М. П. Моклячук

НЕГЛАДКИЙ АНАЛІЗ ТА ОПТИМІЗАЦІЯ

Київ
2008

ЗМІСТ

Передмова

Навчальний посібник написаний на основі лекцій з курсу "Негладкий аналіз та оптимізація", які читаються студентам–магістрам механіко-математичного факультету Київського національного університету імені Тараса Шевченка протягом останніх 10 років.

В першому розділі "Опуклі множини" викладені основні властивості опуклих множин, опуклих многогранників, конусів. Доведені теореми про розділення опуклих множин, про крайні точки опуклої множини.

У другому розділі "Опуклі функції" описані властивості опуклих функцій, субградієнти та субдиференціали, умови оптимальності в задачах математичного програмування, задачі опуклого програмування. Як приклад застосування відповідних теоретичних результатів розглядаються задачі теорії наближень.

Третій розділ "Верхня опукла апроксимація" посібника містить основні властивості функцій, що допускають верхню опуклу апроксимацію або нижню угнуту апроксимацію. Досліджені необхідні умови екстремуму для таких функцій.

Четвертий розділ "Похідні за напрямком" присвячений функціям, які мають похідні за напрямками Діні чи Адамара. Описані апроксимації множин та функцій за допомогою конусів. Сформульовані та доведені умови екстремуму для функцій, що мають похідні за напрямками.

У п'ятому розділі "Субдиференційовні функції" вивчаються узагальнені похідні та субдиференціали. Описані властивості та взаємозв'язок субдиференціалів Кларка, Шора, Мішеля—Пено. Встановлені умови мінімуму для субдиференційовних функцій.

У шостому розділі "Квазідиференційовні функції" вивчаються властивості квазідиференційовних функцій та правила квазідиференційовного числення. Встановлені умови екстремуму квазідиференційовних функцій. Як приклад, описана модель математичної економіки.

Сьомий розділ "Кодиференційовні функції" присвячений вивченню правил кодиференційного числення. Встановлено умови екстремуму першого та другого порядку.

Навчальний посібник підготовлений та виданий за підтримки програми Tempus у рамках проекту TEMPUS PROJECT IB-JET-25054-2004. Автор користується нагодою щоб висловити подяку за підтримку координаторам проекту професору Сільвестрову Дмитру Сергійовичу, Сільвестровій Евеліні Дмитрівні (Малардаленський університет, Швеція), професору Мішурі Юлії Степанівні (Київський національний університет імені Тараса Шевченка).

Вступ

В останні роки невпинно зростає інтерес до дослідження екстремальних задач з параметрами, які не задовольняються стандартним припущенням про гладкість (диференційовність). Це пояснюється як потребами розвитку теорії, так і важливими застосуваннями в економіці, техніці, фізиці, біології, екології (див., наприклад, [4,2,10, 14, 15, 16, 24, 25, 31, 33, 35, 40, 43, 45]).

Необхідність дослідження диференційних властивостей негладких функцій виникає в багатьох розділах математики. Це теорія наближень [29, 53], фунуціональний аналіз [3,45,44], рівняння математичної фізики, дослідження операцій [11,12, 34]. Але найбільша необхідність таких досліджень в теорії екстремальних задач, не негладкості виникають природним чином у зв'язку з використанням операцій максимуму та мінімуму, методів декомпозиції, двоїстості, збурень [13, 20, 26, 28, 54, 58].

Одним з головних понять класичного (гладкого) математичного аналізу є поняття похідної (градієнту). Негладкий аналіз вивчає властивості недиференційовних функцій. Тому виникає проблема замінити градієнт іншим поняттям, що має принаймі частину властивостей градієнту. Різні наукові школи використовують різні узагальнення поняття градієнту, що дозволяють вивчати класи недиференційовних (негладких) функцій і розв'язувати оптимізаційні задачі та інші проблеми негладкого аналізу. Опишемо основні властивості диференційовних функцій та можливі узагальнення поняття градієнту.

1. Гладкий аналіз. Диференційовні функції.

Нехай $f\colon \mathbb{R}^n \to \mathbb{R}$ неперервно диференційовна функція на відкритій множині $S \subset \mathbb{R}^n$. Тоді визначений градієнт $f'\colon S \to \mathbb{R}^n$, що є неперервною функцією на S. За допомогою градієнта ми можемо:
1. Знайти похідні за напрямком

$$f'(x,g) = \lim_{\alpha \downarrow 0} \frac{(f(x+\alpha g) - f(x))}{\alpha} = \langle f'(x), g \rangle. \tag{1}$$

2. Побудувати апроксимацію першого порядку функції f в околі точки x:

$$f(x+\Delta) = f(x) + \langle f'(x), \Delta \rangle + o_x(\Delta), \tag{2}$$

де

$$\lim_{\alpha \downarrow 0} \frac{o_x(\alpha \Delta)}{\alpha} = 0 \quad \forall \Delta \in \mathbb{R}^n. \tag{3}$$

3. Сформулювати необхідні умови екстремуму першого порядку:
3a. Якщо $x^* \in \operatorname{locmin} f(x)$, $x \in \mathbb{R}^n$, то необхідно, щоб

$$f'(x^*) = 0. \tag{4}$$

3b. Якщо $x^{**} \in \operatorname{locmax} f(x)$, $x \in \mathbb{R}^n$, то необхідно, щоб

$$f'(x^{**}) = 0. \tag{5}$$

Зауважимо, що умови (4), (5) співпадають. Точки $x \in \mathbb{R}^n$, що задовольняють умови (4), (5), називаються *стаціонарними* точками функції f.
4. Можемо визначити напрямки найшвидшого спуску та найшвидшого підйому функції f в точці x (якщо точка x не є стаціонарною):
4a. Напрямок $g_0(x) = -\frac{f'(x)}{\|f'(x)\|}, f'(x) \neq 0$ є напрямком найшвидшого спуску функції f в точці x.
4b. Напрямок $g_1(x) = \frac{f'(x)}{\|f'(x)\|}, f'(x) \neq 0$ є напрямком найшвидшого підйому функції f в точці.
Зауважимо, що у гладкому випадку існує лише один напрямок найшвидшого спуску та один напрямок найшвидшого підйому. При цьому

$$g_0(x) = -g_1(x). \tag{6}$$

5. Будувати чисельні методи знаходження екстремальних значень функції f (різні варіанти градієнтного методу).
6. Якщо $f'(x,g) < 0 (> 0)$, то $f'(x, -g) > 0 (< 0)$, тобто функція f спадає (зростає) за деяким напрямком g тоді і тільки тоді, коли вона зростає (спадає) за протилежним напрямком $-g$. При цьому

$$f'(x,g) = -f'(x, -g). \tag{7}$$

7. Якщо $f'(x,g) < 0$, то $f'(x',g) < 0$ для всіх x' з околу точки x, тобто напрямок g є робастним напрямком спадання функції f. Аналогічна властивість справджується і для напрямку зростання функції f.
8. Функція $F(x,\Delta) = \langle f'(x),\Delta \rangle = f'(x,\Delta)$, яка визначає апроксимацію приросту $f(x+\Delta) - f(x)$, неперервна як функція від x.
9. Справджується наступна теорема про середнє: якщо $\operatorname{conv}\{x_1,x_2\} \in S$, то існує таке число $\alpha \in (0,1)$, що

$$f(x_2) - f(x_1) = \langle f'(x_1 + \alpha(x_2 - x_1)), x_2 - x_1 \rangle. \tag{8}$$

10. Диференціальне числення. Знаючи похідні (градієнти) базових функцій, за допомогою правил диференціального числення можна знайти похідні (градієнти) цілого класу гладких функцій, а також доводити основні теореми диференційного числення, такі як теорема про неявну та обернену функції, теорема про нерухому точку.

Отже градієнт – це один із базових інструментів класичного математичного аналізу. Чи існує заміна градієнта в класі недиференційовних функцій? Наведемо список можливих кандидатів (у відповідних класах) та опишемо їх основні властивості.

2. Похідні за напрямком.

Розглянемо клас функцій, що мають похідні за напрямками. Границя
$$f'_D(x,g) = \lim_{\alpha \downarrow 0} \frac{(f(x+\alpha g) - f(x))}{\alpha}, \tag{9}$$
якщо вона існує, називається похідною Діні функції f в точці x за напрямком g.
Функція $f\colon S \to \mathbb{R}$ Діні диференційовна за напрямками, якщо для всіх напрямків g існує і скінченна похідна $f'_D(x,g)$.
Границя
$$f'_H(x,g) = \lim_{[\alpha,g'] \to [+0,g]} \frac{(f(x+\alpha g') - f(x))}{\alpha}, \tag{10}$$
якщо вона існує, називається *похідною Адамара* функції f в точці x за напрямком g.
Функція f диференційовна за напрямками за Адамаром, якщо для всіх напрямків g існує і скінчена похідна $f'_H(x,g)$.
Якщо функція f диференційовна за напрямками за Адамаром, то функція f диференційовна за напрямками за Діні. Обернене твердження не вірне.
Оскільки похідна $f'_D(x,g)$ і похідна $f'_H(x,g)$ додатньо однорідні першого порядку функції від g, тобто $f'_D(x,\lambda g) = \lambda f'_D(x,g)$, $f'_H(x,\lambda g) = \lambda f'_H(x,g)$, $\forall \lambda > 0$, то достатньо брати $g \in S_1 = \{g \in \mathbb{R}^n : \|g\| = 1\}$.
Якщо функція f диференційовна за напрямками за Адамаром, то функція $h(g) = f'_H(x,g)$ неперервна.

Перевіримо, чи справджуються властивості 1-10, які ми сформулювали для диференційовних функцій в тому випадку, коли функція f диферецiйовна в точці x за напрямками (за Адамаром, чи за Діні).
1. Ми можемо знайти похідні за напрямком (за означенням).
2. Побудувати апроксимації першого порядку
$$f(x+\Delta) = f(x) + f'_D(x,\Delta) + o_{x1}(\Delta), \tag{11}$$
$$f(x+\Delta) = f(x) + f'_H(x,\Delta) + o_{x2}(\Delta), \tag{12}$$
де
$$\lim_{\alpha \downarrow 0} \frac{o_{x1}(\alpha\Delta)}{\alpha} = 0 \quad \forall \Delta \in \mathbb{R}^n, \tag{13}$$
$$\lim_{\|\Delta\| \to 0} \frac{o_{x2}(\|\Delta\|)}{\|\Delta\|} = 0. \tag{14}$$
3. Сформулювати необхідні умови екстремуму першого порядку:
3а. Якщо $x^* \in \operatorname{locmin} f$, $x \in \mathbb{R}^n$, то
$$f'_D(x^*,\Delta) \geq 0 \quad \forall \Delta \in \mathbb{R}^n, \tag{15}$$

$$f'_H(x^*,\Delta) \geq 0 \quad \forall \Delta \in \mathbb{R}^n. \tag{16}$$

Якщо

$$f'_H(x^*,\Delta) > 0 \quad \forall \Delta \in \mathbb{R}^n \setminus \{0_n\}, \tag{17}$$

то $x^* \in \operatorname{locmin} f(x), x \in \mathbb{R}^n$.
3b. Якщо $x^{**} \in \operatorname{locmax} f$, $x \in \mathbb{R}^n$, то

$$f'_D(x^{**},\Delta) \leq 0 \quad \forall \Delta \in \mathbb{R}^n, \tag{18}$$

$$f'_H(x^{**},\Delta) \leq 0 \quad \forall \Delta \in \mathbb{R}^n. \tag{19}$$

Якщо

$$f'_H(x^{**},\Delta) < 0 \quad \forall \Delta \in \mathbb{R}^n \setminus \{0_n\}, \tag{20}$$

то $x^* \in \operatorname{locmax} f(x), x \in \mathbb{R}^n$.
Точки $x^* \in \mathbb{R}^n$, що задовольняють умову (15), називаються $D - inf$-*стаціонарними* точками функції f.
Точки $x^* \in \mathbb{R}^n$, що задовольняють умову (16), називаються $H - inf$-*стаціонарними* точками функції f.
Точки $x^{**} \in \mathbb{R}^n$, що задовольняють умову (18), називаються $D - sup$-*стаціонарними* точками функції f.
Точки $x^{**} \in \mathbb{R}^n$, що задовольняють умову (19), називаються $H - sup$-*стаціонарними* точками функції f.
4. Можемо визначити напрямки найшвидшого спуску (якщо точка не є $D - inf$ (чи $H - inf$)–стаціонарною) та найшвидшого підйому (якщо точка не є $D - sup$ (чи $H - sup$)–стаціонарною):
4a. Напрямок $g_0(x) \in S_1$ є напрямком D–найшвидшого спуску функції f в точці x, якщо

$$f'_D(x,g_0(x)) = \inf_{g \in S_1} f'_D(x,g).$$

Напрямок $g_0(x) \in S_1$ є напрямком H–найшвидшого спуску функції f в точці x, якщо

$$f'_H(x,g_0(x)) = \inf_{g \in S_1} f'_H(x,g).$$

4b. Напрямок $g_1(x) \in S_1$ є напрямком D–найшвидшого підйому функції f в точці x, якщо

$$f'_D(x,g_1(x)) = \sup_{g \in S_1} f'_D(x,g).$$

Напрямок $g_1(x) \in S_1$ є напрямком H–найшвидшого підйому функції f в точці x, якщо

$$f'_H(x,g_1(x)) = \sup_{g \in S_1} f'_H(x,g).$$

Зауважимо, що напрямки найшвидшого спуску та найшвидшого підйому не обов'язково єдині.

Властивості 6,7,8 не виконуються.
9. Справедлива теорема про середнє значення: якщо $[x_1,x_2] \in S$, то існує таке $\alpha \in (0,1)$, що

$$f(x_2) - f(x_1) = \langle f'_H(x_1 + \alpha(x_2 - x_1); x_2 - x_1\rangle. \tag{21}$$

З теореми про середнє значення випливає, що диференційовна за напрямками за Адамаром функція визначається однозначно з точністю до константи.
Аналогічну властивість має диференційовна за напрямками за Діні функція. Якщо S опукла множина і

$$f'_{1D}(x;g) = f'_{2D}(x;g) \quad \forall g \in \mathbb{R}^n \quad \forall x \in S,$$

то $f_1(x) = f_2(x) + C \quad \forall x \in S$.
10. Правила обчислення похідних за напрямками від суми, добутку, частки, композиції диференційовних за напрямками функцій аналогічні правилам обчислення похідних для гладких функцій.
Крім того, з'являються нові правила. Наприклад, якщо функції $f_1, \ldots, f_N$ диференційовні за Діні (чи за Адамаром) в точці x за напрямком g, то функції

$$F_1(x) = \max_{i \in I} f_i(x), \quad F_2(x) = \min_{i \in I} f_i(x),$$

де $I = \{1,2,\ldots,N\}$, теж диференційовні в точці x за напрямком g. При цьому

$$F'_1(x;g) = \max_{i \in R(x)} f'_i(x;g), \quad F'_2(x;g) = \min_{i \in Q(x)} f'_i(x;g),$$

де $R(x) = \{i \in I : f_i(x) = F_1(x)\}$, $Q(x) = \{i \in I : f_i(x) = F_2(x)\}$.
Отже в класі диференційовних за напрямками функцій похідна за напрямком є достойна заміна градієнта.

3. Опуклі функції. Субдиференціали.

Опукла функція $f : S \to \mathbb{R}^1$ на опуклій відкритій множині $S \subset \mathbb{R}^n$ неперервна на S і в кожній точці $x \in S$ диференційовна за напрямками за Адамаром. Крім того

$$f'(x,g) = f'_H(x,g) = \max_{v \in \partial f(x)} \langle v,g\rangle \quad \forall g \in \mathbb{R}^n, \tag{22}$$

$$\partial f(x) = \{v \in \mathbb{R}^v\colon f(z) - f(x) \geq \langle v,z - x\rangle\} \tag{23}$$

Множина $\partial f(x) \subset \mathbb{R}^n$, для якої виконується вказане співвідношення, називається *субдиференціалом* функції f в точці x. Ця множина непорожня, опукла, замкнута, обмежена. Елементи цієї множини називаються субградієнтами функції f в точці x. Ж. Моро [39] та Р. Рокафеллар [49] започаткували розвиток теорії опуклих функцій та субдиференціалів.

Гарні підручники з опуклого аналізу написали Р. Рокафеллар [50], Б. М. Пшеничний [47], [48], С. Бойд, Л. Ванденберг [9], Ж. Борвейн, А. Льюис [8], Ж. Ирриар-Урруті, С. Лемаречал [56].

Клас опуклих функцій замкнутий відносно додавання та множення на додатні числа. Максимум опуклих функцій також є опуклою функцією.

Умови екстремуму опуклих функцій можна виразити в термінах субдиференціалу.
Якщо точка $x^* \in S$ є точкою локального мінімуму опуклої функції f, то необхідною умовою є

$$0 \in \partial f(x^*). \tag{24}$$

Умова

$$0 \in \operatorname{int} \partial f(x^*) \tag{25}$$

є достатньою умовою строго локального мінімуму опуклої функції f в точці $x^* \in S$.
Необхідною умовою локального максимуму опуклої функції є

$$\partial f(x^{**}) = \{0\}. \tag{26}$$

Якщо $0 \notin \partial f(x)$, то напрямок

$$g_0(x) = -z(x)/\|z(x)\|, \quad \|z(x)\| = \min_{v \in \partial f(x)} \|v\|,$$

є напрямком найшвидшого спуску функції f в точці x. Напрямок найшвидшого спуску визначається однозначно.
Якщо $\partial f(x) \neq \{0\}$, то напрямок

$$g_1(x) = z_1(x)/\|z_1(x)\|, \quad \|z_1(x)\| = \max_{v \in \partial f(x)} \|v\|,$$

є напрямком найшвидшого підйому функції f в точці x. Напрямок найшвидшого підйому визначається неоднозначно.

Субдиференціальне числення. Можна сформулювати правила субдиференціального числення, що аналогічні правилам диференціального числення (виключення – множення на від'ємні числа).

Кожній опуклій функції f в точці x відповідає компактна множина (субдиференціал) $\partial f(x)$, яка для гладкої функції складається з однієї точки - градієнта. Субдиференціал має багато 'гарних' властивостей, що 'аналогічні' властивостям градієнта. Отже в класі опуклих функцій субдиференціал є достойною заміною градієнта.

4. Субдиференційовні функції.

Властивості субдиференціала опуклих функцій наводять на думку будувати класи функцій, що мають принаймі частину базових властивостей опуклих функцій. Б. М. Пшеничний [47,48] визначив та досліджував

новий клас негладких функцій - субдиференційовні функції (Б. М. Пшеничний називав їх квазідиференційовними).
Функція $f : S \to \mathbb{R}^1$, $S \subset \mathbb{R}^n$, субдиференційовна за Діні в точці $x \in S$, якщо існує така опукла компактна множина $\partial f(x) \subset \mathbb{R}^n$, що

$$f'_D(x,g) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha g) - f(x)}{\alpha} = \max_{v \in \partial f(x)} \langle v,g \rangle \quad \forall g \in \mathbb{R}^n.$$

Множина $\partial f(x) \subset \mathbb{R}^n$, для якої виконується вказане співвідношення, називається *субдиференціалом Діні* функції f в точці $x \in S$.

Функція $f : S \to \mathbb{R}^1$, $S \subset \mathbb{R}^n$, субдиференційовна за Адамаром в точці $x \in S$, якщо існує опукла компактна множина $\partial f(x) \subset \mathbb{R}^n$ така, що

$$f'_H(x,g) = \lim_{[\alpha,g'] \to [+0,g]} \frac{f(x+\alpha g') - f(x)}{\alpha} = \max_{v \in \partial f(x)} \langle v,g \rangle \quad \forall g \in \mathbb{R}^n.$$

Множина $\partial f(x) \subset \mathbb{R}^n$, для якої виконується вказане співвідношення, називається *субдиференціалом Адамара* функції f в точці $x \in S$.
Клас субдиференційовних функцій включає диференційовні функції та опуклі функції. Він замкнутий відносно додавання та множення на додатні числа. Максимум субдиференційовних функцій також є субдиференційовною функцією.

Субдиференційовні функції мають всі властивості диференційовних за напрямком функцій.
Умови екстремуму субдиференційовних функції можна виразити в термінах субдиференціалу.
Якщо точка $x^* \in S$ є точкою локального мінімуму субдиференційовної за Діні чи Адамаром функції f на множині S, то необхідною умовою є

$$0 \in \partial f(x^*).$$

Якщо функція f субдиференційовна за Адамаром в точці $x^* \in S$, то умова

$$0 \in \operatorname{int} \partial f(x^*)$$

є достатньою умовою мінімуму функції f в точці $x^* \in S$.
Якщо точка $x^{**} \in \mathbb{R}^n$ є точкою локального максимуму субдиференційовної за Діні чи Адамаром функції f, то необхідною є умова

$$\partial f(x^{**}) = \{0\}.$$

Якщо $0 \notin \partial f(x)$, то напрямок

$$g_0(x) = -z(x)/\|z(x)\|, \quad \|z(x)\| = \min_{v \in \partial f(x)} \|v\|,$$

є напрямком найшвидшого спуску (за Діні чи Адамаром в залежності від типу субдиференційовності) функції f в точці x. Напрямок найшвидшого спуску визначається однозначно.
Якщо $\partial f(x) \neq \{0\}$, то напрямок

$$g_1(x) = z_1(x)/\|z_1(x)\|, \quad \|z_1(x)\| = \max_{v \in \partial f(x)} \|v\|,$$

є напрямком найшвидшого підйому (за Діні чи Адамаром в залежності від типу субдиференційовності функції) функції f в точці x. Напрямок найшвидшого підйому визначається неоднозначно.

В 1972 році Н. З. Шор [59] запропонував конструкцію майже градієнту. Розглянемо на множині S локально ліпшицеву функцію f. Така функція диференційовна на S майже скрізь. Зафіксуємо $x \in S$ і побудуємо множину

$$\partial_{Sh} f(x) = \left\{ v \in \mathbb{R}^n | \exists \{x_i\} : v = \lim_{i \to \infty} f'(x_i), x_i \to x, x_i \in Q \right\},$$

де $Q \subset S$ – множина точок, в яких функція f диференційовна. Оскільки функція f диференційовна на S майже скрізь, то міра множини $S \setminus Q$ рівна нулю. Множина $\partial_{Sh} f(x)$ називається субдиференціалом Шора, а елементи цієї множини називаються майже градієнтами. Н. З. Шор [60] використовував майже градієнти для побудови чисельних методів мінімізації ліпшицевих функцій. Субдиференціал Шора – компактна множина. Проте ця множина не опукла. Тому вона дає більш "індивідуальну"інформацію про функцію, ніж субдиференціал $\partial f(x)$.

В 1973 році Ф. Кларк (студент Р. Т. Рокафеллара) запропонував конструкцію узагальненого градієнту (див., наприклад, [25]).
Величини

$$f_{\underset{Cl}{\uparrow}}(x,g) = \limsup_{[x',\alpha] \to [x,+0]} \frac{1}{\alpha}(f(x' + \alpha g) - f(x')),$$

$$f_{\underset{Cl}{\downarrow}}(x,g) = \liminf_{[x',\alpha] \to [x,+0]} \frac{1}{\alpha}(f(x' + \alpha g) - f(x'))$$

називаються відповідно верхньою і нижньою похідною Кларка. Якщо функція f локально ліпшицева, то ці величини скінченні. У цьому випадку визначимо субдиференціал Кларка за формулою

$$\partial_{Cl} f(x) = \operatorname{conv} \left\{ v \in \mathbb{R}^n | \exists \{x_i\} : v = \lim_{i \to \infty} f'(x_i), x_i \to x, x_i \in Q \right\},$$

тобто $\partial_{Cl} f(x) = \operatorname{conv} \partial_{Sh} f(x)$. Множина $\partial_{Cl} f(x)$ опукла і компактна. Елементи множини $\partial_{Cl} f(x)$ називаються узагальненими градієнтами. Кожен майже градієнт є узагальненим градієнтом. Обернене твердження невірне. Кларк показав, що

$$f_{\underset{Cl}{\uparrow}}(x,g) = \max_{v \in \partial_{Cl} f(x)} \langle v, g \rangle,$$

$$f_{\underset{Cl}{\downarrow}}(x,g) = \min_{w \in \partial_{Cl} f(x)} \langle w, g \rangle.$$

Умови екстремуму субдиференційовної за Кларком функції f можна виразити в термінах субдиференціала Кларка. Умова

$$0 \in \partial_{Cl} f(x^*)$$

є необхідною умовою (локального чи глобального) мінімуму функції f в точці $x^* \in S$. Ця умова є також необхідною умовою (локального чи глобального) максимуму функції f в точці $x^* \in S$. Точка $x^* \in S$, що задовольняє умову, називається стаціонарною за Кларком. Множина стаціонарних за Кларком точок може мати точки, що не являються точками мінімуму чи максимуму. Більшість властивостей, описаних для диференційовних функцій, не справджуються для субдиференційовних за Кларком функцій. Проте субдиференціал Кларка дозволяє виявити деякі властивості функцій. Наприклад, якщо

Якщо $0 \notin \partial_{Cl} f(x)$, то напрямок

$$g_0(x) = -z(x)/\|z(x)\|, \quad \|z(x)\| = \min_{v \in \partial_{Cl} f(x)} \|v\|,$$

є напрямком найшвидшого спуску функції f в точці x, а напрямок $g_1(x) = -g_0(x)$ є напрямком найшвидшого підйому функції f в точці x.

Р. Т. Рокафеллар [50] визначив таку похідну за напрямком

$$f_{\underset{R}{\uparrow}}(x,g) = \limsup_{[x',g',\alpha] \to [x,g,+0]} \frac{1}{\alpha}(f(x' + \alpha g') - f(x')).$$

Якщо функція f локально ліпшицева, то

$$f_{\underset{R}{\uparrow}}(x,g) = f_{\underset{Cl}{\uparrow}}(x,g).$$

П. Мішель та І. П. Пено (див., наприклад, [32]) запропонували використовувати узагальнені похідні

$$f_{\underset{mp}{\uparrow}}(x,g) = \sup_{q \in \mathbb{R}^n} \overline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha}(f(x + \alpha(g + q)) - f(x + \alpha q)),$$

$$f_{\underset{mp}{\downarrow}}(x,g) = \inf_{q \in \mathbb{R}^n} \underline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha}(f(x + \alpha(g + q)) - f(x + \alpha q)),$$

що називаються відповідно верхньою і нижньою похідною Мішеля–Пено. Якщо функція f локально ліпшицева, то існує така опукла і компактна множина $\partial_{mp} f(x) \subset \mathbb{R}^n$, що

$$f^{\uparrow}_{mp}(x,g) = \max_{v \in \partial_{mp} f(x)} \langle v,g \rangle,$$

$$f^{\downarrow}_{mp}(x,g) = \min_{w \in \partial_{mp} f(x)} \langle w,g \rangle.$$

Множину $\partial_{mp} f(x)$ називають також малим субдиференціалом.

Якщо функція $f : \mathbb{R}^n \to \mathbb{R}^1$, задовольняє умову Ліпшиця з константою K в околі точки x, то похідна за напрямком Діні, похідна Кларка та похідна Мішеля–Пено є сублінійними функціями і задовольняють умову

$$f'_D(x,g) \le f_{\underset{mp}{\uparrow}}(x,g) \le f_{\underset{Cl}{\uparrow}}(x,g) \le K\|g\|.$$

Субдиференціали Діні $\partial_D f(x)$ (може бути порожньою множиною), Кларка $\partial_{Cl} f(x)$ (непорожня множина) та Мішеля–Пено $\partial_{mp} f(x)$ (непорожня множина) є компактними опуклими множинами. Вони задовольняють умову

$$\partial_D f(x) \subset \partial_{mp} f(x) \subset \partial_{Cl} f(x) \subset KB(x).$$

Якщо точка x^* є точкою локального мінімуму субдиференційовної функції f, то виконуються необхідні умови

$$0 \in \partial_D f(x) \subset \partial_{mp} f(x) \subset \partial_{Cl} f(x).$$

Інші конструкції субдиференціалів та їх застосування для вивчення властивостей відповідних класів негладких функцій запропонували Ж. Ириарт-Урруті [55], А. Д. Іоффе [21], А. Я. Кругер [27], В. С. Мордухович [40]. Їх спільною рисою є те, що функції досліджуються за допомогою однієї множини (опуклої чи неопуклої).

5. Апроксимації похідної за напрямком множиною опуклих функцій.

Б. Н. Пшеничный [48] визначив поняття верхньої опуклої апроксимації функції $f(x)$ та нижньої угнутої опуклої апроксимації функції $f(x)$.

Нехай $f(x)$, $x \in X$, функція, що приймає скінченні значення і значення $\pm\infty$. Покладемо

$$\operatorname{dom} f = \{x : |f(x)| < +\infty\}.$$

Нехай $x \in \operatorname{dom} f$, $g \in X$, $g \neq 0$. Величина

$$F(x,g) = \sup_{r(\cdot)} \limsup_{\lambda \downarrow 0} \frac{f(x + \lambda g + r(\lambda)) - f(x)}{\lambda},$$

називається похідною за напрямком Пшеничного. Величина $F(x,g)$ – скінченна або нескінченна – існує завжди. Легко перевірити, що функція $F(x,g)$ додатньо однорідна за g, тобто для $\lambda > 0$

$$F(x,\lambda g) = \lambda F(x,g).$$

Функція $h(x,g)$ називається верхньою опуклою апроксимацією (в.о.а.) функції $f(x)$ в точці x, якщо:

1) $h(x,g) \geqslant F(x,g)$ для всіх $g \neq 0$;

2) $h(x,g)$ – опукла замкнута додатньо однорідна функція аргументу g.

Якщо $h(x,g)$ є в.о.а. для f в точці x, то множина

$$\partial h(x,0) = \{x^* \in X^* : h(x,g) \geqslant \langle x^*,g \rangle, g \in X\} \tag{27}$$

називається субдиференціалом функції f в точці x і позначається $\partial f(x)$. Субдиференціал $\partial h(x,0)$ опуклої замкнутої додатньо однорідної функції $h(x,g)$ відносно g існує завжди. При цьому

$$h(x,g) = \sup_{x^*} \{\langle x^*,g \rangle : x^* \in \partial f(x)\}. \tag{28}$$

Функція $h(x,g)$ і субдиференціал $\partial f(x)$ однозначно визначають одне одного. Якщо $\partial f(x)$– субдиференціал, то $\partial f(x)$– замкнута опукла множина, а функція $h(x,g)$ є в.о.а. для f в точці x. Як і в.о.а., субдиференціал визначений неоднозначно.
У третьому розділі функції досліджуються за допомогою в. о. а.

6. Квазідиференційовні функції.

Визначена на відкритій множині S у просторі $\mathbb{R}^n$ функція $f(x)$ називається квазідиференційовною в точці $x \in S$, якщо вона диференційовна за напрямками в цій точці і її похідну $f'(x,g)$ можна подати у вигляді

$$f'(x,g) = \max_{h \in U}(h,g) + \min_{h \in V}(h,g), \quad g \in \mathbb{R}^n, \tag{29}$$

де U,V – опуклі компактні множини в $\mathbb{R}^n$. Пара опуклих компактів $[U,V]$, що використовується для представлення похідної, визначена неєдиним чином. Клас еквівалентних пар опуклих компактів $[U,V]$, які мають властивість (29) називається *квазідиференціалом* функції f в точці x і позначається символом $\mathfrak{D}f(x)$. Квазідиференційовність функції f у точці x означає, що функція f має похідну за напрямками (похідну Діні) у цій точці і вказана похідна може бути виражена через лінійні функції (квазілінеаризована) за допомогою квазідиференціала $\mathfrak{D}f(x) = [\underline{\partial} f(x),\overline{\partial} f(x)]$.

7. Кодиференційовні функції.

Функція $f(x)$ *кодиференційовна* в точці $x \in X \subset \mathbb{R}^n$, якщо існують опуклі компакти $\underline{d}f(x) \subset \mathbb{R}^{n+1}$ і $\overline{d}f(x) \subset \mathbb{R}^{n+1}$ такі, що

$$f(x+\Delta) = f(x) + \Phi_x(\Delta) + o_x(\Delta), \tag{30}$$

$$\Phi_x(\Delta) = \max_{[a,v] \in \underline{d}f(x)} [a + \langle v,\Delta\rangle] + \min_{[b,w] \in \overline{d}f(x)} [b + \langle w,\Delta\rangle], \tag{31}$$

$$\frac{o_x(\alpha\Delta)}{\alpha} \underset{\alpha \downarrow 0}{\longrightarrow} 0, \quad \forall \Delta \in R^n. \tag{32}$$

де $a,b \in \mathbb{R}^1$; $v,w \in \mathbb{R}^n$, $\operatorname{conv}\{x,x+\Delta\} \subset X$.

Пара множин $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$ називається *кодиференціалом* функції f в точці x, множина $\underline{d}f(x)$ називається *гіподиференціалом*, а множина $\overline{d}f(x)$ — *гіпердиференціалом* функції f у точці x. Кодиференціал, як і квазідиференціал, визначається неоднозначно.

Кодиференційовна в точці x функція $f(x)$ допускає апроксимацію першого порядку в околі точки x.

Детально властивості квазідиференційовних та кодиференційовних функцій описані в книгах [18], [17].

Розділ I
Опуклі множини

1.1. ОЗНАЧЕННЯ ТА ОСНОВНІ ПОНЯТТЯ

Означення 1.1.1. *Відрізком*, що з'єднує точки x^1, x^2 $n-$вимірного лінійного простору $\mathbb{R}^n$, називається множина

$$[x^1,x^2] = \{x \in \mathbb{R}^n \colon x = \lambda x^1 + (1-\lambda)x^2, 0 \le \lambda \le 1\}.$$

Означення 1.1.2. Непорожня множина $X \subset \mathbb{R}^n$ називається *опуклою*, якщо разом з будь-якими двома точками вона містить і відрізок, що з'єднує ці точки. Порожню множину $\emptyset$ будемо вважати опуклою.

Прикладами опуклої множини у просторі $\mathbb{R}^1$ є одноточкові множини, інтервали, півпрямі, прямі. Прикладами опуклої множини в просторі $\mathbb{R}^n$ є сам простір, будь-який його підпростір, одноточкові множини, куля, відрізок, а також:

1) пряма, що проходить через точку $\hat{x}$ в напрямку h

$$l_{\hat{x}h} = \{x \in \mathbb{R}^n \colon x = \hat{x} + \alpha h\,, \alpha \in \mathbb{R}\};$$

2) промінь, що виходить з точки $\hat{x}$ в напрямку h

$$l^+_{\hat{x}h} = \{x \in \mathbb{R}^n \colon x = \hat{x} + \alpha h\,, \alpha \in \mathbb{R}\,, \alpha \ge 0\};$$

3) гіперплощина з нормаллю p

$$H_{p\beta} = \{x \in \mathbb{R}^n \colon \langle p,x\rangle = \beta\};$$

4) півпростори, що породжені цією гіперплощиною

$$H^+_{p\beta} = \{x \in \mathbb{R}^n \colon \langle p,x\rangle \ge \beta\};$$

$$H^-_{p\beta} = \{x \in \mathbb{R}^n \colon \langle p,x\rangle \le \beta\}.$$

Лема 1.1.1. *Нехай I – множина індексів (скінченна чи нескінченна), та нехай $X_i\,, i \in I$, – опуклі множини. Перетин $X = \cap_{i\in I} X_i$ опуклих множин є опуклою множиною.*

Лема 1.1.2. *Нехай $X_1, \ldots, X_m$ - опуклі множини, $a_1, \ldots, a_m$ - довільні числа. Тоді опукла множина*

$$\sum_{i=1}^m a_i X_i = \left\{ x \,\middle|\, x = \sum_{i=1}^m a_i x^i,\ x^i \in X_i,\ \ i = 1, \ldots, m \right\},$$

що називається лінійною комбінацією множин $X_1, \ldots, X_m$.

Як наслідок, опуклими є сума і різниця опуклих множин X_1,X_2:
$$X_1 \pm X_2 = \left\{x | x = x^1 \pm x^2, x^1 \in X_1, x^2 \in X_2\right\}.$$

Означення 1.1.3. Множина $K \subset \mathbb{R}^n$ називається:
а) *конусом*, якщо з того, що $x \in K$ і $\lambda > 0$, випливає, що $\lambda x \in K$;
б) *опуклим конусом*, якщо з того, що $x^1,x^2 \in K$ і $\lambda_1 > 0$, $\lambda_2 > 0$, випливає, що $\lambda_1 x^1 + \lambda_2 x^2 \in K$.

Означення 1.1.4. Множина $X \subset \mathbb{R}^n$ називається *афінною*, якщо $\lambda x^1 + (1-\lambda)\, x^2 \in X$ для всіх $x^1, x^2 \in X, \lambda \in \mathbb{R}$, тобто якщо множина X разом із своїми двома точками x^1, x^2 містить пряму, що проходить через ці точки.

Афінні множини мають зовсім просту стуктуру: вони є зсувами лінійних підпросторів.
Очевидно, що для афінних множин та конусів справедливі аналоги лем 1.1.1 та 1.1.2.

1.2. КОМБІНАЦІЇ ТОЧОК ТА ОБОЛОНКИ МНОЖИН

Означення 1.2.1. Нехай $x^1, \ldots, x^m$ – точки з $\mathbb{R}^n$. Комбінація $\sum\limits_{i=1}^{m} \lambda_i x^i$ точок називається:
1) *опуклою*, якщо $\lambda_i \geq 0, i = 1, \ldots, m, \quad \sum\limits_{i=1}^{m} \lambda_i = 1$;
2) *конічною*, якщо $\lambda_i \geq 0, \quad i = 1, \ldots, m$;
3) *афінною*, якщо $\sum\limits_{i=1}^{m} \lambda_i = 1$;
4) *лінійною*, якщо λ_i – числа з $\mathbb{R}^1$.

Лема 1.2.1. *Опукла множина (опуклий конус, афінна множина, лінійний підпростір) містять всі можливі опуклі (конічні, афінні, лінійні) комбінації своїх точок.*

Доведення. Доведемо твердження для опуклої множини X. Треба показати, що для довільного $m = 1,2, \ldots$ з умов
$$x = \sum_{i=1}^{m} \lambda_i x^i, x^i \in X, \lambda_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^{m} \lambda_i = 1, \tag{1.2.1}$$
випливає $x \in X$. Проведемо індукцію за m. Якщо $m = 1$, то випадок тривіальний. Припустимо, що твердження доведене для $m = k$, і нехай співвідношення (1.2.1) справджується при $m = k + 1$. Якщо $\lambda_{k+1} = 1$, то $\lambda_1 = \ldots = \lambda_k = 0$, і, відповідно, $x = x^{k+1} \in X$. Якщо ж $\lambda_{k+1} < 1$, то ми

можемо записати

$$x = (1 - \lambda_{k+1})\overline{x} + \lambda_{k+1}x^{k+1}, \ \ \overline{x} = \sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} x^i. \qquad (1.2.2)$$

Точка $\overline{x}$ – це опукла комбінація точок $x^1, \ldots, x^k$. Тоді за припущенням індукції $\overline{x} \in X$. З (1.2.2) з урахуванням опуклості X випливає, що $x \in X$. □

Означення 1.2.2. Перетин всіх опуклих множин (опуклих конусів, афінних множин, лінійних підпросторів), що містять дану множину X, називається *опуклою (конічною, афінною, лінійною) оболонкою* множини X и позначається $\operatorname{conv} X$ $(\operatorname{cone} X, \operatorname{aff} X, \operatorname{Lin} X)$.

Означення 1.2.3. $\operatorname{Lin} X$ – це паралельний до афінної оболонки X лінійний підпростір.

Таким чином $\operatorname{Lin} X = \operatorname{aff} X - x^0$, де x^0 - довільна точка X, або навіть з $\operatorname{aff} X$, причому $\operatorname{Lin} X$ визначається однозначно.

Лема 1.2.2. *Лінійний підпростір $Lin\, X$ має такі властивості:*
1) якщо $x^1, x^2 \in X$, то $x^1 - x^2 \in Lin\, X$;
2) якщо $x \in X$, h^1, $h^2 \in Lin\, X$, $\alpha_1, \alpha_2 \in \mathbb{R}$, то $x + \alpha_1 h^1 + \alpha_2 h^2 \in aff\, X$;
3) якщо $0 \in X$, то $Lin\, X = aff\, X$.

Означення 1.2.4. Розмірність простору $\operatorname{Lin} X$ називається *розмірністю опуклої множини X* и позначається $\dim X$.

Лема 1.2.3. *Опукла (конічна, афінна, лінійна) оболонка довільної множини X співпадає з множиною всіх опуклих (конічних, афінних, лінійних) комбінацій точок з X.*

Теорема 1.2.1. Теорема Каратеодорі. *Нехай* $\dim \operatorname{Lin} X = n$. *Кожну точку, що належить опуклій (конічній, афінній, лінійній) оболонці множини X можна подати у вигляді опуклої (конічної, афінної, лінійної) комбінації не більш ніж r точок $x^1, \ldots, x^r$ із X. Для опуклої і афінної оболонки $r = n + 1$ причому точки $x^1, \ldots, x^r$ можна вважати афінно незалежними. Для конічної і лінійної оболонки $r = n$ причому точки $x^1, \ldots, x^r$ можна вважати лінійно незалежними.*

Доведення. Доведемо твердження для опуклої оболонки множини X. Зрозуміло, що центральним місцем у теоремі є твердження про те, що $r \le n + 1$. Візьмемо точку вигляду (1.2.1) і покажемо, що число ненульових доданків у сумі (1.2.1) можна зменшити, якщо $m > n + 1$. Достатньо припустити, що всі $\lambda_i > 0$. Візьмемо $(n+1)$-вимірні вектори $(x^i, 1), i = 1, \ldots, m$, у яких перші n компонент збігаються з відповідними компонентами вектора x^i, а остання компонента дорівнює 1. Оскільки число таких векторів $m > n+1$, то вони лінійно залежні. Тому знайдуться не всі рівні нулю числа $\alpha_i, i = 1, \ldots, m$, такі, що

$$\sum_{i=1}^{m} \alpha_i x^i = 0, \quad \sum_{i=1}^{m} \alpha_i = 0.$$

Серед чисел α_i обов'язково є додатні в силу останнього співвідношення. Покладемо

$$\varepsilon_0 = \min \left\{ \frac{\lambda_i}{\alpha_i} : \alpha_i > 0, i = 1, \ldots, m \right\}.$$

Нехай мінімум досягається при деякому $i = i_0$. Тоді

$$\bar{\lambda}_i = \lambda_i - \varepsilon_0 \alpha_i \geq 0, \quad i = 1, \ldots, m.$$

Це очевидно для $\alpha_i \leq 0$, а для $\alpha_i > 0$ випливає з вибору ε_0. Тепер із співвідношень

$$\sum_{i=1}^{m} \bar{\lambda}_i x^i = \sum_{i=1}^{m} \lambda_i x^i - \varepsilon_0 \sum_{i=1}^{m} \alpha_i x^i = x,$$

$$\sum_{i=1}^{m} \bar{\lambda}_i = \sum_{i=1}^{m} \lambda_i - \varepsilon_0 \sum_{i=1}^{m} \alpha_i = 1$$

випливає, що точку x можна подати у вигляді опуклої комбінації меншої кількості ненульових доданків. Так можна діяти доти, доки $m > n+1$. Звідси і випливає твердження теореми. $\square$

Наслідок 1.2.1. *Якщо X – компактна множина в $\mathbb{R}^n$, то $\operatorname{conv} X$ також компактна множина.*

Доведення. Розглянемо декартів добуток $Y = \Lambda \times X \times \cdots \times X$, де X береться $n+1$ раз, а

$$\Lambda = \left\{ \lambda \in \mathbb{R}^{n+1} \middle| \lambda = (\lambda_1, \ldots, \lambda_{n+1}) \geq 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1 \right\}.$$

Зрозуміло, що Y - це компакт. Визначимо відображення $f : Y \to \mathbb{R}^n$ за формулою $f(y) = \sum\limits_{i=1}^{n+1} \lambda_i x^i$, $y = \left(\lambda, x^1, \ldots, x^{n+1}\right) \in Y$. Воно неперервне на Y. Теорема Каратеодорі стверджує, що $\operatorname{conv} X$ є образом Y при відображенні f, тобто $\operatorname{conv} X = f(Y)$. Але, як відомо, образ компактної множини є компактна множина. $\square$

1.2.1. Замкнуті опуклі оболонки

Перетин будь-якої кількості опуклих множин є множина опукла. Таку ж властивість мають і замкнуті множини. Тому доцільно ввести таке означення.

Означення 1.2.5. Перетин всіх замкнутих опуклих множин, які містять дану множину X, називається *замкнутою опуклою оболонкою* X і позначається $\overline{\mathrm{conv}}\, X$.

Теорема 1.2.2. $\overline{\mathrm{conv}}\, X = \overline{\mathrm{conv}\, X}$.

Доведення. Зрозуміло, що $\overline{\mathrm{conv}}\, X \supseteq \mathrm{conv}\, X$ тому що в процесі утворення $\mathrm{conv}\, X$ беруть участь всі опуклі множини, а не тільки замкнуті. Звідси випливає, що $\overline{\mathrm{conv}}\, X \supseteq \overline{\mathrm{conv}\, X}$. І навпаки, $\overline{\mathrm{conv}\, X}$ є опукла замкнута множина. Тому $\overline{\mathrm{conv}}\, X \subseteq \overline{\mathrm{conv}\, X}$, що завершує доведення. □

Теорема 1.2.3. *Опукла оболонка компакту є компакт.*

Доведення. Нагадаємо, що в $\mathbb{R}^n$ компакт є замкнута обмежена множина. Якщо $x \in \mathrm{conv}\, X$, де X - компакт, то за теоремою 1.2.1 (Каратеодорі)

$$x = \sum_{i=1}^{n+1} \lambda_i x^i, \quad x^i \in X, \quad \lambda_i \geq 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1.$$

Тому

$$\|x\| \leq \sum_{i=1}^{n+1} \lambda_i \left\|x^i\right\| \leq c,$$

де c - така константа, що $\|x\| \leq c$ для будь-якого $x \in X$. Отже, $\mathrm{conv}\, X$ – обмежена множина. Покажемо, що вона замкнута. Нехай

$$x^k = \sum_{i=1}^{n+1} \lambda_{ik} x^{ik}, \quad x^{ik} \in X, \quad \sum_{i=1}^{n+1} \lambda_{ik} = 1.$$

Оскільки послідовності λ_{ik} та x^{ik} обмежені, то з них можна вибрати збіжні підпослідовності. Можна вважати, що $\lambda_{ik} \to \lambda_{i0}$ та $x^{ik} \to x^{i0} \in X$, оскільки X –компакт. Переходячи до границі, отримаємо

$$x^0 = \sum_{i=1}^{n+1} \lambda_{i0} x^{i0}, \quad x^{i0} \in X, \quad \sum_{i=1}^{n+1} \lambda_{i0} = 1.$$

Це означає, що $x^0 \in \mathrm{conv}\, X$. Замкнутість $\mathrm{conv}\, X$ доведена. □

1.3. ТОПОЛОГІЧНІ ВЛАСТИВОСТІ ОПУКЛИХ МНОЖИН

Означення 1.3.1. Точка $x \in X$ називається *внутрішньою точкою* множини X, якщо існує таке $\varepsilon > 0$, що $x + \varepsilon B \subseteq X$, де B - одинична куля в $\mathbb{R}^n$ з центром у початку координат, тобто $B = \{x : \|x\| < 1\}$, $\|x\| = \langle x,x\rangle^{1/2}$. Множина таких точок називається *множиною внутрішніх точок* множини X і позначається $\operatorname{int} X$.

Означення 1.3.2. Точка x називається *граничною точкою* множини X, якщо існує послідовність точок $x^k \in X$, що збігається до x. Сукупність усіх граничних точок множини X називається *замиканням* множини X і позначається $\overline{X}$.

Лема 1.3.1. *Замикання $\overline{X}$ і множина внутрішніх точок $\operatorname{int} X$ опуклої множини X опуклі. При цьому їх афінні оболонки співпадають $\operatorname{aff} X = \operatorname{aff} \overline{X}$.*

Доведення. Якщо X опукла, то із того що $x^1 \in \operatorname{int} X$, $x^2 \in \operatorname{int} X$ випливають включення $x^1 + \varepsilon_1 B \subseteq X$, $x^2 + \varepsilon_2 B \subseteq X$. Нехай $\lambda_1 x^1 + \lambda_2 x^2$ – опукла комбінація точок x^1 і x^2. Тоді $\lambda_1(x^1 + \varepsilon_1 B) + \lambda_2(x^2 + \varepsilon_2 B) = \lambda_1 x^1 + \lambda_2 x^2 + (\lambda_1\varepsilon_1 + \lambda_2\varepsilon_2)B \subseteq X$, тобто $\lambda_1 x^1 + \lambda_2 x^2 \in \operatorname{int} X$. Якщо $x^1, x^2 \in \overline{X}$, то за означенням існують послідовності точок $x^{1k}, x^{2k} \in X$ такі, що $x^{1k} \to x^1$, $x^{2k} \to x^2$. Нехай $\lambda_1 x^{1k} + \lambda_2 x^{2k}$ – опукла комбінація точок x^{1k}, x^{2k}. Тоді $\lambda_1 x^1 + \lambda_2 x^2 = \lim\limits_{k\to\infty} (\lambda_1 x^{1k} + \lambda_2 x^{2k}) \in \overline{X}$, оскільки $\lambda_1 x^{1k} + \lambda_2 x^{2k} \in X$ в силу опуклості останньої.
Афінна множина $\operatorname{aff} X$ замкнута. Тому із співвідношення $X \subset \operatorname{aff} X$ випливає, що $\overline{X} \subset \operatorname{aff} X$. Тому $\operatorname{aff} \overline{X} \subset \operatorname{aff} X$. Обернене включення очевидне. □

Опуклі множини мають ту властивість, що в певному розумінні їх завжди можна помістити в підпростір, в якому вони вже мають внутрішні точки.

Теорема 1.3.1. *Опукла множина X, що лежить в $\mathbb{R}^n$, або має внутрішні точки, або знаходиться в підпросторі меншої розмірності, зміщеному на деякий вектор.*

Доведення. Нехай $x^0 \in X$. Розглянемо всі вектори вигляду $x - x^0, x \in X$. Серед таких векторів є $r \le n$ лінійно незалежних: $x^1 - x^0, \ldots, x^r - x^0$. Можливі два випадки.

а) $r = n$. У цьому випадку є n лінійно незалежних векторів $x^i - x^0, x^i \in X; i = 1, \ldots, n$. Розглянемо множину

$$S^n = \left\{x | x = \lambda_0 x^0 + \cdots + \lambda_n x^n \colon \lambda_i \ge 0, i = 1, \ldots, m; \lambda_0 + \cdots + \lambda_n = 1\right\}.$$

Множина S^n називається *n–вимірним симплексом* породженим точками $x^0,x^1,\ldots,x^n$. За лемою 1.2.1 маємо $S^n \subseteq X$. Якщо буде доведено, що S^n має внутрішні точки, то їх буде мати і множина X. Доведемо, що будь-яка точка $x \in S^n$ із строго додатними коефіцієнтами λ_i належить $\operatorname{int} S^n$. Розглянемо систему рівнянь відносно $\lambda_i, i = 1,\ldots,n$:

$$x - x^0 = \sum_{i=1}^{n} \lambda_i (x^i - x^0).$$

Оскільки вектори $x^i - x^0$ лінійно незалежні, то ця система має єдиний розв'язок $\lambda_i(x), i = 1,\ldots,n$, який неперервно залежить від x (пригадаємо формули Крамера для систем рівнянь з невиродженим визначником). Тому, вважаючи x рівним $\overline{x} = \bar{\lambda}_0 x^0 + \cdots + \bar{\lambda}_n x^n$, $\bar{\lambda}_i > 0$, одержимо, що $\lambda_i\left(\overline{x}\right) = \bar{\lambda}_i > 0\,, i = 1,\ldots,n$, і $\lambda_0 = 1 - \sum\limits_{i=1}^{n} \bar{\lambda}_i > 0$. Звідси випливає, що $\lambda_i\left(x\right) > 0\,, i = 1,\ldots,n$, для всіх x з деякого околу $\overline{x}$ та

$$\lambda_0\left(x\right) = 1 - \sum_{i=1}^{n} \lambda_i\left(x\right) > 0.$$

Тому для всіх точок x з деякого околу $\bar{x}$ справедливе включення

$$x = \sum_{i=1}^{n} \lambda_i\left(x\right) x^i \in S^n,$$

що доводить першу частину теореми.

б) $r < n$. Розглянемо підпростір X^0, що складається з векторів

$$y = \sum_{i=1}^{r} \alpha_i \left(x^i - x^0\right).$$

За побудовою $X - x^0 \subseteq X^0$, тобто $X \subseteq x^0 + X^0$.
Побудований в теоремі підпростір X^0 r-вимірний і в ньому множина $X - x^0$ містить внутрішні точки. Це можна показати так само, як при доведенні теореми 1.3.1. Далі, X^0 не залежить від вибору точки x^0 і векторів $x^i - x^0\,, i = 1,\ldots,r$. Справді, будь-який підпростір, що містить $X - x^0$, має містити вектори $x^i - x^0$, а значить і весь X^0. Звідси випливає, що X^0 є перетином усіх підпросторів, що містять $X - x^0$. Якщо підпростір X^1 містить $X - x^0$ для деякого $x^0 \in X$, то він містить і $X - \overline{x}^0$ для будь-якої іншої точки. Дійсно, $x - \overline{x}^0 = (x - x^0) - (\overline{x}^0 - x^0)$, і оскільки X^1 - підпростір, то він містить різницю будь-яких двох своїх векторів. Таким чином, X^1 містить $X - x^0$ і $X - \overline{x}^0$ одночасно, тобто X не залежить від вибору x^0. $\square$

Тепер можна дати таке означення.

Означення 1.3.3. Точка x називається *відносно внутрішньою точкою* опуклої множини X, якщо $x + \operatorname{Lin} X \cap (\varepsilon B) \subseteq X$, тобто x міститься в X

разом із кулею радіуса $\varepsilon > 0$, яка лежить в $\mathrm{Lin}\, X$. Множина таких точок називається *множиною відносно внутрішніх точок* опуклої множини X і позначається $\mathrm{ri}\, X$. Множина X називається відносно відкритою, якщо $X = \mathrm{ri}\, X$.

Лема 1.3.2. $ri\, \overline{X} = ri\, X$.

Доведення. Оскільки $\mathrm{Lin}\, X$ – це замкнута множина, що містить X, то $\mathrm{Lin}\, X \supseteq \overline{X}$. Легко бачити, що $\mathrm{Lin}\, X = \mathrm{Lin}\overline{X}$. Очевидно, що $\mathrm{ri}\, \overline{X} \supseteq \mathrm{ri}\, X$. Доведемо протилежне включення. Нехай $x \in \mathrm{ri}\, \overline{X}$, а $e^1, \ldots, e^r$– базис у $\mathrm{Lin}\, X$. Тоді при малому ε матимемо

$$y^k = x + \varepsilon \left(e^k - \frac{1}{r+1} e \right) \in \overline{X}, k = 1, \ldots, r;$$

$$y^0 = x - \frac{\varepsilon}{r+1} e \in \overline{X},$$

де $e = e^1 + \cdots + e^r$. Вектори $y^k - y^0 = \varepsilon e^k$ лінійно незалежні і

$$x = \frac{1}{r+1} y^0 + \cdots + \frac{1}{r+1} y^r.$$

Остання рівність означає, що x є внутрішня точка симплекса, породженого точками $y^0, \ldots, y^r$. Якщо взяти достатньо близькі до y^k точки $\overline{y}^k \in X$, то виявляється, що $x \in X$ є внутрішня точка симплекса породженого точками $\overline{y}^k \in X$ і є відносною внутрішньою точкою X. Тому $\mathrm{ri}\, \overline{X} \subseteq \mathrm{ri}\, X$. $\square$

Теорема 1.3.2. *Нехай X - опукла множина. Якщо $x^1 \in \overline{X}, x^2 \in ri\, X$, то при всіх $\lambda \in (0,1]$ точка $(1-\lambda)x^1 + \lambda x^2 \in ri\, X$. Крім того, $\overline{X} = \overline{ri\, X}$.*

Доведення. Якщо $x^2 \in \mathrm{ri}\, X$, то $x^2 + \mathrm{Lin}\, X \cap (\varepsilon B) \subseteq X$. Тому з опуклості X випливає, що

$$(1-\lambda)x^1 + \lambda(x^2 + \mathrm{Lin}\, X \cap (\varepsilon B)) =$$

$$(1-\lambda)x^1 + \lambda x^2 + \mathrm{Lin}\, X \cap (\lambda \varepsilon B) \subseteq \overline{X}.$$

Тобто $(1-\lambda)x^1 + \lambda x^2 \in \mathrm{ri}\, X$ оскільки $\mathrm{ri}\, X = \mathrm{ri}\, \overline{X}$ за лемою 1.3.2.
Нехай $x^0 \in \overline{X}$, $x^k \in X$, $x^k \to x^0$, $\lambda_k \to 0$. Тоді $(1-\lambda_k)x^k + \lambda_k y \in \mathrm{ri}\, X$ якщо $y \in \mathrm{ri}\, X$. Тому x^0 є гранична точка для $\mathrm{ri}\, X$. Отже $\overline{X} \subseteq \overline{\mathrm{ri}\, X}$. Обернене включення очевидне. $\square$

Теорема 1.3.3. *Якщо для опуклих множин X_1 і X_2 виконується умова $ri\, X_1 \cap ri\, X_2 \neq \emptyset$, то*

$$Lin\, X_1 \cap Lin\, X_2 = Lin\, (X_1 \cap X_2),$$

$$ri\, X_1 \cap ri\, X_2 = ri\, (X_1 \cap X_2).$$

Доведення. Будемо вважати, що $0 \in \operatorname{ri} X_1 \cap \operatorname{ri} X_2$. В цьому випадку $\operatorname{Lin} X_1 \supseteq X_1$, $\operatorname{Lin} X_2 \supseteq X_2$. Тому

$$\operatorname{Lin} X_1 \cap \operatorname{Lin} X_2 \supseteq X_1 \cap X_2,$$

і

$$\operatorname{Lin} X_1 \cap \operatorname{Lin} X_2 \supseteq \operatorname{Lin} (X_1 \cap X_2).$$

Навпаки, нехай $z \in \operatorname{Lin} X_1 \cap \operatorname{Lin} X_2$. Тоді при досить малому $\lambda > 0, \lambda z \in X_1$ і $\lambda z \in X_2$, тому що $0 \in \operatorname{ri} X_1 \cap \operatorname{ri} X_2$. Звідси випливає, що $\lambda z \in X_1 \cap X_2$. Отже $\lambda z \in \operatorname{Lin} (X_1 \cap X_2)$. Оскільки $\operatorname{Lin} (X_1 \cap X_2)$ – це підпростір, то $z \in \operatorname{Lin} (X_1 \cap X_2)$.

Доведемо другу частину твердження. Якщо $x \in \operatorname{ri} X_1 \cap \operatorname{ri} X_2$, то $x + \operatorname{Lin} X_1 \cap (\varepsilon B) \subseteq X_1$, $x + \operatorname{Lin} X_2 \cap (\varepsilon B) \subseteq X_2$, так що $(x + \operatorname{Lin} X_1 \cap (\varepsilon B)) \cap (x + \operatorname{Lin} X_2 \cap (\varepsilon B)) = x + \operatorname{Lin} (X_1 \cap X_2) \cap (\varepsilon B) \subseteq X_1 \cap X_2$ і $x \in \operatorname{ri} (X_1 \cap X_2)$. Тому $\operatorname{ri} (X_1 \cap X_2) \supseteq \operatorname{ri} X_1 \cap \operatorname{ri} X_2$.
Нехай тепер $x \in \operatorname{ri} (X_1 \cap X_2)$. Оскільки $0 \in \operatorname{ri} X_j$, то $(1-\lambda)x \in \operatorname{ri} X_j, j = 1,2$ для $0 < \lambda \leq 1$. Тому $(1-\lambda)x \in \operatorname{ri} X_1 \cap \operatorname{ri} X_2$. Спрямовуючи λ до нуля, одержуємо $x \in \overline{\operatorname{ri} X_1 \cap \operatorname{ri} X_2}$. Таким чином,

$$\operatorname{ri} X_1 \cap \operatorname{ri} X_2 \subseteq \operatorname{ri} (X_1 \cap X_2) \subseteq \overline{\operatorname{ri} X_1 \cap \operatorname{ri} X_2}.$$

Нехай $e^1, e^2, \ldots, e^r$ – базис в $\operatorname{Lin} (X_1 \cap X_2)$, $e = e^1 + \cdots + e^r$, $x \in \operatorname{ri} (X_1 \cap X_2)$. Тоді при досить малому $\varepsilon > 0$ всі точки

$$y^k = x + \varepsilon \left(e^k - \frac{1}{r+1} e \right), \; k = 1, \ldots, r, y^0 = x - \frac{\varepsilon}{r+1} e$$

належать $\operatorname{ri} (X_1 \cap X_2)$. Отже належать $\overline{\operatorname{ri} X_1 \cap \operatorname{ri} X_2}$. У той же час

$$x = \frac{1}{r+1} y^0 + \cdots + \frac{1}{r+1} y^r$$

є внутрішньою точкою симплекса, породженого точками $y^0, \ldots, y^r$ відносно підпростору $\operatorname{Lin} (X_1 \cap X_2)$. Якщо взяти досить близькими до y^k точки $\overline{y}^k \in \operatorname{ri} X_1 \cap \operatorname{ri} X_2$ то ясно, що точка x буде внутрішньою точкою симплекса, породженого точками $\overline{y}^k$, а тому буде належати $\operatorname{ri} X_1 \cap \operatorname{ri} X_2$ Тим самим доведено, що будь-яка точка x із $\operatorname{ri} (X_1 \cap X_2)$ належить і $\operatorname{ri} X_1 \cap \operatorname{ri} X_2$. Разом з попереднім це означає, що

$$\operatorname{ri} (X_1 \cap X_2) = \operatorname{ri} X_1 \cap \operatorname{ri} X_2.$$

що і потрібно було довести. □

Теорема 1.3.4. *Нехай X - опукла множина і нехай $x^0 \in \overline{X}$, але $x^0 \notin X$. Тоді у будь-якому околі x^0 знайдуться точки, що не належать $\overline{X}$.*

Доведення. Візьмемо точку $y \in \operatorname{ri} X$. Тоді точки променя

$$l^+_{y, x^0 - y} = \{x \colon x = y + \lambda(x^0 - y), \lambda \geq 0\}$$

при $\lambda > 1$ не належать $\overline{X}$. Справді, якщо при $\lambda > 1$ точка $x^1 = y + \lambda(x^0 - y) \in \overline{X}$, то

$$x^0 = \frac{1}{\lambda}x^1 + \left(1 - \frac{1}{\lambda}\right) y \in \operatorname{ri} X$$

за теоремою 1.3.2, що суперечить тому, що $x^0 \notin X$. $\square$

Теорема 1.3.5. *Нехай X - необмежена замкнута опукла множина в $\mathbb{R}^n$. Тоді:*
1) для будь-якої точки $x^0 \in X$ існує ненульовий вектор $h \in \mathbb{R}^n$ такий, що промінь

$$l^+_{x^0h} = \{x \in \mathbb{R}^n \colon x = x^0 + \alpha h, \alpha \geq 0\}$$

лежить в X;
2) якщо $l^+_{x^0h} \subset X$ при деякому $x^0 \in X$, то $l^+_{xh} \subset X$ при всіх $x \in X$; іншими словами, якщо деякий промінь лежить в X, то промінь з початком в будь-якій точці $x \in X$ у тому ж напрямку h також лежить в X.

Доведення. Нехай $x^0 \in X$. Множина X необмежена. Тому існує послідовність $x^k \in X, k = 1,2\ldots$ така, що $\|x^k\| \to \infty$. Покладемо для $\alpha \geq 0, k = 1,2,\ldots$

$$h^k = \frac{x^k - x^0}{\|x^k - x^0\|}, \lambda_k = \frac{\alpha}{\|x^k - x^0\|}, \overline{x}^k = \lambda_k x^k + (1 - \lambda_k)x^0.$$

Тоді $\|h^k\| = 1$ і можна вважати, що послідовність $\{h^k\}_{k=1}^{\infty}$ збігається до вектора $h \neq 0$. При досить великих k маємо $0 \leq \lambda_k \leq 1$. Оскільки множина X опукла, то для таких k справджується включення $\overline{x}^k \in X$. В той же час $\overline{x}^k = x^0 + \lambda_k(x^k - x^0) = x^0 + \alpha h^k$. Отже $\overline{x}^k \to x^0 + \alpha h$. В силу замкнутості множини X при всіх $\alpha \geq 0$ матимемо $x^0 + \alpha h \in X$, тобто $l^+_{x^0h} \subset X$.
Нехай тепер $l^+_{x^0h} \subset X$ і $x \in X$. При всіх $\alpha \geq 0$ та $k = 1,2,\ldots$ покладемо

$$x^k = x^0 + (\alpha k)h, \overline{x}^k = \frac{1}{k}x^k + \left(1 - \frac{1}{k}\right) x.$$

Тоді $x^k \in l^+_{x^0h} \subset X$ і $\overline{x}^k \in X$ при всіх $k = 1,2\ldots.$ В той же час $\overline{x}^k = x + (x^k - x)/k = x + \alpha h + (x^0 - x)/k$. Тому $\overline{x}^k \to x + \alpha h$. Отже $x + \alpha h \in X$ при всіх $\alpha \geq 0$, тобто $l^+_{xh} \subset X$. $\square$

1.4. РОЗДІЛЕННЯ ДВОХ МНОЖИН

Теорема 1.4.1. *Нехай M – опукла множина і точка $x_0 \notin \overline{M}$. Тоді існує $x^* \neq 0$ і число $\varepsilon > 0$ такі, що*

$$\langle x,x^* \rangle \le \langle x_0,x^* \rangle - \varepsilon$$

для всіх $x \in M$.

Доведення. Нехай $y \in \overline{M}$ – найближча точка до x_0. Оскільки множина замкнута, то така точка існує і задовольняє нерівність

$$\|x - x_0\| \ge \|y - x_0\|, \quad x \in \overline{M}.$$

Точка $\lambda x + (1-\lambda)y = y + \lambda(x-y) \in \overline{M}$ для всіх $\lambda \in [0,1]$ в силу опуклості M. Тому

$$\|\lambda x + (1-\lambda)y - x_0\| = \langle y - x_0 + \lambda(x-y), y - x_0 + \lambda(x-y) \rangle =$$
$$= \|y - x_0\|^2 + 2\lambda\langle x - y, y - x_0\rangle + \lambda^2\|x-y\|^2 \ge \|y - x_0\|^2.$$

Звідси отримаємо

$$2\langle x - y, y - x_0 \rangle + \lambda\|x - y\| \ge 0, \quad \lambda \in [0,1].$$

При $\lambda = 0$ матимемо

$$\langle x - y, y - x_0 \rangle \ge 0.$$

Покладемо $x^* = x_0 - y$, $\varepsilon = \|x^*\|^2$. Оскільки $x_0 \notin \overline{M}$, $y \neq x_0$ і $x^* \neq 0$. Тому $\varepsilon > 0$. Останню нерівність можна записати так

$$\langle x,x^* \rangle \le \langle y,x^* \rangle = \langle x_0,x^* \rangle - \langle x^*,x^* \rangle = \langle x_0,x^* \rangle - \varepsilon.$$

Оскільки точка x була вибрана довільно, то доведення завершено. □

Теорема 1.4.2. *Нехай M – опукла множина і точка $x_0 \notin M$. Тоді існує $x^* \neq 0$, що задовольняє нерівності*

$$\langle x,x^* \rangle \le \langle x_0,x^* \rangle$$

для всіх $x \in M$.

Доведення. Якщо $x_0 \notin \overline{M}$, то твердження є наслідком попередньої теореми. Якщо $x_0 \in \overline{M}$, то за теоремою 1.3.4 існує послідовність $x_k \to x_0$, $x_k \notin \overline{M}$. Застосуємо до M та x_k попередню теорему. Отримаємо

$$\langle x,x_k^* \rangle \le \langle x_k,x_k^* \rangle - \varepsilon_k, \quad \varepsilon_k > 0.$$

Оскільки $\varepsilon_k > 0$, то $x_k^* \neq 0$. З останньої нерівності отримаємо

$$\left\langle x, \frac{x_k^*}{\|x_k^*\|} \right\rangle \le \left\langle x_k, \frac{x_k^*}{\|x_k^*\|} \right\rangle.$$

Можна вважати, що

$$\frac{x_k^*}{\|x_k^*\|} \to x_0^*, \quad \|x_0^*\| = 1.$$

Переходячи до границі в останній нерівності, отримаємо $\langle x, x_0^* \rangle \leq \langle x_0, x_0^* \rangle$. Що і потрібно було показати. $\square$

Теорема 1.4.3. *Нехай M_1, M_2 – опуклі множини, що не перетинаються. Тоді існує така точка $x^* \neq 0$, що виконується нерівність*

$$\langle x_1, x^* \rangle \leq \langle x_2, x^* \rangle$$

для всіх $x_1 \in M_1$, $x_2 \in M_2$.

Доведення. Нехай $M = M_1 - M_2$. Оскільки M_1, M_2 не перетинаються, то точка $x_0 = 0 \notin M$. Застосуємо попередню теорему до M та $x_0 = 0$. Отримаємо, що існує точка $x^* \neq 0$ така, що

$$\langle x, x^* \rangle \leq \langle 0, x^* \rangle$$

для всіх $x \in M$, тобто для всіх $x = x_1 - x_2, x_1 \in M_1, x_2 \in M_2$. З останньої нерівності отримаємо

$$\langle x_1 - x_2, x^* \rangle \leq 0, \quad x_1 \in M_1, x_2 \in M_2.$$

Теорема доведена. $\square$

Теорема 1.4.4. *Нехай M_1, M_2 – замкнуті опуклі множини, що не перетинаються. Нехай одна з множин компактна. Тоді існує точка x^* і число $\varepsilon > 0$ такі, що*

$$\langle x_1, x^* \rangle \leq \langle x_2, x^* \rangle - \varepsilon$$

для всіх $x_1 \in M_1$, $x_2 \in M_2$.

Означення 1.4.1. Функція

$$s(x^*|M) = \sup_x \{\langle x, x^* \rangle : \quad x \in M\}$$

називається *опорною функцією* множини M.

Як наслідок теореми 1.4.1 маємо таке твердження.

Теорема 1.4.5. *Нехай M – замкнута опукла множина. Тоді $x \in M$ тоді і тільки тоді, коли*

$$\langle x, x^* \rangle \leq s(x^*|M)$$

для всіх $x \in M$.

1.4.1. Функція Мінковського

Ця функція представляє інтерес сама по собі, як приклад опуклої функції, що систематично буде вивчатися далі. В цьому параграфі вона буде використана при доведенні теореми про розділення, що базується на теоремі Хана – Банана.

Означення 1.4.2. Нехай M – опукла множина і $0 \in \operatorname{int} M$. Покладемо

$$\mu(x|M) = \inf\left\{\alpha\colon \alpha > 0, \frac{x}{\alpha} \in M\right\}.$$

Функція $\mu(x|M)$ визначена на всьому просторі X і називається *функцією Мінковського*. Така функція позначається ще як $r_M(x)$.

Оскільки $0 \in \operatorname{int} M$, то $\alpha^{-1}x \in M$ при великому α для будь – якого x. Тому $\mu(x|M)$ – скінченне число. Крім цього, за означенням $\mu(x|M) \geq 0$.

Лема 1.4.1. *Нехай M – опукла множина і $0 \in \operatorname{int} M$. Функція Мінковського має такі властивості:*

а) якщо $\lambda \geq 0$, то $\mu(\lambda x|M) = \lambda\mu(x|M)$;

б) якщо $x \in M$, то $\mu(x|M) \leq 1$, а якщо $x \notin M$, то $\mu(x|M) \geq 1$;

в) $\mu(x+y|M) \leq \mu(x|M) + \mu(y|M)$.

Доведення. а) Нехай $\lambda\alpha_1 = \alpha$. Маємо

$$\mu(\lambda x|M) = \inf_{\alpha}\left\{\alpha : \alpha > 0, \frac{\lambda x}{\alpha} \in M\right\} = \inf_{\alpha_1}\left\{\lambda\alpha_1 : \alpha_1 > 0, \frac{x}{\alpha_1} \in M\right\} =$$

$$= \lambda \inf_{\alpha_1}\left\{\alpha_1 : \alpha_1 > 0, \frac{x}{\alpha_1} \in M\right\} = \lambda\mu(x|M).$$

б) Якщо $x \in M$, то $x/1 \in M$, так що $\mu(x|M) \leq 1$. Нехай тепер $x \notin M$. Якщо $\mu(x|M) < 1$, то існує таке $\alpha < 1$, що $\alpha^{-1}x \in M$. Оскільки $0 \in M$ і M опукла, то

$$x = (1-\alpha)\cdot 0 + \alpha(\alpha^{-1}x) \in M,$$

що суперечить припущенню. Отже $\mu(x|M) \geq 1$ для $x \notin M$.

в) Нехай $\gamma > \mu(x|M) + \mu(y|M)$. Тоді знайдуться такі α і β, що $\gamma = \alpha + \beta$,$\alpha > \mu(x|M)$,$\beta > \mu(y|M)$. Із нерівності $\alpha > \mu(x|M)$ випливає, що $\alpha^{-1}x \in M$. Дійсно, за означенням $\mu(x|M)$ існує число α_1 таке, що $\alpha > \alpha_1 > \mu(x|M)$ і $\alpha_1^{-1}x \in M$. Тому $\alpha^{-1}x = (1-\alpha^{-1}\alpha_1)\cdot 0 + \alpha^{-1}\alpha_1(\alpha_1^{-1}x) \in M$ в силу опуклості M і того, що $0 \in M$. Аналогічно доводиться, що $\beta^{-1}y \in M$. Звідси випливає справедливість співвідношення

$$\frac{x+y}{\gamma} = \frac{x+y}{\alpha+\beta} = \frac{\alpha}{\alpha+\beta}(\alpha^{-1}x) + \frac{\beta}{\alpha+\beta}(\beta^{-1}y) \in M.$$

Тому $\mu(x+y|M) \leq \gamma$.

Але оскільки γ – довільне число, яке більше $\mu(x|M) + \mu(y|M)$, то

$$\mu(x+y|M) \leq \mu(x|M) + \mu(y|M).$$

□

1.4.2. Теорема Хана – Банаха

Ця теорема являється однією з основних у функціональному аналізі. Вона практично еквівалентна теоремі про розділення.

Теорема 1.4.6. Теорема Хана–Банаха. *Нехай на просторі X визначена функція $p(x)$ така, що*

$$p(x+y) \le p(x) + p(y), \quad p(\lambda x) = \lambda p(x), \quad \lambda \ge 0.$$

Нехай, крім цього, задана лінійна функція $f(x)$ на підпросторі $X_0 \subseteq X$ і $f(x) \le p(x)$ для всіх $x \in X_0$. Тоді існує лінійна функція $F(x)$, що визначена на всьому просторі X і така, що $f(x) = F(x)$ для $x \in X_0$ і $F(x) \le p(x)$ для $x \in X$.

Доведення. Нехай $x_0 \notin X_0$. За умовою для x_1,$x_2 \in X_0$:

$$f(x_2) - f(x_1) = f(x_2 - x_1) \le p(x_2 - x_1) =$$

$$= p((x_2 + x_0) + (-x_1 - x_0)) \le p(x_2 + x_0) + p(-x_1 - x_0),$$

звідки

$$-p(-x_1 - x_0) - f(x_1) \le p(x_2 + x_0) - f(x_2).$$

Покладемо

$$m = \sup_{x_1 \in X_0} [-p(-x_1 - x_0) - f(x_1)],$$

$$M = \inf_{x_2 \in X_0} [p(x_2 + x_0) - f(x_2)].$$

Очевидно, що числа m та M скінченні і $m \le M$. Якщо взяти r_0 між m та M, то для довільного $x \in X_0$ будемо мати

$$-p(-x - x_0) - f(x) \le r_0 \le p(x + x_0) - f(x). \tag{1.4.1}$$

Розглянемо тепер множину точок X_1 вигляду $x + \alpha x_0$, де $x \in X_0$,$\alpha \in \mathbb{R}^1$. Очевидно, що X_1 - підпростір X, оскільки X_1 замкнутий відносно операцій додавання векторів і множення на число. Далі, якщо $y = x + \alpha x_0$, то x і α визначені однозначно. Дійсно, якщо існує інше представлення $y = x_1 + \alpha_1 x_0$,$x_1 \in X_0$, то

$$x_1 - x = (\alpha - \alpha_1) x_0,$$

і якщо $\alpha - \alpha_1 \ne 0$, то $x_0 \in X_0$, що суперечить припущенню.

Для $y \in X_1$ покладемо $\varphi(y) = f(x) + \alpha r_0$. Функція $\varphi(y)$ визначена однозначно, вона лінійна і $\varphi(y) = f(y)$ при $y \in X_0$. Покажемо, що

$$\varphi(y) \le p(y).$$

Цю нерівність можна переписати у вигляді

$$f(x) + \alpha r_0 \le p(x + \alpha x_0). \tag{1.4.2}$$

Якщо $\alpha > 0$, то нерівність (1.4.2) перетвориться в наступну:

$$r_0 \le p\left(\frac{x}{\alpha} + x_0\right) - f\left(\frac{x}{\alpha}\right),$$

еквівалентну правій нерівності з (1.4.1), якщо замінити в ній x на $\alpha^{-1}x$. Аналогічно, якщо $\alpha < 0$, то нерівність (1.4.2) перетвориться в ліву нерівність із (1.4.1).

Отже показано, що функція $f(x)$ може бути продовжена зі збереженням всіх своїх властивостей на підпростір X_1, розмірність якого на одиницю більша розмірності X_0. Повторюючи дану процедуру потрібну кількість разів, ми побудуємо функцію $f(x)$ в усьому просторі X. □

Покажемо, як з цієї теореми випливає теорема про розділення.

Теорема 1.4.7. *Нехай M – опукла множина,* $\operatorname{int} M \neq \emptyset$ *та* $x_0 \notin M$. *Тоді існує така лінійна функція $F(x)$, що*

$$F(x) \leq F(x_0), \quad F(x_0) \geq 1, \quad \forall x \in M.$$

Доведення. Оскільки $\operatorname{int} M \neq \emptyset$, то можна вважати, що $0 \in \operatorname{int} M$. Інакше можна взяти $\overline{x} \in \operatorname{int} M$ та розглядати $M - \overline{x}$ і точку $x_0 - \overline{x}$.

Оскільки $0 \in \operatorname{int} M$, то визначена функція Мінковського $\mu(x|M)$. За лемою 1.4.1 якщо $x \in M$, то $\mu(x|M) \leq 1$, а якщо $x \notin M$, то $\mu(x|M) \geq 1$. Визначимо тепер на одномірному підпросторі X_0, що складається з елементів $x = \alpha x_0, \alpha \in \mathbb{R}^1$, лінійну функцію $f(x)$, таким чином:

$$f(x) = \alpha\mu(x_0|M).$$

За лемою 1.4.1 функція $\mu(x|M)$ та лінійна функція $f(x)$ на підпросторі X_0 задовольняють всі умови теореми 1.4.6, оскільки легко перевірити нерівність $f(x) \leq \mu(x|M)$, $x \in X_0$.
Застосовуючи терему 1.4.6, отримаємо функцію $F(x)$ таку, що

$$F(x) \leq \mu(x|M), \quad x \in X,$$

$$F(x_0) = f(x_0) = \mu(x_0|M) \geq 1.$$

Але оскільки $\mu(x|M) \leq 1$ при $x \in M$, то

$$F(x) \leq \mu(x|M) \leq 1 \leq \mu(x_0|M) = F(x_0),$$

що й завершує доведення. □

Теорема 1.4.7 майже еквівалентна теоремі 1.4.2, оскільки в скінченновимірному просторі $X = \mathbb{R}^n$ будь-яка лінійна функція задається скалярним добутком. Залишається умова $\operatorname{int} M \neq \emptyset$, не враховувати яку в загальному випадку не можна. Але в скінченновимірному просторі це можна зробити, розглядаючи M в $\operatorname{Lin} M$, де внутрішність множини M непорожня.

1.5. ОПУКЛІ КОНУСИ

Означення 1.5.1. Опукла множина $K \subset \mathbb{R}^n$ називається *опуклим конусом*, якщо з того, що $x \in K$ і $\lambda > 0$, випливає, що $\lambda x \in K$.

Лема 1.5.1. *Опуклий конус містить будь-які лінійні комбінації своїх елементів із додатніми коефіцієнтами. Якщо $x_1,\ldots,x_m \in K$, $\lambda_1 \geq 0,\ldots,\lambda_m \geq 0$, то*

$$\lambda_1 x_1 + \cdots + \lambda_m x_m \in K.$$

Означення 1.5.2. Множина векторів $x^* \in X^*$ таких, що $\langle x,x^* \rangle \geq 0\ \forall x \in K$, називається *спряженим конусом* до конуса K і позначається K^*. Це можна записати так:

$$K^* = \{x^* \in X^* \colon \langle x,x^* \rangle \geq 0, x \in K\}.$$

Наведемо ряд властивостей спряжених конусів.
Очевидно, що K^* – опуклий конус.

Лема 1.5.2. *K^* – замкнутий конус.*

Лема 1.5.3. *Конуси K та $\overline{K}$ мають однакові спряжені конуси, тобто $K^* = (\overline{K})^*$.*

Лема 1.5.4. *Якщо K – замкнутий конус і $\langle x,x^* \rangle \geq 0\ \forall x^* \in K^*$, то $x \in K$.*

Оскільки K^* – опуклий конус, то можна поставити питання про обчислення спряженого до нього конуса $(K^*)^*$, тобто K^{**}.

Лема 1.5.5. *Якщо K – замкнутий конус, то $K^{**} = K$.*

В загальному випадку $K^{**} = \overline{K}$. Це випливає з леми 1.5.3, оскільки $K^* = (\overline{K})^*$ і тому $K^{**} = (\overline{K})^{**} = \overline{K}$.

Лема 1.5.6. *Якщо K_1,K_2 – опуклі конуси, то $K_1 + K_2$ також опуклий конус і*

$$(K_1 + K_2)^* = K_1^* \cap K_2^*.$$

Лема 1.5.7. *Для замкнутих конусів K_1 та K_2 справджується рівність*

$$(K_1 \cap K_2)^* = \overline{K_1^* + K_2^*}.$$

Доведення. Доведення леми формально спирається на попередні результати:

$$(K_1 \cap K_2)^* = (K_1^{**} \cap K_2^{**})^* = ((K_1^* + K_2^*)^*)^* = (K_1^* + K_2^*)^{**} = \overline{K_1^* + K_2^*}.$$

□

Лема 1.5.8. *Якщо добуток $\langle x,x^* \rangle$ обмежений знизу для всіх $x \in K$, то $x^* \in K^*$. Якщо $x \in \operatorname{int} K$, то $\langle x,x^* \rangle > 0$, для всіх $x^* \in K^*$, $x^* \neq 0$.*

Доведення. Доведемо друге твердження. Якщо $x \in \operatorname{int} K$, то існує таке $\varepsilon > 0$, що $x + \varepsilon B \subseteq K$, де B- одинична куля з центром в нулі. Тому $\langle x + \varepsilon z, x^* \rangle \geq 0$ для $x^* \in K^*$ при всіх $z \in B$. Звідси випливає, що

$$\langle x, x^* \rangle \geq \varepsilon \sup_{z \in B} \langle -z, x^* \rangle \geq \varepsilon \left\langle \frac{x^*}{\|x^*\|}, x^* \right\rangle = \varepsilon \|x^*\| > 0.$$

Що й потрібно було довести. □

1.5.1. Розділення опуклих конусів

Теорема 1.5.1. *Нехай* $K_1,\dots,K_m$– *опуклі конуси. Для того, щоб їх перетин був порожнім, необхідно, щоб знайшлись такі* $x_i^* \in K_i^*, i = 1,\dots,m$, *не всі рівні нулю, що*

$$x_1^* + \dots + x_m^* = 0.$$

Доведення. Розглянемо простір X^m, тобто простір $\mathbb{R}^{mn}$, елементи якого мають вигляд $(x_1,\dots,x_m)$, де кожна компонента x_i належить X. Скалярний добуток векторів $(x_1,\dots,x_m)$ і $(x_1^*,\dots,x_m^*)$ в такому просторі можна записати у вигляді суми:

$$\langle x_1, x_1^* \rangle + \dots + \langle x_m, x_m^* \rangle.$$

Розглянемо в X^m два конуси:

$$\widetilde{K} = K_1 \times \dots \times K_m = \{(x_1,\dots,x_m) : x_1 \in K_1,\dots,x_m \in K_m\},$$

$$\widetilde{P} = \{(x,\dots,x) : x \in X\}.$$

Другий конус складається з векторів, всі n – вимірні компоненти яких рівні між собою. Оскільки перетин конусів $K_1,\dots,K_m$ порожній, то $\widetilde{K}$ і $\widetilde{P}$ не перетинаються. Застосовуючи теорему 1.4.3, отримаємо, що існує такий вектор $(x_1^*,\dots,x_m^*)$, що

$$\langle x, x_1^* \rangle + \dots + \langle x, x_m^* \rangle \le \langle x_1, x_1^* \rangle + \dots + \langle x_m, x_m^* \rangle, x \in X, x_1 \in K_1,\dots,x_m \in K_m. \tag{1.5.1}$$

Із нерівності (1.5.1) випливає, що добуток $\langle x_i, x_i^* \rangle$ обмежений на конусі $K_i, i = 1,\dots,m$. Тоді за лемою 1.5.8

$$x_i^* \in K_i^*, \quad i = 1,\dots,m. \tag{1.5.2}$$

Ліва частина нерівності (1.5.1) в свою чергу показує, що добуток

$$\langle x, x_1^* + \dots + x_m^* \rangle$$

обмежений зверху для всіх $x \in X$. Але це може бути лише в тому випадку, коли

$$x_1^* + \dots + x_m^* = 0, \tag{1.5.3}$$

оскільки ненульова лінійна функція може приймати як завгодно великі значення. Із (1.5.2) і (1.5.3) випливає твердження теореми. □

Теорема 1.5.2. *Нехай*

$$K = K_1 \cap \dots \cap K_m,$$

$$K_1 \cap \operatorname{int} K_2 \cap \dots \cap \operatorname{int} K_m \neq \emptyset.$$

Тоді

$$K^* = K_1^* + \dots + K_m^*.$$

Доведення. Справедливість включення
$$K^* \supseteq K_1^* + \dots + K_m^*,$$
перевіряється безпосередньо, виходячи з означення спряженого конуса.

Доведемо обернене включення. Нехай $x^* \in K^*, x^* \neq 0$. Покладемо
$$K_0 = \{x : \langle x, x^* \rangle < 0\}.$$
Конуси K_0 і K не перетинаються, оскільки в протилежному випадку із існування елемента $x_1 \in K \cap K_0 \neq \emptyset$ випливає, що $\langle x_1, x^* \rangle < 0$, оскільки $x_1 \in K_0$, і $\langle x_1, x^* \rangle \geq 0$, оскільки $x_1 \in K, x^* \in K^*$. Розглянемо конус, спряжений до K_0. Нехай із $\langle x, x^* \rangle < 0$ випливає, що $\langle x, y^* \rangle \geq 0$, тобто $y^* \in K_0^*, y^* \neq 0$. Тоді x^* і y^* лінійно залежні. Отже, при деяких не рівних одночасно нулю α_1 і α_2 виконується $\alpha_1 x^* - \alpha_2 y^* = 0$. Оскільки $x^* \neq 0$ і $y^* \neq 0$, то $\alpha_2 \neq 0$ і
$$y^* = \lambda x^*, \quad \lambda = \alpha_1 / \alpha_2.$$
Далі, для $x \in K_0$
$$0 \leq \langle x, y^* \rangle = \lambda \langle x, x^* \rangle.$$
Оскільки $\langle x, x^* \rangle < 0$, то $\lambda < 0$. У тому випадку, коли $y^* = 0$, покладаємо $y^* = 0x^*$.

Отже, доведено, що
$$K_0^* = \{y^* : y^* = \lambda x^*, \lambda \leq 0\}.$$
Оскільки конуси K_0 і K не перетинаються, то за попередньою теоремою існують такі $y^* \in K_0^*, x_i^* \in K_i^*, i = 1, \dots, m$, що
$$y^* + x_1^* + \dots + x_m^* = 0, \tag{1.5.4}$$
причому не всі доданки в цій сумі рівні нулю. Перепишемо рівність (1.5.4) у вигляді
$$-\lambda x^* = x_1^* + \dots + x_m^*, \tag{1.5.5}$$
використовуючи те, що $y^* = \lambda x^*, \lambda \leq 0$. Якщо $\lambda < 0$, то
$$x^* = \left(-\frac{1}{\lambda}\right) x_1^* + \dots + \left(-\frac{1}{\lambda}\right) x_m^* \in K_1^* + \dots + K_m^*,$$
оскільки x_i^* належать конусам $K_i^*, i = 1, \dots, m$. Покажемо, що λ не може бути рівним нулю. Припустимо протилежне. Нехай $\lambda = 0$. Тоді співвідношення (1.5.5) переходить в рівність
$$x_1^* + \dots + x_m^* = 0. \tag{1.5.6}$$
Тут не всі x_i^* рівні нулю. Отже, не рівні нулю принаймні два елемента, наприклад x_1^* і x_2^*. За припущенням теореми існує точка $x_0 \in K_1 \cap \mathrm{int}\ K_2 \cap \dots \cap \mathrm{int}\ K_m$ така, що за лемою 1.5.8
$$\langle x_0, x_2^* \rangle > 0, \quad \langle x_0, x_i^* \rangle \geq 0, i \neq 2.$$
Домножуючи тепер обидві частини (1.5.6) на x_0 скалярно, отримаємо суперечність:
$$0 = \langle x_0, x_1^* + \dots + x_m^* \rangle = \langle x_0, x_1^* \rangle + \dots + \langle x_0, x_m^* \rangle > 0.$$

$\square$

Теорема 1.5.3. *Нехай $K_1, \ldots, K_m$ - опуклі конуси, $K = K_1 \cap \cdots \cap K_m$. Тоді або*

$$K^* = K_1^* + \cdots + K_m^*, \tag{1.5.7}$$

або існують не всі рівні нулю вектори $x_i^* \in K_i^*$ такі, що

$$x_1^* + \cdots + x_m^* = 0.$$

Доведення. Знову звернемося до співвідношення (1.5.5). Якщо для всіх $x^* \in K^*$ в (1.5.5) виявиться, що $\lambda < 0$, то x^* можна представити у вигляді суми елементів $x_i^* \in K_i^*$. Це доводить рівність (1.5.7). Якщо ж для деякого x^* виявляється, що $\lambda = 0$, то виконується рівність (1.5.6).

$\square$

З теореми 1.5.3 випливає, що якщо перетин конусів порожній, то виконується (1.5.6). Поставимо питання: при яких найбільш загальних припущеннях може бути виконана ця рівність?

Означення 1.5.3. Конуси $K_1, \ldots, K_m$ *розділяються*, якщо існують не всі рівні нулю $x_i^* \in K_i^*$ такі, що $x_1^* + \cdots + x_m^* = 0$.

Теорема 1.5.4. *Для того, щоб конуси K_1 і K_2 не розділялись, необхідно і достатньо виконання умов:*

а) $\mathrm{ri}\ K_1 \cap \mathrm{ri}\ K_2 \neq \emptyset$;

б) $\mathrm{Lin}\ K_1 + \mathrm{Lin}\ K_2 = \mathbb{R}^n$.

Доведення. Якщо конуси розділяються, то знайдуться $x_1^* \in K_1^*, x_2^* \in K_2^*, x_1^* \neq 0$ такі, що $x_1^* + x_2^* = 0$. Тоді

$$\langle x_1 - x_2, x_1^* \rangle = \langle x_1, x_1^* \rangle + \langle x_2, x_2^* \rangle \geq 0, \quad x_1 \in K_1, x_2 \in K_2. \tag{1.5.8}$$

Із (1.5.8) випливає, що якщо конуси K_1 і K_2 розділяються, то нуль не може бути внутрішньою точкою $K_1 - K_2$. Дійсно, якщо $0 \in \mathrm{int}\,(K_1 - K_2)$, то $-\varepsilon x_1^* \in K_1 - K_2$ при досить малому $\varepsilon > 0$, тобто $-\varepsilon x_1^* = \overline{x_1} - \overline{x_2}$, $\overline{x_1} \in K_1, \overline{x_2} \in K_2$. Тоді

$$\langle \overline{x_1} - \overline{x_2}, x_1^* \rangle = -\varepsilon \left\| x_1^* \right\|^2 < 0,$$

що суперечить (1.5.8).

І навпаки, якщо $0 \notin \mathrm{int}\,(K_1 - K_2)$, то існує точка $x^* \neq 0$ така, що

$$\langle x_1 - x_2, x^* \rangle \geq 0, \quad x_1 - x_2 \in \mathrm{int}\,(K_1 - K_2).$$

Покладаючи $x_1^* = x^*, x_2^* = -x^*$, отримуємо, що $x_1^* \in K_1^*, x_2^* \in K_2^*$ і $x_1^* + x_2^* = 0$, тобто конуси K_1 і K_2 розділяються.

Отже, необхідною і достатньою умовою нероздільності двох конусів є умова $0 \in \mathrm{int}\,(K_1 - K_2)$.

Якщо ri $K_1 \cap$ri $K_2 = \emptyset$, то K_1 і K_2 розділяються. Для цього достатньо розділити ri K_1 і ri K_2:

$$\langle x_1,x^* \rangle \geq \langle x_2,x^* \rangle\,, \quad x_1 \in \text{ri } K_1, x_2 \in \text{ri } K_2, \tag{1.5.9}$$

і позначити $x_1^* = x^*, x_2^* = -x^*$. Залишається відзначити, що $\overline{K_i}^* = \overline{\text{ri } K_i}$ і $K_i^* = \left(\overline{K_i}\right)^* = (\text{ri } K_i)^*$.

Припустимо тепер, що ri $K_1 \cap$ ri $K_2 \neq \emptyset$, але Lin $K_1 +$ Lin $K_2 \neq \mathbb{R}^n$. Так як $0 \in K$ і Lin $\overline{K} =$ Lin K, то $K \subseteq$ Lin K для конуса. Далі, $K_1 - K_2 \subseteq$ Lin $K_1 -$ Lin K_2. Так як Lin K_2 – підпростір, який разом із елементом x містить і елемент $-x$, то $-$Lin $K_2 =$ Lin K_2. Тому маємо

$$K_1 - K_2 \subseteq \text{Lin } K_1 + \text{Lin } K_2.$$

Але Lin $K_1 +$ Lin $K_2 \neq \mathbb{R}^n$ і тому є власним підпростором $\mathbb{R}^n$. Значить, існує вектор $x^* \neq 0$ ортогональний всім векторам із Lin $K_1 +$ Lin K_2, зокрема,

$$\langle x_1 - x_2,x^* \rangle = 0, \quad x_1 \in K_1,x_2 \in K_2.$$

Остання рівність фактично показує, що K_1 і K_2 розділяються.

Залишилось розглянути випадок, коли ri $K_1 \cap$ ri $K_2 \neq \emptyset$ і Lin $K_1 +$ Lin $K_2 = \mathbb{R}^n$.

Нехай $e_1, \ldots ,e_n$ - базис в $\mathbb{R}^n$. Так як $e_k \in$ Lin $K_1 +$Lin K_2, то існують такі $g_k \in$ Lin K_1 і $f_k \in$ Lin K_2, що $e_k = g_k - f_k$.

Нехай $x_0 \in$ ri $K_1 \cap$ ri K_2. Позначимо через $g = g_1 + \cdots + g_n$, $f = f_1 + \cdots + f_n$. Очевидно, що $g \in$ Lin $K_1, f \in$ Lin K_2. Виберемо тепер $\varepsilon > 0$ настільки малим, щоб виконувалось включення:

$$x_0 + \varepsilon \left(g_k - \frac{1}{n+1} g \right) \in K_1, k = 1, \ldots ,n;\; x_0 - \frac{\varepsilon}{n+1} g \in K_1,$$

$$x_0 + \varepsilon \left(f_k - \frac{1}{n+1} f \right) \in K_2, k = 1, \ldots ,n;\; x_0 - \frac{\varepsilon}{n+1} f \in K_2.$$

Тоді точки

$$y_k = \left(x_0 + \varepsilon \left(g_k - \frac{1}{n+1} g \right) \right) - \left(x_0 + \varepsilon \left(f_k - \frac{1}{n+1} f \right) \right) =$$

$$= \varepsilon \left(e_k - \frac{1}{n+1} e \right) \in K_1 - K_2, k = 1, \ldots ,n.$$

$$y_0 = \left(x_0 - \frac{\varepsilon}{n+1} g \right) - \left(x_0 - \frac{\varepsilon}{n+1} f \right) = -\frac{\varepsilon}{n+1} e \in K_1 - K_2,$$

де $e = g - f = e_1 + \ldots + e_n$, вершини симплексу в $\mathbb{R}^n$, для якого точка нуль – внутрішня. Цей симплекс лежить в $K_1 - K_2$, оскільки його вершини лежать в $K_1 - K_2$. Отже $0 \in$ int $(K_1 - K_2)$ і K_1 та K_2 не розділяються. $\square$

Теорема 1.5.5. *Для того, щоб конуси $K_1,\ldots,K_m$ не розділялись, необхідно і достатньо, щоб виконувались умови:*

а) $\mathrm{ri}\ K_1 \cap \mathrm{ri}\ K_2 \cap \cdots \cap \mathrm{ri}\ K_m \neq \emptyset$;

б) $\mathrm{Lin}\ K_1 \cap \cdots \cap \mathrm{Lin}\ K_{j-1} + \mathrm{Lin}\ K_j = \mathbb{R}^n$ *для всіх* $j = 2,\ldots,m$.

1.5.2. Приклад опуклого конуса

Нехай M – непорожня множина. Позначимо
$$\mathrm{cone}\ M = \{x \colon x = \lambda x_1, x_1 \in M, \lambda > 0\}.$$

Теорема 1.5.6. *Нехай M – опукла множина. Тоді* $\mathrm{cone}\ M$ *– опуклий конус, причому* $x \in \mathrm{int}(\mathrm{cone}\ M)$ *при* $x = \lambda x_1, x_1 \in \mathrm{int}\ M, \lambda > 0$.

Доведення. Візьмемо $x_1,x_2 \in \mathrm{cone}\ M$. Тоді
$$x_1 = \lambda_1 \overline{x}_1, x_2 = \lambda_2 \overline{x}_2, \overline{x}_1, \overline{x}_2 \in M, \lambda_1,\lambda_2 > 0.$$
Нехай $\gamma_1,\gamma_2 > 0$ довільні. Маємо
$$\gamma_1 x_1 + \gamma_2 x_2 = \gamma_1 \lambda_1 \overline{x}_1 + \gamma_2 \lambda_2 \overline{x}_2 =$$
$$= (\gamma_1\lambda_1 + \gamma_2\lambda_2)\left[\frac{\gamma_1\lambda_1}{\gamma_1\lambda_1 + \gamma_2\lambda_2}\overline{x}_1 + \frac{\gamma_2\lambda_2}{\gamma_1\lambda_1 + \gamma_2\lambda_2}\overline{x}_2\right] \in \mathrm{cone}\ M,$$
оскільки в силу опуклості M вираз в квадратних дужках належить M.

Далі, якщо $x = \lambda x_1, x_1 \in \mathrm{int}\ M, \lambda > 0$, то $x_1 + \varepsilon B \subseteq M$ при достатньо малому $\varepsilon > 0$. Тому
$$x + \lambda\varepsilon B = \lambda(x_1 + \varepsilon B) \subseteq \mathrm{cone}\ M,$$
тобто $x \in \mathrm{int}\,\mathrm{cone}\ M$. □

Теорема 1.5.7. *Якщо $M_1,\ldots,M_k$ - опуклі множини і $0 \in M_i$, то*
$$\bigcap_{i=1}^{k} \mathrm{cone}\ M_i = \mathrm{cone}\,(\bigcap_{i=1}^{k} M_i).$$

Доведення. Якщо $x = \lambda x_1$, де $x_1 \in \bigcap\limits_{i=1}^{k} M_i$, то ясно, що $x \in \mathrm{cone}\ M_i$ для всіх $i = 1,\ldots,k$. Навпаки, нехай $x \in \mathrm{cone}\ M_i$, тобто $x = \lambda_i x_i, \lambda_i > 0, x_i \in M_i$, $i = 1,\ldots,k$. Тоді $\lambda_i^{-1} x \in M_i$ і
$$\lambda(\lambda_i^{-1}x) = (1-\lambda)0 + \lambda(\lambda_i^{-1}x) \in M_i$$
при $0 \leq \lambda < 1$. Вибираючи μ із умов $0 < \mu \leq \min \lambda_i^{-1}$, отримаємо
$$\mu x = (\mu\lambda_i)(\lambda_i^{-1}x) \in M_i, \quad i = 1,\ldots,k,$$
тобто
$$\mu x \in \bigcap_{i=1}^{k} M_i, x = \left(\frac{1}{\mu}\right)(\mu x) \in \mathrm{cone}\left(\bigcap_{i=1}^{k} M_i\right),$$

що і потрібно було довести. $\square$

Теорема 1.5.8. *Вектор* $x^* \in (\text{cone } M)^*$ *тоді і тільки тоді, коли виконується нерівність* $\langle x,x^* \rangle \geq 0$ *для всіх* $x \in M$.

Доведення. Дійсно, $x^* \in (\text{cone } M)^*$, тоді і тільки тоді, коли $\langle \lambda x,x^* \rangle \geq 0$ при всіх $\lambda \geq 0$ і всіх $x \in M$. Оскільки $\lambda \geq 0$, то з цієї нерівності і випливає справедливість твердження. $\square$

Теорема 1.5.9. *Нехай* $0 \in M$. *Тоді* $x \in \text{cone } M$ *тоді і тільки тоді, коли* $\lambda x \in M$ *при достатньо малих* $\lambda > 0$.

Доведення. Якщо $x \in \text{cone } M$, то $x = \lambda_1 x_1$, де $\lambda_1 > 0, x_1 \in M$. Тому

$$\lambda x = \lambda \lambda_1 x_1 = (1 - \lambda\lambda_1)0 + \lambda\lambda_1 x_1 \in M,$$

при $\lambda\lambda_1 \leq 1$. Навпаки, якщо $\lambda x \in M, \lambda > 0$, то $x = \frac{1}{\lambda} x_1, x_1 \in M$ і $x \in \text{cone } M$. $\square$

1.6. КРАЙНІ ТОЧКИ ОПУКЛОЇ МНОЖИНИ

Означення 1.6.1. Точка x опуклої множини X називається *крайньою (екстремальною)* точкою, якщо її не можна подати у вигляді опуклої комбінації

$$x = \lambda x^1 + (1 - \lambda)x^2, x^1,x^2 \in X,\ x^1 \neq x^2,\ 0 < \lambda < 1. \tag{1.6.1}$$

Сукупність всіх крайніх точок множини X позначимо через $E(X)$.

Таким чином, точка x є крайньою в X, якщо її не можна помістити на відрізок, кінці якого лежать в X. Наприклад, у трикутника крайніми точками є його вершини, у променя – початок, у круга – всі точки кола. Наведемо лему, яка є корисним інструментом для доведення наступних основних теорем теорії крайніх точок.

Лема 1.6.1. *Нехай* X *– замкнута опукла множина в* $\mathbb{R}^n$, $H = H_{p\beta}$ *– власна опорна до* X *в точці* $\hat{x} \in r\partial\, X = X \setminus ri\, X$ *гіперплощина, тобто виконані умови*

$$\langle p,x \rangle \geq \beta = \langle p,\hat{x} \rangle \quad \textit{при всіх } x \in X, \tag{1.6.2}$$

$$\langle p,\bar{x} \rangle > \beta \quad \textit{при деякому } \bar{x} \in X. \tag{1.6.3}$$

Покладемо $\hat{X} = X \cap H$. *Тоді:*
1) будь-яка крайня точка в $\hat{X}$ *є крайньою і в* X, *тобто* $E(\hat{X}) \subset E(X)$;
2) $dim\, \hat{X} < dim\, X$.

Доведення. 1). Нехай $x \in E(\hat{X})$, але $x \notin E(X)$, тобто x можна подати у вигляді (1.6.1). Користуючись (1.6.2), отримаємо

$$\beta = \langle p,x \rangle = \lambda \langle p,x^1 \rangle + (1-\lambda)\langle p,x^2 \rangle \geq \lambda\beta + (1-\lambda)\beta = \beta.$$

Звідки $\langle p,x^1 \rangle = \langle p,x^2 \rangle = \beta$. Тобто $x^1,x^2 \in \hat{X} = X \cap H$. Разом з (1.6.1) це означає, що $x \notin E(\hat{X})$. Ця суперечність доводить, що $x \in E(X)$. Тобто $E(\hat{X}) \subset E(X)$.

2). Покладемо $M = \operatorname{aff} X$, $\hat{M} = \operatorname{aff} \hat{X}$. Тоді $\hat{M} \subset M$, $\hat{M} \subset H$, оскільки $\hat{X} \subset X$, $\hat{X} \subset H$. Припустимо, що $\hat{M} = M$. Тоді $X \subset M = \hat{M} \subset H$, тобто $\beta = \langle p,x \rangle \ \forall x \in X$, що суперечить (1.6.3). Отже $\hat{M} \neq M$. Паралельні підпростори $L = \operatorname{Lin} X$ та $\hat{L} = \operatorname{Lin} \hat{X}$ пов'язані такими самими співвідношеннями $\hat{L} \subset L$, $\hat{L} \neq L$. Тому базис L має принаймі на один вектор більше, ніж в $\hat{L}$, тобто $\dim \hat{L} < \dim L$. Але за означенням $\dim X = \dim L$, $\dim \hat{X} = \dim \hat{L}$. □

Теорема 1.6.1. Критерій існування крайньої точки. *Нехай X – замкнута опукла множина в $\mathbb{R}^n$. Тоді X має принаймі одну крайню точку тоді і тільки тоді, коли в X не включаються прямі, тобто множини виду*

$$l_{x^0h} = \{x \in \mathbb{R}^n \colon x = x^0 + \alpha h, \alpha \in \mathbb{R}\},$$

де $x^0 \in \mathbb{R}^n, h \in \mathbb{R}^n, h \neq 0$.

Доведення. 1). Нехай $x \in E(X)$, але $l_{x^0h} \subset X$ при деяких x^0 та $h \neq 0$. Тоді за теоремою 1.3.5 маємо $l_{xh} \subset X$, звідки $x_1 = x + h \in X$ та $x_2 = x - h \in X$. При цьому $x = 0{,}5x_1 + 0{,}5x_2, x_1 \neq x_2$, тобто $x \notin E(X)$. Це доводить твердження теореми в один бік.

2). Припустимо тепер, що X не містить прямих. Покажемо, що $E(X) \neq \emptyset$ методом математичної індукції за розмірністю X. Якщо $\dim X = 0$, то $X = \{\hat{x}\}$ – одноточкова множина і $E(X) = \{\hat{x}\} \neq \emptyset$. Нехай твердження справедливе для $\dim X < m$ і $\dim X = m$. Виберемо яку-небудь точку $\hat{x} \in r\partial X$. Нехай $H = H_{p\beta}$ – власна опорна до X в точці $\hat{x}$ гіперплощина. Візьмемо $\hat{X} = X \cap H$. Ця множина замкнута, опукла і не містить прямих. При цьому $\dim \hat{X} < m$. За припущенням $E(\hat{X}) \neq \emptyset$. Але $E(\hat{X}) \subset E(X)$. Тому $E(X) \neq \emptyset$. □

Теорема 1.6.2. Теорема Мінковського про опуклий компакт. *Нехай X – опуклий компакт (опукла замкнута обмежена множина) в $\mathbb{R}^n$. Тоді $X = \operatorname{conv} E(X)$, тобто X співпадає з опуклою оболонкою множини своїх крайніх точок.*

Доведення. Твердження доведемо методом математичної індукції за розмірністю X. Якщо $\dim X = 0$, то теорема очевидна. Нехай твердження справедливе у тому випадку, коли $\dim X < m$ та нехай $\dim X = m$. Нехай $H = H_{p\beta}$ – власна опорна до X у деякій точці $\hat{x} \in r\partial X$ гіперплощина. Візьмемо $\hat{X} = X \cap H$. У даному випадку ця множина - опуклий компакт. При цьому $\dim \hat{X} < m$. Тоді за припущенням індукції $\hat{X} = \operatorname{conv} E(\hat{X})$. Маємо $\hat{x} \in \operatorname{conv} E(\hat{X})$. Але $E(\hat{X}) \subset E(X)$. Тому $\hat{x} \in \operatorname{conv} E(X)$. Отже $r\partial X \subset \operatorname{conv} E(X)$.

Розглянемо тепер довільну точку $\hat{x} \in \operatorname{ri} X$ і вектор $h \in \operatorname{Lin} X$. Тоді пряма $l_{\hat{x}h}$ лежить в $\operatorname{aff} X$. Перетин цієї прямої з X утворює відрізок з кінцями на відносній границі X. Тобто $l_{\hat{x}h} \cap X = \operatorname{conv}\{x^1,x^2\}$, $x^1,x^2 \in r\partial X$. Отже $\hat{x} \in \operatorname{conv}\{x^1,x^2\} \subset \operatorname{conv}(r\partial X) \subset \operatorname{conv}(\operatorname{conv} E(X)) = \operatorname{conv} E(X)$. Таким чином $\operatorname{ri} X \subset \operatorname{conv} E(X)$, $X = r\partial X \cup \operatorname{ri} X \subset \operatorname{conv} E(X)$. Обернене включення $\operatorname{conv} E(X) \subset X$ очевидне, оскільки $E(X) \subset X$ і X - опукла множина. $\square$

Означення 1.6.2. *Поліедром* називається множина розв'язків системи скінченної кількості лінійних нерівностей, тобто перетин скінченної кількості півпросторів:

$$X = \left\{x \in \mathbb{R}^n \colon \langle a^i,x\rangle \le b_i, i \in I = \{1,\ldots,m\}\right\}, \tag{1.6.4}$$

або

$$X = \{x \in \mathbb{R}^n \colon Ax \le b\},$$

де $b = (b_1,\ldots,b_m) \in \mathbb{R}^m$, A - матриця розмірності $m \times n$ з рядками $a^1,a^2,\ldots,a^m \in \mathbb{R}^n$.

Теорема 1.6.3. *Для того щоб точка $\hat{x}$ була крайньою точкою поліедра X, що визначений системою (1.6.4) лінійних нерівностей, необхідно і достатньо, щоб множина*

$$I(\hat{x}) = \left\{i \colon \langle a^i,\hat{x}\rangle = b_i, i \in I\right\}$$

містила підмножину I_0 розмірності n таку, що вектори $a^i, i \in I_0$, лінійно незалежні.

Доведення. Необхідність. Нехай множина $\{a_i \colon \ i \in I(\hat{x})\}$ містить менше ніж n лінійно незалежних елементів. Тоді на підставі відомих теорем лінійної алгебри система лінійних за x рівнянь

$$\langle a^i,x\rangle = 0, i \in I(\hat{x}),$$

має ненульовий розв'язок $\overline{x}$. Із означення множини $I(\hat{x})$ випливає, що

$$\langle \hat{x} \pm \varepsilon\overline{x}, a^i\rangle = b_i, i \in I(\hat{x}),$$

$$\langle \hat{x} \pm \varepsilon\overline{x}, a^i\rangle < b_i, i \in I \backslash I(\hat{x}),$$

при достатньо малому $\varepsilon > 0$, такому що

$$\hat{x} + \varepsilon\overline{x} \in X, \quad \hat{x} - \varepsilon\overline{x} \in X,$$

$$\hat{x} = \frac{1}{2}(\hat{x} + \varepsilon\overline{x}) + \frac{1}{2}(\hat{x} - \varepsilon\overline{x}) \in X,$$

тобто $\hat{x}$ не є крайньою точкою X.

Достатність. Нехай точка $\hat{x} \in X$, розмірність I_0 дорівнює n і для $i \in I_0$ вектори a^i лінійно незалежні. Тоді система нерівностей, що описують множину X, може бути записана в такому виді:

$$\langle \hat{x}, a^i \rangle = b_i, \quad i \in I_0, \tag{1.6.5}$$

$$\langle \hat{x}, a^i \rangle \le b_i, \quad i \in I \backslash I_0. \tag{1.6.6}$$

Припустимо, що

$$\hat{x} = 0{,}5x^1 + 0{,}5x^2, \quad x^1 \in X, \quad x^2 \in X, \quad x_1 \ne x_2. \tag{1.6.7}$$

Оскільки $x^1, x^2 \in X$, то за означенням справедливі нерівності

$$\langle x^k, a^i \rangle \le b_i, \quad i \in I_0, \; k = 1{,}2. \tag{1.6.8}$$

В силу умов (1.6.5), (1.6.8) співвідношення (1.6.7) виконується лише в тому випадку, коли

$$\langle x^k, a^i \rangle = b_i, \quad i \in I_0, \quad k = 1{,}2. \tag{1.6.9}$$

Отже, дві різні точки задовольняють системі n лінійно незалежних рівнянь (1.6.9). Це неможливо в силу відомих теорем. Теорему доведено. □

З останніх двох теорем випливає така теорема.

Теорема 1.6.4. *Обмежений поліедр, що заданий скінченною системою лінійних нерівностей (1.6.4), є опуклою оболонкою своїх крайніх точок, кількість яких скінченне.*

1.7. ОПУКЛІ МНОГОГРАННИКИ

Означення 1.7.1. Опукла оболонка множини, що складається із скінченної кількості точок, називається *опуклим многогранником* породженим цими точками.

Теорема 1.7.1. *Обмежена множина, що задана скінченною системою лінійних нерівностей (1.6.4), є опуклим многогранником.*

Опуклий многогранник $X = \operatorname{conv}\left\{x^1, \ldots, x^m\right\}$ має вигляд

$$X = \left\{ x \in \mathbb{R}^n \,\middle|\, x = \sum_{i=1}^{m} \lambda_i x^i, \lambda_i \ge 0; \; \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Опорна функція многогранника має вигляд

$$s(x^*|M) = \sup_{x \in M} \langle x, x^* \rangle = \max_{1 \le i \le m} \langle x^i, x^* \rangle.$$

Нехай

$$I(x^*) = \{i : \langle x^i, x^* \rangle = s(x^*|M), i = 1, \ldots, m\},$$

тобто $I(x^*)$ – це множина індексів i, для яких скалярний добуток $\langle x^i, x^* \rangle$ досягає максимуму.

Означення 1.7.2. Гіперплощина

$$\{x\colon \langle x, x^* \rangle = s(x^*|M)\}$$

називається *опорою* гіперплощиною. Якщо $j \in I(x^*)$ і серед векторів $\left\langle x^i - x^j, x^* \right\rangle = 0$, $i \in I(x^*)$ є n лінійно незалежних, то опора називається *крайньою*.

Теорема 1.7.2. *Нехай многогранник M містить внутрішні точки Тоді будь-яку точку $x_0 \notin M$ можна розділити за допомогою крайньої опори, тобто знайдеться вектор x^*, який визначає крайню опору такий, що*

$$\langle x, x^* \rangle < \left\langle x^0, x^* \right\rangle, \quad x \in M.$$

Доведення. Будемо вважати, що $0 \in \text{int } M$. За теоремою 1.4.1 існує такий вектор x^*, що

$$\langle x, x^* \rangle \leq \langle x_0, x^* \rangle - \varepsilon \tag{1.7.1}$$

для всіх $x \in M$. Зрозуміло, що

$$s(x^*|M) < \langle x_0, x^* \rangle . \tag{1.7.2}$$

Припустимо, що x^* не визначає крайню опору. Тоді серед векторів $x^i - x^j$ де $i \in I(x^*)$, j – фіксований елемент з $I(x^*)$, існує не більше $n-2$ лінійно незалежних. Виберемо тепер вектор $\overline{x}^*$ так, щоб виконувались співвідношення:

$$\begin{gathered} \left\langle x^i - x^j, \overline{x}^* \right\rangle = 0, \quad i \in I(x^*), \\ \left\langle x^0 - x^j, \overline{x}^* \right\rangle = 0, \\ \left\langle x^0, \overline{x}^* \right\rangle \leq 0. \end{gathered} \tag{1.7.3}$$

Це завжди можна зробити, оскільки серед векторів $x^0 - x^j, x^i - x^j, i \in I(x^*)$, існує не більше $n-1$ лінійно незалежних.

Покажемо, що знайдеться такий індекс $q \in I$, що

$$\left\langle x^q, \overline{x}^* \right\rangle > 0.$$

Дійсно, оскільки за припущенням $0 \in \text{int } M$, то

$$0 < s(\overline{x}^*|M) = \max_{1 \leq i \leq m} \left\langle x^i, \overline{x}^* \right\rangle ,$$

звідки випливає існування вказаного індекса q. В силу (1.7.3) $\left\langle x^i, \overline{x}^* \right\rangle \leq 0$ для $i \in I(x^*)$, тому $q \notin I(x^*)$. Для $i \in I(x^*)$ із тих же співвідношення отримуємо

$$\left\langle x^i x^* + \alpha \overline{x}^* \right\rangle = \left\langle x^i, x^* \right\rangle + \alpha \left\langle x^i, \overline{x}^* \right\rangle = s(x^*|M) + \alpha \left\langle x^0, \overline{x}^* \right\rangle . \tag{1.7.4}$$

Визначимо α_i для $i \notin I(x^*)$ рівнянням

$$\alpha_i = \frac{s(x^*|M) - \left\langle x^i, x^* \right\rangle}{\left\langle x^i, \overline{x}^* \right\rangle - \left\langle x_0, \overline{x}^* \right\rangle}, \quad i \notin I(x^*). \tag{1.7.5}$$

В цій формулі чисельник і знаменник завжди додатні, тому $\alpha_i \neq 0$. Оскільки $\left\langle x^0, \overline{x^*} \right\rangle \leq 0$, а $\langle x^q, \overline{x}^* \rangle > 0$, то $\alpha_q > 0$. Виберемо серед додатних $\alpha_i, i \notin I(x^*)$ найменше: α_p. В силу того, що

$$\left\langle x^i, x^* \right\rangle < s(x^*|M), \quad \notin I(x^*),$$

маємо

$$\left\langle x^i, x^* + \alpha_p \overline{x}^* \right\rangle \leq s(x^*|M) + \alpha_p \left\langle x^0, \overline{x}^* \right\rangle, \quad i \notin I(x^*),$$

причому

$$\langle x^p, x^* + \alpha_p \overline{x}^* \rangle = s(x^*|M) + \alpha_p \left\langle x^0, \overline{x}^* \right\rangle.$$

Порівнюючи останню рівність з співвідношенням (1.7.4), отримуємо

$$s(x^* + \alpha_p \overline{x}^*|M) = s(x^*|M) + \alpha_p \left\langle x^0, \overline{x}^* \right\rangle,$$

$$I(x^* + \alpha_p \overline{x}^*) \supseteq I(x^*) \cup \{p\}.$$

Таким чином, множина $I(x^* + \alpha_p \overline{x}^*)$ містить більше індексів ніж $I(x^*)$. Більш того, вектори $x^i - x^j, i \in I(x^*)$ і $x^p - x^j$ лінійно незалежні, оскільки

$$\left\langle x^i - x^j, \overline{x}^* \right\rangle = 0, \quad i \in I(x^*),$$

а

$$\left\langle x^p - x^j, \overline{x}^* \right\rangle = \langle x^p, \overline{x}^* \rangle - \left\langle x^0, \overline{x}^* \right\rangle > 0,$$

що випливає з формули (1.7.5) і додатності α_p.

Отже показано, що серед векторів $x^i - x^j, i \in I(x^* + \alpha_p \overline{x}^*)$ число лінійно незалежних принаймні на один більше, ніж серед векторів $x^i - x^j, i \in I(x^*)$. Крім цього, для всіх $x \in M$:

$$\langle x, x^* + \alpha_p \overline{x}^* \rangle \leq s(x^* + \alpha_p \overline{x}^*|M) =$$

$$= s(x^*|M) + \alpha_p \langle x_0, \overline{x}^* \rangle < \left\langle x^0, x^* \right\rangle + \alpha_p \left\langle x^0, \overline{x}^* \right\rangle =$$

$$= \left\langle x^0, x^* + \alpha_p \overline{x}^* \right\rangle,$$

тобто вектор $x^* + \alpha_p \overline{x}^*$ відокремлює точку x^0 від M. Якщо серед векторів $x^i - x^j, i \in I(x^* + \alpha_p \overline{x}^*)$, число лінійно незалежних менше ніж $n-1$, то описану вище процедуру можна повторити. □

Очевидно, що крайня опора буде побудована за скінченне число кроків. Неважко побачити, що число крайніх опор скінченне. Дійсно, якщо вектор x^* визначає крайню опору, то він ортогональний до $n-1$ векторів $x^i - x^j, i \in I(x^*)$ і отже визначається однозначно з точністю до скалярного множника. Оскільки $I(x^*)$ є підмножиною скінченної множини I, то число можливих підмножин $I(x^*)$ скінченне, а отже і скінченне число крайніх опор.

Теорема 1.7.3. *Опуклий многогранник може бути заданий скінченною системою лінійних нерівностей.*

Доведення. Нехай $\operatorname{int} M \neq \emptyset$ і вектори $x_k^*, k = 1, \ldots, l$ визначають всі крайні опори до M. Тоді система лінійних нерівностей

$$\langle x, x_k^* \rangle \leq s(x_k^* | M), \quad k = 1, \ldots, l$$

визначає многогранник M. Дійсно, довільна точка $x \in M$ задовольняє цій системі, а довільна точка $x \notin M$ повинна порушувати принаймі одну з цих нерівностей за теоремою 1.7.2.

Якщо $\operatorname{int} M = \emptyset$, то можна розглядати множину $M - x_0, x_0 \in M$, відносно підпростору $\operatorname{Lin} M$, в якому лежить $M - x_0$ і має внутрішні точки. В підпросторі $\operatorname{Lin} M$ многогранник M може бути заданий скінченною системою нерівностей. Сам підпростір $\operatorname{Lin} M$ можна задати скінченним числом лінійних рівнянь. Поєднуючи всі ці співвідношення, отримуємо твердження теореми. □

1.8. МНОГОГРАННІ КОНУСИ

Многогранники представляють собою обмежені множини, які можна задати скінченною системою лінійних нерівностей. Розглянемо важливий клас необмежених множин, які можна задати системою однорідних лінійних нерівностей.

Означення 1.8.1. Конус K називається *многогранним*, якщо існує скінченний набір векторів $x_1, \ldots, x_m$ таких, що для векторів $x \in K$ і тільки для них виконується рівність

$$x = \sum_{i=1}^{m} \lambda_i x_i, \quad \lambda_i \geq 0, \quad i = 1, \ldots, m. \tag{1.8.1}$$

Теорема 1.8.1. *Многогранний конус може бути заданий скінченною кількісттю лінійних однорідних нерівностей*

$$\langle x, x_k^* \rangle \geq 0, \quad k = 1, \ldots, l. \tag{1.8.2}$$

Теорема 1.8.2. *Многогранний конус замкнутий.*

Теорема 1.8.3. *Якщо конус K заданий системою лінійних нерівностей (1.8.2), то спряжений до нього конус K^* є многогранним і складається з елементів*

$$x^* = \sum_{k=1}^{l} \gamma_k x_k^*, \quad \gamma_k \geq 0.$$

Доведення. Розглянемо многогранний конус

$$\widetilde{K} = \left\{ x^* : x^* = \sum_{k=1}^{l} \gamma_k x_k^*, \gamma_k \geq 0 \right\}.$$

Він замкнутий за попередньою теоремою. За означенням $x \in \left(\widetilde{K}\right)^*$, якщо

$$\left\langle x, \sum_{k=1}^{l} \gamma_k x_k^* \right\rangle \geq 0, \quad \gamma_k \geq 0, k = 1, \ldots, l.$$

А це можливо лише в тому випадку, коли

$$\langle x, x_k^* \rangle \geq 0, \quad k = 1, \ldots, l,$$

тобто $x \in K$. Отже $K = \left(\widetilde{K}\right)^*$. Оскільки $\widetilde{K}$ замкнутий, то $K^* = \left(\widetilde{K}\right)^{**} = \widetilde{K}$, що й потрібно було довести. □

Теорема 1.8.4. *Заданий системою лінійних нерівностей конус є многогранним конусом.*

Доведення. Нехай конус K заданий системою нерівностей (1.8.2). Оскільки спряжений до нього конус є многогранним, то існують такі точки $x_i, i = 1, \ldots, m$, що точки $x^* \in K^*$ і тільки вони задовольняють нерівності

$$\langle x_i, x^* \rangle \geq 0, \quad i = 1, \ldots, m.$$

Оскільки конус K замкнутий, то $K = (K^*)^*$ і на основі теореми 1.8.3

$$x = \sum_{i=1}^{m} \lambda_i x_i, \quad \lambda_i \geq 0,$$

тобто K – многогранний конус. □

Теорема 1.8.5. *Сума многограних конусів є многогранним конусом.*

Теорема 1.8.6. *Перетин многогранних конусів є многогранним конусом.*

Теорема 1.8.7. *Спряжений до многогранного конус є многогранним конусом.*

Теорема 1.8.8. *Для многогранних конусів $K_1, \ldots, K_m$ виконується співвідношення*

$$(K_1 \cap \cdots \cap K_m)^* = K_1^* + \cdots + K_m^*.$$

1.9. МНОГОГРАННІ МНОЖИНИ

Означення 1.9.1. Множина точок, що задовольняють систему лінійних нерівностей

$$\langle x,x_i^*\rangle \geq \alpha_i, \quad i=1,\ldots,l, \tag{1.9.1}$$

називається *многогранною*.

Теорема 1.9.1. *Многогранна множина є сумою многогранника і многогранного конуса, і навпаки, сума многогранника і многогранного конуса є многогранною множиною.*

Доведення. Введемо додаткову координату x^0 і запишемо систему нерівностей у такому вигляді

$$\langle x,x_i^*\rangle - x^0\alpha_i \geq 0, \quad i=1,\ldots,l, \quad x^0 \geq 0. \tag{1.9.2}$$

Це однорідна система лінійних нерівностей в $\mathbb{R}^{n+1}$. За теоремою 1.8.4 вона задає многогранний конус, елементи якого можна подати так

$$\begin{pmatrix} x^0 \\ x \end{pmatrix} = \sum_{j=1}^{m} \lambda_j \begin{pmatrix} x_j^0 \\ x_j \end{pmatrix}, \lambda_j \geq 0, \tag{1.9.3}$$

де через $\begin{pmatrix} x^0 \\ x \end{pmatrix}$ позначений $(n+1)$ мірний вектор з компонентами $x^0,\ldots,x^n$. При будь-яких $\lambda_j \geq 0$ вираз в правій частині формули (1.9.3) повинен задовольняти нерівностям (1.9.2). Зокрема, якщо $\lambda_j = 1$, а $\lambda_i = 0$ при $i \neq j$, то з нерівностей випливає, що $x_j^0 \geq 0$. Нехай

$$I^0 = \left\{j : x_j^0 = 0, j=1,\ldots,m\right\},$$

$$I^+ = \left\{j : x_j^0 > 0, j=1,\ldots,m\right\}.$$

Позначимо $y_j = x_j/x_j^0, \gamma_j = \lambda_j x_j^0, j \in I^+$ і перепишемо формули (1.9.3) в наступному вигляді:

$$x^0 = \sum_{j\in I^+} \lambda_j x_j^0 = \sum_{j\in I^+} \gamma_j, \quad \gamma_j \geq 0, \tag{1.9.4}$$

$$x = \sum_{j\in I^0} \lambda_j x_j + \sum_{j\in I^+} \lambda_j x_j^0 \left(\frac{x_j}{x_j^0}\right) =$$

$$= \sum_{j\in I^0} \lambda_j x_j + \sum_{j\in I^+} \gamma_j y_j, \quad \gamma_j \geq 0. \tag{1.9.5}$$

Отже кожний розв'язок системи (1.9.2) можна представити в вигляді (1.9.4), (1.9.5). Але розв'язок системи (1.9.1) отримується з розв'язку системи (1.9.2), якщо покласти $x^0 = 1$. Отже кожен розв'язок системи (1.9.1) можна представити у вигляді

$$x = \sum_{j \in I^0} \lambda_j x_j + \sum_{j \in I^+} \gamma_j y_j, \quad \lambda_j \geq 0 \tag{1.9.6}$$

$$\sum_{j \in I^+} \gamma_j = 1, \quad \gamma_j \geq 0, \quad j \in I^+. \tag{1.9.7}$$

Перша сума в правій частині формули (1.9.6) визначає точки деякого многогранного конуса, друга сума визначає точки деякого многогранника. Те, що сума многогранника і многогранного конуса є многогранною множиною, доводиться аналогічними міркуваннями. □

Теорема 1.9.2. *Нехай K – опуклий конус, K_0 – многогранний конус і нехай* $\mathrm{ri}\, K \cap K_0 \neq \emptyset$. *Тоді*

$$(K \cap K_0)^* = K^* + K_0^*.$$

Доведення. Оскільки $K \subseteq \mathrm{Lin}\ K$, то $K^* \supseteq (\mathrm{Lin}\ K)^\perp$, де $(\mathrm{Lin}\ K)^\perp$– ортогональне доповнення $\mathrm{Lin}\ K$ до всього простору X. Очевидно, що $K \cap K_0 \subseteq \mathrm{Lin}\ K$. Візьмемо як основний простір $\mathrm{Lin}\ K$ і знайдемо спряжений конус $(K \cap K_0)^*$ відносно цього простору. Оскільки $\mathrm{ri}\ K \cap K_0 \neq \emptyset$, то в просторі $\mathrm{Lin}\ K$ можна застосувати теорему 1.5.2, тобто

$$(K \cap K_0)^*_{\mathrm{Lin}\ K} = (K \cap (K_0 \cap \mathrm{Lin}\ K))^*_{\mathrm{Lin}\ K} = K^*_{\mathrm{Lin}\ K} + (K_0 \cap \mathrm{Lin}\ K)^*_{\mathrm{Lin}\ K}.$$

Можна переконатися, що для довільного конуса K_1, що лежить в $\mathrm{Lin}\ K$,

$$K_1^* = (K_1)^*_{\mathrm{Lin}\ K} + (\mathrm{Lin}\ K)^\perp,$$

тому

$$(K \cap K_0)^* = K^*_{\mathrm{Lin}\ K} + (K_0 \cap \mathrm{Lin}\ K)^*_{\mathrm{Lin}\ K} + (\mathrm{Lin}\ K)^\perp =$$

$$= \left(K^*_{\mathrm{Lin}\ K} + (\mathrm{Lin}\ K)^\perp\right) + \left((K_0 \cap \mathrm{Lin}\ K)^*_{\mathrm{Lin}\ K} + (\mathrm{Lin}\ K)^\perp\right) =$$

$$= K^* + (K_0 \cap \mathrm{Lin}\ K)^*.$$

Але $\mathrm{Lin}\ K$- многогранник конус, причому $(\mathrm{Lin}\ K)^* = (\mathrm{Lin}\ K)^\perp$, тому по теоремі 1.8.8

$$(K_0 \cap \mathrm{Lin}\ K)^* = K_0^* + (\mathrm{Lin}\ K)^\perp.$$

Враховуючи, що $K^* \supseteq (\mathrm{Lin}\ K)^\perp$, отримуємо

$$(K \cap K_0)^* = K^* + K_0^* + (\mathrm{Lin}\ K)^\perp = K^* + K_0^*.$$

□

Розділ II
Опуклі функції

2.1. ОЗНАЧЕННЯ ТА ОСНОВНІ ВЛАСТИВОСТІ

Будемо розглядати функції $f\colon \mathbb{R}^n \to \overline{\mathbb{R}} = \mathbb{R}\cup\{-\infty\}\cup\{+\infty\}$, які визначені на просторі $X = \mathbb{R}^n$ і можуть приймати нескінченні значення $-\infty$ та $+\infty$. Для кожної такої функції дамо означення множин, які визначають функцію.

Означення 2.1.1. *Ефективна множина* функції $f(x)$ – це непорожня множина

$$\mathrm{dom}\ f\colon = \{x \in \mathbb{R}^n\colon f(x) < \infty\}$$

точок, в яких функція приймає скінченні значення та значення $-\infty$.

Означення 2.1.2. *Надграфік (епіграф)* функції $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$, яка не дорівнює тотожно $+\infty$, це непорожня множина

$$\mathrm{epi}\ f\colon = \{(x,r) \in \mathbb{R}^n \times \mathbb{R}\colon f(x) \le r\}.$$

Зауважимо, що точка $(x,r) \in \mathbb{R}^n \times \mathbb{R}$ належить epi f лише тоді, коли $x \in \mathrm{dom}\ f$.

Надграфік epi f функції $f(x)$ визначає цю функцію. Дійсно,

$$f(x) = \inf\{r \in \mathbb{R}\colon (x,r) \in \mathrm{epi}\ f\}. \tag{2.1.1}$$

Якщо в просторі $\mathbb{R}^n \times \mathbb{R}$ взяти множину S, яка разом з кожною точкою $(x,r) \in S$ містить всі точки $(x,r'),r' \ge r$, то ця множина буде надграфіком функції, яка визначається співвідношенням (2.1.1). Отже між множинами в просторі $\mathbb{R}^n \times \mathbb{R}$, що мають вказану властивість, і функціями $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ існує тісний зв'язок. Цей зв'язок дозволяє вивчати властивості функцій (множин) знаючи відповідні властивості множин (функцій).

Означення 2.1.3. Функція $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$, що не дорівнює тотожно $+\infty$, називається *опуклою*, якщо надграфік (епіграф) epi f функції f – опукла множина в $\mathbb{R}^n \times \mathbb{R}$.

Означення 2.1.4. Функція $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ називається *власною*, якщо вона не приймає значення $-\infty$ і не дорівнює тотожно $+\infty$.

З означення випливає, що для власної функції $f(x)$ її ефективна множина $\mathrm{dom}\ f \ne \emptyset$ і така функція $f(x)$ приймає скінченні значення в точках $x \in \mathrm{dom}\ f$.

Для власної функції означення опуклості через надграфік можна замінити на більш звичне означення.

Лема 2.1.1. *Для того, щоб власна функція $f(x)$ була опуклою, необхідно і достатньо, щоб справджувалась нерівність*

$$f\left(\lambda_1 x_1 + \lambda_2 x_2\right) \leqslant \lambda_1 f\left(x_1\right) + \lambda_2 f\left(x_2\right) \tag{2.1.2}$$

для всіх $x_1,x_2 \in \mathbb{R}^n$, та всіх $\lambda_1 \geq 0,\lambda_2 \geq 0,\lambda_1 + \lambda_2 = 1$.

Лема 2.1.2. *Ефективна множина* $\operatorname{dom} f$ *опуклої функції $f(x)$ – опукла множина.*

Зауважимо, що ефективна множина $\operatorname{dom} f$ опуклої функції $f(x)$ є опуклою навіть у тому випадку коли функція $f(x)$ невласна. Проте $f(x) = -\infty$ для всіх $x \in \operatorname{ri}\operatorname{dom} f$ якщо функція $f(x)$ невласна. Отже невласна опукла функція може приймати скінченні значення лише на відносній границі своєї області визначення.

Лема 2.1.3. *Нехай f – власна опукла функція. Тоді справджується нерівність Ієнсена*

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leqslant \sum_{i=1}^{m} \lambda_i f\left(x_i\right), \tag{2.1.3}$$

для всіх $m = 1,2\ldots;\ x_i \in X,\ \lambda_i \geqslant 0,\ i = 1,\ldots,m,\ \sum_{i=1}^{m} \lambda_i = 1$.

Лема 2.1.4. *Нехай $f_1(x),\ldots,f_m(x)$ – власні опуклі функції, $\alpha_1,\ldots,\alpha_m$ – невід'ємні числа. Тоді опукла функція*

$$f\left(x\right) = \sum_{i=1}^{m} \alpha_i f_i\left(x\right).$$

Лема 2.1.5. *Нехай $f_1(x),\ldots,f_m(x)$ – власні опуклі функції. Тоді опукла функція*

$$\left(\oplus_{i=1}^{m} f_i\right)(x) = \inf\left\{\sum_{i=1}^{m} f_i\left(x_i\right) :\ \sum_{i=1}^{m} x_i = x\right\},$$

що називається інфімальною конволюцією функцій $f_1(x),\ldots,f_m(x)$.

Лема 2.1.6. *Нехай $f_i(x),i \in I = \{1,\ldots,m\}$, – опуклі функції. Тоді опукла функція*

$$f(x) = \left(\vee_{i\in I} f_i\right)(x) = \sup_{i\in I}\left\{f_i(x)\right\},$$

що називається верхньою гранню функцій $f_i(x)$, $i \in I$.

Лема 2.1.7. *Нехай $f_i(x), i \in I$, – опуклі функції. Тоді опукла функція*

$$f(x) = (\mathrm{conv}\,(\wedge_{i\in I})\,f_i(x)) = \inf\{\alpha \in \mathbb{R}\colon (x,\alpha) \in \mathrm{conv}\,(\cup_{i\in I}\,\mathrm{epi}\;f_i)\}\,,$$

що називається опуклою оболонкою нижньої грані функцій $f_i(x)$, $i \in I$.

Лема 2.1.8. *Нехай $\Lambda\colon X \to Y$ – лінійний оператор, f - опукла функція на X, g - опукла функція на Y. Тоді опуклі функції $g\Lambda$ (називається прообразом функції g при відображенні Λ) та Λf (називається образом функції f при відображенні Λ), які задаються співвідношеннями*

$$(g\Lambda)(x) = g(\Lambda x),$$

$$(\Lambda f)(y) = \inf\{f(x)\colon x \in X, \Lambda x = y\}.$$

Покажемо, для прикладу, які результати дають вказані операції над індикаторними функціями:

$$\delta(\cdot|A_1) + \delta(\cdot|A_2) = \delta(\cdot|A_1) \vee \delta(\cdot|A_2) = \delta(\cdot|A_1 \cap A_2),$$

$$\delta(\cdot|A_1) \oplus \delta(\cdot|A_2) = \delta(\cdot|A_1 + A_2),$$

$$\mathrm{conv}\,(\delta(\cdot|A_1) \wedge \delta(\cdot|A_2)) = \delta(\cdot|\,\mathrm{conv}(A_1 \cup A_2)),$$

$$\delta(\cdot|A)\Lambda = \delta(\cdot|A\Lambda), \quad \Lambda\delta(\cdot|A) = \delta(\cdot|\Lambda A).$$

Наведемо декілька критеріїв, що дозволяють розпізнати опуклу функцію.

Лема 2.1.9. *Нехай $f(x)$ – власна функція, та нехай $g_{x,p}(\alpha) = f(x+\alpha p)$, $\alpha \in \mathbb{R}$, $p \in \mathbb{R}^n$. Функція $f(x)$ опукла тоді і тільки тоді, коли функція $g_{x,p}(\alpha)$ опукла за α для всіх $x \in \mathbb{R}^n$ та $p \in \mathbb{R}^n$.*

Теорема 2.1.1. Критерії опуклості диференційовних функцій. *Нехай функція $f(x)$ диференційовна. Наступні твердження еквівалентні:*

- (i) *функція f опукла;*
- (ii) *для всіх $x,\hat{x} \in \mathbb{R}^n$*
$$f(x) - f(\hat{x}) \geqslant \langle f'(\hat{x}), x - \hat{x}\rangle; \tag{2.1.4}$$
- (iii) *для всіх $x,\hat{x} \in \mathbb{R}^n$*
$$\langle f'(x) - f'(\hat{x}), x - \hat{x}\rangle \geqslant 0; \tag{2.1.5}$$
- (iiii) *функція $\langle p, f'(x + \alpha p)\rangle$ неспадна як функція від $\alpha \in \mathbb{R}$ для всіх $x \in \mathbb{R}^n$ та $p \in \mathbb{R}^n$;*
- (iiiii) *за умови, що функція $f(x)$ двічі неперервно диференційовна:*
$$\langle f''(x)h, h\rangle \geqslant 0 \quad \text{для всіх} \quad x,h \in \mathbb{R}^n; \tag{2.1.6}$$

Лема 2.1.10. *Нерівності для опуклих функцій. Нехай $g(\alpha)$ - опукла функція від $\alpha \in \mathbb{R}$, нехай $\alpha_0 < \alpha_1 < \alpha_2$; α_0,α_1,$\alpha_2 \in \operatorname{dom} g$. Тоді*

$$\frac{g(\alpha_2) - g(\alpha_0)}{\alpha_2 - \alpha_0} \geq \frac{g(\alpha_1) - g(\alpha_0)}{\alpha_1 - \alpha_0},$$

$$\frac{g(\alpha_1) - g(\alpha_0)}{\alpha_1 - \alpha_0} \leq \frac{g(\alpha_2) - g(\alpha_1)}{\alpha_2 - \alpha_1}.$$

Доведення. Покладемо

$$\lambda_1 = \frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0}, \quad \lambda_2 = 1 - \lambda_1 = \frac{\alpha_2 - \alpha_1}{\alpha_2 - \alpha_0}.$$

Тоді

$$\lambda_1 \alpha_2 + \lambda_2 \alpha_0 = \frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0} \alpha_2 + \frac{\alpha_2 - \alpha_1}{\alpha_2 - \alpha_0} \alpha_0 = \alpha_1.$$

Тому

$$g(\alpha_1) = g(\lambda_1 \alpha_2 + \lambda_2 \alpha_0) \leq \frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0} g(\alpha_2) + \frac{\alpha_2 - \alpha_1}{\alpha_2 - \alpha_0} g(\alpha_0).$$

Отриману нерівність можна перетворити кількома способами:

1) Віднімаючи від обох частин $g(\alpha_0)$, отримаємо

$$g(\alpha_1) - g(\alpha_0) \leq \frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0} (g(\alpha_2) - g(\alpha_0)),$$

звідки, після ділення на $\alpha_1 - \alpha_0$, випливає перша нерівність леми.

2) Оскільки $\lambda_1 + \lambda_2 = 1$, то

$$\frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0} g(\alpha_1) + \frac{\alpha_2 - \alpha_1}{\alpha_2 - \alpha_0} g(\alpha_1) \leq \frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0} g(\alpha_2) + \frac{\alpha_2 - \alpha_1}{\alpha_2 - \alpha_0} g(\alpha_0),$$

або

$$\frac{\alpha_2 - \alpha_1}{\alpha_2 - \alpha_0} [g(\alpha_1) - g(\alpha_0)] \leq \frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0} [g(\alpha_2) - g(\alpha_1)],$$

звідки отримаємо другу нерівність леми. $\square$

Зауважимо, що перша з нерівностей, які доведені в лемі, показує, що функція

$$\frac{g(\alpha) - g(\alpha_0)}{\alpha - \alpha_0}$$

монотонно неспадна при зростанні $\alpha \geq \alpha_0$.

2.2. НЕПЕРЕРВНІСТЬ ОПУКЛИХ ФУНКЦІЙ

Теорема 2.2.1. *Нехай власна опукла функція f обмежена зверху в деякому околі точки $\hat{x}$. Тоді функція f неперервна в цій точці.*

Доведення. Нехай, для простоти, $\hat{x} = 0$. Нехай B – відкрита куля з центром в нулі така, що $f(x) \le c_1$ для всіх $x \in B$. Розглянемо функцію $g(\alpha) = f(\alpha x)$ при фіксованому $x \in B$. Взявши в першій нерівності леми 2.1.10 $\alpha_0 = 0, \alpha_1 = \alpha, \alpha_2 = 1$, отримаємо

$$\frac{g(\alpha) - g(0)}{\alpha} \le \frac{g(1) - g(0)}{1}.$$

Оскільки

$$g(1) = f(x) \le c_1, \quad g(0) = f(0) \le c_1,$$

то

$$f(\alpha x) - f(0) \le 2c_1\alpha.$$

Якщо у другій нерівності леми 2.1.10 взяти $\alpha_0 = -1, \alpha_1 = 0, \alpha_2 = \alpha$, то отримаємо

$$\frac{g(0) - g(-1)}{0 - (-1)} \le \frac{g(\alpha) - g(0)}{\alpha},$$

звідки

$$-2c_1\alpha \le f(\alpha x) - f(0).$$

Отже

$$|f(\alpha x) - f(0)| \le 2c_1\alpha. \tag{2.2.1}$$

Візьмемо тепер $\varepsilon > 0$, $\delta = \varepsilon/(2c_1) < 1$, $B_\delta = \delta B$. Нехай $y \in B_\delta$. Тоді існує такий вектор $x \in B$, що $y = \delta x$. Отже

$$|f(y) - f(0)| = |f(\delta x) - f(0)| \le 2c_1\delta = \varepsilon.$$

Це доводить неперервність функції в точці $\hat{x} = 0$. □

Теорема 2.2.2. *Якщо опукла функція f неперервна в точці $\hat{x}$, то вона задовольняє умову Ліпшиця в деякому околі точки $\hat{x}$, тобто*

$$|f(x) - f(\hat{x})| \le L\|x - \hat{x}\|$$

для всіх x з деякого околу точки $\hat{x}$.

Доведення. Нехай, для простоти, $\hat{x} = 0$. Нехай B_r – відкрита куля радіуса r з центром в нулі. Візьмемо $y \in B_r$, $\|y\| < r/2$, $x = \frac{r}{2}\frac{y}{\|y\|}$. З нерівності (2.2.1) випливає, що

$$|f(y) - f(0)| = \left| f\left(\frac{2\|y\|}{r}x\right) - f(0)\right| \le L\|y\|,$$

де $L = 4c/r$. □

Теорема 2.2.3. *Нехай $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ – власна опукла функція. Тоді функція f неперервна на множині* ri dom f.

Означення 2.2.1. Функція $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ називається *замкнутою*, якщо її надграфік epi f - замкнута множина в $\mathbb{R}^n \times \mathbb{R}$.

Теорема 2.2.4. *Наступні властивості функції f еквівалентні:*

- (i) *функція f напівнеперервна знизу на $\mathbb{R}^n$;*
- (ii) *функція f замкнута;*
- (iii)*множини рівня $S_r(f) = \{x \in \mathbb{R}^n\colon f(x) \leqslant r\}$ функції f замкнуті в $\mathbb{R}^n$ для всіх $r \in \mathbb{R}$.*

Теорема 2.2.5. *Якщо $f(x)$ – замкнута опукла функція, що приймає в деякій точці $\hat{x}$ скінченне значення, то $f(x) > -\infty$ для всіх x.*

Означення 2.2.2. *Замикання (напівнеперервна знизу оболонка)* функції $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ визначається співвідношенням

$$\mathrm{cl}\ f(x) = \liminf_{y \to x} f(y) \quad \forall\, x \in \mathbb{R}^n,$$

або (еквівалентно)

$$\mathrm{epi}\,(\mathrm{cl}\ f)\colon = \mathrm{cl}\,(\mathrm{epi}\ f).$$

Теорема 2.2.6. *Замикання власної опуклої функції $f(x)$ можна подати як супремум опорних до $f(x)$ афінних функцій*

$$\mathrm{cl}\ f(x) = \sup_{(s,b) \in \mathbb{R}^n \times \mathbb{R}} \{\langle s,x \rangle - b\colon \langle s,y \rangle - b \leq f(y)\, \forall\, y \in \mathbb{R}^n\}\,. \tag{2.2.2}$$

2.3. СПРЯЖЕНІ ФУНКЦІЇ. ПЕРЕТВОРЕННЯ ЮНГА-ФЕНХЕЛЯ

Означення 2.3.1. Нехай функція $f(x)$ визначена на просторі X. Функція $f^*(x^*)$ на спряженому просторі X^*, яка визначена рівністю

$$f^*(x^*) = \sup_x \{\langle x^*,x \rangle - f(x)\}\,, \tag{2.3.1}$$

називається *перетворенням Юнга-Фенхеля* функції f або *спряженою функцією* до функції f.

Означення 2.3.2. Нехай функція $g(x^*)$ визначена на просторі X^*. Спряжена функція $g^*(x)$ на просторі X визначається рівністю

$$g^*(x) = \sup_{x^*} \{\langle x^*,x \rangle - g(x^*)\}\,. \tag{2.3.2}$$

Спряжена функція $f^{**} = (f^*)^*$ на просторі X називається *другою спряженою функцією* до функції f.

Зауваження 2.3.1. Зауважимо, що у тому випадку, коли $f\,(x)$ – опукла гладка функція на $\mathbb{R}^n$, що росте на нескінченності швидше лінійної функції, перетворення Юнга–Фенхеля – це перетворення Лежандра:

$$f^*\,(x^*) = \langle x^*,x_0 \rangle - f\,(x_0)\,,$$

де x_0 визначається рівністю $x^* = f'(x_0)$.

Приклад 2.3.1. Нехай $A \subset X$ і $f(x) = \delta(x|A)$ – індикаторна функція множини A. Тоді

$$f^*(x^*) = \sup\{\langle x^*,x\rangle : x \in A\} = s(x^*|\, A),$$

тобто спряженою з індикаторною функцією множини A буде опорна функція цієї множини.

Множина

$$A^0 = \{x^* \in X^* \colon s(x^*|\, A) \leqslant 1\}$$

називається полярою множини A. Якщо K – конус, то K^0 теж конус:

$$K^0 = \{x^* \in X^* \colon \langle x^*,x\rangle \leqslant 0, \quad \forall x \in K\}.$$

Конус $K^* = -K^0$ називають спряженим з конусом K. Якщо L – підпростір X, то

$$L^0 = \{x^* \in X^* \colon \langle x^*,x\rangle = 0, \quad \forall x \in L\} = L^{\perp}$$

– анулятор підпростору L.

Приклад 2.3.2. Нехай $A \subset X$. Функція

$$f(x) = \mu(x|\, A) = \begin{cases} 0, & x = 0 \\ \inf\left\{\lambda > 0 \colon \lambda^{-1}x \in A\right\}, & x \neq 0 \end{cases}$$

називається функцією Мінковського множини A. Знайдемо до неї спряжену:

$$f^*(x^*) = \sup\{\langle x^*,x\rangle - f(x) \colon x \in X\} =$$

$$= \sup\left\{\langle x^*,x\rangle - \inf\left\{\lambda > 0 \colon \lambda^{-1}x \in A\right\} : x \in X\right\} =$$

$$= \sup\left\{\langle x^*,x\rangle - \lambda \colon \lambda > 0, \lambda^{-1}x \in A\right\} =$$

$$\sup\{\sup\{\langle x^*,x\rangle \colon x \in \lambda A\} - \lambda \colon \lambda > 0\} =$$

$$= \sup_{\lambda > 0}\left(\lambda\left(\sup_{x \in A}\langle x^*,x\rangle - 1\right)\right) = \delta\left(x^*|\, A^0\right).$$

Отже спряженою до функції Мінковського множини A є індикаторна функція поляри A^0 множини A.

Спряжена функція f^* є супремумом сімейства лінійних функцій від x^*. Тому справджується наступна лемма.

Лема 2.3.1. *Спряжена функція f^* опукла і замкнута (незалежно від того, чи є опуклою сама функція f).*

Із визначення спряженої функції випливає справедливість наступної нерівності, що називається *нерівністю Юнга–Фенхеля*.

Лема 2.3.2. *Для всіх $x \in X$, $x^* \in X^*$ справджується нерівність*

$$f(x) + f^*(x^*) \geqslant \langle x^*, x\rangle . \tag{2.3.3}$$

Лема 2.3.3. *Для всякої функції f справедлива нерівність*

$$f \geqslant f^{**}.$$

Доведення. З нерівністю Юнга–Фенхеля маємо

$$f^{**}(x) = \sup_{x*}\left(\langle x^*, x\rangle - f^*(x^*)\right) \leqslant f(x).$$

□

Лема 2.3.4. *Нехай f – власна замкнута опукла функція на X. Тоді f^* – власна функція.*

Доведення. Якщо $x_0 \in \operatorname{dom} f$, то

$$f^*(x^*) \geqslant \langle x^*, x_0\rangle - f(x_0) > -\infty$$

для всіх x^*. З іншого боку, точка $(f(x_0) - 1, x_0) \notin \operatorname{epi} f$. Тоді за теоремою про розділення існує пара (β_0, y_0^*) така, що

$$\sup_{(\alpha,x)\in \operatorname{epi} f} \left(\beta_0 \alpha + \langle y_0^*, x\rangle\right) < \beta_0 \left(f(x_0) - 1\right) + \langle y_0^*, x_0\rangle .$$

Зрозуміло, що $\beta_0 \neq 0$. З іншого боку, β_0 не може бути додатнім числом, тому що інакше верхня грань зліва дорівнює $+\infty$. Поділивши на $|\beta_0|$, одержуємо:

$$\sup_{x\in \operatorname{dom} f}\left(\left\langle y_0^* |\beta_0|^{-1}, x\right\rangle - f(x)\right) = f^*\left(y_0^* |\beta_0|^{-1}\right)$$

$$< 1 - f(x_0) + \left\langle y_0^* |\beta_0|^{-1}, x_0\right\rangle < \infty.$$

□

З доведеного випливає, що власна замкнута опукла функція $f(x)$ обмежена знизу на кожній обмеженій підмножині простору X, оскільки вона мажорує принаймні одну афінну функцію.

