Знову нагадаємо - статистичні методи не більше, ніж інструмент, і вони не звільняють від необхідності думати, в першу чергу, - про джерела і методи отримання даних. В підрозділі 1.2 вже розглядався приклад "залежність шорсткості деталі від часу обробки". Частотні розподіли можуть бути бімодальними і навіть полімодальними, тобто, мати два максимуми або більше. Причини цього можуть бути різними. Одна з причин - наявність у вибірці об'єктів двох різних класів.

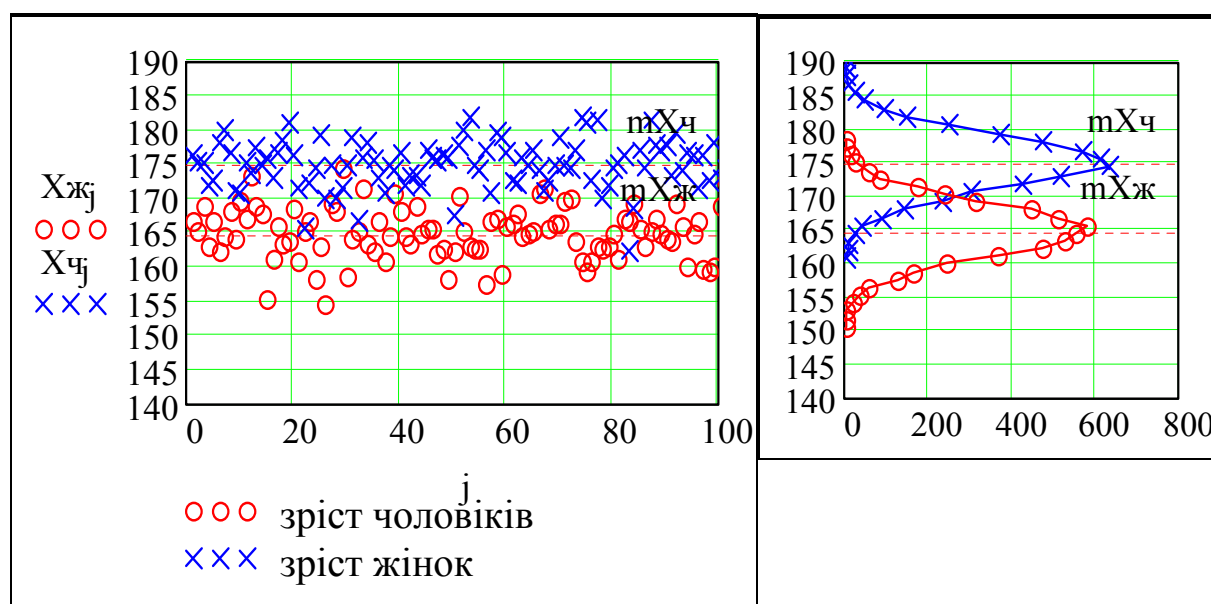
Розглянемо таку задачу: вибірка складається з рівних кількостей чоловіків і жінок. Яким буде частотний розподіл такої змішаної вибірки за ростом, якщо розподіл за ростом у вибірках тільки жінок і тільки чоловіків є гаусовим? Чи може частотний розподіл бути індикатором, що вибірка складається з елементів різних класів? - Конкретно, в яких умовах з'являються два горби (максимуми, моди) на розподілі?

Спочатку зробимо зручний модуль - "стенд" для "експериментального дослідження" задачі, потім пошукаємо аналітичне розв'язання - частотний розподіл "суміші", як функції розподілів однорідних вибірок. Подаємо далі потрібні вирази без коментарів.

```

ORIGIN := 1 ; вибірка := 5000 ; j := 1..вибірка ; КількІ = 25 ; q := 1..КількІ ;
mW := 165 sW := 4 ; mM := 175 ; sM := 4 ;
Xж := gnorm(вибірка,mW,sW)·натF9 ; Xч := gnorm(вибірка,mM,sM)·натF9 ;
mXж := mean(Xж) ; mXч := mean(Xч). Об'єднуємо дві вибірки: Xсм := stack(Xж,Xч).
EmS := histogram(КількІ,Xсм) ;
EmW := histogram(КількІ,Xж) ; EmM := histogram(КількІ,Xч) ;
inw := EmW2,1 - EmW1,1 ; inm := EmM2,1 - EmM1,1 .
EmS := histogram (КількІ ,Xсм ) ;

```



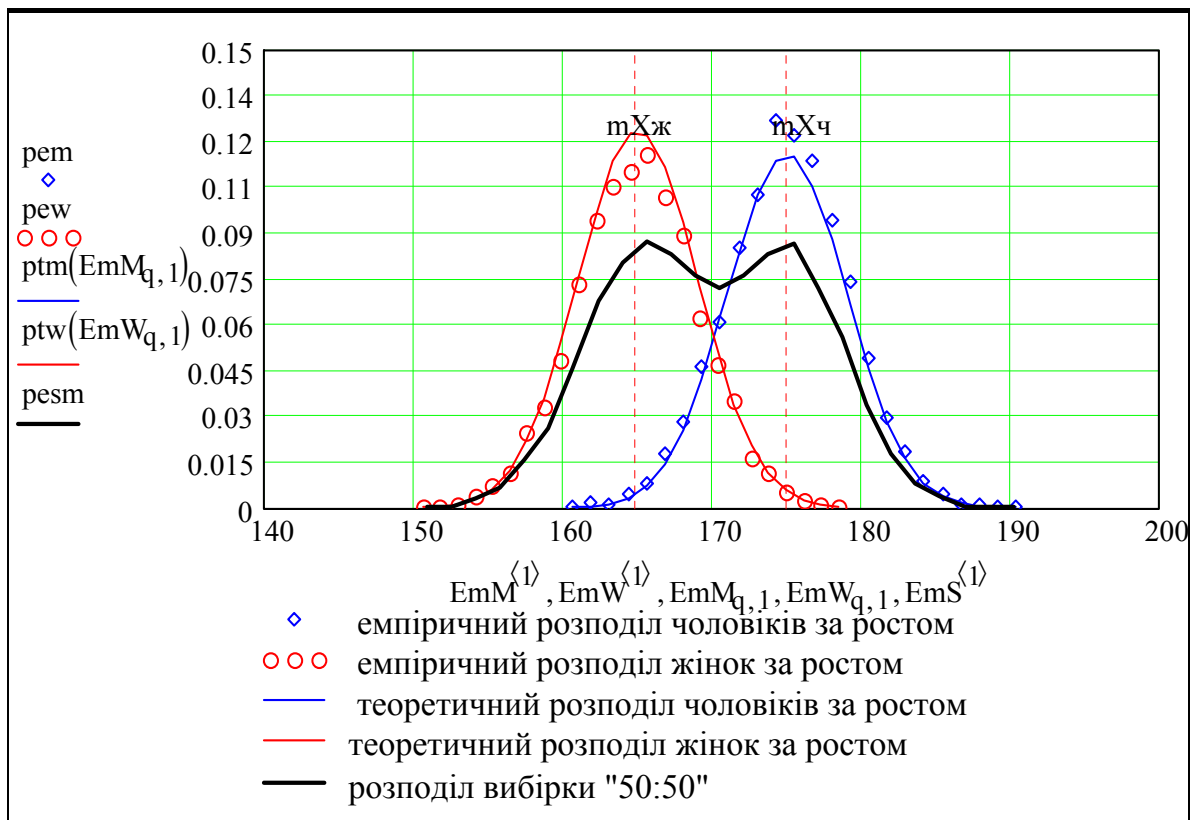
На цих графіках подано разом неагреговані та агреговані дані. Побудуємо тепер разом емпіричні частотні розподіли однорідних вибірок і суміші, а також теоретичні розподіли, що "сховані" у наших вхідних даних про зріст елементів вибірок. Нормуємо і масштабуємо теоретичні та емпіричні розподіли.

```

ptm(x) := dnorm(x,mM,sM)·inw ; ptw(x) := dnorm(x,mW,sW)·inm ;
pem :=  $\frac{EmM^{(2)}}{вибірка}$  ; pesm :=  $EmS^{(2)} \div (2вибірка)$  ; pew :=  $\frac{EmW^{(2)}}{вибірка}$  .

```

Проекспериментуйте трохи: змінюйте кількість інтервалів **КількІ = 25** і просто натискайте **натF9 = 1**. Всі об'єкти вище і нижче буде переобчислюватись.



Завдання по роботі зі стендом. Натискайте відповідну кнопку - побачите які дивні реалізації вибірки можуть **випадково** трапитись - якщо не знати першоджерела, ніколи не здогадатись, що це просто суміш двох вибірок двох гаусових розподілів. Змінійте розмір вибірки, число інтервалів, дисперсії. Зробіть висновки.

Підсумок. Експериментально отримали розподіл з іноді помітними двома модами.

Завдання. Визначте аналітично частотний розподіл вибірки-суміші через розподіл однорідних вибірок.

Зауваження. Не плутайте випадок: "вибірка = вибірка1 + вибірка2" з випадком: "вибірка = множина(елемент = елемент1 + елемент2)". Останній випадок - розподіл від суми випадкових величин з певними розподілами - має назву "**згортка розподілів**". В останньому випадку результатом згортки гаусових розподілів буде знову ж таки гаусовий розподіл з певними параметрами. Згортка розглядається в наступному підрозділі.

1.3.5. Розподіл імовірностей

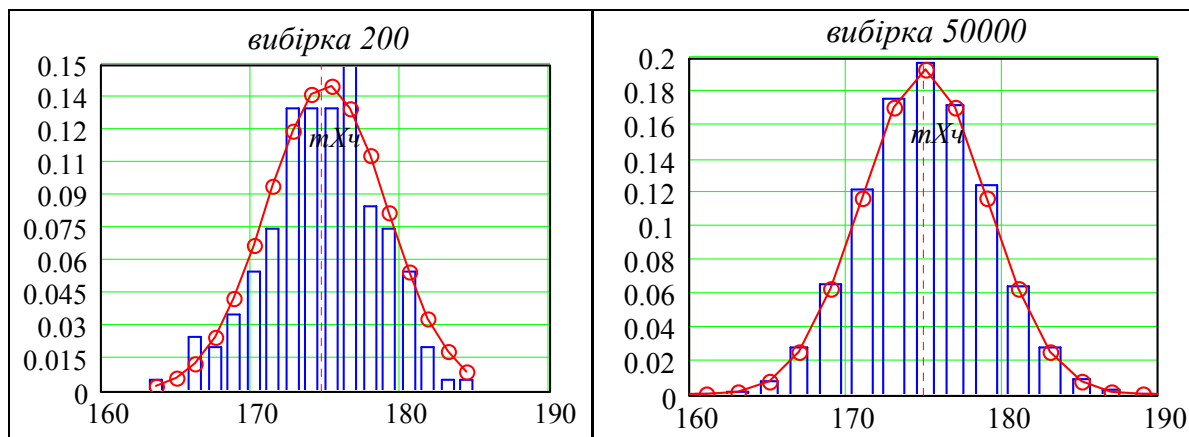
Ми повторили, узагальнили, освоїли елементи статистики, що використовуються при розв'язанні типових задач дослідження операцій. Тепер згадаємо і узагальнимо потрібні нам (знову ж таки - для дослідження операцій) елементи теорії імовірностей.

Між емпіричними розподілами відносних частот і розподілами імовірностей має місце тісний зв'язок. Часто їх не розрізняють, це ніби не є страшною помилкою, але може привести до глибоких помилок.

Під **розподілом імовірностей** можна розуміти граничну форму розподілів відносних частот при необмеженому збільшенні кількості спостережень чи експериментів. Такий розподіл прийнято називати **істинним розподілом імовірностей** для генеральної сукупності даного типу експерименту.

Вибірки з генеральної сукупності утворюють розподіл відносних частот, що являють собою наближені форми істинного розподілу імовірностей. Розподіли імовірностей можуть бути дискретними і неперервними. Вивчення розподілів почнемо з неперервних розподілів імовірностей.

Зауваження. Обережно (не довіряйте і перевіряйте) підходьте до фундаментальних істин, положень, законів. В даному випадку експеримент підтверджує існування "граничної форми розподілів відносних частот при необмеженому збільшенні кількості спостережень"- внизу подано гістограми для вибірок у 200 і 50000 чоловік. Все сходиться...



В житті зустрічаються інші імовірнісні механізми, що породжують принципово інші розподіли гіперболічні. Ці розподіли мають нескінченні дисперсії, емпіричні розподіли з ростом розміру вибірки віддаляються від теоретичних розподілень імовірностей. **Це окрема тема.** Поки пояснимо це *на словах*: якщо взяти якийсь текст, визначити частоти використання усіх слів тексту, то 1) будемо мати довгий "хвіст" із слів, що зустрінуться в тексті тільки один раз, 2) якщо побудувати розподіл у подвійному логарифмічному масштабі, то він на певній ділянці буде лінійним (гіперболічна залежність).

Так от, на відміну від розподілів з обмеженою дисперсією (статистика незалежних елементів), гіперболічні розподіли добре проявляються на відносно невеликих текстах присвячених одній темі, на текстах одного письменника. На великих вибірках, де частини тексту не пов'язані між собою, розподіли стають подібними гаусовим.

На жаль, у підручниках із статистики і теорії імовірностей гіперболічні розподіли звичайно ігноруються.

Властивості розподілів імовірностей

Далі розглядається і досліджується ряд прописних істин з теорії імовірності. Ми отримаємо робочі залежності для визначення характеристик випадкових величин через розподіли імовірностей.

Нагадаємо, що **"робочі моделі"**, "робочі залежності" - це те, що виконується, копіюється, переноситься з документа в документ. В цьому аспекті цей посібник - ланцюг програмних модулів, складність яких зростає. Ці модулі можна включати в курсові і дипломні проекти..(природно, як основу, що можна безмежно вдосконалювати і оптимізувати).

Як наскрізний приклад неперервного розподілу вибираємо так званий логнормальний розподіл. Він асиметричний, визначений на інтервалі $(0, \infty)$. Одна з інтерпретацій цього розподілу - імовірність виконання роботи в певний термін (довгий "хвіст" цього розподілу - це ненульові імовірності затягування роботи дуже на довго).

Записуємо значення параметрів: $NN := 10$ $x := 0, 0.1 \dots NN$; математичне очікування $\mu := 1$; дисперсія $\sigma := 0.6$. Записуємо вбудовані функції розподілів:

кумулятивний розподіл $P(x) := \text{plnorm}(x, \mu, \sigma)$;

щільність розподілу $p(x) := \text{dlnorm}(x, \mu, \sigma)$.

Записуємо аналітичний вираз для щільності розподілу

$$pl(x, \mu, \sigma) := \frac{1}{\sigma \cdot x \cdot \sqrt{2 \cdot \pi}} \cdot \exp \left[\frac{-1}{2 \cdot \sigma^2} \cdot (\ln(x) - \mu)^2 \right].$$

1. Сума імовірностей. Якщо задано діапазон значень випадкової величини (що може бути і нескінченним) $X_{\min} \leq x \leq X_{\max}$, то інтеграл від розподілу (щільності розподілу імовірностей) повинен дорівнювати одиниці. Чому одиниці? - випадкова величина обов'язково прийме одне із значень діапазону. Якщо сума імовірностей (інтеграл) менша одиниці - група подій не повна. А якщо більша - певні події не є незалежними. Ця властивість широко використовується для контролю і отримання оцінок.

повинно бути:

перевіримо:

$$\int_{X_{\min}}^{X_{\max}} p(x) dx = 1 ; \quad \int_0^{27} p(x) dx = 0.999935 ; \quad \int_0^{27} pl(x, \mu, \sigma) dx = 0.999935$$

2. Визначення моди. Ми вже визначали моду за гістограмою дискретного розподілу (розділ 1.1). Зробимо це для неперервного розподілу. Формально нам треба знайти максимум функції $p(x)$. Можна прирівняти до нуля похідну, можна написати програму пошуку екстремуму. Використаємо вбудований пошуковий метод `minerr(x)` - "мінімальна помилка". Цей метод однаково придатний і для пошуку екстремуму і нульових коренів. Записуємо: початкове значення: $x_0 := 1$

Given (дано) рівняння: $1 - p(x_0) = 0$

що шукаємо: `mod2 := MinErr(x_0)`

результат: `mod2 = 1.9`

3. Визначення медіани. Медіана ділить площу (щільності) розподілу навпіл.

Площа під графіком щільності розподілу - це інтеграл: $P(x) = \int_{X_{\min}}^x p(x) dx$.

Нагадуємо, що цей інтеграл називається кумулятивним розподілом. Медіану зручніше визначати на кумулятивному розподілі. Використаємо вбудовану функцію:

"корінь(функція(x) = 0, x, $X_{\min}$, $X_{\max}$)",

`med2 := root[(P(x_0) - 0.5), x_0, 0, 6] ; med2 = 2.72` ,

яка шукає (і знаходить) нульовий корінь заданого рівняння у заданому діапазоні.

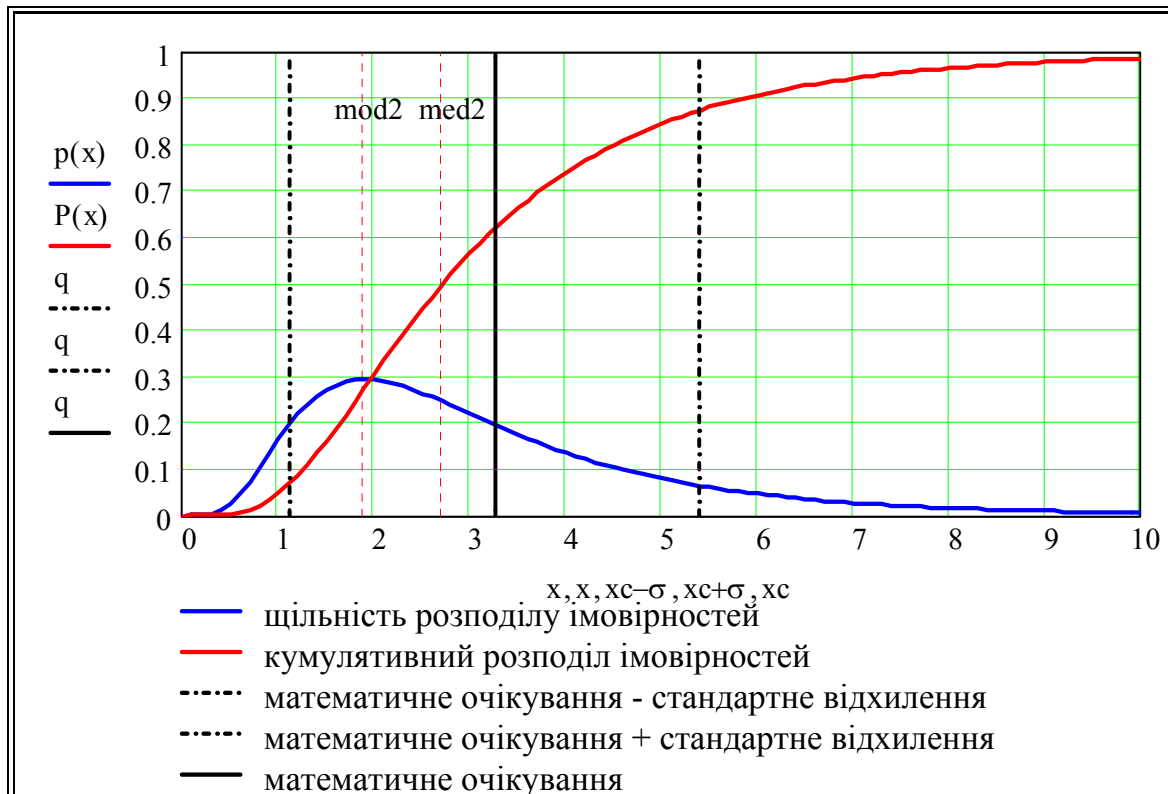
4. Визначення середньоарифметичного і стандартного відхилення. Ми вже визначили середньоарифметичне і середньоквадратичне на неагрегованих даних, і на агрегованих даних - гістограмі. Визначимо тепер ці відхилення за неперервним розподілом імовірностей. Подаємо робочі залежності. $X_{\min} := 0$; $X_{\max} := 30$; $q := 0, 0.01 \dots 1$.

математичне очікування середньоквадратичне відхилення від середнього:

$$x_c := \int_{X_{\min}}^{X_{\max}} p(x) \cdot x dx ; \quad disp := \int_{X_{\min}}^{X_{\max}} p(x) \cdot (x - x_c)^2 dx ; \quad \sigma := \sqrt{disp} .$$

Завдання 1. Знайдіть ці характеристики розподілу на графіку нижче.

`mod2 = 1.9` ; `med2 = 2.71828` ; `x_c = 3.253` `σ = 2.135`



На графіку розподілів бачимо, що мода (пунктирна лінія-маркер) проходить через максимум щільності розподілу, а медіана дійсно проходить через точку $P(x) = 0.5$ кумулятивного розподілу.

Розберемось дещо із середніми. Бачимо, що саме середнє з середніх - медіана. При вибраних параметрах розподілу вона чомусь дорівнює відомому числу "e". Візьміть аналітичний вираз для логнормального розподілу і спробуйте довести, що це не випадковий збіг.

Найбільше середнє - математичне очікування. Найменше середнє - мода. Чи для усіх несиметричних розподілів так? Згадаємо інтерпретацію логнормального розподілу - це розподіл імовірності часу виконання певної роботи (робота не може закінчитись раніше ніж почалась, але може тягнутись до нескінченності - так трактується діапазон...).

Запишемо альтернативні формули для визначення імовірності того, що випадкова величина не вийде із заданого діапазону:

$$P_{\text{дов}} := \int_{x_c - \sigma}^{x_c + \sigma} p(x) dx; \quad P_{\text{дов}} := P(x_c + \sigma) - P(x_c - \sigma);$$

$$P_{\text{дов}} = 0.804; \quad P_{\text{дов}} = 0.804.$$

Завдання 1. Уявіть собі, що x - час обслуговування на АЗС. Запишіть формули і знайдіть значення для таких величин:

- 1) середній час обслуговування клієнта;
- 2) імовірність, що час обслуговування буде в межах $(x_c - \sigma) < x < (x_c + \sigma)$;
- 3) імовірність, що час обслуговування буде більший: $x > x_c + \sigma$;
- 4*) імовірність, що час обслуговування буде більший медіанного значення;
- 5*) найбільш імовірне значення часу обслуговування;
- 6) діапазон часу обслуговування, в який попадає 95% подій (обслуговувань);
- 7) мінімальний інтервал між прибуттям клієнтів, при якому з імовірністю 90% не виникне черги.

Завдання 2. Дайте інтерпретації (в термінах часу обслуговування та імовірностей) таких виразів:

$$\text{що_це} := P(xc + \sigma) - P(xc - \sigma); \quad \text{що_це} = 0.804;$$

$$A_це_шо := P(xc + 3 \cdot \sigma) - P(0); \quad A_це_шо = 0.983;$$

$$\text{Гарант_T1} := P(\text{mod2}) - P(0); \quad \text{Гарант_T1} = 0.274;$$

$$\text{Гарант_T2} := P(\text{med2}) - P(0); \quad \text{Гарант_T2} = 0.5;$$

$$\text{Гарант_T3} := P(xc) - P(0); \quad \text{Гарант_T3} = 0.618;$$

$$зп := \frac{\sqrt{5-1}}{\sqrt{5}+1} \quad \frac{зп}{1} = \frac{1-зп}{зп} = 1; \quad \frac{\sqrt{5-1}}{\sqrt{5}+1} = 0.618; \quad \frac{1-0.618}{0.618} = 0.618.$$

Чомусь змінна Гарант_T3 при вибраних параметрах розподілу дорівнює "золотій пропорції". Спробуйте довести, що це не випадковий збіг, будьте обережні, не вступіть в гармонію, краще вже розв'язувати проблеми за допомогою гороскопів...

Центральна гіпотеза: якщо добре пошукати, у будь якому явищі можна знайти числа "π", "е", "золоту пропорцію" та постійну Планка.

Тест. Що саме обчислюється в поданих нижче рівняннях? Чому отримані різні значення і як вони інтерпретуються у випадку, коли - P(x) - розподіл часу виконання роботи? Що робить функція root(..)?

$$\text{root}[(P(x) - 0.95), x, 0, 10] = 7.293$$

$$\text{root}[(P(x) - 0.90), x, 0, 10] = 5.865$$

Висновок. Якщо ви не можете відповісти на питання цього завдання - уважно прочитайте розділ від початку, придивіться до останнього графіка.

1.3.6. Згортка розподілів імовірностей

На практиці часто виникають такі задачі:

1. Потік клієнтів АЗС (банку, крамниці, митниці, готелю..., тобто системи масового обслуговування) має якийсь вхідний розподіл Рвх(х) інтервалів прибуття. Час обслуговування має розподіл Роб(х). Треба знайти Рвих(х) - розподіл інтервалів між виходом клієнтів із системи.

2. Вхідний потік гравців - клієнтів казино має певний вхідний розподіл капіталів Рвх(х). Діяльність гравців у казино можна описати розподілом імовірності приросту капіталу (виграшів, програшів) Руу(х) (уу - удача + уміння). Треба знайти вихідний розподіл капіталів клієнтів - Рвих(х).

3. Деяка деталь проходить ряд технологічних операцій. На кожному етапі обробки добавляється нова випадкова, незалежна від попередніх помилка розміру. Задано розподіл помилок обробки для усіх етапів, треба визначити розподіл помилок для готової деталі.

Цей список легко продовжити. Спробуйте - придумайте свій приклад.

Далі будемо розглядати, як приклад, задачу 2 - очевидно (обґрунтуйте?) більш важливу для народного господарства (це модель бізнесу).

Поставлену задачу можна узагальнити: "гравці" (елементи системи) проходять через "казино" (перетворювач) неодноразово. Треба визначити, як буде еволюціонувати вихідний розподіл, чи існує усталена форма розподілу?

Загальним для усіх розглянутих прикладів є те, що ми шукаємо розподіл імовіро-ностей для суми незалежних випадкових величин з певними розподілами. Як відомо з теорії імовіроностей, розподіл суми незалежних випадкових величин визначається відомим інтегралом згортки (інтегралом Дюамеля). Трошки далі буде подано логіку виве-дення цього інтеграла (за строгими доведеннями звертайтеся до підручників з матема-тики).

$$P_{vyx}(x) = \int_{X_{min}}^x p_{vx}(\xi) \cdot pob(x - \xi) d\xi$$

Використаємо можливості електронної книги - будемо вивчати властивості згорт-ки "експериментально". Спочатку створимо мікробібліотеку розподілів і модуль для ви-конання послідовних операцій згортки. Виберемо такі відомі розподіли:

Коші: $dcauchy(x, \mu, \sigma)$; $dchisq(x, \sigma)$; експоненційний: $dexp(x, \sigma)$; логнормальний: $dlnorm(x, \mu, \sigma)$; рівномірний: $dunif(x, a, b)$.

В цих розподілах x - значення випадкової величини, μ - математичне очікування, σ - стандартне відхилення, a, b - границі рівномірного розподілу. Задамо значення пара-метрів.

$NN := 20$; $x := 0, 0.1 \dots NN$; $\mu := 0$; $\sigma := 1$; $aa := 1$; $bb := 3$; $X_{max} := 50$

Додамо також тестовий - "трикутний" розподіл:

$p\Delta(x, a, b) :=$	$roz \leftarrow$	$\begin{cases} 0 & \text{if } (x < a) \vee (x > b) \\ \frac{2 \cdot (x - a)}{(b - a)^2} & \text{otherwise} \end{cases}$	.
	roz		

Запишемо робочу модель згортки. Ви можете підставляти (копіювати) у формулу згортки будь-які розподіли з мікробібліотеки або з меню функцій пакета.

Вхідний розподіл

$$p_{vx}(x) := dunif(x, aa, bb)$$

розподіл перетворювача

$$p_{ob}(x) := dlnorm(x, \mu, \sigma)$$

Згортка - обчислення вихідного розподілу:

$$p_{vyx}(x) := \int_{X_{min}}^x p_{vx}(\xi) \cdot p_{ob}(x - \xi) d\xi .$$

Далі подано сервісний модуль. В ньому формуються неперервні і дискретні функ-ції, що потрібні нам для наочного подання операції згортки. Можна пропустити цей модуль, якщо Вас не цікавлять питання програмування в середовищі пакета. Подивіться на синтаксис, семантику і спробуйте визначити прагматику.

Вводимо індексовану змінну: ta та визначаємо вектор значень функції:

$X_{max} := 25$; $dd := 5$; $k := 1 \dots dd \cdot NN$; $xx_k := k \div dd$; $pfl_k := p_{vyx}(xx_k)$

Записуємо вектор (масив) у файл:

читаємо з файлу:

WRITEPRN("13rozp1") := pf1; fp1 := READPRN("13rozp1.prn").

Визначаємо функцію користувача, через вбудовану функцію лінійної інтерполяції та виконуємо контроль точності. Застосовуємо ці операції для результату згортки.

$p_{v1}(x) := \text{interp}(xx, fp1, x)$. Контроль: $\int_{Xmin}^{Xmax} p_{v1}(x) dx = 1$.

$$p_{vv}(x) := \int_{Xmin}^x p_{vx}(\xi) \cdot p_{vx}(x - \xi) d\xi ;$$

pf2k := pvv(xxk); WRITEPRN("13rozp2") := pf2; fp2 := READPRN("13rozp2.prn");

$p_{v2}(x) := \text{interp}(xx, fp2, x)$; q := 1..3 .

sum(x) := pob(x-1)·pΔ(1,0,3) + pob(x-2)·pΔ(2,0,3) + pob(x-3)·pΔ(3,0,3) .

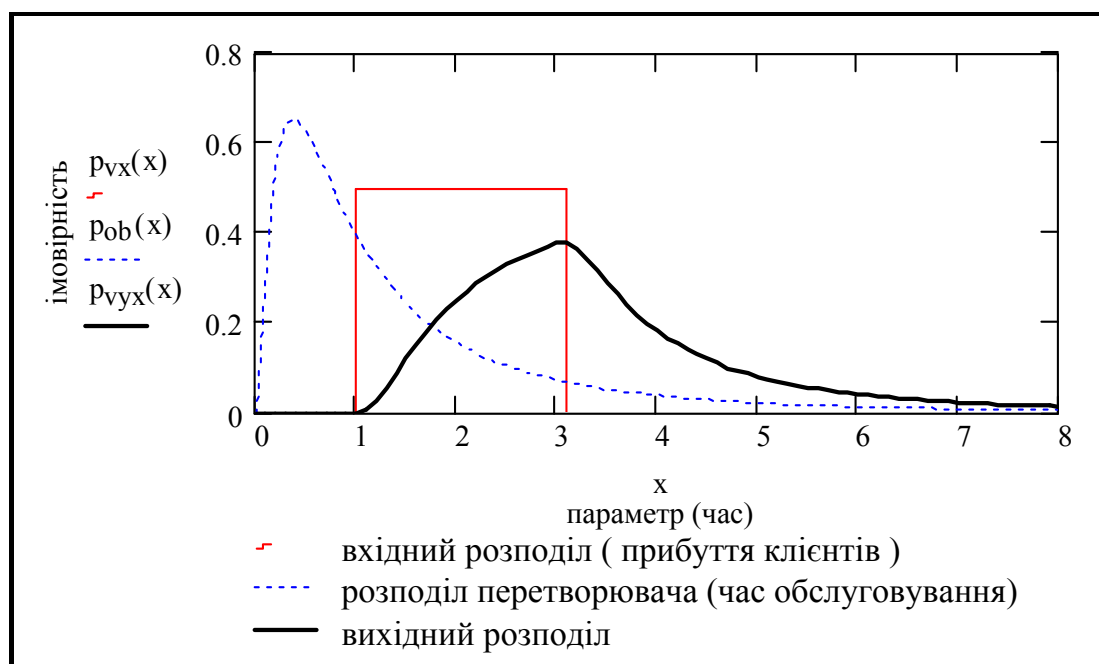
Що спільного у інтеграла згортки та цієї суми? - Ми замінили інтеграл сумою, і дуже приблизно - всього із трьох складових.

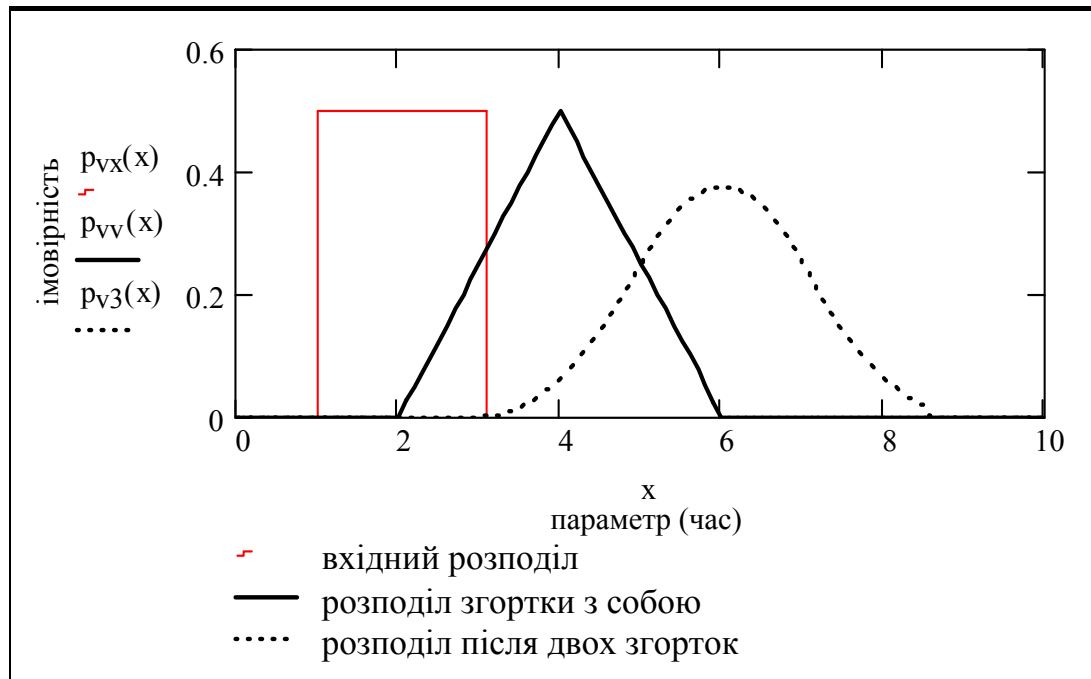
$$p_{v3}(x) := \int_{Xmin}^x p_{vx}(\xi) \cdot p_{v2}(x - \xi) d\xi$$

Розглянемо приклади згорток. **Перевірте свою інтуїцію та ерудицію** - вгадайте та нарисуйте, якими будуть результати (вигляд, а не конкретні значення) згортки розподілів:

- 1) рівномірного та логнормального;
- 2*) логнормального та рівномірного;
- 3) двох однакових рівномірних;
- 4) згортка двох нормальних розподілів з різними середніми;
- 5*) багаторазова згортка рівномірного розподілу з собою;
- 6**) згортка розподілів Коші. (* - складність питання, задачі)

Далі подано графіки - відповіді на деякі із запитань.





Чи здогадалися Ви, що згортка рівномірного розподілу з собою буде "трикутним" розподілом, а подвійна згортка - ну майже гаусовий (нормальний) розподіл?

Завдання 1. Якщо a, b - границі рівномірного розподілу, визначить границі згортки його з собою.

Завдання 2. Подивіться на графіки угорі - з кожною згорткою розподіл зсувається вправо. При яких значеннях параметрів a, b згортки не будуть зсуватись?

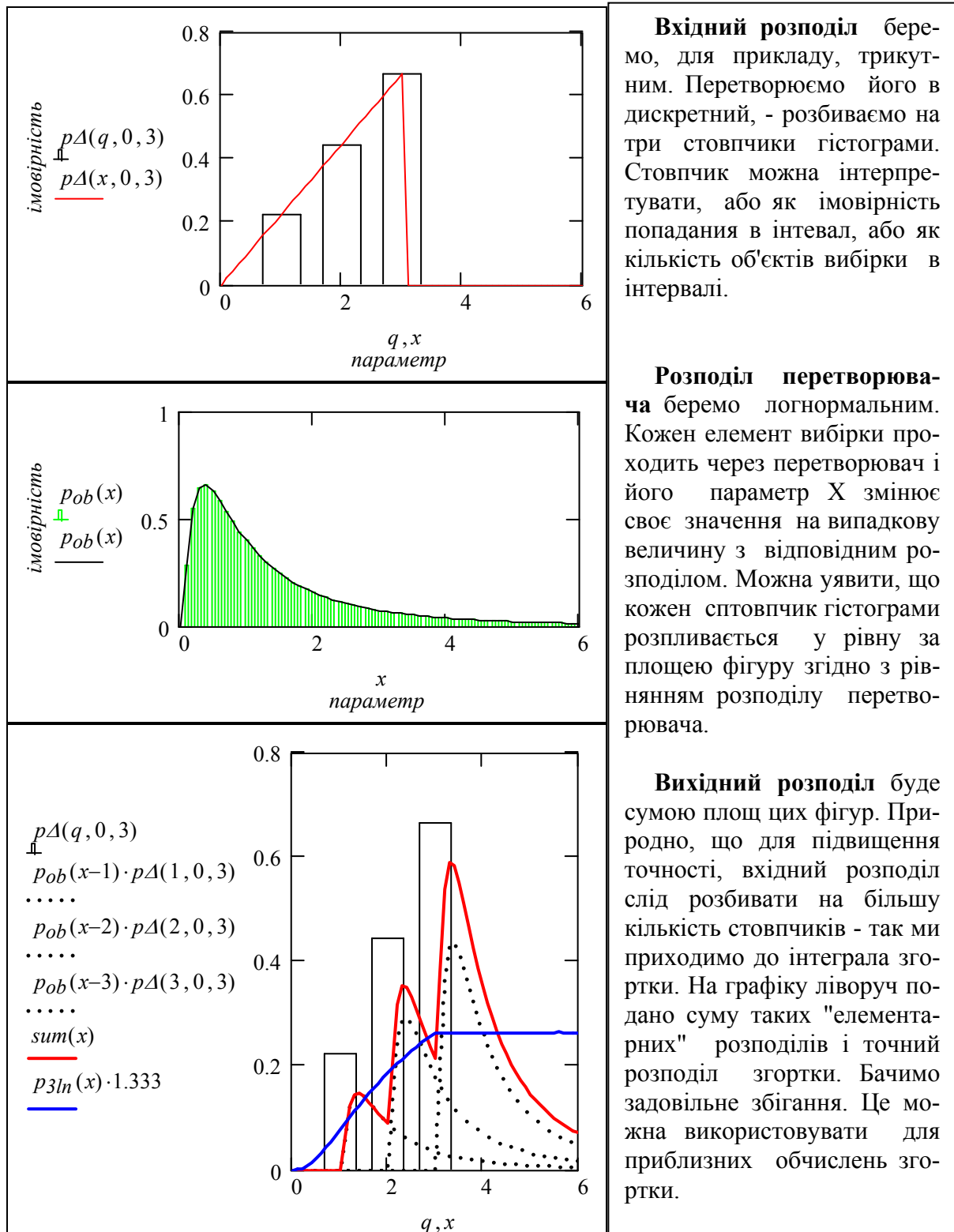
Завдання 3*. Наведіть приклад розподілу, що при згортці з собою не змінюється (змінюються тільки параметри функції розподілу).

А тепер розглянемо згортку з позицій не математика, а інженера, менеджера, чисто по-діловому: що дано, що робимо, що отримуємо.

Далі подана сторінка, що є коміксом на тему "як виконується згортка". Між іншим, в США, Японії, Франції випускаються комікси на технічні теми, наприклад, "сучасні мережі передачі даних", "робототехніка". Це книжки в картинках для освоєння нових науково-технічних досягнень.

Завдання 4. Модифікуйте наведену далі систему графіків для іншої пари функцій розподілів.

Згортка частотних розподілів. Копія екрану.



Вхідний розподіл беремо, для прикладу, трикутним. Перетворюємо його в дискретний, - розбиваємо на три стовпчики гістограми. Стовпчик можна інтерпретувати, або як імовірність попадання в інтервал, або як кількість об'єктів вибірки в інтервалі.

Розподіл перетворювача беремо логнормальним. Кожен елемент вибірки проходить через перетворювач і його параметр X змінює своє значення на випадкову величину з відповідним розподілом. Можна уявити, що кожен стовпчик гістограми розпливається у рівну за площею фігуру згідно з рівнянням розподілу перетворювача.

Вихідний розподіл буде сумою площ цих фігур. Природно, що для підвищення точності, вхідний розподіл слід розбивати на більшу кількість стовпчиків - так ми приходимо до інтеграла згортки. На графіку ліворуч подано суму таких "елементарних" розподілів і точний розподіл згортки. Бачимо задовільне збігання. Це можна використовувати для приблизних обчислень згортки.

Підведемо підсумки розгляду згортки розподілів. Сформулюємо властивості згортки розподілів імовірностей, без доведення - зробіть це самостійно).

1. Запам'ятання формули згортки полегшує її симетрія - можна поміняти місцями вхідний розподіл і розподіл перетворювача. Спробуйте довести цю властивість.

$$\int_{X_{\min}}^{x1} p_{vx}(x1 - \xi) \cdot p_{ob}(\xi) d\xi = \int_{X_{\min}}^{x1} p_{vx}(\xi) \cdot p_{ob}(x1 - \xi) d\xi = 1$$

Уважно подивіться на вираз - а) що там переставлено, б) який зміст має перший знак рівності: це логічне ствердження, яке комп'ютер перевірів, знайшов істинним і дописав "= 1" (згадайте булеву алгебру логіки). Спробуйте це довести самостійно.

2. Розподіл суми випадкових величин з **кінцевими дисперсіями** сходиться до нормального (або гаусового). Доведення шукайте в монографіях з теорії розподілів.

Зауваження. Ясно, що розподіли з обмеженим діапазоном мають обмежену дисперсію. Нехай інтервал є необмеженим $-\infty < x < \infty$. Подивіться на вираз для дисперсії:

$$\text{disp} := \int_{X_{\min}}^{X_{\max}} p(x) \cdot (x - x_c)^2 dx$$

і спробуйте визначити, яку умову повинен задовольняти розподіл $p(x)$, щоб дисперсія була нескінченною. Нагадаємо, що розподіли з нескінченною дисперсією називають негаусовими, або гіперболічними. Пам'ятайте, що економіка, наука, культура мають негаусову статистику. Але це окрема велика тема, що виходить за межі посібника.

3. Нормальний розподіл є стійким - згортка нормального розподілу з собою дає знову ж нормальний розподіл (тільки з іншими параметрами).

1.3.7. Узагальнення. Використання згортки для аналізу динамічних систем

Ми поки розглядаємо інструменти для дослідження операцій. В цьому розділі ми покажемо, що згортка є досить універсальним інструментом, що використовується не тільки в теорії імовірностей, а ще й в теорії управління. Подивимось на згортку, як декларувалось спочатку, "згори", і поєднаємо те, що розійшлося, як колись материки, по різних науках.

В теорії автоматичного управління є задача визначення реакції лінійної динамічної системи на довільний вхідний сигнал. Далі ми будемо розглядати імовірнісні і детерміновані динамічні системи, і ми часто будемо використовувати "інструмент" згортки для обчислення процесів в системі як у просторі (зміна розподілу імовірностей) так і у часі (зміна в часі певної вихідної величини). Розглянемо задачу на прикладі системи, що описується диференціальним 2-го порядку.

$$\frac{d^2}{dt^2}x + a1 \cdot \frac{d}{dt}x + a0 \cdot x = Xvx(t).$$

Необхідно знайти реакцію системи $x(t)$ на зовнішню дію $Xvx(t)$. Розглянемо коротко логіку отримання реакції на довільний вхідний сигнал. Для лінійних систем можна визначити **імпульсну перехідну функцію (ІПХ)** - реакцію системи на імпульс малої тривалості і одиничної інтенсивності. Для лінійних систем є справедливим принцип суперпозиції: реакція на суму вхідних сигналів дорівнює сумі реакцій на кожен окрему складову. Розбиваємо вхідний сигнал на послідовність імпульсів, реакцію на сигнал знаходимо як суму реакцій на окремі імпульси.

Цій словесній формулі відповідає саме інтеграл згортки. Записуємо для порівняння вирази для реакції динамічної системи на вхідний сигнал і "реакції" імовірнісного перетворювача на вхід з певним розподілом імовірностей.

$$X_{vyx}(t) = \int_0^t X_{vx}(\tau) \cdot g(t - \tau) d\tau$$

$$p_{vyx}(x) = \int_{X_{min}}^x p_{vx}(\xi) \cdot p_{ob}(x1 - \xi) d\xi$$

Розглянемо приклади згортки. Задаємо параметри динамічної системи:

$$NN := 50 \quad t := 0, 0.1 \dots NN \quad a1 = 1 \quad a0 = 1 \quad ta := 1 \quad tb := 6 \quad a := 1 \quad b := 3$$

Задаємо варіанти вхідних сигналів:

$$X1_{vx}(t, ta, tb) := \Phi(t - ta) - \Phi(t - tb) \quad X2_{vx}(t, a, b) := p\Delta(t, a, b)$$

Записуємо вираз для ПІХ (це розв'язання дифрівняння):

$$g2(t, a1, a0) := 2 \cdot \exp\left(\frac{-1}{2} \cdot a1 \cdot t\right) \cdot \frac{\sin\left(\frac{1}{2} \cdot \sqrt{4 \cdot a0 - a1^2} \cdot t\right)}{\sqrt{4 \cdot a0 - a1^2}}$$

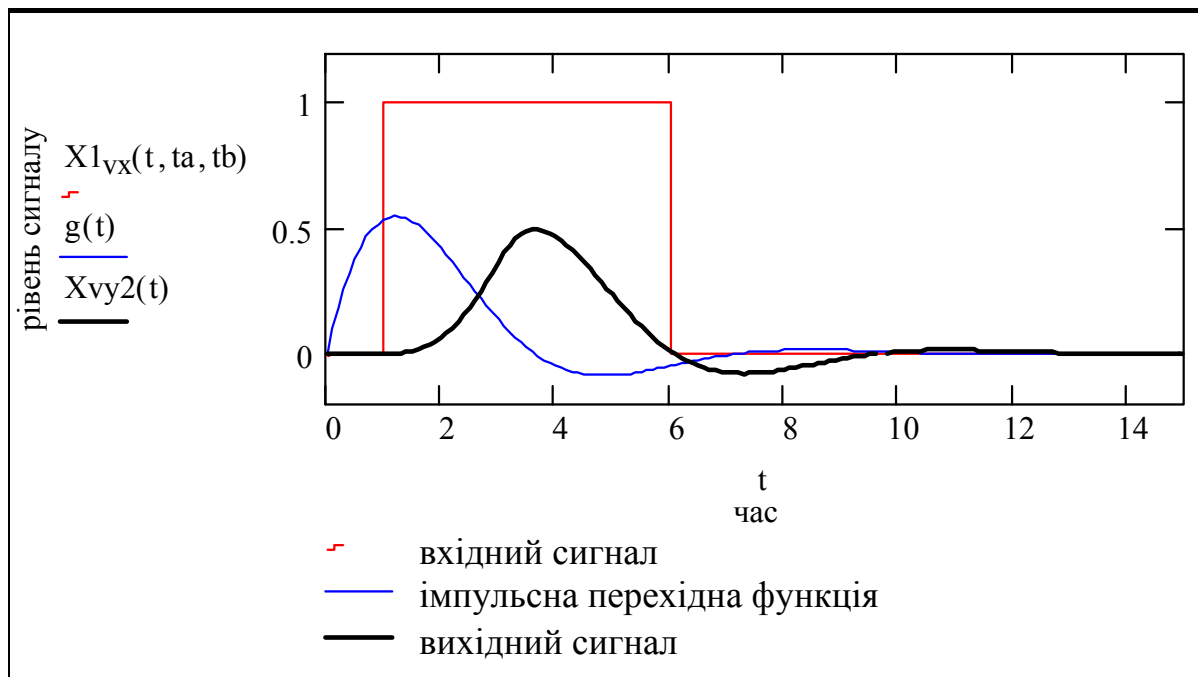
$$g(t) := \begin{cases} 0 & \text{if } t < 0 \\ g2(t, a1, a0) & \text{otherwise} \end{cases}$$

Записуємо вирази для вихідних сигналів і будуємо графіки для вхідного сигналу - імпульсу та реакції динамічної системи на цей вхід.

$$X_{vy1}(t) := \int_0^t X1_{vx}(\tau, ta, tb) \cdot g(t - \tau) d\tau$$

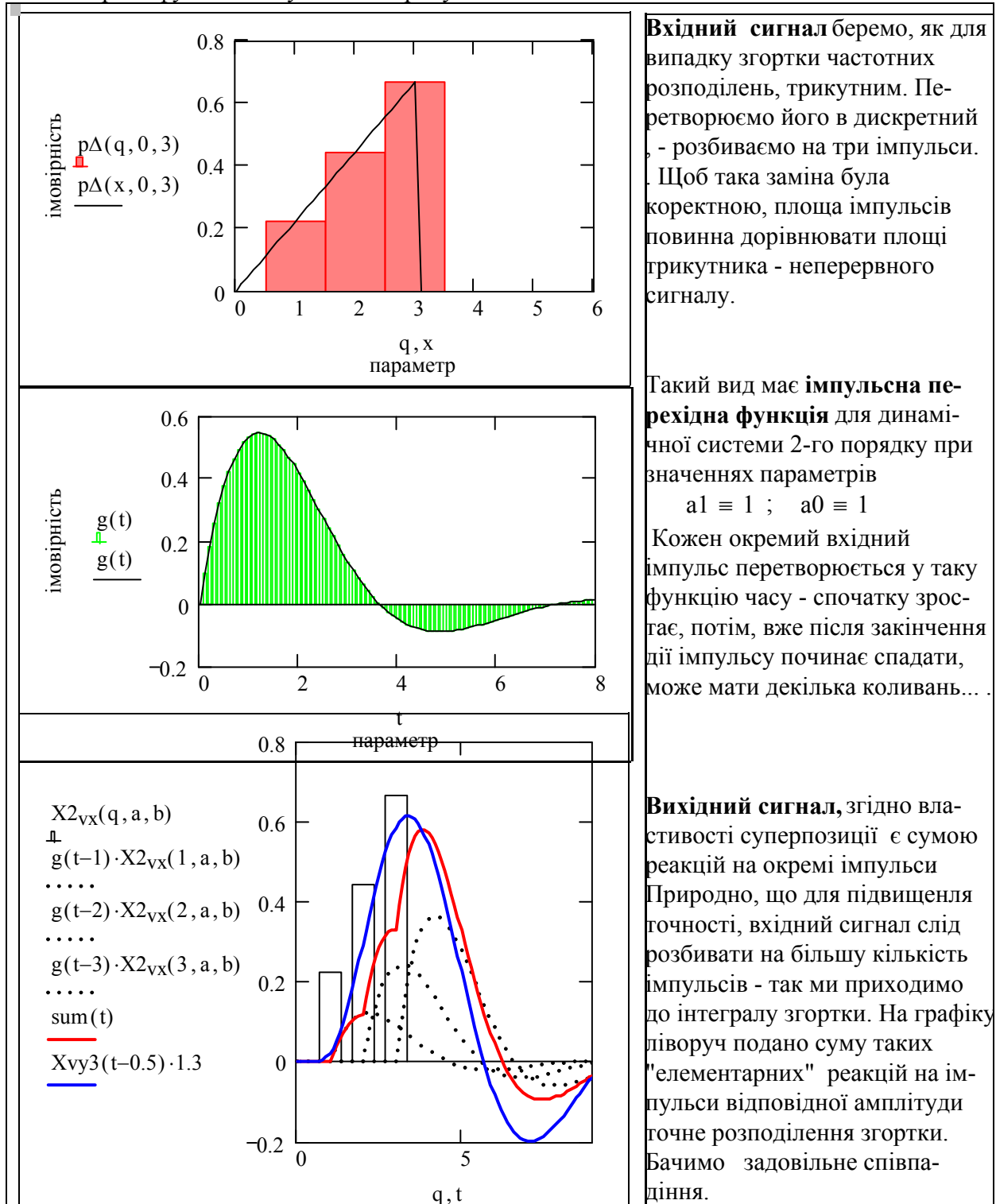
$$X_{vy2}(t) := \int_0^t X2_{vx}(t - \tau, a, b) \cdot g(\tau) d\tau$$

$$q := 1 \dots 3; \text{sum}(t) := g(t - 1) X2_{vx}(1, 0, 3) + g(t - 2) X2_{vx}(2, 0, 3) + g(t - 3) X2_{vx}(3, 0, 3)$$



Розглянемо тепер логіку визначення реакції лінійної динамічної системи на довільний вхідний сигнал – модифікуємо попередню сторінку.

Згортка функцій часу. Копія екрану.



Зауваження. Ми ознайомились із застосуванням згортки а) для обчислення частотного розподілу суми випадкових величин з різними розподілами і б) вихідного сигналу - реакції динамічної системи на певний вхідний сигнал. Ми побачили універсальність математичних моделей, однак, **не переплутайте** - це зовсім різні задачі - розподіл в просторі і процес зміни певної величини в часі.



1.4. Ідентифікація частотних розподілів

Вступ

Ідентифікація емпіричних частотних розподілів - типова, масова операція в наукових дослідженнях, аналізі експериментальних даних. Сьогодні підсистеми, точніше - модулі ідентифікації частотних розподілів вбудовуються у різноманітні автоматизовані системи - диспетчерські, аналітичні помічники біржового маклера чи менеджера постачання.

Під ідентифікацією ми маємо на увазі знаходження деякого "теоретичного" розподілу, що найбільш відповідає експериментальним даним. Відомі теоретичні розподіли - це результат дії певних породжуючих механізмів та фундаментальних законів. Стратегічна ціль ідентифікації - за неповними, зашумленими даними виявити закони і механізми функціонування і на цій основі прогнозувати поведінку системи, управляти системою.

Приклад такої ідентифікації подано вище - спеціалісти компанії FrameStore разом з палеонтологами по скам'янілим відбиткам ідентифікували не тільки зовнішній вигляд, але й ходу, динаміку польоту семиметрових літаючих ящерів. Це була довга, важка, витратна робота.

Матеріал цього розділу - ресурс (програмні модулі, шаблони документів і графіків) для відповідних розділів курсових і дипломних проектів. Тематика проектів може бути найрізноманітнішою - автоматизація задач управління в системах масового обслуговування, управління запасами, управління аграрними підприємствами та ін.

Задачі ідентифікації будемо розглядати на загальнозрозумілому прикладі автозаправної станції (АЗС), де теж корисно мати дані про **частотні розподіли інтервалів між прибуттям клієнтів і частотні розподіли часу обслуговування клієнтів**.

1.4.1. Постановка задачі

Для деякого набору статистичних даних $\{X_j\}$, $j = 1..M$ побудовано гістограму (емпіричний частотний розподіл $Rem[i]$, $i = 1..N$ - число (груп) інтервалів. Розглядається ГІПОТЕЗА про те, що гістограма є реалізацією деякого теоретичного розподілу "засміченого" випадковими незалежними помилками з нормальним розподілом.

ПОТРІБНО ПІДТВЕРДИТИ АБО ВІДКИНУТИ ЦЮ ГІПОТЕЗУ

1.4.2. Порядок розв'язання задачі з використанням критерію

Пірсона

1) Визначаємо параметри теоретичного розподілу, що найкращим чином наближає гістограму.

2) Обчислюємо значення критерію Пірсона.

3) Визначаємо критичне значення критерію для заданої ймовірності істинності гіпотези для даного числа ступенів свободи розподілу χ -квадрат.

4) Порівнюємо значення критерію з критичним. Якщо воно більше критичного, гіпотеза відкидається (ПРИ ДАНОМУ РІВНІ ЙМОВІРНОСТІ!)

Зауваження. Перш, ніж застосовувати цей критерій, слід провести змістовний аналіз джерел і процедур отримання даних, виявити і виключити невинні помилки вимірювань, отримати дані з альтернативних джерел, і пам'ятати, що висновок на підставі критерію Пірсона, як і іншого критерію не є остаточним вироком.

Завжди слід цікавитись механізмами, які саме породжують те, що ми спостерігаємо. Якщо досить точно відомо, що у системи повинен бути певний теоретичний розподіл, а згідно з критерієм Пірсона більш імовірним є якийсь інший, слід критично проаналізувати вхідні дані та процеси їх отримання і одночасно - коректність гіпотези про очікуваний теоретичний розподіл. Виконаємо задачу згідно з цими пунктами.

1.4.3. Побудова гістограми для вхідного потоку (клієнтів)

Побудова гістограми, або емпіричного частотного розподілу є агрегування даних. Це стандартна процедура, вбудована в усі основні пакети програм. ВВОДИМО згідно з завданням **функцію - генератор даних** (власне введення далі)

$X = \text{rpois}(100, 5)$; $N := \text{length}(X)$; $N = 100$; $\text{ORIGIN} := 1$.

Проводимо попередній аналіз даних - визначаємо:

мінімум $\min(X) = 0$, середнє $\text{mean}(X) = 4.7$, максимум $\max(X) = 12$.

1) Кількість інтервалів повинна бути такою, щоб в середньому в інтервал попадало не менше 7 елементів, тобто: $\text{Кільк_інт} := 8$; $q := 1..8$;

В інтервал попадає: $N \div 8 = 12.5$ тобто більше 7.

2) Границі системи інтервалів беремо так, щоб до них входили мінімальне і максимальне значення. **Увага!** для дискретних розподілів інтервали повинні бути обов'язково цілочисленими!

Записуємо вираз для вектора інтервалів:

$\text{Крок} := \text{ceil}\left[\frac{(\max(X) - \min(X)) \cdot 1.2}{\text{Кільк_інт}}\right]$; $\text{Int}_q := \text{ceil}[\text{Крок} \cdot (q - 1) + \min(X) \cdot 0.98]$.

Множники 1.1 та 0.98 - параметри для настроювання вектора інтервалів згідно з пунктом 2). Метою інженерної праці сьогодні є не виконання якогось розрахунку, а створення програми (документа) для розв'язання певного класу задач. В нашому випадку це задачі визначення найкращого наближення гістограми деяким теоретичним розподілом.



Увага! - те, що подано далі - може бути **основою** розділу курсової роботи чи проекту. Маємо вектор емпіричних даних і вектор інтервалів - **обчислюємо гістограму**: (в розділі 1.3 описана нова функція - $\text{histogram}(n, X)$).

$\text{Rem} := \text{hist}(\text{Int}, X)$

Число "стовпчиків" гістограми на 1 менше числа границь інтервалів. Для зручності подальших обчислень додаємо одне, пусте значення до гістограми (для функції $\text{histogram}(n, X)$ цієї проблеми не існує) : $\text{Rem}_{\text{Кільк_інт}} := 0$.

Нормуємо гістограму - ділимо усі значення на об'єм вибірки N . Для контролю обчислюємо суму значень "стовпців" гістограми - вона повинна дорівнювати обсягу вибірки.

$\text{Rem} := \frac{\text{Rem}}{N}$; $N = 100$; $\sum_q \text{Rem}_q = 100$; $\sum_q \text{Rem}_q = 1$; $t := 1..12$.

Обчислюємо **моду, медіану, математичне очікування** двома альтернативними способами - а) за вибіркою і, для контролю, б) за гістограмою [див. підрозділ 1.3].

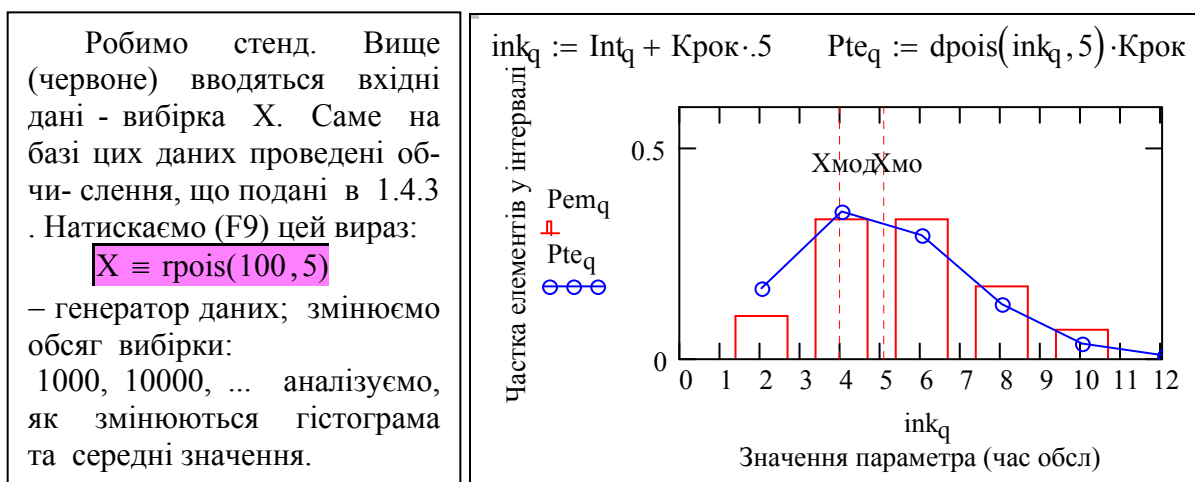
Мода (H_i, I_n) := $\text{csort}(\text{augment}(I_n, -H_i), 2)_{1, 1} + 0.5 \cdot \text{Крок}$

Медіана(Hi, In) :=	$N \leftarrow \text{length}(H_i)$ $\text{sum}_1 \leftarrow H_{i1}$ for $i \in 2..N$ $\text{sum}_i \leftarrow \text{sum}_{i-1} + H_{ii}$ $\text{nom} \leftarrow 1$ while $\text{sum}_{\text{nom}} \leq 0.5 \cdot \text{sum}_N$ $\text{nom} \leftarrow \text{nom} + 1$ $\text{znMed} \leftarrow \text{In}_{\text{nom}} + 0.0 \cdot \text{Крок}$ znMed	$\text{Серед_Ар} := \sum_q \text{Pem}_q \cdot (\text{Int}_q + .5 \cdot \text{Крок})$ Це вбудовані функції визначення середніх по вибірці X $X_{\text{мод}} := \text{mode}(X)$ $X_{\text{мед}} := \text{median}(X)$ $X_{\text{мо}} := \text{mean}(X)$
--------------------	--	--

Виводимо ці дані для контролю і осмислення на тему: що ми втрачаємо, коли агрегуємо дані, що отримуємо? Чому не збігаються відповідні середні, обчислені різними способами?

Мода(Rem, Int) = 6 ; Медіана(Rem, Int) = 5 ; Серед_Ар = 5.56 ;
 $X_{\text{мод}} = 4$; $X_{\text{мед}} = 5$; $X_{\text{мо}} = 5.1$.

Коректуємо систему інтервалів (чому). Будуємо графіки гістограми і "схованого" в емпіричних даних теоретичного розподілу - "гіпотези"



Виконуємо контроль нормованих даних:

$$\sum_q \text{Pem}_q = 1 ; \quad \sum_q \text{dpois}(\text{ink}_q, 5) \cdot \text{Крок} = 1$$

Формуємо і виводимо таблицю даних D2, що буде використана для апроксимації гістограми теоретичним розподілом.

Зауваження. Пам'ятаємо, що вектор інтервалів дає границі інтервалів. Відповідне значення гістограми слід відносити до середини інтервалів.

Пакет має дві вбудовані функції побудови гістограм:

`histogram(вектор_інтервалів, вектор_даних);`

`hist(кількість_інтервалів, вектор_даних).`

Подаємо розшифровку означень:

Int - інтервал-границя, **ink** - інтервал-середина

Pem - розподіл емпіричний, **Pte** - розподіл теоретичний.

Вище розроблено документ побудови гістограми за вхідними неагрегованими даними. Недолік традиційних документів Mathcad - вони немонолітні, тому:

Збираємо тепер те, що було розроблено вище у цей блок обчислення гістограми і формування таблиці даних $\text{Крок} := \text{ceiling}\left[\frac{(\max(Xr) - \min(Xr)) \cdot 1.2}{\text{Кільк_інт}}\right]$. Ми вбудували в цей текстовий блок формули. Тепер цей блок легко копіювати і переносити.

$\text{Int}_q := \text{ceiling}[\text{Крок} \cdot (q - 1) + \min(Xr) \cdot 0.98]$; $\text{Rem} := \text{hist}(\text{Int}, Xr)$; $\text{Rem}_{\text{Кільк_інт}} := 0$.

$\text{Рем} := \frac{\text{Rem}}{N}$ $\text{D2} := \text{augment}(\text{Int} + 0.5 \cdot \text{Крок}, \text{Рем})$.

Виводимо таблицю агрегованих даних - значення гістограми та інтервалів, на яких вона побудована.

$$D2^T =$$

	1	2	3	4	5	6	7	8
1	2	4	6	8	10	12	14	16
2	0.11	0.33	0.27	0.19	0.05	0.05	0	0

1.4.4. Оптимальна апроксимація гістограми теоретичним розподілом

Після виконаної підготовчої роботи переходимо власне до визначення параметрів гіпотетичного теоретичного розподілу, що НАЙКРАЩИМ чином наближає гістограму. В даному випадку вибираємо гіпотезу - розподіл Пуасона (так він же закладений в генераторі даних). Беремо нульове початкове значення індексів масивів: $\text{ORIGIN} := 0$, так зручніше. Вибраний розподіл має тільки один параметр. Виділяємо рамкою і кольором (в електронній версії книги) зону введення **теоретичного розподілу - гіпотези**

$\text{rt}(t, \text{ля}) := \text{dpois}(t, \text{ля})$

Виводимо (тут можна ввести свої дані замість тестових) емпіричні дані і виконуємо контроль їх коректності:

$$D2^T =$$

	0	1	2	3	4	5	6	7
0	2	4	6	8	10	12	14	16
1	0.11	0.33	0.27	0.19	0.05	0.05	0	0

$$; \sum_{q=0}^7 D2_{q,1} = 1$$

Розв'язуємо задачу апроксимації даних заданою функцією. Знаходимо *оптимальні за якимось критерієм відхилення від емпіричного розподілу параметри функції (в якості якогось критерію беремо, зрозуміло чому, квадратичний)*.

Використовуємо блок розв'язання математичного пакета у варіанті $\text{minerr}(\cdot)$, що перекладається так - "знайти значення змінних, що дають мінімальні відхилення від заданих умов". Загальновідомо, що вбудовані методи оптимізації і математичного програмування (=розв'язання нелінійних систем) є досить неефективними і незручними. Найбільш безвідмовним і універсальним є $\text{minerr}(\cdot)$.

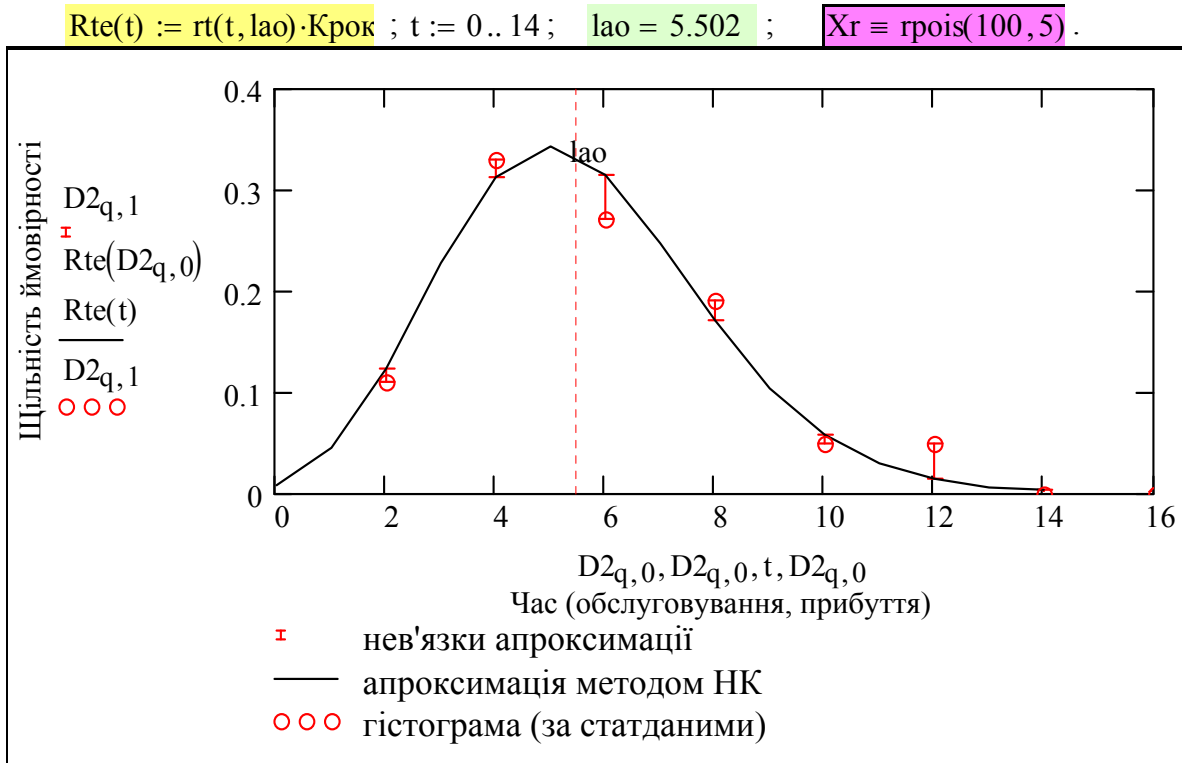
1) Задаємо кількість точок апроксимації і початкові значення параметрів, що оптимізуються: $n := \text{rows}(D2) - 1$; $n = 7$; $q := 0..n$; $\text{ля} := 4$.

2) Після Given записуємо умови - рівняння, нерівності:

$$\text{Given} \quad \sum_{q=0}^n \left(\frac{D2_{q,1}}{\text{Крок}} - \text{rt}(D2_{q,0}, la) \right)^2 = 0; \quad la > 0.$$

Після рівнянь записуємо функцію Minerr() : $la_0 := \text{MinErr}(la)$

Будуємо графік - перевіряємо, чи справді функція наближує емпіричні дані. Виводимо результати - оптимальне значення параметра la_0 . Покладайте по **Xr** і зробіть якісну оцінку розкиду реалізацій гістограми (графік) і значень параметра la_0 .



Ми прийняли гіпотезу, що теоретичний розподіл - **пуасонівський**. Але можуть бути інші, конкурентні гіпотези. Дуже бажано висунути і перевірити ще хоч одну альтернативну гіпотезу. Виконаємо ідентифікацію **для альтернативної гіпотези: "теоретичний розподіл - гаусівський"** Увага! В курсовій роботі слід задачу ідентифікації виконати не менше, ніж для двох гіпотез.

Копіюємо попередній документ, видалимо з нього коментарі. **Вводимо інший теоретичний розподіл** - оце уся різниця. Неважко зробити компактний модуль ідентифікації, входами якого будуть: масив даних та гіпотетичний теоретичний розподіл.

$$\text{rt2}(t, \mu, \text{sgm}) := \text{dnorm}(t, \mu, \text{sgm})$$

1) Задаємо кількість точок апроксимації і початкові значення оптимізуємих параметрів (це зможе зробити самостійно програма ідентифікації):

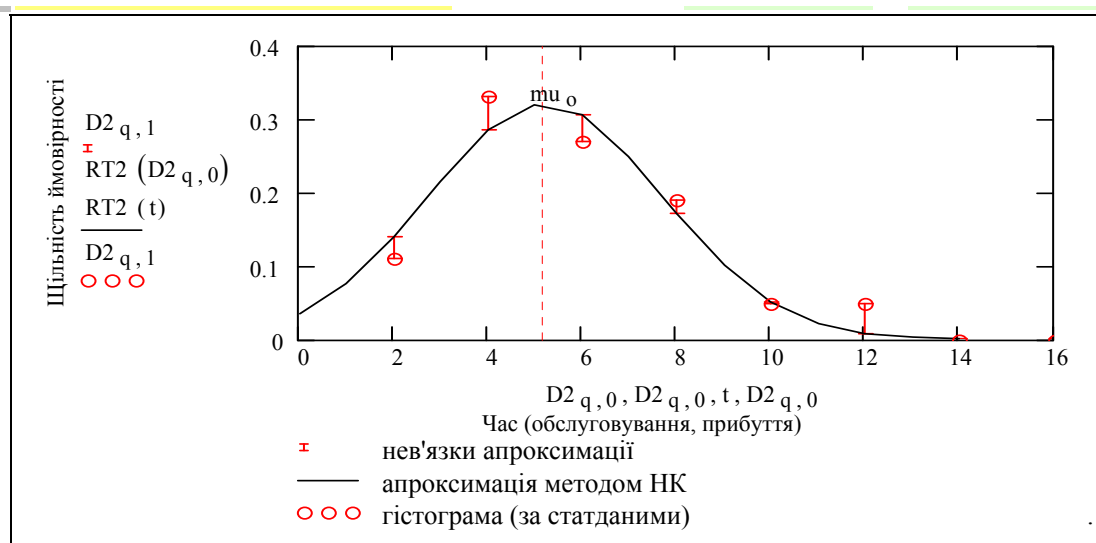
$$n := \text{rows}(D2) - 1 \quad n = 7 \quad q := 0..n \quad \mu := 5 \quad \text{sgm} := 1001$$

2) Записуємо систему рівнянь:

$$\text{Given} \quad \sum_{q=0}^n \left(\frac{D2_{q,1}}{\text{Крок}} - \text{rt2}(D2_{q,0}, \mu, \text{sgm}) \right)^2 = 0 ; \quad \begin{pmatrix} \mu_0 \\ \text{sgm}_0 \end{pmatrix} := \text{MinErr}(\mu, \text{sgm})$$

Будуємо графік - перевіряємо, чи справді функція наближує емпіричні дані. Виводимо результати - оптимальне значення параметрів μ_0 та σ_{m_0} .

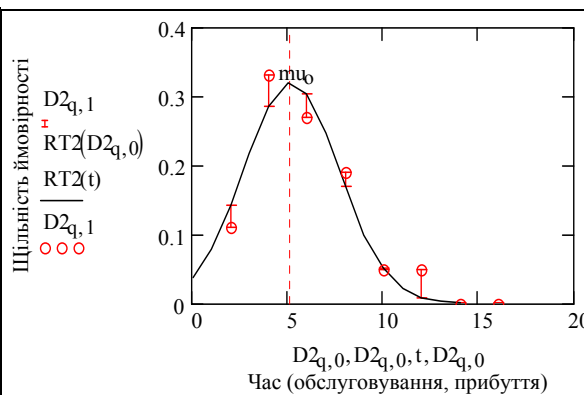
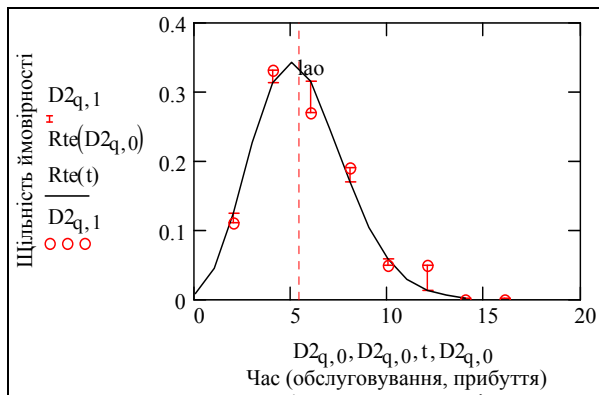
$RT2(t) := rt2(t, \mu_0, \sigma_{m_0}) \cdot \text{Крок}$; $t := 0..14$; $\mu_0 = 5.215$; $\sigma_{m_0} = 2.488$



Виконаємо якісний аналіз отриманих результатів. Порівнюємо графіки апроксимацій (нижче) - і без критеріїв Пірсона і Колмогорова бачимо, що апроксимація нормальним розподілом виглядає не зовсім нормальною, що і очікувалось, оскільки емпіричний розподіл дійсно є пуасонівським.

Гіпотеза: Пуасонівський розподіл

Гіпотеза: Гаусівський розподіл



1.4.5. Перевірка вірогідності гіпотези про теоретичний розподіл за критерієм Пірсона

Проведемо перевірку гіпотез про те, що розподіл поданий емпіричними даними є **а) пуасонівським, чи б) гаусовим** за критерієм Пірсона. Подаємо послідовність кроків з перевірки гіпотези про теоретичний розподіл.

1) Обчислюємо значення критерію за формулою:

$$X_{i1} := \sum_{q=0}^n \frac{\left(\frac{D_{2q,1}}{\text{Крок}} - \text{dpois}(D_{2q,0}, \mu_0) \right)^2}{\text{dpois}(D_{2q,0}, \mu_0)}$$

Для гіпотези: "пуасонівський розподіл" маємо $X_{i1} = 0.059$.

$$Xi2 := \sum_{q=0}^n \frac{\left(\frac{D2_{q,1}}{K_{рок}} - dnorm(D2_{q,0}, m_{\phi}, sgm_{\phi}) \right)^2}{dnorm(D2_{q,0}, m_{\phi}, sgm_{\phi})}$$

Для гіпотези: "гаусівський розподіл" маємо: $Xi2 = 0.125$

Якщо ви будете переобчислювати реалізацію набору випадкових даних, то побачите, що значення критеріїв на різних реалізаціях випадкових даних можуть суттєво змінюватись. Показник для некоректної гіпотези може виявитись кращим, ніж для коректної. Випадкові величини - це випадкові величини, але **в середньому**, значення критерію для першої гіпотези все ж будуть меншими.

2) Визначаємо імовірність такого значення критерію. Для цього спочатку треба отримати розподіл χ -квадрат із заданим числом ступенів вільності.

За таблицею або графіком розподілу визначаємо імовірність такого значення критерію для χ -квадрат розподілу із числом ступенів вільності **dv**.

Число ступенів вільності **dv** визначається як число груп (інтервалів) гістограми **M - 1 - r** (**r** - число параметрів теоретичного розподілу, наприклад, для гаусівського **r = 2**, для пуасонівського **r = 1**).

$M := \text{Кільк_інт}$; $dv1 := M - 1 - 1$; $dv2 := M - 2 - 1$; $xi := .0, .5 .. 24$.

Число ступенів вільності розподілу χ -квадрат для теоретичних розподілів

а) гаусовський $dv2 = 5$; б) пуасонівський $dv1 = 6$.

Використовуємо **вбудовані функції** щільності розподілу χ -квадрат і відповідний кумулятивний розподіл.

$Xi(xi, d) := dchisq(xi, d)$; $IX(xi, d) := pchisq(xi, d)$.

Визначаємо (дехто досі вважає, що це береться з статистичних таблиць) **значення X195, що відповідає ймовірності 95%** - тобто ми встановлюємо істинність гіпотези з 95% ймовірністю.

$xx := 13$; $X195 := \text{root}(dchisq(xx, dv1) - .05, xx)$; $X195 = 9.392$;

$X295 := \text{root}(dchisq(xx, dv2) - .05, xx)$; $X295 = 8.325$.

3) Визначаємо за графіком або таблицею розподілу χ -квадрат **ІМОВІРНІСТЬ** отриманого значення і **ВИРІШУЄМО**, прийняти, чи відкинути гіпотезу...

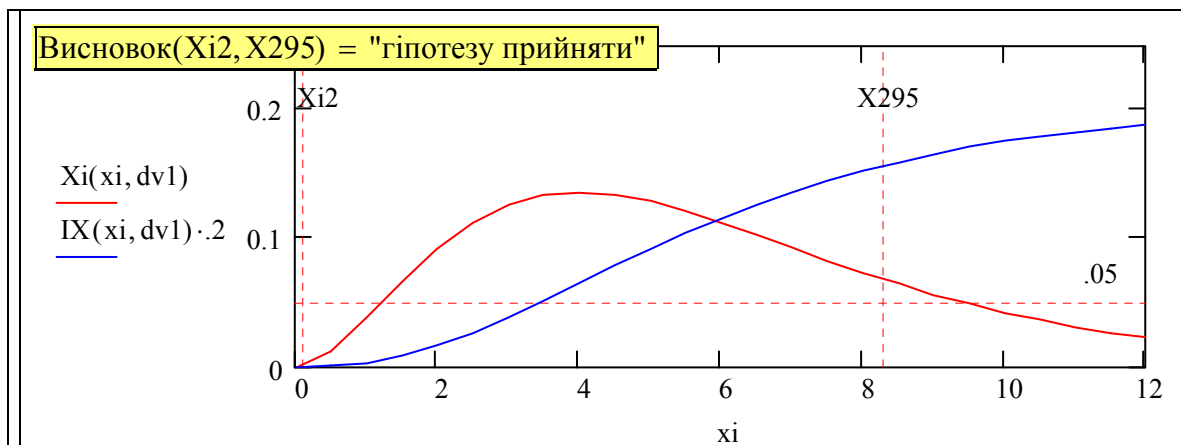
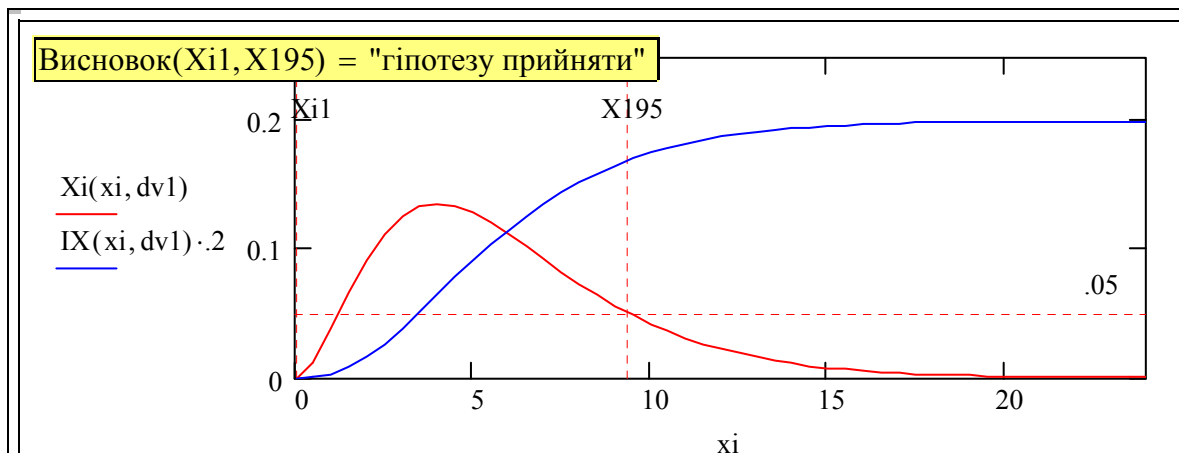
Чим відрізняється електронна книга від паперової? - електронна може бути, як мінімум, - секретарем-референтом, максимум - експертом-порадником. Це дуже просто робиться: напишемо, наприклад, такий модуль для аналізу результатів ідентифікації і, певним чином, думати вже непотрібно.

Висновок(маємо, норма) :=	v1 ← "гіпотезу прийняти"	
	v2 ← "гіпотезу відкинути"	
	виснов ←	v1 if маємо < норма
		v2 otherwise

Тестуємо цю програму:

Висновок(12,2) = "гіпотезу відкинути"; Висновок(2,12) = "гіпотезу прийняти" .

Будуємо графіки з розподілом χ -квадрат і резолюцією програми для першої гіпотези (істинної) і для другої.



Прокоментуємо те, що отримали (обидві гіпотези приймаються) - отримали те, що чекали: пуасонівський розподіл "збігається" (сходиться) до гаусівського і починаючи з певних значень параметрів, дуже подібний гаусівському. Імітовані статистичні дані є відносно дуже чистими - тому отримали такі малі значення критеріїв. Природно, що обидва розподіли є статистично припустимими гіпотезами. Яку гіпотезу все ж прийняти: - ту, що має менше значення критерію (віддамо визначення цього програми):

НашВибір(крит1, крит2) :=	v1 ← гіпотеза1 v2 ← гіпотеза2 виснов ← v1 if крит1 < крит2 v2 otherwise
---------------------------	---

Таким чином: НашВибір(Xi1, Xi2) = "Пуасонівський" розподіл. Значення критерію для нього у разів(Xi1, Xi2) = 2.11 краще, ніж для конкурентної гіпотези.

Працюючи з електронною версією книги можете понатискати зону введення даних і подивитись як буде змінюватись змінна "НашВибір". Не стріляйте в статистику - вона робить те, що може. Іноді дуже багато.

Можна перевірити (при бажанні) гіпотези за критерієм Колмогорова. Для цього обчислюємо максимальне відхилення D та параметр L. Число стовпців (груп) гістограм M = 8 .

$$dF1_q := \frac{D2_{q,1}}{K_{\text{рок}}} - \text{dpois}(D2_{q,0}, \text{lao}) ; D := |\max(dF1)| ; L1 := D \cdot \sqrt{M} ; L1 = 0.052$$

$$dF2_q := \frac{D2_{q,1}}{K_{\text{рок}}} - \text{dnorm}(D2_{q,0}, \mu_0, \text{sgm}_0) ; D := |\max(dF2)| ; L2 := D \cdot \sqrt{M} ;$$

$$L2 = 0.064$$

За таблицями Колмогорова визначаємо величину імовірності $P(L)$ і приймаємо рішення гіпотезу прийняти/не прийняти, в залежності від вибраного нами рівня довіри. Якщо обчислене значення критерію менше значення критерію для заданого рівня довіри - гіпотеза приймається.

1.4.6. Розробка документа для оцінки впливу шумів в емпіричних даних на надійність результатів ідентифікації

А тепер виконаємо задачу апроксимації для інших емпіричних даних (у курсовій роботі необхідно ідентифікувати частотні розподіли мінімум для двох підсистем, процесів, наприклад - для різних ділянок конвеєрної системи, для процесів прибуття клієнтів і обслуговування клієнтів).

Потренуємось у застосуванні **стандартних інформаційних технологій** - скопіюємо та модифікуємо попередні модулі. Поставимо дослідницьку задачу - **визначення впливу рівня шумів і помилок в експериментальних даних на стійкість результатів ідентифікації**

Як конкурентні гіпотези розглянемо **експоненційний** та **геометричний** розподіли. Це теж прототип для виконання розділу курсової роботи. Скопіюємо і ущільнимо відповідні розділи документа.

Побудова гістограми

1) **Вводимо (генеруємо) вхідні дані**

ORIGIN := 0 ; X3 = rgeom(100, .2) ; N := length(X) ; N = 100 .

Проводимо попередній аналіз даних - визначаємо мінімум, максимум, середнє.

$\min(X3) = 0$; $\text{mean}(X3) = 4.5$; $\max(X3) = 18$.

Блок обчислення гістограми. ORIGIN := 0 Вхід: таблиця даних X3	
$K_{\text{рок}} := \text{ceil}\left[\frac{(\max(X3) - \min(X3)) \cdot 1.2}{K_{\text{ілк_інт}}}\right]$	$\text{Int}_q := \text{ceil}[K_{\text{рок}} \cdot (q - 1) + \min(X3) \cdot 0.98]$
$\text{Rem} := \text{hist}(\text{Int}, X3)$	$\text{Rem}_{K_{\text{ілк_інт}}-1} := 0$ Емпіричний розподіл: $\text{Rem} := \frac{\text{Rem}}{N}$
$D3 := \text{augmen}[\text{Int} + 0.5 \cdot K_{\text{рок}}, \text{Rem} \cdot [(1 - \text{sh}) + \text{rnd}(2 \cdot \text{sh})]]$	

В агреговані дані додаємо імітацію шумів вимірювання **sh = 1** - шум (вводиться в кінці документа). Вхідні неагреговані дані (генератор даних), на базі яких блок БГ обчислює таблицю D3 розташовані далі на відповідному стенді.

$D3^T =$		0	1	2	3	4	5	6	7
	0	-1.5	1.5	4.5	7.5	10.5	13.5	16.5	19.5
	1	0	0.739	0.41	0.164	0.131	0.082	0.082	0

Задаємо гіпотетичний розподіл: $rt(t, \mu) := dgeom(t, \mu)$.

Отримали гістограму, вибрали гіпотетичний теоретичний розподіл, тепер переходимо до розв'язання оптимізаційної задачі - шукаємо значення параметрів теоретичного розподілу, що дають мінімум квадратичної помилки

1) Визначаємо кількість точок апроксимації (=число стовпчиків гістограми) і початкові значення оптимізуємих параметрів:

$$n := \text{rows}(D3) - 1; \quad n = 7; \quad q := 0..n; \quad \mu := .2.$$

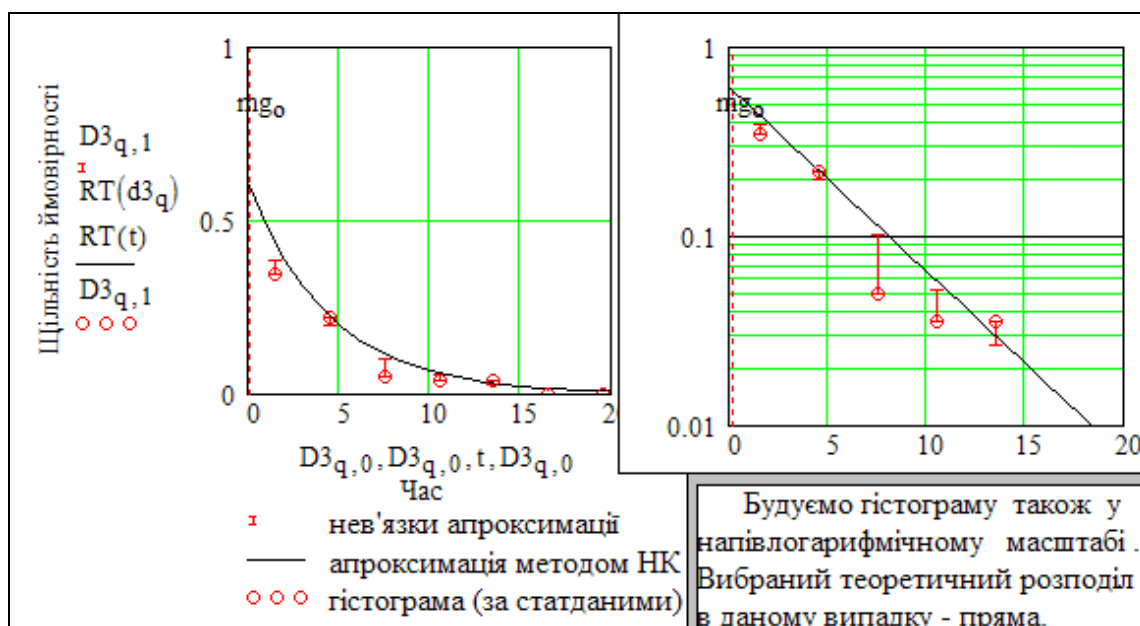
2) Записуємо систему рівнянь (критерій, обмеження, в даному випадку - одне рівняння). Мікрокоментар. Замість прирівнювати похідні до нуля, шукаємо безпосередньо мінімум.

$$\text{Given} \quad \sum_{q=0}^n \left(\frac{D3_{q,1}}{\text{Крок}} - rt(D3_{q,0}, \mu) \right)^2 = 0$$

$$mg_0 := \text{MinEr}(\mu)$$

3) Будуємо графік - перевіряємо, чи справді функція збігається з емпіричними даними. Виводимо оптимальне значення параметру $mg_0 = 0.2$ **Вводимо дані:**

$$RT(t) := rt(t, mg_0) \cdot \text{Крок}; \quad t := 0..20; \quad d3_q := \text{round}(D3_{q,0}, 0); \quad X3 \equiv rgeom(100, .2)$$



Робимо теж саме для конкурентної гіпотези для тих же даних, але без повторень і коментарів.

Побудова гістограми. Виводимо таблицю агрегованих даних - значення гістограми та інтервалів.

$$D3^T =$$

	0	1	2	3	4	5	6	7
0	-1.5	1.5	4.5	7.5	10.5	13.5	16.5	19.5
1	0	0.726	0.259	0.156	0.078	0.052	0	0

Задаємо ГІПОТЕТИЧНИЙ розподіл: $rt(t,r) := \text{dexp}(t,r)$; $\text{ORIGIN} := 0$.

Отримали гістограму, вибрали гіпотетичний теоретичний розподіл, тепер переходимо до розв'язання оптимізаційної задачі - шукаємо значення параметрів теоретичного розподілу, що дають мінімум квадратичної помилки.

Знаходження оптимальної апроксимації

1) Визначаємо кількість точок апроксимації (=число стовпчиків гістограми) і початкові значення оптимізуємих параметрів:

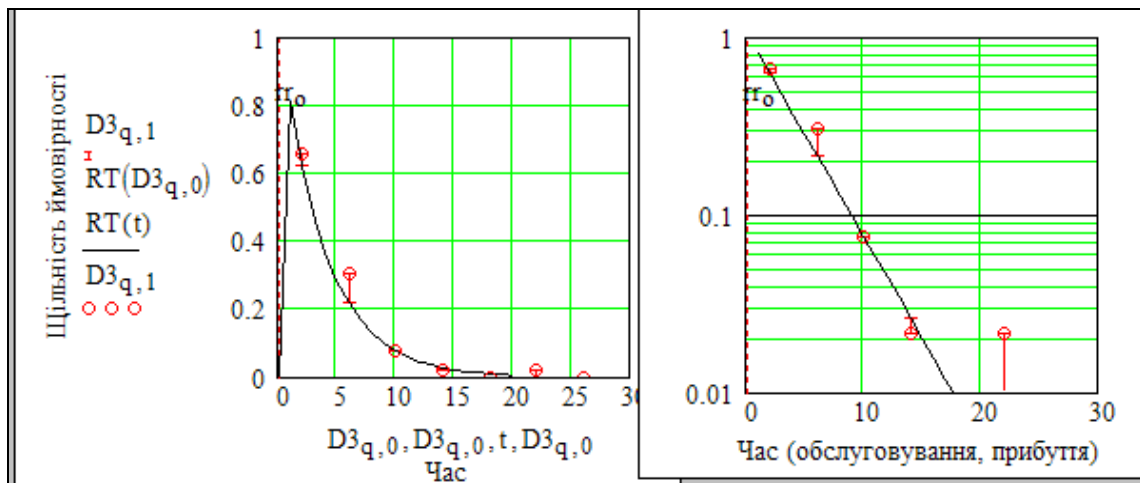
$$n := \text{rows}(D3) - 1 \quad n = 7 \quad q := 1..n \quad \mu := .2.$$

2) Записуємо систему рівнянь (критерій, обмеження, в даному випадку - одне рівняння). Мікрокоментар. Замість прирівнювання похідних до нуля, шукаємо безпосередньо мінімум.

$$\text{Given} \quad \sum_{q=0}^n \left(\frac{D3_{q,1}}{\text{Крок}} - rt(D3_{q,0}, \mu) \right)^2 = 0 \quad ; \quad rr_0 := \text{MinErr}(\mu).$$

3) Будуємо графік - перевіряємо, чи справді функція наближує емпіричні дані. Увага! Використовуємо ті ж самі дані (вище) перевіряємо другу гіпотезу. Виводимо результати - оптимальне значення параметра μ_0 . **Введення вхідних даних:**

$$RT(t) := rt(t, rr_0) \cdot \text{Крок}; \quad t := 0..20; \quad rr_0 = 0.008 \quad (\text{введено вище}).$$



Перевірка вірогідності гіпотези про теоретичний розподіл за критерієм Пірсона. Тепер проведемо перевірку гіпотез про те, що емпіричний є або: гіпотеза1 := "геометричний", або: гіпотеза2 := "експоненційний" за критерієм Пірсона. Подаємо послідовність кроків по перевірці гіпотези про теоретичний розподіл.

