Міністерство освіти і науки України

Львівський національний університет імені Івана Франка

Факультет прикладної математики та інформатики

Кафедра прикладної математики

**Застосування методу Крігінга для побудови сурогатних моделей**

Курсова робота студента 3 курсу групи ПМп-32

Лабенського Данила

Науковий керівник: доц. Щербатий М. В.

Національна шкала\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Кількість балів:\_\_\_\_\_\_\_\_\_\_ Оцінка ECTS:\_\_\_\_\_\_\_

Члени комісії \_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис) (прізвище та ініціали)

Львів 2019

**Зміст**

**Вступ**

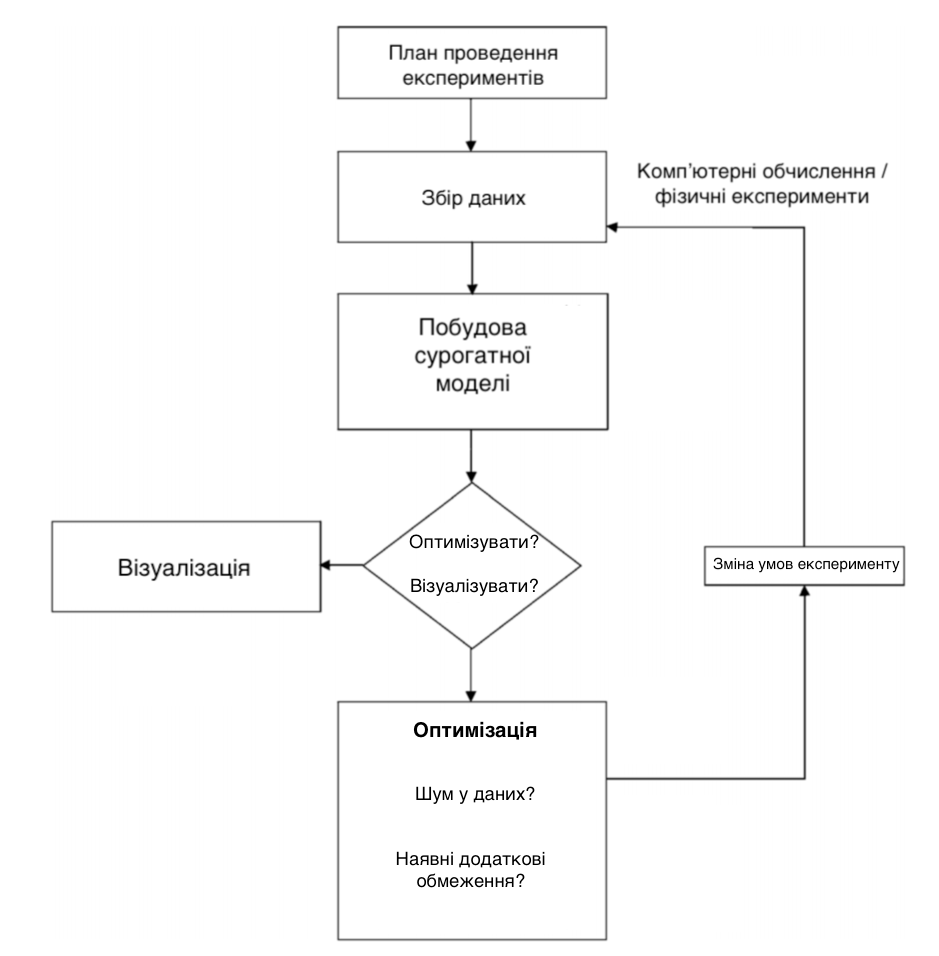
Одним із найпоширеніших застосувань математики у прикладних сферах – наприклад, інженерії – є використання математичних моделей для симуляції комплексних процесів. Водночас, основною проблемою залишається обчислювальна складність. Деякі обчислювальні експерименти можуть займати дуже великі об’єми часу (в рангах годин), і лінійний прогрес у сфері збільшення потужностей обчислювальних машин не в змозі задовільнити потребу у швидкому виконанні обчислень вищих порядків складності.

