Передмова

Теорія та алгоритми оптимізації є важливою частиною арсеналу сучасного прикладного математика. У посібнику викладено основи оптимізаційного підходу в розпізнаванні образів та розглянуто теорію варіаційних нерівностей.

Курс лекцій з розпізнавання образів для студентів спеціальності «Прикладна математика» факультету кібернетики Київського національного університету імені Тараса Шевченка, що викладався у 2012–2014 рр. протягом двох семестрів, складається з двох частин, які умовно можна назвати «Розпізнавання як оптимізація» і «Розпізнавання як перевірка статистичних гіпотез». У посібник увійшли лекції з першої частини курсу — розділи 1–10. Останні два розділи містять матеріал курсу лекцій з теорії варіаційних нерівностей для студентів магістерської програми «Прикладна математика».

Попередні знання, які вимагаються для розуміння матеріалу, відповідають тому середньому рівню, що досягається після трьох років навчання на факультетах кібернетики та прикладної математики класичних університетів України.

Автори глибоко вдячні колегам і студентам за плідні наукові дискусії та зауваження щодо змісту, передусім, В'ячеславові Алєксєєнку, Юрію Маліцькому та Марині Присяжній.

Хочемо вшанувати видатного вченого та унікальну людину, професора Юрія Івановича Петуніна, який був для авторів одним з учителів та з яким вони мали щастя працювати майже 20 років.

Також В. В. Семенов вдячний ДФФД України за підтримку досліджень, результати яких згадуються у посібнику.

Зауваження та побажання можна надсилати електронною поштою: dokmed5@gmail.com, semenov.volodya@gmail.com.

Розділ 1.

Основні поняття розпізнавання образів

1.1. Основні концепції

Розпізнавання образів — це розділ теорії штучного інтелекту, що вивчає методи комп'ютерної класифікації об'єктів, тобто методи автоматичної ідентифікації одного з наперед заданих класів (двох або більше), якому належить об'єкт.

Метою перших досліджень у цьому напрямі була реалізація органів зору в роботів, тому, за традицією, об'єкт, що підлягає класифікації, називається ***образом***. Образом може бути цифрова фотографія (розпізнавання зображень), літера або цифра (розпізнавання символів), аудіозапис (розпізнавання мови) тощо.

Важливою складовою класифікації є процедура машинного навчання, метою якої є побудова ***вирішального правила***, що класифікує об'єкти за ознаками. Відповідно до того, який аспект розпізнавання образів вибирається за основу, машинне навчання часто розглядають або як геометричну теорію, метою якої є пошук лінійної або нелінійної поверхні, що розділяє задані множини прецедентів у просторі ознак, або як статистичну теорію, метою якої є пошук оптимальної апроксимації

вирішальної функції за допомогою функцій із наперед заданої множини навчальних вибірок.

Другою частиною процесу класифікації є ***процедура узагальнення***, яка полягає у класифікації нової вибірки за допомогою вирішального правила, сформульованого в процесі навчання.

Отже, процес класифікації складається з двох етапів: 1) індуктивного, що полягає в навчанні, тобто отриманні загальної інформації (вирішального правила) з часткових спостережень (навчальних вибірок), і 2) дедуктивного, що полягає в застосуванні знайденого вирішального правила для класифікації нових об'єктів.

Розпізнавання = навчання + узагальнення

Перейдемо до опису математичного формалізму, що лежить в основі розпізнавання образів.

Означення 1. Нехай X — множина об'єктів, $\Omega = \{\omega_2, \omega_2, ..., \omega_N\}$ — множина міток класів $C_1, C_2, ..., C_N$, $f : X \to \Omega$ — цільова функція, значення якої відомі лише на скінченній підмножині об'єктів $\{x_1, x_2, ..., x_m\} \subset X$. Пари (x_i, ω_i) називаються ***прецедентами***, а сукупність пар $T = \{(x_i, \omega_i)\}_{i=1}^m$ — ***навчальною вибіркою***.

Задача навчання за прецедентами полягає в тому, щоб за вибіркою T відновити функцію f, тобто побудувати вирішальну (індикаторну) функцію $g : X \to \Omega$, що в певному розумінні найкраще наближає цільову функцію $f : X \to \Omega$ не лише на об'єктах $\{x_1, x_2, ..., x_m\}$, а й на всій множині X. У задачі класифікації на N диз'юнктних класів множина міток визначається однозначно. У випадку $N = 2$ класифікація називається ***бінарною***.

Означення 2. ***Ознака об'єкта*** — це результат вимірювання числової або категорійної характеристики. З формального погляду ознака є відображенням $x : X \to D$, де D — множина допустимих значень ознаки.

Означення 3. Вектор ознак $(x_1, x_2, ..., x_n)$ називається ***ознаковим описом об'єкта***, де $x_i \in D_i$.

В основу теорії розпізнавання образів покладено два постулати.

1. **Постулат про векторну модель:** об'єкт можна подати як елемент векторного простору ознак.

2. **Постулат про компактність:** переважна більшість об'єктів, що належать до одного класу, є ближчими один до одного, ніж до об'єктів іншого класу, і лежать в області з відносно простою межею.

Виходячи з цих постулатів, навчання можна описати як задачу пошуку поверхні, що відокремлює множини розмічених точок у векторному просторі. Оскільки таких поверхонь може бути безліч, то виникає задача про пошук оптимальної за певним критерієм відокремлювальної поверхні. Позначимо множину допустимих відокремлювальних поверхонь як S (наприклад, допустимою множиною поверхонь може бути множина ліній на площині або гіперплощин у багатовимірному евклідовому просторі), а критерій якості розпізнавання, або функцію втрат, як J.

Означення 4. ***Критерій якості розпізнавання*** — це невід'ємний функціонал $\mathcal{J}(s, x)$, який характеризує величину помилки при класифікації об'єкта x за допомогою відокремлювальної поверхні s. Якщо $\mathcal{J}(s, x) = 0$, то класифікація називається правильною.

Тоді задачу навчання можна сформулювати так: знайти

$$\arg \min_{s \in S, x \in T} J(s, x).$$

Зазвичай критерій якості формулюють за допомогою функції середніх втрат, або емпіричного ризику, яка характеризує кількість помилок.

Означення 5. ***Функція втрат, або емпіричний ризик*** на вибірці T вирішального правила, основаного на відокремлюваній поверхні s, має вигляд

$$J(s,T) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}(s, x_i),$$

де функція $\mathcal{L}$ набуває значення 0, якщо об'єкт x класифікується правильно, і 1, якщо об'єкт x класифікується неправильно. У цьому випадку емпіричний ризик $J(s,T)$ дорівнює частоті помилок вирішального правила, основаного на відокремлювальній поверхні s та на об'єктах навчальної вибірки.

Розпізнавання = оптимізація

На практиці якість класифікації часто характеризують такими показниками, як точність, чутливість, специфічність тощо.

Припустимо, що тестова вибірка складається з P об'єктів класу A (позначення P означає з англ. positive — позитивні об'єкти) і N об'єктів класу B (позначення N означає з англ. negative — негативні об'єкти). Якщо об'єкт, про який заздалегідь відомо, що він належить до класу A, класифікується як позитивний, то результат називається ***істинно позитивним***. Якщо об'єкт, про який заздалегідь відомо, що він належить до класу B, класифікується як негативний, то результат називається ***істинно негативним***. Відповідно помилкові результати класифікації називаються ***хибно позитивними*** й ***хибно негативними*** (FN — з англ. false negative).

Позначимо кількість істинно позитивних результатів як TP (true positive), істинно негативних — як TN (true negative), хибно позитивних — як FP (false positive) і хибно негативних результатів — як FN (false negative).

Означення 6. ***Чутливість, або TPR (true positive rate)*** — це частка істинно позитивних результатів серед усіх

позитивних результатів, тобто

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{P}.$$

Означення 7. ***Специфічність, або TNR (true negative rate)*** — це частка істинно негативних результатів серед усіх негативних результатів, тобто

$$TNR = \frac{TN}{TN + FP} = \frac{TN}{N}.$$

Означення 8. ***Точність*** — це частка правильних результатів серед усіх результатів класифікації, тобто

$$PR = \frac{TP + TN}{P + N}.$$

Якщо частку хибно позитивних результатів позначити як $FPR = \frac{FP}{N}$, то легко бачити, що $FPR + TNR = 1$.

Терміни "чутливість", "специфічність" і "точність" характерні для класифікації, яка здійснюється в медичних дослідженнях. У галузі інформаційного пошуку чутливість називають повнотою (recall), специфічність — релевантністю (relevance).

Означення 9. Крива залежності специфічності TPR від величини $1 - TNR$ при варіюванні параметрів вирішальної функції називається ***ROC-кривою.*** Ця крива характеризує якість бінарної класифікації й називається також ***кривою помилок***, а її аналіз — ***ROC-аналізом***.

Якісна інтерпретація ROC-кривої виражається площею, обмеженою ROC-кривою та віссю, на якій відкладаються частки хибних позитивних результатів. Цей показник називається AUC (area under curve — площа під кривою). Якісні класифікатори мають більш високий показник AUC. Якість класифікатора залежно від значення AUC визначається так: від 0,9 до

1,0 — відмінна, від 0,8 до 0,9 — дуже добра, від 0,7 до 0,8 — добра, від 0,6 до 0,7 — задовільна, від 0,5 до 0,6 — незадовільна (випадковий результат).

1.2. Імовірнісні концепції розпізнавання образів

Дані про об'єкти з множини X можуть бути неточними або неповними. У цьому випадку одному опису x можуть відповідати різні розв'язки. Імовірнісна постановка полягає в такому: замість невідомої цільової залежності $f(x)$ припускається існування невідомого ймовірнісного розподілу на множині $X \times \Omega$ із щільністю $p(x, \omega)$, з якого випадково й незалежно вибираються спостереження

$$T = \{(x_i, \omega_i)\}_{i=1}^{m}.$$

Такі вибірки називаються **простими**.

1.2.1. Принцип максимальної правдоподібності

При ймовірнісній постановці задачі замість моделі алгоритмів $g(x, \theta)$, яка апроксимує невідому залежність $f(x)$, задається модель сумісної щільності розподілу об'єктів і розв'язків $\varphi(x, \omega, \theta)$, що апроксимує невідому ймовірність $p(x, \omega)$. Після цього визначається значення параметра θ, при якому вибірка даних T є найбільш правдоподібною, тобто найкраще узгоджується з моделлю щільності. Якщо спостереження у вибірці T є незалежними, то сумісна щільність усіх спостережень дорівнює добутку значень щільності $p(x, \omega)$ для кожного спостереження:

$$p(T) = \prod_{i=1}^{m} p(x_i, \omega_i).$$

Якщо апроксимувати $p(x_i, \omega_i)$ моделлю щільності $\varphi(x_i, \omega_i, \theta)$, то отримуємо функцію правдоподібності

$$L(\theta, T) = \prod_{i=1}^{m} \varphi(x_i, \omega_i, \theta).$$

Що більше значення функції правдоподібности, то краще вибірка узгоджується з моделлю. Отже, треба шукати

$$\arg\max_{\theta} L(\theta, T).$$

Цей метод називається **принципом правдоподібності**.

1.2.2. Мінімізація емпіричного ризику

Замість максимізації функції правдоподібності $L(\theta, T)$ зручніше мінімізувати функціонал $-\ln L(\theta, T)$, оскільки він є адитивним за об'єктами вибірки.

$$-\ln L(\theta, T) = -\sum_{i=1}^{m} \ln \varphi(x_i, \omega_i, \theta) \to \min_{\theta}.$$

Означення 10. ***Імовірнісна функція втрат*** задається формулою

$$\mathcal{L}(a_\theta, x) = -m \ln \varphi(x_i, \omega_i, \theta).$$

Що гірше пара (x_i, ω_i) узгоджується з моделлю φ, то менше значення щільності $\varphi(x_i, y_i, \theta)$ і більше величина втрати $\mathcal{L}(a_\theta, x)$, і навпаки, для багатьох функцій втрат існує така модель щільності $\varphi(x, \omega, \theta)$, що мінімізація емпіричного ризику є еквівалентною максимізації правдоподібності [4].

1.3. Перенавчання та здатність до узагальнення

Мінімізація емпіричного ризику має певні особливості, а саме: якщо мінімум функціонала якості $J(s,T)$ досягається на алгоритмі g, то це не гарантує, що алгоритм g буде добре наближати цільову залежність на довільній контрольній вибірці

$$K = \left\{\left(x_i', \omega_i'\right)\right\}_{i=1}^{n}.$$

Погіршення якості роботи алгоритму на об'єктах, які не входили до навчальної вибірки, може бути наслідком наднавчання.

Приклад 1. Припустимо, що ми розпізнаємо нові об'єкти лише тоді, коли вони точно збігаються з об'єктами з навчальної вибірки. У цьому випадку емпіричний ризик дорівнює нулю, але точність розпізнавання інших вибірок теж дорівнює нулю. Навчання — це не лише запам'ятовування, але й узагальнення.

Означення 11. ***Узагальнена здатність методу*** μ характеризується мінімальною кількістю помилок на простих навчальних і контрольних вибірках, отриманих з однієї генеральної сукупності X.

Означення 12. Метод навчання μ називається ***слушним***, якшо при заданих достатньо малих числах ε і η імовірність того, що узагальнена здатність методу більше ε, менше η.

Означення 13. Параметр ε називається ***точністю методу*** μ, а параметр $1{-}\eta$ — його ***надійністю***.

Коли неможливо отримати теоретичні, застосовуються емпіричні оцінки. Нехай дано вибірку

$$S = \{(x_i, \omega_i)\}_{i=1}^{M}.$$

Розіб'ємо її N способами на диз'юнктні підвибірки: навчальну

$$T_j = \{(x_i, \omega_i)\}_{i=1}^{m}$$

і контрольну

$$K_j = \{(x_i, \omega_i)\}_{i=1}^{n},$$

де $n + m = M$.

Означення 14. Для кожного розбиття $j = 1, 2, ..., N$ побудуємо алгоритм

$$a_j = \mu(T_j)$$

і обчислимо кількість помилок. Середня арифметична кількість помилок по всіх розбиттях називається ***оцінкою кросвалідації***.

Розділ 2.

Байєсівський метод класифікації

2.1. Байєсівська класифікація з мінімальною ймовірністю помилок

Одним з найпопулярніших підходів до класифікації є байєсівський метод, описаний у численних монографіях, наприклад [5, 7, 22, 38]. Нехай $\omega_1, \omega_2, ..., \omega_N$ — мітки класів $C_1, C_2, ..., C_N$ з відомими *апріорними* ймовірностями $p(\omega_1), p(\omega_2), ..., p(\omega_N)$. Метою бійєсівської класифікації є мінімізація помилок на підставі інформації про щільності розподілів кожного класу. Для цього необхідно знайти для заданого об'єкта x найбільш правдоподібний клас, тобто мітку ω_i таку, що виконується умова

$$p(\omega_i|\, x) > p(\omega_j|\, x) \;\; \forall i \neq j, \; i, j = 1, ..., N. \tag{2.1}$$

Інакше кажучи, вирішальне правило відносить об'єкт x до класу C_i, якщо *апостеріорна* ймовірність класу C_i для цього об'єкта є максимальною. Відповідно до цього правила простір ознак Ω розділяється на N областей $\Omega_1, \Omega_2, ..., \Omega_N$: якщо вектор ознак x належить області Ω_i, то він належить класу C_i.

Отже, задача полягає в тому, щоб обчислити апостеріорні ймовірності класів, знаючи їх апріорні ймовірності. Для цього використовується формула Байєса.

Означення 15. ***Формула Байєса***:

$$p(\omega_i|\, x) = \frac{p(x|\omega_j)\, p(\omega_i)}{p(x)} \quad \forall i \neq j,\ i,j = 1,...,N. \tag{2.2}$$

Формула (2.2) дозволяє переписати вирішальне правило (2.1) як правило Байєса з мінімальною ймовірністю помилки.

Означення 16. ***Правило Байєса з мінімальною ймовірністю помилки***

$$p(x|\, \omega_i)p(\omega_i) > p(x|\, \omega_j)p(\omega_j) \quad \forall i \neq j, j = 1,...,N. \tag{2.3}$$

У випадку бінарної класифікації можна обчислити ***відношення правдоподібності***:

$$l(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)}. \tag{2.4}$$

У такому випадку вирішальне правило набуває вигляду

$$l(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{p(\omega_1)}{p(\omega_2)}. \tag{2.5}$$

Покажемо, що байєсівська класифікація дійсно мінімізує ймовірність помилки [38].

Теорема 1. *Правило Байєса є оптимальним, тобто воно мінімізує ймовірність помилки.*

Доведення. Повернемось до загального випадку з класами $C_1, C_2, ..., C_N$ і запишемо ймовірність помилки. Позначимо подію, що означає помилкову класифікацію, як e, а ймовірність помилкової класифікації об'єкта x із класу ω_i — як $p(e|\omega_i)$:

$$p(e) = \sum_{i=1}^{N} p(e|\,\omega_i) p(\omega_i)\,. \tag{2.6}$$

Позначимо через $\Omega \backslash \Omega_i$ доповнення області Ω_i. Тоді ймовірність помилкової класифікації виражається формулою

$$p(e|\,\omega_i) = \int\limits_{\Omega \backslash \Omega_i} p(x|\,\omega_i) dx. \tag{2.7}$$

Формула (2.7) дозволяє записати ймовірність помилок:

$$\begin{aligned} p(e) &= \sum_{i=1}^{N} \int\limits_{\Omega \backslash \Omega_i} p(x|\omega_i)\, p(\omega_i)\, dx = \\ &= \sum_{i=1}^{N} p(\omega_i) \left(1 - \int\limits_{\Omega_i} p(x|\omega_i)\, dx \right) = \\ &= 1 - \sum_{i=1}^{N} p(\omega_i) \int\limits_{\Omega_i} p(x|\omega_i)\, dx. \end{aligned}$$

Звідси випливає, що розбиття області Ω на області Ω_i, $1 = 1, ..., N$, що мінімізує ймовірність помилки, еквівалентне максимізації величини $\sum\limits_{i=1}^{N} p(\omega_i) \int\limits_{\Omega_i} p(x|\omega_i)\, dx$.

Це означає, що мінімізація ймовірності помилки еквівалентна максимізації ймовірності правильної класифікації. Позначимо ймовірність правильної класифікації об'єкта x із класу C_j як q. Для того, щоб вона була максимальною, слід знайти таку область Ω_i, де величина $p(\omega_i)\, p(x|\omega_i)$ є максимальною:

$$q = \int\limits_{\Omega} \max_i p(\omega_i)\, p(x|\omega_i)\, dx.$$

Отже, імовірність помилки байєсівської класифікації

$$p(e) = 1 - q = 1 - \int\limits_{\Omega} \max_i p(\omega_i)\, p(x|\omega_i)\, dx.$$

Що й потрібно було довести. □

2.2. Байєсівська класифікація з мінімальним середнім ризиком

У попередньому підрозділі ми розглянули задачу мінімізації помилкової класифікації, нехтуючи тим фактом, що помилкова класифікація може бути пов'язана з певними витратами, що залежать від класу. Наприклад, з одного боку, якщо здорова людина буде класифікована як хвора, то вона може зазнати витрат на непотрібні ліки або навіть на лікування від їх побічних наслідків, а з іншого боку, класифікація хворої людини як здорової може призвести до дуже серйозних ускладнень. Отже, необхідно розв'язати задачу мінімізації середнього ризику за допомогою байєсівського підходу.

Уведемо до розгляду величину λ_{ij} — вартість наслідків неправильної класифікації об'єкта, що належить класу C_i, коли його відносять до класу C_j.

Означення 17. ***Умовним ризиком*** віднесення об'єкта x до класу C_j називається функціонал

$$\mathcal{L}_j(x) = \sum_{i=1}^{N} \lambda_{ij} p(\omega_i|x)\, dx.$$

Означення 18. ***Середнім ризиком*** віднесення об'єкта x до класу C_j називається функціонал

$$\mathcal{R}_j(x) = \int\limits_{\Omega_j} \mathcal{L}_j(x)\, p(x)\, dx = \int\limits_{\Omega_j} \sum_{i=1}^{N} \lambda_{ij} p(\omega_i|x)\, p(x)\, dx.$$

Означення 19. ***Ризиком*** класифікації об'єкта x називається функціонал

$$\mathcal{R}(x) = \sum_{j=1}^{N} \mathcal{R}_i(x) = \sum_{j=1}^{N} \int_{\Omega_j} \sum_{i=1}^{N} \lambda_{ij} p(\omega_i|x)\, p(x)\, dx.$$

Вирішальне правило формулюється так, щоб мінімізувати ризик на вибраних областях Ω_i. Якщо

$$\int_{\Omega_i} \sum_{j=1}^{N} \lambda_{ij} p(\omega_j|x)\, p(x)\, dx < \int_{\Omega_i} \sum_{j=1}^{N} \lambda_{kj} p(\omega_j|x)\, p(x)\, dx$$

при $k = 1, 2, ..., N$, то $x \in \Omega_i$.

Таким чином, задача зводиться до пошуку областей Ω_i, на яких досягається мінімум ризику:

$$\int_{\Omega} \min_i \sum_{j=1}^{N} \lambda_{ij} p(\omega_j|x)\, p(x)\, dx.$$

2.3. Оцінка щільності розподілу

В обох варіантах байєсівської класифікації — з мінімізацією ймовірності помилок і ризику, відповідно, — ми вважали відомими апріорні ймовірності класів $p(\omega_i)$ і умовні ймовірності $p(x|\omega_i)$. На практиці апріорні ймовірності класів $p(\omega_i)$ вважаються відомими, натомість умовні ймовірності оцінюються на підставі спостережень — багатовимірних вибірок $X_i = (x_{ij})$, $i = 1, ..., N$; $j = 1, ..., n_i$.

Таким чином, постає проблема оцінювання невідомої щільності умовної ймовірності. Для цього існують два підходи: параметричний і непараметричний.

2.3.1. Параметрична оцінка щільності

Параметричний підхід полягає у припущенні, що умовна щільність має певний відомий вигляд $p(x, \theta)$ (наприклад, є рівномірною або нормальною), але з невідомим параметром θ. Цей параметр обчислюється за спостереженнями й наближається числом $\hat{\theta}$.

Приклад 2. Припустимо, що

$$p(x|\omega_i) = \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2}(x-\mu_i)^T det\left(\Sigma_i^{-1}\right)(x-\mu_i)\right\},$$

де μ_i — математичне сподівання класу C_i, Σ_i — коваріаційна матриця класу C_i.

Об'єкт x належить до класу C_i, якщо на цьому класі мінімізується функція

$$\log(p(\omega_i|x)) = \log(p(x|\omega_i)) + \log(p(\omega_i)) - \log(p(x)) =$$

$$= -\frac{1}{2}(x-\mu_i)^T det\left(\Sigma_i^{-1}\right)(x-\mu_i) -$$

$$-\frac{1}{2}\log\left(det\left(\Sigma_i^{-1}\right)\right) - \frac{p}{2}\log(2\pi) + \log(p(\omega_i)) - \log(p(x)).$$

Зважаючи на те, що рішення повинне залежати від класу, ми можемо знехтувати доданками, що не залежать від класу. У результаті отримуємо функцію

$$g_i(x) = \log(p(\omega_i)) - \frac{1}{2}\log\left(det\left(\Sigma_i^{-1}\right)\right) -$$

$$-\frac{1}{2}(x-\mu_i)^T det\left(\Sigma_i^{-1}\right)(x-\mu_i).$$

Вирішальне правило *нормального дискримінантного аналізу* формулюється так: якщо $g_i(x) > g_j(x)\ \forall i \neq j$, то $x \in C_i$.

Маючи навчальну вибірку об'єктів з кожного класу, можемо оцінити математичне сподівання й коваріаційну матрицю

класу:

$$m = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{ik},$$

$$\Sigma_i = \frac{1}{n_i} \sum_{k=1}^{n_i} (x_{ik} - m)(x_{ik} - m)^T.$$

Зауважимо, що в загальному випадку задача пошуку параметрів багатовимірного розподілу може стати дуже складною, тому на практиці часто використовують "наївний" байєсівський підхід, який полягає у припущенні, що всі ознаки об'єктів є незалежними й однаково розподіленими випадковими величинами. Звісно, таке припущення є занадто сильним (тому підхід і називають наївним), але воно спрощує задачу відновлення щільності розподілу, зводячи її до одновимірного випадку. У наступному підрозділі ми розглянемо найзагальніший спосіб розв'язання цієї задачі, в якому немає припущення щодо типу розподілу.

2.3.2. Непараметрична оцінка щільності

Розглянемо класичну оцінку Парзена [22] для одновимірної щільності ймовірності. Нехай $x_1, x_2, ..., x_N$ — незалежні й однаково розподілені випадкові величини. Оцінимо функцію їх розподілу:

$$\hat{P}_N(x) = \frac{\sharp\{x_k \leq x\}}{N}. \tag{2.8}$$

Оцінку щільності ймовірності запишемо відповідно до її означення:

$$\hat{p}_N(x) = \frac{\hat{P}_N(x+h) - \hat{P}_N(x-h)}{2h}. \tag{2.9}$$

Виходячи з властивостей щільності розподілу, необхідно, щоб величина h прямувала до 0, але оптимальний вибір швид-

кості її збіжності повинен бути узгодженим зі статистичними властивостями оцінки.

Перепишемо оцінку (2.9) таким чином:

$$\hat{p}_N(x) = \frac{1}{2h}\int\limits_{x-h}^{x+h} d\hat{P}_N(x) = \int\limits_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-\xi}{h}\right) d\hat{P}_N(\xi) = \\ = \frac{1}{hN}\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right), \quad (2.10)$$

де

$$K(y) = \begin{cases} \frac{1}{2}, & \text{якщо } y \le 1, \\ 0, & \text{якщо } y > 1. \end{cases} \quad (2.11)$$

Означення 20. Функція $K(y)$ називається *ядром* оцінки щільності розподілу.

Залишається з'ясувати, як вибрати величину h (ширину вікна) і які властивості повинне мати ядро. Будемо вимагати, щоб оцінка щільності розподілу була асимптотично незсуненою та слушною, тобто математичне сподівання оцінки (2.8) при $N \to \infty$ прямувало до щільності розподілу (асимптотична незсуненість), а сама оцінка (2.8) — до щільності розподілу при $N \to \infty$ за ймовірністю. Сформулюємо умови, що забезпечують ці властивості [22].

1. $\int\limits_{-\infty}^{\infty} K(z)\,dz = 1.$
2. $\int\limits_{-\infty}^{\infty} |K(z)|\,dz < \infty.$
3. $\sup\limits_{-\infty<z<\infty} |K(z)| < \infty.$
4. $\lim\limits_{h\to 0} |zK(z)| = 0.$

Існує багато ядер, що задовольняють ці властивості. Деякі з них наведені в [4, 38, 22]. Усі вони мають графік з максимумом у точці x, який спадає в її околі до нуля. Форма ядра та швидкість спадання обираються з огляду на бажані властивості їхньої гладкості. Площа фігури, обмеженою цим графіком та віссю Ox, дорівнює одиниці. Наприклад, ядро Парзена (2.11) має прямокутний графік.

Прямокутне ядро:

$$K(y) = \begin{cases} \frac{1}{2}, & \text{якщо } y \leq 1, \\ 0, & \text{якщо } y > 1. \end{cases} \tag{2.12}$$

Трикутне ядро:

$$K(y) = \begin{cases} 1 - |y|, & \text{якщо } y \leq 1, \\ 0, & \text{якщо } y > 1. \end{cases} \tag{2.13}$$

Квартичне ядро:

$$K(y) = \begin{cases} \frac{15}{16}\left(1 - y^2\right) y^2, & \text{якщо } y \leq 1, \\ 0, & \text{якщо } y > 1. \end{cases} \tag{2.14}$$

Гауссове ядро:

$$K(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right), & \text{якщо } y \leq 1, \\ 0, & \text{якщо } y > 1. \end{cases} \tag{2.15}$$

Ядро Єпанечнікова:

$$K(y) = \begin{cases} \frac{3}{4}\left(1 - y^2\right), & \text{якщо } y \leq 1, \\ 0, & \text{якщо } y > 1. \end{cases} \tag{2.16}$$

Універсального рецепта вибору ширини вікна Парзена h не існує, оскільки вона залежить від вибірки, зокрема від щільності розташування точок цієї вибірки на числовій осі. Отже,

для оптимального вибору вікна Парзена проводяться додаткові дослідження, зазвичай методом ковзного контролю [4].

2.4. Застосування наївного байєсівського підходу

Розглянемо задачу класифікації текстів, які виникають у теорії інформаційного пошуку, наприклад при фільтрації спама. Розглянемо два варіанти наївного методу Байєса: мультиноміальну модель і модель Бернуллі. Припустимо, що навчальна вибірка складається з текстів у вигляді неупорядкованих векторів термінів із заданого лексикона, які можуть повторюватися. Оскільки розподіл апостеріорної та умовної ймовірностей у цій задачі є дискретним, то всі обчислення стають елементарними.

1. transfer, greetings, transfer — спам.
2. transfer, transfer, diplomat — спам.
3. transfer, winner, telephone — спам.
4. transfer, name, information — спам.
5. credit, card, transfer — не спам.

Припустимо, що ми отримали лист s, який містить терміни transfer, transfer, transfer, credit, card. Застосуємо наївний байєсівський підхід.

Мультиноміальна байєсівська модель

У цій моделі будемо шукати найімовірніший клас $\omega_i, i = 1, 2$ (ω_1 — спам, ω_2 — не спам) для об'єкта x, який подається як

вектор термінів. Для цього необхідно знайти клас, на якому апостеріорна ймовірність $\hat{P}(\omega_i|x)$ досягає максимуму:

$$\arg\max_{i=1,2} P(\omega_i|\, x) = \arg\max_{i=1,2} P(\omega_i) \prod_{k=1}^{n} P(t_k|\omega_i)\,.$$

Тут $P(t_k|\omega_1)$ — умовна ймовірність появи терміна t_k у листі-спамі, $P(t_k|\omega_2)$ — умовна ймовірність появи терміна t_k у звичайному листі, $P(\omega_1)$ — апріорна ймовірність спама, $P(\omega_2)$ — апріорна ймовірність звичайного листа. Імовірність спама можна або обчислити за навчальною вибіркою, або використати загальні оцінки, що були отримані фахівцями (наприклад 0,5 або навіть 0,8). Будемо використовувати оцінку за навчальними вибірками.

Оскільки в нашому прикладі з п'яти навчальних вибірок чотири належать до спама, то маємо:

$$\hat{P}(\omega_1) = \frac{4}{5},\ \hat{P}(\omega_2) = \frac{1}{5}.$$

Якщо в тестовій вибірці зустрічається термін, якого немає в навчальних вибірках, то його умовна ймовірність дорівнює нулю. Для того, щоб не відкидати листи з рідко вживаними словами, можна використати *згладжування Лапласа*, додавши одиницю до частоти кожного терміна. У такому випадку умовна ймовірність терміна t у класі ω обчислюється за формулою

$$\hat{P}(t|\omega) = \frac{T_{\omega t} + 1}{\sum\limits_{t' \in L} T_{\omega t'} + M},$$

де M — кількість різних термінів у лексиконі.

Оцінимо умовну ймовірність $\hat{P}(t|\omega)$ для кожного класу. У навчальних вибірках, що класифіковані як спам, міститься 12 слів (нагадуємо, що вони можуть повторюватися), з них слово *transfer* зустрічається 6 разів, а слова *credit* і *card* зовсім не

зустрічаються. Лексікон містить 9 різних термінів. Отже, з урахуванням згладжування Лапласа, маємо:

$$\hat{P}\left(transfer|\omega_1\right) = \frac{6+1}{12+9} = \frac{7}{21} = \frac{1}{3},$$

$$\hat{P}\left(credit|\omega_1\right) = \frac{0+1}{12+9} = \frac{1}{21},$$

$$\hat{P}\left(card|\omega_1\right) = \frac{0+1}{12+9} = \frac{1}{21}.$$

У навчальній вибірці, що класифікована як звичайний текст, міститься три слова, з них слова *transfer*, *credit* і *card* зустрічаються по одному разу. Отже:

$$\hat{P}\left(transfer|\omega_2\right) = \frac{0+1}{3+9} = \frac{1}{12},$$

$$\hat{P}\left(credit|\omega_2\right) = \frac{0+1}{3+9} = \frac{1}{12},$$

$$\hat{P}\left(card|\omega_2\right) = \frac{0+1}{3+9} = \frac{1}{12}.$$

Тепер можемо обчислити апостеріорні ймовірності кожного класу.

$$\hat{P}\left(\omega_1|s\right) = \frac{4}{5}\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{21}\frac{1}{21} = 6,72 * 10^{-5}.$$

$$\hat{P}\left(\omega_2|s\right) = \frac{1}{5}\left(\frac{1}{12}\right)^5 = 8,04 * 10^{-7}.$$

Як бачимо, імовірність того, що тестовий лист належить до спама, значно вища, ніж імовірність протилежної події. Відповідно до вирішального правила наївного байєсівського підходу доходимо висновку, що отримали спам.

2.4.1. Модель Бернуллі

Друга модель наївного байєсівського класифікатора — модель Бернуллі — еквівалентна бінарній моделі, що генерує індикатор для кожного терміна словника: 1, якщо термін присутній у документі, і 0, якщо відсутній. У моделі Бернуллі ймовірність $\hat{P}(t|\omega)$ оцінюється як частка об'єктів із класу ω, що мають ознаку t. На противагу їй у мультиноміальній моделі ймовірність $\hat{P}(t|\omega)$ оцінюється як частка ознаки t в об'єктах із класу ω. На відміну від мультиноміальної моделі, при класифікації тестового документа на основі моделі Бернуллі головним є факт наявності ознаки, а кількість її входжень ігнорується. Це є слабкістю моделі Бернуллі, оскільки вона може враховувати рідкі терміни як значущі. З іншого боку, у мультиноміальній моделі ніяк не враховується інформація про терміни, відсутні в тестовому документі, але в моделі Бернуллі, яка для кожного терміна обчислює індикатор 0 або 1, ця інформація має вагу.

Застосуємо модель Бернуллі для класифікації попереднього листа на підставі навчальних вибірок, наведених вище.

Апріорні ймовірності класів не змінюються.

$$\hat{P}(\omega_1) = \frac{4}{5},\ \hat{P}(\omega_2) = \frac{1}{5}.$$

Обчислимо умовні ймовірності в моделі Бернуллі (не забуваймо про згладжування Лапласа). Слово *transfer* зустрічається в усіх чотирьох навчальних вибірках, а слова *greetings, diplomat, winner, telephone, name, information* — по одному разу. Оскільки клас ω_1 містить чотири вибірки, а величина M дорівнює 2 (термін або присутній, або відсутній), то маємо:

$$\hat{P}(transfer|\omega_1) = \frac{4+1}{4+2} = \frac{5}{6},$$
$$\hat{P}(credit|\omega_1) = \frac{0+1}{4+2} = \frac{1}{6},$$

$$\hat{P}\left(card|\omega_1\right) = \frac{0+1}{4+2} = \frac{1}{6},$$
$$\hat{P}\left(greetings|\omega_1\right) = \frac{1+1}{4+2} = \frac{2}{3},$$
$$\hat{P}\left(diplomat|\omega_1\right) = \frac{1+1}{4+2} = \frac{2}{3},$$
$$\hat{P}\left(winner|\omega_1\right) = \frac{1+1}{4+2} = \frac{2}{3},$$
$$\hat{P}\left(telephone|\omega_1\right) = \frac{1+1}{4+2} = \frac{2}{3},$$
$$\hat{P}\left(name|\omega_1\right) = \frac{1+1}{4+2} = \frac{2}{3},$$
$$\hat{P}\left(information|\omega_1\right) = \frac{1+1}{4+2} = \frac{2}{3}.$$

Клас ω_2 містить одну вибірку :

$$\hat{P}\left(transfer|\omega_2\right) = \frac{1+1}{1+2} = \frac{2}{3},$$
$$\hat{P}\left(credit|\omega_2\right) = \frac{1+1}{1+2} = \frac{2}{3},$$
$$\hat{P}\left(card|\omega_2\right) = \frac{1+1}{1+2} = \frac{2}{3},$$
$$\hat{P}\left(greetings|\omega_1\right) = \frac{0+1}{1+2} = \frac{1}{3},$$
$$\hat{P}\left(diplomat|\omega_1\right) = \frac{0+1}{1+2} = \frac{1}{3},$$
$$\hat{P}\left(winner|\omega_1\right) = \frac{0+1}{1+2} = \frac{1}{3},$$
$$\hat{P}\left(telephone|\omega_1\right) = \frac{0+1}{1+2} = \frac{1}{3},$$
$$\hat{P}\left(name|\omega_1\right) = \frac{0+1}{1+2} = \frac{1}{3},$$
$$\hat{P}\left(information|\omega_1\right) = \frac{0+1}{1+2} = \frac{1}{3}.$$

Обчислимо апостеріорні ймовірності кожного класу:

$$\begin{aligned}\hat{P}(\omega_1|s) = \hat{P}(\omega_1) * \\ &* \hat{P}(transfer|\omega_1) * \\ &* \hat{P}(credit|\omega_1) * \\ &* \hat{P}(card|\omega_1) * \\ &* \Big(1 - \hat{P}(greetings|\omega_1)\Big) * \\ &* \Big(1 - \hat{P}(diplomat|\omega_1)\Big) * \\ &* \Big(1 - \hat{P}(winner|\omega_1)\Big) * \\ &* \Big(1 - \hat{P}(telephone|\omega_1)\Big) * \\ &* \Big(1 - \hat{P}(name|\omega_1)\Big) * \\ &* \Big(1 - \hat{P}(information|\omega_1)\Big).\end{aligned}$$

Отже,

$$\hat{P}(\omega_1|s) = \frac{4}{5}\frac{5}{6}\frac{1}{6}\frac{1}{6}\left(\frac{2}{3}\right)^6 = 0,0016.$$

$$\begin{aligned}\hat{P}(\omega_2|s) = \hat{P}(\omega_2) * \hat{P}(transfer|\omega_2) * \\ *\hat{P}(credit|\omega_2) * \\ *\hat{P}(card|\omega_2) * \\ * \Big(1 - \hat{P}(greetings|\omega_2)\Big) * \\ * \Big(1 - \hat{P}(diplomat|\omega_2)\Big) * \\ * \Big(1 - \hat{P}(winner|\omega_2)\Big) * \\ * \Big(1 - \hat{P}(telephone|\omega_2)\Big) * \\ * \Big(1 - \hat{P}(name|\omega_2)\Big) *\end{aligned}$$

$$* \left(1 - \hat{P}\left(information|\omega_2\right)\right).$$

Отже,

$$\hat{P}\left(\omega_2|s\right) = \frac{1}{5}\left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^6 = 3,61 * 10^{-5}.$$

Як і при використанні мультиноміальної моделі, імовірність того, що тестовий лист належить до спама, значно перевищує ймовірність того, що ми отримали звичайний лист. Отже, це спам.