1) **Обчислюємо значення критерію** за формулою

$$X_{G1} := \sum_{q=0}^n \frac{\left(\frac{D3_{q,1}}{\text{Крок}} - \text{dgeom}(D3_{q,0}, mg_0) \right)^2}{\text{dgeom}(D3_{q,0}, mg_0) + 0.005}$$

Для гіпотеза1 = "геометричний" розподіл (цілочислений) значення критерію
 $X_{G1} = 0.059$

$$X_{G2} := \sum_{q=0}^n \frac{\left(\frac{D3_{q,1}}{\text{Крок}} - \text{dexp}(D3_{q,0}, rr_0) \right)^2}{\text{dexp}(D3_{q,0}, rr_0)}$$

Для гіпотеза2 = "експоненційний" розподіл (неперервний) значення критерію
 $X_{G2} = 0.125$

Вводимо випадковий шум $sh = 1$ - так ми перевіряємо працездатність критерію Пірсона. Випадкові величини - це випадкові величини, задаємо (F9) нові реалізації входних даних, і спостерігаємо, як змінюються значення критеріїв.

2) Визначаємо імовірність такого відхилення. Для цього спочатку треба отримати розподіл χ -квадрат із заданим числом ступенів вільності.

За таблицею або графіком розподілу визначаємо імовірність такого значення критерію для χ -квадрат розподілу із числом ступенів вільності dv . Число ступенів вільності dv - це число груп (інтервалів) гістограми $M - 1 - r$ (r - число параметрів теоретичного розподілу, наприклад, для

геометричного $r1 := 1$, для експоненційного теж $r2 := 1$).

$M := \text{Кільк_інт}$; $dv1 := M - r1 - 1$; $dv2 := M - r2 - 1$; $xi := .0, .5 .. 24$.

Число ступенів вільності розподілу χ -квадрат для теоретичних розподілів буде

a) *геометричне* $dv2 = 6$ б) *експоненційне* $dv1 = 6$.

Використовуємо вбудовані функції для χ -квадрат розподілу.

$$Xi(xi, d) := dchisq(xi, d)$$

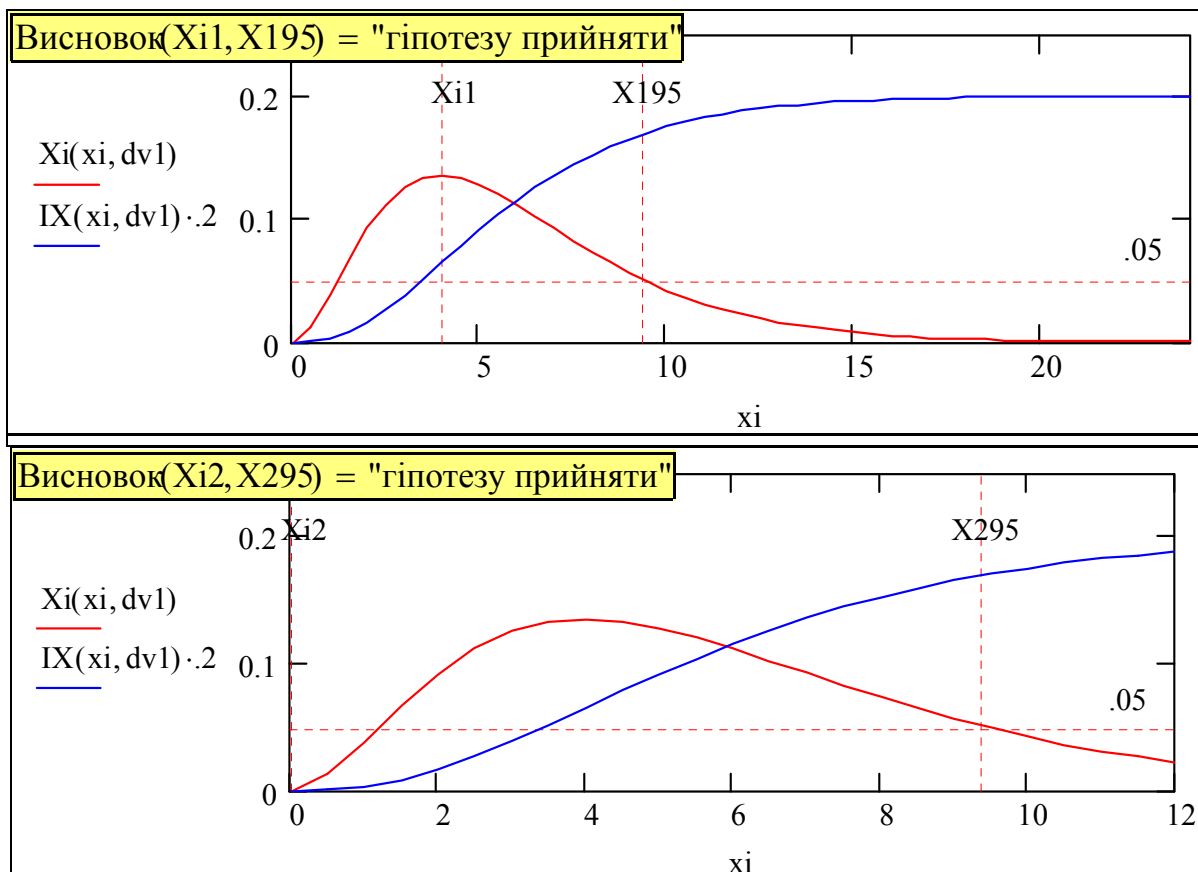
$$IX(xi, d) := pchisq(xi, d)$$

Визначаємо значення $XI95$, що відповідає ймовірності 95%. $xx := 13$;

$XI95 := \text{root}(dchisq(xx, dv1) - .05, xx)$; $XI95 = 9.392$;

$X295 := \text{root}(dchisq(xx, dv2) - .05, xx)$; $X295 = 9.392$.

3) Визначаємо за графіком або таблицею розподілу χ -квадрат ІМОВІРНІСТЬ отриманого значення і ВИРІШУЄМО, прийняти, чи відкинути гіпотезу.. - понатискайте оцю кнопку (F9) $sh \equiv 1$ і думайте, коли користуєтесь витонченими статметодами.



Підсумки і висновки

1. Не довіряйте своїм даним - ретельно проаналізуйте їх - чи можуть там бути великі помилки, чи не є вони сумішшю даних із різних джерел, по різному вимірених, чи не можна їх якось розшарувати, скоректувати.

2. Проаналізуйте можливі **механізми** (процеси) породження даних.

3. Статистика не завжди скомпенсує недостатні і недостовірні дані - головна задача - вдосконалювати техніку і джерела отримання даних, вводити альтернативні та незалежні канали вимірювання та передачі даних.

4. Збільшуйте число точок контролю та управління для вашого складного розподіленого об'єкта.

5. Побудуйте графіки подібні поданим вище. Оцініть розкиди значень критеріїв Пірсона, Колмогорова. Після всього цього можете використати ці критерії для "полірування" вашого аналізу.

6. Після всього цього можете використати обчислені значення критеріїв - Пірсона, Колмогорова - для "полірування" вашого аналізу.

Тепер, на закінчення, пару слів про суть справжньої науки, а саме - "безпринциповість". Однак, як відрізнити справжню науку від несправжньої? От, наприклад, Нобелівські премії в галузі економіки отримують переважно вчені, яким за 60, за роботи, які вони започаткували 20-50 років тому. Тобто, премії дають тоді, коли "безумні теорії" стали нормою практики.

Розглянемо два приклади.

Джемс Мірлі, лауреат 1996 року, вів дослідження оптимальних рівнів оподаткування і прийшов до висновку, що для суспільства краще, якщо багаті будуть платити не 70%, а менше 20%. Він тоді був радником уряду лейбористів. Ось його слова: "Я мушу визнати, що розпочинав це дослідження з метою отримання аргументів на захист високих ставок. А вийшло навпаки".

Рональд Коуз, лауреат 1991 року, створив дуже незвичну теорію фірми, згідно з якою головними факторами утворення фірм є не виробничі, а організаційні та юридичні витрати. Тобто головне - не виробництво, а складання, підписування, пересилання "папірців" - діловодство. Природно, що при обговоренні його книги "Проб лема соціальних витрат" на семінарі спочатку всі вчені були проти, але після дискусії всі були "за".

При користуванні статистичними методами *ну дуже потрібне* саме таке поєднання неупередженості і найвищої кваліфікації.

2

Елементи теорії масового обслуговування

У цьому розділі розглядаються:

- ✓ Елементи теорії масового обслуговування. Розрахунок та моделювання однофазної одноканальної СМО.
- ✓ Розрахунок та моделювання багатофазної одноканальної СМО (конвеєрне виробництво)

Після вивчення цього розділу ви **зможете:**

1. Будувати і аналізувати схеми систем масового обслуговування - багатоканальні, багатофазні, з різними дисциплінами обслуговування.
2. Формувати критерії ефективності систем масового обслуговування, що враховують інтереси клієнта і виконавця.
3. Виконувати розрахунок характеристик СМО.
4. Розробляти програми для імітаційного моделювання СМО.
5. Розробляти програми моделювання СМО на базі моделей марківських процесів.
6. Розробляти програми розрахунку характеристик СМО на базі згорток частотних розподілів.
7. Розробляти інтерфейси для програм класу „аналітичний помічник менеджера”.
8. Виконувати моделювання і дослідження СМО.

9. Дати визначення термінам:

- | | |
|-------------------------------------|--|
| 1. Канал обслуговування. | 5. Частотний розподіл часу обслуговування. |
| 2. Фаза обслуговування. | 6. Частотний розподіл довжини черги. |
| 3. Дисципліна обслуговування. | 7. Пропускна спроможність. |
| 4. Частотний розподіл потоку вимог. | 8. Темп роботи конвеєра. |



2.1 Елементи теорії масового обслуговування

Вступ. Історія

В теорії масового обслуговування основним є поняття черги. Тому її називають ще теорією черг. Кожен знає, що таке черга. Наведіть три-чотири приклади черг. Наприклад, черга в крамниці, черга виробів на конвеєрі біля місць виконання операцій, машини біля світлофорів, літаки в повітряній зоні аеропорту (посадка) та біля злітно-посадкової смуги (зліт), черги до інтернетівських серверів, черги на телефонних станціях, черга ще нерозбомблених цілей (є і воєнні СМО).

В деяких системах черга небажана взагалі, в деяких - припустима обмежена, а в деяких - необхідна (подайте приклад?). Але сьогодні черги стали іншими - черги виробників і постачальників до клієнтів. Спочатку стисло **переглянемо добутки класичної теорії масового обслуговування.**

Між іншим, до чого картинка вгорі? Тому, що це класика і продуктивний напрям, концепція. Тому, що головна ідея там - обслуговування незалежно від якихось там імовірностей завжди повинно бути на висоті.

Основні задачі теорії масового обслуговування (ТМО) - визначення:

- 1) максимальної довжини черги;
- 2) необхідної швидкості обслуговування;
- 3) кількості приладів обслуговування, що працюють паралельно.

Щоб відповісти на ці питання, треба побудувати математичні моделі СМО. Експериментуючи з моделлю СМО, можемо отримати потрібні відповіді. Дані, що отримані при спостереженні реальних систем використовуються для верифікації моделі.

Трохи історії. В 1909-1917 роках датський математик А.К.Ерланг опублікував важливі результати дослідження теоретичних моделей телефонних систем. Потім Д. Кенделом було виконано узагальнення СМО, зокрема виконана класифікація СМО, що дещо застаріла на сьогодні в зв'язку з розширенням задач і схем побудови СМО.

Для аналізу і синтезу СМО можуть бути використані:

- 1) детермінована модель черги (для розрахунку беруться середні значення інтенсивності вхідного потоку і часу обслуговування з деякими "запасами")
- 2) модель на базі теорії розподілів (частотні розподіли для інтенсивності вхідного потоку і часу обслуговування);
- 3) марківська модель черги (виділяються певні стани об'єкта обслуговування та імовірності переходів між ними);
- 4) модель на базі імітаційного моделювання випадкових процесів надходження об'єктів і обслуговування.

Загальний опис СМО

Розглянемо узагальнену схему СМО (Рис.1). В загальному випадку обслуговування заявок може бути багатоканальним і багатofазним. Фази обслуговування - це операції, що виконуються в певній послідовності, відповідними приладами, канали обслуговування - паралельно працюючі прилади певної фази обслуговування.

Неважко дати класифікацію багатофазних і багатоканальних систем, неважко знайти приклади таких систем. Далі буде дано визначення і класифікація СМО.

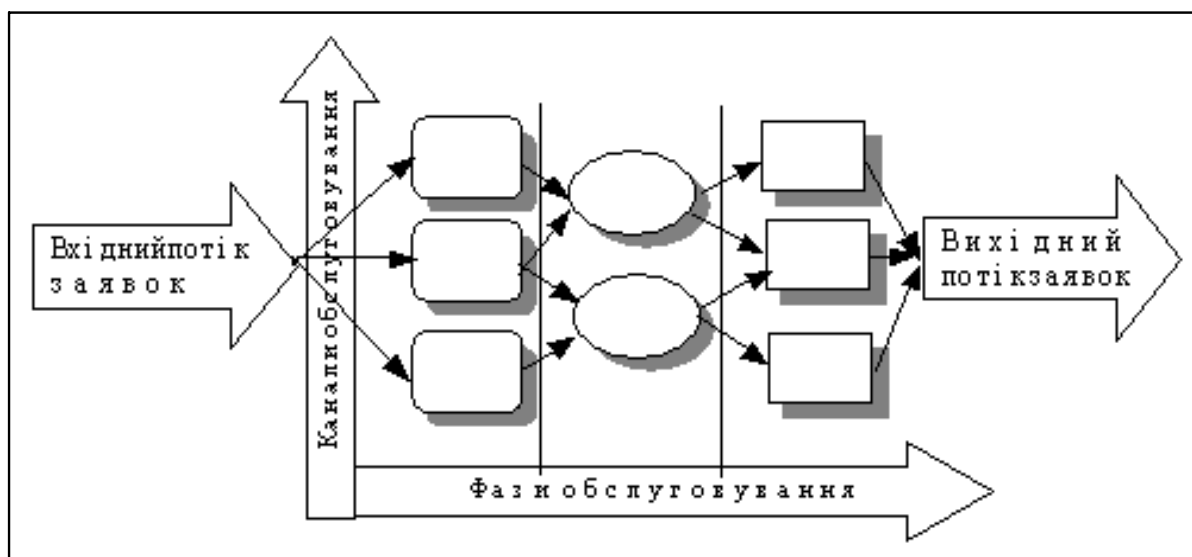


Рис. 1. Узагальнена схема системи масового обслуговування

Виконаємо опис СМО за такими пунктами:

- а) основні компоненти моделі СМО;
- б) основні характеристики СМО.

Основні компоненти моделі СМО

1. Опис вхідного потоку вимог (імовірнісні характеристики, сезонні коливання, тренд...) 2. Опис способу, яким виконується обслуговування - дисципліна обслуговування (імовірнісні характеристики та ін.).

3. Опис дисципліни черги (наприклад: "першим прийшов - першим обслужений", "останнім прийшов - першим обслужений", пріоритети різним категоріям заявок, наприклад - пожежні машини заправляти в першу чергу, міліцію - в другу, усіх інших - без пріоритетів...).

Завжди слід пам'ятати, що неадекватність математичного подання реальної системи, з багатьма спрощеннями та ідеалізаціями - неминуча. Це правило у використанні математики при вивченні реальних процесів, тому треба **критично сприймати** результати, що отримані на моделі і перевіряти їх експериментально.

Основні характеристики СМО

Головний інтерес мають такі характеристики.

- 1) Довжина черги в різні моменти часу (динаміка черги).
- 2) Загальна тривалість знаходження вимоги (клієнта. об'єкта) в системі.
- 3) Доля простоїв, або, навпаки, - інтенсивність завантаження обслуговуючих пристроїв.

Класифікація систем масового обслуговування

Для СМО розроблено стандартну класифікацію за різними ознаками, серед яких найбільш важливими є:

- структура системи;
- закон розподілу потоків вимог, що надходять у СМО;
- закон розподілу часу обслуговування, порядок (дисципліна) обслуговування;
- допустима довжина черги перед обслуговуючими приладами (каналами).

Будь-яка система характеризується своєю структурою, тобто складом її основних частин і функціональними зв'язками між ними. Для СМО такими частинами є обслуговуючі прилади (канали). Якщо система містить тільки один канал, така СМО називається **одноканальною**. Однак в ряді випадків система може мати декілька каналів, призначених для одночасного (паралельного) обслуговування вимог, що надходять в систему. Такі СМО називають **багатоканальними** (рис.1).

Обслуговування може бути спеціалізоване по окремим операціям - фазам. Зрозуміло, що робиться це для підвищення продуктивності системи, підвищення її гнучкості, стійкості до відмов. Такі СМО називаються багатofазними. В кожній фазі може бути декілька каналів, що обслуговують один і той же потік вимог. В багатofазній системі вимога вважається повністю обслугованою тільки тоді, коли послідовно пройде усі фази обслуговування. Кожну фазу можна розглядати як одно - або багатоканальну СМО.

В практиці проектування і експлуатації реальних систем найбільш часто зустрічаються такі закони розподілу: **пуасонівський - М, рівномірний - D, ерлангівський - Е, довільний - G.**

Для характеристики СМО використовується символіка, запропонована **Д. Кенделом** [5,7], яка включає в себе п'ять розрядів, призначених для характеристики відповідно:

- 1) потоку вимог,
- 2) закону часу обслуговування вимог,
- 3) особливості структури,
- 4) особливості черги,
- 5) опис дисципліни обслуговування.

В залежності від особливостей реальних систем і їх призначення на довжину черги перед фазами СМО можуть накладатись обмеження. В цьому випадку розрізняють СМО з обмеженою або необмеженою довжиною черги. Вимоги, що надходять на СМО з обмеженою чергою, вважаються втраченими, якщо в момент їх приходу черга була заповнена. СМО, для яких вимоги можуть втрачатися, називаються СМО з втратами. Для СМО з необмеженою чергою вимоги, що надходять, коли усі канали зайняті, чекають початку свого обслуговування в черзі доти, поки якийсь канал не звільниться від обслуговування. Розрізняють СМО з пріоритетним обслуговуванням і СМО з дисципліною: "першим прийшов - першим обслужений".

В залежності від загального числа джерел вимог U_i ($i=1,..k$), кількості направлених від них вимог і повернення обслужених вимог до джерела, розрізняють замкнені та розімкнені СМО. Замкненими системами (системи з обмеженим вхідним потоком) називають системи, які призначені для обслуговування обмеженого числа джерел вимог. Загальне число вимог, що циркулюють в замкненій системі - обмежене (звичайно - постійне).

В тому випадку, коли число джерел вимог необмежене, або коли одне джерело посилає в систему необмежене число вимог, СМО називають розімкненими.

Розглянемо приклади кодування класів СМО.

Кодове означення "система **М/М/1/∞**", розшифровується так:

М - вхідний потік вимог - пуасонівський;

М - розподіл часу обслуговування - експоненційний;

1 - число каналів = 1;

∞ - обмеження на довжину черги нема (нескінченна допустима черга).

Тепер розшифруйте самостійно: "система E/G/12/5"

(Ерлангівський вхідний потік вимог, довільний закон часу обслуговування, 12 каналів, обмежена черга: "клієнт побачив, що черга біля каси 5 чоловік і передумав обслуговуватись", або: "летіли 5 літаків, їх позбавили, 6-тий літак не встигли обслужити (збити) - пролетів")

Показники ефективності СМО

До показників ефективності СМО відносять:

- розподіл імовірності для числа вимог n в системі;
- середнє число вимог в черзі або в системі ;
- розподіл часу очікування в черзі або часу перебування вимоги в системі ;
- середнє і дисперсія часу очікування ;
- середня тривалість інтервалу зайнятості системи ;
- середнє число обслуговуваних за одиницю часу вимог;
- розподіл імовірності для числа зайнятих каналів системи (ну уявіть собі АТС);
- вартості: каналів обслуговування, втрат від несвоєчасного обслуговування (черг), вартості диспетчеризації (управління, перенаправлення) потоків вимог та ін.

Усі показники якості обслуговування систем можна умовно розділити на три види: **імовірнісні, моментні, економічні**, які тісно пов'язані між собою. З імовірнісних характеристик широко застосовуються такі як:

$P(k = n)$, або, скорочено $P(n)$ - імовірність того, що в системі знаходиться n вимог;

$P(k > n_{\max})$ - імовірність втрати поточної вимоги, та ін.

З моментних характеристик до найбільш важливих відносяться:

- середнє число вимог в системі:

$$n_{\text{середн}} = \sum_n P(n) \cdot n; \quad n = 0 \dots \infty; \quad (1)$$

- середнє число вимог в черзі (s - число каналів СМО)

$$v_{\text{середн}} = \sum_n P(n) \cdot (n - s); \quad n = s + 1 \dots \infty; \quad (2)$$

- середнє число вимог в каналах обслуговування

$$j_{\text{середн}} = \sum_n P(n) \cdot n; \quad n = 0 \dots s \quad (3)$$

- середнє число вільних приладів

$$r_{\text{середн}} = \sum_n P(n) \cdot (s - n); \quad n = 0 \dots s; \quad (4)$$

- середній час очікування в черзі

$$\omega_{\text{середн}} = \int_{t_{\min}}^{t_{\max}} p(\omega) \cdot \omega d\omega; \quad p(\omega) = \frac{d}{d\omega} P(\omega), \quad (5)$$

де ω - змінна інтегрування;

- середня тривалість перебування вимоги в системі

$$u_{\text{середн}} = \int_{t_{\min}}^{t_{\max}} g(\omega) \cdot \omega \, d\omega ; \quad g(\omega) = \frac{d}{d\omega} G(\omega) . \quad (6)$$

В цих формулах прийняті такі означення

$F(t)$ - функція розподілу часу очікування вимоги в черзі;

$G(t)$ - функція розподілу часу перебування вимоги в системі;

$P(n)$ - функція розподілу числа вимог в системі;

Для кожного моменту часу виконуються такі рівняння.

Середнє число обслугованих вимог + середнє число вільних приладів = число каналів системи:

$$j_c + r_c = s. \quad (7)$$

Середнє число вимог в черзі + середнє число вимог на обслуговуванні = середньому числу вимог в системі:

$$v_c + j_c = n_c . \quad (8)$$

Середня тривалість обслуговування вимог + середня тривалість очікування вимоги в черзі = середній тривалості перебування вимоги в системі:

$$t_{ob_c} + \omega_c = u_c. \quad (9)$$

В усталеному стані роботи СМО, якщо такий існує, вихідний потік дорівнює вхідному: $\lambda_{vx} = \lambda_{вих} = \lambda_c$. Якщо, крім того, вхідні вимоги не залежать від вихідних (розімкнена система), то попередні рівняння приводяться до вигляду:

інтенсивність потоку вимог * середня тривалість перебування вимоги в системі = середньому числу вимог в черзі:

$$\lambda_{\text{вимог/хвилину}} \cdot \omega_{\text{хвилин}} = v_{\text{вимог}} , \quad \text{або} \quad \lambda_c \cdot \omega_c = v_c ; \quad (10)$$

інтенсивність потоку вимог * середня тривалість перебування вимоги в системі = середньому числу вимог в системі:

$$\lambda_{\text{вимог/хвилину}} \cdot u_{\text{хвилин}} = n_{\text{вимог}} , \quad \text{або} \quad \lambda_c \cdot u_c = n_c . \quad (11)$$

Очевидно, що чим менше значення таких показників, як **v**, **u**, **n**, тим вища ефективність системи з **точки зору джерела вимог** (клієнта). Однак, зменшення значень цих показників пов'язане з підвищенням **економічних витрат виконавця**.

До складу узагальненого критерію ефективності СМО включають **вартість очікування вимог** в черзі і **вартість простою каналів**. Перша складова - втрати клієнта, друга - витрати виконавця. Чи буде виконавець враховувати першу складову? - В умовах конкуренції і, якщо він планує бізнес надовго, - буде обов'язково.

Розглянемо приклад побудови моделі вартості СМО з **s** каналами. Задаємо період **T**, для якого треба розрахувати середню вартість **Варт(s)**, через **C1черг** означимо вартість (втрати) очікування однієї вимоги за одиницю часу, а через **C2прост** - вартість простою одного каналу в одиницю часу, отримуємо:

$$\text{Варт}(s) = (C1_{\text{черг}} \cdot v_c + C2_{\text{прост}} \cdot r_c) \cdot T. \quad (12)$$

Підставляємо значення середнього числа вимог в черзі (2) та середнього числа вільних приладів (4) отримуємо:

$$\text{Варт}(s) = \left[C1_{\text{черг}} \cdot \sum_{n=s+1}^{\infty} P(n) \cdot (n-s) + C2_{\text{прост}} \cdot \sum_{n=0}^s P(n) \cdot (s-n) \right] \cdot T. \quad (13)$$

Якщо вартості $C1_{\text{черг}}$ и $C2_{\text{прост}}$ постійні в періоді T , то можна використати "повну приведену вартість СМО за одиницю часу" [2-4]:

$$\gamma(s) = C1_{\text{черг}} \cdot \sum_{n=s+1}^{\infty} P(n) \cdot (n-s) + C2_{\text{прост}} \cdot \sum_{n=0}^s P(n) \cdot (s-n). \quad (14)$$

Щоб обчислювати, мінімізувати ці витрати, необхідно отримати функцію $P(n)$ - розподіл імовірностей для кількості вимог в системі (вимоги в черзі + вимоги на обслуговуванні).

Який же з названих вище методів розрахунку СМО вибрати? Коротко охарактеризуємо їх:

- 1) **детермінована модель** черги - дуже проста, але мало реалістична;
- 2) **модель на базі теорії розподілів** ефективна для узагальнень, прогнозів, оптимізації, але такі моделі існують для небагатьох випадків. Такі моделі важко отримати;
- 3) **марківські моделі СМО** - дозволяють отримувати результати як аналітичними методами так і числовими, мають певні обмеження - не завжди є реалістичними;
- 4) **методи на базі імітаційного моделювання** - найбільш поширені в інженерній практиці. Дозволяють враховувати усі особливості СМО. Недолік - дають тільки числові результати. Згідно з совковим стандартом проектних розробок Ви повинні вибрати з цих методів "оптимальний".

Для тих, хто має уяву про методологію науки і сучасні інформаційні технології, неважко вибрати "оптимальний варіант": ці методи взаємодоповнюють, взаємоконтролюють один одного, тому і використовуємо їх разом. Починаємо з імітаційного моделювання, що іноді характеризується так: *"...грубый силовой приём, используемый в ситуациях, когда другие средства не дают результата"*.

Розрахунок та моделювання однофазної одноканальної СМО

Розробимо імітаційну математичну модель СМО. Задаємо такі змінні для описання стану одноканальної однофазної системи.

$S_{1,t}$ - випадкова подія - прибув (1), чи не прибув (0) клієнт;

$S_{2,t}$ - випадкова подія - обслужений (0), чи поки не обслужений (1) клієнт;

$S_{3,t}$ - подія - прилад вільний (1), зайнятий обслуговуванням (0);

$S_{4,t}$ - кількість клієнтів (вимог) у черзі;

$S_{5,t}$ - кількість тактів чергового обслуговування пройшло (і ще не скінчилось).

Завдання. Напишіть сценарій роботи СМО - він немовби очевидний?

Запишемо спочатку загальне матрично-векторне рівняння процесів у СМО.

Його словесна формула: "поточний стан дорівнює операторному перетворенню від попереднього стану", тобто:

$$S^{(t+1)} = \mathfrak{R}(S^{(t)}),$$

де $\mathfrak{R}$ - нелінійний оператор.

Наша задача - деталізувати цей оператор - записати систему рівнянь і реалізувати її у вигляді програмного модуля. Сучасні програмні середовища допускають інтеграцію класичних етапів "математична модель-алгоритм-програма". Але, як і на початку ери програмування, діє закон "програма більша десяти рядків - не відлагоджується".

Тому спочатку розробимо і відтестуємо модулі для імітації випадкових подій (*прибуття клієнта, обслуговування клієнта*). Розглянемо дві версії модуля - перший для довільних розподілів величини (*інтервалу прибуття, часу обслуговування*), другий для окремого, але важливого випадку, коли імовірність події на кожному кроці (такті) функціонування СМО постійна і незалежна від інших подій.

Моделюємо СМО як імпульсну систему, що працює з певним кроком квантування T . Як вибрати цей крок моделювання системи? Використаємо знання з курсу теорії автоматичного управління - там крок квантування для моделювання неперервної динамічної системи ми вибирали досить малим, щоб досить точно передати поведінку неперервної системи. Виберемо крок моделювання так:

$$T_{\text{моделювання}} = \delta \cdot T_{\text{середній_інтервал}}; \quad \delta = 0.1 - 0.2$$

Тобто, беремо крок моделювання в 5-10 разів менший, ніж середній інтервал між прибуттям клієнтів. Тепер, виходячи з цього припущення формуємо логіку роботи модуля: генерується випадкове число int із заданим розподілом. Воно інтерпретується як кількість тактів $T_{\text{мод}}$ до прибуття чергового клієнта. Увага! Це внутрішня, штучна змінна системи моделювання, вона "невідома" системі, її не можна використовувати для прогнозування і планування.

прб :=	int ← 0	-початкове значення змінної залишилося кроків до прибуття	прб =
	for t ∈ 1..12	-цикл по кроках моделювання	
	if int < 1	-якщо інтервал = 0,	1 1
	s1 _t ← 1	-клієнт прибув.	2 1
	int ← rbinom(1,20,.1) ₁	Генерується новий інтервал прибуття	3 0
	otherwise		4 1
	int ← int - 1	-інакше	5 0
	s1 _t ← 0	-до прибуття на крок менше	6 0
	"Endfor"	-прибуття клієнта на кроці моделювання не було	7 1
s1		-програма повертає вектор подій "прибуття клієнта"	8 0
			9 0
			10 1

Вище поданий приклад програмного модуля, що реалізує цю ідею. Як можна бачити, заміною одного рядка - генератора чисел із заданим розподілом модуль можна настроїти на інший розподіл. Можна просто зробити модуль функцією від функції розподілу (тобто параметром функції буде не число, а функція).

Приклади змінних елементів модуля: $| \text{rnorm}(1, 0, 10) |_1$; $\text{rpois}(1, .3)_1$.

Подаємо другу версію програми для випадку, коли імовірність події на кожному кроці (такті) функціонування СМО постійна і незалежна від інших подій.

Зауваження. В програмі два зайвих рядки. Один з них - коментар у лапках - підвищує читабельність програми. Не тільки старше, але й молоде покоління, що крім класичного Паскаля нічого не знає, звикло до блок-схем і структурованих текстів програм, де усі структурні одиниці (цикли, розгалуження) мають початок і кінець (begin - end, for - endfor, if - endif... ets.).

Тому природно дописати коментарі чи вирази, що замінюють "скорочені" операторні дужки. Якщо уважно придивитись, можна побачити, що роль операторних дужок виконують вертикальні риси. Програма оформлена як підпрограма з параметром mu - імовірністю події (прибуття, обслуговування)

prb(my) :=	sto ← 100	prb(0.2) =		1	prb(0.7) =		1
	for t ∈ 1..8		1	1		1	0
	sl _t ←		2	0		2	0
	0 if rnd(1) > my		3	0		3	1
	1 otherwise		4	1		4	1
	"next t"		5	0		5	0
	sl		6	1		6	0
			7	0		7	0
			8	0		8	1

Розробка програми моделювання СМО

Переходимо до розробки повної програми моделювання. Ми принципово не використовуємо викликів підпрограм, а просто зібрати цю програму з модулів, подібних розглянутим вище . Програма невелика, і ми можемо бачити разом фактично всю систему рівнянь СМО в термінах змінних s_1 - s_5 . Ми можемо змінювати все прямо в тексті програми і поряд з текстом програми бачити результати. Це зручно для пошукових досліджень та навчання.

Більш складну програму моделювання слід розбити на функціональні модулі, які можна автономно відлагоджувати і легко замінювати в головній програмі.

Текст нашої програми є гібридом лістингу і блок-схеми. Написано цей текст в основному системою програмування. Оператори, блоки, структурні елементи - все це вибрано з меню програмування.

І саме головне - цю програму легко оформити як функцію відповідних параметрів. Слід пам'ятати, що змінні визначені поза програмою - **глобальні** і в програмі їх **видно**, а змінні, визначені в програмі - **локальні**, їх **видно** тільки в цій програмі і **не видно** за її межами.

Наша програма не іграшкова - вона зібрана з ефективних вбудованих модулів пакета, що написані професіоналами на мові VC++. Задаємо кількість кроків моделювання: $\text{Kolstep} := 200$, $t := 1.. \text{Kolstep}$, $\text{ORIGIN} \equiv 1$.

<pre> Ss := inp ← 0 ino ← 0 sp<1> ← пст for t ∈ 1..Kolstep if inp < 1 sp1,t ← 1 inp ← rbinom(1,N,1-q)1 otherwise inp ← inp - 1 sp1,t ← 0 sp4,t+1 ← sp4,t + sp1,t if sp3,t ∧ sp4,t ino ← rpois(1,ля)1 sp5,t ← 0 sp4,t+1 ← sp4,t - 1 sp3,t+1 ← 0 if sp3,t+1 < 1 ino ← ino - 1 sp5,t+1 ← sp5,t + 1 if ino < 1 sp3,t+1 ← 1 sp5,t ← 0 "endfor" sp </pre>	inp - інтервал часу до прибуття ino - інтервал часу до кінця обслуговування - введення початкового стану. Цикл по кроках моделювання. - вже розглянутий вище модуль імітації випадкової події "прибуття клієнта" Розподіл інтервалів прибуття в даному випадку - біноміальний з параметрами N, q. Для використання іншого розподілу досить замінити цей рядок програми. - черга на наступний період - якщо прилад вільний і черга не нульова генерація випадкового інтервалу обслуговування чергового клієнта - обнулення лічильника тактів обслуговування - черга зменшується - прилад зайнято - якщо прилад зайнято - декремент інтервалу до кінця обслуговування - інкремент лічильника тактів обслуговування - якщо обслуговування закінчилось - прилад вільний - обнулення лічильника тактів обслуговування - ностальгія - те, що програма повертає під іменем Sp
--	---

Побудова стенду для дослідження СМО

Стенд - екранна сторінка, де сконцентровані і зручно розташовані входи і виходи про грами моделювання - числові значення параметрів, таблиці, графіки, текстові кнопки для "включення/відключення" режимів і об'єктів, графіки процесів, обчислених за розробленою програмою. Зверніть увагу, що ці параметри вводяться в режимі абсолютного присвоєння, тому вони відомі програмі, розташованій вище. Пам'ятайте, що цей документ - об'єктна програма, де модулі-об'єкти посилають і приймають повідомлення за певними правилами.

Задаємо розподіл тривалостей інтервалів між появою клієнтів та розподіл часу обслуговування. Обчислюємо середні та стандартні відхилення за відповідними формулами, - в навчальних цілях, але можна було використати вбудовані функції пакета.

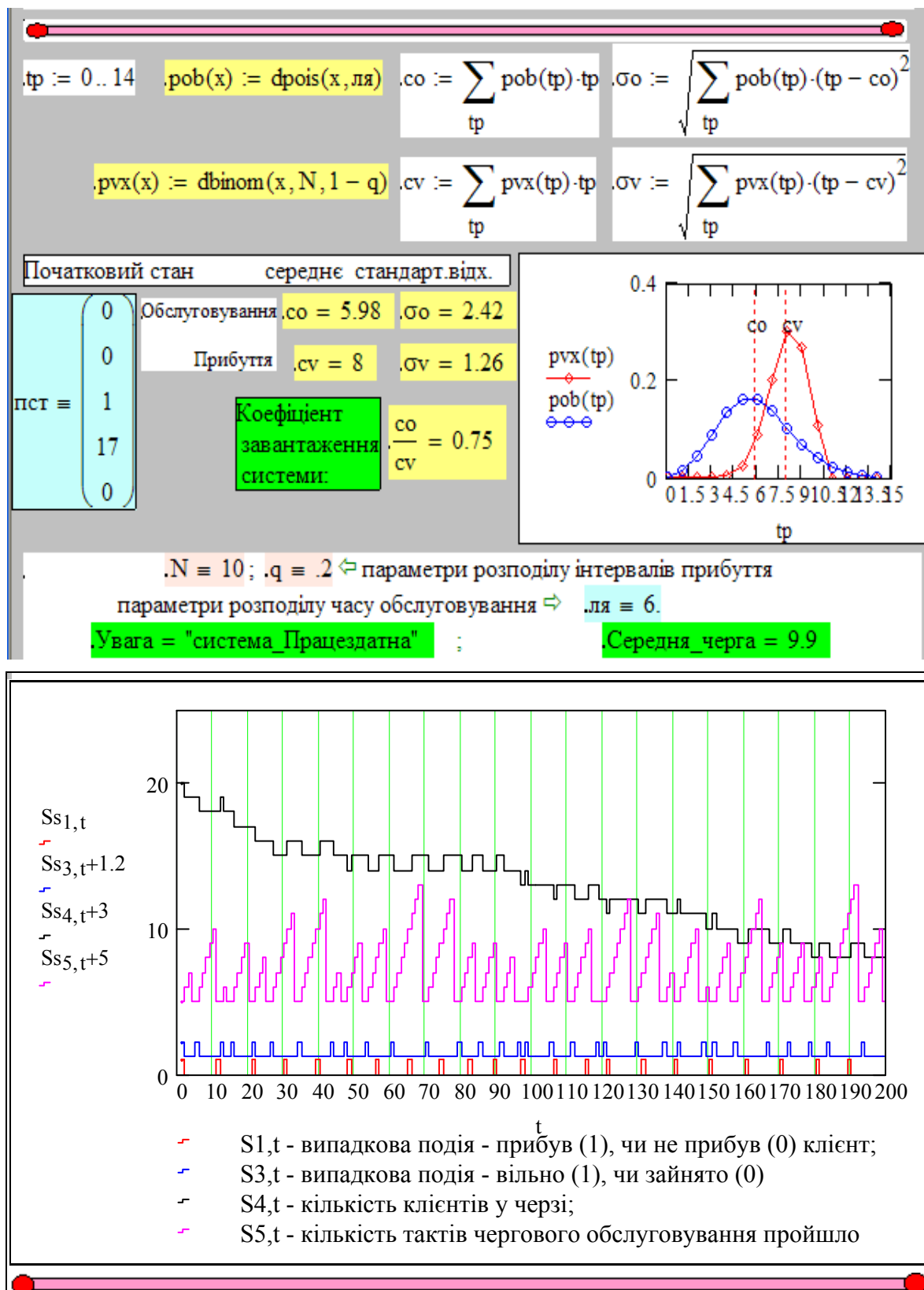


Рис.2 Стенд для дослідження систем масового обслуговування

Подивимось ще раз на "стенд" - екранну сторінку, де зібрані усі "входи" і "виходи" певної задачі або моделі системи. Ви можете вільно все змінювати і одразу ж

бачити наслідки. Це головна відмінність електронної книги від паперової. Сьогодні комп'ютери досягли габаритів книги, мають вбудований Інтернет, в перспективі і ціна їх буде, як у книги... Чи зможемо ми ефективно використати нові можливості. **Інструкція користувачу:** тим, хто занадто довіряє критеріям довіри, корисно поставити курсор миші на зону введення якогось параметра: **N, ля, q** і понатискати F9. При цьому обчислюються **нові реалізації випадкових** процесів прибуття клієнтів та їх обслуговування. Можна побачити, що досить часто виникають "аномальні" реалізації з великими чергами - все це просто *природні властивості випадкових процесів* з певними постійними статистичними характеристиками.

Саме це малося на увазі в класиці "В інший раз буде мені наукою, щоб не довіряв випадковості, а бив самого себе по морді два рази в день, вранці та ввечері, та примовляв: обережність ніколи не буває надмірною, а надмірність - шкодить" (Я. Гашек Пригоди бравого солдата Швейка).

Марківські математичні моделі СМО

Марківські моделі є одним з альтернативних засобів аналізу і синтезу СМО. Згадаємо, з чого ми починали, коли розробляли імітаційну математичну модель СМО. Ми описали стан системи такими змінними:

$S_{1,t}$ - **випадкова** подія: прибув (1), чи не прибув (0) клієнт;

$S_{2,t}$ - **випадкова** подія: обслуговано (0), чи поки не обслуговано (1) клієнта (додаткова змінна);

$S_{3,t}$ - подія: прилад вільний (1), зайнятий (обслуговуванням) (0);

$S_{4,t}$ - кількість клієнтів у черзі;

$S_{5,t}$ - кількість тактів чергового обслуговування пройшло (і ще не скінчилось).

А тепер подивимось на нашу систему під *іншим кутом зору*, виділимо такі стани для нашої системи

$St0$ - нема черги,

$St1$ - черга в одну машину,

$St2$ - черга у дві машини, і..

на першому етапі, щоб не ускладнювати дуже модель, зробимо сюрреалістичне припущення, що третій клієнт, побачивши чергу в дві машини їде до конкурента.

Складаємо матрицю переходів для цих станів

$$\begin{array}{c} \text{Імовірність переходу із стану: } St0 \quad St1 \quad St2 \\ \quad \quad \quad \Downarrow \quad \quad \quad \Downarrow \quad \quad \quad \Downarrow \\ \text{У стан: } \begin{array}{l} St0 \Leftrightarrow \\ St1 \Leftrightarrow \\ St2 \Leftrightarrow \end{array} \left(\begin{array}{ccc} q1 & q1 \cdot p2 & 0 \\ p1 & q1 \cdot q2 + p1 \cdot p2 & q1 \cdot p2 \\ 0 & q2 \cdot p1 & q1 \cdot q2 + p1 \cdot p2 + q2 \cdot p1 \end{array} \right) = M \end{array}$$

Запишіть самостійно аналогічну матрицю для випадку максимальної черги в три машини. Запишемо словесні сценарії для описання елементів матриці переходів.

Приймаємо стандартні означення для імовірностей: імовірність прибуття - p_1 (протягом одного такту, періоду, кроку роботи системи), імовірність неприбуття - $q_1 = 1 - p_1$, імовірність обслуговування - p_2 , імовірність необслуговування - $q_2 = 1 - p_2$.

Зауваження. В першій частині документа були інші означення:

$m_u = p_1$ та $l_a = p_2$.

Напишемо **словесні формули** для елементів матриці переходів M :

m_{11} - (черги не було), не було прибуття - *черга залишилась нульовою*;

m_{12} - (черга в одну машину), прибуття не було, 'і' обслуговування закінчено - черга стала нульовою;

m_{21} - (черги не було), є прибуття - черга стала одиничною;

m_{22} - (черга в одну машину), не було прибуття 'і' не було обслуговування, 'або' було прибуття 'і' було обслуговування - *черга залишилась одиничною*;

m_{23} - (черга в дві машини), прибуття не було 'і' обслуговування закінчено - черга стала одиничною;

m_{32} - (черга в одну машину), є прибуття 'і' обслуговування не закінчено - в черзі стало дві машини;

m_{33} - (черга в дві машини), (не було прибуття 'і' не було обслуговування) 'або' (було прибуття 'і' було обслуговування) 'або' (було прибуття 'і' не було обслуговування, але ж клієнт поїхав далі) - *черга залишилась у 2 машини*. Відповідні елементи матриці переходів нульові, якщо, згідно з умовами задачі, відповідні події неможливі.

Задаємо кількість кроків моделювання

$Kolstep := 20$

$t := 1 .. Kolstep$

$ORIGIN := 1$.

Задаємо початкові умови і значення параметрів вхідного потоку і обслуговування.

Якось не матеріалістично сприймається "імовірність черги в 2 машини". Прийmemo таку (необов'язкову) інтерпретацію вектора X : На комплексі АЗС маємо 10 колонок. На початку періоду по завантаженню вони розподілялись так:

$$X^{(1)} := \begin{pmatrix} 4 \\ 3 \\ 3 \end{pmatrix} \begin{pmatrix} \text{"- бензоколонки з чергою} = 0 \text{"} \\ \text{"- бензоколонки з чергою} = 1 \text{"} \\ \text{"- бензоколонки з чергою} = 2 \text{"} \end{pmatrix}.$$

Записуємо матрицю переходів, виводимо її значення для заданих параметр

$$M_0 := \begin{pmatrix} q_1 & q_1 \cdot p_2 & 0 \\ p_1 & q_1 \cdot q_2 + p_1 \cdot p_2 & q_1 \cdot p_2 \\ 0 & q_2 \cdot p_1 & q_1 \cdot q_2 + p_1 \cdot p_2 + q_2 \cdot p_1 \end{pmatrix}; \quad M_0 = \begin{pmatrix} 0.6 & 0.33 & 0 \\ 0.4 & 0.49 & 0.33 \\ 0 & 0.18 & 0.67 \end{pmatrix};$$

$i := 1 .. 3$;

Контроль матриці переходів:

$$\sum_i M_{0i,1} = 1; \quad \sum_i M_{0i,2} = 1; \quad \sum_i M_{0i,3} = 1.$$

І нарешті записуємо робоче рівняння марківського процесу:

$$X^{(t+1)} := M_0 \cdot X^{(t)}$$

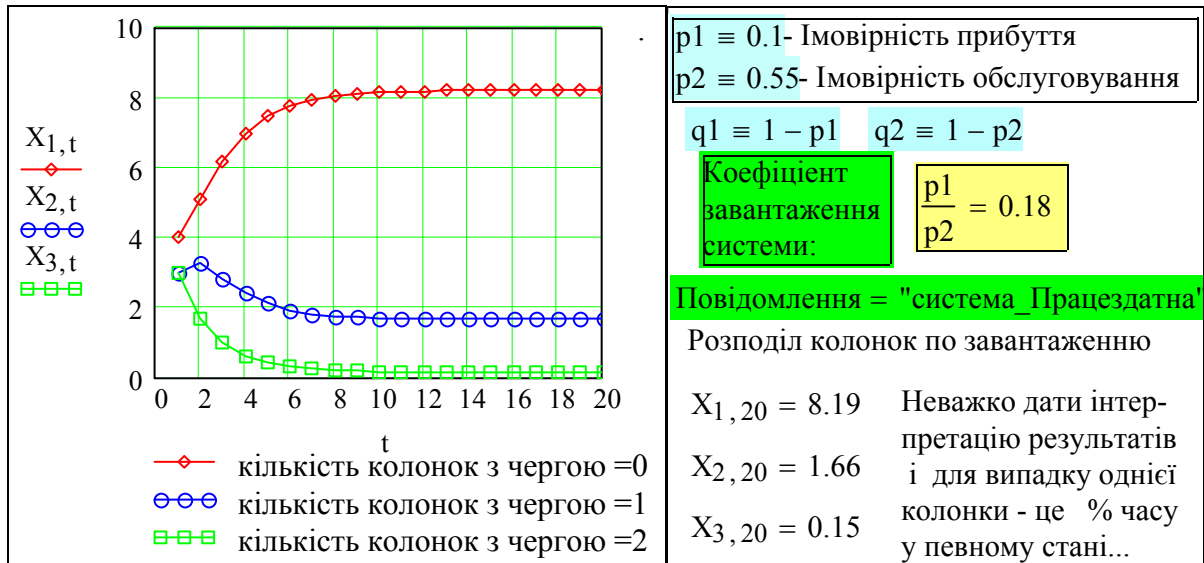


Рис.3 Графіки процесів в системі обслуговування

Проаналізуємо те, що зроблене і отримане. Маємо стенд для аналізу процесів в СМО. Ми можемо змінювати імовірності прибуття та обслуговування і отримувати перехідні процеси. Коли СМО одноканальна, то ми отримуємо процес зміни імовірностей черг довжиною 0, 1, 2 вимоги. Коли каналів досить багато, то результат розрахунку можна трактувати як зміну кількостей приладів обслуговування, що мають черги 0, 1, 2 вимоги. В цілому марківська модель СМО є більш спрощеною порівняно з імітаційною моделлю.

Зупинимось на важливому і очевидному параметрі СМО - **коефіцієнті завантаження**. Для імітаційної моделі ми визначали його як відношення математичних очікувань розподілів часу обслуговування та інтервалу між прибуттям вимог. Для марківської моделі - це відношення відповідних імовірностей (рис.3).

Немовби очевидно, що при завантаженні більше одиниці, черга буде зростати, і навпаки. Але, згідно з властивостями випадкових величин, завантаження менше одиниці не гарантує відсутності черг. Зрозуміло, що при малому завантаженні, імовірність виникнення черги буде малою. Згадаємо, що звичайно в готелі на тисячу місць, ресторан роблять теж на тисячу місць.

Підведемо підсумки. Виконаємо ревізію і аналіз того, що отримали. Ми розглянули дві альтернативні моделі системи масового обслуговування, це

імітаційна модель СМО:

$$S^{(t+1)} = \text{НОП}(S^{(t)})$$

марківська модель СМО:

$$X^{(t+1)} := M_0 \cdot X^{(t)}$$

де S та X - вектори стану, НОП - нелінійний оператор, M_0 - матриця перехідних імовірностей (лінійний оператор). Моделі належать до одного структурного класу, але мають певні відмінності. Подамо далі порівняльний опис цих моделей

Імітаційна модель НЕЛІНІЙНА, універсальна. Реалістичність моделі потенційно не обмежена - можна ураховувати все. Модель не дозволяє отримувати узагальнені аналітичні результати. Може вбудовуватись у реальні автоматизовані системи управління - для прогнозування і планування. Описання вектора стану: $S_{1,t}$ - прибув (1), чи не прибув (0) клієнт; $S_{2,t}$ - обслужений (0), чи не обслужений (1) клієнт (необов'язкова змінна) $S_{3,t}$ - прилад вільний (1), зайнятий (0); $S_{4,t}$ - кількість клієнтів у черзі; $S_{5,t}$ - кількість тактів чергового обслуговування Описання оператора переходу - специфікації операцій в системі в залежності від ситуацій - прибуття, неприбуття, наявність черги, розмір черги, та ін.	Марківська модель ЛІНІЙНА, має обмеження - імовірність переходу з певного стану не залежить від тривалості знаходження в цьому стані Реалістичність моделі суттєво обмежена. Модель дозволяє отримувати узагальнені аналітичні результати (≠ теорія марківських процесів). Може обмежено використатись у реальних автоматизованих системах. Описання вектора стану: $X_{1,t}$ - імовірність перебування в стані "черга = 0"; $X_{2,t}$ - імовірність перебування в стані "черга = 1"; $X_{3,t}$ - імовірність перебування в стані "черга = 2" і т.д. Описання оператора переходу - матриця перехідних імовірностей, де m_{11} - імовірність переходу черга = 0 → черга = 0 на даному інтервалі m_{12} - імовірність переходу черга = 0 → черга = 1 ... і т.д.
--	--

Розгляньте самостійно третю альтернативу - аналіз СМО на базі використання згорток розподілів імовірностей інтервалів прибуття та часу обслуговування.

Контрольні запитання

1. Дайте визначення та приклад одноканальної та багатоканальної системи масового обслуговування.
2. Дайте визначення та приклад однофазної та багатофазної системи масового обслуговування.
3. Сформулюйте умову зменшення черги (в термінах математичних очікувань інтервалу прибуття та часу обслуговування).
4. Сформулюйте умову гарантованого невиникнення черги в термінах параметрів розподілів імовірностей інтервалів потоку вимог та тривалості обслуговування.
5. Запишіть рівняння для марківської моделі. Дайте інтерпретацію а) вектора стану, б) елементів перехідної матриці.
6. Запишіть змінні вектора стану для імітаційної моделі СМО.



2.2 Розрахунок та моделювання багатофазної одноканальної СМО (конвеєрне виробництво)

Вступ. Постановка задачі

Ціль даного розділу - розробка математичної моделі і програми для імітаційного моделювання роботи багатофазної СМО, або конвеєра. В загальному випадку обслуговування заявок (виконання заданих операцій на робочих місцях конвеєра) може бути багатоканальним і багатофазним. Фази обслуговування - це операції, що виконуються в певній послідовності, відповідними приладами, канали обслуговування - паралельно працюючі прилади певної фази обслуговування.

Між іншим, до чого картинка вгорі? Тому, що це класика і продуктивний напрямок. Тому що головна ідея там - безвихідних ситуацій для кого завгодно, тим більше - для людини, не буває. Класична теорія масового обслуговування має 3-5 красивих результатів для 3-5 задач. На практиці зустрічаються переважно задачі, для яких наука може дати тільки один інструмент - імітаційне моделювання.

Конвеєри є окремим випадком СМО. Специфіка конвеєрних виробництв:

- а) вхідний потік "заявок" є детермінованим і рівномірним;
- б) специфічна дисципліна обслуговування - "заявка", що випадає з ритму конвеєра, може бути просто знята, але може бути зупинений весь конвеєр.
- в) "проблеми", що виникають на певній фазі роботи конвеєра, звичайно впливають на інші фази. Одна з головних задач управління - запобігання ситуаціям, коли в певних фазах виникають черги "заявок на обслуговування".

Розробка програмного модуля імітації роботи конвеєра

Приклади поточних виробництв

Класичні приклади поточних виробництв - збирання автомашин, мотоциклів, двигунів, телевізорів, принтерів, виробництво багатьох продтоварів. Відомий приклад поточного виробництва з класики - бройлерна фабрика ім. Луначарського. Ще один приклад - масове виробництво інтелектуальних мікропроцесорних кавоварок (скорочено - ікавок) і чашок в США. (Чашки накопичують статдані, коли і скільки з них пили і передають ікавці, яка варить каву у відповідні моменти, у відповідній кількості). Суперконвеєром було виробництво телевізорів - головний конвеєр починався на складі комплектуючих (1 поверх), де на нього встановлювалось шасі, проходив через три поверхи і закінчувався на складі готової продукції знову на першому поверсі, де на упаковку наклеювався знак якості та ін. Головний конвеєр взаємодіяв з допоміжними - вузлового складання, виробництва друкованих плат, випробування ("випалювання" дефектів).

Тепер на базі цього експрес-аналізу конвеєрних систем сформулюємо задачу управління конвеєром:

Визначення для кожної ділянки ймовірностей виникнення черг, тобто ситуацій "попередній виріб ще не оброблено, а на ділянку надходить наступний виріб".

Вибір потрібної продуктивності ділянки для збалансування з темпом роботи конвеєра (наприклад на певній фазі може працювати 3, 4, або 5 чоловік).

Вибір оптимального темпу роботи конвеєра - компромісу між продуктивністю і імовірністю браку і простоїв.

Можливий ще один спосіб забезпечення нормальної роботи конвеєра - створення на кожній ділянці буферних запасів виробів: коли на черговому виробі потрібну операцію не встигли зробити - на конвеєр ставлять виріб з буферного запасу.

Альтернативні математичні моделі для розподілів інтервалів між виробами

Припускаємо, що відхилення від номінального часу виконання операції на кожному робочому місці конвеєра незалежні. Тому для обчислення розподілу сумарного відхилення після виконання якоїсь операції можна застосувати такі альтернативні ММ:

а) неперервна формула для згортки вхідного розподілу і розподілу "перетворювача"

$$P_{vyx}(n) = \int_{-D}^D X_{vx}(n) \cdot R_p(n - \rho) d\rho; \quad D = (-\infty, \infty);$$

б) дискретна формула для згортки вхідного розподілу і розподілу "перетворювача"

$$P_{vyx_n} = \sum_{\rho} P_{vx_{\rho}} \cdot P_{pr_{n-\rho}}; \quad \rho = (-\infty, \infty);$$

в) обчислення розподілу за даними імітаційного або натурального експерименту, результатом якого є масив випадкових чисел: $M := \{\Delta t_{vx_i} + \Delta t_{p_i}\}; i = 1 \dots N$, де Δt_{vx_i} - випадкове відхилення часу входу i-го виробу; Δt_{p_i} - випадкове відхилення часу виконання операції. Тоді в термінах вбудованої статистичної функції "гістограма" вихідний розподіл буде:

$$P_{vyx}(k) = \text{hist}(in, M) \cdot \frac{1}{N},$$

де in - вектор інтервалів, $k = 1 \dots K$ - індекс інтервалу. Що до вибору "оптимальної моделі", то згідно з сучасними стандартами на інформаційно-обчислювальні системи слід виконувати обчислення і переробку інформації за декількома альтернативними моделями. Це забезпечує контроль і надійність при незначних додаткових витратах на обчислювальні ресурси.

Альтернативні математичні моделі для розподілень часу виконання технологічної операції

Ідеальна, і одночасно - найпростіша модель - детермінований час виконання операції. Щоб досягнути цього ідеалу, розробляють відповідну конструкцію елементів виробу, відповідну технологію, а головне - виключають з цього процесу людину - виконавця операції.

Ще Адам Сміт (Ньютон від економіки) описав, як виключення ручної праці у виробництві булавок дозволила підвищити продуктивність праці у тисячу разів. На жаль, сьогодні не усі виробництва можуть бути перетворені у безлюдні та детерміновані.

Відхилення часу виконання операцій можуть бути результатом десятків причин - зміни температури, настрою персоналу, дня тижня, минулих чи наступних свят та ін.

Відхилення можуть мати як детермінований, так і систематичний характер. Розподілення цих відхилень обумовлені діючими в технологічній операції "механізмами".

Підсумок. Якщо ви не можете зробити операцію детермінованою - займіться статистичним аналізом: вивчіть специфіку техпроцесу, визначте "механізм" породження відхилень для використання його безпосередньо в імітаційних методах, або для вибору відповідного частотного розподілення при використанні методів теорії ймовірності.

Сучасні пакети мають бібліотеки типових розподілень ймовірностей та генераторів випадкових чисел із заданими законами розподілення. Наводимо вирази та стандартні функції для деяких з них: нормального, хі-квадрат, біноміального, пуасонівського, експоненційного, геометричного (логнормальне розглянуто в підрозділі 1.3):

Подаємо для кожного типового розподілення формулу щільності розподілення, відповідну вбудовану функцію, та функцію, що генерує n випадкових чисел із даним розподіленням.

Завдання. Визначте, де саме який розподіл, підпишіть.

$\frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$	$\frac{1}{2^{\frac{d}{2}} \cdot \Gamma\left(\frac{d}{2}\right)} \cdot x^{\frac{d}{2}-1} \cdot e^{-\frac{x}{2}}$	$\frac{N!}{k! \cdot (N-k)!} \cdot p^k \cdot (1-p)^{N-k}$
<code>dnorm(x, μ, σ)</code>	<code>dchisq(x, d)</code>	<code>dbinom(k, N, p)</code>
<code>rnorm(n, μ, σ)</code>	<code>rchisq(n, d)</code>	<code>rbinom(n, N, p)</code>
$r \cdot \exp(-r \cdot x)$	$p \cdot (1-p)^k$	$\frac{e^{-\lambda} \cdot \lambda^k}{k!}$
<code>dexp(x, r)</code>	<code>dgeom(k, p)</code>	<code>dpois(k, λ)</code>
<code>rexp(n, r)</code>	<code>rgeom(n, p)</code>	<code>rpois(n, λ)</code>

Аналіз альтернатив реалізації програми моделювання і вибір програмного середовища

Сучасні технології розробки прикладних програм, незважаючи на зовнішню різноманітність побудовані за єдиною схемою як послідовність етапів:

1) формулювання **списку функцій системи** (специфікація) => 2) ручна або автоматична трансляція специфікації у **діючий макет програми** => 3) повторення п.2 - **розробка альтернативних версій** => відлагодження, випробування та взаємоперевірка версій => 4) ручна або автоматична **трансляція макета програми у код** (.EXE-модуль).

Нашу програму, для випадку інтенсивного використання на робочому місці, можна реалізувати однаково ефективно в Excel, Access, Visual FoxPro, Delphi, в математичних пакетах та ін. До цих продуктів випускаються бібліотеки готових модулів і прототипів, які досить тільки параметризувати і можна включити в свою програму. Для відлагодження та проведення аналітичних досліджень з цілями стратегічного прогнозування краще розробити програму (документ) у середовищі математичного пакета, наприклад Matlab або Mathcad.

Ці пакети дають безмежні можливості досліднику і одночасно мають можливості для створення .EXE-модулів. Сьогодні тільки малі критичні фрагменти програми пишуться "вручну". В ситуації, коли треба швидко отримати конкретні результати дослідження (хіба буває інакше?), все частіше вибирається середовище математичного пакету.

Останній приклад - корпорація Matworks, що випускає пакет Matlab, зробила програмне забезпечення для безпосереднього підключення пакета через стандартні пристрої комп'ютера до будь-яких джерел даних - аналогових, цифрових. Програма забезпечує перетворення та первинну обробку даних, що потім можуть оброблятися потужними засобами пакета Matlab. Таким чином, створений в середовищі математичного пакету макет програми управління якимсь процесом можна перетворити в автоматизовану систему управління реальним, а не модельним процесом. Етап створення програми в кодах може взагалі відпасти.

Оцінка економічної ефективності конвеєра

Побудова моделі і моделювання завжди базуються на постановці певних цілей. Природно поставити ціль економічної ефективності багатофазної СМО (в даному прикладі - конвеєра).

Для аналізу ефективності конвеєра потрібно проаналізувати, врахувати, кількісно оцінити велику кількість нетривіальних факторів, таких як швидкість зміни конструкції виробу, коливання попиту на ринку, швидкість зміни технологій та ін. Оцінимо приріст витрат та прибутків, пов'язаний з використанням розроблюваної автоматизованої системи. Ефект від використання програми проявляється в підвищенні якості рішень користувачами за рахунок точних прогнозів.

Звичайно якість роботи конвеєра знижується при виникненні черги. Для деяких виробництв черга неможлива і конвеєр просто зупиняють, або знімають з конвеєра виріб з незакінченою операцією. Враховуючи це, узагальнену модель ефективності конвеєра подамо так:

$$E_{\text{sum}} = C_i \cdot \left[\frac{T_c \cdot (1 - P_p(T_k))}{T_k} \cdot (1 - P_b(T_k)) \right]$$

де C_i - вартість одного виробу; T_c - номінальна тривалість роботи конвеєра за добу; T_k - темп роботи конвеєра, параметр, величина якого розраховується та оптимізується; $P_b(T_k)$ - імовірність браку; $P_p(T_k)$ - імовірність зупинки конвеєра, як функція темпу роботи конвеєра. Якісно ці залежності описуються так: при підвищенні темпу, тобто зменшенні T_k , підвищується імовірність браку та імовірність зупинки конвеєра при виникненні черги на вході в будь-яку ділянку.

Проблеми безпеки життєдіяльності та забезпечення сприятливих умов роботи на конвеєрі

Для більшості матеріальних виробництв людина є небажаним елементом конвеєра, а конвеєр - небажаним місцем роботи людини. На певних ділянках конвеєрів по збиранню автомашин, двигунів, де працюють зварювальні, монтажні роботи, людина взагалі з'являється тільки при зупиненому конвеєрі для наладки. Там, де одночасно працюють люди і роботи та верстати-автомати, останні обгороджуються сіткою червоного кольору. Факторами небезпеки є електроінструменти, рухомі елементи конвеєра та ін.

Радикальний шлях забезпечення безпеки - вилучення людини з виробничого процесу, якщо це неможливо, - ізоляція людей від виробничого процесу в спецприміщеннях, звідки вони виконують дистанційні контроль і управління. Якщо не можливе і це - слід винести виробничі операції, де потрібні люди, в окремі ділянки.

Розробка блок-схеми і алгоритму програми

Опишемо задачі програми: для заданого варіанта вхідних даних необхідно обчислити частотні розподіли для сумарних відхилень часу обробки виробів після кожної ділянки і на виході конвеєра.

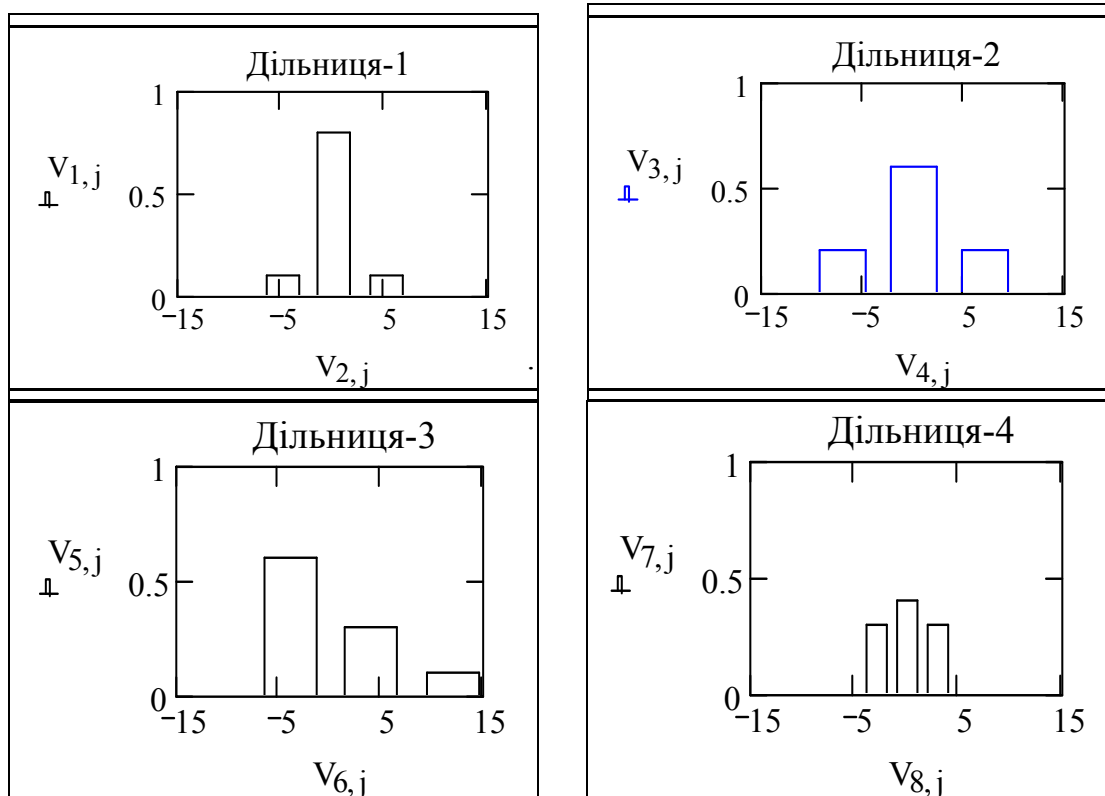
На підставі цих даних обґрунтовано вибираємо оптимальний темп роботи конвеєра (пропускну спроможність). Для підвищення надійності результатів виконуємо розрахунки за **альтернативними** моделями - згортки розподілів та імітаційного моделювання.

$V :=$	$\begin{pmatrix} .1 & .8 & .1 \\ -5 & 0 & 5 \\ .2 & .6 & .2 \\ -7 & 0 & 7 \\ .6 & .3 & .1 \\ -4 & 4 & 12 \\ .3 & .4 & .3 \\ -3 & 0 & 3 \end{pmatrix}$	$M := 200$	$q := 1..4$	$k := 1..M$	$ORIGIN := 1$
	V - це матриця групованих статистичних даних про відхилення часу виконання операції по кожній ділянці. Кожний непарний рядок матриці задає імовірності певних відхилень часу виконання. Кожен парний рядок - величини цих відхилень.				

Контроль даних. Статистичні дані, особливо отримані вручну, слід перевірити на відсутність грубих помилок. Будуємо графіки усіх розподілів. Згідно з умовами задачі **усі розподіли повинні бути центрованими** (мати середнє = 0). Це означає, що при вибраному темпі роботи конвеєра середня продуктивність (пропускна спроможність) усіх ділянок однакова. **Завдання. Обчисліть середні:**

$j := 1..cols(V)$; $cols(V) = 3$; $q := 1..0.5 \cdot rows(V)$; $rows(V) = 8$.

Дивимось на графіки емпіричних розподілів (гістограм) і таблицю (матрицю) даних V та і бачимо структуру даних - непарні рядки матриці - це імовірності, парні - значення інтервалів.



Перевіряємо: нормованість розподілів:

центрованість розподілів:

$$\text{Сума_ймовірностей}_q := \sum_j V_{2q-1,j}$$

$$\sum_j V_{2 \cdot q-1,j} \cdot V_{2 \cdot q,j}$$

0
0
0
0

$$\text{Сума_ймовірностей}^T =$$

	1	2	3	4
1	1	1	1	1

Розробка модуля імітаційного моделювання роботи конвеєра

В сучасних технологіях розробки програм етапи „математична модель-блок-схема-алгоритм-програма” ІНТЕГРУЮТЬСЯ. Математична модель звичайно є виконуваною, тобто формула, рівняння є одночасно і програмою. Блок-схема також часто буває виконуваною програмою - в системах моделювання в реальному часі..

Перше, з чого почнемо - повторимо визначення індексних змінних:

k - індекс (номер) виробу; **q** - номер дільниці; **j** - номер інтервалу частотного розподілу (згідно з умовами задачі усі розподіли для часу операцій складання задано на трьох інтервалах). "Витягуємо" з матриці вхідних даних окремо:

матрицю імовірностей **p** $p_{q,j} := V_{2q-1,j}$

і матрицю значень інтервалів **z** $z_{q,j} := V_{2 \cdot q,j}$.

Формуємо масив **g** із $M \times 4$ (розмір партії виробів*кількість дільниць) випадкових чисел і записуємо вираз, який **вибирає одне із значень інтервалу відповідного дискретного частотного розподілу з відповідною імовірністю**. Те, що подано нижче, власне і є центральний модуль програми імітації роботи конвеєра.

$$g_{k,q} := \text{rnd}(\text{натисни}) \cdot \text{жми}$$

$$dT_{q,k} := \text{if}(g_{k,q} < p_{q,1}, z_{q,1}, \text{if}(g_{k,q} < p_{q,2}, z_{q,2}, z_{q,3})) ;$$

Виводимо значення сформованих даних - для контролю, аналізу і порівняння. **Масив dT інтерпретується так:** стовпець - це виріб з партії, рядок - зсув часу обробки (випередження з мінусом, відставання з плюсом) обробки на кожній дільниці. Наприклад, у деякому стовпці №33 маємо 5, 5, -4, -3. Означає це випередження на 5с на 1-ій дільниці, на 5с - на 2-ій, запізнення на 4с - на 3-ій і знову ж запізнення на 3с на 4-ій дільниці.

$$g =$$

	1	2	3	4
1	0.001	0.193	0.585	0.35
2	0.823	0.174	0.71	0.304
3	0.091	0.147	0.989	0.119
4	0.009	0.532	0.602	0.166
5	0.451	0.057	0.783	0.52
6	0.876	0.956	0.539	0.462

$$dT =$$

	1	2	3	4	5	6	7	8	9	10
1	-5	5	-5	-5	0	5	5	0	-5	0
2	-7	-7	-7	0	-7	7	7	7	0	7
3	-4	12	12	12	12	-4	12	-4	-4	-4
4	0	0	-3	-3	3	3	3	3	3	3

$$Z = \begin{bmatrix} & 1 & 2 & 3 \\ 1 & -5 & 0 & 5 \\ 2 & -7 & 0 & 7 \\ 3 & -4 & 4 & 12 \\ 4 & -3 & 0 & 3 \end{bmatrix}$$

$\Leftarrow$ Матриці z, v
 розпаковані з
 матриці $V \Rightarrow$.
 Знайдіть в
 матриці V рядки і
 стовпці матриць
 z, v .

$$V = \begin{bmatrix} & 1 & 2 & 3 \\ 1 & 0.1 & 0.8 & 0.1 \\ 2 & -5 & 0 & 5 \\ 3 & 0.2 & 0.6 & 0.2 \\ 4 & -7 & 0 & 7 \\ 5 & 0.6 & 0.3 & 0.1 \\ 6 & -4 & 4 & 12 \\ 7 & 0.3 & 0.4 & 0.3 \\ 8 & -3 & 0 & 3 \end{bmatrix}$$

Завдання. Вище подана **найпростіша і обмежена за можливостями альтернатива програми імітаційного моделювання роботи конвеєра**. Можна, наприклад, написати її в мові програмування Mathcad у вигляді процедури, одним з параметрів якої буде вектор із вбудованих функцій частотних розподілів. Така програма буде більш універсальною. Зробіть це.

Модуль обчислення часових інтервалів між виробами на конвеєрі

Транспонуємо масив імітованих статистичних даних $dTT := dT^T$. Записуємо вирази для обчислення відхилень часу проходження k-го виробу через відповідні дільниці:

Ця система рівнянь є другим головним модулем програмної системи для моделювання роботи конвеєра.

Отак:

$$M4^T = \begin{array}{c|cccccccccccccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ \hline 1 & -16 & 10 & -3 & 4 & 8 & 11 & 27 & 6 & -6 & 6 & 0 & 9 & 0 & 9 & 7 & 15 & 6 & 4 \end{array}$$
$$\mathbf{M}_x^T = \begin{array}{c|cccccccccccccccccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 \\ \hline 1 & 0 & 10 & 26 & 20 & 25 & 32 & 51 & 46 & 13 & 13 & 25 & 28 & 28 & 28 & 35 & 35 & 34 & 23 & 39 & 33 & 39 \end{array}$$

Напишемо програму для обчислення кількості ситуацій, коли на 4-ту ділянку "наїжджає" наступний виріб, коли ще не закінчено роботу з попереднім. Подаємо блок з математичних і текстових об'єктів – копію екрану.

```

(kzb4)
(vrzb4) := | Nn ← length(Mx)
              kzb ← 0
              vrzb ← 0
              for i ∈ 1..Nn
                if Mxi ≤ 0
                  | kzb ← kzb + 1
                  | vrzb ← vrzb - Mxi
              ( kzb )
              ( vrzb )

```

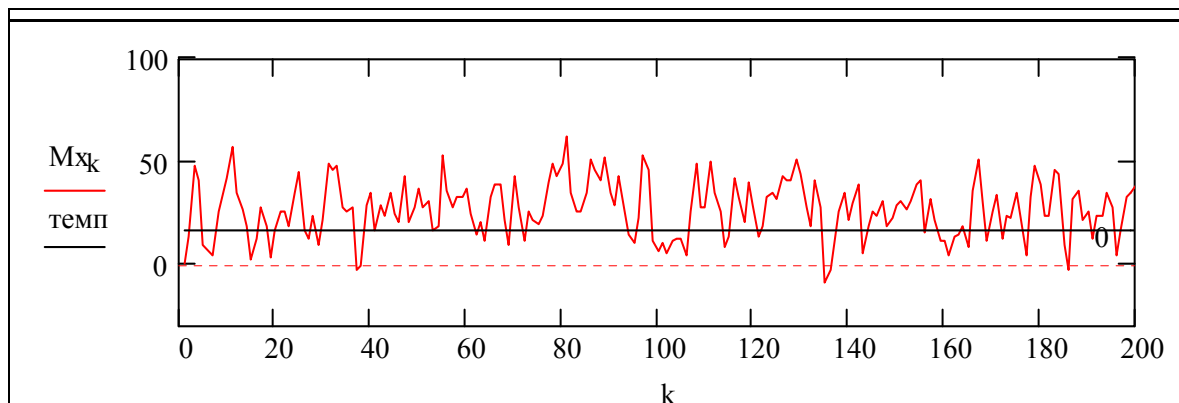
Що бере і що повертає ця програма - прокоментуйте вхідні змінні та видачу.

length(Mx) = 200 kzb4 = 6 vrzb4 = 19

проц_зб := $\frac{kzb4}{length(Mx)}$

простоїв := $\frac{temp \cdot vrzb4 \cdot sec}{length(Mx)}$ +

Ця змінна **натисни ≡ 1** ініціює переобчислення процесу імітації роботи конвеєра - поставте на неї курсор миші, понатискайте F9, подивіться, як змінюється графік.



На цьому графіку подано зміну в часі інтервалу між виробами на виході конвеєра. При **темпі роботи** конвеєра: **temp = 16** (вводиться далі) маємо такі показники:

збоїв: **проц_зб = 3%**, **простоїв:** **простоїв = 1.52 sec**.

У програмі обчислення інтервалів прийнято такі позначення: kzb4 - кількість збоїв, тобто ситуацій, коли часовий інтервал між виробами менший нуля; vrzb4 - час простоїв у секундах (сума усіх від'ємних інтервалів, коли виріб повинен чекати звільнення ділянки).

Контроль. Знайдіть місце, де обчислюються змінні "проц_зб" та "простоїв".

Модуль побудови гістограм для інтервалів між виробами

Обчислюємо середні значення і стандартні відхилення для потоку виробів після кожної ділянки. Будуємо гістограми (=емпіричні частотні розподіли) для інтервалів між виробами для усіх ділянок конвеєра. Для цього використовуємо модуль з розділу 1.3 (копіюємо цей модуль і налаштуємо). В поданих далі виразах використано також вбудовані функції пакету min(.), max(.), hist(., .) - мінімум і максимум на елементах масиву і гістограма для елементів масиву при заданому векторі інтервалів in. Маємо масив даних обсягом M = 200 тактів роботи конвеєра.

Задаємо: кількість інтервалів Kint := 8 і ранжовані змінні для інтервалів: m := 1..(Kint-1) та їх границь: n := 1..Kin.

Формуємо вектори границь інтервалів і обчислюємо гістограми:

$$\begin{aligned} in1_n &:= \min(M1) + \frac{(n-1) \cdot (\max(M1) - \min(M1))}{K_{int} \cdot 8}; & R1 &:= \text{hist}(in1, M1) \cdot \frac{1}{M}; \\ in2_n &:= \min(M2) + \frac{(n-1) \cdot (\max(M2) - \min(M2))}{K_{int} \cdot 8}; & R2 &:= \text{hist}(in2, M2) \cdot \frac{1}{M}; \\ in3_n &:= \min(M3) + \frac{(n-1) \cdot (\max(M3) - \min(M3))}{K_{int} \cdot 8}; & R3 &:= \text{hist}(in3, M3) \cdot \frac{1}{M}; \\ in4_n &:= \min(M4) + \frac{(n-1) \cdot (\max(M4) - \min(M4))}{K_{int} \cdot 8}; & R4 &:= \text{hist}(in4, M4) \cdot \frac{1}{M}. \end{aligned}$$

Контроль даних. Усі дані повинні попадати в діапазон інтервалів, тому суми повинні бути одиничними

$$\sum_m R1_m = 1; \quad \sum_m R2_m = 1; \quad \sum_m R3_m = 1; \quad \sum_m R4_m = 1,$$

однак, діапазон не повинен бути занадто великим - гістограма буде неінформативною.

Подані вище рівняння для інтервалів та гістограм є третім головним модулем нашої програми моделювання роботи конвеєра.

Відлагодження програми, отримання результатів тестування

Виведемо результати розрахунку параметрів техпроцесу для заданих вхідних даних. Будуємо графіки гістограм, наносимо лінії допуску на відхилення часу обробки (це темп роботи конвеєра - змінна Темп), аналізуємо.

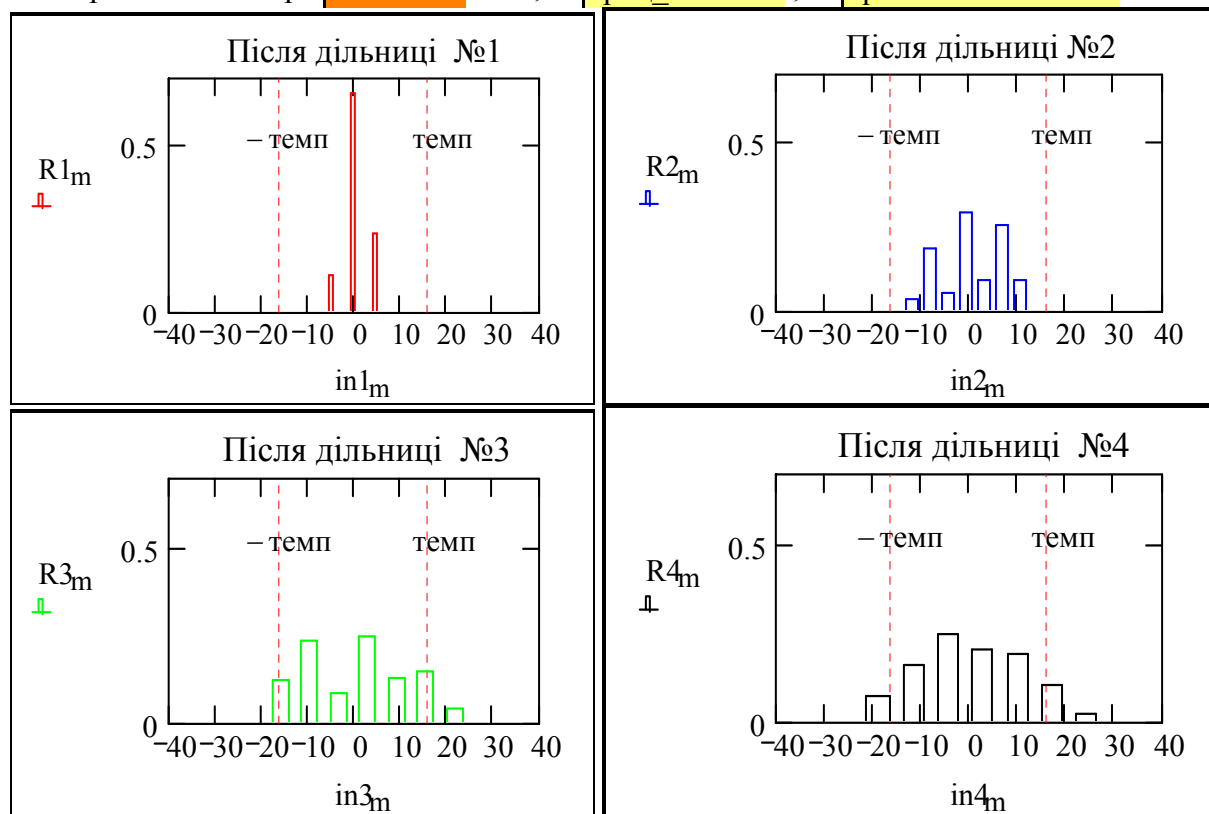
Те, що подано на наступній сторінці – макет стенду, що може бути легко доведений до рівня програмного модуля автоматизованої системи для диспетчера конвеєра - гістограми можуть бути прогнозними, або побудованими за натурними даними.

Диспетчер чи оператор в реальній системі бачить оперативну інформацію - процеси в часі і просторі, інтегральні показники роботи конвеєра. Один з важливих напрямків покращення автоматизованої системи – формування прогнозів і наслідків поточних управлінь. В такому напрямку кінцевим станом вдосконалення системи буде безлюдне виробництво і програма-робот для управління процесом

Диспетчер дивлячись на такі графіки може узагальнено оцінити стабільність і точність роботи ділянок, на базі цих даних він може зменшити або збільшити темп роботи конвеєра.

Завдання. Якщо ви працюєте з електронною книгою, *понатискайте (курсор на "жми", та натискайте F9) змінну "жми" - подивіться як змінюються показники роботи конвеєра, гістограми, статистичні показники, змінійте темп роботи конвеєра.*

Темп роботи конвеєра **темп $\equiv 16$ сек ;** **проц_зб = 6% ;** **простоїв = 4.48 sec**



Виводимо також значення середніх та стандартних відхилень

$\text{mean}(M1) = 0.63$; $\text{mean}(M2) = 1.64$; $\text{mean}(M3) = 3.8$; $\text{mean}(M4) = 4.25$;

$\text{stdev}(M1) = 2.87$; $\text{stdev}(M2) = 6.2$; $\text{stdev}(M3) = 10.39$; $\text{stdev}(M4) = 10.73$.

Висновки-1. Дивлячись на графіки розподілів (в одному масштабі), бачимо:

а) що розподіл після кожної ділянки наближується до нормального і "розпливається" згідно з відповідними положеннями теорії імовірності;

б) що при вибраному темпі роботи (25-30 с) конвеєр працює на грані черги - діапазон відхилення часу досягає ± 25 с. Дійсно, якщо попередній виріб запізнюється на 30с, а наступний випереджає час на 30 с, то виникне черга.

Висновки-2. Отримали програмну систему, що може обробляти як імітовані, так і натуральні дані, і діагностувати і прогнозувати за зміною статистичних характеристик тренди параметрів техпроцесів типу - деградація інструмента, втрата стабільності (наприклад люфти в обладнанні).

Моделювання роботи конвеєра на базі використання згортки розподілів

Розглянемо коротко альтернативний метод обчислення розподілів інтервалів і відповідно альтернативну методологію моделювання. Розглянемо стисло, що стоїть за звичайно пустими термінами "методологія моделювання". В попередній моделі ми оперували з **часовими інтервалами між виробами**, що обробляються на конвеєрі.

В даній моделі об'єкти операцій - **частотні розподіли цих інтервалів**. Якщо випадковий процес із заданим розподілом деякого параметра $P_{\text{ар}}$ проходить через якийсь "перетворювач", де розподіл цього параметра змінюється згідно з розподілом перет-

ворювача.

Вихідний розподіл обчислюється за інтегралом згортки:

для неперервних розподілів

i

для дискретних розподілів

$$P_{vyx}(x) = \int_{-D}^D X_{vx}(\rho) \cdot R_p(x - \rho) d\rho ; D1 = (-\infty .., \infty) ; P_{vyx_n} = \sum_{\rho} P_{vx_{\rho}} \cdot P_{pr_{n-\rho}} .$$

Задаємо діапазони і відповідні ранжовані змінні:

$D := 20$; $ORIGIN := -D$; $x := -D..D$; $\rho := -D..D$; $k := -D..D$; $xx_k := k \cdot 1$.

Задаємо вхідний розподіл $X_{vx}()$ та розподіл (відхилень) часу обробки $R_p(x,d)$ на дільницях конвеєра. Задаємо їх з ряду причин неперервними - це природно, і менше програмних ускладнень.

$X_{vx}(x) :=$	$\begin{cases} 0 & \text{if } x > 2 \\ 0.25 & \text{if } x \leq 2 \end{cases}$	$R_p(x,d) :=$	$\begin{aligned} &hh \leftarrow \frac{1}{d} \\ &kk \leftarrow \frac{hh}{d} \\ &0 & \text{if } x > d \\ &kk \cdot x + hh & \text{if } (x \geq -d) \wedge (x \leq 0) \\ &hh - kk \cdot x & \text{if } (x \leq d) \wedge (x > 0) \end{aligned}$
----------------	--	---------------	--

Подаємо **копію екрану** з робочого документу – блок обчислення згорток

Контролюємо їх:

обчислюємо розподілення після дільниці №1:

$$\int_{-D}^D X_{vx}(x) dx = 1 \quad \int_{-D}^D R_p(x, 5) dx = 1$$

$$P_{vyx}(x) := \int_{-D}^D X_{vx}(\rho) \cdot R_p(x - \rho, d1) d\rho$$

"Редкая птица долетит до середины Днепра", **як відомо**, хоч і неширокий він.

Отак і тут, щоб обчислити (взагалі і в розумний час) розподілення після дільниці №2 - $P_{v2}(x)$, необхідно виконати ряд операцій. (зрозумілих лише тим, хто "бачить", як працюють і взаємодіють "сховані" в програмній платформі програмні модулі, і побачить, що fP_k це не те, що $P_{vyx}(k)$).

$Pf_k := P_{vyx}(k) \quad \text{WRITEPRN}("b19kon") := Pf \quad fP := \text{READPRN}("b19kon.prn")$

$Pv1(x) := \text{interp}(xx, fP, x)$

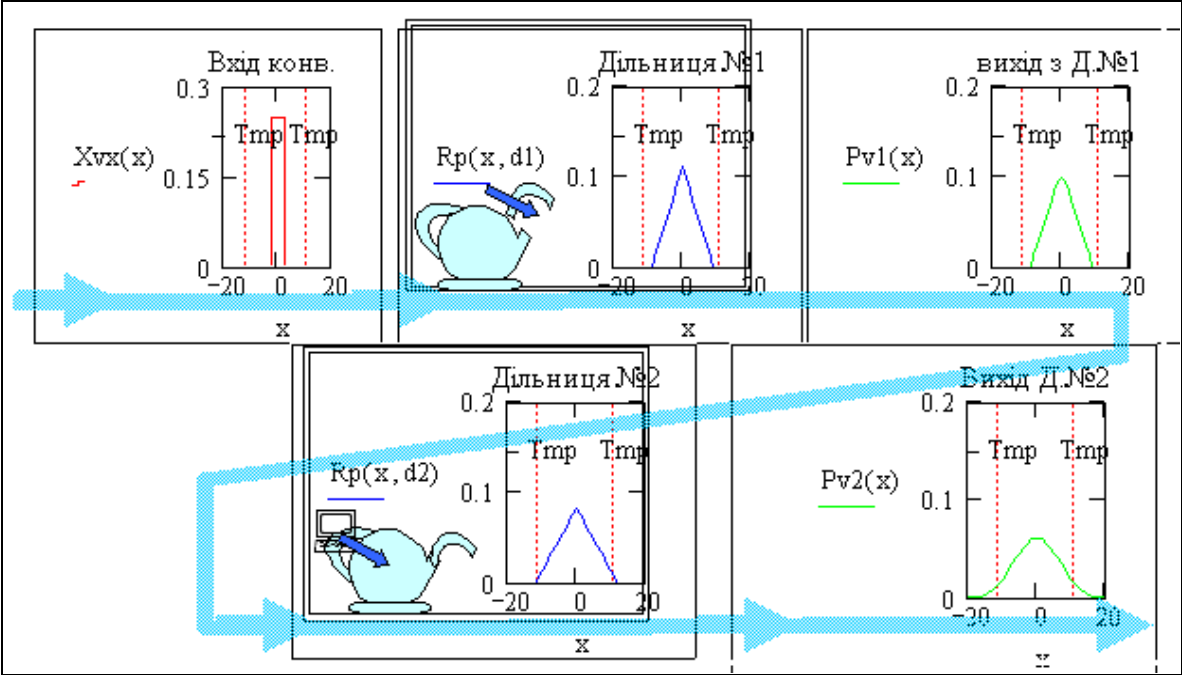
$$Pv2(x) := \int_{-D}^D Pv1(r) \cdot R_p(x - r, d2) dr$$

Увага! Після переобчислень не забувайте натиснути файлову операцію "читати"

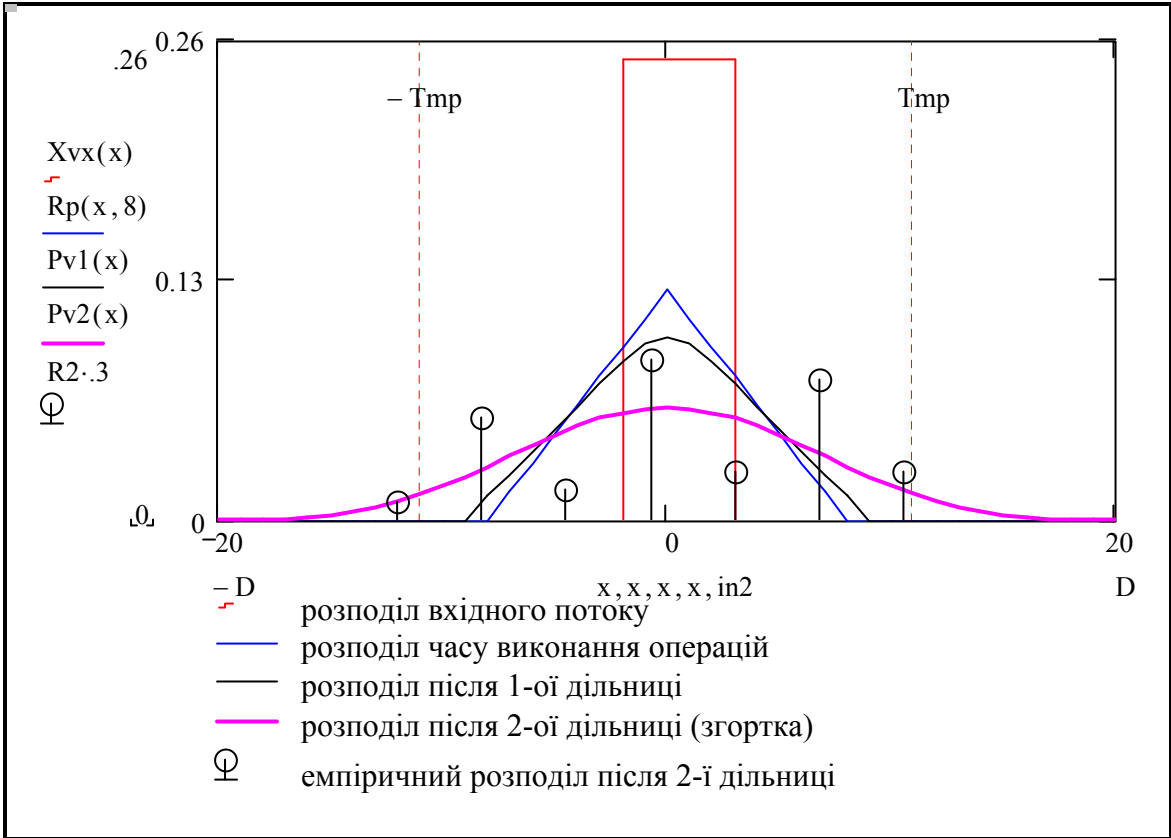
Задаємо темп роботи конвеєра $Tmp := 11$, параметри розподілів для дільниці №1 $d1 \equiv 9$ та №2 $d2 \equiv 12$.

На наступній сторінці подано схему складену з пасивних об'єктів – зображень деталей і активних – графіків частотних розподілень. Не проблема зробити досконалу

схему – імітацію роботи конвеєра, проблема – наявність ліцензованої копії відповідної програмної платформи.



На схемі вище подано приклад оформлення результатів моделювання за методом згортки. Щоб не ускладнювати документ і залишити щось для самостійної роботи, подаємо розподіл тільки для двох дільниць. Для прикладу розглядаємо процес збирання виробу - "ікавка", тобто інтелектуальної кавоварки з мікропроцесорним управлінням.



Перспективні напрямки робіт з підвищення надійності конвеєра

Моделювання, статистика - інструменти аналізу того, що склалося. А про надійність сказано "надійність не вправа з статистики, а пошук конкретних конструкторських рішень та заходів". Розглянемо конвеєрні системи в таких ракурсах.

1. Стратегічний напрямок.

1.1. **Зміна** конструкції виробу в напрямку зменшення кількості складових - блоків, модулів.

1.2. **Спрощення** інтерфейсів - замість гвинтів та гайок - заціпки, кнопки, клей та ін.

1.3. **Ізоляція** критичних технологічних операцій - винесення їх в спецбокси зі стабілізацією усіх умов - температури, вібрацій, газового середовища, запилення.

1.4. **Відмова від** конструкцій і технологій, що є шкідливими, або потребують великих інтелектуальних зусиль для організації настроювання, контролю і управління (адаптивні, з навчанням, генетичними алгоритмами, системами штучного інтелекту та ін.).

2. Тактичні заходи.

2.1. Можливість зупинки конвеєра з кожного робочого місця.

2.2. Можливість просто знімати з конвеєра аномальні вироби.

2.3. Можливість постановки на конвеєр виробів з резерву, що є на кожній ділянці.

Контрольні запитання

1. Дайте визначення та приклад багатофазної системи масового обслуговування.

