Київський національний університет імені Тараса Шевченка

1834

СУБГРАДІЄНТНІ ТА ЕКСТРАГРАДІЄНТНІ АЛГОРИТМИ

ЧАСТИНА 1

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

СУБГРАДІЄНТНІ ТА ЕКСТРАГРАДІЄНТНІ АЛГОРИТМИ

Монографія

Частина 1

КИЇВСЬКИЙ УНІВЕРСИТЕТ
2024

УДК 519.8+519.6+004.424
С89

Автори:
В. В. Семенов (розд. 1, 3–7), П. І. Стецюк (розд. 1, 2),
В. О. Стовба (розд. 1, 2), А. А. Супрун (розд. 2),
О. С. Харьков (розд. 4, 6), О. Ю. Коваленко (розд. 5, 7)

Рецензенти:
чл.-кор. НАН України, д-р фіз.-мат. наук, проф. П. С. Кнопов
(*Інститут кібернетики імені В.М. Глушкова НАН України*);
чл.-кор. НАН України, д-р фіз.-мат. наук, проф. О. М. Кісельова
(*Дніпровський національний університет імені Олеся Гончара*)

Рекомендовано до друку Вченою радою
Інституту кібернетики імені В.М. Глушкова НАН України
(протокол № 13 від 10 вересня 2024 року)

С89 Субградієнтні та екстраградієнтні алгоритми : монографія : у 2-х ч. / В. В. Семенов, П. І. Стецюк, В. О. Стовба та ін. – К. : ВПЦ "Київський університет", 2024. – Ч. 1. – 201 с.

ISBN 978-966-933-292-9 (загальний)
ISBN 978-966-933-293-6 (Ч.1)

Книга складається з нарисів, що присвячені алгоритмам розв'язання задач опуклого програмування, варіаційних нерівностей та загальних задач про рівновагу.

На основі однорангових лінійних операторів побудовано методи фейєрівського типу. Досліджено В-форму методів квазіньютонівського типу, яка дозволяє легко інтерпретувати ці методи, як градієнтні в перетвореному просторі. Розглянуто модифіковані екстраградієнтні методи з динамічним регулюванням величини кроку. Описано нові проксимальні алгоритми рівноважного програмування.

Для фахівців у галузі математичного програмування, обчислювальної математики, а також студентів та аспірантів відповідних спеціальностей.

Публікація книги здійснена підтримки МОН України
(проєкт № 0122U002026), НАН України (проєкт № 0124U002162)
та Volkswagen Foundation (грант № 97775)

УДК 519.8+519.6+004.424

ISBN 978-966-933-292-9 (загальний)
ISBN 978-966-933-293-6 (Ч.1)

ЗМІСТ

Передмова

Даний том складається з семи нарисів на тему алгоритмів розв'язання задач опуклого програмування (гладких та негладких), варіаційних нерівностей та задач рівноважного програмування.

Авторами є колектив дослідників з Київського національного університету імені Тараса Шевченка та Інституту кібернетики імені В.М. Глушкова НАН України, що виконують сумісний науковий проєкт «Нові субградієнтні та екстраградієнтні методи для негладких задач регресії». Проєкт орієнтований на створення з теоретичним обгрунтуванням нових ефективних субградієнтних та екстраградієнтних методів для негладких задач регресії. Задачі регресійного типу з негладкими функціями продовжують бути одним з основних напрямків робіт у математичному програмуванні та його застосуванні.

Однією з ефективних ідей, що використовують автори, є лінійні перетворення простору, які поліпшують субдиференціальні властивості функцій, що оптимізуються. Субградієнтні алгоритми, що розроблені з використанням цієї ідеї Н.З. Шора[1], залишаються одними з найефективніших для мінімізації негладких функцій і, більш того, складають конкуренцію найкращим методам (квазіньютонівські та спряжених напрямів) для гладких функцій.

Альтернативний підхід полягає у тому, що задачі опуклої недиференційовної оптимізації можна розв'язувати з допомогою їх переформулювання у вигляді сідлових (мінімаксних) задач з подальшим застосуванням алгоритмів для варіаційних нерівностей.

Умовно можна розділити зміст книги на дві частини. Перша складається з розділів, у яких приділено увагу дослідженню якісних властивостей методів з перетворенням простору. Друга — з розділів, присвячених збіжності нових алгоритмів розв'язання варіаційних нерівностей та задач рівноважного програмування.

---

[1] Наум Зуселевич Шор (1.01.1937, Київ — 25.02.2006, Київ) — видатний радянський та український математик, один з творців негладкої оптимізації, академік НАН України. Відомий як автор методу послідовного аналізу варіантів, методу узагальненого градієнтного спуску, субградієнтних методів з розтягом простору. Після закінчення у 1958 році механіко-математичного факультету Київського національного університету імені Тараса Шевченка до кінця життя працював в Інституті кібернетики імені В.М. Глушкова НАН України (до 1962 року — Обчислювальний центр АН УРСР). Створив відділ методів негладкої оптимізації.

У розділі 1 розглянуто два однорангові лінійні оператори перетворення простору. Їх використання дозволяє обґрунтувати збіжність сімейства методів змінної метрики для розв'язання задачі опуклого програмування. На основі цих операторів побудовано субградієнтні методи фейєрівського типу для знаходження мінімуму опуклої функції з відомим оптимальним значенням функції. Наведено числові експерименти з цими методами.

У розділі 2 обговорюється спеціальна форма (B-форма) методів квазіньютонівського типу, яка дозволяє легко інтерпретувати ці методи, як градієнтні в перетвореному відповідним чином просторі аргументів. Наведено B-форму методу Давидона–Флетчера–Пауелла та на її основі проведено порівняння цього методу з r-алгоритмами. Для мінімізації гладких опуклих функцій побудовано градієнтний метод з перетворенням простору, що поєднує властивості як квазіньютонівських методів, так і r-алгоритмів. Обговорюються можливі схеми методів такого типу для мінімізації негладких опуклих функцій.

У розділі 3 розглядаються різні варіанти модифікованого екстраградієнтного методу з динамічним регулюванням величини кроку для розв'язання варіаційних нерівностей з монотонними операторами. Також розглянуто варіанти методу для варіаційних нерівностей та операторних рівнянь з апріорною інформацією про розв'язок. У кінці розділу розглянуто деякі застосування варіаційних нерівностей у машинному навчанні. З появою генеруючих змагальних нейронних мереж (generative adversarial networks, GANs) та інших моделей змагального навчання стійкий інтерес до алгоритмів розв'язання варіаційних нерівностей виник і в середовищі спеціалістів в галузі машинного навчання. Створення швидких та надійних алгоритмів для варіаційних нерівностей (особливо для стохастичних постановок) веде до прогресу в навчанні породжуючих змагальних мереж.

У розділі 4 запропоновано нові модифікації методу Л.Д. Попова. Алгоритми цього типу відомі серед спеціалістів з машинного навчання під назвою «Extrapolation from the Past».

У розділі 5 розглядаються задачі про рівновагу з псевдомонотонними біфункціями в гільбертовому просторі. Для цих задач запропоновано та обґрунтовано новий двоетапний проксимальний алгоритм, який є узагальненням методу екстраполяції з минулого пошуку сідлових точок опукло-угнутих функцій.

У розділі 6 розглядається дворівнева задача: варіаційна нерівність на множині розв'язків задачі про рівновагу. Прикладом такої задачі є пошук нормальної рівноваги Неша. Для розв'язання задачі запропоновано алгоритм, що суміщає у собі ідеї двоетапного проксимального методу та ітеративної регуляризації. Крім того, досліджено адаптивний варіант алгоритму з правилом оновлення параметрів, що не використовує значень ліпшицевих констант біфункції. Для монотонних біфункцій ліпшицевого типу та сильно монотонних ліпшицевих операторів доведено теореми про сильну збіжність алгоритмів.

При розв'язанні складних задач дослідження операцій (наприклад, в моделюванні транспортних та телекомунікаційних мереж) та оптимального керування велике значення мають різні декомпозиційні підходи, що дозволяють зводити розв'язання вихідної задачі до розв'язання послідовності задач більш простої структури. У розділі 7 досліджено один з алгоритмів розщеплення (декомпозиції) для варіаційних нерівностей з багатозначними максимальними монотонними операторами, що діють у гільбертовому просторі.

Чимало матеріалів книги вже знайшли відображення в спеціальних курсах, які викладаються авторами на факультеті комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка, Інституті кібернетики імені В.М. Глушкова НАН України та в Київському академічному університеті.

Сподіваємось, що книга зацікавить читачів, які займаються прикладною математикою, — в першу чергу, наукову молодь.

Робота авторів над книгою була підтримана МОН України (проєкт «Обчислювальні алгоритми і оптимізація для штучного інтелекту, медицини та оборони», 0122U002026), Відділенням цільової підготовки Київського національного університету імені Тараса Шевченка при Національній академії наук України (проєкт «Нові субградієнтні та екстраградієнтні методи для негладких задач регресії», 0124U002162) та Volkswagen Foundation (grant № 97775).

Ми щиро вдячні С.І. Ляшку та І.В. Сергієнку за підримку наших наукових досліджень.

Зауваження та побажання можна надсилати електронною поштою:

`semenov.volodya@gmail.com`, `stetsyukp@gmail.com`.

Розділ 1

Однорангові лінійні перетворення та фейєрівські методи опуклого програмування

Розглянуто два однорангові лінійні оператори перетворення простору. Їх використання дозволяє обґрунтувати збіжність сімейства методів змінної метрики для розв'язання задачі опуклого програмування. На основі цих операторів побудовано субградієнтні методи фейєрівського типу для знаходження мінімуму опуклої функції з відомим оптимальним значенням функції. Наведено числові експерименти з цими методами.

1.1. Методи з розтягом простору та методи змінної метрики

Методам субградієнтного типу властива повільна збіжність при мінімізації опуклих функцій яружного типу. Це пов'язано з тим, що антисубградієнт утворює кут, близький до $\pi/2$, з напрямком до точки мінімуму. Оригінальний вихід із цієї ситуації був уперше усвідомлений та запропонований Н.З. Шором у [1]. Він полягає у використанні лінійних неортогональних перетворень простору (а саме оператора розтягу простору), які дозволяють змінювати кути між субградієнтом та напрямком на мінімум. Зокрема, це приводить до методів субградієнтного типу у перетвореному лінійним оператором просторі аргументів. Саме тому такі методи можна вважати методами змінної метрики. Історично першим методом змінної метрики був запропонований Давидоном у [2] і розвинений Флетчером та Пауеллом у [3] метод для мінімізації двічі неперервно диференційовних функцій. Він базується на ідеї квадратичної апроксимації функції, що мінімізується, та фактично імітує метод Ньютона-Рафсона без явного обчислення других похідних [4]. Внаслідок цього за ним закріпилася назва методу квазіньютонівського типу, і в літературі можна зустріти його інтерпретацію як методу змінної метрики. Це зумовлено також формою запису методу Давидона–Флетчера–Пауелла, яка використовує корекцію симетричної матриці, що не сприяє інтерпретації методу у перетвореному просторі аргументів. Ця обставина має місце і для інших алгоритмів квазіньютонівского типу [5], а також для низки алгоритмів типу методу спряжених градієнтів [4].

Надалі методами змінної метрики вважатимемо такі методи, які використовують лінійні неортогональні перетворення простору, і результати збіжності яких базуються на дослідженні поведінки характеристик функції, що мінімізується, в перетвореному просторі аргументів. Найбільшу частину методів змінної метрики становлять методи субградієнтного типу, які використовують операцію розтягу простору як у напрямку субградієнта, так і в напрямку різниці двох послідовних субградієнтів (r-алгоритми) [6, 7]. До перших належить також метод еліпсоїдів, а r-алгоритми зарекомендували себе як досить ефективний засіб для розв'язання негладких задач. До методів змінної метрики можна віднести також два алгоритми [8], які по суті є варіантами r-алгоритмів зі змінним коефіцієнтом розтягу простору та відомим фейєрівським регулюванням крокового множника. Сімейство економних у сенсі обробки інформації методів змінної метрики для задач опуклого програмування дозволяє отримати досить загальну схему методів центрів ваг простих тіл [9], яка разом з розтягом простору передбачає використання й інших лінійних неортогональних перетворень простору.

Результати збіжності вищенаведених методів базуються на вивченні поведінки норми субградієнта функції, що мінімізується, в розтягнутому просторі, дослідженні поведінки функції відстані до точки мінімуму в перетвореному просторі, використанні монотонного зменшення об'єму області локалізації множини екстремумів. Вочевидь, є простий механізм, який дозволяє розширити клас методів змінної метрики. Це можна зробити як шляхом збільшення числа лінійних неортогональних перетворень, так і числа досліджуваних характеристик функції, що мінімізується, у перетвореному просторі аргументів. При цьому, зберігаючи градієнтну природу методів у перетвореному просторі, можна будувати методи, близькі за ефективністю до квазіньютонівських, але застосовні для більш широкого класу функцій, ніж гладкі, включно з негладкими.

Практична ефективність r-алгоритмів для яружних задач зумовлена тим, що розтяг простору спрямований на покращення структури поверхонь рівня функції, що мінімізується, у черговому просторі аргументів. Однак, окрім розтягу простору, для цього можна використовувати й інші лінійні неортогональні оператори. Два з них, які дозволяють покращувати структуру поверхонь рівня яружних функцій, будуть предметом обговорення в цій роботі. Ці перетворення досить прості й для мінімізації опуклих функцій дозволяють будувати та обґрунтовувати методи змінної метрики, які мають наочну геометричну інтерпретацію в перетвореному просторі аргументів.

Окрім перетворення простору побудова таких методів потребує як вибору напрямку руху з поточної точки, так і певного способу регулювання крокового множника у цьому напрямку.

Обмежимося розглядом методів субградієнтного типу з класичним фейєрівським регулюванням крокового множника в напрямку антисубградієнта стосовно задачі мінімізації опуклої функції при відомому значенні функції в точці оптимуму. Основний матеріал розділу вперше був опублікований в статтях [10–12].

Матеріал цього розділу викладено так. У підрозділі 1.2 наведено постановку задачі та коротко проаналізовано фейєрівські методи. У розділі 1.3 розглянуто однорангове еліпсоїдальне перетворення простору та наведено на його основі два простих методи з класичним фейєрівським регулюванням кроку. В розділі 1.4 розглянуто доортогоналізуюче перетворення простору та засноване на ньому сімейство методів ортогонального субградієнтного спуску з класичним фейєрівським регулюванням кроку. Практична ефективність наведених методів підкріплена числовими експериментами. В розділах 1.3 і 1.4 також обговорюються як можливі схеми інших алгоритмів з урахуванням цих перетворень, так і конструктивне розв'язання низки принципових проблем методів фейєрівського типу для задач опуклого програмування.

1.2. Постановка задачі та фейєрівські методи

Постановка задачі. Нехай є задача безумовної мінімізації:

$$\min f(x), \tag{1.1}$$

де $f(x)$ — опукла функція векторного аргументу $x \in X$, $X = \mathbb{R}^n$; $\mathbb{R}^n$ — евклідів простір розмірності n зі скалярним добутком (x, y). X будемо називати вихідним простором аргументів, n — розміром задачі.

Нехай множина екстремумів задачі (1.1) X^* непорожня і відомо значення мінімуму $f(x)$: $f^* = f(x^*)$, $x^* \in X^*$. Не обмежуючи загальності, для зручності міркувань припустимо, що множина X^* складається з єдиної точки x^*.

Така постановка задачі має низку застосувань. Зокрема, до задачі (1.1) легко звести задачу пошуку допустимої точки сумісної системи опуклих нерівностей. Якщо система опуклих нерівностей несумісна, аналіз цієї задачі дозволяє легко отримати достатню умову несумісності.

Оскільки далі мова йтиме про чисельні методи для розв'язання задачі (1.1), сформулюємо, що вважатимемо її розв'язком, або, іншими словами, визначимо критерій зупинки для методів. Нехай

$$X_\varepsilon^* = \left\{x : f(x) - f^* \le \varepsilon\right\},$$

де $\varepsilon > 0$. Точку $x_k \in X_\varepsilon^*$, тобто таку, для якої виконується нерівність

$$f(x_k) - f^* \le \varepsilon,$$

вважатимемо розв'язком задачі безумовної мінімізації (1.1) з точністю ε за функціоналом (ε-розв'язком).

Вхідним параметром методів буде точність ε_f, з якою потрібно розв'язати задачу (1.1) за функціоналом, а їхню зупинку будемо робити, як тільки досягнуто точку $x_k \in X_{\varepsilon_f}^*$.

Про фейєрівські методи. Фейєрівські методи беруть початок в роботах [13, 14] та є одним з варіантів методу субградієнтного спуску, який передбачає певне регулювання крокового множника в напрямку нормованого субградієнта.

Для розв'язання задачі (1.1) фейєрівські методи записуються за допомогою такого ітеративного процесу:

$$x_{k+1} = x_k - h_k \frac{\partial f(x_k)}{\|\partial f(x_k)\|}, \quad h_k = \gamma \frac{\left(f(x_i) - f^*\right)}{\|\partial f(x_k)\|}, \tag{1.2}$$

де $\partial f(x_k)$ — субградієнт $f(x)$ у точці x_k, γ — скаляр, такий, що $0 < \gamma < 2$. Такий вибір значень параметра γ гарантує монотонне зменшення відстані до мінімуму на кожному кроці процесу (1.2) та забезпечує простоту доведення збіжності цих процесів.

Крок h_k при $\gamma = 1$ називатимемо класичним фейєрівським кроком, а відповідний йому метод — класичним фейєрівським методом.

Також метод (1.2) в літературі називають субградієнтним методом з кроком Поляка.

Для розв'язання задачі опуклої мінімізації (1.1) з точністю ε_f за функціоналом класичний фейєрівський метод досить конструктивний і набуває такого вигляду.

На початку процесу маємо $x_0 \in \mathbb{R}^n$, $\varepsilon_f > 0$. Нехай на k-му кроці отримано $x_k \in \mathbb{R}^n$, $f(x_k)$ і $\partial f(x_k)$ — обчислені в ній значення функції та субградієнта. Тоді, якщо $f(x_k) - f^* \leq \varepsilon_f$, то x_k — шукана точка. Інакше обчислюємо чергове наближення

$$x_{k+1} = x_k - h_k \frac{\partial f(x_k)}{\|\partial f(x_k)\|}, \quad h_k = \frac{\big(f(x_i) - f^*\big)}{\|\partial f(x_k)\|}, \tag{1.3}$$

і переходимо до $(k+1)$-го кроку. Справедлива така теорема.

Теорема 1.1. *Нехай на кожному кроці методу (1.3)* $\|\partial f(x_k)\| \leq C$*. Тоді метод (1.3) дозволяє знайти* ε_f*-розв'язок (1.1) не більше, ніж за* K*-кроків, де* $K = \left[\left(\frac{C\|x_0 - x^*\|}{\varepsilon_f}\right)^2\right] + 1$*. Тут* $[a]$ *— ціла частина від числа* a*.*

Доведення наводити не будемо. Воно тривіальне і базується на тому, що на кожному кроці ітераційного методу (1.3) відстань до точки мінімуму зменшується з точністю до величини квадрату зсуву в напрямку нормованого субградієнта (тобто з точністю до h_k^2), і на тому, що крок h_k обмежений знизу величиною (ε_f / C). Зазначимо лише, що умову теореми 1.1 можна посилити, якщо величина $\|\partial f(x)\|$ буде обмеженою в кулі радіусу $\|x_0 - x^*\|$ з центром у точці x^*. Але це суттєво нічого не змінює, лише призведе до певного ускладнення матеріалу даного розділу.

Незважаючи на теоретичну привабливість, метод (1.3) має погану репутацію в практичному значенні навіть для гладких опуклих функцій, що обумовлено його повільною збіжністю для яружних функцій. Зокрема, коли поверхні рівня функції $f(x)$ не сильно витягнуті, його робота цілком задовільна для задач навіть немалих розмірів ($n \sim 100$). Однак, якщо напрямок антисубградієнта в точці досить близький до ортогонального — до напрямку на мінімум, що характерно для яружних функцій, метод майже безнадійний навіть для невеликих задач ($n \sim 10$). Для негладких опуклих функцій, де яружність скоріше правило, ніж виняток, метод (1.3) безнадійний для задач ще менших розмірів ($n \sim 2 - 5$). Зокрема, для мінімізації функції двох змінних $f(x_1, x_2) = |x_1| + k\,|x_2|$ він збігається зі швидкістю геометричної прогресії зі знаменником $\sqrt{1 - 1/k^2}$, який близький до одиниці для невеликих значень k, $(k \sim 10 - 100)$.

Для прискорення збіжності методу (1.3), ґрунтуючись на наближенні функціоналу якості поблизу мінімуму кусково-лінійним, у [15] запропоновано більш загальний ітераційний метод:

$$x_{k+1} = P_{Q_k}(x_k), \quad Q_k = \Big\{x : f(x_i) + (\partial f(x_i), x - x_i) \leq f^*,\ i \in I_k\Big\}. \tag{1.4}$$

Тут I_k — будь-яка підмножина індексів з $0, 1, \ldots, k$, що обов'язково містить k, P_{Q_k} — оператор проектування на Q_k.

Головна проблема в методі (1.4) — це вказати конструктивний спосіб формування множини I_k, для того, щоб обмежити кількість субградієнтів, які зберігаються. Найбільш змістовною відповіддю тут можна вважати метод ортогонального спуску [16]. Він базується на ідеї послідовного занурення нетупокутного конуса, який містить X^*, в прямокутний конус, та використовує не більше, ніж $(n + 1)$ вектор.

Однак, він не завжди дає хороші результати за точністю розв'язання задачі (1.4) за функціоналом. Про це свідчать результати низки чисельних експериментів як з методом ортогонального спуску, так і з його модифікацією, представлені в [17, 18]. Незважаючи на невеликий розмір задач ($n \sim 10$) і те, що вони не дуже яружні, точність їхнього розв'язання за функціоналом часто $\varepsilon_f \sim 10^{-4} - 10^{-2}$.

Ще один метод фейєрівського типу запропоновано в [19]. Він використовує всього два вектори і для побудови чергового напрямку спуску використовує лінійну комбінацію напрямку субградієнта та напрямку руху на попередньому кроці (на кшталт сполучених градієнтів) для того, щоб він становив гостріший кут з напрямком на мінімум. Для яружних функцій він забезпечує більш «плавний» рух вздовж яру, ніж вищенаведені методи з використанням всього двох векторів, поведінка яких «стрибкоподібна». Але він також неефективний для яружних функцій, особливо коли яр багатовимірний.

У значенні точності розв'язання задачі (1.1) покращити ситуацію для методу типу (1.3) здатне змінити лінійне перетворення простору, спрямоване на вирівнювання структури поверхонь рівня функції в перетвореному просторі.

Нехай $Y = AX$ — перетворений за допомогою лінійного оператора A простір аргументів. Тут A — невироджена матриця розмірності $n \times n$.

Нехай B — обернений до A оператор, $B = A^{-1}$. Тоді фейєрівський метод з перетворенням простору в X набуває вигляду

$$x_{k+1} = x_k - h_k B \frac{B^T \partial f(x_k)}{\|B^T \partial f(x_k)\|}, \quad h_k = \frac{\left(f(x_i) - f^*\right)}{\|B^T \partial f(x_k)\|}, \tag{1.5}$$

що відповідає класичному фейєрівському кроку в напрямку антисубградієнта у перетвореному просторі аргументів Y

$$y_{k+1} = y_k - h_k \frac{\partial \varphi(y_k)}{\|\partial \varphi(y_k)\|}, \quad h_k = \frac{(\varphi(y_k) - \varphi^*)}{\|\partial \varphi(y_k)\|}. \tag{1.6}$$

Тут $y = Ax$ — образи точок з X у перетвореному просторі Y, $\varphi(y)$ — опукла функція, визначена в перетвореному просторі Y: $\varphi(y) = f(A^{-1}y) = f(x)$. $\partial\varphi(y_k) = B^T \partial f(x_k)$ — субградієнт у точці $y_k = Ax_k$ для функції $\varphi(y)$.

Нехай $Y_k = A_k X$ — поточний простір аргументів. Тоді, покращуючи структуру поверхонь рівня $\varphi_{k+1}(y)$ у черговому перетвореному просторі аргументів $Y_{k+1} = T_{k+1} Y_k$, природно від методів типу (1.5) очікувати більш ефективної роботи, ніж від методу (1.3). Тут T_{k+1} — невироджена матриця розміру $n \times n$, що задає перетворення з Y_k в Y_{k+1}.

Основні співвідношення, які потрібні для реалізації такого процесу, — це перерахунок матриці B_{k+1} згідно з співвідношенням

$$B_{k+1} = B_k T_{k+1}^{-1}, \tag{1.7}$$

і перерахунок субградієнтів для $\varphi_{k+1}(y)$, визначеної в Y_{k+1}, згідно з співвідношенням

$$\partial\varphi_{k+1}(y) = \left(T_{k+1}^{-1}\right)^T \partial\varphi_k(y). \tag{1.8}$$

Надалі будемо часто використовувати різні тотожності, які випливають із ланцюжка рівностей

$$A_{k+1} = B_{k+1}^{-1} = \left(B_k T_{k+1}^{-1}\right)^{-1} = T_{k+1} B_k^{-1} = T_{k+1} A_k.$$

Враховуючи, що

$$\|B^T \partial f(x_k)\|^2 = \Big(B^T \partial f(x_k),\, B^T \partial f(x_k)\Big) = \Big(\partial f(x_k),\, BB^T \partial f(x_k)\Big),$$

процес (1.5) можна записати у формі

$$x_{k+1} = x_k - \frac{\Big(f(x_k) - f^*\Big)}{\Big(\partial f(x_k), H\partial f(x_k)\Big)} H\partial f(x_k), \tag{1.9}$$

де $H = BB^T$ — додатно визначена симетрична матриця розмірності $n \times n$. Залежно від того, яке зі співвідношень брати за основу, є два способи практичної реалізації методів. Будемо дотримуватись для них тих назв, які закріпилися за ними для методів субградієнтного типу з розтягом простору. Метод на основі (1.5) прийнято називати методом у B-формі, або B-методом, а на основі (1.9) — методом у H-формі або H-методом.

Перевага H-методу полягає в тому, що для зберігання симетричної матриці H він вимагає майже вдвічі менше оперативної пам'яті ($n \times (n+1)/2$), ніж

B-метод ($n \times n$ комірок). Однак, метод у B-формі дозволяє просто інтерпретувати процес у перетвореному просторі Y.

Отже, доцільно будувати методи у B-формі, а вже як наслідок розглядати їх аналоги в H-формі, переважно для економії пам'яті та обчислень. Тому орієнтуватимемося на методи, що базуються на співвідношенні (1.5).

1.3. Одноранговий еліпсоїдальний оператор

Для розв'язання задачі (1.1) в [8] розглянуто два субградієнтні методи з класичним фейєрівським кроком у перетвореному просторі аргументів. Вони базуються на зменшенні об'єму області локалізації множини екстремумів і зберігають монотонність оцінки відстані до оптимуму в перетвореному просторі аргументів. Перетворення простору, що в них використовується, за змістом близьке до такого, що застосовується в r-алгоритмах [6], і зводиться до двох послідовних розтягів простору аргументів в ортогональних напрямках.

Оператор перетворення простору має вигляд

$$T_2(\xi, \eta) = R_{\alpha_1}\left(\frac{\xi + \eta}{\|\xi + \eta\|}\right) R_{\alpha_2}\left(\frac{\xi - \eta}{\|\xi - \eta\|}\right). \tag{1.10}$$

Тут ξ, η — нормовані вектори з $\mathbb{R}^n$, такі, що $(\xi, \eta)^2 \neq 1$. $\alpha_1 = 1/\beta_1$, де $\beta_1 = \sqrt{1 - (\xi, \eta)}$ і $\alpha_2 = 1/\beta_2$, де $\beta_2 = \sqrt{1 + (\xi, \eta)}$.

Надалі оператор (1.10) називатимемо дворанговим еліпсоїдальним оператором.

Методи в [8] використовують незначну інформацію про гіперплощини, які відсікають множину екстремумів. Зокрема, перший метод використовує лише два субградієнти у послідовних точках, отриманих згідно з класичним фейєрівським кроком. Другий метод, окрім двох послідовних субградієнтів, використовує ще й третій вектор агрегатного типу, що є опуклою комбінацією обчислених раніше субградієнтів. Незважаючи на невибагливість у використаній інформації, наведені в [8] обчислювальні експерименти з цими методами для низки як гладких, так і негладких тестових задач виявились не такими вже й поганими. Тому більш доцільні стратегії використання обчислених субградієнтів здатні призвести до більш ефективних методів на основі перетворення (1.10).

Однак, незважаючи на просту геометричну інтерпретацію, дворангове еліпсоїдальне перетворення незручне тим, що для методів у B-формі при переході від k-го до $(k+1)$-ого кроку розрахунок матриці B_{k+1} вимагає двох однорангових корекцій матриці розмірності $n \times n$, що призводить до $4n^2$ арифметичних операцій. Але виявляється, що існує перетворення простору, яке допускає однорангову корекцію матриці B_{k+1}, і за змістом не відрізняється від дворангового еліпсоїдального перетворення. Це не дивно, оскільки методи на кшталт (1.5) визначаються матрицею $H_k = B_k B_k^T$. Зважаючи на неоднозначність розкладу матриці H_k, одному й тому самому методу в H-формі можуть відповідати різні методи в B-формі.

Одноранговий еліпсоїдальний оператор. Нехай ξ, $\eta \in \mathbb{R}^n$ — такі вектори, що $\|\xi\| = 1$, $\|\eta\| = 1$, а їх скалярний добуток задовольняє умову $(\xi, \eta)^2 \neq 1$.

Одноранговим еліпсоїдальним оператором називатимемо лінійний оператор з $\mathbb{R}^n$ в $\mathbb{R}^n$, який у матричній формі записується так:

$$T_1(\xi, \eta) = I - \frac{1}{1 - (\xi, \eta)^2}\left(\left(1 - \sqrt{1 - (\xi, \eta)^2}\right)\eta - (\xi, \eta)\,\xi\right)\eta^T. \qquad (1.11)$$

Тут I — одинична матриця розміру $n \times n$. Цей оператор має низку цікавих властивостей, що дозволяє на його основі будувати та обґрунтовувати різноманітні методи змінної метрики. Для нього справедливі такі характеристики.

Лема 1.1. *Для оператора* $T_1(\xi, \eta)$*, заданого згідно з* (1.11)*, якщо виконується* $(\xi, \eta)^2 \neq 1$*, то існує зворотний* $T_1^{-1}(\xi, \eta)$ *і*

$$T_1^{-1}(\xi, \eta) = I + \frac{1}{\sqrt{1 - (\xi, \eta)^2}}\left(\left(1 - \sqrt{1 - (\xi, \eta)^2}\right)\eta - (\xi, \eta)\,\xi\right)\eta^T. \qquad (1.12)$$

Крім того, для $T_1(\xi, \eta)$ *і* $T_1^{-1}(\xi, \eta)$ *виконуються співвідношення:*

$$T_1^T(\xi, \eta)\; T_1(\xi, \eta) = I + \frac{(\xi, \eta)}{1 - (\xi, \eta)^2}\left(\xi\eta^T + \eta\xi^T\right), \qquad (1.13)$$

$$\begin{aligned} T_1^{-1}(\xi, \eta)\left(T_1^{-1}(\xi, \eta)\right)^T &= I + \frac{(\xi, \eta)^2}{1 - (\xi, \eta)^2}\left(\xi\xi^T + \eta\eta^T\right) - \\ &- \frac{(\xi, \eta)}{1 - (\xi, \eta)^2}\left(\xi\eta^T + \eta\xi^T\right). \end{aligned} \qquad (1.14)$$

Доведення леми 1.1 наводити не будемо через його громіздкість. Зазначимо лише, що у справедливості (1.12)–(1.14) легко переконатися безпосередньою перевіркою. Прийнятний вид співвідношень (1.13) і (1.14) обумовлений тим, що при зведенні подібних членів при однорангових матрицях $\xi\xi^T$, $\xi\eta^T$, $\eta\xi^T$ та $\eta\eta^T$ члени з $\sqrt{1 - (\xi, \eta)^2}$ скорочуються.

Лема 1.2. *Нехай* B_k *— невироджена матриця розмірності* $n \times n$, p_1, p_2 *— n-вимірні вектори, такі, що* $\left(\frac{B_k^T p_1}{\|B_k^T p_1\|}, \frac{B_k^T p_2}{\|B_k^T p_2\|}\right)^2 < 1$. *Нехай*

$$B_{k+1} = B_k T_1^{-1}(\xi, \eta), \qquad \text{де} \quad \xi = \frac{B_k^T p_1}{\|B_k^T p_1\|} \qquad \text{та} \quad \eta = \frac{B_k^T p_2}{\|B_k^T p_2\|}. \tag{1.15}$$

Тоді матриця B_{k+1} *невироджена, і для неї справедливі такі властивості:*

а) $\det(B_{k+1}) = \det(B_k)\sqrt{1 - (\xi, \eta)^2}$;

б) $\left(B_{k+1}^T p_1, B_{k+1}^T p_2\right) = 0$.

Доведення. Невиродженість матриці B_{k+1} та властивість а) випливають з

$$\begin{aligned}
&\det(B_{k+1}) = \det(B_k)\det\Big(T_1^{-1}(\xi, \eta)\Big) = \\
= \ &\det(B_k)\left(1 + \frac{1}{\sqrt{1-(\xi,\eta)^2}}\Big(\big(1 - \sqrt{1-(\xi,\eta)^2}\,\big)\eta - (\xi,\eta)\xi,\ \eta\Big)\right) = \\
= \ &\det(B_k)\left(1 + \frac{1}{\sqrt{1-(\xi,\eta)^2}}\left(1 - \sqrt{1-(\xi,\eta)^2} - (\xi,\eta)^2\right)\right) = \\
= \ &\det(B_k)\sqrt{1-(\xi,\eta)^2} \neq 0.
\end{aligned}$$

Для властивості б) маємо

$$\begin{aligned}
\left(B_{k+1}^T p_1, B_{k+1}^T p_2\right) &= \Big(\big(T_1^{-1}(\xi,\eta)\big)^T B_k^T p_1,\ \big(T_1^{-1}(\xi,\eta)\big)^T B_k^T p_2\Big) = \\
&= \Big(B_k^T p_1,\ T_1^{-1}(\xi,\eta)\big(T_1^{-1}(\xi,\eta)\big)^T B_k^T p_2\Big) = \\
&= \frac{1}{\|B_k^T p_2\| \cdot \|B_k^T p_1\|}\Big(\xi,\ T_1^{-1}(\xi,\eta)\big(T_1^{-1}(\xi,\eta)\big)^T \eta\Big).
\end{aligned}$$

Використовуючи співвідношення (1.14), маємо

$$\begin{aligned}
&T_1^{-1}(\xi,\eta)\big(T_1^{-1}(\xi,\eta)\big)^T \eta = \\
= \ &\left(I + \frac{(\xi,\eta)^2}{1-(\xi,\eta)^2}\left(\xi\xi^T + \eta\eta^T\right) - \frac{(\xi,\eta)}{1-(\xi,\eta)^2}\left(\xi\eta^T + \eta\xi^T\right)\right)\eta = \\
= \ &\left(1 + \frac{(\xi,\eta)^2}{1-(\xi,\eta)^2} - \frac{(\xi,\eta)^2}{1-(\xi,\eta)^2}\right)\eta + \\
&\qquad + (\xi,\eta)\left(\frac{(\xi,\eta)^2}{1-(\xi,\eta)^2} - \frac{1}{1-(\xi,\eta)^2}\right)\xi = \eta - (\xi,\eta)\xi,
\end{aligned}$$

звідки випливає

$$\begin{aligned}\left(B_{k+1}^T p_1, B_{k+1}^T p_2\right) &= \frac{1}{\|B_k^T p_2\| \cdot \|B_k^T p_1\|}\Big(\xi, \eta - (\xi, \eta)\xi\Big) = \\ &= \frac{1}{\|B_k^T p_2\| \cdot \|B_k^T p_1\|}\Big((\xi, \eta) - (\xi, \eta)\Big) = 0.\end{aligned}$$

Властивість б) доведено. □

Однорангове еліпсоїдальне перетворення простору (1.11) має багато спільного з дворанговим еліпсоїдальним перетворенням (1.10). Зокрема, для них обох характерно однакове зменшення детермінанта оберненої матриці, що за певних умов рівносильно зменшенню об'єму області локалізації множини екстремумів. Як перше, так і друге ортогоналізують вектори під час переходу в перетворений простір аргументів.

Далі побачимо, що обидва ці перетворення призводять до одного і того ж методу в H-формі.

Проте, на відміну від лінійного дворангового еліпсоїдального перетворення, перетворення (1.11) вимагає вдвічі менше арифметичних операцій при перерахунку матриці B_k, тобто рівно стільки, скільки й розтяг простору в деякому напрямку.

Тому його використання призводить до більш економних методів у B-формі в обчислювальному плані, ніж застосування дворангового еліпсоїдального перетворення (1.10).

Ґрунтуючись на зменшенні об'єму області локалізації множини екстремумів, одноранговий еліпсоїдальний оператор (1.11) дозволяє легко обґрунтовувати методи для розв'язання задачі (1.1), щоб покращувати структуру поверхонь рівня функції у черговому перетвореному просторі.

Нехай $g_k = B_k^T \partial f(x_k)$ і $g_{k+1} = B_k^T \partial f(x_{k+1})$ — субградієнти $\varphi_k(y)$ в $Y_k = A_k X$ у точках $y_k = A_k x_k$ і $y_{k+1} = A_k x_{k+1}$. Тут y_{k+1} отримана згідно з класичним фейєрівським кроком у перетвореному просторі аргументів Y_k.

Тоді, якщо $(g_k, g_{k+1}) \geq 0$, то перетворення простору не потрібне. Нехай $(g_k, g_{k+1}) < 0$. Тоді перетворення простору $Y_{k+1} = T_1\left(\frac{g_k}{\|g_k\|}, \frac{g_{k+1}}{\|g_{k+1}\|}\right) Y_k$ або $Y_{k+1} = T_1\left(\frac{g_{k+1}}{\|g_{k+1}\|}, \frac{g_k}{\|g_k\|}\right) Y_k$ згідно з лемою 1.2 дозволяє ортогоналізувати образи субградієнтів g_k та g_{k+1} у перетвореному просторі аргументів Y_{k+1}, забезпечивши при цьому кращі поверхні рівня для функції $\varphi_{k+1}(y)$ у Y_{k+1}.

Ці прості міркування призводять до такого методу для розв'язання задачі мінімізації (1.1).

Метод фейєрівського типу з одноранговим еліпсоїдальним перетворенням простору, що використовує два послідовні субградієнти. Перед початком обчислень маємо: $x_0 \in \mathbb{R}^n$; $B_0 = I$ — одинична матриця розмірності $n \times n$; ε_f — точність за функціоналом, з якою потрібно розв'язати задачу (1.1); $f(x_0)$ та $g(x_0) = \partial f(x_0)$ — обчислені в x_0 значення функції та субградієнта. Тоді, якщо $f(x_0) - f^* \leq \varepsilon_f$, то x_0 — шукана точка і ЗУПИНКА.

Нехай на k-му кроці отримано $x_k \in \mathbb{R}^n$, $f(x_k)$, $g(x_k) = \partial f(x_k)$, B_k — матриця розмірності $n \times n$.

1) Обчислимо чергове наближення

$$x_{k+1} = x_k - \frac{f(x_k) - f^*}{\|B_k^T g(x_k)\|} B_k \frac{B_k^T g(x_k)}{\|B_k^T g(x_k)\|}. \tag{1.16}$$

2) Обчислимо $f(x_{k+1})$ та $g(x_{k+1}) = \partial f(x_{k+1})$. Якщо $f(x_{k+1}) - f^* \leq \varepsilon_f$, то x_{k+1} — шукана точка і ЗУПИНКА. Інакше переходимо до пункту 3.

3) Покладемо

$$\xi_1 = \frac{B_k^T g(x_k)}{\|B_k^T g(x_k)\|}, \quad \xi_2 = \frac{B_k^T g(x_{k+1})}{\|B_k^T g(x_{k+1})\|}. \tag{1.17}$$

4) Якщо $(\xi_1, \xi_2) \geq 0$, вважаємо $B_{k+1} = B_k$ і переходимо до пункту 5. Інакше обчислюємо

$$B_{k+1} = B_k T_1^{-1}(\xi_1, \xi_2) \quad \text{або} \quad B_{k+1} = B_k T_1^{-1}(\xi_2, \xi_1). \tag{1.18}$$

5) Переходимо до чергового кроку з x_{k+1}, B_{k+1}, $g(x_{k+1})$, $f(x_{k+1})$.

Справедлива така теорема.

Теорема 1.2. *Послідовність $\{x_{k+1}\}_{k=0}^{\infty}$, яка генерується методом (1.16)–(1.18), задовольняє нерівності*

$$\|A_{k+1}(x_{k+1} - x^*)\|^2 \leq \|A_k(x_k - x^*)\|^2 - \frac{\left(f(x_k) - f^*\right)^2}{\|B_k^T g(x_k)\|^2}. \tag{1.19}$$

Тут $A_k = B_k^{-1}$, $A_{k+1} = B_{k+1}^{-1}$, $k = 0, 1, 2, \ldots$.

Доведення. Спочатку покажемо, що для методу (1.16)–(1.18) завжди виконується нерівність

$$\|A_{k+1}(x_{k+1} - x^*)\|^2 \leq \|A_k(x_{k+1} - x^*)\|^2. \tag{1.20}$$

Якщо $(\xi_1, \xi_2) \geq 0$, то (1.20) виконується як рівність, оскільки $A_{k+1} = A_k$.

Якщо $(\xi_1,\ \xi_2) < 0$, то, використовуючи співвідношення (1.13) леми 1.1 і

$$\xi_1 \ = \frac{B_k^T g(x_k)}{\|B_k^T g(x_k)\|}, \quad \xi_2 = \frac{B_k^T g(x_{k+1})}{\|B_k^T g(x_{k+1})\|},$$

маємо

$$\begin{aligned}
&\|A_{k+1}(x_{k+1} - x^*)\|^2 = \|T_1(\xi_1, \xi_2)\, A_k(x_{k+1} - x^*)\|^2 = \\
=&\Big(T_1(\xi_1, \xi_2)\, A_k(x_{k+1} - x^*),\ T_1(\xi_1, \xi_2)\, A_k(x_{k+1} - x^*)\Big) = \\
=&\Big(A_k(x_{k+1} - x^*), \big(T_1(\xi_1, \xi_2)\big)^T T_1(\xi_1, \xi_2) A_k(x_{k+1} - x^*)\Big) = \\
=&\left(A_k(x_{k+1} - x^*),\ \left(I + \frac{(\xi_1, \xi_2)}{1 - (\xi_1, \xi_2)^2}\left(\xi_1\xi_2^T + \xi_2\xi_1^T\right)\right) A_k(x_{k+1} - x^*)\right) = \\
=&\|A_k(x_{k+1} - x^*)\|^2 + \\
&\quad + \frac{2(\xi_1, \xi_2)}{1 - (\xi_1, \xi_2)^2}\Big(\xi_1,\ A_k(x_{k+1} - x^*)\Big)\Big(\xi_2, A_k(x_{k+1} - x^*)\Big) = \\
=&\|A_k(x_{k+1} - x^*)\|^2 + \\
&\quad + \frac{2(\xi_1,\ \xi_2)}{1 - (\xi_1,\ \xi_2)^2} \cdot \frac{\big(g(x_k),\ x_{k+1} - x^*\big)}{\|B_k^T g(x_k)\|} \cdot \frac{\big(g(x_{k+1}),\ x_{k+1} - x^*\big)}{\|B_k^T g(x_{k+1})\ |}.
\end{aligned} \tag{1.21}$$

Причому рівність (1.21) справедлива як для $B_{k+1} = B_k T_1^{-1}(\xi_1, \xi_2)$, так і для $B_{k+1} = B_k T_1^{-1}(\xi_2, \xi_1)$.

Другий доданок у (1.21) від'ємний, оскільки один з його співмножників

$$\frac{2(\xi_1, \xi_2)}{\big(1 - (\xi_1, \xi_2)^2\big) \cdot \|B_k^T g(x_k)\| \cdot \|B_k^T g(x_{k+1})\|} < 0$$

тому, що $(\xi_1, \xi_2) < 0$, $\|\xi_1\| = 1$, $\|\xi_2\| = 1$, а два інших, враховуючи опуклість функції $f(x)$, додатні

$$\begin{aligned}
&\Big(g(x_k),\ x_{k+1} - x^*\Big) = \\
=\ &\left(g(x_k),\ x_k - x^* - \frac{\big(f(x_i) - f^*\big)}{\|B_k^T g(x_k)\|} \cdot B_k \cdot \frac{B_k^T g(x_k)}{\|B_k^T g(x_k)\|}\right) = \\
=\ &\Big(g(x_k),\ x_k - x^*\Big) - \frac{\big(f(x_i) - f^*\big)}{\|B_k^T g(x_k)\|^2} \cdot \Big(g(x_k),\ B_k B_k^T g(x_k)\Big) =
\end{aligned}$$

$$= \Big(g(x_k),\ x_k - x^*\Big) - \frac{\big(f(x_i) - f^*\big)}{\|B_k^T g(x_k)\|^2} \cdot \Big(B_k^T g(x_k), B_k^T g(x_k)\Big) =$$
$$= \Big(g(x_k),\ x_k - x^*\Big) - \Big(f(x_k) - f^*\Big) \geq 0$$

та

$$\Big(g(x_{k+1}),\ x_{k+1} - x^*\Big) \geq f(x_{k+1}) - f^* \geq 0.$$

Отже, нерівність (1.20) справедлива і для $(\xi_1, \xi_2) < 0$.

Далі, враховуючи, що через опуклість $f(x)$

$$\Big(x_k - x^*, g(x_k)\Big) \geq f(x_k) - f^* \geq 0,$$

нерівність (1.20) можна записати

$$\Big\|A_{k+1}(x_{k+1} - x^*)\Big\|^2 \leq \Big\|A_k(x_{k+1} - x^*)\Big\|^2 =$$
$$= \Big\|A_k\Big(x_k - x^* - \frac{\big(f(x_i) - f^*\big)}{\|B_k^T g(x_k)\|} \cdot B_k \cdot \frac{B_k^T g(x_k)}{\|B_k^T g(x_k)\|}\Big)\Big\|^2 =$$
$$= \Big\|A_k(x_k - x^*) - \frac{\big(f(x_i) - f^*\big)}{\|B_k^T g(x_k)\|^2} B_k^T g(x_k)\Big\|^2 =$$
$$= \Big\|A_k(x_k - x^*)\Big\|^2 -$$
$$-2\frac{\big(f(x_i) - f^*\big)}{\|B_k^T g(x_k)\|^2}\Big(A_k(x_k - x^*),\ B_k^T g(x_k)\Big) + \frac{\big(f(x_i) - f^*\big)^2}{\|B_k^T g(x_k)\|^2} =$$
$$= \Big\|A_k(x_k - x^*)\Big\|^2 - 2\frac{\big(f(x_i) - f^*\big)}{\|B_k^T g(x_k)\|^2}\Big(x_k - x^*, g(x_k)\Big) + \frac{\big(f(x_i) - f^*\big)^2}{\|B_k^T g(x_k)\|^2} \leq$$
$$\leq \Big\|A_k(x_k - x^*)\Big\|^2 - \frac{\big(f(x_i) - f^*\big)^2}{\|B_k^T g(x_k)\|^2},$$

що закінчує доведення теореми. □

Для методу (1.16)–(1.18) теорема 1.2 забезпечує можливість обґрунтування його граничної збіжності до розвязку задачі (1.1) шляхом зменшення об'єму еліпсоїда, що локалізує x^*. При цьому зменшення об'єму тим краще, чим більший класичний фейєрівський крок у перетвореному просторі і чим тупіший кут між послідовними субградієнтами. Але ці обидві обставини залежать від конкретних властивостей $f(x)$ і отримати будь-які оцінки для загальної опуклої $f(x)$ неможливо. Тому зупинимося на такому варіанті доведення збіжності методу (1.16)–(1.18).

Теорема 1.3. *Нехай на кожному кроці методу (1.16)–(1.18) виконуються умови* $\|B_k\| \le c_1$ *та* $\|\partial f(x_k)\| \le c_2$. *Тут* $\|B_k\|$ *— евклідова норма матриці* B_k, *тобто* $\|B_k\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}^k|^2\right)^{1/2}$. *Тоді метод (1.16)–(1.18) розв'язує задачу (1.1) з точністю* ε_f *за функціоналом не більше ніж за* K *кроків, де* $K = \left[\left(\frac{c_1c_2\|x_0 - x^*\|}{\varepsilon_f}\right)^2\right] + 1$. *Тут* $[a]$ *— ціла частина числа* a.

Доведення. З теореми 1.2 випливає справедливість нерівності

$$\|A_k(x_k - x^*)\|^2 \le \|x_0 - x^*\| - \sum_{i=0}^{k-1}\frac{\left(f(x_i) - f^*\right)^2}{\|B_i^T g(x_i)\|^2}.$$

Нехай метод (1.16)–(1.18) за k-кроків не розв'язав задачу (1.1) з точністю ε_f за функціоналом. Тоді для всіх x_i, $i = 0, \dots, k$, виконується $f(x_i) - f^* \ge \varepsilon_f$. Отже,

$$\sum_{i=0}^{k-1}\frac{\left(f(x_i) - f^*\right)^2}{\|B_i^T g(x_i)\|^2} \ge \sum_{i=0}^{k-1}\frac{\left(f(x_i) - f^*\right)^2}{\|B_i\|^2\|g(x_i)\|^2} \ge \frac{1}{c_1^2c_2^2}\sum_{i=0}^{k-1}\varepsilon_f^2 \ge \frac{k\varepsilon_f^2}{c_1^2c_2^2},$$

звідки $\|A_k(x_k - x^*)\|^2 \le \|x_0 - x^*\|^2 - \frac{k\varepsilon_f^2}{c_1^2c_2^2}$. Враховуючи, що $\|A_k(x_k - x^*)\|^2$ має бути невід'ємним, отримуємо, що не більше ніж за $K = \left[\left(\frac{c_1c_2\|x_0 - x^*\|}{\varepsilon_f}\right)^2\right] + 1$ кроків метод (1.16)–(1.18) розв'яже задачу (1.1) з точністю ε_f. □

Отже, якщо виконуються умови теореми 1.3, метод (1.16)–(1.18) з будь-якого початкового наближення збігається до ε_f-розв'язку задачі (1.1) за скінчену кількість ітерацій. Але оцінка максимальної кількості ітерацій, необхідної для досягнення точності ε_f, дуже груба. При реальній роботі методу слід очікувати, що для яружних функцій критерій зупинки спрацьовуватиме з умови

$$f(x_k) - f^* \le \|B_k^T g(x_k)\| \sqrt{\|x_0 - x^*\|^2 - \sum_{i=0}^{k-1}\frac{\left(f(x_i) - f^*\right)^2}{\|B_i^T g(x_i)\|^2}}, \tag{1.22}$$

оскільки величина $\|B_k^T g(x_k)\|$ буде досить малою при великих k через те, що $\det(B_k)$ прямує до нуля при $k \to \infty$.

Тому доцільно посилювати перетворення простору для того, щоб гарантувати сильніше прямування до нуля $\det(B_k)$ при $k \to \infty$. Зокрема, для цього підходить варіант методу як у [8], який використовує для посилення перетворення агрегатний вектор, який є опуклою комбінацією обчислених раніше субградієнтів. При цьому метод трохи ускладниться, і для нього відносно методу (1.16)–(1.18) буде характерне сильніше вирівнювання поверхонь рівня функції у перетвореному просторі, особливо при багатовимірних ярах.

Зосередимо увагу на зручних для реалізації обчислювальних схемах цих двох методів. Хоча слід зазначити, що перетворення (1.11) дозволяє обґрунтовувати ціле сімейство методів фейєрівського типу. Деякі з них обговоримо докладніше пізніше.

Про практичну реалізацію методів. Незважаючи на зручності при доведенні, метод (1.16)–(1.18) незручний для реалізації на ЕОМ. Зокрема, коли реалізується операція перетворення простору, він вимагає обчислювати образ субградієнта $\partial f(x_{k+1})$ як у просторі Y_k – $B_k^T \partial f(x_{k+1})$, так і в просторі Y_{k+1} – $B_{k+1}^T \partial f(x_{k+1})$. Обчислення кожного з них вимагає n^2 арифметичних операцій множення і стільки ж додавання, що призводить загалом до $4n^2$ арифметичних операцій.

Цього можна уникнути, якщо субградієнт при переході в черговий перетворений простір аргументів перераховувати відповідно до співвідношення (1.8). Тим більше, що в розглянутих методах як напрямок руху використовується саме антисубградієнт у перетвореному просторі аргументів.

Для перетворення (1.11) це дозволяє така лема.

Лема 1.3. *Нехай* B_k *— невироджена матриця розмірності* $n \times n$, p_k, p_{k+1} *—* n*-вимірні вектори, такі, що* $\left(\frac{B_k^T p_k}{\|B_k^T p_k\|}, \frac{B_k^T p_{k+1}}{\|B_k^T p_{k+1}\|}\right)^2 < 1$. *Нехай* $B_{k+1} = T_1^{-1}(\xi_k, \xi_{k+1})$, *де* $\xi_k = \frac{B_k^T p_k}{\|B_k^T p_k\|}$, $\xi_{k+1} = \frac{B_k^T p_{k+1}}{\|B_k^T p_{k+1}\|}$. *Тоді*

$$\frac{B_{k+1}^T p_{k+1}}{\|B_{k+1}^T p_{k+1}\|} = \xi_{k+1},$$

$$\frac{B_{k+1}^T p_k}{\|B_{k+1}^T p_k\|} = \frac{1}{\sqrt{1-(\xi_k, \xi_{k+1})^2}}\Big(\xi_k - (\xi_k, \xi_{k+1})\xi_{k+1}\Big), \tag{1.23}$$

$$\|B_{k+1}^T p_i\| = \sqrt{1-(\xi_k, \xi_{k+1})^2}\,\|B_k^T p_i\|, \quad i = k,\ k+1. \tag{1.24}$$

Доведення. Для $B_{k+1}^T p_{k+1}$ маємо

$$
\begin{aligned}
&B_{k+1}^T p_{k+1} = \\
&= \Big(T_1^{-1}(\xi_k, \xi_{k+1})\Big)^T B_k^T p_{k+1} = \|B_k^T p_{k+1}\| \Big(T_1^{-1}(\xi_k, \xi_{k+1})\Big)^T \xi_{k+1} = \\
&= \|B_k^T p_{k+1}\| \Bigg(I + \xi_{k+1}\bigg(\Big(\frac{1}{\sqrt{1-(\xi_k, \xi_{k+1})^2}} - 1\Big)\xi_{k+1} - \\
&\qquad\qquad - \frac{(\xi_k, \xi_{k+1})}{\sqrt{1-(\xi_k, \xi_{k+1})^2}}\xi_k\bigg)^T\Bigg)\xi_{k+1} = \\
&= \|B_k^T p_{k+1}\| \Big(1 + \frac{1}{\sqrt{1-(\xi_k, \xi_{k+1})^2}} - 1 - \frac{(\xi_k, \xi_{k+1})^2}{\sqrt{1-(\xi_k, \xi_{k+1})^2}}\Big)\xi_{k+1} = \\
&= \|B_k^T p_{k+1}\| \sqrt{1-(\xi_k, \xi_{k+1})^2}\,\xi_{k+1},
\end{aligned}
$$

звідки випливає справедливість першої частини (1.23), а також справедливість (1.24) при $i = k+1$.

Аналогічно для $B_{k+1}^T p_k$ маємо

$$
\begin{aligned}
&B_{k+1}^T p_k = \\
&= \Big(T_1^{-1}(\xi_k, \xi_{k+1})\Big)^T B_k^T p_k = \|B_k^T p_k\| \Big(T_1^{-1}(\xi_k, \xi_{k+1})\Big)^T \xi_k = \\
&= \|B_k^T p_k\| \Bigg(I + \xi_{k+1}\bigg(\Big(\frac{1}{\sqrt{1-(\xi_k, \xi_{k+1})^2}} - 1\Big)\xi_{k+1} - \\
&\qquad\qquad - \frac{(\xi_k, \xi_{k+1})}{\sqrt{1-(\xi_k, \xi_{k+1})^2}}\xi_k\bigg)^T\Bigg)\xi_k = \\
&= \|B_k^T p_k\| \Bigg(\xi_k + \Big(\frac{(\xi_k, \xi_{k+1})}{\sqrt{1-(\xi_k, \xi_{k+1})^2}} - (\xi_k, \xi_{k+1}) - \\
&\qquad\qquad - \frac{(\xi_k, \xi_{k+1})}{\sqrt{1-(\xi_k, \xi_{k+1})^2}}\Big)\xi_{k+1}\Bigg) = \\
&= \|B_k^T p_k\| \Big(\xi_k - (\xi_k,\ \xi_{k+1})\xi_{k+1}\Big),
\end{aligned}
$$

звідки, враховуючи $\|\xi_k - (\xi_k, \xi_{k+1})\xi_{k+1}\| = \sqrt{1-(\xi_k, \xi_{k+1})^2}$, випливає справедливість другої частини (1.23), а також (1.24) при $i = k$. □

Для векторів $B_k^T p_k$ та $B_k^T p_{k+1}$ у перетвореному просторі $Y_k = A_k X$ співвідношення (1.23)–(1.24) леми 1.3 забезпечують перерахунок як нормованих напрямів, так і їх норм при переході в черговий перетворений простір аргументів, отриманий в результаті застосування оператора

$$T_1\Big(\frac{B_k^T p_k}{\|B_k^T p_k\|}, \frac{B_k^T p_{k+1}}{\|B_k^T p_{k+1}\|}\Big).$$

Однак, для B-форми методів байдуже, яке з перетворень $T_1\Big(\frac{B_k^T p_k}{\|B_k^T p_k\|}, \frac{B_k^T p_{k+1}}{\|B_k^T p_{k+1}\|}\Big)$ або $T_1\Big(\frac{B_k^T p_{k+1}}{\|B_k^T p_{k+1}\|}, \frac{B_k^T p_k}{\|B_k^T p_k\|}\Big)$ вибирати. Вибір першого більш раціональний, оскільки при цьому забезпечується перерахунок субградієнтів, що дозволяє не накопичувати помилки у нормованих напрямках для здійснення кроку в черговому перетвореному просторі аргументів.

Для реалізації методів потрібен також перерахунок класичного фейєрівського крокового множника у напрямку нормованого субградієнта під час переходу в черговий перетворений простір аргументів. Це не призводить до якихось проблем, оскільки їх перерахунок пов'язаний зі зміною норм субградієнтів і, отже,

$$h_{k+1} = h_k \frac{\|B_k^T g_k\|}{\|B_{k+1}^T g_k\|}.$$

Тут h_{k+1} і h_k — класичні фейєрівські кроки у напрямку нормованого субградієнта у просторах $Y_{k+1} = A_{k+1}X$ та $Y_k = A_k X$.

Отже, лема 1.3 дозволяє уникнути для B-форми методу (1.16)–(1.18) зайвих обчислень і для задачі (1.1) він набуває більш раціонального вигляду в обчислювальному плані.

Перед початком обчислень маємо $\varepsilon_f > 0$, $x_0 \in \mathbb{R}^n$. Тоді, якщо $f(x_0) - f^* \le \varepsilon_f$, то x_0 — шукана точка і ЗУПИНКА. Інакше вважаємо $h_0 = \frac{f(x_0)-f^*}{\|\partial f(x_0)\|}$, $\xi_0 = \frac{\partial f(x_0)}{\|\partial f(x_0)\|} \in \mathbb{R}^n$, $B_0 = I$ — одинична матриця розмірності $n \times n$.

Нехай на k-му кроці отримано $x_k \in \mathbb{R}^n$, h_k, $\xi_k \in \mathbb{R}^n$, B_k — матриця розмірності $n \times n$.

1) Обчислимо чергове наближення

$$x_{k+1} = x_k - h_k B_k \xi_k. \tag{1.25}$$

2) Обчислимо $f(x_{k+1})$, $\partial f(x_{k+1})$. Тоді, якщо $f(x_{k+1}) - f^* \le \varepsilon_f$, то x_{k+1} — шукана точка і ЗУПИНКА. Інакше покладемо

$$\xi_{k+1} = \frac{B_k^T \partial f(x_{k+1})}{\|B_k^T \partial f(x_{k+1})\|}, \quad h_{k+1} = \frac{f(x_{k+1}) - f^*}{\|B_k^T \partial f(x_{k+1})\|}. \tag{1.26}$$

3) Якщо $(\xi_k, \xi_{k+1}) \geq 0$, вважаємо $B_{k+1} = B_k$ і переходимо до пункту 4. Інакше обчислюємо

$$\eta = \Big(\frac{1}{\sqrt{1-(\xi_k,\xi_{k+1})^2}} - 1\Big)\xi_{k+1} - \frac{(\xi_k,\xi_{k+1})}{\sqrt{1-(\xi_k,\xi_{k+1})^2}}\xi_k, \tag{1.27}$$

$$B_{k+1} = B_k\Big(I + \eta\xi_{k+1}^T\Big), \quad h_{k+1} = \frac{h_{k+1}}{\sqrt{1-(\xi_k,\xi_{k+1})^2}}. \tag{1.28}$$

4) Переходимо до чергового кроку з x_{k+1}, B_{k+1}, ξ_{k+1}, h_{k+1}.

Отже, метод (1.25)–(1.28) тотожний методу (1.16)–(1.18), і для нього справедливі теореми 1.2 і 1.3. Однак, він обчислювально раціональніший, ніж метод (1.25)–(1.28), і далі саме його виберемо для чисельних експериментів.

Метод фейєрівського типу з одноранговим еліпсоїдальним перетворенням простору, що використовує два послідовні субградієнти та вектор агрегатного типу. Аналогічно до [8], легко побудувати метод, який використовує два послідовні субградієнти та деякий агрегатний вектор, який дозволяє посилити перетворення простору та забезпечити більш швидку збіжність методу в сенсі (1.22). Використовуючи вищенаведені леми 1.2 та 1.3, а також лему 3 з [8], приходимо до такого методу для розв'язання (1.1).

Перед початком обчислень маємо $\varepsilon_f > 0$, $x_0 \in \mathbb{R}^n$. Тоді, якщо $f(x_0) - f^* \leq \varepsilon_f$, то x_0 – шукана точка і ЗУПИНКА. Інакше вважаємо $h_0 = \dfrac{f(x_0) - f^*}{\|\partial f(x_0)\|}$, $\xi_0 = \dfrac{\partial f(x_0)}{\|\partial f(x_0)\|} \in \mathbb{R}^n$, $p_0 = 0 \in \mathbb{R}^n$, $B_0 = I$ – одинична матриця розмірності $n \times n$.

Нехай на k-му кроці отримано $x_k \in \mathbb{R}^n$, h_k, $\xi_k \in \mathbb{R}^n$, $p_k \in \mathbb{R}^n$, B_k – матриця розмірності $n \times n$.