Лема 2.3.5. *Нехай $\Lambda : X \to Y$ – лінійний гомеоморфізм X на Y, і нехай g – функція на Y. Покладемо*

$$f(x) = \lambda g(\Lambda x + y_0) + \langle x_0^*, x\rangle + \gamma_0,$$

де $y_0 \in Y$, $x_0^* \in X^*$, $\gamma_0 \in \mathbb{R}$, $\lambda > 0$. Тоді

$$f^*(x^*) = \lambda g^*\left(\lambda^{-1}\left(\Lambda^{-1}\right)^*(x^* - x_0^*)\right) - \left\langle x^* - x_0^*, \Lambda^{-1} y_0\right\rangle - \gamma_0.$$

Доведення. Маємо

$$f^*(x^*) = \sup_x\left(\langle x^*, x\rangle - \lambda g(\Lambda x + y_0) - \langle x_0^*, x\rangle - \gamma_0\right) =$$

(покладаючи $y = \Lambda x + y_0$)

$$= \lambda \sup_y \left(\lambda^{-1}\left\langle \left(\Lambda^{-1}\right)^* (x^* - x_0^*), y\right\rangle - g(y)\right) - \left\langle x^* - x_0^*, \Lambda^{-1} y_0\right\rangle - \gamma_0$$

$$= \lambda g^* \left(\lambda^{-1}\left(\Lambda^{-1}\right)^* (x^* - x_0^*)\right) - \left\langle x^* - x_0^*, \Lambda^{-1} y_0\right\rangle - \gamma_0.$$

□

З цієї леми випливають, зокрема, такі властивості

$$f(x) = g(x + x_0) \Rightarrow f^*(x^*) = g^*(x^*) - \langle x^*, x_0\rangle;$$

$$f(x) = g(x) + \langle x_0^*, x\rangle \Rightarrow f^*(x^*) = g^*(x^* - x_0^*);$$

$$f(x) = \lambda g(x), \lambda > 0 \Rightarrow f^*(x^*) = \lambda g^*\left(\lambda^{-1} x^*\right);$$

$$f(x) = \lambda g\left(\lambda^{-1} x\right), \lambda > 0 \Rightarrow f^*(x^*) = \lambda g^*(x^*);$$

$$f(x) = g(\lambda x), \lambda > 0 \Rightarrow f^*(x^*) = g^*\left(\lambda^{-1} x^*\right).$$

Теорема 2.3.1. Теорема Фенхеля–Моро. *Нехай f – функція на X, що всюди більша $-\infty$. Тоді $f = f^{**}$ тоді і тільки тоді, коли f опукла і замкнута.*

Доведення. Якщо $f = f^{**}$, то f опукла і замкнута. Далі, якщо $f(x) \equiv +\infty$, то рівність $f = f^{**}$ очевидна. Нам досить перевірити, що для власної опуклої і замкнутої функції f справедлива нерівність $f \leqslant f^{**}$.
Ідею подальшого доведення і його основних етапів ілюструє рисунок 2.3.1.

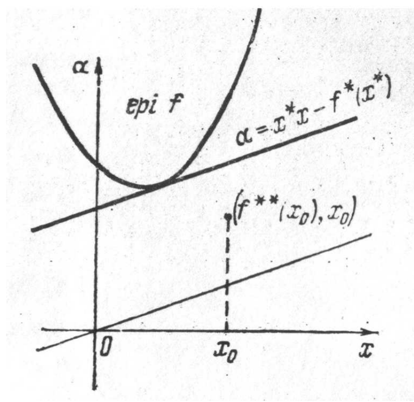


Рис. 2.3.1: Теорема Фенхеля–Моро.

Припустимо, що $f^{**}(x_0) < f(x_0)$ в деякій точці $x_0 \in \operatorname{dom} f^{**}$. Тоді epi f як непорожню опуклу і замкнуту множину можна сильно розділити з точкою $(f^{**}(x_0), x_0)$ ненульовим лінійним функціоналом, що визначається $(\beta, y^*) \in \mathbb{R} \times X^*$, тобто

$$\beta f^{**}(x_0) + \langle y^*, x_0 \rangle > \sup \left\{ \beta\alpha + \langle y^*, y \rangle : \ (\alpha, y) \in \operatorname{epi} f \right\}. \qquad (2.3.4)$$

Величина $\beta \leqslant 0$, оскільки при $\beta > 0$ верхня грань праворуч дорівнює $+\infty$. Якщо $\beta = 0$, то, вибравши $y_1^* \in \operatorname{dom} f^*$ ($\operatorname{dom}\ f^* \neq \emptyset$), одержимо для $t > 0$

$$f^*(y_1^* + ty^*) = \sup \{ \langle y_1^* + ty^*, y \rangle - f(y) \colon y \in \operatorname{dom}\ f \} \leqslant$$

$$\leqslant \sup \left\{ \langle y_1^*, y \rangle - f(y) \colon y \in \operatorname{dom} f \right\} + t \sup \left\{ \langle y^*, y \rangle \colon y \in \operatorname{dom} f \right\} =$$

$$= f^*(y_1^*) + t \sup \{ \langle y^*, y \rangle \colon y \in \operatorname{dom}\ f \}.$$

Враховуючи (2.3.4), отримаємо

$$f^{**}(x_0) \geqslant \langle y_1^* + ty^*, x_0 \rangle - f^*(y_1^* + ty^*) \geqslant$$

$$\geqslant \langle y_1^*, x_0 \rangle - f^*(y_1^*) + t \left[\langle y^*, x_0 \rangle - \sup \left\{ \langle y^*, y \rangle \colon y \in \operatorname{dom} f \right\} \right] \to \infty$$

при $t \to \infty$, тобто $x_0 \notin \operatorname{dom} f^{**}$, що суперечить припущенню. Таким чином, випадок $\beta = 0$ теж виключається. Залишається випадок $\beta < 0$. Поділивши обидві частини нерівності (2.3.4) на $|\beta|$ і покладаючи $x^* = |\beta|^{-1} y^*$, одержимо

$$\langle x^*, x_0 \rangle - f^{**}(x_0) > \sup \left\{ \langle x^*, y \rangle - \alpha \colon \ (\alpha, y) \in \operatorname{epi} f \right\} = f^*(x^*).$$

Це означає, що гіперплощина $\alpha = \langle x^*, x \rangle - f^*(x^*)$ проходить вище точки $(f^{**}(x_0), x_0)$:

$$\langle x^*, x \rangle > f^{**}(x_0) + f^*(x^*).$$

Це суперечить нерівності Юнга–Фенхеля. Теорема доведена. □

З теореми відразу випливає "двоїсте" описання власних замкнутих опуклих функцій.

Наслідок 2.3.1. *Власна замкнута опукла функція $f(x)$ на X збігається з верхньою гранню сімейства всіх неперервних афінних функцій, що не перевищують $f(x)$.*

Доведення. За теоремою Фенхеля–Моро $f(x)$ є верхньою гранню сімейства афінних функцій вигляду

$$x \to \langle x^*, x \rangle - f^*(x^*) \qquad \left(x \in \operatorname{dom} f^* \right)$$

і тим більше всіх неперервних афінних функцій, що не перевищують $f(x)$. $\square$

Наслідок 2.3.2. *Якщо $\overline{\text{conv}}\, f$ – власна функція, то $f^{**} = \overline{\text{conv}}\, f$.*

Доведення. Надграфік epi f^{**} – опукла замкнута множина, яка містить epi f (оскільки $f \geqslant f^{**}$). Тому epi $f \subset \overline{\text{conv}}\,(\text{epi } f) \subset \text{epi } f^{**}$. Це значить, що $f \geqslant \overline{\text{conv}}\, f \geqslant f^{**}$. З лівої нерівності випливає, що $f^* \leqslant (\overline{\text{conv}}\, f)^*$ і, отже, $f^{**} \geqslant (\overline{\text{conv}}\, f)^{**} = \overline{\text{conv}}\, f$, оскільки $\overline{\text{conv}}\, f$ – опукла замкнута функція. $\square$

2.4. ТЕОРЕМИ ДВОЇСТОСТІ

Теореми двоїстості показують, як пов'язані перетворення Юнга–Фенхеля функції, яку отримали в результаті тих чи інших операцій, з перетвореннями Юнга–Фенхеля вихідних функцій. Виявляється, що операції над опуклими функціями розпадаються на пари двоїстих операцій.

Сформулюємо спочатку теореми двоїстості, а потім перейдемо до їх доведень.

Теорема 2.4.1. *Нехай $f_1(x), \ldots, f_n(x)$ – функції на X. Тоді*

$$(f_1 \oplus f_2 \oplus \ldots \oplus f_n)^* = f_1^* + f_2^* + \ldots + f_n^*,$$

$$(f_1 + f_2 + \ldots + f_n)^* \leqslant f_1^* \oplus f_2^* \oplus \ldots \oplus f_n^*.$$

Якщо ж $f_1, \ldots, f_n$ – власні опуклі функції і їх ефективні множини мають спільну точку, в якій всі ці функції, за винятком може однієї, неперервні, то

$$(f_1 + f_2 + \ldots + f_n)^* = f_1^* \oplus f_2^* \oplus \ldots \oplus f_n^*.$$

Більш того, в цьому випадку для всякого $x^* \in \text{dom } (f_1 + \cdots + f_n)^*$ знайдуться такі точки $x_i^* \in \text{dom } f_i^*, i = 1, \ldots, n$, що

$$x^* = x_1^* + \cdots + x_n^*,$$

$$(f_1 + f_2 + \ldots + f_n)^* (x^*) = f_1^* (x_1^*) + f_2^* (x_2^*) + \ldots + f_n^* (x_n^*) .$$

Теорема 2.4.2. *Нехай $f_1, \ldots, f_n$ – функції на X. Тоді*

$$(\text{conv}\,(f_1 \wedge f_2 \wedge \ldots \wedge f_n))^* = f_1^* \vee f_2^* \vee \ldots \vee f_n^*,$$

$$(f_1 \vee f_2 \vee \ldots \vee f_n)^* \leqslant \text{conv}\,(f_1^* \wedge f_2^* \wedge \ldots \wedge f_n^*) .$$

Якщо ж $f_1, \ldots, f_n$ – скінченні на всьому просторі X опуклі функції і всі вони, за винятком може однієї, неперервні, то

$$(f_1 \vee f_2 \vee \ldots \vee f_n)^* = \text{conv}\,(f_1^* \wedge f_2^* \wedge \ldots \wedge f_n^*) .$$

Більш того, в цьому випадку для всякого $x^* \in \mathrm{dom}\,(f_1 \vee \dots \vee f_n)^*$ *знайдуться такі вектори* $x_i^* \in \mathrm{dom}\, f_i^*, i = 1, \dots, n$*, і такі невід'ємні числа* $\alpha_i, i = 1, \dots, n$*, сума яких дорівнює одиниці, що*

$$x^* = \alpha_1 x_1^* + ... + \alpha_n x_n^*,$$

$$(f_1 \vee f_2 \vee \ldots \vee f_n)^* (x^*) = \alpha_1 f_1^* (x_1^*) + \alpha_2 f_2^* (x_2^*) + \ldots + \alpha_n f_n^* (x_n^*) \, .$$

Теорема 2.4.3. *Нехай* $\Lambda : X \to Y$ *– лінійний неперервний оператор. Якщо* g *– функція на* X*,* f *– функція на* Y*, то*

$$(\Lambda g)^* = g^* \Lambda^*, \quad (f\Lambda)^* \leqslant \Lambda^* f^*.$$

Якщо ж f *– неперервна в деякій точці множини* $\mathrm{Im}\,\Lambda$ *опукла функція, то*

$$(f\Lambda)^* = \Lambda^* f^*.$$

Більш того, в цьому випадку для всякого $x^* \in \mathrm{dom}\,(f\Lambda)^*$ *знайдеться такий вектор* $y^* \in Y^*$*, що*

$$x^* = \Lambda^* y^*, \quad (f\Lambda)^* (x^*) = f^* (y^*) \, .$$

Відзначимо, що в кожній із сформульованих теорем перше твердження носить безумовний характер, у той час як друге твердження справедливе лише при додаткових умовах. Причина цього криється в різній природі двоїстих операцій у кожній парі. Одна з них локальна (значення результуючої функції в кожній точці визначається лише значеннями вихідних функцій у відповідній точці) і перетворює замкнуті функції в замкнуті. Такі операції – це сума, верхня грань і прообраз при лінійному неперервному відображенні. Друга операція нелокальна (значення результуючої функції в кожній точці залежить від усієї сукупності значень вихідних функцій) і може не зберігати замкнутість.

Приведемо приклади, які показують, що умови, які фігурують у теоремах, істотні.

Приклад 2.4.1. Нехай f_1 і f_2 - функції на прямій, задані рівностями

$$f_1(x) = ||x| - 1| \, , \quad f_2(x) = |x| \, .$$

Тоді

$$f_1^* (y) = \begin{cases} |y| \, , & |y| \leqslant 1, \\ \infty, & |y| > 1, \end{cases}$$

$$f_2^* (y) = \begin{cases} 0, & |y| \leqslant 1, \\ \infty, & |y| > 1, \end{cases}$$

$$(f_1 + f_2) (x) = \begin{cases} 1, & |x| \leqslant 1, \\ 2 |x| - 1, & |x| > 1, \end{cases}$$

$$(f_1 + f_2)^* (y) = \begin{cases} |y| - 1, & |y| \leqslant 1, \\ \infty, & |y| > 1. \end{cases}$$

З іншого боку,

$$\left(f_1^* \oplus f_2^*\right)(y) = \inf_z \left\{f_1^*(y-z) : |z| \leqslant 1\right\} = \begin{cases} 0, & |y| \leqslant 1, \\ |y|-1, & 1 \leqslant |y| \leqslant 2, \\ \infty, & |y| > 2. \end{cases}$$

Отже перетворення Юнга–Фенхеля суми неопуклих функцій (функція f_1 неопукла) може не збігатися з інфімальною конволюцією спряжених функцій, навіть якщо вихідні функції всюди неперервні.

Приклад 2.4.2. Нехай

$$f_1(x) = \begin{cases} -\sqrt{2x^1x^2}, & x^1 \geqslant 0,\ x^2 \geqslant 0, \\ \infty, & x^1 < 0, \text{ або } x^2 < 0; \end{cases}$$

$$f_2(x) = \begin{cases} 0, & x^1 \geqslant 0, x^2 \geqslant 0, \\ \infty, & x^1 < 0, \text{ або } x^2 < 0; \end{cases}$$

де $x = \left(x^1,x^2\right) \in \mathbb{R}^2$. Обидві функції опуклі, а їхні ефективні множини перетинаються по півосі $x^1 = 0, x^2 \geqslant 0$, у кожній точці якої обидві функції мають розрив. Зокрема,

$$\left(f_1 + f_2\right)(x) = \begin{cases} 0, & x^1 = 0, x^2 \geqslant 0, \\ \infty, & x^1 \neq 0, \text{ або } x^2 < 0; \end{cases}$$

Маємо для $y = \left(y^1,y^2\right) \in \mathbb{R}^2$

$$f_1^*(y) = \begin{cases} 0, & y^1y^2 \geqslant 1,\ y^1 < 0,\ y^2 < 0, \\ \infty, & \text{в інших точках}; \end{cases}$$

$$f_2^*(y) = \begin{cases} 0, & y^1 \geqslant 0,\ y^2 \leqslant 0, \\ \infty, & y^1 < 0, \text{ або } y^2 > 0; \end{cases}$$

$$\left(f_1 + f_2\right)^*(y) = \begin{cases} 0, & y^2 \leqslant 0, \\ \infty, & y^2 > 0. \end{cases}$$

З іншого боку, функція $f_1^* \oplus f_2^*$ як інфімальна конволюція індикаторних функцій дорівнює індикаторній функції суми ефективних множин функцій f_1^* і f_2^*. Сума цих множин складається з усіх точок півплощини $y^2 \leqslant 0$, за винятком точок прямої $y^2 = 0$, тобто вона не збігається з ефективною множиною функції $\left(f_1 + f_2\right)^*$.

Аналогічні приклади можна навести для підтвердження істотності умов інших теорем.

Переходячи до доведень, відмітимо зразу, що перші дві теореми досить довести для випадку двох функцій.

Доведення. Доведення теореми 2.4.1. Перша рівність випливає з того, що

$$\left(f_1 \oplus f_2\right)^*(x^*) = \sup_x \Big\{\langle x^*,x\rangle - \inf_z \left(f_1(x-z) + f_2(z)\right)\Big\} =$$

$$= \sup_{z,y} \{\langle x^*,y\rangle - f_1(y) + \langle x^*,z\rangle - f_2(z)\} = f_1^*(x^*) + f_2^*(x^*).$$

Далі, за нерівністю Юнга – Фенхеля

$$f_1^*(x_1^*) + f_2^*(x_2^*) \geqslant \langle x_1^* + x_2^*,x\rangle - f_1(x) - f_2(x)$$

для будь-яких $x_1^* \in X^*$, $x_2^* \in X^*$, $x \in X$. Тому

$$f_1^*(x_1^*) + f_2^*(x_2^*) \geqslant (f_1 + f_2)^*(x_1^* + x_2^*).$$

Ця нерівність вірна, зокрема, для всіх x_1^* і x_2^* таких, що $x_1^* + x_2^* = x^*$. Тому

$$f_1^* \oplus f_2^* \geqslant (f_1 + f_2)^*.$$

Перша частина теореми доведена. Перейдемо до доведення другої частини. З останньої нерівності, зокрема, випливає, що третя з формул, що доводяться, вірна, якщо

$$\mathrm{dom}\,(f_1 + f_2)^* = \emptyset.$$

Припустимо тепер, що $(f_1 + f_2)^*(x^*) = \alpha_0 < \infty$ і функція f_1 неперервна в деякій точці з $\mathrm{dom}\ f_2$. Тоді $\mathrm{dom}\,(f_1 + f_2) \neq \emptyset$ і, виходить, що $(f_1 + f_2)^*$ всюди більше $-\infty$; зокрема, $-\infty < \alpha_0 < \infty$. Розглянемо множину

$$A = \{(\alpha,x) \in \mathbb{R} \times X \,|\alpha \leqslant \langle x^*,x\rangle - f_2(x) - \alpha_0\}.$$

Ця множина опукла. Покажемо, що $A \cap \mathrm{int}\,(\mathrm{epi}\ f_1) = \emptyset$. Дійсно, якщо $(\alpha,x) \in A \cap \mathrm{int}\,(\mathrm{epi}\ f_1)$, то

$$f_1(x) < \alpha \leqslant \langle x^*,x\rangle - f_2(x) - \alpha_0,$$

тобто

$$\alpha_0 < \langle x^*,x\rangle - f_1(x) - f_2(x) \leqslant (f_1 + f_2)^* = \alpha_0.$$

За теоремою про розділення існує ненульовий лінійний неперервний функціонал, що визначається елементом $(\beta,y^*) \in \mathbb{R} \times X^*$, який розділяє множини A та $\mathrm{int}\,(\mathrm{epi}\ f_1)$, тобто такий, що

$$\sup\{\beta\alpha + \langle y^*,x\rangle :\ (\alpha,x) \in \mathrm{epi}\ f_1\} \leqslant \inf\{\beta\alpha + \langle y^*,x\rangle : (\alpha,x) \in A\}.$$

Ясно, що $\beta \leqslant 0$. Якщо $\beta = 0$, то з останньої нерівності випливає, що y^* розділяє множини $\mathrm{dom}\ f_1$ і $\mathrm{dom}\ f_2$ ($y^* \neq 0$, коли $\beta = 0$). Останнє в силу теореми про розділення неможливо, оскільки за умовою множини $\mathrm{int}\left(\mathrm{dom}\, f_1\right)$ і $\mathrm{dom}\ f_2$ перетинаються.

Отже, $\beta < 0$. Розділивши обидві частини останньої нерівності на $|\beta|$ і поклавши $x_1^* = |\beta|^{-1} y^*$, одержимо

$$f_1^*(x_1^*) = \sup\{\langle x_1^*,x\rangle - f_1(x) : x \in X\} =$$

$$= \sup\{\langle x_1^*,x\rangle - \alpha \colon (\alpha,x) \in \mathrm{epi}\ f_1\} \leqslant \inf\{\langle x_1^*,x\rangle - \alpha \colon (\alpha,x) \in A\} =$$

$$= \inf\{\langle x_1^* - x^*,x\rangle + f_2(x) : x \in \mathrm{dom}\ f_2\} + \alpha_0 = -f_2^*(x^* - x_1^*) + \alpha_0.$$

З цього співвідношення випливає нерівність

$$(f_1^* \oplus f_2^*)(x^*) = f_1^*(x_1^*) + f_2^*(x^* - x_1^*) \leqslant \alpha_0 = (f_1 + f_2)^*(x^*).$$

Це доводить теорему. □

Доведення. Доведення теореми 2.4.3. Як і в попередньому випадку, перше співвідношення випливає з означень, друге випливає з нерівності Юнга–Фенхеля. Нам залишається довести нерівність

$$(\Lambda^* f^*)(x^*) \leqslant (f\Lambda)^*(x)^*$$

у припущенні, що функція f неперервна в деякій точці, що належить множині $\mathrm{Im}\ \Lambda$, і що $x^* \in \mathrm{dom}\,(f\Lambda)^*$. Покладемо $\alpha_0 = (f\Lambda)^*(x^*)$. Оскільки $(\mathrm{dom}\ f) \cap (\mathrm{Im}\,\Lambda) \neq \emptyset$, функція $f\Lambda$ приймає скінченні значення, то виходить, що $(f\Lambda)^*$ всюди більше $-\infty$. Зокрема, α_0 -- скінченне число. Розглянемо в $\mathbb{R} \times X$ лінійний многовид

$$M = \{(\alpha,y) \in \mathbb{R} \times Y\, |\exists x \in X\, :\, \alpha = \langle x^*,x\rangle - \alpha_0, y = \Lambda x\}\,.$$

Він не перетинається з $\mathrm{int\ epi}\ f$, тому що в іншому випадку для деякого $x \in X$ було б

$$f\,(\Lambda x) < \langle x^*,x\rangle - \alpha_0,$$

тобто

$$\alpha_0 < \langle x^*,x\rangle - f\,(\Lambda x) \leqslant (f\Lambda)^*(x)^* = \alpha_0.$$

За теоремою про розділення множини M і $\mathrm{epi}\ f$ можна розділити ненульовим лінійним функціоналом, що визначається елементом $(\beta,y^*) \in \mathbb{R} \times Y^*$, тобто

$$\sup\left\{\beta\alpha + \langle y^*,y\rangle\, :\, (\alpha,y) \in \mathrm{epi}\, f\right\} \leqslant \inf\left\{\beta\alpha + \langle y^*,y\rangle\, :\, (\alpha,y) \in M\right\}.$$

Переконуємося, що $\beta \leqslant 0$. Якщо $\beta = 0$, то y^* не дорівнює нулю та розділяє множини $\mathrm{dom}\ f$ і $\mathrm{Im}\,\Lambda$ всупереч припущенню про те, що $\mathrm{int}\ (\mathrm{dom}\ f) \cap (\mathrm{Im}\ \Lambda) \neq \emptyset$. Отже, $\beta < 0$. Розділивши обидві частини останньої нерівності на $|\beta|$ і поклавши $y_0^* = |\beta|^{-1} y^*$, одержимо

$$\begin{aligned} f^*(y_0^*) &\leqslant \inf\left\{\langle y_0^*,y\rangle - \alpha\colon (\alpha,y) \in M\right\} = \\ &= \inf\left\{\langle y_0^*,\Lambda x\rangle - \langle x^*,x\rangle + \alpha_0\colon x \in X\right\}. \end{aligned}$$

Оскільки $f^*(y^*) > -\infty$ для всіх y^*, то $x^* = \Lambda^* y^*$. В іншому випадку

$$\inf_x\left(\langle y_0^*,\Lambda x\rangle - \langle x^*,x\rangle\right) = \inf_x \langle \Lambda^* y_0^* - x^*,x\rangle = -\infty.$$

Таким чином, $x^* \in \mathrm{Im}\,\Lambda^*$, $x^* = \Lambda^* y_0^*$ і

$$(\Lambda^* f^*)(x^*) \leqslant f^*(y_0^*) \leqslant \alpha_0 = (f\Lambda)^*(x^*)\,.$$

□

2.5. ДОДАТНЬО ОДНОРІДНІ ОПУКЛІ ФУНКЦІЇ

Означення 2.5.1. Функція f називається *додатньо однорідною*, якщо $f(\lambda x) = \lambda f(x)$ для $\lambda > 0$.

Якщо, крім того, функція f опукла, то

$$f(x_1 + x_2) = f\left(2\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right)\right) = 2f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) \le f(x_1) + f(x_2).$$

Так само можна показати, що для додатньо однорідної опуклої функції f справджується нерівність

$$f(x_1 + x_2 + \cdots + x_n) \le f(x_1) + f(x_2) + \cdots + f(x_n).$$

Якщо в попередній нерівності покласти $x_1 = 0$, $x_2 = x$, то отримаємо $f(x) \le f(0) + f(x)$, звідки маємо, що $f(0) \ge 0$.
Нехай функція f додатньо однорідна, опукла і замкнута. В силу замкнутості

$$\lim_{\lambda \downarrow 0} f(\lambda x) = \lim_{\lambda \downarrow 0} \lambda f(x) = 0 \ge f(0).$$

Отже $f(0) = 0$ для замкнутої додатньо однорідної опуклої функції.

Знайдемо спряжену функцію для такої функції

$$f^*(x^*) = \sup_x \{\langle x^*, x\rangle - f(x)\} \ge -f(0) = 0.$$

Нехай існує таке x_1, що

$$\langle x^*, x_1\rangle - f(x_1) > 0.$$

Тоді

$$f^*(x^*) \ge \sup_{\lambda > 0} \{\langle x^*, \lambda x_1\rangle - f(\lambda x_1)\} =$$

$$\sup_{\lambda > 0} \lambda \{\langle x^*, x_1\rangle - f(x_1)\} = +\infty.$$

Отже функція $f^*(x^*)$ приймає лише два значення 0 та $+\infty$. Тому

$$f^*(x^*) = \delta(x^* | \operatorname{dom} f^*) = \begin{cases} 0, & x^* \in \operatorname{dom} f^*; \\ +\infty, & x^* \notin \operatorname{dom} f^*. \end{cases}$$

Теорема 2.5.1. *Нехай f – додатньо однорідна опукла і замкнута функція. Тоді*

$$f^*(x^*) = \delta(x^* | \operatorname{dom} f^*).$$

Множина $\operatorname{dom} f^*$ *замкнута.*

Теорема 2.5.2. *Нехай f – додатньо однорідна опукла і замкнута функція. Тоді*

$$f(x) = \sup_{x^*} \{\langle x^*, x\rangle : x^* \in \operatorname{dom} f^*\}.$$

Доведення. За теоремою Фенхеля-Моро $f(x) = f^{**}(x)$. За попередньою теоремою $f^*(x^*) = \delta(x^* | \operatorname{dom} f^*)$. Тому

$$f^{**}(x) = \sup_{x^*} \{\langle x^*, x\rangle - \delta(x^* | \operatorname{dom} f^*)\} =$$

$$= \sup_{x^*} \{ \langle x^*, x \rangle : x^* \in \operatorname{dom} f^* \}.$$

$\square$

Теорема 2.5.3. *Спряжена функція до опорної функції замкнутої опуклої множини є індикаторною функцією цієї множини.*

Доведення. Нехай C^* – замкнута опукла множина в X^* і

$$f(x) = \sup_{x^*} \{ \langle x^*, x \rangle : x^* \in C^* \}$$

опорна функція множини C^*. Функція $f(x)$ додатньо однорідна, опукла і замкнута.

Розглянемо індикаторну функцію $\delta(x^* | C^*)$ множини C^*. Спряженою до цієї функції буде функція

$$\sup_{x^*} \{ \langle x^*, x \rangle - \delta(x^* | C^*) \} =$$

$$= \sup_{x^*} \{ \langle x^*, x \rangle : x^* \in C^* \} = f(x).$$

Отже функція $f(x)$ є спряженою до індикаторної функції $\delta(x^* | C^*)$ множини C^*. З теореми Фенхеля-Моро випливає, що

$$f^*(x^*) = \delta(x^* | C^*), \quad \operatorname{dom} f^* = C^*.$$

$\square$

2.6. УЗАГАЛЬНЕННЯ ОПУКЛИХ ФУНКЦІЙ

2.6.1. Квазіопуклі функції

Означення 2.6.1. Нехай X – опукла підмножина $\mathbb{R}^n$. Функція $f\colon X \to \mathbb{R}$ називається *квазіопуклою* або *унімодальною*, якщо всі її множини Лебега

$$X_\beta = \{ x \in X \mid f(x) \leqslant \beta \}$$

опуклі. Функція f називається *квазіугнутою*, якщо функція $g = -f$ квазіопукла. Функція, яка одночасно є квазіопуклою та квазіугнутою називається *квазілінійною*.

Опуклі функції мають опуклі множини Лебега. Отже опуклі функції є квазіопуклими. Обернене твердження не вірне.

Приклад 2.6.1. Функція $f(x) = \ln x$ на $\operatorname{int} \mathbb{R}_+$ квазіопукла (та квазіугнута, отже квазілінійна).

Приклад 2.6.2. Функція $f(x) = \mathrm{ceil}(x) = \min\{z \in \mathbb{Z} \,|\, z \geqslant x\}$ квазіопукла (та квазіугнута).

Ці приклади показують, що квазіопукла функція може бути угнутою, навіть розривною. Розглянемо приклади на $\mathbb{R}^n$.

Приклад 2.6.3. *Довжина вектора.* Визначимо довжину вектора $x \in \mathbb{R}^n$ як найбільший індекс ненульової компоненти, тобто

$$f(x) = \max\{k \leqslant n \,|\, x_i = 0, i = k+1, \dots, n\}\,.$$

Ця функція квазіопукла на $\mathbb{R}^n$. Її множини Лебега є підпросторами $\mathbb{R}^n$.

Приклад 2.6.4. Розглянемо функцію $f(x) = x_1 x_2$ на $\mathbb{R}^2_+$. Ця функція не є опуклою або угнутою, бо матриця

$$f''(x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

невизначена. Вона має одне додатнє і одне від'ємне власне число. Але ця функція квазіугнута, оскільки опуклі всі множини вигляду

$$\left\{x \in \mathbb{R}^2_+ \,|\, x_1 x_2 \geqslant \beta\right\}.$$

Приклад 2.6.5. *Дробово-лінійні функції.* Функція

$$f(x) = \frac{\langle a,x\rangle + b}{\langle c,x\rangle + d}$$

з областю визначення $\{x \,|\, \langle c,x\rangle + d > 0\}$ квазіопукла (та квазіугнута), оскільки опуклі її множини Лебега

$$X_\beta = \left\{x \,\middle|\, \langle c,x\rangle + d > 0, \frac{\langle a,x\rangle + b}{\langle c,x\rangle + d} \leqslant \beta\right\} =$$

$$= \left\{x \,\middle|\, \langle c,x\rangle + d > 0, \frac{\langle a - \beta c,x\rangle + b - \beta d}{\langle c,x\rangle + d} \leqslant 0\right\}.$$

Приклад 2.6.6. *Відношення відстаней.* Нехай a, $b \in \mathbb{R}^n$. Визначимо функцію

$$f(x) = \frac{\|x-a\|_2}{\|x-b\|_2},$$

тобто відношення відстаней від точки x до точок a та b. Функція $f(x)$ квазіугнута на $\{x : \|x-a\|_2 \leqslant \|x-b\|_2\}$. Щоб показати це, розглянемо множину Лебега X_β при $\beta \leqslant 1$, оскільки $f(x) \leqslant 1$ на $\{x : \|x-a\|_2 \leqslant \|x-b\|_2\}$. Множина Лебега X_β – це множина точок, що задовольняють умову

$$\|x-a\|_2 \leqslant \beta\|x-b\|_2.$$

Піднесемо обидві частини нерівності до квадрату. Отримуємо

$$(1-\beta^2)\langle x,x\rangle - 2\langle a - \beta^2 b,x\rangle + \beta^2\langle b,b\rangle \leqslant 0.$$

Ця нерівність описує опуклу множину (кулю), якщо $\beta \leqslant 1$.

Наведені вище приклади показують, що квазіопуклість є суттєвим узагальненням опуклості. Проте багато властивостей опуклих функцій зберігаються або мають аналоги для квазіопуклих функцій. Наприклад, аналог нерівності Ієнсена, що характеризує квазіопуклість.

Теорема 2.6.1. *Функція* $f\colon X \to \mathbb{R}$, *де* $X \subset \mathbb{R}^n$ *– опукла множина, квазіопукла тоді і тільки тоді, коли для всіх* x^1, $x^2 \in X$, $0 \leqslant \lambda \leqslant 1$ *виконується нерівність*

$$f(\lambda x^1 + (1-\lambda)x^2) \leqslant \max\{f(x^1), f(x^2)\}. \qquad (2.6.1)$$

Доведення. Нехай функція f квазіопукла, тобто множина X_β опукла для будь-якого β. Зафіксуємо дві довільні точки x^1, $x^2 \in X$ та розглянемо точку $x = \lambda x^1 + (1-\lambda)x^2$, $\lambda \in (0,1)$. Точки x^1, $x^2 \in X_\beta$ при $\beta = \max\{f(x^1), f(x^2)\}$. Оскільки множина X_β опукла, то $x \in X_\beta$, а, отже, $f(x) \leqslant \beta = \max\{f(x^1), f(x^2)\}$, тобто нерівність (2.6.1) виконується.

Нехай тепер виконується (2.6.1). Зафіксуємо довільні точки x^1, $x^2 \in X_\beta$. Тоді $\max\{f(x^1), f(x^2)\} \leqslant \beta$. Оскільки X – опукла, то для будь-якого $\lambda \in (0,1)$ точка $x = \lambda x^1 + (1-\lambda)x^2 \in X$. З нерівності (2.6.1) випливає, що $f(x) \leqslant \max\{f(x^1), f(x^2)\} \leqslant \beta$, тобто $x \in X_\beta$. Оттже, X_β – опукла множина і f – квазіопукла функція.

□

Приклад 2.6.7. *Ранг невід'ємно визначеної матриці.* Функція $f(X) = \mathrm{Rank}(X)$ є квазіугнутою на множині всіх невід'ємно визначених матриць розмірності $n \times n$. Це випливає з нерівності Ієнсена для квазіугнутих функцій (2.6.1)

$$\mathrm{Rank}(X+Y) \geqslant \max\{\mathrm{Rank}(X), \mathrm{Rank}(Y)\},$$

де X, Y – невід'ємно визначені матриці розмірності $n \times n$.

Дамо просту характеризацію квазіопуклих функцій на $\mathbb{R}$. Ми розглянемо неперервні функції, оскільки формулювання теореми в загальному випадку надто складне.

Теорема 2.6.2. *Неперервна функція* $f\colon X \to \mathbb{R}$, *де* X *- опукла множина в* $\mathbb{R}$, *квазіопукла тоді і тільки тоді, коли виконується одна з таких умов: 1)* f *– неспадна; 2)* f *– незростаюча; 3) існує така точка* $c \in X$, *що для всіх* $t \in X$, $t \leqslant c$, *функція* f *незростаюча, і для всіх* $t \in X$, $t \geqslant c$, *функція* f *неспадна.*

Диференційовні квазіопуклі функції

Теорема 2.6.3. *Нехай* $f\colon X \to \mathbb{R}$ *– диференційовна функція на* X, *де* $X \subset \mathbb{R}^n$ *– відкрита опукла множина. Тоді* f *квазіопукла на* X *тоді і* *тільки тоді, коли справджується співвідношення*

$$f(y) \leqslant f(x) \Rightarrow \langle f'(x), y-x\rangle \leqslant 0 \quad \textit{для всіх} \quad x,y \in X. \tag{2.6.2}$$

Доведення. Покажемо, що якщо функція квазіопукла, то виконується (2.6.2). Розглянемо будь-які точки x, $y \in X$ такі, що $f(y) \leqslant f(x)$. З диференційовності $f(x)$ у точці x при $\lambda \in (0,1)$ маємо

$$f(\lambda y + (1-\lambda)x) - f(x) = \lambda\langle f'(x), y-x\rangle + \lambda\|y-x\|\alpha(x;\lambda(y-x)),$$

де $\alpha(x;\lambda(y-x)) \to 0$ при $\lambda \to 0$. Оскільки функція f квазіопукла, то $f(\lambda y + (1-\lambda)x) \leqslant f(x)$. Тоді

$$\lambda\langle f'(x), y-x\rangle + \lambda\|y-x\|\alpha(x;\lambda(y-x)) \leqslant 0.$$

Поділивши цю нерівність на λ та спрямувавши λ до нуля, отримаємо, що $\langle f'(x), y-x\rangle \leqslant 0$.

Нехай виконується твердження (2.6.2). Розглянемо будь-які точки x, $y \in X$, для яких $f(y) \leqslant f(x)$. Потрібно довести, що $f(\lambda y+(1-\lambda)x) \leqslant f(x)$ для будь-яких $\lambda \in (0,1)$. Для цього достатньо показати, що множина

$$L = \{z \,|\, z = \lambda y + (1-\lambda)x, \lambda \in (0,1), f(z) > f(x)\}$$

порожня. Нехай це не так, тобто припустимо, що $x' \in L$. Тоді $x' = \lambda y + (1-\lambda)x$ для деякого $\lambda \in (0,1)$ та $f(x') > f(x)$. Оскільки функція f диференційовна, то вона неперервна, а отже, знайдеться таке $\delta \in (0,1)$, що

$$f(\mu x' + (1-\mu)x) > f(x) \quad \text{для будь-якого} \quad \mu \in [\delta,1],$$

а $f(x') > f(\delta x' + (1-\delta)x)$. З цієї нерівності та теореми про середнє значення отримуємо, що

$$0 < f(x') - f(\delta x' + (1-\delta)x) = (1-\delta)\langle f'(\hat{x}), x'-x\rangle,$$

де $\hat{x} = \hat{\mu}x' + (1-\hat{\mu})x$ для деякого $\hat{\mu} \in (\delta,1)$. Ясно, що $f(\hat{x}) > f(x)$. Поділивши попередню нерівність на $1-\delta > 0$, отримуємо $\langle f'(\hat{x}), x'-x\rangle > 0$. Звідси випливає

$$\langle f'(\hat{x}), y-x\rangle > 0.$$

З іншого боку, $f(\hat{x}) > f(x) \geqslant f(y)$, а точка $\hat{x}$ є опуклою комбінацією точок x та y, $\hat{x} = \hat{\lambda}y+(1-\hat{\lambda})x$, $\hat{\lambda} \in (0,1)$. За припущенням теореми маємо $\langle f'(\hat{x}), y-\hat{x}\rangle \leqslant 0$. Тому повинно виконуватись співвідношення

$$0 \geqslant \langle f'(\hat{x}), y-\hat{x}\rangle = (1-\hat{\lambda})\langle f'(\hat{x}), y-x\rangle.$$

Ця нерівність несумісна з нерівністю $\langle f'(\hat{x}), y-x\rangle > 0$. Отже, $L = \varnothing$. □

Умова (2.6.2) має просту геометричну інтерпретацію коли $f'(x) \neq 0$. Вона стверджує, що $f'(x)$ визначає опорну гіперплощину до множини Лебега $\{y \,|\, f(y) \leqslant f(x)\}$.

Наслідок 2.6.1. *Нехай f – двічі диференційовна функція. Якщо f квазіопукла на X, то виконується умова:*

$$\langle y, f'(x)\rangle = 0 \Rightarrow \langle f''(x)y, y\rangle \geqslant 0, \textit{для всіх}\ x \in X,\ y \in \mathbb{R}^n. \tag{2.6.3}$$

Для квазіопуклих функцій на $\mathbb{R}$ це твердження перетворюється на просту умову:

$$f'(x) = 0 \Rightarrow f''(x) \geqslant 0,$$

тобто в точці з нульовим нахилом друга похідна невід'ємна. Для квазіопуклих функцій на $\mathbb{R}^n$ інтерпретація умови (2.6.3) дещо складніша. Як і в одновимірному випадку, якщо $f'(x) = 0$, то повинно виконуватись $f''(x) \geqslant 0$. Якщо $f'(x) \neq 0$, то умова (2.6.3) означає, що $f''(x)$ невід'ємно визначена на $(n-1)$-вимірному підпросторі $f'(x)^{\perp}$. Звідси випливає, що матриця $f''(x)$ повинна мати принаймі одне від'ємне власне число. Навпаки, якщо f задовольняє умову

$$\langle y, f'(x)\rangle = 0 \Rightarrow \langle f''(x)y, y\rangle > 0$$

для всіх $x \in X$ та $y \in \mathbb{R}^n$, $y \neq 0$, то f квазіопукла. Ця умова додатньої визначеності $f''(x)$ в кожній точці x, де $f'(x) = 0$, а в інших точках – це умова додатньої визначеності $f''(x)$ на підпросторі $f'(x)^{\perp}$.

Операції, що зберігають квазіопуклість

1. Максимум зважених квазіопуклих функцій з невід'ємними вагами, тобто

$$f(x) = \max\{w_1 f_1(x), \ldots, w_n f_n(x)\}$$

з $w_i > 0$ та f_i – квазіопуклими, є квазіопуклим.

2. Супремум сім'ї функцій

$$f(x) = \sup_{y \in C} w(y) g(x,y),$$

де $w(y) \geqslant 0$, $g(x,y)$ – квазіопукла за x для будь-якого $y \in C$ функція. Це твердження можна легко перевірити: $f(x) \leqslant \beta$ тоді і тільки тоді, коли

$$w(y) g(x,y) \leqslant \beta \quad \text{для всіх} \quad y \in C,$$

тобто множина Лебега X_β функції $f(x)$ є перетином множин Лебега $Z(y)_\beta$ функцій $w(y)g(x,y)$ за змінною x.

Приклад 2.6.8. *Узагальнені власні числа.* Найбільше узагальнене власне число пари матриць (X,Y), де $Y > 0$, визначається наступним чином

$$\lambda_{\max}(X,Y) = \sup_{u \neq 0} \frac{\langle Xu, u\rangle}{\langle Yu, u\rangle} = \sup\{\lambda \mid \det(\lambda Y - X) = 0\}.$$

Ця функція квазіопукла на множині $\{(X,Y)\}$, де X – симетрична $n \times n$ матриця, Y – симетрична додатньо визначена $n \times n$ матриця. Щоб довести це, зауважимо, що для кожного $u \neq 0$, функція $\langle Xu, u\rangle / \langle Yu, u\rangle$ дробово-лінійна, а отже, квазіопукла. Таким чином $\lambda_{\max}$ – супремум сім'ї квазіопуклих функцій.

3. Якщо h: $X \to \mathbb{R}$, де $X \subset \mathbb{R}^n$, квазіопукла функція та g: $Y \to \mathbb{R}$, де $Y \subset \mathbb{R}$, неспадна функція, то $f(x) = g(h(x))$ – квазіопукла функція.

4. Квазіопукла функція від афінної або дробово-лінійної функції є квазіопуклою.

5. Якщо g – квазіопукла функція, то $f(x) = g(Ax + b)$ квазіопукла функція, та $f(x) = g((Ax+b)/(\langle c,x\rangle+d))$ квазіопукла функція на множині

$$\left\{ x \,\middle|\, \frac{Ax + b}{\langle c,x\rangle + d} \in Y, \langle c,x\rangle + d > 0 \right\}.$$

Теорема 2.6.4. *Якщо $g(x,y)$ квазіопукла за (x,y) функція та C – опукла множина, то квазіопукла функція*

$$f(x) = \inf_{y \in C} g(x,y).$$

Доведення. За визначенням функції f, $f(x) \leqslant \beta$ тоді і тільки тоді, коли для будь-якого $\epsilon > 0$ існує $y \in C$ таке, що $g(x,y) \leqslant \beta + \epsilon$. Нехай X_β – множина Лебега функції f. Тоді для будь-якого $\epsilon > 0$ існують y_1, $y_2 \in C$ такі, що

$$g(x_1,y_1) \leqslant \beta + \epsilon, \quad g(x_2,y_2) \leqslant \beta + \epsilon.$$

Оскільки g – квазіопукла за (x,y) функція, то

$$g(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \leqslant \beta + \epsilon,$$

для $0 \leqslant \theta \leqslant 1$. Отже, $f(\theta x_1 + (1-\theta)x_2) \leqslant \beta$. □

Представлення у вигляді сім'ї опуклих функцій.

Ми хочемо знайти сім'ю опуклих функцій ϕ_t: $\mathbb{R}^n \to \mathbb{R}$ (індекс t пробігає дійсні значення) таких, що

$$f(x) \leqslant t \Leftrightarrow \phi_t(x) \leqslant 0.$$

Тобто, якщо X_β, $\beta \geqslant 0$ – множини Лебега функції f, а $Y_{t,\beta}$ – множини Лебега функцій ϕ_t, то $X_t = Y_{t,0}$. Очевидно, що функції ϕ_t повинні задовольняти нерівність $\phi_t(x) \leqslant 0 \Rightarrow \phi_s(x) \leqslant 0$ для $s \geqslant t$ та всіх $x \in \mathbb{R}^n$. Ця нерівність виконується, якщо для будь-якого x функція $\phi_t(x)$ незростаюча за t.

Щоб показати, що таке представлення існує, покладемо

$$\phi_t(x) = \begin{cases} 0, & f(x) \leqslant t; \\ \infty, & \text{в іншому випадку.} \end{cases}$$

Таке представлення не єдине. Наприклад, якщо множини Лебега замкнуті, то ми можемо покласти

$$\phi_t(x) = \operatorname{dist}(x, \{z \mid f(z) \leqslant t\}).$$

Звичайно нас цікавлять сім'ї ϕ_t з "хорошими" властивостями, наприклад, диференційовністю.

Приклад 2.6.9. Нехай p – опукла функція, а q – угнута функція, причому такі, що $p(x) \geqslant 0$ та $q(x) > 0$ на опуклій множині C. Тоді функція $f(x) = p(x)/q(x)$ квазіопукла на C. Для такої функції ми маємо

$$f(x) \leqslant t \Leftrightarrow p(x) - tq(x) \leqslant 0,$$

отже як ϕ_t ми можемо обрати функцію $\phi_t(x) = p(x) - tq(x)$ для $t \geqslant 0$. Для кожного t функція $\phi_t(x)$ опукла, а для кожного x функція $\phi_t(x)$ спадна за t.

Задача максимізації для квазіопуклих функцій

Теорема 2.6.5. *Нехай X – компактна многогранна множина в $\mathbb{R}^n$, функція $f\colon \mathbb{R}^n \to \mathbb{R}$ – квазіопукла та неперервна на X. Розглянемо задачу*

$$f(x) \to \max\,, \quad x \in X.$$

Серед розв'язків цієї задачі обов'язково існує крайня точка $\bar{x}$.

Доведення. Оскільки функція f неперервна, то вона досягає максимума на X в деякій точці $x' \in X$. Якщо існує крайня точка, в якій значення цільової функції дорівнює $f(x')$, то твердження справедливе. Припустимо, що це не так, тобто $f(x') > f(x^j)$, де x^j, $j = 1,\ldots,k$ – крайні точки множини X. За теоремою 2.2.9 (теорема Мінковського про опуклий компакт) точка x' може бути представлена у вигляді

$$x' = \sum_{j=1}^{k} \lambda_j x^j, \quad \sum_{j=1}^{k} \lambda_j = 1, \quad \lambda_j \geqslant 0, j = 1,\ldots,k,$$

де x^j, $j = 1,\ldots,k$ – крайні точки множини X. Оскільки $f(x') > f(x^j)$ для всіх j, то

$$f(x') > \max_{1 \leqslant j \leqslant k} f(x^j) = \beta.$$

Розглянемо множину $X_\beta = \{x \,|\, f(x) \leqslant \beta\}$. Зауважимо, що $x^j \in X_\beta$ при $j = 1,\ldots,k$ та X_β – опукла множина Лебега функції $f(x)$. Отже, $x' = \sum_{j=1}^{k} \lambda_j x^j \in X_\beta$. Звідси $f(x') \leqslant \beta$. Ми прийшли до суперечності. Це показує що $f(x') = f(x^j)$ для деякої крайньої точки x^j. $\square$

Строго квазіопуклі функції

Означення 2.6.2. Нехай X – опукла множина в $\mathbb{R}^n$. Функція $f\colon X \to \mathbb{R}$ називається *строго квазіопуклою,* якщо для будь-яких x_1, $x_2 \in X$, таких, що $f(x_1) \neq f(x_2)$, для всіх $\lambda \in (0,1)$ виконується нерівність

$$f(\lambda x_1 + (1 - \lambda)x_2) < \max\{f(x_1), f(x_2)\}.$$

Функція f називається *строго квазіугнутою*, якщо функція $g = -f$ строго квазіопукла.

З означення випливає, що будь-яка опукла функція є також строго квазіопуклою. За означенням 3.1.1 строго опукла функція є опуклою. Але строго квазіопукла функція не обов'язково квазіопукла. Наведемо приклад, що запропонований Карамардіаном:

$$f(x) = \begin{cases} 1 & \text{при} \quad x = 0, \\ 0 & \text{при} \quad x \neq 0. \end{cases}$$

За означенням 2.6.2 функція $f(x)$ строго квазіопукла. Але вона не є квазіопуклою, оскільки при $x_1 = 1$, $x_2 = -1$ маємо $f(x_1) = f(x_2) = 0$, а $f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) = f(0) = 1 > f(x_2)$.

Якщо ж функція f напівнеперервна знизу, то з її строгої квазіопуклості випливає звичайна квазіопуклість. Наступна теорема показує, що будь-який локальний мінімум строго квазіопуклої функції на опуклій множині є також її глобальним мінімумом. Квазіопуклі функції такої властивості не мають.

Теорема 2.6.6. *Нехай $f\colon \mathbb{R}^n \to \mathbb{R}$ – строго квазіопукла функція. Розглянемо задачу мінімізації $f(x)$ при умові, що $x \in X$, де X – опукла множина в $\mathbb{R}^n$. Нехай $\hat{x}$ – точка локального мінімуму задачі, тоді точка $\hat{x}$ є точкою глобального мінімуму задачі.*

Доведення. Припустимо що твердження теореми не вірне. Нехай існує точка $\bar{x} \in X$, для якої $f(\bar{x}) < f(\hat{x})$. З опуклості X випливає, що точка $\lambda\bar{x} + (1-\lambda)\hat{x} \in X$ для будь-якого $\lambda \in (0,1)$. Оскільки $\hat{x}$ – точка локального мінімуму, то $f(\hat{x}) \leqslant f\left(\lambda\bar{x} + (1-\lambda)\hat{x}\right)$ при всіх $\lambda \in (0,\delta)$ для деякого $\delta \in (0,1)$. Внаслідок строгої квазіопуклості функції f та нерівності $f(\bar{x}) < f(\hat{x})$ отримуємо, що $f(\lambda\bar{x} + (1-\lambda)\hat{x}) < f(\hat{x})$ при всіх $\lambda \in (0,1)$. Отримана суперечність доводить теорему. □

Лема 2.6.1. *Нехай X – опукла множина в $\mathbb{R}^n$, $f\colon X \to \mathbb{R}$ – строго квазіопукла напівнеперервна знизу функція. Тоді функція f квазіопукла.*

Доведення. Нехай x_1, $x_2 \in X$. Якщо $f(x_1) \neq f(x_2)$, то за означенням строгої квазіопуклості для кожного $\lambda \in (0,1)$ маємо $f(\lambda x_1 + (1-\lambda)x_2) < \max\{f(x_1),f(x_2)\}$. Нехай тепер $f(x_1) = f(x_2)$. Щоб впевнитися, що функція квазіопукла, треба показати, що $f(\lambda x_1 + (1-\lambda)x_2) < f(x_1)$ для всіх $\lambda \in (0,1)$. Припустимо протилежне, тобто нехай $f(\mu x_1 + (1-\mu)x_2) > f(x_1)$ при деякому $\mu \in (0,1)$. Розглянемо точку $x = \mu x_1 + (1-\mu)x_2$. Оскільки функція f напівнеперервна знизу, то існує таке $\lambda \in (0,1)$, що

$$f(x) > f(\lambda x_1 + (1-\lambda)x) > f(x_1) = f(x_2). \tag{2.6.4}$$

Зауважимо, що точка x може бути зображена у вигляді опуклої комбінації точок $\lambda x_1 + (1-\lambda)x$ і x_2. Тоді, оскільки функція f строго квазіопукла і $f(\lambda x_1 + (1-\lambda)x) > f(x_2)$, маємо $f(x) < f(\lambda x_1 + (1-\lambda)x) > f(x_2)$. Це суперечить (2.6.4). □

Сильно квазіопуклі функції

Означення 2.6.3. Нехай $f\colon X \to \mathbb{R}$, де X – опукла множина в $\mathbb{R}^n$. Функція f називається *сильно квазіопуклою,* якщо для будь-яких x_1, $x_2 \in X$, $x_1 \neq x_2$, для всіх $\lambda \in (0,1)$ виконується нерівність

$$f(\lambda x_1 + (1-\lambda)x_2) < \max\{f(x_1), f(x_2)\}.$$

Функція f називається *сильно квазіугнутою,* якщо функція $g = -f$ сильно квазіопукла.

З означень випливають такі твердження:

1. Строго опукла функція є сильно квазіопуклою.
2. Сильно квазіопукла функція є строго квазіопуклою.
3. Сильно квазіопукла функція є квазіопуклою, навіть якщо вона не є напівнеперервною знизу.

Теорема 2.6.7. *Нехай $f\colon \mathbb{R}^n \to \mathbb{R}$ – сильно квазіопукла функція. Розглянемо задачу мінімізації $f(x)$ при умові, що $x \in X$, де X – опукла множина в $\mathbb{R}^n$. Якщо $\hat{x}$ – точка локального мінімуму $f(x)$ на X, то вона є єдиним глобальним оптимальним розв'язком цієї задачі.*

Доведення. Оскільки $\hat{x}$ – локальний оптимальний розв'язок задачі, то існує такий ε–окіл $N_\varepsilon(\hat{x})$ точки $\hat{x}$, що $f(\hat{x}) \leqslant f(x)$ для всіх $x \in X \cap N_\varepsilon(\hat{x})$. Припустимо, що твердження теореми не вірне, тобто існує така точка $\bar{x} \in X$, що $\bar{x} \neq \hat{x}$ та $f(\bar{x}) < f(\hat{x})$. З сильної квазіопуклості f випливає, що

$$f(\lambda\bar{x} + (1-\lambda)\hat{x}) < \max\{f(\bar{x}), f(\hat{x})\} = f(\hat{x})$$

для всіх $\lambda \in (0,1)$. Але якщо λ достатньо мале, то $\lambda\bar{x} + (1-\lambda)\hat{x} \in X \cap N_\varepsilon(\hat{x})$. Остання нерівність суперечить локальній оптимальності $\hat{x}$ □

2.6.2. Псевдоопуклі функції

Означення 2.6.4. Нехай X – відкрита множина в $\mathbb{R}^n$, $f\colon X \to \mathbb{R}$ – диференційовна функція. Функція f називається *псевдоопуклою*, якщо для будь-яких x_1, $x_2 \in X$ таких, що $\langle f'(x_1), x_2 - x_1\rangle \geqslant 0$ виконується нерівність $f(x_2) \geqslant f(x_1)$, або, що еквівалентно, якщо $f(x_2) < f(x_1)$, то $\langle f'(x_1), x_2 - x_1\rangle < 0$.

Функція f називається *псевдоугнутою,* якщо функція $g = -f$ псевдоопукла.

Означення 2.6.5. Функція f називається *строго псевдоопуклою*, якщо для будь-яких різних $x_1,\ x_2 \in X$ таких, що $\langle f'(x_1), x_2 - x_1\rangle \geqslant 0$, виконується нерівність $f(x_2) > f(x_1)$, або, що еквівалентно, якщо для будь-яких різних $x_1,\ x_2 \in X$ з нерівності $f(x_2) \leqslant f(x_1)$ випливає нерівність $\langle f'(x_1), x_2 - x_1\rangle < 0$.
Функція f називається *строго псевдоугнутою,* якщо функція $g = -f$ строго псевдоопукла.

Теорема 2.6.8. *Нехай X – відкрита опукла множина в $\mathbb{R}^n$. Нехай $f\colon X \to \mathbb{R}$ – диференційовна псевдоопукла функція. Тоді f строго квазіопукла та квазіопукла.*

Доведення. Покажемо спочатку, що f – строго квазіопукла функція. Припустимо, що це не так, тобто існують такі $x_1,\ x_2 \in X$, що $f(x_1) \neq f(x_2)$ та $f(x') \geqslant \max\{f(x_1), f(x_2)\}$, де $x' = \lambda x_1 + (1-\lambda)x_2$ для деякого $\lambda \in (0{,}1)$. Нехай $f(x_1) < f(x_2)$. Отже

$$f(x') \geqslant f(x_2) > f(x_1). \tag{2.6.5}$$

Із означення псевдоопуклості функції f випливає, що $\langle f'(x'),\, x_1 - x'\rangle < 0$. Оскільки $\langle f'(x'), x_1 - x'\rangle < 0$ та $x_1 - x' = -\frac{1-\lambda}{\lambda}(x_2 - x')$, то $\langle f'(x'), x_2 - x'\rangle > 0$. Знову використовуючи псевдоопуклість f, отримуємо, що $f(x_2) \geqslant f(x')$. Тоді з (2.6.5) випливає, що $f(x_2) = f(x')$. Оскільки $\langle f'(x'), x_2 - x'\rangle > 0$, то знайдеться така точка $\hat{x} = \mu x' + (1-\mu)x_2$, $\mu \in (0{,}1)$, що

$$f(\hat{x}) > f(x') = f(x_2).$$

Аналогічним чином, використовуючи псевдоопуклість f, легко впевнитися в тому, що $\langle f'(x), x_2 - \hat{x}\rangle < 0$ та $\langle f'(\hat{x}), x' - \hat{x}\rangle < 0$. Зауважимо, що $x_2 - \hat{x} = \frac{\mu}{1-\mu}(\hat{x} - x')$. Отже, дві останні нерівності несумісні. Отримана суперечність показує, що припущення було невірним, тобто функція f строго квазіопукла. За попередньою лемою вона є також квазіопуклою. □

Теорема 2.6.9. *Нехай X – відкрита опукла множина в $\mathbb{R}^n$. Нехай $f\colon X \to \mathbb{R}$ – диференційовна строго псевдоопукла функція. Тоді f сильно квазіопукла.*

Доведення. Припустимо, що твердження теореми невірне, тобто існують різні $x_1,\ x_2 \in X$ та $\lambda \in (0{,}1)$, такі, що $f(x) \geqslant \max\{f(x_1), f(x_2)\}$, де $x = \lambda x_1 + (1-\lambda)x_2$. Оскільки $f(x_1) \leqslant f(x)$, то зі строгої псевдоопуклості функції f випливає, що $\langle f'(x), x_1 - x\rangle < 0$. Звідси

$$\langle f'(x), x_1 - x_2\rangle < 0.$$

Крім того, оскільки $f(x_2) \leqslant f(x)$, то

$$\langle f'(x), x_2 - x_1\rangle < 0.$$

Дві останні нерівності суперечливі. Отже, f є сильно квазіопуклою. □

Приклад 2.6.10. Функція

$$f_1(x) = \begin{cases} 0, & \text{при} \quad x \in \mathbb{R} \backslash \mathbb{R}_+; \\ -x^n, & \text{при} \quad x \in \mathbb{R}_+, n \in \mathbb{N}, n \geq 2, \end{cases}$$

квазіопукла на множині $\mathbb{R}$, проте вона не строго квазіопукла, не псевдоопукла, не строго псевдоопукла, не опукла.

Приклад 2.6.11. Функція

$$f_2(x) = -x^n, x \in \mathbb{R}_+, n \in \mathbb{N}, n \geq 2,$$

квазіопукла і строго квазіопукла на множині $\mathbb{R}_+$, проте вона не псевдоопукла, не строго псевдоопукла, не опукла.

Приклад 2.6.12. Функція

$$f_3(x) = -(x_1 + 1)^2 + 1, \quad x = (x_1, x_2) \in \mathbb{R}^2_+,$$

псевдоопукла, строго квазіопукла, квазіопукла на множині $\mathbb{R}^2_+$, проте вона не строго псевдоопукла, не опукла.

Приклад 2.6.13. Функція

$$f_4(x) = ax + b, \quad x \in \mathbb{R}, \quad a < 0,$$

опукла, строго псевдоопукла, псевдоопукла, строго квазіопукла, квазіопукла на множині $\mathbb{R}$, проте вона не строго опукла.

Приклад 2.6.14. Функція

$$f_5(x) = x^n, \quad x \in \mathbb{R}_+, n \in \mathbb{N}, n \geq 2,$$

строго опукла, опукла, строго псевдоопукла, псевдоопукла, строго квазіопукла, квазіопукла на множині $\mathbb{R}_+$.

2.6.3. Логарифмічно опуклі функції

Означення 2.6.6. Функція f: $X \to \mathbb{R}$, $X \subset \mathbb{R}^n$ називається *логарифмічно опуклою*, якщо $f(x) > 0$ для всіх $x \in X$ та функція $\ln f(x)$ опукла. Функція f називається *логарифмічно угнутою*, якщо функція $g = -\ln f(x)$ опукла. Таким чином f логарифмічно угнута тоді і тільки тоді, коли $1/f$ логарифмічно опукла.

З того що функція $g = e^f$ опукла коли функція f опукла випливає, що логарифмічно опукла функція є опуклою. Аналогічно, угнута функція логарифмічно угнута. Крім того, логарифмічно опукла функція квазіопукла, а логарифмічно угнута функція – квазіугнута.

Декілька простих прикладів:

1. Аффіна функція $f(x) = \langle a,x \rangle + b$ логарифмічно угнута на $\{x \mid \langle a,x \rangle + b > 0\}$.

2. Степенева функція $f(x) = x^a$ на $\operatorname{int} \mathbb{R}_+$ логарифмічно опукла при $a \leqslant 0$ і логарифмічно угнута при $a \geqslant 0$.

3. Експоненційна функція $f(x) = e^{ax}$ логарифмічно опукла та логарифмічно угнута.

4. Функція нормального розподілу $F(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-u^2/2} du$ логарифмічно угнута.

5. Гамма функція $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ логарифмічно опукла при $x \geqslant 1$.

6. Визначник $\det X$ логарифмічно опуклий на множині всіх додатньо визначених матриць S^n_{++}.

Логарифмічно угнуті щільності розподілів. Багато ймовірностних розподілів мають логарифмічно угнуті щільності.

1. Щільність багатовимірного нормального розподілу

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})},$$

$\bar{x} \in \mathbb{R}^n$, Σ – додатньо визначена матриця розміру $n \times n$, $\Sigma \in S^n_{++}$.

2. Щільність показникового розподілу на $\mathbb{R}^n_+$

$$f(x) = \left(\prod_{i=1}^{n} \lambda_i \right) e^{-\langle \lambda, x \rangle}, \lambda \geqslant 0.$$

3. Щільність рівномірного розподілу на опуклій множині C:

$$f(x) = \begin{cases} \frac{1}{\alpha}, & x \in C; \\ 0, & x \notin C, \end{cases}$$

де α – міра Лебега множини C ($\ln f(x) = -\infty$,$x \notin C$).

4. Щільність розподілу Вішарта, який визначається наступним чином. Нехай $x_1, \ldots ,x_p \in \mathbb{R}^n$,$p > n$, – незалежні нормально розподілені випадкові вектори з нульовим середнім та коваріаційною матрицею Σ. Випадкова матриця $X = \sum_{i=1}^{p} x_i x_i^T$ має розподіл Вішарта зі щільністю

$$f(X) = a(\det X)^{(p-n-1)/2} e^{-\frac{1}{2} \operatorname{Tr} \Sigma^{-1} X},$$

де X – додатньо визначена матриця, a – додатня стала. Щільність Вішарта є логарифмічно угнутою, оскільки функція

$$\ln f(X) = \ln a + \frac{p-n-1}{2} \ln \det X - \frac{1}{2} \operatorname{Tr} \Sigma^{-1} X$$

угнута за X.

5. Щільність гамма розподілу:

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, \quad x \geqslant 0,$$

де $\lambda \geqslant 1$, $\alpha > 0$.

6. Щільність багатовимірного гіперболічного розподілу

$$f(x) = ce^{-\alpha\left(\delta+\langle\Sigma^{-1}(x-\bar{x}),(x-\bar{x})\rangle\right)^{1/2}+\langle\beta,x-\bar{x}\rangle},$$

де Σ – додатньо визначена симетрична матриця, $\beta \in \mathbb{R}^n$, α та c – додатні сталі.

7. Щільність розподілу Діріхле

$$f(x) = \frac{\mathit{\Gamma}(\lambda)}{\mathit{\Gamma}(\lambda_1)\cdot\ldots\cdot\mathit{\Gamma}(\lambda_{n+1})} x_1^{\lambda_1-1}\cdot\ldots\cdot x_n^{\lambda_n-1}\left(1-\sum_{i=1}^{n} x_i\right)^{\lambda_{n+1}-1}$$

на множині $\{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i \leqslant 1\}$. Тут $\lambda \geqslant 1$.

Властивості логарифмічно опуклих функцій

Теорема 2.6.10. *Нехай функція $f\colon X \to \mathbb{R}$ – двічі диференційовна на опуклій множині X. Функція f логарифмічно опукла тоді і тільки тоді, коли*

$$f(x)f''(x) \geqslant f'(x)(f'(x))^T \quad \textit{для всіх} \quad x \in X,$$

і логарифмічно угнута тоді і тільки тоді, коли

$$f(x)f''(x) \leqslant f'(x)(f'(x))^T \quad \textit{для всіх} \quad x \in X.$$

Очевидно, що добуток та заміна масштабу зберігають логарифмічну опуклість та логарифмічну угнутість. Прості приклади показують, що сума логарифмічно угнутих функцій може не бути логарифмічно угнутою функцією. Але сума зберігає логарифмічну опуклість. Нехай f, g – логарифмічно опуклі функції, тобто $F = \ln f$ та $G = \ln g$ – опуклі функції. За теоремою про складну функцію

$$\ln(\exp F + \exp G) = \ln(f+g)$$

опукла. Отже сума двох логарифмічно опуклих функцій є логарифмічно опуклою функцією.
Взагалі, якщо функція $f(x,y)$ – логарифмічно опукла за x для всіх $y \in C$, то функція

$$g(x) = \int_C f(x,y)\,dy$$

логарифмічно опукла.

Приклад 2.6.15. *Перетворення Лапласа невід'ємної функції, твірна функція моментів, твірна функція кумулянт.* Нехай $p\colon \mathbb{R}^n \to \mathbb{R}$, $p(x) \geqslant 0$ для всіх x. Перетворення Лапласа функції p

$$P(z) = \int p(x)e^{-\langle z,x\rangle}dx,$$

є логарифмічно опуклим. Нехай тепер $p(x)$ є щільністю розподілу, тобто $\int p(x)dx = 1$. Функція $M(z) = P(-z)$ називається *твірною функцією* *моментів* щільності $p(x)$. Вона має таку назву тому що моменти випадкової величини ξ зі щільністю $p(x)$ є значеннями похідних від твірної функції мометів в точці 0, тобто

$$M'(0) = \mathbf{E}\xi, \quad M''(0) = \mathbf{E}\xi\xi^T.$$

Функція $\ln M(z)$, яка є опуклою, називається *твірною функцією кумулянт* щільності $p(x)$, оскільки її похідні дають кумулянти щільності. Наприклад, перша та друга похідні твірної функції кумулянт в нулі, є, відповідно, середнім та матрицею котарацій випадкової величини ξ:

$$(\ln M(x))'\big|_{x=0} = \mathbf{E}\xi, \quad (\ln M(x))''\big|_{x=0} = \mathbf{E}(\xi - \mathbf{E}\xi)(\xi - \mathbf{E}\xi)^T.$$

Інтегрування логарифмічно угнутих функцій

Теорема 2.6.11. *Нехай $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ – логарифмічно угнута за x для всіх y, тоді функція*

$$g(x) = \int f(x,y)dy \tag{2.6.6}$$

є логарифмічно угнутою на $\mathbb{R}^n$.

Ця теорема має багато важливих наслідків. Наприклад, з неї випливає, що маргинальні розподіли випадкових величин з логарифмічно угнутими щільностями є логарифмічно угнутими. Ще декілька наслідків:

Наслідок 2.6.2. *Нехай функції f та g – логарифмічно угнуті на $\mathbb{R}^n$. Тоді згортка функцій f та g*

$$(f * g)(x) = \int f(x-y)g(y)dy$$

теж логарифмічно угнута на $\mathbb{R}^n$.

Наслідок 2.6.3. *Нехай $C \subset \mathbb{R}^n$ – опукла множина, ξ – випадковий вектор з $\mathbb{R}^n$ із логарифмічно угнутою щільністю $p(x)$. Тоді функція*

$$f(x) = \mathbf{P}(x + \xi \in C)$$

логарифмічно угнута по x.

Доведення. Щоб показати це, запишемо f у вигляді

$$f(x) = \int g(x+z)p(z)dz,$$

де g визначається таким чином

$$g(u) = \begin{cases} 1, & u \in C; \\ 0, & u \notin C \end{cases}$$

(g – логарифмічно угнута). Залишилось використати (2.6.6). □

Приклад 2.6.16. *Функція розподілу випадкової величини.* Нехай ξ – випадкова величина з функцією розподілу $F(x)$, тобто
$$F(x) = \mathbf{P}(\xi < x).$$
Якщо щільність $f(x)$ логарифмічно угнута, то функція розподілу $F(x)$ теж логарифмічно угнута.

Приклад 2.6.17. *Об'єм n-вимірного многогранника.* Нехай A – матриця розміру $m \times n$. Визначимо
$$\mathcal{P}_u = \{x \in \mathbb{R}^n \mid Ax \leqslant u\}.$$
Тоді об'єм $\operatorname{vol} \mathcal{P}_u$ – логарифмічно угнута функція від u. Щоб довести це, зауважимо, що функція
$$\Psi(x,u) = \begin{cases} 1, & Ax \leqslant u, \\ 0, & \text{в іншому випадку} \end{cases}$$
логарифмічно угнута. З (2.6.6) випливає, що об'єм многогранника
$$\operatorname{vol} \mathcal{P}_u = \int \Psi(x,u)\, du$$
логарифмічно угнутий.

2.6.4. Опуклість за відношенням порядку

Означення 2.6.7. Нехай $K \subset \mathbb{R}^n$ – правильний конус, який задає відношення порядку $\preceq_K$ на $K \subset \mathbb{R}^n$. Функція $f : \mathbb{R}^n \to \mathbb{R}$ називається *K-неспадною*, якщо
$$x \preceq_K y \Rightarrow f(x) \leqslant f(y),$$
і *K-зростаючою*, якщо
$$x \preceq_K y, x \neq y \Rightarrow f(x) < f(y).$$
K-незростаючі та *K-спадні* функції визначаються аналогічним чином.

Приклад 2.6.18. Функція $f\colon \mathbb{R}^n \to \mathbb{R}$ неспадна відповідно до $\mathbb{R}^n_+$, якщо
$$x_1 \leqslant y_1, \ldots ,x_n \leqslant y_n \Rightarrow f(x) \leqslant f(y)$$
для всіх x, y. Це все одно, що f неспадна за будь-якою своєю компонентою.

Матрично монотонні функції. Функція $f : \mathbb{S}^n \to \mathbb{R}$ називається *матрично монотонною* (зростаючою, спадною), якщо вона монотонна (зростаюча, спадна) по відношенню до конуса невід'ємно визначених матриць.

Наведемо декілька прикладів матрично монотонних функцій від $X \in \mathbb{S}^n$:

1. Функція $\operatorname{Tr} WX$, де $W \in \mathbb{S}^n$ матрично неспадна, якщо W невід'ємно визначена, і матрично зростаюча, якщо W додатньо визначена (вона є матрично незростаючою, якщо W недодатньо визначена, і матрично спадна, якщо W від'ємно визначена)

2. Функція $\operatorname{Tr} X^{-1}$ матрично спадна на множині додатньо визначених матриць.

3. Функція $\det X$ матрично зростаюча на множині невід'ємно визначених матриць.

Теорема 2.6.12. *Диференційовна функція* $f\colon X \to \mathbb{R}$ *є* K*–неспадною тоді і тільки тоді, коли*

$$f'(x) \succeq_{K^*} 0 \quad \textit{для всіх} \quad x \in X. \tag{2.6.7}$$

Вкажемо на відмінність від скалярного випадку: похідна повинна бути невід'ємною відповідно до спряженого відношення порядку. У випадку строгої монотонності маємо: якщо

$$f'(x) \succ_{K^*} 0 \quad \text{для всіх} \quad x \in X, \tag{2.6.8}$$

то f є K-зростаючою. Як і у скалярному випадку обернене твердження не вірне.

Доведення. По-перше, припустимо, що виконується (2.6.7) для всіх x, але f не є K–неспадною, тобто існують такі x, y, що $x \preceq_{K^*} y$ та $f(y) < f(x)$. З диференційовності f випливає, що існує таке $t \in [0,1]$, що

$$\frac{d}{dt} f(x + t(y-x)) = \langle f'(x + t(y-x)), y - x \rangle < 0.$$

Оскільки $y - x \in K$, то це означає, що

$$-f'(x + t(y-x)) \notin K^*.$$

Це суперечить припущенню, що (2.6.7) виконується для всіх $x \in X$. Можна також довести, що з (2.6.8) випливає K-зростання f.
Легко показати, що виконання (2.6.7) всюди є необхідним для K–неспадання f. Припустимо, що (2.6.8) не виконується для $x = z$. Тоді за означенням спряженого конуса існує таке $v \in K$, що

$$\langle f'(z), v \rangle < 0.$$

Тепер розглянемо $h(t) = f(z + tv)$ як функцію від t. Оскільки $h'(0) = \langle f'(z), v \rangle < 0$, то існує таке $t > 0$, що $h(t) = f(z + tv) < h(0) = f(z)$. А це означає, що f не є K–неспадною. ☐

Означення 2.6.8. Нехай $K \subset \mathbb{R}^m$ – правильний конус, який задає відношення порядку $\preceq_K$. Функція $f\colon \mathbb{R}^n \to \mathbb{R}^m$ називається K-*опуклою*, якщо для всіх x, y та $0 \leqslant \lambda \leqslant 1$ виконується

$$f(\lambda x + (1-\lambda) y) \preceq_K \lambda f(x) + (1-\lambda) f(y).$$

Функція f називається *строго K-опуклою*, якщо

$$f(\lambda x + (1-\lambda)y) \prec_K \lambda f(x) + (1-\lambda) f(y).$$

для всіх x, y та $0 < \lambda < 1$.

Ці означення перетворюються на звичайні означення опуклості та строгої опуклості у випадку $m = 1$ та $K = \mathbb{R}_+$.

Приклад 2.6.19. Функція f: $\mathbb{R}^n \to \mathbb{R}^m$ покомпонентно опукла, якщо для всіх x, y та $0 \leqslant \lambda \leqslant 1$

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y),$$

тобто коли кожна компонента $f_i, i = 1, \ldots, m$, є опуклою функцією. Функція f – строго покомпонентно опукла, якщо кожна її компонента є строго опуклою. При цьому відношення порядку задається конусом $\mathbb{R}^m_+$

Матрична опуклість. Нехай функція f приймає значення на множині симетричних матриць, f: $\mathbb{R}^n \to \mathbb{S}^m$. Функція f опукла відповідно до матричного відношення порядку (тобто, відповідно до конусу невід'ємно визначених матриць), якщо

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y)$$

для всіх x, y та $\lambda \in [0,1]$. Це іноді називають *матричною опуклістю*. Еквівалентне означення: функція $\langle f(x)z, z\rangle$ опукла для всіх $z \in \mathbb{R}^m$. Це означення вказує спосіб перевірки матричної опуклості. Функція f є строго матрично опуклою, якщо

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

для всіх $x \neq y$ та $0 < \lambda < 1$ або, якщо $\langle f(x)z, z\rangle$ строго опукла для всіх $z \neq 0$.

Наведемо декілька прикладів:

1. Функція $f(X) = XX^T$, де X – матриця розміру $n \times m$, опукла, оскільки для фіксованого z функція $\langle XX^T z, z\rangle = \|X^T z\|^2$ є опуклою квадратичною формою від компонент матриці X. З цієї ж причини, $f(X) = X^2$ опукла на $\mathbb{S}^n$.

2. Функція X^p матрично опукла на множині додатньо визначених матриць, якщо $1 \leqslant p \leqslant 2$ або $-1 \leqslant p \leqslant 0$, та матрично угнута, якщо $0 \leqslant p \leqslant 1$

3. Функція $f(X) = e^X$ не є опуклою на множині симетричних матриць.

Теорема 2.6.13. *Функція f є K-опуклою тоді і тільки тоді, коли для кожного $w \succeq_K 0$ функція $\langle f, w\rangle$ є опуклою. Функція f є строго K-опуклою тоді і тільки тоді, коли для кожного ненульового $w \succeq_K 0$ функція $\langle f, w\rangle$ є строго опуклою.*

Доведення. Твердження випливає з означення та властивостей спряженого відношення порядку. $\square$

Теорема 2.6.14. *Диференційовна функція $f\colon X \to \mathbb{R}^m$ є K-опуклою тоді і тільки тоді, коли для всіх $x,\ y \in X$*

$$f(y) \succeq_K f(x) + \frac{\partial f(x)}{\partial x}(y-x).$$

Функція f є строго K-опуклою, якщо для всіх $x,\ y \in X,\ x \neq y$

$$f(y) \succ_K f(x) + \frac{\partial f(x)}{\partial x}(y-x).$$

Теорема 2.6.15. *Нехай $g\colon X \to \mathbb{R}^p$, $X \subset \mathbb{R}^n$– K-опукла функція, а $h\colon U \to \mathbb{R}$ – опукла K-неспадна функція на опуклій множині $U \subset \mathbb{R}^m$, $g\,(X) \subset U$. Тоді функція $h(g(x))$ опукла на X.*

Доведення. Функція g – K-опукла. Отже, $g(\lambda x_1 + (1-\lambda x_2)) \preceq_K \lambda g(x_1) + (1-\lambda)g(x_2)$. Функція h – K-зростаюча та опукла. Таким чином

$$h(g(\lambda x_1 + (1-\lambda)x_2)) \leqslant h(\lambda g(x_1) + (1-\lambda)g(x_2))) \leqslant$$
$$\leqslant \lambda h(g(x_1)) + (1-\lambda)h(g(x_2)).$$

$\square$

Приклад 2.6.20. Квадратична форма від матриці:

$$g(X) = X^T AX + B^T X + X^T B + C,$$

де $A \in \mathbb{S}^m$, B – матриця розміру $m \times n$, $C \in \mathbb{S}^n$, є опуклою, якщо A додатньо визначена.

Функція $h\colon \mathbb{S}^n \to \mathbb{R}$, $h(Y) = -\ln\det(-Y)$ опукла та зростаюча на множині всіх симетричних від'ємно визначених матриць. За попередньою теоремою функція

$$f(X) = -\ln\det\big(-(X^T AX + B^T X + X^T B + C)\big)$$

опукла на множині

$$\{X \mid X^T AX + B^T X + X^T B + C < 0\}.$$

Співвідношення між різними типами опуклості.

$$\begin{tikzcd}
\boxed{\text{Строга опуклість}} \arrow[r] \arrow[d] & \boxed{\text{Опуклість}} \arrow[d] \\
\boxed{\text{Строга псевдоопуклість}} \arrow[r] \arrow[d] & \boxed{\text{Псевдоопуклість}} \arrow[d, "\text{Диференційовність}"'] \\
\boxed{\text{Сильна квазіопуклість}} \arrow[r] \arrow[d] & \boxed{\text{Строга квазіопуклість}} \arrow[d, "\text{Напівнеперервність знизу}"'] \\
\boxed{\text{Квазіопуклість}} \arrow[r, equal] & \boxed{\text{Квазіопуклість}}
\end{tikzcd}$$

Задачі

1. Нехай f – опукла функція на опуклій множині X. Довести, що $f(\lambda x^1 + (1-\lambda)x^2) \geqslant \lambda f(x^1) + (1-\lambda)f(x^2)$ для довільних $x^1, x^2 \in X$, $\lambda \notin [0,1]$, для яких $\lambda x^1 + (1-\lambda)x^2 \in X$.
2. Нехай E – опукла обмежена множина в $\mathbb{R}^n \times \mathbb{R}$ і X – її проекція на $\mathbb{R}^n$. Довести, що функція $f(x) = \inf\{\beta \,|\, (x,\beta) \in E\}$ опукла на X.
3. Нехай f – двічі диференційовна функція на опуклому компакті X, причому $\langle f''(\hat{x})h,h\rangle > 0$ для довільних $\hat{x} \in X$, $h \in \mathbb{R}^n$, $h \neq 0$. Довести, що f сильно опукла на X.
4. Нехай f опукла функція на опуклій множині $X \subset \mathbb{R}^n$. Показати, що величина $f'(x;h)$ як функція на $\operatorname{ri} X \times \operatorname{Lin} X$ має властивості:

   а) $f'(x;h)$ опукла за h на $\operatorname{Lin} X \quad \forall x \in \operatorname{ri} X$;

   б) $f'(x;h) \geqslant -f'(x;-h) \quad \forall h \in \operatorname{Lin} X \, \forall x \in \operatorname{ri} X$;

   в) $f'(x;h)$ напівнеперервна зверху на $\operatorname{ri} X \times \operatorname{Lin} X$.
5. Нехай f – сильно опукла функція з константою $\theta > 0$ на опуклій множині X, $\hat{x}$ - точка мінімуму f на X. Отримати такі оцінки:

   а) $\theta\|x - \hat{x}\|^2 \leqslant f(x) - f(\hat{x}) \; \forall x \in X$;

   б) якщо f диференційовна в точці $x \in X$, то $2\theta\|x - \hat{x}\| \leqslant \|f'(x)\|$ та $4\theta(f(x) - f(\hat{x})) \leqslant \|f'(x)\|^2$.
6. *Нерівність Юнга.* Нехай f – зростаюча функція на $\mathbb{R}$, $f(0) = 0$, g – обернена до f функція. Визначимо F та G таким чином
$$F(x) = \int_0^x f(a)da, \quad G(y) = \int_0^y g(a)da.$$

Довести, що F та G спряжені. Дати геометричну інтерпретацію нерівності Юнга $xy \leqslant F(x) + G(y)$.

7. Нехай дані функції $f_0, \ldots, f_n \colon \mathbb{R}_+ \to \mathbb{R}$. Розглядається задача наближення функції f_0 лінійними комбінаціями функцій $f_1, \ldots, f_n$. Для $x \in \mathbb{R}^n$ $f = x_1 f_1 + \ldots x_n f_n$ наближає функцію f_0 з точністю $\varepsilon > 0$ на інтервалі $[0,T]$, якщо $|f(t) - f(t_0)| \leqslant \varepsilon$ для $0 \leqslant t \leqslant T$. Тепер зафіксуємо точність $\varepsilon > 0$ і визначимо *довжину наближення* як найбільше T, для якого f наближає f_0 на інтервалі $[0,T]$:
$$W(x) = \sup\{T : |x_1 f_1(t) + \ldots + x_n f_n(t) - f_0(t)| \leqslant \varepsilon, 0 \leqslant t \leqslant T\}.$$
Показати, що функція W квазіугнута.

8. Розглянемо квадратичну функцію $f \colon \mathbb{R}^n \to \mathbb{R}$ вигляду $f(x) = \langle Hx,x \rangle$. Функція f називається *додатньо субвизначеною*, якщо з того, що $\langle Hx,x \rangle < 0$, випливає, що $Hx \geqslant 0$, або $Hx \leqslant 0$ для будь-якого $x \in \mathbb{R}^n$. Довести, що функція f квазіопукла на $\mathbb{R}^n_+$ тоді і тільки тоді, коли вона додатньо субвизначена.

9. Розглянемо квадратичну функцію $f \colon \mathbb{R}^n \to \mathbb{R}$ вигляду $f(x) = \langle Hx,x \rangle$. Функція f називається *строго додатньо субвизначеною*, якщо з того, що $\langle Hx,x \rangle < 0$, випливає, що $Hx \geqslant 0$, або $Hx \leqslant 0$ для будь-якого $x \in \mathbb{R}^n$. Довести, що функція f псевдоопукла на $\mathbb{R}^n_+/\{0\}$ тоді і тільки тоді, коли вона строго додатньо субвизначена.

10. а) Показати, що необхідну умову опуклості двічі диференційовних функцій:
$$\langle y, f'(x) \rangle = 0 \Rightarrow \langle f''(x)y, y \rangle \geqslant 0 \quad \text{для всіх} \quad x \in X.$$
можна виразити такими двома еквівалентними способами:

1) для всіх $x \in X$ існує $\lambda(x) \geqslant 0$ таке, що
$$f''(x) + \lambda(x) f'(x) \left(f'(x)\right)^T \geqslant 0;$$
2) для всіх $x \in X$ матриця
$$\begin{pmatrix} f''(x) & f'(x) \\ (f'(x))^T & 0 \end{pmatrix}$$
має не більше одного від'ємного власного числа.

б) Показати, що достатню умову опуклості двічі диференційовних функцій:
$$\langle y, f'(x) \rangle = 0 \Rightarrow \langle f''(x)y, y \rangle > 0 \quad \text{для всіх} \quad x \in X, y \in \mathbb{R}^n.$$
можна виразити такими двома еквівалентними способами:

1) для всіх $x \in X$ існує $\lambda(x) \geqslant 0$ таке, що
$$f''(x) + \lambda(x) f'(x) \left(f'(x)\right)^T > 0;$$

2) для всіх $x \in X$ матриця

$$\begin{pmatrix} f''(x) & f'(x) \\ (f'(x))^T & 0 \end{pmatrix}$$

має одне невід'ємне та n додатних власних чисел.

11. Перевірити квазіопуклість функції $f(x) = -x_1 x_2$ на $\operatorname{int} \mathbb{R}^2_+$.

12. Квазілінійна функція на $\mathbb{R}$ (квазіопукла та квазіугнута) монотонна. Розглянемо узагальнення на випадок функцій на $\mathbb{R}^n$. Нехай функція $f\colon \mathbb{R}^n \to \mathbb{R}$ – квазілінійна. Вважатимемо її неперервною. Показати, що її можна зобразити у вигляді $f(x) = g(\langle a,x \rangle)$, де $g\colon \mathbb{R} \to \mathbb{R}$ – монотонна, $a \in \mathbb{R}^n$. Іншими словами, квазілінійна функція є монотонною функцією від лінійної. (Обернене також вірне.)

13. Нехай $g\colon X \to \mathbb{R}$, $h\colon X \to \mathbb{R}$, де X – опукла множина в $\mathbb{R}^n$. Показати, що функція $f\colon X \to \mathbb{R}$, вигляду $f(x) = g(x)/h(x)$ квазіопукла, якщо виконуються дві таких умови:

    а) g – опукла на X та $g(x) \geqslant 0$ для будь-якого $x \in X$;

    б) h – угнута на X та $h(x) > 0$ для будь-якого $x \in X$.

14. Нехай $g\colon X \to \mathbb{R}$, $h\colon X \to \mathbb{R}$, де X – опукла множина в $\mathbb{R}^n$. Показати, що функція $f\colon X \to \mathbb{R}$, вигляду $f(x) = g(x)/h(x)$ квазіопукла, якщо виконуються дві таких умови:

    а) g – опукла на X та $g(x) \leqslant 0$ для будь-якого $x \in X$;

    б) h – опукла на X та $h(x) > 0$ для будь-якого $x \in X$.

15. Нехай $g\colon X \to \mathbb{R}$, $h\colon X \to \mathbb{R}$, де X – опукла множина в $\mathbb{R}^n$. Показати, що функція $f\colon X \to \mathbb{R}$, вигляду $f(x) = g(x)h(x)$ квазіопукла, якщо виконуються дві таких умови:

    а) g – опукла на X та $g(x) \leqslant 0$ для будь-якого $x \in X$;

    б) h – угнута на X та $h(x) > 0$ для будь-якого $x \in X$.

16. Нехай $f\colon \mathbb{R}^n \to \mathbb{R}^m$, $g\colon \mathbb{R}^n \to \mathbb{R}^k$ – диференційовні та опуклі функції, функція $\phi\colon \mathbb{R}^{m+k} \to \mathbb{R}$ має таку властивість: якщо $a_2 \geqslant a_1$ та $b_2 \geqslant b_1$, то $\phi(a_2,b_2) \geqslant \phi(a_1,b_1)$. Розглянемо функцію $h: \mathbb{R}^n \to \mathbb{R}$ вигляду $h(x) = \phi(f(x),g(x))$. Показати, що:

    а) якщо ϕ – опукла, то h – опукла функція;

    б) якщо ϕ – псевдоопукла, то й h - псевдоопукла;

    в) якщо ϕ – квазіпукла, то й h - квазіоопукла.

17. Показати, що функція $f(x) = e^x/(1 + e^x)$, яку іноді називають *логістичною функцією*, логарифмічно угнута.

18. Довести, що середнє гармонічне

$$H(x) = \frac{1}{\dfrac{1}{x_1} + \ldots + \dfrac{1}{x_n}}$$

величин $x_1, \ldots, x_n > 0$ логарифмічно угнуте.

19. Нехай $f\colon \mathbb{R}_+ \to \mathbb{R}$ – невід'ємна. Для $x \geqslant 0$ визначимо

$$M(x) = \int_0^\infty u^x f(u)du.$$

Коли x – додатнє ціле число та f – щільність розподілу, $M(x)$ є x-тим моментом випадкової величини зі щільністю f. Показати, що M логарифмічно угнута.

Підказка: для кожного $u \geqslant 0$ фунуція u^x логарифмічно опукла на $\operatorname{int} \mathbb{R}_+$.

Використати доведене для того щоб показати, що гамма функція

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$$

логарифмічно опукла при $x \geqslant 1$.

20. Функція нормального розподілу

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

логарифмічно угнута. Це випливає із загального результату про те, що згортка двох логарифмічно угнутих функцій логарифмічно угнута. В цій задачі ми надамо просте доведення того, що f логарифмічно угнута без посилань на цей результат. Нагадаємо, що f є логарифмічно угнутою тоді і тільки тоді, коли $f''(x)f(x) \leqslant (f'(x))^2$ для всіх x.

а) Перевірити, що f логарифмічно угнута при $x \geqslant 0$.

б) Перевірити, що для будь-якого t та x виконується $t^2/2 \geqslant -x^2/2 + xt$.

в) Використовуючи б) показати, що $e^{-t^2/2} \leqslant e^{x^2/2 - xt}$. Вивести звідси

$$\int_{-\infty}^x e^{-t^2/2} dt \leqslant e^{x^2/2} \int_{-\infty}^x e^{-xt} dt.$$

г) Використати в) для перевірки $f''(x)f(x) \leqslant (f'(x))^2$ при $x \leqslant 0$.

21. Нехай $g(t) = \exp(-h(t))$ – диференційовна логарифмічно угнута щільність розподілу і

$$f(x) = \int_{-\infty}^x g(t)dt = \int_{-\infty}^x e^{-h(t)} dt$$

її функція розподілу. Покажемо, що f логарифмічно угнута, тобто вона задовольняє співвідношення $f''(x)f(x) \leqslant (f'(x))^2$.

а) Виразити похідні f через функцію h та її похідні. Перевірити те, що f логарифмічно угнута якщо $h'(x) \geqslant 0$.

б) Припустити, що $h'(x) < 0$. Використати нерівність

$$h(t) \geqslant h(x) + h'(x)(t - x)$$

(яка випливає з опуклості h), щоб довести, що

$$\int_{-\infty}^{x} e^{-h(t)} dt \leqslant \frac{e^{-h(x)}}{-h'(x)}.$$

Використати цю нерівність для перевірки логарифмічної угнутості f.

22. Ймовірнісна міра π на $\mathbb{R}^n$ – логарифмічно угнута, якщо

$$\pi((1-\lambda)C_1 + \lambda C_2) \geqslant \pi(C_1)^{1-\lambda} \pi(C_2)^{\lambda}$$

для всіх опуклих підмножин C_1 та C_2 з $\mathbb{R}^n$ та всіх $\lambda \in [0,1]$.

Показати, що коли міра π породжена щільністю p, тобто, $\pi(A) = \int_A p(x)dx$, то вона є логарифмічно угнутою тоді і тільки тоді, коли щільність p логарифмічно угнута.

2.7. ПОХІДНІ ЗА НАПРЯМКОМ І СУБДИФЕРЕНЦІАЛИ

Означення 2.7.1. Нехай $f(x)$ – функція на просторі X. Величина

$$f'(x,p) = \lim_{\lambda \downarrow 0} \frac{f(x+\lambda p) - f(x)}{\lambda} \tag{2.7.1}$$

називається *похідною за напрямком* $p \in X$ функції f в точці $x \in X$.

Лема 2.7.1. *Нехай $f(x)$ – власна опукла функція, $x_0 \in \operatorname{dom} f$. Тоді величина $f'(x_0,p)$ (скінченна чи нескінченна) існує для всіх $p \in X$.*

Доведення. Якщо $x_0 + \lambda p \notin \operatorname{dom} f$ при всіх $\lambda > 0$, то $f(x_0 + \lambda p) = +\infty$. Отже $f'(x_0,p) = +\infty$. Якщо $x_0 + \lambda p \in \operatorname{dom} f$ при малих $\lambda > 0$, то за лемою 2.1.10 відношення

$$\frac{f(x_0 + \lambda p) - f(x_0)}{\lambda}$$

є незростаюча функція λ, коли λ прямує до нуля. Тому границя

$$f'(x_0,p) = \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda p) - f(x_0)}{\lambda}$$

існує. □

Лема 2.7.2. Про скінченність похідної за напрямком. *Якщо величина $x_0 + \lambda p \in \operatorname{dom} f$ для $\lambda \in [-\varepsilon,\varepsilon]$, $\varepsilon > 0$, то похідна за напрямком $f'(x_0,p)$ скінченна.*

Доведення. З другої нерівності леми 2.1.10 при $\alpha_0 = -\varepsilon, \alpha_1 = 0, \alpha_2 = \lambda$ маємо

$$\frac{f(x_0) - f(x_0 - \varepsilon p)}{\varepsilon} \le \frac{f(x_0 + \lambda p) - f(x_0)}{\lambda},$$

Переходячи до границі при $\lambda \downarrow 0$, отримаємо

$$\frac{f(x_0 + \lambda p) - f(x_0)}{\lambda} \ge f'(x_0,p) \ge \frac{f(x_0) - f(x_0 - \varepsilon p)}{\varepsilon}. \tag{2.7.2}$$

□

Лема 2.7.3. *Похідна за напрямком $f'(x,p)$ – додатньо однорідна функція напрямку $p \in X$. Якщо функція $f(x)$ опукла, то $f'(x,p)$ – додатньо однорідна опукла функція напрямку $p \in X$.*

Означення 2.7.2. Нехай $f(x)$ – функція на просторі X. Вектор $x^* \in X^*$ називається *субградієнтом* функції $f(x)$ в точці $x_0 \in X$, якщо справджується нерівність

$$f(x) - f(x_0) \geqslant \langle x^*, x - x_0 \rangle \quad \forall x \in X. \tag{2.7.3}$$

Множина всіх субградієнтів називається *субдиференціалом* функції f в точці x_0 і позначається $\partial f(x_0)$.

Лема 2.7.4. *Вектор x^* є субградієнтом функції $f(x)$ в точці x_0 тоді і тільки тоді, коли*

$$f'(x_0,p) \ge \langle x^*, p \rangle \tag{2.7.4}$$

для всіх $p \in X$.

Доведення. Якщо x^* – субградієнт, то $f(x_0+\lambda p)-f(x_0) \ge \lambda\langle x^*, p\rangle$, звідки

$$\frac{f(x_0 + \lambda p) - f(x_0)}{\lambda} \ge \langle x^*, p \rangle$$

і $f'(x_0,p) \ge \langle x^*, p \rangle$. Навпаки, якщо нерівність (2.7.4) виконана, то для $0 < \lambda < 1$,

$$f(x) - f(x_0) = \frac{f(x_0 + 1(x - x_0)) - f(x_0)}{1} \ge \frac{f(x_0 + \lambda(x - x_0)) - f(x_0)}{\lambda}$$
$$\ge f'(x_0, x - x_0) \ge \langle x^*, x - x_0 \rangle,$$

тобто x^* - субградієнт. □

Наслідок 2.7.1. *З леми 2.7.4 випливає, що*

$$\partial f(x_0) = \partial_p f'(x_0,0), \tag{2.7.5}$$

де $\partial_p f'(x_0,p)$ означає субдиференціал функції $f'(x_0,p)$ за аргументом $p \in X$.

Теорема 2.7.1. *Якщо $f'(x_0,p)$ – замкнута опукла функція, $p \in X$, то множина $\partial f(x_0)$ непорожня і*
$$f'(x_0,p) = \sup_{x^*}\{\langle x^*,p\rangle \colon x^* \in \partial f(x_0)\}.$$

Доведення. Функція $f'(x,p)$ замкнута додатньо однорідна опукла функція напрямку p. Тому можна застосувати теорему 2.5.2. За цією теоремою
$$f'(x_0,p) = \sup_{x^*}\{\langle x^*,p\rangle \colon x^* \in \operatorname{dom}(f'(x_0,\cdot))^*\}, \qquad (2.7.6)$$
де функція $(f'(x_0,\cdot))^*$ спряжена до функції $f'(x_0,p)$ відносно p, тобто
$$(f'(x_0,\cdot))^*(x^*) = \sup_{p}\{\langle x^*,p\rangle - f'(x_0,p)\}. \qquad (2.7.7)$$
З іншого боку
$$(f'(x_0,\cdot))^*(x^*) = \begin{cases} 0, & x^* \in \operatorname{dom}(f'(x_0,\cdot))^*; \\ +\infty, & x^* \notin \operatorname{dom}(f'(x_0,\cdot))^*. \end{cases}$$
Враховуючи співвідношення (2.7.7), отримаємо, що $x^* \in \operatorname{dom}(f'(x_0,\cdot))^*$ тоді і тільки тоді, коли
$$0 \geq \langle x^*,p\rangle - f'(x_0,p).$$
Звідки
$$\partial f(x_0) = \partial_p f'(x_0,0) = \operatorname{dom}(f'(x_0,\cdot))^*.$$
Це співвідношення разом із співвідношенням (2.7.6) доводить справедливість твердження теореми. □

Безпосередньо з означення випливає справедливість наступної теореми.

Теорема 2.7.2. *Субдиференціал $\partial f(x_0)$ – це замкнута опукла множина.*

Теорема 2.7.3. *Включення $x^* \in \partial f(x_0)$ справджується тоді і тільки тоді, коли*
$$\langle x^*,x_0\rangle - f(x_0) = f^*(x^*).$$

Доведення. За означенням $x^* \in \partial f(x_0)$ тоді, коли
$$\langle x^*,x_0\rangle - f(x_0) \geq \langle x^*,x\rangle - f(x).$$
Взявши супремум в правій частині нерівності, отримаємо
$$\langle x^*,x_0\rangle - f(x_0) \geq f^*(x^*).$$
За нерівністю Юнга-Фенхеля
$$\langle x^*,x_0\rangle - f(x_0) \leq f^*(x^*).$$
Тому
$$\langle x^*,x_0\rangle - f(x_0) = f^*(x^*).$$
□

Теорема 2.7.4. *Якщо опукла функція f диференційовна в точці x_0, то $\partial f(x_0) = \{f'(x_0)\}$, тобто градієнт $f'(x_0)$ є єдиним субградієнтом функції f в точці x_0.*

Доведення. Нехай $a \in \partial f(x_0)$. Якщо опукла функція диференційовна в точці x_0, то для всіх p маємо

$$f'(x_0,p) = \langle f'(x_0),p\rangle \geqslant \langle a,p\rangle,$$

тобто

$$\langle f'(x_0) - a,p\rangle \geqslant 0.$$

Якщо взяти $p = a - f'(x_0)$, то отримаємо, що $-\|f'(x_0) - a\|^2 \geqslant 0$, тобто $a = f'(x_0)$. Отже $\partial f(x_0) = \{f'(x_0)\}$. □

Теорема 2.7.5. Про субдиференціал неперервної функції. *Нехай опукла функція f неперервна в точці x_0. Тоді $\partial f(x_0)$ - непорожня замкнута опукла обмежена множина і*

$$f'(x_0,p) = \max_{x^*} \{\langle p,x^*\rangle \colon x^* \in \partial f(x_0)\}.$$

Доведення. Оскільки $f(x_0)$ неперервна в точці x_0, то за лемою 2.7.2 про скінченність похідної за напрямком $f'(x_0,p)$ – скінченна величина. Розглянемо в $\mathbb{R} \times X$ промінь

$$\mathcal{L} = \{(\alpha,x) \colon \alpha = f(x_0) + \lambda f'(x_0,p), x = x_0 + \lambda p, \lambda \geq 0\}$$

і множину

$$A = \{(\alpha,x) \colon \alpha > f(x)\}.$$

Множини A, $\mathcal{L}$ опуклі і не перетинаються. Доведемо це. Опуклість A одразу випливає з опуклості f. З нерівності

$$\frac{f(x_0 + \lambda p) - f(x_0)}{\lambda} \geq f'(x_0,p)$$

випливає, що

$$f(x_0 + \lambda p) \geq f(x_0) + \lambda f'(x_0,p).$$

Якщо ж точка $(f(x_0 + \lambda f'(x_0,p), x_0 + \lambda p)$ променя $\mathcal{L}$ належить A, то

$$f(x_0) + \lambda f'(x_0,p) > f(x_0 + \lambda p),$$

що суперечить попередній нерівності.

За теоремою про розділення існує такий не рівний нулю вектор $(\beta,x^*) \in \mathbb{R} \times X$, що для $(y^0,y) \in \mathcal{L}, (x^0,x) \in A$

$$y^0\beta + \langle y,x^*\rangle \leq x^0\beta + \langle x,x^*\rangle. \tag{2.7.8}$$

Оскільки $x^0 > f(x)$ для $(x^0,x) \in A$, то x^0 можна спрямувати до нескінченності. Нерівність (2.7.8) тоді порушується у випадку $\beta < 0$. Тому $\beta \geq 0$. Якщо $\beta = 0$, то (2.7.8) переходить в нерівність

$$\langle y,x^*\rangle \leq \langle x,x^*\rangle, \quad (y^0,y) \in \mathcal{L}, \quad x \in \operatorname{dom} f. \tag{2.7.9}$$

Точка $(f(x_0),x_0) \in \mathcal{L}$, з іншого боку $x_0 + \varepsilon z \in \mathrm{dom}\, f, z \in B$, при досить малому $\varepsilon > 0$, оскільки x_0 - точка неперервності f. Отже, $x_0 \in \mathrm{dom}\, f$. Тому з (2.7.9) випливає

$$\langle x_0,x^* \rangle \leq \langle x_0,x^* \rangle + \varepsilon \langle z,x^* \rangle,$$

тобто

$$\langle z,x^* \rangle \geq 0, \quad z \in B,$$

де B - одинична куля з центром в нулі. Останнє можливе лише в тому випадку, коли $x^* = 0$ тобто $(\beta,x^*) = (0,0)$. Отримали суперечність.

Отже $\beta > 0$ і можна покласти $\beta = 1$ (інакше потрібно поділити обидві частини нерівності (2.7.8) на β). Підставимо в нерівність (2.7.8) замість $(y^0,y) \in \mathcal{L}$ вираз із означення, а в правій частині покладемо $x^0 = f(x)$. Останнє можливе за неперервністю, оскільки нерівність (2.7.8) справедлива для всіх $x^0 > f(x)$. Тепер вона має вигляд:

$$f(x_0) + \lambda f'(x_0,p) + (x_0 + \lambda p,x^*) \leq f(x) + (x,x^*), \lambda \geq 0. \tag{2.7.10}$$

Покладемо $\lambda = 0$, отримаємо:

$$\langle x - x_0, -x^* \rangle \leq f(x) - f(x_0),$$

тобто $x_0^* = -x^* \in \partial f(x_0)$. Таким чином $\partial f(x_0) \neq \emptyset$. Враховуючи це, можна переписати нерівність (2.7.10) у наступному вигляді:

$$f(x) - f(x_0) \geq \langle x - x_0,x_0^* \rangle + \lambda[f'(x_0,p) - \langle p,x_0^* \rangle].$$

Покладаючи $x = x_0$, отримаємо

$$f'(x_0,p) \leq \langle p,x_0^* \rangle, \quad x_0^* \in \partial f(x_0). \tag{2.7.11}$$

За лемою 2.7.4

$$f'(x_0,p) \geq \langle p,x_0^* \rangle, \quad x_0^* \in \partial f(x_0). \tag{2.7.12}$$

Порівнявши нерівності (2.7.11), (2.7.12), отримаємо

$$f'(x_0,p) = \langle p,x_0^* \rangle = \max_{x^*} \{\langle p,x^* \rangle : x^* \in \partial f(x_0)\}. \tag{2.7.13}$$

Залишилось довести, що $\partial f(x_0)$ є обмежена множина. Функція f задовольняє умову Ліпшиця в точці x_0, тому

$$f'(x_0,p) \leq \frac{f(x_0 + \lambda p) - f(x_0)}{\lambda} \leq L\|p\|.$$

Звідки, з урахуванням (2.7.13), отримаємо

$$\langle p,x^* \rangle \leq L\|p\|$$

для всіх $x^* \in \partial f(x_0)$ і довільних p. Покладемо $p = x^*$. Отримаємо

$$\|x^*\| \leq L.$$

□

Теорема 2.7.6. *Нехай f – опукла функція. Якщо $x_0 \in \mathrm{ri}\,\mathrm{dom}\, f$, то $\partial f(x_0)$ - непорожня множина і*

$$f'(x_0,p) = \sup_{x^*} \{\langle p,x^* \rangle : x^* \in \partial f(x_0)\}.$$

Теорема 2.7.7. *Нехай $f(x)$ – опукла функція, $\alpha > 0$. Тоді*
$$\partial (\alpha f(x)) = \alpha \partial f(x).$$

Теорема 2.7.8. Теорема Моро–Рокафеллара. *Нехай $f_1, \ldots, f_n$ – власні опуклі функції на X. Тоді для кожної точки $x \in X$*
$$\partial f_1(x) + \cdots + \partial f_n(x) \subset \partial (f_1 + \cdots + f_n)(x). \qquad (2.7.14)$$
Якщо в деякій точці $x_1 \in (\mathrm{dom}\ f_1) \cap \cdots \cap (\mathrm{dom}\ f_n)$ функції $f_1(x), \ldots, f_n(x)$, за виключенням може однієї, неперервні, то
$$\partial f_1(x) + \cdots + \partial f_n(x) = \partial (f_1 + \cdots + f_n)(x) \qquad (2.7.15)$$
для всіх $x \in X$.

Доведення. Справедливість включення (2.7.14) випливає з означення субдиференціалу. Доведемо зворотне включення в точці $x = 0$. Будемо вважати, що $n = 2$, $f_1(0) = 0$, $f_2(0) = 0$. Нехай $x^* \in \partial (f_1 + f_2)(0)$. За означенням це означає, що
$$f_1(x) + f_2(x) \geq \langle x, x^* \rangle \quad \forall x \in X.$$
Запишемо цю нерівність у вигляді
$$f_1(x) - \langle x, x^* \rangle \geq -f_2(x) \quad \forall x \in X. \qquad (2.7.16)$$
Визначимо у просторі $\mathbb{R} \times X$ опуклі множини
$$A = \{(\alpha, x)\colon \alpha > f_1(x) - \langle x, x^* \rangle\},$$
$$B = \{(\gamma, y)\colon \gamma < -f_2(y)\}.$$
Опуклість цих множин випливає із опуклості функцій. З нерівності (2.7.16) випливає, що множини A, B не перетинаються. Тому їх можна розділити, тобто існує ненульовий вектор (β, x_0^*) такий, що
$$\gamma\beta + \langle y, x_0^* \rangle \leq \alpha\beta + \langle x, x_0^* \rangle, \quad (\alpha, x) \in A, (\gamma, y) \in B. \qquad (2.7.17)$$
Величина β не може бути від'ємною. Покажемо, що $\beta > 0$. Дійсно, якщо $\beta = 0$, то нерівність (2.7.17) можна записати у вигляді
$$\langle y, x_0^* \rangle \leq \langle x, x_0^* \rangle, \quad x \in \mathrm{dom}\ f_1, y \in \mathrm{dom}\ f_2. \qquad (2.7.18)$$
Покладемо $y = x_1$. Оскільки $x_1 \in \mathrm{dom}\ f_1$ – точка неперервності функції f_1, то $x_1 + \varepsilon z \in \mathrm{dom}\ f_1$, $\|z\| \leq 1$, при досить малих ε. Підставимо $y = x_1$, $x = x_1 + \varepsilon z$ в нерівність (2.7.18). Отримаємо
$$0 \leq \varepsilon \langle z, x_0^* \rangle, \quad \|z\| \leq 1.$$
Це можливо лише при $x_0^* = 0$, що суперечить теоремі про розділення. Отже $\beta > 0$ і можна брати $\beta = 1$. Тепер нерівність (2.7.17) можна записати у вигляді
$$-f_2(y) + \langle y, x_0^* \rangle \leq f_1(x) - \langle x, x^* \rangle + \langle x, x_0^* \rangle, x \in \mathrm{dom}\ f_1, y \in \mathrm{dom}\ f_2. \qquad (2.7.19)$$
Оскільки за межами $\mathrm{dom}\ f_1$ та $\mathrm{dom}\ f_2$ значення функцій нескінченні, то можна вважати, що нерівність (2.7.19) справджується для всіх x, y. Взявши $y = 0$, отримаємо
$$\langle x, x^* - x_0^* \rangle \leq f_1(x),$$

тобто $(x^* - x_0^*) \in \partial f_1(0)$. Взявши $x = 0$, отримаємо

$$\langle y, x_0^* \rangle \leq f_2(y),$$

тобто $x_0^* \in \partial f_2(0)$. Тому

$$x^* = (x^* - x_0^*) + x_0^* \in \partial f_1(0) + \partial f_2(0).$$

Оскільки x^* – довільний елемент $\partial(f_1 +_2)(0)$, то

$$\partial f(0) \subset \partial f_1(0) + \partial f_2(0).$$

Враховуючи протилежне включення (2.7.14), отримаємо

$$\partial f(0) = \partial f_1(0) + \partial f_2(0).$$

□

Теорема 2.7.9. *Нехай* $f_1, \ldots, f_n$ *– власні опуклі функції на* X. *Нехай виконується умова* $(\mathrm{ri\,dom}\ f_1) \cap \cdots \cap (\mathrm{ri\,dom}\ f_n) \neq \emptyset$. *Тоді*

$$\partial f_1(x) + \cdots + \partial f_n(x) = \partial \left(f_1 + \cdots + f_n\right)(x)$$

для всіх $x \in (\mathrm{dom}\ f_1) \cap \cdots \cap (\mathrm{dom}\ f_n)$.

Застосуємо теореми 2.7.8, 2.7.9 до обчислення субдиференціалів індикаторних функцій опуклих множин. Нехай M–опукла множина, та нехай $\delta(x|M)$ – індикаторна функція множини M. Обчислимо субдиференціал $\partial\delta(x_0|M)$, $x_0 \in M$. За означенням $x^* \in \partial\delta(x_0|M)$ тоді і тільки тоді коли

$$\delta(x|M) - \delta(x_0|M) \geq \langle x^*, x - x_0 \rangle.$$

Але $\delta(x_0|M) = 0$. Якщо $x \notin M$, то $\delta(x|M) = +\infty$ і нерівність виконується завжди. Для $x \in M$ отримаємо

$$0 \geq \langle x^*, x - x_0 \rangle. \tag{2.7.20}$$

Нехай

$$\mathrm{cone}\,(M - x_0) = \{p \colon p = \lambda(x - x_0), \lambda > 0, x \in M\}.$$

Тоді нерівність (2.7.20) еквівалентна тому, що

$$-x^* \in (\mathrm{cone}\,(M - x_0))^*$$

Отже

$$\partial\delta(x_0|M) = -\left(\mathrm{cone}\,(M - x_0)\right)^*. \tag{2.7.21}$$

Нехай тепер $K_i, i = 1, \ldots, m$, – опуклі конуси і нехай $0 \in K_i$. Тоді $\mathrm{cone}\,(K_i - 0) = K_i$. Тому

$$\partial\delta(0|K_i) = -K_i^*. \tag{2.7.22}$$

Зауважимо, що

$$\delta\left(x | \cap_{i=1}^m K_i\right) = \delta\left(x|K_1\right) + \cdots + \delta\left(x|K_m\right). \tag{2.7.23}$$

Якщо застосувати до $\delta\left(x | \cap_{i=1}^m K_i\right)$ теорему 2.7.9, то отримаємо такий результат

Теорема 2.7.10. *Нехай* $K_i, i = 1, \ldots, m,$ *– опуклі конуси і нехай* $(\mathrm{ri}\ K_1) \cap \cdots \cap (\mathrm{ri}\ K_m) \neq \emptyset$. *Тоді*

$$(K_1 \cap \cdots \cap K_m)^* = K_1^* + \cdots + K_m^*.$$

Доведення. Ясно, що $\mathrm{dom}\ \delta\,(\cdot|K_i) = K_i$. Тому з умов теореми випливає, що

$$(\mathrm{ri\ dom}\ \delta\,(\cdot|K_1)) \cap \cdots \cap (\mathrm{ri\ dom}\ \delta\,(\cdot|K_m)) \neq \emptyset.$$

Можна застосувати теорему 2.7.8 до функції (2.7.23). Враховуючи рівність (2.7.22), отримаємо

$$-\left(\cap_{i=1}^m K_i\right)^* = \partial\delta\,(0|\cap_{i=1}^m K_i) =$$

$$= \partial\delta\,(0|K_1) + \cdots + \partial\delta\,(0|K_m) = -K_1^* - \cdots - K_m^*.$$

□

Теорема 2.7.11. *Нехай* M^* *– замкнута опукла множина,*

$$f(x) = \sup_{x^*} \{\langle x^*, x\rangle : x^* \in M^*\}.$$

Тоді

$$\partial f(x_0) = \{x^* \in M^* : \langle x^*, x_0\rangle = f(x_0)\}.$$

Якщо $x_0 = 0$, *то* $\partial f(0) = M^*$.

Доведення. Якщо $x^* \in M^*$, $\langle x^*, x_0\rangle = f(x_0)$, то

$$f(x) - f(x_0) \geq \langle x^*, x\rangle - \langle x^*, x_0\rangle = \langle x^*, x - x_0\rangle,$$

Тобто $x^* \in \partial f(x_0)$.

Нехай $x_0^* \in \partial f(x_0)$. Припустимо, що $x_0^* \notin M^*$. Тоді за теоремою про розділення існує такий вектор p, що

$$\sup_{x^*} \{\langle x^*, p\rangle : x^* \in M^*\} < \langle x_0^*, p\rangle. \tag{2.7.24}$$

З іншого боку

$$\sup_{x^*} \{\langle x^*, x - x_0\rangle : x^* \in M^*\} \geq f(x) - f(x_0) \geq \langle x_0^*, x - x_0\rangle.$$

Покладаючи $x = x_0 + p$, отримаємо

$$\sup_{x^*} \{\langle x^*, p\rangle : x^* \in M^*\} \geq \langle x_0^*, p\rangle.$$

Це суперечить нерівності (2.7.24). Отже $x_0^* \in M^*$. За припущенням $x_0^* \in \partial f(x_0)$. Тому

$$f(x) - \langle x_0^*, x\rangle \geq f(x_0) - \langle x_0^*, x_0\rangle.$$

Покладемо $x = 0$. Отримаємо $\langle x_0^*, x_0\rangle \geq f(x_0)$. Але $x_0^* \in M^*$. Тому $\langle x_0^*, x_0\rangle \leq f(x_0)$. Отже

$$\langle x_0^*, x_0\rangle = f(x_0).$$

Якщо $x_0 = 0$, то $f(0) = \langle x^*, 0\rangle = 0$ для всіх $x^* \in M^*$. Тому $\partial f(0) = M^*$. □

Теорема 2.7.12. *Нехай $f(x)$ – додатньо однорідна замкнута опукла функція. Тоді*

$$\partial f(x_0) = \{x^* \in \operatorname{dom} f^* \colon \langle x^*, x_0 \rangle = f(x_0)\}\,.$$

Нехай $f(x,\alpha)$ – опуклі при кожному $\alpha \in A$ функції від x, A – деяка множина індексів. Функція

$$f(x) = \sup_{\alpha} \{f(x,\alpha) \colon \alpha \in A\}$$

опукла. Вияснимо як субдиференціал функції $f(x)$ виражається через субдиференціали функцій $f(x,\alpha)$.

Лема 2.7.5. *Нехай множина A компактна і функція $f(x,\alpha)$ неперервна за x та α в околі точки x_0 і $\alpha \in A$. Тоді функція*

$$f(x) = \max_{\alpha} \{f(x,\alpha) \colon \alpha \in A\} \tag{2.7.25}$$

неперервна в точці x_0.

Доведення. Позначимо

$$A(x) = \{\alpha \in A \colon f(x,\alpha) = f(x)\}\,.$$

Тоді

$$f(x,\alpha) = f(x) \geq f(x,\alpha_0),$$

$$f(x_0,\alpha) \leq f(x_0) = f(x_0,\alpha_0)$$

при $\alpha \in A(x)$, $\alpha_0 \in A(x_0)$. Віднімаючи ці нерівності, отримаємо

$$f(x,\alpha) - f(x_0,\alpha) \geq f(x) - f(x_0) \geq f(x,\alpha_0) - f(x_0,\alpha_0). \tag{2.7.26}$$

При $x \to x_0$ вираз в правій частині прямує до нуля. Покажемо, що і ліва частина прямує до нуля. Нехай $x_k \to x_0$ і $\alpha_k \in A(x_k) \subset A$,

$$|f(x_k,\alpha_k) - f(x_0,\alpha_k)| \geq \varepsilon > 0.$$

Оскільки A – компактна множина, то можна вважати, що $\alpha_k \to \alpha \in A$. Тоді

$$f(x_k,\alpha_k) - f(x_0,\alpha_k) \to f(x_0,\alpha) - f(x_0,\alpha) = 0.$$

Це суперечить припущенню. Отже при $x \to x_0$ і ліва і права частини нерівності (2.7.26) прямують до нуля. Отже $f(x) \to f(x_0)$. □

Лема 2.7.6. *Нехай виконуються умови леми 2.7.5. Тоді для кожної відкритої множини C такої, що $A(x_0) \subset C \subset A$ існує такий окіл точки x_0, що для всіх точок із цього околу виконується включення*

$$A(x) \subset C.$$

Доведення. Нехай для деякої відкритої множини C_0, $A(x_0) \subset C_0$, існує така послідовність $x_k \to x_0$, що знайдуться точки $\alpha_k \in A(x_k)$, $\alpha_k \notin C_0$.

Оскільки A – компактна множина, то можна вважати, що $\alpha_k \to \alpha_0$. За означенням

$$f(x_k) = f(x_k,\alpha_k).$$

Переходячи до границі, за лемою 2.7.5 отримаємо

$$f(x_0) = f(x_0,\alpha_0),$$

Тобто $\alpha_0 \in A(x_0) \subset C_0$. Тоді α_k при достатньо великих k мають належати відкритій множині C_0 – околу точки α_0. Це суперечить тому, що $\alpha_k \notin C_0$. □

Теорема 2.7.13. *Нехай множина A компактна і функція $f(x,\alpha)$ опукла за x при кожному $\alpha \in A$. Нехай функція $f(x + \lambda p,\alpha)$ неперервна за λ та α для $\lambda \in [-\delta,\delta]$, $\delta > 0$, $\alpha \in A$. Тоді функція*

$$f(x) = \max_{\alpha} \{f(x,\alpha) \colon \alpha \in A\}$$

диференційовна за напрямком p і

$$f'(x_0,p) = \max_{\alpha} \{f'(x_0,p,\alpha) \colon \alpha \in A(x_0)\},$$

де $f'(x_0,p,\alpha)$ – похідна функції $f(x,\alpha)$ за напрямком p в точці x_0.

Доведення. Нехай

$$g(\lambda) = f(x_0 + \lambda p), \quad g(\lambda,\alpha) = f(x_0 + \lambda p,\alpha), \quad A(\lambda) = A(x_0 + \lambda p).$$

Тоді

$$f'(x_0,p) = g'(0), \quad f'(x_0,p,\alpha) = g'(0,\alpha).$$

Функція

$$g(\lambda) = \max_{\alpha} \{g(\lambda,\alpha) \colon \alpha \in A\}$$

скінченна при $\lambda \in [-\delta,\delta]$ оскільки $g(\lambda,\alpha)$ неперервна за α, а множина A – компактна. Тому похідна $g'(0)$ існує і скінченна. Поклавши в (2.7.26) $x = x_0 + \lambda p$, $\lambda > 0$, отримаємо

$$\frac{g(\lambda,\alpha) - g(0,\alpha)}{\lambda} \geq \frac{g(\lambda) - g(0)}{\lambda} \geq \frac{g(\lambda,\alpha_0) - g(0,\alpha_0)}{\lambda}, \tag{2.7.27}$$

для всіх $\alpha \in A(\lambda)$, $\quad \alpha_0 \in A(0)$. З правої нерівності (2.7.27) отримаємо

$$\frac{g(\lambda) - g(0)}{\lambda} \geq g'(0,\alpha_0), \quad \alpha_0 \in A(0).$$

Тому

$$g'(0) \geq \sup_{\alpha_0} \{g'(0,\alpha_0) \colon \alpha_0 \in A(0)\}.$$

Розглянемо ліву нерівність (2.7.27). За лемою 2.7.6 для відкритої множини $C \supset A(x_0) = A(0)$ знайдеться таке $\varepsilon > 0$, що $A(\lambda) \subset C$ при $0 \leq \lambda < \varepsilon$. Тому при $\lambda < \varepsilon$ з нерівності (2.7.27) отримаємо

$$g'(0) \leq \frac{g(\lambda) - g(0)}{\lambda} \leq \sup_{\alpha} \left\{ \frac{g(\lambda,\alpha) - g(0,\alpha)}{\lambda} \colon \alpha \in C \right\}. \tag{2.7.28}$$

Розглянемо точну верхню грань виразу, що стоїть в правій частині нерівності (2.7.28) по всіх $C \supset A(0)$. Оскільки функція $g(\lambda,\alpha)$ неперервна за α і $A(0)$ замкнута підмножина компактної множини, то можна переконатися, що

$$\inf_{C \supset A(0)} \sup_{\alpha} \left\{ \frac{g(\lambda,\alpha) - g(0,\alpha)}{\lambda} : \alpha \in C \right\} =$$
$$= \max_{\alpha} \left\{ \frac{g(\lambda,\alpha) - g(0,\alpha)}{\lambda} : \alpha \in A(0) \right\}.$$

Тому

$$g'(0) \leq \max_{\alpha} \left\{ \frac{g(\lambda,\alpha) - g(0,\alpha)}{\lambda} : \alpha \in A(0) \right\}. \tag{2.7.29}$$

Нехай $\lambda_k \to 0$ і $\alpha_k \in A(0)$ такі, що

$$\frac{g(\lambda_k,\alpha_k) - g(0,\alpha_k)}{\lambda_k} =$$
$$= \max_{\alpha} \left\{ \frac{g(\lambda_k,\alpha) - g(0,\alpha)}{\lambda_k} : \alpha \in A(0) \right\}. \tag{2.7.30}$$

Можна вважати,що $\alpha_k \to \alpha_0 \in A(0)$. За лемою 2.1.10

$$\frac{g(0,\alpha_k) - g(-\delta,\alpha_k)}{\delta} \leq \frac{g(\lambda_k,\alpha_k) - g(0,\alpha_k)}{\lambda_k} \leq \frac{g(\delta,\alpha_k) - g(0,\alpha_k)}{\delta},$$

Оскільки функція $g(\lambda,\alpha)$ неперервна за α і множина A компактна, то відношення

$$\frac{g(\lambda_k,\alpha_k) - g(0,\alpha_k)}{\lambda_k}$$

обмежене. Можна вважати, що воно прямує до деякої величини μ.

Фіксуємо $\lambda > 0$. Тоді $\lambda_k < \lambda$ при великих k і за лемою 2.1.10

$$\frac{g(\lambda,\alpha_k) - g(0,\alpha_k)}{\lambda} \geq \frac{g(\lambda_k,\alpha_k) - g(0,\alpha_k)}{\lambda_k}.$$

Переходячи до границі при $k \to \infty$ отримаємо, що

$$\frac{g(\lambda,\alpha_0) - g(0,\alpha_0)}{\lambda} \geq \mu, \tag{2.7.31}$$

звідки при $\lambda \to 0$ матимемо $g'(0,\alpha_0) \geq \mu$. Скористаємося нерівностями (2.7.29), (2.7.30). Отримаємо

$$g'(0) \leq \mu \leq g'(0,\alpha_0), \quad \alpha_0 \in A(0).$$

Аналіз всіх нерівностей приводить до висновку, що

$$g'(0,\alpha_0) \geq g'(0) \geq \sup_{\alpha_0} \{g'(0,\alpha_0) \colon \alpha_0 \in A(0)\}, \quad \alpha_0 \in A(0),$$

тобто

$$g'(0) = \max_{\alpha_0} \{g'(0,\alpha_0) \colon \alpha_0 \in A(0)\}.$$

$\square$

Теорема 2.7.14. *Нехай множина A компактна і функція $f(x,\alpha)$ опукла за x при кожному $\alpha \in A$ і неперервна за x та α для x з деякого околу точки x_0 та $\alpha \in A$. Тоді*

$$\partial f(x_0) = \overline{\mathrm{conv}} \left(\bigcup_{\alpha \in A(x_0)} \partial f(x_0,\alpha) \right).$$

Доведення. За теоремою 2.7.5

$$f'(x_0,p,\alpha) = \max_{x^*} \{\langle p,x^* \rangle : x^* \in \partial f(x_0,\alpha)\}.$$

З попередньої теореми випливає, що

$$f'(x_0,p) = \max_{\alpha \in A(x_0)} \max_{x^* \in \partial f(x_0,\alpha)} \langle p,x^* \rangle =$$

$$= \max_{x^*} \left\{ \langle p,x^* \rangle : x^* \in \bigcup_{\alpha \in A(x_0)} \partial f(x_0,\alpha) \right\} =$$

$$= \sup_{x^*} \left\{ \langle p,x^* \rangle : x^* \in \overline{\mathrm{conv}} \bigcup_{\alpha \in A(x_0)} \partial f(x_0,\alpha) \right\}. \tag{2.7.32}$$

З теореми 2.7.5 випливає, що

$$f'(x_0,p,\alpha) = \sup_{x^*} \{\langle p,x^* \rangle : x^* \in \partial f(x_0,\alpha)\}. \tag{2.7.33}$$

Співвідношення (2.7.32), (2.7.33) показують, що опорні функції двох опуклих замкнутих множин співпадають. Тому самі множини теж співпадають. Теорема доведена. $\square$

Теорему корисно переформулювати для максимуму скінченної кількості опуклих функцій.

Теорема 2.7.15. Теорема Дубовицького-—Мілютіна про субдиференціал максимума. *Нехай функції $f_1(x),\dots,f_m(x)$ опуклі і неперервні в точці x_0. Тоді субдиференціал функції $f(x) = \max\limits_{i=1,\dots,m} f_i(x)$ має вигляд*

$$\partial f(x_0) = \mathrm{conv} \left\{ \bigcup_{i \in I(x_0)} \partial f_i(x_0) \right\}, \tag{2.7.34}$$

де $I(x_0) = \{i \mid f_i(x_0) = f(x_0)\}$.

Використовуючи опуклість множин $\partial f_i(x_0)$, неважко показати, що формулу (2.7.34) можна подати у вигляді

$$\partial f(x_0) = \left\{ x^* \colon x^* = \sum_{i \in I(x)} \lambda_i x_i^*, x_i^* \in \partial f_i(x_0), \lambda_i \geq 0, \sum_{i \in I(x)} \lambda_i = 1 \right\} \tag{2.7.35}$$

Зокрема, якщо функції $f_1(x), \ldots, f_m(x)$ диференційовні, то

$$\partial f(x_0) = \left\{ x^* \colon x^* = \sum_{i \in I(x)} \lambda_i f_i'(x_0), \lambda_i \geq 0, \sum_{i \in I(x)} \lambda_i = 1 \right\}. \tag{2.7.36}$$

Теорема 2.7.16. *Нехай $\Lambda \colon X \to Y$ – неперервний лінійний оператор, f - опукла функція на Y. Тоді для кожної точки $x \in X$ справджується включення*

$$\Lambda^* \partial f(\Lambda x) \subset \partial (f\Lambda)(x).$$

Якщо функція f неперервна в деякій точці $y_0 \in Im\,\Lambda$, то для кожної точки $x \in X$ справджується рівність

$$\Lambda^* \partial f(\Lambda x) = \partial (f\Lambda)(x). \tag{2.7.37}$$

Теорема 2.7.17. *Нехай $g_1, \ldots, g_m$ – опуклі функції на X, $g = (g_1, \ldots, g_m)$ – утворена з них вектор-функція, φ – монотонно неспадна опукла функція на відкритій опуклій множині $U \subset \mathbb{R}^m$, $g(X) \subset U$. Якщо функція $g(x)$ неперервна в точці x_0, а функція φ неперервна в точці $g(x_0)$, то субдиференціал функції $f(x_0) = \varphi(g(x_0))$ має вигляд*

$$\partial f(x_0) = \bigcup_{p \in \partial\varphi(u)} \left(\sum_{i=1}^{m} p_i \partial g_i(x_0) \right), \tag{2.7.38}$$

де $u = g(x_0)$.

Зокрема, якщо функція φ диференційовна в точці $u = g(x_0)$, то формула (2.7.38) матиме вигляд

$$\partial f(x_0) = \sum_{i=1}^{m} \frac{\partial \varphi}{\partial u_i}(g(x_0)) \partial g_i(x_0). \tag{2.7.39}$$

Якщо, крім того, функції $g_1, \ldots, g_m$ диференційовні в точці x_0, то отримаємо відому формулу для градієнта суперпозиції диференційовних функцій:

$$f'(x_0) = \sum_{i=1}^{m} \frac{\partial \varphi}{\partial u_i}(g(x_0)) g_i'(x_0).$$

Теорема 2.7.18. *Нехай φ – опукла функція на відкритій опуклій множині $U \subset \mathbb{R}^m$,A – матриця розміру $m \times n$, $b \in \mathbb{R}^m$, причому множина $X = \{x \in \mathbb{R}^n \mid Ax + b \in U\}$ непорожня. Тоді субдиференціал функції $f(x) = \varphi(Ax + b)$ має вигляд*

$$\partial f(x) = \partial\varphi(u)A \stackrel{def}{=} \{a \in \mathbb{R}^n \mid a = pA, p \in \partial\varphi(u)\}, u = Ax + b.$$

За допомогою теореми Каратеодорі 1.2.1 можна довести посилений варіант теореми 2.7.14 для простору $X = \mathbb{R}^n$.

Теорема 2.7.19. Теорема про очистку. *Якщо за умов теореми 2.7.14 $X = \mathbb{R}^n$, то кожний елемент $x^* \in \partial f(x_0)$ можна представити у вигляді*

$$x^* = \lambda_1 x_1^* + \cdots + \lambda_r x_r^*, r \leq n + 1,$$

$$\sum_{i=1}^{r} \lambda_i = 1, \lambda_i > 0, x_i^* \in \partial f(x_0,\alpha_i), \alpha_i \in A(x_0), i = 1, \ldots, r.$$

Доведення. Позначимо через

$$P = \bigcup_{\alpha \in A(x_0)} \partial f(x_0,\alpha).$$

Покажемо, що множина P обмежена і замкнута. Тоді $\overline{\operatorname{conv}} P = \operatorname{conv} P$ і твердження теореми випливатиме з теореми Каратеодорі. За лемою 2.7.5 функція $f(x)$ неперервна. За теоремою 2.7.5 субдиференціал $\partial f(x_0)$ неперервної функції обмежений. Оскільки субдиференціал $\partial f(x_0)$ включає множину P, то вона обмежена.

Покажемо, що множина P замкнута. Нехай послідовність елементів $z_1, z_2, \ldots$ множини P збігається до z:

$$z_k \in \partial f(x_0,\alpha_k), \quad \alpha_k \in A(x_0).$$

Оскільки множина $A(x_0)$ компактна, то послідовність α_k має граничну точку $\alpha_0 \in A(x_0)$. Функція неперервна за α. Тому для всіх $x \in \mathbb{R}^n$

$$f(x,\alpha_0) - f(x_0,\alpha_0) = f(x,\alpha_0) - f(x_0) \geq \overline{\lim}_{k\to\infty} f(x,\alpha_k) - f(x_0) =$$

$$= \overline{\lim}_{k\to\infty} [f(x,\alpha_k) - f(x_0,\alpha_k)] \geq \lim_{k\to\infty} \langle z_k, x - x_0\rangle = \langle z, x - x_0\rangle.$$

Отже $z \in \partial f(x_0,\alpha_0) \subset P$. Теорема доведена. □

2.7.1. Субдиференціал відстані до множини

Нехай M - опукла множина, C - опукла, обмежена множина, $0 \in \operatorname{int} C$. Тоді для всіх $x \in X$ визначена функція

$$d_C(x|M) = \inf_{\rho}\{\rho \geq 0 : x \in M + \rho C\}. \qquad (2.7.40)$$

Якщо $C = B$, де B - одинична куля з центром в нулі, то $d_B(x|M)$ – це відстань від точки x до множини M. Якщо $M = \{0\}$, то

$$d_C(x|\{0\}) = \inf_{\rho}\{(\rho > 0\colon x \in \rho C\} = \inf_{\rho}\{(\rho > 0\colon \frac{x}{\rho} \in C\} = \mu(x|C),$$

де $\mu(x|C)$ – функція Мінковського.

Теорема 2.7.20. *Справедлива формула*

$$d_C(x|M) = \sup_{x^*}\{(x^*,x) - s(x^*|M)\colon s(x^*|C) \le 1\},$$

де $s(x^*|C)$ *– опорна функція множини* C.

Доведення. Оскільки $0 \in \operatorname{int} C$, то множини $\overline{M + \rho C,\rho} \ge 0$, вкладені одна в одну. Можна переконатися, що якщо $x \in \overline{(M + \rho C)}$, то $d_C(x|M) \le \rho$. За теоремою 1.4.5 про опорну функцію, $x \in \overline{(M + \rho C)}$ тоді і лише тоді, коли

$$\langle x^*,x\rangle \le s(x^*|M + \rho C) = s(x^*|M) + \rho s(x^*|C) \tag{2.7.41}$$

при всіх x^*. Якщо $x \in \overline{M}$, то $d_C(x|M) = 0$ і

$$\langle x^*,x\rangle \le s(x^*|M).$$

Тому

$$\sup_{x^*}\{\langle x^*,x\rangle - s(x^*|M)\colon s(x^*|C) \le 1\} = 0,$$

причому точна верхня грань в лівій частині останнього співвідношення досягається при $x^* = 0$. Таким чином, якщо $d_C(x|M) = 0$, то твердження теореми доведене.

Нехай тепер $x \in \overline{(M + \rho C)}$ і $\rho \ge d_C(x|M) > 0$. При $x^* \ne 0$ маємо $s(x^*|C) > 0$, оскільки $0 \in \operatorname{int} C$. Нерівність (2.7.41) можна переписати в такому вигляді:

$$\langle x^*,x\rangle \le s(x^*|M) + \rho, \quad s(x^*|C) = 1.$$

Тому

$$\rho \ge \sup_{x^*}\{\langle x^*,x\rangle - s(x^*|M)\colon s(x^*|C) = 1\}. \tag{2.7.42}$$

Нехай ρ задовольняє нерівність (2.7.42). Тоді $x \in \overline{(M + \rho C)}$. Покажемо, що $\rho \ge d_C(x|M)$. Оскільки $0 \in \operatorname{int} C$, то для довільного $\varepsilon > 0$ виконано співвідношення $(x - \varepsilon C) \cap (M + \rho C) \ne \emptyset$, тобто $x \in (M + (\rho + \varepsilon)C)$, а тому $d_C(x|M) \le \rho + \varepsilon$. Оскільки ε - довільне, то $d_C(x|M) \le \rho$. Якщо ж ρ не задовольняє нерівність (2.7.42), то $(M + \rho C) \cap (x - \varepsilon_0 C) = \emptyset$, при деякому $\varepsilon_0 > 0$ і x не може належати $M + (\rho + \varepsilon_0)C$, тобто $\rho + \varepsilon_0 \le d_C(x|M)$. Звідси випливає, що $d_C(x|M)$ співпадає з правою частиною нерівності (2.7.42). □

З теореми випливає, що функція $d_C(x|M)$ опукла і обмежена в довільній обмеженій області і тому неперервна.
Обчислимо субдиференціал функції $d_C(x|M)$.

Теорема 2.7.21. *Справедливі формули:*

$$\partial d_C(x|M) = \partial\delta(x|\overline{M}) \cap \{x^*\colon s(x^*|C) \le 1\}$$

при $d_C(x|M) = 0$.

$$\partial d_C(x|M) = \partial\delta(x|\overline{(M + d_C(x|M)C)}) \cap \{x^*\colon s(x^*|C) = 1\}$$

при $d_C(x|M) > 0$, *де* $\delta(x^*|C)$ – *індикаторна функція множини* C.

Доведення. За означенням $x^* \in \partial d_C(x_0|M)$ тоді і лише тоді, коли

$$d_C(x|M) - d_C(x_0|M) \ge \langle x^*, x - x_0\rangle \qquad (2.7.43)$$

при всіх x. Будь-яка точка x може бути представлена у вигляді $y + \rho z$ при деяких $y \in M, z \in C, \rho \ge 0$, причому $d_C(x|M)$ є нижня грань всіх таких ρ. Тому нерівність (2.7.43) еквівалентна наступній нерівності:

$$\rho - d_C(x_0|M) \ge s(x^*|M) + \rho s(x^*|C) - \langle x^*, x_0\rangle \qquad (2.7.44)$$

Остання нерівність справедлива для будь-яких $\rho \ge 0$ лише тоді, коли:

$$s(x^*|C) \le 1. \qquad (2.7.45)$$

Тоді її можна переписати в еквівалентній формі:

$$-d_C(x_0|M) \ge s(x^*|M) - \langle x^*, x_0\rangle. \qquad (2.7.46)$$

Якщо $d_C(x_0|M) = 0$, то $x_0 \in \overline{M}$, а нерівність (2.7.46) можна записати у вигляді:

$$0 \ge \langle x^*, x - x_0\rangle, \quad x \in \overline{M},$$

або

$$-x^* \in [\mathrm{cone}\,(\overline{M} - x_0)]^* = -\partial\delta(x_0|\overline{M}).$$

Порівнявши це з формулою (2.7.46) одержуємо твердження теореми для цього випадку.

Нехай тепер $d_c(x_0|M) > 0$. Тоді $x_0 \in \overline{(M + \rho_0 C)}, \rho_0 = d_C(x_0|M)$, і нерівність (2.7.41) справедлива при $x = x_0, \rho = \rho_0$:

$$-d_C(x_0|M)s(x^*|C) \le s(x^*|M) - \langle x^*, x_0\rangle.$$

Віднявши його від нерівності (2.7.46), отримаємо:

$$d_c(x_0|M)(s(x^*|C) - 1) \ge 0.$$

Враховуючи умови (2.7.45), це дає $s(x^*|C) = 1$. Тому нерівність (2.7.46) можна переписати у вигляді

$$0 \ge s(x^*|M) + d_C(x_0|M)s(x^*|C) - \langle x^*, x_0\rangle,$$

або

$$0 \ge \langle x^*, x - x_0\rangle, \quad x \in \overline{(M + d_C(x_0|M)C)},$$

Звідки

$$-x^* \in [\mathrm{cone}(\overline{(M + d_c(x_0|M)C)} - x_0)]^* = -\partial\delta(x_0|\overline{(M + d_c(x_0|M)C)}).$$

Теорему доведено. □

2.7.2. Конус допустимих напрямків та субдиференціали

Конус допустимих напрямків

$$\operatorname{cone}(M - x_0) = \{p\colon p = \lambda(x - x_0), \lambda > 0, x \in M\}$$

визначений для опуклої множини M і точки $x_0 \in M$. В задачах на екстремум такі конуси зустрічаються постійно і обчислення спряжених до таких конусів є однією із основних задач. Покажемо як можна обчислити спряжений конус $[\operatorname{cone}(M - x_0)]^*$ у тому випадку, коли множина M задається за допомогою нерівності для опуклої функції.

Теорема 2.7.22. *Нехай $f(x)$ – опукла функція, нехай множина M задається за допомогою нерівності*

$$M = \{x\colon f(x) \le 0\},$$

та нехай існує точка $x_1 \in M$ така, що $f(x_1) < 0$. Якщо $f(x_0) = 0$ і $f'(x_0, p)$ – замкнута функція, то

$$[\operatorname{cone}(M - x_0)]^* = -\operatorname{cone}\partial f(x_0).$$

Доведення. Оскільки

$$0 > f(x_1) = f(x_1) - f(x_0) \ge \langle x^*, x_1 - x_0\rangle, \quad x^* \in \partial f(x_0),$$

то $0 \notin \partial f(x_0)$. Покажемо, що

$$\operatorname{cone}\partial f(x_0) = \{x^*: x^* = \lambda x_0^*,\ \lambda \ge 0,\ x_0^* \in \partial f(x_0)\}$$

є замкнута множина. Дійсно, нехай послідовності λ_k, x_k^* такі, що $\lambda_k x_k^* \to x^*$, $\lambda_k > 0$, $x_k^* \in \partial f(x_0)$, $x^* \ne 0$. Тоді послідовність λ_k обмежена. Якщо припустити протилежне, тобто $\lambda_k \to +\infty$, то $\lambda_k \|x_k^*\| \to \|x^*\|$. Це можливо лише при $\|x_k^*\| \to 0$. А цього не може бути, оскільки $0 \notin \partial f(x_0)$. В силу обмеженості послідовності λ_k можна вважати, що $\lambda_k \to \lambda_0 > 0$. Але тоді послідовність x_k^* збігається до $\lambda_0^{-1}x^*$. А так як множина $\partial f(x_0)$ замкнута, то $\lambda_0^{-1}x^* \in \partial f(x_0)$. Отже $x^* \in \operatorname{cone}\partial f(x_0)$, що доводить замкнутість $\operatorname{cone}\partial f(x_0)$.

За означенням $p \in [-\operatorname{cone}\partial f(x_0)]^*$ тоді і тільки тоді, коли

$$\langle p, -\lambda x_0^*\rangle \ge 0, \quad \lambda > 0, \quad x_0^* \in \partial f(x_0),$$

тобто

$$\sup_{x_0^*}\{\langle x_0^*, p\rangle : x_0^* \in \partial f(x_0)\} \le 0. \tag{2.7.47}$$

Оскільки за умов теореми $f'(x_0, p)$ – замкнута функція, то скориставшись теоремою 2.7.1 і нерівністю (2.7.47), можна переконатися у справедливості рівності

$$[-\operatorname{cone}\partial f(x_0)]^* = \{p\colon f'(x_0, p) \le 0\}.$$

Покладемо $p_1 = x_1 - x_0$. Тоді для $0 < \lambda < 1$

$$f'(x_0, p_1) \le \frac{f(x_0 + \lambda p_1) - f(x_0)}{\lambda} \le f(x_1) - f(x_0) < 0$$

і в силу опуклості $f'(x_0,p)$ за p

$$f'(x_0,\lambda p_1+(1-\lambda)p)\le\lambda f'(x_0,p_1)+(1-\lambda)f'(x_0,p)<0 \qquad (2.7.48)$$

якщо $f'(x_0,p)\le 0$.
Нехай тепер $p=\lambda(x-x_0),x\in M,\lambda>0$, тобто $p\in\operatorname{cone}(M-x_0)$. Тоді $f(x)\le 0$ і

$$0\ge f(x)-f(x_0)\ge\langle x^*,x-x_0\rangle,\quad x^*\in\partial f(x_0).$$

Тому

$$\sup_{x^*}\{\langle x^*,p\rangle\colon x^*\in\partial f(x_0)\}=\lambda\sup_{x^*}\{\langle x^*,x-x_0\rangle\colon x^*\in\partial f(x_0)\}\le 0.$$

Звідси випливає, що

$$\operatorname{cone}(M-x_0)\subset\{p\colon f'(x_0,p)\le 0\}.$$

З іншого боку, якщо $f'(x_0,p)<0$, то для малих $\lambda>0$

$$f(x_0+\lambda p)=f(x_0)+\lambda f'(x_0,p)+o(\lambda)=\lambda f'(x_0,p)+o(\lambda)<0.$$

Тому $x_0+\lambda p\in M$. Це означає, що $p\in\operatorname{cone}(M-x_0)$. Тепер з нерівності (2.7.48) випливає, що

$$\lambda p_1+(1-\lambda)p\in\operatorname{cone}(M-x_0),$$

якщо $p\in\{p\colon f'(x_0,p)\le 0\}$. Спрямовуючи λ до нуля, отримаємо, що будь-яка точка конуса $\{p\colon f'(x_0,p)\le 0\}$ є граничною точкою для конуса $\operatorname{cone}(M-x_0)$. Конус $\{p\colon f'(x_0,p)\le 0\}$ замкнутий, оскільки $f'(x_0,p)$ – замкнута функція і справедлива теорема 2.2.4, а також

$$\overline{(\operatorname{cone}(M-x_0))}\supset\{p\colon f'(x_0,p)\le 0\}.$$

Тому

$$\overline{(\operatorname{cone}(M-x_0))}=\{p\colon f'(x_0,p)\le 0\}.$$

Раніше було показано, що

$$[-\operatorname{cone}\partial f(x_0)]^*=\{p\colon f'(x_0,p)\le 0\}.$$

Тому справджується рівність

$$[-\operatorname{cone}\partial f(x_0)]^*=\overline{(\operatorname{cone}(M-x_0))}.$$

Тепер отримаємо

$$-\operatorname{cone}\partial f(x_0)=[-\operatorname{cone}\partial f(x_0)]^{**}=$$

$$=[\overline{(\operatorname{cone}(M-x_0))}]^*=[(\operatorname{cone}(M-x_0))]^*,$$

що і потрібно було показати. □

Теорема 2.7.23. *Нехай опукла функція $f(x)$ неперервна в точці $x_0\in M$, $M=\{x\colon f(x)\le 0\}$, та нехай існує така точка x_1, що $f(x_1)<0$. Тоді*

$$[\operatorname{cone}(M-x_0)]^*=\begin{cases}\{0\}, & f(x_0)<0,\\ -\operatorname{cone}\partial f(x_0), & f(x_0)=0.\end{cases}$$

Доведення. Якщо $f(x_0) < 0$, то в силу неперервності f в точці x_0 отримаємо, що $x_0 + \lambda p \in M$ для кожного p при малих $\lambda > 0$. Тому $\mathrm{cone}\,(M - x_0) = X$. Тоді спряжений до $\mathrm{cone}\,(M - x_0)$ конус містить лише одну точку – початок координат. Це доводить справедливість першого твердження теореми. Друге твердження теореми є наслідком теорем 2.7.5 та 2.7.23. □

Теорема 2.7.24. *Нехай $f(x)$ – власна опукла функція та нехай існують такі точки x_0, x_1, що $f(x_1) < 0$, $x_0 \in \mathrm{ri}\,\mathrm{dom}\, f$ та $x_0 \in M = \{x\colon f(x) \le 0\}$. Тоді*

$$[\mathrm{cone}\,(M - x_0)]^* = \begin{cases} (\mathrm{Lin}\ \mathrm{dom}\ f)^{\perp}, & f(x_0) < 0, \\ -\,\mathrm{cone}\ \partial f(x_0), & f(x_0) = 0. \end{cases}$$

Доведення. Друге твердження теореми є наслідком теорем 2.7.6 та 2.7.23. Перше твердження теореми отримаємо, якщо помітимо, що в силу теореми 2.2.3 для $x_0 \in \mathrm{ri}\ \mathrm{dom}\ f$ виконується співвідношення

$$\mathrm{cone}\,(M - x_0) = \mathrm{Lin}\ \mathrm{dom}\ f. \tag{2.7.49}$$

Тому

$$[\mathrm{cone}\,(M - x_0)]^* = [\mathrm{Lin}\ \mathrm{dom}\ f]^*.$$

Якщо конус $K = L$, де L – підпростір X, то $x^* \in X^*$ тоді і тільки тоді, коли

$$\langle x, x^* \rangle \ge 0, \quad x \in L.$$

Якщо ж $x \in L$, то $-x \in L$. Отже

$$\langle -x, x^* \rangle \ge 0, \quad x \in L.$$

Тому

$$\langle x, x^* \rangle = 0, \quad x \in L,$$

Тобто $x^* \in L^{\perp}$. Отже, якщо конус $K = L$, то $K^* = L^{\perp}$.
Застосовуючи це твердження до рівності (2.7.49), отримаємо твердження теореми. □

2.8. ОПЕРАЦІЇ НАД ОПУКЛИМИ ОБ'ЄКТАМИ

2.8.1. Операції над опуклими функціями

На множині опуклих функцій $\mathrm{Conv}\,(X,\overline{\mathbb{R}})$ визначені такі операції:
1) сума

$$(f_1 + f_2)(x) = f_1(x) + f_2(x);$$

2) конволюція

$$(f_1 \oplus f_2)(x) = \inf_{x=x_1+x_2} (f_1(x_1) + f_2(x_2))\,;$$

3) максимум

$$(f_1 \vee f_2)(x) = \max\{f_1(x), f_2(x)\};$$

4) опукла оболонка мінімуму

$$(f_1 \operatorname{conv} \wedge f_2)(x) = \operatorname{conv}(\min\{f_1(\cdot), f_2(\cdot)\})(x);$$

5) прообраз функції при лінійному відображенні $\Lambda\colon X \to Y$

$$(f\Lambda)(x) = f(\Lambda x);$$

6) образ функції при лінійному відображенні $\Lambda\colon X \to Y$

$$(\Lambda f)(y) = \inf\{f(x)\colon x \in X, \Lambda x = y\}.$$

Всі перераховані операції переводять опуклі функції в опуклі функції. Операції 1), 2) допускають природне розширення на скінченну кількість функцій. Операції 3), 4) допускають природне розширення на будь-яку кількість функцій:

$$(\vee_{\alpha \in A} f_\alpha)(x) = \sup_{\alpha \in A} f_\alpha(x),$$

$$\operatorname{conv}(\wedge_{\alpha \in A} f_\alpha)(x) = \operatorname{conv}\left(\inf_{\alpha \in A} f_\alpha(x)\right).$$

На множині опуклих однорідних функцій $\mathrm{SL}(X, \overline{\mathbb{R}})$ визначені такі ж операції 1)—6). Крім того, визначають ще одну операцію, яка переводить опуклі однорідні функції в опуклу однорідну функцію

$$p_1 \nabla p_2 = \vee\{\alpha_1 p_1 + \alpha_2 p_2\colon \alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_1 + \alpha_2 = 1\}.$$

2.8.2. Операції над опуклими множинами

На множині $\operatorname{Conv}(X)$ опуклих підмножин простору X визначені такі операції:

1) сума

$$A_1 + A_2 = \{x\colon x = x_1 + x_2, x_1 \in A_1, x_2 \in A_2\};$$

2) конволюція, або сума Келлі

$$A_1 \boxplus A_2 = \cup\{\alpha_1 A_1 \cap \alpha_2 A_2\colon \alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_1 + \alpha_2 = 1\};$$

3) опукла оболонка об'єднання

$$A_1 \operatorname{conv} \cup A_2 = \operatorname{conv}(A_1 \cup A_2) =$$

$$= \{x\colon x = \alpha_1 x_1 + \alpha_2 x_2, x_1 \in A_1, x_2 \in A_2, \alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_1 + \alpha_2 = 1\};$$

4) перетин

$$A_1 \cap A_2 = \{x\colon x \in A_1, x \in A_2\};$$

5) прообраз при лінійному відображенні

$$A\Lambda = \Lambda^{-1}A;$$

6) образ при лінійному відображенні

$$\Lambda A.$$

Всі перераховані операції переводять опуклі множини в опуклі множини. Операції 1), 2) допускають природне розширення на скінченну кількість множин. Операції 3), 4) допускають природне розширення на будь-яку кількість множин.