2. Як залежить середнє значення часу перебування об'єкта в багатофазній системі і стандартного відхилення цього часу від математичних очікувань та стандартних відхилень для часу перебування в кожній окремій фазі.

3. Якщо фази обслуговування в одноканальній СМО можна міняти місцями, то куди краще поставити операцію з найбільшим розкидом часу виконання - першою, чи останньою.

4. Для чого при обчисленні згортки введені операції запису і читання файлу?

Для виконується інтерполяція зчитаних з файлу даних?

5. Чи зберігає операція згортки розподілів нормованість вихідного розподілу?

6. Як експериментально перевірити нормованість згортки розподілів?

7. Яким буде вихідний розподіл після багаторазової згортки прямокутних розподілів з нульовим середнім?



Задачі управління запасами

У цьому розділі розглядаються:

- ✓ Теорія управління запасами. Базові математичні моделі управління запасами
- ✓ Задача управління запасами при розподілених поставках.
- ✓ Задача управління запасами зі штрафами
- ✓ Розробка узагальнених математичних моделей управління запасами

Після вивчення цього розділу ви зможете:

1. Формалізувати реальні задачі управління запасами і відносити їх до одного із класів задач.
2. Розв'язувати задачі управління запасами з „моментальними поставками”.
3. Розв'язувати задачі управління запасами з розподіленими (виробничими) поставками.
4. Розв'язувати задачі управління запасами з штрафами.
5. Розв'язувати задачі **оптимального управління запасами при невідомому попиті**.
6. Розробляти програми моделювання систем управління запасами.
7. Розробляти інтерфейси для програм класу „аналітичний помічник менеджера”.
8. Виконувати моделювання і дослідження систем управління запасами

9. Дати визначення термінам:

- | | |
|----------------------------------|-----------------------------|
| 1. інтенсивність попиту; | 6. розмір партії; |
| 2. організаційні витрати; | 7. критерій оптимізації; |
| 3. вартість товару; | 8. змінна управління; |
| 4. витрати на утримання запасів; | 9. функція штрафу; |
| 5. функція Беллмана; | 10. функціональне рівняння. |



3.1. Управління запасами. Базові математичні моделі управління запасами

Вступ. Постановка задачі

Управління запасами - царина знань однаково важлива і для теорії і для практики. Ми розглянемо спочатку базові, класичні моделі управління запасами. Цим моделям 50-90 років, вони занадто прості, але, якщо навіть трохи наблизити їх до реальності - то вони стають такими, що знайти їх розв'язання під силу Р. Беллману чи Р. Акофу. Приклад такої задачі розглянемо в підрозділі 3.3.

Як наскрізний приклад подамо загально зрозумілу і очевидну задачу (Р. Акоф): кіоскер замовляє на кожен день газети. Попит - випадковий. В кінці дня газети можуть бути в надлишку, або в дефіциті. Здоровий глузд підказує, що і те, і те - недобре: або втрачаємо клієнтів, або витрачаємо гроші на макулатуру. Приблизно те ж і з хлібом, овочами і мобільними телефонами теж (хліб старіє фізично, мобільник - морально).

А що робити, коли ціни і попит на товари мають тренди, сезонні коливання, або випадкові зльоти і обвали? Ми побачимо, як важко знайти аналітичні розв'язання оптимізаційних задач такого типу.

В цьому розділі ставиться задача розробки документа класу "що буде, якщо" для класичних, базових задач управління запасами. На основі цих задач будуються більш реалістичні і більш складні діалогові програмні системи - помічники менеджера у прогнозуванні і плануванні постачання та запасів.

Основа таких систем - набір базових робочих математичних моделей для типових ситуацій управління запасами. Базовою вважаємо математичну модель, що є джерелом моделей для певного класу задач. Робоча модель - це модель, що записана згідно зі стандартами математики і виконується в середовищі якогось математичного, статистичного та ін. пакета.

Природно, робоча модель повинна бути

- а) математично коректною,
- б) задовольняти додаткові вимоги конкретного пакета до запису моделі.

Базовою будемо вважати задачу забезпечення детермінованого постійного попиту при заданих цінах товару, поставок і витратах на зберігання певних товарів чи ресурсів.

В базовій моделі **витрати на поставки вважаємо лінійними функціями** від кількості партій, витрати на зберігання вважаємо лінійними функціями від "товароднів" зберігання. Цільову функцію - сумарні витрати на забезпечення попиту теж вважаємо лінійною функцією усіх компонентів витрат - просто сумою.

На основі базової моделі потім будуть будуватись більш реалістичні моделі процесів забезпечення попиту для таких випадків:

- а) випадкові коливання попиту і цін товарів;
- б) нелінійні функції витрат на партії та зберігання;
- в) урахування затримок постачання, штрафів за затримки, оптових знижок.

Інтелектуальність розроблюваної системи не в тому, що вона приймає рішення за менеджера, а в тому, що вона, може показати, як отримується це рішення, дозволяє

простим, природним способом змінювати параметри, математичні моделі і правила.

Розроблювана система є не тільки інструментом розрахунку оптимальних стратегій, але й тренажером для навчання та набуття досвіду в управлінні реальними системами в реальних умовах за допомогою імітаційного моделювання.

Модель з "моментальними" поставками. Формулювання задачі

Цей документ є не тільки розділом книги, але й ще автоматизованою системою навчання користувача - не легендарного "оператора АСУ", що діє згідно з "Інструкцією оператора", а спеціаліста, що на базі даної розробки може поступово створити, настроїти "персонального помічника".

Для цього необхідно мати доступ до робочих математичних моделей і можливість їх змінювати, а користувач повинен вільно орієнтуватись в моделях та програмах. Саме така концепція систем - засобів автоматизації роботи спеціаліста є світовим стандартом. Можна не дотримуватись цього стандарту, якщо поставити за мету створити нікому непотрібну систему.

Цей документ має дві форми існування і використання

- це звичайний паперовий документ - книга, техдокументація на автоматизовану систему;

- електронний документ з елементами управління та навігації, "живими формулами" і графіками.

В зв'язку з цим проведемо "варіантний аналіз" - розглянемо варіанти побудови інтерфейсу. Зробимо це на прикладі таблиці з описом вхідних даних базової задачі управління запасами.

Зведемо в таблицю змінні, параметри, їх визначення, і припущення відносно правил і закономірностей їх зміни. Розглянемо альтернативи створення таблиці в а) Mathcad, б) Word, в) Excel. Ні один з цих пакетів сьогодні, сам по собі, не може забезпечити потрібних можливостей.

Тому використаємо **компонентну технологію**. Суть компонентної технології в тому, що в документ, який розроблено в певному програмному середовищі, можна включити **компоненти** зроблені в інших програмних середовищах. Імпортований компонент буде працювати в іншому документі і може (не завжди) брати дані з цього середовища, обробляти їх і повертати в цей документ.

За все треба платити. За можливості компонентної технології ми платимо необхідністю купити ліцензовані версії компонентів, використовувати більш потужний комп'ютер. Розглянемо компонентну технологію на конкретному прикладі.

Для створення модулів автоматизованої системи підтримки рішень в управлінні запасами вхідні і вихідні дані слід подавати в таблицях, де подано разом текст, числові дані, графіку.

Табличний процесор Excel є неефективним і недружнім середовищем для систем з відкритими математичними моделями і програмами. Ідеальним у цьому плані є пакет Mathcad, але таблиці в ньому дозволяють вводити і виводити тільки дані без коментарів, "легенд". Тому бажано взяти за основну програмну платформу Mathcad, а таблиці створювати в інших платформах, наприклад, Axum, Word, Excel. Розглянемо ці альтернативи.

Пакет професійної графіки Axum створений тією ж корпорацією (MathSoft), що і Mathcad і має єдиний недолік - високу ціну.

Текстовий редактор Word є "безкоштовним" - він обов'язково є на комп'ютері. Нижче подано, як зразок, таблицю вхідних даних базової задачі управління запасами.

Альтернатива 1. Таблиця вхідних даних створена в Word

Величина	Позначення	Одиниці вимірювання	Припущення
Інтенсивність попиту	d	Одиниць товару за рік	Попит постійний і неперервний; увесь попит задовольняється
Організаційні витрати	s	Гривень за одну партію	Організаційні витрати постійні, не залежать від розміру партії
Вартість товару	c	Гривень за одиницю товару	Ціна одиниці товару постійна; розглядається тільки один вид товару
Витрати утримання запасів	h	Гривень за одиницю товару за рік	Вартість зберігання одиниці товару протягом року постійна
Розмір партії	q	Одиниць товару в одній партії	Постійна величина; нова партія надходить миттєво, як тільки рівень запасу стає нульовим

Пакет Word має великі можливості в створенні таблиць, але не дозволяє ефективно обробляти дані.

Пакет Excel є одним з головних компонентів пакету Mathcad. Причина цього, в поширеності табличного процесора Excel - з ним **сьогодні** працюють сотні мільйонів користувачів. Розглянемо технологію (інформаційну) експорту таблиць Excel в математичний пакет і навпаки. Excel - не оптимальне середовище для складних математичних обчислень і програм. Математичні моделі і програми розташовуються по комірках так, що їх важко бачити як одне ціле, важко контролювати. Виконуються такі програми досить повільно.

Вибираємо експорт таблиць Excel у середовище математичного пакета. Розглянемо особливості компонентної технології цього експорту. На рис. 1 подано послідовні стадії роботи з компонентним "візардом":

перший крок вибираємо компонент - це може бути будь-яка поширена програмна платформа; для прикладу вибрали Excel;

другий крок - вибираємо або готовий документ Excel, або створюємо таблицю "з нуля" в середовищі Excel;

третій крок - описуємо **входи і виходи** компонента - що він бере від документа-імпортера" і що повертає в цей документ.

Спеціаліст знає, що за цими маніпуляціями стоїть налаштування відповідних підпрограм-модулів звичайно написаних на VC++.



Рис. 1. Стадії створення програмного компонента

Альтернатива 2. Таблиця вхідних даних створена в Excel.

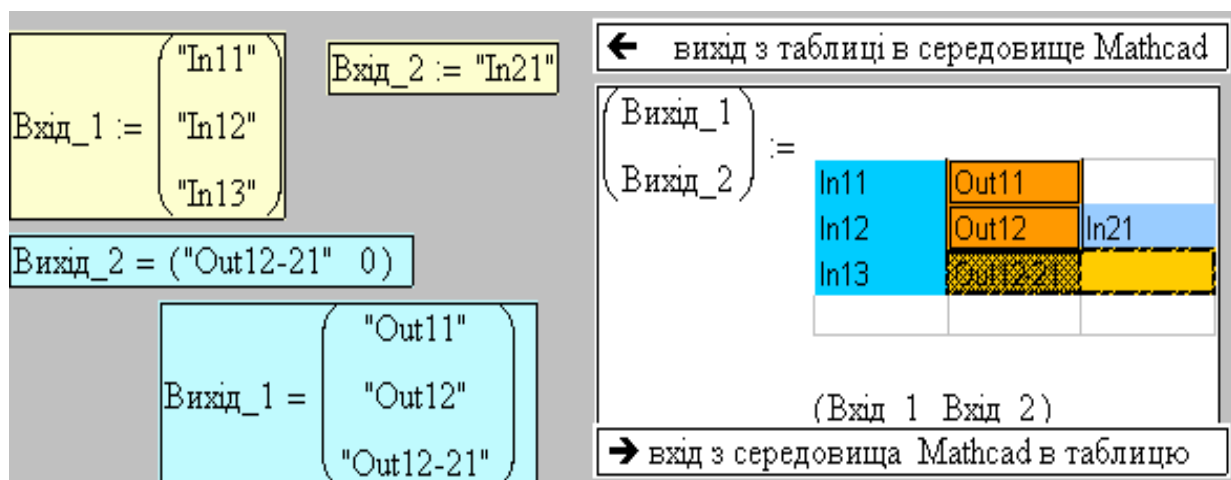
Таблиця_1 :=

Величина	Позначення	Одиниці вимірювання	Припущення
Інтенсивність попиту	d	Одиниць товару за рік	Попит постійний і неперервний; увесь попит задовольняється
Організаційні витрати	s	Гривень за одну партію	Організаційні витрати постійні, не залежать від розміру партії
Вартість товару	c	Гривень за одиницю товару	Ціна одиниці товару постійна; розглядається тільки один вид товару
Витрати на утримання запасів	h	Гривень за одиницю товару за рік	Вартість збереження одиниці товару протягом року постійна
Розмір партії	q	Одиниць товару в одній партії	Постійна величина; поставка товару виконується одразу, як тільки рівень запасів стане рівним нулю

Подана вище таблиця створена в Excel. З цієї таблиці передається в змінну "Таблиця_1" те, що описано в меню "properties", а саме - "діапазон" A1:B2 (для прикладу).

$$\text{Таблиця}_1 = \begin{pmatrix} \text{"Величина"} & \text{"Позначення"} \\ \text{"Інтенсивність попиту"} & \text{"d"} \end{pmatrix}; \quad \text{ORIGIN} := 1$$

Імпортована таблиця може мати декілька таких виходів - так дані **експортуються з таблиці Excel**. Можна також створити декілька входів - так дані **імпортуються в таблицю Excel з Mathcad**, наприклад, можлива така структура:



Така компонентна технологія дозволяє швидко створювати зручні надійні інтерфейси для системи управління запасами. Розгляньте уважно подані вище матриці, вектори, таблиці. Все це "об'єкти", що обмінюються "повідомленнями".

Постановка задачі управління запасами

Задача управління запасами формулюється так: **визначити розмір партії товару q** (що періодично замовляється) **при якій сумарні річні витрати** (на забезпечення поставок) **будуть мінімальними.**

Побудуємо **графік зміни запасів** для нашої базової задачі. Розробляємо програму, що відповідає поданим в таблиці 1 припущенням щодо параметрів задачі.

Запас(q, d) :=	ORIGIN $\leftarrow 1$ "N- глобальна змінна" розход $\leftarrow d \div N$ зап ₁ $\leftarrow q$ for $k \in 2..N$ ostat $\leftarrow$ зап _{$k-1$} – розход зап _{k} $\leftarrow$ ostat if ostat > 0 q + зап _{$k-1$} otherwise zap
--------------------	---

Тестуємо програму:

Запас (q, d) ^T =		1	2	3	4	5	6	7	8	9	10
	1	208	204	200	196	192	188	184	180	176	172

Далі створюємо ще один об'єкт із сучасного арсеналу програмування - об'єкт класу "текст", до якого вбудовано програмні об'єкти класів "зона вводу", "зона виводу" та ін. Беремо його в рамку.

Згідно з сучасними стандартами програмування і створення документів в усіх програмних платформах (вони ж усі об'єктно орієнтовані) передбачені опції "подати границю об'єкта (групи об'єктів)", "виділити кольором об'єкт (групу об'єктів)".

Такий об'єкт можна переносити з документа в документ як звичайний програмний модуль. Він дуже зручний для сприйняття та контролю людиною, бо він має звичну людині форму.

Задаємо для прикладу: попит за рік $d = 1092$; розмір партії $q \equiv 208$; вартість одиниці товару $c = 1$; витрати на зберігання одиниці $h = 4$; оргвитрати на партію $s = 40$; кроків моделювання $N \equiv 273$ $k := 1..N$

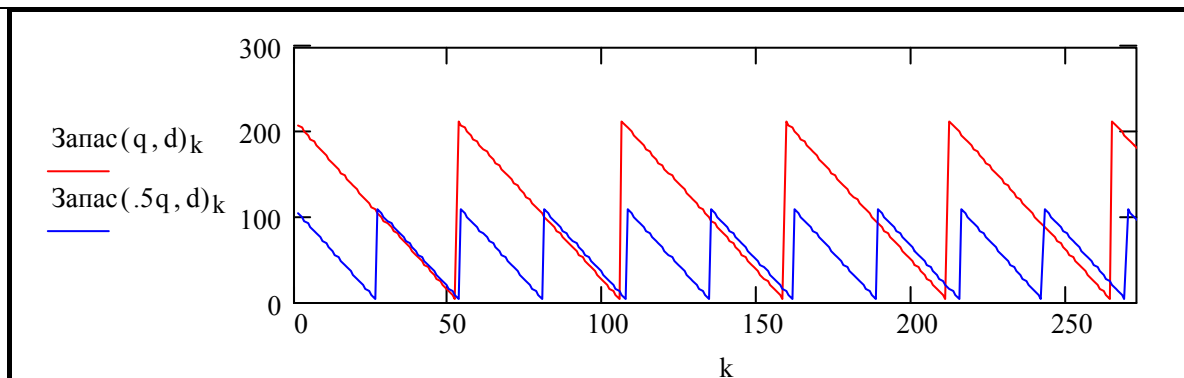


Рис.2. Графік зміни запасів

На цьому графіку подано два процеси зміни запасів для партії $q = 208$ одиниць і для вдвічі меншої партії. Дивлячись на графіки, неважко вивести основні залежності для процесу управління запасами.

1. Для повного задоволення попиту в розмірі d при розмірі поставки q потрібно зробити d/q поставок за рік, тобто:

$$\text{Поставок_за_рік} := \frac{d}{q}$$

$$\text{Поставок_за_рік} = 5.25$$

2. Середній рівень запасів буде:

$q/2 = \text{площа_під_графіком_за_цикл_}/\text{тривалість_циклу}$.

Середній рівень запасів за рік визначаємо вбудованою функцією середнього:

$\text{Середнє_за_рік} := \text{mean}(\text{Запас}(q, d));$

$$\text{Середнє_за_рік} = 110.52$$

3. Терміни "партія" і "поставка" в цих задачах - синоніми.

Отримання рівняння витрат

Сумарні витрати складаються з

1) C1- організаційних витрат на поставки, пропорційних кількості поставок;

2) C2 - вартості запасів;

3) C3 - витрат на зберігання, що пропорційні середньому рівню запасів.

Запишемо рівняння витрат, що пов'язані з утриманням потрібних запасів протягом року (або іншого планового періоду). Зробимо це у трьох альтернативних формах.

$$\text{В традиційній: } C1 := s \cdot \frac{d}{q}; \quad C2 := c \cdot d; \quad C3 := h \cdot \frac{q}{2}.$$

В сучасній: $\text{загальні_оргвитрати} := C1;$

$\text{вартість_товарів} := C2;$

$\text{загальні_витрати_зберігання} := C3.$

В сучасних системах програмування ім'я змінної є часто є описом змінної, допускаються національні алфавіти, фактично не обмежена довжина імені.

Запишемо формулу, що майже не потребує коментарів.

$C_{\text{sum}} := \text{загальні_оргвитрати} + \text{вартість_товарів} + \text{загальні_витрати_зберігання};$

$$C_{\text{сумарн}} := C1 + C2 + C3; \quad C_{\text{сум_витр}} := s \cdot \frac{d}{q} + c \cdot d + h \cdot \frac{q}{2}$$

$$C_{\text{sum}} = 1718; \quad C_{\text{сумарн}} = 1718; \quad C_{\text{сум_витр}} = 1718.$$

Визначення оптимального розміру партії q

Для знаходження оптимального розміру партії знайдемо похідну від сумарних витрат та прирівняємо її до нуля. Зробимо це в двох альтернативних формах - символічній та числовій. Визначаємо сумарні витрати як функцію користувача

$$Cs(s, d, q, c, h) := s \cdot \frac{d}{q} + c \cdot d + h \cdot \frac{q}{2}.$$

Використовуємо символьний процесор для операцій "диференціювати" та "знайти нульовий корінь рівняння". Процесор видає все, що може і знає - два кореня, але нам потрібен лише один - додатний. Увага! Щоб отримати саме **символьний** результат, треба а) або не присвоювати змінним числові значення, б) або "зняти" їх, отак:

$$s := s; \quad c := c; \quad d := d; \quad q := q \cdot h := h$$

$$\frac{d}{dq} Cs(s, d, q, c, h) \text{ expand, } q \rightarrow -s \cdot \frac{d}{q^2} + \frac{1}{2} \cdot h \text{ solve, } q \rightarrow \left[\begin{array}{l} \frac{1}{h} \cdot \sqrt{2} \cdot (h \cdot s \cdot d)^{\frac{1}{2}} \\ \frac{-1}{h} \cdot \sqrt{2} \cdot (h \cdot s \cdot d)^{\frac{1}{2}} \end{array} \right]$$

Зауваження 1. Згідно з традиційними стандартами проектування слід виконувати "варіантний аналіз", потім вибрати **єдиний** "оптимальний" варіант. Для інформаційних та інших подібних систем працює інша парадигма оптимізації "замість вибору найкращої альтернативи - найкращий набір альтернатив". Наприклад, в пакеті Mathcad паралельно співіснують дві альтернативні версії символьного процесора - результат роботи однієї подано вище, в другій (меню Symbolics) тіж самі дії мають такий вигляд:

$$s \cdot \frac{d}{q} + c \cdot d + h \cdot \frac{q}{2} \text{ by differentiation, yields } -s \cdot \frac{d}{q^2} + \frac{1}{2} \cdot h ;$$

$$-s \cdot \frac{d}{q^2} + \frac{1}{2} \cdot h \text{ has solution(s) : } \left[\begin{array}{l} \frac{1}{h} \cdot \sqrt{2} \cdot (h \cdot s \cdot d)^{\left(\frac{1}{2}\right)} \\ \frac{-1}{h} \cdot \sqrt{2} \cdot (h \cdot s \cdot d)^{\left(\frac{1}{2}\right)} \end{array} \right]$$

Між іншим, на базі такої парадигми побудовані природні мови - для певного поняття існує декілька варіантів, синонімів. Вони збагачують мову і роблять її більш інформативною, наприклад, в англійській мові не менше чотирьох синонімів для поняття "ціль", в американській - набагато більше, українській, поки два: ціль та мета.

Зауваження 2. Згідно з традиційними стандартами призначення засобів автоматизації праці спеціаліста - підвищення продуктивності праці, скорочення персоналу. Згідно з сучасними світовими стандартами це:

1) можливість спеціалісту мати вільний час для експериментів, всебічного випробування новацій, що закладаються у проект;

2) виключення можливості для розробника робити помилки.

Чому не можна виконувати елементарні перетворення вручну? Помилки в паперових книгах були певним чином "не летальними". Той, хто брав формулу з книги для обчислень, міг помітити помилку.

В електронних документах формули є активними - те, що написано, обчислюється. В США проведена перевірка програм автоматизації бізнес-процесів, саморобних, зроблених в середовищі Excel. Була виявлена велика кількість "непомітних" помилок.

Сьогодні документ, де формули і рівняння набрані вручну, в текстовому редакторі, а не за допомогою символьного процесора, з високою імовірністю містить помилки.

Перевіримо альтернативні вирази для похідної (задані вище).

числовий:

$$\text{Пох}(s, d, q, c, h) := \frac{d}{dq} Cs(s, d, q, c, h) ;$$

$$\text{Пох}(s, d, q, c, h) = 0.99 ;$$

аналітичний:

$$\text{Про}(s, d, q, c, h) := -s \cdot \frac{d}{q^2} + \frac{1}{2} \cdot h$$

$$\text{Про}(s, d, q, c, h) = 0.99$$

Записуємо вирази для **оптимального розміру партії (вперше отримано Ф. Харрісом у 1915 році)** так як видав символічний процесор і в звичній для нас формі:

$$Q_{op}(s, h) := \frac{1}{h} \cdot \sqrt{2} (h \cdot sd)^{\left(\frac{1}{2}\right)} ;$$

$$Q_{op}(s, h) = \sqrt{\frac{2 \cdot s \cdot d}{h}}$$

В практичних задачах функції часто задаються в табличному вигляді і аналітичний вираз не завжди можливо отримати. Подаємо більш універсальний вираз для знаходження оптимальної партії - з використанням числового методу знаходження кореня. Порівнюємо результати отримані альтернативними методами.

$$\text{Оппартія} := \text{root}(\text{Пох}(s, d, q, c, h), q, 1, d) :$$

$$\text{Смін} := Cs(s, d, Q_{op}(s, h), c, h)$$

$$\text{Оппартія} = 148 ; \quad \boxed{Q_{op}(s, h) = 148} ; \quad \boxed{\text{Смін} = 1683.1} .$$

Побудуємо графіки для сумарних витрат та їх компонентів. Задаємо ранжовану змінну $qq := \text{ceil}(0.01 \cdot d), 2 \cdot \text{ceil}(0.01 \cdot d) .. d$. Визначаємо компоненти сумарних витрат як функції відповідних параметрів.

$$C1v(s, d, q) := s \cdot \frac{d}{q} - \text{організаційні витрати на поставки партій виробів};$$

$$C2v(c, d) := c \cdot d - \text{вартість виробів};$$

$$C3v(h, q) := h \cdot \frac{q}{2} - \text{витрати на зберігання виробів}.$$

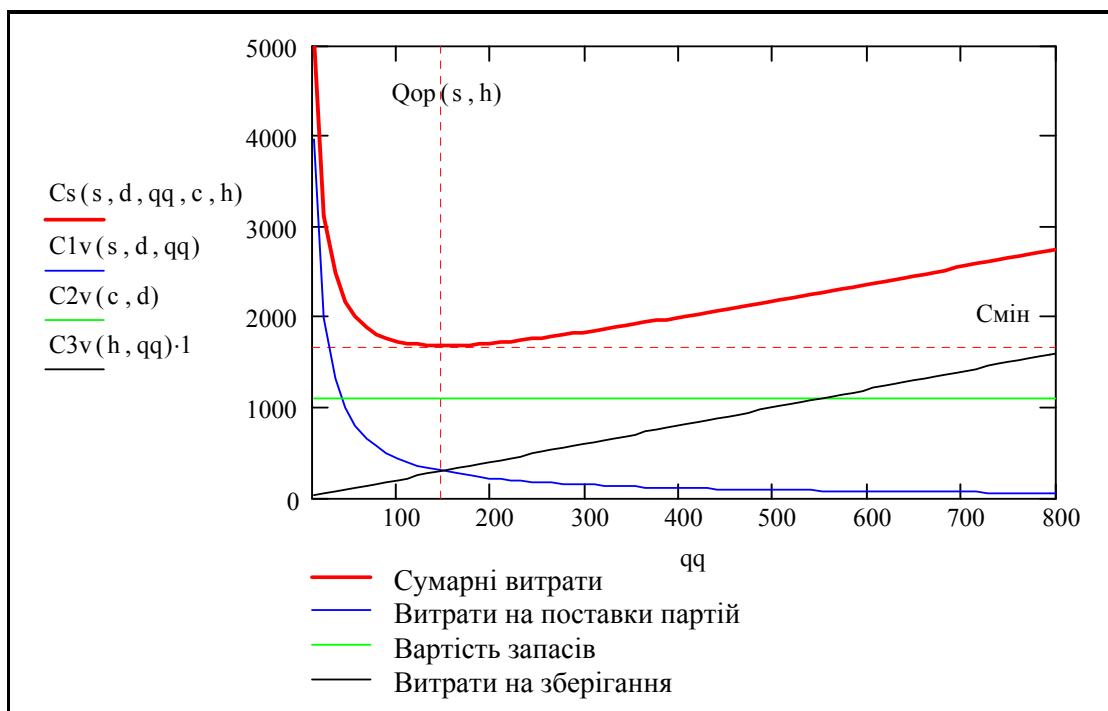


Рис.3. Залежність витрат від розміру партії

Задача оптимізації розміру партії при моментальних поставках

Задача оптимізації розміру партії при моментальних поставках

Вхідні дані: попит за рік: $d \equiv 1092$; вартість одиниці товару: $c \equiv 1$; витрати на зберігання одиниці: $h \equiv 4$; оргвитрати на партію: $s \equiv 40$.

Результати оптимізації: розмір партії: $Q_{op}(s, h) = 148$; число партій (поставок за рік): $d \div Q_{op}(s, h) = 7.39$; сумарні витрати: $C_{мін} = 1683.1$ грн.

Те, що вище у рамці (конгломерат тексту, зон введення та виведення) - останній фрагмент стенду для задачі управління запасами. Цей стенд для аналізу та розв'язання задачі управління запасами при разових поставках складається з трьох модулів: графіка, модуля введення даних та модуля виведення результатів. Зрозуміло, що цей стенд можна безмежно вдосконалювати і розширювати.

Зауваження. Термін "моментальні" поставки означає, що замовлену партію ви отримуете одразу, в повному розмірі. В наступному розділі ми розглянемо інший спосіб постачання - розподілені поставки, що виконуються на протязі певного періоду

Завдання для самостійного виконання

Вибрано і розроблено компоненти і програмні модулі для базової задачі управління запасами при "моментальних" поставках. На базі цього документа можна розробляти більш складні модулі для системи управління запасами.

1. Розробіть стенд "що буде, якщо" для аналізу типових ситуацій у вигляді "комп'ютеризованих" висловлювань типу: "Якщо оргвитрати на одну партію зміняться на $QQ\%$, то сумарні витрати зміняться на $QKRQ\%$. Рекомендуємо оптимальну партію $OpPa$.

2. Розв'яжіть оптимізаційну задачу для випадку лінійного (тренду) **росту/падіння** цін на товар на $pp\%$ (товар закуповується у постачальників і продається клієнтам за поточними цінами - це вже не тільки управління запасами, але й біржова гра?).

3. Розв'яжіть оптимізаційну задачу для випадку сезонних коливань попиту на

$$\text{продукт за законом } c(t) = c_0 \cdot \left(1 + qq \cdot \sin\left(\frac{2 \cdot \pi \cdot t}{T_{річн}}\right) \right).$$

Контрольні запитання

1. Назвіть компоненти витрат на забезпечення постачання.
2. Змінні управління для задачі постачання
3. Приклади задач і систем управління запасами.
4. Критерій оптимізації для задачі управління запасами.
5. Метод знаходження оптимального розміру партії.
6. Запишіть формули для компонентів сумарних витрат на постачання.
7. Дайте постановку задачі оптимізації розміру партії для випадку, коли ціна продукту постачання лінійно росте/падає.
8. Що таке компонентні технології розробки програм?



3.2. Задача управління запасами при розподілених поставках

Вступ. Постановка задачі

В підрозділі 3.1 ми розглянули математичну модель управління запасами при разових, "моментальних" поставках. Тепер розглянемо дещо іншу модель з іншою організацією поставок. Ця модель описує випадки, коли і виробництво і попит і система забезпечення попиту належать до одного підприємства. Розглянемо процес постачання у цьому випадку.

Як і в попередній задачі задано річний попит d одиниць на рік, шукається оптимальний розмір партії q одиниць. На початку циклу постачання склад пустий. Видається замовлення і запускається певна виробнича лінія чи агрегат, що видає потрібні вироби в кількості p одиниць на рік.

Лінія працює деякий період T_p , поки не виробить q одиниць виробів. Частина цих виробів іде безпосередньо на забезпечення попиту, а частина - на склад. Потім лінія виключається, постачання ведеться зі складу, аж поки запас не стане нульовим (пороговим). Потім цикл повторюється. На рис.1 подано графік зміни запасів. Така модель називається ще **моделлю виробничих поставок**.

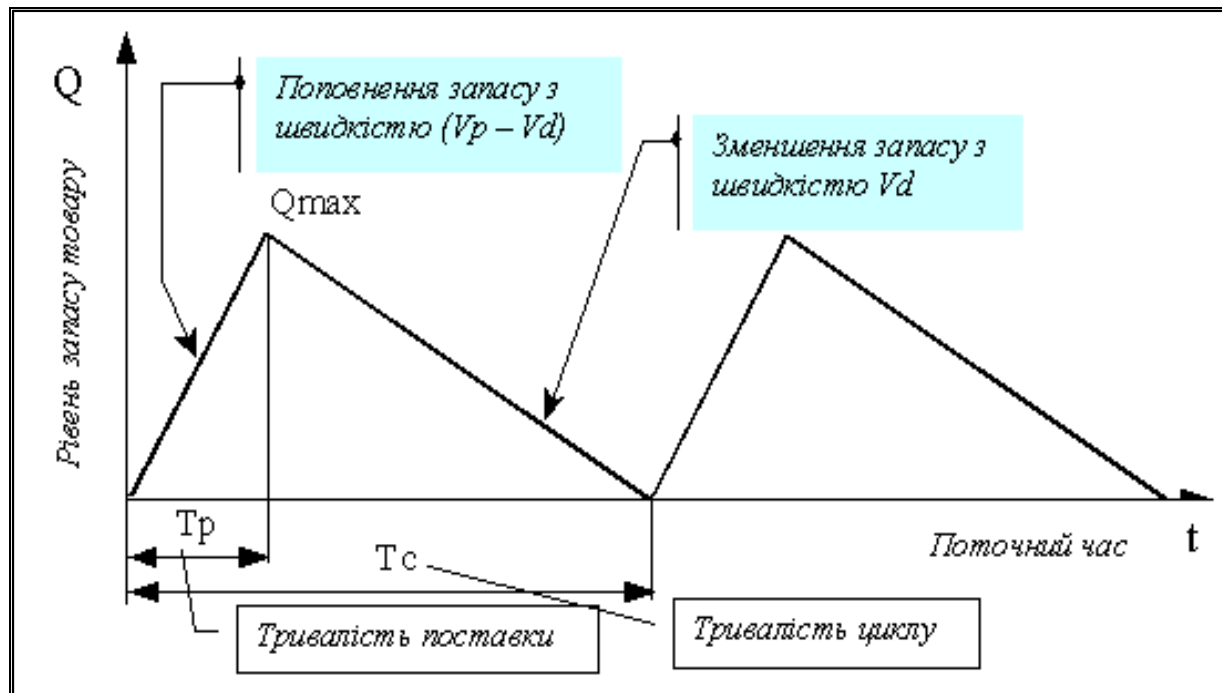


Рис. 1. Процес з неперервними поставками

Чи не надумана така організація постачання? А чому не можна зробити так, щоб виробнича лінія видавала вироби в темпі, що забезпечує потребу, тобто $p = d$? Сучасні виробничі лінії, агрегати базуються на високих технологіях, їх продуктивність оптимізована, її не можна змінювати - це приведе до браку, або збільшення витрат.

Записуємо вираз для сумарних витрат.

$$C.\Sigma = C1 + C2 + C3, \text{ де} \quad (1)$$

C1 - витрати на організацію поставок (оформлення замовлення, налагодження виробничої лінії);

C2 - вартість річної кількості виробів (товарів);

C3 - витрати на зберігання виробів на складі. Визначимо усі ці компоненти.

$C1 = n \cdot s = \frac{d}{q} \cdot s = \frac{s \cdot d}{q}$, де $n = \frac{d}{q}$ - кількість поставок за рік, s - витрати на одну партію.

$C2 = c \cdot d$, де c - вартість одиниці виробу.

$C3 = \text{середній_рівень_запасів} \cdot h$, де h - вартість зберігання одного виробу протягом року.

При розробці автоматизованої системи - **помічника менеджера** - будемо дотримуватись гнучких компромісів - використовувати традиційні, короткі змінні в текстах, постановочних розділах, і довгі - в програмах. Дивлячись на рис. 1 визначимо середній рівень запасів. Максимальний рівень запасів дорівнює:

$$Q_{\max} = (p - d) \cdot T_p, \text{ де}$$

T_p - період, протягом якого виробнича лінія виробить q одиниць виробів (звичайно він менший року, тобто $T_p < 1$). Згідно з визначенням $T_p = \frac{q}{p}$, тому маємо:

$$Q_{\max} = (p - d) \cdot \frac{q}{p}.$$

Ще раз дивимось на трикутник на рис. 1 і записуємо:

$$\text{середній_рівень_запасів} = \frac{Q_{\max}}{2} = \frac{(p - d) \cdot q}{2 \cdot p}.$$

Записуємо нарешті вираз для сумарних витрат в системі управління запасами.

$$C_{\text{сум}}(d, q, p, s, c, h) := \frac{s \cdot d}{q} + 0.05c \cdot d + \frac{(p - d) \cdot q \cdot h}{2 \cdot p} \quad (2)$$

Єдина (поки) змінна управління (= те, що від нас залежить) - q . Необхідна умова екстремуму - похідна по цій змінній від сумарних витрат повинна бути нульовою. Дивимось на (2) і бачимо, що це саме мінімум. Знайдемо оптимальний розмір партії у символічному та числовому вигляді. Знаходимо похідну

$$\frac{d}{dq} \left[\frac{s \cdot d}{q} + 0.05c \cdot d + \frac{(p - d) \cdot q \cdot h}{2 \cdot p} \right] \rightarrow -s \cdot \frac{d}{q^2} + \frac{1}{2} \cdot \frac{(p - d)}{p}.$$

Прирівнюємо до нуля та розв'язуємо відносно q

$$-s \cdot \frac{d}{q^2} + \frac{1}{2} \cdot (p - d) \cdot \frac{h}{p} = 0 \text{ solve, } q \rightarrow \left[\begin{array}{l} \frac{1}{(-p + d)} \cdot [-(-2 \cdot p + 2 \cdot d) \cdot s \cdot d \cdot p]^{\frac{1}{2}} \\ \frac{-1}{(-p + d)} \cdot [-(-2 \cdot p + 2 \cdot d) \cdot s \cdot d \cdot p]^{\frac{1}{2}} \end{array} \right]$$

З того, що видав символічний процесор беремо додатний корінь, потім приводимо до звичного вигляду і порівнюємо дві версії формули на довільних значеннях параметрів

$$s := 1 ; \quad h := 2 ; \quad p := 4000 ; \quad d := 2400 ; \quad q := 200 ; \quad c := 1$$

$$\frac{-1}{(-h \cdot p + h \cdot d)} \cdot [-(-2 \cdot p + 2 \cdot d) \cdot h \cdot s \cdot d \cdot p]^{\left(\frac{1}{2}\right)} = \sqrt{\frac{(2 \cdot p \cdot s \cdot d)}{(p - d) \cdot h}} ;$$

$$q_{\text{opsym}}(h) := \sqrt{\frac{(2 \cdot p \cdot s \cdot d)}{(p - d) \cdot h}} \quad (3)$$

Робимо теж саме у числовому вигляді - створюємо функцію - похідну:

$$DCsum(d, q, p, s, c, h) := \frac{d}{dq} Csum(d, q, p, s, c, h).$$

Знаходимо нульовий додатний корінь за допомогою числового методу

$$\text{Given} \quad DCsum(d, q, p, s, c, h) = 0 \quad q > 0$$

$$q_{\text{opnum}}(h) := \text{Find}(q) \quad (4)$$

Записуємо остаточні результати, порівнюємо

$$q_{\text{opsym}}(h) = 77.46 ;$$

$$q_{\text{opnum}}(h) = 77.46$$

Проаналізуємо альтернативні методи розв'язання задачі. Незалежно від результатів аналізу, на питання: "який метод кращий, яку альтернативу вибрати", ми повинні відповісти собі "обидві", нехай взаємодоповнюють, взаємоконтролюють одна одну. Це ж інформаційні системи, а інформація майже "невагома".

Символьне розв'язання зручне тим, що дає явну аналітичну формулу - це зручно і для теорії і для практики. Але так мало випадків, коли ми маємо аналітичне розв'язання.

Символьне розв'язання зручне своєю універсальністю - ми можемо задавати вище виразу (2), де визначається функція користувача, будь які моделі витрат та постачання - все, що нижче не змінюється.

Для прикладу ми зробили обидва розв'язання (3) та (4) функціями змінної **h**. При бажанні список змінних ми можемо розширити. Плата за розширення можливостей, узагальнення результатів - досить потужний мікропроцесор та пам'ять - сьогодні не така велика. Побудуємо графіки функції сумарних витрат та її складових. Введемо нові позначення для параметрів і змінних:

$$dm := 2400 ; \quad q := 1, 3 \dots dm ; \quad p := 4800 ; \quad s := 2 ; \quad h := 1 ; \quad c := 1.$$

Для того, щоб зробити цю сторінку автономною переписуємо (копіюємо) вирази для функції витрат та її компонентів. Вводимо "інформативні" імена змінних.

$$Csum(d, q, p, s, c, h) := \frac{s \cdot d}{q} + 0.02c \cdot d + \frac{(p - d) \cdot q \cdot h}{2 \cdot p}$$

організаційні витрати $C1(s, d, q) := \frac{s \cdot d}{q}$; вартість партії виробів $C2(c, d) := 0.02c \cdot d$;

витрати на зберігання $C3(d, p, q, h) := \frac{(p - d) \cdot q \cdot h}{2 \cdot p}$.

$$q_{\text{opsym}}(h) := \sqrt{\frac{(2 \cdot p \cdot s \cdot d)}{(p - d) \cdot h}} ; \text{ Опт_кільк_партій} := \frac{d}{q_{\text{opsym}}(h)} ;$$

Опт_витрати := $C_{\text{сум}}(d, q_{\text{opsym}}(h), p, s, c, h)$; Оп_партія := $q_{\text{opsym}}(h)$.

Результати оптимізації $\boxed{\text{Опт_витрати} = 117.3 \text{ грн.}} ; \boxed{\text{Оп_партія} = 139 \text{ шт.}} ;$
 $\boxed{\text{Опт_кільк_партій} = 17}$.

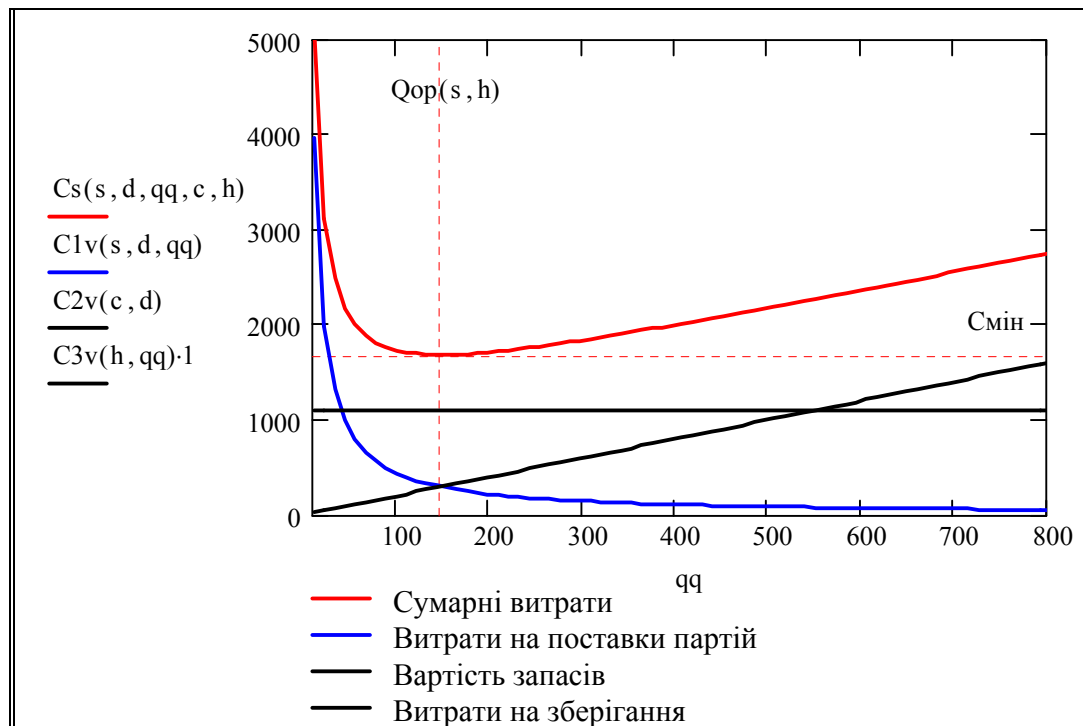


Рис.2. Функція сумарних витрат та її складові

Розробка програми для моделювання процесу з виробничими поставками

Отримали аналітичні формули для розрахунку оптимальної партії, що працюють тільки у випадку постійних цін і попиту. Тепер напишемо програму для моделювання процесу з виробничими поставками. Програму будемо так, щоб вона була придатна і для довільних моделей попиту і поставок. Дивимось на трикутник одного циклу роботи системи (рис.1) і записуємо умови продовження режимів "поставки" та "витрати" (коли нема поставок), а також умови переключення режимів - усього 4 умови.

Програму робимо функцією трьох параметрів: **d** – річного (темпу) попиту; **p** – темпу постачання (в перерахунку на рік); **qs** – розміру партії (беремо іншу назву тому, що вище q було ранжованою змінною). Вводимо кількість кроків моделювання – **Nkr**, задаємо ранжовану змінну – номер кроку. Внутрішня змінна **flag** характеризує стан системи (1 – є поставка, 0 – нема поставок).

Вводимо вхідні дані програми:

ORIGIN := 1 ; Nkr := 100 ; k := 1..Nkr ; d = 2400 ; p := 10000 ; qs := 700 .

Програма моделювання процесу зміни запасів при виробничих поставках.

$Po(d, p, qs) :=$	$Q_{\max} \leftarrow (p - d) \cdot \frac{qs}{p}$ $qq_1 \leftarrow 0$ $flag \leftarrow 1$ for $j \in 1 \dots N_{kr}$ "Якщо запас менший максимуму і режим поставок - продовжувати" $qq_{j+1} \leftarrow qq_j + \frac{(p - d)}{N_{kr}} \quad \text{if } (qq_j < Q_{\max}) \cdot flag > 0$ "Якщо запас менший максимуму і режим витрат - продовжувати" $qq_{j+1} \leftarrow qq_j - \frac{d}{N_{kr}} \quad \text{if } (qq_j < Q_{\max}) \cdot flag < 1$ "Якщо запас менший мінімуму - почати режим поставки" if $qq_j < 0$ $flag \leftarrow 1$ $qq_{j+1} \leftarrow qq_j + \frac{(p - d)}{N_{kr}}$ "Якщо запас більший максимуму - зупинити виробництво" if $qq_j > Q_{\max}$ $flag \leftarrow 0$ $qq_{j+1} \leftarrow qq_j - \frac{d}{N_{kr}}$
	qq

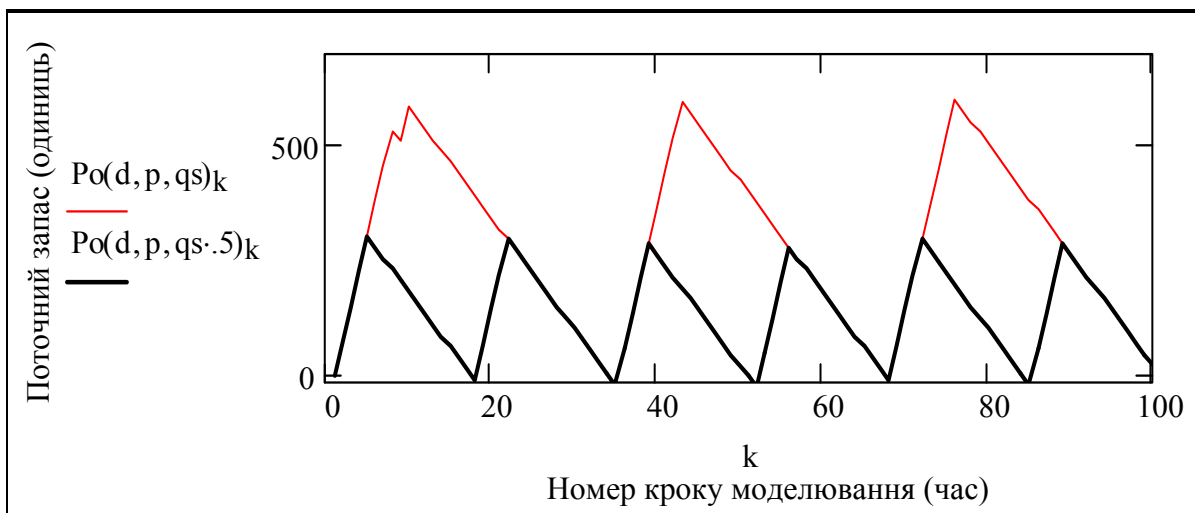


Рис.3. Видача програми: графіки зміни запасів при виробничих поставках

3.3. Задача управління запасами зі штрафами

Вступ. Постановка задачі

Розглянемо ще одну модифікацію базової моделі управління запасами - модель зі штрафами. Припустимо, що коли виникає дефіцит - на складі нема товару під поточний попит, то він задовольняється трохи пізніше - з наступної поставки, а клієнтам платять штраф за несвоєчасну поставку. Така ситуація здається реальною, конкретизуємо її [2].

Деяка фірма згідно з контрактом постачає q одиниць товару протягом періоду TL . Темп попиту d одиниць за одиницю часу (за рік). Партія в q одиниць витрачається за період TL : $q = TL \cdot d$.

Припустимо, що на початку періоду TL створено запас тільки в y одиниць товару: $y < q$. Тоді до закінчення періоду виникне дефіцит і черга, тому що поточні вимоги нічим задовольняти. В кінці періоду TL буде виконано поставку розміром q . З цієї поставки буде "закрита" черга і будуть виконуватись поточні вимоги, але неважко показати, що знов виникне дефіцит, і все повториться на новому циклі. Слід тільки пам'ятати, що за затримку поставок сплачується штраф, який залежить від розміру дефіциту та часу затримки.

На рис. 4 подано графік такого процесу. Тепер, дивлячись на графік, спробуйте дати відповідь - чи вигідно мати в кінці кожного періоду черги і штрафи, але зменшити рівень запасів на складі? Природно, і це слід пам'ятати, що крім витрат в моделі зі штрафами існують і втрати, наприклад клієнт, що двічі попав під дефіцит, на третій раз візьме товар в іншому місці. Але можна знайти рішення, коли дефіцити не приводять до втрати клієнтів. Придумайте, як це організувати.

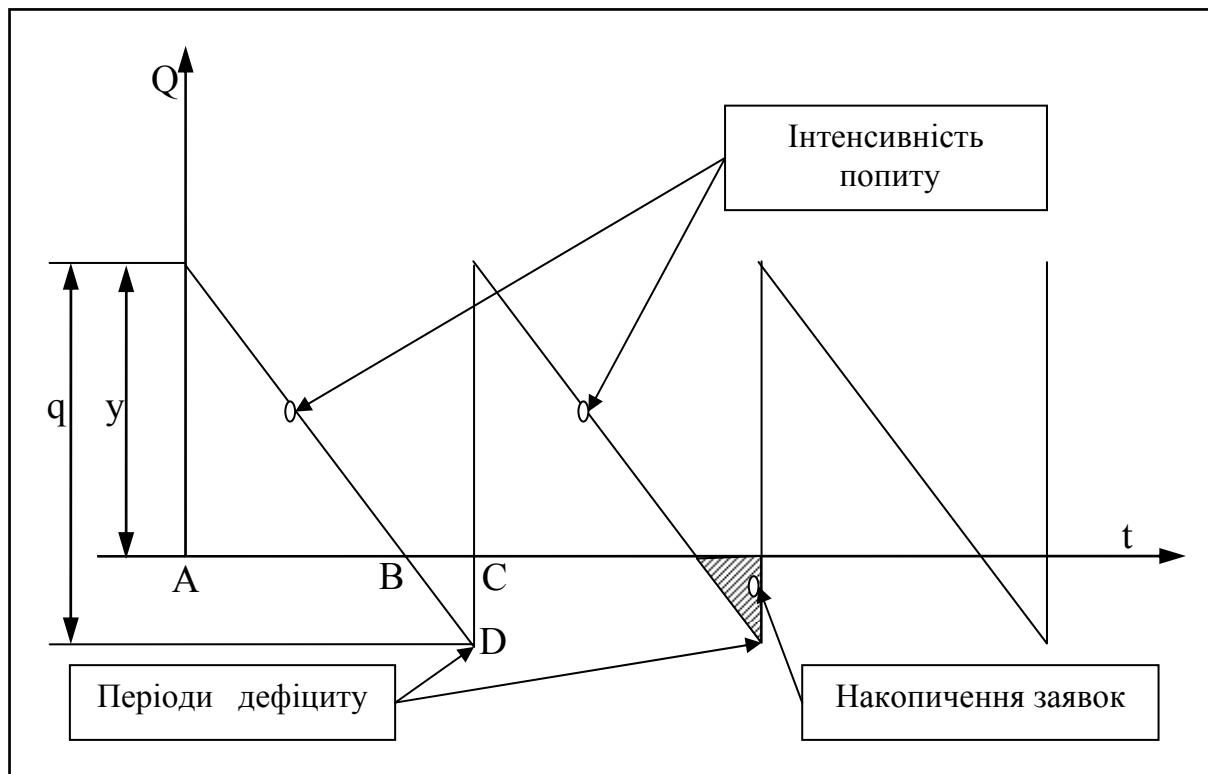


Рис. 4. Графік зміни запасів для випадку поставок зі штрафами

В цій задачі враховуємо такі витрати:

h - витрати на зберігання одиниці товару за одиницю часу;

P - витрати на штраф за затримку постачання одиниці товару за одиницю часу.

Запишемо вираз для сумарних витрат.

$$C.\Sigma = C1 + C2, \text{ де}$$

C1 - загальні витрати на підтримку запасів;

C2 - загальні витрати на штрафи.

Визначимо ці складові витрат. Спочатку витрати на зберігання. Товари знаходяться на складі протягом періоду АВ (див рис. 4), середній рівень запасів очевидно буде $y/2$, тривалість періоду $AB = y/d$. Таким чином,

$$C1 = h \cdot \frac{y}{2} \cdot \frac{y}{d} = \frac{h \cdot y^2}{2 \cdot d}.$$

Тепер визначимо витрати на штрафи. Штраф виплачується за невиконання поставок протягом періоду ВС (див рис. 4). Загальне число "товароднів", за які треба виплатити штраф, дорівнює площі заштрихованого трикутника BCD. Ця площа дорівнює

$$\left(\frac{q-y}{d} \right) \cdot \left(\frac{q-y}{2} \right). \text{ Записуємо вираз для } C2 \quad C2 = \frac{P \cdot (q-y)^2}{2d}.$$

Таким чином, сумарні витрати будуть

$$C_{\text{сум3}}(d, q, y, h, P) := \left[\frac{h \cdot y^2}{2 \cdot d} + \frac{P \cdot (q-y)^2}{2d} \right].$$

На відміну від попередніх задач змінною оптимізації тут є не розмір партії, а максимальний запас y . Визначимо величину $U_{\text{опт}}$, що мінімізує сумарні витрати. Знайдемо оптимальний розмір партії у символьному та числовому вигляді. Спочатку у символьному вигляді (вводимо імена змінних, що не мають числових значень)

$$\frac{d}{dy1} \left[\frac{h3 \cdot y1^2}{2 \cdot d3} + \frac{Pp \cdot (q3 - y1)^2}{2d3} \right] \rightarrow h3 \cdot \frac{y1}{d3} - Pp \cdot \frac{(q3 - y1)}{d3}.$$

Прирівнюємо до нуля та розв'язуємо відносно y

$$h3 \cdot \frac{y}{d3} - P \cdot \frac{(q3 - y)}{d3} = 0 \text{ solve, } y \rightarrow P \cdot \frac{q3}{(h3 + P)}.$$

Приводимо отриманий вираз до звичного вигляду і порівнюємо дві версії формули - символьну і числову на довільних значеннях параметрів

$$P := 5 ; h := 1 ; d := 2400 ; q := 200 ; c := 1 ; y := 100 ;$$

$$y_{\text{opsym}}(h) := P \cdot \frac{q}{(h + P)} ;$$

Робимо теж саме у числовому вигляді - створюємо функцію користувача - похідну:

$$DC_{\text{сум3}}(d, q, y, h, P) := \frac{d}{dy} C_{\text{сум3}}(d, q, y, h, P).$$

Знаходимо нульовий додатний корінь за допомогою числового методу

Given

$$DC_{\text{сум3}}(d, q, y, h, P) = 0$$

$$y_{\text{опnum}}(h) := \text{Find}(y) .$$

Записуємо остаточні результати, порівнюємо

$$y_{\text{opsym}}(h) = 166.667 ;$$

$$y_{\text{опnum}}(h) = 166.667 .$$

Проаналізуємо використання альтернативних методів розв'язання задачі: збіглися результати - з високою ймовірністю результати коректні, якби не збіглися, слід було б шукати помилки. Для прикладу ми зробили обидва розв'язання функціями змінної **h**. При бажанні список змінних ми можемо розширити. Побудуємо графіки функції сумарних витрат та її складових. Введемо нові означення для параметрів і змінних. (Дійсний ввід там, біля графіків)

$$dm = 2400 ; \quad y = 1, 3 .. dm ; \quad P = 5 ; \quad h = 1 .$$

Для того, щоб зробити цю сторінку автономною переписуємо (копіюємо) вирази для функції витрат та її компонентів. Вводимо естетичні імена змінних.

$$C1(h, d, y) := \frac{h \cdot y^2}{2 \cdot d} ;$$

$$C2(P, d, y, q) := \frac{P \cdot (q - y)^2}{2d} ;$$

$$C_{\text{сум3}}(d, q, y, h, P) := \left[\frac{h \cdot y^2}{2 \cdot d} + \frac{P \cdot (q - y)^2}{2d} \right] ;$$

$$y_{\text{opsym}}(h) := P \cdot \frac{q}{(h + P)} ;$$

$$\text{Опт_витрати} := C_{\text{сум3}}(d, q, y_{\text{opsym}}(h), h, P) ;$$

$$\text{Оп_запас} := y_{\text{opsym}}(h) ;$$

$$\text{партія} := q ;$$

$$\text{кільк_партій} := \frac{d}{q} .$$

.

Подаємо результати оптимізації:

$$\overline{\text{Опт_витрати}} = 6.9 \text{ грн.}; \quad \overline{\text{Оп_запас}} = 166.7 \text{ шт.}; \quad \overline{\text{кільк_партій}} = 12$$

Будуємо графіки залежності витрат від розміру партії

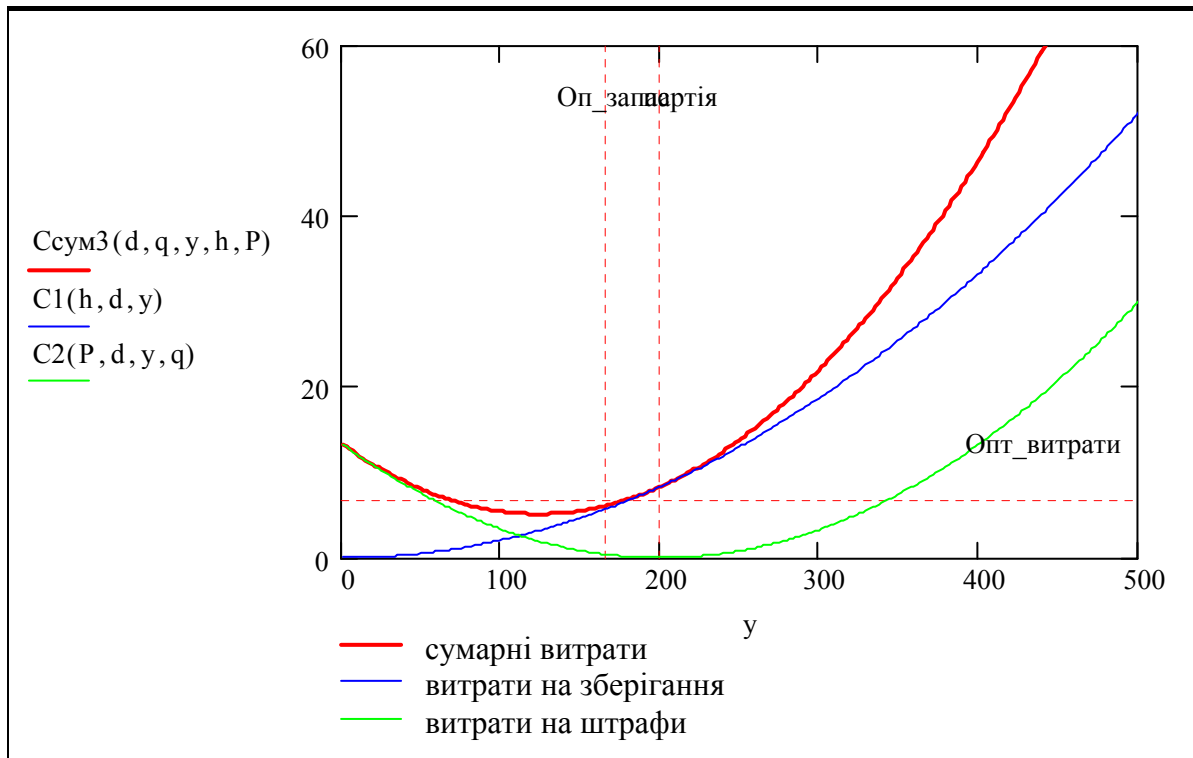


Рис.5. Графіки зміни запасів при виробничих поставках

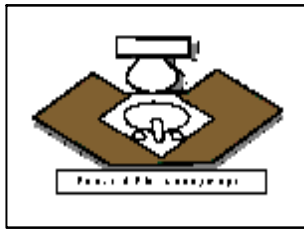
Висновки

Моделі управління запасами, що були розглянуті (модель з моментальними поставками, модель з розподіленими поставками, модель зі штрафами) є простішими, базовими. Вони корисні для розуміння і як перші наближення. Ці моделі допускають узагальнення і розвиток.

В яких напрямках можна узагальнювати і розвивати ці моделі? На практиці інтенсивність попиту не постійна - вона має різні тренди, сезонні коливання, випадкові коливання. Випадкова компонента попиту має певні розподіли імовірностей. І процес постачання теж може бути складним, імовірнісним. Наприклад, ми не враховували оптові знижки, які суттєво змінюють характер оптимізаційної задачі. Ми не враховували зміну цін товарів та послуг (постачання), інфляційних процесів та ін. Реальні системи управління запасами - багатопродуктові. В наступному розділі ми розглянемо більш реалістичну постановку задачі, зроблену Р. Беллманом. Це буде задача нового класу - варіаційна.

Контрольні запитання

1. Сформулюйте задачу управління запасами з моментальними поставками.
2. Сформулюйте задачу управління запасами з розподіленими поставками.
3. Сформулюйте задачу управління запасами зі штрафами.
4. Змінні управління задачі управління запасами зі штрафами.
5. Назвіть компоненти сумарних витрат для задачі з розподіленими поставками.
6. Назвіть компоненти сумарних витрат для задачі зі штрафами.
7. Критерії оптимізації в задачах управління запасами.
8. Методи знаходження екстремуму в задачах управління запасами.



3.4 Розробка узагальнених математичних моделей управління запасами

Вступ. Постановка задачі

В підрозділах 3.1, 3.2, 3.3 цієї теми ми розглянули три класи детермінованих задач управління запасами - для випадків разових, і розподілених поставок, поставок зі штрафами. Для постійного детермінованого попиту ми отримали досить прості аналітичні розв'язання, а сама задача зводилася до знаходження екстремуму функції однієї змінної. Тепер узагальнимо цю задачу - будемо розглядати випадковий попит з довільним розподілом імовірності.

Спочатку трохи історії. Задача управління запасами відноситься до класу варіаційних задач, тому що шукається не один якийсь параметр - розмір партії та ін., а певна функція часу - оптимальна стратегія. Як ми побачимо далі, оптимальне розв'язання в багатьох випадках має розривний характер - "або замовити в оптимальному розмірі, або нічого не замовляти".

Подані далі постановки і розв'язання задач управління запасами були отримані ще 40 років тому Р.Беллманом [3]. Це фундаментальна класика. Саме на базі розгляду таких задач Беллман створив метод динамічного програмування.

Жорсткі обмеження на витрати, обсяги виробництва, терміни не завжди адекватні реальності - не випадково Р. Беллман разом із Л. Заде створили новий науковий напрям теорію розмитих множин та нечітку логіку. Р. Беллман розробив методику розв'язання для цілої низки задач, що поступово ускладнюються, і розв'язання однієї задачі використовується як перше наближення для наступної.

Розглянемо запропоновані Беллманом методи і використаємо їх для побудови програмних модулів оптимізації стратегій управління запасами.

Постановка узагальненої задачі управління запасами

В загальній постановці, задача управління запасами - це окремий випадок проблеми прийняття рішень в умовах невизначеності. Опишемо ситуацію управління запасами. Попит на певні товари є випадковим з певним розподілом імовірностей.

Для задоволення попиту в певні моменти видаються замовлення в розмірах, що розраховуються або задаються незмінними. Створюється певний рівень запасів, але допускається в певні моменти виникнення дефіциту він ліквідується за певними штрафними розцінками.

Що може бути у цій задачі управлінням? З певних причин (подумайте, чому) поповнювати запаси вигідно "імпульсно" партіями певного розміру. Але в якому розмірі і коли видавати замовлення на поставку чергових партій для поповнення запасів? На якому рівні слід підтримувати середні і максимальні запаси за умови, що попит має випадкові коливання і тренди? Чи є якісь закономірності у попиті на певні групи товарів?

Що до закономірностей, то для певних товарів на певних інтервалах часу є такі закономірності - це так звані "ринкові вікна" для товарів з високим темпом морального старіння, наприклад, моделі мобільних телефонів змінюються кожні півроку.

Це означає, що створювати запас мобільних телефонів більше, ніж на 1-2 місяці - нерозумно, адже вони швидко знецінюються. **Нові інформаційні технології** дають можливі засоби виживання в сучасному оточенні: *без статистики, з постійними обвальними змінами*. Це

1) Інтернет – можливість в онлайні отримувати хоч кожную хвилину інформацію про стан ринків, і тут же аналізувати її власною програмою класу "аналітичним помічником менеджера", але чи допоможе це, адже конкуренти будуть мати ту ж зброю;

2) можливості комп'ютерної симуляції - імітаційного моделювання. Ми можемо експериментувати, набирати досвід, працюючи з віртуальною реальністю, створюючи теж онлайн програми для "що буде, якщо аналізу", ризик-аналізу.

Починаючи розробку програмних засобів оптимального управління напишемо програму моделювання типового процесу зміни запасів, де замість управлінь стоять "заглушки" – постійні і випадкові параметри і оцінимо ефективність традиційного управління.

Подаємо приклад такої програми. Вона прозора - не потребує коментарів і працює - в електронній версії – питання можна задавати самій програмі.

PR :=	<pre> ORIGIN ← 1 запас1 ← поч_зап попит1 ← 10 for i ∈ 1..kromod попит_{i+1} ← 8 + 0.1·i + rnd(колив_попиту) партія ← 105 + rnd(37) + i mom ← $\begin{cases} 1 & \text{if } \text{mod}[i, (9 + \text{cei}(\text{rnd}(1)))] = 0 \\ 0 & \text{otherwise} \end{cases}$ запас_{i+1} ← запас_i – попит_i + партія·mom вихід ← stack(попит^T, запас^T) вихід </pre>	<pre> ORIGIN := 1 kromod ≡ 100 i := 1..kromod </pre>
-------	---	--

Зверніть увагу на формування виходу програми: ми транспонували вектори попиту і запасу, а потім об'єднали їх функцією stack(масив1, масив2). Як формувати вихід інакше?

PR =		1	2	3	4	5	6	7	8	9	10
	1	10	8.1	11	14	9.6	8.6	9.9	15	16.4	15.8
	2	100	90	81.9	70.9	56.9	47.3	38.8	28.8	13.9	-2.6

На рис.1 подано графіки попиту та зміни запасів, так щоб відобразити можливі ситуації зміни запасів - зміни попиту, виникнення дефіцитів. Працюючи з електронною версією цього документа можна "пограти в рулетку". Поставте курсор на ці зону введення: $\text{колив_попиту} \equiv 8$ | $\text{поч_зап} \equiv 100$ і натисніть F9 - отримаєте реалізації випадкового процесу. Ви побачите, що графік зміни запасів досить значно змінюється, часто виникають ситуації з великими дефіцитами. На рис.2 подано приклади реалізацій процесів.

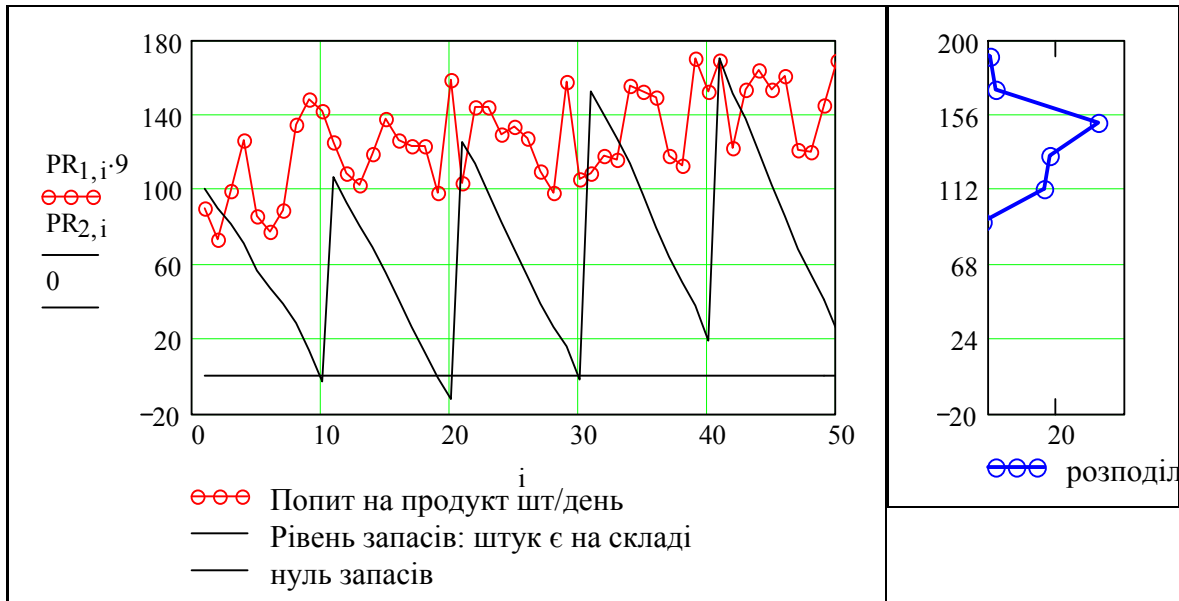


Рис. 1. Приклад процесів зміни попиту та рівня запасів

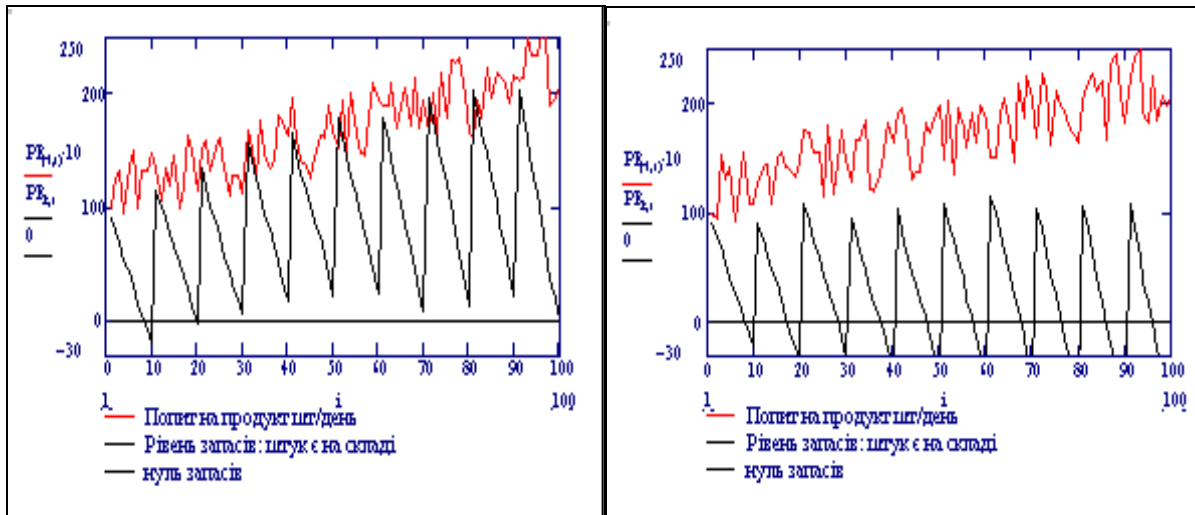


Рис.2. Приклади реалізації процесів зміни запасів при випадкових коливаннях попиту

Очевидно (подивіться на рис.2), що вже не можна розпланувати поставки на рік вперед - треба приймати рішення на кожному кроці процесу відносно значень вибраних змінних управління. Це можуть бути:

1. Розмір партії q ; 2. Моменти видачі замовлення - $T_{зм}[k]$
3. Максимальний або мінімальний, або середній рівень запасів q_{max} , q_{min} , q_c .

Очевидно, що ці змінні пов'язані між собою. В залежності від специфіки задачі певні змінні управління можуть бути фіксованими:

моменти видачі замовлень $T_{зм}[k]$; розмір партій q .

Критерієм управління природно взяти сумарні витрати $C_{сум}$. Складовими витрат можуть бути витрати на організацію поставок C_1 , витрати на закупівлю товарів C_2 , витрати на зберігання запасів C_3 . До витрат на зберігання можуть бути включені витрати від морального старіння виробів, зміни цін, "права випасу" та ін. Розглянемо послідовно задачі Беллмана, на основі яких будується система управління запасами.

Задача управління запасами при кінцевому інтервалі процесу постачання

Подаємо цю задачу **точно в постановці Беллмана**, в оригінальних означеннях. Ми хочемо показати, що справжня наука не застаріває. Вважаємо, що нам відомі такі функції:

- а) $\phi(s) \cdot ds$ - імовірність того, що попит буде знаходитись в інтервалі $s \leq \text{demand} \leq s + ds$;
- б) $k(z)$ - вартість номінального замовлення розміром z на початку циклу поставання;
- в) $p(z)$ - вартість "штрафного" замовлення розміром z , зробленого у кінці циклу, якщо виявилось перевищення попиту над запасами на складі (див. рис.3).

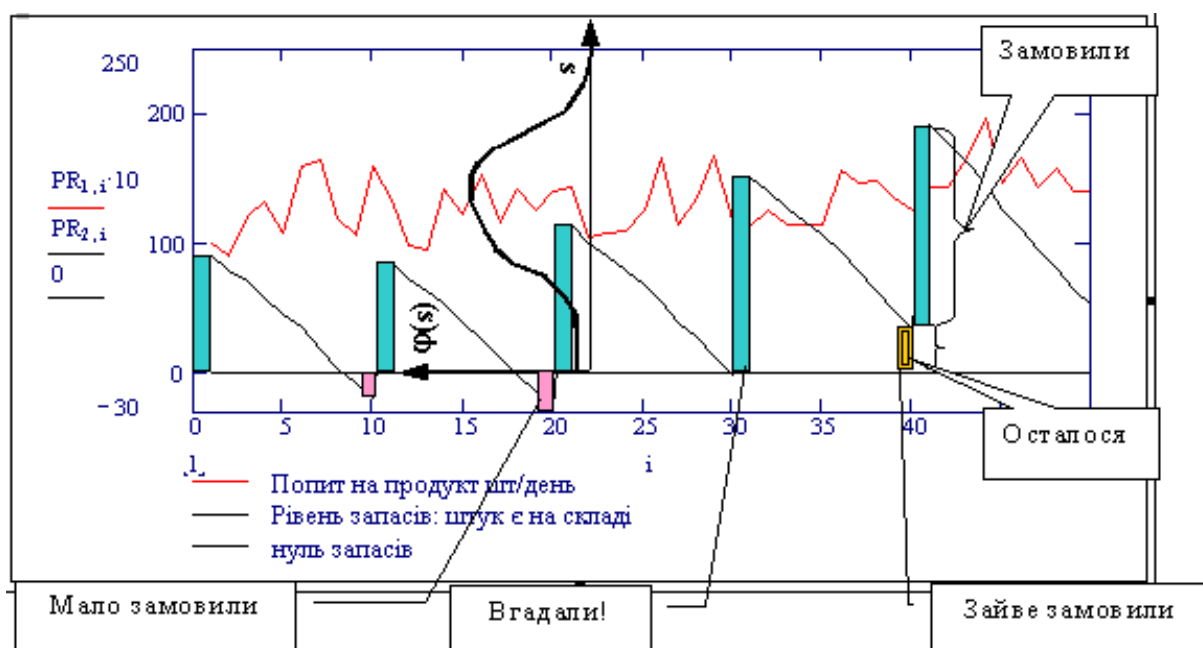


Рис.3. Управління запасами при невизначеному попиті

Позначимо через X рівень запасів на початку циклу (рис.3) забезпечення попиту. Один цикл складається з таких подій:

- **на початку циклу** менеджер виходячи із досвіду, чи розрахунків визначає бажаний рівень запасів **y**, потім замовляє $Z = y - X$ одиниць товару, де X - залишок від попереднього циклу, якщо **y < x**, то нічого не замовляється;
- **потім протягом циклу** запас витрачається відповідно до попиту;
- **в кінці циклу**, якщо виявляється дефіцит, то попит задовольняється за штрафною вартістю (це може мати різні інтерпретації - термінове замовлення, премія покупцю за те, що він почекає...), якщо залишок - то він переходить у наступний цикл. Припустимо, що процес складається з n етапів (циклів). На першому створюється запас **y1** одиниць товару, на другому - **y2** і т.д.

Послідовність замовлень - вектор $(y_1, y_2, \dots, y_n)$ назвемо **стратегією управління**. Кожній стратегії відповідає певна сумарна очікувана вартість, що складається з вартості номінального замовлення та вартості штрафного замовлення, якщо воно потрібне.

Поставимо задачу визначення стратегії, що мінімізує очікувану (середню) сумарну вартість $C_{смс}$.

Друга більш важлива задача - мінімізація імовірності того, що вартість перейде деякий заданий рівень. Це фактично задача **мінімізації витрат при заданій надійності забезпечення попиту**. Наша задача може мати і іншу інтерпретацію - технічну: штрафні поставки - це ремонт після відмови. З математичної точки зору рівняння задачі управління запасами відносяться до класу задач "відновлення" (восстановления).

Стратегія, що веде до мінімуму очікуваної сумарної вартості, називається оптимальною.

На будь-якому етапі результат оптимізації - стратегія - повністю визначається двома змінними - початковим запасом x та кількістю етапів до кінця процесу n . Це відомий **принцип оптимальності** Беллмана, який базується на відсутності "післядії". Визначимо цю функцію (це і є **Функція Беллмана**), на якій базується метод динамічного програмування (правильна назва: "метод динамічного планування"):

$f(n, x)$ - очікувана сумарна вартість для n -етапного процесу з початковим рівнем запасу x при використанні оптимальної стратегії видачі замовлень. Подамо стисло процес отримання функціонального рівняння для визначення функції $f(n, x)$.

Зауваження. У Беллмана ця функція записана як $f_n(x)$, позначення $f(n, x)$ може бути некоректним, якщо для різних n функції $f_n(x)$ належатимуть до різних структурних класів. Але в нашому випадку це допустиме. Якщо на першому кроці замовлено $z = y - x \geq 0$, то очікувані витрати будуть:

$$\text{Сум_витрати}(1, x) = k(y - x) + \int_y^{\infty} p(s - y) \cdot \phi(s) ds \quad (1)$$

Зауваження. Звідки взявся інтеграл у формулі витрат і що він означає? - Перегляньте розділ 1, особливо, підрозділ 1.3 – "Міри групування даних – частотні розподіли". Інтеграл дає оцінку штрафних витрат. Дійсно - ми розглядаємо усі можливі випадки попиту, що перевищує запас - від значення y до "нескінченності". Кожне значення можливого штрафу $p(s - y)$ ($p = \text{penalty}$) множимо на імовірність відповідного попиту – $\phi(s)$, складаємо – отримуємо оцінку середнього штрафу. Тоді функція мінімальних витрат визначиться так:

$$f(1, x) = \min \left(k(y - x) + \int_y^{\infty} p(s - y) \cdot \phi(s) ds \right); \quad y \geq x. \quad (2)$$

Для процесів з числом кроків 2 і більше маємо рекурентну залежність:

$$f(n, x) = \min \left[k(y - x) + \int_y^{\infty} p(s - y) \cdot \phi(s) ds + f[(n - 1), 0] \cdot \int_y^{\infty} \phi(s) ds \dots \right. \\ \left. + \int_0^y f[(n - 1), (y - s)] \cdot \phi(s) ds \right]; \quad y \geq x. \quad (3)$$

Задача управління запасами при необмеженому інтервалі часу

Для того щоб розглядати необмежені інтервали функціонування системи слід ввести показник дисконтування грошових потоків (знецінення майбутніх витрат і прибутків) - a . Цей показник має економічне обґрунтування і просто необхідний для отримання математично коректної задачі - без нескінченних витрат.

$$f(x) = \min \left[\begin{aligned} &k(y-x) + a \cdot \int_y^\infty p(s-y) \cdot \phi(s) ds + a \cdot f(0) \cdot \int_y^\infty \phi(s) ds \dots \\ &+ \int_0^y f(y-s) \cdot \phi(s) ds \end{aligned} \right]; \quad y \geq x. \quad (4)$$

Рівняння (4) більш просте, ніж (3) в ньому треба визначити тільки одну функцію $f(x)$ а не цілу послідовність з n функцій.

Проаналізуємо те, що ми отримали. Ми отримали функціональне рівняння, - певну залежність для невідомої функції $f(x)$. Ми часто працюємо з певним класом функціональних рівнянь - диференціальними і не замислюємось, як же ми їх розв'язуємо. Лінійні диференціальні рівняння розв'язуються на базі підстановки Ейлера, згадаємо: "розв'язання шукаємо у вигляді $x(t) = C \cdot e^{a \cdot t}$ ". Метод розв'язання довільних диференціальних рівнянь в математичному фольклорі подається так: "дивитись на диференціальне рівняння, поки його розв'язання не прийде в голову".

Беллман додав своє до цього фольклору: "Тепер ми будемо піднімати самі себе за волосся!" [Беллман Р., Гликсберг И., Гросс О. Некоторые вопросы математической теории управления. - М.:, Издат. Иностран. Литер., 1962. -233 с.].

Але, коли для певного класу задач знайдено вигляд функції оптимального управління, то розв'язання їх стає простим, формалізованим, загальнодоступним - ну як розв'язання лінійних диференціальних рівнянь. Розглянемо випадки, коли явний вигляд розв'язання можна знайти.

Задача управління запасами при пропорційній вартості витрат

Припущення про те, що витрати на замовлення партії прямо пропорційні обсягу партії, а штрафні витрати - пропорційні обсягу штрафного замовлення, є досить реалістичними - потім ми додамо організаційні витрати на партію, оптові знижки, розглянемо нелінійні функції вартості.

Прийняті нами припущення суттєво спрощують задачу. Властивості розв'язання визначаються теоремою, що **подається точно, як у Р. Беллмана**.

Теорема. Розглянемо рівняння

$$f(x) = \min \left[\begin{aligned} &k \cdot (y-x) + a \cdot \int_y^\infty p \cdot (s-y) \cdot \phi(s) ds + a \cdot f(0) \cdot \int_y^\infty \phi(s) ds \dots \\ &+ \int_0^y f(y-s) \cdot \phi(s) ds \end{aligned} \right]; \quad y \geq x \quad (5)$$

Припускаємо, що виконуються такі умови:

(а) k та p - позитивні константи;

(б) $\phi(s) > 0$; $\int_0^\infty \phi(s) ds = 1$; $\int_0^\infty s \cdot \phi(s) ds < \infty$ - розподіл імовірностей має кінцеву дисперсію;

(в) $0 < a < 1$ - грошові потоки дисконтуються - приводяться до одного часового горизонту;

(г) $a \cdot p > k$ - навіть дисконтована штрафна вартість поставок повинна бути більша номінальної вартості - інакше вигідно все поставляти в штрафному режимі. Нехай $h_{ек}$ - єдиний корінь рівняння:

$$k = a \cdot p \cdot \int_y^\infty \phi(s) ds + a \cdot k \cdot \int_0^y \phi(s) ds \quad (6)$$

Тоді оптимальна стратегія має вигляд:

$$(a) \quad y = h_{ек} - \text{при } 0 \leq x \leq h_{ек} \quad (7)$$

$$(б) \quad y = x - \text{при } x > h_{ек}$$

інакше кажучи, оптимальний рівень запасів дорівнює $h_{ек}$ (x - Єдиний Корінь рівняння (б)). Якщо $a \cdot p \leq k$ то розв'язок при $x \geq 0$ має вигляд $y = x$, тобто в цьому випадку нічого не треба замовляти.

Доведення теореми. Наводимо доведення так, як його подано Беллманом це приклад методики пошуку розв'язань функціональних рівнянь. З перших рук! Використовується евристичний (інтелектуальний) підхід: звідкись приходить (у голову) правдоподібне розв'язання, ми його перевіряємо підстановкою в (5), якщо воно задовольняє, то за теоремою про єдине, це і буде оптимальне розв'язання. Воно згідно з (5) повинно задовольняти необхідну умову екстремуму (похідна дорівнює нулю).

$$\frac{d}{dy} \left[k \cdot (y - x) + a \cdot \int_y^\infty p \cdot (s - y) \cdot \phi(s) ds + a \cdot f(0) \cdot \int_y^\infty \phi(s) ds + \int_0^y f(y - s) \cdot \phi(s) ds \right] = 0.$$

Визначаємо складові похідної

$$\frac{d}{dy} k \cdot (y - x) \rightarrow k \quad ; \quad \frac{d}{dy} a \cdot \int_y^\infty p \cdot (s - y) \cdot \phi(s) ds \rightarrow a \cdot \int_y^\infty -p \cdot \phi(s) ds.$$

А як же брати похідну від ще невідомої функції? Згідно з декларованою стратегією управління запасами, функція Белмана для цієї задачі повинна мати вигляд $f(x) = -k \cdot x$, тому відповідні похідні будуть такими:

$$\frac{d}{dy} \int_0^y f(y - s) \cdot \phi(s) ds = a \cdot k \cdot \int_0^y \phi(s) ds \quad ; \quad \frac{d}{dy} a \cdot f(0) \cdot \int_y^\infty \phi(s) ds = 0.$$

Враховуємо, що (сума імовірностей повної групи подій) $\int_0^{\infty} \phi(s) ds = 1$, тоді рівняння (6) приводиться до вигляду

$$\int_y^{\infty} \phi(s) ds = \frac{a \cdot p - k}{a(p - k)} \quad (8)$$

Неважко бачити, що рівняння (8) має тільки один корінь - теорему доведено.

Розробка робочих програм та інтерфейсу для системи управління запасами

Вступ. Постановка задачі

Ми вибрали і обґрунтували методи оптимального управління запасами в умовах невизначеності попиту. Теоретична основа розробки - методи і методологія розроблені Беллманом - принцип оптимальності, метод функціональних рівнянь, метод функціональної апроксимації та їх синтез - метод динамічного програмування. За рахунок введення дисконтування для необмеженого інтервалу часу функціонування системи ми отримали відносно просте функціональне рівняння з однією невідомою функцією - оптимальною стратегією.

Попередньою гарантією ефективності вибраної нами методології є той факт, що Беллман постійно підкреслював непродуктивність і неконструктивність методу динамічного програмування. Наприклад, "динамічне програмування не метод, а стан розуму", "Нарешті, ми вкажемо вигляд загального розв'язання, хоч не матимемо з цього особливої користі", "розглянемо властивості цього розв'язання у всій його неприкритій простоті".

В цьому розділі подано розробку базової версії автоматизованої системи - помічника менеджера постачання. Автоматизована система для довільних законів розподілу імовірностей попиту, вартостей планового і термінового - штрафного постачання - повинна розраховувати параметри оптимального процесу забезпечення попиту (розмір партії, середній рівень запасів та ін.). Крім цього система повинна забезпечити:

- а) можливість проведення обчислювального експерименту;
- б) можливість візуального контролю кожного елемента моделей і програмних модулів;
- в) відкритість для модифікації - користувач може все (майже) бачити, змінювати і сам може вибирати, які фрагменти системи слід закрити паролями від змін і переглядання.

До комплекту даної системи повинен входити модуль ідентифікації частотних розподілів, приклад якого подано в розділі 1.4.

Розробка модуля для визначення оптимального рівня запасів

Будуємо модуль для розв'язання рівняння (8) при довільних частотних розподілах попиту. Вважаємо, що вигляд і параметри частотного розподілу ідентифіковані (можна використати і просто емпіричний частотний розподіл, але тоді буде закрита можливість для аналітичних досліджень). Створюємо зону введення параметрів (в рамці – копія екрану), задаємо їх значення (для "контрольного прикладу"). Задаємо (вибираємо) тип розподілу.

Назва розподілу $N := \text{"нормальний"}; \text{Середнє } \mu := 12; \text{ дисперсія } \sigma := 3;$
 норма дисконту $\rho := 0.1; a := 1 - \rho; a = 0.9$
 вартості: звичайної поставки $k := 5; \text{ штрафної поставки } p := 10.$
 Контроль рівня штрафу $a \cdot p > k = 1$
 Вибраний розподіл $\phi(x, \sigma) := \text{dnorm}(x, \mu, \sigma)$ $x := 0..30; y := 0..30.$

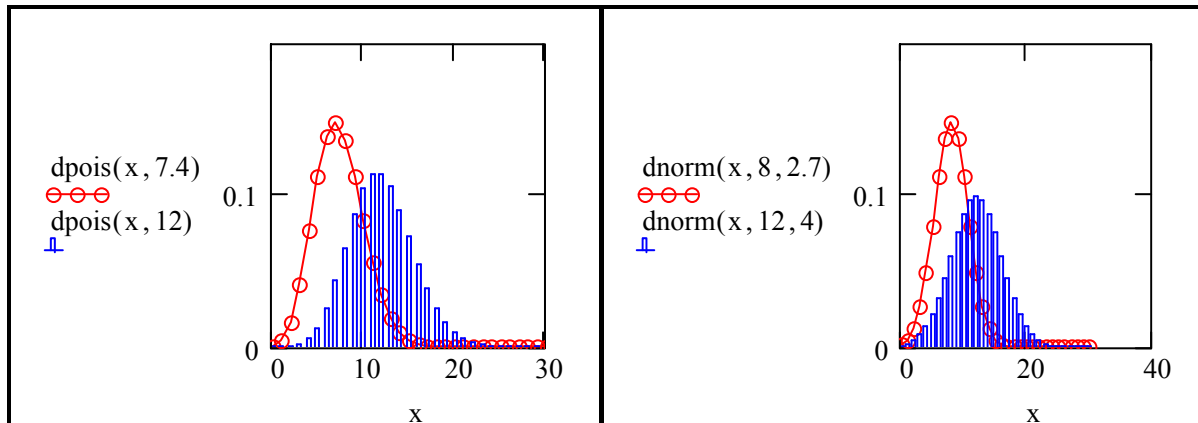


Рис. 4 Приклади частотних розподілів - альтернативних гіпотез

Зауваження. В реальній системі модуль ідентифікації частотних розподілів попиту чи цін може працювати на онлайнних даних, що надходять по мережах. Технічних проблем тут нема, є проблеми організаційно-фінансові: наявність джерел даних, можливість підписки та оплати. Виконаємо алгебраїчні перетворення рівняння (8) і побудуємо графіки - для аналізу і графічного розв'язання рівняння.

$$k = a \cdot p \cdot \int_y^\infty \phi(s, \sigma) ds + a \cdot k \cdot \int_0^y \phi(s, \sigma) ds; \quad \frac{1}{a} - \int_0^y \phi(s, \sigma) ds = \frac{p}{k} \cdot \int_y^\infty \phi(s, \sigma) ds;$$

$$F1(y, \sigma) := \frac{1}{a} - \int_0^y \phi(s, \sigma) ds; F2(y, \sigma) := \frac{p}{k} \cdot \int_y^\infty \phi(s, \sigma) ds; F1(y, \sigma) - F2(y, \sigma) = 0.$$

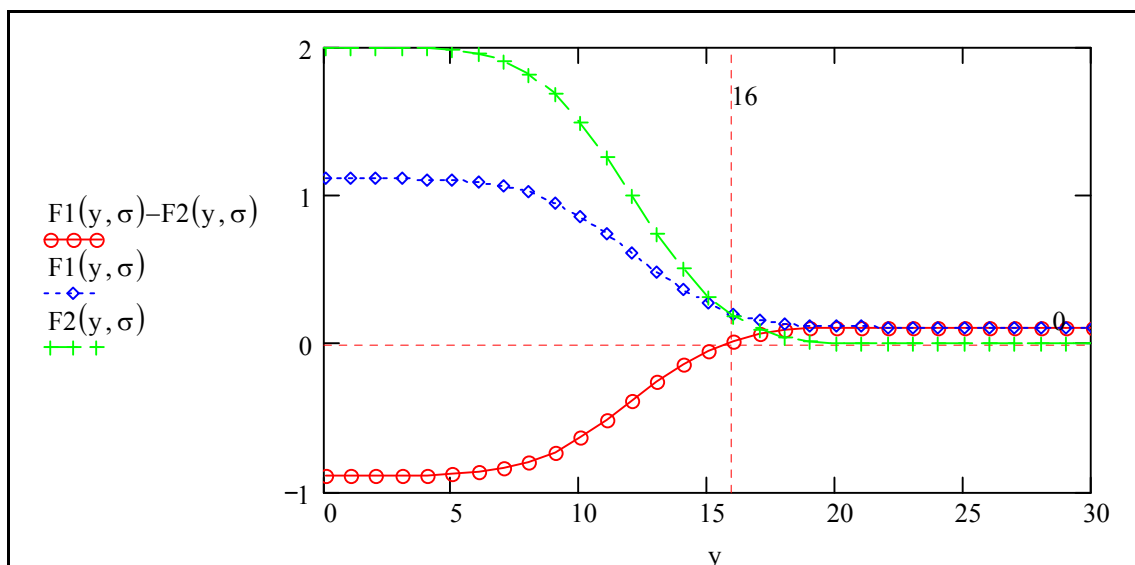


Рис. 5. До пошуку нульового кореня рівняння (8)

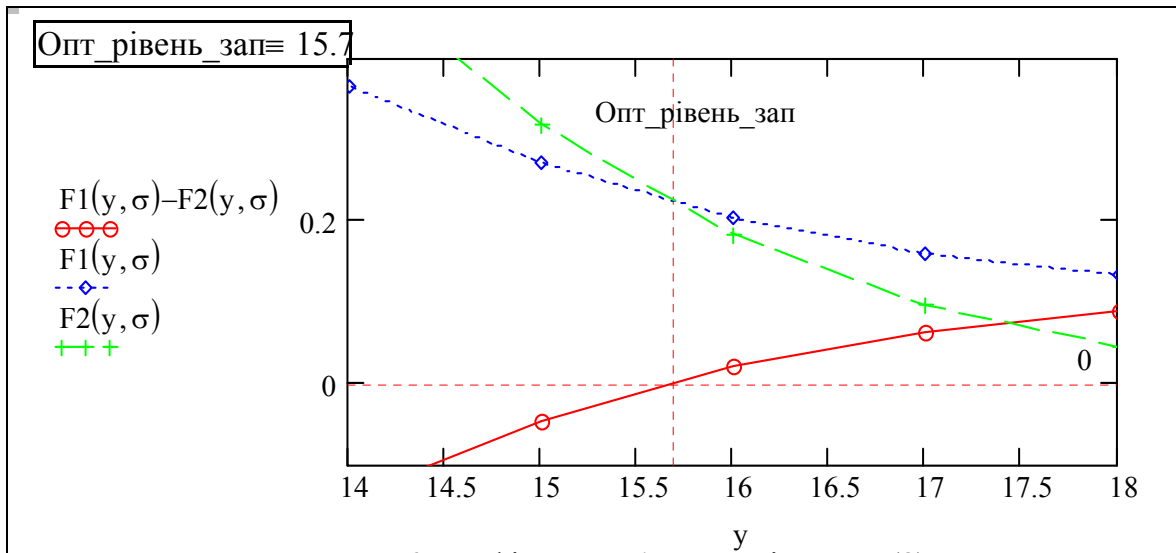


Рис. 6. Графічне розв'язання рівняння (8)

Знайдемо корінь рівняння (8) вбудованим числовим пошуковим методом. Задаємо стартове (для пошуку) значення $y_p := 12$. Після ключового слова "Дано" пишемо систему рівнянь

Given

$$\frac{1}{a} - \int_0^{y_p} \phi(s, \sigma) ds = \frac{p}{k} \cdot \int_{y_p}^{\infty} \phi(s, \sigma) ds; \quad y_p \geq 0. \quad (9)$$

Увага! Результат розв'язання формуємо як функцію параметрів σ , p , k . З точки зору традиційної математики те, що написано нижче - рівність функцій від різних змінних. Якщо ці змінні незалежні, то ці функції повинні дорівнювати константі. А в даному випадку це визначення функції користувача в термінах розв'язального блока "Given-Find". Це типова конструкція мови C++. Інтерпретатору і компілятору це зрозуміло.

$$\boxed{Xek(\sigma, p, k) := Find(y_p)};$$

$$\sigma p := 0.2, 0.7..3.2$$

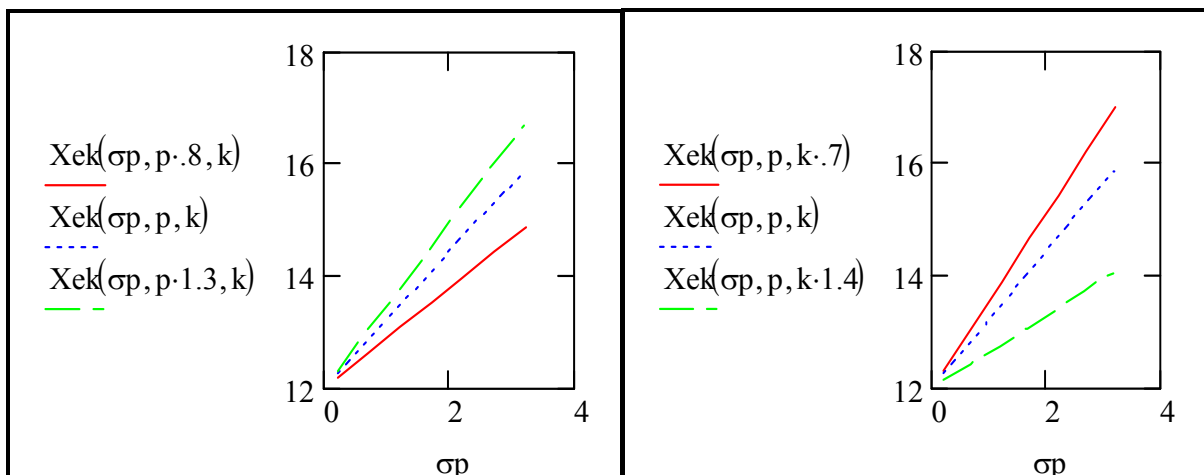


Рис. 7. Залежності оптимального середнього рівня запасів від дисперсії та ціни замовлень

Бачимо, що залежності для оптимального рівня запасів дуже прості - так і повинно бути, тому що функції витрат - **лінійні**. Легко узагальнити задачу на випадок нелінійних - випуклих функцій витрат.

