

ЛЕКЦІЯ 5

ОБ'ЄДНАННЯ РЕЗУЛЬТАТІВ РІЗНИХ ВИМІРЮВАНЬ. ЗВАЖЕНІ СЕРЕДНІ. МЕТОД НАЙМЕНШИХ КВАДРАТІВ

Розглянемо питання об'єднання результатів двох або більше окремих і незалежних вимірювань однієї фізичної величини. Покажемо, що *найкращою оцінкою цієї величини, ґрунтованої на декількох вимірюваннях, буде відповідне зважене середнє цих вимірювань.*

5.1 Зважені середні

5.1.1 Проблема об'єднання результатів різних вимірювань

Часто буває так, що одна фізична величина вимірюється кілька разів, можливо навіть в декількох незалежних лабораторіях, і тоді виникає питання, *як об'єднати ці результати, щоб отримати єдину найкращу оцінку.*

Припустимо, наприклад, що дві людини A та B ретельно вимірюють величину x і отримують такі результати:

$$A: x = x_A \pm \sigma_A \quad (5.1)$$

та

$$B: x = x_B \pm \sigma_B. \quad (5.2)$$

Кожен з цих результатів, імовірно, й сам по собі є результатом кількох вимірювань; наприклад, x_A може бути середнім всіх вимірювань A та σ_A – стандартне відхилення цього середнього (і аналогічно для x_B та σ_B). Питання полягає в тому, як найкраще об'єднати x_A та x_B , щоб отримати єдину найкращу оцінку для x .

Зазначимо, що якщо різниця $|x_A - x_B|$ між двома вимірюваннями набагато більша обох похибок σ_A та σ_B , то, імовірно, щось не в порядку, принаймні, у випадку одного з вимірювань. В цій ситуації можна сказати, що два вимірювання є *суперечливими*, і необхідно ретельно проаналізувати обидва

вимірювання, щоб перевірити, чи не зазнає одне з них (або обидва) непомічених систематичних помилок.

Припустимо, що два вимірювання (5.1) і (5.2) є *несуперечливими*, тобто різниця $|x_A - x_B|$ не набагато більша за будь-яку з похибок σ_A і σ_B . У цьому випадку доцільно запитати, якою є найкраща оцінка $x_{\text{найкр}}$ істинного значення X , ґрунтована на цих двох вимірюваннях. Першою реакцією могло б бути обчислення середнього значення $\frac{(x_A - x_B)}{2}$ двох вимірювань. Однак *цей шлях не підходить, якщо дві похибки σ_A і σ_B не рівні*. Обчислення простого середнього $\frac{(x_A - x_B)}{2}$ робить однаково важливими обидва вимірювання, в той час як точніший результат є більш значущим і, отже, повинен мати більшу вагу.

5.1.2 Зважене середнє

Можна розв'язати нашу задачу, використовуючи *принцип максимальної правдоподібності*. Якщо припустити, що результати обох вимірювань підпадають під розподіл Гауса, і позначити невідоме істинне значення величини x як X , то імовірність того, що A отримає своє значення x_A , дорівнює

$$P_X(x_A) \sim \frac{1}{\sigma_A} e^{-\frac{(x_A - X)^2}{2\sigma_A^2}} \quad (5.3)$$

та імовірність того, що B отримає своє значення x_B , дорівнює

$$P_X(x_B) \sim \frac{1}{\sigma_B} e^{-\frac{(x_B - X)^2}{2\sigma_B^2}}. \quad (5.4)$$

Якщо ввести індекс X , ми вкажемо явно, що ці імовірності залежать від невідомого істинного значення. (Вони також залежать від відповідної ширини σ_A і σ_B , але ми цього не вказали).

Імовірність того, що A отримає значення x_A та B отримає значення x_B , дорівнює добутку двох імовірностей (5.3) та (5.4). Цей добуток буде експоненціальною функцією з показником, рівним сумі двох показників у (5.3) та (5.4). Запишемо це як

$$P_X(x_A, x_B) = P_X(x_A)P_X(x_B) \sim \frac{1}{\sigma_A \sigma_B} e^{-\frac{\chi^2}{2}}, \quad (5.5)$$

де

$$\chi^2 = \left(\frac{x_A - X}{\sigma_A} \right)^2 + \left(\frac{x_B - X}{\sigma_B} \right)^2. \quad (5.6)$$

Ця величина є сумою квадратів відхилень від X результатів двох вимірювань, поділених на відповідні похибки. Її іноді називають **сумою квадратів**.

Принцип максимальної правдоподібності стверджує, що найкращою оцінкою для невідомого істинного значення X буде таке значення, для якого фактично отримані величини x_A та x_B є найімовірнішими. *Іншими словами, найкращою оцінкою для X буде значення, при якому імовірність (5.5) досягає максимуму, або, що еквівалентно, показник χ^2 є мінімальним.* (Оскільки максимізація імовірності тягне за собою мінімізацію «суми квадратів» χ^2 , то цей метод оцінювання X називають **методом найменших квадратів (МНК)**.)

Таким чином, щоб визначити найкращу оцінку, продиференціюємо (5.6) по X та дорівнюємо похідну до нуля: $2 \frac{x_A - X}{\sigma_A^2} + 2 \frac{x_B - X}{\sigma_B^2} = 0$.

Розв'язок цього рівняння відносно X є найкращою оцінкою $x_{\text{найкр}}$, і вона

дорівнює

$$x_{\text{найкр}} = \frac{\left(\frac{x_A}{\sigma_A^2} + \frac{x_B}{\sigma_B^2} \right)}{\left(\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} \right)}. \quad (5.7)$$

Цей досить громіздкий результат можна записати компактніше, якщо визначити *вагові коефіцієнти*

$$\omega_A = \frac{1}{\sigma_A^2} \quad \text{і} \quad \omega_B = \frac{1}{\sigma_B^2}. \quad (5.8)$$

Підставляючи їх у (5.7), отримаємо

$$x_{\text{найкр}} = \frac{\omega_A x_A + \omega_B x_B}{\omega_A + \omega_B}. \quad (5.9)$$

Якщо два вихідних вимірювання є однаково точними ($\sigma_A = \sigma_B$ і, отже, $\omega_A = \omega_B$), то наша відповідь зводиться до простого середнього значення $\frac{(x_A + x_B)}{2}$. У загальному випадку вираз (5.9) дає **зважене середнє** (воно аналогічне формулі для центра ваги двох тіл, коли ω_A і ω_B – дійсні вагові коефіцієнти двох тіл, а x_A і x_B – їх координати). У даному випадку **вагові коефіцієнти** є оберненими значеннями квадратів похибок у вихідних вимірюваннях, як видно з (5.8). **Якщо вимірювання А є точнішими за В, тобто $\sigma_A < \sigma_B$ і, отже, $\omega_A > \omega_B$; то найкраща оцінка $x_{\text{найкр}}$ буде ближчою до x_A , ніж до x_B .**

Застосування для випадку, коли об'єднуються декілька вимірювань однієї величини. Припустимо, що в нас є N окремих вимірювань величини x : $x_1 \pm \sigma_1$, $x_2 \pm \sigma_2$, ..., $x_N \pm \sigma_N$ з відповідними похибками σ_1 , σ_2 , ..., σ_N . Міркуючи аналогічно викладеному вище, отримаємо, що найкраща оцінка, ґрунтована на цих вимірюваннях, дорівнює зваженому середньому

$$x_{\text{наил}} = \frac{\sum_{i=1}^N \omega_i x_i}{\sum_{i=1}^N \omega_i}, \quad (5.10)$$

де **вагові коефіцієнти** ω_i – обернені значення квадратів відповідних похибок

$$\omega_i = \frac{1}{\sigma_i^2} \quad \text{для } i = 1, 2, \dots, N. \quad (5.11)$$

Оскільки ваговий коефіцієнт $\omega_i = \frac{1}{\sigma_i^2}$, пов'язаний з кожним вимірюванням, містить квадрат відповідної похибки σ_i , то будь-яке вимірювання, суттєво менш точне за інші, внесе набагато менший внесок у кінцевий результат (5.10).