На множині $\mathrm{Cone}\,(X)$ конусів, множині $\mathrm{Aff}\,(X)$ афінних підмножин простору X, множині $\mathrm{Lin}\,(X)$ визначені такі ж операції як і на множині $\mathrm{Conv}\,(X)$. Проте необхідно приймати до уваги, що для цих множин конволюція (сума Келлі) – це перетин множин: $A_1 \boxplus A_2 = A_1 \cap A_2$, а опукла оболонка об'єднання – це сума множин: $A_1 \operatorname{conv} \cup A_2 = A_1 + A_2$.

2.8.3. Основні оператори опуклого аналізу

Нехай $X, Y = X^*$ простори, f– функція на X, A – підмножина в X, K – конус в X, p – однорідна функція на X. Визначимо найважливіші оператори опуклого аналізу.

Функція

$$lf(y) = \sup_x(\langle x,y\rangle - f(x))$$

називається *перетворенням Юнга-Фенхеля* функції f.

Множина

$$\pi A = \{y \in Y \colon \langle x,y\rangle \leq 1, \forall x \in A\}$$

називається *полярою* множини A.

Поляра підпростору L називається анулятором підпростору L і позначається $L^{\perp}$.

Конус

$$\sigma K = \{y \in Y \colon \langle x,y\rangle \geq 0, \forall x \in K\}$$

називається *спряженим* до конуса K.

Множина

$$\partial p = \{y \in Y \colon \langle x,y\rangle \leq p(x), \forall x \in X\}$$

називається *субдиференціалом* однорідної функції p.

Функцію

$$sA(y) = \sup_{x\in A}\langle x,y\rangle$$

називають *опорною функцією* множини A.

Множина

$$\partial f(\hat{x}) = \{y \in Y \colon f(x) - f(\hat{x}) \geq \langle x - \hat{x},y\rangle, \forall x \in X\}$$

називається *субдиференціалом функції* f в точці $\hat{x}$.
Аналогічно визначаються всі перераховані оператори для функції g на Y, підмножини B, конуса L, опуклої однорідної функції q на Y:

$$lg(x) = \sup_{y}(\langle x,y\rangle - g(y)),$$

$$\pi B = \{x \in X \colon \langle x,y\rangle \le 1, \forall y \in B\},$$

$$\sigma L = \{x \in X \colon \langle x,y\rangle \ge 0, \forall y \in L\},$$

$$\partial q = \{x \in X \colon \langle x,y\rangle \le q(y), \forall y \in Y\},$$

$$sB(x) = \sup_{y \in B}\langle x,y\rangle,$$

$$\partial g(\hat{y}) = \{x \in X \colon g(y) - g(\hat{y}) \ge \langle y - \hat{y},x\rangle, \forall y \in Y\}.$$

Для визначених операторів використовують і традиційні їх позначення, наприклад $lf(x^*) = f^*(x^*)$. Запропоновані позначення, однак, спрощують формули опуклого аналізу. Далі будемо користуватися наступними позначеннями: $l^2 f = l(lf)$, $\pi^2 A = \pi(\pi A)$, $\sigma^2 K = \sigma(\sigma K)$, $\partial s A = \partial(sA)$, $s\partial p = s(\partial p)$.

Теорема 2.8.1. Теорема про інволютивності.

- *Для того щоб*
  $$L^{\perp\perp} = L,$$
  необхідно і достатньо, щоб підпростір був замкнутим:
  $$L \in \mathrm{Lin}\,(X) \Rightarrow L^{\perp\perp} = L \Leftrightarrow L \in \mathrm{Lin}\,(X) \cap \mathrm{Cl}\,(X).$$
- *Для того щоб*
  $$\pi^2 A = A,$$
  необхідно і достатньо, щоб множина A була замкнута, опукла, і містила нуль:
  $$\pi^2 A = A \Leftrightarrow A \in \mathrm{Conv}\,(X) \cap \mathrm{Cl}\,(X), \quad 0 \in A.$$
- *Для того щоб*
  $$\sigma^2 K = K,$$
  необхідно і достатньо, щоб конус K був опуклий і замкнутий:
  $$\sigma^2 K = K, \Leftrightarrow K \in \mathrm{Cone}\,(X) \cap \mathrm{Cl}\,(X).$$
- *Для того щоб*
  $$l^2 f = f,$$
  необхідно і достатньо, щоб власна функція f була опукла і замкнута:
  $$l^2 f = f \Leftrightarrow f \in \mathrm{ClConv}\,(X,\overline{\mathbb{R}}).$$

Теорема 2.8.2. Про обернені оператори.

- *Для того щоб при* $\partial p \neq \emptyset$
$$s\partial p = p,$$
необхідно і достатньо, щоб однорідна функція p *була опуклою і замкнутою:*
$$s\partial p = p \Leftrightarrow p \in \mathrm{Cl}\,\mathrm{SL}\,(X,\overline{\mathbb{R}}).$$

- *Для того щоб*
$$\partial sA = A,$$
необхідно і достатньо, щоб множина A *була опукла і замкнута:*
$$\partial sA = A, \Leftrightarrow A \in \mathrm{Cl}\,(X) \cap \mathrm{Conv}\,(X).$$

Теорема 2.8.3. *Про перетворення Юнга–Фенхеля.*

1. $l(f_1 \oplus f_2) = lf_1 + lf_2$.
2. $l(f_1 + f_2) \approxeq lf_1 \oplus lf_2$.
3. $l(f_1 \operatorname{conv} \wedge f_2) = lf_1 \vee lf_2$.
4. $l(f_1 \vee f_2) \approxeq lf_1 \operatorname{conv} \wedge lf_2$.
5. $l(\Lambda f) = lf\Lambda^*$.
6. $l(f\Lambda) \approxeq \Lambda^* lf$.
7. $l^2 f \leq f$.
8. $f \leq g \Rightarrow lf \geq lg, \quad l^2 f \leq l^2 g$.

Для того щоб була рівність в 2, достатньо, щоб існувала точка $x_0 \in \mathrm{dom}\, f_2$*, в якій* f_1 *неперервна. Для того щоб була рівність в 4, достатньо, щоб функції* f_1 *та* f_2 *були скінченні на* X *і функція* f_1 *була неперервна. Для того щоб була рівність в 6, достатньо, щоб існувала точка* $z_0 \in Im\,\Lambda$*, в якій* f_1 *неперервна.*

2.8.4. Деякі співвідношення між основними операторами

Нехай A,A_1,A_2 – опуклі множини, K – опуклий конус, p – опукла однорідна функція. Справджуються наступні рівності:

1. $l\delta A = sA$.
2. $l\mu A = \delta\pi A$.
3. $0 \in A \Rightarrow \mu\pi A = sA$.
4. $lp = \delta\partial p$.
5. $\sigma K = -\pi K$.
6. $sK = \delta\sigma K$.
7. $\delta\pi K = sK$.
8. $\delta(A_1 + A_2) = \delta A_1 \oplus \delta A_2$.
9. $\delta(A_1 \cap A_2) = \delta A_1 \vee \delta A_2 = \delta A_1 + \delta A_2$.

10. $\delta(A_1 \operatorname{conv} \cup A_2) = \delta A_1 \operatorname{conv} \wedge \delta A_2$.
11. $0 \in (A_1 \cap A_2) \Rightarrow \mu A_1 + \mu A_2 = \mu(A_1 \boxplus A_2)$.
12. $\mu(A_1 \cap A_2) = \mu A_1 \vee \mu A_2$.
13. $\mu(A_1 \operatorname{conv} \cap A_2) = \mu A_1 \operatorname{conv} \wedge \mu A_2$.

Формули для субдиференціалів однорідних функцій:

1. $\partial(p_1 + p_2) \approxeq \partial p_1 + \partial p_2$.
2. $\partial(p_1 \triangledown p_2) = \partial p_1 \boxplus \partial p_2$.
3. $\partial(p_1 \vee p_2) \approxeq \partial p_1 \operatorname{conv} \cup \partial p_2$.
4. $\partial(p_1 \operatorname{conv} \wedge p_2) = \partial p_1 \cap \partial p_2$.

Формули для поляр:

1. $\pi(A_1 + A_2) = \pi A_1 \boxplus \pi A_2$.
2. $\pi(A_1 \boxplus A_2) \approxeq \pi A_1 + \pi A_2$.
3. $\pi(A_1 \cap A_2) \approxeq \pi A_1 \operatorname{conv} \cup \pi A_2$.
4. $\pi(A_1 \operatorname{conv} \cup A_2) = \pi A_1 \cap \pi A_2$.

Формули для опорних функцій:

1. $s(A_1 + A_2) = sA_1 + sA_2$.
2. $s(A_1 \boxplus A_2) \approxeq sA_1 \triangledown sA_2$.
3. $s(A_1 \cap A_2) \approxeq sA_1 \operatorname{conv} \wedge sA_2 \approxeq sA_1 \oplus sA_2$.
4. $s(A_1 \operatorname{conv} \cup A_2) = sA_1 \vee sA_2$.

Формули для ануляторів;

1. $(L_1 + L_2)^{\perp} = L_1^{\perp} \cap L_2^{\perp}$.
2. $(L_1 \cap L_2)^{\perp} \approxeq L_1^{\perp} + L_2^{\perp}$.

Формули для спряжених конусів:

1. $\sigma(K_1 + K_2) = \sigma K_1 \cap \sigma K_2 \Leftrightarrow (K_1 + K_2)^* = K_1^* \cap K_2^*$.
2. $\sigma(K_1 \cap K_2) \approxeq \sigma K_1 + \sigma K_2 \Leftrightarrow (K_1 \cap K_2)^* \approxeq K_1^* + K_2^*$.

Властивості опорних функцій:

1. $sA(\lambda y) = \lambda sA(y), \lambda \geq 0$.
2. $sA(y_1 + y_2) \leq sA(y_1) + sA(y_2)$.
3. $s(A_1 + A_2)(y) = sA_1(y) + sA_2(y)$.
4. $X = \mathbb{R}^n$, $\Lambda \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, $s\Lambda A(y) = sA(\Lambda^* y)$.
5. $s(\lambda A)(y) = \lambda sA(y), \lambda \geq 0$.
6. $X = \mathbb{R}^n$, $\operatorname{conv} A = \cap_{y \in S^{n-1}} \{x \colon \langle x, y \rangle \leq sA(y)\}$.
7. $A_1, A_2 \in \mathrm{Comp}\,(X)$, $A_1 = A_2 \Leftrightarrow sA_1 = sA_2$.
8. $\operatorname{conv} A \subset \operatorname{conv} B \Leftrightarrow sA \leq sB$.
9. $A \in \mathrm{Comp}\,(\mathbb{R}^n)$, $x \in \operatorname{int} \operatorname{conv} A \Leftrightarrow \langle x, y \rangle \leq sA(y) \forall y \in S^{n-1}$.

10. $A_1,A_2 \in \mathrm{Comp}\,(\mathbb{R}^n) \Rightarrow \mathrm{conv}\, A_1 \cap \mathrm{conv}\, A_2 \neq \emptyset \Leftrightarrow sA_1(y) + sA_2(-y) \geq 0$.

11. $p \in \mathrm{SL}(\mathbb{R}^n,\mathbb{R}^m) \Leftrightarrow p(y) = sA(y), A = \partial p$.

12. $A_1,A_2 \in \mathrm{Comp}\,(\mathbb{R}^n) \Rightarrow |sA_1(y_1) - sA_2(y_2)| \leq |y_1|h(A_1,A_2) +$ Diam $A_2|y_1 - y_2|$.

13. $A_1,A_2 \in \mathrm{Comp}\,(\mathbb{R}^n) \Rightarrow h(\mathrm{conv}\ A_1, \mathrm{conv}\ A_2) = \max_{y\in S^{n-1}} |sA_1(y) - sA_2(y)|$.

14. $s(A_1 \cup A_2) = \max\{sA_1(y),sA_2(y)\} \Leftrightarrow s(A_1 \cup A_2) = sA_1 \vee sA_2$.

15. $s(A_1 \cap A_2) = \inf_z\{sA_1(y-z) + sA_2(y)\} = (sA_1 \oplus sA_2)(y)$.

16. $\rho(A_1,A_2) = -\min_{y\in S^{n-1}}(sA_1(y) + sA_2(-y))$.

17. $A \in \mathrm{Comp}\,(\mathbb{R}^n) \Rightarrow \mathrm{Diam}\, A = \max_{y\in S^{n-1}}(sA(y) + sA(-y))$.

2.8.5. Приклади

1. Перетворення Юнга–Фенхеля.

а) $X = Y = \mathbb{R},\ f(x) = |x|^p/p, p > 1 \Rightarrow lf(y) = |y|^q/q, 1/p + 1/q = 1$;
б) $X = Y = \mathbb{R},\ f(x) = |x| \Rightarrow lf(y) = \delta[-1,1](y)$;
в) $X = Y = \mathbb{R},\ f(x) = \exp(x) \Rightarrow$

$$lf(y) = \begin{cases} y(\ln(y) - 1), & y > 0, \\ 0, & y = 0, \\ +\infty, & y < 0, \end{cases}$$

г) $X = Y = \mathbb{R}^n,\ f(x) = \sum_{i=1}^n |x_i|^{p_i}/p_i, p_i > 1 \Rightarrow lf(y) = \sum_{i=1}^n |y_i|^{q_i}/q_i$, $q_i > 1, 1/p_i + 1/q_i = 1$;
д) $X = Y = \mathbb{R}^n,\ f(x) = \max\{x_1,\ldots,x_n\} \Rightarrow lf(y) = \delta\left\{y \in \mathbb{R}^n_+ \colon \sum_{i=1}^n y_i = 1\right\}$;
е) $X = Y$ – Гільбертів простір, $lf(x) = f(x) \Leftrightarrow f(x) = \langle x,x\rangle/2$;
ж) X – лінійний нормований простір, X^* – спряжений простір, $f(x) = \|x\| \Rightarrow lf(y^*) = \delta B^*(y^*), B^* = \{x^* \colon \|x^*\| \leq 1\}$.

2. Поляри

а) $X = Y = \mathbb{R}^n,\ A = B^n_p = \{x \in \mathbb{R}^n \colon \sum_{i=1}^n |x_i|^p \leq 1\}, p > 1 \Rightarrow \pi A = B^n_q, 1/p + 1/q = 1$;
б) $X = Y = \mathbb{R}^n,\ A = B^n_\infty(a) = \{x \in \mathbb{R}^n \colon |x_i| \leq a_i, i = 1,\ldots,n\} \Rightarrow \pi A = B^n_1(a^{-1}) = \{y \in \mathbb{R}^n \colon \sum_{i=1}^n |a_i y_i| \leq 1\}$;
в) $X = Y = \mathbb{R}^n,\ A = B^n_1(a) = \{x \in \mathbb{R}^n \colon \sum_{i=1}^n |x_i/a_i| \leq 1\} \Rightarrow \pi A = B^n_\infty(a) = \left\{y \in \mathbb{R}^n \colon |y_i| \leq a_i^{-1}, i = 1,\ldots,n\right\}$;
г) $X = Y$ – Гільбертів простір, $\pi A = A \Leftrightarrow A = \{x \colon \langle x,x\rangle \leq 1\}$;
д) X – лінійний нормований простір, X^* – спряжений простір, $B = \{x \colon \|x\| \leq 1\}$, $B^* = \{x^* \colon \|x^*\| \leq 1\}$, $\pi B = B^*$.

3. Спряжені конуси

а) $X = Y = \mathbb{R}^n$, $K = \mathbb{R}^n_+ \Rightarrow \sigma K = \mathbb{R}^n_+$;

б) $X = Y$ – Гільбертів простір, $K = K_a(\alpha) = \{x\colon \langle x,a\rangle \geq \|x\| \cos(\alpha), \|a\| = 1, 0 < \alpha < \pi/2\} \Rightarrow \sigma K = K_a(\pi/2 - \alpha)$;

в) X – лінійний нормований простір, X^* – спряжений простір, $K = \Pi_0 \xi^* = \{x\colon \langle \xi^*,x\rangle \geq 0\} \Rightarrow \sigma K = \lambda\xi^* = \{x^*\colon x^* = \alpha\xi^*, \alpha \geq 0\}$.

4. Субдиференціали однорідних функцій

а) $X = Y = \mathbb{R}$, $p(x) = p(x; a,b) = \{bx, x \geq 0;\ ax, x \leq 0, a < b\} \Rightarrow \partial p = [a,b]$;

б) $X = Y = \mathbb{R}^n$, $p(x) = \max\{x_1, \ldots, x_n\} \Rightarrow \partial p = \left\{y \in \mathbb{R}^n_+\colon \sum_{i=1}^n y_i = 1\right\}$;

в) X – лінійний нормований простір, X^* – спряжений простір, $p(x) = \|x\| \Rightarrow \partial p = B^*, B^* = \{x^*\colon \|x^*\| \leq 1\}$.

5. Опорні функції

а) $X = Y = \mathbb{R}^n$, $A = B^n_p = \{x \in \mathbb{R}^n\colon \sum_{i=1}^n |x_i|^p \leq 1\}, p > 1 \Rightarrow sA(y) = \|y\|_q = (\sum_{i=1}^n |y_i|^q)^{1/q}, 1/p + 1/q = 1$;

б) $X = Y = \mathbb{R}^n$, $A = B^n_\infty(a) \Rightarrow sA(y) = \sum_{i=1}^n |a_i y_i|$;

в) $X = Y = \mathbb{R}^n$, $A = B^n_1(a) \Rightarrow sA(y) = \max_{i=1,\ldots,n} |a_i y_i|$.

6. Субдиференціали опуклих функцій

а) $X = Y = \mathbb{R}$, $f(x) = |x| \Rightarrow$

$$\partial f(x) = \begin{cases} -1, & x < 0, \\ [-1,1], & x = 0, \\ 1, & x > 0; \end{cases}$$

б) $X = Y = \mathbb{R}$, $f(x) = \max\{e^x, 1 - x\} \Rightarrow$

$$\partial f(x) = \begin{cases} e^x, & x > 0, \\ [-1,1], & x = 0, \\ -1, & x < 0; \end{cases}$$

в) $X = Y$ – Гільбертів простір, $f(x) = \|x\| \Rightarrow$

$$\partial f(x) = \begin{cases} x\|x\|^{-1}, & x \neq 0, \\ y\colon \|y\| \leq 1, & x = 0. \end{cases}$$

2.9. ЗАДАЧІ ОПУКЛОГО ПРОГРАМУВАННЯ

Теорема 2.9.1. *Нехай множина M опукла і функція f опукла на M. Тоді локальний розв'язок задачі на мінімум*

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

є також глобальним розв'язком задачі.

Доведення. Нехай x_0 – локальний розв'язок задачі. Тоді при деякому $\varepsilon > 0$ виконується нерівність

$$f(x_0) \le f(x) \quad \text{при всіх} \quad x \in M \cap B_\varepsilon(x_0).$$

Для будь-якої точки $x \in M, x \neq x_0$, візьмемо $\lambda = \min\{\varepsilon/\|x - x_0\|, 1\}$. Тоді $\lambda x + (1-\lambda)x_0 \in M \cap B_\varepsilon(x_0)$ і

$$f(x_0) \le f(\lambda x + (1-\lambda)x_0) \le \lambda f(x) + (1-\lambda) f(x_0).$$

Звідси

$$f(x_0) \le f(x) \quad \text{при всіх} \quad x \in M.$$

Тобто x_0 - глобальний розв'язок задачі. □

Отже для опуклих задач поняття локального і глобального розв'язків не відрізняються і можна говорити просто про розв'язок задачі.

Іншу важливу властивість опуклих задач можна сформулювати у вигляді такого загального принципу: необхідні умови оптимальності в тому чи іншому класі задач оптимізації при відповідних припущеннях опуклості виявляються і достатніми. Як приклад наведемо таку теорему.

Теорема 2.9.2. *Нехай функція f опукла на $\mathbb{R}^n$ і диференційовна в точці $x_0 \in \mathbb{R}^n$. Якщо $f'(x_0) = 0$, то x_0 – точка мінімуму функції $f(x)$.*

Доведення. Для всіх $x \in \mathbb{R}^n$ та $\lambda \in (0,1]$ маємо

$$f(\lambda x + (1-\lambda)x_0) \le \lambda f(x) + (1-\lambda) f(x_0).$$

Користуючись тим, що функція f диференційовна в точці x_0, отримаємо

$$f(x) - f(x_0) \ge \frac{f(x_0 + \lambda(x - x_0)) - f(x_0)}{\lambda} =$$

$$= \frac{\langle f'(x_0), \lambda(x - x_0)\rangle + o(\lambda)}{\lambda} = \frac{o(\lambda)}{\lambda}.$$

Переходячи до границі при $\lambda \to 0$, матимемо $f(x) \ge f(x_0)$. Тобто x_0 – глобальний розв'язок задачі. □

Наведемо ще одну властивість опуклих задач.

Теорема 2.9.3. *Нехай множина M опукла і функція f опукла на M. Тоді множина розв'язків $\hat{M} = Arg\min_{x \in M} f(x)$ задачі на мінімум опукла. Якщо при цьому функція f строго опукла на M, то розв'язок задачі єдиний.*

Доведення. Нехай $x_1, x_2 \in \hat{M}, \lambda \in [0,1]$. Тоді $f(x_1) = f(x_2) = \hat{f}$. При цьому виконується нерівність

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2) = \hat{f}.$$

За визначенням $\hat{M}$ тут може бути лише рівність. Отже $\lambda x_1 + (1-\lambda)x_2 \in \hat{M}$, тобто $\hat{M}$ - опукла множина.
Нехай f строго опукла. Якщо припустити, що в $\hat{M}$ існують дві різні точки x_1, x_2, $x_1 \neq x_2$, то при $\lambda \in (0,1)$ в останньому співвідношенні нерівність повинна бути строгою, що неможливо. □

2.9.1. Обмеження, що задаються опуклими множинами

Нехай $f(x)$ – власна опукла функція, x_0 – точка мінімуму функції $f(x)$ на всьому просторі X. Тоді справджується нерівність

$$f(x) - f(x_0) \geq 0 \quad \forall x \in X.$$

яку можна переписати у вигляді

$$f(x) - f(x_0) \geq \langle x - x_0, 0\rangle, \quad 0 \in X^*, \quad \forall x \in X. \tag{2.9.1}$$

Дана нерівність означає, що $0 \in \partial f(x_0)$. Отже справедлива наступна теорема.

Теорема 2.9.4. *Нехай $f(x)$ – власна опукла функція. Для того щоб точка x_0 була точкою мінімуму функції $f(x)$ на всьому просторі X, необхідно і достатньо, щоб $0 \in \partial f(x_0)$.*

Ця проста умова разом із доведеними раніше теоремами дозволяє отримати цілий ряд результатів.

Розглянемо задачу мінімізації функції $f(x)$ на множині M:

$$f(x) \to \min, \quad x \in M. \tag{2.9.2}$$

Теорема 2.9.5. *Нехай $f(x)$ – власна опукла функція. Нехай M – опукла множина та нехай існує точка $x_1 \in M$, в якій функція $f(x)$ неперервна. Для того щоб точка x_0 була точкою мінімуму функції $f(x)$ на множині M, необхідно і достатньо, щоб*

$$\partial f(x_0) \cap K_M^*(x_0) \neq \emptyset, \tag{2.9.3}$$

$$K_M(x_0) = \operatorname{cone}(M - x_0) = \{p\colon p = \lambda(x - x_0), \lambda > 0, x \in M\}.$$

Доведення. Оскільки за означенням

$$\delta(x|M) = \begin{cases} 0, & x \in M, \\ +\infty, & x \notin M, \end{cases}$$

то можна побачити, що задача знаходження мінімуму функції $f(x)$ на множині M еквівалентна задачі знаходження мінімуму функції $f(x) + \delta(x|M)$ на всьому просторі. За теоремою 2.7.8 субдиференціал функції $f(x) + \delta(x|M)$ дорівнює сумі субдиференціалів функцій $f(x)$ та $\delta(x|M)$. За формулою (2.7.21)

$$\partial\delta(x_0|M) = -K_M^*(x_0). \tag{2.9.4}$$

Отже субдиференціал функції $f(x)+\delta(x|M)$ в точці x_0 дорівнює $\partial f(x_0)-K_M^*(x_0)$. За теоремою 2.9.4 для того щоб точка x_0 була точкою мінімуму функції $f(x)+\delta(x|M)$, необхідно і достатньо, щоб

$$0 \in \partial f(x_0) - K_M^*(x_0).$$

Ця умова еквівалентна умові (2.9.3). Теорема доведена. □

Теорема 2.9.6. *Нехай виконуються умови теореми 2.9.5 і*

$$M = M_1 \cap M_2 \cap \cdots \cap M_k,$$

де $M_i, i = 1, \ldots, k$ – опуклі множини, причому

$$\text{int } M_1 \cap \text{int } M_2 \cap \cdots \cap \text{int } M_{k-1} \cap M_k \neq \emptyset.$$

Для того щоб точка x_0 була точкою мінімуму функції $f(x)$ на множині M, необхідно і достатньо, щоб знайшлись точки $x_i^* \in K_{M_i}^*(x_0), i = 1, \ldots, k$ і точка $x^* \in \partial f(x_0)$ такі, що

$$x^* = x_1^* + \cdots + x_k^*.$$

Доведення. З теорем 1.5.7, 1.5.6 та умов теореми випливає, що

$$K_M(x_0) = \cap_{i=1}^k K_{M_i}(x_0), \tag{2.9.5}$$

і якщо $x_2 \in \text{int } M_i, i = 1, \ldots, k-1; x_2 \in M_k$, то

$$x_2 - x_0 \in \text{int } K_{M_i}, i = 1, \ldots, k-1; \quad x_2 - x_0 \in K_{M_k}.$$

Тому

$$\text{int } K_{M_1}(x_0) \cap \text{int } K_{M_2}(x_0) \cap \cdots \cap \text{int } K_{M_{k-1}}(x_0) \cap K_{M_k}(x_0) \neq \emptyset. \tag{2.9.6}$$

Із співвідношень (2.9.5), (2.9.6) та теореми 1.5.2 випливає, що

$$K_M^*(x_0) = K_{M_1}^*(x_0) + K_{M_2}^*(x_0) + \cdots + K_{M_k}^*(x_0). \tag{2.9.7}$$

Тоді точка $x^* \in \partial f(x_0) \cap K_M^*(x_0)$, яка існує за теоремою 2.9.5, допускає розклад

$$x^* = x_1^* + \cdots + x_k^*, \quad x_i^* \in K_{M_i}^*(x_0), i = 1, \ldots, k. \tag{2.9.8}$$

Це доводить теорему. □

Теорема 2.9.7. *Нехай $f(x)$ – власна опукла функція. Нехай*

$$M = \cap_{i=1}^k M_i,$$

$M_i, i = 1, \ldots, k$ – опуклі множини та нехай існує точка $x_1 \in M$, в якій функція $f(x)$ неперервна. Для того щоб точка x_0 була точкою мінімуму функції $f(x)$ на множині M, необхідно і достатньо, щоб знайшлись такі точки $x^* \in \partial f(x_0)$, $x_i^* \in K_{M_i}^*(x_0), i = 1, \ldots, k$ і число λ, рівне нулю або одиниці, що

$$\lambda x^* = x_1^* + \cdots + x_k^*. \tag{2.9.9}$$

Причому, якщо $\lambda = 0$, то серед точок $x_1^*, \ldots, x_k^*$ є принаймі одна ненульова. Якщо ж $\lambda = 1$, то ці умови є достатніми.

Доведення. За теоремою 2.9.5 існує вектор $x^* \in \partial f(x_0)$, такий, що $x^* \in K^*_M(x_0)$. Оскільки справедлива формула (2.9.5), то за теоремою 1.5.3 можливі два випадки.

1). Справедлива формула (2.9.7). Отже справедлива формула (2.9.8), яка співпадає з (2.9.9) при $\lambda = 1$. Оскільки теорема 2.9.5 дає необхідні і достатні умови, то в цьому випадку умова (2.9.9) є необхідною і достатньою.

2). Існують такі $x_i^* \in K^*_{M_i}(x_0), i = 1, \ldots, k$, не всі рівні нулю, що

$$x_1^* + \cdots + x_k^* = 0.$$

В цьому випадку формула (2.9.9) справджується при $\lambda = 0$. □

2.9.2. Обмеження, що задаються опуклими нерівностями

Нехай M – опукла множина, $f_i, i = 0, 1, \ldots, m$ – власні опуклі функції. Будемо далі вважати, що функції $f_i, i = 0, 1, \ldots, m$ неперервні в точках множини M.

Розглянемо тепер задачу мінімізації функції $f_0(x)$

$$\min_x \{f_0(x) \colon f_i(x) \leq 0, i = 1, \ldots, m, \; x \in M\}. \tag{2.9.10}$$

Розглядатимемо більш загальну задачу, що залежить від параметра. Нехай

$$V(u) = \inf_x \{f_0(x) \colon f_i(x) \leq u_i, i = 1, \ldots, m, \; x \in M\}. \tag{2.9.11}$$

Тоді вихідна задача (2.9.10) є частинний випадок задачі (2.9.11) при $u = 0$.

Лема 2.9.1. *Функція $V(u)$ опукла. Якщо існує така точка $\bar{x} \in M$, що*

$$f_i(\bar{x}) < 0, \quad i = 1, \ldots, m,$$

і $V(0)$ – скінченне число, то $V(u)$ неперервна в точці $u = 0$ і $\partial V(0) \neq \emptyset$.

Доведення. Нехай $\beta_1 > V(u^1)$, $\beta_2 > V(u^2)$. Тоді існують такі точки $x_1 \in M, x_2 \in M$, що

$$f_0(x_1) < \beta_1, f_i(x_1) \leq u_i^1, i = 1, \ldots, m; \; x_1 \in M,$$

$$f_0(x_2) < \beta_2, f_i(x_2) \leq u_i^2, i = 1, \ldots, m; \; x_2 \in M.$$

В силу опуклості функцій та множини

$$f_0(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f_0(x_1) + \lambda_2 f_0(x_2) < \lambda_1 \beta_1 + \lambda_2 \beta_2,$$

$$f_i(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 u_i^1 + \lambda_2 u_i^2, \quad i = 1, \ldots, m,$$

$$\lambda_1 x_1 + \lambda_2 x_2 \in M, \quad \lambda_1 + \lambda_2 = 1, \quad \lambda_1 \geq 0, \lambda_2 \geq 0.$$

Звідси

$$V(\lambda_1 u_i^1 + \lambda_2 u_i^2) < \lambda_1 \beta_1 + \lambda_2 \beta_2.$$

Отже множина $\{(u,\beta)\colon \beta > V(u)\}$ опукла. Тому опукла множина

$$\text{epi}\, V = \{(u,\beta)\colon \beta \geq V(u)\}.$$

Якщо існує точка $\bar{x} \in M$, що задовольняє умови леми, то

$$f_i(\bar{x}) < u_i, \quad i = 1,\ldots,m,$$

для всіх u з деякого околу нуля. Тому

$$V(u) \leq f_0(\bar{x}),$$

і $0 \in \text{int dom}\, V$. Звідси випливає, що функція $V(u)$ в околі нуля приймає скінченні значення і обмежена зверху. Тому вона неперервна в точці $u = 0$ і $\partial V(0) \neq \emptyset$. □

Означення 2.9.1. Функція $L(x,y)$, $x \in X$, $y \in \mathbb{R}^m$, де

$$L(x,y) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x),$$

називається (регулярною) *функцією Лагранжа* задачі опуклого програмування (2.9.10).

Означення 2.9.2. Вектор $y \in \mathbb{R}^m$ називається вектором Куна-Таккера задачі опуклого програмування (2.9.10) якщо $y \geq 0$ і

$$V(0) = \inf_x \{L(x,y)\colon x \in M\}. \tag{2.9.12}$$

Теорема 2.9.8. *Нехай $V(0)$ – скінченне число. Вектор $y \in \mathbb{R}^m$ буде вектором Куна–Таккера задачі опуклого програмування (2.9.10) тоді і тільки тоді, коли $-y \in \partial V(0)$.*

Доведення. За означенням $-y \in \partial V(0)$ тоді і тільки тоді, коли

$$V(u) \geq V(0) - \langle y,u\rangle, \quad \forall u.$$

Тобто, тоді і тільки тоді, коли

$$\inf_u \{V(u) + \langle y,u\rangle\} = V(0).$$

Підставляючи в цю формулу означення $V(u)$, отримаємо

$$\inf_u \inf_x \{f_0(x) + \langle y,u\rangle\colon f_i(x) \leq u_i, i = 1,\ldots,m,\ x \in M\} = V(0). \tag{2.9.13}$$

Якщо $y_i < 0$ при деякому i, то, мінімізуючи за u_i, отримаємо $-\infty$. Отже $y_i \geq 0$ і формулу можна переписати у такому вигляді

$$\inf_x \left\{ f_0(x) + \sum_{i=1}^{m} y_i f_i(x)\colon\ x \in M \right\} = \inf_x \{L(x,y)\colon\ x \in M\}.$$

Отже $-y \in \partial V(0)$ тоді і тільки тоді, коли

$$\inf_x \{L(x,y)\colon\ x \in M\} = V(0),$$

тобто, коли $y \in \mathbb{R}^m$ є вектором Куна-Таккера задачі опуклого програмування (2.9.10). □

Теорема 2.9.9. *Нехай існує така точка $\bar{x} \in M$, що $f_i(\bar{x}) < 0, \quad i = 1,\ldots,m$, і $V(0)$ – скінченне число. Тоді існує вектор Куна-Таккера $y \in \mathbb{R}^m$ задачі опуклого програмування (2.9.10). Якщо x_0 — точка мінімуму задачі (2.9.10), то виконуються умови*

$$f_0(x_0) = L(x_0,y) \le L(x,y), \quad x \in M, y \ge 0,$$

$$y_i f_i(x_0) = 0, \quad i = 1,\ldots,m. \tag{2.9.14}$$

Ці умови є достатніми для того, щоб точка x_0 була розв'язком задачі опуклого програмування (2.9.10).

Доведення. Існування вектора Куна-Таккера $y \in \mathbb{R}^m$ задачі опуклого програмування (2.9.10) випливає з леми 2.9.1 та теореми 2.9.8. Тому переходимо до доведення останньої частини теореми.

Нехай x_0 — точка мінімуму задачі (2.9.10). Тоді

$$f_i(x_0) \le 0, \quad i = 1,\ldots,m; \quad f_0(x_0) = V(0). \tag{2.9.15}$$

Оскільки $y \ge 0$, то

$$V(0) = f_0(x_0) \ge f_0(x_0) + \sum_{i=1}^{m} y_i f_i(x_0) = L(x_0,y).$$

За означенням $V(0) \le L(x_0,y)$. Тому

$$f_0(x_0) = L(x_0,y) = V(0) \le L(x,y), \quad x \in M.$$

Крім того, $\sum_{i=1}^{m} y_i f_i(x_0) = 0$. Оскільки $y_i \ge 0$, $f_i(x_0) \le 0, i = 1,\ldots,m$, то це можливе лише при $y_i f_i(x_0) = 0, i = 1,\ldots,m$.

Покажемо, що умови теореми (2.9.14) є достатніми для того, щоб точка x_0 була розв'язком задачі опуклого програмування (2.9.10). Нехай x – точка, що задовольняє умови задачі (2.9.10). В силу (2.9.14) справджуються такі співвідношення.

$$f_0(x_0) = f_0(x_0) + \sum_{i=1}^{m} y_i f_i(x_0) \le L(x,y) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x) \le f_0(x).$$

Отже точка x_0 є розв'язком задачі опуклого програмування (2.9.10). □

Теорема 2.9.10. *Нехай виконуються умови теореми 2.9.9. Для того, щоб точка x_0 була розв'язком задачі опуклого програмування (2.9.10) необхідно і достатньо, щоб існував вектор Куна–Таккера $y \in \mathbb{R}^m$ задачі (2.9.10) і виконувались умови*

$$\partial_x L(x_0,y) \cap [\mathrm{cone}\,(M - x_0)]^* \ne \emptyset,$$

$$y \ge 0, \quad y_i f_i(x_0) = 0, \quad i = 1,\ldots,m. \tag{2.9.16}$$

Доведення. Умови (2.9.14) теореми 2.9.9 означають, що функція $L(x,y)$ досягає мінімуму за змінною x на множині M в точці x_0. Далі застосовуємо теорему 2.9.5. □

Побудуємо спряжену функцію до функції $V(u)$. За означенням

$$V^*(y) = \sup_u \{\langle y,u\rangle - V(u)\} =$$

$$= \sup_u \left\{\langle y,u\rangle - \inf_x \{f_0(x) \colon f_i(x) \le u_i, i=1,\dots,m; x \in M\}\right\} =$$

$$= \sup_x \sup_u \left\{ -f_0(x) + \sum_{i=1}^m y_i u_i \colon f_i(x) \le u_i, i=1,\dots,m; x \in M \right\}.$$

Можна записати

$$V^*(y) = \begin{cases} \sup_x \{-f_0(x) + \sum_{i=1}^m y_i f_i(x) \colon x \in M\}, & y \le 0, \\ +\infty, & y_i > 0. \end{cases}$$

Або

$$V^*(y) = \begin{cases} -\inf_x \{L(x,-y) \colon x \in M\}, & y \le 0, \\ +\infty, & y_i > 0. \end{cases}$$

Теорема 2.9.11. Теорема двоїстості. *Якщо функція $V(u)$ напівнеперервна знизу при $u = 0$ (або виконуються умови теореми 2.9.9), то*

$$V(0) = \sup_{y \ge 0} \inf_x \{L(x,y) \colon x \in M\}.$$

Вектор Куна–Таккера y_0 є розв'язком задачі

$$\varphi(y_0) = \max_y \{\varphi(y) \colon y \ge 0\},$$

$$\varphi(y) = \inf_x \{L(x,y) \colon x \in M\}.$$

Доведення. Маємо

$$V(0) = V^{**}(0) = \sup_y \{\langle y,0\rangle - V^*(y)\} =$$

$$= \sup_{y \le 0} \inf_x \{L(x,-y) \colon x \in M\}.$$

Заміна $-y$ на y дає перше твердження теореми.

Нехай тепер $y \ge 0$ — довільний вектор. Тоді

$$\varphi(y) = \inf_x \left\{ f_0(x) + \sum_{i=1}^m y_i f_i(x) \colon x \in M \right\} \le$$

$$\le \inf_x \{f_0(x) \colon f_i(x) \le 0, i=1,\dots,m; x \in M\} = V(0).$$

Отже для кожного $y \ge 0$ справджується нерівність $\varphi(y) \le V(0)$.

Якщо ж y_0 – вектор Куна–Таккера, то

$$\varphi(y) \le V(0) = \varphi(y_0), \quad y \ge 0.$$

Теорема доведена. □

2.10. ЗАДАЧІ ТЕОРІЇ НАБЛИЖЕНЬ

Нехай C – опукла компактна множина в X і нехай $0 \in \operatorname{int} C$. За означенням опорної функції

$$s(x^*|C) = \max_x \{\langle x,x^*\rangle\colon\ x \in C\}\,.$$

Оскільки $0 \in \operatorname{int} C$, то $C \supseteq \varepsilon B$, де B – одинична куля, а $\varepsilon > 0$. Тому

$$s(x^*|C) \geq \max_x \{\langle x,x^*\rangle\colon\ x \in \varepsilon B\} = \varepsilon \|x^*\|\,. \tag{2.10.1}$$

Останнє співвідношення випливає з того, що з відомої формули $\langle x,x^*\rangle \leq \|x\|\,\|x^*\|$ поклавши $x_0 = \varepsilon \|x^*\|^{-1} x^*$, отримаємо $x_0 \in \varepsilon B$, $\langle x_0,x^*\rangle = \varepsilon \|x^*\|$.

Розглянемо тепер функцію Мінковського множини C:

$$r_C(x) = \inf_{\rho \geq 0} \{\rho\colon\ x \in \rho C\}\,.$$

Функція Мінковського $r_C(x)$ додатньо однорідна, опукла и скінченна для всіх x, тому вона неперервна як функція від x. Оскільки C – компактна множина, то $r_C(x) > 0$ для $x \neq 0$. Отже $r_C(x)$ можна розглядати як узагальнену відстань від точки x до точки 0.

Визначимо для компактної опуклої множини A і замкнутої опуклої множини M величину

$$d_C(A|M) = \inf_{\rho \geq 0} \{\rho\colon\ A \subseteq M + \rho C\}\,. \tag{2.10.2}$$

Зокрема, якщо A складається з однієї точки x, то отримаємо функцію відстані $d_C(x|M)$, яка вивчалася в попередніх розділах.
Обчислення функції $d_C(x|M)$ зводиться до оцінки відстані від точки x до множини M за умови, що метрика задається за допомогою функції Мінковського $r_C(x)$. Це стане більш зрозумілим, якщо помітимо, що

$$d_C(x|M) = \inf_{\rho \geq 0} \{\rho\colon x \in M + \rho C\} =$$

$$= \inf_{x_1 \in M} \inf_{\rho \geq 0} \{\rho\colon x \in x_1 + \rho C\} =$$

$$= \inf_{x_1 \in M} \inf_{\rho \geq 0} \{\rho\colon\ (x - x_1) \in \rho C\} = \inf_{x_1 \in M} r_C(x - x_1)\,,$$

тобто

$$d_C(x|\,M) = \inf_{x_1} \{r_C(x - x_1)\colon\ x_1 \in\ M\}\,. \tag{2.10.3}$$

В загальному випадку

$$d_C(A|M) = \inf_{\rho} \{\rho \geq 0\colon A \subseteq M + \rho C\} =$$

$$= \sup_{x_2 \in A} \inf_{\rho \geq 0} \{\rho\colon x_2 \in M + \rho C\} = \sup_{x_2 \in A} d_C(x_2|M)\,. \tag{2.10.4}$$

Отже обчислення функції $d_C(A|M)$ зводиться до знаходження в A точки, яка найгіршим чином апроксимується множиною M в метриці, що визначається множиною C. Якщо $C = B$, де B – одинична куля, то $r_B(x) = \|x\|$, так що у цьому випадку наближення відбувається у природній метриці простору X.
Нехай тепер множина M складається з однієї точки x. Тоді

$$d_C(A|x) = \inf_{\rho \geq 0} \{\rho \colon A \subseteq x + \rho C\} = \inf_{\rho \geq 0} \{\rho \colon A - x \subseteq \rho C\}.$$

Звідси видно, що обчислення $d_C(A|x)$ зводиться до знаходження найменшого коефіцієнту розтягу, при якому множина ρC містить $A - x$. Якщо поставити задачу знаходження мінімуму $d_C(A|x)$ за x, то геометрично задача зведеться до знаходження такого зміщення множини A, при якому множину можна помістити у множину ρC з найменшим коефіцієнтом розтягу. Зокрема, якщо $C = B$ – куля, то задача зводиться до задачі знаходження кулі найменшого радіуса, описаної навколо A.
Вияснимо, що дає теорія, яка описана в попередніх параграфах, для таких задач.

Теорема 2.10.1. *Справедлива формула*

$$d_C(A|M) = \sup_{x^*} \{s(x^*|A) - s(x^*|M) \colon s(x^*|C) \leq 1\}.$$

Доведення. Згідно теореми 2.7.20

$$d_C(x|\, M) = \sup_{x^*} \{\langle x, x^* \rangle - s(x^*|M) \colon s(x^*|C) \leq 1\}. \tag{2.10.5}$$

Підставляючи цей вираз у формулу (2.10.4), отримаємо

$$d_C(A|M) = \sup_{x_2 \in A} \sup_{x^*} \{\langle x_2, x^* \rangle - s(x^*|M) \colon s(x^*|C) \leq 1\} =$$

$$= \sup_{x^*} \left\{ \sup_{x_2 \in A} \{\langle x_2, x^* \rangle - s(x^*|M) \colon s(x^*|C) \leq 1\} \right\},$$

що і потрібно було довести. □

Повернемося тепер до задачі про відшукання умов, які характеризують точку x_1, в якій досягається мінімум у правій частині формули (2.10.3), тобто які характеризують точку множини M, найближчу до x у метриці $r_C(x)$.
Оскільки $r_C(x) = d_C(x|\{0\})$, то за теоремою 2.7.21 справедливі формули:

$$\partial r_C(x_0) = \partial \delta(x_0|\{0\}) \cap \{x^* \colon\ s(x^*|C) \leq 1\}, \tag{2.10.6}$$

при $r_C(x_0) = 0$;

$$\partial r_C(x_0) = \partial \delta(x_0|r_C(x_0)\, C) \cap \{x^* \colon\ s(x^*|C) = 1\}, \tag{2.10.7}$$

при $r_C(x_0) > 0$.

Але $r_C(x_0)=0$ при $x_0=0$. Оскільки
$$\partial\delta(x_0|\,M)=-(\mathrm{cone}\;(M-x_0))^*\,, \qquad (2.10.8)$$
а $(\mathrm{cone}\{0\})^*=X^*$, то
$$\partial\delta(0|\;\{0\})=-X^*=X^*,$$
і в силу формул (2.10.6), (2.10.7) виконується співвідношення
$$\partial r_C(0)=\{x^*\colon s(x^*|C)\le 1\}\,. \qquad (2.10.9)$$
Нехай тепер $x_0\ne 0$, так що $r_C(x_0)>0$. Тоді
$$\mathrm{cone}\;(r_C(x_0)\,C-x_0)=\{\lambda\,(r_C(x_0)\,x-x_0)\,:\,\lambda>0,x\in C\}\,.$$
Тому спряжений до нього конус складається з тих точок x^*, для яких
$$\langle r_C(x_0)\,x-x_0,x^*\rangle\ge 0,\quad x\in C,$$
або
$$-\langle x_0,x^*\rangle\ge r_C(x_0)\,s\,(-x^*|\,C)\,. \qquad (2.10.10)$$
Враховуючи формули (2.10.8) і (2.10.6), (2.10.7), отримуємо

$$\partial r_C(x_0)=\{x^*\colon\,\langle x_0,x^*\rangle\ge r_C(x_0)\,,s\,(x^*|\,C)=1\}\,. \qquad (2.10.11)$$
Але $x_0\in r_C(x_0)\,C$ за означенням $r_C(x_0)$ і в силу замкнутості C. Тому
$$\langle x_0,x^*\rangle\le r_C(x_0)\,s\,(x^*|\,C)\,,$$
а, значить, у формулі (2.10.11) має місце рівність
$$\langle x_0,x^*\rangle=r_C(x_0)\,.$$
Отже справджується така теорема.

Теорема 2.10.2. *Якщо C - компактна опукла множина і $0\in\mathrm{int}\,C$, то*
$$\partial r_C(x_0)=\begin{cases}\{x^*\colon s\,(x^*|\,C)\le 1\}\,, & x_0=0,\\ \{x^*\colon\,\langle x_0,x^*\rangle=r_C(x_0)\,,s\,(x^*|\,C)=1\}\,, & x_0\ne 0.\end{cases} \qquad (2.10.12)$$

Тепер можна сформулювати умови, які характеризують точку x_1, в якій функція $r_C(x-x_1)$ досягає свого мінімуму $d_C(x|M)$ на множині M.

Теорема 2.10.3. *Для того щоб точка $x_1\in M$ була найкращим наближенням до точки $x\notin M$ в метриці $r_C(x-x_1)$, необхідно і достатньо, щоб знайшовся такий вектор $x*\in K_M^*(x_1)$, що*
$$\langle x_1-x,x^*\rangle=r_C(x-x_1)=d_C(x|M)\,,\quad s\,(-x^*|C)=1.$$

Доведення. Точка найкращого наближення x_1 за означенням є точкою мінімуму функції $g(x_2)=r_C(x-x_2)$ при $x_2\in M$. Для цього необхідно і достатньо, щоб знайшовся такий вектор $x^*\in K_M^*(x_1)$, що $x^*\in\partial g(x_1)$. Але $x^*\in\partial g(x_1)$ тоді і тільки тоді, коли
$$g(x_2)-g(x_1)=r_C(x-x_2)-r_C(x-x_1)\ge\langle x_2-x_1,x^*\rangle$$

для всіх x_2. Позначимо $z = x - x_2, z_0 = x - x_1$. Отримаємо

$$r_C(z) - r_C(z_0) \geq \langle z - z_0, -x^* \rangle ,$$

тобто $-x^* \in \partial r_C(x - x_1)$. За теоремою 2.10.3 для виконання цієї нерівності необхідно і достатньо, щоб

$$\langle x - x_1, -x^* \rangle = r_C(x - x_1) = d_C(x|M), \qquad s(-x^*|C) = 1,$$

що і доводить теорему. □

Множину $x + \rho C$ будемо для зручності називати *узагальненою кулею радіуса ρ з центром у точці x*. Розглянемо задачу знаходження кулі мінімального радіуса, яка містить даний компакт A. Ця задача зводиться до мінімізації функції $d_C(A|x)$ за x. За означенням

$$d_C(A|x) = \inf_{\rho \geq 0} \{\rho \colon A \subseteq x + \rho C\} = \inf_{\rho \geq 0} \{\rho \colon A - x \subseteq \rho C\} =$$

$$= \sup_{x_1 \in A} \inf_{\rho \geq 0} \{\rho \colon x_1 - x \in \rho C\} = \sup_{x_1 \in A} r_C(x - x_1).$$

Отже, поставлена задача зводиться до знаходження мінімуму функції

$$f(x) = \sup_{x_1 \in A} r_C(x - x_1). \tag{2.10.13}$$

Помітимо, що $f(x)$ – це радіус мінімальної кулі з фіксованим центром x, яка містить множину A. Оскільки A – компакт, а r_C – неперервна функція, то верхня грань у правій частині формули (2.10.13) досягається.

Позначимо

$$A(x) = \{x_1 \colon r_C(x - x_1) = f(x), x_1 \in A\}.$$

За теоремою 2.7.14, всі умови якої у даному випадку виконуються,

$$\partial f(x) = \overline{\operatorname{conv}} \left(\bigcup_{x_1 \in A(x)} \partial_x r_C(x - x_1) \right). \tag{2.10.14}$$

З викладок, зроблених у попередній теоремі, і формули (2.10.12) випливає, що

$$\partial_x r_C(x - x_1) = \{x^* \colon \langle x_1 - x, x^* \rangle = r_C(x - x_1), s(-x^*|C) = 1\}. \tag{2.10.15}$$

Оскільки множина тих x^*, для яких $s(-x^*|C) = 1$, в силу нерівності (2.10.1) обмежена, а множина $A(x)$ компактна, то можна показати, використовуючи теореми 1.2.2 та 1.2.1, що у формулі (2.10.14) замість замкнутої опуклої оболонки можна взяти просто опуклу оболонку, тобто

$$\partial f(x) = \operatorname{conv} \left(\bigcup_{x_1 \in A(x)} \partial_x r_C(x - x_1) \right). \tag{2.10.16}$$

Теорема 2.9.4 стверджує, що точка x_0 є точкою мінімуму функції $f(x)$ тоді і тільки тоді, коли $0 \in \partial f(x)$. З формул (2.10.16), (2.10.15) і теореми

1.2.1 тепер випливає, що якщо x_0 є точкою мінімуму $f(x)$, то існують такі числа $\lambda_i \geq 0$, $i = 0,\ldots,n$, і такі вектори x_i^*, $x_{1i} \in A(x_0)$, $i = 0,\ldots,n$, що

$$\sum_{i=0}^{n} \lambda_i x_i^* = 0, \quad \sum_{i=0}^{n} \lambda_i = 1, \tag{2.10.17}$$

$$\langle x_{i1} - x_0, x_i^* \rangle = r_C(x_0 - x_{1i}) = f(x_0),$$

$$s(-x_i^* | C) = 1, \quad i = 0,\ldots,n.$$

Сформулюємо отриманий результат.

Теорема 2.10.4. *Для того, щоб узагальнена куля $x_0 + f(x_0)C$ була кулею мінімального радіуса, описаного навколо компактної множини A, необхідно і достатньо, щоб виконувалися умови* (2.10.17).

Помітимо, що у попередніх міркуваннях ніяк не використовується опуклість множини A, так що цієї умови нема і у формулюванні теореми.

Теорема 2.10.5. *Нехай A – компактна множина. Тоді існує така її підмножина A_0, яка складається не більш ніж з $n+1$ точок, що радіус мінімальної кулі, описаної навколо A_0, співпадає з радіусом мінімальної кулі, описаної навколо A.*

Доведення. Якщо розглянути множину A_0, яка складається з точок x_{1i}, $i = 0,\ldots,n$, які фігурують в рівнянннях (2.10.17), то, оскільки ці умови є достатніми, куля $x_0 + f(x_0)C$ буде кулею мінімального радіуса, описаного навколо A_0. □

2.11. ЗАДАЧІ НАЙКРАЩОГО РІВНОМІРНОГО НАБЛИЖЕННЯ

Класичною задачею найкращого рівномірного наближення є задача апроксимації многочленами неперервної функції на відрізку.
Нехай $g(t), t \in [0,1]$ – неперервна функція,

$$P_n(x,t) = x^1 + x^2 t + \cdots + x^n t^{n-1},$$

де x – вектор з компонентами $x^1,\ldots,x^n$, які є коефіцієнтами многочлена $P_n(x,t)$ степеня $n-1$ відносно змінної t. Найкраще рівномірне наближення зводиться до відшукання такого многочлена, для якого величина

$$f(x) = \max_{t\in[0,1]} |g(t) - P_n(x,t)| \tag{2.11.1}$$

мінімальна.

Друга класична задача, яка розглядалася з інших позицій у попередньому параграфі, це задача знаходження центру кулі мінімального радіуса, описаної навколо компактної множини $A \subseteq \mathbb{R}^n$. Неважко побачити, що така задача зводиться до задачі мінімізації функції

$$f(x) = \max_y \{\|x - y\| : y \in A\}.$$

На перший погляд, природа сформульованих задач різна, але вони можуть бути розв'язані одним і тим же методом. Цей метод базується на теоремі 2.7.14.

Нехай $f(x,\alpha)$, $\alpha \in A$, де A – компактна множина, сімейство опуклих за $x \subseteq \mathbb{R}^n$, неперервних за x і α функцій. Будемо також припускати, що при кожному $\alpha \in A$ існує градієнт функції $f(x,\alpha)$ по x, тобто існує вектор $f'_x(x,\alpha)$. Припустимо, що він неперервний за α при кожному x. Покладемо

$$f(x) = \max_\alpha \{f(x,\alpha) : \alpha \in A\}, \tag{2.11.2}$$

$$A(x) = \{\alpha \in A\colon f(x,\alpha) = f(x)\}.$$

Функція $f(x)$ опукла, множина $A(x)$ компактна. За теоремою 2.7.14

$$\partial f(x) = \overline{\mathrm{conv}} \left(\bigcup_{\alpha \in A(x)} f'_x(x,\alpha) \right).$$

В силу зроблених припущень всі умови теореми 2.7.14 виконані, а так як функції $f(x,\alpha)$ диференційовні за x, то $\partial f(x,\alpha) = \{f'_x(x,\alpha)\}$. Множина

$$\bigcup_{\alpha \in A(x)} f'_x(x,\alpha)$$

є образом множини $A(x)$ при відображенні $f'_x(x,\alpha)$: $A \to \mathbb{R}^n$, тобто

$$\bigcup_{\alpha \in A(x)} f'_x(x,\alpha) = f'_x(x,A(x)).$$

Але $A(x)$ – компакт, а відображення $f'_x(x,\alpha)$ неперервне за α в силу зроблених припущень. Тому $f'_x(x,A(x))$ – компактна множина. Отже,

$$\partial f(x) = \overline{conv}\, f'_x(x,A(x)).$$

В силу теореми 1.2.3 (якщо M – компакт, то $convM = \overline{conv}M$) маємо

$$\partial f(x) = conv\, f'_x(x,A(x)).$$

Застосовуючи теорему 1.2.1, остаточно отримаємо

$$\partial f(x) = \left\{ \sum_{i=0}^{n} \lambda_i f'_x(x,\alpha_i) : \alpha_i \in A(x), \lambda_i \geq 0, i = 0, \ldots, n, \sum_{i=0}^{n} \lambda_i = 1 \right\}. \tag{2.11.3}$$

З формули (2.11.3) і теореми 2.9.5 безпосередньо випливає наступна теорема.

Теорема 2.11.1. *Для того, щоб функція* $f(x)$*, що визначена співвідношенням* (2.11.2)*, досягала свого мінімуму на опуклій множині* M *в точці* x_0*, необхідно і достатньо, щоб існували такі точки* $\alpha_i \in A(x_0)$*,* $i = 0,\ldots,n$*, і такі числа* λ_i*, що*

$$\sum_{i=0}^{n} \lambda_i {f'}_x(x_0,\alpha_i) \in K_M^*(x_0)\,, \ \lambda_i \geq 0,\ i = 0,\ldots,n,\ \sum_{i=0}^{n} \lambda_i = 1. \qquad (2.11.4)$$

Розглянемо тепер множину точок $\alpha_i \in A(x_0)$, $i = 0,\ldots,n$, які фігурують в теоремі, і позначимо її через A_0:

$$A_0 = \{\alpha_i \colon i = 0,\ldots,n\}\,.$$

Покладемо

$$f_0(x) = \max_{\alpha} \{f(x,\alpha) : \alpha \in A_0\}\,. \qquad (2.11.5)$$

Якщо точка x_0 є точкою мінімуму функції $f(x)$, то виконані умови (2.11.4), які є необхідними і достатніми умовами мінімуму. Застосування теореми 2.11.1 до функції $f_0(x)$ показує, що x_0 є також точкою мінімуму і для функції $f_0(x)$.

Отже, справедлива наступна теорема.

Теорема 2.11.2. *При зроблених вище припущеннях існує підмножина* $A_0 \subset A$*, яка складається не більш ніж з* $n+1$ *точок* α_i*, і така, що точка мінімуму функції* $f(x)$ *є одночасно точкою мінімуму функції* $f_0(x)$*, яка визначена співвідношенням* (2.11.5).

Використаємо отримані результати для розв'язання поставлених задач. Нехай

$$f(x,y) = \|x - y\|\,, y \in A,$$

де A – компакт. Якщо $x \neq y$, тоді

$${f'}_x(x,y) = \frac{x-y}{\|x-y\|} = \frac{x-y}{f(x,y)}. \qquad (2.11.6)$$

Нехай

$$f(x) = \max_{y} \{\|x-y\| : y \in A\}\,, \qquad (2.11.7)$$

$$A(x) = \{y \in A\colon\ \|x-y\| = f(x)\}\,. \qquad (2.11.8)$$

Застосуємо теорему 2.11.1 у тому випадку, коли $M = \mathbb{R}^n$. Отримаємо, що точка x_0 тоді і тільки тоді є точкою мінімуму функції (2.11.7), коли існують точки $y_i \in A$, $i = 0,\ldots,n$, і числа $\lambda_i \geq 0$ такі, що

$$\sum_{i=0}^{n} \lambda_i \frac{x_0 - y_i}{\|x_0 - y_i\|} = 0, \quad \sum_{i=0}^{n} \lambda_i = 1, \quad \|x_0 - y_i\| = f(x_0)\,.$$

Звідси легко отримуємо

$$x_0 = \sum_{i=0}^{n} \lambda_i y_i, \lambda_i \geq 0, \|x_0 - y_i\| = f(x_0), i = 0, \ldots, n, \sum_{i=0}^{n} \lambda_i = 1. \quad (2.11.9)$$

Нагадаємо, що ми розглядаємо задачу про пошук центру кулі найменшого радіуса, описаної навколо компакта A. Геометрична інтерпретація умов (2.11.9) дозволяє сформулювати наступну теорему.

Теорема 2.11.3. *Для того, щоб точка x_0 була центром кулі найменшого радіуса, описаної навколо компакта A, необхідно і достатньо, щоб існували такі точки $y_i \in A$, $i = 0, \ldots, n$, які лежать на поверхні кулі, що x_0 належить симплексу, породженому цими точками.*

Той факт, що точки y_i лежать на поверхні кулі, виражається другим із співвідношень (2.11.9). Геометрична інтерпретація теореми 2.11.2 для даного випадку достатньо очевидна.

Теорема 2.11.4. *В компактній множині $A \subseteq \mathbb{R}^n$ існує така підмножина A_0, що складається не більш ніж з $n+1$ точки, що куля мінімального радіуса, яка описана навколо A, одночасно є кулею мінімального радіуса, яка описана навколо множини A_0.*

Нехай тепер на деякому компакті $\Omega \subseteq \mathbb{R}^m$ задані неперервні функції $\varphi_i(y)$, $y \in \Omega$, $i = 1, \ldots, n$. Назвемо *узагальненим многочленом* вираз

$$P_n(x,y) = \sum_{i=1}^{n} x^i \varphi_i(y). \quad (2.11.10)$$

Якщо $g(y)$ – довільна неперервна функція, то задача наближення функції $g(y)$ за допомогою узагальненого многочлена полягає у мінімізації функції

$$f(x) = \max_{y \in \Omega} |g(y) - P_n(x,y)| \quad (2.11.11)$$

за $x \subseteq \mathbb{R}^n$. Нехай

$$A = \left\{ (y,\xi) \in \mathbb{R}^{m+1} : y \in \Omega, |\xi| \leq 1 \right\}.$$

Покладемо

$$f(x,y,\xi) = \xi \left(g(y) - P_n(x,y) \right).$$

Тоді

$$f'_x(x,y,\xi) = -\xi \varphi(y), \quad (2.11.12)$$

де $\varphi(y)$ – n-вимірний вектор з компонентами $\varphi_i(y)$. Можна переконатися в тому, що

$$f(x) = \max_{y,\xi} \left\{ f(x,y,\xi) : (y,\xi) \in A \right\}. \quad (2.11.13)$$

Далі, якщо $(y_1,\xi_1) \in A(x)$, тобто
$$\xi_1 \left(g(y_1) - P_n(x,y_1)\right) = \max_{(y,\xi)} \left\{\xi \left(g(y) - P_n(x,y)\right) : y \in \Omega, |\xi| \le 1\right\},$$
то
$$\xi_1 = \operatorname{sign}\left(g(y_1) - P_n(x,y_1)\right). \tag{2.11.14}$$
Застосування теореми 2.11.1 до функції $f(x)$, яка визначається співвідношенням (2.11.13), дає наступний результат.

Точка x_0 дає найменше значення функції $f(x)$ тоді і тільки тоді, коли існують такі точки $y_i \in \Omega$, $i = 0,\ldots,n$, що
$$\sum_{i=0}^{n} \lambda_i \xi_i g(y_i) = 0, \tag{2.11.15}$$
$$\xi_i = sign\left(g(y_i) - P_n(x_0,y_i)\right), \tag{2.11.16}$$
$$|g(y_i) - P_n(x_0,y_i)| = f(x_0),\ i = 0,1,\ldots,n, \tag{2.11.17}$$
$$\sum_{i=0}^{n} \lambda_i = 1,\ \lambda_i \ge 0. \tag{2.11.18}$$

Теорема 2.11.5. *Для того, щоб узагальнений многочлен $P_n(x_0,y)$ був многочленом найкращого рівномірного наближення неперервної функції $g(y)$ на компакті Ω, необхідно і достатньо, щоб існували такі точки $y_i \in \Omega$, $i = 0,\ldots,n$, у яких відхилення многочлена від функції $g(y)$ максимальне по модулю і виконані умови*
$$\sum_{i=0}^{n} \lambda_i \xi_i g(y_i) = 0,$$
$$\xi_i = sign\left(g(y_i) - P_n(x_0,y_i)\right),$$
при деяких невід'ємних λ_i, які не дорівнюють нулю одночасно.

Доведення. Доведення теореми зводиться до аналізу формул (2.11.15) – (2.11.18). Те, що числа λ_i невід'ємні і не всі рівні нулю, випливає з (2.11.18). Співвідношення (2.11.17) виражає той факт, що в точках y_i досягається максимальне відхилення, оскільки $f(x_0)$ визначається з формули (2.11.11). □

Застосовуючи теорему 2.11.2 до задачі, отримаємо таке твердження.

Теорема 2.11.6. *Існує підмножина Ω_0 компакту Ω, яка складається не більш ніж з $n+1$ точок і така, що многочлен найкращого наближення функції $g(y)$ на Ω є одночасно многочленом найкращого наближення функції $g(y)$ на Ω_0.*

Розділ III
Верхня опукла апроксимація

Для того, щоб можна було виписати необхідні умови екстремуму для задач, які задаються множинами та функціями, що не мають відповідних властивостей опуклості та диференційовності, необхідно, щоб множини та функції, що фігурують в задачі, мали певні локальні властивості. В околі точки мінімума функції мають допускати апроксимацію деякими більш простими функціями. Так, наприклад, гладкі функції допускають лінійну апроксимацію. Опуклі функції добре апроксимуються опуклими додатньо однорідними функціями – похідними за напрямком. Проте негладкі і неопуклі функції вже неможливо наблизити в околі деякої точки опуклими додатньо однорідними функціями. Для таких функцій Б. М. Пшеничний [47] визначив поняття верхньої опуклої апроксимації та нижньої угнутої апроксимації.

3.1. КОНУСИ ДОТИЧНИХ НАПРЯМКІВ ТА ШАТРА

3.1.1. Конуси дотичних напрямків

Нехай M– довільна множина в просторі X.

Означення 3.1.1. Вектор $g \in X$ називається дотичним до множини M в точці $x \in M$, якщо існує така функція $\varphi(\lambda) \in X$, що

$$x + \lambda g + \varphi(\lambda) \in M$$

при достатньо малих $\lambda \geqslant 0$ і $\lambda^{-1}\varphi(\lambda) \to 0$ коли $\lambda \downarrow 0$.

Дотичні напрямки утворюють деякі множини – як правило конуси: якщо g – дотичний вектор, то і вектор αg, $\alpha \geqslant 0$, теж є дотичним.

Означення 3.1.2. Опуклий конус $K(x,M)$ називається конусом дотичних напрямків до множини M в точці x, якщо з включення $g \in K(x,M)$ випливає, що g – дотичний вектор до множини M в точці $x \in M$.

Якщо M– опукла множина, то

$$K(x,M) = \mathrm{cone}\,(M - x) = \{g : g = \lambda(y - x), y \in M, \lambda > 0\} \qquad (3.1.1)$$

є конусом дотичних напрямків до M. Для цього достатньо покласти $\varphi(\lambda) \equiv 0$ в означенні (3.1.1).
Наведемо приклади конусів дотичних напрямків.

Приклад 3.1.1. Нехай множина M задана системою рівнянь

$$f_i(x) = 0, \quad i \in I, \tag{3.1.2}$$

де I – скінченна множина індексів, а функції $f_i(x)$ неперервно диференційовні.

Лема 3.1.1. *Якщо $x_0 \in M$, тобто x_0 задовольняє системі рівнянь (3.1.2), і градієнти $f_i'(x), i \in I$, лінійно незалежні, то*

$$K(x_0,M) = \{g : \langle g,f_i'(x_0)\rangle = 0, i \in I\} \tag{3.1.3}$$

є конусом дотичних напрямків.

Доведення. Нехай $f'(x_0)$ – матриця, складена з вектор–рядків $f_i'(x_0), i \in I$. Нехай розмірність множини I дорівнює m. Тоді $f'(x_0)$ має розмірність $m \times n$ ($X = \mathbb{R}^n$). Умову $g \in K(x,M)$ можна записати у вигляді

$$f'(x_0)g = 0.$$

Нехай $(f'(x_0))^*$– транспонована до $f'(x_0)$ матриця і нехай

$$A = f'(x_0)(f'(x_0))^*.$$

Квадратна матриця A розмірності $m \times m$ невироджена. Дійсно, якщо припустити, що вона вироджена, то знайдеться ненульовий вектор $y \in \mathbb{R}^m$ такий, що $Ay = 0$. Тоді

$$\langle y,Ay\rangle = \|(f'(x_0))^*y\|^2 = \|\sum_{j\in I} f_j'(x_0)y^j\|^2 = 0,$$

тобто

$$\sum_{j\in I} f_j'(x_0)y^j = 0$$

при умові, що не всі y^j дорівнюють нулю. Остання рівність означає, що вектори $f_i'(x_0), i \in I$, лінійно залежні, що суперечить припущенню. Отже матриця A невироджена.

Розглянемо систему рівнянь

$$u_i(\lambda,y) \equiv f_i(x_0 + \lambda g + (f'(x_0))^*y) = 0, i \in I, \tag{3.1.4}$$

де y - невідомі величини, а λ – змінний параметр. Неважко порахувати, що

$$\frac{\partial u_i(0,0)}{\partial \lambda} = \langle g,f'(x_0)\rangle = 0, i \in I,$$

а матриця з елементами $\frac{\partial u_i(0,0)}{\partial y^j}$; $i,j = 1,\ldots,m$, співпадає з A і тому невироджена. За теоремою, що доводиться нижче, система (3.1.4) має при достатньо малих $\lambda > 0$ розв'язок $y(\lambda)$, причому $\lambda^{-1}y(\lambda) \to 0$ при $\lambda \to 0$. Покладемо

$$\varphi(\lambda) = (f'(x_0))^*y(\lambda).$$

Тоді $\lambda^{-1}\varphi(\lambda) \to 0$ при $\lambda \to 0$ і згідно з системою (3.1.4)

$$x_0 + \lambda g + \varphi(\lambda) \in M$$

при достатньо малих $\lambda > 0$, тобто x – дотичний вектор. □

Знайдемо конус, який спряжений до конуса, що задається співвідношенням (3.1.3). Оскільки система рівнянь

$$\langle g, f'(x_0) \rangle = 0, i \in I,$$

еквівалентна системі нерівностей

$$\langle g, f'(x_0) \rangle \geqslant 0, i \in I,$$

$$\langle g, -f'(x_0) \rangle \geqslant 0, i \in I,$$

то за теоремою 1.8.3 конус $K^*(x_0, M)$ складається з елементів, які мають вигляд

$$x^* = \sum_{i \in I} \lambda_i^+ f_i'(x_0) - \sum_{i \in I} \lambda_i^- f_i'(x_0), \quad \lambda_i^+ \geqslant 0, \lambda_i^- \geqslant 0.$$

Позначивши $\lambda_i = \lambda_i^+ - \lambda_i^-$, отримаємо

$$K^*(x_0, M) = \{x^* : x^* = \sum_{i \in I} \lambda_i f_i'(x_0), \lambda_i \in R^1, i \in I\}. \tag{3.1.5}$$

Приклад 3.1.2. Нехай множина M задана системою рівнянь та нерівностей

$$M = \{x : f_i(x) \leqslant 0,\ i \in I^-,\ f_i(x) = 0,\ i \in I\},$$

де I та I^- - скінченні множини індексів, функції $f_i(x)$ неперервно диференційовні. Нехай $x_0 \in M$ і

$$I^-(x_0) = \{i \in I^- : f_i(x_0) = 0\}.$$

Якщо вектори $f_i(x_0), i \in I$, лінійно незалежні, то напрямок g, що задовольняє співвідношення

$$\langle g, f_i'(x_0) \rangle < 0, \quad i \in I^-(x_0); \tag{3.1.6}$$

$$\langle g, f_i'(x_0) \rangle = 0, \quad i \in I,$$

є дотичним. Дійсно, згідно з попереднім прикладом існує така функція $\varphi(\lambda)$, $\lambda^{-1}\varphi(\lambda) \to 0$ якщо $\lambda \downarrow 0$, що

$$f_i(x + \lambda g + \varphi(\lambda)) = 0, i \in I.$$

З іншого боку, для $i \in I^-$ за відомою формулою з аналізу маємо

$$f_i(x + \lambda g + \varphi(\lambda)) = f_i(x_0) + \lambda \left\langle g + \lambda^{-1}\varphi(\lambda), f_i'(\xi_i) \right\rangle, \tag{3.1.7}$$

де ξ_i - точка відрізку, що з'єднує x_0 і $x_0 + \lambda g + \varphi(\lambda)$. З формули (3.1.7) випливає, що якщо $i \in I^- \backslash I^-(x_0)$, то $f_i(x_0) < 0$, і

$$f_i(x + \lambda g + \varphi(\lambda)) < 0$$

при достатньо малих $\lambda > 0$. Якщо ж $i \in I^-(x_0)$, то з формули (3.1.7) маємо

$$f_i(x + \lambda g + \varphi(\lambda)) = \lambda \langle g, f_i'(x_0) \rangle + \lambda \langle g, f_i'(\xi_i) - f_i'(x_0) \rangle + \langle \varphi(\lambda), f_i'(\xi_i) \rangle .$$

Оскільки $\xi_i \to x_0$ при $\lambda \to 0$, то останні два доданки прямують до нуля, причому швидше за λ. Тому в силу нерівності (3.1.6)

$$f_i(x + \lambda g + \varphi(\lambda)) < 0,\ i \in I^-(x_0)$$

при малих $\lambda > 0$.

Отже, при достатньо малих $\lambda > 0$ виконуються співвідношення

$$f_i(x_0 + \lambda g + \varphi(\lambda)) < 0,\ i \in I^-;$$

$$f_i(x_0 + \lambda g + \varphi(\lambda)) = 0,\ i \in I,$$

тобто $x_0 + \lambda g + \varphi(\lambda) \in M$. Отримане включення означає, що вектор g – дотичний напрямок до множини M. Таким чином,

$$K(x_0, M) = \{g : \langle g, f_i'(x_0) \rangle < 0, i \in I^-(x_0), \langle g, f_i'(x_0) \rangle = 0, i \in I\}. \quad (3.1.8)$$

Застосовуючи теорему 1.8.3, неважко показати, що

$$K^*(x_0, M) = \{x^* : x^* = -\sum_{i \in I^-} \lambda_i f_i'(x_0) - \sum_{i \in I} \lambda_i f_i'(x_0), \lambda_i \geq 0, i \in I^-(x_0)\}. \quad (3.1.9)$$

3.1.2. Теорема про неявні функції

Як видно з наведених прикладів, обчислення дотичних напрямків неможливе без застосування теорем про існування розв'язків тих чи інших систем нелінійних рівнянь. Теореми такого типу будуть необхідні у подальшому для побудови шатрів і обгрунтування необхідних умов екстремуму.

Теорема 3.1.1. *Нехай $x \in \mathbb{R}^n$, $y \in \mathbb{R}^k$ і $g(y,x)$ – неперервна векторна функція з компонентами $g_i(y,x), i = 1, \ldots, n$, що задовольняють умову: існує така невироджена $n \times n$–матриця g_x', що*

$$\|g(y,x) - g_x' x\| \leqslant \overline{r}(\sqrt{\|x\|^2 + \|y\|^2}),$$

де $\overline{r}(\lambda) = o(\lambda)$. Тоді при достатньо малих y система рівнянь $g(y,x) = 0$ має розв'язки відносно x, причому існує такий розв'язок $x(y)$, що

$$\lim_{y \to 0} \frac{\|x(y)\|}{\|y\|} = 0.$$

Зауваження 3.1.1. З умов теореми випливає, що:

а) $g(0,0) = 0$;

б) існують перші частинні похідні функції $g(y,x)$ в точці $y = 0$, $x = 0$, причому

$$g_{iy}'(0,0) = 0,\ i = 1, \ldots, n;$$

в) матриця частинних похідних $g'_x = \frac{\partial g_i(0,0)}{\partial x_j}; j,i = 1,\ldots,n$ – невироджена.

Якщо функції $g_i(y,x)$ неперервно диференційовні в околі точки $y = 0$, $x = 0$, то легко перевірити, що з умов а), б), в) випливає виконання умов теореми 3.1.1. Більш того, в цьому випадку вона співпадає зі звичайною теоремою про неявні функції, з якої випливає, що $x(y)$ є неперервно диференційовною функцією y в околі початку координат.

Доведення. З умов теореми випливає, що

$$g(y,x) = g'_x x + r(y,x), \tag{3.1.10}$$

$$\|r(y,x)\| \leqslant \overline{r}(\sqrt{\|x\|^2 + \|y\|^2}).$$

Розглянемо відображення

$$f(y,x) = x - (g'_x)^{-1} g(y,x).$$

Використовуючи відношення (3.1.10), отримаємо

$$f(y,x) = -(g'_x)^{-1} r(y,x), \tag{3.1.11}$$

$$\|f(y,x)\| \leqslant \|(g'_x)^{-1}\| \ \overline{r}(\sqrt{\|x\|^2 + \|y\|^2}).$$

Відмітимо, що можна вважати $\overline{r}(\lambda)$ неспадною функцією λ. Справді, якщо це не так, то замінимо функцію $\overline{r}(\lambda)$ на функцію

$$w(\lambda) = \sup_{0 \leqslant t \leqslant \lambda} \overline{r}(t),$$

покажемо, що $w(\lambda)$– неспадна функція і

$$\frac{w(\lambda)}{\lambda} \to 0.$$

Справді, оскільки $\lambda^{-1}\overline{r}(\lambda) \to 0$, то для довільного $\varepsilon > 0$ знайдеться таке $\delta > 0$, що

$$\frac{\overline{r}(t)}{t} \leqslant \varepsilon,$$

при $t \leqslant \delta$. Якщо тепер $\lambda < \delta$, то

$$\frac{w(\lambda)}{\lambda} = \frac{1}{\lambda} \sup_{0 \leqslant t \leqslant \lambda} \overline{r}(t) \leqslant \sup_{0 \leqslant t \leqslant \lambda} \frac{\overline{r}(t)}{t} \leqslant \varepsilon,$$

тобто $\frac{w(\lambda)}{\lambda} \to 0$.

Нехай тепер

$$\tau(y) = \inf_{\tau} \{\tau : c\overline{r}(\sqrt{\|y\|^2 + \tau^2}) \leqslant \tau\},$$

де $c = \|(g'_x)^{-1}\|$. При достатньо малих y виконується нерівність

$$c\overline{r}(\sqrt{\|y\|^2 + \|y\|^2}) = c\overline{r}(\sqrt{2}\|y\|) \leqslant \|y\|,$$

тому $\tau(y) \leqslant \|y\|$. Крім того, за означенням точної нижньої грані для довільного y існує таке $\tau^*(y)$, що

$$\tau(y) \leqslant \tau^*(y) \leqslant \tau(y) + \|y\|^2,$$

$$c\overline{r}(\sqrt{\|y\|^2+(\tau^*(y))^2}) \leqslant \tau^*(y). \tag{3.1.12}$$
Покажемо, що
$$\frac{\tau(y)}{\|y\|} \to 0,$$
при $y \to 0$. За означенням $\tau(y)$ маємо
$$\tau(y)-\|y\|^2 < c\overline{r}\sqrt{\|y\|^2+(\tau(y)-\|y\|^2)^2} \leqslant c\overline{r}(\|y\|+|\tau(y)-\|y\|^2|), \tag{3.1.13}$$
де використано те, що $\overline{r}(\lambda)$– неспадна функція і
$$\sqrt{\alpha^2+\beta^2} \leqslant \alpha+\beta, \alpha>0, \beta>0.$$
Далі, оскільки $\tau(y) \leqslant \|y\|$ для малих y, то праву частину нерівності (3.1.13) можна оцінити наступним чином:
$$c\overline{r}(\|y\|+|\tau(y)-\|y\|^2|) \leqslant c\overline{r}(2\|y\|),$$
звідки
$$\tau(y)-\|y\|^2 \leqslant c\overline{r}(2\|y\|),$$
або
$$\frac{\tau(y)}{\|y\|} \leqslant 2c\frac{\overline{r}(2\|y\|)}{2\|y\|}+\|y\|,$$
тобто
$$\frac{\tau(y)}{\|y\|} \to 0$$
при $y \to 0$. Але тоді, з нерівності (3.1.12) випливає, що
$$\frac{\tau^*(y)}{\|y\|} \to 0.$$
Якщо тепер
$$\|x\| \leqslant \tau^*(y),$$
то згідно до нерівностей (3.1.11) та (3.1.12) справедливі нерівності
$$\|f(y,x)\| \leqslant c\overline{r}\sqrt{\|y\|^2+\|x\|\|^2} \leqslant c\overline{r}\sqrt{\|y\|^2+(\tau^*(y))^2} \leqslant \tau^*(y).$$
Звідси випливає, що неперервне відображення $f(y,x)$ відображає кулю $\|x\| \leqslant \tau^*(y)$ в себе. На основі теореми Брауера можна стверджувати, що відображення $f(y,x)$ має нерухому точку в цій кулі, тобто існує така точка $x(y)$, що
$$x(y)=f(y,x), \|x(y)\| \leqslant \tau^*(y).$$
Але з означення $f(y,x)$ випливає, що
$$f(y,x(y))=0.$$
Крім того,
$$\frac{\|x\|}{\|y\|} \leqslant \frac{\tau^*(y)}{\|y\|},$$

тобто $\|y\|^{-1}x(y) \to 0$ при $y \to 0$. Теорему доведено. $\square$

Покажемо, що перевірка тверджень теореми може бути спрощена.

Лема 3.1.2. *Нехай функція $f(x), x \in \mathbb{R}^n$, має в точці x_0 градієнт $f'(x_0)$ і*

$$\lim_{\lambda \to 0} \frac{f(x_0 + \lambda g + \varphi(\lambda)) - f(x_0)}{\lambda} = \langle g, f'(x_0) \rangle$$

для довільного вектора $g \in \mathbb{R}^n$ і довільної функції $\varphi(\lambda) \in \mathbb{R}^n$, яка визначена для малих $\lambda > 0$ і така, що задовольняє умову $\lambda^{-1}\varphi(\lambda) \to 0$, при $\lambda \downarrow 0$. Тоді існує така функція $\overline{r}(\alpha)$, $\alpha^{-1}\overline{r}(\alpha) \to 0$, при $\alpha \downarrow 0$, що

$$|f(x) - f(x_0) - \langle x - x_0, f'(x_0) \rangle| \leqslant \overline{r}(\|x - x_0\|).$$

Доведення. Будемо вважати, що $x_0 = 0, f(x_0) = 0$, так що

$$\lim_{\lambda \to 0} \frac{f(\lambda g + \varphi(\lambda))}{\lambda} = \langle g, f'(0) \rangle\,. \tag{3.1.14}$$

Припустимо, що лема невірна. Тоді знайдеться така послідовність точок $x_k \to 0$, що

$$\frac{|f(x_k) - \langle x_k, f'(0) \rangle|}{\|x_k\|} \geqslant \varepsilon > 0. \tag{3.1.15}$$

Вектори

$$\overline{x_k} = \frac{x_k}{\|x_k\|}$$

мають одиничну норму, тому можна вибрати збіжну підпослідовність. Нехай $\overline{x}_k \to g$. Покладемо

$$\lambda_k = \|x_k\|, \quad \varphi(\lambda_k) = \lambda_k(\overline{x}_k - g). \tag{3.1.16}$$

Можна вважати, що $\lambda_k > \lambda_{k+1}$. На відрізках $[\lambda_{k+1}, \lambda_k]$ задамо функцію $\varphi(\lambda)$ за допомогою лінійної інтерполяції. Тоді із співвідношень (3.1.16) і того, що $\overline{x}_k \to g$, випливає, що $\lambda^{-1}\varphi(\lambda) \to 0$. Тому справедлива формула (3.1.14). З іншого боку,

$$x_k = \lambda_k g + \varphi(\lambda_k),$$

і в силу оцінки (3.1.15) виконується нерівність

$$\left| \lim_{k \to \infty} \frac{f(\lambda_k g + \varphi(\lambda_k))}{\lambda_k} - \langle g, f'(0) \rangle \right| \geqslant \varepsilon > 0,$$

що суперечить співвідношенню (3.1.14). Отримана суперечність доводить лему. $\square$

З результатів леми 3.1.2 випливає, що якщо функції $g_i(y,x)$ задовольняють умови а), б), в) зауваження до теореми 3.1.1 і, крім того, виконується співвідношення

$$\lim_{\lambda \to 0} \frac{g_i(\lambda \overline{y} + \varphi_1(\lambda), \lambda \overline{x} + \varphi_2(\lambda))}{\lambda} = \langle \overline{x}, g'_{ix}(0,0) \rangle$$

для будь-яких $\varphi_1(\lambda)$ і $\varphi_2(\lambda)$, що прямують до нуля швидше, ніж λ, то справедлива теорема 3.1.1.

3.1.3. Шатра

Конус дотичних напрямків містить напрямки, для кожного з яких в означенні 3.1.1 існує своя функція $\varphi(\lambda)$. Цього виявляється недостатньо для того, щоб робити висновки про властивості множини M.

Означення 3.1.3. Конус дотичних напрямків $K(x_0,M)$ в точці $x_0 \in M$ називається локальним шатром, якщо для кожного напрямку $g \in \mathrm{ri}\, K(x_0,M)$ існує опуклий конус Q і визначене в околі початку координат неперервне відображення $\psi(g)$ такі, що:

а) $g \in \mathrm{ri}\, Q$, $\mathrm{Lin}\ Q = \mathrm{Lin}\ K(x_0,M)$, $Q \subseteq K(x_0,M)$;

б) $\psi(g) = g + r(g)$, $\|g\|^{-1} r(g) \to 0$ при $g \to 0$;

в) $x_0 + \psi(g) \in M$ для $g \in Q \cap (\varepsilon B)$ при деякому $\varepsilon > 0$.

(Нагадаємо, що $B-$ одинична куля в просторі $X = \mathbb{R}^n$.)

Приклад 3.1.3. Нехай множина M задана так як в прикладі 3.1.1 і нехай в точці x_0 виконуються умови леми 3.1.1. Покажемо, що конус $K(x_0,M)$, визначений формулою (3.1.3), є локальним шатром. Для цього складемо систему рівнянь

$$g_i(\overline{x},y) \equiv f_i(x_0 + \overline{x} + (f'(x_0))^* y) - \langle \overline{x}, f_i'(x_0) \rangle = 0, i \in I, \tag{3.1.17}$$

в якій невідомим є y, а $\overline{x}$ – параметр. Неважко помітити, що

$$g_i(0,0) = 0, i \in I;$$

$$g'_{i\overline{x}}(0,0) = 0, i \in I,$$

а матриця похідних $\frac{\partial g_i(0,0)}{\partial y^j}, i,j \in I$ співпадає з матрицею

$$A = f'(x_0)(f'(x_0))^*$$

і тому є невиродженою. Оскільки функції $f_i(x)$ неперервно диференційовні, то виконуються умови теореми 3.1.1. Згідно зауваження до теореми, при достатньо малих $\overline{x}$ існує гладкий розв'язок $y(\overline{x})$ такий, що

$$\frac{y(x)}{||\overline{x}||} \to 0. \tag{3.1.18}$$

Покладемо $Q = K_m(x_0)$ і

$$\psi(\overline{x}) = \overline{x} + (f_i'(x_0))^* y(\overline{x}). \tag{3.1.19}$$

Візьмемо $\overline{x} \in Q \cap (\varepsilon B)$, де $\varepsilon > 0$, так, щоб $y(\overline{x})$ було визначене в шатрі εB. В силу формули (3.1.3) співвідношення (3.1.17) переходить в рівність

$$f_i(x_0 + \psi(\overline{x})) = 0, \quad i \in I,$$

тобто $x_0 + \psi(\overline{x}) \in M$. Звідси та з формул (3.1.18) (3.1.19) випливає, що $K(x_0,M)-$ локальне шатро.

Приклад 3.1.4. Нехай множина M задана як у прикладі 3.1.2. Покажемо, що конус $K(x_0,M)$, що визначений формулою (3.1.8), є локальним шатром. Неважко побачити, що $\operatorname{ri} K(x_0,M) = K(x_0,M)$.

Нехай $g_0 \in K(x_0,M)$. Покладемо

$$\delta = \frac{1}{\|g\|} \max_i \{\langle g_0, f_i'(x_0)\rangle : i \in I^-(x_0)\} < 0$$

$$Q = \{g : \langle g, f_i'(x_0)\rangle \leqslant \delta\|g\|, i \in I^-(x_0), \langle g, f_i'(x_0)\rangle = 0, \quad i \in I\}. \quad (3.1.20)$$

Тоді $g_0 \in \operatorname{ri} Q$ і $Q \subseteq K(x_0,M)$. Виберемо функцію $\psi(g)$ так як в прикладі 3.1.3. За теоремою про середнє неважко переконатися в тому, що

$$f_i(x_0 + \psi(g)) = f_i(x_0) + \langle g, f_i'(x_0)\rangle + o(\|x\|), \quad i \in I^-. \quad (3.1.21)$$

Якщо $i \in I^- \backslash I^-(x_0)$, то $f_i(x_0) < 0$. Тоді з формули (3.1.21) випливає, що

$$f_i(x_0 + \psi(g)) < 0, \quad i \in I^- \backslash I^-(x_0), \quad (3.1.22)$$

при малих g. Якщо $i \in I^-(x_0)$, то для $g \in Q$

$$f_i(x_0 + \psi(g)) \leqslant \delta\|g\| + o(\|g\|), \quad i \in I^-(x_0), \quad (3.1.23)$$

при малих g. Нарешті,

$$f_i(x_0 + \psi(g)) = 0, \quad i \in I, \quad (3.1.24)$$

за побудовою функції $\psi(g)$. З формул (3.1.22)—(3.1.24) випливає, що при малих $g \in Q$ сума $x_0 + \psi(g) \in M$, тобто $K(x_0,M)$– локальне шатро. Отже, якщо M задана, як в прикладі 3.1.2, і в точці x_0 вектори $f_i'(x_0), i \in I$, лінійно незалежні, то конус $K(x_0,M)$, що задається формулою (3.1.8), є локальним шатром.

Приклад 3.1.5. Нехай M– опукла множина. Тоді

$$K(x_0,M) = \operatorname{cone}(M - x_0), \ x_0 \in M,$$

є локальним шатром.

Можна вважати, що $x_0 = 0$, $0 \in M$ і M, має внутрішні точки. В тому випадку, коли $\operatorname{int} M = \emptyset$, потрібно розглядати M відносно простору $\operatorname{Lin} M$. Отже

$$K(0,M) = \{g : g = \lambda x, \lambda > 0, x \in M\}.$$

Нехай $g_0 \in \operatorname{int}\ K(0,M)$, $g_0 \neq 0$. Тоді

$$g_0 + \varepsilon B \subseteq K(0,M), \quad 0 \notin (g_0 + \varepsilon B)$$

для достатньо малих $\varepsilon > 0$ (нагадаємо, що B- одинична куля з центром в нулі). Виберемо точки $g_i, i = 1, \ldots, n+1$, так, щоб $g_i \in (g_0 + \varepsilon B)$ і щоб точка g_0 була центром симплекса S, утвореного точками g_i. Тоді існує таке $\varepsilon_1 > 0$, що

$$g_0 + \varepsilon_1 B \subseteq S \subseteq K(0,M).$$

Оскільки $g_i \in K(0,M)$, то $g_i = \lambda x_i, x_i \in M$. Зрозуміло, що $\lambda_i > 0$, оскільки $g_i \neq 0$, так як $0 \notin (g_0 + \varepsilon B)$, а $g_i \in (g_0 + \varepsilon B)$. Покладемо

$$\lambda_0 = \min\left\{\frac{1}{\lambda_i} : i = 1, \ldots, n+1\right\}.$$

Оскільки $0 \in M$, $g_i = \lambda_i$, $x_i \in M$, то
$$\lambda_0 g_i = (1 - \lambda_0\lambda_i) \cdot 0 + (\lambda_0\lambda_i)x_i \in M,$$
Оскільки $\lambda_0\lambda_i \leqslant 1$ за означенням λ_0. Тому $\lambda_0 S \subseteq M$. Отже,
$$\lambda_0(g_0 + \varepsilon_1 B) \subseteq \lambda_0 S \subseteq M.$$
Покладемо тепер
$$Q = \operatorname{cone}(g_0 + \varepsilon_1 B) = \{g : g = \gamma(g_0 + \varepsilon_1 u), u \in B, \gamma > 0\}.$$
Очевидно, що $g_0 \in \operatorname{int} Q$, $Q \subseteq K(0,M)$. Так як $0 \notin (g_0 + \varepsilon B)$, і $\varepsilon_1 \leqslant \varepsilon$, то
$$\delta = \min_u \{\|g_0 + \varepsilon_1 u\| : u \in B\} > 0.$$
Тому для $g \in Q$ виконується співвідношення
$$\|g\| = \gamma\|g_0 + \varepsilon_1 u\| \geqslant \gamma\delta,$$
або
$$\gamma \leqslant \frac{1}{\delta}\|g\|.$$
Оберемо тепер ε_2 настільки малим, щоб $\varepsilon_2/\delta < \lambda_0$. Тоді, якщо $g \in Q$, $\|g\| < \varepsilon_2$, то
$$g = \gamma(g_0 + \varepsilon_1 u), \quad \gamma \leqslant \frac{\|g\|}{\delta} \leqslant \frac{\varepsilon_2}{\delta} \leqslant \lambda_0.$$
Тому
$$g = \gamma(g_0 + \varepsilon_1 u) = \left(\frac{\gamma}{\lambda_0}\right)(\lambda_0(g_0 + \varepsilon_1 u)) \in \left(\frac{\gamma}{\lambda_0}\right)(\lambda_0(g_0 + \varepsilon_1 B)) \subseteq \left(\frac{\gamma}{\lambda_0}\right) M,$$
звідки в силу того, що $\gamma < \lambda_0$, $0 \in M$ отримуємо
$$g \in \left(\frac{\gamma}{\lambda_0}\right) M \subseteq M,$$
тобто $g \in M$. Тим самим доведено, що якщо $g \in Q$, $\|g\| < \varepsilon_2$, то $\psi(g) \in M$. Це означає, що $K(0,M)$ – гладке локальне шатро до M в точці ноль. Якщо $g_0 \in K(0,M)$, $g_0 = 0$, то повторюючи попередні міркування, переконуємося, що $\varepsilon_1 B \subseteq M$ для достатньо малих $\varepsilon_1 > 0$, а $K(0,M) = X$. Тому, поклавши в данному випадку $Q = X$ для g, $\|g\| < \varepsilon_1$, отримаємо, що $\psi(g) = g \in M$, що закінчує доведення.

Важливою властивістю шатер є те, що при достатньо загальних припущеннях перетин локальних шатер є шатром.

Теорема 3.1.2. *Нехай $M_i, i \in I$, – множини в $\mathbb{R}^n$, I – скінченна множина індексів, $x_0 \in M = \bigcap_{i \in I} M_i$ і нехай $K(x_0,M_i), i \in I$, – локальні шатра множин M_i в точці x_0. Якщо шатра $K(x_0,M_i), i \in I$, нероздільні, то для довільного вектора $\overline{x}_0 \in \operatorname{ri} K$, де*
$$K = \bigcap_{i \in I} K(x_0,M_i)$$

існує такий конус Q і така функція $\psi(\overline{x})$, що:

а) $\overline{x}_0 \in \operatorname{ri} Q, \operatorname{Lin} Q = \operatorname{Lin} K$ *і* $Q \subseteq K$*;*

б) $\psi(\overline{x}) = \overline{x} + r(\overline{x})$*,* $\|\overline{x}\|^{-1} r(\overline{x}) \to 0$ *при* $\overline{x} \to 0$*;*

в) $x_0 + \psi(\overline{x}) \in M$ *для* $\overline{x} \in Q \cap (\varepsilon B)$ *при деякому* $\varepsilon > 0$.

Зауваження 3.1.2. Функція $\psi(\overline{x})$ може не бути неперервною. Тому K, взагалі кажучи, не є локальним шатром для M в точці x_0.

Доведення. Обмежимося випадком двох множин M_1 і M_2. Для спрощення будемо вважати, що $x_0 = 0$ і позначимо $K_i = K(x_0, M_i), i = 1,2$. Розглянемо конус $K = K_1 \cap K_2$. Так як конуси K_1 і K_2 нероздільні, то $\operatorname{ri} K_1 \cap \operatorname{ri} K_2 \neq \emptyset$ і за теоремою 1.3.3

$$\operatorname{Lin} K = \operatorname{Lin} K_1 \cap \operatorname{Lin} K_2, \quad \operatorname{ri} K = \operatorname{ri}, K_1 \cap \operatorname{ri} K_2. \tag{3.1.25}$$

Крім того, за теоремою 1.5.4

$$\operatorname{Lin} K_1 + \operatorname{Lin} K_2 = \mathbb{R}^n \tag{3.1.26}$$

Нехай тепер $x_0 \in \operatorname{ri} K$. Згідно формулам (3.1.25) маємо $\overline{x}_0 \in \operatorname{ri} K_i, i = 1,2$. Оскільки K_i– локальне шатро для M_i в точці нуль, то існують такі конуси Q_i, що

$$x_0 \in \operatorname{ri} Q_i, \quad \operatorname{Lin} Q_i = \operatorname{Lin} K_i, \quad Q_i \subseteq K_i, i = 1,2.$$

Крім того, існують функції

$$\psi_i(\overline{x}) = \overline{x} + r_i(\overline{x}), \quad \|x\|^{-1} r_i(\overline{x}) \to 0, \overline{x} \to 0$$

такі, що

$$\psi_i(\overline{x}) \in M_i, \quad \overline{x} \in Q_i \cap (\varepsilon B) \tag{3.1.27}$$

Нехай тепер вектори $e_1, \ldots, e_k$ утворюють базис в $\operatorname{Lin} Q_1 = \operatorname{Lin} K_1$, а вектори $e_{k+1}, \ldots, e_n$ належать $\operatorname{Lin} Q_2 = \operatorname{Lin} K_2$ і доповнюють базис $e_1, \ldots, e_k$ до базиса у всьому $\mathbb{R}^n$. Це можливе за формулою (3.1.26).

Складемо систему рівнянь, записавши її у векторні формі:

$$g(\overline{x}, y) = \psi_1(y^1 e_1 + \cdots + y^k e_k + \overline{x}) - \psi_2(\overline{x} - y^{k+1} e_{k+1} - \cdots - y^n e_n) = 0. \tag{3.1.28}$$

Невідомим тут є вектор y з компонентами $y^i, i = 1, \ldots, n$, а вектор $\overline{x}$ є параметром. Якщо позначити через $\mathcal{E}$ матрицю із стовпчиками $e_i, i = 1, \ldots, n$, то має місце формула

$$g(\overline{x}, y) = \mathcal{E} y + r_1(\overline{x} + \sum_{i=1}^{k} y^i e_i)) - r_2(\overline{x} - \sum_{i=k+1}^{n} y^i e_i)), \tag{3.1.29}$$

так як $\psi_i(\overline{x}) = \overline{x} + r_i(\overline{x})$. В силу властивостей функції r_1

$$\frac{r_1(\overline{x} + \sum\limits_{i=1}^{k} y^i e_i))}{\sqrt{\|\overline{x}\|^2 + \|y\|^2}} \to 0, \tag{3.1.30}$$

коли x і y прямують до нуля. Аналогічні властивості має функція r_2. Тому

$$\|g(\overline{x},y) - \mathcal{E}y\| \leqslant \overline{r}\sqrt{\|\overline{x}\|^2 + \|y\|^2}.$$

Оскільки $e_1,\dots,e_n$– базис в $\mathbb{R}^n$, то матриця $\mathcal{E}$ невироджена і можна застосувати теорему 3.1.1, з якої випливає, що існує така $y(\overline{x})$, що задовольняє системі (3.1.28) і така, що

$$\frac{y(\overline{x})}{\|\overline{x}\|} \to 0. \tag{3.1.31}$$

Покладемо тепер

$$\psi(\overline{x}) = \psi_1(\overline{x} + \sum_{i=1}^{k} y^i(\overline{x})e_i) = \psi_2(\overline{x} - \sum_{i=k+1}^{n} y^i(\overline{x})e_i). \tag{3.1.32}$$

Тоді

$$\psi(\overline{x}) = \psi_1(\overline{x} + \sum_{i=1}^{k} y^i(\overline{x})e_i) = \overline{x} + \sum_{i=1}^{n} y^i(\overline{x})e_i + r_1(\overline{x} + \sum_{i=1}^{n} y^i(\overline{x})e_i) = \overline{x} + r(\overline{x}), \tag{3.1.33}$$

де

$$r(\overline{x}) = \sum_{i=1}^{n} y^i(\overline{x})e_i + r_1(\overline{x} + \sum_{i=1}^{n} y^i(\overline{x})e_i).$$

В силу (3.1.31) і того, що $\|x\|^{-1}r_1(\overline{x}) \to 0$ при $\overline{x} \to 0$, $r(\overline{x})$ має таку ж властивість.

Припустимо, що $\overline{x}_0 \neq 0$. Оскільки $\overline{x}_0 \in \operatorname{ri} Q_i$, то існує таке $\varepsilon > 0$, що $x_0 + \operatorname{Lin} Q_i \cap (\varepsilon_0 B) \subseteq Q_i$. Розглянемо конус

$$K_\delta = \left\{ \overline{x} : \langle \overline{x},\overline{x}_0 \rangle > ||\overline{x}|| \cdot ||\overline{x}_0||(1 - \frac{\delta^2}{2||x_0||^2}) \right\}. \tag{3.1.34}$$

Неважко помітити, що якщо $\overline{x} \in K_\delta$, то

$$\left\| \overline{x}\frac{||\overline{x}_0||}{||\overline{x}||} - \overline{x}_0 \right\| < \delta. \tag{3.1.35}$$

Тому, якщо $\overline{x} \in \operatorname{Lin} Q_i \cap K_{\varepsilon_0}$, то в силу нерівності (3.1.35) отримуємо, що

$$\frac{\|\overline{x}_0\|}{\|\overline{x}\|}\overline{x} = \overline{x}_0 + \left(\frac{\|\overline{x}_0\|}{\|\overline{x}\|}\overline{x} - x \right) \in \overline{x}_0 + \operatorname{Lin} Q_i \cap (\varepsilon_0 B),$$

і тому $\overline{x} \in Q_i$.

Нехай тепер

$$Q = Q_1 \cap Q_2 \cap K_{\frac{1}{2}\varepsilon_0}.$$

Введемо наступні позначення:

$$z_1(\overline{x}) = \overline{x} + \sum_{i=1}^{k} y^i(\overline{x})e_i,\ z_2(\overline{x}) = \overline{x} - \sum_{i=k+1}^{n} y^i(\overline{x})e_i.$$

В силу рівності (3.1.32)

$$\psi(\overline{x}) = \psi_1(z_1(\overline{x})) = \psi_2(z_2(\overline{x})). \tag{3.1.36}$$

виберемо $\varepsilon_1 > 0$ настільки малим, щоб з умов $\|\overline{x}\| < \varepsilon_1$, $\overline{x} \in Q$ випливало, що

$$\|z_i(\overline{x})\| < \varepsilon, \quad z_i(\overline{x}) \in K_{\varepsilon_0}. \tag{3.1.37}$$

Із співвідношень (3.1.31) і (3.1.34) отримаємо можливість такого вибору. Зауважимо тепер, що так як вектори $e_i, i = 1, \ldots, k$ належать $\mathrm{Lin}\, Q_1$, то $z_1(\overline{x}) \in \mathrm{Lin}\, Q_1$ для $\overline{x} \in Q$.
Аналогічно, якщо $\overline{x} \in Q$, то $z_2(\overline{x}) \in \mathrm{Lin}\, Q_2$. Тому для

$$\overline{x} \in Q \cap (\varepsilon_1 B)$$

з формул (3.1.37) і (3.1.27) випливає, що

$$z_i(\overline{x}) \in \mathrm{Lin}\, Q_i \cap K_\varepsilon, \quad z_i(\overline{x}) \in Q_i \cap (\varepsilon B),$$

$$\psi_i(z_i(\overline{x})) \in M_i, i = 1,2.$$

Але тоді рівність (3.1.36) показує, що $\psi(\overline{x}) \in M_1 \cap M_2$ для $\overline{x} \in Q \cap (\varepsilon_1 B)$, тобто для $\overline{x}_0 \in \mathrm{ri}\, K, x_0 \neq 0$, побудовані множина Q і функція ψ, що задовольняють умовам теореми.
Якщо $x_0 = 0$, то $\mathrm{Lin}\, Q_i \cap (\varepsilon B) \subseteq Q_i$ при достатньо малому $\varepsilon_0 > 0$, а звідси одразу випливає, що $Q_i = \mathrm{Lin}\, Q_i$, тобто Q_i– підпростір. Причому, так як $\mathrm{Lin}\, K_i = \mathrm{Lin}\, Q_i$, то $Q_i = \mathrm{Lin}\, K_i$.
Покладемо

$$Q = Q_1 \cap Q_2 = \mathrm{Lin}\, Q_1 \cap \mathrm{Lin}\, Q_2$$

і оберемо $\varepsilon_1 > 0$ настільки малим, щоб з $\|\overline{x}\| < \varepsilon_1$ випливала нерівність $\|z_i(\overline{x})\| < \varepsilon$. Тоді для $\overline{x} \in Q \cap (\varepsilon_1 B)$ виконується включення

$$z_i(\overline{x}) \in Q \cap (\varepsilon B), i = 1,2,$$

і формули (3.1.27) показують, що

$$\psi(z_i(\overline{x})) \in M_i, i = 1,2.$$

Тепер з рівності (3.1.36) отримуємо співвідношення

$$\psi(\overline{x}) = \psi_1(z_1(\overline{x})) = \psi_2(z_2(\overline{x})) \in M_1 \cap M_2.$$

Теорему доведено. □

Теорема 3.1.3. *Нехай виконані умови теореми 3.1.2 і принаймі один з конусів $K_{M_i}(x_0)$ не є підпростором. Тоді існує точка $x_1 \in M$, яка відмінна від x_0.*

Доведення. Нехай $x_1 \in M$ і $K_i \equiv K_{M_i}(0)$. Оскільки конуси K_i нероздільмі, то

$$\mathrm{ri}\, K = \bigcap_{i \in I} \mathrm{ri}\, K_i \neq \emptyset, \quad K = \bigcap_{i \in I} K_i.$$

Припустимо, що конус $K_j, j \in I$, не є підпростором. Тоді $0 \notin \operatorname{ri} K_j$, оскільки з умови $0 \in \operatorname{ri} K_j$ випливає рівність $K_j = \operatorname{Lin} K_j$. Оскільки $\operatorname{ri} K \neq \emptyset$, то можна вказати напрямок $\overline{x}_0 \in \operatorname{ri} K$ такий, що $\overline{x}_0 \neq 0$.
За попередньою теоремою існує такий конус Q, що

$$\overline{x}_0 \in \operatorname{ri} Q, Q \subseteq K, \operatorname{Lin} K = \operatorname{Lin} Q,$$

і така функція $\psi(\overline{x}) = \overline{x} + r(\overline{x})$, що $\|\overline{x}\|^{-1} r(\overline{x}) \to 0$ при $\overline{x} \to 0$ і

$$\psi(\overline{x}) \in M, \overline{x} \in Q \cap (\varepsilon_1 B), \varepsilon > 0.$$

Оскільки $\|\overline{x}\|^{-1} r(\overline{x}) \to 0$ при $\overline{x} \to 0$, то $\psi(\overline{x}) \neq 0$ при достатньо малих $\overline{x}$. Тому, при достатньо малих $\overline{x}$, що належать Q, $\psi(\overline{x}) \in M$, $\psi(\overline{x}) \neq 0$, що і потрібно було довести. □

Теорема 3.1.4. *Нехай виконані всі умови теореми 3.1.2 і нехай, крім того, $K_{M_i}(x_0)$ – гладкі локальні шатра, тобто відповідні функції $\psi_i(\overline{x})$, що фігурують в означенні 3.1.3 локального шатра, неперервно диференційовні в околі початку координат. Тоді конус*

$$K = \bigcap_{i \in I} K_{M_i}(x_0)$$

є гладким локальним шатром до M в точці x_0.

Доведення. Доведення майже повністю співпадає з доведенням теореми 3.1.2. Однак, в силу зауваження до теореми 3.1.1 можна стверджувати, що побудована при доведенні теореми 3.1.2 функція $\psi(\overline{x})$ гладка, і тому K – локальне шатро. □

3.2. ФУНКЦІЇ, ЩО ДОПУСКАЮТЬ ВЕРХНЮ ОПУКЛУ АПРОКСИМАЦІЮ

3.2.1. Означення верхньої опуклої апроксимації та субдиференціала

Нехай $f(x)$, $x \in X$, функція, що приймає скінченні значення і значення $\pm\infty$. Покладемо

$$\operatorname{dom} f = \{x : |f(x)| < +\infty\}.$$

Означення 3.2.1. Нехай $x \in \operatorname{dom} f$, $g \in X$, $g \neq 0$. Величину

$$F(x,g) = \sup_{r(\cdot)} \limsup_{\lambda \downarrow 0} \frac{f(x + \lambda g + r(\lambda)) - f(x)}{\lambda},$$

де зовнішня верхня грань береться по всім функціям $r(\lambda) \in X$ таким, що $\lambda^{-1} r(\lambda) \to 0$ при $\lambda \downarrow 0$, будемо називати похідною за напрямком Пшеничного. Величина $F(x,g)$ – скінченна або нескінченна – існує завжди.

Легко також перевірити, що функція $F(x,g)$ додатньо однорідна за g, тобто для $\lambda > 0$

$$F(x,\lambda g) = \lambda F(x,g).$$

Зауваження 3.2.1. Якщо функція $f(x)$ в околі точки x задовольняє умову Ліпшиця, то

$$|f(x+\lambda g + r(\lambda)) - f(x+\lambda g)| \leqslant L\|r(\lambda)\|.$$

Тому

$$\frac{f(x+\lambda g + r(\lambda)) - f(x+\lambda g)}{\lambda} \to 0,$$

$$F(x,g) = \sup_{r(\cdot)} \limsup_{\lambda\downarrow 0} [\frac{f(x+\lambda g) - f(x)}{\lambda} + \frac{f(x+\lambda g + r(\lambda)) - f(x+\lambda g)}{\lambda}] =$$

$$= \limsup_{\lambda\downarrow 0} \frac{f(x+\lambda g) - f(x)}{\lambda},$$

тобто для функцій, що задовольняють умову Ліпшиця, справедлива формула

$$F(x,g) = \limsup_{\lambda\downarrow 0} \frac{f(x+\lambda g) - f(x)}{\lambda}.$$

Означення 3.2.2. Функція $h(x,g)$ називається верхньою опуклою апроксимацією функції $f(x)$ в точці x, якщо:

1) $h(x,g) \geqslant F(x,g)$ для всіх $g \neq 0$;

2) $h(x,g)$ – опукла замкнута додатньо однорідна функція аргументу g.

Для спрощення, в подальшому будемо замість "верхня опукла апроксимація" писати просто в.о.а. Зрозуміло, що в.о.а. для функції $f(x)$ в точці x визначена неоднозначно і може існувати багато різних в.о.а.

Лема 3.2.1. *Якщо $h_1(x,g)$ і $h_2(x,g)$ в.о.а. для $f(x)$ в точці x, то функція $h(x,g) = \lambda_1 h_1(x,g) + \lambda_2 h_2(x,g)$, $\lambda_1 + \lambda_2 = 1, \lambda_1, \lambda_2 \geqslant 0$, і функція $h(x,g) = \max\{h_1(x,g),h_2(x,g)\}$ також є в.о.а. для $f(x)$ в точці x.*

Означення 3.2.3. Якщо $h(x,g)$ є в.о.а. для f в точці x, то множина

$$\partial h(x,0) = \{x^* \in X^* : h(x,g) \geqslant \langle x^*,g\rangle\,, g \in X\} \tag{3.2.1}$$

називається субдиференціалом функції f в точці x і позначається $\partial f(x)$.

Субдиференціал $\partial h(x,0)$ опуклої замкнутої додатньо однорідної функції $h(x,g)$ відносно g існує завжди. При цьому

$$\partial h(x,0) = \operatorname{dom} h^*(\cdot,x), \tag{3.2.2}$$

де

$$h^*(x^*,x) = \sup_g \{\langle g,x^*\rangle - h(x,g)\}.$$

З іншого боку, із співвідношення (3.2.2) і теореми 2.5.2 випливає, що

$$h(x,g) = \sup_{x^*}\{\langle x^*,g\rangle : x^* \in \partial h(x,0)\},$$

або, з урахуванням означення 3.2.3,

$$h(x,g) = \sup_{x^*}\{\langle x^*,g\rangle : x^* \in \partial f(x)\}. \tag{3.2.3}$$

При цьому субдиференціал $\partial f(x)$ в формулі (3.2.3) обчислений для функції $h(x,g)$. З формули (3.2.3) також видно, що $h(x,g)$ є опорною функцією множини $\partial f(x)$.
Формули (3.2.1) та (3.2.3) показують, що функція $h(x,g)$ і субдиференціал $\partial f(x)$ однозначно визначають одне одного. Якщо $\partial f(x)$– субдиференціал, то $\partial f(x)$– замкнута опукла множина, а функція $h(x,g)$, що визначена за допомогою формули (3.2.3), є в.о.а. для f в точці x. Як і в.о.а., субдиференціал визначений неоднозначно.

Лема 3.2.2. *Якщо $h_1(x,g)$ і $h_2(x,g)$ – в.о.а. для $f(x)$ в точці x і $h_1(x,g) \geqslant h_2(x,g)$, то*

$$\partial_1 f(x) \subseteq \partial_2 f(x),$$

де $\partial_1 f(x)$ і $\partial_2 f(x)$– субдиференціали, що визначені $h_1(x,g)$ і $h_2(x,g)$ відповідно.

Доведення. Дійсно, згідно з означенням 3.2.3 і формулою 3.2.3), $x^* \in \partial_2 f(x)$ тоді і тільки тоді, коли

$$\langle x^*,g\rangle \leqslant h_2(x,g), \quad g \in X.$$

В силу умови леми, звідси випливає, що

$$\langle x^*,g\rangle \leqslant h_2(x,g) \leqslant h_1(x,g), \quad g \in X,$$

тобто $x^* \in \partial_1 f(x)$. □

Приклад 3.2.1. Нехай $x \in \mathbb{R}^1$,

$$f_c(x) = \begin{cases} 0, & x < 0, \\ c, & x \geqslant 0. \end{cases}$$

Неважко підрахувати, що для $c > 0$

$$F(0,g) = \begin{cases} -\infty, & g < 0, \\ 0, & g > 0. \end{cases}$$

Тому будь-яка функція

$$h(0,g) = ag, \quad a \geqslant 0,$$

є в.о.а. для $f_c(x)$ в точці $x = 0$. Відповідно $\partial f_c(0) = \{a\}, a \geqslant 0$. Отже, якщо $c > 0$, то $f_c(x)$ в точці нуль має цілу множину субдиференціалів кожний з яких складається з єдиного числа $a \geqslant 0$.
Якщо $c < 0$, то

$$F(0,g) = \begin{cases} +\infty, & g < 0, \\ 0, & g > 0. \end{cases}$$

Покладемо

$$h_0(0,g) = \begin{cases} +\infty, & g < 0, \\ 0, & g \geqslant 0. \end{cases}$$

Функція $h_0(0,g)$ опукла, замкнута і додатньо однорідна, так що h_0 є в.о.а. для $f_c(x), c < 0$, в точці нуль. Просте обчислення показує, що

$$\partial f_c(0) \equiv \partial h_0(0,0) = (-\infty,0].$$

Отже, множина $(-\infty,0]$ є субдиференціалом $f_c(x)$ в точці нуль. Так як

$$h_0(0,g) = F(0,g), \quad g \neq 0,$$

то для довільної в.о.а. h буде виконуватись нерівність $h \geqslant h_0$. З леми 3.2.2 випливає, що

$$\partial f_c(0) \supseteq (-\infty,0]$$

для довільного субдиференціала $\partial f_c(0)$.

Приклад показує, що субдиференціали можуть існувати навіть для розривних функцій.

З означень випливає, що $\partial(cf(x)) = c\partial f(x)$ для $c \geqslant 0$.

Теорема 3.2.1. *Нехай $f(x) = f_1(x) + f_2(x)$, $h_1(x,g)$ і $h_2(x,g)$– в.о.а. для $f_1(x)$ і $f_2(x)$ в точці x. Тоді*

$$h(x,g) = h_1(x,g) + h_2(x,g)$$

є в.о.а. для $f(x)$ в точці x. Якщо при цьому

$$\operatorname{int} \operatorname{dom} h_1(x,\cdot) \cap h_2(x,\cdot) \neq \emptyset,$$

то

$$\partial f(x) = \partial f_1(x) + \partial f_2(x).$$

Доведення. За означенням

$$F(x,g) =$$

$$= \sup_{r(\cdot)} \limsup_{\lambda \downarrow 0} \left[\frac{f_1(x + \lambda g + r(\lambda))}{\lambda} + \frac{f_2(x + \lambda g + r(\lambda)) - f_1(x) - f_2(x)}{\lambda} \right] \leqslant$$

$$\leqslant \sup_{r(\cdot)} \limsup_{\lambda \downarrow 0} \frac{f_1(x + \lambda gx + r_1(\lambda)) - f_1(x)}{\lambda} +$$

$$+\sup_{r(\cdot)} \limsup_{\lambda \downarrow 0} \frac{f_2(x+\lambda g + r_2(\lambda)) - f_2(x)}{\lambda} =$$

$$= F_1(x,g) + F_2(x,g).$$

Оскільки $h_1 \geqslant F_1$, $h_2 \geqslant F_2$, то $h \geqslant F$. Крім того, функція h опукла додатньо однорідна і замкнута як сума опуклих додатньо однорідних і замкнутих функцій. Зі сказаного випливає, що $h(\overline{x},0)$ є в.о.а. для f в точці x і перша частина теореми доведена.

На основі теорем 2.7.8 і 2.2.3 маємо

$$\partial h(x,0) = \partial h_1(x,0) + \partial h_2(x,0),$$

тобто

$$\partial f(x) = \partial f_1(x) + \partial f_2(x).$$

□

3.2.2. Класи функцій, що допускають верхню опуклу апроксимацію

Теорема 3.2.2. *Нехай для даної функції $f(x)$ в точці x функція $F(x,g)$ опукла за g і замкнута, якщо покласти $F(x,0) = 0$. Тоді $\partial F(x,0)$ є субдиференціалом функції f в точці x і для довільного іншого субдиференціала $\partial f(x)$ виконується включення*

$$\partial f(x) \supseteq \partial F(x,0).$$

Доведення випливає з того, що в припущеннях теореми $F(x,g)$ є в.о.а. і для довільної іншої в.о.а. h виконується нерівність $h \geqslant F$.

Наслідок 3.2.1. *Якщо функція $f(x)$ неперервно диференційовна в точці x, то*

$$h(x,g) = \langle f'(x),g\rangle$$

є в.о.а. При цьому $\partial f(x) = \{f'(x)\}$.

Наслідок 3.2.2. *Якщо f – неперервна в точці x опукла функція, то*

$$h(x,g) = f'(x,g) = \lim_{\lambda \downarrow 0} \frac{f(x+\lambda g) - f(x)}{\lambda}$$

є в.о.а., а звичайний субдиференціал опуклої функції є субдиференціалом в сенсі означення 3.2.3.

Доведення. Оскільки за теоремою 2.2.2 неперервна опукла функція задовольняє умові Ліпшиця, то легко перевірити, що

$$F(x,g) = f'(x,g).$$

Далі, за теоремами 2.7.2, 2.7.5 $\partial f(x)$ є обмеженою замкнутою опуклою множиною і
$$f'(x,g) = \max_{x^*}\{\langle x^*,g\rangle : x^* \in \partial f(x)\}.$$
Оскільки $f'(x,g)$ є верхня грань лінійних (а отже неперервних і тим більш замкнутих) функцій, то $f'(x,g)$ є замкнутою функцією g. Застосування теореми 2.7.11 тепер показує, що
$$\partial F(x,0) = \partial f'(x,0) = \partial f(x).$$
Отже, $\partial F(x,g)$ є опуклою додатньо однорідною замкнутою функцією g, тобто вона задовольняє умовам теореми 3.2.2, а її субдиференціал співпадає зі звичайним субдиференціалом опуклої функції f. □

Теорема 3.2.3. *Нехай f – неперервна в точці x угнута функція. Тоді довільна функція вигляду*
$$h(x,g) = -\langle x^*,g\rangle\,, \quad x^* \in \partial(-f(x))$$
є в.о.а. для f в точці x, а $\{-x^*\}$,$x^* \in \partial(-f(x))$, є субдиференціалом f в точці x. Тут $\partial(-f(x))$ означає звичайний субдиференціал опуклої функції $-f$.

Доведення. За означенням угнутості функція $f_0(x) = -f(x)$ опукла. Оскільки функція $f_0(x)$ задовольняє умову Ліпшиця, то $f(x)$ також задовольняє умову Ліпшиця. Тому
$$F(x,g) = \lim_{\lambda\downarrow 0}\frac{f(x+\lambda g)-f(x)}{\lambda} = f'(x,g).$$
Але тоді
$$F(x,g) = f'(x,g) = -f_0'(x,g) =$$
$$= -\max_{x^*}\{\langle x^*,g\rangle : x^* \in \partial f_0(x)\} = \min_{x^*}\{-\langle x^*,g\rangle : x^* \in \partial f_0(x)\},$$
отже
$$F(x,g) \leqslant -\langle x^*,g\rangle\,, \quad x^* \in \partial f_0(x).$$
Тому $h(x,g) = -\langle x^*,g\rangle$ є в.о.а. Так як
$$\partial h(x,0) = \{-x^*\},$$
то $\{-x^*\}$ – субдиференціал, що і потрібно було довести. □

Теорема 3.2.4. *Нехай I – довільна множина індексів і нехай при кожному $i \in I$ функція $f_i(x)$ має в точці x в.о.а. $h_i(x,g)$. Нехай*
$$f(x) = \inf_i\{f_i(x) : i \in I\},$$
$$I(x) = \{i \in I : f_i(x) = f(x)\}.$$
Тоді $h_i(x,g)$,$i \in I(x)$, є в.о.а. функції f в точці x, а $\partial f_i(x)$ – субдиференціал $f_i(x)$, що відповідає $h_i(x,g)$,$i \in I(x)$,– одночасно є субдиференціалом функції f в x.

Доведення. Оскільки

$$\frac{f(x+\lambda g+r(\lambda))-f(x)}{\lambda}\leqslant\frac{f_i(x+\lambda g+r(\lambda))-f_i(x)}{\lambda},\ i\in I(x),$$

то

$$F(x,g)\leqslant F_i(x,g),\quad i\in I(x).$$

Тому $h_i(x,g),i\in I(x)$, є в.о.а., а $\partial f_i(x)$ - субдиференціал функції f. □

Теорема 3.2.5. *Нехай A – компактна множина і для довільного $\alpha\in A$ визначена функція $f(x,\alpha)$. Покладемо*

$$f(x)=\sup_{\alpha}\{f(x,\alpha):\alpha\in A\}. \tag{3.2.4}$$

Нехай функції $f(x,\alpha)$ неперервні за сукупністю аргументів, коли x змінюється в деякому околі точки x_0, а $\alpha\in A$. Нехай, крім того, при кожному $\alpha\in A$ існують похідні за напрямком g в точці x_0, тобто визначені $f'(x_0,g,\alpha)$, причому відношення

$$\frac{f(x_0+\lambda g,\alpha)-f(x_0,\alpha)}{\lambda} \tag{3.2.5}$$

прямує до $f'(x_0,g,\alpha)$ рівномірно за $\alpha\in A$ при $\lambda\downarrow 0$. Тоді функція $f(x)$, що визначена формулою (3.2.4), диференційовна за напрямком g і

$$f'(x_0,g)=\max_{\alpha}\{f'(x_0,g,\alpha):\alpha\in A(x_0)\},$$

де $A(x_0)=\{\alpha\in A:f(x,\alpha)=f(x)\}$.

Доведення. Позначимо

$$\gamma(\lambda,\alpha)=\frac{f(x_0+\lambda g,\alpha)-f(x_0,\alpha)}{\lambda}-f'(x_0,g,\alpha). \tag{3.2.6}$$

Оскільки при будь-якому $\lambda>0$ відношення (3.2.5) неперервне за α і збігається до $f'(x_0,g,\alpha)$ рівномірно на компакті A, то $f'(x_0,g,\alpha)$– неперервна функція α.

Нехай тепер $x(\lambda)=x_0+\lambda g$. Легко перевірити, що має місце нерівність

$$\frac{f(x(\lambda),\alpha)-f(x_0,\alpha)}{\lambda}\leqslant\frac{f(x(\lambda))-f(x_0)}{\lambda}\leqslant\frac{f(x(\lambda),\alpha_0)-f(x_0,\alpha_0)}{\lambda} \tag{3.2.7}$$

для довільних $\alpha\in A(x(\lambda))$, $\alpha_0\in A(x_0)$. З правої нерівності, спрямувавши λ до нуля, отримуємо

$$\liminf_{\lambda\downarrow 0}\frac{f(x(\lambda))-f(x_0)}{\lambda}\geqslant f'(x_0,g,\alpha),$$

або, оскільки α_0 – довільний елемент $A(x_0)$, то

$$\liminf_{\lambda\downarrow 0}\frac{f(x(\lambda))-f(x_0)}{\lambda}\geqslant\max_{\alpha}\{f'(x_0,g,\alpha):\alpha\in A(x_0)\}. \tag{3.2.8}$$

Оберемо тепер довільну відкриту підмножину C множини A, яка містить $A(x_0)$, $C \supseteq A(x_0)$, і покажемо, що

$$\inf_{C \supseteq A(x_0)} \sup_{\alpha \in C} f'(x_0,g,\alpha) = \max_{\alpha \in A(x_0)} f'(x_0,g,\alpha). \tag{3.2.9}$$

Оскільки $f'(x_0,g,\alpha)$ неперервна за α функція, то для довільного $C \supseteq A(x_0)$ виконується нерівність

$$\sup_{\alpha \in C} f'(x_0,g,\alpha) \geqslant \max_{\alpha \in A(x_0)} f'(x_0,g,\alpha)$$

і тому ліва частина співвідношення (3.2.9) завжди не менше правої. З іншого боку, $A(x_0)$ є замкнутою підмножиною компактної множини A і тому множина $A(x_0)$ компактна. Оберемо $\varepsilon > 0$ і поставимо у відповідність кожному $\alpha \in A(x_0)$ окіл U_α так, щоб

$$|f'(x_0,g,\overline{\alpha}) - f'(x_0,g,\alpha)| < \varepsilon, \ \overline{\alpha} \in U_\alpha. \tag{3.2.10}$$

Оскільки $A(x_0)$ – компактна множина, а об'єднання відкритих множин U_α покриває $A(x_0)$, то існує скінченний набір множин U_α, наприклад, $U_{\alpha_1}, \ldots U_{\alpha_m}$, що покривають $A(x_0)$. Позначимо

$$C_\varepsilon = \bigcup_{i=1}^{m} U_{\alpha_i}.$$

Для будь-якої точки C_ε знайдеться такий номер i, що $\alpha \in U_{\alpha_i}$. В силу нерівності (3.2.10) можно записати

$$f'(x_0,g,\alpha) \leqslant f'(x_0,g,\alpha_i) + \varepsilon \leqslant \max_{i=1,\ldots,m} f'(x_0,g,\alpha_i) + \varepsilon \leqslant$$

$$\leqslant \max_{\alpha \in A(x_0)} f'(x_0,g,\alpha_i) + \varepsilon.$$

Тому

$$\inf_{C \supseteq A(x_0)} \sup_{\alpha \in C_\varepsilon} f'(x_0,g,\alpha) \leqslant \sup_{\alpha \in C_\varepsilon} f'(x_0,g,\alpha)$$

$$\leqslant \max_{\alpha \in A(x_0)} f'(x_0,g,\alpha) + \varepsilon.$$

Оскільки $\varepsilon > 0$ довільне, то виходить, що ліва частина співвідношення (3.2.9) не більша за праву. Співставляючи це з отриманою раніше протилежною нерівністю, отримуємо рівність (3.2.9).

Для закінчення доведення теореми скористаємося лемою 2.7.6, в силу якої при достатньо малому λ справедливе включення $A(x(\lambda)) \subseteq C$. Тому з лівої нерівності (3.2.7) випливає, що

$$\frac{f(x(\lambda)) - f(x_0)}{\lambda} \leqslant \sup_{\alpha \in C} \frac{f(x(\lambda),\alpha) - f(x_0,\alpha)}{\lambda}$$

$$= \sup_{\alpha \in C} [f'(x_0,g,\alpha) + \gamma(\lambda,\alpha)].$$

За припущенням теореми $\gamma(\lambda,\alpha) \leqslant r(\lambda), r(\lambda) \to 0$. Оскільки співвідношення (3.2.5) прямує до $f'(x_0,g,\alpha)$ рівномірно по $\alpha \in A$, то

$$\frac{f(x(\lambda)) - f(x_0)}{\lambda} \leqslant \sup_{\alpha \in C} f'(x_0,g,\alpha) + r(\lambda).$$

Перейшовши до границі спочатку по $\lambda \downarrow 0$, а потім беручи нижню грань по всім $C \supseteq A(x_0)$, отримаємо, використовуючи рівність (3.2.9)

$$\limsup_{\lambda \downarrow 0} \frac{f(x(\lambda)) - f(x_0)}{\lambda} \leqslant \max_{\alpha \in A(x_0)} f'(x_0,g,\alpha). \tag{3.2.11}$$

Порівнюючи формули (3.2.8) і (3.2.11), отримаємо доведення теореми. □

Теорема 3.2.6. *Нехай A – компакт, функції $f(x,\alpha)$, $\alpha \in A$ диференційовні за x в околі точки x_0 і градієнт $f'_x(x,\alpha)$ неперервний за сукупністю аргументів x і α. Тоді для функції $f(x)$, що визначається формулою (3.2.4), виконується співвідношення*

$$f'(x_0,g) = \max_{\alpha} \{ \langle f'_x(x_0,\alpha), g \rangle : \alpha \in A(x_0) \}. \tag{3.2.12}$$

Більш того, $f'(x_0,g)$ є в.о.а. для $f(x)$ в точці x_0, і відповідний субдиференціал може бути заданий формулою

$$\partial f(x_0) = \operatorname{conv}(\bigcup_{\alpha \in A(x_0)} f'_x(x_0,\alpha)) = \operatorname{conv} f'_x(x_0, A(x_0)). \tag{3.2.13}$$

Доведення. Оскільки $f'_x(x_0,\alpha)$ неперервно залежить від x в околі x_0 і $\alpha \in A$, причому A – компакт, то

$$\| f'_x(x,\alpha) - f'_x(x_0,\alpha) \| \leqslant \varepsilon(\| x - x_0 \|),$$

де функція $\varepsilon(\lambda)$ монотонно спадаючи прямує до нуля при $\lambda \downarrow 0$.
Використовуючи теорему про середнє, отримуємо, що

$$\frac{f(x_0 + \lambda g,\alpha) - f(x_0,\alpha)}{\lambda} = \langle g, f'(x_0 + \theta_\alpha \lambda g, \alpha) \rangle , \quad 0 \leqslant \theta_\alpha \leqslant 1.$$

Оскільки в даному випадку

$$f'(x_0,g,\alpha) = \langle g, f'(x_0,\alpha) \rangle , \tag{3.2.14}$$

то з співвідношення (3.2.6) отримуємо, що

$$|\gamma(\lambda,\alpha)| = | \langle g, f'(x_0 + \theta_\alpha \lambda g, \alpha) - f'(x_0,\alpha) \rangle | \leqslant$$

$$\leqslant \|g\| \cdot \| f'(x_0 + \theta_\alpha \lambda g,\alpha) - f'(x_0,\alpha \| \leqslant$$

$$\leqslant \|g\| \varepsilon(\theta_\alpha \lambda \|g\|) \leqslant \|g\| \varepsilon(\lambda \|g\|).$$

Звідси випливає, що величина $\gamma(\lambda,\alpha)$ рівномірно прямує до нуля при $\lambda \downarrow 0$ і всі умови теореми 3.2.5 виконані. Використовуючи результат цієї теореми і формулу (3.2.14), приходимо до співвідношення (3.2.12).

Зауважимо тепер, що оскільки $f'_x(x,\alpha)$ неперервно залежить від своїх аргументів і A компакт, то

$$\|f'_x(x,\alpha)\| \leqslant L$$

для всіх x з малого околу x_0 і $\alpha \in A$. Використавши знову теорему про середнє значення отримаємо, що

$$|f(x,\alpha) - f(y,\alpha)| = |\,\langle x-y, f'(y+\theta(x-y),\alpha)\rangle\,| \leqslant L\|x-y\|,$$

тобто функція $f(x,\alpha)$ задовольняє в околі x_0 умові Ліпшиця. Але

$$f(x,\alpha) - f(y,\alpha) \leqslant f(x) - f(y) \leqslant f(x,\alpha_0) - f(y,\alpha_0),$$

$$\alpha \in A(y), \quad \alpha_0 \in A(x).$$

Оскільки

$$f(x,\alpha) - f(y,\alpha) \geqslant -L\|x-y\|,$$

$$f(x,\alpha_0) - f(y,\alpha_0) \leqslant L\|x-y\|,$$

то

$$|f(x) - f(y)| \leqslant L\|x-y\|,$$

тобто $f(x)$ також задовольняє умову Ліпшиця. В силу зауваження 3.2.1 і того факту, що $f'(x_0,g)$ існує і визначається формулою (3.2.12), отримуємо

$$F(x_0,g) = f'_x(x_0,g) = \max_{\alpha}\{\langle g, f'_x(x_0,\alpha)\rangle : \alpha \in A(x_0)\}.$$

Останню формулу можна записати інакше:

$$F(x_0,g) = \max_{x^*}\{\langle g, x^*\rangle : x^* \in f'_x(x_0,A(x_0))\}.$$

Зробимо декілька зауважень:

1) множина $f'_x(x_0,A(x_0))$ компактна як неперервний образ компактної множини $A(x_0)$;

2) за теоремою 1.2.3 множина $\operatorname{conv} f'_x(x_0,A(x_0))$ є опуклою і компактною;

3) максимум лінійної функції на множині і на її опуклій оболонці один і той самий.

Тому

$$F(x_0,g) = \max_{x^*}\left\{\langle x^*, g\rangle : x^* \in \operatorname{conv} f'_x(x_0,A(x_0))\right\}.$$

Отже, $F(x_0,g)$ є опуклою додатньо однородною і замкнутою функцією g. Вона навіть неперервна, оскільки з компактності $\operatorname{conv} f'_x(x_0,A(x_0))$ випливає, що $F(x_0,g)$ визначена для всіх g і, отже, за теоремою 1.3.3 неперервна. Зрозуміло, що функцію $F(x_0,g)$ можна розглядати як верхню опуклу апроксимацію функції f в точці x_0.

За теоремою 2.7.11 $x^* \in \partial F(x_0,0)$ тоді і тільки тоді, коли

$$x^* \in \operatorname{conv} f'_x(x_0,A(x_0))$$

і $\langle x^*,0\rangle \in \partial F(x_0,0)$. Але останнє виконується очевидно для всіх x^*, оскільки $F(x_0,0) = 0$. Тому

$$\partial f(x_0) = \partial F(x_0,0) = \text{conv}\, f'_x(x_0,A(x_0),$$

що і потрібно було довести. $\square$

3.2.3 Функція відстані до множини

Нехай M – довільна множина. Розглянемо функцію:

$$d(x|M) = \inf_y\{\|x-y\| : y \in M\}, \tag{3.2.15}$$

де $\|x-y\| = \langle x-y,x-y\rangle^{1/2}$ – звичайна евклідова відстань між точками. Тоді $d(x|M)$ є відстанню від точки x до множини M. Оскільки відстань до множини і до замикання однакова, то вважатимемо множину M замкнутою. Це дозволяє переписати формулу (3.2.15) у наступному вигляді:

$$d(x|M) = \min_y\{\|x-y\| : y \in M\}.$$

Множина

$$M(x) = \{y \in M : \|x-y\| = d(x|M)\}$$

замкнута і обмежена, а тому компактна.

Лема 3.2.3. *Функція $d(x|M)$ задовольняє умові Ліпшиця з константою $L = 1$.*

Доведення. Якщо $y_1 \in M(x_1), y_2 \in M(x_2)$, то

$$\|x_1-y_1\| - \|x_2-y_1\| \leqslant d(x_1|M) - d(x_2|M) \leqslant \|x_1-y_2\| - \|x_2-y_2\|.$$

З нерівності трикутника випливає, що

$$\|x_1-y_2\| - \|x_2-y_2\| \leqslant \|x_1-x_2\|,$$

$$\|x_2-y_1\| - \|x_1-y_1\| \leqslant \|x_1-x_2\|,$$

тому

$$|d(x_1|M) - d(x_2|M)| \leqslant \|x_1-x_2\|,$$

що і потрібно було довести. $\square$

Теорема 3.2.7. *Якщо $d(x|M) > 0$, то*

$$d'(x,g|M) = \lim_{\lambda\downarrow 0}\frac{d(x+\lambda g|M) - d(x|M)}{\lambda} =$$

$$= \min_y\left\{\frac{\langle x-y,g\rangle}{d(x|M)} : y \in M(x)\right\}.$$

Доведення. Якщо $x \neq y$, то $\|x-y\|$ є диференційовною функцією x, градієнт якої дорівнює

$$\|x-y\|'_x = \frac{x-y}{\|x-y\|}. \tag{3.2.16}$$

Можна вважати, що множина M компактна, тобто обмежена, оскільки, якщо відкинути точки M, що лежать достатньо далеко, то відстань до знову отриманої множини буде співпадати з $d(x|M)$. Записавши $d(x|M)$ у вигляді

$$d(x|M) = -\max_y\{||x-y|| : y \in M\}$$

і скориставшись теоремою 3.2.6, формулою (3.2.12) і виразом (3.2.16) для градієнта, отримаємо твердження теореми. □

Теорема 3.2.8. *Якщо* $d(x|M) > 0$*, то функція*

$$h(x,g) = \frac{\langle x-y,g\rangle}{d(x|M)}, \quad y \in M(x), \tag{3.2.17}$$

є в.о.а. для $d(x|M)$ *в точці* x*, а відповідний субдиференціал дорівнює* $[d(x|M)]^{-1}(x-y)$.

Доведення. Оскільки функція $d(x|M)$ задовольняє умові Ліпшиця, то

$$F(x,g) = d'(x,g|M) = \min_y \left\{ \frac{\langle x-y,g\rangle}{d(x|M)} : y \in M(x) \right\}. \tag{3.2.18}$$

З цього виразу випливає твердження теореми. □

Теорема 3.2.9. *Нехай* $d(x|M) = 0$*, і нехай* $K(x,M)$ *– опуклий конус дотичних напрямків для* M *в точці* x*. Тоді функція*

$$h(x,g) = \inf_y \{\|g-y\| : y \in K(x,M)\} = d(g|K(x,M))$$

є верхньою опуклою апроксимацією для $d(x|M)$ *в точці* x*, а відповідний субдиференціал дорівнює* $B \cap (-K^*(x,M))$.

Доведення. Нехай $y \in K(x,M)$. За означенням існує така функція $r(\lambda)$, $\lambda^{-1}r(\lambda) \to 0$ при $\lambda \downarrow 0$, що $x + \lambda y + r(\lambda) \in M$ при малих λ.
Оскільки $d(x|M)$ задовольняє умову Ліпшиця, то

$$F(x,g) = \limsup_{\lambda \downarrow 0} \frac{d(x+\lambda g|M) - d(x|M)}{\lambda} \leqslant$$

$$\leqslant \limsup_{\lambda \downarrow 0} \frac{||(x+\lambda g) - x + \lambda y + r(\lambda)||}{\lambda} \leqslant$$

$$\leqslant \|g-y\| + \lim_{\lambda \downarrow 0} \frac{\|r(\lambda)\|}{\lambda} = \|g-y\|.$$

Оскільки y - довільний елемент $K(x,M)$, то

$$F(x,g) \leqslant \inf_y \{\|g-y\| : y \in K(x,M)\} = d(g|\overline{K(x,M)}). \qquad (3.2.19)$$

Функція $d(g|\overline{K_m(x)})$ опукла. Ця функція є точна верхня грань лінійних функцій. Тому вона замкнута. Додатня однорідність $d(g|\overline{K_m(x)})$ випливає з того, що $K(x,M)$– конус. В цих умовах формула (3.2.19) показує, що функція $d(g|\overline{K(x,M)})$ є в.о.а. для $d(x,M)$ в точці x.

Обчислимо субдиференціал функції $d(g|\overline{K(x,M)})$ при $g = 0$. Зауважимо, що ця функція співпадає з визначеною для опуклого випадку функцією $d_c(\cdot|\overline{K(x,M)})$, коли $C = B$, тобто C є одиничною кулею. Тому можна скористатись теоремою 2.7.21, яка стверджує, що

$$\partial d(0|\overline{K(x,M)}) = \partial\delta(0|\overline{K(x,M)}) \cap \{x^* : s(x^*|B) \leq 1\}$$

Згідно з формулами (2.7.21) і (2.7.22)

$$s(x^*|B) = \max_x \{\langle x,x^*\rangle : \|x\| \leqslant 1\} = \|x^*\|,$$

$$\partial\delta(0|\overline{K(x,M)}) = -K^*(x,M),$$

тому

$$\partial d(0|\overline{K(x,M)}) = (-K^*(x,M)) \cap \{x^* : \|x^*\| \leqslant 1\},$$

що і потрібно було довести. □

3.2.4 Головні верхні апроксимації та головні субдиференціали

Як показують наведені вище приклади, для даної функції $f(x)$ в точці x може існувати багато верхніх опуклих апроксимацій. Проте, природньо, що якщо $h_1(x,g)$ і $h_2(x,g)$ - в.о.а. функції $f(x)$ в точці x і $h_1(x,g) \geqslant h_2(x,g)$, то $h_1(x,g)$ гірше наближає функцію $f(x)$ в околі точки x. Це мотивує наступне означення.

Означення 3.2.4. В.о.а. $h(x,g)$ функції $f(x)$ в точці x називається головною, якщо не існує іншої в.о.а. $h_1(x,g)$, такої, що

$$h(x,g) \geqslant h_1(x,g) \quad \forall x.$$

Відповідний в.о.а. $h(x,g)$ субдиференціал називається головним.

Теорема 3.2.10. *Якщо $F(x,g)$ після довизначення $F(x,0) = 0$ є опуклою замкнутою функцією, то існує єдиний головний субдиференціал. Якщо f - опукла неперервна функція, то її звичайний субдиференціал є єдиним головним субдиференціалом.*

В подальшому завжди, коли виконані умови теореми 3.2.10, під субдиференціалом розуміють головний субдиференціал.

Теорема 3.2.11. *Якщо* $h(x,g) = \langle x^*,g\rangle$ *є в.о.а. функції* f *в точці* x*, то* $h(x,g)$ *– головна в.о.а., а* x^* *– головний субдиференціал.*

Доведення. Нехай існує в.о.а. $h_1(x,g)$ така, що

$$\langle x^*,g\rangle \geqslant h_1(x,g)$$

для всіх g. Візьмемо $x_1^* \in \partial f(x)$, де $\partial f(x)$ відповідає $h_1(x,g)$. Тоді в силу формули (3.2.3) отримуємо, що

$$\langle x^*,g\rangle \geqslant h_1(x,g) = \langle x_1^*,g\rangle\,.$$

Неважко переконатися, що одна лінійна функція може бути всюду більше іншої лише тоді, коли вони співпадають. Тому

$$\langle x^*,g\rangle = h_1(x,g) = \langle x_1^*,g\rangle\,,$$

тобто $h = h_1$. □

3.3. НЕОБХІДНІ УМОВИ МІНІМУМУ

Побудова необхідних умов экстремуму тісно пов'язана з класами множин і функцій, що беруть участь у задачі. У попередніх параграфах такі класи були введені і досліджені, так що тепер формулювання необхідних умов экстремума може бути дане порівняно просто.

3.3.1. Обмеження, що задаються довільними множинами

Теорема 3.3.1. *Нехай* x_0 *- точка мінімуму функції* $f(x)$ *на множині* M *і нехай* $h(x_0,g)$*– верхня опукла апроксимація* f *у точці* x_0*. Тоді, якщо виконана умова*

$$\operatorname{int}(\operatorname{dom} h(x_0,\cdot)) \cap K(x_0,M) \neq \emptyset,$$

то

$$\partial f(x_0) \cap K^*(x_0,M) \neq \emptyset.$$

Доведення. Оскільки x_0– точка мінімуму, то $h(x_0,g) \geqslant 0$ для всіх $g \in K(x_0,M)$. Справді, якщо $g \in K(x_0,M)$ і $h(x_0,g) < 0$, то існує така функція $r(\lambda)$, $\lambda^{-1}r(\lambda) \to 0$ при $\lambda \downarrow 0$, що

$$x_0 + \lambda g + r(\lambda) \in M.$$

Тому

$$\limsup_{\lambda\downarrow 0} \frac{f(x_0 + \lambda g + r(\lambda)) - f(x_0)}{\lambda} \leqslant F(x_0,g) \leqslant h(x_0,g) < 0,$$

тобто при досить малих $\lambda > 0$, $f(x_0 + \lambda g + r(\lambda)) < f(x_0)$ всупереч тому, що x_0 – точка мінімуму.

Отже, опукла функція $h(x_0,g)$ досягає свого мінімуму на опуклій множині $K(x_0,M)$ в точці $g = 0$. Це можливо лише, якщо

$$\partial f(x_0) \cap K^*(x_0,M) \neq \emptyset,$$

що і треба було довести. □

Наслідок 3.3.1. *Нехай функція $f(x)$ допускає верхню опуклу апроксимацію $h(x_0,g)$ у точці x_0. Тоді для того, щоб точка x_0 була точкою мінімуму функції $f(x)$ необхідно виконання умови $0 \in \partial f(x_0)$.*

Важливо відзначити, що ця умова повинна виконуватися для кожного субдиференціала. Для ілюстрації розглянемо функцію

$$f_c(x) = \begin{cases} 0, & x < 0, \\ c, & x \geqslant 0. \end{cases}$$

Тоді для $c > 0$ будь-яке число $a \geqslant 0$ є субдиференціалом. Отже умова $0 \in \partial f_c(0)$ не виконана. З іншого боку, якщо $c < 0$, то будь-який субдиференціал $\partial f_c(0)$ містить множину $(-\infty,0]$, тому $0 \in \partial f_c(0)$. Цей приклад показує, що отримані необхідні умови мінімуму можуть бути ефективними навіть для розривних функцій.

Теорема 3.3.2. *Нехай x_0 – точка мінімуму функції $f(x)$ на множині*

$$M = \bigcap_{i=1}^{m} M_i.$$

Нехай $f(x)$ допускає в x_0 верхню опуклу апроксимацію $h(x_0,g)$, конуси $K(x_0,M_i)$ є локальними шатрами і

$$\operatorname{int}(\operatorname{dom} h(x_0,\cdot)) \cap \left(\bigcap_{i=1}^{m} K(x_0,M_i) \right) \neq \emptyset.$$

Тоді існує таке число $\lambda \geqslant 0$ і такі вектори $x_i^* \in K^*(x_0,M_i)$, не всі одночасно рівні нулю, що

$$\lambda x_0^* = \sum_{i=1}^{m} x_i^*, \quad x_0^* \in \partial f(x_0).$$

Доведення. Якщо конуси $K_i \equiv K(x_0,M_i)$ розділяються, то існують такі не всі рівні нулю вектори $x_i^*, i = 1,\ldots,m$, що

$$\sum_{i=1}^{m} x_i^* = 0, \quad x_i^* \in K_i^*,$$

і результат теореми отримуємо, якщо покласти $\lambda = 0$.

Якщо ж конуси $K_i, i = 1,\ldots,m$, не розділяються, то відповідно до теореми 3.1.2 для будь-якого вектора

$$\overline{x}_0 \in \operatorname{ri} K, \quad K = \cap_{i=1}^{m} K_i,$$

існують такі конус Q і функція $\psi(\bar{x})$, що

$$\bar{x}_0 \in \text{ri } Q, \quad \text{Lin } Q = \text{Lin } K, \ Q \subseteq K,$$

$$\psi(\bar{x}) = \bar{x} + r(\bar{x}), \ \ \|\bar{x}\|^{-1} r(\bar{x}) \to 0,$$

якщо $\bar{x} \to 0$, і, крім того, $x_0 + \psi(\bar{x}) \in M$ для $\bar{x} \in Q \cap (\varepsilon B)$, $\varepsilon > 0$. Тому при досить малих $\lambda > 0$ виконується співвідношення

$$x_0 + \psi(\lambda\bar{x}_0) = x_0 + \lambda\bar{x}_0 + r(\lambda\bar{x}_0) \in M,$$

$\lambda^{-1} r(\lambda\bar{x}_0) \to 0$ при $\lambda \downarrow 0$. Виходить, що $\bar{x}_0$– дотичний напрямок.

Таким чином, конус ri K є конусом дотичних напрямків до M в точці x_0. За припущеннями теореми

$$\text{int } (\text{dom } h(x_0,\cdot)) \cap K \neq \emptyset,$$

тобто існує такий вектор $\bar{x}_1 \in \text{int } (\text{dom } h(x_0,\cdot))$, що $\bar{x}_1 \in K$. Оскільки будь-який вектор з опуклої множини може бути наближений векторами з його відносної внутрішності, а $\bar{x}_1$ є внутрішня точка dom $h(x_0,\cdot)$, то знайдеться вектор $\bar{x}_2 \in \text{ri } K$, що належить int dom $h(x_0,\cdot)$. Таким чином, ri K є конус дотичних напрямків до M в точці x_0, для якого виконані припущення теореми 3.3.1. Тому

$$\partial f(x_0) \cap (\text{ri } K)^* \neq \emptyset.$$

Але відповідно до леми 1.3.2 $\overline{K} = \overline{(\text{ri } K)}$ і $(\text{ri } K)^* = K^*$. Отже,

$$\partial f(x_0) \cap K^* \neq \emptyset.$$

Оскільки конуси $K_i, i = 1, \ldots, m$, не розділяються, то у відповідності з теоремою 1.5.3 можемо записати, що

$$K^* = \sum_{i=1}^{m} K_i^*.$$

Це означає, що знайдеться такий вектор $x_0^* \in \partial f(x_0)$ і такі вектори $x_i^* \in K_i^*$, що

$$x_0^* = \sum_{i=1}^{m} x_i^*.$$

$\square$

Наслідок 3.3.2. *Нехай x_0 – точка мінімуму функції $f_0(x)$ при додаткових умовах*

$$f_i(x) = 0, \quad i = 1, \ldots, m,$$

де $f_i(x)$– неперервно диференційовні функції для $i = 0,1,...,m$. Тоді існують такі числа $y^i, i = 0,1,...,m$, що

$$\sum_{i=1}^{m} y^i f_i'(x_0) = 0, \ \ y^0 \geqslant 0,$$

і числа y^i не дорівнюють нулю одночасно.

Доведення. Розглянемо множину

$$M = \{x : f_i(x) = 0, \quad i = 0,1, \ldots, m\}.$$

Можливі два випадки: градієнти $f_i'(x_0), i = 1, \ldots, m$, лінійно залежні і градієнти $f_i'(x_0), i = 1, \ldots, m$, лінійно незалежні. У першому випадку покладемо $y^0 = 0$, а $y^i, i = 1, \ldots, m$, візьмемо з лінійної комбінації

$$\sum_{i=1}^{m} y^i f_i'(x_0) = 0.$$

У другому випадкі, згідно прикладу 3.1.1, маємо

$$K(x_0, M) = \{x : \langle x, f_i'(x_0)\rangle = 0, \ \ i = 1,...,m\},$$

$$K^*(x_0, M) = \left\{x^* : x^* = \sum_{i=1}^{m} \lambda f_i'(x_0), \ \ \lambda_i \in R^1, \ \ i = 1,...,m\right\}.$$

Функція $f_0(x)$ диференційовна. Тому існують $\partial f_0(x_0) = \{f_0'(x_0)\}$ і застосування теореми 3.3.1 дає рівність

$$f_0'(x_0) = \sum_{i=1}^{m} \lambda_i f_i'(x_0),$$

звідки отримаємо необхідний результат, якщо покладемо

$$y^0 = 1, \ y^i = -\lambda_i, \quad i = 1, \ldots, m.$$

□

Наслідок 3.3.3. *Нехай x_0 – точка мінімуму функції $f_0(x)$ при обмеженнях*

$$f_i(x) \leqslant 0, \quad i \in I^-; \quad f_i(x) = 0, \quad i \in I,$$

де функції $f_i(x), i \in \{0\} \cup I^- \cup I$, неперервно диференційовні. Тоді існують такі числа $y^i, i \in \{0\} \cup I^- \cup I$, що

$$\sum_{i \in \{0\} \cup I^- \cup I} y^i f_i'(x_0) = 0,$$

причому, не всі числа y^i дорівнюють нулю, і

$$y^i \geqslant 0, \, i \in \{0\} \cup I^-, \ \ y^i f_i(x_0) = 0, \ \ i \in I^-.$$

Доведення. Розглянемо множини

$$M_i = \begin{cases} \{x : f_i(x) \leqslant 0\}, & i \in I^-, \quad f_i(x_0) = 0, \\ X, & i \in I^-, \quad f_i(x_0) < 0, \\ \{x : f_i(x) = 0\}, & i \in I. \end{cases}$$

У відповідності з прикладами 3.1.3, 3.1.4 конуси

$$K_i = \begin{cases} \{\bar{x} : \langle \bar{x}, f_i'(x)\rangle < 0\}, & i \in I^-, \quad f_i(x_0) = 0, \\ X, & i \in I^-, \quad f_i(x_0) < 0, \\ \{\bar{x} : \langle \bar{x}, f_i'(x)\rangle = 0\}, & i \in I. \end{cases}$$

є локальними шатрами до множин M_i у точці x_0, якщо $f_i'(x_0) \neq 0$ для $i \in I^-$, $f_i(x_0) = 0$, чи $i \in I$. Скориставшись тим, що

$$K_i^* = \begin{cases} \{-\lambda_i f_i'(x_0) : \lambda_i \geqslant 0\}, & i \in I^-, \quad f_i(x_0) = 0, \\ \{0\}, & i \in I^-, \quad f_i(x_0) < 0, \\ \left\{-\lambda_i f_i'(x_0) : \lambda_i \in \mathbb{R}^1\right\}, & i \in I. \end{cases}$$

результат теореми 3.3.2 можна переписати у вигляді

$$\lambda x_0^* = -\sum_{i \in \{0\} \cup I^- \cup I} \lambda_i f_i'(x_0),$$

$$\lambda_i \geqslant 0, \; i \in I^-, \; \lambda_i f_i(x_0) = 0, \; i \in I^-.$$

Поклавши $y^0 = \lambda$, $y^i = \lambda_i$, одержуємо необхідний результат.