Виводимо для порівняння графічне і числове розв'язання рівняння (8).

$$\text{Опт_рівень_зап} = 15.7 ; \quad \text{Хек}(\sigma, p, k) = 15.661 .$$

Бачимо, що $\sigma + \text{mu} = 15$ є гарною оцінкою оптимального рівня запасів. Чому?

Розробка модуля імітації довільних функцій попиту

Потрібну для обчислювального експерименту імітацію функцій попиту з довільними розподілами робимо на базі вбудованих генераторів випадкових чисел. Незавжди за аналогією побудувати процеси з нестационарними характеристиками, трендами та ін.

$\text{mu} = 12 ; \quad \sigma = 3 ; \quad \text{Nкроків_процесу} := 100 ; \quad i := 1 .. \text{Nкроків_процесу} ;$

$\text{Попит} := \text{rnorm}(\text{Nкроків_процесу}, \text{mu}, \sigma) ;$

$\text{Pop2} := \text{rnorm}(\text{Nкроків_процесу}, \text{mu}, \sigma) ; \quad \text{Pop3} := \text{rnorm}(\text{Nкроків_процесу}, 2\text{mu}, 1) .$

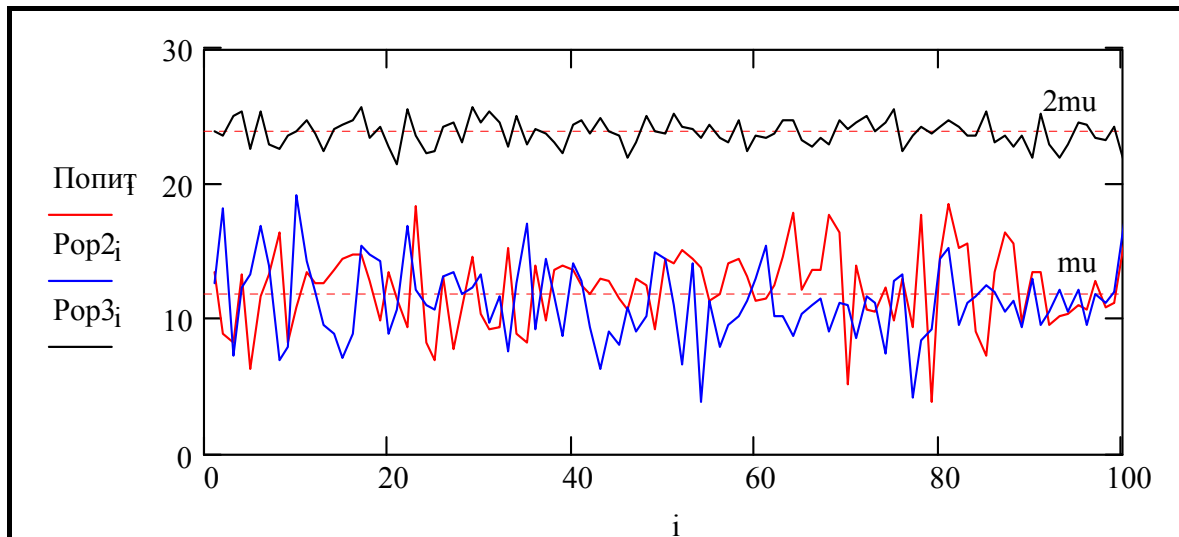


Рис. 8. Приклади реалізацій функції попиту

Розробка модуля імітації процесів постачання

Напишемо модуль за лінійною логікою, але останню частину - стенд зробимо за об'єктною: спочатку подамо результати моделювання, а за ними - введення вхідних даних. Подаємо специфікацію вхідних даних:

Номинальна вартість замовлення одиниці товару $\text{km} = 5 ;$

Штрафна вартість замовлення одиниці товару $\text{pm} = 6 ;$

Норма дисконтування грошових потоків $\text{am} = 0.9 ;$

Запас на початку періоду одиниць $\text{X0} = 50 ;$ Кількість замовлень $\text{nzak} = 20 ;$

Розподіл імовірностей рівнів попиту $\boxed{\text{fi}(x, \text{mu}, \sigma) := \text{dnorm}(x, \text{mu}, \sigma)}$.

Визначаємо (копіюємо і модифікуємо (9)) в термінах розв'язального блоку функцію: Хорт(середнє , розкид , штрафна_вартість , номінальна_вартість) .

Ця функція визначає оптимальний рівень запасів.

$$y_p := 21 ; \text{ Given } \frac{1}{a_m} - \int_0^{y_p} f(s, m, \sigma) ds = \frac{p_m}{k_m} \cdot \int_{y_p}^{\infty} f(s, m, \sigma) ds ; \quad y_p \geq 0 .$$

$$X_{opt}(m, \sigma, p, k) := \text{Find}(y_p) ; \quad X_{opt}(m, \sigma, p, k) = 11.455$$

Тепер пишемо програму (копіюємо та модифікуємо модуль з розділу 3.3).

OPk =	запас₁ ← X₀ "Визначаємо інтервал між партіями (оптимізація - в розд.2)" $\text{interval} \leftarrow \text{ceil} \left(\frac{\text{Nкроків_процесу}}{\text{nzak}} \right)$ "Обчислюємо імітацію попиту на всі кроки процесу" попит ← norm(Nкроків_процесу + 1, m, σ) "Визначаємо оптимальний розмір запасів" Opzap ← X_{opt}(m, σ, p, k) "Визначаємо розмір партії" Opzak ← Opzap · interval "Цикл по крокам процесу (наприклад - дням)" for i ∈ 1 .. Nкроків_процесу "Логіка оптимальної стратегії згідно з (8) з розділу 3" $\text{партія} \leftarrow \begin{cases} (\text{Opzak} - \text{запас}_i) & \text{if } \text{запас}_i < \text{Opzak} \\ 0 & \text{if } \text{запас}_i \geq \text{Opzak} \end{cases}$ "Логіка видачі (і отримання) партій товару" $\text{momento} \leftarrow \begin{cases} 1 & \text{if } \text{mod}(i, \text{interval}) = 0 \\ 0 & \text{otherwise} \end{cases}$ "Обчислення поточного запасу" $\text{запас}_{i+1} \leftarrow \text{запас}_i - \text{попит}_i + \text{партія} \cdot \text{momento}$ "Формування вихідного масиву" $\text{вихід} \leftarrow \text{stack}(\text{попит}^T, \text{запас}^T)$ вихід
-------	---

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
OPk =	1	13	10	7	17	8	9	24	17	21	18	21	15	23	7	19	17	20
	2	50	37	27	19	2	63	54	30	13	-8	65	44	29	5	-2	68	52

Подамо короткий опис програми моделювання.

1. Програма використовує як підпрограму функцію $X_{opt}(\mu, \sigma, p, k)$.
2. Програма використовує певні глобальні змінні документа - вони вводяться далі - на стенді в режимі глобального присвоєння.
3. У програмі визначені певні локальні змінні - вони діють тільки в цій програмі і "невидимі" у головній програмі - цьому документі.
4. Програма повертає у головну програму (цей документ) масив розміром $2 \times N$ - значення поточних попиту та рівня запасів.

Вихід програми можна розширити - виводити значення показників якості управління запасами (сумарних витрат, штрафів та ін.)

Те, що подано далі - стенд: усі виходи і входи зібрані на одній екранній сторінці.

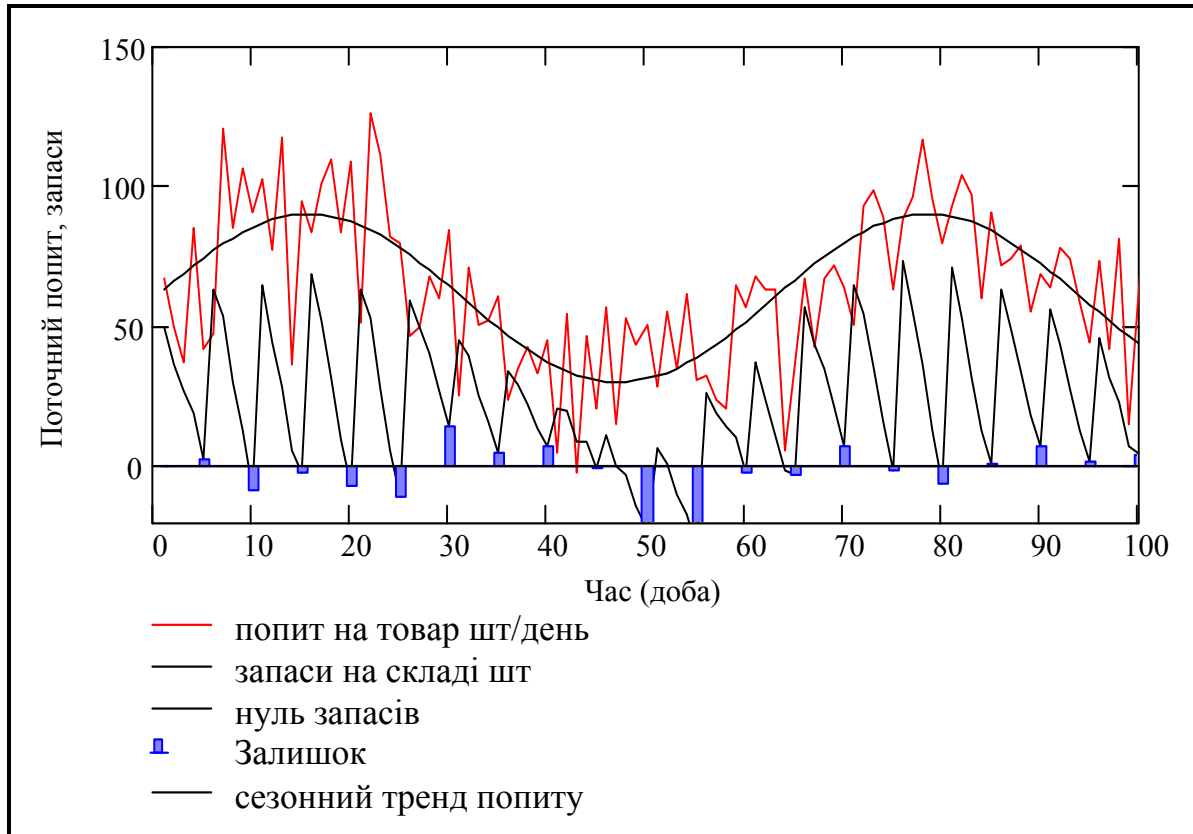


Рис. 9. Реалізація оптимального процесу постачання (управління запасами)

Сезонний тренд попиту	$AmplTr \equiv 0.5$	Розподіл попиту	типР = "гаусовий"
Середній попит	$\mu m \equiv 12$	Розкид попиту	$\sigma m \equiv 3.7$
Нормальна вартість замовлення одиниці	$km \equiv 5.0$	Норма дисконтування	$am \equiv 0.9$
Штрафна вартість замовлення одиниці	$pm \equiv 6.0$		
Запас початковий	$X0 \equiv 50$ тон.	Кількість замовлень за період	$nzak \equiv 20$
Оптимальний рівень запасів*	$Опр3 = 22$ тон.	Витрати	$Витр = 12300$ уо
Запас початковий	$X0 \equiv 50$	Середнє значення попиту	$\mu m = 12$
Оптимальний середній рівень запасів	$X_{opt}(\mu m, \sigma m, pm, km) = 11.455$		
"Плата" за мінімізацію ризику	$\frac{X_{opt}(\mu m, \sigma m, pm, km)}{\mu m} = 0.955$		

Завдання. Спробуйте змінювати рівень штрафу за недопостачання - подивіться, як змінюється оптимальний середній рівень запасів - абсолютний і відносно середнього рівня попиту. Може тоді ви побачите "в неприхованій простоті" результат, до якого може дійти менеджер і без високої математики: **середній запас** при майже детермінованому попиті повинен дорівнювати **середньому попиту**. При високих штрафах рівень запасів зростає, при малих штрафах - навпаки. Таким чином, тепер, коли ми побудували аналітичну оптимальну систему, можна побудувати і пошукову систему оптимізації.

Увага! Вище подано текст програми для випадку постійного середнього попиту. Графіки побудовані за модифікованою програмою (вона - для самостійного виконання), що враховує і тренд. Зокрема, імітується (синусоїдою) сезонний тренд попиту.

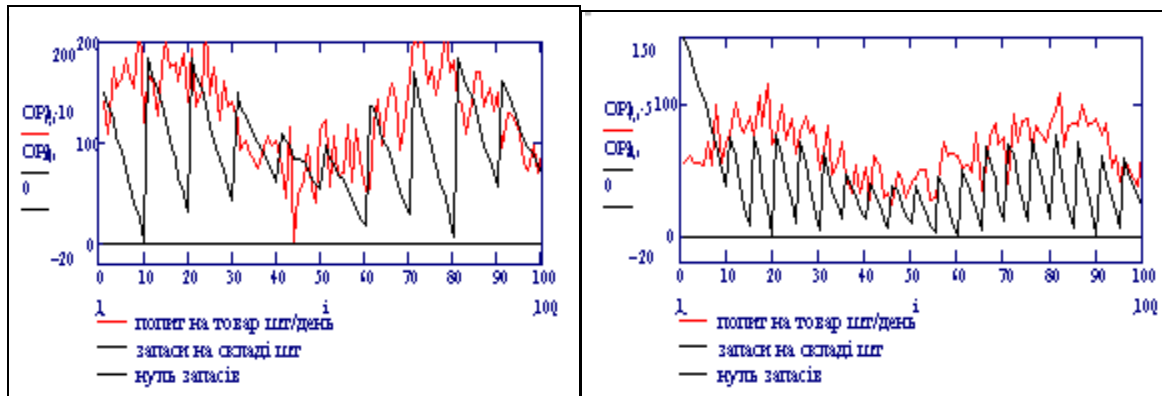


Рис. 10. Реалізації оптимального процесу управління запасами для випадку великого сезонного тренду (10 і 20 партій за період)

Проведіть дослідження системи для таких граничних ситуацій:

- 1) велика невизначеність - великі штрафи;
- 2) велика невизначеність - малі штрафи;
- 3) мала невизначеність - великі штрафи;
- 4) мала невизначеність - малі штрафи.

Висновки

Розглянуто базову модель управління запасами в умовах невизначеності. Для оптимізації вибрано метод динамічного програмування. Отримано програмну систему для розрахунку оптимального за критерієм сумарних витрат управління запасами і моделювання процесів постачання (забезпечення попиту).

Проведено випробування системи, що підтвердило працездатність програмної системи. Проведено також випробування на модифікацію і розширення функцій. Отримана версія з розрахунком оптимального рівня запасів на кожному циклі постачання.

4

Оптимізація планів виробництва та перевезень

У цьому розділі розглядаються:

- ✓ Розробка базової математичної моделі для оптимізації транспортної задачі
- ✓ Розробка інтерфейсу користувача з підсистемою оптимізації планів перевезень
- ✓ Масштабування програми оптимізації планів перевезень. "Що-буде-якщо" аналіз
- ✓ Розробка системи для розв'язання задач лінійного програмування альтернативними методами
- ✓ Дискретні задачі оптимізації при обмеженнях. Задача розподілу виробничих ресурсів в часі

Після вивчення цього розділу ви **зможете:**

1. Формалізувати реальні задачі, як задачі лінійного програмування – виділити змінні, записати обмеження і цільову функцію (критерій).
2. Записувати задачі лінійного програмування у скалярному та матрично-векторному вигляді.
3. Записувати задачу, що є двоїстою до заданої задачі лінійного програмування (ЛП).
4. Використовувати вбудовані функції пакетів для розв'язання задач ЛП.
5. Будувати програмні інтерфейси для „що буде, якщо” аналізу транспортних та виробничих задач.
6. Переводити задачу ЛП у нечіткий (розмитий) формат – конструювати функції належності для цілі, обмежень і загальну функцію належності. Отримувати розв'язання „розмитої задачі”.

7. Дати визначення термінам:

- | | |
|--|-------------------------------------|
| 1) лінійне програмування (ЛП) | 6) двоїста (спряжена) задача ЛП; |
| 2) цільова функція, критерій, обмеження; | 7) комбінування функцій належності; |
| 3) функція належності обмеження; | 8) оптимальний план перевезень; |
| 4) виробнича задача ЛП; | 9) оптимальний план виробництва; |
| 5) транспортна задача ЛП; | 10) оптимальна оцінка ресурсів. |

4.1. Розробка базової математичної моделі для оптимізації транспортної задачі

Вступ. Постановка задачі

Задачі лінійного програмування (ЛП) складали і досі складають значну частину "набору інструментів" інженера і менеджера. ЛП - не просто класика, а ціла епоха. В цьому розділі виконується аналіз і певна ревізія типових задач лінійного програмування (ЛП) в таких напрямках: а) використання "нових інформаційних технологій", застосування нових методів і навіть нових математичних моделей реальних задач.

Почнемо з так званої транспортної задачі і розглянемо класичну - базову математичну моделі транспортної задачі. В базовій моделі витрати на перевезення вважаємо лінійними функціями відстаней і об'ємів перевезень (= тонокілометрів). Цільову функцію - сумарні витрати на перевезення - вважаємо теж лінійною функцією об'єму перевезень.

Базова модель не враховує такі фактори, як

- а) випадкові коливання потреби в товарах;
- б) нелінійні функції витрат на перевезення;
- в) з урахуванням затримок на перевезення, штрафів за затримки та ін.

Відомо, що для розв'язання деякої оптимізаційної задачі можуть використовуватись різні методи оптимізації. Далі для розв'язання задачі лінійного програмування (ЛП) буде застосовано метод оптимізації на базі розмитої логіки.

Ми побачимо, що цей метод:

- а) змінює саму суть задачі і приводить до інших результатів, більш коректних з точки зору техніко-економічних реалій;
- б) узагальнює задачу ЛП настільки, що фактично закриває ЛП як прикладний напрямок;
- в) узагальнює задачу багатокритеріальної оптимізації - критерії і обмеження стають рівноправними.

Але, згідно з законом проблемології, розв'язання однієї проблеми породжує не менше двох нових (закон розмноження проблем). Коротше кажучи, за все треба платити. Саме про це попереджали засновники нового напрямку - Р. Беллман та Л. Заде - теорія розмитих множин і розмитої логіки ставить занадто важкі інтелектуальні задачі.

Класична транспортна задача

Лінгвістична модель

Розподілена система постачання складається з N баз (складів, фабрик...) та M "магазинів" (пунктів обслуговування).

Задано: потреби пунктів обслуговування V_j , $j = 1..M$; можливості баз A_i , $i = 1..N$; вартості перевезення одиниці вимірювання вантажу від кожної бази до кожного магазину C_{ij} . **Потрібно** визначити план перевезень, що дає мінімум сумарних витрат.

Математична модель

Зауваження. Далі використовуються два типи змінних пакету, наприклад: A_i - запаси продукту на i -й базі, змінна, що має присвоєне значення; As_i - та ж сама змінна, що НЕ має присвоєного значення, вона використовується у символічних обчисленнях.

Беремо для прикладу $N := 3$; $M := 4$; $i := 1..N$; $j := 1..M$. Виконуємо такі присвоєння: $X_{i,j} := X_{i,j}$; $C_{i,j} := C_{i,j}$; $A_i := A_i$; $A := A^T$; $B_j := B_j$; $B := B^T$. Так ми задаємо змінні без значення, (це використовується в мові C++).

При символічних обчисленнях символічний процесор підставляє значення змінних, а потрібні саме символічні вирази. У виразах, що подані далі, вирази після стрілки написав комп'ютер. Ця дрібничка - елемент "нових інформаційних технологій".

Дано: запаси на базах $A \rightarrow (A_1 \ A_2 \ A_3)$

потреби магазинів $B \rightarrow (B_1 \ B_2 \ B_3 \ B_4)$

ціни перевезень $C \rightarrow \begin{pmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} \\ C_{2,1} & C_{2,2} & C_{2,3} & C_{2,4} \\ C_{3,1} & C_{3,2} & C_{3,3} & C_{3,4} \end{pmatrix}$

Знайти: план перевезень $X \rightarrow \begin{pmatrix} X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} \\ X_{2,1} & X_{2,2} & X_{2,3} & X_{2,4} \\ X_{3,1} & X_{3,2} & X_{3,3} & X_{3,4} \end{pmatrix}$, такий, що дає мінімум

сумарних витрат $C_{\text{opt}} = \min \left(\sum_{i=1}^N \sum_{j=1}^M C_{i,j} \cdot X_{i,j} \right)$ при обмеженнях $\sum_i X_{i,j} = B_j$ -

потреб магазинів та $\sum_j X_{i,j} = A_i$ - запасів на базах.

Дивлячись на математичну модель, де ми використали векторно-матричні позначення, можемо конкретизувати словесну модель.

Знайти план перевезень - таку матрицю X (кожен елемент якої X_{ij} задає кількість продукції, яка доставляється з i -ї бази в j -тий магазин), що забезпечує **мінімум** цільової функції C - вартості усіх перевезень при **$n + m$ обмеженнях** на потреби магазинів та поставки з баз.

Більшість сучасних пакетів для техніко-економічних розрахунків має вбудовані методи розв'язання задач ЛП. Від користувача вимагається тільки коректно записати умови задачі, краще у розгорнутому вигляді - для контролю. Зробимо це для нашого прикладу. В цілому маємо систему з $N + M + N \cdot M = 19$ обмежень. Записуємо цільову

функцію $C_{\text{sum}}(X) = \sum_{i=1}^N \sum_{j=1}^M C_{i,j} \cdot X_{i,j}$. Задаємо початкові наближення для плану пе-

ревозень.

$$X_{1,1} = X_{011}; \ X_{1,2} = X_{012}; \ X_{1,3} = X_{013}; \ X_{1,4} = X_{014};$$

$$X_{2,1} = X_{021}; \ X_{2,2} = X_{022}; \ X_{2,3} = X_{023}; \ X_{2,4} = X_{024};$$

$$X_{3,1} = X_{031}; \ X_{3,2} = X_{032}; \ X_{3,3} = X_{033}; \ X_{3,4} = X_{034}.$$

Записуємо усі обмеження

$$\begin{array}{lcl}
 \left. \begin{array}{l}
 X_{1,1} + X_{1,2} + X_{1,3} + X_{1,4} = A_1 \\
 X_{2,1} + X_{2,2} + X_{2,3} + X_{2,4} = A_2 \\
 X_{3,1} + X_{3,2} + X_{3,3} + X_{3,4} = A_3
 \end{array} \right\} & \begin{array}{l}
 N = 3 \text{ рівняння} \\
 \text{балансу для кожної} \\
 \text{бази (постачання)}
 \end{array} \\
 \\
 \left. \begin{array}{l}
 X_{1,1} + X_{2,1} + X_{3,1} = B_1 \\
 X_{1,2} + X_{2,2} + X_{3,2} = B_2 \\
 X_{1,3} + X_{2,3} + X_{3,3} = B_3 \\
 X_{1,4} + X_{2,4} + X_{3,4} = B_4
 \end{array} \right\} & \begin{array}{l}
 M = 4 \text{ рівняння} \\
 \text{балансу для кожної} \\
 \text{крамниці (споживання,} \\
 \text{використання)}
 \end{array} \\
 \\
 \left. \begin{array}{l}
 X_{1,1} \geq 0; X_{1,2} \geq 0; X_{1,3} \geq 0; X_{1,4} \geq 0 \\
 X_{2,1} \geq 0; X_{2,2} \geq 0; X_{2,3} \geq 0; X_{2,4} \geq 0 \\
 X_{3,1} \geq 0; X_{3,2} \geq 0; X_{3,3} \geq 0; X_{3,4} \geq 0
 \end{array} \right\} & \begin{array}{l}
 N \times M \text{ нерівностей -} \\
 \text{невід'ємність усіх} \\
 \text{елементів плану перевезень}
 \end{array}
 \end{array}$$

Нарешті записуємо, що потрібно мінімізувати, за яким критерієм і по яких змінних. З точки зору програмування тут колективне присвоєння - складові оптимального плану перевезень дорівнюють відповідній вбудованій функції (підпрограмі).

$$\left(\begin{array}{cccccc}
 \text{Хор}_{1,1} & \text{Хор}_{1,2} & \text{Хор}_{1,3} & \text{Хор}_{1,4} & \text{Хор}_{1,5} & \text{Хор}_{1,6} \\
 \text{Хор}_{2,1} & \text{Хор}_{2,2} & \text{Хор}_{2,3} & \text{Хор}_{2,4} & \text{Хор}_{2,5} & \text{Хор}_{2,6}
 \end{array} \right) = \text{Minimize}(\text{Csum}, X)$$

Звичайно **алгоритм методу розв'язання транспортної задачі**, складається з таких кроків

- 1) Формування задачі, що є двоїстою до заданої.
- 2) Визначення початкового (такого, що задовольняє обмеженням) плану перевезень.
- 3) Перевірка отриманого плану на оптимальність (за допомогою двоїстої задачі).
- 4) Перехід до нового плану перевезень.
- 5) Перехід до пункту 3) та повторення пунктів 3), 4) до тих пір, поки не буде отриманий оптимальний план, або виявиться, що дана задача не має оптимального розв'язання.

Пункти 1-5 неважко реалізувати як програму у будь-якому програмному середовищі. Звичайно користуються готовими бібліотечними програмними модулями. Потрібно визначити план перевезень, що дає мінімум сумарних витрат.

Узагальнення математичної моделі транспортної задачі

Задачі лінійного (і нелінійного теж) програмування часто мають велику розмірність. Зведемо транспортну задачу до загальної задачі лінійного програмування і запишемо її в канонічній, компактній матрично-векторній формі. Вводимо:

Об'єднаний вектор обмежень

$$\text{Obm} := \text{stack}(A^T, B^T); \quad \text{Obm}^T \rightarrow (A_1 \ A_2 \ A_3 \ B_1 \ B_2 \ B_3 \ B_4).$$

Вектор шуканих змінних (плану)

$$Xv^T \rightarrow (X_{1,1} \ X_{1,2} \ X_{1,3} \ X_{1,4} \ X_{2,1} \ X_{2,2} \ X_{2,3} \ X_{2,4} \ X_{3,1} \ X_{3,2} \ X_{3,3} \ X_{3,4}).$$

Вектор цін

$$Cv^T \rightarrow (C_{1,1} \ C_{1,2} \ C_{1,3} \ C_{1,4} \ C_{2,1} \ C_{2,2} \ C_{2,3} \ C_{2,4} \ C_{3,1} \ C_{3,2} \ C_{3,3} \ C_{3,4}).$$

Тепер можемо записати критерій як скалярний добуток двох останніх векторів:

$$Xv^T \cdot Cv \rightarrow X_{1,1} \cdot C_{1,1} + X_{1,2} \cdot C_{1,2} + X_{1,3} \cdot C_{1,3} + X_{1,4} \cdot C_{1,4} + X_{2,1} \cdot C_{2,1} + X_{2,2} \cdot C_{2,2} + X_{2,3} \cdot C_{2,3} + X_{2,4} \cdot C_{2,4} + X_{3,1} \cdot C_{3,1} + X_{3,2} \cdot C_{3,2} + X_{3,3} \cdot C_{3,3} + X_{3,4} \cdot C_{3,4}$$

та об'єднану матрицю балансів постачання та споживання (**копія екрану!**):

$MP := \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$	В цих позначеннях наша система обмежень запишеться так: $MP \cdot Xsv = \text{Obm}$ Не довіряємо - перевіряємо (розкриваємо цей вираз символьним процесором) і маємо такий результат
$MP \cdot Xv = \text{Obm} \rightarrow \begin{pmatrix} X_{1,1} + X_{1,2} + X_{1,3} + X_{1,4} \\ X_{2,1} + X_{2,2} + X_{2,3} + X_{2,4} \\ X_{3,1} + X_{3,2} + X_{3,3} + X_{3,4} \\ X_{1,1} + X_{2,1} + X_{3,1} \\ X_{1,2} + X_{2,2} + X_{3,2} \\ X_{1,3} + X_{2,3} + X_{3,3} \\ X_{1,4} + X_{2,4} + X_{3,4} \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix}$	Бачимо, що символьний процесор підставив вирази для матриці Мр та векторів Хv та Obm і перемножив їх символьному вигляді.

Таким чином, все збігається з системою обмежень, що записана вище, вручну. Записуємо нарешті задачу у **канонічній формі** разом із **двоїстою задачею**. Спробуйте самостійно дати словесне визначення двоїстої задачі.

Пряма задача. Знайти $V_{\text{sum}_{\text{bpt}}} = \min(C^T \cdot X)$ при обмеженнях $MP \cdot X \geq B$. Двоїста задача. Знайти $D_{\text{sum}_{\text{bpt}}} = \max(B^T \cdot Y)$ при обмеженнях $MP^T \cdot Y \leq C$. Основна теорема двоїстості: $\min(C^T \cdot X) = \max(B^T \cdot Y)$.
--

Методика розв'язання транспортної задачі лінійного програмування

Комп'ютеризація припиняє використання певних математичних методів, або ховає їх у вбудовані функції. Це давня тенденція - хто зараз сам пече хліб, робить тканину, ліпить горщики. В цьому є плюси і мінуси. Коли людина в голові чи на папірці не може перемножити $37 \times 3 = (111)$ або поділити $4/9 : 2/3 = (2/3)$ - це трагедія. Коли спеціаліст не знає, як обчислюється синус, в чому суть симплекс-методу - це теж якийсь не порядок. Потрібен якийсь конструктивний компроміс? Тому розглянемо класичну методику "ручного" розв'язання транспортної задачі, перед тим, як відмовитись від ЛП взагалі.

Алгоритм методу розв'язання транспортної задачі складається з таких кроків:

- 1) Формування задачі, що є двоїстою до заданої.
- 2) Визначення початкового (такого, що задовольняє обмеження) плану перевезень.
- 3) Перевірка отриманого плану на оптимальність.
- 4) Перехід до нового плану перевезень.
- 5) Перехід до пункту 3) та повторення пунктів 3), 4) до тих пір, поки не буде отриманий оптимальний план, або виявиться, що дана задача не має оптимального розв'язання.

Пункти 1-5 неважко реалізувати як програму у будь-якому програмному середовищі. Ще раз нагадаємо, що звичайно користуються готовими бібліотечними програмними модулями.

Економічна інтерпретація отриманого математичного розв'язання

Розв'яжемо задачу крок за кроком згідно з вище поданим алгоритмом.

1) Формування двоїстої задачі. Вводимо змінні u_1, u_2, u_3 (за числом баз) та v_1, v_2, v_3, v_4 (за числом крамниць) та складаємо таблицю, рядки якої відповідають обмеженням прямої задачі (це потім будуть коефіцієнти при двоїстих невідомих змінних), а останній рядок - коефіцієнти рівняння цільової функції. Задамо числові значення параметрів. В першому стовпці запишемо позначення спряжених - двоїстих змінних, в останньому стовпці - обмеження B_j . Ця таблиця полегшує запис та осмислення двоїстої (спряженої) задачі. Поряд з таблицею виводимо її джерела - матриці A, B, C . В перших двох рядках подано припустимі (відносно A та B) плани перевезень.

Копія екрану.

Таблиця 1												
	X_{11}	X_{12}	X_{13}	X_{14}	X_{21}	X_{22}	X_{23}	X_{24}	X_{31}	X_{32}	X_{33}	X_{34}
план-1:	15	15					40	10				20
план-0	15		15				25	25		15		5
u_1	1	1	1	1	0	0	0	0	0	0	0	30
u_2	0	0	0	0	1	1	1	1	0	0	0	50
u_3	0	0	0	0	0	0	0	0	1	1	1	20
v_1	1	0	0	0	1	0	0	0	1	0	0	15
v_2	0	1	0	0	0	1	0	0	0	1	0	15
v_3	0	0	1	0	0	0	1	0	0	0	1	40
v_4	0	0	0	1	0	0	0	1	0	0	0	30
	1	8	2	3	4	7	3	1	5	3	4	11

$A^T = \begin{bmatrix} & 1 \\ 1 & 30 \\ 2 & 50 \\ 3 & 20 \end{bmatrix}$
 $B^T = \begin{bmatrix} & 1 \\ 1 & 15 \\ 2 & 15 \\ 3 & 40 \\ 4 & 30 \end{bmatrix}$
 $C = \begin{bmatrix} & 1 & 2 & 3 & 4 \\ 1 & 1 & 8 & 2 & 3 \\ 2 & 4 & 7 & 3 & 1 \\ 3 & 5 & 3 & 4 & 11 \end{bmatrix}$

Формулювання і запис двоїстої задачі лінійного програмування (ЛП)

Двоїста задача ЛП формулюється так: Знайти розв'язання системи обмежень (це суми по стовпцях поданої вище таблиці)

$$\left. \begin{array}{lll} u_1 + v_1 \leq 1 & u_2 + v_1 \leq 4 & u_3 + v_1 \leq 5 \\ u_1 + v_2 \leq 8 & u_2 + v_2 \leq 7 & u_3 + v_2 \leq 3 \\ u_1 + v_3 \leq 2 & u_2 + v_3 \leq 3 & u_3 + v_3 \leq 4 \\ u_1 + v_4 \leq 3 & u_2 + v_4 \leq 1 & u_3 + v_4 \leq 11 \end{array} \right\} C^T = \begin{array}{|c|c|c|c|} \hline & 1 & 2 & 3 \\ \hline 1 & 1 & 4 & 5 \\ \hline 2 & 8 & 7 & 3 \\ \hline 3 & 2 & 3 & 4 \\ \hline 4 & 3 & 1 & 11 \\ \hline \end{array}$$

яке максимізує таку цільову функцію

$$F = 30 \cdot u_1 + 50 \cdot u_2 + 20 \cdot u_3 + 15 \cdot v_1 + 15 \cdot v_2 + 40 \cdot v_3 + 30 \cdot v_4$$

2) Визначення початкового плану перевезень. Використовуємо спосіб "північно-західний кут" (або "верхній лівий"). Згідно з цим способом спочатку максимально можлива кількість товару з бази розміщується у верхній лівій клітинці таблиці, потім в наступній клітинці в тому ж рядку..., аж поки не буде використана 1-ша база постачання, потім наступні клітинки в другому рядку (тобто розподіляється товар з другої бази) і т.д.

Виконується контроль допустимості плану - суми по рядках та стовпцях повинні дорівнювати відповідним обмеженням. Записуємо матрицю вартостей перевезень і обчислюємо значення цільової функції.

$$\begin{pmatrix} pp & Mag1 & Mag2 & Mag3 & Mag4 & ock1 \\ Ck11 & 15 & 15 & 0 & 0 & 30 \\ Ck12 & 0 & 0 & 40 & 10 & 50 \\ Ck13 & 0 & 0 & 0 & 20 & 20 \\ omag & 15 & 15 & 40 & 30 & ooo \end{pmatrix}$$

$$F1 := 1 \cdot 15 + 8 \cdot 15 + 3 \cdot 40 + 1 \cdot 10 + 11 \cdot 20 \quad F1 = 485$$

3) Перевірка отриманого плану на оптимальність. Повертаємось до початкової таблиці 1 і у верхньому рядку записуємо перший план перевезень.

Випишуємо обмеження для тих стовпців, де $X_{ij} \neq 0$, задаємо нульові значення зайвим змінним, знаходимо значення ненульових змінних і підставляємо їх в інші нерівності. Якщо не усі з нерівностей виконуються - значить план ще не оптимальний. Порівнюємо дійсні значення нерівностей з потрібними.

$$u_1 := 0 ; u_2 := -10 ; u_3 := 0 ; \quad v_1 := 1 ; v_2 := 8 ; v_3 := 3 ; v_4 := 11 .$$

	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₃₁	X ₃₂	X ₃₃	X ₃₄
план-1 :	15	15				40	10				20	
план-о :	15		15			25	25		15		5	
u1	1	1	1	1	0	0	0	0	0	0	0	30
u2	0	0	0	0	1	1	1	1	0	0	0	50
u3	0	0	0	0	0	0	0	0	1	1	1	20
v1	1	0	0	0	1	0	0	0	1	0	0	15
v2	0	1	0	0	0	1	0	0	1	0	0	15
v3	0	0	1	0	0	1	0	0	0	1	0	40
v4	0	0	0	1	0	0	1	0	0	0	1	30
	1	8	2	3	4	7	3	1	5	3	4	11

Маємо:

Потрібно:

$$u1 + v1 = 1 \quad u1 + v1 \leq 1$$

$$u1 + v2 = 8 \quad u1 + v2 \leq 8$$

$$u2 + v3 = -7 \quad u2 + v3 \leq 3$$

$$u2 + v4 = 1 \quad u2 + v4 \leq 1$$

$$u3 + v4 = 11 \quad u3 + v4 \leq 11$$

$$u3 + v3 = 3 \quad u3 + v3 \leq 4$$

Маємо:

Потрібно:

$$u2 + v1 = -9 \quad u2 + v1 \leq 4 \quad \text{nev}$$

$$u2 + v2 = -2 \quad u2 + v2 \leq 7$$

$$u1 + v3 = 3 \quad u1 + v3 \leq 2 \quad \text{nev}$$

$$u1 + v4 = 11 \quad u1 + v4 \leq 3$$

$$u3 + v1 = 1 \quad u3 + v1 \leq 5$$

$$u3 + v2 = 8 \quad u3 + v2 \leq 3 \quad \text{nev}$$

4) Перехід до нового плану перевезень. Тепер, зберігаючи баланс перевезень, покращуємо план, - виносимо перевезення у клітинки з малою вартістю перевезень. Записуємо отриманий після двох циклів (кроків) перерозподілу план у вхідну таблицю 1 та перевіряємо його на оптимальність. Бачимо, що сумарні витрати зменшились майже вдвічі відносно початкового плану.

$$F2 := 1 \cdot 15 + 2 \cdot 15 + 3 \cdot 25 + 1 \cdot 25 + 3 \cdot 15 + 11 \cdot 5;$$

$$F1 = 485 ; \quad F2 = 245 ; \quad F1 - F2 = 240 .$$

Перевіримо цей план на оптимальність. Використаємо один з вбудованих методів - функцію (підпрограму) розв'язання числовим методом довільних систем рівностей та нерівностей і інших відношень - так званий розв'язальний блок Given- Find. Ми задаємо для цільової функції відношення "краще" (тобто менше витрат).

Можливо використати вбудовану функцію оптимізації (Optimize(...)) - ми можемо вибрати метод оптимізації, спосіб пошуку, точність та ін. Це ми зробимо в наступних розділах.

Задаємо початкові наближення рівними другому плану перевезень.

$$Z_{tr} := 200 \quad X_{11} := 15 \quad X_{12} := 15 \quad X_{13} := 0 \quad X_{14} := 0 \quad X_{21} := 0 \quad X_{22} := 0$$

$$X_{31} := 0 \quad X_{32} := 0 \quad X_{33} := 0 \quad X_{34} := 0 \quad X_{21} := 0 \quad X_{23} := 0 \quad X_{24} := 0$$

Після ключового слова Given записуємо систему рівнянь прямої задачі для даних контрольного прикладу.

Given

$$\begin{aligned}
 &1 \cdot X_{11} + 8 \cdot X_{12} + 2 \cdot X_{13} + 3 \cdot X_{14} + 4 \cdot X_{21} + 7 \cdot X_{22} + 3 \cdot X_{23} \dots \leq Z_{tr} \\
 &+ 1 \cdot X_{24} + 5 \cdot X_{31} + 3 \cdot X_{32} + 4 \cdot X_{33} + 11 \cdot X_{34} \\
 &X_{11} + X_{12} + X_{13} + X_{14} = 30; \quad X_{11} \geq 0; \quad X_{21} \geq 0; \quad X_{31} \geq 0; \\
 &X_{11} + X_{21} + X_{31} = 15; \\
 &X_{21} + X_{22} + X_{23} + X_{24} = 50; \quad X_{12} \geq 0; \quad X_{22} \geq 0; \quad X_{32} \geq 0; \\
 &X_{12} + X_{22} + X_{32} = 15; \quad X_{13} \geq 0; \quad X_{23} \geq 0; \quad X_{33} \geq 0; \\
 &X_{31} + X_{32} + X_{33} + X_{34} = 20; \\
 &X_{13} + X_{23} + X_{33} = 40; \quad X_{14} \geq 0; \quad X_{24} \geq 0; \quad X_{34} \geq 0; \\
 &X_{14} + X_{24} + X_{34} = 30;
 \end{aligned}$$

Після системи рівнянь пишемо ключову функцію Find() - знайти, що замикає розв'язувальний блок. Виводимо обидва плани.

Оптим_план := Find(X11,X12,X13,X14,X21,X22,X23,X24,X31,X32,X33,X34) .

Оптим_план = Наш_план - Оптим_план - Наш_план =

	1
1	15
2	0
3	15
4	0
5	-0
6	0
7	20
8	30
9	0
10	15
11	5
12	0

	1
1	15
2	0
3	15
4	0
5	0
6	0
7	25
8	25
9	0
10	15
11	0
12	5

	1
1	0
2	0
3	-0
4	0
5	-0
6	0
7	-5
8	5
9	0
10	0
11	5
12	-5

Ось тут ми маємо в числах:
а) плани перевезень;
б) матрицю цін перевезень.

Помножимо, як у школі,
обсяги перевезень на ціни -
отримаємо витрати
на перевезення

C =

	1	2	3	4
1	1	8	2	3
2	4	7	3	1
3	5	3	4	11

$$F2 := 1 \cdot 15 + 8 \cdot 0 + 2 \cdot 15 + 3 \cdot 0 + 4 \cdot 0 + 7 \cdot 0 + 3 \cdot 25 + 1 \cdot 25 + 5 \cdot 0 + 3 \cdot 15 + 4 \cdot 0 + 11 \cdot 5$$

$$For := 1 \cdot 15 + 8 \cdot 0 + 2 \cdot 15 + 3 \cdot 0 + 4 \cdot 0 + 7 \cdot 0 + 3 \cdot 20 + 1 \cdot 30 + 5 \cdot 0 + 3 \cdot 15 + 4 \cdot 5 + 11 \cdot 0$$

Подивимось (зконцентруємося, зфокусуємося, як пишуть у певних підручниках з менеджменту) на плани перевезень - стартовий "південно-західний", наш покращений і знайдений "комп'ютером" (= вбудованою програмою пакету).

Бачимо, що наш план помітно відрізняється за структурою перевезень від оптимального плану перевезень. Витрати за цими планами дорівнюють:

стартовий (опорний) $F1 = 485$;

покращений $F2 = 245$;

оптимальний $For = 200$.

Але навколо оптимального плану є декілька дуже близьких до нього "субоптимальних". Це добре знають практики і математики - система обмежень класичної транспортної задачі - випукла і тому близькі до оптимуму плани не можуть бути значно гіршими.

Менеджеру, між іншим, потрібен не оптимальний план типу "крок вправо, крок вліво - ..." , а підмножина близьких до оптимального планів - простір для маневрів. І взагалі, якщо врахувати реальні рівні невизначеності цін, потреб, то все, що ми розглянули, може здаватись просто математичною іграшкою.

Зауваження. Подивіться на вхідні дані для блока "Given - Find", там прийнято: $Z_{tr} = 200$ - ми поставили умову, щоб сумарні витрати не перевищували це значення. Чому ми не взяли $Z_{tr} = 150$ або 300 ? Вибране число - результат такої процедури: задаємо якесь велике значення Z_{tr} , програма знаходить нам план, що задовольняє це обмеження, потім зменшуємо Z_{tr} , аж поки програма "Given - Find" не видасть червоне повідомлення: "розв'язок не знайдено". Так ми приходимо до мінімальних витрат і відповідного плану. Ніхто не забороняє Вам написати програму, що виконує таку процедуру пошуку екстремуму. Це буде альтернатива симплекс-методу.

Контрольні запитання

1. Сформулюйте транспортну задачу: обмеження, критерій, змінні оптимізації, ціль.
2. Запишіть задачу лінійного програмування (ЛП) у матрично-векторному вигляді.
3. Запишіть у матрично-векторному вигляді пряму і двоїсту задачу ЛП.
4. Основний результат (теорема) теорії двоїстості.
5. Загальна методика розв'язання транспортної задачі.
6. Що таке метод "північно-західного кута"?
7. Що дає (крім зайвого клопоту) використання символьного процесора?
8. Запишіть у розгорнутому вигляді систему обмежень (нерівностей) транспортної задачі довільної розмірності.
9. Дайте інтерпретацію матрицям даних A, B, C для транспортної задачі.
10. Для чого у розв'язанні транспортної задачі використовується двоїста задача?



4.2. Розробка інтерфейсу користувача з підсистемою оптимізації планів перевезень

Вступ. Постановка задачі

Головними елементами, що відрізняють експертну, інтелектуальну систему від просто автоматизованої системи є:

1) відкритість усіх моделей і "механізмів" для аналізу (переглядання) та модифікації - кваліфікований спеціаліст не буде працювати з "чорним ящиком";

2) розв'язання задач (в першу чергу - оптимізаційних) паралельно декількома альтернативними методами (згідно з "варіантним аналізом" слід знайти декілька варіантів проектного рішення і вибрати з них "оптимальний", для інформаційних систем оптимальним варіантом буде **сума** усіх працездатних альтернатив);

3) усі вхідні і вихідні дані, системи рівнянь, цільові функції - все це повинно даватися у зручному графічному вигляді.

Головна задача цього розділу - розробка структури і дизайну базового графічного модуля для задач планування перевезень. Термін "базовий модуль" означає стандартний програмний об'єкт, який можна копіювати, модифікувати його сценарії та вигляд, вбудовувати в інші стандартні документи.

Під **стандартними документами** маються на увазі документи з графікою, анімацією, аудіо-відео, які можна просто **переносити між стандартними програмними платформами і пересилати Інтернетом**.

Розробка підсистеми графічного подання задачі

Обмежимося досить обмеженими графічними можливостями пакету Mathcad. Побудуємо систему для розв'язання та дослідження транспортних задач. Спочатку розробимо базовий модуль мінімальної розмірності - 2 бази та 3 крамниці. Потім спробуємо його масштабувати - збільшити розмірність задачі.

Дано $n = 2$ магазини, на які щоденно доставляються з $m = 2$ складів товари.

Задано:

потрібні кількості для кожного магазину - M_i ;

можливості постачання кожного складу - C_{kl} ;

ціни перевезень з кожного складу до кожного магазину - $c_{i,j}$.

Треба знайти план перевезень (= по скільки возити з кожного складу в кожний магазин), що давав би **МІНІМАЛЬНІ** сумарні витрати на перевезення при виконанні вимог за обсягами поставок. І не просто знайти, а розробити зручну діалогову систему, що дозволяла би вводити значення параметрів задачі і отримувати рішення у зручному вигляді.

Розробка модуля формування схеми транспортної задачі

Спочатку організуємо сервіс - задаємо параметри графіків, що відобразять карту розташування магазинів і складів:

ORIGIN := 1 - це асемблерна змінна - "початкове значення індексу масивів".

Задаємо координати розташування крамниць та складів матрицями координат відповідних пунктів:

Крамниць := 2 ; Складів := 2 ; i := 1..Крамниць ; j := 1..Складів ;

$$Ma := \begin{pmatrix} 1 & 3 \\ 2.7 & .5 \end{pmatrix} ; \quad Ckl := \begin{pmatrix} 1.2 & 3.2 \\ 1.3 & 3.5 \end{pmatrix} ; \quad k := 1..Складів + 1 .$$

В робочій версії системи положення буде задавати користувач в режимі "рисунка", а координати буде визначати відповідний модуль. Формуємо матриці для виведення схеми:

$$Vz2 := \begin{pmatrix} Ma_{1,1} & Ckl_{1,2} & Ma_{1,2} \\ Ma_{2,1} & Ckl_{2,2} & Ma_{2,2} \end{pmatrix} ; \quad Vz1 := \begin{pmatrix} Ma_{1,1} & Ckl_{1,1} & Ma_{1,2} \\ Ma_{2,1} & Ckl_{2,1} & Ma_{2,2} \end{pmatrix}$$

Розробка модуля розв'язання транспортної задачі

Формалізуємо задачу - задаємо (введення нижче - на стенді)

матрицю цін: вектори потреб магазинів: та можливостей складів :

$$ci = \begin{pmatrix} 12 & 16 \\ 6 & 10 \end{pmatrix} ; \quad mag = \begin{pmatrix} 40 \\ 80 \end{pmatrix} ; \quad skl = \begin{pmatrix} 50 \\ 70 \end{pmatrix} ;$$

Задаємо **цільову функцію** (те, що будемо оптимізувати):

$$\boxed{\text{Опф}(v11, v12, v21, v22) := v11 \cdot ci_{1,1} + v12 \cdot ci_{1,2} + v21 \cdot ci_{2,1} + v22 \cdot ci_{2,2} .}$$

Контроль. Згідно з минулим стандартом - треба виконати все вручну:

можливості: $50 + 70 = 120$; потреби: $40 + 80 = 120$.

Згідно зі світовими стандартами - зараз **забороняється обчислення та алгебраїчні перетворення виконувати вручну** - людина є головним джерелом дуже дорогих помилок. Комп'ютером же перевіряємо те, що закладаємо у комп'ютер:

$$\sum_{i=1}^{\text{Крамниць}} mag_i = 120 ; \quad \sum_{j=1}^{\text{Складів}} skl_j = 120 .$$

Краще записати елемент контролю у вигляді логічного висловлювання:

$$\sum_{i=1}^{\text{Крамниць}} mag_i = \sum_{j=1}^{\text{Складів}} skl_j = 1 .$$

Якщо балансу "попит-запаси" немає, це висловлювання буде дорівнювати нулю. Можна написати елемент контролю з виведенням світлового (блимаючого) та звукового попередження, можна через DLL взяти заповнити своїм текстом стандартне попередження з Windows.

Задаємо початкові значення змінних, що оптимізуються - обсягів перевезень:

$$v11 := 0 ; \quad v12 := 0 ; \quad v21 := 0 ; \quad v22 := 0 .$$

Нарешті записуємо саму оптимізаційну задачу згідно з синтаксисом пакета - після ключового слова „Given” записуємо усі обмеження задачі (рівності, нерівності, лінійні, нелінійні).

Given

$v11 + v21 = \text{mag}_1$ - потреби 1-го магазину ;

$v12 + v22 = \text{mag}_2$ - потреби 2-го магазину ;

$v11 + v12 = \text{skl}_1$ - можливості 1-го складу ;

$v21 + v22 = \text{skl}_2$ - можливості 2-го складу ;

$v11 \geq 0$; $v21 \geq 0$ - а це обмеження здорового глузду

$v12 \geq 0$; $v22 \geq 0$ - змінні не можуть бути від'ємними?

В цьому документі використовуються як "саморобні", так і вбудовані методи оптимізації. З певних причин (авторські права та ін.) вбудовані методи оптимізації, не завжди ефективні, але вони "завжди готові". Подаємо копію екрана з системою меню налаштування методу і параметрів пошукової оптимізації - на фоні нашої задачі.

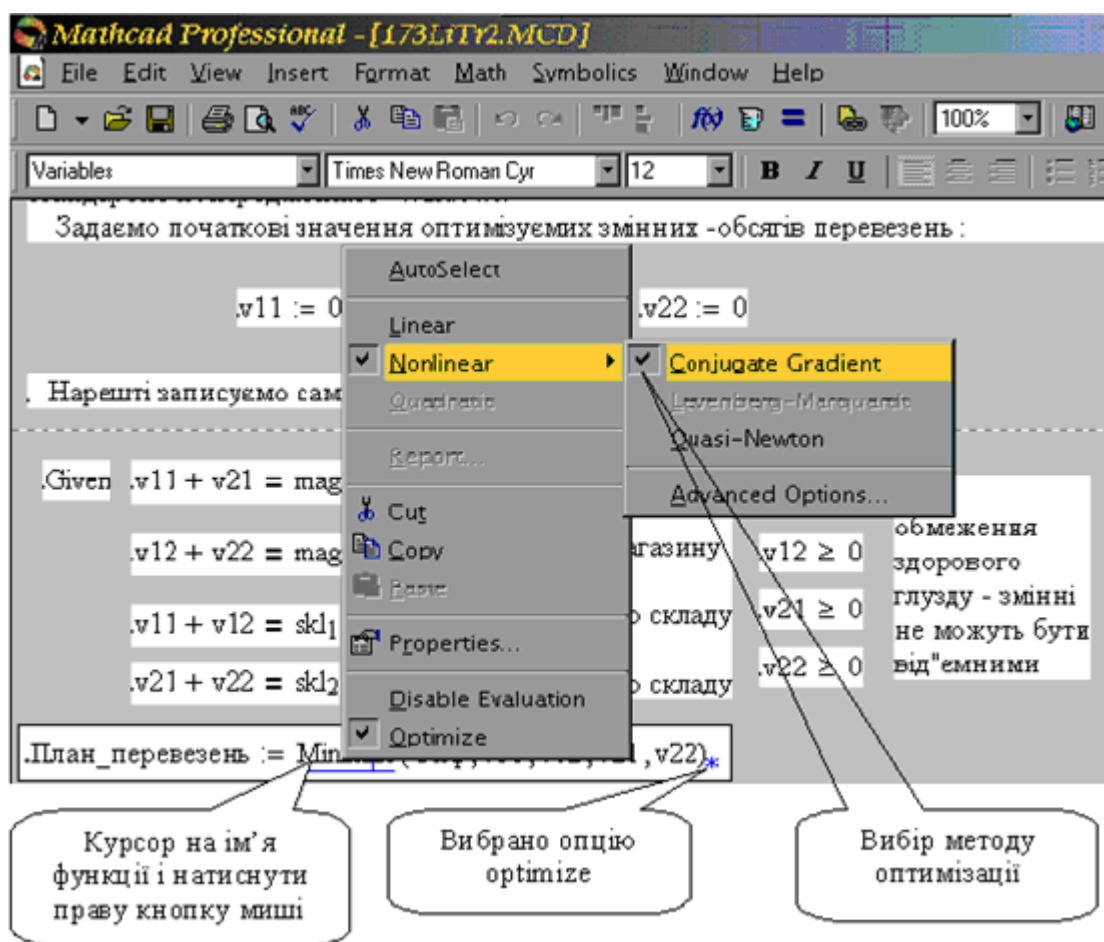


Рис. 1. Налаштування оптимізаційної задачі – копія екрану

Тепер після системи рівнянь (обмежень) записуємо ім'я поліморфної вбудованої функції оптимізації (згідно з канонами об'єктно-орієнтованого програмування, поліморфізм - це, коли в залежності від типу даних, виконується різна обробка даних)

План_перевезень := Minimize(Опф, v11, v12, v21, v22)*

Цей останній вираз "блока розв'язання" задає: що робити (мінімізація), для якої цільової функції та по яких змінних. Отримали РЕЗУЛЬТАТ - оптимальний план перевезень. Формуємо з вектора матрицю плану перевезень.

$$\text{План_перевезень} = \begin{pmatrix} 7.7 \\ 42.3 \\ 32.3 \\ 37.7 \end{pmatrix} \quad \begin{array}{ll} \text{pp1,1} := \text{План_перевезень}_1 & \text{pp1,2} := \text{План_перевезень}_2 \\ \text{pp2,1} := \text{План_перевезень}_3 & \text{pp2,2} := \text{План_перевезень}_4 \end{array}$$

Розробка системи формування схеми транспортної задачі

Те, що подано далі, - стенд для оптимізації та аналізу плану перевезень. Це підсумковий **результат** даного розділу. З переглянутих вище способів графічного подання задачі вибрано простіший - аплікація елементів і виведення на графік - схему розташування баз і країн. Виводяться потреби магазинів, можливості баз, ціни перевезень і власне оптимальний план - звідки і скільки перевозиться.

Поряд зі схемою перевезень подано зони введення параметрів задачі - цін, потреб магазинів, можливостей складів. В нижній частині стенду подано результати оптимізації - мінімальне значення витрат та план перевезень, що їх забезпечує.

Цей стенд дозволяє подивитись, як змінюється оптимальний план із зміною цін чи потреб. Особливість лінійної задачі з обмеженнями в тому, що при неперервній зміні ціни перевезень план змінюється стрибком - так і повинно бути, тому що оптимальне розв'язання дискретне - це одна з вершин симплексу обмежень.

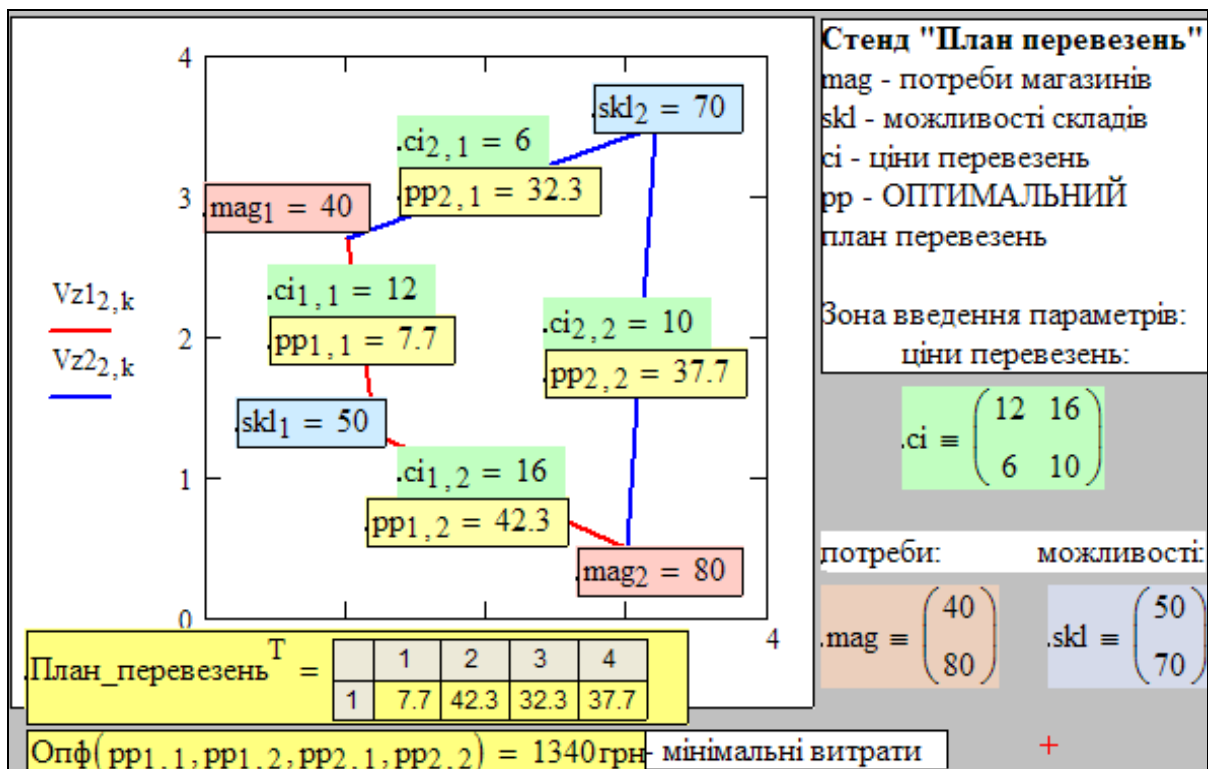


Рис. 2. Базовий варіант стенду для оптимізації та аналізу планів перевезень

В електронній версії те, що подано вище - не картинка, а працююча діалогова система для розв'язання певного класу задач.

Вхідні дані:

- а) може вводити користувач;
- б) можна вводити з мереж різного рівня;
- в) можуть передаватись безпосередньо від стандартних передавачів.

Корпорація Matworks випустила на ринок продукт Mathworks Test & Measurement Suite що забезпечує збір та обробку інформації від різних приладів та джерел. Ці дані без проблем адаптуються і обробляється в приложеннях, що створені в середовищі пакету Matlab цієї ж корпорації.

Інструкція користувача стенду.

1. У виділених зонах введення (рис.2, праворуч, матриці mag , skl , ci) задається значення попиту, запасів та цін перевезень. Ліворуч виводяться план перевезень та витрати згідно з цим планом.

2. Для зміни назв змінних (для узгодження із змістом задачі та естетичними поглядами слід внести відповідні зміни в документі. При бажанні можна потім будь-яку частину документа захистити від змін, або взагалі закрити від перегляду власним паролем.

Висновки. Напрямки модифікацій системи

Розроблено базовий модуль для підсистеми планування перевезень. Проведено випробування системи. Модуль при незначних доробках може бути використаний в інших задачах для імітаційного моделювання, досліджень, аналізу та тренінгу.

Можна виділити такі напрямки модифікації базового модуля:

- 1. Масштабування транспортної задачі.
- 2. Модифікація стенду для дослідження впливу варіацій цін (5%, 10%, 25%...) на стійкість оптимального плану.
- 3. Розробка програми для визначення критичних значень, коли при малих змінах параметрів в оптимальному плані перевезення відбуваються великі раптові зміни.
- 4. Модифікація системи для дослідження залежності витрат від варіацій (5%, 10%, 25%...) параметрів оптимального плану без порушення балансу (наприклад, збільшити перевезення в одну крамницю на 5% і на стільки ж знизити в іншу).
- 5. Модифікація системи для оптимізації плану поставок декількох типів товарів (ціну перевезення типів товарів можна вважати однаковою для спрощення задачі.

Контрольні запитання

- 1. Сформулюйте транспортну задачу: обмеження, критерій, змінні оптимізації, ціль.
- 2. Запишіть задачу лінійного програмування (ЛП) у матрично-векторному вигляді.
- 3. Скільки обмежень має транспортна задача, що характеризують ці обмеження?
- 4. Для розглянутої транспортної задачі запишіть двоїсту задачу.
- 5. Запишіть розв'язальний блок для випадків:
 - оптимізації при обмеженнях;
 - знаходження коренів системи рівнянь;
 - знаходження розв'язання системи рівнянь з мінімальною помилкою.
- 6. Що роблять вбудовані функції `find()`, `minerr()`, `minimize()`, `maximize()`?



4.3. Масштабування програми оптимізації планів перевезень. Що буде якщо аналіз

Вступ. Постановка задачі

Цілі розділу - 1) розробка діалогової системи для оптимізації транспортних задач; 2) демонстрація на конкретному прикладі того, що називають "новими інформаційними технологіями".

Виконується модифікація базової задачі з розділу 4.2 - ми збільшуємо її розмірність з "2 крамниці x 2 склади" до "6 крамниць x 2 склади"). Така операція називається масштабуванням задачі чи програми.

Стратегічна ціль масштабування - зробити програму оптимізації планів перевезень досить універсальною, так, щоб зміна розмірності задачі зводилась до простого введення нових значень розмірності задачі без переробки модулів програми.

Кінцева ціль (за межами цього посібника) - зробити документ класу "аналітичний помічник менеджера" для аналізу ситуацій перевезень (зміна цін, коливання попиту) та оптимізації планів перевезень.

Розмірність задачі повинна бути вхідним параметром документа, як і ціни, потреби крамниць та можливості складів. Цей документ повинен легко вбудовуватись в існуючі автоматизовані робочі місця (АРМ) менеджерів.

При масштабуванні задачі може виявитись, що обсяг обчислень зростає занадто швидко - слід ретельно вибирати і розробляти обчислювальні методи для задач великої розмірності. Наше проектне рішення - використання відомих та розробка нових ітеративних схем, розбитої логіки, декомпозиції, агрегування і редукції (спрощення) задач високої розмірності.

Розробка підсистеми формування схеми транспортної задачі

Дано $n = 6$ магазинів, на які щоденно доставляються з $m = 2$ складів продукти стійкого попиту - ну, **столи і стільці**. Задано: потрібні кількості для кожної магазину - Ma і можливості постачання кожного складу - Sk . Задано ціни перевезень з кожного складу до кожного магазину - ci .

Треба знайти план перевезень (= по скільки возити з кожного складу в кожну крамницю), що давав би МІНІМАЛЬНІ сумарні витрати на перевезення при виконанні вимог щодо обсягів поставок. Схема перевезень, вхідні і вихідні дані повинні розташовуватись на одній-двох екранних сторінках.

Модифікуємо документ розроблений в попередньому розділі. Розробляємо сервісний модуль, що будуватиме схему (граф перевезень). Задаємо координати розташування крамниць та складів.

ORIGIN:= 1;

Крамниць := 6; Складів:= 2; $i := 1..$ Крамниць; $k := 1..2$ Крамниць; $j := 1..$ Складів.

Зробимо програму, яка формує робочі масиви для побудови ліній "магазин-склад". Спочатку робимо два окремих модуля:

vz1 :=	for i ∈ 1..2 · Крамниць
	$ko^{(i)} \leftarrow \begin{cases} Ma^{(i+1) \cdot 0.5} & \text{if } \text{mod}(i, 2) \neq 0 \\ Ckl^{(1)} & \text{otherwise} \end{cases}$
	"qq"
	ko
vz2 :=	for i ∈ 1..2 · Крамниць
	$ko^{(i)} \leftarrow \begin{cases} Ma^{(i+1) \cdot 0.5} & \text{if } \text{mod}(i, 2) \neq 0 \\ Ckl^{(2)} & \text{otherwise} \end{cases}$
	"qq"
	ko

Тепер збираємо ці модулі в один. Використовуємо можливості пакета - групове присвоєння. В даному випадку "ім'я" програми - вектор з двох компонентів. Кожний компонент - матриця 2 x 12. Перша матриця задає шляхи до кожного магазину з першої бази, друга - шляхи з другої.

$\begin{pmatrix} vz1 \\ vz2 \end{pmatrix} :=$	for i ∈ 1..2 · Крамниць
	$ko^{(i)} \leftarrow \begin{cases} Ma^{(i+1) \cdot 0.5} & \text{if } \text{mod}(i, 2) \neq 0 \\ Ckl^{(1)} & \text{otherwise} \end{cases}$
	$ku^{(i)} \leftarrow \begin{cases} Ma^{(i+1) \cdot 0.5} & \text{if } \text{mod}(i, 2) \neq 0 \\ Ckl^{(2)} & \text{otherwise} \end{cases}$
	$\begin{pmatrix} ko \\ ku \end{pmatrix}$

Подаємо вихід програми. Зауважимо, що, якби ми дали скалярне ім'я програмі, нам би прийшлося розпаковувати її вихід по компонентах.

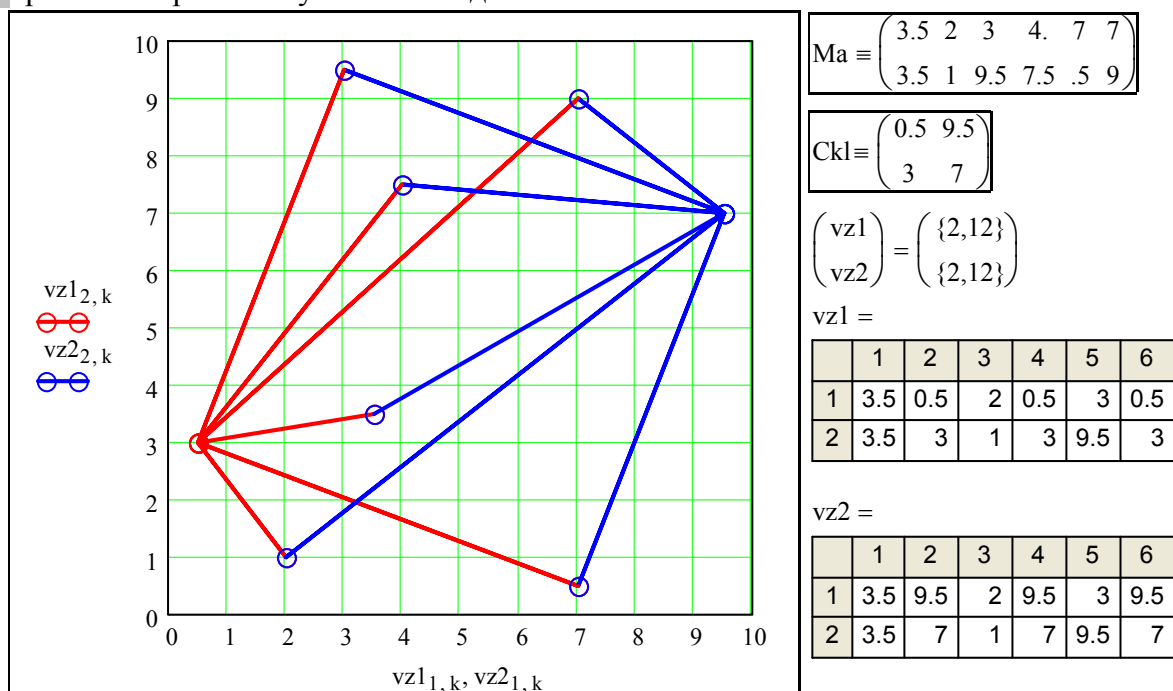


Рис.1. Стенд "формування схеми перевезень"

Будуємо графік - схему перевезень (рис.1). Біля нього розташовуємо матриці координат складів та магазинів. Завдяки абсолютному присвоєнню ці дані "бачать" програми vz1 vz2, що розташовані вище. Ми можемо коректувати ці матриці і бачити відразу результати. В робочій версії системи положення буде задавати користувач в режимі "рисунка", а координати буде визначати відповідний модуль. Формуємо матриці Ma і Skl для виведення схеми (рис.1).

Розробка підсистеми розв'язання транспортної задачі

Модифікуємо документ розроблений в попередньому розділі. Формалізуємо задачу - задаємо (згідно з логікою послідовного подання задачі, дійсне введення - далі на стенді):

матрицю цін: вектори потреб магазинів: та можливостей складів :

$$c = \begin{pmatrix} 9 & 8 & 16 & 11 & 12 & 6 \\ 8 & 10 & 16 & 7 & 11 & 9 \end{pmatrix} ; \quad \text{mag}^T = (30 \ 50 \ 100 \ 15 \ 55 \ 50) ; \quad \text{skl} = \begin{pmatrix} 110 \\ 190 \end{pmatrix} .$$

Задаємо **цільову функцію**, або критерій - числову міру переваги. В даному випадку це сумарні витрати на перевезення. Ціль оптимізації - мінімізація цих витрат.

Зауваження. Така ціль неочевидна, у перевізника можуть бути інші цілі. І критерій теж не враховує усіх факторів забезпечення пунктів обслуговування необхідними ресурсами. Але зупиняємось на загальноприйнятому критерії.

$$\text{Опф}(P_p) := \sum_{j=1}^{\text{Крамниць Складів}} \sum_{i=1}^{\text{Складів}} P_{p,i,j} \cdot c_{i,j} .$$

Контроль. Згідно з минулим стандартом - треба виконати все вручну, отак:

$$30 + 50 + 100 + 15 + 55 + 50 = 300 ; \quad 110 + 190 = 300 .$$

Ще раз нагадаємо, що вкрай небажано **обчислення та алгебраїчні перетворення довіряти людині** - вона головне джерело дуже дорогих помилок. Перевіряємо те, що вводимо у комп'ютер. Отак:

$$\text{mag}_1 + \text{mag}_2 + \text{mag}_3 + \text{mag}_4 + \text{mag}_5 + \text{mag}_6 = 300 ; \quad \text{skl}_1 + \text{skl}_2 = 300 .$$

$$\text{Краще отак:} \quad \sum_{i=1}^{\text{Крамниць}} \text{mag}_i = 300 ; \quad \sum_{j=1}^{\text{Складів}} \text{skl}_j = 300 .$$

Ще краще записати елемент контролю у вигляді логічного висловлювання:

$$\sum_{i=1}^{\text{Крамниць}} \text{mag}_i = \sum_{j=1}^{\text{Складів}} \text{skl}_j = 1 ; \quad \sum_{i=1}^{\text{Крамниць}} (\text{mag}_i + .001) = \sum_{j=1}^{\text{Складів}} \text{skl}_j = 0$$

Якщо баланс "попит-запаси" немає, це висловлювання буде дорівнювати нулю. Можна взагалі написати програмні модулі контролю даних і обчислень з ефективними повідомленнями, наприклад, можна через DLL взяти заповнити своїм текстом стандартне попередження з Windows.

Задаємо початкові значення змінних, що оптимізуються - обсягів перевезень:

$$Pr_{1,1} := 0; \quad Pr_{1,2} := 0; \quad Pr_{1,3} := 0; \quad Pr_{1,4} := 0; \quad Pr_{1,5} := 0; \quad Pr_{1,6} := 0;$$

$$Pr_{2,1} := 1; \quad Pr_{2,2} := 2; \quad Pr_{2,3} := 10; \quad Pr_{2,4} := 0; \quad Pr_{2,5} := 4; \quad Pr_{2,6} := 10;$$

Нарешті записуємо саму оптимізаційну задачу згідно з синтаксисом пакета. Після ключового слова `given` записуємо обмеження - рівняння, нерівності.

Given

Умови задоволення потреб магазинів. В даному випадку їх 6.

$$Pr_{1,1} + Pr_{2,1} = mag_1; \quad Pr_{1,3} + Pr_{2,3} = mag_3; \quad Pr_{1,5} + Pr_{2,5} = mag_5;$$

$$Pr_{1,2} + Pr_{2,2} = mag_2; \quad Pr_{1,4} + Pr_{2,4} = mag_4; \quad Pr_{1,6} + Pr_{2,6} = mag_6.$$

Умови балансу постачання для кожної бази постачання.

В даному випадку баз - 2.

$$Pr_{1,1} + Pr_{1,2} + Pr_{1,3} + Pr_{1,4} + Pr_{1,5} + Pr_{1,6} = skl_1;$$

$$Pr_{2,1} + Pr_{2,2} + Pr_{2,3} + Pr_{2,4} + Pr_{2,5} + Pr_{2,6} = skl_2.$$

Обмеження здорового глузду - змінні не можуть бути від'ємними

$$Pr_{1,1} \geq 0; \quad Pr_{1,2} \geq 0; \quad Pr_{1,3} \geq 0; \quad Pr_{1,4} \geq 0; \quad Pr_{1,5} \geq 0; \quad Pr_{1,6} \geq 0;$$

$$Pr_{2,1} \geq 0; \quad Pr_{2,2} \geq 0; \quad Pr_{2,3} \geq 0; \quad Pr_{2,4} \geq 0; \quad Pr_{2,5} \geq 0; \quad Pr_{2,6} \geq 0;$$

Після цього записуємо вираз, що вказує, а) що робити (мінімізація); б) для якої цільової функції (вона визначена вище); г) по яких змінних. Зверніть увагу на те, що ми пишемо тільки ім'я цільової функції - Опф.

$$\begin{pmatrix} Pr_{1,1} & Pr_{1,2} & Pr_{1,3} & Pr_{1,4} & Pr_{1,5} & Pr_{1,6} \\ Pr_{2,1} & Pr_{2,2} & Pr_{2,3} & Pr_{2,4} & Pr_{2,5} & Pr_{2,6} \end{pmatrix} := \text{Minimize}(\text{Опф}, Pr).$$

Отримуємо РЕЗУЛЬТАТ - оптимальний план перевезень.

$$\begin{pmatrix} Pr_{1,1} & Pr_{1,2} & Pr_{1,3} & Pr_{1,4} & Pr_{1,5} & Pr_{1,6} \\ Pr_{2,1} & Pr_{2,2} & Pr_{2,3} & Pr_{2,4} & Pr_{2,5} & Pr_{2,6} \end{pmatrix} = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 50 & 10 & 0 & 0 & 50 \\ 2 & 30 & 0 & 90 & 15 & 55 & 0 \end{array}.$$

А тепер скомпонуємо входи і виходи задачі в зручному для користувача інтерфейсі. Виконаємо декоративне присвоєння (замість того, щоб писати коментарі до змінної, можна зробити довге ім'я, що саме себе описує:

$$\text{План_перевезень} := \begin{pmatrix} Pr_{1,1} & Pr_{1,2} & Pr_{1,3} & Pr_{1,4} & Pr_{1,5} & Pr_{1,6} \\ Pr_{2,1} & Pr_{2,2} & Pr_{2,3} & Pr_{2,4} & Pr_{2,5} & Pr_{2,6} \end{pmatrix}.$$

Зауваження. Далі будуть розглянуті альтернативні (нечітка оптимізація) методи розв'язання оптимізаційних задач. Сучасні програмні платформи мають вбудовані методи оптимізації, не завжди вдалі та ефективні, але "завжди готові". Вибираємо в даному випадку саме вбудований метод оптимізації - користувачу потрібні надійні результати, а деталі методу - необов'язкові.

Далі подано версію стенда для отримання та аналізу розв'язання транспортних задач. Користувач легко може змінити матрицю вартостей перевезень, потреби

магазинів (попит) і можливості складів (пропозицію) - просто треба змінити елементи відповідних матриць. Дещо складніше змінити розмірність задачі - треба модифікувати відповідні матриці на іншу розмірність, модифікувати систему рівнянь і нерівностей блока оптимізації. В дійсності баланс попиту і пропозиції може порушуватись - на складах можуть бути залишки або запаси товарів, можуть бути періоди дефіциту. Попит може мати випадкові і детерміновані коливання. Далі будуть розглянуті більш реалістичні моделі.

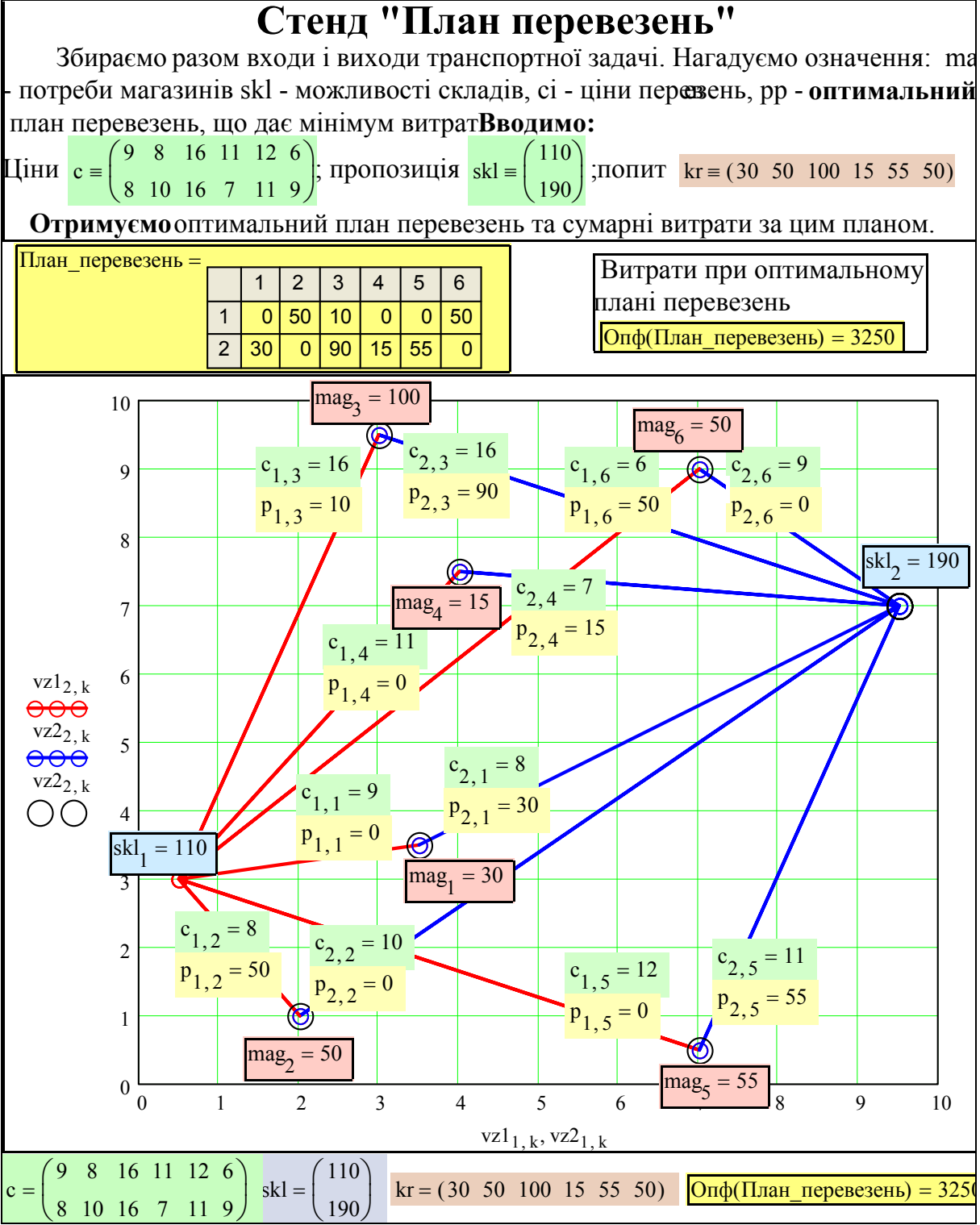


Рис. 2. Базовий варіант стенду для оптимізації та аналізу планів перевезень

В електронній версії те, що подано на рис.2 - не картинка, а працююча діалогова система для розв'язання певного класу задач. Називаємо таку екранну сторінку стендом, за аналогією з стендами для випробування механічних та електронних систем. На цій сторінці зібрані всі входи і виходи системи.

Вхідні дані:

- а) може вводити користувач;
- б) можуть вводиться з мереж різного рівня;
- в) можуть передаватись безпосередньо від стандартних передавачів.

Корпорація Matworks випустила на ринок продукт Mathworks Test & Measurement Suite, що забезпечує збір та обробку інформації від різних приладів та джерел. Ці дані без проблем адаптуються і обробляються в **прикладних програмах** (application, app, прикладная программа, приложение - створимо відповідний термін і в українській мові, інакше не матиме вона застосування), що створені в середовищі пакету Matlab цієї ж корпорації.

Розробка програми для аналізу чутливості планів перевезень до коливань цін і попиту

Для використання в умовах оперативного управління бажано мати версію системи для "що буде, якщо" аналізу (what if analysis). Можна виділити такі задачі:

1. Оцінка впливу варіацій окремих цін (5%, 10%, 25%...) на стійкість і витрати оптимального плану.
2. Оцінка впливу детермінованої зміни попиту (сезонні коливання, тренд цін) на стійкість і витрати оптимального плану.
3. Оцінка чутливості оптимального плану до випадкових коливань попиту.

Копіюємо та модифікуємо модуль оптимізації з попереднього розділу. Усі змінні даного документа - глобальні. Щоб виключити небажану "інтерференцію" між змінними розділів, змінюємо імена змінних. Задаємо:

матрицю цін: вектори потреб магазинів: та можливостей складів :

$$c_i = \begin{pmatrix} 10 & 12 & 10 & 11 & 12 & 6 \\ 8 & 10 & 8 & 7 & 11 & 9 \end{pmatrix}; \quad \text{Mag}^T = (30 \ 50 \ 100 \ 15 \ 55 \ 50); \quad \text{Skl} = \begin{pmatrix} 110 \\ 190 \end{pmatrix};$$

Крам := 6; Скл := 2; i := 1..Скл; j := 1..Крам

Задаємо цільову функцію. В даному випадку це сумарні витрати на перевезення. Ціль оптимізації - мінімізація цих витрат.

Зауваження. Зверніть увагу на відмінність запису цільової функції (замість посилення ми просто подамо попередній вираз тут поряд для порівняння). Можна бачити, що цільова функція у нас визначена як функція, що залежить не тільки від плану перевезень, а ще і від матриці цін.

беремо:

замість:

$$\text{For}(\text{Op}, c_i) := \sum_{j=1}^{\text{Крамниць}} \sum_{i=1}^{\text{Складів}} \text{Op}_{i,j} \cdot c_{i,j}$$

$$\text{Опф}(\text{Pr}) = \sum_{j=1}^{\text{Крамниць}} \sum_{i=1}^{\text{Складів}} \text{Pr}_{i,j} \cdot c_{i,j}$$

Зверніть увагу на те, що праві частини цих виразів однакові!

Задаємо початкові значення оптимізованих змінних - обсягів перевезень (Op - оптимальний план) :

$$\text{Op}_{i,j} := 1$$

Записуємо саму оптимізаційну задачу.

Given

Умови задоволення потреб магазинів. В даному випадку їх 6.

$$\begin{aligned} \text{Op}_{1,1} + \text{Op}_{2,1} &= \text{mag}_1 ; & \text{Op}_{1,3} + \text{Op}_{2,3} &= \text{mag}_3 ; & \text{Op}_{1,5} + \text{Op}_{2,5} &= \text{mag}_5 ; \\ \text{Op}_{1,2} + \text{Op}_{2,2} &= \text{mag}_2 ; & \text{Op}_{1,4} + \text{Op}_{2,4} &= \text{mag}_4 ; & \text{Op}_{1,6} + \text{Op}_{2,6} &= \text{mag}_6 . \end{aligned}$$

Умови балансу постачання для кожної бази постачання. В даному випадку баз - 2.

$$\begin{aligned} \text{Op}_{1,1} + \text{Op}_{1,2} + \text{Op}_{1,3} + \text{Op}_{1,4} + \text{Op}_{1,5} + \text{Op}_{1,6} &= \text{skl}_1 ; \\ \text{Op}_{2,1} + \text{Op}_{2,2} + \text{Op}_{2,3} + \text{Op}_{2,4} + \text{Op}_{2,5} + \text{Op}_{2,6} &= \text{skl}_2 . \end{aligned}$$

Обмеження здорового глузду - змінні не можуть бути від'ємними

$$\begin{aligned} \text{Op}_{1,1} \geq 0 ; & \text{Op}_{1,2} \geq 0 ; & \text{Op}_{1,3} \geq 0 ; & \text{Op}_{1,4} \geq 0 ; & \text{Op}_{1,5} \geq 0 ; & \text{Op}_{1,6} \geq 0 ; \\ \text{Op}_{2,1} \geq 0 ; & \text{Op}_{2,2} \geq 0 ; & \text{Op}_{2,3} \geq 0 ; & \text{Op}_{2,4} \geq 0 ; & \text{Op}_{2,5} \geq 0 ; & \text{Op}_{2,6} \geq 0 ; \end{aligned}$$

Після цього записуємо вираз, що вказує, а) що робити (мінімізація); б) для якої цільової функції (вона визначена вище); г) по яких змінних. Зверніть увагу на те, що ми пишемо тільки ім'я цільової функції. Знову порівнюємо вирази

беремо:

замість:

$$\boxed{\text{Opp}(\text{ci}) := \text{Minimize}(\text{Fop}, \text{Op})} \quad \left(\begin{array}{cccccc} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} & p_{1,5} & p_{1,6} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & p_{2,5} & p_{2,6} \end{array} \right) = \text{Minimize}(\text{Opp}, \text{Op}).$$

В першому виразі ліва частина - ім'я функції цін, в другому - ім'я змінної. В даному випадку, пакет дозволяє визначити функцію користувача в термінах блока оптимізації. Звернемо увагу на синтаксис виразів. В першому - ліва частина - матриця з дванадцяти ($2 \times 6 = 12$) функцій і кожна така функція залежить від 12 змінних - цін перевезень. За цим коротким виразом стоїть декілька сторінок програмного коду на VC++. В другому виразі змінна лівої частини - матриця, що подана в розгорнутому вигляді. Тут ми ясно бачимо "колективне присвоєння". Останнє питання - звідки блок Minimize() знає, що повертати - матрицю, вектор, і якої розмірності?

Повернемося на початок розділу і побачимо, як послідовно визначені функція Fop() і матриця Op - саме вони і задають структуру і параметри оптимізаційної задачі. У підсумку бачимо, що ми написали для нашої задачі все те ж саме, що потрібно було писати в програмах на Паскалі чи C++. Фактично ми займалися непрямым програмуванням на C++. Це суттєво зменшило витрати часу на програмування, але дуже бажано знати мову на якій написано пакет - це дозволяє в складних випадках діяти не наосліп. Слід пам'ятати, що не всі синтаксично коректні вирази будуть виконуватися. Останнє слово за - випробуванням. А тепер скомпонуємо входи і виходи задачі в зручному для користувача інтерфейсі.

Введення номінальних значень параметрів для задачі "План перевезень"

Позначення змінних : mag - потреби магазинів skl - можливості складів, ci - ціни перевезень, pp - **оптимальний** план перевезень, що дає мінімум витрат. **Вводимо:**

Ціни $\text{ci} = \begin{pmatrix} 10 & 12 & 10 & 11 & 12 & 6 \\ 8 & 10 & 8 & 7 & 11 & 9 \end{pmatrix}$; пропозиція $\text{skl} = \begin{pmatrix} 110 \\ 190 \end{pmatrix}$; попит $\text{Kr} = (30 \ 50 \ 100 \ 15 \ 55 \ 50)$

Отримуємо оптимальний план перевезень та сумарні витрати за цим планом

Opp(c _i) =						
	1	2	3	4	5	6
1	5	0	0	0	55	50
2	25	50	100	15	0	0

Витрати при оптимальному плані перевезень
 $Fop(Opp(c_i), c_i) = 2615$

Тепер на базі **функціональної залежності оптимального плану від матриці цін перевезень** користувач може самостійно будувати персональний набір модулів "що буде якщо" для ситуацій постачання, що зустрічаються у його практиці. Розглянемо це на прикладі, що є елементом допомоги користувачу.

Приклад

Сценарій: Постачання ведеться з двох баз. Ціни перевезень з другої бази мають коливання. Знайти, при яких відхиленнях ціни зміниться оптимальний план перевезень. Задаємо зміну цін для перевезень з другого складу. Формуємо матриці цін для випадків зменшення C_n і збільшення C_v на величину Δc . $\Delta c := 0.35$.

$$C_o := c_i ; \quad C_{n,i,j} := c_{i,j} \cdot [1 - \Delta c \cdot (i > 1)] ; \quad C_{v,i,j} := c_{i,j} \cdot [1 + \Delta c \cdot (i > 1)] .$$

Формуємо блок вихідних даних для аналізу.

Opp(C _n) =						
	1	2	3	4	5	6
1	30	0	30	0	0	50
2	0	50	70	15	55	0

Витрати при оптимальному плані перевезень та зниженні цін по 2-му складу на $\Delta c = 35\%$

$$\frac{Fop(Opp(C_n), C_n)}{Fop(Opp(c_i), c_i)} = 78\%$$

Opp(c_i) =						
	1	2	3	4	5	6
1	5	0	0	0	55	50
2	25	50	100	15	0	0

Витрати при оптимальному плані перевезень номінальні
 $Fop(Opp(c_i), c_i) = 2615$ $\Delta c = 0.35$

Opp(C_v) =						
	1	2	3	4	5	6
1	0	5	0	0	55	50
2	30	45	100	15	0	0

Витрати при оптимальному плані перевезень та підвищення цін по 2-му складу на $\Delta c = 35\%$

$$\frac{Fop(Opp(C_v), C_v)}{Fop(Opp(c_i), c_i)} = 121\%$$

Інструкція користувачу: змінюйте параметр Δc поки плани для нижнього і верхнього відхилення цін не зміняться відносно номінального плану. В даному випадку **план стійкий при змінах цін перевезень з другого складу в діапазоні -20% - +35%**

Контрольні запитання

1. Сформулюйте транспортну задачу: обмеження, критерій, змінні оптимізації, ціль.
2. Запишіть задачу лінійного програмування (ЛП) у матрично-векторному вигляді.
3. Скільки обмежень має транспортна задача, що характеризують ці обмеження?
4. Для розглянутої транспортної задачі запишіть двоїсту задачу.
5. Прокоментуйте вираз $Opp(c_i) := Minimize(Fop, Op)$ - синтаксис, семантику, прагматику.
6. Як визначити функцію користувача через блок оптимізації пакета?
7. Які товари перевозяться зі складів у магазини?



Рис.1 "Каков стол, таков стул"[тнп].

4.4. Розробка системи для розв'язання задач лінійного програмування альтернативними методами

Вступ. Постановка задачі

В попередньому розділі розглянуто задачу оптимізації плану перевезень **столів і стільців**. А тепер розглянемо виробництво цих товарів. На цьому класичному прикладі подається набір з альтернативних методів і базових модулів розв'язання задач лінійного програмування (ЛП). Додаткове призначення цього документа - навчання користувача автоматизованих систем підтримки рішень.

Нагадаємо, що сучасні системи підтримки рішень призначені не для "оператора АСУ", а для користувача - професіонала, менеджера, що бажає працювати з прозорою системою, мати доступ до кожного рівняння і можливість все переробляти під свої задачі, бачити не тільки числа, але й візуальну картину задачі та її розв'язання.

Цілі цього розділу:

Освоєння методів оптимізації на базі теорії нечіткості.

Освоєння сучасних інформаційних технологій, зокрема: багатоверсійного програмування, створення програм на базі бібліотечних модулів.