1) Обчислимо чергове наближення

$$x_{k+1} = x_k - h_k B_k \xi_k. \tag{1.29}$$

2) Обчислимо $f(x_{k+1})$, $\partial f(x_{k+1})$. Тоді, якщо $f(x_{k+1}) - f^* \leq \varepsilon_f$, то x_{k+1} – шукана точка і ЗУПИНКА. Інакше покладемо

$$\xi_{k+1} = \frac{B_k^T \partial f(x_{k+1})}{\|B_k^T \partial f(x_{k+1})\|}, \quad h_{k+1} = \frac{f(x_{k+1}) - f^*}{\|B_k^T \partial f(x_{k+1})\|}. \tag{1.30}$$

3) Обчислимо

$$\lambda_1 = -\frac{(p_k, \xi_{k+1})}{\sqrt{(p_k, \xi_{k+1})^2 + (\xi_k, \xi_{k+1})^2}}, \;\; \lambda_2 = -\frac{(\xi_k, \xi_{k+1})}{\sqrt{(p_k, \xi_{k+1})^2 + (\xi_k, \xi_{k+1})^2}}$$

і покладемо

$$p_{k+1} = \begin{cases} \lambda_1 p_k + \lambda_2 \xi_k, & \text{якщо} \;\; \lambda_1 > 0 \;\; \text{і} \;\; \lambda_2 > 0, \\ p_k, & \text{якщо} \;\; \lambda_1 > 0 \;\; \text{і} \;\; \lambda_2 \leq 0, \\ \xi_k, & \text{якщо} \;\; \lambda_1 \leq 0 \;\; \text{і} \;\; \lambda_2 > 0, \\ 0, & \text{якщо} \;\; \lambda_1 \leq 0 \;\; \text{і} \;\; \lambda_2 \leq 0. \end{cases} \tag{1.31}$$

4) Якщо $(p_{k+1}, \xi_{k+1}) \geq 0$, то $B_{k+1} = B_k$ і переходимо до пункту 5. Інакше обчислюємо

$$\eta = \left(\frac{1}{\sqrt{1 - (p_{k+1}, \xi_{k+1})^2}} - 1\right)\xi_{k+1} - \frac{(p_{k+1}, \xi_{k+1})}{\sqrt{1 - (p_{k+1}, \xi_{k+1})^2}} p_{k+1}, \tag{1.32}$$

$$B_{k+1} = B_k\Big(I + \eta \xi_{k+1}^T\Big), \quad h_{k+1} = \frac{h_{k+1}}{\sqrt{1 - (p_{k+1}, \xi_{k+1})^2}}, \tag{1.33}$$

$$p_{k+1} = \frac{1}{\sqrt{1 - (p_{k+1}, \xi_{k+1})^2}}\Big(p_{k+1} - (p_{k+1}, \xi_{k+1})\xi_{k+1}\Big). \tag{1.34}$$

5) Переходимо до наступного кроку з x_{k+1}, B_{k+1}, ξ_{k+1}, h_{k+1}, p_{k+1}.

Для методу (1.29)–(1.34) властиве використання мінімуму інформації про поведінку функції зі збереженням класичного фейєрівського кроку у напрямку останнього обчисленого субградієнта. Зокрема, крім двох останніх субградієнтів, він використовує також вектор агрегатного типу p_k, який є опуклою комбінацією обчислених раніше субградієнтів, і задає гіперплощину, яка відсікає від образу точки x_k множину екстремумів у перетвореному просторі аргументів.

При цьому проблема оновлення вектора p_{k+1} вирішується автоматично за допомогою аналізу кутів між останніми субградієнтами та поточним p_k.

Для методу (1.29)–(1.34) справедливі аналоги теорем 1.2 і 1.3, як і для методу (1.25)–(1.28). Зокрема, справедлива така теорема.

Теорема 1.4. *Нехай на кожному кроці методу* (1.29)–(1.34) $\|B_k\| \leq c_1$ *та* $\|\partial f(x_k)\| \leq c_2$. *Тоді метод* (1.29)–(1.34) *розв'язує задачу* (1.1) *з точністю* ε_f *за функціоналом не більше, ніж за* K*-кроків, де* $K = \left[\left(\frac{c_1 c_2 \|x_0 - x^*\|}{\varepsilon_f}\right)^2\right] + 1.$

Вона гарантує його скінчену збіжність до ε_f-розв'язку задачі (1.1) з довільного початкового наближення. Для яружних функцій відносно методу (1.25)–(1.28) він забезпечує більше швидке прямування до нуля детермінанта матриці B_k при $k \to \infty$. Тому від нього слід очікувати більш ефективної практичної роботи в сенсі (1.22), ніж від методу (1.25)–(1.28).

Про H-форму наведених методів. Розглянуті вище методи, які використовують однорангове еліпсоїдальне перетворення простору, близькі до методів [8] на базі дворангового еліпсоїдального перетворення. Зокрема, першому з них відповідає метод (1.28)–(1.33), другому — метод (37)–(44) з [8].

Нехай всі ці методи стартують з однієї й тієї початкової точки. Тоді B-форми методів (1.25)–(1.28), (1.29)–(1.34) і B-форми відповідних їм аналогів з [8] різні в тому розумінні, що вони генерують різні послідовності матриць B_k. При цьому відрізняються також напрямки руху у перетворених просторах.

Однак, послідовності точок $\{x_k\}_{k=0}^{\infty}$, які генерують методи (1.25)–(1.28) і (1.29)–(1.34) збігаються з такими ж послідовностями їх аналогів з [8]. Річ у тім, що H-форми цих методів співпадають, тобто послідовності матриць $\{H_k\}_{k=0}^{\infty}$ для них однакові.

Насправді H-форму вищенаведених алгоритмів визначає співвідношення (1.14) леми 1.1 для $T_1^{-1}(\xi,\eta)\big(T_1^{-1}(\xi,\eta)\big)^T$. Аналогічний добуток для $T_2(\xi,\eta)$ має вигляд

$$
\begin{aligned}
&T_2^{-1}(\xi,\eta)\Big(T_2^{-1}(\xi,\eta)\Big)^T = \\
&= R_{\beta_1}\Big(\frac{\xi+\eta}{\|\xi+\eta\|}\Big)R_{\beta_2}\Big(\frac{\xi-\eta}{\|\xi-\eta\|}\Big)R_{\beta_1}^T\Big(\frac{\xi+\eta}{\|\xi+\eta\|}\Big)R_{\beta_2}^T\Big(\frac{\xi-\eta}{\|\xi-\eta\|}\Big) = \\
&= R_{\beta_1^2}\Big(\frac{\xi+\eta}{\|\xi+\eta\|}\Big)R_{\beta_2^2}\Big(\frac{\xi-\eta}{\|\xi-\eta\|}\Big) = \Big(I+(\beta_1^2-1)\Big(\frac{\xi+\eta}{\|\xi+\eta\|}\Big)\Big(\frac{\xi+\eta}{\|\xi+\eta\|}\Big)^T\Big)\times \\
&\qquad\qquad \times\Big(I+(\beta_2^2-1)\Big(\frac{\xi-\eta}{\|\xi-\eta\|}\Big)\Big(\frac{\xi-\eta}{\|\xi-\eta\|}\Big)^T\Big) = \\
&= I+(\beta_1^2-1)\Big(\frac{\xi+\eta}{\|\xi+\eta\|}\Big)\Big(\frac{\xi+\eta}{\|\xi+\eta\|}\Big)^T+(\beta_2^2-1)\Big(\frac{\xi-\eta}{\|\xi-\eta\|}\Big)\Big(\frac{\xi-\eta}{\|\xi-\eta\|}\Big)^T = \\
&= I-\frac{(\xi,\eta)}{2\big(1+(\xi,\eta)\big)}(\xi+\eta)(\xi+\eta)^T+\frac{(\xi,\eta)}{2\big(1-(\xi,\eta)\big)}(\xi-\eta)(\xi-\eta)^T = \\
&= I+\Big(\frac{(\xi,\eta)}{2\big(1-(\xi,\eta)\big)}-\frac{(\xi,\eta)}{2\big(1+(\xi,\eta)\big)}\Big)\Big(\xi\xi^T+\eta\eta^T\Big)-
\end{aligned}
$$

$$-\Big(\frac{(\xi,\eta)}{2(1+(\xi,\eta))}+\frac{(\xi,\eta)}{2(1-(\xi,\eta))}\Big)\Big(\xi\eta^T+\eta\xi^T\Big)=$$

$$=I+\frac{(\xi,\eta)^2}{1-(\xi,\eta)^2}\Big(\xi\xi^T+\eta\eta^T\Big)-\frac{(\xi,\eta)}{1-(\xi,\eta)^2}\Big(\xi\eta^T+\eta\xi^T\Big)$$

і точно відповідає співвідношенню (1.14) для $T_1^{-1}(\xi,\eta)\big(T_1^{-1}(\xi,\eta)\big)^T$.

Прямолінійне використання співвідношення (1.14) для перерахунку матриці H_{k+1} вимагає чотирьох однорангових корекцій матриць розмірності $n\times n$. Однак, згрупувавши відповідним чином члени при однорангових матрицях, для перерахунку H_{k+1} достатньо двох однорангових корекцій. Зокрема, для H-форми, яка відповідає методу (1.25)–(1.28), формули для перерахунку матриці H_{k+1} набувають однієї з форм

$$\begin{gathered}H_{k+1}=H_k-\frac{H_kg_{k+1}g_{k+1}^TH_k}{(g_{k+1},H_kg_{k+1})}+\frac{H_kpp^TH_k}{(p,H_kp)},\\ p=\frac{(g_k,H_kg_{k+1})}{(g_k,H_kg_k)}\,g_k-g_{k+1},\end{gathered}\qquad(1.35)$$

або

$$\begin{gathered}H_{k+1}=H_k-\frac{H_kg_kg_k^TH_k}{(g_k,H_kg_k)}+\frac{H_kpp^TH_k}{(p,H_kp)},\\ p=\frac{(g_k,H_kg_{k+1})}{(g_{k+1},H_kg_{k+1})}\,g_{k+1}-g_k.\end{gathered}\qquad(1.36)$$

Тут $g_k=\partial f(x_k)$ і $g_{k+1}=\partial f(x_{k+1})$ — субградієнти $f(x)$ у послідовних точках. Для H-форми, яка відповідає методу (1.29)–(1.34), формули для перерахунку матриці H_{k+1} трохи ускладняться.

Окрім економії пам'яті, необхідної для зберігання матриці, використання H-форми призводить також до більш економних методів в обчислювальному плані, ніж вищенаведені B-методи.

Зокрема, дворангова корекція матриці H_{k+1}, враховуючи її симетричність, вимагає не більше операцій, ніж одноранговa корекція матриці B_{k+1}, а замість двох множень матриці на вектор, яких вимагають методи (1.25)–(1.28) та (1.29)–(1.34), для H-форми достатньо й одного.

Цілком можливо, що і методика обґрунтування H-методів може бути не менш гарною, ніж B-методів. Однак, ці питання вимагають спеціального дослідження, яке тут проводитись не буде. Зазначимо лише, що формули (1.35) та (1.36) для перерахунку H_{k+1} дуже нагадують перерахунок аналогічних матриць у квазіньютонівських методах [5].

Числові експерименти. Оскільки методи (1.25)–(1.28) та (1.29)–(1.34) ідентичні методам (28)–(33) та (37)–(44) з [8], то немає особливої потреби в числовому дослідженні їхньої поведінки залежно від розмірів задач. Можна вважати, що для справедливі результати, наведені в [8] (табл. 1 і 2). Хоча тут потрібне деяке застереження. Справа в тому, що набір тестових задач у [8] характеризується сильною яружністю. Враховуючи, що різним B-методам відповідають різні способи накопичення помилок у матриці B_k, чисельні розрахунки можуть не співпадати повністю, тобто можливі незначні розбіжності.

Тому в першій серії експериментів обмежимося задачами невеликих розмірів $n \sim 5-10$ і перевіримо числову стійкість вищенаведених методів відносно точності розв'язання задач за функціоналом. Як тестові негладкі задачі виберемо два найчастіше згадувані приклади кусково-квадратичних функцій:

$$\mathsf{Shor}(n = 5, \quad f^* = 22.6001620958) \text{ та}$$

$$\mathsf{Maxquad}(n = 10, \quad f^* = -0.841408334596),$$

які вважаються поганими задачами для субградієнтних методів. Така скрупульозна вказівка f^* для задач Shor і $\mathsf{Maxquad}$ обумовлена тим, щоб перевірити роботу методів при досить малих значеннях ε_f. Як тестові гладкі задачі виберемо квадратичні задачі $\mathsf{Quad}(t)$ з [8] з різним ступенем яружності ($\mathsf{Quad}(10.)$ і $\mathsf{Quad}(3.)$) і при різних розмірах ($n = 5$ і $n = 10$). Тут $f^* = 0$. Всі ці задачі характеризує єдина точка мінімуму, але для негладких задач вона визначена точністю задання f^*. Як ε_f візьмемо ε_0 та ε_0^2. Тут $\varepsilon_0 = 10^{-5}$ для негладких задач та 10^{-10} для квадратичних задач.

Результати роботи методів наведені в таблиці 1.1. Тут $\mathsf{iter}(\varepsilon)$ – кількість обчислень значень функції та її субградієнта, необхідне для досягнення точності ε за функціоналом.

У дужках наведено кількість перетворень простору, які при цьому реалізуються. Як видно з таблиці 1.1, робота методів (1.25)–(1.28) та (1.29)–(1.34) не настільки безнадійна, як робота класичного фейєрівського методу без перетворення простору.

Це певною мірою підтверджує, що лінійне перетворення простору, спрямоване на вирівнювання поверхонь рівня функції, здатне значно збільшити точність розв'язання задач за функціоналом для ітераційних методів фейєрівського типу.

Таблиця 1.1: ε_f-збіжність методів (1.25)–(1.28) та (1.29)–(1.34)

Задача	n	метод (1.25)–(1.28)		метод (1.29)–(1.34)	
		$\mathrm{iter}(\varepsilon_0)$	$\mathrm{iter}(\varepsilon_0^2)$	$\mathrm{iter}(\varepsilon_0)$	$\mathrm{iter}(\varepsilon_0^2)$
Shor	5	112(109)	227(224)	38(36)	70(68)
$\mathrm{Maxquad}$	10	120(113)	293(286)	41(35)	85(79)
$\mathrm{Quad}(3)$	5	40(11)	73(11)	40(11)	73(11)
$\mathrm{Quad}(3)$	10	82(60)	115(74)	76(59)	109(80)
$\mathrm{Quad}(10)$	5	60(22)	93(22)	57(21)	90(21)
$\mathrm{Quad}(10)$	10	187(141)	220(152)	148(124)	181(141)

Таблиця 1.2: ε_f-збіжність методу (1.29)–(1.34)

Задача	n	$f(x_0)-f^*$	$\mathrm{iter}(10^{-5})$	$\mathrm{iter}(10^{-10})$	$\mathrm{iter}(10^{-20})$
$\mathrm{Quad}(1.1)$	50	581.955	42(32)	65(49)	102(73)
$\mathrm{Sabs}(1.1)$	50	1163.909	176	279	347
$\mathrm{Quad}(1.05)$	100	1305.013	51(43)	79(65)	124(97)
$\mathrm{Sabs}(1.05)$	100	2610.025	318	424	614

Витрачені зусилля на розв'язання цих задач з точністю $\varepsilon_f = \varepsilon_0^2$ за функціоналом часто вдвічі більші, ніж з точністю $\varepsilon_f = \varepsilon_0$. Однак, цей розрив суттєво скорочується зі збільшенням розмірів задач. Про це свідчить друга серія експериментів з методом (1.29)–(1.34) для слабо яружних задач $\mathrm{Quad}(t)$ та $\mathrm{Sabs}(t)$ з [8]. Результати наведено у таблиці 1.2. Для задах $\mathrm{Sabs}(t)$ перетворення простору реалізується практично на кожному кроці методу (1.29)–(1.34). Тому для них у таблиці 1.2 наведено лише загальне число ітерацій.

Деякі висновки та загальні зауваження. Отже, однорангове перетворення простору — гідна заміна дворангового перетворення в тому розумінні, що B-методи на його основі вимагають меншої кількості операцій. Воно дозволяє будувати різноманітні методи змінної метрики, що мають просту інтерпретацію, і допускає зручний механізм доведення методів у B-формі, що базується на зовнішній апроксимації множини екстремумів, яка зменшується за об'ємом, еліпсоїдом X, який є образом кулі в Y_k. Цей факт дозволяє отримати задовільні відповіді на низку питань в опуклому програмуванні.

Зокрема, теорема 1.3 вказує про принципову можливість побудови конструктивного критерію зупинки для несумісних систем опуклих нерівностей шляхом аналізу послідовності $\sum_{k=0}^{\infty} h_k^2$. Якщо $\sum_{k=0}^{K} h_k^2 \geq \ r^2$ при деякому K, це означає, що система опуклих нерівностей не має допустимої точки в кулі радіусу r з центром в точці x_0. Отже, вищенаведені методи за скінчену кількість кроків або знаходять допустиму точку системи опуклих нерівностей, або одержують достатню умову її несумісності.

При побудові процесів знаходження мінімуму опуклої функції $f(x)$ при невідомому значенні f^* радіус кулі, що локалізує множину екстремумів у перетвореному просторі, дозволяє простий механізм уточнення нижньої оцінки f^*. Для отримання верхньої оцінки f^* можна використовувати як її «грубе» уточнення шляхом поєднання з кроками, близькими до класичного методу еліпсоїдів, і більш «точне» за допомогою процедур типу найшвидшого спуску. Як перший, так і другий шлях дозволяє для доведення збіжності методів зберегти критерій зменшення об'єму області локалізації множини екстремумів. Але, якщо перший у низці випадків дозволяє зберегти також можливість руху антисубградієнтом і приводить до субградієнтних методів фейєрівського типу зі зміною початкового наближення, то другий вимагатиме зміни напрямку дослідження поведінки функції (наприклад, як у ε-субградієнтних методах) і приведе до методів на кшталт спряжених градієнтів, але у перетвореному просторі аргументів.

Зокрема, керування параметром γ: $0 \leq \gamma < 1$ і для фейєрівських процесів при відомому f^* дозволяє будувати методи, близькі до методів на кшталт спряжених градієнтів.

Наприклад, як у [19], тобто як напрямок руху з точки в перетвореному просторі використовується агрегатний вектор, який є опуклою комбінацією обчислених раніше субградієнтів і задає кращий напрямок на точку оптимуму, ніж останній субградієнт. Це приводить до досить розумних схем методів навіть в умовах мінімуму інформації, яка використовується.

Зокрема, для одного з них достатньо використовувати останній обчислений субградієнт та деякий агрегатний вектор, по якому здійснюється рух з точки. Для другого достатньо двох агрегатних векторів та останнього субградієнта. При цьому автоматично оновлюватимуться вектори агрегатного типу і замінюватимуться останніми обчисленими субградієнтами.

Враховуючи, що при цьому також є можливість обчислювати субградієнт у точці, в яку перенесені вектори агрегатного типу, в обчислювальному розумінні ці методи можуть виявитися кращими, ніж метод (1.29)–(1.34).

Але, в будь-якому випадку, особливо для задач великих розмірів, не обійтися без деякої процедури накопичення субградієнтів, які б характеризували яружність функції. Очевидно, що метод (1.29)–(1.34) можна поліпшити за допомогою використання великої кількості обчислених раніше субградієнтів з переведенням їх у черговий перетворений простір аргументів. Однак, при цьому складно будувати конструктивні правила відсіву зайвих субградієнтів. Тому доцільно будувати схеми методів так, щоб механізм відсіву був простим. Найкраще для цього підходить процедура послідовного проектування деякої точки в перетвореному просторі, наприклад центру кулі, яка локалізує множину екстремумів, на множину як у (1.4), починаючи з обчисленого в цій точці $\varphi(y)$ та $\partial\varphi(y)$, і до тих пір, поки задача проектування буде розв'язуватись аналітично. Зрозуміло, що таким чином не можна накопичити більше ніж $m \leq (n+1)$ субградієнтів, перші $(m-1)$ з яких задаватимуть конус, який локалізує множину екстремумів, а m-й субградієнт буде обчислено у вершині цього конуса, що знаходиться на найкоротшій відстані від центра кулі. Далі переходимо у вершину конуса і за допомогою перетворення (1.11) позбавляємося тупих кутів між субградієнтами. Використовуючи як нову точку вершину конуса і обчислений у ній субградієнт, повторюємо зазначену процедуру. Якщо при цьому точку, що проектується, замінювати на вершину поточного конуса, то отримаємо процедуру накопичення субградієнтів як у методі ортогонального спуску [16], при цьому кут між будь-якими двома субградієнтами буде негострим. Такі схеми методів приводять до досить простого розв'язання проблем проектування у фейєрівських методах, проблем відсіву зайвих субградієнтів, і, крім того, дозволяють забезпечити більш розумні поверхні рівня функції в наступних перетворених просторах. Досить стійка робота методу (1.29)–(1.34) дозволяє сподіватися на ефективність роботи таких методів для задач чималих розмірів.

1.4. Доортогоналізуючий одноранговий оператор

Одноранговий оператор не дозволяє просто працювати з більше ніж двома-трьома векторами. Нехай $K(g_1, g_2, \ldots, g_m)$, $m \leq n$ — гострокутний конус, який локалізує множину екстремумів. Тут $g_1, g_2, \ldots, g_m$ — лінійно-незалежні субградієнти, причому $(g_i, g_j) < 0$ для всіх $i = \overline{1, m}$, $j = \overline{1, m}$, таких, що $i \neq j$. Для ефективного перетворення конуса $K(g_1, g_2, \ldots, g_m)$ за допомогою однорангового оператора потрібно в кожному з перетворених просторів розв'язувати задачі занурення такого типу конусів в простіші конуси $K(p_1, p_2)$, такі, що $(p_1, p_2) < 0$. Ці задачі не є проблемою, а їх розв'язання потребує незначних обчислювальних витрат. Однак, цього можна уникнути, якщо на кроці методу забезпечити роботу з простішим конусом, наприклад, ортогональним.

Таку можливість легко реалізувати і для перетворення простору можна використовувати лінійний неортогональний оператор, описаний нижче.

Доортогоналізуючий одноранговий оператор. Нехай в $\mathbb{R}^n$ задано набір векторів

$$P = \{p_1, p_2, ..., p_m\}, \quad \|p_i\| = 1, \quad i = 1, \ldots, m, \quad m \leq n, \tag{1.37}$$

для яких виконуються такі умови:

$$(p_i, p_j) = 0, \quad i \neq j, \quad i = 1, 2, \ldots, m-1, \quad j = 1, 2, \ldots, m-1, \tag{1.38}$$

$$\left\| p_m - \sum_{i=1}^{m-1} (p_m, p_i) p_i \right\|^2 = 1 - \sum_{i=1}^{m-1} (p_m, p_i)^2 > 0. \tag{1.39}$$

тобто вектори з P лінійно незалежні, причому перші $(m-1)$ векторів взаємно-ортогональні, а p_m не обов'язково ортогональний до всіх попередніх.

Позначимо

$$p = \sum_{i=1}^{m-1} (p_m, p_i) p_i.$$

Для векторів (1.37), що задовольняють (1.38) і (1.39), виконуються рівності:

$$\begin{aligned} &a) \quad (p, p_m - p) = 0; \\ &b) \quad (p_i, p_m - p) = 0, \quad i = 1, \ldots, m-1; \\ &c) \quad (p_m, p_m - p) = \|p_m - p\|^2. \end{aligned} \tag{1.40}$$

Ними часто користуватимемося надалі.

Розглянемо лінійний оператор з $\mathbb{R}^n$ в $\mathbb{R}^n$, який у матричному вигляді представимо так:

$$T_\lambda(p_m, p) = I + \frac{p_m - p}{\|p_m - p\|^2} \left(\frac{1}{\lambda} p_m + p \right)^T, \tag{1.41}$$

де I – одинична матриця розмірності $n \times n$, λ – деякий скаляр, такий, що

$$\lambda(\lambda + 1) \neq 0.$$

Вибір параметра λ і його сенс обговоримо трохи пізніше. Оператор (1.41) будемо називати доортогоналізуючим одноранговим оператором. Для нього справедливі такі твердження.

Лема 1.4. *Для оператора* $T_\lambda(p_m, p)$ *при* $\lambda(\lambda+1) \neq 0$ *існує обернений* $T_\lambda^{-1}(p_m, p)$ *і*

$$T_\lambda^{-1}(p_m, p) = I - \frac{p_m - p}{\|p_m - p\|^2}\Big(\frac{1}{\lambda+1}p_m + \frac{\lambda}{\lambda+1}p\Big)^T. \tag{1.42}$$

Крім того,

$$\begin{aligned} &T_\lambda^T(p_m, p)\, T_\lambda(p_m, p) = \\ &\qquad = I + \frac{1}{\|p_m - p\|^2}\left(\Big(\frac{2}{\lambda} + \frac{1}{\lambda^2}\Big)p_m p_m^T - pp^T + p_m p^T + p p_m^T\right). \end{aligned} \tag{1.43}$$

Доведення. Справедливість (1.42) та (1.43) легко показати безпосередньою перевіркою. Зокрема, враховуючи, що в силу (1.40a) та (1.40c), має місце $\left(\frac{1}{\lambda}p_m + p,\ p_m - p\right) = \frac{1}{\lambda}\,(p_m,\ p_m - p) + (p,\ p_m - p) = \frac{1}{\lambda}\|p_m - p\|^2$, для перевірки (1.42) маємо

$$\begin{aligned} T_\lambda(p_m, p)\, T_\lambda^{-1}(p_m, p) &= \left(I + \frac{p_m - p}{\|p_m - p\|^2}\Big(\frac{1}{\lambda}p_m + p\Big)^T\right) \times \\ &\qquad \times \left(I - \frac{p_m - p}{\|p_m - p\|^2}\Big(\frac{1}{\lambda+1}p_m + \frac{\lambda}{\lambda+1}p\Big)^T\right) = \\ = \left(I + \frac{p_m - p}{\|p_m - p\|^2}\Big(\frac{1}{\lambda}p_m + p\Big)^T\right) &\left(I - \frac{\lambda}{\lambda+1}\frac{p_m - p}{\|p_m - p\|^2}\Big(\frac{1}{\lambda}p_m + p\Big)^T\right) = \\ = I + \left(1 - \frac{\lambda}{\lambda+1} - \frac{\lambda}{\lambda+1}\frac{\left(\frac{1}{\lambda}p_m + p, p_m - p\right)}{\|p_m - p\|^2}\right)&\frac{p_m - p}{\|p_m - p\|^2} \times \\ &\times \left(\frac{1}{\lambda}p_m + p\right)^T = \\ = I + \left(\frac{1}{\lambda+1} - \frac{\lambda}{\lambda+1}\,\frac{1}{\lambda}\,\frac{p_m - p}{\|p_m - p\|^2}\right)&\frac{p_m - p}{\|p_m - p\|^2}\left(\frac{1}{\lambda}p_m + p\right)^T = I. \end{aligned}$$

Справедливість (1.43) випливає з

$$\begin{aligned} &T_\lambda^T(p_m, p) T_\lambda(p_m, p) = \\ &= \left(I + \left(\frac{1}{\lambda}p_m + p\right)\left(\frac{p_m - p}{\|p_m - p\|^2}\right)^T\right)\left(I + \frac{p_m - p}{\|p_m - p\|^2}\left(\frac{1}{\lambda}p_m + p\right)^T\right) = \end{aligned}$$

$$
\begin{aligned}
=&I + \frac{1}{\|p_m - p\|^2}\left(\left(\frac{1}{\lambda}p_m + p\right)(p_m - p) + (p_m - p)\left(\frac{1}{\lambda}p_m + p\right)^T + \right.\\
&\left. + \left(\frac{1}{\lambda}p_m + p\right)\left(\frac{1}{\lambda}p_m + p\right)^T\right) = \\
=&I + \frac{1}{\|p_m - p\|^2}\left(\frac{2}{\lambda} + \frac{1}{\lambda^2}\right)p_m p_m^T + (-1 - 1 + 1)pp^T + \\
&+ \left(-\frac{1}{\lambda} + \frac{1}{\lambda+1}\right)p_m p^T + \left(\frac{1}{\lambda} - \frac{1}{\lambda+1}\right)pp_m^T = \\
=&I + \frac{1}{\|p_m - p\|^2}\left(\left(\frac{2}{\lambda} + \frac{1}{\lambda^2}\right)p_m p_m^T - pp^T + p_m p^T + pp_m^T\right).
\end{aligned}
$$

Справедливість (1.42) та (1.43) показали. □

Лема 1.5. *Нехай набір векторів (1.37) задовольняє (1.38) та (1.39), і* $\tilde{P} = \{\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_m\}$ *— набір векторів, отриманих з (1.37), як* $\tilde{p}_i = (T_\lambda^{-1}(p_m, p))^T p_i$ *для всіх* $i = \overline{1, m}$ *при* $\lambda(\lambda + 1) \neq 0$*. Тоді жоден з векторів* $\tilde{p}_i$ *тотожно не рівний нульовому вектору,* $(\tilde{p}_i, \tilde{p}_j) = 0$*,* $i \neq j$*,* $i = \overline{1, m}$*,* $j = \overline{1, m}$*, і, окрім того,*

$$
\tilde{p}_i = p_i, \quad i = \overline{1, m-1}, \quad \tilde{p}_m = \tfrac{\lambda}{\lambda+1}(p_m - p). \tag{1.44}
$$

Доведення. Перерахунок векторів p_i для всіх $i = \overline{1, m}$ виконується згідно з співвідношенням

$$
\begin{aligned}
\tilde{p}_i = (T_\lambda^{-1}(p_m, p))^T p_i \;&=\; \left(I - \Big(\frac{1}{\lambda+1}p_m + \frac{\lambda}{\lambda+1}p\Big)\frac{(p_m - p)^T}{\|p_m - p\|^2}\right)p_i = \\
&=\; p_i - \Big(\frac{1}{\lambda+1}p_m + \frac{\lambda}{\lambda+1}p\Big)\frac{(p_m - p, p_i)}{\|p_m - p\|^2}.
\end{aligned}
$$

Звідси випливає справедливість першої частини (1.44), оскільки $(p_m - p, p_i) = 0$ для всіх $i = \overline{1, m-1}$ в силу властивості (1.40b). Для $\tilde{p}_m$ з урахуванням (1.40c) отримуємо

$$
\tilde{p}_m = p_m - \Big(\frac{1}{\lambda+1}p_m + \frac{\lambda}{\lambda+1}p\Big)\frac{\|p_m - p\|^2}{\|p_m - p\|^2} = \frac{\lambda}{\lambda+1}(p_m - p).
$$

Той факт, що жоден з векторів $\tilde{p}_i$, $i = \overline{1, m}$, тотожно не дорівнює нульовому вектору, випливає з того, що $\|\tilde{p}_i\| = \|p_i\| = 1$, $i = \overline{1, m-1}$, а також з умови (1.39) і $\lambda(\lambda + 1) \neq 0$, що рівносильно $\|\tilde{p}_m\| \neq 0$.

Далі, $(\tilde{p}_i, \tilde{p}_j) = (p_i, p_j) = 0$, $i = \overline{1, m-1}$, $j = \overline{1, m-1}$, $i \neq j$, згідно з (1.38) і $(\tilde{p}_m, \tilde{p}_i) = \dfrac{\lambda}{\lambda+1}(p_m - p, p_i) = 0$ для всіх $i = \overline{1, m-1}$ згідно з (1.40b). $\square$

Отже, оператор (1.41) допускає однорангову корекцію матриці B_{k+1} при переході в черговий перетворений простір аргументів і, більше того, забезпечує достатньо зручний механізм роботи з конусом, який визначається скінченим набором векторів $m \leq n$. Його використання дозволяє будувати конструктивні методи змінної метрики з досить наочною геометричною інтерпретацією у перетвореному просторі аргументів.

1. Якщо як вектори вибирати субградієнти, можна отримати різноманітні методи, на кшталт спряжених градієнтів для гладких функцій. Зокрема, для безумовної мінімізації додатно визначених квадратичних функцій проста процедура послідовного накопичення градієнтів з їхньою доортогоналізацією приводить до методів цього типу, які вимагають не більше ніж $(n+1)$ обчислень значення функції та градієнта. При цьому за допомогою вибору кроків необхідно лише гарантувати лінійну незалежність векторів $y_0 - y_k$, $k = \overline{1, m}$, $m \leq n$, де $y_0 = A_m x_0$, $y_k = A_m x_k$, а на останньому m-му кроці методу, коли отримано лінійно залежний градієнт, для знаходження точки мінімуму розв'язати в перетвореному просторі просту систему лінійних рівнянь з діагональною матрицею розмірності $m \times m$. Керування параметром λ в операторі (1.41) дозволяє зберігати додатну визначеність матриці $H_k = B_k B_k^T$ та керувати ступенем її виродженості. Це дозволяє будувати методи на кшталт спряжених градієнтів для негладких задач.

2. Використання скаляра $\left(p_m, \dfrac{p}{\|p\|}\right)$ задає досить зручний механізм для уточнення нижньої оцінки f^* у перетвореному просторі аргументів. Ця обставина дозволяє забезпечити конструктивний критерій зупинки у розумінні наближеного виконання умов оптимальності задач опуклого програмування. Його можна використовувати для побудови методів змінної метрики, використовуючи процедури на кшталт ε-субградієнтних [20], але у перетвореному просторі аргументів. Враховуючи, що при певному значенні параметра λ зберігається зовнішня апроксимація множини екстремумів у перетвореному просторі кулею без збільшення радіусу, такі процедури підходять і для реалізації внутрішнього алгоритму МЦВПТ [9].

3. За допомогою перетворення (1.41) можна спробувати отримати обчислювально стійкі методи для розв'язання систем лінійних рівнянь в рамках методів ортогоналізації, які зарекомендували себе чисельно нестійкими [21]. При цьому, окрім самого перетворення, є низка механізмів як на рівні

B-форми, так і H-форми, для забезпечення стійкого перерахунку відстаней до гіперплощин у перетвореному просторі аргументів.

4. Перетворення (1.41) дозволяє просто будувати методи змінної метрики, які базуються на локалізації множини екстремумів конусом в $\mathbb{R}^n$. На одному з таких варіантів методів для розв'язання задачі (1.1), який допускає класичний фейєрівський крок у напрямку антисубградієнта, і зосередимося в цьому розділі. В його основі лежить простий геометричний факт. Нехай y_k – вершина ортогонального конуса $K(g_1, \dots, g_m)$, який локалізує множину екстремумів, і g_k — обчислений у ній субградієнт. Нехай $K_1(g'_1, \dots, g'_{m_1}, g_k)$ — побудований з урахуванням цієї інформації черговий конус з вершиною в y_k. Тут $g'_1, \dots, g'_{m_1}$ — ті з ортогональних субградієнтів конуса $K(g_1, \dots, g_m)$, для яких виконується нерівність $(g_i, g_k) < 0$. За побудовою цей конус міститиме множину екстремумів задачі (1.1). Конус $K_1(g'_1, \dots, g'_{m_1}, g_k)$ — нетупокутний, а його ребра утворюють між собою гострі кути. Перетворення його в ортогональний конус природно використовувати для покращення структури поверхонь рівня яружних функцій.

Оператор (1.41) і така покрокова стратегія побудови конуса приводить до низки методів змінної метрики, які базуються на локалізації множини екстремумів у перетвореному просторі ортогональним конусом, який визначається обчисленими раніше субградієнтами. При цьому автоматично забезпечується вирішення проблеми обмеження кількості субградієнтів, які зберігаються. Більш того, їхнє відсіювання відбувається не простим «забуванням», а кожен з цих векторів робить певний внесок у перетворення простору, яке спрямоване на розширення конуса можливих напрямків спадання функції.

Для процесів з класичним фейєрівським кроком у напрямку антисубградієнта та описаною стратегією побудови конуса, який локалізує множину екстремумів, є дві можливості їхньої реалізації. Перша — спочатку робити перетворення простору, а потім у перетвореному просторі реалізовувати класичний феєрівський крок у напрямку $\tilde{g}_k$. Тут $\tilde{g}_k$ — образ субградієнта g_k у перетвореному просторі аргументів. Друга — спочатку робити класичний фейєрівський крок у напрямку g_k, а потім перетворювати простір аргументів. В обох випадках отримаємо локалізацію множини екстремумів ортогональним конусом у перетвореному просторі аргументів. У першому випадку для всіх перенесених в поточну точку обчислених раніше субградієнтів буде реалізовано максимальний зсув по опуклості $f(x)$. У другому — кожен з перенесених у поточну точку субградієнтів, крім останнього, допускає додатний крок у напрямку антисубградієнта так, щоб не порушити локалізації множини екстремумів ортогональним конусом. Отже, у другому випадку метод можна покращити у розумінні локалізації множини екстремумів за рахунок кроку по опуклій комбінації ортогональних субградієнтів, не включаючи останнього. Це приводить до напрямку руху, відмінного від антисубградієнта, і другий шлях не розглядатимемо. Очевидно, що без кроку по опуклій комбінації попередніх субградієнтів він програватиме першому.

Зупинимось на першому варіанті та описаній покроковій стратегії побудови конуса. Такі процеси для розв'язання задачі (1.1) називатимемо методами ортогонального субградієнтного спуску з класичним феєрівським кроком через те, що рух з точки в перетвореному просторі визначається субградієнтом у ній, який ортогональний до накопичених субградієнтів.

Метод ортогонального субградієнтного спуску з класичним фейєрівським кроком (ORTGF). Перш ніж переходити до його опису, визначимо ще деякі параметри, які мають суто практичне значення та конкретизують правило побудови чергового конуса. Зокрема, окрім точності розв'язання задачі (1.1) за функціоналом — $\varepsilon_f > 0$, використовуватимемо ще два досить малі додатні скаляри: ε_K та ε_R, а також параметр $m_0 \le n-1$. Тут $\varepsilon_K > 0$ визначатиме правило відсіву субградієнтів, тобто при переході до $(k+1)$-го кроку залишатимемо тільки ті ортогональні субградієнти, для яких виконується умова $\left(\frac{g_i}{\|g_i\|}, \frac{g_k}{\|g_k\|}\right) < -\varepsilon_K$, $\varepsilon_R > 0$ визначатиме додатковий відсів субградієнтів, щоб уникнути накопичення помилок обчислень у процесі доортогоналізації, тобто ті ортогональні субградієнти, для яких $\left|\left(\frac{g_i}{\|g_i\|}, \frac{g_j}{\|g_j\|}\right)\right| > \varepsilon_R$, виключатимемо. Параметр m_0 задаватиме обмеження на максимальну кількість субградієнтів, які зберігаються, і це будуть субградієнти обчислені останніми. Множину субградієнтів, які зберігаються, на k-му кроці позначатимемо P_k, а її розмір, тобто кількість накопичених субградієнтів — $size(P_k)$. Зокрема, якщо $P_k = \emptyset$, то $size(P_k) = 0$. Для роботи з P_k використовуватимемо операції суми ($\cup$) та різниці ($\setminus$).

Для спрощення викладення матеріалу підемо на ще деякі обмеження у методі ORTGF. Зокрема, зберігатимемо тільки нормовані субградієнти і, крім того, параметр λ вважатимемо постійним. Хоча слід зауважити, що використовуючи норми субградієнтів та змінний параметр λ_k, можна забезпечити більш змістовні методи ортогонального субградієнтного спуску. Наприклад, за допомогою використання оператора розтягу простору та оператора (1.41) можна забезпечити перетворення простору таким чином, щоб напрямок руху в Y_k збігався з найкоротшим вектором до опуклої оболонка накопичених субградієнтів. Такий напрямок руху використовується в ε-субградієнтних методах і вважається вдалою заміною ньютонівському напрямку.

Враховуючи вищесказане, $\mathsf{ORTGF}(x_0, \varepsilon_f, \lambda, \varepsilon_K, \varepsilon_R, m_0)$ — метод ортогонального субградієнтного спуску з класичним фейєрівським кроком — набуває такого вигляду.

Ініціалізація: $B_0 = I_n$, $P_0 = \emptyset$.

Нехай на k-му кроці отримано x_k, B_k та P_k. Обчислимо $f(x_k)$ та $\partial f(x_k)$. Якщо $f(x_k) - f^* \le \varepsilon_f$, то x_k — шукана точка та ЗУПИНКА. Інакше переходимо до $(k+1)$-го кроку.

1) Покладемо $\xi_k = \dfrac{B_k^T \partial f(x_k)}{\|B_k^T \partial f(x_k)\|}$ та $h_k = \dfrac{f(x_k) - f^*}{\|B_k^T \partial f(x_k)\|}$.

2) Сформуємо множину $\tilde{P}_k = \{p_i \in P_k : \ (p_i, \xi_k) < -\varepsilon_K\}$, зберігаючи порядок слідування p_i в P_k.

3) Якщо $\mathit{size}(P_k) = 0$, то покладемо $B_{k+1} = B_k, \xi_{k+1} = \xi_k, h_{k+1} = h_k$ та переходимо до пункту 4. Інакше обчислюємо вектор $\tilde{p}_k = \sum\limits_{p_i \in \tilde{P}_k} (p_i, \xi_k) p_i$

та перераховуємо параметри: $B_{k+1} = B_k T_\lambda^{-1}(\xi_k, \tilde{p}_k)$,

$$\xi_{k+1} = \frac{B_{k+1}^T \partial f(x_k)}{\|B_{k+1}^T \partial f(x_k)\|}, \quad h_{k+1} = \frac{f(x_k) - f^*}{\|B_{k+1}^T \partial f(x_k)\|}.$$

В обчислювальному плані це рівносильно $B_{k+1} = B_k(I - \eta_1 \eta_2^T)$, де

$$\eta_1 = \frac{\xi_k - \tilde{p}_k}{\|\xi_k - \tilde{p}_k\|^2}, \qquad \eta_2 = \frac{1}{\lambda+1}\xi_k + \frac{\lambda}{\lambda+1}\tilde{p}_k.$$

$$\xi_{k+1} = \frac{\frac{\lambda}{\lambda+1}(\xi_k - \tilde{p}_k)}{\|\frac{\lambda}{\lambda+1}(\xi_k - \tilde{p}_k)\|}, \qquad h_{k+1} = \frac{h_k}{\|\frac{\lambda}{\lambda+1}(\xi_k - \tilde{p}_k)\|}.$$

4) Обчислимо наступне наближення $x_{k+1} = x_k - h_{k+1} B_{k+1} \xi_{k+1}$.

5) Сформуємо наступну множину

$$P_{k+1} = \{p_i \in \tilde{P}_k : \ |(p_i, \xi_{k+1})| < \varepsilon_R\} \cup \xi_{k+1},$$

згідно з порядком слідування p_i в $\tilde{P}_k$, ξ_{k+1} буде останнім вектором в P_{k+1}. Якщо $\mathit{size}(P_{k+1}) > m_0$, то $P_{k+1} = P_{k+1} \setminus \{p_1 \in P_{k+1}\}$, де p_1 – перший вектор в P_{k+1}.

6) Переходимо до наступного кроку з x_{k+1}, B_{k+1}, P_{k+1}.

Теорема 1.5. *Для* k*-го кроку методу* ORTGF *виконується*

$$\big(A_k(x_k - x^*), \xi_k\big) \geq \frac{f(x_k) - f^*}{\|B_k^T \partial f(x_k)\|} \geq 0, \tag{1.45}$$

де $A_k = B_k^{-1}$. *Коли* $P_k \neq \emptyset$, *то* $\|p_i\| = 1$ *для всіх* $p_i \in P_k$*;* $(p_i, p_j) = 0$ *для* $p_i, p_j \in P_k$ *таких, що* $i \neq j$*; окрім того, виконуються нерівності*

$$(A_k(x_k - x^*), p_i) \geq 0, \quad p_i \in P_k, \tag{1.46}$$

$$(A_k(x_k - x^*), \tilde{p}_k) \leq 0. \tag{1.47}$$

Доведення. Справедливість (1.45) випливає з

$$\begin{aligned}(A_k(x_k - x^*), \xi_k) &= \left(A_k(x_k - x^*), \frac{B_k^T \partial f(x_k)}{\|B_k^T \partial f(x_k)\|}\right) = \\ &= \frac{(x_k - x^*, \partial f(x_k))}{\|B_k^T \partial f(x_k)\|} \geq \frac{f(x_k) - f^*}{\|B_k^T \partial f(x_k)\|}\end{aligned}$$

через опуклість $f(x)$.

Ортогональність та нормованість векторів у $P_k = \tilde{P}_{k-1} \cup \xi_k$ забезпечує оператор $T_\lambda(\xi_{k-1}, \tilde{p}_{k-1})$ за лемою 1.5, оскільки вектори в $\tilde{P}_{k-1}$ ортогональні та нормовані, а $\xi_k = \dfrac{\big(T_\lambda^{-1}(\xi_{k-1}, \tilde{p}_{k-1})\big)^T \xi_{k-1}}{\| \big(T_\lambda^{-1}(\xi_{k-1}, \tilde{p}_{k-1})\big)^T \xi_{k-1}\|}$.

Для тієї частини P_k, яка сформована з $\tilde{P}_{k-1}$, справедливість (1.46) випливає за індукцією. Дійсно, нехай

$$(A_{k-1}(x_{k-1} - x^*), p_i) \geq 0, \quad p_i \in P_{k-1}.$$

Враховуючи, що за лемою 1.5 $\big(T_\lambda^{-1}(\xi_{k-1}, \tilde{p}_{k-1})\big)^T p_i = p_i$ та $(\xi_k, p_i) = 0$ для всіх $p_i \in \tilde{P}_{k-1}$, маємо

$$\begin{aligned}&(A_k(x_k - x^*), p_i) = (A_k(x_{k-1} - x^* - h_k B_k \xi_k), p_i) = \\ &= (A_k(x_{k-1} - x^*) - h_k \xi_k, p_i) = (A_k(x_{k-1} - x^*), p_i) - h_k(\xi_k, p_i) = \\ &= (A_k(x_{k-1} - x^*), p_i) = \Big(A_k(x_{k-1} - x^*), \big(T_\lambda^{-1}(\xi_{k-1}, \tilde{p}_{k-1})\big)^T p_i\Big) = \\ &= \big(T_\lambda^{-1}(\xi_{k-1}, \tilde{p}_{k-1}) A_k(x_{k-1} - x^*), p_i\big) = (A_{k-1}(x_{k-1} - x^*), p_i),\end{aligned}$$

звідки випливає справедливість (1.46) для тієї частини P_k, яка сформована з $\tilde{P}_{k-1}$.

Для ξ_k, оскільки $\xi_k = \frac{B_k^T \partial f(x_{k-1})}{\|B_k^T \partial f(x_{k-1})\|}$ і $h_k = \frac{f(x_{k-1}) - f^*}{\|B_k^T \partial f(x_{k-1})\|}$, справедливість (1.46) випливає з

$$
\begin{aligned}
& \Big(A_k(x_k - x^*),\ \xi_k\Big) = \Big(A_k(x_{k-1} - x^* - h_k B_k \xi_k),\ \xi_k\Big) = \\
= \ & \Big(A_k(x_{k-1} - x^*) - h_k \xi_k, \xi_k\Big) = \Big(A_k(x_{k-1} - x^*),\ \xi_k\Big) - h_k = \\
= \ & \Big(A_k(x_{k-1} - x^*),\ \frac{B_k^T \partial f(x_{k-1})}{\|B_k^T \partial f(x_{k-1})\|}\Big) - \frac{f(x_{k-1}) - f^*}{\|B_k^T \partial f(x_{k-1})\|} = \\
= \ & \frac{\big(x_{k-1} - x^*,\ \partial f(x_{k-1})\big) - \big(f(x_{k-1}) - f^*\big)}{\|B_k^T \partial f(x_{k-1})\|} \geq 0
\end{aligned}
$$

через опуклість $f(x)$.

Нарешті справедливість (1.47) забезпечують нерівності (1.46) та спосіб формування множини $\tilde{P}_k$, який для вектора $\tilde{p}_k$ гарантує вибір від'ємних коефіцієнтів за відповідних p_i. □

Теорема 1.6. *Послідовність* $\{x_{k+1}\}_{k=0}^{\infty}$, *що генерується алгоритмом* ORTGF *при* $\lambda = -1/2$, *задовольняє нерівності*

$$
\|A_{k+1}(x_{k+1} - x^*)\|^2 \leq \|A_k(x_k - x^*)\|^2 - \frac{(f(x_k) - f^*)^2}{\|B_k^T \partial f(x_k)\|^2}. \tag{1.48}
$$

Тут $A_k = B_k^{-1}$, $A_{k+1} = B_{k+1}^{-1}$, $k \geq 0$.

Доведення. Якщо $P_k = \varnothing$, то $\xi_{k+1} = \xi_k = \frac{B_k^T \partial f(x_k)}{\|B_k^T \partial f(x_k)\|}$, $B_{k+1} = B_k$, $h_{k+1} = h_k = \frac{f(x_k) - f^*}{\|B_k^T \partial f(x_k)\|}$, і нерівність (1.48) випливає з

$$
\begin{aligned}
& \|A_{k+1}(x_{k+1} - x^*)\|^2 = \\
= \ & \|A_k(x_k - x^* - h_k B_k \xi_k)\|^2 = \|A_k(x_k - x^*) - h_k \xi_k\|^2 = \\
= \ & \|A_k(x_k - x^*)\|^2 - 2h_k(A_k(x_k - x^*), \xi_k) + h_k^2 = \\
= \ & \|A_k(x_k - x^*)\|^2 - 2\frac{f(x_k) - f^*}{\|B_k^T \partial f(x_k)\|}(A_k(x_k - x^*), \xi_k) + \frac{(f(x_k) - f^*)^2}{\|B_k^T \partial f(x_k)\|^2} \leq \\
\leq \ & \|A_k(x_k - x^*)\|^2 - \frac{\big(f(x_k) - f^*\big)^2}{\|B_k^T \partial f(x_k)\|^2}
\end{aligned}
$$

через співвідношення (1.45) теореми 1.5.

Коли $P_k \neq \emptyset$, через опуклість $f(x)$ маємо

$$\begin{aligned}
&\|A_{k+1}(x_{k+1} - x^*)\|^2 = \\
= &\left\|A_{k+1}(x_k - x^*) - \frac{f(x_k) - f^*}{\|B_{k+1}^T \partial f(x_k)\|} \cdot \frac{B_{k+1}^T \partial f(x_k)}{\|B_{k+1}^T \partial f(x_k)\|}\right\|^2 = \\
= &\|A_{k+1}(x_k - x^*)\|^2 - 2\frac{f(x_k) - f^*}{\|B_{k+1}^T \partial f(x_k)\|^2}(x_k - x^*, \partial f(x_k)) + \\
&+ \frac{(f(x_k) - f^*)^2}{\|B_{k+1}^T \partial f(x_k)\|^2} \leq \|A_{k+1}(x_k - x^*)\|^2 - \frac{(f(x_k) - f^*)^2}{\|B_{k+1}^T \partial f(x_k)\|^2}.
\end{aligned}$$

Оцінимо обидва доданки в (1.49). Для першого, використовуючи співвідношення (1.43) леми 1.4 при $\lambda = -1/2$, маємо

$$\begin{aligned}
&\|A_{k+1}(x_k - x^*)\|^2 = \|T_{-1/2}(\xi_k, \tilde{p}_k)\ A_k(x_k - x^*)\|^2 = \\
= &\Big(A_k(x_k - x^*),\ T_{-1/2}^T(\xi_k, \tilde{p}_k)\ T_{-1/2}(\xi_k, \tilde{p}_k)\ A_k(x_k - x^*)\Big) = \\
= &\left(A_k(x_k - x^*),\ \left(I + \frac{\left(\tilde{p}_k \xi^T + \xi \tilde{p}_k^T - \tilde{p}_k \tilde{p}_k^T\right)}{\|\xi_k - \tilde{p}_k\|^2}\right) A_k(x_k - x^*)\right) = \\
= &\|A_k(x_k - x^*)\|^2 + \\
&+ \frac{2\Big(A_k(x_k - x^*), \xi_k\Big)\Big(A_k(x_k - x^*), \tilde{p}_k\Big)}{\|\xi_k - \tilde{p}_k\|^2} - \frac{\Big(A_k(x_k - x^*), \tilde{p}_k\Big)^2}{\|\xi_k - \tilde{p}_k\|^2} \leq \\
\leq &\|A_k(x_k - x^*)\|^2 + \\
&+ \frac{2\Big(A_k(x_k - x^*), \xi_k\Big)\Big(A_k(x_k - x^*), \tilde{p}_k\Big)}{\|\xi_k - \tilde{p}_k\|^2} \leq \|A_k(x_k - x^*)\|^2.
\end{aligned} \tag{1.49}$$

Остання нерівність в (1.49) випливає через співвідношення (1.45) та (1.47) теореми 1.5, оскільки $(A_k(x_k - x^*), \tilde{p}_k) \leq 0$, а $(A_k(x_k - x^*), \xi_k) \geq 0$.

Окрім того,

$$\begin{aligned}
\frac{(f(x_k) - f^*)^2}{\|B_{k+1}^T \partial f(x_k)\|^2} = &\frac{\|B_k^T \partial f(x_k)\|^2}{\|B_{k+1}^T \partial f(x_k)\|^2} \cdot \frac{(f(x_k) - f^*)^2}{\|B_k^T \partial f(x_k)\|^2} = \\
= &\frac{1}{\|\xi_k - \tilde{p}_k\|^2} \cdot \frac{(f(x_k) - f^*)^2}{\|B_k^T \partial f(x_k)\|^2} \geq \frac{(f(x_k) - f^*)^2}{\|B_k^T \partial f(x_k)\|^2},
\end{aligned} \tag{1.50}$$

оскільки $\|\xi_k - \tilde{p}_k\|^2 \leq 1$. Враховуючи (1.49), (1.50), з (1.49) випливає (1.48) також в тому випадку, коли $P_k \neq \emptyset$, що завершує доведення теореми. □

Теорема 1.6 означає, що вибір параметра $\lambda = -1/2$ для методу ORTGF дозволяє зберегти зовнішню апроксимацію множини екстремумів кулею зі зменшенням радіусу при переході в перетворений простір аргументів. Враховуючи, що $|\det(B_k)| = 1$, такий вибір параметра λ дозволяє обґрунтовувати збіжність методу ORTGF в розумінні зовнішньої локалізації множини екстремумів еліпсоїдом зі зменшенням об'єму. Однак, зменшення об'єму тут буде невеликим, оскільки його забезпечує лише класичний фейєрівський крок з центру кулі. Тим не менш, оператор (1.41) при $\lambda = -1/2$ (на жаль він єдиний) дозволяє покращувати структуру поверхонь рівня функції, не збільшуючи об'єм області локалізації множини екстремумів. Цей факт дозволяє обґрунтовувати методи змінної метрики і за будь-яких інших стратегій побудови нетупокутного конуса, що локалізує множину екстремумів.

При цьому перетворення (1.41) набуває вигляду

$$\begin{aligned} T_{-1/2}(p_m, p) &= T^{-1}_{-1/2}(p_m, p) = I - \frac{p_m - p}{\|p_m - p\|^2}(2p_m - p)^T = \\ &= \left(I - 2\frac{p_m - p}{\|p_m - p\|}\Big(\frac{p_m - p}{\|p_m - p\|}\Big)^T\right)\left(I + \frac{p_m - p}{\|p_m - p\|^2}p^T\right) \end{aligned}$$

і містить як «чисте» відображення, яке визначається ортогональною матрицею, так і перетворення, що діє за типом «розтягу». Відносно відображення субградієнтні методи, а також методи на кшталт відсікань інваріантні, а перетворення на кшталт «розтягу» здатне покращувати структуру поверхонь рівня функції, що оптимізується.

Недолік оператора (1.41) при $\lambda = -1/2$ полягає в тому, що він не є стискаючим для простору субградієнтів, тобто $\det(B_k B_k^T) = 1$. Це сприяє накопиченню помилок при обчисленні нормованих напрямків і може призвести до нестійкої роботи методів для сильно яружних функцій.

З теореми (1.6) випливає теорема (1.7).

Теорема 1.7. *Нехай на кожному етапі методу* ORTGF *при* $\lambda = -1/2$ $\|B_k\| \le c_1$ *і* $\|\partial f(x_k)\| \le c_2$*. Тоді метод* ORTGF *розв'язує задачу (1.1) з точністю* ε_f *не більше, ніж за* K *кроків, де* $K = \left[\left(\frac{c_1 c_2 \|x_0 - x^*\|}{\varepsilon_f}\right)^2\right] + 1.$

Її доведення аналогічне доведенню теореми 1.3.

Однак, параметр $\lambda = -1/2$ єдиний, який дозволяє, ґрунтуючись на розбіжності суми ряду $\sum\limits_{i=0}^{k} \frac{(f(x_i) - f^*)^2}{\|B_i^T \partial f(x_i)\|^2}$, обґрунтувати збіжність для методу ORTGF у розумінні ε_f-збіжності за функціоналом. Для обґрунтування збіжності

ORTGF при інших значеннях параметра λ, зокрема змінних, потрібно використати інший апарат. Найбільше тут підходить інтерпретація цих методів як методів на кшталт спряжених напрямків [4]. Однак це питання потребує окремого обговорення. Тут обмежимося лише чисельною перевіркою методу ORTGF за двома значеннями параметра λ: $\lambda = -1/2$ — $\mathsf{ORTGF}(-0.5)$ і $\lambda = 1$ — $\mathsf{ORTGF}(1.0)$. Однак, якщо для задачі (1.1) збіжність методу $\mathsf{ORTGF}(-0.5)$ до ε_f-розв'язку обґрунтована, то метод $\mathsf{ORTGF}(1.0)$ вимагає більш чіткого обґрунтування.

Поясненням другого вибору параметра λ може слугувати таке геометричне пояснення. Нехай у точці y_k на k-му кроці методу $\mathsf{ORTGF}(-0.5)$ $\tilde{P}_k \neq \emptyset$. Тоді, перш ніж перетворювати конус, що локалізує множину екстремумів, за допомогою оператора $T_{-1/2}(\xi_k, \tilde{p}_k)$, поставимо собі за мету зменшити об'єм області локалізації множини екстремумів, не зіпсувавши при цьому структури конуса. Це просто зробити в рамках еліпсоїда на кшталт [8], де одна з осей співпадає з напрямком $(\xi_k - \tilde{p}_k)$, а друга — з тим $p_i \in \tilde{P}_k$, для якого $(\xi_k, p_i) \leq -1/\sqrt{2}$. Це, як правило, матиме місце для сильно яружних функцій. Перетворення такого еліпсоїда в кулю не змінює структури конуса, який локалізує множину екстремумів. Воно вимагає розтягу простору в напрямку $(\xi_k - \tilde{p}_k)$ і стиснення в ортогональному напрямку p_i, й призводить до зменшення норми останнього субградієнта в перетвореному просторі аргументів. Вибір параметра $\lambda = 1$ фактично надає методу цей сенс, але без додаткових перетворень простору. При цьому норма останнього субградієнта зменшуватиметься у два рази відносно того, що забезпечував би оператор $T_{-1/2}(\xi_k, \tilde{p}_k)$.

Числові експерименти. У першій серії експериментів перевіримо чисельну стійкість методів $\mathsf{ORTGF}(-0.5)$ та $\mathsf{ORTGF}(1.0)$ для того ж набору тестових задач, що і в розділі 2, і для досягнення тієї ж точності ε_f за функціоналом. Як параметри формування конуса виберемо: $\varepsilon_K = 10^{-4}$, $\varepsilon_R = 10^{-8}$, $m_0 = n-1$.

Результати наведено в таблиці 1.3. Тут перше число в дужках позначає кількість перетворень, друге — максимальне число накопичених субградієнтів. Як видно з таблиці, збіжність методів до ε_f-розв'язку досить стійка та майже така ж, як і для для методу (1.29)–(1.34).

Хоча метод ORTGF побудовано з урахуванням вирівнювання структури поверхонь рівня $f(x)$ для подолання її яру, проте він не завжди досягатиме цієї мети. Справа в тому, що яр $\varphi(y)$ у точці y_k характеризує локальна поведінка $\varphi(y)$ поблизу y_k.

Однак при використанні класичного феєрівського кроку та прийнятого способу накопичення субградієнтів існує небезпека, що накопичені субградієнти, особливо «найстаріші», тільки псують картину яружності в поточній

Таблиця 1.3: ε_f-збіжність методів $\mathsf{ORTGF}(-0.5)$ та $\mathsf{ORTGF}(1.0)$.

Задача	n	OTRGF(−0.5)		OTRGF(1.)	
		$\mathrm{iter}(\varepsilon_0)$	$\mathrm{iter}(\varepsilon_0^2)$	$\mathrm{iter}(\varepsilon_0)$	$\mathrm{iter}(\varepsilon_0^2)$
Shor	5	33(30,4)	59(56,4)	33(30,4)	69(66,4)
Maxquad	10	45(37,5)	95(87,5)	42(35,5)	88(79,5)
Quad(3.)	5	40(9,2)	71(9,2)	52(30,3)	96(58,3)
Quad(3.)	10	80(61,5)	113(62,5)	86(68,3)	141(109,3)
Quad(10.)	5	57(26,3)	90(26,3)	50(22,3)	74(22,3)
Quad(10.)	10	156(123,8)	189(128,8)	131(109,4)	193(161,4)

точці.

Для гладких функцій це, зазвичай, не спрацьовуватиме, оскільки збіжності методу допомагає прямування до нуля норми субградієнта. Це забезпечуватиме і накопичення субградієнтів, які розумно характеризують яр. Однак для кусково-лінійних функцій це не так. Тут збіжність швидше забезпечуватиметься завдяки обчисленню вершини конуса, який визначається числом шматків, близьких до n.

Враховуючи вищесказане, для кусково-лінійних функцій не варто очікувати хорошої практичної роботи від методу $\mathsf{ORTGF}(-0.5)$ при $m_0 \ll n$. Для сильно яружних функцій ситуація погіршиться також завдяки накопиченню помилок у матриці B_k. Звісно, тут можна покращити ситуацію як завдяки додатковим розтягам простору, спрямованим на зменшення об'єму, так і завдяки процедурі «відновлення»та удосконалення способу накопичення субградієнтів. Для методу $\mathsf{ORTGF}(1.0)$ перетворення спрямоване на більш сильний розтяг конуса, і в нього значно більше шансів на хорошу практичну роботу при невеликих значеннях m_0.

У другій серії експериментів перевіримо роботу методів для мінімізації опуклої кусково-лінійної функції з дуже великим числом шматків. Для цього виберемо «погану»для субградієнтних методів задачу $\mathsf{TR48}$ ($n = 48$, $f^* = -683565$) з [22, с. 161].

Результати для двох різних початкових точок з [22] наведено в таблиці 1.4. Тут в дужках вказано максимальну кількість накопичених субградієнтів. Параметри ε_K та ε_R такі ж, як і раніше. Параметр $\varepsilon_f = 50$ для першої початкової точки вибраний так, щоб число обчислень $f(x)$ і $\partial f(x)$ можна було порівняти з результатами, представленими в [23], де як прийнятний розв'язок вважалися точки, де $f(x) \le -683500$.

Як видно з таблиці 1.4, тут також характерна стійкість методів у сенсі ε_f-збіжності за функціоналом.

Нарешті в останній третій серії експериментів перевіримо чисельну стійкість методу ORTGF(1.0) для сильно яружних задач середніх розмірів ($n \sim 30$—100). Як тестові завдання виберемо $Quad(t)$ та $Sabs(t)$. Результати наведено в таблиці 1.5. Тут $iter(\varepsilon)$ — кількість обчислень $f(x)$ та $\partial f(x)$. В дужках наведено число перетворень, а для ORTGF(1.0) при $m_0 = n - 1$ — і максимальна кількість накопичених субградієнтів. Якщо вважати однією великою ітерацією методу ORTGF(1.0) n обчислень $f(x)$ та $\partial f(x)$, що за трудомісткістю рівнозначно однієї ітерації методу Ньютона, то, як видно з таблиці 1.5, кількість великих ітерацій не така вже й велика і становить ~ 10.