Якщо $f_{i_0}'(x_0) = 0$ для деякого $i_0 \in I^-$, $f_i(x_0) = 0$ чи $i \in I$, то досить покласти $y^{i_0} = 1$ і $y^i = 0$ для всіх інших i. □

Наслідок 3.3.4. *Нехай x_0 – точка мінімуму функції $f_0(x)$ при обмеженнях*

$$f_i(x) \leqslant 0, \quad i \in I^-,$$

$$f_i(x) \leqslant 0, \quad i \in I^-, \quad x \in M,$$

де функції $f_i(x)$ неперервно диференційовні, а множина M опукла. Тоді існують такі не всі рівні нулю числа $y^i, i \in \{0\} \cup I^- \cup I$, що

$$\sum_{i \in \{0\} \cup I^- \cup I} y^i f_i'(x_0) \in (\mathrm{cone}\,(M - x_0))^*,$$

$$y^i \geqslant 0, \quad i \in \{0\} \cup I^-, \quad y^i f_i(x_0) = 0, \quad i \in I^-.$$

Доведення. Доведення отримаємо з теореми 3.3.2 аналогічно попередньому наслідку, якщо крім множин M_i розглянути і множину M, для якої, відповідно до прикладу 3.1.5, $\mathrm{cone}\,(M - x_0)$ є локальне шатро. Дійсно, у цьому випадку знайдуться такі числа $\lambda_i \geqslant 0$, $i \in I^-$, що

$$\lambda f_0'(x_0) = -\sum_{i \in \{0\} \cup I^- \cup I} \lambda_i f_i'(x_0) + x^*, \qquad x^* \in (\mathrm{cone}\,(M - x_0))^*,$$

і $\lambda_i = 0$, якщо $f_i(x_0) < 0$. З цієї формули і випливає потрібний результат, якщо покласти $y^0 = \lambda, y^i = \lambda_i$. □

3.3.2. Обмеження, що задаються рівностями і нерівностями

Розглянемо тепер задачу мінімізації функції $f_0(x)$ при обмеженнях

$$f_i(x) \leqslant 0, \;\; i \in I^-; \;\; f_i(x) = 0, \;\; i \in I; \;\; x \in M. \tag{3.3.1}$$

Сформулюємо припущення, при яких буде розв'язуватися поставлена задача. Нехай x_0 – точка мінімуму.

П р и п у щ е н н я 1. Функції $f_i(x), i \in \{0\} \cup I^-$, допускають у точці x_0 верхню опуклу апроксимацію $h_i(x_0,g)$.

П р и п у щ е н н я 2. Функції $f_i(x), i \in I$, неперервно диференційовні в околі точки x_0, тобто мають неперервні градієнти $f_i'(x)$.

П р и п у щ е н н я 3. У точці x_0 існує опуклий конус $K(x_0,M)$ дотичних напрямків до множини M, яка має наступну властивість. Нехай $L = \mathrm{Lin}(K(x_0,M))$ – найменший лінійний підпростір, що містить $K(x_0,M)$, а $\bar{x}_0 \in K(x_0,M), \bar{x}_0 \neq 0$, і $e_j \in L$, $j = 1,...,m$,– фіксовані вектори. Позначимо через L_0 лінійний підпростір, що містить $\bar{x}_0$ і $e_j, j = 1,...,m$, тобто множина всіх векторів $\bar{x}$ вигляду

$$\bar{x} = \gamma\bar{x}_0 + \sum_{j=1}^{m} \delta_j e_j,$$

де γ і δ_j – дійсні числа. Тоді існують такі $\varepsilon_1 > 0$ і $\varepsilon_2 > 0$, що:

а) множина

$$\Omega = \left\{ \bar{x} = \bar{x}_0 + \sum_{j=1}^{m} \delta_j e_j : |\delta_j| \leqslant \varepsilon_1,\ j = 1,...,m \right\}$$

міститься в $K(x_0,M)$;

б) існує функція $\psi(\bar{x})$, яка визначена для всіх досить малих $\bar{x} \in L_0$, неперервно диференційовна в області визначення і

$$\psi(\bar{x}) = \bar{x} + r(\bar{x}),\quad x_0 + \psi(\bar{x}) \in M$$

для $\bar{x} \in Q$ і $\|\bar{x}\| < \varepsilon_2$, де $r(\bar{x})$ така, що $\|\bar{x}\|^{-1} r(\bar{x}) \to 0$ при $\bar{x} \to 0,\ \bar{x} \in L_0$, а множина Q визначається співвідношенням

$$\Omega = \mathrm{cone}\,\Omega = \left\{ \bar{x} : \bar{x} = \gamma\left(\bar{x}_0 + \sum_{j=1}^{m} \delta_j e_j \right),\quad \gamma \geqslant 0,\quad |\delta_j| \leqslant \varepsilon_1 \right\}.$$

З цих трьох припущень найбільш важко перевіряється припущення 3. Однак воно необхідне для цілого ряду задач.

Відзначимо, що в тому випадку, коли $K(x_0,M)$ є шатро множини M, яке лежить у скінченновимірному просторі $\mathbb{R}^n$, припущення 3 виконується. Це випливає з означення шатра, якщо замість $K(x_0,M)$ взяти $\mathrm{ri}\,K_M(x_0)$.

П р и п у щ е н н я 4.

$$\left(\cap_{i\in\{0\}\cup I^-} \mathrm{dom}\ h_i(x_0,\cdot)\right) \cap K(x_0,M) \neq \emptyset.$$

Лема 3.3.1. *Нехай $l_i(x)$, $i = 1,...,m$,– лінійні функції, визначені на підпросторі $L \subseteq X$. Тоді або існують такі не всі рівні нулю числа α_i, що*

$$\sum_{i=1}^{m} \alpha_i l_i(x) = 0,\ \ x \in L, \tag{3.3.2}$$

або існують такі вектори $e_j \in L,\ j = 1,...,m,$ *що*

$$l_i(e_j) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \qquad (3.3.3)$$

Доведення. Розглянемо вектор-функцію $l(x) \in \mathbb{R}^m$ з компонентами $l_i(x)$. Нехай $A = l(L)$– образ підпростору L при відображенні l.

Якщо $A = \mathbb{R}^m$, то будь-який вектор $y \in \mathbb{R}^m$ можна представити у вигляді

$$y = l(x),\ x \in L.$$

Зокрема, якщо $q_j \in \mathbb{R}^m$– одиничні орти $\mathbb{R}^m$, тобто для компонент $q_j^i, i = 1,...,m$, виконується співвідношення $q_j^i = \delta_{ij}$, то існують такі вектори $e_j \in L$, що

$$q_j = l(e_j),$$

або, у покомпонентному вигляді, $l_i(e_j) = \delta_{ij}$, так що співвідношення (3.3.3) виконуються. При цьому співвідношення (3.3.2) можливо лише при нульових α_i, оскільки

$$\sum_{i=1}^{m} \alpha_i l_i(e_j) = \alpha_i = 0,\ j = 1,...,m.$$

Якщо ж A не збігається з усім $\mathbb{R}^m$, то A є деякий власний підпростір простору $\mathbb{R}^m$. Звідси випливає, що A лежить у деякій гіперплощині

$$\langle y,y^* \rangle = 0,\ y \in A,\ y^* \neq 0.$$

Враховуючи, що $y = l(x),\ x \in L$, одержуємо

$$\sum_{i=1}^{m} y^{i*} l_i(x) = 0,\ x \in L.$$

Покладаючи $\alpha_i = y^{i*}$, приходимо до співвідношення (3.3.2). □

Лема 3.3.2. *Нехай виконані припущення 2 і 3. Тоді або*

$$\sum_{i \in I} \alpha_i \langle \bar{x}, f_i'(x_0) \rangle = 0,\ \sum_{i \in I} |\alpha_i| = 1, \qquad (3.3.4)$$

для всіх $\bar{x} \in L, L = Lin\, K(x_0,M)$, *або для всякого* $\bar{x}_0 \in K(x_0,M), \bar{x}_0 \neq 0$, *що задовольняє умовам*

$$\langle \bar{x}_0, f_i'(x_0) \rangle = 0,\ \ i = 1,...,m, \qquad (3.3.5)$$

існує функція $r_0(\gamma) \in X$, *визначена для всіх досить малих* $\gamma \geqslant 0$ *і така, що*

$$x_0 + \gamma \bar{x}_0 + r_0(\gamma) \in M,$$

$$f_i(x_0 + \gamma \bar{x}_0 + r_0(\gamma)) = 0,\ \ i = 1,...,m,$$

і $\gamma^{-1} r_0(\gamma) \to 0$ *при* $\gamma \downarrow 0$.

Доведення. Відповідно до попередньої леми або виконано співвідношення (3.3.4), або знайдуться такі вектори $e_j \in L$, $j \in I$, що
$$\langle e_j, f_i'(x_0)\rangle = \delta_{ij}, \quad i,j \in I. \tag{3.3.6}$$
Зрозуміло, що необхідно розглянути лише другу можливість. Візьмемо функцію $\psi(\bar{x})$, визначену на підпросторі L_0 (див. припущення 3), і складемо систему рівнянь
$$g_i(\gamma,\delta) \equiv f_i\left(x_0 + \psi\left(\gamma\bar{x}_0 + \sum_{j\in I}\delta_j e_j\right)\right) = 0, \quad i \in I. \tag{3.3.7}$$
Тут $\delta-$ вектор з компонентами δ_j. Згідно з припущеннями 2 і 3 функції $g_i(\gamma,\delta)$ неперервно диференційовні при всіх досить малих γ і δ. Обчислимо перші похідні функцій g_i при $\gamma = 0$, $\delta = 0$. Для цього відмітимо, що в силу властивостой функції $\psi(\bar{x})$ справедлива формула
$$\psi_\gamma'(0) \equiv \frac{\partial}{\partial\gamma}\psi\left(\gamma\bar{x}_0 + \sum_{j\in I}\delta_j e_j\right)\Bigg|_{\gamma=0,\ \delta=0} =$$
$$= \lim_{\gamma\to 0}\frac{\psi(\gamma\bar{x}_0) - \psi(0)}{\gamma} = \lim_{\gamma\to 0}\left(\bar{x}_0 + \frac{r(\gamma\bar{x}_0)}{\gamma}\right) = \bar{x}_0. \tag{3.3.8}$$
Аналогічно отримуємо
$$\psi_{\delta_j}'(0) = \frac{\partial}{\partial\delta_j}\psi\left(\gamma\bar{x}_0 + \sum_{j\in I}\delta_j e_j\right)\Bigg|_{\gamma=0,\ \delta=0} = e_j \tag{3.3.9}$$
Тому за правилом диференціювання складної функції маємо
$$\frac{\partial}{\partial\gamma}g_i(\gamma,\delta)\big|_{\gamma=0,\ \delta=0} = \left\langle\psi_{\delta_j}'(0), f_i'(x_0)\right\rangle = \langle\bar{x}_0, f_i'(x_0)\rangle = 0,\ i \in I, \tag{3.3.10}$$
в силу умов (3.3.5). Аналогічно одержуємо
$$\frac{\partial}{\partial\delta_j}g_i(\gamma,\delta)\big|_{\gamma=0,\ \delta=0} = \left\langle\psi_{\delta_j}'(0), f_i'(x_0)\right\rangle = \langle e_j, f_i'(x_0)\rangle = \delta_{ij}. \tag{3.3.11}$$
Якщо скористатися тепер зауваженням до теореми 3.1.1, то можна зробити висновок, що для досить малих γ визначена неперервна функція $\delta(\gamma)$, причому така, що виконується співвідношення
$$\lim_{\gamma\to 0}\frac{\delta(\gamma)}{\gamma} = 0$$
і задовольняються рівняння (3.3.7). В силу припущення 3 маємо
$$\psi\left(\gamma\bar{x}_0 + \sum_{i\in I}\delta_j(\gamma)e_j\right) = \gamma\bar{x}_0 + \sum_{j\in I}\delta_j(\gamma)e_j +$$

$$+r\left(\gamma\bar{x}_0+\sum_{i\in I}\delta_j\left(\gamma\right)e_j\right)=\gamma\bar{x}_0+r_0\left(\gamma\right),$$

$$r_0\left(\gamma\right)=\sum_{j\in I}\delta_j\left(\gamma\right)e_j+r\left(\gamma\bar{x}_0+\sum_{i\in I}\delta_j\left(\gamma\right)e_j\right).$$

Із властивостей функцій $\delta\left(\gamma\right)$ і $r\left(\bar{x}\right)$ (вони прямують до нуля швидше, ніж γ і $\bar{x}$) випливає, що

$$\frac{r_0\left(\gamma\right)}{\gamma}\to 0, \tag{3.3.12}$$

і при досить малих $\gamma\geqslant 0$

$$\gamma\bar{x}_0+\sum_{j\in I}\delta_j\left(\gamma\right)e_j=\gamma\left(x_0+\sum_{j\in I}\frac{\delta_j}{\gamma}e_j\right)\in\gamma\Omega\subseteq Q, \tag{3.3.13}$$

оскільки

$$\left|\frac{\delta_j\left(\gamma\right)}{\gamma}\right|\leqslant\varepsilon_1,\ \ j\in I.$$

Згідно з припущенням 3 з включення (3.3.13) випливає, що

$$x_0+\psi\left(\gamma\bar{x}_0+\sum_{i\in I}\delta_j\left(\gamma\right)e_j\right)\in M,$$

при малому γ, тобто

$$x_0+\gamma\bar{x}_0+r_0\left(\gamma\right)\in M. \tag{3.3.14}$$

Враховуючи, що система (3.3.7) може бути переписана у вигляді

$$f_i\left(x_0+\gamma\bar{x}_0+r_0\left(\gamma\right)\right)=0,\ \ i\in I, \tag{3.3.15}$$

зі співвідношень (3.3.14) і (3.3.12) одержуємо усі твердження леми. □

Теорема 3.3.3. *Нехай точка x_0 є точкою мінімуму функції $f_0\left(x\right)$ при обмеженнях (3.3.1). Нехай виконані припущення 1–4. Тоді існують такі числа y^i, $i\in\{0\}\cup I^-\cup I$, що*

$$\sum_{i\in I^-}y^ih_i\left(x_0,g\right)+\sum_{i\in I}y^i\left\langle g,f_i'\left(x_0\right)\right\rangle\geqslant 0 \tag{3.3.16}$$

для всіх

$$\bar{x}\in K\left(x_0,M\right)\cap\left(\bigcap_{i\in\{0\}\cup I^-}dom\,h\left(x_0,\cdot\right)\right).$$

При цьому $y^i\geqslant 0$ для $i\in\{0\}\cup I^-$ та $y^i=0$ для $i\in I$, де

$$I_0^-=\left\{i\in I^-:f_i\left(x_0\right)<0\right\}.$$

Доведення. Якщо співвідношення (3.3.4) виконується, то достатньо для $i \in I$ покласти $y^i = \alpha_i$, взявши α_i із співвідношення (3.3.4), а всі інші y^i вибрати рівними нулю.

Припустимо тепер, що співвідношення (3.3.4) не виконується. Введемо наступні позначення:

$$J^- = \{0\} \cup \left(I^- \backslash I_0^-\right), \; J = J^- \cup I,$$

$$K = K\left(x_0, M\right) \cap \left(\bigcap_{i \in \{0\} \cup I^-} dom\, h\left(x_0, \cdot\right)\right).$$

Нехай b – число індексів у J. Визначимо множину P у такий спосіб: $\bar{y} \in P, \; P \subseteq \mathbb{R}^b$, тоді і тільки тоді, коли існує такий вектор $\bar{x} \in K$, що

$$\bar{y}^i > h_i\left(x_0, \bar{x}\right), \;\; i \in J^-, \;\; \bar{y}^i = \left\langle \bar{x}, f_i'\left(x_0\right)\right\rangle, \; i \in I.$$

З опуклості функцій h_i легко випливає, що P– опукла множина. В силу припущення 4 вона не порожня.

Розглянемо множину

$$N = \left\{\bar{y} \in \mathbb{R}^b : \;\; \bar{y}^i \leqslant 0, \;\; i \in J^-, \;\; \bar{y}^i = 0, \;\; i \in I\right\}$$

і покажемо, що множини P і N не перетинаються. Припустимо протилежне. Тоді знайдуться вектор $\bar{y}_0 \in P \cap N$ і вектор $\bar{x}_0 \in K$, такі, що

$$\begin{array}{lcl} h_i\left(x_0, \bar{x}_0\right) < y & \leqslant & 0, i \in J^- \\ \left\langle \bar{x}_0, f_i'\left(x_0\right)\right\rangle & = & 0, i \in I. \end{array} \qquad (3.3.17)$$

Оскільки $\bar{x}_0 \in K$, а $K \subseteq K_M\left(x_0\right)$, то $\bar{x}_0 \in K_M\left(x_0\right)$. З першого співвідношення (3.3.17) випливає, що $\bar{x}_0 \neq 0$, оскільки $h_i\left(0, x_0\right) = 0$.

З другого співвідношення (3.3.17) і леми 3.3.2 випливає існування такої функції $r_0\left(\gamma\right)$, що $\gamma^{-1} r_0\left(\gamma\right) \to 0$ при $\gamma \downarrow 0$, для якої виконані співвідношення (3.3.14), (3.3.15). В силу припущення 1 для $i \in J^-$ одержуємо, що

$$\limsup_{\gamma \downarrow 0} \frac{f_i\left(x_0 + \gamma \bar{x}_0 + r_0\left(\gamma\right)\right) - f_i\left(x_0\right)}{\gamma} \leqslant h_i\left(x_0, \bar{x}_0\right) < 0,$$

тобто для всіх досить малих $\gamma > 0$ виконується

$$f_i\left(x_0 + \gamma \bar{x}_0 + r_0\left(\gamma\right)\right) \leqslant f_i\left(x_0\right) + \frac{1}{2}\gamma h_i\left(x_0, \bar{x}_0\right), \; i \in J^-.$$

Звідси, так як $h_i\left(x_0, \bar{x}_0\right) < 0$ для $i \in J^-$, при досить малих $\gamma > 0$ одержуємо наступні співвідношення:

$$f_0\left(x_0 + \gamma \bar{x}_0 + r_0\left(\gamma\right)\right) < f_0\left(x_0\right), \qquad (3.3.18)$$

$$f_i\left(x_0 + \gamma \bar{x}_0 + r_0\left(\gamma\right)\right) < 0, \; i \in \left(I^- \backslash I_0^-\right). \qquad (3.3.19)$$

Величини $h_i\left(x_0, \bar{x}_0\right), \, i \in I_0^-$, скінченні, оскільки $\bar{x}_0 \in K$, $K \subseteq \mathrm{dom}\; h_i\left(x_0, \cdot\right)$. Тому

$$\limsup_{\gamma \downarrow 0} \frac{f_i\left(x_0 + \gamma \bar{x}_0 + r_0\left(\gamma\right)\right) - f_i\left(x_0\right)}{\gamma} \leqslant h_i\left(x_0, \bar{x}_0\right).$$

Звідки випливає, що нерівність

$$f_i\left(x_0+\gamma\bar{x}_0+r_0\left(\gamma\right)\right)\leqslant f_i\left(x_0\right)+\gamma\left(h_i\left(x_0,\bar{x}_0\right)+\varepsilon\right)<0,\ \ i\in I_0, \qquad (3.3.20)$$

виконується для будь-якого $\varepsilon>0$ і при досить малих $\gamma>0$ в силу того, що $f_i\left(x_0\right)<0$ для $i\in I_0^-$.

Зіставляючи співвідношення (3.3.14), (3.3.15), (3.3.18)–(3.3.20), бачимо, що знайдуться точки $x=x_0+\gamma\bar{x}_0+r_0\left(\gamma\right)$, відмінні від x_0, в яких значення $f_0\left(x\right)$ менше, а всі обмеження (3.3.1) виконуються. Це суперечить тому, що x_0 – точка мінімуму функції $f_0\left(x\right)$. Таким чином, введені вище множини P і N не перетинаються. На підставі теореми про розділення існує такий вектор $y\in\mathbb{R}^b$, що

$$\langle y_2,y\rangle\geqslant\langle y_1,y\rangle \qquad (3.3.21)$$

для всіх $y_2\in P$, $y_1\in N$, або, у покомпонентному записі,

$$\sum_{i\in J^-}y_2^iy^i+\sum_{i\in I}y_2^iy^i\geqslant\sum_{i\in J^-}y_1^iy^i+\sum_{i\in J}y_1^iy^i,\quad y_2\in P,\quad \bar{y}_1\in N. \qquad (3.3.22)$$

З визначення множини P випливає, що величини y_2^i, $i\in J^-$, можуть необмежено зростати при фіксованому $\bar{x}$. Тому $y^i\geqslant 0$, $i\in J^-$, оскільки протилежна нерівність приводила б до суперечності з (3.3.22) при $y_2^i\to+\infty$. Отже,

$$y^i\geqslant 0,\ \ i\in\{0\}\cup\left(I^-\backslash I_0^-\right). \qquad (3.3.23)$$

Нехай $\bar{x}\in K$. Покладемо в нерівності (3.3.22) $y_1=0$. Тоді спрямувавши y_2^i до $h_i\left(x_0,\bar{x}\right)$ для $i\in J^-$ і поклавши для $i\in I$ $y_2^i=\langle\bar{x},\,f_i'\left(x_0\right)\rangle$, отримаємо

$$\sum_{i\in J^-}y^ih_i\left(x_0,\bar{x}\right)+\sum_{i\in I}y^i\left\langle\bar{x},f_i'\left(x_0\right)\right\rangle\geqslant 0$$

для $\bar{x}\in K$. Вибравши $y^i=0$ для $i\in I_0^-$, отримаємо

$$\sum_{i\in\{0\}\cup I^-}y^ih_i\left(x_0,\bar{x}\right)+\sum_{i\in I}y^i\left\langle\bar{x},f_i'\left(x_0\right)\right\rangle\geqslant 0,\ \ \bar{x}\in K.$$

$$y^i\geqslant 0,\ i\in\{0\}\cup I^-;$$

$$y^i=0,\ \ i\in I_0^-,$$

що і було потрібно довести. □

Теорема 3.3.4. *Нехай виконані припущення 1-3 і нехай існує точка $\bar{x}_1\in K\left(x_0,M\right)$, у якій функції $h_i\left(x_0,\bar{x}\right),i\in\{0\}\cup I^-$, неперервні. Тоді існують не всі рівні нулю числа $y^i,i\in\{0\}\cup I^-\cup I$, і вектори*

$$x_i^*\in\partial f_i\left(x_0\right),\ \ i\in\{0\}\cup I^{-1},\quad x_{i0}^*\in\left(dom\,h_i\left(x_0,\cdot\right)\right)^*\ \ x^*\in K^*\left(x_0,M\right),$$

такі, що

$$\sum_{i\in\{0\}\cup I^-}y^ix_i^*+\sum_{i\in I}y^if_i'\left(x_0\right)=x^*+\sum_{i\in\{0\}\cup I^-}x_{i0}^*, \qquad (3.3.24)$$

$$y^i \geqslant 0,\ \ i \in \{0\} \cup I^-; \quad y^i = 0,\ \ i \in I_0^-.$$

Доведення. В силу припущень виконані всі умови теореми 3.3.3. Тому знайдуться такі числа y^i, що $y^i \geqslant 0,\ i \in \{0\} \cup I^-; \quad y^i = 0,\ i \in I_0^-$ і буде виконана нерівність (3.3.16). Позначимо функцію в лівій частині нерівності (3.3.16) через $h(\bar{x})$. Зрозуміло, що

$$\mathrm{dom}\ h = \bigcap_{i \in \{0\} \cup I^-} \mathrm{int\ dom}\ h_i(x_0, \cdot),$$

і в силу умов теореми виконане включення

$$\bar{x}_1 \in \mathrm{dom}\ h, \tag{3.3.25}$$

функція $h(\bar{x})$ неперервна в точці $\bar{x}_1$.

Нерівність (3.3.16) показує, що $\bar{x} = 0$ є точка мінімуму функції $h(\bar{x})$ на конусі

$$K = K(x_0, M) \cap \left(\bigcap_{i \in \{0\} \cup I^-} \mathrm{int\ dom}\ h_i(x_0, \cdot) \right).$$

Відповідно до теореми 2.9.5

$$\partial h(0) \cap K^* \neq \emptyset, \tag{3.3.26}$$

а в силу теореми 2.7.8

$$\partial h(0) = \sum_{i \in \{0\} \cup I^-} y^i \partial f_i(x) + \sum_{i \in I} y^i f_i'(x) \tag{3.3.27}$$

тому що відповідно до зроблених припущень функції h_i неперервні в точці $\bar{x}_1$ і

$$\bar{x}_1 \in K(x_0, M) \cap \left(\bigcap_{i \in \{0\} \cup I^-} \mathrm{int\ dom}\ h_i(x_0, \cdot) \right),$$

то за теоремою 1.5.3 одержуємо, що

$$K^* = K^*(x_0, M) + \sum_{i \in \{0\} \cup I^-} (\mathrm{dom}\ h_i(x_0, \cdot))^*. \tag{3.3.28}$$

Зі співвідношень (3.3.26)–(3.3.28) випливає рівність (3.3.24). □

Наслідок 3.3.5. *Якщо функція $h_i(x_0, \bar{x})$ неперервна за $\bar{x}$ на всьому просторі X, то за припущень 1–3 вірна теорема 3.3.4, а рівність (3.3.24) може бути записана у вигляді*

$$\sum_{i \in \{0\} \cup I^-} y^i x_i^* + \sum_{i \in I} y^i f_i'(x_0) = x^*.$$

Доведення. Справді, якщо $h_i(x_0, \bar{x})$ неперервна на всьому просторі, то

$$\mathrm{dom}\ h_i(x_0, \cdot) = X, \quad (dom\, h_i(x_0, \cdot))^* = \{0\},$$

і з включення $x_{i0}^* \in (\mathrm{dom}\ h_i(x_0, \cdot))^*$ випливає, що $x_{i0}^* = 0$. □

Розділ IV
Похідні за напрямком

4.1. ПОХІДНІ ДІНІ ТА АДАМАРА

Нехай функція $f\colon \mathbb{R}^n \to \mathbb{R}^1$ визначена на деякій відкритій множині $S \subset \mathbb{R}^n$ і приймає лише скінченні значення.

Означення 4.1.1. Величина
$$f_D^{\uparrow}(x,g) = \overline{\lim_{\alpha\downarrow 0}} \frac{1}{\alpha}\left(f(x+\alpha g) - f(x)\right)$$
називається *верхньою похідною Діні* функції f в точці $x \in S$ за напрямком $g \in \mathbb{R}^n$.

Означення 4.1.2. Величина
$$f_D^{\downarrow}(x,g) = \underline{\lim_{\alpha\to+0}} \frac{1}{\alpha}(f(x+\alpha g) - f(x))$$
називається *нижньою похідною Діні* функції f в точці $x \in S$ за напрямком $g \in \mathbb{R}^n$.

Означення 4.1.3. Границя
$$f'(x,g) = \lim_{\alpha\to+0} \frac{1}{\alpha}(f(x+\alpha g) - f(x)), \tag{4.1.1}$$
якщо вона існує, називається *похідною функції f в точці $x \in S$ за напрямком $g \in \mathbb{R}^n$*. Цю границю також називають *похідною Діні* і позначають через $f_D'(x,g)$.

У тому випадку, коли ця похідна існує та скінченна, будемо казати, що функція f *диференційовна* (*за Діні*) *в точці $x \in S$ за напрямком $g \in \mathbb{R}^n$*. Функція f *диференційовна за напрямками* (або *диференційовна за Діні*), якщо границя (4.1.1) існує і скінченна для всіх $g \in \mathbb{R}^n$. Разом із символом $f'(x,g)$ будемо використовувати для похідної за напрямком позначення $f_x'(g)$.

Безпосередньо з означень випливає, що похідна Діні є додатньо однорідною першого порядку функцією напрямку:
$$f'(x,\lambda g) = \lambda f'(x,g) \quad \forall\lambda > 0.$$
Це стосується і верхньої і нижньої похідної.

Зафіксуємо точку $x \in S$, напрямок $g \in \mathbb{R}^n$, і розглянемо функцію однієї дійсної змінної
$$\varphi(\alpha) = f(x+\alpha g),$$

що визначена для $\alpha \geq 0$. Верхня похідна Діні $f_D^{\uparrow}(x,g)$ функції $f(x)$ співпадає з верхньою правою похідною $\bar{\varphi}'_+(0)$ функції φ в точці $\alpha = 0$. Нижня похідна Діні $f_D^{\downarrow}(x,g)$ функції $f(x)$ співпадає з нижньою правою похідною $\underline{\varphi}'_+(0)$ функції φ в точці $\alpha = 0$. Похідна Діні $f'(x,g)$ функції $f(x)$ співпадає з правою похідною $\varphi'_+(0)$ функції φ.

Нагадаємо, що за означенням

$$\bar{\varphi}'_+(\beta) = \overline{\lim_{\alpha\downarrow\beta}} \frac{1}{\alpha - \beta}(\varphi(\alpha) - \varphi(\beta)),$$

$$\underline{\varphi}'_+(\beta) = \underline{\lim}_{\alpha\downarrow\beta} \frac{1}{\alpha - \beta}(\varphi(\alpha) - \varphi(\beta)),$$

$$\varphi'_+(\beta) = \lim_{\alpha\downarrow\beta} \frac{1}{\alpha - \beta}(\varphi(\alpha) - \varphi(\beta)).$$

Отже, похідна Діні (відповідно, верхня та нижня похідні Діні) є односторонньою похідною (відповідно, верхньою та нижньою односторонньою похідною) звичайної числової функції. Тому можна використовувати методи аналізу функцій однієї змінної при дослідженні похідних Діні. Це дозволяє говорити про диференційне числення для похідних за напрямками. Точніше кажучи, справедливе наступне твердження.

Теорема 4.1.1. *Нехай функції $f_1(x)$ і $f_2(x)$ диференційовні за напрямками в точці x. Тоді їх сума і добуток, а також частка (якщо $f_2(x) \neq 0$) також диференційовні за напрямками в цій точці. При цьому*

$$(\lambda f)'(x,g) = \lambda f'(x,g), \lambda \in \mathbb{R}^1;$$

$$(f_1 + f_2)'(x,g) = f_1'(x,g) + f_2'(x,g);$$

$$(f_1 - f_2)'(x,g) = f_1'(x,g) - f_2'(x,g);$$

$$(f_1 f_2)'(x,g) = f_1(x) f_2'(x,g) + f_2(x) f_1'(x,g);$$

$$\left(\frac{f_1}{f_2}\right)'(x,g) = \frac{-1}{[f_2(x)]^2}\left(f_1(x) f_2'(x,g) - f_2(x) f_1'(x,g)\right).$$

Доведення. Для доведення слід розглянути функції $\varphi_1(\alpha) = f_1(x + \alpha g)$ і $\varphi_2(\alpha) = f_2(x + \alpha g)$ та застосувати відповідні теореми диференційного числення функцій однієї змінної. □

Клас диференційовних за напрямками функцій досить широкий. Він містить всі гладкі (диференційовні за Гато) функції. Теорема 4.1.1 показує, що сукупність диференційовних за напрямками функцій є лінійним простором, який разом із двома своїми елементами містить їх добуток і частку (якщо дільник відмінний від нуля). Більш того, з існування похідної Діні функцій $f_1(x)$ і $f_2(x)$ випливає існування похідної Діні у функцій $\underline{f}(x)$ і $\bar{f}(x)$, де

$$\underline{f}(x) = \min\{f_1(x), f_2(x)\}, \quad \bar{f}(x) = \max\{f_1(x), f_2(x)\}.$$

Сформулюємо аналог теореми про середнє для диференційовних за напрямками функцій.

Теорема 4.1.2. *Нехай функція $f(x)$ визначена і неперервна на деякій відкритій множині $S \subset \mathbb{R}^n$ і відрізок $\{y\colon\ y = x + \alpha g, \alpha \in [0,\alpha_0]\}$ повністю міститься в S. Покладемо*

$$m = \inf_{\alpha \in [0,\alpha_0]} f_D^{\downarrow}(x + \alpha g, g), \quad M = \sup_{\alpha \in [0,\alpha_0]} f_D^{\uparrow}(x + \alpha g, g).$$

Тоді

$$m\alpha_0 \le f(x + \alpha_0 g) - f(x) \le M\alpha_0.$$

Доведення теореми спирається на наступну лему.

Лема 4.1.1. *Нехай функція $h(\alpha)$ визначена і неперервна на відрізку $[a,b]$. Тоді: 1) якщо $\underline{h}'_+(\alpha) \ge 0$ для всіх $\alpha \in (a,b)$, то $h(b) \ge h(a)$; 2) якщо $\bar{h}'_+(\alpha) \le 0$ для всіх $\alpha \in (a,b)$, то $h(b) \le h(a)$.*

Доведення. Обмежимося доведенням першої частина леми. Зафіксуємо $\varepsilon > 0$ і розглянемо підмножину A відрізка $[a,b]$, що складається з чисел α, які мають таку властивість: якщо $a \le \beta \le \alpha$, то

$$h(\beta) - h(\alpha) \ge -\varepsilon(\beta - a). \tag{4.1.2}$$

З означення випливає, що множина A з кожною своєю точкою α містить відрізок $[a,\alpha]$. Крім того, $a \in A$. Отже, A є непорожнім проміжком. Нехай γ — правий кінець цього проміжку. Оскільки функція h неперервна, то, переходячи в нерівності (4.1.2) до границі при $\beta \uparrow \gamma$, одержимо

$$h(\gamma) - h(a) \ge -\varepsilon(\gamma - a), \tag{4.1.3}$$

звідки випливає, що $\gamma \in A$. Таким чином $A = [a,\gamma]$. Покажемо, що $\gamma = b$. Припустимо протилежне: для кожного $\delta > 0$ знайдеться таке число $\alpha_\delta \in (0,\delta)$, що

$$h(\gamma + \alpha_\delta) - h(a) < -\varepsilon(\gamma + \alpha_\delta - a). \tag{4.1.4}$$

Із (4.1.3) і (4.1.4) випливає нерівність $\frac{1}{\alpha_\delta}[h(\gamma + \alpha_\delta) - h(\gamma)] < -\varepsilon$, яка показує, що

$$\underline{h}'_+(\gamma) = \varliminf_{al \downarrow 0} \frac{1}{\alpha}[h(\gamma + \alpha) - h(\gamma)] \le -\varepsilon < 0.$$

Це суперечить умові леми. Отже $\gamma = b$. Тому $h(b) - h(a) \ge -\varepsilon(b - a)$. Спрямовуючи ε до нуля, переконуємося в справедливості леми. $\square$

Доведення. Доведення теореми 4.1.2. Покладемо $h(\alpha) = f(x+\alpha g)$, $\alpha \in [0,\alpha_0]$. Тоді $\bar{h}'_+(\alpha) = f_D^{\uparrow}(x+\alpha g,g)$. Тому $\bar{h}'_+(\alpha) \le M$ при всіх $\alpha \in (0,\alpha_0)$. Покладемо $h_1(\alpha) = M\alpha - h(\alpha)$. Оскільки $(\underline{h}_1)'_+(\alpha) = M - \bar{h}'_+(\alpha) \ge 0$, то в силу леми 4.1.1 одержимо $h_1(\alpha_0) \ge h_1(0)$ або, що те саме, $M\alpha_0 - h(\alpha_0) \ge -h(0)$. Останню нерівність можна переписати у вигляді $f(x+\alpha_0 g) - f(x) \le M\alpha_0$. Аналогічно, за допомогою функції $h_2(\alpha) = h(\alpha) - m\alpha$ доводиться нерівність $f(x+\alpha_0 g) - f(x) \ge m\alpha_0$. □

Наслідок 4.1.1. *Нехай функція $f(x)$ визначена на відкритій опуклій множині S, та нехай її верхня похідна Діні обмежена на цій множині зверху, а нижня похідна Діні обмежена знизу. Інакше кажучи, існує таке число $L > 0$, що*

$$\sup_{x\in S,\|g\|=1} f_D^{\uparrow}(x,g) \le L, \qquad \inf_{x\in S,\|g\|=1} f_D^{\downarrow}(x,g) \ge -L.$$

Тоді функція f задовольняє на множині S умову Ліпшиця, причому константа Ліпшиця співпадає з числом L.

Доведення. Дійсно, нехай $x,y \in S$ і $x = y + \alpha g$, де $\|g\| = 1$, $\alpha = \|x-y\|$. Тоді, як безпосередньо випливає з теореми 4.1.2,

$$-L\|x-y\| \le f(y) - f(x) \le L\|x-y\|.$$

□

Означення 4.1.4. Функція $f(x)$, що визначена на відкритій множині S, називається *рівномірно диференційовною за напрямками* (або *рівномірно диференційовною за Діні*) *в точці* $x \in S$, якщо вона має в цій точці скінченну похідну $f'(x,g)$ за всіма напрямками $g \in \mathbb{R}^n$ та існує таке число $\alpha_0 > 0$, що

$$\frac{1}{\alpha}|f(x+\alpha g) - f(x) - \alpha f'(x,g)| < \varepsilon \quad \forall \alpha \in (0,\alpha_0), \quad \forall g \in B, \tag{4.1.5}$$

де $B = \{g : \|g\| = 1\}$ — одинична сфера.

Покладаючи $\alpha g = v$, нерівність (4.1.5) можна переписати так:

$$|f(x+v) - f(x) - f'_x(v)| < \varepsilon\|v\| \quad \forall v : \|v\| \le \alpha_0. \tag{4.1.6}$$

Таким чином, рівномірна диференційовність за напрямками означає, що

$$\frac{1}{\|v\|}|f(x+v) - f(x) - f'_x(v)| \to 0, \quad \|v\| \to 0. \tag{4.1.7}$$

В подальшому, як правило, будуть розглядатися лише функції, в яких похідна за напрямком $f'_x(g) = f'(x,g)$ в точці x неперервна як функція напрямку g. Наступний приклад показує, що рівномірна диференційовність не забезпечує цю властивість.

Приклад 4.1.1. Нехай $\{x_k\}$ — послідовність різних точок на одиничному колі простору $\mathbb{R}^2$. Покладемо

$$f(x) = \begin{cases} k\lambda, & \text{якщо } x = \lambda x_k \text{ при деяких } k \text{ і } \lambda > 0, \\ 0, & \text{в протилежному випадку.} \end{cases}$$

Функція $f(x)$ рівномірно диференційовна за напрямками в точці $x = 0$, проте похідна $f'_x(g)$ розривна.

В той же час рівномірна диференційовність за напрямками разом із неперервністю похідної еквівалентні диференційовності за Адамаром (тобто, існуванню і скінченності похідної Адамара за всіма напрямками).

Означення 4.1.5. Величина

$$f^{\uparrow}_H(x,g) = \varlimsup_{\alpha\downarrow 0,\, g'\to g} \frac{1}{\alpha}(f(x+\alpha g') - f(x))$$

називається *верхньою похідною Адамара* функції f в точці $x \in S$ за напрямком $g \in \mathbb{R}^n$.

Означення 4.1.6. Величина

$$f^{\downarrow}_H(x,g) = \varliminf_{\alpha\downarrow 0,\, g'\to g} \frac{1}{\alpha}(f(x+\alpha g') - f(x))$$

називається *нижньою похідною Адамара* функції f в точці $x \in S$ за напрямком $g \in \mathbb{R}^n$.

Означення 4.1.7. Границя

$$f'_H(x,g) = \lim_{\alpha\downarrow 0,\, g'\to g} \frac{1}{\alpha}(f(x+\alpha g') - f(x)), \tag{4.1.8}$$

якщо вона існує, називається *похідною Адамара функції f в точці $x \in S$ за напрямком $g \in \mathbb{R}^n$.*

Теорема 4.1.3. *Функція $f(x)$ диференційовна за Адамаром в точці x тоді і тільки тоді, коли вона рівномірно диференційовна за Діні і її похідна $f'(x,g)$ неперервна як функція напрямку.*

Доведення. 1). Нехай функція f диференційовна за Адамаром. Тоді вона диференційовна за Діні. Неперервність $f'(x,g)$ за g доводиться від супротивного. Покажемо рівномірну диференційовність функції f. Для довільного $\varepsilon > 0$ і довільного напрямку g з одиничної сфери B знайдуться такі числа $\eta_g > 0$ і $\alpha_g > 0$, що $|f(x+\alpha g') - f(x) - \alpha f'_H(x,g)| < \alpha \cdot \varepsilon/2$ для всіх $\alpha \in (0,\alpha_g)$ і $g' \in V(g,\eta_g) = \{\tilde{g} | \|g - \tilde{g}\| < \eta_g\}$. Оскільки функція $g \mapsto f'(x,g)$ неперервна, то знайдеться таке $\xi_g > 0$, що $|f'(x,g) - f'(x,g')| < \varepsilon/2$ при $g' \in V(g,\xi_g)$. Покладемо $\delta_g = \min(\eta_g,\xi_g)$. Сім'я околів $\{V(g,\delta_g)\}$ є відкритим покриттям сфери B. Скориставшись компактністю сфери в просторі $\mathbb{R}^n$, виберемо із цього покриття скінченне підпокриття, визначене точками $g_1,g_2,\dots,g_m$. Покладемо $\alpha_0 = \min(\alpha_{g_1},\dots,\alpha_{g_m})$. Якщо $g \in S$, то знайдеться таке k, що $g \in V(g_k,\delta_{g_k})$. При $\alpha < \alpha_0 \leq \alpha_{g_k}$ виконуються нерівності (див.(4.1.8))

$$|f(x+\alpha g) - f(x) - \alpha f'_H(x,g)| \leq$$

$$\leq |f(x+\alpha g) - f(x) - \alpha f'_H(x,g_k)| +$$

$$+\alpha |f'_H(x,g_k) - f'_H(x,g)| < \alpha\varepsilon.$$

Отже $f(x)$ рівномірно диференційовна за напрямками в точці x.

2). Нехай функція $f(x)$ рівномірно диференційовна за напрямками і похідна $f'(x,g)$ неперервна. Маємо

$$|\frac{1}{\alpha}(f(x+\alpha g') - f(x)) - f'(x,g)| \leq$$

$$\leq |\frac{1}{\alpha}(f(x+\alpha g') - f(x)) - f'(x,g')| + |f'(x,g') - f'(x,g)|. \tag{4.1.9}$$

Зафіксуємо $\varepsilon > 0$. Використовуючи рівномірну диференційовність знайдемо таке $\beta_0 > 0$, що справджується нерівність

$$\left| \frac{1}{\beta}(f(x+\beta v) - f(x)) - f'(x,v) \right| < \varepsilon \quad \forall \beta \in (0,\beta_0) \quad \forall v \in B. \tag{4.1.10}$$

Нехай $g' \neq 0$ і $v = g'/\|g'\|$. Тоді, використовуючи додатню однорідність похідної, маємо

$$\frac{1}{\alpha}(f(x+\alpha g') - f(x)) - f'(x,v) =$$

$$= \|g'\| \left[\frac{1}{\alpha \|g'\|}(f(x+\alpha\|g'\|v) - f(x)) - f'(x,v) \right]. \tag{4.1.11}$$

Покладаючи $\alpha\|g'\| = \beta$, $\alpha_0 = \beta_0/\|g\|$, одержимо, використовуючи (4.1.10)

$$\left| \frac{1}{\alpha}(f(x+\alpha g') - f(x)) - f'(x,g') \right| \leq \varepsilon \|g'\| \quad \forall \alpha \in (0,\alpha_0).$$

Це співвідношення вірне і при $g' = 0$. Таким чином, перший доданок у правій частині (4.1.9) може бути зроблений як завгодно малим одразу для всіх g' із деякої кулі з центром в точці g. Другий доданок буде малий при близькому g' до g в силу неперервності функції $g \mapsto f'(x,g)$. Таким чином, f диференційовна в точці x за Адамаром. $\square$

Зауваження 4.1.1. Якщо функція f диференційовна за Адамаром в точці x, то вона неперервна в цій точці. Це випливає, наприклад, із (4.1.7). Таким чином, із рівномірної диференційовності за Діні і неперервності похідної як функції напрямку випливає неперервність самої функції. Якщо диференційовність функції f в точці x не рівномірна, то функція в цій точці може мати розрив, навіть якщо похідна неперервна як функція напрямку. Наведемо відповідний приклад.

Приклад 4.1.2. Розглянемо на площині з координатами (u,v) множину: $S = S_1 \cup S_2 \cup S_3$, де S_1 — опукла множина, обмежена параболою $v = u^2$, S_2 — опукла множина, обмежена параболою $v = -u^2$, S_3 — спільна дотична до цих парабол в точці $x_0 = (0,0)$. Нехай

$$f(x) = \begin{cases} 1, & \text{якщо } x \in S, \\ 0, & \text{якщо } x \notin S. \end{cases}$$

Функція f диференційовна за напрямками в точці x_0 (причому не рівномірно), $f'(x_0,g) = 0$ при всіх g. В той же час функція $f(x)$ розривна в нулі.

Нагадаємо, що вектор $a \in \mathbb{R}^n$ називається *похідною Гато (слабкою похідною)* функції $f(x)$ в точці x, якщо для довільного $g \in \mathbb{R}^n$ виконується співвідношення

$$f(x + \alpha g) = f(x) + \alpha \langle a,g \rangle + o_g(\alpha), \quad \frac{o_g(\alpha)}{\alpha} \underset{\alpha \to 0}{\longrightarrow} 0.$$

Іншими словами, похідна Гато — це похідна Діні, яка лінійно залежить від $g \in \mathbb{R}^n$. Похідна Гато функції $f(x)$ в точці x співпадає з градієнтом $a = \nabla f(x)$.

Функція $f(x)$ *диференційовна за Фреше* в точці x, якщо справджується співвідношення

$$f(x + v) = f(x) + \langle \nabla f(x),v \rangle + o(v), \quad \frac{o(v)}{\|v\|} \underset{\|v\| \to 0}{\longrightarrow} 0. \tag{4.1.12}$$

В силу (4.1.7) це означає, що $f(x)$ рівномірно диференційовна за напрямками, причому похідна $f'(x,g)$ залежить від $g \in \mathbb{R}^n$ лінійно, отже неперервно. Теорема 4.1.3 показує, що диференційовність за Фреше рівносильна “лінійній диференційовності” за Адамаром.

При деяких припущеннях диференційовна за Діні функція $f(x)$ буде диференційовною за Адамаром. Наведемо результат такого роду.

Теорема 4.1.4. *Нехай функція $f(x)$ диференційовна за напрямками в точці x і задовольняє умову Ліпшиця з константою L в деякому околі цієї точки. Тоді:*
1) похідна $f'(x,g)$ задовольняє на всьому просторі $\mathbb{R}^n$ умову Ліпшиця з тією ж константою L;
2) функція f диференційовна за Адамаром.

Доведення. При достатньо малих α маємо

$$|f(x + \alpha g) - f(x + \alpha g')| \le \alpha L \|g - g'\|.$$

Тому
$$\left|\frac{1}{\alpha}(f(x+\alpha g)-f(x))-\frac{1}{\alpha}(f(x+\alpha g')-f(x))\right| \le L\|g-g'\|. \tag{4.1.13}$$
Переходячи до границі при $\alpha\downarrow 0$, одержимо з (4.1.13)
$$|f'(x,g)-f'(x,g')| \le L\|g-g'\|.$$

Перевіримо, що f диференційовна за Адамаром. Маємо
$$\lim_{\alpha\downarrow 0, g'\to g} \tfrac{1}{\alpha}(f(x+\alpha g')-f(x)) =$$
$$= \lim_{\alpha\downarrow 0, g'\to g} \tfrac{1}{\alpha}(f(x+\alpha g')-f(x+\alpha g))+$$
$$+\lim_{\alpha\downarrow 0} \tfrac{1}{\alpha}(f(x+\alpha g)-f(x)).$$

Оскільки f задовольняє умову Ліпшиця, то перша із виписаних праворуч границь існує і дорівнює нулю. Друга границя, що дорівнює $f'(x,g)$, також існує. Звідси і випливає диференційовність за Адамаром. □

Сформулюємо теорему про похідну композиції функцій. Нехай задано відображення $H : S\to\mathbb{R}^m$, де S — відкрита множина в $\mathbb{R}^n$. Координатні функції цього відображення позначимо через $h_1(x),\ldots,h_m(x)$, так що $H(x)=(h_1(x),\ldots,h_m(x))$. Похідна $H'(x,g)$ відображення H в точці x за напрямком g визначається як границя
$$\lim_{\alpha\downarrow 0}\frac{1}{\alpha}(H(x+\alpha g)-H(x)).$$
Як і у випадку дійсної функції, використовується також термін *похідна Діні*. Границя
$$\lim_{\alpha\downarrow 0, g'\to g}\frac{1}{\alpha}(H(x+\alpha g')-H(x)),$$
яку ми позначимо таким же символом $H'(x,g)$, називається похідною Адамара відображення H в точці x за напрямком g. Зрозуміло, що відображення H в точці x диференційовне за Діні або за Адамаром тоді і тільки тоді, коли координатні функції h_i диференційовні відповідно за Діні або за Адамаром. При цьому $H'(x,g)=(h'_1(x,g),\ldots,h'_m(x,g))$.

Теорема 4.1.5. *Нехай S_1 — відкрита множина в просторі $\mathbb{R}^n$, S_2 — відкрита множина в $\mathbb{R}^m$, відображення $H : S_1\to S_2$ диференційовне за напрямками в точці $x\in S_1$, функція f визначена на S_2 і диференційовна за Адамаром в точці $y=H(x)$. Тоді функція $\varphi(x)=f(H(x))$ диференційовна в точці x за напрямками, причому*
$$\varphi'(x,g)=f'(H(x),H'(x,g)). \tag{4.1.14}$$

Якщо, крім того, відображення H диференційовне за Адамаром, то і φ диференційовне за Адамаром.

Доведення. Покладемо

$$\omega(\alpha,g) = \frac{1}{\alpha}[H(x+\alpha g) - H(x) - \alpha H'(x,g)],$$

$$\psi(u) = \frac{1}{\|u\|}(f(H(x)+u) - f(H(x)) - f'(H(x),u)), \quad u \neq 0,$$

$$u(\alpha,g) = H'(x,g) + \omega(\alpha,g).$$

Нехай відображення H диференційовне за напрямками. Тоді $\omega(\alpha,g) \xrightarrow[\alpha\downarrow 0]{} 0$ і тим самим $u(\alpha,g) \xrightarrow[\alpha\downarrow 0]{} H'(x,g)$. Маємо

$$\varphi(x+\alpha g) = f(H(x+\alpha g)) = f(H(x) + \alpha H'(x,g) + \alpha\omega(\alpha,g)) =$$
$$= f(H(x)) + \alpha f'(H(x),u(\alpha,g)) + \psi(\alpha u(\alpha,g)) \cdot \|\alpha u(\alpha,g)\|. \qquad (4.1.15)$$

Оскільки із диференційовності за Адамаром функції f випливає її рівномірна диференційовність і неперервність похідної $f'(Hx,u)$ як функції напрямку u, то з (4.1.15) випливає існування похідної Діні $\varphi'(x,g)$ функції $\varphi(x)$ і рівність (4.1.14).

Припустимо тепер, що відображення H диференційовне за Адамаром. Тоді, якщо $g' \to g$ і $\alpha \downarrow 0$, то $\omega(\alpha,g') \to 0$, $H'(x,g') \to H'(x,g)$ і, відповідно, $u(\alpha,g') \to H'(x,g)$. Застосовуючи (4.1.15), переконуємося в існуванні похідної Адамара $\varphi'(x,g)$. □

Далі дослідимо диференціальні властивості *функції максимуму*, тобто функції виду

$$\varphi(x) = \max_{y\in Y} f(x,y), \quad x \in X. \qquad (4.1.16)$$

Вважатимемо, що $f(x,y)$ — неперервна за сукупністю змінних функція, що визначена на $X \times Y$, де X — відкрита множина в $\mathbb{R}^n$, а Y — компактна множина в $\mathbb{R}^m$. Відзначимо насамперед, що функція φ неперервна. Дійсно, нехай $x \in X$ і V — компактний окіл точки x. Функція f рівномірно неперервна на компактній множині $V \times Y$, звідки випливає, що для будь-якого $\varepsilon > 0$ знайдеться таке $\delta > 0$, що

$$f(x',y) - \varepsilon < f(x,y) < f(x',y) + \varepsilon \quad \forall y \in Y, x' \in B_\delta(x).$$

Переходячи в цих нерівностях до максимуму по y, переконуємося в неперервності функції φ на X.

Для точки $x \in X$ позначимо через $R(x)$ множину точок $y \in Y$, на яких в (4.1.16) досягається максимум:

$$R(x) = \{y \in Y \,|\, \varphi(x) = f(x,y)\}. \qquad (4.1.17)$$

Із неперервності функцій φ та f випливає, що множина $R(x)$ замкнута і компактна. Більш того, справедливе наступне твердження.

Теорема 4.1.6. *Багатозначне відображення $x \mapsto R(x)$ замкнуте.*

Доведення. Нехай $x_k \to x$, $y_k \to y$, $y_k \in R(x_k)$. Тоді $\varphi(x_k) = f(x_k,y_k)$. Оскільки функції f та φ неперервні, то $\varphi(x) = f(x,y)$, звідки і випливає справедливість теореми. □

При дослідженні диференціальних властивостей функції максимуму важливу роль відіграє рівномірна відносно параметра диференційовність за напрямками. Нехай $f(x,y)$ — неперервна функція двох змінних, що визначена на $X \times Y$, де X — відкрита множина, Y — компактна множина. Припустимо, що для кожного y функція $x' \mapsto f(x',y)$ диференційовна за деяким напрямком g в точці x (за Діні чи за Адамаром). Позначимо її похідну за цим напрямком через $f'(x,y,g)$. Нехай

$$\omega_{x,g}(\alpha,y) = f(x+\alpha g,y) - f(x,y) - \alpha f'(x,y,g). \tag{4.1.18}$$

За означенням, для довільного $\varepsilon > 0$ і для довільного $y \in Y$ існує таке число $\delta(\varepsilon,y)$, що $|\omega_{x,g}(\alpha,y)| < \varepsilon\alpha$ при $0 < \alpha < \delta(\varepsilon,y)$.

Функція f *диференційовна за напрямком g в точці x рівномірно відносно параметра y*, якщо величину $\delta(\varepsilon,y)$ можна вибрати незалежно від y, тобто за довільним $\varepsilon > 0$ знайдеться таке $\delta(\varepsilon) > 0$, що $|\omega_{x,g}(\alpha,y)| < \varepsilon\alpha$ при всіх y і всіх $\alpha \in (0,\delta(\varepsilon))$.

З диференційовності функції $f(x,y)$ за напрямком g в точці x рівномірно відносно y випливає неперервність функції $y \mapsto f'(x,y,g)$. Дійсно, ця функція є рівномірною границею при $\alpha \downarrow 0$ сім'ї неперервних функцій $h_\alpha(y) = \frac{1}{\alpha}(f(x+\alpha g,y) - f(x,y))$.

Теорема 4.1.7. *Нехай*

$$\varphi(x) = \max_{y \in Y} f(x,y) \quad \forall x \in X,$$

де X — відкрита множина в $\mathbb{R}^n$, Y — компактна множина в $\mathbb{R}^m$, функція $f(x,y)$ неперервна за сукупністю змінних і диференційовна за напрямком g в точці x рівномірно відносно параметра y. Тоді функція φ диференційовна за напрямком g, причому

$$\varphi'(x,g) = \max_{y \in R(x)} f'(x,y,g),$$

де множина $R(x)$ визначена формулою (4.1.17).

Доведення. Якщо $y \in R(x)$, то

$$\frac{1}{\alpha}(\varphi(x+\alpha g) - \varphi(x)) \geq \frac{1}{\alpha}(f(x+\alpha g,y) - f(x,y)).$$

Переходячи до границі при $\alpha \downarrow 0$, отримаємо $\varphi_D^{\downarrow}(x,g) \geq f'(x,y,g)$, звідки, зважаючи на довільність $y \in R(x)$, випливає нерівність

$$\varphi_D^{\downarrow}(x,g) \geq \max_{y \in R(x)} f'(x,y,g) \tag{4.1.19}$$

(максимум в правій частині (4.1.19) досягається через компактність множини $R(x)$ і неперервність функції $y \mapsto f'(x,y,g)$).

Розглянемо тепер елемент y_α із множини $R(x+\alpha g), \alpha > 0$. Маємо

$$f(x+\alpha g,y_\alpha) - f(x,y_\alpha) = \alpha f'(x,y_\alpha,g) - \omega_{x,g}(\alpha,y_\alpha),$$

де $\omega_{x,g}$— функція, що визначена формулою (4.1.18). Враховуючи означення множини $R(x+\alpha g)$, маємо

$$\frac{1}{\alpha}(\varphi(x+\alpha g) - \varphi(x)) \le \frac{1}{\alpha}(f(x+\alpha g,y_\alpha) - f(x,y_\alpha)) =$$

$$= f'(x,y_\alpha,g) - \frac{1}{\alpha}\omega_{x,g}(\alpha,y_\alpha). \qquad (4.1.20)$$

Оскільки Y — компактна множина, то можна вважати, що існує границя $\lim_{\alpha\downarrow 0} y_\alpha = y'$. Оскільки відображення R замкнуте (див. теорему 4.1.6), то $y' \in R(x)$. Із неперервності функції $y \mapsto f'(x,y,g)$ випливає, що

$$f'(x,y_\alpha,g) \to f'(x,y',g) \le \max_{y\in R(x)} f'(x,y,g).$$

Використовуючи рівномірну диференційовність функції f отримаємо, що для довільного $\varepsilon > 0$ при достатньо малих $\alpha > 0$ виконується нерівність $\frac{1}{\alpha}\omega_{x,g}(\alpha,y_\alpha) < \varepsilon$; тому $\lim\limits_{\alpha\downarrow 0} \alpha^{-1}\omega_{x,g}(\alpha,y_\alpha) = 0$. Враховуючи сказане і переходячи в (4.1.20) до верхньої границі, отримаємо

$$\varphi_D^{\uparrow}(x,g) \le \max_{y\in R(x)} f'(x,y,g). \qquad (4.1.21)$$

Доведення теореми випливає з нерівності (4.1.19) і нерівності (4.1.21). □

Зауваження 4.1.2. Якщо функція f задовольняє умову Ліпшиця з константою L за сукупністю змінних, то і функція φ задовольняє цю умову з тією ж константою. Якщо в такому випадку умови теореми виконані для всіх напрямків g, то функція φ диференційовна за Адамаром. Це випливає з теореми 4.1.6.

Зауваження 4.1.3. Якщо диференційовність функції f не рівномірна за параметром, то теорема 4.1.7 перестає бути справедливою. Можна навести відповідний приклад. Диференційовність φ можна гарантувати і без рівномірної диференційовності f. При цьому можна одержати формулу для похідної, що відмінна від тієї, яка наведена в теоремі 4.1.7.

Наслідок 4.1.2. *Нехай*

$$\varphi(x) = \max_{i=1,\dots,m} f_i(x), \quad x \in X,$$

де функції f_i визначені на відкритій множині X, неперервні і диференційовні в точці x за напрямком g. Тоді похідна $\varphi'(x,g)$ існує і обчислюється за формулою

$$\varphi'(x,g) = \max_{i\in R(x)} f_i'(x,g).$$

Наслідок 4.1.3. *Нехай X — відкрита, а Y — компактна множини,*

$$\varphi(x) = \max_{y \in Y} f(x,y), \quad x \in X,$$

де f — визначена на $X \times Y$ неперервна функція, що має неперервну частинну похідну $f'_x(x,y)$. Тоді функція φ диференційовна за будь-яким напрямком g, причому

$$\varphi'(x,g) = \max_{y \in R(x)} \langle f'_x(x,y), g \rangle. \tag{4.1.22}$$

Функцію $g \mapsto \varphi'(x,g)$, де $\varphi'(x,g)$ визначена формулою (4.1.22), можна представити як максимум деякої множини лінійних функцій. Звідси одразу випливає, що вона субадитивна: $\varphi'(x,g_1+g_2) \leq \varphi'(x,g_1)+\varphi'(x,g_2)$ для всіх g_1, g_2. Оскільки ця функція, як похідна, додатньо однорідна, то вона і сублінійна.

4.2. АПРОКСИМАЦІЯ МНОЖИН ЗА ДОПОМОГОЮ КОНУСІВ

Локальну апроксимацію множин можна здійснюватись різними способами. Найбільш прості та вживані серед них використовують конуси як апроксимуючі об'єкти.

Розглянемо множину $S \in \mathbb{R}^n$. Нехай точка x належить замиканню $\mathrm{cl}\, S$ цієї множини.

Означення 4.2.1. Вектор g називається *допустимим напрямком* множини S в точці x, якщо знайдеться таке число $\alpha_g > 0$, що

$$x + \alpha g \in S \quad \forall \alpha \in (0,\alpha_g).$$

Сукупність всіх допустимих в точці x напрямків множини S позначимо через $\gamma(x,S)$.

Означення 4.2.2. Вектор g називається *дотичним напрямком* в точці x до множини S, якщо знайдуться $\alpha_g > 0$ і така функція $\psi_g(\alpha) : [0,\alpha_g] \to \mathbb{R}^n$, що

$$x + \alpha g + \psi_g(\alpha) \in S \quad \forall \alpha \in (0,\alpha_g); \quad \frac{\psi_g(\alpha)}{\alpha} \to 0,\ \alpha \downarrow 0. \tag{4.2.1}$$

Іноді замість (4.2.1) використовують запис $x + \alpha g + o(\alpha) \in S$. Сукупність всіх дотичних в точці x до множини S векторів позначимо через $K(x,S)$.

Означення 4.2.3. Вектор g називається *можливим напрямком* множини S в точці x, якщо знайдуться такі послідовності $\{g_k\}$ і $\{\alpha_k\}$, що

$$g_k \in \mathbb{R}^n, \quad \alpha_k \in \mathbb{R}^1_+, \quad g_k \to g, \quad \alpha_k \downarrow 0, \quad x + \alpha_k g_k \in S. \tag{4.2.2}$$

Сукупність всіх можливих напрямків множини S в точці x позначимо через $\Gamma(x,S)$.

Напрямок $g \neq 0$ є можливим тоді і тільки тоді, коли знайдеться послідовність $\{x_k\}$, яка має такі властивості:

$$x_k \in S, \quad x_k \neq x, \quad x_k \to x, \quad \frac{x_k - x}{\| x_k - x \|} \to \frac{g}{\| g \|}. \tag{4.2.3}$$

Якщо виконується (4.2.2) і $g \neq 0$, то покладаючи $x_k = x + \alpha_k g_k$, переконуємося в справедливості (4.2.3). Якщо ж має місце (4.2.3), то для послідовностей $\{\alpha_k\}$, $\{g_k\}$, де $\alpha_k = \|x_k - x\|/\|g\|, g_k = (x_k - x)/\alpha_k$, виконується (4.2.2).
В термінах околів означення можливого напрямку таке: для будь-яких чисел $\alpha_0 > 0, \varepsilon > 0$ знайдуться елемент $w \in g + B_\varepsilon(0)$ та число $\alpha \in (0,\alpha_0)$, при яких $x + \alpha w \in S$.

Безпосередньо з означень випливає, що множини $\gamma(x,S)$, $K(x,S)$ і $\Gamma(x,S)$ це конуси (тобто вони містять разом із кожним своїм елементом g і весь промінь $\{\lambda g | \lambda \geq 0\}$). Будемо називати ці множини *конусами, що апроксимують множину S поблизу точки x*. При цьому $K(x,S)$ часто називають *дотичним конусом* множини S в точці x, а $\Gamma(x,S)$– *конусом Булігана* множини S в точці x. Конуси $\gamma(x,S)$, $K(x,S)$ і $\Gamma(x,S)$ іноді записують за допомогою формул, в яких безпосередньо фігурують S та x. Для того щоб записати їх, помітимо, що при $\alpha \neq 0$ співвідношення $x + \alpha g \in S$ можна переписати у вигляді $g \in \frac{1}{\alpha}(S - x)$. Звідси одразу випливає рівність

$$\gamma(x,S) = \bigcup_{\alpha_0 > 0} \bigcap_{0 < \alpha < \alpha_o} \frac{1}{\alpha}(S - x). \tag{4.2.4}$$

Крім того

$$\Gamma(x,S) = \bigcap_{\varepsilon > 0} \bigcap_{\alpha_0 > 0} \bigcup_{\alpha \in (0,\alpha_o)} \left(\frac{1}{\alpha}(S - x) + \varepsilon B_1(0) \right), \tag{4.2.5}$$

де $B_1(0)$– одинична куля з центром в нулі.

Щоб записати подібним чином дотичний конус $K(x,S)$ розглянемо множину Ψ_n, яка складається з функцій ψ, визначених на деякому відрізку $[0,\alpha_0]$, які діють в $\mathbb{R}^n$ і мають властивість: $\alpha^{-1}\psi(\alpha) \to 0$ при $\alpha \downarrow 0$. Назвемо Ψ_n *множиною нескінченно малих першого порядку*. Включення $g \in K(x,S)$ рівносильне тому, що $x + \alpha g + \psi(\alpha) \in S$, де $\alpha^{-1}\psi(\alpha) \to 0$ при $\alpha \downarrow 0$. Можна вважати, що область визначення функцій ψ співпадає з деяким відрізком $[0,\alpha_0]$, що не залежить від g, тому $\psi \in \Psi_n$. Безпосередньо з означення випливає рівність

$$K(x,S) = \bigcup_{\psi \in \Psi_n} \bigcup_{\alpha_0 > 0} \bigcap_{0 < \alpha < \alpha_o} \frac{1}{\alpha}(S - x - \psi(\alpha)). \tag{4.2.6}$$

Зрозуміло, що

$$\gamma(x,S) \subset K(x,S) \subset \Gamma(x,S). \tag{4.2.7}$$

Якщо x – внутрішня точка S, то всі конуси, які апроксимують S в околі точки x, співпадають з простором $\mathbb{R}^n$. Якщо ж x – ізольована точка, то ці конуси складаються лише з нуля. Всі три вказані конуси здійснюють локальну апроксимацію S. Це означає, що при будь-якому $\varepsilon > 0$ мають місце рівності

$$\gamma(x,S) = \gamma(x,S \cap B_\varepsilon(x)),$$

$$K(x,S) = K(x,S \cap B_\varepsilon(x)),$$

$$\Gamma(x,S) = \Gamma(x,S \cap B_\varepsilon(x)).$$

Справедливість цих рівностей випливає безпосередньо з означень.

Конус Булігана $\Gamma(x,S)$ завжди замкнутий. Дійсно, нехай $g \in \operatorname{cl}\Gamma(x,S)$. Зафіксуємо числа $\varepsilon > 0$ та $\alpha_0 > 0$ і знайдемо такий елемент $u \in \Gamma(x,S)$, що $\|u-g\| < \varepsilon/2$. Оскільки $u \in \Gamma(x,S)$, то знайдуться елемент $v \in B_{\varepsilon/2}(u)$ і число $\alpha \in (0,\alpha_0)$, при яких виконується включення $x+\alpha v \in S$. Оскільки $\|v - g\| < \varepsilon$, $\alpha \in (0,\alpha_0)$, де ε, α_0 – довільні додатні числа, то $g \in \Gamma(x,S)$. Множина $\gamma(x,S)$ є найбільш простою із трьох конусів. Але в багатьох випадках вона здійснює "неповну апроксимацію" множини S і несе мало інформації про те, як влаштована S поблизу точки x. Для деяких важливих в застосуваннях множин S конус $\gamma(x,S)$ складається лише з нуля. Це відноситься, наприклад, до поверхні $S = \{y | f(y) = 0\}$, де f – строго опукла функція. В зв'язку зі сказаним стає зрозуміла необхідність використання конусів $K(x,S)$, $\Gamma(x,S)$, які, однак, облаштовані набагато складніше, ніж $\gamma(x,S)$.

Корисно знайти ті множини, для яких "складний" конус $\Gamma(x,S)$ відновлюється за більш простим конусом $\gamma(x,S)$. Для цього корисно використовувати включення:

$$\gamma(x,S) \subset \operatorname{cone}(S - x), \quad \Gamma(x,S) \subset \operatorname{cl}\operatorname{cone}(S - x). \tag{4.2.8}$$

Співвідношення (4.2.8) випливають безпосередньо з означень. Використовуючи (4.2.7) і (4.2.8), легко переконатися в справедливості наступного твердження.

Теорема 4.2.1. *Якщо* $\gamma(x,S) = \operatorname{cone}(S - x)$*, то* $\Gamma(x,S) = \operatorname{cl}\gamma(x,S)$.

Найбільш просто влаштованими підмножинами векторних просторів є опуклі множини. Апроксимацію опуклої множини S поблизу точки $x \in \operatorname{cl} S$ здійснюють, як правило, за допомогою *опорного конуса* $K_x(S)$. За означенням

$$K_x(S) = \operatorname{cone}(S - x) = \bigcup_{\lambda > 0} \lambda(S - x).$$

Те, що $K_x(S)$ – конус, випливає з означення. Покажемо, що цей конус є опуклим. Нехай $u,v \in K_x(S)$. Тоді при деяких $\lambda > 0, \mu > 0$ виконується

$u/\lambda \in (S-x)$,$v/\mu \in (S-x)$. Оскільки $(S-x)$ – опукла множина, то

$$\frac{u+v}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}\frac{u}{\lambda} + \frac{\mu}{\lambda+\mu}\frac{v}{\mu} \in (S-x)$$

і тому $u+v \in (\lambda+\mu)(S-x) \subset K_x(S)$.

Теорема 4.2.2. *Нехай S – опукла множина, $x \in S$. Тоді*

$$\gamma(x,S) = K_x(S), \quad \Gamma(x,S) = \operatorname{cl} K_x(S).$$

Доведення. 1) Як показує формула (4.2.8), $\gamma(x,S) \subset K_x(S)$. Перевіримо протилежне включення. Нехай $g \in K_x(S)$ і число $\lambda > 0$ таке, що $g \in \lambda(S-x)$. Тоді $x+\frac{1}{\lambda}g \in S$. Нехай $\alpha < 1/\lambda$. Тоді $x+\alpha g = \beta(x+\frac{1}{\lambda}g)+(1-\beta)x$, де $\beta = \alpha\lambda$. Оскільки S опукла, $x \in S$,$\beta \in (0,1)$, то $x+\alpha g \in S$. Таким чином $g \in \gamma(x,S)$, отже $K_x(S) \subset \gamma(x,S)$.
2) Рівняння $\Gamma(x,S) = \operatorname{cl} K_x(S)$ випливає з теореми 4.2.1. □

Зауваження 4.2.1. Нижче показано, що $K(x,S) = \operatorname{cl} K_x(S)$ якщо S – опукла множина та $x \in S$.

Нормальний конус $N_x(S)$ до опуклої множини S в точці $x \in S$ визначається співвідношенням

$$N_x(S) = \{v \in \mathbb{R}^n | \langle v,x\rangle = \max_{y\in S}\langle v,y\rangle\}.$$

Іншими словами,

$$N_x(S) = \{v \in \mathbb{R}^n | \langle v,x\rangle = p_S(v)\},$$

де $p_S(v) = \sup\limits_{y\in S}\langle v,y\rangle$ – опорна функція опуклої множини S.

Теорема 4.2.3. *Нехай S – опукла множина, $x \in S$. Тоді*

$$N_x(S) = -K_x^+(S).$$

Тут $^+$ позначає перехід до спряженого конуса.

Доведення. 1) Нехай $v \in (-K_x^+(S))$. Оскільки $K_x(S) \supset S-x$, то для всіх $y \in S$ виконується нерівність $\langle v,y-x\rangle \le 0$, з якої випливає, що $p_S(v) = \langle v,x\rangle$.
2) Якщо $g \in K_x(S)$, то при достатньо малих $\alpha > 0$ виконується включення $x+\alpha g \in S$. Тому якщо $v \in N_x(S)$, то $\langle v,x+\alpha g\rangle \le \langle v,x\rangle$, звідки випливає нерівність $\langle v,g\rangle \le 0$. Отже $N_x(S) \subset -K_x^+(S)$. □

Наслідок 4.2.1. *Множина $N_x(S)$ є замкнутим опуклим конусом.*

Апроксимація множини S в околі точки x здійснюється конусами локально. Можна казати, що на кожному з променів, що входять в конус, апроксимація із заданою точністю здійснюється лише на деякому відрізку, довжина якого залежить від цього променя. Наприклад, якщо g – дотичний напрямок, то справедливе співвідношення $x + \alpha g + \psi_g(\alpha) \in S$, де $\alpha^{-1}\psi_g(\alpha) \to 0$ при $\alpha \downarrow 0$, тобто для довільного $\delta > 0$ знайдеться таке $\alpha_g > 0$, що $\|\alpha^{-1}\psi_g(\alpha)\| < \delta$ при $\alpha \in (0,\alpha_g]$. Зрозуміло, що α_g залежить від g і може не знайтися жодного такого α_0, що $\|\alpha^{-1}\psi_g(\alpha)\| < \delta$ при всіх $g \in K(x,S)$ і $\alpha \in (0,\alpha_0]$.

Опишемо конуси, які здійснюють рівномірну локальну апроксимацію з точністю до величин першого порядку малості. Дамо відповідне означення.

Замкнутий конус $\mathcal{K}(x,S)$ є *рівномірною апроксимацією першого порядку множини S в околі точки $x \in \operatorname{cl} S$*, якщо

$$\rho(S \cap B_\delta(x),(x + \mathcal{K}(x,S)) \cap B_\delta(x)) = o(\delta). \tag{4.2.9}$$

Тут ρ – метрика Хаусдорфа

$$\rho(S_1,S_2) = \max\{\max_{x\in S_1}\min_{y\in S_2}\|x-y\|;\ \max_{y\in S_2}\min_{x\in S_1}\|x-y\|\}.$$

Відзначимо, що на відміну від конусів $\gamma(x,S)$, $K(x,S)$, $\Gamma(x,S)$, визначених як об'єднання променів, що мають деякі властивості, рівність (4.2.9) визначає одразу весь конус $\mathcal{K}(x,S)$.

Теорема 4.2.4. *Нехай існує конус $\mathcal{K}(x,S)$, який є рівномірною апроксимацією першого порядку множини S в околі точки x. Тоді:*

1) $\mathcal{K}(x,S) = K(x,S) = \Gamma(x,S)$;

2) для будь-якого $g \in K(x,S)$ можна знайти таку функцію $\psi_g(\alpha)$, що $x + \alpha g + \psi_g(\alpha) \in S$ і $\alpha^{-1}\psi_g(\alpha) \to 0$ при $\alpha \downarrow 0$ рівномірно по g.

Доведення. Спочатку покажемо, що $\Gamma(x,S) \subset \mathcal{K}(x,S)$. Припускаючи протилежне, знайдемо елемент $g \in \Gamma(x,S)$, який не міститься в $\mathcal{K}(x,S)$. Оскільки $g \in \Gamma(x,S)$, то знайдуться послідовності $g_k \to g$, $\alpha_k \downarrow 0$, для яких справедливе включення $x + \alpha_k g_k \in S$. Оскільки $g \notin \mathcal{K}(x,S)$ і конус $\mathcal{K}(x,S)$ замкнутий, то $\rho(g,\mathcal{K}(x,S)) = \gamma > 0$. При достатньо великих k виконується нерівність $\rho(g_k,\mathcal{K}(x,S)) > \gamma/2$. Множина $\mathcal{K}(x,S)$ є конусом, тому $\rho(\alpha_k g_k,\mathcal{K}(x,S)) > \frac{\gamma}{2}\alpha_k$, звідки

$$\rho(x + \alpha_k g_k, x + \mathcal{K}(x,S)) > \frac{\gamma}{2}\alpha_k. \tag{4.2.10}$$

Оскільки $x + \alpha_k g_k \in S \cap B_{\alpha_k\|g_k\|}(x)$, то нерівність (4.2.10) означає, що

$$\rho(S \cap B_{\alpha_k\|g_k\|}(x),(x + \mathcal{K}(x,S)) \cap B_{\alpha_k\|g_k\|}(x)) > \frac{\gamma}{2}\alpha_k.$$

Це суперечить співвідношенню (4.2.9), що визначає конус $\mathcal{K}(x,S)$. Таким чином включення $\Gamma(x,S) \subset \mathcal{K}(x,S)$ доведене.

Покажемо тепер, що $\mathcal{K}(x,S) \subset K(x,S)$. Покладемо $\varphi(\delta) = \rho(S \cap B_\delta,(x + \mathcal{K}(x,S)) \cap B_\delta)$. Тоді $\delta^{-1}\varphi(\delta) \to 0$ при $\delta \downarrow 0$. Нехай $g \in \mathcal{K}(x,S)$, $\|g\| = 1$.

Для довільного $\delta > 0$ знайдеться такий елемент $x_\delta \in S$, що $\|x_\delta - x\| \le \delta$ і $\|x_\delta - (x + \delta g)\| = \varphi(\delta)$. Покладемо $\psi_g(\delta) = x_\delta - x - \delta g$. Тоді $x_\delta = x + \delta g + \psi_g(\delta) \in S$ і

$$\|\psi_g(\delta)\|/\delta = \varphi(\delta)/\delta \to 0, \quad \delta \downarrow 0. \tag{4.2.11}$$

Ці співвідношення показують, що $g \in K(x,S)$. Таким чином, $\mathcal{K}(x,S) \subset K(x,S)$. Оскільки, крім того, $K(x,S) \subset \Gamma(x,S)$, то

$$\mathcal{K}(x,S) = K(x,S) = \Gamma(x,S)$$

і перша частина теореми доведена. Друга частина одразу випливає із співвідношення (4.2.11). □

Використовуючи теорему 4.2.4, легко навести приклад множини, для якої не існує конуса, який є рівномірною апроксимацією першого порядку.

Приклад 4.2.1. Нехай $\lambda_k \to 0$, причому $\lim \frac{\lambda_{k-1}}{\lambda_k} = c < 1$; $\lambda_0 = 0$. Покладемо $S = \{\lambda_k g | k = 0,1, \ldots, +\infty\}$, де g – деякий фіксований елемент із $\mathbb{R}^n$. Нехай $x = 0$. Знайдемо конус $K(x,S)$. Якщо ненульовий елемент h входить в цей конус, то $\alpha h + o(\alpha) \in S$. Позначимо через L_g промінь з вершиною в нулі, що проходить через точку g. Припускаючи, що $h \notin L_g$, одержуємо $\rho(\alpha h,S) \ge \rho(\alpha h,L_g) = \alpha\rho(h,L_g)$. Ця обставина показує, що не знайдеться функцій $\psi(\alpha)$, які задовольняють співвідношення $\alpha h + \psi(\alpha) \in S$, $\psi(\alpha)/\alpha \to 0, \alpha \downarrow 0$. Таким чином, конус $K(x,S)$ міститься в промені L_g. Покажемо, що $g \notin K(x,S)$. Дійсно, нехай число α_k співпадає з серединою відрізка $[\lambda_{k-1},\lambda_k]$, тобто $\alpha_k = (\lambda_{k-1} + \lambda_k)/2$. Припустимо, що $\alpha_k g + \psi(\alpha_k) \in S$. Тоді $\|\psi(\alpha_k)\| \ge (\alpha_k - \lambda_{k-1})\|g\| = (\lambda_k - \lambda_{k-1})\|g\|/2$. Маємо

$$\frac{\|\psi(\alpha_k)\|}{\alpha_k} = \frac{\lambda_k - \lambda_{k-1}}{\lambda_k + \lambda_{k-1}}\|g\| = \frac{1 - \lambda_{k-1}/\lambda_k}{1 + \lambda_{k-1}/\lambda_k}\|g\| \to \frac{1-c}{1+c}\|g\| \ne 0.$$

Таким чином $g \notin K(x,S)$. Звідси випливає, що $\lambda g \notin K(x,S)$ при $\lambda > 0$. Зі сказаного випливає рівність $K(x,S) = \{0\}$. В той же час, як неважко перевірити, $\Gamma(x,S) = L_g$. Ми показали, що $K(x,S) \ne \Gamma(x,S)$ і тому, використовуючи теорему 4.2.4, конуса, що здійснює рівномірну апроксимацію першого порядку, не існує.

4.3. АПРОКСИМАЦІЯ ФУНКЦІЙ ЗА ДОПОМОГОЮ КОНУСІВ

Розглянемо питання про апроксимацію заданої функції в околі деякої точки додатньо однорідною функцією напрямку за допомогою конусів, які апроксимують множини, що пов'язані з функцією – її надграфік і підграфік.

Нехай $f(x)$ – визначена на $\mathbb{R}^n$ функція, яка приймає значення із розширеної числової прямої $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$. Множина

$$\mathrm{dom} f = \{x \in \mathbb{R}^n : |f(x)| < +\infty\}$$

називається *ефективною множиною* функції f, а множини

$$\mathrm{gr}\, f = \{[x,\mu] \in \mathbb{R}^n \times \mathbb{R} : \mu = f(x)\}$$

$$\mathrm{epi}\, f = \{[x,\mu] \in \mathbb{R}^n \times \mathbb{R} : \mu \geq f(x)\}$$

$$\mathrm{hyp}\, f = \{[x,\mu] \in \mathbb{R}^n \times \mathbb{R} : \mu \leq f(x)\}$$

називаються відповідно *графіком, надграфіком* і *підграфіком* цієї функції. Зрозуміло, що $(\mathrm{epi} f) \cap (\mathrm{hyp} f) = \mathrm{gr} f$. Якщо $f(x) = +\infty$, то вертикальна пряма $\Pi_x = \{(x,\mu)|\mu \in \mathbb{R}\}$ не перетинається з надграфіком і повністю міститься в підграфіку; якщо $f(x) = -\infty$, то ця пряма повністю міститься в надграфіку і не перетинається з надграфіком; якщо $x \in \mathrm{dom}\, f$, то пряма Π_x може бути представлена як об'єднання двох променів: Π_x^+ і Π_x^-,

$$\Pi_x^+ = \{[x,\mu] \in \Pi_x : \mu \geq f(x)\},$$

$$\Pi_x^- = \{[x,\mu] \in \Pi_x : \mu \leq f(x)\},$$

де перший міститься в надграфіку $\mathrm{epi}\, f$ функції f, а другий міститься в підграфіку $\mathrm{hyp}\, f$ функції f.

Множину $S \subset \mathbb{R}^n \times \mathbb{R}$ будемо називати *стійкою вгору*, якщо із співвідношення $[x,\mu] \in S$, $\mu' \geq \mu$ випливає, що $[x,\mu'] \in S$. Зрозуміло, що надграфік $\mathrm{epi}\, f$ – стійка вгору множина.

Нехай S — довільна стійка вгору множина, $x \in \mathbb{R}^n$. Розглянемо пряму $\Pi_x = \{[x,\mu]|\mu \in \mathbb{R}\}$. Можливі наступні випадки:

1) $\Pi_x \cap S = \emptyset$; 2) $\Pi_x \subset S$; 3) $\Pi_x \cap S \neq \emptyset$, $\Pi_x \not\subset S$.

В третьому випадку існує таке число λ, що $[x,\mu] \in S$ при $\mu > \lambda$ і $[x,\mu] \notin S$ при $\mu < \lambda$. Визначимо функцію $\varphi : \mathbb{R}^n \to \bar{\mathbb{R}}$, поклавши $\varphi(x) = +\infty$ в першому випадку, $\varphi(x) = -\infty$ — в другому і $\varphi(x) = \lambda$ — в третьому. В усіх трьох випадках можна скористатись формулою

$$\varphi(x) = \inf\{\mu|[x,\mu] \in S\}. \tag{4.3.1}$$

Вважатимемо для зручності, що inf порожньої множини співпадає з $+\infty$.

Зрозуміло, що надграфік $\mathrm{epi}\, \varphi$ так визначеної функції $\varphi(x)$ може відрізнятися від S лише точками $(x,\varphi(x))$, що лежать на графіку $\mathrm{gr}\, \varphi$. (Ці точки можуть і не належати S.) Про функцію $\varphi(x)$, побудовану за формулою (4.3.1), будемо говорити, що вона породжена стійкою вгору множиною.

Множину $S \subset \mathbb{R}^n \times \mathbb{R}$ назвемо *стійкою вниз*, якщо з кожною своєю точкою $[x,\mu]$ вона містить і точки $[x,\mu']$ при $\mu' \leq \mu$. Підграфік $\mathrm{hyp}\, f$

функції f стійкий вниз. Якщо S — стійка вниз множина, то про функцію

$$\varphi(x) = \sup\{\mu | [x,\mu] \in S\}$$

будемо казати, що вона породжена множиною S.

4.3.1. Конус допустимих напрямків

Нехай $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ — деяка функція і $x \in \operatorname{dom} f$. Величина

$$f_D^{\uparrow}(x,g) = \overline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha} \left(f(x + \alpha g) - f(x) \right) \tag{4.3.2}$$

називається *верхньою похідною Діні* функції f в точці x за напрямком g. Границя в (4.3.2) існує завжди, але не обов'язково скінченна.

Теорема 4.3.1. *Нехай $x \in \operatorname{dom} f$, $y = [x, f(x)]$ — відповідна точка графіка $\operatorname{gr} f$ функції f. Тоді конус допустимих напрямків $\gamma(y, \operatorname{epi} f)$ до надграфіка $\operatorname{epi} f$ стійкий вгору і породжує функцію $g \mapsto f_D^{\uparrow}(x,g)$.*

Доведення. Включення $[g,\mu] \in \gamma(y, \operatorname{epi} f)$ має місце тоді і тільки тоді, коли при деякому $\alpha_0 > 0$ виконується співвідношення $[x, f(x)] + \alpha[g,\mu] \in \operatorname{epi} f \quad \forall \alpha \in (0, \alpha_0)$, або, що те саме,

$$f(x + \alpha g) \le f(x) + \alpha \mu \quad \forall \alpha \in (0, \alpha_0).$$

Останню нерівність можна переписати так:

$$\sup_{0 < \alpha < \alpha_0} \frac{1}{\alpha} (f(x + \alpha g) - f(x)) \le \mu. \tag{4.3.3}$$

Співвідношення (4.3.3) показує, що конус $\gamma(y, \operatorname{epi} f)$ стійкий вгору, тому має сенс породжена ним функція

$$\varphi(g) = \inf\{\mu | \, [g,\mu] \in \gamma(y, \operatorname{epi} f)\}. \tag{4.3.4}$$

Перевіримо рівність $\varphi(g) = f_D^{\uparrow}(x,g)$. Нехай $\mu > f_D^{\uparrow}(x,g)$. Тоді, як випливає з означення верхньої границі, знайдеться таке $\alpha_0 > 0$, при якому виконується нерівність (4.3.3). Тому $[g,\mu] \in \gamma(y, \operatorname{epi} f)$, тобто $\mu \ge \varphi(g)$. Таким чином, $\varphi(g) \le f_D^{\uparrow}(x,g)$. Перевіримо протилежну нерівність. Нехай $\mu > \varphi(g)$. Тоді $[g,\mu] \in \gamma(y, \operatorname{epi} f)$ і тому при деякому $\alpha_0 > 0$ виконується (4.3.3). Маємо

$$f_D^{\uparrow}(x,g) \le \sup_{0 < \alpha < \alpha_0} \frac{1}{\alpha} (f(x + \alpha g) - f(x)) \le \mu,$$

звідки і випливає твердження теореми. □

Розглянемо *нижню похідну Діні* $f_D^{\downarrow}(x,g)$ функції $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ в точці $x \in \operatorname{dom} f$. За означенням

$$f_D^{\downarrow}(x,g) = \underline{\lim}_{\alpha \to +0} \frac{1}{\alpha} (f(x + \alpha g) - f(x)).$$

Аналогічно твердженню 4.3.1 доводиться наступне твердження.

Теорема 4.3.2. *Нехай* $x \in \mathrm{dom} f$, $y = [x,f(x)]$. *Тоді конус* $\gamma(y,\mathrm{hyp} f)$ *стійкий вниз і породжує функцію* $g \mapsto f_D^{\downarrow}(x,g)$.

Границя
$$f'(x,g) = \lim_{\alpha\downarrow 0} \frac{1}{\alpha}(f(x+\alpha g) - f(x)),$$
якщо вона існує, називається *похідною функції* f *в точці* x *за напрямком* g. Цю границю також називають *похідною Діні* і позначають через $f'_D(x,g)$.

Зрозуміло, що існування похідної Діні за напрямком g рівносильне тому, що числа $f_D^{\downarrow}(x,g)$ і $f_D^{\uparrow}(x,g)$ співпадають. Геометрична інтерпретація цього факту в припущенні скінченності вказаних чисел така: на прямій $\Pi_g = \{[g,\mu] \,|\, \mu \in \mathbb{R}\}$ знайдеться точка $[g,\bar{\mu}]$ така, що визначені нею відкриті промені $\{[g,\mu] \,|\, \mu > \bar{\mu}\}$ і $\{[g,\mu] \,|\, \mu < \bar{\mu}\}$ входять в конуси $\gamma(y,\mathrm{epi}\, f)$ і $\gamma(y,\mathrm{hyp}\, f)$ відповідно. Існування похідної функції f в точці x за всіма напрямками рівносильне тому, що весь простір $\mathbb{R}^n$ може бути представлений як об'єднання трьох частин: конуса $\gamma(y,\mathrm{epi}\, f)$, конуса $\gamma(y,\mathrm{hyp}\, f)$ і поверхні, яка обмежує ці конуси. Вказана поверхня може перетинатися з кожним із конусів.

4.3.2. Конус можливих напрямків

Вияснимо, яку апроксимацію функції $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ породжує апроксимація її надграфіку і підграфіку за допомогою конуса можливих напрямків.

Верхньою та нижньою похідними Адамара функції $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ *в точці* $x \in \mathrm{dom}\, f$ *за напрямком* g називаються величини
$$f_H^{\uparrow}(x,g) = \overline{\lim_{\alpha\downarrow 0,\ g'\to g}} \frac{1}{\alpha}(f(x+\alpha g') - f(x)),$$
$$f_H^{\downarrow}(x,g) = \underline{\lim_{\alpha\downarrow 0,\ g'\to g}} \frac{1}{\alpha}(f(x+\alpha g') - f(x)).$$

Теорема 4.3.3. *Нехай* $x \in \mathrm{dom}\, f$, $y = [x,f(x)]$ — *відповідна точка графіка* $\mathrm{gr}\, f$ *функції* f. *Тоді конус можливих напрямків* $\Gamma(y,\mathrm{epi}\, f)$ *в точці* y *до надграфіка* $\mathrm{epi}\, f$ *є стійким вгору і породжує функцію* $g \mapsto f_H^{\downarrow}(x,g)$.

Доведення. Включення $[g,\mu] \in \Gamma(y,\mathrm{epi}\, f)$ справедливе тоді і тільки тоді, коли існують послідовності $\{[g_k,\mu_k]\}$, $[g_k,\mu_k] \to [g,\mu]$ та $\{\alpha_k\}$, $\alpha_k \downarrow 0$, при яких виконується включення $[x,f(x)] + \alpha_k[g_k,\mu_k] \in \mathrm{epi}\, f$. Це включення рівносильне нерівності
$$\frac{1}{\alpha_k}(f(x+\alpha_k g_k) - f(x)) \le \mu_k. \tag{4.3.5}$$

Переходячи в (4.3.5) до границі, отримаємо

$$\varliminf_{k\to+\infty} \frac{1}{\alpha_k}(f(x+\alpha_k g_k) - f(x)) \le \mu. \qquad (4.3.6)$$

Із (4.3.6) безпосередньо випливає, що конус $\Gamma(y,\mathrm{epi}\, f)$ стійкий вгору.

Нехай φ — породжена цим конусом функція. Використовуючи (4.3.6), отримаємо

$$f_H^{\downarrow}(x,g) = \varliminf_{\alpha\downarrow 0,\, g'\to g} \frac{1}{\alpha}(f(x+\alpha g') - f(x)) \le \varliminf_{k\to+\infty} \frac{1}{\alpha_k}(f(x+\alpha_k g_k) - f(x)) \le \mu.$$

Оскільки $\varphi(g) = \inf\{\mu|\, [g,\mu] \in \Gamma(y,\mathrm{epi} f)\}$, то $f_H^{\downarrow}(x,g) \le \varphi(g)$. Перевіримо протилежну нерівність. Нехай $\mu > f_H^{\downarrow}(x,g)$. Тоді знайдуться такі послідовності $\{g_k\}$ і $\{\alpha_k\}$, що $g_k \to g$, $\alpha_k \downarrow 0$, $\frac{1}{\alpha_k}(f(x+\alpha_k g_k) - f(x)) \le \mu$, тобто $[x,f(x)] + \alpha_k[g_k,\mu] \in \mathrm{epi} f$. Звідси випливає співвідношення $[g,\mu] \in \Gamma(y,\mathrm{epi} f)$, яке показує, що $\mu \ge \varphi(g)$, а також справедливість нерівності $f_H^{\downarrow}(x,g) \ge \varphi(g)$, звідки випливає твердження теореми. □

Зауваження 4.3.1. Звернемо увагу на те, що конус допустимих напрямків до надграфіку породжує верхню похідну Діні, а конус можливих напрямків до надграфіку породжує нижню похідну Адамара.

Аналогічно попередньому твердженню доводиться таке твердження.

Теорема 4.3.4. *Нехай* $x \in \mathrm{dom}\, f$, $y = [x,f(x)]$. *Тоді конус* $\Gamma(y,\mathrm{hyp}\, f)$ *стійкий вниз і породжує функцію* $g \mapsto f_H^{\uparrow}(x,g)$.

Похідною Адамара функції $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ *в точці* $x \in \mathrm{dom}\, f$ *за напрямком* g називається границя

$$\lim_{\alpha\downarrow 0,\, g'\to g} \frac{1}{\alpha}(f(x+\alpha g') - f(x)). \qquad (4.3.7)$$

Зрозуміло, що із існування похідної Адамара випливає існування похідної Діні, причому обидві вказані похідні співпадають. Існування похідної Адамара за напрямком g рівносильна наступному: перетин променів $\{[g,\mu]\,|\mu \ge \bar{\mu}\}$ і $\{[g,\mu]\,|\mu \le \underline{\mu}\}$, що входять в конуси $\Gamma(y,\mathrm{epi} f)$ і $\Gamma(y,\mathrm{hyp}\, f)$, складається лише з однієї точки, а не з відрізка, тобто $\bar{\mu} = \underline{\mu}$. Та обставина, що ці промені перетинаються, випливає із тверджень 4.3.3 і 4.3.4.

4.3.3. Конус дотичних напрямків

Розглянемо апроксимацію надграфіка і підграфіка функції за допомогою конуса дотичних напрямків. Нехай $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ — деяка функція.

Для $x \in \mathrm{dom} f$ і $g \in \mathbb{R}^n$ розглянемо наступні величини:

$$f^*(x,g) = \inf_{\psi \in \Psi_n} \overline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha}(f(x + \alpha g + \psi(\alpha)) - f(x)),$$

$$f_*(x,g) = \sup_{\psi \in \Psi_n} \underline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha}(f(x + \alpha g + \psi(\alpha)) - f(x)),$$

де Ψ_n — множина нескінченно малих першого порядку.

Теорема 4.3.5. *Нехай $x \in \mathrm{dom}\, f$ та $y = [x,f(x)]$. Тоді конус $K(y,\mathrm{epi}\, f)$ дотичних напрямів в точці y до надграфіка $\mathrm{epi}\, f$ є стійким вгору і породжує функцію $g \mapsto f^*(x,g)$.*

Доведення. Нехай $[g,\mu] \in K(y,\mathrm{epi} f)$. Тоді знайдуться функції $\psi \in \Psi_n$ і $S \in \Psi_1$, для яких виконується співвідношення

$$[x,f(x)] + \alpha[g,\mu] + [\psi(\alpha),S(\alpha)] \in \mathrm{epi}\, f. \tag{4.3.8}$$

Це співвідношення може бути переписано у вигляді

$$\frac{1}{\alpha}(f(x + \alpha g + \psi(\alpha)) - f(x)) \le \mu + S(\alpha)/\alpha. \tag{4.3.9}$$

Із (4.3.9) безпосередньо випливає, що конус $K(y,\mathrm{epi}\, f)$ стійкий вгору. Крім того, із (4.3.9) випливає нерівність

$$f^*(x,g) = \inf_{\psi \in \Psi_n} \overline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha}(f(x + \alpha g + \psi(\alpha)) - f(x)) \le \mu,$$

яка показує, що $f^*(x,g) \le \varphi(g)$. З іншого боку, якщо $\mu > f^*(x,g)$, то знайдеться така функція $\psi \in \Psi_n$, що

$$\overline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha}(f(x + \alpha g + \psi(\alpha)) - f(x)) < \mu. \tag{4.3.10}$$

Нехай $S(\alpha) = 0$. Із нерівності (4.3.10) при достатньо малих α випливає нерівність (4.3.9) звідки, в свою чергу, випливає включення (4.3.8). Справедливість цього включення показує, що $[g,\mu] \in K(y,\mathrm{epi}\, f)$, тобто $\mu \ge \varphi(g)$. □

Теорема 4.3.6. *Нехай $x \in \mathrm{dom}\, f$ та $y = [x,f(x)]$. Тоді конус $K(y,\mathrm{hyp}\, f)$ стійкий вниз і породжує функцію $g \mapsto f_*(x,g)$.*

Доведення проводиться за допомогою тих же міркувань, що й доведення попереднього твердження.

4.3.4. Апроксимуючі конуси та функція відстані

Верхні і нижні похідні можна визначити за допомогою конусів, які апроксимують множини. Інший підхід: спочатку визначити похідні, а потім за їх допомогою визначити апроксимуючі конуси. При цьому використовується функція $\rho_S(x)$, що показує відстань до множини S. За означенням

$$\rho_S(x) = \inf_{z \in S} \|z - x\|.$$

Зрозуміло, що $\rho_S(x) = 0$ тоді і тільки тоді, коли $x \in \operatorname{cl} S$. Ця обставина дозволяє описати конус допустимих напрямків $\gamma(x,S)$ в термінах функції $\rho_S(x)$. А саме: $g \in \gamma(x,S)$ тоді і тільки тоді, коли $\rho_S(x + \alpha g) = 0$ при всіх досить малих $\alpha > 0$.

Конуси дотичних і можливих напрямків можна описати за допомогою верхньої та нижньої похідної Діні функції ρ_S.

Нехай $x \in \operatorname{cl} S$, $g \in \mathbb{R}^n$ — деякий напрямок. Оскільки $\rho_S(x) = 0$, то

$$(\rho_S)^{\uparrow}_D(x,g) = \overline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha} \rho_S(x + \alpha g), \tag{4.3.11}$$

$$(\rho_S)^{\downarrow}_D(x,g) = \underline{\lim_{\alpha \downarrow 0}} \frac{1}{\alpha} \rho_S(x + \alpha g), \tag{4.3.12}$$

$$(\rho_S)'(x,g) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \rho_S(x + \alpha g), \tag{4.3.13}$$

Оскільки, крім того, $\rho_S(x + \alpha g) \geq 0$, то $(\rho_S)^{\downarrow}_D(x,g) \geq 0$ при всіх g. Тому із співвідношення $(\rho_S)^{\uparrow}_D(x,g) = 0$ випливає існування похідної $(\rho_S)'(x,g)$ і рівність її нулю.

Теорема 4.3.7. *Нехай* $x \in \operatorname{cl} S$. *Тоді*

$$K(x,S) = \{g \mid (\rho_S)^{\uparrow}_D(x,g) = 0\} = \{g \mid (\rho_S)'(x,g) = 0\}, \tag{4.3.14}$$

$$\Gamma(x,S) = \{g \mid (\rho_S)^{\downarrow}_D(x,g) = 0\}. \tag{4.3.15}$$

Доведення. Встановимо справедливість рівності (4.3.14). Враховуючи означення функції ρ_S, для довільного $\alpha > 0$ можна знайти такий елемент $v_\alpha \in S$, що

$$\|x + \alpha g - v_\alpha\| \leq \rho_S(x + \alpha g) + \alpha^2. \tag{4.3.16}$$

Покладемо $\psi(\alpha) = x + \alpha g - v_\alpha$. Якщо $(\rho_S)'(x,g) = 0$ то, як випливає з (4.3.13) і (4.3.16), $\psi(\alpha)$ є нескінченно малою першого порядку. Оскільки $x + \alpha g + \psi(\alpha) = v_\alpha \in S$, то $g \in K(x,S)$. З іншого боку, якщо $g \in K(x,S)$, то знайдеться така функція $\psi(\alpha)$, що $x + \alpha g + \psi(\alpha) \in S$ і $\alpha^{-1}\psi(\alpha) \to 0$. Маємо

$$\frac{1}{\alpha}\rho_S(x + \alpha g) \leq \frac{1}{\alpha}\|(x + \alpha g) - (x + \alpha g + \psi(\alpha))\| = \alpha^{-1}\|\psi(\alpha)\| \to 0.$$

Тим самим нерівність (4.3.14) доведена.

Перейдемо до доведення формули (4.3.15). Скористаємося рівнянням (4.3.12). Нехай нижня границя в формулі (4.3.12) рівна нулю і нехай $\{\alpha_k\}$

така послідовність, що $\alpha_k \downarrow 0$ і $\frac{1}{\alpha_k}\rho_S(x + \alpha_k g) \to 0$. Знайдемо елементи $v_k \in S$, які мають властивість $\|x + \alpha_k g - v_k\| < \alpha_k/k$, і покладемо $S_k = x + \alpha_k g - v_k$, $g - \frac{S_k}{\alpha_k} = g_k$. Тоді $x + \alpha_k g_k = v_k \in S$, $g_k - g \to 0$. Звідки випливає включення $g \in \Gamma(x,S)$.

Припустимо тепер, що g — можливий напрямок в точці x відносно множини S. Тоді знайдуться такі послідовності $\{\alpha_k\}$, $\{g_k\}$, що $\alpha_k \downarrow 0$, $g_k \to g$, $x + \alpha_k g_k \in S$. Останнє включення можна переписати у вигляді $x + \alpha_k g + \alpha_k(g_k - g) \in S$. Маємо

$$\rho_S(x + \alpha_k g) \leq \|(x + \alpha_k g) - (x + \alpha_k g + \alpha_k(g_k - g))\| \leq \alpha_k\|g_k - g\|.$$

Таким чином $\frac{1}{\alpha_k}\rho_S(x + \alpha_k g) \to 0$ і, відповідно, $\varliminf\limits_{\alpha\downarrow 0}\rho_S(x + \alpha g) = 0$. $\square$

Зауваження 4.3.2. Використовуючи теорему 4.3.6, покажемо, що конус дотичних напрямків $K(x,S)$ до опуклої множини S в точці $x \in \operatorname{cl} S$ в точності співпадає з замиканням $\operatorname{cl} K_x(S)$ опорного конуса $K_x(S)$. Для цього скористаємося наступними твердженнями:

1) якщо S опукла, то функція ρ_S опукла;

2) опукла функція має похідну за напрямками.

Із цих тверджень випливає, що $\{g|\ \rho_S'(x,g) = 0\} = \{g|\ \rho_S^{\downarrow}(x,g) = 0\}$. Отже, в силу теореми 4.3.7, $K(x,S) = \Gamma(x,S)$. А також (див. теорему 4.2.2) $\Gamma(x,S) = \operatorname{cl} K_x(S)$.

4.4. АПРОКСИМАЦІЯ МНОЖИН, ЩО ЗАДАНІ ЗА ДОПОМОГОЮ НЕРІВНОСТЕЙ І РІВНЯНЬ. УМОВИ РЕГУЛЯРНОСТІ

Апроксимуючі конуси часто доводиться знаходити для множин, які задаються як сукупність розв'язків нерівності $h(x) \leq 0$ або рівняння $h(x) = 0$, де h — деяка неперервна функція. В теорії екстремальних задач такі обмеження іноді називають "обмеження типу нерівності" та "обмеження типу рівності" відповідно. В деяких випадках описати конуси, що апроксимують вказані множини, вдається за допомогою похідної за напрямком. Розглянемо спочатку множини, які можуть бути описаними за допомогою "обмежень типу нерівності".

Нехай $S = \{x \in X|\ h(x) \leq 0\}$, де h — неперервна функція, що визначена на відкритій множині X в $\mathbb{R}^n$. Якщо $h(x) < 0$, то $x \in \operatorname{int} S$ і $\gamma(x,S) = K(x,S) = \Gamma(x,S) = \mathbb{R}^n$. Тому нас буде цікавити лише точка x, для якої $h(x) = 0$. Припустимо, що h диференційовна за Адамаром в точці x і розглянемо множини

$$\gamma_1(x) = \{g|\ h'(x,g) < 0\}, \quad \Gamma_1(x) = \{g|\ h'(x,g) \leq 0\}. \tag{4.4.1}$$

Оскільки похідна $h'(x,g)$ є неперервною функцією напрямку, то множина $\gamma_1(x)$ відкрита, а множина $\Gamma_1(x)$ замкнута. Оскільки функція $h'(x,g)$ додатньо однорідна за g, то ці множини є конусами. Опишемо зв'язки конусів $\gamma_1(x)$ і $\Gamma_1(x)$ з конусом допустимих напрямків $\gamma(x,S)$ і конусом Булігана $\Gamma(x,S)$.

Теорема 4.4.1. *Справедливі включення*
$$\gamma_1(x) \subset \gamma(x,S), \quad \Gamma(x,S) \subset \Gamma_1(x).$$

Доведення. Нехай $g \in \gamma_1(x)$. Тоді
$$h(x+\alpha g) = h(x) + \alpha h'(x,g) + o(\alpha) \leq 0$$
при достатньо малих $\alpha \geq 0$, і тому $x + \alpha g \in S$ при таких α. Це означає, що $g \in \gamma(x,S)$. Розглянемо тепер елемент g конуса $\Gamma(x,S)$. Знайдуться такі послідовності $g_k \to g$ і $\alpha_k \downarrow 0$, що $h(x+\alpha_k g_k) \leq 0$. Оскільки
$$h(x+\alpha_k g_k) = h(x) + \alpha_k h'(x,g_k) + o(\alpha_k g_k),$$
де $h(x) = 0$, $o(v)/\|v\| \to 0$, похідна $h'(x,g)$ неперервна за напрямком, то $h'(x,g) \leq 0$, тобто $g \in \Gamma_1(x)$. $\square$

Говорять, що в точці $x \in S$, для якої $h(x) = 0$, виконана *умова регулярності*, якщо
$$\mathrm{cl}\gamma_1(x) = \Gamma_1(x). \tag{4.4.2}$$

Теорема 4.4.2. *Якщо в точці x виконана умова регулярності, то*
$$\Gamma(x,S) = \mathrm{cl}\gamma(x,S) = \Gamma_1(x).$$

Доведення випливає з теореми 4.4.1.

Ситуація, в якій умова регулярності не виконана, виникає, наприклад, у тому випадку, коли функція h досягає в точці x глобального максимуму на всій множині X, тобто з нерівності $h(y) \leq 0$ випливає рівність $h(y) = 0$ (нагадаємо, що $h(x) = 0$). Дійсно, в цьому випадку $\gamma_1(x)$ порожня, а множина $\Gamma_1(x)$ завжди непорожня, вона містить нуль.

Опишемо один простий випадок, в якому можна гарантувати виконання умови регулярності.

Теорема 4.4.3. *Нехай похідна $h'(x,g)$ сублінійна за g і при деякому v виконується нерівність $h'(x,v) < 0$. Тоді справедлива умова регулярності* (4.4.2).

Доведення. Нехай $g \in \Gamma_1(x)$ і $h'(x,g) = 0$. Маємо $h'(x,g+\alpha v) \leq h'(x,g) + \alpha h'(x,v) < 0$. Тоді $g + \alpha v \in \gamma_1(x)$ при всіх $\alpha > 0$. В той же час $g + \alpha v \to g$ при $\alpha \downarrow 0$. $\square$

Зауважимо, що теореми 4.4.1—4.4.3 справедливі і при відсутності неперервності h.

Розглянемо тепер функцію $G(x,y)$, яка визначена на $X \times Y$, де X — відкрита, а Y — компактна множини. Покладемо

$$S = \{x|\, G(x,y) \leq 0, y \in Y\}. \tag{4.4.3}$$

Зрозуміло, що $S = \bigcap\limits_{y \in Y} S_y$, де $S_y = \{x|\, G(x,y) \leq 0\}$. Тобто S являє собою перетин "обмежень типу нерівностей". Множину S можна представити як одне "обмеження типу нерівності", а саме $S = \{x|\, h(x) \leq 0\}$, де

$$h(x) = \max_{y \in Y} G(x,y). \tag{4.4.4}$$

Даний прийом дозволяє звести визначення множини за допомогою системи нерівностей вигляду $G(x,y) \leq 0, y \in Y$, до її визначення за допомогою однієї нерівності $h(x) \leq 0$. При цьому, однак, деякі властивості функції G, яка задає вказану систему, можуть бути втрачені. Наприклад, якщо G неперервно диференційовна за x, то функція h вже цієї властивості може не мати, можна говорити лише про диференційовність h за напрямком.

Теорема 4.4.4. *Нехай множина S визначена співвідношенням* (4.4.3)*, де $G(x,y)$ — неперервна разом зі своєю частинною похідною $G'_x(x,y)$ функція, причому функція $G'_x(x,y)$ обмежена в області визначення $X \times Y$ функції $G(x,y)$* (нагадаємо, що X — відкрита, а Y — компактна множини)*. Нехай також $x \in S$ і $G(x,y) = 0$ при всіх $y \in R(x)$, тобто справедлива рівність*

$$R(x) = \{y \in Y|\, G(x,y) \geq G(x,y')\; \forall y' \in Y\}.$$

Тоді, якщо існує така точка v, що $\langle G'_x(x,y),v\rangle < 0$ при всіх $y \in R(x)$, то справедливі рівності

$$\operatorname{cl}\gamma(x,S) = \Gamma(x,S) = \{g|\, \langle G'_x(x,y),g\rangle \leq 0 \quad \forall y \in R(x)\}\,.$$

Доведення. Нехай функція h задана формулою (4.4.4). Наслідок 4.1.3 показує, що

$$h'(x,g) = \max_{y \in R(x)} \langle G'_x(x,y),g\rangle. \tag{4.4.5}$$

З обмеженості частинної похідної $G'_x(x,y)$ відразу випливає, що функція h задовольняє умові Ліпшиця і тому похідна $h'(x,g)$ є похідною Адамара. Формула (4.4.5) показує, що ця похідна сублінійна за g, і крім того виконується нерівність $h'(x,g) < 0$. Тому можна застосувати теорему 4.4.2. Використовуючи формулу (4.4.5), переконуємлся в справедливості даного твердження. □

Покажемо, що за умови регулярності конус Булігана здійснює рівномірну апроксимацію першого порядку множини $S = \{x|\, h(x) \leq 0\}$.

Теорема 4.4.5. *Нехай точка $x \in S$ така, що $h(x) = 0$, функція h диференційовна за Адамаром у цій точці і виконані умови регулярності* (4.4.2). *Тоді конус $\Gamma(x,S)$ є рівномірною апроксимацією першого порядку множини S поблизу точки x.*

Доведення. Для $\delta > 0$ покладемо

$$S_\delta(x) = S \cap B_\delta(x), \quad \Gamma_\delta(x) = (x + \Gamma(x,S)) \cap B_\delta(x).$$

Потрібно довести, що

$$\rho(S_\delta(x),\Gamma_\delta(x)) =$$

$$= \max\left\{ \max_{z\in\Gamma_\delta(x)} \min_{y\in S_\delta(x)} \|z-y\|, \max_{y\in S_\delta(x)} \min_{z\in\Gamma_\delta(x)} \|z-y\| \right\} = o(\delta). \tag{4.4.6}$$

Припустимо, що співвідношення (4.4.6) не виконується. Тоді можливі два випадки:

1) існують такі числа $a > 0$ і послідовності $\{\delta_k\}$, $\{z_k\}$, що $\delta_k \downarrow 0$, $z_k \in \Gamma_{\delta_k}(x)$,

$$\min_{y\in S_{\delta_k}(x)} \|y - z_k\| \geq a\delta_k; \tag{4.4.7}$$

2) існують такі числа $a > 0$ і послідовності $\{\delta_k\}$, $\{y_k\}$, що $\delta_k \downarrow 0$, $y_k \in S_{\delta_k}(x)$,

$$\min_{z\in\Gamma_{\delta_k}(x)} \|z - y_k\| \geq a\delta_k. \tag{4.4.8}$$

Розглянемо перший випадок. Оскільки $z_k \in \Gamma_{\delta_k}(x)$, то z_k можна представити у вигляді $z_k = x + \delta_k v_k$, де $\|v_k\| \leq 1$, $v_k \in \Gamma(x,S)$. Вважатимемо, що $v_k \to v$. Оскільки конус $\Gamma(x,S)$ замкнутий, то $v \in \Gamma(x,S)$. Тому (див. теорему 4.4.2) $h'(x,v) \leq 0$. Використовуючи умову регулярності (4.4.2), знайдемо такий елемент v', що $\|v - v'\| < a/2$ і $h'(x,v') < 0$. Маємо $h(x + \delta_k v') = h(x) + \delta_k h'(x,v') + o(\delta_k)$. Тому при достатньо великих k виконується нерівність $h(x + \delta_k v') \leq 0$, яка показує, що $x + \delta_k v' \in S$. Оскільки, з іншого боку, $x + \delta_k v_k \in B_{\delta_k}(x)$, то $x + \delta_k v' \in S_{\delta_k}$. Застосовуючи (4.4.7), одержимо

$$\|x + \delta_k v' - z_k\| \geq a\delta_k. \tag{4.4.9}$$

Оскільки $z_k = x + \delta_k v_k$, то нерівність (4.4.9) може бути переписана так:

$$\|v' - v_k\| \geq a. \tag{4.4.10}$$

Із (4.4.10) випливає, що $\|v' - v\| \geq a$, а це суперечить вибору елемента v'. Отже випадок 1) неможливий.

Перейдемо до другого випадку. Оскільки $y_k \in S_{\delta_k}(x)$, то $\|y_k - x\| \leq \delta_k$. Тому знайдуться такі вектори $u_k \in B$, що $y_k = x + \delta_k u_k$. Але $y_k \in S$ і тому $h(x + \delta_k u_k) \leq 0$. Функція h диференційовна в точці x за Адамаром, тому

$$h(x + \delta_k u_k) = h(x) + \delta_k h'(x,u_k) + o(\delta_k u_k), \tag{4.4.11}$$

Оскільки $h(x) = 0$, то із (4.4.11) випливає, що

$$h'(x,u_k) + \frac{o(\delta_k u_k)}{\|\delta_k u_k\|}\|u_k\| \le 0. \tag{4.4.12}$$

Вважатимемо, що існує границя $\lim u_k = u$. Переходячи в (4.4.12) до границі, отримаємо, що $h'(x,u) \le 0$. Ця нерівність показує, що $u \in \Gamma(x,S)$. Оскільки $\|u\| \le 1$, то $x + \delta_k u \in \Gamma_{\delta_k}$. Тому, як випливає зі (4.4.8),

$$\|x + \delta_k u - y_k\| = \|x + \delta_k u - (x + \delta_k u_k)\| \ge \delta_k a. \tag{4.4.13}$$

Нерівність (4.4.13) показує, що $\|u - u_k\| \ge a$, а це суперечить визначенню u. Таким чином випадок 2) також неможливий. □

Наслідок 4.4.1. *В умовах теореми конус дотичних напрямків* $K(x,S)$ *співпадає із конусом Булігана* $\Gamma(x,S)$. *При цьому для будь-якого* $g \in \Gamma(x,S)$ *можна знайти таку функцію* $\psi_g(\alpha)$, *що* $x + \alpha g + \psi_g(\alpha) \in S$ *і* $\alpha^{-1}\psi_g(\alpha) \to 0$ *при* $\alpha \downarrow 0$ *рівномірно по* g.

Перейдемо до “обмежень типу рівності”. Нехай множина S має вигляд

$$S = \{y \in X \mid h(y) = 0\}.$$

Тут X — відкрита множина в $\mathbb{R}^n$, h — неперервна на X диференційовна за Адамаром в точці $x \in X$ функція. Розглянемо, як і раніше, конуси

$$\gamma_1(x) = \{g \mid h'(x,g) < 0\}, \quad \Gamma_1(x) = \{g \mid h'(x,g) \le 0\},$$

а також конуси

$$\gamma_2(x) = \{g \mid h'(x,g) > 0\}, \quad \Gamma_2(x) = \{g \mid h'(x,g) \ge 0\}.$$

Встановимо зв’язок конуса Булігана $\Gamma(x,S)$ з конусами $\mathrm{cl}\,\gamma_i(x)$ і $\Gamma_i(x)$ $(i = 1,2)$.

Теорема 4.4.6. *Справедливі наступні включення:*

$$(\mathrm{cl}\,\gamma_1(x)) \cap (\mathrm{cl}\,\gamma_2(x)) \subset \Gamma(x,S) \subset \Gamma_1(x) \cap \Gamma_2(x). \tag{4.4.14}$$

Доведення. Друге включення із (4.4.14) очевидне, тому перевіримо лише перше. Нехай $g \in \mathrm{cl}\,\gamma_i(x)$ $(i = 1,2)$. Тоді знайдуться такі послідовності $g_k^1 \to g$ і $g_k^2 \to g$, що

$$h'(x,g_k^1) < 0, \quad h'(x,g_k^2) > 0. \tag{4.4.15}$$

Нерівності (4.4.15) показують, що для кожного k при достатньо малих $\alpha > 0$ виконуються співвідношення

$$h(x + \alpha g_k^1) < 0, \quad h(x + \alpha g_k^2) > 0.$$

Нехай α_k — одне із цих чисел. Розглянемо відрізок $[x + \alpha_k g_k^1, x + \alpha_k g_k^2]$. Оскільки функція h приймає на його кінцях значення різних знаків, то знайдеться така точка y на цьому відрізку, що $h(y) = 0$. Зрозуміло, що y можна представити у вигляді $y = x + \alpha_k g_k$, де g_k — опукла комбінація векторів g_k^1 і g_k^2. Оскільки $g_k \to g$, $\alpha \downarrow 0$ то $g \in \Gamma(x,S)$. □

Кажуть, що в точці $x \in S$ виконана *умова регулярності* (*відносно обмеження типу рівності*), якщо

$$\operatorname{cl}\gamma_1(x) = \Gamma_1(x), \quad \operatorname{cl}\gamma_2(x) = \Gamma_2(x). \tag{4.4.16}$$

Теорема 4.4.7. *Якщо в точці x виконана умова регулярності* (4.4.16)*, то*

$$\Gamma(x,S) = \{g|\, h'(x,g) = 0\}.$$

Доведення випливає з теореми 4.4.6.

Зауваження 4.4.1. Множину

$$S = \{x|\, h_i(x) \le 0, i = 1,\ldots,k\}$$

можна записати у вигляді одного обмеження – нерівності $S = \{x|\, h(x) \le 0\}$, де $h(x) = \max_i h_i(x)$. Якщо для функції h_i виконані умови регулярності (4.4.2), то при природних припущеннях ці умови виконуються і для функції h.

Якщо множина S має вигляд

$$S = \{x|\, h_i(x) = 0,\ i = 1,\ldots,k\},$$

то її можна задати одним обмеженням типу рівності $S = \{x|\, h(x) = 0\}$, де $h(x) = \max_i |h_i(x)|$. При цьому, однак, для функції h умова регулярності (4.4.16) не виконується. Тому випадок декількох обмежень типу рівностей потрібно розглядати ретельно. При його дослідженні використовують, як правило, ту чи іншу теорему про неявну функцію.

Зауваження 4.4.2. Умову регулярності можна записати і у випадку ліпшицевих, але не диференційовних за напрямками функцій. Будемо розглядати лише обмеженнями типу рівності $S = \{x|\, h(x) = 0\}$. Покладемо

$$\gamma_{1D}(x) = \{g|\, h_D^{\downarrow}(x,g) < 0\}, \quad \Gamma_{1D}(x) = \{g|\, h_D^{\downarrow}(x,g) \le 0\},$$

$$\gamma_{2D}(x) = \{g|\, h_D^{\uparrow}(x,g) > 0\}, \quad \Gamma_{2D}(x) = \{g|\, h_D^{\uparrow}(x,g) \ge 0\}.$$

Будемо казати, що в точці $x \in S$ виконана умова регулярності, якщо

$$\operatorname{cl}\gamma_{1D}(x) = \Gamma_{1D}(x), \quad \operatorname{cl}\gamma_{2D}(x) = \Gamma_{2D}(x).$$

Можна показати, що при виконанні умови регулярності справедливі рівності

$$\Gamma(x,S) = \Gamma_{1D}(x) \cap \Gamma_{2D}(x) = \operatorname{cl}\gamma_{1D}(x) \cap \operatorname{cl}\gamma_{2D}(x).$$

4.5. УМОВИ ЭКСТРЕМУМУ ДИФЕРЕНЦІЙОВНИХ ЗА НАПРЯМКАМИ ФУНКЦІЙ

Нехай функція f задана на відкритій множині X$\subset \mathbb{R}^n$. Розглянемо задачу мінімізації функції f на множині $S \subset$X. Наша мета — встановити необхідні (а по можливості і достатні) умови мінімуму функції f в точці x.

Означення 4.5.1. Точка $x^* \in S$ називається *точкою локального мінімуму* функції f на множині S, якщо існує таке δ>0, що виконується нерівність

$$f(x) \geq f(x^*) \quad \forall x \in S \cap B_\delta(x^*), \tag{4.5.1}$$

де $B_\delta\left(x^*\right) = \{x \in \mathbb{R}^n \colon\ \|x - x^*\| \leq \delta\}$.

Означення 4.5.2. Якщо нерівність (4.5.1) виконується при $\delta = \infty$, то точка x^* називається *точкою глобального мінімуму* функції f на множині S. Тоді нерівність (4.5.1) має вигляд

$$f\left(x\right) \geq f\left(x^*\right) \quad \forall x \in S. \tag{4.5.2}$$

Означення 4.5.3. Точка $x^* \in S$ називається точкою *строгого локального мінімуму функції* f на множині S, якщо існує таке δ >0, що

$$f\left(x\right) > f\left(x^*\right) \quad \forall x \in S \cap B_\delta\left(x^*\right), x \neq x^*. \tag{4.5.3}$$

4.5.1. Умови экстремуму функції на $\mathbb{R}^n$

Розглянемо задачу мінімізації функції f на всьому просторі $\mathbb{R}^n$ (тобто $S = \mathbb{R}^n$).

Лема 4.5.1. *Нехай функція f диференційовна за Діні в точці $x^* \in S$. Для того, щоб точка x^* була точкою локального мінімуму функції f на $\mathbb{R}^n$ необхідно, щоб виконувалася умова*

$$f'_D\left(x^*,g\right) \geq 0 \ \ \forall g \in \mathbb{R}^n. \tag{4.5.4}$$

Якщо функція f задовольняє умову Ліпшиця в околі точки x^* і

$$f'_D\left(x^*,g\right) > 0 \ \ \forall g \in \mathbb{R}^n,\, g \neq 0, \tag{4.5.5}$$

то x^* — точка строгого локального мінімуму функції f на $\mathbb{R}^n$ (тобто умова (4.5.5)— достатня умова локального мінімуму для ліпшицевої функції).

Доведення. Необхідність. Оскільки функція f диференційовна в точці x^* за напрямками, то

$$f\left(x^* + \alpha g\right) = f\left(x^*\right) + \alpha f'\left(x^*,g\right) + o\left(\alpha\right) \quad \forall \alpha \geq 0,\ \forall g \in \mathbb{R}^n,$$

де

$$o(\alpha) = o(\alpha, x^*, g), \quad \frac{o(\alpha, x^*, g)}{\alpha} \underset{\alpha \downarrow 0}{\longrightarrow} 0 \ \forall g \in \mathbb{R}^n.$$

Звідси відразу випливає умова (4.5.4).

Достатність. Нехай функція f задовольняє умову Ліпшиця в околі точки x^* і виконана умова (4.5.5). Тоді x^* — точка строгого локального мінімуму f на $\mathbb{R}^n$. Припустимо протилежне. Тоді для кожного $\delta_k > 0$ знайдеться таке g_k, що

$$f(x^* + g_k) < f(x^*), \ \ \|g_k\| \le \delta_k. \tag{4.5.6}$$

Можемо вважати, що $\bar{g}_k = \frac{g_k}{\|g_k\|} \to g$. Покладемо $\alpha_k = \|g_k\|$. Тоді

$$\alpha_k \downarrow 0, \quad k \to \infty, \tag{4.5.7}$$

$$f(x^* + g_k) - f(x^*) = f(x^* + \alpha_k \bar{g}_k) - f(x^*) =$$

$$= f(x^* + \alpha_k g) - f(x^*) + f(x^* + \alpha_k \bar{g}_k) - f(x^* + \alpha_k g). \tag{4.5.8}$$

Оскільки функція f локально ліпшицева, то

$$|f(x^* + \alpha_k \bar{g}_k) - f(x^* + \alpha_k g)| \le L\,\alpha_k \ \|\bar{g}_k - g\|, \tag{4.5.9}$$

де $L < \infty$ — константа Ліпшиця. З (4.5.6), (4.5.8), (4.5.9) одержуємо

$$\frac{1}{\alpha_k}\left(f(x^* + \alpha_k g) - f(x^*)\right) \le L\ \|\bar{g}_k - g\| \tag{4.5.10}$$

Переходячи в (4.5.10) до границі при $k \to \infty$ і враховуючи (4.5.7), маємо $f'(x^*, g) \le 0$, що суперечить умові (4.5.5). □

Лема 4.5.2. *Нехай функція f диференційовна за Діні в точці $x^{**} \in S$. Для того щоб точка x^{**} була точкою локального максимуму функції f на $\mathbb{R}^n$ необхідно, щоб виконувалась умова*

$$f'_D(x^{**}, g) \le 0 \ \ \forall g \in \mathbb{R}^n. \tag{4.5.11}$$

Якщо функція f задовольняє умову Ліпшиця в околі точки x^{**} і

$$f'_D(x^{**}, g) < 0 \ \ \forall g \in \mathbb{R}^n,\ g \ne 0, \tag{4.5.12}$$

то x^{**} — точка строгого локального максимуму функції f на $\mathbb{R}^n$ (тобто умова (4.5.12)— достатня умова локального максимуму для ліпшицевої функції).

Означення 4.5.4. Точка x^*, що задовольняє умові (4.5.4), називається *Діні-inf–стаціонарною точкою* функції f на $\mathbb{R}^n$.

Означення 4.5.5. Точка x^{**}, що задовольняє умові (4.5.11), називається *Діні-sup–стаціонарною точкою* функції f на $\mathbb{R}^n$.

Аналогічно доводяться такі твердження.

Лема 4.5.3. *Нехай функція f диференційовна за Адамаром в точці $x^* \in S$. Для того щоб точка x^* була точкою локального мінімуму функції f на $\mathbb{R}^n$ необхідно, щоб виконувалася умова*

$$f'_H\left(x^*,g\right) \geq 0 \ \ \forall g \in \mathbb{R}^n. \tag{4.5.13}$$

Якщо функція f задовольняє умову

$$f'_H\left(x^*,g\right) > 0 \ \ \forall g \in \mathbb{R}^n,\, g \neq 0, \tag{4.5.14}$$

то x^* — точка строгого локального мінімуму функції f на $\mathbb{R}^n$ (тобто умова (4.5.14)— достатня умова локального мінімуму).

Лема 4.5.4. *Нехай функція f диференційовна за Адамаром в точці $x^{**} \in S$. Для того щоб точка x^{**} була точкою локального максимуму функції f на $\mathbb{R}^n$ необхідно, щоб виконувалася умова*

$$f'_H\left(x^{**},g\right) \leq 0 \ \ \forall g \in \mathbb{R}^n. \tag{4.5.15}$$

Якщо функція f задовольняє умову

$$f'_H\left(x^{**},g\right) < 0 \ \ \forall g \in \mathbb{R}^n,\, g \neq 0, \tag{4.5.16}$$

то x^{**} — точка строгого локального максимуму функції f на $\mathbb{R}^n$ (тобто умова (4.5.16)— достатня умова локального максимуму).

Означення 4.5.6. Точка x^*, що задовольняє умові (4.5.13), називається *Адамар-inf–стаціонарною точкою* функції f на $\mathbb{R}^n$.

Означення 4.5.7. Точка x^{**}, що задовольняє умові (4.5.15), називається *Адамар-sup–стаціонарною точкою* функції f на $\mathbb{R}^n$.

Умови (4.5.4) — (4.5.16) є необхідними умовами першого порядку. Використовуючи апроксимації другого порядку, можна одержати більш тонкі необхідні умови.

Зауваження 4.5.1. Відзначимо, що в гладкому випадку (тобто в тому випадку, коли f – диференційовна функція) $f'(x,g) = \langle f'(x),g\rangle$, де $f'(x)$ – градієнт функції f в точці x. Тоді умови (4.5.4), (4.5.11), (4.5.13), (4.5.15) мають відповідно вигляд $f'(x^*) = 0$ та $f'(x^{**}) = 0$ і тому випадки (4.5.5), (4.5.12) та (4.5.14), (4.5.16) неможливі, тобто умови (4.5.5), (4.5.12) та (4.5.14), (4.5.16) мають місце лише для істотно негладких функцій.

Зауваження 4.5.2. Леми 4.5.1, 4.5.2, 4.5.3, 4.5.4 справедливі і для задачі мінімізації та максимізації на множині S в тих випадках коли точка $x^* \in \operatorname{int} S$ і точка $x^{**} \in \operatorname{int} S$.

Означення 4.5.8. Нехай функція f диференційовна за Діні в точці $x_0 \in \mathbb{R}^n$ і в точці x_0 не виконана необхідна умова мінімуму (4.5.4). Якщо $\|g_0\| = 1$ та

$$f'_D\left(x_0,g_0\right) = \inf_{\|g\|=1} f'_D\left(x_0,g\right), \tag{4.5.17}$$

то напрямок g_0 називається *Діні–напрямком найшвидшого спуску* (н. н. с.) функції f в точці x_0. Якщо $||g_0|| = 1$ та

$$f'_D(x_0,g_0) = \sup_{\|g\|=1} f'_D(x_0,g), \tag{4.5.18}$$

то напрямок g_0 називається *Діні–напрямком найшвидшого підйому* (н. н. п.) функції f в точці x_0.

Означення 4.5.9. Нехай функція f диференційовна за Адамаром в точці $x_0 \in \mathbb{R}^n$ і в точці x_0 не виконана необхідна умова мінімуму (4.5.13). Якщо $\|g_0\| = 1$ та

$$f'_H(x_0,g_0) = \inf_{\|g\|=1} f'_H(x_0,g), \tag{4.5.19}$$

то напрямок g_0 називається *Адамар–напрямком найшвидшого спуску* (н. н. с.) функції f в точці x_0. Якщо $||g_0|| = 1$ і

$$f'_H(x_0,g_0) = \sup_{\|g\|=1} f'_H(x_0,g), \tag{4.5.20}$$

то напрямок g_0 називається *Адамар–напрямком найшвидшого підйому* (н. н. п.) функції f в точці x_0.

Якщо функція f локально ліпшицева, то функція $f'(x_0,g)$ неперервна за g, тому інфімум в (4.5.17), (4.5.19) і супремум в (4.5.18), (4.5.20) досягаються, тобто н. н. с. та н. н. п. існують.

Зауваження 4.5.3. Висловлювання “напрямок спуску функції f у точці x_0” є загальноприйнятим, хоча правильніше було б казати: “... з точки x_0”.

4.5.2. Умови экстремуму функції на множині $S \subset \mathbb{R}^n$

Перейдемо до задачі “умовної” мінімізації, тобто до задачі мінімізації функції f на множині S. Припускаємо, що функція f задана на відкритій множині $X \subset \mathbb{R}^n$ та $S \subset X$.

Лема 4.5.5. *Нехай функція f локально ліпшицева в околі точки x^* і диференційовна за напрямками у точці $x^* \in S$. Для того, щоб у точці x^* функція f досягала найменшого на S значення, необхідно, щоб*

$$f'(x^*,g) \geq 0 \ \ \forall g \in \Gamma(x^*,S), \tag{4.5.21}$$

де $\Gamma(x^*,S)$ *– конус Булігана.*
Якщо виявилося, що

$$f'(x^*,g) > 0 \ \ \forall g \in \Gamma(x^*,S), \, g \neq 0_n, \tag{4.5.22}$$

то x^* є точкою строгого локального мінімуму функції f на множині S.

Доведення. Нехай x^* — точка мінімуму f на S. Припустимо, що умова (4.5.21) не виконана. Тоді знайдеться $g_0 \in \Gamma(x^*,S)$ таке, що

$$f'(x^*,g_0) = -a < 0. \tag{4.5.23}$$

За означенням $\Gamma(x^*,S)$ існує послідовність точок $\{x_k\}$ така, що $x_k \in S$, $g_k = \frac{x_k - x^*}{\|x_k - x^*\|} \to g_0$. Покладемо $\alpha_k = \|x_k - x^*\|$. Тоді

$$f(x_k) = f(x^* + (x_k - x^*)) =$$
$$= f(x^* + \alpha_k g_k) = f(x^* + \alpha_k g_0) + \varphi_k(\alpha_k), \tag{4.5.24}$$

де $\varphi_k(\alpha_k) = f(x^* + \alpha_k g_k) - f(x^* + \alpha_k g_0)$. Оскільки f – ліпшицева функція, то $|\varphi_k(\alpha_k)| \le L\alpha_k \|g_k - g_0\|$. З (4.5.23) та (4.5.24) маємо

$$f(x_k) = f(x^*) + \alpha_k f'(x^*,g_0) + o(\alpha_k) + \varphi_k(\alpha_k) \le$$
$$\le f(x^*) + \alpha_k(-a + L\|g_k - g_0\|) + o(\alpha_k)$$

При досить малих α_k будемо мати

$$f(x_k) < f(x^*) - \frac{1}{2}\alpha_k a. \tag{4.5.25}$$

Оскільки $x_k \in S$, то (4.5.25) суперечить тому, що x^* — точка мінімуму.

Достатність умови (4.5.22) встановлюється так само, як достатність умови (4.5.5) в лемі 4.5.1 (з урахуванням того, що отриманий при доведенні від супротивного напрямок належить конусу $\Gamma(x^*,S)$). □

Означення 4.5.10. Точка $x^* \in S$, в якій виконана умова (4.5.21), називається *inf–стаціонарною точкою* функції f на множині S. Якщо точка $x_0 \in S$ не є inf–стаціонарною точкою функції f на множині S, то напрямок $g_0 \in \Gamma(x_0,S)$, $\|g_0\| = 1$, такий, що

$$f'(x_0,g_0) = \inf_{\|g\|=1, g\in\Gamma(x_0,S)} f'(x_0,g),$$

називається *напрямком найшвидшого спуску* (н. н. с.) функції f на множині S в точці x_0. Якщо функція f ліпшицева, то $f'(x_0,g)$ є неперервною функцією напрямку g. Тому інфімум досягається і н. н. с. існує.

Аналогічно формулюються необхідна та достатня умови максимуму і даються означення sup-стаціонарної точки і напрямки найшвидшого підйому функції f на множині S.

Нехай тепер f – довільна ліпшицева функція (не обов'язково диференційовна за напрямками), яка задана на відкритій множині $X \subset \mathbb{R}^n$. Тоді для будь-яких $x \in X$ і $g \in \mathbb{R}^n$ існують скінченні верхня і нижня похідні Діні

$$f_D^{\uparrow}(x,g) = \overline{\lim_{\alpha\downarrow 0}} \frac{1}{\alpha}[f(x+\alpha g) - f(x)],$$

$$f_D^{\downarrow}(x,g) = \underline{\lim_{\alpha\downarrow 0}} \frac{1}{\alpha}[f(x+\alpha g) - f(x)].$$

При цьому

$$f(x+\alpha g) = f(x) + \alpha f_D^{\uparrow}(x,g) + \bar{o}(\alpha) \quad \forall \alpha > 0, \tag{4.5.26}$$

$$f(x+\alpha g) = f(x) + \alpha f_D^{\downarrow}(x,g) + \underline{o}(\alpha) \quad \forall \alpha > 0, \tag{4.5.27}$$

де

$$\bar{o}(\alpha) = \bar{o}(\alpha,g,x), \quad \overline{\lim_{\alpha\downarrow 0}} \frac{\bar{o}(\alpha)}{\alpha} = 0, \tag{4.5.28}$$

$$\underline{o}(\alpha) = \underline{o}(\alpha,g,x), \quad \underline{\lim}_{\alpha\downarrow 0} \frac{\underline{o}(\alpha)}{\alpha} = 0. \tag{4.5.29}$$

Лема 4.5.6. *Нехай у точці $x^* \in S$ множина S допускає рівномірну конічну апроксимацію першого порядку, а функція f локально ліпшицева в околі точки x^*. Для того, щоб у точці x^* функція f досягала свого найменшого на S значення, необхідно, щоб*

$$f_D^{\downarrow}(x^*,g) \geq 0 \;\; \forall g \in \Gamma(x^*,S). \tag{4.5.30}$$

Якщо виявилося, що

$$f_D^{\downarrow}(x^*,g) > 0 \;\; \forall g \in \Gamma(x^*,S),\, g \neq 0_n, \tag{4.5.31}$$

то x^* є точкою строгого локального мінімуму функції f на S.

Доведення. Необхідність. Припустимо протилежне. Тоді знайдеться $g_0 \in \Gamma(x^*,S)$ таке, що

$$f_D^{\downarrow}(x^*,g_0) = -a < 0. \tag{4.5.32}$$

Нехай послідовність $\{\alpha_k\}$ така, що $\alpha_k \downarrow 0$,

$$\frac{1}{\alpha_k}\left[f(x^*+\alpha_k g_0) - f(x^*)\right] \to \underline{\lim}_{\alpha\downarrow 0} \frac{1}{\alpha}\left[f(x^*+\alpha g_0) - f(x^*)\right].$$

Можна вважати, що (див. (4.5.32))

$$f(x^*+\alpha_k g_0) - f(x^*) \leq -\frac{1}{2}\alpha_k a. \tag{4.5.33}$$

З іншого боку, оскільки $g_0 \in \Gamma(x^*,S)$, а множина S в точці x^* допускає рівномірну конічну апроксимацію першого порядку, то знайдеться послідовність $\{x_k\}$ така, що $x_k \in S$, $\|x_k - x^* - \alpha_k g_0\| = o(\alpha_k)$. Маємо

$$f(x_k) = f(x^*+\alpha_k g_0) + (f(x_k) - f(x^*+\alpha_k g_0)). \tag{4.5.34}$$

Оскільки функція f ліпшицева, то

$$|f(x_k) - f(x^*+\alpha_k g_0)| \leq L\|x_k - x^* - \alpha_k g_0\| = L \cdot o(\alpha_k).$$

Тому з (4.5.33) і (4.5.34) одержуємо $f(x_k) \leq f(x^*) - \frac{1}{2}a\alpha_k + L \cdot o(\alpha_k)$. При досить великих k буде $f(x_k) < f(x^*) - \frac{1}{4}a\alpha_k$, що суперечить тому, що x^* — точка мінімуму функції f на S (тому що $x_k \in S$).

Достатність. Нехай виконана (4.5.31). Доведемо, що тоді x^* — точка строгого локального мінімуму функції f на множині S. Припустимо протилежне. Тоді знайдеться послідовність точок $\{x_k\}$ така, що $x_k \in S$, $x_k \to x^*$,

$$f(x_k) \le f(x^*) \tag{4.5.35}$$

Можна вважати, що $g_k = \frac{x_k - x^*}{\|x_k - x^*\|} \to g_0$. Зрозуміло, що $g_0 \in \Gamma(x^*,S)$. Покладемо $\alpha_k = \|x_k - x^*\|$. Тоді $x_k = x^* + \alpha_k g_k$. З (4.5.35) маємо

$$0 \ge f(x_k) - f(x^*) = f(x_k) - f(x^* + \alpha_k g_0) + f(x^* + \alpha_k g_0) - f(x^*) .$$

Звідси

$$0 \ge \frac{1}{\alpha_k}\left[f(x^* + \alpha_k g_k) - f(x^* + \alpha_k g_0)\right] + \frac{1}{\alpha_k}\left[f(x^* + \alpha_k g_0) - f(x^*)\right] \tag{4.5.36}$$

Оскільки функція f ліпшицева, то

$$|f(x^* + \alpha_k g_k) - f(x^* + \alpha_k g_0)| \le L\alpha_k \|g_k - g_0\| .$$

Звідси і з (4.5.36) випливає

$$f_D^{\downarrow}(x^*,g_0) \le 0,$$

що суперечить (4.5.31). □

Зауваження 4.5.4. При доведенні достатності не використовувався той факт, що множина S допускає рівномірну конічну апроксимацію в точці x^*.

Лема 4.5.7. *Нехай у точці $x^{**} \in S$ множина S допускає рівномірну конічну апроксимацію першого порядку, а функція f локально ліпшицева в околі точки x^{**}. Для того, щоб у точці x^{**} функція f досягала свого найбільшого на множині S значення, необхідно, щоб*

$$f_D^{\uparrow}(x^{**},g) \le 0 \quad \forall g \in \Gamma(x^{**},S), \tag{4.5.37}$$

Якщо виявилося, що

$$f_D^{\uparrow}(x^{**},g) < 0 \quad \forall g \in \Gamma(x^{**},S),\ g \ne 0_n, \tag{4.5.38}$$

то x^{**} є точкою строгого локального максимуму функції f на множині S.

Означення 4.5.11. Точка $x \in S$, що задовольняє умову (4.5.30), називається *Діні-inf–стаціонарною точкою функції f на множині S.*

Означення 4.5.12. Якщо $x_0 \in S$ не є Діні-inf–стаціонарною точкою функції f на множині S, то напрямок $g_0 \in \Gamma(x_0,S)$, $\|g_0\| = 1$, для якого

$$f_D^{\downarrow}(x_0,g_0) = \inf_{g \in \Gamma(x_0,S), \|g\|=1} f_D^{\downarrow}(x_0,g),$$

називається *напрямком найшвидшого спуску (н. н. с.) функції f на множині S.*

Означення 4.5.13. Точка $x^{**} \in S$, що задовольняє умові (4.5.37), називається *Діні-sup–стаціонарною точкою функції f на множині S.*

Означення 4.5.14. Якщо точка $x_0 \in S$ не є Діні-sup–стаціонарною точкою функції f на множині S, то напрямок $g_0 \in \Gamma(x_0)$, $\|g_0\| = 1$, такий, що

$$f_D^{\uparrow}(x_0,g_0) = \sup_{g \in \Gamma(x_0,S), \|g\|=1} f_D^{\uparrow}(x_0,g),$$

називається *напрямком найшвидшого підйому (н. н. п.) функції f на множині S.*

Зауваження 4.5.5. З викладеного вище випливає, що необхідні умови екстремуму близькі до достатніх умов локального екстремуму. Це свідчить про те, що похідні за напрямками (а якщо вони не існують, то похідні Діні) дозволяють будувати цілком задовільні апроксимації функцій. Раніше відмічалося, що і конічні апроксимації теж описуються за допомогою похідних Діні функцій, що задають множину S.

Зауваження 4.5.6. Отримані вище умови екстремуму в наведеному вигляді не досить конструктивні. Для їхнього практичного використання необхідно вміти обчислювати похідні за напрямками (чи похідні Діні) і будувати конуси Булігана. Це робиться для конкретних класів функцій і множин. Деякі з них розглядаються в наступних розділах.

Розділ V
Субдиференційовні функції

5.1. УЗАГАЛЬНЕНІ ПОХІДНІ ТА СУБДИФЕРЕНЦІАЛИ

Субдиференціал $\partial f(x)$ опуклої функції $f(x)$ має багато корисних властивостей похідної. Наприклад, за допомогою субдиференціала задається необхідна умова $0 \in \partial f(x)$ того, що точка x є точкою (локального) мінімуму функції $f(x)$. Ця умова спрощується до умови $f'(x) = 0$ в тому випадку, коли функція $f(x)$ диференційовна в точці x. Субдиференціал в багатьох випадках задовольняє умову адитивності $\partial(\lambda f + \mu g)(x) = \lambda\partial f(x) + \mu\partial g(x), \lambda > 0, \mu \geq 0$. Властивості субдиференціала опуклих функцій наводять на думку будувати класи функцій, що мають принаймні частину базових властивостей опуклих функцій. Б. М. Пшеничний [47,48] визначив та досліджував новий клас негладких функцій – субдиференційовні функції. (Б. М. Пшеничний називав їх квазідиференційовними.)

Означення 5.1.1. Функція $f : S \to \mathbb{R}^1$, $S \subset \mathbb{R}^n$, субдиференційовна за Діні в точці $x \in S$, якщо існує така опукла компактна множина $\partial f(x) \subset \mathbb{R}^n$, що

$$f'_D(x,g) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha g) - f(x)}{\alpha} = \max_{v \in \partial f(x)} \langle v,g \rangle \quad \forall g \in \mathbb{R}^n.$$

Множина $\partial f(x) \subset \mathbb{R}^n$, для якої виконується вказане співвідношення, називається *субдиференціалом Діні* функції f в точці $x \in S$.

Означення 5.1.2. Функція $f : S \to \mathbb{R}^1$, $S \subset \mathbb{R}^n$, субдиференційовна за Адамаром в точці $x \in S$, якщо існує опукла компактна множина $\partial f(x) \subset \mathbb{R}^n$ така, що

$$f'_H(x,g) = \lim_{[\alpha,g'] \to [+0,g]} \frac{f(x+\alpha g') - f(x)}{\alpha} = \max_{v \in \partial f(x)} \langle v,g \rangle \quad \forall g \in \mathbb{R}^n.$$

Множина $\partial f(x) \subset \mathbb{R}^n$, для якої виконується вказане співвідношення, називається *субдиференціалом Адамара* функції f в точці $x \in S$.

Клас субдиференційовних функцій включає диференційовні функції, опуклі функції. Він замкнутий відносно додавання та множення на додатні числа. Максимум субдиференційовних функцій також є субдиференційовною функцією. Субдиференційовні функції мають всі властивості диференційовних за напрямком функцій.

Умови екстремуму субдиференційовних функції можна виразити в термінах субдиференціалу.

Якщо точка $x^* \in S$ є точкою локального мінімуму субдиференційовної за Діні чи Адамаром функції f на множині S, то необхідною є умова

$$0 \in \partial f(x^*).$$

Якщо функція f субдиференційовна за Адамаром в точці $x^* \in S$, то умова

$$0 \in \operatorname{int} \partial f(x^*)$$

є достатньою умовою мінімуму функції f в точці $x^* \in S$.

Якщо точка $x^{**} \in \mathbb{R}^n$ є точкою локального максимуму субдиференційовної за Діні чи Адамаром функції f, то необхідною є умова

$$\partial f(x^{**}) = \{0\}.$$

Якщо $0 \notin \partial f(x)$, то напрямок

$$g_0(x) = -z(x)/\|z(x)\|, \quad \|z(x)\| = \min_{v \in \partial f(x)} \|v\|,$$

є напрямком найшвидшого спуску (за Діні чи Адамаром в залежності від типу субдиференційовності) функції f в точці x. Напрямок найшвидшого спуску визначається однозначно.

Якщо $\partial f(x) \neq \{0\}$, то напрямок

$$g_1(x) = z_1(x)/\|z_1(x)\|, \quad \|z_1(x)\| = \max_{v \in \partial f(x)} \|v\|,$$

є напрямком найшвидшого підйому (за Діні чи Адамаром в залежності від типу субдиференційовності функції) функції f в точці x. Напрямок найшвидшого підйому визначається неоднозначно.

В 1972 році Н.З.Шор [59] запропонував конструкцію майже градієнту. Множина майже градієнтів будується наступним чином. Розглянемо на множині S локально ліпшицеву функцію f. Така функція диференційовна на S майже скрізь. Зафіксуємо $x \in S$ і побудуємо множину

$$\partial_{Sh} f(x) = \left\{ v \in \mathbb{R}^n | \exists \{x_i\} : v = \lim_{i \to \infty} f'(x_i), x_i \to x, x_i \in Q \right\},$$

де $Q \subset S$ – множина точок, в яких функція f диференційовна. Оскільки функція f диференційовна на S майже скрізь, то міра множини $S \setminus Q$ рівна нулю. Множина $\partial_{Sh} f(x)$ називається субдиференціалом Шора, а елементи цієї множини називаються майже градієнтами.

В 1973 році Ф. Кларк (студент Р. Т. Рокафеллара) запропонував конструкцію узагальненого градієнту [25].

Означення 5.1.3. Величини

$$f_{\underset{Cl}{\uparrow}}(x,g) = \limsup_{[x',\alpha] \to [x,+0]} \frac{1}{\alpha}(f(x'+\alpha g) - f(x')),$$

$$f_{\underset{Cl}{\downarrow}}(x,g) = \liminf_{[x',\alpha] \to [x,+0]} \frac{1}{\alpha}(f(x'+\alpha g) - f(x'))$$

називаються відповідно верхньою і нижньою похідною Кларка.

Якщо функція f локально ліпшицева, то ці величини скінченні. У цьому випадку визначимо субдиференціал Кларка за формулою

$$\partial_{Cl} f(x) = \operatorname{conv} \left\{ v \in \mathbb{R}^n | \exists \{x_i\} : v = \lim_{i \to \infty} f'(x_i), x_i \to x, x_i \in Q \right\},$$

тобто $\partial_{Cl} f(x) = \operatorname{conv} \partial_{Sh} f(x)$. Множина $\partial_{Cl} f(x)$ опукла і компактна. Елементи множини $\partial_{Cl} f(x)$ називаються узагальненими градієнтами. Кожен майже градієнт є узагальненим градієнтом. Зворотнє твердження невірне. Кларк показав, що

$$f_{\substack{\uparrow\\ Cl}}(x,g) = \max_{v \in \partial_{Cl} f(x)} \langle v,g \rangle,$$

$$f_{\substack{\downarrow\\ Cl}}(x,g) = \min_{w \in \partial_{Cl} f(x)} \langle w,g \rangle.$$

Більш детально властивості узагальнених градієнтів та субдиференціалу Кларка описані далі.

Р. Т. Рокафеллар [50] визначив таку похідну за напрямком

$$f_{\substack{\uparrow\\ R}}(x,g) = \limsup_{[x',g',\alpha] \to [x,g,+0]} \frac{1}{\alpha}(f(x' + \alpha g') - f(x')).$$

Якщо функція f локально ліпшицева, то

$$f_{\substack{\uparrow\\ R}}(x,g) = f_{\substack{\uparrow\\ Cl}}(x,g).$$

П. Мішель та І. П. Пено [32] запропонували використовувати узагальнені похідні

$$f_{\substack{\uparrow\\ mp}}(x,g) = \sup_{q \in \mathbb{R}^n} \overline{\lim}_{\alpha \downarrow 0} \frac{1}{\alpha}(f(x + \alpha(g+q)) - f(x + \alpha q)),$$

$$f_{\substack{\downarrow\\ mp}}(x,g) = \inf_{q \in \mathbb{R}^n} \underline{\lim}_{\alpha \downarrow 0} \frac{1}{\alpha}(f(x + \alpha(g+q)) - f(x + \alpha q)),$$

які називаються відповідно верхньою і нижньою похідною Мішеля–Пено. Якщо функція f локально ліпшицева, то існує така опукла і компактна множина $\partial_{mp} f(x) \subset \mathbb{R}^n$, що

$$f^{\uparrow}_{mp}(x,g) = \max_{v \in \partial_{mp} f(x)} \langle v,g \rangle,$$

$$f^{\downarrow}_{mp}(x,g) = \min_{w \in \partial_{mp} f(x)} \langle w,g \rangle.$$

Множину $\partial_{mp} f(x)$ називають також малим субдиференціалом.