Розробка програмної системи класу "що буде якщо" для аналізу і розв'язання задачі лінійного програмування.

Використання теорії двоїстості для аналізу практичних задач.

Ревізія значення ЛП для сучасних науки і практики.

Задачі лінійного програмування зустрічаються у практичній діяльності спеціалістів будь-якого профілю. В ЛП склалися певні класи задач - транспортна, виробнича, визначення оптимального раціону та ін. В цьому розділі аналіз і розв'язання задачі ЛП виконується на класичному (традиційному) прикладі. Подамо описання цієї конкретної задачі.

Деяка фірма постачає два ($i = 1..2$, задача потім масштабується) види продукції - столи і стільці. Для виробництва використовуються три ($j = 1..3$) види ресурсів (сировини і комплектуючих), наприклад, - дуб, лак, цвяхи.

Задано **технологічну матрицю А**, елемент якої $A_{i,j}$ визначає, скільки i -го ресурсу потрібно для виготовлення одиниці виміру j -ої продукції.

Задані **обмеження** по кожному виду ресурсів (B_1, B_2, B_3) - вектор запасів ресурсів.

Задано **ціни** одиниці кожного виду продукції - стола, стільця - (C_1, C_2) - вектор цін.

Задано **цільову функцію** - сумарний дохід:

$$S_{\text{дох}}(x_1, x_2) := C_1 \cdot x_1 + C_2 \cdot x_2,$$

де x_1, x_2 - кількості виготовлених столів і стільців.

Ціль оптимізації - знайти план випуску ($x_{1\text{оп}}, x_{2\text{оп}}$), що задовольняє обмеження по ресурсах і дає максимальний сумарний дохід.

Наступна сторінка – **копія екрану** з робочого документу

Розробка стенду для графічного розв'язання задачі

Записуємо вхідні дані згідно з постановкою задачі ЛП.

Технологічна матриця: наявні ресурси: ціни: цільова функція (критерій):

$$A = \begin{bmatrix} & 0 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 3 \\ 2 & 1 & 1 \end{bmatrix} ; \quad B = \begin{bmatrix} & 0 \\ 0 & 20 \\ 1 & 22 \\ 2 & 9 \end{bmatrix} ; \quad C = \begin{bmatrix} & 0 \\ 0 & 3 \\ 1 & 2 \end{bmatrix} ; \quad \boxed{S_{\text{dox}}(x_1, x_2) := C_1 \cdot x_1 + C_2 \cdot x_2}$$

Обмеження по ресурсах:

$$A_{1,1} \cdot x_1 + A_{1,2} \cdot x_2 \leq B_1, \quad A_{2,1} \cdot x_1 + A_{2,2} \cdot x_2 \leq B_2, \quad A_{3,1} \cdot x_1 + A_{3,2} \cdot x_2 \leq B_3.$$

Обмеження здорового глузду: $x_1 \geq 0, \quad x_2 \geq 0$.

Обмеження записуємо як рівності, розв'язуємо їх і цільову функцію відносно змінної x_2 - кількості стільців. Отримуємо такі функції від x_1 - кількості столів.

$$x_1 := 0, 0.2 \dots 25 \quad \text{Dox}(x_1) := (S_{\text{dxcg}} - C_1 \cdot x_1) \div C_2$$

$$\text{Обме1}(x_1) := \frac{(B_1 - A_{1,1} \cdot x_1)}{A_{1,2}} ; \quad \text{Обме2}(x_1) := \frac{(B_2 - A_{2,1} \cdot x_1)}{A_{2,2}} ; \quad \text{Обме3}(x_1) := \frac{(B_3 - A_{3,1} \cdot x_1)}{A_{3,2}}.$$

Формуємо границю області (= симплекс), точки якої (x_1, x_2) задовольняють усі обмеження.

$$\text{Обмс}(x_1) := \min(\text{Обме1}(x_1), \text{Обме2}(x_1), \text{Обме3}(x_1)).$$

Для побудови тривимірних графіків вводимо дискретні змінні $X1_i$ та $X2_j$.

$$\text{Діапазон} := 50 ; \quad i := 1 \dots \text{Діапазон} ; \quad j := 1 \dots \text{Діапазон} ; \quad h := .15 ; \quad X1_i := h \cdot i ; \quad X2_j := h \cdot j.$$

Формуємо дискретну функцію двох змінних - цільову функцію "обрізану" обмеженнями.

$$\text{Dox}_{i,j} := \text{if}[(A_{1,1} \cdot X1_i + A_{1,2} \cdot X2_j \leq B_1) \cdot (A_{2,1} \cdot X1_i + A_{2,2} \cdot X2_j \leq B_2) \cdot (A_{3,1} \cdot X1_i + A_{3,2} \cdot X2_j \leq B_3), S_{\text{dxcg}}(X1_i, X2_j), 0]$$

$$x1_{\text{op}} \equiv 5.4 ; \quad x2_{\text{op}} \equiv 3.7 ; \quad S_{\text{dxcg}} \equiv 23.5 ; \quad xc1_{i,j} := i ; \quad xc2_{i,j} := \left(\frac{S_{\text{dxcg}}}{h} - C_1 \cdot xc1_{i,j} \right) \div C_2 ; \quad Z_{i,j} := j \cdot .5 ;$$

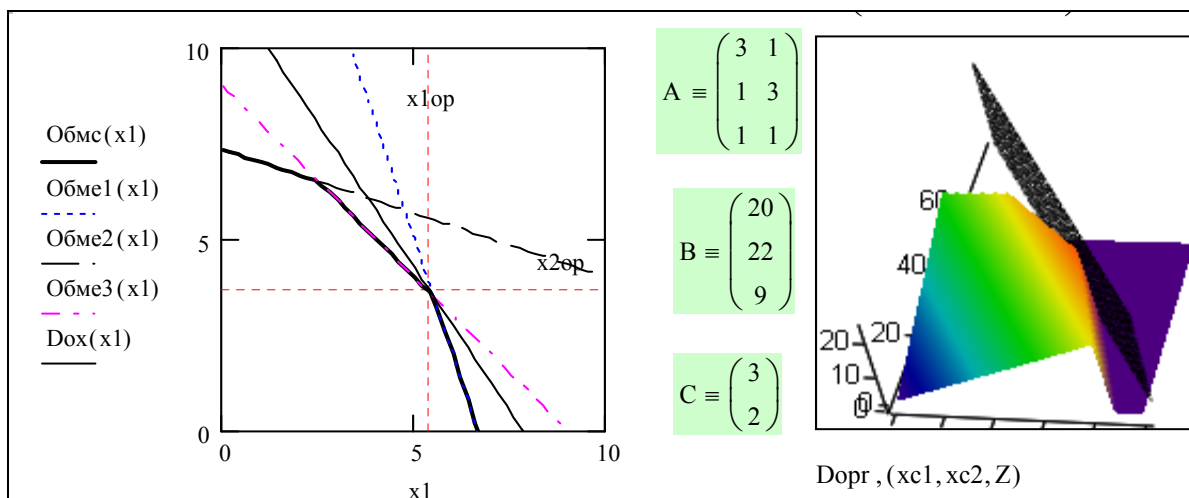


Рис. 1. Графічне подання задачі ЛП

Електронні і паперові книги мають певні відмінності, зокрема графічні елементи електронних книг - це не рисунки в тексті паперових книг, а органічні елементи (=об'єкти) певної об'єктної програми. Графіки не можна довільно переставляти відносно інших об'єктів. Те, що подано вище, - **стенд**. В електронній версії ви можете змінювати параметри задачі і одразу бачити наслідки. Не подаємо легенди на кожну залежність на графіках - дуже корисно самотійно розібратись, де яка залежність. Ясно, що товста лінія - симплекс з обмежень, що і як він обмежує, видно на правому графіку, де по осях x та y відкладаються x_1 та x_2 , а по осі z - відповідні значення цільової функції = сумарного доходу. На правому графіку подано також площину "постійне значення цільової функції", що є дотичною до симплексу. Точка дотику - оптимум.

Інструкція користувача стенду

Стабільно популярні теми гумору - "інструкція з користування молотком" та ін. І до нашого стенду потрібна інструкція.

П1. Змінюй параметр Sd_{xg} (дохід), поки графік цільової функції не стане дотикатись до симплекса.

П2. Підбери значення x_{1op} та x_{2op} (кількості столів та стільців) так, щоб горизонтальний і вертикальний маркери проходили через точку дотику. Отримані значення цих трьох параметрів (Sd_{hg} , x_{1op} , x_{2op}) і є розв'язком задачі оптимізації.

П3. Напиши програму ("робот"), яка зробить все це без тебе.

Розв'язання задачі ЛП числовим методом

Сьогодні методи розв'язання оптимізаційних задач, взагалі, і задачі ЛП, зокрема, часто вбудовуються в математичні пакети. Від користувача вимагається **коректно записати свою задачу** і сподіватися на успіх, який не гарантовано (методи пошуку екстремуму не завжди працюють). Використаємо можливості пакета. Копіюємо готовий блок - фрейм розв'язання задачі і налаштуємо його.

Записуємо функцію, екстремум якої шукаємо:		$D(xx1, xx2) := C_1 \cdot xx1 + C_2 \cdot xx2$	
Записуємо початкові значення шуканих змінних:		$xx1 := 1$	$xx2 := 1$
Після ключового слова <i>Given</i> записується система обмежень (рівностей, нерівностей)			
Given	$A_{1,1} \cdot xx1 + A_{1,2} \cdot xx2 \leq B_1$ $A_{2,1} \cdot xx1 + A_{2,2} \cdot xx2 \leq B_2$ $A_{3,1} \cdot xx1 + A_{3,2} \cdot xx2 \leq B_3$		
і отримуємо (не завжди) розв'язок:		$\begin{pmatrix} x_{1opd} \\ x_{2opd} \end{pmatrix} := \text{Maximize}(D, xx1, xx2)$	
$Sd_{xd} := Sd_{ox}(x_{1opd}, x_{2opd})$	$x_{1opd} = 5.5$	$x_{2opd} = 3.5$	$Sd_{xd} = 23.5$

Зауваження. Те, що вище взято в рамку - це програмний модуль. Його можна копіювати, переносити з документа в документ. Рамки і виділення кольором - це стандарт усіх сучасних продуктів. В логіко-семантичному аспекті - це інверсія відносно традиційного подання текстів (лістингів) програм - там в тест програми вкраплювались коментарі, тут в текст коментарю "вкраплюються" елементи програми.

Нагадаємо: стандарт програмування в C++ дозволяє вводити користувачу свої оператори та ін. Потім ці "спецслова" обробляються транслятором у стандартний текст C++. Все призначено для того, щоб створювати тексти програм, що подібні до звичайних текстів - для людини, а не для машини.

Знову ж нагадаємо, що поданий вище блок - "верхівка айсберга" - під ним модулі, що написані на C++. В цілому ж повна програма цього модуля на 70% складається з викликів функцій Windows, на 25% - виклики функцій Mathcad, і останні 5% - специфіка нашого модуля - параметри, описання розташування елементів та ін.

Розв'язання задачі ЛП методом нечіткої оптимізації

Ми зневажливо відносимось до теорії „теплороду”, середньовічних методів множення чисел, астрології, хіромантії і не помічаємо, що місце розв'язаних забобів і ритуалів займають нові. Більшість постановок оптимізаційних задач і методів їх розв'язання, є або а) штучними, але корисними (як теорія „теплороду”), або некорисними, але звичними.

У ранньому середньовіччі не багато студентів освоювали множення "корабликом", сьогодні не більше студентів освоюють симплекс-метод.

А чи не міг Нобелівський Лауреат академік Канторович трошки інакше поставити задачу ЛП і використати інший метод? Розглянемо методику постановки і розв'язання задачі ЛП як задачі нечіткої оптимізації. Зауважимо, що успіх, продуктивність, довге життя певного наукового напрямку залежить і від його назви.

Теорія нечітких множин теж перестане бути модою, даватиме стійкі практичні результати, коли для неї знайдуть відповідну назву, наприклад, "фузня" (від fuzzy theory, співзвучно терміну "знання", „моделивання”). Сьогодні вживається термін "фазифікація обмежень", що є калькою з англійського терміна. А можна сказати "розмивання обмежень"?

Поглянемо на обмеження нашої задачі ЛП з точки зору практики: чи збанкрутує фабрика, якщо замість $B = 100$ одиниць 1-го ресурсу у виробництві буде використано 97 або 103 одиниці? Може й збанкрутує, може стане корпорацією, але в даному конкретному випадку немає фізичних законів, що забороняють "розмити" обмеження. В реальному житті малі відхилення (100 ± 3) ніяк не позначаються на результатах бізнесу, середні (100 ± 20) - змінять доход, а великі (100 ± 70) – можуть звести до нуля.

Зробимо вирішальний крок - замінимо **"жорстке" обмеження на "функцію штрафу"**. Розглянемо це на прикладі. На рис.2 подано графік цільової функції $u = k \cdot x$, "обрізаной" чітким обмеженням $x \leq x_1$ і два варіанти "розмивання обмеження" – симетрична функція штрафу, що карає і за перевищення і за недобір обмеження і несиметрична функція штрафу, яка "карає" тільки за перевищення обмеження. Нормуємо функції штрафу так, щоб для випадку $x = x_1$ вони дорівнювали 1. Нульове значення функції штрафу означає "дане значення x є неприпустимим".

Зробимо наступний крок – визначимо **"розмиту" цільову функцію**. Для цього треба логічно скомбінувати цільову функцію і розмиті обмеження – функції штрафу. Перше, що приходить у голову – взяти добуток цільової функції і нечітких обмежень. На рис.3 нижче подано розмиті цільові функції, їх називають **"функції належності цілі"**.

Ступінь розмитості можна змінювати – від дуже малої, тоді ми маємо задачу з чіткими обмеженнями, до великої. Розмитість може визначатись специфікою задачі - реальними штрафами за недопостачання, та кредитними відсотками, а може бути просто параметром процесу оптимізації – збільшення розмитості покращує сходиність пошукових методів, але і зменшує точність знаходження екстремуму.

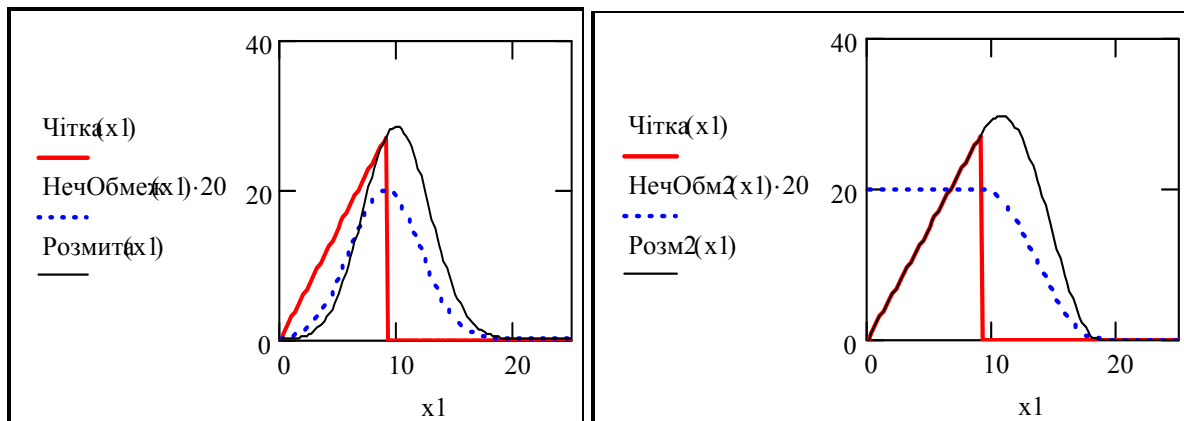


Рис. 2. Альтернативи розмивання обмежень

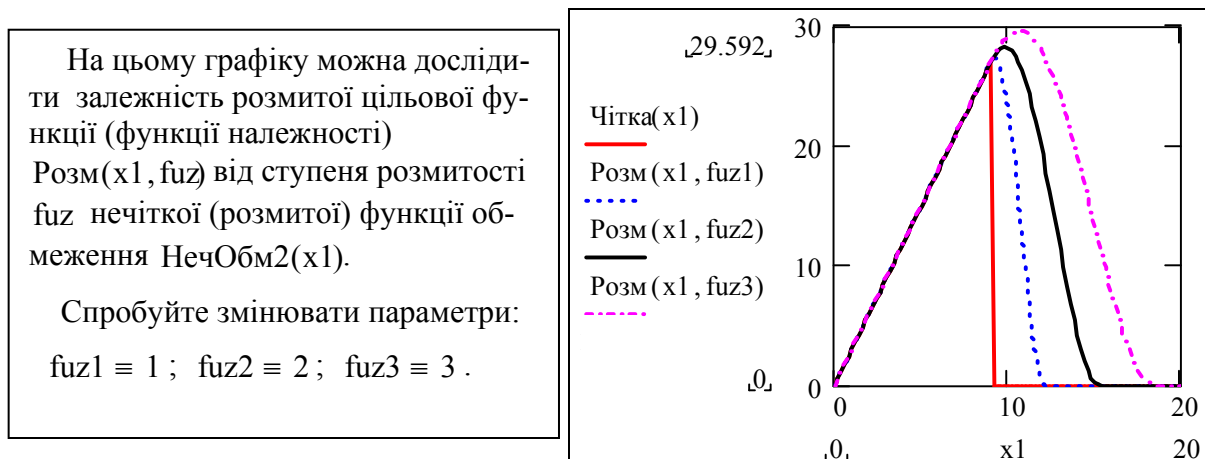


Рис.3. Функція належності цілі при різних рівнях розмитості

Не усі ентузіасти нечітких множин і нечіткої логіки знають, що - **теорія нечіткості не є логічно завершеною** - в ній не визначені однозначно правила роботи з функціями належності. Кожен раз вони конструюються виходячи із специфіки задачі, здорового глузду і формальної логіки. В поданих вище графіках функції належності сконструйовані з функції гаусового розподілу $\text{dnorm}(\cdot)$ та $\cos(\cdot)$. Ці функції закриті (без пароллю) і доступні в електронній версії книги.

В теорії нечітких множин існують операції над множинами, аналоги, відповідних операцій над звичайними множинами – об'єднання нечітких множин (диз'юнкція), перетин множин (кон'юнкція), належність, або включення та ін. **Загальноприйнятого єдиного визначення для цих відношень поки немає.**

На рис.2 подано дві версії функція належності обмеження. Одна з них "карає" як за перевищення обмеження, так і за зменшення; друга "карає" тільки за перевищення. Яку з них вибрати – залежить від техніко-економічної специфіки задачі. Звернемо увагу на те, що екстремум функції належності цілі (рис.3) не збігається з екстремумом "чіткої" цільової функції. Неформальне обґрунтування цього просте: вигідно брати більше ресурсу і виробляти більше продукції, поки розмір штрафу не стане більшим виграшу.

Нагадаємо, що терміни "функція належності обмеження" і "функція штрафу" (за відхилення від обмеження) приблизно еквівалентні. Хто в реальній ситуації накладає ці штрафи? В ринковій економіці – ринок: беремо більше ресурсу, – потрібні додаткові кошти - кредити на закупівлю, додаткова продукція "збиває" вниз ціни. В адміністративній економіці – адміністрація.

Переведення задачі ЛП у нечіткий формат

Формуємо функції належності для кожного обмеження, а потім функцію належності усіх обмежень. Беремо вхідний матеріал - відповідні "чіткі" функції:

$$\begin{aligned} \text{Обмеж1}(x_1) &:= \frac{(B_1 - A_{1,1} \cdot x_1)}{A_{1,2}} ; & \text{Обмеж2}(x_1) &:= \frac{(B_2 - A_{2,1} \cdot x_1)}{A_{2,2}} ; \\ \text{Обмеж3}(x_1) &:= \frac{(B_3 - A_{3,1} \cdot x_1)}{A_{3,2}} ; & \text{Sdox}(x_1, x_2) &:= C_1 \cdot x_1 + C_2 \cdot x_2 . \end{aligned}$$

Записуємо рядок параметрів (кількість кроків, кроки обчислень) для відповідних дискретних функцій:

$$\text{Диап} := 40 ; \quad i := 1.. \text{Диап} ; \quad h := 0.4 \max(B) \div \text{Диап} ; \quad X1_i := h \cdot i ; \quad X2_j := h \cdot j .$$

Формуємо дискретну цільову функцію "обрізану" "жорсткими" обмеженнями:

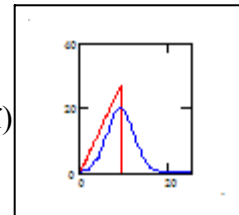
$$\text{Dop}_{i,j} := \text{if} \left[(A_{1,1} \cdot X1_i + A_{1,2} \cdot X2_j \leq B_1) \cdot (A_{2,1} \cdot X1_i + A_{2,2} \cdot X2_j \leq B_2) \cdot (A_{3,1} \cdot X1_i + A_{3,2} \cdot X2_j \leq B_3), \text{Sdox}(X1_i, X2_j), 0 \right]$$

Записуємо функції належності обмежень (= переводимо обмеження у нечіткий формат, або: "виконуємо фазифікацію обмежень")

Альтернатива 1. Симетрична функція належності.

Використаємо таку функцію належності обмежень $\implies$

Реалізуємо її (мають значення тільки топологічні властивості функції) на базі гаусового розподілу, а мірою нечіткості буде змінна fuzь - дисперсія розподілу. Зокрема, при $\text{fuzь} = 0$ маємо чітке обмеження.



Розмиваємо обмеження:

$$\text{НенормованеНечіткеОбмеження1}(x_1, x_2) := \text{dnorm}(B_1 - A_{1,1} \cdot x_1 - A_{1,2} \cdot x_2, 0, \text{fuzь}) ;$$

$$\text{НенормованеНечіткеОбмеження2}(x_1, x_2) := \text{dnorm}(B_2 - A_{2,1} \cdot x_1 - A_{2,2} \cdot x_2, 0, \text{fuzь}) ;$$

$$\text{НенормованеНечіткеОбмеження3}(x_1, x_2) := \text{dnorm}(B_3 - A_{3,1} \cdot x_1 - A_{3,2} \cdot x_2, 0, \text{fuzь} \cdot 0.5) ;$$

Дискретизуємо і нормуємо ці функції так, щоб максимуми були одиничними.

$$\text{NNO1}_{i,j} := \text{НенормованеНечіткеОбмеження1}(X1_i, X2_j) ; \quad \max(\text{NNO1}) = 0.199 ;$$

$$\text{NNO2}_{i,j} := \text{НенормованеНечіткеОбмеження2}(X1_i, X2_j) ; \quad \max(\text{NNO2}) = 0.199 ;$$

$$\text{NNO3}_{i,j} := \text{НенормованеНечіткеОбмеження3}(X1_i, X2_j) ; \quad \max(\text{NNO3}) = 0.399 ;$$

$$\text{NO1} := \text{NNO1} \div \max(\text{NNO1}) ; \quad \text{NO2} := \text{NNO2} \div \max(\text{NNO2}) ;$$

$$\text{NO3} := \text{NNO3} \div \max(\text{NNO3}) ; \quad \max(\text{NO3}) .$$

Виконуємо згортку **функції належності обмежень**. Для нашої задачі ЛП вони повинні виконуватись усі одночасно (кон'юнкція). Це можна зробити по різному.

Перша альтернатива згортки - перемножити функції належності.

Це можна записати так:

або так: (з використанням векторизації)

$$\text{ОбфузC}_{i,j} = \frac{(\text{NO1}_{i,j} \cdot \text{NO2}_{i,j} \cdot \text{NO3}_{i,j})}{\max(\text{ОбфузC})}; \quad \text{ОбфузC} := \frac{\overrightarrow{(\text{NO3} \cdot \text{NO1} \cdot \text{NO2})}}{\max(\overrightarrow{(\text{NO1} \cdot \text{NO2} \cdot \text{NO3})})}.$$

Записуємо функцію належності цілі:

$$\boxed{\text{Критфуз}_{i,j} := \text{ОбфузC}_{i,j} \cdot \text{Sdox}(X1_i, X2_j)}$$

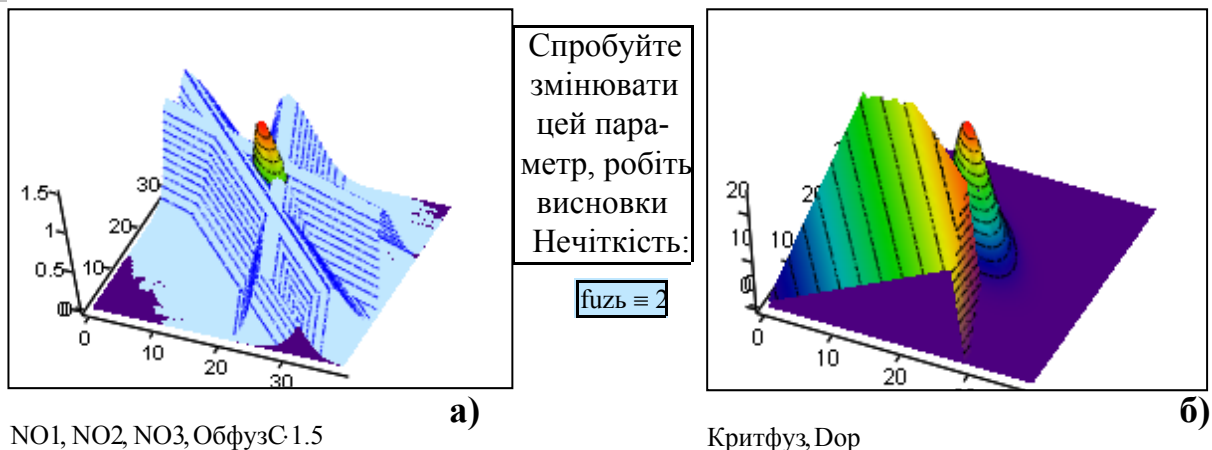


Рис. 4. Порівняння чіткого і розмитого розв'язання - альтернатива 1

Проаналізуємо те, що бачимо на графіках. На рис.4а подано розмиті обмеження і згортку цих обмежень - нечітку кон'юнкцію. Бачимо, що функція належності цілі має максимум не на границі симплексу, а в "клітці" - перетині розмитих обмежень. На рис.4б побудовано поверхні чіткої та нечіткої цільової функції. Не збігаються вони, але, так і повинно бути - ми взяли симетричну функцію належності, що накладає "штраф" не тільки за перебільшення обмеження, але і за зменшення.

Це не помилки - ми просто поставили задачу з іншими критеріями "бажано - небажано", з іншою "системою цінностей" і отримали дещо інше розв'язання. Виникає чітка підозра, що класичні постановки задач ЛП є менш реалістичними, ніж "нечіткі". Обидві задачі коректні відносно своїх умов.

Що робити менеджеру? - вибрати ту трактовку задачі, що більше відповідає реальній ситуації - йому ж відповідати за наслідки. Саме це ми мали на увазі, коли говорили на початку розділу про ревізію підвалин лінійного програмування.

Друга альтернатива згортки - в кожній точці (x1, x2) простору рішень брати мінімум з трьох функцій:

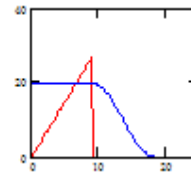
$$\boxed{\text{ОбфузC}_{i,j} := \min(\text{NO1}_{i,j}, \text{NO2}_{i,j}, \text{NO3}_{i,j})}$$

Ця альтернатива згортки дає приблизно такі ж результати. Перевірте **самостійно**. Для цього достатньо замінити функцію $\text{ОбфузC}_{i,j}$ у відповідному місці документа.

Подана далі сторінка – копія екрану з електронної книги.

Альтернатива 2. Несиметрична функція належності.

Використаємо таку функцію належності обмежень $\Rightarrow$
 Реалізуємо її на базі тригонометричної функції (це майже не приципово - значення мають топологічні властивості функції), а мрою нечіткості буде змінна $fyз$ - період синусоїди.



Як можна бачити з рисунку, ця функція "штрафує" тільки за перевищення обмеження. Тобто тут логіка ближче до логіки, що використана у класичній постановці задачі.

$$\text{НечОб}l(x1, x2) := \begin{cases} \text{Тьфу} \leftarrow B_1 - A_{1,1} \cdot x1 - A_{1,2} \cdot x2 \\ 1 & \text{if } \text{Тьфу} > 0 \\ [0.5(1 + \cos(\text{Тьфу} \div \text{фюз}))] & \text{if } -\pi < \frac{\text{Тьфу}}{\text{фюз}} \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Нижче закрито (сколапсовано) без пароля функції належності обмежень. Вони такі ж, як і пода на ліворуч. Можна винайти щось більш природне.



Дискретизуємо і нормуємо ці функції, але нормування в даному випадку зайве, функції належності згідно з визначенням мають одиничні максимальні значення.

$$\text{Критфуз}H(x1, x2) := \text{НечОб}l(x1, x2) \cdot \text{НечОб}2(x1, x2) \cdot \text{НечОб}3(x1, x2) \cdot \text{Sдох}(x1, x2)$$

$$\text{NNO1}_{1,j} := \text{НечОб}1(X1_i, X2_j) \quad \text{NNO2}_{1,j} := \text{НечОб}2(X1_i, X2_j) \quad \text{NNO3}_{1,j} := \text{НечОб}3(X1_i, X2_j)$$

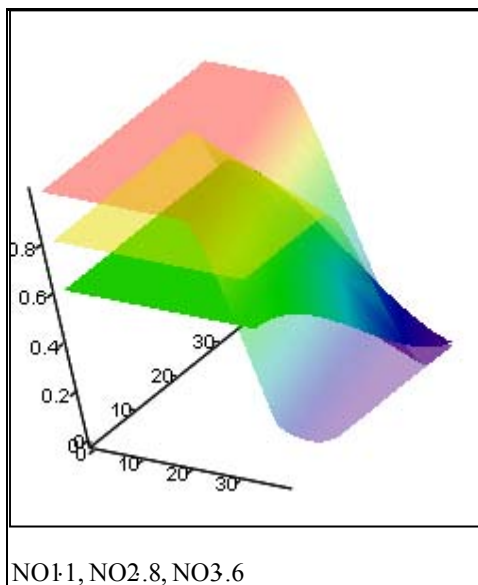
$$\text{NO1} := \text{NNO1} \div \max(\text{NNO1}) \quad \text{NO2} := \text{NNO2} \div \max(\text{NNO2}) \quad \text{NO3} := \text{NNO3} \div \max(\text{NNO3})$$

$$\text{Контроль: } \max(\text{NNO1}) = 1 \quad \max(\text{NNO2}) = 1$$

Записуємо загальну функцію належності обмежень (її теж можна не нормувати)

$$\text{ОбфузC} := \frac{\text{NO1} \cdot \text{NO2} \cdot \text{NO3}}{\max(\text{NO1} \cdot \text{NO2} \cdot \text{NO3})}$$

$$\text{I, нарешті, записуємо функцію належності ці} \quad \text{Критфуз}_{1,j} := \text{ОбфузC}_{1,j} \cdot \text{Sдох}(X1_i, X2_j)$$



Спробуйте змінювати цей параметр, робіть висновки Нечіткість:

$$\text{фюз} = 3$$

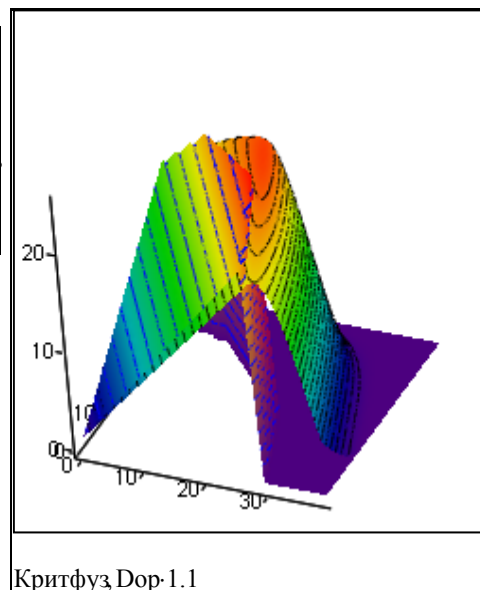


Рис. 5. Порівняння чіткого і розмитого розв'язання - альтернатива 2

Проаналізуємо те, що бачимо на рис. 5. На рис.5а подано розмиті обмеження. На рис.5б подано "чітку" цільову функцію з урахуванням чітких обмежень "менше можна – більше не можна" і нечітку цільову функцію, що є кон'юнкцією усіх обмежень і функції цілі (сумарного доходу). На рис. 6 подано вигляд цільової функції при різних значеннях параметра розмитості. Бачимо – при малих значеннях розмитості приходимо до чіткої задачі, отримаємо те ж саме розв'язання. При великій розмитості розв'язання **суттєво не збігається** із розв'язанням відповідної чіткої задачі.

Інструкція користувача: змінюй параметр розмитості f_{yz} від 0.1 до 7. Дивись на результати з різних точок зору. Для цього мишею повертай графіки.

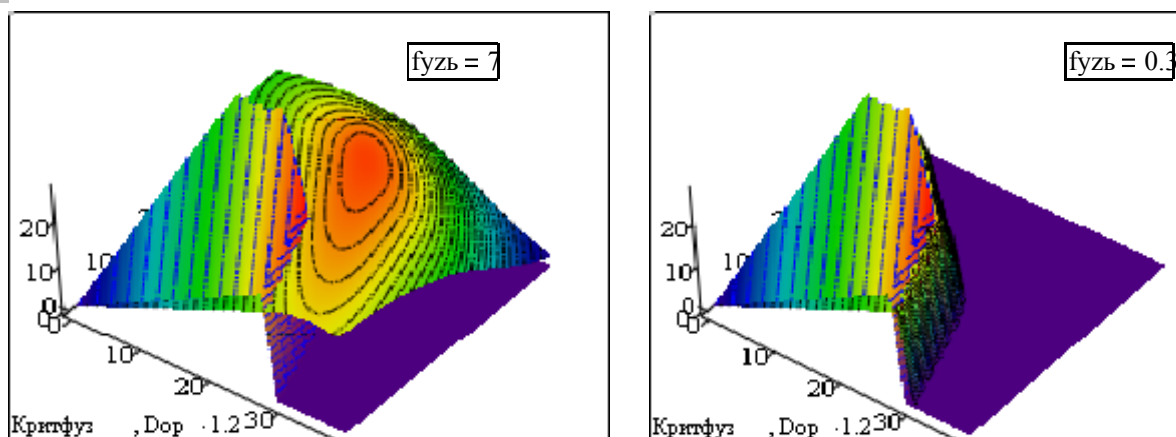


Рис. 6. Вплив ступеня розмитості на нечітке розв'язання

Висновки

Ми проаналізували дві альтернативи переведення задачі ЛП у нечіткий формат. Тепер ми повинні визначити, який з них є коректним. Неважко бачити, що обидві альтернативи коректні - кожна для своєї постановки задачі. Отримали те, що і повинні отримати: при зменшенні рівня розмитості її розв'язання сходиться до розв'язання "чіткої задачі". Отримали додатковий результат - задача ЛП - занадто далека від реальності.

А тепер обчислимо оптимальні плани за альтернативними методами.

Записуємо функцію, екстремум якої шукаємо: $D(x1, x2) := \text{КритфузН}(x1, x2)$

Записуємо початкові значення шуканих змінних: $xx1 := 1 \quad xx2 := 1$

і отримуємо (не завжди) розв'язок: $\begin{pmatrix} x1_{\text{опн2}} \\ x2_{\text{опн2}} \end{pmatrix} := \text{Maximiz}(D, xx1, xx2)$

$Sdxn2 := Sdox(x1_{\text{опн2}}, x2_{\text{опн2}}) \quad x1_{\text{опн2}} = 5.1 \quad x2_{\text{опн2}} = 4.9 \quad Sdxn2 = 25.3$

Тут ми використали (скопювали і модифікували) блок оптимізації, за допомогою якого ми знайшли розв'язання для "чіткої" задачі ЛП.

Завдання. Порівняйте той і цей блоки, знайдіть, в чому їх відмінність, поясніть чому.

Підсумки

Формуємо і виводимо таблицю результатів оптимізації. Якщо ви не знаєте як виробити 4,1 стола, а) займіться цілочисленим програмуванням; б) займіться раціоналізацією технологій, так щоб при даних обмеженнях виробити 5 столів.

Результати_оптимізації=				
	1	2	3	4
1	"Метод оптим."	"столів"	"стілців"	"доход"
2	"Графічний"	5.4	3.7	23.5
3	"Розмитий"	5.5	3.5	23.5
4	"Нечіткий"	5.1	4.9	25.3

"Розмитий" дохід трохи більший - так і повинно бути, бо там не заборонено переступати через обмеження.

Постановка і розв'язання двоїстої задачі

Не варто зовсім відмінати ЛП. Ще дуже багато може дати використання результатів теорії двоїстості в ЛП, якщо оволодіти цією теорією. Для деякої (= прямої) задачі ЛП (лінійного програмування):

$$\max \{C^T \cdot X\} \quad \text{при } A \cdot X \leq B.$$

Двоїстою (спряженою) називається така задача ЛП

$$\min \{B^T \cdot Y\} \quad \text{при } A^T \cdot Y \geq C.$$

Головний результат теорії двоїстості - теорема про мінімакс:

$$\max \{C^T \cdot X\} = \min \{B^T \cdot Y\}$$

Чим цей результат корисний для практики?

Ситуація №1. Маємо задачу ЛП великої розмірності і одне якесь допустиме розв'язання $X_{\text{прям}}$. Чи далеко воно від бажаного оптимуму? - записуємо двоїсту (спряжену) задачу ЛП, знаходимо якесь припустиме її розв'язання $X_{\text{спряж}}$. На основі теореми про мінімакс можна стверджувати, що оптимальне розв'язання знаходиться в діапазоні:

$$X_{\text{спряж}} \leq X_{\text{опт}} \leq X_{\text{прям}}$$

Ситуація №2. Спряжена задача може бути більш легкою для розв'язання (в яких випадках, чому - дійдіть до цього самостійно)

Спряжена задача - абстракція, чи вона має теж техніко-економічну інтерпретацію як і основна, пряма задача ЛП? Корисно знайти інтерпретацію самостійно. Подамо підказку. В лінійній економіці, без доданої вартості, теорема про мінімакс стверджує, що максимально можливий прибуток дорівнює мінімально можливим витратам. Якщо змінні $x_1, x_2, \dots$ - план випуску, то змінні $y_1, y_2, \dots$ - ?). Запишемо пряму та спряжену задачі в розгорнутому вигляді.

Витрати ресурсів на один виріб	Запаси ресурсів обмеження	Ціна одиниці	План випуску (шукаємо)
$Aa := \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{pmatrix};$	$Bb := \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \begin{pmatrix} \text{дуб} \\ \text{лак.} \\ \text{цвяхи} \end{pmatrix};$	$Cc := \begin{pmatrix} Cc_1 \\ Cc_2 \end{pmatrix} \begin{pmatrix} \text{"грн/стіл"} \\ \text{"грн/стілець"} \end{pmatrix};$	$Xx := \begin{pmatrix} Xx_1 \\ Xx_2 \end{pmatrix} \begin{pmatrix} \text{столів} \\ \text{стілців} \end{pmatrix}$

Далі подана копія екрану сторінки з електронної книги

Подаємо у цих позначеннях **пряму** задачу ЛП.

Знайти максимум цільової функції:

при обмеженнях

$$\boxed{C^T \cdot X \rightarrow C_1 \cdot X_1 + C_2 \cdot X_2}; \quad A \cdot X \leq B \rightarrow \begin{pmatrix} a_{1,1} \cdot X_1 + a_{1,2} \cdot X_2 \\ a_{2,1} \cdot X_1 + a_{2,2} \cdot X_2 \\ a_{3,1} \cdot X_1 + a_{3,2} \cdot X_2 \end{pmatrix} \leq \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Робимо теж саме для спряженої задачі

Витрати ресурсів
на один виріб

обмеження на
вартість виробів

кількості
ресурсів

Ціни ресурсів
(шукаємо)

$$A^T \rightarrow \begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \end{pmatrix}; \quad C := \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \begin{pmatrix} \text{"грн/стіл"} \\ \text{"грн/стілець"} \end{pmatrix}; \quad B^T \rightarrow \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix}; \quad Y := \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}.$$

Подаємо у цих же означеннях **двоїсту** (спряжену) задачу ЛП.

Знайти мінімум цільової функції:

при обмеженнях

$$\boxed{B^T \cdot Y \rightarrow b_1 \cdot Y_1 + b_2 \cdot Y_2 + b_3 \cdot Y_3}; \quad A \cdot Y \geq C \rightarrow \begin{pmatrix} a_{1,1} \cdot Y_1 + a_{2,1} \cdot Y_2 + a_{3,1} \cdot Y_3 \\ a_{1,2} \cdot Y_1 + a_{2,2} \cdot Y_2 + a_{3,2} \cdot Y_3 \end{pmatrix} \geq \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$$

Зауваження. Були використані позначення X замість x , A замість A та ін. щоб використати апарат символьних обчислень - там потрібні змінні без числових значень.

Стенд для знаходження розв'язань спряжених задач ЛП

Для набуття практичних навичок і осмислення двоїстості виконаємо (якщо Ви працюєте з електронною версією книги) практичне заняття - знаходження розв'язання задачі ЛП методом проб і помилок. Спочатку записуємо два модулі для розв'язання прямої і двоїстої задач ЛП за умови, що значення цільових функцій КритГ та КритД задані.

Задача ЛП розв'язується так: починаємо з малого значення КритП та великого для КритД і змінюємо їх у відповідному напрямку, поки функція Find(.) не почервоніє (буде повідомлення "не знайдено розв'язання", якщо курсор на Find(.)).

Пряма задача - максимізація доходу від виробництва

Початкові значення змінних: $x_1 := 1$; $x_2 := 1$; КритП = 23.3

Given

$$A \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq B; \quad C^T \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \text{КритГ}; \quad x_1 \geq 0; \quad x_2 \geq 0;$$

$$\begin{pmatrix} x_{1opp} \\ x_{2opp} \end{pmatrix} := \text{Find}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right); \quad \begin{pmatrix} x_{1opp} \\ x_{2opp} \end{pmatrix} = \begin{pmatrix} 1 & 5.3 \\ 2 & 3.7 \end{pmatrix}; \quad \begin{pmatrix} x_{1opd} \\ x_{2opd} \\ Sdxg \end{pmatrix} = \begin{pmatrix} 1 & 5.5 \\ 2 & 3.5 \\ 3 & 23.5 \end{pmatrix} \quad - \text{підказка.}$$

Двоїста задача - мінімізація витрат на виробництво

Початкові значення змінних: $y_1 := 1$; $y_2 := 1$; $y_3 := 1$; КритД = 24.35 .

Given

$$A^T \cdot \begin{pmatrix} y1 \\ y2 \\ y3 \end{pmatrix} \geq C ; \quad B^T \cdot \begin{pmatrix} y1 \\ y2 \\ y3 \end{pmatrix} = \text{КритД} ; \quad y1 \geq 0 ; y2 \geq 0 ; y3 \geq 0 .$$

$$\begin{pmatrix} y1_{op} \\ y2_{op} \\ y3_{op} \end{pmatrix} := \text{Find} \left(\begin{pmatrix} y1 \\ y2 \\ y3 \end{pmatrix} \right) ; \quad \begin{pmatrix} y1_{op} \\ y2_{op} \\ y3_{op} \end{pmatrix} = \begin{array}{|c|c|} \hline & 1 \\ \hline 1 & 0.6 \\ \hline 2 & 0.1 \\ \hline 3 & 0.9 \\ \hline \end{array} .$$

Будуємо графіки (копіюємо і редагуємо модулі з першого розділу). Обмеження записуємо як рівності, розв'язуємо їх відносно змінної $x2$ і отримуємо такі функції від $x1$ - кількості столів.

$$x1 := 0, 0.2 .. 25 ; \quad \text{Dox}(x1) := (\text{КритП} - C_1 \cdot x1) \div C_2 ;$$

$$\text{Обме1}(x1) := \frac{(B_1 - A_{1,1} \cdot x1)}{A_{1,2}} ; \text{Обме2}(x1) := \frac{(B_2 - A_{2,1} \cdot x1)}{A_{2,2}} ; \text{Обме3}(x1) := \frac{(B_3 - A_{3,1} \cdot x1)}{A_{3,2}}$$

Формуємо границю області (симплекс), точки якої $(x1, x2)$ задовольняють усі обмеження.

$$\text{ОбмП}(x1) := \min(\text{Обме1}(x1), \text{Обме2}(x1), \text{Обме3}(x1)) .$$

Те ж саме робимо для спряженої задачі:

$$\text{Roz}(y1, y2) := (\text{КритД} - B_1 \cdot y1 - B_2 \cdot y2) \div B_3 ;$$

$$\text{Обм1}(y1, y2) := \frac{(C_1 - A_{1,1} \cdot y1 - A_{2,1} \cdot y2)}{A_{3,1}} ; \text{Обм2}(y1, y2) := \frac{(C_2 - A_{1,2} \cdot y1 - A_{2,2} \cdot y2)}{A_{3,2}} ;$$

$$\text{ОбД}(y1, y2) := \max(\text{Обм1}(y1, y2), \text{Обм2}(y1, y2)) .$$

Для побудови тривимірних графіків вводимо дискретні (індексовані) змінні $X1_i, X2_j, Y1_i$.

$$\text{Діап} := 50 ; i := 1 .. \text{Діап} ; j := 1 .. \text{Діап} ; h := 0.15 ; X1_i := h \cdot i ; X2_j := h \cdot j ;$$

$$\text{Sdox}(x1, x2) := C_1 \cdot x1 + C_2 \cdot x2 ; h2 := 0.05 ; Y1_i := h2 \cdot i ; Y2_j := h2 \cdot j ; .$$

Формуємо дискретну функцію двох змінних - цільову функцію прямої задачі "обрізану" обмеженнями.

$$\text{Dopr}_{i,j} := \text{if} \left[(A_{1,1} \cdot X1_i + A_{1,2} \cdot X2_j \leq B_1) \cdot (A_{2,1} \cdot X1_i + A_{2,2} \cdot X2_j \leq B_2) \cdot (A_{3,1} \cdot X1_i + A_{3,2} \cdot X2_j \leq B_3), \text{Sdox}(X1_i, X2_j), 0 \right]$$

Формуємо відповідні дискретні функції для спряженої задачі

$$\text{Roz}_{i,j} := \frac{\text{Roz}(Y1_i, Y2_j)}{h2} ; \text{Обм1}_{i,j} := \frac{\text{Обм1}(Y1_i, Y2_j)}{h2} ;$$

$$\text{Обм2}_{i,j} := \frac{\text{Обм2}(Y1_i, Y2_j)}{h2} ; \text{ОбД}_{i,j} := \frac{\text{ОбД}(Y1_i, Y2_j)}{h2} .$$

Зареєструйся - введи ім'я
 Користувач := "Дебагер"

$$xc1_{i,j} := i \quad xc2_{i,j} := \left(\frac{\text{КритП}}{h} - C_1 \cdot xc1_{i,j} \right) \div C_2 \quad Z_{i,j} := j \cdot 5$$

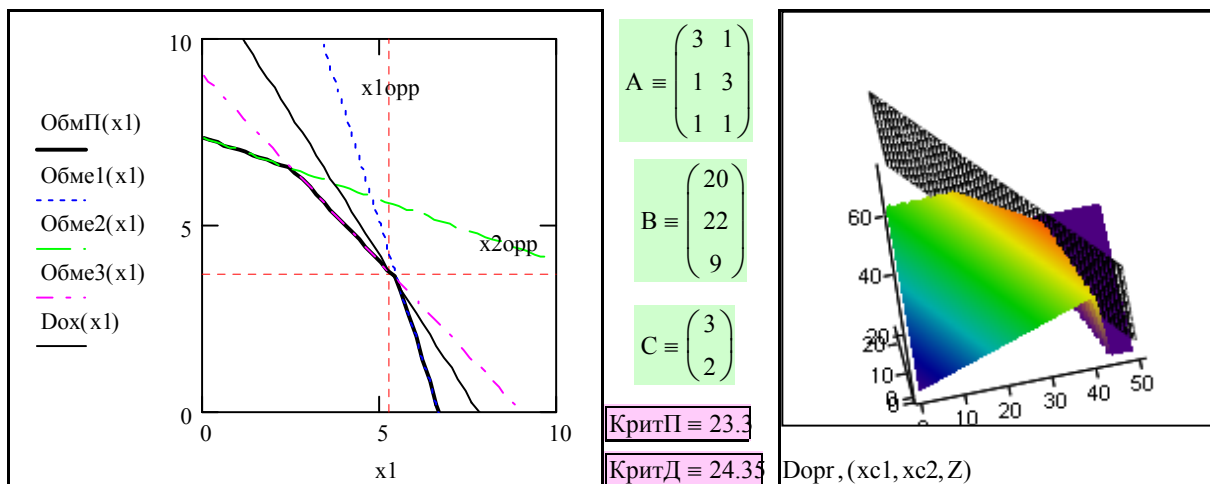


Рис. 7. Графічний розв'язок прямої задачі (2 змінних, 3 обмеження)

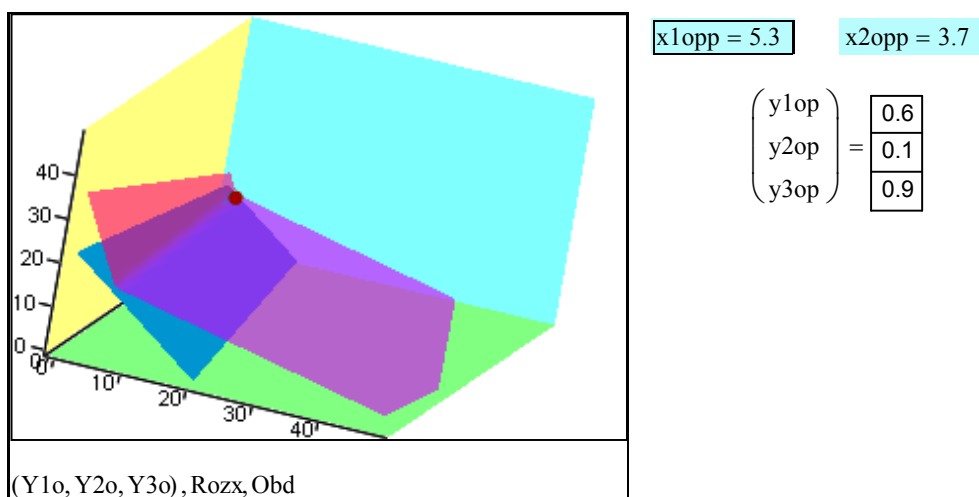


Рис.8 Графічне розв'язання двоїстої задачі (3 змінних, 2 обмеження)

Інструкція користувача стенда. Шановний Користувач = "Дебагер" !

1. **Введи дані** своєї задачі - A, B, C.
 2. **Встанови значення** цільових функцій КритП - малим (для Прямої виробничої задачі), КритД - великим (для Двоїстої задачі).
 3. **Змінюй** (у відповідному напрямі) ці значення, дивлячись на графіки, не виходячи за обмеження, **поки** не досягнеш мінімаксу.
 4. **Можна** повертати 3D-графіки мишею у зручний ракурс.
- Зверніть увагу на одну особливість нашої конкретної задачі - в точці оптимуму одна з координат вектора цін ресурсів у двоїстій задачі стає нульовою. Тобто відповідний ресурс є "безкоштовним". Що це означає, розглянемо далі.

Розробка модуля для розв'язання та аналізу пари спряжених задач ЛП

На основі виконаних досліджень зробимо модуль зручний для користування і переносу в інші документи. **Входи модуля** - певна задача ЛП. В модулі автоматично формується спряжена (двоїста) задача. **Виходи модуля** а) розв'язання пари задач; г) приклад висновків, що може зробити менеджер, дивлячись на ці розв'язки.

Вхідні дані: технологічна матриця:

$$Aa := \begin{pmatrix} 3 & 1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix}$$

обмеження:

$$Bb := \begin{pmatrix} 20 \\ 22 \\ 9 \end{pmatrix}$$

ціни:

$$Cc := \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Пряма задача. Записуємо цільову функцію, екстремум якої шукаємо:

$$Dp(xx1, xx2) := Cc_1 \cdot xx1 + Cc_2 \cdot xx2$$

Записуємо початкові значення шуканих змінних: $xx1 := 1$ $xx2 := 1$

Записуємо систему обмежень (рівностей, нерівностей)

$$\text{Given } xx1 \geq 0 \quad xx2 \geq 0$$

$$Aa_{1,1} \cdot xx1 + Aa_{1,2} \cdot xx2 \leq Bb_1$$

$$Aa_{2,1} \cdot xx1 + Aa_{2,2} \cdot xx2 \leq Bb_2$$

$$Aa_{3,1} \cdot xx1 + Aa_{3,2} \cdot xx2 \leq Bb_3$$

і отримуємо (не завжди) розв'язання: $\begin{pmatrix} x1opp \\ x2opp \end{pmatrix} := \text{Maximiz}(Dp, xx1, xx2)$

$$\text{Сумдох} := Dp(x1opp, x2opp) \quad x1opd = 5.5 \quad x2opd = 3.5 \quad \text{Сумдох} = 23.5$$

Двоїста задача. Записуємо цільову функцію, екстремум якої шукаємо:

$$Dd(yy1, yy2, yy3) := Bb_1 \cdot yy1 + Bb_2 \cdot yy2 + Bb_3 \cdot yy3$$

Записуємо початкові значення шуканих змінних: $yy1 := 1$ $yy2 := 1$ $yy3 := 1$

Записуємо систему обмежень (рівностей, нерівностей)

$$\text{Given } yy1 \geq 0 \quad yy2 \geq 0 \quad yy3 \geq 0$$

$$Aa_{1,1} \cdot yy1 + Aa_{2,1} \cdot yy2 + Aa_{3,1} \cdot yy3 \geq Cc_1$$

$$Aa_{1,2} \cdot yy1 + Aa_{2,2} \cdot yy2 + Aa_{3,2} \cdot yy3 \geq Cc_2$$

і отримуємо (не завжди) розв'язання: $\begin{pmatrix} y1op \\ y2op \\ y3op \end{pmatrix} := \text{Minimiz}(Dd, yy1, yy2, yy3)$

$$\text{Сумвитр} := Dd(y1op, y2op, y3op) \quad y1op = 0.5 \quad y2op = 0 \quad y3op = 1.5 \quad \text{Сумвитр} = 23.5$$

Подаємо приклад висновків, що може зробити менеджер, дивлячись на ці результати:

Зайве_дуб := $Bb_1 - Aa_{1,1} \cdot x1opp - Aa_{1,2} \cdot x2opp$; Зайве_дуб = 0 - ресурс вибрано повністю;

Зайве_лак := $Bb_2 - Aa_{2,1} \cdot x1opp - Aa_{2,2} \cdot x2opp$; Зайве_лак = 6 - є залишок ресурсу;

Зайве_цвях := $Bb_3 - Aa_{3,1} \cdot x1opp - Aa_{3,2} \cdot x2opp$; Зайве_цвях = 0 - ресурс вибрано повністю.

Рішення: **продати надлишок** Зайве_лак = 6

Таким чином, якщо у двоїстій задачі якась змінна = оцінка ресурсу дорівнює нулю, то відповідний ресурс не використовується повністю у оптимальному плані виробництва. Можна зменшити запас ресурсу на відповідну величину, а сумарний дохід не зменшиться. Перевірте!

В реальних умовах ціни "скачуть", технології вдосконалюються, ресурси застарівають. Тому потрібна програмна система не тільки для розрахунку "оптимального" плану, а ще й для аналізу **"що буде, якщо"**. Потрібна система для ризик-аналізу, не на базі **"інтервалів довіри"**, а на підставі конкретних розрахунків для можливих альтернатив зовнішніх умов та власних дій.

Висновки

Розроблено набір базових модулів для комплексного аналізу і розв'язання задач ЛП альтернативними методами: графічним, пошуковим, нечітким, на базі теорії двоїстості. Усі модулі випробувані і вони можуть досить легко бути модифіковані на вищі розмірності задач. Розроблені модулі не використовують симплекс-метод і тому дозволяють розв'язувати і нелінійні задачі. Модулі можна переносити простим копіюванням в інші документи.

Контрольні запитання

1. Сформулюйте виробничу задачу: обмеження, критерій, змінні оптимізації, ціль.
2. Сформулюйте задачу, двоїсту до виробничої: обмеження, критерій, змінні, ціль.
3. "Фізичний смисл" елементів технологічної матриці.
4. Що таке "функція належності", належність чого до чого вона задає?
5. Як формуються функції належності, на базі реальних вимог та умов реальної задачі, що характеризують ці обмеження? Як комбінуються функції належності?
6. Альтернативи згортки (комбінування) декількох нечітких цілей і обмежень.
7. Чому і в яку сторону відрізняється розв'язання нечіткої задачі ЛП від чіткої?
8. Певна змінна у розв'язанні двоїстої задачі дорівнює нулю. Як це інтерпретується? Які рішення може прийняти в такій ситуації менеджер?
9. Які товари виробляються на фабриці?

5

Оптимізація розвитку систем

У цьому розділі розглядаються:

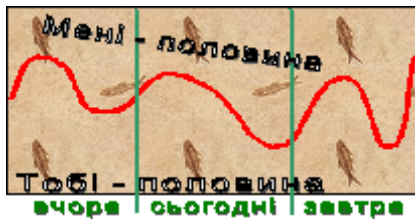
- ✓ Одновимірна задача Марковиця. Варіаційне обчислення.
- ✓ Двовимірна задача Марковиця.
- ✓ Узагальнення задачі Марковиця для випадку нелінійних інвестиційних функцій.

Після вивчення цього розділу ви **зможете:**

1. Знаходити в своїй професійній діяльності варіаційні задачі і виконувати їх формалізацію – подання у стандартній формі – критерій, обмеження, граничні умови.
2. Для типових варіаційних задач записувати рівняння Ейлера-Лагранжа і знаходити їх аналітичні, або числові розв'язання.
3. Записувати задачі максимізації сумарного прибутку виробничих систем довільної розмірності і з довільними функціями віддачі інвестицій.
4. „Конструювати” функції Гамільтона для задач максимізації сумарного прибутку з нелінійними функціями віддачі інвестицій.
5. Розробляти програмні модулі оптимізації функції Гамільтона.
6. Розробляти програми моделювання розвитку виробничих систем з оптимальним за критерієм сумарного прибутку управлінням.
7. Виконувати дослідження на моделях розвитку виробничих систем.

8. Дати визначення термінам:

- | | |
|-----------------------------------|--|
| 1) функція, функціонал, оператор; | 6) варіаційна задача; |
| 2) функція віддачі інвестицій; | 7) рівняння Ейлера – Лагранжа; |
| 3) інтегральний функціонал; | 8) неперервна/розривна функція управління; |
| 4) функція Гамільтона; | 9) випукла/ невивукла функція; |
| 5) варіація функціоналу; | 10) оптимальна стратегія |



5.1. Одновимірна задача Марковиця. Варіаційне обчислення

Вступ. Варіаційні задачі

Досі ми розглядали оптимізаційні задачі, де треба було знайти екстремум функції 2-ох, 3-ох .. змінних. Спробуємо осмислити і освоїти новий клас оптимізаційних задач. На практиці зустрічаються оптимізаційні задачі, де треба знайти не декілька чисел, а цілу функцію - оптимальну стратегію управління. Формально, це задача знаходження екстремуму функції нескінченного числа змінних (формально неперервна функція задається нескінченним числом своїх значень).

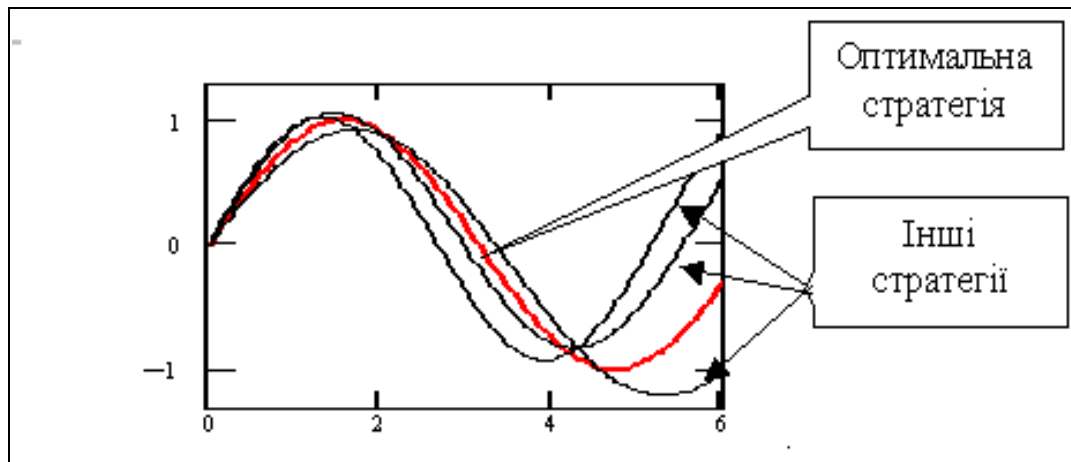


Рис.1. Варіаційна задача - пошук функцій, що оптимізують певні критерії

Функцію, що є розв'язанням варіаційної задачі, можна наблизити рядом дискретних значень, можна розкласти в ряд за певними базисними функціями (наприклад, в ряд Фур'є) - тоді варіаційна задача зведеться до звичної задачі оптимізації деякої функції $N < \infty$ змінних. Особливість варіаційних задач в тому, що розв'язання їх **дуже легко** знаходиться "саме собою", природним способом, і **дуже важко** - математичними методами. Розглянемо декілька прикладів варіаційних задач у нашому оточенні.

Ейфелева, Останкінська башти, башта Шухова мають форму, що дає мінімум ваги при заданій висоті та міцності.

Дроти між стовпами провисають між стовпами не просто так, а по так званій "ланцюговій лінії", що мінімізує зусилля в дроті.

Краплини води на склі приймають форму, яка мінімізує зусилля на поверхні. Подібну форму надають сучасним великим резервуарам для нафти.

Математичне розв'язання варіаційних задач потребує найвищої кваліфікації, винахідливості і просто везіння. Нові розв'язання варіаційних задач з'являються дуже рідко, а ті, що відомі - майже усі - **іменні**, наприклад, задача Ейлера, задача Лагранжа, задача Больца, метод динамічного програмування Беллмана, метод принципу максимуму Понтрягіна. Ми не ставимо задачі систематичного вивчення варіаційного обчислення - це задачі університетського курсу вищої математики. Ми послідовно розглянемо конкретні розв'язання конкретних задач. Першою розглянемо так звану **задачу Марковиця**. Вона дуже просто розв'язується для окремого випадку і, як побачимо, є актуальною для практики.

Постановка задачі

Задачі максимізації повного прибутку, отримання максимальної кількості продукції ніколи не переставали бути актуальними. Традиційно, починаючи з робіт Р. Беллмана, викладення дискретного методу динамічного програмування ведеться на прикладі оптимізації сумарного прибутку у багатокроковому виробничому процесі.

Цікаво, що Беллман створив дискретний метод динамічного програмування для "поганих" варіаційних задач - дискретних, з "жорсткими" обмеженнями. Він же, разом із Л. Заде запропонував підхід на базі "нечіткої логіки". Цей підхід дозволяє задачі з обмеженнями типу нерівностей трансформувати у такі, що можуть бути розв'язані класичними варіаційними методами. Між іншим динамічне програмування зовсім не програмування, а "динамічне планування", якщо точно перевести, а от control - це не контроль, а управління типу "регулювання положення автомашини на дорозі", а контроль - це checking, а управління типу "вибір дороги" - це guiding, а керування - це "руководство", і скоріше "менеджмент", а дослідження операцій певним чином вже не існує, тепер це називається "що_буде_якщо" аналіз, ризик аналіз та ін. От така кібернетика (вона ж розглядає всі ці задачі в комплексі) з термінологією.

Розглянемо два види варіаційних задач: 1) з обмеженнями на припустимі витрати (управляюча дія) типу нерівності; 2) класичні варіаційні задачі без обмежень.

Задачі максимізації повного прибутку. Задача Марковиця

Розглянемо дещо ідеалізовану задачу з математичної економіки. Маємо економічну систему, в якій виробляються N видів продукції, що випускаються з темпами $x_1(t)$, $x_2(t)$, $x_3(t)$,... $x_1(t)$, $x_N(t)$ (одиниць вимірювання за місяць, квартал, рік). Залежності приростів темпів випуску від капіталовкладень (інвестицій у розширення виробництва) будемо вважати лінійними для усіх видів виробництв. Задаємо також обмеження на параметри і початкову умову для виробництва. Динаміка розширення виробництва:

$$\text{ORIGIN} := 1 \quad \frac{d}{dt} x_i = \sum_{j=1}^N a_{i,j} \cdot y(t)_j \quad \begin{matrix} a_{i,j} > 0 \\ i := 1..N \end{matrix} \quad \boxed{y(t)_j \leq x(t)_j} \quad \begin{matrix} x(0)_j = c_j \\ j := 1..N \end{matrix}$$

Звернемо увагу на виділену нерівність - вона інтерпретується як "поточний розмір інвестицій не повинен бути більше поточного доходу (обсягу виробництва)". Потім на основі отриманого розв'язання задачі, можна буде розглянути більш реалістичні постановки з кредитами, акціонерним капіталом та ін. Згідно з умовами задачі, потік продукції $x_{t,i}$ розподіляється на дві складові: $x_{t,i}$ та $z_{t,i}$ - темп (потік) інвестицій у розширення виробництва і темп прибутку, що використовується за межами системи, тобто:

$$z_{t,i} = x_{t,i} - y_{t,i}.$$

Потрібно визначити оптимальну стратегію інвестицій (не плутати з поточними витратами на виробництво видів продукції), що максимізує повний прибуток:

$$J_N = \int_0^T \sum_{j=1}^N (x_{t,i} - y_{t,i}) dt$$

Одновимірна задача

Цей розділ, як і підрозділ 3.4 є в значній мірі перенесенням у середовище сучасного математичного пакета першоджерела - відповідного розділу з книги: Беллман Р., Гликсберг И., Гросс О. Некоторые вопросы математической теории управления. - М.: Издат. Иностран. Литер., 1962. -233 с. На основі цієї задачі потім (підрозділи 5.2, 5.3) спробуємо розв'язати актуальні задачі управління розвитком виробничих систем.

Розглянемо розв'язання одновимірної задачі. Розв'язання багатовимірних задач має ту ж структуру і властивості, але відповідно воно більш трудомістке - більше змінних, більше рівнянь. Оскільки ми робимо робочий документ, що потім буде модифіковано для випадку N змінних, то використовуємо означення з індексами (єдиний в одновимірній задачі елемент технологічної матриці ми позначаємо як a_{11}).

Записуємо диференціальне рівняння: $\frac{d}{dt}x(t)_1 = a_{11} \cdot y(t)_1$ (це обмеження, що задане диференціальним рівнянням), обмеження типу нерівності для коефіцієнта a_{11} : $a_{11} > 0$, початкова умова - значення для $x(t)$ при $t = 0$: $x(0)_1 = c_1$ і, нарешті, **функціонал** (це відображення функції в

точку)

$$J_1 = \int_0^T (x(t)_1 - y(t)_1) dt.$$

Ціль оптимізації - максимум функціоналу. Управляюча змінна (те, що від нас залежить) - функція інвестицій - $y(t)$.

Розв'язання задачі

Розв'яжемо диференціальне рівняння відносно x :

$$\frac{d}{dt}x(t)_1 = a_{1,1} \cdot y(t)_1 \quad dx_1 = a_{1,1} \cdot y(t)_1 \cdot dt \quad x(t)_1 = c_1 + a_{1,1} \cdot \int_0^t y(t)_1 dt \quad (1)$$

Підставимо вираз для $x(t)_1$ у функціонал:

$$J_1 = \int_0^T (x(t)_1 - y(t)_1) dt ; \quad J_1 = \int_0^T \left(c_1 + a_{1,1} \cdot \int_0^t y(t_1)_1 dt_1 - y(t)_1 \right) dt ; \quad (2)$$

$$\text{де } y(t)_1 \text{ задовольняє умову } 0 \leq y(t)_1 \leq c_1 + a_{1,1} \cdot \int_0^t y(t)_1 dt. \quad (3)$$

Ця умова інтерпретується так: не можна інвестувати більше, ніж заробляється. Ми бачимо, що права частина нерівності - це вираз для x_1 (3).

Завдання. Часто оптимальне управління є очевидним - його можна знайти виходячи тільки з інтуїції. Опишіть характер оптимального управління (пропорційно, спочатку збільшувати, потім зменшувати та ін.) $Y_{opt}(t)$.

Змінюємо порядок інтегрування у (2) і отримуємо:

$$J_1 = c_1 \cdot T + \int_0^T [a_{1,1} \cdot (T - t) - 1] \cdot y(t)_1 dt \quad (4)$$

Зауваження. Відносно отриманого виразу може виникнути ряд питань. Чи можна змінювати порядок інтегрування? - Можна, бо оператор інтегрування лінійний. Дещо заважає осмисленню те, що внутрішній і зовнішній інтеграли (у виразі (2)) беруться по змінних, які мають однакову розмірність - час. Слід тільки дотримуватись певних правил при визначенні границь інтегрування. Неважко дати змістовну інтерпретацію кратному інтегралу при зміні порядку інтегрування, якщо замінити його подвійною сумою, а змінні t, t_1 на індекси k, l . Тоді ми побачимо, що ми замінюємо підсумовування стовпців на підсумовування рядків по тій же області.

Завдання для самостійного виконання. Виведіть (4) з (2).

Проаналізуємо інтеграл (4). Щоб функціонал був максимальним, підінтегральний вираз в кожний момент часу повинен бути максимально можливим. Як цього досягнути? Бачимо - підінтегральний вираз - це добуток виразів $[a_{1,1} \cdot (T - t) - 1]$ та $y(t)_1$. Перший множник - функція часу, на яку ми не маємо впливу, другий множник - $y(t)_1$ - це шукана функція управління - розмір інвестицій у розширення виробництва. Як же слід сформулювати цю функцію? Перший множник - функція часу, що в момент $t = 0$, при умові $a_{1,1} \cdot T > 1$, стає меншою нуля. На підставі цього вибираємо стратегію управління техніко-економічним процесом "інвестиції-виробництво".

При заданих нами додаткових умовах відносно a_{11} і T , що є природними, **функція $Y_{opt}(t)$ повинна бути максимально можливою там, де перший множник додатний, і нульовою, там де він від'ємний.**

Тобто, в деякий момент часу T_1 має місце "переключення" управління з стратегії "усе у виробництво" в стратегію "усе у накопичення". Увага! Далі для скорочення позначимо **$Y_{opt}(t) = y_1(t)$** . Враховуючи обмеження (3) на $y_1(t)$ маємо:

$$y_1(t) = \begin{cases} \left(c_1 + a_{1,1} \cdot \int_0^T y_1(t) dt \right) & \text{if } 0 \leq t \leq T_1 \\ 0 & \text{if } T_1 \leq t \leq T \end{cases} \quad (5)$$

Отримаємо явний вираз для $y_1(t)$. Продиференціюємо рівняння (5):

$$y_1(t) = c_1 + a_{1,1} \cdot \int_0^T y_1(t) dt; \Rightarrow \frac{d}{dt} y_1(t) = a_{1,1} \cdot y_1(t); \Rightarrow \frac{d}{dt} y_1(t) - a_{1,1} \cdot y_1(t) = 0$$

Розв'язуємо отримане нами звичайне лінійне однорідне диференціальне рівняння першого порядку (САМОСТІЙНО?). Записуємо характеристичне рівняння $s - a_{11} = 0$ і знаходимо його корінь $s = a_{11}$.

Записуємо розв'язання диференціальне, враховуємо граничне (початкове) значення $x_1(0)$ і те, що оптимальне управління починається з $y_1(0) = x_1(0)$:

$$y_1(t) = c_1 \cdot e^{a_{11} \cdot t} \quad \text{- до моменту переключення } T_1. \quad (6)$$

Розробка стенда для аналізу інвестиційної політики

Сьогодні нераціонально і дуже замало отримати якесь конкретне розв'язання інженерно-економічної задачі. Прикладну програму розрахунку слід оформити у вигляді стенду для "що-буде-якщо" аналізу. Назва "стенд" є певним перебільшенням - ми розглядаємо поки лінійну одновимірну задачу Марковіца. Вводимо значення параметрів:

$$a_{11} := 0.27; \quad c_1 := 2; \quad T := 12; \quad t := 0, 0.5..12.$$

Визначимо момент T_1 переключення управління (див. (4)) з стратегії "усе у виробництво" в стратегію "усе у накопичення/споживання".

$$a_{11} \cdot (T_1 - t_1) - 1 = 0 \text{ solve, } t_1 \rightarrow \frac{(a_{11} \cdot T_1 - 1)}{a_{11}} ; T_1 := \frac{(a_{11} \cdot T - 1)}{a_{11}} ; T_1 = 8.3 .$$

Ми використали символний процесор для розв'язання рівняння відносно змінної T_1 . Але нам довелося ввести штучні змінні без числових значень, тому що пакет примусово підставляє у символні вирази значення тих змінних, які ці значення мають. Потім ми скопіювали результат (у другий вираз), повернулися до старих змінних і вивели числовий результат $T_1 = 8.3$.

Записуємо рівняння для темпу виробництва $x_1(t)$, інвестицій (підставили (6) у (5)) $y_1(t)$, "прибутку" $z_1(t)$, сумарного прибутку $Sz(t)$:

$$T_1 := \frac{(a_{11} \cdot T - 1)}{a_{11}} ; y_1(t, T_1, a_{11}) := \begin{cases} \left(c_1 + a_{11} \cdot \int_0^t c_1 \cdot e^{a_{11} \cdot t} dt \right) & \text{if } 0 \leq t \leq T_1 \\ 0 & \text{if } T_1 \leq t \leq T \end{cases} ;$$

$$x_1(t, T_1, a_{11}) := \begin{cases} y_1(t, T_1, a_{11}) & \text{if } 0 \leq t \leq T_1 \\ y_1(T_1 - .01, T_1, a_{11}) & \text{if } T_1 \leq t \leq T \end{cases} ;$$

$$z_1(t, T_1, a_{11}) := x_1(t, T_1, a_{11}) - y_1(t, T_1, a_{11}); Sz(t, T_1, a_{11}) := \int_0^t z_1(t, T_1, a_{11}) dt .$$

Збираємо входи і виходи в стенд:

$$c_1 = 2 \quad a_{11} = 0.27 \quad T = 12 \quad T_1 = 8.3 \quad Sz(T, T_1, a_{11}) = 69.4 \text{ уо}$$

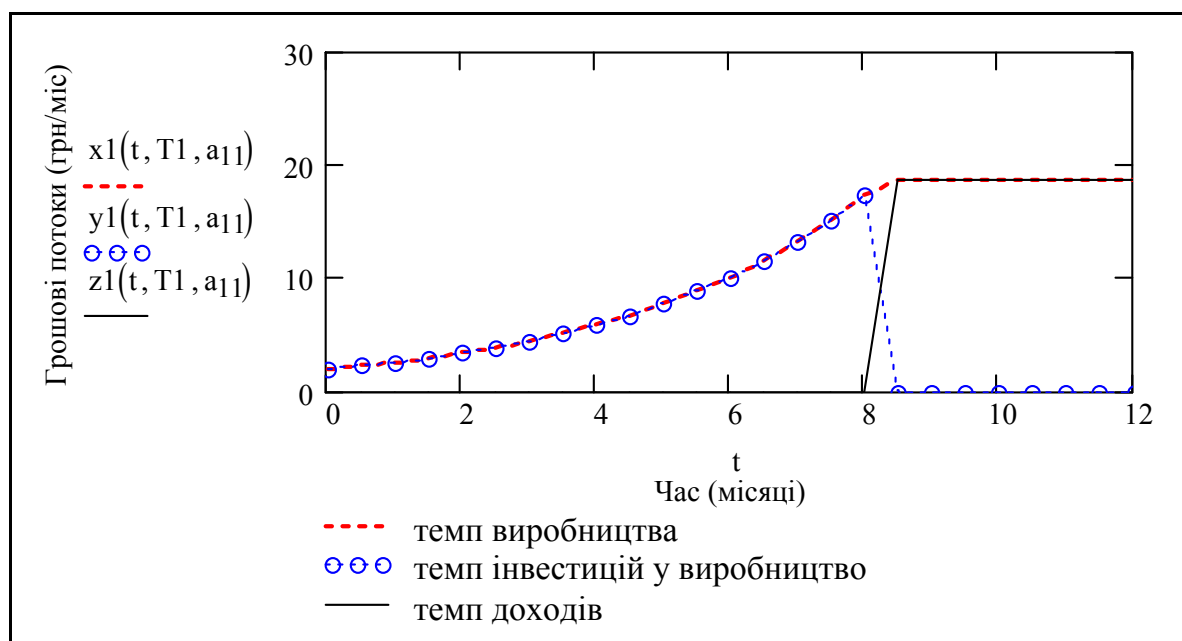


Рис.2. Оптимальний процес розвитку

Зауваження. Числа в прикладі взято умовні. Подаємо кілька копій екрану для інших значень параметрів

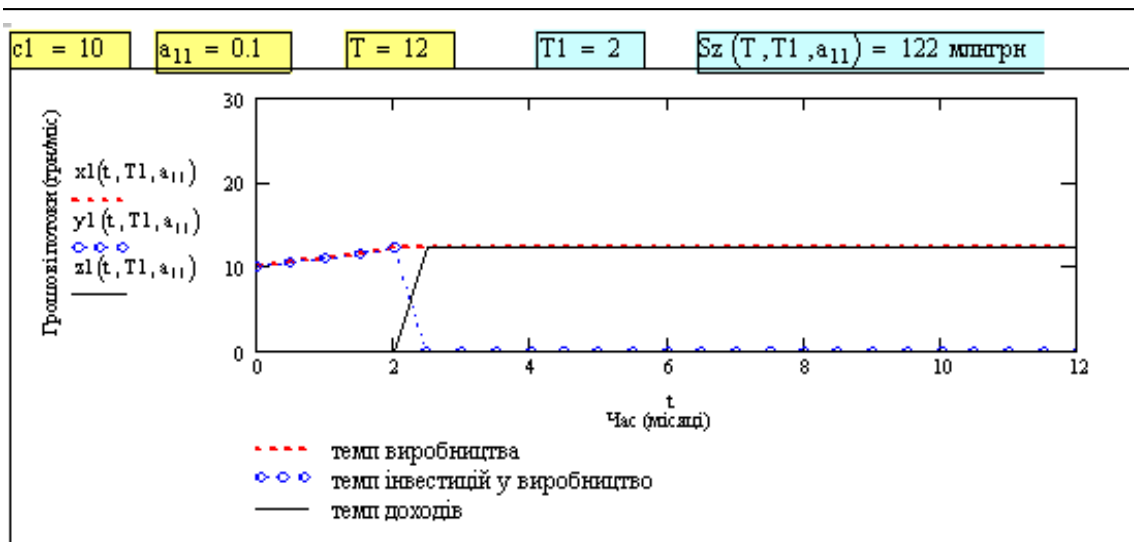
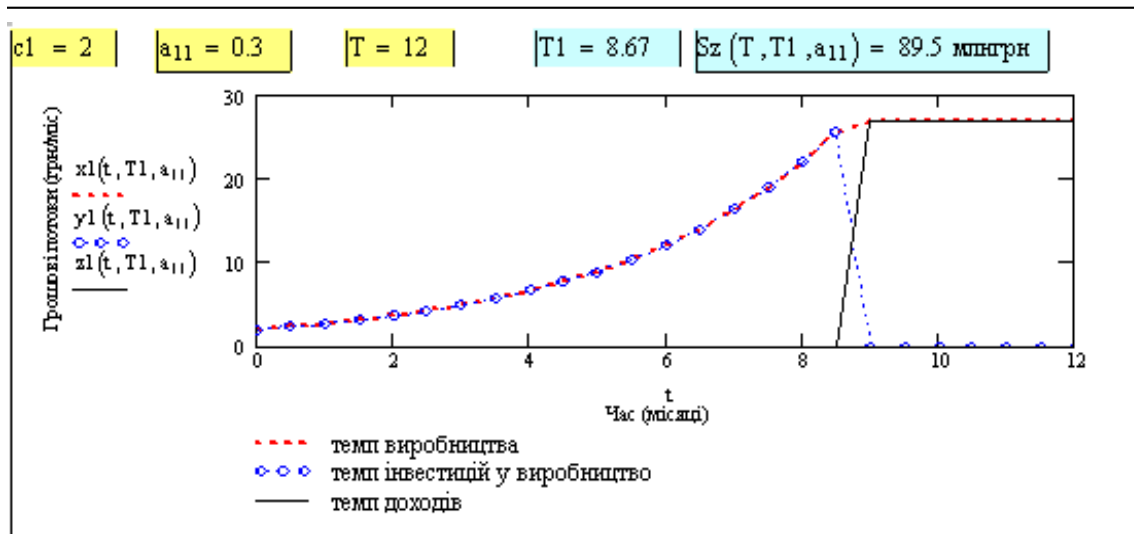


Рис.3. Оптимальні процеси при різних значеннях коефіцієнта віддачі інвестицій

Отримане розв'язання дійсно оптимальне - спробуємо змінити момент переключення:

$$Sz(T, T_1 \cdot 0.9, a_{11}) = 67.9 \quad ; \quad Sz(T, T_1, a_{11}) = 69.4 \quad ; \quad Sz(T, T_1 \cdot 1.1, a_{11}) = 67.4$$

Бачимо, що дійсно "крок вправо, крок вліво - втеча", від максимального прибутку.

Завдання для самостійного виконання

1. Спробуйте розв'язати задачу для випадку, коли бюджетне обмеження $y_1(t) \leq x_1(t)$ є м'яким (нечітким) - 55 балів.

2. Розробіть програму для розв'язання задачі (методом динамічного програмування та ін.) для випадку, коли диференціальна віддача інвестицій у розширення виробництва є нелінійним:

$\frac{d}{dt} x(t)_i = f_n(y(t)_1)$, де $f_n(y(t)_1)$ - довільна нелінійна функція - 299 балів.

3. Модифікуйте стенд так, щоб вхідні параметри вводилися біля графіка - 15 балів.

4. Модифікуйте стенд так, щоб виводились одночасно три процеси - номінальний і для $+\Delta p$, $-\Delta p$ - заданого розкиду деякого заданого параметра - 25 балів.

5. Розробіть програму для розв'язання двовимірної задачі Марковіца - 75 балів.

6. Розробіть програму для розв'язання тривимірної задачі Марковіца - 100 балів.

Класичні варіаційні задачі

Як було сказано на початку розділу, нові наукові підходи, нові програмно-апаратні можливості **дозволяють ефективно використовувати класичні варіаційні методи** для розв'язання задач прогнозування, планування і управління соціо-техніко-економічними системами. Для того, щоб ефективно застосовувати ці методи, треба їх освоїти до рівня напівавтоматичних навичок. Далі подано розв'язання типових, "навчальних" варіаційних задач. Новим тут є застосування засобів математичного пакета: **"нові інформаційні технології"**. Не розглядаються якісь техніко-економічні інтерпретації цих задач - це залишено Вам, для самостійної роботи.

Задача Ейлера. Приклад 1

Постановка задачі

Потрібно знайти функцію $Y_{opt}(t)$, що при заданих граничних умовах: $y(t_0) = 2$ $y(t_k) = 0$ або в мові пакета: $y_0 := 2$; $y_k := 0$; $t_0 := 0$; $t_k := 1$ ORIGIN := 0 дає мінімум інтегральному функціоналу:

$$J(y(t)) = \int_0^T \left[y(t)^2 + 4 \cdot \left(\frac{d}{dt} y(t) \right)^2 \right] dt; \quad \max \{ J(y(t)) \} ??$$

Розв'язання задачі

1. Записуємо функцію - підінтегральний вираз функціонала (функцію y та її похідну записуємо як звичайні змінні: y та py).

$$F(y, py) := y^2 + 4py^2$$

2. Отримуємо відповідні похідні від функції

$$\frac{\partial}{\partial y} F(y, py) \rightarrow 2 \cdot y \qquad \frac{\partial}{\partial py} F(y, py) \rightarrow 8 \cdot py$$

3. Записуємо рівняння Ейлера у загальному вигляді, підставляємо вирази для похідних, виконуємо алгебраїчні перетворення.

$$\frac{\partial}{\partial y} F(y, py) - \frac{d}{dt} \frac{\partial}{\partial py} F(y, py) = 0 \Leftrightarrow 2 \cdot y(t) - 8 \cdot \frac{\partial^2}{\partial t^2} y(t) = 0 \Leftrightarrow \frac{\partial^2}{\partial t^2} y(t) - 0.25 \cdot y(t) = 0.$$

4. Розв'язуємо отримане дифрівняння (в даному випадку - звичайне, лінійне, другого порядку, це ж штучний приклад), використовуємо символний процесор для знаходження коренів характеристичного рівняння s_1 та s_2 .

$$s^2 - 0.25 = 0 \quad \left| \begin{array}{l} \text{solve, s} \\ \text{float, 1} \end{array} \rightarrow \begin{pmatrix} .5 \\ -.5 \end{pmatrix} \right; \quad s_1 := 0.5; \quad s_2 := -0.5.$$

Запишемо розв'язання цього рівняння у загальному вигляді:

$$y(t) = C_1 \cdot e^{s_1 \cdot t} + C_2 \cdot e^{s_2 \cdot t}.$$

3 граничних умов визначаємо постійні C_1 та C_2 .

$$\text{при } t_0 := 0 \quad y_0 := 2, \text{ тобто } y(0) = C_1 \cdot e^{s_1 \cdot 0} + C_2 \cdot e^{s_2 \cdot 0} = C_1 + C_2$$

при $t_k := 1$ $y_k := 0$, тобто $y(1) = C1 \cdot e^{s1 \cdot 1} + C2 \cdot e^{s2 \cdot 1} = C1 \cdot e^{0.5} + C2 \cdot e^{-0.5}$

Використовуємо розв'язувальний блок і знаходимо C1 та C2 у числовому вигляді.

Початкові значення $Cv1 := 1$; $Cv2 := 2$;

$$\left. \begin{array}{l} \text{Given} \quad C_{v1} \cdot e^{s1 \cdot t0} + C_{v2} \cdot e^{s2 \cdot t0} = y0 ; \\ C_{v1} \cdot e^{s1 \cdot tk} + C_{v2} \cdot e^{s2 \cdot tk} = yk ; \end{array} \right\}$$

$$\begin{pmatrix} C1 \\ C2 \end{pmatrix} := \text{Find}(Cv1, Cv2) ; \quad \begin{pmatrix} C1 \\ C2 \end{pmatrix} = \begin{pmatrix} -1.16 \\ 3.16 \end{pmatrix} .$$

Для узагальнення, впевненості та освоєння символьних обчислень визначаємо C1 та C2 у символьному вигляді:

$$\left. \begin{array}{l} \text{Given ;} \quad C1 \cdot e^{S1 \cdot T0} + C2 \cdot e^{S2 \cdot T0} = Y0 ; \\ C1 \cdot e^{S1 \cdot Tk} + C2 \cdot e^{S2 \cdot Tk} = Yk ; \end{array} \right\}$$

$$\text{Find}(C1, C2) \rightarrow \left[\begin{array}{l} \frac{(\exp(S2 \cdot Tk) \cdot Y0 - Yk \cdot \exp(S2 \cdot T0))}{(\exp(S1 \cdot T0 + S2 \cdot Tk) - \exp(S1 \cdot Tk + S2 \cdot T0))} \\ \frac{(\exp(S1 \cdot T0) \cdot Yk - Y0 \cdot \exp(S1 \cdot Tk))}{(\exp(S1 \cdot T0 + S2 \cdot Tk) - \exp(S1 \cdot Tk + S2 \cdot T0))} \end{array} \right] ;$$

$$\begin{pmatrix} C1sim \\ C2sim \end{pmatrix} := \left[\begin{array}{l} \frac{(\exp(s2 \cdot tk) \cdot y0 - yk \cdot \exp(s2 \cdot t0))}{(\exp(s1 \cdot t0 + s2 \cdot tk) - \exp(s2 \cdot t0 + s1 \cdot tk))} \\ \frac{-(-\exp(s1 \cdot t0) \cdot yk + y0 \cdot \exp(s1 \cdot tk))}{(\exp(s1 \cdot t0 + s2 \cdot tk) - \exp(s2 \cdot t0 + s1 \cdot tk))} \end{array} \right] ;$$

$$\text{порівнюємо:} \quad \begin{pmatrix} C1 \\ C2 \end{pmatrix} = \begin{pmatrix} -1.16 \\ 3.16 \end{pmatrix} ; \quad \begin{pmatrix} C1sim \\ C2sim \end{pmatrix} = \begin{pmatrix} -1.16 \\ 3.16 \end{pmatrix} .$$

Бачимо, що отримані альтернативними методами "розв'язки" для C1 та C2 збігаються. Записуємо робочі вирази для отриманої оптимальної функції (= стратегії, політики, управління) $Y_{opt}(t)$, похідної $pY(t)$ підінтегрального виразу $piv(t)$ і будуємо графіки. Бачимо, що функція $Y_{opt}(t)$, що дає мінімум функціоналу, досить близька до прямої. Чи не можна було прийти до такого розв'язання більш легкими шляхами?

$$Y_{opt}(t) := (C1 \cdot e^{s1 \cdot t} + C2 \cdot e^{s2 \cdot t}); \quad Nn := 25; \quad t := t0, t0 + \frac{(tk - t0)}{Nn} .. tk$$

$$pqy(t) := \left(\frac{d}{dt} Y_{opt}(t) \right)^2 ; \quad piv(t) := \frac{d}{dt} Y_{opt}(t) + \left(\frac{d}{dt} Y_{opt}(t) \right)^2 ; \quad pY(t) := \frac{d}{dt} Y_{opt}(t).$$

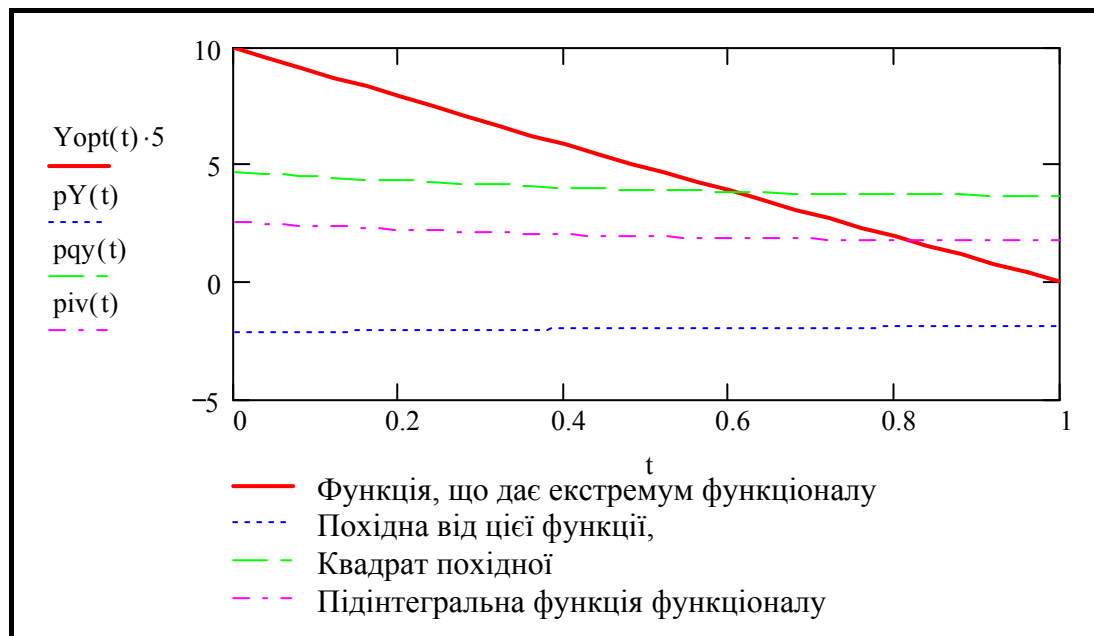


Рис.4. Оптимальний за інтегральним критерієм процес

Задача Ейлера. Приклад 2

Розв'яжемо ще одну подібну варіаційну задачу. Знову використаємо при цьому "новітні інформаційні технології", що в перекладі на побутову мову означає - скопіюємо і відредагуємо попередній документ. Перевіримо, чи може цей документ доведений до рівня фрейма - програми, куди згідно сценарію можна підставляти різні параметри задач класу "варіаційна задача Ейлера". Ще раз нагадаємо - задача спеціаліста сьогодні не виконання якогось розрахунку і отримання "розв'язку", а створення програми для виконання певного класу розрахунків, отримання загальних розв'язань. В даному випадку проблема в тому, що рівняння Ейлера-Лагранжа для варіаційної задачі дуже не завжди мають аналітичне розв'язання.

Постановка задачі

Потрібно знайти функцію $Y_{opt}(t)$, що при заданих граничних умовах: $ORIGIN := 0 \quad y(t_0) = 2 \quad y(t_k) = 0$ або в мові пакета: $y_0 := 2; \quad y_k := 0; \quad t_0 := 1; \quad t_k := 2$ дає мінімум інтегральному функціоналу:

$$J(y(t)) = \int_0^T \left[y(t)^2 + 2 \cdot y(t) \cdot \frac{d}{dt} y(t) + \left(\frac{d}{dt} y(t) \right)^2 \right] dt ; \quad \max \{ J(y(t)) \} ?$$

Розв'язання задачі

1. Записуємо функцію - підінтегральний вираз функціонала (функцію y та її похідну записуємо як звичайні змінні: Y та PY):

$$F(Y, PY) := Y^2 + 2Y \cdot PY + PY^2.$$

2. Отримуємо відповідні похідні від функції, що входять до рівняння Ейлера:

$$\frac{\partial}{\partial Y} F(Y, PY) \rightarrow 2 \cdot Y + 2 \cdot PY; \quad \frac{\partial}{\partial PY} F(Y, PY) \rightarrow 2 \cdot Y + 2 \cdot PY.$$

Потрібні певні навички у користуванні символьним процесором. Отак, у два етапи, знаходимо складну похідну тому, що цей вираз не під силу символьному процесору:

$$\frac{d}{dt} \frac{\partial}{\partial PY} F(Y, PY) = \frac{d}{dt} (2 \cdot Y(t) + 2 \cdot PY(t)) \rightarrow$$

3. Записуємо рівняння Ейлера у загальному вигляді, підставляємо вирази для похідних, виконуємо алгебраїчні перетворення.

$$\frac{\partial}{\partial Y} F(Y, PY) - \frac{\partial}{\partial t} \frac{\partial}{\partial PY} F(Y, PY) = 0 \Leftrightarrow 2 \cdot Y + 2 \cdot \frac{d}{dt} Y - 2 \cdot \frac{d}{dt} Y - 2 \cdot \frac{d^2}{dt^2} Y = 0 \Leftrightarrow$$

$$\frac{d^2}{dt^2} Y - Y = 0. \text{ Записуємо характеристичне рівняння } s^2 - 1 = 0.$$

4. Розв'язуємо отримане дифрівняння (в даному випадку - звичайне, лінійне, другого порядку), використовуємо символьний процесор для знаходження коренів характеристичного рівняння s_1 та s_2 .

$$s^2 - 1 = 0 \text{ solve, } s \rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix}; \quad s_1 := 1; \quad s_2 := -1.$$

Запишемо розв'язання цього рівняння у загальному вигляді

$$y(t) = C_1 \cdot e^{s_1 \cdot t} + C_2 \cdot e^{s_2 \cdot t}$$

З граничних умов визначаємо константи C_1 та C_2 .

$$\text{при } t_0 := 1 \quad y_0 := 2, \text{ тобто } y(0) = C_1 \cdot e^{s_1 \cdot 1} + C_2 \cdot e^{s_2 \cdot 1};$$

$$\text{при } t_k := 2 \quad y_k := 0, \text{ тобто } y(1) = C_1 \cdot e^{s_1 \cdot 2} + C_2 \cdot e^{s_2 \cdot 2}.$$

Маємо систему з двох рівнянь з двома невідомими. Використовуємо розв'язувальний блок і знаходимо C_1 та C_2 спочатку у числовому вигляді.