Якщо використання наївного байєсівського класифікатора передбачає дві можливі моделі — мультиноміальну та Бернуллі — то виникає питання, як правильно вибрати потрібну модель. З погляду сили припущень, які лежать в основі моделі, вони еквівалентні — припускається, що всі ознаки є незалежними та ймовірність їх появи на будь-якій позиції однакова. Обидві моделі дають дуже неточну оцінку справжньої умовної ймовірності термінів. Утім, незважаючи на це, вони є достатньо точними щодо рішення про класифікацію [17]. Отже, на передній план виходить їх ефективність при роботі з різними обсягами вибірок і ознак. З цього боку вони розрізняються: мультиноміальна модель краще працює з довгими вибірками й великою кількістю ознак, а модель Бернуллі — з короткими вибірками й невеликою кількістю ознак.

Розділ 3.

Дискримінант Фішера

Однією з основних проблем є залежність складності алгоритму розпізнавання від розмірності простору ознак. Що більше розмірність простору ознак, то більше складність алгоритму. З цієї причини бажано зменшити простір ознак, в ідеалі — до числової прямої. Саме ця ідея покладена в основу методу, основаного на використанні лінійної дискримінантної функції Фішера.

3.1. Лінійний дискримінант Фішера

Розглянемо дві множини навчальних вибірок з p-вимірного простору: $X_1, X_2, \ldots, X_{n_1}$ та $X_{n_1+1}, X_{n_1+2}, \ldots, X_{n_1+n_2}$. Вважатимемо, що $X_1, X_2, \ldots, X_{n_1}, X_{n_1+1}, X_{n_1+2}, \ldots, X_{n_1+n_2}$ — вектори чисел у просторі розмірністю p. Дискримінантний метод Фішера полягає в проектуванні цих векторів із простору розмірністю p на числову пряму за допомогою лінійної функції $l(X) = w^T X$ і відокремленні двох генеральних сукупностей якомога далі одна від одної за допомогою вектора w із простору розмірністю p.

Необхідно знайти вектор $\hat{w}$, що максимізує функціонал

$J(w)$, де

$$J(w) = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right)^2}{S_Y^2},\ \bar{Y}_1 = \frac{\sum\limits_{i=1}^{n_1} Y_i}{n_1},\ \bar{Y}_2 = \frac{\sum\limits_{i=n_1+1}^{n_1+n_2} Y_i}{n_2},$$

$$S_Y^2 = \frac{\sum\limits_{i=1}^{n_1} (Y_i - \bar{Y}_1)^2 + \sum\limits_{i=n_1+1}^{n_1+n_2} (Y_i - \bar{Y}_2)^2}{n_1 + n_2 - 2},$$

$$Y_i = w^T \vec{X}_i = \left(w, \vec{X}_i\right),\ \ i = 1, 2, \ldots, n_1 + n_2.$$

З інтуїтивного погляду функція критерію $J(w) = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right)^2}{S_Y^2}$ оцінює різницю між середніми проекцій $\bar{Y}_1 - \bar{Y}_2$ щодо стандартного відхилення S_Y. Якщо проекції $Y_1, Y_2, \ldots, Y_{n_1}$ і $Y_{n_1+1}, Y_{n_1+2}, \ldots, Y_{n_1+n_2}$ можна відокремити повністю, то величина $\left(\bar{Y}_1 - \bar{Y}_2\right)^2$ повинна бути великою щодо середнього відхилення S_Y.

Теорема 2. *Вектор $\hat{w}$, що максимізує величину $J(w) = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right)^2}{S_Y^2}$, має вигляд $S_W^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$, де*

$$S_W = \frac{(n_1 - 1) S_1 + (n_2 - 1) S_2}{n_1 + n_2 - 2},$$

$$S_1 = \frac{\sum\limits_{i=1}^{n_1} \left(X_i - \bar{X}_1\right)\left(X_i - \bar{X}_1\right)^T}{n_1 - 1},$$

$$S_2 = \frac{\sum\limits_{i=n_1+1}^{n_1+n_2} \left(X_i - \bar{X}_2\right)\left(X_i - \bar{X}_2\right)^T}{n_2 - 1},$$

$$\bar{X}_1 = \frac{\sum\limits_{i=1}^{n_1} X_i}{n_1},\ \bar{X}_2 = \frac{\sum\limits_{i=n_1+1}^{n_1+n_2} X_i}{n_2}.$$

Доведення. Запишемо середнє вибіркове значення проекції першої вибірки як

$$\bar{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(w, \vec{X}_i\right) = \\ = \left(w, \frac{1}{n_1}\sum_{i=1}^{n_1} X_i\right) = \left(w, \bar{X}_1\right).$$

Аналогічно $\bar{Y}_2 = \left(w, \bar{X}_2\right)$. Отже,

$$\sum_{i=1}^{n_1}\left(Y_i - \bar{Y}_1\right)^2 = \sum_{i=1}^{n_1}\left(w^T\vec{X}_i - w^T\bar{X}_1\right)^2 = \\ = \sum_{i=1}^{n_1}\left(w^T X_i - w^T\bar{X}_1\right)\left(w^T X_i - w^T\bar{X}_1\right)^T = \\ = \sum_{i=1}^{n_1} w^T\left(X_i - \bar{X}_1\right)\left(X_i - \bar{X}_1\right)^T w = \\ = w^T\left[\sum_{i=1}^{n_1}\left(X_i - \bar{X}_1\right)\left(X_i - \bar{X}_1\right)^T\right] w.$$

Крім того,

$$\sum_{i=n_1+1}^{n_1+n_2}\left(Y_i - \bar{Y}_2\right)^2 = w^T\left[\sum_{i=n_1+1}^{n_1+n_2}\left(X_i - \bar{X}_2\right)\left(X_i - \bar{X}_2\right)^T\right] w.$$

Таким чином,

$$S_Y^2 = \frac{\sum\limits_{i=1}^{n_1}\left(Y_i - \bar{Y}_1\right)^2 + \sum\limits_{i=n_1+1}^{n_1+n_2}\left(Y_i - \bar{Y}_2\right)^2}{n_1 + n_2 - 2} =$$

$$= \frac{w^T \left[\sum\limits_{i=1}^{n_1} \left(X_i - \bar{X}_1\right) \left(X_i - \bar{X}_1\right)^T \right] w}{n_1 + n_2 - 2} +$$

$$+ \frac{w^T \left[\sum\limits_{i=n_1+1}^{n_1+n_2} \left(X_i - \bar{X}_2\right) \left(X_i - \bar{X}_2\right)^T \right] w}{n_1 + n_2 - 2} =$$

$$= w^T \left[\frac{\sum\limits_{i=1}^{n_1} \left(X_i - \bar{X}_1\right) \left(X_i - \bar{X}_1\right)^T}{n_1 + n_2 - 2} \right] w +$$

$$+ w^T \left[\frac{\sum\limits_{i=n_1+1}^{n_1+n_2} \left(X_i - \bar{X}_2\right) \left(X_i - \bar{X}_2\right)^T}{n_1 + n_2 - 2} \right] w =$$

$$= w^T \left[\frac{(n_1 - 1) S_1 + (n_2 - 1) S_2}{n_1 + n_2 - 2} \right] w = w^T S_W w.$$

Отже,

$$J(w) = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right)^2}{S_Y^2} = \frac{\left(w^T \left(\bar{X}_1 - \bar{X}_2\right)\right)^2}{w^T S_W w}.$$

Вектор $\hat{w}$ можна знайти, розв'язавши рівняння

$$\frac{\partial J(w)}{\partial w} = 2 \frac{\left(w^T \left(\bar{X}_1 - \bar{X}_2\right)\right) \left(\bar{X}_1 - \bar{X}_2\right) \left(w^T S_W w\right)}{\left(w^T S_W w\right)^2} -$$

$$- \left(w^T \left(\bar{X}_1 - \bar{X}_2\right)\right)^2 \frac{2 S_W w}{\left(w^T S_W w\right)^2} = 0.$$

Подальші перетворення приводять до рівняння

$$\left(\bar{X}_1 - \bar{X}_2\right) = \left[\frac{w^T \left(\bar{X}_1 - \bar{X}_2\right)}{w^T S_W w} \right] S_W w.$$

Множачи обидві частини цього рівняння на матрицю S_W^{-1}, отримуємо:

$$S_W^{-1}\left(\bar{X}_1 - \bar{X}_2\right) = \left[\frac{w^T\left(\bar{X}_1 - \bar{X}_2\right)}{w^T S_W w}\right] w.$$

Оскільки $\frac{w^T\left(\bar{X}_1 - \bar{X}_2\right)}{w^T S_W w}$ — дійсне число, то маємо $\hat{w} = cS_W^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$, де c — деяка константа.

Матриця S_W називається *матрицею розкиду всередині класу*. Матриця $S_B = \left(\bar{X}_1 - \bar{X}_2\right)\left(\bar{X}_1 - \bar{X}_2\right)^T$ називається *матрицею розкиду між класами*. □

За допомогою матриць S_W і S_B функцію критерію в методі Фішера можна сформулювати за допомогою узагальненого відношення Релея:

$$J(w) = \frac{w^T S_B w}{w^T S_W w}.$$

Припустимо, що маємо спостереження X_0 і для визначеності припустимо, що $\bar{Y}_1 < \bar{Y}_2$. Спираючись на дискримінантну функцію $l(X) = \hat{w}^T X$, яку знайшли вище, можемо віднести це спостереження до певного класу. Вирішальне правило формулюється так: спостереження X_0 належить до другої генеральної сукупності, якщо

$$\hat{Y}_0 = \hat{w}^T X_0 = \left(\bar{X}_1 - \bar{X}_2\right)^T S_W^{-1} X_0 \geq \frac{1}{2}\hat{w}^T(\bar{X}_1 + \bar{X}_2) =$$
$$= \frac{1}{2}\left(\bar{Y}_1 + \bar{Y}_2\right) = \frac{1}{2}\left(\bar{X}_1 - \bar{X}_2\right)^T S_W^{-1}\left(\bar{X}_1 + \bar{X}_2\right).$$

У протилежному випадку

$$\hat{Y}_0 = (\bar{X}_1 - \bar{X}_2)^T S_W^{-1} X_0 < \frac{1}{2}\left(\bar{X}_1 - \bar{X}_2\right)^T S_W^{-1}\left(\bar{X}_1 + \bar{X}_2\right)$$

і X_0 належить до першої генеральної сукупності.

Зауваження 1. Значне відділення не означає гарної класифікації. З іншого боку, якщо відділення не є значним, то здійснювати класифікацію немає сенсу.

3.2. Нелінійний дискримінант Фішера

Цікаво, що лінійний дискримімант Фішера був винайдений у 1936 р. [28], а його узагальнення на нелінійні випадки — лише у 1999 р. [32].

Якщо множини точок з навчальної вибірки не допускають лінійного розділення, то можна скористатися відображенням вихідного простору ознак X у новий простір ознак F за допомогою деякої функції φ так, що $w = \varphi(x)$. У такому випадку задача класифікації зводиться до максимізації функціонала

$$J(\vec{w}) = \frac{\vec{w}^{\mathrm{T}} \vec{S}_B^{\varphi} \vec{w}}{\vec{w}^{\mathrm{T}} \vec{S}_W^{\varphi} \vec{w}},$$

де

$$\vec{S}_B^{\varphi} = (\vec{m}_2^{\varphi} - \vec{m}_1^{\varphi})(\vec{m}_2^{\varphi} - \vec{m}_1^{\varphi})^{\mathrm{T}},$$

$$\vec{S}_W^{\varphi} = \sum_{i=1}^{2} \sum_{n=1}^{l_i} (\varphi(\vec{x}_n^i) - \vec{m}_i^{\varphi})(\varphi(\vec{x}_n^i) - \vec{m}_i^{\varphi})^{\mathrm{T}},$$

$$\vec{m}_i^{\varphi} = \frac{1}{l_i} \sum_{j=1}^{l_i} \varphi(\vec{x}_j^i).$$

Безпосереднє застосування лінійного дискримінанта Фішера у просторі F може виявитися недоречним з огляду на складність обчислень або велику вимірність простору F. Тому введемо до розгляду ядро

$$K(\vec{x}, \vec{y}) = \varphi(\vec{x}) \cdot \varphi(\vec{y})$$

і скористаємось розвиненням

$$\vec{w} = \sum_{i=1}^{l} \alpha_i \varphi(\vec{x}_i).$$

Зважаючи на рівність

$$\vec{w}^{\mathrm{T}} \vec{m}_i^{\varphi} = \frac{1}{l_i} \sum_{j=1}^{l} \sum_{k=1}^{l_i} \alpha_j K(\vec{x}_j, \vec{x}_k^i) = \vec{\alpha}^{\mathrm{T}} \vec{M}_i,$$

де $(\vec{M}_i)_j = \frac{1}{l_i} \sum_{k=1}^{l_i} k(\vec{x}_j, \vec{x}_k^i)$, перепишемо чисельник функціонала $J(\vec{w})$ таким чином:

$$\vec{w}^{\mathrm{T}} \vec{S}_B^{\varphi} \vec{w} = \vec{w}^{\mathrm{T}} (\vec{m}_2^{\varphi} - \vec{m}_1^{\varphi})(\vec{m}_2^{\varphi} - \vec{m}_1^{\varphi})^{\mathrm{T}} \vec{w} = \vec{\alpha}^{\mathrm{T}} \vec{M} \vec{\alpha},$$

де $\vec{M} = (\vec{M}_2 - \vec{M}_1)(\vec{M}_2 - \vec{M}_1)^{\mathrm{T}}$.

Знаменник можна переписати аналогічно:

$$\vec{w}^{\mathrm{T}} \vec{S}_W^{\varphi} \vec{w} = \vec{\alpha}^{\mathrm{T}} \vec{N} \vec{\alpha},$$

де $\vec{N} = \sum_{j=1}^{2} \vec{K}_j (I - L) \vec{K}_j^{\mathrm{T}}$, а n-й і m-й компоненти $\vec{K}_j$ визначаються як $K(\vec{x}_n, \vec{x}_m^j)$, I — одинична матриця, L — матриця, заповнена числами $1/l_j$. Цю тотожність можна вивести, застосувавши розвинення $\vec{w}$ і означення S_W^{φ} та $\vec{m}_i^{\varphi}$ у виразі

$$\vec{w}^{\mathrm{T}} \vec{S}_W^{\varphi} \vec{w}.$$

Зауваження 2. Виконайте цю вправу самостійно!

Перепишемо рівняння для функціонала J як

$$J(\vec{\alpha}) = \frac{\vec{\alpha}^{\mathrm{T}} \vec{M} \vec{\alpha}}{\vec{\alpha}^{\mathrm{T}} \vec{N} \vec{\alpha}}.$$

Застосовуючи умови Ейлера, отримуємо рівняння

$$(\vec{\alpha}^{\mathrm{T}} \vec{M} \vec{\alpha}) \vec{N} \vec{\alpha} = (\vec{\alpha}^{\mathrm{T}} \vec{N} \vec{\alpha}) \vec{M} \vec{\alpha}.$$

Розв'язуючи це рівняння, отримуємо:

$$\vec{\alpha} = \vec{N}^{-1}(\vec{M}_2 - \vec{M}_1).$$

Для того, щоб уникнути сингулярності, додамо незначне збурення [32]:

$$\vec{N}_\epsilon = \vec{N} + \epsilon\vec{I}.$$

Отже, образом точки $\vec{x}$ є

$$\vec{y} = (\vec{w}, \varphi(\vec{x})) = \sum_{i=1}^{l} \alpha_i k(\vec{x}_i, \vec{x}).$$

Розділ 4.

Метод опорних векторів із жорстким зазором

У двокласовому дискримінантному методі Фішера на площині ми шукали таку пряму, проекція на яку забезпечувала б максимальне розділення точок. Природно поставити нову задачу: знайти таку гіперплощину, яка б розділяла два класи так, щоб відстань від неї до найближчої точки з кожного класу була максимальною. Ця ідея інтуїтивно зрозуміла — якщо є вибір ліній, які розділяють дві множини точок, то найкращим класифікатором була б лінія, максимально віддалена від цих множин. У такому випадку емпірична помилка класифікації стає мінімальною, а здатність алгоритму до узагальнення — максимальною.

Припустимо, що N навчальних вибірок із класів C_1 і C_2, які мають мітку ω, що дорівнює 1, якщо $x_i \in C_1$ і $\omega = -1$, якщо $x_i \in C_1$ відповідно, містять n вибіркових значень, тобто $\vec{x}_i = \{x_1, x_2, ..., x_n\}$, $i = 1, ..., N$. Для простоти припустимо, що множини точок є лінійно роздільними, і позначимо мітку точки x_i як ω_i.

Розглянемо роздільну лінію

$$g(\vec{x}) = (\vec{w}, \vec{x}) + b, \tag{4.1}$$

де $\vec{w}$ — n-вимірний вектор, ортогональний роздільній гіперплощині, b — параметр зсуву цієї гіперплощини та

$$g(\vec{x}_i) = \omega_i, i = 1, ..., N. \tag{4.2}$$

Таким чином, функція $g(x)$ набуває значення 1 на навчальних вибірках із класу C_1 і значення -1 на навчальних вибірках із класу C_2. До того ж, оскільки ми припустили, що множини навчальних вибірок є лінійно роздільними, то жодна з точок не лежить на прямій, тобто

$$g(\vec{x}_i) \neq 0, i = 1, ..., N.$$

Уведемо до розгляду важливі поняття, які стануть у нагоді в майбутньому.

Означення 21. Точки, що є найближчими до відокремлювальної гіперплощини, називаються ***опорними***.

Означення 22. Величина $\omega((\vec{w}, \vec{x}_i) + b)$ називається ***функціональним зазором***.

Означення 23. Відстань від відокремлювальної гіперплощини до найближчої навчальної точки називається ***зазором***.

Означення 24. Максимальна ширина смуги, яку можна провести паралельно відокремлювальній гіперплощині через опорні точки, називається ***геометричним зазором***.

Позначимо євклідову відстань між деякою точкою $\vec{x}_i$ та відокремлювальною гіперплощиною символом r_i. Оскільки найменша відстань між точкою й гіперплощиною визначається перпендикуляром до площини, паралельним вектору $\vec{w}$, то одиничний вектор у цьому напрямку має вигляд $\frac{\vec{w}}{|\vec{w}|}$. Проекція точки $\vec{x}_i$ на відокремлювальну гіперплощину обчислюється за формулою

$$\vec{x}_i' = \vec{x}_i - \omega_i r \frac{\vec{w}}{|\vec{w}|}.$$

Проекція $\vec{x}_i'$ лежить на гіперплощині й задовольняє рівняння $(\vec{w}, \vec{x}_i') + b = 0$. Отже,

$$\left(\vec{w}, \vec{x}_i - \omega_i r \frac{\vec{w}}{|\vec{w}|}\right) + b = 0. \tag{4.3}$$

Звідси випливає, що

$$r_i = \omega_i \frac{(\vec{w}, \vec{x}_i) + b}{|\vec{w}|}. \tag{4.4}$$

Множник ω_i визначає положення точки x_i відносно відокремлювальної гіперплощини: якщо $\omega_i = 1$, то $x_i \in C_1$, а якщо $\omega_i = -1$, то $x_i \in C_2$. Множення правої частини рівняння (4.4) на будь-яке додатне число не впливає на точність класифікації, отже, можна для зручності оптимізації провести нормалізацію геометричного зазору так, щоб він не був менше одиниці на жодній навчальній точці й досягав одиниці на опорних точках. Цього можна досягти, поклавши $|\vec{w}| = 1$. Тоді геометричний й функціональний зазори збігаються і можна сформулювати таке обмеження:

$$\omega_i \left((\vec{w}, \vec{x}_i) + b\right) \geq 1. \tag{4.5}$$

Виходячи з рівності (4.5), доходимо висновку, що геометричний зазор $\rho = \frac{2}{|\vec{w}|}$.

Таким чином, отримуємо задачу оптимізації

$$\arg \max_{\omega_i((\vec{w},\vec{x}_i)+b)\geq 1} \frac{2}{|\vec{w}|}. \tag{4.6}$$

Це еквівалентно такій задачі мінімізації:

$$\arg \min_{\omega_i((\vec{w},\vec{x}_i)+b)\geq 1} \frac{(\vec{w}, \vec{w})}{2}. \tag{4.7}$$

Задача мінімізації квадратичної функції за лінійних обмежень називається *задачею квадратичної оптимізації*.

Стандартним способом розв'язання задачі квадратичної мінімізації є її зведення до двоїстої задачі, у якій кожному обмеженню $\omega_i\left((\vec{w},\vec{x}_i)+b\right) \geq 1$ прямої задачі ставиться у відповідність шуканий множник Лагранжа λ_i. За теоремою Куна – Таккера [8], задача (4.7) є еквівалентною опуклій задачі пошуку сідлової точки функції Лагранжа без обмежень.

4.1. Двоїста задача без обмежень

Треба знайти множники Лагранжа λ_i, $i = 1, ..., N$, що задовольняють умову

$$\begin{aligned} & J\left(\vec{w}, b, \vec{\lambda}\right) = \\ & = \frac{(\vec{w},\vec{w})}{2} - \sum_{i=1}^{N} \lambda_i \left(\omega_i\left((\vec{w},\vec{x}_i)+b\right)-1\right) \rightarrow \\ & \rightarrow \min_{\vec{w},b} \max_{\lambda_i}. \end{aligned} \tag{4.8}$$

Тут $(\vec{x},\vec{y})$ — це скалярний добуток векторів $\vec{x}$ та $\vec{y}$. Відповідно до теореми Куна – Таккера, розв'язком цієї задачі є сідлова точка. Ця точка задовольняє такі умови:

$$\frac{\partial J\left(\vec{w}, b, \vec{\lambda}\right)}{\partial \omega} = 0, \tag{4.9}$$

$$\frac{\partial J\left(\vec{w}, b, \vec{\lambda}\right)}{\partial b} = 0, \tag{4.10}$$

$$\lambda_i\left(\omega_i\left((\vec{w},\vec{x}_i)+b\right)-1\right) = 0, \quad i = 1, ..., N, \tag{4.11}$$

$$\lambda_i \geq 0 \quad i = 1, ..., N. \tag{4.12}$$

Обчислюючи частинні похідні (4.9), (4.10), маємо

$$\vec{w} = \sum_{i=1}^{N} \lambda_i \omega_i x_i, \tag{4.13}$$

$$\sum_{i=1}^{N} \lambda_i \omega_i = 0. \tag{4.14}$$

Підставляючи ці вирази в (4.8), отримуємо еквівалентну двоїсту задачу щодо множників Лагранжа λ_i.

4.2. Двоїста задача щодо множників Лагранжа

Треба знайти множники Лагранжа λ_i, $i = 1, ..., N$, за яких величина

$$J(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \omega_i \omega_j (\vec{x}_i, \vec{x}_j) \tag{4.15}$$

досягає максимуму за умов:

1. $\sum\limits_{i=1}^{N} \lambda_i \omega_i = 0$;
2. $\lambda_i \geq 0 \;\; i = 1, ..., N$.

Знайшовши множники Лагранжа, можемо визначити розв'язок задачі (4.7):

$$\begin{aligned} \vec{w} &= \sum_{i=1}^{N} \lambda_i \omega_i \vec{x}_i, \\ b &= \omega_k - (\vec{w}, \vec{x}_k) \quad \text{для таких } x_k, \text{ що } \lambda_k \neq 0. \end{aligned} \tag{4.16}$$

Зважаючи на умову (4.11), легко бачити, що всі множники Лагранжа, які не відповідають опорним точкам, тобто точкам, що задовольняють умову $\omega_i ((\vec{w}, \vec{x}_i) + b) - 1 = 0$, дорівнюють

нулю. Тому розмірність задачі оптимізації визначається лише кількістю опорних точок, яка зазвичай є невеликою. Урешті, вирішальна функція набуває вигляду

$$g(\vec{x}) = sign\left(\sum_{k=1}^{M} \lambda_k \omega_k (\vec{x}_k, \vec{x}) + b\right), \tag{4.17}$$

де M — кількість опорних точок.

Досі ми припускали, що точки можна розділити гіперплощиною. Утім це ідеальний випадок, який рідко зустрічається на практиці. Для того, щоб класифікувати множини, які не можна розділити гіперплощиною, є два підходи: 1) штрафувати помилки (метод опорних векторів з м'яким зазором) і 2) використати ядро, щоб відобразити точки у випрямний простір (нелінійний метод опорних векторів).

Розділ 5.

Метод опорних векторів з м'яким зазором

Якщо застосувати до лінійно нероздільних множин точок метод опорних векторів із жорстким зазором, описаний у попередньому розділі, то задача класифікації не матиме розв'язку. Для того, щоб урахувати лінійну нероздільність навчальних вибірок, уведемо в обмеження (4.5) невід'ємні фіктивні змінні, які відіграють роль штрафу.

$$\omega_i \left((\vec{w}, \vec{x}_i) + b \right) \geq 1 - \xi_i, \quad i = 1, ..., N. \tag{5.1}$$

Завдяки фіктивним змінним розв'язок оптимізаційної задачі завжди існує. У цьому випадку смуга, що розділяє навчальні вибірки, не буде порожньою, як у методі опорних векторів із жорстким зазором. Вона буде містити навчальні вибірки, які можуть бути класифіковані неправильно. Якщо $0 < \xi_i < 1$, то навчальні вибірки класифікуються правильно, хоча максимум функціонального зазору не досягається. Якщо $\xi_i \geq 1$, то навчальна вибірка класифікується неправильно. Для того, щоб

мінімізувати кількість помилок, треба розв'язати задачу

$$J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}, \vec{\mu}\right) = \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{N}\xi_i^p - \\ -\sum_{k=1}^{N}\lambda_i\left(\omega_i\left((\vec{w}, \vec{x}_i) + b\right) - 1 + \xi_i\right) - \\ -\sum_{k=1}^{N}\mu_i\xi_i \to \min_{\vec{w},b}\max_{\lambda_i\mu_i}, \tag{5.2}$$

де $\vec{\lambda}$ та $\vec{\mu}$ — вектори невід'ємних множників Лагранжа.

Показник p може набувати два значення. Якщо $p = 1$, то описуваний метод називається L_1-методом опорних векторів, а якщо $p = 2$, то L_2-методом [24]. Вибір цього параметра залежить від характеру навчальних вибірок, тому поки приділимо основну увагу загальній схемі методів, не віддаючи переваги жодному з них.

5.1. L_1-метод опорних векторів з м'яким зазором

Відповідно до теореми Куна – Таккера, розв'язком задачі (5.2) є сідлова точка, яка задовольняє такі умови:

$$\frac{\partial J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}, \vec{\mu}\right)}{\partial\omega} = 0, \tag{5.3}$$

$$\frac{\partial J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}, \vec{\mu}\right)}{\partial b} = 0, \tag{5.4}$$

$$\frac{\partial J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}, \vec{\mu}\right)}{\partial\xi} = 0, \tag{5.5}$$

$$\lambda_i\left(\omega_i\left((\vec{w}, \vec{x}_i) + b\right) - 1 + \xi_i\right) = 0, \quad i = 1, ..., N, \tag{5.6}$$

$$\mu_i \xi_i = 0, \quad i = 1, ..., N, \tag{5.7}$$

$$\lambda_i \geq 0, \mu_i \geq 0, \xi_i \geq 0, \quad i = 1, ..., N. \tag{5.8}$$

Обчислюючи частинні похідні (5.3) – (5.5), маємо:

$$\vec{w} = \sum_{i=1}^{N} \lambda_i \omega_i x_i, \tag{5.9}$$

$$\sum_{i=1}^{N} \lambda_i \omega_i = 0, \tag{5.10}$$

$$\lambda_i + \mu_i = C, \quad i = 1, ..., N. \tag{5.11}$$

Підставляючи ці вирази в (5.2), отримуємо еквівалентну двоїсту задачу відносно множників Лагранжа λ_i.

Двоїста задача відносно множників Лагранжа. Знайти множники Лагранжа λ_i, $i = 1, ..., N$, за яких величина

$$J(\lambda) = \sum_{1=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \omega_i \omega_j (\vec{x}_i, \vec{x}_j) \tag{5.12}$$

досягає максимуму за умов:

1. $\sum\limits_{i=1}^{N} \lambda_i \omega_i = 0$;
2. $C \geq \lambda_i \geq 0 \;\; i = 1, ..., N$.

Як бачимо, L_1-метод опорних векторів з м'яким зазором відрізняєься від методу з жорстким зазором лише обмеженням $C \geq \lambda_i \geq 0 \;\; i = 1, ..., N$.

Розглянемо можливі варіанти значень множників Лагранжа λ_i.

1. $\lambda_i = 0$. У цьому випадку точка x_i класифікується правильно.

2. $0 < \lambda_i < C$. Тоді $\omega_i((\vec{w}, \vec{x}_i)+b)-1+\xi_i = 0$, $\xi_i = 0$. Отже, $\omega_i\left((\vec{w}, \vec{x}_i) + b\right) = 1$, тобто точка $\vec{x}_i$ є опорним вектором. Назвемо його *вільним опорним вектором.*

3. $\lambda_i = C$. Тоді $\omega_i\left((\vec{w}, \vec{x}_i) + b\right) - 1 + \xi_i = 0$, $\xi_i \geq 0$. Отже, знову маємо $\omega_i\left((\vec{w}, \vec{x}_i) + b\right) = 1$, тобто точка $\vec{x}_i$ є опорним вектором. На відміну від попереднього випадку, назвемо його *зв'язаним опорним вектором.*

Вирішальна функція в методі опорних векторів із м'яким зазором має такий самий вигляд, як і в методі з жорстким зазором:

$$\begin{aligned} g\left(\vec{x}\right) &= \sum_{k=1}^{M} \lambda_k \omega_k \left(\vec{x}_k, \vec{x}\right) + b, \\ b &= \omega_k - \left(\vec{w}, \vec{x}_k\right), \end{aligned} \tag{5.13}$$

де $\vec{x}_k$ — один з вільних опорних векторів. Оскільки $\lambda_i \neq 0$ лише для опорних векторів, то підсумовування у вирішальній функції відбувається лише по опорних векторах. Вирішальне правило формулюється так:

$$\vec{x}_i \in C_1, \text{ якщо } g\left(\vec{x}_i\right) > 0, \tag{5.14}$$

$$\vec{x}_i \in C_2, \text{ якщо } g\left(\vec{x}_i\right) < 0. \tag{5.15}$$

Якщо $g\left(\vec{x}_i\right) = 0$, то вектор $\vec{x}_i$ не класифікується.

5.2. L_2-метод опорних векторів з м'яким зазором

Як було зазначено вище, вибираючи в (5.2) параметр $p = 2$, можна побудувати метод L_2-метод опорних векторів з м'яким зазором. У цьому методі задача набуває такого вигляду:

$$J\left(\vec{w}, b, \vec{\lambda}\right) =$$

$$= \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{N}\xi_i^2 - \frac{C}{2}\sum_{k=1}^{N}\lambda_i^2 \to \min_{\vec{w},b}\max_{\lambda_i} \tag{5.16}$$

за умови

$$\omega_i\left((\vec{w}, h\left(\vec{x}_i\right)) + b\right) \geq 1 - \xi_i, i =, ..., N, \tag{5.17}$$

де $\vec{w} = (w_1, ..., w_l)$ — l-вимірний вектор, а $h\left(\vec{x}\right)$ — функція, що відображає вектор $\vec{x} = (x_1, ..., x_m)$ у l-вимірний простір, ξ_i — фіктивні змінні, C — параметр зазору. Отже, ідея L_2-методу опорних векторів з м'яким зазором полягає у зменшенні вимірності задачі.

Задача мінімізації функціонала в L_2-методі опорних векторів з м'яким зазором формулюється так:

$$J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}\right) = \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{N}\xi_i^2 -$$

$$-\sum_{i=1}^{N}\lambda_i\left(\omega_i\left((\vec{w}, h\left(\vec{x}_i\right)) + b\right) - 1 + \xi_i\right). \tag{5.18}$$

На відміну від L_1-методу опорних векторів з м'яким зазором у L_2-методі множники Лагранжа для фіктивних змінних не потрібні, оскільки для оптимального розв'язку виконується умова $C\xi_i = \lambda_i \geq 0$.

Відповідно до теореми Куна – Таккера, розв'язком задачі (5.18) є сідлова точка, яка задовольняє такі умови:

$$\frac{\partial J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}\right)}{\partial\omega} = 0, \tag{5.19}$$

$$\frac{\partial J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}\right)}{\partial b} = 0, \tag{5.20}$$

$$\frac{\partial J\left(\vec{w}, b, \vec{\xi}, \vec{\lambda}\right)}{\partial\xi} = 0, \tag{5.21}$$

$$\lambda_i \left(\omega_i \left(\left(\vec{w}, h\left(\vec{x}_i\right)\right) + b\right) - 1 + \xi_i\right) = 0, \quad i = 1, ..., N. \tag{5.22}$$

Обчислюючи частинні похідні (5.19)–(5.21), маємо:

$$\vec{w} = \sum_{i=1}^{N} \lambda_i \omega_i h\left(\vec{x}_i\right), \tag{5.23}$$

$$C\xi_i - \lambda_i = 0, \tag{5.24}$$

$$\sum_{i=1}^{N} \omega_i \lambda_i = 0. \tag{5.25}$$

Підставляючи ці вирази в (5.18), отримуємо еквівалентну двоїсту задачу відносно множників Лагранжа λ_i.

З обмежень (5.23–5.25) доходимо висновку, що сідлова точка повинна задовольняти умову $\lambda_j = 0$ або

$$\omega_j \left(\sum_{i=1}^{N} \omega_i \lambda_i \left(K\left(\vec{x}_i, \vec{x}_j\right) + \frac{\delta_{ij}}{C}\right) + b\right) - 1 = 0, \tag{5.26}$$

де $K\left(\vec{x}, \vec{y}\right) = \left(h\left(\vec{x}\right), h\left(\vec{y}\right)\right)$ — ядро, δ_{ij} — символ Кронекера. Як бачимо, ядро є матрицею, розмір якої визначається розміром вектора $g\left(\vec{x}\right)$ у *просторі ознак*. Приклади різних ядер, у тому числі таких, що не є матрицею, будуть наведені в наступному розділі.

Відповідно вирішальна функція має вигляд

$$\begin{aligned} g\left(\vec{x}\right) &= \sum_{i=1}^{N} \lambda_i \omega_i K\left(\vec{x}_i, \vec{x}\right) + b, \\ b &= \omega_k - \sum_{i=1}^{N} \lambda_i \omega_i \left(K\left(\vec{x}_k, \vec{x}_i\right) + \frac{\delta_{ij}}{C}\right). \end{aligned} \tag{5.27}$$

Підставляючи вирази (5.23)–(5.25) у (5.18), отримуємо дво-

їсту задачу відносно множників Лагранжа.

$$J\left(\vec{\lambda}\right) = \sum_{i=1}^{N} \lambda_i - \\ - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j \lambda_i \lambda_j \left(K\left(\vec{x}_i, \vec{x}_j\right) + \frac{\delta_{ij}}{C} \right) \to \max_{\lambda_i} \tag{5.28}$$

за умови, що

$$\sum_{i=1}^{N} \lambda_i \omega_i = 0,; \;\; \lambda_i \geq 0, \;\; i = 1, ..., N. \tag{5.29}$$

Як бачимо, відмінність L_2-методу опорних векторів із м'яким зазором від методу із жорстким зазором полягає в доданку $\frac{\delta_{ij}}{C}$ і зменшенні виміру вихідної задачі. Завдяки тому, що доданок $\frac{\delta_{ij}}{C}$ додається до кожного діагонального елемента матриці $K\left(\vec{x}_i, \vec{x}_j\right)$, вона є додатно визначеною. Це значно підвищує стійкість обчислень порівняно з L_1-методом опорних векторів із м'яким зазором.

5.3. Нелінійний метод опорних векторів

У попередньому розділі ми зіткнулися з поняттям ядра $K\left(\vec{x}, \vec{y}\right)$, що породжувалося скалярним добутком $\left(h^T\left(\vec{x}\right), h\left(\vec{y}\right)\right)$, за допомогою якого можна було перейти з вихідного n-вимірного простору у l-вимірний *випрямний простір ознак*. Якщо ядро задовольняє умови Мерсера:

1. $K\left(\vec{x}, \vec{y}\right) = K\left(\vec{y}, \vec{x}\right)$(симетричність);
2. $\int\limits_X K\left(\vec{x}, \vec{y}\right) h\left(\vec{x}\right) h\left(\vec{y}\right) d\vec{x} d\vec{y} \geq 0$ для функцій h, що діють із простору ознак у простір $\mathbb{R}$ (додатна визначеність),

то навчаючі вибірки без дублікатів допускають розділення поліномами у випрямному просторі [4, 14].

За допомогою ядра двоїсту задачу у випрямному просторі можна сформулювати так:

$$J\left(\vec{\lambda}\right) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j \lambda_i \lambda_j K\left(\vec{x}_i, \vec{x}_j\right) \tag{5.30}$$

за умови, що

$$\sum_{i=1}^{N} \lambda_i \omega_i = 0,; \quad 0 \leq \lambda_i \leq C, \quad i = 1, ..., N. \tag{5.31}$$

Завдяки додатній визначеності ядра задача квадратичної оптимізації (5.30), (5.31) є опуклою, а отже, має єдиний розв'язок.

Умови Куна – Таккера у випрямному просторі набувають такого вигляду:

$$\lambda_i \left(\omega_i \left(\sum_{j=1}^{N} \omega_j \lambda_j K\left(\vec{x}_i, \vec{x}_j\right) + b - 1 \right) + \xi_i \right) = 0, \quad i = 1, ..., N, \tag{5.32}$$

$$\left(C - \lambda_i\right) \xi_i = 0 \quad i = 1, ..., N, \tag{5.33}$$

$$\lambda_i \geq 0, \xi_i \geq 0, \quad i = 1, ..., N. \tag{5.34}$$

Вирішальна функція задається формулою

$$g\left(\vec{x}\right) = \sum_{k=1}^{M} \lambda_k \omega_k K\left(\vec{x}_i, \vec{x}\right) + b,$$

$$b = \omega_k - \sum_{i=1}^{M} \lambda_k \omega_k K\left(\vec{x}_i, \vec{x}_k\right), \tag{5.35}$$

де x_k — деякий вільний опорний вектор, M — кількість опорних векторів.

Вирішальне правило формулюється так:

$$\vec{x}_i \in C_1, \text{ якщо } g(\vec{x}_i) > 0, \tag{5.36}$$

$$\vec{x}_i \in C_2, \text{ якщо } g(\vec{x}_i) < 0. \tag{5.37}$$

Якщо $g(\vec{x}_i) = 0$, то вектор $\vec{x}_i$ не класифікується.

Наведемо список найбільш широко вживаних ядер.

1. Лінійне ядро:
   $K(\vec{x}, \vec{y}) = (\vec{x}, \vec{x})$.

2. Поліноміальне ядро степеня d:
   $K(\vec{x}, \vec{y}) = ((\vec{x}, \vec{x}) + 1)^d$, $d \in \mathbb{N}$.

3. Радіальне ядро:
   $\exp\left(-\gamma||\vec{x} - \vec{y}||^2\right)$, $\gamma > 0$.