Так якщо одне вимірювання у чотири рази є менш точним за інші, то його ваговий коефіцієнт у 16 разів менший за інші вагові коефіцієнти, і у багатьох випадках це вимірювання можна просто ігнорувати.

Оскільки кінцевий результат (5.10) для $x_{\text{найкр}}$ – це проста функція вихідних значень x_1, x_2, \dots, x_N , то похибка нашого результату методом розрахунку помилок у непрямих вимірюваннях дорівнює

$$\sigma_{x_{\text{найкр}}} = \left(\sum_{i=1}^N \omega_i \right)^{-\frac{1}{2}}, \quad (5.12)$$

де $\omega_i = \frac{1}{\sigma_i^2}$.

5.2 Апроксимація методом найменших квадратів

5.2.1 Дані, які повинні розміщуватись на прямій

Один з найбільш загальних експериментів полягає у вимірюванні кількох значень двох різних фізичних змінних для дослідження їх математичного зв'язку. Наприклад, можна кидати камінь з різних висот h_1, \dots, h_N та вимірювати відповідний час падіння t_1, \dots, t_N , щоб перевірити, чи пов'язані ці значення висоти та часу очікуваним співвідношенням $h = \frac{1}{2}gt^2$.

Імовірно, найважливішими експериментами такого типу є ті, де очікуваний зв'язок є лінійним, як це реалізується у наступному прикладі. Якщо ми допускаємо, що тіло падає з постійним прискоренням вільного падіння g , то його швидкість v повинна бути лінійною функцією часу t : $v = v_0 + gt$.

У загальному випадку розглядатимемо будь-які дві фізичні змінні x та y , які, як ми вважаємо, пов'язані лінійною залежністю вигляду

$$y = A + Bx, \quad (5.13)$$

де A та B – постійні.

Якщо дві змінні y та x пов'язані лінійною залежністю вигляду (5.13), то графік y від x повинен бути прямою з нахилом B , яка перетинає вісь y в точці

$y=A$. Якщо б ми виміряли N різних значень x_1, \dots, x_N та відповідних значень y_1, \dots, y_N і якщо б результати наших вимірювань не містили похибок, то кожна точка (x_i, y_i) розмістилася б точно на лінії $y=A+Bx$, як наведено на рис.5.1,а. На практиці ж завжди є похибки, і найбільшого, чого ми можемо очікувати, – це того, що відстань кожної точки (x_i, y_i) від лінії повинна бути порівняна в розумних межах з похибками, як наведено на рис.5.1,б.

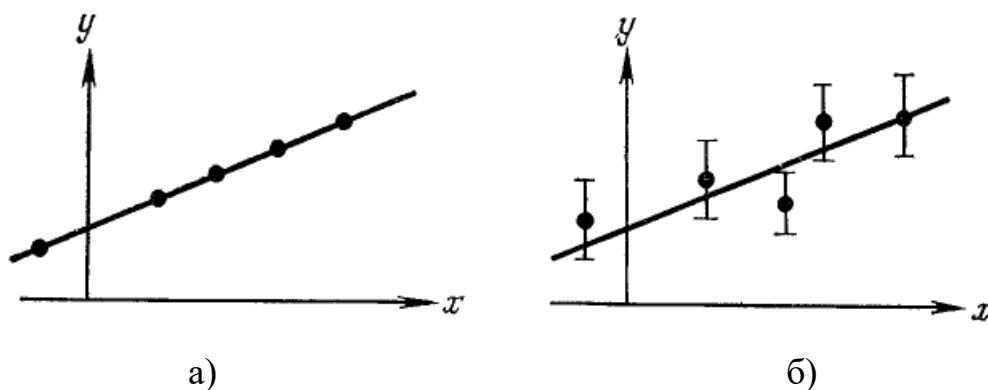


Рисунок 5.1

При проведенні ряду вимірювань описаного типу виникають два питання. Перше: якщо ми приймемо за факт те, що y та x дійсно пов'язані лінійно, то прийдемо до задачі визначення прямої $y=A+Bx$, яка найкраще апроксимує результати вимірювань, тобто – до задачі визначення найкращих оцінок постійних A та B , ґрунтованих на даних $(x_1, y_1), \dots, (x_N, y_N)$. Ця задача може бути розв'язана як графічно, так і аналітично за допомогою методу максимальної правдоподібності. Рівняння, що визначає «найкращу» (у сенсі математичного очікування) залежність y від x , називається **регресією y на x** . Як відомо, аналітичний метод визначення найкращої прямої лінії, яка апроксимує серію експериментальних точок, називається **лінійною регресією** або **апроксимацією прямої МНК**.

Друге питання: чи дійсно виміряні значення $(x_1, y_1), \dots, (x_N, y_N)$ відповідають наших очікуванням, що функція y є лінійною по x . Спочатку ми можемо визначити лінію, що найкраще апроксимує дані, але потім ми повинні

запропонувати деяку міру, яка показала б, наскільки добре ця лінія апроксимує дані.

5.2.2 Розрахунок постійних A та B

Розглянемо перше питання – задачу визначення найкращої прямої $y=A+Bx$, що апроксимує набір експериментальних точок $(x_1, y_1), \dots, (x_N, y_N)$. Для спрощення припустимо, що, хоча результати наших вимірювань x та y мають деякі похибки, похибка вимірювань x є нехтовно малою. (Це – розумне припущення, оскільки похибки в одній змінній часто є набагато більшими за похибки в іншій, які ми, отже, можемо ігнорувати.) Далі припускатимемо, що всі похибки в y однакові за величиною. Якщо говорити точніше, припустимо, що результат вимірювання кожного y_i підпадає під розподіл Гауса з однаковою шириною σ_y в усіх вимірюваннях.

Знаючи постійні A та B , можна для будь-якого даного значення x_i (яке, за нашими припущеннями, не має похибки) обчислити істинне значення відповідної величини y_i : (істинне значення y_i) $= A + Bx_i$. (5.14)

Результат вимірювання y_i підпадає під нормальний розподіл з центром в істинному значенні та з шириною σ_y . Отже, імовірність отримання значення y_i

$$\text{дорівнює} \quad P_{A,B}(y_i) \sim \frac{1}{\sigma_y} e^{-\frac{(y_i - A - Bx_i)^2}{2\sigma_y^2}}, \quad (5.15)$$

де нижні індекси A та B вказують на те, що ця імовірність залежить від (невдомих) значень A та B . Імовірність отримання всього набору результатів вимірювань y_1, \dots, y_N дорівнює добутку

$$P_{A,B}(y_1, \dots, y_N) = P_{A,B}(y_1) \times \dots \times P_{A,B}(y_N) \sim \frac{1}{\sigma_y^N} e^{-\frac{\chi^2}{2}}, \quad (5.16)$$

де показник визначається формулою

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - A - Bx_i)^2}{\sigma_y^2}. \quad (5.17)$$

Відомо, що найкращі оцінки для невідомих постійних A та B , ґрунтовані на даних вимірюваннях, – це такі значення A та B , для яких імовірність $P_{A,B}(y_1, \dots, y_N)$ є максимальною або для яких сума квадратів χ^2 (5.17) є мінімальною (тому цей метод відомий як апроксимація МНК). Щоб знайти ці значення, продиференціюємо χ^2 по A та B і дорівнюємо ці похідні нулю:

$$\frac{\partial \chi^2}{\partial A} = \left(-\frac{2}{\sigma_y^2} \right) \sum_{i=1}^N (y_i - A - Bx_i) = 0, \quad (5.18)$$

$$\frac{\partial \chi^2}{\partial B} = \left(-\frac{2}{\sigma_y^2} \right) \sum_{i=1}^N x_i (y_i - A - Bx_i) = 0. \quad (5.19)$$

Вирази (5.18) та (5.19) можна переписати як систему рівнянь для A та B :

$$AN + B \sum x_i = \sum y_i, \quad (5.20)$$

$$A \sum x_i + B \sum x_i^2 = \sum x_i y_i. \quad (5.21)$$

Ці два рівняння (**нормальні рівняння**) легко розв'язуються й дають **оцінки МНК для постійних A та B** :

$$A = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{\Delta}, \quad (5.22)$$

$$B = \frac{N(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\Delta}, \quad (5.23)$$

$$\text{де} \quad \Delta = N(\sum x_i^2) - (\sum x_i)^2. \quad (5.24)$$

Формули (5.22) і (5.23) дають найкращі оцінки постійних A та B для прямої $y = A + Bx$, ґрунтовані на вимірних точках $(x_1, y_1), \dots, (x_N, y_N)$. Отримана лінія називається **лінією апроксимації МНК** цих даних або **лінією регресії y від x** .