Початкові значення шуканих змінних $Cv_1 := 1$; $Cv_2 := 2$;

$$\left. \begin{aligned} \text{Given} \quad ; \quad Cv_1 \cdot e^{s_1 \cdot t_0} + Cv_2 \cdot e^{s_2 \cdot t_0} &= y_0; \\ Cv_1 \cdot e^{s_1 \cdot t_k} + Cv_2 \cdot e^{s_2 \cdot t_k} &= y_k; \end{aligned} \right\}$$

$$\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} := \text{Find}(Cv_1, Cv_2); \quad \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} -0.12 \\ 6.29 \end{pmatrix}.$$

Для узагальнення і впевненості - визначаємо C_1 та C_2 у символьному вигляді:

Given

$$\left. \begin{aligned} C_1 \cdot e^{s_1 \cdot T_0} + C_2 \cdot e^{s_2 \cdot T_0} &= Y_0; \\ C_1 \cdot e^{s_1 \cdot T_k} + C_2 \cdot e^{s_2 \cdot T_k} &= Y_k; \end{aligned} \right\}$$

$$\text{Find}(C_1, C_2) \rightarrow \left[\begin{aligned} &\frac{(\exp(s_2 \cdot T_k) \cdot Y_0 - Y_k \cdot \exp(s_2 \cdot T_0))}{(\exp(s_1 \cdot T_0 + s_2 \cdot T_k) - \exp(s_1 \cdot T_k + s_2 \cdot T_0))} \\ &\frac{(\exp(s_1 \cdot T_0) \cdot Y_k - Y_0 \cdot \exp(s_1 \cdot T_k))}{(\exp(s_1 \cdot T_0 + s_2 \cdot T_k) - \exp(s_1 \cdot T_k + s_2 \cdot T_0))} \end{aligned} \right];$$

$$\begin{pmatrix} C1_{sim} \\ C2_{sim} \end{pmatrix} := \begin{bmatrix} \frac{(\exp(s2 \cdot tk) \cdot y0 - yk \cdot \exp(s2 \cdot t0))}{(\exp(s1 \cdot t0 + s2 \cdot tk) - \exp(s2 \cdot t0 + s1 \cdot tk))} \\ \frac{-(-\exp(s1 \cdot t0) \cdot yk + y0 \cdot \exp(s1 \cdot tk))}{(\exp(s1 \cdot t0 + s2 \cdot tk) - \exp(s2 \cdot t0 + s1 \cdot tk))} \end{bmatrix} \quad \begin{pmatrix} C1 \\ C2 \end{pmatrix} = \begin{pmatrix} -1.16 \\ 3.16 \end{pmatrix}$$

$$\begin{pmatrix} C1_{sim} \\ C2_{sim} \end{pmatrix} = \begin{pmatrix} -1.16 \\ 3.16 \end{pmatrix}$$

Бачимо, що отримані альтернативними методами розв'язки для $C1$ та $C2$ збігаються. Записуємо робочі вирази для отриманої оптимальної функції $Y_{opt}(t)$, похідної $pY(t)$ підінтегрального виразу $piv(t)$ і будуємо графіки. Бачимо, що функція $Y_{opt}(t)$, що дає мінімум функціоналу, знову досить близька до прямої. Знову задаємо питання - чи не можна було прийти до цього розв'язання більш легкими шляхами?

$$Y_{opt}(t) := (C1 \cdot e^{s1 \cdot t} + C2 \cdot e^{s2 \cdot t}); \quad Nn := 25; \quad t := t0, t0 + \frac{(tk - t0)}{Nn} .. tk$$

$$piv(t) := Y_{opt}(t) + 2 \cdot Y_{opt}(t) \frac{d}{dt} Y_{opt}(t) + \left(\frac{d}{dt} Y_{opt}(t) \right)^2 \quad pY(t) := \frac{d}{dt} Y_{opt}(t)$$

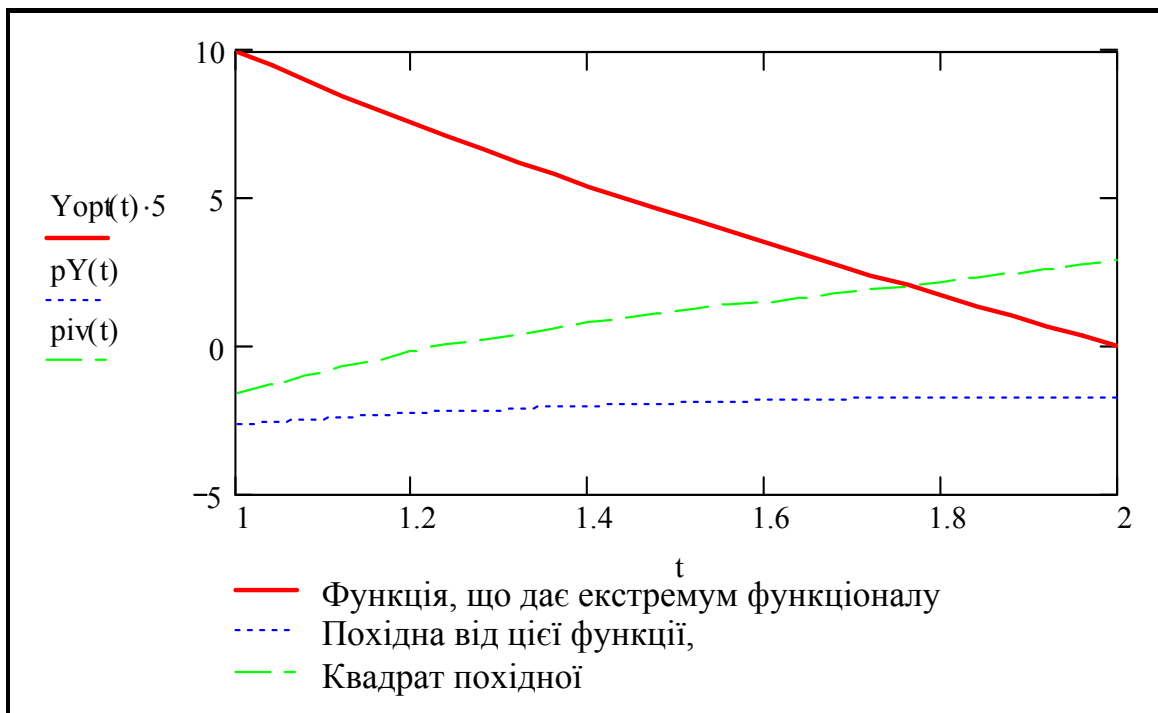
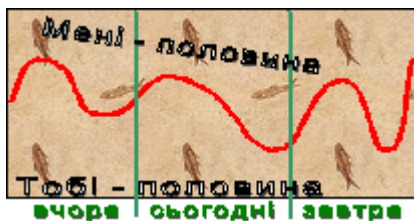


Рис.5. Оптимальний за інтегральним критерієм процес

Контрольні запитання

1. Визначення і приклади функції, функціоналу, оператора.
2. Визначення і приклади варіації функції.
3. Визначення і приклади варіації функціоналу.
3. Порядок розв'язання класичної варіаційної задачі.
4. Порядок розв'язання класичної варіаційної задачі при наявності обмежень.
5. Класифікація обмежень у оптимізаційних задачах.
6. Типові інтегральні критерії оптимізаційних задач техніки та економіки.



5.2. Двовимірна задача Марковіца

Вступ

В цьому розділі продовжується дослідження задачі Марковіца - вже для випадків, коли виробляється два і більше продуктів. З точки зору інформаційних технологій, зокрема, технологій програмування, це приклад "масштабування" задачі. З точки зору теорії управління і дослідження операцій - це задача оптимального управління розвитком виробничої системи. Суть управління - **розподіл ресурсу в просторі** (між виробництвами) і **часі**. Поки ми розглядали задачі розглядали задачі з розподілом обмеженого ресурсу тільки в просторі, або тільки в часі (розділ 5.1)

В розділі 5.1 розроблено документ для випадку однопродуктового виробництва. Щоб зробити документ досить автономним, ми повторюємо постановку задачі Марковіца, а потім копіюємо і модифікуємо програмні модулі з розділу 5.1.

Постановка задачі

Розглядаємо ідеалізовану задачу з математичної економіки. Маємо економічну систему, в якій виробляються N видів продукції, що випускаються з темпами $x_1(t)$, $x_2(t)$, $x_3(t)$, ..., $x_N(t)$ (одиниць вимірювання продукції за місяць, квартал, рік). Англomовний термін "темп виробництва" - це синонім "виробничої потужності". Записуємо умови задачі.

$$\frac{d}{dt}x(t)_i = \sum_{j=1}^N a_{i,j} \cdot y(t)_j \quad \text{- рівняння для виробничих потужностей;}$$

$x(t)_i$ - поточна виробнича потужність по i -му продукту;

$y(t)_j$ - поточні інвестиції j -го продукту в розширення виробничих потужностей по i -му;

$a_{i,j} \geq 0$ - параметр рівняння динаміки виробничих потужностей;

$x(0)_j = c_j$ - початковий рівень виробничих потужностей по i -му продукту;

$y(t)_j \leq x(t)_j$ - обмеження рівня інвестицій. $\text{ORIGIN} := 1$; $i := 1..N$ $j := 1..N$.

Звернемо увагу на виділену нерівність - вона інтерпретується як "поточний розмір інвестицій не повинен бути більшим поточного доходу (обсягу виробництва)". Потім, на основі отриманого розв'язання задачі, можна буде розглянути більш реалістичні задачі - з кредитами, акціонерним капіталом та ін. Згідно з умовами задачі, потік продукції $x_{t,i}$ розподіляється на дві складові: $y_{t,i}$ та $z_{t,i}$ - темп (потік) інвестицій у розширення виробництва і темп прибутку, що використовується десь за межами даного виробництва, тобто:

$$z_{t,i} = x_{t,i} - y_{t,i}$$

Потрібно визначити оптимальну стратегію інвестицій у виробництво (не плутати з поточними витратами на виробництво видів продукції), що максимізує сумарний прибуток за певний період:

$$JN = \int_0^T \sum_{j=1}^N (x_{t,i} - y_{t,i}) dt.$$

Двовимірна задача

Розглянемо розв'язання двовимірної задачі. Розв'язання багатомірних задач має ту ж структуру і властивості, але воно більш трудомістке - більше змінних, більше рівнянь.

Записуємо дифрівняння: $\frac{d}{dt}x_1 = a_{11} \cdot y_1 + a_{12} \cdot y_2$; $\frac{d}{dt}x_2 = a_{21} \cdot y_1 + a_{22} \cdot y_2$;

- це обмеження задачі оптимізації прибутку, що задані дифрівняннями;

обмеження на позитивність усіх коефіцієнтів: $a_{ij} \geq 0$, $i=1,2$, $j=1,2$

початкові умови - значення для x_1, x_2 при $t=0$: $x(0)_1 = c_1$; $x(0)_2 = c_2$.

Вводимо позначення $z_1 = x_1 - y_1$, $z_2 = x_2 - y_2$ і, нарешті, записуємо функціонал

$$J_2 = \int_0^T z_1 + z_2 dt \quad (1)$$

Ціль оптимізації - максимум функціоналу. **Змінні управління** (те, що від нас залежить) - функції інвестицій $y_{1op}(t), y_{2op}(t)$.

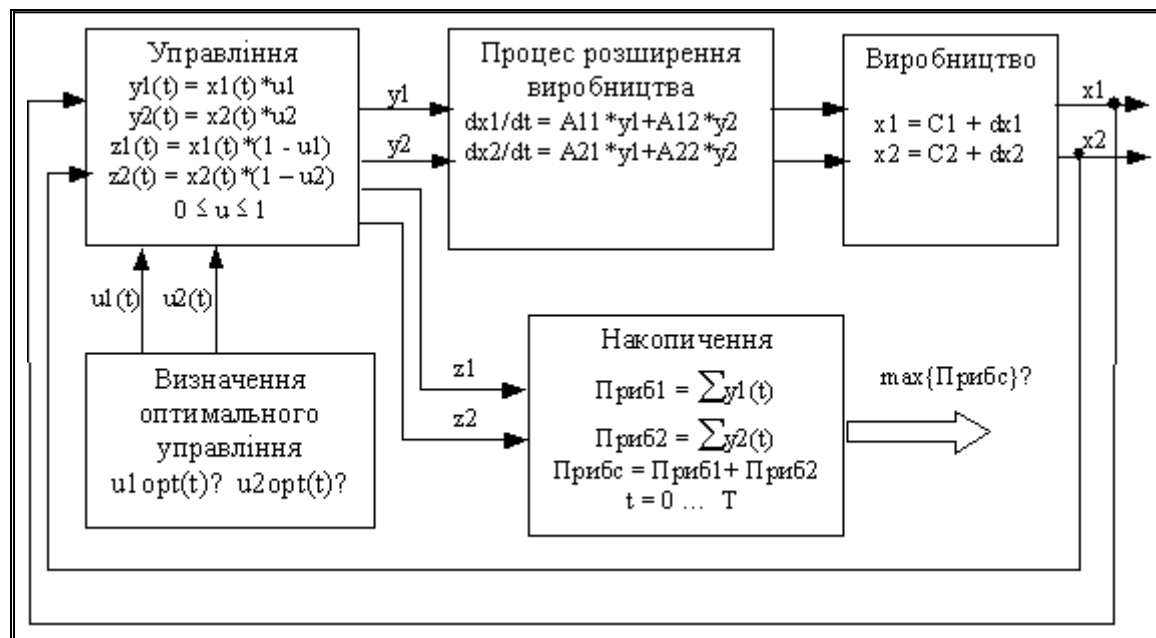


Рис.1. Схема задачі Марковіца оптимізації сумарного прибутку.

Випадок двотоварного виробництва

Розв'язання задачі

Розв'яжемо дифрівняння росту виробничих потужностей відносно x_1 та x_2 :

$$\frac{d}{dt}x_i = a_{i,1} \cdot y_1 + a_{i,2} \cdot y_2 ; \quad dx_i = a_{i,1} \cdot y(t)_1 \cdot dt + a_{i,2} \cdot y(t)_2 \cdot dt \quad i = 1..2$$

$$x(t)_1 = c_1 + a_{1,1} \cdot \int_0^t y(t)_1 dt + a_{1,2} \cdot \int_0^t y(t)_2 dt \quad (2)$$

Тут сколапсовано (без пароля) робочі залежності для рис. 2:

Підставимо вираз (2) у функціонал (1):

$$J_2 = \int_0^T \left(c_1 + a_{11} \cdot \int_0^t y_1 dt_1 + a_{12} \cdot \int_0^t y_2 dt_1 - y_1 \right) dt \dots \quad (3)$$

$$+ \int_0^T \left(c_2 + a_{21} \cdot \int_0^t y_1 dt_1 + a_{22} \cdot \int_0^t y_2 dt_1 - y_2 \right) dt$$

Дивимось на вираз (3). Інтеграл в дужках (по змінній t_1) - це обчислення виробничих потужностей на поточний момент часу. Зовнішній інтеграл (по змінній t) характеризує сумарну кількість продукції, що буде вироблена на поточних потужностях до кінця періоду - моменту T . До такого подвійного інтеграла можна застосувати операцію зміни порядку інтегрування (див. курс вищої математики). Навіщо це? - підінтегральні вирази можна розділити на множники, що залежать тільки від часу, і множники, що залежать тільки від управління - тоді легко буде зробити висновки про структуру оптимального управління. Далі подано остаточний вираз (4) після зміни порядку інтегрування (виведіть самостійно). Як підказку (для тих, "хто не любить математику") даємо дискретну графічну інтерпретацію подвійного інтеграла (рис.2).

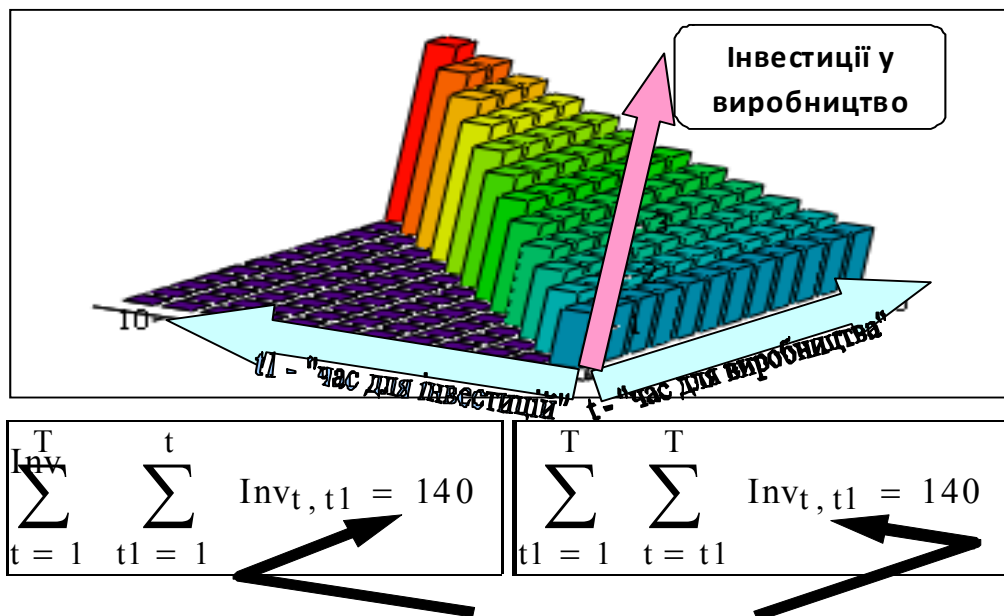


Рис.2. Графічна інтерпретація подвійного інтеграла

Дивимось на вираз для критерію (4), що отриманий після зміни порядку інтегрування, і бачимо, що вирази у дужках під інтегралами:

$$[(a_{11} + a_{12})(T - t) - 1], [(a_{12} + a_{22})(T - t) - 1]$$

є функціями часу, що не залежать від управління.

$$J_2 = (c_1 + c_2) \cdot T \cdot \left[\int_0^T [(a_{11} + a_{12})(T - t) - 1] \cdot y_1 dt + \int_0^T [(a_{12} + a_{22})(T - t) - 1] \cdot y_2 dt \right] \quad (4)$$

Для максимізації функціоналу (4) потрібно управління y_1 та y_2 вибирати таким, щоб величина їх була максимально можливою, а знак збігався із знаком виразів у дужках. Але, згідно з умовами задачі, y_1 та y_2 не можуть бути від'ємними, тому, коли вирази у дужках стають від'ємними, то y_1, y_2 повинні бути нульовими.

Таким чином, стратегія управління має бути такою: **спочатку все вироблене (і продане) йде на інвестиції у розширення виробництва, починаючи з певних моментів T_1, T_2 все вже йде у накопичення, а виробничі потужності залишаються постійними.** Визначимо ці моменти "переключення".

Визначення моментів переключення

Виведемо формули для моментів переключення управління T_1, T_2 (з стратегії "усе у виробництво" в стратегію "усе у накопичення/споживання"). Використовуємо для цього символічне розв'язання системи алгебраїчних рівнянь.

$$\text{Given} \quad (a_{11} + a_{21})(T - T_1) - 1 = 0 ; \quad (a_{12} + a_{22})(T - T_2) - 1 = 0 ;$$

$$\text{Find}(T_1, T_2) \rightarrow \left[\begin{array}{c} \frac{(12 \cdot a_{11} + 12 \cdot a_{21} - 1)}{(a_{11} + a_{21})} \\ \frac{(12 \cdot a_{12} + 12 \cdot a_{22} - 1)}{(a_{12} + a_{22})} \end{array} \right] ; \quad \text{звідки маємо:}$$

$$T_1 = \frac{(a_{11} + a_{21}) \cdot T - 1}{a_{11} + a_{21}} ; \quad T_2 = \frac{(a_{12} + a_{22}) \cdot T - 1}{a_{12} + a_{22}} ; \quad (5)$$

Припустимо, що $T_2 > T_1$. Тоді оптимальна стратегія управління буде такою:

$$\begin{array}{ll} \text{весь 1-ий продукт у інвестиції: } z_1 = 0 & \text{при } 0 \leq t \leq T_1 \\ \text{весь 2-ий продукт у інвестиції: } z_2 = 0 & \text{при } 0 \leq t \leq T_2 \\ \text{все у накопичення} & y_1 = y_2 = 0 \text{ при } T_2 \leq t \leq T \end{array} \quad (6)$$

Нам потрібно визначити оптимальне управління першим виробництвом в інтервалі $T_1 \leq t \leq T_2$. Забігаючи вперед скажемо, що до певного моменту $T_1 \leq T_3 \leq T_2$ все вироблене - $x_1(t)$ повинно йти в інвестиції, потім - все у накопичення. Довести це досить важко, в першоджерелі доведення - це дві сторінки формул. Отримати аналітичний вираз для моменту переключення T_3 - ще важче (там не наводиться). Тому **момент переключення T_3 будемо знаходити числовим - пошуковим методом.**

Тільки для бажаючих! та інших ОАА наводимо доведення виразу для оптимального управління на інтервалі $T_1 \leq t \leq T_2$ - **точно за першоджерелом.**

Для визначення значення стратегії z_1 при $T_1 \leq t \leq T_2$ використаємо (4) і (6):

$$J_2 = \int_{T_1}^{T_2} z_1 dt + (T - T_2) \cdot (x_1(T_2) + x_2(T_2)) \cdot \quad (7)$$

Враховуємо (2) для визначення $x_1(T_2)$:

$$x(T_2)_1 = c_1 + a_{1,1} \cdot \int_0^{T_2} y(t)_1 dt + a_{1,2} \cdot \int_0^{T_2} y(t)_2 dt = c_3 + a_{1,1} \cdot \int_{T_1}^{T_2} y(t)_1 dt + a_{1,2} \cdot \int_{T_1}^{T_2} x(t)_2 dt$$

де c_3 - константа, що не залежить від значення $y(t)_2$ на інтервалі $[T_1, T_2]$.

Аналогічно знаходимо

$$x(T_2)_1 = c_4 + a_{2,1} \cdot \int_{T_1}^{T_2} y(t)_1 dt + a_{2,2} \cdot \int_{T_1}^{T_2} x(t)_2 dt.$$

Визначаємо накопичення z_1 через x_1 - виробництво та y_1 – інвестиції

$$z_1 = x_1 - y_1 = c_1 + a_{1,1} \cdot \int_0^t y(t)_1 dt + a_{1,2} \cdot \int_0^t y(t)_2 dt - y_1.$$

Підставляємо у вираз (7) для критерію і після перетворень отримуємо

$$J_2 = c_5 + \int_{T_1}^{T_2} (c_6(T-t) - 1) \cdot y_1 dt + a_{1,2} \cdot \int_{T_1}^{T_2} \int_0^t y_2 dt_1 dt + (a_{1,2} + a_{2,1}) \cdot \int_{T_1}^{T_2} y_2 dt_1 \quad (8)$$

Визначимо інтеграл $\int_0^t y_2 dt$ на інтервалі $T_1 \leq t \leq T_2$:

$$\frac{d}{dt} \left(\int_0^t y_2 dt \right) = y_2 = c_1 + a_{2,1} \cdot \int_0^t y_1 dt + a_{2,2} \cdot \int_0^t y_2 dt.$$

Розв'язуємо цей вираз як диференціальне рівняння із змінною $\left(\int_0^t y_2 dt \right)$:

$$\begin{aligned} \int_0^t y_2 dt &= e^{a_{2,2} \cdot t} \cdot \int_0^t \left(c_2 \cdot e^{-a_{2,2} \cdot s} + a_{2,1} \cdot \int_0^s y_1 dt_1 \right) ds = \\ &= \phi_1(t) + a_{2,1} \cdot e^{-a_{2,2} \cdot t} \cdot \int_{T_1}^t \left(e^{-a_{2,2} \cdot s} \cdot \int_{T_1}^s y_1 dt_1 \right) ds. \end{aligned}$$

Функція $\phi_1(t)$ не залежить від значення y_1 на інтервалі $T_1 \leq t \leq T_2$, тому маємо:

$$\int_{T_1}^{T_2} x_2 dt = \int_{T_1}^{T_2} y_2 dt = c_7 + a_{2,1} \cdot e^{a_{2,2} \cdot T_2} \cdot \int_{T_1}^{T_2} \left(e^{-a_{2,2} \cdot s} \cdot \int_{T_1}^s y_1 dt_1 \right) ds.$$

Змінюємо порядок інтегрування і виконуємо алгебраїчні перетворення

$$\int_{T1}^{T2} y_2 dt = c_7 + a_{2,1} \cdot e^{a_{2,2} \cdot T2} \cdot \int_{T1}^{T2} \left(y_1(t_1) \cdot \int_{t_1}^{T2} e^{-a_{2,2} \cdot s} ds \right) dt_1.$$

Дивимось на коефіцієнт при $y_1(t_1)$: $a_{2,1} \cdot e^{a_{2,2} \cdot T2} \cdot \int_{t_1}^{T2} e^{-a_{2,2} \cdot s} ds$ і бачимо, що він є функцією t_1 , що монотонно зменшується. Виконаємо спрощення виразу

$$\int_{T1}^{T2} \left(\int_0^t y_2(t) dt \right) dt, \text{ що входить до критерію } J_2 \text{ у виразі (8).}$$

$$\begin{aligned} \int_{T1}^{T2} \left(\int_0^t y_2(t) dt \right) dt &= c_8 + a_{2,1} \cdot \int_{T1}^{T2} \left[e^{a_{2,2} \cdot t} \cdot \int_{T1}^t \left(e^{-a_{2,2} \cdot s} \cdot \int_{T1}^s y_1(t_1) dt_1 \right) ds \right] dt = \\ &= c_8 + a_{2,1} \cdot \int_{T1}^{T2} \left[e^{-a_{2,2} \cdot t} \cdot \int_{T1}^t \left(y_1(t_1) \cdot \int_{t_1}^t e^{-a_{2,2} \cdot s} ds \right) dt_1 \right] dt. \end{aligned}$$

Останній інтеграл можна подати у вигляді

$$\int_{T1}^{T2} e^{a_{2,2} \cdot t} \cdot \left(\int_{T1}^t y_1(t_1) \cdot \psi(t, t_1) dt_1 \right) dt = \int_{T1}^{T2} y_1(t_1) \cdot \left(\int_{t_1}^{T2} e^{a_{2,2} \cdot t} \cdot \psi(t, t_1) dt \right) dt_1 \quad (9)$$

Уважно дивимось на (9) і бачимо - під інтегралом функція управління $y_1(t_1)$ помножена на функцію часу. Дослідимо її - візьмемо похідну по t_1 .

$$\frac{d}{dt_1} \left(\int_{t_1}^{T2} e^{a_{2,2} \cdot t} \cdot \psi(t, t_1) dt \right) = e^{a_{2,2} \cdot t} \cdot \psi(t_1, t_1) + \int_{t_1}^{T2} e^{a_{2,2} \cdot t} \cdot \frac{d}{dt_1} \psi(t, t_1) dt.$$

Ураховуємо визначення функції $\psi(t, t_1)$ (див вище) і нарешті отримуємо

$$e^{a_{2,2} \cdot t} \cdot \psi(t_1, t_1) + \int_{t_1}^{T2} e^{a_{2,2} \cdot t} \cdot \frac{d}{dt_1} \psi(t, t_1) dt = 0 + \int_{t_1}^{T2} e^{a_{2,2} \cdot t} \cdot (-e^{-a_{2,2} \cdot t_1}) dt_1.$$

Тепер чітко бачимо, що коефіцієнт при $y_1(t_1)$ в інтегралі (9) - функція часу, що монотонно зменшується. Дивимосся на вираз (8) для критерію, бачимо також, що $c_6 > 0$, тому функція $c_6 \cdot (T - t) - 1$ монотонно зменшується по t і може змінювати знак.

Таким чином, повний коефіцієнт при y_1 зменшується в інтервалі (T_1, T_2) . Тепер можемо записати те, за що ми боролися на двох сторінках з інтегралами:

$$J_2 = c_8 + \int_{T_1}^{T_2} k(t_1) \cdot y_1(t_1) dt_1 .$$

Тому, для максимізації прибутку слід вибирати y_1 найбільшим можливим на інтервалі $[T_1, T_3]$, де $k_1(t_1) > 0$ і нульовим на інтервалі $[T_3, T_2]$, де $k_1(t_1) < 0$.

Оскільки величина $c_6 \cdot (T - t) - 1$ (див. (8)) буде від'ємною при $t > T_1$, а інші коефіцієнти стануть рівними нулю при $t > T_2$, то дійсно $T_1 < T_3 < T_2$.

Зауваження.

Для тих, хто більш схильний до математики. Типовий прийом у варіаційному обчисленні, методах принципу максимуму, динамічного програмування - так перетворити інтегральний критерій, щоб під інтегралом був добуток функції часу і функції управління. Це дає можливість визначити структуру і властивості оптимального управління, "вгадати" його вигляд, побудувати методи його визначення.

Як же привести критерій до бажаної форми? - Іноді, хтось, якимось додумується, і з'являються очевидні для усіх "підстановка Ейлера", "задача Марковіца" та ін.

Для тих, хто більш схильний до менеджменту. Ми витрачали багато зусиль на те, щоб під інтегралом був добуток функції управління, що залежить від нас і функції часу, яка від нас не залежить, іноді ще й невизначена. В термінальних задачах управління, коли треба досягнути певної цілі на певний момент часу, функція, що від нас не залежить, зменшується з наближенням терміну. Це означає, що, чим менше часу до кінця процесу, тим більшими стають потрібні для досягнення одного й того ж результату управління. Менеджер знає це і без математики - із досвіду. Та не завжди цей досвід є. Математик може витратити 20 років на отримання розв'язання задачі, на яку менеджеру дається година.

Таким чином, показано, що максимальний прибуток за час T при **обмеженнях** (1) ОСТАТОЧНО визначається такими умовами:

$$\left. \begin{array}{ll} \text{нема} & z_1 = 0 \quad \text{при} \quad 0 \leq t \leq T_3 \\ \text{накопичень} & z_2 = 0 \quad \text{при} \quad 0 \leq t \leq T_2 \\ \text{нема} & y_1 = 0 \quad \text{при} \quad T_3 \leq t \leq T \\ \text{інвестицій} & y_2 = 0 \quad \text{при} \quad T_2 \leq t \leq T \end{array} \right\} \quad (10)$$

За рахунок перестановки рівнянь інвестиційних процесів і вибору тривалості процесу T завжди можна зробити: $T_3 \leq T_2 \leq T$

Розробка робочої програми для розв'язання задачі Марковіца

Ми бажаємо не тільки обчислити моменти переключення управління для певної задачі, але й моделювати процес функціонування виробничої системи. Нам потрібно обчислювати поточні виробничі потужності $x_1(t)$, $x_2(t)$, інвестиції $y_1(t)$, $y_2(t)$, "прибутки" $z_1(t)$, $z_2(t)$, сумарний прибуток $S_z(t)$. Отримаємо відповідні вирази і зберемо з них програму. Введемо значення параметрів:

$$A := \begin{pmatrix} 0.05 & 0.1 \\ 0.08 & 0.15 \end{pmatrix}; C := \begin{pmatrix} 17 \\ 2 \end{pmatrix}; T = 12; Kh = 60; k := 1..Kh + 1; t_k := \frac{T}{Kh} \cdot (k - 1).$$

Записуємо вирази для моментів переключення $T1, T2$ (5), і приблизне значення $T3$

$$T1 := \frac{(A_{1,1} + A_{2,1}) \cdot T - 1}{A_{1,1} + A_{2,1}}; T2 := \frac{(A_{1,2} + A_{2,2}) \cdot T - 1}{A_{1,2} + A_{2,2}}; T3o := (T1 + T2) \cdot 0.5;$$

$$\Delta t := T \div Kh \quad T1 = 4.31; \quad T2 = 8; \quad T3o = 6.15.$$

Ще раз нагадаємо, що аналітичний вираз для моменту $T3$ отримати важко, але відомо, що $T1 < T3 < T2$. Пам'ятаємо, що умови задачі звичайно точно невідомі. Візьмемо поки приблизне значення $T3$, яке потім можна буде уточнити пошуковим методом.

Побудуємо програмний модуль та графік переключень управління.

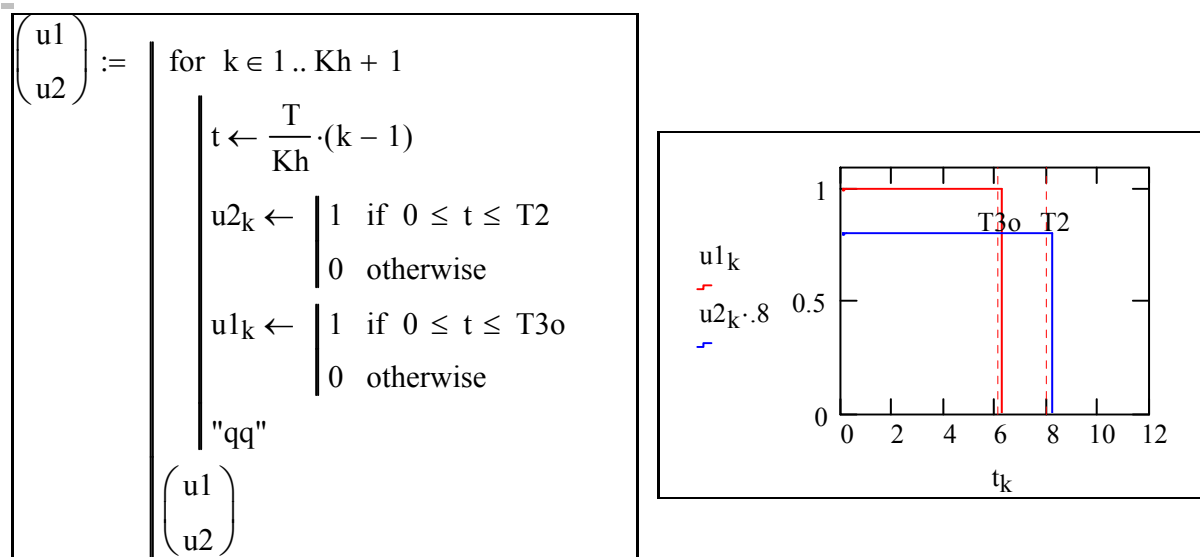


Рис.3. Управління процесом

Потрібні нам залежності для виробничих потужностей та інвестицій в даному випадку можна отримати в

- аналітичній формі (приклад подано в підрозділі 3.1);
- числовими методами інтегрування диференціальних рівнянь.

В загальному випадку аналітичні розв'язання отримати неможливо, тому вибираємо універсальні числові методи. Програму моделювання $MP(A, C, pr)$ оформляємо, як функцію від параметрів: A - матриці параметрів інвестиційного процесу, C - вектора початкових виробничих потужностей, pr - параметра, що задає значення $T3$: $T3 = T1$ при $pr = 0$ та $T3 = T2$ при $pr = 1$. Вхідні дані задаються там далі, **в абсолютному присвоєнні**

Зауваження. В пакеті Mathcad можливо вибирати зображення для основних математичних операторів - зокрема, можливі такі альтернативні форми подання операції ділення

$$\frac{T}{Kh} = 0.2; T \div Kh = 0.2.$$

Нагадуємо також, що змінні, які визначені в середині Mathcad-програми є локальними змінними тільки цієї програми. Змінні, що визначені в документі, є глобальними, Mathcad-програма їх "бачить". Наприклад, змінні T, Kh - глобальні, змінні $Te1, Te2$ визначені в програмі - вони локальні

MP(A, C, pr) :=	$\Delta t \leftarrow T \div Kh$ $x1_1 \leftarrow C_1$ $x2_1 \leftarrow C_2$ $Te1 \leftarrow \frac{(A_{1,1} + A_{2,1}) \cdot T - 1}{A_{1,1} + A_{2,1}}$ $Te2 \leftarrow \frac{(A_{1,2} + A_{2,2}) \cdot T - 1}{A_{1,2} + A_{2,2}}$ $Te3 \leftarrow Te1 \cdot (1 - pr) + Te2 \cdot pr$ for k ∈ 1 .. Kh + 1 $t \leftarrow \frac{T}{Kh} \cdot (k - 1)$ $u1 \leftarrow \begin{cases} 1 & \text{if } 0 \leq t \leq Te3 \\ 0 & \text{otherwise} \end{cases}$ $u2 \leftarrow \begin{cases} 1 & \text{if } 0 \leq t \leq Te2 \\ 0 & \text{otherwise} \end{cases}$ $x1_{k+1} \leftarrow x1_k + (A_{1,1} \cdot x1_k \cdot u1 + A_{1,2} \cdot x2_k \cdot u2) \cdot \Delta t$ $x2_{k+1} \leftarrow x2_k + (A_{2,1} \cdot x1_k \cdot u1 + A_{2,2} \cdot x2_k \cdot u2) \cdot \Delta t$ $z1_k \leftarrow x1_k \cdot (1 - u1)$ $z2_k \leftarrow x2_k \cdot (1 - u2)$ $V_{yx}^{(k)} \leftarrow \begin{pmatrix} x1_k \\ z1_k \\ x2_k \\ z2_k \end{pmatrix}$
	Vyx

Нагадуємо також, що Mathcad-програма повертає в документ те, що записано в останньому рядку. В даному випадку це масив поточних значень інвестицій і прибутків.

MP (A1 , C1 , .6) =

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	13	13	13	13	14	14	14	14	14	15	15	15	15	15	16	16	16
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	2	3	3	3	3	4	4	5	5	5	6	6	7	7	8	8
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Зробимо модуль, що не тільки знаходить оптимальне значення моменту переключення T3, але і дає залежність сумарного прибутку від значення T3.

OpT3(A, C) :=	$\text{Nkrok} \leftarrow 25$ $\text{Te1} \leftarrow \frac{(A_{1,1} + A_{2,1}) \cdot T - 1}{A_{1,1} + A_{2,1}}$ $\text{Te2} \leftarrow \frac{(A_{1,2} + A_{2,2}) \cdot T - 1}{A_{1,2} + A_{2,2}}$ $\text{maxpryb} \leftarrow 0$ $\text{krok} \leftarrow 1 \div \text{Nkrok}$ for i ∈ 1 .. Nkrok $\text{prp} \leftarrow 0 + \text{krok} \cdot (i - 1)$ $\text{bb} \leftarrow \text{MP}(A, C, \text{prp})$ $\text{pryb}_i \leftarrow \sum_{k=1}^{Kh+1} \text{bb}_{2,k} + \sum_{k=1}^{Kh+1} \text{bb}_{4,k}$ if $\text{pryb}_i > \text{maxpryb}$ $\text{maxpryb} \leftarrow \text{pryb}_i$ $\text{prop} \leftarrow \text{prp}$ $\text{Te3}_i \leftarrow \text{Te1} \cdot (1 - \text{prp}) + \text{Te2} \cdot \text{prp}$ $\text{Vyx}^{(i)} \leftarrow \begin{pmatrix} \text{pryb}_i \\ \text{Te3}_i \end{pmatrix}$ $\text{T3o} \leftarrow \text{Te1} \cdot (1 - \text{prop}) + \text{Te2} \cdot \text{prop}$ $\begin{pmatrix} \text{Vyx} \\ \text{maxpryb} \\ \text{prop} \\ \text{T3o} \end{pmatrix}$
---------------	--

Тестуємо модуль, на двох наборах вхідних даних (A,C) та (A1,C1). Наша програма повертає структуру даних з масиву Nkrok x 2 та двох скалярів. **Розпаковуємо** виходи програми для кожного набору даних.

"пропорція" для T3: значення T3: масив поточних даних: i := 1..25

propq := OpT3(A1,C1)₃ ; T3q := OpT3(A1,C1)₄ ; qq := OpT3(A1,C1)₁ ;

propk := OpT3(A,C)₃ ; T3k := OpT3(A,C)₄ ; qk := OpT3(A,C)₁ .

Значення максимальних прибутків:

maxprbq := OpT3(A1,C1)₂ ; maxprbk := OpT3(A,C)₂ . Виводимо результати.

	1	2	3	4	5	6	7	8	9	10	11	12	13	
qk =	1	1607	1610	1612	1612	1614	1614	1614	1613	1611	1608	1608	1604	1599
	2	4	4	5	5	5	5	5	5	5	6	6	6	6

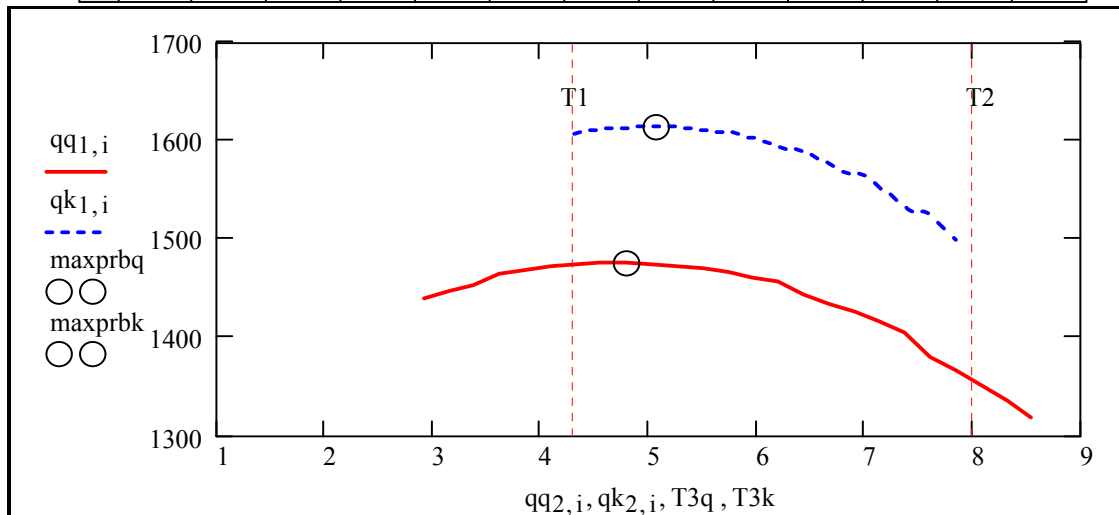


Рис.4. Визначення оптимальних моментів переключення виробництва

Дивимось на графіки і робимо перший висновок - момент T_3 можна визначати приблизно - екстремум розмитий. Другий висновок - програма працює - вона дійсно ловить екстремум. Слід однак вбудувати в програму "запобіжник" для випадків $T_2 < T_1$, $T_1 < 0$. Це робиться за допомогою команди: `on error`, або функції `error("x повинно бути позитивним") if x ≤ 0`.

Тепер на базі розроблених програм побудуємо **стенд** для дослідження виробництва та інвестиційних процесів. Вхідні дані A та C вводимо в абсолютному присвоєнні. Це дозволяє розмістити їх поряд з графіками та вихідними даними.

Обчислюємо потрібні дані, наприклад, моменти переключення T_1 , T_2 . Їх можна було б обчислити в програмі моделювання, але ми не поступаємо принципом - все повинно бути відкритим для користувача, і записане по можливості в "природному вигляді" - як писали Ньютон та Ейлер (це копія екрану електронної книги).

$$T_1 := \frac{(A_{1,1} + A_{1,2}) \cdot T - 1}{A_{1,1} + A_{1,2}} \quad T_2 := \frac{(A_{1,2} + A_{1,2}) \cdot T - 1}{A_{1,2} + A_{1,2}}$$

Обчислюємо оптимальне T_3 та процес (інвестицій та накопичень) для заданих A , C .

$$T_3 := \text{OpT3}(A, C) \quad \text{prop} := \text{OpT3}(A, C) \quad \text{byb} := \text{MP}(A, C, \text{prop})$$

Обчислюємо прибутки - по окремих продуктах та сумарний.

$$\text{Прибу1} := \sum_{k=1}^{Kh+1} \text{byb}_{2,k} \quad \text{Прибу2} := \sum_{k=1}^{Kh+1} \text{byb}_{4,k} \quad \text{TotalPr} := \text{Прибу1} + \text{Прибу2}$$

Тривалість періоду $T \equiv 12$; кількість кроків моделювання $Kh \equiv 60$; стартові темпи виробництв	$C1 \equiv \begin{pmatrix} 13 \\ 2 \end{pmatrix}$; матриця інвестиційного процесу:	$A1 \equiv \begin{pmatrix} 0.05 & 0.06 \\ 0.06 & 0.25 \end{pmatrix}$
--	---	--

На рис. 5 подано копію екрану "результати моделювання"

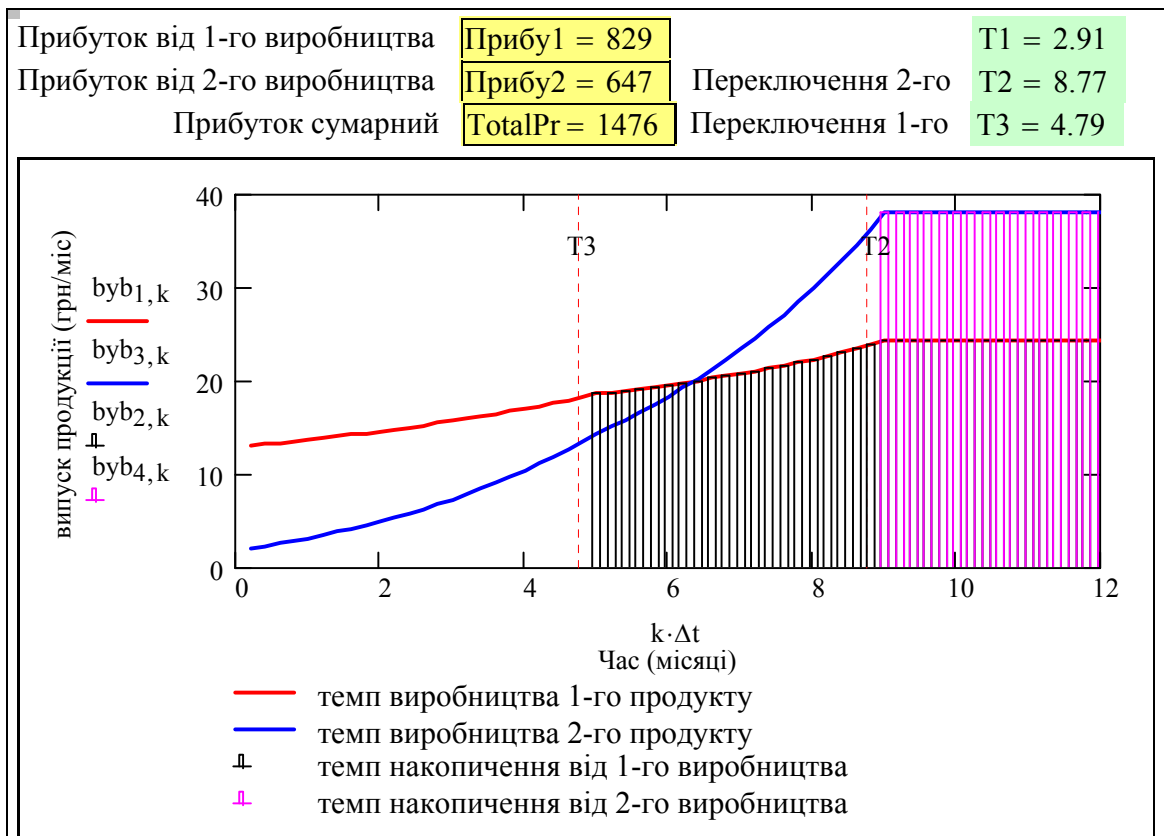


Рис.5. Процеси функціонування виробничої системи

Подаємо результати моделювання у вигляді тривимірної стовпчикової діаграми. Якщо ще підписати рядки, така діаграма полегшує сприймання даних багатомірних задач. Для цього слід використати один з дорогих професійних графічних пакетів.

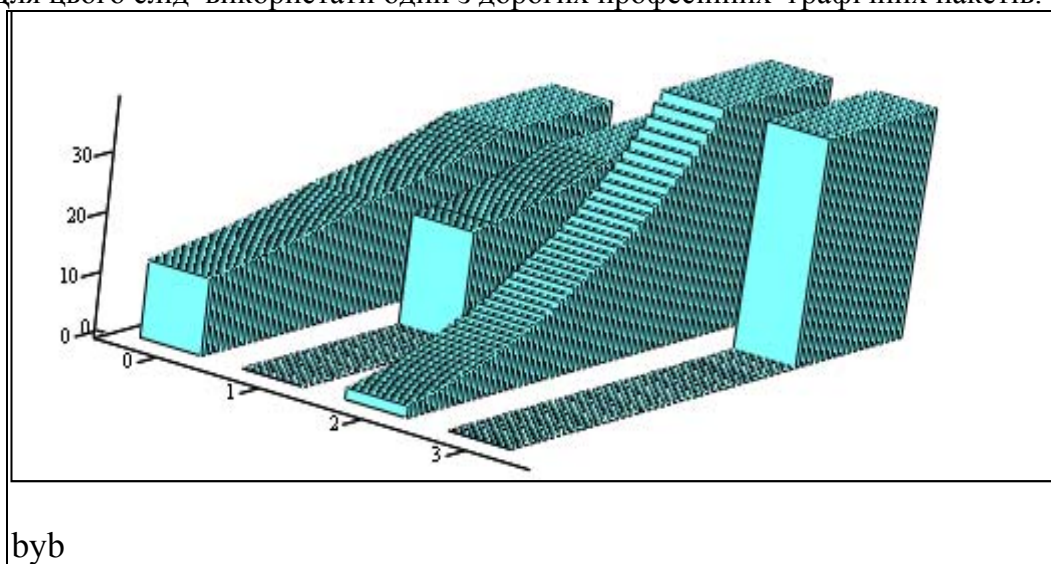


Рис.6. Процеси функціонування виробничої системи

На цьому графіку подано таку ситуацію - перше виробництво – з малою віддачею інвестицій і великим початковим рівнем; друге виробництво – з великою віддачею інвестицій і малим початковим рівнем. Те, що бачимо, інтерпретується так: досить рано $T3 = 4.79$ вихід 1-го виробництва забирається у накопичення (другий рядок на графіку); вихід 2-го ще йде в інвестиції – обидва виробництва (рядки 1 і 3) зростають, але дещо повільніше. З моменту $T2 = 8.77$ все йде в накопичення

Організація досліджень "що буде якщо". Приклад

Далі подано завдання для **самостійного** виконання. Для підкріплення самостійності подаємо приклад використання розроблених програм для аналізу чутливості оптимального управління. Менеджера завжди цікавить розкид вихідних величин при можливих розкидах вхідних величин. Дослідимо вплив відхилення моменту переключення T2 на розмір прибутку. Для цього дещо модифікуємо програму моделювання.

VP(A, C, vT2) :=	$\Delta t \leftarrow T \div Kh$ $x1_1 \leftarrow C_1$ $x2_1 \leftarrow C_2$ $Te2 \leftarrow \frac{(A_{1,2} + A_{2,2}) \cdot T - 1}{A_{1,2} + A_{2,2}}$ $Te3 \leftarrow OpT3(A, C)_4$ $\text{for } k \in 1 .. Kh + 1$ $t \leftarrow \frac{T}{Kh} \cdot (k - 1)$ $u1 \leftarrow \begin{cases} 1 & \text{if } 0 \leq t \leq Te3 \\ 0 & \text{otherwise} \end{cases}$ $u2 \leftarrow \begin{cases} 1 & \text{if } 0 \leq t \leq Te2 \cdot vT2 \\ 0 & \text{otherwise} \end{cases}$ $x1_{k+1} \leftarrow x1_k + (A_{1,1} \cdot x1_k \cdot u1 + A_{1,2} \cdot x2_k \cdot u2) \cdot \Delta t$ $x2_{k+1} \leftarrow x2_k + (A_{2,1} \cdot x1_k \cdot u1 + A_{2,2} \cdot x2_k \cdot u2) \cdot \Delta t$ $z1_k \leftarrow x1_k \cdot (1 - u1)$ $z2_k \leftarrow x2_k \cdot (1 - u2)$ $pryb \leftarrow \sum_{k=1}^{Kh+1} z1_k + \sum_{k=1}^{Kh+1} z2_k$ $pryb$
------------------	--

Обчислюємо нормовані прибутки при відхиленнях параметра T2

$$\text{Minus}\Phi(A, C, dd) := \frac{VP[A, C, (1 - dd)]}{VP(A, C, 1)}; \quad \text{Plus}\Phi(A, C, dd) := \frac{VP[A, C, (1 + dd)]}{VP(A, C, 1)};$$

$$va := 0.5, 0.55 .. T \div T2; \quad T \div T2 = 1.37; \quad T = 12; \quad Te2 := \frac{(A_{1,2} + A_{2,2}) \cdot T - 1}{A_{1,2} + A_{2,2}}.$$

$$A = \begin{pmatrix} 0.05 & 0.1 \\ 0.08 & 0.15 \end{pmatrix}; C = \begin{pmatrix} 17 \\ 2 \end{pmatrix}; \quad A1 = \begin{pmatrix} 0.05 & 0.06 \\ 0.06 & 0.25 \end{pmatrix}; C1 = \begin{pmatrix} 13 \\ 2 \end{pmatrix};$$

$$\text{Розкид}T2 := 0.2$$

$$\text{Minus}\Phi(A, C, \text{Розкид}T2) = 0.98; \quad \text{Minus}\Phi(A1, C1, \text{Розкид}T2) = 0.96;$$

$$\text{Plus}\Phi(A, C, \text{Розкид}T2) = 0.97; \quad \text{Plus}\Phi(A1, C1, \text{Розкид}T2) = 0.92.$$

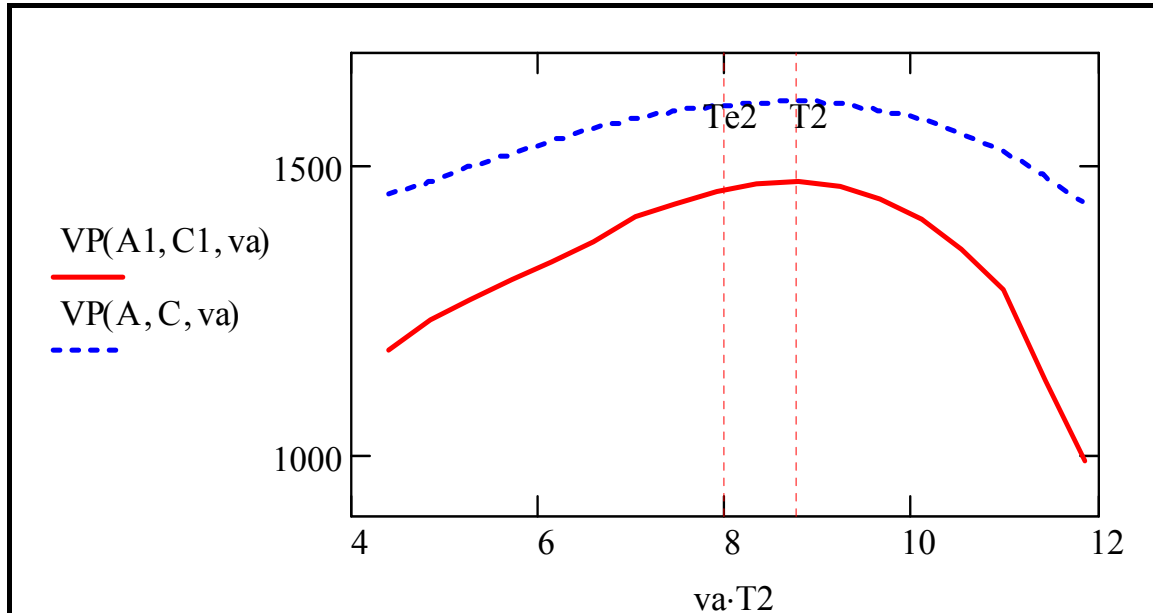
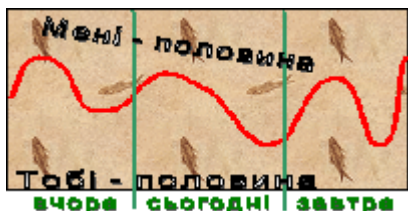


Рис.7. Вплив відхилень моменту переключення $T2$ на сумарний прибуток

Робимо попередні висновки. Чутливість оптимального управління залежить від параметрів матриці A . Запізнення переключення дає більші втрати, ніж випередження. Система досить "гуманна" - при 20% розкиді $T2$ втрати доходу становлять 2-8%.

Контрольні запитання

1. Постановка задачі Марковіца - критерій, обмеження, змінні управління, граничні умови
2. Визначення функції, функціоналу, оператора.
3. Визначення варіаційної задачі. Що потрібно знайти?
4. Порядок розв'язання задачі Марковіца.
5. Структура оптимального управління. Чим власне ми управляємо?
6. Які програми використані в цьому документі? Назвіть їх "входи" і "виходи".
7. Зробіть специфікацію програми моделювання.
8. В яких ситуаціях вигідно прибутки одного виробництва вкладати в розширення іншого виробництва?
9. Чи виправдана в сучасній економіці орієнтація на отримання максимального прибутку за визначений термін?
10. В чому загальна причина того, що пропорція розподілу доходу в оптимальній стратегії не є неперервна в часі, а релейна - спочатку все в накопичення, потім - все в розвиток?



5.3. Узагальнення задачі Марковіца для випадку нелінійних інвестиційних функцій

Вступ

В цьому розділі продовжуємо розгляд задачі Марковіца - вже для випадків, коли **функції віддачі інвестицій - нелінійні, випуклі**. Лінійність чи нелінійність, випуклість чи невивуклість відповідних функцій варіаційної задачі радикально змінює структуру оптимальних розв'язань, може суттєво збільшити трудомісткість розв'язання. В попередньому розділі знайдено оптимальне управління для випадку лінійних функцій інвестицій. Воно виявилось релейним: спочатку все в інвестиції, потім - все в накопичення.

Для варіаційних задач, як відомо, аналітичні рішення існують для небагатьох окремих випадків. Для керівника, менеджера **математичні проблеми** оптимального управління стоять десь на другому плані - їм потрібен універсальний, зручний, надійний **інструмент для знаходження і дослідження оптимальних стратегій** управління. Цей розділ є черговим кроком в узагальненні задачі розподілу ресурсів між інвестиціями у різні виробництва та накопиченням. Щоб зробити документ досить автономним, ми повторюємо постановку базової задачі Марковіца, а потім узагальнюємо та деталізуємо задачу.

Постановка задачі

Розглядаємо ідеалізовану задачу з математичної економіки. Маємо економічну систему, в якій виробляються N видів продукції, що випускаються з темпами $x_1(t)$, $x_2(t)$, $x_3(t)$,... $x_1(t)$, $x_N(t)$ (одиниць вимірювання продукції за місяць, квартал, рік). **Темп виробництва** - це синонім терміна **виробнича потужність**.

Залежності приростів темпів випуску від інвестицій у розширення виробництва тепер будемо вважати нелійними - випуклими (вгору) для усіх видів виробництв. Для порівняння подаємо постановку лінійної задачі з попереднього розділу.

$$\text{Рівняння динаміки виробничих потужностей} \quad \frac{d}{dt}x(t)_i = \sum_{j=1}^N a_{i,j} \cdot y(t)_j, \text{ де}$$

$x(t)_i$ - поточна виробнича потужність по i -му продукту;

$y(t)_j$ - поточні інвестиції j -го продукту в розширення виробництва;

$a_{i,j} \geq 0$ - параметр рівняння динаміки виробничих потужностей;

$x(0)_j = C_j$ - початковий рівень виробничих потужностей по j -му продукту.

Обмеження рівня інвестицій задачі: $y(t)_j \leq x(t)_j$;

А тепер подамо постановку задачі цього розділу.

Рівняння динаміки виробничих потужностей

$$\frac{d}{dt}x(t)_i = \text{fin}(y(t)_i, i) = \text{fin}(x(t) \cdot u_i, i), \text{ де}$$

$\text{fin}(y(t)_i, i)$ - **функція віддачі інвестицій** для i -го виробництва, що належить до класу монотонно зростаючих вигнутих (вгору) функцій;

$$x_s(t) = \sum_{j=1}^N x(t)_j ; - \text{сумарне виробництво (в грошових одиницях) в момент } t;$$

$0 \leq u(t)_i \leq 1$ - управління, за змістом, це доля сумарних поточних ресурсів, що виділяється в поточний момент для розширення виробничих потужностей по i -му продукту. Для управлінь виконується умова нормування:

$$\sum_{j=1}^N u(t)_j + u_{ak}(t) = 1, \text{ де } u_{ak}(t) - \text{частка ресурсів, що йде в накопичення.}$$

Зауваження. В літературі поки існує розбіжність у термінології. Ми беремо те визначення, що відповідає стандартам математичних публікацій – "випуклий". Трмін "опуклий" є місечковим, та інформативно невдалим ("опуклий" – "опухлий"?). Дайте самостійно математичне визначення випуклості функції.

Потрібно визначити оптимальну стратегію інвестицій (не плутати з поточними витратами на виробництво видів продукції), що максимізує сумарний прибуток за

певний період: $JN = \int_0^T x_s(t) \cdot u_{ak}(t) dt$

Щоб краще бачити, чим ми власне управляємо, відобразимо постановку задачі у схемі виробничої системи (рис.1).

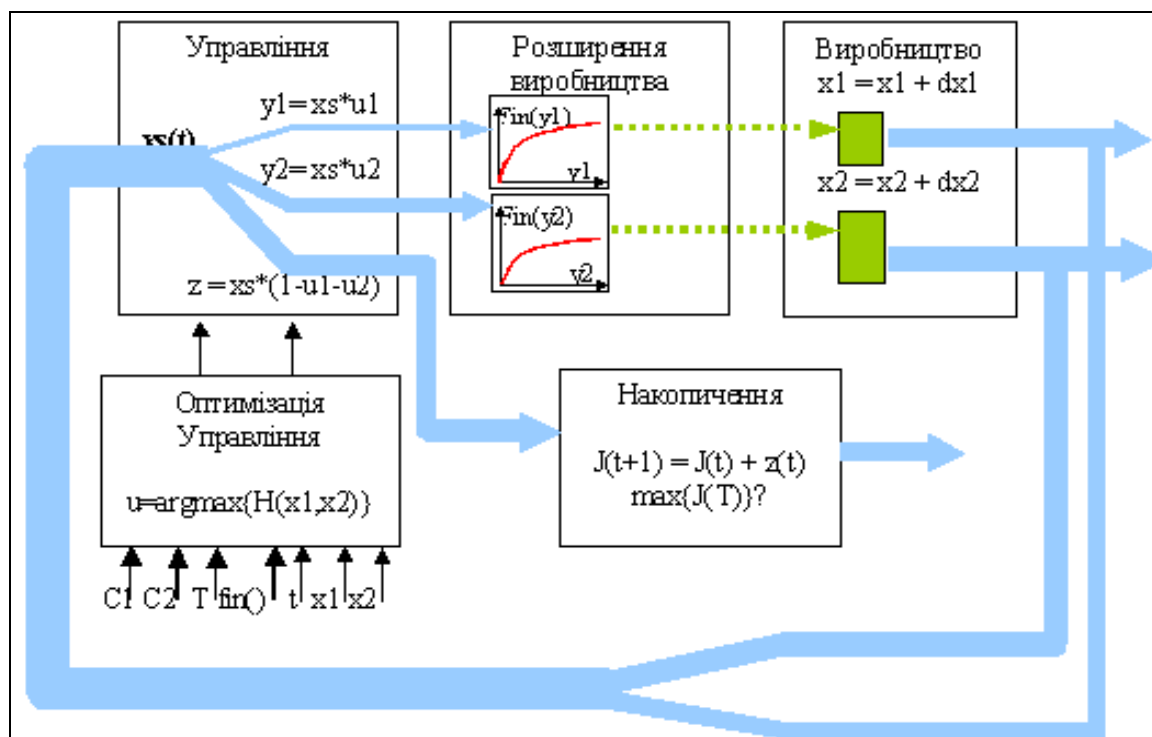


Рис.1. Схема функціонування виробничої системи

Очевидно (рис.1), у двовимірній задачі ми шукаємо **дві функції часу** $u(t)_1$ $u(t)_2$ такі, що дають максимум функціоналу і задовольняють усі обмеження.

$$J2 = \int_0^T (x(t)_1 + x(t)_2) \cdot u_{ak}(t) dt = \int_0^T (x(t)_1 + x(t)_2) \cdot (1 - u(t)_1 - u(t)_2) dt$$

Прокоментуємо зміни, що ми внесли до класичної задачі Марковіца. Чи не є вони штучними? Чи маємо ми право усі ресурси вимірювати в грошових одиницях і вважати необмежено конвертованими в гроші і навпаки (ліквідними). В періоди відносно стабільного стану виробництва і економіки в цілому, коли продукція має стабільний попит за стабільними цінами, припущення про конвертованість продуктів виробництва коректне. В сучасній глобалізованій і високотехнологічній економіці *майже* завжди стандартний товар перетворюється в гроші, а за гроші *практично* завжди можна отримати потрібні для розширення виробництва ресурси та послуги. Дефіцити, державні обмеження на торгівлю у світі в цілому зменшуються.

Ми припускаємо, що при збільшенні інвестицій приріст виробничих потужностей зменшується. Чи так це? Відомо, що у "новій", високотехнологічній, глобалізованій, мережевій економіці, навпаки, виникають ефекти позитивних зворотних зв'язків, експоненційного зростання випуску та експоненційного зменшення витрат.

Перша можлива причина зменшення ефекту інвестицій - перевантаження організацій, що конкретно виконують розширення виробництва - проєктують, будують, монтують та відлагоджують обладнання, навчають новий персонал.

Друга причина - це десятки наслідків насичення ринку та конкуренції. Ми вже зробили припущення: весь випущений продукт стабільно реалізується. Якою ціною забезпечується ця стабільність? Ну збільшуємо ми випуск певного харчового чи програмного продукту на 13%. Так одночасно треба інвестувати в розширення реклами, розбудову нових ринків збуту, підвищення якості та ін., - інакше ці 13% залишаться на складі.

Таким чином, якщо інвестиційний процес розглядати системно, з урахуванням особливостей зовнішнього середовища для нашої виробничої системи, то виявиться, що існують ситуації, коли приріст виробничих потужностей може ставати все дорожчим з їх зростанням.

Залишається питання - чи не є така модель виробництва фундаментом песимізму - рано чи пізно, розширення виробництва стає неефективним. Що робити? - Шукати нові можливості - нові продукти, нові технології, нові ринки. Задача Марковіца саме орієнтована на таку стратегію: оцінити термін T , коли дане виробництво вичерпає свій потенціал росту, і до моменту T взяти максимум накопичень від нього - для інвестицій у щось зовсім інше - з безмежним ринком без конкурентів. В десятому, підсумковому законі нової економіки з книги С. Келлі "Нові правила для нової економіки" [47] сказано "не займайся оптимізацією, лови можливості". Будемо пам'ятати це при розв'язанні нашої оптимізаційної задачі.

Розв'язання узагальненої оптимізаційної задачі

Розглядаємо, як і в попередньому розділі двовимірну задачу. **Згадаємо** схему розв'язання лінійної двовимірної задачі:

1) записали розв'язання системи лінійних диференціальних рівнянь у загальному вигляді

$$\frac{d}{dt}x(t)_i = a_{1,1} \cdot y(t)_1; \quad dx_1 = a_{1,1} \cdot y(t)_1 \cdot dt; \quad x(t)_1 = c_1 + a_{1,1} \cdot \int_0^t y(t)_1 dt;$$

2) підставили це розв'язання у вираз для критерію

$$J_1 = \int_0^T (x(t)_1 - y(t)_1) dt; \quad J_1 = \int_0^T \left(c_1 + a_{1,1} \cdot \int_0^t y(t_1)_1 dt_1 - y(t)_1 \right) dt;$$

3) в отриманому виразі змінили порядок інтегрування

$$J1 = c_1 \cdot T + \int_0^T [a_{1,1} \cdot (T - t) - 1] \cdot y(t)_1 dt.$$

Внутрішній інтеграл внаслідок лінійності задачі отримали в кінцевому вигляді. Тому в останньому виразі під інтегралом маємо добуток функції часу і функції управління. Функція часу монотонно зменшується. Звідси витікає стратегія управління - поки функція часу додатна, управління (рівень інвестицій) вибрати максимально можливим додатним, коли функція часу стане від'ємною - взяти управління нульовим.

Задача зводиться до визначення моментів переключення управлінь по кожному з виробництв. А чи можна управління зробити від'ємним? - Можна! Це будуть інвестиції у згорання виробництва - продаж фондів, переведення персоналу в інші підрозділи та ін. Це окрема задача.

Отримати за описаною методикою розв'язання задачі з довільними нелінійностями практично неможливо. Можна спробувати знайти аналітичне розв'язання для конкретного виду функцій інвестицій - степеневих, логарифмічних, експоненційних. Але імовірність успіху дуже мала.

Будемо шукати розв'язання з позицій менеджера, що трохи озброєний знаннями про методи розв'язання варіаційних задач та навичками моделювання процесів. Почнемо з розгляду лінійної одновимірної задачі. Ми знаємо, що при виконанні певної умови: $T > \frac{1}{a_{11}}$, початкове (оптимальне) управління буде $u = 1$. В якийсь певний момент

воно змінюється на $u = 0$. Чи можна з поточних змінних процесу сформулювати умову (індикатор) переключення. Якщо умова така існує, то задача менеджера (чи автоматизованої системи) буде: після старту процесу (виробництва продукції та розширення виробничих потужностей) періодично вимірювати поточний стан процесу та перевіряти умову переключення. При виконанні її - переключитись на стратегію "все в накопичення".

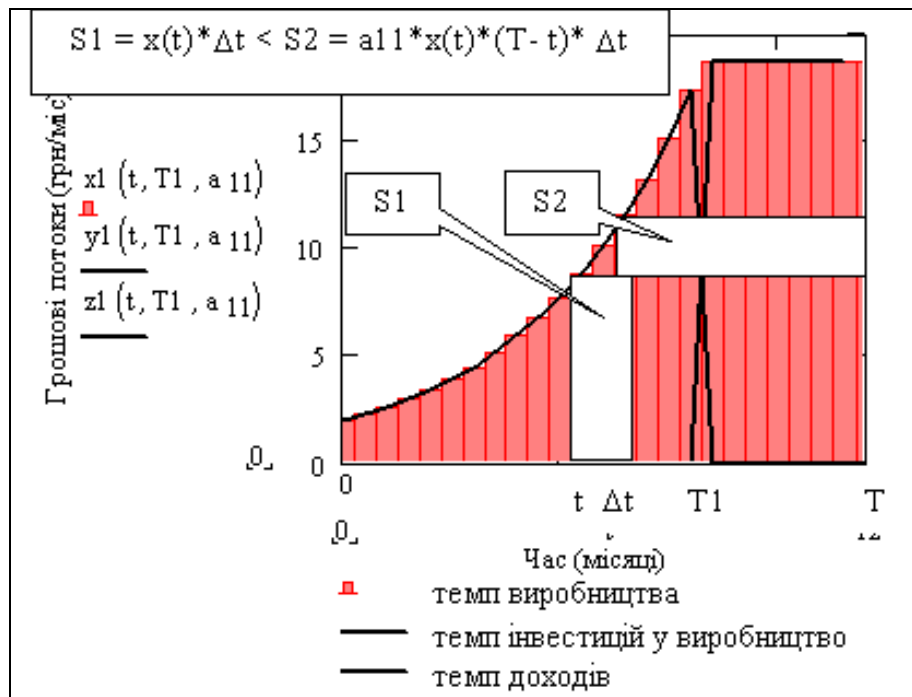


Рис.2. До виведення умови переключення управління

Визначимо цю умову. На рис. 2 подано графіки зміни виробничих потужностей, інвестицій та накопичень. Спробуємо для довільної поточної точки визначити, що вигідніше - чергову (виручку) продукцію $x(t) \cdot \Delta t$ за малий період часу Δt направити в накопичення чи інвестувати у розширення виробництва на $\Delta x = a_{11} \cdot x(t) \cdot \Delta t$ (виведено з: $dx/dt = a_{11} \cdot x$). Визначимо приріст критерію для цих альтернатив.

Бачимо, що для альтернативи "все в накопичення", приріст критерію буде:

$\Delta J_{\text{нак}} = x(t) \cdot \Delta t$; для альтернативи "все в інвестиції" приріст критерію з наступного кроку і до кінця процесу буде: $\Delta J_{\text{інв}} = a_{11} \cdot x(t) \cdot (T - t) \cdot \Delta t$. Тепер можемо записати умову переключення: $\boxed{\text{Умова} = \Delta J_{\text{нак}} - \Delta J_{\text{інв}}}$.

Узагальнимо цю умову для випадків нелінійних випуклих функцій інвестицій. Зробимо таке *інтелектуальне зусилля*: замість вибору "або тільки інвестування, або тільки накопичення", будемо вибирати певну **оптимальну пропорцію ("золоту"?)** інвестування та накопичення, що дає максимальний приріст значення критерію. Будемо вибирати управління $u(t)$ так, щоб вираз $H(x, u) = (\Delta J_{\text{нак}} + \Delta J_{\text{інв}})$ був максимальним, тобто:

$$\boxed{H(x, u) = x(t) \cdot (1 - u(t)) + f_{\text{ін}}(x(t) \cdot u(t)) \cdot (T - t)} \rightarrow \max_u \quad (1)$$

Далі виникають дві проблеми.

Перша, маленька - чи буде знайдене розв'язання оптимальним у випадку лінійних функцій інвестицій? Буде. Це просто довести. Функція $H(x, u)$ буде лінійною по u , що обмежене діапазоном $0 \leq u(t) \leq 1$. Максимальне значення (= точна верхня границя) буде або при $u = 0$, або $u = 1$. В момент переключення буде $H(x, 0) = H(x, 1)$.

Неважко довести також, що для випадку, коли $f_{\text{ін}}(y)$ - випукла (convex) функція, то в інтервалі $(0, T)$ може існувати інтервал (t_a, t_b) , де оптимальне управління буде в середині допустимого інтервалу: $0 < u(t) < 1$. Якщо функція $f_{\text{ін}}(y)$ має неперервну першу похідну, то управління $u(t)$ на інтервалі $t_a \leq t \leq t_b$ буде неперервним.

Друга, велика (проблема) - строге математичне доведення оптимальності управління, отриманого запропонованим методом для **довільних** функцій інвестицій $f_{\text{ін}}(y)$. Це окрема задача, що виходить за межі розділу. Скажемо тільки, що таке управління є досить близьким до оптимального. – воно дає сумарний прибуток на 2-5% менший. Однак, вхідні дані для задачі розвитку виробництва мають велику невизначеність. Залишаємо це доведення для самостійного виконання. Для підказки, назовемо терміни і розділи вищої математики, що мають пряме відношення до нашої задачі: канонічна форма варіаційної задачі Лагранжа, функція Гамільтона, динамічне програмування, принцип максимуму.

Розробка робочої програми для розв'язання узагальненої задачі Марковіца

Для розробки програми з попередніх п'яти сторінок нам власне потрібна формула (1). Усе інше можна було б і не читати? Рутинна робота - отримати аналогічні вирази для випадків двовимірної, ... N-вимірної задачі, при умові, що ці п'ять сторінок дійсно засвоєні.

Програма, що подана далі, з одного боку - інструмент знаходження оптимального управління, з іншого - засіб для імітаційного моделювання організаційно-виробничої системи. Ми не тільки обчислюємо оптимальне управління, але й моделюємо процес функціонування виробничої системи - розширення виробництва, виробництво. Програма відповідає блок-схемі на рис.1. При бажанні можна деталізувати, ускладнювати її функціональні модулі. **Напишіть коментарі до програми - це тест.**

Вводимо значення параметрів: період, число кроків, крок квантування.

$T := 12 \quad Kh := 60 \quad \Delta t := T \div Kh \quad \text{ORIGIN} := 1; \quad i := 1..N \quad j := 1..N$

$\left(\begin{array}{c} \text{MF} \\ \text{SP} \end{array} \right) :=$	$\Delta t \leftarrow T \div Kh$ $x1_1 \leftarrow C1$ $x2_1 \leftarrow C2$ $u1op \leftarrow 0.5$ $u2op \leftarrow 0.5$ $sdox \leftarrow 0$ $setka \leftarrow 80$ $\Delta u \leftarrow 1 \div setka$ for $k \in 1..Kh + 1$ <table> <tr> <td>$t \leftarrow \Delta t \cdot (k - 1)$</td><td></td></tr> <tr> <td>$maHm \leftarrow 0$</td><td></td></tr> <tr> <td>$xs \leftarrow x1_k + x2_k$</td><td></td></tr> <tr> <td>for $i \in 1..setka$</td><td></td></tr> <tr> <td>$u1 \leftarrow (i - 1) \cdot \Delta u$</td><td></td></tr> <tr> <td>for $j \in 1..(setka - i + 1)$</td><td></td></tr> <tr> <td>$u2 \leftarrow (j - 1) \Delta u$</td><td></td></tr> <tr> <td>$unak \leftarrow 1 - u1 - u2$</td><td></td></tr> <tr> <td>$potm \leftarrow (\text{fin}(u1 \cdot xs, \alpha 1, A1) + \text{fin}(u2 \cdot xs, \alpha 2, A2))$</td><td></td></tr> <tr> <td>$zaraz \leftarrow unak \cdot xs$</td><td></td></tr> <tr> <td>$fHm \leftarrow zaraz + potm \cdot (T - t)$</td><td></td></tr> <tr> <td>$ymo \leftarrow maHm < fHm$</td><td></td></tr> <tr> <td>$maHm \leftarrow maHm \cdot (1 - ymo) + fHm \cdot ymo$</td><td></td></tr> <tr> <td>$u1op \leftarrow u1op \cdot (1 - ymo) + u1 \cdot ymo$</td><td></td></tr> <tr> <td>$u2op \leftarrow u2op \cdot (1 - ymo) + u2 \cdot ymo$</td><td></td></tr> </table> $u1_k \leftarrow u1op$ $u2_k \leftarrow u2op$ $x1_{k+1} \leftarrow x1_k + \text{fin}(u1_k \cdot xs, \alpha 1, A1) \cdot \Delta t$ $x2_{k+1} \leftarrow x2_k + \text{fin}(u2_k \cdot xs, \alpha 2, A2) \cdot \Delta t$ $z_k \leftarrow xs \cdot (1 - u1_k - u2_k)$ $xs \leftarrow x1_{k+1} + x2_{k+1}$ $sdox \leftarrow sdox + z_k$	$t \leftarrow \Delta t \cdot (k - 1)$		$maHm \leftarrow 0$		$xs \leftarrow x1_k + x2_k$		for $i \in 1..setka$		$u1 \leftarrow (i - 1) \cdot \Delta u$		for $j \in 1..(setka - i + 1)$		$u2 \leftarrow (j - 1) \Delta u$		$unak \leftarrow 1 - u1 - u2$		$potm \leftarrow (\text{fin}(u1 \cdot xs, \alpha 1, A1) + \text{fin}(u2 \cdot xs, \alpha 2, A2))$		$zaraz \leftarrow unak \cdot xs$		$fHm \leftarrow zaraz + potm \cdot (T - t)$		$ymo \leftarrow maHm < fHm$		$maHm \leftarrow maHm \cdot (1 - ymo) + fHm \cdot ymo$		$u1op \leftarrow u1op \cdot (1 - ymo) + u1 \cdot ymo$		$u2op \leftarrow u2op \cdot (1 - ymo) + u2 \cdot ymo$	
$t \leftarrow \Delta t \cdot (k - 1)$																															
$maHm \leftarrow 0$																															
$xs \leftarrow x1_k + x2_k$																															
for $i \in 1..setka$																															
$u1 \leftarrow (i - 1) \cdot \Delta u$																															
for $j \in 1..(setka - i + 1)$																															
$u2 \leftarrow (j - 1) \Delta u$																															
$unak \leftarrow 1 - u1 - u2$																															
$potm \leftarrow (\text{fin}(u1 \cdot xs, \alpha 1, A1) + \text{fin}(u2 \cdot xs, \alpha 2, A2))$																															
$zaraz \leftarrow unak \cdot xs$																															
$fHm \leftarrow zaraz + potm \cdot (T - t)$																															
$ymo \leftarrow maHm < fHm$																															
$maHm \leftarrow maHm \cdot (1 - ymo) + fHm \cdot ymo$																															
$u1op \leftarrow u1op \cdot (1 - ymo) + u1 \cdot ymo$																															
$u2op \leftarrow u2op \cdot (1 - ymo) + u2 \cdot ymo$																															

$$v_{yx}^{(k)} \leftarrow \begin{pmatrix} u1_k \\ u2_k \\ 1 - u1_k - u2_k \\ x1_k \\ x2_k \\ z_k \end{pmatrix}$$

$$V_{yx} \leftarrow \begin{pmatrix} v_{yx} \\ sdox \end{pmatrix}$$

В програмі обчислюються поточні виробничі потужності $x1(t)$, $x2(t)$, інвестиції $y1(t)$, $y2(t)$, "прибуток" $z(t)$, сумарний прибуток $Sdox(t)$. Робимо стенд - екранну сторінку, де зібрані усі входи і виходи виробничої системи (рис.3).

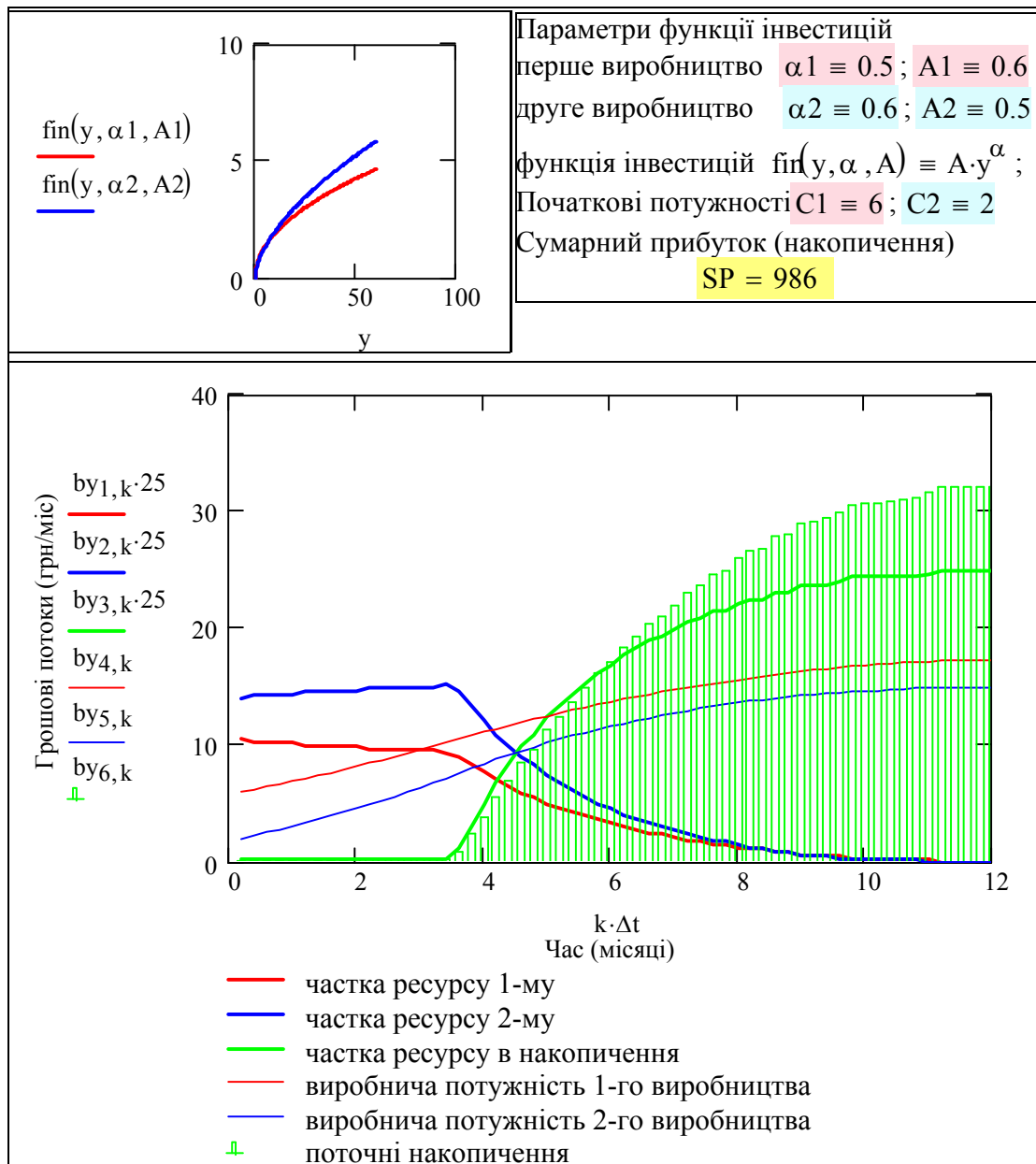


Рис.3. Оптимальний процес розвитку двопродуктового виробництва.

Подаємо копії екранів - результати оптимізації для випадків а) обидва виробництва лінійні, б) одне з виробництв - лінійне. Проаналізуйте, те що бачите на рис.3, 4.

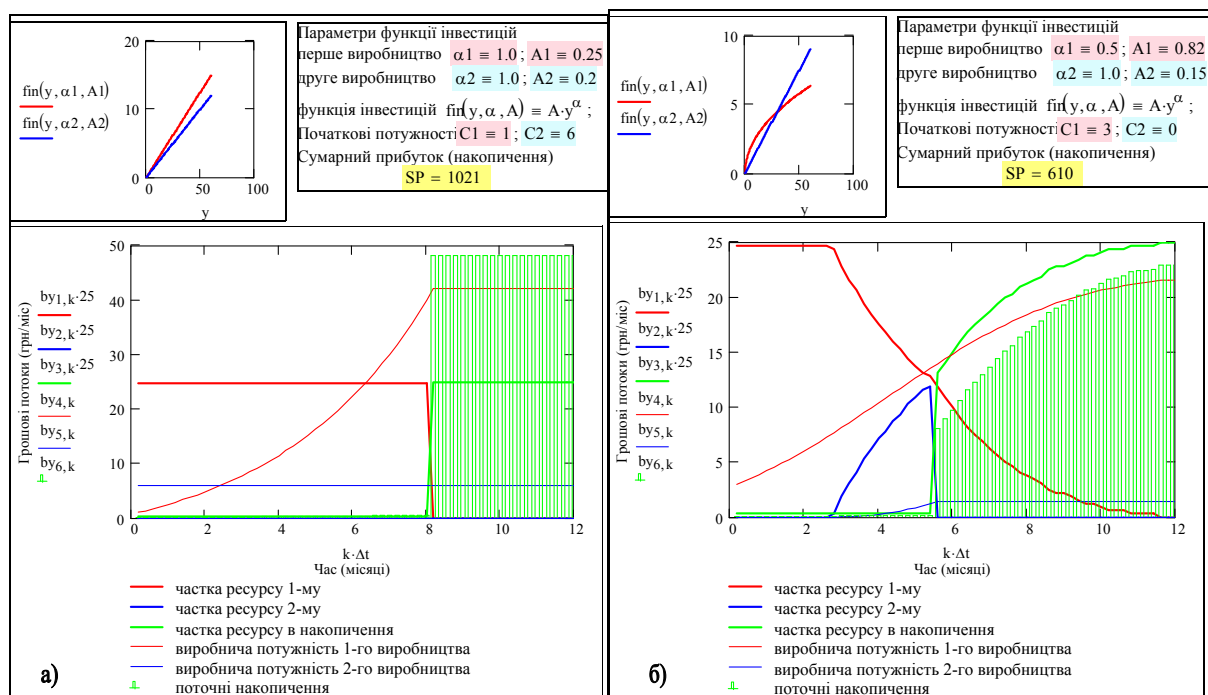


Рис.4. Оптимальні процеси розвитку для випадку лінійних функцій інвестицій

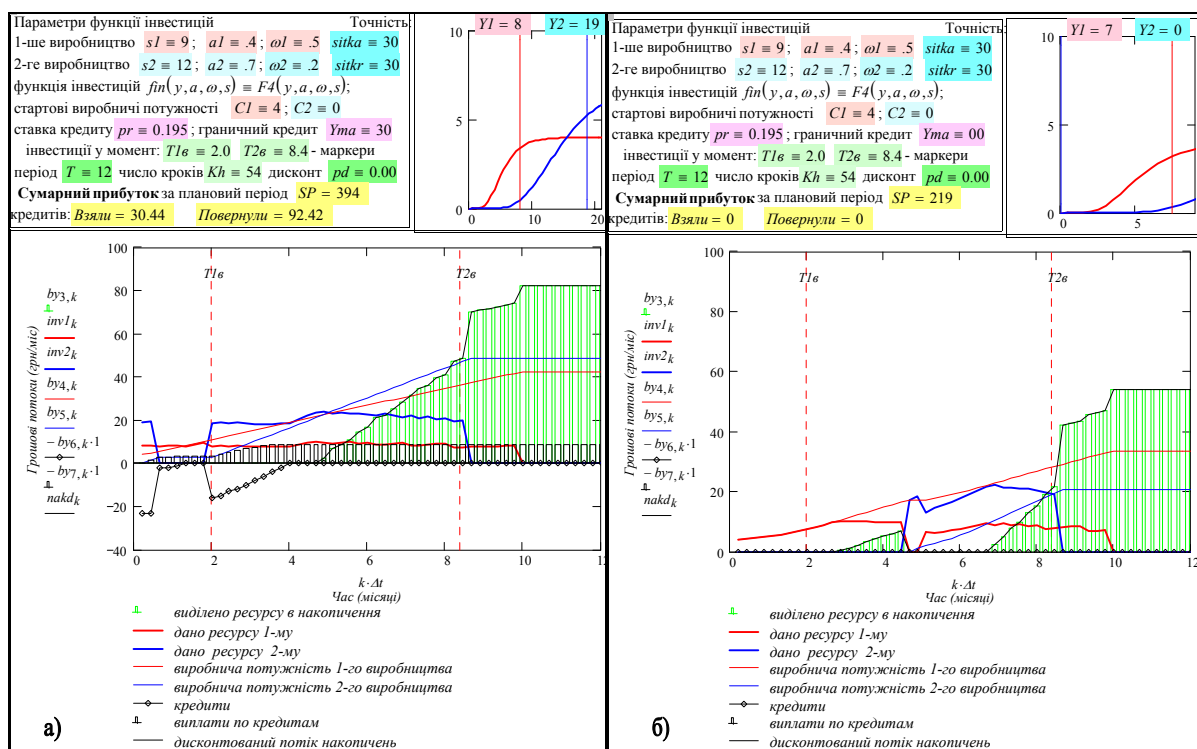


Рис.5. Порівняння двох стратегій. Оцінка ефективності кредитів

На базі даного розділу розроблено багато систем оптимізації і моделювання для більш складних і реалістичних задач [4-12, 22, 23, 39-43]. На рис.5 подано приклад такої задачі - оптимізація сумарного прибутку при довільних функціях інвестицій і кредитах. На рис.5а бачимо процес розвитку при мінімальному стартовому рівні і без кредитів. Прибуток 701у.о. На рис.5б - випадок використання кредитів під 10%. Результат - прибуток 2583у.о. і на додаток - прості управління і ріст системи - майже лінійний, без стрибків.

Експериментальна перевірка оптимальності управління

Згадаємо, що у варіаційній задачі шукають не числа, а функції, що дають екстремум функціоналу. Якщо знайдені нами управління - оптимальні, то при будь-яких варіаціях функцій управління значення функціоналу не покращується (в нашому випадку - не збільшується).

Дослідимо зміни функціоналу - сумарного прибутку при варіаціях управлінь. Варіації управлінь - функції певних класів, що задовольняють граничні умови і обмеження для управління. Вигляд варіацій будемо вибирати виходячи з практичних аспектів реалізації управління. Перевірка на оптимальність одночасно є і аналізом чутливості оптимального управління. Розглянемо два приклади варіацій.