Приклад 3. Нехай $x = (x_1, x_2)$ і d=2. У такому випадку поліноміальне ядро записується як

$$\begin{aligned} K(x, y) &= 1 + 2x_1y_1 + 2x_2y_2 + 2x_1y_1x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 = \\ &= (h, (x), h(y)), \end{aligned}$$

де

$$h(x) = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2\right).$$

Отже, воно задовольняє умови Мерсера.

Розділ 6.

Нейронні мережі

Ідея застосувати для класифікації штучні нейронні мережі, що імітують природні процеси, які відбуваються в мозку людини, досить слушна. Зважаючи на складність мозку, це можливо здійснити, лише зробивши серйозні спрощення. Нейронним мережам присвячено дуже багато робіт, у тому числі [21, 23]. Ми зосередимося на оптимізаційній складовій цієї теорії.

6.1. Персептрон Розенблатта

Модель нейронної мережі схематично описує роботу мозку як сукупності нейронів, які можуть перебувати у двох станах: збудженому й незбудженому. Кожний нейрон отримує сигнали від інших нейронів, обчислює їх лінійну комбінацію й відповідно змінює свій потенціал. Якщо обчислений потенціал перевищує порогове значення, то нейрон збуджується. У такій моделі навчання можна звести до обчислення коефіцієнтів лінійних комбінацій потенціалів, що змінюються в часі.

На вхід нейрона надходить вектор ознак $\vec{x} = (x_1, x_2, ..., x_n)$. На підставі цього вектора і вектора ваг $\vec{w} = (w_1, w_2, ..., w_n)$ обчислюється лінійна комбінація

$$g(\vec{x}) = \sum_{i=1}^{n} w_i x_i + w_0.$$

Нейрон переходить у збуджений стан, якщо вихідний сигнал перевищує порогове значення. Ступінь збудження монотонно залежить від стану, обмежений знизу і зверху і стрімко змінюється в інтервалі значень від $\min \sum_{i=1}^{N} w_i x_i$ до $\max \sum_{i=1}^{N} w_i x_i$. Прикладами активаційних функцій є сходинка, сходинка з лінійним порогом, гіперболічний тангенс та сігмоїдна функція.

Результат класифікації обчислюється за таким вирішальним правилом:

$$\omega = \begin{cases} 1, \text{ якщо } g(\vec{x}) > 0, \\ 0, \text{ якщо } g(\vec{x}) < 0. \end{cases} \tag{6.1}$$

Отже, нейрон працює як лінійна дискримінантна функція.

6.1.1. Алгоритм Розенблатта

Одна з перших нейронних мереж була запропонована Ф. Розенблаттом у 1957 р. і отримала назву *персептрон*. Алгоритм Розенблатта призначений для поступового навчання персептрона безпомилковій бінарній класифікації об'єктів шляхом ітераційного уточнення ваг лінійної комбінації. Особливістю алгоритму Розенблатта є його циклічність — об'єкти для класифікації подаються від першого до останнього $(x_1, \omega_1), ..., (x_N, \omega_N)$, потім знову з першого до останнього тощо, поки кількість помилок не буде дорівнювати нулю.

Алгоритм Розенблатта

1. $\vec{w}^{(0)} := 0$ (вектор ваг).
2. $l := 0$ (лічильник корекцій).
3. $k := 0$ (лічильник ітерацій).

4. $m := N$ (лічильник помилок).

5. Якщо $m > 0,\ m := 0,\ k := k\ mod\ N + 1$, то подати на вхід вектор ознак k-го об'єкта $\vec{x}^{(k)}$; інакше перейти на крок 11.

6. Обчислити лінійну комбінацію

$$g\left(\vec{x}^{(k)}\right) = \sum_{i=1}^{n} w_i^{(k)} x_i^{(k)} + w_0^{(k)}.$$

7. Якщо $\vec{x}^{(k)} \in C_1$ і $g\left(\vec{x}^{(k)}\right) > 0$, то перейти на крок 5.

8. Якщо $\vec{x}^{(k)} \in C_2$ і $g\left(\vec{x}^{(k)}\right) < 0$, то перейти на крок 5.

9. Якщо $\vec{x}^{(k)} \in C_1$ і $g\left(\vec{x}^{(k)}\right) < 0$ (помилкова класифікація), то $m := m + 1; l := l + 1; w^{(k)} := w^{(k-1)} + x^{(k)}$ і перейти на крок 5.

10. Якщо $\vec{x}^{(k)} \in C_2$ і $g\left(\vec{x}^{(k)}\right) > 0$ (помилкова класифікація), то $m := m + 1; l := l + 1; w^{(k)} := w^{(k-1)} - x^{(k)}$ і перейти на крок 5.

11. Вихід: вектор w.

Відповідь на запитання, чи збігається процес навчання за алгоритмом Розенблатта, тобто чи дорівнює кількість помилок нулю після навчання, дає теорема Новікова [14].

Теорема 3. *Якщо навчальні вибірки можна розділити смугою шириною* 2δ *і всі вони лежать у кулі радіусом* $R = \max\limits_{1 \le i \le N} ||\vec{x}_i||$, *то навчання за алгоритмом Розенблатта збігається і кількість корекцій не перевищує* $\left(\frac{R}{\delta}\right)^2$.

Доведення. Позначимо напрямний вектор роздільної смуги як v, а мітки класів C_1 і C_2 як $\omega_1 = -1$ та $\omega_2 = 1$. Це означає, що

корекція вектора ваг $w^{(k-1)}$ здійснюється шляхом додавання до нього вектора $\omega_i \vec{x}_i$ такого, що

$$\left(\vec{w}^{(k-1)}, \vec{v}\right) < 0, \;\; ||\omega_i \vec{x}_i|| \leq R.$$

Отже,

$$\left(\vec{w}^{(k)}, \vec{v}\right) = \left(\vec{w}^{(k-1)}, \vec{v}\right) + (\omega_i \vec{x}_i, \vec{v}) \geq \left(\vec{w}^{(k-1)}, \vec{v}\right) + \delta||\vec{v}||.$$

Звідси випливає, що $\left(\vec{w}^{(k)}, \vec{v}\right) \geq k\delta||\vec{v}||$. З нерівності Коші – Буняковського маємо:

$$||\vec{w}^{(k)}|| \geq k\delta. \tag{6.2}$$

З іншого боку,

$$\begin{aligned} ||\vec{w}^{(k)}||^2 &= ||\vec{w}^{(k)} + \omega_i \vec{x}_i||^2 \leq \\ &\leq ||\vec{w}^{(k-1)}||^2 + ||\omega_i \vec{x}_i||^2 \leq ||\vec{w}^{(k-1)}||^2 + R^2. \end{aligned} \tag{6.3}$$

Нерівності (6.2) і (6.3) є сумісними лише за умови, що

$$k \leq \left(\frac{R}{\delta}\right)^2.$$

Таким чином, кількість кроків алгоритму є скінченною. □

6.1.2. Оптимізаційна трактовка персептрона Розенблатта

Алгоритм Розенблатта допускає оптимізаційну інтерпретацію [15]. Уведемо до розгляду кусково-лінійну функцію

$$J(w) = \sum_{x \in Y} \delta_x (w, x),$$

де Y — множина неправильно класифікованих об'єктів,

$$\delta_x = \begin{cases} -1, \text{ якщо } x \in C_1, \\ 1, \text{ якщо } x \in C_2. \end{cases}$$

У такому випадку класифікація об'єктів за допомогою роздільної гіперплощини, яка визначається коефіцієнтами $w = (w_0, w_1, ..., w_n)$, еквівалентна задачі

$$J(w) = \sum_{x \in Y} \delta_x (w, x) \to \min_w .$$

Мінімізуємо цю функцію за методом градієнтного спуску:

$$w^{(k)} = w^{(k-1)} - \rho^{(k-1)} \frac{dJ(w)}{dw} = w^{(k-1)} - \rho^{(k-1)} \sum_{x \in Y} x \delta_x.$$

Таким чином, алгоритм Розенблатта є різновидом алгоритму градієнтного спуску. Для його збіжності ряд $\sum\limits_{k=0}^{\infty} |\rho_k|$ повинен розбігатися, а ряд $\sum\limits_{k=0}^{\infty} |\rho_k^2|$ — збігатися.

6.2. Багатошаровий персептрон

Шаром персептрона є суматор, який імітує роботу нейронів. Узагальнюючи цей факт, можна дати таке означення.

Означення 25. ***Шаром*** штучної нейронної мережі є сукупність її елементів, що імітують накопичення потенціалу нейрона.

Зауваження 3. Елементи штучної нейронної мережі, на які надходять ознаки об'єкта, а також елементи, що зберігають результати розпізнавання, також називають вхідним і вихідним шарами, але вони не використовуються для класифі-

кації мереж. Тому, незважаючи на наявність трьох фізичних шарів, персептрон Розенблатта називають **одношаровим**.

Теорема Новікова дає підстави сподіватися, що нейронні мережі можна навчити розв'язувати складні задачі розпізнавання образів. Утім нагадаємо, що основним припущенням цієї теореми є лінійна роздільність навчальних множин. У випадку множин, які не можна розділити гіперплощиною, виникають серйозні проблеми. Розглянемо, наприклад, чотири точки: $A_1 = (0,0)$, $A_2 = (1,1)$, $B_1 = (0,1)$ і $B_2 = (1,0)$. Будемо вважати, що точки A_1 та A_2 належать класу A, а точки B_1 і B_2 — класу B. У цих точках легко впізнати значення логічної функції XOR, а в класах A та B — класи одиниць і нулів. Ці класи неможливо розділити лінією на площині. Намагаючись побудувати лінійну роздільну функцію для цієї задачі за допомогою одношарового персептрона Розенблатта, будемо перебирати всі лінійні комбінації точок $w_1x_1 + w_2x_2$, і жодна з них не зможе описати таку лінію, щоб точки A_1 та A_2 лежали з одного боку, а точки B_1 та B_2 — з іншого. Для виходу з цієї ситуації було запропоновано побудувати багатошарові штучні нейронні мережі.

Згадаємо, що операцію XOR можна виразити за допомогою операцій OR і AND.

$$A\ XOR\ B = A\ OR\ B\ AND\ (NOT\ A\ AND\ NOT\ B)$$

Логічна функція OR набуває значення 0, якому відповідає точка $A_1 = (0,0)$, що утворює клас A, і 1, якому відповідають точки $B_1 = (0,1)$, $B_2 = (1,0)$ і $B_3 = (1,1)$. Легко зауважити, що ці точки розділяються прямою $x_1 + x + 2 = \frac{1}{2}$.

Логічна функція AND набуває значення 0, якому відповідають точки $A_1 = (0,0), A_2 = (1,0), A_3 = (0,1)$, що утворюють клас A, і 1, якому відповідає точка $B_1 = (1,1)$. Легко зауважити, що ці точки розділяються прямою $x_1 + x + 2 = \frac{3}{2}$.

Таким чином, застосувавши суперпозицію двох одношарових персептронів Розенблатта, розділимо точки із задачі про

функцію XOR двома лініями. Спочатку побудуємо лінійну роздільну функцію для операції OR, на виході якої отримуємо значення y_1 та y_2, які надходять на вхід наступного персептрона (додаткового шару нейронів). У термінах значень y_1 та y_2 розв'язок задачі про функцію XOR записується як $y_1 - y_2 = \frac{1}{2}$.

6.2.1. Оптимізаційна трактовка багатошарової нейронної мережі

Побудову багатошарової штучної нейронної мережі для багатокласової класифікації можна описати як оптимізаційну задачу [12]. Отримана таким чином штучна нейронна мережа називається персептроном Румпельхарта. Позначимо кількість класів як m, кількість навчальних вибірок — як N, кількість прихованих шарів у нейронній мережі — як L, кількість нейронів на l-му шарі — як K_l. На вхід нейронної мережі подається вектор ознак $\vec{x}_i = \left(x_1^{(i)}, ..., x_n^{(i)}\right)$, $i = 1, ..., N$, на виході очікується вектор $\vec{\omega}_i = \left(\omega_1^{(i)}, ..., \omega_m^{(i)}\right)$, $i = 1, ..., N$.

Нехай на i-й навчальній вибірці мережа видає результат $\hat{\vec{\omega}}_i$. Він може збігатися або не збігатися з очікуваним результатом $\vec{\omega}_i$. Позначимо помилку на i-й навчальній вибірці як

$$\epsilon_i = \frac{1}{2} \sum_{j=1}^{m} \left(\hat{\vec{\omega}}_j^{(i)} - \vec{\omega}_j^{(i)}\right)^2,$$

а функціонал помилок — як

$$J(w) = \sum_{i=1}^{N} \epsilon_i = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{m} \left(\hat{\vec{\omega}}_j^{(i)} - \vec{\omega}_j^{(i)}\right)^2. \qquad (6.4)$$

Навчання мережі зводиться до розв'язання оптимізаційної задачі: знайти

$$\arg\min_w J(w).$$

6.2.2. Алгоритм зворотного розповсюдження помилки

Позначимо функцію активації нейрона як $f(s)$. Перетворення вхідного сигналу на j-му нейроні l-го шару обчислимо як $x_j^{(l)} = f\left(s_j^{(l)}\right)$, де $s_j^{(l)} = \left(w_j^{(l)}, x\right)$. Корекцію ваги w_{ij} на l-му шарі обчислимо за методом градієнтного спуску:

$$\Delta w_{ij}^{(l)} = -\rho \frac{\partial J}{\partial w_{ij}^{(l)}} = -\rho \frac{\partial J}{\partial x_j^{(l)}} \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = -\rho \delta_j^{(l)} x_i^{(l)},$$

де $\delta_j^{(l)} = \frac{\partial J}{\partial x_j^{(l)}} \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}}$ — помилка на j-му нейроні l-го шару.

Виразимо помилку на j-му нейроні l-го шару через помилку на $(l+1)$-му шарі.

$$\delta_j^{(l)} = \frac{\partial J}{\partial x_j^{(l)}} \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \left(\sum_{k=1}^{K_{l+1}} \frac{\partial J}{\partial x_k^{(l+1)}} \frac{\partial x_k^{(l+1)}}{\partial s_k^{(l+1)}} \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l+1)}} \right) \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}}.$$

Ураховуючи, що $\frac{\partial s_k^{(l+1)}}{\partial x_j^{(l+1)}} = w_{jk}^{(l+1)}$ і $\frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = f'\left(s_j^{(l)}\right)$, маємо:

$$\delta_j^l = \left(\sum_{k=1}^{K_{l+1}} \delta_k^{(l+1)} w_{jk}^{(l+1)} \right) f'\left(s_j^{(l)}\right).$$

Отже,

$$\delta_j^l = \left(\sum_{k=1}^{K_{l+1}} \delta_k^{(l+1)} w_{jk}^{(l+1)} \right) f'\left(s_j^{(l)}\right), \quad l = L-1, L-2, ..., 1.$$

На останньому шарі

$$\delta_j^L = \left(x_j^{(L)} - \omega_j\right) f'\left(s_j^{(l)}\right).$$

Назва алгоритму зворотного розповсюдження помилки пояснюється тим, что корекція ваги на кожному шарі обчислюється через корекції на наступних шарах, що можна трактувати як зворотне розповсюдження помилки від останнього шару до першого.

6.2.3. Приклади активаційних функцій

Для того, щоб обчислити помилку для кожного нейрона, треба продиференціювати його активаційну функцію. Знаючи цю функцію заздалегідь, можна спростити обчислення в алгоритмі зворотного розповсюдження помилок [23].

Логістична функція

Логістична функція j-го нейрона на l-му шарі має вигляд

$$f\left(s_j^{(l)}\right) = \frac{1}{1+\exp\left(-as_j^{(l)}\right)},\ a>0. \tag{6.5}$$

Похідна цієї функції

$$f'\left(s_j^{(l)}\right) = \frac{a\exp\left(-as_j^{(l)}\right)}{\left(1+\exp\left(-as_j^{(l)}\right)\right)^2},\ a>0. \tag{6.6}$$

Ураховуючи, що за означенням $f\left(s_j^{(l)}\right) = x_j^{(l)}$, можна переписати похідну:

$$f'\left(s_j^{(l)}\right) = ax_j^{(l)}\left(1-x_j^{(l)}\right). \tag{6.7}$$

Отже, для довільного нейрона на прихованому шарі в алгоритмі зворотного розповсюдження помилки з логістичною функцією активації маємо:

$$\delta_j^l = \left(\sum_{k=1}^{K_{l+1}} \delta_k^{(l+1)} w_{jk}^{l+1} \right) f'\left(s_j^{(l)}\right) =$$

$$= a x_j^{(l)} \left(1 - x_j^{(l)}\right) \sum_{k=1}^{K_{l+1}} \delta_k^{(l+1)} w_{jk}^{(l+1)}. \tag{6.8}$$

Гіперболічний тангенс

Активаційна функція j-го нейрона на l-му шарі, що задається гіперболічним тангенсом, описується формулою

$$f\left(s_j^{(l)}\right) = a \tanh\left(b s_j^{(l)}\right),\ a > 0,\ b > 0. \tag{6.9}$$

Похідна цієї функції

$$\begin{aligned} f'\left(s_j^{(l)}\right) &= ab\left(1 - \tanh^2\left(b s_j^{(l)}\right)\right) = \\ &= \frac{b}{a}\left(a - s_j^{(l)}\right)\left(a + s_j^{(l)}\right). \end{aligned} \tag{6.10}$$

Отже, для довільного нейрона на прихованому шарі в алгоритмі зворотного розповсюдження помилки з логістичною функцією активації маємо:

$$\begin{aligned} \delta_j^l &= \left(\sum_{k=1}^{K_{l+1}} \delta_k^{(l+1)} w_{jk}^{(l+1)} \right) f'\left(s_j^{(l)}\right) = \\ &= \frac{b}{a}\left(a - s_j^{(l)}\right)\left(a + s_j^{(l)}\right) \sum_{k=1}^{K_{l+1}} \delta_k^{(l+1)} w_{jk}^{(l+1)}. \end{aligned} \tag{6.11}$$

6.2.4. Метод стохастичного градієнта

Мінімізація функціонала помилок (6.4) — не єдиний спосіб навчання штучних нейронних мереж. Замість мінімізації сумарної помилки можна мінімізувати емпіричний ризик, що задається диференційовною функцією [4]:

$$\arg\min_{w} J(w) = \sum_{i=1}^{N} \mathcal{L}\left(\omega_i\left(w, \vec{x}_i\right)\right), \tag{6.12}$$

де N — кількість навчальних вибірок, ω_i — мітка i-го класу.

Застосуємо для розв'язання оптимізаційної задачі (6.12) метод градієнтного спуску:

$$\begin{aligned} w^{(k+1)} &= w^k - \rho_k \frac{\partial J(\vec{w})}{\partial \vec{w}} = \\ &= w^k - \rho_k \mathcal{L}'\left(\omega_i\left(w, \vec{x}_i\right)\right) x_i \omega_i, \quad k = 1, 2, ... \end{aligned} \tag{6.13}$$

де $\frac{\partial J(\vec{w})}{\partial \vec{w}} = \left(\frac{\partial J(\vec{w})}{\partial w_j}, ..., \frac{\partial J(\vec{w})}{\partial w_n}\right)^T$.

Для цього методу справджується теорема [14].

Теорема 4. *Якщо виконуються умови:*

1) *навчальні вибірки* $(\vec{x}_i; \omega_i)$ *є незалежними й мають однаковий розподіл* F*;*

2) *для будь-якого* w *випадкова величина* $\frac{\partial J(\vec{w})}{\partial \vec{w}}$*, що залежить від випадкового вектора* $(\vec{x}; \omega)$ *з розподілом* F*, має скінченне математичне сподівання та дисперсію;*

3) $\rho_k > 0$ *при всіх* k *і* $\rho_k \to 0$ *при* $k \to \infty$*;*

4) *ряд* $\sum\limits_{k=1}^{\infty} \rho_i$ *розбігається;*

5) *ряд* $\sum\limits_{k=1}^{\infty} \rho_i^2$ *збігається;*

то з імовірністю 1 ітераційна процедура стохастичного градієнтного спуску (6.13) збігається до локального мінімуму математичного сподівання помилки.

Зауваження 4. Підкреслимо, що класичний персептрон Розенблатта і теорема Новікова ґрунтуються на припущенні про лінійну роздільність множин навчальних вибірок. Якщо ж ці множини не є лінійно роздільними (див. задачу про функцію XOR), то персептрон Розенблатта не є ефективним. Обмеження персептрона Розенблатта були виявлені в [16].

Розділ 7.

Метод потенціальних функцій

Ідея методу потенціальних функцій має фізичну природу. Як відомо, згідно із законом Кулона вплив заряду на точку спадає пропорційно квадрату відстані до неї. Отже, потенціал може використовуватися як оцінка віддаленості точки від заряду. У цьому розумінні метод потенціальних функцій є різновидом методу найближчого сусіда. Якщо поле утворене кількома зарядами, то потенціал у кожній точці поля дорівнює сумі потенціалів, створюваних у цій точці кожним із зарядів. Якщо заряди, що утворюють поле, розташовані компактно, то потенціал поля досягає максимального значення в центрі тяжіння множини точок і спадає в міру віддалення від нього. Таким чином, якщо точку в просторі ознак інтерпретувати як заряд, то ці міркування можна формалізувати та звести до обчислення певної суми функцій, що описують потенціали зарядів — точок навчальної множини.

Розглянемо задачу бінарної класифікації. Кожному образу поставимо в однозначну відповідність точку в просторі ознак X. Далі будемо припускати, що класи C_1 і C_2 не перетинаються. Зазвичай топологічні властивості простору ознак допускають застосування малої леми Урисона, з якої випливає, що в просторі ознак X існує неперервна функція Φ, яка набуває значення більше нуля в точках класу C_1 і менше нуля —

у точках класу C_2. Зауважимо, що таких функцій може бути багато (і навіть нескінченно багато). У процесі навчання для кожного образу R_k, якому відповідає точка x_k, обчислюється функція $K(x, x_k)$, яка називається потенціальною, а побудова роздільної функції Φ за навчальною послідовністю образів $A_1, A_2, ..., A_k, ...$ зводиться до побудови послідовності потенціальних функцій $K(x, x_1), K(x, x_2), ..., K(x, x_k), ...$, що збігаються до функції Φ.

Інтуїтивна ідея методу полягає в тому, щоб обчислити загальні потенціали обох класів:

$$K_1(x) = \sum_{x_k \in C_1} K(x, x_k),$$

$$K_2(x) = \sum_{x_k \in C_2} K(x, x_k),$$

а потім знайти роздільну функцію у вигляді їх різниці:

$$\Phi(x) = K_1(x) - K_2(x).$$

Якщо при класифікації нової точки x^* виконується умова

$$K_1(x) > K_2(x),$$

то $x^* \in C_1$, інакше $x^* \in C_2$.

Звідси випливає, що роздільна функція є додатною на точках із класу C_1 і від'ємною на точках із класу C_2.

Математична формалізація цієї ідеї зводиться до відновлення певної функції за допомогою рекурентних процедур і обґрунтування їх збіжності.

7.1. Загальна схема

Метод потенціальних функцій ґрунтується на припущенні, що існує система функцій $\varphi_1, \varphi_2, ..., \varphi_k, ...$, яка дозволяє для

будь-яких двох диз'юнктних класів знайти таке число N, що функцію Φ можна подати формулою

$$\Phi(x) = \sum_{k=1}^{N} c_k \varphi_k(x). \tag{7.1}$$

Зазвичай простір X є гільбертовим, тому в новому просторі існує повна система функцій $\varphi_1, \varphi_2, ..., \varphi_k, ...$, за допомогою яких функцію Φ можна подати у вигляді ряду Фур'є

$$\Phi(x) = \sum_{k=1}^{\infty} c_k \varphi_k(x). \tag{7.2}$$

Прагнучи подати роздільну функцію Φ у вигляді суми (а не ряду), уведемо до розгляду N-вимірний простір Y, на який простір ознак X відображається за правилом $y_k = \varphi_k(x)$, $k = 1, 2, ..., N$. Таким чином, функція Φ відображається в лінійну функцію $\sum\limits_{k=1}^{N} a_k y_k(x)$ і набуває значення більше нуля в точках класу C_1 і менше нуля — у точках класу C_2. Легко бачити, що в просторі Y функція Φ є лінійною (відносно точок y).

Потенціальна функція розглядається як функція двох векторів:

$$K(x, x^*) = \sum_{k=1}^{\infty} \alpha_k^2 \varphi_k(x) \varphi_k(x^*), \tag{7.3}$$

де $\{\varphi_k\}$ — лінійно незалежна система функцій; α_k^2 — дійсні числа, що не обертаються на нуль одночасно; x^* — точка, яка обчислюється в процесі навчання.

Вважатимемо, що функції φ_k і $K(x, x^*)$ обмежені на просторі ознак. Першому образу P_1 ставимо у відповідність точку ознак x_1, за якою будується потенціальна функція

$$\Phi_1(x) = \begin{cases} K(x, x_1), & \text{якщо } x \in C_1, \\ -K(x, x_1), & \text{якщо } x \in C_2. \end{cases} \tag{7.4}$$

На k-му кроці отримуємо потенціальну функцію $\Phi_k(x)$. На (k+1)-му кроці, обчислюючи потенціал у точці x_{k+1}, можемо отримати такі значення потенціалу:

$$\begin{aligned}
&x_{k+1} \in K_1 \Rightarrow \Phi_k(x_{k+1}) > 0 \Rightarrow \Phi_{k+1}(x) = \Phi_k(x), \\
&x_{k+1} \in K_2 \Rightarrow \Phi_k(x_{k+1}) < 0 \Rightarrow \Phi_{k+1}(x) = \Phi_k(x), \\
&x_{k+1} \in K_1 \Rightarrow \Phi_k(x_{k+1}) < 0 \Rightarrow \\
&\qquad \Rightarrow \Phi_{k+1}(x) = \Phi_k(x) + \Phi(x, x_{k+1}), \\
&x_{k+1} \in K_2 \Rightarrow \Phi_k(x_{k+1}) > 0 \Rightarrow \\
&\qquad \Rightarrow \Phi_{k+1}(x) = \Phi_k(x) - \Phi(x, x_{k+1}).
\end{aligned}$$

Таким чином,

$$\Phi_k(x) = \sum_{x_i^- \in V_1} K(x, x_i) + \sum_{x_j^+ \in V_2} K(x, x_j), \tag{7.5}$$

де x_i — точки з класу C_1, на яких класифікатор робить помилку.

Корекцію роздільної функції можна здійснювати порізному, тому існують кілька алгоритмів, заснованих на методі потенціальних функцій.

Вважатимемо, що початкове наближення функції $\Phi_0(x)$ у точці $x_{(0)}$ дорівнює нулю. Тоді корекцію роздільної функції можна здійснити так:

$$\Phi_{k+1}(x) = \Phi_k(x) +$$

$$+\alpha_{k+1}\operatorname{sgn}\left[\Phi(x_{k+1}) - \Phi_k(x_{k+1})\right] K(x, x_{k+1}), \tag{7.6}$$

де $\Phi(x_{k+1})$ — істинне значення роздільної функції у точці x_{k+1}; (α_k) — послідовність чисел, що задовольняє умови:

$$\alpha_k \to 0, \ \sum_{k=1}^{\infty} \alpha_k = \infty, \ \sum_{k=1}^{\infty} \alpha_k^2 < +\infty.$$

Інакше кажучи, $\{\alpha_k\} \in (c_0 \cap \ell_2) \setminus \ell_1$. Це еквівалентно обчисленню коефіцієнтів розвинення функції $\Phi(x)$ за правилом:

$$c_k^{(n+1)} = c_{k+1}^{(n)} + \\ + \alpha_{k+1} \mathrm{sgn}\left[\Phi(x_{k+1}) - \sum_{k=1}^{N} c_k \varphi_k(x_{k+1})\right] \varphi_k(x_{k+1}). \quad (7.7)$$

Збіжність цього ітераційного процесу обґрунтовується теоремою 5.

Теорема 5. *Нехай x_k — послідовність незалежних випадкових точок із простору ознак X, $P(x)$ — щільність імовірності появи цих точок, а $\Phi(x)$ — функція, що записується формулою*

$$\Phi(x) = \sum_{k=1}^{N} c_k \varphi_k(x).$$

Тоді послідовність функцій $\Phi_k(x)$, $k = 1, 2, ...$, що задаються рекурентними співвідношеннями (7.6), задовольняє умову

$$P\left\{\lim_{k\to\infty} \int_X |\Phi(x) - \Phi_i(x)| P(x)\, dx = 0\right\} = 1. \quad (7.8)$$

Другий варіант:

$$\Phi_0(x) = 0,$$

$$\Phi_{k+1}(x) = \Phi_k(x) + \frac{1}{\lambda}\left[\Phi(x_{k+1}) - \Phi_k(x_{k+1})\right] K(x, x_{k+1}), \quad (7.9)$$

де $\lambda > \frac{1}{2} \max K(x, x^*)$.

Це еквівалентно обчисленню коефіцієнтів розвинення фун-

кції $\Phi(x)$ за таким правилом:

$$c_{k+1}^{(n+1)} = c_{k+1}^{(n)} + \\ + \frac{1}{\lambda}\left[\Phi(x_{k+1}) - \sum_{k=1}^{N} c_k \varphi_k(x_{k+1})\right] \varphi_k(x_{k+1}). \tag{7.10}$$

Збіжність цієї процедури випливає з теореми 6.

Теорема 6. *За умов теореми 5 послідовність функцій* $\Phi_k(x)$, $k = 1, 2, ...,$ *що визначаються співвідношенням (7.8), задовольняє умову*

$$P\left\{\lim_{k\to\infty} \int_X (\Phi(x) - \Phi_i(x))^2 P(x)\, dx = 0\right\} = 1. \tag{7.11}$$

Зауваження 5. Варіанти методу потенціальних функцій, які ґрунтуються на формулах (7.6) і (7.9), називаються *машинною* реалізацією, а варіанти, які ґрунтуються на формулах (7.7) і (7.10) — *персептронною* реалізацією. Вибір назви для персептронної реалізації пояснюється тим, що цю схему можно описати за допомогою персептрона Розенблатта.

7.2. Геометрична інтерпретація

Нехай у просторі X існує функція $\Phi(x)$, що розділяє множини $A \subset C_1$ та $B \subset C_2$ і задовольняє умови (7.1) та (7.2). Тоді в просторі Y існує роздільна гіперплощина Γ, що проходить через початок координат з напрямним вектором $\vec{c}$. Відобразимо множину B симетрично щодо початку координат у множину B' (тобто замінимо кожний вектор x вектором $-x$) і отримаємо множину $S = A \cup B'$. За припущенням множини A та B' розділяються гіперплощиною Γ, тобто множина S лежить по один бік від площини Γ. Розглянемо послідовності точок $M = \{x_1, x_2, ..., x_n, ...\}$ із простору X та їх образи

$M^* = \{y_1, y_2, ..., y_n, ...\}$ у просторі Y.

Потенціальна функція в просторі Y може бути подана як

$$U(Z, Z^*) = ZZ^*, \tag{7.12}$$

де

$$Z = \{z_1, z_2, ...\} = \{\alpha_1\varphi_1(x), \alpha_2\varphi_2(x), ...\}$$

і

$$Z^* = \{z_1^*, z_2^*, ...\} = \{\alpha_1\varphi_1(x^*), \alpha_2\varphi_2(x^*), ...\}.$$

Отже, співвідношення (7.5) можна переписати:

$$\Phi_k(Z) = \sum_{Z^{p(-)} \in M^*} (Z, Z^p). \tag{7.13}$$

Виправлення помилки відбувається тоді, коли $\Phi_k(Z) < 0$. Отже, корекція помилки на (k+1)-му кроці відбувається, коли

$$z_{k+1} \sum_{m=1}^{k} z_m < 0. \tag{7.14}$$

Таким чином, перша точка $Z^{(1)}$ з множини M^* зумовлює побудову площини $U_1(Z) = \left(Z, Z^1\right) = 0$ з напрямним вектором $Z^{(1)}$. Якщо наступна точка з множини M^* лежить у тому самому підпросторі, що й напрямний вектор Z^1, то помилка відсутня й роздільна площина не уточнюється. Якщо ж наступна точка потрапляє в протилежний підпростір, то відбувається корекція. При цьому попередній напрямний вектор складається із вектором точки, що вимагала корекції, та їхня сума береться як новий напрямний вектор.

Після k корекцій напрямний вектор дорівнює $\sum\limits_{m=1}^{k} Z_m$ і гіперплощина, що проходить через початок координат, приймається як роздільна.

7.3. Оптимізаційна інтерпретація

Розглянемо N функцій $\Phi(\vec{c}, x_i), i = 1, ..., N$, де x — випадкова величина і $\vec{c} = (c_1, c_2, ..., c_N)$. Запишемо систему рівнянь регресії щодо коефіцієнтів c_i:

$$M(\Phi(\vec{c}, x_i)) = 0, \tag{7.15}$$

де $M(\Phi(\vec{c}, x_i))$ — математичне сподівання функції $\Phi(\vec{c}, x_i)$ за величиною x.

Припустимо, що розподіл імовірності випадкової величини x є невідомим, але за послідовними точками $x_1, x_2, ...$ для будь-якого числа c можна обчислити значення функції $\Phi(\vec{c}, x_i)$. За таких умов Г. Роббінс і С. Монро запропонували спеціальну рекурентну процедуру:

$$c_i^{(n+1)} = c_i^{(n)} + \gamma_i \Phi(\vec{c}, x_i), i = 1, 2, ..., N, \tag{7.16}$$

де $\sum\limits_{i=1}^{\infty} \gamma_i = +\infty$ і $\sum\limits_{i=1}^{\infty} \gamma_i^2 < +\infty$.

У скінченновимірному випадку персептронний варіант методу потенціальних функцій може бути поданий як спеціальний варіант процедури Роббінса – Монро і, відповідно, зведений до мінімізації функціонала

$$J(c) \to \min_c, \tag{7.17}$$

де $J(c) = M(G(\vec{c}, x))$ і G — невід'ємна функція, що обертається на нуль, коли $\sum\limits_{i=1}^{N} c_i \varphi_i(x) = \Phi(x)$.

Легко бачити, що в такому випадку персептронний варіант методу потенціальних функцій є різновидом методу стохастичного градієнта.

Розділ 8.

Логістична регресія

Розглянемо новий різновид лінійного класифікатора — метод логістичної регресії, який дозволяє оцінити ймовірність успіху в схемі випадкових випробувань (множині незалежних спостережень) і класифікувати об'єкти за цією ймовірністю. Метод має два варіанти: бінарний і мультиноміальний.

8.1. Бінарна логістична регресія

Уведемо позначення: $Y_i = 1$ — подія відбулася; $Y_i = 0$ — подія не відбулася; $P(Y_i = 1) = p_i\,, \quad P(Y_i = 0) = 1 - p_i$ — імовірності подій $Y_i = 1$ та $Y_i = 0$; $E(Y_i) = 0 \cdot (1-p_i) + 1 \cdot (p_i) = p_i$ — математичне сподівання події Y_i.

Задача полягає в прогнозуванні $P(Y_i = 1) = p_i$ на основі попередніх відомостей. Чи можна тут скористатися лінійною регресією $E(Y_i\,|X_i) = \beta_0 + \beta_1 X_i = p_i$? На жаль, за умови дискретності даних виникають деякі проблеми, а саме:

1. Лінійна регресія не задовольняє умову $0 \leq E\,(Y_i|X_i) = p_i \leq 0$.
2. Необхідною умовою для лінійної регресії є сталість дисперсії відгуків. У випадку ж дискретних даних маємо, що $D(Y_i) = p_i(1-p_i)$, тобто її значення залежить від X_i.

3. Ще одна необхідна умова лінійної регресії – нормальний розподіл похибки: $\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_i)$. Коли ж $Y_i = 1$, маємо: $\varepsilon_i = 1 - (\beta_0 + \beta_1 X_i)$.

Коли дані є дискретними, графік очікуваних відгуків має вигляд деякої S-кривої, яку називають логістичною кривою. Модель логістичної регресії описується формулою

$$E(Y_i \,|X_i) = p_i = \frac{e^{(\beta_0+\beta_1 X_i)}}{1 + e^{(\beta_0+\beta_1 X_i)}}.$$

Досліджувані параметри є в степені експоненти, тому спочатку потрібно звести модель до вигляду, коли p_i будуть залежати від X_i лінійно, а потім повернутися до оригінального вигляду моделі, щоб відобразити реальну залежність. Такі перетворення мають вигляд:

$$p = \frac{e^{(\beta_0+\beta_1 X)}}{1 + e^{(\beta_0+\beta_1 X)}},$$

$$1 - p = \frac{1 + e^{(\beta_0+\beta_1 X)}}{1 + e^{(\beta_0+\beta_1 X)}} - \frac{e^{(\beta_0+\beta_1 X)}}{1 + e^{(\beta_0+\beta_1 X)}} = \frac{1}{1 + e^{(\beta_0+\beta_1 X)}},$$

$$\frac{p}{1-p} = e^{(\beta_0+\beta_1 X)},$$

$$\frac{p}{1-p} \neq 1,$$

тому $\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X$.

Отже, маємо лінійну залежність від X. Тепер наблизимо ймовірність події p_i її частотою h_i. З рівності

$$\ln(\frac{h_i}{1-h_i}) = \beta_0 + \beta_1 X$$

знайдемо регресійні коефіцієнти й, повернувшись до вихідної

моделі, зможемо обчислити передбачувану ймовірність

$$\hat{p} = \frac{e^{(\beta_0+\beta_1 X)}}{1+e^{(\beta_0+\beta_1 X)}}.$$

8.2. Оптимізаційна інтерпретація

Найпоширенішим способом обчислення коефіцієнтів логістичної регресії є метод максимальної правдоподібності. Цей метод полягає в обчисленні максимуму функції правдоподібності, що виражає ймовірність спільної появи результатів вибірки $Y_1, Y_2, ..., Y_k$.

$$\arg\max_{\theta} L\left(Y_1, Y_2, ..., Y_n, \theta\right),$$

де θ — невідомий параметр.

Функція ln є монотонною, тому можна максимізувати не функцію L, а її натуральний логарифм $\ln(L)$, оскільки максимуми обох функцій збігаються. Апроксимуємо ймовірність події логіт-функцією:

$$P\left(Y=1|x\right) = f\left(z\right) = \frac{1}{1+e^{-z}},$$

де $z = \sum_{i=1}^{N} w_i x_i$ і w_i — регресійні коефіцієнти.

У бінарному випадку логарифмічна функція правдоподібності

$$L\left(\vec{w}\right) = \sum_{i=1}^{N} \left(Y_i \ln P_i\left(\vec{w}\right) + \left(1-Y_i\right) P_i \ln\left(1-\vec{w}\right)\right).$$

Градієнт функції правдоподібності

$$g = \sum_{i=1}^{N} \left(Y_i - P_i\right) X_i,$$

а гесіан функції правдоподібності

$$H = -\sum_{i=1}^{N} P_i \left(1 - P_i\right) \left(X_i, X_i\right),$$

де X_i — рядок матриці пояснювальних змінних.