Якщо функція $f : \mathbb{R}^n \to \mathbb{R}^1$, задовольняє умову Ліпшиця з константою L в околі точки x, то похідна за напрямком Діні, похідна Кларка та похідна Мішеля–Пено є сублінійними функціями і задовольняють умову

$$f'_D(x,g) \le f_{\substack{\uparrow\\ mp}}(x,g) \le f_{\substack{\uparrow\\ Cl}}(x,g) \le L\|g\|.$$

Субдиференціали Діні $\partial_- f(x)$ (може бути порожньою множиною), Кларка $\partial_{Cl} f(x)$ (непорожня множина) та Мішеля–Пено $\partial_{mp} f(x)$ (непорожня множина) є компактними опуклими множинами, що задовольняють умову

$$\partial_- f(x) \subset \partial_{mp} f(x) \subset \partial_{Cl} f(x) \subset LB(0).$$

Якщо точка x^* є точкою локального мінімуму субдиференційовної функції f, то виконуються необхідні умови

$$0 \in \partial_- f(x) \subset \partial_{mp} f(x) \subset \partial_{Cl} f(x).$$

5.2. СУБДИФЕРЕНЦІАЛ КЛАРКА

5.2.1. Верхня і нижня регуляризації функцій

У попередніх розділах похідна $f'(x,g) = f'_x(g)$ функції f в точці x за напрямком g розглядалась в основному як функція напрямку при фіксованій точці x. В класичному аналізі, як правило, розглядають неперервно диференційовні функції, тобто функції, для яких відображення $f' : x \to f'_x$ є неперервним. З точки зору негладкого аналізу неперервність похідної, а іноді навіть її напівнеперервність вважається досить жорсткою умовою. Виявляється, що топологічні властивості відображення $f'(x,g)$ в деякі точці x тісно пов'язані з алгебраїчними властивостями похідної $f'_x(x,g)$ як функції напрямку g. Так, напівнеперевність зверху (відповідно знизу) відображення f' веде до сублінійністі (відповідно суперлінійність) похідної f'_x. Звідси випливає, що неперервність f' веде до лінійності f'_x і тим самим до існування диференціала.

Дослідимо більш детально напівнеперервність похідної. Насамперед зауважимо, що в тому випадку, коли похідна не є напівнеперервна зверху (відповідно знизу), доцільно розглянути її верхню (відповідно нижню) регуляризацію.

Нехай $f : \mathbb{R}^n \to \mathbb{R}^1$ – дійсна функція, що визначена на підмножині X простору $\mathbb{R}^n$, і нехай $x \in \operatorname{cl} X$.

Означення 5.2.1. Функції

$$\bar{f}(x) = \inf_{\delta > 0} \sup_{\|x' - x\| < \delta, x' \in X} f(x'),$$

$$\underline{f}(x) = \sup_{\delta > 0} \inf_{\|x' - x\| < \delta, x' \in X} f(x').$$

називаються відповідно *верхньою і нижньою регуляризаціями функції f на множині X*.

Зрозуміло, що виконуються рівності

$$\overline{f}(x) = \lim_{\delta \downarrow 0} \sup_{\|x'-x\|<\delta, x' \in X} f(x'),$$

$$\underline{f}(x) = \lim_{\delta \downarrow 0} \inf_{\|x'-x\|<\delta, x' \in X} f(x').$$

Нагадаємо, що верхня і нижня границі функції f визначені на X в точці $x \in \mathrm{cl}\, X$, і їх можна представити у вигляді

$$\varlimsup_{x' \to x} f(x') = \lim_{\delta \downarrow 0} \sup_{\|x'-x\|<\delta, x' \in X, x' \neq x} f(x'),$$

$$\varliminf_{x' \to x} f(x') = \lim_{\delta \downarrow 0} \inf_{\|x'-x\|<\delta, x' \in X, x' \neq x} f(x').$$

Звідси випливає, що

$$\overline{f}(x) = \max\left(f(x), \varlimsup_{x' \to x} f(x')\right),$$

$$\underline{f}(x) = \min\left(f(x), \varliminf_{x' \to x} f(x')\right).$$

Завжди справджуються співвідношення

$$\underline{f}(x) \le f(x) \le \overline{f}(x).$$

Якщо функція f напівнеперервна зверху в точці x, то $\overline{f}(x) = f(x)$. Дійсно, припустимо, що $\overline{f}(x) > f(x)$. Тоді в силу напівнеперервності зверху знайдеться таке $\delta > 0$, що $\overline{f}(x) > f(x')$ для всіх $x' \in X \cap B_\delta(x)$, а це суперечить визначенню функції $\overline{f}$. Таким же способом перевіряється, що для напівнеперервної знизу функції f виконується рівність $\underline{f}(x) = f(x)$.

Теорема 5.2.1. *Для кожної функції f її верхня регуляризація $\overline{f}$ напівнепервна зверху, а нижня регуляризація $\underline{f}$ напівнеперервна знизу.*

Доведення. Обмежимося випадком функції $\overline{f}$. Розглянемо точку $x \in X$ і зафіксуємо $\varepsilon > 0$. За визначенням верхньої регуляризації знайдеться таке $\delta > 0$, що

$$\sup_{\|x'-x\|<\delta, x' \in X} f(x') < \overline{f}(x) + \varepsilon$$

Нехай $y \in B_{\delta/2}(x)$ і $x' \in B_{\delta/2}(y) \cap X$. Тоді $x' \in B_\delta(x)$ і тому

$$\sup_{\|x'-y\|<\delta/2, x' \in X} f(x') < \overline{f}(x) + \varepsilon.$$

Звідси випливає співвідношення

$$\overline{f}(y) < \overline{f}(x) + \varepsilon \quad \forall y \in B_{\delta/2}(x) \cap X,$$

яке показує, що функція $\overline{f}$ напівнеперервна зверху. $\square$

Виявляється, що верхня регуляризація $\overline{f}$ є найменшою напівнеперервною зверху функцією, яка мажорує f, а нижня регуляризація $\underline{f}$ є найбільшою напівнеперервною знизу функцією, яка мінорує f. Іншими словами, справедливе наступне твердження.

Теорема 5.2.2. *Якщо $f_1(x)$ – напівнеперервна зверху функція і $f_1(x) \geq f(x)$ при всіх x, то $f_1(x) \geq \overline{f}(x)$. Якщо $f_2(x)$ – напівнеперервна знизу функція і $f_2(x) \leq f(x)$ при всіх x, то $f_2(x) \leq \underline{f}(x)$.*

Доведення. Обмежимося функцією $\overline{f}$. Якщо $f_1(x) \geq f(x)$ при всіх x, тоді з означення випливає, що $\overline{f_1}(x) \geq \overline{f}(x)$. Оскільки функція $f_1(x)$ напівнеперервна зверху, то $\overline{f_1}(x) = f_1(x)$ і, відповідно, $f_1(x) \geq \overline{f}(x)$. $\square$

5.2.2. Верхня і нижня похідні Кларка

Нехай функція f визначена на відкритій множині X в просторі $\mathbb{R}^n$. Розглянемо її верхню та нижню похідні Діні за деяким фіксованим напрямком g при всіх x із X, тобто функції $x \to f_{\underset{D}{\uparrow}}(x,g)$, $x \in X$ та $x \to f_{\underset{D}{\downarrow}}(x,g)$, $x \in X$. Визначимо верхню регуляризацію

$$\overline{f}_{\underset{D}{\uparrow}}(x,g) = \max\left(f_{\underset{D}{\uparrow}}(x,g), \overline{\lim_{x' \to x}} f_{\underset{D}{\uparrow}}(x',g)\right)$$

верхньої похідної Діні $f_{\underset{D}{\uparrow}}(x,g)$ і нижню регуляризацію

$$\underline{f}_{\underset{D}{\downarrow}}(x,g) = \min\left(f_{\underset{D}{\downarrow}}(x,g), \underline{\lim_{x' \to x}} f_{\underset{D}{\downarrow}}(x',g)\right)$$

нижньої похідної Діні $f_{\underset{D}{\downarrow}}(x,g)$.

Будемо вважати, що функція f *локально ліпшицева* на множині X. Останнє означає, що для довільної точки $x \in X$ існує окіл V цієї точки, в якому функція f задовольняє умову Ліпшиця з деякою константою (яка, можливо, залежить від цього околу). З локальної ліпшицевості випливає, що верхня похідна Діні $f_{\underset{D}{\uparrow}}(x,g)$ і нижня похідна Діні $f_{\underset{D}{\downarrow}}(x,g)$ при довільному g обмежені в деякому околі точки x і числа $\overline{f}_{\underset{D}{\uparrow}}(x,g)$ та $\underline{f}_{\underset{D}{\downarrow}}(x,g)$ скінченні. Зауважимо, що в даному випадку похідна за напрямком (якщо вона існує) співпадає з похідною Адамара.

Теорема 5.2.3. *Справедливі рівності*

$$\overline{f}_{\underset{D}{\uparrow}}(x,g) = \overline{\lim_{x' \to x, \alpha \downarrow 0}} \frac{1}{\alpha}(f(x' + \alpha g) - f(x')),$$

$$\underline{f}_{\underset{D}{\downarrow}}(x,g) = \varliminf_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g) - f(x')).$$

Доведення. Обмежимося перевіркою першої рівності. Нехай

$$A_1 = \varlimsup_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g) - f(x')),$$

$$A_2 = \overline{f}_{\underset{D}{\uparrow}}(x,g) = \inf_{\delta \downarrow 0} \sup_{\|x' - x\| < \delta} f_{\underset{D}{\uparrow}}(x',g).$$

Зафіксуємо $\varepsilon > 0$ і знайдемо таку послідовність $\{x_k\}$, що $x_k \to x$,

$$\overline{f}_{\underset{D}{\uparrow}}(x_k,g) = \varlimsup_{\alpha \downarrow 0} \frac{1}{\alpha}(f(x_k + \alpha g) - f(x_k)) > A_2 - \varepsilon.$$

Для довільного k знайдеться таке число α_k, що

$$\frac{1}{\alpha_k}(f(x_k + \alpha_k g) - f(x_k)) > A_2 - \varepsilon, \quad 0 < \alpha_k < 1/k. \tag{5.2.1}$$

Нерівність (5.2.1) показує, що $A_1 \geq A_2 - \varepsilon$. Оскільки ε – довільне додатне число, то $A_1 \geq A_2$.

З визначення числа A_2 випливає, що для довільного $\varepsilon > 0$ знайдеться $\delta > 0$, яке має таку властивість: якщо y належить кулі $B_\delta(x)$ радіуса δ з центром в точці x, то $f_{\underset{D}{\uparrow}}(y,g) < A_2 + \varepsilon$. Якщо $\|x - x'\| < \delta/2$ і $0 < \alpha < \frac{\delta}{2\|g\|}$, то точки x' та $x' + \alpha g$ входять в $B_\delta(x)$. Тому, використовуючи теорему 4.1.2, отримаємо нерівність $f(x' + \alpha g) - f(x') \leq \alpha(A_2 + \varepsilon)$. Звідси випливає нерівність $A_1 \leq A_2 + \varepsilon$, з якої, оскільки ε довільне, випливає нерівність $A_1 \leq A_2$. $\square$

Зауважимо, що величина $\varlimsup_{x' \to x} f_{\underset{D}{\uparrow}}(x',g)$ є повторною верхньою границею

$$\varlimsup_{x' \to x} \varlimsup_{\alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g) - f(x')).$$

Оскільки

$$\overline{f}_{\underset{D}{\uparrow}}(x,g) = \max\left(f_{\underset{D}{\uparrow}}(x,g), \varlimsup_{x' \to x} f_{\underset{D}{\uparrow}}(x',g)\right),$$

то теорема 5.2.3 показує, що у тому випадку, коли $f_{\underset{D}{\uparrow}}(x,g) \leq \varlimsup_{x' \to x} f_{\underset{D}{\uparrow}}(x',g)$, подвійна верхня границя

$$\varlimsup_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g) - f(x'))$$

співпадає з вказаною повторною верхньою границею.

Означення 5.2.2. Величини

$$f_{\underset{Cl}{\uparrow}}(x,g) = \varlimsup_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g) - f(x')),$$

$$f_{\underset{Cl}{\downarrow}}(x,g) = \underline{\lim}_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g) - f(x'))$$

називаються відповідно *верхньою і нижньою похідною Кларка.*

Теорема 5.2.3 показує, що при фіксованому напрямку g верхня похідна Кларка є верхньою регуляризацією верхньої похідної Діні, а нижня похідна Кларка є нижньою регуляризацією нижньої похідної Діні.

Теорема 5.2.4. *При будь-якому g функція $x \to f_{\underset{Cl}{\uparrow}}(x,g)$ напівнеперервна зверху, а функція $x \to f_{\underset{Cl}{\downarrow}}(x,g)$ напівнеперервна знизу.*

Доведення. Доведення випливає з тверджень 5.2.1 та 5.2.3. □

Сформулюємо одну з основних властивостей похідних Кларка.

Теорема 5.2.5. *При довільному x функція $g \to f_{\underset{Cl}{\uparrow}}(x,g)$ сублінійна, а функція $g \to f_{\underset{Cl}{\downarrow}}(x,g)$ суперлінійна.*

Доведення. Розглянемо функцію $g \to f_{\underset{Cl}{\uparrow}}(x,g)$. З рівності

$$f_{\underset{Cl}{\uparrow}}(x,\lambda g) = \overline{\lim}_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha\lambda g) - f(x')) =$$

$$= \overline{\lim}_{x' \to x, \alpha \downarrow 0} \lambda \frac{1}{\alpha\lambda}(f(x' + \alpha\lambda g) - f(x')) = \lambda f_{\underset{Cl}{\uparrow}}(x,g)$$

випливає, що дана функція додатньо однорідна. Покажемо, що вона субадитивна. Використовуючи субадитивність верхньої границі, маємо

$$f_{\underset{Cl}{\uparrow}}(x,g_1 + g_2) = \overline{\lim}_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}\left[\, f(x' + \alpha g_1 + \alpha g_2) - f(x')\,\right] \le$$

$$\le \overline{\lim}_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g_1 + \alpha g_2) - f(x' + \alpha g_1)) +$$

$$+ \overline{\lim}_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g_1) - f(x')) = f_{\underset{Cl}{\uparrow}}(x,g_2) + f_{\underset{Cl}{\uparrow}}(x,g_1).$$

□

Оскільки функція $g \to f_{\underset{Cl}{\uparrow}}(x,g)$ сублінійна, то справедлива рівність

$$f_{\underset{Cl}{\uparrow}}(x,g) = \max_{l \in \underline{\partial} f_{\underset{Cl}{\uparrow}}(x)} \langle l,g \rangle, \tag{5.2.2}$$

де $\underline{\partial} f_{\underset{Cl}{\uparrow}}(x)$– субдиференціал цієї функції. Аналогічно

$$f_{\underset{Cl}{\downarrow}}(x,g) = \min_{l \in \overline{\partial} f_{\underset{Cl}{\downarrow}}(x)} \langle l,g \rangle, \tag{5.2.3}$$

де $\overline{\partial} f_{\underset{Cl}{\downarrow}}(x)$– супердиференціал суперлінійної функції $g \to f_{\underset{Cl}{\downarrow}}(x,g)$. Покажемо, що множини $\underline{\partial} f_{\underset{Cl}{\uparrow}}(x)$ і $\overline{\partial} f_{\underset{Cl}{\downarrow}}(x)$ співпадають. Для цього перевіримо виконання наступних тверджень.

Теорема 5.2.6. *Має місце рівність*

$$f_{\underset{Cl}{\uparrow}}(x,-g) = (-f)_{\underset{Cl}{\uparrow}}(x,g)$$

Доведення. За визначенням

$$f_{\underset{Cl}{\uparrow}}(x,-g) = \varlimsup_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x'-\alpha g) - f(x')).$$

Покладемо $x' - \alpha g = z$. Тоді $x' = z + \alpha g$ і, відповідно,

$$f_{\underset{Cl}{\uparrow}}(x,-g) = \varlimsup_{z \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(z) - f(z+\alpha g)) =$$

$$= \varlimsup_{z \to x, \alpha \downarrow 0} \frac{1}{\alpha}((-f)(z+\alpha g) - (-f)(z)) = (-f)_{\underset{Cl}{\uparrow}}(x,g).$$

□

Теорема 5.2.7. *Справедлива рівність*

$$f_{\underset{Cl}{\downarrow}}(x,g) = -f_{\underset{Cl}{\uparrow}}(x,-g).$$

Доведення. Доведення випливає з теореми 5.2.6 і рівності

$$f_{\underset{Cl}{\downarrow}}(x,g) = \varliminf_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x'+\alpha g) - f(x')) =$$

$$= -\varlimsup_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(-f(x'+\alpha g) - (-f)(x')) = -(-f)_{\underset{Cl}{\uparrow}}(x,g).$$

□

Використовуючи теорему 5.2.7 і формули (5.2.2) і (5.2.3), маємо при всіх $g \in R^n$

$$\max_{l \in \underline{\partial} f_{\underset{Cl}{\uparrow}}(x)} \langle l,g \rangle = f_{\underset{Cl}{\uparrow}}(x,g) =$$

$$= -f_{\underset{Cl}{\downarrow}}(x,-g) = -\min_{l \in \overline{\partial} f_{\underset{Cl}{\downarrow}}(x)} \langle l,-g \rangle = \max_{l \in \overline{\partial} f_{\underset{Cl}{\downarrow}}(x)} \langle l,g \rangle$$

Ці рівності показують, що опорні функції компактів $\underline{\partial} f_{\underset{Cl}{\uparrow}}(x)$ і $\overline{\partial} f_{\underset{Cl}{\downarrow}}(x)$ співпадають. Звідси випливає, що відповідні компакти теж співпадають.

Означення 5.2.3. Опуклий компакт, який одночасно є субдиференціалом верхньої похідної Кларка в точці x і супердиференціалом нижньої похідної Кларка в цій точці, називається *субдиференціалом Кларка* функції f в точці x і позначається символом $\partial_{Cl} f(x)$.

Отже. за означенням

$$f_{\underset{Cl}{\uparrow}}(x,g) = \max_{l\in\partial_{Cl}f(x)} \langle l,g\rangle, \quad f_{\underset{Cl}{\downarrow}}(x,g) = \min_{l\in\partial_{Cl}f(x)} \langle l,g\rangle \quad \forall g \in \mathbb{R}^n.$$

Означення 5.2.4. Елементи субдиференціала Кларка $\partial_{Cl} f(x)$ називаються *узагальненими градієнтами* функції f в точці x.

Теорема 5.2.8. *Багатозначне відображення $x \to \partial_{Cl} f(x)$ напівнеперервне зверху.*

5.2.3. Взаємозвязок похідної Кларка та похідної Діні

Теорема 5.2.9. *Якщо локально ліпшицева функція f визначена на відкритій множині X і її верхня похідна Діні $f_{\underset{D}{\uparrow}}(x,g)$ напівнеперервна зверху як функція точки x для довільного фіксованого напрямку g, то $f_{\underset{D}{\uparrow}}(x,g) = f_{\underset{Cl}{\uparrow}}(x,g)$ для всіх x та g.*

Доведення. Оскільки функція $x \to f_{\underset{D}{\uparrow}}(x,g)$ напівнеперервна зверху, то вона співпадає зі своєю верхньою регуляризацією $x \to f_{\underset{Cl}{\uparrow}}(x,g)$. □

Теорема 5.2.10. *Якщо нижня похідна Діні $f_{\underset{D}{\downarrow}}(x,g)$ локально ліпшицевої функції f, визначеної на відкритій множині X, напівнеперервна знизу як функція точки x при довільному фіксованому напрямку g, то ця похідна при всіх x та g співпадає з нижньою похідною Кларка $f_{\underset{Cl}{\downarrow}}(x,g)$.*

Наслідок 5.2.1. *З теорем 5.2.5 і 5.2.9 випливає така властивість похідної Діні. Якщо верхня похідна Діні напівнеперервна зверху як функція точки x, то вона сублінійна як функція напрямку g. Якщо функція $x \to f_{\underset{D}{\downarrow}}(x,g)$ напівнеперервна знизу при всіх g, то функція $g \to f_{\underset{D}{\downarrow}}(x,g)$ суперлінійна при всіх x.*

Якщо функція f диференційовна за напрямками і похідна $f'(x,g)$ при всіх g неперервна за x, то вона лінійна за g при всіх x і, тим самим, функція f неперервно диференційовна в класичному розумінні.

Наведені твердження допомагають з'ясувати співвідношення між верхньою і нижньою похідними Кларка, з одного боку, і похідною за напрямком – з іншого. Нехай похідна за напрямком $f'(x,g)$ напівнеперервна зверху при довільному g. Тоді справедливі рівності

$$f'(x,g) = f_{\underset{Cl}{\uparrow}}(x,g) = \max_{l\in\partial_{Cl}f(x)} \langle l,g\rangle,$$

$$f_{\underset{Cl}{\downarrow}}(x,g) = \min_{l \in \partial_{Cl} f(x)} \langle l, g \rangle.$$

Отже в даному випадку верхня похідна Кларка співпадає з $f'(x,g)$. В той же час нижня похідна Кларка відрізняється від похідної за напрямком в усіх тих точках x, де функція f не диференційовна за Гато. Важливі не лише числові відмінності між цими похідними, але, що істотно, принципові відмінності в їх властивостях: похідна $f'(x,g)$ субадитивна, вона виражається через лінійні функції за допомогою операції максимума; похідна ж $f_{\underset{Cl}{\downarrow}}(x,g)$ супeрадитивна і виражається через лінійні функції за допомогою операції взяття мінімума.

Якщо похідна $f'(x,g)$ напівнеперервна знизу при всіх g, то

$$f'(x,g) = f_{\underset{Cl}{\downarrow}}(x,g) = \min_{l \in \partial_{Cl} f(x)} \langle l, g \rangle,$$

$$f_{\underset{Cl}{\uparrow}}(x,g) = \max_{l \in \partial_{Cl} f(x)} \langle l, g \rangle.$$

В даному випадку похідна Діні функції f в тих точках де вона не диференційовна за Гато відрізняється від верхньої похідної Кларка.

У тому випадку, коли похідна $f'(x,g)$ не є напівнеперервною ні знизу, ні зверху, можна лише стверджувати, що

$$f_{\underset{Cl}{\downarrow}}(x,g) \leq f'(x,g) \leq f_{\underset{Cl}{\uparrow}}(x,g).$$

Отже, верхня і нижня похідні Кларка є відповідно сублінійною мажорантою і суперлінійною мінорантою похідної $f'(x,g)$.

Похідна за напрямком (яка співпадає для локально ліпшицевих функцій з похідною Адамара) визначає точну апроксимацію першого порядку функції f в точці x в тому розумінні, що

$$\frac{1}{\|h\|}(f(x+h) - f(x)) - f'(x,h) \to 0 \quad \text{при} \quad h \to 0$$

Із сказаного випливає, що в тих точках x, де f не диференційовна за Гато, принаймі одна із похідних $f_{\underset{Cl}{\uparrow}}(x,g)$, $f_{\underset{Cl}{\downarrow}}(x,g)$ не дає апроксимації першого порядку. Можна лише стверджувати, що вказані похідні є верхня і нижня границі апроксимації.

5.2.4. Субдиференціал Кларка. Приклади

Приклад 5.2.1. Нехай f – опукла функція. Тоді функція f диференційовна за напрямками і функція $x \to f'(x,g)$ напівнеперервна зверху при будь-якому g. Таким чином, для опуклих функцій похідна за напрямком співпадає з верхньою похідною Кларка. Звідси випливає, що субдиференціал Кларка $\partial_{Cl} f(x)$ співпадає з субдиференціалом $\underline{\partial} f(x) = \partial f(x)$ опуклої функції f в точці x.

Приклад 5.2.2. Нехай f – угнута функція. Тоді функція $x \to f'(x,g)$ напівнеперервна знизу для довільного g. Тому похідна за напрямком співпадає з нижньою похідною Кларка, а субдиференціал Кларка $\partial_{Cl} f(x)$ співпадає з супердиференціалом $\overline{\partial} f(x)$ угнутої функції f в точці x.

Приклад 5.2.3. Нехай $f(x) = f_1(x) + f_2(x)$, де $f_1(x)$ – опукла, а $f_2(x)$ – угнута функції. В цьому випадку похідна за напрямками $f'(x,g)$ існує, проте вона не обов'язково напівнеперервна зверху або знизу (як функція точки x). Щодо цієї похідної можна стверджувати, що

$$f_{\underset{Cl}{\downarrow}}(x,g) \le f'(x,g) \le f_{\underset{Cl}{\uparrow}}(x,g),$$

при цьому "відстань" $f_{\underset{Cl}{\uparrow}}(x,g) - f_{\underset{Cl}{\downarrow}}(x,g)$ може бути досить великою. Оцінка субдиференціала Кларка $\partial_{Cl} f(x)$ через субдиференціал $\underline{\partial} f_1(x)$ і супердиференціал $\overline{\partial} f_2(x)$ наводиться далі.

Приклад 5.2.4. Нехай $f(x) = \max_{y \in S} \varphi(x,y), x \in X$, де X – відкрита множина, S – компакт, φ – неперервна разом з частинною похідною $\varphi'_x(x,y)$ функція на $X \times Y$. Тоді (див. наслідок 4.1.3)

$$f'(x,g) = \max_{y \in R(x)} \langle \varphi'_x(x,y), g \rangle \,, \tag{5.2.4}$$

де $R(x) = \{y \,|\, f(x) = \varphi(x,y)\}$. Нагадаємо (див. теорему 4.1.6), що відображення $R(x)$ напівнеперервне зверху. Можна перевірити, використовуючи неперервність функції $\varphi'_x(x,y)$ і напівнеперервність відображення $R(x)$, що похідна $f'(x,g)$ напівнеперевна зверху. Тому $f'(x,g)$ співпадає з верхньою похідною Кларка. Позначимо через $\underline{\partial} f(x)$ опуклу оболонку множини векторів $\{\varphi'_x(x,y) \,|\, y \in R(x)\}$. З (5.2.4) випливає, що

$$f'(x,g) = \max_{l \in \underline{\partial} f(x)} \langle l, g \rangle$$

при всіх g. Ця рівність показує, що

$$\partial_{Cl} f(x) = \underline{\partial} f(x).$$

Приклад 5.2.5. Нехай $f(x) = \min_{y \in S} \varphi(x,y)\, x \in X$, де множини X, S і функція φ такі ж, як і в попередньому прикладі. Тоді, використовуючи рівність

$$\min_{y \in S} \varphi(x,y) = -\max_{y \in S} (-\varphi(x,y)),$$

можна переконатися, що функція f диференційовна за напрямками, причому при довільних g функція $x \to f'(x,g)$ напівнеперервна знизу і справедливе представлення

$$f'(x,g) = \min_{y \in Q(x)} \langle \varphi'_x(x,y), g \rangle = \min_{l \in \overline{\partial} f(x)} \langle l, g \rangle,$$

де $Q(x) = \{y \in S \,|\, f(x) = \varphi(x,y)\}$, $\overline{\partial} f(x) = \mathrm{conv}\, \{\varphi'_x(x,y) \,|\, y \in Q(x)\}$. Звідси випливає, що субдиренціал Кларка співпадає з множиною $\overline{\partial} f(x)$, а нижня похідна Кларка – з похідною за напрямком.

Теорема 5.2.11. *Якщо функція f диференційовна за Гато в точці x, то $\nabla f(x) \in \underline{\partial}_{Cl} f(x)$.*

Доведення. Оскільки верхня похідна Кларка мажорує похідну за напрямком, то $\langle \nabla f(x), g \rangle \leq f_{\underset{Cl}{\uparrow}}(x,g)$ при всіх g. Це означає, що вектор $\nabla f(x)$ входить в субдиференціал сублінійної функції $g \to f_{\underset{Cl}{\uparrow}}(x,g)$, який співпадає з $\partial_{Cl} f(x)$. □

Кажуть, що функція f строго диференційовна в точці x, якщо вона має градієнт в цій точці і

$$\langle \nabla f(x), g \rangle = \lim_{x' \to x, \alpha \downarrow 0} \frac{1}{\alpha}(f(x' + \alpha g) - f(x')).$$

Можна перевірити, що неперервна диференційовність в точці x приводить до строгої диференційовності в цій точці. Безпосередньо з визначення випливає, що верхня похідна Кларка строго диференційовної функції f в точці x співпадає з лінійною функцією $g \to \langle \nabla f(x), g \rangle$, і тому в цьому випадку субдиференціал Кларка $\partial_{Cl} f(x)$ складається з одного елемента $\nabla f(x)$. При відсутності строгої диференційовності рівність $\partial_{Cl} f(x) = \nabla f(x)$ виконується не завжди. Відповідний приклад наведено нижче.

5.2.5. Субдиференціал Шора та субдиференціал Кларка

Відомо [51], що якщо функція f визначена на відкритій множині $X \subset \mathbb{R}^n$ і задовольняє умові Ліпшиця, то вона диференційовна майже всюди на X, тобто в всіх точках $x \in X$, за винятком, можливо, множини нульової лебегової міри, існує диференціал функції f. (Надалі будемо позначати цей диференціал або символом $df(x)$, або символом $\nabla f(x)$.) Використовуючи це, опишемо верхню похідну Кларка в термінах диференціального відображення $x \to df(x)$. Надалі n–вимірну міру Лебега позначимо символом μ_n. Нам будуть потрібні наступні твердження з теорії міри.

Лема 5.2.1. *Нехай $Z \subset \mathbb{R}^n$, $\mu_n(Z) = 0$, $\Pi_g = \left\{\alpha g \,\middle|\, \alpha \in \mathbb{R}^1 \right\}$ – пряма, яка проходить через нуль і точку $g \in \mathbb{R}^n$ $(g \neq 0_n)$. Тоді для майже всіх $y \in \mathbb{R}^n$ множина $(y + \Pi_g) \cap Z$ має нульову міру.*

Доведення. Нехай χ_z – характеристична функція множини Z: $\chi_z(z) = 0$, якщо $z \notin Z$; $\chi_z(z) = 1$, якщо $z \in Z$. Через P позначимо ортогональну вектору g гіперплощину. Ототожнимо точку $z \in \mathbb{R}^n$ з парою (y,α), де y– проекція z на P, а α – проекція на пряму Π_g. Використовуючи теорему

Фубіні [51], маємо

$$0 = \mu_n(Z) = \int\limits_{\mathbb{R}^n} \chi_z(z)dz = \int\limits_P \int\limits_{\mathbb{R}^1} \chi_z(y,\alpha)dyd\alpha =$$

$$= \int\limits_P dy \int\limits_{\mathbb{R}^1} \chi_z(y,\alpha)d\alpha = \int\limits_P \mu_1((y + \Pi_g) \cap Z)dy. \qquad (5.2.5)$$

Покладемо $E = \{y \in P \,|\mu_1((y + \Pi_g) \cap Z) > 0\}$. З (5.2.5) випливає, що $\mu_{n-1}(E) = 0$. Звідси випливає справедливість леми. $\square$

Лема 5.2.2. *Нехай функція $h(\alpha)$ визначена на відрізку $[a,b]$ і задовольняє на цьому відрізку умову Ліпшиця. Тоді*

$$\int\limits_a^b h'(\alpha)d\alpha = h(b) - h(a).$$

Доведення. Доведення випливає з абсолютної неперервності функції, яка задовольняє умові Ліпшиця (див., наприклад, [51]). Нагадаємо, що абсолютно неперервна функція майже всюди диференційовна і може бути визначена як первісна від своєї похідної. $\square$

Теорема 5.2.12. *Нехай функція f визначена на відкритій множині $X \subset \mathbb{R}^n$ і задовольняє умову Ліпшиця в деякому околі точки $x \in X$. Нехай Q – підмножина повної міри в цьому околі. Тоді для $g \in \mathbb{R}^n$ виконується рівність*

$$f_{\underset{Cl}{\uparrow}}(x,g) = \varlimsup_{x' \to x, x' \in Q} f_{\underset{D}{\uparrow}}(x',g).$$

Доведення. Використовуючи теорему 5.2.3, маємо

$$f_{\underset{Cl}{\uparrow}}(x,g) = \varlimsup_{x' \to x} f_{\underset{D}{\uparrow}}(x',g) \geq \varlimsup_{x' \to x, x' \in Q} f_{\underset{D}{\uparrow}}(x',g).$$

Припустимо, що $f_{\underset{Cl}{\uparrow}}(x,g) > \varlimsup\limits_{x' \to x, x' \in Q} f_{\underset{D}{\uparrow}}(x',g)$. Тоді знайдуться такі числа $\delta > 0$ і c, що

$$f_{\underset{Cl}{\uparrow}}(x,g) > c > f_{\underset{D}{\uparrow}}(x',g) \qquad (5.2.6)$$

при $x' \in B_\delta(x) \cap Q$. Для $y \in B_{\delta/2}(x)$ покладемо $M_y = B_\delta(x) \cap (y + \Pi_g)$, де $\Pi_g = \left\{\alpha g \,\middle|\, \alpha \in \mathbb{R}^1\right\}$. Множина M_y є непорожнім інтервалом, який лежить на прямій $y + \Pi_g$. З леми 5.2.1 випливає, що для майже всіх $y \in B_{\delta/2}(x)$ перетин інтервала M_y з множиною нульової міри $B_\delta(x) \setminus Q$ має нульову міру. Розглянемо лише точки вказаного перетину. Функція $h_y = f(y+\alpha g)$ задовольняє умову Ліпшиця і тому майже всюди диференційовна. Якщо в точці α існує похідна $h'_y(\alpha)$, то

$$h'_y(\alpha) = \lim_{\beta \downarrow 0} \frac{1}{\beta}(h_y(\alpha + \beta) - h_y(\alpha)) =$$

$$= \lim_{\beta\downarrow 0} \frac{1}{\beta}(f(y+\alpha g+\beta g) - f(y+\alpha g)) = f'(y+\alpha g,g).$$

Використовуючи лему 5.2.2 і нерівність (5.2.6), отримаємо для майже всіх $y \in B_{\delta/2}(x)$ і $\alpha \in \left(0, \frac{\delta}{2\|g\|}\right)$ (вважаємо, що $g \neq 0$)

$$f(y+\alpha g) - f(y) = h_y(\alpha) - h_y(0) = \int\limits_0^{\alpha} h'_y(\beta)d\beta =$$

$$= \int\limits_0^{\alpha} f'(y+\beta g,g)d\beta < \int\limits_0^{\alpha} cd\beta = c\cdot\alpha.$$

Оскільки функція f неперервна, то нерівність $f(y+\alpha g) - f(y) \leq c \cdot \alpha$ виконується при всіх $y \in B_{\delta/2}(x)$ і $\alpha \in \left(0, \frac{\delta}{2\|g\|}\right)$. Звідси випливає нерівність $f_{\underset{Cl}{\uparrow}}(x,g) \leq c$, яка суперечить співвідношенню (5.2.6). □

Наслідок 5.2.2. *Нехай Q – множина повної міри і нехай в точках $x' \in Q$ існує диференціал. Тоді*

$$f_{\underset{Cl}{\uparrow}}(x,g) = \varlimsup_{x'\to x, x'\in Q} \langle \nabla f(x'),g\rangle.$$

Нехай функція f задовольняє умову Ліпшиця в деякому околі точки x і нехай Q – деяка множина повної міри, в точках якої функція f має диференціал. Визначимо субдиференціал Шора як множину

$$\partial_{Sh} f(x) = \left\{ l \,|\exists\, \{x'_i\} : l = \lim_{i\to\infty} \nabla f(x'_i), x'_i \to x, x'_i \in Q \right\}.$$

Оскільки f – ліпшицева функція, то градієнти $\nabla f(y), y \in Q$, обмежені в сукупності і, відповідно, множина $\partial_{Sh} f(x)$ обмежена. Можна перевірити, що множина $\partial_{Sh} f(x)$ замкнута і тому компактна. Розглянемо опуклу оболонку $\operatorname{conv} \partial_{Sh} f(x)$ множини $\partial_{Sh} f(x)$. Оскільки $\partial_{Sh} f(x)$ – компактна множина, то і множина $\operatorname{conv} \partial_{Sh} f(x)$ компактна.

Теорема 5.2.13. *Справедлива рівність*

$$\partial_{Cl} f(x) = \operatorname{conv} \partial_{Sh} f(x).$$

Доведення. Безпосередньо з означення множини $\partial_{Sh} f(x)$ випливає, що для кожного $g \in \mathbb{R}^n$ справджується рівність

$$\max_{l \in \partial_{Sh} f(x)} \langle l,g\rangle = \varlimsup_{x'\to x, x'\in Q} \langle \nabla f(x'),g\rangle.$$

З теореми 5.2.12 випливає, що $\max\limits_{l \in \partial_{Sh} f(x)} \langle l,g\rangle = f_{\underset{Cl}{\uparrow}}(x,g)$. Оскільки субдиференціал сублінійної функції $g \to f_{\underset{Cl}{\uparrow}}(x,g)$ співпадає з множиною $\partial_{Cl} f(x)$, то, використовуючи двоїстість Мінковського, приходимо до необхідної рівності $\partial_{Cl} f(x) = \operatorname{conv} \partial_{Sh} f(x)$. □

Наведемо приклади застосування теореми до знаходження субдиференціала Кларка.

Приклад 5.2.6. Розглянемо функцію f, що визначена на площині рівністю

$$f(z) = ||x| - |y||\,, \quad z = (x,y) \in \mathbb{R}^2.$$

Нехай Q складається з точок $z = (x,y)$, які не лежать на прямих $x = y$ і $x = -y$. Зрозуміло, що $\mu(\mathbb{R}^2 \backslash Q) = 0$. В точках множини Q функція f має диференціал. Якщо $(x,y) \in Q$, то число $f(x,y)$ співпадає з одним із чисел $x + y,\, - x - y,\, - x + y, x - y$. В перетині довільного околу точки $z_0 = (0,0)$ і множини Q функція f має градієнтами вектори $a_1 = (1,1)$, $a_2 = (1,-1)$, $a_3 = (-1,-1)$, $a_4 = (-1,1)$, і лише ці вектори. Тоді з теореми 5.2.13 випливає, що субдиференціал Кларка $\partial_{Cl} f(x)$ функції f в точці z_0 співпадає з квадратом, вершинами якого є вказані точки.

Приклад 5.2.7. Для $z = (x,y) \in \mathbb{R}^2$ покладемо

$$f(x,y) = \max(y - x^2,0) + \min(y + x^2,0).$$

Нехай $S_1 = \left\{(x,y) \in \mathbb{R}^2 \left| y > x^2 \right.\right\}$– внутрішність параболи $y = x^2$, $S_2 = \left\{(x,y) \in \mathbb{R}^2 \left| y < -x^2 \right.\right\}$– внутрішність параболи $y = -x^2$, $S_3 = \mathbb{R}^2 \backslash (S_1 \cup S_2)$. Тоді

$$f(x,y) = \begin{cases} y - x^2, & \forall (x,y) \in S_1, \\ y + x^2, & \forall (x,y) \in S_2, \\ 0, & \forall (x,y) \in S_3. \end{cases}$$

Нехай Q складається з точок площини, які не належать параболам $y = x^2$ і $y = -x^2$. Міра множини $\mathbb{R}^2 \backslash Q$ дорівнює нулю і в точках множини Q функція f неперервно диференційовна. При цьому

$$\nabla f(x,y) = \begin{cases} (-2x,1), & \forall (x,y) \in S_1, \\ (2x,1), & \forall (x,y) \in S_2, \\ (0,0), & \forall (x,y) \in S_3 \cap Q. \end{cases}$$

Нехай $z_k = (x_k,y_k) \to (0,0)$, $z_k \in Q$. Тоді послідовність $\nabla f(z_k)$ може мати граничними точками лише орт $e_2 = (0,1)$ і початок координат $z_0 = (0,0)$. Звідси випливає, що субдиференціал Кларка $\partial_{Cl} f(z_0)$ співпадає з відрізком, кінці якого є $e_2 = (0,1)$ і $z_0 = (0,0)$.

Можна перевірити, що функція f має в нулі похідну за будь-яким напрямком $g = (g^1,g^2)$ і $f'(z_0,g) = g^2 = \langle e_2,g \rangle$. Таким чином, функція f має в точці z_0 градієнт, який співпадає з ортом e_2. Відмітимо, що субдиференціал $\partial_{Cl} f(z_0)$ містить градієнт $\nabla f(z_0) = e_2$ а також інші точки.

5.2.6. Властивості субдиференціала Кларка

Теорема 5.2.14. *Нехай* $f = f_1 + f_2$. *Тоді*

$$\partial_{Cl} f(x) \subset \partial_{Cl} f_1(x) + \partial_{Cl} f_2(x).$$

Доведення. Субадитивність верхньої границі показує, що при всіх g виконується нерівність

$$\overline{\lim_{x'\to x, \alpha\downarrow 0}} \frac{1}{\alpha} [\, f(x'+\alpha g) - f(x')\,] \leq$$

$$\leq \overline{\lim_{x'\to x, \alpha\downarrow 0}} \frac{1}{\alpha} (f_1(x'+\alpha g) - f_1(x')) + \overline{\lim_{x'\to x, \alpha\downarrow 0}} \frac{1}{\alpha} (f_2(x'+\alpha g) - f_2(x')),$$

тобто $f_{\underset{Cl}{\uparrow}}(x,g) \leq (f_1)_{\underset{Cl}{\uparrow}}(x,g) + (f_2)_{\underset{Cl}{\uparrow}}(x,g)$.
Використовуючи двоїстість Мінковського, отримаємо

$$\partial_{Cl} f(x) \subset \partial_{Cl} f_1(x) + \partial_{Cl} f_2(x).$$

□

Означення 5.2.5. Локально ліпшицева функція f називається *регулярною в точці* x, якщо вона диференційовна за напрямком в цій точці і її похідна за напрямком $f'(x,g)$ співпадає з похідною Кларка $f_{\underset{Cl}{\uparrow}}(x,g)$.

Наслідок 5.2.3. *Якщо функції* f_1, f_2 *регулярні, то*

$$\partial_{Cl} f(x) = \partial_{Cl} f_1(x) + \partial_{Cl} f_2(x). \tag{5.2.7}$$

Доведення. Дійсно, використовуючи регулярність, маємо

$$f_{\underset{Cl}{\uparrow}}(x,g) \geq f'(x,g) = f_1'(x,g) + f_2'(x,g) = (f_1)_{\underset{Cl}{\uparrow}}(x,g) + (f_2)_{\underset{Cl}{\uparrow}}(x,g),$$

звідси випливає $\partial_{Cl} f(x) \supset \partial_{Cl} f_1(x) + \partial_{Cl} f_2(x)$. З цього включення і теореми 5.2.14 випливає рівність (5.2.7). □

Зауважимо, що гарантувати рівність (5.2.7) без додаткових припущень неможливо. Наведемо відповідний приклад.

Приклад 5.2.8. Нехай f– парна функція ($f(x) = f(-x)$), що визначена на $\mathbb{R}^n$. Покажемо, що верхня похідна Кларка цієї функції в нулі є парною функцією напрямку. Дійсно,

$$f_{\underset{Cl}{\uparrow}}(0,g) = \overline{\lim_{x'\to 0, \alpha\downarrow 0}} \frac{1}{\alpha} [f(x'+\alpha g) - f(x')] =$$

$$= \overline{\lim_{x'\to 0, \alpha\downarrow 0}} \frac{1}{\alpha} [f(-x'+\alpha(-g)) - f(-x')] = f_{\underset{Cl}{\uparrow}}(0, -g).$$

В силу теореми 5.2.6 має місце рівність $f_{\underset{Cl}{\uparrow}}(0,-g)=(-f)_{\underset{Cl}{\uparrow}}(0,g)$. Використовуючи парність похідної $f_{\underset{Cl}{\uparrow}}(0,g)$, перейдемо до рівності

$$f_{\underset{Cl}{\uparrow}}(0,-g)=(-f)_{\underset{Cl}{\uparrow}}(0,g).$$

Зі сказаного випливає, що

$$\partial_{Cl}f(0)+\partial_{Cl}(-f)(0)=2\partial_{Cl}f(0). \tag{5.2.8}$$

В той же час $\partial_{Cl}(f+(-f))(0)=\{0\}$. Якщо, наприклад, $f(x)=\|x\|$, то в правій частині формули (5.2.8) стоїть множина $2B$, де $B-$ одинична куля. В цьому випадку

$$\partial_{Cl}f(0)+\partial_{Cl}(-f)(0)=2B\neq\{0\}=\partial_{Cl}(f+(-f))(0).$$

Теорема 5.2.15. *Нехай відображення $H:\mathbb{R}^n\to\mathbb{R}^m$ неперервно диференційовне в деякому околі $B_\delta(x)$ точки x і, крім того, образ кожного околу цієї точки включає деякий окіл точки $y=Hx$. Припустимо, далі, що ліпшицева функція f визначена на деякій відкритій множині Y, яка містить точку y. Тоді функція*

$$\varphi(x')=f(H(x')),\quad x'\in B_\delta(x)$$

задовольняє умові Ліпшиця в околі $B_\delta(x)$ і

$$\partial_{Cl}\varphi(x)=(H'_x)^*(\partial_{Cl}f(y)),\quad y=H(x).$$

Іншими словами,

$$\partial_{Cl}\varphi(x)=\{l\,|l=(H'_x)^*l',\ l'\in\partial_{Cl}f(y)\}.$$

(*Через $(H'_x)^*$ позначена матриця, транспонована до матриці частинних похідних H'_x відображення H в точці x.*)

Доведення. Оскільки відображення H неперервно диференційовне, то воно задовольняє умову Ліпшиця, а тому функція φ задовольняє умову Ліпшиця. Зафіксуємо напрямок g і покладемо

$$S(x',\alpha)=\frac{1}{\alpha}(H(x'+\alpha g)-H(x'))-H'_x(g).$$

Оскільки відображення H неперервно диференційовне, то

$$S(x',\alpha)\underset{x'\to x,\alpha\downarrow 0}{\longrightarrow}0.$$

Оскільки функція f задовольняє умову Ліпшиця, то маємо

$$\lim_{x'\to x,\alpha\downarrow 0}\frac{1}{\alpha}\left[f(H(x')+\alpha H'_x(g))-f(H(x'+\alpha g))\right]=0. \tag{5.2.9}$$

Використовуючи (5.2.9) і те, що образ кожного околу при відображенні H містить окіл, знайдемо верхню похідну Кларка функції φ

$$\varphi_{\underset{Cl}{\uparrow}}(x,g)=\overline{\lim_{x'\to x,\alpha\downarrow 0}}\frac{1}{\alpha}(f(H(x'+\alpha g))-f(H(x')))=$$

$$= \overline{\lim_{x' \to x, \alpha \downarrow 0}} \frac{1}{\alpha}(f(H(x') + \alpha H'_x(g)) - f(H(x')) =$$

$$\overline{\lim_{y' \to y, \alpha \downarrow 0}} \frac{1}{\alpha}(f(y' + \alpha H'_x(g)) - f(y')) = f_{\underset{Cl}{\uparrow}}(y, H'_x(g)). \tag{5.2.10}$$

З (5.2.10) випливає, що похідна $\varphi_{\underset{Cl}{\uparrow}}(x,g)$ при всіх g співпадає з функцією $q(g) = p(Ag)$, де $p(\nu) = f_{\underset{Cl}{\uparrow}}(y,\nu)$, $A = H'_x$. Функція q сублінійна. Покажемо, що її субдиференціал $\underline{\partial} q$ співпадає з множиною

$$A^*(\underline{\partial} p) = \{l' \,|l' = A^* l, l \in \underline{\partial} p\}.$$

Дійсно, множина $A^*(\underline{\partial} p)$ опукла і компактна. Крім того

$$\max_{l' \in A^*(\underline{\partial} p)} \langle l', g \rangle = \max_{l \in \underline{\partial} p} \langle A^* l, g \rangle = \max_{l \in \underline{\partial} p} \langle l, Ag \rangle = p(Ag).$$

Тому рівність $\underline{\partial} q = A^*(\underline{\partial} p)$ справедлива в силу двоїстості Мінковського. Для завершення доведення залишилося зауважити, що

$$\underline{\partial} p = \partial_{Cl} f(y), \quad \underline{\partial} q = \partial_{Cl} \varphi(x).$$

$\square$

Зауваження 5.2.1. З доведення теореми 5.2.15 випливає, що для будь якого неперервно диференційовного відображення H справедливе включення

$$\partial_{Cl} \varphi(x) \subset (H'_x)^*(\partial_{Cl} f(y)).$$

Рівність $\partial_{Cl} \varphi(x) = (H'_x)^*(\partial_{Cl} f(y))$ виконується не тільки в умовах теореми, але й у тому випадку, коли функція f регулярна, тобто $f_{\underset{Cl}{\uparrow}}(y,g) = f'(y,g)$ при всіх g. Щоб встановити це, слід використати співвідношення $\varphi'(x,g) = f'(H(x), H'_x(g))$, яке справедливе в силу теореми 4.1.5

Як наслідок маємо такі твердження.

Теорема 5.2.16. *Нехай $f_i, i = 1, \ldots, n$ – ліпшицеві функції в деякому околі $B_\delta(x)$ точки x та нехай $f = \max\{f_i : i = 1, \ldots, n\}$. Тоді*

$$\partial_{Cl} f(x) \subset conv\, \{\partial_{Cl} f_i(x) : i \in I(x)\},$$

де $I(x) = \{i : f(x) = f_i(x)\}$. Якщо функції $f_i, i = 1, \ldots, n$ регулярні в точці x, то функція $f = \max\{f_i : i = 1, \ldots, n\}$ регулярна в точці x і

$$\partial_{Cl} f(x) = conv\, \{\partial_{Cl} f_i(x) : i \in I(x)\}.$$

Теорема 5.2.17. *Нехай f_1, f_2 – ліпшицеві функції в деякому околі $B_\delta(x)$ точки x. Тоді функція $(f_1 \cdot f_2)(x)$ ліпшицева в околі точки x і*

$$\partial_{Cl}\, (f_1 \cdot f_2)\, (x) \subset f_2(x)\, \partial_{Cl} f_1(x) + f_1(x)\, \partial_{Cl} f_2(x).$$

Якщо, крім того, $f_1(x) \geq 0, f_2(x) \geq 0$ і функції f_1, f_2 регулярні в точці x, то функція $f = f_1 \cdot f_2$ регулярна в точці x і

$$\partial_{Cl}\, (f_1 \cdot f_2)\, (x) = f_2(x)\, \partial_{Cl} f_1(x) + f_1(x)\, \partial_{Cl} f_2(x).$$

Теорема 5.2.18. *Нехай f_1,f_2 – ліпшицеві функції в деякому околі $B_\delta(x)$ точки x, причому $f_2(x) \neq 0$. Тоді функція $(f_1/f_2)(x)$ ліпшицева в околі точки x і*

$$\partial_{Cl}\left(\frac{f_1}{f_2}\right)(x) \subset \frac{f_2(x)\,\partial_{Cl}f_1(x) - f_1(x)\,\partial_{Cl}f_2(x)}{f_2^2(x)}.$$

Якщо, крім того, $f_1(x) \geq 0,f_2(x) > 0$ і функції f_1,f_2 регулярні в точці x, то функція $f = f_1/f_2$ регулярна в точці x і

$$\partial_{Cl}\left(\frac{f_1}{f_2}\right)(x) = \frac{f_2(x)\,\partial_{Cl}f_1(x) - f_1(x)\,\partial_{Cl}f_2(x)}{f_2^2(x)}.$$

Наведемо теорему про середнє для похідних Кларка. Нам знадобляться наступні необхідні умови екстремума.

Теорема 5.2.19. *Якщо визначена на відкритій множині X локально ліпшицева функція f має локальний екстремум в точці $x \in X$, то* $0 \in \partial_{Cl}f(x)$.

Доведення. Нехай x – точка локального мінімуму функції f, тобто $f(x') \geq f(x)$ для всіх достатньо близьких до x точок x'. Тоді при будь-якому g і достатньо малих $\alpha > 0$ виконується нерівність $f(x + \alpha g) - f(x) \geq 0$, а тому

$$f_{\underset{Cl}{\uparrow}}(x,g) = \overline{\lim_{x' \to x, \alpha \downarrow 0}}\ \frac{1}{\alpha}(f(x' + \alpha g) - f(x')) \geq$$

$$\geq \overline{\lim_{\alpha \downarrow 0}}\ \frac{1}{\alpha}(f(x + \alpha g) - f(x)) \geq 0.$$

Оскільки сублінійна функція $g \to f_{\underset{Cl}{\uparrow}}(x,g)$ невід'ємна, то 0 входить в її субдиференціал $\partial_{Cl}f(x)$.

Якщо x– точка локального максимуму функції f, то аналогічно перевіряється, що $f_{\underset{Cl}{\downarrow}}(x,g) \leq 0$ при всіх g, звідси також випливає, що $0 \in \partial_{Cl}f(x)$. □

Теорема 5.2.20. *Нехай X– відкрита множина і нехай точки x^1,x^2 містяться в X разом з відрізком, що їх з'єднує. Нехай f– локально ліпшицева функція, що визначена на X. Тоді знайдеться таке $\theta \in$* $(0,1)$, *що*

$$f_{\underset{Cl}{\downarrow}}(x_\theta,x^2 - x^1) \leq f(x^2) - f(x^1) \leq f_{\underset{Cl}{\uparrow}}(x_\theta,x^2 - x^1).$$

Доведення. Нехай $x_t = x^1 + t(x^2 - x^1)$. Покладемо

$$g(t) = f(x_t), \quad h(t) = g(t) + t\left[f(x^1) - f(x^2)\right].$$

Оскільки $h(0) = h(1)$, то в деякій точці θ інтервала $(0,1)$ функція h досягає екстремуму на цьому інтервалі. З теореми 5.2.15 випливає, що $0 \in \partial_{Cl} h(\theta)$, тобто

$$h_{\underset{Cl}{\downarrow}}(\theta,v) \le 0 \le h_{\underset{Cl}{\uparrow}}(\theta,v) \tag{5.2.11}$$

при всіх $v \in \mathbb{R}^1$. Оцінимо похідні Кларка функції h. Для цього оцінимо спочатку похідні Кларка функції g. Маємо

$$g_{\underset{Cl}{\uparrow}}(\theta,v) = \overline{\lim_{t\to\theta,\alpha\downarrow 0}} \frac{1}{\alpha}(g(t+\alpha v) - g(t)) =$$

$$= \overline{\lim_{t\to\theta,\alpha\downarrow 0}} \frac{1}{\alpha}(f(x^1 + (t+\alpha v)(x^2 - x^1)) - f(x^1 + t(x^2 - x^1))) \le$$

$$\le \overline{\lim_{x'\to x_\theta,\alpha\downarrow 0}} \frac{1}{\alpha}(f(x' + \alpha v(x^2 - x^1)) - f(x')) = f_{\underset{Cl}{\uparrow}}(x_\theta, v(x^2 - x^1)).$$

Аналогічно отримаємо

$$g_{\underset{Cl}{\downarrow}}(\theta,v) \ge f_{\underset{Cl}{\downarrow}}(x_\theta, v(x^2 - x^1)).$$

Звідси випливає, що

$$h_{\underset{Cl}{\uparrow}}(\theta,v) = g_{\underset{Cl}{\uparrow}}(\theta,v) + v\left[f(x^1) - f(x^2)\right] \le$$

$$\le f_{\underset{Cl}{\uparrow}}(x_\theta, v(x^2 - x^1)) + v\left[f(x^1) - f(x^2)\right], \tag{5.2.12}$$

$$h_{\underset{Cl}{\downarrow}}(\theta,v) \ge f_{\underset{Cl}{\downarrow}}(x_\theta, v(x^2 - x^1)) + v\left[f(x^1) - f(x^2)\right]. \tag{5.2.13}$$

Покладемо в (5.2.12) і (5.2.13) $v = 1$ і, використовуючи формулу (5.2.11), переконаємося в справедливості теореми. $\square$

Зауваження 5.2.2. За умов теореми знайдеться таке $\theta \in (0,1)$ і такий елемент $w \in \partial_{Cl} f(x_\theta)$, що

$$f(x^2) - f(x^1) = \langle w, x^2 - x^1 \rangle.$$

Дійсно, як w тут виступає точка $\alpha w^1 + (1-\alpha)w^2$, де w^1 і w^2– елементи субдиференціала Кларка $\partial_{Cl} f(x_\theta)$, що задовольняють умови:

$$f_{\underset{Cl}{\uparrow}}(x_\theta, x^2 - x^1) = \langle w^1, x^2 - x^1 \rangle,$$

$$f_{\underset{Cl}{\downarrow}}(x_\theta, x^2 - x^1) = \langle w^2, x^2 - x^1 \rangle,$$

а число α знаходиться з рівності

$$f(x^2) - f(x^1) = \alpha f_{\underset{Cl}{\uparrow}}(x_\theta, x^2 - x^1) + (1-\alpha) f_{\underset{Cl}{\downarrow}}(x_\theta, x^2 - x^1).$$

5.3. ДОТИЧНИЙ КОНУС КЛАРКА

Конуси $K(x,S)$ та $\Gamma(x,S)$, які апроксимують множину S поблизу точки $x \in \operatorname{cl} S$, можуть бути описані за допомогою похідних Діні функції ρ_S, де $\rho_S(x)$ – відстань від точки x до множини S. З теореми 4.3.7 випливає, що

$$K(x,S) = \{g \,|\rho'_S(x,g) = 0\} = \left\{ g \,\middle|\, (\rho_S)^{\uparrow}_{D}(x,g) = 0 \right\}.$$

За аналогією з конусом дотичних напрямків $K(x,S)$, доцільно розглянути множину

$$T(x,S) = \left\{ g \,\middle|\, (\rho_S)^{\uparrow}_{Cl}(x,g) = 0 \right\}, \tag{5.3.1}$$

де $x \in \operatorname{cl} S$. Насамперед покажемо, що функція ρ_S задовольняє умові Ліпшиця і тому говорити про її похідні Кларка коректно.

Теорема 5.3.1. *Для $x,y \in \mathbb{R}^n$ виконується нерівність*

$$|\rho_S(x) - \rho_S(y)| \le \|x - y\|\,.$$

Доведення. Зафіксуємо довільне число $\varepsilon > 0$. Нехай елемент $z \in S$ такий, що $\rho_S(y) \ge \|y - z\| - \varepsilon$. Маємо

$$\rho_S(x) \le \|x - z\| \le \|x - y\| + \|y - z\| \le \|x - y\| + \rho_S(y) + \varepsilon.$$

Звідси, оскільки ε – довільне число, отримаємо $\rho_S(x) - \rho_S(y) \le \|x - y\|$. Аналогічно доводиться нерівність $\rho_S(y) - \rho_S(x) \le \|x - y\|$. □

Оскільки $(\rho_S)^{\uparrow}_{Cl}(x,\lambda g) = \lambda(\rho_S)^{\uparrow}_{Cl}(x,g)$ при $\lambda \ge 0$ (це випливає з визначення похідної Кларка), то множина $T(x,S)$, що визначена рівністю (5.3.1), є конусом. Вона називається *дотичним конусом Кларка* або просто *конусом Кларка*.

Теорема 5.3.2. *Наступні умови еквівалентні:*
(а) елемент g належить конусу $T(x,S)$;
(б) для довільного $\varepsilon > 0$ існує таке $\delta > 0$, що при всіх $x' \in B_\delta(x)$ і $\alpha \in (0,\delta)$ існує елемент $g' \in B_\varepsilon(g)$, такий, що виконується включення $x' + \alpha g' \in S$;
(в) для довільних послідовностей $\{x'_k\}$ і $\{\alpha_k\}$ таких, що $x'_k \to x, \alpha_k \to \alpha$, існує послідовність $\{g_k\}$, для якої $g_k \to g, x'_k + \alpha_k g_k \in S$.

Доведення. Твердження (б) і (в) – це переформулювання на мові "$\varepsilon - \delta$" і на мові послідовностей рівності

$$\overline{\lim_{x' \to x, \alpha \downarrow 0}} \inf_{y \in S} \frac{1}{\alpha} \|y - (x' + \alpha g)\| = 0, \tag{5.3.2}$$

яка є розшифровкою співвідношення $(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) = 0$, що визначає конус Кларка. Перевіримо, наприклад, імплікацію (а)$\Rightarrow$ (б). Нехай $g \in T(x,S)$, тобто виконується (5.3.1). Оскільки під знаком $\overline{\lim}$ в (5.3.2) стоїть невід'ємна функція, то при довільному $\varepsilon > 0$ знайдеться таке $\delta > 0$, що $\inf\limits_{y \in S} \frac{1}{\alpha} \|y - (x' + \alpha g)\| < \varepsilon$ якщо $x' \in \mathcal{B}_\delta(x)$ та $\alpha \in (0,\delta)$; іншими словами, для всіх $x' \in \mathcal{B}_\delta(x)$ та $\alpha \in (0,\delta)$ знайдеться елемент $y' \in S$ такий, що справджується нерівність $\left\| \frac{1}{\alpha}(y' - x') - g \right\| < \varepsilon$. Покладемо $g' = \frac{1}{\alpha}(y' - x')$. З одного боку отримаємо нерівність $\|g' - g\| < \varepsilon$, а з іншого боку – включення $x' + \alpha g' = y' \in S$. □

Оскільки $x \in \mathrm{cl}\, S$, то

$$(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) = \overline{\lim_{x' \to x, \alpha \downarrow 0}} \frac{1}{\alpha} \rho_S(x' + \alpha g) \geq 0.$$

Це означає, що співвідношення $(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) = 0$ та $(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) \leq 0$ еквівалентні. Тим самим конус $T(x,S)$ можна записати у вигляді

$$T(x,S) = \quad \{g \mid (\rho_S)_{\underset{Cl}{\uparrow}}(x,g) \leq 0 \}. \tag{5.3.3}$$

Теорема 5.3.3. *Конус Кларка $T(x,S)$ опуклий і замкнутий.*

Доведення. Оскільки функція $g \to (\rho_S)_{\underset{Cl}{\uparrow}}(x,g)$ сублінійна, то множина $\{g \mid (\rho_S)_{\underset{Cl}{\uparrow}}(x,g) \leq 0 \}$ опукла і замкнута. Для завершення доведення потрібно використати формулу (5.3.3). □

На відміну від дотичного конуса Кларка $T(x,S)$ конус дотичних напрямків $K(x,S)$ і конус Булігана $\Gamma(x,S)$ не завжди опуклі. Проте конуси $K(x,S)$ і $\Gamma(x,S)$ визначають локальну апроксимацію множини S поблизу точки x, в той час як конус Кларка може і не мати ніякого відношення до множини S поблизу цієї точки. Наведемо відповідний приклад.

Приклад 5.3.1. Нехай множина S в просторі $\mathbb{R}^2$ – це об'єднання від'ємної частини осі абсцис і бісектриси першого координатного кута. Можна перевірити, що в точці $x = (0,0)$ виконується рівність

$$K(x,S) = \Gamma(x,S) = S, \quad T(x,S) = \{(0,0)\}.$$

Цей приклад показує, що ціна, яку потрібно платити за опуклість, іноді занадто висока.

Далі аналіз конуса Кларка базується на означенні нижньої границі багатозначного відображення. Нехай X– деяка підмножина скінченного простору, a – відображення X в 2^Y, де Y – скінченновимірний простір. Під *нижньою границею відображення* a в точці $x \in \mathrm{cl}\, X$ розуміють множину $\lim\limits_{x'\to x}\inf a(x') = \varliminf\limits_{x'\to x} a(x')$, яка складається з усіх елементів g, які мають наступну властивість: для довільного $\varepsilon > 0$ існує таке $\delta > 0$, що $\rho(g,a(x')) \le \varepsilon$, якщо $x' \in \mathcal{B}_\delta(x) \cap X$.

Теорема 5.3.4. *Нехай S – деяка множина в $\mathbb{R}^n$. Для $\alpha > 0$, $x \in \mathbb{R}^n$ покладемо*

$$a(x,\alpha) = \frac{1}{\alpha}(S - x).$$

Тоді

$$T(x,S) = \lim_{x'\to x}\inf_{\alpha\downarrow 0}\, a(x',\alpha).$$

Доведення. Включення $g \in \lim\limits_{x'\to x}\inf\limits_{\alpha\downarrow 0} a(x',\alpha)$ еквівалентно такому: для довільного $\varepsilon > 0$ існує таке $\delta > 0$, що при всіх $x' \in B_\delta(x)$ і $\alpha \in (0,\delta)$ виконується співвідношення $\rho(g,\frac{1}{\alpha}(S - x')) \le \varepsilon$. Остання нерівність означає, що знайдеться елемент

$$g' \in \frac{1}{\alpha}(S - x'), \tag{5.3.4}$$

для якого $\|g - g'\| \le \varepsilon$. Переписавши співвідношення (5.3.4) у вигляді $x' + \alpha g' \in S$ і використовуючи твердження 5.3.2, переконаємося в справедливості твердження. $\square$

Дотичний конус Кларка $T(x,S)$ може бути представлений як нижня границя конусів Булігана $\Gamma(x,S)$. При цьому припускається, що множина S замкнута, а багатозначне відображення $x \to \Gamma(x,S)$ визначене на S.

Дамо строге формулювання відповідного твердження.

Теорема 5.3.5. *Нехай S– замкнута множина в $\mathbb{R}^n$, $x \in S$. Тоді*

$$T(x,S) = \liminf_{x'\to x} \Gamma(x',S).$$

Доведення цієї теореми спирається на ряд лем.

Лема 5.3.1. *Справедливе включення*

$$T(x,S) \subset \liminf_{x'\to x} \Gamma(x',S).$$

Доведення. Нехай $g \in T(x,S)$. Тоді для кожного $\varepsilon > 0$ знайдеться $\delta > 0$, що має таку властивість: для довільних $x' \in \mathcal{B}_\delta(x)$ і $\alpha \in (0,\delta)$ знайдеться такий елемент $g'(x',\alpha)$, що $\|g'(x',\alpha) - g\| < \varepsilon$, $x' + \alpha g'(x',\alpha) \in S$. Зафіксуємо $x' \in B_\delta(x) \cap S$ і виберемо деяку послідовність $\alpha_k \downarrow 0$. Покладемо $g'_k = g'(x',\alpha_k)$. Нехай існує $\lim g'_k = g'$. Оскільки $x' + \alpha_k g'_k \in S$, то $g' \in \Gamma(x',S)$. В той же час $\|g' - g\| < \varepsilon$. Отже $\rho(g',\Gamma(x',S)) < \varepsilon$. Це означає, шо $g \in \liminf\limits_{x'\to x} \Gamma(x',S)$. $\square$

Для $v \notin S$ покладемо

$$\pi(v) = \{w \in S \colon \|w - v\| = \rho_S(v)\}\,.$$

Через $\Delta(w,S)$ позначимо замкнуту опуклу оболонку конуса Булінга $\Gamma(w,S)$.

Лема 5.3.2. *Якщо $w \in \pi(v),g \in \Delta(w,S)$, то $\langle v - w,g\rangle \leq 0$.*

Доведення. Елемент w є розв'язком наступної екстремальної задачі: знайти мінімум функції $f_v(x) = \|v - x\|^2$ на множині S. Згідно з необхідними умовами екстремуму в точці w має виконуватися співвідношення $(f_v)'(w,g) \geq 0$ для всіх $g \in \Gamma(w,S)$. Оскільки

$$(f_v)'(w,g) = 2(w - v,g),$$

то $(w - v,g) \geq 0$ для всіх $g \in \Gamma(w,S)$. Звідси випливає твердження леми.

□

Зафіксуємо $x \in S$ і $g \in \mathbb{R}^n$. Розглянемо функцію

$$\varphi(\alpha) = \frac{1}{2}\rho_S^2(x + \alpha g), \quad \alpha \geq 0 \tag{5.3.5}$$

Оскільки функція ρ_S ліпшицева, то і функція φ ліпшицева, тому вона майже всюди диференційовна.

Лема 5.3.3. *Якщо в точці $\alpha > 0$ функція $\varphi(\alpha)$, що визначена формулою (5.3.5), диференційовна, то*

$$\varphi'(\alpha) \leq \rho_S(x + \alpha g) \cdot \rho(g,\Delta(z,S)),$$

де z– довільний елемент множини $\pi(x + \alpha g)$.

Доведення. Покладемо $x + \alpha g = y$. Маємо

$$\varphi'(\alpha) = \lim_{\beta \to 0} \frac{1}{\beta}(\varphi(\alpha + \beta) - \varphi(\alpha)) = \lim_{\beta \to 0} \frac{1}{2\beta}(\rho_S^2(y + \beta g) - \rho_S^2(y)).$$

Нехай $z \in \pi(y)$. Тоді $\rho_S(y) = \|y - z\|\,,\ \rho_S(y + \beta g) \leq \|y + \beta g - z\|$. Тому

$$\varphi'(\alpha) = \lim_{\beta \to 0} \frac{1}{2\beta}(\|y + \beta g - z\|^2 - \|y - z\|^2) = \frac{1}{2}(f_z)'(y,g) = (y - z,g').$$

Тут, як і в лемі 5.3.2, $f_z(y) = \|z - y\|^2$. Нехай $g' \in \Delta(z,S)$. Тоді, як випливає з леми 5.3.2, $(y - z,g') \leq 0$ і тому $\varphi'(\alpha) \leq (y - z,g - g')$. Оскільки

$$(y - z,g - g') \leq \|y - z\| \cdot \|g - g'\| = \rho_S(y)\,\|g - g'\|\,,$$

то

$$\varphi'(\alpha) \leq \rho_S(y) \inf_{g' \in \Delta(z,S)} \|g - g'\| = \rho_S(y)\,\rho(g,\Delta(z,S)).$$

□

Лема 5.3.4. *Справедливе включення*

$$\liminf_{x' \to x} \Delta(x',S) \subset T(x,S).$$

Доведення. Нехай $g \in \liminf\limits_{x' \to x} \Delta(x',S)$. Тоді для довільного $\varepsilon > 0$ знайдеться таке $\delta > 0$, що для $x' \in B_\delta(x) \cap S$ виконується нерівність

$$\rho(g,\Delta(x',S)) \leq \varepsilon \tag{5.3.6}$$

Нехай числа $\eta > 0, \alpha_0 > 0$ такі, що

$$2(\eta + \alpha_0 \|g\|) < \delta. \tag{5.3.7}$$

Тоді для $x' \in B_\eta(x), \quad \alpha \in (0,\alpha_0)$ виконується включення

$$\pi(x' + \alpha g) \subset \mathcal{B}_\delta(x). \tag{5.3.8}$$

Дійсно, візьмемо $x' \in B_\eta(x), \alpha \in (0,\alpha_0)$. Маємо

$$\rho_S(x' + \alpha g) \leq \|x' + \alpha g - x\| \leq \|x' - x\| + \alpha \|g\| < \eta + \alpha_0 \|g\|.$$

Тому для $z_\alpha \in \pi(x' + \alpha g)$ виконуються нерівності

$$\|z_\alpha - x\| = \|z_\alpha - (x' + \alpha g) + (x' - x) + \alpha g\| \leq$$
$$\leq \|z_\alpha - (x' + \alpha g)\| + \|(x' - x)\| + \alpha \|g\| =$$
$$= \rho_S(x' + \alpha g) + \|(x' - x)\| + \alpha \|g\| \leq 2(\eta + \alpha_0 \|g\|),$$

звідси і випливає (5.3.8). Нехай для чисел η і α_0 виконується нерівність (5.3.7), $x' \in B_\eta(x) \cap S$. Розглянемо при $\alpha \in (0,\alpha_0)$ функцію $\varphi(\alpha) = \frac{1}{2}\rho_S^2(x' + \alpha g)$. Нехай як і вище, $z_\alpha \in \pi(x' + \alpha g)$. З леми 5.3.3 випливає, що в тих точках α, де існує похідна $\varphi'(\alpha)$, виконується нерівність

$$\varphi'(\alpha) \leq \rho_S(x' + \alpha g)\, \rho(g,\Delta(z_\varepsilon,S)).$$

В силу (5.3.8) $z_\alpha \in B_\delta(x)$, тому з нерівності (5.3.6) випливає співвідношення $\rho(g,\Delta(z_\alpha,S) \leq \varepsilon$. В той же час, оскільки $x' \in S$, то $\rho_S(x' + \alpha g) \leq \|(x' + \alpha g) - x'\| \leq \alpha \|g\|$. Отже $\varphi'(\alpha) \leq \varepsilon\alpha \|g\|$. Оскільки $\varphi(0) = 0$, то

$$\varphi(\alpha) = \varphi(\alpha) - \varphi(0) = \int\limits_0^\alpha \varphi'(u)du \leq \varepsilon \|g\| \frac{\alpha^2}{2}.$$

Ми показали, що при $x' \in B_\eta(x), \alpha \in (0,\alpha_0)$ виконується нерівність $\rho_S^2(x' + \alpha g) \leq \varepsilon \|g\| \alpha^2$. Звідси

$$\rho_S(x' + \alpha g) - \rho_S(x') \leq \rho_S(x' + \alpha g) \leq \alpha\sqrt{\varepsilon \|g\|}. \tag{5.3.9}$$

Таким чином,

$$(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) = \overline{\lim_{x' \to x, \alpha \downarrow 0}} \frac{1}{\alpha}(\rho_S(x' + \alpha g) - \rho_S(x')) \leq \sqrt{\varepsilon \|g\|}.$$

В силу довільності ε

$$(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) = 0.$$

За визначенням конуса Кларка це означає, що $g \in T(x,S)$. $\square$

Зауваження 5.3.1. Нехай Q – множина повної міри, яка належить деякому околу точки x. Тоді

$$\lim_{x' \to x} \inf_{x' \in Q} \Delta(x',S) \subset T(x,S). \tag{5.3.10}$$

Доведення. Візьмемо $\varepsilon > 0$ і найдемо число $\delta > 0$ так, щоб для $x' \in B_\delta(x) \cap S$ виконувалась нерівність (5.3.6). Потім за δ знайдемо числа $\eta > 0$ і $\alpha_0 > 0$, при яких справедливе (5.3.7). Використовуючи доведення леми 5.3.4, отримаємо, що для $x' \in B_\eta(x) \cap Q$ виконується нерівність (5.3.9). З цієї нерівності, завдяки тому, що ε довільне маємо співвідношення $(\rho_S)_{\underset{D}{\uparrow}}(x',g) = 0$. Оскільки в силу теореми 5.2.12.

$$(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) = \overline{\lim_{x' \to x, x' \in Q}} (\rho_S)_{\underset{D}{\uparrow}}(x',g),$$

то $(\rho_S)_{\underset{Cl}{\uparrow}}(x,g) = 0$. Звідси і випливає включення (5.3.10). □

Доведення теореми 5.3.5 безпосередньо випливає з лем 5.3.1 і 5.3.4, якщо врахувати включення $\Delta(x,S) \supset \Gamma(x,S)$.

Зауваження 5.3.2. Нехай, як і в зауваженні 5.3.1, Q– множина повної міри, яка належить деякому околу точки x. Тоді

$$\lim_{x' \to x} \inf_{\Gamma}(x,S) = \lim_{x' \to x} \inf_{\Delta}(x,S) = \lim_{x' \to x} \inf_{x' \in Q} \Delta(x,S).$$

Це випливає з леми 5.3.4 і зауваження 5.3.1.

Багатозначне відображення a, яке визначене на множині X, називається *напівнеперервним знизу*, якщо $\lim\limits_{x' \to x} \inf a(x') = a(x)$. З теореми 5.3.5 випливає такий наслідок.

Наслідок 5.3.1. *Рівність $T(x,S) = \Gamma(x,S)$ справедлива тоді і тільки тоді, коли багатозначне відображення $x' \to \Gamma(x',S)$ напівнеперервне знизу в точці x.*

5.4. УМОВИ МІНІМУМУ СУБДИФЕРЕНЦІЙОВНОЇ ФУНКЦІЇ

Нехай функція f задана і субдиференційовна на відкритій множині $X \subset \mathbb{R}^n$, тобто в кожній точці $x \in X$ функція f диференційовна за напрямками, причому існує опуклий компакт $\partial f\,(x) \subset \mathbb{R}^n$ такий, що

$$f'\,(x,g) = \max_{v \in \partial f(x)} \langle v,g \rangle \tag{5.4.1}$$

Розглянемо задачу мінімізації функції f на замкнутій множині $S \subset X$. З леми 4.5.1 випливає справедливість наступного твердження.

Лема 5.4.1. *Для того, щоб у точці $x^* \in S$ функція f досягала свого найменшого на S значення, необхідно, щоб*

$$\max_{v \in \partial f(x^*)} \langle v,g \rangle \geq 0 \;\; \forall g \in \Gamma\,(x^*,S) \tag{5.4.2}$$

Зафіксуємо $x \in S$. Нехай $A \subset \Gamma(x,S)$ — такий замкнутий опуклий конус, що $0_n \in A$, і нехай $\mathfrak{U}(x)$— сімейство опуклих конусів таких, що

$$A \subset \Gamma(x,S) \quad \forall A \in \mathfrak{U}(x), \quad \bigcup_{A \in \mathfrak{U}(x)} A = \Gamma(x,S). \qquad (5.4.3)$$

Завжди можна знайти сімейство $\mathfrak{U}(x)$, що задовольняє (5.4.3). Наприклад, можна взяти

$$\mathfrak{U}(x) = \{l \in \mathbb{R}^n \,|l = \{g = \lambda g_0 \,|\, \lambda > 0\},\ g_0 \in \Gamma(x,S),\ \|g_0\| = 1\}.$$

Через A^+ позначимо спряжений конус до конуса A:

Лема 5.4.2. *Умова (5.4.2) еквівалентна умові*

$$\partial f(x^*) \cap A^+ \neq \emptyset \quad \forall A \in \mathfrak{U}(x^*). \qquad (5.4.4)$$

Доведення. Припустимо протилежне. Нехай

$$\partial f(x^*) \cap A^+ = \emptyset. \qquad (5.4.5)$$

Знайдемо

$$\min_{w \in A^+,\, v \in \partial f(x^*)} \|w - v\| = \|w_0 - v_0\| = a. \qquad (5.4.6)$$

Очевидно, що в силу (5.4.5) маємо $a > 0$. Покладемо $g_0 = \frac{w_0 - v_0}{\|w_0 - v_0\|}$. З (5.4.6) маємо

$$\max_{v \in \partial f(x^*)} \langle g_0, v\rangle = -a < 0, \qquad (5.4.7)$$

$$\langle g_0, q\rangle \geq 0 \quad \forall q \in A^+. \qquad (5.4.8)$$

З (5.4.8) випливає $g_0 \in A^{++} = A$, тобто $g_0 \in \Gamma(x^*,S)$. Тепер, як і при доведенні леми 4.5.2, знаходимо такі точки $x_k \in S$, що $f(x_k) < f(x^*)$, що суперечить тому, що x^* — точка мінімуму. □

Наслідок 5.4.1. *Якщо $x^* \in int\, S$, то умова (5.4.4) має вигляд*

$$0 \in \partial f(x^*). \qquad (5.4.9)$$

Припустимо, що $x \in S$ не є стаціонарною точкою функції f на S. Тоді знайдеться такий конус $A \in \mathfrak{U}(x)$, що $\partial f(x) \cap A^+ = \emptyset$. Знайдемо

$$\min_{w \in A^+,\, v \in \partial f(x)} \|v - w\| = \|v_A(x) - w_A(x)\| = a_A(x). \qquad (5.4.10)$$

Очевидно, що $a_A(x) > 0$.

Оскільки A^+ і $\partial f(x)$— опуклі множини, то вектор $v_A(x) - w_A(x)$ єдиний (хоча може існувати ціле сімейство пар точок $[v_A(x), w_A(x)]$, які задовольняють (5.4.10)). Неважко бачити, що напрямок

$$g_A(x) = \frac{1}{a_A(x)} (w_A(x) - v_A(x))$$

– це напрямок спуску функції f на множині S, тобто

$$g_A(x) \in \Gamma(x), \qquad \max_{v \in \partial f(x)} \langle v, g_A(x)\rangle = -a_A(x) < 0.$$

Знайдемо

$$\sup_{A\in\mathfrak{U}(x)} a_A(x) = a(x). \tag{5.4.11}$$

Можна показати, що супремум в (5.4.10) досягається. Нехай він досягається на $A(x)$. Тоді напрямок

$$g_0(x) = \frac{1}{a(x)}\left(w_{A(x)}(x) - \upsilon_{A(x)}(x)\right)$$

є напрямком найшвидшого спуску функції f на множині S у точці x. Оскільки елемент $A(x)$ не обов'язково єдиний, то і напрямків найшвидшого спуску може бути декілька.

Зауваження 5.4.1. Якщо $x \in$int S, то із (5.4.9) одержимо, що напрямок

$$g(x) = -\frac{z(x)}{\|z(x)\|},$$

єдиний н. н. с.

Таким чином, перевірка необхідної умови мінімуму зводиться до розв'язання задачі математичного програмування (5.4.10). Кількість цих задач залежить від кількості множин A в сімействі $\mathfrak{U}(x)$. Звичайно, ми зацікавлені в тому, щоб їхнє число було якнайменше. Якщо, наприклад, $\Gamma(x,S)$ — опуклий конус, то за $\mathfrak{U}(x)$ можна взяти множину конусів, що містить лише конус $\Gamma(x,S)$. Якщо конус $\Gamma(x,S)$ опуклий (і точка $x \in S$ не є стаціонарною), то існує єдиний напрямок найшвидшого спуску. Це пов'язано з тим, що в (5.4.9) норма евклідова. Якщо взяти, наприклад, max-норму: $||x||_m = \max\limits_{i\in 1:n} |x^{(i)}|$, де $x = (x^1, \ldots, x^n)$, то н. н. с. може бути багато. Проте множина напрямків найшвидшого спуску є опуклою.

Розглянемо випадок, коли f— гладка функція. Тоді умова (5.4.4) еквівалентна умові

$$f'(x^*) \in \mathfrak{L}(x^*), \tag{5.4.12}$$

де $f'(x)$ - градієнт функції f в точці x, $\mathfrak{L}(x) = \bigcap\limits_{a\in\mathfrak{U}(x)} A^+$. Якщо, зокрема, виявилося, що $\mathfrak{L}(x^*) = \{0\}$, то умова (5.4.12) має вигляд $f'(x^*) = 0$.

Приклад 5.4.1. Нехай $x = (x^1,x^2) \in \mathbb{R}^2$, $x_0 = (0,0)$, $S = S_1 \cup S_2$, де $S_1 = \mathbb{R}^2_+ = \{x \in \mathbb{R}^2 | x^1 \geq 0, x^2 \geq 0\}, S_2 = \{x = (-\alpha, -\alpha) | \alpha \in [0,1]\}$. Очевидно, що

$$\Gamma(x_0) = A_1 \cup A_2,$$

де $A_1 = \mathbb{R}^2_+, A_2 = \{x = (-\alpha, -\alpha) | \alpha \geq 0\}$. Ясно, що $A_1^+ = A_1$, $A_2^+ = \{x \in \mathbb{R}^2 | (x,l) \geq 0\}, l = (-1, -1)$. Маємо $\mathfrak{L}(x_0) = A_1^+ \cap A_2^+ = \{(0,0)\}$.
Тому. для того, щоб в точці $x_0 = (0,0)$ гладка функція f досягала найменшого значення, необхідно, щоб $f'(x_0) = 0$.

Порівняємо необхідні умови (5.4.4) і (5.4.12). Вони еквівалентні. Отже, для гладкої функції перевіряти точку на екстремум можна за допомогою співвідношення (5.4.4) і за допомогою співвідношення (5.4.12). Вище вже було показано, як за допомогою (5.4.4) можна знайти напрямок (чи напрямки, якщо їх багато) найшвидшого спуску функції f на S. Чи можна це зробити за допомогою умови (5.4.12), діючи так само, як у випадку (5.4.4)? Знайдемо

$$\min_{\upsilon \in \mathfrak{L}(x)} \|\upsilon - f'(x)\| = \|\upsilon(x) - f'(x)\|.$$

Однак напрямок

$$g = \frac{\upsilon(x) - f'(x)}{||\upsilon(x) - f'(x)||}$$

не має нічого спільного з напрямком найшвидшого спуску. Він взагалі може не належати конусу можливих напрямків.

Приклад 5.4.2. Нехай $S \in \mathbb{R}^2$, $x_0 = (0{,}0)$, $S = S_1 \cup S_2$, де

$$S_1 = \{x \in \mathbb{R}^2 |\, x = (x^1{,}0), x^1 \in [0{,}1]\},$$

$$S_2 = \{x \in \mathbb{R}^2 |\, x = (0{,}x^2), x^2 \in [0{,}1)\}.$$

Зрозуміло, що $\Gamma(x_0) = \mathrm{A}_1 \cup \mathrm{A}_2$, де

$$\mathrm{A}_1 = \{x \in \mathbb{R}^2 |\, x = \left(x^1{,}0\right), x^1 \geq 0\},$$

$$\mathrm{A}_2 = \{x \in \mathbb{R}^2 |\, x = \left(0{,}x^2\right), x^2 \geq 0\}.$$

Маємо

$$\mathrm{A}_1^+ = \left\{x \in \mathbb{R}^2 |\, (x{,}l_1) \geq 0, \quad l_1 = (1{,}0)\right\},$$

тобто

$$\mathrm{A}_1^+ = \{x = \left(x^1{,}x^2\right) |x^1 \geq 0\}.$$

$$\mathrm{A}_2^+ = \left\{x \in \mathbb{R}^2 |\, (x{,}l_2) \geq 0, \quad l_2 = (0{,}1)\right\},$$

тобто

$$\mathrm{A}_2^+ = \{x = \left(x^1{,}x^2\right) |x^2 \geq 0\}.$$

Тоді

$$\mathfrak{L}\left(x_0\right) = \mathrm{A}_1^+ \cap \mathrm{A}_2^+ = \mathbb{R}_+^2 = \{x = \left(x^1{,}x^2\right) |x^1 \geq 0{,}x^2 \geq 0\}.$$

Візьмемо функцію $f\left(x\right) = -x^1 - x^2$. Ясно, що $f'\left(x_0\right) = (-1, -1)$, і умова (5.4.12) не виконана. Знайдемо (див. рис. 5.4.1)

$$\min_{\upsilon \in \Im(x_0)} ||\upsilon - f'\left(x_0\right)|| = || - f'\left(x_0\right)||.$$

Напрямок

$$g = -f'\left(x_0\right) / ||f'\left(x_0\right)|| = \left(1 \Big/ \sqrt{2}{,}1 \Big/ \sqrt{2}\right)$$

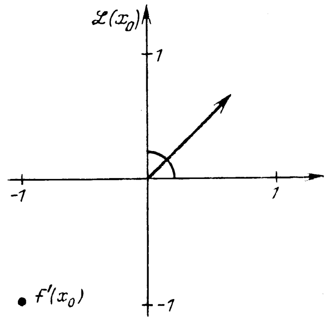


Рис. 5.4.1:

не належить конусу $\Gamma(x_0)$. В той же час, із умови (5.4.4) знайдемо

$$\min_{v \in A_1^+} ||f'(x_0) - v|| = ||f'(x_0) - v_1||, \quad \text{де} \quad v_1 = (0, -1),$$

$$\min_{v \in A_2^+} ||f'(x_0) - v|| = ||f'(x_0) - v_2||, \quad \text{де} \quad v_2 = (-1,0).$$

Звідси маємо два напрямки

$$g_1 = \frac{v_1 - f'(x_0)}{||v_1 - f'(x_0)||} = (1,0), \quad g_2 = \frac{v_2 - f'(x_0)}{||v_2 - f'(x_0)||} = (0,1).$$

При цьому

$$f'(x_0,g_1) = (f'(x_0), g_1) = -1, \quad f'(x_0,g_2) = (f'(x_0), g_2) = -1,$$

тобто обидва напрямки є н. н. с. За допомогою умови (5.4.12) цього факту нам встановити не вдалося.

Отже, не кожна необхідна умова зручна не тільки для перевірки того, чи є точка стаціонарною, але і для знаходження напрямків найшвидшого спуску.

Вище було показано, як за допомогою (5.4.4) знайти н. н. с. субдиференційовної функції f на множині S в точці $x \in S$. Цей напрямок $g(x)$ належить конусу можливих напрямків $\Gamma(x)$. Однак він може виявитися недопустимим, тобто може виявитися, що $x + \alpha g(x) \notin S \quad \forall \alpha > 0$.

Лема 5.4.3. *Якщо* $A \subset \mathbb{R}^n$*– опуклий конус і*

$$\mathit{int}\, A \neq \emptyset, \tag{5.4.13}$$

$B \subset \mathbb{R}^n$*– опуклий компакт,* $\eta > 0$ *– довільне число, то умова*

$$B \cap A^+ \neq \emptyset \tag{5.4.14}$$

еквівалентна умові

$$0 \in \operatorname{conv}\{\mathrm{B} \cup \mathrm{T}_\eta(\mathrm{A})\}, \tag{5.4.15}$$

де

$$\mathrm{T}_\eta(\mathrm{A}) = \{v \in \mathbb{R}^n | v \in \left[-\mathrm{A}^+\right], \quad ||v|| = \eta\}.$$

Доведення. Спочатку покажемо, що з (5.4.14) випливає (5.4.15). Якщо має місце (5.4.14), то існує таке v, що $v \in \mathrm{B}, v \in \mathrm{A}^+$, тому $v_1 = \beta v \in [-\mathrm{A}^+]$, де $\beta = -2/||v||$. Складемо опуклу комбінацію

$$S_\alpha = \alpha v + (1-\alpha) v_1 = \alpha v + (1-\alpha)\beta v = ((1-\beta)\alpha + \beta) v.$$

При

$$\alpha_0 = -\frac{\beta}{1-\beta} = \frac{\eta/||v||}{1+\eta/||v||}$$

буде $S_{\alpha_0} = 0$. При цьому $\alpha_0 \in (0,1)$, тобто $S_\alpha \in \{\mathrm{B} \cup \mathrm{T}_\eta(\mathrm{A})\}$.
Нехай тепер виконано (5.4.15). Тоді знайдуться такі $\alpha \in [0,1]$, $v_1 \in \mathrm{B}$ і $v_2 \in \mathrm{T}_\eta(\mathrm{A})$, що

$$\alpha v_1 + (1-\alpha) v_2 = 0_n. \tag{5.4.16}$$

Покажемо, що $\alpha \neq 0$. Припустимо, що це не так. Тоді $v_2 = 0$, тобто

$$0 \in \operatorname{conv} \mathrm{T}_\eta(\mathrm{A}). \tag{5.4.17}$$

За припущенням має місце умова (5.4.13).
Нехай $S \in \operatorname{int}\mathrm{A}$, тобто існує $r > 0$ таке, що $B_r(S) \subset \mathrm{A}$. Нехай $\sin\gamma = r/||S||$ (випадок $||S|| = 0$ тривіальний).
Тоді $\Gamma = \operatorname{cone} B_r(S) \subset \mathrm{A}$. Оскільки $\Gamma^+ \supset A^+$, то

$$\mathrm{T}_\eta(\mathrm{A}) \subset \mathrm{T}_\eta(\Gamma). \tag{5.4.18}$$

Але з рис. 5.4.2 видно, що

$$||v|| \geq \eta \sin\gamma \quad \forall v \in \mathrm{T}_\eta(\Gamma).$$

Звідси. і з (5.4.18) випливає, що $0 \notin \mathrm{T}_\eta(\mathrm{A})$, а це суперечить (5.4.17). Отже, в (5.4.16) $\alpha \neq 0$. Тоді

$$v_1 = -\frac{1-\alpha}{\alpha} v_2. \tag{5.4.19}$$

Але $v_1 \in \mathrm{B}$, а $v_2 \in \mathrm{T}_\eta(\mathrm{A})$, тому $v = -\frac{1-\alpha}{\alpha} v_2 \in \mathrm{A}^+$. Отже, з (5.4.19) випливає (5.4.14). □

Зауваження 5.4.2. Умова (5.4.13) суттєва, що видно з наступного прикладу.

Приклад 5.4.3. Нехай $x = \left(x^{(1)}, x^{(2)}\right) \in \mathbb{R}^2$, $\mathrm{A} = \{x \in \mathbb{R}^2 | \, x = \left(x^1, 0\right), x^1 \geq 0\}$. Тоді (див. рис. 5.4.2) $\inf \mathrm{A} = \emptyset$, $\mathrm{A}^+ = \{v = (v^{(1)}, v^{(2)}) | v^{(1)} \geq 0\}$. Оскільки $0 \in \operatorname{conv} \mathrm{T}_\eta(\mathrm{A}) \quad \forall \eta > 0$, то умова (5.4.15) справджується для будь-якого $B \subset \mathbb{R}^n$.

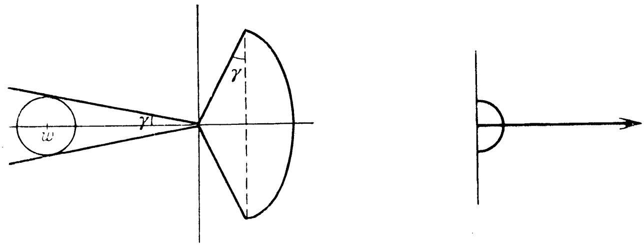


Рис. 5.4.2:

Наслідок 5.4.2. *Якщо*

$$int\,\mathrm{A} \neq \emptyset \quad \forall \mathrm{A} \in \mathfrak{U}\,(x^*)\,, \tag{5.4.20}$$

то умова (5.4.4) еквівалентна умові

$$0 \in \mathrm{conv}\,\{\partial f\,(x^*) \cup \mathrm{T}_\eta\,(\mathrm{A})\} \quad \forall \mathrm{A} \in U\,(x^*)\,. \tag{5.4.21}$$

Тут $\eta > 0$ *— довільне (але фіксоване) число.*

Нехай $x \in S$ не є стаціонарною точкою, тобто умова (5.4.20) (чи, еквівалентна умова (5.4.4)) не виконана. Для кожного $\mathrm{A} \in \mathfrak{U}\,(x)$ знайдемо $\min\limits_{v \in C(\mathrm{A})} ||v|| = ||v_{\mathrm{A}}\,(x)\,||$, де $C\,(\mathrm{A}) = \mathrm{conv}\,\{\partial f\,(x) \cap \mathrm{T}_\eta\,(\mathrm{A})\}$. Нехай

$$\max_{\mathrm{A} \in \mathfrak{U}(x)} ||v_{\mathrm{A}}\,(x)\,|| = ||v_{\mathrm{A}(x)}\,(x)\,||.$$

Тоді напрямок

$$g_\eta\,(x) = -v_{\mathrm{A}(x)}\,(x)/||v_{\mathrm{A}(x)}\,(x)\,|| \tag{5.4.22}$$

є допустимий напрямок спуску, тобто

$$(v, g_\eta\,(x)) \leq -||v_{\mathrm{A}(x)}\,(x)\,|| \quad \forall v \in \partial f\,(x)\,, \tag{5.4.23}$$

$$g_\eta\,(x) \in \mathrm{int}\,\mathrm{A}\,(x) \subset \mathrm{int}\,\Gamma\,(x)\,. \tag{5.4.24}$$

В ряді випадків (наприклад, множина S задається за допомогою нерівностей) з умов (5.4.23), (5.4.24) випливає, що існує таке $\alpha_0 > 0$, що $x + \alpha g_\eta \in S \quad \forall \alpha \in [0, \alpha_0]$.

При чисельному розв'язуванні задач мінімізації умову (5.4.21) використовувати зручніше, ніж умову (5.4.14). Відзначимо, що у тому випадку коли $\mathfrak{U}(x)$ містить більш ніж одну множину A, напрямок $g_\eta\,(x)$ може виявитися неєдиним.

Зауваження 5.4.3. Неважко показати, що для $g_\eta\,(x)$ (див. (5.4.22)) має місце твердження: якщо

$$g_{\eta_k}(x) \underset{\eta_k \to \infty}{\longrightarrow} g(x), \tag{5.4.25}$$

то $g(x)$ — напрямок найшвидшого спуску функції f на S в точці x. В силу неєдиності $g_\eta(x)$ границя в (5.4.25) теж може бути неєдиною.

Зауваження 5.4.4. Нехай множина S визначається за допомогою нерівностей
$$S = \{x \in \mathbb{R}^n |\, h_i(x) \le 0 \quad \forall i \in I\},$$
де $h_i \in C^1$,$I = \{1,\ldots,N\}$. Якщо $x \in S$, $\max_{i\in I} h_i(x) = 0$ і
$$0 \notin \mathrm{conv}\,\{h'_i(x)|\, i \in Q(x)\}, \tag{5.4.26}$$
де $Q(x) = \{i \in I|\, h_i(x) = 0\}$, то $\Gamma(x)$— опуклий конус, при цьому
$$(\Gamma(x))^+ = \Gamma^+(x) = \mathrm{cone}\,\{-h'_i(x)|\, i \in Q(x)\}. \tag{5.4.27}$$
В цьому випадку умова (5.4.21) еквівалентна умові
$$0 \in \mathrm{conv}\,\left\{\{\partial f(x)\} \cup \{h'_i(x)|\, i \in Q(x)\}\right\}. \tag{5.4.28}$$
Умова (5.4.26) називається *умовою Слейтера.*

Зауваження 5.4.5. Якщо (див. зауваження 5.4.4) функції f і h_i опуклі на $\mathbb{R}^n$ і має місце умова Слейтера (5.4.26), то умова (5.4.28) є достатньою умовою мінімуму.

Зауваження 5.4.6. Очевидно, умова (5.4.14) еквівалентна умові
$$0 \in [\mathrm{B} - \mathrm{A}^+]. \tag{5.4.29}$$
Більш того, якщо (5.4.14) не виконана, то при розв'язанні задач
$$\min_{\upsilon\in \mathrm{B},\, w\in \mathrm{A}^+} ||\upsilon - w|| = ||\upsilon_0 - w_0||,$$
$$\min_{z\in[\mathrm{B}-\mathrm{A}^+]} ||z|| = ||z_0||$$
одержуємо один і той самий вектор $z_0 = \upsilon_0 - w_0$.

Аналогічно, (5.4.4) еквівалентна умові
$$0 \in [\partial f(x^*) - \mathrm{A}^+] \quad \forall \mathrm{A} \in \mathfrak{U}(x^*),$$
і для знаходження напрямків найшвидшого спуску потрібно знайти
$$\max_{\mathrm{A}\in\mathfrak{U}(x_0)} \min_{z\in[\partial f(x_0)-\mathrm{A}^+]} ||z||.$$

Лема 5.4.4. *Достатня умова мінімуму*
$$\max_{\upsilon\in\partial f(x^*)} \langle \upsilon, g\rangle > 0 \quad \forall g \in \Gamma(x^*), \quad g \ne 0_n, \tag{5.4.30}$$
за умов (5.4.3) еквівалентна умові
$$0 \in \mathit{int}\,[\partial f(x^*) - \mathrm{A}^+] \quad \forall \mathrm{A} \in \mathfrak{U}(x^*). \tag{5.4.31}$$

Розділ VI
Квазідиференційовні функції.

6.1. КВАЗІДИФЕРЕНЦІАЛЬНЕ ЧИСЛЕННЯ

6.1.1. Означення та приклади квазідиференційовних функцій

Нехай функція f визначена на відкритій множині X у просторі $\mathbb{R}^n$ і має похідну за напрямками $f'(x,g)$ у точці $x \in X$. Нагадаємо, що функція f називається субдиференційовною у точці x, якщо $f'(x,g)$ є сублінійною функцією або, що те ж саме, якщо знайдеться такий опуклий компакт U, що

$$f'(x,g) = \max_{h \in U} \langle h,g \rangle, \quad g \in \mathbb{R}^n. \tag{6.1.1}$$

Форма (6.1.1) дуже зручна для вивчення похідної. Її можна розглядати як свого роду лінеаризацію похідної, вираз її через лінійні функції. Разом із субдиференційовними можна розглядати супердиференційовні функції, тобто функції, похідні $f'(x,g)$ яких можна записати так:

$$f'(x,g) = \min_{h \in V} \langle h,g \rangle, \quad g \in \mathbb{R}^n,$$

де V – опуклий компакт. Зрозуміло, що функція f супердиференційовна тоді і тільки тоді, коли функція $f_1(x) = -f(x)$ субдиференційовна.

Клас субдиференційовних функцій досить широкий, зокрема, як відзначалося у попередньому розділі, якщо похідна $f'(x,g)$ напівнеперервна зверху за x при кожному g, то функція $f(x)$ субдиференційовна. Сума і максимум (поточковий супремум) скінченної кількості субдиференційовних функцій також є субдиференційовними функціями. У той же час множина субдиференційовних функцій не є лінійним простором. Ця обставина наводить на думку розглянути множину функцій, у яких похідна може бути записана у вигляді різниці двох сублінійних функцій або суми сублінійної і суперлінійної функцій. Сукупність таких функцій вже є лінійним простором.

Означення 6.1.1. Визначена на відкритій множині X у просторі $\mathbb{R}^n$ функція $f(x)$ називається *квазідиференційовною* в точці $x \in X$, якщо вона диференційовна за напрямками в цій точці і її похідну $f'(x,g)$ можна подати у вигляді

$$f'(x,g) = \max_{h\in U}\langle h,g\rangle + \min_{h\in V}\langle h,g\rangle, \quad g \in \mathbb{R}^n, \tag{6.1.2}$$

де U,V – опуклі компактні множини в $\mathbb{R}^n$.

Покладемо

$$\underline{f}'(x,g) = \max_{h\in U}\langle h,g\rangle, \quad \overline{f}'(x,g) = \min_{h\in V}\langle h,g\rangle, \quad l(g) = \underline{f}'(x,g) + \overline{f}'(x,g).$$

Функція $\underline{f}'(x,g)$ сублінійна, а функція $\overline{f}'(x,g)$ – суперлінійна. Тому похідна $f'(x,g) = l(g)$ є елементом простору L функцій, які можна подати у вигляді суми сублінійної та суперлінійної функції. Пара опуклих компактів $[U,V]$, що використовується у формулі (6.1.2) для запису похідної визначена неєдиним чином. Оскільки похідна $f'(x,g)$ є елементом простору L, то їй відповідає елемент простору опуклих множин, тобто клас еквівалентних пар опуклих компактів. Пари $[U_1,V_1]$ і $[U_2,V_2]$, де V_i, U_i, $i = 1,2$, – опуклі компакти називаються еквівалентними, якщо $U_1 - V_2 = U_2 - V_1$.

Означення 6.1.2. Нехай функція f квазідиференційовна в точці x. Клас еквівалентних пар опуклих компактів $[U,V]$, які мають властивість

$$f'(x,g) = \max_{h\in U}\langle h,g\rangle + \min_{h\in V}\langle h,g\rangle \quad \forall g \in \mathbb{R}^n,$$

називається *квазідиференціалом функції f в точці x* і позначається символом $\mathfrak{D}f(x)$. Кожну пару множин, що належить цьому класу, також будемо називати квазідиференціалом і позначатимемо тим же символом $\mathfrak{D}f(x)$. Якщо $\mathfrak{D}f(x) = [U,V]$, то множину U назвемо *субдиференціалом функції f у точці x* і позначатимемо символом $\underline{\partial} f(x)$, а множину V назвемо *супердиференціалом функції f в точці x* і позначатимемо через $\overline{\partial} f(x)$. Отже, $\mathfrak{D}f(x) = [\underline{\partial} f(x),\overline{\partial} f(x)]$.

Підкреслимо, що множини $\underline{\partial} f(x)$ та $\overline{\partial} f(x)$ не можна розглядати окремо. Сенс має лише пара $\mathfrak{D}f(x) = [\underline{\partial} f(x),\overline{\partial} f(x)]$ як представник класу еквівалентних пар простору опуклих множин.

Квазідиференційовність функції f у точці x означає, що функція f має похідну за напрямками (похідну Діні) у цій точці, і вказана похідна може бути виражена через лінійні функції (квазілінеаризована) за допомогою квазідиференціала $\mathfrak{D}f(x) = [\underline{\partial} f(x),\overline{\partial} f(x)]$ за формулою

$$f'(x,g) = \max_{h\in\underline{\partial} f(x)}\langle h,g\rangle + \min_{h\in\overline{\partial} f(x)}\langle h,g\rangle.$$

У тому випадку, коли функція f диференційовна в точці x за Адамаром і квазідиференційовна, будемо говорити, що *f квазідиференційовна за Адамаром*.

Зауважимо, що квазідиференційовна локально ліпшицева функція квазідиференційовна за Адамаром.

Нехай функція f диференційовна за напрямками на відкритій множині X. Якщо функцію f можна представити у вигляді суми двох функцій, причому похідна однієї з них напівнеперервна знизу, а похідна іншої - напівнеперервна зверху при кожному напрямку g, то функція f квазідиференційовна на X (тобто в кожній точці $x \in X$). Припустимо, що квазідиференціал функції f у точці $x \in X$ невироджений в тому сенсі, що не існує квазідиференціала вигляду $[\underline{\partial} f(x),\{0\}]$ або $[\{0\},\overline{\partial} f(x)]$, тобто функція f не є субдиференційовною чи супердиференційовною. Тоді $f'(x,g)$ не є напівнеперервною зверху чи знизу. Точніше кажучи, принаймі при одному $g \in \mathbb{R}^n$ функція $x \to f'(x,g)$ не є напівнеперервною зверху чи напівнеперервною знизу в точці x.

Субдиференційовність (супердиференційовність) квазідиференційовної функції f рівносильна тому, що ця функція має квазідиференціал вигляду $\mathfrak{D} f(x) = [\underline{\partial} f(x),\{0\}]$ (вигляду $\mathfrak{D} f(x) = [\{0\},\overline{\partial} f(x)]$).

Клас квазідиференційовних функцій досить широкий. Приведемо декілька прикладів таких функцій.

Приклад 6.1.1. Якщо функція f має градієнт у точці x, то вона квазідиференційовна в цій точці. Її квазідиференціалами є, наприклад, пари $[\nabla f(x),\{0\}]$ чи $[\{0\},\nabla f(x)]$. Таким чином, функція f є одночасно і субдиференційовною і супердиференційовною.

Приклад 6.1.2. Нехай f – визначена на відкритій множині $X \subset \mathbb{R}^n$ опукла функція. Тоді функція f диференційовна за напрямками у точці $x \in X$, причому

$$f'(x,g) = \max_{h \in \underline{\partial} f(x)} \langle h,g \rangle,$$

де

$$\underline{\partial} f(x) = \{v \in \mathbb{R}^n \mid f(z) - f(x) \geq \langle v,z-x \rangle, \ \forall z \in X\}$$

субдиференціал функції f у точці x. Звідси випливає, що функція f - квазідиференційовна в точці x. Як квазідиференціал функції f в цій точці можна взяти пару $\mathfrak{D} f(x) = [\underline{\partial} f(x),\{0\}]$. Це означає, що опукла функція субдиференційовна.

Аналогічно показується, що визначена на відкритій множині $X \subset \mathbb{R}^n$ угнута функція f є квазідиференційовною (і навіть супердиференційовною) в точках $x \in X$. Як квазідиференціал можна взяти пару $\mathfrak{D} f(x) = [\{0\},\overline{\partial} f(x)]$, де $\overline{\partial} f(x)$ – супердиференціал функції f

$$\overline{\partial} f(x) = \{w \in \mathbb{R}^n \mid f(z) - f(x) \leq \langle w,z-x \rangle, \ \forall z \in X\}.$$

Приклад 6.1.3. Нехай X – відкрита множина в $\mathbb{R}^n$, Y – компакт в $\mathbb{R}^m$, функція $\varphi(x,y)$ визначена на $X \times Y$ і неперервна разом зі своєю частинною похідною $\varphi'_x(x,y)$. Покладемо

$$f_1(x) = \max_{y \in Y} \varphi(x,y), \ \ f_2(x) = \min_{y \in Y} \varphi(x,y).$$

Оскільки

$$f_1'(x,g) = \max_{y \in R(x)} \langle \varphi_x'(x,y),g \rangle, \quad f_2'(x,g) = \min_{x \in Q(x)} \langle \varphi_x'(x,y),g \rangle,$$

де $R(x) = \{y \in G | \varphi(x,y) = f_1(x)\}$, $Q(x) = \{y \in G | \varphi(x,y) = f_2(x)\}$, то функції f_1 і f_2 квазідиференційовні в точці $x \in X$, причому як квазідиференціали виступають відповідно пари

$$\mathfrak{D} f_1(x) = [\underline{\partial} f_1(x),\{0\}], \quad \mathfrak{D} f_2(x) = [\{0\},\overline{\partial} f_2(x)],$$

де

$$\underline{\partial} f_1(x) = \operatorname{conv} \{v \in \mathbb{R}^n | v = \varphi_x'(x,y),\ y \in R(x)\},$$
$$\overline{\partial} f_2(x) = \operatorname{conv} \{w \in \mathbb{R}^m | w = \varphi_x'(x,y),\ y \in Q(x)\}.$$

Звідси випливає, що функція максимуму $f_1(x)$ субдиференційовна, а функція мінімуму $f_2(x)$ супердиференційовна на X.

6.1.2. Основні формули квазідиференційного числення

Вкажемо властивості квазідиференційовних функцій, які можна трактувати як квазідиференціальне числення. При цьому алгебраїчні операції над парами опуклих компактів (у тому числі і над квазідиференціалами) будемо розуміти так:

$$[U_1,V_1] + [U_2,V_2] = [U_1 + U_2,V_1 + V_2],$$

$$\lambda[U,V] = \begin{cases} [\lambda U,\lambda V], & \lambda \geq 0, \\ [\lambda V,\lambda U], & \lambda \leq 0. \end{cases}$$

Теорема 6.1.1. 1). *Нехай функції $f_1(x)$ і $f_2(x)$ квазідиференційовні в точці x. Тоді сума і добуток цих функцій квазідиференційовні в цій точці. При цьому*

$$\mathfrak{D}(f_1 + f_2)(x) = \mathfrak{D} f_1(x) + \mathfrak{D} f_2(x), \tag{6.1.3}$$
$$\mathfrak{D}(f_1 \cdot f_2)(x) = f_1(x)\mathfrak{D} f_2(x) + f_2(x)\mathfrak{D} f_1(x). \tag{6.1.4}$$

Іншими словами, якщо $[\underline{\partial} f_1(x),\overline{\partial} f_1(x)]$, $[\underline{\partial} f_2(x),\overline{\partial} f_2(x)]$ - квазідиференціали функцій f_1 і f_2 в точці x, то квазідиференціали $[\underline{\partial}(f_1{+}f_2)(x),\overline{\partial}(f_1{+}f_2)(x)]$, $[\underline{\partial}(f_1 \cdot f_2)(x),\overline{\partial}(f_1 \cdot f_2)(x)]$ функцій $f_1 + f_2$ та $f_1 \cdot f_2$ обчислюються за формулами

$$\underline{\partial}(f_1 + f_2)(x) = \underline{\partial} f_1(x) + \underline{\partial} f_2(x),$$

$$\overline{\partial}(f_1 + f_2)(x) = \overline{\partial} f_1(x) + \overline{\partial} f_2(x),$$

$$\underline{\partial}(f_1\cdot f_2)(x)=\begin{cases} f_1(x)\underline{\partial}f_2(x)+f_2(x)\underline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\geq 0, f_2(x)\geq 0,\\ f_1(x)\overline{\partial}f_2(x)+f_2(x)\underline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\leq 0, f_2(x)\geq 0,\\ f_1(x)\overline{\partial}f_2(x)+f_2(x)\overline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\leq 0, f_2(x)\leq 0,\\ f_1(x)\underline{\partial}f_2(x)+f_2(x)\overline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\geq 0, f_2(x)\leq 0,\end{cases}$$

$$\overline{\partial}(f_1\cdot f_2)(x)=\begin{cases} f_1(x)\overline{\partial}f_2(x)+f_2(x)\overline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\geq 0, f_2(x)\geq 0,\\ f_1(x)\underline{\partial}f_2(x)+f_2(x)\overline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\leq 0, f_2(x)\geq 0,\\ f_1(x)\underline{\partial}f_2(x)+f_2(x)\underline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\leq 0, f_2(x)\leq 0,\\ f_1(x)\overline{\partial}f_2(x)+f_2(x)\underline{\partial}f_1(x), & \text{якщо}\quad f_1(x)\geq 0, f_2(x)\leq 0.\end{cases}$$

2). *Нехай функція $f(x)$ квазідиференційовна в точці x. Тоді при будь-якому дійсному λ функція $\lambda f(x)$ також квазідиференційовна в цій точці, причому*

$$\mathfrak{D}(\lambda f)(x)=\lambda\mathfrak{D}f(x). \tag{6.1.5}$$

Іншими словами, якщо $[\underline{\partial}f(x),\overline{\partial}f(x)]$ – квазідиференціал функції $f(x)$ в точці x, то для квазідиференціала $[\underline{\partial}(\lambda f)(x),\overline{\partial}(\lambda f)(x)]$ функції $\lambda f(x)$ в цій точці справедлива рівність

$$\underline{\partial}(\lambda f)(x)=\begin{cases}\lambda\underline{\partial}f(x), & \lambda\geq 0,\\ \lambda\overline{\partial}f(x), & \lambda\leq 0,\end{cases}$$

$$\overline{\partial}(\lambda f)(x)=\begin{cases}\lambda\overline{\partial}f(x), & \lambda\geq 0,\\ \lambda\underline{\partial}f(x), & \lambda\leq 0.\end{cases}$$

3). *Нехай функція $f(x)$ квазідиференційовна в точці x, причому $f(x)\neq 0$. Тоді функція $(\frac{1}{f})(x)=\frac{1}{f(x)}$ квазідиференційовна в точці x, і*

$$\mathfrak{D}\left(\frac{1}{f}\right)(x)=-\frac{1}{f^2(x)}\mathfrak{D}f(x). \tag{6.1.6}$$

Іншими словами, якщо $[\underline{\partial}f(x),\overline{\partial}f(x)]$– квазідиференціал функції $f(x)$ в точці x, то для квазідиференціала $[\underline{\partial}\left(\frac{1}{f}\right)(x),\overline{\partial}\left(\frac{1}{f}\right)(x)]$ функції $\frac{1}{f}$ в точці x виконуються рівності

$$\underline{\partial}\left(\frac{1}{f}\right)(x)=-\frac{1}{f^2(x)}\overline{\partial}f(x),\quad \overline{\partial}\left(\frac{1}{f}\right)(x)=-\frac{1}{f^2(x)}\underline{\partial}f(x).$$

Доведення. Спочатку відмітимо, що функція $f_1(x)+f_2,(x)$ $f_1(x)\cdot f_2(x)$, $\lambda f(x)$ і $\frac{1}{f}(x)$ диференційовні в точці x за напрямками, причому для всіх

$g \in \mathbb{R}^n$ виконуються рівності

$$(f_1 + f_2)'(x,g) = f_1'(x,g) + f_2'(x,g),$$
$$(f_1 \cdot f_2)'(x,g) = f_1(x)f_2'(x,g) + f_2(x)f_1'(x,g),$$
$$(\lambda f)'(x,g) = \lambda f'(x,g),$$
$$\left(\frac{1}{f}\right)'(x,g) = -\frac{1}{f^2(x)}f'(x,g).$$

Розглянемо, наприклад, функцію $f_1 \cdot f_2$. Її похідна представляється у вигляді $\mu_1 l_2(g) + \mu_2 l_1(g)$, де числа μ_1, μ_2 співпадають з $f_1(x)$, $f_2(x)$ відповідно, а функції $l_1(g)$, $l_2(g)$ – з похідними $f_1'(x,g)$, $f_2'(x,g)$. За означенням квазідиференційовності функції $l_i(g)$ $i = 1,2$, є елементами простору L, тобто їх можна представити у вигляді суми сублінійної та суперлінійної функцій. Оскільки L – лінійний простір, то функція $\mu_1 l_2(g) + \mu_2 l_1(g)$ також входить в L. Це означає, що $f_1 \cdot f_2$ – квазідиференційовна функція. Виразимо її квазідиференціал $\mathfrak{D}(f_1 \cdot f_2)(x)$ через квазідиференціали $\mathfrak{D}f_i(x)$ функцій $f_i(x)$, $i = 1,2$. Нехай ψ – відображення, яке ставить у відповідність функції $l \in L$ елемент $\alpha = [U,V]$ простору опуклих множин при якому справедлива рівність

$$l(g) = \max_{h \in U}\langle h,g\rangle + \min_{h \in V}\langle h,g\rangle.$$

Тоді $\psi(l_1) = \mathfrak{D}f_1(x)$, $\psi(l_2) = \mathfrak{D}f_2(x)$. Оскільки відображення ψ лінійне, то $\psi(\mu_1 l_2 + \mu_2 l_1) = \mu_1\mathfrak{D}f_2(x) + \mu_2\mathfrak{D}f_1(x)$. В той же час $\psi(\mu_1 l_2 + \mu_2 l_1) = \mathfrak{D}(f_1 \cdot f_2)(x)$. Отже,

$$\mathfrak{D}(f_1 \cdot f_2)(x) = f_1(x)\mathfrak{D}f_2(x) + f_2(x)\mathfrak{D}f_1(x).$$

Тим самим формула (6.1.4) доведена. Аналогічно перевіряються формули (6.1.3), (6.1.5) і (6.1.6). Їх роз'яснення, дане в теоремі, одразу випливає з визначення алгебраїчних операцій у просторі опуклих множин. □

Наслідок 6.1.1. *Якщо функції $f_1(x)$ і $f_2(x)$ квазідиференційовні в точці x, то їх різниця $f_1(x) - f_2(x)$ квазідиференційовна в цій точці. При цьому, як випливає з* (6.1.3) *і* (6.1.5)*,*

$$\mathfrak{D}(f_1 - f_2)(x) = \mathfrak{D}f_1(x) - \mathfrak{D}f_2(x).$$

Якщо $f_2(x) \neq 0$, то частка $\frac{f_1(x)}{f_2(x)}$ квазідиференційовна в точці x, причому

$$\mathfrak{D}\left(\frac{f_1}{f_2}\right)(x) = \frac{1}{f_2^2(x)}\left(f_2(x)\mathfrak{D}f_1(x) - f_1(x)\mathfrak{D}f_2(x)\right).$$

Ця рівність випливає з формул (6.1.4) *і* (6.1.6).

Зауваження 6.1.1. Формули квазідиференціального числення, запропоновані в теоремі 6.1.1 (і нижче), мають наступний зміст: серед квазідиференціалів функції в деякій точці існує квазідиференціал вказаного вигляду (інші квазідиференціали є еквівалентними йому парами).

Перейдемо до операції знаходження мінімуму та максимуму.

Теорема 6.1.2. *Нехай функції* $f_1(x),\ldots,f_m(x)$ *визначені на відкритій множині* X *та квазідиференційовні в точці* $x \in X$. *Нехай*

$$\varphi_1(x) = \max_{i=\overline{1,m}} f_i(x), \quad \varphi_2(x) = \min_{i=\overline{1,m}} f_i(x).$$

Тоді функції $\varphi_1(x)$ *і* $\varphi_2(x)$ *квазідиференційовні в точці* x *і*

$$\mathfrak{D}\varphi_1(x) = [\underline{\partial}\varphi_1(x),\overline{\partial}\varphi_1(x)], \;\; \mathfrak{D}\varphi_2(x) = [\underline{\partial}\varphi_2(x),\overline{\partial}\varphi_2(x)],$$

де

$$\underline{\partial}\varphi_1(x) = \operatorname{conv} \bigcup_{k\in R(x)} \left(\underline{\partial} f_k(x) - \sum_{i\in R(x), i\neq k} \overline{\partial} f_i(x) \right), \tag{6.1.7}$$

$$\overline{\partial}\varphi_1(x) = \sum_{k\in R(x)} \overline{\partial} f_k(x), \quad \underline{\partial}\varphi_2(x) = \sum_{k\in Q(x)} \underline{\partial} f_k(x), \tag{6.1.8}$$

$$\overline{\partial}\varphi_2(x) = \operatorname{conv} \bigcup_{k\in Q(x)} \left(\overline{\partial} f_k(x) - \sum_{i\in Q(x), i\neq k} \underline{\partial} f_i(x) \right). \tag{6.1.9}$$

Тут $[\underline{\partial} f_k(x),\overline{\partial} f_k(x)]$ *квазідиференціал функції* f_k *в точці* x, $R(x) = \{i \in I | f_i(x) = \varphi_1(x)\}$, $Q(x) = \{i \in I | f_i(x) = \varphi_2(x)\}$, $I = \{1,\ldots,m\}$

Доведення. Обмежимося випадком функції φ_1. Ця функція диференційовна за напрямками, причому

$$(\varphi_1)'(x,g) = \max_{i\in R(x)} f_i'(x,g). \tag{6.1.10}$$

Функції $l_i(g) = f_i'(x,g)$ входять у простір L, функція $l(g) = \max\limits_{i\in R(x)} l_i(g)$ також входить у L. Ця обставина разом з формулою (6.1.10) показує, що функція φ_1 квазідиференційовна. Формули (6.1.7) — (6.1.9) випливають з теореми 6.1.1. $\square$

Зауваження 6.1.2. Нижче наводиться інше доведення цієї теореми.

Істотне місце у квазідиференціальному численні займає теорема про квазідиференційовність композиції.

Теорема 6.1.3. *Нехай* X *- відкрита множина в* $\mathbb{R}^n$, Y *- відкрита множина в* $\mathbb{R}^m$, *і нехай відображення* $H(x) = (h_1(x),\ldots,h_m(x))$ *визначене на* X, *приймає значення в* Y, *і його координатні функції* $h_i(x)$ *квазідиференційовні в точці* $x_0 \in X$. *Нехай функція* f *визначена на* Y *і квазідиференційовна за Адамаром в точці* $y_0 = H(x_0)$. *Тоді функція*

$$\varphi(x) = f(H(x))$$

квазідиференційовна в точці x_0. *При цьому квазідиференціал* $\mathfrak{D}\varphi(x_0) = [\underline{\partial}\varphi(x_0),\overline{\partial}\varphi(x_0)]$ *описується наступним чином:*

$$\underline{\partial}\varphi(x_0) = \left\{p|p = \sum_{i=1}^{m}(\nu^{(i)}(\lambda_i + \mu_i) - \underline{\nu}^{(i)}\lambda_i - \overline{\nu}^{(i)}\mu_i),\right.$$

$$\left.\nu = (\nu^{(1)},\ldots,\nu^{(m)}) \in \underline{\partial}f(y_0),\ \lambda_i \in \underline{\partial}h_i(x_0),\ \mu_i \in \overline{\partial}h_i(x_0)\right\}, \qquad (6.1.11)$$

$$\overline{\partial}\varphi(x_0) = \left\{l|l = \sum_{i=1}^{m}(\nu^{(i)}(\lambda_i + \mu_i) + \underline{\nu}^{(i)}\lambda_i + \overline{\nu}^{(i)}\mu_i),\right.$$

$$\left.\nu = (\nu^{(1)},...,\nu^{(m)}) \in \overline{\partial}f(y_0),\ \lambda_i \in \underline{\partial}h_i(x_0),\ \mu_i \in \overline{\partial}h_i(x_0)\right\}. \qquad (6.1.12)$$

Тут $\overline{\nu}$ *і* $\underline{\nu}$ *будь-які вектори, які задовольняють нерівність* $\underline{\nu} \leqslant \nu \leqslant \overline{\nu}$ *для всіх* $\nu \in \underline{\partial}f(y_0) \cup (-\overline{\partial}f(y_0))$.

Зауваження 6.1.3. Пара множин $[\underline{\partial}\varphi(x_0),\overline{\partial}\varphi(x_0)]$, про яку йде мова в теоремі, залежить від векторів $\overline{\nu}$ і $\underline{\nu}$, які визначаються неоднозначно. Вибираючи ці вектори різними способами, будемо одержувати різні пари опуклих компактів. Однак всі отримані таким способом пари еквівалентні між собою і представляють один і той же елемент простору опуклих множин.

Теорема 6.1.3 випливає з теореми 7.4.1.

Наведемо приклади обчислення квазідиференціала композиції.

Приклад 6.1.4. Нехай функції $h_1(x),\ldots,h_m(x), x \in \mathbb{R}^n$, квазідиференційовні в точці $x_0 \in X$, а функція $f(y)$ неперервно диференційовна в точці $y_0 = (h_1(x_0),\ldots,h_m(x_0)) \in \mathbb{R}^m$. Покладемо

$$\varphi(x) = f(h_1(x),\ldots,h_m(x)),\ x \in X,$$

і виразимо квазідиференціал $\mathfrak{D}\varphi(x_0) = [\underline{\partial}\varphi(x_0),\overline{\partial}\varphi(x_0)]$ функції $\varphi(x)$ в точці x_0 через квазідиференціал $\mathfrak{D}h_i(x_0) = [\underline{\partial}h_i(x_0),\overline{\partial}h_i(x_0)]$ функції $h_i(x)$ в точці x_0 і градієнт $\nabla f(y_0) = \left(\frac{\partial f}{\partial y^{(1)}}(y_0),\ldots,\frac{\partial f}{\partial y^{(m)}}(y_0)\right)$ функції f. Вважатимемо, що квазідиференціал цієї функції f в точці y_0 виражається парою $\left[\{\frac{1}{2}\nabla f(y_0)\},\{\frac{1}{2}\nabla f(y_0)\}\right]$. Нехай $I = \{1,2,\ldots,m\}$,

$$I_1 = \{i \in I|\ \frac{\partial f}{\partial y^{(i)}}(y_0) < 0\},\ \ I_2 = \{i \in I|\ \frac{\partial f}{\partial y^{(i)}}(y_0) \geqslant 0\}.$$

Покладемо

$$\underline{\nu}^{(i)} = \begin{cases} +\frac{1}{2}\frac{\partial f}{\partial y^{(i)}}(y_0), & i \in I_1, \\ -\frac{1}{2}\frac{\partial f}{\partial y^{(i)}}(y_0), & i \in I_2, \end{cases}$$

$$\overline{\nu}^{(i)} = \begin{cases} -\frac{1}{2}\frac{\partial f}{\partial y^{(i)}}(y_0), & i \in I_1, \\ +\frac{1}{2}\frac{\partial f}{\partial y^{(i)}}(y_0), & i \in I_2. \end{cases}$$

Використавши теорему 6.1.3, отримаємо

$$\underline{\partial}\varphi(x_0) = \left\{ l | l = \sum_{i \in I_1} \frac{\partial f}{\partial y^{(i)}}(y_0)\mu_i + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y_0)\lambda_i, \right.$$

$$\left. \lambda_i \in \underline{\partial} h_i(x_0), \mu_i \in \overline{\partial} h_i(x_0), i \in I \right\},$$

$$\underline{\partial}\varphi(x_0) = \left\{ l | l = \sum_{i \in I_1} \frac{\partial f}{\partial y^{(i)}}(y_0)\lambda_i + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y_0)\mu_i, \right.$$

$$\left. \lambda_i \in \underline{\partial} h_i(x_0), \mu_i \in \overline{\partial} h_i(x_0), i \in I \right\}.$$

Приклад 6.1.5. Нехай функції $h_1(x), \ldots, h_m(x)$ визначені на відкритій множині $X \subset \mathbb{R}^n$ і супердиференційовні в точці $x_0 \in X$, тобто ці функції диференційовні за напрямками у точці x_0 і для їх похідних $h_i'(x_0,g)$ справедливі представлення

$$h_i'(x_0,g) = \min_{p \in V_i} \langle p, g \rangle,$$

де $V_i, i \in I$, – опуклі компакти. Зрозуміло, що як квазідиференціал $\mathfrak{D}h_i(x_0)$ можна розглядати пару $[\{0\}, V_i]$. Нехай функція f визначена на відкритій множині Y, що містить точку $y_0 = (h_1(x_0), \ldots, h_m(x_0))$, і функція f субдиференційовна у цій точці, іншими словами, існує квазідиференціал $\mathfrak{D}f(x_0)$ вигляду $[U, \{0\}]$. Покладемо

$$\varphi(x) = f(h_1(x), \ldots, h_m(x)) \tag{6.1.13}$$

і знайдемо квазідиференціал $\mathfrak{D}\varphi(x_0)$. Нехай вектор $\overline{\nu} = (\overline{\nu}^{(1)}, \ldots, \overline{\nu}^{(m)}) \in \mathbb{R}^m$ такий, що

$$0 \leqslant \overline{\nu},\ \nu \leqslant \overline{\nu} \quad \forall\ \nu \in U.$$

Застосовуючи теорему 6.1.3, отримаємо

$$\mathfrak{D}\varphi(x_0) = [\underline{\partial}\varphi(x_0), \overline{\partial}\varphi(x_0)],$$

де

$$\underline{\partial}\varphi(x_0) = \left\{ w |\ w = \sum_{i=1}^{m} (\nu^{(i)} - \overline{\nu}^{(i)}) v_i, \right.$$

$$\left. v_i \in V_i,\ \ i \in I,\ \nu = (\nu^{(1)}, \ldots, \nu^{(m)}) \in U \right\}, \tag{6.1.14}$$

$$\overline{\partial}\varphi(x_0) = \left\{ w |\ w = \sum_{i=1}^{m} \overline{\nu}^{(i)} v_i, v_i \in V_i \right\} = \sum_{i=1}^{m} \overline{\nu}^{(i)} V_i. \tag{6.1.15}$$

Зауваження 6.1.4. Нехай компакт U повністю лежить у конусі $\mathbb{R}^m_-$, що складається з векторів з недодатніми компонентами. Тоді як $\overline{\nu}$ можна взяти нульовий вектор. У цьому випадку множина $\overline{\partial}\varphi(x_0)$, що визначена формулою (6.1.15), складається лише з нуля, тобто функція φ

субдиференційовна. Її субдиференціал визначається формулою (6.1.14) при $\overline{\nu}^{(i)} = 0, i = 1,\ldots,m$. Розглянемо випадок, коли f – опукла функція, а $h_1(x),\ldots,h_m(x)$ – угнуті функції. Тоді функція f субдиференційовна, $h_1(x),\ldots,h_m(x)$ супердиференційовні. Квазідиференціал композиції (6.1.13) знаходиться за допомогою формул (6.1.14) і (6.1.15), де $U = \underline{\partial} f(y_0)$ – субдиференціал функції f у точці y_0, а $V_i = \overline{\partial} h_i(x_0)$ – супердиференціал функції h_i в точці x_0.

Приклад 6.1.6. Нехай f – така ж функція, як і в прикладі 6.1.5, а функції $h_1(x),\ldots,h_m(x)$ субдиференційовні. Нехай квазідиференціал $\mathfrak{D}h_i(x_0)$ функції $h_i(x)$ у точці x_0 збігається з парою $[V_i,\{0\}]$. Обчислимо квазідиференціал $\mathfrak{D}\varphi(x_0)$ функції $\varphi(x)$, визначеної формулою (6.1.11), у точці x_0. Для цього розглянемо вектор $\underline{\nu} = (\underline{\nu}^{(1)},\ldots,\underline{\nu}^{(m)}) \in \mathbb{R}^m$, що задовольняє нерівності

$$\underline{\nu} \leqslant 0,\ \underline{\nu} \leqslant \nu \quad \forall\ \nu \in U,$$

де U – опуклий компакт, який має властивість

$$f'(y_0,g) = \max_{p \in U} \langle p,g \rangle.$$

Застосовуючи теорему 6.1.3, одержимо

$$\mathfrak{D}\varphi(x_0) = [\underline{\partial}\varphi(x_0), \overline{\partial}\varphi(x_0)],$$

де

$$\underline{\partial}\varphi(x_0) = \Bigg\{ w \mid w = \sum_{i=1}^{m} (\nu^{(i)} - \underline{\nu}^{(i)}) v_i,$$

$$v_i \in V_i, i = 1,\ldots,m,\ \nu = (\nu^{(1)},\ldots,\nu^{(m)}) \in U \Bigg\}, \tag{6.1.16}$$

$$\overline{\partial}\varphi(x_0) = \sum_{i=1}^{m} \underline{\nu}^{(i)} V_i. \tag{6.1.17}$$

Зауваження 6.1.5. Якщо множина U повністю лежить у конусі $\mathbb{R}^m_+$, який складається з векторів з невід'ємними компонентами, то можна вважати, що $\underline{\nu} = 0$. У цьому випадку функція φ субдиференційовна.

Якщо $f(y)$ і $h_1(x),\ldots,h_m(x)$ – опуклі функції, тоді їх композиція (6.1.13) квазідиференційовна і її квазідиференціал знаходиться за допомогою формул (6.1.16) і (6.1.17), де $V_i = \underline{\partial} h_i(x_0)$, $U = \underline{\partial} f(y_0)$ – субдиференціали відповідних функцій. Якщо, крім того, функція $f(y)$ зростає (тобто нерівність $y_1 \geqslant y_2$ приводить до $f(y_1) \geqslant f(y_2)$), то функція φ опукла. З монотонності $f(y)$ випливає, що субдиференціал $\underline{\partial} f(y_0)$ лежить у конусі $\mathbb{R}^m_+$. Тому можна вважати, що $\underline{\nu} = 0$. У даному випадку субдиференціал $\underline{\partial}\varphi(x_0)$ опуклої функції $\underline{\varphi}(x) = f(h_1(x),\ldots,h_m(x))$ в точці x_0

може бути обчислений за допомогою формули (6.1.14) при $\underline{\nu} = 0$:

$$\underline{\partial}\varphi(x_0) = \left\{ w | w = \sum_{i=1}^{m} \nu^{(i)} v_i,\ v_i \in \underline{\partial} h_i(x_0),\ \nu = (\nu^{(1)}, \ldots, \nu^{(m)}) \in \underline{\partial} f(y_0) \right\}.$$

Використовуючи теорему 6.1.3, виведемо формулу для квазідиференціала функції максимуму. Нехай

$$\varphi(x) = \max\{h_1(x), \ldots, h_m(x)\},$$

де $h_1(x), \ldots, h_m(x)$– квазідиференційовні функції, які визначені на відкритій множині X і мають у точці $x_0 \in X$ квазідиференціал $\mathfrak{D} h_i(x_0) = [\underline{\partial} h_i(x_0), \overline{\partial} h_i(x_0)]$. Представимо функцію φ як композицію відображення $H : X \to \mathbb{R}^m$, визначеного рівністю $H(x) = (h_1(x), \ldots, h_m(x))$ і функції f, де

$$f(y) = \max_{i=1,\ldots,m} y^{(i)},$$

Тут $y = (y^{(1)}, \ldots, y^{(m)}) \in \mathbb{R}^m$. Функція f сублінійна. Її похідна $f'(y,g)$ обчислюється за формулою

$$f'(y,g) = \max_{i \in R(y)} g^{(i)},$$

де $g = (g^{(1)}, \ldots, g^{(m)}) \in \mathbb{R}^m$, $R(y) = \{i \in I |\ y^{(i)} = f(y)\}$. Нехай $y_0 = (h_1(x_0), \ldots, h_m(x_0))$. Легко перевірити, що субдиференціал $\underline{\partial} f(y_0)$ має вигляд

$$\underline{\partial} f(y_0) = \left\{ \nu | \nu = (\nu^{(1)}, \ldots, \nu^{(m)}) : \sum_{i=1}^{m} \nu^{(i)} = 1, \nu_{(i)} \geqslant 0; \nu^{(i)} = 0, i \notin R(y_0) \right\},$$

а квазідиференціал $\mathfrak{D} f(x_0)$ збігається з парою $[\underline{\partial f}(x_0), \{0\}]$.

Знайдемо квазідиференціал $\mathfrak{D}\varphi(x_0) = [\underline{\partial}\varphi(x_0), \overline{\partial}\varphi(x_0)]$ функції φ в точці x_0. З цією метою скористаємося формулами (6.1.11) і (6.1.12), поклавши $\underline{\nu} = 0$, $\overline{\nu} = (\overline{\nu}^{(1)}, \ldots, \overline{\nu}^{(m)})$, де

$$\overline{\nu}^{(i)} = \begin{cases} 0, & \text{якщо } \ i \notin R(y_0), \\ 1, & \text{якщо } \ i \in R(y_0). \end{cases}$$

Маємо

$$\underline{\partial}\varphi(x_0) = \left\{ l |\ l = \sum_{i \in R(y_0)} \nu^{(i)} (\alpha_i + \beta_i) - \sum_{i \in R(y_0)} \beta_i, \right.$$

$$\left. \alpha_i \in \underline{\partial} h_i(x_0),\ \beta_i \in \overline{\partial} h_i(x_0),\ \sum_{i \in R(y_0)} \nu^i = 1,\ \nu^i \geqslant 0,\ \forall\ i \in R(y_0) \right\},$$

$$\overline{\partial}\varphi(x_0) = \sum_{i\in R(y_0)} \overline{\partial}h_i(x_0). \tag{6.1.18}$$

Якщо $w \in \underline{\partial}\varphi(x_0)$, то

$$w = \sum_{i\in R(y_0)} \nu^{(i)}(\alpha_i + \beta_i) - \left(\sum_{i\in R(y_0)} \nu^{(i)}\right)\left(\sum_{k\in R(y_0)} \beta_k\right) =$$

$$= \sum_{i\in R(y_0)} \nu^{(i)}\left(\alpha_i - \sum_{k\in R(y_0)} \beta_k + \beta_i\right) = \sum_{i\in R(y_0)} \nu^{(i)}\left(\alpha_i - \sum_{k\in R(y_0),k\neq i} \beta_k\right)$$

Тут $\nu^{(i)} \geqslant 0$, $\sum_{i\in R(y_0)} \nu^{(i)} = 1$, $\alpha_i \in \underline{\partial}h_i(x_0)$, $\beta_i \in \overline{\partial}h_i(x_0)$. Зі сказаного випливає, що множина $\underline{\partial}\varphi(x_0)$ може бути представлена у вигляді

$$\underline{\partial}\varphi(x_0) = \operatorname{conv} \bigcup_{i\in R(y_0)} \left(\underline{\partial}h_i(x_0) - \sum_{k\in R(y_0),k\neq i} \overline{\partial}h_k(x_0)\right). \tag{6.1.19}$$

Формули (6.1.18) і (6.1.19) збігаються з формулами для обчислення квазідиференціала функції максимуму, наведеними в теоремі 6.1.2.

6.1.3. Приклади обчислення квазідиференціалів

Приклад 6.1.7. Нехай $f(x) = |x|$, $x \in \mathbb{R}^1$, $x_0 = 0$. Маємо

$$f(x) = \max\{\varphi_1(x),\varphi_2(x)\} = \max_{i\in I} \varphi_i(x)$$

$$\varphi_1(x) = x,\ \varphi_2(x) = -x,\ I = \{1,2\}.$$

Ясно, що $R(x_0) = \{1,2\}$, функції φ_1, φ_2 - гладкі. Візьмемо $\mathfrak{D}\varphi_1(x_0) = [\underline{\partial}\varphi_1(x_0),\overline{\partial}\varphi_1(x_0)]$, $\mathfrak{D}\varphi_2(x_0) = [\underline{\partial}\varphi_2(x_0),\overline{\partial}\varphi_2(x_0)]$,

$$\underline{\partial}\varphi_1(x_0) = \{\varphi_1'(x_0)\} = \{1\},\ \ \overline{\partial}\varphi_1(x_0) = \{0\},$$

$$\underline{\partial}\varphi_2(x_0) = \{\varphi_2'(x_0)\} = \{-1\},\ \ \overline{\partial}\varphi_2(x_0) = \{0\}.$$

За формулами (6.1.7) - (6.1.8) маємо

$$\mathfrak{D}f(x_0) = [\underline{\partial}f(x_0),\overline{\partial}f(x_0)], \tag{6.1.20}$$

$$\begin{aligned}\underline{\partial}f(x_0) = \operatorname{conv}\{\underline{\partial}\varphi_1(x_0) - \overline{\partial}\varphi_2(x_0),\underline{\partial}\varphi_2(x_0) - \overline{\partial}\varphi_1(x_0)\} &= \\ = \operatorname{conv}\{\{1\},\{-1\}\} &= [-1,1],\end{aligned}$$

$$\overline{\partial}f(x_0) = \overline{\partial}\varphi_1(x_0) + \overline{\partial}\varphi_2(x_0) = \{0\} + \{0\} = \{0\}.$$

Якщо ж як $\mathfrak{D}\varphi_1(x_0)$ і $\mathfrak{D}\varphi_2(x_0)$ взяти пари

$$\mathfrak{D}\varphi_1(x_0) = [\underline{\partial}\varphi_1(x_0),\overline{\partial}\varphi_1(x_0)],$$

$$\mathfrak{D}\varphi_2(x_0) = [\underline{\partial}\varphi_2(x_0),\overline{\partial}\varphi_2(x_0)],$$

$$\underline{\partial}\varphi_1(x_0) = \{0\},\ \ \overline{\partial}\varphi_1(x_0) = \{\varphi_1'(x_0)\} = \{1\},$$
$$\underline{\partial}\varphi_2(x_0) = \{\varphi_2'(x_0)\} = \{-1\},\ \ \overline{\partial}\varphi_2(x_0) = \{0\},$$

тоді отримаємо

$$\mathfrak{D}f(x_0) = [\underline{\partial}f(x_0),\overline{\partial}f(x_0)], \tag{6.1.21}$$

$$\underline{\partial}f(x_0) = \operatorname{conv}\{\underline{\partial}\varphi_1(x_0) - \overline{\partial}\varphi_2(x_0),\ \ \underline{\partial}\varphi_2(x_0) - \overline{\partial}\varphi_1(x_0)\} =$$
$$= \operatorname{conv}\{\{0\} - \{0\},\{-1\} - \{1\}\} = \operatorname{conv}\{\{0\},\{-2\}\} = [-2,0],$$
$$\overline{\partial}f(x_0) = \overline{\partial}\varphi_1(x_0) + \overline{\partial}\varphi_2(x_0) = \{1\} + \{0\} = \{1\}.$$

Очевидно, що пари (6.1.20) і (6.1.21) еквівалентні.
Для точок $x \neq 0$ маємо

$$f(x) = \begin{cases} x, & x > 0, \\ -x, & x < 0. \end{cases}$$

Тому можна взяти

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)], \tag{6.1.22}$$

$$\underline{\partial}f(x) = \begin{cases} \{1\}, & x > 0, \\ \{-1\}, & x < 0, \end{cases}$$
$$\overline{\partial}f(x) = \{0\}\ \ \forall\ x \neq 0.$$

Об‘єднуючи (6.1.22) і (6.1.20), можна взяти

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)]$$

$$\underline{\partial}f(x) = \begin{cases} \{1\}, & x > 0, \\ \{-1\}, & x < 0, \\ [-1,1], & x = 0, \end{cases}$$
$$\overline{\partial}f(x) = \{0\}\ \ \forall\ x \in \mathbb{R}^1.$$

Приклад 6.1.8. Нехай $f(x) = -|x|,\ \ x \in \mathbb{R}^1$. Використовуючи приклад 6.1.7, можна взяти

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)],$$
$$\underline{\partial}f(x) = \{0\}\ \ \forall\ x \in \mathbb{R}^1,$$
$$\overline{\partial}f(x) = \begin{cases} \{-1\}, & x > 0, \\ \{1\}, & x < 0, \\ [-1,1], & x = 0. \end{cases}$$

Приклад 6.1.9. Нехай $x \in \mathbb{R}^1$,

$$f(x) = \text{sat}\ x = \begin{cases} x, & |x| \leqslant 1, \\ 1, & x > 1, \\ -1, & x < -1, \end{cases}$$

Покладемо $\varphi_1(x) = x$, $\varphi_2(x) = 1$, $\varphi_3(x) = -1$, $\varphi_4(x) = \max\{\varphi_1(x),\varphi_3(x)\}$. Тоді, як неважко бачити,

$$f(x) = \min\{\varphi_2(x),\varphi_4(x)\}.$$

Оскільки функції $\varphi_1(x),\varphi_2(x)$ і $\varphi_3(x)$ диференційовні, то можна взяти

$$\mathfrak{D}\varphi_1(x) = [\{1\},\{0\}],\ \mathfrak{D}\varphi_2(x) = [\{0\},\{0\}],\ \mathfrak{D}\varphi_3(x) = [\{0\},\{0\}].$$

Тоді за формулами (6.1.7) - (6.1.8) маємо

$$\mathfrak{D}\varphi_4(x) = [\underline{\partial}\varphi_4(x),\ \ \overline{\partial}\varphi_4(x)],$$

$$\underline{\partial}\varphi_4(x) = \begin{cases} \{1\}, & x > -1, \\ \{0\}, & x < -1, \\ [0,1], & x = -1, \end{cases} \quad \overline{\partial}\varphi_4(x) = \{0\}.$$

Застосовуючи тепер формули (6.1.8) — (6.1.9), одержуємо

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)],$$

$$\underline{\partial}f(x) = \begin{cases} \{0\}, & |x| > 1, \\ [0,1], & x = 1, \\ \{1\}, & x \in (-1,1], \end{cases} \qquad \overline{\partial}f(x) = \begin{cases} \{0\}, & x \neq 1, \\ [-1,0], & x = 1. \end{cases}$$

Приклад 6.1.10. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $f(x) = |x^{(1)}| - |x^{(2)}|$. Тоді $f(x) = \varphi_1(x) + \varphi_2(x)$, де $\varphi_1(x) = |x^{(1)}|$, $\varphi_2(x) = -|x^{(2)}|$.

З прикладів 6.1.7 і 6.1.8 неважко побачити, що можна взяти

$$\mathfrak{D}\varphi_1(x) = [\underline{\partial}\varphi_1(x),\overline{\partial}\varphi_1(x)], \quad \mathfrak{D}\varphi_2(x) = [\underline{\partial}\varphi_2(x),\overline{\partial}\varphi_2(x)],$$

$$\underline{\partial}\varphi_i(x), \quad \overline{\partial}\varphi_i(x) \subset \mathbb{R}^2,\ i = 1,2,$$

$$\underline{\partial}\varphi_1(x) = \begin{cases} \{(1,0)\}, & x^{(1)} > 0, \\ \{(-1,0)\}, & x^{(1)} < 0, \\ \text{conv}\,\{(-1,0),(1,0)\}, & x^{(1)} = 0, \end{cases}$$

$$\overline{\partial}\varphi_1(x) = \{(0,0)\} \ \ \forall x \in \mathbb{R}^2,$$

$$\underline{\partial}\varphi_2(x) = \{(0,0)\} \ \ \forall x \in \mathbb{R}^2,$$

$$\overline{\partial}\varphi_2(x) = \begin{cases} \{(0,-1)\}, & x^{(2)} > 0, \\ \{(0,1)\}, & x^{(2)} < 0, \\ \text{conv}\,\{(0,-1),(0,1)\}, & x^{(2)} = 0, \end{cases}$$

За теоремою 6.1.1 маємо $\mathfrak{D}f(x)(x) = [\underline{\partial} f(x), \overline{\partial} f(x)]$,

$$\underline{\partial} f(x) = \underline{\partial}\varphi_1(x) + \underline{\partial}\varphi_2(x) = \underline{\partial}\varphi_1(x),$$

$$\overline{\partial} f(x) = \overline{\partial}\varphi_1(x) + \overline{\partial}\varphi_2(x) = \overline{\partial}\varphi_2(x).$$

Приклад 6.1.11. Нехай $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2$, $f(x) = |x^{(1)}| + |x^{(2)}|$. Тоді $f(x) = \varphi_1(x) + \varphi_2(x)$, де $\varphi_1(x) = |x^{(1)}|$, $\varphi_2(x) = |x^{(2)}|$.

Аналогічно до прикладу 6.1.10 можна взяти

$$\mathfrak{D}\varphi_1(x) = [\underline{\partial}\varphi_1(x), \overline{\partial}\varphi_1(x)], \mathfrak{D}\varphi_2(x) = [\underline{\partial}\varphi_2(x), \overline{\partial}\varphi_2(x)],$$

$$\underline{\partial}\varphi_i(x), \quad \overline{\partial}\varphi_i(x) \subset \mathbb{R}^2,\ i = 1{,}2,$$

$$\underline{\partial}\varphi_1(x) = \begin{cases} \{(1{,}0)\}, & x^{(1)} > 0, \\ \{(-1{,}0)\}, & x^{(1)} < 0, \\ \operatorname{conv}\{(-1{,}0),(1{,}0)\}, & x^{(1)} = 0, \end{cases}$$

$$\overline{\partial}\varphi_1(x) = \{(0{,}0)\} \quad \forall x \in \mathbb{R}^2,$$

$$\overline{\partial}\varphi_2(x) = \{(0{,}0)\} \quad \forall x \in \mathbb{R}^2,$$

$$\underline{\partial}\varphi_2(x) = \begin{cases} \{(0{,}1)\}, & x^{(2)} > 0, \\ \{(0,-1)\}, & x^{(2)} < 0, \\ \operatorname{conv}\{(0,-1),(0{,}1)\}, & x^{(2)} = 0, \end{cases}$$

За теоремою 6.1.1 $\mathfrak{D}f(x)(x) = [\underline{\partial} f(x), \overline{\partial} f(x)]$,

$$\underline{\partial} f(x) = \underline{\partial}\varphi_1(x) + \underline{\partial}\varphi_2(x),$$

$$\overline{\partial} f(x) = \overline{\partial}\varphi_1(x) + \overline{\partial}\varphi_2(x).$$

Приклад 6.1.12. Нехай $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2$, $\varphi_1(x) = |x^{(1)}| - |x^{(2)}|$, $\varphi_2(x) = |x^{(2)}| - |x^{(1)}|$. Покладемо $f(x) = \max\{\varphi_1(x), \varphi_2(x)\} = \left||x^{(1)}| - |x^{(2)}|\right|$. За прикладом 6.1.10

$$\mathfrak{D}\varphi_1(x) = [\underline{\partial}\varphi_1(x), \overline{\partial}\varphi_1(x)],$$

$$\underline{\partial}\varphi_1(x) = \begin{cases} \{(1{,}0)\}, & x^{(1)} > 0, \\ \{(-1{,}0)\}, & x^{(1)} < 0, \\ \operatorname{conv}\{(-1{,}0),(1{,}0)\}, & x^{(1)} = 0, \end{cases}$$

$$\overline{\partial}\varphi_1(x) = \begin{cases} \{(0,-1)\}, & x^{(2)} > 0, \\ \{(0{,}1)\}, & x^{(2)} < 0, \\ \operatorname{conv}\{(0,-1),(0{,}1)\}, & x^{(2)} = 0. \end{cases}$$

Застосовуючи теорему 6.1.1 отримаємо

$$\mathfrak{D}\varphi_2(x) = [\underline{\partial}\varphi_2(x), \overline{\partial}\varphi_2(x)],$$

$$\underline{\partial}\varphi_2(x) = -\overline{\partial}\varphi_1(x), \quad \overline{\partial}\varphi_2(x) = -\underline{\partial}\varphi_1(x)$$

Застосовуючи тепер формули (6.1.8) –(6.1.9), одержуємо

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)],$$

$$\underline{\partial}f(x) = \text{conv } \{\underline{\partial}\varphi_1(x) - \overline{\partial}\varphi_2(x),\underline{\partial}\varphi_2(x) - \overline{\partial}\varphi_1(x)\} =$$
$$= \text{conv } \{2\underline{\partial}\varphi_1(x), -2\overline{\partial}\varphi_1(x)\}$$
$$\overline{\partial}f(x) = \overline{\partial}\varphi_1(x) + \overline{\partial}\varphi_2(x) = \overline{\partial}\varphi_1(x) - \underline{\partial}\varphi_1(x).$$

Приклад 6.1.13. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $\varphi_1(x) = |x^{(1)}| + |x^{(2)}|$, $\varphi_2(x) = |x^{(1)}| - |x^{(2)}|$. Покладемо $f(x) = \max\{\varphi_1(x),\varphi_2(x)\}$. Аналогічно до попереднього

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)],$$

де

$$\underline{\partial}f(x) = \text{conv } \{\underline{\partial}\varphi_1(x) - \overline{\partial}\varphi_2(x),\underline{\partial}\varphi_2(x) - \overline{\partial}\varphi_1(x)\}$$
$$\overline{\partial}f(x) = \overline{\partial}\varphi_1(x) + \overline{\partial}\varphi_2(x).$$

Квазідиференціали $\mathfrak{D}\varphi_1(x) = [\underline{\partial}\varphi_1(x),\overline{\partial}\varphi_1(x)]$, $\mathfrak{D}\varphi_2(x) = [\underline{\partial}\varphi_2(x),\overline{\partial}\varphi_2(x)]$ знайдені в прикладах 6.1.11 і 6.1.10 відповідно.