Таблиця 1.4: ORTGF(−0.5) та ORTGF(1.0) для задачі TR48

Метод	m_0	$f(x_0) = -464816.$		$f(x_0) = -638524.94$	
		$iter(50.)$	$iter(10^{-5})$	$iter(1.)$	$iter(10^{-5})$
ORTGF(−0.5)	n-1	139(28)	222(31)	72(34)	151(34)
ORTGF(1.)	n-1	170(24)	344(24)	97(30)	248(30)
ORTGF(1.)	20	172	358	162	303
ORTGF(1.)	10	166	345	200	340
ORTGF(1.)	5	199	412	207	357

Загальні зауваження. Отже, на основі оператора (1.41) також легко будувати методи змінної метрики з простою геометричною інтерпретацією у перетвореному просторі аргументів. При цьому, для загальної задачі опуклого програмування цілком можливе створення практично ефективних методів зі строгим обґрунтуванням їх збіжності, використовуючи зменшення об'єму області локалізації множини екстремумів.

Проте саме перетворення — це лише одна складова методів. Друга вимагає вибору напряму дослідження функції та способу регулювання крокового множника. Способи, які використовуються в ORTGF, навряд чи дуже вдалі, оскільки обов'язковий рух з точки антисубградієнта сильно ускладнює аналіз ліній рівня $\varphi(y)$.

Тому, раціональнішими є схеми алгоритмів, де лінії рівня $\varphi(y)$ аналізуються або уточнюються відносно фіксованої на деякий час точки.

Таблиця 1.5: ε_f-збіжність методу ORTGF(1.0) для сильно яружних задач

Задача	n	$\mathsf{OTRGF}(1.0), m_0 = n-1$		$\mathsf{OTRGF}(1.0), m_0 = 10$	
		$iter(10^{-10})$	$iter(10^{-20})$	$iter(10^{-10})$	$iter(10^{-20})$
$Quad(2.)$	30	236(229,5)	332(315,5)	236(229)	332(315)
$Sabs(2.)$	30	476(464,12)	527(524,12)	462(459)	523(520)
$Quad(1.2)$	60	188(183,5)	277(259,5)	188(183)	277(259)
$Sabs(1.2)$	60	464(462,21)	541(539,21)	469(467)	544(542)
$Quad(1.2)$	100	428(422,8)	542(534,8)	428(422)	542(534)
$Sabs(1.2)$	100	1480(1475,19)	1564(1559,19)	1293(1291)	1375(1373)

При цьому для обчислення $f(x)$ та $\partial f(x)$ можна забезпечити прості процедури типу ε-найшвидшого спуску в перетвореному просторі аргументів, використовуючи ортогональність субградієнтів, які зберігаються, або їх опуклих комбінацій. Просто вирішуються питання відсіву зайвих субградієнтів, легко уточнюються верхня та нижня оцінки f^* тощо.

Перетворення простору природно спрямувати те що, щоб напрямок дослідження функції $f(x)$ з точки наближався до напрямку на оптимум. Тоді, змінюючи за деяким правилом фіксовану точку (наприклад, на точку рекорду $f(x)$, або проєкцію фіксованої точки на множину

$$Q_k = \left\{ y : \varphi(y_i) + \big(\partial\varphi(y_i), y - y_i\big) \le f^{record}, \quad i \in I_k \right\},$$

і, повторюючи таку ж процедуру відносно нової точки, приходимо до методів на кшталт ньютонівських (квазіньютонівських), які будуть застосовні як для гладких, так і для негладких функцій.

1.5. Висновки

Отже, застосування описаних перетворень простору, спрямоване на вирівнювання структури поверхонь рівня функції, дозволяє значно поліпшити працездатність субградієнтних методів.

Для задачі опуклого програмування використання операторів (1.11) і (1.41) у поєднанні з оператором розтягу простору дозволяє побудувати цілу серію ефективних методів змінної метрики.

При цьому для методів B-формі можна дати просту геометричну інтерпретацію процесу в перетвореному просторі аргументів.

Розглянуті перетворення навіть близько не вичерпують перелік ефективних однорангових перетворень простору.

Зокрема, цікавим є оптимальне одноранговe перетворення на кшталт (1.11), яке дозволило б покращити збіжність у розумінні зменшення об'єму області локалізації множини екстремумів.

У рамках доортогоналізуючих одноранговиx перетворень перспективним є перетворення на кшталт симплексного, яке дозволяє будувати методи для опуклих задач досить великих розмірів, обійшовшись при цьому зберіганням кінцевого набору векторів, а не повної матриці розмірності $n \times n$.

Крім того, можна побудувати й низку інших цікавих лінійних операторів ортогоналізуючого типу.

Розділ 2

B-форма методу Давидона–Флетчера–Пауелла

Обговорюється спеціальна форма (B-форма) методів квазіньютонівського типу, яка дозволяє легко інтерпретувати ці методи, як градієнтні в перетвореному відповідним чином просторі аргументів. Наведено B-форму методу Давидона–Флетчера–Пауелла та на її основі проведено порівняння цього методу з r-алгоритмами. Для мінімізації гладких опуклих функцій побудовано градієнтний метод з перетворенням простору, що поєднує властивості як квазіньютонівських методів, так і r-алгоритмів. Обговорюються можливі схеми методів такого типу для мінімізації негладких опуклих функцій.

2.1. Квазіньютонівські методи

Теорія квазіньютонівських методів [4, 5, 24–27] спирається на можливість апроксимації кривини нелінійної функції без явного формування її матриці Гессе, тобто дані про матрицю Гессе накопичуються в них на основі спостереження за зміною градієнта функції під час ітерацій спуску за напрямком. Вони вимагають на ітерації значно меншого обсягу обчислень ніж ньютонівські методи, і за певних умов здатні забезпечити швидкість збіжності, притаманну для методів другого порядку.

Перший метод квазіньютонівського типу (далі ДФП-метод), був запропонований Давидоном у [2] та розвинений Флетчером і Пауеллом у статті [3]. ДФП-метод є одним із ефективних серед методів квазіньютонівського типу. Його перевершує тільки метод Бройдена–Флетчера–Гольдфарба–Шанно, варіант якого, відомий як L-BFGS [28], активно використовується в машинному навчанні та обробці великих об'ємів даних.

Сімейство методів квазіньютонівського типу можна розширити, використовуючи методики з [4] та [5], які полягають у виборі коректив за допомогою матриць першого або другого рангу для корекції симетричної матриці, що використовується на наступному кроці методу. Проте аналіз деяких квазіньютонівських методів можна спростити, якщо від кроку до кроку використовувати однорангову корекцію несиметричної матриці [29].

Інакше кажучи, є дві можливості представлення одного й того ж квазіньютонівського методу. Для них будемо дотримуватися тих назв, які закріпилися за схожими схемами методів субградієнтного типу з розтягом простору [6]. Методи, де коригується матриця B — матриця оберненого перетворення простору (вона може бути несиметричною), називатимемо методами в B-формі, а методи, що використовують корекцію симетричної матриці BB^T, — методами в H-формі. Кожна з цих форм має як переваги, так і недоліки. Поєднання обох форм дозволяє проводити більш ґрунтовний аналіз методів квазіньютонівського типу.

У розділі розглядається B-форма ДФП-методу, яка дозволяє інтерпретувати його як градієнтний метод у перетвореному просторі аргументів.

Матеріал викладено у такому порядку. У другому підрозділі розглядаються H та B-форми квазіньютонівських методів, у третьому підрозділі ці форми наводяться для ДФП-методу. Четвертий підрозділ присвячено обговоренню подібності ДФП-методу та r-алгоритмів. У п'ятому підрозділі побудовано $DFPR(\alpha)$-алгоритм, який є в певному сенсі проміжним між ДФП-методом та $r_\mu(\alpha)$-алгоритмом [6]. У шостому підрозділі обговорюється перенесення основних принципів $DFPR(\alpha)$-алгоритму на загальний випадок задачі мінімізації опуклої функції.

2.2. H- та B-форми квазіньютонівських методів

Нехай є задача

$$\min f(x), \tag{2.1}$$

де $f(x)$ — опукла двічі диференційовна функція векторного аргументу $x \in X$, $X = \mathbb{R}^n$. $\mathbb{R}^n$ — евклідів простір розмірності n зі скалярним добутком (x, y); x^* — екстремальна точка задачі (2.1).

Нехай початкова точка x_0 вибрана в досить малому околі точки мінімуму x^*, тобто $f(x)$ буде добре апроксимуватись квадратичною функцією

$$\varphi(x) = \frac{1}{2}\Big(\nabla^2 f(x^*)\,(x - x^*),\ x - x^*\Big) + f(x^*).$$

Тут $\nabla^2 f(x^*)$ — матриця Гессе в точці мінімуму.

Нехай H_0 — симетрична додатно визначена матриця розмірності $n \times n$. Зазвичай покладають $H_0 = I$, де I — одинична матриця.

Методи квазіньютонівського типу в H-формі генерують послідовність точок $\{x_k, k = 0, 1, \ldots, n\}$ за таким правилом:

$$x_{k+1} = x_k - h_k H_k \nabla f(x_k), \tag{2.2}$$

де $\nabla f(x_k)$ — градієнт $f(x)$ в точці x_k; H_k — симетрична матриця розмірності $n \times n$; h_k — крок, який відповідає мінімуму $f(x)$ в напрямку $-H_k \nabla f(x_k)$, тобто

$$h_k = \operatorname{argmin}_{h \geq 0} \Big\{ f\Big(x_k - hH_k \nabla f(x_k)\Big) \Big\}. \tag{2.3}$$

Перерахунок матриці H від кроку до кроку здійснюється таким чином:

$$H_{k+1} = H_k + \Delta H_k, \tag{2.4}$$

де ΔH_k — матриця невеликого рангу, побудована на основі поведінки градієнтів $\nabla f(x_k)$ и $\nabla f(x_{k+1})$ так, щоб $H_n \approx \Big[\nabla^2 f(x^*)\Big]^{-1}$. Можливість такого перерахунку забезпечується так званою квазіньютонівською умовою

$$H_{k+1}\Big(\nabla f(x_{k+1}) - \nabla f(x_k)\Big) = -h_k H_k \nabla f(x_k). \tag{2.5}$$

Для квазіньютонівських методів, у яких умову закінчення процесу за n кроків при мінімізації квадратичних функцій відкинуто, умова (2.5) забезпечує виконання умови, що полягає в тому, що власні числа матриці H_k прямують до власних чисел матриці $\Big[\nabla^2 f(x^*)\Big]^{-1}$. Фактична різниця між методами квазіньютонівського типу в H-формі полягає у різних формулах перерахунку матриці H_{k+1}, які задовольняють квазіньютонівську умову (2.5) та один з вищенаведених способів наближення H_k до $\Big[\nabla^2 f(x^*)\Big]^{-1}$.

Додатно визначену симетричну матрицю H_k завжди можна представити у вигляді $H_k = B_k B_k^T$, де B_k — невироджена матриця розмірності $n \times n$. Співвідношення (2.2), яке реалізує перехід до наступної точки для H-форми квазіньютонівських методів, можна записати так:

$$x_{k+1} = x_k - h_k B_k B_k^T \nabla f(x_k) \tag{2.6}$$

або

$$y_{k+1} = y_k - h_k B_k^T \nabla f(x_k) = y_k - h_k \nabla \varphi_k(y_k), \tag{2.7}$$

де $y_{k+1} = B_k^{-1} x_{k+1}$ і $y_k = B_k^{-1} x_k$ — образи точок x_{k+1} и x_k из X в перетвореному за допомогою лінійного оператора $A_k = B_k^{-1}$ просторі аргументів, $\nabla \varphi_k(y_k) = B_k^T \nabla f(x_k)$ — градієнт функції $\varphi_k(y) = f(B_k y)$, визначеної в просторі аргументів $Y_k = A_k X$, в точці y_k.

Процес (2.7) можна інтерпретувати як градієнтний метод найшвидшого спуску в перетвореному просторі аргументів $Y_k = A_k X$ для мінімізації опуклої гладкої функції $\varphi_k(y) = f(B_k y)$. Співвідношення (2.3), яке визначає вибір крокового множника в квазіньютонівських методах, рівносильне такому:

$$\begin{aligned} h_k = \operatorname{argmin}_{h \geq 0} \left\{ \varphi_k\Big(y_k - h\, B_k^T \nabla f(x_k)\Big) \right\} = \\ = \operatorname{argmin}_{h \geq 0} \left\{ \varphi_k\Big(y_k - h \nabla \varphi_k(y_k)\Big) \right\}, \end{aligned} \tag{2.8}$$

де $\varphi_k(y) = f(B_k y)$ — функція, визначена в Y_k.

Співвідношення (2.6)–(2.8) у поєднанні з процедурою малорангової корекції матриці B_k дозволяють описувати квазіньютонівські методи у B-формі. При цьому такі методи мають достатньо просту градієнтну природу у перетвореному просторі аргументів.

Зауважимо, що, зважаючи на неоднозначність розкладу $H_k = B_k B_k^T$, для конкретного методу в H-формі випливає існування різних методів у B-формі. Але це не так вже й погано, тому що при побудові B-методу легко врахувати й ту умову, яка забезпечує чисельну стійкість методу. Наприклад, при одноранговій корекції матриці $B_{k+1} = B_k T_k$ матрицю T_k вибираємо так, щоб відношення

$$\lambda_{max}(T_k) \,/\, \lambda_{min}(T_k)$$

було якомога меншим. Тут $\lambda_{max}(T_k)$ $\big(\lambda_{min}(T_k)\big)$ — максимальне (мінімальне) власне число матриці T_k.

Однак чисельна стійкість B-форми методів поступається двом перевагам H-форми: більш економному зберіганню матриці та можливості обійтися меншою кількістю арифметичних операцій на ітерації. Тому потрібен певний компроміс між дослідженням методів та їхньою реалізацією на ЕОМ. Таким компромісом може слугувати розробка чисельно стійкого методу квазіньютонівського типу у B-формі, та розгляд його H-форми як наслідку цього методу з метою економії пам'яті та обчислень. Тим більше, що розробці такого методу нічого не заважає. Дійсно, квазіньютонівську умову можна записати так:

$$B_{k+1} B_{k+1}^T \Big(\nabla f(x_{k+1}) - \nabla f(x_k)\Big) = -h_k\, B_k B_k^T \nabla f(x_k).$$

Позначимо $\xi = B_k^T \nabla f(x_{k+1}) - B_k^T \nabla f(x_k)$ та $\eta = B_k^T \nabla f(x_k)$ — відповідно вектори різниці послідовних градієнтів та поточного градієнта в перетвореному просторі аргументів.

Нехай для перерахунку B_{k+1} використовується одноранґова корекція досить загального виду:

$$B_{k+1} = B_k \left(I + t_1 \left(\xi + t_2 \eta \right) \left(\xi + t_3 \eta \right)^T \right),$$

де t_1, t_2, t_3 — невідомі скалярні параметри. Тоді для виконання квазіньютонівської умови параметри t_1, t_2, t_3 мають задовольняти таке співвідношення:

$$\left(I + t_1 (\xi + t_2 \eta) \left(\xi + t_3 \eta \right)^T \right) \left(I + t_1 (\xi + t_3 \eta) \ \left(\xi + t_2 \eta \right)^T \right) \xi = -h_k \eta, \quad (2.9)$$

яке пов'язує вектори ξ и η в перетвореному просторі аргументів.

Зі співвідношення (2.9) випливають два рівняння для трьох невідомих параметрів t_1, t_2 і t_3. Вибір певних із них дає вже відомі методи квазіньютонівського типу.

Наприклад, поклавши $t_3 = 0$ та однозначно визначивши t_1 і t_2, отримаємо ДФП-метод. Але вибір цих параметрів породжує і низку нових методів квазіньютонівського типу, які використовують одноранґову корекцію матриці B_k, і цілком можливо, що серед цих методів можна відшукати й такі, що за своєю ефективністю не будуть уступати методу Бройдена–Флетчера–Гольдфарба–Шанно.

2.3. H- та B-форми ДФП-методу

Обидві форми ДФП-методу для задачі (2.1) опишемо, припускаючи, що x_0 — початкове наближення з досить малого околу точки мінімуму x^*. H-форму ДФП-методу наведемо згідно з [24] з точністю до незначних перепозначень.

H-форма ДФП-методу

Крок 0. Вибрати $x_0 \in \mathbb{R}^n$. Якщо $\nabla f(x_0) = 0$, то зупинитись та покласти $x^* = x_0$. Інакше покласти $H_0 = I$, де I — одинична матриця розмірності $n \times n$, $g_0 = \nabla f(x_0)$, $k = 0$ та перейти до кроку 1.

Крок 1. Покласти

$$\xi_k = H_k \, g_k. \quad (2.10)$$

Крок 2. Обчислити

$$h_k = \mathrm{argmin}_{h \geq 0} \left\{ f\left(x_k - h\xi_k\right) \right\}. \tag{2.11}$$

Крок 3. Покласти

$$x_{k+1} = x_k - h_k\, \xi_k. \tag{2.12}$$

Крок 4. Якщо $\nabla f(x_{k+1}) = 0$, то зупинитись і покласти $x^* = x_{k+1}$. Інакше покласти

$$g_{k+1} = \nabla f\left(x_{k+1}\right), \quad \Delta g_k = g_{k+1} - g_k, \quad \Delta x_k = x_{k+1} - x_k, \tag{2.13}$$

$$H_{k+1} = H_k - \frac{H_k \Delta g_k (\Delta g_k)^T H_k}{(\Delta g_k,\ H_k \Delta g_k)} = \frac{\Delta x_k (\Delta x_k)^T}{(\Delta g_k,\ \Delta x_k)}. \tag{2.14}$$

Крок 5. Покласти $k = k + 1$ та перейти до кроку 1.

Для опису B-форми ДФП-методу нам знадобляться певні допоміжні результати. Зокрема, використовуючи (2.12) та (2.13), перерахунок матриці H_{k+1}, згідно з (2.14), можна записати так:

$$\begin{aligned} H_{k+1} &= H_k - \frac{H_k \Delta g_k (\Delta g_k)^T H_k}{(\Delta g_k,\ H_k \Delta g_k)} + h_k \frac{H_k g_k\, g_k^T H_k}{(\Delta g_k,\ H_k g_k)} = \\ &= B_k B_k^T - \frac{B_k B_k^T \Delta g_k (\Delta g_k)^T B_k B_k^T}{(\Delta g_k,\ B_k B_k^T \Delta g_k)} + h_k \frac{B_k B_k^T g_k\, g_k^T B_k B_k^T}{\left(\Delta g_k,\ B_k B_k^T g_k\right)} = \\ &= B_k \left(I - \xi_k\, \xi_k^T + t_k^2 \eta_k\, \eta_k^T \right) B_k^T = B_{k+1} B_{k+1}^T, \end{aligned} \tag{2.15}$$

де $\xi_k = \frac{B_k^T g_{k+1} - B_k^T g_k}{\|B_k^T g_{k+1} - B_k^T g_k\|}$, $\eta_k = \frac{B_k^T g_k}{\|B_k^T g_k\|}$ та $t_k^2 = h_k \frac{\|B_k^T g_k\|^2}{\left(B_k^T g_k,\ B_k^T g_k - B_k^T g_{k+1}\right)}$.

Додатність параметра t_k^2 забезпечується вибором h_k з умови найшвидшого спуску для гладкої функції в перетвореному просторі аргументів. При точній реалізації найшвидшого спуску $t_k^2 = h_k$, оскільки $(B_k^T g_k, B_k^T g_{k+1}) = 0$. Для того, щоб забезпечити перерахунок матриці H_{k+1} як в (2.15), достатньо коригувати $B_{k+1} = B_k T_k$, де $T_k = \left(I - \left(\xi_k + t_k \eta_k\right) \xi_k^T \right)$. Насправді, при такій корекції B_{k+1} для $T_k T_k^T$ маємо:

$$T_k T_k^T = \left(I - \left(\xi_k + t_k \eta_k\right) \xi_k^T \right) \left(I - \xi_k \left(\xi_k + t_k \eta_k\right)^T \right) = I - \xi_k\, \xi_k^T + t_k^2 \eta_k\, \eta_k^T,$$

що забезпечує середній множник в правій частині (2.15).

Отже, для однорангової корекції матриці B_{k+1} підходить така формула:

$$B_{k+1} = B_k \left(I - \left(\frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|} + \sqrt{h_k \frac{\|\widetilde{g}_k\|^2}{(\widetilde{g}_k, \widetilde{g}_k - \widetilde{g}_{k+1})}} \frac{\widetilde{g}_k}{\|\widetilde{g}_k\|} \right) \times \right. \\ \left. \times \left(\frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|} \right)^T \right), \tag{2.16}$$

де $\widetilde{g}_{k+1}$ и $\widetilde{g}_k$ – субградієнти функції $\varphi_k(y) = f(B_k y)$ в точках $y_{k+1} = B_k^{-1} x_{x+1}$ и $y_k = B_k^{-1} x_k$ відповідно. Враховуючи (2.16), B-форма ДФП-методу набуває такого вигляду.

B-форма ДФП-методу

Крок 0. Вибрати $x_0 \in \mathbb{R}^n$. Якщо $\nabla f(x_0) = 0$, то зупинитись та покласти $x^* = x_0$. Інакше покласти

$$B_0 = I,$$

де I — одинична матриця розмірності $n \times n$, $g_0 = \nabla f(x_0)$, $k = 0$ та перейти до кроку 1.

Крок 1. Покласти

$$\widetilde{g}_k = B_k^T g_k. \tag{2.17}$$

Крок 2. Покласти

$$\xi_k = B_k \widetilde{g}_k. \tag{2.18}$$

Крок 3. Обчислити

$$h_k = \operatorname{argmin}_{h \geq 0} \Big\{ f(x_k - h\xi_k) \Big\}. \tag{2.19}$$

Крок 4. Покласти

$$x_{k+1} = x_k - h_k \eta_k. \tag{2.20}$$

Крок 5. Якщо $\nabla f(x_{k+1}) = 0$, то зупинитись та покласти $x^* = x_{k+1}$. Інакше покласти $g_{k+1} = \nabla f(x_{k+1})$, $\quad \widetilde{g}_{k+1} = B_k^T g_{k+1}$,

$$\xi_k = \frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|} + \sqrt{h_k \frac{\|\widetilde{g}_k\|^2}{(\widetilde{g}_k,\ \widetilde{g}_k - \widetilde{g}_{k+1})}}\ \frac{\widetilde{g}_k}{\|\widetilde{g}_k\|}, \quad \eta_k = \frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|}, \tag{2.21}$$

$$B_{k+1} = B_k \left(I - \xi_k \eta_k^T \right). \tag{2.22}$$

Крок 6. Покласти $k = k + 1$ та перейти до кроку 1.

Наведений ДФП-метод у B-формі (2.17)–(2.22) не є оптимальним за використанням арифметичних операцій. Його можна покращити, прибравши операцію множення матриці на вектор в (2.17) за рахунок того, що

$$B_{k+1}^T g_{k+1} = \left(I - \eta_k\, \xi_k^T\right) B_k^T g_{k+1} = \widetilde{g}_{k+1} - \eta_k(\xi_k,\ \widetilde{g}_{k+1}).$$

Але навіть в такому випадку він буде поступатись ДФП-методу в H-формі (2.10)–(2.14) за кількістю арифметичних операцій, хоча цей розрив буде незначним ($4n^2$ множень як порівняти з $3n^2$ множеннями).

Тим не менш B-форма дозволяє прояснити низку моментів для ДФП-методу.

По-перше, додатна визначеність матриці

$$H_{k+1} = B_{k+1} B_{k+1}^T$$

є наслідком невиродженості матриці B_{k+1}, якщо матриця B_k — невироджена. Насправді, визначник матриці B_{k+1} буде таким:

$$\begin{aligned}
\det\left(B_{k+1}\right) &= \det\left(B_k\left(I - \xi_k \eta_k^T\right)\right) = \det\left(B_k\right)\left(1 - \left(\xi_k,\ \eta_k\right)\right) = \\
&= \det\left(B_k\right)\left(1 - \left(\frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|} + \right.\right. \\
&\quad \left.\left. + \sqrt{h_k \frac{\|\widetilde{g}_k\|^2}{(\widetilde{g}_k,\ \widetilde{g}_k - \widetilde{g}_{k+1})}}\, \frac{\widetilde{g}_k}{\|\widetilde{g}_k\|},\ \frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|}\right)\right) = \\
&= \det\left(B_k\right)\sqrt{h_k \frac{\|\widetilde{g}_k\|^2}{(\widetilde{g}_k,\ \widetilde{g}_k - \widetilde{g}_{k+1})}}\, \frac{(\widetilde{g}_k,\ \widetilde{g}_k - \widetilde{g}_{k+1})}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\| \cdot \|\widetilde{g}_k\|} = \\
&= \det\left(B_k\right)\sqrt{h_k \frac{(\widetilde{g}_k,\ \widetilde{g}_k - \widetilde{g}_{k+1})}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|^2}} \neq 0.
\end{aligned}$$

При точному найшвидшому спуску

$$\det\left(B_{k+1}\right) = \det\left(B_k\right)\sqrt{h_k \tfrac{\|\widetilde{g}_k\|^2}{\|\widetilde{g}_k\|^2 + \|\widetilde{g}_{k+1}\|^2}}$$

та він рівний нулю тоді, коли або $h_k = 0$, або $\|\widetilde{g}_k\| = 0$, що для гладких опуклих функцій означає виконання достатньої умови оптимальності в перетвореному просторі аргументів.

По-друге, параметр $\Delta_k = \dfrac{\left(\widetilde{g}_k,\ \widetilde{g}_{k+1}\right)}{\|\widetilde{g}_k\|^2} = \dfrac{\left(g_k,\ H_k g_{k+1}\right)}{\left(g_k,\ H_k g_k\right)}$, який випливає з (2.21) при $\dfrac{\widetilde{g}_k}{\|\widetilde{g}_k\|}$, задає точність виконання умови найшвидшого спуску в перетвореному просторі аргументів. Цю або близьку до неї умову доцільно використовувати в квазіньютонівських методах з наближеним обчисленням мінімуму функції за напрямком як критерій зупинки при одновимірному спуску за напрямком.

По-третє, Δ_k в комбінації з кроком найшвидшого спуску h_k розумно використовувати для процедури повторного старту методів квазіньютонівського типу, враховуючи, що в процесі перерахунку матриці B_{k+1} можливе накопичення помилок. Саме ці параметри та в більшій мірі крок найшвидшого спуску h_k впливають на точність перерахунку матриці B_{k+1}.

2.4. ДФП-метод та r-алгоритми

Для H-алгоритмів [6,30,31] підкреслено їхню «близькість» до ДФП-методу в плані перерахунку матриці H_{k+1}. Використовуючи B-форму ДФП-методу, можна інтерпретувати цю «близькість» більш змістовно шляхом аналізу того, що відбувається в перетвореному просторі аргументів. Виконаємо це для ДФП-методу (2.17)–(2.22) і тих варіантів r-алгоритмів, які використовують точний пошук мінімуму функції за напрямком – $r_\mu(\alpha)$-алгоритм ($\alpha \in [2, 4]$) та граничний варіант r-алгоритму ($\alpha = \infty$) [6]. Як характеристику для порівняння виберемо зміну кута між двома послідовними субградієнтами, яка характерна для однієї ітерації цих методів при переході з Y_k в Y_{k+1}. Очевидно, такі порівняння справедливі лише для гладких функцій.

Нехай g_k і g_{k+1} — субградієнти функції $\varphi_k(y) = f(B_k y)$, визначеної в перетвореному просторі аргументів $Y_k = B_k^{-1}X$, отримані згідно з точним кроком найшвидшого спуску h_k^* в просторі аргументів Y_k. Тоді їхні образи $\widetilde{g}_k$ та $\widetilde{g}_{k+1}$ в перетвореному просторі $Y_{k+1} = B_{k+1}^{-1}X$, враховуючи, що $\left(g_k, g_{k+1}\right) = 0$, будуть

$$\widetilde{g}_k = \left(I - \frac{g_{k+1} - g_k}{\|g_{k+1} - g_k\|}\left(\frac{g_{k+1} - g_k}{\|g_{k+1} - g_k\|} + \sqrt{h_k^*}\,\frac{g_k}{\|g_k\|}\right)^T\right) g_k =$$

$$= g_k + \left(\frac{\|g_k\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2} - \sqrt{h_k^*\,\frac{\|g_k\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2}}\right)(g_{k+1} - g_k) =$$

$$
= \left(\frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2} + \sqrt{h_k^* \frac{\|g_k\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2}} \right) g_k +
$$

$$
+ \left(\frac{\|g_k\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2} - \sqrt{h_k^* \frac{\|g_k\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2}} \right) g_{k+1}, \quad (2.23)
$$

$$
\begin{aligned}
\widetilde{g}_{k+1} &= \left(I - \frac{g_{k+1} - g_k}{\|g_{k+1} - g_k\|} \left(\frac{g_{k+1} - g_k}{\|g_{k+1} - g_k\|} + \sqrt{h_k^*} \frac{g_k}{\|g_k\|} \right)^T \right) g_{k+1} = \\
&= g_{k+1} - \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2} (g_{k+1} - g_k) = \\
&= \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2} g_k + \frac{\|g_k\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2} g_{k+1}. \qquad (2.24)
\end{aligned}
$$

З (2.23), (2.24) випливає, що косинус кута ψ_k між векторами $\widetilde{g}_k$ та $\widetilde{g}_{k+1}$ буде

$$
\cos\psi_k = \left(\frac{\widetilde{g}_k}{\|\widetilde{g}_k\|}, \frac{\widetilde{g}_{k+1}}{\|\widetilde{g}_{k+1}\|} \right) = 1 \Big/ \sqrt{1 + h_k \frac{\|g_k\|^2 + \|g_{k+1}\|^2}{\|g_k\|^2}}. \qquad (2.25)
$$

Отже, один крок ДФП-методу з Y_k до Y_{k+1} призводить до зменшення кута між послідовними субградієнтами від $\pi/2$ до гострого ψ_k, який визначається з умови (2.25). Ця властивість ДФП-методу характерна і для r-алгоритмів, побудованих на ідеї «розширити» конус можливих напрямків спадання функції шляхом операції розтягу простору аргументів в напрямку різниці двох послідовних субградієнтів.

З (2.25) випливає, що коли $h_k \ll 1$, тоді ітерація ДФП-методу близька до ітерації граничного варіанту r-алгоритмів, який перетворює субградієнти g_k та g_{k+1} так, щоб у перетвореному просторі $\cos\psi_k = 1$. Цей же факт випливає з (2.16), оскільки $h_k \ll 1$, то операція перетворення простору в ДФП-методів буде близькою до розтягу простору з дуже великим коефіцієнтом в напрямку різниці двох послідовних субградієнтів.

Така ситуація буде мати місце для сильно яружних функцій, коли ітераційний процес знаходиться в точці, близькій до «дна» яру.

Для $r_\mu(\alpha)$-алгоритму, який використовує постійний коефіцієнт розтягу простору в напрямку $g_{k+1} - g_k$, характерне перетворення прямого кута в гострий ψ_k, косинус якого рівний

$$\cos\psi_k = (\alpha^2 - 1) \Big/ \sqrt{\alpha^4 + \alpha^2\left(\frac{\|g_k\|^2}{\|g_{k+1}\|^2} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right) + 1} \leq 1 - \frac{2}{\alpha^2 + 1}.$$

Тут перетворення кута між послідовними градієнтами прямо не пов'язано з h_k^*, хоча непрямий зв'язок забезпечує співвідношення між нормами векторів g_k та g_{k+1}.

Тому для гладких функцій перетворення простору, що використовується в ДФП-методі, виглядає більш логічним в тому сенсі, що коли крок найшвидшого спуску великий, то слабше «розтягується» конус підходящих напрямків спадання функції $f(x)$, а коли крок найшвидшого спуску малий, то він тягне за собою більш сильний «розтяг» конусу. Але для негладких функцій, де h_k може бути рівним нулю, таке перетворення загалом не застосовне.

Тим не менш, цю обставину можна використовувати для модифікації $r_\mu(\alpha)$-алгоритму, щоб розширити область його застосування для класу майже диференційовних функцій. Насправді, «пастки» для мінімізуючої послідовності $r_\mu(\alpha)$-алгоритму [32, ст. 110–115] базуються на використанні ним постійного коефіцієнту розтягу простору. Замінивши на кроці $r_\mu(\alpha)$-алгоритму постійний коефіцієнт розтягу на змінний, який залежить від параметрів h_k^*, $\|g_k\|$ та $\|g_{k+1}\|$, для такого варіанту r-алгоритмів цілком можливо забезпечити вихід з «поганих» кутових точок для майже диференційовних функцій, при цьому зберігаючи для нього максимальну «близькість» до $r_\mu(\alpha)$-алгоритму.

2.5. DFPR(α)-алгоритм

Для задачі (2.1), за умови, що ми вміємо точно реалізовувати процедуру найшвидшого спуску, побудуємо метод «змінної метрики» — проміжний між ДФП-методом та $r_\mu(\alpha)$-алгоритмом. При цьому збережемо переваги як першого (вибір наступного напрямку руху), так і другого (метод можна перенести на негладкий випадок).

Нехай g_k та g_{k+1} — градієнти $\varphi_k(y) = f(B_k y)$ в точках y_k та y_{k+1}. Тут y_{k+1} отримана згідно з кроком найшвидшого спуску в напрямку $-g_k$ з точки y_k в просторі аргументів $Y_k = B_k^{-1}X$. Тоді $\left(g_k,\ g_{k+1}\right) = 0$.

Перетворення простору з Y_k до Y_{k+1}, що лежить в основі ДФП-методу, має одну «чудову» властивість — незалежно від параметра t_k, рух у просторі Y_k, що відповідає напрямку образу градієнта g_{k+1} в Y_{k+1}, виконується по найкоротшому вектору опуклої комбінації g_k та g_{k+1}.

Насправді, напрямок руху в Y_k, який відповідає руху по вектору $\widetilde{g}_{k+1}$ в Y_{k+1} визначається:

$$\begin{aligned}
p_k &= T_k\big(g_k,\ g_{k+1}\big)T_k^T\big(g_k,\ g_{k+1}\big)g_{k+1} = \\
&= \left(I - \Big(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|} + t_k\frac{g_k}{\|g_k\|}\Big)\Big(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|}\Big)^T\right)\times \\
&\qquad \times\left(I - \Big(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|}\Big)\Big(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|} + t_k\frac{g_k}{\|g_k\|}\Big)^T\right)g_{k+1} = \\
&= \left(I - \Big(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|}\Big)\Big(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|}\Big)^T + t_k^2\frac{g_k}{\|g_k\|}\frac{g_k}{\|g_k\|}^T\right)g_{k+1} = \\
&= g_{k+1} - (g_{k+1}-g_k)\,\frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2} = \\
&= \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2}g_k + \frac{\|g_k\|^2}{\|g_{k+1}\|^2 + \|g_k\|^2}g_{k+1}, \qquad (2.26)
\end{aligned}$$

через те, що $\big(g_k,\ g_{k+1}\big) = 0$. Вектор p_k, який визначається згідно з (2.26), є нічим іншим, як розв'язком такої задачі:

$$\min \frac{1}{2}\left\|\lambda_1 g_k + \lambda_2 g_{k+1}\right\|^2, \qquad (2.27)$$

$$\big(g_k,\ g_{k+1}\big) = 0; \quad \lambda_1 + \lambda_2 = 1; \quad \lambda_1 \geq 0; \quad \lambda_2 \geq 0. \qquad (2.28)$$

Саме тому в основу методу доцільно покласти перетворення, на якому базується ДФП-метод, оскільки воно забезпечує досить успішний напрямок руху на кроці в Y_k. Зокрема, для квадратичних функцій $f(x)$ з розмірністю простору $n = 2$ такий напрямок руху з точки найшвидшого спуску задаватиме точний напрямок на мінімум незалежно від початкової точки.

Тому для таких функцій при довільному виборі параметра t_k метод потребуватиме не більше ніж два кроки найшвидшого спуску.

Однак, збіжність методу за n кроків для квадратичних функцій при $n \geq 3$ (аналогічно до ДФП-методу) забезпечуватиметься далеко не кожним вибором параметра t_k.

Тому відкинемо умову завершення процесу мінімізації за n кроків для квадратичних цільових функцій та на вибір параметра t_k накладемо умову, що метод має наближатись до $r_\mu(\alpha)$-алгоритму в тому сенсі, що перетворення простору було «стискаючим» простір субградієнтів так, як це відбувається в $r_\mu(\alpha)$-алгоритмі, тобто

$$\det(B_{k+1}) = \frac{1}{\alpha}\det(B_k), \quad \alpha > 1, \quad \alpha \in [2,4].$$

Для цього достатньо покласти $t_k = \dfrac{1}{\alpha}\dfrac{\|g_{k+1}-g_k\|}{\|g_k\|}$ і для корекції матриці $B_{k+1} = B_k T_k\big(g_k, g_{k+1}\big)$ вибрати

$$T_k\big(g_k,\ g_{k+1}\big) = I - \left(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|} + \frac{1}{\alpha}\,\frac{\|g_{k+1}-g_k\|}{\|g_k\|}\,\frac{g_k}{\|g_k\|}\right)\times$$
$$\times\left(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|}\right)^T, \quad (2.29)$$

оскільки

$$\begin{aligned}
&\det\Big(T_k(g_k,\ g_{k+1})\Big) = \\
&= 1 - \left(\frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|} + \frac{1}{\alpha}\,\frac{\|g_{k+1}-g_k\|}{\|g_k\|}\,\frac{g_k}{\|g_k\|},\ \frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|}\right) = \\
&= 1 - 1 - \frac{1}{\alpha}\,\frac{\|g_{k+1}-g_k\|}{\|g_k\|}\,\left(\frac{g_k}{\|g_k\|},\ \frac{g_{k+1}-g_k}{\|g_{k+1}-g_k\|}\right) = \\
&= \frac{1}{\alpha}\,\frac{\|g_{k+1}-g_k\|}{\|g_k\|}\,\frac{\|g_k\|}{\|g_{k+1}-g_k\|} = \frac{1}{\alpha}.
\end{aligned}$$

З вищенаведених міркувань отримуємо такий метод «змінної метрики» (будемо називати його $\mathrm{DFPR}(\alpha)$-алгоритмом) для розв'язання задачі (2.1).

DFPR(α)-алгоритм

Крок 0. Перед початком обчислень маємо $\alpha > 1$, $\varepsilon_g > 0$, $x_0 \in \mathbb{R}^n$, $B_0 = I_n$ — одинична матриця розмірності $n \times n$, $g_0 = \nabla f(x_0)$. Тут ε_g — досить мале число, яке задає критерій зупинки за нормою субградієнта. Якщо $\|g_0\| \le \varepsilon_g$, то x_0 — шукана точка та зупинка. Інакше переходимо до наступного кроку.

Нехай на k-му кроці маємо $x_k \in \mathbb{R}^n$, $g_k \in \mathbb{R}^n$, B_k — матриця $n \times n$. Тут $g_k = \nabla f(x_k)$ — градієнт $f(x)$ в точці x_k. Тоді $(k+1)$-й крок характеризує така послідовність операцій.

Крок 1. Покласти $\widetilde{g}_k = B_k^T g_k$.

Крок 2. Покласти $\xi_k = B_k \widetilde{g}_k$.

Крок 3. Обчислити $h_k = \mathrm{argmin}_{h \geq 0}\Big\{f(x_k - h\xi_k)\Big\}$.

Крок 4. Обчислити чергове наближення $x_{k+1} = x_k - h_k\eta_k$.

Крок 5. Покласти $g_{k+1} = \nabla f(x_{k+1})$. Якщо $\|g_{k+1}\| \leq \varepsilon_g$, то зупинка та x_{k+1} — шукана точка. Інакше обчислимо

$$\widetilde{g}_{k+1} = B_k^T g_{k+1}, \quad t_k = \frac{1}{\alpha}\sqrt{1 + \frac{\|\widetilde{g}_{k+1}\|^2}{\|\widetilde{g}_k\|^2}},$$

$$\eta_1 = \frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|} + t_k\frac{\widetilde{g}_k}{\|\widetilde{g}_k\|}, \quad \eta_2 = \frac{\widetilde{g}_{k+1} - \widetilde{g}_k}{\|\widetilde{g}_{k+1} - \widetilde{g}_k\|}, \quad B_{k+1} = B_k\Big(I - \eta_1\eta_2^T\Big).$$

Крок 6. Переходимо до чергового кроку з x_{k+1}, B_{k+1}, g_{k+1}.

Зрозуміло, що збіжність $\mathsf{DFPR}(\alpha)$-алгоритму до розв'язку x^* для задачі (5.5) забезпечуватиме процедура найшвидшого спуску в напрямку антиградієнта. Крім того, він, як і ДФП-метод та r-алгоритми, використовує перетворення простору для того, щоб розширити конус можливих напрямків спадання функції на наступному кроці методу. Насправді, кут між послідовними градієнтами в Y_{k+1} визначається таким співвідношенням:

$$\cos\psi_k = 1 \Bigg/ \sqrt{1 + \left(\frac{1}{\alpha}\left(\frac{\|g_k\|}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{\|g_k\|}\right)\right)^2}, \tag{2.30}$$

і чим більше α, тим меншим буде цей кут.

Якщо $\alpha = \infty$, то $\mathsf{DFPR}(\alpha)$-алгоритм буде рівносильним граничному варіанту r-алгоритмів і для квадратичних функцій забезпечуватиме збіжність до x^* за n кроків.

З (2.30) випливає, що якщо в $\mathsf{DFPR}(\alpha)$-алгоритмі замінити постійний коефіцієнт α на змінний $\alpha_k = \dfrac{\|g_k\|}{\|g_{k+1}\|} + \dfrac{\|g_{k+1}\|}{\|g_k\|} \geq 2$, то кут між послідовними градієнтами зменшуватиметься рівно вдвічі, тобто з $\pi/2$ до $\pi/4$.

Ще одну модифікацію $\mathsf{DFPR}(\alpha)$-алгоритму забезпечує такий вибір α_k:

$$\alpha_k = \frac{\|g_{k+1}\|}{\|g_k\|}\left(\frac{\|g_k\|}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{\|g_k\|}\right) = 1 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} > 1,$$

при якому кут між послідовними градієнтами зменшуватиметься згідно зі співвідношенням

$$\cos\psi_k = 1 \Bigg/ \sqrt{1 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2}}.$$

При такому виборі α_k дуже простого вигляду набувають як однорангова матриця оберненого перетворення $T_k(g_k,\ g_{k+1})$

$$T_k(g_k,\ g_{k+1}) = I - \frac{1}{\|g_{k+1} - g_k\|^2} g_{k+1}(g_{k+1} - g_k)^T,$$

так і матриця перетворення простору аргументів

$$T_k^{-1}(g_k,\ g_{k+1}) = I - \frac{1}{\|g_k\|^2} g_{k+1}(g_{k+1} - g_k)^T.$$

Враховуючи, що перетворення простору спрямовано на зменшення кута між двома послідовними градієнтами, $DFPR(\alpha)$-алгоритм має бути ефективним для мінімізації гладких яружних функцій в широкому діапазоні значень параметра α, в тому числі при $\alpha \in [2, 4]$.

Підтвердженням цього можуть слугувати наведені в таблицях 2.1 та 2.2 результати чисельних експериментів для мінімізації квадратичних функцій

$$f(x) = \frac{1}{2} x^T A x = \frac{1}{2} \sum_i^n q^{i-1} x_i^2 = Quad(q, n),$$

в тому числі й сильно яружних.

Для порівняння тут наведено також результати роботи $r_\mu(\alpha)$-алгоритму за тих же значень α та параметрі $\mu = 0$ ($r_0(\alpha)$-алгоритм).

Обидва методи працювали в однакових умовах: за стартову точку обирали $x_0 = (1, 1, \ldots, 1)$; критерій зупинки — $\varepsilon_g = 10^{-10}$; крок найшвидшого спуску обчислювався аналітично

$$h_k = \frac{(g_k,\ \xi_k)}{(A\xi_k,\ \xi_k)},$$

і при цьому точність виконання найшвидшого спуску в була досить високою, тобто

$$\left| \left(\frac{\widetilde{g}_k}{\|\widetilde{g}_k\|}, \frac{\widetilde{g}_{k+1}}{\|\widetilde{g}_{k+1}\|} \right) \right| < 10^{-14}.$$

Отже, можна вважати, що результати експериментів, наведені в таблицях 2.1 та 2.2, показують, наскільки перетворення типу ДФП-методу краще, ніж розтяг простору в напрямку різниці двох послідовних градієнтів, другий з яких отримано згідно з кроком найшвидшого спуску в перетвореному просторі аргументів.

Для несильно яружних квадратичних функцій $Quad(1.1, 70)$ та $Quad(1.2, 50)$ робота $DFPR(\alpha)$-алгоритму при невеликих α близька до роботи ДФП-методу в тому сенсі, що забезпечує збіжність за кількість кроків, дуже близьку до n.

Таблиця 2.1: Чисельні експерименти при малих значеннях α [29]

$\mathrm{Quad}(q, n)$	**DFPR(α)-алгоритм**			**$r_0(\alpha)$-алгоритм**		
	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$
$\mathrm{Quad}(1.1, 200)$	732	581	508	1168	885	775
$\mathrm{Quad}(1.1, 130)$	288	241	218	496	419	398
$\mathrm{Quad}(1.1, 70)$	88	79	74	178	176	183
$\mathrm{Quad}(1.2, 100)$	337	273	239	627	494	457
$\mathrm{Quad}(1.2, 50)$	80	69	66	191	176	173
$\mathrm{Quad}(2.0, 30)$	103	83	76	273	218	206

Таблиця 2.2: Чисельні експерименти при великих значеннях α [29]

$\mathrm{Quad}(q, n)$	**DFPR(α)-алгоритм**			**$r_0(\alpha)$-алгоритм**		
	$\alpha = 10$	$\alpha = 100$	$\alpha = 1000$	$\alpha = 10$	$\alpha = 100$	$\alpha = 1000$
$\mathrm{Quad}(1.1, 200)$	379	271	221	702	692	550
$\mathrm{Quad}(1.1, 130)$	177	131	130	391	384	290
$\mathrm{Quad}(1.1, 70)$	70	70	70	212	177	140
$\mathrm{Quad}(1.2, 100)$	181	133	107	422	365	276
$\mathrm{Quad}(1.2, 50)$	54	50	50	185	143	106
$\mathrm{Quad}(2.0, 30)$	58	42	36	178	120	87

Звісно, розрив у кількості ітерацій між $\mathrm{DFPR}(\alpha)$-алгоритмом та $r_0(\alpha)$-алгоритмом має зменшуватись при збільшенні α, але це має місце при дуже великих значеннях α. Зокрема, при $\alpha = 1000$ (таблиця 2.2), яке не можна вважати малим, цей розрив досі є досить великим. Якщо $\mathrm{DFPR}(\alpha)$-алгоритм гарантує збіжність до x^* практично за n кроків, то для $r_0(\alpha)$-алгоритмом це не так. Однією з причин такої поведінки методів є занадто велика точність розв'язання задач $\varepsilon_g = 10^{-10}$, що рівносильно

$$f(x_k) - f^* < 10^{-20}.$$

У випадку менш жорсткого критерію зупинки розрив буде меншим, але й кількість ітерацій для $\mathsf{DFPR}(\alpha)$-алгоритму буде меншою. Але така точність для функції, що мінімізується, свідчить про стійкість $\mathsf{DFPR}(\alpha)$-алгоритму до точності розв'язку навіть для сильно яружних задач.

Отже, для погано обумовлених задач $\mathsf{DFPR}(\alpha)$-алгоритм можна вважати більш стійким, ніж $r_0(\alpha)$-алгоритм, точніше – лінійне перетворення простору типу ДФП-методу варто признати більш раціональним, ніж розтяг простору в напрямку різниці двох послідовних градієнтів.

2.6. Про субградієнтні методи типу ДФП-методу

$\mathsf{DFPR}(\alpha)$-алгоритм має «ідеалізований» характер у тому значенні, що він передбачає точну процедуру найшвидшого спуску, яка нереалізовна навіть для гладких функцій. Крім того, для негладких функцій точна процедура найшвидшого спуску гарантує сильнішу умову, ніж існування антисубградієнта, ортогонального до напрямку спуску. Зокрема, кут може бути тупим, тобто $\left(g_k, g_{k+1}\right) < 0$. Тому перенесення основних принципів розглянутого $\mathsf{DFPR}(\alpha)$-алгоритму на загальний випадок мінімізації опуклих функцій вимагає вирішення двох головних питань.

Перше — заміна перетворення (2.29) перетворенням, яке б забезпечувало вибір напрямку руху, як в $\mathsf{DFPR}(\alpha)$-алгоритмі, не лише при ортогональних послідовних субградієнтах. Друге — заміна точного найшвидшого спуску в напрямку субградієнта іншим регулюванням крокового множника, яке б потребувало невеликої кількості обчислень $f(x)$ та $\partial f(x)$.

Перше питання не є проблемою, оскільки незалежно від кута між g_k та g_{k+1} в $Y_k = B_k^{-1}X$, вибір напрямку руху як в $\mathsf{DFPR}(\alpha)$-алгоритмі можна зберегти, якщо для корекції матриці $B_{k+1} = B_k T_k(g_k,\ g_{k+1})$ вибрати

$$T_k(g_k,\ g_{k+1}) = I - \left(\frac{g_{k+1} - g_k}{\|g_{k+1} - g_k\|} + t_k\left(g_k - \frac{(g_{k+1},\ g_k)}{\|g_{k+1}\|^2} g_{k+1}\right)\right) \times \\ \times \left(\frac{g_{k+1} - g_k}{\|g_{k+1} - g_k\|}\right)^T. \tag{2.31}$$

Тут t_k — певний скалярний параметр, від вибору якого залежить збіжність методів.

Друге питання пов'язане з регулюванням крокового множника в напрямку субградієнту та є більш складним і призводить до різного роду модифікацій. Зокрема, класичний феєрівський крок в напрямку субградієнта або певний його аналог, якщо f^* відоме, призводить до немонотонних за функціоналом субградієнтних методів «змінної метрики» на основі перетворення (2.31). При цьому очевидно, що перетворення простору потрібне на тих кроках методу, коли для найкоротшої опуклої комбінації векторів g_k та g_{k+1} виконуються $\lambda_1 > 0$ та $\lambda_2 > 0$, як в (2.27)–(2.28).

Лінійне перетворення простору (2.31) сильніше розширює конус можливих напрямків спадання функції, ніж «ортогоналізація» в подібних методах [32, ст. 239–244]. Обґрунтування збіжності таких методів можна побудувати аналогічно до того, як це виконано в [32], прибравши навіть вимогу на обмеженість евклідової норми матриці B_k. Такий спосіб регулювання крокового множника потребує не більше одного обчислення $f(x)$ та $\partial f(x)$ на кожному кроці методу і дозволяє за скінчену кількість кроків K або знайти $f(x_K)$ зі значенням функції, яке потребується, або отримати достатні умови розбіжності фейєрівського процесу, коли замість f^* використовується занижене значення.

Адаптивний спосіб регулювання кроку [7], який часто використовується в $r(\alpha)$-алгоритмі, дозволяє отримати майже монотонні по функціоналу методи. Формально він призводить до заміни точного пошуку мінімуму за напрямком на наближений, тобто $h_k > h_k^*$, але так, щоб h_k було близьким до h_k^*. Тут h_k^* — точний крок найшвидшого спуску в напрямку антисубградієнта. Враховуючи, що гарантоване зменшення норми субградієнта в перетвореному просторі дозволяє просто уточнити крок найшвидшого спуску в перетвореному просторі, за такого регулювання крокового множника достатньо на кожному кроці методу використовувати в середньому 2-3 обчислення $f(x)$ та $\partial f(x)$. При адаптивному регулюванні крокового множника можна переходити і в точки з меншим кроком, ніж h_k^*. Таким чином, одночасне виконання умов $\|g_k\|^2 > \left(g_k, g_{k+1}\right)$ та $\|g_{k+1}\|^2 > \left(g_k, g_{k+1}\right)$ дозволяє застосовувати перетворення (2.31) і коли $\left(g_k, g_{k+1}\right) > 0$.

Отже, для мінімізації опуклих функцій як перший, так і другий способи регулювання крокового множника дозволяють отримати практично реалізовні субградієнтні методи «змінної метрики» на основі перетворення типу ДФП-методу в умовах мінімального використання інформації (тільки два послідовних субградієнти).

Відмітимо, що обговорені вище субградієнтні методи «змінної метрики» більше відповідають назві роботи [33], ніж методи, які запропоновано в цій роботі, які стали основою для створення ε-субградієнтних методів в негладкій оптимізації. Насправді, в ε-субградієнтних методах з ДФП-методу був запозичений лише принцип руху за напрямком, протилежним до найкоротшого вектора до опуклої оболонки двох послідовних субградієнтів, і узагальнений на випадок більшої кількості векторів. Але зміст ДФП-методу полягає не скільки у виборі такого напрямку руху, скільки в лінійному перетворенні простору, який застосовується, і яке дозволяє покращувати структуру поверхонь рівня функції, що мінімізується.

2.7. Висновки

В роботі обговорюється B-форма методу Давидона–Флетчера–Пауелла, яка дозволяє інтерпретувати його як градієнтний метод в перетвореному просторі аргументів. На її основі проведено порівняння цього методу з r-алгоритмами. Для мінімізації гладких опуклих функцій побудовано $\mathsf{DFPR}(\alpha)$-алгоритм — градієнтний метод з перетворенням простору, який поєднує властивості як квазіньютонівських методів, так і r-алгоритмів. Обговорюються можливі схеми такого типу методів для мінімізації негладких опуклих функцій.

Відмітимо, що алгоритм, схожий на $\mathsf{DFPR}(\alpha)$-алгоритм, можна побудувати також на основі лінійного однорангового оператора для методу Бройдена–Флетчера–Гольдфарба–Шанно [34]. Цей оператор має вид

$$T_k(g_k, g_{k+1}) =$$
$$= I + \frac{g_k}{(g_k - g_{k+1}, g_k)} \left(\sqrt{\frac{h_k(g_k, g_k - g_{k+1})}{(g_k, g_k)}}\, g_k + g_{k+1} - g_k \right)^T,$$

і використовується для перерахунку матриці, оберненої до матриці перетворення простору. Такі алгоритми тісно пов'язані з методом Левенберга–Марквардта [26, 35, 36], який є комбінацією методу Ньютона та методу градієнтного спуску.

Розділ 3

Модифіковані екстраградієнтні алгоритми

В розділі розглядаються різні варіанти модифікованого екстраградієнтного методу з динамічним регулюванням величини кроку для розв'язання варіаційних нерівностей з монотонними операторами. Основний матеріал розділу вперше був опублікований в статтях [37, 38].

Розділ побудовано наступним чином. В підрозділі 3.1 наведено необхідний мінімум відомостей стосовно варіаційних нерівностей та факти, що відіграють важливу роль у доведеннях результатів розділу. В підрозділі 3.2 описано класичний екстраградієнтний алгоритм Корпелевич та доведено його збіжність. В підрозділі 3.3 описано модифікацію субградієнтного екстраградієнтного алгоритму з динамічним регулюванням величини кроку для варіаційних нерівностей з монотонними неліпшицевими операторами та доведено його збіжність. Підрозділ 3.4 присвячений алгоритму для пошуку розв'язку варіаційної нерівності з апріорною інформацією, що описана у вигляді включення до множини нерухомих точок квазінерозтягуючого оператора. Розглянуто алгоритм для операторних рівнянь з апріорною інформацією про розв'язок. В підрозділі 3.5 розгянуто сильно збіжні екстраградієнтні методи з динамічним регулюванням величини кроку для розв'язання монотонних варіаційних нерівностей та операторних рівнянь з апріорною інформацією. Для регуляризації використано схему Гальперна, проекційну CQ-схему Nakajo–Takahashi та метод Takahashi–Takeuchi–Kubota. У кінці розділу розглянуто деякі застосування варіаційних нерівностей у машинному навчанні.

3.1. Попередні відомості та допоміжні твердження

Нехай H — дійсний гільбертовий простір із заданим скалярним добутком $(\cdot,\cdot)$ та нормою $\|\cdot\|$, що породжена цим скалярним добутком. Сильну збіжність будемо позначати символом $\to$, а слабку $\rightharpoonup$.

Нехай C — непорожня опукла і замкнена підмножина простору H та $A : H \to H$ – деякий оператор. У даному розділі ми розглянемо екстраградієнтні методи розв'язання варіаційних нерівностей вигляду:

$$\text{знайти } x \in C: \ (Ax, y - x) \geq 0 \quad \forall y \in C. \tag{3.1}$$

Будемо позначати $VI(A, C)$ множину розв'язків варіаційної нерівності (3.1).

Варіаційні нерівності один з центральних об'єктів у прикладному нелінійному аналізі. Дослідження варіаційних нерівностей розпочалося в 1960-х роках [39–42]. Ці роботи виникли при спробах розв'язання проблем варіаційного числення, оптимального керування та теорії крайових задач з односторонніми умовами. Гарний огляд перших застосувань варіаційних нерівностей дають книги [43–45].

Дуже швидко варіаційні нерівності стали потужним інструментом для розв'язання проблем оптимізації, оптимального керування, економіки, математичної фізики. В оптимізації та оптимальному керуванні варіаційні нерівності є перш за все зручним способом запису умов оптимальності. Відома монографія Ж.-Л. Ліонса [46] є систематичним викладенням теорії оптимального керування системами з розподіленими параметрами з точки зору варіаційних нерівностей.

Окремі задачі опуклої недиференційовної оптимізації можуть ефективно розв'язуватися, якщо їх переформулювати у вигляді сідлових (мінімаксних) задач і застосувати алгоритми розв'язання варіаційних нерівностей [47]. Нещодавно був розвинутий такий варіант побудови швидких алгоритмів для задач опуклого програмування: за допомогою теорії двоїстості переходимо до деякої опукло-угнутої сідлової задачі (гра Фенхеля) та застосовуємо екстраградієнтні алгоритми розв'язання варіаційних нерівностей [48].

З появою генеруючих змагальних нейронних мереж (generative adversarial networks, GANs) та інших моделей змагального навчання [90,91] стійкий інтерес до алгоритмів розв'язання варіаційних нерівностей виник і в середовищі спеціалістів в галузі машинного навчання [49–53].

Нагадаємо декілька відомих та потрібних нам фактів.

Нехай P_C — оператор метричного проектування на замкнену опуклу підмножину $C \subseteq H$, тобто $P_C x$ — єдиний елемент C, що володіє властивістю

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

Елемент $P_C x$ можна охарактеризувати таким чином [44]:

$$y = P_C x \quad \Leftrightarrow \quad y \in C \text{ та } (y - x, z - y) \geq 0 \ \forall z \in C, \tag{3.2}$$

$$y = P_C x \quad \Leftrightarrow \quad y \in C \text{ та } \|y - z\|^2 \leq \|x - z\|^2 - \|y - x\|^2 \ \forall z \in C. \tag{3.3}$$

Із нерівності (4.4) випливає, що $x \in VI(A, C)$ тоді і тільки тоді, коли

$$x = P_C (x - \lambda A x),$$

де $\lambda > 0$ [44].

Означення 3.1. Оператор $A : H \to H$ називаємо монотонним, якщо

$$(Ax - Ay, x - y) \geq 0 \quad \forall x, \ y \in H.$$

Має місце

Лема 3.1 (лема Мінті). *Якщо оператор $A : H \to H$ — монотонний та неперервний, а множина $C \subseteq H$ — опукла та замкнена, то $x \in VI(A, C)$ тоді і тільки тоді, коли $x \in C$ та*

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

Зокрема, множина $VI(A, C)$ опукла та замкнена.

Зауваження 3.1. Варіаційна нерівність (3.1) записується у формі операторного включення [54]:

$$\text{знайти } x \in H : \ 0 \in Ax + N_C x,$$

де $N_C x$ — нормальний конус множини C в точці x:

$$N_C x = \begin{cases} \{z \in H : \ (z, y - x) \leq 0 \ \forall y \in C\}, & \text{якщо } x \in C, \\ \emptyset, & \text{інакше}. \end{cases}$$

При доведенні слабкої збіжності послідовностей елементів гільбертового простору будемо використовувати відому лему Опяла.

Лема 3.2 (Opial, [55]). *Нехай послідовність (x_n) елементів гільбертового простору H слабко збігається до елемента $x \in H$. Тоді для всіх $y \in H \backslash \{x\}$ маємо*

$$\liminf_{n \to \infty} \|x_n - x\| < \liminf_{n \to \infty} \|x_n - y\|.$$

Наступні факти також відіграють важливу роль у доведеннях основних результатів розділу.

Лема 3.3. *Нехай послідовність невід'ємних чисел* (ξ_n) *задовольняє рекурентну нерівність*

$$\xi_{n+1} \leq (1 - \alpha_n)\,\xi_n + \alpha_n\beta_n + \gamma_n, \quad n \geq 1,$$

де послідовності (α_n) *та* (β_n) *мають властивості:*

1) $\alpha_n \in (0, 1)$, $\sum_{n=1}^{\infty} \alpha_n = +\infty$*;*

2) $\limsup_{n\to\infty} \beta_n \leq 0$*;*

3) $\gamma_n \in [0, +\infty)$, $\sum_{n=1}^{\infty} \gamma_n < +\infty$*.*

Тоді $\lim\limits_{n\to\infty} \xi_n = 0$.

Лема 3.4 (Mainge, [56])**.** *Нехай числова послідовність* (a_n) *має підпослідовність* (a_{n_k})*, яка володіє властивістю*

$$a_{n_k} < a_{n_k+1} \quad \forall k \in \mathbb{N}.$$

Тоді існує така неспадна послідовність (m_k) *натуральних чисел, що*

$$m_k \to +\infty \quad \textit{i} \quad a_{m_k} \leq a_{m_k+1},\ a_k \leq a_{m_k+1}\ \forall k \geq n_1.$$

Зауваження 3.2. Лема 3.4 є ефективним інструментом дослідження збіжності ітераційних процесів, що не володіють фейєрівською властивістю відносно множини розв'язків [56–58].

Розглянемо функцію

$$t \mapsto \|x - P_C\,(x - tAx)\|\,,\ \ t \in \mathbb{R}.$$

Вона володіє наступними корисними властивостями.