Завдяки негативній визначеності гесіана логарифмічна функція має єдиний глобальний максимум. Для розв'язання оптимізаційної задачі можна застосовувати метод Ньютона – Рафсона або градієнтні методи, як от: метод градієнта, метод покоординатного спуску або метод спряжених градієнтів.

Зауваження 6. Неважно помітити, що логістичну регресію можна подати у вигляді персептрона Розенблатта із сигмоїдною функцією, активацією та вагами, що є регресійними коефіцієнтами.

Приклад 4. Для ілюстрації покажемо, як метод логістичної регресії можна використати для класифікації. Для цього введемо точку відсікання, відносно якої здійснюється класифікація. Це ймовірність успіху така, що ймовірність успіху для навчальних вибірок з першого класу менша, ніж для вибірок другого класу. За умовчанням, точка відсікання дорівнює 0,5.

Розглянемо одну задачу з медичної практики. Проаналізуємо вибірку хворих, які були прооперовані з приводу раку передміхурової залози. Після операції в крові цих хворих був виміряний показник CA-125 — відомий діагностичний маркер раку передміхурової залози. Оцінимо ймовірність успіху, яким у цій схемі є відсутність метастазів протягом п'яти років після операції. Метод логістичної регресії дозволяє для кожного пацієнта обчислити ймовірність появи метастазів, а також розділити групи хворих на два класи: групу ризику, у якій імовірність появи метастазів перевищує 0,5, і групу позитивного прогнозу, у членів якої вона не перевищує 0,5.

В аналізі використовувалися дані про маркер CA-125 і виживаність 60 хворих. Точка відсікання дорівнювала 0.5. Обчи-

слення показали, що цій точці відповідає концентрація CA-125, що дорівнює 12,171.

Таблиця. Специфічність і чутливість діагностики.

Подія	N	0	1	Усього	Показник
0	38	0	38	100,00	Специфічність
1	5	17	22	77,27	Чутливість
Усього	43	17	60	91,67	Точність

8.3. Множинна логістична регресія

Множинна логістична регресія використовує лінійну функцію відклику $f(\beta_k, x_i)$, яка оцінює ймовірність того, що i-та навчальна вибірка відповідає k-му результату:

$$f(\beta_k, x_i) = \vec{\beta}_k \vec{x}_i,$$

де $\vec{\beta}_k$ — рядок регресійних коефіцієнтів, що відповідає k-му результату, а $\vec{x}_i$ — i-та навчальна вибірка.

8.3.1. Сукупність незалежних бінарних моделей

Множинну логістичну регресію з N результатами можна подати як сукупність $N-1$ незалежних моделей бінарної логістичної регресії, у яких кожному результату відповідає окрема множина пояснювальних змінних.

$$\ln \frac{P(Y_i = 1)}{P(Y_i = K)} = \beta_1 X_i,$$
$$\ln \frac{P(Y_i = 2)}{P(Y_i = K)} = \beta_2 X_i,$$
$$\ln \frac{P(Y_i = K-1)}{P(Y_i = K)} = \beta_{K-1} X_i.$$

Елементарні перетворення дають такі результати:

$$P(Y_i = 1) = P(Y_i = K)e^{\boldsymbol{\beta}_1 \cdot \vec{X}_i},$$
$$P(Y_i = 2) = P(Y_i = K)e^{\boldsymbol{\beta}_2 \cdot \vec{X}_i},$$
$$\cdots\cdots$$
$$P(Y_i = K-1) = P(Y_i = K)e^{\boldsymbol{\beta}_{K-1} \cdot \vec{X}_i}.$$

Оскільки сума всіх імовірностей має дорівнювати одиниці, то отримуємо:

$$P(Y_i = K) = \frac{1}{1 + \sum\limits_{k=1}^{K-1} e^{\beta_k X_i}}.$$

Отже,

$$P(Y_i = 1) = \frac{e^{\beta_1 \cdot \vec{X}_i}}{1 + \sum\limits_{k=1}^{K-1} e^{\beta_k \cdot \vec{X}_i}},$$
$$P(Y_i = 2) = \frac{e^{\beta_2 \cdot \vec{X}_i}}{1 + \sum\limits_{k=1}^{K-1} e^{\beta_k \cdot \vec{X}_i}},$$
$$\cdots\cdots$$
$$P(Y_i = K-1) = \frac{e^{\beta_{K-1} \cdot \vec{X}_i}}{1 + \sum\limits_{k=1}^{K-1} e^{\beta_k \cdot \vec{X}_i}}.$$

Як і в методі бінарної логістичної регресії, оцінка параметрів множинної логістичної регресії здійснюється за методом максимальної правдоподібності, тобто зводиться до розв'язання оптимізаційної задачі щодо регресійних коефіцієнтів.

8.3.2. Узагальнення бінарної моделі

Інший спосіб побудови моделі множинної логістичної регресії полягає в безпосередньому узагальненні бінарної моделі за допомогою множника нормалізації.

$$\ln P(Y_i = 1) = \beta_1 \cdot \vec{X}_i - \ln Z,$$
$$\ln P(Y_i = 2) = \beta_2 \cdot \vec{X}_i - \ln Z,$$
$$\cdots\cdots$$
$$\ln P(Y_i = K) = \beta_K \cdot \vec{X}_i - \ln Z.$$

Нормалізуючий множник дозволяє задовольнити умову

$$\sum_{k=1}^{K} P(Y_i = k) = 1.$$

Шляхом елементарних перетворень легко отримати такі рівняння:

$$P(Y_i = 1) = \frac{1}{Z} e^{\beta_1 \cdot \vec{X}_i},$$
$$\cdots\cdots$$
$$P(Y_i = K) = \frac{1}{Z} e^{\beta_K \cdot \vec{X}_i}.$$

Підсумовуючи ці рівняння та прирівнюючи їх до одиниці, маємо:

$$1 = \sum_{k=1}^{K} P(Y_i = k) = \sum_{k=1}^{K} \frac{1}{Z} e^{\beta_k \cdot \vec{X}_i} = \frac{1}{Z} \sum_{k=1}^{K} e^{\beta_k \cdot \vec{X}_i}.$$

Таким чином, знаходимо множник нормалізації:

$$Z = \sum_{k=1}^{K} e^{\beta_k \cdot \vec{X}_i}.$$

За рівняннями (8.1)—(8.4) знаходимо ймовірності кожного результату:

$$P(Y_i = 1) = \frac{e^{\beta_1 \cdot \vec{X}_i}}{\sum\limits_{k=1}^{K} e^{\beta_k \cdot \vec{X}_i}},$$

$$P(Y_i = 2) = \frac{e^{\beta_2 \cdot \vec{X}_i}}{\sum\limits_{k=1}^{K} e^{\beta_k \cdot \vec{X}_i}},$$

$$\cdots\cdots$$

$$P(Y_i = K) = \frac{e^{\beta_K \cdot \vec{X}_i}}{\sum\limits_{k=1}^{K} e^{\beta_k \cdot \vec{X}_i}}.$$

Інакше кажучи,

$$P(Y_i = c) = \frac{e^{\beta_c \cdot \vec{X}_i}}{\sum\limits_{k=1}^{K} e^{\beta_k \cdot \vec{X}_i}}.$$

Таким чином, задача знаходження регресійних коефіцієнтів зводиться до оптимізаційної задачі:

$$\arg\max_{\beta_k} S(c, \beta_1 \cdot \vec{X}_i, \ldots, \beta_K \cdot \vec{X}_i).$$

Розділ 9.

Метод найближчого сусіда

Перш ніж розпочати аналіз методу найближчого сусіда, нагадаємо, що наші міркування завжди ґрунтуються на двох постулатах:

1. **Постулат про векторну модель:** об'єкт можна подати як елемент векторного простору ознак.
2. **Постулат про компактність:** переважна більшість об'єктів, що належать до одного класу, є ближчими один до одного, ніж до об'єктів іншого класу, і лежать в області з відносно простою межею.

Метод найближчого сусіда є найбільш інтуїтивно зрозумілим алгоритмом класифікації, адже порівняння за принципом подібності — це дуже розповсюджений і природний спосіб розпізнавання (якщо хтось ходить як качка і крякає як качка, то це і є качка). За правилом найближчого сусіда об'єкт x належить тому класу, якому належить найближчий об'єкт із навчальної вибірки (прецедент).

Очевидно, що основними поняттями методу найближчого сусіда є поняття, які описують "близькість" об'єктів у просторі ознак. Це можуть бути або метрика, або міра близькості.

Нагадаємо означення метрики.

Означення 26. Нехай X — довільна множина. Відображення $\rho : X \times X \to R^+$ називається ***метрикою***, якщо $\forall x, y, z \in X$ воно має такі властивості (аксіоми метрики):

1. $\rho(x, y) = 0 \Leftrightarrow x = y$ (аксіома тотожності).
2. $\rho(x, y) = \rho(y, x)$ (аксіома симетрії).
3. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ (нерівність трикутника).

Означення 27. Значення метрики $\rho(x, y)$ називається ***відстанню*** між векторами x та y.

Відповідно до постулату про векторну модель, кожному об'єкту можна поставити в однозначну відповідність вектор $\vec{x} = (x_1, x_2, ..., x_n)$ у векторному просторі ознак X і ввести в цьому просторі метрику, перетворивши його на метричний простір. Відстань між схожими об'єктами повинна бути малою, а між несхожими — великою. Отже, відносно кожного об'єкта y елементи навчальної вибірки можна упорядкувати за зростанням відстані до нього.

$$\rho(x_{(1)}, y) \leq \rho(x_{(2)}, y) \leq \cdots \leq \rho(x_{(n)}, y),$$

де $x_{(i)}$ — i-й елемент варіаційного ряду, або i-та порядкова статистика. Інакше кажучи, $x_{(i)}$ — об'єкт навчальної вибірки, що є i-м сусідом об'єкта y.

Отже, задачу пошуку найближчого сусіда можна сформулювати так:

$$\arg\min_{x_i \in X} \rho(x_i, y). \tag{9.1}$$

Нормалізація даних. Зауважимо, що ознаки, які використовуються при обчисленні відстані (наприклад, за евклідовою метрикою або будь-якою іншою), можуть коливатися в досить широкому діапазоні. Якщо одна з ознак коливається в ширшому діапазоні, ніж інші, то вона буде мати більший вплив на відстань, ніж інші. Якщо ми виходимо з припущення, що ознаки мають однакову вагу, то така ситуація є небажаною. Для

того, щоб уникнути її, перед обчисленням відстані виконують нормалізацію даних. Існує кілька способів нормалізації даним. Найпоширенішими є мінімаксна нормалізація та стандартизація.

Позначимо мінімальне значення k-ї ознаки в навчальній вибірці як $x_k^{\min} = \min\limits_{i=1,\dots,n} x_k^i$, а максимальне — як $x_k^{\max} = \max\limits_{i=1,\dots,n} x_k^i$.

Мінімаксна нормалізація k-ї ознаки i-го об'єкта здійснюється так:

$$\overline{x}_k^{(i)} = \frac{x_k - x_k^{\min}}{x_k^{\max} - x_k^{\min}}.$$

Позначимо середнє вибіркове k-ї ознаки як m_k:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_k^{(i)},$$

а оцінку стандартного відхилення k-ї ознаки — як s_k:

$$s_k = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_k^{(i)} - m_k\right)^2}.$$

Стандартизація k-ї ознаки:

$$\overline{x}_k = \frac{x_k - m_k}{s_k}.$$

Інакше кажучи, m_k — це середнє значення k-ї ознаки в навчальній вибірці, а s_k — його стандартне відхилення.

Після нормалізації всі ознаки об'єктів будуть змінюватися у відрізку $[0, 1]$. Іноді зручно, щоб ознаки коливалися в інтервалі $[-1, 1]$ із центром у нулі. Для цього використовують іншу форму мінімаксної нормалізації:

$$\overline{x}_k^{(i)} = \frac{x_k - \frac{x_k^{\max} + x_k^{\min}}{2}}{x_k^{\max} - x_k^{\min}}.$$

Після стандартизації всі ознаки об'єктів будуть мати наближено стандартизований нормальний розподіл, тобто приблизно нормальний розподіл із нульовим математичним сподіванням і одиничною дисперсією.

Нормалізація та стандартизація даних урівноважують вплив кожної ознаки. Однак, незважаючи на те, що обидва методи є достатньо ефективними, перевагу слід віддати стандартизації, тому що вона дає можливість застосувати статистичний аналіз і отримати обґрунтовані оцінки щодо відхилення об'єктів один від одного (наприклад, виявити викиди — об'єкти, що статистично значущо відрізняються від об'єктів того ж самого класу).

9.1. Метод k найближчих сусідів

Переваги й недоліки методу найближчого сусіда очевидні. Основні переваги — простота та швидкість. Основний недолік — нестійкість. Дійсно, зважаючи на те, що об'єкти навчальної вибірки є випадковими, немає жодних гарантій, що найближчий прецедент – це не випадковий сусід, а неодмінний супутник об'єкта, що класифікується. З цієї причини для підвищення точності класифікації часто визначають не одного, а k найближчих сусідів і відносять об'єкт до того класу, до якого належить більшість серед сусідів. Цю схему можна інтерпретувати як "голосування".

Є два способи голосування: без ваг та з вагами.

Якщо всі об'єкти вважаються рівноправними, то можна нехтувати вагами й застосувати алгоритм (9.1).

Якщо ж треба урахувати вагу, то доцільно зважити на фізичний закон обернених квадратів:

$$\arg\max_{x_i \in X} \sum_{i=1}^{n} \frac{1}{\rho^2(x_i, y)}.$$

9.2. Вибір кількості сусідів k

Вибір кількості сусідів — важлива задача, від якої залежить стійкість алгоритму. Єдиного рецепта вибору параметра k немає, оскільки він сильно залежить від даних. Якщо вибрати граничні значення, то метод найближчих сусідів стає або нестійким ($k = 1$), або виродженим ($k = m$), тобто втрачає здатність до узагальнення. Для пошуку параметра k, близького до оптимального, зазвичай використовують метод кросвалідації по k-блоках. Для цього навчальна вибірка випадковим чином розбивається на k диз'юнктних блоків. Один з блоків стає тестовою вибіркою, а решта — навчальною.

Позначимо вихідну навчальну вибірку як X, а вибірку, що грає роль тестової — як $\vec{x}_k$. Тоді навчальна вибірка в процесі крос-валідації — це різниця $X \setminus \vec{x}_k$. Шуканий параметр k визначається як розв'язок задачі

$$\arg\min_k \epsilon_k,$$

де $\epsilon_k = \frac{1}{k}\sum_{i=1}^{k} \mathcal{J}\left(X \setminus \vec{x}_k, x_k\right)$ — середня помилка класифікації на тестових вибірках.

9.3. Розпізнавання викидів

На практиці постулат про компактність виконується лише наближено, тобто деякі об'єкти можуть бути оточені сусідами того самого класу, а деякі бути далеко від інших.

Означення 28. Об'єкт, що оточений сусідами того самого класу, називається ***типовим***.

Означення 29. Об'єкт, що віддалений від інших об'єктів того самого класу, називається ***викидом***, або ***шумом***.

Розпізнавання типових об'єктів і шуму має велике значення. По-перше, типові об'єкти можна видалили з навчальної ви-

бірки без зменшення точності класифікації. Це дозволяє зменшити розмір задачі. По-друге, шум сильно впливає на точність класифікації, тому видалення шуму значно підвищує специфічність і чутливість алгоритму.

Часто об'єкти задаються не вектором, а матрицею, де стовпчики матриці є випадковими вибірками (наприклад, набором вимірювань k-ї ознаки). Це типова ситуація в медичних дослідженнях, коли в пацієнта беруть n клітин, у яких вимірюють m ознак. Визначення викидів і впорядкування таких об'єктів стає набагато складнішою задачею. Для розв'язання цієї задачі використовується, зокрема, поняття статистичної глибини.

Розділ 10.

Класифікація за статистичною глибиною

Як показано в попередньому розділі, задачі класифікації об'єктів часто зводяться до ранжування багатовимірних вибірок. Відповідно до загальновизнаної термінології, методи багатовимірного ранжування розділяються на маргінальні, редуковані, часткові та умовні. Маргінальні методи впорядковують вибірки за окремими компонентами. Редуковані методи обчислюють відстань кожної вибірки від центра розподілу. Часткове ранжування припускає розділення вибірок на групи однакових підвибірок. В умовних методах здійснюється впорядкування вибірок за обраним компонентом, що впливає на інші.

Велику популярність серед методів багатовимірного ранжування отримав підхід, що ґрунтується на концепції статистичної глибини вибірок відносно центра розподілу й відповідних методах пілінгу. Ці методи дозволяють урахувати геометричні властивості багатовимірних розподілів і є відносно простими для обчислень. Розглянемо один з них, що використовує еліпсоїд Петуніна.

10.1. Еліпсоїд Петуніна

Не обмежуючи загальності, опишемо алгоритм побудови еліпса Петуніна на площині, а потім перенесемо його в простір R^m при $m > 2$. Вихідними даними для алгоритму є множина багатовимірних векторів $M_n = \{\vec{x}_1, ..., \vec{x}_n\}$, де $\vec{x}_n = (x_n, y_n)$.

Еліпс Петуніна. На першому етапі побудуємо опуклу оболонку точок $M_n = \{(x_1, y_1), ..., (x_n, y_n)\}$. Знайдемо вершини опуклої оболонки (x_k, y_k) та (x_l, y_l), що лежать на діаметрі опуклої оболонки, тобто вершини, найбільш віддалені одна від одної. З'єднаємо точки (x_k, y_k) та (x_l, y_l) відрізком L. Знайдемо вершини опуклої оболонки (x_r, y_r) та (x_q, y_q), найбільш віддалені від L. З'єднаємо точки (x_r, y_r) та (x_q, y_q) відрізками L_1 та L_2, паралельними відрізку L. Проведемо через точки (x_k, y_k) та (x_l, y_l) відрізки L_3 та L_4, перпендикулярні відрізку L. Перетини відрізків L_1, L_2, L_3 і L_4 утворюють прямокутник П, сторони якого мають довжини a та b. Будемо вважати, що $a \leq b$. Переведемо лівий нижній кут прямокутника в початок нової системи координат з осями Ox' та Oy' за допомогою повороту й паралельного перенесення. Точки (x_1, y_1), ..., (x_n, y_n) перейдуть у точки (x'_1, y'_1), ..., (x'_n, y'_n). Відобразимо точки (x'_1, y'_1), (x'_2, y'_2), ..., (x'_n, y'_n) у точки $(\alpha x'_1, y'_1)$, $(\alpha x'_2, y'_2)$, ..., $(\alpha x'_n, y'_n)$, де $\alpha = \frac{a}{b}$. У результаті отримуємо сукупність точок, що лежать у квадраті S. Обчислимо центр (x'_0, y'_0) квадрата S і знайдемо відстані $r_1, r_2, ..., r_n$ від нього до кожної точки $(\alpha x'_1, y'_1)$, $(\alpha x'_2, y'_2)$, ..., $(\alpha x'_n, y'_n)$. Найбільше число $R = \max(r_1, r_2, ..., r_n)$ визначає коло з центром у точці (x'_0, y'_0) і радіусом R. У результаті всі точки $(\alpha x'_1, y'_1)$, $(\alpha x'_2, y'_2)$, ..., $(\alpha x'_n, y'_n)$ опиняються всередині кола з радіусом R. Розтягуючи це коло вздовж осі Ox' з коефіцієнтом $\beta = \frac{1}{\alpha}$ і виконуючи зворотні перетворення повороту й переносення, отримуємо еліпс Петуніна.

Еліпсоїд Петуніна. У m-вимірному просторі на першому кроці знайдемо дві вершини опуклої оболонки $\vec{x}_k$ та $\vec{x}_l$, що лежать на її діаметрі. З'єднаємо точки $\vec{x}_k$ та $\vec{x}_l$ відрізком L. Повернемо й перенесемо систему координат, щоб ді-

аметр опуклої оболонки лежав на осі Ox_1'. Побудуємо найменший прямокутний паралелепіпед, що містить точки $\vec{x}_1', ..., \vec{x}_n'$. Стискаючи прямокутний паралелепіпед, відобразимо точки в гіперкуб. Знайдемо центр $\vec{x}_0$ гіперкуба й обчислимо відстані $r_1, r_2, ..., r_n$ від нього до кожної точки. Знайдемо найбільше число $R = \max(r_1, r_2, ..., r_n)$ і побудуємо гіперкулю з центром у точці $\vec{x}_0$ і радіусом R. Застосовуючи до цієї гіперкулі зворотні операції розтягування, повороту й перенесення, одержимо еліпсоїд Петуніна у m-вимірному просторі. У результаті на кожному вкладеному еліпсоїді лежатиме по одній точці з вибірки, тобто відбуватиметься їхнє ранжування.

Теорема 7. *Якщо $\vec{x}_1, \vec{x}_2, ..., \vec{x}_n$ є незалежними й однаково розподіленими випадковими векторами з генеральної сукупності G, E_n — еліпсоїд Петуніна, що містить точки $\vec{x}_1, \vec{x}_2, ..., \vec{x}_n$, та $\overrightarrow{x}_{n+1} \in G$, то $P(\vec{x}_{n+1} \in E_n) = \frac{n}{n+1}$.*

10.2. Міра близькості

За аналогією з p-статистикою, для одновимірного випадку сконструюємо міру близькості, використовуючи як варіаційний ряд багатовимірні вибірки, побудовані при ранжуванні за допомогою еліпсоїдів Петуніна. Варіаційному ряду вибірок $\vec{x}_{(1)} \preceq \vec{x}_{(2)} \preceq ... \preceq \vec{x}_{(n)}$ поставимо у відповідність послідовність вкладених еліпсоїдів $E_{(1)} \subset E_{(2)} \subset ... \subset E_{(n)}$. Як випливає з теореми, імовірність того, що вибірка $\vec{x}$ з багатовимірної генеральної сукупності G задовольняє умові $\vec{x}_{(i)} \preceq \vec{x} \preceq \vec{x}_{(j)}$, дорівнює імовірності потрапити між еліпсами $E_{(i)}$ та $E_{(j)}$, тобто $\frac{j-i}{n+1}$. Ця обставина дозволяє побудувати p-статистику для багатовимірного випадку. Нагадаємо основні означення. Нехай $x = (x_1, ..., x_n) \in G$ — вибірка з генеральної сукупності G і p — деякий відомий чи невідомий показник, значення якого можуть залежати від вибірки x. Розглянемо дві неперервні функції $a(u_1, ..., u_n)$ та $b(u_1, ..., u_n)$ від n змінних $u_1, ..., u_n$, що задовольняють нерівності $a(u_1, ..., u_n) \leq b(u_1, ..., u_n)\ \forall (u_1, ..., u_n) \in$

R^n. Випадковий інтервал $(a(u_1,...,u_n), b(u_1,...,u_n)) = (a,b)$ називається довірчим інтервалом для p, що відповідає рівню значущості β, якщо $P(p \in (a,b)) = 1-\beta, (0 \le \beta \le 1)$; при цьому числа $a = a(x_1,...,x_n)$, $b = b(x_1,...,x_n)$ називаються довірчими межами для p, що відповідають рівню значущості β.

Означення 30. Інтервали

$$(a_k, b_k) = (a_k(x_1,...,x_k), b_k(x_1,...,x_k)),$$

де $k = 1,2,...$, називаються ***асимптотичними інтервалами*** для показників $p_i, i = 1,2,...,k,...$, що відповідають рівню значущості β, якщо

$$\underline{\lim}_{k\to\infty} P(p_k \in (a_k(x_1,...,x_k), b_k(x_1,...,x_k))) = 1-\beta, \qquad (10.1)$$

а кінці цих інтервалів $a_k(x_1,...,x_k)$ та $b_k(x_1,...,x_k)$ називаються ***асимптотичними довірчими межами***.

Означення 31. Величина β називається ***асимптотичним рівнем значущості*** послідовності (a_k, b_k), $k = 1,2,....$

Означення 32. Якщо $p_k = p \ \forall k = 1,2,...$, то інтервал (a_k, b_k) називається ***асимптотичним довірчим інтервалом*** показника p, а величина β — ***асимптотичним рівнем значущості інтервалу*** (a_k, b_k).

Позначимо як H гіпотезу про рівність неперервних функцій розподілу $F_1(u)$ та $F_2(u)$ генеральних сукупностей багатовимірних випадкових величин G_1 та G_1, відповідно. Нехай $(\vec{x}_1, \vec{x}_2, ..., \vec{x}_n) \in G_1$ та $(\vec{x}'_1, \vec{x}'_2, ..., \vec{x}'_n) \in G_2$, $\vec{x}_{(1)} \preceq \vec{x}_{(2)} \preceq ... \preceq \vec{x}_{(n)}$, $\vec{x}'_{(1)} \preceq \vec{x}'_{(2)} \preceq ... \preceq \vec{x}'_{(n)}$ — відповідні варіаційні ряди. Припустимо, що $F_1(u)$=$F_2(u)$. Позначимо як A_{ij}, $k = 1,2,...,m$ випадкову подію, яка полягає у тому, що $\vec{x}'_k$ потрапляє в область $E_{(j)} \backslash E_{(i)}$. Якщо $F_1(u)$=$F_2(u)$, то ймовірність p_{ij} цієї події об-

числюється за формулою

$$p_{ij}^{(n)} = \frac{j-i}{n+1}. \tag{10.2}$$

Покладемо

$$p_{ij}^{(1)} = \frac{h_{ij}m + g^2/2 - g\sqrt{h_{ij}(1-h_{ij})m + g^2/4}}{m+g^2}, \tag{10.3}$$

$$p_{ij}^{(2)} = \frac{h_{ij}m + g^2/2 + g\sqrt{h_{ij}(1-h_{ij})m + g^2/4}}{m+g^2}, \tag{10.4}$$

де h_{ij} — частота події A_{ij} у m випробуваннях і g=3. Розглянемо довірчі інтервали $I_{ij}^{(n)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right)$. Загальна кількість інтервалів $I_{ij}^{(n)}$ становить $N = \frac{n(n-1)}{2}$. Позначимо через L кількість тих інтервалів $I_{ij}^{(n)}$, що містять імовірності $p_{ij}^{(n)}$. Покладемо $h = \rho\left(\vec{x}, \vec{x}'\right) = \frac{L}{N}$. Оскільки h — частота випадкової події $B = \left\{p_{ij}^{(n)} \in I_{ij}^{(n)}\right\}$, що має ймовірність $p(B) = 1 - \beta$, то, поклавши $h_{ij} = h, m = N$ і $g = 3$ у формулах (10.3), (10.4), одержуємо довірчий інтервал $I^{(n)} = \left(p^{(1)}, p^{(2)}\right)$ для ймовірності $p(B)$. Статистика h називається р-статистикою. Вона є мірою близькості $\rho\left(\vec{x}, \vec{x}'\right)$ між вибірками $\vec{x}$ та $\vec{x}'$. Якщо гіпотеза H є істинною, то схема випробувань, у якій можуть з'являтися події $A_{ij}^{(k)}$, називається узагальненою схемою Бернуллі, а якщо гіпотеза H є хибною, то схема випробувань називається модифікованою схемою Бернуллі. У загальному випадку, коли може бути істинною будь-яка гіпотеза, як $F_1(u) = F_2(u)$, так і $F_1(u), \neq F_2(u)$, ця схема випробувань називається МР-схемою.

Теорема 8. *Якщо в узагальненій схемі випробувань Бернуллі виконуються умови* $n = m$, $0 < \lim\limits_{n\to\infty} p_{ij}^{(n)} = p_0 < 1$ *і* $0 < \lim\limits_{n\to\infty} \frac{i}{n+1} = p^* < 1$, *то асимптотичний рівень значущо-*

сті β *послідовності довірчих інтервалів* $I_{ij}^{(n)}$ *для ймовірностей* $p_{ij}^{(n)}$, *побудованих за правилом* $3s$, *не перевищує* $0,05$.

Теорема 9. *Якщо вибірки* $\vec{x} = (\vec{x}_1, ..., \vec{x}_n) \in G_1$ *і* $\vec{x}' = (\vec{x}'_1, ..., \vec{x}'_m) \in G_2$ *мають однаковий розмір, то асимптотичний рівень значущості інтервалу* $I^{(n)} = \left(p^{(1)}, p^{(2)}\right)$, *побудований за правилом* $3s$ *при* $g = 3$ *за допомогою формул (10.3), (10.4), не перевищує* $0,05$.

10.3. Обчислювальний експеримент

У межах обчислювального експерименту за допомогою статистичного пакету R було проведено попарне порівняння вибірок з генеральних сукупностей, що мають двовимірний нормальний розподіл, вектори математичних сподівань яких дорівнюють (0,0), (1,1), (2,2) і (3,3), відповідно, а коваріаційна матриця є одиничною. Крім того, було проведено попарне порівняння вибірок з генеральних сукупностей, що мають двовимірний нормальний розподіл, вектори математичних сподівань яких $(0, 0)$, а на діагоналі коваріаційної матриці стоять дисперсії $(1, 1)$, $(2, 2)$, $(3, 3)$ і $(4, 4)$ (позадіагональні елементи дорівнюють нулю). Для експериментів генерувалися вибірки обсягом 300 елементів і обчислена середня міра близькості. Як бачимо, міра близькості монотонно спадає при збільшенні відстані між центрами розподілів при фіксованій дисперсії, а також при збільшенні дисперсії, якщо центр є фіксованим.

Таблиця. Усереднена міра близькості.

Центри	Міра близькості	Дисперсія	Міра близькості
(0,0)—(0,0)	0,922	(1,1)—(1,1)	0,928
(0,0)—(1,1)	0,395	(1,1)—(2,2)	0,428
(0,0)—(2,2)	0,120	(1,1)—(3,3)	0,289
(0,0)—(3,3)	0,063	(1,1)—(4,4)	0,223

Розділ 11.

Базові положення теорії варіаційних нерівностей

Варіаційні нерівності — один з центральних об'єктів прикладного нелінійного аналізу. Вони є зручною загальною формою запису та дослідження різних нелінійних задач. Зокрема, у вигляді задачі розв'язання варіаційної нерівності можуть бути сформульовані задачі розв'язання рівнянь, знаходження екстремуму функціоналів, знаходження точок рівноваги Неша в грі тощо. Цей та наступний розділи містять базові положення теорії варіаційних нерівностей і опис основних проекційних методів їх розв'язання. Для детальнішого ознайомлення з варіаційними нерівностями та відповідними алгоритмами пропонуємо роботи [2, 3, 9, 13, 18, 25, 27].

11.1. Проекція на опуклу множину

Нехай H — дійсний гільбертовий простір зі скалярним добутком $(\cdot,\cdot)$ та нормою $\|\cdot\|$. Сильну збіжність позначатимемо через $\to$, а слабку — $\rightharpoonup$.

Теорема 10. *Нехай $C \subseteq H$ — опукла замкнена множина. Тоді для будь-якого елемента x простору H в C існує єдиний*

найближчий до x елемент. Іншими словами, існує єдиний елемент $z \in C$, для якого

$$\|z - x\| = \min_{y \in C} \|y - x\| .$$

Перше доведення. Нехай $z_k \in C$ – мінімізуюча послідовність, тобто

$$\lim_{k \to \infty} \|z_k - x\| = d = \inf_{y \in C} \|y - x\| . \tag{11.1}$$

За правилом паралелограма (елементарний наслідок властивостей скалярного добутку)

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \ , \ \ x, y \in H,$$

можемо записати

$$\|z_k - z_l\|^2 = 2\|x - z_k\|^2 + 2\|x - z_l\|^2 - \\ - 4\left\|x - \frac{z_k + z_l}{2}\right\|^2 . \tag{11.2}$$

З опуклості C випливає $\frac{1}{2}z_k + \frac{1}{2}z_l \in C$, тому

$$d^2 \le \left\|x - \frac{z_k + z_l}{2}\right\|^2 .$$

Отже,

$$\|z_k - z_l\|^2 \le 2\|x - z_k\|^2 + 2\|x - z_l\|^2 - 4d^2.$$

З (11.1) випливає, що $\lim_{k,l \to \infty} \|z_k - z_l\| = 0$.

Оскільки простір H повний, множина C замкнена, то існує елемент $z \in C$ такий, що $z_k \to z$. Крім того,

$$\|z - x\| = \lim_{k \to \infty} \|z_k - x\| = d.$$

Для доведення єдиності найближчого елемента слід тільки

зазначити, що після підстановки в (11.2) замість z_k і z_l довільних двох елементів $z, z' \in C$, які задовольняють $\|z - x\| = \|z' - x\| = d$, одержимо

$$\left\|z - z'\right\|^2 = 2\left\|x - z\right\|^2 + 2\left\|x - z'\right\|^2 - \\ - 4\left\|x - \frac{z + z'}{2}\right\|^2 \leq 4d^2 - 4d^2 = 0,$$

звідки випливає, що $z = z'$. □

Друге доведення. Оскільки $d(x, C) = \inf_{y \in C} \|y - x\| = d(0, C - x)$, то теорему достатньо довести для випадку $x = 0$. Позначимо $d(0, C)$ через d і розглянемо множини

$$C_n = \left\{ y \in C : \ \|y\| \leq d + \frac{1}{n} \right\}.$$

Перетин усіх C_n — це множина елементів, які розташовані на відстані d від нуля. Отже, потрібно довести, що $\bigcap_{n=1}^{\infty} C_n$ складається з однієї точки. Скористаємось принципом вкладених множин. Множини C_n утворюють спадний ланцюг опуклих замкнених множин. Потрібно показати, що $\mathrm{diam} C_n \to 0$ $(n \to \infty)$. Для $x, y \in C_n$ маємо:

$$\|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 - \|x + y\|^2 \leq$$

$$\leq 2\left(\left(d + \frac{1}{n}\right)^2 + \left(d + \frac{1}{n}\right)^2 - 2d^2\right) = \frac{8d}{n} + \frac{4}{n^2}.$$

Отже, $\mathrm{diam} C_n \leq \sqrt{\frac{8d}{n} + \frac{4}{n^2}} \to 0$ $(n \to \infty)$. □

Теорема 10 дозволяє коректно ввести оператор P_C метричного проектування (проекції) простору H на опуклу замкнену множину $C \subseteq H$, який ставить у відповідність елементу

$x \in H$ єдиний елемент $P_C x \in C$, для якого

$$\|P_C x - x\| = \min_{y \in C} \|y - x\| . \tag{11.3}$$

Оператор P_C можна охарактеризувати таким чином.

Теорема 11. *Нехай $C \subseteq H$ — опукла замкнена множина, $x \in H$, $z \in C$. Такі умови рівносильні:*

(i) $z = P_C x$.

(ii) $(z - x, y - z) \geq 0 \quad \forall y \in C$.

(iii) $\|z - y\|^2 \leq \|x - y\|^2 - \|z - x\|^2 \quad \forall y \in C$.

Доведення. Рівносильність (ii) та (iii) випливає з тотожності

$$2(z - y, z - x) = \|z - y\|^2 + \|z - x\|^2 - \|x - y\|^2 .$$

Доведемо рівносильність (i) та (ii). Нехай $x \in H$ і $z = P_C x$. Оскільки множина C опукла, то

$$(1 - t)z + ty = z + t(y - z) \in C \quad \forall z \in C \;\; 0 \leq t \leq 1.$$

Унаслідок (11.3) функція

$$\begin{aligned} \phi(t) = \|x - z - t(y - z)\|^2 = \\ = \|x - z\|^2 - 2t(x - z, y - z) + t^2 \|z - y\|^2 \end{aligned}$$

досягає мінімуму в точці $t = 0$. Це означає, що $\frac{d}{dt}\phi(0) \geq 0$, тобто $(x - z, y - z) \leq 0$ для всіх $y \in C$, або $(z - x, y - z) \geq 0$ для всіх $y \in C$. З іншого боку, якщо $z \in C$ і $(z - x, y - z) \geq 0$ при $y \in C$, то

$$0 \leq (z - x, y - x + x - z) = - \|x - z\|^2 + (z - x, y - x).$$

Отже, $\|x - z\|^2 \leq (z - x, y - x) \leq \|z - x\| \, \|y - x\|$ і тим самим

$$\|x - z\| \leq \|y - x\| \quad \forall y \in C,$$

тобто $z = P_C x$. $\square$

З доведеної теореми випливає, що оператор метричного проектування P_C нерозтягуючий, тобто

$$\|P_C x - P_C y\| \leq \|x - y\|, \quad \forall x, y \in H.$$

Для деяких опуклих замкнених множин C відомі явні формули обчислення проекції P_C. Наприклад, для кулі $B(x_0, R) = \{y \in H : \|y - x_0\| \leq R\}$ та $x \notin B(x_0, R)$ маємо

$$P_{B(x_0,R)} x = x_0 + R\frac{x - x_0}{\|x - x_0\|},$$

а для гіперплощини $L = \{y \in H : (x_0, y) = c\}$ $(x_0 \neq 0,\ c \in \mathbb{R})$ проекція $P_L x$ обчислюється за формулою

$$P_L x = x + (c - (x_0, x))\frac{x_0}{\|x_0\|^2}.$$

Доведемо дві асимптотичні властивості проекції.

Твердження 1. *Нехай $C_1 \subseteq C_2 \subseteq \ldots$ — неспадна послідовність непорожніх опуклих замкнених множин. Покладемо $C = \overline{\bigcup_{n=1}^{\infty} C_n}$ і нехай $x \in H$. Тоді $P_{C_n} x \to P_C x$.*

Доведення. Множина C — непорожня, опукла та замкнена. Розглянемо таку послідовність точок $y_n \in C_n$, що $y_n \to P_C x$ $(n \to \infty)$. Маємо

$$\|x - P_C x\| \leq \|x - P_{C_n} x\| \leq \|x - y_n\| \quad \forall n \in \mathbb{N}.$$

Таким чином,

$$\|x - P_{C_n} x\| \to \|x - P_C x\|. \tag{11.4}$$

Покажемо, що $P_{C_n} x \rightharpoonup P_C x$, тоді з (11.4) випливатиме $P_{C_n} x \to P_C x$. Нехай z — часткова слабка границя послідовності $(P_{C_n} x)$.

Тоді $z \in C$ та

$$\|P_C x - x\| \le \|z - x\| \le \lim_{n\to\infty} \|P_{C_n} x - x\|.$$

Отже, $\|z - x\| = \|P_C x - x\|$. Це означає, що $z = P_C x$ — єдина часткова слабка границя обмеженої послідовності $(P_{C_n} x)$. Тому $P_{C_n} x \rightharpoonup P_C x$. □

Твердження 2. *Нехай $C_1 \supseteq C_2 \supseteq \ldots$ — незростаюча послідовність опуклих замкнених множин. Припустимо, що $C = \bigcap_{n=1}^{\infty} C_n \neq \emptyset$, і нехай $x \in H$. Тоді $P_{C_n} x \to P_C x$.*

Доведення. Покладемо $y_n = P_{C_n} x$. Тоді для всіх $n \in \mathbb{N}$

$$\|x - y_n\| \le \|x - y_{n+1}\| \le \|x - P_C x\|.$$

Таким чином, послідовність (y_n) обмежена та існує границя $\lim_{n\to\infty} \|x - y_n\|$. Доведемо фундаментальність послідовності (y_n). Для $k \ge n$ маємо (при $k, n \to \infty$):

$$\|y_k - y_n\|^2 = 2\left(\|y_k - x\|^2 + \|y_n - x\|^2\right) - \\ - 4\left\|\frac{y_k + y_n}{2} - x\right\|^2 \le 2\left(\|y_k - x\|^2 - \|y_n - x\|^2\right) \to 0.$$

Використали включення $\frac{y_k + y_n}{2} \in C_n$. Послідовність (y_n) фундаментальна. Тому $y_n \to y \in H$. Оскільки $y_k \in C_n$ для $k \ge n$ і множина C_n замкнена, то $y \in C_n$. Отже, $y \in C$ та

$$\|x - P_C x\| \le \|x - y\| = \lim_{n\to\infty} \|x - y_n\| \le \|x - P_C x\|.$$

Оскільки множина C опукла й замкнена, то $y = P_C x$. □

Зауваження 7. Якщо у твердженні 2 множина C_1 обмежена, то автоматично $\bigcap_{n=1}^{\infty} C_n \neq \emptyset$.