Приклад 6.1.14. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $\varphi_1(x) = |x^{(1)}|+|x^{(2)}|$, $\varphi_2(x) = |x^{(1)}|-|x^{(2)}|$, $x_0 = 0$. Покладемо $f(x) = \min\{\varphi_1(x),\varphi_2(x)\}$. За прикладами 6.1.10, 6.1.11 і теоремою 6.1.2

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)],$$

де

$$\underline{\partial}f(x) = \underline{\partial}\varphi_1(x) + \underline{\partial}\varphi_2(x),$$
$$\overline{\partial}f(x) = \text{conv } \{\overline{\partial}\varphi_1(x) - \underline{\partial}\varphi_2(x),\overline{\partial}\varphi_2(x) - \underline{\partial}\varphi_1(x)\}.$$

Зокрема

$$\underline{\partial}f(x_0) = \text{conv}\{(-1,0),(1,0)\} + \text{conv}\{(-1,0),(1,0)\}+$$
$$+\text{conv}\{(0,-1),(0,1)\},$$
$$\overline{\partial}f(x_0) = \text{conv}\{\text{conv}\{(0,-1),(0,1)\} - \text{conv}\{(-1,0),(1,0)\}-$$
$$-\text{conv}\{(0,-1),(0,1)\},\text{conv}\{(-1,0),(1,0)\}\} =$$
$$= \text{conv}\{\text{conv}\{(0,-1),(0,1)\} + \text{conv}\{(-1,0),(1,0)\}+$$
$$+\text{conv}\{(0,-1),(0,1)\},\text{conv}\{(-1,0),(1,0)\}\}.$$

Остання рівність має місце оскільки

$$\text{conv}\{(-1,0),(1,0)\} = -\text{conv}\{(-1,0),(1,0)\},$$
$$\text{conv}\{(0,-1),(0,1)\} = -\text{conv}\{(0,-1),(0,1)\}.$$

Приклад 6.1.15. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $f(x) = ||x^{(1)}| + x^{(2)}|$. Тоді $f(x) = \max\{\varphi_1(x),\varphi_2(x)\}$, де $\varphi_1(x) = |x^{(1)}| + x^{(2)}$, $\varphi_2(x) = -|x^{(1)}| - x^{(2)}$.

Визначимо $\varphi_3(x) = x^{(2)}$. Для неї квазідиференціал буде $\mathfrak{D}\varphi_3(x) = [\{(0{,}1)\},\{(0{,}0)\}]$. Тоді за прикладами 6.1.7, 6.1.8

$$\mathfrak{D}\varphi_1(x) = [\underline{\partial}\varphi_1(x),\overline{\partial}\varphi_1(x)],$$

$$\underline{\partial}\varphi_1(x) = \begin{cases} \{(1{,}0)\} + \{(0{,}1)\}, & x^{(1)} > 0, \\ \{(-1{,}0)\} + \{(0{,}1)\}, & x^{(1)} < 0, \\ \operatorname{conv}\{(-1{,}0),(1{,}0)\} + \{(0{,}1)\}, & x^{(1)} = 0, \end{cases}$$

$$\overline{\partial}\varphi_1(x) = \{(0{,}0)\}$$

$$\mathfrak{D}\varphi_2(x) = [\underline{\partial}\varphi_2(x),\overline{\partial}\varphi_2(x)],$$

$$\underline{\partial}\varphi_2(x) = \{(0{,}0)\}$$

$$\overline{\partial}\varphi_2(x) = \begin{cases} \{(-1{,}0)\} - \{(0{,}1)\}, & x^{(1)} > 0, \\ \{(1{,}0)\} - \{(0{,}1)\}, & x^{(1)} < 0, \\ -\operatorname{conv}\{(-1{,}0),(1{,}0)\} - \{(0{,}1)\}, & x^{(1)} = 0, \end{cases}$$

За теоремою 6.1.2

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)],$$

$$\underline{\partial}f(x) = \operatorname{conv}\ \{\underline{\partial}\varphi_1(x) - \overline{\partial}\varphi_2(x),\underline{\partial}\varphi_2(x) - \overline{\partial}\varphi_1(x)\},$$

$$\overline{\partial}f(x) = \overline{\partial}\varphi_1(x) + \overline{\partial}\varphi_2(x).$$

Зокрема

$$\underline{\partial}f(x_0) = \operatorname{conv}\{\operatorname{conv}\{(-1{,}0),(1{,}0)\} + \{(0{,}1)\}+ \\ + \operatorname{conv}\{(-1{,}0),(1{,}0)\} + \{(0{,}1)\},\{(0{,}0)\}\},$$

$$\overline{\partial}f(x_0) = -\operatorname{conv}\{(-1{,}0),(1{,}0)\} - \{(0{,}1)\}.$$

Приклад 6.1.16. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $x_0 = 0$, $f(x) = |x^{(1)} + x^{(2)}|$. Тоді $f(x) = \max\{\varphi_1(x),\varphi_2(x)\}$, де $\varphi_1(x) = x^{(1)} + x^{(2)}$, $\varphi_2(x) = -x^{(1)} - x^{(2)}$. Аналогічно до попередніх прикладів

$$\mathfrak{D}f(x_0) = [\underline{\partial}f(x_0),\overline{\partial}f(x_0)],$$

$$\underline{\partial}f(x_0) = \operatorname{conv}\{(-1,-1),(1{,}1)\},$$

$$\overline{\partial}f(x_0) = \{(0{,}0)\}.$$

Приклад 6.1.17. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $x_0 = 0$, $\varphi_1(x) = 1 + |x^{(1)}|$, $\varphi_2(x) = 1+|x^{(1)}+x^{(2)}|$. Покладемо $f(x) = \varphi_1(x)\cdot\varphi_2(x)$. Якщо c- константа, то $\mathfrak{D}[f(x) + c] = \mathfrak{D}f(x)$. Тому

$$\mathfrak{D}\varphi_1(x_0) = [\underline{\partial}\varphi_1(x_0),\overline{\partial}\varphi_1(x_0)],$$

$$\underline{\partial}\varphi_1(x_0) = \operatorname{conv}\{(-1{,}0),(1{,}0)\},$$

$$\overline{\partial}\varphi_1(x_0) = \{(0{,}0)\}.$$

За прикладом 6.1.16

$$\mathfrak{D}\varphi_2(x_0)=[\underline{\partial}f_2(x_0),\overline{\partial}\varphi_2(x_0)],$$
$$\underline{\partial}\varphi_2(x_0)=\operatorname{conv}\{(-1,-1),(1{,}1)\},$$
$$\overline{\partial}\varphi_2(x_0)=\{(0{,}0)\}.$$

За теоремою 6.1.1

$$\mathfrak{D}f(x_0)=[\underline{\partial}f(x_0),\overline{\partial}f(x_0)],$$
$$\underline{\partial}f(x_0)=\varphi_1(x_0)\cdot\underline{\partial}\varphi_2(x_0)+\varphi_2(x_0)\cdot\underline{\partial}\varphi_1(x_0)=$$
$$=\operatorname{conv}\{(-1{,}0),(1{,}0)\}+\operatorname{conv}\{(-1,-1),(1{,}1)\},$$
$$\overline{\partial}f(x_0)=\{(0{,}0)\}.$$

Приклад 6.1.18. Нехай $x=(x^{(1)},x^{(2)})\in\mathbb{R}^2$, $\varphi_1(x)=|x^{(1)}|+x^{(2)}+1$, $\varphi_2(x)=[\varphi_3(x)]^{-1}$, $\varphi_3(x)=|x^{(1)}|-|x^{(2)}|+1$. Покладемо $f(x)=\varphi_1(x)\cdot\varphi_2(x)$. Оскільки $\mathfrak{D}\varphi_1(x)=\mathfrak{D}\varphi_4(x)$, де $\varphi_4(x)=|x^{(1)}|+x^{(2)}$, то за прикладом 6.1.15

$$\mathfrak{D}\varphi_1(x)=[\underline{\partial}\varphi_1(x),\overline{\partial}\varphi_1(x)],$$
$$\underline{\partial}\varphi_1(x)=\begin{cases}\{(1{,}0)\}+\{(0{,}1)\}, & x^{(1)}>0,\\ \{(-1{,}0)\}+\{(0{,}1)\}, & x^{(1)}<0,\\ \operatorname{conv}\{(-1{,}0),(1{,}0)\}+\{(0{,}1)\}, & x^{(1)}=0,\end{cases}$$
$$\overline{\partial}\varphi_1(x)=\{(0{,}0)\}$$

Оскільки $\mathfrak{D}\varphi_3(x)=\mathfrak{D}\varphi_5(x)$, де $\varphi_5(x)=|x^{(1)}|-|x^{(2)}|$, то за прикладом 6.1.10

$$\mathfrak{D}\varphi_3(x)=[\underline{\partial}\varphi_3(x),\overline{\partial}\varphi_3(x)],$$
$$\underline{\partial}\varphi_3(x)=\begin{cases}\{(1{,}0)\}, & x^{(1)}>0,\\ \{(-1{,}0)\}, & x^{(1)}<0,\\ \operatorname{conv}\{(-1{,}0),(1{,}0)\}, & x^{(1)}=0,\end{cases}$$
$$\overline{\partial}\varphi_3(x)=\begin{cases}\{(0,-1)\}, & x^{(2)}>0,\\ \{(0{,}1)\}, & x^{(2)}<0,\\ \operatorname{conv}\{(0,-1),(0{,}1)\}, & x^{(2)}=0,\end{cases}$$

За теоремою 6.1.1

$$\mathfrak{D}\varphi_2(x)=[\underline{\partial}\varphi_2(x),\overline{\partial}\varphi_2(x)],$$
$$\underline{\partial}\varphi_2(x)=-(\varphi_3(x))^{-2}\overline{\partial}\varphi_3(x),\quad \overline{\partial}\varphi_2(x)=-(\varphi_3(x))^{-2}\underline{\partial}\varphi_3(x).$$

Знову використовуючи теорему 6.1.1, остаточно отримаємо

$$\mathfrak{D}f(x)=[\underline{\partial}f(x),\overline{\partial}f(x)],$$
$$\underline{\partial}f(x)=\varphi_1(x)\cdot\underline{\partial}\varphi_2(x)+\varphi_2(x)\cdot\underline{\partial}\varphi_1(x),$$
$$\overline{\partial}f(x)=\varphi_1(x)\cdot\overline{\partial}\varphi_2(x)+\varphi_2(x)\cdot\overline{\partial}\varphi_1(x).$$

Приклад 6.1.19. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $\varphi_1(x) = |x^{(1)}| + x^{(2)} + 1$. Покладемо $f(x) = [\varphi_1(x)]^2$.

Як зазначалося в попередньому прикладі

$$\mathfrak{D}\varphi_1(x) = [\underline{\partial}\varphi_1(x),\overline{\partial}\varphi_1(x)],$$

$$\underline{\partial}\varphi_1(x) = \begin{cases} \{(1,0)\} + \{(0,1)\}, & x^{(1)} > 0, \\ \{(-1,0)\} + \{(0,1)\}, & x^{(1)} < 0, \\ \mathrm{conv}\,\{(-1,0),(1,0)\} + \{(0,1)\}, & x^{(1)} = 0, \end{cases}$$

$$\overline{\partial}\varphi_1(x) = \{(0,0)\}$$

За теоремою 6.1.1

$$\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)],$$

$$\underline{\partial}f(x) = \varphi_1(x)\cdot\underline{\partial}\varphi_1(x) + \varphi_1(x)\cdot\underline{\partial}\varphi_1(x),$$

$$\overline{\partial}f(x) = \varphi_1(x)\cdot\overline{\partial}\varphi_1(x) + \varphi_1(x)\cdot\overline{\partial}\varphi_1(x).$$

Приклад 6.1.20. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $f(x) = ||x||$, де $||x|| = \sqrt{(x^{(1)})^2 + (x^{(2)})^2}$, $x_0 = 0$. Очевидно, що $\frac{\partial f(x_0)}{\partial g} = \lim_{\alpha\to+0}\alpha^{-1}[f(x_0 + \alpha g) - f(x_0)] = \alpha\cdot\alpha^{-1}||g|| = ||g||$. Отже $\frac{\partial f(x_0)}{\partial g} = \max_{v\in S_1((0,0))}(v,g)$, де $S_1((0,0)) = \{v \in \mathbb{R}^2|\ ||v|| \le 1\}$. Тому функція $f(x)$ є квазідиференційовною (навіть субдиференційовною), причому

$$\mathfrak{D}f(x_0) = [\underline{\partial}f(x_0),\overline{\partial}f(x_0)],$$

де

$$\underline{\partial}f(x_0) = S_1((0,0)), \quad \overline{\partial}f(x_0) = \{(0,0)\}$$

6.2. РІЗНИЦІ ОПУКЛИХ КОМПАКТІВ

При вивченні сублінійних функцій і опуклих компактів зручно використовувати *спряженість Мінковського*, тобто відображення φ, що ставить у відповідність кожній сублінійній функції її субдиференціал. Відображення φ здійснює взаємно однозначну відповідність між сукупністю P усіх визначених на $\mathbb{R}^n$ сублінійних функцій і сукупністю N всіх непорожніх опуклих компактних підмножин простору $\mathbb{R}^n$. При цьому сумі функцій відповідає сума (за Мінковським) множин, добутку функції на невід'ємне число – добуток множини на те ж число, максимуму сублінійних функцій – опукла оболонка об'єднання множин. Обернене до φ відображення відновлює сублінійну функцію p_V на опуклому компакті V за допомогою операції взяття максимуму:

$$p_V(g) = \max_{l\in U}\langle l,g\rangle.$$

Спряженість Мінковського можна розглядати і на сукупності суперлінійних функцій. Тут вона має такі ж властивості, як і у сублінійному випадку, за тим виключенням, що опукла оболонка об'єднання компактів відповідає мінімуму, а не максимуму відповідних функцій, а обернене відображення відновлює суперлінійну функцію q_V на опуклому компакті V за допомогою операції взяття мінімуму.

Відображення φ відіграє основну роль при дослідження геометричної інтерпретації похідної. Так, верхня похідна Кларка завжди є сублінійною функцією. Тому в силу спряженості Мінковського їй відповідає опуклий компакт, який є субдиференціалом Кларка. Зокрема, якщо похідна за напрямками $f'(x,g)$ напівнеперервна зверху як функція точки x, то вона сублінійна як функція напрямку g. Тому їй відповідає опуклий компакт – її субдиференціал, за яким вона відновлюється за допомогою взяття максимуму. Якщо ж похідна $f'(x,g)$ напівнеперервна знизу (за x), то вона суперлінійна як функція напрямку. Тому їй відповідає супердиференціал, за яким вона відновлюється за допомогою взяття мінімуму. Однак припущення про сублінійність чи суперлінійність похідної як функції напрямку, зокрема, про напівнеперервність зверху чи знизу цієї похідної як функції точки, часто виявляється дуже жорстким. Так, функція, яка може бути представлена у вигляді різниці опуклих функцій, вже не завжди має потрібні властивості. Геометричну інтерпретацію її похідної за напрямками не можна дати на основі класичних методів.

Спробуємо поширити спряженість Мінковського на ширший ніж множина сублінійних чи суперлінійних функцій клас. Основна перешкода тут полягає в тому, що множина сублінійних функцій P не є лінійним простором. Різниця сублінійних функцій не завжди сублінійна.

На множині непорожніх опуклих компактів N визначимо *віднімання за Мінковським* таким чином: якщо $U,V \in N$, то $U - V = \{y - z | y \in U, z \in V\}$. Ця операція, однак, не узгоджена зі звичними операціями в N, точніше кажучи, віднімання за Мінковським не є операцією, оберненою до додавання. Дійсно, якщо множина U містить більш однієї точки, то $U - U \neq \{0\}$ (множина $U - U$ містить елементи вигляду $y - x$ і $x - y$, де $x,y \in U, x \neq y$). Віднімання за Мінковським не представляє особливого інтересу при вивченні спряженості Мінковського. Опорна функція p_{U-V} різниці $U - V = U + (-V)$ збігається з сумою (а не різницею) опорних функцій p_U та p_V компактів U та $-V$. Якщо множина V симетрична, наприклад, $V = B$ – одинична куля, то $V = -V$ і $p_{U-V} = p_U + p_V$.

У зв'язку зі сказаним робились різні спроби визначити операцію віднімання на сукупності опуклих компактів так, щоб вона в деякому сенсі була більш близькою до операції, оберненої до додавання, ніж різниця за Мінковським (наприклад, мала б ту властивість, що $U - U = \{0\}$ для будь-якого $U \in N$). На жаль, якщо залишатися в рамках простору

N, то говорити про обернену до додавання операцію не можна. Однак, використовуючи загальноприйняті в алгебрі підходи, можна розширити простір N до деякого лінійного простору M. У цьому просторі в силу його лінійності операція, обернена до додавання, існує, тим самим вона визначена і для елементів N. Однак її результат вже не обов'язково лежить в N, він може знаходитися серед доданих елементів. Ми розглянемо зазначену конструкцію спочатку для сублінійних функцій, а потім, використовуючи спряженість Мінковського, застосуємо її до вивчення опуклих компактів.

Розглянемо множину L всіх визначених на $\mathbb{R}^n$ функцій, які можуть бути представлені у вигляді суми сублінійної і суперлінійної функцій або у вигляді різниці двох сублінійних функцій. Іншими словами, $L = P+Q$, де P – сукупність усіх сублінійних, а Q – сукупність усіх суперлінійних функцій.

Оскільки сублінійна функція, яка визначена на всьому просторі, неперервна, то елементи множини L є неперервними функціями, тобто $L \subset C(\mathbb{R}^n)$, де $C(\mathbb{R}^n)$ — простір неперервних на $\mathbb{R}^n$ функцій. Можна перевірити, що L є лінійним підпростором у просторі $C(\mathbb{R}^n)$, і, тим самим, сам є лінійним простором: якщо $l_1,l_2 \in L$, то і $\lambda_1 l_1 + \lambda_2 l_2 \in L$ для будь-яких дійсних λ_1,λ_2. Дійсно, нехай $l_i = p_i + q_i$, де $p_i \in P$, $q_i \in Q, i = 1{,}2$. Маємо

$$\lambda_1 l_1 + \lambda_2 l_2 = \lambda_1 p_1 + \lambda_1 q_1 + \lambda_2 p_2 + \lambda_2 q_2. \qquad (6.2.1)$$

Припустимо, наприклад, що $\lambda_1 > 0,\lambda_2 < 0$. Тоді функція $\lambda_1 p_1 + \lambda_2 q_2$ сублінійна, а функція $\lambda_1 q_1 + \lambda_2 p_2$ суперлінійна і тому, як випливає з (6.2.1), $\lambda_1 l_1 + \lambda_2 l_2 = (\lambda_1 p_1 + \lambda_2 q_2) + (\lambda_1 q_1 + \lambda_2 p_2) \in L$. Аналогічно розглядаються три інші можливі випадки

$$1)\lambda_1 > 0,\lambda_2 > 0;\ \ 2)\lambda_1 < 0,\lambda_2 > 0;\ \ 3)\lambda_1 < 0,\lambda_2 < 0.$$

Визначимо в L природне відношення порядку: якщо $l_1,l_2 \in L$, то співвідношення $l_1 \geq l_2$ означає, що $l_1(x) \geq l_2(x)$ для всіх $x \in \mathbb{R}^n$. Властивості простору L описуються в термінах цього відношення порядку.

Одна з найважливіших властивостей простору L полягає в тому, що з будь-якими своїми елементами $f_1,...,f_k$ він містить їх поточковий максимум $\max\{f_1(x),...,f_k(x)\}$ а також поточковий мінімум $\min\{f_1(x),...,f_k(x)\}$ функцій $f_1,\ldots,f_k$. Щоб показати це, скористаємося наступною лемою.

Лема 6.2.1. *Нехай $f_1,\ldots,f_k$ визначені на деякій множині X функції, причому $f_i = \varphi_i + \psi_i, i = 1,\ldots,k$. Тоді*

$$\max_{i=1,\ldots,k} f_i(x) = \max_{j=1,\ldots,k} \left(\varphi_j(x) - \sum_{i \neq j} \psi_i(x) \right) + \sum_{i=1}^{k} \psi_i(x), \qquad (6.2.2)$$

$$\min_{i=1,\dots,k} f_i(x) = \sum_{i=1}^{k} \varphi_i(x) + \min_{j=1,\dots,k} \left(\psi_j(x) - \sum_{i \neq j} \varphi_i(x) \right). \qquad (6.2.3)$$

Доведення. Обмежимося перевіркою формули (6.2.2). Розглянемо спочатку випадок $k = 2$. Тоді формула (6.2.2) виглядає так:

$$\max\{f_1(x), f_2(x)\} =$$

$$= \max\{\varphi_1(x) - \psi_2(x), \varphi_2(x) - \psi_1(x)\} + (\psi_1(x) + \psi_2(x)). \qquad (6.2.4)$$

Якщо $f_1(x) \geq f_2(x)$, то права і ліва частини формули (6.2.4) збігаються з $f_1(x)$, якщо ж $f_2(x) \geq f_1(x)$, то ці частини збігаються з $f_2(x)$. Таким чином, рівність (6.2.4) правильна. Скористаємося тепер індукцією за k. Припустимо, що для множин з $k-1$ функцій лема справедлива і розглянемо набір з k функцій: $f_1, \dots, f_{k-1}, f_k$. Покладемо $f_*(x) = \max\limits_{i=1,\dots,k-1} f_i(x)$. Тоді, за індукційним припущенням, $f_* = \varphi_* + \psi_*$, де

$$\varphi_*(x) = \max_{j=1,\dots,k-1} \{\varphi_j(x) - \sum_{i=1, i \neq j}^{k-1} \psi_i(x)\}, \quad \psi_*(x) = \sum_{i=1}^{k-1} \psi_i(x).$$

Оскільки $\max\limits_{i=1,\dots,k} f_i(x) = \max\{\max\limits_{i=1,\dots,k-1} f_i(x), f_k(x)\} = \max\{f_*(x), f_k(x)\}$, то, використовуючи (6.2.4), маємо

$$\max_{i=1,\dots,k} f_i(x) = \max\{\alpha(x), \beta(x)\} + \sum_{i=1}^{k-1} \psi_i(x) + \psi_k(x), \qquad (6.2.5)$$

$$\alpha(x) = \max_{j=1,\dots,k-1} \left\{ \varphi_j(x) - \sum_{i=1,\dots,k-1; i \neq j} \psi_i(x) - \psi_k(x) \right\},$$

$$\beta(x) = \varphi_k(x) - \sum_{i=1}^{k-1} \psi_i(x).$$

Розглянемо величину

$$\gamma(x) = \max_{j=1,\dots,k} \left\{ (\varphi_j(x) - \sum_{i=1,\dots,k; i \neq j} \psi_i(x) \right\}. \qquad (6.2.6)$$

Якщо точка x така, що максимум у (6.2.6) досягається на індексі k, то $\gamma(x) = \varphi_k(x) - \sum\limits_{i=1}^{k-1} \psi_i(x) = \beta(x)$. В той же час, $\gamma(x) \geq \varphi_j(x) - \sum\limits_{i \in 1:k, i \neq j} \psi_i(x)$ при всіх $j \neq k$, і, тим самим, $\gamma(x) \geq \alpha(x)$. Таким чином, у даному випадку $\gamma(x) = \max\{\alpha(x), \beta(x)\}$.

Нехай максимум в (6.2.6) досягається на індексі $j < k$. Тоді

$$\gamma(x) = \varphi_j(x) - \sum_{i=1,\dots,k;i\neq j} \psi_i(x)$$

і, в той же час,

$$\gamma(x) \geq \varphi_m(x) - \sum_{i=1,\dots,k;i\neq j} \psi_i(x) \ (m \neq j).$$

Звідси випливає, що $\gamma(x) = \alpha(x),\gamma(x) \geq \beta(x)$, тобто і в даному випадку $\gamma(x) = \max\{\alpha(x),\beta(x)\}$. Використовуючи формулу (6.2.5), переконаємося в справедливості леми. $\square$

Теорема 6.2.1. *Нехай функції $l_1,\dots,l_k$ належать простору L, причому $l_i = p_i + q_i$, де $p_i \in P,q_i \in Q$. Тоді поточковий максимум $\max\{l_i(x),i = 1,\dots,k\}$ і поточковий мінімум $\min\{l_i(x),i = 1,\dots,k\}$ цих функцій також належать простору L, причому*

$$\max_{i=1,\dots,k} l_i(x) = \max_{j=1,\dots,k} \left\{ p_j(x) - \sum_{i=1,\dots,k;i\neq j} q_i(x) \right\} + \sum_{i=1}^{k} q_i(x), \qquad (6.2.7)$$

$$\min_{i=1,\dots,k} l_i(x) = \sum_{i=1}^{k} p_i(x) + \min_{j=1,\dots,k} \left\{ q_j(x) - \sum_{i=1,\dots,k;i\neq j} p_i(x) \right\}. \qquad (6.2.8)$$

Доведення. Справедливість формул (6.2.7) і (6.2.8) випливає з леми 6.2.1. Покладемо

$$p(x) = \max_j \left\{ p_j(x) - \sum_{i\neq j} q_i(x) \right\}, \quad q(x) = \sum_{i-1}^{k} q_i(x).$$

Функція p сублінійна, а q – суперлінійна, тому $\max_i\{l_i(x)\}$ входить в L. Таким же чином перевіряється, що $\min_i\{l_i(x)\}$ є елементом L. $\square$

Нехай $l \in L$ і $l = p+q$, де p – сублінійна, а q – суперлінійна функції. В силу спряженості Мінковського функції p відповідає її субдиференціал $\underline{\partial} p = U$, а функції q – її супердиференціал $\overline{\partial} q = V$. При цьому p відновлюється за множиною U за допомогою операції взяття максимуму, а q відновлюється за множиною V за допомогою операції взяття мінімуму:

$$p(x) = \max_{h\in U} \langle h,x \rangle, \quad q(x) = \min_{h\in V} \langle h,x \rangle. \qquad (6.2.9)$$

Отже, кожній функції $l \in L$ відповідає впорядкована пара опуклих компактів $[U,V]$, причому l відновлюється за цією парою за допомогою операцій взяття максимуму за першою компонентою і мінімуму за другою:

$$l(x) = \max_{h\in U}\langle h,x \rangle + \min_{h\in V}\langle h,x \rangle \ \forall x \in R^n. \qquad (6.2.10)$$

Представлення функцій l у вигляді суми елементів з P і Q неєдине. Так, якщо $l = p + q$, де $p \in P, q \in Q$, то l можна представити й у вигляді $l = (p + p') + (q - p')$, де p' – будь-який елемент із P. При цьому $p + p' \in P, q - p' \in Q$.
Нехай $l = p_1 + q_1, l = p_2 + q_2$, де $p_i \in P, q_i \in Q, i = 1,2$. Покладемо $U_i = \underline{\partial} p_i, V_i = \overline{\partial} q_i, i = 1,2$. Оскільки $p_1 + q_1 = p_2 + q_2$, то $p_1 + (-q_2) = p_2 + (-q_1)$. Функція $-q_1$ сублінійна і її субдиференціал збігається з множиною $-\overline{\partial} q_i = -V_i, i = 1,2$. Це випливає з рівності

$$-q_i(x) = -\min_{h \in V_i} \langle h,x \rangle = \max_{h \in V_i} \langle -h,x \rangle = \max_{h \in -V_i} \langle h,x \rangle.$$

Застосовуючи спряженість Мінковського, одержимо, що

$$U_1 - V_2 = U_2 - V_1. \tag{6.2.11}$$

Отже, якщо дві пари опуклих компактів $[U_1,V_1]$ і $[U_2,V_2]$ такі, що за ними відновлюється за допомогою формули (6.2.10) одна й та ж функція l, то виконується рівність (6.2.11). Міркуючи подібним чином, легко перевірити і зворотне: якщо виконана рівність (6.2.11), то

$$\max_{h \in U_1} \langle h,x \rangle + \min_{h \in V_1} \langle h,x \rangle = \max_{h \in U_2} \langle h,x \rangle + \min_{h \in V_2} \langle h,x \rangle \ \forall x \in R^n,$$

тобто $l_1 = l_2$.

Впорядковані пари опуклих компактів будемо називати *еквівалентними* (позначення: $[U_1,V_1] \approx [U_2,V_2]$), якщо $U_1 - V_2 = U_2 - V_1$. Наведені вище міркування показують, як можна поширити спряженість Мінковського на функції з простору L. При цьому через наявність еквівалентних пар, які не співпадають між собою, це поширення проводиться технічно досить складно.

Розглянемо множину N^2 – декартів квадрат множини непорожніх опуклих компактів N. Елементами N^2 є впорядковані пари $[U,V]$, де $U,V \in N$. На множині N^2 природним чином вводяться *операції додавання* і *множення на додатне число*:

$$[U_1,V_1] + [U_2,V_2] = [U_1 + U_2, V_1 + V_2], \quad \lambda[U,V] = [\lambda U, \lambda V] \ \forall \lambda > 0.$$

Кожній парі $[U,V] \in N^2$ поставимо у відповідність функцію $l_{[U,V]}$, що визначена формулою (6.2.10):

$$l_{[U,V]}(x) = \max_{h \in U} \langle h,x \rangle + \min_{h \in V} \langle h,x \rangle. \tag{6.2.12}$$

Можна перевірити, що

$$l_{[U_1,V_1]} + l_{[U_2,V_2]} = l_{[U_1,V_1]+[U_2,V_2]}, \tag{6.2.13}$$

$$\lambda l_{[U,V]} = l_{\lambda[U,V]} \ \forall \lambda > 0. \tag{6.2.14}$$

Перевіримо, наприклад, формулу (6.2.13). Маємо

$$l_{[U_1,V_1]}(x) + l_{[U_2,V_2]}(x) = \max_{h \in U_1} \langle h,x \rangle + \min_{h \in V_1} \langle h,x \rangle + \max_{h \in U_2} \langle h,x \rangle + \min_{h \in V_2} \langle h,x \rangle =$$

$$= \max_{h \in U_1 + U_2} \langle h,x \rangle + \min_{h \in V_1 + V_2} \langle h,x \rangle = l_{[U_1+U_2, V_1+V_2]} = l_{[U_1,V_1]+[U_2,V_2]}.$$

Знайдемо тепер пару множин, яка відповідає функції $\lambda l_{[U,V]}$ при $\lambda < 0$. Спочатку візьмемо $\lambda = -1$. Матимемо

$$-l_{[U,V]}(x) = -(\max_{h\in U}\langle h,x\rangle + \min_{h\in V}\langle h,x\rangle) = \min_{h\in U}\langle -h,x\rangle + \max_{h\in V}\langle -h,x\rangle =$$

$$= \max_{h\in -V}\langle h,x\rangle + \min_{h\in -U}\langle h,x\rangle = l_{[-V,-U]}. \qquad (6.2.15)$$

Нехай тепер λ – довільне від'ємне число. Тоді, використовуючи (6.2.14) і (6.2.15), маємо

$$\lambda l_{[U,V]} = |\lambda|(-l_{[U,V]}) = |\lambda| l_{[-V,-U]} = l_{[\lambda V,\lambda U]}. \qquad (6.2.16)$$

На базі формули (6.2.16) природно ввести наступне визначення: якщо $[U,V] \in N^2, \lambda < 0$, то $\lambda[U,V] = [\lambda V,\lambda U]$. Тим самим в N^2 визначене множення на всі дійсні числа, однак це множення не узгоджене з операцією додавання, і тому N^2 не є лінійним простором. Так, наприклад,

$$[U,V] - [U,V] = [U,V] + [-V,-U] = [U-V,V-U].$$

Таким чином, різниця $[U,V] - [U,V]$ не обов'язково співпадає з нулем.

Щоб перетворити N^2 у лінійний простір, потрібно ототожнити еквівалентні між собою пари, розглядаючи їх як один елемент, оскільки сукупність еквівалентних пар описується однією функцією з L, причому алгебраїчні операції над функціями відповідають алгебраїчним операціям над парами.

Розглянемо множину M, елементами якої є класи α еквівалентних пар. (Іншими словами, M є фактор-множиною множини N^2 за відношенням еквівалентності $\approx$.) Елемент $\alpha \in M$, що містить пару $[U,V]$, будемо позначати через $[[U,V]]$. Таким чином, $[[U_1,V_1]] = [[U_2,V_2]]$ тоді і тільки тоді, коли $U_1 - V_2 = V_2 - U_1$. Клас α, що містить пару $[\{0\},\{0\}]$, назвемо нулем і позначимо символом 0. Зрозуміло, що $[[A,B]] = 0$ тоді і тільки тоді, коли $A = -B$.

Нехай $l \in L$. Через $\psi(l)$ позначимо клас $\alpha \in M$, що складається з усіх пар $[U,V] \in N^2$, за яких справедлива рівність $l = l_{[U,V]}$ (де $l_{[U,V]}$ визначене формулою (6.2.12)). Зі сказаного випливає, що відображення ψ здійснює взаємно однозначну відповідність між L і M. Під сумою $\alpha_1+\alpha_2$ елементів $\alpha_1 = [[U_1,V_1]]$ і $\alpha_2 = [[U_2,V_2]]$ простору M будемо розуміти клас $\alpha = [[U_1+U_2,V_1+V_2]]$. Добуток елемента $\alpha = [[U,V]]$ на дійсне число λ визначається так: $\lambda\alpha = [[\lambda[U,V]]]$. Можна показати, що ці визначення коректні в тому сенсі, що сума $\alpha_1 + \alpha_2$ не залежить від вибору конкретних пар із класів α_1,α_2, а добуток $\lambda\alpha$ не залежить від вибору пари $[U,V]$ із класу α. Це твердження легко перевіряється безпосередньо. Його можна перевірити і використовуючи відображення ψ. Дійсно, якщо функції l_1,l_2 такі, що $\alpha_1 = \psi(l_1),\alpha_2 = \psi(l_2)$, то

$$\alpha_1 + \alpha_2 = \psi(l_1 + l_2). \qquad (6.2.17)$$

Якщо $\alpha = \psi(l),\lambda$ – дійсне число, то

$$\lambda\alpha = \psi(\lambda l). \tag{6.2.18}$$

Рівності (6.2.17) і (6.2.18) дають представлення класів $\alpha_1+\alpha_2$ і $\lambda\alpha$, що не базується на використанні конкретних пар, які містяться в цих класах. Ці рівності показують також, що взаємно однозначне відображення $\psi : L \to M$ лінійне. Звідси і з лінійності простору L випливає лінійність простору M.

Клас α, що містить пари вигляду$[U,\mathbf{0}]$, де $U \in N,\mathbf{0} = \{0\}$, позначимо через α_U. Помітимо, що якщо $[U,\mathbf{0}] \in \alpha$ і $[V,\mathbf{0}] \in \alpha$, то $[U,\mathbf{0}] \approx [V,\mathbf{0}]$, тобто $U = V$. Таким чином, рівності $\alpha_U = \alpha_V$ і $U = V$ рівносильні. Зрозуміло, що функція $l_{[U,\mathbf{0}]}$ співпадає з опорною функцією компакта U. Тому, ототожнюючи компакт U і клас α_U, можна сказати, що відображення ψ співпадає на конусі P сублінійних функцій з спряженістю Мінковського φ: якщо $p \in P$, то $\varphi(p) = \underline{\partial} p,\psi(p) = \alpha_{\underline{\partial} p}$.

За допомогою простору M можна визначити різницю опуклих компактів. Нехай $U,V \in N$. Розглянемо класи α_U,α_V. Різниця цих класів (у просторі M) $\alpha_U - \alpha_V$ обчислюється так:

$$\alpha_U - \alpha_V = [[U,\mathbf{0}]] - [[V,\mathbf{0}]] = [[U,\mathbf{0}]] + [[\mathbf{0}, -V]] = [[U, -V]].$$

Якщо ототожнити компакти U і V з класами α_U,α_V, то можна сказати, що різниця U і V є класом еквівалентних пар, що містить пару $[U, -V]$. Помітимо, що кожен елемент α простору M виражається через пари вигляду α_U і, тим самим, через елементи простору N. Дійсно, якщо $\alpha = [U,V]$, то

$$\alpha = [[U,\mathbf{0}]] + [[\mathbf{0},V]] = [[U,\mathbf{0}]] - (-1)[[\mathbf{0},V]] = [[U,\mathbf{0}]] - [[-V,\mathbf{0}]] = \alpha_U - \alpha_{-V}.$$

В зв'язку з цим M часто називають *простором опуклих множин* або, точніше, *простором опуклих компактів*.

Визначимо в просторі M відношення порядку. Нехай $\alpha_1,\alpha_2 \in M$, причому $\alpha_1 = \psi(l_1),\alpha_2 = \psi(l_2)$. Вважаємо, що $\alpha_1 \geq \alpha_2$ тоді і тільки тоді, коли $l_1 \geq l_2$. Припустимо, що $\alpha_1 = [[U_1,V_1]],\alpha_2 = [[U_2,V_2]]$; позначимо

$$p_{U_i}(x) = \max_{h \in U_i} \langle h,x \rangle, \quad q_{V_i}(x) = \min_{h \in V_i} \langle h,x \rangle,\ i = 1,2.$$

Тоді $l_1 = p_{U_1} + q_{V_1},l_2 = p_{U_2} + q_{V_2}$. Нерівність $l_1 \geq l_2$ рівносильна нерівності

$$p_{U_1} - q_{V_2} \geq p_{U_2} - q_{V_1}. \tag{6.2.19}$$

Сублінійна функція $(-q_{V_1})$ має субдиференціалом множину $-V_i,(i = 1,2)$, тому в силу спряженості Мінковського субдиференціали сублінійних функцій $p_{U_1} - q_{V_2}$ і $p_{U_2} - q_{V_1}$ збігаються відповідно з компактами $U_1 - V_2$ і $U_2 - V_1$, а нерівність (6.2.19) рівносильна включенню

$$U_1 - V_2 \supset U_2 - V_1. \tag{6.2.20}$$

Таким чином, нерівність $\alpha_1 \geq \alpha_2$, де $\alpha_i = [[U_i,V_i]]$, еквівалентна співвідношенню (6.2.20). Безпосередньо з визначення випливає, що відображення ψ зберігає порядок: нерівності $l_1 \geq l_2$ і $\psi(l_1) \geq \psi(l_2)$ рівносильні. Наведемо два твердження, які будуть використовуватися.

Теорема 6.2.2. *Нехай $l_1,\ldots,l_k \in L$, причому $\psi(l_i) = [[U_i,V_i]]$. Нехай $\overline{l}(x) = \max\limits_{i=1,\ldots,k}\{l_i(x)\}$, $\underline{l}(x) = \min\limits_{i=1,\ldots,k}\{l_i(x)\}$. Тоді $\psi(\overline{l}) = [[\overline{U},\overline{V}]]$, $\psi(\underline{l}) = [[\underline{U},\underline{V}]]$, де*

$$\overline{U} = \operatorname{conv}\bigcup_{i=1}^{k}\left(U_i - \sum_{j=1,\ldots,k;j\neq i} V_j\right), \quad \overline{V} = \sum_{i=1}^{k} V_i,$$

$$\underline{U} = \sum_{i=1}^{k} U_i, \quad \underline{V} = \operatorname{conv}\bigcup_{i=1}^{k}\left(V_i - \sum_{j=1,\ldots,k;j\neq i} U_i\right).$$

Доведення. Оскільки $\psi(l_i) = [[U_i,V_i]]$, то $l_i = p_i + q_i$, де

$$p_i(x) = \max_{h\in U_i}\langle h,x\rangle, \quad q_i(x) = \min_{h\in V_i}\langle h,x\rangle, \quad i = 1,\ldots,k.$$

Скориставшись формулами (6.2.7) і (6.2.8), які дають вирази $\overline{l}(x)$ та $\underline{l}(x)$ через $p_i,q_i,i = 1,\ldots,k$, та застосовуючи спряженість Мінковського, переконаємося в справедливості твердження. □

Теорема 6.2.3. *Нехай $U,V,W \in N$ і $U + V \supset W + V$. Тоді $U \supset W$.*

Доведення. Нехай p_U,p_V,p_W – опорні функції компактів U,V,W відповідно. Використовуючи спряженість Мінковського, маємо

$$p_U + p_V = p_{U+V} \geq p_{W+V} = p_W + p_V,$$

звідки випливає нерівність $p_U \geq p_W$, рівносильна необхідному включенню. □

Наслідок 6.2.1. *Якщо $U,V,W \in N$ і $U + V = W + V$, то $U = W$.*

Повернемося до визначення різниці на сукупності N опуклих компактів з простору $\mathbb{R}^n$. Один із способів визначення різниці, що базується на використанні операції віднімання в лінійному просторі опуклих множин, вказаний у попередньому пункті. При цьому віднімання є операцією, оберненою до додавання за Мінковським. Однак її результат вже не обов'язково лежить у вихідній сукупності N. Він належить більш широкому простору M. За самим визначенням віднімання в лінійному просторі для будь-яких $U,V \in N$ справедлива рівність

$$(U - V) + V = U. \tag{6.2.21}$$

Тут $U - V$ — елемент $[[U,-V]]$ простору опуклих множин, що є різницею елементів $[[U,\mathbf{0}]]$ і $[[V,\mathbf{0}]]$, яким належать множини U і V відповідно. Рівність (6.2.21) є символічним записом рівності

$$[[U,-V]] + [[V,\mathbf{0}]] = [[U,\mathbf{0}]].$$

Та обставина, що різниця, яка визначається на просторі опуклих множин, може виходити за межі сукупності N, є іноді незручною. Тому розглядаються й інші способи визначення різниці. При цьому під різницею розуміється операція, що визначена для довільних $U,V \in N$. Її результатом є елемент сукупності N або порожня множина. Вона обернена до додавання за Мінковським в наступному сенсі: якщо $U = V + W$, то різниця множин U і V повинна співпадати з W. Якщо ж множина U не може бути представленою у вигляді $V + W$, то різниця компактів U і V може залежати від способу визначення. (Різні визначення ведуть до різних результатів.) Іншими словами, виконання рівності вигляду (6.2.21) потрібне лише у тому випадку, коли $U = V + W$.

Розглянемо два способи визначення різниці.

1. Операція $\div$.
Нехай $U,V \in N$. Покладемо

$$U \div V = \{x | x + V \subset U\}.$$

Результатом операції $\div$ може бути і порожня множина, тим самим вона може виводити за межі N. Можна показати, що множина $U \div V$ є опуклою і компактною. Перевіримо, наприклад, опуклість. Припустимо, що $U \div V$ непорожня і нехай $x,y \in U \div V$. Тоді для будь-яких $z \in V$ маємо $x + z \in U, y + z \in U$, а тому при $0 \le \alpha \le 1$ виконується співвідношення

$$\alpha(x + z) + (1 - \alpha)(y + z) = \alpha x + (1 - \alpha)y + z \in U,$$

яке показує, що $\alpha x + (1 - \alpha)y \in U \div V$. Операцію $\div$ можна розглядати як віднімання. Точніше кажучи, справедливе наступне твердження.

Теорема 6.2.4. *Якщо $U = V + W$, то $U \div V = W$.*

Доведення. Нехай $x \in U \div V$, тобто $x + V \subset U = V + W$. Використовуючи попереднє твердження, одержимо, що $x \in W$. Якщо ж $x \in W$, то $x + V \subset W + V = U$, тобто $x \in U \div V$. □

Наведемо два приклади.

Приклад 6.2.1. Нехай $V \supset U$. Тоді множина $U \div V$ порожня. У той же час, $0 \in V \div U$.

Приклад 6.2.2. Нехай U – куля, V – діаметр цієї кулі. Тоді $U \div V = \{0\}$. Якщо V_1 – довільна підмножина кулі, що містить діаметр, то також $U \div V_1 = \{0\}$.

Покажемо, як операція $\div$ пов'язана з відношенням еквівалентності.

Теорема 6.2.5. *Нехай пари $[U, -V]$ і $[W, -Z]$ еквівалентні. Тоді $U \div V = W \div Z$.*

Доведення. Співвідношення $[U, -V] \approx [W, -Z]$ означає, що $U + Z = V + W$. Тому справедливі наступні імплікації:

$$x \in U \div V \Rightarrow x + V \subset U \Rightarrow x + V + Z \subset U + Z \Rightarrow x + V + Z \subset V + W.$$

Твердження 6.2.3 означає, що включення $x+V+Z \subset V+W$ рівносильне включенню $x+Z \subset W$, яке показує, що $x \in W \div Z$. Отже, $U \div V \subset W \div Z$. Аналогічно доводиться обернене включення. $\square$

Зауваження 6.2.1. Виразимо еквівалентність пар $[U,-V]$ і $[W,-Z]$ мовою простору L. Нехай елемент $l \in L$ такий, що $\psi(l) = [[U, -V]]$. Тоді

$$l(x) = \max_{h \in U}\langle h,x\rangle + \min_{h \in -V}\langle h,x\rangle = \max_{h \in U}\langle h,x\rangle - \max_{h \in V}\langle h,x\rangle.$$

Таким же чином, якщо $\psi(\tilde{l}) = [[W, -Z]]$, то $\tilde{l}(x) = \max\limits_{h \in W}\langle h,x\rangle - \max\limits_{h \in Z}\langle h,x\rangle$. Будемо ставити у відповідність парі компактів (A_1,A_2) функцію $l \in L$ за таким правилом:

$$l(x) = \max_{h \in A_1}(h,x) - \max_{h \in A_2}(h,x).$$

Із сказаного вище випливає, що еквівалентність пар $[U, -V]$ і $[W, -Z]$ рівносильна тому, що пари $[U,V]$ і $[W,Z]$ породжують за вказаним вище правилом ту ж функцію $l \in L$.

2. Операція $\dot{-}$.

Перш ніж визначити цю операцію, наведемо ряд фактів.

Нехай U – опуклий компакт і p_U – його опорна функція. Для $x \in \mathbb{R}^n$ покладемо

$$G_x(U) = \{h \in U | \langle h,x\rangle = \max_{g \in U}\langle x,g\rangle\}.$$

Множину $G_x(U)$ назвемо $\max$ *–границею* компакту U, породжену елементом x. Нехай p_U – опорна функція компакту U. Ця функція сублінійна і тому її субдиференціал $\underline{\partial} p_U(x)$ в точці x обчислюється за формулою $\underline{\partial} p_U(x) = \{h \in \partial p_U | \langle h,x\rangle = p_U(x)\}$. Іншими словами, $\max$–границя $G_x(U)$ співпадає з субдиференціалом $\underline{\partial} p_U(x)$. Якщо $G_x(U)$ складається з однієї точки, то p_U диференційовна в точці x і $G_x(U) = \{\nabla p_U(x)\}$. Опукла функція майже всюди диференційовна. Тому множина тих x, для яких $\max$-границя $G_x(U)$ складається більш ніж з однієї точки, має міру нуль.

Відзначимо, що сублінійна функція p (її субдиференціал) повністю відновлюється за значеннями градієнта $\nabla p(x)$ на деякій множині повної міри T, де цей градієнт існує. Дійсно, множина $\underline{\partial} p$ співпадає з субдиференціалом Кларка функції p в нулі, а тому $\underline{\partial} p$ співпадає з опуклим замиканням множини $\{\nabla f(x) | x \in T\}$.

Розглянемо опуклі компакти U,V і нехай T – деяка множина повної міри, у точках x якої існують градієнти $\nabla p_U(x)$ і $\nabla p_V(x)$. Розглянемо множину різниць $\{\nabla p_U(x) - \nabla p_V(x)\}$. Її замкнуту опуклу оболонку позначимо символом $U \dot{-} V$. Таким чином, за означенням,

$$U \dot{-} V = \mathrm{cl\,conv}\{\nabla p_U(x) - \nabla p_V(x) \colon\ x \in T\}.$$

Допускаючи деяку вільність, можна сказати, що різниця $U \dot{-} V$ визначається різницями точок з опуклих компактів U і V, у яких гіперплощина, яка ортогональна одному й тому ж вектору $x \in T$, "дотикається" до цих компактів.

Насамперед варто перевірити, що дане означення коректне, тобто $U \dot{-} V$ не залежить від вибору множини T, що має зазначені вище властивості. Нехай $T_{U,V}$ – сукупність усіх векторів x, в яких існують градієнти $\nabla p_U(x)$ і $\nabla p_V(x)$, і множина $T \subset T_{U,V}$ така, що її замикання містить $T_{U,V}$. Нехай

$$W_1 = \mathrm{cl\,conv}\{\nabla p_U(x) - \nabla p_V \colon\ x \in T_{U,V}\},$$

$$W_2 = \mathrm{cl\,conv}\{\nabla p_U(x) - \nabla p_V \colon\ x \in T\}.$$

Зрозуміло, що $W_1 \supset W_2$. З іншого боку, нехай $x \in T_{U,V}$ і $x_k \in T, x_k \to x$. Оскільки в точці x існує градієнт $\nabla p_U(x)$, то $\nabla p_U(x_k) \to \nabla p_U(x)$. Це легко випливає, наприклад, з напівнеперервності зверху субдиференціального відображення $y \mapsto \underline{\partial} p_U(y)$ і того факту, що $\underline{\partial} p_U(x) = \{\nabla p_U(x)\}$. Подібним же чином $\nabla p_V(x_k) \to \nabla p_V(x)$. Звідси випливає, що $\nabla p_U(x) - \nabla p_V(x) \in W_2$, а звідси слідує справедливість включення $W_1 \subset W_2$. Зі сказаного випливає, що якщо $T_1, T_2 \subset T_{U,V}$ і T_1, T_2 – множин повної міри, то

$$\mathrm{cl\,conv}\{\nabla p_U(x) - \nabla p_V \colon\ x \in T_1\} =$$

$$= \mathrm{cl\,conv}\{\nabla p_U(x) - \nabla p_V(x) \colon\ x \in T_2\}.$$

Отже, множина $U \dot{-} V$ визначена коректно.

За визначенням, множина $U \dot{-} V$ опукла і замкнута. Зрозуміло, що ця множина міститься в різниці Мінковського $U - V$ і тому є обмеженою. Таким чином, $U \dot{-} V$ – опуклий компакт. Це показує, що на відміну від операції $\div$ операція $\dot{-}$ завжди може бути здійсненою в сукупності N. Покажемо, що так само, як і $\div$, операція $\dot{-}$ є відніманням у тому сенсі, що $(V + W) \dot{-} W = V$. Точніше кажучи, справедливе наступне твердження.

Теорема 6.2.6. *Якщо $U = V + W$, то $U \dot{-} V = W$.*

Доведення. Нехай множина T складається з точок x, в яких існують градієнти опорних функцій p_V і p_W компактів V і W відповідно. В кожній точці $x \in T$ існує градієнт функції $p_V + p_W$, що співпадає з опорною функцією p_U компакта $U = V + W$. Маємо

$$U \dot{-} V = \mathrm{cl\,conv}\{\nabla p_U(x) - \nabla p_V(x) \colon\ x \in T\} =$$

$$= \mathrm{cl\,conv}\{\nabla p_V(x) + \nabla p_W(x) - \nabla p_V(x) \colon\ x \in T\} =$$

$$= \mathrm{cl\,conv}\{\nabla p_W(x):\ x \in T\}.$$

Як відмічалося вище, опукле замикання множини $\{\nabla p_W(x)|\ x \in T\}$ співпадає з субдиференціалом $\underline{\partial} p_W$ функції p_W, тобто з компактом W. □

Приклад 6.2.3. Розглянемо площину $\mathbb{R}^2$ з одиничними ортами e_1, e_2. Нехай U – одиничний круг з центром в нулі, $V = \{\lambda e_2 | \lambda \in [-1{,}1]\}$ – його діаметр. Покладемо $T = \{x = (x^{(1)}, x^{(2)})|\ x^{(2)} \neq 0\}$. Для всіх $x \in T$ max-границя $G_x(U)$ складається з точки x. Якщо $x^{(2)} > 0$, то $G_x(V) = \{e_2\}$, якщо $x^{(2)} < 0$, то $G_x(V) = \{-e_2\}$. Звідси легко випливає, що множина $U - V$ співпадає з описаним навколо круга квадратом $[-1{,}1] \times [-1{,}1]$. (Нагадаємо, що, як відзначалося вище, $U \dot{-} V = \{0\}$.)

Теорема 6.2.7. *Якщо пари* $[U, -V]$ *і* $[W, -Z]$ *еквівалентні, то* $U \dot{-} V = W \dot{-} Z$.

Доведення. Нехай множина T складається із всіх елементів x, у яких опорні функції p_U, p_V, p_W, p_Z компактів U, V, W, Z мають градієнти. За визначенням операції $\dot{-}$ одержимо, що компакти $U \dot{-} V$ і $W \dot{-} Z$ співпадають із замкнутою опуклою оболонкою множин $\{\nabla p_U(x) - \nabla p_V(x)|\ x \in T\}$ і $\{\nabla p_W(x) - \nabla p_Z(x)|\ x \in T\}$ відповідно. З еквівалентності пар $U \dot{-} V$ і $W \dot{-} Z$ випливає рівність $U + Z = V + W$. Використовуючи спряженість Мінковського, одержимо $p_U + p_Z = p_V + p_W$ чи, що те ж саме, $p_U - p_V = p_W - p_Z$. Із сказаного випливає рівність

$$\nabla p_U(x) - \nabla p_V(x) = \nabla p_W(x) - \nabla p_Z(x)\ \forall x \in T,$$

яка і показує, що $U \dot{-} V = W \dot{-} Z$. □

6.3. УМОВИ РЕГУЛЯРНОСТІ ДЛЯ МНОЖИН, ЯКІ ВИЗНАЧЕНІ КВАЗІДИФЕРЕНЦІЙОВНИМИ ФУНКЦІЯМИ

За допомогою квазідиференціалу вдається одержати умови регулярності, що легко перевіряються, для множин, які задаються нерівностями вигляду $\{x | h(x) \leq 0\}$ або рівностями вигляду $\{x | h(x) = 0\}$, де $h(x)$ – квазідиференційовна функція.

Надалі нам знадобиться поняття max–границі опуклого компакту V, що визначається точкою x. Нагадаємо, max–границя $G_x(V)$ опуклого компакту V, яка визначається точкою x, має вигляд

$$G_x(V) = \left\{ v \in V | \langle v, x\rangle = \max_{v' \in V} \langle v', x\rangle \right\}.$$

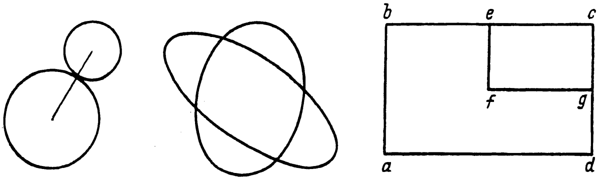


Рис. 6.3.1:

Вона співпадає з субдиференціалом $\partial p_V(x)$ опорної функції компакту V в точці x. Зокрема, якщо $x = 0$, то $G_x(V) = V$.

Означення 6.3.1. Впорядкована пара опуклих компактів $[U, V]$ *знаходиться у спільному стані*, якщо для кожного g max–границя $G_g(V)$ не міститься в max–границі $G_g(U)$.

Приклад 6.3.1. Якщо множини U,V не перетинаються, то пара $[U, V]$, як і пара $[V, U]$ знаходяться в спільному стані.

Приклад 6.3.2. Якщо $U \subset int\, V$, то пара $[U, V]$ знаходиться в спільному стані, а пара $[V, U]$ не знаходиться в спільному стані.

Приклад 6.3.3. Нехай U співпадає з одиничною кулею $B_1(0)$ (в евклідовій нормі), $V = B_{1/2}(x)$, де $\|x\| = 3/2$ (див. рис. 6.3.1, на якому ці множини зображені при $n = 2$). Тоді ні пара $[V, U]$, ні пара $[U, V]$ не знаходяться в спільному стані.

Приклад 6.3.4. Нехай множини U та V тілесні, їх межі перетинаються і в точках перетину вони не мають спільних опорних площин (див. рис. 6.3.1). Тоді як пара $[U, V]$, так і пара $[V, U]$ знаходяться в спільному стані.

Приклад 6.3.5. Нехай U співпадає з квадратом $abcd$, а V - з квадратом $ecgf$, зображеними на рис. 6.3.1. Тоді пара $[V, U]$ знаходиться в спільному стані, а пара $[U, V]$ не має цієї властивості.

Нехай $[U, V]$ - пара опуклих компактів і нехай функція $l_{[U,V]}$ визначена рівністю

$$l_{[U,V]}(g) = \max_{h \in U}\langle h,g\rangle + \min_{h \in V}\langle h,g\rangle =$$
$$= \max_{h \in U}\langle h,g\rangle - \min_{h \in -V}\langle h,g\rangle,\ g \in \mathbb{R}^n. \tag{6.3.1}$$

Функція $l_{[U,V]}$ належить простору L функцій, які можна подати у вигляді суми сублінійної і суперлінійної функцій, і є елементом цього простору, який відповідає при природному ізоморфізмі парі $[U, V]$.

Теорема 6.3.1. *Пара* $[U,-V]$ *знаходиться в спільному стані тоді і тільки тоді, коли для будь-якого* $g \in \mathbb{R}^n$ *знайдеться елемент* v *такий, що* $\left(l_{[U,V]}\right)'(g,v) < 0$.

Доведення. Нехай пара $[U,-V]$ знаходиться в спільному стані. Тоді для будь-якого елемента g знайдеться елемент $w \in G_g(-V)$ такий, що $w \notin G_g(U)$. Застосовуючи теорему про розділення, одержимо, що при деякому $v \in \mathbb{R}^n$ виконується нерівність

$$\langle w,v\rangle > \max_{w' \in G_g(U)} \langle w',v\rangle.$$

Тим більше

$$\max_{\tilde{w} \in G_g(-V)} \langle \tilde{w},v\rangle > \max_{w' \in G_g(U)} \langle w',v\rangle. \tag{6.3.2}$$

Як відзначалося вище, границя $G_g(U)$ компакту U співпадає з субдиференціалом $\underline{\partial} p_U(g)$ опорної функції p_U цього компакту в точці g. Тому

$$\max_{w' \in G_g(U)} \langle w',v\rangle = \max_{w' \in \underline{\partial} p_U(g)} \langle w',v\rangle = (p_U)'(g,v).$$

(Тут ми скористалися тим, що для опуклої функції p_U похідна за напрямком v відновлюється як максимум скалярних добутків v на елементи субдиференціалу.) Аналогічно

$$\max_{\tilde{w} \in G_g(-V)} \langle \tilde{w},v\rangle = (p_{-V})'(g,v),$$

де p_{-V} - опорна функція компакту $-V$. Нерівність (6.3.2) показує, що

$$\left(l_{[U,V]}\right)'(g,v) = (p_U - p_{-V})'(g,v) = (p_U)'(g,v) - (p_{-V})'(g,v) < 0.$$

Припустимо тепер, що для кожного g знайшовся напрямок v, при якому

$$\left(l_{[U,V]}\right)'(g,v) < 0.$$

Тоді справедлива нерівність $(p_U)'(g,v) < (p_{-V})'(g,v)$, яка є рівносильною нерівності (6.3.2). Остання показує, що $G_g(-V)$ не міститься в $G_g(U)$. $\square$

Теорема 6.3.1 дозволяє з'ясувати, як пов'язаний спільний стан з поняттям еквівалентності пар, яке було визначене для опуклих множин. Тому можна стверджувати, що якщо $[U_1,V_1]$ та $[U_2,V_2]$ – еквівалентні пари опуклих компактів, то пара $[U_1,-V_1]$ знаходиться у спільному стані тоді і тільки тоді, коли пара $[U_2,-V_2]$ має цю властивість. Дійсно, з еквівалентності $[U_1,V_1]$ та $[U_2,V_2]$ випливає, що побудовані за формулою (6.3.1) функції $l_{[U_1,V_1]}$ та $l_{[U_2,V_2]}$ співпадають. Тому властивість бути в спільному стані для пар $[U_1,-V_1]$ і $[U_2,-V_2]$ виконується чи не виконується одночасно.

Розглянемо множину $S = \{x|\ h(x) \le 0\}$, де h - квазідиференційовна локально ліпшицева в деякій точці x функція, причому $h(x) = 0$.

Покладемо

$$\gamma_1(x) = \{x|\, h'(x,g) < 0\}, \quad \Gamma_1(x) = \{x|\, h'(x,g) \le 0\}. \qquad (6.3.3)$$

Нагадаємо, що умова регулярності для множини S в точці x означає, що $\mathrm{cl}\,\gamma_1(x) = \Gamma_1(x)$. Якщо ця умова виконується, то конус Булігана $\Gamma(x, S)$ допускає просту інтерпретацію: $\Gamma(x, S) = \Gamma_1(x)$.

Теорема 6.3.2. *Нехай $\left[\underline{\partial} h(x), \overline{\partial} h(x)\right]$ – квазідиференціал функції h в точці x. Якщо пара $\left[\underline{\partial} h(x), -\overline{\partial} h(x)\right]$ знаходиться в спільному стані, то в точці x виконується умова регулярності.*

Доведення. Покладемо $U = \underline{\partial} h(x)$, $V = \overline{\partial} h(x)$. Тоді похідна $h'(x,g)$ співпадає з функцією $l_{[U,V]}$, що визначається формулою (6.3.1). Нехай $h'(x,g) = 0$. З теореми 6.3.1 випливає, що існує таке v, що $\left(l_{[U,V]}\right)'(g,v) < 0$. Тому при достатньо малих $\alpha > 0$ виконується нерівність

$$h'(x,g + \alpha v) = l_{[U,V]}(g + \alpha v) =$$
$$= l_{[U,V]}(g) + \alpha l'_{[U,V]}(g,v) + o(\alpha) <$$
$$< l_{[U,V]}(g) = h'(x,g) = 0.$$

Звідки $g + \alpha v \in \gamma_1(x)$. Спрямовуючи α до нуля, можна переконатися в тому, що $g \in \mathrm{cl}\gamma_1(x)$. □

Зауваження 6.3.1. Із результатів попереднього пункту випливає, що властивість пари $\left[\underline{\partial} f(x), -\overline{\partial} f(x)\right]$ бути в спільному стані зберігається при переході до еквівалентної пари. Тому вона визначається квазідиференціалом, як класом еквівалентних пар.

Розглянемо тепер множину S, що задається співвідношенням

$$S = \{x|\ h(x) = 0\}. \qquad (6.3.4)$$

Нехай конуси $\gamma_1(x)$ та $\Gamma_1(x)$ визначені рівностями (6.3.3). Розглянемо, конуси

$$\gamma_2(x) = \{g|\ h'(x,g) > 0\}, \quad \Gamma_2(x) = \{g|\ h'(x,g) \ge 0\}.$$

Нагадаємо, що для множини вигляду (6.3.4) умова регулярності в точці x має вигляд:

$$\mathrm{cl}\,\gamma_1(x) = \Gamma_1(x), \quad \mathrm{cl}\,\gamma_2(x) = \Gamma_2(x).$$

Теорема 6.3.3. *Нехай локально ліпшицева функція h квазідиференційовна в точці $x \in S$, де S задається формулою (6.3.4). Тоді, якщо пари $\left[\underline{\partial} h(x), \overline{\partial} h(x)\right]$ та $\left[\overline{\partial} h(x), -\underline{\partial} h(x)\right]$ знаходяться в спільному стані, то в точці x виконується умова регулярності.*

Доведення. Справедливість теореми випливає з теореми 6.3.2. Спочатку застосовуємо її до функції h, а потім до функції $-h$. □

6.4. УМОВИ ЭКСТРЕМУМУ КВАЗІДИФЕРЕНЦІЙОВНОЇ ФУНКЦІЇ

6.4.1. Умови экстремуму квазідиференційовної функції на $\mathbb{R}^n$

Нехай функція f задана і квазідиференційовна на $\mathbb{R}^n$. Нехай $\mathfrak{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)]$– квазідиференціал функції f в точці x. З лем 4.5.1 і 4.5.2 випливає наступна теорема.

Теорема 6.4.1. *Для того, щоб функція $f(x)$ досягала в точці x^* свого найменшого на $\mathbb{R}^n$ значення, необхідно, щоб*

$$-\overline{\partial}f(x^*) \subset \underline{\partial}f(x^*). \tag{6.4.1}$$

Якщо

$$-\overline{\partial}f(x^*) \subset int\ \underline{\partial}f(x^*), \tag{6.4.2}$$

то x^*— точка строгого локального мінімуму функції f на $\mathbb{R}^n$. Для того, щоб функція f досягала в точці x^{**} свого найбільшого значення, необхідно, щоб

$$-\underline{\partial}f(x^{**}) \subset \overline{\partial}f(x^{**}). \tag{6.4.3}$$

Умова

$$-\overline{\partial}f(x^{**}) \subset int\ \underline{\partial}f(x^{**}) \tag{6.4.4}$$

є достатньою умовою строго локального максимуму функції f на $\mathbb{R}^n$.

Доведення. Доведемо перше твердження. Якщо функція $f(x)$ диференційовна за напрямками в точці x^* і x^*– точка мінімуму функції $f(x)$ на $\mathbb{R}^n$, то

$$f'(x^*,g) \geq 0 \quad \forall g \in \mathbb{R}^n. \tag{6.4.5}$$

Оскільки в даному випадку

$$f'(x^*,g) = \max_{v\in\underline{\partial}f(x^*)}\langle v,g\rangle + \min_{w\in\overline{\partial}f(x^*)}\langle w,g\rangle. \tag{6.4.6}$$

то з (6.4.5), (6.4.6) маємо

$$\inf_{g\in B_1(0)}\left[\max_{v\in\underline{\partial}f(x^*)}\langle v,g\rangle + \min_{w\in\overline{\partial}f(x^*)}\langle w,g\rangle\right] = 0. \tag{6.4.7}$$

де $B_1(0) = \{g \in \mathbb{R}^n \colon \|g\| \leq 1\}$.

З (6.4.7) випливає, що

$$\min_{g\in B_1(0)}\min_{w\in\overline{\partial}f(x^*)}\max_{v\in\underline{\partial}f(x^*)}\langle v+w,g\rangle =$$

$$\min_{g\in B_1(0)}\min_{w\in\overline{\partial}f(x^*)}\max_{v\in\underline{\partial}f(x^*)+w}\langle v,g\rangle = 0. \tag{6.4.8}$$

Оскільки множини $B_1(0)$ та $\overline{\partial} f(x^*)$ опуклі та компакні, то з (6.4.8) маємо

$$\min_{w\in\overline{\partial} f(x^*)}\ \min_{g\in B_1(0)}\ \max_{v\in\underline{\partial} f(x^*)+w} \langle v,g\rangle = 0, \tag{6.4.9}$$

тобто

$$\min_{g\in B_1(0)}\ \max_{v\in\underline{\partial} f(x^*)+w} \langle v,g\rangle = 0 \quad \forall w\in\overline{\partial} f(x^*). \tag{6.4.10}$$

Умова (6.4.10) еквівалентна умові

$$0\in\underline{\partial} f(x^*)+w \quad \forall w\in\overline{\partial} f(x^*). \tag{6.4.11}$$

Ця умова означає, що

$$-w\in\underline{\partial} f(x^*) \quad \forall w\in\overline{\partial} f(x^*), \tag{6.4.12}$$

тобто

$$-\overline{\partial} f(x^*)\subset\underline{\partial} f(x^*). \tag{6.4.13}$$

□

Зауваження 6.4.1. Теорема 6.4.1 справедлива і для випадку оптимізації f на множині $S\subset\mathbb{R}^n$, якщо $x^*\in\operatorname{int} S$.

Нехай $x\in\mathbb{R}^n$ не є inf–стаціонарною точкою функції f на $\mathbb{R}^n$ (тобто умова (6.4.1) не виконується). Візьмемо $w\in\underline{\partial} f(x)$ і обчислимо

$$\min_{v\in\underline{\partial} f(x)} \|v+w\| = \|v(w)+w\| = \rho_1(w).$$

Оскільки $\underline{\partial} f(x)$— опуклий компакт, то $v(w)$ — єдина точка. Знайдемо $\max\limits_{w\in\overline{\partial} f(x)} \rho_1(w) = \rho_1(w(x))$. Точка $w(x)$ не обов'язково визначається однозначно. Оскільки x не є inf–стаціонарною точкою, то $\rho_1(w(x))>0$. Напрямок

$$g_1(x) = -\frac{v(w(x))+w(x)}{\|v(w(x))+w(x)\|} \tag{6.4.14}$$

є напрямком найшвидшого спуску функції f на $\mathbb{R}^n$ в точці x. При цьому $f'(x,g_1(x)) = -\rho_1(w(x))$. Таких напрямків може бути багато (їх множина не обов'язково опукла).

Нехай тепер $x\in\mathbb{R}^n$ не є sup–стаціонарною точкою функції f на $\mathbb{R}^n$ (тобто (6.4.3) не має місця). Візьмемо $v\in\underline{\partial} f(x)$ і знайдемо

$$\min_{w\in\overline{\partial} f(x)} \|v+w\| = \|v+w(v)\| = \rho_2(v),$$

$$\max_{v\in\underline{\partial} f(x)} \rho_2(v) = \rho_2(v(x)). \tag{6.4.15}$$

Напрямок

$$g_2(x) = \frac{v(x)+w(v(x))}{\|v(x)+w(v(x))\|} = \frac{v(x)+w(v(x))}{\rho_2(v(x))}$$

є напрямком найшвидшого підйому функції f на $\mathbb{R}^n$ в точці x і при цьому $f'(x,g_2(x)) = \rho_2(v(x))$. Таких напрямків також може бути багато.

Приклад 6.4.1. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2, x_0 = (0,0), f(x) = (|x^{(1)}| - |x^{(2)}| + 1)(|x^{(1)}| + 2|x^{(2)}| + 1)$. Знайдемо $\mathfrak{D}f(x_0)$. Маємо $f(x) = f_1(x)f_2(x)$, де

$$f_1(x) = |x^{(1)}| - |x^{(2)}| + 1, \quad f_2(x) = |x^{(1)}| + 2|x^{(2)}| + 1.$$

За правилами обчислення квазідиференціалів маємо

$$\mathfrak{D}f_1(x_0) = \left[\underline{\partial} f_1(x_0), \overline{\partial} f_1(x_0)\right],$$

$$\mathfrak{D}f_2(x_0) = \left[\underline{\partial} f_2(x_0), \overline{\partial} f_2(x_0)\right],$$

$$\underline{\partial} f_1(x_0) = \operatorname{conv}\{(1,0),(-1,0)\}, \quad \overline{\partial} f_1(x_0) = \operatorname{conv}\{(0,1),(0,-1)\},$$

$$\underline{\partial} f_2(x_0) = \operatorname{conv}\{(1,2),(1,-2),(-1,2),(-1,-2)\}, \ \overline{\partial} f_2(x_0) = \{(0,0)\}.$$

За правилами квазідиференціального числення

$$\mathfrak{D}f(x) = f_2(x)\mathfrak{D}f_1(x) + f_1(x)\mathfrak{D}f_2(x),$$

а $f_1(x_0) = f_2(x_0) = 1$, тому

$$\mathfrak{D}f(x_0) = \mathfrak{D}f_1(x_0) + \mathfrak{D}f_2(x_0) = \left[\underline{\partial} f(x_0), \overline{\partial} f(x_0)\right],$$

$$\underline{\partial} f(x_0) = \underline{\partial} f_1(x_0) + \underline{\partial} f_2(x_0) =$$

$$\operatorname{conv}\{(2,2),(2,-2),(0,2),(0,-2),(-2,2),(-2,-2)\},$$

$$\overline{\partial} f(x_0) = \overline{\partial} f_1(x_0) + \overline{\partial} f_2(x_0) = \operatorname{conv}\{(0,1),(0,-1)\}.$$

З рис. 6.4.1 видно, що $-\overline{\partial} f(x_0) \subset \operatorname{int} \underline{\partial} f(x_0)$, тобто точка x_0 є inf–стаціонарною точкою функції f. Умова (6.4.3) не виконується, при цьому існує чотири розв'язки задачі (6.4.15):

$$v_1(x_0) = (2,2) \quad \text{при} \quad w(v_1(x_0)) = (0,1),$$

$$v_2(x_0) = (-2,2) \quad \text{при} \quad w(v_2(x_0)) = (0,1),$$

$$v_3(x_0) = (-2,-2) \quad \text{при} \quad w(v_3(x_0)) = (0,-1),$$

$$v_4(x_0) = (2,-2) \quad \text{при} \quad w(v_4(x_0)) = (0,-1),$$

і, відповідно, чотири напрямки найскорішого підйому

$$g_1(x_0) = \frac{v_1(x_0) + w(v_1(x_0))}{\|v_1(x_0) + w(v_1(x_0))\|} = (2\Big/\sqrt{13}, 3\Big/\sqrt{13}),$$

$$g_2(x_0) = (-2/\sqrt{13}, 3/\sqrt{13}),$$

$$g_3(x_0) = (-2/\sqrt{13}, -3/\sqrt{13}),$$

$$g_4(x_0) = (2/\sqrt{13}, -3/\sqrt{13}).$$

Швидкість найшвидшого підйому $f'(x_0, g_i(x_0)) = \sqrt{13}$, $i = 1, \ldots, 4$.

Зауваження 6.4.2. Умова (6.4.1) може бути записана у вигляді

$$-w \in \underline{\partial} f(x^*) \quad \forall w \in \overline{\partial} f(x^*)$$

або

$$0 \in \underline{\partial} f(x^*) + w \quad \forall w \in \overline{\partial} f(x^*).$$

Звідси

$$0 \in \bigcap_{w \in \overline{\partial} f(x^*)} [\underline{\partial} f(x^*) + w] \equiv L'(x^*). \tag{6.4.16}$$

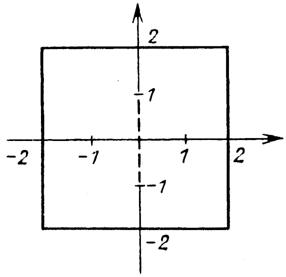


Рис. 6.4.1:

Можна показати, що хоча умови (6.4.1) і (6.4.16) еквівалентні, проте, якщо (6.4.16) не виконується для деякої точки x, то за допомогою (6.4.16) не вдається одержати не тільки напрямок найшвидшого спуску, але й простий напрямок спуску (може трапитися, що множина $L'(x)$ порожня).

Приклад 6.4.2. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2, x_0 = (0,0), f(x) = \left|x^{(1)}\right| - \left|x^{(2)}\right|$. Тоді $\mathfrak{D} f(x_0) = \left[\underline{\partial} f(x_0), \overline{\partial} f(x_0)\right]$, де

$$\underline{\partial} f(x_0) = \operatorname{conv}\{(1,0),(-1,0)\}, \quad \overline{\partial} f(x_0) = \operatorname{conv}\{(0,1),(0,-1)\}.$$

Очевидно, що множина

$$L'(x_0) = \bigcap_{w \in \overline{\partial} f(x_0)} [\underline{\partial} f(x_0) + w]$$

порожня. Таким чином, умова (6.4.16) в точці, що не є inf–стаціонарною, не дає інформації про поведінку функції в точці x_0 (дійсно, ми встановлюємо, що точка x_0 не є inf–стаціонарною, але нічого про напрямки спуску дізнатися не можемо).

Приклад 6.4.3. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2, x_0 = (0,0)$,

$$f(x) = \min\left\{|x^{(1)}| + |x^{(2)}|, |x^{(1)}| - |x^{(2)}|\right\}.$$

Ця функція розглядалась в прикладі 6.1.14. Там було показано, що

$$\mathfrak{D} f(x) = [\underline{\partial} f(x), \overline{\partial} f(x)],$$

$$\underline{\partial} f(x_0) = \operatorname{conv}\{(2,1),(-2,1),(-2,-1),(2,-1)\},$$
$$\overline{\partial} f(x_0) = \operatorname{conv}\{(1,2),(-1,2),(-1,-2),(1,-2)\}.$$

Очевидно, що в силу симетричності $\overline{\partial} f(x_0) = -\overline{\partial} f(x_0)$. Існує два напрямки найшвидшого спуску

$$g_1 = (0,1), \quad g_2 = (0,-1).$$

і два напрямки найшвидшого підйому

$$g_3 = (1,0), \quad g_4 = (-1,0).$$

Очевидно, що

$$f'(x_0, g_i(x_0)) = -1,\ i = 1,2; \quad f'(x_0, g_i(x_0)) = 1,\ i = 3,4.$$

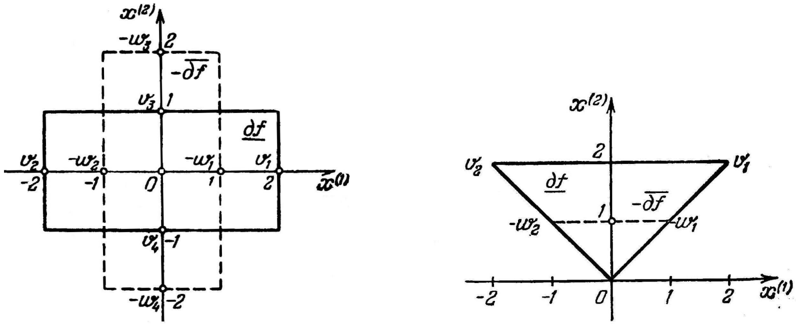


Рис. 6.4.2:

Приклад 6.4.4. Нехай $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2, x_0 = (0,0)$, $f(x) = ||x^{(1)}| + x^{(2)}|$. Ця функція розглядалася у прикладі 6.1.15. Там було показано, що

$$\mathfrak{D} f(x) = [\underline{\partial} f(x), \overline{\partial} f(x)],$$

$$\underline{\partial} f(x_0) = \operatorname{conv}\{(0,0),(2,2),(-2,2)\}, \quad \overline{\partial} f(x_0) = \operatorname{conv}\{(-1,-1),(1,-1)\}.$$

Тоді $-\overline{\partial} f(x_0) = \operatorname{conv}\{(1,1),(-1,1)\}$.

Точка x_0 є inf–стаціонарною. Існує два напрямки найшвидшого підйому

$$g_1 = (\sqrt{2}/2, \sqrt{2}/2), \quad g_2 = (-\sqrt{2}/2, \sqrt{2}/2).$$

При цьому швидкість найшвидшого підйому $f'(x_0, g_i(x_0)) = \sqrt{2}$, $i = 1,2$.

6.4.2. Квазідиференційовні множини та умови оптимальності.

Означення 6.4.1. Множина $S \subset \mathbb{R}^n$ називається *квазідиференційовною*, якщо вона може бути представлена у вигляді

$$S = \{x \in \mathbb{R}^n | h(x) \leq 0\}, \tag{6.4.17}$$

де $h(x)$ - квазідиференційовна функція.

Візьмемо $x \in S$. Розглянемо конуси

$$\gamma_1(x) = \{g \in \mathbb{R}^n | h'(x,g) < 0\},$$

$$\Gamma_1(x) = \{g \in \mathbb{R}^n | h'(x,g) \leq 0\}.$$

Нехай точка x така, що $h(x) = 0$. Кажуть, що в точці x виконується *умова регулярності*, якщо

$$\mathrm{cl}\ \gamma_1(x) = \Gamma_1(x) \tag{6.4.18}$$

Через $\Gamma(x)$ позначимо конус можливих напрямків множини S в точці $x \in S$. Вище вже було відзначено, що якщо $h(x) < 0$, то $\Gamma(x) = \mathbb{R}^n$, а якщо $h(x) = 0$ і виконується умова регулярності (6.4.18), то

$$\Gamma(x) = \Gamma_1(x). \tag{6.4.19}$$

Розглянемо задачу мінімізації на множині S квазідиференційовної на відкритій множині $X \supset S$ функції f. Нехай $x^* \in S$ – точка, яка підозріла на мінімум. Якщо $h(x^*) < 0$ (тоді $x^* \in \operatorname{int} S$), то (див. теорему 6.4.1 і зауваження 6.4.1) має місце умова

$$f'(x,g) = \max_{v \in \partial f(x)} \langle v,g \rangle. \tag{6.4.20}$$

Тому розглянемо випадок $h(x^*) = 0$.

Теорема 6.4.2. *Нехай функції $f(x)$ і $h(x)$ ліпшицеві і квазідиференційовні в деякому околі точки $x^* \in S$ і $h(x^*) = 0$. Припустимо також, що в точці x^* виконується умова регулярності (6.4.18). Для того, щоб функція f досягала в точці x^* свого найменшого на S значення, необхідно, щоб*

$$(\underline{\partial} f(x^*) + w) \cap [-\,\mathrm{cl}\ (\mathrm{cone}\ (\underline{\partial} h(x^*) + w'))] \neq \emptyset \tag{6.4.21}$$

для всіх $w \in \overline{\partial} f(x^*)$, $w' \in \overline{\partial} h(x^*)$.

Доведення. За лемою 4.5.5

$$f'(x^*,g) \geq 0 \quad \forall g \in \Gamma(x^*).$$

З (6.4.6) і (6.4.19)

$$\min_{w \in \overline{\partial} f(x^*)} \left[\max_{v \in \underline{\partial} f(x^*)} \langle v + w,g \rangle \right] \geq 0 \quad \forall g \in \Gamma_1(x^*). \tag{6.4.22}$$

Оскільки
$$\Gamma_1(x^*) = \left\{ g \in \mathbb{R}^n | \min_{w' \in \overline{\partial} h(x^*)} \left[\max_{v' \in \underline{\partial} h(x^*)} \langle v' + w', g \rangle \right] \le 0 \right\} = \cup_{w' \in \overline{\partial} f(x^*)} \Gamma_{1w'},$$
де
$$\Gamma_{1w'} = \left\{ g \in \mathbb{R}^n | \max_{v \in [\underline{\partial} h(x^*) + w']} \langle v, g \rangle \le 0 \right\},$$
то (6.4.22) еквівалентна умові
$$\max_{v \in [\underline{\partial} f(x^*) + S]} \langle v, g \rangle \ge 0 \quad \forall g \in \Gamma_{1w'} \tag{6.4.23}$$
для всіх $w \in \overline{\partial} f(x^*)$, $w' \in \overline{\partial} h(x^*)$. Застосовуючи тепер лему 5.4.2, одержуємо
$$(\underline{\partial} f(x^*) + w) \cap \Gamma_{1w'}^+ \ne \emptyset \quad \forall w \in \overline{\partial} f(x^*), \quad \forall w' \in \overline{\partial} h(x^*).$$
Залишилося помітити, що
$$\Gamma_{1w'}^+ = -\operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x^*) + w'\right)\right).$$
□

Теорема 6.4.3. *Умова (6.4.21) еквівалентна умові*
$$-\overline{\partial} f(x^*) \subset L(x^*), \tag{6.4.24}$$
де
$$L(x) = \bigcap_{w \in \overline{\partial} h(x)} \left[\underline{\partial} f(x) + \operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x) + w\right)\right)\right]. \tag{6.4.25}$$

Доведення. Нехай виконується умова (6.4.21). Візьмемо $w \in \overline{\partial} f(x^*)$, $w' \in \overline{\partial} h(x^*)$. З (6.4.21) випливає, що існує такий вектор $v \in \underline{\partial} f(x^*)$, що
$$v + w \in \left[-\operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x^*) + w'\right)\right)\right].$$
Звідки
$$-w \in \left[v + \operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x^*) + w'\right)\right)\right] \subset \underline{\partial} f(x^*) + \operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x^*) + w'\right)\right).$$
Оскільки вказане співвідношення справджується для кожного вектора $w' \in \overline{\partial} h(x^*)$, то
$$-w \in \bigcap_{w' \in \overline{\partial} h(x^*)} \left[\underline{\partial} f(x^*) + \operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x^*) + w'\right)\right)\right].$$
Оскільки вказане співвідношення справджується для кожного вектора $w \in \overline{\partial} f(x^*)$, то справедливе співвідношення (6.4.24).

Нехай справджується співвідношення (6.4.24). Тоді для всіх $w \in \overline{\partial} f(x^*)$ та всіх $w' \in \overline{\partial} h(x^*)$ справедливе співвідношення
$$-w \in \left[\underline{\partial} f(x^*) + \operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x^*) + w'\right)\right)\right].$$
Звідки
$$0 \in \left[\underline{\partial} f(x^*) + w + \operatorname{cl}\left(\operatorname{cone}\left(\underline{\partial} h(x^*) + w'\right)\right)\right].$$

Тобто

$$[\underline{\partial} f(x^*) + w] \cap [\mathrm{cl}\ (-\,\mathrm{cone}\ (\underline{\partial} h(x^*) + w'))] \neq \emptyset.$$

$\square$

Відзначимо, що $L(x)$ — опукла множина. Вона непорожня, оскільки

$$\underline{\partial} f(x) \subset L(x). \tag{6.4.26}$$

Неважко бачити, що умова (4.5.22) еквівалентна умові

$$-\overline{\partial} f(x^*) \subset \mathrm{int}\, L(x^*), \tag{6.4.27}$$

Нехай $x \in S$, $h(x) = 0$. Якщо точка x не є inf–стаціонарною, то знайдемо

$$\min_{\substack{z \in \underline{\partial} f(x) + w \\ z' \in \mathrm{cl}\ (\mathrm{cone}\ (\underline{\partial} h(x) + w'))}} \|z + z'\| = \|z(w,w') + z'(w,w')\| = d(w,w').$$

Тепер нехай

$$\rho(x) = \max_{w \in \overline{\partial} f(x), w' \in \overline{\partial} h(x)} d(w,w') = d(w_0,w_0').$$

Оскільки умова (6.4.21) не виконана (x – не є inf–стаціонарною точкою), то $\rho(x) > 0$.

Теорема 6.4.4. *Якщо $h(x) = 0$ і виконана умова регулярності (6.4.18), то напрямок*

$$g_0 = -\frac{v_0 + w(v_0)}{\|v_0 + w(v_0)\|}$$

є напрямком найшвидшого спуску функції f на множині S в точці x, причому

$$f'(x,g_0) = \min_{g \in \Gamma(x), \|g\|=1} f'(x,g) = -\|v_0 + w(v_0)\| = -\rho(x).$$

Використаємо тепер умову (6.4.24). Якщо $x \in S$ не є inf–стаціонарною точкою, тобто $-\overline{\partial} f(x) \not\subset L(x)$, то знайдемо напрямок

$$g = -\frac{v(x) + w(v(x))}{\|v(x) + w(v(x))\|},$$

де

$$v(x) = \arg \max_{v \in \overline{\partial} f(x)} \rho(v), \quad \rho(v) = \min_{w \in L(x)} \|v + w\| = \|v + w(v)\|.$$

Однак цей напрямок може не бути допустимим (див. приклад 6.4.5 нижче).

Отже, і умова (6.4.21), і умова (6.4.24) дозволяють одержати напрямки найшвидшого спуску, якщо точка не є inf–стаціонарною. Грубо кажучи, ці умови еквівалентні не тільки в inf-стаціонарних точках, але й в інших точках (в тому розумінні, що ми одержуємо за їх допомогою ту саму інформацію про напрямки найшвидшого спуску).

Зауваження 6.4.3. Як і в зауваженні 6.4.2, можна показати, що умова (6.4.21) еквівалентна умові

$$0 \in \bigcap_{w \in \overline{\partial} f(x^*), w' \in \overline{\partial} h(x^*)} [\underline{\partial} f(x^*) + w + \mathrm{cl}\ (\mathrm{cone}\ (\underline{\partial} h(x^*) + w'))] \equiv L'(x^*). \tag{6.4.28}$$

Якщо точка $x \in S$ не є inf–стаціонарною (тобто $0 \notin L'(x)$) і якщо множина $L'(x)$ непорожня, то можна знайти

$$\min_{z \in L'(x)} \|z\| = \|z(x)\|.$$

Однак, напрямок $g(x) = z(x)/\|z(x)\|$ не є, взагалі кажучи, не тільки напрямком найшвидшого спуску, але навіть і можливим напрямком. Як і в прикладі 6.4.2, множина $L'(x)$ може виявитися порожньою.

Зауваження 6.4.4. Покладемо для $x \in S$, $w' \in \overline{\partial} h(x)$

$$A(w',x) = -\,\mathrm{cl}\ (\mathrm{cone}\ (\underline{\partial} h(x^*) + w'))\,.$$

Якщо $\mathrm{int}\, A(w',x) \neq \emptyset$, то (див. лему 5.4.3 і зауваження 5.4.3) можна показати, що умова

$$(\underline{\partial} f(x^*) + w) \cap A(w',x) \neq \emptyset, \quad w \in \overline{\partial} f(x^*)$$

еквівалентна умові

$$0 \in \mathrm{conv}\ \{(\underline{\partial} f(x^*) + w) \cup (\underline{\partial} h(x^*) + w')\} \equiv \mathfrak{L}(x^*,w,w'). \tag{6.4.29}$$

Справджується наступна лема.

Лема 6.4.1. *Якщо для всіх* $w' \in \overline{\partial} h(x^*)$

$$\mathrm{int}\ (\mathrm{cone}\ (\underline{\partial} h(x^*) + w')) \neq \emptyset, \tag{6.4.30}$$

то умова (6.4.21) еквівалентна умові

$$0 \in \mathfrak{L}(x^*,w,w') \quad \forall w \in \overline{\partial} f(x^*), \quad \forall w' \in \overline{\partial} h(x^*) \tag{6.4.31}$$

Нагадаємо, що тут розглядається випадок $h\,(x^*) = 0$.
Нехай для точки $x \in S$ такої, що $h(x) = 0$, і для всіх $w' \in \overline{\partial} h\,(x)$ виконана умова (6.4.30). Знайдемо

$$\min_{z \in \mathfrak{L}(x,w,w')} \|z\| = \|z(x,w,w')\|$$

і

$$\max_{w \in \overline{\partial} f(x), w' \in \overline{\partial} h(x)} \|z(x,w,w')\| = \|z(x,w(x),w'(x))\|. \tag{6.4.32}$$

Тоді напрямок

$$g(x) = -\frac{z(x,w(x),w'(x))}{\|z(x,w(x),w'(x))\|} \tag{6.4.33}$$

є напрямком спуску функції f на S в точці x, і при цьому цей напрямок допустимий, тобто при достатньо малих α буде $[x + \alpha g(x)] \in S$.

Зауваження 6.4.5. Умову (6.4.31) зручно перевіряти в тому випадку, коли $\overline{\partial} f(x)$ і $\overline{\partial} h(x)$ – опуклі оболонки скінченної кількості точок. Тоді потрібно перевірити (6.4.31) лише для точок w,w', які є вершинами $\overline{\partial} f(x)$ і $\overline{\partial} h(x)$ відповідно.

Зауваження 6.4.6. Умова (6.4.30) виконана, наприклад, у тому випадку, коли

$$\operatorname{int} \underline{\partial} h(x^*) \neq \emptyset. \tag{6.4.34}$$

Але умова (6.4.30) може бути виконана і тоді, коли умова (6.4.34) не виконується.

Приклад 6.4.5. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $x_0 = (0,0)$, $h(x) = -|x^{(1)}| + 2x^{(1)} + |x^{(2)}|$. Маємо

$$\mathfrak{D} h(x_0) = \left[\underline{\partial} h(x_0), \overline{\partial} h(x_0)\right],$$

де (див. рис. 6.4.3)

$$\underline{\partial} h(x_0) = \operatorname{conv}\{(0,1),(0,-1)\}, \quad \overline{\partial} h(x_0) = \operatorname{conv}\{(1,0),(3,0)\}.$$

Очевидно, що умова (6.4.30) виконана. На рис. 6.4.3 зображені конуси

$$\operatorname{cone}\left(\underline{\partial} h(x_0) + (1,0)\right), \quad \operatorname{cone}\left(\underline{\partial} h(x_0) + (2,0)\right), \quad \operatorname{cone}\left(\underline{\partial} h(x_0) + (3,0)\right).$$

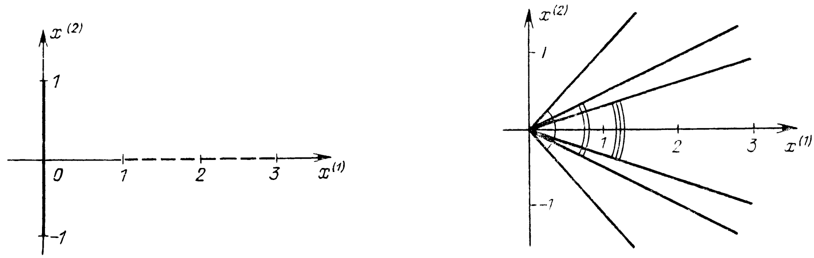


Рис. 6.4.3:

Приклад 6.4.6. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2$, $x_0 = (0,0)$, $h(x) = -|x^{(1)}| + |x^{(2)}|$. Маємо

$$\mathfrak{D} h(x_0) = \left[\underline{\partial} h(x_0), \overline{\partial} h(x_0)\right],$$

де

$$\underline{\partial} h(x_0) = \operatorname{conv}\{(0,1),(0,-1)\}, \quad \overline{\partial} h(x_0) = \operatorname{conv}\{(-1,0),(1,0)\}.$$

Ясно, що при $S_0 = (0,0) \in \overline{\partial} h(x_0)$ буде

$$\operatorname{cone}\left(\underline{\partial} h(x_0) + S_0\right) = \left\{x = (0,x^{(2)}) | x^{(2)} \in \mathbb{R}^1\right\} \equiv P.$$

Оскільки $\operatorname{int} P = \emptyset$, то умова (6.4.30) не виконана.

Проаналізуємо детальніше умову (6.4.31). Доведемо спочатку наступні леми.

Лема 6.4.2. *Нехай $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^n$ – опуклі множини. Якщо*

$$0 \in \operatorname{conv}\{(\mathrm{A} + w_1) \cup \mathrm{B}\}, \quad 0 \in \operatorname{conv}\{(\mathrm{A} + w_2) \cup \mathrm{B}\}, \tag{6.4.35}$$

то для будь-якого $\alpha \in [0,1]$

$$0 \in \operatorname{conv}\{(\mathrm{A} + w_\alpha) \cup \mathrm{B}\}, \tag{6.4.36}$$

де $w_\alpha = \alpha w_1 + (1-\alpha) w_2$.

Доведення. З умови (6.4.35) випливає, що існують $\alpha_1 \in [0,1], \alpha_2 \in [0,1]$ і $a_1, a_2 \in \mathrm{A}; b_1, b_2 \in \mathrm{B}$ такі, що

$$\alpha_1(a_1 + w_1) + (1-\alpha_1) b_1 = 0, \tag{6.4.37}$$

$$\alpha_2(a_2 + w_2) + (1-\alpha_2) b_2 = 0. \tag{6.4.38}$$

Розглянемо більш складний випадок $\alpha \in (0,1), \alpha_1 \in (0,1), \alpha_2 \in (0,1)$ (якщо одне чи декілька з чисел $\alpha, \alpha_1, \alpha_2$ рівні 0 чи 1, то доведення спрощується). Помножимо рівність (6.4.37) на γ_1, рівність (6.4.38) помножимо на γ_2 і додамо отримані рівності. Одержимо

$$\begin{gathered}(\alpha_1\gamma_1 a_1 + \alpha_2\gamma_2 a_2) + (\alpha_1\gamma_1 w_1 + \alpha_2\gamma_2 w_2) + \\ +(1-\alpha_1)\gamma_1 b_1 + (1-\alpha_2)\gamma_2 b_2 = 0.\end{gathered} \tag{6.4.39}$$

Спробуємо підібрати $\beta \in [0,1], \beta_1 \in [0,1], \gamma_1$ і γ_2 так, щоб (6.4.39) мало вигляд

$$\beta\,(a_\alpha + w_\alpha) + (1-\alpha)(\beta_1 b_1 + (1-\beta_1) b_2) = 0. \tag{6.4.40}$$

Тут $a_\alpha = \alpha a_1 + (1-\alpha) a_2$. Прирівнюючи коефіцієнти при a_1, a_2, b_1, b_2, w_1 і w_2, в (6.4.39) і (6.4.40), маємо рівності

$$\alpha_1\gamma_1 = \alpha\beta, \quad (1-\alpha_1)\gamma_1 = (1-\beta)\beta_1,$$

$$\alpha_2\gamma_2 = (1-\alpha)\beta, \quad (1-\alpha_2)\gamma_2 = (1-\beta)(1-\beta_1).$$

Розв'язуючи цю нелінійну систему 4 рівнянь з 4 невідомими, одержимо

$$\beta_1 = \frac{(1-\alpha_1)\alpha\alpha_2}{(1-\alpha_1)\alpha\alpha_2 + (1-\alpha_2)(1-\alpha)\alpha_1}, \quad \beta = \frac{\alpha_1\beta_1}{(1-\alpha_1)\alpha + \alpha_1\beta_1},$$

$$\gamma_1 = \frac{\alpha\beta}{\alpha_1}, \gamma_2 = \frac{(1-\alpha)\beta}{\alpha_2}.$$

Очевидно, що $\beta \in (0,1), \beta_1 \in (0,1)$. Отже знайдені $\beta, \beta_1, \gamma_1, \gamma_2$, що задовольняють (6.4.40), а це означає, що має місце (6.4.36). $\square$

Наслідок 6.4.1. *Якщо*

$$0 \in \operatorname{conv}\{(\mathrm{A} + w) \cup \mathrm{B}\} \quad \forall w \in E(C), \tag{6.4.41}$$

де $E(C)$— множина всіх крайніх точок опуклої множини C, то

$$0 \in \operatorname{conv}\{(\mathrm{A} + w) \cup \mathrm{B}\} \quad \forall w \in C. \tag{6.4.42}$$

Тут, як і в лемі 6.4.2, А,В $\subset \mathbb{R}^n$– опуклі множини.

Лема 6.4.3. *Нехай* А,В $\subset \mathbb{R}^n$— *опуклі множини. Якщо*

$$0 \in \text{conv}\{(\text{A} + w) \cup (\text{B} + w')\} \quad \forall w \in E(C), \quad \forall w' \in E(D), \qquad (6.4.43)$$

де $E(C)$ *і* $E(D)$— *множини всіх крайніх точок опуклих множин* C *і* D *відповідно, то*

$$0 \in \text{conv}\{(\text{A} + w) \cup (\text{B} + w')\} \quad \forall w \in C, \quad \forall w' \in D. \qquad (6.4.44)$$

Доведення. Зафіксуємо будь-яке $w' \in E(D)$. За лемою 6.4.2

$$0 \in \text{conv}\{(\text{A} + w) \cup (\text{B} + w')\} \quad \forall w \in C. \qquad (6.4.45)$$

Умова (6.4.42) виконується для будь-якого $w' \in E(D)$. Фіксуючи тепер $w \in C$ і застосовуючи лему 6.4.2 (де роль A тепер грає B, а роль B — множина $\text{A} + w$), одержимо, що

$$0 \in \text{conv}\{(\text{A} + w) \cup (\text{B} + w')\} \quad \forall w' \in D. \qquad (6.4.46)$$

Лема доведена, оскільки умова(6.4.46) справедлива для всіх $w \in C$. □

Наслідок 6.4.2. *Умову (6.4.31) досить перевіряти лише для крайніх точок множин* $\overline{\partial} f\,(x^*)$ *і* $\overline{\partial} h\,(x^*)$ *. Ця процедура особливо спрощується в тому випадку, коли* $\overline{\partial} f\,(x^*)$ *і* $\overline{\partial} h\,(x^*)$— *многогранники. Тоді умову (6.4.31) потрібно перевірити лише для вершин цих многогранників. Аналогічно можна показати, що й умову (6.4.21) досить перевірити лише для крайніх точок множин* $\overline{\partial} f\,(x^*)$ *і* $\overline{\partial} h\,(x^*)$ *.*

Сформулюємо ще одну умову мінімуму.

Припустимо, що $S = \{x|\ \ h\,(x) \leq 0\}$.

Теорема 6.4.5. *Нехай* $x^* \in S$ *і* $h(x^*) = 0$. *Припустимо, що функції* f *і* h *квазідиференційовні на* $\mathbb{R}^n$. *Для того щоб в точці* x^* *функція* f *досягала свого найменшого на* S *значення, необхідно, щоб*

$$L_1(x^*) \subset L_2(x^*), \qquad (6.4.47)$$

де

$$L_1(x) = -[\overline{\partial} f(x) + \overline{\partial} h(x)], \qquad (6.4.48)$$

$$L_2(x) = \text{conv}\{\underline{\partial} f(x) - \overline{\partial} h(x), \underline{\partial} h(x) - \overline{\partial} f(x)\}. \qquad (6.4.49)$$

Доведення. Нехай $f^* = f(x^*) = \min\limits_{x \in S} f(x), h(x^*) = 0$. Розглянемо функцію $F(x) = \max\{f(x) - f^*, h(x)\}$. Зрозуміло, що $F(x) \geq 0 \quad \forall x \in \mathbb{R}^n$. Оскільки $F(x^*) = 0$, то точка x^*— точка мінімуму функції F на $\mathbb{R}^n$. За теоремою 6.4.1 має виконуватись умова

$$-\overline{\partial} F(x^*) \subset \underline{\partial} F(x^*). \qquad (6.4.50)$$

Обчислюючи квазідиференціал функції F, одержуємо

$$\mathfrak{D} F(x^*) = [\underline{\partial} F(x^*), \overline{\partial} F(x^*)],$$

$$\underline{\partial}F(x^*) = \mathrm{conv}\{\underline{\partial}f(x^*) - \overline{\partial}h(x^*), \underline{\partial}h(x^*) - \overline{\partial}f(x^*)\}, \tag{6.4.51}$$

$$\overline{\partial}F(x^*) = \overline{\partial}f(x^*) + \overline{\partial}h(x^*). \tag{6.4.52}$$

Звідси і з (6.4.50) випливає (6.4.47). □

Теорема 6.4.6. *Якщо $h(x^*) = 0$ і виявилося, що*

$$L_1(x^*) \subset \mathrm{int}\, L_2(x^*), \tag{6.4.53}$$

де $L_1(x)$ і $L_2(x)$ задані співвідношеннями (6.4.48) і (6.4.49), то точка x^* є точкою строгого локального мінімуму функції f на множині S, тобто, умова (6.4.53)– достатня умова строгого локального мінімуму.

Доведення випливає з (6.4.2) і співвідношень (6.4.18), (6.4.51), (6.4.52).

Теорема 6.4.7. *Нехай $h(x^*) = 0$. Припустимо, що в точці x^* виконана умова регулярності (6.4.18). Тоді умова (6.4.47) еквівалентна умові*

$$f'(x^*,g) \geq 0 \quad \forall g \in \Gamma(x^*), \tag{6.4.54}$$

де $\Gamma(x^*)$— конус можливих напрямків множини S в точці x.

Доведення. Нехай виконана умова (6.4.47). Тоді

$$\max_{v \in L_1(x^*)} \langle v,g \rangle \leq \max_{v \in L_2(x^*)} \langle v,g \rangle \quad \forall g \in \mathbb{R}^n,$$

тобто

$$\max_{v \in -[\overline{\partial}f + \overline{\partial}h]} \langle v,g \rangle \leq \max_{v \in \mathrm{conv}\{\underline{\partial}f - \overline{\partial}h, \underline{\partial}h - \overline{\partial}f\}} \langle v,g \rangle \quad \forall g \in \mathbb{R}^n. \tag{6.4.55}$$

Тут для скорочення запису позначено $\overline{\partial}f = \overline{\partial}f(x^*)$ і т. п. З (6.4.55) випливає, що $\forall g \in \mathbb{R}^n$

$$\max_{\substack{w \in \overline{\partial}f \\ w' \in \overline{\partial}h}} \langle -w - w',g \rangle \leq \max_{\substack{\alpha \in [0,1] \\ v \in \underline{\partial}f, w \in \overline{\partial}f \\ v' \in \underline{\partial}h, w' \in \overline{\partial}h}} \langle \alpha(v - w') + (1-\alpha)(v' - w),g \rangle.$$

Оскільки всі зазначені множини одна від одної не залежать, то

$$\max_{w \in \overline{\partial}f} \langle -w,g \rangle + \max_{w' \in \overline{\partial}h} \langle -w',g \rangle \leq$$

$$\leq \max_{\alpha \in [0,1]} \left[\alpha \max_{v \in \underline{\partial}f} \langle v,g \rangle + (1-\alpha) \max_{w \in \overline{\partial}f} \langle -w,g \rangle + \right.$$

$$\left. +(1-\alpha) \max_{v' \in \underline{\partial}h} \langle -v',g \rangle + \alpha \max_{w' \in \overline{\partial}h} \langle -w',g \rangle \right] \quad \forall g \in \mathbb{R}^n. \tag{6.4.56}$$

Але

$$\max_{w \in A} \langle -w,g \rangle = -\min_{w \in A} \langle w,g \rangle.$$

Тому з (6.4.56) маємо

$$0 \leq \max_{\alpha\in[0,1]} \left[\alpha\left(\max_{v\in\underline{\partial} f}\langle v,g\rangle + \min_{w\in\overline{\partial} f}\langle w,g\rangle\right)+\right.$$

$$\left.+(1-\alpha)\left(\max_{v'\in\underline{\partial} h}\langle v',g\rangle + \min_{w'\in\overline{\partial} h}\langle w',g\rangle\right)\right] \quad \forall g \in \mathbb{R}^n. \qquad (6.4.57)$$

Якщо $g \in \Gamma(x^*)$ то $h'(x^*,g) \leq 0$. Тому

$$\max_{\alpha\in[0,1]} \alpha\left(\max_{v\in\underline{\partial} f}\langle v,g\rangle + \min_{w\in\overline{\partial} f}\langle w,g\rangle\right) \geq 0 \quad \forall g \in \Gamma(x^*).$$

Звідси

$$f'(x^*,g) = \max_{v\in\underline{\partial} f}\langle v,g\rangle + \min_{w\in\overline{\partial} f}\langle w,g\rangle \geq 0 \quad \forall g \in \Gamma(x^*). \qquad (6.4.58)$$

Дійсно, якби виявилося $f'(x^*,g) = -a < 0$, то в силу того, що має місце (6.4.18) і (6.4.19), знайдеться $g' = \gamma_1(x)$ досить близьке до g і таке, що $f'(x^*,g') \leq -a/2 < 0$, при цьому $h'(x^*,g) = -b < 0$. Тому

$$\max_{\alpha\in[0,1]} \left[\alpha f'(x^*,g') + (1-\alpha)h'(x^*,g')\right] =$$

$$\max\{f'(x^*,g'),h'(x^*,g')\} \leq \max\{-a/2,-b\} < 0,$$

що суперечить (6.4.57). Тим самим (6.4.58) встановлена, тобто має місце (6.4.54).

Нехай тепер виконана (6.4.54). Припустимо, що при цьому (6.4.47) не має місця, тобто знайдеться таке $\overline{z} \in L_1(x^*)$, що $\overline{z} \notin L_2(x^*)$. За теоремою про розділення знайдуться $\overline{g} \in \mathbb{R}^n$ і $a > 0$ такі, що

$$\langle\overline{z},\overline{g}\rangle \geq \langle z,\overline{g}\rangle + a \quad \forall z \in L_2(x^*). \qquad (6.4.59)$$

З (6.4.48), (6.4.49), (6.4.59) маємо для $\overline{z} = \overline{w} + \overline{w}'$, де $\overline{w} \in \overline{\partial} f$, $\overline{w}' \in \overline{\partial} h$,

$$-\langle\overline{w}+\overline{w}',\overline{g}\rangle \geq \langle\alpha(v-w') + (1-\alpha)(v'-w),\overline{g}\rangle + a$$

$$\forall\alpha \in [0,1], \quad v \in \underline{\partial} f, \quad w \in \overline{\partial} f, \quad v' \in \underline{\partial} h, \quad w' \in \overline{\partial} h.$$

Тим більше це вірно при $w = \overline{w}, w' = \overline{w}'$ тобто

$$\alpha\langle v+\overline{w},\overline{g}\rangle + (1-\alpha)\langle v'+\overline{w}'),\overline{g}\rangle \leq -a \quad \forall\alpha \in [0,1], v \in \underline{\partial} f, v' \in \underline{\partial} h. \qquad (6.4.60)$$

При $\alpha = 1$ з (6.4.60) одержуємо

$$\langle v+\overline{w},\overline{g}\rangle \leq -a \quad v \in \underline{\partial} f,$$

тобто

$$\max_{v\in\underline{\partial} f}\langle v,\overline{g}\rangle + \langle\overline{w},\overline{g}\rangle \leq -a.$$

Тим більше

$$\max_{v\in\underline{\partial} f}\langle v,\overline{g}\rangle + \min_{w\in\overline{\partial} f}\langle w,\overline{g}\rangle \leq -a,$$

тобто

$$f'(x^*,\overline{g}) \le -a. \tag{6.4.61}$$

При $\alpha = 0$ з (6.4.60) одержуємо

$$\langle v' + w',\overline{g}\rangle \le -a \quad \forall v' \in \underline{\partial} h.$$

Як і вище, звідси

$$h'(x^*,\overline{g}) \le -a. \tag{6.4.62}$$

тобто $\overline{g} \in \Gamma_1(x^*)$. Але (6.4.61) суперечить (6.4.54). Одержана суперечність доводить теорему. □

Зауваження 6.4.7. Умова (6.4.47) отримана без врахування умови регулярності в точці x^*.

Зауваження 6.4.8. Нехай $x \in S$, $h(x) = 0$. Припустимо, що умова (6.4.47) не має місця. Знайдемо

$$d(x) = \max_{v \in L_1(x)} \rho(x) = \rho(v(x)), \tag{6.4.63}$$

де

$$\rho(v) = \min_{w \in L_2(x)} \|v - w\| = \|v - w(v)\|. \tag{6.4.64}$$

Оскільки умова (6.4.47) не виконується, то $\rho\left(v\left(x\right)\right) > 0$. Множини $L_1(x)$ і $L_2(x)$ опуклі, тому для кожного $v \in L_1(x)$ існує єдина точка $w\left(v\right)$, що задовольняє (6.4.64) (нагадаємо, що в нас евклідова норма, але $v\left(x\right)$, що задовольняє (6.4.63), вже не обов'язково єдине.

Теорема 6.4.8. *Напрямок*

$$g_0 = \frac{v(x) - w(v(x))}{\|v(x) - w(v(x))\|} \tag{6.4.65}$$

є напрямком спуску функції f на множині S в точці x.

Доведення дослівно повторює доведення другої частини теореми 6.4.7.

Зауваження 6.4.9. Напрямок (6.4.65) не обов'язково єдиний.

Відмітимо, що множину S можна записати у вигляді

$$S = \left\{x \in \mathbb{R}^n | h_\eta(x) \le 0\right\},$$

де $h_\eta(x) = \eta h(x),\eta > 0$. Тому з (6.4.47) одержуємо таку необхідну умову мінімуму:

$$L_{1\eta}(x^*) \subset L_{2\eta}(x^*), \tag{6.4.66}$$

де

$$L_{1\eta}(x) = -\left[\overline{\partial} f(x) + \eta \overline{\partial} h(x)\right],$$

$$L_{2\eta}(x) = \operatorname{conv}\left\{\underline{\partial} f(x) - \eta \overline{\partial} h(x), \eta \underline{\partial} h(x) - \overline{\partial} f(x)\right\}.$$

Якщо $h(x) = 0$, а точка x не задовольняє умову (6.4.66), то можна побудувати напрямок спуску $g_\eta(x)$ за формулами, аналогічними (6.4.63)—(6.4.65). Можна також показати, що якщо

$$g_{\eta_k}(x) \to g_0, \quad \eta_k \to \infty,$$

то g_0 - напрямок найшвидшого спуску функції f на множині S. При цьому для кожного $\eta > 0$ напрямок $g_\eta(x)$ є допустимим, тобто при досить малих $\alpha > 0$ буде $x + \alpha g_\eta(x) \in S$.

Зауваження 6.4.10. Всі описані вище умови мають місце незалежно від того, які взяти квазідиференціали функцій f і h (адже квазідиференціал визначається неєдиним чином). Звичайно, з обчислювальної точки зору краще взяти найбільш прості в якомусь розумінні квазідиференціали.

Розглянемо декілька ілюстративних прикладів застосування отриманих вище результатів.

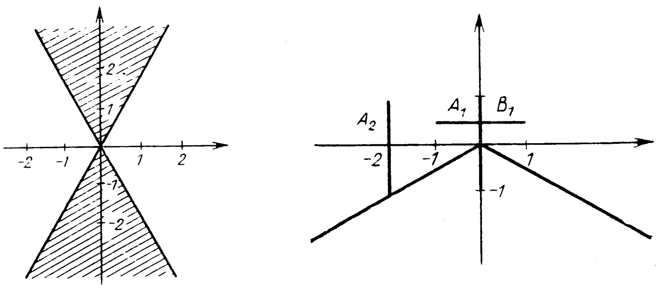


Рис. 6.4.4:

Приклад 6.4.7. Нехай $x = \left(x^{(1)},x^{(2)}\right) \in \mathbb{R}^2, x_0 = (0,0)$, $h(x) = \left|x^{(1)}\right| - \frac{1}{2}\left|x^{(2)}\right|$, $S = \{x \in \mathbb{R}^2 | h(x) \le 0\}$. Множина S зображена на рис. 6.4.4. Очевидно, що $x_0 \in S, \mathfrak{D}h(x_0) = \left[\underline{\partial} h(x_0), \overline{\partial} h(x_0)\right]$, де

$$\underline{\partial} h(x_0) = \text{conv}\{(1,0),(-1,0)\}$$

$$\overline{\partial} h(x_0) = \text{conv}\{(0,1/2),(0,-1/2)\}.$$

В точці $x_0 \in S$ виконується умова регулярності (6.4.18). Розглянемо функцію $f(x) = -\left|x^{(1)}\right| + \left|x^{(2)}\right| - x^{(1)}$. Маємо

$$\mathfrak{D}f(x_0) = \left[\underline{\partial} f(x_0), \overline{\partial} f(x_0)\right],$$

$$\underline{\partial} f(x_0) = \text{conv}\{(-1,1),(-1,-1)\},$$

$$\overline{\partial} f\,(x_0) = \mathrm{conv}\{(1{,}0),(-1{,}0)\}.$$

Перевіримо виконання умови мінімуму (6.4.21) в точці x_0. Умову (6.4.21) потрібно перевірити лише для крайніх точок множин $\overline{\partial} f(x_0)$ та $\overline{\partial} h(x_0)$. Покладемо $w_1 = (1{,}0)$, $w_2 = (-1{,}0)$, $w_1' = (0{,}1/2)$, $w_2' = (0,-1/2)$. Маємо $A_1 = \underline{\partial} f\,(x_0) + w_1 = \mathrm{conv}\{(-1{,}1)\,,(-1,-1)\} + (1{,}0) = \mathrm{conv}\{(0{,}1),(0,-1)\}$, $A_2 = \underline{\partial} f\,(x_0) + w_2 = \mathrm{conv}\{(-1{,}1)\,,(-1,-1)\} + (-1{,}0) = \mathrm{conv}\{(-2{,}1),(-2,-1)\}$, $B_1 = \underline{\partial} h\,(x_0) + w_1' = \mathrm{conv}\{(1{,}0)\,,(-1{,}0)\} + (0{,}1/2) = \mathrm{conv}\{(1{,}1/2),(-1{,}1/2)\}$, $B_2 = \underline{\partial} h\,(x_0) + w_2' = \mathrm{conv}\{(1{,}0)\,,(-1{,}0)\} + (0,-1/2) = \mathrm{conv}\{(1,-1/2),(-1,-1/2)\}$,

$$C_1 = -\,\mathrm{cl}\,(\mathrm{cone}\,B_1)\,,C_2 = -\,\mathrm{cl}\,(\mathrm{cone}\,B_2)\,.$$

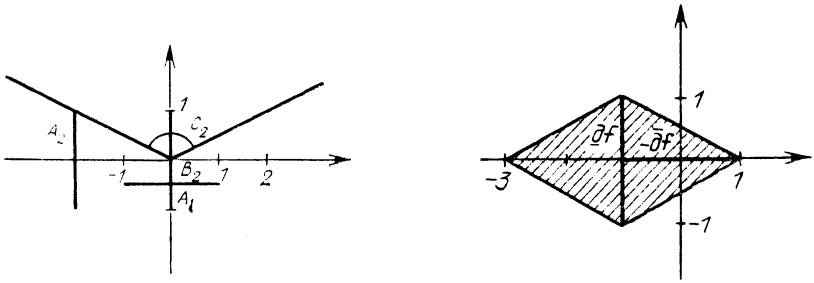


Рис. 6.4.5:

На рис. 6.4.5 видно, що $A_i \cap C_j \neq \emptyset\ i{,}j = 1{,}2$, тобто умова (6.4.21) виконується.

Перевіримо умову (6.4.24). Маємо (див. (6.4.25))

$$L(x_0) = [\underline{\partial} f(x_0) - C_1] \cap [\underline{\partial} f(x_0) - C_2]\,.$$

На рис. 6.4.5 видно, що

$$L(x_0) = \mathrm{conv}\{(1{,}0),(-3{,}0),(-1{,}1),(-1,-1)\}.$$

Зрозуміло, що

$$-\overline{\partial} f(x_0) = \mathrm{conv}\{(-1{,}0),(1{,}0)\} \subset L(x_0),$$

тобто умова (6.4.24) також виконується.

Перевіримо умову (6.4.31). Маємо

$$\mathfrak{L}_1 = \mathfrak{L}\,(x_0,w_1,w_1') = \mathrm{conv}\{\underline{\partial} f(x_0) + w_1, \underline{\partial} h(x_0) + w_1'\} = \mathrm{conv}\{A_1,B_1\} = \\ = \mathrm{conv}\{(0{,}1),(0,-1),(1{,}1/2),(-1{,}1/2)\},$$

$$\mathfrak{L}_2 = \mathfrak{L}\,(x_0,w_1,w_2') = \mathrm{conv}\{A_1,B_2\} = \mathrm{conv}\{(0{,}1),(0{,}-1),(1{,}-1/2),(-1{,}-1/2)\},$$

$\mathfrak{L}_3 = \mathfrak{L}\,(x_0,w_2,w_1') = \mathrm{conv}\{A_2,B_1\} = \mathrm{conv}\{(-2,1),(-2,-1),(1,1/2),(-1,1/2)\}$, $\mathfrak{L}_4 = \mathfrak{L}\,(x_0,w_2,w_2') = \mathrm{conv}\{A_2,B_2\} = \mathrm{conv}\{(-2,1),(-2,-1),(1,-1/2),$ $(-1,-$ $1/2)\}$. На рис. 6.4.6 видно, що $0 \in \mathfrak{L}_i \quad i = 1,\ldots,4$, тобто умова (6.4.31) виконана. Перевіримо тепер умову (6.4.47). Маємо

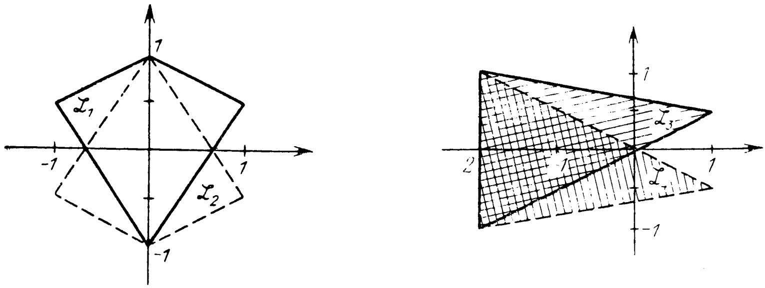


Рис. 6.4.6:

$$L_1(x_0) = -\left[\overline{\partial} f(x_0) + \overline{\partial} h(x_0)\right] =$$
$$= -\left[\mathrm{conv}\{(1,0),(-1,0)\} + \mathrm{conv}\{(0,1/2),(0,-1/2)\}\right] =$$
$$= -\,\mathrm{conv}\{(1,1/2),(-1,1/2),(1,-1/2),(-1,-1/2)\} =$$
$$\mathrm{conv}\{(-1,-1/2),(1,-1/2),(-1,1/2),(1,1/2)\},$$
$$\underline{\partial} f(x_0) - \overline{\partial} h(x_0) = \mathrm{conv}\{(-1,1),(-1,-1)\} - \mathrm{conv}\{(0,1/2),(0,-1/2)\} =$$
$$= \mathrm{conv}\{(-1,3/2),(-1,1/2),(-1,-1/2),(-1,-3/2)\},$$
$$\underline{\partial} h(x_0) - \overline{\partial} f(x_0) = \mathrm{conv}\{(1,0),(-1,0)\} - \mathrm{conv}\{(1,0),(-1,0)\} =$$
$$\mathrm{conv}\{(2,0),(0,0),(-2,0)\} = \mathrm{conv}\{(2,0),(-2,0)\},$$
$$L_2(x_0) = \mathrm{conv}\{\underline{\partial} f(x_0) - \overline{\partial} h(x_0), \underline{\partial} h(x_0) - \overline{\partial} f(x_0)\} =$$
$$\mathrm{conv}\{(-1,3/2),(-1,1/2),(-1,-1/2),(-1,-3/2),(0,0),(2,0),(-2,0)\}.$$

На рис. 6.4.7 видно, що $L_1(x_0) \subset L_2(x_0)$, тобто умова (6.4.47) також виконана. Нарешті, побудуємо множину $L'(x_0)$ (див. (6.4.28)):
$$L'(x_0) = (A_1 - C_1) \cap (A_1 - C_2) \cap (A_2 - C_1) \cap (A_2 - C_2).$$
Маємо
$$P_1 = (A_1 - C_1) \cap (A_1 - C_2) = \mathrm{conv}\{(2,0),(-2,0),(0,1),(0,-1)\},$$
$$P_1 = (A_2 - C_1) \cap (A_2 - C_2) = \mathrm{conv}\{(-4,0),(0,0),(-2,1),(-2,-1)\},$$
Звідси
$$L'(x_0) = P_1 \cap P_2 = \mathrm{conv}\{(-2,0),(0,0),(-1,1/2),(-1,-1/2)\},$$
тобто (див. рис. 6.4.7) $0 \in L'(x_0)$, тобто умова (6.4.28) теж виконана.

Приклад 6.4.8. Нехай S — множина, яка визначена в прикладі 6.4.7. Візьмемо $f(x) = \left|x^{(1)}\right| - \left|x^{(2)}\right| - 2x^{(1)}$, $x_0 = (0{,}0) \in S$. Тоді

$$\underline{\partial} f(x_0) = \operatorname{conv}\{(-3{,}0),(-1{,}0)\}, \quad \overline{\partial} f(x_0) = \operatorname{conv}\{(0{,}1),(0,-1)\}.$$

Перевіримо, чи виконується умова (6.4.21).

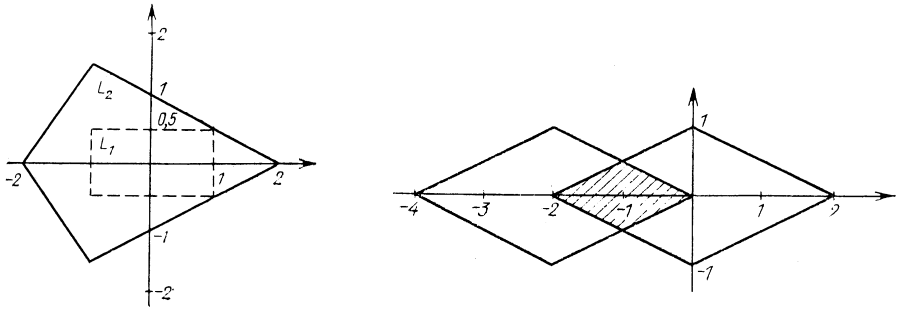


Рис. 6.4.7:

Досить перевірити (6.4.21) лише для крайніх точок множин $\overline{\partial} f(x_0)$, $\overline{\partial} h(x_0)$. Покладемо

$$w_1 = (0{,}1), w_2 = (0,-1); w_1' = (0{,}1/2), w_2' = (0,-1/2),$$

$$A_1 = \underline{\partial} f(x_0) + w_1 = \operatorname{conv}\{(-3{,}1),(-1{,}1)\},$$

$$A_2 = \underline{\partial} f(x_0) + w_2 = \operatorname{conv}\{(-3,-1),(-1,-1)\},$$

$$B_1 = \underline{\partial} h(x_0) + w_1' = \operatorname{conv}\{(1{,}1/2),(-1{,}1/2)\},$$

$$B_2 = \underline{\partial} h(x_0) + w_2' = \operatorname{conv}\{(1,-1/2),(-1,-1/2)\},$$

$$C_1 = -\operatorname{cone} B_1, \quad C_2 = -\operatorname{cone} B_2.$$

З рис. 6.4.8 ясно, що $A_1 \cap C_1 = \emptyset$, $A_1 \cap C_2 \neq \emptyset$, $A_2 \cap C_1 \neq \emptyset$, $A_2 \cap C_2 = \emptyset$. Знайдемо

$$\max_{v \in A_1} \min_{w \in C_1} \|v - w\| = \min_{w \in C_1} \|v_1 - w\| = \|v_1 - w_1\|,$$

$$\max_{v \in A_2} \min_{w \in C_2} \|v - w\| = \min_{w \in C_2} \|v_2 - w\| = \|v_2 - w_2\|,$$

де

$$v_1 = (-1{,}1),\ w_1 = (-2/5,-1/5),\ v_2 = (-1,-1),\ w_2 = (-2/5{,}1/5).$$

Маємо

$$\|v_1 - w_1\| = \|(-3/5{,}6/5)\| = 3\sqrt{5}/5,$$

$$\|v_2 - w_2\| = \|(-3/5,-6/5)\| = 3\sqrt{5}/5.$$

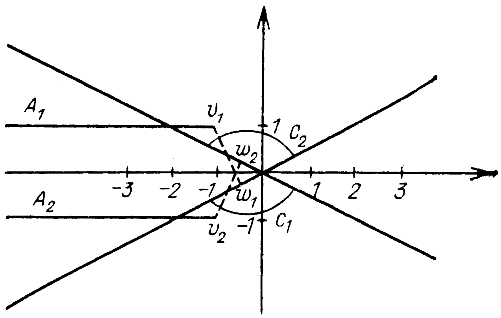


Рис. 6.4.8:

Отже, існує два напрямки найшвидшого спуску

$$g_1 = (1/\sqrt{5}, -2/\sqrt{5}), \quad g_2 = (1/\sqrt{5}, 2/\sqrt{5}), \qquad (6.4.67)$$

при цьому швидкість найшвидшого спуску

$$f'(x_0, g_1) = f'(x_0, g_2) = -3\sqrt{5}/5.$$

Перевіримо умову (6.4.24). Маємо

$$L(x_0) = (\underline{\partial} f(x_0) - C_1) \cap (\underline{\partial} f(x_0) - C_2).$$

З рис. 6.4.9 видно, що $L(x_0) = \mathrm{conv}\{(-3,0),(-1,0)\}$. Розв'язуючи задачу

$$\min_{w \in L(x_0)} \|v - w\| \to \max_{v \in [-\bar{\partial} f(x_0)]} \rho(v),$$

одержимо (див. рис. 6.4.9) два напрямки

$$g_1' = (\sqrt{2}/2, \sqrt{2}/2), g_2' = (\sqrt{2}/2, -\sqrt{2}/2).$$

Проте жоден з цих напрямків не є допустимим напрямком (вони навіть не належать $\Gamma(x_0)$).

Перевіримо тепер умову (6.4.31). Маємо

$$\mathfrak{L}_1 = \mathfrak{L}(x_0, w_1, w_1') = \mathrm{conv}\{\underline{\partial} f(x_0) + w_1, \underline{\partial} h(x_0) + w_1'\} = \mathrm{conv}\{A_1, B_1\} =$$

$$= \mathrm{conv}\{(-3,1),(-1,1),(1,1/2),(-1,1/2)\};$$

$$\mathfrak{L}_2 = \mathfrak{L}(x_0, w_1, w_2') = \mathrm{conv}\{A_1, B_2\} =$$

$$= \mathrm{conv}\{(-3,1),(-1,1),(1,-1/2),(-1,-1/2)\};$$

$$\mathfrak{L}_3 = \mathfrak{L}(x_0, w_2, w_1') = \mathrm{conv}\{A_2, B_1\} =$$

$$= \mathrm{conv}\{(-3,-1),(-1,-1),(1,1/2),(-1,1/2)\};$$

$$\mathfrak{L}_4 = \mathfrak{L}(x_0, w_2, w_2') = \mathrm{conv}\{A_2, B_2\} =$$

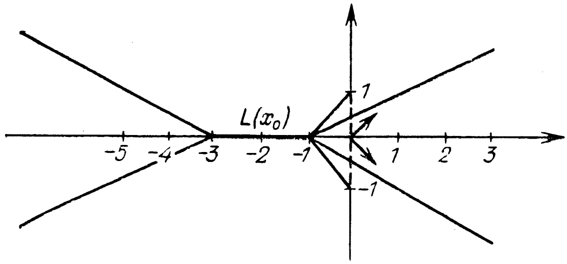


Рис. 6.4.9:

$$= \text{conv}\{(-3,-1),(-1,-1),(1,-1/2),(-1,-1/2)\};$$

З рис. 6.4.10 ясно, що $0 \in \mathfrak{L}_2$, $0 \in \mathfrak{L}_3$, але $0 \notin \mathfrak{L}_1$, $0 \notin \mathfrak{L}_4$. З рис. 6.4.10 одержуємо два напрямки спуску $g_1' = (0{,}1)$, $g_2' = (0,-1)$. Неважко обчислити, що

$$f'(x_0,g_1) = f'(x_0,g_2) = -1.$$

При цьому $g_1 \in \operatorname{int}\Gamma(x_0), g_2 \in \operatorname{int}\Gamma(x_0)$.

Раніше ми вже знайшли напрямки найшвидшого спуску (див. (6.4.67)) і швидкість найшвидшого спуску $-3/\sqrt{5}$.

Звернемося до умов (6.4.47). Маємо

$$L_1(x_0) = -[\overline{\partial} f(x_0) + \overline{\partial} h(x_0)] =$$

$$-[\text{conv}\{(0{,}1),(0,-1)\} + \text{conv}\{(0{,}1/2),(0,-1/2)\}] =$$

$$= \text{conv}\{(0{,}3/2),(0,-1/2),(0{,}1/2),(0,-3/2)\} = \text{conv}\{(0,-3/2),(0{,}3/2)\},$$

$$Q_1 \equiv \underline{\partial} f(x_0) - \overline{\partial} h(x_0) = \text{conv}\{(-3{,}0),(-1{,}0)\} - \text{conv}\{(0{,}1/2),(0,-1/2)\} =$$

$$= \text{conv}\{(-3,-1/2),(-3{,}1/2),(-1,-1/2),(-1{,}1/2)\},$$

$$Q_2 \equiv \underline{\partial} h(x_0) - \overline{\partial} f(x_0) =$$

$$= \text{conv}\{(1{,}0),(-1{,}0)\} - \text{conv}\{(0{,}1),(0,-1)\} =$$

$$\text{conv}\{(1{,}1),(1,-1),(-1{,}1),(-1,-1)\},$$

$$L_2(x_0) = \text{conv}\{Q_1,Q_2\} =$$

$$= \text{conv}\{(-3,-1/2),(-3{,}1/2),(-1,-1/2),$$

$$(-1{,}1/2),(1{,}1),(1,-1),(-1{,}1),(-1,-1)\}.$$

З рис. 6.4.11 ясно, що $L_1(x_0) \not\subset L_2(x_0)$. Знайдемо

$$\max_{z \in L_1(x_0)} \min_{z' \in L_2(x_0)} \|z - z'\| = \min_{z' \in L_2(x_0)} \|z_0 - z'\| = \|z_0 - z'(z_0)\|. \quad (6.4.68)$$

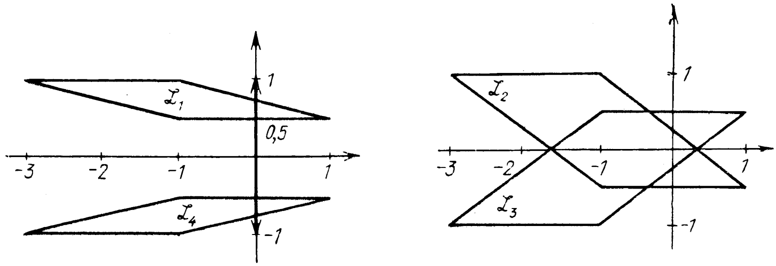


Рис. 6.4.10:

Очевидно, що існує дві точки z_0, що задовольняють (6.4.68): $z_0 = (0{,}3/2)$ (її відповідає $z'(z_0) = (0{,}1)$) і
$\bar{z}_0 = (0, -3/2)$ (з $z'(z_0) = (0, -1)$). Звідси знаходимо два напрямки спуску $g_1'' = (0{,}1)$ та $g_2'' = (0, -1)$ (що збігаються зі знайденими вище напрямками g_1' і g_2' отриманими за допомогою умови (6.4.31)).

Нарешті, перевіримо умову (6.4.28). Побудуємо

$$L'(x_0) = (A_1 - C_1) \cap (A_1 - C_2) \cap (A_2 - C_1) \cap (A_2 - C_2).$$

Так як $(A_1 - C_1) \cap (A_1 - C_2) = A_1$, $(A_2 - C_1) \cap (A_2 - C_2) = A_2$ (див. рис. 6.4.11), а множини A_1 і A_2 не перетинаються, то $L'(x_0) = \emptyset$.
Таким чином, необхідна умова (6.4.28) не виконана, але більше ніякої інформації ми отримати не можемо.
Отже, за допомогою всіх розглянутих умов ми знайшли, що точка x_0 не є стаціонарною. Умова (6.4.21) дозволила знайти всі напрямки най-

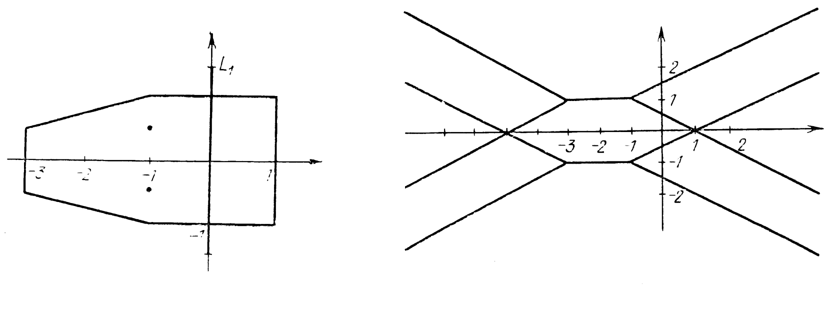


Рис. 6.4.11:

швидшого спуску, а за допомогою умов (6.4.31) і (6.4.47) були знайдені допустимі напрямки спуску, що ведуть всередину множини S. Умова (6.4.28) лише "попередила" про те, що точка x_0 не є стаціонарною, а умова (6.4.24), крім того, привела до напрямку, який не є навіть можливим. Тому цією умовою треба користуватися з обережністю. Практично можна рекомендувати умови (6.4.21) (для знаходження н. н. с.) і (6.4.31) і (6.4.47), (для знаходження допустимих напрямків спуску).

Приклад 6.4.9. Нехай $x = (x^{(1)},x^{(2)}) \in \mathbb{R}^2, x_0 = (0,0)$, $f(x) = |x^{(1)}| - |x^{(2)}| + x^{(2)}$, $h(x) = -\frac{1}{2}|x^{(1)}| - x^{(2)}$. Тоді $f(x) = f_1(x) + f_2(x) + f_3(x)$, $h(x) = h_1(x) + h_2(x)$, де $f_1(x) = |x^{(1)}|$, $f_2(x) = -|x^{(2)}|$, $f_3(x) = x^{(2)}$, $h_1(x) = -\frac{1}{2}|x^{(1)}|$, $h_2(x) = -x^{(2)}$. Побудуємо множину

$$S = \{x \in \mathbb{R}^2 | h(x) \leq 0\}.$$

За прикладами 6.1.4-6.1.5

$$\begin{aligned}
\underline{\partial} f_1(x_0) &= \operatorname{conv}\{(-1,0),(1,0)\},\\
\overline{\partial} f_1(x_0) &= \{(0,0)\},\\
\underline{\partial} f_2(x_0) &= \{(0,0)\},\\
\overline{\partial} f_2(x_0) &= \operatorname{conv}\{(0,-1),(0,1)\},\\
\underline{\partial} f_3(x_0) &= \{(0,1)\},\\
\overline{\partial} f_3(x_0) &= \{(0,0)\},\\
\underline{\partial} h_1(x_0) &= \{(0,0)\},\\
\overline{\partial} h_1(x_0) &= \operatorname{conv}\{\frac{1}{2}(-1,0),\frac{1}{2}(1,0)\},\\
\underline{\partial} h_2(x_0) &= \{(0,-1)\},\\
\overline{\partial} h_2(x_0) &= \{(0,0)\}.
\end{aligned}$$

Тоді за правилами квазідиференціального числення

$$\begin{aligned}
\underline{\partial} f(x_0) &= \operatorname{conv}\{(-1,0),(1,0)\} + \{(0,1)\},\\
\overline{\partial} f(x_0) &= \operatorname{conv}\{(0,-1),(0,1)\},\\
\underline{\partial} h(x_0) &= \{(0,-1)\},\\
\overline{\partial} h(x_0) &= \operatorname{conv}\left\{\frac{1}{2}(-1,0),\frac{1}{2}(1,0)\right\}.
\end{aligned}$$

Перевіримо, що в точці виконується умова регулярності. Маємо

$$\frac{\partial h(x_0)}{\partial g} = \langle (0,-1),g\rangle + \min\left\{\langle\frac{1}{2}(1,0),g\rangle,\langle\frac{1}{2}(-1,0),g\rangle\right\}.$$

Знайдемо ті $g \in \mathbb{R}^2$, $||g|| = 1$, для яких $\frac{\partial h(x_0)}{\partial g} = 0$. Очевидно, що

$$\frac{\partial h(x_0)}{\partial g_1} = \frac{\partial h(x_0)}{\partial g_2} = 0,$$

$$g_1 = (2/\sqrt{5}, -1/\sqrt{5}),\ g_2 = (-2/\sqrt{5}, -1/\sqrt{5}).$$

Зауважимо, що

$$\gamma_1 = \beta_1 \cup \beta_2 \cup \beta_3,$$

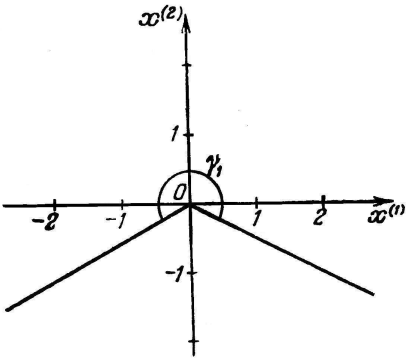
$$\beta_1 = \{g \in \mathbb{R}^2 | g = \lambda(-2,z), \lambda > 0, z \in (-1,0]\},$$

$$\beta_2 = \{g \in \mathbb{R}^2 | g = \lambda(y,1), \lambda > 0, y \in [-1,1]\},$$

$$\beta_1 = \{g \in \mathbb{R}^2 | g = \lambda(2,z), \lambda > 0, z \in (-1,0]\}.$$

$$\Gamma_1(x_0) = \gamma_1(x_0) \cup \{g = \lambda g_1, \lambda \ge 0\} \cup \{g = \lambda g_2, \lambda \ge 0\} = \mathrm{cl}\, \gamma_1(x_0).$$

Тобто в точці x_0 виконується умова регулярності (6.4.18).



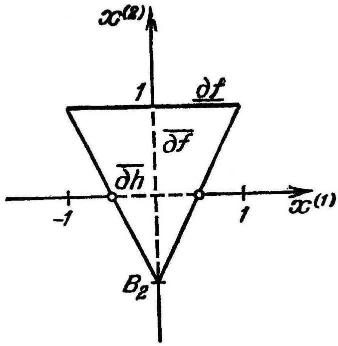


Рис. 6.4.12:

З рис. 6.4.12 видно, що

$$H \equiv \bigcap_{w \in \overline{\partial} h(x_0)} [\underline{\partial} f(x_0) + \mathrm{cl}(\mathrm{cone}(\underline{\partial} h(x_0) + w))] = \mathrm{conv}\{(-1,1),(1,1),(0,-1)\}.$$

Тоді $-\overline{\partial} f(x_0) \subset H$, тобто умова (6.4.21) виконується. Отже точка x_0 є стаціонарною для функції $f(x) = |x^{(1)}| - |x^{(2)}| + x^{(2)}$ на множині $S = \{x | -\frac{1}{2}|x^{(1)}| - x^{(2)} \le 0\}$. Ця точка є точкою мінімума для функції f на множині S.

Приклад 6.4.10. Нехай $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2, x_0 = (0,0)$, $f(x) = |x^{(1)}| - |x^{(2)}| + x^{(2)}$, $h(x) = -|x^{(1)}| - x^{(2)}$. Побудуємо множину

$$S = \{x \in \mathbb{R}^2 | h(x) \le 0\}.$$

Множина S зображена на рис. 6.4.13. За прикладом 6.4.4

$$\underline{\partial} f(x_0) = \mathrm{conv}\,\{(-1,0),(1,0)\} + \{(0,1)\},$$

$$\overline{\partial} f(x_0) = \mathrm{conv}\,\{(0,-1),(0,1)\}.$$

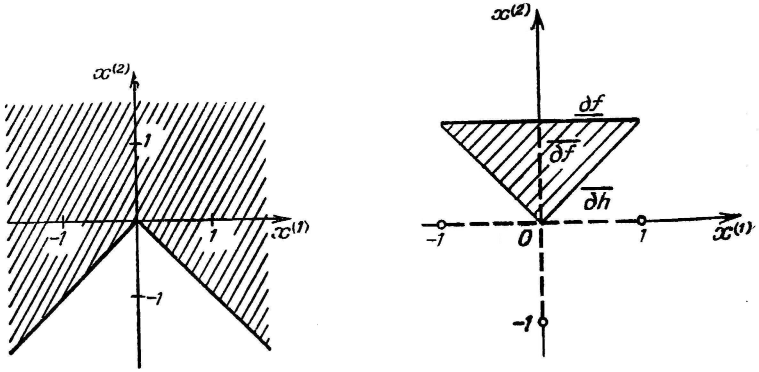


Рис. 6.4.13:

Легко отримати, що

$$\underline{\partial} h(x_0) = \{(0, -1)\}$$
$$\overline{\partial} h(x_0) = \text{conv}\,\{(-1,0),(1,0)\}.$$

Аналогічно до минулого прикладу, перевіряється умова регулярності в точці x_0. Множина H зображена на рис. 6.4.13.

$$H \equiv \bigcap_{w\in\overline{\partial} h(x_0)} [\underline{\partial} f(x_0) + \text{cl}(\text{cone}(\underline{\partial} h(x_0) + w))] = \text{conv}\{(-1,1),(1,1),(0,0)\}.$$

Очевидно, що

$$-\underline{\partial} f(x_0) = -\,\text{conv}\{(0,-1),(0,1)\} = \text{conv}\{(0,-1),(0,1)\} \not\subseteq H.$$

Знайдемо напрямок найскорішого спуску. Маємо для $w \in \overline{\partial} f(x_0), w' \in \overline{\partial} h(x_0)$.

$$\overline{\partial} f(x_0) + w = \text{conv}\{(1,0) + (0,1) + w, (-1,0) + (0,1) + w\} \equiv U_S,$$

$$-\,\text{cl}(\text{cone}(\underline{\partial} h(x_0)+w'))] = \{z = \lambda(-(0,-1)-w') = \lambda((0,1)-w') | \lambda \geq 0\} \equiv B_{w'}.$$

Множини U_w – це відрізки, які паралельні відрізку conv$\{(-1,0),(1,0)\}$ і заповнюють прямокутник з вершинами $(-1,0),(-1,2),(1,2),(1,0)$. Множини $B_{w'}$ – це промені з центром в початку координат, які проходять через точки відрізка, що з'єднує точки $(-1,1)$, $(1,1)$. Знайдемо $\rho(x_0)$

$$\rho(x_0) = \max_{w\in\overline{\partial} f(x_0), w'\in\overline{\partial} h(x_0)} d(w,w') = d(w_0,w'_0) = d(w_1,w'_1),$$

де

$$w_0 = (0,1), w'_0 = (1,0), w_1 = w_0 = (0,1), w'_1 = (-1,0).$$

Маємо два напрямки найскорішого спуску

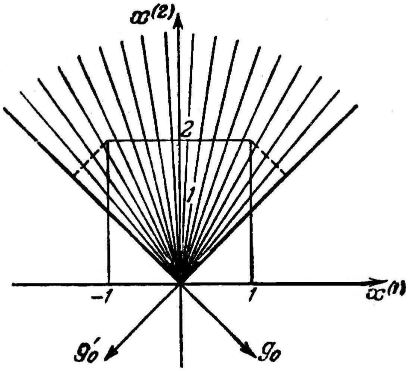


Рис. 6.4.14:

$$g_0 = (\sqrt{2}/2, -\sqrt{2}/2), g_0' = (-\sqrt{2}/2, -\sqrt{2}/2).$$

При цьому

$$\frac{\partial f(x_0)}{\partial g_0} = \max_{v \in \underline{\partial} f(x_0)} \langle v, g_0 \rangle + \min_{w \in \overline{\partial} f(x_0)} \langle w, g_0 \rangle = -\sqrt{2}/2.$$

Аналогічно

$$\frac{\partial f(x_0)}{\partial g_0'} = -\sqrt{2}/2.$$

Очевидно, що

$$\frac{\partial f(x_0)}{\partial g_0} = -\rho(x_0) = -\sqrt{2}/2.$$

Приклад 6.4.11. Нехай $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2, x_0 = (0,0)$, $f(x) = |x^{(1)}| - \frac{1}{2}|x^{(2)}| + x^{(2)}$, $h(x) = -\frac{1}{4}|x^{(1)}| - x^{(2)}$. Побудуємо множину

$$S = \{x \in \mathbb{R}^2 | h(x) \leq 0\}.$$

Як і в попередніх прикладах, отримаємо (рис. 6.4.15), що

$$\underline{\partial} f(x_0) = \operatorname{conv}\{(-1,0),(1,0)\} + \{(0,1)\} = \operatorname{conv}\{(1,1),(-1,1)\},$$

$$\overline{\partial} f(x_0) = \operatorname{conv}\{(0,\frac{1}{2}),(0,-\frac{1}{2})\},$$

$$\underline{\partial} h(x_0) = \{(0,-1)\},$$

$$\overline{\partial} h(x_0) = \operatorname{conv}\{(-\frac{1}{4},0),(\frac{1}{4},0)\}.$$

Візьмемо $w \in \overline{\partial} f(x_0)$ і $w' \in \overline{\partial} h(x_0)$. Тоді

$$\underline{\partial} f(x_0) + w = \operatorname{conv}\{(1,0) + (0,1) + w, (-1,0) + (0,1) + w\} \equiv U_w,$$

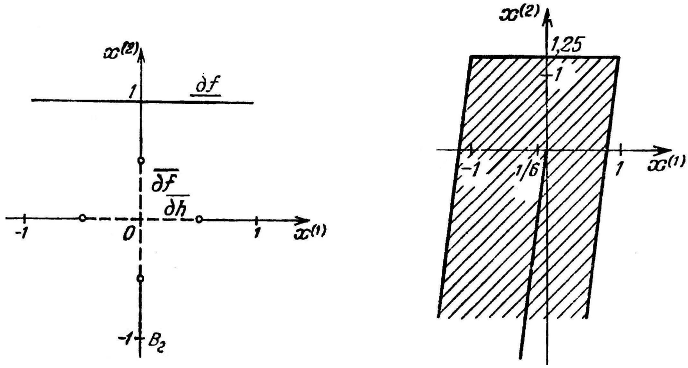


Рис. 6.4.15:

$$-\operatorname{cl}(\operatorname{cone}(\underline{\partial} h(x_0) + w'))] = \{z = \lambda((0, -1) + w')|\lambda \geq 0\} \equiv B_{w'}.$$

Множина $\mathfrak{L}(w,w')$ зображена на рис. 6.4.15 для $w = (0,\frac{1}{4})$, $w' = (-\frac{1}{8},0)$. Промені l_1, l_2, l_3 паралельні. Очевидно, що

$$0 \in \mathfrak{L}(w,w'), \forall w \in \overline{\partial} f(x_0), \forall w' \in \overline{\partial} h(x_0).$$

Неважко помітити, що

$$r = \min_{w \in \overline{\partial} f(x_0), w' \in \overline{\partial} h(x_0)} r(w,w').$$

Мінімум досягається при $w = (0,\frac{1}{2})$, $w' = (\frac{1}{4},0)$ і при $w = (0,\frac{1}{2})$, $w' = (-\frac{1}{4},0)$. Підрахувавши, маємо $r = \sqrt{425}/34 \approx 0{,}6063$.

6.5. ЗВ'ЯЗОК КВАЗІДИФЕРЕНЦІАЛА З СУБДИФЕРЕНЦІАЛАМИ МІШЕЛЯ–ПЕНО І КЛАРКА

Нехай функція f визначена на відкритій множині $X \subset \mathbb{R}^n$ і квазідиференційовна в точці $x \in X$. Нагадаємо, що концепція квазідиференційовності пов'язана з представленням похідної у вигляді суми сублінійної та суперлінійної функцій (різниці двох сублінійних функцій). Якщо $\mathfrak{D} f(x) = \left[\underline{\partial} f(x), \overline{\partial} f(x)\right]$ - квазідиференціал функції f в точці x, то

$$f'(x,g) = \max_{h \in \underline{\partial} f(x)} \langle h, g\rangle + \min_{h \in \overline{\partial} f(x)} \langle h, g\rangle = \max_{h \in \underline{\partial} f(x)} \langle h, g\rangle - \max_{h \in -\overline{\partial} f(x)} \langle h, g\rangle.$$

За допомогою пари множин $\left[\underline{\partial} f(x), -\overline{\partial} f(x)\right]$ похідна відновлюється як різниця двох максимумів. Геометрична інтерпретація похідної при цьому зводиться до опису різниці опуклих множин. Цю різницю можна розуміти неоднозначно. Найбільш природний шлях, при якому віднімання визначається як операція, обернена до додавання в лінійному просторі, приводить до різниці в просторі опуклих множин. Якщо α_1,α_2 - елементи цього простору, що відповідають множинам $\underline{\partial} f(x)$ та $\left(-\overline{\partial} f(x)\right)$ відповідно, то різниця цих елементів має вигляд

$$\alpha_1 - \alpha_2 = [\underline{\partial} f(x), \{0\}] + \left[\{0\}, -\left(-\overline{\partial} f(x)\right)\right] = \left[\underline{\partial} f(x), \overline{\partial} f(x)\right] = \mathfrak{D} f(x).$$

Отже різниця множин $\underline{\partial} f(x)$ та $-\overline{\partial} f(x)$, обчислена в просторі опуклих множин, співпадає з квазідиференціалом.

На відміну від різниці, що визначається операцією віднімання в просторі опуклих множин, можна розглядати й інші операції взяття різниці. Ми зупинимося на двох з них: операція $\div$ та операція $\dot{-}$.

Операція $\div$.

Нехай функція f – локально ліпшицева і диференційовна за напрямками в точці x.

Означення 6.5.1. *Субдиференціалом Мішеля–Пено функції* f в точці x називається множина

$$\partial_{mp} f(x) = \{h |\, \langle h,g\rangle \le f'(x,g),\ g \in \mathbb{R}^n\}.$$

Супердиференціалом Мішеля–Пено функції f в точці x називається множина

$$\partial^{mp} f(x) = \{h |\, \langle h,g\rangle \ge f'(x,g),\ g \in \mathbb{R}^n\}.$$

Субдиференціал та супердиференціал Мішеля–Пено є опуклими замкнутими множинами. Ці множини можуть бути і порожніми. Припустимо, що одна з цих множин, наприклад $\partial_{mp} f(x)$, непорожня. Оскільки функція f ліпшицева в околі точки x, то похідна $f'(x,g)$ також задовольняє умову Ліпшиця відносно g, тому $|f'(x,g)| \le L\,\|g\|$, де L – константа Ліпшиця, $g \in \mathbb{R}^n$. Якщо $h \in \partial_{mp} f(x)$, то

$$\langle h,g\rangle \le f'(x,g) \le L\,\|g\|, \quad g \in \mathbb{R}^n,$$

тому

$$\|h\| = \max_{\|g\|=1} \langle h,g\rangle \le L.$$

Отже субдиференціал Мішеля–Пено $\partial_{mp} f(x)$ обмежений. Тому він є опуклим компактом. Покладемо

$$p(g) = \max_{h \in \partial_{mp} f(x)} \langle h,g\rangle.$$

Функція p є неперервною та сублінійною. При цьому $p(g) \le f'(x,g) \equiv f'_x(g)$. Покажемо, що p - найбільша сублінійна функція, яка мажорується похідною $f'_x(g)$: якщо $\tilde{p}$ – сублінійна функція, $\tilde{p}(g) \le f'_x(g)$, то $\tilde{p}(g) \le p(g)$ для всіх g. Дійсно, якщо $h \in \partial\tilde{p}$, то $\langle h,g\rangle \le \tilde{p}(g) \le f'_x(g)$, а тому $h \in \underline{\partial} p$. Включення $\underline{\partial}\tilde{p} \subset \underline{\partial} p$ в силу двоїстості Мінковського приводить до потрібної нерівності. Із сказаного випливає, що непорожність субдиференціала Мішеля–Пено рівносильна існуванню сублінійної функції, що мажорується похідною $f'_x(g)$. Аналогічно, непорожність супердиференціала Пено рівносильна існуванню суперлінійної функції, що мінорується похідною $f'_x(g)$. При цьому супердиференціал Мішеля–Пено є опуклим компактом.