Для **МНК з ваговими коефіцієнтами** формули (5.22)–(5.24) виглядатимуть таким чином

$$A = \frac{(\sum \omega_i x_i^2)(\sum \omega_i y_i) - (\sum \omega_i x_i)(\sum \omega_i x_i y_i)}{\Delta}, \quad (5.25)$$

$$B = \frac{(\sum \omega_i)(\sum \omega_i x_i y_i) - (\sum \omega_i x_i)(\sum \omega_i y_i)}{\Delta}, \quad (5.26)$$

де $\omega_i = \frac{1}{\sigma_i^2}$ – вагові коефіцієнти,

$$\Delta = (\sum \omega_i)(\sum \omega_i x_i^2) - (\sum \omega_i x_i)^2. \quad (5.27)$$

Далі розглянемо похибки у наших оцінках A та B . Для цього необхідно розглянути похибки σ_y у вихідних вимірюваннях y_1, \dots, y_N .

5.2.3 Похибка у вимірюваннях y

Слід пам'ятати, що числа y_1, \dots, y_N не є N результатами вимірювань однієї величини. (Вони могли б бути, наприклад, значеннями часу, протягом яких камінь падає з N різних висот). Таким чином, ми не отримаємо жодної інформації про надійність цих результатів, досліджуючи розкид в їх значеннях.

Однак можна легко оцінити похибку σ_y в числах y_1, \dots, y_N . Результат вимірювання кожного y_i (за нашими припущеннями) розподілений нормально біля істинного значення $A + Bx_i$ з параметром ширини σ_y . Таким чином, відхилення $y_i - A - Bx_i$ розподілені нормально, причому всі з однаковим центральним значенням 0 і однаковою шириною σ_y . Це одразу ж призводить до припущення, що хороша оцінка σ_y могла б визначатися сумою квадратів у відомому нам вигляді

$$\sigma_y^2 = \frac{1}{N} \sum (y_i - A - Bx_i)^2. \quad (5.28)$$

Цей результат може бути підтверджений за допомогою принципу максимальної правдоподібності. (Найкраща оцінка для потрібного нам параметра (в даному випадку σ_y) є значенням, для якого імовірність (5.16) отримання значень y_1, \dots, y_N є максимальною. Диференціюючи (5.16) за σ_y та

дорівнюючи похідну нулю, отримаємо, що цією найкращою оцінкою буде вираз (5.28)).

Однак оцінка (5.28) для σ_y^2 не є остаточною. Числа A та B у (5.28) – невідомі істинні значення постійних A та B . На практиці вони повинні бути замінені найкращими оцінками для A та B , а саме – виразами (5.22) та (5.23), і така заміна дещо змінює значення (5.28). Ця зміна компенсується заміною фактору N у знаменнику на $(N-2)$. Таким чином, наша остаточна відповідь для похибки у вимірюваннях y_1, \dots, y_N є

$$\sigma_y^2 = \frac{1}{N-2} \sum (y_i - A - Bx_i)^2, \quad (5.29)$$

де A та B даються виразами (5.22) і (5.23).

5.2.4 Похибка у постійних A та B

Визначивши похибку σ_y в отриманих числах y_1, \dots, y_N , можна розрахувати похибки оцінок постійних A та B . Оцінки (5.22) та (5.23) для A та B – точно визначені функції вимірюваних значень y_1, \dots, y_N . Отже, похибки в A та B визначають простим розрахунком помилок в непрямих вимірюваннях,

виходячи з похибок в y_1, \dots, y_N :

$$\sigma_A^2 = \frac{\sigma_y^2 \sum x_i^2}{\Delta}, \quad (5.30)$$

$$\sigma_B^2 = \frac{N\sigma_y^2}{\Delta}, \quad (5.31)$$

де Δ визначається з (5.24).

Приклад 5.1. Якщо об'єм деякої кількості ідеального газу підтримувати постійним, то його температура T буде лінійною функцією тиску в газі P :

$$T = A + BP. \quad (5.32)$$

Тут постійна A – температура, за якої тиск P падає до нуля (якщо газ не сконденсується спочатку у рідину); вона називається абсолютним нулем температури і має прийняте значення $A = -273.15^\circ\text{C}$.

Постійна B залежить від природи газу, його маси й об'єму. Вимірюючи ряд значень T та P , ми можемо визначити найкращі оцінки для постійних A та B . Зокрема, постійна A дає абсолютний нуль температури.

Одна система п'яти вимірювань P та T , отримана експериментально, наведена у трьох перших стовпцях табл.5.1. Вважатимемо, що похибка у вимірюваннях P нехтовно мала, а похибки в T всі однакові та дорівнюють «кільком градусам». Припускаючи, що його точки повинні розміщуватися на прямій типу (5.32), обчислюємо найкращу оцінку постійної A (абсолютного нуля) та її похибку. Які можна зробити висновки?

Таблиця 5.1 – Вимірювання тиску та температури

Номер досліду i	Тиск P_i , мм рт.ст.	Температура T_i , $^\circ\text{C}$	$A + BP_i$
1.	65	-20	-22.2
2.	75	17	14.9
3.	85	42	52.0
4.	95	94	89.1
5.	105	127	126.2

Використовуємо формули (5.22) і (5.30), замінюючи x_i на P_i та y_i на T_i , щоб розрахувати всі величини, що представляють інтерес. Для цього потрібне обчислення сум $\sum P_i$, $\sum P_i^2$, $\sum T_i$, $\sum P_i T_i$. Згідно з табл.5.1, можна обчислити $\sum P_i = 425$, $\sum P_i^2 = 37125$, $\sum T_i = 260$, $\sum P_i T_i = 25810$, $\Delta = 5000$, де $\Delta = N(\sum P_i^2) - (\sum P_i)^2$.

У розрахунках такого роду важливо залишати багато значущих цифр, оскільки нам доведеться обчислювати різниці цих великих чисел. Знаючи ці суми, одразу ж можна обчислити найкращі оцінки постійних A та B :

$$A = \frac{(\sum P_i^2)(\sum T_i) - (\sum P_i)(\sum P_i T_i)}{\Delta} = -263,35 \text{ та } B = \frac{N(\sum P_i T_i) - (\sum P_i)(\sum T_i)}{\Delta} = 3,71.$$

Це дає найкращу оцінку для абсолютного нуля $A = -263,35^\circ\text{C}$.

Знаючи постійні A та B , можна розрахувати числа $A + BP_i$, тобто температури, «очікувані» на основі нашої найкращої апроксимації співвідношення $T = A + BP$. Вони наведені у правому стовбці табл.5.1, і, як ми і сподівались, всі узгоджуються в розумних межах з температурами, що спостерігаються. Далі можна розрахувати різниці між числами у двох останніх стовпцях та отримати $\sigma_T^2 = \frac{1}{N-2} \sum (T_i - A - BP_i)^2 = 44.6333$ і, отже, і стандартне відхилення $\sigma_T = 6.6808$.

Це значення добре узгоджується з оцінкою, згідно з якою вимірювання температури не визначені з точністю до «кількох градусів».

Зрештою, ми можемо розрахувати похибку в A , використовуючи (5.30):

$$\sigma_A^2 = \frac{\sigma_T^2 \sum P_i^2}{\Delta} = 331.4025 \text{ або } \sigma_A = 18.2045.$$

Отже, округлений підсумковий висновок має вигляд:

$$\text{абсолютний нуль } A = -263.35 \pm 18.21^\circ\text{C},$$

що задовільно узгоджується з прийнятим значенням -273.15°C .