11.2. Пошук спільної точки опуклих множин

Розглянемо скінченний набір замкнених опуклих множин C_1, C_2,..., C_r. Для $x_0 \in H$ визначимо послідовність (x_n) за таким правилом:

$$x_0 \overset{P_{C_1}}{\mapsto} x_1 \overset{P_{C_2}}{\mapsto} x_2 \overset{P_{C_3}}{\mapsto} x_3 \overset{P_{C_4}}{\mapsto} ... \overset{P_{C_r}}{\mapsto} x_r \overset{P_{C_1}}{\mapsto} x_{r+1} \overset{P_{C_2}}{\mapsto} ..., \qquad (11.5)$$

тобто

$$x_{n+1} = P_{C_{n \bmod r+1}} x_n.$$

Має місце

Теорема 12 (Л. Брегман, 1965). *Нехай C_1,..., $C_r \subseteq H$ — замкнені опуклі множини з непорожнім перетином. Тоді*

$$x_n \rightharpoonup \bar{x} \in \cap_{i=1}^r C_i \quad при \ n \to \infty.$$

Доведення. Нехай $x \in \cap_{i=1}^r C_i$. Позначимо через (y_k^i) підпослідовність (x_{kr+i}), де $i \in \{1, ..., r\}$. З означення метричної проекції випливають нерівності

$$\|x_0 - x\| \geq \|x_1 - x\| \geq \|x_2 - x\| \geq \|x_3 - x\| \geq ... \geq 0.$$

Отже, послідовність (x_n) обмежена та існує границя $\lim_{n\to\infty} \|x_n - x\|$. Зокрема, обмеженою є послідовність (y_k^1). Тому існує така підпослідовність $(y_{k_l}^1)$, що $y_{k_l}^1 \rightharpoonup \bar{x} \in H$ $(l \to \infty)$. Оскільки $y_{k_l}^1 \in C_1$, а множина C_1 слабко замкнена, то $\bar{x} \in C_1$. Для $y_{k_l}^2 = P_{C_2} y_{k_l}^1$ маємо

$$\left\|y_{k_l}^1 - y_{k_l}^2\right\|^2 \leq \left\|y_{k_l}^1 - x\right\|^2 - \left\|y_{k_l}^2 - x\right\|^2.$$

Тому $\lim_{l\to\infty} \left\|y_{k_l}^1 - y_{k_l}^2\right\| = 0$. Звідси $y_{k_l}^2 \rightharpoonup \bar{x}$. Отримали, що $\bar{x} \in C_2$. Продовжуючи аналогічно, доходимо висновку, що має місце включення $\bar{x} \in \cap_{i=1}^r C_i$.

Залишилось показати, що вся послідовність (x_n) слабко збігається до $\bar{x}$. Доводимо від супротивного. Припустимо, що (x_n) не збігається слабко до $\bar{x}$. Тоді існує така підпослідовніть (x_{n_j}), що $x_{n_j} \rightharpoonup \tilde{x} \neq \bar{x}$ $(j \to \infty)$. Підпослідовність (x_{n_j}) містить принаймні одну підпослідовність вигляду $(y^i_{k_l})$ для деякого $i \in \{1, ..., r\}$. Міркуючи як і раніше, отримаємо $\tilde{x} \in \cap_{i=1}^r C_i$. Розглянемо числову послідовність

$$\alpha_n = \|x_n - \bar{x}\|^2 - \|x_n - \tilde{x}\|^2 = \|\bar{x}\|^2 - \|\tilde{x}\|^2 + 2(x_n, \tilde{x} - \bar{x}).$$

Послідовність (α_n) збіжна. Нехай $\alpha = \lim_{n\to\infty} \alpha_n$. З одного боку, розглянувши підпослідовність (α_{n_j}), отримаємо $\alpha = \|\bar{x} - \tilde{x}\|^2$. З іншого, розглянувши $(y^1_{k_l})$, отримаємо $\alpha = -\|\bar{x} - \tilde{x}\|^2$. Таким чином, $\alpha = 0$ та $\bar{x} = \tilde{x}$. □

Має місце

Теорема 13 (І. Гальперін, 1962). *Нехай E_1,..., $E_r \subseteq H$ — замкнені лінійні підпростори. Тоді для всіх $x \in H$*

$$(P_{E_r} ... P_{E_2} P_{E_1})^n x \to P_{\cap_{i=1}^r E_i} x \quad при \ n \to \infty.$$

Опишемо тепер метод усереднених проекцій знаходження спільної точки скінченного набору замкнених опуклих множин C_1, C_2,..., C_r. Для $x_0 \in H$ будуємо послідовність (x_n) за правилом:

$$x_{n+1} = \frac{1}{r} \sum_{i=1}^{r} P_{C_i} x_n. \tag{11.6}$$

Обчислення проекцій у (11.6) можна організувати паралельно.

Має місце

Теорема 14. *Нехай C_1, C_2,..., $C_r \subseteq H$ — замкнені опуклі множини з непорожнім перетином, (x_n) — послідовність, побудована методом усереднених проекцій* (11.6). *Тоді*

$$x_n \rightharpoonup \bar{x} \in \cap_{i=1}^r C_i \quad при \ n \to \infty.$$

Доведення. Розглянемо гільбертовий простір

$$\mathcal{H} = \underbrace{H \times H \times ... \times H}_{r}$$

зі скалярним добутком $\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{r} (x_i, y_i)$. Задамо дві множини:

$$\mathcal{C} = \{\vec{x} = (x_1, x_2, ..., x_r) \in \mathcal{H} : \ x_i \in C_i\},$$
$$\mathcal{D} = \{(x, x, ..., x) \in \mathcal{H} : \ x \in H\}.$$

Множина $\mathcal{C}$ — опукла та замкнена, а $\mathcal{D}$ — замкнений лінійний підпростір. Більш того, $\mathcal{C} \cap \mathcal{D} \neq \emptyset$ тоді й тільки тоді, коли $\cap_{i=1}^{r} C_i \neq \emptyset$. Неважко переконатись у правильності таких формул:

$$P_{\mathcal{C}}\vec{x} = (P_{C_1}x_1, P_{C_2}x_2, ..., P_{C_r}x_r),$$
$$P_{\mathcal{D}}\vec{x} = (x, x, ..., x), \ \text{де} \ x = \frac{\sum_{i=1}^{r} x_i}{r}.$$

Для $\vec{x} = (x, x, ..., x) \in \mathcal{D}$ розглянемо $P_{\mathcal{D}}P_{\mathcal{C}}\vec{x}$. Маємо

$$P_{\mathcal{D}}P_{\mathcal{C}}\vec{x} = \left(\frac{1}{r}\sum_{i=1}^{r} P_{C_i}x, \frac{1}{r}\sum_{i=1}^{r} P_{C_i}x, ..., \frac{1}{r}\sum_{i=1}^{r} P_{C_i}x \right).$$

Застосувавши теорему 12 для процеса

$$\vec{x}_{n+1} = P_{\mathcal{D}}P_{\mathcal{C}}\vec{x}_n, \ \ \vec{x}_0 = (x_0, x_0, ..., x_0) \in \mathcal{H},$$

отримаємо потрібне. □

Зауваження 8. Якщо, додатково, C_1, C_2,..., $C_r \subseteq H$ — замкнені лінійні підпростори, то

$$x_n \to P_{\cap_{i=1}^{r} C_i} x_0 \ \ \text{при} \ \ n \to \infty,$$

де (x_n) — послідовність, побудована методом усереднених проекцій (11.6).

11.3. Нерухомі точки

11.3.1. Аналітичне доведення теореми Брауера

Будемо використовувати позначення:

- $B^n = \{x \in \mathbb{R}^n : \ \|x\| \leq 1\}$ — одинична куля;
- $S^{n-1} = \{x \in \mathbb{R}^n : \ \|x\| = 1\}$ — одинична сфера.

Нехай $f : X \to X$ — деяке відображення. Точку $x \in X$ називають нерухомою точкою відображення f, якщо $f(x) = x$.

Теорема 15 (Брауер). *Довільне неперервне відображення $f : B^n \to B^n$ має нерухому точку.*

Наслідок 1. *Нехай $K \subseteq \mathbb{R}^n$ — компактна опукла множина та $f : K \to K$ — неперервне відображення. Тоді f має нерухому точку.*

Теорему Брауера виведемо з теореми про відсутність ретракції кулі B^n на сферу S^{n-1}.

Означення 33. Нехай $A \subseteq X$. Неперервне відображення $r : X \to A$ називають ***ретракцією*** X на A, якщо $r(x) = x$ для всіх $x \in A$.

Теорема 16. *Не існує ретракції кулі B^n на сферу S^{n-1}.*

Доведемо теорему 16 для випадку гладкого відображення. Перейти до неперервних відображень можна за допомогою апроксимації неперервних відображень гладкими. Дійсно, припустимо, що існує неперервна ретракція $r : B^n \to S^{n-1}$. Покажемо, що тоді існує гладка ретракція $r_0 : B^n \to S^{n-1}$. Якщо $\|x\| = 1$, то $r(x) = x$. Тому для довільного $\varepsilon_1 > 0$ існує таке $\delta > 0$, що $\|r(x) - x\| \leq \varepsilon_1$ при $1 - \delta \leq \|x\| \leq 1$. За теоремою Стоуна—Вейєрштрасса існують таке гладке відображення $f : \mathbb{R}^n \to \mathbb{R}^n$, що $\|f(x) - (r(x) - x)\| \leq \varepsilon_1$ при $\|x\| \leq 1$, і така гладка функція ψ, що $0 \leq \psi(t) \leq 1$ при $0 \leq t \leq 1$, $\psi(1) = 0$ і

$1-\varepsilon_2 \le \psi(t)$ при $t^2 \le 1-\delta$. Покладемо $g(x) = x + \phi(x)f(x)$, де $\phi(x) = \psi(\|x\|^2)$. Якщо $\|x\| \le 1-\delta$, то

$$\begin{aligned}\|g(x)\| &= \|x + \phi(x)f(x)\| = \\ &= \|r(x) + \phi(x)(f(x) - r(x) + x) + (\phi(x) - 1)(r(x) - x)\| \ge \\ &\ge \|r(x)\| - \phi(x)\|f(x) - r(x) + x\| - (1 - \phi(x))\|r(x) - x\| \ge \\ &\ge 1 - 1 \cdot \varepsilon_1 - \varepsilon_2 \cdot 2 = 1 - \varepsilon_1 - 2\varepsilon_2.\end{aligned}$$

Якщо $1-\delta \le \|x\| \le 1$, то

$$\|g(x)\| = \|x + \phi(x)f(x)\| =$$

$$\begin{aligned}&= \|x + \phi(x)(f(x) - r(x) + x) + \phi(x)(r(x) - x)\| \ge \\ &\ge \|x\| - \phi(x)\|f(x) - r(x) + x\| - \phi(x)\|r(x) - x\| \ge \\ &\ge 1 - \delta - 1 \cdot \varepsilon_1 - 1 \cdot \varepsilon_1 = 1 - \delta - 2\varepsilon_1.\end{aligned}$$

При $\varepsilon_1 \to 0$ маємо $\delta \to 0$. Тому можна вважати, що ε_1, ε_2, $\delta \le \frac{1}{4}$. У цьому випадку $\|g(x)\| \ge \frac{1}{4} > 0$ для всіх $x \in B^n$. Якщо $\|x\| = 1$, то $\phi(x) = 0$ і $g(x) = x$. Потрібна гладка ретракція $r_0 : B^n \to S^{n-1}$ задається формулою $r_0(x) = \frac{g(x)}{\|g(x)\|}$.

Доведення теореми 16. Припустимо, що існує неперервно диференційовна ретракція $r : B^n \to S^{n-1}$. Для $x \in B^n$ і $0 \le t \le 1$ покладемо

$$\begin{aligned}&g(x) = r(x) - x, \\ &r_t(x) = x + tg(x) = (1-t)x + tr(x).\end{aligned}$$

З неперервної диференційовності відображення g випливає існування такої сталої $L > 0$, що

$$\|g(x) - g(y)\| \le L\|x - y\| \quad \forall x, y \in B^n.$$

Відображення $r_t : B^n \to B^n$ ін'єктивне при $0 \le t < \frac{1}{L}$. Дійсно,

якщо $x \neq y$, то

$$\|r_t(x) - r_t(y)\| \geq \|x - y\| - t\|g(x) - g(y)\| \geq \\ \geq (1 - tL)\|x - y\| > 0.$$

Частинні похідні відображення g рівномірно обмежені, тому матриця Якобі

$$r_t'(x) = I_n + t \cdot g'(x) \tag{11.7}$$

за малих t оборотна. Отже, за теоремою про обернене відображення r_t при $t \in [0, t_0]$ відображає $\mathrm{int}B^n$ на деяку відкриту множину $G_t \subseteq \mathrm{int}B^n$. Нехай $e \in B^n \backslash G_t$. З'єднаємо відрізком точку e з довільною точкою множини G_t і розглянемо точку b, у якій цей відрізок перетинає межу множини G_t. Множина B^n компактна, тому $b = r_t(x)$ для деякої точки $x \in B^n$. Оскільки $b \notin G_t = r_t(\mathrm{int}B^n)$, то $x \notin \mathrm{int}B^n$, тобто $x \in S^{n-1}$. Тому $b = x$ і $e = b = x \in S^{n-1}$. Таким чином, r_t сюр'єктивно відображає $\mathrm{int}B^n$ на $\mathrm{int}B^n$. Крім того, r_t бієктивно відображає S^{n-1} на S^{n-1} (на сфері S^{n-1} маємо $r_t(x) = x$) і, як було показано раніше, r_t ін'єктивно відображає B^n у B^n. Тому при $t \in [0, t_0]$ відображення r_t — бієкція B^n на B^n.

Розглянемо інтеграл

$$V(t) = \int_{B^n} \det\left(r_t'(x)\right) dx = \int_{B^n} \det\left(I_n + t \cdot g'(x)\right) dx.$$

При $0 \leq t \leq t_0$ цей інтеграл дорівнює об'єму кулі B^n. Формула (11.7) показує, що $V(t)$ — багаточлен від t. Тому $V(t)$ — додатна стала при $0 \leq t \leq 1$, зокрема $V(1) = \int_{B^n} \det\left(r'(x)\right) dx > 0$.

З іншого боку, $r(x) \in S^{n-1}$ для всіх $x \in B^n$. Оскільки $(r(x), r(x)) = \|r(x)\|^2 = 1$ для всіх $x \in B^n$, то

$$0 = \frac{d}{dt}\left(r(x + th), r(x + th)\right)|_{t=0} = 2(r(x), r'(x)h) \ \forall h \in \mathbb{R}^n.$$

Отже, $R(r'(x)) \subseteq \{r(x)\}^{\perp}$ і тому $\det\left(r'(x)\right) = 0$. Проте в цьому випадку $V(1) = 0$. Отримали протиріччя. $\square$

Доведення теореми Брауера. Припустимо, що існує неперервне відображення $f : B^n \to B^n$ без нерухомих точок. Побудуємо за допомогою f ретракцію r кулі B^n на сферу S^{n-1}.

Для кожної точки $x \in B^n$ розглянемо промінь із початком $f(x) \neq x$, що проходить через x. Нехай $r(x)$ — точка, у якій цей промінь перетинає сферу S^{n-1} (зробіть рисунок!). Ясно, що побудоване відображення r — ретракція B^n на S^{n-1}. $\square$

Теорема Брауера та теорема 16 еквівалентні. Дійсно, нехай існує ретракція $r : B^n \to S^{n-1}$. Тоді відображення $-r : B^n \to S^{n-1} \subseteq B^n$ не має нерухомих точок, що суперечить теоремі Брауера.

Доведемо один важливий результат, який за традицією називають лемою Кнастера – Куратовського – Мазуркевича.

Лема 1 (Кнастера – Куратовського – Мазуркевича, ККМ)**.** *Нехай X — довільна множина в $\mathbb{R}^n$. Кожній точці $x \in X$ поставлено у відповідність компактну множину $F(x) \subseteq \mathbb{R}^n$, так що для довільної скінченної множини $\{x_1, x_2, ..., x_p\} \subseteq X$*

$$\operatorname{conv}\{x_1, x_2, ..., x_p\} \subseteq \bigcup_{i=1}^{p} F(x_i).$$

Тоді

$$\bigcap_{x \in X} F(x) \neq \emptyset.$$

Доведення. Достатньо довести, що сім'я множин $\{F(x)\}_{x \in X}$ центрована. Міркуємо від супротивного. Припустимо, що існує така множина $\{x_1, x_2, ..., x_p\} \subseteq X$, що

$$\bigcap_{i=1}^{p} F(x_i) = \emptyset.$$

Тоді

$$\bigcup_{i=1}^{p} O_i = \mathbb{R}^n,$$

де $O_i = \mathbb{R}^n \backslash F(x_i)$. Нехай $\{\phi_i\}$ — неперервне розбиття одиниці в $\mathbb{R}^n$, узгоджене з відкритим покриттям $\{O_i\}$, тобто

$$\phi_i \in C(\mathbb{R}^n),\ \ 0 \le \phi_i \le 1,\ \ \sum_{i=1}^{p} \phi_i = 1,\ \ \mathrm{supp}\phi_i \subseteq O_i.$$

Розглянемо відображення

$$\phi(x) = \sum_{i=1}^{p} \phi_i(x) x_i.$$

Ясно, що $\phi(K) \subseteq K$, де $K = \mathrm{conv}\,\{x_1, x_2, ..., x_p\}$. За теоремою Брауера про нерухому точку існує така точка $y \in K$, що

$$y = \sum_{i=1}^{p} \phi_i(y) x_i.$$

Можна вважати, що для деякого $k \le p$ маємо

$$\phi_i(y) \begin{cases} > 0, & \text{якщо } i \le k, \\ = 0, & \text{якщо } i > k, \end{cases}$$

тобто $y \in \mathrm{conv}\,\{x_1, x_2, ..., x_k\}$. Тоді, за умовою леми, $y \in \bigcup_{i=1}^{k} F(x_i)$, а отже, для деякого $i \le k$ маємо $y \in F(x_i)$. Звідси випливає, що $y \notin O_i$, тобто $\phi_i(y) = 0$. Дійшли протиріччя. $\square$

Зауваження 9. У лемі 1 умову компактності всіх множин $F(x)$ можна послабити до їх замкненості та існування принаймні однієї компактної множини $F(x_0)$, $x_0 \in X$.

11.3.2. Опуклість чебишовських множин

Відстанню від заданої точки x до заданої множини C називають величину

$$d(x, C) = \inf_{y \in C} \|x - y\|.$$

Під ***елементом найкращого наближення*** або найближчою точкою для заданої точки x будемо розуміти таку точку $y_0 \in C$, для якої $\|x-y_0\| = d(x,C)$, тобто $\|x-y_0\| \le \|x-y\|$ для всіх $y \in C$. Множину всіх найближчих точок з C для заданої точки x позначимо $P_C x$.

Означення 34. Непорожню множину C називають ***чебишовською***, якщо довільна точка x має точно одну найближчу в C, тобто

$$\forall x \text{ множина } P_C \text{ складається з однієї точки.}$$

Якщо C — чебишовська множина, то відображення P_C, що ставить у відповідність точці x її найближчу точку $P_C x$ із C, називається ***метричною проекцією*** на C.

Опишемо всі чебишовські множини евклідового (скінченновимірного гільбертового) простору.

Теорема 17 (Л. Бунт, Т. Моцкін)**.** *Чебишовська множина евклідового простору є замкненою та опуклою.*

Ключову роль у розглянутому нижче доведенні В. Клі – В. І. Бердишева грає теорема Брауера про нерухому точку.

Лема 2. *Метрична проекція P_C на чебишовську множину C є неперервним відображенням.*

Доведення. Спочатку зазначимо, що для $x, y \in \mathbb{R}^n$

$$|d(x,C) - d(y,C)| \le \|x-y\|. \tag{11.8}$$

Припустимо, що метрична проекція P_C розривна в деякій точці x, тобто знайдуться число $\varepsilon > 0$ і послідовність (x_n), $x_n \to x$, такі, що $\|P_C x_n - P_C x\| \ge \varepsilon$ для всіх $n \in \mathbb{N}$. Унаслідок (11.8) послідовність $(P_C x_n)$ обмежена. Нехай y — її часткова границя. Ясно, що $y \ne P_C x$. За (11.8)

$$\|x-y\| = \lim_{k\to\infty} \|x_{n_k} - P_C x_{n_k}\| = \lim_{k\to\infty} d(x_{n_k}, C) = d(x,C).$$

Оскільки множина C замкнена, то $y \in C$, тобто точка y також є найближчою до x. Дістали протиріччя з тим, що C — чебишовська множина. □

Означення 35. Чебишовську множину C називають ***чебишовським сонцем***, якщо

$$P_C(P_C x + t(x - P_C x)) = P_C x \quad \forall x \notin C \ \ \forall t \geq 0.$$

Лема 3. *Чебишовська множина є чебишовським сонцем.*

Доведення. Нехай C — чебишовська множина та $x \notin C$. Розглянемо промінь ℓ, що виходить з $y = P_C x \in C$ та проходить через x. Покладемо $r = \|x - y\|$ та розглянемо кулю $B = B(x, r)$. Задамо відображення $f : B \to B$ за формулою

$$f(z) = x + r\frac{x - P_C z}{\|x - P_C z\|}, \quad z \in B.$$

Точка $f(z)$ — точка перетину сфери $S(x, r)$ та променю, що виходить з x у напрямку $x - P_C z$.

Відображення f неперервне (випливає з неперервності метричної проекції на чебишовську множину). За теоремою Брауера існує точка $z_0 \in B$ така, що $f(z_0) = z_0$. З означення $f(z_0)$ випливає, що точка x лежить на відрізку, що з'єднує z_0 з її проекцією $P_C z_0$. При цьому найближчим елементом до кожної точки відрізка $[z_0, P_C z_0]$ буде точка $P_C z_0$ (випливає з нерівності трикутника та чебишовості C). Однак для x найближча точка $y = P_C x$, і вона єдина. Отже, $P_C z_0 = y$. Таким чином, для всіх точок із $[y, z_0]$ найближчою точкою із C є точка y.

Застосовуючи попередні міркування, проведені для точки x, до точки z_0, ми ще далі просунемось по променю ℓ. У результаті для кожної точки $p \in \ell$ точка y буде єдиною найближчою з множини C. □

Лема 4. *Чебишовське сонце є опуклою множиною.*

Доведення. Нехай x, $y \in C$ та $z = \lambda x + (1-\lambda)y$, де $\lambda \in (0,1)$. Оскільки C — чебишовське сонце, то при $z \notin C$ кожна точка $z + t(z - P_C z)$ для $t > 0$ має точку $P_C z$ своєю найближчою із C. Тому

$$\|z + t(z - P_C z) - P_C z\|^2 \leq \|z + t(z - P_C z) - x\|^2 . \qquad (11.9)$$

Використовуючи властивості скалярного добутку, отримуємо

$$\frac{1}{t}\|z - P_C z\|^2 + 2(z - P_C z, z - P_C z) \leq$$
$$\leq \frac{1}{t}\|z - x\|^2 + 2(z - P_C z, z - x).$$

Спрямовуючи $t \to +\infty$ у цій нерівності, отримуємо

$$(z - P_C z, z - x) \geq \|z - P_C z\|^2.$$

Аналогічно, підставляючи y замість x у (11.9), маємо

$$(z - P_C z, z - y) \geq \|z - P_C z\|^2.$$

Звідси

$$\|z - P_C z\|^2 \leq \lambda(z - P_C z, z - x) + (1-\lambda)(z - P_C z, z - y) = 0.$$

Отже, $z = P_C z \in C$, тобто множина C опукла. □

У нескінченновимірних лінійних просторах про опуклість чебишовських множин відомо небагато. Наприклад, на момент написання цієї книги не відомо, чи опукла довільна чебишовська множина в нескінченновимірному гільбертовому просторі.

Проблема 1 (В. Клі, 1961)**.** Довести чи спростувати, що довільна чебишовська множина в нескінченновимірному гільбертовому просторі опукла.

11.4. Варіаційні нерівності в $\mathbb{R}^n$

Нехай дано непорожню підмножину C простору $\mathbb{R}^n$ та оператор $A : C \to \mathbb{R}^n$. Розглядатимемо задачу:

$$\text{знайти } x \in C : \ (Ax, y - x) \geq 0 \quad \forall y \in C. \tag{11.10}$$

Нерівності вигляду (11.10) називають варіаційними. Чому варіаційними? Відповідь на це запитання пов'язана з тим, що у вигляді (11.10) можна записати критерій оптимальності в задачі

$$f(x) \to \min, \quad x \in C,$$

де $C \subseteq \mathbb{R}^n$— опукла, замкнена множина, $f : \mathbb{R}^n \to \mathbb{R}$ — опукла та диференційовна функція. Дійсно,

$$f(x) = \min_{y \in C} f(y) \ \Leftrightarrow \ x \in C, \ (\nabla f(x), y - x) \geq 0 \ \forall y \in C.$$

Зазначимо, що коли x — розв'язок нерівності (11.10) — належить внутрішності $\text{int}C$ множини C, то $Ax = 0$. Дійсно, якщо $x \in \text{int}C$, то точки $y - x$ утворюють окіл нуля, коли y пробігає C, тобто для довільного $z \in \mathbb{R}^n$ знайдуться $\varepsilon > 0$ та $y \in C$ такі, що $z = \varepsilon(y - x)$. Отже,

$$(Ax, z) = \varepsilon(Ax, y - x) \geq 0 \quad \forall z \in \mathbb{R}^n,$$

тому $Ax = 0$.

Нехай $C \subseteq \mathbb{R}^n$ — опукла множина. Якщо x — розв'язок нерівності (11.10) — належить межі $\text{fr}C$ множини C та $Ax \neq 0$, то елемент Ax задає опорну гіперплощину до C.

Зауваження 10. Нехай $C \subseteq \mathbb{R}^n$ — опукла множина та $x \in \text{fr}C$. Нагадаємо, що гіперплощина $L = \{y \in \mathbb{R}^n : (a, y - x) = 0\}$, $a \neq 0$, називається ***опорною*** до множини C, якщо $(a, y - x) \geq 0$ для всіх $y \in C$.

Будемо позначати $VI(A, C)$ множину розв'язків варіаційної нерівності (11.10).

Варіаційну нерівність (11.10) можна сформулювати у вигляді задачі пошуку нерухомої точки. Точка $x \in C$ є розв'язком варіаційної нерівності тоді й тільки тоді, коли

$$x = P_C(I - \lambda A)x, \tag{11.11}$$

де $\lambda > 0$, P_C — оператор метричного проектування на множину C. Дійсно, нехай $\lambda > 0$. З теореми 11 випливає

$$x = P_C(I - \lambda A)x \;\Leftrightarrow\; (x - x + \lambda Ax, y - x) \geq 0 \quad \forall y \in C.$$

Має місце

Теорема 18. *Нехай $C \subseteq \mathbb{R}^n$ — компактна опукла множина та $A : C \to \mathbb{R}^n$ — неперервний оператор. Тоді існує розв'язок варіаційної нерівності (11.10).*

Доведення. Оператор $T = P_C(I - \lambda A) : C \to C$ — неперервний, тому, за теоремою Брауера, у нього є нерухома точка $x \in C$, тобто виконується (11.11), що рівносильно (11.10). □

Якщо допустима множина C необмежена, то задача (11.10) може не мати розв'язків. Умови існування розв'язків у необмеженому випадку отримують шляхом уведення додаткових припущень про властивості задачі, наприклад, обмеженість потенційної множини розв'язків, коерцитивність, сильну монотонність тощо. Розглянемо загальну ідею виявлення таких властивостей. Для даної замкненої опуклої множини C покладемо $C_R = C \cap B_R$, де $B_R = \{y \in \mathbb{R}^n : \|y\| \leq R\}$ — замкнена куля з радіусом $R > 0$ та центром у нулі. Для неперервного оператора $A : C \to \mathbb{R}^n$ за попередньою теоремою при будь-якому $C_R \neq \emptyset$ знайдеться така точка $x_R \in C_R$, що

$$(Ax_R, y - x_R) \geq 0 \quad \forall y \in C_R. \tag{11.12}$$

Має місце

Теорема 19. *Нехай $C \subseteq \mathbb{R}^n$ — замкнена опукла множина та $A : C \to \mathbb{R}^n$ — неперервний оператор. Для існування розв'язків варіаційної нерівності (11.10) необхідно та достатньо існування такого $R > 0$, що $\|x_R\| < R$ для деякого $x_R \in VI(A, C_R)$.*

Доведення. Ясно, що якщо x — розв'язок нерівності (11.10), то x є і розв'язком (11.12) за умови $\|x\| < R$.

Припустимо тепер, що для $x_R \in VI(A, C_R)$ виконується $\|x_R\| < R$. Для кожного $y \in C$ оберемо таке число $\lambda > 0$, що $z = x_R + \lambda(y - x_R) \in C_R$. Маємо:

$$0 \leq (Ax_R, z - x_R) = \lambda(Ax_R, y - x_R) \quad \forall y \in C.$$

Отже, $x_R \in VI(A, C)$. □

З теореми 19 можна отримати достатні умови існування розв'язків. Сформулюємо лише одну з них, яка пов'язана з поняттям коерцитивності.

Означення 36. Оператор $A : C \to \mathbb{R}^n$ називаємо ***коерцитивним***, якщо для деякого $x_0 \in C$ виконано

$$\frac{(Ax, x - x_0)}{\|x\|} \to +\infty \quad \text{при} \quad \|x\| \to +\infty,\ x \in C.$$

Теорема 20. *Нехай $C \subseteq \mathbb{R}^n$ — замкнена опукла множина та $A : C \to \mathbb{R}^n$ — неперервний коерцитивний оператор. Тоді існує розв'язок варіаційної нерівності (11.10).*

Доведення. З коерцитивності A випливає, що для кожного $M > 0$ існує достатньо велике $R_M > 0$ таке, що

$$(Ax, x - x_0) \geq M\|x\| \quad \forall x \in C,\ \|x\| \geq R_M,$$

де $x_0 \in C_{R_M}$ не залежить від M та R_M.

Унаслідок теореми 18 існує точка $x_{R_M} \in C_{R_M}$ така, що

$$(Ax_{R_M}, y - x_{R_M}) \geq 0 \quad \forall y \in C_{R_M}.$$

Якщо $\|x_{R_M}\| < R_M$, то за теоремою 19 точка x_{R_M} є розв'язком варіаційної нерівності (11.10).

Якщо $\|x_{R_M}\| = R_M$, то отримуємо

$$(Ax_{R_M}, x_0 - x_{R_M}) \leq -M \|x_{R_M}\| = -CR_M < 0,$$

що суперечить означенню x_{R_M}. □

З попередньої теореми випливає

Наслідок 2. *Нехай оператор $A : \mathbb{R}^n \to \mathbb{R}^n$ неперервний та коерцитивний у сенсі*

$$\frac{(Ax, x)}{\|x\|} \to +\infty \quad \textit{при} \quad \|x\| \to +\infty.$$

Тоді існує елемент $x \in \mathbb{R}^n$, для якого $Ax = 0$.

Уведемо важливе поняття.

Означення 37. Оператор $A : C \to \mathbb{R}^n$ називаємо ***монотонним***, якщо

$$(Ax - Ay, x - y) \geq 0 \quad \forall x, y \in C.$$

Похідна опуклої диференційовної функції $f : \mathbb{R}^n \to \mathbb{R}$ є монотонним оператором.

Коли оператор A монотонний, то множина розв'язків варіаційної нерівності (11.10) має одну важливу властивість.

Твердження 3. *Нехай $C \subseteq \mathbb{R}^n$ — замкнена опукла множина та $A : C \to \mathbb{R}^n$ — неперервний монотонний оператор. Тоді множина розв'язків варіаційної нерівності (11.10) є замкненою та опуклою.*

Доведення. Замкненість та опуклість множини $VI(A, C)$ випливає з такої рівносильності:

$$x \in VI(A, C) \;\Leftrightarrow\; x \in C : (Ay, y - x) \geq 0 \; \forall y \in C. \qquad (11.13)$$

Доведемо (11.13). Нехай спочатку $x \in VI(A,C)$. Унаслідок монотонності A,

$$(Ay, y-x) \geq (Ax, y-x) \geq 0 \quad \forall y \in C.$$

Тепер припустимо, що $x \in C$ та $(Ay, y-x) \geq 0$ для всіх $y \in C$. Для довільної точки $z \in C$ маємо $y = x + \lambda(z-x) \in C$, де $\lambda \in [0,1]$, через опуклість C. Тому при $\lambda \in (0,1)$ маємо

$$(A(x+\lambda(z-x)), \lambda(z-x)) \geq 0,$$

або

$$(A(x+\lambda(z-x)), z-x) \geq 0 \quad \forall z \in C.$$

Оскільки оператор A неперервний, то при $\lambda \to 0$ отримуємо, що $(Ax, z-x) \geq 0$ для всіх $z \in C$. Отже, $x \in VI(A,C)$. $\square$

Узагалі варіаційні нерівності можуть мати багато розв'язків. Строга монотонність оператора A гарантує єдиність можливого розв'язку (11.10).

Означення 38. Оператор $A : C \to \mathbb{R}^n$ називаємо ***строго монотонним***, якщо

$$(Ax - Ay, x-y) > 0 \quad \forall x, y \in C,\ x \neq y.$$

Твердження 4. *Якщо оператор A строго монотонний, то варіаційна нерівність (11.10) має не більше одного розв'язку.*

Доведення. Якщо x_1, x_2 — два розв'язки (11.10), то

$$(Ax_1, y-x_1) \geq 0 \quad \forall y \in C,$$
$$(Ax_2, y-x_2) \geq 0 \quad \forall y \in C.$$

Поклавши $y = x_2$ (відповідно $y = x_1$) у першій (відповідно у другій) нерівності та склавши, отримаємо

$$(Ax_1 - Ax_2, x_1 - x_2) \leq 0,$$

звідки $x_1 = x_2$. $\square$

11.5. Варіаційні нерівності в гільбертовому просторі

У попередньому підрозділі ми вже зустрічали монотонні оператори у скінченновимірному просторі. Там строга монотонність була використана для доведення єдиності розв'язку варіаційної нерівності, а монотонність — опуклості множини розв'язків. У даному підрозділі ми розглянемо варіаційні нерівності в нескінченновимірному гільбертовому просторі. Властивість монотонності знайде важливе застосування в питанні існування розв'язків.

Нехай H — гільбертовий простір, $C \subseteq H$ — замкнена опукла множина.

Означення 39. Оператор $A : C \to H$ називаємо ***монотонним***, якщо

$$(Ax - Ay, x - y) \geq 0 \quad \forall x, y \in C.$$

Монотонний оператор A називаємо ***строго монотонним***, якщо

$$(Ax - Ay, x - y) = 0 \Rightarrow x = y.$$

Означення 40. Оператор $A : C \to H$ називаємо ***неперервним на скінченновимірних підпросторах***, якщо для кожного скінченновимірного підпростору $E \subseteq H$ оператор $A : C \cap E \to H$ слабко неперервний.

Теорема 21. *Нехай $C \subseteq H$ — непорожня замкнена опукла обмежена множина та $A : C \to H$ — монотонний та неперервний на скінченновимірних підпросторах оператор. Тоді існує елемент*

$$x \in C : \ (Ax, y - x) \geq 0 \quad \forall y \in C. \tag{11.14}$$

Доведення. Спочатку доведемо лему Мінті.

Лема 5 (Мінті, 1962). *Нехай $C \subseteq H$ — непорожня замкнена опукла обмежена множина та $A : C \to H$ — монотонний і неперервний на скінченновимірних підпросторах оператор. Тоді нерівність (11.14) рівносильна нерівності*

$$x \in C : \ (Ay, y - x) \geq 0 \quad \forall y \in C. \tag{11.15}$$

Доведення. Нехай спочатку $x \in C$ — розв'язок (11.14). Унаслідок монотонності оператора A

$$(Ay, y - x) \geq (Ax, y - x) \geq 0 \quad \forall y \in C.$$

Тепер припустимо, що $x \in C$ та $(Ay, y - x) \geq 0$ для всіх $y \in C$. Для довільної точки $z \in C$ маємо $y = x + \lambda(z - x) \in C$, де $\lambda \in [0, 1]$, через опуклість C. Тому при $\lambda \in (0, 1)$ маємо

$$(A(x + \lambda(z - x)), z - x) \geq 0 \quad \forall z \in C.$$

Оскільки оператор A слабко неперервний на перетині C з лінійною оболонкою множини $\{x, z\}$, то при $\lambda \to 0$ отримуємо, що $(Ax, z - x) \geq 0$ для всіх $z \in C$. □

Перейдемо до доведення теореми 21. Без обмеження загальності можна вважати, що $0 \in C$. Нехай $E \subseteq H$ — скінченновимірний підпростір, $\pi : E \to H$ — оператор вкладення та $\pi^* : H \to E$ — оператор, спряжений до π. Позначимо $C_E = C \cap E = C \cap \pi E$ та розглянемо оператор $\pi^* A \pi : C_E \to E$. Оскільки C_E — компактна опукла підмножина E та $\pi^* A \pi$ — неперервний оператор, то існує такий елемент $x_E \in C_E$, що $(\pi^* A \pi x_E, y - x_E) \geq 0$ для всіх $y \in C_E$, або

$$(Ax_E, y - x_E) \geq 0 \quad \forall y \in C_E.$$

За лемою Мінті

$$(Ay, y - x_E) \geq 0 \quad \forall y \in C_E. \tag{11.16}$$

Покладемо $S(y) = \{x \in C : (Ay, y - x) \geq 0\}$. Ясно, що для кожного $y \in C$ множина $S(y)$ слабко компактна. Отже, множина $\cap_{y \in C} S(y)$ слабко компактна. Покажемо, що вона непорожня. Для цього достатньо довести центрованість системи $\{S(y)\}_{y \in C}$, тобто показати, що

$$S(y_1) \cap S(y_2) \cap \ldots \cap S(y_m) \neq \emptyset \tag{11.17}$$

для довільного скінченного набору $\{y_1, \ldots, y_m\} \subseteq C$.