Проблема полягає також у тому, що під час процесу пошуку певних оптимальних значень для комплексної системи, характер дослідження змушує нас повторювати затратні обчислення багато разів на всій множині визначення вхідних параметрів, постійно шукаючи оптимальніше співвідношення між ними.

На таких міркуваннях базується ідея використання «сурогатної моделі» - знаходження швидкої для обчислення апроксимації для заданої складної математичної моделі. В цій роботі я розгляну основні етапи при роботі із сурогатними моделями, та опишу використання методу Крігінгу для побудови таких моделей.

1. **Процес роботи з сурогатними моделями**

Інженери використовують сурогатні моделі у багатьох сценаріях: прогноз поведінки складної для обчислень функції в області визначення; калібрація математичних моделей фізичних процесів з метою покращення точності без збільшення затрат на їх обчислення; очищення даних, отриманих внаслідок фізичних експериментів, від шуму; передбачення результатів для вхідних даних, для яких неможливо провести експеримент.

В усіх цих випадках алгоритм роботи з сурогатними моделями аналогічний. Його можна описати такою схемою: ****

1. **1 Вибір плану проведення експериментів**

Розглядатимемо певну неперервну функцію , де , для якої ми намагаємось побудувати сурогатну модель . Окрім припущення неперервності, все що ми можемо дізнатись про набір дискретних спостережень

.

Оскільки ці значення складні для обчислень, наше завдання – підібрати оптимальний для побудови хорошої моделі план проведення експериментів . Погано спроектований набір даних може привести до моделі, яка акуратна в своїх передбаченнях лише на певних ділянках .

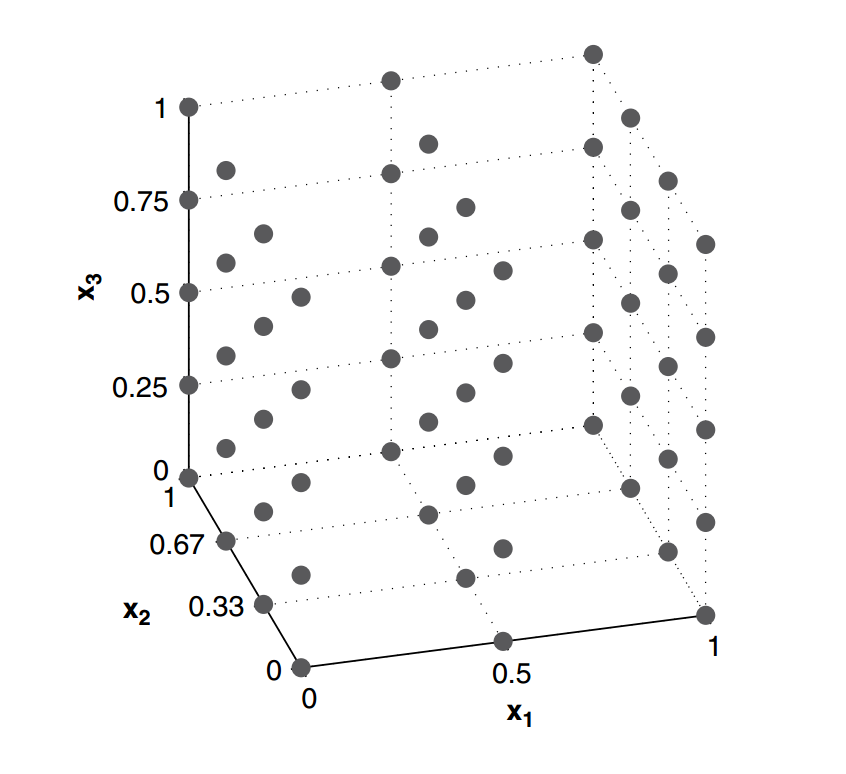
Загальною характеристикою всіх методів побудови сурогатних моделей є те, що вони значно точніші у своїх передбаченнях в околах Тому, основною метою при створенні плану проведення експериментів є рівномірний розподіл точок по всьому . План, що задовільняє такій вимозі, називатимемо таким, що заповнює простір (space-filling). (*Forrester A.I.J., Keane A.J. Engineering Design via Surrogate Modelling. 2008 – 1.4.1*)