Побудуємо графік за отриманими результатами (рис.5.2). П'ять експериментальних точок разом з їх похибками $\pm 7^\circ$ за шкалою T наведені праворуч зверху. Найкраща пряма проходить через чотири рисочки помилок і близько від п'ятої.

Щоб знайти значення абсолютного нуля, лінію слід продовжити за межі області, де лежать всі експериментальні точки, до її перетину з віссю T . Цей процес *екстраполяції* (продовження кривої за межі точок, за якими вона визначається) може призвести до великих похибок, як це видно з рисунку. Дуже невелика зміна у нахилі лінії призведе до великих змін у положенні точки перетину цієї лінії з віссю T , оскільки ця вісь дуже віддалена від експериментальних точок. Таким чином, *будь-яка похибка у даних значно збільшується, якщо ми змушені екстраполювати на великі відстані*. Це

пояснює, чому похибка у значенні абсолютного нуля ($\pm 18^\circ$) набагато більша, ніж у вихідних вимірюваннях температури ($\pm 7^\circ$).

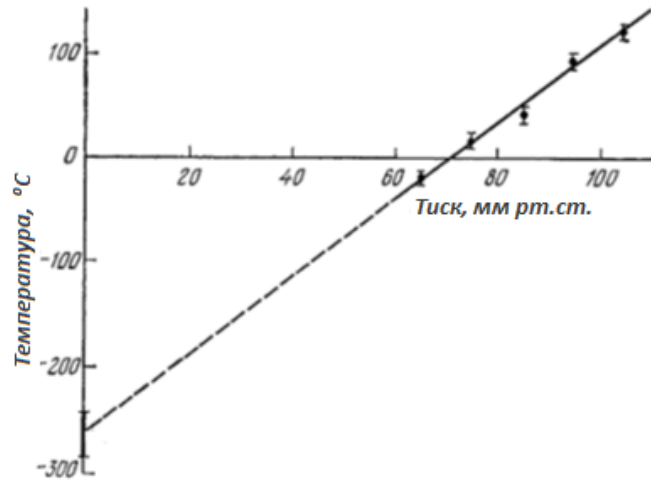


Рисунок 5.2 – Графік залежності температури T від тиску P для газу при постійному об'ємі. Рисочки помилок мають довжину в одне стандартне відхилення σ_T з кожного боку від кожної з п'яти експериментальних точок, а лінія – це найкраща апроксимація, отримана МНК. Абсолютний нуль температури можна знайти, якщо екстраполювати лінію до її перетину з віссю T

5.2.5 Апроксимація іншими кривими методом найменших квадратів

Апроксимація поліномом. Часто одна змінна y виражається через поліном від іншої змінної x :

$$y = A + Bx + Cx^2 + \dots + Hx^n. \quad (5.33)$$

Так можна очікувати, що висота y , якою пролітає тіло, яке падає, квадратично залежить від часу t : $y = y_0 + v_0 t - \frac{1}{2} g t^2$, де y_0 та v_0 – початкові висота та швидкість, g – прискорення вільного падіння. Якщо задана сукупність спостережень двох змінних, то можна визначити найкращі оцінки постійних A, B, \dots, H в (5.33) аналогічно проведеному у п.5.1.2.

Для спрощення припустимо, що поліном (5.33) є квадратичним, тобто

$$y = A + Bx + Cx^2. \quad (5.34)$$

Припустимо, що в нас є серія вимірювань (x_i, y_i) , $i = \overline{1, N}$, де у всіх y_i однакові похибки і всі x_i є точними. Для кожного x_i відповідне істинне значення y_i задається формулою (5.34), де параметри A , B і C поки невідомі. Нехай результати вимірювань y_i підпадають під нормальний розподіл, кожний з центром на відповідному істинному значенні і всі з однією шириною σ_y . Це дозволяє представити імовірність отримання значень спостереження y_1, \dots, y_N у

вигляді:

$$P(y_1, \dots, y_N) \sim e^{-\frac{\chi^2}{2}}, \quad (5.35)$$

де

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - A - Bx_i - Cx_i^2)^2}{\sigma_y^2}. \quad (5.36)$$

(Це відповідає виразу (5.17) у лінійному випадку.) Найкращі оцінки A , B і C – такі, для яких імовірність $P(y_1, \dots, y_N)$ є максимальною або величина χ^2 є мінімальною. Диференціюючи χ^2 по A , B і C та дорівнюючи ці похідні нулю, ми отримаємо три рівняння:

$$\begin{aligned} AN + B \sum x_i + C \sum x_i^2 &= \sum y_i, \\ A \sum x_i + B \sum x_i^2 + C \sum x_i^3 &= \sum x_i y_i, \\ A \sum x_i^2 + B \sum x_i^3 + C \sum x_i^4 &= \sum x_i^2 y_i. \end{aligned} \quad (5.37)$$

Для будь-якого даного набору результатів вимірювань (x_i, y_i) ця система нормальних рівнянь для A , B і C може бути розв'язана і, таким чином, знайдені найкращі оцінки для A , B і C . Зі знайденими таким шляхом значеннями A , B , C вираз $y = A + Bx + Cx^2$ називається **поліноміальною апроксимацією, отриманою МНК**, або **поліноміальною регресією** для даних результатів вимірювань.

Метод поліноміальної регресії легко узагальнюється для поліномів будь-якої степені, хоча й отримані нормальні рівняння стають дуже громіздкими у випадку поліномів високої степені. Аналогічний метод може бути застосований до будь-якої функції $y=f(x)$, що залежить від різних невідомих параметрів A , B , На жаль, отримані нормальні рівняння, які визначають найкращі оцінки для A , B , ..., складно, а часом і неможливо розв'язати. Однак

є один клас задач, які завжди можна розв'язати, а саме – задачі, в яких функція $y=f(x)$ лінійно залежить від параметрів A, B, \dots . Цей клас включає всі поліноми, а також багато інших функцій. Наприклад, у випадку деяких задач у подається як сума тригонометричних функцій:

$$y = A \sin x + B \cos x. \quad (5.38)$$

Для цієї функції й фактично для будь-якої функції, лінійної відносно параметрів A, B, \dots , нормальні рівняння, що визначають найкращі оцінки для A, B, \dots , – це система лінійних рівнянь, яка завжди може бути розв'язана.

Експоненціальні функції. Одна з найважливіших функцій у фізиці – експоненціальна функція $y = Ae^{Bx}$, (5.39)
де A і B – постійні. Так, інтенсивність I випромінювання після проходження відстані x через перешкоду спадає експоненціально: $I = I_0 e^{-\mu x}$, де I_0 – початкова інтенсивність, а μ , характеризує поглинання у перешкоді. Заряд на замкнутому через опір конденсаторі спадає експоненціально: $Q = Q_0 e^{-\lambda t}$, де Q_0 – початковий заряд, $\lambda = \frac{1}{RC}$, R і C – опір та ємність, відповідно.

Якщо є серія вимірювань (x_i, y_i) , то для кожного y_i можна обчислити $z_i = \ln y_i$. Тоді точки (x_i, z_i) повинні розміщуватись на лінії

$$z = \ln A + Bx \quad (5.40).$$

Ця лінія може бути розрахована МНК, і таким чином будуть отримані найкращі оцінки для постійних $\ln A$ (за якою ми можемо знайти A) та B .

Приклад 5.2. Багато популяцій (людей, бактерій, радіоактивних ядер тощо) змінюються з часом експоненціально. Якщо якась популяція N зменшується з часом експоненціально, ми записуємо

$$N = N_0 e^{-\frac{t}{\tau}}, \quad (5.41)$$

де τ називається *середнім часом життя популяції*, який тісно пов'язаний з часом піврозпаду $t_{\frac{1}{2}}$ (дійсно, $t_{\frac{1}{2}} = 0.693\tau$). Нехай якийсь біолог вважає, що популяція бактерій зменшується з часом експоненціально згідно з (5.41). Він вимірює чисельність популяції на кожний з трьох послідовних днів і отримує результати, наведені в табл.5.2. Яку найкращу оцінку середнього часу життя τ він отримає за цими даними?