Лема 3.5. *Для* $x \in H$ *і чисел* $\alpha \geq \beta > 0$ *мають місце нерівності*

$$\frac{\|x - P_C\,(x - \alpha Ax)\|}{\alpha} \leq \frac{\|x - P_C\,(x - \beta Ax)\|}{\beta},$$

$$\|x - P_C\,(x - \beta Ax)\| \leq \|x - P_C\,(x - \alpha Ax)\|\,.$$

Доведення. Покладемо $x_\alpha = P_C(x - \alpha A x)$, $x_\beta = P_C(x - \beta A x)$. Маємо

$$\left(\frac{x_\alpha - x + \alpha A x}{\alpha}, x_\beta - x_\alpha\right) \geq 0,$$

$$\left(\frac{x_\beta - x + \beta A x}{\beta}, x_\alpha - x_\beta\right) \geq 0.$$

Склавши ці нерівності, отримаємо

$$0 \leq \left(\frac{x - x_\alpha}{\alpha} - \frac{x - x_\beta}{\beta}, x_\alpha - x_\beta\right) = \left(\frac{x - x_\alpha}{\alpha} - \frac{x - x_\beta}{\beta}, (x - x_\beta) - (x - x_\alpha)\right).$$

Звідки,

$$0 \leq -\|x - x_\alpha\|^2 - \frac{\alpha}{\beta}\|x - x_\beta\|^2 + \|x - x_\alpha\|\,\|x - x_\beta\| + \frac{\alpha}{\beta}\|x - x_\alpha\|\,\|x - x_\beta\|.$$

Отже,

$$0 \geq \left(\|x - x_\alpha\| - \frac{\alpha}{\beta}\|x - x_\beta\|\right)(\|x - x_\alpha\| - \|x - x_\beta\|).$$

Звідки випливає

$$\|x - x_\alpha\| - \frac{\alpha}{\beta}\|x - x_\beta\| \leq 0, \quad \|x - x_\beta\| \leq \|x - x_\alpha\|,$$

що і потрібно було довести. □

Означення 3.2. Оператор $T : H \to H$ називають нерозтягуючим, якщо

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

Оператор проектування P_C — нерозтягуючий.

Означення 3.3. Оператор $T : H \to H$ називають квазінерозтягуючим (фейєрівським), якщо $F(T) = \{x \in H : \ Tx = x\} \neq \emptyset$ та

$$\|Tx - y\| \leq \|x - y\| \quad \forall x \in H\ \forall y \in F(T).$$

Відомо, що множина нерухомих точок $F(T)$ квазінерозтягуючого оператора замкнена та опукла [54, 59].

Зауваження 3.3. Квазінерозтягуючі (фейєрівські) оператори та породжені ними ітераційні процеси мають велике значення в оптимізаційній алгоритмиці (див. [59–61]). Але в багатьох інших областях обчислювальної математики ці плідні конструкції не достатньо відомі. Книга [59] — одне з найкращих джерел по ітераційним алгоритмам з фейєрівською властивістю для розв'язання рівнянь, нерівностей та задач оптимізації.

Означення 3.4. Оператор $S : C \to H$ називають демізамкненим в точці $y \in H$ якщо для послідовності точок $x_n \in C$ із $x_n \rightharpoonup x$ і $Sx_n \to y$ випливає $Sx = y$.

Відомо, що для нерозтягуючого оператора $T : C \to H$ оператор $I - T$ демізамкнений в нулі [54].

Основними схемами, що забезпечують апроксимацію нерухомих точок нерозтягуючих операторів (в припущенні їх існування) є алгоритми Красносельського–Манна [55, 62, 63], Гальперна [64–71], гібридний алгоритм Nakajo–Takahashi [72] та алгоритм скорочення (shrinking algorithm) Takahashi–Takeuchi–Kubota [73]. Розглянемо їх докладніше.

Метод Красносельського–Манна генерує послідовність (x_n) за допомогою схеми

$$x_{n+1} = \alpha x_n + (1 - \alpha)Tx_n,$$

де $\alpha \in (0, 1)$. Відомо, що (x_n) слабко збігається до деякої точки з $F(T)$ [55].

У 1967 р. Б. Гальперн [64] дослідив збіжність методу

$$x_{n+1} = \alpha_n y + (1 - \alpha_n)Tx_n, \tag{3.4}$$

де $y \in C$, а (α_n) — послідовність чисел з $(0, 1)$. Він показав, що умови

$$\lim_{n\to\infty} \alpha_n = 0, \quad \sum_{n=1}^{\infty} \alpha_n = +\infty \tag{3.5}$$

є необхідними для сильної збіжності послідовності (x_n) до нерухомої точки нерозтягуючого оператора T, та отримав достатні умови сильної збіжності. Зокрема, Гальперн довів збіжність (3.4) при $\alpha_n = \frac{1}{(n+1)^q}$, де $q \in (0, 1)$.

У 1977 р. П.–Л. Ліонс [65] покращив результат Гальперна, довівши сильну збіжність (3.4), якщо послідовність (α_n) задовольняє умови (3.5) та

$$\lim_{n\to\infty} (\alpha_{n+1} - \alpha_n)/\alpha_{n+1}^2 = 0. \tag{3.6}$$

У 1992 р. Р. Віттман [66] довів сильну збіжність методу (3.4) до нерухомої точки оператора T за виконання умов (3.5) та

$$\sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n| < +\infty. \tag{3.7}$$

Нарешті, у 2002 р. H.K. Xu [67, 68] довів теорему про сильну збіжність методу (3.4) за виконання умов (3.5) та

$$\lim_{n\to\infty} (\alpha_{n+1} - \alpha_n)/\alpha_{n+1} = 0. \tag{3.8}$$

У додатку до розділу наведемо доведення збіжності методу (3.4).

Відомо, що у загальному випадку метод Красносельського–Манна лише слабко збігається. Існує кілька сильно збіжних його модифікацій. Однією з найпопулярніших є CQ–метод (відомий під назвою «гібридний метод»), що запропонований у 2003 р. K. Nakajo та W. Takahashi [72]. Метод має вигляд:

$$\begin{cases} x_1 = x \in C, \\ y_n = \alpha_n x_n + (1-\alpha_n) T x_n, \\ C_n = \{z \in C : \|z - y_n\| \le \|z - x_n\|\}, \\ Q_n = \{z \in C : (x_n - z, x_0 - x_n) \ge 0\}, \\ x_{n+1} = P_{C_n \cap Q_n} x, \end{cases}$$

де (α_n) — послідовність чисел із $[0, \alpha]$ для деякого $\alpha \in [0, 1)$. У роботі [72] доведено сильну збіжність послідовності (x_n) до точки $z = P_{F(T)} x$.

У 2008 р. W. Takahashi, Y. Takeuchi, та R. Kubota [73] запропонували такий алгоритм (shrinking algorithm)

$$\begin{cases} x_1 = x \in C, \ \ C_1 = C, \\ y_n = \alpha_n x_n + (1-\alpha_n) T x_n, \\ C_{n+1} = \{z \in C_n : \|z - y_n\| \le \|z - x_n\|\}, \\ x_{n+1} = P_{C_{n+1}} x, \end{cases}$$

де (α_n) — послідовність чисел із $[0, \alpha]$ для деякого $\alpha \in [0, 1)$. Вони довели, породжена цим алгоритмом послідовність (x_n) сильно збігається до точки $z = P_{F(T)} x$.

3.2. Екстраградієнтний алгоритм Корпелевич

Рівносильність варіаційної нерівності (3.1) та задачі пошуку нерухомої точки оператора $P_C (I - \lambda A)$, де $\lambda > 0$, веде до очевидного методу, що генерує послідовність (x_n) за правилом

$$x_{n+1} = P_C (x_n - \lambda A x_n), \tag{3.9}$$

де $\lambda > 0$.

Для збіжності (x_n) до розв'язку (3.1) необхідно накласти на A більш сильні умови, ніж просто монотонність. Зокрема, якщо A — сильно монотонний та ліпшицевий або обернено сильно монотонний (ко-коерцитивний), то послідовність (x_n) збігається[1] до розв'язку (3.1) (у першому випадку — сильно, у другому — слабко).

У випадку лише монотонності оператора A найпростіша схема (3.9) не гарантує збіжності. У 1976 р. Г.М. Корпелевич запропонувала для розв'язання (3.1) в евклідовому просторі $\mathbb{R}^m$, так званий, екстраградієнтний алгоритм [74] (див. також роботу А.С. Антіпіна [75]):

$$\begin{cases} y_n = P_C\left(x_n - \lambda A x_n\right), \\ x_{n+1} = P_C\left(x_n - \lambda A y_n\right), \end{cases}$$

де $\lambda > 0$.

Докладно розглянемо цей популярний алгоритм. Спираємось на матеріал книги [76].

Будемо припускати виконаними такі умови:

A1) множина $C \subseteq H$ замкнена та опукла;

A2) оператор $A : H \to H$ монотонний та ліпшицевий (зі сталою $L > 0$) на множині C;

A3) $VI(A, C) \neq \emptyset$.

Алгоритм 3.1 (Екстраградієнтний алгоритм Корпелевич).

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.*

---

[1] Величина крокового множнику λ у першому випадку обирається з інтервалу $\left(0, 2\mu/L^2\right)$, де $\mu > 0$, $L > 0$ — сталі сильної монотонності та Ліпшиця оператора A, відповідно, у другому випадку — з $(0, 2\alpha)$, де $\alpha > 0$ стала ко-коерцитивності A. Нагадаємо, що оператор $A : H \to H$ називають ко–коерцитивним (обернено сильно монотонним), якщо існує додатня константа α така, що

$$(Ax - Ay, x - y) \geq \alpha \|Ax - Ay\|^2 \quad \forall x, y \in H.$$

Наступна лема обґрунтовує правило зупинки.

Лема 3.6. *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
$$\left(y_n - x_n + \lambda A x_n, x - y_n\right) \geq 0 \quad \forall x \in C.$$
З урахуванням умови $x_n = y_n$ маємо $x_n \in VI(A, C)$. □

Принциповим елементом доведення збіжності є нерівність з наступної леми. Саме вона гарантує фейєрівську властивість відносно множини розв'язків.

Лема 3.7. *Для* $z \in VI(A, C)$ *і породжених алгоритмом 3.1 послідовностей* (x_n), (y_n) *виконується нерівність*
$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \left(1 - \lambda^2 L^2\right) \|x_n - y_n\|^2. \tag{3.10}$$

Доведення. Маємо
$$\begin{aligned}\|x_{n+1} - z\|^2 \leq \|x_n - \lambda A y_n - z\|^2 - \|x_n - \lambda A y_n - x_{n+1}\|^2 = \\ = \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + 2\lambda\left(A y_n, z - x_{n+1}\right). \end{aligned} \tag{3.11}$$

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
$$\left(A y_n, z - x_{n+1}\right) \leq \left(A y_n, y_n - x_{n+1}\right). \tag{3.12}$$

Оцінивши праву частину (3.11) за допомогою (3.12), отримаємо
$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + 2\lambda\left(A y_n, y_n - x_{n+1}\right).$$

Далі
$$\begin{aligned}\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|\left(x_n - y_n\right) + \left(y_n - x_{n+1}\right)\|^2 + \\ + 2\lambda\left(A y_n, y_n - x_{n+1}\right) = \|x_n - z\|^2 - \|x_n - y_n\|^2 - \\ - \|y_n - x_{n+1}\|^2 + 2\left(x_n - \lambda A y_n - y_n, x_{n+1} - y_n\right).\end{aligned}$$

Оскільки $x_{n+1} \in C$, 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,

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_n - y_n\|^2 - \\ - \|y_n - x_{n+1}\|^2 + 2\lambda (A x_n - A y_n, x_{n+1} - y_n) . \quad (3.13)$$

Перейдемо до оцінки $2\lambda (A x_n - A y_n, x_{n+1} - y_n)$. Маємо

$$2\lambda (A x_n - A y_n, x_{n+1} - y_n) \leq 2\lambda L \|x_n - y_n\| \|x_{n+1} - y_n\| \leq \\ \leq \lambda^2 L^2 \|x_n - y_n\|^2 + \|x_{n+1} - y_n\|^2 . \quad (3.14)$$

Використовуючи (3.14) у (3.13), 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 3.1. *Породжені алгоритмом 3.1 послідовності* (x_n), (y_n) *слабко збігаються до розв'язку (3.1).*

Доведення. З нерівності (3.10) випливає існування $\lim_{n\to\infty} \|x_n - z\|$ для всіх $z \in VI(A, C)$. Зокрема, послідовність (x_n) обмежена.

Запишемо (3.10) у вигляді

$$(1 - \lambda^2 L^2) \|x_n - y_n\|^2 \leq \|x_n - z\|^2 - \|x_{n+1} - z\|^2 .$$

Маємо

$$\sum_{n=0}^{\infty} \|x_n - y_n\|^2 < +\infty.$$

Звідси

$$\lim_{n\to\infty} \|x_n - y_n\| = 0.$$

Отже, (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

$$\begin{aligned} 0 \le (y_{n_k} - x_{n_k}, x - y_{n_k}) + \lambda (Ax_{n_k}, x_{n_k} - y_{n_k}) + \lambda (Ax_{n_k}, x - x_{n_k}) \le \\ \le (y_{n_k} - x_{n_k}, x - y_{n_k}) + \lambda (Ax_{n_k}, x_{n_k} - y_{n_k}) + \lambda (Ax, x - x_{n_k}) . \end{aligned} \tag{3.15}$$

Перейшовши до границі у (3.15), отримаємо

$$(Ax, x - x^*) \ge 0 \quad \forall x \in C,$$

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 Опяла (лема 3.2) випливає:

$$\lim_{n\to\infty} \|x_n - p\| = \lim_{k\to\infty} \|x_{n_k} - p\| < \lim_{k\to\infty} \|x_{n_k} - q\| = \lim_{n\to\infty} \|x_n - q\| .$$

Повторивши це міркування, приходимо до абсурдної нерівності

$$\lim_{n\to\infty} \|x_n - p\| < \lim_{n\to\infty} \|x_n - p\| .$$

Отже, породжена алгоритмом 3.1 послідовність (x_n) слабко збігається до точки з множини $VI(A, C)$. □

У 2011 р. Y. Censor, A. Gibali та S. Reich [77] запропонували модифікацію алгоритму Корпелевич з одним метричним проектуванням на допустиму множину C — так званий, субградієнтний екстраградієнтний алгоритм, що має вигляд:

$$\begin{cases} y_n = P_C (x_n - \lambda A x_n) , \\ T_n = \{z \in H : \ (x_n - \lambda A x_n - y_n, z - y_n) \le 0\}, \\ x_{n+1} = P_{T_n} (x_n - \lambda A y_n) , \end{cases} \tag{3.16}$$

де $\lambda \in (0, 1/L)$. Друга проекція в (3.16) має явну форму. Тому цей метод є привабливим, коли проектування на допустиму множину C є складною задачею.

Зазначимо, що незалежно алгоритм (3.16) був запропонований в [78] в контексті задач рівноважного програмування.

В роботах [77, 78] доведена слабка збіжність породжених цим алгоритмом послідовностей (x_n) і (y_n) до деякого розв'язку варіаційної нерівності (3.1).

Близьким до алгоритму Корпелевич є P. Tseng [79]

$$\begin{cases} y_n = P_C\left(x_n - \lambda A x_n\right), \\ x_{n+1} = y_n + \lambda\left(A x_n - A y_n\right), \end{cases} \tag{3.17}$$

де $\lambda \in (0, 1/L)$. В роботі [79] показано, що послідовності (x_n) і (y_n) слабко збіжні до деякого розв'язку нерівності (4.1).

Обидва алгоритми (3.16) і (3.17) мають однакову складність ітераційного кроку: одне проектування на C та обчислення двох значень оператора A.

Зауваження 3.4. Існує така адаптивна модифікація екстраградієнтного методу [80].

1) Задаємо $x_1 \in C$, $\tau \in (0,1)$, $\lambda_1 \in (0, +\infty)$.

2) Для x_n *обчислюємо* $y_n = P_C\left(x_n - \lambda_n A x_n\right)$.

3) Якщо $x_n = y_n$, *то СТОП, інакше обчислюємо*

$$x_{n+1} = P_C\left(x_n - \lambda_n A y_n\right).$$

4) Обчислюємо

$$\lambda_{n+1} = \begin{cases} \min\left\{\lambda_n, \tau\frac{\|x_n - y_n\|}{\|Ax_n - Ay_n\|}\right\}, & \text{якщо } \ Ax_n \neq Ay_n, \\ \lambda_n, & \text{інакше.} \end{cases}$$

покладаємо $n := n+1$ *та переходимо на крок 2.*

Перейдемо до розгляду модифікацій екстраградієнтного методу Корпелевич для монотонних варіаційних нерівностей з нелінійним оператором, що не має властивості глобальної ліпшицевості.

3.3. Модифікований екстраградієнтний алгоритм

Нижче ми розглянемо модифікацію субградієнтного екстраградієнтного алгоритму з динамічним регулюванням величини кроку.

Далі будемо вважати, що виконуються такі умови:

A1) $VI(A, C) \neq \emptyset$;

A2) оператор $A : H \to H$ — монотонний, рівномірно неперервний на обмежених множинах.

Алгоритм 3.2. *Задаємо числові параметри* $\sigma > 0$, $\tau \in (0,1)$, $\theta \in (0,1)$ *і елемент* $x_0 \in H$.

Ітераційний крок. *Для* $x_n \in H$ *обчислюємо*

$$y_n = P_C\left(x_n - \lambda_n A x_n\right),$$

де λ_n *отримуємо із умови*

$$\begin{cases} j\,(n) = \min\{j \geq 0 : \; \left\|AP_C\left(x_n - \sigma\tau^j A x_n\right) - Ax_n\right\| \leq \\ \qquad\qquad\qquad\qquad \leq \frac{\theta}{\sigma\tau^j}\left\|P_C\left(x_n - \sigma\tau^j A x_n\right) - x_n\right\|\}, \\ \lambda_n = \sigma\tau^{j(n)}. \end{cases} \tag{3.18}$$

Якщо $y_n = x_n$, *то закінчуємо, в противному випадку обчислюємо*

$$x_{n+1} = P_{T_n}\left(x_n - \lambda_n A y_n\right),$$

де

$$T_n = \{z \in H : \; (x_n - \lambda_n A x_n - y_n, z - y_n) \leq 0\}.$$

Зрозуміло, що якщо $y_n = x_n$, то точка x_n належить множині C і є розв'язком варіаційної нерівності.

Покажемо, що процедура (3.18) завжди закінчується за скінченну кількість кроків. Має місце

Лема 3.8. *Правило* (3.18) *вибору параметра* λ_n *коректне, тобто*

$$j\,(n) < +\infty.$$

Доведення. Нехай $x_n \in VI(A, C)$. Тоді

$$x_n = P_C\left(x_n - \sigma A x_n\right) \quad \text{та} \quad j\,(n) = 0.$$

Розглянемо ситуацію $x_n \notin VI(A, C)$ і припустимо, що для всіх $j \in \mathbb{N}$ виконується нерівність

$$\sigma\tau^j\left\|AP_C\left(x_n - \sigma\tau^j A x_n\right) - Ax_n\right\| > \theta\left\|P_C\left(x_n - \sigma\tau^j A x_n\right) - x_n\right\|. \tag{3.19}$$

Якщо $x_n \in C$, то

$$\lim_{j\to\infty}\left\|P_C\left(x_n - \sigma\tau^j A x_n\right) - x_n\right\| = 0.$$

З рівномірної неперервності оператора A на обмежених множинах випливає

$$\lim_{j\to\infty} \left\| AP_C\left(x_n - \sigma\tau^j Ax_n\right) - Ax_n \right\| = 0.$$

Таким чином,

$$\lim_{j\to\infty} \frac{\left\| P_C\left(x_n - \sigma\tau^j Ax_n\right) - x_n \right\|}{\sigma\tau^j} = 0. \tag{3.20}$$

Покладемо $y_n^j = P_C\left(x_n - \sigma\tau^j Ax_n\right)$. Маємо

$$\left(\frac{y_n^j - x_n}{\sigma\tau^j}, x - y_n^j \right) + \left(Ax_n, x - y_n^j \right) \geq 0 \quad \forall x \in C. \tag{3.21}$$

Здійснивши граничний перехід в (3.21) з урахуванням асимптотики (3.20), отримаємо $(Ax_n, x - x_n) \geq 0 \quad \forall x \in C$, тобто, $x_n \in VI(A, C)$. Протиріччя.

У випадку $x_n \notin C$ маємо

$$\lim_{j\to\infty} \left\| P_C\left(x_n - \sigma\tau^j Ax_n\right) - x_n \right\| = \| P_C x_n - x_n \| > 0.$$

та

$$\lim_{j\to\infty} \sigma\tau^j \left\| AP_C\left(x_n - \sigma\tau^j Ax_n\right) - Ax_n \right\| = 0.$$

Знову отримали протиріччя (див. нерівність (3.19)). □

Зауваження 3.5. При доведенні леми 3.8 зовсім не використовувалась монотонність оператора A.

Перейдемо тепер до доведення слабкої збіжності алгоритму 3.2.

Лема 3.9. *Для послідовностей* (x_n), (y_n), *породжених алгоритмом 3.2, має місце нерівність*

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \left(1 - \theta^2\right) \|x_n - y_n\|^2, \tag{3.22}$$

де $z \in VI(A, C)$.

Доведення. Нехай $z \in VI(A, C)$. Маємо

$$\begin{aligned} \|x_{n+1} - z\|^2 &= \|P_{T_n}(x_n - \lambda_n Ay_n) - z\|^2 = \\ &= \|P_{T_n}(x_n - \lambda_n Ay_n) - (x_n - \lambda_n Ay_n) + (x_n - \lambda_n Ay_n) - z\|^2 = \\ &= \|(x_n - \lambda_n Ay_n) - z\|^2 + \|P_{T_n}(x_n - \lambda_n Ay_n) - (x_n - \lambda_n Ay_n)\|^2 + \\ &\quad + 2\left(P_{T_n}(x_n - \lambda_n Ay_n) - (x_n - \lambda_n Ay_n), (x_n - \lambda_n Ay_n) - z\right). \end{aligned}$$

Оскільки

$$\begin{aligned}2\left\|P_{T_n}(x_n-\lambda_n Ay_n)-(x_n-\lambda_n Ay_n)\right\|^2+\\ +2\left(P_{T_n}(x_n-\lambda_n Ay_n)-(x_n-\lambda_n Ay_n),(x_n-\lambda_n Ay_n)-z\right)=\\ =2\left((x_n-\lambda_n Ay_n)-P_{T_n}(x_n-\lambda_n Ay_n),z-P_{T_n}(x_n-\lambda_n Ay_n)\right)\leq 0,\end{aligned}$$

то для всіх $n\in\mathbb{N}$ маємо

$$\begin{aligned}\|x_{n+1}-z\|^2\leq\|x_n-\lambda_n Ay_n-z\|^2-\|P_{T_n}(x_n-\lambda_n Ay_n)-x_n+\lambda_n Ay_n\|^2=\\ =\|(x_n-\lambda_n Ay_n)-z\|^2-\|x_{n+1}-(x_n-\lambda_n Ay_n)\|^2=\\ =\|x_n-z\|^2-\|x_n-x_{n+1}\|^2+2\lambda_n(z-x_{n+1},Ay_n).\end{aligned}$$

З монотонності оператора A, включення $z\in VI(A,C)$ та леми Мінті випливає

$$\begin{aligned}0\leq(Ay_n-Az,y_n-z)=\\ =(Ay_n,y_n-z)-\underbrace{(Az,y_n-z)}_{\geq 0}\leq\\ \leq(Ay_n,y_n-z)=(Ay_n,y_n-x_{n+1})+(Ay_n,x_{n+1}-z).\end{aligned}$$

Тобто,

$$(Ay_n,z-x_{n+1})\leq(Ay_n,y_n-x_{n+1}).$$

Таким чином,

$$\|x_{n+1}-z\|^2\leq\|x_n-z\|^2-\|x_n-x_{n+1}\|^2+2\lambda_n(Ay_n,y_n-x_{n+1}).$$

Далі,

$$\begin{aligned}\|x_{n+1}-z\|^2\leq\|x_n-z\|^2-\|(x_n-y_n)+(y_n-x_{n+1})\|^2+\\ +2\lambda_n(Ay_n,y_n-x_{n+1})=\|x_n-z\|^2-\|x_n-y_n\|^2-\|y_n-x_{n+1}\|^2+\\ +2(x_n-\lambda_n Ay_n-y_n,x_{n+1}-y_n).\end{aligned}$$

Оскільки $x_{n+1}\in T_n$, то

$$\begin{aligned}(x_n-\lambda_n Ay_n-y_n,x_{n+1}-y_n)=\\ =\underbrace{(x_n-\lambda_n Ax_n-y_n,x_{n+1}-y_n)}_{\leq 0}+\\ +\lambda_n(Ax_n-Ay_n,x_{n+1}-y_n)\leq\lambda_n(Ax_n-Ay_n,x_{n+1}-y_n).\end{aligned}$$

Отже, приходимо до нерівності

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda_n (Ax_n - Ay_n, x_{n+1} - y_n). \quad (3.23)$$

Доданок $2\lambda_n (Ax_n - Ay_n, x_{n+1} - y_n)$ в (3.23) оцінимо наступним чином

$$2\lambda_n (Ax_n - Ay_n, x_{n+1} - y_n) \leq 2\lambda_n \|Ax_n - Ay_n\| \|x_{n+1} - y_n\| \leq \\ \leq 2\theta \|x_n - y_n\| \|x_{n+1} - y_n\| \leq \theta^2 \|x_n - y_n\|^2 + \|x_{n+1} - y_n\|^2. \quad (3.24)$$

Враховуючи оцінку (3.24) в (3.23), приходимо до нерівності (3.22). □

Зауваження 3.6. Оцінивши доданок $2\lambda_n (Ax_n - Ay_n, x_{n+1} - y_n)$ в (3.23) інакше, отримаємо корисну нерівність ($z \in VI(A, C)$)

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \theta) \|x_n - y_n\|^2 - (1 - \theta) \|x_{n+1} - y_n\|^2. \quad (3.25)$$

Теорема 3.2. *Послідовності* (x_n) *і* (y_n)*, породженні алгоритмом 3.2, слабко збігаються до деякої точки* $z \in VI(A, C)$.

Доведення. Із нерівності (3.22) випливає, що послідовність (x_n) фейєрівська відносно множини $VI(A, C)$, тобто

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 \quad \forall n \in \mathbb{N} \ \forall z \in VI(A, C).$$

Зокрема, послідовність (x_n) обмежена.

Зафіксуємо номер $N \in \mathbb{N}$ і розглянемо нерівності (3.22) для всіх номерів $1, 2, \ldots, N$. Додавши їх, отримаємо

$$\|x_{N+1} - z\|^2 \leq \|x_1 - z\|^2 - \left(1 - \theta^2\right) \sum_{n=1}^{N} \|x_n - y_n\|^2. \quad (3.26)$$

Із нерівності (3.26) випливає збіжність числового ряду

$$\sum_n \|x_n - y_n\|^2.$$

Таким чином,

$$\lim_{n \to \infty} \|x_n - y_n\| = 0. \quad (3.27)$$

Розглянемо підпослідовність (x_{n_k}), що слабко збігається до деякої точки $z \in H$. Тоді $y_{n_k} \rightharpoonup z$ і $z \in C$. Покажемо, що $z \in VI(A, C)$.

Можливі два варіанти:

1) числова послідовність (λ_{n_k}) не прямує до нуля;

2) $\lim\limits_{k\to\infty} \lambda_{n_k} = 0$.

Спочатку розглянемо варіант 1).

Можна вважати, що $\lambda_{n_k} \geq \lambda$ для всіх достатньо великих k і деякого $\lambda > 0$. Маємо,

$$(y_{n_k} - x_{n_k} + \lambda_{n_k} A x_{n_k}, x - y_{n_k}) \geq 0 \quad \forall x \in C.$$

Звідки, використовуючи монотонність оператора A, виводимо оцінку

$$\begin{aligned} 0 \leq \frac{(y_{n_k} - x_{n_k} + \lambda_{n_k} A x_{n_k}, x - y_{n_k})}{\lambda_{n_k}} &= \\ = \frac{(y_{n_k} - x_{n_k}, x - y_{n_k})}{\lambda_{n_k}} + (A x_{n_k}, x_{n_k} - y_{n_k}) + (A x_{n_k}, x - x_{n_k}) &\leq \\ \leq \frac{(y_{n_k} - x_{n_k}, x - y_{n_k})}{\lambda_{n_k}} + (A x_{n_k}, x_{n_k} - y_{n_k}) + (A x, x - x_{n_k}) &. \end{aligned}$$

Здійснивши граничний перехід із врахуванням (3.27), отримаємо

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

Отже, $z \in VI(A, C)$.

Розглянемо варіант 2).

Нехай $\lim\limits_{k\to\infty} \lambda_{n_k} = 0$. Покладемо

$$z_{n_k} = P_C (x_{n_k} - \mu_{n_k} A x_{n_k}),$$

де $\mu_{n_k} = \lambda_{n_k} \tau^{-1} = \sigma \tau^{j(n_k)-1} > \lambda_{n_k} > 0$. Застосуємо лему 3.5. Маємо,

$$\|x_{n_k} - z_{n_k}\| \leq \frac{1}{\tau} \|x_{n_k} - y_{n_k}\| \to 0.$$

Зокрема, послідовність (z_{n_k}) обмежена і $z_{n_k} \rightharpoonup z$. Із рівномірної неперервності оператора A на обмежених множинах випливає $\|A x_{n_k} - A z_{n_k}\| \to 0$. А нерівність

$$\mu_{n_k} \|A z_{n_k} - A x_{n_k}\| > \theta \|z_{n_k} - x_{n_k}\|$$

тягне асимптотику

$$\frac{\|x_{n_k} - z_{n_k}\|}{\mu_{n_k}} \to 0. \tag{3.28}$$

Далі маємо,

$$(z_{n_k} - x_{n_k} + \mu_{n_k} A x_{n_k}, x - z_{n_k}) \geq 0 \quad \forall x \in C.$$

Звідки виводимо оцінку

$$0 \le \frac{(z_{n_k} - x_{n_k}, x - z_{n_k})}{\mu_{n_k}} + (Ax_{n_k}, x_{n_k} - z_{n_k}) + (Ax, x - x_{n_k})\,.$$

Здійснивши граничний перехід із врахуванням (3.28), отримаємо

$$(Ax, x - z) \ge 0 \quad \forall x \in C,$$

Звідки, $z \in VI\,(A, C)$.

Покажемо тепер, що $x_n \rightharpoonup z$. Тоді із (3.27) буде випливати $y_n \rightharpoonup z$. Міркуємо від супротивного. Нехай існує підпослідовність (x_{m_k}), така, що $x_{m_k} \rightharpoonup z'$ і $z \ne z'$. Зрозуміло, що $z' \in VI\,(A, C)$. Застосуємо двічі лему 3.2. Маємо,

$$\begin{aligned}\lim_{n\to\infty} \|x_n - z\| = \liminf_{k\to\infty} \|x_{n_k} - z\| < \liminf_{k\to\infty} \|x_{n_k} - z'\| = \lim_{n\to\infty} \|x_n - z'\| = \\ = \liminf_{k\to\infty} \|x_{m_k} - z'\| < \liminf_{k\to\infty} \|x_{m_k} - z\| = \lim_{n\to\infty} \|x_n - z\|\,,\end{aligned}$$

що неможливо. Таким чином, $z = z'$. □

Зауваження 3.7. Слабка границя z породженної алгоритмом 3.2 фейєрівської послідовності (x_n) має властивість [54]: $P_{VI(A,C)}x_n \to z$. Якщо множина $VI\,(A, C)$ є афінним многовидом, то $x_n \to P_{VI(A,C)}x_0$ [54].

Зауваження 3.8. Асимптотику (3.27) можна уточнити до наступної:

$$\liminf_{n\to\infty} \sqrt{n}\,\|x_n - y_n\| = 0. \tag{3.29}$$

Дійсно, якщо (3.29) не виконується, то

$$\|x_n - y_n\| \ge \mu n^{-1/2}$$

для деякого $\mu > 0$ і всіх достатньо великих номерів n. Отже числовий ряд $\sum_n \|x_n - y_n\|^2$ розбігається. Отримали протиріччя.

3.4. Модифікований екстраградієнтний алгоритм для варіаційних нерівностей та операторних рівнянь з апріорною інформацією

Розглянемо варіант методу для пошуку розв'язку варіаційної нерівності (3.1), що додатково є нерухомою точкою заданого оператора.

Нехай $S : H \to H$ — квазінерозтягуючий оператор з множиною нерухомих точок

$$F(S) = \{x \in H : \ Sx = x\}.$$

Припустимо, що оператор $I - S$ — демізамкнений в нулі. Крім того, нехай має місце:

A3) $VI(A, C) \cap F(S) \neq \emptyset$.

Зауваження 3.9. Нехай $g : H \to \mathbb{R}$ — опукла диференційовна функція. Якщо множина

$$D = \{x \in H : \ \ g(x) \leq 0\}$$

непорожня, то її можна трактувати як множину нерухомих точок квазінерозтягуючого оператора

$$Sx = \begin{cases} x - \frac{g(x)}{\|\nabla g(x)\|^2} \nabla g(x), & \text{якщо } x \notin D, \\ x, & \text{якщо } x \in D, \end{cases}$$

де $\nabla g(x) \in H$ — похідна g в точці $x \in H$ [59]. Для демізамкненості в нулі оператора $I - S$ достатньо обмеженості g на будь-якій обмеженій множині [59].

Для пошуку елементів множини $VI(A, C) \cap F(S)$ розглянемо наступний алгоритм.

Алгоритм 3.3. *Задаємо числові параметри* $\sigma > 0$, $\tau \in (0, 1)$, $\theta \in (0, 1)$, *елемент* $x_0 \in H$ *і послідовність* $(\delta_n) \subseteq [a, b] \subseteq (0, 1)$.

Ітераційний крок. *Для* $x_n \in H$ *обчислюємо*

$$y_n = P_C(x_n - \lambda_n A x_n),$$

де λ_n *отримуємо із умови*

$$\begin{cases} j(n) = \min\{j \geq 0 : & \left\| AP_C\left(x_n - \sigma\tau^j A x_n\right) - Ax_n \right\| \leq \\ & \leq \frac{\theta}{\sigma\tau^j} \left\| P_C\left(x_n - \sigma\tau^j A x_n\right) - x_n \right\| \}, \\ \lambda_n = \sigma\tau^{j(n)}. & \end{cases}$$

Обчислюємо

$$x_{n+1} = \delta_n x_n + (1 - \delta_n) S P_{T_n}(x_n - \lambda_n A y_n),$$

де

$$T_n = \{z \in H : \ (x_n - \lambda_n A x_n - y_n, z - y_n) \leq 0\}.$$

Має місце наступна теорема.

Теорема 3.3. *Послідовності* (x_n) *і* (y_n)*, що породжені алгоритмом 3.3, слабко збігаються до деякої точки* $z \in VI(A, C) \cap F(S)$*.*

Доведення. Покладемо $z_n = P_{T_n}(x_n - \lambda_n A y_n)$. Оскільки оператор S квазінерозтягуючий, то для всіх точок $z \in VI(A, C) \cap F(S)$ отримуємо

$$\begin{aligned}\|x_{n+1} - z\|^2 &\le \|\delta_n (x_n - z) + (1 - \delta_n)(Sz_n - z)\|^2 = \\ &= \delta_n \|x_n - z\|^2 + (1 - \delta_n)\|Sz_n - z\|^2 - \delta_n (1 - \delta_n)\|x_n - Sz_n\|^2 \le \\ &\le \delta_n \|x_n - z\|^2 + (1 - \delta_n)\|z_n - z\|^2 - \delta_n (1 - \delta_n)\|x_n - Sz_n\|^2 . \end{aligned} \tag{3.30}$$

Використовуючи (3.25) в (3.30) для оцінки доданка $(1 - \delta_n)\|z_n - z\|^2$, приходимо до нерівності

$$\begin{aligned}\|x_{n+1} - z\|^2 \le \|x_n - z\|^2 - (1 - \delta_n)(1 - \theta)\|x_n - y_n\|^2 - \\ - (1 - \delta_n)(1 - \theta)\|z_n - y_n\|^2 - \delta_n (1 - \delta_n)\|x_n - Sz_n\|^2 . \end{aligned} \tag{3.31}$$

Із нерівності (3.31) випливає, що послідовність (x_n) фейєрівська відносно множини $VI(A, C) \cap F(S)$, тобто

$$\|x_{n+1} - z\|^2 \le \|x_n - z\|^2 \quad \forall n \in \mathbb{N} \ \forall z \in VI(A, C) \cap F(S) .$$

Зокрема, послідовність (x_n) обмежена. Крім того, справедливі нерівності

$$\|x_n - y_n\|^2 + \|z_n - y_n\|^2 \le \frac{\|x_n - z\|^2 - \|x_{n+1} - z\|^2}{(1 - \delta_n)(1 - \theta)},$$

$$\|x_n - Sz_n\|^2 \le \frac{\|x_n - z\|^2 - \|x_{n+1} - z\|^2}{\delta_n (1 - \delta_n)}.$$

Звідки випливає

$$\lim_{n\to\infty} \|x_n - y_n\| = \lim_{n\to\infty} \|z_n - y_n\| = \lim_{n\to\infty} \|x_n - Sz_n\| = 0. \tag{3.32}$$

Розглянемо підпослідовність (x_{n_k}), що слабко збігається до деякої точки $z \in H$. Тоді (y_{n_k}), (z_{n_k}) слабко збігається до z та $z \in C$. Міркуючи як в доведенні теореми 3.2 отримуємо, що $z \in VI(A, C)$.

Залишилось показати, що $z \in F(S)$. Оскільки

$$\|z_n - Sz_n\| \le \|z_n - y_n\| + \|y_n - x_n\| + \|x_n - Sz_n\| ,$$

то із (3.32) випливає

$$\lim_{n\to\infty} \|z_n - Sz_n\| = 0.$$

Оператор $I - S$ демізамкнений в нулі. Отже із

$$z_{n_k} \rightharpoonup z \text{ та } z_{n_k} - Sz_{n_k} \to 0$$

отримуємо що $z \in F(S)$.

Аналогічно доведенню теореми 3.2 показуємо, що $x_n \rightharpoonup z$. Тоді із (3.32) буде випливати $y_n \rightharpoonup z$. □

Розглянемо тепер операторне рівняння з апріорною інформацією, яка задана у вигляді множини нерухомих точок квазінерозтягуючого оператора $T : H \to H$:

$$Ax = 0, \quad x \in F(T). \tag{3.33}$$

Подібні задачі розглядались в [59].

Алгоритм 3.3 для задачі (3.33) набуде такого вигляду.

Алгоритм 3.4. *Задаємо числові параметри* $\sigma > 0$, $\tau \in (0, 1)$, $\theta \in (0, 1)$, *елемент* $x_0 \in H$ *і послідовність* $(\delta_n) \subseteq [a, b] \subseteq (0, 1)$.

Ітераційний крок. *Для* $x_n \in H$ *обчислюємо*

$$y_n = x_n - \lambda_n A x_n,$$

де λ_n *отримуємо із умови*

$$\begin{cases} j(n) = \min\{j \geq 0 : \left\|A\left(x_n - \sigma\tau^j A x_n\right) - A x_n\right\| \leq \theta \|A x_n\|\}, \\ \lambda_n = \sigma\tau^{j(n)}. \end{cases}$$

Обчислюємо

$$z_n = x_n - \lambda_n A y_n,$$
$$x_{n+1} = \delta_n x_n + (1 - \delta_n) T z_n.$$

Частковим випадком теореми 3.3 є наступний результат.

Теорема 3.4. *Нехай оператор* $A : H \to H$ *— монотонний, рівномірно неперервний на обмежених множинах. Нехай оператор* $T : H \to H$ *— квазінерозтягуючий, причому оператор* $I - T$ *демізамкнений в нулі. Припустимо, що* $A^{-1}0 \cap F(T) \neq \emptyset$. *Тоді послідовності* (x_n), (y_n) *та* (z_n), *породжені алгоритмом 3.4, слабко збігаються до деякої точки* $z \in A^{-1}0 \cap F(T)$.

3.5. Сильно збіжний модифікований екстраградієнтний алгоритм

Розглянемо сильно збіжний модифікований екстраградієнтний метод з динамічним регулюванням величини кроку.

Відносно операторів, як і раніше, не припускаємо їх глобальної ліпшицевості. Також розглянемо методи для нерівностей і операторних рівнянь з апріорною інформацією про розв'язок, яка задана у вигляді множини нерухомих точок квазінерозтягуючого оператора.

Для регуляризації модифікованого екстраградієнтного алгоритму використовувалась проста схема Гальперна [64], яка по суті співпадає зі схемою ітеративної регуляризації [81]. Також розглянуто інші регуляризації.

Алгоритм 3.5. *Задаємо числові параметри* $\sigma > 0$, $\tau \in (0,1)$, $\theta \in (0,1)$, *елемент* $x_0 \in H$ *і послідовність* $(\alpha_n) \subseteq (0,1)$, *таку, що* $\lim\limits_{n\to\infty} \alpha_n = 0$, $\sum_{n=1}^{\infty} \alpha_n = +\infty$.

Ітераційний крок. *Для* $x_n \in H$ *обчислюємо*

$$y_n = P_C\left(x_n - \lambda_n A x_n\right),$$

де λ_n *отримуємо із умови*

$$\begin{cases} j(n) = \min\{j \geq 0 : & \left\|AP_C\left(x_n - \sigma\tau^j A x_n\right) - Ax_n\right\| \leq \\ & \leq \frac{\theta}{\sigma\tau^j}\left\|P_C\left(x_n - \sigma\tau^j A x_n\right) - x_n\right\|\}, \\ \lambda_n = \sigma\tau^{j(n)}. & \end{cases}$$

Обчислюємо

$$x_{n+1} = \alpha_n x_0 + (1-\alpha_n) P_{T_n}\left(x_n - \lambda_n A y_n\right),$$

де $T_n = \{z \in H : \ (x_n - \lambda_n A x_n - y_n, z - y_n) \leq 0\}$.

Покладемо

$$z_n = P_{T_n}\left(x_n - \lambda_n A y_n\right). \tag{3.34}$$

Має місце лема.

Лема 3.10. *Для послідовностей* (x_n), (y_n) *і* (z_n), *що породжені алгоритмом 3.5, має місце нерівність*

$$\|z_n - z\|^2 \leq \|x_n - z\|^2 - (1-\theta)\|x_n - y_n\|^2 - (1-\theta)\|z_n - y_n\|^2, \tag{3.35}$$

де $z \in VI(A, C)$.

Доведення. Повторивши міркування леми 3.9 приходимо до нерівності

$$\|z_n - z\|^2 \le \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - z_n\|^2 + \\ + 2\lambda_n (Ax_n - Ay_n, z_n - y_n). \quad (3.36)$$

Доданок $2\lambda_n (Ax_n - Ay_n, z_n - y_n)$ в (3.36) оцінимо наступним чином

$$2\lambda_n (Ax_n - Ay_n, z_n - y_n) \le 2\lambda_n \|Ax_n - Ay_n\| \|z_n - y_n\| \le \\ \le 2\theta \|x_n - y_n\| \|z_n - y_n\| \le \theta \|x_n - y_n\|^2 + \theta \|z_n - y_n\|^2. \quad (3.37)$$

Враховуючи оцінку (3.37) в (3.36), приходимо до нерівності (3.35). □

Лема 3.11. *Для породжених алгоритмом 3.5 послідовностей* (x_n), (y_n) *та* (z_n) *має місце нерівність*

$$\|x_{n+1} - z\|^2 - \|x_n - z\|^2 + (1-\theta) \|x_n - y_n\|^2 + \\ + (1-\theta) \|z_n - y_n\|^2 \le 2\alpha_n (x_0 - z, x_{n+1} - z).$$

де $z \in VI(A, C)$.

Доведення. Нехай $z \in VI(A, C)$. Застосуємо елементарну нерівність

$$\|a + b\|^2 \le \|a\|^2 + 2(b, a+b).$$

Маємо

$$\|x_{n+1} - z\|^2 = \|\alpha_n (x_0 - z) + (1-\alpha_n)(z_n - z)\|^2 \le \\ \le (1-\alpha_n)^2 \|z_n - z\|^2 + 2\alpha_n (x_0 - z, x_{n+1} - z) \le \\ \le \|z_n - z\|^2 + 2\alpha_n (x_0 - z, x_{n+1} - z). \quad (3.38)$$

Враховуючи в (3.38) нерівність (3.35) приходимо до бажаної оцінки. □

Доведемо обмеженість породжених алгоритмом послідовностей. Має місце

Лема 3.12. *Породжені алгоритмом 3.5 послідовності* (x_n), (y_n) *і* (z_n) *— обмежені.*

Доведення. Нехай $z \in VI(A, C)$. Маємо

$$\|x_{n+1} - z\| = \|\alpha_n x_0 + (1-\alpha_n) z_n - z\| = \\ = \|\alpha_n (x_0 - z) + (1-\alpha_n)(z_n - z)\| \le \\ \le \alpha_n \|x_0 - z\| + (1-\alpha_n) \|z_n - z\|.$$

Скориставшись нерівністю леми 3.10, отримаємо

$$\|x_{n+1} - z\| \leq \alpha_n \|x_0 - z\| + (1 - \alpha_n) \|x_n - z\| \leq \max\{\|x_0 - z\|, \|x_n - z\|\}.$$

Отже, $\|x_{n+1} - z\| \leq \max\{\|x_0 - z\|, \|x_1 - z\|\} \quad \forall n \in \mathbb{N}$. Таким чином, послідовність (x_n) – обмежена. Обмеженність послідовностей (y_n) і (z_n) випливає із обмеженності (x_n) і леми 3.10. □

Має місце наступна теорема.

Теорема 3.5. *Нехай множина* $C \subseteq H$ *— опукла і замкнена, оператор* $A : H \to H$ *— монотонний, рівномірно неперервний на обмежених множинах. Припустимо, що* $VI(A, C) \neq \emptyset$. *Тоді послідовності* (x_n), (y_n) *та* (z_n), *породжені алгоритмом 3.5, сильно збігаються до точки* $\bar{z} = P_{VI(A,C)} x_0$.

Доведення. Розглянемо елемент $\bar{z} = P_{VI(A,C)} x_0$. Із леми 3.12 випливає існування такого числа $M > 0$, що

$$|(x_0 - \bar{z}, x_{n+1} - \bar{z})| \leq M \quad \forall n \in \mathbb{N}.$$

Тоді із нерівності леми 3.11 отримаємо оцінку

$$\|x_{n+1} - \bar{z}\|^2 - \|x_n - \bar{z}\|^2 + (1 - \theta) \|x_n - y_n\|^2 + \\ + (1 - \theta) \|z_n - y_n\|^2 \leq 2\alpha_n M. \quad (3.39)$$

Розглянемо числову послідовність $(\|x_n - \bar{z}\|)$. Можливі два варіанти:

a) існує номер $\bar{n} \in \mathbb{N}$ такий, що $\|x_{n+1} - \bar{z}\| \leq \|x_n - \bar{z}\|$ для всіх $n \geq \bar{n}$;

b) існує зростаюча послідовність номерів (n_k) така, що $\|x_{n_k+1} - \bar{z}\| > \|x_{n_k} - \bar{z}\|$ для всіх $k \in \mathbb{N}$.

Спочатку розглянемо варіант a). В цьому випадку існує $\lim\limits_{n\to\infty} \|x_n - \bar{z}\| \in \mathbb{R}$. Оскільки $\|x_{n+1} - \bar{z}\|^2 - \|x_n - \bar{z}\|^2 \to 0$ та $\alpha_n \to 0$, то при $n \to \infty$ маємо

$$\|x_n - y_n\| \to 0, \quad (3.40)$$

$$\|z_n - y_n\| \to 0. \quad (3.41)$$

Із обмеженності (x_n) випливає існування підпослідовності (x_{n_k}), слабко збіжної до точки $w \in H$. Із (3.40) випливає, що $y_{n_k} \rightharpoonup w$. Отже, $w \in C$. Покажемо, що обов'язково $w \in VI(A, C)$.

Можливі два варіанти:

1) числова послідовність (λ_{n_k}) не прямує до нуля;

2) $\lim\limits_{k\to\infty} \lambda_{n_k} = 0$.

Спочатку розглянемо варіант 1). Можна вважати, що $\lambda_{n_k} \geq \lambda$ для всіх достатньо великих k і деякого $\lambda > 0$. Маємо,

$$(y_{n_k} - x_{n_k} + \lambda_{n_k} A x_{n_k}, x - y_{n_k}) \geq 0 \quad \forall x \in C.$$

Звідки, використовуючи монотонність оператора A, виводимо оцінку

$$\begin{aligned} 0 \leq \frac{(y_{n_k} - x_{n_k} + \lambda_{n_k} A x_{n_k}, x - y_{n_k})}{\lambda_{n_k}} &= \frac{(y_{n_k} - x_{n_k}, x - y_{n_k})}{\lambda_{n_k}} + \\ &+ (A x_{n_k}, x_{n_k} - y_{n_k}) + (A x_{n_k}, x - x_{n_k}) \leq \\ &\leq \frac{(y_{n_k} - x_{n_k}, x - y_{n_k})}{\lambda_{n_k}} + (A x_{n_k}, x_{n_k} - y_{n_k}) + (A x, x - x_{n_k}) \,. \end{aligned}$$

Здійснивши граничний перехід з урахуванням (3.40), отримаємо

$$(Ax, x - w) \geq 0 \quad \forall x \in C.$$

Отже, $w \in VI(A, C)$.

Розглянемо варіант 2). Нехай $\lim\limits_{k\to\infty} \lambda_{n_k} = 0$. Покладемо

$$w_{n_k} = P_C (x_{n_k} - \mu_{n_k} A x_{n_k}) \,,$$

де $\mu_{n_k} = \lambda_{n_k} \tau^{-1} = \sigma \tau^{j(n_k)-1} > \lambda_{n_k} > 0$. Використаємо лему 3.5. Маємо,

$$\|x_{n_k} - w_{n_k}\| \leq \frac{1}{\tau} \|x_{n_k} - y_{n_k}\| \to 0.$$

Зокрема, послідовність (w_{n_k}) обмежена і слабко збігається до w. Із рівномірної неперервності оператора A на обмежених множинах випливає

$$\|A x_{n_k} - A w_{n_k}\| \to 0.$$

А з нерівності

$$\mu_{n_k} \|A w_{n_k} - A x_{n_k}\| > \theta \|w_{n_k} - x_{n_k}\|$$

випливає асимптотика

$$\frac{\|x_{n_k} - w_{n_k}\|}{\mu_{n_k}} \to 0. \tag{3.42}$$

Далі маємо,

$$\left(w_{n_k} - x_{n_k} + \mu_{n_k} A x_{n_k}, x - w_{n_k}\right) \geq 0 \quad \forall x \in C.$$

Звідки виводимо оцінку

$$0 \leq \frac{\left(w_{n_k} - x_{n_k}, x - w_{n_k}\right)}{\mu_{n_k}} + \left(Ax_{n_k}, x_{n_k} - w_{n_k}\right) + \left(Ax, x - x_{n_k}\right).$$

Здійснивши граничний перехід з урахуванням (3.42), отримаємо

$$(Ax, x - w) \geq 0 \quad \forall x \in C,$$

звідки, $w \in VI(A, C)$.

Доведемо, що

$$\limsup_{n\to\infty} (x_0 - \bar{z}, x_{n+1} - \bar{z}) \leq 0. \tag{3.43}$$

Розглянемо таку підпослідовність (x_{n_k}), що

$$\lim_{k\to\infty} \left(x_0 - \bar{z}, x_{n_k} - \bar{z}\right) = \limsup_{n\to\infty} (x_0 - \bar{z}, x_{n+1} - \bar{z}).$$

Можна вважати, що $x_{n_k} \rightharpoonup w \in VI(A, C)$. Тоді отримаємо

$$\lim_{k\to\infty} \left(x_0 - \bar{z}, x_{n_k} - \bar{z}\right) = (x_0 - \bar{z}, w - \bar{z}) = \\ = \left(x_0 - P_{VI(A,C)}x_0, w - P_{VI(A,C)}x_0\right) \leq 0,$$

чим і доводимо (3.43).

Тепер із (3.43), оцінки

$$\|x_{n+1} - \bar{z}\|^2 = \|\alpha_n (x_0 - \bar{z}) + (1 - \alpha_n)(z_n - \bar{z})\|^2 \leq \\ \leq (1 - \alpha_n)^2 \|z_n - \bar{z}\|^2 + 2\alpha_n (x_0 - \bar{z}, x_{n+1} - \bar{z}) \leq \\ \leq (1 - \alpha_n) \|x_n - \bar{z}\|^2 + 2\alpha_n (x_0 - \bar{z}, x_{n+1} - \bar{z})$$

та леми 3.3 робимо висновок, що $\|x_n - \bar{z}\| \to 0$. Із (3.40), (3.41) отримуємо $\|y_n - \bar{z}\| \to 0$ і $\|z_n - \bar{z}\| \to 0$.

Вивчимо варіант b). В цьому випадку розглянемо послідовність номерів (m_k) із властивостями (див. лему 3.4):

(i) $m_k \nearrow +\infty$;

(ii) $\|x_{m_k+1} - \bar{z}\| \geq \|x_{m_k} - \bar{z}\|$ для всіх $k \geq n_1$;

(iii) $\|x_{m_k+1} - \bar{z}\| \geq \|x_k - \bar{z}\|$ для всіх $k \geq n_1$.

Із нерівності леми 3.11 та (ii) випливає

$$(1-\theta)\|x_{m_k} - y_{m_k}\|^2 + (1-\theta)\|z_{m_k} - y_{m_k}\|^2 \leq$$
$$\leq 2\alpha_{m_k}(x_0 - \bar{z}, x_{m_k+1} - \bar{z}) \leq 2\alpha_{m_k} M.$$

Звідки

$$\lim_{k\to\infty}\|x_{m_k} - y_{m_k}\| = \lim_{k\to\infty}\|z_{m_k} - y_{m_k}\| = 0.$$

Міркуваннями, подібними до викладених вище, показуємо, що часткові слабкі границі послідовностей (x_{m_k}) і (y_{m_k}) належать множині $VI(A, C)$. Із нерівності

$$\|x_{m_k+1} - x_{m_k}\| = \|\alpha_{m_k}x_0 + (1-\alpha_{m_k})z_{m_k} - x_{m_k}\| \leq$$
$$\leq \alpha_{m_k}\|x_0 - x_{m_k}\| + (1-\alpha_{m_k})\|z_{m_k} - x_{m_k}\| \leq$$
$$\leq \alpha_{m_k}\|x_0 - x_{m_k}\| + (1-\alpha_{m_k})(\|z_{m_k} - y_{m_k}\| + \|y_{m_k} - x_{m_k}\|)$$

випливає $\|x_{m_k+1} - x_{m_k}\| \to 0$. Із тотожності

$$(x_0 - \bar{z}, x_{m_k+1} - \bar{z}) = (x_0 - \bar{z}, x_{m_k} - \bar{z}) + (x_0 - \bar{z}, x_{m_k+1} - x_{m_k})$$

отримуємо $\limsup_{k\to\infty}(x_0 - \bar{z}, x_{m_k+1} - \bar{z}) = \limsup_{k\to\infty}(x_0 - \bar{z}, x_{m_k} - \bar{z})$. Як і раніше отримуємо

$$\limsup_{k\to\infty}(x_0 - \bar{z}, x_{m_k+1} - \bar{z}) \leq 0.$$

Далі матимемо

$$\|x_{m_k+1} - \bar{z}\|^2 \leq (1-\alpha_{m_k})\|x_{m_k} - \bar{z}\|^2 + 2\alpha_{m_k}(x_0 - \bar{z}, x_{m_k+1} - \bar{z}) \leq$$
$$\leq (1-\alpha_{m_k})\|x_{m_k+1} - \bar{z}\|^2 + 2\alpha_{m_k}(x_0 - \bar{z}, x_{m_k+1} - \bar{z}).$$

Звідки, враховуючи умову (iii), отримуємо

$$\|x_k - \bar{z}\|^2 \leq \|x_{m_k+1} - \bar{z}\|^2 \leq 2(x_0 - \bar{z}, x_{m_k+1} - \bar{z}).$$

Таким чином,

$$\limsup_{k\to\infty}\|x_k - \bar{z}\|^2 \leq 2\limsup_{k\to\infty}(x_0 - \bar{z}, x_{m_k+1} - \bar{z}) \leq 0.$$

Отже,

$$\lim_{n\to\infty}\|x_n - \bar{z}\| = 0$$

і, в свою чергу, $\lim\limits_{n\to\infty}\|y_n-\bar{z}\|=\lim\limits_{n\to\infty}\|z_n-\bar{z}\|=0.$ □

Розглянемо ще один варіант сильно збіжного модифікованого екстраградієнтного алгоритму для розв'язання варіаційної нерівності (3.1).

Алгоритм 3.6. *Задаємо числові параметри* $\sigma>0$, $\tau\in(0,1)$, $\theta\in(0,1)$ *та елемент* $x_0\in H$.

Ітераційний крок. *Для* $x_n\in H$ *обчислюємо*

$$y_n=P_C\left(x_n-\lambda_n A x_n\right),$$

де λ_n *отримуємо із умови*

$$\begin{cases} j\left(n\right)=\min\{j\geq 0: & \left\|AP_C\left(x_n-\sigma\tau^j Ax_n\right)-Ax_n\right\|\leq \\ & \leq \frac{\theta}{\sigma\tau^j}\left\|P_C\left(x_n-\sigma\tau^j Ax_n\right)-x_n\right\|\}, \\ \lambda_n=\sigma\tau^{j(n)}. & \end{cases}$$

Обчислюємо

$$z_n=P_{T_n}\left(x_n-\lambda_n A y_n\right),$$

де

$$T_n=\{z\in H:\ (x_n-\lambda_n Ax_n-y_n, z-y_n)\leq 0\}.$$

Обчислюємо

$$x_{n+1}=P_{C_n\cap Q_n}x_0,$$

де

$$C_n=\{z\in H:\ \|z_n-z\|\leq\|x_n-z\|\}$$

та

$$Q_n=\{z\in H:\ (x_n-z, x_0-x_n)\geq 0\}.$$

Множини Q_n, C_n — замкнені півпростори. Проектування точки на їх перетин $C_n\cap Q_n$ є елементарною задачею, яка має аналітичний розв'язок [54].

Зауваження 3.10. Алгоритм 3.6 утворений введенням процедури динамічного регулювання величини кроку та проектування на опорний для множини C півпростір в схему гібридного методу [82], який в свою чергу є регуляризацією методу Корпелевич за допомогою проекційної CQ-схеми апроксимації нерухомих точок нерозтягуючих операторів, що запропонована японськими математиками K. Nakajo та W. Takahashi [72].

Має місце наступна теорема.

Теорема 3.6. *Нехай множина* $C \subseteq H$ *— опукла і замкнена, оператор* $A : H \to H$ *— монотонний, рівномірно неперервний на обмежених множинах. Припустимо, що* $VI(A, C) \neq \emptyset$. *Тоді послідовності* (x_n), (y_n) *та* (z_n), *породжені алгоритмом 3.6, сильно збігаються до точки* $\bar{z} = P_{VI(A,C)} x_0$.

Доведення. Нехай $z \in VI(A, C)$. Для послідовностей (x_n), (y_n) і (z_n), які породжені алгоритмом 3.6, має місце нерівність

$$\|z_n - z\|^2 \leq \|x_n - z\|^2 - (1-\theta)\|x_n - y_n\|^2 - (1-\theta)\|z_n - y_n\|^2. \tag{3.44}$$

З нерівності (3.44) випливає $z \in C_n$ для всіх $n \in \mathbb{N}$. Отже, $VI(A, C) \subseteq C_n$ для всіх $n \in \mathbb{N} \cup \{0\}$.

Тепер за допомогою індукції покажемо, що для кожного $n \in \mathbb{N} \cup \{0\}$ 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 $(x_{k+1} - z, x_0 - x_{k+1}) \geq 0$ для всіх $z \in C_k \cap Q_k$. 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 $\bar{z} \in VI(A, C)$, така, що $\bar{z} = P_{VI(A,C)} x_0$. 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 $\bar{z} \in VI(A, C) \subseteq C_n \cap Q_n$, то

$$\|x_{n+1} - x_0\| \leq \|\bar{z} - x_0\|. \tag{3.45}$$

Звідки випливає обмеженість (x_n).

Доведемо, що

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0. \tag{3.46}$$

З $x_{n+1} \in C_n \cap Q_n \subseteq Q_n$ та $x_n = P_{Q_n} x_0$, випливає $\|x_{n+1} - x_0\| \geq \|x_n - x_0\|$ для всіх $n \in \mathbb{N}$. Числова 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 $(x_n - x_{n+1}, x_0 - x_n) \geq 0$ та

$$\begin{aligned}\|x_n - x_{n+1}\|^2 &= \|(x_n - x_0) - (x_{n+1} - x_0)\|^2 = \|x_n - x_0\|^2 - \\ &- 2(x_n - x_0, x_{n+1} - x_0) + \|x_{n+1} - x_0\|^2 = \|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2 - \\ &- 2(x_n - x_{n+1}, x_0 - x_n) \leq \|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2.\end{aligned}$$

Звідки випливає (3.46).

Оскільки $x_{n+1} \in C_n$, то $\|z_n - x_{n+1}\| \leq \|x_n - x_{n+1}\|$. Звідки

$$\|z_n - x_n\| \leq \|z_n - x_{n+1}\| + \|x_{n+1} - x_n\| \leq 2\|x_n - x_{n+1}\| \to 0. \tag{3.47}$$

Використовуючи нерівність (3.44) отримуємо

$$\begin{aligned}\|x_n - y_n\|^2 \leq \frac{\|x_n - z\|^2 - \|z_n - z\|^2}{(1-\theta)} = \\ = \frac{(\|x_n - z\| - \|z_n - z\|)(\|x_n - z\| + \|z_n - z\|)}{(1-\theta)} \leq \\ \leq \frac{(\|x_n - z\| + \|z_n - z\|)}{(1-\theta)} \|x_n - z_n\| = O(\|x_n - z_n\|),\end{aligned}$$

$$\begin{aligned}\|z_n - y_n\|^2 \leq \frac{\|x_n - z\|^2 - \|z_n - z\|^2}{(1-\theta)} = \\ = \frac{(\|x_n - z\| - \|z_n - z\|)(\|x_n - z\| + \|z_n - z\|)}{(1-\theta)} \leq \\ \leq \frac{(\|x_n - z\| + \|z_n - z\|)}{(1-\theta)} \|x_n - z_n\| = O(\|x_n - z_n\|),\end{aligned}$$

де $z \in VI(A, C)$. З (3.47) випливає

$$\|x_n - y_n\| \to 0, \tag{3.48}$$

$$\|z_n - y_n\| \to 0. \tag{3.49}$$

Із обмеженності (x_n) випливає існування підпослідовності (x_{n_k}), слабко збіжної до точки $w \in H$. Із (3.48) випливає, що $y_{n_k} \rightharpoonup w$. Отже, $w \in C$. Аналогічно міркуванням теореми 3.5, можна показати, що $w \in VI(A, C)$.

Для $\bar{z} = P_{VI(A,C)}x_0$ з нерівності (3.45) випливає

$$\|x_0 - \bar{z}\| \leq \|x_0 - w\| \leq \liminf_{k\to\infty} \|x_0 - x_{n_k}\| \leq \limsup_{k\to\infty} \|x_0 - x_{n_k}\| \leq \|x_0 - \bar{z}\|.$$

Тобто, ми отримали $\lim_{k\to\infty} \|x_0 - x_{n_k}\| = \|x_0 - w\| = \|x_0 - \bar{z}\|$. Звідки, $x_{n_k} \to w = \bar{z}$. Отже, $x_n \to \bar{z}$. З (3.48) і (3.49) випливає $y_n \to \bar{z}$ і $z_n \to \bar{z}$, що і треба було довести. □

З допомогою методу Takahashi–Takeuchi–Kubota [73] можна побудувати такий сильно збіжний алгоритм для розв'язання варіаційної нерівності (3.1).