1. **1. 1 Рівномірне розбиття простору**

Найпростіший спосіб задати подібне розбиття простору – заповнити його -вимірною сіткою рівновіддалених точок. Такий план, очевидно, заповнить простір, однак він містить ряд недоліків.

По-перше, якщо для досягнення певної точності необхідно точок у одному вимірі, для простору кількість обчислень функції зростає до . Такий план називається повним факторним експериментом. Оскільки ми прагнемо мінімізувати затрати на обчислення при підготовці даних, така тенденція є великим мінусом.

По-друге, план, створений описаним методом, негнучкий у своєму заповненні простору. На діаграмі 1.1 зображено рівномірне розбиття трьохвимірного простору. При проекції на осі, точки з цього розбиття накладатимуться, тому можна стверджувати, що для кожної із трьох змінних план може бути покращений певним зсувом точок вдовж відповідної осі.

****

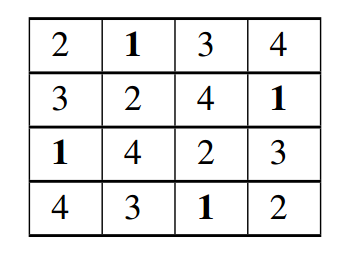
**Діаграма 1.1**

Цього можна досягти, розбивши відрізок допустимих значень цієї змінної на велику кількість рівних інтервалів, та заповнивши їх точками експерименту випадковим чином. Отриманий таким чином план називають *випадковою стратифікованою пробою*. Узагальнюючи цю ідею для всіх змінних, отримаємо техніку побудови плану експерименту, що називається *Латинський гіперкуб*.

1. **1. 2 Латинський гіперкуб**

Для опису побудови Латинського гіперкубу для -вимірного випадку, спершу проілюструємо його побудову для двохвимірного простору.

Якщо для досягнення бажаної точності моделі необхідно точок проведення експерименту, простір розбивається на підпростори рівномірною сіткою, після чого кожен рядок заповняється перестановкою так, щоб кожне число з’являлось у кожному рядку і стовпці лише один раз (Діаграма 1.2).

**

**Діаграма 1.2**

Після цього з кожного підпростору, позначеного, наприклад, **1**, вибирається одна точка у план проведення експерименту довільним чином. Такий квадрат називається *Латинським квадратом*, його відповідний -вимірний аналог – Латинським гіперкубом.

Побудований таким методом план гарантує, що при проекції на жодну з осей точки не будуть накладатись одна на одну. Це, однак, не гарантує того, що кожен такий гіперкуб заповнюватиме простір: наприклад, ми можемо отримати куб, де всі точки експерименту розміщені на головній діагоналі. Щоб характеризувати те, наскільки гіперкуб заповнює простір, скористаємось метрикою *мінімакс* (*Johnson et al. 1990*).

Нехай унікальні значення відстаней між усіма можливими парами точок у , посортовані у порядку зростання; набір значень, де кількість пар точок з , відстань між якими рівна . Ми називатимемо ***мінімаксним*** із всіх можливих планів, якщо в нього максимальне можливе і, серед всіх планів для яких ця умова виконується, мінімальне .

Відстанню між двома точками у вважатимемо *p-норму*:

У літературі немає переконливих демонстрацій переваги норми для певного *p* у контексті оцінки мінімаксом (*Forrester A.I.J., Keane A.J.*). Оскільки обчислення даної норми є найлегші при , надалі використовуватимемо її, якщо не вказано інакше.