Таблиця 5.2 – Чисельність популяції бактерій

Час t_i , дні	Чисельність популяції N_i	$z_i = \ln N_i$
0	153000	11.94
1	137000	11.83
2	128000	11.76

Якщо N змінюється згідно з (5.41), то змінна $z = \ln N$ повинна лінійно залежати від t :

$$z = \ln N = \ln N_0 - \frac{t}{\tau}. \quad (5.42)$$

Отже, біолог розраховує три числа $z_i = \ln N_i$ ($i=0,1,2$), наведені у третьому стовпці табл.5.2. Використовуючи ці три числа, він розраховує пряму (5.42) МНК і отримує такі найкращі оцінки для коефіцієнтів $\ln N_0$ та $-\frac{1}{\tau}$: $\ln N_0 = 11.93$ та $-\frac{1}{\tau} = -0.085$ день⁻¹.

З другого значення виходить, що найкраща оцінка для середнього часу життя дорівнює $\tau = 11,7$ дня.

Множинна регресія. Дотепер ми розглядали випадок, коли спостерігаються лише дві змінні x та y , і аналізували їх зв'язок. У багатьох реальних задачах необхідно розглядати більшу кількість змінних. Наприклад, вимірюючи тиск газу P , можна виявити, що він залежить від об'єму V та температури T , і тому слід досліджувати P як функцію V і T . Найпростішим прикладом є задача, коли одна змінна z залежить лінійно від двох інших x і y :

$$z = A + Bx + Cy. \quad (5.43)$$

Ця задача може бути розв'язана методом, що є прямим узагальненням випадку двох змінних. Якщо у нас є серія вимірювань (x_i, y_i, z_i) , $i = \overline{1, N}$ (де похибки всіх z_i однакові, а x_i та y_i є точними), то ми можемо використати принцип максимальної правдоподібності й отримати, що найкращі оцінки постійних A , B , C визначаються нормальними рівняннями такого вигляду:

$$\begin{aligned} AN + B \sum x_i + C \sum y_i &= \sum z_i, \\ A \sum x_i + B \sum x_i^2 + C \sum x_i y_i &= \sum x_i z_i, \\ A \sum y_i + B \sum x_i y_i + C \sum y_i^2 &= \sum y_i z_i. \end{aligned} \quad (5.44)$$

Ці рівняння можна розв'язати відносно A , B та C і отримати найкращу апроксимацію для співвідношення (5.43).

Розглянутий випадок називається **множинною регресією**, оскільки тут більше двох змінних.

5.3 Реалізація задач регресійного аналізу в системі MatLab

Далі розглянемо розв'язання задач МНК в системі MatLab: спочатку для функції однієї змінної, потім – для функції кількох змінних. Система MatLab містить найрозповсюдженіші статистичні функції, які є частиною як ядра системи, так і Statistics Toolbox MatLab.

5.3.1 Степенева апроксимація

Зазвичай, якщо ми нічого не можемо сказати про характер поведінки теоретичної кривої, то перше, що спадає на думку – спробувати побудувати її у вигляді степеневі функції. (Як відомо, для степеневі залежності $y = b_0 + b_1x + b_2x^2 + \dots + b_mx^m$ базисними функціями є: $1, x, x^2, \dots, x^m$, і всього їх $m + 1$.) Для розв'язання цієї задачі в MatLab є функції `polyfit` та `polyval`. Перша визначає коефіцієнти апроксимуючого поліному заданого степеня, а друга –

обчислює значення цього полінома у заданих точках. Якщо ж нам потрібно знайти оптимальний (мінімально можливий) степінь апроксимуючого поліному, потрібно побудувати ортонормований базис, оскільки базисні функції $1, x, x^2, \dots, x^m$ у загальному випадку не утворюють ортонормований базис.

Нехай вихідні дані знаходяться в текстовому файлі у вигляді двох стовпців: 1-й – аргументи x_i , 2-й – експериментальні значення y_i . Спочатку введемо дані з цього файлу та знайдемо розмір вибірки n :

```
clear all % очищуємо пам'ять
data='filename.txt'; % ім'я файлу з даними
xy=load(data); % вводимо дані з файлу – 2 стовпця
x=xy(:,1); % аргументи
y=xy(:,2); % функції
n=length(x); % кількість точок
fprintf('Кількість експериментальних точок n=%d\n',n);
```

Задамо рівень значущості q і максимально допустимий степінь полінома. Як сказано вище, базисні функції $1, x, x^2, \dots, x^m$, у загальному випадку, не утворюють ортонормований базис. Щоб його побудувати, застосовується метод Соніна-Шмідта. Для розв'язання задачі МНК в MatLab використовуємо функцію `regress`. Матрицю експерименту для неї складемо з ортонормованих поліномів. Знайдемо параметри моделі та довірчі інтервали для них (для ортонормованого базису вони повинні вийти однакової ширини). Проаналізуємо модель. Якщо довірчий інтервал для коефіцієнта апроксимації включає в себе нуль, то такий доданок можна відкинути на обраному рівні значущості q . Знайдемо ті базисні функції, які потрібно враховувати, тобто *оберемо оптимальний степінь апроксимуючого полінома*. Вважатимемо, що якщо x^m враховувати не потрібно, то й вищі степені також не враховуватимемо (хоча на практиці це не завжди так):

```
q=0.01; % рівень значущості
kmax=20; % максимальний показник степеня
k=[0:kmax]; % всі степені
X=repmat(x,1,kmax+1).^repmat(k,n,1); % матриця базисних функцій
```

```

X(:,1)=X(:,1)/norm(X(:,1)); % нормуємо 1-й стовпець
for ki=2:kmax+1 % ортонормуємо стовпці
    Sum=0; % накопичуємо суму
    for kk=1:ki-1
        Sum=Sum+(X(:,ki)'*X(:,kk))*X(:,kk);
    end
    X(:,ki)=X(:,ki)-Sum;    X(:,ki)=X(:,ki)/norm(X(:,ki));
end
[b,bint]=regress(y,X,q); % регресійна модель
disp('Довірчі інтервали для параметрів моделі');
fprintf('на рівні значущості q=%5.2f\n',q)
disp(' k   ЛіваГр   ПраваГр');
fprintf('%2.0f %12.7f %12.7f\n',[k;bint]);
np=find(prod(bint,2)>0); % враховуємо степені
fprintf('Враховуємо степені:');
fprintf(' %d',k(np));
fprintf('\n'); % перехід до нового рядка

```

Тепер можна було б побудувати аналітичні вирази для ортонормованих базисних функцій у символічному вигляді і записати вираз для апроксимуючого полінома. Але тепер, коли оптимальний ступінь апроксимуючого поліному визначений, простіше його побудувати, наприклад, за допомогою функцій *MatLab*. Тому обчислюємо коефіцієнти полінома знайденого оптимального степеня (тобто фактично заново розв'язуємо задачу), а потім будуємо графік, подібний до наведеного на рис.5.3, тобто виводимо теоретичну криву та експериментальні точки:

```

omax=max(k(np)); % оптимальний ступінь
p=polyfit(x,y,omax); % коефіцієнти поліному
fprintf('Апроксимуючий поліном %d-го степеня:\nu(x)=',omax);
fprintf('%+f12*x^%d',[p;[omax:-1:0]]);
fprintf('\n'); % перехід до нового рядка
yt=polyval(p,x); % теоретичні значення
plot(x,y,'k.',x,yt,'b-');
set(get(gcf,'CurrentAxes'),'FontName','Times New Roman Cyr','FontSize',10)
title('\bfАпроксимація степеневими поліномами')
xlabel('\itx') % мітка осі ОХ
ylabel('\ity') % мітка осі ОУ