Алгоритм 3.7. *Задаємо числові параметри* $\sigma > 0$, $\tau \in (0,1)$, $\theta \in (0,1)$ *та елемент* $x_0 \in H$.

Ітераційний крок. *Для* $x_n \in H$ *обчислюємо*

$$y_n = P_C\left(x_n - \lambda_n A x_n\right),$$

де λ_n *отримуємо із умови*

$$\begin{cases} j\left(n\right) = \min\{j \geq 0 : & \left\|AP_C\left(x_n - \sigma\tau^j A x_n\right) - Ax_n\right\| \leq \\ & \leq \frac{\theta}{\sigma\tau^j}\left\|P_C\left(x_n - \sigma\tau^j A x_n\right) - x_n\right\|\}, \\ \lambda_n = \sigma\tau^{j(n)}. & \end{cases}$$

Обчислюємо

$$z_n = P_{T_n}\left(x_n - \lambda_n A y_n\right),$$

де

$$T_n = \{z \in H : \ (x_n - \lambda_n A x_n - y_n, z - y_n) \leq 0\}.$$

Обчислюємо

$$x_{n+1} = P_{C_{n+1}} x_0,$$

де

$$C_{n+1} = \{z \in C_n : \ \|z_n - z\| \leq \|x_n - z\|\}, \quad C_0 = H.$$

Має місце

Теорема 3.7. *Нехай виконані умови A1) та A2). Тоді послідовності* (x_n), (y_n) *і* (z_n), *породжені алгоритмом 3.7, сильно збігаються до* $P_{VI(A,C)}x_0$.

Доведення. Для послідовностей (x_n), (y_n) і (z_n), які породжені алгоритмом 3.7, має місце нерівність (3.44). Покажемо, що алгоритм 3.7 породжує ланцюжок вкладень

$$H = C_0 \supseteq C_1 \supseteq \ldots \supseteq C_n \supseteq \ldots \supseteq VI(A, C).$$

Ясно, що

$$VI\left(A, C\right) \subseteq C_0 = H.$$

Припустимо, що $VI\left(A, C\right) \subseteq C_n$. Нехай $z \in VI\left(A, C\right)$. З (3.44) маємо

$$\|z_n - z\| \leq \|x_n - z\|.$$

Отже, $z \in C_{n+1}$. Тому $VI\left(A, C\right) \subseteq C_{n+1} \subseteq C_n$ для всіх $n \in \mathbb{N}$.

Покажемо, що існує $\lim_{n\to\infty}\|x_n - x_0\| \in \mathbb{R}$. Оскільки

$$x_n = P_{C_n}x_0, \quad VI(A,C) \subseteq C_n,$$

то для всіх $z \in VI(A,C)$ має місце нерівність

$$\|x_n - z\|^2 \le \|x_0 - z\|^2 - \|x_n - x_0\|^2 .$$

Звідки

$$\|x_n - x_0\|^2 \le \|x_0 - z\|^2 - \|x_n - z\|^2 \le \|x_0 - z\|^2 .$$

Звідки випливає обмеженість зверху послідовності $(\|x_n - x_0\|)$. Оскільки $C_{n+1} \subseteq C_n$, то

$$\|x_{n+1} - x_0\| \ge \|x_n - x_0\| .$$

Отже, $(\|x_n - x_0\|)$ — обмежена зверху та неспадна послідовність, тому існує скінченна границя $\lim_{n\to\infty}\|x_n - x_0\|$.

Покажемо фундаментальність послідовності (x_n). Для довільного $m \in \mathbb{N}$ маємо (враховуємо $C_{n+m} \subseteq C_n$)

$$\|x_{n+m} - x_n\|^2 = \|x_{n+m} - P_{C_n}x_0\|^2 \le \|x_0 - x_{n+m}\|^2 - \|x_n - x_0\|^2 .$$

Звідки випливає фундаментальність послідовності (x_n). Таким чином,

$$x_n \to \bar{z} \in H \ \text{ при } \ n \to \infty.$$

Покажемо, що $\bar{z} \in VI(A,C)$. Оскільки $x_{n+1} \in C_{n+1}$, то

$$\|z_n - x_{n+1}\| \le \|x_n - x_{n+1}\| .$$

Звідки

$$\|z_n - x_n\| \le \|z_n - x_{n+1}\| + \|x_n - x_{n+1}\| \le 2\|x_n - x_{n+1}\| \to 0.$$

Використовуючи нерівність (3.44), для всіх $z \in VI(A,C)$ одержимо асимптотичні співвідношення

$$\|x_n - y_n\|^2 = O(\|x_n - z_n\|),$$

$$\|z_n - y_n\|^2 = O(\|x_n - z_n\|),$$

Звідки

$$\lim_{n\to\infty}\|x_n - y_n\| = \lim_{n\to\infty}\|y_n - z_n\| = 0.$$

З одержаних асимптотик випливає, що $y_n \to \bar{z}$ та $z_n \to \bar{z}$ при $n \to \infty$. Отже, $\bar{z} \in C$. Аналогічно міркуванням теореми 3.5, можна показати, що

$$\bar{z} \in VI(A, C) .$$

Оскільки $x_n = P_{C_n} x_0$ і $VI(A, C) \subseteq C_n$, то

$$(x_n - x_0, z - x_n) \geq 0 \quad \forall z \in VI(A, C) .$$

Здійснивши граничний перехід в цій нерівності, одержимо

$$(\bar{z} - x_0, z - \bar{z}) \geq 0 \quad \forall z \in VI(A, C) ,$$

тобто $\bar{z} = P_{VI(A,C)} x_0$. Як видно з наведених міркувань послідовності (y_n) та (z_n) сильно збігаються до $\bar{z} = P_{VI(A,C)} x_0$. □

Розглянемо квазінерозтягуючий оператор $S : H \to H$ з множиною нерухомих точок $F(S)$ та задачу

$$\text{знайти } x \in VI(A, C) \cap F(S) . \tag{3.50}$$

Припустимо, що оператор $I - S$ – демізамкнений в нулі та виконується умова A3).

Для пошуку розв'язків задачі (3.50) розглянемо такий алгоритм.

Алгоритм 3.8. *Задаємо числові параметри* $\sigma > 0$, $\tau \in (0, 1)$, $\theta \in (0, 1)$, *елемент* $x_0 \in H$, *послідовність* $(\delta_n) \subseteq [a, b] \subseteq (0, 1)$ *і послідовність* $(\alpha_n) \subseteq (0, 1)$, *таку, що* $\lim\limits_{n\to\infty} \alpha_n = 0$, $\sum_{n=1}^{\infty} \alpha_n = +\infty$.

Ітераційний крок. *Для* $x_n \in H$ *обчислюємо*

$$y_n = P_C (x_n - \lambda_n A x_n) ,$$

де λ_n *отримуємо із умови*

$$\begin{cases} j(n) = \min\{j \geq 0 : & \left\| A P_C \left(x_n - \sigma\tau^j A x_n \right) - A x_n \right\| \leq \\ & \leq \frac{\theta}{\sigma\tau^j} \left\| P_C \left(x_n - \sigma\tau^j A x_n \right) - x_n \right\| \}, \\ \lambda_n = \sigma\tau^{j(n)}. \end{cases}$$

Обчислюємо

$$x_{n+1} = \alpha_n x_0 + (1 - \alpha_n) \left(\delta_n x_n + (1 - \delta_n) S P_{T_n} (x_n - \lambda_n A y_n) \right) ,$$

де $T_n = \{z \in H : (x_n - \lambda_n A x_n - y_n, z - y_n) \leq 0\}$.

При аналізі алгоритму 3.8 будемо продовжувати використовувати позначення (3.34).

Лема 3.13. *Для породжених алгоритмом 3.8 послідовностей* (x_n), (y_n) *та* (z_n) *має місце нерівність*

$$\begin{aligned}\|x_{n+1}-z\|^2-\|x_n-z\|^2+(1-\delta_n)(1-\theta)\|x_n-y_n\|^2+\\ +(1-\delta_n)(1-\theta)\|z_n-y_n\|^2+\delta_n(1-\delta_n)\|x_n-Sz_n\|^2\le\\ \le 2\alpha_n(x_0-z,x_{n+1}-z)\,.\end{aligned}$$

де $z\in VI(A,C)\cap F(S)$.

Доведення. Нехай $z\in VI(A,C)\cap F(S)$. Маємо

$$\begin{aligned}\|x_{n+1}-z\|^2&=\|\alpha_n(x_0-z)+(1-\alpha_n)(\delta_n x_n+(1-\delta_n)Sz_n-z)\|^2\le\\ &\le(1-\alpha_n)^2\|\delta_n x_n+(1-\delta_n)Sz_n-z\|^2+2\alpha_n(x_0-z,x_{n+1}-z)\le\\ &\le\|\delta_n(x_n-z)+(1-\delta_n)(Sz_n-z)\|^2+2\alpha_n(x_0-z,x_{n+1}-z)=\\ &=\delta_n\|x_n-z\|^2+(1-\delta_n)\|Sz_n-z\|^2-\delta_n(1-\delta_n)\|x_n-Sz_n\|^2+\\ &\quad+2\alpha_n(x_0-z,x_{n+1}-z)\le\\ &\le\delta_n\|x_n-z\|^2+(1-\delta_n)\|z_n-z\|^2-\delta_n(1-\delta_n)\|x_n-Sz_n\|^2+\\ &\quad+2\alpha_n(x_0-z,x_{n+1}-z)\,.\end{aligned}\tag{3.51}$$

Використовуючи лему 3.10 для оцінки доданка $(1-\delta_n)\|z_n-z\|^2$ в (3.51), приходимо до нерівності

$$\begin{aligned}\|x_{n+1}-z\|^2\le\|x_n-z\|^2-(1-\delta_n)(1-\theta)\|x_n-y_n\|^2-\\ -(1-\delta_n)(1-\theta)\|z_n-y_n\|^2-\delta_n(1-\delta_n)\|x_n-Sz_n\|^2+\\ +2\alpha_n(x_0-z,x_{n+1}-z)\,.\end{aligned}$$

що і потрібно було довести. □

Лема 3.14. *Послідовності* (x_n), (y_n) *та* (z_n) *— обмежені.*

Доведення. Нехай $z\in VI(A,C)\cap F(S)$. Маємо

$$\begin{aligned}&\|x_{n+1}-z\|\le\\ &\le\alpha_n\|x_0-z\|+(1-\alpha_n)\delta_n\|x_n-z\|+(1-\alpha_n)(1-\delta_n)\|Sz_n-z\|\le\\ &\le\alpha_n\|x_0-z\|+(1-\alpha_n)\delta_n\|x_n-z\|+(1-\alpha_n)(1-\delta_n)\|z_n-z\|\,.\end{aligned}$$

Скориставшись нерівністю (3.35), отримаємо

$$\|x_{n+1}-z\| \le \alpha_n \|x_0-z\| + (1-\alpha_n)\|x_n-z\| \le \max\{\|x_0-z\|\,,\ \|x_n-z\|\}.$$

Отже,

$$\|x_{n+1}-z\| \le \max\{\|x_0-z\|\,,\ \|x_1-z\|\}$$

для всіх $n \in \mathbb{N}$.

Таким чином, послідовність (x_n) — обмежена. Обмеженність послідовностей (y_n) і (z_n) випливає із обмеженності (x_n) і нерівності (3.35). □

Має місце наступна теорема.

Теорема 3.8. *Нехай множина $C \subseteq H$ — опукла і замкнена, оператор $A : H \to H$ — монотоний, рівномірно неперервний на обмежених множинах. Нехай оператор $S : H \to H$ — квазінерозтягуючий, причому оператор $I-S$ демізамкнений в нулі. Припустимо, що $VI(A,C) \cap F(S) \neq \emptyset$. Тоді послідовності (x_n), (y_n) та (z_n), породжені алгоритмом 3.8, сильно збігаються до точки $\bar{z} = P_{VI(A,C)\cap F(S)} x_0$.*

Доведення. Розглянемо елемент $\bar{z} = P_{VI(A,C)\cap F(S)} x_0$. Із леми 3.14 випливає існування такого числа $M > 0$, що

$$|(x_0-\bar{z}, x_{n+1}-\bar{z})| \le M \quad \forall n \in \mathbb{N}.$$

Тоді із нерівності леми 3.13 отримаємо оцінку

$$\|x_{n+1}-\bar{z}\|^2 - \|x_n-\bar{z}\|^2 + (1-\delta_n)(1-\theta)\|x_n-y_n\|^2 + \\ + (1-\delta_n)(1-\theta)\|z_n-y_n\|^2 + \delta_n(1-\delta_n)\|x_n-Sz_n\|^2 \le 2\alpha_n M. \quad (3.52)$$

Розглянемо числову послідовність $(\|x_n-\bar{z}\|)$. Можливі два варіанти:

a) існує номер $\bar{n} \in \mathbb{N}$ такий, що $\|x_{n+1}-\bar{z}\| \le \|x_n-\bar{z}\|$ для всіх $n \ge \bar{n}$;

b) існує зростаюча послідовність номерів (n_k) така, що

$$\|x_{n_k+1}-\bar{z}\| > \|x_{n_k}-\bar{z}\|$$

для всіх $k \in \mathbb{N}$.

Спочатку розглянемо варіант a). В цьому випадку існує границя $\lim\limits_{n\to\infty} \|x_n-\bar{z}\| \in \mathbb{R}$. Оскільки

$$\|x_{n+1}-\bar{z}\|^2 - \|x_n-\bar{z}\|^2 \to 0$$

і $\alpha_n \to 0$, то при $n \to \infty$ маємо

$$\|x_n - y_n\| \to 0, \tag{3.53}$$

$$\|z_n - y_n\| \to 0, \tag{3.54}$$

$$\|x_n - Sz_n\| \to 0. \tag{3.55}$$

Із обмеженості послідовності (x_n) випливає існування підпослідовності (x_{n_k}), яка слабко збігається до $w \in H$. Із (3.53) випливає, що (y_{n_k}) слабко збігається до $w \in H$.

Отже, $w \in C$. Міркуючи як в доведенні теореми 3.5 отримуємо, що

$$w \in VI\,(A, C)\,.$$

Залишилось показати, що $w \in F(S)$. Оскільки

$$\|z_n - Sz_n\| \leq \|z_n - y_n\| + \|y_n - x_n\| + \|x_n - Sz_n\|\,,$$

то із (3.53), (3.54) та (3.55) випливає $\lim\limits_{n\to\infty} \|z_n - Sz_n\| = 0$.

Оператор $I - S$ демізамкнений в нулі. Отже з $z_{n_k} \rightharpoonup w$ та

$$\lim_{k\to\infty} \|z_{n_k} - Sz_{n_k}\| = 0$$

отримуємо, що $w \in F(S)$.

Як і при доведенні теореми 3.5 отримуємо

$$\limsup_{n\to\infty}\,(x_0 - \bar{z}, x_{n+1} - \bar{z}) \leq 0. \tag{3.56}$$

Із (3.56), нерівності

$$\begin{aligned}
\|x_{n+1} - \bar{z}\|^2 \leq (1-\alpha_n)^2 \,\|\delta_n x_n + (1-\delta_n)\,Sz_n - \bar{z}\|^2 + \\
+ 2\alpha_n\,(x_0 - \bar{z}, x_{n+1} - \bar{z}) \leq \\
\leq (1-\alpha_n)\,\delta_n\,\|x_n - \bar{z}\|^2 + (1-\alpha_n)\,(1-\delta_n)\,\|Sz_n - \bar{z}\|^2 + \\
+ 2\alpha_n\,(x_0 - \bar{z}, x_{n+1} - \bar{z}) \leq \\
\leq (1-\alpha_n)\,\delta_n\,\|x_n - \bar{z}\|^2 + (1-\alpha_n)\,(1-\delta_n)\,\|z_n - \bar{z}\|^2 + \\
+ 2\alpha_n\,(x_0 - \bar{z}, x_{n+1} - \bar{z}) \leq (1-\alpha_n)\,\|x_n - \bar{z}\|^2 + 2\alpha_n\,(x_0 - \bar{z}, x_{n+1} - \bar{z})
\end{aligned}$$

і леми 3.3 робимо висновок, що $\|x_n - \bar{z}\| \to 0$. Із (3.53), (3.54) отримуємо

$$\|y_n - \bar{z}\| \to 0, \quad \|z_n - \bar{z}\| \to 0.$$

Вивчимо варіант b). В цьому випадку розглянемо послідовність номерів (m_k) , що володіють властивостями:

(i) $m_k \nearrow +\infty$;

(ii) $\|x_{m_k+1} - \bar{z}\| \ge \|x_{m_k} - \bar{z}\|$ для всіх $k \ge n_1$;

(iii) $\|x_{m_k+1} - \bar{z}\| \ge \|x_k - \bar{z}\|$ для всіх $k \ge n_1$.

Із (ii) випливає

$$(1 - \delta_{m_k})(1 - \theta)\|x_{m_k} - y_{m_k}\|^2 + (1 - \delta_{m_k})(1 - \theta)\|z_{m_k} - y_{m_k}\|^2 + \\ + \delta_{m_k}(1 - \delta_{m_k})\|x_{m_k} - Sz_{m_k}\|^2 \le 2\alpha_{m_k}(x_0 - \bar{z}, x_{m_k+1} - \bar{z}) \le 2\alpha_{m_k} M.$$

Звідки

$$\lim_{k\to\infty}\|x_{m_k} - y_{m_k}\| = \lim_{k\to\infty}\|z_{m_k} - y_{m_k}\| = \lim_{k\to\infty}\|x_{m_k} - Sz_{m_k}\| = 0.$$

Міркуваннями подібними до викладених вище, показуємо, що часткові слабкі границі послідовностей (x_{m_k}), (y_{m_k}) і (z_{m_k}) належать множині $VI(A, C) \cap F(S)$. Із нерівності

$$\|x_{m_k+1} - x_{m_k}\| = \|\alpha_{m_k}(x_0 - x_{m_k}) + (1 - \alpha_{m_k})(1 - \delta_{m_k})(Sz_{m_k} - x_{m_k})\| \le \\ \le \alpha_{m_k}\|x_0 - x_{m_k}\| + (1 - \alpha_{m_k})(1 - \delta_{m_k})\|Sz_{m_k} - x_{m_k}\|.$$

випливає $\|x_{m_k+1} - x_{m_k}\| \to 0$. Далі повторюємо доведення теореми 3.5. □

Розглянемо тепер операторне рівняння з апріорною інформацією, що задана у вигляді множини нерухомих точок оператора $T : H \to H$:

$$Ax = 0, \quad x \in F(T). \tag{3.57}$$

Алгоритм 3.8 для задачі (3.57) приймає наступний вигляд.

Алгоритм 3.9. *Задаємо числові параметри* $\sigma > 0$, $\tau \in (0, 1)$, $\theta \in (0, 1)$, *елемент* $x_0 \in H$, *послідовність* $(\delta_n) \subseteq [a, b] \subseteq (0, 1)$ *і послідовність* $(\alpha_n) \subseteq (0, 1)$, *таку, що* $\lim_{n\to\infty}\alpha_n = 0$, $\sum_{n=1}^{\infty}\alpha_n = +\infty$.

Ітераційний крок. *Для* $x_n \in H$ *обчислюємо*

$$y_n = x_n - \lambda_n A x_n,$$

де λ_n *отримуємо із умови*

$$\begin{cases} j(n) = \min\{j \geq 0 : \left\|A\left(x_n - \sigma\tau^j A x_n\right) - Ax_n\right\| \leq \theta \|Ax_n\|\}, \\ \lambda_n = \sigma\tau^{j(n)}. \end{cases}$$

Обчислюємо

$$x_{n+1} = \alpha_n x_0 + (1-\alpha_n)\left(\delta_n x_n + (1-\delta_n) T(x_n - \lambda_n A y_n)\right).$$

Частковим випадком теореми 3.8 є наступний результат.

Теорема 3.9. *Нехай оператор* $A : H \to H$ *— монотонний, рівномірно неперервний на обмежених множинах. Нехай оператор* $T : H \to H$ *— квазінерозтягуючий, причому оператор* $I - T$ *демізамкнений в нулі. Припустимо, що* $A^{-1}0 \cap F(T) \neq \emptyset$*. Тоді послідовності* (x_n), (y_n) *і* (z_n)*, породжені алгоритмом 3.9, сильно збігаються до точки* $\bar{z} = P_{A^{-1}0 \cap F(T)} x_0$.

3.6. Застосування варіаційних нерівностей у машинному навчанні

Розглянемо декілька задач, що зв'язані з машинним навчанням та ведуть до варіаційних нерівностей.

Сідлові формулювання. Задача мінімізації функції максимуму

$$\max_{i=1,\ldots,m} f_i(x) \to \min_{x \in C},$$

де $C \subseteq \mathbb{R}^n$, може бути розглянута у вигляді сідлової

$$\min_{x \in C} \max_{y \in \Delta^m} \left(\sum_{i=1}^{m} y_i f_i(x) \right), \tag{3.58}$$

де $\Delta^m = \{y \in \mathbb{R}^m : \ y_i \geq 0, \ \sum_{i=1}^m y_i = 1\}$ [83]. До (3.58) задачі можна застосовувати алгоритми екстраградієнтного типу.

Деякі задачі аналізу даних та машинного навчання можна сформулювати у вигляді

$$F(Kx) + G(x) \to \min_{x \in \mathbb{R}^n}, \tag{3.59}$$

де $F : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$, $G : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — власні замкнені та опуклі функції, $K : \mathbb{R}^n \to \mathbb{R}^m$ — лінійний оператор [84].

Задачу (3.59) та її двоїсту

$$-G^*(-K^*y) - F^*(y) \to \max_{y\in\mathbb{R}^m}$$

можна сформулювати у вигляді сідлової задачі (використали відоме прямо-двоїсте формулювання)

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m}\left(\langle Kx, y\rangle + G(x) - F^*(y)\right). \tag{3.60}$$

Тут $\varphi^*(y) = \sup_{x\in\mathbb{R}^n}\left(\langle y, x\rangle - \varphi(x)\right)$ — спряжена функція до $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Деталі можна знайти в [84, 85].

Припустимо, що задача (3.61) має розв'язок $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$. Тоді він задовольняє операторне включення (що рівносильне варіаційній нерівності)

$$\begin{cases} 0 \in K^*\bar{y} + \partial G(\bar{x}), \\ 0 \in -K\bar{x} + \partial F^*(\bar{y}), \end{cases} \tag{3.61}$$

де ∂G та ∂F^* — субдиференціали функцій G та F^* [83–85]. А до (3.61) підходимо з алгоритмами для варіаційних нерівностей.

У вигляді (3.59) формулюються ROF та TV-L1 моделі в задачах відновлення зашумлених зображень [86, 87].

Наведемо дві задачі вигляду (3.59).

1. ℓ_2-регуляризована ℓ_2-регресія. Розглянемо задачу

$$\frac{1}{2}\|Kx - b\|_2^2 + \frac{\lambda}{2}\|x\|_2^2 \to \min_{x\in\mathbb{R}^n},$$

де $b \in \mathbb{R}^m$, $K : \mathbb{R}^n \to \mathbb{R}^m$ — лінійний оператор, $\lambda > 0$. Сідлове формулювання має вигляд

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m}\left(\langle Kx, y\rangle + \frac{\lambda}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle b, y\rangle\right).$$

Отримуємо таку умову оптимальності

$$\begin{cases} 0 = K^*\bar{y} + \lambda\bar{x}, \\ 0 = -K\bar{x} + \bar{y} + b. \end{cases}$$

2. ℓ_1-регресія з умовою невід'ємності. Розглянемо задачу

$$\|Kx - b\|_1 \to \min_{x\in\mathbb{R}^n}, \quad x \geq 0,$$

де $b \in \mathbb{R}^m$, $K : \mathbb{R}^n \to \mathbb{R}^m$ — лінійний оператор. Сідлове формулювання має вигляд

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} \left(\langle Kx, y \rangle + \delta_{\mathbb{R}^n_+}(x) - \delta_{B_\infty(\mathbb{R}^m)}(y) - \langle b, y \rangle \right),$$

де $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : \ x_i \geq 0\}$, $B_\infty(\mathbb{R}^m) = \{y \in \mathbb{R}^m : \ \|y\|_\infty = \max_i |y_i| \leq 1\}$, δ_M — індикаторна функція множини M.

Отримуємо таку умову оптимальності

$$\begin{cases} 0 \in K^* \bar{y} + N_{\mathbb{R}^n_+} \bar{x}, \\ 0 \in -K\bar{x} + b + N_{B_\infty(\mathbb{R}^m)} \bar{y}, \end{cases}$$

де $N_C z$ — нормальний конус множини C в точці z [83].

Робастне навчання. Мінімаксні задачі мають довгу історію успішного застосування у робастній оптимізації [88]. Вони також знайшли застосування в навчанні глибоких нейронних мереж, які виявились вразливими до суперечливих прикладів — вхідних даних, які майже неможливо відрізнити від навчальних даних і все ж неправильно класифікованих мережею.

Робастне навчання [89] має на меті подолання цих проблем шляхом розв'язання мінімаксної задачі

$$\min_x \mathbb{E}_{(u,v)} \left(\max_{y \in S} L\left(u + y, v, x\right) \right),$$

де u — вектор ознак об'єкту, v — мітка класу цього об'єкту, x — вектор параметрів моделі, що тренується, y — спотворення (шум) для змагальної модифікації об'єкту, S — множина допустимих спотворень.

Породжуючі змагальні мережі. Розглянемо ігрову задачу для побудови породжуючої змагальної мережі (Generative Adversarial Network (GAN)).

Така нейронна мережа є комбінацією двох мереж, одна з яких — генератор G — генерує зразки, а друга — дискримінатор D — відрізняє справжні зразки з деякої бази даних від штучно згенерованих генератором G.

Оскільки нейронні мережі G та D мають протилежні задачі — генерувати зразки, що не відрізняються від справжніх, та відсіювати штучні зразки, то між ними виникає антагоністична гра.

Породжуючі змагальні мережі були вперше запропоновані в роботі [90].

Наведемо математичний опис цієї гри.

Нехай задано ймовірносний розподіл p_{data} на множині всіх зразків. Зразки зображаються векторами простору $\mathbb{R}^d$.

Нехай X та Y — опуклі підмножини просторів $\mathbb{R}^k$ та $\mathbb{R}^m$, а відображення G_x — генератор, де $x \in X$ — вектор його параметрів.

Розподіл p_x, що моделюється на множині усіх зразків, породжується генератором з рівномірного (або нормального) розподілу μ:

$$p_x(A) = \mu\left(G_x^{-1}(A)\right).$$

Нехай функція D_y — дискримінатор, вона визначена на множині зразків та примає значення в сегменті $[0,1]$, $y \in Y$ — вектор її параметрів.

Генератор та дискримінатор реалізуються у вигляді нейронних мереж певної архітектури.

Значення $D_y(u)$ розуміється як ймовірність того, що зразок згенерований розподілом p_{data}.

Навчаємо нейронну мережу D_y так, щоб максимізувати ймовірність привильної класифікації зразків. Одночасно навчаємо нейронну мережу G_x так, щоб мінімізувати середнє значення величини

$$\log\left(1 - D_y\left(G_x(z)\right)\right)$$

при $z \sim \mu$.

Між генератором та дискримінатором виникає антагоністична гра з деякою платіжною функцією

$$L : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}.$$

В роботі [90] запропоновано таку платіжну функцію

$$L(x,y) = \mathbb{E}_{u \sim p_{data}}\left(\log D_y(u)\right) + \mathbb{E}_{z\sim\mu}\left(\log\left(1 - D_y\left(G_x(z)\right)\right)\right). \tag{3.62}$$

Зауваження 3.11. Існують інші способи обрати цільову функцію для навчання [91].

Для побудови оптимальних мереж розв'язується задача пошуку сідлової точки функції (3.62) на $X \times Y$:

$$\min_{x\in X}\max_{y\in Y} L(x,y) =$$
$$= \min_{x\in X}\max_{y\in Y}\left(\mathbb{E}_{u \sim p_{data}}\left(\log D_y(u)\right) + \mathbb{E}_{z\sim\mu}\left(\log\left(1 - D_y\left(G_x(z)\right)\right)\right)\right). \tag{3.63}$$

Зауваження 3.12. Функція $y \mapsto L(x,y)$ є угнутою, якщо дискримінатор D_y реалізовано у вигляді нейронної мережі з одним внутрішнім шаром та пороговою функцією

$$D_y(u) = \frac{1}{1 + e^{-(y,u)}}.$$

Локальні необхідні умови, що характеризують розв'язок $(x^*, y^*) \in X \times Y$ задачі (3.63) мають вигляд:

$$(\nabla_x L(x^*, y^*), x - x^*) + (-\nabla_y L(x^*, y^*), y - y^*) \geq 0 \quad \forall (x, y) \in X \times Y. \quad (3.64)$$

За відсутності обмежень ($X = \mathbb{R}^k$, $Y = R^m$) варіаційна нерівність (3.64) набуває вигляду:

$$\nabla_x L(x^*, y^*) = 0 \quad \text{та} \quad -\nabla_y L(x^*, y^*) = 0.$$

В роботах [49–53] алгоритми для варіаційних нерівностей застосовувались для навчання генеруючих змагальних нейронних мереж.

Зрозуміло, що створення більш швидких та стійких алгоритмів для варіаційних нерівностей (особливо для стохастичних постановок) веде до прогресу в навчанні породжуючих змагальних мереж.

3.7. Додаток: збіжність методу Гальперна

Нехай H — гільбертовий простір, $C \subseteq H$ — непорожня опукла замкнена множина, $T : C \to C$ — нерозтягуючий оператор. Припустимо, що $F(T) \neq \emptyset$. Розглядаємо таку ітераційну процедуру.

Алгоритм 3.10. Задаємо $x_0 \in C$ та $y \in C$, генеруємо послідовність елементів $x_n \in C$ за допомогою ітераційної схеми

$$x_{n+1} = \alpha_n y + (1 - \alpha_n) T x_n,$$

де (α_n) — послідовність чисел з $(0, 1)$.

Теорема 3.10 (H.K. Xu, [67]). *Нехай $T : C \to C$ — нерозтягуючий оператор із $F(T) \neq \emptyset$, $y \in C$. Нехай послідовність чисел $\alpha_n \in (0, 1)$ задовольняє умови:*

1) $\lim_{n\to\infty} \alpha_n = 0$,

2) $\sum_{n=0}^{\infty} \alpha_n = +\infty$,

3) $\sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n| < +\infty$ *або* $\lim_{n\to\infty} \frac{\alpha_{n+1} - \alpha_n}{\alpha_{n+1}} = 0$.

Тоді згенерована алгоритмом 3.10 послідовність (x_n) сильно збігається до точки $P_{F(T)} y$.

Доведення. Покажемо обмеженість послідовності (x_n). Для $p \in F(T)$ маємо

$$\|x_{n+1} - p\| \leq (1-\alpha_n)\|Tx_n - p\| + \alpha_n \|y - p\| \leq$$
$$\leq (1-\alpha_n)\|x_n - p\| + \alpha_n \|y - p\| \leq \max\{\|x_n - p\|, \|y - p\|\}.$$

Звідси індукцією отримуємо $\|x_n - p\| \leq \{\|x_0 - p\|, \|y - p\|\}$, $n \geq 0$.

Таким чином, послідовність (x_n) обмежена, а разом з нею — також послідовність (Tx_n).

Покажемо, що

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0 \tag{3.65}$$

та

$$\lim_{n\to\infty} \|x_n - Tx_n\| = 0. \tag{3.66}$$

Використаємо лему 3.3 про числові послідовності. Маємо

$$\|x_{n+1} - x_n\| = \|(\alpha_n - \alpha_{n-1})(y - Tx_{n-1}) + (1-\alpha_n)(Tx_n - Tx_{n-1})\| \leq$$
$$\leq (1-\alpha_n)\|x_n - x_{n-1}\| + |\alpha_n - \alpha_{n-1}| \cdot \|y - Tx_{n-1}\|.$$

З леми 3.3 про числові послідовності випливає (3.65). Тепер ми можемо отримати (3.66) із (3.65).

$$\|x_n - Tx_n\| \leq \|x_n - x_{n+1}\| + \alpha_n \|y - Tx_n\| \to 0.$$

Доведемо, що

$$\limsup_{n\to\infty} (P_{F(T)}y - x_n, P_{F(T)}y - y) \leq 0. \tag{3.67}$$

Виділимо з (x_n) підпослідовність (x_{n_k}) таку, що

$$\limsup_{n\to\infty} (P_{F(T)}y - x_n, P_{F(T)}y - y) = \lim_{k\to\infty} (P_{F(T)}y - x_{n_k}, P_{F(T)}y - y).$$

Можна вважати, що $x_{n_k} \rightharpoonup \tilde{x}$. Із (3.66) випливає включення $\tilde{x} \in F(T)$. Тому одержуємо

$$\lim_{k\to\infty} (P_{F(T)}y - x_{n_k}, P_{F(T)}y - y) = (P_{F(T)}y - \tilde{x}, P_{F(T)}y - y) \leq 0,$$

чим і доводимо (3.67).

Покажемо тепер, що $x_n \to P_{F(T)}y$. Маємо

$$\left\|x_{n+1} - P_{F(T)}y\right\|^2 \leq (1-\alpha_n)\left\|x_n - P_{F(T)}y\right\|^2 +$$
$$+ 2\alpha_n \left(y - P_{F(T)}y, x_{n+1} - P_{F(T)}y\right). \tag{3.68}$$

Застосувавши до одержаної рекурентної нерівності (3.68) лему 3.3 про числові послідовності, робимо висновок, що $\left\|x_n - P_{F(T)}y\right\| \to 0$. $\square$

Зауваження 3.13. Розглянута схема є узагальненням усереднення за Чезаро. Дійсно, якщо T — лінійний оператор, $\alpha_n = \frac{1}{n+1}$ та $x_0 = y$, то породжена алгоритмом Гальперна послідовність має вигляд

$$x_n = \frac{1}{n+1}\sum_{k=0}^{n} T^k y,$$

а за умови $\|T\| \leq 1$ послідовність (x_n) сильно збігається до $P_{N(I-T)}y$. Отже, з теореми 3.10 випливає ергодична теорема фон Неймана.

Нехай оператор $T : H \to H$ — нерозтягуючий, оператор $A : H \to H$ — ліпшицевий та сильно монотонний[2]. Розглянемо ітераційну схему:

$$\begin{cases} y_n = Tx_n, \\ x_{n+1} = y_n - \alpha_n A y_n, \end{cases}$$

де послідовність чисел $\alpha_n \in (0,1)$ задовольняє умови: 1) $\lim_{n\to\infty} \alpha_n = 0$; 2) $\sum_{n=0}^{\infty} \alpha_n = +\infty$; 3) $\sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n| < +\infty$.

Аналогічно теоремі 3.10 доводиться, що послідовність (x_n) сильно збігається до єдиного розв'язку варіаційної нерівності:

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

Розглянемо задачу

$$f(x) \to \min, \quad x \in \bigcap_{i=1}^{N} C_i, \tag{3.69}$$

де f — опукла функція з ліпшицевим градієнтом, C_i — опуклі замкнені множини. Указана схема дозволяє для (3.69) обгрунтувати такий метод

$$\begin{cases} y_n = \left(\frac{1}{N}\sum_{i=1}^{N} P_{C_i}\right) x_n, \\ x_{n+1} = y_n - \alpha_n \nabla f(y_n), \end{cases}$$

де $\alpha_n > 0$ задовольняє наведені вище умови.

[2]Нагадаємо, що оператор $A : H \to H$ називають сильно монотонним на множині $C \subseteq H$, якщо існує така стала $\mu > 0$, що $(Ax - Ay, x - y) \geq \mu\|x - y\|^2 \quad \forall x,\ y \in C$.

Розділ 4

Алгоритм екстраполяції з минулого

Найвідомішим чисельним методом розв'язання варіаційних нерівностей є екстраградієнтний метод, що запропонований Г.М. Корпелевич в 1970-х роках [74]. В 1980 році Л.Д. Попов [92] запропонував для пошуку сідлових точок опукло-угнутих функцій цікаву модифікацію класичного методу Ерроу–Гурвіца, яка стала джерелом багатьох сучасних алгоритмів прикладного нелінійного аналізу. Алгоритми цього типу відомі серед спеціалістів з машинного навчання під назвою «Extrapolation from the Past» [49].

У розділі розглядаються варіаційні нерівності з операторами, що діють в гільбертовому просторі. Для цих задач запропоновано та обгрунтовано нові модифікації методу Л.Д. Попова.

4.1. Постановка задачі та допоміжні відомості

Нехай H — дійсний гільбертовий простір з скалярним добутком $(\cdot, \cdot)$ та породженою нормою $\|\cdot\|$.

Нехай C — непорожня опукла і замкнена підмножина простору H та $A : H \to H$ – деякий оператор.

Означення 4.1. Варіаційною нерівністю називаємо таку задачу:

$$\text{знайти } x \in C : \ (Ax, y - x) \geq 0 \quad \forall y \in C. \tag{4.1}$$

Множину розв'язків варіаційної неравності (4.1) позначимо через S.

У вигляді варіаційної нерівності можуть бути сформульовані задачі розв'язання рівнянь, знаходження екстремуму функціоналів, знаходження точок рівноваги тощо. Наведемо ряд типових постановок.

(1) Задача розв'язання операторного рівняння

$$\text{знайти } x \in H: \ Ax = 0$$

рівносильна варіаційній нерівності

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

(2) Нехай f — диференційовна опукла функція, C — опукла замкнена множина. Критерій оптимальності першого порядку для задачі

$$f \to \min_C$$

має вигляд

$$x \in C \quad \text{та} \quad (\nabla f(x), y - x) \geq 0 \quad \forall y \in C.$$

(3) Нехай $F : X \times Y \to \mathbb{R}$ — диференційовна опукло-угнута функція, X, Y — опуклі замкнені множини. Точка $(x^*, y^*) \in X \times Y$ називається сідловою точкою функції F, якщо

$$F(x^*, y) \leq F(x^*, y^*) \leq F(x, y^*) \quad \forall x \in X \ \forall y \in Y. \tag{4.2}$$

Задача пошуку сідлової точки (4.2) рівносильна варіаційній нерівності:

$$\left(\begin{pmatrix} \nabla_x F(x^*, y^*) \\ -\nabla_y F(x^*, y^*) \end{pmatrix}, \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix} \right) \geq 0 \quad \forall (x, y) \in X \times Y.$$

(4) Нехай X, Y — опуклі замкнені множини, $f : X \times Y \to \mathbb{R}$ ($g : X \times Y \to \mathbb{R}$) — диференційовна за x (по y) при фіксованому y (при фіксованому x) функція. Точка $(x^*, y^*) \in X \times Y$ називається рівновагою Неша, якщо

$$f(x^*, y^*) \leq f(x, y^*) \quad \forall x \in X, \quad g(x^*, y^*) \leq g(x^*, y) \quad \forall y \in Y. \tag{4.3}$$

Якщо функція $f(\cdot, y)$ опукла на X для всіх $y \in Y$, а функція $g(x, \cdot)$ опукла на Y для всіх $x \in X$, то задача пошуку рівноваги Неша (4.3) рівносильна варіаційній нерівності:

$$\left(\begin{pmatrix} \nabla_x f(x^*, y^*) \\ \nabla_y g(x^*, y^*) \end{pmatrix}, \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix} \right) \geq 0 \quad \forall (x, y) \in X \times Y.$$

Нагадаємо основні означення [54, 76, 93].

Означення 4.2. Оператор $A : H \to H$ називаємо псевдомонотонним на множині $C \subseteq H$, якщо для x, $y \in C$

$$(Ax, y - x) \geq 0 \quad \Rightarrow \quad (Ay, x - y) \leq 0.$$

Означення 4.3. Оператор $A : H \to H$ називаємо монотонним, якщо

$$(Ax - Ay, x - y) \geq 0 \quad \forall x,\ y \in H.$$

Означення 4.4. Оператор $A : H \to H$ називаємо обернено сильно монотонним на множині $C \subseteq H$, якщо існує така стала $\alpha > 0$, що

$$(Ax - Ay, x - y) \geq \alpha \|Ax - Ay\|^2 \quad \forall x,\ y \in C.$$

У цьому випадку кажуть, що оператор A — α-обернено сильно монотонним (ко-коерцитивним, α-ко-коерцитивним).

Означення 4.5. Оператор $A : H \to H$ називаємо рівномірно монотонним на множині $C \subseteq H$, якщо існує така зростаюча функція $\phi : [0, +\infty) \to [0, +\infty)$, $\phi(0) = 0$ та $\phi(t) > 0$ для $t > 0$, що

$$(Ax - Ay, x - y) \geq \phi\left(\|x - y\|\right) \quad \forall x,\ y \in C.$$

Означення 4.6. Оператор $A : H \to H$ називаємо сильно монотонним на множині $C \subseteq H$, якщо існує така стала $\mu > 0$, що

$$(Ax - Ay, x - y) \geq \mu \|x - y\|^2 \quad \forall x,\ y \in C.$$

У цьому випадку кажуть, що оператор A — μ-сильно монотонний.

Скрізь у цьому розділі будемо вважати, що оператор $A : H \to H$ — ліпшицевий на множині C (з константою $L > 0$), тобто

$$\|Ax - Ay\| \leq L \|x - y\| \quad \forall x, y \in C,$$

та $S \neq \emptyset$.

Для замкненої опуклої множини C та ліпшицевого сильно монотонного оператора A множина S не порожня та складається з одного елемента [44].

Нагадаємо декілька відомих та потрібних нам фактів.

Нехай P_C — оператор метричного проектування на замкнену опуклу підмножину $C \subseteq H$, тобто $P_C x$ — єдиний елемент C, що володіє властивістю

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

Елемент $P_C x$ можна охарактеризувати таким чином [44, 76]:

$$y = P_C x \quad \Leftrightarrow \quad y \in C \text{ та } (y - x, z - y) \geq 0 \ \forall z \in C, \tag{4.4}$$

$$y = P_C x \quad \Leftrightarrow \quad y \in C \text{ та } \|y - z\|^2 \leq \|x - z\|^2 - \|y - x\|^2 \ \forall z \in C. \tag{4.5}$$

Оператор метричного проектування P_C є нерозтягуючий, тобто

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

точніше, обернено сильно монотонним (1-ко-коерцитивним)

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

Варіаційну нерівність (4.1) можна сформулювати як задачу пошуку нерухомої точки [44, 76]:

$$x = P_C (x - \lambda A x), \tag{4.6}$$

де $\lambda > 0$.

Формулювання (4.6) корисне, оскільки веде до ітераційної схеми

$$x_{n+1} = P_C (x_n - \lambda A x_n), \tag{4.7}$$

яка сильно збіжна для ліпшицевих сильно монотонних операторів та слабко збіжна для обернено сильно монотонних (ко-коерцитивних) операторів [54, 76]. Але для ліпшицевих монотонних операторів схема (4.7) в загальному випадку не збігається.

Розглянемо дуальну варіаційну нерівність:

$$\text{знайти } x \in C : \ (Ay, y - x) \geq 0 \quad \forall y \in C. \tag{4.8}$$

Множину розв'язків варіаційної нерівності (4.8) позначимо S^d. Відомо, що множина S^d опукла та замкнена [44]. Нерівність (4.8) називають слабким або дуальним формулюванням варіаційної нерівності (4.1) (або нерівністю типу Мінті), а розв'язки нерівності (4.8) — слабкими розв'язками варіаційної нерівності (4.1).

Для монотонних (псевдомонотонних) операторів A завжди маємо

$$S \subseteq S^d.$$

А коли оператор A монотонний та неперервний маємо (лема Мінті, [44])

$$S^d = S.$$

Однією з основних теоретичних задач є оцінка числа ітерацій алгоритму, що необхідне для отримання наближеного розв'язку заданої якості. Якість наближеного розв'язку $x \in C$ варіаційної нерівності (4.1) у монотонному випадку будемо оцінювати за допомогою невід'ємної функції зазору [47, 94]

$$\operatorname{gap}(x) = \sup_{y \in C} (Ay, x - y). \tag{4.9}$$

Очевидно, що для коректності означення функції зазору (4.9) необхідна обмеженість допустимої множини C.

Лема 4.1 (Nesterov Y., [94]). *Нехай оператор* A *— монотонний. Якщо* $x \in C$ *— розв'язок* (4.1), *то*

$$\operatorname{gap}(x) = 0.$$

Навпаки, якщо для $x \in C$ *маємо*

$$\operatorname{gap}(x) = 0,$$

то x *— розв'язок* (4.1).

Нагадаємо відомі леми про числові нерівності.

Лема 4.2. *Нехай невід'ємні послідовності* (α_n), (β_n), *такі, що*

$$\alpha_{n+1} \le \alpha_n - \beta_n, \quad n \ge 1.$$

Тоді існує границя $\lim_{n\to\infty} \alpha_n \in \mathbb{R}$ *та* $\sum_{n=1}^{\infty} \beta_n < +\infty$.

Лема 4.3. *Нехай послідовність невід'ємних чисел* (ξ_n) *задовольняє рекурентній нерівності*

$$\xi_{n+1} \le (1 - \alpha_n)\,\xi_n + \alpha_n\beta_n + \gamma_n, \quad n \ge 1,$$

де послідовності (α_n) *та* (β_n) *мають властивості:*

1) $\alpha_n \in (0, 1)$, $\sum_{n=1}^{\infty} \alpha_n = +\infty$;

2) $\limsup_{n\to\infty} \beta_n \le 0$;

3) $\gamma_n \in [0, +\infty)$, $\sum_{n=1}^{\infty} \gamma_n < +\infty$.

Тоді $\lim\limits_{n\to\infty} \xi_n = 0$.

Лема 4.4 (P.-E. Mainge, [56]). *Нехай числова послідовність* (α_n) *має підпослідовність* (α_{n_k}) *з властивістю* $\alpha_{n_k} < a_{n_k+1}$ *для всіх* $k \geq 1$. *Тоді існує така неспадна послідовність* (m_k) *натуральних чисел, що*

$$m_k \to +\infty, \quad \alpha_{m_k} \leq \alpha_{m_k+1}, \quad \alpha_k \leq \alpha_{m_k+1}$$

для всіх $k \geq n_1$.

При доведенні слабкої збіжності послідовностей елементів гільбертового простору будемо використовувати відому лему Опяла.

Лема 4.5 (Z. Opial, [55]). *Нехай послідовність* (x_n) *елементів гільбертового простору* H *слабко збігається до точки* $x \in H$. *Тоді для всіх* $y \in H \setminus \{x\}$ *маємо*

$$\liminf_{n\to\infty} \|x_n - x\| < \liminf_{n\to\infty} \|x_n - y\|.$$

4.2. Алгоритм екстраполяції з минулого

Для розв'язання задачі (4.1) розглянемо такий ітераційний алгоритм.

Алгоритм 4.1. Для $x_1 = y_0 \in C$ генерируємо послідовність елементів x_n, $y_n \in C$ за допомогою ітераційної схеми

$$\begin{cases} y_n = P_C\left(x_n - \lambda_n A y_{n-1}\right), \\ x_{n+1} = P_C\left(x_n - \lambda_n A y_n\right), \end{cases}$$

де $\lambda_n > 0$.

Зауваження 4.1. Частнинний випадок алгоритму 4.1 запропонований Л.Д. Поповим [92] для пошуку сідлових точок. Останнім часом цей метод став відомим у середовищі спеціалістів з машинного навчання під назвою «Extrapolation from the Past» (екстраполяція з минулого) [49]. Модифікація алгоритму з одним метричним проектуванням на допустиму множину запропонована в [95], модифікації з бегманівською відстанню досліджено в [96–102].

При виконанні для деякого $n \in \mathbb{N}$ в алгоритмі 4.1 рівностей

$$y_n = y_{n-1} = x_n \quad \text{або} \quad x_{n+1} = x_n = y_n \tag{4.10}$$

має місце включення $y_n \in S$. Дійсно, рівність

$$x_{n+1} = P_C\left(x_n - \lambda_n A y_n\right)$$

рівносильна нерівності

$$(Ay_n, y - x_{n+1}) + \frac{(x_{n+1} - x_n, y - x_{n+1})}{\lambda_n} \geq 0 \quad \forall y \in C.$$

З другої рівності (4.10) випливає

$$(Ay_n, y - y_n) \geq 0 \ \ \forall y \in C,$$

тобто, $y_n \in S$.

Аналогічно, з

$$(Ay_{n-1}, y - y_n) + \frac{(y_n - x_n, y - y_n)}{\lambda_n} \geq 0 \quad \forall y \in C$$

при першій рівності в (4.10) отримуємо $y_n \in S$.

Далі припустимо, що для всіх номерів $n \in \mathbb{N}$ умова (4.10) не має місця.

4.3. Слабка збіжність та сублінійна оцінка

Припустимо, що оператор $A : H \to H$ є псевдомонотонним та L-ліпшицевим на множині C.

Лема 4.6. *Для породжених алгоритмом 4.1 послідовностей* (x_n), (y_n) *та точки* $z \in S$ *виконується нерівність*

$$\begin{aligned} \|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - \lambda_n L) \|x_{n+1} - y_n\|^2 - \\ - (1 - 2\lambda_n L) \|y_n - x_n\|^2 + 2\lambda_n L \|x_n - y_{n-1}\|^2 . \end{aligned} \quad (4.11)$$

Доведення. Маємо

$$\begin{aligned} &\|x_{n+1} - z\|^2 = \\ &= \|x_n - z\|^2 - \|x_n - x_{n+1}\|^2 + 2\,(x_{n+1} - x_n, x_{n+1} - z) = \\ &= \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 - \\ &- 2\,(x_n - y_n, y_n - x_{n+1}) + 2\,(x_{n+1} - x_n, x_{n+1} - z) . \end{aligned} \quad (4.12)$$

З правила породження точок x_{n+1} та y_n випливає

$$\lambda_n(Ay_n, z - x_{n+1}) \geq (x_{n+1} - x_n, x_{n+1} - z), \quad (4.13)$$

$$\lambda_n(Ay_{n-1}, x_{n+1} - y_n) \geq -(x_n - y_n, y_n - x_{n+1}). \quad (4.14)$$

Використавши нерівності (4.13), (4.14) для оцінки скалярних добутків в (4.12), отримуємо

$$\begin{aligned}\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - \|x_n-y_n\|^2 - \|y_n-x_{n+1}\|^2 + \\ + 2\lambda_n\{(Ay_{n-1}, x_{n+1}-y_n) + (Ay_n, z-x_{n+1})\} = \\ = \|x_n-z\|^2 - \|x_n-y_n\|^2 - \|y_n-x_{n+1}\|^2 + \\ + 2\lambda_n\{(Ay_n, z-y_n) + (Ay_{n-1}, x_{n+1}-y_n) + (Ay_n, y_n-x_{n+1})\}.\end{aligned} \tag{4.15}$$

З псевдомонотонності оператора A та включення $z \in S$ випливає

$$(Ay_n, z-y_n) \le 0,$$

а ліпшицевість F гарантує виконання нерівності

$$\begin{aligned}(Ay_{n-1}-Ay_n, x_{n+1}-y_n) \le \\ \le L\,\|y_{n-1}-y_n\|\,\|x_{n+1}-y_n\| \le \\ \le \frac{L}{2}\,\|y_{n-1}-y_n\|^2 + \frac{L}{2}\,\|y_n-x_{n+1}\|^2.\end{aligned}$$

Використавши вищенаведені оцінки в (5.10), отримуємо

$$\begin{aligned}\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - \|x_n-y_n\|^2 - \|y_n-x_{n+1}\|^2 + \\ + \lambda_n L\,\|y_{n-1}-y_n\|^2 + \lambda_n L\,\|y_n-x_{n+1}\|^2.\end{aligned} \tag{4.16}$$

Член $\|y_{n-1}-y_n\|^2$ оцінимо таким чином

$$\|y_{n-1}-y_n\|^2 \le 2\,\|y_{n-1}-x_n\|^2 + 2\,\|y_n-x_n\|^2.$$

Ураховуючи цю оцінку в (4.16), приходимо до нерівності

$$\begin{aligned}\|x_{n+1}-z\|^2 \le \|x_n-z\|^2 - \|x_n-y_n\|^2 - \|y_n-x_{n+1}\|^2 + \\ + 2\lambda_n L\,\|y_{n-1}-x_n\|^2 + 2\lambda_n L\,\|y_n-x_n\|^2 + \lambda_n L\,\|y_n-x_{n+1}\|^2,\end{aligned}$$

тобто, до нерівності (4.11). □

Перейдемо безпосередньо до доведення слабкої збіжності алгоритму 4.1 у випадку монотонності A.

Нехай $z \in S$. Покладемо

$$\begin{aligned}&\alpha_n = \|x_n-z\|^2 + 2\lambda_n L\,\|y_{n-1}-x_n\|^2, \\ &\beta_n = (1-2\lambda_n L)\,\|y_n-x_n\|^2 + (1-2\lambda_{n+1}L-\lambda_n L)\,\|y_n-x_{n+1}\|^2.\end{aligned}$$

Тоді (4.11) приймає вигляд

$$\alpha_{n+1} \le \alpha_n - \beta_n.$$

Вимагатимемо виконання умови

$$0 < \underline{\lambda} \le \lambda_n \le \bar{\lambda} < \frac{1}{3L}.$$

Тоді з леми 4.2 можемо зробитии висновок, що існує границя

$$\lim_{n\to\infty} \left(\|x_n - z\|^2 + 2\lambda_n L \|y_{n-1} - x_n\|^2 \right)$$

та

$$\lim_{n\to\infty} \left((1 - 2\lambda_n L) \|y_n - x_n\|^2 + (1 - 2\lambda_{n+1} L - \lambda_n L) \|y_n - x_{n+1}\|^2 \right) = 0.$$

Звідки отримуємо

$$\lim_{n\to\infty} \|y_n - x_n\| = \lim_{n\to\infty} \|y_n - x_{n+1}\| = \lim_{n\to\infty} \|x_n - x_{n+1}\| = 0 \tag{4.17}$$

та збіжність числових послідовностей $(\|x_n - z\|)$, $(\|y_n - z\|)$ для всіх $z \in S$. Зокрема, послідовності (x_n), (y_n) обмежені.

З (4.17) випливає

$$\lim_{n\to\infty} (Ay_n, x_{n+1} - y_n) = 0. \tag{4.18}$$

Розглянемо підпослідовність (x_{n_k}), слабко збіжну до деякої точки $\bar{z} \in C$. Тоді з (4.17) випливає, що й

$$y_{n_k} \rightharpoonup \bar{z}.$$

Припустимо, що оператор A монотонний та покажемо, що $\bar{z} \in S$. Маємо

$$(Ay_n, y - y_n) \ge (Ay_n, x_{n+1} - y_n) + \frac{(x_{n+1} - x_n, x_{n+1} - y)}{\lambda_n} \quad \forall y \in C. \tag{4.19}$$

Здійснивши граничний перехід в (4.19) з урахуванням (5.13), (5.14) та умови монотонності оператора A, отримаємо

$$\begin{aligned}(Ay, y - \bar{z}) = \lim_{k\to\infty} (Ay, y - y_{n_k}) \ge \limsup_{k\to\infty} (Ay_{n_k}, y - y_{n_k}) \ge \\ \ge \lim_{k\to\infty} \left\{ (Ay_{n_k}, x_{n_k+1} - y_{n_k}) + \frac{(x_{n_k+1} - x_{n_k}, x_{n_k+1} - y)}{\lambda_{n_k}} \right\} = 0 \quad \forall y \in C,\end{aligned}$$

тобто, $\bar{z} \in S$.

Покажемо тепер, що $x_n \rightharpoonup \bar{z}$. Тоді з $x_n - y_n \to 0$ випливає, що й послідовність (y_n) слабко збігається $\bar{z}$. Міркуємо від супротивного. Нехай існує така підпослідовність (x_{m_k}), що $x_{m_k} \rightharpoonup \tilde{z}$ та $\tilde{z} \neq \bar{z}$. Ясно, що $\tilde{z} \in S$. Застосуємо двічі лему Опяла. Маємо

$$\lim_{n\to\infty} \|x_n - \bar{z}\| = \lim_{k\to\infty} \|x_{n_k} - \bar{z}\| < \lim_{k\to\infty} \|x_{n_k} - \tilde{z}\| = \lim_{n\to\infty} \|x_n - \tilde{z}\| = \\ = \lim_{k\to\infty} \|x_{m_k} - \tilde{z}\| < \lim_{k\to\infty} \|x_{m_k} - \bar{z}\| = \lim_{n\to\infty} \|x_n - \bar{z}\|,$$

що неможливо. Отже, $x_n \rightharpoonup \bar{z}$.

Сформулюємо отриманий результат.

Теорема 4.1. *Нехай H — гільбертовий простір, $C \subseteq H$ — непорожня опукла замкнена множина, $A : H \to H$ — монотонний та L-ліпшицевий на множині C оператор та $S \neq \emptyset$. Припустимо, що*

$$\lambda_n \in \left[\underline{\lambda}, \overline{\lambda}\right] \subseteq \left(0, \tfrac{1}{3L}\right).$$

Тоді породжені алгоритмом 4.1 послідовності (x_n), (y_n) слабко збігаються до розв'язку $\bar{z} \in S$ варіаційної нерівності (4.1), причому

$$\lim_{n\to\infty} \|x_n - y_n\| = 0.$$

У випадку обмеженості допустимої множини C доведемо, що алгоритму 4.1 необхідно зробити

$$O\left(\frac{LD^2}{\varepsilon}\right)$$

ітерацій для отримання точки $x \in C$ з

$$\operatorname{gap}(x) \leq \varepsilon,$$

де $\varepsilon > 0$, $D = \sup_{a,b\in C} \|a - b\| < +\infty$.

Нехай $y \in C$. Маємо

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 - (1 - \lambda_n L) \|x_{n+1} - y_n\|^2 - \\ - (1 - 2\lambda_n L) \|y_n - x_n\|^2 + 2\lambda_n L \|x_n - y_{n-1}\|^2 + 2\lambda_n (Ay_n, y - y_n). \quad (4.20)$$

З монотонності оператора A та (4.20) випливає

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 - (1 - \lambda_n L) \|x_{n+1} - y_n\|^2 - \\ - (1 - 2\lambda_n L) \|y_n - x_n\|^2 + 2\lambda_n L \|x_n - y_{n-1}\|^2 + 2\lambda_n (Ay, y - y_n). \quad (4.21)$$

Перепишемо (4.21) у вигляді

$$2\lambda_n(Ay, y_n - y) \le \left(\|x_n - y\|^2 + 2\lambda_n L\,\|x_n - y_{n-1}\|^2\right) -$$
$$- \left(\|x_{n+1} - y\|^2 + 2\lambda_{n+1} L\,\|x_{n+1} - y_n\|^2\right) -$$
$$- (1 - 2\lambda_{n+1}L - \lambda_n L)\,\|x_{n+1} - y_n\|^2 - (1 - 2\lambda_n L)\,\|y_n - x_n\|^2. \quad (4.22)$$

Припустимо, що $\lambda_n \in (0, 1/3L]$. Тоді з (4.22) випливає

$$2\lambda_n(Ay, y_n - y) \le \left(\|x_n - y\|^2 + 2\lambda_n L\,\|x_n - y_{n-1}\|^2\right) -$$
$$- \left(\|x_{n+1} - y\|^2 + 2\lambda_{n+1} L\,\|x_{n+1} - y_n\|^2\right). \quad (4.23)$$

Просумувавши (4.23) по n від 1 до N отримаємо

$$2\sum_{n=1}^{N} \lambda_n(Ay, y_n - y) \le \|x_1 - y\|^2 + 2\lambda_1 L\,\|x_1 - y_0\|^2,$$

та

$$(Ay, z_N - y) \le \frac{\|x_1 - y\|^2}{2\sum_{n=1}^{N} \lambda_n}, \quad (4.24)$$

де $z_N = \frac{\sum_{n=1}^{N} \lambda_n y_n}{\sum_{n=1}^{N} \lambda_n}$. Переходимо до супремуму за $y \in C$ в (4.23)

$$\mathrm{gap}(z_N) \le \frac{\sup_{y \in C} \|x_1 - y\|^2}{2\sum_{n=1}^{N} \lambda_n}.$$

Таким чином, має місце

Теорема 4.2. *Нехай $C \subseteq H$ — непорожня опукла замкнена обмежена множина, $A : H \to H$ — монотонний та L-ліпшицевий на C оператор. Нехай (y_n) — послідовність, що породжена ітераційним алгоритмом 4.1 з $\lambda_n = 1/3L$, тобто,*

$$\begin{cases} x_1 = y_0 \in C, \\ y_n = P_C\left(x_n - \frac{1}{3L} A y_{n-1}\right), \\ x_{n+1} = P_C\left(x_n - \frac{1}{3L} A y_n\right). \end{cases}$$

Тоді для послідовності середніх $z_N = \frac{1}{N}\sum_{n=1}^{N} y_n$ має місце оцінка

$$\mathrm{gap}\,(z_N) \le \frac{3L \sup_{y \in C} \|x_1 - y\|^2}{2N}.$$

4.4. Сильна збіжність

Припустимо, що оператор $A : H \to H$ є рівномірно монотонним на обмежених підмножинах множини $C \subseteq H$. Тоді варіаційна нерівність (4.1) має єдиний розв'язок $\bar{z} \in C$.

Покажемо, що породжені алгоритмом 4.1 послідовності (x_n), (y_n) сильно збігаються до $\bar{z}$.

Оскільки множина $\{y_n\} \cup \{\bar{z}\} \subseteq C$ обмежена, то якщо існує така зростаюча функція $\phi : [0, +\infty) \to [0, +\infty)$, $\phi(0) = 0$ та $\phi(t) > 0$ для $t > 0$, що

$$(Ay_n, \bar{z} - y_n) \leq -\phi \left(\|\bar{z} - y_n\| \right).$$

Замість нерівності леми 5.1 запишемо її уточнену версію

$$\begin{aligned} \|x_{n+1} - \bar{z}\|^2 + 2\lambda_n \phi \left(\|\bar{z} - y_n\| \right) \leq \|x_n - \bar{z}\|^2 - (1 - \lambda_n L) \|x_{n+1} - y_n\|^2 - \\ - (1 - 2\lambda_n L) \|y_n - x_n\|^2 + 2\lambda_n L \|x_n - y_{n-1}\|^2. \quad (4.25) \end{aligned}$$

Перепишемо (4.25) у вигляді

$$\begin{aligned} 2\lambda_n \phi \left(\|\bar{z} - y_n\| \right) \leq \Big(\|x_n - \bar{z}\|^2 + 2\lambda_n L \|x_n - y_{n-1}\|^2 \Big) - \\ - \Big(\|x_{n+1} - \bar{z}\|^2 + 2\lambda_{n+1} L \|x_{n+1} - y_n\|^2 \Big) - \\ - (1 - 2\lambda_{n+1} L - \lambda_n L) \|x_{n+1} - y_n\|^2 - \\ - (1 - 2\lambda_n L) \|y_n - x_n\|^2. \quad (4.26) \end{aligned}$$

Нерівність (4.26) та припущення $\lambda_n \in \left[\underline{\lambda}, \overline{\lambda}\right] \subseteq \left(0, \frac{1}{3L}\right)$ дають

$$\sum_{n=1}^{\infty} \lambda_n \phi \left(\|\bar{z} - y_n\| \right) < +\infty \quad \text{та} \quad \lim_{n \to \infty} \phi \left(\|\bar{z} - y_n\| \right) = 0.$$

Отже, $\|\bar{z} - y_n\| \to 0$ та $\|\bar{z} - x_n\| \to 0$.