Позначимо через E лінійну оболонку векторів $\{y_1, \ldots, y_m\}$, і нехай, як і раніше, $C_E = C \cap E$. За наведеними при доведенні (11.16) міркуваннями існує такий елемент $x_E \in C_E$, що $(Ay, y - x_E) \geq 0$ для всіх $y \in C_E$. Зокрема, $(Ay_i, y_i - x_E) \geq 0$, $i = 1, 2, \ldots, m$, і отже, $x_E \in S(y_i)$, $i = 1, 2, \ldots, m$. Таким чином, (11.17) виконується для довільного скінченного набору і тим самим існує елемент $x \in \cap_{y \in C} S(y)$. Ясно, що $(Ay, y - x) \geq 0$ для всіх $y \in C$, тому знову за лемою Мінті $(Ax, y - x) \geq 0$ для всіх $y \in C$. □

Наслідок 3. *В умовах леми 5 множина $VI(A, C)$ є опуклою та замкненою.*

Означення 41. Оператор $A : C \to H$ називаємо ***коерцитивним***, якщо для деякого $x_0 \in C$ виконано

$$\frac{(Ax, x - x_0)}{\|x\|} \to +\infty \quad \text{при} \quad \|x\| \to +\infty, \ x \in C.$$

Означення 42. Оператор $A : C \to H$ називають ***сильно монотонним***, якщо існує додатна константа m така, що

$$(Ax - Ay, x - y) \geq m \|x - y\|^2 \quad \forall x, y \in C.$$

Сильно монотонний оператор є коерцитивним.

Теорема 22. *Нехай $C \subseteq H$ — непорожня замкнена опукла множина та $A : C \to H$ — монотонний, неперервний на*

скінченновимірних підпросторах та коерцитивний оператор. Тоді існує елемент $x \in C$, для якого виконується (11.14).

Доведення. Нехай $r > 0$ та $C_r = C \cap B_r$, де $B_r = \{y \in H : \|y\| \le r\}$. Ясно, що C_r — обмежена, замкнена, опукла та при достатньо великих r ще й непорожня множина. З доведеного вище випливає існування такого $x_r \in C_r$, що

$$(Ax_r, y - x_r) \ge 0 \ \ \forall y \in C_r. \tag{11.18}$$

При $r \ge \|x_0\|$ покладемо в (11.18) $y = x_0$. Одержимо

$$\frac{(Ax_r, x_r - x_0)}{\|x_r\|} \le 0.$$

З умови коерцитивності A випливає обмеженість множини $\{x_r\}$. Виділимо з $\{x_r\}$ послідовність елементів x_{r_k}, $r_k \to +\infty$, що збіжна в H до деякого елемента x. Ясно, що $x \in C$ — розв'язок (11.14). Дійсно, для довільного $y \in C$ унаслідок леми Мінті для всіх достатньо великих номерів k маємо $(Ay, y - x_{r_k}) \ge 0$. Після граничного переходу отримаємо $(Ay, y - x) \ge 0$, а з леми Мінті випливає $x \in VI(A, C)$. □

Наслідок 4. *Нехай $A : H \to H$ — монотонний, неперервний на скінченновимірних підпросторах та коерцитивний оператор. Тоді для довільного $f \in H$ існує елемент $x \in H$, для якого виконується $Ax = f$.*

Доведення. Оскільки для довільного $f \in H$ оператор $Ax - f$ теж монотонний, неперервний на скінченновимірних підпросторах та коерцитивний, то достатньо переконатись у розв'язності рівняння $Ax = 0$. У цьому переконуємось за допомогою теореми 22. □

Доведемо сильніший варіант цього наслідку.

Наслідок 5. *Нехай оператор $A : H \to H$ монотонний, неперервний на скінченновимірних підпросторах та*

$$\|Ax\| \to +\infty \quad \textit{рівномірно при} \quad \|x\| \to +\infty. \tag{11.19}$$

Тоді для довільного $f \in H$ існує елемент $x \in H$, для якого виконується $Ax = f$.

Доведення. При $\varepsilon > 0$ оператор $A_\varepsilon = A + \varepsilon I$ монотонний та

$$\frac{(A_\varepsilon x, x)}{\|x\|} = \varepsilon \|x\| + \frac{(Ax, x)}{\|x\|} \geq \varepsilon \|x\| + \frac{(A(0), x)}{\|x\|}. \tag{11.20}$$

Оскільки відношення $\frac{(A(0),x)}{\|x\|}$ обмежене, то права частина (11.20) прямує до $+\infty$ при $\|x\| \to +\infty$. Згідно з попереднім наслідком, оператор A_ε сюр'єктивний. Нехай x_ε — розв'язок рівняння $A_\varepsilon x = f \in H$.

Покажемо, що норми $\|x_\varepsilon\|$ рівномірно обмежені за всіх $\varepsilon > 0$. Маємо

$$\frac{(f, x_\varepsilon)}{\|x_\varepsilon\|} = \frac{(A_\varepsilon x_\varepsilon, x_\varepsilon)}{\|x_\varepsilon\|} \geq \varepsilon \|x_\varepsilon\| - \|A(0)\|,$$

тому

$$\varepsilon \|x_\varepsilon\| \leq \|A(0)\| + \|f\| = K,$$

де K не залежить від ε. Оскільки $Ax_\varepsilon + \varepsilon x_\varepsilon = f$, то $\|Ax_\varepsilon\| \leq \varepsilon \|x_\varepsilon\| + \|f\| \leq \text{const}$, де константа не залежить від ε. Унаслідок (11.19) звідси випливає, що $\|x_\varepsilon\| \leq \text{const}$, де константа знову не залежить від ε.

Нехай $\varepsilon_k > 0$, $\varepsilon_k \to 0$. Тоді з послідовності (x_{ε_k}) можна виділити слабко збіжну підпослідовність (яку знову позначимо через (x_{ε_k})), границю якої позначимо через x. Оскільки $Ax_{\varepsilon_k} + \varepsilon_k x_{\varepsilon_k} = f$, то $Ax_{\varepsilon_k} \to f$. Унаслідок монотонності A

$$(Ax_{\varepsilon_k} - Ay, x_{\varepsilon_k} - y) \geq 0 \quad \forall y \in H.$$

Після граничного переходу отримуємо

$$(f - Ay, x - y) \geq 0 \quad \forall y \in H.$$

Унаслідок леми Мінті маємо

$$(f - Ax, x - y) \geq 0 \quad \forall y \in H.$$

Звідси випливає $Ax - f = 0$. □

Зауваження 11. Якщо оператор A строго монотонний, то в теоремах 21, 22 і наслідках 4, 5 має місце єдиність розв'язку.

Зауваження 12. Якщо в наслідку 4 оператор A додатково сильно монотонний, то обернений оператор A^{-1} є ліпшицевим. А якщо оператор A ще й ліпшицевий, то A^{-1} — сильно монотонний.

Означення 43. Оператор $A : C \to H$ називаємо ***хемінеперервним***, якщо для всіх $x, y \in C$, $z \in H$ функція

$$[0, 1] \ni \lambda \mapsto (A(x + \lambda(y - x)), z) \in \mathbb{R}$$

неперервна, тобто звуження оператора A на відрізки, що лежать у C, неперервне в слабкій топології H.

Зауваження 13. Усі твердження цього підрозділу справедливі для монотонних хемінеперервних операторів. Інструментом доведення в цій ситуції є лема Кнастера – Куратовського – Мазуркевича.

11.6. Апроксимація Браудера – Тихонова

У цьому підрозділі розглядається схема апроксимації розв'язків (11.14), якщо вони існують, послідовністю розв'язків деяких допоміжних нерівностей, що мають ліпші властивості. Використання такої апроксимації для побудови регуляризуючих алгоритмів для задач оптимізації було запропоновано

А. Тихоновим. Аналогічні загальніші твердження щодо спеціальної апроксимації розв'язків варіаційних нерівностей були доведені Ф. Браудером. Тому згідно з [3] називатимемо подібні апроксимації апроксимаціями Браудера – Тихонова.

Нехай C — замкнена опукла підмножина гільбертового простору H, $A : C \to H$ — монотонний хемінеперервний оператор. Розглянемо варіаційну нерівність:

$$\text{знайти } x \in C : \ (Ax, y - x) \geq 0 \quad \forall y \in C. \tag{11.21}$$

Припустимо, що $VI(A, C) \neq \emptyset$.

Зауваження 14. Множина $VI(A, C)$ замкнена та опукла.

Зафіксуємо сильно монотонний хемінеперервний оператор $R : C \to H$. Розв'язки нерівності (11.21) будемо наближати розв'язками нерівностей з операторами $A + \varepsilon R$, $\varepsilon > 0$:

$$\text{знайти } x_\varepsilon \in C : \ (Ax_\varepsilon + \varepsilon Rx_\varepsilon, y - x_\varepsilon) \geq 0 \ \forall y \in C. \tag{11.22}$$

Ясно, що варіаційна нерівність (11.22) має єдиний розв'язок. Наступна теорема описує основну асимптотичну властивість x_ε.

Теорема 23. *Нехай C — замкнена опукла підмножина гільбертового простору H, $A : C \to H$ — монотонний хемінеперервний оператор, $R : C \to H$ — сильно монотонний хемінеперервний оператор. Якщо $VI(A, C) \neq \emptyset$, то*

$$\lim_{\varepsilon \to 0} \|x_\varepsilon - z\| = 0,$$

де $z \in VI(A, C)$ — єдиний розв'язок варіаційної нерівності

$$z \in VI(A, C) : \ (Rz, y - z) \geq 0 \ \forall y \in VI(A, C). \tag{11.23}$$

Доведення. Доведемо обмеженість $\{x_\varepsilon\}_{\varepsilon > 0}$. Підставимо еле-

мент $y \in VI(A,C)$ в (11.22). За лемою Мінті маємо

$$0 \geq (Ax_\varepsilon, y - x_\varepsilon) \geq \varepsilon(Rx_\varepsilon, x_\varepsilon - y). \tag{11.24}$$

Сильна монотонність R дає

$$0 \geq \varepsilon(Rx_\varepsilon, x_\varepsilon - y) \geq \varepsilon(Ry, x_\varepsilon - y) + \varepsilon m \left\|x_\varepsilon - y\right\|^2,$$

де $m > 0$ — константа сильної монотонності оператора R. Звідси

$$m \left\|x_\varepsilon - y\right\| \leq \left\|Ry\right\|.$$

Таким чином, множина $\{x_\varepsilon\}_{\varepsilon>0}$ обмежена.

Нехай $\varepsilon_k > 0$, $\varepsilon_k \to 0$. Тоді з послідовності (x_{ε_k}) можна виділити слабко збіжну підпослідовність (яку знову позначимо через (x_{ε_k})), границю якої позначимо через x^*. Ясно, що $x^* \in C$. Перейшовши в нерівності

$$(Ay, y - x_{\varepsilon_k}) + \varepsilon_k(Ry, y - x_{\varepsilon_k}) \geq 0 \quad \forall y \in C$$

до границі, отримаємо $(Ay, y - x^*) \geq 0 \ \forall y \in C$, тобто $x^* \in VI(A,C)$.

Покажемо, що (x_{ε_k}) сильно збігається до z — єдиного розв'язку варіаційної нерівності (11.23). Унаслідок сильної монотонності R це випливатиме з

$$\lim_{k\to\infty}(Rx_{\varepsilon_k} - Rz, x_{\varepsilon_k} - z) = 0. \tag{11.25}$$

Завдяки (11.24) маємо

$$0 \leq (Rx_{\varepsilon_k} - Rz, x_{\varepsilon_k} - z) \leq -(Rz, x_{\varepsilon_k} - z) = (Rz, z - x_{\varepsilon_k}).$$

Звідси

$$\limsup_{k\to\infty}(Rx_{\varepsilon_k} - Rz, x_{\varepsilon_k} - z) \leq (Rz, z - x^*) \leq 0.$$

Таким чином, виконується (11.25).

З єдиності елемента z випливає сильна збіжність кривої $\varepsilon \mapsto x_\varepsilon$ до z при $\varepsilon \to 0$. $\square$

Якщо $Rx = x - a$, $a \in H$, то теорема 23 дає

$$x_\varepsilon \to P_{VI(A,C)}a \quad \text{при} \quad \varepsilon \to 0.$$

З теореми 23 випливає

Наслідок 6. *Нехай оператор $A : H \to H$ монотонний, хемінеперервний, $f \in H$ та $A^{-1}f \neq \emptyset$. Тоді*

$$x_\varepsilon = (A + \varepsilon I)^{-1} f \to z = P_{A^{-1}f}0 \quad \textit{при} \quad \varepsilon \to 0,\ \varepsilon > 0.$$

Елемент $z = P_{A^{-1}f}0$ називають ***нормальним розв'язком*** операторного рівняння $Ax = f$ (розв'язком з мінімальною нормою).

11.7. Проксимальний метод

Нехай C — замкнена опукла підмножина гільбертового простору H, $A : C \to H$ — монотонний хемінеперервний оператор.

Означення 44. ***Проксимальним оператором*** (щодо A) називають оператор $J_A : H \to 2^C$:

$$x \mapsto J_A x = \{z \in C :\ (Az, y - z) + \\ +(z - x, y - z) \geq 0\ \ \forall y \in C\}. \quad (11.26)$$

Проксимальний оператор J_A всюди визначений, однозначний та має місце нерівність

$$\|J_A x - J_A y\|^2 \leq (J_A x - J_A y, x - y) \quad \forall x, y \in H,$$

яка рівносильна такій:

$$\|J_A x - J_A y\|^2 \le \|x - y\|^2 - \\ - \|(x - J_A x) - (y - J_A y)\|^2 \quad \forall x, y \in H. \quad (11.27)$$

Крім того, множина нерухомих точок оператора J_A збігається з $VI(A, C)$.

Розглянемо ітераційний процес, породжений проксимальним оператором, а саме: нехай

$$x_{n+1} = J_A x_n, \quad n = 0, 1, \ldots, \quad (11.28)$$

де $x_0 \in H$ — довільна початкова точка.

Процес (11.28) називають проксимальним методом; він є, по суті, лише методом простої ітерації для пошуку нерухомої точки оператора J_A. Зазначимо, що проксимальний метод є дворівневим: на кожній його ітерації слід розв'язати допоміжну варіаційну нерівність (але із сильно монотонним оператором $x \mapsto Ax + x - x_n$), що вимагає свого, узагалі кажучи, нескінченного обчислювального процесу.

Припустимо, що $VI(A, C) \ne \emptyset$. З нерівності (11.27) для $\|x_{n+1} - z\|^2$, де $z \in VI(A, C)$, отримуємо

$$\|x_{n+1} - z\|^2 = \|J_A x_n - J_A z\|^2 \le$$

$$\le \|x_n - z\|^2 - \|x_n - J_A x_n\|^2 = \|x_n - z\|^2 - \|x_n - x_{n+1}\|^2.$$

Звідси

$$\|x_{n+1} - z\| \le \|x_n - z\|$$

та

$$\sum_{k=0}^{n} \|x_{k+1} - x_k\|^2 \le \|x_0 - z\|^2 \quad \forall n \in \mathbb{N}. \quad (11.29)$$

Таким чином, існує $\lim_{n\to\infty} \|x_n - z\| \in \mathbb{R}$, послідовність (x_n)

обмежена, а з (11.29) випливає

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0. \tag{11.30}$$

Покажемо за допомогою (11.30), що всі часткові слабкі границі послідовності (x_n) належать $VI(A, C)$. Нехай підпослідовність (x_{n_k}) слабко збігається до $z \in H$. Очевидно, що $z \in C$. Маємо

$$(Ax_{n_k+1}, x - x_{n_k+1}) + (x_{n_k+1} - x_{n_k}, x - x_{n_k+1}) \geq 0 \ \forall x \in C.$$

Звідси для всіх $x \in C$ отримуємо

$$\begin{aligned} 0 \leq (x_{n_k+1} - x_{n_k}, x - x_{n_k+1}) + (Ax, x - x_{n_k+1}) = \\ = (x_{n_k+1} - x_{n_k}, x - x_{n_k+1}) + (Ax, x - x_{n_k}) + \\ + (Ax, x_{n_k} - x_{n_k+1}). \end{aligned} \tag{11.31}$$

Перейшовши до границі при $k \to \infty$ в (11.31), отримаємо

$$(Ax, x - z) \geq 0 \quad \forall x \in C,$$

тобто $z \in VI(A, C)$.

Покажемо, що (x_n) слабко збігається до деякого елемента $z \in VI(A, C)$. Нехай $a, b \in VI(A, C)$ — дві слабкі часткові границі послідовності (x_n). Припустимо, що $a \neq b$, $x_{n_k} \rightharpoonup a$, $x_{n_l} \rightharpoonup b$. Тоді

$$\begin{aligned} \lim_{n\to\infty} \|x_n - a\| = \lim_{k\to\infty} \|x_{n_k} - a\| < \lim_{k\to\infty} \|x_{n_k} - b\| = \\ = \lim_{l\to\infty} \|x_{n_l} - b\| < \lim_{l\to\infty} \|x_{n_l} - a\| = \lim_{n\to\infty} \|x_n - a\|. \end{aligned}$$

Абсурдна нерівність указує на те, що $a = b$. Отже, послідовність (x_n) слабко збіжна.

Має місце

Теорема 24. *Нехай C — замкнена опукла підмножина гільбертового простору H, $A : C \to H$ — монотонний хемінеперервний оператор. Якщо $VI(A, C) \neq \emptyset$, то породжена проксимальним методом (11.28) послідовність (x_n) слабко збігається до деякої точки з $VI(A, C)$.*

З теореми 24 випливає

Наслідок 7. *Нехай оператор $A : H \to H$ монотонний, хемінеперервний, $f \in H$ та $A^{-1}f \neq \emptyset$. Тоді послідовність (x_n), що задана формулою*

$$x_{n+1} = (A + I)^{-1}(x_n + f), \;\; x_0 \in H,$$

слабко збігається до деякого розв'язку рівняння $Ax = f$.

Розділ 12.

Проекційні методи розв'язання варіаційних нерівностей

Нехай C — непорожня підмножина дійсного гільбертового простору H, $A : H \to H$ — оператор, що діє в H. У даному розділі розглянемо основні проекційні методи розв'язання варіаційних нерівностей вигляду:

$$\text{знайти } x \in C : \ (Ax, y - x) \geq 0 \quad \forall y \in C. \tag{12.1}$$

Якщо не вказано інше, то будемо припускати виконаними такі умови:

A1) множина $C \subseteq H$ замкнена та опукла;

A2) оператор $A : H \to H$ монотонний і ліпшицевий (зі сталою $L > 0$) на множині C;

A3) $VI(A, C) \neq \emptyset$.

При написанні розділу використані роботи [11, 19, 20, 26, 29-31, 33-35, 37].

12.1. Допоміжні твердження

Лема 6 (З. Оп'ял, 1967 [34]). *Якщо послідовність (x_n) точок гільбертового простору H слабко збігається до точки $x \in H$, то для довільної точки $y \in H \backslash \{x\}$ має місце нерівність*

$$\liminf_{n\to\infty} \|x_n - x\| < \liminf_{n\to\infty} \|x_n - y\| . \tag{12.2}$$

Доведення. Для доведення нерівності (12.2) достатньо зазначити, що в рівності

$$\|x_n - y\|^2 = \|x_n - x\|^2 + \|x - y\|^2 + 2(x_n - x, x - y)$$

останній доданок прямує до нуля при $n \to \infty$. □

Лема 7 (Г. Б. Пасті, 1979 [35]). *Нехай H — гільбертовий простір; $F \subseteq H$ — непорожня множина; (x_n) — послідовність елементів H та $z_n = \frac{\sum_{k=1}^{n} \lambda_k x_k}{\sum_{k=1}^{n} \lambda_k}$, де (λ_n) — послідовність додатних чисел така, що $\sum_{n=1}^{\infty} \lambda_n = +\infty$. Припустимо, що: 1) границя довільної слабко збіжної підпослідовності (z_{n_k}) належить F; 2) для довільної точки $y \in F$ існує $\lim_{n\to\infty} \|x_n - y\| \in \mathbb{R}$. Тоді послідовність (z_n) слабко збігається до деякої точки $z \in F$.*

Наведемо кілька фактів про числові послідовності, які будуть використані при доведенні збіжності методів.

Лема 8. *Нехай (a_n), (b_n) — послідовності невід'ємних чисел, що задовольняють рекурентну нерівність*

$$a_{n+1} \le a_n - b_n.$$

Тоді (a_n) збіжна та $\lim_{n\to\infty} b_n = 0$.

Лема 9. *Нехай (a_n), (b_n) — послідовності невід'ємних чисел такі, що $a_{n+1} \le a_n + b_n$, $\sum_{n=1}^{\infty} b_n < +\infty$. Тоді існує границя $\lim_{n\to\infty} a_n \in \mathbb{R}$.*

Лема 10. *Нехай (x_n) — послідовність невід'ємних чисел, що задовольняє рекурентну нерівність*

$$x_{n+1} \leq (1 - a_n)x_n + a_n b_n + c_n,$$

де послідовності (a_n), (b_n) і (c_n) мають властивості:

1) $a_n \in [0, 1)$ та $\sum_{n=1}^{\infty} a_n = +\infty$;

2) $\limsup_{n\to\infty} b_n \leq 0$;

3) $c_n \in [0, +\infty)$ та $\sum_{n=1}^{\infty} c_n < +\infty$.

Тоді $\lim_{n\to\infty} x_n = 0$.

Доведення. Для довільного $\epsilon > 0$ існує таке $N \in \mathbb{N}$, що для всіх $n \geq N$

$$b_n < \epsilon\,, \ \sum_{n=N}^{\infty} c_n < \epsilon.$$

Для $n > N$ з рекурентної нерівності отримуємо

$$x_{n+1} \leq \left(\prod_{k=N}^{n}(1 - a_k)\right) x_N + \left(1 - \prod_{k=N}^{n}(1 - a_k)\right)\epsilon + \sum_{k=N}^{n} c_k.$$

Беручи до уваги рівносильність умов $\sum_{n=0}^{\infty} a_n = +\infty$ та $\prod_{n=0}^{\infty}(1 - a_n) = 0$, отримуємо, що $\limsup_{n\to\infty} x_n \leq 2\epsilon$. Звідси $\lim_{n\to\infty} x_n = 0$. □

Лема 11 (П.-Е. Маж, 2008 [31]). *Нехай числова послідовність (a_n) має підпослідовність (a_{n_k}) із властивістю $a_{n_k} < a_{n_k+1}$ для всіх $k \in \mathbb{N}$. Тоді існує така неспадна послідовність (m_k) натуральних чисел, що $m_k \to +\infty$ та $a_{m_k} \leq a_{m_k+1}$, $a_k \leq a_{m_k+1}$ для всіх $k \geq n_1$.*

Доведення. Для $k \geq n_1$ покладемо

$$m_k = \max\{i \leq k : \, a_i < a_{i+1}\}\,.$$

Послідовність чисел m_k коректно визначена, неспадна та $m_k \to +\infty$. Нерівність $a_{m_k} \leq a_{m_k+1}$ очевидна.

Доведемо виконання для всіх $k \geq n_1$ нерівності

$$a_k \leq a_{m_k+1}. \tag{12.3}$$

Розглянемо три можливі випадки:

(i) $m_k = k$;

(ii) $m_k = k-1$;

(iii) $m_k < k-1$.

У випадку (i) нерівність (12.3) збігається з $a_{m_k} \leq a_{m_k+1}$. У випадку (ii) нерівність (12.3) очевидна. У випадку (iii) для всіх $i \in \{m_k+1, m_k+2, ..., k-1\}$ виконується $a_i \geq a_{i+1}$, точніше $a_{m_k+1} \geq a_{m_k+2} \geq ... \geq a_{k-1} \geq a_k$. □

12.2. Найпростіший проекційний метод

Розглянемо ітераційний процес.

Алгоритм 1.

1) Задаємо $x_0 \in C$, $\lambda > 0$.

2) Для x_n обчислюємо

$$x_{n+1} = P_C\left(x_n - \lambda A x_n\right).$$

3) Якщо $x_n = x_{n+1}$, то СТОП, інакше покладаємо $n := n+1$ та переходимо на крок 2.

При зупинці алгоритму 1 отримуємо роз'язок варіаційної нерівності (12.1). Дійсно, рівність

$$x_{n+1} = P_C\left(x_n - \lambda A x_n\right)$$

рівносильна варіаційній нерівності

$$(x_{n+1} - x_n + \lambda A x_n, x - x_{n+1}) \geq 0 \quad \forall x \in C.$$

З урахуванням умови $x_n = x_{n+1}$ маємо $x_n \in VI(A, C)$.

Розглянемо спочатку випадок ліпшицевого та сильно монотонного оператора A, тобто вважаємо, що

$$\|Ax - Ay\| \leq L \|x - y\| \quad \forall x, y \in C,$$
$$(Ax - Ay, x - y) \geq \alpha \|x - y\|^2 \quad \forall x, y \in C,$$

де L, $\alpha > 0$. У цьому випадку варіаційна нерівність (12.1) має єдиний розв'язок $z \in C$.

Для $\lambda > 0$ введемо оператор

$$Tx = P_C(x - \lambda A x),$$

що діє із C в C. Покажемо, що він є стискаючим при $0 < \lambda < 2\alpha L^{-2}$. Маємо:

$$\begin{aligned}\|Tx - Ty\|^2 &= \|P_C(x - \lambda Ax) - P_C(y - \lambda Ay)\|^2 \leq \\ &\leq \|x - \lambda Ax - y + \lambda Ay\|^2 = \|x - y\|^2 + \lambda^2 \|Ax - Ay\|^2 - \\ &\quad - 2\lambda(Ax - Ay, x - y) \leq (1 + \lambda^2 L^2 - 2\alpha\lambda) \|x - y\|^2,\end{aligned}$$

тобто

$$\|Tx - Ty\| \leq q(\lambda) \|x - y\| \quad \forall x, y \in C, \tag{12.4}$$

де $q(\lambda) = \sqrt{1 + \lambda^2 L^2 - 2\alpha\lambda}$. Оскільки $0 < \lambda < 2\alpha L^{-2}$, то $0 < q(\lambda) < 1$. Це означає, що оператор T стискаючий. Зазначимо, що замкнена множина $C \subseteq H$ є повним метричним простором з метрикою $d(x, y) = \|x - y\|$. Алгоритм 1 при $0 < \lambda < 2\alpha L^{-2}$, записаний у вигляді $x_{n+1} = Tx_n$, є класичним процесом пошуку нерухомої точки стискаючого оператора T, яка існує, єдина та є розв'язком (12.1). З (12.4) випливає

$$\|x_n - x_k\| \leq q(\lambda)^n \|x_0 - x_{k-n}\| \quad \forall k \geq n.$$

Звідси при $k \to \infty$ отримаємо оцінку

$$\|x_n - z\| \le q(\lambda)^n \|x_0 - z\| \quad n = 0, 1, ...$$

Отже, має місце така теорема.

Теорема 25. *Нехай C — опукла замкнена множина, оператор A ліпшицевий і сильно монотонний на C. Нехай $0 < \lambda < 2\alpha L^{-2}$, де L, α — сталі ліпшицевості та сильної монотонності оператора A, відповідно. Тоді породжена алгоритмом 1 послідовність (x_n) сильно збігається до єдиного розв'язку (12.1) z, причому справедлива оцінка*

$$\|x_n - z\| \le q(\lambda)^n \|x_0 - z\| \quad n = 0, 1, ...,$$

де $q(\lambda) = \sqrt{1 - 2\alpha\lambda + \lambda^2 L^2} \in (0, 1)$.

Зауваження 15. Найменше значення $q(\lambda)$ при $0 < \lambda < 2\alpha L^{-2}$ досягається в $\lambda_0 = \alpha L^{-2}$ та становить

$$q(\lambda) = \sqrt{1 - \alpha^2 L^{-2}}.$$

Тепер розглянемо алгоритм 1 у випадку варіаційної нерівності з обернено сильно монотонним оператором A.

Означення 45. Оператор $A : C \to H$ називають ***обернено сильно монотонним*** (***ко-коерцитивним***), якщо існує додатна константа α така, що

$$(Ax - Ay, x - y) \ge \alpha \|Ax - Ay\|^2 \quad \forall x, y \in C.$$

У цьому випадку кажуть, що A — α-обернено сильно монотонний.

Важливим прикладом обернено сильно монотонних операторів є ліпшицеві похідні опуклих функціоналів.

Теорема 26. *Нехай функціонал $f : H \to \mathbb{R}$ диференційовний за Фреше. Тоді такі умови рівносильні:*

1) функціонал f опуклий та похідна ∇f задовольняє умову Ліпшиця зі сталою $L > 0$;

2) для всіх $x, y \in H$:

$$f(y) \geq f(x) + (\nabla f(x), y - x) + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2;$$

3) для всіх $x, y \in H$:

$$(\nabla f(x) - \nabla f(y), x - y) \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2.$$

Доведення. Покажемо, що 1) $\Rightarrow$ 2) $\Rightarrow$ 3) $\Rightarrow$ 1).
1) $\Rightarrow$ 2). Має місце

Лема 12. *Нехай функціонал $f : H \to \mathbb{R}$ — обмежений знизу та диференційовний за Фреше; похідна ∇f задовольняє умову Ліпшиця зі сталою $L > 0$. Тоді*

$$\inf_H f \leq f(x) - \frac{1}{2L} \|\nabla f(x)\|^2 \quad \forall x \in H. \tag{12.5}$$

Доведення. Для $x, y \in H$ маємо

$$f(y) = f(x) + \int_0^1 (\nabla f(x + \tau(y - x)), y - x) d\tau = f(x) +$$
$$+ (\nabla f(x), y - x) + \int_0^1 (\nabla f(x + \tau(y - x)) - \nabla f(x), y - x) d\tau.$$

Оцінимо підінтегральний вираз

$$(\nabla f(x + \tau(y - x)) - \nabla f(x), y - x) \leq$$
$$\leq \|\nabla f(x + \tau(y - x)) - \nabla f(x)\| \, \|y - x\| \leq$$
$$\leq \tau \cdot L \cdot \|y - x\|^2.$$

Отримуємо

$$f(y) \le f(x) + (\nabla f(x), y - x) + \frac{L}{2} \|y - x\|^2 .$$

Поклавши $y = x - \frac{1}{L}\nabla f(x)$, матимемо

$$f\left(x - \frac{1}{L}\nabla f(x)\right) \le f(x) - \frac{1}{2L} \|\nabla f(x)\|^2 .$$

Оцінивши ліву частину через $\inf_H f$, отримуємо нерівність (12.5). □

Для $x \in H$ розглянемо допоміжний функціонал

$$y \mapsto \phi(y) = f(y) - (\nabla f(x), y).$$

Функціонал ϕ опуклий, похідна $\nabla\phi(y) = \nabla f(y) - \nabla f(x)$ задовольняє умову Ліпшица зі сталою $L > 0$. Крім того, $x \in \operatorname{argmin}_{y\in H} \phi(y)$[1]. Нерівність (12.5) для ϕ дає

$$f(x) - (\nabla f(x), x) \le f(y) - (\nabla f(x), y) - \\ - (2L)^{-1} \|\nabla f(y) - \nabla f(x)\|^2 ,$$

що і треба було довести.

2) ⇒ 3). Маємо

$$f(y) - f(x) \ge (\nabla f(x), y - x) + (2L)^{-1} \|\nabla f(x) - \nabla f(y)\|^2 ,$$
$$f(x) - f(y) \ge (\nabla f(y), x - y) + (2L)^{-1} \|\nabla f(y) - \nabla f(x)\|^2 .$$

Склавши ці нерівності, отримаємо

$$0 \ge (\nabla f(x) - \nabla f(y), y - x) + (2L)^{-1} \|\nabla f(x) - \nabla f(y)\|^2 ,$$

[1] Нагадаємо, що з опуклості f випливає

$$f(y) \ge f(x) + (\nabla f(x), y - x) \quad \forall x, y \in H.$$

що і треба було довести.

3) $\Rightarrow$ 1). Очевидно. $\square$

Має місце

Теорема 27. *Нехай C — опукла замкнена множина, оператор A обернено сильно монотонний на C. Нехай $0 < \lambda < 2\alpha$, де α — стала оберненої сильної монотонності оператора A. Тоді породжена алгоритмом 1 послідовність (x_n) слабко збігається до деякого розв'язку (12.1).*

Доведення. Для $0 < \lambda < 2\alpha$ розглянемо оператори $Tx = P_C Sx$, $Sx = x - \lambda Ax$. Множина нерухомих точок оператора T збігається з множиною $VI(A, C)$ розв'язків варіаційної нерівності (12.1), а алгоритм 1 записуєтья у вигляді $x_{n+1} = Tx_n$.

Для оператора S має місце оцінка

$$\|Sx - Sy\|^2 \le \|x - y\|^2 - \\ - \left(\frac{2\alpha}{\lambda} - 1\right) \|(x - Sx) - (y - Sy)\|^2 \quad \forall x, y \in C.$$

Ураховуючи опуклість функції $\|\cdot\|^2$, отримаємо

$$\begin{aligned} &\|(x - Tx) - (y - Ty)\|^2 / 2 = \\ &\quad = \|(x - P_C Sx) - (y - P_C Sy)\|^2 / 2 = \|(x - Sx) - \\ &\quad -(y - Sy) + (Sx - P_C Sx) - (Sy - P_C Sy)\|^2 / 2 \le \\ &\qquad \le \|(x - Sx) - (y - Sy)\|^2 + \\ &\qquad + \|(Sx - P_C Sx) - (Sy - P_C Sy)\|^2 \le \\ &\quad \le \lambda(2\alpha - \lambda)^{-1} \left(\|x - y\|^2 - \|Sx - Sy\|^2\right) + \\ &\qquad + \|Sx - Sy\|^2 - \|P_C Sx - P_C Sy\|^2 \le \\ &\le \max\left\{\lambda(2\alpha - \lambda)^{-1}, 1\right\} \left(\|x - y\|^2 - \|P_C Sx - P_C Sy\|^2\right). \end{aligned}$$

Таким чином,

$$\|Tx - Ty\|^2 \le \|x - y\|^2 - \\ - \mu \|(x - Tx) - (y - Ty)\|^2 \quad \forall x, y \in C, \quad (12.6)$$

де $\mu = \left(\max\left\{2\lambda(2\alpha - \lambda)^{-1}, 2\right\}\right)^{-1} > 0$.

Використовуючи (12.6), оцінимо зверху $\|x_{n+1} - z\|^2$, де $z \in VI(A, C)$,

$$\|x_{n+1} - z\|^2 = \|Tx_n - Tz\|^2 \le \|x_n - z\|^2 - \\ - \mu \|x_n - Tx_n\|^2 = \|x_n - z\|^2 - \mu \|x_n - x_{n+1}\|^2 .$$

Звідси

$$\|x_{n+1} - z\| \le \|x_n - z\| ,$$

та

$$\mu \sum_{k=0}^{n} \|x_{k+1} - x_k\|^2 \le \|x_0 - z\|^2 \quad \forall n \in \mathbb{N}. \qquad (12.7)$$

Таким чином, існує $\lim_{n\to\infty} \|x_n - z\| \in \mathbb{R}$, послідовність (x_n) обмежена, а з (12.7) випливає

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0. \qquad (12.8)$$

Покажемо за допомогою (12.8), що всі часткові слабкі границі послідовності (x_n) належать $VI(A, C)$. Нехай підпослідовність (x_{n_k}) слабко збігається до $z \in H$. Очевидно, що $z \in C$. Маємо

$$(x_{n_k+1} - x_{n_k} + \lambda A x_{n_k}, x - x_{n_k+1}) \ge 0 \quad \forall x \in C.$$

Звідси для всіх $x \in C$ отримуємо

$$0 \le (x_{n_k+1} - x_{n_k}, x - x_{n_k+1}) + \lambda (Ax_{n_k}, x_{n_k} - x_{n_k+1}) +$$

$$+\lambda\left(Ax_{n_k}, x - x_{n_k}\right) \le \left(x_{n_k+1} - x_{n_k}, x - x_{n_k+1}\right) +$$
$$+\lambda\left(Ax_{n_k}, x_{n_k} - x_{n_k+1}\right) + \lambda\left(Ax, x - x_{n_k}\right). \qquad (12.9)$$

Перейшовши до границі при $k \to \infty$ в (12.9), отримаємо

$$(Ax, x - z) \ge 0 \quad \forall x \in C,$$

тобто $z \in VI(A, C)$.

Покажемо, що (x_n) слабко збігається до деякого елемента $z \in VI(A, C)$. Нехай $a, b \in VI(A, C)$ — дві слабкі часткові границі послідовності (x_n). Припустимо, що $a \ne b$, $x_{n_k} \rightharpoonup a$, $x_{n_l} \rightharpoonup b$. Тоді

$$\lim_{n\to\infty} \|x_n - a\| = \lim_{k\to\infty} \|x_{n_k} - a\| < \lim_{k\to\infty} \|x_{n_k} - b\| =$$
$$= \lim_{l\to\infty} \|x_{n_l} - b\| < \lim_{l\to\infty} \|x_{n_l} - a\| = \lim_{n\to\infty} \|x_n - a\|.$$

Абсурдна нерівність указує на те, що $a = b$. Отже, послідовність (x_n) слабко збіжна. □

Розглянемо регуляризований варіант найпростішого проекційного методу для нерівностей з обернено сильно монотонними ліпшицевими операторами. Зафіксуємо число $\lambda \in (0, 2\alpha)$, де α — стала оберненої сильної монотонності оператора A, і послідовність чисел (α_n) таку, що $\alpha_n \in (0, 1)$, $\lim_{n\to\infty} \alpha_n = 0$, $\sum_{n=0}^{\infty} \alpha_n = +\infty$, $\sum_{n=1}^{\infty} |\alpha_{n+1} - \alpha_n| < +\infty$ або $\lim_{n\to\infty}(\alpha_{n+1} - \alpha_n)/\alpha_{n+1} = 0$.

Алгоритм 2.

1) Задаємо $x_0 \in H$.

2) Для x_n обчислюємо

$$y_n = P_C(x_n - \lambda A x_n),$$
$$x_{n+1} = P_C(1 - \alpha_n) y_n.$$

3) Покладаємо $n := n + 1$ та переходимо на крок 2.

Покажемо, що послідовності (x_n), (y_n) обмежені. Для $z \in VI(A,C)$ маємо

$$\begin{aligned}\|x_{n+1}-z\| &= \|P_C(1-\alpha_n)y_n - P_C z\| \leq \\ &\leq \|(1-\alpha_n)y_n - z\| \leq (1-\alpha_n)\|y_n - z\| + \alpha_n\|z\| = \\ &= (1-\alpha_n)\|P_C(x_n - \lambda A x_n) - P_C(z-\lambda A z)\| + \alpha_n\|z\| \leq \\ &\leq (1-\alpha_n)\|x_n - z\| + \alpha_n\|z\| \leq \max\{\|x_n - z\|, \|z\|\}.\end{aligned}$$

Тому

$$\|x_{n+1}-z\| \leq \max\{\|x_0 - z\|, \|z\|\}. \tag{12.10}$$

З (12.10) випливає обмеженість послідовності (x_n). Обмеженість (y_n) випливає з формули $y_n = P_C(x_n - \lambda A x_n)$.