Повернемося до квазідиференційовних локально ліпшицевих функцій. Нехай функція f має в точці x квазідиференціал

$$\mathfrak{D}f(x) = \left[\underline{\partial}f(x), \overline{\partial}f(x)\right] = \left[\underline{\partial}f(x), -\left(-\overline{\partial}f(x)\right)\right].$$

Обчислимо різниці $\underline{\partial}f(x) \div (-\overline{\partial}f(x))$ та $(-\overline{\partial}f(x)) \div \underline{\partial}f(x)$. (Нагадаємо, що $U \div V = \{h \mid h + V \subset U\}$.)

Теорема 6.5.1. *Справедливі наступні рівності*

$$\underline{\partial}f(x) \div (-\overline{\partial}f(x)) = \partial_{mp}f(x), \tag{6.5.1}$$

$$-\overline{\partial}f(x) \div (\underline{\partial}f(x)) = \partial^{mp}f(x). \tag{6.5.2}$$

Доведення. Обмежимося доведенням рівності (6.5.1). Нехай $h \in \underline{\partial}f(x) \div (-\overline{\partial}f(x))$. Тоді $h - \overline{\partial}f(x) \subset \underline{\partial}f(x)$ і, отже, для всіх $g \in \mathbb{R}^n$ виконується нерівність

$$\langle h,g\rangle + \max_{h' \in (-\overline{\partial}f(x))} \langle h',g\rangle \le \max_{h' \in \underline{\partial}f(x)} \langle h',g\rangle,$$

або, що те ж саме,

$$\langle h,g\rangle \le \max_{h' \in \underline{\partial}f(x)} \langle h',g\rangle - \max_{h' \in (-\overline{\partial}f(x))} \langle h',g\rangle = f'(x,g),$$

тобто $h \in \partial_{mp}f(x)$. Тим самим ми перевірили включення

$$\partial_{mp}f(x) \supset \underline{\partial}f(x) \div (-\overline{\partial}f(x)).$$

Обернене включення перевіряється таким же чином. $\square$

Операція $\dot{-}$.

Виявляється, що для досить широкого класу функцій ця операція, застосована до множин $\underline{\partial}f(x)$ та $(-\overline{\partial}f(x))$, приводить до субдиференціала Кларка. Нагадаємо, що множина $U \dot{-} V$, де U та V - опуклі компакти, визначається так. Розглядається деяка множина T повної міри, в точках g якої існують градієнти $\nabla p_U(g)$ та $\nabla p_V(g)$ опорних функцій p_U та p_V компактів U та V відповідно, потім складаються різниці $\nabla p_U(g) - \nabla p_V(g)$. Опукле замикання цих різниць і співпадає з множиною $U \dot{-} V$. При цьому існування градієнта $\nabla p_U(g)$ рівносильне тому, що лінійна функція $\langle h,g\rangle$ досягає максимуму на опуклому компакті U в єдиній точці. Ця точка співпадає з градієнтом $\nabla p_U(g)$.

Нехай локально ліпшицева функція f визначена на відкритій множині X, квазідиференційовна в точці $x \in X$ і має в цій точці квазідиференціал

$\mathfrak{D}f(x) = \left[\underline{\partial} f(x), \overline{\partial} f(x)\right] \equiv [U,V]$. Нехай елемент g такий, що лінійна функція $(g,h)\langle h,g\rangle$ досягає максимуму на кожній з множин $U = \underline{\partial} f(x)$ та $-V = -\overline{\partial} f(x)$ в єдиній точці. Позначимо ці точки відповідно через $\varphi_U(g)$ та $-\psi_V(g)$. Із сказаного вище випливає, що $\varphi_U(g) = \nabla p_V(g)$, $-\psi_V(g) = \nabla p_{(-V)}(g)$. Легко перевірити, що $\psi_V(g)$ - єдина точка компакту $V = \overline{\partial} f(x)$, в якій досягає мінімуму на V лінійна функція $\langle h,g\rangle$.

Нехай $T \subset \mathbb{R}^n$. Будемо казати, що множина T має властивість $(\mathcal{E})$ відносно пари $[U,V] = [\underline{\partial} f(x), \overline{\partial} f(x)]$, якщо:

1) лебегова міра множини $\mathbb{R}^n \backslash T$ дорівнює нулю,

2) при $g \in T$ лінійна функція $\langle h,g\rangle$ досягає максимуму на множині $U = \underline{\partial} f(x)$ в єдиній точці $\varphi_U(g)$ і досягає мінімуму на множині $V = \overline{\partial} f(x)$ в єдиній точці $\psi_V(g)$.

Із вище сказаного випливає, що множина $\underline{\partial} f \dot{-} (-\overline{\partial} f(x))$ співпадає з замкнутою опуклою оболонкою сум $\varphi_U(g) + \psi_V(g)$, де g пробігає деяку множину T, що має властивість $(\mathcal{E})$ відносно пари $[U,V]$.

Позначимо через $M(x)$ сукупність функцій f, визначених на відкритій множині X, яка містить точку x, що мають наступні властивості:

а) f задовольняє умову Ліпшиця в деякому околі $B_\delta(x)$ точки x; звідси випливає, що множина V_f точок $y \in B_\delta(x)$, де існує градієнт $\nabla f(y)$, має повну міру;

б) $f-$ квазідиференційовна в точці x;

в) знайдуться такі підмножина $Q \subset V_f$, що має в околі $B_\delta(x)$ повну міру, та квазідиференціал $\mathfrak{D}f(x) = \left[\underline{\partial} f(x), \overline{\partial} f(x)\right] \equiv [U,V]$ функції f в точці x і множина T, що має властивість $(\mathcal{E})$ відносно пари $[U,V]$, що із співвідношень

$$g_k \to g, \quad \alpha_k \downarrow 0, \quad x_k = x + \alpha_k g_k \in Q, \quad g \in T$$

випливає

$$\nabla f(x_k) \to \varphi_U(g) + \psi_V(g),$$

де $\varphi_U(g)$ та $\psi_V(g)$ - визначені вище відображення.

Справедливість властивості в) визначається вибором деякої конкретної пари $\left[\underline{\partial} f(x), \overline{\partial} f(x)\right]$ із квазідиференціалу $\mathfrak{D}f(x)$ і відповідної множини T. При переході до еквівалентної пари та інших T ця властивість може не виконуватись (див. приклад нижче).

Нехай функція f задовольняє умову Ліпшиця в околі $B_\delta(x)$ точки x, $Q \subset B_\delta(x)$, $T \subset \mathbb{R}^n$, причому

$$\mu(B_\delta(x) \backslash Q) = 0, \quad \mu(\mathbb{R}^n \backslash T) = 0,$$

де μ - міра Лебега. Покладемо

$$\partial_T f(x) = \mathrm{cl\ conv}\ D_T,$$

де

$$D_T = \{v |\ \exists x_k = x + \alpha_k g_k \in Q, \quad g_k \to g \in T, \ \ \alpha_k \downarrow 0, \ \nabla f(x_k) \to v\}.$$

Теорема 6.5.2. *Нехай функція f входить в $M(x)$, множина $Q \subset V_f$, квазідиференціал $[U,V] = \left[\underline{\partial} f(x), \overline{\partial} f(x)\right]$ і множина T, що має властивість $(\mathcal{E})$ відносно пари $[U,V]$, такі, що для них виконана умова в) з означення класу $M(x)$. Тоді*

$$\partial_T f(x) = \underline{\partial} f(x) \dot{-} \left(-\overline{\partial} f(x)\right).$$

Доведення. Нехай $v \in D_T$. Тоді знайдуться такі послідовності g_k та α_k, що $g_k \to g$, $\alpha_k \downarrow 0$, $x_k = x + \alpha_k g_k \in Q$, $g \in T$, $\nabla f(x_k) \to v$. Оскільки, з іншого боку,$f \in M(x)$, то $\nabla f(x_k) \to \varphi_U(g) + \psi_V(g)$, тобто $v = \varphi_U(g) + \psi_V(g)$. Таким чином, $D_T \subset \{\varphi_U(g) + \psi_V(g) \,|\, g \in T\}$ і, отже,

$$\partial_T f(x) \subset \text{cl conv}\{\varphi_U(g) + \psi_V(g) \,|\, g \in T\} = \underline{\partial} f(x) \dot{-} \left(-\overline{\partial} f(x)\right).$$

Перевіримо обернене включення. Нехай $g \in \mathbb{R}^n$,$\alpha_k \downarrow 0$,$\varepsilon_k \downarrow 0$. Оскільки перетин Q з кулею $x+\alpha_k B_{\varepsilon_k}(g)$ непорожній, то знайдеться такий елемент $g_k \in B_{\varepsilon_k}(g)$, що

$$x + \alpha_k g_k \in Q. \tag{6.5.3}$$

Таким чином, існує послідовність $\{g_k\}$, для якої $g_k \to g$ і виконана умова (6.5.3). Нехай $g \in T$. Розглянемо елемент $v = \varphi_V(g) + \psi_U(g)$ і послідовності $\{g_k\}$ і $\{\alpha_k\}$, для яких $g_k \to g$, $\alpha_k \downarrow 0$ і виконана умова (6.5.3). За означенням $M(x)$ є справедливим співвідношення $\nabla f(x_k) \to v$, тобто $v \in D_T$. Із цієї обставини одразу випливає, що

$$\partial_T f(x) \supset \underline{\partial} f(x) \dot{-} \left(-\overline{\partial} f(x)\right).$$

□

Наслідок 6.5.1. *Нехай $f \in M(x)$. Тоді*

$$\partial_{Cl} f(x) \supset \underline{\partial} f(x) \dot{-} \left(-\overline{\partial} f(x)\right). \tag{6.5.4}$$

Це твердження одразу випливає з теореми 5.2.13 і теореми 6.5.2.

Наслідок 6.5.1 показує, що для функції з класу $M(x)$ квазідиференціал дозволяє оцінити субдиференціал Кларка знизу.
Якщо знайдеться пара $\left[\underline{\partial} f(x), \overline{\partial} f(x)\right]$, для якої виконується властивість в) з означення $M(x)$ коли множина T співпадає з усім $\mathbb{R}^n$, то

$$\partial_{Cl} f(x) = \underline{\partial} f(x) \dot{-} \left(-\overline{\partial} f(x)\right).$$

Рівність $T = \mathbb{R}^n$ виконується, якщо множини $\underline{\partial} f(x)$ та $\overline{\partial} f(x)$ строго опуклі або складаються з однієї точки.

Покажемо, що множина $M(x)$ досить широка. Насамперед перевіримо, що опуклі функції входять в $M(x)$. З цією метою встановимо справедливість наступного твердження.

Лема 6.5.1. *Нехай f – опукла функція, що визначена на відкритій опуклій множині X, яка містить точку x. Нехай $x_k = x + \alpha_k g_k$, де $\alpha_k \downarrow 0$, $g_k \to g$, причому в точках x_k існує градієнт $\nabla f(x_k)$. Тоді, якщо*

$\nabla f(x_k) \to v$, то $\langle v,g\rangle = \max\limits_{h\in\partial f(x)} \langle h,g\rangle$, де $\partial f(x)$ – субдиференціал функції f в точці x.

Доведення. Нехай елементи $z_k \in \partial f(x)$ такі, що $f'(x,g_k) = \max\limits_{h\in\partial f(x)} \langle h,g_k\rangle = \langle z_k,g_k\rangle$. Вважаємо, що $z_k \to z$. Тоді $f'(x,g) = \langle z,g\rangle$. Розглянемо функцію дійсної змінної $\varphi_k(\alpha) = f(x+\alpha g_k), k = 1,2,\ldots$ Ця функція опукла, тому її права похідна $(\varphi_k)'_+(\alpha)$ неспадна. Звідси випливає, що $(\varphi_k)'_+(\alpha) \geq (\varphi_k)'_+(0)$. Оскільки

$$(\varphi_k)'_+(\alpha) = \lim_{\beta\downarrow 0} \frac{1}{\beta}(\varphi_k(\alpha+\beta) - \varphi_k(\alpha)) = f'(x+\alpha g_k,\, g_k),$$

то

$$f'(x_k,g_k) \geq f'(x,g_k).$$

Пригадуючи, що $f'(x,g_k) = \langle z_k,g_k\rangle$, $f'(x_k,g_k) = \langle \nabla f(x_k), g_k\rangle$, одержимо

$$\langle \nabla f(x_k), g_k\rangle \geq \langle z_k,g_k\rangle.$$

Переходячи в цій нерівності до границі, маємо

$$\langle v,g\rangle \geq \langle z,g\rangle = f'(x,g) = \max_{h\in\underline{\partial} f(x)} \langle h,g\rangle. \tag{6.5.5}$$

Оскільки $v \in \partial f(x)$, то співвідношення (6.5.5) показують, що

$$\langle v,g\rangle = \max_{h\in\partial f(x)} \langle h,g\rangle.$$

□

Теорема 6.5.3. *Якщо функція f опукла в деякому опуклому околі точки x, то $f \in M(x)$.*

Доведення. Нехай Q – множина повної міри, в точках x' якої існує градієнт $\nabla f(x')$, множина T має властивість $(\mathcal{E})$ в точці x. В даному випадку це означає, що при $g \in T$ лінійна функція $\langle h,g\rangle$ досягає максимуму на множині $U = \partial f(x)$ в єдиній точці $\varphi_U(x)$. (Як завжди, у випадку опуклої функції розглядаємо її квазідиференціал вигляду $[\underline{\partial} f(x), \{0\}]$.) Розглянемо такі послідовності g_k та α_k, що $g_k \to g$, $\alpha_k \downarrow 0$, $x_k = x + \alpha_k g_k \in Q$, $g \in T$. Вважаємо, що існує границя $\lim \nabla f(x_k) = v$. Тоді, в силу леми 6.5.1, лінійна функція (g,h) досягає максимуму на множині $\underline{\partial} f(x)$ в точці v. Оскільки $g \in T$, то $v = \varphi_U(g)$. Звідси і випливає співвідношення $f \in M(x)$. □

Теорема 6.5.4. *Нехай функції $h_1(x), \ldots, h_m(x)$ належать $M(x)$, нехай $y = (h_1(x), \ldots, h_m(x))$, і нехай функція $f(y)$ неперервно диференційовна в точці y. Тоді функція $F(x) = f(h_1(x), \ldots, h_m(x))$ належить множині $M(x)$.*

Доведення. Оскільки $h_i \in M(x)$, то для кожного $i \in I = \{1, \ldots, m\}$ знайдуться множина Q_i, квазідиференціал $[U_i, V_i] = \left[\underline{\partial} h_i(x), \overline{\partial} h_i(x)\right]$ функції h_i в точці x та множина T_i, що має властивість $(\mathcal{E})$ відносно пари $[U_i, V_i]$ такі, що
1) Q_i міститься в деякому околі $B_\delta(x)$, має там повну міру і в точках $x' \in Q_i$ існує градієнт $\nabla h_i(x_i')$;
2) із співвідношень

$$g_k \to g, \quad \alpha_k \downarrow 0, \quad x_k = x + \alpha_k g_k \in Q_i, \quad g \in T_i$$

випливає

$$\nabla h_i(x_k) \to \varphi_{U_i}(g) + \psi_{V_i}(g),$$

де $\varphi_{U_i}(g)$ (відповідно, $\psi_{V_i}(g)$) — єдина точка, в якій лінійна функція $l \mapsto \langle g, l\rangle$ досягає максимуму на U_i (відповідно, мінімуму на V_i). Покладемо $Q = Q_1 \cap Q_2 \cap \ldots \cap Q_m$. Зрозуміло, що Q - множина повної міри в $B_\delta(x)$. За теоремою 6.1.3 функція F квазідиференційовна. Приклад 6.1.4 показує, що квазідиференціал $\mathfrak{D}F(x) = \left[\underline{\partial} F(x), \overline{\partial} F(x)\right]$ може бути представлений у вигляді

$$\underline{\partial} F(x) = \left\{ w \middle| \; w = \sum_{i \in I_1} \frac{\partial f}{\partial y^{(i)}}(y)\, \mu_i + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y)\, \lambda_i : \right.$$

$$\left. \lambda_i \in \underline{\partial} h_i(x), \mu_i \in \overline{\partial} h_i(x) \;\; \forall i \in 1:m \right\}, \tag{6.5.6}$$

$$\overline{\partial} F(x) = \left\{ w \middle| \; w = \sum_{i \in I_1} \frac{\partial f}{\partial y^{(i)}}(y)\, \lambda_i + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y)\, \mu_i : \right.$$

$$\left. \lambda_i \in \underline{\partial} h_i(x), \mu_i \in \overline{\partial} h_i(x) \;\; \forall i \in 1:m \right\}, \tag{6.5.7}$$

$$I_1 = \left\{ i \colon \frac{\partial f}{\partial y^{(i)}}(y) < 0 \right\}, \quad I_2 = \left\{ i \colon \frac{\partial f}{\partial y^{(i)}}(y) \geq 0 \right\}.$$

Покажемо, що множина $T = \bigcap_i T_i$ має властивість $(\mathcal{E})$ відносно пари $[U_i, V_i] = \left[\underline{\partial} F(x), \overline{\partial} F(x)\right]$. З цією метою представимо множини $\underline{\partial} F(x)$ та $\overline{\partial} F(x)$ у вигляді

$$\underline{\partial} F(x) = \sum_{i \in I_1} \left| \frac{\partial f}{\partial y^{(i)}}(y) \right| (-\overline{\partial} h_i(x)) + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y) \underline{\partial} h_i(x),$$

$$\overline{\partial} F(x) = \sum_{i \in I_1} \left| \frac{\partial f}{\partial y^{(i)}}(y) \right| (-\underline{\partial} h_i(x)) + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y) \overline{\partial} h_i(x).$$

Нехай $g \in T$. Тоді

$$\max_{l \in \underline{\partial} F(x)} \langle g,l \rangle = \sum_{i \in I_1} \left| \frac{\partial f}{\partial y^{(i)}}(y) \right| \max_{l \in -\overline{\partial} h_i(x)} \langle g,l \rangle + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y) \max_{l \in \underline{\partial} h_i(x)} \langle g,l \rangle,$$

$$\min_{l \in \overline{\partial} F(x)} \langle g,l \rangle = \sum_{i \in I_1} \left| \frac{\partial f}{\partial y^{(i)}}(y) \right| \min_{l \in -\underline{\partial} h_i(x)} \langle g,l \rangle + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y) \min_{l \in \overline{\partial} h_i(x)} \langle g,l \rangle.$$

Оскільки функція $l \mapsto \langle g,l \rangle$ досягає максимуму на множині $U_i = \underline{\partial} h_i(x)$ в єдиній точці $\varphi_{U_i}(g)$, а на множині $-V_i = -\overline{\partial} h_i(x)$ - в єдиній точці $-\psi_{V_i}(g)$, то і на множині $U = \underline{\partial} F(x)$ ця функція досягає максимуму в єдиній точці $\varphi_U(g)$. При цьому

$$\varphi_U(g) = \sum_{i \in I_1} \frac{\partial f}{\partial y^{(i)}}(y)\, \psi_{V_i}(g) + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y)\, \varphi_{U_i}(g).$$

Таким же чином перевіряється, що функція $l \mapsto (g,l)$ досягає мінімуму на множині $V = \overline{\partial} F(x)$ в єдиній точці $\psi_V(g)$, де

$$\psi_V(g) = \sum_{i \in I_1} \frac{\partial f}{\partial y^{(i)}}(y)\, \varphi_{U_i}(g) + \sum_{i \in I_2} \frac{\partial f}{\partial y^{(i)}}(y)\, \psi_{V_i}(g).$$

Зауважимо, що

$$\varphi_U(g) + \psi_V(g) = \sum_{i=1}^{m} \frac{\partial f}{\partial y^{(i)}}(y)\, (\varphi_{U_i}(g) + \psi_{V_i}(g)). \tag{6.5.8}$$

Припустимо тепер, що дано послідовності $\{g_k\}$ та $\{\alpha_k\}$, що мають такі властивості:

$$g_k \to g, \quad \alpha_k \downarrow 0, \quad x_k = x + \alpha_k g_k \in Q, \quad g \in T.$$

Оскільки в точці x_k множини Q існують градієнти $\nabla h_i(x_k)$, то в цій же точці існує градієнт $\nabla F(x_k)$ функції F, при цьому

$$\nabla F(x_k) = \sum_{i=1}^{m} \frac{\partial f}{\partial y^{(i)}}(y_k) \nabla h_i(x_k), \tag{6.5.9}$$

де покладено $y_k = (h_1(x_k),...,h_m(x_k))$. Оскільки функція f є неперервно диференційовною, а $\nabla h_i(x_k) \to \varphi_{U_i}(g) + \psi_{V_i}(g)$ (це справедливо, оскільки $h_i \in M(x)$), то, переходячи в (6.5.9) до границі, одержимо

$$\nabla F(x_k) \to \sum_{i=1}^{m} \frac{\partial f}{\partial y^{(i)}}(y)(\varphi_{U_i}(g) + \psi_{V_i}(g)). \tag{6.5.10}$$

Застосовуючи (6.5.8) та (6.5.10), прийдемо до співвідношення

$$\nabla F(x_k) \to \varphi_U(g) + \psi_V(g)\,,$$

яке і показує, що $F \in M(x)$. □

Наслідок 6.5.2. *Множина $M(x)$ разом з кожними двома своїми елементами містить їх суму, добуток та частку (якщо знаменник відмінний від нуля).*

Доведення. Для доведення достатньо розглянути в просторі $\mathbb{R}^2$ функції $f(y^{(1)},y^{(2)}) = y^{(1)} + y^{(2)}$, $f(y^{(1)},y^{(2)}) = y^{(1)} \cdot y^{(2)}$, $f(y^{(1)},y^{(2)}) = y^{(1)}/y^{(2)}$ та застосувати теорему. $\square$

Зауваження 6.5.1. Із доведення теореми 6.5.4 випливає, що один з квазідиференціалів $\left[\underline{\partial}F(x), \overline{\partial}F(x)\right]$ функції F, при якому виконується властивість в) із означення класу $M(x)$, обчислюється за формулами (6.5.6), (6.5.7). При цьому множина $T = \cap_i T_i$.

Зауваження 6.5.2. Нехай $f = f_1 + f_2$, де f_1 – опукла, f_2 – угнута функції. Оскільки $\partial_{Cl} f_1(x)$ співпадає з субдиференціалом $\underline{\partial} f_1(x)$, а $\partial_{Cl} f_2(x)$ – з супердиференціалом $\overline{\partial} f_2(x)$, то (див. теорему 5.2.14) сума $\underline{\partial} f_1(x) + \overline{\partial} f_2(x)$ дає оцінку субдиференціалу Кларка згори:

$$\partial_{Cl} f(x) \subset \underline{\partial} f_1(x) + \overline{\partial} f_2(x).$$

В той же час, як випливає з тверджень 6.5.3, 6.5.4, зауваження 6.5.1 та наслідку 6.5.2, функція f входить в $M(x)$, причому властивість в) із означення $M(x)$ виконується для пари $\left[\underline{\partial} f(x), \overline{\partial} f(x)\right]$. Звідси випливає (див. наслідок 6.5.2), що різниця $\underline{\partial} f_1(x) \dot{-} \left(-\overline{\partial} f_2(x)\right)$ дає оцінку субдиференціалу Кларка знизу. Таким чином,

$$\underline{\partial} f_1(x) \dot{-} \left(-\overline{\partial} f_2(x)\right) \subset \partial_{Cl}(f_1 + f_2)(x) \subset \underline{\partial} f_1(x) + \overline{\partial} f_2(x).$$

Теорема 6.5.5. *Нехай функції $h_1(x), \ldots, h_m(x)$, що визначені на відкритій множині $X \subset \mathbb{R}^n$, належать $M(x)$, де $x \in X$ та*

$$\underline{h}(x') = \max_{i \in \overline{1,m}} h_i(x'), \quad \overline{h}(x') = \min_{i \in \overline{1,m}} h_i(x') \quad \forall x' \in X.$$

Тоді функції $\underline{h}$ та $\overline{h}$ містяться в $M(x)$.

Доведення. Обмежимося доведенням для $\underline{h}$. Нехай множини Q_i, квазідиференціали $[U_i,V_i] = [\underline{\partial} h_i(x),\overline{\partial} h_i(x)]$ та множини T_i визначені за функцією h_i так само, як і при доведенні теореми 6.5.4. Нагадаємо, що один з квазідиференціалів функції $\underline{h}$ в точці x має вигляд $[\underline{\partial} h(x),\overline{\partial} h(x)] = [U,V]$, де

$$U = \operatorname{conv} \bigcup_{i \in R(x)} \left\{ \underline{\partial} h_i(x) - \sum_{j \in R(x), j \neq i} \overline{\partial} h_j(x) \right\},$$

$$V = \sum_{i \in R(x)} \overline{\partial} h_i(x); \qquad (6.5.11)$$

тут $R(x) = \{i \,|\, h_i(x) = \underline{h}(x)\}$.

Нехай $\tilde{T}$ – множина в $\mathbb{R}^n$, що має властивість $(\mathcal{E})$ відносно пари $[U,V]$. Покладемо $Q = \bigcap\limits_i Q_i$, $T = (\bigcap\limits_i T_i) \bigcap \tilde{T}$. Розглянемо послідовності $\{g_k\}$ та $\{\alpha_k\}$, такі, що

$$g_k \to g, \quad \alpha_k \downarrow 0, \quad x_k = x + \alpha_k g_k \in Q, \quad g \in T.$$

Оскільки $h_i \in M(x)$, то

$$\nabla h_i(x_k) \to \varphi_{U_i}(g) + \psi_{V_i}(g) \quad \forall i \in 1:m, \tag{6.5.12}$$

де $\varphi_{U_i}(g)$ та $\psi_{V_i}(g)$ - точки максимуму і мінімуму лінійної функції $l \mapsto (g,l)$ на множинах U_i та V_i відповідно. Оберемо такий індекс $i(g) \in R(x)$, що

$$\left(g,\varphi_{U_{i(g)}}(g) + \psi_{V_{i(g)}}(g)\right) = \max_{i \in R(x)} \left(g,\varphi_{U_i}(g) + \psi_{V_i}(g)\right). \tag{6.5.13}$$

Безпосередньо з означення множин U та V випливає, що

$$\varphi_U(g) + \psi_V(g) = \varphi_{U_{i(g)}}(g) + \psi_{V_{i(g)}}(g). \tag{6.5.14}$$

Оскільки $\underline{h}'(x,g) = \max\limits_{i \in R(x)} h_i'(x,g)$, то

$$\underline{h}(x + \alpha g) = \underline{h}(x) + \alpha \max_{i \in R(x)} h_i'(x,g) + o(\alpha,g), \tag{6.5.15}$$

де $\frac{1}{\alpha} o(\alpha,g) \underset{\alpha\downarrow 0}{\longrightarrow} 0$ рівномірно по g. (Останнє випливає із ліпшицевості функції $\underline{h}$.) Рівність (6.5.15) показує, що $i(g) \in R(x_k)$ при достатньо великих k. Крім того, з цієї рівності випливає, що $i(g) \in R(x')$ для x', досить близьких до вказаних x_k. Припустимо, що в точках x_k існує градієнт $\nabla \underline{h}(x_k)$ функції $\underline{h}$. Тоді, як показують наведені вище міркування,

$$\underline{h}(x_k) = h_{i(g)}(x_k), \quad \nabla \underline{h}(x_k) = \nabla h_{i(g)}(x_k),$$

Використовуючи (6.5.13) та (6.5.14), одержимо

$$\nabla \underline{h}(x_k) = \nabla h_{i(g)}(x_k) \to \varphi_{U_{i(g)}}(g) + \psi_{V_{i(g)}}(g) = \varphi_U(g) + \psi_V(g).$$

$\square$

Зауваження 6.5.3. Один із квазідиференціалів $[\underline{\partial h}(x),\overline{\partial}\underline{h}(x)] = [U,V]$ функції $\underline{h}$, при якому виконується властивість в) із означення класу $M(x)$, обчислюється за формулами (6.5.11). Подібне твердження справедливе і для функції мінімуму $\overline{h}(x)$. При цьому в якості T обирається $(\bigcap\limits_i T_i) \bigcap \tilde{T}$.

Приклад 6.5.1. Для точки $z = (x,y) \in \mathbb{R}^2$ покладемо

$$f(x,y) = \max\left\{\min\{x, -y\}; x - y\right\}.$$

Нас буде цікавити поведінка цієї функції в околі точки $z_0 = (0,0)$. Множина $\partial_{Cl} f(z_0)$ співпадає з трикутником, породженим точками (1,0), (0,-1), (-1,1) (див. рис. 6.5.1)

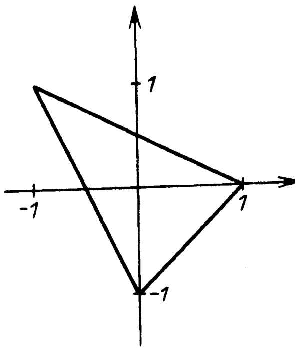


Рис. 6.5.1:

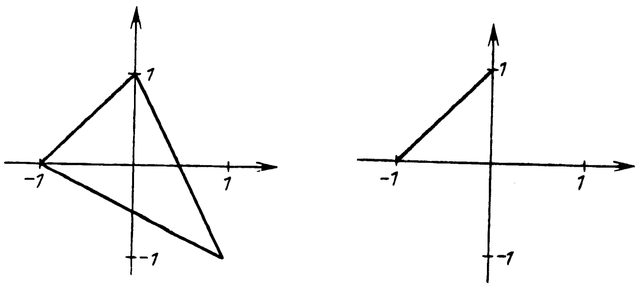


Рис. 6.5.2:

Покладемо

$$f_1(x,y) = \min\{x, -y\}, \quad f_2(x,y) = x - y,$$

$$f_3(x,y) = x, \quad f_4(x,y) = -y.$$

Функції f_3 та f_4 неперервно диференційовні і тому входять в $M(z_0)$. Із теореми 6.5.5 випливає, що функції $f_1(x,y)$ та $f(x,y)$ також входять в клас $M(z_0)$. Обчислимо квазідиференціали цих функцій, використовуючи формули для обчислення квазідиференціала максимуму та мінімуму, які були наведені в теоремі 6.1.2. В результаті одержимо квазідиференціал $[U,V] = [\underline{\partial} f(z_0), \overline{\partial} f(z_0)]$ функції f в точці z_0, де $\underline{\partial} f(z_0)$ співпадає з трикутником, породженим точками $(1, -1), (0,1), (-1,0)$, а $\overline{\partial} f(z_0)$ – з відрізком, кінцями якого є точки (0,1) та (-1,0) (див. рис. 6.5.2).

Зауваження 6.5.3 показує, що властивість в) із означення класу $M\,(x)$ виконується, якщо за квазідиференціал взяти вказану пару $[U,V]$. Множина T в даному випадку складається з таких точок g, що лінійна функція $l \mapsto (g,l)$ досягає максимуму на трикутнику $\underline{\partial} f(z_0)$ лише в його вершинах і мінімуму на відрізку $\overline{\partial} f(z_0)$ лише на його кінцях. Неважко перевірити, що T співпадає з усією площиною, за виключенням променів з вершиною в нулі, що проходять через точки $\left(-1/\sqrt{2}, 1/\sqrt{2}\right)$, $\left(2/\sqrt{5}, 1/\sqrt{5}\right)$, $\left(1/\sqrt{2}, -1\sqrt{2}\right)$, $\left(-1/\sqrt{5}, 2/\sqrt{5}\right)$. Простий підрахунок показує, що елементи $\varphi_U(g) + \psi_V(g)$ для вказаних g співпадають з однією з точок $(0, -1)$, $(1, 0)$, $(-1,1)$. Тому множина $\underline{\partial} f_1(z_0) \dot{-} \left(-\overline{\partial} f_2(z_0)\right)$ представляє собою трикутник з вершинами в даних точках. У даному випадку ця множина співпадає з субдиференціалом Кларка.

Приклад 6.5.2. Нехай функція f визначена на площині $\mathbb{R}^2$ формулою

$$f(x,y) = \max\left\{y - x^2, 0\right\} + \min\left\{y + x^2, 0\right\}.$$

Ця функція в точці $z_0 = (0,0)$ вона має градієнт $\nabla f(z_0) = e_2$, а її субдиференціал Кларка $\partial_{Cl} f(z_0)$ представляє собою відрізок U з кінцями z_0 та e_2. Тут $e_2 = (0,1)$ — одиничний орт.

З тверджень 6.5.4, 6.5.5 та зауважень 6.5.1, 6.5.3 випливає, що $f \in M(z_0)$, причому в квазідиференціалі $[\underline{\partial} f(z_0), \overline{\partial} f(z_0)]$, відносно якого виконується властивість в) із визначення класу $M(x)$, як субдиференціал $\underline{\partial} f(z_0)$, так і супердиференціал $\overline{\partial} f(z_0)$ співпадають з відрізком U. Множина T в цьому випадку складається з векторів $g = \left(g^{(1)}, g^{(2)}\right)$, у яких $g^{(2)} \neq 0$. Для будь-яких $g \in T$ виконується рівність

$$\varphi_U(g) + \psi_V(g) = e_2.$$

Тому в даному випадку

$$\underline{\partial} f_1(z_0) \dot{-} \left(-\overline{\partial} f_2(z_0)\right) = \{e_2\}.$$

При цьому

$$e_2 = \nabla f(z_0) \subset U = \partial_{Cl} f(z_0),$$

але

$$\underline{\partial} f_1(z_0) \dot{-} \left(-\overline{\partial} f_2(z_0)\right) \neq \partial_{Cl} f(z_0).$$

Розглянемо інший квазідиференціал функції f в точці z_0, а саме квазідиференціал $[U,V] = [\{e_2\},\{0\}]$. Тут як T можна взяти множину ненульових векторів $g \in \mathbb{R}^2$. При цьому для будь-якого $g \in T$ виконується рівність

$$\varphi_U(g) + \psi_V(g) = e_2.$$

Нехай $e_1 = (1,0)$, $g = e_1$, $g_k = e_1$ при всіх k, $\alpha_k \downarrow 0, z_k = z_0 + \alpha_k g_k = \alpha_k e_1$. Тоді $\nabla f(z_k) = 0$, отже $\nabla f(z_k)$ не збігається до вектора $\varphi_U(g) + \psi_V(g) = e_2$. Таким чином, відносно пари $[U,V]$ та вказаної множини T властивість в) із означення класу $M(x)$ не виконується. Однак ця властивість буде виконуватись, якщо звузити множину T, відкинувши з неї вектори $g = (g^1, g^2)$, для яких $g^2 \neq 0$.

6.6. ВЕРХНІ ОПУКЛІ АПРОКСИМАЦІЇ ТА КВАЗІДИФЕРЕНЦІАЛИ

Розглянемо визначену на відкритій множині X в просторі $\mathbb{R}^n$ функцію f, яка має в точці $x \in X$ похідну за напрямком $f'(x,g) = f'_x(g)$. Якщо ця похідна сублінійна, то її можна "квазілінеаризувати", тобто виразити через лінійні функції за допомогою операції максимуму:

$$f'(x,g) = \max_{h \in U} \langle h,g \rangle \quad \forall g \in \mathbb{R}^n. \tag{6.6.1}$$

У загальному випадку така квазілінеаризація не завжди можлива. Однак для багатьох задач досить замість рівності мати принаймні нерівність

$$f'(x,g) \leq \max_{h \in U} \langle h,g \rangle,$$

де U — опуклий компакт. Іншими словами, досить вміти мажорувати функцію $f'(x,g)$ деякою сублінійною функцією $p(g)$. Пояснимо це на простому прикладі. Нехай відома сублінійна функція $p(g)$, яка мажорує похідну $f'(x,g)$. Випишемо за її допомогою необхідні умови мінімуму. Якщо функція f досягає локального мінімуму в точці x, то $f'(x,g) \geq 0$ при всіх $g \in \mathbb{R}^n$. Тому $p\,(g) \geq 0$ при всіх $g \in \mathbb{R}^n$, а це означає, що $0 \in \underline{\partial} p$, де $\underline{\partial} p$ — субдиференціал функції p. Тим самим необхідні умови записані у звичній формі: деяка множина повинна містити нуль. Звичайно, чим менше $p(g)$ відрізняється від $f'(x,g)$, тим точніші зазначені умови. Перейдемо до точних означень.

Означення 6.6.1. Сублінійна функція p називається *верхньою опуклою апроксимацією (в. о. а.) функції f в точці $x \in X$*, якщо існує похідна за напрямками $f'(x,g)$ і виконується умова $p(g) \geq f'(x,g)$ для всіх $g \in \mathbb{R}^n$.

Безпосередньо з означення випливає, що

$$f\,(x + \alpha g) \leq f\,(x) + \alpha p\,(g) + o_g\,(\alpha)\,, \tag{6.6.2}$$

де $o_g\,(\alpha)\,/\alpha \to 0$ при $\alpha \downarrow 0$.

Відзначимо, що термін "верхня опукла апроксимація" не зовсім точно відображає суть справи. Як вже відзначалося вище, точніше було б говорити про "верхню опуклу мажорацію" і називати p опуклою мажорантою.

Аналогічно вводиться поняття *нижньої угнутої апроксимації (н. у. а.)*. Так називається суперлінійна функція q, яка має ту властивість, що $q\,(g) \leq f'(x,g)$ при всіх g.
Зрозуміло, що в. о. а. і н. у. а. функції f в точці x визначаються неоднозначно. Наведемо кілька прикладів.

Приклад 6.6.1. Нехай f — опукла функція. Оскільки похідна $f'(x,g)$ сублінійна, то вона є в. о. а. При цьому $f'(x,g)$ — найменша в. о. а. Як н. у. а. в даному випадку можна взяти лінійні функції $q\,(g) = \langle h,g \rangle$, де h

— елемент субдиференціала $\underline{\partial} f(x)$ функції f. Зрозуміло, що найбільшої н. у. а. в даному випадку не існує.

Для угнутої функції f картина симетрична. Як н. у. а. можна взяти похідну $f'(x,g)$, а в. о. а. породжуються елементами супердиференціала $\overline{\partial} f(x)$ функції f.

Приклад 6.6.2. Нехай f — локально ліпшицева функція, яка має похідну за напрямками в точці x. Тоді верхня похідна Кларка $g \to f_{Cl}^{\uparrow}(x,g)$ і нижня похідна Кларка $g \to f_{Cl}^{\downarrow}(x,g)$ є відповідно в. о. а. і н. у. а. функції f в точці x. Дійсно, функція $f_{Cl}^{\uparrow}(x,g)$ сублінійна, а функція $f_{Cl}^{\downarrow}(x,g)$ суперлінійна за g. Крім того,

$$f_{Cl}^{\downarrow}(x,g) \leq f'(x,g) \leq f_{Cl}^{\uparrow}(x,g).$$

Приклад 6.6.3. Нехай знову f — локально ліпшицева функція, яка має похідну за напрямками $f'(x,g)$. Функція $f'(x,g)$ задовольняє умову Ліпшиця, тому має сенс розглянути похідну Кларка функції $f_x'(g) = f'(x,g)$ в нулі: $p(g) = (f_x')_{Cl}^{\uparrow}(0,g)$. Зрозуміло, що p — сублінійна функція. При цьому $p(g) \geq f_x'(g) = f'(x,g)$ (ця нерівність випливає, наприклад, з результатів наступного пункту). Отже, p — в. о. а. функції f в точці x.

Функція p цікава сама по собі. З'ясуємо деякі її властивості.

Нехай h — додатньо однорідна першого ступеня локально ліпшицева функція, що визначена на $\mathbb{R}^n$. Покладемо

$$p_h(g) = \sup_{u \in \mathbb{R}^n} (h(u+g) - h(u)). \tag{6.6.3}$$

Лема 6.6.1. *Нехай ε,δ – додатні числа. Тоді*

$$\sup_{\|u\|<\varepsilon, 0<\alpha<\delta} \frac{1}{\alpha}(h(u+\alpha g) - h(u)) = p_h(g). \tag{6.6.4}$$

Доведення. Використовуючи додатню днорідність функції h, маємо

$$\frac{1}{\alpha}(h(u+\alpha g) - h(u)) = h\left(\frac{u}{\alpha} + g\right) - h\left(\frac{u}{\alpha}\right).$$

Довільний елемент $u \in \mathbb{R}^n$ можна записати у вигляді $u = u'/\alpha$, де $\|u'\| < \varepsilon, \alpha < \delta$. Тому

$$\sup_{\|u\|<\varepsilon, \alpha \in (0,\delta)} \frac{1}{\alpha}(h(u+\alpha g) - h(u)) =$$

$$= \sup_{\|u\|<\varepsilon, \alpha \in (0,\delta)} h\left(\frac{u}{\alpha} + g\right) - h\left(\frac{u}{\alpha}\right) =$$

$$= \sup_{u \in \mathbb{R}^n} (h(u+g) - h(u)) = p_h(g).$$

$\square$

Теорема 6.6.1. *Справедлива рівність*

$$p_h(g) = h^{\uparrow}_{Cl}(0,g).$$

Доведення. За означенням

$$h^{\uparrow}_{Cl}(0,g) = \inf_{\varepsilon\downarrow 0,\delta\downarrow 0} \sup_{\|u\|<\varepsilon,\alpha\in(0,\delta)} \frac{1}{\alpha}\left(h(u+\alpha g) - h(u)\right).$$

Тому справедливість твердження випливає з леми 6.6.1. □

Наслідок 6.6.1. *Визначена формулою (6.6.3) функція $p_h(g)$ сублінійна.*

Наслідок 6.6.2. *При всіх $g \in \mathbb{R}^n$ виконується нерівність*

$$h^{\uparrow}_{Cl}(0,g) \geq h(g).$$

Доведення. Дійсно, $h^{\uparrow}_{Cl}(0,g) = p_h(g)$, а нерівність $p_h(g) \geq h(g)$ випливає безпосередньо з (6.6.3):

$$\sup_u \left(h(u+g) - h(u)\right) \geq h(u_0+g) - h(u_0), u_0 = 0.$$

□

Відзначимо ще одну цікаву властивість функції p_h. Нагадаємо, що для опуклих компактів була визначена операція віднімання $\div$. Ця операція може бути визначена і для довільних множин. Якщо U,V — деякі підмножини $\mathbb{R}^n$, то

$$U \div V = \{x | x + V \subset U\}.$$

Нагадаємо, що для функції h, заданої на $\mathbb{R}^n$, надграфік epi h визначається рівністю epi $h = \{[x,\lambda] | \lambda \geq h(x)\}$. Якщо h — додатньо однорідна першого ступеня, то epi h — конус. Якщо функція h сублінійна, то epi h — опуклий конус.

Теорема 6.6.2. *Справедлива рівність*

$$\text{epi}\, p_h = \text{epi}\, h \div \text{epi}\, h.$$

Доведення. 1) Нехай $[g,\lambda] \in \text{epi}\, h \div \text{epi}\, h$. Тоді $[g,\lambda] + \text{epi}\, h \subset \text{epi}\, h$, тобто

$$[g,\lambda] + [u,\mu] \in \text{epi}\, h \ \ \forall [u,\mu] \in \text{epi}\, h.$$

Це включення показує, що $\lambda+\mu \geq h(g+u)$, якщо $\mu \geq h(u)$. Зокрема, при $\mu = h(u)$ маємо $\lambda \geq p(g+u) - p(u)$ для будь-якого u, звідки випливає нерівність

$$\lambda \geq \sup_u \left(p(g+u) - p(u)\right) = p_h(g). \tag{6.6.5}$$

Співвідношення (6.6.5) означає, що $[g,\lambda] \in \text{epi}\, p_h$. Цим самим доведено включення epi $h \div$ epi $h \subset$ epi p_h.

2) Перевіримо зворотне включення. Нехай $[g,\lambda] \in \text{epi}\ p_h$. Тоді для всіх $u \in \mathbb{R}^n$ виконується нерівність $\lambda \geq h(g+u) - h(u)$, котра показує, що

$$[g+u,\lambda+h(u)] = [g,\lambda] + [u,h(u)] \in \text{epi}\ h.$$

Якщо $[u,\mu] \in \text{epi}\ h$, то $\mu \geq h(u)$ і тому $[g,\lambda]+[u,\mu] \in \text{epi}\ h$. Це означає, що $[g,\lambda] + \text{epi}\ h \subset \text{epi}\ h$, тобто $[g,\lambda] \in \text{epi}\ h \dotdiv \text{epi}\ h$. □

Повернемося до прикладів в. о. а. З результатів попереднього пункту випливає, що в. о. а. функції $p(g) = (f'_x)^{\uparrow}_{Cl}(0,g)$, яку було розглянуто в прикладі 6.6.3, може бути записана так:

$$p(g) = \sup_{u \in \mathbb{R}^n} \left(f'_x(g+u) - f'_x(u) \right).$$

Приклад 6.6.4. Нехай функція f квазідиференційовна в точці x, тобто

$$f'_x(g) = \max_{l \in U} \langle l,g \rangle + \min_{l' \in V} \langle l',g \rangle,$$

де $[U,V] = \left[\underline{\partial} f(x), \bar{\partial} f(x)\right]$ — квазідиференціал функції f у точці x. Покладемо

$$p^U(g) = \max_{l \in U} \langle l,g \rangle, \quad q^V(g) = \min_{l' \in V} \langle l',g \rangle.$$

Зрозуміло, що

$$q^V(g) + \langle l,g \rangle \leq f'_x(g) \leq p^U(g) + \langle l',g \rangle$$

для всіх $l \in U$, $l' \in V$. Оскільки функція $q_l(g) = q^V(g) + \langle l,g \rangle$ суперлінійна, а функція $p_{l'}(g) = p^U(g) + \langle l',g \rangle$ сублінійна, то кожна з функцій $q_l, l \in U$, є н. у. а. функції f в точці x, а кожна з функцій $p_{l'}, l' \in V$, — в. о. а. функції f у цій точці.

Таким чином, квазідиференціал $\left[\underline{\partial} f(x), \bar{\partial} f(x)\right] = [U,V]$ функції f в точці x породжує цілий набір в. о. а. $\{p_{l'} | l' \in V\}$ і н. у. а. $\{q_l | l \in U\}$. Ці набори, як неважко перевірити, мають наступну властивість: для кожного $g \in \mathbb{R}^n$ знайдуться такі $l' \in V$ і $l \in U$, що

$$f'_x(g) = \langle l',g \rangle + p^U(g) = \langle l,g \rangle + q^V(g).$$

Це співвідношення показує, що

$$f'_x(g) = \min_{l' \in V} p_{l'}(g), \ f'_x(g) = \max_{l \in U} q_l(g).$$

Розглянемо деяку сукупність в. о. а. $\{p_\lambda | \lambda \in \Lambda\}$ функції f в точці x. Назвемо цю сукупність *вичерпним сімейством в. о. а.*, якщо $\inf\limits_{\lambda} p_\lambda(g) = f'_x(g)$ для всіх $g \in \mathbb{R}^n$. Аналогічно, сукупність н. у. а. $\{q_\lambda | \lambda \in \Lambda\}$ функції f в точці x називається *вичерпним сімейством н. у. а.*, якщо $\sup\limits_{\lambda} q_\lambda(g) = f'_x(g)$.

Сімейства $\{p_{l'} | l' \in V\}$ і $\{q_l | l \in U\}$, які побудовані в прикладі 6.6.4 для квазідиференційовної функції, представляють собою приклади вичерпних сімейств в. о. а. та н. у. а. Чи можна гарантувати існування таких сімейств без припущення про квазідиференційовність? Відповідь дається наступною теоремою.

Теорема 6.6.3. *Нехай функція f визначена на відкритій множині X і має похідну за напрямками $f'_x(g)$ в точці $x \in X$, яка є обмеженою в тому сенсі, що $\sup\limits_{\|g\| \le 1} f'_x(g) < +\infty$. Тоді, якщо похідна $f'_x(g)$ напівнеперервна зверху як функція напрямку g, то функція f допускає вичерпне сімейство в. о. а.; якщо ж f'_x напівнеперервна знизу, то функція f допускає вичерпне сімейство н. у. а.*

Доведення випливає з наступної леми.

Лема 6.6.2. *Нехай задана на $\mathbb{R}^n$ функція $h(g)$ додатньо однорідна першого степеня і обмежена в тому сенсі, що $\sup\limits_{\|g\| \le 1} h(g) < \infty$. Тоді, якщо $h(g)$ напівнеперервна зверху, то існує сімейство таких сублінійних функцій $\{p_\lambda | \lambda \in \Lambda\}$, що $h(g) = \inf\limits_{\lambda \in \Lambda} p_\lambda(g)$. Якщо ж h напівнеперервна знизу, то існує сімейство таких суперлінійних функцій $\{q_\lambda | \lambda \in \Lambda\}$, що $h(g) = \sup\limits_{\lambda \in \Lambda} q_\lambda(g)$. При цьому множину індексів Λ можна вибрати однаковою для всіх $h(g)$.*

Доведення. Обмежимося випадком напівнеперервної знизу функції. Покладемо $\Lambda = S \times (0,1)$, де $S = \{g| \; \|g\| = 1\}$ і сконструюємо сімейство функцій $\{l_\lambda | \lambda \in \Lambda\}$, яка має наступну властивість

$$h(g) = \sup_{\lambda \in \Lambda} l_\lambda(g), \quad g \in S. \tag{6.6.6}$$

Нехай $\lambda \in \Lambda$, тобто $\lambda = [u,\varepsilon]$, де $u \in S, \varepsilon \in (0,1)$. Оскільки функція h напівнеперервна знизу в точці u, то знайдеться таке δ, що $h(g) > h(u) - \varepsilon$ при $g \in B_\delta(u)$. Покладемо

$$l_\lambda(g) = -\frac{1}{\delta^2}\left(h(u) - \varepsilon - M\right)\|g - u\|^2 + \left(h(u) - \varepsilon\right), \quad g \in S,$$

де число M вибрано так, щоб

$$\inf_{g \in S} h(g) > M + 1, \quad \left(1 - \delta^2/2\right)\left(h(u) - \varepsilon\right) > M. \tag{6.6.7}$$

Вкажемо деякі властивості функції l_λ:

$$1) \quad l_\lambda(u) = h(u) - \varepsilon; \tag{6.6.8}$$

$$2) \; l_\lambda(g) \le h(g) \quad \forall g \in S. \tag{6.6.9}$$

Дійсно, якщо $g \in B_\delta(u)$, то $h(g) > h(u) - \varepsilon \ge l_\lambda(g)$. Тут використовується нерівність $h(u) - \varepsilon > \inf\limits_{g \in S} h(g) - 1 > M$. Якщо ж $\|g - u\| > \delta$, то

$$l_\lambda(g) = -\frac{1}{\delta^2}\|u - g\|^2\left(h(u) - \varepsilon - M\right) + \left(h(u) - \varepsilon\right) \le$$

$$-\left(h\left(u\right)-\varepsilon-M\right)+\left(h\left(u\right)-\varepsilon\right)=M<h\left(g\right).$$

З (6.6.8) і (6.6.9) випливає рівність (6.6.6). Нехай $\lambda=[u,\varepsilon]\in\Lambda$. Тоді $u\in S$. Тому, якщо $g\in S$, то $\|u-g\|^2=2-2\langle u,g\rangle$. Використовуючи це, отримуємо

$$\begin{aligned}&l_\lambda\left(g\right)=-\tfrac{1}{\delta^2}\left(h\left(u\right)-\varepsilon-M\right)\left(2-2\langle u,g\rangle\right)+\left(h\left(u\right)-\varepsilon\right)=\\&=\tfrac{2}{\delta^2}\left(h\left(u\right)-\varepsilon-M\right)\langle u,g\rangle-\tfrac{2}{\delta^2}\left(\left(h\left(u\right)-\varepsilon\right)\left(1-\tfrac{\delta^2}{2}\right)-M\right).\end{aligned}\tag{6.6.10}$$

Покладемо

$$\beta_1=\frac{2}{\delta^2}\left(h\left(u\right)-\varepsilon-M\right),\quad \beta_2=-\frac{2}{\delta^2}\left(\left(h\left(u\right)-\varepsilon\right)\left(1-\frac{\delta^2}{2}\right)-M\right).$$

Тоді, як випливає з (6.6.7), $\beta_1>0,\beta_2<0$. Визначимо на $\mathbb{R}^n$ функцію q_λ, поклавши

$$q_\lambda\left(g\right)=\left(u,g\right)\beta_1+\|g\|\beta_2.$$

Оскільки $\beta_2<0$, то функція q_λ суперлінійна. В той же час з (6.6.10) випливає, що $q_\lambda\left(g\right)=l_\lambda\left(g\right)$ при $g\in S$. Тому з додатної однорідності функцій h і q_λ отримуємо, при всіх g з $\mathbb{R}^n$ виконується рівність

$$h\left(g\right)=\sup_{\lambda\in\Lambda}q_\lambda\left(g\right).$$

$\square$

6.7. ε–КВАЗІДИФЕРЕНЦІАЛИ

Якщо функція f квазідиференційовна в точці x, то її похідна може бути представлена у вигляді суми максимуму і мінімуму лінійних функцій. У тому випадку коли таке представлення неможливе, можна спробувати дати наближене з точністю до заданого $\varepsilon>0$ представлення похідної у вигляді зазначеної суми. Подібна спроба приводить до поняття ε–квазідиференціалу та апроксимативної квазідиференційовності. Перейдемо до точних означень.

Розглянемо визначену на відкритій множині $X\subset\mathbb{R}^n$ диференційовну за напрямками в точці $x\in X$ функцію f. Нехай $\varepsilon>0$.

Означення 6.7.1. Пара компактих множин $\left[\underline{\partial}_\varepsilon f\left(x\right),\bar{\partial}_\varepsilon f\left(x\right)\right]$ називається *ε–квазідиференціалом функції f в точці x*, якщо

$$\left|f'\left(x,g\right)-\left(\max_{\upsilon\in\underline{\partial}_\varepsilon f(x)}\langle\upsilon,g\rangle+\min_{w\in\bar{\partial}_\varepsilon f(x)}\langle w,g\rangle\right)\right|\le\varepsilon\|g\|\quad\forall g\in\mathbb{R}^n.\tag{6.7.1}$$

З формули (6.7.1) випливає, що пара $\left[\underline{\partial}_\varepsilon f\left(x\right),\bar{\partial}_\varepsilon f\left(x\right)\right]$ є ε–квазідиференціалом тоді і тільки тоді, коли будь-яка еквівалентна їй пара є

ε–квазідиференціалом. Отже, ε–квазідиференціал можна розглядати як елемент простору опуклих множин. Зрозуміло, що існують і нееквівалентні пари, які є ε–квазідиференціалами функції f в точці x. Кожен ε–квазідиференціал є ε'–квазідиференціалом, якщо $\varepsilon' > \varepsilon$.

Означення 6.7.2. Функція f називається апроксимативно квазідиференційовною в точці x, якщо існує ε – квазідиференціал цієї функції в даній точці для кожного $\varepsilon > 0$.

Зрозуміло, що кожна квазідиференційовна функція є апроксимативно квазідиференційовною; зворотне твердження невірне. Наступна теорема показує, що клас апроксимативно квазідиференційовних функцій дуже широкий.

Теорема 6.7.1. *Визначена на відкритій множині $X \subset \mathbb{R}^n$ і диференційовна за напрямками в точці $x \in X$ функція f є апроксимативно квазідиференційовною тоді і тільки тоді, коли її похідна за напрямом $f'(x,g) = f'_x(g)$ неперервна як функція напрямку.*

Доведення теореми 6.7.1 є наслідком з відомої теореми Стоуна–Вейерштрасса, яку ми сформулюємо в наступному вигляді.

Теорема 6.7.2. *Нехай Z – лінійний підпростір простору $C(Q)$ всіх неперервних дійснозначних функцій, визначених на компакті Q. Припустимо, що Z містить константи і є решіткою, тобто разом із будь-якими двома своїми елементами z_1,z_2 цей підпростір містить функції*

$$\underline{z}(x) = \max\{z_1(x), z_2(x)\}, \quad \bar{z}(x) = \min\{z_1(x), z_2(x)\}.$$

Тоді підпростір Z щільний в $C(Q)$ за рівномірною нормою, тобто для будь-якої неперервної на Q функції h і будь-якого $\varepsilon > 0$ знайдеться такий елемент $z \in Z$, що $|z(x) - h(x)| < \varepsilon$ при всіх $x \in Q$.

Доведення. Доведення теореми 6.7.1.

1) Розглянемо простір L визначених на $\mathbb{R}^n$ функцій, які можуть бути представлені у вигляді суми сублінійної та суперлінійної функцій. Оскільки функції з L додатньо однорідні, то вони повністю визначаються своїми значеннями на одиничній сфері $S = \{g \in \mathbb{R}^n | \|g\| = 1\}$. Позначимо через Z сукупність визначених на S функцій, які є звуженням на S функцій з L, іншими словами, $z \in Z$, якщо $z(g) = l(g)$ для деякої функції $l \in L$ і $g \in S$. Зрозуміло, що $Z \subset C(S)$. Оскільки L є лінійним простором і решіткою, то і Z має зазначені властивості. Нарешті, Z містить константи: функція $z(g) = c$ є звуженням на S функції $s(g) = c\|g\|$, котра є сублінійною при $c \geq 0$ і суперлінійною при $c \leq 0$. Таким чином, всі умови теореми 6.7.2 виконані. В силу цієї теореми кожну неперервну на S

функцію можна з будь-якою точністю рівномірно наблизити елементами з Z.

2) Нехай похідна за напрямком $f'(x,g)$ функції f в точці x неперервна як функція напрямку. Зі сказаного випливає, що для кожного $\varepsilon > 0$ знайдеться така функція $z \in Z$, що

$$|f'(x,g) - z\,(g)| < \varepsilon \quad \forall g \in S. \tag{6.7.2}$$

За означенням Z знайдеться така функція $l \in L$, що $z\,(g) = l\,(g)$, якщо $\|g\| = 1$. Оскільки функції $f'(x,g)$ і $l(g)$ додатньо однорідні, то нерівність (6.7.2) показує, що

$$|f'(x,g) - l\,(g)| < \varepsilon\|g\| \quad \forall g \in \mathbb{R}^n. \tag{6.7.3}$$

Нехай

$$l\,(g) = \max_{u \in U}\langle u,g\rangle + \min_{v \in V}\langle v,g\rangle.$$

Тоді, як випливає з (6.7.3), пара $[U,V]$ є ε–квазідиференціалом функції f в точці x. В силу довільності $\varepsilon > 0$ переконуємося в тому, що функція f апроксимативно квазідиференційовна.

3) Нехай функція f апроксимативно квазідиференційовна. Тоді для будь-якого k знайдеться така функція $l_k \in L$, що

$$|f'(x,g) - l_k\,(g)| < \varepsilon \;\; \forall g \in B.$$

Таким чином, на одиничній кулі B функція $f'(x,g)$ може бути представлена як рівномірна границя послідовності неперервних функцій l_k і тому є неперервною. Оскільки $f'(x,g)$ додатньо однорідна, то з неперервності цієї функції на кулі випливає її неперервність на всьому просторі. $\square$

Сформулюємо і доведемо теорему про ε–квазідиференційовність композиції функцій.

Нехай функції $h_1(x), \ldots, h_m(x)$ визначені і диференційовні за напрямками на відкритій множині $X \subset \mathbb{R}^n$, нехай відображення $H : X \to \mathbb{R}^m$, що має $h_1, \ldots, h_m$ своїми координатними функціями:

$$H\,(x) = (h_1\,(x)\,, \ldots, h_m\,(x))\,, \quad x \in X.$$

Нехай $x \in X, y = H\,(x)$ і функція f диференційовна за Адамаром у точці y. Визначимо вектор $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m) \in \mathbb{R}^m$, де $\varepsilon_i > 0$, а також число $\tilde{\varepsilon} > 0$. Розглянемо ε_i–квазідиференціали $\mathfrak{D}_{\varepsilon_i} h_i\,(x)$ функцій h_i в точці x і $\tilde{\varepsilon}$–квазідиференціал $\mathfrak{D}_{\tilde{\varepsilon}} f\,(y)$ функції f у точці y:

$$\mathfrak{D}_{\varepsilon_i} h_i\,(x) = \left[\underline{\partial}_{\varepsilon_i} h_i\,(x)\,, \bar{\partial}_{\varepsilon_i} h_i\,(x)\right] \qquad i = 1, \ldots, m,$$

$$\mathfrak{D}_{\tilde{\varepsilon}} f\,(y) = \left[\underline{\partial}_{\tilde{\varepsilon}} f\,(y)\,, \bar{\partial}_{\tilde{\varepsilon}} f\,(y)\right].$$

Покладемо

$$M = \max_{\|g\| \leq 1} \left(\sum_{i=1}^{m} \left(h_i'\,(x,g)\right)^2\right)^{1/2}, \tag{6.7.4}$$

$$C_1 = \max_{\upsilon \in \underline{\partial}_{\tilde{\varepsilon}} f(y)} \|\upsilon\|, \quad C_2 = \max_{\upsilon \in \bar{\partial}_{\tilde{\varepsilon}} f(y)} \|\upsilon\|, \quad C = \max\left(C_1, C_2\right). \tag{6.7.5}$$

Припустимо, що число M, яке визначене формулою (6.7.4), скінченне. Це справджується в тому випадку, коли похідні $h_i'(x,g)$ неперервні як функції напрямку. Визначимо число

$$\varepsilon_0 = \tilde{\varepsilon} M + \|\varepsilon\| C,$$

де $\|\varepsilon\|$ — евклідова норма вектора ε. Визначимо вектори $\underline{\nu} = (\underline{\nu}_1, \ldots, \underline{\nu}_m)$ та $\bar{\nu} = (\bar{\nu}_1, \ldots, \bar{\nu}_m)$, що задовольняють наступну властивість:

$$\underline{\nu} \le \nu \le \bar{\nu}, \quad \nu \in \underline{\partial}_{\tilde{\varepsilon}} f(y) \cup \left(-\bar{\partial}_{\tilde{\varepsilon}} f(y)\right).$$

Теорема 6.7.3. *За зроблених вище припущень пара множин $[U,V]$, де*

$$U = \Bigg\{ u | u = \sum_{i=1}^{m} \nu_i \left(\lambda_i + \mu_i\right) - \underline{\nu}_i \lambda_i - \bar{\nu}_i \mu_i :$$

$$\nu = (\nu_1, \ldots, \nu_m) \in \underline{\partial}_{\tilde{\varepsilon}} f(y); \lambda_i \in \underline{\partial}_{\varepsilon_i} h_i(x), \mu_i \in \bar{\partial}_{\varepsilon_i} h_i(x) \Bigg\}, \tag{6.7.6}$$

$$V = \Bigg\{ \upsilon | \upsilon = \sum_{i=1}^{m} \nu_i \left(\lambda_i + \mu_i\right) + \underline{\nu}_i \lambda_i + \bar{\nu}_i \mu_i :$$

$$\nu = (\nu_1, \ldots, \nu_m) \in \bar{\partial}_{\tilde{\varepsilon}} f(y); \lambda_i \in \underline{\partial}_{\tilde{\varepsilon}} f(y), \ \mu_i \in \bar{\partial}_{\varepsilon_i} h_i(x) \Bigg\} \tag{6.7.7}$$

є ε_0–квазідиференціалом функції

$$\varphi(x) = f\left(h_1(x), \ldots, h_m(x)\right) \tag{6.7.8}$$

в точці $x \in X$.

Доведення. Визначена формулою (6.7.7) функція φ диференційовна за напрямками в точці x, до того ж

$$\varphi'(x,g) = f'\left(y, H'(x,g)\right),$$

де $H'(x,g) = \left(h_1'(x,g), \ldots, h_m'(x,g)\right)$. Покладемо

$$\underline{A} = \underline{\partial}_{\tilde{\varepsilon}} f(y), \ \bar{A} = \bar{\partial}_{\tilde{\varepsilon}} f(y),$$

$$\underline{a}(g) = \max_{\mu \in \underline{A}} (\mu, g), \ \bar{a}(g) = \min_{\mu \in \bar{A}} (\mu, g), \ a(g) = \underline{a}(g) + \bar{a}(g).$$

Тоді, як випливає безпосередньо з означення $\tilde{\varepsilon}$ – квазідиференціала, виконується нерівність

$$|f'(y,g) - a(g)| \le \tilde{\varepsilon} \|g\| \quad \forall g \in \mathbb{R}^n.$$

Покажемо, що $a(g)$ задовольняє умові Ліпшиця з константою C, яка визначена формулою (6.7.5). З цією метою зазначимо, що для кожного $g \in \mathbb{R}^n$ знайдеться такий елемент $\mu_1 \in A$, що $a(g) = \langle \mu_1, g \rangle$. Тому $a(g) \le \|\mu_1\| \ \|g\| \le C_1 \|g\|$, де C_1 також визначено формулою (6.7.6).

Оскільки функція $\underline{a}(g)$ сублінійна, то для $g_1,g_2 \in \mathbb{R}^n$ маємо

$$\underline{a}(g_1) = \underline{a}((g_1 - g_2) + g_2) \leq \underline{a}(g_1 - g_2) + \underline{a}(g_2),$$

звідки випливає нерівність

$$\underline{a}(g_1) - \underline{a}(g_2) \leq \underline{a}(g_1 - g_2) \leq C_1 \|g_1 - g_2\| \quad . \tag{6.7.9}$$

Аналогічно встановлюється, що

$$\underline{a}(g_2) - \underline{a}(g_1) \leq C_1 \|g_1 - g_2\| \tag{6.7.10}$$

Нерівності (6.7.9) і (6.7.10) показують, що a задовольняє умову Ліпшица з константою C_1. Подібним же чином показуємо, що $\bar{a}$ задовольняє умову Ліпшица з константою C_2, яка визначена формулою (6.7.5). Щоб переконатися в цьому, досить розглянути сублінійну функцію $-\bar{a}$ і застосувати до неї запропоновані вище міркування. Зі сказаного випливає, що функція $a = \bar{a} + \underline{a}$ задовольняє умову Ліпшица з константою $C = \max(C_1,C_2)$. Покладемо

$$U_i = \underline{\partial}_{\varepsilon_i} h_i(x),\ V_i = \bar{\partial}_{\varepsilon_i} h_i(x),\ l_i(g) = \max_{\mu \in U_i} \langle \mu,g \rangle + \min_{\nu \in V_i} \langle \nu,g \rangle.$$

Нехай $\Im$ – відображення $\mathbb{R}^n \to \mathbb{R}^m$ з координатними функціями l_i:

$$\Im(g) = (l_1(g), \ldots, l_m(g)).$$

З означення ε_i–квазідиференціала маємо:

$$\|H'(x,g) - \Im(g)\| = \left(\sum_{i=1}^{m} |h'_i(x,g) - l_i(g)|^2 \right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^{m} \varepsilon_i^2 \|g\|^2 \right)^{\frac{1}{2}} = \|\varepsilon\| \|g\|.$$

Використовуючи цю нерівність, ліпшицевість (з константою C) функції a і означення числа M, для $g \in \mathbb{R}^n$ одержуємо

$$\begin{aligned}
&\|f'(y,H'(x,g)) - a(\Im(g))\| \leq |f'(y,H'(x,g)) - a(H'(x,g))| + \\
&+|a(H'(x,g)) - a(\Im(g))| \leq \tilde{\varepsilon}\|H'(x,g)\| + C\|H'(x,g) - \Im(g)\| \leq \\
&\leq \tilde{\varepsilon} \left(\textstyle\sum_{i=1}^{m} (h'_i(x,g))^2 \right)^{1/2} + C\|\varepsilon\|\ \|g\| \leq \tilde{\varepsilon} M \|g\| + C\|\varepsilon\|\ \|g\| = \varepsilon_0 \|g\|.
\end{aligned}$$

Для закінчення доведення слід перевірити, що функція $\psi(g) = a(\Im(g))$ може бути представлена у вигляді

$$\psi(g) = \max_{u \in U} \langle u,g \rangle + \min_{v \in V} \langle v,g \rangle,$$

де множини U і V визначені рівностями (6.7.7) і (6.7.6) відповідно. Оскільки функція ψ додатньо однорідна, то

$$\psi'(0,g) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} [\psi(\alpha g) - \psi(0)] = \psi(g).$$

Аналогічно

$$l'_i(0,g) = l_i(g),\ a'(0,g) = a(g).$$

Використовуючи теорему про квазідиференціал композиції, маємо

$$\psi(g) = \psi'(0,g) = \max_{u \in \underline{\partial}\psi(0)} \langle u,g \rangle + \min_{v \in \bar{\partial}\psi(0)} \langle v,g \rangle,$$

$$\underline{\partial}\psi(0) = \Bigg\{ u | u = \sum_{i=1}^{m} (\nu_i(\lambda_i + \mu_i) - \underline{\nu}_i\lambda_i - \bar{\nu}_i\mu_i) : \nu = (\nu_1, \ldots, \nu_m) \in \underline{\partial}a(0),$$

$$\lambda_i \in \underline{\partial}l_i(0), \mu_i \in \bar{\partial}l_i(0), \; i = 1 = 1, \ldots, m \Bigg\};$$

$$\bar{\partial}\psi(0) = \Bigg\{ \upsilon | \upsilon = \sum_{i=1}^{m} (\nu_i(\lambda_i + \mu_i) + \underline{\nu}_i\lambda_i + \bar{\nu}_i\mu_i) : \nu = (\nu_1, \ldots, \nu_m) \in \bar{\partial}a(0),$$

$$\lambda_i \in \underline{\partial}l_i(0), \mu_i \in \bar{\partial}l_i(0), \; i = 1 = 1, \ldots, m \Bigg\},$$

а вектори $\underline{\nu}$ і $\bar{\nu}$ такі, що $\underline{\nu} \leq \nu \leq \bar{\nu}$ для всіх $\nu \in \underline{\partial}a(0) \cup \left(-\bar{\partial}a(0)\right)$. Оскільки пара $\left[\underline{A}, \bar{A}\right] = \left[\underline{\partial}_{\tilde{\varepsilon}} f(y), \bar{\partial}_{\tilde{\varepsilon}} f(y)\right]$ є квазідиференціалом функції a в нулі, а пара $[U_i, V_i] = \left[\underline{\partial}_{\varepsilon_i} h_i(x), \bar{\partial}_{\varepsilon_i} h_i(x)\right]$ є квазідиференціалом функції l_i в нулі, то $\underline{\partial}\psi(0)$ співпадає з множиною U, а $\bar{\partial}\psi(0)$ співпадає з множиною V. $\square$

Теорема 6.7.3 дозволяє встановити формули для обчислення ε–квазідиференціалів суми, добутку, частки, поточкового максимуму і мінімуму ε–квазідиференційовних функцій.

6.8. ДОСЛІДЖЕННЯ МОДЕЛІ МАТЕМАТИЧНОЇ ЕКОНОМІКИ

Розглянемо економічну система, в якій є m учасників і n видів продуктів. Учасник з номером i характеризується своєю функцією корисності U_i і початковим запасом товарів $w_i = \left(w_i^{(1)}, \ldots, w_i^{(n)}\right) \in \mathbb{R}_+^n$. Функція корисності U_i задана на деякій множині, що містить $\mathbb{R}_+^n$ і описує переваги учасника з номером i. Його мета полягає в максимізації своєї функції корисності при так званих бюджетних обмеженнях, які задаються за допомогою вектора цін p. Бюджетне обмеження $Z_i(p)$ i–го учасника при цінах p – це множина

$$Z_i(p) = \left\{ x \in \mathbb{R}_+^n | \langle p, x \rangle \leqslant \langle p, w_i \rangle \right\}, \tag{6.8.1}$$

що складається з усіх векторів (наборів продуктів) $x \geqslant 0$, вартість яких за цінами p не перевищує кількості грошей (p, w_i), які i–й учасник виручить, продавши свій початковий запас за цими цінами. Передбачається, що кожен учасник, не піклуючись про інших, максимізує свою корисність, тобто прагне мати вектор x_i, який є розв'язком екстремальної задачі

$$U_i \to \max, \quad x \in Z_i(p).$$

Може трапитися так, що $\sum\limits_i x_i$ перевищує вектор w, де $w = \sum\limits_i^n w_i$— загальна кількість продуктів. У цьому випадку мета принаймі одного з учасників недосяжна. Якщо $\sum\limits_i^n x_i \leqslant w$ (нерівність покоординатна), то кажуть, що система перебуває в напіврівновазі. Якщо ж $\sum\limits_i^n x_i = w$, то система перебуває в рівновазі.

Набір векторів $(\bar{p}; \bar{x}_1, \ldots, \bar{x}_m)$ називається станом напіврівноваги моделі, якщо:

1) вектор $\bar{p}$ має додатні координати;

2) $\bar{x}_i \in Z_i(\bar{p})$ і

$$\langle \bar{p}, \bar{x}_i \rangle = \max_{x_i \in Z_i(\bar{p})} \langle \bar{p}, x_i \rangle, \quad i = 1, \ldots, m; \tag{6.8.2}$$

3)

$$\sum_{i=1}^m \bar{x}_i \leqslant w. \tag{6.8.3}$$

Вектор $\bar{p}$, що входить у цей набір, називається напіврівноважними цінами. Якщо замість нерівності (6.8.3) реалізується рівність

$$\sum_{i=1}^m \bar{x}_i = w, \tag{6.8.4}$$

то набір $(\bar{p}; \bar{x}_1, \ldots, \bar{x}_m)$ називається станом рівноваги, а вектор $\bar{p}$ –— рівноважними цінами.

Можна показати (див., наприклад, [61], що при деяких природних припущеннях стан рівноваги існує. Як його відшукати? В математичній економіці часто припускають, що рівноважні ціни встановлюються шляхом переговорів між учасниками, які тим чи іншим способом домовляються про ціни між собою, скажемо, підвищують ціни на товар j, якщо цей товар дефіцитний, тобто $\sum\limits_{i=1}^m x_i^{(j)} \geqslant w^{(j)}$ і знижують, якщо він надлишковий, тобто $\sum\limits_{i=1}^m x_i^{(j)} < w^{(j)}$. Ці дії не завжди приводять до рівноважних цін. Можливий інший підхід до проблеми відшукання рівноваги. Припустимо, що в економічній системі існує деякий ціноутворюючий орган, який зацікавлений у встановленні рівноважних цін. З цією метою зазначений орган повинен аналізувати величину відхилення наявних цін від рівноважних. Припустимо, що в нього є деякі уявлення про порівняльний внесок кожного з учасників в економіку в цілому. Тоді для всякого стану $(x_1, \ldots, x_m)$, $x_i \in \mathbb{R}_+^n$ він зможе визначити “сумарну корисність”

$\sum\limits_{i=1}^{m} \rho_i U_i(x_i)$ цього стану, де ρ_i відображають "порівняльну значимість" учасників. За допомогою "сумарної корисності" вдається вказати один із способів оцінки відхилення наявних цін від рівноважних. Опишемо його. При цьому, для простоти, будемо говорити про напіврівноважні ціни (У тому випадку, коли функції U_i строго монотонні, вони співпадуть з рівноважними).

Визначимо множину

$$W = \left\{ X = (x_1, \ldots, x_m) \in \left(\mathbb{R}_+^n\right)^m \mid \sum_{i=1}^{m} x_i \leqslant w \right\}.$$

Елементи конуса $\left(\mathbb{R}_+^n\right)^m$ називаються розподілами, а елементи множини W називаються допустимими розподілами. Розглянемо багатозначні відображення $\bar{Z}(p)$ і $Z(p)$, визначені на конусі $\operatorname{int} \mathbb{R}_+^n$ всіх векторів з додатними координатами:

$$\bar{Z}(p) = \left\{ X = (x_1, \ldots, x_m) \in \left(\mathbb{R}_+^n\right)^m \mid \langle p, x_i \rangle \leqslant \langle p, w_i \rangle, \quad i = 1, \ldots, m \right\}, \tag{6.8.5}$$

$$Z(p) = \bar{Z}(p) \cap W. \tag{6.8.6}$$

Нехай $\rho = (\rho_1, \ldots, \rho_m)$— вектор з додатними координатами. Для розподілу $X = (x_1, \ldots, x_m)$ покладемо

$$U^\rho(X) = \sum_i \rho_i U_i(x_i) \tag{6.8.7}$$

Розглянемо функції

$$\phi_1^\rho(p) = \max_{X \in \bar{Z}(p)} U^\rho(X), \ \phi_2^\rho(p) = \max_{X \in Z(p)} U^\rho(X), \tag{6.8.8}$$

$$\phi^\rho(p) = \phi_1^\rho(p) - \phi_2^\rho(p), \quad p \in \operatorname{int} R_+^n. \tag{6.8.9}$$

Ясно, що множина $\bar{Z}(p)$ складається з таких розподілів $X = \{x_1, \ldots, x_m\}$, що $x_i \in Z_i(p)$, $i = 1, \ldots, m$. Звідси випливає, що

$$\phi_1^\rho(p) = \max_{x_i \in Z_i(p)} \sum_i \rho_i U_i(x_i) = \sum_i \rho_i \max_{x_i \in Z_i(p)} U_i(x_i),$$

тобто $\phi_1^\rho(p)$ – це максимальна сумарна корисність (при вагових коефіцієнтів $\rho_1, \ldots, \rho_m$), яку може досягнути вся спільнота, якщо учасники пов'язані лише бюджетними обмеженнями і не піклуються про допустимість розподілів. Множина $Z(p)$ складається з допустимих розподілів, що входять в $\bar{Z}(p)$. Величина $\phi_2^\rho(p)$ показує максимальну сумарну корисність (з вагами $\rho_1, \ldots, \rho_m$), яку суспільство досягне, якщо воно враховує і бюджетні обмеження учасників і допустимість розподілів. Оскільки $Z(p) \subset \bar{Z}(p)$, то $\phi_1^\rho(p) \geqslant \phi_2^\rho(p)$ і тому $\phi^\rho(p) \geqslant 0$ при всіх p.

Лема 6.8.1. *Рівність $\phi^\rho(\bar{p}) = 0$ виконується тоді й тільки тоді, коли $\bar{p}$ -— напіврівноважні ціни.*

Доведення. Нехай $\phi^{\rho}(\bar{p}) = 0$ і розподіл $\bar{X} = (\bar{x}_1, \ldots, \bar{x}_m) \in Z(p)$ такий, що $\phi_2^{\rho}(\bar{p}) = U^{\rho}\left(\bar{X}\right)$. Тоді

$$\phi_1^{\rho}(\bar{p}) = \phi_2^{\rho}(\bar{p}) = U^{\rho}\left(\bar{X}\right) = \sum_i \rho_i U_i(\bar{x}_i).$$

В той же час

$$\phi_1^{\rho}(p) = \sum_i \rho_i \max_{x_i \in Z_i(p)} U_i(x_i).$$

Величини $\bar{x}_i \in Z_i(p)$, тому

$$U_i(\bar{x}_i) = \max_{x_i \in Z_i(p)} U_i(x_i). \tag{6.8.10}$$

Оскільки розподіл $(\bar{x}_1, \ldots, \bar{x}_m)$ допустимий, то набір $(\bar{p}; \bar{x}_1, \ldots, \bar{x}_m)$ є станом напіврівноваги.

Нехай, навпаки $(\bar{p}; \bar{x}_1, \ldots, \bar{x}_m)$ є станом напіврівноваги. Тоді виконується (6.8.10) і тому

$$\phi_1^{\rho}(\bar{p}) = \sum_i \rho_i \max_{x_i \in Z_i(p)} U_i(\bar{x}_i).$$

В той же час, оскільки розподіл $(\bar{x}_1, \ldots, \bar{x}_m)$ допустимий, то

$$\phi_2^{\rho}(\bar{p}) \geqslant \sum_i \rho_i \max_{x_i \in Z_i(p)} U_i(\bar{x}_i).$$

Таким чином

$$\phi_2^{\rho}(\bar{p}) \geqslant \phi_1^{\rho}(\bar{p}).$$

Зворотна нерівність справедлива при всіх p. □

Лема 6.8.1 показує, що напіврівноважні ціни $\bar{p}$ є точками мінімуму функції ϕ^{ρ} при будь-яких $\rho \in (\rho_1, \ldots, \rho_m) \in \operatorname{int} \mathbb{R}_+^m$. Число $\phi^{\rho}(p)$ можна розглядати як деяку міру відхилення цін p від множини рівноважних цін. Якщо $\phi^{\rho} > 0$, то ціноутворюючий орган прагне змінити вектор цін p так, щоб зменшити значення функції ϕ^{ρ}. Напрямок спадання цієї функції найпростіше знайти, використовуючи її похідну за напрямками. Тому виникають питання про диференційовність за напрямками функції ϕ^{ρ}, про умови, при яких вона квазідиференційовна, про обчислення її квазідиференціала у тому випадку, коли ці умови виконані. Ці питання досліджуються нижче.

При дослідженні моделі обміну як правило припускають, що функції U_i увігнуті або квазіувігнуті, а також монотонні в тому чи іншому сенсі. Будемо вважати, що ці функції увігнуті і сильно монотонні: з покоординатної нерівності $x \geqslant y$ випливає нерівність $U_i(x) \geqslant U_i(y)$, причому, якщо $x^{(k)} > y^{(k)}$ принаймі при одному k, то $U_i(x) > U_i(y)$. Вважатимемо також, що U_i визначені на деякій відкритій множині, яка містить $\mathbb{R}_+^n$. Зафіксуємо числа $\rho_i > 0$ і позначимо функцію $\rho_i U_i(x)$ через $F_i(x)$. Для

$X = (x_1, \ldots, x_m) \in (\mathbb{R}^n)^m$ покладемо $F(X) = \sum\limits_i F_i(x_i)$. Для спрощення записів індекс ρ у функцій $\phi_1^\rho, \phi_2^\rho, \phi^\rho$ будемо опускати. Маємо

$$\phi_1(p) = \max_{X \in \bar{Z}(p)} F(x) = \sum_{i=1}^{m} \phi_{1i}(p), \tag{6.8.11}$$

де

$$\phi_{1i}(p) = \max_{x \in \bar{Z}_i(p)} F_i(x), \tag{6.8.12}$$

$$\phi_2(p) = \max_{X \in Z(p)} F(x), \quad \phi(p) = \phi_1(p) - \phi_2(p). \tag{6.8.13}$$

Оскільки $\phi_{1i}(p)$ і $\phi_2(p)$ – це функції максимуму при зв'язаних обмеженнях, то для обчислення їх похідних за напрямками можна застосувати результати відповідного розділу. Багатозначне відображення $Z_i(p)$ (відповідно $Z(p)$) допускає апроксимацію першого порядку в точках $p \in \operatorname{int}\mathbb{R}^n_+$, $y \in Z_i(p)$ (відповідно, $Y \in Z(p)$) відносно відображення $A_g(y') = \operatorname{cl}\gamma(x,y',g)$. Оскільки $\operatorname{cl}\gamma(x,y',g) = \Gamma(x,y',g)$, то $\operatorname{cl}\gamma(x,y',g) = \tilde{K}(x,y,g)$, де множина $\tilde{K}(x,y,g)$ визначена рівністю

$$\tilde{K}(x,y,g) = \{v | y + \alpha v + o(\alpha) \in a(x + \alpha g)\}. \tag{6.8.14}$$

Позначимо через $\tilde{K}_i$ (відповідно K_Z) відображення, що побудоване за формулою (6.8.14) для відображення $a = Z_i$, $i = 1, \ldots, m$ (відповідно, відображення $a = Z$). Для $p \in \operatorname{int}\mathbb{R}^n_+$ покладемо

$$R_i(p) = \{\bar{y}_i \in Z_i(p) | F_i(\bar{y}_i) = \phi_{1i}(p)\}, \quad i = 1, \ldots, m,$$

$$R_Z(p) = \{Y = (y_1, \ldots, y_m) \in Z(p) | F(Y) = \phi_2(p)\}.$$

Функції ϕ_{1i}, ϕ_2 диференційовні за напрямками в точці $p \in \operatorname{int}\mathbb{R}^n_+$ і

$$\phi'_{1i}(p,g) = \sup_{\bar{y}_i \in Z_i(p)} \sup_{v_i \in \tilde{K}_i(p,\tilde{y}_i,g)} F'_i(\tilde{y}_i, v_i), \tag{6.8.15}$$

$$\phi'_2(p,g) = \sup_{Y=(y_1,\ldots,y_m) \in R_Z(p)} \sup_{V=(v_1,\ldots,v_m) \in \tilde{K}_Z(p,y,g)} \sum_i F'_i(y_i, v_i). \tag{6.8.16}$$

Щоб спростити вираз для похідної $\phi'_{1i}(p,g)$, опишемо множину $\tilde{K}_i(p,\tilde{y}_i,g)$. З цією метою для вектора $y \in \mathbb{R}^n_+$ покладемо

$$I^1(y) = \left\{k \in I = \{1, \ldots, n\} \colon y^{(k)} > 0\right\}, I^2(y) = \left\{k \in I \colon y^{(k)} = 0\right\} \tag{6.8.17}$$

Нехай вектори $p \in \operatorname{int}\mathbb{R}^n_+$, $g \in \mathbb{R}^n$ фіксовані. Для $y \in \mathbb{R}^n_+$ покладемо

$$c_i(y) = \langle g, w_i - y \rangle, \tag{6.8.18}$$

$$\Delta_i(y) = \left\{v \in \mathbb{R}^n \colon \langle p,v \rangle \leqslant c_i(y), v^{(k)} \geqslant 0, \in I^2(y)\right\}. \tag{6.8.19}$$

Лема 6.8.2. *Нехай $\bar{y} \in R_i(p)$. Тоді $\tilde{K}_i(p,\bar{y},g) = \Delta_i(\bar{y})$.*

Доведення. 1) Нехай $v \in \tilde{K}_i\left(p,\bar{y},g\right)$, тобто $\bar{y}+\alpha v+o\left(\alpha\right) \in Z_i\left(p+\alpha g\right)$ при досить малих α. Звідси випливає, що при малих α

$$\langle p+\alpha g,\bar{y}+\alpha v+o\left(\alpha\right)\rangle \leqslant \langle p+\alpha g,w_i\rangle. \tag{6.8.20}$$

Строга монотонність функції U_i (і функції F_i) показує, що в точках $\bar{y}$ множини $R_i\left(p\right)$ виконується рівність $\left(p,\bar{y}\right)=\left(p,w_i\right)$. Тому, користуючись (6.8.20), маємо

$$\langle p,v\rangle+\langle g,\bar{y}\rangle+\alpha\langle g,v\rangle+\frac{o\left(\alpha\right)}{\alpha} \leq \langle g,w_i\rangle.$$

Переходячи до границі при $\alpha \downarrow 0$, переконаємося в справедливості нерівності $\langle p,v\rangle \leqslant c_i\left(\bar{y}\right)$. Оскільки $\bar{y}+\alpha v+o\left(\alpha\right) \in Z\left(p+\alpha g\right)$, то $\bar{y}+\alpha v+o\left(\alpha\right) \geqslant 0$. Із цього співвідношення випливає нерівність $v^{(k)} \geqslant 0$ при $k \in I^2\left(\bar{y}\right)$.

2) Нехай $v \in \Delta_i\left(\bar{y}\right)$. Враховуючи визначення множини $\Delta_i\left(\bar{y}\right)$, рівність $\langle p,\bar{y}\rangle=\langle p,w_i\rangle$ і формулу (6.8.18), маємо

$$\langle p+\alpha g,\bar{y}+\alpha v\rangle \leqslant \langle p+\alpha g,w_i\rangle+\alpha^2\langle g,v\rangle.$$

Виберемо елемент $\bar{v}$ так, щоб виконувалися нерівності $\langle p,\bar{v}\rangle+\langle g,v\rangle<0$ і $\bar{v}^{(k)} \geqslant 0$ при $k \in I^2\left(\bar{y}\right)$. Оскільки всі координати вектора p додатні і множина $I^1\left(\bar{y}\right)$ непорожня (це випливає з монотонності функції F_i і включення $\bar{y} \in R_i\left(p\right)$), то такий вектор v існує. При досить малих α маємо

$$\langle p+\alpha g,\bar{y}+\alpha v+\alpha^2\bar{v}\rangle \leqslant \langle p+\alpha g,w_i\rangle+$$
$$+\alpha^2\left(\langle g,v\rangle+\langle p,\bar{v}\rangle\right)+\alpha^3\langle g,\bar{v}\rangle \leqslant \langle p+\alpha g,w_i\rangle.$$

В той же час при досить малих α величина $y+\alpha v+\alpha^2\bar{v} \geqslant 0$. Отже $y+\alpha v+\alpha^2\bar{v} \in Z_i\left(p\right)$ і тому $v \in K_i\left(p,\bar{y},g\right)$.

Нехай вектори $p \in \operatorname{int}\mathbb{R}^n_+$ і $g \in \mathbb{R}^n_+$. Для векторів $l \in \mathbb{R}^n_+$, $y \in \mathbb{R}^n_+$ покладемо

$$s_i\left(y,l\right)=\frac{l^{(k)}}{p^{(k)}}c_i\left(y\right), \quad k \in I^1\left(y\right),$$

якщо

$$\frac{l^{(k)}}{p^{(k)}}c_i\left(y\right) \leqslant \frac{l^{(j)}}{p^{(j)}}c_i\left(y\right), \quad \forall k \in I^1\left(y\right), \quad \forall j=1,\ldots,n$$

(в тому випадку, коли числа $\frac{l^{(k)}}{p^{(k)}}c_i\left(y\right)$ співпадають при всіх $k \in I$), $s_i\left(y,l\right)=\infty$ в протилежному випадку. За допомогою величини $s_i\left(y,l\right)$ виразимо похідну $\phi'_{1i}\left(p,g\right)$ функції ϕ_{1i}. $\square$

Теорема 6.8.1. *Справедлива рівність*

$$\phi'_{1i}\left(p,g\right)=\sup_{\bar{y}_i \in R_i\left(p\right)}\inf_{l \in \bar{\partial}F_i\left(\bar{y}_i\right)} s_i\left(\bar{y}_i,l\right),$$

де $\bar{\partial}F_i\left(\bar{y}_i\right)$— субдиференціал функції F_i в точці $\bar{y}_i$.

Доведення. Використовуючи формулу (6.8.15), лему 6.8.2 і теорему про мінімакс, отримаємо

$$\phi'_{1i}(p,g) = \sup_{\bar{y}_i \in R_i(p)} \sup_{v_i \in \Delta_i(\bar{y}_i)} \inf_{l \in \partial F_i(\bar{y}_i)} \langle l,v_i \rangle = \sup_{\bar{y}_i \in R_i(p)} \inf_{l \in \partial F_i(\bar{y}_i)} \sup_{v_i \in \Delta_i(\bar{y}_i)} \langle l,v_i \rangle. \tag{6.8.21}$$

Величина

$$\sup_{v_i \in \Delta_i(\bar{y}_i)} \langle l,v_i \rangle$$

є розв'язком задачі лінійного програмування

$$\langle l,v_i \rangle \to \max, \tag{6.8.22}$$

$$\langle p,v_i \rangle \leqslant c_i(\bar{y}_i), \; v_i^{(k)} \geqslant c, \, k \in I^2(y_i). \tag{6.8.23}$$

Запишемо двоїсту задачу

$$c_i(\bar{y}_i)\, z \to \min \tag{6.8.24}$$

$$z \geqslant 0, \quad p^{(k)} z \geqslant l^{(k)}, \quad k \in I^2(y_i) \tag{6.8.25}$$

$$p^{(k)} z = l^{(k)}, \quad k \in I^1(y_i) \tag{6.8.26}$$

Нехай елемент l такий, що значення задачі (6.8.22)–(6.8.23) скінченне. Тоді двоїста задача сумісна і її значення також скінченне. Множина $I^1(\bar{y})$ непорожня. Це випливає зі строгої монотонності функції F_i і включення $\bar{y} \in R_i(p)$. Тому співвідношення (6.8.25) і (6.8.26) показують, що існує таке число $z_i \geqslant 0$, що для даного l виконується співвідношення

$$\frac{l^{(k)}}{p^{(k)}} = z_i \; \forall k \in I^1(\bar{y}_i), \tag{6.8.27}$$

$$\frac{l^{(k)}}{p^{(k)}} \geqslant z_i \;\; \forall k \in I^2(\bar{y}_i). \tag{6.8.28}$$

При цьому значення двоїстої задачі, а отже й значення прямої задачі збігається з числом $z_i c_i(\bar{y}_i)$. Навпаки, якщо число z_i, при якому виконано (6.8.27)–(6.8.28), існує, що значення задачі (6.8.24)—(6.8.26) скінченне й збігається з $z_i c_i(\bar{y}_i)$. Якщо ж такого числа не існує, то двоїста задача несумісна і значення задачі (6.8.24)—(6.8.26) дорівнює $+\infty$. Зі сказаного та рівності (6.8.21) випливає справедливість твердження. □

Зауваження 6.8.1. Оскільки похідна $\phi'_{1i}(p,g)$ скінченна, то для будь-якого $\bar{y}_i \in R_i(p)$ знайдеться елемент $l \in \partial F_i(\bar{y}_i)$, при якому $s_i(g,l) < +\infty$.

Зауваження 6.8.2. Як показано при доведенні твердження 6.8.1,

$$s_i(\bar{y}_i,l) = \sup_{v \in \Delta_i(\bar{y}_i)} \langle l,v \rangle.$$

Перейдемо до похідної $\phi'_2(p,g)$. Насамперед нам знадобиться наступне твердження.

Лема 6.8.3. *Нехай* $Y = (y_1, \ldots, y_m) \in R_z(p)$. *Тоді* $\sum y_i = w$ *і* $\langle p, y_i \rangle = \langle p, w_i \rangle, i = 1, \ldots, m$.

Доведення. Оскільки $Y \in Z(p)$, то $\sum_i y_i \leqslant w$ та $\langle p, y_i \rangle \leqslant \langle p, w_i \rangle$ $i = 1, \ldots, m$. Оскільки вектор p строго додатній, то з нерівності $\sum y_i \neq w$ випливає співвідношення

$$\sum_i \langle p, y_i \rangle = \langle p, \sum_i y_i \rangle < \langle p, w \rangle = \sum_i \langle p, w_i \rangle.$$

Тому знайдеться індекс j, при якому $\langle p, y_j \rangle < \langle p, w_j \rangle$. Із сказаного випливає, що існує такий вектор $\tilde{Y} = (\tilde{y}_1, \ldots, \tilde{y}_m) \in Z(p)$, що $\tilde{y}_i = y_i$ при $i \neq j$, $\tilde{y}_j > y_j$. Оскільки функція F_j строго монотонна, то $F_j(\tilde{y}_j) > F_j(y_j)$, а тому $F(\tilde{Y}) > F(Y)$. Це суперечить тому, що $Y \in R_z(p)$. Тим самим рівність $\sum_i y_i = w$ доведена. Із цієї рівності випливає, що $\langle p, y_i \rangle = \langle p, w_i \rangle$ $\forall i = 1, \ldots, m$.. □

Опишемо множину $\tilde{K}_z(p, Y, g)$. При фіксованих p, g для $Y = (y_1, \ldots, y_m) \in (\mathbb{R}^n_+)^m$ покладемо

$$\Delta_z(Y) = \left\{ V = (v_1, \ldots, v_m) \,|\, \sum_i v_i \leqslant 0, \langle p, v_i \rangle \leqslant c_i(y_i), v_i^{(k)} \geqslant 0, k \in I^2(y_i) \right\}.$$

Лема 6.8.4. *Нехай p і g фіксовані,* $Y \in R(p)$. *Тоді* $\tilde{K}_z(p, Y, g) = \Delta_z(Y)$.

Доведення. 1) Нехай $V = (v_1, \ldots, v_m) \in \tilde{K}_z(p, Y, g)$, тобто $Y + \alpha V \in Z(p + \alpha g)$. Це включення означає, що

$$\sum_i (y_i + \alpha v_i) = \sum_i y_i + \alpha \sum_i v_i \leqslant w, \tag{6.8.29}$$

$$\langle p + \alpha g, y_i + \alpha v_i \rangle \leqslant \langle p + \alpha g, w_i \rangle, \quad i = 1, \ldots, m, \tag{6.8.30}$$

$$y_i + \alpha v_i \geqslant 0, \quad i = 1, \ldots, m. \tag{6.8.31}$$

Оскільки в силу попередньої леми $\sum y_i = w$, то з (6.8.29) випливає, що $\sum v_i \leqslant 0$. Оскільки в силу тієї ж леми

$$\langle p, y_i \rangle = \langle p, w_i \rangle \;\; \forall i = 1, \ldots, m,$$

то співвідношення (6.8.30) приводить до нерівності

$$\langle g, y_i \rangle + \langle p, v_i \rangle + \alpha \langle g, v_i \rangle \leqslant \langle g, w_i \rangle, \quad i = 1, \ldots, m,$$

з якої випливає, що $\langle p, v_i \rangle \leqslant c_i(y)$ при $\alpha \downarrow 0$. Нерівності (6.8.31) справджуються при досить малих $\alpha > 0$ лише у тому випадку, коли $v_i^{(k)} \geqslant 0$ при $k \in I^2(y_i)$. Отже $V \in \Delta_z(Y)$.

2) Нехай $Y \in R_z(p)$, $V \in \Delta_z(Y)$. Покладемо $\bar{V} = (\bar{v}_1, \ldots, \bar{v}_m)$, де вектори $\bar{v}_i$ обрані так, що

$$\bar{v}_i \leqslant 0, \quad \langle g, v_i \rangle + \langle p, \bar{v}_i \rangle < 0, \quad \bar{v}_i^{(k)} = 0, \quad k \in I^2(y_i), \quad i = 1, \ldots, m.$$

Оскільки множина $I^1(y_i)$ непорожня і всі координати вектора p додатні, то такий вибір можливий. Використовуючи співвідношення

$$c_i(y_i) = \langle g, w_i - y_i\rangle,\ \langle p, y_i\rangle = \langle p, w_i\rangle,\ \langle p, v_i\rangle \geqslant c_i(y_i),$$

маємо

$$\langle p + \alpha g, y_i + \alpha v_i\rangle \leqslant \langle p + \alpha g, w_i\rangle + \alpha^2\langle g, v_i\rangle.$$

Тому при досить малих α

$$\langle p + \alpha g, y_i + \alpha v_i + \alpha^2 \bar{v}_i\rangle \leqslant$$

$$\leq \langle p + \alpha g, w_i\rangle + \alpha^2(\langle g, v_i\rangle + \langle p, \bar{v}_i\rangle) + \alpha^3\langle g, \bar{v}_i\rangle \leqslant \langle p + \alpha g, w_i\rangle.$$

В той же час

$$\sum_i (y_i + \alpha v_i + \alpha^2 \bar{v}_i) \leqslant \sum_i y_i \leqslant w$$

і при досить малих α

$$y_i + \alpha v_i + \alpha^2 \bar{v}_i \geqslant 0.$$

Ми показали, що

$$Y + \alpha V + \alpha^2 \bar{V} \in Z(p + \alpha g).$$

Звідси випливає включення $V \in \tilde{K}_z(p, Y, g)$. □

Введемо ряд позначень. Для $Y = (y_1, \ldots, y_m) \in R_z(p)$ і $L = (l_1, \ldots, l_m)$ покладемо

$$s_Z(Y, L) = \sup_{V \in \Delta_Z(Y)} \sum_i \langle l, v_i\rangle.$$

Елементу Y поставимо у відповідність множину

$$\partial F(Y) = \left\{L = (l_1, \ldots, l_m) : l_i \in \bar{\partial} F_i(y_i)\right\}.$$

Тут $\bar{\partial} F_i(y_i)$— супердиференціал функції F_i в точці y_i.

Теорема 6.8.2. *Справедлива рівність*

$$\phi_2'(p, g) = \sup_{Y \in R_Z(p)} \inf_{L \in \partial F(Y)} s_Z(Y, L).$$

Доведення. Використовуючи формулу (6.8.16), лему 6.8.4 і теорему про мінімакс, маємо

$$\phi_2'(p, g) = \sup_{Y \in R_Z(p)} \sup_{V \in \tilde{K}_Z(p, Y, g)} \sum_i \inf_{l_i \in \bar{\partial} F_i(y_i)} \langle l_i, v_i\rangle$$

$$= \sup_{Y \in R_Z(p)} \sup_{V \in \Delta_Z(Y)} \inf_{l_i \in \bar{\partial} F_i(y_i)} \sum_i \langle l_i, v_i\rangle = \sup_{Y \in R_Z(p)} \inf_{L \in \partial F(Y)} s_Z(Y, L).$$

□

Для $Y \in R_z(p)$ покладемо

$$J(Y) = \left\{ k \in \{1,\ldots,n\} \colon y_i^{(k)} > 0 \quad \forall i \in \{1,\ldots,m\} \right\}.$$

Іншими словами,

$$J(Y) = \bigcap_{i=1}^{m} I^1(y_i).$$

Якщо індекс k входить в $J(Y)$, то продукт з цим індексом в тій чи іншій кількості споживається кожним учасником. Наявність таких продуктів, тобто непорожність множини $J(Y)$ з економічної точки зору є досить природним. Припускаючи, що множина $J(Y)$ непорожня, обчислимо величину $s_Z(Y,L)$.

Теорема 6.8.3. *Нехай $Y \in R_Z(p)$ і множина $J(Y)$ непорожня. Тоді при кожному $L = (l_1,\ldots,l_m)$ виконується одне із двох: або $s_Z(Y,L) = +\infty$, або*

$$s_Z(Y,Z) = \frac{1}{p^{(k)}} \sum_i c_i(y_i)\, l_i^{(k)}, \; k \in J(Y).$$

Доведення. Припустимо, що $s_Z(Y,L) < +\infty$ і розглянемо число $s_Z(Y,L)$ як значення задачі лінійного програмування

$$\sum_i \langle l_i, v_i \rangle \to \max; \tag{6.8.32}$$

$$\sum_k p^{(k)} v_i^{(k)} \leqslant c_i(y_i), \quad i = 1,\ldots,m; \tag{6.8.33}$$

$$\sum_i v_i^{(k)} \leqslant 0, \quad k = 1,\ldots,n; \tag{6.8.34}$$

$$v_i^{(k)} \geqslant 0, \quad k \in I^2(y_i), \quad i = 1,\ldots,m. \tag{6.8.35}$$

Двоїста до (6.8.32)–(6.8.33) задача має вигляд

$$\sum_{i=1}^{m} c_i(y_i)\, z_i \to \min; \tag{6.8.36}$$

$$p^{(k)} z_i + t_k \geqslant l_i^{(k)}, \quad k \in I^2(y_i), \quad i = 1,\ldots,m; \tag{6.8.37}$$

$$p^{(k)} z_i + t_k = l_i^{(k)}, \quad k \in I^1(y_i), \quad i = 1,\ldots,m; \tag{6.8.38}$$

$$z_i \geqslant 0, \quad i = 1,\ldots,m; \quad t_k \geqslant 0, \quad k = 1,\ldots,n. \tag{6.8.39}$$

Оскільки значення задачі (6.8.32)–(6.8.33) скінченне, то задача (6.8.36)–(6.8.39) сумісна і має розв'язок. Нехай k— деякий елемент множини

$J(Y)$. Тоді, як випливає з (6.8.38), змінні z_i пов'язані із числом t_k системою рівнянь

$$p^{(k)} z_i + t_k = l_i^{(k)}, \quad i = 1, \ldots, m.$$

Виражаючи з цієї системи t_k, одержимо, що

$$z_i = z_1 + \frac{1}{p^{(k)}} \left(l_i^{(k)} - l_1^{(k)} \right), \quad i = 1, \ldots, m.$$

Звідси

$$\sum_{i=1}^{m} c_i (y_i) z_i = \sum_{i=1}^{m} c_i (y_i) \left(z_1 + \frac{1}{p^{(k)}} \left(l_i^{(k)} - l_1^{(k)} \right) \right).$$

Згадуючи визначення величини $c_i(y_i)$, маємо

$$\sum_{i=1}^{m} c_i (y_i) = \sum_{i=1}^{m} \langle g, w_i - y_i \rangle = \langle g, \sum_i w_i \rangle - \langle g, \sum_i y_i \rangle.$$

Оскільки $Y \in R_Z(p)$, то

$$\sum_i y_i = \sum_i w_i,$$

тобто лінійна функція

$$Y \to \sum_{i=1}^{m} c_i (y_i) z_i$$

на множині (6.8.37)–(6.8.39) стала. Застосовуючи теорему двоїстості лінійного програмування, приходимо до рівності

$$s_Z (Y, Z) = \sup_{V \in \Delta_Z(Y)} \sum_i \langle l_i, v_i \rangle = \frac{1}{p^{(k)}} \sum_i c_i (y_i) l_i^{(k)}.$$

$\square$

Результати попереднього пункту дозволяють описати похідну функції ϕ^ρ. Зафіксуємо вектори $p \in \operatorname{int} \mathbb{R}_+^n$, $g \in \mathbb{R}^n$ і введемо наступні позначення

$$\widetilde{\partial} U_i (\bar{y}_i) = \left\{ l \in \bar{\partial} U_i (\bar{y}_i) \,|\, \sup_{v \in \Delta_i(\bar{y}_i) \bar{y}_i \in R_i(p) i \in \{1, \ldots, m\}} \langle l, v \rangle < +\infty \right\},$$

$$\widetilde{\partial} U^\rho (Y) = \left\{ L = (l_1, \ldots, l_m) | l_i \in \bar{\partial} U_i (y_i) : \sup_{V \in \Delta_Z(Y)} \sum_i \langle l_i, v_i \rangle < +\infty \right\},$$

$$R_i (p) = \left\{ y \in Z_i (p) \,| U_i (y) = \max_{x \in Z_i(p)} U_i (x) \right\}, \tag{6.8.40}$$

$$R_{Z,\rho} (p) = \left\{ Y \in Z (p) \,| U^\rho (y) = \max_{X \in Z(p)} U^\rho (X) \right\}. \tag{6.8.41}$$

Функція U^ρ визначена формулою (6.8.7).

Теорема 6.8.4. *Нехай задані вектори* $p \in \operatorname{int}\mathbb{R}^n_+$, $g \in \mathbb{R}^n$, *функції* U_i *і числа* ρ_i, *причому:*

1) функції U_i *увігнуті, строго монотонні й визначені на деякій відкритій множині, що містить конус* $\mathbb{R}^n_+$*; числа* ρ_i *додатні;*

2) точка $p \in \operatorname{int}\mathbb{R}^n_+$ *така, що для кожного* $Y \in R_{Z,\rho}$ *множина індексів* $J(Y)$ *непорожня.*

Тоді функція ϕ^ρ, *що визначена формулою (6.8.9), диференційовна в точці* p *за напрямком* g *і*

$$(\phi^\rho)'(p,g) = (\phi_1^\rho)'(p,g) - (\phi_2^\rho)'(p,g),$$

де

$$(\phi_1^\rho)'(p,g) = \sum_{i=1}^{m} \sup_{\bar{y}_i \in R_i(p)} \inf_{l_i \in \overline{\partial} U_i(\bar{y}_i)} \frac{p_i l_i^{(k_i)}}{p^{(k_i)}} \langle g, w_i - \bar{y}_i \rangle, \ \ k \in I^1(\bar{y}_i),$$

$$(\phi_2^\rho)'(p,g) = \sup_{Y \in R_{Z,\rho}(p)} \inf_{L \in \overline{\partial} U^\rho(Y)} \sum_{i=1}^{m} \frac{p_i l_i^{(k)}}{p^{(k)}} \langle g, w_i - \bar{y}_i \rangle, \ \ k \in J(Y).$$

Вкажемо умови, при яких функція ϕ^ρ квазідиференційовна.

Наслідок 6.8.1. *Нехай в умовах теореми функції* U_i *диференційовні, тобто супердиференціал* $\overline{\partial} U_i(y)$ *складається з одного вектора*

$$\nabla U_i(y) = \left(\frac{\partial U_i}{\partial x^{(1)}}(y), \dots, \frac{\partial U_i}{\partial x^{(n)}}(y) \right).$$

Тоді

$$(\phi^\rho)'(p,g) = \sum_{i=1}^{m} \sup_{\bar{y}_i \in R_i(p)} \left\langle g, \frac{\rho_i}{p^{(k_i)}} \frac{\partial U_i(\bar{y}_i)}{\partial x^{(k_i)}} (w_i - \bar{y}_i) \right\rangle$$

$$- \sup_{Y \in R_{Z,\rho}(p)} \left\langle g, \sum_{i=1}^{m} \frac{\rho_i}{p^{(k)}} \frac{\partial U_i(y_i)}{\partial x^{(k)}} (w_i - y_i) \right\rangle, \tag{6.8.42}$$

де k_i *— довільний індекс із множини* $I^1(\bar{y}_i)$, k*— довільний індекс із множини* $J(Y)$. *У даному випадку функція* ϕ^ρ *квазідиференційовна в точці* p. *З (6.8.42) легко випливає, що квазідиференциал функції* ϕ^ρ *має вигляд*

$$\mathfrak{D}\phi^\rho = \left[\underline{\partial}\phi^\rho(p), \overline{\partial}\phi^\rho(p)\right],$$

де $\underline{\partial}\phi^\rho(p)$*— опукла оболонка множини*

$$\left\{ \sum_{i=1}^{m} \frac{\rho_i}{p^{(k_i)}} \frac{\partial U_i(\bar{y}_i)}{\partial x^{(k_i)}} (w_i - \bar{y}_i) \,|\, \bar{y}_i \in R_i(p), k_i \in I^1(\bar{y}_i) \right\},$$

$\overline{\partial}\phi^\rho(p)$*— опукла оболонка множини*

$$\left\{ -\sum_{i=1}^{m} \frac{\rho_i}{p^{(k)}} \frac{\partial U_i(y_i)}{\partial x^{(k)}} (w_i - y_i) \,|\, Y \in R_{Z,\rho}(p), k \in J(Y) \right\}.$$

Наслідок 6.8.2. *Нехай в умовах теореми 6.8.4 функції U_i строго угнуті. Тоді множина $R_i\,(p)$ складається з одного елемента $\bar{y}_i\,(p)$, а множина $R_{Z,\rho}\,(p)$ — з одного елемента $Y\,(p,\rho) = (y_1(p,\rho),\ldots,y_m(p,\rho))$. В цьому випадку*

$$(\phi^{\rho})'\,(p,g) = \sum_{i=1}^{m} \inf_{l_i \in \widetilde{\partial} U_i(\bar{y}_i(p))} \left\langle g, \frac{\rho_i l_i^{(k)}}{p^{(k_i)}}\,(w_i - \bar{y}_i(p)) \right\rangle -$$

$$- \inf_{L \in \widetilde{\partial} U^{\rho}(Y(p,\rho))} \sum_{i=1}^{m} \left\langle g, \frac{\rho l_i^{(k)}}{p^{(k_i)}}\,(w_i - y_i\,(p,\rho)) \right\rangle. \qquad (6.8.43)$$

В даному випадку функція ϕ^{ρ} також квазідиференційовна. Використовуючи формулу (6.8.43), легко виписати її квазідиференціал.

Наслідок 6.8.3. *Нехай в умовах теореми функції U_i строго угнуті й диференційовні. Тоді функція ϕ^{ρ} диференційовна в точці p і її градієнт $\nabla\phi^{\rho}$ має вигляд*

$$\nabla\phi^{gr}\,(p) = \sum_{i=1}^{m} \frac{\rho_i}{p^{(k_i)}} \frac{\partial U_i\,(\bar{y}_i\,(p))}{\partial x^{(k_i)}}\,(w_i - \bar{y}_i\,(p)) -$$

$$- \sum_{i=1}^{m} \frac{\rho_i}{p^{(k)}} \frac{\partial U_i\,(y_i\,(\rho,p))}{\partial x^{(k)}}\,(w_i - y_i\,(\rho,p)).$$

Тут вектори $\bar{y}_i\,(p)$, $Y\,(\rho,p)$, індекси k_i, k визначаються так само, як і в наслідку 6.8.2.

6.9. АНАЛІЗ КООПЕРАТИВНИХ ІГОР

У цьому параграфі ми покажемо як субдиференціал Кларка та квазідиференциал дозволяють розглянути з єдиної точки зору деякі проблеми теорії кооперативних ігор.

Докладний виклад результатів теорії ігор, які будуть використовуватися, можна знайти, наприклад, в [2,46].

Отже, нехай декілька учасників деякого процесу (це може бути процес виробництва, обміну, або просто гра) можуть діяти як поодинці, так і створювати коаліції, одержуючи при цьому якісь виграші. Будемо вважати, що кожен учасник може оцінити свій виграш, тобто в кожного учасника є функція корисності. Припускаємо також, що корисності учасників (гравців) виміряються по одній шкалі (наприклад, мова може йти просто про гроші) і можуть передаватися один одному (побічні платежі) без втрат.

Кооперативна гра (з побічними платежами), або класична кооперативна гра — це пара $[I{,}v]$, що складається із скінченної множини I, яка називається множиною гравців, і дійсної функції $v : 2^I \to \mathbb{R}$, що ставить у відповідність кожній підмножині $J \subset I$, яка називається коаліцією, число $v(J)$, що інтерпретується як максимальний виграш (максимальна корисність), який можуть забезпечити собі гравці з коаліції J, діючи спільно. Одне з центральних питань теорії кооперативних ігор — питання про визначення розв'язку гри — коротко можна сформулювати так. Як розподілити між гравцями виграш (корисність) $v(I)$ так, щоб вважати цей розподіл у деякому розумінні справедливим, розумним, оптимальним?

Серед різних підходів до розв'язку цієї задачі ми виділимо два, які, з одного боку, належать до числа найбільш істотних, привабливих і відомих, а з іншого боку, дозволять продемонструвати можливість використання субдиференціала Кларка й квазідиференціала.

Перший підхід полягає в тому, що кожній грі $[I{,}v]$ ставиться у відповідність деякий вектор $\Phi(v) \in \mathbb{R}^I$, компоненти якого описують виграші гравців у цій грі. Природно, що для того щоб такий вектор вважався в якомусь сенсі справедливим, він повинен задовольняти деяким вимогам. Вперше систему аксіом, яким повинна задовольняти функція $\Phi(v)$, сформулював Л. Шеплі [57]. Сформулюємо цю систему в зручному для нас вигляді.

А1 (симетричність). Якщо π — така перестановка множини I, що для будь-якої коаліції J справджується рівність $v\left(\pi J\right) = v\left(J\right)$, то для будь-якого $i \in I$ справджується рівність $\Phi_{\pi i}\left(v\right) = \Phi_i\left(v\right)$;

А2 (Парето–оптимальністъ). $\sum\limits_{i \in I} \Phi_i\left(v\right) = v\left(I\right)$;

А3 (несуттєвого гравця). Якщо у грі v гравець i_0 такий, що для будь-якої коаліції J справджується рівність $v\left(J \cup \{i_0\}\right) = v\left(J\right)$, то $\Phi_{i_0}\left(v\right) = 0$;

А4 (лінійність). Якщо для всіх J справджується $w\left(J\right) = u\left(J\right) + v\left(J\right)$, то $\Phi\left(w\right) = \Phi\left(u\right) + \Phi\left(v\right)$.

Л. Шеплі розглядав $\Phi_i\left(v\right)$ як апріорну оцінку гравцем i вигідності для нього гри v. Зміст цих аксіом такий. Перша аксіома стверджує, що оцінка гравцем гри не повинна залежати від того, яким індексом він позначений. Друга аксіома стверджує, що гравці повинні розподілити між собою всю корисність $v(I)$. Третя аксіома стверджує, що гравець, який ніяк не впливає на виграш ні однієї з коаліцій, до якої він приєднується, не повинен щось одержувати.

Даним аксіомам задовольняє єдина функція Φ, названа функцією (значенням) Шеплі, а саме:

$$\Phi_i\left(v\right) = \sum_{i \notin J} \frac{|J|!\left(|I| - |J| - 1\right)!}{|I|!}\left(v\left(J \cup \{i\}\right) - v\left(J\right)\right),$$

де $|J|$-— число елементів множини J. Функцію Φ можна інтерпретувати як функцію, що задає середній виграш гравців при наступній ймовірнісній схемі. Рівноймовірно вибираємо кожного із гравців. Далі рівноймовірно приєднуємо до нього кожного із гравців, що залишилися, і продовжуємо цей процес, поки не вичерпаємо всю множину I. При цьому, якщо гравець i вибирається після того, що утворено коаліцію J, то він одержує величину $v\left(J \cup \{i\}\right) - v\left(J\right)$. Вектор $\Phi\left(v\right)$ називається вектором (значенням) Шеплі гри v.

Зауваження 6.9.1. Вектор Шеплі є розв'язком задачі мінімізації функції

$$Q\left(x,v\right) = \sum_J \left(|J| - 1\right)! \left(|I| - |J| - 1\right)! \left(v\left(J\right) - x\left(J\right)\right)^2$$

на множині

$$\left\{x \in \mathbb{R}^I | x\left(I\right) = v\left(I\right)\right\},$$

де $x\left(J\right) = \sum\limits_{i \in J} x_i$.

Цікаве сімейство нелінійних аналогів функції Шеплі, зокрема багатозначних, можна одержати, замінивши в $Q(x,v)$ опуклу функцію $(v(J) - (x(J))^2$ довільною опуклою функцією $\psi\left(v\left(J\right) - x\left(J\right)\right)$.

Другий підхід до визначення розв'язку гри полягає в тому, що розв'язком гри v називається множина "недомінуючих поділів", тобто множина

$$c\left(v\right) = \left\{x \in \mathbb{R}^I | x\left(I\right) = v\left(I\right), x\left(J\right) \geqslant v\left(J\right) \quad \forall J \subset I\right\},$$

яка називається c-ядром гри v. На відміну від вектора Шеплі, який завжди існує і єдиний, c-ядро може бути порожнім, однак допускає зрозумілу інтерпретацію. А саме, жодна коаліція не може одержати сумарну корисність більше, ніж їй дає довільний елемент c-ядра.

Надзвичайно цікавим як для застосувань (наприклад, при вивченні економічної рівноваги), так і з чисто технічних міркувань (вдається досить просто одержати деякі класичні результати), виявилося вивчення "нечітких кооперативних ігор", суть яких полягає в наступному.

Якщо I—– множина гравців, то кожна коаліція J може бути ототожнена з її характеристичним вектором e^J, що ставить у відповідність кожному гравцю $i \in I$ "інтенсивність його участі" $e_i^J \in \{0,1\}$ у цій коаліції;

$$e_i^J = \begin{cases} 1, i \in J, \\ 0, i \notin J. \end{cases}$$

У цьому випадку множина всіх коаліцій є $\{0,1\}^n$. Множина нечітких коаліцій, за визначенням, є $[0,1]^n$. Іншими словами, нечітка коаліція $\tau \in [0,1]^n$ ставить у відповідність кожному гравцю $i \in I$ інтенсивність його участі $\tau_i \in [0,1]$. Нечітка кооперативна гра — це додатньо однорідна функція на $[0,1]^n$. Таким чином, якщо класична кооперативна гра являє собою функцію, визначену в вершинах одиничного куба, то нечітка гра — це функція на всьому одиничному кубі. Ясно, що в силу додатньої однорідності функцію $v : [0,1]^n \to \mathbb{R}^1$ можна продовжити на $\mathbb{R}^n_+$, поклавши

$$v(0) = 0,\ v(\tau) = \left(\sum_{i=1}^{n} \tau_i\right) v\left(\frac{\tau}{\sum\limits_{i=1}^{n} \tau_i}\right).$$

Будемо розглядати простір $\mathbb{R}^n$ як простір корисностей. Корисність вектора $x \in \mathbb{R}^n$ для коаліції τ визначається величиною

$$(\tau,x) = \sum_{i-1}^{n} \tau_i x_i.$$

Якщо $J \subset I$, то ця корисність дорівнює $\left(e^J,x\right) = \sum\limits_{i\in J} x_i$.

Визначимо розв'язок на деяких класах нечітких ігор. Під розв'язком розуміється відображення S, яке ставить у відповідність кожній грі v деяку множину $S(v)$ і задовольняє ряд властивостей. Нехай v — нечітка гра, $S(v)$ — множина векторів.

1. Говорять, що множина $S(v)$ Парето–оптималъна, якщо

$$\sum_{i=1}^{n} x_i = v(\mathbf{1}) \quad \forall x \in S(v).$$

2. Для довільної перестановки π множини $I = \{1,\ldots,n\}$ покладемо

$$(\pi * v)(\tau) = v\left(\tau_{\pi^{-1}(1)},\ldots,\tau_{\pi^{-1}(n)}\right).$$

Множина $S(v)$ називається симметричною, якщо для будь-якої перестановки π виконується $S(\pi * v) = \pi(S(v))$.

3. Нехай $P = (S_1,\ldots,S_m)$— деяке розбиття множини I на m непорожніх коаліцій. Позначимо через $P \otimes v$ гру m осіб, визначену формулою

$$(P \otimes v)(\theta_1,\ldots,\theta_m) = v(\tau_1,\ldots,\tau_n),$$

де $\tau_k = \theta_j$ при $k \in S_j$. Нехай, далі, $P \otimes x \in \mathbb{R}^m$, визначена рівністю

$$(P \otimes x)_j = \sum_{i\in S_j} x_i.$$

Говорять, що для гри v виконується властивість розбиття, якщо $S(P \otimes v) = P \otimes S(v)$ при всіх розбиттях P.

4. Говорять, що виконується властивість несуттєвого гравця, якщо з рівності

$$v(\tau) = v\left(\tau^{I\setminus\{i\}}\right) \quad \forall \tau \in [0,1]^n$$

випливає співвідношення $x_i = 0 \ \ \forall x \in S\,(v)$.

Під розв'язком на деякому класі ігор природно розуміти відображення S, що має на цьому класі зазначені вище властивості (або їх модифікації). Деякі із цих властивостей можуть виконуватися не на всьому класі, а лише на якій-небудь його частині.

Розглянемо насамперед клас неперервно диференційовних нечітких ігор. Узагальненим значенням Шеплі функції v із цього класу називається вектор $S(v) = \nabla_v\,(\mathbf{1})$. Можна перевірити, що цей вектор (точніше кажучи, множина, що складається з нього) має властивості Парето-оптимальності, симетричності, розбиття та несуттєвого гравця. Крім того, виконується властивість адитивності $S(v_1 + v_2) = S\,(v_1) + S\,(v_2)$.

Доведення цього твердження наведено нижче в більш загальній ситуації.

Виявляється, що якщо відображення S ставить у відповідність неперервно диференційовній грі v вектор $S(v)$, який має властивості Парето–оптимальності, симетричності й розбиття, то $S(v) = \nabla_v\,(\mathbf{1})$. Отже, узагальнений розв'язок Шеплі претендує на роль розв'язку в класі неперервно диференційовних ігор.

Нехай $u : \{0{,}1\}^n \to \mathbb{R}^1$— класична кооперативна гра. Можна показати, що існують її розширення до нечіткої неперервно диференційовної гри, узагальнене значення Шеплі якої збігається зі значенням Шеплі вихідної гри. Одне з таких розширень wu виглядає так:

$$wu\,(\tau) = \sum_S \alpha_S\,(u) \left(\prod_{i \in S} \tau_i\right)^{\frac{1}{|S|}},$$

де

$$\alpha_S\,(u) = \sum_{T \subset S} (-1)^{|S|-|T|} v\,(T).$$

Узагальнимо на нечіткі ігри поняття c–ядра. Нехай v— нечітка кооперативна гра. c–ядром такої гри називається множина

$$c\,(v) = \left\{ x \in \mathbb{R}^I |\sum_{i \in I} x_i = v\,(\mathbf{1})\,, \sum_{i \in I} \tau_i x_i \geqslant v\,(\tau)\,, \tau \in [0{,}1]^n \right\}.$$

Очевидно, що $x \in c\,(v)$ тоді й тільки тоді, коли $x \in \bar{\partial} v\,(\mathbf{1})$, де $\bar{\partial} v\,(\mathbf{1})$ супердиференціал функції v в точці $\mathbf{1}$. c-ядро може бути й порожнім, однак, якщо v– увігнута (і, тим самим, суперлінійна) функція, то c–ядро непорожнє. Можна показати (це випливає з наведених нижче результатів), що для угнутих ігор c-ядро Парето–оптимальне, симетричне, має властивості розбиття й несуттєвого гравця. Таким чином, у класі угнутих ігор на роль розв'язку претендує c-ядро.

Особливий інтерес представляє собою так звана супераддитивна оболонка πu класичної гри, яка визначається формулою

$$\pi u\left(\tau\right)=\sup_{\mu_S\geq 0,\,\sum\mu_S e^S=\tau}\sum_S\mu_S u\left(S\right).$$

Функція $\pi u:[0,1]^n\to\mathbb{R}^1$-— найменша додатньо однорідна угнута функція, яка більша за u. Гра u називається збалансованою, якщо $\pi u\left(\mathbf{1}\right)=u\left(I\right)$. Оскільки $c(v)=\bar{\partial}v\left(\mathbf{1}\right)$, то справедлива наступна теорема.

Теорема 6.9.1. *Класична коооперативна гра u має непорожнє c-ядро тоді і тільки тоді, коли вона збалансована. У цьому випадку її c-ядро збігається з c-ядром нечіткої гри πu.*

Перша частина цієї теореми є один із класичних результатів про непорожність c-ядра.

Можна розглядати зазначені вище розв'язки (узагальнене значення Шеплі для неперервно диференційовних ігор та c–ядро для угнутих ігор) в рамках єдиного підходу. Один з таких підходів базується на використанні субдиференціала Кларка. Розглянемо клас всіх локально ліпшицевих нечітких ігор. Розв'язком $S(v)$ гри v із цього класу називається субдиференціал Кларка $\partial_{c1}v\left(\mathbf{1}\right)$ функції v у точці $\mathbf{1}$. Зрозуміло, що для неперервно диференційовної гри таким чином визначений розв'язок збігається з узагальненим значенням Шеплі, а для угнутої гри він збігається з c-ядром.

Теорема 6.9.2. *Множина розв'язків $S(v)$ локально ліпшицевої гри v непорожня, компактна і опукла. Вона має властивості Парето–оптимальності, симетричності та несуттєвого гравця. Крім того,*

a) $S\left(\lambda v\right)=\lambda S\left(v\right),\quad\forall\lambda\geqslant 0,$

b) $S(u+v)\subset S\left(u\right)+S(v),$

c) якщо v монотонна функція, то $S\left(v\right)\subset\mathbb{R}^n_+,$

d) якщо $H:\mathbb{R}^m\to\mathbb{R}^n$ лінійне відображення і $H\left(\mathbf{1}\right)=\mathbf{1}_n$, то $S\left(v\circ H\right)\subset H^*S\left(v\right).$

Якщо, крім того, $H\left(\mathbb{R}^m\right)=\mathbb{R}^n$, або функція v регулярна в точці $x=\mathbf{1}$, то $S\left(v\circ H\right)=H^*S\left(v\right).$

e) якщо функція v регулярна, то S задовольняє умові розбиття.

Доведення. Властивість a) очевидна, b) випливає з теореми 5.2.14, d) випливає з теореми 5.2.15 і зауваження 5.2.1.
Доведемо властивість c). Нехай v зростаюча функція, $g\geq 0$. Тоді

$$v\left(x'+\alpha g\right)-v\left(x'\right)\geqslant 0\ \ \forall\alpha\geqslant 0,$$

тому

$$v^{\uparrow}_{Cl}\left(\mathbf{1},g\right)=\lim_{\alpha\downarrow 0,x'\to 1}\frac{1}{\alpha}\left[v\left(x'+\alpha g\right)-v\left(x'\right)\right]\geqslant 0.$$

Нагадаємо, що $S(v) = \partial_{Cl} v\,(\mathbf{1})$ є супердиференціалом суперлінійної функції $v^{\downarrow}_{Cl}\,(\mathbf{1},g)$. Тому, якщо $l \in S\,(v)\,,\ g \geq 0$, то

$$\langle l,g\rangle \geqslant v^{\downarrow}_{Cl}\,(\mathbf{1},g) \geqslant 0.$$

Звідси випливає включення $l \in \mathbb{R}^n_+$.

Властивість e) випливає з d). Властивості симетричності й несуттєвого гравця можуть бути також доведені за допомогою властивості d). Нам залишається лише перевірити Парето–оптимальність. Використовуючи додатню однорідність функції v, маємо

$$v\,(\mathbf{1}) = \frac{v\,(\mathbf{1}+\alpha\mathbf{1}) - v\,(\mathbf{1})}{\alpha} \leqslant \liminf_{\substack{\tau\to 1\\ \alpha\downarrow 0}} \frac{1}{\alpha}\left[v\,(\tau+\alpha\mathbf{1}) - v\,(\tau)\right] = v^{\uparrow}_{Cl}\,(\mathbf{1},\mathbf{1})\,.$$

Нехай $L-$ ліпшицева константа локально ліпшицевої функції в околі точки $x = \mathbf{1}$. Тоді

$$v\,(\tau+\alpha\mathbf{1}) - v\,(\mathbf{1}+\alpha\mathbf{1}) \leqslant L\,\|\tau-\mathbf{1}\|\,,$$

$$v\,(\mathbf{1}) - v\,(\tau) \leqslant L\,\|\tau-\mathbf{1}\|\,.$$

Тому

$$v\,(\tau+\alpha\mathbf{1}) - v\,(\tau) \leqslant v\,(\mathbf{1}+\alpha\mathbf{1}) - v\,(\mathbf{1}) + \frac{2}{\tau-\mathbf{1}},$$

звідки випливають співвідношення

$$v^{\uparrow}_{Cl}\,(\mathbf{1},\mathbf{1}) = \liminf_{\substack{\tau\to \mathbf{1}\\ \alpha\downarrow 0}} \frac{1}{\alpha}\left[v\,(\tau+\alpha\mathbf{1}) - v\,(\tau)\right] \leqslant \frac{v\,(\mathbf{1}+\alpha\mathbf{1}) - v\,(\mathbf{1})}{\alpha} = v\,(\mathbf{1})\,.$$

Отже

$$v^{\uparrow}_{Cl}\,(\mathbf{1},\mathbf{1}) = v\,(\mathbf{1})$$

і, якщо $x \in S\,(v)\,,$ то

$$\langle \mathbf{1},x\rangle \leqslant v^{\uparrow}_{Cl}\,(\mathbf{1},\mathbf{1}) = v\,(\mathbf{1})\,.$$

Оскільки функція $(-v)$ також додатньо однорідна, то $\langle \mathbf{1},-x\rangle \leqslant -v\,(\mathbf{1})$ для $x \in S\,(v)\,.$ Таким чином $\langle \mathbf{1},x\rangle = v\,(\mathbf{1})\,.$ □

Інший підхід до розв'язання гри, що дозволяє із загальної точки зору розглядати як неперервно диференційовні, так і увігнуті ігри, заснований на понятті квазідиференціала. Під квазідиференційовною нечіткою грою будемо розуміти додатньо однорідну функцію на конусі $\mathbb{R}^n_+$ яка квазідиференційовна в точці $\mathbf{1}$. Нам знадобляться деякі додаткові результати, що стосуються додатньо однорідних квазідиференційовних функций.

Нехай K опуклий конус в $\mathbb{R}^n$ з компактною основою T (тобто $K = \cup_{\lambda\geq 0}\lambda T$) та непорожньою внутрішністю $\mathrm{ri}\,T$. Функція $f\colon T\to\mathbb{R}^1$ називається квазідиференційовною в точці $x \in \mathrm{ri}\,T$, якщо вона диференційовна в цій точці за кожним напрямком $g \in \tilde{R}_T = R_T - x$ та існують такі опуклі компакти

$$\underline{\partial}_T f(x) \subset \tilde{R}_T, \quad \overline{\partial}_T f(x) \subset \tilde{R}_T,$$

що

$$f'(x,g) = \max_{y \in \underline{\partial}_T f(x)} \langle y,g \rangle + \min_{z \in \overline{\partial}_T f(x)} \langle z,g \rangle \quad \forall g \in \tilde{R}_T.$$

Теорема 6.9.3. *Нехай функція $f : K \to \mathbb{R}^1$ квазідиференційрвна в точці $x \in \operatorname{int} K$. Тоді функція $f|_{T(x,p)}$, де*

$$T(x,p) = \{z \in K | \langle z - x,p \rangle = 0\},$$

квазідиференційовна в точці x і її квазідиференціал визначається парою $[A,B]$, де

$$A = \operatorname{Pr}_p \left(\underline{\partial}_T f(x)\right), \ \ B = \operatorname{Pr}_p \left(\overline{\partial}_T f(x)\right),$$

а $\operatorname{Pr}_p C$— ортогональна проекція множини C на гіперплощину

$$H_p = \{z \in \mathbb{R}^n | \langle z,p \rangle = 0\}.$$

Теорема 6.9.4. *Нехай функція $\bar{f}$ є додатньо однорідним продовженням функції f, визначеній на множині $T(x,x)$, $x \in \operatorname{int} K$. Нехай f— квадизідиференційовна в x. Функція $\bar{f}$ квазідиференційовна і*

$$\mathfrak{D}\bar{f}(x) = \left[\underline{\partial}_T p(x) + x\frac{f(x)}{\|x\|^2}, \overline{\partial} f(x)\right].$$

Особливу роль при вивченні нечітких ігор грає коаліція всіх гравців, тобто точка $\mathbf{1}$. Тому надалі, говорячи про квазідиференційовність, ми будемо мати на увазі квазідиференційовнічть в точці $\mathbf{1}$. Розглянемо квазідиференційовну гру v. Нехай $\left[\underline{\partial} v(\mathbf{1}), \overline{\partial} v(\mathbf{1})\right]$ її квазідиференціал в точці $\mathbf{1}$. Із твердження 6.9.3 випливає, що функція $v_1 = v|_{T(\mathbf{1},\mathbf{1})}$ також квазідиференційовна і її квазідиференціал визначається парою

$$\left[\operatorname{Pr}_1 \underline{\partial} v(\mathbf{1}), \operatorname{Pr}_1 \overline{\partial} v(\mathbf{1})\right].$$

Ясно, що додатньо однорідне продовження $\bar{v}$ функції v_1 збігається з v, а її квазідиференціал можна визначити також за теоремою 6.9.4:

$$\left[\operatorname{Pr}_1 \underline{\partial} v(\mathbf{1}) + \frac{v(\mathbf{1})}{\|\mathbf{1}\|^2}\mathbf{1}, \ \operatorname{Pr}_1 \overline{\partial} v(\mathbf{1})\right].$$

Ця пара задовольняє властивості Парето–оптимальності в тому сенсі, що для

$$x \in \operatorname{Pr}_1 \underline{\partial} v(\mathbf{1}) + \frac{v(\mathbf{1})}{\|\mathbf{1}\|^2}\mathbf{1}, \quad y \in \operatorname{Pr}_1 \overline{\partial} v(\mathbf{1})$$

має місце рівність

$$\sum_{i=1}^n \langle x_i + y_i \rangle = \langle x + y, \mathbf{1} \rangle = v(\mathbf{1}).$$

Аналогічно можна сформулювати властивість несуттєвого гравця.

Квазірозв'язком гри v назвемо квазідиференціал $\mathfrak{D}^{\pi} v\left(\mathbf{1}\right)$ функції v в точці 1, який має властивість Парето–оптимальності.

Теорема 6.9.5. *Нехай гра v квазідиференційовна. Тоді квазірозв'язок має властивості Парето–оптимальності, симетричності й несуттєвого гравця. Крім того,*

a) квазірозв'язок лінійний за v,

b) якщо $H : \mathbb{R}^m \to \mathbb{R}^n$ лінійне відображення і $H\left(\mathbf{1}\right) = \mathbf{1}$, то

$$\mathfrak{D}^{\pi}\left(v \circ H\right)\left(\mathbf{1}\right) = H^{*}\mathfrak{D}^{\pi}\left(v\left(\mathbf{1}\right)\right),$$

c) якщо v неперервно диференційовна в $\mathbf{1}$, то $\mathfrak{D}^{\pi} v\left(\mathbf{1}\right) = \left[\nabla_{v}\left(\mathbf{1}\right),0\right]$, і квазірозв'язок можна ототожнити з узагальненим значенням Шеплі гри v.

d) якщо v увігнута, то

$$\mathfrak{D}^{\pi} v\left(\mathbf{1}\right) = [0,\bar{\partial} v\left(\mathbf{1}\right)],$$

де $\bar{\partial} v\left(\mathbf{1}\right)$– супердиференціал увігнутої функції. Отже його можна ототожнити з с-ядром гри v.

Визначимо розв'язок квазідиференційовної гри, який визначається однозначно (на відміну від квазірозв'язку, визначеного з точністю до класу еквівалентності).

Точкою Штойнера опуклого компакту K в $\mathbb{R}^n$ називається елемент

$$s\left(K\right) = \frac{1}{\sigma_n}\int\limits_{S^{n-1}} \alpha\rho\left(K,\alpha\right) d\lambda,$$

де λ – міра Лебега на одиничній сфері S^{n-1}, σ_n — об'єм одиничної кулі, α — змінний вектор і $p\left(K,\alpha\right)$– опорна функція множини K. Відзначимо, що відображення s лінійне за K і $s\left(K\right) \in K$.

Нехай гра v квазідиференційовна і $\left[\underline{\partial} v(\mathbf{1}),\overline{\partial} v(\mathbf{1})\right]$ – її квазірозв'язок. st–розв'язком квазідиференційовної гри v називається вектор $st\left(v\right) = s(\underline{\partial} v\left(\mathbf{1}\right)) + s(\overline{\partial} v\left(\mathbf{1}\right))$. Можна перевірити, що це визначення коректне, тобто не залежить від того, яка пара визначає квазідиференціал $\mathfrak{D}^{\pi} v\left(\mathbf{1}\right)$.

Теорема 6.9.6. *Якщо гра v квазідиференційовна, то відображення $st : v \to st\left(v\right)$ лінійне за v, st–розв'язок Парето–оптимальний. Якщо v неперервно диференційовна, то $st\left(v\right) = \nabla_{v}\left(1\right)$ і st-розв'язок збігається з узагальненим значенням Шеплі гри v. Якщо v увігнута, то $st(v)$ являєтся точкою Штойнера с-ядра гри v. st–розв'язок симетричний. st–розв'язок має властивість несуттєвого гравця.*

Доведення див. в [19].

Відзначимо, що вектор $st\left(v\right)$ можна інтерпретувати як вектор середніх виграшів гравців.

Розділ VII
Кодиференційовні функції

7.1. ОЗНАЧЕННЯ І ПРИКЛАДИ КОДИФЕРЕНЦІЙОВНИХ ФУНКЦІЙ

Нехай $X \subset \mathbb{R}^n$ — відкрита множина і нехай функція f – визначена і скінченна на X.

Означення 7.1.1. Функція $f(x)$ *кодиференційовна в точці* x, якщо існують опуклі компакти $\underline{d}f(x) \subset \mathbb{R}^{n+1}$ і $\overline{d}f(x) \subset \mathbb{R}^{n+1}$ такі, що

$$f(x+\Delta) = f(x) + \Phi_x(\Delta) + o_x(\Delta), \tag{7.1.1}$$

$$\Phi_x(\Delta) = \max_{[a,v]\in\underline{d}f(x)}[a + \langle v,\Delta\rangle] + \min_{[b,w]\in\overline{d}f(x)}[b + \langle w,\Delta\rangle], \tag{7.1.2}$$

$$\frac{o_x(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0, \quad \forall\Delta \in \mathbb{R}^n. \tag{7.1.3}$$

де $a,b \in \mathbb{R}^1$; $v,w \in \mathbb{R}^n$, $\operatorname{conv}\{x,x+\Delta\} \subset X$.

Якщо співвідношення (7.1.3) справджується рівномірно за $\Delta \in B = \{\Delta \in \mathbb{R}^n : \|\Delta\| = 1\}$, то функція f *кодиференційовна в точці x рівномірно за напрямками*.

Пара множин $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$ називається *кодиференціалом* функції f в точці x, множина $\underline{d}f(x)$ називається *гіподиференціалом*, а множина $\overline{d}f(x)$ — *гіпердиференціалом* функції f в точці x.

Якщо функція f кодиференційовна в деякому околі точки x, то відображення Df назвемо кодиференціальним.

Функція f називається *неперервно кодиференційовною у точці* x, якщо вона кодиференційовна в деякому околі точки x і якщо існує неперервне (за Хаусдорфом) кодиференціальне відображення $Df(x)$ в цій точці.

Функція f називається *гіподиференційовною в точці* x, якщо існує кодиференціал вигляду $Df(x) = [\underline{d}f(x),\{0_{n+1}\}]$.

Функція f називається *гіпердиференційовною в точці* x, якщо існує кодиференціал вигляду $Df(x) = [\{0_{n+1}\},\overline{d}f(x)]$. Тут 0_m— нульовий елемент простору $\mathbb{R}^m$.

Аналогічно визначаються неперервно гіподиференційовні функції і неперервно гіпердиференційовні функції.

Кодиференціал, як і квазідиференціал, визначається неоднозначно.

Із (7.1.1)—(7.1.2) видно, що кодиференційовна в точці x функція $f(x)$ допускає апроксимацію першого порядку в околі точки x. При цьому як апроксимацію можна взяти функцію $F_x(\Delta) = f(x) + \Phi_x(\Delta)$. Якщо функція f кодиференційовна в точці x рівномірно за напрямками, то $F_x(\Delta)$ є апроксимацією функції f в околі точки x рівномірно за напрямками. Для неперервно кодиференційовної функції апроксимація $F_x(\Delta)$, що побудована за допомогою кодиференціала, є неперервною функцією за x. Це в деякому сенсі поріднює неперервно кодиференційовні функції з неперервно диференційовними функціями.

Теорема 7.1.1. *Множина кодиференційовних в точці x функцій f співпадає з множиною квазідиференційовних функцій.*

Доведення. Дійсно, нехай функція f квазідиференційовна в точці x. Тоді

$$f(x+\Delta) = f(x) + \max_{v \in \underline{\partial} f(x)} \langle v,\Delta\rangle + \min_{w \in \overline{\partial} f(x)} \langle w,\Delta\rangle + o_x(\Delta),$$

$$\frac{o_x(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0 \quad \forall\Delta \in \mathbb{R}^n,$$

де $\underline{\partial} f(x) \subset \mathbb{R}^n$, $\overline{\partial} f(x) \subset \mathbb{R}^n$– опуклі компакти. Звідси випливає, що функція f кодиференційовна в точці x і як кодиференціал можна брати $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де

$$\underline{d}f(x) = \{[a,v] | a = 0, v \in \underline{\partial} f(x)\} \subset \mathbb{R}^{n+1}, \tag{7.1.4}$$

$$\overline{d}f(x) = \{[b,w] | b = 0, w \in \overline{\partial} f(x)\} \subset \mathbb{R}^{n+1}. \tag{7.1.5}$$

Навпаки, якщо функція f кодиференційовна в точці x, то має місце (7.1.1) – (7.1.3). Тоді

$$f'(x,g) = \lim_{\alpha\downarrow 0} \frac{1}{\alpha}[f(x+\alpha g) - f(x)] = \max_{v \in \underline{\partial} f(x)} \langle v,g\rangle + \min_{w \in \overline{\partial} f(x)} \langle w,g\rangle,$$

де

$$\underline{\partial} f(x) = \{v \in \mathbb{R}^n | [\overline{a},v] \in \underline{d}f(x)\}, \tag{7.1.6}$$

$$\overline{\partial} f(x) = \{w \in \mathbb{R}^n | [\overline{b},w] \in \overline{d}f(x)\}, \tag{7.1.7}$$

$$\overline{a} = \max\{a | [a,v] \in \underline{d}f(x)\}, \quad \overline{b} = \min\{b | [b,w] \in \overline{d}f(x)\}.$$

А це означає, що функція f квазідиференційовна в точці x, причому як квазідиференціал можна взяти $\mathrm{D}(x) = [\underline{\partial} f(x),\overline{\partial} f(x)]$, де множини $\underline{\partial} f(x)$, $\overline{\partial} f(x)$ визначені співвідношеннями (7.1.6)–(7.1.7). □

Наслідок 7.1.1. *Будь-яка гіподиференційовна функція є субдиференційовною, а будь-яка гіпердиференційовна функція є супердиференційовною. Справджується й обернене твердження.*

Зауваження 7.1.1. Клас кодиференційовних функцій співпадає с класом квазідиференційовних функцій. Проте використання поняття кодиференціала дозволяє виділити підмножину неперервно кодиференційовних функцій, в той час як для негладких функцій квазідиференціальне відображення є, взагалі кажучи, розривним.

Відмітимо, що квазідиференціал – це пара опуклих компактів у просторі $\mathbb{R}^n$, а кодиференціал – пара опуклих компактів у просторі $\mathbb{R}^{n+1}$.

Важливе значення (особливо з обчислювальної точки зору) мають кодиференційовні функції, в яких гіподиференціали та гіпердиференціали є опуклі оболонки скінченної кількості точок. Такі кодиференціали будемо називати *многогранними*.

Приклад 7.1.1. Нехай функція f неперервно диференційовна в околі точки $x \in X$. Тоді

$$f(x+\Delta) = f(x) + \langle f'(x),\Delta\rangle + o_x(\Delta), \quad \frac{o_x(\Delta)}{\|\Delta\|} \underset{\|\Delta\|\to 0}{\longrightarrow} 0,$$

де $f'(x)$ - градієнт функції f в точці x. Очевидно, що функція f неперервно кодиференційовна в околі точки x рівномірно за напрямками і як кодиференціал можна взяти $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де

$$\underline{d}f(x) = \{[0,f'(x)]\} \subset \mathbb{R}^{n+1}, \quad \overline{d}f(x) = \{0_{n+1}\} \subset \mathbb{R}^{n+1}.$$

Відображення $Df(x)$ неперервне. Пару множин $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де

$$\underline{d}f(x) = \{0_{n+1}\}, \quad \overline{d}f(x) = \{[0,f'(x)]\},$$

також можна взяти як кодиференціал. Отже, функція f є одночасно гіподиференційовною і гіпердиференційовною функцією (навіть неперервно гіподиференційовною і гіпердиференційовною функцією).

Приклад 7.1.2. Нехай f – опукла і скінченна на X функція, $x \in X$. Нехай X_0 – довільна замкнута обмежена підмножина X, що містить x як внутрішню точку. Оскільки

$$f(x) = \max_{z\in X_0}[f(z) + \langle v(z),x-z\rangle],$$

де $v(z)$ – вектор із $\underline{\partial}f(z)$, а $x+\Delta \in \operatorname{int} X_0$ при досить малих $\|\Delta\|$, і тому

$$f(x+\Delta) = f(x) + \max_{z\in X_0}[f(z) - f(x) + \langle v(z),x-z\rangle + \langle v(z),\Delta\rangle] =$$

$$= f(x) + \max_{[a,v]\in \underline{d}f(x)}[a + \langle v(z),\Delta\rangle],$$

де

$$\underline{d}f(x) = \operatorname{conv}\{[a,v]\colon a = f(z) - f(x) + \langle v(z),x-z\rangle, v(z) \in \underline{\partial}f(z), z \in X_0\}. \tag{7.1.8}$$

Отже опукла функція є неперервно кодиференційовною (і навіть гіподиференційвною). Як кодиференціал можна взяти $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де $\underline{d}f(x)$ задано (7.1.8), а $\overline{d}f(x) = \{0_{n+1}\}$.

Аналогічно встановлюється, що угнута функція неперервно кодиференційовна (і навіть неперервно гіпердиференційовна).

Приклад 7.1.3. Нехай $f(x) = \|x\| = \sqrt{\langle x,x\rangle}$. Оскільки

$$f(x) = \max_{\langle v,v\rangle \le 1} \langle x,v\rangle,$$

то

$$f(x+\Delta) = f(x) + \max_{\langle v,v\rangle \le 1} (\langle x,v\rangle - f(x) + \langle v,\Delta\rangle).$$

Звідси

$$f(x+\Delta) = f(x) + \max_{[a,v]\in \underline{d}f(x)} [a + \langle v,\Delta\rangle], \tag{7.1.9}$$

де

$$\underline{d}f(x) = \{[a,v] | a = \langle x,v\rangle - f(x), \langle v,v\rangle \le 1\}. \tag{7.1.10}$$

Отже, f є неперервно кодиференційовною (навіть гіподиференційовною) функцією, причому як кодиференціал в точці x можна взяти пару множин $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де $\underline{d}f(x)$ задано співвідношенням (7.1.10), а $\overline{d}f(x) = \{0_{n+1}\}$. Очевидно, що множина $\underline{d}f(x)$ є опуклою.

Приклад 7.1.4. Нехай

$$f(x) = \max_{y\in G} \varphi(x,y), \tag{7.1.11}$$

де G - компакт, а функція φ неперервна на $X \times G$ і неперервно диференційовна за x на X при кожному фіксованому $y \in G$. Тоді

$$f(x+\Delta) = \max_{y\in G} [\varphi(x,y) + \langle \varphi'_x(x,y),\Delta\rangle + o_x(\Delta,y)], \tag{7.1.12}$$

$$\frac{o_x(\Delta,y)}{\|\Delta\|} \underset{\|\Delta\|\to 0}{\longrightarrow} 0.$$

Якщо (7.1.12) має місце рівномірно за $y \in G$, то

$$f(x+\Delta) = f(x) + \max_{[a,v]\in \underline{d}f(x)} [a + \langle v,\Delta\rangle] + o_x(\Delta), \tag{7.1.13}$$

$$\frac{o_x(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0$$

рівномірно за $\Delta \in B$, де

$$\underline{d}f(x) = \operatorname{conv}\{[a,v] \colon a = \varphi(x,y) - f(x), v = \varphi'_x(x,y), y \in G\}, \tag{7.1.14}$$

Звідси випливає, що визначена співвідношенням (7.1.11) функція максимуму f є неперервно кодиференційвною на X, а її кодиференціал дорівнює

$$Df(x) = [\underline{d}f(x),\overline{d}f(x)], \tag{7.1.15}$$

де $\underline{d}f(x)$ задано співвідношенням (7.1.14), а $\overline{d}f(x) = \{0_{n+1}\}$. Отже функція максимуму гіподиференційовна.

Приклад 7.1.5. Аналогічно показується, що функція

$$f(x) = \min_{y \in G} \varphi(x,y),$$

де функція φ, множини X та G такі ж, як і в попередньому прикладі, є неперервно кодиференційовною, причому як $Df(x)$ можна взяти пару множин $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де

$$\underline{d}f(x) = \{0_{n+1}\},$$

$$\overline{d}f(x) = \operatorname{conv}\{[b,w]\colon b = \varphi(x,y) - f(x), w = \varphi'_x(x,y), y \in G\}.$$

Отже функція мінімуму гіпердиференційовна.

Приклад 7.1.6. Пояснимо наведені поняття на прикладі функції $f(x) = |x|$, де $x \in \mathbb{R}^1$. Оскільки $f(x) = \max\{x,-x\}$, то функція f диференційовна за напрямками на $\mathbb{R}^1$, причому

$$f'(x,g) = \begin{cases} g, & x > 0, \\ -g, & x < 0, \\ |g|, & x = 0. \end{cases}$$

Тут $g \in \mathbb{R}^1$. Функція f квазідиференційовна (навіть субдиференційовна) на $\mathbb{R}^1$, причому як її квазідиференціал можна взяти $\mathcal{D}f(x) = [\underline{\partial}f(x),\overline{\partial}f(x)]$, де

$$\underline{\partial}f(x) = \begin{cases} 1, & x > 0, \\ -1, & x < 0, \\ [-1,1], & x = 0, \end{cases} \qquad \overline{\partial}f(x) = \{0\}.$$

Очевидно, що квазідиференційовне відображення $\mathcal{D}f$ має розрив у точці $x = 0$.

Оскільки f - функція максимуму, то вона кодиференційовна і за формулою (7.1.15) $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де

$$\underline{d}f(x) = \operatorname{conv}\{(x - |x|\,,1),(-x - |x|\,, -1)\}, \quad \overline{d}f(x) = \{(0,0)\}.$$

Неважко переконатися в тому, що

$$\underline{d}f(x) = \begin{cases} \operatorname{conv}\{(0,1),(-2x, -1)\}, & x > 0, \\ \operatorname{conv}\{(2x,1),(0, -1)\}, & x \le 0. \end{cases}$$

Відображення Df є неперервним на $\mathbb{R}^1$. Маємо

$$f(x + \Delta) = \begin{cases} f(x) + \max\{\Delta, -2x - \Delta\}, & x > 0, \\ f(x) + \max\{2x + \Delta, -\Delta\}, & x \le 0, \end{cases}$$

тобто

$$f(x + \Delta) = \begin{cases} x + \max\{\Delta, -2x - \Delta\}, & x > 0, \\ -x + \max\{2x + \Delta, -\Delta\}, & x \le 0. \end{cases} \tag{7.1.16}$$

Відмітимо, що доданок $o_x(\Delta)$ тут відсутній, тобто за допомогою кодиференціала функцію f можна записати у вигляді (7.1.16) точно.