Варто зазначити, що таке означення не завжди однозначно визначає мінімаксний план. Тому, скористаємося доповненням до означення, запропонованих *Моррісом і Мітчелом у 1995*:

– ***мінімаксний*** із всіх можливих планів, якщо в нього максимальне можливе , серед всіх планів, для яких ця умова виконується, мінімальне , серед всіх планів, для яких ця умова виконується, максимальне , … , мінімальне

Залишається додати, що задля спрощення обчислень та уникнення проблем із великими числами, перед початком проведення будь-яких обчислень ми нормалізовуємо до одиничного кубу .

1. **2 Побудова сурогатної моделі**

Отримавши з кроку **1.1** набір з пар значень , ми шукаємо функцію з простору функцій . Процес такого пошуку називається *контрольованим навчанням*.

Звісно, переважна більшість функцій з цього простору наближатимуть дуже погано; ми можемо значно звузити коло пошуку певними обмеженнями – наприклад, покласти що повинна бути неперервна – однак нам все одно доведеться обирати із безлічі функцій, що потенційно нам підходять.

Слід зазначити, що з метою тестування слід обрати певний набір тестових точок, незалежний від тренувального набору , та обчислити точне значення у цих точках. Його рекомендований обсяг - . При моделюванні реальних задач, ці додаткові затрати на обчислення потрібно враховувати при оцінці допустимого обсягу . Ці дані не слід використовувати під час тренування моделі, адже вони потрібні, щоб дати оцінку, наскільки добре передбачає значення у точках, на яких не проводилось навчання. Цей набір обиратимемо також за допомогою Латинського гіперкубу.

Для оцінки того, наскільки добре знайдена наближає , існує цілий ряд характеристик. Найчастіше вживаними із них є:

* Максимальне абсолютне відхилення ( розмір тестового набору даних):
* Середньоквадратичне відхилення:
* Коефіцієнт кореляції:

Щодо RMSE, можна вважати, що якщо його значення, поділене на розкид значень серед тестових , менше *0,1* (ця величина матиме значення від 0 до 1), модель дає достатньо хороші передбачення. Метрика не залежить від області значень *f* ; вона набуває значень від 0 до 1 та характеризує топологію функцій, не їх конкретні значення. Чим ближче до 1, тим краще модель передбачує функцію; приблизно відповідає RMSE = 0.1.

1. **Крігінг**

Метод *Крігінгу* – частковий випадок методу *радіальних базисних функцій* з базисною функцією виду

(Опис методу радіальних базисних функцій (надалі РБФ) знаходиться за межами даної роботи, його можна знайти у, наприклад, *Zhong-Hua Han and Ke-Shi Zhang (2012). Surrogate-Based Optimization, Real-World Applications of Genetic Algorithms*).

Особлива відмінність цієї базисної функції – наявність наборів коефіцієнтів та , що дозволяє їй змінювати свою поведінку для кожної змінної. Наприклад, при та константою для всіх просторів, метод Крігінга перетворюється у метод РБФ з базисною функцією Гауса. Таким чином, Крігінг накладає найменше припущень на завдяки більшій кількості параметрів моделі – припускається, що вона гладка та нерозривна.

Варто зазначити, що розглядається *інтерполяція Крігінгом.* Відома в топології *регресія Крігінгом* лежить поза межами даної роботи.

**2.1 Побудова моделі Крігінгом**

Ідея методу Крігінгу полягає в тому, що ми розглядаємо отримані значення як значення, отримані внаслідок стохастичного (випадкового) процесу. Позначимо це набором випадкових векторів

Нехай цей набір має середнє значення , вектор-стовпець одиниць. Випадкові змінні корелюються одна з одною через базисну функцію:

Відповідно, ми можемо побудувати матрицю кореляції для всіх вхідних даних:

Саме для введення такої кореляції між даними нам необхідне припущення про гладкість та неперервність *f*. Значення у матриці залежать від абсолютних відстаней між точками та параметрів і . Більші значення відповідають більшій кореляції між точками по *j*-ій змінній; параметри характеризують вплив *j*-ої змінної на Y чим більший , тим більший вплив має на значення Y *j*-та змінна.