Покажемо, що

$$\lim_{n\to\infty}\|x_{n+1} - x_n\| = 0 \tag{12.11}$$

та

$$\lim_{n\to\infty}\|x_n - y_n\| = 0. \tag{12.12}$$

Маємо

$$\begin{aligned}\|x_{n+1}-x_n\| &= \|P_C(1-\alpha_n)y_n - P_C(1-\alpha_{n-1})y_{n-1}\| \leq \\ &\leq \|(1-\alpha_n)y_n - (1-\alpha_{n-1})y_{n-1}\| \leq \\ &\leq (1-\alpha_n)\|x_n - x_{n-1}\| + |\alpha_n - \alpha_{n-1}|\cdot\|y_{n-1}\|.\end{aligned}$$

З леми 10 випливає (12.11). Тепер можемо отримати (12.12) із (12.11).

$$\begin{aligned}\|x_n - y_n\| &\leq \|x_n - x_{n+1}\| + \|x_{n+1} - y_n\| = \|x_n - x_{n+1}\| + \\ &+ \|P_C(1-\alpha_n)y_n - y_n\| \leq \|x_n - x_{n+1}\| + \alpha_n\|y_n\| \to 0.\end{aligned}$$

З обмеженості (x_n) випливає існування підпослідовності (x_{n_k}), що слабко збігається до деякої точки $w \in H$. Покажемо, що $w \in VI(A,C)$. Ясно, що $w \in C$. Маємо

$$(y_{n_k} - x_{n_k} + \lambda A x_{n_k}, x - y_{n_k}) \geq 0 \quad \forall x \in C.$$

Звідси

$$(y_{n_k} - x_{n_k}, x - y_{n_k}) + \lambda\,(Ax_{n_k}, x_{n_k} - y_{n_k}) + \\ + \lambda\,(Ax, x - x_{n_k}) \geq 0 \quad \forall x \in C.$$

Після граничного переходу отримаємо

$$(Ax, x - w) \geq 0 \quad \forall x \in C,$$

тобто $w \in VI(A, C)$.

Розглянемо елемент $x^* = P_{VI(A,C)}0$. Доведемо, що

$$\limsup_{n\to\infty}(-x^*, x_n - x^*) \leq 0. \tag{12.13}$$

Виділимо з (x_n) підпослідовність (x_{n_k}) таку, що

$$\limsup_{n\to\infty}(-x^*, x_n - x^*) = \lim_{k\to\infty}(-x^*, x_{n_k} - x^*).$$

Можна вважати, що $x_{n_k} \rightharpoonup w \in VI(A, C)$. Тому маємо

$$\lim_{k\to\infty}(-x^*, x_{n_k} - x^*) = (-x^*, w - x^*) \leq 0,$$

чим і доводимо (12.13).

Покажемо тепер, що $x_n \to x^*$. Маємо

$$\|x_{n+1} - x^*\|^2 = \|P_C(1 - \alpha_n)y_n - x^*\|^2 = \\ = \|P_C(1 - \alpha_n)y_n - P_C x^*\|^2 \leq \\ \leq \|(1 - \alpha_n)(y_n - x^*) - \alpha_n x^*\|^2 = (1 - \alpha_n)^2\,\|y_n - x^*\|^2 + \\ + 2(1 - \alpha_n)\alpha_n(x^* - y_n, x^*) + \alpha_n^2\,\|x^*\|^2 \leq (1 - \alpha_n)\,\|x_n - x^*\|^2 + \\ + 2(1 - \alpha_n)\alpha_n\,\{(x^* - x_n, x^*) +$$

$$+(x_n - y_n, x^*) + \alpha_n \|x^*\|^2 \Big\}. \qquad (12.14)$$

Застосувавши до рекурентної нерівності (12.14) лему 10, робимо висновок, що $x_n \to x^*$ при $n \to \infty$. А з (12.12) випливає $y_n \to x^*$.

Отже, має місце

Теорема 28. *Нехай C — опукла замкнена множина, оператор A обернено сильно монотонний на C. Нехай $0 < \lambda < 2\alpha$, де α — стала оберненої сильної монотонності оператора A, послідовність чисел (α_n) така, що $\alpha_n \in (0,1)$, $\lim_{n\to\infty} \alpha_n = 0$, $\sum_{n=0}^{\infty} \alpha_n = +\infty$, $\sum_{n=1}^{\infty} |\alpha_{n+1} - \alpha_n| < +\infty$ або $\lim_{n\to\infty}(\alpha_{n+1} - \alpha_n)/\alpha_{n+1} = 0$. Тоді породжені алгоритмом 2 послідовності (x_n) та (y_n) сильно збігаються до єдиного нормального розв'язку $x^* = P_{VI(A,C)}0$ задачі (12.1).*

Перейдемо до вивчення ітераційного алгоритму 1 у випадку лише монотонних і хемінеперервних операторів A.

Зафіксуємо послідовність додатних чисел (λ_n), що задовольняє умову

$$\sum_{n=0}^{\infty} \lambda_n = +\infty\ ,\ \sum_{n=0}^{\infty} \lambda_n^2 < +\infty. \qquad (12.15)$$

Розглянемо

Алгоритм 3.

1) Задаємо $x_0 \in C$.

2) Для x_n обчислюємо

$$x_{n+1} = P_C\left(x_n - \lambda_n A x_n\right).$$

3) Якщо $x_n = x_{n+1}$, то СТОП, інакше покладаємо $n := n+1$ та переходимо на крок 2.

Розглянемо послідовність середніх за Чезаро:

$$z_n = \frac{\sum_{k=0}^{n} \lambda_k x_k}{\sum_{k=0}^{n} \lambda_k}.$$

Щодо оператора A зробимо таке припущення:

$$\text{послідовність } (Ax_n) \text{ обмежена.} \tag{12.16}$$

Лема 13. *Для породженої алгоритмом 3 послідовності* (x_n) *і точки* $y \in C$ *виконується нерівність*

$$\|x_{n+1} - y\|^2 \le \|x_n - y\|^2 + \lambda_n^2 \|Ax_n\|^2 - \\ - 2\lambda_n (Ay, x_n - y). \tag{12.17}$$

Доведення. Для $y \in C$ та x_{n+1} маємо

$$\|x_{n+1} - y\|^2 = \|P_C (x_n - \lambda_n A x_n) - y\|^2 \le \\ \le \|x_n - y\|^2 + \lambda_n^2 \|Ax_n\|^2 - 2\lambda_n (Ax_n, x_n - y).$$

Оператор A — монотонний. Тому $(Ax_n, x_n - y) \ge (Ay, x_n - y)$. Отже,

$$\|x_{n+1} - y\|^2 \le \|x_n - y\|^2 + \lambda_n^2 \|Ax_n\|^2 - 2\lambda_n (Ay, x_n - y),$$

що і треба було довести. □

Лема 14. *Для породженої алгоритмом 3 послідовності* (x_n)*, послідовності середніх* (z_n) *і точки* $y \in C$ *виконується нерівність*

$$\frac{\|x_{n+1} - y\|^2 - \|x_0 - y\|^2}{\sum_{k=0}^{n} \lambda_k} \le 2 (Ay, y - z_n) + \\ + \frac{\sum_{k=0}^{n} \lambda_k^2 \|Ax_k\|^2}{\sum_{k=0}^{n} \lambda_k}. \tag{12.18}$$

Доведення. Подамо нерівність леми 13 у вигляді

$$\|x_{k+1}-y\|^2-\|x_k-y\|^2\leq \\ \leq 2\left(Ay,\lambda_k y-\lambda_k x_k\right)+\lambda_k^2\|Ax_k\|^2. \quad (12.19)$$

Підсумувавши (12.19) за k від 0 до $n\in\mathbb{N}$, отримаємо

$$\|x_{n+1}-y\|^2-\|x_0-y\|^2\leq 2\left(Ay,\sum_{k=0}^{n}\lambda_k y-\sum_{k=0}^{n}\lambda_k x_k\right)+ \\ +\sum_{k=0}^{n}\lambda_k^2\|Ax_k\|^2. \quad (12.20)$$

Розділивши (12.20) на $\sum_{k=0}^{n}\lambda_k$, приходимо до (12.18). □

Припустимо, що $VI(A,C)\neq\emptyset$. Має місце

Лема 15. *Нехай (x_n) — породжена алгоритмом 3 послідовність. Тоді для довільної точки $y\in VI(A,C)$ існує скінченна границя $\lim_{n\to\infty}\|x_n-y\|$. Зокрема, послідовність (x_n) обмежена.*

Доведення. Використаємо леми 9 та 13. У нерівності (12.17) покладемо $y\in VI(A,C)$. Отримаємо

$$\|x_{n+1}-y\|^2\leq\|x_n-y\|^2+\lambda_n^2\|Ax_n\|^2, \quad (12.21)$$

оскільки $(Ay,x_n-y)\geq 0$. З нерівності (12.21), припущення (12.16) та умови $(\lambda_n)\in\ell_2$ випливає існування границі $\lim_{n\to\infty}\|x_n-y\|\in\mathbb{R}$. □

Обмеженість послідовності (x_n) зумовлює обмеженість послідовності середніх (z_n). А з леми 14 випливає

Лема 16. *Усі слабкі часткові границі послідовності середніх (z_n) належать множині $VI(A,C)$.*

Доведення. Розглянемо слабко збіжну підпослідовність (z_{n_l}) послідовності (z_n). Нехай $z \in H$ — слабка границя (z_{n_l}). Ясно, що z належить множині C. Записавши нерівність (12.18) для елементів z_{n_l}, після граничного переходу при $l \to \infty$ отримаємо $(Ay, y - z) \geq 0 \ \forall y \in C$, що за лемою 5 рівносильно включенню $z \in VI(A, C)$. □

Сформулюємо теорему про ергодичну збіжність.

Теорема 29. *Справедливі твердження:*

1) якщо $VI(A, C) \neq \emptyset$, то послідовність середніх за Чезаро (z_n) слабко збігається до деякої точки $x \in VI(A, C)$;

2) якщо $VI(A, C) = \emptyset$, то $\|z_n\| \to +\infty$.

Доведення. З лем 15 та 16 випливає, що у випадку $VI(A, C) \neq \emptyset$ для згенерованої алгоритмом 3 послідовності (x_n) і множини $F = VI(A, C)$ виконані умови леми 7. Отже, послідовність (z_n) слабко збігається до деякої точки $x \in VI(A, C)$.

Припустимо, що $VI(A, C) = \emptyset$. Тоді $\|z_n\| \to +\infty$. Дійсно, інакше послідовність (z_n) має слабку граничну точку z, яка, як було показано раніше, належить множині $VI(A, C)$. □

За деяких додаткових умов має місце сильна збіжність послідовності (x_n).

Теорема 30. *Нехай оператор A сильно монотонний. Тоді породжена алгоритмом 3 послідовність (x_n) сильно збігається до єдиного розв'язку (12.1).*

Доведення. Нехай $z \in C$ — розв'язок (12.1), A — сильно монотонний оператор з константою $\mu > 0$. Маємо

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 + \lambda_n^2 \|Ax_n\|^2 - \\ - 2\lambda_n (Ax_n, x_n - y), \ \ y \in C.$$

Завдяки сильній монотонності оператора A отримуємо

$$(Ax_n, x_n - y) \geq (Ay, x_n - y) + \mu \|x_n - y\|^2 .$$

Отже,

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 + \lambda_n^2 \|Ax_n\|^2 - \\ - 2\lambda_n (Ay, x_n - y) - 2\mu\lambda_n \|x_n - y\|^2 . \quad (12.22)$$

Розглянувши у (12.22) варіант $y = z$, отримуємо нерівність

$$2\mu\lambda_n \|x_n - z\|^2 \leq \|x_n - z\|^2 - \|x_{n+1} - z\|^2 + \\ + \lambda_n^2 \|Ax_n\|^2 . \quad (12.23)$$

Просумувавши (12.23) за n від 0 до N, отримаємо

$$2\mu \sum_{n=0}^{N} \lambda_n \|x_n - z\|^2 \leq \|x_0 - z\|^2 + \sum_{n=0}^{N} \lambda_n^2 \|Ax_n\|^2 .$$

Звідси $\sum_{n=0}^{\infty} \lambda_n \|x_n - z\|^2 < +\infty$. Оскільки $(\lambda_n) \notin \ell_1$ та існує $\lim_{n\to\infty} \|x_n - z\|$, то маємо $\lim_{n\to\infty} \|x_n - z\| = 0$. □

Теорема 31. *Нехай* $\mathrm{int} VI(A, C) \neq \emptyset$. *Тоді породжена алгоритмом 3 послідовність* (x_n) *сильно збігається до розв'язку (12.1).*

Доведення. Візьмемо елемент $y \in \mathrm{int} VI(A, C)$. Тоді існує замкнена куля $B(y, r) \subseteq VI(A, C)$, $r > 0$. Запишемо для $y_n = y - r\frac{x_{n+1} - x_n}{\|x_{n+1} - x_n\|} \in B(y, r)$ нерівність (12.21)

$$\|x_{n+1} - y_n\|^2 \leq \|x_n - y_n\|^2 + \lambda_n^2 \|Ax_n\|^2 .$$

Цю нерівність можна записати у такому вигляді:

$$2r \|x_{n+1} - x_n\| \leq \|x_n - y\|^2 - \|x_{n+1} - y\|^2 + \lambda_n^2 \|Ax_n\|^2 .$$

Для довільних $m > n$ маємо

$$\|x_m - x_n\| \le \sum_{k=n}^{m-1} \|x_{k+1} - x_k\| \le \frac{\|x_n - y\|^2 - \|x_m - y\|^2}{2r} + \\ + \frac{1}{2r} \sum_{k=n}^{m-1} \lambda_k^2 \|Ax_k\|^2 .$$

З припущення (12.16), $(\lambda_n) \in \ell_2$ та леми 15 випливає фундаментальність послідовності (x_n). Нехай $z \in H$ — сильна границя (x_n). Тоді послідовність середніх (z_n) сильно збігається до точки z. Включення $z \in VI(A, C)$ випливає з леми 16. $\square$

Включивши операцію усереднення у схему обчислень, отримаємо такий алгоритм.

Алгоритм 4.

1) Задаємо $x_0 = z_0 \in C$; покладаємо $\sigma_0 := \lambda_0$, $n := 0$.

2) Для x_n знаходимо $x_{n+1} = P_C(x_n - \lambda_n A x_n)$.

3) Якщо $x_{n+1} = x_n$, то СТОП та $x_n \in VI(A, C)$. Інакше переходимо на крок 4.

4) Покладаємо

$$\sigma_{n+1} = \sigma_n + \lambda_{n+1},$$
$$z_{n+1} = \left(1 - \frac{\lambda_{n+1}}{\sigma_{n+1}}\right) z_n + \frac{\lambda_{n+1}}{\sigma_{n+1}} x_{n+1},$$

$n := n + 1$, переходимо на крок 2.

Має місце

Теорема 32. *Справедливі твердження:*

1) якщо $VI(A, C) \ne \emptyset$, то послідовність (z_n) слабко збігається до деякого елемента $x \in VI(A, C)$;

2) якщо $VI(A, C) = \emptyset$, то $\|z_n\| \to +\infty$.

12.3. Метод Корпелевич

Одним з найпопулярніших методів розв'язання варіаційних нерівностей є екстраградієнтний метод Корпелевич.

Алгоритм 5 ([11]).

1) Задаємо $x_0 \in C$, $\lambda \in \left(0, \frac{1}{L}\right)$.

2) Для x_n *обчислюємо* $y_n = P_C\left(x_n - \lambda A x_n\right)$.

3) Якщо $x_n = y_n$, *то СТОП, інакше обчислюємо*

$$x_{n+1} = P_C\left(x_n - \lambda A y_n\right),$$

покладаємо $n := n + 1$ *та переходимо на крок 2.*

Лема 17. *Якщо* $x_n = y_n$, *то* $x_n \in VI(A, C)$.

Доведення. Рівність $y_n = P_C\left(x_n - \lambda A x_n\right)$ рівносильна варіаційній нерівності $\left(y_n - x_n + \lambda A x_n, x - y_n\right) \geq 0 \ \forall x \in C$. З урахуванням умови $x_n = y_n$ маємо $x_n \in VI(A, C)$. □

Лема 18. *Для* $z \in VI(A, C)$ *і породжених алгоритмом 5 послідовностей* (x_n), (y_n) *виконується нерівність*

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \left(1 - \lambda^2 L^2\right)\|x_n - y_n\|^2. \qquad (12.24)$$

Доведення. Маємо

$$\begin{aligned}\|x_{n+1} - z\|^2 &\leq \|x_n - \lambda A y_n - z\|^2 - \\ &\quad - \|x_n - \lambda A y_n - x_{n+1}\|^2 = \\ &= \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + 2\lambda\left(A y_n, z - x_{n+1}\right). \qquad (12.25)\end{aligned}$$

З монотонності оператора A та $z \in VI(A, C)$ випливає

$$\begin{aligned}0 \leq \left(A y_n - A z, y_n - z\right) &= \left(A y_n, y_n - z\right) - \left(A z, y_n - z\right) \leq \\ &\leq \left(A y_n, y_n - z\right) = \left(A y_n, y_n - x_{n+1}\right) + \left(A y_n, x_{n+1} - z\right),\end{aligned}$$

тобто

$$(Ay_n, z - x_{n+1}) \leq (Ay_n, y_n - x_{n+1}). \qquad (12.26)$$

Оцінивши праву частину (12.25) за допомогою (12.26), отримаємо

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + \\ + 2\lambda (Ay_n, y_n - x_{n+1}).$$

Далі

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|(x_n - y_n) + (y_n - x_{n+1})\|^2 + \\ + 2\lambda (Ay_n, y_n - x_{n+1}) = \|x_n - z\|^2 - \|x_n - y_n\|^2 - \\ - \|y_n - x_{n+1}\|^2 + 2 (x_n - \lambda Ay_n - y_n, x_{n+1} - y_n).$$

Оскільки $x_{n+1} \in C$, то

$$(x_n - \lambda Ay_n - y_n, x_{n+1} - y_n) = \\ = (x_n - \lambda Ax_n - y_n, x_{n+1} - y_n) + \\ + \lambda (Ax_n - Ay_n, x_{n+1} - y_n) \leq \lambda (Ax_n - Ay_n, x_{n+1} - y_n).$$

Отже,

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_n - y_n\|^2 - \\ - \|y_n - x_{n+1}\|^2 + 2\lambda (Ax_n - Ay_n, x_{n+1} - y_n). \qquad (12.27)$$

Перейдемо до оцінки $2\lambda (Ax_n - Ay_n, x_{n+1} - y_n)$. Маємо

$$2\lambda (Ax_n - Ay_n, x_{n+1} - y_n) \leq \\ \leq 2\lambda L \|x_n - y_n\| \|x_{n+1} - y_n\| \leq \\ \leq \lambda^2 L^2 \|x_n - y_n\|^2 + \|x_{n+1} - y_n\|^2. \qquad (12.28)$$

Використовуючи (12.28) у (12.27), отримаємо нерівність

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda^2 L^2) \|x_n - y_n\|^2,$$

чим і завершуємо доведення. □

Теорема 33. *Породжені алгоритмом 5 послідовності* (x_n), (y_n) *слабко збігаються до розв'язку (12.1).*

Доведення. З нерівності (12.24) випливає існування $\lim_{n\to\infty} \|x_n - z\|$ для всіх $z \in VI(A, C)$. Зокрема, послідовність (x_n) обмежена. Запишемо нерівність (12.24) у вигляді

$$(1 - \lambda^2 L^2) \|x_n - y_n\|^2 \leq \|x_n - z\|^2 - \|x_{n+1} - z\|^2.$$

Маємо

$$\sum_{n=0}^{m} \|x_n - y_n\|^2 \leq \frac{\|x_0 - z\|^2}{1 - \lambda^2 L^2} \quad \forall m \in \mathbb{N}.$$

Звідси $\lim_{n\to\infty} \|x_n - y_n\| = 0$. Отже, послідовність (y_n) також обмежена.

Покажемо, що всі слабкі часткові границі послідовності (x_n) належать множині $VI(A, C)$. Нехай підпослідовність (x_{n_k}) слабко збігається до $x^* \in H$. Очевидно, що $x^* \in C$ та $y_{n_k} \rightharpoonup x^*$. Маємо

$$(y_{n_k} - x_{n_k} + \lambda A x_{n_k}, x - y_{n_k}) \geq 0 \quad \forall x \in C.$$

Звідси для всіх $x \in C$ отримуємо

$$\begin{aligned} 0 \leq (y_{n_k} - x_{n_k}, x - y_{n_k}) + \lambda (Ax_{n_k}, x_{n_k} - y_{n_k}) + \\ + \lambda (Ax_{n_k}, x - x_{n_k}) \leq (y_{n_k} - x_{n_k}, x - y_{n_k}) + \\ + \lambda (Ax_{n_k}, x_{n_k} - y_{n_k}) + \lambda (Ax, x - x_{n_k}). \end{aligned} \quad (12.29)$$

Перейшовши до границі при $k \to \infty$ у (12.29), отримаємо $(Ax, x - x^*) \geq 0$ $\forall x \in C$, тобто $x^* \in VI(A, C)$.

Покажемо, що послідовність (x_n) слабко збігається. Припустимо, що послідовність (x_n) має принаймні дві різні слабкі часткові границі p та q. За доведеним $p,\ q \in VI(A,C)$. Нехай (x_{n_k}) — підпослідовність, що слабко збігається до p. Тоді з леми Оп'яла (лема 6) випливає:

$$\lim_{n\to\infty}\|x_n-p\| = \lim_{k\to\infty}\|x_{n_k}-p\| = \liminf_{k\to\infty}\|x_{n_k}-p\| <$$
$$< \liminf_{k\to\infty}\|x_{n_k}-q\| = \lim_{k\to\infty}\|x_{n_k}-q\| = \lim_{n\to\infty}\|x_n-q\|.$$

Повторивши це міркування, приходимо до абсурдної нерівності $\lim_{n\to\infty}\|x_n-p\| < \lim_{n\to\infty}\|x_n-p\|$. Отже, послідовність (x_n) слабко збігається до точки з множини $VI(A,C)$. □

12.4. Метод Цзена

У цьому підрозділі ми припустимо, що оператор $A: H \to H$ монотонний та ліпшицевий на всьому просторі H.

Алгоритм 6 ([37]).

1) Задаємо $x_0 \in H$, $\lambda \in \left(0, \frac{1}{L}\right)$.

2) Для x_n *обчислюємо*

$$y_n = P_C\left(x_n - \lambda A x_n\right).$$

3) Якщо $x_n = y_n$, *то СТОП, інакше обчислюємо*

$$x_{n+1} = y_n - \lambda\left(Ay_n - Ax_n\right),$$

покладаємо $n := n+1$ *та переходимо на крок 2.*

Зауважимо, що при виконанні умови $x_n = y_n$ маємо $x_n \in VI(A,C)$. Дійсно, у цьому випадку рівність $y_n = P_C\left(x_n - \lambda A x_n\right)$ рівносильна варіаційній нерівності

$$(y_n - x_n + \lambda A x_n, x - y_n) = \lambda\left(Ax_n, x - x_n\right) \geq 0 \quad \forall x \in C.$$

Отже, $x_n \in VI(A,C)$.

Має місце

Лема 19. *Для породжених алгоритмом 6 послідовностей* (x_n), (y_n) *має місце нерівність*

$$\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - \left(1-\lambda^2 L^2\right)\|x_n-y_n\|^2, \qquad (12.30)$$

де $z \in VI(A,C)$.

Доведення. Нехай $z \in VI(A,C)$. Маємо

$$\begin{aligned}\|x_{n+1}-z\|^2 &= \|y_n-\lambda\left(Ay_n-Ax_n\right)-z\|^2 = \\ &= \|(y_n-z)+\lambda\left(Ax_n-Ay_n\right)\|^2 = \|y_n-z\|^2 + \\ &+ 2\lambda(Ax_n-Ay_n,y_n-z)+\lambda^2\|Ax_n-Ay_n\|^2. \qquad (12.31)\end{aligned}$$

Зі включення $z \in VI(A,C)$ та рівності $y_n = P_C\left(x_n-\lambda Ax_n\right)$ випливає

$$(y_n-(x_n-\lambda Ax_n)-\lambda Az, y_n-z) \le 0.$$

З монотонності оператора A випливає

$$(\lambda Az-\lambda Ay_n, y_n-z) \le 0.$$

Склавши ці нерівності, отримаємо

$$(y_n-(x_n-\lambda Ax_n)-\lambda Ay_n, y_n-z) \le 0. \qquad (12.32)$$

З (12.32) випливає нерівність

$$2\lambda(Ax_n-Ay_n,y_n-z) = 2(y_n-(x_n-\lambda Ax_n)-\lambda Ay_n,y_n-z)+$$

$$\begin{aligned}&+2(x_n-y_n,y_n-z) \le 2(x_n-y_n,y_n-z) = \\ &= \|x_n-z\|^2-\|y_n-z\|^2-\|y_n-x_n\|^2. \qquad (12.33)\end{aligned}$$

Урахувавши (12.33) у (12.31), отримаємо

$$\begin{aligned}\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - \\ - \|y_n-x_n\|^2 + \lambda^2\|Ax_n-Ay_n\|^2 \le \\ \le \|x_n-z\|^2 - (1-L^2\lambda^2)\|x_n-y_n\|^2,\end{aligned}$$

що і треба було довести. □

Теорема 34. *Породжені алгоритмом 6 послідовності* (x_n), (y_n) *слабко збігаються до розв'язку (12.1).*

Доведення. З нерівності (12.30) випливає існування $\lim_{n\to\infty}\|x_n-z\|$ для всіх $z \in VI(A,C)$ та рівність $\lim_{n\to\infty}\|x_n-y_n\| = 0$. Зокрема, послідовності (x_n), (y_n) обмежені.

Усі слабкі часткові границі послідовності (x_n) належать множині $VI(A,C)$. Дійсно, нехай підпослідовність (x_{n_k}) слабко збігається до $x^* \in H$. Очевидно, що $y_{n_k} \rightharpoonup x^*$ та $x^* \in C$. Маємо

$$(y_{n_k} - x_{n_k} + \lambda Ax_{n_k}, x - y_{n_k}) \ge 0 \quad \forall x \in C.$$

Звідси для всіх $x \in C$ отримуємо

$$0 \le (y_{n_k}-x_{n_k}, x-y_{n_k}) + \lambda(Ax_{n_k}, x_{n_k}-y_{n_k}) +$$

$$\begin{aligned}+\lambda(Ax_{n_k}, x-x_{n_k}) \le (y_{n_k}-x_{n_k}, x-y_{n_k}) + \\ + \lambda(Ax_{n_k}, x_{n_k}-y_{n_k}) + \lambda(Ax, x-x_{n_k}). \quad (12.34)\end{aligned}$$

Перейшовши до границі при $k\to\infty$ в (12.34), отримаємо

$$(Ax, x-x^*) \ge 0 \quad \forall x \in C,$$

тобто $x^* \in VI(A,C)$.

Припустимо, що послідовність (x_n) має принаймні дві різні слабкі часткові границі p та q. За доведеним p, $q \in VI(A,C)$.

Нехай (x_{n_k}), (x_{m_k}) — підпослідовності, що слабко збігаються до p, q, відповідно. Тоді з леми Оп'яла випливає

$$\lim_{n\to\infty}\|x_n-p\|=\lim_{k\to\infty}\|x_{n_k}-p\|<\lim_{k\to\infty}\|x_{n_k}-q\|=$$
$$=\lim_{n\to\infty}\|x_n-q\|=\lim_{k\to\infty}\|x_{m_k}-q\|<\lim_{n\to\infty}\|x_n-p\|.$$

Отримали абсурдну нерівність. Отже, послідовність (x_n) слабко збігається до точки з множини $VI(A,C)$. □

Розглянемо два варіанти регуляризації алгоритму 6.

Алгоритм 7.

1) Задаємо $x_0\in H$, $a\in H$, $\lambda\in\left(0,\frac{1}{L}\right)$, $\alpha_n\in(0,1)$, $\lim_{n\to\infty}\alpha_n=0$, $\sum_{n=0}^{\infty}\alpha_n=+\infty$.

2) Для x_n *обчислюємо*

$$y_n=P_C\left(x_n-\lambda Ax_n\right).$$

3) Обчислюємо

$$z_n=y_n-\lambda\left(Ay_n-Ax_n\right),$$
$$x_{n+1}=\alpha_n a+(1-\alpha_n)z_n,$$

покладаємо $n:=n+1$ *та переходимо на крок 2.*

Має місце

Лема 20. *Для породжених алгоритмом 7 послідовностей* (x_n), (y_n) *та* (z_n) *має місце нерівність*

$$\|x_{n+1}-z\|^2-\|x_n-z\|^2+\|x_{n+1}-z_n\|^2+$$
$$+\left(1-\lambda^2L^2\right)\|x_n-y_n\|^2\le$$
$$\le-2\alpha_n(z_n-a,x_{n+1}-z),\quad(12.35)$$

де $z\in VI(A,C)$.

Доведення. Маємо

$$\|x_{n+1}-z\|^2 = \|\alpha_n a + (1-\alpha_n)z_n - z\|^2 = \\ = \|z_n - z\|^2 - 2\alpha_n(z_n - a, z_n - z) + \alpha_n^2\|z_n - a\|^2 = \\ = \|z_n - z\|^2 - 2\alpha_n(z_n - a, x_{n+1} - z) - \|x_{n+1} - z_n\|^2 .$$

Оцінимо зверху $\|z_n - z\|^2$, використовуючи лему 19. Отримаємо

$$\|x_{n+1}-z\|^2 \le \|x_n - z\|^2 - \left(1-\lambda^2 L^2\right)\|x_n - y_n\|^2 - \\ - \|x_{n+1} - z_n\|^2 - 2\alpha_n(z_n - a, x_{n+1} - z).$$

Звідси випливає нерівність (12.35). □

Лема 21. *Породжені алгоритмом 7 послідовності* (x_n), (y_n) *та* (z_n) *обмежені.*

Доведення. Нехай $z \in VI(A,C)$. Маємо

$$\|x_{n+1}-z\| = \|\alpha_n a + (1-\alpha_n)z_n - z\| = \|\alpha_n(a-z)+ \\ +(1-\alpha_n)(z_n - z)\| \le \alpha_n\|a-z\| + (1-\alpha_n)\|z_n - z\| .$$

Використавши нерівність леми 19, отримаємо

$$\|x_{n+1}-z\| \le \alpha_n\|a-z\| + (1-\alpha_n)\|x_n - z\| \le \\ \le \max\{\|a-z\|, \|x_n - z\|\} .$$

Отже,

$$\|x_{n+1}-z\| \le \max\{\|a-z\|, \|x_0 - z\|\} \quad \forall n \in \mathbb{N}.$$

Таким чином, послідовність (x_n) обмежена. Обмеженість послідовностей (y_n) та (z_n) випливає з нерівності

$$\|z_n - z\|^2 \le \|x_n - z\|^2 - \left(1-\lambda^2 L^2\right)\|x_n - y_n\|^2$$

та обмеженості (x_n). $\square$

Тепер сформулюємо теорему про сильну збіжність алгоритму 7.

Теорема 35. *Породжені алгоритмом 7 послідовності* (x_n), (y_n) *та* (z_n) *сильно збігаються до точки* $x^* = P_{VI(A,C)}a$.

Доведення. Розглянемо елемент $x^* = P_{VI(A,C)}a$. З леми 21 випливає існування такого $M > 0$, що $|(z_n - a, x_{n+1} - x^*)| \leq M$ для всіх $n \in \mathbb{N}$. Тоді з леми 12.35 одержуємо оцінку

$$\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 + \|x_{n+1} - z_n\|^2 + \\ + \left(1 - \lambda^2 L^2\right) \|x_n - y_n\|^2 \leq 2\alpha_n M. \quad (12.36)$$

Розглянемо числову послідовність $(\|x_n - x^*\|)$. Можливі два варіанти:

a) існує номер $\bar{n} \in \mathbb{N}$ такий, що

$$\|x_{n+1} - x^*\| \leq \|x_n - x^*\| \quad \forall\, n \geq \bar{n};$$

b) існує зростаюча послідовність номерів (n_k) така, що

$$\|x_{n_k+1} - x^*\| > \|x_{n_k} - x^*\| \quad \forall\, k \in \mathbb{N}.$$

Розглянемо варіант a). У цьому випадку існує $\lim_{n\to\infty} \|x_n - x^*\| \in \mathbb{R}$. Оскільки $\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 \to 0$ та $\alpha_n \to 0$, то при $n \to \infty$ маємо

$$\|x_{n+1} - z_n\| \to 0, \quad \|x_n - y_n\| \to 0. \quad (12.37)$$

З обмеженості (x_n) випливає існування підпослідовності (x_{n_k}), що слабко збігається до деякої точки $w \in H$. Покажемо, що $w \in VI(A, C)$. З (12.37) випливає $y_{n_k} \rightharpoonup w$ та $w \in C$. Маємо

$$(y_{n_k} - x_{n_k} + \lambda A x_{n_k}, x - y_{n_k}) \geq 0 \quad \forall x \in C.$$

Звідси

$$(y_{n_k} - x_{n_k}, x - y_{n_k}) + \lambda\,(Ax_{n_k}, x_{n_k} - y_{n_k}) + \\ + \lambda\,(Ax, x - x_{n_k}) \geq 0 \quad \forall x \in C.$$

Після граничного переходу отримаємо

$$(Ax, x - w) \geq 0 \quad \forall x \in C,$$

тобто $w \in VI(A, C)$.

Доведемо, що

$$\limsup_{n\to\infty}(a - x^*, x_{n+1} - x^*) \leq 0. \tag{12.38}$$

Виділимо з (x_n) підпослідовність (x_{n_k}) таку, що

$$\limsup_{n\to\infty}(a - x^*, x_{n+1} - x^*) = \lim_{k\to\infty}(a - x^*, x_{n_k} - x^*).$$

Можна вважати, що $x_{n_k} \rightharpoonup w \in VI(A, C)$. Тому

$$\lim_{k\to\infty}(a - x^*, x_{n_k} - x^*) = (a - x^*, w - x^*) \leq 0,$$

чим і доводимо (12.38).

З (12.38) та нерівності

$$\|x_{n+1} - x^*\|^2 = \|\alpha_n(a - x^*) + (1 - \alpha_n)(z_n - x^*)\|^2 \leq \\ \leq (1 - \alpha_n)^2\,\|z_n - x^*\|^2 + 2\alpha_n\,(a - x^*, x_{n+1} - x^*) \leq \\ \leq (1 - \alpha_n)\,\|x_n - x^*\|^2 + 2\alpha_n\,(a - x^*, x_{n+1} - x^*)\,,$$

узявши до уваги лему 10, робимо висновок, що $\|x_n - x^*\| \to 0$.

З (12.37) випливає

$$\lim_{n\to\infty}\|y_n - x^*\| = \lim_{n\to\infty}\|z_n - x^*\| = 0.$$

Розглянемо варіант b). Використаємо лему 11. У цьому ви-

падку розглянемо послідовність номерів (m_k) із властивостями:

(i) $m_k \nearrow +\infty$;

(ii) $\|x_{m_k+1} - x^*\| \geq \|x_{m_k} - x^*\|$ для всіх $k \geq n_1$;

(iii) $\|x_{m_k+1} - x^*\| \geq \|x_k - x^*\|$ для всіх $k \geq n_1$.

З (12.36) та (ii) випливає

$$\|x_{m_k+1} - z_{m_k}\|^2 + \left(1 - \lambda^2 L^2\right) \|x_{m_k} - y_{m_k}\|^2 \leq \\ \leq -2\alpha_{m_k}(z_{m_k} - a, x_{m_k+1} - z) \leq 2\alpha_{m_k} M. \quad (12.39)$$

Звідси $\lim_{k\to\infty} \|x_{m_k+1} - z_{m_k}\| = \lim_{k\to\infty} \|x_{m_k} - y_{m_k}\| = 0$. Крім того, $\lim_{k\to\infty} \|z_{m_k} - y_{m_k}\| = 0$.

Як і в попередньому випадку, показуємо, що часткові слабкі границі послідовностей (x_{m_k}) та (y_{m_k}) належать множині $VI(A, C)$.

Нехай $z_{m_{k_j}} \rightharpoonup w$. Тоді $y_{m_{k_j}} \rightharpoonup w$ та $w \in VI(A, C)$. З (12.39) випливає

$$(z_{m_k} - a, x_{m_k+1} - z) \leq 0 \quad \forall k \geq n_1. \quad (12.40)$$

Запишемо тотожність

$$\|z_{m_k} - x^*\|^2 = (z_{m_k} - a, x_{m_k+1} - x^*) + \\ + (z_{m_k} - a, z_{m_k} - x_{m_k+1}) - (x^* - a, z_{m_k} - x^*).$$

Ураховуючи нерівність (12.40), отримуємо

$$\|z_{m_k} - x^*\|^2 \leq (z_{m_k} - a, z_{m_k} - x_{m_k+1}) - \\ - (x^* - a, z_{m_k} - x^*). \quad (12.41)$$

Оскільки

$$(z_{m_k} - a, z_{m_k} - x_{m_k+1}) \to 0,$$
$$(x^* - a, z_{m_{k_j}} - x^*) \to (x^* - a, w - x^*) \geq 0,$$

то з (12.41) випливає

$$\limsup_{j\to\infty}\left\|z_{m_{k_j}}-x^*\right\|^2\le\limsup_{j\to\infty}\left\{-(x^*-a,z_{m_{k_j}}-x^*)\right\}\le 0.$$

Таким чином,

$$\lim_{j\to\infty}\left\|z_{m_{k_j}}-x^*\right\|=0.$$

З обмеженості (z_{m_k}) та єдиності $x^*=P_{VI(A,C)}a$ випливає

$$\lim_{k\to\infty}\|z_{m_k}-x^*\|=0.$$

Далі маємо

$$\lim_{k\to\infty}\|x_{m_k+1}-x^*\|=0.$$

Ураховуючи умову (iii), отримуємо $\lim_{n\to\infty}\|x_n-x^*\|=0$. Звідси, у свою чергу, випливає $\lim_{n\to\infty}\|y_n-x^*\|=\lim_{n\to\infty}\|z_n-x^*\|=0$. □

Алгоритм 8.

1) Задаємо $x_0\in H$, $\lambda\in\left(0,\frac{1}{L}\right)$.

2) Для x_n *обчислюємо*

$$y_n=P_C\left(x_n-\lambda Ax_n\right),$$
$$z_n=y_n-\lambda\left(Ay_n-Ax_n\right).$$

3) Будуємо напівпростори

$$C_n=\{z\in H:\|z_n-z\|\le\|x_n-z\|\},$$
$$Q_n=\{z\in H:(x_n-z,x_0-x_n)\ge 0\},$$

обчислюємо

$$x_{n+1}=P_{C_n\cap Q_n}x_0,$$

покладаємо $n:=n+1$ *та переходимо на крок 2.*

Теорема 36. *Породжені алгоритмом 8 послідовності* (x_n), (y_n) *та* (z_n) *сильно збігаються до точки* $x^* = P_{VI(A,C)}x_0$.