Варіація рівня інвестицій. Інтерпретація: дивимось, що буде якщо одному виробництву трохи збільшити інвестиції, другому настільки ж зменшити. Множники у дужках - це урахування обмежень на управління. Робимо таку функцію:

$$up1v(r\delta, u1o, u2o) := u1o + r\delta \cdot (0 \leq u1o + r\delta \leq 1) \cdot (u2o > r\delta),$$

де $r\delta$ - розмір варіації управління. Беремо номінальне управління розраховане вище і формуємо функції управління з варіаціями, будуємо їх графіки.

$$bv1,k := up1v(0.05, by1,k, by2,k) ; \quad bv2,k := 1 - bv1,k - by3,k$$

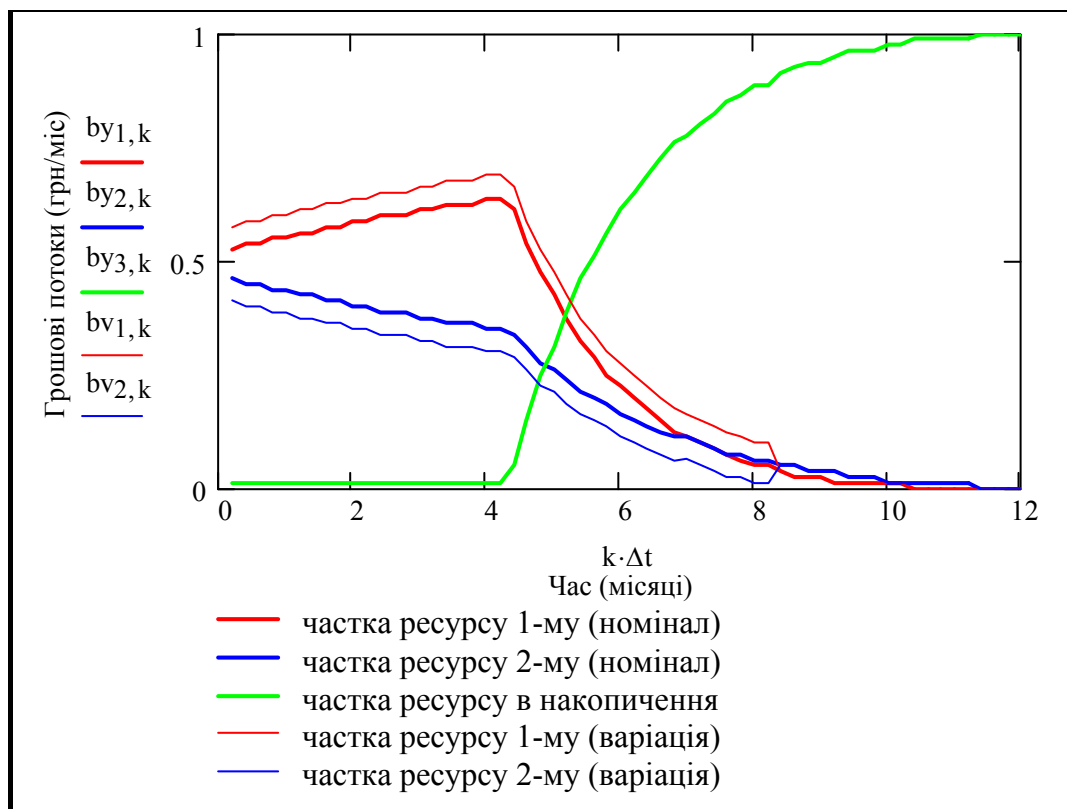


Рис.6. Варіація управління. Перерозподіл рівнів інвестицій у виробництва

Варіація моменту переключення. Інтерпретація: дивимось, що буде якщо момент переключення змінити на величину δT . Множник у дужках - обмеження управління.

$$uo2v(t, \delta T, Tno, u2o) := u2o \cdot (t \leq Tno + \delta T);$$

$$bv_{2,k} := uo2v(k \cdot \Delta t, -3 \cdot \Delta t, 18 \cdot \Delta t, by_{2,k}) ; \quad bv_{3,k} := 1 - bv_{2,k} - by_{1,k} .$$

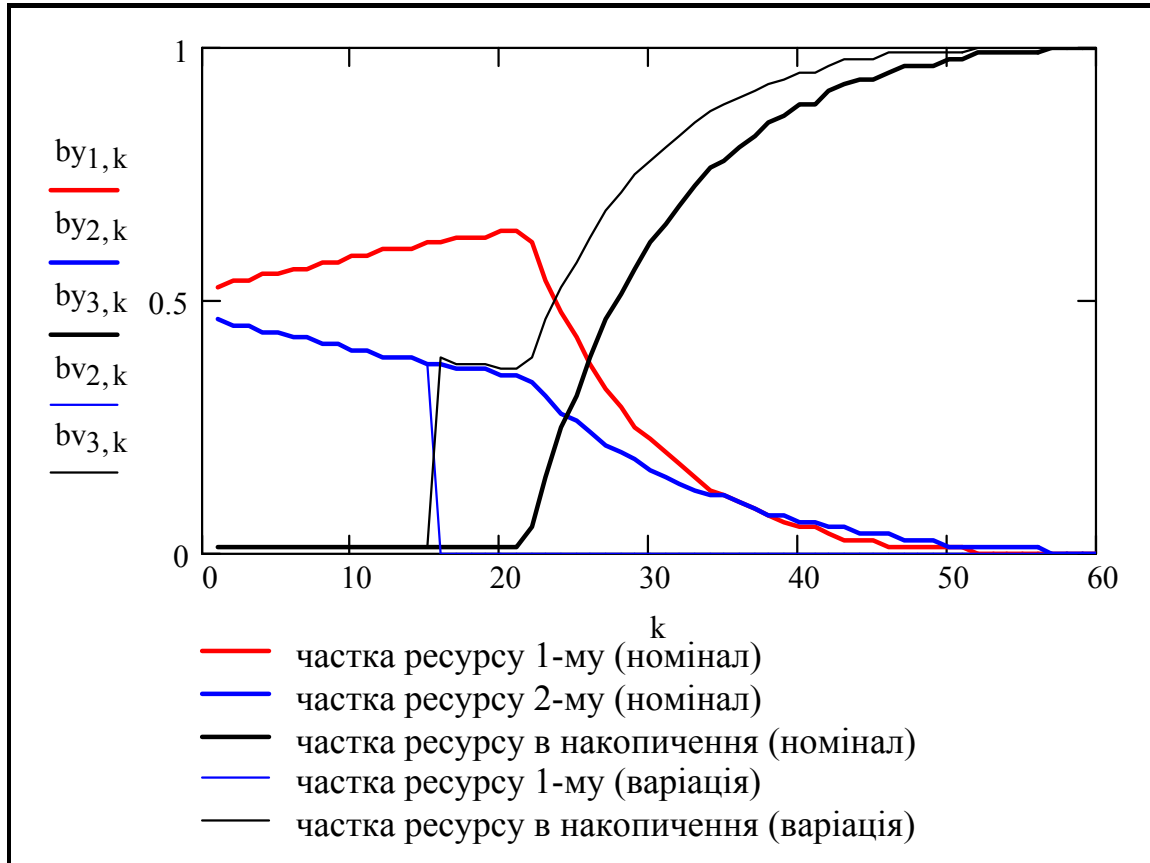


Рис.7. Варіація управління. Зміна моменту переключення

Залишається чисто технічна проблема - зробити ефективну програму генерації варіацій управління та дослідження їх впливу. Далі подано приклад такої програми (вона сколапсована без пароля), в електронній версії посібника її можна відкрити.

Як видно, для випадку варіації класу "перерозподіл ресурсу між 1-им та 2-им виробництвом", знайдене управління є оптимальним - воно дає максимальний прибуток. Але це не доведення. Спробуйте знайти аналітико-експериментальне доведення оптимальності (або квазіоптимальності) управління, що знайдене за нашим методом.

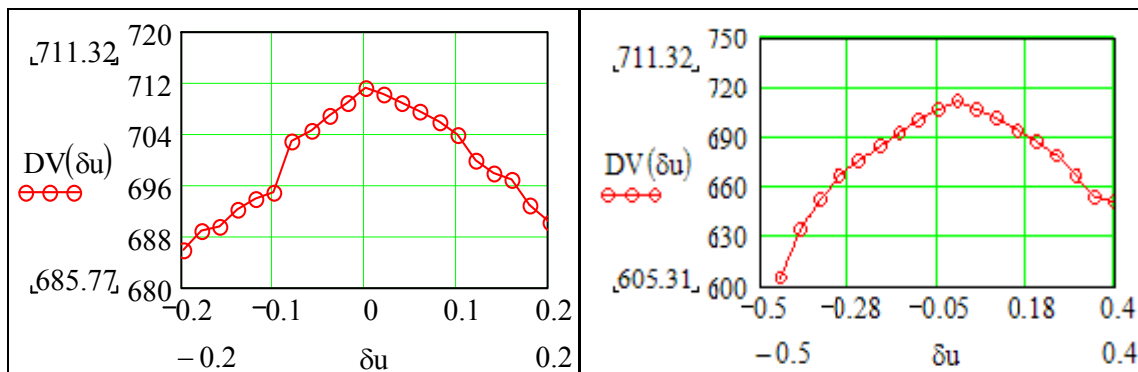


Рис.8. Дослідження впливу варіації управління (для 10% та 20% діапазонів)

Бачимо, що при зміні номінального управління в будь-якому напрямку значення критерію погіршується. Однак, це тільки "передостання істина".

Висновки

Конкретні висновки зробіть самостійно. Подамо тільки приклад того, що побачить у цій задачі менеджер, природні здібності якого підшліфовані конструктивним вивченням математики.

Варіаційна задача еквівалентна задачі пошуку екстремуму функції нескінченного числа змінних. Можна задати функцію оптимального управління тільки в N точках. Тоді задача знаходження екстремуму інтегрального функціоналу (в нашому випадку - максимуму сумарного доходу) зведеться до задачі знаходження екстремуму функції N змінних. Ми розбили процес на N кроків у часі. На кожному кроці ми знаходили оптимальне управління із умови екстремуму деякої функції $H(x,u)$. Тобто, задачу одноразового знаходження екстремуму функції N змінних ми замінили N -разовим знаходженням функції однієї змінної - u . Як відомо ще з до біблійських часів, зробити це набагато легше.

Контрольні запитання

1. Постановка задачі Марковіца - критерій, обмеження, змінні управління, граничні умови.
2. Чим відрізняється постановка оптимізаційної задачі даного розділу від класичної задачі Марковіца?
3. Дайте визначення функції, функціоналу оператора. Знайдіть в тексті документу приклади функції _____, функціоналу _____, оператора _____.
4. Дайте формулювання варіаційної задачі.
5. Порядок розв'язання класичної задачі Марковіца.
6. Порядок розв'язання модифікованої задачі Марковіца з нелінійними функціями інвестицій.
7. Назвіть "входи" і "виходи" програми моделювання та оптимізації процесу розвитку виробництва.
8. Дайте визначення випуклості та невивуклості функції.
9. Можливі причини, що обумовлюють випуклість функції інвестицій.
10. Чи є виправданою в економіці нетрадиційна орієнтація на отримання максимального прибутку за визначений термін?
11. Головне практичне питання. Яку структуру має оптимальне управління для двопродуктової виробничої системи (неперервність, монотонне збільшення/зменшення, наявність і число релейних переключень) для випадків:
а) обидві функції інвестицій лінійні, б) обидві функції інвестицій нелінійні, вигнуті, в) одна функція інвестицій лінійна.

Завдання для самостійного виконання

1. Спробуйте розв'язати задачу для випадку, коли бюджетне обмеження $y_1(t) \leq x_1(t)$ є м'яким (нечітким). - 55 балів
2. Розробіть більш ефективний алгоритм знаходження максимуму. - 25 балів
3. Модифікуйте стенд так, щоб виводились одночасно три процеси:
- номінальний і для $+\Delta p$, $-\Delta p$ - заданого розкиду деякого параметра. - 25 балів
4. Модифікуйте задачу для урахування: витрат на виробництво і ціни продажу.
Виконайте дослідження залежності сумарного прибутку від терміну T . - 150 балів
5. Розробіть програму для розв'язання тривимірної задачі Марковіца. - 30 балів



Задачі оптимального управління динамічними системами при обмеженнях

У цьому розділі розглядаються:

- ✓ Метод принципу максимуму Понтрягіна. Оптимальне управління інерційним об'єктом.
- ✓ Метод принципу максимуму Понтрягіна. Оптимальне управління об'єктом з демпфуванням.
- ✓ Метод принципу максимуму Понтрягіна. Оптимальне управління об'єктом-маятником.

Після вивчення цього розділу ви **зможете:**

1. Знаходити в своїй професійній діяльності задачі оптимального за швидкодією управління при обмеженнях і виконувати їх формалізацію - запис у стандартній формі – критерій, обмеження, граничні умови.
2. Для типових варіаційних задач записувати рівняння для спряжених функцій, функції Гамільтона, та визначати умову її максимуму в аналітичному або числовому вигляді.
3. Визначати оптимальне управління в координатній формі, отримувати рівняння для ліній переключення.
4. Будувати перехідні процеси у фазовому просторі.
5. Розробляти програми моделювання процесів в оптимальних за швидкодією динамічних системах.
6. Виконувати дослідження на моделях оптимальних динамічних систем.

7. Дати визначення термінам:

- | | |
|---|---|
| 1) критерій, обмеження, граничні умови, ціль оптимізації; | 5) принцип максимуму; |
| 2) форма Коші запису рівняння динаміки об'єкта; | 6) функція переключення, лінія переключення; |
| 3) спряжена система рівнянь, спряжені функції; | 7) фазова площина (простір), фазова траєкторія; |
| 4) функція Гамільтона, гамільтоніан; | 8) перехідний процес. |

6.1. Метод принципу максимуму Понтрягіна. Оптимальне управління інерційним об'єктом

Вступ

Що є найбільш характерним для задач управління об'єктами будь-якої природи? Кожний одразу і без вагань здогадається: обмеження на ресурси управління. В системі управління літака - це обмеження потужності двигуна, кутів відхилення аеродинамічних органів управління, в техніко-економічній системі - це обмеження виробничих потужностей, фінансових ресурсів, кадрів.

Який критерій оптимізації є найбільш характерним для задач управління. Знову ж таки, кожен скаже: швидкодія (зрозуміло, при обмеженнях на економічність, безпечність...). Особливо у бізнесі: "другий повинен бути в десять разів кращий за першого", що в перекладі означає - запізнення з виходом на ринок з певним товаром суттєво знижує, або взагалі елімінує доходи.

А тепер спробуйте, спираючись на власні знання, досвід та інтуїцію здогадатися, яку структуру має оптимальне за швидкодією управління при обмеженнях на управління? Або простіше, - сформулюйте принцип такого управління? Інженери, практики винайшли і застосовували цей принцип ще до того, як була створена математична теорія оптимального управління. Це так званий принцип "максимальний розгін - максимальне гальмування". Суть його в тому, що оптимальне за швидкодією управління повинно бути в кожен момент часу максимально можливим.

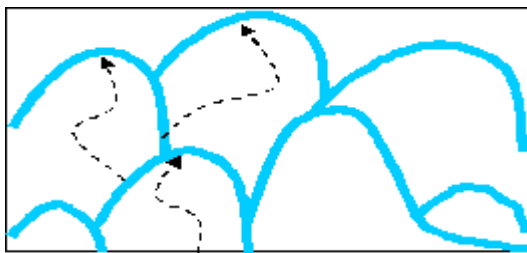
Вибирається тільки напрямок дії. Функція оптимального управління $U_{opt}(t)$ звичайно має розриви - "переключення". Для таких задач важко застосувати методи класичного варіаційного обчислення - вони (методи, а не задачі) вимагають існування відповідних похідних. Саме для таких задач колективом вчених під керівництвом академіка Л. Понтрягіна на початку 1950-их років був розроблений метод під назвою **принцип максимуму**.

А тепер, враховуючи, що:

а) практики і без теоретиків винайшли структуру оптимального за швидкодією управління;

б) в сучасному світі глобалізованої, високотехнологічної, пов'язаної мережами економіки головним правилом є "не займайся оптимізацією, лови можливості"! (бестселер: Стівен Келлі "Нові правила для нової економіки. Десять радикальних стратегій для конектизованого світу"[48]).

Дайте відповідь на питання - чи комусь, крім сотні математиків, потрібно детально вивчати досить складні методи, результати яких очевидні, іноді? Виявляється – так, – професія математика входить до першої десятки престижних професій поряд з професіями "дизайнерів софту". Сьогодні дійсно треба ловити, вирощувати (памперсувати, по-їхньому) можливості. Але, як легко втратити ці можливості, якщо ретельно їх не дослідити і оптимізувати.



Ось фрагмент книги Келлі (рисунок і коментар).

Turbulent times mean that local success is not global success. A company may be at peak efficiency? but on the wrong mountain. The trick is to select a high-potential area to excel in.

Метод принципу максимуму досить широко використовується для розв'язання задач оптимального управління при наявності обмежень типу нерівності. Ціль даного розділу - **освоєння техніки постановки і розв'язання оптимізаційних задач методом принципу максимуму**. Розглянемо класичні, загальновідомі задачі. Розв'язання цих задач використовуються як базові управління для побудови адаптивних, "істинно кібернетичних" систем. Об'єднаємо постановку задачі з алгоритмом її розв'язання.

Порядок розв'язання оптимізаційних задач методом принципу максимуму

Дано:

критерій: $J = \int_0^T f(x, u, t) dt$;

ORIGIN := 0

рівняння динаміки процесу (об'єкта): $\frac{d}{dt}x(t)_i = f(x, u, t)_i$; $i := 0..n$; $j := 0..m$

обмеження на управління: $u(t) \in U$, наприклад:

$$-u_{0j} \leq u_j \leq u_{0j};$$

x - вектор стану об'єкта: $x = \begin{pmatrix} x_1 \\ x_i \\ x_n \end{pmatrix}$; u - вектор управління: $u = \begin{pmatrix} u_1 \\ u_j \\ u_m \end{pmatrix}$.

1. Беремо похідну від критерію і додаємо це рівняння до рівнянь динаміки процесу:

$$x(t)_0 = \frac{d}{dt}J; \quad \frac{d}{dt}x(t)_i = f(x, u, t)_i; \quad \boxed{i := 0..n}$$

2. Складаємо функцію $H(..)$ - гамільтоніан: $H = \sum_{i=0}^n \psi_i \cdot f_i$.

3. Визначаємо структуру оптимального управління, що максимізує функцію $H(..)$:

$$\max_u \{H\}.$$

4. Складаємо рівняння для визначення спряжених функцій :

$$\frac{d}{dt}\psi(t)_i = -\frac{d}{dx_i}H.$$

Через ці функції виражається оптимальне управління. Особливість принципу максимуму в тому, що **варіаційна задача знаходження функції**, яка дає екстремум функціоналу **J** замінюється **задачею математичного аналізу** з визначення параметрів оптимального управління u , структура якого, в наслідок обмежень, виявляється простою.

В цілому ж задача визначення оптимального управління зводиться до розв'язання системи з $n + n + m$ рівнянь (дифференціальних процесу, дифференціальних для спряжених функцій, рівняння для оптимальних управлінь).

Приклад 1. Синтез оптимального за швидкістю управління

Розглядаємо об'єкт другого порядку з одним (скалярним) управлінням. *Це електронна книга, використовуємо відповідні інформаційні технології: копіюємо і редагуємо "порядок розв'язання задачі"*

Дано: критерій: $J = \int_0^T 1 \, dt$; дифференціальний процес: $\frac{d^2}{dt^2} x(t) = u$;

обмеження на управління (нормоване): $|u| \leq 1$; $n := 2$ $m := 1$.

1. Записуємо розширену систему рівнянь динаміки процесу:

$$\frac{d}{dt} x_0 = f(x, u, t)_0 = 1; \quad \frac{d}{dt} x_1 = f(x, u, t)_1 = x_2; \quad \frac{d}{dt} x_2 = f(x, u, t)_2 = u.$$

Рівняння 2 і 3 отримані приведенням до форми Коші вхідного дифференціального процесу.

2. Записуємо гамильтоніан (символьний процесор не працює з індексованими змінними, тому замінюємо на прості: ψ_0 на ψ_0 і т.д.)

$$H = \sum_{i=0}^n \psi_i \cdot f_i = \psi_0 \cdot 1 + \psi_1 \cdot x_2 + \psi_2 \cdot u; \quad H(\psi, x, u) := \psi_0 \cdot 1 + \psi_1 \cdot x_2 + \psi_2 \cdot u$$

3. Аналізуємо умови, за яких функція $H(\psi, x, u)$ досягає **максимуму**: ну так очевидно, що а) тільки остання складова залежить від управління; б) для максимізації цієї складової треба, щоб управління $u(t)$ мало б1) той самий знак, що і ψ_2 , б2) максимально можливий модуль. Таким чином, управління буде релейним. тобто таким, що у певні моменти змінює знак і має граничні за модулем значення.

$$\max_u (H) \Rightarrow \max_u (\psi_2 \cdot u) \Rightarrow u_{\text{оп}} = \text{sign}(\psi(t)_2)$$

4. Записуємо і розв'язуємо рівняння для спряжених функцій (подаємо схему отримання цих рівнянь).

$$\begin{aligned} \frac{d}{dt} \psi(t)_i = -\frac{d}{dx_i} H & \Rightarrow \begin{cases} \frac{d}{dx_0} H(\psi, x, u) \rightarrow 0; & \frac{d}{dt} \psi_0 = 0; & \psi_0 = \text{const} = C_0; \\ \frac{d}{dx_1} H(\psi, x, u) \rightarrow 0; & \frac{d}{dt} \psi_1 = 0; & \psi_1 = \text{const} = C_1; \\ \frac{d}{dx_2} H(\psi, x, u) \rightarrow \psi_1; & \frac{d}{dt} \psi_2 = -\psi_1; & \psi_2 = C_2 - C_1 \cdot t. \end{cases} \end{aligned}$$

Отримання розв'язання в координатній формі

Формально ми отримали розв'язання задачі управління у вигляді:

$$u_{\text{оп}} = \text{sign}(C_2 - C_1 \cdot t)$$

Незручно це, з усіх сторін: 1) C_1 і C_2 визначаються як функції граничних умов, щоб їх обчислити, треба виміряти початковий стан (x_{10} , x_{20}) та задати кінцевий (x_{1k} , x_{2k}); 2), якщо під час перехідного процесу, під впливом збурень номінальна траєкторія зміниться, необхідно переобчислювати параметри програмного управління C_1 , C_2 .

Зручно мати управління в координатній формі - не як функцію часу а як функцію поточних координат:

$$u_{opt} = \text{sign}(x_1(t), x_2(t)) \quad (1)$$

Порядок розв'язання задачі отримання управління в координатній формі

1. Підставити в рівняння динаміки оптимальне управління u_{opt} (воно при наявності обмежень матиме 2-3 значення, наприклад: $-u_{max}$, 0 , $+u_{max}$, тобто фактично буде релейним) і знайти (побудувати відповідні фазові траєкторії.)
2. Використовуючи інформацію про число та тривалість інтервалів знакопостійності функції $\text{sign}(u(t))$ побудувати **лінію переключення**. Ця лінія ділить фазовий простір на області з різним (за знаком і величиною) управлінням.
3. Описати алгоритмічно лінію переключення як функцію поточних значень вектора стану (написати програму для моделювання чи обчислення в мікроконтролері реальної системи управління).

Виконаємо послідовно усі ці пункти для взятого прикладу. Зауважимо, що задача побудови фазових траєкторій та ліній переключення не є (на жаль, на щастя?) повністю формалізованою - серед альтернативних програм і виразів можна знайти більш компактний, елегантний варіант. Можна знайти якусь ефектну підстановку (відомо декілька підстановок Ейлера), що значно спрощує вираз для лінії переключення. Від розробника іноді потрібні такі компоненти як інтуїція, естетичне бачення задачі.

1. Побудова фазових траєкторій

Записуємо диференціальні рівняння процесу, інтегруємо їх і отримуємо такі аналітичні розв'язання:

$$\frac{d}{dt} x_1 = x_2 \Rightarrow \int_{x_{10}}^{x_1} 1 dx_1 = \int_0^t x_2(t) dt \Rightarrow x_1(t) = x_{10} + x_{20} \cdot t + \frac{t^2}{2} \quad (2)$$

$$\frac{d}{dt} x_2 = u = \pm 1 \Rightarrow \int_{x_{20}}^{x_2} 1 dx_2 = \int_0^t \pm 1 dt \Rightarrow$$

$$\text{так записуємо розв'язання для } +u \text{ та } -u: \quad x_2(t) = x_{20} + t; \quad x_2(t) = x_{20} - t \quad (3)$$

Виключаємо час з рівняння (2), для чого використовуємо (3). Згідно з вимогами сучасних стандартів, використовуємо символічний процесор для виконання підстановки.

$$t = x_2 - x_{20};$$

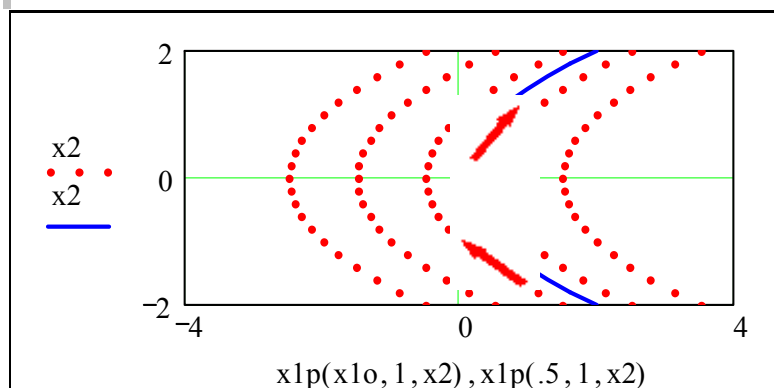
$$x_{10} + x_{20} \cdot t + \frac{t^2}{2} \text{ substitute } t = x_2 - x_{20} \rightarrow x_{10} + x_{20} \cdot (x_2 - x_{20}) + \frac{1}{2} \cdot (x_2 - x_{20})^2.$$

Записуємо рівняння для фазових траєкторій, як функцій x_2 та початкового стану x_{10}, x_{20} :

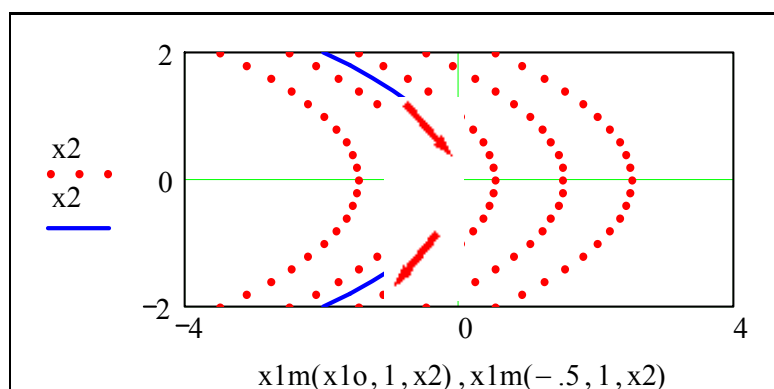
$$x_{1p}(x_{10}, x_{20}, x_2) := x_{10} + x_{20} \cdot (x_2 - x_{20}) + \frac{1}{2} \cdot (x_2 - x_{20})^2 \quad (4)$$

$$x_{1m}(x_{10}, x_{20}, x_2) := x_{10} - x_{20} \cdot (x_2 - x_{20}) - \frac{1}{2} \cdot (x_2 - x_{20})^2 \quad (5)$$

Рівняння (4) і (5) задають фазові траєкторії для управлінь $u+1$ та -1 . Отримані вирази можна спростити. Майже неважко бачити, що це рівняння парабол, симетричних відносно осі x_1 . Будуємо графіки фазових траєкторій (для цього задаємо дві ранжовані змінні – x_2 – "швидкість" та x_{10} – початкове значення x_1) $x_{10} := -2, -1..2$; $x_2 := -2, -1.8..2$.



Траєкторії для "+" управління - усі траєкторії йдуть у перший квадрант - і швидкість, і координата зростають.



Траєкторії для "-" управління - усі траєкторії йдуть у третій квадрант - і швидкість, і координата зменшуються.

Дивимось на ці графіки, бачимо, що для обох можливих управлінь можна виділити по одній траєкторії, що проходять через точку $(0,0)$ - бажаний стан рівноваги (помилка $= 0$, швидкість $= 0$). Сконструюємо лінію переключення саме з цих траєкторій.

2. Побудова ліній переключення

Виділяємо з двох траєкторій, що проходять через точку $(0,0)$ фазової площини ділянки, які ведуть (входять) до цієї точки. Склеїмо графічно і алгоритмічно з них ЛІНІЮ ПЕРЕКЛЮЧЕННЯ.

Дайте відповідь на питання: яким повинно бути управління, якщо ми (ну, не ми, а стан процесу...) знаходимось "вище" цієї лінії? Відповідь: $u = -1$. Аналогічно, якщо ми знаходимось "нижче" лінії переключення, то управління повинно бути $u = +1$.

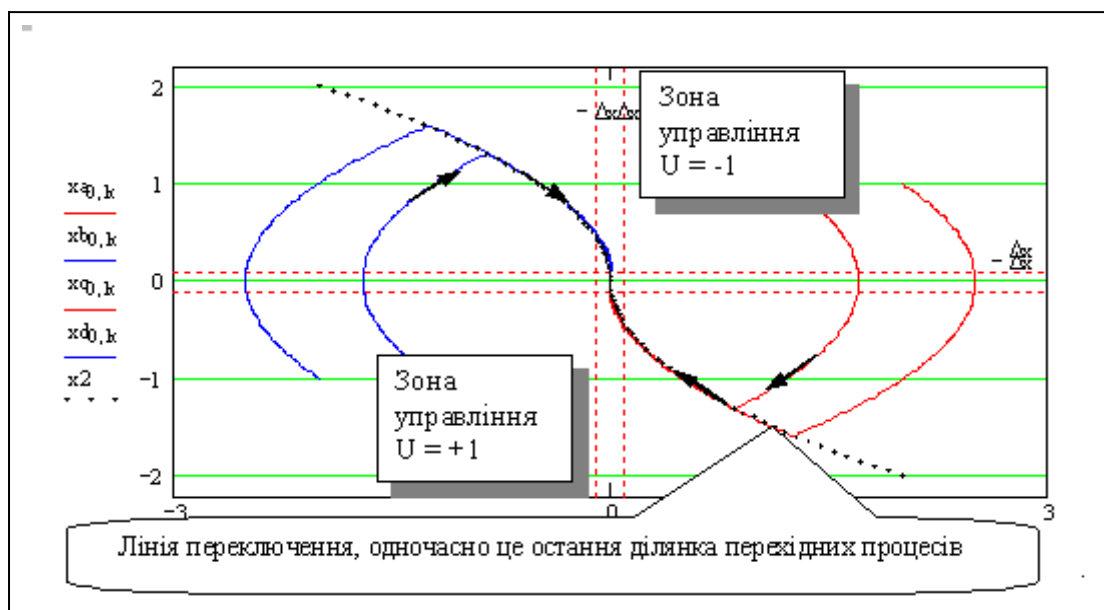
Розглянемо релейне управління з різних точок зору. В теорії автоматичного управління релейні системи "виникають" як результат того, що виконавчі елементи мають тільки два або три режими - "ввімкнено +", "вимкнено", "ввімкнено -". Це електромагнітні і пневматичні релейні елементи, тиристорні підсилювачі, ракетні двигуни. Закони управління в релейних системах, точніше закони переключення часто вибираються лінійними, або кусочно-лінійними.

В теорії оптимального управління, кібернетиці, математичній економіці ми приходимо до релейного управління з інших позицій - управління, наприклад, рівень фінансування може мати які завгодно значення - від нуля до максимально допустимого, але логіка оптимального управління теж виявляється релейною.

Дайте відповідь на питання: чим відрізняється "неоптимальне" релейне управління, зокрема з лінійними законами переключення, від оптимального? - Відповідь: в оптимальному за швидкодією управлінні при обмеженні рівня управління **лінія переключення є однією із фазових траєкторій**.

Подивіться на рисунок внизу - кінцевий етап для перехідного процесу з будь-якого початкового стану завжди проходить по лінії переключення. Але це тільки номінально. Далі ми розглянемо, яким може бути процес з оптимальним управлінням, якщо параметри процесу (об'єкта управління) будуть відрізнятися від номінальних, або управління буде реалізовуватись з певними помилками.

Слід пам'ятати, що оптимальне управління з причини надмірної чутливості до невизначеності параметрів, зовнішніх збурень і помилок управління в реальних умовах може виявитись гіршим "неоптимального".



Негативний приклад. На початкових стадіях якихось новацій завжди вважається, що науково-технічний прогрес - абсолютне благо. Наприклад, подальша еволюція Інтернету може привести до відновлення професії кур'єрів та поштових голубів - через масивні атаки хакерів та інше.

Теж саме з декларованими всюди і, зокрема, в цьому посібнику "новими інформаційними технологіями" типу "використовуй символьний процесор", "використовуй вбудовані методи розв'язання усіх рівнянь і проблем" - обережно.

Використаємо замість аналізу і логіки можливості математичного пакету для визначення рівнянь тих траєкторій, що проходять через точку (0,0) - початок координат фазової площини. Задаємо координату x20 довільної початкової точки, потім визначаємо таку точку x10, щоб траєкторія проходила через початок координат.

$$x20 := -1; \quad x2 := 0; \quad x2 := 0; \quad x10 := 11;$$

$$\text{Give1} \quad x10 + x20 \cdot (x2 - x20) + \frac{1}{2} \cdot (x2 - x20)^2 = 0;$$

$$X10 := \text{Find}(x10); \quad X10 = 0.5;$$

$$X10 + x20 \cdot (X2 - x20) + \frac{1}{2} \cdot (X2 - x20)^2 \rightarrow \frac{-1}{2} - X2 + \frac{1}{2} \cdot (X2 + 1)^2 \text{ simplify} \rightarrow \frac{1}{2} \cdot X2^2$$

Що тут власне зроблено? - знайдено X_{10} , X_{20} - початкову точку, з якої фазова траєкторія попадає у початок координат; X_{10} , X_{20} підставлено у загальне рівняння фазової траєкторії; виконано спрощення виразу і нарешті ми отримали:

$$X_1 = \frac{1}{2}X_2^2$$

- це можна було одразу записати, глянувши на графік.

А тепер запишемо вираз (програму) для лінії переключення:

LP(X_2) :=	$0.5 \cdot X_2^2 \quad \text{if } X_2 < 0$ $-0.5 \cdot X_2^2 \quad \text{if } X_2 \geq 0$
----------------	--

Можна значно спростити цей вираз - спробуйте, підставте свою відповідь у праву частину(за 13 балів) $lp(x_2) := 0.5x_2 \cdot |x_2 - 13|$ Пароль expand.

Відповідь = "не те Ви написали"

3. Побудова перехідних процесів

По різному можна розраховувати, досліджувати перехідні процеси і будувати їх графіки. Далі подана задовільна версія програми і стенду для досліджень. В даному прикладі фазові траєкторії і лінії переключення отримані у досить простому аналітичному вигляді. Для побудови перехідних процесів візьмемо більш близьку до реальності (мікропроцесорні системи управління) математичну модель з квантованим у часі управлінням. І в економіці процеси можна вважати квантованими: по роках, кварталах, місяцях, днях.

Як елемент реальності введемо умови закінчення процесу: якщо процес увійшов у допуск, тобто $|x_1| < \text{допуск}_1$ і одночасно $|x_2| < \text{допуск}_2$ - управління відключається.

Пишемо програмний модуль - функцію від початкового стану (x_{10} , x_{20}). Інші параметри задаються з головної програми (це глобальні змінні).

PP(x_{10}, x_{20}) :=	<pre> ORIGIN ← 0 x10 ← x10 x20 ← x20 k ← 0 dt ← Крок while (x1k > Δx ∨ x2k > Δx) ∧ k ≤ Nz ymov ← -x1k - 0.5 · x2k · x2k x2k+1 ← x2k + (Збач · sign(ymov) + Збуп) · dt x1k+1 ← x1k + x2k · dt Вектор_стану^(k) ← (x2k x1k) k ← k + 1 Вектор_стану </pre>
---------------------------	---

Особливість програми - наявність умов закінчення процесу: або вхід в допуск, або перевищення максимального числа кроків моделювання програми. Очевидно, для чого додана друга умова.

Тестуємо

програму: $PP(2, 1) =$

	0	1	2	3	4	5	6	7	8	9
0	1	0.98	0.95	0.93	0.9	0.88	0.86	0.83	0.81	0.78
1	2	2.04	2.08	2.12	2.15	2.19	2.23	2.26	2.29	2.33

Створюємо стенд - біля графіка вводимо (в режимі абсолютного присвоєння) вхідні дані, що використовуються програмою моделювання $PP()$:

$x1o := 2$; $x2o := 1$; $Nd := \text{cols}(PP(x1o, x2o)) - 1$; $k := 0..Nd$; $x2 := -2, -1.95..2$

;

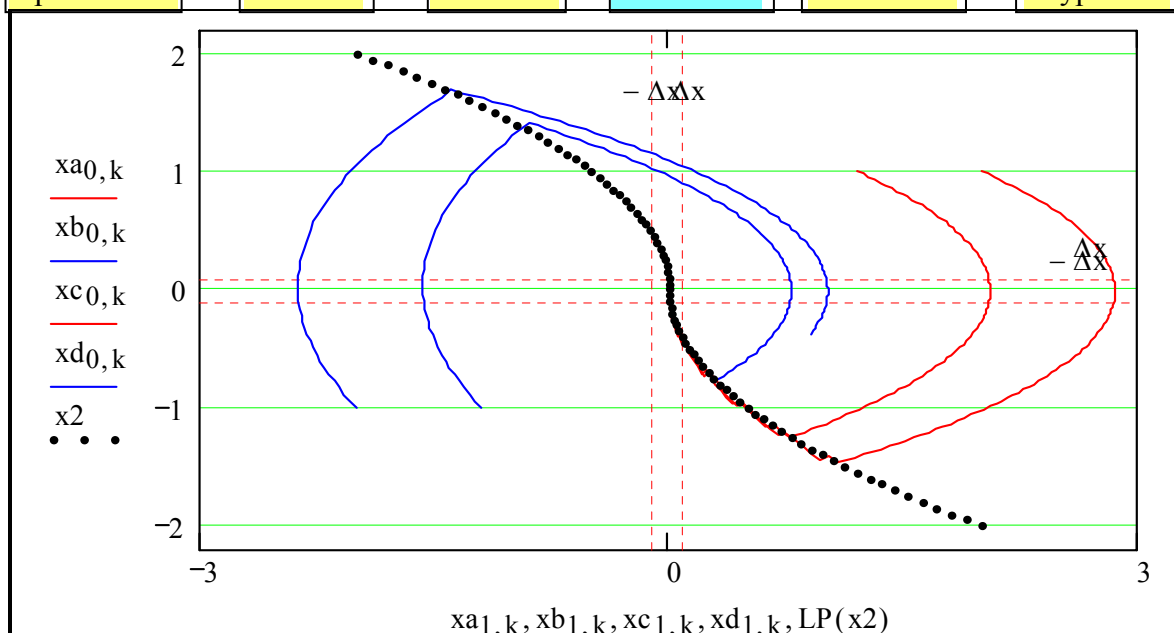
Коментар до наступного рядка. Документ швидше виконується коли попередньо "витягнути" з підпрограми дані у масиви xa, xb, xc, xd , а потім будувати графіки.

$xa := PP(x1o, x2o)$; $xb := PP(-x1o, -x2o)$;

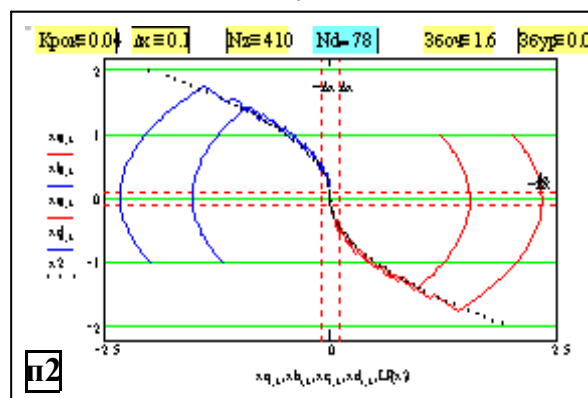
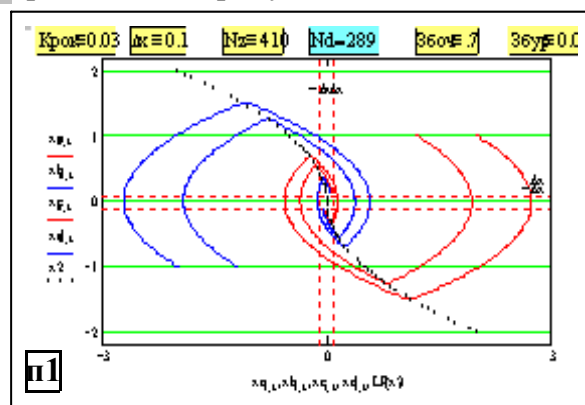
$xc := PP(0.6 \cdot x1o, x2o)$ $xd := PP(-0.6 \cdot x1o, -x2o)$

Крок обчислень, точність, число кроків: задане, дійсне, відхилення управління, зовнішнє збурення

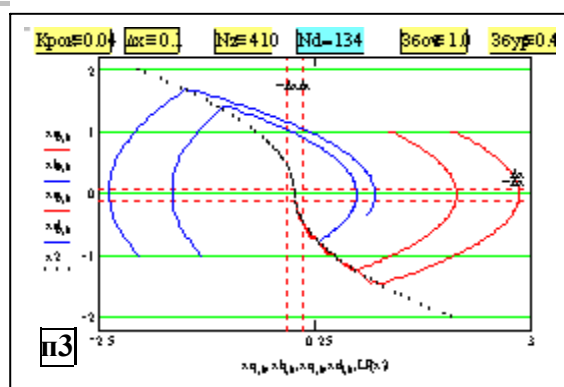
Крок $\equiv 0.04$ $\Delta x \equiv 0.1$ $Nz \equiv 410$ $Nd = 134$ $\Delta u \equiv 1.0$ $\Delta y \equiv 0.4$



Проведемо попереднє, якісне, дослідження чутливості оптимального управління до зовнішніх збурень та відхилень управління від номінального значення. Далі подано декілька копій екрана. Дивимося на ці графіки, бачимо, що отримано досить простий, але продуктивний стенд для вивчення певного класу оптимальних систем.



Ми можемо якісно і кількісно оцінювати вплив різних відхилень, збурень, запізнень, покращень або спрощень алгоритму управління. Цей стенд слід доповнити виведенням тривалості перехідного процесу, витрат енергії (пропорційно u^2), витрат робочого тіла (пропорційно $|u|$).



Контрольні запитання

1. Яким ситуаціям управління відповідають перехідні процеси п1, п2, п3 ?
2. При яких значеннях параметрів зовнішніх і внутрішніх збурень (процеси п1, п2, п3) система стає нестійкою?
3. Визначіть тривалості перехідних процесів п1, п2, п3
4. Чому дорівнюють початкові значення координат для усіх процесів ?

5. Чим відрізняється метод принципу максимуму від класичних методів варіаційного обчислення?

6. Фізичний смисл спряжених функцій?

7. Порядок розв'язання задачі методом принципу максимуму.

8. Порядок отримання управління в координатній формі: $u(t) \Rightarrow u(x1, x2)$?

9. Можливі ускладнення при реалізації оптимального управління.

Контрольні завдання

1. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^2x/dt^2 = u + x$ - 32 б.
2. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^2x/dt^2 = u - x - dx/dt$ - 34 б.
3. Використовуючи цей документ визначити залежність тривалості перехідного процесу від величини параметра Збур. - 19 б.
4. Використовуючи цей документ визначити залежність тривалості перехідного процесу від величини параметра Збач - 19 б.
5. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^3x/dt^3 = u$. (Фазові траєкторії у просторі) - 52 б.
6. Побудуйте графіки перехідних процесів у функції часу - 8 б.

6.2. Метод принципу максимуму Понтрягіна. Оптимальне управління об'єктом з демпфуванням

Цей документ є другою частиною теми "Метод принципу максимуму". В першій частині розглянуто синтез управління для об'єктів виду:

$$1) \quad d^2x/dt^2 = u ;$$

В цій частині розглядається об'єкт виду

$$2) \quad d^2x/dt^2 = u - dx/dt.$$

Майже неважко побачити, що це дифрівняння описує рух матеріальної точки у рідині з опором, пропорційним швидкості руху. Задача має багато застосувань в техніці, економіці, екології. Що нового дає цей приклад? - практичне і концептуальне освоєння нових інформаційних технологій (що є і не такими й новими).

Ми беремо попередній документ і підставляємо в нього нові математичні моделі, бачимо, що більшість програм і рівнянь "автоматично" обробляють дані нової задачі, але певні частини документа є специфічними - їх необхідно переробити. Чим менше таких частин тим більш досконалим є документ - програмний комплекс для вирішення певного класу задач.

Метод принципу максимуму досить широко використовується при розв'язанні задач оптимального управління при наявності обмежень типу нерівності. Ціль даного розділу - **освоєння техніки постановки і розв'язання оптимізаційних задач методом принципу максимуму**. Розв'язання задачі цього розділу використовуються як базове управління для побудови адаптивних - "кібернетичних" систем. Перш, ніж розробляти і досліджувати "інтелектуальні" системи на базі "нейронних мереж", "генетичних алгоритмів", нечіткої логіки, слід твердо освоїти дійсно фундаментальні методи, тому що звичайно "інтелектуальні" системи шукають **лінію переключення** для об'єкта з невідомою, нестаціонарною динамікою. Ефективність таких систем суттєво залежить від якості першого наближення. Тобто, спочатку треба дослідити поведінку об'єкта, отримати розумно спрощену - узагальнену модель, для цієї моделі отримати оптимальне управління і потім використати його, як базове в "інтелектуальній" системі. Слід пам'ятати, що певні "очевидні" терміни ("інтелектуальні системи" в першу чергу) не тільки не мають строгих визначень, а іноді, і референта. Об'єднаємо постановку задачі з алгоритмом її розв'язання.

Порядок розв'язання оптимізаційних задач методом принципу максимуму

$$\text{Дано: критерій: } J = \int_0^T f(x, u, t) dt ; \quad \text{ORIGIN} := 0$$

$$\text{рівняння динаміки процесу (об'єкта): } \frac{d}{dt} x(t)_i = f(x, u, t)_i ; i := 1..n ; j := 0..m$$

$$\text{обмеження на управління: } u(t) \in U, \text{ наприклад: } -u_{0j} \leq u_j \leq u_{0j} ;$$

$$x - \text{вектор стану об'єкта: } x = \begin{pmatrix} x_1 \\ x_i \\ x_n \end{pmatrix} ; u - \text{вектор управління: } u = \begin{pmatrix} u_1 \\ u_j \\ u_m \end{pmatrix} .$$

1. Беремо похідну від критерію і додаємо це рівняння до рівнянь динаміки процесу:

$$x(t)_0 = \frac{d}{dt} J; \quad \frac{d}{dt} x(t)_i = f(x, u, t)_i; \quad i := 0..n.$$

2. Складаємо функцію $H(\cdot)$ - гамільтоніан: $H = \sum_{i=0}^n \psi_i \cdot f_i$

3. Визначаємо структуру оптимального управління, що максимізує функцію $H(\cdot)$:

$$\max_u \{H\}.$$

4. Складаємо рівняння для визначення спряжених функцій :

$$\frac{d}{dt} \psi(t)_i = -\frac{\partial}{\partial x_i} H.$$

Через ці функції виражається оптимальне управління. Особливість принципу максимуму в тому, що **варіаційна задача знаходження функції**, яка дає екстремум функціоналу J , замінюється **задачею математичного аналізу** з визначення параметрів оптимального управління u , структура якого, в наслідок обмежень, виявляється простою.

В цілому ж задача визначення оптимального управління зводиться до розв'язання системи з $n + n + m$ рівнянь (диференціальних процесу, диференціальних для спряжених функцій, рівняння для оптимальних управлінь).

Синтез оптимального за швидкодією управління

Розглядаємо об'єкт другого порядку з одним управлінням. *Цей документ - електронна книга, копіюємо і редагуємо "порядок розв'язання задачі" з підрозділу 6.1.*

Дано:

критерій: диференціальний процесу : обмеження на управління (нормоване):

$$J = \int_0^T 1 dt; \quad \frac{d^2}{dt^2} x(t) = u - x(t); \quad |u| \leq 1; \quad n := 2 \quad m := 1.$$

1. Записуємо розширену систему рівнянь динаміки процесу :

$$\frac{d}{dt} x_0 = f(x, u, t)_0 = 1; \quad \frac{d}{dt} x_1 = f(x, u, t)_1 = x_2; \quad \frac{d}{dt} x_2 = f(x, u, t)_2 = u - x_2.$$

Рівняння 2 і 3 отримані приведенням до форми Коші вхідного диференціального рівняння.

2. Записуємо гамільтоніан. Далі подаємо те, що він видав у більш звичній формі запису.

$$H = \sum_{i=0}^n \psi_i \cdot f_i \rightarrow H = \psi_0 \cdot f_0 + \psi_1 \cdot f_1 + \psi_2 \cdot f_2;$$

$$\psi_0 \cdot f_0 + \psi_1 \cdot f_1 + \psi_2 \cdot f_2 \left| \begin{array}{l} \text{substitute, } f_0 = 1 \\ \text{substitute, } f_1 = x_2 \\ \text{substitute, } f_2 = u - x_2 \end{array} \right. \rightarrow \psi_0 + \psi_1 \cdot x_2 + \psi_2 \cdot (u - x_2).$$

$$H = \sum_{i=0}^n \psi_i \cdot f_i = \psi_0 \cdot 1 + \psi_1 \cdot x_2 + \psi_2 \cdot u \quad H(\psi, x, u) := \psi_0 \cdot 1 + \psi_1 \cdot x_2 + \psi_2 \cdot (u - x_2);$$

3. Аналізуємо умови, за яких функція $H(\psi, x, u)$ досягає **максимуму**: очевидно, що а) тільки остання складова залежить від управління; б) для максимізації цієї складової треба, щоб управління $u(t)$ мало б1) той самий знак, що і ψ_2 , б2) максимально можливий модуль. Таким чином, управління буде релейним, тобто таким, що має тільки (в даному випадку) два значення.

$$\overline{\max_u(H)} \Leftrightarrow \overline{\max_u[\psi_2 \cdot (u - x_2)]} \Leftrightarrow \boxed{u_{\text{оп}} = \text{sign}(\psi(t)_2)}$$

4. Записуємо і розв'язуємо рівняння для спряжених функцій (подаємо схему отримання цих рівнянь).

$$\begin{aligned} \frac{d}{dt} \psi(t)_i = -\frac{d}{dx_i} H & \begin{cases} \xrightarrow{\text{d}} \frac{d}{dx_0} H(\psi, x, u) \rightarrow 0; \quad \frac{d}{dt} \psi_0 = 0; \quad \psi_0 = \text{const} = C_0; \\ \xrightarrow{\text{d}} \frac{d}{dx_1} H(\psi, x, u) \rightarrow 0; \quad \frac{d}{dt} \psi_1 = 0; \quad \psi_1 = \text{const} = C_1; \\ \xrightarrow{\text{d}} \frac{d}{dx_2} H(\psi, x, u) \rightarrow \psi_1 - \psi_2; \quad \frac{d}{dt} \psi_2 = \psi_1 - \psi_2; \quad \psi_2 = C_2 - C_1 \cdot t \end{cases} \end{aligned}$$

Отримання розв'язання в координатній формі

Формально ми отримали розв'язання задачі - управління у вигляді:

$$\boxed{u_{\text{оп}} = \text{sign}(C_2 - C_1 \cdot t)}$$

Незручно це, з двох причин - а) C_1 і C_2 визначаються як функції граничних умов, щоб їх обчислити, треба виміряти початковий стан (x_{10}, x_{20}) та задати кінцевий (x_{1k}, x_{2k}); б) якщо під час перехідного процесу, під впливом збурень номінальна траєкторія зміниться, необхідно переобчислювати параметри програмного управління C_1, C_2 .

Зручно мати управління в координатній формі - не як функцію часу, а як функцію поточних координат:

$$\boxed{u_{\text{оп}} = \text{sign}(x_1(t), x_2(t))} \quad (1)$$

Порядок розв'язання задачі отримання управління в координатній формі

1. Підставити в рівняння динаміки оптимальне управління $u_{\text{оп}}$ (воно при наявності обмежень матиме 2-3 значення, наприклад: $-u_{\text{мах}}, 0, +u_{\text{мах}}$, тобто фактично буде релейним) і знайти (побудувати відповідні фазові траєкторії.)

2. Використовуючи інформацію про число та тривалість інтервалів знакопостійності функції $\text{sign}(\psi_2(t))$ побудувати лінію переключення. Ця лінія ділить фазовий простір на області з різними (за знаком і величиною) управліннями.

3. Описати алгоритмічно лінію переключення як функцію поточних значень вектора стану (написати програму для моделювання чи обчислення в мікроконтролері реальної системи управління).

Виконаємо послідовно усі ці пункти визначення управління в координатній формі для взятого прикладу.

1. Побудова фазових траєкторій

Записуємо диференціальний процес, інтегруємо їх і отримуємо аналітичні розв'язання (це елементарні рівняння):

$$\frac{d}{dt}x_2 = u - x_2 \Leftrightarrow \frac{d}{dt}x_2 + x_2 = u \Leftrightarrow s + 1 = 0 \Leftrightarrow s = -1 \Leftrightarrow x_2 = C_1 \cdot e^{-1 \cdot t} + u.$$

Визначаємо константу C для випадків $u = 1$ та $u = -1$

$$x_{20} = C_1 \cdot e^{-1 \cdot 0} + 1 \text{ solve, } C_1 \rightarrow x_{20} - 1; \quad x_{20} = C_1 \cdot e^{-1 \cdot 0} - 1 \text{ solve, } C_1 \rightarrow x_{20} + 1.$$

Отримуємо, нарешті вирази для координати x_2 траєкторій при $u = 1$, $u = -1$.

$$x_{2plus} = (x_{20} - 1) \cdot e^{-1 \cdot t} + 1; \quad x_{2minus} = (x_{20} + 1) \cdot e^{-1 \cdot t} + 1 \quad (2)$$

Знаходимо вирази для координати x_1 .

$$\frac{d}{dt}x_1 = x_2 \Leftrightarrow \int_{x_{10}}^{x_1} 1 dx_1 = \int_0^t x_2(t) dt \Leftrightarrow x_1(t) = x_{10} + \int_0^t x_2(t) dt \quad (3)$$

Можна виключити час з рівняння (3), використавши (2), але в даному випадку зручніше подати фазову траєкторію у параметричному вигляді.

Записуємо x_1 та x_2 як функції часу t та початкового стану x_{10} , x_{20} . Виконуємо інтегрування та алгебраїчні перетворення і отримуємо рівняння траєкторій для $u = +1$, $u = -1$. Використовуємо символічний процесор.

$$\int_0^t [(x_{20} - 1) \cdot e^{-1 \cdot tt} + 1] dt \rightarrow (-x_{20} + 1 + t \cdot \exp(t)) \cdot \exp(-t) + x_{20} - 1$$

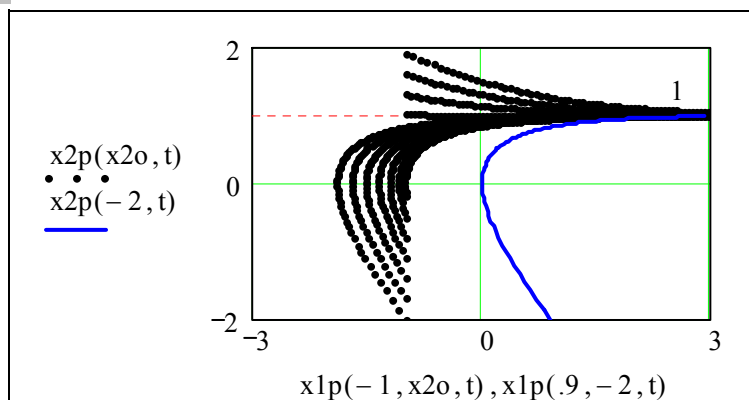
$$\left. \begin{aligned} \text{Згідно з (3) маємо} \quad x_{2p}(x_{20}, t) &:= [(x_{20} - 1) \cdot e^{-1 \cdot t} + 1] \\ x_{1p}(x_{10}, x_{20}, t) &:= x_{10} + [(-x_{20} + 1 + t \cdot \exp(t)) \cdot \exp(-t) + x_{20} - 1] \end{aligned} \right\} \quad (4)$$

$$\int_0^t [(x_{20} + 1) \cdot e^{-1 \cdot tt} - 1] dt \rightarrow (-x_{20} - 1 - t \cdot \exp(t)) \cdot \exp(-t) + x_{20} + 1$$

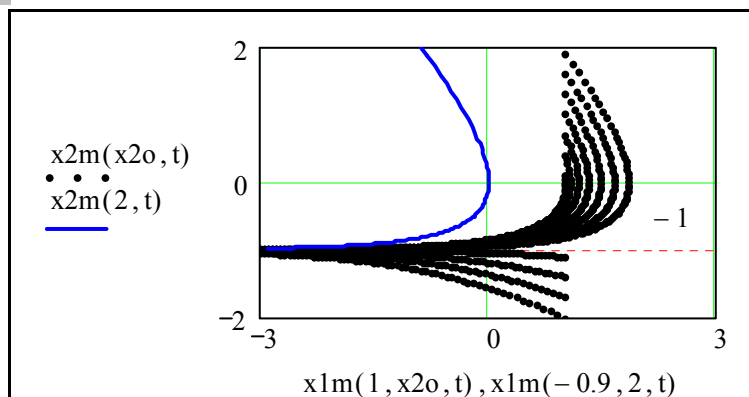
$$\left. \begin{aligned} x_{2m}(x_{20}, t) &:= (x_{20} + 1) \cdot e^{-1 \cdot t} - 1 \\ x_{1m}(x_{10}, x_{20}, t) &:= x_{10} + [(-x_{20} - 1 - t \cdot \exp(t)) \cdot \exp(-t) + x_{20} + 1] \end{aligned} \right\} \quad (5)$$

Будуємо графіки фазових траєкторій для різних початкових умов (для цього задаємо ранжовані змінні x_2 та t - початкову координату і час.)

$$x_{2o} := -2, -1.7 \dots 2; \quad t := 0, 0.05 \dots 5$$



Траєкторії для "+" управління - траєкторії, що починаються у 3-му і 4-му квадрантах і йдуть в перший квадрант, де швидкість $x_2 > 0$ і координата $x_1 > 0$.
Для усіх траєкторій встановлюється одна й та ж швидкість $x_2(\infty) = 1$.



Траєкторії для "-" управління - траєкторії, що починаються у 1-му і 2-му квадрантах і йдуть в 3-ій квадрант, де швидкість $x_2 < 0$ і координата $x_1 < 0$.
Для усіх траєкторій встановлюється одна й та ж швидкість $x_2(\infty) = -1$.

Для обох можливих управлінь можна виділити по одній траєкторії (виділені лінії), що проходять через точку (0,0) - бажаний стан рівноваги (помилка = 0, швидкість = 0). Сконструємо лінію переключення саме з цих траєкторій.

2. Побудова ліній переключення

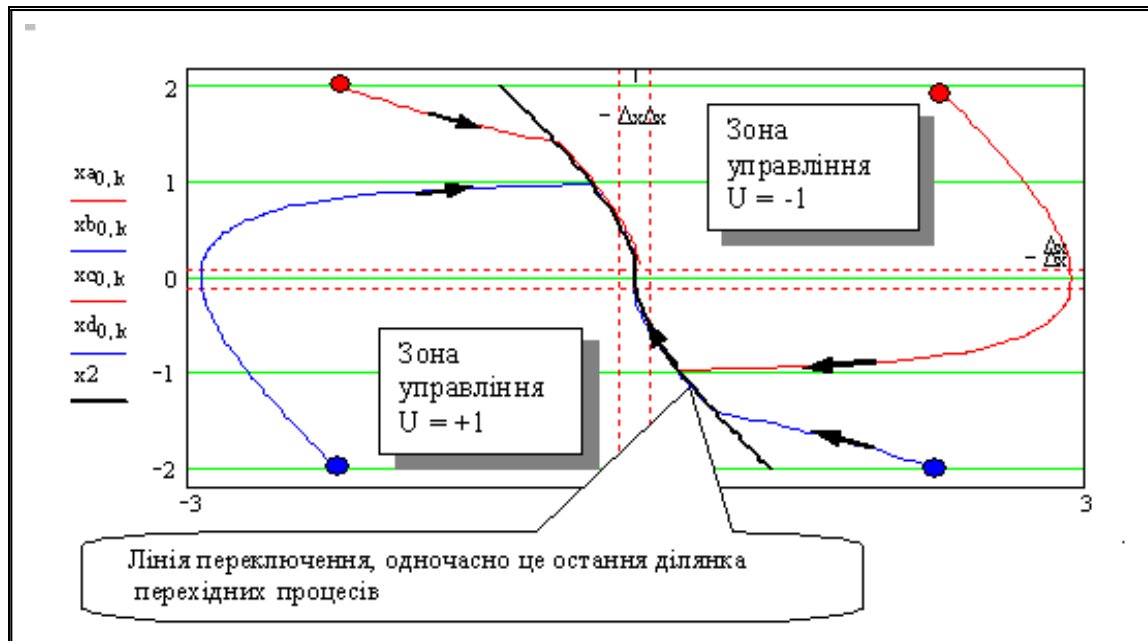
Виділяємо з двох траєкторій, що проходять через точку (0,0) фазової площини ділянки, які ведуть (входять) до цієї точки. Склеїмо графічно і алгоритмічно з них лінію переключення.

Дайте відповідь на питання: яким повинно бути управління, якщо ми (ну, не ми, а стан процесу...) знаходимось "вище" лінії переключення? Відповідь: $u = -1$ (див. нижче). Аналогічно, якщо ми знаходимось "нижче" лінії переключення, то управління повинно бути $u = +1$.

В теорії автоматичного управління іноді релейне управління розглядається, як результат використання простих дешевих регуляторів і підсилювачів, як ерзац "пропорційного" управління. В теорії оптимального управління, кібернетиці, математичній економіці ми приходимо до релейного управління з інших позицій - управляюча дія (наприклад, рівень фінансування) може мати які завгодно значення - від нуля до максимально допустимого, але логіка оптимального управління теж виявляється релейною, наприклад: "спочатку все в інвестиції, потім все у споживання".

Дайте відповідь на питання: чим відрізняється "неоптимальне" релейне управління, зокрема з лінійними законами переключення, від оптимального? (336). Відповідь: в оптимальному за швидкодією управлінні при обмеженні управління лінія переключення є однією із фазових траєкторій.

Подивіться на рисунок вгорі - кінцевий етап для **оптимального** перехідного процесу з будь-якого початкового стану завжди проходить по лінії переключення.



На цьому рисунку подано лінії переключення та 4 перехідних процеси.

Візьмемо в (4-5) початковий стан $X10 = 0$, $X20 = 0$. Щоб отримати рівняння лінії переключення, що зв'язує $x1$ та $x2$, спочатку з (4а) визначимо t , потім підставимо у (4б). Виконаємо все це за допомогою символьного процесора.

$$X2 = (x20 - 1) \cdot e^{-1 \cdot t} + 1 \text{ substitute } x20 = 0 \rightarrow \text{solve, } t \rightarrow$$

$$X1p = x10 + [(-x20 + 1 + t \cdot \exp(t)) \cdot \exp(-t) + x20 - 1]$$

$$x10 + [(-x20 + 1 + t \cdot \exp(t)) \cdot \exp(-t) + x20 - 1] \left| \begin{array}{l} \text{substitute, } x10 = 0 \\ \text{substitute, } x20 = 0 \end{array} \right. \rightarrow (1 + t \cdot \exp(t)) \cdot \exp(-t) - 1$$

$$(1 + t \cdot \exp(t)) \cdot \exp(-t) - 1 \text{ substitute } t = -\ln(-X2 + 1) \rightarrow \left[1 - \frac{\ln(-X2 + 1)}{(-X2 + 1)} \right] \cdot (-X2 + 1) - 1$$

$$X1p = \left[1 - \frac{\ln(-X2 + 1)}{(-X2 + 1)} \right] \cdot (-X2 + 1) - 1$$

Виконаємо теж саме для лінії переключення в зоні $X1 > 0$:

$$X2 = (x20 + 1) \cdot e^{-1 \cdot t} - 1 \text{ substitute } x20 = 0 \rightarrow X2 = \exp(-t) - 1 \text{ solve, } t \rightarrow -\ln(X2 + 1)$$

$$X1m = x10 + [(-x20 - 1 - t \cdot \exp(t)) \cdot \exp(-t) + x20 + 1]$$

$$x10 + [(-x20 - 1 - t \cdot \exp(t)) \cdot \exp(-t) + x20 + 1] \left| \begin{array}{l} \text{substitute, } x10 = 0 \\ \text{substitute, } x20 = 0 \end{array} \right. \rightarrow (-1 - t \cdot \exp(t)) \cdot \exp(-t) + 1$$

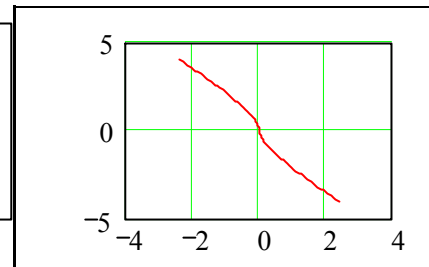
$$(-1 - t \cdot \exp(t)) \cdot \exp(-t) + 1 \text{ substitute } t = -\ln(X2 + 1) \rightarrow \left[-1 + \frac{\ln(X2 + 1)}{(X2 + 1)} \right] \cdot (X2 + 1) + 1$$

$$X1m = \left[-1 + \frac{\ln(X2 + 1)}{(X2 + 1)} \right] \cdot (X2 + 1) + 1$$

$$xx2 := -4, -3.9 \dots 4$$

"Склеюємо" з X1m та X1p єдину лінію переключення і тут же тестуємо:

$$LP(X2) := \begin{cases} 1 - \frac{\ln(-X2 + 1)}{(-X2 + 1)} \cdot (-X2 + 1) - 1 & \text{if } X2 < 0 \\ -1 + \frac{\ln(X2 + 1)}{(X2 + 1)} \cdot (X2 + 1) + 1 & \text{if } X2 \geq 0 \end{cases}$$



3. Побудова перехідних процесів

По різному можна розраховувати, досліджувати перехідні процеси і будувати їх графіки. В нашому прикладі фазові траєкторії та лінії переключення отримані у відносно простому аналітичному вигляді, що не завжди можливе. Для побудови перехідних процесів візьмемо більш близьку до реальності дискретну математичну модель з квантованим у часі управлінням.

Як елемент реальності введемо умови закінчення процесу: якщо процес увійшов у допуск, тобто $|x1| < \text{допуск}x1$ і одночасно $|x2| < \text{допуск}x2$ - управління відключається. Робимо програмний модуль функцією від параметрів початкового стану $x10, x20$. Інші параметри задаються з головної програми (це глобальні змінні).

PP (x10 ,x20) :=	$x1_0 \leftarrow x10$ $x2_0 \leftarrow x20$ $k \leftarrow 0$ $dt \leftarrow \text{Крок}$ "далі буде" while $\left(x1_k > \Delta x \vee x2_k > \Delta x \right) \wedge k \leq Nz$ $y_{mov} \leftarrow -x1_k + LP(x2_k) + 0.00 \cdot \text{sign}(x1_k)$ $x2_{k+1} \leftarrow x2_k + \left[\text{sign}(y_{mov}) - 3\text{боч} \cdot (x2_k + 3\text{буп}) \right] \cdot dt$ $x1_{k+1} \leftarrow x1_k + x2_k \cdot dt$ Вектор_стану $\langle k \rangle \leftarrow \begin{pmatrix} x2_k \\ x1_k \\ \text{sign}(y_{mov}) \end{pmatrix}$ $k \leftarrow k + 1$ Вектор стану
------------------	--

Тестуємо програму:

$$PP(2, 1) =$$

	0	1	2	3	4	5	6	7	8	9
0	1	0.9	0.8	0.71	0.62	0.54	0.46	0.38	0.3	0.23
1	2	2.04	2.08	2.11	2.14	2.16	2.18	2.2	2.22	2.23
2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Створюємо стенд - біля графіка вводимо (в режимі абсолютного присвоєння) вхідні дані, що використовуються програмою моделювання PP()

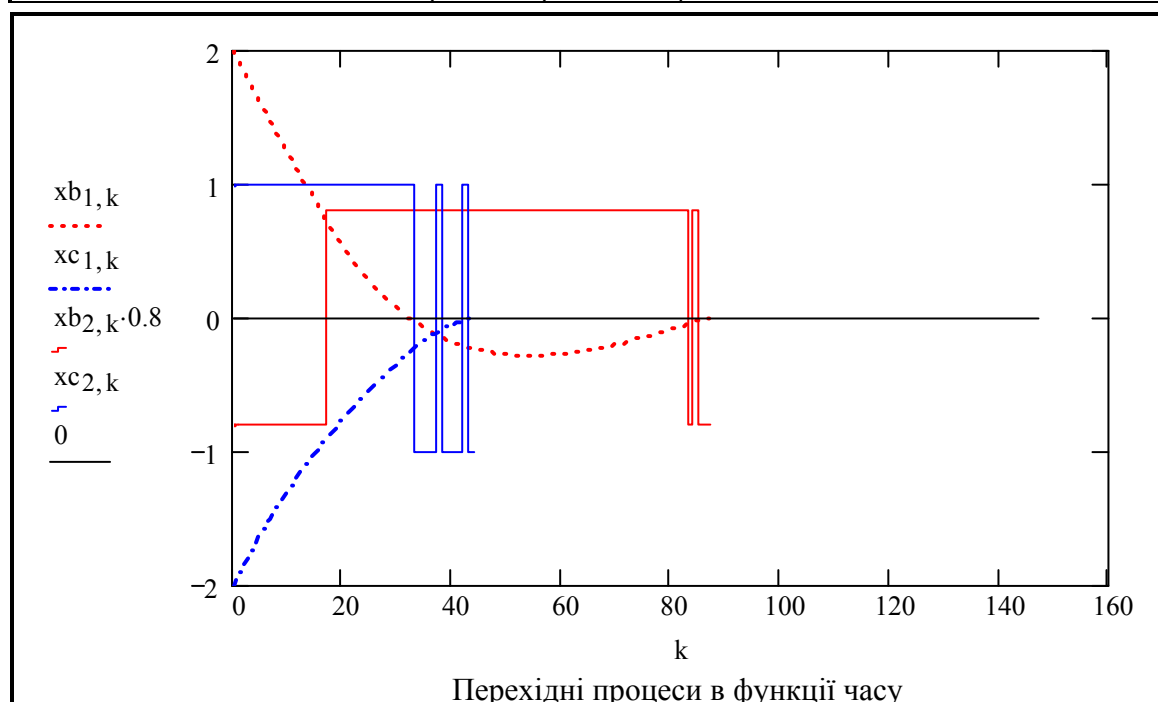
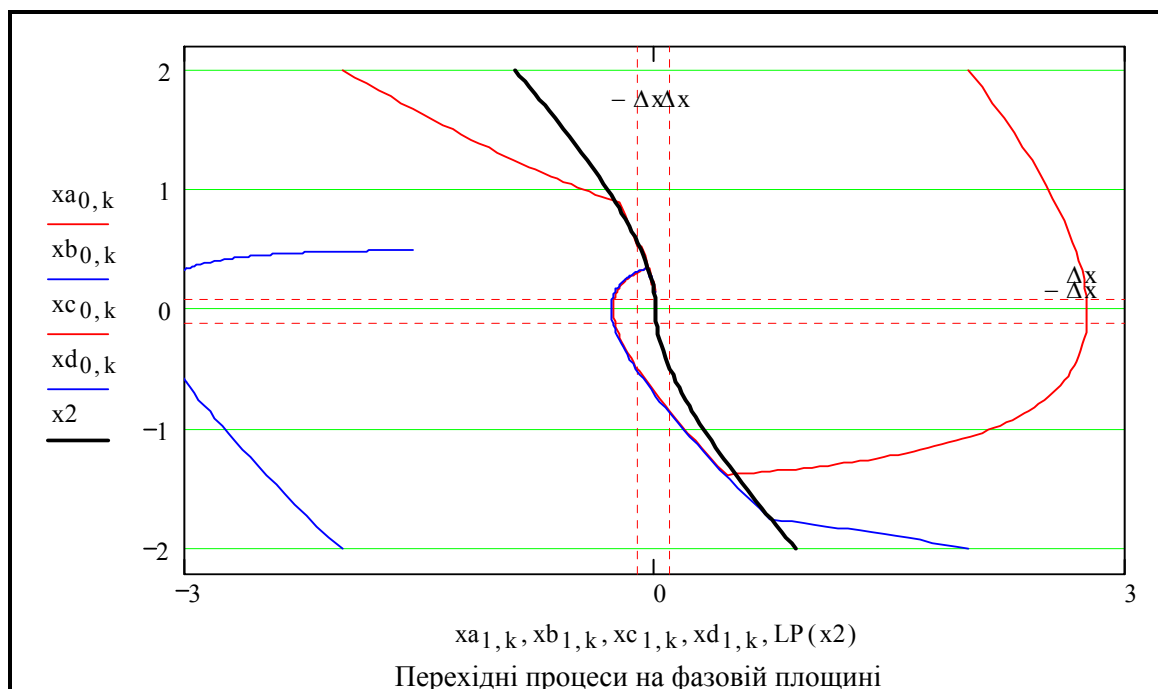
$x1o := 2;$ $x2o := 2;$ $Nd := \text{cols}(\text{PP}(x1o, x2o)) - 1;$ $k := 0..Nd;$ $x2 := -2, -1.95..2$

Коментар до наступного рядка. Документ швидше виконується, якщо "витягти" з підпрограми дані у масиви **ха**, **xb**, **xc**, **xd**, а потім будувати графіки. Уявіть собі, якщо цього не зробити, то, кожен раз, коли потрібен якийсь рядок чи стовпець вихідної матриці PP(x1o,x2o), програмний модуль знову обчислює весь процес.

$xa := \text{PP}(x1o, x2o)$ $xb := \text{PP}(x1o, -x2o)$ $xc := \text{PP}(-x1o, x2o)$ $xd := \text{PP}(-x1o, -x2o)$

Крок обчислень, точність, число кроків: задане, дійсне, відхилення управління, зовнішнє збурення.

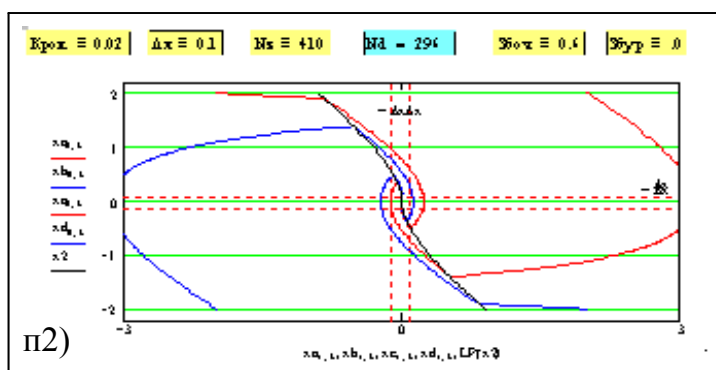
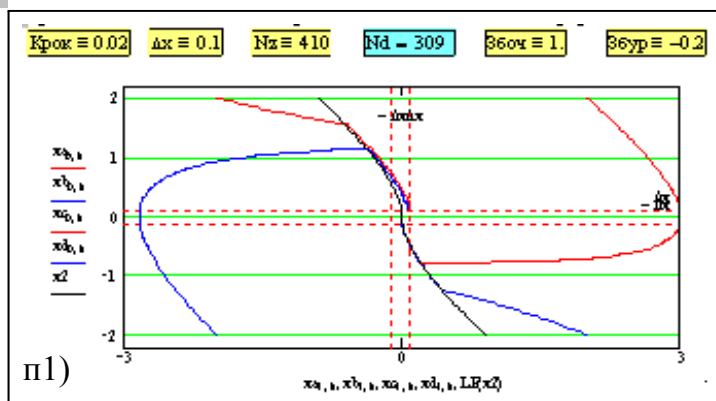
Крок $\equiv 0.04$ $\Delta x \equiv 0.1$ $Nz \equiv 410$ $Nd = 147$ $Зб\text{оч} \equiv 1.$ $Зб\text{ур} \equiv .5$



Завдання. Подивіться на подані вище графіки перехідних процесів - на фазовій площині та в функції часу.

1. На верхньому графіку побудовано 4 процеси, які саме з них подано на нижньому графіку? 2. Змініть щось у зоні введення, щоб два перехідних процеси на останньому графіку стали симетричними (збурення?). 3. Чому в кінці збурені перехідні процеси мають багато переключень (при "Збур" = 0.5).

Подаємо декілька копій екрану.



Контрольні запитання

1. Яким ситуаціям управління відповідають перехідні процеси п 1, п2 ?
2. При яких значеннях параметрів зовнішніх і внутрішніх збурень (процеси п1, п2) система стає нестійкою?
3. Визначіть тривалості перехідних процесів для процесів п 1, п2.
4. Чому дорівнюють початкові значення координат для усіх процесів ?

Контрольні завдання

1. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^2x/dt^2 = u + x$ - 32 бали.
2. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^2x/dt^2 = u - x - dx/dt$ - 33 бали.
3. Використовуючи цей документ визначити залежність тривалості перехідного процесу від величини параметра Збур. - 17 балів.
4. Використовуючи цей документ визначити залежність тривалості перехідного процесу від величини параметра Збоч. - 17 балів.
5. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^3x/dt^3 = u$.(Фазові траєкторії у просторі) - 52 бали.
6. Побудуйте графіки перехідних процесів у функції часу - 1 бал.

6.3. Метод принципу максимуму Понтрягіна. Оптимальне управління об'єктом-маятником

Вступ. Постановка задачі

Цей документ є третьою частиною розділу. В перших двох частинах розглянуто синтез управління для об'єктів виду:

$$1) \quad d^2x/dt^2 = u; \quad 2) \quad d^2x/dt^2 = u - dx/dt.$$

В цій частині розглядається об'єкт виду

$$3) \quad d^2x/dt^2 = u - x.$$

Майже неважко побачити, що це дифрівняння є рівнянням динаміки маятника без розсіювання енергії коливань. Що нового дає цей третій приклад?

Для тих, хто любить математику: в прикладах 1 і 2 корені характеристичних рівнянь були дійсними, а в прикладі 3 - уявні. Як побачимо, це породжує дещо іншу структуру оптимального управління. Принципова відмінність цього прикладу від попередніх в тому, що функція управління може мати як завгодно велике число інтервалів знакопостійності - чим більше початкове відхилення від стану рівноваги, тим більше число переключень управління.

Ціль даного розділу - освоєння і закріплення техніки постановки і розв'язання оптимізаційних задач методом принципу максимуму. Щоб зробити цей документ незалежним від інших розділів, повторюємо методичну частину попередніх документів (не завжди сучасний користувач може поставити на своєму комп'ютері повну електронну книгу). Об'єднаємо постановку задачі з алгоритмом її розв'язання.

Порядок розв'язання оптимізаційних задач методом принципу максимуму

Дано: критерій: $J = \int_0^T f(x, u, t) dt;$

рівняння динаміки процесу (об'єкта): $\frac{d}{dt}x(t)_i = f(x, u, t)_i; \quad i := 1..n; \quad j := 0..m.$

обмеження на управління: $u(t) \in U$, наприклад: $-u_{0j} \leq u_j \leq u_{0j};$

x - вектор стану об'єкта: $x = \begin{pmatrix} x_1 \\ x_i \\ x_n \end{pmatrix};$ u - вектор управління: $u = \begin{pmatrix} u_1 \\ u_j \\ u_m \end{pmatrix}.$

1. Беремо похідну від критерію і додаємо це рівняння до рівнянь динаміки процесу:

$$x(t)_0 = \frac{d}{dt}J; \quad \frac{d}{dt}x(t)_i = f(x, u, t)_i; \quad i := 0..n.$$

2. Складаємо функцію $H(\cdot)$ - гамільтоніан:
$$H = \sum_{i=0}^n \psi_i \cdot f_i.$$

3. Визначаємо структуру оптимального управління, що максимізує функцію
 $H(\cdot): \max_u \{H\}$.

4. Складаємо рівняння для визначення спряжених функцій :

$$\frac{d}{dt} \psi(t)_i = -\frac{d}{dx_i} H.$$

Через ці функції виражається оптимальне управління. Особливість принципу максимуму в тому, що **варіаційна задача знаходження функції**, яка дає екстремум функціоналу J замінюється задачею математичного аналізу з визначення параметрів оптимального управління u , структура якого, в наслідок обмежень, виявляється простою. В цілому ж задача визначення оптимального управління зводиться до розв'язання системи з $2 \cdot n + m$ рівнянь (диференціальних процесу, диференціальних для спряжених функцій, рівняння для оптимальних управлінь).

Синтез оптимального за швидкістю управління для коливального об'єкта

Розглядаємо об'єкт другого порядку з одним (скалярним) управлінням. Застосуємо до нього загальну методику.

Дано: критерій: $J = \int_0^T 1 dt$; диференціальний процесу: $\frac{d^2}{dt^2} x(t) = u - x$;

обмеження на управління (нормоване): $|u| \leq 1$; $n := 2$; $m := 1$

1. Записуємо розширену систему рівнянь динаміки процесу :

$$\frac{d}{dt} x_0 = f(x, u, t)_0 = 1; \quad \frac{d}{dt} x_1 = f(x, u, t)_1 = x_2; \quad \frac{d}{dt} x_2 = f(x, u, t)_2 = u - x.$$

Рівняння 2 і 3 отримані приведенням до форми Коші заданого диференціального об'єкта управління.

2. Записуємо гамільтоніан (символьний процесор не працює з індексованими змінними, тому замінюємо на прості: ψ_0 на ψ_0 і т.д.)

$$H = \sum_{i=0}^n \psi_i \cdot f_i = \psi_0 \cdot 1 + \psi_1 \cdot x_2 + \psi_2 \cdot (u - x_1);$$

$$H(\psi, x, u) := \psi_0 \cdot 1 + \psi_1 \cdot x_2 + \psi_2 \cdot (u - x_1);$$

$$\psi_0 \cdot 1 + \psi_1 \cdot x_2 + \psi_2 \cdot (u - x_1) \text{ expand, } u \rightarrow \psi_0 + \psi_1 \cdot x_2 + \psi_2 \cdot u - \psi_2 \cdot x_1$$

3. Аналізуємо умови, за яких функція $H(\psi, x, u)$ досягає **максимуму**: майже очевидно, що а) тільки складова $\psi_2 \cdot u$ залежить від управління; б) для максимізації цієї складової треба, щоб управління $u(t)$ мало: б1) той самий знак, що і ψ_2 , б2) а також максимально можливий модуль. Таким чином, управління буде релейним, тобто таким, що має тільки два можливих значення: $u = +1$, $u = -1$.

$$\boxed{\max_u (H)} \Leftrightarrow \boxed{\max_u (\psi_2 \cdot u)} \Leftrightarrow \boxed{u_{op} = \text{sign}(\psi(t)_2)}$$

4. Записуємо і розв'язуємо рівняння для спряжених функцій (подаємо схему отримання цих рівнянь). Тут R - константа, що визначається з граничних умов.

$$\frac{d}{dt} \psi(t)_i = -\frac{d}{dx_i} H \begin{cases} \xrightarrow{\quad} \frac{d}{dx_0} H(\psi, x, u) \rightarrow 0 ; \quad \frac{d}{dt} \psi_0 = 0 ; \quad \psi_0 = \text{const} = C_0 ; \\ \xrightarrow{\quad} \frac{d}{dx_1} H(\psi, x, u) \rightarrow -\psi_2 ; \quad \frac{d}{dt} \psi_1 = \psi_2 ; \quad \psi_1 = R \cdot \sin(t) ; \\ \xrightarrow{\quad} \frac{d}{dx_2} H(\psi, x, u) \rightarrow \psi_1 ; \quad \frac{d}{dt} \psi_2 = -\psi_1 ; \quad \psi_2 = R \cdot \cos(t) \end{cases}$$

Отримання розв'язання в координатній формі

Формально ми отримали розв'язання задачі - управління у вигляді:

$$u_{\text{opt}} = \text{sign}(C_2 - C_1 \cdot t)$$

Таке управління називається програмним. Воно неефективне, якщо параметри об'єкта неточно відомі, змінюються з часом, якщо на об'єкт діють зовнішні збурення.

Дійсно, оптимальне управління, розраховане до початку процесу, потім може стати неадекватним. В таких випадках використовується управління із зворотним зв'язком, або управління в координатній формі:

$$u_{\text{opt}} = \text{sign}(x_1(t), x_2(t)) \quad (1)$$

Порядок розв'язання задачі отримання управління в координатній формі

Подаємо спочатку перелік кроків процедури отримання управління у вигляді (1)

1. Підставити в рівняння динаміки оптимальне управління u_{opt} (воно при наявності обмежень матиме 2-3 значення, наприклад: $-u_{\text{max}}$, 0, $+u_{\text{max}}$, тобто фактично буде релейним) і знайти (побудувати відповідні фазові траєкторії.)
2. Використовуючи інформацію про число та тривалість інтервалів знакопостійності функції $\text{sign}(\psi_2(t))$ побудувати лінію переключення. Ця лінія ділить фазовий простір на області з різними за знаком і величиною управліннями.
3. Описати алгоритмічно лінію переключення як функцію поточних значень вектора стану.

Ідеально було б створити програму для отримання управління в координатній формі. Спробуйте зробити таку програму, після ознайомлення з наступними трьома сторінками.

1. Побудова фазових траєкторій

Застосуємо подані вище кроки до В цьому прикладі рівняння фазових траєкторій визначимо іншим методом. В прикладах 1 і 2 ми спочатку інтегрували диференціальні рівняння процесу (вони були елементарними з роздільними змінними), потім виключали час (знаходили з одного рівняння час і підставляли в інше).

Зробимо так: поділимо друге рівняння на перше:

$$\frac{d}{dt} x_2 = u - x_1 = \pm 1 - x_1; \quad \frac{d}{dt} x_1 = x_2 ; \quad (2)$$

розділимо змінні, запишемо рівняння для випадків $u = +1$ та $u = -1$:

$$x_2 \cdot dx_2 = (u - x_1) \cdot dx_1 \quad \Rightarrow \quad \text{a) } x_2 \cdot dx_2 = (1 - x_1) \cdot dx_1; \quad \text{b) } x_2 \cdot dx_2 = (-1 - x_1) \cdot dx_1$$

і проінтегруємо праву і ліву частину рівнянь по своїх змінних:

$$\int_{x_{20}}^{x_2} x_2 dx_2 = \int_{x_{10}}^{x_1} 1 dx_1 - \int_{x_{10}}^{x_1} x_1 dx_1 \Rightarrow \frac{x_2^2}{2} - \frac{x_{20}^2}{2} = x_1 - x_{10} - \frac{x_1^2}{2} + \frac{x_{10}^2}{2}; \quad (3)$$

виконуємо алгебраїчні перетворення і от маємо: $\boxed{(x_1 + 1)^2 + x_2^2 = R^2}$,

де $R^2 = x_{20}^2 + 1 - 2 \cdot x_{10} + x_{10}^2 = x_{20}^2 + (x_{10} + 1)^2$

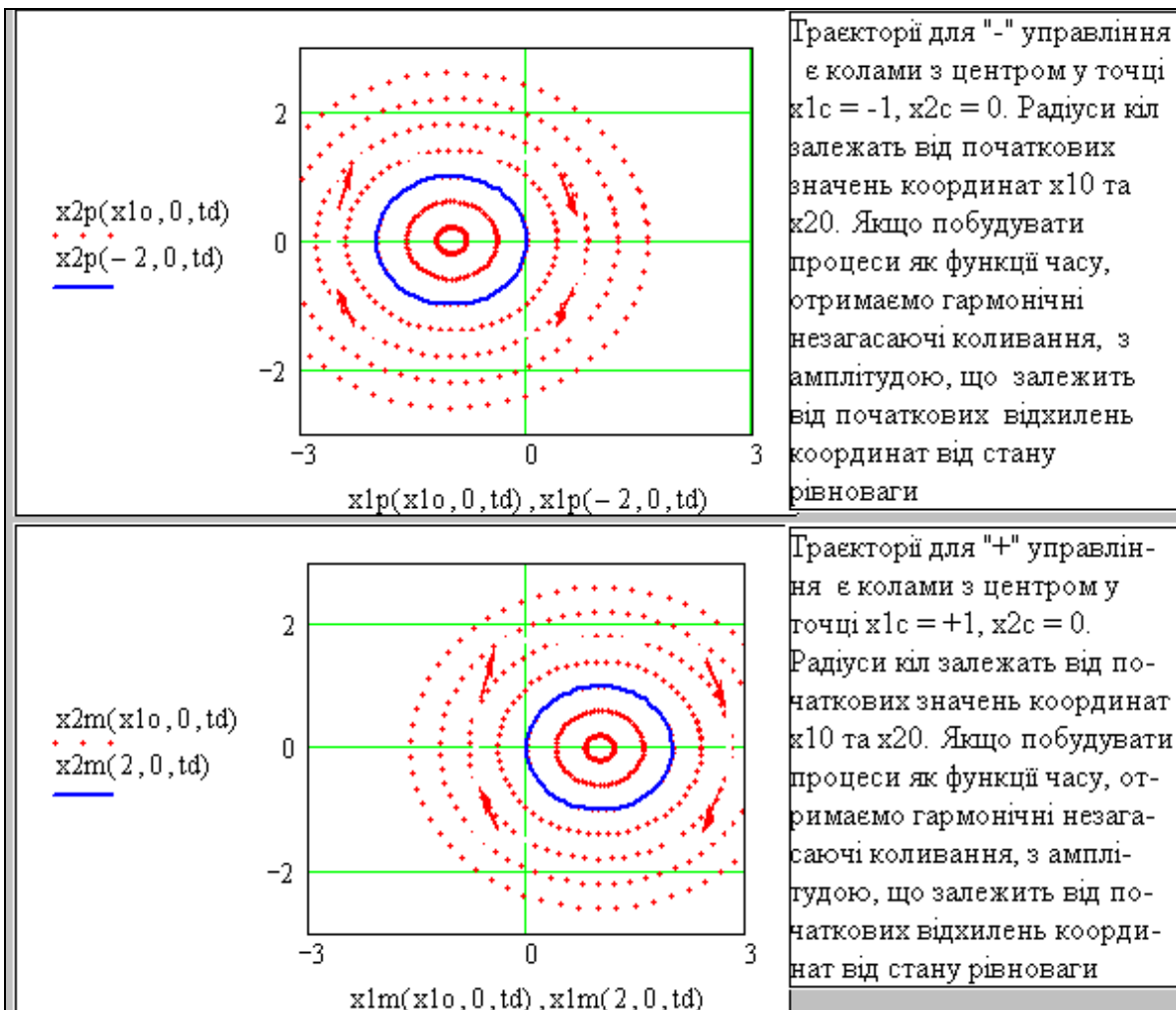
Записуємо рівняння для фазових траєкторій. Це зробимо теж інакше, ніж у розділах 6.1 і 6.2, а саме, задамо x_{1p} та x_{2p} як функції параметра td - часу та початкового стану x_{10}, x_{20} :

$$x_{1p}(x_{10}, x_{20}, td) = R \cdot \cos(td) - 1; \quad x_{2p}(x_{10}, x_{20}, td) = R \cdot \sin(td). \quad (4)$$

Підставляємо вираз для R і отримуємо робочі вирази для фазових траєкторій.

$$\left. \begin{aligned} x_{1p}(x_{10}, x_{20}, td) &:= \sqrt{x_{20}^2 + (x_{10} + 1)^2} \cdot \cos(td) - 1 \\ x_{2p}(x_{10}, x_{20}, td) &:= \sqrt{x_{20}^2 + (x_{10} + 1)^2} \cdot \sin(td) \\ x_{1m}(x_{10}, x_{20}, td) &:= \sqrt{x_{20}^2 + (x_{10} - 1)^2} \cdot \cos(td) + 1 \\ x_{2m}(x_{10}, x_{20}, td) &:= \sqrt{x_{20}^2 + (x_{10} - 1)^2} \cdot \sin(td) \end{aligned} \right\} \quad (5)$$

Будуємо графіки фазових траєкторій. Для цього задаємо дві ранжовані змінні x_{20}, x_{10} - початкові значення. Маємо рівняння кіл, симетричних відносно осі x_1 . Задаємо ранжовані змінні для побудови графіків: $x_{10} := -1.6, -1.2 \dots 1.6$; $td := 0, 0.1 \dots 6.6$.

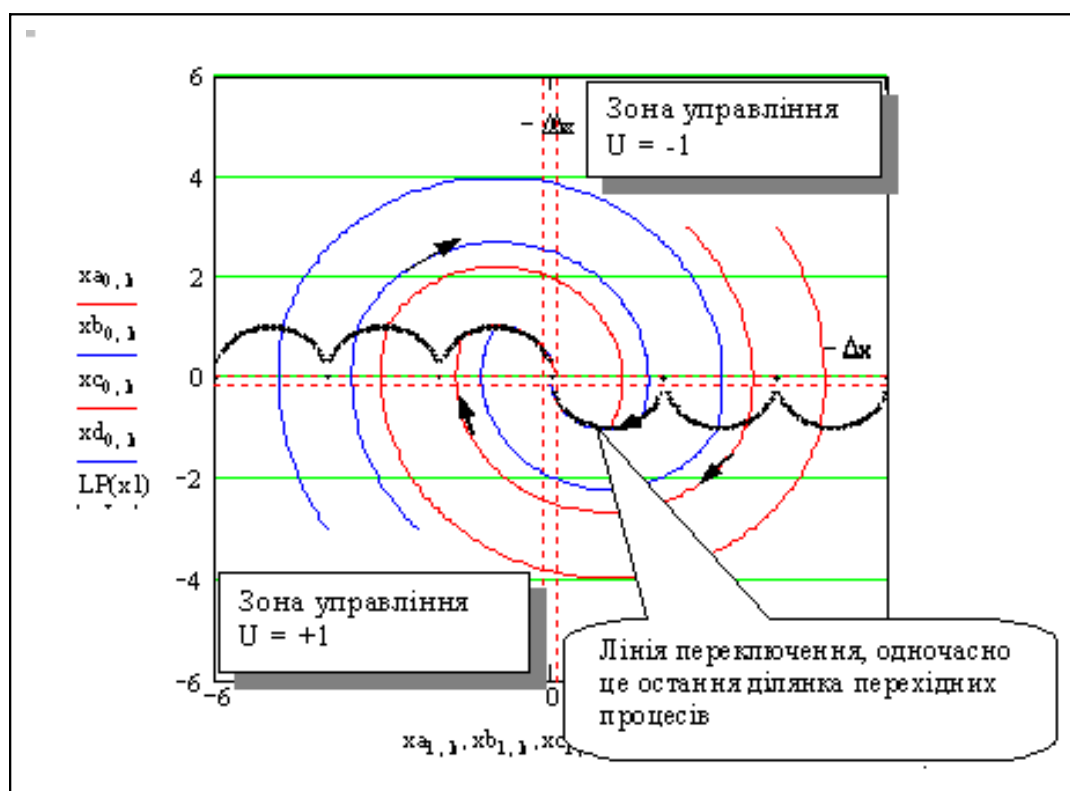


Таким чином фазові траєкторії утворюють два сімейства кіл з центрами в точках з координатами $x_1 = -1$ та $x_1 = +1$. Радіуси кіл залежать від початкових умов. Для обох можливих управлінь можна виділити по одній траєкторії, що проходять через точку $(0,0)$ - бажаний стан рівноваги (помилка = 0, швидкість = 0). Сконструємо лінію переключення саме з цих траєкторій.

2. Побудова ліній переключення

Виділяємо з двох траєкторій, що проходять через точку $(0,0)$ фазової площини ділянки, які ведуть до цієї точки. Склеїмо графічно і алгоритмічно з них ЛП. В даному випадку ЛП принципово відрізняється від ліній переключення для прикладів з підрозділів 6.1 та 6.2. Подивіться на графік внизу. **Спробуйте самостійно осмислити, чому саме такими є лінії переключення для коливального об'єкта управління.**

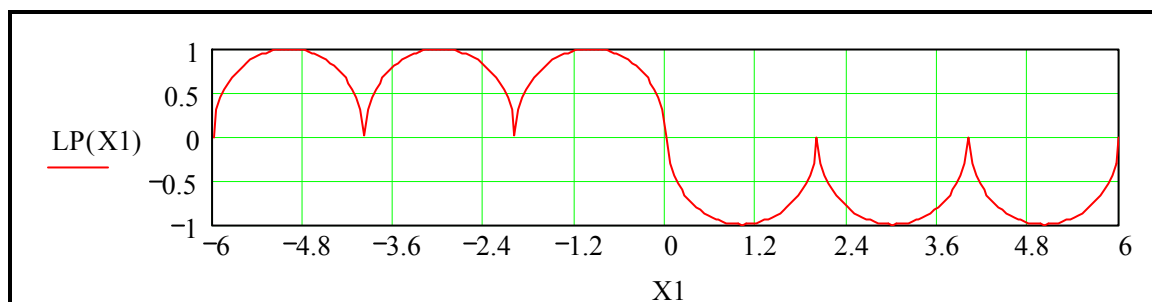
Кінцевий етап для номінального перехідного процесу з будь-якого початкового стану завжди проходить по лінії переключення.