Таким чином, має місце

Теорема 4.3. *Нехай H — гільбертовий простір, $C \subseteq H$ — непорожня опукла замкнена множина, $A : H \to H$ — рівномірно монотонний на обмежених підмножинах множини $C \subseteq H$ та L-ліпшицевий на множині C оператор, $\bar{z} \in C$ — єдиний розв'язок варіаційної нерівності (4.1). Припустимо, що $\lambda_n \in \left[\underline{\lambda}, \overline{\lambda}\right] \subseteq \left(0, \frac{1}{3L}\right)$. Тоді породжені алгоритмом 4.1 послідовності (x_n), (y_n) сильно збігаються до $\bar{z}$.*

4.5. Лінійна швидкість збіжності

Припустимо, що оператор A є L-ліпшицевим та μ-сильно монотонним на множині C. Тоді існує єдиний розв'язок $z \in C$ варіаційної нерівності (4.1) та виконується нерівність

$$(Ax, x - z) \geq \mu \|x - z\|^2 \quad \forall x \in C \tag{4.27}$$

для деякого $\mu > 0$.

Розглянемо варіант алгоритму 4.1 з лінійною швидкістю збіжності.

Алгоритм 4.2. Для $x_1 = y_0 \in C$ генерируємо послідовність елементів x_n, $y_n \in C$ за допомогою ітераційної схеми

$$\begin{cases} y_n = P_C\left(x_n - \frac{1}{4L} A y_{n-1}\right), \\ x_{n+1} = P_C\left(x_n - \frac{1}{4L} A y_n\right). \end{cases}$$

Покажемо, що

$$\|x_{n+1} - z\|^2 = O\left(\left(1 - \frac{\mu}{4L}\right)^n\right), \quad \|y_n - x_{n+1}\|^2 = \left(\left(1 - \frac{\mu}{4L}\right)^n\right).$$

Запишемо нерівність (5.10):

$$\begin{aligned} \|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda\{(Ay_n, z - y_n) + (Ay_{n-1} - Ay_n, x_{n+1} - y_n)\}. \end{aligned}$$

З умови (4.27) та нерівності $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ випливає

$$(Ay_n, y_n - z) \geq \mu \|y_n - z\|^2 \geq \mu \left(\frac{1}{2}\|x_n - z\|^2 - \|y_n - x_n\|^2\right),$$

а ліпшицевість A гарантує виконання нерівності

$$\begin{aligned} (Ay_{n-1} - Ay_n, x_{n+1} - y_n) \leq L \|y_{n-1} - y_n\| \|x_{n+1} - y_n\| \leq \\ \leq \frac{L}{2} \|y_{n-1} - y_n\|^2 + \frac{L}{2} \|y_n - x_{n+1}\|^2. \end{aligned}$$

Використавши вищенаведені оцінки, отримуємо

$$\begin{aligned} \|x_{n+1} - z\|^2 \leq (1 - \lambda\mu) \|x_n - z\|^2 - (1 - 2\lambda\mu) \|x_n - y_n\|^2 - \\ - \|y_n - x_{n+1}\|^2 + \lambda L \|y_{n-1} - y_n\|^2 + \lambda L \|y_n - x_{n+1}\|^2. \end{aligned} \tag{4.28}$$

Член $\|y_{n-1} - y_n\|^2$ оцінимо таким чином

$$\|y_{n-1} - y_n\|^2 \le 2\,\|y_{n-1} - x_n\|^2 + 2\,\|y_n - x_n\|^2 .$$

Враховуючи цю оцінку в (4.28), приходимо до нерівності

$$\|x_{n+1} - z\|^2 \le (1 - \lambda\mu)\,\|x_n - z\|^2 - (1 - 2\lambda L - 2\lambda\mu)\,\|x_n - y_n\|^2 - \\ - (1 - 2\lambda L)\,\|y_n - x_{n+1}\|^2 + 2\lambda L\,\|y_{n-1} - x_n\|^2 .$$

Перепишемо останню нерівність у вигляді

$$\|x_{n+1} - z\|^2 + \frac{2\lambda L}{1-\lambda\mu}\,\|y_n - x_{n+1}\|^2 \le \\ \le (1 - \lambda\mu)\left(\|x_n - z\|^2 + \frac{2\lambda L}{1-\lambda\mu}\,\|y_{n-1} - x_n\|^2\right) - \\ - (1 - 2\lambda L - 2\lambda\mu)\,\|x_n - y_n\|^2 - \left(1 - 2\lambda L - \frac{2\lambda L}{1-\lambda\mu}\right)\|y_n - x_{n+1}\|^2 .$$

Для $\lambda = \frac{1}{4L}$ отримуємо

$$\|x_{n+1} - z\|^2 + \frac{2L}{4L-\mu}\,\|y_n - x_{n+1}\|^2 \le \\ \le \left(1 - \frac{\mu}{4L}\right)\left(\|x_n - z\|^2 + \frac{2L}{4L-\mu}\,\|y_{n-1} - x_n\|^2\right).$$

Таким чином, має місце

Теорема 4.4. *Нехай* $C \subseteq H$ *— непорожня опукла замкнена множина,* $A : H \to H$ *—* L*-ліпшицевий та* μ*-сильно монотонний на множині* C *оператор,* $z \in C$ *— єдиний розв'язок варіаціної нерівності (4.1). Тоді для породжених алгоритмом 4.2 послідовностей* (x_n)*,* (y_n) *виконується оцінка*

$$\|x_{n+1} - z\|^2 + \frac{1}{2}\,\|y_n - x_{n+1}\|^2 \le \left(1 - \frac{\mu}{4L}\right)^n \|x_1 - z\|^2, \quad n \ge 1.$$

Зауваження 4.2. З теореми 4.4 випливають оцінки

$$\|x_{n+1} - z\|^2 \le e^{-\frac{\mu}{4L}n}\,\|x_1 - z\|^2, \quad \|y_n - x_{n+1}\|^2 \le e^{-\frac{\mu}{4L}n} 2\,\|x_1 - z\|^2 .$$

Розглянемо ще один алгоритм з лінійною швидкістю збіжності для варіаціної нерівності (4.1) з L-ліпшицевим та μ-сильно монотонним на множині C оператором [103].

Алгоритм 4.3. Для $x_1 = x_0 \in C$ генеруємо послідовність точок $x_n \in C$ за допомогою ітераційної схеми

$$x_{n+1} = P_C\left(x_n - \tfrac{1}{2L}Ax_n - \tfrac{1}{2(L+\mu)}\left(Ax_n - Ax_{n-1}\right)\right).$$

Покажемо, що має місце оцінка

$$\|x_{n+1} - z\|^2 = O\left(\left(1 - \frac{\mu}{L+\mu}\right)^n\right).$$

Розглянемо схему

$$x_{n+1} = P_C\left(x_n - \alpha Ax_n - \beta\left(Ax_n - Ax_{n-1}\right)\right),$$

де $x_1 = x_0 \in C$, $\alpha > 0$, $\beta > 0$.

Нехай z — розв'язок варіаційної нерівності (4.1). Маємо

$$\begin{aligned}\|x_{n+1} - z\|^2 \le \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + \\ + 2\left(\alpha Ax_n + \beta\left(Ax_n - Ax_{n-1}\right), z - x_{n+1}\right).\end{aligned} \tag{4.29}$$

З умови (4.27) випливає

$$\begin{aligned}&\left(\alpha Ax_n + \beta\left(Ax_n - Ax_{n-1}\right), z - x_{n+1}\right) = \\ &\quad = \alpha\left(Ax_n - Ax_{n+1}, z - x_{n+1}\right) + \\ &\quad + \beta\left(Ax_n - Ax_{n-1}, z - x_{n+1}\right) + \alpha\left(Ax_{n+1}, z - x_{n+1}\right) \le \\ &\quad \le \alpha\left(Ax_n - Ax_{n+1}, z - x_{n+1}\right) + \beta\left(Ax_n - Ax_{n-1}, z - x_n\right) + \\ &\quad + \beta\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right) - \\ &\quad - \alpha\mu\|x_{n+1} - z\|^2.\end{aligned} \tag{4.30}$$

Застосуємо нерівність (4.30) для оцінки правої частини (4.29) та отримаємо таку нерівність

$$\begin{aligned}&(1 + 2\alpha\mu)\|x_{n+1} - z\|^2 \le \\ &\quad \le \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + \\ &\quad + 2\alpha\left(Ax_n - Ax_{n+1}, z - x_{n+1}\right) + 2\beta\left(Ax_n - Ax_{n-1}, z - x_n\right) + \\ &\quad + 2\beta\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right).\end{aligned} \tag{4.31}$$

Оцінимо зверху вираз $2\beta\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right)$ в (4.31). Маємо

$$
\begin{aligned}
2\beta\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right) \le& \\
&\le 2\beta\left\|Ax_n - Ax_{n-1}\right\|\left\|x_n - x_{n+1}\right\| \le \\
&\le 2\beta L\left\|x_n - x_{n-1}\right\|\left\|x_{n+1} - x_n\right\| \le \\
&leq \beta L\left\|x_n - x_{n-1}\right\|^2 + \beta L\left\|x_n - x_{n+1}\right\|^2.
\end{aligned}
$$

Отримаємо

$$
\begin{aligned}
(1+2\alpha\mu)\left\|x_{n+1} - z\right\|^2 + 2\alpha\left(Ax_n - Ax_{n+1}, x_{n+1} - z\right) + \alpha L\|x_{n+1} - x_n\|^2 \le& \\
\le \left\|x_n - z\right\|^2 + 2\beta\left(Ax_{n-1} - Ax_n, x_n - z\right) + \beta L\|x_n - x_{n-1}\|^2 -& \\
- (1 - \alpha L - \beta L)\left\|x_{n+1} - x_n\right\|^2&.
\end{aligned}
$$

Покладемо $\alpha = \frac{1}{2L}$:

$$
\begin{aligned}
\left(1+\frac{\mu}{L}\right)\left\|x_{n+1} - z\right\|^2 + \frac{1}{L}\left(Ax_n - Ax_{n+1}, x_{n+1} - z\right) + \frac{1}{2}\|x_{n+1} - x_n\|^2 =& \\
= \left(1+\frac{\mu}{L}\right)\Bigg(\left\|x_{n+1} - z\right\|^2 + \frac{1}{L\left(1+\frac{\mu}{L}\right)}\left(Ax_n - Ax_{n+1}, x_{n+1} - z\right) +& \\
+\frac{1}{2\left(1+\frac{\mu}{L}\right)}\|x_{n+1} - x_n\|^2\Bigg) \le& \\
\le \left\|x_n - z\right\|^2 + 2\beta\left(Ax_{n-1} - Ax_n, x_n - z\right) + \beta L\|x_n - x_{n-1}\|^2 -& \\
- \left(\frac{1}{2} - \beta L\right)\left\|x_{n+1} - x_n\right\|^2&.
\end{aligned}
$$

Покладемо $\beta = \frac{1}{2L\left(1+\frac{\mu}{L}\right)}$. Тоді для

$$
W_n = \left\|x_n - z\right\|^2 + \frac{1}{L\left(1+\frac{\mu}{L}\right)}\left(Ax_{n-1} - Ax_n, x_n - z\right) + \frac{1}{2\left(1+\frac{\mu}{L}\right)}\|x_n - x_{n-1}\|^2
$$

маємо

$$
\left(1+\frac{\mu}{L}\right)W_{n+1} \le W_n - \left(\frac{1}{2} - \frac{1}{2\left(1+\frac{\mu}{L}\right)}\right)\left\|x_{n+1} - x_n\right\|^2 \le W_n.
$$

Звідси

$$
W_{n+1} \le \left(1+\frac{\mu}{L}\right)^{-1} W_n \le \ldots \le \left(1+\frac{\mu}{L}\right)^{-n}\left\|x_1 - z\right\|^2. \qquad (4.32)
$$

Оскільки

$$
\begin{aligned}
W_n = \|x_n - z\|^2 + \\
+ \frac{1}{L\left(1+\frac{\mu}{L}\right)}\left(Ax_{n-1} - Ax_n, x_n - z\right) + \frac{1}{2\left(1+\frac{\mu}{L}\right)}\|x_n - x_{n-1}\|^2 \geq \\
\geq \|x_n - z\|^2 - \\
- \frac{1}{L\left(1+\frac{\mu}{L}\right)}\|Ax_{n-1} - Ax_n\|\,\|x_n - z\| + \frac{1}{2\left(1+\frac{\mu}{L}\right)}\|x_{n-1} - x_n\|^2 \geq \\
\geq \|x_n - z\|^2 - \\
- \frac{1}{\left(1+\frac{\mu}{L}\right)}\|x_n - x_{n-1}\|\,\|x_n - z\| + \frac{1}{2\left(1+\frac{\mu}{L}\right)}\|x_{n-1} - x_n\|^2 \geq \\
\geq \left(1 - \frac{1}{2\left(1+\frac{\mu}{L}\right)}\right)\|x_n - z\|^2 \geq \frac{1}{2}\|x_n - z\|^2,
\end{aligned}
$$

то з (4.32) випливає

$$\|x_{n+1} - z\|^2 \leq \left(1 - \frac{\mu}{L+\mu}\right)^n 2\|x_1 - z\|^2.$$

Таким чином, має місце

Теорема 4.5. *Нехай* H *— гільбертовий простір,* $C \subseteq H$ *— непорожня опукла замкнена множина,* $A : H \to H$ *—* L*-ліпшицевий та* μ*-сильно монотонний на множині* C *оператор,* $z \in C$ *— єдиний розв'язок варіаціної нерівності (4.1). Тоді для породженої алгоритмом 4.3 послідовності* (x_n) *виконується оцінка*

$$\|x_{n+1} - z\|^2 \leq \left(1 - \frac{\mu}{L+\mu}\right)^n 2\|x_1 - z\|^2, \quad n \geq 1.$$

Зауваження 4.3. З теореми 4.5 випливає оцінка

$$\|x_{n+1} - z\|^2 \leq e^{-\frac{\mu}{2L}n} 2\,\|x_1 - z\|^2.$$

Зауваження 4.4. Алгоритм 4.3 є варіантом методу операторної екстраполяції [104]. За об'ємом необхідних для здійснення ітераційного кроку обчислень він має перевагу над екстаградієнтним методом Корпелевич та методом екстраполяції з минулого. Під назвою «forward-reflected-backward algorithm» близький метод був запропонований в [105].

4.6. Адаптивний алгоритм екстраполяції з минулого

Для наближеного розв'язання задачі (4.1) розглянемо алгоритм екстраполяції з минулого з адаптивним вибором λ_n.

Алгоритм 4.4. Обираємо елементи $x_1 = y_0 \in C$, $\tau \in \left(0, \frac{1}{3}\right)$, $\lambda_1 \in (0, +\infty)$. Покладаємо $n = 1$.

1. Обчислити
$$y_n = P_C(x_n - \lambda_n A y_{n-1}).$$

2. Обчислити
$$x_{n+1} = P_C(x_n - \lambda_n A y_n).$$

Якщо $x_{n+1} = x_n = y_n$, то зупинити та x_n — розв'язок. Інакше перейти на крок **3**.

3. Обчислити
$$\lambda_{n+1} = \begin{cases} \lambda_n, & \text{якщо } (Ay_{n-1} - Ay_n, x_{n+1} - y_n) \le 0, \\ \min\left\{\lambda_n, \frac{\tau}{2} \frac{\|y_{n-1} - y_n\|^2 + \|x_{n+1} - y_n\|^2)}{(Ay_{n-1} - Ay_n, x_{n+1} - y_n)}\right\}, & \text{інакше.} \end{cases}$$

Покласти $n := n + 1$ та перейти на **1**.

Зауваження 4.5. Схема, що подібна алгоритму 4.4, запропонована для задач про рівновагу в просторах Адамара в [106].

Зауваження 4.6. Результати, аналогічні наведеним нижче, мають місце для модифікації алгоритму 4.4 з заміною інструкції перерахунку λ_n на таку
$$\lambda_{n+1} = \begin{cases} \min\left\{\lambda_n, \tau\frac{\|y_{n-1} - y_n\|}{\|Ay_{n-1} - Ay_n\|}\right\}, & \text{якщо } Ay_{n-1} \ne Ay_n, \\ \lambda_n, & \text{інакше.} \end{cases}$$
де $\tau \in \left(0, \frac{1}{3}\right)$ [101].

Параметр λ_{n+1} залежить від розташування точок y_{n-1}, y_n, x_{n+1}, значень Ay_{n-1}, Ay_n. Інформація про константу L не використовується. Очевидно, що послідовність (λ_n) незростаюча. Також вона обмежена знизу числом
$$\min\left\{\lambda_1, \frac{\tau}{L}\right\}.$$

Дійсно, маємо
$$(Ay_{n-1} - Ay_n, x_{n+1} - y_n) \le \frac{L}{2}\left(\|y_{n-1} - y_n\|^2 + \|y_n - x_{n+1}\|^2\right).$$

Для послідовностей (x_n), (y_n), породжених алгоритмом 4.4, мають місце нерівності

$$\begin{aligned}(Ay_{n-1}, y_n - y) &\leq \\ &\leq \frac{1}{2\lambda_n}\left(\|y - x_n\|^2 - \|x_n - y_n\|^2 - \|y_n - y\|^2\right) \quad \forall y \in C, \end{aligned} \tag{4.33}$$

$$\begin{aligned}(Ay_n, x_{n+1} - y) &\leq \\ &\leq \frac{1}{2\lambda_n}\left(\|y - x_n\|^2 - \|x_n - x_{n+1}\|^2 - \|x_{n+1} - y\|^2\right) \quad \forall y \in C. \end{aligned} \tag{4.34}$$

Зауваження 4.7. Зокрема, нерівність (4.34) дає обґрунтування правила зупинки алгоритму 4.4. Дійсно, при $x_{n+1} = x_n = y_n$ із (4.34) випливає

$$(Ay_n, y_n - y) \leq 0 \qquad \forall y \in C,$$

тобто $x_n = y_n \in S$.

Зауваження 4.8. Можна використовувати для зупинки алгоритму 4.4 правило $x_n = y_n = y_{n-1}$, яке гарантує $x_n \in S$.

Доведемо основну оцінку, яка пов'язує відстані між породженими алгоритмом 4.4 точками і довільним елементом множини розв'язків S.

Припустимо додатково, що оператор A є псевдомонотонним.

Лема 4.7. *Для послідовностей* (x_n), (y_n), *породжених алгоритмом 4.4, має місце нерівність*

$$\begin{aligned}\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \left(1 - \tau\frac{\lambda_n}{\lambda_{n+1}}\right)\|x_{n+1} - y_n\|^2 - \\ - \left(1 - 2\tau\frac{\lambda_n}{\lambda_{n+1}}\right)\|x_n - y_n\|^2 + 2\tau\frac{\lambda_n}{\lambda_{n+1}}\|x_n - y_{n-1}\|^2, \end{aligned} \tag{4.35}$$

де $z \in S$.

Доведення. Нехай $z \in S$. Із псевдомонотонності оператора A випливає

$$(Ay_n, z - y_n) \leq 0. \tag{4.36}$$

Із (4.36) і (4.34) випливає

$$2\lambda_n(Ay_n, x_{n+1} - y_n) \leq \|z - x_n\|^2 - \|x_n - x_{n+1}\|^2 - \|x_{n+1} - z\|^2. \tag{4.37}$$

Склавши нерівності (4.37) та нерівність, що випливає з (4.33)

$$2\lambda_n(Ay_{n-1}, y_n - x_{n+1}) \le \|x_{n+1} - x_n\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2$$

маємо

$$2\lambda_n(Ay_{n-1} - Ay_n, y_n - x_{n+1}) \le \|z - x_n\|^2 - \\ - \|z - x_{n+1}\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - y_n\|^2. \quad (4.38)$$

З правила обчислення λ_{n+1} випливає нерівність

$$(Ay_{n-1} - Ay_n, x_{n+1} - y_n) \le \\ \le \frac{\tau}{2\lambda_{n+1}} \left(\|y_{n-1} - y_n\|^2 + \|y_n - x_{n+1}\|^2 \right). \quad (4.39)$$

Для оцінки виразу $(Ay_{n-1} - Ay_n, y_n - x_{n+1})$ в (4.38) скористаємося (4.39). Отримаємо

$$\|z - x_{n+1}\|^2 \le \|z - x_n\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - y_n\|^2 + \\ + \tau \frac{\lambda_n}{\lambda_{n+1}} \left(\|y_{n-1} - y_n\|^2 + \|y_n - x_{n+1}\|^2 \right).$$

Оскільки

$$\|y_{n-1} - y_n\|^2 \le 2\|y_{n-1} - x_n\|^2 + 2\|y_n - x_n\|^2,$$

то

$$\|z - x_{n+1}\|^2 \le \|z - x_n\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - y_n\|^2 + \\ + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|y_{n-1} - x_n\|^2 + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|y_n - x_n\|^2 + \tau \frac{\lambda_n}{\lambda_{n+1}} \|y_n - x_{n+1}\|^2,$$

що і треба було довести. □

Сформулюємо основний результат.

Теорема 4.6. *Нехай* H *— гільбертовий простір,* $C \subseteq H$ *— непорожня опукла замкнена множина,* $A : H \to H$ *— монотонний та* L*-ліпшицевий на множині* C *оператор та* $S \neq \emptyset$*. Тоді породжені алгоритмом 4.4 послідовності* (x_n)*,* (y_n) *слабко збігаються до розв'язку* $\bar{z} \in S$ *варіаційної нерівності (4.1), причому*

$$\lim_{n\to\infty} \|x_n - y_n\| = \lim_{n\to\infty} \|x_{n+1} - y_n\| = 0.$$

Доведення. Нехай $z' \in S$. Покладемо

$$\alpha_n = \|x_n - z'\|^2 + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|x_n - y_{n-1}\|^2,$$

$$\beta_n = \left(1 - 2\tau \frac{\lambda_n}{\lambda_{n+1}}\right) \|y_n - x_n\|^2 + \left(1 - 2\tau \frac{\lambda_{n+1}}{\lambda_{n+2}} - \tau \frac{\lambda_n}{\lambda_{n+1}}\right) \|x_{n+1} - y_n\|^2.$$

Нерівність (4.35) набуває вигляду

$$\alpha_{n+1} \le \alpha_n - \beta_n.$$

Оскільки існує $\lim_{n\to\infty} \lambda_n > 0$, то

$$1 - 2\tau \frac{\lambda_n}{\lambda_{n+1}} \to 1 - 2\tau \in (0,1) \quad \text{і} \quad 1 - 2\tau \frac{\lambda_{n+1}}{\lambda_{n+2}} - \tau \frac{\lambda_n}{\lambda_{n+1}} \to 1 - 3\tau \in (0,1)$$

при $n \to \infty$. З леми 4.2 можемо зробити висновок, що існує границя

$$\lim_{n\to\infty} \left(\|x_n - z'\|^2 + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|x_n - y_{n-1}\|^2 \right)$$

та збігається числовий ряд

$$\sum_n \left(\left(1 - 2\tau \frac{\lambda_n}{\lambda_{n+1}}\right) \|y_n - x_n\|^2 + \left(1 - 2\tau \frac{\lambda_{n+1}}{\lambda_{n+2}} - \tau \frac{\lambda_n}{\lambda_{n+1}}\right) \|x_{n+1} - y_n\|^2 \right).$$

Звідки отримуємо

$$\lim_{n\to\infty} \|y_n - x_n\| = \lim_{n\to\infty} \|x_{n+1} - y_n\| = \lim_{n\to\infty} \|x_{n+1} - x_n\| = 0 \tag{4.40}$$

і збіжність числових послідовностей $\left(\|x_n - z'\|\right)$, $\left(\|y_n - z'\|\right)$ для всіх $z' \in S$. Зокрема, послідовності (x_n), (y_n) обмежені.

Розглянемо підпослідовність (x_{n_k}), слабко збіжну до деякої точки $\bar{z} \in C$. Тоді з (4.40) випливає, що й $y_{n_k} \rightharpoonup \bar{z}$.

Покажемо, що $\bar{z} \in S$. Маємо

$$(Ay_n, y - y_n) \ge (Ay_n, x_{n+1} - y_n) + \frac{(x_{n+1} - x_n, x_{n+1} - y)}{\lambda_n} \quad \forall y \in C. \tag{4.41}$$

З (4.40) випливає

$$\lim_{n\to\infty} (Ay_n, x_{n+1} - y_n) = 0. \tag{4.42}$$

Здійснивши граничний перехід в (4.41) з урахуванням (4.40), (4.42) та умови монотонності оператора A, отримаємо

$$(Ay, y - \bar{z}) = \lim_{k\to\infty}(Ay, y - y_{n_k}) \geq \limsup_{k\to\infty}(Ay_{n_k}, y - y_{n_k}) \geq$$
$$\geq \lim_{k\to\infty}\left\{(Ay_{n_k}, x_{n_k+1} - y_{n_k}) + \frac{(x_{n_k+1} - x_{n_k}, x_{n_k+1} - y)}{\lambda_{n_k}}\right\} = 0 \quad \forall y \in C,$$

тобто, $\bar{z} \in S$.

Покажемо тепер, що $x_n \rightharpoonup \bar{z} \in S$. Тоді з $x_n - y_n \to 0$ випливає, що й послідовність (y_n) слабко збігається $\bar{z}$. Міркуємо від супротивного. Нехай існує така підпослідовність (x_{m_k}), що $x_{m_k} \rightharpoonup \tilde{z}$ та $\tilde{z} \neq \bar{z}$. Ясно, що $\tilde{z} \in S$. Застосуємо двічі лему Опяла. Маємо

$$\lim_{n\to\infty}\|x_n - \bar{z}\| = \lim_{k\to\infty}\|x_{n_k} - \bar{z}\| < \lim_{k\to\infty}\|x_{n_k} - \tilde{z}\| = \lim_{n\to\infty}\|x_n - \tilde{z}\| =$$
$$= \lim_{k\to\infty}\|x_{m_k} - \tilde{z}\| < \lim_{k\to\infty}\|x_{m_k} - \bar{z}\| = \lim_{n\to\infty}\|x_n - \bar{z}\|,$$

що неможливо. Отже, $x_n \rightharpoonup \bar{z}$. □

4.7. Регуляризований алгоритм

Для непорожньої опуклої замкненої множини $C \subseteq H$, монотонного ліпшицевого оператора $A : H \to H$ та $z \in H$ розглянемо задачу:

$$\text{знайти} \;\; x^* = P_S z, \tag{4.43}$$

де $S = \{x \in C : \; (Ax, y - x) \geq 0 \;\; \forall y \in C\}$.

Будемо звичайно припускати, що $S \neq \emptyset$. Задача (4.43) має єдиний розв'язок [44]. Відомим окремим випадком (4.43) є задача пошуку нормального розв'язку варіаційної нерівності (при $z = 0$).

Метод регуляризації Тихонова полягає в апроксимації задачі (4.43) варіаційною нерівністю:

$$\text{знайти} \;\; x \in C : \quad (Ax, y - x) + \varepsilon(x - z, y - x) \geq 0 \quad \forall y \in C, \tag{4.44}$$

де $\varepsilon > 0$.

Відомо, що існує єдиний розв'язок $x_\varepsilon \in C$ задачі (4.44) для довільного $\varepsilon > 0$. А при прямуванні додатного параметру ε до нуля елементи x_ε сильно збігаються до розв'язку задачі (4.43) [107], тобто

$$\lim_{\varepsilon\to 0}\|x_\varepsilon - x^*\| = 0.$$

Відштовхуючись від алгоритму 4.1, для розв'язання задачі (4.43) пропонуємо такий

Алгоритм 4.5. Для $z \in H$, $x_1 = y_0 \in C$ генерируємо послідовність елементів x_n, $y_n \in C$ за допомогою ітераційної схеми

$$\begin{cases} y_n = P_C\left(\alpha_n\lambda_n z + (1-\alpha_n\lambda_n)x_n - \lambda_n A y_{n-1}\right), \\ x_{n+1} = P_C\left(\alpha_n\lambda_n z + (1-\alpha_n\lambda_n)x_n - \lambda_n A y_n\right), \end{cases}$$

де $\lambda_n > 0$, $\alpha_n > 0$.

Зауваження 4.9. Даний метод при $z = 0$ розглядався в [108].

Відносно параметрів алгоритму 4.5 будемо припускати, що виконані такі умови:

(A1) $\lambda_n \in \left[\underline{\lambda}, \overline{\lambda}\right] \subseteq \left(0, \frac{1}{3L}\right)$, де $L > 0$ — стала Ліпшиця оператора A;

(A2) $\lim_{n\to\infty} \alpha_n = 0$;

(A3) $\sum_{n=1}^{\infty} \alpha_n = +\infty$;

(A4) $\lim_{n\to\infty} \frac{\alpha_{n+1}-\alpha_n}{\alpha_n^2} = 0$.

Алгоритм 4.5 є результатом застосування до методу екстраполяції з минулого (алгоритм 4.1) схеми ітеративної регуляризації [81]. Доведення його сильної збіжності проводиться за такою схемою. Нехай x_{α_n} — розв'язок задачі (4.44) при $\varepsilon = \alpha_n$. Оскільки

$$\|x_n - x^*\| \le \|x_n - x_{\alpha_n}\| + \|x_{\alpha_n} - x^*\|, \quad \lim_{n\to\infty} \|x_{\alpha_n} - x^*\| = 0,$$

то достатньо показати, що породжена алгоритмом 4.5 послідовність (x_n) має властивість

$$\lim_{n\to\infty} \|x_n - x_{\alpha_n}\| = 0.$$

Для загальнішої задачі отримаємо подібні результати в розділі 6. З результатів розділу 6 випливає такий результат.

Теорема 4.7. *Нехай $A : H \to H$ — монотонний та L-ліпшицевий на множині C оператор та $S \neq \emptyset$. Нехай виконуються умови (A1)–(A4). Тоді для породжених алгоритмом 4.5 послідовностей (x_n), (y_n) має місце*

$$\lim_{n\to\infty} \|x_n - x^*\| = \lim_{n\to\infty} \|y_n - x^*\| = 0, \tag{4.45}$$

де $x^* \in H$ — единий розв'язок задачі (4.43).

Побудуємо для схеми

$$x_{n+1} = P_C\left(x_n - \alpha A x_n - \beta\left(A x_n - A x_{n-1}\right)\right)$$

сильно збіжний варіант. Використовуємо метод Гальперна апроксимації нерухомих точок нерозтягуючих операторів.

Алгоритм 4.6. Задаємо елементи $y \in H$, $x_0 = x_1 \in C$, послідовність додатніх чисел (λ_n) та таку послідовність (α_n), що

$$\alpha_n \in (0,1)\,, \quad \lim_{n\to\infty} \alpha_n = 0, \quad \sum_{n=1}^{\infty} \alpha_n = +\infty.$$

Ітерації. Генеруємо послідовність (x_n) за допомогою ітераційної схеми

$$x_{n+1} = P_C\left(\alpha_n y + (1-\alpha_n)\, x_n - \lambda_n A x_n - (1-\alpha_n)\,\lambda_{n-1}\left(A x_n - A x_{n-1}\right)\right).$$

Відносно додатніх параметрів λ_n припустимо виконання такої умови:

$$0 < \inf_n \lambda_n \le \sup_n \lambda_n < \tfrac{1}{2L}. \tag{4.46}$$

Має місце

Лема 4.8. *Для послідовності (x_n), що породжена алгоритмом 4.6, виконується нерівність*

$$\begin{aligned} &\|x_{n+1} - z\|^2 + 2\lambda_n\left(A x_n - A x_{n+1}, x_{n+1} - z\right) + \tfrac{1}{2}\|x_{n+1} - x_n\|^2 \le \\ &\le (1-\alpha_n)\left(\|x_n - z\|^2 + 2\lambda_{n-1}\left(A x_{n-1} - A x_n, x_n - z\right) + \tfrac{1}{2}\|x_n - x_{n-1}\|^2\right) + \\ &\qquad + \alpha_n\|y - z\|^2 - \alpha_n\|y - x_{n+1}\|^2 - \\ &\qquad - \left(\tfrac{1}{2} - \alpha_n - (1-\alpha_n)\,\lambda_{n-1} L\right)\|x_{n+1} - x_n\|^2 - \\ &\qquad\qquad - (1-\alpha_n)\left(\tfrac{1}{2} - \lambda_{n-1} L\right)\|x_n - x_{n-1}\|^2, \end{aligned} \tag{4.47}$$

де $z \in S$.

Доведення. Нехай $z \in S$. Маємо

$$\begin{aligned} &\left(x_{n+1} - \alpha_n y - (1-\alpha_n)\, x_n + \lambda_n A x_n + \right. \\ &\qquad \left. + (1-\alpha_n)\,\lambda_{n-1}\left(A x_n - A x_{n-1}\right), z - x_{n+1}\right) \ge 0. \end{aligned} \tag{4.48}$$

Монотонність оператора A та включення $z \in S$ дає

$$
\begin{aligned}
&(\lambda_n A x_n + (1-\alpha_n)\lambda_{n-1}(Ax_n - Ax_{n-1}), z - x_{n+1}) = \\
&\qquad = \lambda_n (Ax_n - Ax_{n+1}, z - x_{n+1}) + \\
&\qquad + (1-\alpha_n)\lambda_{n-1}(Ax_n - Ax_{n-1}, z - x_{n+1}) + \underbrace{\lambda_n (Ax_{n+1}, z - x_{n+1})}_{\le 0} \le \\
&\le \lambda_n (Ax_n - Ax_{n+1}, z - x_{n+1}) + (1-\alpha_n)\lambda_{n-1}(Ax_n - Ax_{n-1}, z - x_n) + \\
&\qquad + (1-\alpha_n)\lambda_{n-1}(Ax_n - Ax_{n-1}, x_n - x_{n+1}). \qquad (4.49)
\end{aligned}
$$

Використавши (4.49) в (4.48), отримаємо

$$
\begin{aligned}
0 \le\ & 2(x_{n+1} - \alpha_n y - (1-\alpha_n)x_n, z - x_{n+1}) + \\
&+ 2\lambda_n (Ax_n - Ax_{n+1}, z - x_{n+1}) + \\
&+ 2(1-\alpha_n)\lambda_{n-1}(Ax_n - Ax_{n-1}, z - x_n) + \\
&+ 2(1-\alpha_n)\lambda_{n-1}(Ax_n - Ax_{n-1}, x_n - x_{n+1}). \qquad (4.50)
\end{aligned}
$$

Оцінимо зверху доданок $2\lambda_{n-1}(Ax_n - Ax_{n-1}, x_n - x_{n+1})$ в (4.50). Маємо

$$
\begin{aligned}
2\lambda_{n-1}(Ax_n - Ax_{n-1}, x_n - x_{n+1}) &\le 2\lambda_{n-1}\|Ax_n - Ax_{n-1}\| \cdot \|x_n - x_{n+1}\| \le \\
&\le 2\lambda_{n-1} L \|x_n - x_{n-1}\| \cdot \|x_{n+1} - x_n\| \le \\
&\le \lambda_{n-1} L \|x_n - x_{n-1}\|^2 + \lambda_{n-1} L \|x_n - x_{n+1}\|^2.
\end{aligned}
$$

Перетворимо доданок $2(x_{n+1} - \alpha_n y - (1-\alpha_n)x_n, z - x_{n+1})$ в (4.50). Маємо

$$
\begin{aligned}
2(x_{n+1} - \alpha_n y - (1-\alpha_n)x_n, z - x_{n+1}) &= \|\alpha_n y + (1-\alpha_n)x_n - z\|^2 - \\
&- \|x_{n+1} - z\|^2 - \|x_{n+1} - \alpha_n y - (1-\alpha_n)x_n\|^2. \qquad (4.51)
\end{aligned}
$$

Для перетворення різниці

$$
\|\alpha_n y + (1-\alpha_n)x_n - z\|^2 - \|\alpha_n y + (1-\alpha_n)x_n - x_{n+1}\|^2
$$

в (4.51) використаємо таку тотожність

$$
\begin{aligned}
\|\alpha u + (1-\alpha)v - w\|^2 &= \|v - w - \alpha(v - u)\|^2 = \\
&= \|v - w\|^2 - 2\alpha(v - w, v - u) + \alpha^2\|v - u\|^2 = \\
&= \|v - w\|^2 - \alpha\|v - u\|^2 - \alpha\|v - w\|^2 + \alpha\|u - w\|^2 + \alpha^2\|v - u\|^2,
\end{aligned}
$$

де u, v, $w \in H$, $\alpha > 0$.

Отримаємо рівність

$$\begin{aligned}
&\|\alpha_n y + (1-\alpha_n)x_n - z\|^2 - \|\alpha_n y + (1-\alpha_n)x_n - x_{n+1}\|^2 = \\
&= \|x_n - z\|^2 - \alpha_n\|x_n - y\|^2 - \alpha_n\|x_n - z\|^2 + \\
&+ \alpha_n\|y - z\|^2 + \alpha_n^2\|x_n - y\|^2 - \\
&- \|x_n - x_{n+1}\|^2 + \alpha_n\|x_n - y\|^2 + \alpha_n\|x_n - x_{n+1}\|^2 - \\
&- \alpha_n\|y - x_{n+1}\|^2 - \alpha_n^2\|x_n - y\|^2 = \\
&= (1-\alpha_n)\|x_n - z\|^2 - (1-\alpha_n)\|x_n - x_{n+1}\|^2 + \\
&+ \alpha_n\|y - z\|^2 - \alpha_n\|y - x_{n+1}\|^2.
\end{aligned}$$

Прийшли до нерівності

$$\begin{aligned}
&0 \le (1-\alpha_n)\|x_n - z\|^2 - (1-\alpha_n)\|x_n - x_{n+1}\|^2 + \alpha_n\|y - z\|^2 - \\
&- \alpha_n\|y - x_{n+1}\|^2 - \|x_{n+1} - z\|^2 + 2\lambda_n(Ax_n - Ax_{n+1}, z - x_{n+1}) + \\
&+ 2(1-\alpha_n)\lambda_{n-1}(Ax_n - Ax_{n-1}, z - x_n) + \\
&+ (1-\alpha_n)\lambda_{n-1}L\|x_n - x_{n-1}\|^2 + (1-\alpha_n)\lambda_{n-1}L\|x_n - x_{n+1}\|^2.
\end{aligned} \tag{4.52}$$

Перегрупуємо члени в (4.52) та отримаємо

$$\begin{aligned}
&\|x_{n+1} - z\|^2 + 2\lambda_n(Ax_n - Ax_{n+1}, x_{n+1} - z) + \tfrac{1}{2}\|x_{n+1} - x_n\|^2 \le \\
&\le (1-\alpha_n)\left(\|x_n - z\|^2 + 2\lambda_{n-1}(Ax_{n-1} - Ax_n, x_n - z) + \tfrac{1}{2}\|x_n - x_{n-1}\|^2\right) + \\
&+ \alpha_n\|y - z\|^2 - \alpha_n\|y - x_{n+1}\|^2 - \left(\tfrac{1}{2} - \alpha_n - (1-\alpha_n)\lambda_{n-1}L\right)\|x_{n+1} - x_n\|^2 - \\
&- (1-\alpha_n)\left(\tfrac{1}{2} - \lambda_{n-1}L\right)\|x_n - x_{n-1}\|^2,
\end{aligned}$$

що й потрібно було довести. □

Лема 4.9. *Для послідовності* (x_n)*, що породжена алгоритмом 4.6, виконується нерівність*

$$\begin{aligned}
&\|x_{n+1} - z\|^2 + 2\lambda_n(Ax_n - Ax_{n+1}, x_{n+1} - z) + \frac{1}{2}\|x_{n+1} - x_n\|^2 \le \\
&\le (1-\alpha_n)\left(\|x_n - z\|^2 + 2\lambda_{n-1}(Ax_{n-1} - Ax_n, x_n - z) + \frac{1}{2}\|x_n - x_{n-1}\|^2\right) + \\
&+ 2\alpha_n(y - z, x_{n+1} - z) - \left(\tfrac{1}{2} - \alpha_n - (1-\alpha_n)\lambda_{n-1}L\right)\|x_{n+1} - x_n\|^2 - \\
&- (1-\alpha_n)\left(\tfrac{1}{2} - \lambda_{n-1}L\right)\|x_n - x_{n-1}\|^2, \quad z \in S.
\end{aligned} \tag{4.53}$$

Доведення. Застосуємо елементарну нерівність

$$\|a+b\|^2 \le \|a\|^2 + 2\,(b, a+b)\,.$$

Маємо

$$\|y-z\|^2 = \|y-x_{n+1}+x_{n+1}-z\|^2 \le \\ \le \|y-x_{n+1}\|^2 + 2\,(y-z, x_{n+1}-z)\,. \quad (4.54)$$

Використавши (4.54) в нерівності (4.47), отримаємо (4.53), що і потрібно було довести. □

Доведемо обмеженість послідовності (x_n).

Лема 4.10. *Нехай виконана умова* (4.46). *Тоді породжена ітераційним алгоритмом 4.6 послідовність* (x_n) *є обмеженою.*

Доведення. Оскільки $0 < \inf_n \lambda_n \le \sup_n \lambda_n < \frac{1}{2L}$, $\lim\limits_{n\to\infty} \alpha_n = 0$, то існує такий номер $n_0 \ge 1$, що

$$\tfrac{1}{2} - \alpha_n - (1-\alpha_n)\,\lambda_{n-1}L = \tfrac{1}{2} - \lambda_{n-1}L - \alpha_n\,(1-\lambda_{n-1}L) > 0 \quad (4.55)$$

та

$$(1-\alpha_n)\left(\tfrac{1}{2} - \lambda_{n-1}L\right) > 0. \quad (4.56)$$

З (4.47) та (4.55), (4.56) випливає, що для $n \ge n_0$ виконується нерівність

$$\xi_{n+1} \le (1-\alpha_n)\,\xi_n + \alpha_n\,\|y-z\|^2\,, \quad (4.57)$$

де

$$\xi_n = \|x_n - z\|^2 + 2\lambda_{n-1}\,(Ax_{n-1} - Ax_n, x_n - z) + \tfrac{1}{2}\,\|x_n - x_{n-1}\|^2\,, \quad z \in S.$$

Оцінимо знизу ξ_n. Маємо

$$\xi_n = \|x_n - z\|^2 + 2\lambda_{n-1}\,(Ax_{n-1} - Ax_n, x_n - z) + \tfrac{1}{2}\,\|x_n - x_{n-1}\|^2 \ge \\ \ge \|x_n - z\|^2 - 2\lambda_{n-1}\,\|Ax_{n-1} - Ax_n\|\,\|x_n - z\| + \tfrac{1}{2}\,\|x_{n-1} - x_n\|^2 \ge \\ \ge \|x_n - z\|^2 - 2\lambda_{n-1}L\,\|x_{n-1} - x_n\|\,\|x_n - z\| + \tfrac{1}{2}\,\|x_{n-1} - x_n\|^2 \ge \\ \ge (1-\lambda_{n-1}L)\,\|x_n - z\|^2 + \left(\tfrac{1}{2} - \lambda_{n-1}L\right)\|x_n - x_{n-1}\|^2 \ge 0. \quad (4.58)$$

З нерівностей (4.57), (4.58) випливає обмеженість послідовностей (ξ_n) та (x_n), що і потрібно було довести. □

Сформулюємо основний результат.

Теорема 4.8. *Нехай* C *— опукла та замкнена підмножина гільбертового простору* H*,* $A : H \to H$ *— монотонний та ліпшицевий на множині* C *оператор,* $S \neq \emptyset$*,* $y \in H$*, виконується умова (4.46). Тоді послідовність* (x_n)*, що породжена алгоритмом 4.6, сильно збігається до точки* $z = P_S y$*.*

Доведення. Нехай $z = P_S y$. З леми 4.10 випливає існування такого числа $M \geq 0$, що

$$|(y - z, x_{n+1} - z)| \leq M$$

для всіх $n \geq 1$. Тоді з леми 4.9 випливає нерівність

$$\xi_{n+1} - \xi_n + \alpha_n \xi_n + \left(\tfrac{1}{2} - \alpha_n - (1 - \alpha_n)\lambda_{n-1} L\right) \|x_{n+1} - x_n\|^2 + \\ + (1 - \alpha_n)\left(\tfrac{1}{2} - \lambda_{n-1} L\right) \|x_n - x_{n-1}\|^2 \leq 2\alpha_n M, \quad (4.59)$$

де

$$\xi_n = \|x_n - z\|^2 + 2\lambda_{n-1}(Ax_{n-1} - Ax_n, x_n - z) + \frac{1}{2}\|x_n - x_{n-1}\|^2.$$

Розглянемо числову послідовність (ξ_n). Можливі два варіанти:

a) існує номер $\bar{n} \geq 1$ такий, що $\xi_{n+1} \leq \xi_n$ для всіх $n \geq \bar{n}$;

b) існує зростаюча послідовність номерів (n_k) така, що $\xi_{n_k+1} > \xi_{n_k}$ для всіх $k \geq 1$.

Спочатку розглянемо варіант a). В цьому випадку існує $\lim\limits_{n\to\infty} \xi_n$. Оскільки

$$\xi_{n+1} - \xi_n \to 0, \quad \alpha_n \to 0$$

та виконуються нерівності (4.59), то при $n \to \infty$ маємо

$$\|x_{n+1} - x_n\| \to 0. \quad (4.60)$$

Покажемо, що всі слабкі часткові границі послідовності (x_n) належать допустимій множині S.

Розглянемо підпослідовність (x_{n_k}), що слабко збігається до деякої точки $w \in H$. Ясно, що $w \in C$. Покажемо, що $w \in S$. Маємо

$$(x_{n_k+1} - \alpha_{n_k} y - (1 - \alpha_{n_k}) x_{n_k} + \lambda_{n_k} A x_{n_k} + \\ + (1 - \alpha_{n_k}) \lambda_{n_k-1} (A x_{n_k} - A x_{n_k-1}), y - x_{n_k+1}) \geq 0 \quad \forall y \in C.$$

Звідки, використовуючи монотонність оператора A, виводимо оцінку

$$\begin{aligned}(Ay, y - x_{n_k}) + (Ax_{n_k}, x_{n_k} - x_{n_k+1}) \geq (Ax_{n_k}, y - x_{n_k+1}) \geq \\ \geq \frac{1}{\lambda_{n_k}} \left(\alpha_{n_k}\left(y - x_{n_k}\right) + x_{n_k} - x_{n_k+1}, y - x_{n_k+1}\right) - \\ - \left(1 - \alpha_{n_k}\right) \frac{\lambda_{n_k-1}}{\lambda_{n_k}} \left(Ax_{n_k} - Ax_{n_k-1}, y - x_{n_k+1}\right) \quad \forall y \in C.\end{aligned}$$

З $\lim\limits_{n\to\infty} \alpha_n = 0$, обмеженості послідовності (x_n), (4.60) та ліпшицевості оператора A випливає

$$\liminf_{k\to\infty} (Ay, y - x_{n_k}) \geq 0 \quad \forall y \in C.$$

З іншого боку

$$(Ay, y - w) = \lim_{k\to\infty} (Ay, y - x_{n_k}) = \liminf_{k\to\infty} (Ay, y - x_{n_k}) \geq 0 \quad \forall y \in C.$$

Отже, $w \in S$.

Доведемо, що

$$\limsup_{n\to\infty} (y - z, x_{n+1} - z) \leq 0. \tag{4.61}$$

Розглянемо таку підпослідовність (x_{n_k}), що

$$\lim_{k\to\infty} (y - z, x_{n_k} - z) = \limsup_{n\to\infty} (y - z, x_{n+1} - z).$$

Можна вважати, що $x_{n_k} \rightharpoonup w \in S$. Тоді отримуємо

$$\lim_{k\to\infty} (y - z, x_{n_k} - z) = (y - z, w - z) = (y - P_S y, w - P_S y) \leq 0,$$

чим доводимо (4.61).

Тепер з (4.61), нерівності

$$\xi_{n+1} \leq (1 - \alpha_n)\, \xi_n + 2\alpha_n (y - z, x_{n+1} - z),$$

що виконується для достатньо великих n, та леми 4.3 робимо висновок, що

$$\xi_n = \|x_n - z\|^2 + 2\lambda_{n-1} (Ax_{n-1} - Ax_n, x_n - z) + \tfrac{1}{2} \|x_n - x_{n-1}\|^2 \to 0.$$

З (4.58) отримуємо

$$\lim_{n\to\infty} \|x_n - z\| = 0.$$

Розглянемо варіант b). В цьому випадку існує неспадна послідовність номерів (m_k) з властивостями (лема 4.4):

i) $m_k \to +\infty$;

ii) $\xi_{m_k+1} \geq \xi_{m_k}$ для всіх $k \geq n_1$;

iii) $\xi_{m_k+1} \geq \xi_k$ для всіх $k \geq n_1$.

З нерівності леми 4.9, (4.58) та ii) випливає

$$\left(\tfrac{1}{2} - \alpha_{m_k} - (1-\alpha_{m_k})\,\lambda_{m_k-1}L\right)\|x_{m_k+1} - x_{m_k}\|^2 + \\ + (1-\alpha_{m_k})\left(\tfrac{1}{2} - \lambda_{m_k-1}L\right)\|x_{m_k} - x_{m_k-1}\|^2 \leq 2\alpha_{m_k}M.$$

Звідки

$$\lim_{k\to\infty}\|x_{m_k+1} - x_{m_k}\| = \lim_{k\to\infty}\|x_{m_k} - x_{m_k-1}\| = 0.$$

Міркуваннями, що подібні вищевикладеним, доводимо, що слабкі часткові границі послідовності (x_{m_k}) належать множині S. З тотожності

$$(y - z, x_{m_k+1} - z) = (y - z, x_{m_k} - z) + (y - z, x_{m_k+1} - x_{m_k})$$

отримуємо

$$\limsup_{k\to\infty}(y - z, x_{m_k+1} - z) = \limsup_{k\to\infty}(y - z, x_{m_k} - z)\,.$$

Як і раніше отримуємо нерівність

$$\limsup_{k\to\infty}(y - z, x_{m_k+1} - z) \leq 0.$$

Далі маємо

$$\xi_{m_k+1} \leq (1-\alpha_{m_k})\,\xi_{m_k} + 2\alpha_{m_k}(y - z, x_{m_k+1} - z) \leq \\ \leq (1-\alpha_{m_k})\,\xi_{m_k+1} + 2\alpha_{m_k}(y - z, x_{m_k+1} - z)\,.$$

Звідки, ураховуючи умову iii), отримуємо

$$\xi_k \leq \xi_{m_k+1} \leq 2\,(y - z, x_{m_k+1} - z)\,.$$

Таким чином,

$$\limsup_{k\to\infty}\xi_k \leq 2\limsup_{k\to\infty}(y - z, x_{m_k+1} - z) \leq 0.$$

Отже, $\lim_{n\to\infty} \xi_n = 0$ та, в свою чергу, $\lim_{n\to\infty} \|x_n - z\| = 0$, що і потрібно було довести. $\square$

Зауваження 4.10. Параметри λ_n алгоритму 4.6 задавалась виходячи з умови (4.46). Тобто, явно використовувалась інформація про константи ліпшицевості оператора A. Схема з статті [104] дозволяє побудувати модифікацію алгоритму 4.6 з адаптивним вибором величини λ_n, що не вимагає знання ліпшицевих констант операторів та процедур типу лінійного пошуку:

$$\begin{cases} x_{n+1} = P_C\left(\alpha_n y + (1-\alpha_n)\, x_n - \lambda_n A x_n - (1-\alpha_n)\, \lambda_{n-1}\left(A x_n - A x_{n-1}\right)\right), \\ \lambda_{n+1} = \begin{cases} \min\left\{\lambda_n, \tau \frac{\|x_{n+1}-x_n\|}{\|A x_{n+1} - A x_n\|}\right\}, & \text{якщо } A x_{n+1} \neq A x_n, \\ \lambda_n, & \text{інакше}, \end{cases} \end{cases}$$

де $\tau \in \left(0, \frac{1}{2}\right)$.

Розділ 5

Двоетапний проксимальний алгоритм для задач про рівновагу

У цьому розділі розглядаються задачі про рівновагу з псевдомонотонними біфункціями в гільбертовому просторі. Для цих задач запропоновано та обґрунтовано новий двоетапний проксимальний алгоритм, який є узагальненням методу Л.Д. Попова пошуку сідлових точок опукло-угнутих функцій (метод екстраполяції з минулого).

Розділ побудовано наступним чином. В підрозілі 5.1 сформульовано загальну задачу про рівновагу (задачу рівноважного програмування) в дійсному гільбертовому просторі, наведено основні припущення та двоетапний проксимальний алгоритм. В підрозділі 5.2 ми доводимо основну нерівність на якій грунтується доведення збіжності алгоритму. Підрозділ 5.3 містить теорему про слабку збіжність двоетапного проксимального алгоритму.

5.1. Постановка задачі про рівновагу та опис алгоритму

Всюди далі H — дійсний гільбертовий простір з скалярним добутком $(\cdot,\cdot)$ та породженою нормою $\|\cdot\|$.

Нехай $g : H \to \mathbb{R} \cup \{+\infty\}$ — власна опукла напівнеперервна знизу функція. Нагадаємо, що проксимальним оператором [54], асоційованим з функцією g, називають оператор

$$H \ni x \mapsto \mathrm{prox}_g x = \mathrm{argmin}_{y \in \mathrm{dom}\, g} \left(g(y) + \frac{1}{2} \|y - x\|^2 \right) \in \mathrm{dom}\, g.$$

Оператор prox_g — міцно нерозтягуючий (firmly nonexpansive), тобто,

$$\|\mathrm{prox}_g x - \mathrm{prox}_g y\|^2 \le \|x - y\|^2 - \|(x - \mathrm{prox}_g x) - (y - \mathrm{prox}_g y)\|^2 \;\; \forall\, x, y \in H,$$

та

$$g(y) - g(z) + (z - x, y - z) \geq 0 \ \ \forall y \in \operatorname{dom} g \ \Leftrightarrow \ \ z = \operatorname{prox}_g x.$$

Зокрема,

$$F(\operatorname{prox}_g) = \operatorname{argmin}_{y \in \operatorname{dom} g} g(y),$$

де $F(\operatorname{prox}_g) = \left\{x \in H : \ x = \operatorname{prox}_g x\right\}$.

Для непорожньої опуклої замкненої множини $C \subseteq H$ та біфункції $F : C \times C \to \mathbb{R}$ розглянемо задачу про рівновагу [109]:

$$\text{знайти} \ \ x \in C : \ F(x, y) \geq 0 \ \ \forall y \in C. \tag{5.1}$$

Задача (5.1) — зручна загальна форма запису та дослідження різних задач, що виникають в математичній фізиці, дослідженні операцій та оптимізації. Наведемо ряд типових постановок.

(1) Якщо $F(x, y) = g(y) - g(x)$, де $g : C \to \mathbb{R}$, то задача (5.1) є задачею умовної мінімізації:

$$g \to \min_C .$$

(2) Якщо $F(x, y) = (Ax, y - x)$, де $A : C \to H$, то задача (5.1) зводиться до класичної варіаційної нерівності:

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

(3) Нехай C_1, C_2 — опуклі підмножини H, $C = C_1 \times C_2$, $L : C \to \mathbb{R}$ — опукла-угнута функція. Точка $(x_1, x_2) \in C$ називається сідловою точкою функції L, якщо

$$L(x_1, y_2) \leq L(x_1, x_2) \leq L(y_1, x_2) \quad \forall (y_1, y_2) \in C. \tag{5.2}$$

Покладемо $F(x, y) = L(y_1, x_2) - L(x_1, y_2)$, де $x = (x_1, x_2)$, $y = (y_1, y_2)$. Тоді задача пошуку сідлової точки (5.2) рівносильна задачі про рівновагу вигляду (5.1):

$$\text{знайти} \ \ (x_1, x_2) \in C : \ L(y_1, x_2) - L(x_1, y_2) \geq 0 \quad \forall (y_1, y_2) \in C.$$

(4) Нехай I — скінченна множина індексів. Для кожного $i \in I$ задано множину C_i та функцію $f_i : C \to \mathbb{R}$, де $C = \prod_{i \in I} C_i$. Для $x = (x_i)_{i \in I} \in C$ позначимо $x^i = (x_j)_{j \in I, j \neq i}$. Точка $\bar{x} = (\bar{x}_i)_{i \in I}$ називається рівновагою Неша, якщо для всіх $i \in I$ справедливі нерівності

$$f_i(\bar{x}) \leq f_i(\bar{x}^i, y_i) \ \ \forall y_i \in C_i.$$

Визначимо функцію $F : C \times C \to \mathbb{R}$ таким чином

$$F(x, y) = \sum_{i \in I} \left(f_i(x^i, y_i) - f_i(x) \right).$$

Точка $\bar{x} \in C$ є рівновагою Неша тоді і тільки тоді, коли $\bar{x}$ є розв'язком задачі (5.1).

Формулювання задачі про рівновагу, яке вважають класичним, було наведено ще в роботах Х. Нікайдо та К. Ісоди, виконаних в 1950-х роках [110] та пов'язаних з доведенням існування точок рівноваги за Нешем в некооперативних іграх.

Увагу дослідників до задач рівноважного програмування у 1990-х привернули роботи W. Oettli [111, 112]. Пріоритетні результати, повязані з ітераційними проксимальними методами рівноважного програмування, належать А.С. Антіпіну [113–116].

Припустимо, що виконуються такі умови:

(A1) $F(x, x) = 0$ для всіх $x \in C$;

(A2) для всіх x, $y \in C$ з $F(x, y) \geq 0$ випливає $F(y, x) \leq 0$ (псевдомонотонність);

(A3) для всіх $x \in C$ функція $F(x, \cdot)$ напівнеперервна знизу та опукла на множині C;

(A4) для всіх $y \in C$ функція $F(\cdot, y)$ слабко напівнеперервна зверху на множині C;

(A5) для всіх x, y, $z \in C$ має місце

$$F(x, y) \leq F(x, z) + F(z, y) + a \|x - z\|^2 + b \|z - y\|^2,$$

де a, b — додатні константи (ліпшицевість).

Зауваження 5.1. Умова (A5) типу ліпшицевості введена G. Mastroeni [117]. Наприклад, біфункція $F(x, y) = (Ax, y - x)$ з ліпшицевим оператором $A : C \to H$ задовольняє (A5) з з $a = \frac{L\varepsilon}{2}$, $b = \frac{L}{2\varepsilon}$, де $\varepsilon > 0$. Дійсно,

$$\begin{aligned} F(x, y) - F(x, z) - F(z, y) &= (Ax, y - x) - (Ax, z - x) - (Az, y - z) = \\ &= (Ax - Az, y - z) \leq \|Ax - Az\| \, \|y - z\| \leq L \|x - z\| \, \|y - z\| \leq \\ &\leq \frac{L\varepsilon}{2} \|x - z\|^2 + \frac{L}{2\varepsilon} \|y - z\|^2. \end{aligned}$$

Розлянемо, так звану, дуальну [109] (для задачі (5.1)) задачу про рівновагу:

$$\text{знайти } x \in C: \ F(y,x) \le 0 \ \ \forall y \in C. \tag{5.3}$$

Множини розв'язків задач (5.1) та (5.3) позначимо S та S^*, відповідно. При виконанні умов (A1)–(A4) маємо $S = S^*$ [109]. Зокрема, множина S — опукла та замкнена, оскільки

$$S = S^* = \{x \in C: \ F(y,x) \le 0 \quad \forall y \in C\}.$$

Далі будемо припускати, що $S \ne \emptyset$.

У 2008 р. Quoc, Muu та Hien, грунтуючись на ідеях [117], запропонували у роботі [118] аналог екстраградієнтного методу

$$\begin{cases} y_n = \operatorname{prox}_{\lambda_n F(x_n,\cdot)} x_n, \\ x_{n+1} = \operatorname{prox}_{\lambda_n F(y_n,\cdot)} x_n, \end{cases} \tag{5.4}$$

де $\lambda_n > 0$. Автори [118] довели при певних умовах збіжність методу (5.4) та його аналогу з відстанню Брегмана замість евклідової. Дана робота отримала продовження в багатьох дослідників [119–126]. Наприклад, автори [120] поєднавши ідеї [118] та [72, 73] запропонували сильно збіжні методи. А в роботі [125], відштовхуючись від двокрокового екстраградієнтного алгоритму Зикіної та Меленьчука [127], запропоновано та досліджено такий алгоритм

$$\begin{cases} y_n = \operatorname{prox}_{\lambda_n \cdot F(x_n,\cdot)} x_n, \\ z_n = \operatorname{prox}_{\lambda_n \cdot F(y_n,\cdot)} y_n, \\ x_{n+1} = \operatorname{prox}_{\lambda_n \cdot F(z_n,\cdot)} x_n, \end{cases}$$

де $\lambda_n > 0$.

Для розв'язання задачі (5.1) розглянемо такий ітераційний алгоритм.

Алгоритм 5.1 (С. І. Ляшко, В. В. Семенов, [128])**.** Для x_1, $y_0 \in C$ генеруємо послідовність елементів x_n, $y_n \in C$ за допомогою ітераційної схеми

$$\begin{cases} y_n = \operatorname{prox}_{\lambda_n F(y_{n-1},\cdot)} x_n, \\ x_{n+1} = \operatorname{prox}_{\lambda_n F(y_n,\cdot)} x_n, \end{cases}$$

де $\lambda_n > 0$.

На кожному кроці алгоритму 5.1 слід розв'язати дві задачі мінімізації (на множині C) з сильно опуклими функціями. Припустимо, що їх можна ефективно розв'язати.

Алгоритм 5.1 був запропонований у [128]. Далі, у [129] був запропонований варіант алгоритму з дивергенцією Брегмана замість евклідової метрики. У цій роботі алгоритм 5.1 розглядається при слабших припущеннях ніж в [128].

Зауваження 5.2. Якщо $F(x, y) = (Ax, y - x)$, то алгоритм 5.1 приймає вигляд:

$$\begin{cases} x_1 \in C, \ y_0 \in C, \\ y_n = P_C(x_n - \lambda_n A y_{n-1}), \\ x_{n+1} = P_C(x_n - \lambda_n A y_n), \end{cases}$$

де P_C — оператор метричного проектування на множину C.

Для сідлової задачі (5.2) алгоритм 5.1 набуває наступного вигляду. Для

$$(x_1^1, x_2^1), \ \ (y_1^0, y_2^0) \in C_1 \times C_2$$

генеруємо послідовності пар (x_1^n, x_2^n), $(y_1^n, y_2^n) \in C_1 \times C_2$ за допомогою ітераційної схеми:

$$\begin{aligned} (y_1^n, y_2^n) = \operatorname{argmin}_{y_1 \in C_1, y_2 \in C_2} \Big\{ L(y_1, y_2^{n-1}) - L(y_1^{n-1}, y_2) + \tfrac{1}{2\lambda_n} \|y_1 - x_1^n\|^2 + \\ + \tfrac{1}{2\lambda_n} \|y_2 - x_2^n\|^2 \Big\}, \\ (x_1^{n+1}, x_2^{n+1}) = \operatorname{argmin}_{y_1 \in C_1, y_2 \in C_2} \Big\{ L(y_1, y_2^n) - L(y_1^n, y_2) + \tfrac{1}{2\lambda_n} \|y_1 - x_1^n\|^2 + \\ + \tfrac{1}{2\lambda_n} \|y_2 - x_2^n\|^2 \Big\}, \end{aligned}$$

де $\lambda_n > 0$.