```

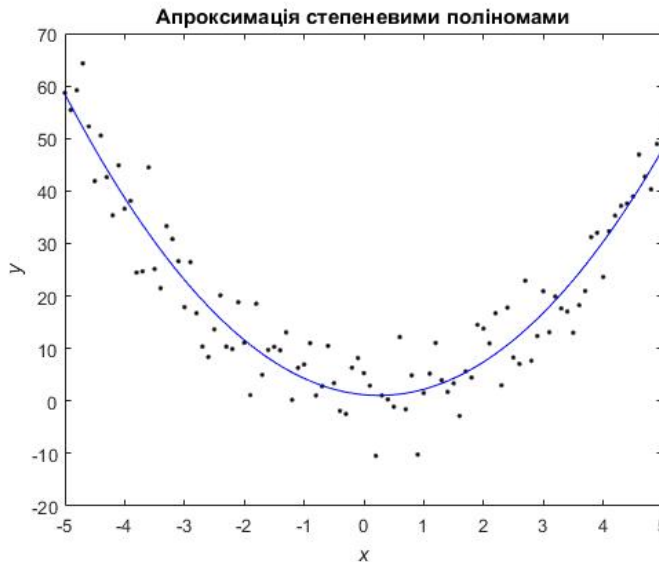


Рисунок 5.3

5.3.2 Тригонометрична апроксимація

Іноді з теорії відомо, що теоретична функція $y(x)$ є періодичною із заданим періодом T . Наприклад, температура або деформація на ободі диску ($T = 2\pi$) або атмосферний тиск протягом доби ($T = 1$ доба) тощо. У цьому випадку теоретичну криву також зручно шукати у вигляді періодичної функції з періодом T . (З гармонічного аналізу відомо, що практично будь-яку періодичну функцію можна розкласти на прості гармоніки за допомогою тригонометричного ряду – ряду Фур'є). Такими функціями є синус і косинус, тому рівняння теоретичної кривої шукаємо у вигляді відрізка ряду Фур'є.

Так, наприклад, у тригонометричному многочлені

$$y = \frac{a_0}{2} + \sum_{j=1}^m \left(a_j \cos \frac{j\pi x}{l} + b_j \sin \frac{j\pi x}{l} \right)$$

невідомі коефіцієнти позначені як $a_0, a_1, b_1,$

$a_2, b_2, \dots, a_m, b_m$, тобто всього $2m+1$ параметрів і стільки ж базисних функцій:

$$\frac{1}{2}; \cos \frac{j\pi x}{l}; \sin \frac{j\pi x}{l}.$$

На їх основі досить просто побудувати ортонормований

базис. Якщо вдало обрати точки x_i , то ці функції вже самі по собі будуть ортогональними, і їх залишиться лише пронормувати. А вдало обрати точки x_i

можна розбивши відрізок $[0; T]$ на n рівних інтервалів довжиною $\frac{T}{n}$ та взявши за x_i кінці цих інтервалів (або початки, оскільки заміна останньої точки на першу нічого не змінює): $x_i = \frac{iT}{n}$

Виконаємо завдання з тригонометричної апроксимації. Вихідні дані для нього – також текстовий файл з двома стовпцями чисел: x_i та y_i , причому для x_i повинна виконуватись умова $x_i = \frac{iT}{n}$. Вводимо початкові дані, перевіряємо виконання вказаної вище умови та знаходимо розмір вибірки n (аналогічно випадку степеневі апроксимації).

Далі задаємо рівень значущості та максимальний номер гармоніки (гармонічної складової ряду Фур'є). Для розв'язання задачі також використовуватимемо функцію `regress`. Будуємо для неї матрицю ортонормованих базисних функцій. Знаходимо параметри моделі та довірчі інтервали (однакової ширини) для них. Аналізуємо модель. Якщо довірчий інтервал для коефіцієнта апроксимації включає в себе нуль, то такий доданок можна відкинути на обраному рівні значущості q . Знаходимо ті гармоніки, які потрібно враховувати:

```
% q=0.01; % рівень значущості
% kmax=10; % максимальний номер гармоніки
T=max(x); % період
X=ones(n,1)/n^0.5; % 1-й стовпець матриці базисних функцій
for k=1:kmax % додаємо ортонормовані стовпці
    X=[X [cos(2*pi*k*x/T) sin(2*pi*k*x/T)]*(2/n)^0.5];
end
k=[0 reshape(repmat([1:kmax],2,1),2*kmax,1)']; % всі гармоніки
[b,bint]=regress(y,X,q); % регресійна модель
disp('Довірчі інтервали для параметрів моделі');
fprintf('на рівні значущості q=%5.2f\n',q)
disp('гарм    НижняГр    ВерхняГр');
fprintf('%2.0f %12.7f %12.7f\n',[k;bint]);
np=find(prod(bint,2)>0); % враховуємо базисні функції
fprintf('Враховуємо гармоніки:');
```

```
fprintf(' %d',unique(k(np)));
fprintf('\n');
```

Будуємо результуючий тригонометричний поліном. Зазвичай, його виражають не через нормовані тригонометричні функції, а через вихідні, як у прикладі, що ми і робимо. Обчислюємо також теоретичну функцію для виводу графіка:

```
mmax=max(k(np)); % максимальна гармоніка
a0=b(1)/(n/4)^0.5; % a0
ak=b(2:2:2*mmax)/(n/2)^0.5; % a(k)
bk=b(3:2:2*mmax+1)/(n/2)^0.5; % b(k)
fprintf('Тригонометричний поліном %d-го степеня:\n',mmax)
fprintf('y(x)=%f12',a0/2); % середній рівень
w=[1:mmax;ones(1,mmax)*T]; % допоміжний масив
fprintf('%+f12*cos(2*pi*d*x/%d)%+f12*sin(2*pi*d*x/%d)',[ak;w;bk;w]);
fprintf('\n');
yt=a0*ones(size(x))/2; % теоретичні значення
for k=1:mmax
    yt=yt+ak(k)*cos(2*pi*x*k/T)+bk(k)*sin(2*pi*x*k/T);
end
```

Зрештою, виводимо в одне вікно теоретичну криву та експериментальні точки (отримуємо графік, подібний до наведеного на рис.5.4):

```
plot(x,y,'.k',[0;x],[yt(end);yt],'-b')
set(get(gcf,'CurrentAxes'),'FontName','Times New Roman Cyr','FontSize',10);
title('\bfТригонометрична апроксимація')
xlabel('\itx') % мітка осі OX
ylabel('\ity') % мітка осі OY
```

Далі розглянемо розв’язання задач МНК для функції кількох змінних. Особливістю апроксимації функції кількох змінних є те, що тут ортогональність базисних функцій досягається не лише за рахунок вибору самих функцій, а й за рахунок вибору точок для експерименту (**рівнів факторів**). Тобто для багатofакторного експерименту необхідне правильне планування.

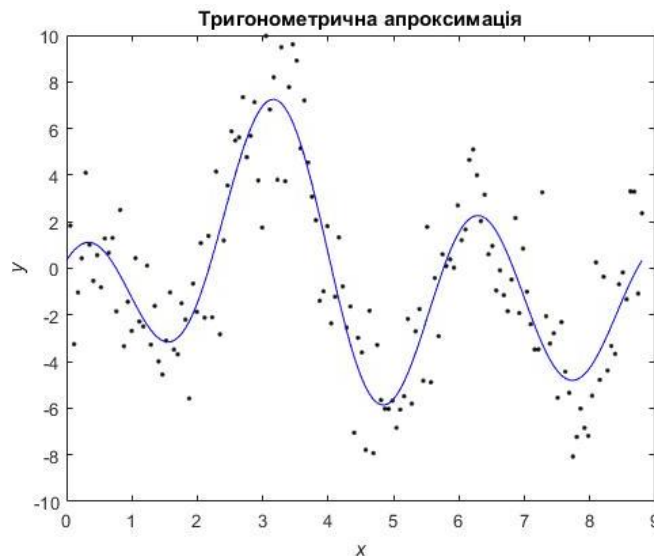


Рисунок 5.4

5.3.3 Лінійна модель для двофакторного експерименту

Перш, ніж будувати лінійну модель для 2-факторного експерименту, повторіть матеріал лекції 1.