Доведення. Має місце нерівність

$$\|z_n - z\|^2 \leq \|x_n - z\|^2 - \left(1 - \lambda^2 L^2\right) \|x_n - y_n\|^2, \qquad (12.42)$$

де $z \in VI(A,C)$.

Множини Q_n, C_n — опуклі та замкнені. Нехай $z \in VI(A,C)$. З нерівності (12.42) випливає $z \in C_n$ для всіх $n \geq 0$. Отже, $VI(A,C) \subseteq C_n$ для всіх $n \geq 0$.

Тепер за допомогою математичної індукції покажемо, що для всіх $n \geq 0$ має місце вкладення $VI(A,C) \subseteq C_n \cap Q_n$. Для $n = 0$ маємо $Q_n = H$. Тому $VI(A,C) \subseteq C_0 \cap Q_0$. Нехай для деякого $k \in \mathbb{N}$ маємо $VI(A,C) \subseteq C_k \cap Q_k$. Тоді існує єдина точка $x_{k+1} \in C_k \cap Q_k$ така, що $x_{k+1} = P_{C_k \cap Q_k} x_0$. З $x_{k+1} = P_{C_k \cap Q_k} x_0$ випливає

$$(x_{k+1} - z, x - x_{k+1}) \geq 0 \quad \forall z \in C_k \cap Q_k.$$

Оскільки $VI(A,C) \subseteq C_k \cap Q_k$, то $VI(A,C) \subseteq Q_{k+1}$. Таким чином, $VI(A,C) \subseteq C_{k+1} \cap Q_{k+1}$.

Покажемо, що послідовність (x_n) обмежена. Існує єдина точка $x^* \in VI(A,C)$, така, що $x^* = P_{VI(A,C)}x_0$. З $x_{n+1} = P_{C_n \cap Q_n} x_0$ випливає

$$\|x_{n+1} - x_0\| \leq \|z - x_0\| \quad \forall z \in C_n \cap Q_n.$$

Оскільки $x^* \in VI(A,C) \subseteq C_n \cap Q_n$, то

$$\|x_{n+1} - x_0\| \leq \|x^* - x_0\| \quad \forall n \in \mathbb{N}. \qquad (12.43)$$

Звідси випливає обмеженість (x_n).

Доведемо, що

$$\lim_{n \to \infty} \|x_{n+1} - x_n\| = 0. \qquad (12.44)$$

З $x_{n+1} \in C_n \cap Q_n \subseteq Q_n$ та $x_n = P_{Q_n} x_0$ випливає

$$\|x_{n+1} - x_0\| \geq \|x_n - x_0\| \quad \forall n \in \mathbb{N}.$$

Послідовність $(\|x_n - x_0\|)$ обмежена та неспадна. Тому існує скінченна границя $\lim_{n\to\infty} \|x_n - x_0\|$. З іншого боку, оскільки $x_{n+1} \in Q_n$, то $(x_n - x_{n+1}, x_0 - x_n) \geq 0$ і

$$\begin{aligned}\|x_n - x_{n+1}\|^2 &= \|(x_n - x_0) - (x_{n+1} - x_0)\|^2 = \\ &= \|x_n - x_0\|^2 - 2(x_n - x_0, x_{n+1} - x_0) + \|x_{n+1} - x_0\|^2 = \\ &= \|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2 - 2(x_n - x_{n+1}, x_0 - x_n) \leq \\ &\leq \|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2 .\end{aligned}$$

Звідси випливає (12.44).

Оскільки $x_{n+1} \in C_n$, то $\|z_n - x_{n+1}\| \leq \|x_n - x_{n+1}\|$. Звідси

$$\|z_n - x_n\| \leq 2\|x_n - x_{n+1}\| \to 0. \tag{12.45}$$

Використовуючи нерівність (12.42), отримуємо

$$\begin{aligned}\|x_n - y_n\|^2 &\leq \frac{\|x_n - z\|^2 - \|z_n - z\|^2}{(1 - \lambda^2 L^2)} = \\ &= \frac{(\|x_n - z\| - \|z_n - z\|)(\|x_n - z\| + \|z_n - z\|)}{(1 - \lambda^2 L^2)} \leq \\ &\leq \frac{(\|x_n - z\| + \|z_n - z\|)}{(1 - \lambda^2 L^2)} \|x_n - z_n\| ,\end{aligned}$$

де $z \in VI(A, C)$. З (12.45) випливає

$$\lim_{n\to\infty} \|x_n - y_n\| = 0. \tag{12.46}$$

Розглянемо довільну підпослідовність (x_{n_k}), що слабко збігається до деякої точки $w \in H$. Покажемо, що $w \in VI(A, C)$. З (12.46) випливає $y_{n_k} \rightharpoonup w$ та $w \in C$. Маємо

$$(y_{n_k} - x_{n_k} + \lambda A x_{n_k}, x - y_{n_k}) \geq 0 \quad \forall x \in C.$$

Звідси

$$(y_{n_k} - x_{n_k}, x - y_{n_k}) + \lambda (A x_{n_k}, x_{n_k} - y_{n_k}) + \\ + \lambda (A x, x - x_{n_k}) \geq 0 \quad \forall x \in C.$$

Після граничного переходу отримаємо $(Ax, x - w) \geq 0 \; \forall x \in C$, тобто $w \in VI(A, C)$.

Для $x^* = P_{VI(A,C)} x_0$ з нерівності (12.43) випливає

$$\|x_0 - x^*\| \leq \|x_0 - w\| \leq \liminf_{k \to \infty} \|x_0 - x_{n_k}\| \leq \\ \leq \limsup_{k \to \infty} \|x_0 - x_{n_k}\| \leq \|x_0 - x^*\| .$$

Отримали $\lim_{k\to\infty} \|x_0 - x_{n_k}\| = \|x_0 - w\| = \|x_0 - x^*\|$. Звідси $x_{n_k} \to w = x^*$. Отже, $x_n \to x^*$. З (12.45) і (12.46) випливає $z_n \to x^*$ та $y_n \to x^*$, що і треба було довести. □

12.5. Метод Попова

Певною альтернативою методу Г. М. Корпелевич є метод Л. Д. Попова, опублікований у 1980 р.

Алгоритм 9 ([19]).

1) Задаємо $x_0 = y_0 \in C$, $\lambda \in \left(0, \frac{1}{3L}\right)$.

2) Для x_n *та* y_n *обчислюємо*

$$x_{n+1} = P_C (x_n - \lambda A y_n) , \\ y_{n+1} = P_C (x_{n+1} - \lambda A y_n) .$$

3) Якщо $x_n = y_n = x_{n+1}$, *то СТОП, інакше покладаємо* $n := n + 1$ *та переходимо на крок 2.*

Лема 22. *Для $z \in VI(A,C)$ і породжених алгоритмом 9 послідовностей (x_n), (y_n) виконується нерівність*

$$\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - (1-\lambda L)\|x_n-y_n\|^2 - \\ -(1-2\lambda L)\|x_{n+1}-y_n\|^2 + \lambda L\|x_n-y_{n-1}\|^2. \quad (12.47)$$

Доведення. Маємо

$$\|x_{n+1}-z\|^2 \le \|x_n-\lambda Ay_n-z\|^2 - \\ -\|x_n-\lambda Ay_n-x_{n+1}\|^2 = \|x_n-z\|^2 - \\ -\|x_{n+1}-x_n\|^2 - 2\lambda(Ay_n, x_{n+1}-z). \quad (12.48)$$

До правої частини (12.48) додамо $2\lambda(Ay_n, y_n-z) \ge 0$. Отримаємо

$$\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - \|x_{n+1}-x_n\|^2 - \\ -2\lambda(Ay_n, x_{n+1}-y_n) = \|x_n-z\|^2 - \|y_n-x_n\|^2 -$$

$$-\|x_{n+1}-y_n\|^2 - 2(x_n-y_n, y_n-x_{n+1}) - \\ -2\lambda(Ay_n, x_{n+1}-y_n) = \|x_n-z\|^2 - \|y_n-x_n\|^2 - \\ -\|x_{n+1}-y_n\|^2 + 2\lambda(Ay_{n-1}-Ay_n, x_{n+1}-y_n) + \\ +2(x_n-\lambda Ay_{n-1}-y_n, x_{n+1}-y_n). \quad (12.49)$$

Оцінимо четвертий та п'ятий доданки в правій частині (12.49). Почнемо з п'ятого. Оскільки $x_{n+1} \in C$, то

$$(x_n-\lambda Ay_{n-1}-y_n, x_{n+1}-y_n) \le 0. \quad (12.50)$$

Перейдемо до оцінки $(Ay_{n-1}-Ay_n, x_{n+1}-y_n)$. Маємо

$$2\lambda(Ay_{n-1}-Ay_n, x_{n+1}-y_n) \le$$

$$\leq 2\lambda L \|y_{n-1} - y_n\| \|x_{n+1} - y_n\| \leq$$
$$\leq 2\lambda L \left(\|y_n - x_n\| + \|x_n - y_{n-1}\|\right) \|x_{n+1} - y_n\| \leq$$
$$\leq \lambda L \Big(\|y_n - x_n\|^2 + \|x_n - y_{n-1}\|^2 + 2\|x_{n+1} - y_n\|^2\Big). \quad (12.51)$$

Використовуючи (12.50), (12.51) у (12.49), отримуємо нерівність

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda L)\|x_n - y_n\|^2 - (1 - 2\lambda L)\|x_{n+1} - y_n\|^2 + \lambda L \|x_n - y_{n-1}\|^2,$$

чим і завершуємо доведення. □

Має місце

Теорема 37. *Породжені алгоритмом 9 послідовності* (x_n), (y_n) *слабко збігаються до розв'язку (12.1).*

Доведення. Доведемо обмеженість послідовності (x_n). Зафіксуємо номер $N \in \mathbb{N}$ та розглянемо нерівності (12.47) для всіх номерів N, $N+1$, ..., M, де $M > N$. Додавши їх, отримаємо

$$\|x_{M+1} - z\|^2 \leq \|x_N - z\|^2 - (1 - \lambda L)\sum_{n=N}^{M} \|x_n - y_n\|^2 - (1 - 3\lambda L)\sum_{n=N}^{M} \|x_{n+1} - y_n\|^2 + \lambda L \|x_N - y_{N-1}\|^2. \quad (12.52)$$

Звідси випливає обмеженість послідовності (x_n). Крім того, з нерівності (12.52) отримуємо збіжність рядів $\sum_n \|x_n - y_n\|^2$ та $\sum_n \|x_{n+1} - y_n\|^2$. Таким чином,

$$\lim_{n\to\infty} \|x_n - y_n\| = \lim_{n\to\infty} \|x_{n+1} - y_n\| = 0.$$

Покажемо, що всі слабкі часткові границі послідовності (x_n) належать множині $VI(A,C)$. Нехай підпослідовність (x_{n_k}) слабко збігається до $x^* \in H$. Очевидно, що $x^* \in C$ та $y_{n_k} \rightharpoonup x^*$. Маємо

$$(y_{n_k+1} - x_{n_k+1} + \lambda A y_{n_k}, y - y_{n_k+1}) \geq 0 \quad \forall y \in C.$$

Звідси для всіх $y \in C$ отримуємо

$$\begin{aligned} 0 \leq (y_{n_k+1} - x_{n_k+1}, y - y_{n_k+1}) + \lambda\,(Ay_{n_k}, y_{n_k} - y_{n_k+1}) + \\ + \lambda\,(Ay_{n_k}, y - y_{n_k}) \leq (y_{n_k+1} - x_{n_k+1}, y - y_{n_k+1}) + \\ + \lambda\,(Ay_{n_k}, y_{n_k} - x_{n_k+1}) + \lambda\,(Ay_{n_k}, x_{n_k+1} - y_{n_k+1}) + \\ + \lambda\,(Ay, y - y_{n_k})\,. \quad (12.53) \end{aligned}$$

Перейшовши до границі при $k \to \infty$ у (12.53), отримаємо $(Ay, y - x^*) \geq 0\ \forall y \in C$, тобто $x^* \in VI(A,C)$.

Покажемо, що послідовність (x_n) слабко збігається. Оберемо довільне $z \in VI(A,C)$. Оскільки $1 - 2\lambda L \geq \lambda L$, то з нерівності (12.47) випливає

$$\begin{aligned} \|x_{n+1} - z\|^2 + \lambda L \|x_{n+1} - y_n\|^2 \leq \\ \leq \|x_n - z\|^2 + \lambda L \|x_n - y_{n-1}\|^2\,. \end{aligned}$$

Отже, існує границя послідовності

$$\left(\|x_n - z\|^2 + \lambda L \|x_n - y_{n-1}\|^2\right).$$

А з рівності $\lim_{n\to\infty} \|x_n - y_{n-1}\| = 0$ випливає збіжність числової послідовності $(\|x_n - z\|)$. Припустимо тепер, що послідовність (x_n) має принаймні дві різні слабкі часткові границі p та q. За доведеним $p, q \in VI(A,C)$. Нехай (x_{n_k}), (x_{m_k}) — підпослідовності, що слабко збігаються до p, q, відповідно. Тоді з

леми Оп'яла випливає

$$\lim_{n\to\infty}\|x_n-p\|=\lim_{k\to\infty}\|x_{n_k}-p\|<\lim_{k\to\infty}\|x_{n_k}-q\|=$$
$$=\lim_{k\to\infty}\|x_{m_k}-q\|<\lim_{k\to\infty}\|x_{m_k}-p\|=\lim_{n\to\infty}\|x_n-p\|.$$

Отримали абсурдну нерівність. Отже, послідовність (x_n) слабко збігається до точки з множини $VI(A,C)$. □

12.6. Субградієнтний екстраградієнтний метод

Припустимо, що оператор $A:H\to H$ монотонний на множині C та ліпшицевий на H.

Алгоритм 10 ([26, 29])**.**

1) Задаємо $x_0\in C$, $\lambda\in\left(0,\frac{1}{L}\right)$.

2) Для x_n *обчислюємо*

$$y_n=P_C\left(x_n-\lambda Ax_n\right).$$

3) Якщо $x_n=y_n$, *то СТОП, інакше будуємо півпростір*

$$T_n=\{z\in H:\,(x_n-\lambda Ax_n-y_n,z-y_n)\le 0\}$$

та обчислюємо

$$x_{n+1}=P_{T_n}\left(x_n-\lambda Ay_n\right),$$

покладаємо $n:=n+1$ *та переходимо на крок 2.*

Маємо $C\subseteq T_n$. Дійсно, якщо припустити існування точки $z\in C\setminus T_n$, то нерівність $(x_n-\lambda Ay_n-y_n,z-y_n)>0$ буде суперечити рівності $y_n=P_C\left(x_n-\lambda Ay_n\right)$.

Ясно, що коли $x_n=y_n$, то $x_n\in VI(A,C)$.

Лема 23. *Для $z \in VI(A,C)$ і породжених алгоритмом 10 послідовностей (x_n), (y_n) виконується нерівність*

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda^2 L^2) \|x_n - y_n\|^2 . \qquad (12.54)$$

Доведення. Аналогічно доведенню леми 18 отримуємо

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2 (x_n - \lambda A y_n - y_n, x_{n+1} - y_n) .$$

Оскільки $x_{n+1} \in T_n$, то

$$(x_n - \lambda A y_n - y_n, x_{n+1} - y_n) = \\ = (x_n - \lambda A x_n - y_n, x_{n+1} - y_n) + \\ + \lambda (A x_n - A y_n, x_{n+1} - y_n) \leq \lambda (A x_n - A y_n, x_{n+1} - y_n) .$$

Отже,

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda (A x_n - A y_n, x_{n+1} - y_n) . \quad (12.55)$$

Перейдемо до оцінки $2\lambda (A x_n - A y_n, x_{n+1} - y_n)$. Маємо

$$2\lambda (A x_n - A y_n, x_{n+1} - y_n) \leq \\ \leq 2\lambda L \|x_n - y_n\| \|x_{n+1} - y_n\| \leq \\ \leq \lambda^2 L^2 \|x_n - y_n\|^2 + \|x_{n+1} - y_n\|^2 . \quad (12.56)$$

Використовуючи (12.56) у (12.55), отримуємо нерівність

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda^2 L^2) \|x_n - y_n\|^2 ,$$

чим і завершуємо доведення. □

Теорема 38. *Породжені алгоритмом 10 послідовності (x_n), (y_n) слабко збігаються до розв'язку (12.1).*

Доведення. З нерівності (12.54) випливає існування $\lim_{n\to\infty}\|x_n - z\|$ для всіх $z \in VI(A,C)$. Зокрема, послідовність (x_n) обмежена. Запишемо нерівність (12.54) у вигляді

$$(1-\lambda^2L^2)\|x_n - y_n\|^2 \le \|x_n - z\|^2 - \|x_{n+1} - z\|^2.$$

Маємо

$$\sum_{n=0}^{m}\|x_n - y_n\|^2 \le \frac{\|x_0 - z\|^2}{1-\lambda^2L^2} \quad \forall m \in \mathbb{N}.$$

Звідси $\lim_{n\to\infty}\|x_n - y_n\| = 0$. Отже, послідовность (y_n) також обмежена.

Покажемо, що всі слабкі часткові границі послідовності (x_n) належать множині $VI(A,C)$. Нехай підпослідовність (x_{n_k}) слабко збігається до $x^* \in H$. Очевидно, що $y_{n_k} \rightharpoonup x^*$ та $x^* \in C$. Маємо

$$(y_{n_k} - x_{n_k} + \lambda Ax_{n_k}, x - y_{n_k}) \ge 0 \quad \forall x \in C.$$

Звідси для всіх $x \in C$ отримуємо

$$\begin{aligned} 0 \le (y_{n_k} - x_{n_k}, x - y_{n_k}) + \lambda\,(Ax_{n_k}, x_{n_k} - y_{n_k}) + \\ + \lambda\,(Ax_{n_k}, x - x_{n_k}) \le (y_{n_k} - x_{n_k}, x - y_{n_k}) + \\ + \lambda\,(Ax_{n_k}, x_{n_k} - y_{n_k}) + \lambda\,(Ax, x - x_{n_k}). \end{aligned} \quad (12.57)$$

Перейшовши до границі при $k \to \infty$ у (12.57), отримаємо

$$(Ax, x - x^*) \ge 0 \quad \forall x \in C,$$

тобто $x^* \in VI(A,C)$.

За допомогою леми Оп'яла доводимо, що послідовність (x_n) слабко збігається до точки з множини $VI(A,C)$ (тоді з $\|x_n - y_n\| \to 0$ випливає слабка збіжність (y_n) до тієї самої границі). □

12.7. Метод Маліцького – Семенова

Розглянемо

Алгоритм 11 ([30]).

1) Задаємо x_0, y_0, $y_{-1} \in C$, $\lambda \in \left(0, \frac{1}{3L}\right)$.

2) Для x_n, y_n *та* y_{n-1} *будуємо півпростір*

$$T_n = \{z \in H : (x_n - \lambda A y_{n-1} - y_n, z - y_n) \leq 0\}$$

та обчислюємо

$$x_{n+1} = P_{T_n}(x_n - \lambda A y_n),$$
$$y_{n+1} = P_C(x_{n+1} - \lambda A y_n).$$

3) Якщо $x_{n+1} = y_n = y_{n+1}$, *то СТОП, інакше покладаємо* $n := n + 1$ *та переходимо на крок 2.*

Перш за все зазначимо, що $C \subseteq T_n$. Дійсно, якщо припустити існування точки $z \in C \setminus T_n$, то нерівність $(x_n - \lambda A y_{n-1} - y_n, z - y_n) > 0$ буде суперечити факту $y_n = P_C(x_n - \lambda A y_{n-1})$.

Перейдемо до доведення збіжності алгоритму.

Лема 24. *Для* $z \in VI(A, C)$ *і породжених алгоритмом 11 послідовностей* (x_n), (y_n) *виконується нерівність*

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda L)\|x_n - y_n\|^2 - (1 - 2\lambda L)\|x_{n+1} - y_n\|^2 + \lambda L \|x_n - y_{n-1}\|^2.$$

Доведення. Оскільки $z \in VI(A, C) \subseteq T_n$, то з $x_{n+1} = P_{T_n}(x_n - \lambda A y_n)$ випливає

$$\|x_{n+1} - z\|^2 \leq \|x_n - \lambda A y_n - z\|^2 - \|x_n - \lambda A y_n - x_{n+1}\|^2 =$$
$$= \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 - 2\lambda (A y_n, x_{n+1} - z). \qquad (12.58)$$

До правої частини (12.58) додамо $2\lambda\left(Ay_n, y_n - z\right) \geq 0$. Отримаємо

$$\begin{aligned}\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 - \\ - 2\lambda\left(Ay_n, x_{n+1} - y_n\right) = \|x_n - z\|^2 - \|y_n - x_n\|^2 - \\ - \|x_{n+1} - y_n\|^2 - 2(x_n - y_n, y_n - x_{n+1}) - \\ - 2\lambda\left(Ay_n, x_{n+1} - y_n\right) = \|x_n - z\|^2 - \|y_n - x_n\|^2 - \\ - \|x_{n+1} - y_n\|^2 + 2\lambda\left(Ay_{n-1} - Ay_n, x_{n+1} - y_n\right) + \\ + 2\left(x_n - \lambda Ay_{n-1} - y_n, x_{n+1} - y_n\right). \quad (12.59)\end{aligned}$$

Оцінимо четвертий та п'ятий доданки в правій частині (12.59). Почнемо з п'ятого. Оскільки $x_{n+1} \in T_n$, то

$$\left(x_n - \lambda Ay_{n-1} - y_n, x_{n+1} - y_n\right) \leq 0. \quad (12.60)$$

Перейдемо до оцінки $\left(Ay_{n-1} - Ay_n, x_{n+1} - y_n\right)$. Маємо

$$\begin{aligned}2\lambda\left(Ay_{n-1} - Ay_n, x_{n+1} - y_n\right) \leq \\ \leq 2\lambda L\left\|y_{n-1} - y_n\right\|\left\|x_{n+1} - y_n\right\| \leq \\ \leq 2\lambda L\left(\left\|y_n - x_n\right\| + \left\|x_n - y_{n-1}\right\|\right)\left\|x_{n+1} - y_n\right\| \leq \\ \leq \lambda L\left(\left\|y_n - x_n\right\|^2 + \left\|x_n - y_{n-1}\right\|^2 + \right. \\ \left. + 2\left\|x_{n+1} - y_n\right\|^2\right). \quad (12.61)\end{aligned}$$

Використовуючи (12.60), (12.61) у (12.59), отримуємо нерівність

$$\begin{aligned}\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda L)\|x_n - y_n\|^2 - \\ - (1 - 2\lambda L)\|x_{n+1} - y_n\|^2 + \lambda L\|x_n - y_{n-1}\|^2,\end{aligned}$$

чим і завершуємо доведення. □

Теорема 39. *Породжені алгоритмом 11 послідовності* (x_n), (y_n) *слабко збігаються до розв'язку (12.1).*

Доведення. Аналогічне доведенню теореми 37. □

12.8. Метод проектування з відбиттям

У цьому підрозділі припустимо, що оператор $A : H \to H$ монотонний та ліпшицевий на всьому просторі H.

Алгоритм 12.

1) Задаємо $x_0 = y_0 \in C$, $\lambda \in \left(0, \frac{1}{3L}\right]$.

2) Для x_n *та* y_n *обчислюємо*

$$x_{n+1} = P_C\left(x_n - \lambda A y_n\right).$$

3) Якщо $x_n = y_n = x_{n+1}$, *то СТОП, інакше обчислюємо*

$$y_{n+1} = 2x_{n+1} - x_n,$$

покладаємо $n := n + 1$ *та переходимо на крок 2.*

Зауваження 16. Цей метод запропонував у 2014 р. аспірант Ю. В. Маліцький. Точка y_{n+1} є результатом дзеркального відбиття точки x_n від гіперплощини $T = \{y \in H : (x_{n+1} - x_n, y - x_{n+1}) = 0\}$.

Лема 25. *Для* $z \in VI(A, C)$ *і породжених алгоритмом 12 послідовностей* (x_n), (y_n) *виконується нерівність*

$$\|x_{n+1} - z\|^2 \le \|x_n - z\|^2 - (1 - \lambda L)\|x_n - x_{n-1}\|^2 - \\ - (1 - 2\lambda L)\|x_{n+1} - y_n\|^2 + \lambda L\|x_n - y_{n-1}\|^2 - \\ - 2\lambda\left(Az, y_n - z\right). \quad (12.62)$$

Доведення. Маємо

$$\|x_{n+1}-z\|^2 \le \|x_n-\lambda Ay_n-z\|^2 - \|x_n-\lambda Ay_n-x_{n+1}\|^2 = \\ = \|x_n-z\|^2 - \|x_{n+1}-x_n\|^2 - 2\lambda\,(Ay_n, x_{n+1}-z)\,. \quad (12.63)$$

До правої частини нерівності (12.63) додамо $2\lambda\,(Ay_n - Az, y_n - z) \ge 0$. Отримаємо

$$\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - \|x_{n+1}-x_n\|^2 + \\ + 2\lambda\,(Ay_n, y_n - x_{n+1}) - 2\lambda(Az, y_n - z) = \|x_n-z\|^2 - \\ - \|x_{n+1}-x_n\|^2 + 2\lambda\,(Ay_n - Ay_{n-1}, y_n - x_{n+1}) + \\ + 2\lambda\,(Ay_{n-1}, y_n - x_{n+1}) - 2\lambda(Az, y_n - z). \quad (12.64)$$

Оцінимо третій та четвертий доданки в правій частині (12.64). Почнемо з четвертого. Оскільки $x_{n+1},\ x_{n-1} \in C$, то

$$(x_n - x_{n-1} + \lambda Ay_{n-1}, x_n - x_{n+1}) \le 0,$$

$$(x_n - x_{n-1} + \lambda Ay_{n-1}, x_n - x_{n-1}) \le 0.$$

Склавши ці нерівності, отримаємо

$$(x_n - x_{n-1} + \lambda Ay_{n-1}, y_n - x_{n+1}) \le 0,$$

звідки випливає оцінка

$$2\lambda\,(Ay_{n-1}, y_n - x_{n+1}) \le 2\,(x_n - x_{n-1}, x_{n+1} - y_n) = \\ = \|x_{n+1}-x_n\|^2 - \|x_n-y_n\|^2 - \|x_{n+1}-y_n\|^2\,, \quad (12.65)$$

оскільки $x_n - x_{n-1} = y_n - x_n$.

Перейдемо до оцінки $(Ay_n - Ay_{n-1}, y_n - x_{n+1})$. Маємо

$$2\lambda\,(Ay_n - Ay_{n-1}, y_n - x_{n+1}) \le$$

$$\le 2\lambda L\,\|y_n - y_{n-1}\|\,\|y_n - x_{n+1}\| \le$$

$$\leq 2\lambda L\left(\|y_n - x_n\| + \|x_n - y_{n-1}\|\right)\|y_n - x_{n+1}\| \leq$$
$$\leq \lambda L\left(\|y_n - x_n\|^2 + \|x_n - y_{n-1}\|^2 + \right.$$
$$\left. +2\|y_n - x_{n+1}\|^2\right). \quad (12.66)$$

Використовуючи (12.65), (12.66) у (12.64), отримуємо нерівність

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda L)\|y_n - x_n\|^2 -$$
$$- (1 - 2\lambda L)\|x_{n+1} - y_n\|^2 + \lambda L\|x_n - y_{n-1}\|^2 -$$
$$- 2\lambda (Az, y_n - z),$$

чим і завершуємо доведення. □

Має місце

Теорема 40. *Породжені алгоритмом 12 послідовності* (x_n), (y_n) *слабко збігаються до розв'язку (12.1).*

Доведення. Доведемо обмеженість послідовності (x_n). Оберемо $z \in VI(A, C)$. Оскільки $1 - 2\lambda L \geq \lambda L$ та

$$(Az, y_n - z) = 2(Az, x_n - z) - (Az, x_{n-1} - z) \geq$$
$$\geq (Az, x_n - z) - (Az, x_{n-1} - z),$$

то з нерівності (12.62) випливає

$$\|x_{n+1} - z\|^2 + \lambda L\|x_{n+1} - y_n\|^2 + 2\lambda (Az, x_n - z) \leq$$
$$\leq \|x_n - z\|^2 + \lambda L\|x_n - y_{n-1}\|^2 + 2\lambda (Az, x_{n-1} - z) -$$
$$- (1 - \lambda L)\|x_n - x_{n-1}\|^2. \quad (12.67)$$

Покладемо

$$a_n = \|x_n - z\|^2 + \lambda L\|x_n - y_{n-1}\|^2 + 2\lambda (Az, x_{n-1} - z),$$
$$b_n = (1 - \lambda L)\|x_n - x_{n-1}\|^2.$$

Тоді нерівність (12.67) можна записати у формі $a_{n+1} \leq a_n - b_n$. За лемою 8 послідовність (a_n) має границю та $\lim_{n\to\infty} \|x_n - x_{n-1}\| = 0$. Таким чином, послідовність $(\|x_n - z\|)$ обмежена, звідки випливає обмеженість (x_n).

З нерівності

$$\|x_{n+1} - y_n\| \leq \|x_{n+1} - x_n\| + \|x_n - y_n\| = \\ = \|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|$$

випливає $\lim_{n\to\infty} \|x_{n+1} - y_n\| = 0$.

Покажемо, що всі слабкі часткові границі послідовності (x_n) належать множині $VI(A, C)$. Нехай підпослідовність (x_{n_k}) слабко збігається до $x^* \in H$. Очевидно, що $x^* \in C$ та $y_{n_k} \rightharpoonup x^*$. Маємо

$$(x_{n_k+1} - x_{n_k} + \lambda A y_{n_k}, y - x_{n_k+1}) \geq 0 \quad \forall y \in C.$$

Звідси для всіх $y \in C$ отримуємо

$$0 \leq (x_{n_k+1} - x_{n_k}, y - x_{n_k+1}) + \lambda\,(Ay_{n_k}, y - y_{n_k}) +$$

$$+ \lambda\,(Ay_{n_k}, y_{n_k} - x_{n_k+1}) \leq (x_{n_k+1} - x_{n_k}, y - x_{n_k+1}) + \\ + \lambda\,(Ay, y - y_{n_k}) + \lambda\,(Ay_{n_k}, y_{n_k} - x_{n_k+1}). \quad (12.68)$$

Перейшовши до границі при $k \to \infty$ у (12.68), отримаємо

$$(Ay, y - x^*) \geq 0 \quad \forall y \in C,$$

тобто $x^* \in VI(A, C)$.

Покажемо, що послідовність (x_n) слабко збігається. Для довільного $z \in VI(A, C)$ числова послідовність $\left(\|x_n - z\|^2 + 2\lambda\,(Az, x_n - z)\right)$ має границю. Припустимо, що послідовність (x_n) має принаймні дві різні слабкі часткові границі p та q. За доведеним $p,\ q \in VI(A, C)$. Нехай (x_{n_k}) — під-

послідовність, що слабко збігається до p. Тоді з леми Оп'яла випливає

$$\begin{aligned}
&\lim_{n\to\infty}\left(\|x_n-p\|^2+2\lambda\,(Ap,x_n-p)\right)=\\
&=\lim_{k\to\infty}\left(\|x_{n_k}-p\|^2+2\lambda\,(Ap,x_{n_k}-p)\right)=\lim_{k\to\infty}\|x_{n_k}-p\|^2<\\
&\qquad<\liminf_{k\to\infty}\|x_{n_k}-q\|^2\le\liminf_{k\to\infty}\|x_{n_k}-q\|^2+\\
&\qquad\qquad+\liminf_{k\to\infty}2\lambda\,(Aq,x_{n_k}-q)\le\\
&\qquad\le\liminf_{k\to\infty}\left(\|x_{n_k}-q\|^2+2\lambda\,(Aq,x_{n_k}-q)\right)=\\
&\qquad\qquad=\lim_{n\to\infty}\left(\|x_n-q\|^2+2\lambda\,(Aq,x_n-q)\right).
\end{aligned}$$

Повторивши це міркування, отримуємо абсурдну нерівність

$$\begin{aligned}
&\lim_{n\to\infty}\left(\|x_n-p\|^2+2\lambda\,(Ap,x_n-p)\right)<\\
&\qquad\qquad<\lim_{n\to\infty}\left(\|x_n-p\|^2+2\lambda\,(Ap,x_n-p)\right).
\end{aligned}$$

Отже, послідовність (x_n) слабко збігається до точки з множини $VI(A,C)$. □

Література

[1] **Айзерман М. А.** Метод потенциальных функций в теории обучения машин / М. А. Айзерман, Э. М. Браверман, Л. И. Розоноэр. – М.: Наука, 1970.

[2] **Байоки К.** Вариационные и квазивариационные неравенства / К. Байоки, А. Капело. – М.: Мир, 1988.

[3] **Бакушинский А. Б.** Некорректные задачи. Численные методы и приложения / А. Б. Бакушинский, А. В. Гончарский. – М.: Изд-во МГУ, 1989.

[4] **Воронцов К. В.** Машинное обучение: курс лекций [Електронний ресурс] / К. В. Воронцов. – М.: ВМиК МГУ. – 2009. – Режим доступу: www.machinelearning.com

[5] **Вапник В. Н.** Теория распознавания образов / В. Н. Вапник, А. Я. Червоненкис. – М.: Наука, 1974.

[6] **Главач В.** Десять лекций по статистическому и структурному распознаванию образов / В. Главач, М. И. Шлезингер. – К.: Наукова думка, 2004.

[7] **Дуда Р.** Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – М.: Мир, 1976.

[8] Математические методы исследования операций / Ю. М. Ермольев, И. И. Ляшко, В. С. Михалевич, В. И. Тюптя. – К: Вища школа, 1979.

[9] **Киндерлерер Д.** Введение в вариационные неравенства и их приложения / Д. Киндерлерер, Г. Стампаккья. – М.: Мир, 1983.

[10] **Клюшин Д. А.** Доказательная медицина: применение статистических методов / Д. А. Клюшин, Ю. И. Петунин. – М.: Вильямс, 2008.

[11] **Корпелевич Г. М.** Экстраградиентный метод для отыскания седловых точек и других задач / Г. М. Корпелевич // Экономика и матем. методы. – 1976. – Т. 12, № 4. – С. 747–756.

[12] **Лепский А. Е.** Математические методы распознавания образов: курс лекций / А. Е. Лепский, А. Г. Броневич. – Таганрог: Южный федеральный университет, 2009.

[13] **Лионс Ж.-Л.** Некоторые методы решения нелинейных краевых задач / Ж.-Л. Лионс. – М.: Мир, 1972. – 587 с.

[14] **Мерков А. Б.** Распознавание образов. Введение в методы статистического обучения / А. Б. Мерков. – М.: Едиториал УРСС, 2011.

[15] **Местецкий Л. М.** Математические методы распознавания образов / Л. М. Местецкий. – М.: Изд-во МГУ, 2004.

[16] **Минский М.** Персептроны / М. Минский, С. Пейперт. – М.: Мир, 1971.

[17] **Маннинг К.** Введение в информационный поиск / К. Маннинг, П. Рагхаван, Х. Шютце. – М: Вильямс, 2011.

[18] **Обен Ж. П.** Прикладной нелинейный анализ / Ж. П. Обен, И. Экланд. – М.: Мир, 1988.

[19] **Попов Л. Д.** Модификация метода Эрроу – Гурвица поиска седловых точек / Л. Д. Попов // Математические заметки. — 1980. – Т. 28, № 5. – С. 777–784.

[20] **Семенов В. В.** Явный алгоритм расщепления для вариационных неравенств с монотонными операторами / В. В. Семенов // Журн. обчисл. та прикл. матем. – 2013. – № 2 (112). – С. 42–52.

[21] **Уоссермен Ф.** Нейрокомпьютерная техника: теория и практика / Ф. Уоссермен. – М.: Мир, 1992.

[22] **Фукунага К.** Введение в статистическую теорию распознавания образов / К. Фукунага. – М.: Наука, 1979.

[23] **Хайкин С.** Нейронные сети: полный курс / С. Хайкин. – М.: Вильямс, 2006.

[24] **Abe S.** Support vector machines for pattern classification / S. Abe. – London: Springer, 2010.

[25] **Bauschke H. H.** Convex analysis and monotone operator theory in Hilbert spaces / H. H. Bauschke, P. L. Combettes. – New York: Springer, 2011.

[26] **Censor Y.** The subgradient extragradient method for solving variational inequalities in Hilbert space / Y. Censor, A. Gibali, S. Reich // J. of Optimization Theory and Applications. – 2011. – Vol. 148. – P. 318–335.

[27] **Facchinei F.** Finite-Dimensional Variational Inequalities and Complementarity Problem / F. Facchinei, J.-S. Pang. – New York: Springer, 2003. – Vol. 2.

[28] **Fisher R. A.** The Use of Multiple Measurements in Taxonomic Problems / R. A. Fisher // Annals of Eugenics. – 1936. – Vol. 7. – С. 179–188.

[29] **Lyashko S. I.** Low-cost modification of Korpelevich's methods for monotone equilibrium problems / S. I. Lyashko, V. V. Semenov, T. A. Voitova // Cybernetics and Systems Analysis. – 2011. – Vol. 47. — P. 631–639.

[30] **Malitsky Yu. V.** An extragradient algorithm for monotone variational inequalities / Yu. V. Malitsky, V. V. Semenov // Cybernetics and Systems Analysis. – 2014. – Vol. 50. – P. 271–277.

[31] **Mainge P.-E.** Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization / P.-E. Mainge // Set–Valued Analysis. – 2008. – Vol. 16. – P. 899–912.

[32] Fisher discriminant analysis with kernels / S. Mika, G. Raetsch, J. Weston et al. // Neural Networks for Signal Processing IX. — NJ: IEEE, 1999. — P. 41–48.

[33] **Nadezhkina N.** Strong convergence theorem by a hybrid method for nonexpansive mappings and Lipschitz-continuous monotone mappings / N. Nadezhkina, W. Takahashi // SIAM J. Optim. – 2006. – Vol. 16. – P. 1230–1241.

[34] **Opial Z.** Weak convergence of the sequence of successive approximations for nonexpansive mappings / Z. Opial // Bull. Amer. Math. Soc. – 1967. – Vol. 73. – P. 591–597.

[35] **Passty G. B.** Ergodic convergence to a zero of the sum of monotone operators in Hilbert spaces / G. B. Passty // J. of Mathematical Analysis and Applications. – 1979. – Vol. 72. – P. 383–390.

[36] **Robbins H.** A stochastic approximation method / H. Robbins, S. Monro // Ann. Math. Statist. – 1951. – Vol. 51. – P.400–407.

[37] **Tseng P.** A modified forward-backward splitting method for maximal monotone mappings / P. Tseng // SIAM J. Control Optim. – 2000. – Vol. 38. – P. 431–446.

[38] **Webb A.** Statistical Pattern Recognition / A. Webb. – NY: John Wiley and Sons, 2002.

ЗМІСТ