Для пошуку рівноваги Неша алгоритм 5.1 можна подати у такому вигляді. Для

$$(x_i^1)_{i \in I}, \ \ (y_i^0)_{i \in I} \in C = \prod_{i \in I} C_i$$

генеруємо послідовності $(x_i^n)_{i \in I}$, $(y_i^n)_{i \in I} \in C$ за допомогою ітераційної схеми:

$$(y_i^n)_{i \in I} = \operatorname{argmin}_{(y_i)_{i \in I} \in C} \left\{ \sum_{i \in I} f_i(y^{i,n-1}, y_i) + \frac{1}{2\lambda_n} \sum_{i \in I} \|y_i - x_i^n\|^2 \right\},$$

$$(x_i^{n+1})_{i \in I} = \operatorname{argmin}_{(y_i)_{i \in I} \in C} \left\{ \sum_{i \in I} f_i(y^{i,n}, y_i) + \frac{1}{2\lambda_n} \sum_{i \in I} \|y_i - x_i^n\|^2 \right\},$$

де $y^{i,n} = (y_j^n)_{j \in I, j \neq i}$, $\lambda_n > 0$.

Зауваження 5.3. Зрозуміло, що допоміжні задачі ітераційних кроків варіантів алгоритму 5.1 для сідлових задач та задач рівноваги Неша мають сепарабельну структуру та допускають розщеплення. Так, для «сідлового випадку» маємо

$$\begin{aligned} y_1^n &= \mathrm{prox}_{\lambda_n \cdot L(\cdot, y_2^{n-1})} x_1^n, \\ y_2^n &= \mathrm{prox}_{-\lambda_n \cdot L(y_1^{n-1}, \cdot)} x_2^n, \\ x_1^{n+1} &= \mathrm{prox}_{\lambda_n \cdot L(\cdot, y_2^n)} x_1^n, \\ x_2^{n+1} &= \mathrm{prox}_{-\lambda_n \cdot L(y_1^n, \cdot)} x_2^n. \end{aligned}$$

Відносно обгрунтуваня збіжності алгоритму 5.1 зауважимо, що за виконання для деякого $n \in \mathbb{N}$ рівностей

$$y_n = y_{n-1} = x_n \quad \text{або} \quad x_{n+1} = x_n = y_n \tag{5.5}$$

має місце включення $y_n \in S$. Дійсно, рівність

$$x_{n+1} = \mathrm{prox}_{\lambda_n F(y_n, \cdot)} x_n$$

рівносильна нерівності

$$F(y_n, y) - F(y_n, x_{n+1}) + \frac{(x_{n+1} - x_n, y - x_{n+1})}{\lambda_n} \geq 0 \ \ \forall y \in C.$$

З другої рівності (5.5) випливає

$$F(y_n, y) \geq 0 \ \ \forall y \in C,$$

тобто, $y_n \in S$.

Аналогічно, з

$$F(y_{n-1}, y) - F(y_{n-1}, y_n) + \frac{(y_n - x_n, y - y_n)}{\lambda_n} \geq 0 \ \ \forall y \in C$$

при першій рівності в (5.5) отримуємо $y_n \in S$.

Зауваження 5.4. Вищенаведене спостереження є просто безпосередній наслідок критерію:

$$x \in S \quad \Leftrightarrow \quad x = \mathrm{prox}_{\lambda F(x, \cdot)} x, \ \ \lambda > 0.$$

Далі припустимо, що для всіх номерів $n \in \mathbb{N}$ умова (5.5) не має місця.

5.2. Основна нерівність

Доведення збіжності алгоритму 5.1 почнемо з доведення важливої нерівності для послідовностей (x_n), (y_n), що породжуються ним.

Лема 5.1. *Для породжених алгоритмом 5.1 послідовностей* (x_n), (y_n) *та елементу* $z \in S$ *виконується нерівність*

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - (1 - 2\lambda_n b) \|x_{n+1} - y_n\|^2 - \\ - (1 - 4\lambda_n a) \|y_n - x_n\|^2 + 4\lambda_n a \|x_n - y_{n-1}\|^2 . \quad (5.6)$$

Зауваження 5.5. За певного вибору λ_n з нерівності (5.6) випливає один з основних елементів доведення збіжності алгоритму 5.1: незростання послідовності $\left(\|x_n - z\|^2 + 4\lambda_n a \|y_{n-1} - x_n\|^2\right)$ для всіх $z \in S$.

Доведення. Маємо

$$\|x_{n+1} - z\|^2 = \|x_n - z\|^2 - \|x_n - x_{n+1}\|^2 + 2 (x_{n+1} - x_n, x_{n+1} - z) = \\ = \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 - \\ - 2 (x_n - y_n, y_n - x_{n+1}) + 2 (x_{n+1} - x_n, x_{n+1} - z) . \quad (5.7)$$

З правила породження точок x_{n+1} та y_n випливає

$$\lambda_n F(y_n, z) - \lambda_n F(y_n, x_{n+1}) \geq (x_{n+1} - x_n, x_{n+1} - z), \quad (5.8)$$

$$\lambda_n F(y_{n-1}, x_{n+1}) - \lambda_n F(y_{n-1}, y_n) \geq -(x_n - y_n, y_n - x_{n+1}). \quad (5.9)$$

Використавши нерівності (5.8), (5.9) для оцінки скалярних добутків в (5.7), отримуємо

$$\|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda_n \{F(y_n, z) - F(y_n, x_{n+1}) + F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n)\} . \quad (5.10)$$

З псевдомонотонності біфункції F та включення $z \in S$ випливає

$$F(y_n, z) \leq 0,$$

а ліпшицевість F гарантує виконання нерівності

$$- F(y_n, x_{n+1}) + F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) \leq \\ \leq a \|y_{n-1} - y_n\|^2 + b \|y_n - x_{n+1}\|^2 . \quad (5.11)$$

Використавши вищенаведені оцінки в (5.10), отримуємо

$$\|x_{n+1}-z\|^2 \le \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda_n a \|y_{n-1} - y_n\|^2 + 2\lambda_n b \|y_n - x_{n+1}\|^2. \quad (5.12)$$

Член $\|y_{n-1} - y_n\|^2$ оцінимо таким чином

$$\|y_{n-1} - y_n\|^2 \le 2\|y_{n-1} - x_n\|^2 + 2\|y_n - x_n\|^2.$$

Ураховуючи цю оцінку в (5.12), приходимо до нерівності

$$\|x_{n+1}-z\|^2 \le \|x_n - z\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 4\lambda_n a \|y_{n-1} - x_n\|^2 + 4\lambda_n a \|y_n - x_n\|^2 + 2\lambda_n b \|y_n - x_{n+1}\|^2,$$

тобто, до нерівності (5.6). □

5.3. Збіжність алгоритму

Перейдемо безпосередньо до доведення слабкої збіжності алгоритму 5.1. Нам будуть потрібні такі факти.

Лема 5.2. *Нехай невід'ємні послідовності* (a_n), (b_n), *такі, що*

$$a_{n+1} \le a_n - b_n.$$

Тоді існує границя $\lim_{n\to\infty} a_n \in \mathbb{R}$ *та* $\sum_{n=1}^{\infty} b_n < +\infty$.

Лема 5.3 (Z. Opial, [55]). *Нехай послідовність* (x_n) *елементів гільбертового простору* H *слабко збігається до точки* $x \in H$. *Тоді для всіх* $y \in H\backslash\{x\}$ *маємо*

$$\liminf_{n\to\infty} \|x_n - x\| < \liminf_{n\to\infty} \|x_n - y\|.$$

Нехай $z \in S$. Покладемо

$$a_n = \|x_n - z\|^2 + 4\lambda_n a \|y_{n-1} - x_n\|^2, \\ b_n = (1 - 4\lambda_n a)\|y_n - x_n\|^2 + (1 - 4\lambda_n a - 2\lambda_n b)\|y_n - x_{n+1}\|^2.$$

Тоді (5.6) приймає вигляд

$$a_{n+1} \le a_n - b_n.$$

Вимагатимемо виконання умови

$$0 < \underline{\lambda} \le \lambda_n \le \overline{\lambda} < \frac{1}{2(2a+b)}.$$

Тоді з леми 5.2 можемо зробитии висновок, що існує границя

$$\lim_{n\to\infty}\left(\|x_n - z\|^2 + 4\lambda_n a\,\|y_{n-1} - x_n\|^2\right)$$

та

$$\lim_{n\to\infty}\left((1-4\lambda_n a)\,\|y_n - x_n\|^2 + (1-4\lambda_n a - 2\lambda_n b)\,\|y_n - x_{n+1}\|^2\right) = 0.$$

Звідки отримуємо

$$\lim_{n\to\infty}\|y_n - x_n\| = \lim_{n\to\infty}\|y_n - x_{n+1}\| = \lim_{n\to\infty}\|x_n - x_{n+1}\| = 0 \tag{5.13}$$

та збіжність числових послідовностей $(\|x_n - z\|)$, $(\|y_n - z\|)$ для всіх $z \in S$. Зокрема, послідовності (x_n), (y_n) обмежені.

З (5.8), (5.9), (5.10), (5.11) та (5.13) випливає

$$\limsup_{n\to\infty} F(y_n, x_{n+1}) \le 0, \quad \limsup_{n\to\infty}\left(F(y_{n-1}, y_n) - F(y_{n-1}, x_{n+1})\right) \le 0,$$
$$\lim_{n\to\infty}\left(F(y_n, x_{n+1}) + F(y_{n-1}, y_n) - F(y_{n-1}, x_{n+1})\right) = 0.$$

Звідки, зокрема, отримуємо

$$\lim_{n\to\infty} F(y_n, x_{n+1}) = 0. \tag{5.14}$$

Розглянемо підпослідовність (x_{n_k}), слабко збіжну до деякої точки $\bar{z} \in C$. Тоді з (5.13) випливає, що й $y_{n_k} \rightharpoonup \bar{z}$. Покажемо, що $\bar{z} \in S$. Маємо

$$F(y_n, y) \ge F(y_n, x_{n+1}) + \frac{(x_{n+1} - x_n, x_{n+1} - y)}{\lambda} \quad \forall y \in C. \tag{5.15}$$

Здійснивши граничний перехід в нерівності (5.15) з урахуванням (5.13), (5.14) та умови (A4), отримаємо

$$F(\bar{z}, y) \ge \limsup_{k\to\infty} F(y_{n_k}, y) \ge$$
$$\ge \lim_{k\to\infty}\left\{F(y_{n_k}, x_{n_k+1}) + \frac{(x_{n_k+1} - x_{n_k}, x_{n_k+1} - y)}{\lambda}\right\} = 0 \quad \forall y \in C,$$

тобто, $\bar{z} \in S$.

Покажемо тепер, що $x_n \rightharpoonup \bar{z}$. Тоді з $x_n - y_n \to 0$ випливає, що й послідовність (y_n) слабко збігається $\bar{z}$. Міркуємо від супротивного. Нехай існує така підпослідовність (x_{m_k}), що $x_{m_k} \rightharpoonup \tilde{z}$ та $\tilde{z} \neq \bar{z}$. Ясно, що $\tilde{z} \in S$. Застосуємо двічі лему 5.3. Маємо

$$\lim_{n\to\infty} \|x_n - \bar{z}\| = \lim_{k\to\infty} \|x_{n_k} - \bar{z}\| < \lim_{k\to\infty} \|x_{n_k} - \tilde{z}\| = \lim_{n\to\infty} \|x_n - \tilde{z}\| = \\ = \lim_{k\to\infty} \|x_{m_k} - \tilde{z}\| < \lim_{k\to\infty} \|x_{m_k} - \bar{z}\| = \lim_{n\to\infty} \|x_n - \bar{z}\|,$$

що неможливо. Отже, $x_n \rightharpoonup \bar{z}$.

Таким чином, має місце

Теорема 5.1. *Нехай H — гільбертовий простір, $C \subseteq H$ — непорожня опукла замкнена множина, для біфункції $F : C \times C \to \mathbb{R}$ виконані умови (A1)–(A5) та $S \neq \emptyset$. Припустимо, що*

$$\lambda_n \in \left[\underline{\lambda}, \overline{\lambda}\right] \subseteq \left(0, \frac{1}{2(2a+b)}\right).$$

Тоді породжені алгоритмом 5.1 послідовності (x_n), (y_n) слабко збігаються до розв'язку $\bar{z} \in S$ задачі про рівновагу (5.1), причому

$$\lim_{n\to\infty} \|x_n - y_n\| = 0.$$

Зауваження 5.6. Ми довели слабку збіжність алгоритму 5.1. Питання сильно збіжних модифікацій алгоритму розібрано в наступному розділі. Крім того, отримана модифікація алгоритму 5.1, яка не потребує знання констант з умови (A5).

Зауваження 5.7. Останнім часом зростає активність в дослідженні задач про рівновагу в метричних просторах Адамара [106, 130–134]. В роботах [106,134] для псевдомонотонних задач про рівновагу в просторах Адамара запропоновано та досліджено аналоги алгоритму 5.1.

Розділ 6

Дворівневі задачі та двоетапний проксимальний алгоритм

В дослідженні операцій виникають задачі оптимізації за послідовно заданими критеріями (лексикографічна, послідовна оптимізація) [135, 136]. Дворівневі варіаційні нерівності виникли як природне узагальнення задач лексикографічної оптимізації з двома критеріями, а також при аналізі звичайних оптимізаційних задач з обмеженнями у формі варіаційної нерівності. Як самостійний математичний об'єкт, дворівнева варіаційна нерівність у скінченновимірному випадку розглядалась у [137]. Розв'язності більш загальних n-рівневих варіаційних нерівностей та побудові одноетапних алгоритмів їх розв'язання присвячено роботи [138, 139]. У [58, 107, 140, 141] розглядались варіаційні нерівності на множині розв'язків задачі про рівновагу.

У розділі розглядається дворівнева задача: варіаційна нерівність на множині розв'язків задачі про рівновагу. Прикладом такої задачі є пошук нормальної рівноваги Неша. Подібна задача розглядалась в роботі [58], де був запропонований сильно збіжний алгоритм, який використовував операцію обчислення значення резольвенти біфункції. Останнє істотно збільшувало обчислювальну складність алгоритму.

Для розв'язання задачі запропоновано алгоритм, що суміщає у собі ідеї двоетапного проксимального методу та ітеративної регуляризації. Крім того, досліджено адаптивний варіант алгоритму з правилом оновлення параметрів, що не використовує значень ліпшицевих констант біфункції. Для монотонних біфункцій ліпшицевого типу та сильно монотонних ліпшицевих операторів доведено теореми про сильну збіжність алгоритмів. У розділі уточнюються результати робіт [107, 140, 141].

Розділ побудовано наступним чином. В підрозділі 6.1 сформульовано задачу та наведено основні припущення. В підрозділі 6.2 розлянуто конструкцію апроксимації Браудера–Тихонова для варіаційної нерівності на множині розв'язків задачі про рівновагу та доведено основні її властивості. Підрозділ 6.3 містить опис запропонованого алгоритму, а підрозділ 6.4 — теорему про сильну збіжність алгоритму. Нарешті, підрозділи 6.5 та 6.6 присвячені дослідженню адаптивного варіанту алгоритму.

6.1. Постановка задачі

Нехай H — дійсний гільбертовий простір з скалярним добутком $(\cdot,\cdot)$ та породженою нормою $\|\cdot\|$. Для оператора $A : H \to H$, множини $M \subseteq H$ та біфункції $F : H \times H \to \mathbb{R}$ позначимо $VI(A, M)$ та $EP(F, M)$ множини

$$\{x \in M : \ (Ax, y - x) \geq 0 \ \ \forall y \in M\} \ \text{ та } \ \{x \in M : \ F(x, y) \geq 0 \ \ \forall y \in M\},$$

відповідно.

Для непорожньої опуклої замкненої множини $C \subseteq H$ розглянемо дворівневу задачу:

$$\text{знайти} \ \ x \in VI(A, EP(F, C)). \tag{6.1}$$

Припустимо, що виконуються такі умови:

(A1) $F(x, x) = 0$ для всіх $x \in C$;

(A2) $F(x, y) + F(y, x) \leq 0$ для всіх x, $y \in C$ (монотонність);

(A3) для всіх $x \in C$ функція $F(x, \cdot)$ напівнеперервна знизу та опукла на множині C;

(A4) для всіх $y \in C$ функція $F(\cdot, y)$ слабко напівнеперервна зверху на множині C;

(A5) для всіх x, y, $z \in C$ має місце

$$F(x, y) \leq F(x, z) + F(z, y) + a\|x - z\|^2 + b\|z - y\|^2,$$

де a, b — додатні константи (ліпшицевість).

(A6) $EP(F, C) \neq \emptyset$.

(A7) $A : C \to H$ — μ–сильно монотонний та L–ліпшицевий оператор.

За цих умов множина $EP(F, C)$ опукла та замкнена [142], а задача (6.1) має єдиний розв'язок $x^* \in H$ [44].

Зауважимо, що біфункція $F(x, y) = (Ax, y - x)$ з ліпшицевим оператором $A : C \to H$ задовольняє (A5) з $a = \frac{L\varepsilon}{2}$, $b = \frac{L}{2\varepsilon}$, де $\varepsilon > 0$.

Зауваження 6.1. Найвідомішим окремим випадком (6.1) є задача пошуку нормального розв'язку варіаційної нерівності (при $Ax = x$, $F(x, y) = (Bx, y - x)$, де $B : C \to H$):

$$\|x\|^2 \to \min, \quad x \in C : (Bx, y - x) \geq 0 \quad \forall y \in C.$$

Апроксимуємо задачу (6.1) однорівневою та більш регулярною задачею про рівновагу.

6.2. Апроксимація Тихонова–Браудера

Розглянемо допоміжну задачу про рівновагу:

$$\text{знайти } \ x \in C : \quad F(x, y) + \varepsilon(Ax, y - x) \geq 0 \quad \forall y \in C, \tag{6.2}$$

де $\varepsilon > 0$.

Наслідуючи А.Б. Бакушинському [81], назвемо задачу (6.2) апроксимацією Тихонова–Браудера дворівневої задачі (6.1).

Зауваження 6.2. Для розв'язання екстремальних задач подібна апроксимація була запропонована А.М. Тихоновим для побудови регуляризуючих алгоритмів, а пізніше F. Browder [143, 144] застосував подібну схему для стійкої апроксимації нормального розв'язку варіаційної нерівності або проекції заданої точки на множину нерухомих точок нерозтягуючих операторів.

З результатів [142] випливає існування та єдиність розв'язку $x_\varepsilon \in C$ задачі (6.2) для довільного $\varepsilon > 0$.

Елементи $x_\varepsilon \in C$ мають декілька важливих властивостей.

Лема 6.1. *Справедливі такі нерівності:*

(i) $\|x_\varepsilon\| \leq \frac{1}{\mu}\|Ax^*\| + \|x^*\|$ *для всіх* $\varepsilon > 0$;

(ii) $\|x_\varepsilon - x_\delta\| \leq \frac{|\varepsilon - \delta|}{\varepsilon}\frac{1}{\mu}\left(1 + \frac{L}{\mu}\right)\|Ax^*\|$ *для всіх* ε, $\delta > 0$.

Доведення. Нехай $\varepsilon > 0$. Для x_ε — розв'язку задачі (6.2) та довільного елементу $\hat{x} \in EP(F, C)$ маємо

$$F(x_\varepsilon, \hat{x}) + \varepsilon(Ax_\varepsilon, \hat{x} - x_\varepsilon) \geq 0 \ \text{ и } \ F(\hat{x}, x_\varepsilon) \geq 0.$$

Склавши нерівності та скориставшись монотонністю біфункції F, отримаємо

$$(Ax_\varepsilon, \hat{x} - x_\varepsilon) \geq 0,$$

тобто

$$(Ax_\varepsilon - A\hat{x}, x_\varepsilon - \hat{x}) \le (A\hat{x}, \hat{x} - x_\varepsilon).$$

Сильна монотонність оператора A та нерівність Шварца дають

$$\mu\|x_\varepsilon - \hat{x}\| \le \|A\hat{x}\|, \tag{6.3}$$

звідки і випливає (i).

Доведемо (ii). Нехай x_ε та x_δ — розв'язки задачі (6.2) з $\varepsilon > 0$ та $\delta > 0$, відповідно. Маємо

$$F(x_\varepsilon, x_\delta) + \varepsilon(Ax_\varepsilon, x_\delta - x_\varepsilon) \ge 0 \ \text{ та } \ F(x_\delta, x_\varepsilon) + \delta(Ax_\delta, x_\varepsilon - x_\delta) \ge 0.$$

Склавши нерівності та скориставшись монотонністю біфункції F, отримаємо

$$\varepsilon(Ax_\varepsilon, x_\delta - x_\varepsilon) + \delta(Ax_\delta, x_\varepsilon - x_\delta) \ge 0.$$

Перепишемо останню нерівність у вигляді

$$\varepsilon(Ax_\varepsilon - Ax_\delta, x_\varepsilon - x_\delta) \le (\delta - \varepsilon)(Ax_\delta, x_\varepsilon - x_\delta).$$

Скориставшись сильною монотонністю оператора A, отримаємо

$$\varepsilon\mu\|x_\varepsilon - x_\delta\|^2 \le |\delta - \varepsilon|\|Ax_\delta\|\|x_\varepsilon - x_\delta\|,$$

тобто,

$$\|x_\varepsilon - x_\delta\| \le \frac{|\delta - \varepsilon|}{\varepsilon\mu}\|Ax_\delta\|. \tag{6.4}$$

Оцінимо зверху за допомогою (6.3) норму $\|Ax_\delta\|$

$$\begin{aligned} \|Ax_\delta\| \le \|Ax^*\| + \|Ax_\delta - Ax^*\| \le \\ \le \|Ax^*\| + L\|x_\delta - x^*\| \le \left(1 + \frac{L}{\mu}\right)\|Ax^*\|. \end{aligned} \tag{6.5}$$

Використавши оцінку (6.5) в (6.4) приходимо до (ii). □

При прямуванні малого додатнього параметру ε до нуля елементи x_ε сильно збігаються до розв'язку задачі (6.1).

Лема 6.2. *Нехай виконуються умови (A1)–(A4) та (A6), (A7). Тоді*

$$\lim_{\varepsilon \to 0} \|x_\varepsilon - x^*\| = 0.$$

Доведення. В силу пункту (i) леми 6.1 з $\{x_\varepsilon\}_{\varepsilon>0}$ можна виділити слабко збіжну до $w \in C$ послідовність (x_{ε_n}) $(\varepsilon_n \to 0)$.

Скориставшись слабкою напівнеперервністю зверху функції $F(\cdot, y)$, перейдемо до границі в

$$F(x_{\varepsilon_n}, y) + \varepsilon_n (Ax_{\varepsilon_n}, y - x_{\varepsilon_n}) \geq 0 \quad \forall y \in C.$$

Отримаємо

$$F(w, y) \geq 0 \quad \forall y \in C,$$

тобто, $w \in EP(F, C)$. А перейшовши в нерівності

$$(A\hat{x}, \hat{x} - x_{\varepsilon_n}) \geq (Ax_{\varepsilon_n}, \hat{x} - x_{\varepsilon_n}) \geq 0 \quad \forall \hat{x} \in EP(F, C)$$

до границі, отримаємо

$$(A\hat{x}, \hat{x} - w) \geq 0 \quad \forall \hat{x} \in EP(F, C),$$

тобто, $w = x^*$.

Покажемо, що

$$\lim_{n\to\infty} \|x_{\varepsilon_n} - x^*\| = 0.$$

Це випливає з нерівності

$$\mu \|x_{\varepsilon_n} - x^*\|^2 \leq (Ax_{\varepsilon_n} - Ax^*, x_{\varepsilon_n} - x^*) \leq (Ax^*, x^* - x_{\varepsilon_n}).$$

З єдиності елемента x^* отримуємо $\lim_{\varepsilon\to 0} \|x_\varepsilon - x^*\| = 0$. □

Перейдемо до опису алгоритму розв'язання дворівневої задачі (6.1).

6.3. Алгоритм

У роботі [128] (див. також розділ 5) для апроксимації елементів множини $EP(F, C)$ був запропонований двоетапний проксимальний алгоритм вигляду

$$\begin{cases} y_n = \mathrm{prox}_{\lambda_n F(y_{n-1}, \cdot)} x_n, \\ x_{n+1} = \mathrm{prox}_{\lambda_n F(y_n, \cdot)} x_n, \end{cases} \tag{6.6}$$

де $\lambda_n > 0$. Відштовхуючись від схеми (6.6), для розв'язання дварівневої задачі (6.1) пропонуємо такий алгоритм.

Алгоритм 6.1. Для x_1, $y_0 \in C$ генерируємо послідовність елементів x_n, $y_n \in C$ за допомогою ітераційної схеми

$$\begin{cases} z_n = x_n - \alpha_n \lambda_n A x_n, \\ y_n = \operatorname{prox}_{\lambda_n F(y_{n-1},\cdot)} z_n = \operatorname{argmin}_{y \in C} \left(\lambda_n F(y_{n-1}, y) + \frac{1}{2} \|y - z_n\|^2 \right), \\ x_{n+1} = \operatorname{prox}_{\lambda_n F(y_n,\cdot)} z_n = \operatorname{argmin}_{y \in C} \left(\lambda_n F(y_n, y) + \frac{1}{2} \|y - z_n\|^2 \right), \end{cases}$$

де $\lambda_n > 0$, $\alpha_n > 0$.

На кожному кроці алгоритму 6.1 слід розв'язати дві опуклі задачі мінімізації з сильно опуклими функціями.

Відносно параметрів алгоритму 6.1 будемо припускати, що виконані такі умови:

(B1) $\lambda_n \in \left[\underline{\lambda}, \overline{\lambda}\right] \subseteq \left(0, \frac{1}{2(2a+b)}\right)$;

(B2) $\lim_{n\to\infty} \alpha_n = 0$;

(B3) $\sum_{n=1}^{\infty} \alpha_n = +\infty$;

(B4) $\lim_{n\to\infty} \frac{\alpha_{n+1} - \alpha_n}{\alpha_n^2} = 0$.

Зауваження 6.3. В якості допустимої послідовності (α_n) можна обрати таку:

$$\alpha_n = \frac{1}{n^p}, \; p \in (0, 1).$$

Зауваження 6.4. Якщо $F(x, y) = (Bx, y - x)$, то алгоритм 6.1 набуває вигляду:

$$\begin{cases} x_1 \in C, \; y_0 \in C, \\ z_n = x_n - \alpha_n \lambda_n A x_n, \\ y_n = P_C(z_n - \lambda_n B y_{n-1}), \\ x_{n+1} = P_C(z_n - \lambda_n B y_n), \end{cases}$$

де P_C — оператор метричного проектування на множину C. Даний метод при $A = I$ був досліджений в [108].

Зауваження 6.5. В [145] для дворівневої варіаційної нерівності був запропонований та обгрунтований близький алгоритм:

$$\begin{cases} x_1 \in C, \\ y_n = P_C(x_n - \lambda_n B x_n), \\ z_n = P_C(x_n - \lambda_n B y_n), \\ x_{n+1} = z_n - \alpha_n A z_n. \end{cases}$$

Алгоритм 6.1 поєднує у собі ідеї двоетапного проксимального методу (розділ 5, [128]) та ітеративної регуляризації [81].

Доведення його сильної збіжності проведемо за такою схемою.

Нехай x_{α_n} — розв'язок задачі (6.2) при $\varepsilon = \alpha_n$. Оскільки

$$\|x_n - x^*\| \le \|x_n - x_{\alpha_n}\| + \|x_{\alpha_n} - x^*\|, \quad \lim_{n\to\infty} \|x_{\alpha_n} - x^*\| = 0,$$

то достатньо показати, що породжена алгоритмом 6.1 послідовність (x_n) має властивість

$$\lim_{n\to\infty} \|x_n - x_{\alpha_n}\| = 0.$$

6.4. Доведення збіжності

Доведення збіжності алгоритму 6.1 почнемо з доведення важливої нерівності для згенерованих ним послідовностей (x_n), (y_n) та елементів x_{α_n}.

Лема 6.3. *Для породжених алгоритмом 6.1 послідовностей* (x_n), (y_n) *та елементів* x_{α_n} *виконується нерівність*

$$\begin{aligned} \|x_{n+1} - x_{\alpha_n}\|^2 &\le (1 - \alpha_n \lambda_n \mu) \|x_n - x_{\alpha_n}\|^2 - \\ &- \left(1 - 4\lambda_n a - \alpha_n \lambda_n \frac{L^2}{\mu}\right) \|x_n - y_n\|^2 - (1 - 2\lambda_n b) \|y_n - x_{n+1}\|^2 + \\ &+ 4\lambda_n a \|y_{n-1} - x_n\|^2. \end{aligned} \tag{6.7}$$

Доведення. Маємо

$$\begin{aligned} &\|x_{n+1} - x_{\alpha_n}\|^2 = \\ &= \|x_n - x_{\alpha_n}\|^2 - \|x_n - x_{n+1}\|^2 + 2(x_{n+1} - x_n, x_{n+1} - x_{\alpha_n}) = \\ &= \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 - \\ &- 2(x_n - y_n, y_n - x_{n+1}) + 2(x_{n+1} - x_n, x_{n+1} - x_{\alpha_n}). \end{aligned} \tag{6.8}$$

З визначення точок x_{n+1} та y_n випливає

$$\lambda_n F(y_n, x_{\alpha_n}) - \lambda_n F(y_n, x_{n+1}) \ge (x_{n+1} - x_n + \alpha_n \lambda_n A x_n, x_{n+1} - x_{\alpha_n}), \tag{6.9}$$

$$\begin{aligned} &\lambda_n F(y_{n-1}, x_{n+1}) - \lambda_n F(y_{n-1}, y_n) \ge \\ &\ge -(x_n - \alpha_n \lambda_n A x_n - y_n, y_n - x_{n+1}). \end{aligned} \tag{6.10}$$

Використовуючи нерівності (6.9), (6.10) для оцінки скалярних добутків в (6.8), отримуємо

$$\|x_{n+1} - x_{\alpha_n}\|^2 \le \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda_n \{F(y_n, x_{\alpha_n}) - F(y_n, x_{n+1}) + F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n)\} + \\ + 2\alpha_n\lambda_n(Ax_n, x_{\alpha_n} - y_n). \quad (6.11)$$

Ліпшицевість біфункції F гарантує виконання нерівності

$$-F(y_n, x_{n+1}) + F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) \le \\ \le a\|y_{n-1} - y_n\|^2 + b\|y_n - x_{n+1}\|^2.$$

Використавши вищенаведену оцінку в (6.11), отримуємо

$$\|x_{n+1} - x_{\alpha_n}\|^2 \le \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda_n a\|y_{n-1} - y_n\|^2 + 2\lambda_n b\|y_n - x_{n+1}\|^2 + \\ + 2\alpha_n\lambda_n(Ax_n, x_{\alpha_n} - y_n) + 2\lambda_n F(y_n, x_{\alpha_n}). \quad (6.12)$$

Член $\|y_{n-1} - y_n\|^2$ оцінимо таким чином

$$\|y_{n-1} - y_n\|^2 \le 2\|y_{n-1} - x_n\|^2 + 2\|y_n - x_n\|^2.$$

Маємо

$$\|x_{n+1} - x_{\alpha_n}\|^2 \le \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 4\lambda_n a\|y_{n-1} - x_n\|^2 + 4\lambda_n a\|y_n - x_n\|^2 + 2\lambda_n b\|y_n - x_{n+1}\|^2 + \\ + 2\alpha_n\lambda_n(Ax_n, x_{\alpha_n} - y_n) + 2\lambda_n F(y_n, x_{\alpha_n}). \quad (6.13)$$

З монотонності біфункції F випливає

$$F(y_n, x_{\alpha_n}) \le -F(x_{\alpha_n}, y_n),$$

звідки

$$F(y_n, x_{\alpha_n}) - \alpha_n(Ax_{\alpha_n}, y_n - x_{\alpha_n}) \le -F(x_{\alpha_n}, y_n) - \alpha_n(Ax_{\alpha_n}, y_n - x_{\alpha_n}).$$

Оскільки

$$F(x_{\alpha_n}, y_n) + \alpha_n(Ax_{\alpha_n}, y_n - x_{\alpha_n}) \ge 0,$$

то

$$F(y_n, x_{\alpha_n}) \le \alpha_n(Ax_{\alpha_n}, y_n - x_{\alpha_n}).$$

Урахувавши останню оцінку в (6.13), приходимо до нерівності

$$\begin{aligned}\|x_{n+1}-x_{\alpha_n}\|^2 \le \|x_n-x_{\alpha_n}\|^2 - \|x_n-y_n\|^2 - \|y_n-x_{n+1}\|^2 + \\ +4\lambda_n a\|y_{n-1}-x_n\|^2 + 4\lambda_n a\|y_n-x_n\|^2 + 2\lambda_n b\|y_n-x_{n+1}\|^2 + \\ +2\alpha_n\lambda_n(Ax_n - Ax_{\alpha_n}, x_{\alpha_n}-y_n).\end{aligned} \tag{6.14}$$

Оцінимо зверху член $(Ax_n - Ax_{\alpha_n}, x_{\alpha_n}-y_n)$. Маємо

$$\begin{aligned}(Ax_n - Ax_{\alpha_n}, x_{\alpha_n}-y_n) = (Ax_n - Ax_{\alpha_n}, x_{\alpha_n}-x_n) + \\ +(Ax_n - Ax_{\alpha_n}, x_n-y_n) \le -\mu\|x_{\alpha_n}-x_n\|^2 + L\|x_n-x_{\alpha_n}\|\,\|x_n-y_n\| \le \\ \le -\mu\|x_{\alpha_n}-x_n\|^2 + \frac{\mu}{2}\|x_n-x_{\alpha_n}\|^2 + \frac{L^2}{2\mu}\|x_n-y_n\|^2 = \\ = -\frac{\mu}{2}\|x_n-x_{\alpha_n}\|^2 + \frac{L^2}{2\mu}\|x_n-y_n\|^2.\end{aligned} \tag{6.15}$$

З нерівностей (6.14) та (6.15) отримуємо

$$\begin{aligned}\|x_{n+1}-x_{\alpha_n}\|^2 \le (1-\alpha_n\lambda_n\mu)\|x_n-x_{\alpha_n}\|^2 - \\ -\left(1-4\lambda_n a - \alpha_n\lambda_n\frac{L^2}{\mu}\right)\|x_n-y_n\|^2 - (1-2\lambda_n b)\|y_n-x_{n+1}\|^2 + \\ +4\lambda_n a\|y_{n-1}-x_n\|^2.\end{aligned}$$

що й було потрібно. □

Доведемо оцінку з якої випливає збіжність до нулю послідовностей

$$(\|x_n-x_{\alpha_n}\|) \text{ та } (\|y_{n-1}-x_n\|).$$

Лема 6.4. *Для породжених алгоритмом 6.1 послідовностей* (x_n), (y_n) *та елементів* x_{α_n} *при великих* $n \in \mathbb{N}$ *виконується нерівність*

$$\begin{aligned}\|x_{n+1}-x_{\alpha_{n+1}}\|^2 + \frac{4\lambda_{n+1}a}{1-\alpha_{n+1}\lambda_{n+1}\mu}\|y_n-x_{n+1}\|^2 \le \\ \le \left(1-\frac{\alpha_n\lambda_n\mu}{2}\right)\left(\|x_n-x_{\alpha_n}\|^2 + \frac{4\lambda_n a}{1-\alpha_n\lambda_n\mu}\|y_{n-1}-x_n\|^2\right) + \\ +\frac{2M}{\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^3},\end{aligned} \tag{6.16}$$

де $M = \frac{1}{\mu}\left(1+\frac{L}{\mu}\right)\|Ax^*\|$.

Доведення. Маємо

$$\|x_{n+1}-x_{\alpha_n}\|^2 =$$
$$= \|x_{n+1}-x_{\alpha_{n+1}}\|^2 + \|x_{\alpha_{n+1}}-x_{\alpha_n}\|^2 +$$
$$+ 2\left(x_{n+1}-x_{\alpha_{n+1}}, x_{\alpha_{n+1}}-x_{\alpha_n}\right) \geq \|x_{n+1}-x_{\alpha_{n+1}}\|^2 +$$
$$+ \|x_{\alpha_{n+1}}-x_{\alpha_n}\|^2 - 2\|x_{n+1}-x_{\alpha_{n+1}}\|\,\|x_{\alpha_{n+1}}-x_{\alpha_n}\| \geq$$
$$\geq (1-\varepsilon)\|x_{n+1}-x_{\alpha_{n+1}}\|^2 + \frac{(\varepsilon-1)}{\varepsilon}\|x_{\alpha_{n+1}}-x_{\alpha_n}\|^2, \quad (6.17)$$

де $\varepsilon > 0$. Покладемо в (6.17) $\varepsilon = \frac{1}{2}\alpha_n\lambda_n\mu$. Отримаємо

$$\|x_{n+1}-x_{\alpha_n}\|^2 \geq \frac{2-\alpha_n\lambda_n\mu}{2}\|x_{n+1}-x_{\alpha_{n+1}}\|^2 -$$
$$- \frac{2-\alpha_n\lambda_n\mu}{\alpha_n\lambda_n\mu}\|x_{\alpha_{n+1}}-x_{\alpha_n}\|^2. \quad (6.18)$$

В силу правил узгодження значень параметрів α_n, λ_n при великих n маємо $1-\alpha_n\lambda_n\mu > 0$. З урахуванням другої нерівності леми 6.1 з (6.18) виводимо

$$\|x_{n+1}-x_{\alpha_n}\|^2 \geq \frac{2-\alpha_n\lambda_n\mu}{2}\|x_{n+1}-x_{\alpha_{n+1}}\|^2 -$$
$$- \frac{(2-\alpha_n\lambda_n\mu)}{\alpha_n\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^2}\frac{1}{\mu}\left(1+\frac{L}{\mu}\right)\|Ax^*\|, \quad (6.19)$$

для всіх $n \geq n_0$. Використавши (6.19) в (6.7), отримаємо (для $n \geq n_0$)

$$\frac{2-\alpha_n\lambda_n\mu}{2}\|x_{n+1}-x_{\alpha_{n+1}}\|^2 \leq$$
$$\leq (1-\alpha_n\lambda_n\mu)\|x_n-x_{\alpha_n}\|^2 -$$
$$- \left(1-4\lambda_n a-\alpha_n\lambda_n\frac{L^2}{\mu}\right)\|x_n-y_n\|^2 - (1-2\lambda_n b)\|y_n-x_{n+1}\|^2 +$$
$$+ 4\lambda_n a\|y_{n-1}-x_n\|^2 + \frac{(2-\alpha_n\lambda_n\mu)}{\alpha_n\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^2}M,$$

де $M = \frac{1}{\mu}\left(1+\frac{L}{\mu}\right)\|Ax^*\|$. Звідки випливає нерівність

$$\|x_{n+1}-x_{\alpha_{n+1}}\|^2 \leq \frac{2-2\alpha_n\lambda_n\mu}{2-\alpha_n\lambda_n\mu}\|x_n-x_{\alpha_n}\|^2 -$$

$$
-\frac{2\left(1-4\lambda_n a-\alpha_n\lambda_n\frac{L^2}{\mu}\right)}{2-\alpha_n\lambda_n\mu}\|x_n-y_n\|^2-\frac{2(1-2\lambda_n b)}{2-\alpha_n\lambda_n\mu}\|y_n-x_{n+1}\|^2+
$$

$$
+\frac{8\lambda_n a}{2-\alpha_n\lambda_n\mu}\|y_{n-1}-x_n\|^2+\frac{2M}{\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^3}. \quad (6.20)
$$

Перегрупувавши члени в (6.20), отримаємо

$$
\|x_{n+1}-x_{\alpha_{n+1}}\|^2+\frac{4\lambda_{n+1}a}{1-\alpha_{n+1}\lambda_{n+1}\mu}\|y_n-x_{n+1}\|^2\leq
$$

$$
\leq\frac{2-2\alpha_n\lambda_n\mu}{2-\alpha_n\lambda_n\mu}\left(\|x_n-x_{\alpha_n}\|^2+\frac{4\lambda_n a}{1-\alpha_n\lambda_n\mu}\|y_{n-1}-x_n\|^2\right)-
$$

$$
-\frac{2\left(1-4\lambda_n a-\alpha_n\lambda_n\frac{L^2}{\mu}\right)}{2-\alpha_n\lambda_n\mu}\|x_n-y_n\|^2-
$$

$$
-\left(\frac{1-2\lambda_n b}{1-\frac{\alpha_n\lambda_n\mu}{2}}-\frac{4\lambda_{n+1}a}{1-\alpha_{n+1}\lambda_{n+1}\mu}\right)\|y_n-x_{n+1}\|^2+
$$

$$
+\frac{2M}{\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^3}. \quad (6.21)
$$

Оскільки

$$
\lambda_n\in[\underline{\lambda},\overline{\lambda}]\subseteq\left(0,\frac{1}{2(2a+b)}\right)\quad\text{та}\quad \alpha_n>0,\ \lim_{n\to\infty}\alpha_n=0,
$$

то починаючи з деякого номера n_1 будуть виконуватися нерівності

$$
\frac{1-4\lambda_n a-\alpha_n\lambda_n\frac{L^2}{\mu}}{2-\alpha_n\lambda_n\mu}>0,\quad \frac{1-2\lambda_n b}{1-\frac{\alpha_n\lambda_n\mu}{2}}-\frac{4\lambda_{n+1}a}{1-\alpha_{n+1}\lambda_{n+1}\mu}>0,
$$

та

$$
\frac{2-2\alpha_n\lambda_n\mu}{2-\alpha_n\lambda_n\mu}<1-\frac{\alpha_n\lambda_n\mu}{2}.
$$

Таким чином, для номерів $n\geq N=\max\{n_0,n_1\}$ з (6.21) випливає потрібна нерівність (6.16). □

Нагадаємо відомий результат про числові послідовності [81].

Лема 6.5. *Нехай послідовність невід'ємних чисел* ξ_n *задовольняє нерівність*

$$
\xi_{n+1}\leq(1-\alpha_n)\xi_n+\beta_n,
$$

де послідовності (α_n) *та* (β_n) *мають властивості:*

1) $\alpha_n \in (0,1)$ *та* $\sum_{n=1}^{\infty} \alpha_n = +\infty$*;*

2) $\limsup_{n\to\infty} \frac{\beta_n}{\alpha_n} \le 0$.

Тоді $\lim_{n\to\infty} \xi_n = 0$.

Сформулюємо основний результат.

Теорема 6.1. *Нехай виконуються умови (A1)–(A7) та (B1)–(B4). Тоді для породжених алгоритмом 6.1 послідовностей* (x_n), (y_n) *має місце*

$$\lim_{n\to\infty} \|x_n - x^*\| = \lim_{n\to\infty} \|y_n - x^*\| = 0, \tag{6.22}$$

де $x^* \in H$ *— єдиний розв'язок задачі* (6.1).

Доведення. В силу леми 6.5 та нерівності (6.16) маємо

$$\lim_{n\to\infty} \left(\|x_n - x_{\alpha_n}\|^2 + \frac{4\lambda_n a}{1-\alpha_n\lambda_n\mu} \|y_{n-1} - x_n\|^2 \right) = 0.$$

Звідки

$$\lim_{n\to\infty} \|x_n - x_{\alpha_n}\| = \lim_{n\to\infty} \|y_{n-1} - x_n\| = 0. \tag{6.23}$$

З нерівності

$$\|x_n - x^*\| \le \|x_n - x_{\alpha_n}\| + \|x_{\alpha_n} - x^*\|,$$

леми 6.2 та (6.23) отримуємо шукані співвідношення (6.22). □

Зауваження 6.6. При $A = I$ задача (6.1) співпадає з задачею пошуку $P_{EP(F,C)}0$. Таким чином, в цьому випадку алгоритм 6.1 є сильно збіжною схемою обчислення нормального (з мінімальною нормою) розв'язку задачі про рівновагу.

6.5. Адаптивний алгоритм

Відштовхуючись від схеми двоетапного проксимального методу, для розв'язання задачі (6.1) було запропоновано такий метод (алгоритм 6.1):

$$\begin{cases} z_n = x_n - \alpha_n\lambda_n A x_n, \\ y_n = \operatorname{argmin}_{y\in C} \left(\lambda_n F(y_{n-1}, y) + \frac{1}{2}\|y - z_n\|^2 \right), \\ x_{n+1} = \operatorname{argmin}_{y\in C} \left(\lambda_n F(y_n, y) + \frac{1}{2}\|y - z_n\|^2 \right), \end{cases} \tag{6.24}$$

де λ_n задавалися виходячи з вимоги

$$\lambda_n \in [\underline{\lambda}, \overline{\lambda}] \subseteq \left(0, \frac{1}{2(2a+b)}\right),$$

а додатня послідовність (α_n) така, що

$$\lim_{n\to\infty} \alpha_n = 0, \quad \sum_{n=1}^{\infty} \alpha_n = +\infty, \quad \lim_{n\to\infty} \frac{\alpha_{n+1} - \alpha_n}{\alpha_n^2} = 0.$$

З метою позбутися явного використання в (6.24) інформації про значення констант a і b при заданні меж для λ_n розглянемо такий алгоритм з адаптивним вибором λ_n.

Алгоритм 6.2. Обираємо елементи x_1, $y_0 \in C$, $\tau \in \left(0, \frac{1}{3}\right)$, $\lambda_1 \in (0, +\infty)$. Покладаємо $n = 1$.

1. Обчислити

$$z_n = x_n - \alpha_n \lambda_n A x_n.$$

2. Обчислити

$$y_n = \mathrm{prox}_{\lambda_n F(y_{n-1}, \cdot)} x_n.$$

3. Обчислити

$$x_{n+1} = \mathrm{prox}_{\lambda_n F(y_n, \cdot)} x_n.$$

4. Обчислити

$$\lambda_{n+1} = \begin{cases} \lambda_n, & \text{якщо } F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) - F(y_n, x_{n+1}) \le 0, \\ \min\left\{\lambda_n, \frac{\tau}{2} \frac{\|y_{n-1} - y_n\|^2 + \|y_n - x_{n+1}\|^2}{(F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) - F(y_n, x_{n+1}))}\right\}, & \text{інакше}. \end{cases}$$

Покласти $n := n + 1$ та перейти на **1**.

У пропонованому алгоритмі параметр λ_{n+1} залежить від розташування точок y_{n-1}, y_n, x_{n+1}, значень $F(y_{n-1}, x_{n+1})$, $F(y_{n-1}, y_n)$ і $F(y_n, x_{n+1})$. Ніяка інформація про константи a і b не використовується. Очевидно, що послідовність (λ_n) незростаюча. Також вона обмежена знизу числом

$$\min\left\{\lambda_1, \frac{\tau}{2\max\{a, b\}}\right\}.$$

Дійсно, маємо

$$F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) - F(y_n, x_{n+1}) \le$$
$$\le \max\{a, b\} \left(\|y_{n-1} - y_n\|^2 + \|y_n - x_{n+1}\|^2 \right).$$

Щодо додатніх параметрів α_n будемо припускати виконаними наступні умови:

(C1) $\lim_{n\to\infty} \alpha_n = 0$;

(C2) $\sum_{n=1}^{\infty} \alpha_n = +\infty$;

(C3) $\lim_{n\to\infty} \frac{\alpha_{n+1}-\alpha_n}{\alpha_n^2} = 0$.

Перейдемо до обґрунтування алгоритму 6.2. Доведення збіжності проведемо за стандартною схемою. А саме, покажемо, що

$$\lim_{n\to\infty} \|x_n - x_{\alpha_n}\| = 0,$$

де x_{α_n} — розв'язок задачі (6.2) при $\varepsilon = \alpha_n$.

6.6. Доведення збіжності адаптивного алгоритму

Спочатку доведемо важливу нерівність для послідовностей (x_n), (y_n) та елементів x_{α_n}.

Лема 6.6. *Для породжених алгоритмом 6.2 послідовностей* (x_n), (y_n) *та елементів* x_{α_n} *виконується нерівність*

$$\|x_{n+1} - x_{\alpha_n}\|^2 \leqslant (1 - \alpha_n\lambda_n\mu) \|x_n - x_{\alpha_n}\|^2 -$$
$$- \left(1 - \tau\frac{\lambda_n}{\lambda_{n+1}}\right) \|x_{n+1} - y_n\|^2 - \left(1 - 2\tau\frac{\lambda_n}{\lambda_{n+1}} - \alpha_n\lambda_n\frac{L^2}{\mu}\right) \|y_n - x_n\|^2 +$$
$$+ 2\tau\frac{\lambda_n}{\lambda_{n+1}} \|x_n - y_{n-1}\|^2. \quad (6.25)$$

Доведення. Маємо

$$\|x_{n+1} - x_{\alpha_n}\|^2 =$$
$$= \|x_n - x_{\alpha_n}\|^2 - \|x_n - x_{n+1}\|^2 + 2\big(x_{n+1} - x_n, x_{n+1} - x_{\alpha_n}\big) =$$
$$= \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 -$$
$$- 2\big(x_n - y_n, y_n - x_{n+1}\big) + 2\big(x_{n+1} - x_n, x_{n+1} - x_{\alpha_n}\big). \quad (6.26)$$

З визначення точок x_{n+1} і y_n випливає

$$\lambda_n F(y_n, x_{\alpha_n}) - \lambda_n F(y_n, x_{n+1}) \geqslant \big(x_{n+1} - x_n + \alpha_n \lambda_n A x_n, x_{n+1} - x_{\alpha_n}\big), \quad (6.27)$$

$$\lambda_n F(y_{n-1}, x_{n+1}) - \lambda_n F(y_{n-1}, y_n) \geqslant -\big(x_n + \alpha_n \lambda_n A x_n - y_n, y_n - x_{n+1}\big). \quad (6.28)$$

Використавши нерівності (6.27), (6.28) для оцінки скалярних добутків в (6.26), отримуємо

$$\begin{aligned} \|x_{n+1} - x_{\alpha_n}\|^2 \leqslant \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\lambda_n \left(F(y_n, x_{\alpha_n}) - F(y_n, x_{n+1}) + F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) \right) + \\ + 2\alpha_n \lambda_n \left(A x_n, x_{\alpha_n} - y_n \right). \quad (6.29) \end{aligned}$$

З правила обчислення λ_{n+1} випливає нерівність

$$\begin{aligned} F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) - F(y_n, x_{n+1}) \leqslant \\ \leqslant \frac{\tau}{2\lambda_{n+1}} \left(\|y_{n-1} - y_n\|^2 - \|x_{n+1} - y_n\|^2 \right). \quad (6.30) \end{aligned}$$

Для оцінки виразу

$$F(y_{n-1}, x_{n+1}) - F(y_{n-1}, y_n) - F(y_n, x_{n+1})$$

в (6.29) скористаємося (6.30). Отримаємо

$$\begin{aligned} \|x_{n+1} - x_{\alpha_n}\|^2 \leqslant \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + \tau \frac{\lambda_n}{\lambda_{n+1}} \|y_{n-1} - y_n\|^2 + \tau \frac{\lambda_n}{\lambda_{n+1}} \|x_{n+1} - y_n\|^2 + \\ + 2\lambda_n F(y_n, x_{\alpha_n}) + 2\alpha_n \lambda_n \big(A x_n, x_{\alpha_n} - y_n \big). \end{aligned}$$

Член $\|y_{n-1} - y_n\|^2$ оцінимо наступним чином:

$$\|y_{n-1} - y_n\|^2 \leqslant 2\|y_{n-1} - x_n\|^2 + 2\|x_n - y_n\|^2.$$

Маємо

$$\begin{aligned} \|x_{n+1} - x_{\alpha_n}\|^2 \leqslant \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|y_{n-1} - x_n\|^2 + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|x_n - y_n\|^2 + \tau \frac{\lambda_n}{\lambda_{n+1}} \|x_{n+1} - y_n\|^2 + \\ + 2\lambda_n F(y_n, x_{\alpha_n}) + 2\alpha_n \lambda_n \big(A x_n, x_{\alpha_n} - y_n \big). \quad (6.31) \end{aligned}$$

Із монотонності біфункції F випливає

$$F(y_n, x_{\alpha_n}) \leqslant F(x_{\alpha_n}, y_n),$$

звідки

$$F(y_n, x_{\alpha_n}) - \alpha_n (A x_{\alpha_n}, y_n - x_{\alpha_n}) \leqslant -F(x_{\alpha_n}, y_n) - \alpha_n (A x_{\alpha_n}, y_n - x_{\alpha_n}).$$

Оскільки

$$F(x_{\alpha_n}, y_n) + \alpha_n (A x_{\alpha_n}, y_n - x_{\alpha_n}) \geqslant 0,$$

то

$$F(y_n, x_{\alpha_n}) \leqslant \alpha_n (A x_{\alpha_n}, y_n - x_{\alpha_n}).$$

Врахувавши останню оцінку в (6.31), приходимо до нерівності

$$\begin{aligned} \|x_{n+1} - x_{\alpha_n}\|^2 \leqslant \|x_n - x_{\alpha_n}\|^2 - \|x_n - y_n\|^2 - \|y_n - x_{n+1}\|^2 + \\ + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|y_{n-1} - x_n\|^2 + 2\tau \frac{\lambda_n}{\lambda_{n+1}} \|x_n - y_n\|^2 + \\ + \tau \frac{\lambda_n}{\lambda_{n+1}} \|x_{n+1} - y_n\|^2 + 2\alpha_n \lambda_n \left(A x_n - A x_{\alpha_n}, x_{\alpha_n} - y_n \right). \end{aligned} \tag{6.32}$$

Оцінимо зверху член $(Ax_n - Ax_{\alpha_n}, x_{\alpha_n} - y_n)$. Маємо

$$\begin{aligned} (Ax_n - Ax_{\alpha_n}, x_{\alpha_n} - y_n) = \\ = (Ax_n - Ax_{\alpha_n}, x_{\alpha_n} - x_n) + (Ax_n - Ax_{\alpha_n}, x_n - y_n) \leqslant \\ \leqslant -\mu \|x_{\alpha_n} - x_n\|^2 + L \|x_n - x_{\alpha_n}\| \, \|x_n - y_n\| \leqslant \\ \leqslant -\mu \|x_{\alpha_n} - x_n\|^2 + \frac{\mu}{2} \|x_n - x_{\alpha_n}\|^2 + \frac{L^2}{2\mu} \|x_n - y_n\|^2 = \\ = -\frac{\mu}{2} \|x_n - x_{\alpha_n}\|^2 + \frac{L^2}{2\mu} \|x_n - y_n\|^2. \end{aligned} \tag{6.33}$$

З нерівностей (6.32) і (6.33) отримуємо

$$\begin{aligned} \|x_{n+1} - x_{\alpha_n}\|^2 \leqslant (1 - \alpha_n \lambda_n \mu) \|x_n - x_{\alpha_n}\|^2 - \\ - \left(1 - \tau \frac{\lambda_n}{\lambda_{n+1}} \right) \|x_{n+1} - y_n\|^2 - \left(1 - 2\tau \frac{\lambda_n}{\lambda_{n+1}} - \alpha_n \lambda_n \frac{L^2}{\mu} \right) \|y_n - x_n\|^2 + \\ + 2\tau \lambda_n \frac{\lambda_n}{\lambda_{n+1}} \|x_n - y_{n-1}\|^2, \end{aligned}$$

що і треба було довести. □

Доведемо тепер оцінку, з якої буде слідувати збіжність до нуля послідовностей $(\|x_n - x_{\alpha_n}\|)$ і $(\|x_n - y_{n-1}\|)$.

Лема 6.7. *Для породжених алгоритмом 6.2 послідовностей* (x_n), (y_n) *і елементів* x_{α_n} *при великих* $n \in \mathbb{N}$ *виконується нерівність*

$$
\begin{aligned}
&\|x_{n+1} - x_{\alpha_{n+1}}\|^2 + \frac{2\tau\lambda_{n+1}\lambda_{n+2}^{-1}}{1-\alpha_{n+1}\lambda_{n+1}\mu}\|x_{n+1} - y_n\|^2 \leqslant \\
&\qquad \leqslant \left(1 - \frac{\alpha_n\lambda_n\mu}{2}\right)\left(\|x_n - x_{\alpha_n}\|^2 + \frac{2\tau\lambda_n\lambda_{n+1}^{-1}}{1-\alpha_n\lambda_n\mu}\|x_n - y_{n-1}\|^2\right) + \\
&\qquad\qquad + \frac{2M}{\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^3}, \qquad (6.34)
\end{aligned}
$$

де $M = \mu^{-1}(1 + L\mu^{-1})\|Ax^*\|$.

Доведення. Маємо

$$
\begin{aligned}
&\|x_{n+1} - x_{\alpha_n}\|^2 = \\
&\quad = \|x_{n+1} - x_{\alpha_{n+1}}\|^2 + \|x_{\alpha_{n+1}} - x_{\alpha_n}\|^2 + 2(x_{n+1} - x_{\alpha_{n+1}}, x_{\alpha_{n+1}} - x_{\alpha_n}) \geqslant \\
&\quad \geqslant \|x_{n+1} - x_{\alpha_{n+1}}\|^2 + \|x_{\alpha_{n+1}} - x_{\alpha_n}\|^2 - 2\|x_{n+1} - x_{\alpha_{n+1}}\|\,\|x_{\alpha_{n+1}} - x_{\alpha_n}\| \geqslant \\
&\quad \geqslant (1-\varepsilon)\|x_{n+1} - x_{\alpha_{n+1}}\|^2 + \frac{\varepsilon - 1}{\varepsilon}\|x_{\alpha_{n+1}} - x_{\alpha_n}\|^2, \qquad (6.35)
\end{aligned}
$$

де $\varepsilon > 0$. Покладемо в (6.35) $\varepsilon = \frac{1}{2}\alpha_n\lambda_n\mu$. Отримаємо

$$
\begin{aligned}
\|x_{n+1} - x_{\alpha_n}\|^2 \geqslant \frac{2-\alpha_n\lambda_n\mu}{2}&\|x_{n+1} - x_{\alpha_{n+1}}\|^2 - \\
&- \frac{2-\alpha_n\lambda_n\mu}{\alpha_n\lambda_n\mu}\|x_{\alpha_{n+1}} - x_{\alpha_n}\|^2. \qquad (6.36)
\end{aligned}
$$

В силу правил узгодження значень параметрів α_n, λ_n при великих номерах $n \in \mathbb{N}$ маємо $1 - \alpha_n\lambda_n\mu > 0$. З урахуванням другої нерівності леми 6.1 з (6.36) виводимо

$$
\begin{aligned}
\|x_{n+1} - x_{\alpha_n}\|^2 \geqslant \frac{2-\alpha_n\lambda_n\mu}{2}&\|x_{n+1} - x_{\alpha_{n+1}}\|^2 - \\
&- \frac{(2-\alpha_n\lambda_n\mu)}{\alpha_n\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^2}\mu^{-1}(1+L\mu^{-1})\|Ax^*\|. \qquad (6.37)
\end{aligned}
$$

для всіх $n \geqslant n_0$. Використавши (6.37) в (6.25), отримаємо (для $n \geqslant n_0$)

$$\frac{2-\alpha_n\lambda_n\mu}{2}\|x_{n+1}-x_{\alpha_{n+1}}\|^2 \leqslant (1-\alpha_n\lambda_n\mu)\|x_n-x_{\alpha_n}\|^2 - \\ -\left(1-\tau\frac{\lambda_n}{\lambda_{n+1}}\right)\|x_{n+1}-y_n\|^2 - \left(1-2\tau\frac{\lambda_n}{\lambda_{n+1}}-\alpha_n\lambda_n\frac{L^2}{\mu}\right)\|y_n-x_n\|^2 + \\ +2\tau\frac{\lambda_n}{\lambda_{n+1}}\|x_n-y_{n-1}\|^2+\frac{(2-\alpha_n\lambda_n\mu)}{\alpha_n\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^2}M,$$

де $M = \mu^{-1}(1+L\mu^{-1})\|Ax^*\|$. Звідки випливає нерівність

$$\|x_{n+1}-x_{\alpha_{n+1}}\|^2 \geqslant \frac{2-2\alpha_n\lambda_n\mu}{2-\alpha_n\lambda_n\mu}\|x_n-x_{\alpha_n}\|^2 - \\ -\frac{2(1-\tau\lambda_n\lambda_{n+1}^{-1})}{2-\alpha_n\lambda_n\mu}\|x_{n+1}-y_n\|^2 - \\ -\frac{2(1-2\tau\lambda_n\lambda_{n+1}^{-1}-\alpha_n\lambda_n L^2\mu^{-1})}{2-\alpha_n\lambda_n\mu}\|y_n-x_n\|^2 + \\ +\frac{4\tau\lambda_n\lambda_{n+1}^{-1}}{2-\alpha_n\lambda_n\mu}\|x_n-y_{n-1}\|^2+\frac{2M}{\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^3}. \quad (6.38)$$

Перегрупувавши члени в (6.38), отримаємо

$$\|x_{n+1}-x_{\alpha_{n+1}}\|^2 + \frac{2\tau\lambda_{n+1}\lambda_{n+2}^{-1}}{1-\alpha_{n+1}\lambda_{n+1}\mu}\|x_{n+1}-y_n\|^2 \leqslant \\ \leqslant \frac{2-2\alpha_n\lambda_n\mu}{2-\alpha_n\lambda_n\mu}\left(\|x_n-x_{\alpha_n}\|^2+\frac{2\tau\lambda_n\lambda_{n+1}^{-1}}{1-\alpha_n\lambda_n\mu}\|x_n-y_{n-1}\|^2\right) - \\ -\frac{2(1-2\tau\lambda_n\lambda_{n+1}^{-1}-\alpha_n\lambda_n L^2\mu^{-1})}{2-\alpha_n\lambda_n\mu}\|y_n-x_n\|^2 - \\ -\left(\frac{1-\tau\lambda_n\lambda_{n+1}^{-1}}{1-\frac{1}{2}\alpha_n\lambda_n\mu}-\frac{2\tau\lambda_{n+1}\lambda_{n+2}^{-1}}{1-\alpha_{n+1}\lambda_{n+1}\mu}\right)\|x_{n+1}-y_n\|^2 + \\ +\frac{2M}{\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^3}. \quad (6.39)$$

Оскільки $\tau \in \left(0, \frac{1}{3}\right)$, існує $\lim\limits_{n\to\infty}\lambda_n > 0$ і $\lim\limits_{n\to\infty}\alpha_n = 0$, то починаючи з деякого номера $n_1 \in \mathbb{N}$ будуть виконуватися нерівності

$$\frac{1-2\tau\lambda_n\lambda_{n+1}^{-1}-\alpha_n\lambda_n L^2\mu^{-1}}{2-\alpha_n\lambda_n\mu}>0,\ \ \frac{1-\tau\lambda_n\lambda_{n+1}^{-1}}{1-\frac{1}{2}\alpha_n\lambda_n\mu}-\frac{2\tau\lambda_{n+1}\lambda_{n+2}^{-1}}{1-\alpha_{n+1}\lambda_{n+1}\mu}>0,$$

$$\frac{2-2\alpha_n\lambda_n\mu}{2-\alpha_n\lambda_n\mu}<1-\frac{\alpha_n\lambda_n\mu}{2}.$$

Таким чином, для $n\geqslant N=\max\{n_0,n_1\}$ з (6.39) випливає нерівність

$$\|x_{n+1}-x_{\alpha_{n+1}}\|^2+\frac{2\tau\lambda_{n+1}\lambda_{n+2}^{-1}}{1-\alpha_{n+1}\lambda_{n+1}\mu}\|x_{n+1}-y_n\|^2\leqslant$$
$$\leqslant\left(1-\frac{\alpha_n\lambda_n\mu}{2}\right)\left(\|x_n-x_{\alpha_n}\|^2-\frac{2\tau\lambda_n\lambda_{n+1}^{-1}}{1-\alpha_n\lambda_n\mu}\|x_n-y_{n-1}\|^2\right)+$$
$$+\frac{2M}{\lambda_n\mu}\frac{(\alpha_{n+1}-\alpha_n)^2}{\alpha_n^3},$$

що і потрібно було довести. □

Сформулюємо основний результат про збіжність адаптивного алгоритму.