А тепер запишемо вираз (функцію користувача) для ЛП:

$$D := 6; h := .05; X1 := -D, (-D + h) .. D; LP(X1) := -\text{sign}(X1) \cdot \sqrt{1 - (\text{mod}(|X1|, 2) - 1)^2}$$

Будуємо графік лінії переключення



3. Побудова перехідних процесів

Далі подана задовільна альтернатива програми і стенда для досліджень. В даному прикладі фазові траєкторії і лінії переключення отримані у досить простому аналітичному вигляді. Вибираємо більш близьку до реальності (мікропроцесорні системи управління) математичну модель з квантованим у часі управлінням. Введемо умови закінчення процесу: якщо процес увійшов у допуск, тобто $|x_1| < \text{допуск_}x_1$ і одночасно $|x_2| < \text{допуск_}x_2$ - моделювання процесу закінчується. Формуємо програмний модуль як функцію з параметрами x_{10} , x_{20} - координатами початкового стану. Інші параметри задаються з головної програми (це глобальні змінні).

PP(x10, x20) :=	<pre> ORIGIN ← 0 x10 ← x10 x20 ← x20 k ← 0 dt ← Крок while (x1_k > Δx ∨ x2_k > Δx) ∧ k ≤ Nz ymova ← sign(x1_k) · √(1 - (mod(x1_k , 2) - 1)²) + x2_k x2_{k+1} ← x2_k + (-x1_k + -3боч · sign(ymova) + 3буп) · dt x1_{k+1} ← x1_k + x2_k · dt Вектор_стану^{⟨k⟩} ← $\begin{pmatrix} x2_k \\ x1_k \\ \text{sign}(ymova) \end{pmatrix}$ k ← k + 1 Вектор_стану </pre>
-----------------	--

Тестуємо програму - виводимо таблицю (матрицю, масив) даних процесу

PP(2, 1) =

	0	1	2	3	4	5	6	7	8	9
0	1	0.85	0.7	0.54	0.39	0.23	0.07	-0.09	-0.25	-0.41
1	2	2.05	2.09	2.13	2.15	2.17	2.19	2.19	2.18	2.17
2	1	1	1	1	1	1	1	1	1	1

Створюємо стенд - будуємо графік, біля графіка вводимо (в режимі абсолютного присвоєння) вхідні дані для програми моделювання. Обчислюємо і виводимо одразу чотири процеси з різними початковими умовами.

$x_{10} := 4$; $x_{20} := 3$; $N_d := \text{cols}(\text{PP}(x_{10}, x_{20})) - 1$; $k := 0..N_d$; $x_1 := -6, -5.95..6$.

Коментар до наступного рядка. Документ швидше виконується коли попередньо "витягти" з підпрограми дані у масиви x_a , x_b , x_c , x_d , а потім вже будувати графіки.

$x_a := \text{PP}(x_{10}, x_{20})$; $x_b := \text{PP}(-x_{10}, -x_{20})$; $x_c := \text{PP}(0.6 \cdot x_{10}, x_{20})$; $x_d := \text{PP}(-0.6 \cdot x_{10}, -x_{20})$

Крок обчислень, точність, число кроків: задане, дійсне, відхилення управління, зовнішнє збурення

Крок $\equiv 0.05$

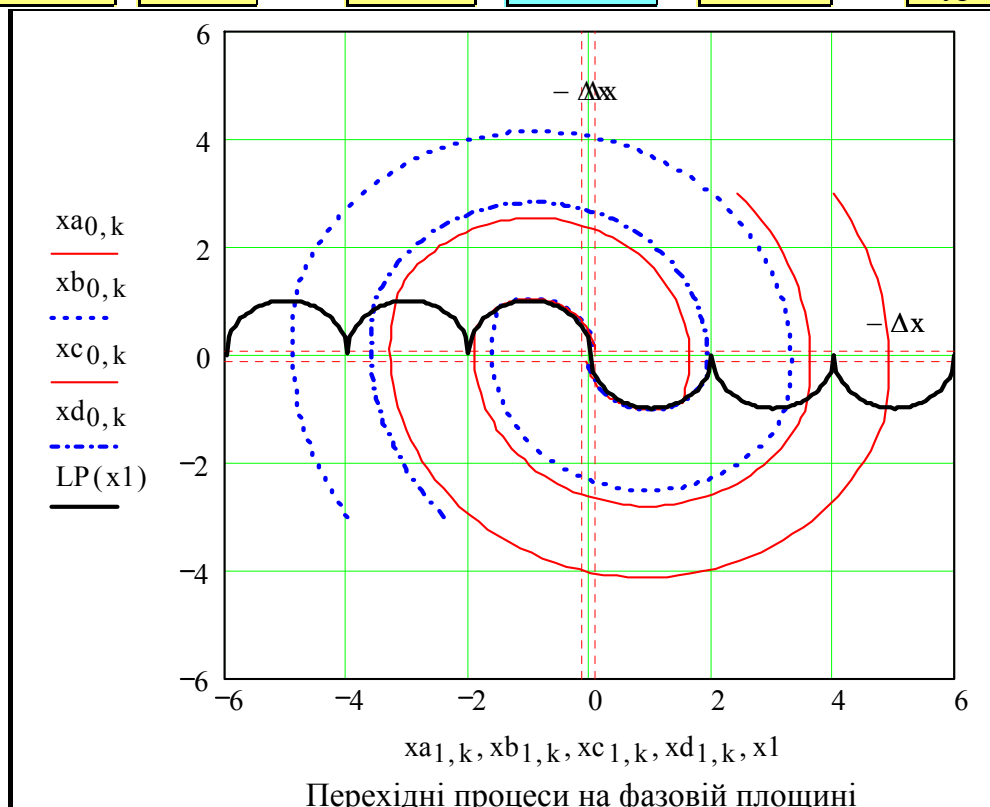
$\Delta x \equiv 0.1$

$N_z \equiv 410$

$N_d = 176$

$Зб_{оч} \equiv 1$

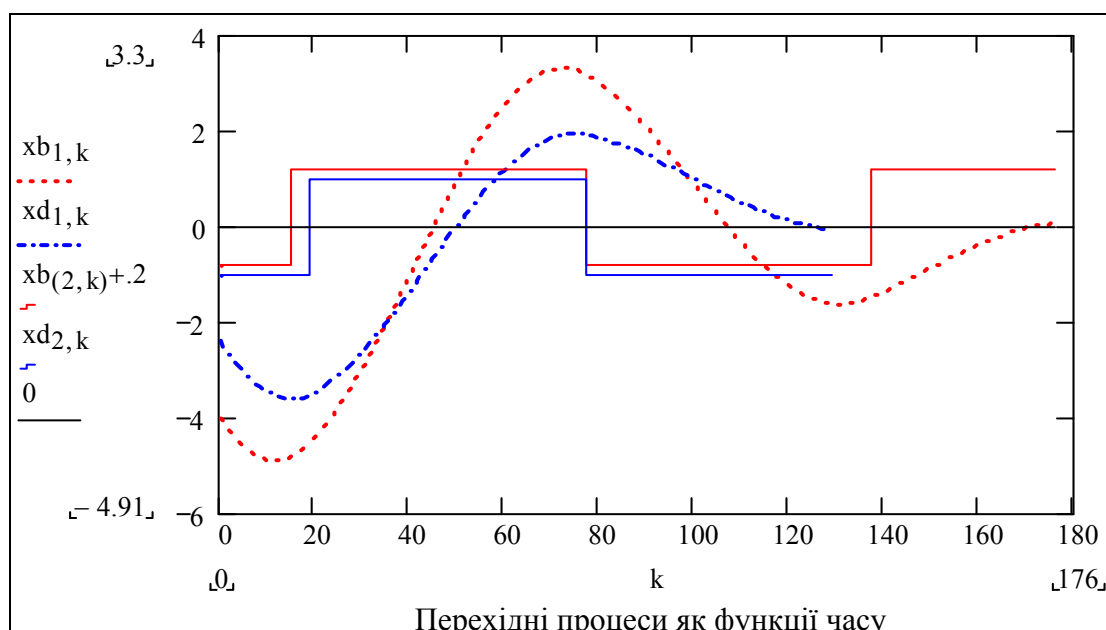
$Зб_{ур} \equiv 0$



Завдання. Подивіться на подані вище графіки перехідних процесів - на фазовій площині та в функції часу. 1. На верхньому графіку побудовано 4 процеси, які саме з них подано на нижньому графіку? 2. Як показати на графіку моменти переключення? Модифікуйте вихід програми моделювання, так щоб виводилось і управління.

Побудуємо також графіки перехідних процесів як функції часу. Сформуємо функцію визначення тривалості перехідного процесу:

$$\text{Тривалість_ПП}(x1, x2) := \text{cols}(\text{PP}(x1, x2)) \cdot \text{Крок}$$



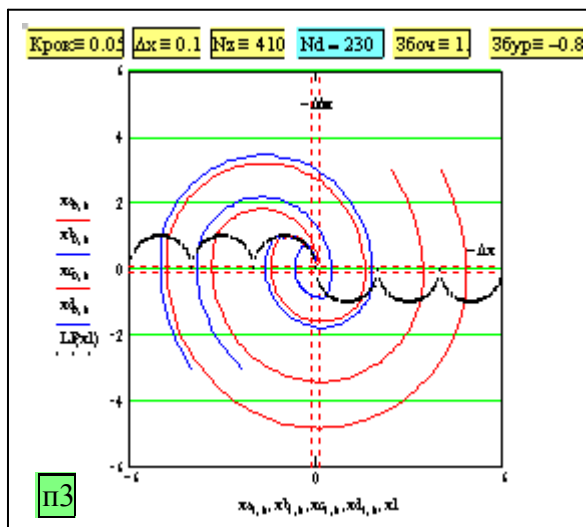
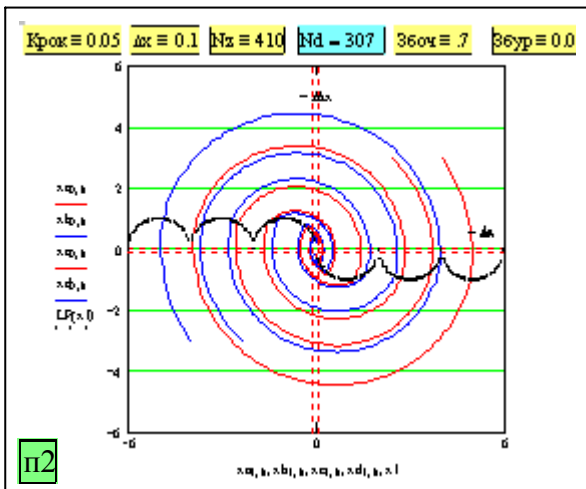
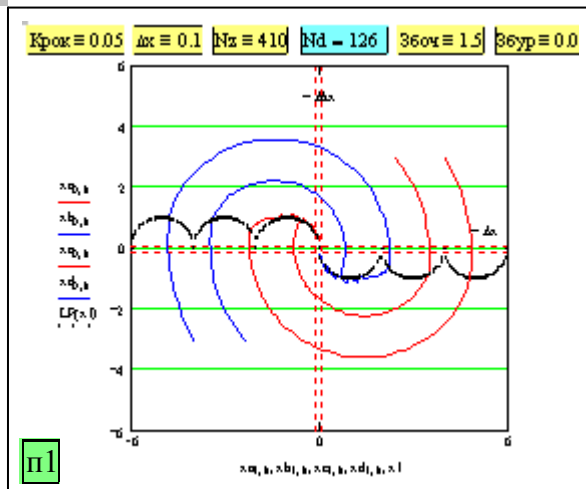
Виводимо тривалості для серії з 4-ох ПП, що подані вище на графіку.

Тривалість_ПП(x_{10}, x_{20}) ;

Тривалість_ПП($0.6 \cdot x_{10}, x_{20}$) ;

Тривалість_ПП($-x_{10}, -x_{20}$) ;

Тривалість_ПП($-0.6 \cdot x_{10}, -x_{20}$)



Контрольні запитання

1. Яким ситуаціям управління відповідають перехідні процеси C1, C2, C3 ?
2. При яких значеннях параметрів зовнішніх і внутрішніх збурень (ситуації C1, C2, C3) система стає нестійкою?
3. Визначіть тривалості перехідних процесів для ситуацій C1, C2, C3.
4. Чому дорівнюють початкові значення координат для усіх процесів ?

Завдання для самостійного виконання

1. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^2x/dt^2 = u + x$.
2. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^2x/dt^2 = u - x - dx/dt$.
3. Використовуючи цей документ визначити залежність тривалості перехідного процесу від величини параметра Збур.
4. Використовуючи цей документ визначити залежність тривалості перехідного процесу від величини параметра Збоч.
5. Модифікувати документ для синтезу та дослідження оптимального за швидкодією управління для об'єкта: $d^3x/dt^3 = u$. (Фазові траєкторії у просторі).

7

Оптимізація на базі нечіткої логіки

У цьому розділі розглядаються:

- ✓ Приклад нечіткої оптимізації на базі експертних оцінок.
- ✓ Оптимізація надійності систем альтернативними методами.

Після вивчення цього розділу ви **зможете:**

1. Знаходити в своїй професійній діяльності задачі оптимізації системи за мультиплікативним критерієм при обмеженні ресурсу – задачі надійності, якості, пропускної спроможності. Виконувати формалізацію цих задач.
2. Розробляти програми дискретної (цілочислової оптимізації).
3. Переводити оптимізаційну задачу у нечіткий формат: вибирати вид і параметри функцій належності для цілі та обмежень.
4. Вибирати метод побудови загальної функції належності виходячи із специфіки задачі.
5. Виконувати обробку експертних оцінок і побудову відповідних функцій належності (виконувати квантифікацію емпіричних залежностей нечіткими методами).
6. Виконувати дослідження „що_буде_якщо” і аналіз чутливості для оптимізаційної задачі.

7. Дати визначення термінам:

- | | |
|---|---|
| 1) розмите обмеження; | 5) експертна оцінка; |
| 2) функція належності обмеження; | 6) метод невизначених множників Лагранжа; |
| 3) функція належності цілі; | 7) метод динамічного програмування; |
| 4) комбінування (згортка) функцій належності; | 8) надійність, відмовостійкість |

7.1. Приклад нечіткої оптимізації на базі експертних оцінок

Вступ

Ефективність діяльності інженера, менеджера визначається не стільки знанням окремих методів і окремих задач, а тим, наскільки він легко, впевнено застосовує ці методи для розв'язання своїх задач. В цьому розділі для псевдоконкретної задачі комплексно використовуються: ідентифікація емпіричних залежностей, оптимізація на базі розмитої логіки, графічні методи аналізу задачі та ін. Оптимізація на базі розмитої логіки вже була використана для задач лінійного програмування. Тепер розглянемо цей метод на прикладі задачі з нелінійною цільовою функцією і обмеженнями.

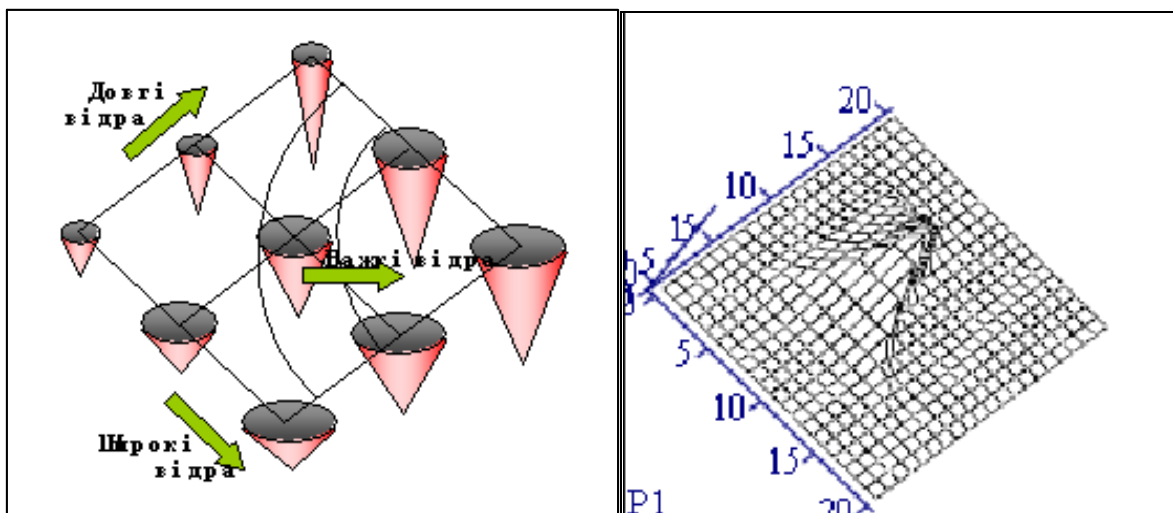
Постановка задачі

Детально ознайомитись з об'єктом оптимізації можна 1) по книзі В.Ф.Очкова "Mathcad 8 Pro для инженеров и студентов", 2) переглянувши серію "Маски на пожежі", 3) оглянувши пожежний щит десь на підприємстві.

Пожежне відро (рисунок ліворуч) повинно **задовольняти суперечливі і нечіткі вимоги** - бути:

а) в міру широким, б) помірно високим, в) не дуже важким, але досить вагомим для гасіння пожежі. Треба:

а) формалізувати нечіткі вимоги, б) сконструювати цільову функцію і знайти оптимальні розміри відра (рисунок праворуч)



Конструювання функції належності для радіуса відра

Задаємо спочатку функцію, якою будемо наближати експертні оцінки - нормальний розподіл. Чому саме нормальний? - принциповим є тільки топологія, можна взяти довільну гладку додатну функцію з максимумом в середині діапазону.

$$\mu(\text{Par}, A, B) := \begin{cases} \text{"Саморобна ф-ція норм. розподілу (є вбудована dnorm(.))"} \\ \exp\left[-A \cdot (B - \text{Par})^2\right] \end{cases}$$

Подаємо дані опитування експертів про оптимальний радіус відра

$$\mu_r := \begin{pmatrix} 25 & 50 & 75 & 100 & 125 & 150 & 175 & 200 & 225 & 250 & \text{"Радіус відра (мм)"} \\ .05 & .12 & .22 & .42 & .8 & 1 & .65 & .47 & .29 & .11 & \text{"0-незручно, 1-зручно"} \end{pmatrix}$$

Розв'язуємо задачу апроксимації даних опитування у вибраному двопараметричному класі функцій. Це буде **функція належності** для радіуса відра. Підбираємо оптимальні за якимось (= квадратичним) критерієм відхилення від емпіричного розподілу параметри функції. Використовуємо розв'язувальний блок. Задаємо початкові значення параметрів, що оптимізуються:

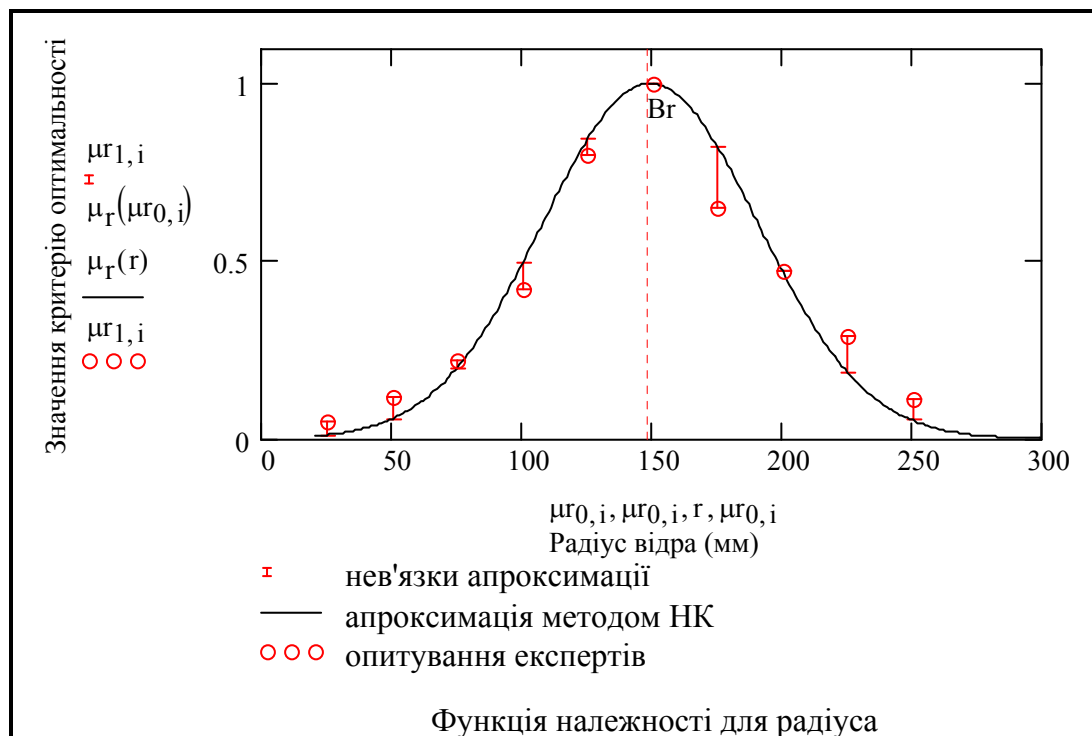
$$n := \text{cols}(\mu_r) - 1 ; n = 10 ; A := -0.013 ; B := 100 .$$

Мінімізуємо суму квадратів відхилень:

$$\text{Give} \sum_{i=0}^{n-1} (\mu_{r1,i} - \mu(\mu_{r0,i}, A, B))^2 = 0 ; \begin{pmatrix} A_r \\ B_r \end{pmatrix} := \text{MinErr}(A, B) .$$

Виводимо результати. Будуємо графік - перевіряємо, чи справді функція наближує емпіричні дані.

$$A_r = -0.000292 ; B_r = 149.212 ; \mu_r(r) := \mu(r, A_r, B_r) ; i := 0..n-1 \quad r := 20..300 .$$



Конструювання функції належності для висоти відра

Аналогічно отримуємо функцію належності для висоти відра. Вводимо дані опитування експертів про оптимальну висоту відра.

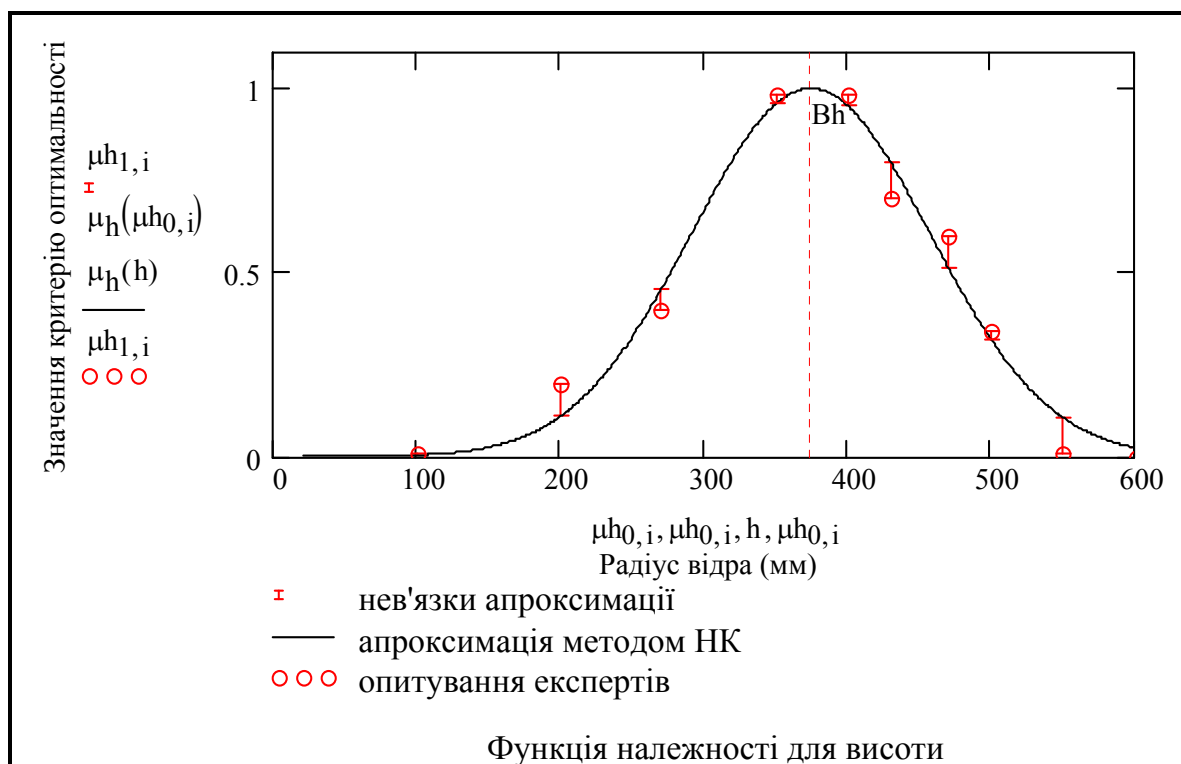
$$\mu_h := \begin{pmatrix} 100 & 200 & 270 & 350 & 400 & 430 & 470 & 500 & 550 & 600 & \text{"Висота відра (мм)"} \\ .01 & .2 & .4 & .98 & .98 & .7 & .6 & .34 & .01 & .0 & \text{"0-незручно, 1-зручно"} \end{pmatrix} ;$$

Цінні результати (алгоритми програми) можна закрити під паролем. Це робиться так: Ставимо курсор на вільне місце, в меню **Insert** вибираємо пункт **Area** - з'являться дві лінії. Їх можна пересувати мишею так, щоб охопити певну зону. Потім у меню **Format** вибираємо теж **Area**. Там будуть пункти Lock (блокувати від змін, Unlock - розблокувати, Collapse - сховати, Expand - розкрити сховане. Ясно, коли ці пункти потрібні, і що з ними робити. Як позбавитись від зони - клацнути мишею на лінії відкритої зони. Головне правило - **завжди пишіть пароль поряд із зоною**. Не закриваємо цю зону:

$$A := -0.01 \quad B := 400 \quad \text{Given}_{n-1} \quad \sum_{i=0} (\mu_{h1,i} - \mu(\mu_{h0,i}, A, B))^2 = 0 \quad \begin{pmatrix} Ah \\ Bh \end{pmatrix} := \text{MinErr}(A, B)$$

Виводимо результати. Будуємо графік - перевіряємо апроксимацію.

$Ah = -0.000073$; $Bh = 374.148$; $\mu_h(h) := \mu(h, Ah, Bh)$; $i := 0..n-1$ $h := 20..600$.



Конструювання функції належності для об'єму відра

Ще раз повторюємо процедуру - формуємо функцію належності для об'єму відра. Вводимо дані опитування експертів про оптимальну масу (вагу) відра.

$$\mu_v := \begin{pmatrix} 3 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 13 & 15 & \text{"Об'єм води (л)"} \\ 1 & 1 & 1 & .9 & .65 & .33 & .11 & .05 & .01 & .0 & \text{"1-відро легке, 0-відро важке"} \end{pmatrix}.$$

Функція належності тут має інший топологічний характер. Щоб менше думати і використати попереднє, штучно робимо дані симетричними і на них визначаємо параметри, потім "підправляємо" функцію належності. Ніхто не забороняє Вам зробити це краще.

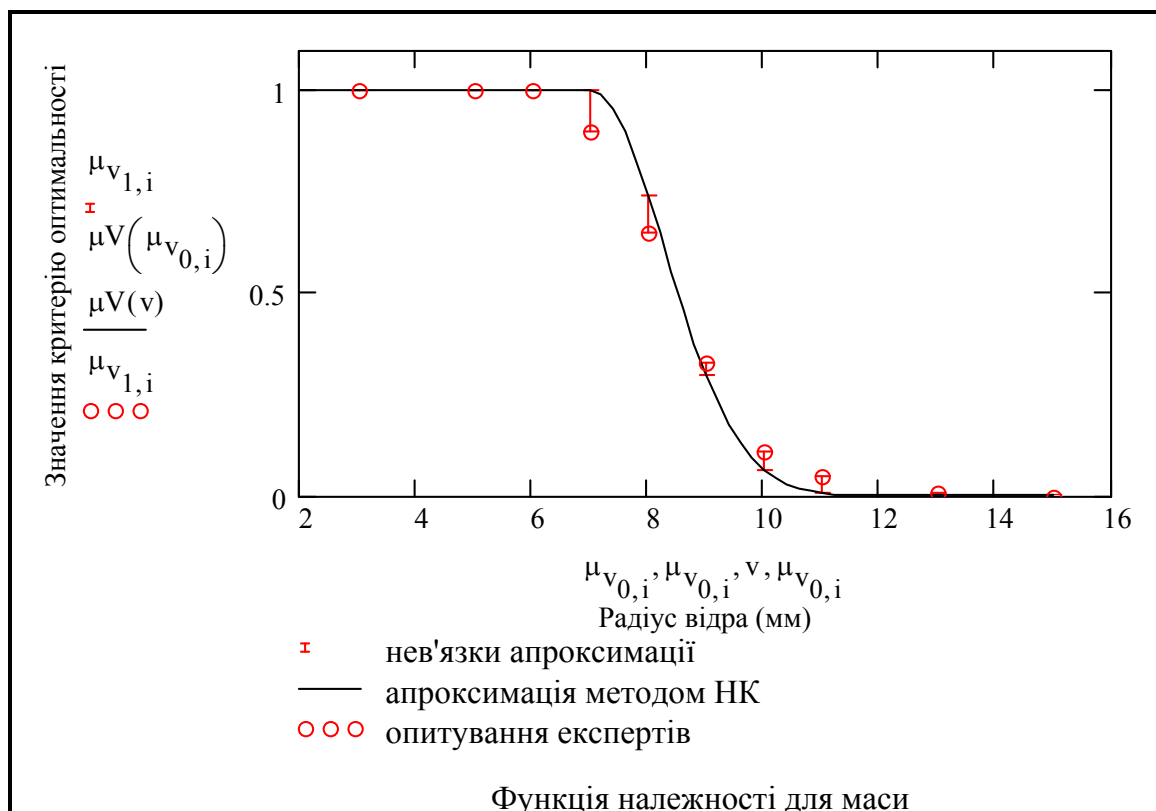
$$\mu_{1v} := \begin{pmatrix} 3 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 13 & 15 & \text{"Об'єм води (л)"} \\ .11 & .33 & .65 & .9 & .65 & .33 & .11 & .05 & .01 & 0 & \text{"симетр.для узгодження"} \end{pmatrix}$$

$$n := \text{cols}(\mu_v) - 1 ; n = 10 ; A := -0.1 ; B := 5$$

$$\text{Given } \sum_{i=0}^{n-1} (\mu_{1v_{1,i}} - \mu(\mu_{1v_{0,i}}, A, B))^2 = 0 ; \begin{pmatrix} A_v \\ B_v \end{pmatrix} := \text{MinErr}(A, B)$$

$$\mu V(v) := \text{if}(v < B_v, 1, \mu(v, A_v, B_v)) ;$$

$$A_v = -0.31 ; B_v = 7.01 ; i := 0..n-1 ; v := 2, 2.2..15$$



4. Конструювання двовимірної функції належності для радіуса і висоти

Згідно з методологією нечіткої оптимізації для формування загальної функції належності КОМБІНУЄМО усі функції належності. Ця процедура не повністю формалізована. Тут треба врахувати змістовний характер задачі, специфіку.

В даному випадку використовуємо операцію "**мінімальний із**". Тобто, глобальна функція належності в кожній точці простору параметрів приймає менше з трьох функцій належності значення.

Може виникнути питання звідки взялася формула для об'єму відра - це ж формула для об'єму конуса. Третя змінна оптимізації - об'єм, повністю визначається іншими змінними - висотою та радіусом

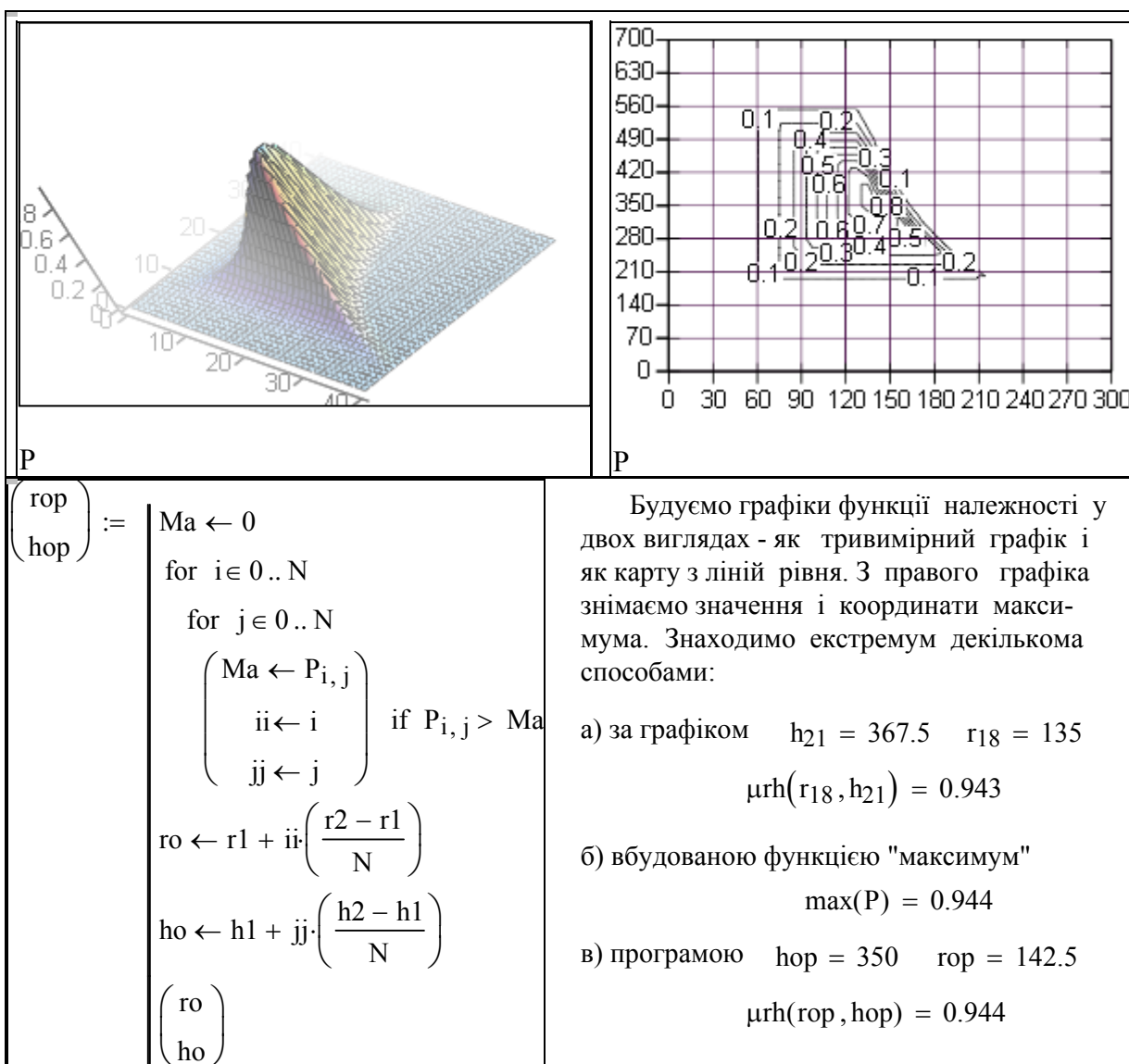
$$\mu_{rh}(r, h) := \min \left[\begin{array}{l} \mu_r(r) \\ \mu_h(h) \\ \mu_V \left[v \leftarrow \frac{\left(\frac{\pi}{3} \cdot r^2 \cdot h \right)}{100^2 \cdot 100} \right] \end{array} \right]$$

Для побудови тривимірних графіків функція двох змінних задається матрицею (не обов'язково). Вводимо індекси і визначаємо відповідні ранжовані змінні.

$N := 40$; $i := 0..N$; $j := 0..N$; $r1 := 0$; $r2 := 300$; $h1 := 0$; $h2 := 700$;

$$r_i := r1 + \frac{r2 - r1}{N} \cdot i; h_j := h1 + \frac{h2 - h1}{N} \cdot j; \quad P_{i,j} := \mu_{rh}(r_i, h_j) \quad .$$

Формуємо блок "програма розрахунку" + результати розрахунку". Нижче подано копію екрану з електронної книги.



Висновки

Знайдено оптимальне відро. Науковим методом - за допомогою експертного оцінювання і нечіткої логіки. Зрозуміло, що задача не є найважливішою для народного господарства. Але саме на таких задачах можна навчитися вільно, творчо, а головне, - критично застосувати знання у народному господарстві. На всяк випадок, нагадаємо, що вкрай нерозумно використовувати метод експертних оцінок для оцінки майбутнього. Експертне оцінювання слід застосовувати тільки в чітких стандартних ситуаціях, наприклад: дегустація чаю, конкурси краси.

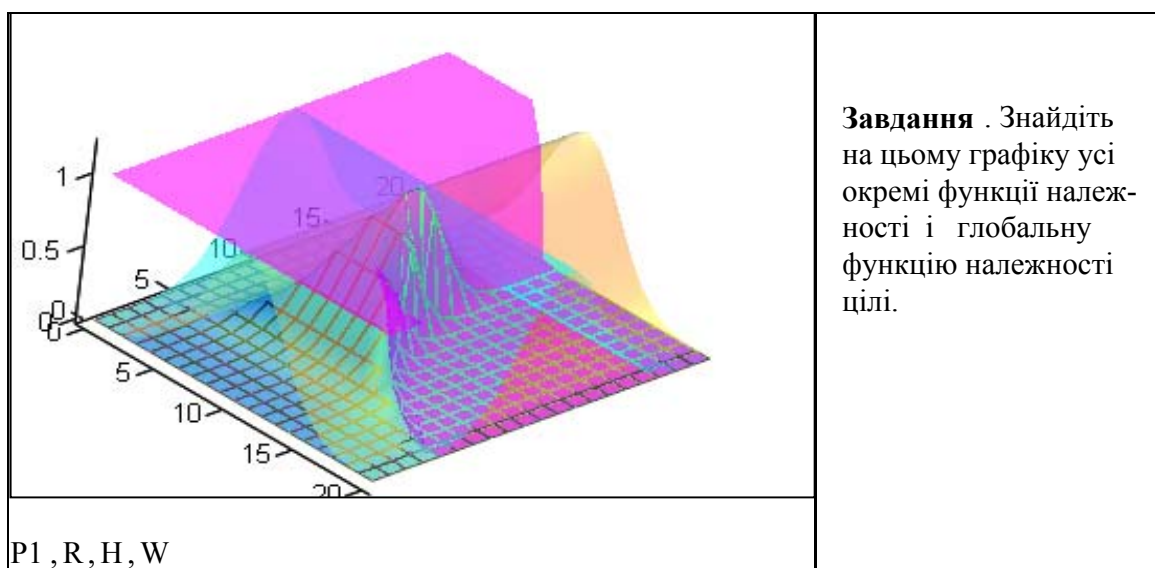
Зробимо останній погляд на функцію належності. А саме, подивимось, як вона формується з окремих функцій належності.

$$N := 20; i := 0..N; j := 0..N; r1 := 0; r2 := 300; h1 := 0; h2 := 700;$$

$$r_i := r1 + \frac{r2 - r1}{N} \cdot i; h_j := h1 + \frac{h2 - h1}{N} \cdot j; R_{i,j} := \mu_r(r_i); H_{i,j} := \mu_h(h_j); P_{i,j} := \mu_{rh}(r_i, h_j);$$

$$W_{i,j} := \mu V \left[v \leftarrow \frac{\left[\frac{\pi}{3} \cdot (r_i)^2 \cdot h_j \right]}{100^2 \cdot 100} \right].$$

Між іншим, придивіться до останнього виразу - в дужках замість константи чи змінної стоїть присвоєння. Програмування таке дозволяє.



Контрольні запитання

1. Проаналізуйте поняття "нечітка ймовірність", "нечітка надійність", "систематична система..."- що у них загального.
2. Джерела та інтерпретації нечіткості.
3. Опишіть процедуру апроксимації експертних оцінок певними функціями. Яка оптимізаційна задача тут розв'язується.
4. Класифікація ситуацій де можна/неможна застосовувати експертні оцінки.
5. Постановка задачі оптимізації функції N змінних при чітких обмеженнях.
6. Постановка задачі оптимізації функції N змінних при нечітких обмеженнях.



7.2. Оптимізація надійності систем альтернативними методами

Вступ. Формулювання цілей

Цей розділ є підсумковим не тільки для теми оптимізації та прийняття рішень на базі **розмитої логіки**, але і для курсу в цілому. В цьому розділі ми на прикладі досить типової задачі повторимо **методи оптимізації та моделювання**.

Природно, що ми поставимо за мету не розв'язання якоїсь окремої задачі, навіть альтернативними методами, а створення діалогового документа для розв'язання певного класу задач.

Складні системи і процеси, якими займаються кібернетика та дослідження операцій, не можуть бути чітко визначені. Складні системи можна визначити тільки через серію розгорнутих прикладів та "розмов з приводу". Стратегічна ціль цього розділу - дати приклад методології системних наук, до яких відносяться кібернетика і дослідження операцій, тактична - створити ресурс для виконання курсових робіт.

Постановка задачі

Розглядається система з трьох (для прикладу) **послідовно** поєднаних підсистем. Ефективність системи залежить від ефективностей підсистем. Кожна підсистема може мати в своєму складі 1, 2, 3...елементів, що працюють **паралельно**. Ефективність підсистеми залежить від ефективностей її елементів. Природно припустити, що у випадку непрацездатності усіх елементів підсистеми і підсистема, і система в цілому будуть непрацездатними. Сумарна вартість системи обмежується.

Потрібно вибрати кількість елементів в кожній підсистемі так, щоб максимізувати ефективність системи в цілому. Можлива зворотна задача - мінімізація вартості при обмеженні ефективності.

Розглянемо можливі інтерпретації задачі.

1. Багатофазна система масового обслуговування, чи виробнича система. Потрібно в рамках бюджету вибрати кількості "верстатів" на кожній фазі обслуговування чи виробництва так, щоб ефективність (це може бути надійність обслуговування, кількість і якість продукції) була максимальною.

2. Технічна система з трьох послідовно поєднаних підсистем. Потрібно в межах допустимої вартості вибрати кратність резервування для кожної підсистеми так, щоб надійність системи була максимальною. В умовах високих технологій та інтелектуалізації технічних систем теорія, методи істинної надійності все далі відходять від цієї найважчої задачі. **Надійність - не імовірнісна, а детермінована категорія, надійність завжди була не вправою з статистики, а пошуком конкретних конструкторських розв'язань.**

Зрозуміло, що методи імовірнісної теорії надійності мають досить застосувань, наприклад, для оптимізації надійності сірників.

Але є традиція - розглядати важливі задачі на псевдореалістичних злегка дурнуваних прикладах - це настраює на конструктивно-критичне відношення до науки.

Математична модель надійності і вартості резервованої системи

Подаємо **робочу математичну модель** "надійності" і вартості резервованої системи. Кожен тип елементів системи характеризується показниками надійності **г_і** (ймовірність працездатності на заданому періоді) і вартості **с_і**

$$\text{ORIGIN} := 1; \quad i := 1..3; \quad r_1 := 0.8; \quad c_1 := 1; \quad r_2 := 0.65; \quad c_2 := 2; \quad r_3 := 0.7; \quad c_3 := 3.$$

Задається обмеження сумарної вартості системи: $C_0 := 15$.

Записуємо у матрично-векторному вигляді вхідні дані і виконуємо **контроль** даних $q_i := 1 - r_i$ - ймовірність відмови на заданому періоді;

$$\text{"надійність"} \quad r = \begin{pmatrix} 0.8 \\ 0.65 \\ 0.7 \end{pmatrix}; \quad \text{"ненадійність"} \quad q = \begin{pmatrix} 0.2 \\ 0.35 \\ 0.3 \end{pmatrix}; \quad \text{повна група подій} \quad (r + q) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Записуємо математичну модель надійності резервованої системи із заданою конфігурацією (послідовне поєднання ідентично резервованих елементів) у вигляді функції користувача, що потім буде використана у програмі.

$$R(x_1, x_2, x_3, q_1, q_2, q_3) := [1 - (q_1)^{x_1+1}] \cdot [1 - (q_2)^{x_2+1}] \cdot [1 - (q_3)^{x_3+1}]$$

Ми назвали задачу оптимізації резервування не зовсім розумною. Вибрана нами модель надійності базується на припущенні про незалежність відмов елементів. В житті, зазвичай, елементи виходять з ладу "колективно": вони можуть бути з однієї дефектної партії, на них може подіяти одне й те ж збурення, елемент, що "згорів", може "підпалити" інші.

Записуємо математичну модель вартості резервованої системи теж у вигляді функції користувача

$$C(x_1, x_2, x_3, c_1, c_2, c_3) := c_1 \cdot (x_1 + 1) + c_2 \cdot (x_2 + 1) + c_3 \cdot (x_3 + 1)$$

і визначаємо частку вартості системи, яка може бути використана на резервування в рамках обмеження сумарної вартості: $CR := C_0 - c_1 - c_2 - c_3$.

Постановка оптимізаційної задачі. Обґрунтування вибору методу оптимізації

Постановка задачі: Для системи з трьох послідовно поєднаних підсистем із заданими вартостями і надійностями, **знайти кратність резервування** кожної підсистеми, що забезпечує максимальну надійність **при обмеженні сумарної вартості** системи.

Вибір методу оптимізації. Вибираємо метод узгоджений з математичними властивостями задачі та інженерними потребами і сучасними принципами конструювання програм, а саме:

1. Програма повинна розв'язувати досить широкий клас задач, мати інтерфейс в межах екранної сторінки, дозволяючи вводити вхідні змінні і отримувати виходи у зручній формі.

2. В даній задачі цільова функція (надійність) монотонно залежить від ресурсу (кратності резервування), тому саме на підмножині варіантів з граничною вартістю (з урахуванням дискретності) знаходиться точка екстремуму.

Альтернативи методів оптимізації. З великої кількості відомих методів оптимізації при обмеженнях вибираємо:

- 1) метод динамічного програмування,
- 2) метод невизначених множників Лагранжа;
- 3) метод оптимізації на базі нечіткої логіки.

Метод динамічного програмування. Записуємо базове функціональне рівняння Беллмана для дискретної задачі з адитивним критерієм

$$f(N, p_1) = \max(g(p_1, q_1) + f(N - 1, T(p_1, q_1))),$$

де управління $q_1 \in QQ$ береться з області допустимих управлінь.

У цьому рівнянні $f(N, p_1)$ - шукана функція, що дає залежність **оптимального значення** критеріальної функції від **початкового стану p_1** і **кількості кроків процесу N** . $g(p_i, q_i)$ - функція доходу, що залежить від p_i - поточного стану, і q_i - "управління", "стратегія", "політика" - тобто те, що ми можемо вибирати. Для нашої задачі оптимізації "управлінням" є вибір кратності резервування i -го елемента. Р.Беллман вважав динамічне програмування (ДП) не методом а "станом розуму". Знайти найкраще подання деякої задачі у термінах ДП - це штука, винахідницька задача.

Подаємо в стислому формулюванні логіку моделі процесу оптимізації та програми, що реалізує цей процес для нашого задачі: **упорядковано, починаючи з певного "кута" відбираємо граничні варіанти, одночасно на них виділяємо оптимальний. В кінці роботи програми маємо масив граничних варіантів, та координату і показник оптимального варіанта.** Далі подано реалізацію цього методу.

Обґрунтування вибору програмного середовища і технології розробки системи

Сучасні технології розробки прикладних програм, незважаючи на зовнішню різноманітність, побудовані за єдиною схемою:

- 1) формулювання списку функцій системи (специфікація);
- 2) ручна або автоматична трансляція специфікації у діючий макет програми;
- 3) повторення п.2 - розробка альтернативних версій, відлагодження, випробування та взаємоперевірка версій;
- 4) ручна або автоматична трансляція макета програми у код (.EXE- модуль).

Ідеальне програмне середовище - таке, що має максимум готових модулів, зручні засоби збирання програми з модулів, і одночасно дозволяє користувачу створювати власні модулі.

Математичні моделі економічної ефективності систем з резервуванням

Взагалі, чи завжди потрібна надійність? Почнемо з лінгвістичної моделі - сценарію виробництва, продажу і експлуатації системи. Нехай деяка система (чайник, наприклад) випускається масово, відмова системи ніяких втрат окрім втрати самої системи не викликає.

Для масових недорогих систем можемо застосувати методи статистичної теорії надійності: протягом року вийде з ладу $P \cdot M$ систем, де P - імовірність виходу з ладу системи протягом року, M - річний обсяг виробництва (і продаж). Додаткові сумарні витрати

суспільства (користувача, якщо купляє не там; виробника, якщо дається річна гарантія) дорівнюють: $P \cdot M \cdot C_0$, де C_0 - вартість існуючої системи.

Нехай проведено деякі дослідження, капіталовкладення в технології з постійними витратами $C_{\text{пост}}$, введено резервування елементів системи з витратами dC і в результаті надійність підвищилась на dP . В підсумку суспільство (споживачі, виробники) має економію:

$$P \cdot M \cdot C_0 - (P - dP) \cdot M \cdot (C + dC + K_a \cdot C_{\text{пост}}) > 0 ?$$

Перший крок аналізу ефективності - перевірка позитивності ефекту, потім на базі даної моделі виконується більш детальний аналіз. Слід зауважити, що результатом досліджень може бути повна зміна конструкції і технології системи з одночасним зниженням вартості і підвищенням надійності.

Порівняльний аналіз сучасних методів забезпечення надійності

В традиційній теорії надійності підвищення надійності монотонно залежить від вартості системи. Для високих технологій ця залежність може бути зворотною - **чим дешевша система, тим вона надійніша і якісніша.**

Але це потребує пошуку структурних, а не параметричних змін в системі, наприклад:

☞ замінити технологію виготовлення друкованої плати, кристалу на нову, бездефектну;

☞ замінити три ненадійних друкованих плати, на 12 монокристальних систем з такою ж вартістю і абсолютною надійністю, об'єднати їх у відмовостійку конфігурацію, а додатково отримати "премію" - закрити екологічно шкідливу ділянку травлення друкованих плат, припинити розробку "інтелектуальних систем" для оптимізації процесу травлення.

☞ Можна зробити систему з автоматичною реконфігурацією (сучасні сервери), з автоматичною діагностикою і компенсацією люфтів і дрейфу параметрів (сучасні принтери і верстати).

☞ Можна ввести неідентичне резервування (сучасні великі системи і спецсистеми), організувати інформаційні банки (сучасні інформаційно-вимірювальні системи). Відповідні напрямки називаються - гарантоспроможність, відмовостійкість, живучість, структурна надійність.

☞ Останній, простіший засіб у боротьбі за надійність і якість - ідентичне резервування. Єдина проблема тут - розрахунок оптимального резервування елементів системи. Тимчасові труднощі в цій задачі - дискретність та розмірність.

Розробка блок-схеми роботи користувача в автоматизованій системі

Для проектування інтерфейсу програми необхідно:

- визначити задачу, що буде розв'язувати користувач;
- вибрати оптимальну послідовність кроків;
- скласти блок-схему роботи користувача.

А ще необхідно пам'ятати, що в сучасних програмних системах, починаючи з 1С-бухгалтерії, сам користувач настраює собі інтерфейс. В даному документі можна змінювати все, що не закрито паролем від змін і читання.

Розробка алгоритму і програми для оптимізації конфігурації системи

Далі подано коментовану блок-схему - текст програми оптимізації (це копія екрану з електронної книги).

$Cr \leftarrow Cs - c1 - c2 - c3$ break if $Cr < 0$ $x1m \leftarrow \text{floor}\left(\frac{Cr}{c1}\right)$ $x2m \leftarrow \text{floor}\left(\frac{Cr}{c2}\right)$ $x3m \leftarrow \text{floor}\left(\frac{Cr}{c3}\right)$ $n \leftarrow 0$ $Rmax \leftarrow 0$ $x3 \leftarrow 0$ $x2 \leftarrow 0$ $x11 \leftarrow x1m$ $u \leftarrow 0$ while $x11 > 0$ $u \leftarrow u + 1$ $x11 \leftarrow x1m - \text{ceil}\left(\frac{c3}{c1}\right) \cdot (u - 1)$ $n \leftarrow n + 1$ $x3 \leftarrow u - 1$ $Ma^{(n)} \leftarrow \begin{pmatrix} x11 \\ x2 \\ x3 \end{pmatrix}$ if $Rmax < R(x11, x2, x3, q_1, q_2, q_3)$ $Rmax \leftarrow R(x11, x2, x3, q_1, q_2, q_3)$ $nmax \leftarrow n$ $x1 \leftarrow x11$ while $x1 > 0$ $x1 \leftarrow x1 - 1$ $dc1 \leftarrow Cs - C(x1, x2, x3, c1, c2, c3)$ $vx2 \leftarrow \text{floor}\left(\frac{dc1}{c2}\right)$ if $vx2 > 0$ $n \leftarrow n + 1$ $Ma^{(n)} \leftarrow \begin{pmatrix} x1 \\ x2 + vx2 \\ x3 \end{pmatrix}$ if $Rmax < R(x1, x2 + vx2, x3, q_1, q_2, q_3)$
--

			"..."	
			$R_{\max} \leftarrow R(x_1, x_2 + vx_2, x_3, q_1, q_2, q_3)$	- перевірка на екстремум.
			$n_{\max} \leftarrow n$	
		$\begin{pmatrix} Ma \\ R_{\max} \\ n_{\max} \end{pmatrix}$		- те, що повертає програма: масив варіантів, максимальна надійність та номер оптимальної конфігурації.

Розробка схеми і програми аналізу даних оптимізації. Відлагодження і тестування програми

Виконуємо ці пункти разом, згідно з тенденціями інтеграції етапів розробки програми. Те, що подано вище - гібрид лістингу, схеми, може ще й алгоритму. Проводимо тести програми оптимізації. Бачимо, що для критичного випадку $C_s=6$, при $c_1 + c_2 + c_3 = 6$ - коли немає ресурсу для резервування - програма повертає нулі. Для $C_s=7$ вже повертається масив з 2-ох варіантів. Для обмеження $c=133$ надійність=1, число переглянутих варіантів - 2753. Обчислюється ці варіанти за 5-10 сек. роботи комп'ютера

$$Xr(6, c_1, c_2, c_3) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad Xr(7, c_1, c_2, c_3) = \begin{pmatrix} \{3, 2\} \\ 0.437 \\ 1 \end{pmatrix} \quad C_s = 15 \quad c = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad r = \begin{pmatrix} 0.8 \\ 0.65 \\ 0.7 \end{pmatrix}$$

$$Xr(C_s, c_1, c_2, c_3) = \begin{pmatrix} \{3, 19\} \\ 0.864 \\ 13 \end{pmatrix} \quad Xr(33, 1, 2, 3) = \begin{pmatrix} \{3, 136\} \\ 0.996 \\ 100 \end{pmatrix} \quad Xr(133, 1, 2, 3) = \begin{pmatrix} \{3, 2753\} \\ 1 \\ 2085 \end{pmatrix}$$

Наша програма оформлена, як процедура, від параметру C_x - вартості системи. Програма повертає певну структуру даних. Розпакуємо тепер видачу програми.

$Rm(C_x) := Xr(C_x, c_1, c_2, c_3)_2$ - надійність оптимального варіанта;

$Nm(C_x) := Xr(C_x, c_1, c_2, c_3)_3$ - номер оптимального варіанта;

$xR(C_x) := Xr(C_x, c_1, c_2, c_3)_1$ - масив граничних (задовільних) варіантів;

$xop(C_x) := xR(C_x)^{Nm(C_x)}$ - параметри оптимального варіанта.

Виводимо результати оптимізації для заданих даних (=контрольного прикладу). Маємо: задані надійності, ціни, оптимальне резервування та надійність

$$q = \begin{pmatrix} 0.2 \\ 0.35 \\ 0.3 \end{pmatrix}; \quad c = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}; \quad xop(15) = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}; \quad Rm(C_c) = 0.864; \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} := xop(15); \quad \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} := q; \quad \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} := c$$

Виконаємо контроль результатів. Перевіряємо усе, що було написано і використане, - підставляємо дані у праві, ліві частини визначень функцій, процедур

$$R(x_1, x_2, x_3, q_1, q_2, q_3) := [1 - (q_1)^{x_1+1}] \cdot [1 - (q_2)^{x_2+1}] \cdot [1 - (q_3)^{x_3+1}];$$

$$C(x_1, x_2, x_3, c_1, c_2, c_3) := c_1 \cdot (x_1 + 1) + c_2 \cdot (x_2 + 1) + c_3 \cdot (x_3 + 1).$$

Бачимо, немовби все збігає

$$R(x_1, x_2, x_3, q_1, q_2, q_3) = 0.864 \quad c_1 \cdot (x_1 + 1) + c_2 \cdot (x_2 + 1) + c_3 \cdot (x_3 + 1) = 15$$

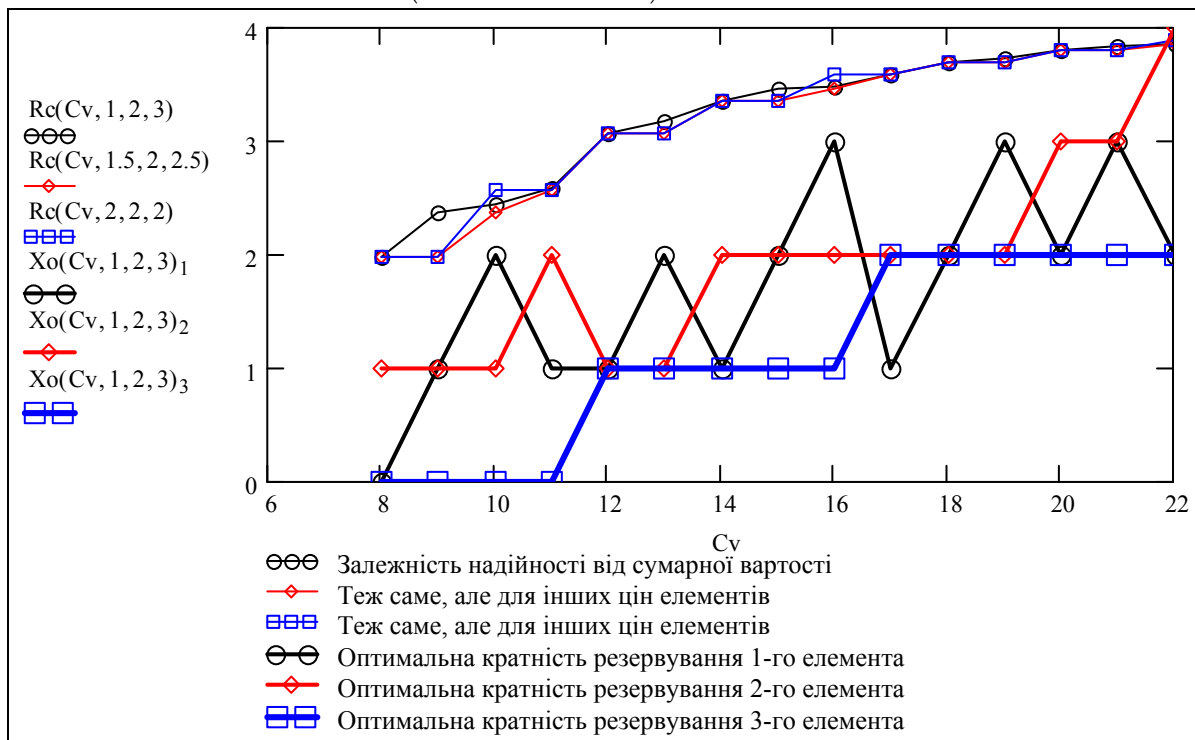
$$\left[1 - (q_1)^{x_1+1}\right] \cdot \left[1 - (q_2)^{x_2+1}\right] \cdot \left[1 - (q_3)^{x_3+1}\right] = 0.864 \quad C(x_1, x_2, x_3, c_1, c_2, c_3) = 15$$

Розробка інтерфейсу для аналізу ефективності резервування

З множини допустимих варіантів резервування перший модуль програми відібрав підмножину граничних, відносно обмеження варіантів, тобто варіантів, вартість яких найближча до граничної. Формуємо масив даних для проведення аналізу.

$X_o(C_x, c_1, c_2, c_3) := \left(X_r(C_x, c_1, c_2, c_3)_1 \right)^{\langle X_r(C_x, c_1, c_2, c_3) \rangle_3}$ → Вектор-функція кратностей резервування

$R_c(C_x, c_1, c_2, c_3) := \left(X_r(C_x, c_1, c_2, c_3)_2 \right) \cdot 4$ → Функція надійності системи

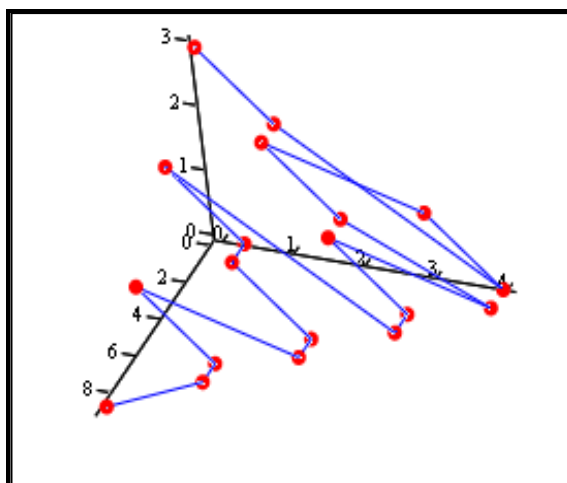


Дивлячись на графіки, бачимо, що досить точні розв'язання можна отримати, якщо задачу розглядати як неперервну. Тобто, бачимо загальні властивості задачі і підозрюємо, як Віні Пух (прочитайте першу сторінку бестселера), що є інші способи роботи з цією задачею.

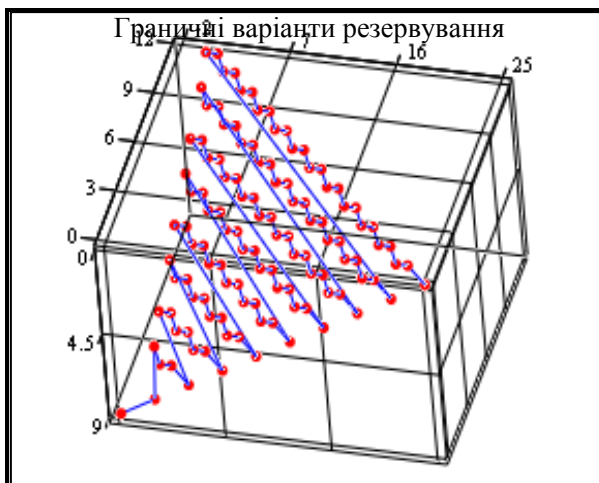
Маємо інтерфейс, що не тільки дає якісь числа (= "розв'язок"), а дозволяє глянути на задачу "згори", все проаналізувати (= розв'язання). Зробимо ще крок у цьому напрямку - "витягнемо" та побудуємо поверхню - границю допустимих варіантів резервування.

Записуємо потрібні вирази, будуємо графіки.

$$X1(C_x) := \left(xR(C_x)^T \right)^{\langle 1 \rangle}; \quad X2(C_x) := \left(xR(C_x)^T \right)^{\langle 2 \rangle}; \quad X3(C_x) := \left(xR(C_x)^T \right)^{\langle 3 \rangle}.$$



(X1(15), X2(15), X3(15))



(X1(31), X2(31), X3(31))

Ці графіки побудовано для двох значень обмежень сумарної вартості: 15 та 31 одиниця.

Контроль програми. Програма досить складна і може мати непомітний дефект - пропускати оптимум. Перевіримо її використавши альтернативний метод знаходження оптимального варіанта. Варіант з максимальною надійністю можна знайти, використовуючи вбудовану функцію пакета для сортування стовпців матриці X (див. вище), по значенню останнього - 5-го рядка - значенню надійності. Це робиться так: Додаємо до матриці варіантів резервування вектор-рядок надійності цих варіантів і сортуємо по значенню останнього рядка, виводимо таблицю, де варіанти ранжовані за надійністю.

$$\text{cols}(xR(15)) = 19 \quad u := 1.. \text{cols}(xR(15)) \quad R_{zu} := R(xR(15)_{1,u}, xR(15)_{2,u}, xR(15)_{3,u}, q_1, q_2, q_3) \quad R_z := R_z^T$$

$$Mx(C) := \text{stack}(xR(C), R_z)$$

$$Mxsort := \text{rsort}(Mx(15), 4)$$

Mxsort =		10	11	12	13	14	15	16	17	18	19
	1	5	0	2	3	0	3	4	1	1	2
	2	2	1	3	3	3	1	1	1	2	2
	3	0	2	0	0	1	1	1	2	1	1
	4	0.67	0.683	0.684	0.688	0.717	0.797	0.798	0.82	0.836	0.864

Контроль
(пересунь
движок)

$$Rm(15) = 0.864$$

Достоїнства такого способу - простота і швидкість, а також те, що ми маємо не одну точку екстремуму, а можемо бачити близькі до оптимального варіанти резервування.

Недоліки - простота, неможливість побудувати залежність оптимальної надійності від рівня обмеження чи цін елементів.

Напрямки модифікації програми для захисту від помилок вхідних даних

Можливі напрямки модифікації програми для захисту від помилок вхідних даних:

- 1) перевірка умов допустимості усіх вхідних величин у програмі,
- 2) альтернативні методи обчислення,
- 3) графічне виведення.

Частина цих напрямків вже реалізована при розробці і відлагодженні програми.

Розробка програми оптимізації для випадку неперервних змінних

Пам'ятаючи головну тезу даного посібника - найкраща альтернатива - це (по можливості) сума альтернатив, розв'яжемо оптимізаційну задачу як неперервну. Тобто допустимо, щоб кратності резервування x_1 , x_2 , x_3 приймали не цілочислові значення. Отримані розв'язання повинні бути близькими до цілочислових розв'язань, їх можна використати як перші наближення і для контролю. Згадаємо головну інтерпретацію задачі - оптимальний розподіл ресурсу по етапах процесу при мультиплікативному критерії - де x_1 , x_2 , x_3 – не цілочислові. Використовуємо два альтернативних методи оптимізації:

- метод невизначених множників Лагранжа і
- метод оптимізації на базі розмитої логіки.

Метод невизначених множників Лагранжа

Маємо **цільову функцію** $R(x_1, x_2, x_3, q_1, q_2, q_3)$ - надійність системи і **обмеження** $C(x_1, x_2, x_3, c_1, c_2, c_3) = 0$ на вартість системи. Записуємо **функцію Лагранжа**:

$$L_1(x_1, x_2, x_3, q_1, q_2, q_3, \lambda) := R(x_1, x_2, x_3, q_1, q_2, q_3) + \lambda \cdot C(x_1, x_2, x_3, c_1, c_2, c_3)$$

Необхідна умова екстремуму - похідні від функції Лагранжа по усіх змінних оптимізації повинні бути нульовими. В даному випадку можна знайти аналітичні вирази для похідних і навіть розв'язання в символьному вигляді, але ми використаємо вбудовані числові методи пакета - аналітичні результати існують не завжди.

Задаємо початкові значення змінних: $x_1 := 1$; $x_2 := 2$; $x_3 := 3$; $\lambda := 1$; $q := 1..3$.

Записуємо систему рівнянь (це копія екрану) - три похідних і обмеження вартості системи:

Given

$\frac{d}{dx_1} L_1(x_1, x_2, x_3, q_1, q_2, q_3, \lambda) = 0$	$x_1 \geq 0$	}
	$x_2 \geq 0$	
$\frac{d}{dx_2} L_1(x_1, x_2, x_3, q_1, q_2, q_3, \lambda) = 0$	$x_3 \geq 0$	
$\frac{d}{dx_3} L_1(x_1, x_2, x_3, q_1, q_2, q_3, \lambda) = 0$		
$c_1 \cdot (x_1 + 1) + c_2 \cdot (x_2 + 1) + c_3 \cdot (x_3 + 1) - C_0 = 0$		

і отримуємо розв'язання: $\text{RezLagr} := \text{Find}(x_1, x_2, x_3, \lambda)$

RezLagr =

	1
1	2
2	2
3	1
4	-0

Mxsort_{q, 19}

	1
1	2
2	2
3	1

Результат ми навмисне округлили до цілого, щоб краще бачити збігання з результатом дискретної оптимізації.

Зауваження. Вбудовані в Mathcad числові методи оптимізації та розв'язання рівнянь не є всемогутніми. Іноді функція `find(.)` не працює (оптимізована функція має незручну структуру, наприклад, "гірський хребет"). Функція `minerr(.)` дає розв'язок досить близький до дискретного.

Оптимізація на базі нечіткої логіки

Теорія нечіткості була запропонована для моделювання задач з розмитими, нечітко визначеними границями, умовами і цілями. Ефективним є застосування цієї теорії для розв'язання задач з обмеженнями. Навіть там, де все чітко визначено, введення штучної нечіткості підвищує ефективність числових методів пошуку екстремуму.

Обмеження вартості системи природним чином є нечітким - завжди допускається деяке відхилення від заданої величини. Таким же природно нечітким є обмеження маси, наприклад, в авіакосмічній техніці конструкторів преміюють/штрафують за зменшення чи збільшення маси розроблюваної системи. Ще раз нагадаємо, що це ефективний напрям, тому, бережіться "сурогатів", наприклад, зустрічається вам робота, де є "нечітка імовірність". Так знайте, що імовірність і є видом нечіткості.

Запишемо цільову функцію і номінальне обмеження:

$$Rlb(x_1, x_2, x_3) := \left[1 - (q_1)^{x_1+1} \right] \cdot \left[1 - (q_2)^{x_2+1} \right] \cdot \left[1 - (q_3)^{x_3+1} \right]$$

$$Ogr(x_1, x_2, x_3) := c_1 \cdot (x_1 + 1) + c_2 \cdot (x_2 + 1) + c_3 \cdot (x_3 + 1) - Cog$$

Перетворюємо обмеження в нечітке (розмите, гумове...) - конструюємо **функцію належності** для обмеження. Для цього використовуємо вбудовану функцію нормального розподілу. Чому нормального? - Це не принципово, вирішальними є певні топологічні властивості функції, можна було взяти іншу функцію):

$$Fpog(x_1, x_2, x_3) := dnorm(Ogr(x_1, x_2, x_3), 0, fuznja)$$

Задаємо ранжовані змінні, діапазони і крок зміни x_1, x_2, x_3

$$QQ := 20; i := 1..QQ; j := 1..QQ; k := 1..QQ; h := .15; X1_i := h \cdot i; X2_j := h \cdot j; X3_k := h \cdot k; xf := 1.$$

Конструюємо **загальну функцію належності**, або **функцію належності цілі**. Для цього ми виконуємо над усіма нечіткими обмеженнями і цільовою функцією операцію "нечітка кон'юнкція". Ця операція в теорії нечітких множин визначається ситуативно, в залежності від "фізичного смислу" цільової функції і обмежень - розмірності та ін.

Розглядаємо дві альтернативи:

а) добуток (беремо декілька варіантів з різним рівнем нечіткості)

$$OPf1_{i,j} := Rlb(X1_i, X2_j, 1) \cdot Fpog(X1_i, X2_j, 1);$$

$$OPf2_{i,j} := Rlb(X1_i, X2_j, 1) \cdot Fpog(X1_i, X2_j, 1);$$

$$OPf3_{i,j} := Rlb(X1_i, X2_j, 3) \cdot Fpog(X1_i, X2_j, 3)$$

б) мінімум від функцій належності.

$$OPfo_{i,j} := \min(Rlb(X1_i, X2_j, 2), Fpog(X1_i, X2_j, 2) \cdot 4)$$

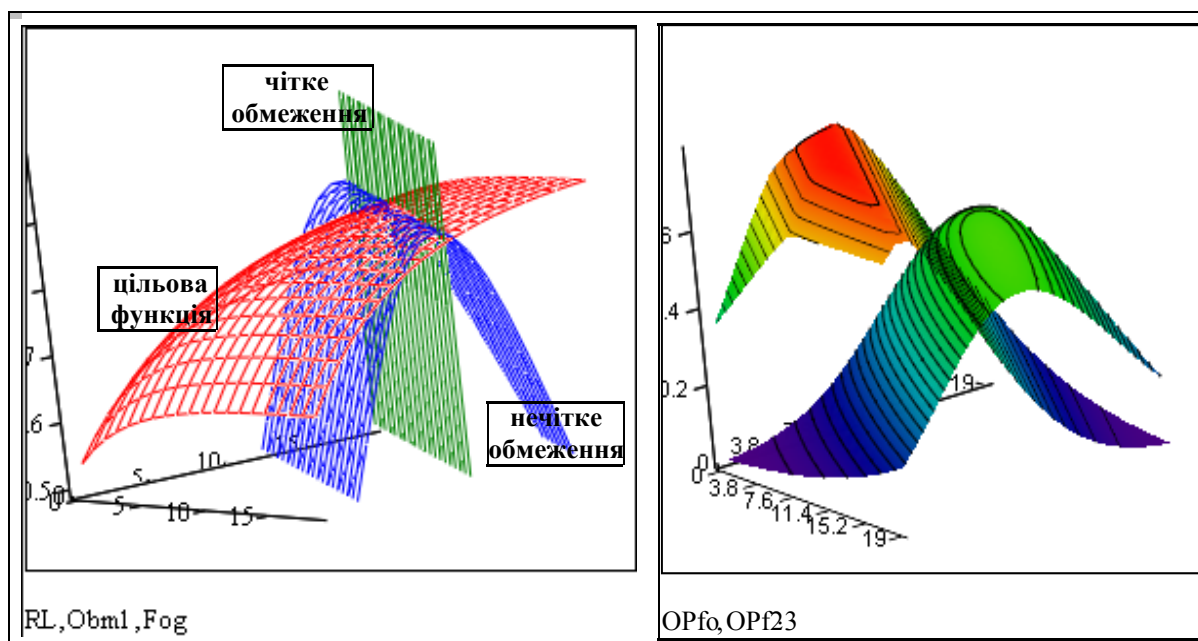
Задаємо, ще ряд функцій для побудови графіків

$$RL_{i,j} := Rlb(X1_i, X2_j, 1) \cdot 1; Fog_{i,j} := Fpog(X1_i, X2_j, 1) \cdot 4.35; Obm1_{i,j} := Ogr(X1_i, X2_j, 1) \cdot 5.$$

Для чого ми визначили три функції належності - $OPf1, OPf2, OPf3$? Важливе значення має можливість "побачити власними очима" об'єкти математичних операцій. Якби ми могли наочно бачити функції багатьох змінних у гіперпросторі, це б можливо дозволило зробити безліч відкриттів і точно вже не зробити безліч помилок. Наприклад, уявіть собі - майже весь об'єм гіперсфери у 10, 100 вимірів лежить біля її поверхні.

В даному випадку маємо функцію трьох змінних, що потребує чотиривимірного простору. Тому беремо три фіксованих значення змінної: $x_3 = 1$, що означає однократне резервування, та $x_3 = 2$, $x_3 = 3$. Далі побудовано три тривимірних графіки - три "зрізи" чотиривимірної поверхні.

На графіку ліворуч подано 1) цільову функцію, 2) чітке обмеження, 3) нечітке обмеження. Зверніть увагу - усі три поверхні проходять через точку екстремуму. На правому графіку - подано два варіанти конструювання загальної функції належності - через добуток і через мінімум. Функції мають екстремум приблизно в одній точці. Розглядати ці функції **дуже бажано на комп'ютері** - мишею можна повернути зображення в який завгодно ракурс, Ctrl+мишею змінювати розмір зображення та ін.



Побудували ми функцію належності, тепер виконуємо власне саму оптимізацію. Це вираз для функції належності: Це визначення максимуму цієї функції:

$$OPf(x_1, x_2, x_3) := Rlb(x_1, x_2, x_3) \cdot Fpog(x_1, x_2, x_3)$$

$$RezFuz := \text{Maximize}(OPf, x_1, x_2, x_3)$$

Максимум визначаємо вбудованим "Квазі-Ньютоновим методом". Методи вибираються через меню, що викликається правою кнопкою миші, курсор на ім'я функції.

Подаємо разом результати оптимізації за усіма альтернативними методами. Три кращих дискретних варіанти, варіант за методом Лагранжа і "нечіткий".

Mxsort ⁽¹⁷⁾ =	Mxsort ⁽¹⁸⁾ =	Mxsort ⁽¹⁹⁾ =	RezLagr =	RezFuz =																																																
<table><tr><td></td><td>1</td></tr><tr><td>1</td><td>1</td></tr><tr><td>2</td><td>1</td></tr><tr><td>3</td><td>2</td></tr><tr><td>4</td><td>0.82</td></tr></table>		1	1	1	2	1	3	2	4	0.82	<table><tr><td></td><td>1</td></tr><tr><td>1</td><td>1</td></tr><tr><td>2</td><td>2</td></tr><tr><td>3</td><td>1</td></tr><tr><td>4</td><td>0.836</td></tr></table>		1	1	1	2	2	3	1	4	0.836	<table><tr><td></td><td>1</td></tr><tr><td>1</td><td>2</td></tr><tr><td>2</td><td>2</td></tr><tr><td>3</td><td>1</td></tr><tr><td>4</td><td>0.864</td></tr></table>		1	1	2	2	2	3	1	4	0.864	<table><tr><td></td><td>1</td></tr><tr><td>1</td><td>1.5</td></tr><tr><td>2</td><td>1.8</td></tr><tr><td>3</td><td>1.3</td></tr><tr><td>4</td><td>-0</td></tr></table>		1	1	1.5	2	1.8	3	1.3	4	-0	<table><tr><td></td><td>1</td></tr><tr><td>1</td><td>1.5</td></tr><tr><td>2</td><td>1.9</td></tr><tr><td>3</td><td>1.3</td></tr></table>		1	1	1.5	2	1.9	3	1.3
	1																																																			
1	1																																																			
2	1																																																			
3	2																																																			
4	0.82																																																			
	1																																																			
1	1																																																			
2	2																																																			
3	1																																																			
4	0.836																																																			
	1																																																			
1	2																																																			
2	2																																																			
3	1																																																			
4	0.864																																																			
	1																																																			
1	1.5																																																			
2	1.8																																																			
3	1.3																																																			
4	-0																																																			
	1																																																			
1	1.5																																																			
2	1.9																																																			
3	1.3																																																			

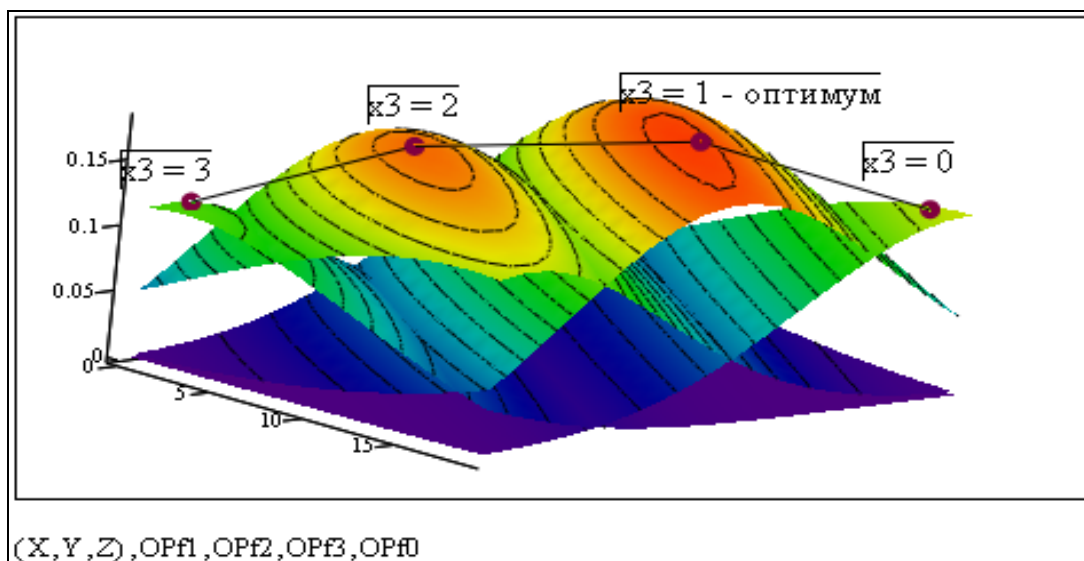
Розмитість: **fuznja = 2**

Обмеження вартості системи **Cog = 15**

Спробуйте змінювати значення розмитості та обмеження вартості

На графіку внизу подано три "зрізи" чотиривимірної функції належності - для таких значень третьої змінної $x_3 = 0$, 1 (оптимальне резервування), 2, 3. На цьому ж

графіку подано положення точок екстремуму. Бачимо, що дійсно найбільше значення функція має при $x_3 = 1$. Природно, що із збільшенням x_3 , в силу обмеження за ресурсом, зменшуються значення x_1 та x_2 - точка максимуму наближується до осі z .

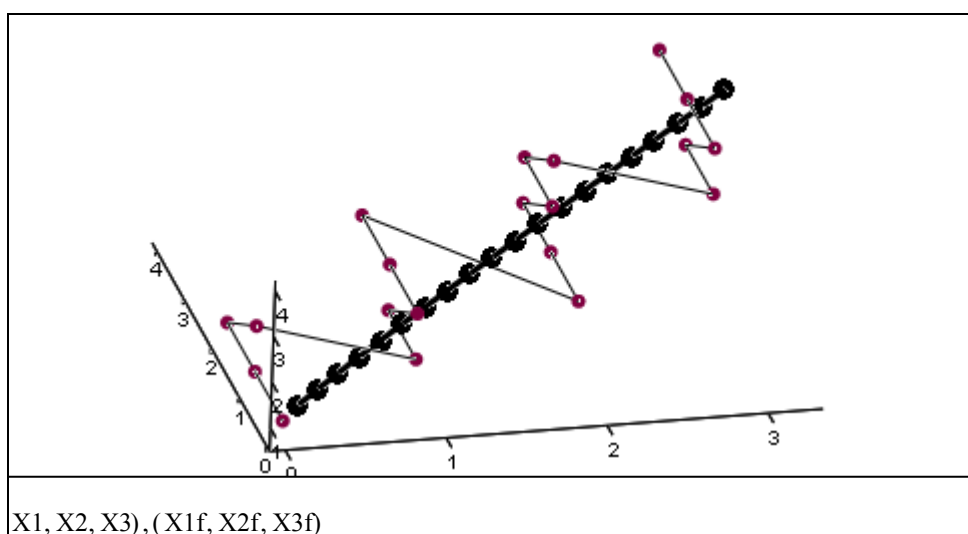


Чітко про нечіткість

Застосували ми, при розв'язанні наших задач, без перебільшення, десь 0.1% можливостей і результатів розмитості логіки. Кинемо ще один, нечіткий, погляд на дискретне розв'язання задачі оптимізації надійності. Дещо переробимо функцію належності. Побудуємо в просторі конфігурацій системи (x_1, x_2, x_3) залежності оптимальної конфігурації від рівня обмеження за ресурсом - змінюємо обмеження 8 до 20.

$$q := 1 \dots 20 ; X1_q := Xo(q + 7, 1, 2, 3)_1 ; X2_q := Xo(q + 7, 1, 2, 3)_2 ; X3_q := Xo(q + 7, 1, 2, 3)_3 ;$$

$$X1f_q := Rzfz(q + 7)_1 ; X2f_q := Rzfz(q + 7)_2 ; X3f_q := Rzfz(q + 7)_3 .$$

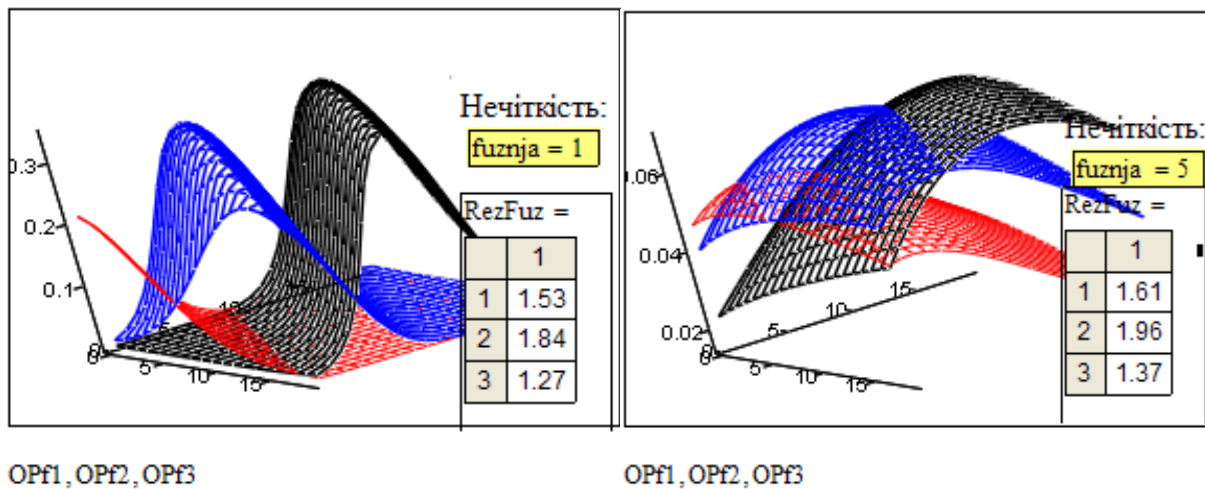


Бачимо ми, що розмите і неперервне розв'язання є гладким, інтуїтивно прогнозним (добавився ресурс – розподіляється він усім пропорційно їх ефективності. Проблема оптимізації – лише визначити цю пропорцію). Дискретні оптимальні конфігурації лише стрибають навколо "генеральної лінії", **чітко накресленої нечітким методом.**

Підсумки

Перше враження людини з нормальною психікою від нечітких методів – чиста алхімія, крім того, більш ніж підозріла термінологія, велика кількість інформаційного шуму. Попереджали ж Р. Беллман і Л. Заде – засновники цього напрямку, що все це занадто складне для осмислення.

Перевіряємо чи дійсно це дає стійкі результати, і не є штучно підібраним. Змінюємо параметр **функції належності** обмежень - *fuznja*, що характеризує ступінь "розмитості" обмеження (дисперсія це в теорії ймовірності). Бачимо (на графіках і в таблицях), що в діапазоні $1 < \text{fuznja} < 5$ результати досить стабільні - вони вказують на оптимальне резервування 2-2-1.



Висновки

Розроблено базову програмну систему для розв'язання і дослідження певного класу оптимізаційних задач нелінійного програмування. Використано альтернативні методи оптимізації. Система може бути модифікована для задач більшої розмірності.

Контрольні запитання

1. Порядок розв'язання задачі дискретним методом динамічного програмування.
2. Порядок розв'язання задачі методом множників Лагранжа.
3. Порядок розв'язання задачі методом нечіткої логіки.
4. Інтерпретації нечіткості.
5. Розробіть самостійно визначення понять:
 - надійність, безвідмовність.
 - відмовонечутливість.
 - відмовостійкість.
 - відмовобезпечність.

Література

1. Бадьора С.П., Колесник І.С. Інтеграція навчання, наукових досліджень і практики на прикладі узагальнень задачі Марковіца // Доповіді МНК "Інтернет – освіта – наука – 2002. Том 2", - Вінниця: Універсум-Вінниця, 2002. — С 275-279.
2. Беллман Р. Процессы регулирования с адаптацией. — М.: Наука, 1964. — 317 с.
3. Беллман Р., Гликсберг И., Гросс О. Некоторые вопросы математической теории управления. — М.: Издат. Иностран. Литер., 1962. — 233 с.
4. Беллман Р., Калаба Р. Динамическое программирование и современная теория управления. — М.: Наука, 1969. — 131 с.
5. Боровская Т.Н., Северилов В.А., Северилов П.В. Что будет, если? Имитационное моделирование в Mathcad // Компьютеры + Программы. — 2000. — №12. — С. 37— 41.
6. Боровская Т.Н., Северилов В.А., Колесник И.С. Детская экономика. Моделирование и оптимизация производственных систем // Компьютеры + Программы. — 2002. — №2. — С. 43 — 47.
7. Боровська Т.М. Декомпозиційні структури для прикладних програм синтезу регуляторів // Вісник ВДТУ. — 2000. — № 1. — С. 17-22.
8. Боровська Т.М., Рибіна І.В., Северилов В.А. Орієнтація на обчислювальний експеримент – зміна парадигми вищої освіти // Доповіді НМК "Проблеми підручника вищої школи", Універсум-Вінниця, 2001. — С. 132-137.
9. Боровська Т.М., Северилов В.А. Одержання та обробка оцінок в модульно-рейтинговій системі організації навчання // Вісник ВДТУ. — 1997. — № 3. — С. 96-104.
10. Боровська Т.М., Северилов В.А., Стахов О.П. Ориентированный на компьютер лабораторный практикум: исследование нелинейных динамических систем // Вибрации в технике и технологиях. — 1998. — №5. — С. 77-81.
11. Боровська Т.М., Северілов В.А. Електронна книга "Моделювання у менеджменті". Технології навчання, орієнтовані на моделювання // Доповіді МНК "Інтернет – освіта – наука – 2002. Том 2", - Вінниця: Універсум-Вінниця, 2002. — С 285-288.
12. Боровська Т.М., Северілов В.А. Електронна книга "Управління проектами". Технології навчання, орієнтовані на моделювання // Доповіді МНК "Інтернет – освіта – наука – 2002. Том 2", - Вінниця: Універсум-Вінниця, 2002. — С 289-294.
13. Вентцель Е.С. Исследование операций. — М.: Советское радио, 1972. — 356 с.
14. Вентцель Е.С. Теория вероятностей. — М.: Наука, 1969. — 371 с.
15. Геловэй Л. Операционный менеджмент. Принципы и практика.—

- Москва-Харьков: «Питер», 2001. — 273 с.
16. Дж.Кемени, Дж.Снелл Кибернетическое моделирование. — М.: Советское радио, 1972. — 223 с.
 17. Дибб С., Симкин Л. Практическое руководство по сегментированию рынка. — Москва-Харьков: «Питер», 2001. — 231 с.
 18. Исследование операций.Т1. Методологические основы и математические методы / Под ред. Дж Моудера. — М.: Мир, 1981. — 233 с.
 19. Исследование операций.Т2. Модели и методы / Под ред. Дж Моудера. — М.: Мир, 1981. — 417 с.
 20. Клейнрок Л. Теория массового обслуживания. Пер. с англ. Под ред. В.И. Неймана. — М.: Мир, 1979. — 432 с.
 21. Клиланд Д., Кинг В. Системный анализ и менеджмент проектов. — М.: Советское радио, 1974. — 412 с.
 22. Колесник І.С., Бадьора С.П. Інтернет-орієнтовані технології виробництва інтелектуальної продукції. Структура малих дослідницьких груп // Доповіді МНК “Інтернет – освіта – наука – 2002. Том 2”, - Вінниця: Універсум-Вінниця, 2002. — С 285-288.
 23. Колесник, І.С., Северілов В.А. Оптимальне управління розподіленням ресурсів в децентралізованих системах // Доповіді МНК “Контроль і управління в технічних системах”, Універсум-Вінниця, 2001. — С 73-78.
 24. Коршунов Ю.М. Математические основы кибернетики. — М.: Энергия, 1980. — 475 с.
 25. Котлер Ф. Основы маркетинга: Пер. с англ. / Общ. ред. Пиньковой.— М.: Прогресс, 1990.—736 с.
 26. Кофман А. Модели и методы исследования операций. — М.: Мир, 1966. — 370 с.
 27. Крушвиц Л. Финансирование и инвестиции: Базовый курс. Учебник для вузов. — Москва-Харьков: «Питер», 2000. — 389 с.
 28. Кузин Л.Т. Основы кибернетики. — М.: Наука, 1969. — 413 с.
 29. Мак-Дональд М. Стратегическое планирование маркетинга. — Москва-Харьков: «Питер», 2001. — 267 с.
 30. Макконелл А., Брю Т. Экономикс. — М.: Республика, 1991. — 562 с.
 31. Маркетинг. Сборник переводов. — К.: Видавництво „Україна”, 1995, — 240 с.
 32. Нэгл Томас Т. Стратегия и тактика ценообразования. Руководство для принятия решений приносящих прибыль. — Москва-Харьков: «Питер», 2001. — 375 с.

35. Опойцев В.И. Равновесие и устойчивость в моделях коллективного поведения. — М.: Наука, 1977. — 311 с.
36. Пешель М. Моделирование сигналов и систем. - М.: Мир, 1981.-301 с.
37. Робертс Ф.С. Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам. — М.: Наука, 1986. — 496 с.
38. Саати Т.Л. Элементы теории массового обслуживания и её приложения. — М.: Советское радио, 1972. — 277 с.
39. Северилов В.А. Детская экономика. Роль везения и умения в развитии производственных систем // Компьютеры+Программы. — 2003. — №1, — С. 46- 50
40. Северилов В.А., Боровская Т.Н., Мельник Е.Н. Эволюционная технология разработки экспертных систем // Теория автоматизированного проектирования.— Харьков: ХАИ, - 1987. — Вып. 4 — С. 33-38.
41. Северілов В.А., Колесник І.С. Оптимальне управління розподіленням ресурсів в децентралізованих системах. // Сб. доповідей МНК “Контроль і управління в технічних системах”, Вінниця, 2001. — С. 142 -146.
42. Северілов В.А., Колесник І.С., Бадьора С.М. Електронна книга “Моделювання та оптимізація в економіці”. Проблема трьох “не” - нелінійності, нестаціонарності та не випуклості // Сб. Доповідей НМК “Проблеми підручника вищої школи”, Універсум-Вінниця, 2001. — С. 138-141.
43. Северілов В.А., Січко Т.В. Технологія створення електронних книг на прикладі посібника з спеціальних розділів вищої математики // Сб.Доповідей НМК “Проблеми підручника вищої школи”, Універсум-Вінниця, 2001. — С. 217- 221.
44. Тёрнер Д. Вероятность, статистика и исследование операций. — М.: Статистика, 1976. — 317 с.
45. Цыпкин Я.З. Адаптация и обучение в автоматических системах. — М.: Наука, 1968 — 400 с.
46. Чуев Ю.В., Спехова Г.П. Технические задачи исследования операций. — М.: Советское радио, 1981. — 176 с.
47. Экланд И. Элементы математической экономики. — М.: Мир, 1983. — 248 с.
48. Kelly K. New Rules for the New Economy. 10 radical strategies for a connected world.— Penguin books, 1999. — 180 p.
49. Robert R. Prechter. Jr The wave principle of human social behavior and the new Conquer the Crash.. You can survive and prosper in deflationary depression. — John Willey&sons, LTD, 2002 — 278p.

Навчальне видання

**Таїса Миколаївна Боровська,
Ірина Сергіївна Колесник,
Віктор Андрійович Северілов**

**ОСНОВИ ТЕОРІЇ УПРАВЛІННЯ
ТА ДОСЛІДЖЕННЯ ОПЕРАЦІЙ**

Навчальний посібник

Оригінал-макет підготовлено Т. М. Боровською

Редактор В. О. Дружиніна

Видавництво ВНТУ "Універсум-Вінниця"
Свідоцтво Держкомінформу України
серія ДК №746 від 25.12.2001
21021, м. Вінниця, Хмельницьке шосе, 95, ВНТУ

Підписано до друку
Формат 29,7×42¹/₄
Друк різнографічний
Тираж _____ прим.
Зам. №

Гарнітура Times New Roman
Папір офсетний
Ум. друк. арк. 5.37

Віддруковано в комп'ютерному інформаційно-видавничому центрі
Вінницького національного технічного університету
Свідоцтво Держкомінформу України
серія ДК №746 від 25.12.2001
21021, м. Вінниця, Хмельницьке шосе, 95, ВНТУ