Позначимо незалежні змінні через x та y , а результат експерименту – через z . Нехай, з певних теоретичних міркувань, z повинна бути лінійною функцією $z = b_0 + b_x x + b_y y$. На основі експериментальних даних (x_1, y_1, z_1) , (x_2, y_2, z_2) , ..., (x_n, y_n, z_n) потрібно за МНК визначити параметри лінійної моделі b_0 , b_x та b_y , тобто провести лінійну апроксимацію функції двох змінних.

Вихідними даними є експериментальні точки, і ми використовуватимемо сітку повного факторного експерименту (ПФЕ). Дані розміщуються у текстовому файлі у вигляді матриці розміром $(l+1) \times (m+1)$, причому її перший стовпець (починаючи з 2-го елемента) – це рівні фактору x , а перший рядок (починаючи з 2-го елемента) – рівні фактору y . Інша частина матриці – експериментальні значення функції z_{ij} . Спочатку введемо вихідні дані та визначимо розмірності задачі.

```
% вводимо дані та визначаємо розмірність задачі
data='filename.txt'; % ім'я файлу з даними
od=load(data); % вводимо дані
x=od(2:end,1); % рівні фактору x
y=od(1,2:end); % рівні фактору y
z=od(2:end,2:end); % результати експерименту
l=length(x);
m=length(y);
fprintf('Кількість рівнів фактору x l=%d\n',l);
fprintf('Кількість рівнів фактору y m=%d\n',m);
```

За заданими рівнями факторів побудуємо сітку ПФЕ. Для функції regress потрібно задати матрицю, кожний стовпець якої є значеннями однієї базисної функції в потрібних точках. Побудуємо таку матрицю за сіткою ПФЕ. За допомогою функції regress знайдемо коефіцієнти апроксимації та довірчі інтервали для них. Надрукуємо розв'язок:

```
% будуємо сітку повного факторного експерименту
[X,Y]=meshgrid(y,x); % сітка вузлових точок
XY=[ones(l*m,1) X(:) Y(:)]; % матриця експерименту
q=0.05; % рівень значущості
[b,bint]=regress(z(:),XY,q); % лінійна модель
disp('Довірчі інтервали для параметрів моделі');
fprintf('на урівні значущості q=%5.2f\n',q)
disp(' N   НижняГр   ВерхняГр');
fprintf('%2.0f %12.7f %12.7f\n',[[1:3];bint]);
disp('Лінійна модель:')
fprintf('z(x,y)=%f12%+f12*x%+f12*y\n',b);
```

Обчислимо за цією формулою теоретичні значення z . Зобразимо в одному графічному вікні теоретичну поверхню (площину) та експериментальні точки. Оформимо результат (напишемо заголовок, позначимо мітки осей, додамо сітку та обмежуючий паралелепіпед, оберемо точку перегляду):

```
% обчислення теоретичних значень
zt=b(1)+b(2)*X+b(3)*Y; % теоретичні аплікати
figure;
surf(X,Y,zt); % теоретична поверхня
hold on
plot3(X,Y,z,'b. '); % експериментальні точки
```



```

set(get(gcf,'CurrentAxes'),'FontName','Times New Roman Cyr','FontSize',10)
title('\bfЛінійна модель для 2-факторного експерименту')
xlabel('\itx') % мітка осі OX
ylabel('\ity') % мітка осі OY
zlabel('\itz') % мітка осі OZ
box on % обмежуючий прямокутник
grid on % сітка
view(-130,25) % обрали точку перегляду

```

В результаті отримуємо кількість рівнів факторів, довірчі інтервали для параметрів моделі на заданому рівні значущості q , аналітичний вигляд лінійної моделі, а також графік, подібний до наведеного на рис.5.5:

Кількість рівнів фактору x $l=15$

Кількість рівнів фактору y $m=18$

Довірчі інтервали для параметрів моделі
на рівні значущості $q=0.05$

N	НижняГр	ВерхняГр
1	-0.6315761	0.1016331
2	-2.8678667	-2.7157021
3	2.8352935	2.9983662

Лінійна модель:

$$z(x,y) = -0.26497212 - 2.79178412 \cdot x + 2.91683012 \cdot y$$

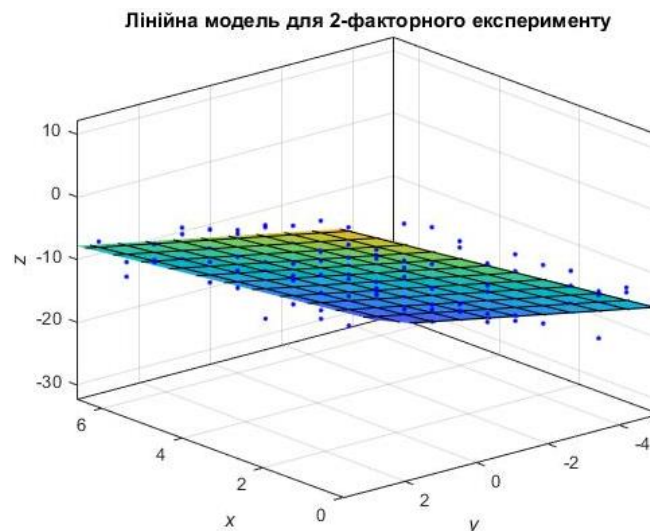


Рисунок 5.5

5.3.4 Нелінійна залежність від параметрів апроксимації

Досі в ми використовували теоретичну залежність виду $y = \sum_{j=1}^m b_j \psi_j(x)$, де

b_j – параметри апроксимації, $\psi_j(x)$ – відомі (наперед задані) функції.

Теоретична крива (або поверхня) могла нелінійно залежати від аргументів, але не завжди лінійно залежала від параметрів апроксимації b_j . Іноді теоретичні міркування говорять про інше. Наприклад, для радіоактивного розпаду швидкість розпаду речовини у даний момент часу прямо пропорційна її кількості у даний момент. Теоретична залежність маси y від часу x має вигляд $y = b_1 e^{-b_2 x}$, де $b_1 > 0$ – початкова маса, $b_2 > 0$ – параметр швидкості розпаду. Якщо задані експериментальні точки, то рівняння теоретичної кривої потрібно шукати у вигляді наведеної вище експоненціальної функції.

Розглянемо ще один приклад – насичення. За цим законом змінюється, зокрема, струм під час перехідного процесу при включенні напруги та великому опорі у ланцюзі. Рівняння кривої насичення має вигляд $y = b_1(1 - e^{-b_2 x})$, де $b_1 > 0$ – граничне значення, до якого прямує функція, $b_2 > 0$ – параметр швидкості досягнення граничного значення. Зрозуміло, можуть бути й інші види теоретичних залежностей, нелінійних відносно параметрів b_j .

Побудуємо криву, що нелінійно залежить від параметрів. Теоретичну залежність шукатимемо у вигляді кривої насичення (див. вище). Вихідні дані зберігаються у текстовому файлі у вигляді двох стовпців. Перший стовпець – це x_i , другий – y_i . Спочатку введемо дані та знайдемо обсяг вибірки (аналогічно п.5.3.1).