Теорема 6.2. *Нехай виконуються умови (A1)–(A7) і (C1)–(C3). Тоді для породжених алгоритмом 6.2 послідовностей* (x_n), (y_n) *має місце*

$$\lim_{n\to\infty}\|x_n-x^*\|=\lim_{n\to\infty}\|y_n-x^*\|=0, \tag{6.40}$$

де $x^*\in H$ *— єдиний розв'язок задачі* (6.1).

Доведення. В силу леми 6.5 і нерівності (6.34) маємо

$$\|x_n-x_{\alpha_n}\|^2+2\tau\frac{\lambda_n\lambda_{n+1}^{-1}}{1-\alpha_n\lambda_n\mu}\|x_n-y_{n-1}\|^2\to 0\quad\text{при}\quad n\to\infty.$$

Звідки

$$\lim_{n\to\infty}\|x_n-x_{\alpha_n}\|=\lim_{n\to\infty}\|y_{n-1}-x_n\|=0. \tag{6.41}$$

З нерівності

$$\|x_n-x^*\|\leqslant\|x_n-x_{\alpha_n}\|+\|x_{\alpha_n}-x^*\|,$$

леми 6.2 і (6.41) отримуємо рівності (6.40). □

Зауваження 6.7. Ясно, що послідовність (z_n) також сильно збігається до $x^*\in H$.

Розглянемо тепер окремий випадок задачі (6.1): дворівневу варіаційну нерівність в гільбертовому просторі H:

$$\text{знайти} \quad x \in VI(A_2, VI(A_1, C)). \tag{6.42}$$

Нехай виконані наступні умови: множина $C \subseteq H$ опукла та замкнена; оператор $A_1 : C \to H$ монотонний і ліпшицевий; множина $VI(A_1, C)$ непорожня; оператор $A_2 : C \to H$ сильно монотонний і ліпшицевий. Нехай P_C — оператор метричного проектування на опуклу замкнену множину C, тобто $P_C x$ — єдиний елемент множини C з властивістю

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

Для задачі (6.42) алгоритм 6.2 приймає такий вигляд.

Алгоритм 6.3 (Варіант для варіаційних нерівностей)**.** Обираємо елементи x_1, $y_0 \in C$, $\tau \in \left(0, \frac{1}{3}\right)$, $\lambda_1 \in (0, +\infty)$. Покладаємо $n = 1$.

1. Обчислити

$$z_n = x_n - \alpha_n \lambda_n A_2 x_n.$$

2. Обчислити

$$y_n = P_C \left(z_n - \lambda_n A_1 y_{n-1}\right).$$

3. Обчислити

$$x_{n+1} = P_C \left(z_n - \lambda_n A_1 y_n\right).$$

4. Обчислити

$$\lambda_{n+1} = \begin{cases} \lambda_n, & \text{якщо} \ \ (A_1 y_{n-1} - A_1 y_n, x_{n+1} - y_n) \le 0, \\ \min\left\{\lambda_n, \frac{\tau}{2} \frac{\|y_{n-1} - y_n\|^2 + \|y_n - x_{n+1}\|^2}{(A_1 y_{n-1} - A_1 y_n, x_{n+1} - y_n)}\right\}, & \text{інакше.} \end{cases}$$

Покласти $n := n + 1$ та перейти на **1**.

З теореми 6.2 випливає наступний результат.

Теорема 6.3. *Нехай* H *— гільбертовий простір,* $C \subseteq H$ *— непорожня опукла замкнена множина; оператор* $A_1 : C \to H$ *монотонний і ліпшицевий; множина* $VI(A_1, C)$ *непорожня; оператор* $A_2 : C \to H$ *сильно монотонний і ліпшицевий; виконуються умови (C1)–(C3). Тоді породжені алгоритмом 6.3 послідовності* (x_n) *і* (y_n) *сильно збігаються до єдиного розв'язку задачі* (6.42).

Зауваження 6.8. Аналогічний теоремі 6.3 результат має місце для модифікації алгоритму 6.3 з заміною інструкції перерахунку λ_n на наступну

$$\lambda_{n+1} = \begin{cases} \min\left\{\lambda_n, \tau\frac{\|y_{n-1}-y_n\|}{\|A_1 y_{n-1}-A_1 y_n\|}\right\}, & \text{якщо} \quad A_1 y_{n-1} \neq A_1 y_n, \\ \lambda_n, & \text{інакше}. \end{cases}$$

де $\tau \in \left(0, \frac{1}{3}\right)$.

Зауваження 6.9. Спираючись на результати [95], можна побудувати економну модифікацію алгоритму 6.3. Слід змінити крок 3, поклавши

$$x_{n+1} = P_{T_n}(z_n - \lambda_n A_1 y_n),$$

де

$$T_n = \{z \in H : (z_n - \lambda_n A_1 y_{n-1} - y_n, z - y_n) \leqslant 0\}.$$

Розділ 7

Алгоритм розщеплення

При розв'язанні складних задач дослідження операцій (наприклад, в моделюванні транспортних та телекомунікаційних мереж) та оптимального керування велике значення мають різні декомпозиційні підходи, що дозволяють зводити розв'язання вихідної задачі до розв'язання послідовності задач більш простої структури.

Популярні алгоритми розщеплення для варіаційних нерівностей

$$\text{знайти } x \in C: \ \exists u \in \sum_i A_i x \ \text{ і } \ (u, y - x) \geq 0 \quad \forall y \in C,$$

або більш загальних задач пошуку нулів суми монотонних операторів

$$\text{знайти } x \in H: \ 0 \in \sum_i T_i x,$$

часто використовують на кожному кроці резольвенти багатозначних операторів [54]. Явний крок в алгоритмах цитованих робіт використовується тільки для однозначних операторів-доданків. Завдяки неявному характеру ці методи мають достатній запас стійкості, однак обчислення резольвенти часто вимагає великих обчислювальних витрат[1].

В алгоритміці опуклої оптимізації для задач вигляду

$$f_1 + f_2 + ... + f_n \to \min_C$$

з'явилась серія субградієнтих алгоритмів розщеплення [146, 147].

В роботі [148] один з цих алгоритмів адаптовано для варіаційних нерівностей. Також в [148] (Зауваження 2 цитованої статті) описано без обґрунту-

[1]Нагадаємо, що резольвентою оператора $A : H \to 2^H$ називають оператор $J_A = (I + A)^{-1} : H \to 2^H$. Відомо, що у випадку максимальної монотонності оператора A резольвента J_A є однозначним, всюди заданим та міцно нерозтягуючим (firmly nonexpansive) оператором [54].

вання метод, що узагальнює відомий в оптимізації «incremental subgradient method» [146].

Нижче ми досліджуємо згаданий алгоритм розщеплення (декомпозиції) для варіаційних нерівностей з багатозначними максимальними монотонними операторами, що діють у гільбертовому просторі. Основні результати розділу було опубліковано в статті [149].

Розділ побудовано наступним чином. В підрозділі 7.1 наведено постановку задачі та сформульовані основні припущення. В підрозділі 7.2 описано алгоритм розщеплення. В підрозділі 7.3 отримано основні оцінки, з допомогою яких в підрозділі 7.4 доведено теорему про слабку збіжність чезарівських середніх елементів, що згенеровані алгоритмом. Також в підрозділі 7.4 при додаткових умовах доведено сильну збіжність алгоритму.

7.1. Попередні відомості та постановка задачі

Введемо позначення та сформулюємо задачу. Всюди далі H — дійсний гільбертовий простір з скалярним добутком $(\cdot,\cdot)$ та породженою нормою $\|\cdot\|$. Як завжди символом P_K позначимо оператор метричної проекції простору H на замкнену опуклу множину $K \subseteq H$.

Нагадаємо деякі поняття [54]. Нехай $A : H \to 2^H$ — оператор з графіком $\mathrm{gr}(A) = \left\{(x,u) \in H^2 :\ y \in Ax\right\}$.

Означення 7.1. Оператор $A : H \to 2^H$ називають монотонним, якщо для всіх (x,u), $(y,v) \in \mathrm{gr}(A)$ виконується нерівність $(u-v, x-y) \geq 0$.

Означення 7.2. Оператор $A : H \to 2^H$ називають сильно монотонним з константою $\mu > 0$, якщо для всіх (x,u), $(y,v) \in \mathrm{gr}(A)$ виконується нерівність

$$(u-v, x-y) \geq \mu \|x-y\|^2 .$$

Означення 7.3. Оператор $A : H \to 2^H$ називають максимальним монотонним, якщо для довільного монотонного оператора $B : H \to 2^H$ із співвідношення $\mathrm{gr}(A) \subseteq \mathrm{gr}(B)$ випливає $\mathrm{gr}(A) = \mathrm{gr}(B)$.

Нехай:

- $A_i : H \to 2^H$ — монотонний оператор, $i = \overline{1,p}$;
- $A = \sum_{i=1}^{p} A_i$ — максимальний монотонний оператор;
- $C \subseteq \bigcap_{i=1}^{p} \mathrm{dom}(A_i)$ — замкнена опукла множина.

Варіаційна нерівність з оператором A на множині C формулюється таким чином:

$$\text{знайти } x \in C: \ \exists u \in Ax \ \text{ та } \ (u, y - x) \geq 0 \quad \forall y \in C. \tag{7.1}$$

Зауваження 7.1. У вигляді (7.1) можна сформулювати задачу

$$f_1 + f_2 + ... + f_p \to \min_C,$$

де f_i — неперервні опуклі функції. Тут $A_i = \partial f_i$ — субдиференціал функції f_i:

$$\partial f_i(x) = \{v \in H: \ f_i(y) - f_i(x) \geq (v, y - x) \ \forall y \in H\}.$$

Множину розв'язків задачі (7.1) позначимо $VI(A, C)$.

Важливим фактом відносно структури множини розв'язків варіаційної нерівності є наступна

Лема 7.1. *Якщо оператор* $A : H \to 2^H$ *— максимальний монотонний, то*

$$VI(A, C) = \{x \in C: \ (v, y - x) \geq 0 \ \forall y \in C \ \forall v \in Ay\}.$$

Зокрема, множина $VI(A, C)$ *опукла та замкнена.*

Наша мета полягає в дослідженні явного алгоритму розщеплення для розв'язання варіаційної нерівності (7.1), що узагальнює відомий в опуклій оптимізації «incremental subgradient method» [146].

Під явним ми розуміємо алгоритм, що не містить операцій обчислення резольвент операторів

$$A_i + N_C,$$

де N_C — нормальний конус множини C в точці-аргументі, тобто

$$N_C x = \begin{cases} \{w \in H: \ (w, y - x) \leq 0\}, & x \in C, \\ \emptyset, & \text{інакше.} \end{cases}$$

7.2. Алгоритм розщеплення

Опишемо алгоритм розщеплення для варіаційної нерівності (7.1) [148].

Зафіксуємо послідовність додатніх чисел (λ_n), що задовольняє умови:

$$\sum_{n=1}^{\infty} \lambda_n = +\infty \ , \ \sum_{n=1}^{\infty} \lambda_n^2 < +\infty. \tag{7.2}$$

Алгоритм 7.1. Задано $(\lambda_n) \in \ell_2 \setminus \ell_1$.

Крок 1. Задаємо $x_1 \in C$; $n := 1$.

Крок 2. Починаючи з $y_{(n,0)} = x_n$ послідовно знаходимо елементи:

$$y_{(n,i)} = P_C\left(y_{(n,i-1)} - \lambda_n u_{(n,i)}\right) =$$
$$= \operatorname{argmin}_{y \in C}\left\{\lambda_n\left(u_{(n,i)}, y - y_{(n,i-1)}\right) + \frac{1}{2}\|y - y_{(n,i-1)}\|^2\right\},$$
$$u_{(n,i)} \in A_i y_{(n,i-1)}, \quad i = \overline{1,p}.$$

Крок 3. Якщо $y_{(n,i)} = y_{(n,i-1)}$ для всіх $i = \overline{1,p}$, то СТОП та $x_n \in VI(A,C)$. Інакше переходимо на **Крок 4**.

Крок 4. Покладаємо
$$x_{n+1} = y_{(n,p)},$$
$n := n + 1$, переходимо на **Крок 2**.

Покажемо, що якщо $y_{(n,i)} = x_n$ для всіх $i = \overline{1,p}$, то
$$x_n \in VI(A,C).$$
Дійсно, нехай $x_n = y_{(n,1)} = P_C\left(x_n - \lambda_n u_{(n,1)}\right)$. Тоді,
$$\left(x_n - \left(x_n - \lambda_n u_{(n,1)}\right), y - x_n\right) = \lambda_n\left(u_{(n,1)}, y - x_n\right) \geq 0 \quad \forall y \in C.$$
Звідки,
$$\left(u_{(n,1)}, y - x_n\right) \geq 0 \quad \forall y \in C.$$
Аналогічно отримуємо для всіх $i = \overline{2,p}$
$$\left(u_{(n,i)}, y - x_n\right) \geq 0 \quad \forall y \in C.$$
Звідки,
$$(u_n, y - x_n) \geq 0 \quad \forall y \in C.$$
де
$$u_n = \sum_{i=1}^{p} u_{(n,i)} \in \sum_{i=1}^{p} A_i x_n = A x_n.$$
Тобто, $x_n \in VI(A,C)$.

Далі розглянемо ситуацію, коли алгоритм 7.1 породжує нескінченну послідовність. Навіть у випадку $p = 1$, коли алгоритм 7.1 співпадає з класичним «субградієнтним методом», на такому рівні загальності не очікується збіжність послідовності (x_n). Нашою основною метою є доведення слабкої збіжності в H послідовності чезарівських середніх

$$z_n = \frac{\sum_{k=1}^{n} \lambda_k x_k}{\sum_{k=1}^{n} \lambda_k}.$$

Результати такого типу традиційно називають теоремами чезарівської або ергодичної збіжності [150–152].

Зробимо відносно операторів A_i наступне припущення:

$$\text{множини } \bigcup_{i=1}^{p} A_i y_{(n,i-1)} \text{ рівномірно обмежені.} \tag{7.3}$$

Зауваження 7.2. Припущення (7.3) — аналог використаного в [146, 147] «subgradient boundedness assumption».

Для доведення збіжності алгоритму нам будуть потрібні наступні факти.

Лема 7.2. *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 7.3 (Passty, [152])**.** *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 множині* 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* $z \in F$.

7.3. Основні оцінки

Аналіз збіжності алгоритму почнемо з доведення двох важливих оцінок для послідовностей (x_n) та (z_n).

Лема 7.4. *Для породженої алгоритмом 7.1 послідовності* (x_n) *та елемента* $y \in C$ *виконується нерівність*

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 - 2\lambda_n (v, x_n - y) + \\ + \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 + 2\lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\|, \quad (7.4)$$

де $v = \sum_{i=1}^{p} v_i$, $v_i \in A_i y$, $i = \overline{1, p}$.

Доведення. Для $y \in C$ и x_{n+1} маємо

$$\|x_{n+1} - y\|^2 = \left\|P_C\left(y_{(n,p-1)} - \lambda_n u_{(n,p)}\right) - y\right\|^2 \leq \left\|y_{(n,p-1)} - \lambda_n u_{(n,p)} - y\right\|^2 = \\ = \left\|y_{(n,p-1)} - y\right\|^2 + \lambda_n^2 \left\|u_{(n,p)}\right\|^2 - 2\lambda_n \left(u_{(n,p)}, y_{(n,p-1)} - y\right).$$

Візьмемо $v_p \in A_p y$. Завдяки монотонності оператора A_p отримуємо

$$\left(u_{(n,p)}, y_{(n,p-1)} - y\right) \geq \left(v_p, y_{(n,p-1)} - y\right).$$

Отже,

$$\|x_{n+1} - y\|^2 \leq \left\|y_{(n,p-1)} - y\right\|^2 + \lambda_n^2 \left\|u_{(n,p)}\right\|^2 - 2\lambda_n \left(v_p, y_{(n,p-1)} - y\right). \quad (7.5)$$

Аналогічно отримуємо нерівності

$$\left\|y_{(n,i)} - y\right\|^2 \leq \left\|y_{(n,i-1)} - y\right\|^2 + \lambda_n^2 \left\|u_{(n,i)}\right\|^2 - 2\lambda_n \left(v_i, y_{(n,i-1)} - y\right), \quad (7.6)$$

де $v_i \in A_i y$, $i = \overline{1, p-1}$. Ураховуючи (7.6) в (7.5), отримуємо

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 + \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 - 2\lambda_n \sum_{i=1}^{p} \left(v_i, y_{(n,i-1)} - y\right).$$

Маємо,

$$\sum_{i=1}^{p} \left(v_i, y_{(n,i-1)} - y\right) = (v, x_n - y) + \sum_{i=2}^{p} \left(v_i, y_{(n,i-1)} - x_n\right).$$

де $v = \sum_{i=1}^{p} v_i \in \sum_{i=1}^{p} A_i y = Ay$. Оскільки

$$\left|\left(v_i, y_{(n,i-1)} - x_n\right)\right| \leq \|v_i\| \|y_{(n,i-1)} - x_n\| \leq \|v_i\| \lambda_n \sum_{k=1}^{i} \left\|u_{(n,k)}\right\|,$$

то

$$\|x_{n+1} - y\|^2 \le$$
$$\le \|x_n - y\|^2 - 2\lambda_n (v, x_n - y) + \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 +$$
$$+ 2\lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\|,$$

що і потрібно було довести. □

Лема 7.5. *Для породженої алгоритмом 7.1 послідовності* (x_n)*, послідовності середніх* (z_n) *та елемента* $y \in C$ *виконується нерівність*

$$\frac{\|x_{m+1} - y\|^2 - \|x_1 - y\|^2}{\sum_{n=1}^{m} \lambda_n} \le 2\,(v, y - z_m) +$$
$$+ \frac{\sum_{n=1}^{m} \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2}{\sum_{n=1}^{m} \lambda_n} + \frac{2 \sum_{n=1}^{m} \lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\|}{\sum_{n=1}^{m} \lambda_n}, \qquad (7.7)$$

де $v = \sum_{i=1}^{p} v_i$*,* $v_i \in A_i y$*,* $i = \overline{1,p}$*.*

Доведення. Запишемо нерівність леми 7.4 у вигляді

$$\|x_{n+1} - y\|^2 - \|x_n - y\|^2 \le$$
$$\le 2\,(v, \lambda_n y - \lambda_n x_n) + \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 +$$
$$+ 2\lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\|, \qquad (7.8)$$

Сумуючи (7.8) по n від 1 до $m \in \mathbb{N}$, отримуємо

$$\|x_{m+1} - y\|^2 - \|x_1 - y\|^2 \le$$
$$\le 2\left(v, \sum_{n=1}^{m} \lambda_n y - \sum_{n=1}^{m} \lambda_n x_n\right) + \sum_{n=1}^{m} \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 +$$
$$+ 2 \sum_{n=1}^{m} \lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\|. \qquad (7.9)$$

Розділивши (7.9) на $\sum_{n=1}^{m} \lambda_n$, приходимо до нерівності (7.7). □

7.4. Теореми збіжності алгоритму

Припустимо, що $\mathrm{VI}(A, C) \neq \emptyset$. Має місце

Лема 7.6. *Нехай* (x_n) *— породжена алгоритмом 7.1 послідовність. Тоді для довільного елемента* $y \in \mathrm{VI}(A, C)$ *існує скінченна границя*

$$\lim_{n\to\infty} \|x_n - y\| .$$

Зокрема, послідовність (x_n) *обмежена.*

Доведення. Скористаємось лемами 7.2 та 7.4. В нерівності (7.4) припустимо, що $y \in \mathrm{VI}(A, C)$. Нехай $v_i \in A_i y$, $i = \overline{1, p}$. Отримаємо

$$\|x_{n+1} - y\|^2 \le \|x_n - y\|^2 + \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 + 2\lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\| , \quad (7.10)$$

оскільки $(v, x_n - y) \ge 0$, $v = \sum_{i=1}^{p} v_i \in Ay$.

З нерівності (7.10), припущення (7.3) та умови $(\lambda_n) \in \ell_2$ випливає існування границі $\lim_{n\to\infty} \|x_n - y\| \in \mathbb{R}$. □

Обмеженість послідовності (x_n) тягне за собою обмеженість послідовності середніх (z_n). А з леми 7.5 випливає

Лема 7.7. *Всі слабкі часткові границі послідовності середніх* (z_n) *належить множині* $\mathrm{VI}(A, C)$.

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 (7.7) для елементів z_{n_l}, после граничного переходу при $l \to \infty$, отримаємо

$$(v, y - z) \ge 0 \quad \forall y \in C \ \forall v \in Ay,$$

що в силу леми 7.1 рівносильно включенню $z \in \mathrm{VI}(A, C)$. □

Сформулюємо один з основних результатів.

Теорема 7.1. *Справедливі твердження:*

(1) якщо $\mathrm{VI}(A, C) \neq \emptyset$, *то послідовність середніх за Чезаро* (z_n) *слабко збігається до деякого елемента* $x \in \mathrm{VI}(A, C)$;

(2) якщо $\mathrm{VI}(A, C) = \emptyset$, *то* $\|z_n\| \to +\infty$.

Доведення. З лем 7.6 та 7.7 випливає, що у випадку $\mathrm{VI}(A, C) \neq \emptyset$ для згенерованої алгоритмом 7.1 послідовності (x_n) та для множини $F = \mathrm{VI}(A, C)$ виконано умови леми 7.3. Отже, послідовність (z_n) слабко збігається до деякого елемента $x \in \mathrm{VI}(A, C)$.

Припустимо, що $\mathrm{VI}(A, C) = \emptyset$. Тоді $\|z_n\| \to +\infty$. Дійсно, інакше послідовність (z_n) має слабку граничну точку z, яка, як було показано вище, належить множині $\mathrm{VI}(A, C)$. □

При деяких додаткових умовах, що пов'язані з операторами задачі, має місце сильна збіжність послідовності (x_n).

Теорема 7.2. *Нехай оператор A_1 сильно монотонний. Тоді породжена алгоритмом 7.1 послідовність (x_n) сильно збігається до єдиного розв'язку варіаційної нерівності (7.1).*

Доведення. Нехай $z \in C$ — розв'язок (7.1), A_1 — сильно монотонний оператор з константою $\mu > 0$. При $i \neq 1$ маємо

$$\left\|y_{(n,i)} - y\right\|^2 \le \left\|y_{(n,i-1)} - y\right\|^2 + \lambda_n^2 \left\|u_{(n,i)}\right\|^2 - 2\lambda_n \left(v_i, y_{(n,i-1)} - y\right),$$

де $y \in C$, $v_i \in A_i y$.

Візьмемо $v_1 \in A_1 y$. Завдяки сильній монотонності оператора A_1 отримуємо

$$\left(u_{(n,1)}, x_n - y\right) \ge (v_1, x_n - y) + \mu \|x_n - y\|^2.$$

Отже,

$$\left\|y_{(n,1)} - y\right\|^2 \le \|x_n - y\|^2 + \lambda_n^2 \left\|u_{(n,1)}\right\|^2 - 2\lambda_n (v_1, x_n - y) - 2\mu\lambda_n \|x_n - y\|^2.$$

Таким чином, маємо

$$\begin{aligned} \|x_{n+1} - y\|^2 \le \|x_n - y\|^2 - 2\lambda_n (v, x_n - y) + \\ + \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 + 2\lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\| - 2\mu\lambda_n \|x_n - y\|^2, \end{aligned} \quad (7.11)$$

де $v = \sum_{i=1}^{p} v_i \in \sum_{i=1}^{p} A_i y = Ay$. Розглянувши в (7.11) варіант $y = z$, приходимо до нерівності

$$\begin{aligned} 2\mu\lambda_n \|x_n - z\|^2 \le \|x_n - z\|^2 - \|x_{n+1} - z\|^2 + \\ + \lambda_n^2 \sum_{i=1}^{p} \left\|u_{(n,i)}\right\|^2 + 2\lambda_n^2 \sum_{i=2}^{p} \|v_i\| \sum_{k=1}^{i} \left\|u_{(n,k)}\right\|. \end{aligned} \quad (7.12)$$

Просумувавши (7.12) по n від 1 до N, отримаємо

$$2\mu\sum_{n=1}^{N}\lambda_n\|x_n-z\|^2\le\|x_1-z\|^2-\|x_{N+1}-z\|^2+$$
$$+\sum_{n=1}^{N}\lambda_n^2\left(\sum_{i=1}^{p}\left\|u_{(n,i)}\right\|^2+2\sum_{i=2}^{p}\|v_i\|\sum_{k=1}^{i}\left\|u_{(n,k)}\right\|\right).$$

Звідки випливає $\sum_{n=1}^{\infty}\lambda_n\|x_n-z\|^2<+\infty$. Оскільки $\sum_{n=1}^{\infty}\lambda_n=+\infty$ та існує границя $\lim_{n\to\infty}\|x_n-z\|$, то маємо $\lim_{n\to\infty}\|x_n-z\|=0$. □

Теорема 7.3. *Нехай оператори* $A_i : H\to 2^H$ *максимамальні монотонні* ($i=\overline{1,p}$) *та*

$$\mathrm{int}VI(A,C)\neq\emptyset.$$

Тоді породжена алгоритмом 7.1 послідовність (x_n) *сильно збігається до розв'язку (7.1).*

Доведення. Візьмемо елемент $y\in\mathrm{int}VI(A,C)$. Оскільки оператори A_i локально обмежені [54], то існує така куля

$$B(y,r)=\{z\in H:\ \|z-y\|\le r\}\subseteq VI(A,C)\ \ (r>0)$$

та число $M>0$, що для $w\in B(y,r)$ маємо (див. нерівність (7.10))

$$\|x_{n+1}-w\|^2\le\|x_n-w\|^2+\lambda_n^2\sum_{i=1}^{p}\left\|u_{(n,i)}\right\|^2+2\lambda_n^2M\sum_{i=2}^{p}\sum_{k=1}^{i}\left\|u_{(n,k)}\right\|.$$

Запишемо для

$$y_n=y-r\frac{x_{n+1}-x_n}{\|x_{n+1}-x_n\|}\in B(y,r)$$

попередню нерівність

$$\|x_{n+1}-y_n\|^2\le\|x_n-y_n\|^2+\lambda_n^2\sum_{i=1}^{p}\left\|u_{(n,i)}\right\|^2+2\lambda_n^2M\sum_{i=2}^{p}\sum_{k=1}^{i}\left\|u_{(n,k)}\right\|.$$

Цій нерівності можна надати такого вигляду

$$2r\|x_{n+1}-x_n\|\le\|x_n-y\|^2-\|x_{n+1}-y\|^2+$$
$$+\lambda_n^2\sum_{i=1}^{p}\left\|u_{(n,i)}\right\|^2+2\lambda_n^2M\sum_{i=2}^{p}\sum_{k=1}^{i}\left\|u_{(n,k)}\right\|.$$

Для довільних $m > n$ маємо

$$\|x_m - x_n\| \le \sum_{l=n}^{m-1} \|x_{l+1} - x_l\| \le \frac{\|x_n - y\|^2 - \|x_m - y\|^2}{2r} +$$
$$+ \frac{1}{2r} \sum_{l=n}^{m-1} \lambda_l^2 \left(\sum_{i=1}^{p} \left\|u_{(l,i)}\right\|^2 + 2M \sum_{i=2}^{p} \sum_{k=1}^{i} \left\|u_{(l,k)}\right\| \right).$$

З припущення (7.3), $(\lambda_n) \in \ell_2$ та леми 7.6 випливає фундаментальність послідовності (x_n). Нехай $z \in H$ — сильна границя (x_n). Тоді послідовність середніх (z_n) сильно збігається до z. Включення $z \in VI(A, C)$ випливає з леми 7.7. $\square$

Аналогічні результати мають місце для схеми розщеплення з паралельною організацією обчислень [148].

Алгоритм 7.2. Задано $(\lambda_n) \in \ell_2 \setminus \ell_1$.

Крок 1. Задаємо $x_1 \in C$; $n := 1$.

Крок 2. Для x_n знаходимо елементи:

$$y_{(n,i)} = P_C\left(x_n - \lambda_n u_{(n,i)}\right) =$$
$$= \operatorname{argmin}_{y \in C} \left\{ \lambda_n \left(u_{(n,i)}, y - x_n \right) + \frac{1}{2}\|y - x_n\|^2 \right\},$$
$$u_{(n,i)} \in A_i x_n, \;\; i = \overline{1,p}.$$

Крок 3. Якщо $y_{(n,i)} = x_n$ для всіх $i = \overline{1,p}$, то СТОП та $x_n \in VI(A, C)$. Інакше переходимо на **Крок 4**.

Крок 4. Покладаємо

$$x_{n+1} = \frac{1}{p} y_{n,1} + \frac{1}{p} y_{n,2} + \ldots + \frac{1}{p} y_{n,p},$$

$n := n + 1$, переходимо на **Крок 2**.

Зауваження 7.3. Крок 2 дозволяє паралельну реалізацію на різних обчислювальних пристроях.

7.5. Заключні коментарі

Включивши операцію усереднення в схему обчислень, отримаємо наступний явний алгоритм розщеплення.

Алгоритм 7.3. Задано $(\lambda_n) \in \ell_2 \setminus \ell_1$.

Крок 1. Задаємо $x_1 = z_1 \in C$; $\sigma_1 := \lambda_1$, $n := 1$.

Крок 2. Покладаємо $y_{(n,0)} = x_n$ та послідовно знаходимо елементи:

$$y_{(n,i)} = P_C\left(y_{(n,i-1)} - \lambda_n u_{(n,i)}\right), \ \ u_{(n,i)} \in A_i y_{(n,i-1)}, \ \ i = \overline{1,p}.$$

Крок 3. Якщо $y_{(n,i)} = y_{(n,i-1)}$ для всіх $i = \overline{1,p}$, то СТОП та $x_n \in VI(A,C)$. Інакше переходимо на **Крок 4**.

Крок 4. Покладаємо

$$\begin{aligned} x_{n+1} &= y_{(n,p)}, \\ \sigma_{n+1} &= \sigma_n + \lambda_{n+1}, \\ z_{n+1} &= \left(1 - \frac{\lambda_{n+1}}{\sigma_{n+1}}\right) z_n + \frac{\lambda_{n+1}}{\sigma_{n+1}} x_{n+1}, \end{aligned}$$

$n := n+1$, переходимо на **Крок 2**.

Якщо $VI(A,C) \neq \emptyset$, то послідовність (z_n), породжена цим алгоритмом, слабко збігається до деякого елемента $x \in VI(A,C)$, інакше $\|z_n\| \to +\infty$.

Наведемо варіант алгоритму 7.3 для задачі пошуку сідлової точки:

$$\text{знайти} \ \ (x,y) \in X \times Y: \ L(x,\eta) \leq L(x,y) \leq L(\xi,y) \quad \forall (\xi,\eta) \in X \times Y,$$

де X, Y — опуклі замкнені підмножини відповідних гільбертових просторів, $L = \sum_{i=1}^{p} L_i$, $L_i : X \times Y \to \mathbb{R}$ — неперервні опукло-угнуті функції.

Алгоритм 7.4. Задано $(\lambda_n) \in \ell_2 \setminus \ell_1$.

Крок 1. Задаємо $x_1 = \overline{x}_1 \in X$, $y_1 = \overline{y}_1 \in Y$; $\sigma_1 := \lambda_1$, $n := 1$.

Крок 2. Покладаємо $\xi_{(n,0)} = x_n$, $\eta_{(n,0)} = y_n$ та послідовно знаходимо елементи:

$$\xi_{(n,i)} = P_X\left(\xi_{(n,i-1)} - \lambda_n u_{(n,i)}\right), \ \ u_{(n,i)} \in \partial_1 L_i\left(\xi_{(n,i-1)}, \eta_{(n,i-1)}\right), \ \ i = \overline{1,p}.$$

$$\eta_{(n,i)} = P_Y\left(\eta_{(n,i-1)} + \lambda_n v_{(n,i)}\right), \ \ v_{(n,i)} \in \partial_2 L_i\left(\xi_{(n,i-1)}, \eta_{(n,i-1)}\right), \ \ i = \overline{1,p}.$$

Крок 3. Якщо $\xi_{(n,i)} = \xi_{(n,i-1)}$ і $\eta_{(n,i)} = \eta_{(n,i-1)}$ для всіх $i = \overline{1,p}$, то СТОП та (x_n, y_n) — сідлова точка. Інакше переходимо на **Крок 4**.

Крок 4. Покладаємо

$$
\begin{aligned}
x_{n+1} &= \xi_{(n,p)}, \\
y_{n+1} &= \eta_{(n,p)}, \\
\sigma_{n+1} &= \sigma_n + \lambda_{n+1}, \\
\overline{x}_{n+1} &= \left(1 - \frac{\lambda_{n+1}}{\sigma_{n+1}}\right)\overline{x}_n + \frac{\lambda_{n+1}}{\sigma_{n+1}} x_{n+1}, \\
\overline{y}_{n+1} &= \left(1 - \frac{\lambda_{n+1}}{\sigma_{n+1}}\right)\overline{y}_n + \frac{\lambda_{n+1}}{\sigma_{n+1}} x_{n+1},
\end{aligned}
$$

$n := n + 1$, переходимо на **Крок 2**.

На завершення зауважимо, що цікавою та актуальною проблемою є побудова та обгрунтування схем розщеплення (декомпозиції) для варіаційних нерівностей вигляду:

$$
\text{знайти } x \in \bigcap_k F(T_k) : \exists u \in \sum_i A_i x \;\text{ і }\; (u, y - x) \geq 0 \;\forall y \in \bigcap_k F(T_k), \tag{7.13}
$$

де $F(T_k)$ — множина нерухомих точок (квазі)нерозтягуючого оператора T_k. У випадку $A_i \equiv 0$ задача (7.13) переходить в класичну задачу пошуку елемента множини $\bigcap_k F(T_k)$ — спільної нерухомої точки операторів T_k (common fixed point problem), яка має багату алгоритміку [54].

Література

[1] Шор Н.З. Использование операций растяжения пространства в задачах минимизации выпуклых функций. Кибернетика. 1970. № 1. С. 6–12.

[2] Davidon W.C. Variable metric methods for minimization. AEC Research and Development Rept. ANL 5990 (Rev.). 1959.

[3] Fletcher R., Powell M.J.D. A rapidly convergent descent method for minimization. Comput. J. 1963. 6, № 2. P. 163–168.

[4] Пшеничный Б.Н., Данилин Ю.М. Численные методы в экстремальных задачах. М.: Наука. 1975. 319 с.

[5] Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1983. 384 с.

[6] Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. Киев: Наук. думка. 1979. 199 с.

[7] Шор Н.З., Стеценко С.И. Квадратичные экстремальные задачи и недифференцируемая оптимизация. Киев: Наук. думка. 1989. 208 с.

[8] Стецюк П.И. r-алгоритмы и эллипсоиды. Кибернетика и системный анализ. 1996. № 1. С. 113–134.

[9] Стецюк П.И. Метод центров тяжести простых тел. Кибернетика и системный анализ. 1996. № 5. С. 117–138.

[10] Stetsyuk P.I. Orthogonalizing linear operators in convex programming. I. Cybernetics and Systems Analysis. 1997. Vol. 33. Issue 3. P. 386–401.

[11] Stetsyuk P.I. Orthogonalizing linear operators in convex programming. II. Cybernetics and Systems Analysis. 1997. Vol. 33. Issue 5. P. 700–709.

[12] Semenov V., Stetsyuk P., Stovba V., Velarde Cantu J.M. One-Rank Linear Transformations and Fejer-Type Methods: An Overview. Mathematics. 2024. 12(10):1527. https://doi.org/10.3390/math12101527.

[13] Agmon S. The relaxation method for linear inequalities. Canadien Journal of Mathematics. 1954. N 6. P. 382–392.

[14] Motzkin T., Schoenberg I.J. The relaxation method for linear inequalities. Canadien Journal of Mathematics. 1954. N 6. P. 393–404.

[15] Поляк Б.Т. Минимизация негладких функционалов. Вычисл. математика и мат. физика. 1969. 9, № 3. С. 507–521.

[16] Щепакин М.Б. О методе ортогонального спуска. Кибернетика. 1987. № 1. С. 58–62.

[17] Skokov V.A., Shchepakin M.B. Numerical analysis of the orthogonal descent method. Cybernetics and Systems Analysis. 1994. Vol. 30. P. 274–282.

[18] Щепакин М.Б., Шубенкова И.А. Исследование модифицированного метода ортогонального спуска для поиска нуля выпуклой функции. Кибернетика и системный анализ. 1993. № 4. С. 63–72.

[19] Camerini P., Fratta L., Maffioli F. On improving relaxation methods by modified gradient techniques. Math. Program. 1975. Study 3. P. 26–34.

[20] Hiriart-Urruty J.B., Lemarechal C. Convex analysis and minimization Algorithms. Berlin: Springer-Verlag. 1994. Vol I-II.

[21] Воеводин В.В. Численные методы алгебры. М.: Наука. 1966. 248 с.

[22] Nonsmooth optimization. Eds. Lemarechal C., Mifflin R. Oxford: Pergamon Press. 1978. 186 p.

[23] Lemarechal C. Numerical experiments in nonsmooth optimization. In: Progress in nondifferentiable optimization. Ed. Nurminski E.A. CP–82–S2, International Institute for Applied System Analysis: Laxenburg, Austria, 1982. P. 61–84.

[24] Полак Э. Численные методы оптимизации. Единый подход. М.: Мир. 1974. 376 с.

[25] Химмельблау Д. Прикладное нелинейное программирование. М.: Мир, 1975. 535 с.

[26] Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. М.: Мир, 1985. 509 с.

[27] Журбенко Н.Г. Квазиньютоновские алгоритмы минимизации на основе использования оператора растяжения пространства. Теория оптимальных решений. Киев, 1999. С. 45–50.

[28] Byrd R.H., Lu P., Nocedal J. Limited Memory Algorithm for Bound Constrained Optimization. SIAM Journal of Scientific and Statistical Computing. 1995. No. 16(5). P. 1190–1208.

[29] Стецюк П.И. Квазиньютоновские методи и r-алгоритмы. Препринт НАН Украины. Ин-т кибернетики им. В.М. Глушкова. Киев. 1996. 96–10. 21 с.

[30] Shor N.Z. Nondifferentiable optimization and polynomial problems. Boston; Dordrecht; London: Kluwer Academic Publishers. 1998. 412 p.

[31] Стецюк П.И. Теория и программные реализации r-алгоритмов Шора. Кибернетика и системный анализ. 2017. № 5. С. 43–57.

[32] Стецюк П.И. Методы эллипсоидов и r-алгоритмы. Кишинэу: Эврика. 2014. 488 с.

[33] Lemarechal C. An extension of Davidon methods to nondifferentiable problems. Math. Progr. Study 3. 1975. P. 95–109.

[34] Стецюк П.И. Линейные операторы в квазиньютоновских методах. Теория и приложения методов оптимизации. Киев. 1998. С. 3–8.

[35] Polyak R. Introduction to continuous optimization. Cham, Switzerland: Springer, 2021. 541 p.

[36] Измаилов А.Ф., Куренной А.С., Стецюк П.И. Метод Левенберга–Марквардта для задач безусловной оптимизации. Вестник Тамбовского университета. Серия: Естественные и технические науки. 2019. Т. 24, №125. С. 60–74.

[37] Denisov S.V., Semenov V.V., Chabak L.M. Convergence of the Modified Extragradient Method for Variational Inequalities with Non-Lipschitz Operators. Cybernetics and Systems Analysis. 2015. Vol. 51. P. 757–765.

[38] Verlan D.A., Semenov V.V., Chabak L.M. A Strongly Convergent Modified Extragradient Method for Variational Inequalities with Non-Lipschitz Operators. Journal of Automation and Information Sciences. 2015. Vol. 47. Issue 7. P. 31–46.

[39] Lions J.L., Stampacchia G. Variational inequalities. Commun. Pure Appl. Math. 1967. Vol. XX. P. 493–519.

[40] Stampacchia G. Le probleme de Dirichlet pour les equations elliptiques du second ordre a coefcients discontinus. Ann. Inst. Fourier. 1965. Vol. 15, no. 1. P. 189–257.

[41] Fichera G. Problemi Elastostatici con Vincoli Unilaterali, il Problema di Signorini con Ambigue Condizioni al Contorno. Mem. Accad. Naz. Lincei. 1964. Vol. VIII, no. 7. P. 91–140.

[42] Hartman P., Stampacchia G. On some nonlinear elliptic di erential functional equations. Acta Math. 1966. Vol. 115. P. 153–188.

[43] Лионс Ж.-Л. Некоторые методы решения нелинейных краевых задач. Москва: Мир, 1972. 587 с.

[44] Киндерлерер Д., Стампаккья Г. Введение в вариационные неравенства и их приложения. Москва: Мир, 1983. 256 с.

[45] Байокки К., Капело А. Вариационные и квазивариационные неравенства. Москва: Наука, 1988. 448 с.

[46] Лионс Ж.-Л. Оптимальное управление системами, описываемыми уравнениями с частными производными. Москва: Мир, 1972. 414 с.

[47] Nemirovski A. Prox-method with rate of convergence O(1/T) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. on Optim. 2004. Vol. 15. P. 229–251.

[48] Wang J.-K., Abernethy J., Levy K. Y. No-Regret Dynamics in the Fenchel Game: A Unified Framework for Algorithmic Convex Optimization. arXiv preprint arXiv:2111.11309. 2021.

[49] Gidel G., Berard H., Vincent P., Lacoste-Julien S. A Variational Inequality Perspective on Generative Adversarial Networks. arXiv preprint arXiv:1802.10551. 2018.

[50] Daskalakis C., Ilyas A., Syrgkanis V., Zeng H. Training GANs with optimism. arXiv preprint arXiv:1711.00141. 2018.

[51] Chavdarova T., Gidel G., Fleuret F., Lacoste-Julien S. Reducing Noise in GAN Training with Variance Reduced Extragradient. arXiv preprint arXiv:1904.08598. 2019.

[52] Mishchenko K., Kovalev D., Shulgin E., Richtarik P., Malitsky Y. Revisiting Stochastic Extragradient. arXiv preprint arXiv:1905.11373. 2019.

[53] Liu M., Zhang W., Mroueh Y., Cui X., Ross J., Yang T., Das P. A decentralized parallel algorithm for training generative adversarial nets. arXiv preprint arXiv:1910.12999. 2019.

[54] Bauschke H.H., Combettes P.L. Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Berlin, Heidelberg, NY: Springer, 2011. 408 p.

[55] Opial Z. Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Amer. Math. Soc. 1967. Vol. 73. P. 591–597.

[56] Mainge P.-E. Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. Set-Valued Analysis. 2008. Vol. 16. P. 899–912.

[57] Semenov V.V. A Strongly Convergent Splitting Method for Systems of Operator Inclusions with Monotone Operators. Journal of Automation and Information Sciences. 2014. Vol. 46, No. 5. P. 45–56.

[58] Semenov V.V. Strongly Convergent Algorithms for Variational Inequality Problem Over the Set of Solutions the Equilibrium Problems. In: M.Z. Zgurovsky and V.A. Sadovnichiy (eds.), Continuous and Distributed Systems, Solid Mechanics and Its Applications, Volume 211, Springer International Publishing Switzerland, 2014, pp. 131–146.

[59] Васин В.В., Еремин И.И. Операторы и итерационные процессы фейеровского типа. (Теория и приложения). Москва-Ижевск: Регулярная и хаотическая динамика, 2005. 200 с.

[60] Еремин И.И., Мазуров В.Д. Нестационарные процессы математического программирования. Москва: Наука, 1979. 228 с.

[61] Нурминский Е.А. Использование дополнительных малых воздействий в фейеровских моделях итеративных алгоритмов. Ж. вычисл. матем. и матем. физ. 2008. Т. 48, № 12. С. 2121–2128.

[62] Mann W.R. Mean value methods in iteration. Proc. Amer. Math. Soc. 1953. P. 506–510.

[63] Красносельский М.А. Два замечания о методе последовательных приближений. УМН. 1955. Том 10. Выпуск 1(63). С. 123–127.

[64] Halpern B. Fixed points of nonexpanding maps. Bull. Amer. Math. Soc. 1967. 73. P. 957–961.

[65] Lions P.-L. Approximation de points fixes de contractions. C. R. Acad. Sci. Paris, Ser. A-B. 1977. 284. P. A1357–A1359.

[66] Wittmann R. Approximation of fixed points of nonexpansive mappings. Arch. Math. 1992. 58. P. 486–491.

[67] Xu H.K. Another control condition in an iterative method for nonexpansive mappings. Bull. Austral. Math. Soc. 2002. 65 P. 109–113.

[68] Xu H.K. Iterative algorithms for nonlinear operators. J. London Math. Soc. 2002. 2. P. 240-256.

[69] Suzuki T. A sufficient and necessary condition for Halpern-type strong convergence to fixed points of nonexpansive mappings. Proc. of the AMS. 2007. V. 135. N. 1. P. 99–106.

[70] Bauschke H.H. The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space. J. Math. Anal. Appl. 1996. Vol. 202. P. 150–159.

[71] Семенов В.В. Два методи апроксимації нерухомої точки фейєрівського оператора. Журнал обчисл. та прикл. матем. 2013. № 1(111). С. 46–56.

[72] Nakajo K., Takahashi W. Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups. J. Math. Anal. Appl. 2003. 279. P. 372–379.

[73] Takahashi W., Takeuchi Y., Kubota R. Strong convergence theorems by hybrid methods for families of nonexpansive mappings in Hilbert spaces. J. Math. Anal. Appl. 2008. 341. P. 276—286.

[74] Корпелевич Г.М. Экстраградиентный метод для отыскания седловых точек и других задач. Экономика и математические методы. 1976. № 4. С. 747–756.

[75] Антипин А. С. О методе выпуклого программирования, использующем симметрическую модификацию функции Лагранжа. Экономика и математические методы. 1976. № 6. С. 1164–1173.

[76] Семенов В.В. Варіаційні нерівності: теорія та алгоритми. Київ: ВПЦ «Київський університет», 2021. 167 с.

[77] Censor Y., Gibali A., Reich S. The subgradient extragradient method for solving variational inequalities in Hilbert space. Journal of Optitmization Theory and Applications. 2011. 148. P. 318–335.

[78] Lyashko S.I., Semenov V.V., Voitova T.A. Low-cost modification of Korpelevich's methods for monotone equilibrium problems. Cybernetics and Systems Analysis. 2011. Vol. 47. Issue 4. P. 631–639.

[79] Tseng P. A modified forward-backward splitting method for maximal monotone mappings. SIAM Journal on Optimization. 2000. Vol. 38. P. 431–446.

[80] Denisov S.V., Semenov V.V., Stetsyuk P.I. Bregman Extragradient Method with Monotone Rule of Step Adjustment. Cybernetics and Systems Analysis. 2019. Vol. 55. Issue 3. P. 377–383.

[81] Бакушинский А.Б., Гончарский А.В. Некорректные задачи. Численные методы и приложения. Москва: Изд-во МГУ, 1989. 200 с.

[82] Nadezhkina N., Takahashi W. Strong convergence theorem by a hybrid method for nonexpansive mappings and Lipschitz-continuous monotone mappings. SIAM J. Optim. 2006. Vol. 16. P. 1230–1241.

[83] Beck A. First-Order Methods in Optimization. Philadelphia: Society for Industrial and Applied Mathematics, 2017. 479 p.

[84] Ryu E. K., Yin W. Large-Scale Convex Optimization. Algorithms and Analyses via Monotone Operators. Cambridge: Cambridge University Press, 2023. 303 p.

[85] Экланд И., Темам Р. Выпуклый анализ и вариационные проблемы. Москва: Мир, 1979. 399 с.

[86] Rudin L. I., Osher S., Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena. 1992. Vol. 60. N. 1–4. P. 259–268.

[87] Chambolle A., Lions P.-L. Image recovery via total variation minimization and related problems. Numerische Mathematik. 1997. Vol. 76. N. 2. P. 167–188.

[88] Ben-Tal A., El Ghaoui L., Nemirovski A. Robust optimization. Princeton University Press, 2009. 576 p.

[89] Madry A., Makelov A., Schmidt L., Tsipras D., Vladu A. Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083. 2017.

[90] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair Sh., Courville A., Bengio Y. Generative Adversarial Networks. Advances in Neural Information Processing Systems. 2014. P. 2672–2680.

[91] Arjovsky M., Chintala S., Bottou L. Wasserstein GAN. arXiv preprint arXiv:1701.07875. 2017.

[92] Попов Л.Д. Модификация метода Эрроу-Гурвица поиска седловых точек. Математические заметки. 1980. Т. 28, № 5. С. 777–784.

[93] Konnov I.V. Combined relaxation methods for variational inequalities. Berlin, Heidelberg, New York: Springer–Verlag, 2001. 181 p.

[94] Nesterov Y. Dual Extrapolation and Its Applications to Solving Variational Inequalities and Related Problems. Math. Program. 2007. V. 109. No. 2-3. P. 319–344.

[95] Malitsky Yu.V., Semenov V.V. An extragradient algorithm for monotone variational inequalities. Cybernetics and Systems Analysis. 2014. Vol. 50. P. 271–277.

[96] Semenov V.V. A Version of the Mirror descent Method to Solve Variational Inequalities. Cybernetics and Systems Analysis. 2017. Vol. 53. P. 234–243.

[97] Semenov V.V. A variant of mirror descent method for solving variational inequalities. In: Polyakova, L.N. (ed.) Constructive Nonsmooth Analysis and Related Topics (dedicated to the memory of V.F. Demyanov). IEEE, 2017. P. 281–284.

[98] Denisov S.V., Dudar V.V., Semenov V.V., Vedel Y.I. A New Mirror-prox Algorithm For Variational Inequalities. Журнал обчисл. та прикл. матем. 2017. № 1(124). С. 15–29.

[99] Ведель Я.И., Семенов В.В. Двухэтапный метод с расхождением Брэгмана для решения вариационных неравенств. Журнал обчисл. та прикл. матем. 2018. № 3(129). С. 18–27.

[100] Семенов В.В., Денисов С.В., Сирык Д.С., Харьков О.С. Сходимость метода экстраполяции из прошлого и метода операторной экстраполяции. Міжнародний науково-технічний журнал «Проблеми керування та інформатики». 2021. Том 66. № 3. С. 58–72.

[101] Semenov V.V., Denisov S.V., Kravets A.V. Adaptive Two-Stage Bregman Method for Variational Inequalities. Cybernetics and Systems Analysis. 2021. Vol. 57. Issue 6. P. 959–967.

[102] Харьков О.С. Оцінки ефективності для методів з дивергенцією Брегмана. Журнал обчисл. та прикл. матем. 2023. № 2. С. 83–93.

[103] Харьков О., Семенов В. Швидкість збіжності нових алгоритмів для варіаційних нерівностей. Modeling, control and information technologies: Proceedings of VI International scientific and practical conference. P. 159–160.

[104] Semenov V.V., Denisov S.V., Sandrakov G.V., Kharkov O.S. Convergence of the Operator Extrapolation Method for Variational Inequalities in Banach Spaces. Cybernetics and Systems Analysis. 2022. Vol. 58. Issue 5. P. 740–753.

[105] Malitsky Y., Tam M.K. A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity. SIAM J. on Optimization. 2020. Vol. 30. P. 1451–1472.

[106] Vedel Y.I., Sandrakov G.V., Semenov V.V. An Adaptive Two-Stage Proximal Algorithm for Equilibrium Problems in Hadamard Spaces. Cybernetics and Systems Analysis. 2020. Vol. 56. Issue 6. P. 978–989.

[107] Ведель Я.И., Денисов С.В., Семёнов В.В. Алгоритм для вариационного неравенства на множестве решений задачи о равновесии. Журнал обчисл. та прикл. матем. 2020. № 1(133). С. 18–30.

[108] Popov L.D. On schemes for the formation of a master sequence in a regularized extragradient method for solving variational inequalities. Russian Mathematics. 2004. Vol. 48. Issue 1. P. 67–76.

[109] Kassay G., Radulescu V.D. Equilibrium Problems and Applications. London: Academic Press, 2019. xx + 419 p.

[110] Nikaido H., Isoda K. Note on noncooperative convex games. Pacific Journal of Mathematics. 1955. Vol. 5. P. 807–815.

[111] Blum E., Oettli W. From optimization and variational inequalities to equilibrium problems. Math. Stud. 1994. 63. P. 123–145.

[112] Muu L.D., Oettli W. Convergence of an adaptive penalty scheme for finding constrained equilibria. Nonlinear Anal. TMA. 1992. 18. P. 1159–1166.

[113] Антипин А.С. Равновесное программирование: проксимальные методы. Журн. вычисл. матем. и матем. физ. 1997. 37, № 11. С. 1327—1339.

[114] Antipin A.S. Equilibrium programming: gradient methods. Part 2. Autom. Remote Control. 1997. 58 (8). P. 1337–1347.

[115] Antipin A. Equilibrium programming problems: prox-regularization and prox-methods. In: Recent Advances in Optimization. Lecture Notes in Economics and Mathematical Systems, vol. 452, Springer, Heidelberg, 1997. P. 1–18.

[116] Antipin A.S., Flam S.D. Equilibrium programming using proximal-like algorithms. Math. Program. 1997. 78. P. 29–41.

[117] Mastroeni G. On auxiliary principle for equilibrium problems. In: Daniele, P. et al. (eds.) Equilibrium Problems and Variational Models. Kluwer Academic Publishers, Dordrecht, 2003. P. 289–298.

[118] Quoc T.D., Muu L.D., Hien N.V. Extragradient algorithms extended to equilibrium problems. Optimization. 2008. Vol. 57. P. 749–776.

[119] Van N.T.T., Strodiot J.J., Nguyen V.H. A bundle method for solving equilibrium problems. Math. Program. 2009. 116(1–2), Ser. B. P. 529–552.

[120] Войтова Т.А., Денисов С.В., Семенов В.В. Сильно збіжний модифікований варіант методу Корпелевич для задач рівноважного програмування. Журнал обчисл. та прикл. матем. 2011. № 1(104). С. 10–23.

[121] Anh P.N. Strong convergence theorems for nonexpansive mappings and Ky Fan inequalities. J. Optim. Theory Appl. 2012. 154. P. 303–320.

[122] Vuong P.T., Strodiot J.J, Nguyen V.H. Extragradient methods and linesearch algorithms for solving Ky Fan inequalities and fixed point problems. J. Optim. Theory Appl. 2012. 155. P. 605–627.

[123] Quoc T.D., Anh P.N., Muu L.D. Dual extragradient algorithms to equilibrium problems. J. Glob. Optim. 2012. 53. P. 139–159.

[124] Vuong P.T., Strodiot J.J., Nguyen V.H. On extragradient-viscosity methods for solving equilibrium and fixed point problems in a Hilbert space. Optimization. 2015. 64(2). P. 429–451.

[125] Nguyen T.P.D., Strodiot J.J., Nguyen V.H., Nguyen T.T.V. A family of extragradient methods for solving equilibrium problems. J. Ind. Manag. Optim. 2015. 11. P. 619–630.

[126] Anh P.N., Hai T.N., Tuan P.M. On Ergodic Algorithms for Equilibrium Problems. J. Glob. Optim. 2016. 64(1). P. 179–195.

[127] Зыкина А. В., Меленьчук Н. В. Двухшаговый экстраградиентный метод для вариационных неравенств. Изв. вузов. Матем. 2010. № 9. С. 82–85.

[128] Lyashko S.I., Semenov V.V. A New Two-Step Proximal Algorithm of Solving the Problem of Equilibrium Programming. In: B. Goldengorin (ed.) Optimization and Its Applications in Control and Data Sciences. Springer Optimization and Its Applications, vol. 115. Springer, Cham, 2016. P. 315–325.

[129] Chabak L., Semenov V., Vedel Y. A New Non-Euclidean Proximal Method for Equilibrium Problems. In: Chertov O., Mylovanov T., Kondratenko Y., Kacprzyk J., Kreinovich V., Stefanuk V. (eds) Recent Developments in Data Science and Intelligent Analysis of Information. ICDSIAI 2018. Advances in Intelligent Systems and Computing, vol. 836. Springer, Cham, 2019. P. 50–58.

[130] Colao V., Lopez G., Marino G., Martin-Marquez V. Equilibrium problems in Hadamard manifolds. Journal of Mathematical Analysis and Applications. 2012. Vol. 388. P. 61–77.

[131] Khatibzadeh H., Mohebbi V. Approximating solutions of equilibrium problems in Hadamard spaces. Miskolc Mathematical Notes. 2019. Vol. 20. No. 1. P. 281–297. https://doi.org/10.18514/MMN.2019.2361.

[132] Vedel Y.I., Denisov S.V., Semenov V.V. Regularized Adaptive Extra-Proximal Algorithm for Equilibrium Problem in Hadamard Spaces. Journal of Automation and Information Sciences. 2020. Vol. 52. Issue 9. P. 12–26.

[133] Vedel Y., Semenov V. Adaptive Extraproximal Algorithm for the Equilibrium Problem in Hadamard Spaces. In: Olenev N., Evtushenko Y., Khachay M., Malkova V. (eds) Optimization and Applications. OPTIMA 2020. Lecture Notes in Computer Science, vol. 12422. Springer, Cham, 2020. P 287–300.

[134] Vedel Y. I., Sandrakov G. V., Semenov V. V., Chabak L. M. Convergence of a Two-Stage Proximal Algorithm for the Equilibrium Problem in Hadamard Spaces. Cybernetics and Systems Analysis. 2020. Vol. 56. Issue 5. P. 784–792.

[135] Еремин И.И. О задачах последовательного программирования. Сиб. мат. журн. 1973. 14, № 1. С. 53–63.

[136] Подиновский В.В., Гаврилов В.Н. Оптимизация по последовательно применяемым критериям. Москва: Советское радио, 1975. 192 с.

[137] Калашников В.В., Калашникова Н.И. Решение двухуровневого вариационного неравенства. Кибернетика и системный анализ. 1994. № 4. С. 178–180.

[138] Коннов И.В. О системах вариационных неравенств. Изв. вузов. Матем. 1997. № 12. С. 79–88.

[139] Попов Л.Д. Лексикографические вариационные неравенства и некоторые приложения. Математическое программирование. Регуляризация и аппроксимация, Сборник статей. Тр. ИММ. 8, № 1. 2002. С. 103–115.

[140] Denisov S., Semenov V., Vedel Y. Adaptive Algorithm for Variational Inequality Problem Over the Set of Solutions the Equilibrium Problem. 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2020. P. 325–329.

[141] Vedel Y.I., Denisov S.V., Semenov V.V. An Adaptive Algorithm for the Variational Inequality Over the Set of Solutions of the Equilibrium Problem. Cybernetics and Systems Analysis. 2021. Vol. 57. Issue 1. P. 91–100.

[142] Combettes P.L., Hirstoaga S.A. Equilibrium Programming in Hilbert Spaces. J. Nonlinear Convex Anal. 2005. Vol. 6. P. 117–136.

[143] Browder F. Existence and approximation of solutions of nonlinear variational inequalities. Proc. Nat. Acad. Sci. USA. 1966. Vol. 56. No. 4. P. 1080–1086.

[144] Browder F.E. Convergence of approximants of fixed points of nonexpansive non-linear mappings in Banach spaces. Arch. Rational Mech. Anal. 1967. Vol. 24. P. 82–90.

[145] Апостол Р.Я., Гриненко А.А., Семенов В.В. Ітераційні алгоритми для монотонних дворівневих варіаційних нерівностей. Журнал обчисл. та прикл. матем. 2012. № 1(107). С. 3–14.

[146] Nedic A., Bertsekas D.P. Incremental subgradient methods for nondifferentiable optimization. SIAM J. Optim. 2001. Vol. 12, No. 1. P. 109–138.

[147] Bertsekas D.P. Incremental proximal methods for large scale convex optimization. Math. Program., Ser. B. 2011. 129. P. 163–195.

[148] Семенов В.В. Явный алгоритм расщепления для вариационных неравенств с монотонными операторами. Журнал обчисл. та прикл. матем. 2013. № 2(112). С. 42–52.

[149] Ляшко Н.И., Семенов В.В., Чабак Л.М., Алгоритм расщепления для вариационных неравенств с максимальными монотонными операторами. Журнал обчисл. та прикл. матем. 2014. № 3(117). С. 131–139.

[150] Bruck R.E. On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space. Journal of Mathematical Analysis and Applications. 1977. 61. P. 159–164.

[151] Немировский А.С., Юдин Д.Б. Чезаровская сходимость градиентного метода аппроксимации седловых точек выпукло-вогнутых функций. Доклады АН СССР. 1978. т. 239, вып. 5. 1056–1059.

[152] Passty G.B. Ergodic Convergence to a Zero of the Sum of Monotone Operators in Hilbert Spaces. Journal of Mathematical Analysis and Applications. 1979. 72. P. 383–390.

Інформація про авторів

Семенов Володимир Вікторович — доктор фізико-математичних наук, професор, професор кафедри обчислювальної математики факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка **(розділи 1, 3–7)**.

Стецюк Петро Іванович — член–кореспондент НАН України, доктор фізико-математичних наук, старший науковий співробітник, завідувач відділу методів негладкої оптимізації Інституту кібернетики імені В.М. Глушкова НАН України **(розділи 1, 2)**.

Стовба Віктор Олександрович — доктор філософії за спеціальністю «Прикладна математика», науковий співробітник відділу методів негладкої оптимізації Інституту кібернетики імені В.М. Глушкова НАН України, асистент кафедри інтелектуальних програмних систем факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка **(розділи 1, 2)**.

Супрун Антон Андрійович — доктор філософії за спеціальністю «Прикладна математика», у 2023 році закінчив аспірантуру Інституту кібернетики імені В.М. Глушкова НАН України **(розділ 2)**.

Харьков Олег Сергійович — доктор філософії за спеціальністю «Прикладна математика», у 2024 році закінчив аспірантуру факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка **(розділи 4, 6)**.

Коваленко Олександра Юріївна — аспірантка факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка **(розділи 5, 7)**.

Наукове видання

СЕМЕНОВ Володимир Вікторович
СТЕЦЮК Петро Іванович
СТОВБА Віктор Олександрович
СУПРУН Антон Андрійович
ХАРЬКОВ Олег Сергійович
КОВАЛЕНКО Олександра Юріївна

СУБГРАДІЄНТНІ ТА ЕКСТРАГРАДІЄНТНІ АЛГОРИТМИ

Монографія

Частина 1

Друкується за авторською редакцією

Формат $60x84^{1/16}$. Ум. друк. арк. 11,74. Наклад 300. Зам. № 224-11127.
Гарнітура Computer Modern. Папір офсетний. Друк офсетний.
Підписано до друку 18.10.24

Видавець і виготовлювач
ВПЦ "Київський університет"
Б-р Тараса Шевченка, 14, м. Київ, 01601, Україна
☎ (38044) 239 32 22; (38044) 239 31 58; (38044) 239 31 28
e-mail: vpc@knu.ua; vpc_div.chief@univ.net.ua; redaktor@univ.net.ua
http: vpc.knu.ua
Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02

Freepik.com