Щоб побудувати сурогатну модель, ми намагатимемося обрати такі і , щоб максимізувати вірогідність **y**, отримуючи таким чином , що максимізує вірогідність отримати правильні значення *y* на всьому .

Для подальших міркувань, опишемо процес оцінки моделі за допомогою методу максимальної вірогідності.

Вірогідність – величина, що характеризує імовірність того, що набір даних було отримано із функції , де певний невеликий сталий окіл допустимої похибки кожного із значень. Наприклад, припустивши що випадково незалежно розподілені згідно нормального розподілу із стандартним відхиленням , вірогідність набору даних є

Ми прагнемо максимізувати правдоподібність моделі для набору даних **x**, а для зручності обчислень – мінімізувати її від’ємний натуральний логарифм:

Оскільки ми припускаємо що модель, побудована інтерполяцією Крігінга, не містить похибки , імовірність можна записати наступним чином:

Що можна перевиразити через наші дані як

Запишемо логарифм, який необхідно максимізувати:

Прирівнюючи похідні рівняння (2.3) до нуля, отримуємо наступні найбільші вірогідності для і

Тепер ми можемо підставити отримані значення назад у (2.3) і прибрати доданки-константи для отримання *функції концентрованої логарифмічної вірогідності*:

Значення цієї функції залежить від шуканих нами і Для її максимізації, оптимально буде використати техніку числової оптимізації – наприклад, глобальний пошук, що уже реалізований в середовищі Matlab. (*Forrester A.I.J., Keane A.J.*)

**2.2 Передбачення значень Крігінгом**

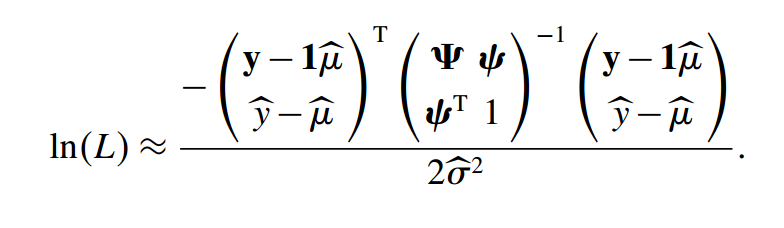
Кожне нове значення , передбачене моделлю Крігінгу, повинне сходитися з уже обчисленими для тренувального набору даних параметрами кореляції. Тому, ми додаємо до нового вектора та означаємо новий вектор кореляцій між тренувальними даними та

Тепер, сконструюємо доповнену матрицю кореляцій:

Останній елемент одиниця, адже головна діагональ заповнена одиницями ().

Запишемо логарифмічну імовірність доповнених даних:

Тільки останній доданок цього виразу залежить від тому ми можемо розглядати тільки його для максимізації. Підставивши значення і отримаємо

****

Щоб максимізувати цю рівність, нам треба обчислити Для цього, можна використати, наприклад, метод знаходження оберненої матриці Тейла (*1971*). Знайшовши розклад через та виконавши відповідні спрощення, знаходимо, що найкраща вірогідність для досягається при

Не одразу є очевидним, як рівняння (2.4) випливає з РБФ (2.1). У рівнянні (2.4), базисні функції містяться у векторі Модель побудована таким чином, що вона проходить через усі точки вхідних даних (інтерполює їх) : якщо ми спробуємо передбачити значення для ми отримаємо значення

1. **Програмна реалізація Крігінгу**

**3. Програмна реалізація**

Опис програмної реалізації (Матлаб, то-то, структура рішення)

Як запустити, що отримається

Для моделі, приймемо p=2;

Шукатимемо сігма по логарифмічній шкалі

1. **Результати числових досліджень**

Функція однієї змінної: Чи можна застосувати Крігінг?

Засікати час?

Графіки – в точці з найгіршим результатом

**Висновки**

**Список використаної літератури**