Далі обчислимо параметри найкращої кривої насичення та довірчі інтервали для них. Для розв’язання цієї задачі у MatLab є функція `nlinfit`, в яку одним з вхідних параметрів потрібно передати вигляд теоретичної нелінійної

функції. Це можна зробити, створивши окрему m-функцію або використовуючи дескриптор функції @. Для складних функцій зручніше перший шлях: функція для обчислення $y(x)$ записується до m-файлу у вигляді звичайної функції MatLab, а посилання на ім'я файлу використовується як вхідний параметр при виклику функції nlinfit.

```
% обчислення параметрів найкращої кривої насичення
% ff-теоретична нелінійна функція b(1)*(1-exp(-b(2)*x))
[b,r,J]=nlinfit(x,y,'ff',[1 1]); % підбір параметрів
disp('Крива насичення:');
fprintf('y(x)=%10.7f*(1-exp(%+10.7f*x))\n',b(1),-b(2));
ci=nlparci(b,r,J); % довірчі інтервали
disp('Довірчі інтервали для параметрів моделі:');
disp(' N    НижГр    ВерхГр');
fprintf('%2.0f %10.7f %10.7f\n',[[1;2],ci]);
```

Зрештою, обчислимо теоретичні значення в заданих точках x_i , а також виведемо в одне графічне вікно експериментальні точки та теоретичну криву насичення:

```
% обчислення теоретичних значень у заданих точках
yt=ff(b,x);
plot(x,y,'k.',x,yt,'b-'); % графік кривої насичення
set(get(gcf,'CurrentAxes'),'FontName','Times New Roman Cyr','FontSize',10)
title('\bfКрива насичення')
xlabel('\itx') % мітка осі OX
ylabel('\ity') % мітка осі OY
```

В результаті отримаємо значення розмірності n , аналітичний вигляд кривої насичення, довірчі інтервали для параметрів моделі а також графік, подібний до наведеного на рис.5.6.

5.5.5 Метод найменших, але не квадратів

Якщо розподіл випадкових помилок є нормальним, а вимірювання — незалежними, то з принципу максимальної правдоподібності виходить МНК.

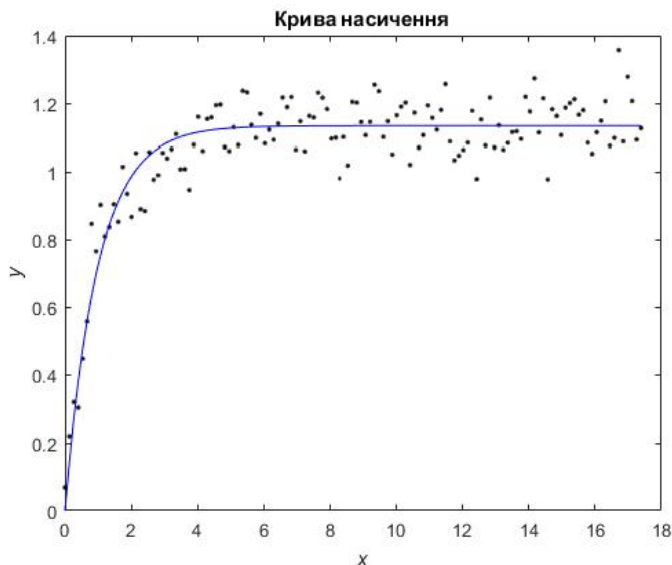


Рисунок 5.6

Проте не завжди виконуються наступні гіпотези: всі досліди незалежні; при вимірюваннях немає систематичних помилок та промахів (грубих помилок); помилка вимірювань розподілена за нормальним законом; всі досліди однаково точні. Тому іноді замість мінімізації суми квадратів відхилень експериментальних точок від теоретичних значень доводиться мінімізувати зовсім іншу величину: суму модулів відхилень; максимальне за модулем відхилення; суму модулів кубів або вищих степенів відхилень; іншу величину, що характеризує відхилення експериментальних точок від теоретичної кривої.

Будь-яку з цих задач легко можна розв'язати, наприклад, за допомогою MatLab. Так, розглянемо перші дві з наведених вище задач – мінімізацію суми модулів відхилень та максимального за модулем відхилення. При розв'язанні обмежимося поліномом 2-го степеня. Як і в п.5.3.1, вважатимемо, що вихідні дані знаходяться в текстовому файлі у вигляді двох стовпців: 1-й – аргументи x_i , 2-й – експериментальні значення y_i .

Спочатку введемо вихідні дані та знайдемо параметри квадратичної моделі за МНК. Нас цікавить максимальний утримуваний степінь, оскільки

для інших задач потрібно буде задавати функцію, що мінімізується, й ми задаватимемо її у вигляді поліному такого ж степеня.

Як і у п.п.5.3.1 та 5.3.2, виконуємо наступні дії:

```
% очищуємо пам'ять
% вказуємо ім'я файлу з даними
% вводимо дані з файлу – 2 стовпця: аргументи та функції
% знаходимо кількість точок – обсяг вибірки n
...
p=polyfit(x,y,2); % знаходимо параметри квадратичної моделі за МНК
fprintf('Поліном 2-го степеня за МНК:\ny(x)=')
fprintf('%+f14*x^%d',[p:[2:-1:0]]);
fprintf('\n');
```

Створимо m-функцію `fun1` для обчислення суми абсолютних відхилень експериментальних точок від теоретичних значень. Ця функція мінімізуватиметься по аргументах b , а змінні x та y – параметри. Мінімізуємо її за допомогою `fminsearch`. За початкові наближення коефіцієнтів b_j беремо значення, отримані за МНК. Друкуємо знайдений апроксимуючий поліном.

```
function ff=fun1(b,x,y)
ff=sum(abs(y-b(1)*x.^2-b(2)*x-b(3)));
end

% мінімізація суми модулів відхилень
p1=fminsearch(@fun1,p,[],x,y);
% друк знайденого апроксимуючого поліному 2-го степеня за сумою модулів
% відхилень
...

```

Аналогічно чинимо і в другому випадку. Тільки мінімізуватиметься не сума абсолютних відхилень, а максимальне абсолютне відхилення.

```
function ff=fun2(b,x,y)
ff=max(abs(y-b(1)*x.^2-b(2)*x-b(3)));
end

% мінімізація максимального за модулем відхилення
p2=fminsearch(fun2,p,[],x,y);
% друк знайденого апроксимуючого поліному 2-го степеня за максимумом
% модуля
...

```

Далі обчислюємо теоретичні значення та виводимо експериментальні точки й будуємо всі 3 графіки в одному графічному вікні.

% обчислення теоретичних значень

yt=polyval(p,x); % теоретичні значення за МНК

yt1=polyval(p1,x); % теоретичні значення за сумою модулів

yt2=polyval(p2,x); % теоретичні значення за max модуля

% побудова експериментальних точок та графіків теоретичних значень

...

В результаті отримуємо аналітичний вигляд поліномів 2-го степеня за: МНК, мінімумом суми модулів та мінімумом максимуму модуля, а також графік, подібний до наведеного на рис.5.7.

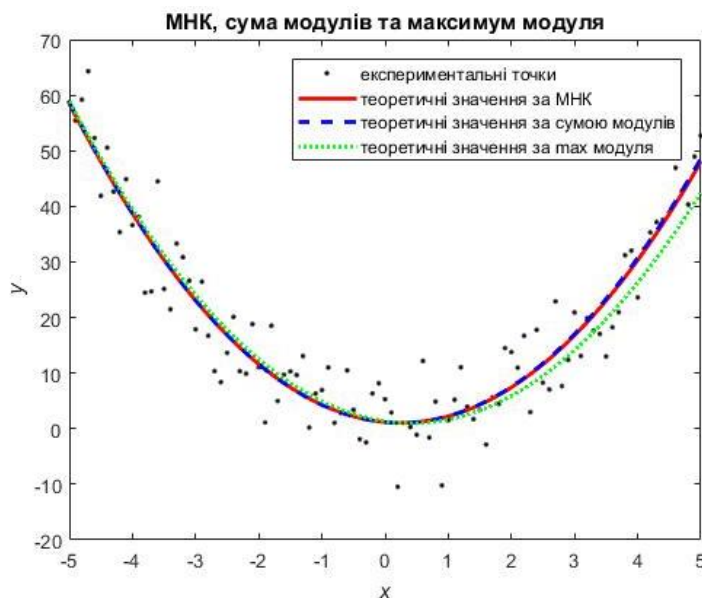


Рисунок 5.7

Подібний до наведеного вище аналізу даних можна також проводити за допомогою Curve Fitting Toolbox.

Контрольні питання

1. Проблема об'єднання результатів різних вимірювань.
2. Зважене середнє.
3. Принцип максимальної правдоподібності.
4. Апроксимація методом найменших квадратів.
5. Реалізація задач регресійного аналізу в системі MatLab.