7.2. ОСНОВНІ ФОРМУЛИ КОДИФЕРЕНЦІАЛЬНОГО ЧИСЛЕННЯ

Далі розглядатимемо скінченні функції, що визначені на відкритій множині $X \subset \mathbb{R}^n$.

Лема 7.2.1. *Якщо $f_1(x),\ldots,f_N(x)$ - кодиференційовні (неперервно кодиференційовні) функції в точці $x \in X$, то функція $f(x) = \sum\limits_{i=1}^{N} c_i f_i(x)$, де $c_i \in \mathbb{R}^1$, кодиференційовна (неперервно кодиференційовна) в точці x. При цьому*

$$Df(x) = \sum_{i=1}^{N} c_i Df_i(x), \tag{7.2.1}$$

де $Df_i(x) = [\underline{d}f_i(x),\overline{d}f_i(x)]$ - кодиференціал функції $f_i(x)$ в точці x.

Доведення. Доведення очевидне, якщо врахувати, що

$$f_i(x+\Delta) = f_i(x) + \max_{[a,v]\in \underline{d}f_i(x)} [a + \langle v,\Delta\rangle] +$$
$$+ \min_{[b,w]\in \overline{d}f_i(x)} [b + \langle w,\Delta\rangle] + o_{ix}(\Delta).$$

□

Зауваження 7.2.1. Точніше кажучи, лему 7.2.1 потрібно сформулювати так:

Якщо $f_1(x),\ldots,f_N(x)$ – кодиференційовні (неперервно кодиференційовні) функції в точці $x \in X$, то і функція $f(x) = \sum\limits_{i=1}^{N} c_i f_i(x)$, де $c_i \in \mathbb{R}^1$, також кодиференційовна (неперервно кодиференційовна) в цій же точці. При цьому, якщо $Df_i(x)$, $i = 1,\ldots,N$, – кодиференціал функції $f_i(x)$ в точці x, то серед кодиференціалів функції $f(x)$ є кодиференціал вигляду

$$Df(x) = \sum_{i=1}^{N} c_i Df_i(x).$$

Якщо до того ж відображення $Df_i(x)$, $i = 1,\ldots,N$, неперервні, то й кодиференціальне відображення функції $f(x)$ вигляду (7.2.1) також неперервне.

Аналогічні зауваження можна зробити стосовно лем 7.2.2—7.2.4.

Лема 7.2.2. *Якщо функції $f_1(x)$ та $f_2(x)$ кодиференційовні (неперервно кодиференційовні) в точці $x \in X$, то функція $f(x) = f_1(x) \cdot f_2(x)$ кодиференційовна (неперервно кодиференційовна) в точці x. Якщо функції f_1, f_2 кодиференційовні рівномірно за напрямками, то і f кодиференційовна рівномірно за напрямками. Якщо функції $f_1(x)$ та $f_2(x)$ неперервно кодиференційовні в точці x, то функція $f(x)$ також неперервно кодиференційовна в цій точці. При цьому*

$$Df(x) = f_1(x)Df_2(x) + f_2(x)Df_1(x). \tag{7.2.2}$$

Доведення. Нехай

$$f_1(x + \Delta) = f_1(x) + \Phi_{1x}(\Delta) + o_1(\Delta), \tag{7.2.3}$$

$$f_2(x + \Delta) = f_2(x) + \Phi_{2x}(\Delta) + o_2(\Delta), \tag{7.2.4}$$

де

$$\Phi_{ix}(\Delta) = \max_{[a,v]\in \underline{d}f_i(x)}[a + \langle v,\Delta\rangle] + \min_{[b,w]\in \overline{d}f_i(x)}[b + \langle w,\Delta\rangle],$$

$$\frac{o_i(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0, \quad \forall\Delta \in \mathbb{R}^n.$$

Оскільки $\Phi_{ix}(0) = 0$, то

$$\max_{[a,v]\in \underline{d}f_i(x)} a + \min_{[b,w]\in \overline{d}f_i(x)} b = 0.$$

Можна вважати, що

$$\max_{[a,v]\in \underline{d}f_i(x)} a = \min_{[b,w]\in \overline{d}f_i(x)} b = 0. \tag{7.2.5}$$

Покладемо

$$\underline{\Phi}_{ix}(\Delta) = \max_{[a,v]\in \underline{d}f_i(x)}[a + \langle v,\Delta\rangle], \quad \overline{\Phi}_{ix}(\Delta) = \min_{[b,w]\in \overline{d}f_i(x)}[b + \langle w,\Delta\rangle].$$

В силу (7.2.5)

$$\max_{[0,v]\in \underline{d}f_i(x)}\langle v,\Delta\rangle \le \underline{\Phi}_{ix}(\Delta) \le \max_{[a,v]\in \underline{d}f_i(x)}\langle v,\Delta\rangle.$$

Звідси

$$|\underline{\Phi}_{ix}(\Delta)| \le \max_{[a,v]\in \underline{d}f_i(x)} |\langle v,\Delta\rangle| \le L_{1i}\left\|\Delta\right\|, \tag{7.2.6}$$

де $L_{1i} = \max\limits_{[a,v]\in \underline{d}f_i(x)} \|v\|$. Аналогічно, в силу (7.2.5),

$$\min_{[0,w]\in \overline{d}f_i(x)}\langle w,\Delta\rangle \ge \overline{\Phi}_{ix}(\Delta) \ge \min_{[b,w]\in \overline{d}f_i(x)}\langle w,\Delta\rangle.$$

Звідси

$$\left|\overline{\Phi}_{ix}(\Delta)\right| \le \max_{[b,w]\in [-\overline{d}f_i(x)]} |\langle w,\Delta\rangle| = \max_{[b,w]\in \overline{d}f_i(x)} |\langle w,\Delta\rangle| \le L_{2i}\left\|\Delta\right\|, \tag{7.2.7}$$

де $L_{2i} = \max\limits_{[b,w]\in\overline{d}f_i(x)} \|w\|$. З (7.2.6) та (7.2.7) випливає, що

$$|\Phi_{1x}(\Delta)| \leq L_1 \|\Delta\|, \quad |\Phi_{2x}(\Delta)| \leq L_2 \|\Delta\|, \tag{7.2.8}$$

де $L_i = L_{1i} + L_{2i}$, $i = 1{,}2$. Очевидно, що в силу компактності $\underline{d}f_i(x)$ та $\overline{d}f_i(x)$, будемо мати $L_i < \infty$, $i = 1{,}2$.

Відмітимо, що співвідношення (7.2.8) істинні незалежно від умови (7.2.5).

Із співвідношень (7.2.3) та (7.2.4) маємо

$$f(x+\Delta) = f_1(x+\Delta)f_2(x+\Delta) = f_1(x)f_2(x) + f_1(x)\Phi_{2x}(\Delta)+$$

$$+f_2(x)\Phi_{1x}(\Delta) + f_1(x)o_2(\Delta) + f_2(x)o_1(\Delta) + \Phi_{1x}(\Delta)\Phi_{2x}(\Delta) + o_1(\Delta)o_2(\Delta).$$

В силу (7.2.8) виконується співвідношення

$$f(x+\Delta) = f(x) + f_1(x)\Phi_{2x}(\Delta) + f_2(x)\Phi_{1x}(\Delta) + o(\Delta),$$

$$\frac{o(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0, \quad \forall\Delta \in \mathbb{R}^n.$$

Враховуючи вигляд $\Phi_{ix}(\Delta)$, маємо

$$f(x+\Delta) = f(x) + f_1(x)\left[\max_{[a,v]\in\underline{d}f_2(x)}(a + \langle v,\Delta\rangle) + \min_{[b,w]\in\overline{d}f_2(x)}(b + \langle w,\Delta\rangle)\right] +$$

$$+f_2(x)\left[\max_{[a,v]\in\underline{d}f_1(x)}(a + \langle v,\Delta\rangle) + \min_{[b,w]\in\overline{d}f_1(x)}(b + \langle w,\Delta\rangle)\right] + o(\Delta).$$

Відповідно до правил додавання та множення на число в просторі множин матимемо

$$f(x+\Delta) = f(x) + \max_{[a,v]\in\underline{d}f(x)}(a + \langle v,\Delta\rangle) + \min_{[b,w]\in\overline{d}f(x)}(b + \langle w,\Delta\rangle) + o(\Delta),$$

$$Df(x) = [\underline{d}f(x),\overline{d}f(x)] = f_1(x)Df_2(x) + f_2(x)Df_1(x),$$

тобто функція f кодиференційовна в точці x. □

Лема 7.2.3. *Якщо функція $f_1(x)$ кодиференційовна (неперервно кодиференційовна) в точці x і $f_1(x) \neq 0$, то функція $f(x) = 1/f_1(x)$ кодиференційовна (неперервно кодиференційовна) в точці x, при цьому*

$$Df(x) = -\frac{1}{f_1^2(x)}Df_1(x). \tag{7.2.9}$$

Доведення. Нехай

$$f_1(x+\Delta) = f_1(x) + \Phi_{1x}(\Delta) + o_1(\Delta),$$

де

$$\Phi_{1x}(\Delta) = \max_{[a,v]\in\underline{d}f_1(x)}(a + \langle v,\Delta\rangle) + \min_{[b,w]\in\overline{d}f_1(x)}(b + \langle w,\Delta\rangle),$$

$$o_1(\Delta) = o_1(x,\Delta), \quad \frac{o_1(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0, \quad \forall\Delta \in R^n. \tag{7.2.10}$$

Як і при доведенні леми 7.2.2, показуємо, що

$$|\Phi_{1x}(\Delta)| \le L_1 \left\|\Delta\right\|, \quad \text{де } L_1 < \infty. \tag{7.2.11}$$

Далі

$$f(x+\Delta) - f(x) = \frac{1}{f_1(x) + \ \Phi_{1x}(\Delta) + o_1(\Delta)} - \frac{1}{f_1(x)} =$$

$$= \frac{-\Phi_{1x}(\Delta) - o_1(\Delta)}{f_1(x)(f_1(x) + \ \Phi_{1x}(\Delta) + o_1(\Delta))} =$$

$$= -\frac{\Phi_{1x}(\Delta)}{f_1^2(x)} + \frac{\Phi_{1x}^2(\Delta) + \Phi_{1x}(\Delta)o_1(\Delta) - o_1(\Delta)f_1(x)}{f_1^2(x)(f_1(x) + \ \Phi_{1x}(\Delta) + o_1(\Delta))}. \tag{7.2.12}$$

Із (7.2.11) випливає, що $\Phi_{1x}^2(\Delta) = o(\Delta)$, тому із (7.2.12) матимемо

$$f(x+\Delta) = f(x) - \frac{1}{f_1^2(x)}\Phi_{1x}(\Delta) + o(\Delta), \tag{7.2.13}$$

$$\frac{o(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0 \quad \forall\Delta \in \mathbb{R}^n.$$

Із (7.2.13), (7.2.10) випливає співвідношення (7.2.9). $\square$

Лема 7.2.4. *Якщо функції $\varphi_i(x), i \in I = \{1,\ldots,N\}$ кодиференційовні (неперервно кодиференційовні) в точці x, то функції*

$$f_1(x) = \max_{i\in I} \varphi_i(x), \quad f_2(x) = \min_{i\in I} \varphi_i(x)$$

також кодиференційовні (неперервно кодиференційовні) в точці x. При цьому

$$Df_1(x) = [\underline{d}f_1(x),\overline{d}f_1(x)], \quad Df_2(x) = [\underline{d}f_2(x),\overline{d}f_2(x)],$$

де

$$\underline{d}f_1(x) = \operatorname{conv}\{\underline{d}\varphi_k(x) - \sum_{i\in I\setminus\{k\}} \overline{d}\varphi_i(x) + \{[\varphi_k(x) - f_1(x),0_n]\}|k \in I\},$$

$$\overline{d}f_1(x) = \sum_{i\in I} \overline{d}\varphi_i(x), \quad \underline{d}f_2(x) = \sum_{i\in I} \underline{d}\varphi_i(x),$$

$$\overline{d}f_2(x) = \operatorname{conv}\{\overline{d}\varphi_k(x) - \sum_{i\in I\setminus\{k\}} \underline{d}\varphi_i(x) + \{[\varphi_k(x) - f_2(x),0_n]\}|k \in I\}.$$

Доведення. Розглянемо лише функцію $f_1(x)$. Маємо

$$\varphi_i(x+\Delta) = \varphi_i(x) + \Phi_{ix}(\Delta) + o_{ix}(\Delta), \quad i \in I,$$

де

$$\Phi_{ix}(\Delta) = \max_{[a,v]\in\underline{d}\varphi_i(x)} [a + \langle v,\Delta\rangle] + \min_{[b,w]\in\overline{d}\varphi_i(x)} [b + \langle w,\Delta\rangle], \tag{7.2.14}$$

$$\frac{o_{ix}(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0, \quad \forall\Delta \in \mathbb{R}^n.$$

Тоді

$$f_1(x+\Delta) = \max_{i\in I} \varphi_i(x+\Delta) =$$

$$= f_1(x) + \max_{i\in I}\{\Phi_{ix}(\Delta) + \varphi_i(x) - f_1(x) + o_{ix}(\Delta)\} =$$

$$= f_1(x) + \max_{i\in I}\{\Phi_{ix}(\Delta) + \varphi_i(x) - f_1(x)\} + o_x(\Delta),$$

$$o_x(\Delta) \in [\min_{i\in I} o_{ix}(\Delta), \max_{i\in I} o_{ix}(\Delta)].$$

Очевидно, що

$$\frac{o_x(\alpha\Delta)}{\alpha} \underset{\alpha\downarrow 0}{\longrightarrow} 0, \quad \forall\Delta \in \mathbb{R}^n.$$

Тепер залишилось, враховуючи (7.2.14), скористатись теоремою 6.2.1. $\square$

Зауважимо, що в тому випадку, коли для кодиференціалів функцій $\varphi_i(x)$ виконується умова (7.2.5), то і для кодиференціалів функцій $f_1(x)$ та $f_2(x)$ також виконуються аналогічні умови, оскільки для функції $f_1(x)$ виконується умова $\varphi_k(x) - f_1(x) \leq 0$, а для функції $f_2(x)$ виконується умова $\varphi_k(x) - f_2(x) \geq 0$.

Далі буде показано, що суперпозиція кодиференційовних (неперервно кодиференційовних) функцій є кодиференційовною (неперервно кодиференційовною) функцією.

Отже клас кодиференційовних (неперервно кодиференційовних) функцій – це лінійний простір, замкнутий відносно всіх "гладких" операцій і, що особливо важливо, відносно операцій взяття поточкового максимуму та поточкового мінімуму від скінченної кількості функцій.

7.3. ПРИКЛАДИ ЗАСТОСУВАННЯ ФОРМУЛ КОДИФЕРЕЦІАЛЬНОГО ЧИСЛЕННЯ

Приклад 7.3.1. Нехай $f(x) = \mathrm{sat}\, f_1(x)$, де $f_1(x)$ – диференційовна на $\mathbb{R}^n$ функція,

$$\mathrm{sat}\, a = \begin{cases} a, & \text{якщо} \quad |a| \leq 1, \\ 1, & \text{якщо} \quad a > 1, \\ -1, & \text{якщо} \quad a < -1. \end{cases}$$

Функцію $f(x)$ можна записати у вигляді $f(x) = \min\{1, \max\{-1, f_1(x)\}\}$. Покладемо $\varphi_1(x) = 1$, $\varphi_2(x) = -1$, $\varphi_3(x) = \max\{\varphi_2(x), f_1(x)\}$. Тоді $f(x) = \min\{\varphi_1(x), \varphi_3(x)\}$. Оскільки f_1, φ_1 та φ_2 диференційовні функції, то

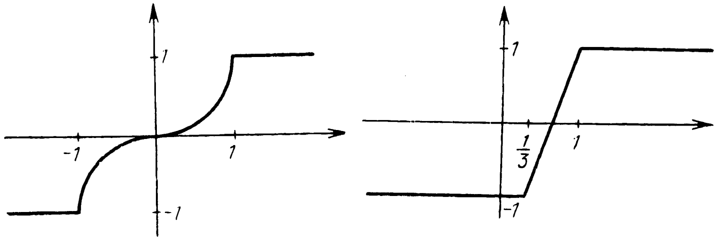


Рис. 7.3.1:

можна взяти $Df_1(x) = [\underline{d}f_1(x),\overline{d}f_1(x)]$, де $\underline{d}f_1(x) = \{[0,f_1'(x)]\} \subset \mathbb{R}^{n+1}$, $\overline{d}f_1(x) = \{0_{n+1}\} \subset \mathbb{R}^{n+1}$, $f_1'(x)$ - градієнт функції f_1 в точці x,

$$D\varphi_1(x) = [\underline{d}\varphi_1(x), \overline{d}\varphi_1(x)], \quad D\varphi_2(x) = [\underline{d}\varphi_2(x),\overline{d}\varphi_2(x)],$$

$$\underline{d}\varphi_1(x) = \{[0,0_n]\}, \ \overline{d}\varphi_1(x) = \overline{d}\varphi_2(x) = \{0_{n+1}\}, \ \underline{d}\varphi_2(x) = \{[0,0_n]\}.$$

За правилом обчислення кодиференціала функції максимуму (лема 7.2.4) $D\varphi_3(x) = [\underline{d}\varphi_3(x),\overline{d}\varphi_3(x)]$, де

$$\underline{d}\varphi_3(x) = \operatorname{conv}\{\underline{d}\varphi_2(x) - \overline{d}f_1(x) + \{[\varphi_2(x) - \varphi_3(x),0_n]\},$$

$$\underline{d}f_1(x) - \overline{d}\varphi_2(x) + \{[f_1(x) - \varphi_3(x),0_n]\}\} =$$

$$= \operatorname{conv}\{-\overline{d}f_1(x) + \{[-1 - \varphi_3(x),0_n]\}, \underline{d}f_1(x) + \{[f_1(x) - \varphi_3(x),0_n]\}\},$$

$$\overline{d}\varphi_3(x) = \overline{d}\varphi_2(x) + \overline{d}f_1(x) = \{0_{n+1}\}.$$

За правилом обчислення кодиференціала функції мінімуму (лема 7.2.4) $Df(x) = [\underline{d}f(x),\overline{d}f(x)]$, де

$$\underline{d}f(x) = \underline{d}\varphi_1(x) + \underline{d}\varphi_3(x) = \underline{d}\varphi_3(x) \tag{7.3.1}$$

$$\overline{d}f(x) = \operatorname{conv}\{\overline{d}\varphi_1(x) - \underline{d}\varphi_3(x) + \{[\varphi_1(x) - f(x),0_n]\}, \tag{7.3.2}$$

$$\overline{d}\varphi_3(x) - \underline{d}\varphi_1(x) + \{[\varphi_3(x) - f(x),0_n]\}\} =$$

$$= \operatorname{conv}\{-\underline{d}\varphi_3(x) + \{[1 - f(x),0_n]\},\{[\varphi_3(x) - f(x),0_n]\}\}. \tag{7.3.3}$$

Застосуємо формули (7.3.1)–(7.3.3) до функції $f_1(x) = x^3$, $x \in \mathbb{R}^1$. Тоді (див. рис. 7.3.1) $\underline{d}f_1(x) = \{(0,3x^2)\} \subset \mathbb{R}^2$, $\overline{d}f_1(x) = \{(0,0)\} \subset \mathbb{R}^2$. Для $x_0 = 1$ маємо

$$\varphi_3(x_0) = \max\{-1,x_0^3\} = \max\{-1,1\} = 1, \quad f_1(x_0) = 1, \quad f(x_0) = 1,$$

$$\underline{d}f(x_0) = \underline{d}\varphi_3(x_0) = \operatorname{conv}\{\{(-1-1,0)\},\{(0,3)+(1-1,0)\}\} = \operatorname{conv}\{(-2,0),(0,3)\},$$

$$\overline{d}f(x_0) = \operatorname{conv}\{-\underline{d}\varphi_3(x_0)+\{(1-1,0)\},\{(1-1,0)\}\} = \operatorname{conv}\{(2,0),(0,-3),(0,0)\}.$$

Тоді за формулами (7.1.1), (7.1.2) маємо

$$f(x_0 + \Delta) = f(x_0) + \Phi_{x_0}(\Delta) + o_x(\Delta),$$

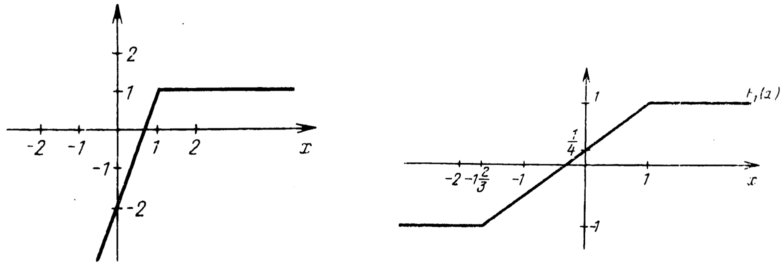


Рис. 7.3.2:

де

$$\Phi_x(\Delta) = \max\{-2; 3\Delta\} + \min\{2, -3\Delta, 0\}.$$

На рис. 7.3.1 зображений графік функції $F_1(x) = f(x_0) + \Phi_{x_0}(x - x_0)$. Неважко показати, що для цієї функції квазідиференціал в точці $x_0 = 1$ має вигляд $\mathfrak{D} f(x_0) = [\underline{\partial} f(x_0), \overline{\partial} f(x_0)]$, де

$$\underline{\partial} f(x_0) = \{f_1'(x_0)\} = \{3\} \subset \mathbb{R}^1,$$

$$\overline{\partial} f(x_0) = \operatorname{conv}\{0, -f_1'(x_0)\} = [0, -3].$$

Тому

$$f(x_0 + \Delta) = f(x_0) + \max_{v \in \underline{\partial} f(x_0)} \langle v, \Delta \rangle + \min_{w \in \overline{\partial} f(x_0)} \langle w, \Delta \rangle + o(\Delta) =$$

$$= f(x_0) + \overline{\Phi}_{x_0}(\Delta) + o(\Delta),$$

де

$$\overline{\Phi}_x(\Delta) = 3\Delta + \min\{0, -3\Delta\}.$$

На рис. 7.3.2 зображений графік функції $\overline{F}_1(x) = f(x_0) + \overline{\Phi}_{x_0}(x - x_0)$, що є апроксимацією f в околі точки $x_0 = 1$, яка отримана за допомогою квазідиференціала. Порівнюючи $F_1(x)$ та $\overline{F}_1(x)$, бачимо, що в околі точки $x_0 = 1$ функції $F_1(x)$ та $\overline{F}_1(x)$ співпадають.

Розглянемо тепер точку $x_0 = -1/2$. Маємо

$$\varphi_3(x_0) = \max\{-1, -1/8\} = -1/8, f_1(x_0) = -1/8, f(x_0) = -1/8,$$

$$\underline{d} f_1(x_0) = \{(0,3/4)\}, \overline{d} f_1(x_0) = \{(0,0)\},$$

$$\underline{d}\varphi_3(x_0) = \operatorname{conv}\{\{(-1 + 1/8, 0)\}, \{(0,3/4) + \{(-1/8 + 1/8, 0)\}\} =$$

$$= \operatorname{conv}\{(-7/8, 0), (0, 3/4)\},$$

$$\underline{d} f(x_0) = \underline{d}\varphi_3(x_0) = \operatorname{conv}\{(-7/8, 0), (0, 3/4)\},$$

$$\overline{d} f(x_0) = \operatorname{conv}\{-\underline{d}\varphi_3(x_0) + \{(1 + 1/8, 0)\}, (-1/8 + 1/8, 0)\} =$$

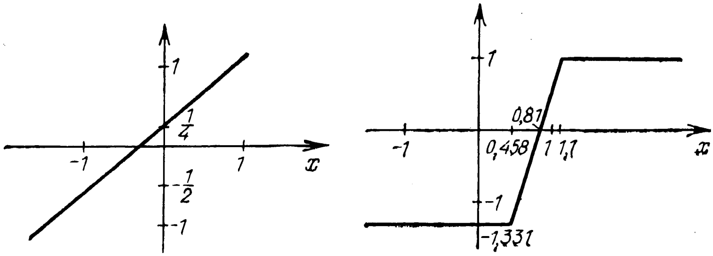


Рис. 7.3.3:

$$= \mathrm{conv}\{(2{,}0),(9/8,-3/4),(0{,}0)\},$$

$$\Phi_{x_0}(\Delta) = \max\{-7/8{,}3\Delta/4\} + \min\{2{,}0{,}9/8 - 3\Delta/4\};$$

$$f(x_0+\Delta) = f(x_0) + \Phi_{x_0}(\Delta) + o_x(\Delta).$$

На рис. 7.3.2 зображений графік функції $F_1(x) = f(-1/2) + \Phi_{-1/2}(x + 1/2)$, значення якої в точці $x_0 = -1/2$ співпадає зі значенням функції $f(x)$.

Побудуємо тепер квазідиференціал функції f в точці $x_0 = -1/2$. Маємо $\mathfrak{D}f(x_0) = [\underline{\partial}f(x_0),\overline{\partial}f(x_0)]$, де

$$\underline{\partial}f(x_0) = \{f_1'(x_0)\} = \{3x_0^2\} = \{3/4\}, \overline{\partial}f(x_0) = \{0\}.$$

Тому

$$f(x_0+\Delta) = f(x_0) + \max_{v\in\underline{\partial}f(x_0)}\langle v,\Delta\rangle + \min_{w\in\overline{\partial}f(x_0)}\langle w,\Delta\rangle + o(\Delta) =$$

$$= -\frac{1}{8} + \overline{\Phi}_{x_0}(\Delta) + o(\Delta), \quad \overline{\Phi}_{x_0}(\Delta) = \frac{3}{4}\Delta.$$

На рис. 7.3.3 зображений графік функції

$$\overline{F}_1(x) = f(x_0) + \overline{\Phi}_{x_0}(x - x_0) = -\frac{1}{8} + \frac{3}{4}\left(x + \frac{1}{2}\right) = \frac{3}{4}x + \frac{1}{4},$$

що є апроксимацією функції f в околі точки $x_0 = -1/2$, яка отримана за допомогою квазідиференціала.

Нарешті, розглянемо точку $x_0 = 1{,}1$. Маємо

$$\varphi_3(x_0) = \max\{\varphi_2(x_0),f_1(x_0)\} = \max\{-1; 1{,}331\} = 1{,}331,$$

$$f_1(x_0) = 1{,}331, \quad f(x_0) = 1,$$

$$\underline{d}f_1(x_0) = \{(0; 3{,}63)\}, \quad \overline{d}f_1(x_0) = \{(0{,}0)\},$$

$$\underline{d}\varphi_3(x_0) = \mathrm{conv}\{\{(-1 - 1{,}331; 0),(0; 3{,}63) + (1{,}331 - 1{,}331; 0)\} =$$

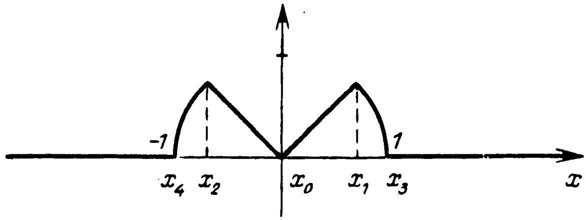


Рис. 7.3.4:

$$= \operatorname{conv}\{(-2{,}331; 0),(0; 3{,}63)\}.$$

$$\overline{d}\varphi_3(x_0) = \{(0{,}0)\},$$

$$\underline{d}f(x_0) = \underline{d}\varphi_3(x_0) = \operatorname{conv}\{(-2{,}331; 0); (0; 3{,}63)\},$$

$$\overline{d}f(x_0) = \operatorname{conv}\{\operatorname{conv}\{(2{,}331; 0); (0; -3{,}63)\} + \{(1-1; 0)\}; (0{,}331; 0)\} =$$

$$= \operatorname{conv}\{(2{,}331; 0); (0; -3{,}63); (0{,}331; 0)\}.$$

За формулами (7.1.1), (7.1.2) маємо

$$f(x_0 + \Delta) = f(x_0) + \Phi_{x_0}(\Delta) + o_x(\Delta),$$

де

$$\Phi_{x_0}(\Delta) = \max\{-2{,}331; 3{,}63\Delta\} + \min\{2{,}331; 0{,}331; -3{,}63\Delta\}.$$

На рис. 7.3.3 зображений графік функції $F_1(x) = f(x_0) + \Phi_{x_0}(x - x_0)$, що є апроксимацією функції f в околі точки $x_0 = 1{,}1$, яка отримана за допомогою кодиференціала.
Квазидиференціалом функції f в точці $x_0 = 1{,}1$ є $\mathfrak{D}f(x_0) = [\{(0)\},\{(0)\}]$, тобто $\overline{\Phi}_1(\Delta) = 0$ та $\overline{F}_1(x) = 1 \quad \forall x$. Ясно, що $F_1(x)$ краще апроксимує f в околі точки $x_0 = 1{,}1$, ніж функція $\overline{F}_1(x)$.

Приклад 7.3.2. Нехай (рис. 7.3.4)

$$f_1(x) = x,\ \ f_2(x) = -x,\ \ f_3(x) = 0,\ \ f_4(x) = 1 - x^2,$$

$$f_5(x) = \max\{f_1(x), f_2(x)\},\ \ f_6(x) = \max\{f_3(x), f_4(x)\},$$

$$f(x) = \min\{f_5(x), f_6(x)\}.$$

Оскільки функції f_1, f_2, f_3 та f_4 — диференційовні, покладемо

$$Df_1(x) = [\underline{d}f_1(x), \overline{d}f_1(x)] = [\{(0{,}1)\},\{(0{,}0)\}],$$

$$Df_2(x) = [\underline{d}f_2(x), \overline{d}f_2(x)] = [\{(0, -1)\},\{(0{,}0)\}],$$

$$Df_3(x) = [\underline{d}f_3(x), \overline{d}f_3(x)] = [\{(0{,}0)\}, \{(0{,}0)\}],$$
$$Df_4(x) = [\underline{d}f_4(x), \overline{d}f_4(x)] = [\{(0, -2x)\}, \{(0{,}0)\}].$$

Використовуючи лему 7.2.4, маємо

$$Df_5(x) = [\underline{d}f_5(x), \overline{d}f_5(x)],$$

де

$$\underline{d}f_5(x) = \text{conv}\{\underline{d}f_1(x) - \overline{d}f_2(x) + \{(f_1(x) - f_5(x), 0)\},$$
$$\underline{d}f_2(x) - \overline{d}f_1(x) + \{(f_2(x) - f_5(x), 0)\}\} = \text{conv}\{(0,1) - (0,0) + \{(x - f_5(x), 0)\},$$
$$\{(0, -1) - (0,0) + (-x - f_5(x), 0)\}\} = \text{conv}\{(x - f_5(x), 1),\ (-x - f_5(x), -1)\},$$
$$\overline{d}f_5(x) = \overline{d}f_1(x) + \overline{d}f_2(x) = \{(0,0)\};$$
$$Df_6(x) = [\underline{d}f_6(x), \overline{d}f_6(x)],$$

де

$$\underline{d}f_6(x) = \text{conv}\{\underline{d}f_3(x) - \overline{d}f_4(x) + \{(f_3(x) - f_6(x), 0)\},$$
$$\underline{d}f_4(x) - \overline{d}f_3(x) + \{(f_4(x) - f_6(x), 0)\}\} = \text{conv}\{(0,0) - (0,0) + \{(0 - f_6(x), 0)\},$$
$$(0, -2x) + \{(1 - x^2 - f_6(x), 0)\}\} = \text{conv}\{(-f_6(x), 0), (1 - x^2 - f_6(x), -2x)\},$$
$$\overline{d}f_6(x) = \overline{d}f_3(x) + \overline{d}f_4(x) = \{(0,0)\};$$
$$Df(x) = [\underline{d}f(x), \overline{d}f(x)],$$

де

$$\underline{d}f(x) = \underline{d}f_5(x) + \underline{d}f_6(x) = \text{conv}\{(x - f_5(x), 1),\ (-x - f_5(x), -1)\} +$$
$$+ \text{conv}\{(-f_6(x), 0),\ (1 - x^2 - f_6(x), -2x)\} =$$
$$= \text{conv}\{(x - f_5(x) - f_6(x), 1),\ (1 + x - x^2 - f_5(x) - f_6(x), 1 - 2x),$$
$$(-x - f_5(x) - f_6(x), -1), (1 - x - x^2 - f_5(x) - f_6(x), -1 - 2x)\}$$
$$\overline{d}f(x) = \text{conv}\{\overline{d}f_5(x) - \underline{d}f_6(x) + \{(f_5(x) - f(x), 0)\},$$
$$\overline{d}f_6(x) - \underline{d}f_5(x) + \{(f_6(x) - f(x), 0)\}\} =$$
$$= \text{conv}\{(0,0) - \text{conv}\{(-f_6(x), 0),\ (1 - x^2 - f_6(x), -2x)\} + \{(f_5(x) - f(x), 0)\},$$
$$(0,0) - \text{conv}\{(x - f_5(x), 1),\ (-x - f_5(x), -1)\} + \{(f_6(x) - f(x), 0)\}\} =$$
$$= \text{conv}\{(f_5(x) + f_6(x) - f(x), 0),\ (-1 + x^2 + f_6(x) + f_5(x) - f(x), 2x),$$
$$(f_6(x) - f(x) - x + f_5(x), -1),\ (f_6(x) - f(x) + x + f_5(x), 1)\}.$$

Розглянемо точку $x_0 = 1/2$. Маємо

$$f_1(x_0) = 1/2,\ \ f_2(x_0) = -1/2,\ \ f_3(x_0) = 0,\ \ f_4(x_0) = 1 - 1/4 = 3/4,$$
$$f_5(x_0) = \max\{f_1(x_0), f_2(x_0)\} = 1/2,\ \ f_6(x_0) = \max\{f_3(x_0), f_4(x_0)\} = 3/4,$$

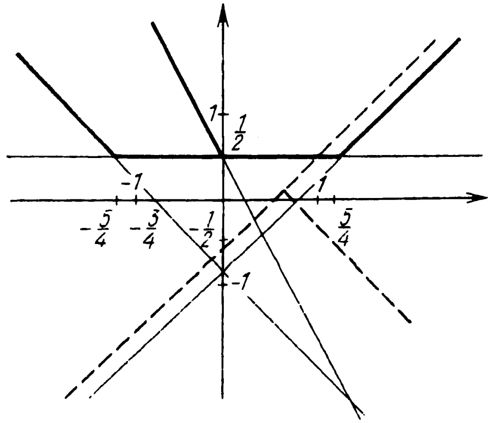


Рис. 7.3.5:

$$f(x_0) = \min\{f_5(x_0), f_6(x_0)\} = 1/2,$$

$$\underline{d}f(x_0) = \operatorname{conv}\{(1/2 - 1/2 - 3/4, 1),\ (1 + 1/2 - 1/4 - 1/2 - 3/4, 1 - 2 \cdot 1/2),$$

$$(-1/2 - 1/2 - 3/4,\ -1),\ (1 - 1/2 - 1/4 - 1/2 - 3/4,\ -1 - 1)\} =$$

$$= \operatorname{conv}\{(-3/4, 1),\ (0, 0),\ (-7/4,\ -1),\ (-1,\ -2)\},$$

$$\overline{d}f(x_0) = \operatorname{conv}\{(1/2 + 3/4 - 1/2, 0),\ (-1 + 1/4 + 3/4 + 1/2 - 1/2, 1),$$

$$(3/4 - 1/2 - 1/2 + 1/2,\ -1),\ (3/4 - 1/2 + 1/2 + 1/2, 1)\} =$$

$$= \operatorname{conv}\{(3/4, 0),\ (0, 1),\ (1/4,\ -1),\ (5/4, 1)\}.$$

Звідси (рис. 7.3.5)

$$f(x_0 + \Delta) = f(x_0) + \Phi_{x_0}(\Delta) + o(\Delta) = 1/2 + \Phi_{x_0}(\Delta) + o(\Delta),$$

де

$$\Phi_{x_0}(\Delta) = \max\{-3/4 + \Delta, 0,\ -7/4 - \Delta,\ -1 - 2\Delta\} +$$

$$\min\{3/4, \Delta, 1/4 - \Delta, 5/4 + \Delta\}.$$

На рис. 7.3.6 зображено графік функції $F_1(x) = f(x_0) + \Phi_{x_0}(x - x_0)$, яка є апроксимацією функції f в околі точки $x_0 = 1/2$, одержаної за допомогою кодиференціалу. У той же час апроксимацією функції f в околі тієї ж самої точки x_0, одержаної за допомогою квазідиференціалу, є функція $\overline{F}_1(x) = x$.

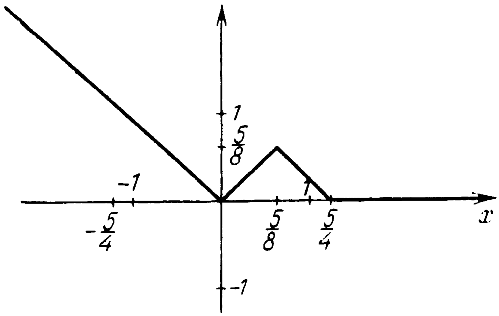


Рис. 7.3.6:

7.4. КОДИФЕРЕНЦІЙОВНІСТЬ СУПЕРПОЗИЦІЇ

Нехай $z \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ і нехай $y_1(x), \ldots, y_m(x)$ – кодиференційовні в точці x_0 функції. Припустимо, що функція
$F(z)$ кодиференційовна в точці $z_0 = (y_1(x_0), \ldots, y_m(x_0))$. Тоді

$$y_i(x_0 + \Delta x) = y_i(x_0) + \max_{[\tilde{a}_i,\, \tilde{v}_i] \in \underline{d} y_i(x_0)} [\tilde{a}_i + \langle \tilde{v}_i, \Delta x \rangle] +$$

$$+ \min_{[\tilde{b}_i, \tilde{w}_i] \in \overline{d} y_i(x_0)} [\tilde{b}_i + \langle \tilde{w}_i, \Delta x \rangle] + o_i(\Delta x) \quad \forall i \in \{1, \ldots, m\}, \tag{7.4.1}$$

$$F(z_0 + \Delta z) = F(z_0) + \max_{[a,\, v] \in \underline{d} F(z_0)} [a + \langle v, \Delta z \rangle] +$$

$$+ \min_{[b,\, w] \in \overline{d} F(z_0)} [b + \langle w, \Delta z \rangle] + o(\Delta z). \tag{7.4.2}$$

Тут

$$\frac{o_i(\alpha \Delta x)}{\alpha} \underset{\alpha \downarrow 0}{\longrightarrow} 0 \ \ \forall \Delta x \in \mathbb{R}^n, \quad \frac{o(\alpha \Delta x)}{\alpha} \underset{\alpha \downarrow 0}{\longrightarrow} 0 \ \ \forall \Delta z \in \mathbb{R}^m,$$

$\underline{d} y_i(x_0)$, $\overline{d} y_i(x_0)$— опуклі компакти у $\mathbb{R}^{n+1}$, $\underline{d} F(z_0)$, $\overline{d} F(z_0)$— опуклі компакти у $\mathbb{R}^{m+1}$.

Теорема 7.4.1. *Якщо функції $y_1(x), \ldots, y_m(x)$ кодиференційовні в точці x_0, функція $F(z)$ кодиференційовна в точці z_0, то функція $f(x) = F(y_1(x), \ldots, y_m(x))$ кодиференційовна в точці x_0. Якщо функції $y_1(x), \ldots,$ $y_m(x)$ неперервно кодиференційовні в околі точки x_0, а функція $F(z)$ неперервно кодиференційовна в околі точки z_0, то і функція $f(x)$ також неперервно кодиференційовна в околі точки x_0.*

Доведення. Покладемо

$$r_i'(\Delta x) = \max_{[\tilde a_i,\, \tilde v_i]\, \in \underline{d} y_i(x_0)} [\tilde a_i + \langle \tilde v_i, \Delta x\rangle],$$
$$r_i''(\Delta x) = \min_{[\tilde b_i,\, \tilde w_i]\, \in \overline{d} y_i(x_0)} [\tilde b_i + \langle \tilde w_i, \Delta x\rangle],$$

$$r_i(\Delta x) = r_i'(\Delta x) + r_i''(\Delta x), \;\; r'(\Delta z) = \max_{[a,\, v]\, \in \underline{d} F(z_0)} [a + \langle v, \Delta z\rangle],$$
$$r''(\Delta z) = \min_{[b,\; w]\, \in \overline{d} F(z_0)} [b + \langle w, \Delta z\rangle], \;\; r(\Delta z) = r'(\Delta z) + r''(\Delta z).$$

З (7.4.1) та (7.4.2) робимо висновок, що

$$\max_{[\tilde a_i,\, \tilde v_i]\, \in \underline{d} y_i(x_0)} \tilde a_i + \min_{[\tilde b_i,\, \tilde w_i]\, \in \overline{d} y_i(x_0)} \tilde b_i = 0 \quad \forall i \in 1, ..., m,$$
$$\max_{[a,\, v]\, \in \underline{d} F(z_0)} a + \min_{[b,\, w]\, \in \overline{d} F(z_0)} b = 0.$$

Звідси можна отримати (див. доведення леми 7.2.2), що

$$|r_i(\Delta x)| \le K_i ||\Delta x||, \tag{7.4.3}$$

$$|r(\Delta z)| \le K ||\Delta z||, \tag{7.4.4}$$

де

$$K_i = \max_{[\tilde a_i,\, \tilde v_i]\, \in \underline{d} y_i(x_0)} \|\tilde v_i\| + \min_{[\tilde b_i,\, \tilde w_i] \in \overline{d} y_i(x_0)} \|\tilde w_i\|,$$
$$K = \max_{[a,\, v]\, \in \underline{d} F(z_0)} \|v\| + \min_{[b,\, w] \in \overline{d} F(z_0)} \|w\|.$$

Маємо

$$f(x_0 + \Delta x) = F(y_1(x_0 + \Delta x), \ldots, y_m(x_0 + \Delta x)) =$$
$$= F(y_1(x_0) + r_1(\Delta x) + o_1(\Delta x), \ldots, y_m(x_0) + r_m(\Delta x) + o_m(\Delta x)) =$$
$$= F(z_0 + \Delta z), \tag{7.4.5}$$

де $\Delta z = (r_1(\Delta x) + o_1(\Delta x), \ldots, r_m(\Delta x) + o_m(\Delta x))$. З (7.4.3) зрозуміло, що для достатньо малих $||\Delta x||$ можна знайти K_0 таке, що

$$\|\Delta z\| \le K_0 \|\Delta x\|. \tag{7.4.6}$$

З (7.4.5) та (7.4.2) маємо

$$f(x_0 + \Delta x) = f(x_0) + \max_{[a,\, v]\, \in \underline{d} F(z_0)} \left\{ a + \sum_{i=1}^{m} v_i [r_i(\Delta x) + o_i(\Delta x)] \right\} +$$

$$+ \min_{[b,\, w] \in \overline{d} F(z_0)} \left\{ b + \sum_{i=1}^{m} w_i [r_i(\Delta x) + o_i(\Delta x)] \right\} + o(\Delta z). \tag{7.4.7}$$

Тут $v = (v_1, \ldots, v_m) \in \mathbb{R}^m$, $\;w = (w_1, \ldots, w_m) \in \mathbb{R}^m$. З (7.4.6) випливає, що в (7.4.7)

$$o(\Delta z) = o(\Delta x). \tag{7.4.8}$$

З (7.4.4), (7.4.6), (7.4.7) та (7.4.8) робимо висновок, що

$$f(x_0+\Delta x)=f(x_0)+\max_{[a,\,\upsilon]\,\in\underline{d}F(z_0)}\left[a+\sum_{i=1}^{m}\upsilon_i r_i(\Delta x)\right]+$$

$$+\min_{[b,\,w]\in\overline{d}F(z_0)}\left[b+\sum_{i=1}^{m}w_i r_i(\Delta x)\right]+o(\Delta x). \qquad (7.4.9)$$

Виберемо додатні числа R_i', R_i'' такі, що

$$-R_i'<\upsilon_i<R_i''\,,\ -R_i'<-w_i<R_i''\ \ \forall i\in\{1,\dots,m\}. \qquad (7.4.10)$$

З (7.4.9) маємо

$$f(x_0+\Delta x)=f(x_0)+\max_{[a,\,\upsilon]\,\in\underline{d}F(z_0)}\left\{a+\sum_{i=1}^{m}(\upsilon_i+R_i')r_i'(\Delta x)+\right.$$

$$\left.+\sum_{i=1}^{m}(\upsilon_i-R_i''\,)r_i''\,(\Delta x)\right\}-\sum_{i=1}^{m}R_i'\,r_i'(\Delta x)+\sum_{i=1}^{m}R_i''\,r_i''\,(\Delta x)+$$

$$+\min_{[b,\,w]\in\overline{d}F(z_0)}\left\{b+\sum_{i=1}^{m}(w_i-R_i')r_i'(\Delta x)+\sum_{i=1}^{m}(w_i+R_i''\,)r_i''\,(\Delta x)\right\}+$$

$$+\sum_{i=1}^{m}R_i'\,r_i'(\Delta x)-\sum_{i=1}^{m}R_i''\,r_i''\,(\Delta x)+o(\Delta x)=$$

$$=f(x_0)+\max_{[a,\,\upsilon]\,\in\underline{d}F(z_0)}\left\{a+\sum_{i=1}^{m}(\upsilon_i+R_i')r_i'(\Delta x)+\sum_{i=1}^{m}(\upsilon_i-R_i''\,)r_i''\,(\Delta x)\right\}+$$

$$+\min_{[b,\,w]\in\overline{d}F(z_0)}\left\{b+\sum_{i=1}^{m}(w_i-R_i')r_i'(\Delta x)+\sum_{i=1}^{m}(w_i+R_i''\,)r_i''\,(\Delta x)\right\}+o(\Delta x). \qquad (7.4.11)$$

Оскільки з (7.4.10) випливає $\upsilon_i+R_i'>0$, $\upsilon_i-R_i''<0$, $\ w_i+R_i''>0$, $\ w_i-R_i'<0$, то з (7.4.11) отримуємо

$$f(x_0+\Delta x)=f(x_0)+$$

$$\max_{[a,\,\upsilon]\,\in\underline{d}F(z_0)}\left\{a+\sum_{i=1}^{m}\left[\max_{[\tilde a_i,\,\tilde\upsilon_i]\,\in\underline{d}y_i(x_0)}[(\upsilon_i+R_i')(\tilde a_i+\langle\tilde\upsilon_i,\Delta x\rangle)]+\right.\right.$$

$$\left.\left.+\max_{[\tilde b_i,\,\tilde w_i]\,\in\overline{d}y_i(x_0)}[(\upsilon_i-R_i''\,)(\tilde b_i+\langle\tilde w_i,\Delta x\rangle)]\right]\right\}+$$

$$+\min_{[b,\,w]\in\overline{d}F(z_0)}\left\{b+\sum_{i=1}^{m}\left[\min_{[\tilde a_i,\,\tilde\upsilon_i]\,\in\underline{d}y_i(x_0)}[(w_i-R_i')(\tilde a_i+\langle\tilde\upsilon_i,\Delta x\rangle)]+\right.\right.$$

$$+\min_{[\tilde{b}_i,\,\tilde{w}_i]\,\in\overline{d}y_i(x_0)}[(w_i+R_i'')(\tilde{b}_i+\langle\tilde{w}_i,\Delta x\rangle)]\Big]\Big\}+o(\Delta x)=$$

$$=f(x_0)+\max_{[a,\,\upsilon]\,\in\underline{d}F(z_0)}\Bigg\{\max_{\substack{[\tilde{a}_i,\tilde{\upsilon}_i]\,\in\,\underline{d}y_i(x_0)\\ [\tilde{b}_i,\tilde{w}_i]\,\in\,\overline{d}y_i(x_0)}}\Bigg[a+\sum_{i=1}^{m}((\upsilon_i+R_i')\tilde{a}_i+(\upsilon_i-R_i'')\tilde{b}_i)+$$

$$+\sum_{i=1}^{m}\langle(\upsilon_i+R_i')\,\tilde{\upsilon}_i+(\upsilon_i-R_i'')\tilde{w}_i,\Delta x\rangle\Bigg]\Bigg\}+$$

$$+\min_{[b,\,w]\in\overline{d}F(z_0)}\Bigg\{\min_{\substack{[\tilde{a}_i,\tilde{\upsilon}_i]\,\in\,\underline{d}y_i(x_0)\\ [\tilde{b}_i,\tilde{w}_i]\,\in\,\overline{d}y_i(x_0)}}\Bigg[b+\sum_{i=1}^{m}((\upsilon_i+R_i')\tilde{a}_i+(w_i+R_i'')\tilde{b}_i)+$$

$$+\sum_{i=1}^{m}\langle(w_i-R_i')\tilde{\upsilon}_i+(w_i+R_i'')\tilde{w}_i,\Delta x\rangle\Bigg]\Bigg\}+o(\Delta x).$$

Звідси

$$f(x_0+\Delta x)=f(x_0)+\max_{[a',\,\upsilon']\,\in\underline{d}f(x_0)}(a'+\langle\upsilon',\Delta x\rangle)+$$
$$+\min_{[b',\,w']\in\overline{d}f(x_0)}(b'+\langle w',\Delta x\rangle)+o(\Delta x), \qquad (7.4.12)$$

де

$$\underline{d}f(x_0)=\Big\{[a',\upsilon']\in\mathbb{R}^{n+1}|a'=a+\sum_{i=1}^{m}((\upsilon_i+R_i')\tilde{a}_i+(\upsilon_i-R_i'')\tilde{b}_i);$$
$$\upsilon'=\sum_{i=1}^{m}((\upsilon_i+R_i')\tilde{\upsilon}_i+(\upsilon_i-R_i'')\tilde{w}_i);\ \ [a,\upsilon]\in\underline{d}F(z_0);$$
$$[\tilde{a}_i,\tilde{\upsilon}_i]\in\underline{d}y_i(x_0);\ \ [\tilde{b}_i,\tilde{w}_i]\in\overline{d}y_i(x_0)\Big\}, \qquad (7.4.13)$$

$$\overline{d}f(x_0)=\Big\{[b',w']\in\mathbb{R}^{n+1}|b'=b+\sum_{i=1}^{m}((w_i-R_i')\tilde{a}_i+(w_i+R_i'')\tilde{b}_i),$$
$$w'=\sum_{i=1}^{m}((w_i-R_i')\tilde{\upsilon}_i+(w_i+R_i'')\tilde{w}_i);\ [b,w]\in\overline{d}F(z_0),$$
$$[\tilde{a}_i,\tilde{\upsilon}_i]\in\underline{d}y_i(x_0),\ [\tilde{b}_i,\tilde{w}_i]\in\overline{d}y_i(x_0)\Big\}. \qquad (7.4.14)$$

Залишилось показати, що множини $\underline{d}f(x_0)$ та $\overline{d}f(x_0)$, які визначені за формулами (7.4.13) та (7.4.14), є опуклими. Покажемо, наприклад, опуклість множини $\underline{d}f(x_0)$. Нехай $[a_1',\upsilon_1']\in\underline{d}f(x_0)$ і $[a_2',\upsilon_2']\in\underline{d}f(x_0)$. Треба показати, що для всіх $\alpha\in[0,1]$ справедливо $[a_\alpha',\upsilon_\alpha']\in\underline{d}f(x_0)$, де

$a'_\alpha = \alpha a'_1 + (1-\alpha)a'_2, \quad \upsilon'_\alpha = \alpha\upsilon'_1 + (1-\alpha)\upsilon'_2$. Маємо

$$a'_1 = a_1 + \sum_{i=1}^{m} \left((\upsilon_{1i} + R'_i)\tilde{a}_{1i} + (\upsilon_{1i} - R''_i)\tilde{b}_{1i}\right),$$
$$\upsilon'_1 = \sum_{i=1}^{m} ((\upsilon_{1i} + R'_i)\tilde{\upsilon}_{1i} + (\upsilon_{1i} - R''_i)\tilde{w}_{1i}),$$

$$a'_2 = a_2 + \sum_{i=1}^{m} \left((\upsilon_{2i} + R'_i)\tilde{a}_{2i} + (\upsilon_{2i} - R''_i)\tilde{b}_{2i}\right),$$
$$\upsilon'_2 = \sum_{i=1}^{m} ((\upsilon_{2i} + R'_i)\tilde{\upsilon}_{2i} + (\upsilon_{2i} - R''_i)\tilde{w}_{2i}),$$

де $[a_j, \upsilon_j] \in \underline{d}F(z_0)$, $[\tilde{a}_{ji}, \tilde{\upsilon}_{ji}] \in \underline{d}y_i(x_0)$, $[\tilde{b}_{ji}, \tilde{w}_{ji}] \in \overline{d}y_i(x_0)$, $j = 1, 2$. Тоді

$$a'_\alpha = \alpha a_1 + (1-\alpha)a_2 + \sum_{i=1}^{m} [\alpha(\upsilon_{1i} + R'_i)\tilde{a}_{1i} + (1-\alpha)(\upsilon_{2i} + R'_i)\tilde{a}_{2i} +$$
$$+\alpha(\upsilon_{1i} - R''_i)\tilde{b}_{1i} + (1-\alpha)(\upsilon_{2i} - R''_i)\tilde{b}_{2i}] =$$
$$= \alpha a_1 + (1-\alpha)a_2 + \sum_{i=1}^{m} [(\alpha(\upsilon_{1i} + R'_i) + (1-\alpha)(\upsilon_{2i} + R'_i))\times$$
$$\left(\frac{\alpha(\upsilon_{1i} + R'_i)}{\alpha(\upsilon_{1i} + R'_i) + (1-\alpha)(\upsilon_{2i} + R'_i)}\tilde{a}_{1i} + \frac{(1-\alpha)(\upsilon_{2i} + R'_i)}{\alpha(\upsilon_{1i} + R'_i) + (1-\alpha)(\upsilon_{2i} + R'_i)}\tilde{a}_{2i}\right) +$$
$$+(\alpha(\upsilon_{1i} - R''_i) + (1-\alpha)(\upsilon_{2i} - R''_i))\times$$
$$\times\left(\frac{\alpha(\upsilon_{1i} - R''_i)}{\alpha(\upsilon_{1i} - R''_i) + (1-\alpha)(\upsilon_{2i} - R''_i)}\tilde{b}_{1i} + \right.$$
$$\left. + \frac{(1-\alpha)(\upsilon_{2i} - R''_i)}{\alpha(\upsilon_{1i} - R''_i) + (1-\alpha)(\upsilon_{2i} - R''_i)}\tilde{b}_{2i}\right)\Bigg]. \qquad (7.4.15)$$

Покладемо

$$c_{1i} = \alpha(\upsilon_{1i} + R'_i) + (1-\alpha)(\upsilon_{2i} + R'_i) = \alpha\upsilon_{1i} + (1-\alpha)\upsilon_{2i} + R'_i,$$
$$\alpha_{1i} = \frac{\alpha(\upsilon_{1i} + R'_i)}{c_{1i}}, \ \beta_{1i} = \frac{(1-\alpha)(\upsilon_{2i} + R'_i)}{c_{1i}},$$
$$c_{2i} = \alpha(\upsilon_{1i} - R''_i) + (1-\alpha)(\upsilon_{2i} - R''_i) = \alpha\upsilon_{1i} + (1-\alpha)\upsilon_{2i} - R''_i,$$
$$\alpha_{2i} = \frac{\alpha(\upsilon_{1i} - R''_i)}{c_{2i}}, \ \beta_{2i} = \frac{(1-\alpha)(\upsilon_{2i} - R''_i)}{c_{2i}}, \ i = 1, \ldots, m.$$

Врахувавши (7.4.10)

$$\begin{array}{l} \alpha_{1i} \geq 0, \ \beta_{1i} \geq 0, \ \alpha_{1i} + \beta_{1i} = 1, \ c_{1i} > 0, \ i = 1, \ldots, m, \\ \alpha_{2i} \geq 0, \ \beta_{2i} \geq 0, \ \alpha_{2i} + \beta_{2i} = 1, \ c_{2i} < 0, \ i = 1, \ldots, m. \end{array} \qquad (7.4.16)$$

Тому з (7.4.15) отримуємо

$$a'_\alpha = a_\alpha + \sum_{i=1}^{m} [(\upsilon_{\alpha i} + R'_i)(\alpha_{1i}\tilde{a}_{1i} + (1-\alpha_{1i})\tilde{a}_{2i}) +$$

$$+(\upsilon_{\alpha i} - R''_i)(\alpha_{2i}\tilde{b}_{1i} + (1-\alpha_{2i})\tilde{b}_{2i})], \qquad (7.4.17)$$

де

$$a_\alpha = \alpha a_1 + (1-\alpha)a_2,\ \upsilon_{\alpha i} = \alpha \upsilon_{1i} + (1-\alpha)\upsilon_{2i},\ i = 1,\ldots,m.$$

Аналогічно отримуємо

$$\upsilon'_\alpha = \sum_{i=1}^{m} [(\upsilon_{\alpha i} + R'_i)(\alpha_{1i}\tilde{\upsilon}_{1i} + (1-\alpha_{1i})\tilde{\upsilon}_{2i}) +$$

$$+(\upsilon_{\alpha i} - R''_i)(\alpha_{2i}\tilde{w}_{1i} + (1-\alpha_{2i})\tilde{w}_{2i}), \qquad (7.4.18)$$

Врахувавши (7.4.16) та опуклість множин $\underline{d}F(z_0)$, $\underline{d}y_i(x_0)$ та $\overline{d}y_i(x_0)$ отримаємо $[a_\alpha, \upsilon_\alpha] \in \underline{d}F(z_0)$, $[\tilde{a}_{\alpha i}, \tilde{\upsilon}_{\alpha i}] \in \underline{d}y_i(x_0)$, $[\tilde{b}_{\alpha i}, \tilde{w}_{\alpha i}] \in \overline{d}y_i(x_0)$. Тут $\upsilon_\alpha = (\upsilon_{\alpha 1}, ..., \upsilon_{\alpha m})$,

$$\tilde{a}_{\alpha i} = \alpha_{1i}\tilde{a}_{1i} + (1-\alpha_{1i})\tilde{a}_{2i},\ \tilde{\upsilon}_{\alpha i} = \alpha_{1i}\tilde{\upsilon}_{1i} + (1-\alpha_{1i})\tilde{\upsilon}_{2i},$$
$$\tilde{b}_{\alpha i} = \alpha_{2i}\tilde{b}_{1i} + (1-\alpha_{2i})\tilde{b}_{2i},\ \tilde{w}_{\alpha i} = \alpha_{2i}\tilde{w}_{1i} + (1-\alpha_{2i})\tilde{w}_{2i}.$$

Звідси, та з (7.4.13), (7.4.17), (7.4.18) робимо висновок, що

$$[a'_\alpha, \upsilon'_\alpha] \in \underline{d}f(x_0) \quad \forall \alpha \in [0,1].$$

Але це означає, що множина $\underline{d}f(x_0)$ є опукла. Аналогічно встановлюється опуклість множини $\overline{d}f(x_0)$.

З (7.4.12) робимо висновок, що функція f є кодиференційовна у точці x_0, а її кодиференціал підраховується за формулами (7.4.13) та (7.4.14).

З (7.4.13), (7.4.14) зрозуміло також, що якщо відображення

$$DF(z) = [\underline{d}F(z), \overline{d}F(z)],\ Dy_i(x) = [\underline{d}y_i(x), \overline{d}y_i(x)],\ i = 1,\ldots,m$$

неперервні (за Хаусдорфом) відповідно у точках z_0 та x_0, то і відображення $Df(x) = [\underline{d}f(x), \overline{d}f(x)]$ також неперервне у точці x_0. □

7.5. НЕПЕРЕРВНО КОДИФЕРЕНЦІЙОВНІ МНОЖИНИ

Нехай множина S визначається співвідношенням

$$S = \{x \in \mathbb{R}^n | h(x) \le 0\}, \qquad (7.5.1)$$

де $h : \mathbb{R}^n \to \mathbb{R}^1$ – кодиференційовна на X функція, $X \subset \mathbb{R}^n$ – відкрита множина, $S \subset X$.

Зафіксуємо $x \in X$. Покладемо

$$\Gamma(x) = x + K(x),\ \ K(x) = \{\Delta \in \mathbb{R}^n | h(x) + h_1(x,\Delta) \le 0\}, \qquad (7.5.2)$$

де
$$h_1(x,\Delta) = h_1(\Delta) = \max_{[a,\,v]\,\in \underline{d}h(z)} [a + \langle v,\Delta\rangle] + \min_{[b,\,w]\,\in \overline{d}h(z)} [b + \langle w,\Delta\rangle]. \quad (7.5.3)$$
Розглянемо також множини
$$\bar{\gamma}_1(x) = \{\Delta \in \mathbb{R}^n | h(x) + h_1(x,\Delta) < 0\}, \quad (7.5.4)$$
$$\bar{\Gamma}_1(x) = \{\Delta \in \mathbb{R}^n | h(x) + h_1(x,\Delta) \le 0\}. \quad (7.5.5)$$
Раніше були визначені множини
$$\gamma_1(x) = \{\Delta \in \mathbb{R}^n | h'(x,\Delta) < 0\}, \quad (7.5.6)$$
$$\Gamma_1(x) = \{\Delta \in \mathbb{R}^n | h'(x,\Delta) \le 0\}, \quad (7.5.7)$$
де
$$h'(x,\Delta) = \max_{v\,\in \underline{\partial}h(z)} \langle v,\Delta\rangle + \min_{w\,\in \overline{\partial}h(z)} \langle w,\Delta\rangle,$$
$$\mathfrak{D}h(x) = [\underline{\partial}h(x), \overline{\partial}h(x)]$$
– квазідиференціал функції h в точці x.

Означення 7.5.1. Говорять, що в точці x виконана умова *регулярності*, якщо $\mathrm{cl}\ \gamma_1(x) = \Gamma(x)$.

Лема 7.5.1. *Нехай точка x така, що $h(x) = 0$. Якщо в точці x виконана умова регулярності, то визначена співвідношенням (7.5.2) множина $\Gamma(x)$ є апроксимацією першого порядку множини S в околі точки x.*

Доведення. За умовою регулярності
$$\mathrm{cl}\ \gamma_1(x) = \Gamma_1(x). \quad (7.5.8)$$
Оскільки для $\Delta_0 = 0$ виконується $h_1'(\Delta_0,g) = h'(\Delta_0,g)$, то за теоремою 4.4.5 множина $\Gamma_1(x)$ є апроксимацією першого порядку множини $K(x)$ в околі точки $\Delta_0 = 0$. Тоді множина $\Gamma(x) = x + K(x)$ є апроксимацією першого порядку множини S в околі точки x. □

Означення 7.5.2. В точці x виконана умова *корегулярності*, якщо
$$\mathrm{cl}\ \overline{\gamma}_1(x) = \overline{\Gamma}_1(x). \quad (7.5.9)$$

Лема 7.5.2. *Нехай $x \in X$. Якщо функція h неперервно кодиференційовна в точці x і виконана умова корегулярності, то відображення K неперервне за Какутані в точці x.*

Доведення. Встановимо спочатку напівнеперервність зверху. Нехай $x_k \to x, \Delta_k \to \Delta, \Delta_k \in K(x_k)$, тобто
$$h(x_k) + h_1(x_k,\, \Delta_k) \le 0. \quad (7.5.10)$$

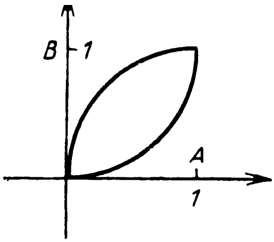


Рис. 7.5.1:

Оскільки відображення Dh неперервне за Хаусдорфом, то (див. (7.5.3)) функція h_1 неперервна за сукупністю x, Δ. Тому з (7.5.10) випливає $h(x) + h_1(x, \Delta) \leq 0$, тобто відображення K напівнеперервне зверху в точці x. Відмітимо, що при доведенні напівнеперервності зверху умова (7.5.9) не використовувалась.

Візьмемо $\Delta \in K(x)$. Виберемо довільну послідовність $\{x_k\} : x_k \to x$. Якщо $h(x)+h_1(x, \Delta) < 0$, то в силу неперервності h та Dh при достатньо великих k отримаємо $h(x_k) + h_1(x_k, \Delta) \leq 0$, тобто $\Delta \in K(x_k)$.
Якщо ж $h(x) + h_1(x, \Delta) = 0$, то в силу умови корегулярності (7.5.9) знайдеться послідовність $\{\Delta^{(s)}\}$ така, що $\Delta^{(s)} \to 0, \Delta^{(s)} \in \overline{\gamma}_1(x)$. Візьмемо $\Delta^{(1)}$. Оскільки $h(x) + h_1(x, \Delta^{(1)}) < 0$, то знайдеться k_1 таке, що $h(x_k) + h_1(x_k, \Delta^{(1)}) < 0 \ \forall k > k_1$. Тепер візьмемо $\Delta^{(2)}$. Знайдеться $k_2 > k_1$ таке, що $h(x) + h_1(x, \Delta^{(2)}) < 0$. Далі робимо аналогічно: нехай вже визначено k_l. Візьмемо $\Delta^{(l+1)}$. Знайдеться $k_{l+1} > k_l$ таке, що $h(x) + h_1(x, \Delta^{(l+1)}) < 0$. Покладемо

$$\Delta_k = \begin{cases} \Delta^{(0)}, & k < k_1, \\ \Delta^{(l)}, & k \in [k_l; k_{l+1} - 1]. \end{cases}$$

Очевидно, що $\Delta_k \in K(x_k)$, $\Delta_k \to \Delta$. Це означає, що відображення K напівнеперервне зверху в точці x. $\square$

Зауваження 7.5.1. Відображення K визначене і для $x \notin S$.

Приклад 7.5.1. Нехай $x \in \mathbb{R}^2, A = (1, 0), B = (0,1)$, $h_1(x) = (x - A)^2 - 1, h_2(x) = (x - B)^2 - 1, h(x) = \max\{h_1(x), h_2(x)\}, S = \{x \in \mathbb{R}^2 | h(x) \leq 0\}$ (рис. 7.5.1).
Оскільки h_1, h_2 – гладкі функції, то їх кодиференціали дорівнюють

$$Dh_1(x) = [\underline{d}h_1(x), \overline{d}h_1(x)], \ \ Dh_2(x) = [\underline{d}h_2(x), \overline{d}h_2(x)],$$

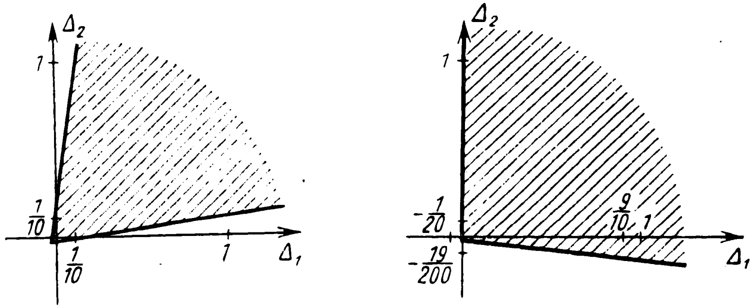


Рис. 7.5.2:

де

$$\begin{array}{ll} \underline{d}h_1(x) = \{[0,\, h_1'(x)]\}, & \overline{d}h_1(x) = \{0_3\}, \\ \underline{d}h_2(x) = \{[0,\, h_2'(x)]\}, & \overline{d}h_2(x) = \{0_3\}. \end{array}$$

Звідси

$$\underline{d}h(x) = \operatorname{conv}\left\{\underline{d}h_1(x) - \overline{d}h_2(x) + \{[h_1(x) - h(x), 0_2]\},\right.$$

$$\left.\underline{d}h_2(x) - \overline{d}h_1(x) + \{[h_2(x) - h(x), 0_2]\}\right\};$$

$$\overline{d}h(x) = \overline{d}h_1(x) + \overline{d}h_2(x) = \{0_3\},$$

тобто

$$\underline{d}h(x) = \operatorname{conv}\{[h_1(x) - h(x); 2(x - A)], [h_2(x) - h(x); 2(x - B)]\},$$

$$K(x) = \{\Delta \in \mathbb{R}^2 |\max\{h_1(x) + 2(x - A, \Delta), h_2(x) + 2(x - B,\Delta)\} \le 0\}.$$

При $x_0 = (1/10{,}1/10)$ маємо $h_1(x) = -18/100$, $h_2(x) = -18/100$. Очевидно, що $x_0 \in S$, $K(x_0) =$

$$= \left\{\Delta \in \mathbb{R}^2 |\max\left\{-\frac{18}{100} - \frac{18}{10}\Delta_1 + \frac{2}{10}\Delta_2; -\frac{18}{100} + \frac{2}{10}\Delta_1 - \frac{18}{10}\Delta_2\right\} \le 0\right\}.$$

На рис. 7.5.2а зображена множина $\Gamma(x_0) = x_0 + K(x_0)$. При $\overline{x}_0 = (-1/10, 0)$ маємо $h_1(\overline{x}_0) = 21/100$, $h_2(\overline{x}_0) = 1/100$. Очевидно, що $\overline{x}_0 \notin S$,

$$K(\overline{x}_0) = \left\{\Delta \in \mathbb{R}^2 | \frac{21}{100} - \frac{21}{10}\Delta_1 \le 0,\ \frac{1}{100} - \frac{2}{10}\Delta_1 - 2\Delta_2 \le 0\right\}.$$

На рис. 7.5.2б зображена множина $\Gamma(\overline{x}_0) = \overline{x}_0 + K(\overline{x}_0)$.

7.6. ДВІЧІ КОДИФЕРЕНЦІЙОВНІ ФУНКЦІЇ.

Апроксимації другого порядку диференційовних функцій будуються за допомогою перших і других похідних. Апроксимації недиференційовних функцій базуються на іншому підході.

Нехай функція f задана і скінченна на відкритій множині $X \subset \mathbb{R}^n$.

Означення 7.6.1. Функція f *двічі кодиференційовна* в точці $x \in X$, якщо існують опуклі компакти $\underline{d}^2 f(x)$, $\overline{d}^2 f(x) \subset \mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^{n\times n}$ такі, що

$$f(x+\Delta) = f(x) + \max_{[a,v,A]\in\underline{d}^2 f(x)} [a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle] +$$

$$+ \min_{[b,w,B]\in\overline{d}^2 f(x)} [b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle] + o(\Delta^2), \qquad (7.6.1)$$

де

$$\frac{o((\alpha\Delta)^2)}{\alpha^2} \underset{\alpha\downarrow 0}{\longrightarrow} 0.$$

Тут $\mathbb{R}^{n\times n}$ - простір дійсних $(n\times n)$ – матриць, відрізок $\operatorname{conv}\{x,x+\Delta\} \subset X$. Пара множин $D^2 f(x) = [\underline{d}^2 f(x),\overline{d}^2 f(x)]$ називається *другим кодиференціалом функції f в точці x*. Множина $\underline{d}^2 f(x)$ називається *другим гіподиференціалом*, а множина $\overline{d}^2 f(x)$ називається *другим гіпердиференціалом* функції f в точці x. Якщо функція f двічі кодиференційовна в деякому околі точки x і відображення $D^2 f$ неперервне в метриці Хаусдорфа в точці x, то функція f називається *двічі неперервно кодиференційовною в точці x*.

Очевидно, що другий кодиференціал визначається неоднозначно.

Якщо серед других кодиференціалів функції f в точці x є кодиференціал вигляду $D^2 f(x) = [\underline{d}^2 f(x),\{\mathrm{O}\}]$, то функція f називається *двічі гіподиференційовною*.

Тут $\mathrm{O} = [0,0_n,0_{n\times n}] \in \mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^{n\times n}$.

Якщо серед других кодиференціалів функції f в точці x є другий кодиференціал вигляду $D^2 f(x) = [\{\mathrm{O}\},\overline{d}^2 f(x)]$, то функція f називається *двічі гіпердиференційовною*.

Зауваження 7.6.1. Можна переконатися, що двічі кодиференційовна функція (двічі неперервно кодиференційовна функція) є кодиференційовною (неперервно кодиференційовною). Дійсно, з (7.6.1) видно, що

$$f(x+\Delta) = f(x) + \max_{[a,v]\in\underline{d}f(x)} [a + (v,\Delta)] + \min_{[b,w]\in\overline{d}f(x)} [b + (w,\Delta)] + o(\Delta),$$

$$\underline{d}f(x) = \{[a,v] |\ \exists A \in \mathbb{R}^{n\times n} : [a,v,A] \in \underline{d}^2 f(x)\},$$

$$\overline{d}f(x) = \{[b,w] |\ \exists B \in \mathbb{R}^{n\times n} : [b,w,B] \in \overline{d}^2 f(x)\}.$$

7.6.1. Приклади двічі кодиференційовних функцій

Приклад 7.6.1. Нехай f - опукла скінченна на X функція, $x \in X$ і нехай X_0 – довільна замкнута обмежена підмножина X така, що $x \in \operatorname{int} X_0$. Оскільки

$$f(x+\Delta) = f(x) + \max_{[a,v]\in \underline{d}f(x)} [a + \langle v,\Delta\rangle], \tag{7.6.2}$$

де

$$\underline{d}f(x) = \operatorname{conv}\{[a,v] | a = f(z) + \langle v,x-z\rangle, v(z) \in \underline{\partial} f(z), z \in X_0\},$$

то (7.6.2) можна переписати у вигляді

$$f(x+\Delta) = f(x) + \max_{[a,v,A]\in \underline{d}^2 f(x)} [a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle], \tag{7.6.3}$$

де

$$\underline{d}^2 f(x) = \operatorname{conv}\{[a,v,A] | a = f(z) + \langle v,x-z\rangle, v(z) \in \underline{\partial} f(z), A = 0_{n\times n}, z \in X_0\}.$$

Тут $v(z) \in \underline{\partial} f(z)$ - довільне, але фіксоване для кожного $z \in X_0$, $\underline{\partial} f(z)$ – субдиференціал функції f в точці z.

Отже, кожна опукла функція є двічі неперервно кодиференційовною (і навіть двічі неперервно гіподиференційовною), причому

$$D^2 f(x) = [\underline{d}^2 f(x), \{\mathrm{O}\}].$$

Приклад 7.6.2. Кожна угнута функція є двічі неперервно кодиференційовною (і навіть двічі неперервно гіпердиференційовною).

Приклад 7.6.3. Будь-яка кодиференційовна в точці x функція f така, що

$$f(x+\Delta) = f(x) + \max_{[a,v]\in \underline{d}f(x)} [a + \langle v,\Delta\rangle] + \min_{[b,w]\in \overline{d}f(x)} [b + \langle w,\Delta\rangle],$$

(тобто в (7.1.1) $o_x(\Delta) = 0$) є двічі кодиференційовною, причому

$$D^2 f(x) = [\underline{d}^2 f(x), \overline{d}^2 f(x)],$$

де

$$\underline{d}^2 f(x) = \{[a,v,A] \colon [a,v] \in \underline{d}f(x), A = 0_{n\times n}\},$$

$$\overline{d}^2 f(x) = \{[b,w,B] \colon [b,w] \in \overline{d}f(x), B = 0_{n\times n}\}.$$

Зокрема, функція $f(x) = \|x\| = \sqrt{(x,x)}$ є двічі неперервно кодиференційовною. При цьому (див. (7.1.10))

$$D^2 f(x) = [\underline{d}^2 f(x), \overline{d}^2 f(x)],$$

де

$$\underline{d}^2 f(x) = \{[a,v,A] \colon a = \langle x,v\rangle - f(x), \langle v,v\rangle \le 1, A = 0_{n\times n}\},$$

$$\overline{d}^2 f(x) = \{[0,0_n,0_{n\times n}]\}.$$

Приклад 7.6.4. Нехай f — двічі неперервно диференційовна функція, тобто

$$f(x+\Delta) = f(x) + \langle f'(x),\Delta\rangle + \frac{1}{2}\langle f''(x)\Delta,\Delta\rangle + o(\Delta^2).$$

Тут $f''(x)$ - матриця других похідних функції f в точці x. Очевидно, що f – двічі неперервно кодиференційовна функція, причому як $D^2 f$ можна взяти пару множин

$$D^2 f(x) = [\underline{d}^2 f(x),\overline{d}^2 f(x)],$$

де

$$\underline{d}^2 f(x) = \{[0,f'(x),f''(x)]\}, \quad \overline{d}^2 f(x) = \{[0,0_n,0_{n\times n}]\}, \tag{7.6.4}$$

а також пару множин

$$D^2 f(x) = [\underline{d}^2 f(x),\overline{d}^2 f(x)],$$

де

$$\underline{d}^2 f(x) = \{[0,0_n,0_{n\times n}]\}, \quad \overline{d}^2 f(x) = \{[0,f'(x),f''(x)]\}. \tag{7.6.5}$$

Отже, двічі неперервно диференційовна функція є і двічі гіподиференційовною, і двічі гіпердиференційовною.

Приклад 7.6.5. Нехай

$$f(x) = \max_{y\in G} \varphi(x,y), \tag{7.6.6}$$

де G - компактна множина, функція φ неперервна на $X \times G$ та двічі неперервно диференційовна за x на X при кожному фіксованому $y \in G$. Тоді

$$f(x+\Delta) = \max_{y\in G}[\varphi(x,y)+\langle\varphi'_x(x,y),\Delta\rangle+\frac{1}{2}\langle\varphi''_{xx}(x,y)\Delta,\Delta\rangle+o_x(\Delta^2,y)], \tag{7.6.7}$$

де $\varphi'_x(x,y)$ - градієнт функції φ відносно x, а $\varphi''_{xx}(x,y)$ - матриця других похідних відносно x при фіксованому y,

$$\frac{o_x(\Delta^2,y)}{\Delta^2} \underset{\|\Delta\|\to 0}{\longrightarrow} 0. \tag{7.6.8}$$

Якщо умова (7.6.8) виконується рівномірно відносно $y \in G$, то

$$f(x+\Delta) = f(x) + \max_{[a,v,A]\in\underline{d}^2 f(x)}[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle] + o_x(\Delta^2), \tag{7.6.9}$$

де

$$\underline{d}^2 f(x) = \operatorname{conv}\{[a,v,A]\colon a = \varphi(x,y) - f(x), v = \varphi'_x(x,y), A = \varphi''_{xx}(x,y), y \in G\}, \tag{7.6.10}$$

$$\frac{o_x((\alpha\Delta)^2)}{\alpha^2} \underset{\alpha\downarrow 0}{\longrightarrow} 0$$

рівномірно за $\Delta \in \mathrm{B} = \{\Delta \in \mathbb{R}^n\colon \|\Delta\| \leq 1\}$. Отже функція (7.6.6) є двічі неперервно кодиференційовною (навіть гіподиференційовною).

Приклад 7.6.6. Аналогічно показується, що функція

$$f(x) = \min_{y \in G} \varphi(x,y), \tag{7.6.11}$$

де φ і G - такі ж, як і в попередньому прикладі, є двічі неперервно кодиференційовною, причому як $D^2 f(x)$ можна взяти пару множин

$$D^2 f(x) = [\underline{d}^2 f(x), \overline{d}^2 f(x)],$$

де

$$\underline{d}^2 f(x) = \{[0, 0_n, 0_{n \times n}]\},$$

$$\overline{d}^2 f(x) = \mathrm{conv}\{[b,w,B] \colon b = \varphi(x,y) - f(x), w = \varphi'_x(x,y), B = \varphi''_{xx}(x,y), y \in G\}.$$

Отже функція мінімуму (7.6.11) двічі неперервно гіпердиференційовна.

7.6.2. Формули обчислення кодиференціалів другого порядку

Покажемо, що основні "гладкі" операції зберігають властивість двічі кодиференційовності. Нехай

$$D_1^2 = [\underline{C}_1, \overline{C}_1], \quad D_2^2 = [\underline{C}_2, \overline{C}_2],$$

$$\underline{C}_i, \overline{C}_i \subset \mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^{n \times n}, \quad i = 1,2.$$

Визначимо, як звичайно, суму пар множин і добуток на дійсне число $D_1^2 + D_2^2 = [\underline{C}, \overline{C}]$, де $\underline{C} = \underline{C}_1 + \underline{C}_2$, $\overline{C} = \overline{C}_1 + \overline{C}_2$,

$$\lambda D_1^2 = \begin{array}{l} [\lambda \underline{C}_1, \lambda \overline{C}_1], \text{якщо } \lambda \geq 0, \\ [\lambda \overline{C}_1, \lambda \underline{C}_1], \text{якщо } \lambda < 0. \end{array}$$

Лема 7.6.1. *Якщо функції $f_i(x), i \in I = \{1, \ldots, N\}$, двічі неперервно кодиференційовні в точці $x \in X$, то функція $f(x) = \sum_{i \in I} c_i f_i(x)$, де $c_i \in \mathbb{R}^1$, теж двічі неперервно кодиференційовна в точці x, причому*

$$D^2 f(x) = \sum_{i \in I} c_i D^2 f_i(x). \tag{7.6.12}$$

Лема 7.6.2. *Якщо функції $f_1(x)$ та $f_2(x)$ двічі кодиференційовні (двічі неперервно кодиференційовні) в точці $x \in X$, то і функція $f(x) = f_1(x) \cdot f_2(x)$ є двічі кодиференційовна (двічі неперервно кодиференційовна) в точці x.*

Доведення. Маємо

$$f_1(x + \Delta) = f_1(x) + r'_1(\Delta) + r''_1(\Delta) + o_1(\Delta^2), \tag{7.6.13}$$

$$f_2(x + \Delta) = f_2(x) + r'_2(\Delta) + r''_2(\Delta) + o_2(\Delta^2), \tag{7.6.14}$$

де для $i = 1{,}2$

$$r_i'(\Delta) \equiv r_i'(x,\Delta) = \max_{[a_i,v_i,A_i]\in \underline{d}^2 f_i(x)} [a_i + \langle v_i,\Delta\rangle + \frac{1}{2}\langle A_i\Delta,\Delta\rangle], \qquad (7.6.15)$$

$$r_i''(\Delta) \equiv r_i''(x,\Delta) = \min_{[b_i,w_i,B_i]\in \overline{d}^2 f_i(x)} [b_i + \langle w_i,\Delta\rangle + \frac{1}{2}\langle B_i\Delta,\Delta\rangle]. \qquad (7.6.16)$$

Покладемо $r_i(\Delta) = r_i'(\Delta) + r_i''(\Delta)$, $i = 1{,}2$. Тоді

$$f(x+\Delta) = f_1(x+\Delta)f_2(x+\Delta) =$$

$$f_1(x)f_2(x) + f_1(x)r_2(\Delta) + f_2(x)r_1(\Delta) + r_1(\Delta)r_2(\Delta) + o(\Delta^2). \qquad (7.6.17)$$

Обчислення в (7.6.17) усіх доданків, крім $r_1(\Delta)r_2(\Delta)$, не становить труднощів. Знайдемо $r_1(\Delta)r_2(\Delta)$. Маємо

$$r_1(\Delta)r_2(\Delta) = r_1'(\Delta)r_2'(\Delta) + r_1'(\Delta)r_2''(\Delta) + r_1''(\Delta)r_2'(\Delta) + r_1''(\Delta)r_2''(\Delta). \qquad (7.6.18)$$

Для довільного $\delta > 0$ знайдеться таке $R < \infty$, що для $i = 1{,}2$

$$|\gamma_i'| \equiv \left| a_i + \langle v_i,\Delta\rangle + \frac{1}{2}\langle A_i\Delta,\Delta\rangle \right| < R,\ [a_i,v_i,A_i] \in \underline{d}^2 f_i(x), \Delta \in B_\delta, \qquad (7.6.19)$$

$$|\gamma_i''| \equiv \left| b_i + \langle w_i,\Delta\rangle + \frac{1}{2}\langle B_i\Delta,\Delta\rangle \right| < R,\ [b_i,w_i,B_i] \in \overline{d}^2 f_i(x), \Delta \in B_\delta. \qquad (7.6.20)$$

Зафіксуємо деяке $\delta > 0$. Оскільки

$$r_1'(\Delta)r_2'(\Delta) = \max\gamma_1' \cdot \max\gamma_2' =$$

$$\max_{[a_1,v_1,A_1]\in \underline{d}^2 f_1(x)} [a_1 + \langle v_1,\Delta\rangle + \frac{1}{2}\langle A_1\Delta,\Delta\rangle]\times$$

$$\times \max_{[a_2,v_2,A_2]\in \overline{d}^2 f_2(x)} [a_2 + \langle v_2,\Delta\rangle + \frac{1}{2}\langle A_2\Delta,\Delta\rangle] =$$

$$= (\max(\gamma_1' + R) - R)(\max(\gamma_2' + R) - R) =$$

$$= \max(\gamma_1' + R)\max(\gamma_2' + R) - R\max(\gamma_1' + R) - R\max(\gamma_2' + R) + R^2,$$

то, враховуючи, що $\gamma_1' + R > 0$, $\gamma_2' + R > 0$ (див. (7.6.19), (7.6.20)), одержимо

$$r_1'(\Delta)r_2'(\Delta) = \max_{\gamma_1',\gamma_2'}[\gamma_1' \cdot \gamma_2' + R(\gamma_1' + \gamma_2')] + \min_{\gamma_1',\gamma_2'}[-R(\gamma_1' + \gamma_2')],$$

тобто

$$r_1'(\Delta)r_2'(\Delta) = \max_{\substack{[a_1,v_1,A_1] \in \underline{d}^2 f_1(x) \\ [a_2,v_2,A_2] \in \underline{d}^2 f_2(x)}} \left[\left(a_1 + \langle v_1,\Delta\rangle + \frac{1}{2}\langle A_1\Delta,\Delta\rangle\right) \times\right.$$

$$\times\left(a_2 + \langle v_2,\Delta\rangle + \frac{1}{2}\langle A_2\Delta,\Delta\rangle\right) +$$

$$+ R\left(a_1 + \langle v_1,\Delta\rangle + \frac{1}{2}\langle A_1\Delta,\Delta\rangle + a_2 + \langle v_2,\Delta\rangle + \frac{1}{2}\langle A_2\Delta,\Delta\rangle\right)\Bigg] +$$

$$+ \min_{\substack{[a_1,v_1,A_1] \in \underline{d}^2 f_1(x)\\ [a_2,v_2,A_2] \in \underline{d}^2 f_2(x)}} \Bigg[- R\Bigg(a_1 + \langle v_1,\Delta\rangle + \frac{1}{2}\langle A_1\Delta,\Delta\rangle +$$

$$a_2 + \langle v_2,\Delta\rangle + \frac{1}{2}\langle A_2\Delta,\Delta\rangle\Bigg)\Bigg] =$$

$$= \max_{\substack{[a_1,v_1,A_1] \in \underline{d}^2 f_1(x)\\ [a_2,v_2,A_2] \in \underline{d}^2 f_2(x)}} \Bigg[a_1a_2 + R(a_1+a_2) + \langle a_2v_1 + a_1v_2 + Rv_1 + Rv_2,\Delta\rangle +$$

$$+\left\langle\left(\frac{1}{2}a_1A_2 + \frac{1}{2}a_2A_1 + v_1v_2 + \frac{1}{2}RA_1 + \frac{1}{2}RA_2\right)\Delta,\Delta\right\rangle\Bigg] + o(\Delta^2) +$$

$$\min_{\substack{[a_1,v_1,A_1] \in \underline{d}^2 f_1(x)\\ [a_2,v_2,A_2] \in \underline{d}^2 f_2(x)}} \left[-R(a_1+a_2) - R\langle v_1+v_2,\Delta\rangle - \frac{1}{2}\langle R(A_1+A_2)\Delta,\Delta\rangle\right].$$

Отже, остаточно отримуємо:

$$r_1'(\Delta)r_2'(\Delta) = \max_{[a,v,A]\in\underline{d}^2(r_1',r_2')} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle\right] +$$

$$+ \min_{[b,w,B]\in\overline{d}^2(r_1',r_2')} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle\right] + o(\Delta^2), \quad \Delta \in B_\delta, \qquad (7.6.21)$$

де

$$\underline{d}^2(r_1',r_2') = \Big\{[a,v,A] | a = a_1a_2 + R(a_1+a_2),$$

$$v = a_2v_1 + a_1v_2 + Rv_1 + Rv_2,$$

$$A = (a_1+R)A_2 + (a_2+R)A_1 + 2v_1v_2,$$

$$[a_1,v_1,A_1] \in \underline{d}^2 f_1(x), \quad [a_2,v_2,A_2] \in \underline{d}^2 f_2(x)\Big\},$$

$$\overline{d}^2(r_1',r_2') = \Big\{[b,w,B] | b = -R(a_1+a_2),$$

$$w = -R(v_1+v_2), \quad B = -R(A_1+A_2),$$

$$[a_1,v_1,A_1] \in \underline{d}^2 f_1(x), \quad [a_2,v_2,A_2] \in \underline{d}^2 f_2(x)\Big\}.$$

Аналогічно,

$$r_1'(\Delta)r_2''(\Delta) = \min[\gamma_1'\gamma_2'' + R(\gamma_2'' - \gamma_1')] + \max[R(\gamma_1' - \gamma_2'')] =$$

$$= \min_{\substack{[a_1,v_1,A_1] \in \underline{d}^2 f_1(x) \\ [b_2,w_2,A_2] \in \overline{d}^2 f_2(x)}} \left[\left(a_1 + \langle v_1,\Delta\rangle + \frac{1}{2}\langle A_1\Delta,\Delta\rangle\right) \times\right.$$

$$\left(b_2 + \langle w_2,\Delta\rangle + \frac{1}{2}\langle B_2\Delta,\Delta\rangle\right) +$$

$$\left. +R\left(b_2 + \langle w_2,\Delta\rangle + \frac{1}{2}\langle B_2\Delta,\Delta\rangle - a_1 - \langle v_1,\Delta\rangle - \frac{1}{2}\langle A_1\Delta,\Delta\rangle\right)\right] +$$

$$+ \max_{\substack{[a_1,v_1,A_1] \in \underline{d}^2 f_1(x) \\ [d_2,w_2,A_2] \in \overline{d}^2 f_2(x)}} \left[R\left(a_1 + \langle v_1,\Delta\rangle + \frac{1}{2}\langle A_1\Delta,\Delta\rangle -\right.\right.$$

$$\left.\left. -b_2 - \langle w_2,\Delta\rangle - \frac{1}{2}\langle B_2\Delta,\Delta\rangle\right)\right] =$$

$$= \max\left[R(a_1 - b_2) + \langle R(v_1 - w_2),\Delta\rangle + \frac{1}{2}\langle R(A_1 - B_2)\Delta,\Delta\rangle\right] +$$

$$+ \min\left[a_1b_2 - R(a_1 - b_2) + \langle a_1w_2 + b_2v_1 + R(w_2 - v_1),\Delta\rangle +\right.$$

$$\left. + \frac{1}{2}\langle(b_2A_1 + a_1B_2 + 2v_1w_2 + RB_2 - RA_1)\Delta,\Delta\rangle\right] + o(\Delta^2).$$

Отже,

$$r_1'(\Delta)r_2''(\Delta) = \max_{[a,v,A]\in\underline{d}^2(r_1',r_2'')} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle\right] +$$

$$\min_{[b,w,B]\in\overline{d}^2(r_1',r_2'')} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle\right] + o(\Delta^2), \tag{7.6.22}$$

де

$$\underline{d}^2(r_1',r_2'') = \Big\{[a,v,A] | a = R(a_1 - b_2), v = R(v_1 - w_2), A = R(A_1 - B_2),$$

$$[a_1,v_1,A_1] \in \underline{d}^2 f_1(x), \quad [b_2,w_2,B_2] \in \overline{d}^2 f_2(x)\Big\},$$

$$\overline{d}^2(r_1',r_2'') = \Big\{[b,w,B]\colon b = a_1b_2 - R(a_1 - b_2),$$

$$w = a_1 w_2 + b_2 v_1 + R(w_2 - v_1), \quad B = (b_2 - R)A_1 + (a_1 + R)B_2 + 2v_1 w_2,$$

$$\left.[a_1,v_1,A_1] \in \underline{d}^2 f_1(x), \quad [b_2,w_2,B_2] \in \overline{d}^2 f_2(x)\right\}.$$

Подібним чином одержуємо:

$$r_1''(\Delta)r_2'(\Delta) = \max_{[a,v,A]\in\underline{d}^2(r_1'',r_2')} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle\right] +$$

$$+ \min_{[b,w,B]\in\overline{d}^2(r_1'',r_2')} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle\right] + o(\Delta^2), \tag{7.6.23}$$

де

$$\underline{d}^2(r_1'',r_2') = \Big\{[a,v,A] | a = R(a_2 - b_1), v = R(v_2 - w_1), A = R(A_2 - B_1),$$

$$\left.[a_2,v_2,A_2] \in \underline{d}^2 f_2(x), \quad [b_1,w_1,B_1] \in \overline{d}^2 f_1(x)\right\},$$

$$\overline{d}^2(r_1'',r_2') = \Big\{[b,w,B]\colon b = a_2 b_1 - R(a_2 - b_1), w = a_2 w_1 + b_1 v_2 + R(w_2 - v_1),$$

$$\left.B = (b_1 - R)A_2 + (a_2 + R)B_1 + 2v_2 w_1, [a_2,v_2,A_2] \in \underline{d}^2 f_2(x), [b_1,w_1,B_1] \in \overline{d}^2 f_1(x)\right\}.$$

Нарешті,

$$r_1''(\Delta)r_2''(\Delta) = \min \gamma_1'' \ \min \gamma_2'' =$$

$$= \max[\gamma_1''\gamma_2'' - R(\gamma_1'' + \gamma_2'')] + \min[R(\gamma_1'' + \gamma_2'')] =$$

$$= \max_{[a,v,A]\in\underline{d}^2(r_1'',r_2'')} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle\right] +$$

$$\min_{[b,w,B]\in\overline{d}^2(r_1'',r_2'')} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle\right] + o(\Delta^2), \tag{7.6.24}$$

де

$$\underline{d}^2(r_1'',r_2'') = \Big\{[a,v,A]\colon a = b_1 b_2 - R(b_1 + b_2),$$

$$v = (b_2 - R)w_1 + (b_1 - R)w_2,$$

$$A = (b_2 - R)B_1 + (b_1 - R)B_2 + 2w_1 w_2,$$

$$\left.[b_1,w_1,B_1] \in \overline{d}^2 f_1(x), \quad [b_2,w_2,B_2] \in \overline{d}^2 f_2(x)\right\},$$

$$\overline{d}^2(r_1'',r_2'') = \Big\{[b,w,B]\colon b = R(b_1 + b_2),$$

$$w = R(w_1 + w_2), B = R(B_1 + B_2),$$

$$\left.[b_1, w_1, B_1] \in \overline{d}^2 f_1(x), \quad [b_2, w_2, B_2] \in \overline{d}^2 f_2(x)\right\}.$$

З (7.6.17) – (7.6.24) остаточно робимо висновок, що при $\Delta \in B_\delta$ має місце представлення

$$f(x+\Delta) = f(x) + \max_{[a,v,A]\in\underline{d}^2 f(x)} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle\right] +$$

$$\min_{[b,w,B]\in\overline{d}^2 f(x)} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle\right] + o(\Delta^2), \tag{7.6.25}$$

де

$$D^2 f(x) = f_1(x) D^2 f_2(x) + f_2(x) D^2 f_1(x) + D^2(r_1, r_2), \tag{7.6.26}$$

$$D^2[r_1, r_2] = [\underline{d}^2(r_1, r_2), \overline{d}^2(r_1, r_2)]$$

$$\underline{d}^2(r_1, r_2) = \underline{d}^2(r_1', r_2') + \underline{d}^2(r_1', r_2'') + \underline{d}^2(r_1'', r_2') + \underline{d}^2(r_1'', r_2''),$$

$$\overline{d}^2(r_1, r_2) = \overline{d}^2(r_1', r_2') + \overline{d}^2(r_1', r_2'') + \overline{d}^2(r_1'', r_2') + \overline{d}^2(r_1'', r_2'').$$

З (7.6.25) – (7.6.26) ясно, що якщо $f_1(x), f_2(x)$ - двічі неперервно кодиференційовні функції, то функція $f(x) = f_1(x) f_2(x)$ двічі неперервно кодиференційовна в точці x. □

Лема 7.6.3. *Нехай $f_1(x)$– двічі кодиференційовна (двічі неперервно кодиференційовна) в точці x функція і $f_1(x) \neq 0$. Тоді функція $f(x) = \frac{1}{f_1(x)}$ двічі кодиференційовна (двічі неперервно кодиференційовна) в точці x.*

Доведення. Маємо

$$f_1(x+\Delta) = f_1(x) + r(\Delta) + o(\Delta^2), \quad r(\Delta) = r' + r'',$$

де

$$r' = \max_{[a,v,A]\in\underline{d}^2 f_1(x)} [a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle],$$

$$r'' = \min_{[b,w,B]\in\overline{d}^2 f_1(x)} [b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle].$$

Оскільки

$$\frac{1}{z+\alpha} = \frac{1}{z} - \frac{1}{z^2}\alpha + \frac{2}{z^3}\alpha^2 + o(\alpha^2),$$

то

$$f(x+\Delta) = \frac{1}{f_1(x+\Delta)} = f(x) - \frac{r(\Delta)}{(f_1(x))^2} + \frac{2r^2(\Delta)}{(f_1(x))^3} + o(\Delta^2). \tag{7.6.27}$$

Тут враховано те, що (див. доведення леми 7.2.2.) $|r(\Delta)| \le L \|\Delta\|$. Міркуючи, як при доведенні леми 7.6.2, з (7.6.27) одержуємо для $\Delta \in B_\delta$:

$$f(x+\Delta) = f(x) + \max_{[a,v,A]\in \underline{d}^2 f(x)} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle \right] +$$

$$\min_{[b,w,B]\in \overline{d}^2 f(x)} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle \right] + o(\Delta^2), \tag{7.6.28}$$

де

$$D^2 f(x) = [\underline{d}^2 f(x),\overline{d}^2 f(x)] = -\frac{1}{(f_1(x))^2} D^2 f_1(x) + \frac{2}{(f_1(x))^3} D^2(r^2),$$

$$D^2(r^2) = [\underline{d}^2 r^2,\overline{d}^2 r^2],$$

$$\underline{d}^2(r^2) = \underline{d}^2(r',r') + \underline{d}^2(r',r'') + \underline{d}^2(r'',r') + \underline{d}^2(r'',r''),$$

$$\overline{d}^2(r^2) = \overline{d}^2(r',r') + \overline{d}^2(r',r'') + \overline{d}^2(r'',r') + \overline{d}^2(r'',r''),$$

$$\underline{d}^2(r',r') = \Big\{ [a,v,A] \colon a = a_1 a_2 + R(a_1+a_2), v = a_2 v_1 + a_1 v_2 + R v_1 + R v_2,$$

$$A = (a_1+R)A_2 + (a_2+R)A_1 + 2v_1 v_2, [a_1,v_1,A_1] \in \underline{d}^2 f_1(x),$$

$$[a_2,v_2,A_2] \in \underline{d}^2 f_1(x) \Big\},$$

$$\underline{d}^2(r',r'') = \Big\{ [a,v,A] \colon a = R(a_1-b_1), v = R(v_1-w_1), A = R(A_1-B_1),$$

$$[a_1,v_1,A_1] \in \underline{d}^2 f_1(x), [b_1,w_1,B_1] \in \overline{d}^2 f_1(x) \Big\},$$

$$\underline{d}^2(r'',r') = \Big\{ [a,v,A] \colon a = R(a_1-b_1), v = R(v_1-w_1), A = R(A_1-B_1),$$

$$[a_1,v_1,A_1] \in \underline{d}^2 f_1(x), [b_1,w_1,B_1] \in \overline{d}^2 f_1(x) \Big\},$$

$$\underline{d}^2(r'',r'') = \Big\{ [a,v,A] \colon a = b_1 b_2 - R(b_1+b_2), v = (b_2-R)w_1 + (b_1-R)w_2,$$

$$A = (b_2-R)B_1 + (b_1-R)B_2 + 2w_1 w_2, [b_1,w_1,B_1] \in \overline{d}^2 f_1(x),$$

$$[b_2,w_2,B_2] \in \overline{d}^2 f_1(x) \Big\},$$

$$\overline{d}^2(r',r') = \Big\{ [b,w,B] \colon b = -R(a_1+a_2), w = -R(v_1+v_2),$$

$$B=-R(A_1+A_2),[a_1,v_1,A_1]\in \underline{d}^2 f_1(x),[a_2,v_2,A_2]\in \underline{d}^2 f_1(x)\Big\},$$

$$\overline{d}^2(r',r'')=\Big\{[b,w,B]\colon b=a_1b_1-R(a_1-b_1),w=a_1w_1+b_1v_1+R(w_1-v_1),$$

$$B=(b_1-R)A_1+(a_1+R)B_1+2v_1w_1,[a_1,v_1,A_1]\in \underline{d}^2 f_1(x),$$

$$[b_1,w_1,B_1]\in \overline{d}^2 f_1(x)\Big\},$$

$$\overline{d}^2(r'',r')=\Big\{[b,w,B]\colon b=a_1b_1-R(a_1-b_1),w=a_1w_1+b_1v_1+R(w_1-v_1),$$

$$B=(b_1-R)A_1+(a_1+R)B_1+2v_1w_1,[a_1,v_1,A_1]\in \underline{d}^2 f_1(x),$$

$$[b_1,w_1,B_1]\in \overline{d}^2 f_1(x)\Big\},$$

$$\overline{d}^2(r'',r'')=\Big\{[b,w,B]\colon b=R(b_1+b_2),w=R(w_1+w_2),B=R(B_1+B_2),$$

$$[b_1,w_1,B_1]\in \overline{d}^2 f_1(x),[b_2,w_2,B_2]\in \overline{d}^2 f_1(x)\Big\}.$$

Тут число $R<\infty$ задовольняє, як у (7.6.17), нерівності

$$\left|a+\langle v,\Delta\rangle+\frac{1}{2}\langle A\Delta,\Delta\rangle\right|<R\quad \forall [a,v,A]\in \underline{d}^2 f_1(x),\quad \forall \Delta\in B_\delta,$$

$$\left|b+\langle w,\Delta\rangle+\frac{1}{2}\langle B\Delta,\Delta\rangle\right|<R\quad \forall [b,w,B]\in \overline{d}^2 f_1(x),\quad \forall \Delta\in B_\delta.$$

З (7.6.28) робимо висновок, що функція f двічі кодиференційовна (двічі неперервно кодиференційовна) в точці x. ☐

Лема 7.6.4. *Якщо функції $\varphi_i(x)$, $i\in I=\{1,\ldots,N\}$ двічі (неперервно) кодиференційовні в точці x, то і функції*

$$f_1(x)=\max_{i\in I}\varphi_i(x),\quad f_2(x)=\min_{i\in I}\varphi_i(x)$$

теж двічі (неперервно) кодиференційовні в цій точці. При цьому

$$D^2 f_1(x)=[\underline{d}^2 f_1(x),\overline{d}^2 f_1(x)],\quad D^2 f_2(x)=[\underline{d}^2 f_2(x),\overline{d}^2 f_2(x)], \tag{7.6.29}$$

де

$$\underline{d}^2 f_1(x)=$$

$$\operatorname{conv}\Big\{\underline{d}^2\varphi_k(x)-\sum_{i\in I\setminus\{k\}}\overline{d}^2\varphi_i(x)+\{[\varphi_k(x)-f_1(x);0_n;0_{n\times n}]\}\colon k\in I\Big\},$$

$$\overline{d}^2 f_1(x) = \sum_{i\in I} \overline{d}^2 \varphi_i(x), \quad \underline{d}^2 f_2(x) = \sum_{i\in I} \underline{d}^2 \varphi_i(x),$$

$$\overline{d}^2 f_2(x) =$$

$$= \operatorname{conv}\left\{\overline{d}^2 \varphi_k(x) - \sum_{i\in I\setminus\{k\}} \underline{d}^2 \varphi_i(x) + \{[\varphi_k(x) - f_2(x); 0_n; 0_{n\times n}]\} \colon k \in I\right\}.$$

Доведення проводиться так само, як доведення леми 7.2.4.

Лема 7.6.5. *Якщо функції $\varphi_1(x),\ldots,\varphi_m(x)$ - двічі (неперервно) кодиференційовні в точці x_0, а функція $F(z) = F(z_1,\ldots,z_m)$ двічі (неперервно) диференційовна в точці $z^0 = (\varphi_1(x_0),\ldots,\varphi_m(x_0))$, то функція $f(x) = F(\varphi_1(x),\ldots,\varphi_m(x))$ двічі (неперервно) кодиференційовна в точці x_0.*

Доведення. Маємо

$$F(z^0 + \Delta z) = F(z_1^0 + \Delta z_1,\ldots,z_m^0 + \Delta z_m) =$$

$$= F(z^0) + \langle F'_z(z^0),\Delta z\rangle + \frac{1}{2}\langle F''_{zz}(z^0)\Delta z,\Delta z\rangle + o((\Delta z)^2), \tag{7.6.30}$$

$$\varphi_i(x_0 + \Delta) = \varphi_i(x_0) + r_i(\Delta) + o_i(\Delta^2), \tag{7.6.31}$$

де

$$\Delta \in R^n, \quad \Delta z = (\Delta z_1,\ldots,\Delta z_m) \in R^m, \quad r_i(\Delta) = r'_i + r''_i,$$

$$r'_i = \max_{[a_i,v_i,A_i]\in\underline{d}^2\varphi_i(x_0)} [a_i + \langle v_i,\Delta\rangle + \frac{1}{2}\langle A_i\Delta,\Delta\rangle],$$

$$r''_i = \min_{[b_i,w_i,B_i]\in\overline{d}^2\varphi_i(x_0)} [b_i + \langle w_i,\Delta\rangle + \frac{1}{2}\langle B_i\Delta,\Delta\rangle],$$

$F'_z(z^0)$ - градієнт функції F в точці z^0, а $F''_{zz}(z^0)$ - матриця других похідних функції F в точці z^0:

$$F'_z = (F'_{z_1},\ldots,F'_{z_m}), \quad F''_{zz} = \begin{pmatrix} F''_{z_1 z_1} & \cdots & F''_{z_1 z_m} \\ \vdots & \ddots & \vdots \\ F''_{z_m z_1} & \cdots & F''_{z_m z_m} \end{pmatrix}.$$

З (7.6.30), (7.6.31) маємо

$$f(x_0 + \Delta) = F(z_0 + \Delta z) = F(z_0) + \sum_{i=1}^m F'_{z_i} r_i(\Delta) +$$

$$\sum_{j,k=1}^m F''_{z_j z_k} r_j(\Delta) r_k(\Delta) + o(\Delta^2) + o((\Delta z)^2). \tag{7.6.32}$$

Оскільки $|r_i(\Delta)| \leq L_i \|\Delta\|$, де $L_i < \infty$, то при $\Delta z_i = r_i(\Delta) + o_i(\Delta^2)$ маємо

$$o((\Delta z)^2) = o(\Delta^2). \tag{7.6.33}$$

З (7.6.32), (7.6.33) маємо

$$f(x_0+\Delta) = f(x_0)+\sum_{i=1}^{m} F'_{z_i} r_i(\Delta) + \sum_{j,k=1}^{m} F''_{z_j z_k} r_j(\Delta) r_k(\Delta) + o(\Delta^2). \tag{7.6.34}$$

Для довільного $\delta > 0$ знайдеться $R < \infty$ таке, що

$$\left| a_i + \langle v_i,\Delta\rangle + \frac{1}{2}\langle A_i\Delta,\Delta\rangle \right| < R \quad \forall [a_i,v_i,A_i] \in \underline{d}^2\varphi_i(x_0), \tag{7.6.35}$$

$$\left| b_i + \langle w_i,\Delta\rangle + \frac{1}{2}\langle B_i\Delta,\Delta\rangle \right| < R \quad \forall [b_i,w_i,B_i] \in \overline{d}^2\varphi_i(x_0). \tag{7.6.36}$$

Нерівності (7.6.35), (7.6.36) справджуються для всіх $i \in \{1,\ldots,m\}$, $\Delta \in B_\delta$. Зафіксуємо деяке $\delta > 0$. Міркуючи, як при доведенні леми 7.6.2, одержуємо при $\Delta \in B_\delta$

$$f(x_0+\Delta) = f(x_0) + \max_{[a,v,A]\in \underline{d}^2 f(x_0)} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle \right] +$$

$$\min_{[b,w,B]\in \overline{d}^2 f(x_0)} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle \right] + o(\Delta^2), \tag{7.6.37}$$

де

$$D^2 f(x_0) = [\underline{d}^2 f(x_0), \overline{d}^2 f(x_0)] =$$

$$= \sum_{i=1}^{m} F'_{z_i}(z^0) D^2\varphi_i(x_0) + \frac{1}{2}\sum_{j,k=1}^{m} F''_{z_j z_k}(z^0) D^2[r_j,r_k],$$

$$D^2[r_j,r_k] = [\underline{d}^2(r_j,r_k), \overline{d}^2(r_j,r_k)],$$

$$\underline{d}^2(r_j,r_k) = \underline{d}^2(r'_j,r'_k) + \underline{d}^2(r'_j,r''_k) + \underline{d}^2(r''_j,r'_k) + \underline{d}^2(r''_j,r''_k),$$

$$\overline{d}^2(r_j,r_k) = \overline{d}^2(r'_j,r'_k) + \overline{d}^2(r'_j,r''_k) + \overline{d}^2(r''_j,r'_k) + \overline{d}^2(r''_j,r''_k),$$

$$\underline{d}^2(r'_j,r'_k) = \Big\{ [a,v,A]\colon a = a_j a_k + R(a_j + a_k), v = a_k v_j + a_j v_k + R(v_j + v_k),$$

$$A = (a_j + R)A_k + (a_k + R)A_j + 2v_j v_k, [a_j,v_j,A_j] \in \underline{d}^2\varphi_j(x_0),$$

$$[a_k,v_k,A_k] \in \underline{d}^2\varphi_k(x_0) \Big\},$$

$$\underline{d}^2(r'_j,r''_k) = \Big\{ [a,v,A]\colon a = R(a_j - b_k), v = R(v_j - w_k), A = R(A_j - B_k),$$

$$[a_j, v_j, A_j] \in \underline{d}^2\varphi_j(x_0), [b_k, w_k, B_k] \in \overline{d}^2\varphi_k(x_0)\Big\},$$

$$\underline{d}^2(r_j'', r_k') = \Big\{[a, v, A]\colon a = R(a_k - b_j), v = R(v_k - w_j), A = R(A_k - B_j),$$

$$[a_k, v_k, A_k] \in \underline{d}^2\varphi_k(x_0), [b_j, w_j, B_j] \in \overline{d}^2\varphi_j(x_0)\Big\},$$

$$\underline{d}^2(r_j'', r_k'') = \Big\{[a, v, A]\colon a = b_j b_k - R(b_j + b_k), v = (b_k - R)w_j + (b_j - R)w_k,$$

$$A = (b_k - R)B_j + (b_j - R)B_k + 2w_j w_k, [b_j, w_j, B_j] \in \overline{d}^2\varphi_j(x_0),$$

$$[b_k, w_k, B_k] \in \overline{d}^2\varphi_k(x_0)\Big\},$$

$$\overline{d}^2(r_j', r_k') = \Big\{[b, w, B]\colon b = -R(a_j + a_k), w = -R(v_j + v_k),$$

$$B = -R(A_j + A_k), [a_j, v_j, A_j] \in \underline{d}^2\varphi_j(x_0), [a_k, v_k, A_k] \in \underline{d}^2\varphi_k(x_0)\Big\},$$

$$\overline{d}^2(r_j', r_k'') = \Big\{[b, w, B]\colon b = a_j b_k + R(a_j - b_k), w = a_j w_k + b_k v_j + R(w_k - v_j),$$

$$B = (b_k - R)A_j + (a_j + R)B_k + 2v_j w_k, [a_j, v_j, A_j] \in \underline{d}^2\varphi_j(x_0),$$

$$[b_k, w_k, B_k] \in \overline{d}^2\varphi_k(x_0)\Big\},$$

$$\overline{d}^2(r_j'', r_k') = \Big\{[b, w, B]\colon b = a_k b_j + R(a_k - b_j),$$

$$w = a_k w_j + b_j v_k + R(w_j - v_k),$$

$$B = (b_j - R)A_k + (a_k + R)B_j + 2v_k w_j, [a_k, v_k, A_k] \in \underline{d}^2\varphi_k(x_0),$$

$$[b_j, w_j, B_j] \in \overline{d}^2\varphi_j(x_0)\Big\},$$

$$\overline{d}^2(r_j'', r_k'') = \Big\{[b, w, B]\colon b = R(b_j + b_k), w = R(w_j + w_k), B = R(B_j + B_k),$$

$$[b_j, w_j, B_j] \in \overline{d}^2\varphi_j(x_0), [b_k, w_k, B_k] \in \overline{d}^2\varphi_k(x_0)\Big\}. \tag{7.6.38}$$

$\square$

Зауваження 7.6.2. Множина двічі кодиференційовних функцій (двічі неперервно кодиференційовних функцій) утворює лінійний простір, замкнутий відносно всіх "гладких" операцій і, що особливо важливо, щодо операцій поточкового максимуму і поточкового мінімуму від скінченної кількості функцій. Можна також довести кодиференційовність суперпозиції кодиференційовних функцій.

Зауваження 7.6.3. Можна також ввести поняття k раз кодиференційовної функції: f називається *k раз кодиференційовною* в точці x, якщо

$$f(x+\Delta) = f(x) + \max_{\varphi(\cdot)\in A} \varphi(\Delta) + \min_{\psi(\cdot)\in B} \psi(\Delta) + o(\|\Delta\|^k),$$

де $\varphi(\Delta)$ і $\psi(\Delta)$ - багатовимірні поліноми (від Δ) степеня k, а A і B - опуклі компакти таких поліномів. Можна побудувати і відповідне числення.

7.7. УМОВИ МІНІМУМУ ГІПОДИФЕРЕНЦІЙОВНОЇ ФУНКЦІЇ

Нехай функція f задана і неперервно кодиференційовна на відкритій множині $X \subset \mathbb{R}^n$, і нехай множину $S \subset X$ задано співвідношенням

$$S = \{x \in \mathbb{R}^n |\, h(x) \le 0\}, \tag{7.7.1}$$

де h – неперервно кодиференційовна на X функція.

Оскільки будь-яка кодиференційовна функція є квазідиференційовною, то виконуються всі результати попереднього розділу. Якщо точка $x \in S$ не є inf–стаціонарною, то одержані в попередньому розділі напрямки найшвидшого спуску і допустимі напрямки є, взагалі кажучи, розривними (як функція x). Використання кодиференціалів у випадку неперервно кодиференційовних функцій f і h дозволяє отримати напрямки спуску, які неперервно залежать від x.

Покажемо це для неперервно гіподиференційовних функцій f і h.

Нехай функції f і h є ліпшицевими і неперервно гіподиференційовними на X, тобто

$$f(x+\Delta) = f(x) + \max_{[a,v]\in df(x)} [a + \langle v,\Delta\rangle] + o_1(\Delta), \tag{7.7.2}$$

$$h(x+\Delta) = h(x) + \max_{[a',v']\in dh(x)} [a' + \langle v',\Delta\rangle] + o_2(\Delta), \tag{7.7.3}$$

$$\frac{o_i(\alpha\Delta)}{\alpha} \to_{\alpha\downarrow 0} 0,\ i = 1,2,$$

$df(x), dh(x) \subset \mathbb{R}^{n+1}$ – компакти, причому відображення df і dh неперервні (в метриці Хаусдорфа) на X.

З (7.7.2) і (7.7.3) випливає, що
$$\max_{[a,v]\in df(x)} a = \max_{[a',v']\in dh(x)} a' = 0. \tag{7.7.4}$$
Відмітимо, що тоді функції f і h субдиференційовні, тобто
$$f(x+\alpha g) = f(x) + \alpha \max_{v\in\partial f(x)} \langle v,g\rangle + o_1(\alpha),$$
$$h(x+\alpha g) = h(x) + \alpha \max_{v'\in\partial h(x)} \langle v',g\rangle + o_2(\alpha),$$
де $\partial f(x)$ і $\partial h(x) \subset \mathbb{R}^n$ – опуклі компакти (субдиференціали функцій f і h відповідно), причому
$$\partial f(x) = \{v \in \mathbb{R}^n \mid [0,v] \in df(x)\},$$
$$\partial h(x) = \{v' \in \mathbb{R}^n \mid [0,v'] \in dh(x)\}.$$

Теорема 7.7.1. *Для того, щоб в точці $x^* \in S$ функція f досягала свого найменшого на S значення, необхідно, щоб виконувалось умова*
$$0_{n+1} \in \operatorname{conv}\{df(x^*),\, dh(x^*) + [h(x),0_n]\} \equiv L(x^*). \tag{7.7.5}$$

Доведення. Припустимо, що x^* – точка мінімуму, але умова (7.7.5) не виконується. Тоді за теоремою про розділення знайдуться число $\gamma > 0$ і вектор $\bar g = [\delta,g] \in \mathbb{R}^{n+1}$ (де $\delta \in \mathbb{R}^1$, $g \in \mathbb{R}^n$) такі, що
$$\langle \bar v,\bar g\rangle \le -\gamma \quad \forall \bar v \in L(x^*). \tag{7.7.6}$$
При цьому можна вибрати $\bar g$ таке, що $\delta \ge 0$. Дійсно, з того, що $a \le 0, a' \le 0, h(x^*) \le 0$, і узявши $\bar g = -\bar z/\|\bar z\|$, де $\|\bar z\| = \min\limits_{z\in L(x^*)} \|z\|$, робимо висновок, що $\bar g = [\delta,g]$, де $\delta \ge 0$, тому що перші компоненти всіх векторів з $L(x^*)$ недодатні, а тому і перша компонента $\bar z$ також недодатня, а тоді $\delta \ge 0$. З (7.7.6) маємо:
$$a\delta + \langle v,g\rangle \le -\gamma \quad \forall\, [a,v] \in df(x^*), \tag{7.7.7}$$
$$(a' + h(x^*))\delta + (v',g) \le -\gamma \quad \forall\, [a',v'] \in dh(x^*). \tag{7.7.8}$$
Оскільки з (7.7.7) випливає $(v,g) \le -\gamma < 0\ \forall\, v \in \partial f(x^*)$, то $g \ne 0_n$. З (7.7.2) і (7.7.7) одержуємо
$$f(x^*+\alpha g) = f(x^*) + \max_{[a,v]\in df(x^*)} [a + \alpha\langle v,g\rangle] + o_1(\alpha) \le$$
$$\le f(x^*) + \max_{[a,v]\in df(x^*)} [a + \alpha(-\gamma - a\delta)] + o_1(\alpha) =$$
$$= f(x^*) - \alpha\gamma + \max_{[a,v]\in df(x^*)} [a(1-\alpha\delta)] + o_1(\alpha). \tag{7.7.9}$$
При достатньо малих $\alpha > 0$ буде $1 - \alpha\delta > 0$, тому з (7.7.4) випливає, що
$$\max_{[a,v]\in df(x^*)} [a(1-\alpha\delta)] = (1-\alpha\delta) \max_{[a,v]\in df(x^*)} a = 0.$$

Звідси і з (7.7.9) одержуємо

$$f(x^* + \alpha g) \leq f(x^*) - \alpha\gamma + o_1(\alpha). \tag{7.7.10}$$

Аналогічно з (7.7.3) і (7.7.8) маємо

$$h(x^* + \alpha g) \leq h(x^*) +$$

$$+ \max_{[a',v'] \in dh(x^*)} [a' + \alpha(-\gamma - (a' + h(x^*))\delta] + o_2(\alpha) =$$

$$= h(x^*) + \max_{[a',v'] \in dh(x^*)} [a' (1 - \alpha\delta)] - \alpha(\gamma + h(x^*)\delta) + o_2(\alpha). \tag{7.7.11}$$

Для досить малих $\alpha > 0$ буде $1 - \alpha\delta > 0$, і з (7.7.4) маємо

$$\max_{[a',v'] \in dh(x^*)} [a'(1 - \alpha\delta)] = (1 - \alpha\delta) \max_{[a',v'] \in dh(x^*)} a' = 0.$$

Тому з (7.7.11) випливає, що

$$h(x^* + \alpha g) \leq h(x^*) - \alpha(\gamma + h(x^*)\delta) + o_2(\alpha). \tag{7.7.12}$$

Якщо $h(x^*) < 0$, то для досить малих $\alpha \geq 0$ з (7.7.12) випливає, що $h(x^* + \alpha g) < h(x^*)/2 < 0$. Якщо ж $h(x^*) = 0$, то з (7.7.12) випливає, що $h(x^* + \alpha g) \leq -\alpha\gamma + o_2(\alpha)$, і знову при досить малих $\alpha > 0$ одержуємо

$$h(x^* + \alpha g) < 0, \tag{7.7.13}$$

тобто $x^* + \alpha g \in S$. Нерівності (7.7.10) і (7.7.13) суперечать тому, що x^* – точка мінімуму f на S. □

Зауваження 7.7.1. Ми не припускали, що $h(x^*) = 0$, тобто не розрізняли випадки $h(x^*) = 0$ і $h(x^*) \neq 0$.

Наслідок 7.7.1. *Якщо $h(x^*) < 0$, то з (7.7.4) і (7.7.5) отримаємо умову*

$$0_{n+1} \in df(x^*) \tag{7.7.14}$$

(елементи з $dh(x^*) + [h(x^*), 0_n]$ не можуть брати участь в опуклій комбінації, яка дає 0_{n+1}), а оскільки $\max\limits_{[a,v] \in df(x^*)} a = 0$, то з (7.7.14) випливає

$$0_n \in \partial f(x^*), \tag{7.7.15}$$

де $\partial f(x)$ – субдиференціал функції f в точці x.
Якщо ж $h(x^*) = 0$, то з (7.7.4) і (7.7.5) випливає, що

$$0_n \in \operatorname{conv}\{\partial f(x^*), \partial h(x^*)\}. \tag{7.7.16}$$

Із (7.7.15) і (7.7.16) легко бачити, що з (7.7.5) випливають стандартні умови мінімуму субдиференційовної функції на субдиференційовній множині.

Нехай тепер умова (7.7.5) в точці $x \in S$ не виконується. Знайдемо

$$\min_{\overline{z} \in L(x)} \|\overline{z}\| = \|\overline{z}(x)\|.$$

Зрозуміло, що $\|\overline{z}(x)\| > 0$. Нехай $\overline{z}(x) = [\eta(x), z(x)], \eta(x) \in \mathbb{R}^1, z(x) \in \mathbb{R}^n$.

Лема 7.7.1. *Напрямок*

$$g(x) = \frac{-z(x)}{\|z(x)\|} \tag{7.7.17}$$

є напрямком спуску функції f на множині S. При цьому

$$f'(x,g(x)) \leq -\|\overline{z}(x)\|. \tag{7.7.18}$$

Доведення. Лему фактично було доведено при доведенні теореми 7.7.1. Там же було показано, що $z(x) \neq 0_n$. □

Зауваження 7.7.2. В силу припущення функції f і h неперервно кодиференційовні, тому $g(x)$ (а також $z(x)$)– неперервні вектор-функції. Цей факт є суттєвим і може бути використаний для розробки чисельних методів оптимізації.

7.8. МЕТОД КОДИФЕРЕНЦІАЛЬНОГО СПУСКУ

7.8.1. Безумовна мінімізація

Нехай f – ліпшицева неперервно кодиференційовна на відкритій множині $X \subset \mathbb{R}^n$ функція. Потрібно знайти мінімум функції f на множині $S \subset X$. Спочатку розглянемо випадок $S = \mathbb{R}^n, \ X = \mathbb{R}^n$. У цьому випадку

$$f(x + \Delta) = f(x) + \max_{[a,v] \in \underline{d}f(x)} [a + \langle v,\Delta \rangle] +$$

$$+ \min_{[b,w] \in \overline{d}f(x)} [b + \langle w,\Delta \rangle] + o(\Delta), \tag{7.8.1}$$

$\underline{d}f(x), \ \overline{d}f(x) \subset \mathbb{R}^n$ – опуклі компакти, а відображення $Df = [\underline{d}f,\overline{d}f]$ неперервне за Хаусдорфом, $\frac{o(\alpha\Delta)}{\alpha} \to 0, \ \alpha \downarrow 0 \ \forall \Delta \in \mathbb{R}^n$.
Можна вважати, що виконується умова (7.7.4). Тоді необхідна умова мінімуму (6.4.1) має вигляд

$$0_{n+1} \in \{\underline{d}f(x^*) + [0,w]\} \ \ \forall [0,w] \in \overline{d}f(x^*). \tag{7.8.2}$$

Нехай в точці $x \in \mathbb{R}^n$ умова (7.8.2) не виконується. Тоді знайдеться таке $\overline{w} = [0,w] \in \overline{d}f(x)$, що

$$0_{n+1} \notin \{\underline{d}f(x) + \overline{w}\} \equiv L_{\overline{w}}(x). \tag{7.8.3}$$

Знайдемо $\min\limits_{\overline{z} \in L_{\overline{w}(x)}} \|\overline{z}\| = \|\overline{z}_{\overline{w}}(x)\|$. З (7.8.3) випливає, що

$$\overline{z}_{\overline{w}}(x) = [\eta_{\overline{w}}(x),z_{\overline{w}}(x)] \neq 0_{n+1}.$$

Як і при доведенні теореми 7.7.1 можемо зробити висновок, що $z_{\overline{w}}(x) \neq 0_n$, і що для напрямку $g_{\overline{w}}(x) = -z_{\overline{w}}(x)/\|z_{\overline{w}}(x)\|$ буде

$$f'(x,g_{\overline{w}}(x)) \leq -\|\overline{z}_{\overline{w}}(x)\|.$$

Опишемо наступний "принциповий" ("концептуальний", за термінологією Е. Полака) алгоритм:

Зафіксуємо довільне $\mu > 0$. Оберемо довільне $x_0 \in \mathbb{R}^n$. Нехай вже побудовано $x_k \in \mathbb{R}^n$. Якщо в точці x_k виконується (7.8.2), то точка x_k – inf-стаціонарна, і процес припиняється. В протилежному випадку для кожного $\overline{w} \in \overline{d}_\mu f(x_k)$, де

$$\overline{d}_\mu f(x) = \{\overline{w} \in \overline{d}f(x) | \ \overline{w} = (s,w), \ 0 \leq s \leq \mu\}, \tag{7.8.4}$$

знаходимо

$$\min_{\overline{z} \in L_{\overline{w}}(x_k)} \|\overline{z}\| = \|\overline{z}_{k\overline{w}}\|, \tag{7.8.5}$$

де

$$\overline{z}_{k\overline{w}} = [\eta_{k\overline{w}},z_{k\overline{w}}], \ L_{\overline{w}}(x_k) = [\underline{d}f(x_k) + \overline{w}].$$

З (7.8.1) випливає, що

$$f(x_k - \alpha z_{k\overline{w}}) \le f(x_k) + \max_{[a,v]\in \underline{d}f(x_k)} [(a+s) - \alpha\langle v+w, z_{k\overline{w}}\rangle] + o(\alpha). \quad (7.8.6)$$

З (7.8.5) маємо

$$\langle \overline{z}, -\overline{z}_{k\overline{w}}\rangle \le -\|\overline{z}_{k\overline{w}}\|^2 \ \ \forall \overline{z} \in L_{\overline{w}}(x_k) \quad (7.8.7)$$

тобто

$$(a+s)(-\eta_{k\overline{w}}) - \langle v+w, z_{k\overline{w}}\rangle \le -\|\overline{z}_{k\overline{w}}\|^2, \quad (7.8.8)$$

або

$$-\langle v+w, z_{k\overline{w}}\rangle \le -\|z_{k\overline{w}}\|^2 + (a+s)\eta_{k\overline{w}} \ \ \forall [a,v] \in \underline{d}f(x_k).$$

З (7.8.6), (7.8.8) маємо

$$f(x_k - \alpha z_{k\overline{w}}) \le$$

$$\le f(x_k) + \max_{[a,v]\in \underline{d}f(x_k)} [(a+s) - \alpha\|\overline{z}_{k\overline{w}}\|^2 + \alpha(a+s)\eta_{k\overline{w}}] + o_k(\alpha) =$$

$$= f(x_k) - \alpha\|\overline{z}_{k\overline{w}}\|^2 + \max_{[a,v]\in \underline{d}f(x_k)} [(a+s)(1+\alpha\eta_{k\overline{w}})] + o_k(\alpha). \quad (7.8.9)$$

При досить малих $\alpha > 0$ буде $(1+\alpha\eta_{k\overline{w}}) > 0$, тому з (7.8.9) одержуємо

$$f(x_k - \alpha z_{k\overline{w}}) \le f(x_k) - \alpha\|\overline{z}_{k\overline{w}}\|^2 +$$

$$+(1+\alpha\eta_{k\overline{w}}) \max_{[a,v]\in \underline{d}f(x_k)} [(a+s)] + o_k(\alpha) =$$

$$= f(x_k) - \alpha\|\overline{z}_{k\overline{w}}\|^2 + (1+\alpha\eta_{k\overline{w}})s + o_k(\alpha). \quad (7.8.10)$$

Відмітимо, що тут $s \in [0,\mu]$, тому напрямок $-z_{k\overline{w}}$ може й не бути напрямком спуску (навіть якщо $\|\overline{z}_{k\overline{w}}\| > 0$). Але оскільки в точці x_k не виконується умова (7.8.2), то знайдеться принаймні одне $\overline{w}_0 \in \overline{\partial} f(x_k)$ таке, що $\|\overline{z}_{kw_0}\| > 0$, а тоді (в цьому випадку $w_0 = 0$) з (7.8.10) бачимо, що напрямок $-z_{k\overline{w_0}}$ – це напрямок спуску.
Тепер знайдемо спочатку для кожного $\overline{w} \in \overline{d}_\mu f(x_k)$

$$\min_{\alpha>0} f(x_k - \alpha z_{k\overline{w}}) = f(x_k - \alpha_{k\overline{w}} z_{k\overline{w}}), \quad (7.8.11)$$

а потім

$$\min_{\overline{w}\in \overline{d}_\mu f(x_k)} f(x_k - \alpha_{k\overline{w}} z_{k\overline{w}}) = f(x_k - \alpha_{k\overline{w}_k} z_{k\overline{w}_k}). \quad (7.8.12)$$

Покладемо $x_{k+1} = x_k - \alpha_{k\overline{w}_k} z_{k\overline{w}_k}$. Далі продовжуємо аналогічно. В результаті будуємо послідовність $\{x_k\}$ таку, що

$$f(x_{k+1}) < f(x_k). \quad (7.8.13)$$

Зауваження 7.8.1. З (7.8.10) випливає, що напрямок $x_{k+1} - x_k$ може і не бути напрямком спуску (в цьому напрямку функція може спочатку зростати, а потім спадати, тобто алгоритм дозволяє "виходити" з деяких точок локального мінімуму).

Зауваження 7.8.2. Для практичного використання описаного алгоритму необхідно вміти ефективно розв'язувати задачі (7.8.5), (7.8.11), (7.8.12). Задача (7.8.5) є задачею квадратичного програмування, а якщо замість евклідової норми взяти m-норму, то вона перетворюється на задачу лінійного програмування (якщо, крім того, $\underline{d}f$ – многогранник, то досить розглядати лише його вершини).

7.8.2. Умовна мінімізація

Нехай функції f і h ліпшицеві і неперервно кодиференційовні на відкритій множині $X \subset \mathbb{R}^n$. Нехай

$$S = \{x \in \mathbb{R}^n |\ \ h(x) \leq 0\}. \tag{7.8.14}$$

Тоді

$$\begin{gathered} f(x+\Delta) = f(x) + \\ + \max_{[a,v] \in \underline{d}f(x)} [a + \langle v,\Delta\rangle] + \min_{[b,w] \in \overline{d}f(x)} [b + \langle w,\Delta\rangle] + o_1(\Delta), \end{gathered} \tag{7.8.15}$$

$$\begin{gathered} h(x+\Delta) = h(x) + \\ + \max_{[a',v'] \in \underline{d}h(x)} [a' + \langle v',\Delta\rangle] + \min_{[b',w'] \in \overline{d}h(x)} [b' + \langle w',\Delta\rangle] + o_2(\Delta). \end{gathered} \tag{7.8.16}$$

Можна довести наступне твердження.

Теорема 7.8.1. *Для того, щоб функція f в точці $x^* \in S$ досягала свого найменшого на S значення, необхідно, щоб*

$$\begin{gathered} 0_{n+1} \in \operatorname{conv}\{\underline{d}f(x^*) + [0,w], \underline{d}h(x^*) + [0,w'] + [h(x^*),0_n]\} \equiv \\ \equiv L(x^*,w,w') \ \ \forall w \in \overline{\partial}f(x^*),\ \ w' \in \overline{\partial}h(x^*), \end{gathered} \tag{7.8.17}$$

де

$$\overline{\partial}f(x) = \{w \in \mathbb{R}^n\colon\ \ [0,w] \in \overline{d}f(x)\},$$
$$\overline{\partial}h(x) = \{w' \in \mathbb{R}^n\colon\ \ [0,w'] \in \overline{d}h(x)\}.$$

Доведення. Припустимо протилежне. Тоді знайдуться $w_0 \in \overline{\partial}f(x^*)$ і $w_0' \in \overline{\partial}h(x^*)$ такі, що

$$0_{n+1} \notin L(x^*,w_0,w_0') \equiv L_0. \tag{7.8.18}$$

Знайдемо

$$\min_{\overline{z} \in L_0} \|\overline{z}\| = \|\overline{z}_0\|, \tag{7.8.19}$$

де $\overline{z}_0 = [\eta_0,z_0]$. З (7.8.18)

$$\|\overline{z}_0\| > 0, \tag{7.8.20}$$

Візьмемо $\overline{g}_0 = -\overline{z_0} = [\delta, g_0]$. З (7.8.19) отримуємо, що

$$\langle \overline{z}, g_0 \rangle \le -\|\overline{z}_0\|^2 \quad \forall \overline{z} \in L_0.$$

Звідси та з визначення L_0 маємо

$$\langle \overline{z}, \overline{g_0} \rangle \le -\|\overline{z}_0\|^2 \quad \forall \overline{z} \in \{\underline{d}f(x^*) + [0, w_0]\}, \tag{7.8.21}$$

$$\langle \overline{z}, g_0 \rangle \le -\|\overline{z}_0\|^2 \quad \forall \overline{z} \in \{\underline{d}h(x^*) + [h(x^*), w_0']\}, \tag{7.8.22}$$

З (7.8.21), (7.8.22) одержуємо

$$a\delta + \langle v + w_0, g_0 \rangle \le -\|\overline{z}_0\|^2 \quad \forall [a, v] \in \underline{d}f(x^*),$$

$$(a' + h(x^*))\delta + \langle v' + w_0', g_0 \rangle \le -\|\overline{z}_0\|^2 \quad \forall [a', v'] \in \underline{d}h(x^*).$$

Звідси

$$\langle v + w_0, g_0 \rangle \le -\|\overline{z}_0\|^2 - a\delta \quad \forall [a, v] \in \underline{d}f(x^*), \tag{7.8.23}$$

$$\langle v' + w_0', g_0 \rangle \le -\|\overline{z}_0\|^2 - (a' + h(x^*))\delta \quad \forall [a', v'] \in \underline{d}h(x^*). \tag{7.8.24}$$

З (7.8.15), (7.8.23) маємо

$$f(x^* + \alpha g_0) \le$$
$$\le f(x^*) + \max_{[a,v] \in \underline{d}f(x^*)} [a - \alpha\|\overline{z}_0\|^2 - \alpha a \delta] + o_1(\alpha) =$$
$$= f(x^*) - \alpha\|\overline{z_0}\|^2 + \max_{[a,v] \in \underline{d}f(x^*)} [(1 - \alpha\delta)a] + o_1(\alpha). \tag{7.8.25}$$

З (7.8.16), (7.8.24) отримуємо

$$h(x^* + \alpha g_0) \le$$
$$\le h(x^*) + \max_{[a',v'] \in \underline{d}h(x^*)} [a' - \alpha\|\overline{z}_0\|^2 - \alpha(a' + h(x^*))\delta] + o_2(\alpha) =$$
$$= h(x^*) - \alpha\|\overline{z_0}\|^2 + \max_{[a',v'] \in \underline{d}h(x^*)} [a'(1 - \alpha\delta)] - \alpha h(x^*)\delta + o_2(\alpha). \tag{7.8.26}$$

Оскільки при досить малих $\alpha > 0$ матимемо $(1 - \alpha\delta) > 0$, то з (7.7.4), (7.8.25), (7.8.26) отримуємо

$$f(x^* + \alpha g_0) \le f(x^*) - \alpha\|\overline{z_0}\|^2 + (1 - \alpha\delta) \max_{[a,v] \in \underline{d}f(x^*)} a + o_1(\alpha) =$$
$$= f(x^*) - \alpha\|\overline{z_0}\|^2 + o_1(\alpha), \tag{7.8.27}$$

$$h(x^* + \alpha g_0) \le h(x^*) - \alpha\|\overline{z_0}\|^2 - \alpha h(x^*)\delta + o_2(\alpha). \tag{7.8.28}$$

Якщо $h(x^*) < 0$, то при досить малих $\alpha > 0$ буде

$$h(x^* + \alpha g_0) \le \frac{1}{2} h(x^*) < 0, \tag{7.8.29}$$

а якщо $h(x^*) = 0$, то з (7.8.28) випливає

$$h(x^* + \alpha g_0) \le -\alpha\|\overline{z_0}\|^2 + o_2(\alpha). \tag{7.8.30}$$

З (7.8.29), (7.8.30) отримуємо, що при досить малих $\alpha > 0$ буде $h(x^* + \alpha g_0) < 0$, тобто $x^* + \alpha g_0 \in S$, а з (7.8.27) $f(x^* + \alpha g_0) < f(x^*)$, що суперечить тому, що x^* – точка мінімуму f на S. $\square$

7.9. УМОВИ МІНІМУМУ ДРУГОГО ПОРЯДКУ ДЛЯ ДВІЧІ КОДИФЕРЕНЦІЙОВНОЇ ФУНКЦІЇ

Нехай функція f двічі кодиференційовна на $\mathbb{R}^n$, тобто

$$f(x+\Delta) = f(x) + \max_{[a,v,A]\in \underline{d}^2 f(x)} \left[a + \langle v,\Delta\rangle + \frac{1}{2}\langle A\Delta,\Delta\rangle \right] +$$

$$+ \min_{[b,w,B]\in \overline{d}^2 f(x)} \left[b + \langle w,\Delta\rangle + \frac{1}{2}\langle B\Delta,\Delta\rangle \right] + o(\Delta^2), \qquad (7.9.1)$$

де

$$D^2 f(x) = [\underline{d}^2 f(x), \overline{d}^2 f(x)],\ \underline{d}^2 f(x),\ \overline{d}^2 f(x) \subset \mathbb{R}^1 \times \mathbb{R}^n \times \mathbb{R}^{n\times n},$$

$$\frac{o((\alpha\Delta)^2)}{\alpha^2} \to 0,\ \alpha \downarrow 0\ \ \forall \Delta \in \mathbb{R}^n.$$

Тут $\mathbb{R}^{n\times n}$ – простір дійсних $n \times n$ матриць.
Будемо вважати, що

$$\max_{[a,v,A]\in \underline{d}^2 f(x)} a = \min_{[b,w,B]\in \overline{d}^2 f(x)} b = 0. \qquad (7.9.2)$$

Нагадаємо, що за допомогою другого кодиференціалу $D^2 f$ легко побудувати і перший кодиференціал Df, і квазідиференціал $\mathfrak{D} f$. Так, наприклад,

$$\mathfrak{D} f(x) = [\underline{\partial} f(x), \overline{\partial} f(x)],$$

$$\underline{\partial} f(x) = \{v \in \mathbb{R}^n |\ \exists A \in R^{n\times n}:\ \ [0,v,A] \in \underline{d}^2 f(x)\},$$

$$\overline{\partial} f(x) = \{w \in \mathbb{R}^n |\ \exists B \in R^{n\times n}:\ \ [0,w,B] \in \overline{d}^2 f(x)\}.$$

Розглянемо задачу мінімізації f на $\mathbb{R}^n$. З теореми 6.4.1 маємо таку необхідну умову: для того, щоб в точці $x^* \in \mathbb{R}^n$ функція f досягала свого найменшого значення, необхідно, щоб

$$-\overline{\partial} f(x^*) \subset \underline{\partial} f(x^*). \qquad (7.9.3)$$

Точка x^*, для якої виконується (7.9.3), називається $\inf$*–стаціонарною точкою першого порядку функції f на $\mathbb{R}^n$.*

Теорема 7.9.1. *Для того, щоб в точці $x^* \in \mathbb{R}^n$ функція f досягала свого найменшого на $\mathbb{R}^n$ значення, необхідно, щоб 1) виконувалось включення (7.9.3) і 2) для кожного $g \in \mathbb{R}^n$, $\|g\| = 1$, було б або а) $f'(x^*,g) > 0$, або б) (якщо $f'(x^*,g) = 0$) для довільного $w \in \underline{\partial} f(x^*)$ має місце нерівність $\max\limits_{v\in\underline{\partial} f(x^*)} \langle v + w,g\rangle > 0$, або, якщо $\max\limits_{v\in\underline{\partial} f(x^*)} \langle v + w,g\rangle = 0$, нерівність*

$$\max_{[v,A]\in C(w,g)} [\langle v + w,q\rangle + \langle (A + B)g,g\rangle] \geq 0 \qquad (7.9.4)$$

$$\forall q \in \mathbb{R}^n,\ \forall B:\ \ [0,w,B] \in \overline{d}^2 f(x^*),$$

де

$$C(x^*,w,g) \equiv C(w,g) \equiv \{[v,A]|\ \ [0,v,A] \in \underline{d}^2 f(x^*),\ \langle v+w,g\rangle = 0\}. \quad (7.9.5)$$

Доведення. Для довільних $g,q \in \mathbb{R}^n$ з (7.9.1)

$$f(x^* + \alpha g + \frac{1}{2}\alpha^2 q) =$$

$$= f(x^*) + \max_{[a,v,A]\in \underline{d}^2 f(x^*)} \left[a + \alpha\langle v,g\rangle + \frac{1}{2}\alpha^2\left(\langle v,q\rangle + \langle Ag,g\rangle\right)\right] +$$

$$+ \min_{[b,w,B]\in \overline{d}^2 f(x^*)} \left[b + \alpha\langle w,g\rangle + \frac{1}{2}\alpha^2\left(\langle w,q\rangle + \langle Bg,g\rangle\right)\right] + o(\alpha^2). \quad (7.9.6)$$

Всі твердження теореми, крім (7.9.4), випливають з (7.9.3). Доведемо (7.9.4). Припустимо протилежне. Тоді знайдуться $q \in \mathbb{R}^n$, $w \in \underline{\partial} f(x^*)$ і $B \in \mathbb{R}^{n\times n}$ такі, що $[0,w,B] \in \overline{d}^2 f(x^*)$, $\max\limits_{v\in\underline{\partial} f(x^*)} \langle v+w,g\rangle = 0$ і

$$\max_{[v,A]\in C(w,g)} [\langle v+w,q\rangle + \langle (A+B)g,g\rangle] = -\gamma < 0. \quad (7.9.7)$$

Розглянемо відображення

$$\psi(\varepsilon) = \Big\{[a,v,A] \in \underline{d}^2 f(x^*):\quad -\varepsilon \le a \le 0,$$

$$\langle v,g\rangle = \max_{[a,v' A]\in \underline{d}^2 f(x^*)} \langle v',g\rangle - \varepsilon\Big\},\ \ \varepsilon \ge 0. \quad (7.9.8)$$

Оскільки $\underline{d}^2 f(x^*)$ – опуклий компакт, то відображення $\psi(\varepsilon)$ неперервне в точці $\varepsilon = 0$, тому з (7.9.7) отримуємо, що при досить малих $\varepsilon > 0$ буде

$$\langle v+w,q\rangle + \langle (A+B)g,g\rangle \le -\frac{\gamma}{2}\ \ \forall A: [a,v,A] \in \psi(\varepsilon), \quad (7.9.9)$$

а тоді з (7.9.6) отримуємо

$$f(x^* + \alpha g) \le f(x^*) + \max\{\varphi_1(\alpha),\varphi_2(\alpha)\} + o(\alpha^2), \quad (7.9.10)$$

де

$$\varphi_1(\alpha) = \max_{[a,v,A]\in\psi(\varepsilon)} [a + \alpha\langle v+w,g\rangle + \frac{1}{2}\alpha^2(\langle v+w,q\rangle + \langle (A+B)g,g\rangle)],$$

$$\varphi_2(\alpha) = \sup_{[a,v,A]\in \underline{d}^2 f(x^*)\setminus\psi(\varepsilon)} [a + \alpha\langle v+w,g\rangle + \frac{1}{2}\alpha^2(\langle v+w,q\rangle + \langle (A+B)g,g\rangle)].$$

Але з (7.9.9) маємо

$$\varphi_1(\alpha) \le -\frac{1}{2}\alpha^2\gamma, \quad (7.9.11)$$

а з (7.9.8) маємо

$$\varphi_2(\alpha) \le -\varepsilon - \alpha\varepsilon + \frac{1}{2}\alpha^2 \max_{[a,v,A]\in \underline{d}^2 f(x^*)} [\langle v+w,q\rangle + \langle (A+B)g,g\rangle].$$

При досить малих $\alpha > 0$ матимемо

$$\varphi_2(\alpha) < -\frac{\varepsilon}{2}. \tag{7.9.12}$$

З (7.9.10), (7.9.12) отримуємо, що при досить малих $\alpha > 0$

$$f(x^* + \alpha g) < f(x^*).$$

Отримана суперечність завершує доведення теореми. □

Означення 7.9.1. Точка $x^* \in \mathbb{R}^n$, що задовольняє умову (7.9.4), називається inf–*стаціонарною точкою другого порядку функції f на* $\mathbb{R}^n$.

Зауваження 7.9.1. Якщо $\max\limits_{v \in \underline{\partial} f(x^*)} \langle v + w, g \rangle = 0$, то при $q = 0$ із (7.9.4) маємо умову

$$\max_{A \in C_0(w,g)} [\langle (A + B)g, g \rangle] \geq 0 \quad \forall B: \ [0, w, B] \in \overline{d}^2 f(x^*), \tag{7.9.13}$$

де

$$C_0(w,g) = \left\{ A \in \mathbb{R}^{n \times n} \middle| \ [0, v, A] \in \underline{d}^2 f(x^*), \langle v + w, g \rangle = 0 \right\}.$$

Умова (7.9.4) отримана з (7.9.1) при $\Delta = \alpha g + \frac{1}{2}\alpha^2 q$, а умова (7.9.13) отримана при $\Delta = \alpha g$. Умова (7.9.4) дозволяє провести більш детальне дослідження умов мінімуму, ніж (7.9.13). Це можна побачити з наступного прикладу.

Приклад 7.9.1. Розглянемо функцію

$$f(x) = f(x^{(1)}, x^{(2)}) = \max\{f_1(x), f_2(x)\}, \tag{7.9.14}$$

де

$$f_1(x) = x^{(2)} + \frac{1}{2}(x^{(1)})^2, \ f_2(x) = -\frac{1}{2}x^{(2)} - \frac{1}{2}(x^{(1)})^2.$$

Візьмемо $x_0 = (0,0), \ \Delta = (\Delta^{(1)}, \Delta^{(2)})$. Маємо

$$f_1(x_0 + \Delta) = \Delta^{(2)} + \frac{1}{2}(\Delta^{(1)})^2, \tag{7.9.15}$$

$$f_2(x_0 + \Delta) = -\frac{1}{2}\Delta^{(2)} - \frac{1}{2}(\Delta^{(1)})^2, \tag{7.9.16}$$

тобто

$$\underline{d}^2 f_1(x_0) = \left\{ \left[0; (0,1); \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\},$$

$$\overline{d}^2 f_1(x_0) = \left\{ \left[0; (0,0); \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\},$$

$$\underline{d}^2 f_2(x_0) = \left\{ \left[0; (0, -\frac{1}{2}); \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\},$$

$$\overline{d}^2 f_2(x_0) = \left\{ \left[0; (0{,}0); \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\},$$

Звідси

$$\underline{d}^2 f(x_0) = \mathrm{conv} \left\{ \left[0; (0{,}1); \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right], \left[0; (0, -1/2); \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\},$$

$$\overline{d}^2 f(x_0) = \left\{ \left[0; (0{,}0); \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\}.$$

Маємо

$$\underline{\partial} f(x_0) = \mathrm{conv}\{(0{,}1); (0, -1/2)\}, \ \overline{\partial} f(x_0) = \{(0{,}0)\},$$

тобто $-\overline{\partial} f(x_0) \subset \underline{\partial} f(x_0)$, і необхідна умова першого порядку (7.9.3) виконана. Очевидно, що якщо $g = (g^{(1)}, g^{(2)})$, то $f'(x_0,g) = \max\{g^{(2)}, -\frac{1}{2} g^{(2)}\}$, тому $f'(x_0,g) = 0$, якщо $g^{(2)} = 0$. Для цього випадку $w = (0{,}0)$,

$$B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad C_0(w,g) = \mathrm{conv} \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \right\},$$

якщо $g_1 = (1{,}0)$ і $g_2 = (-1{,}0)$. Нагадаємо, що ми розглядаємо g такі, що $\|g\| = 1$. Маємо

$$\psi(g) = \max \left\{ \left\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} g, g \right\rangle, \left\langle \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} g, g \right\rangle \right\} =$$

$$= \max\{(g^{(1)})^2, -(g^{(1)})^2\} = (g^{(1)})^2.$$

Тому $\psi(g_1) = \psi(g_2) = 1 > 0$, тобто необхідна умова (7.9.13) виконується. Перевіримо тепер умову (7.9.4). Маємо для g_1 і g_2:

$$w = (0{,}0), \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$C_0(x_0,w,g) = \mathrm{conv} \left\{ \left[(0{,}1), \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right], \left[(0, -\frac{1}{2}), \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\},$$

Покладемо $q = (q^{(1)}, q^{(2)})$. Знайдемо

$$\psi(g,q) = \max_{[v,A] \in C(x_0,w,g)} \{\langle v,q \rangle + \langle Ag,g \rangle\} =$$

$$= \max\{q^{(2)} + (g^{(1)})^2, -\frac{1}{2} q^{(2)} - (g^{(1)})^2\}.$$

Для $g_1 = (1{,}0)$ і $g_2 = (-1{,}0)$ матимемо

$$\psi(g_1,q) = \psi(g_2,q) = \max\{q^{(2)} + 1, -\frac{1}{2} q^{(2)} - 1\}.$$

Взявши $q_0 = (0, -3/2)$, отримаємо

$$\psi(g_1,q_0) = \psi(g_2,q_0) = \max\{-1/2, -1/4\} = -1/4 < 0,$$

тобто умова (7.9.4) не виконується, і точка x_0 не є точкою мінімуму. Дійсно, взявши

$$\Delta = \alpha g_1 + \frac{1}{2}\alpha^2 q_0 = \alpha(1{,}0) + \frac{1}{2}\alpha^2(0, -\frac{3}{2}) = (\alpha, -\frac{3}{4}\alpha^2),$$

маємо

$$f(x_0 + \Delta) = \max\left\{-\frac{3}{4}\alpha^2 + \frac{1}{2}\alpha^2,\ \frac{1}{2}\cdot\frac{3}{4}\alpha^2 - \frac{1}{2}\alpha^2\right\} =$$

$$= \alpha^2 \max\left\{-\frac{1}{4}, -\frac{1}{8}\right\} = -\frac{1}{8}\alpha^2 < 0\ \ \forall\alpha > 0,$$

а для $\Delta = \alpha g_1$, наприклад, буде

$$f(x_0 + \Delta) = f(x_0 + \alpha g_1) = \max\left\{\frac{1}{2}\alpha^2, -\frac{1}{2}\alpha^2\right\} = \frac{1}{2}\alpha^2 > 0\ \ \forall\alpha > 0,$$

тобто за допомогою “променів” знайти менше значення функції не вдається.

ПОКАЖЧИК ТЕРМІНІВ

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. *Алексеев В. В., Тихомиров В. М., Фомин С. В.* Оптимальное управление. – М., Наука, 1979. – 430 с.
2. *Aubin J.–P.* Mathematical methods of game and economic theory. – Amsterdam, North-Holland, 1979. – 619 p.
3. *Aubin J.–P., Cellina A.* Differential inclusions. – N. Y., Berlin, Springer, 1984. – 342 p.
4. *Aubin J.–P.* Optima and equilibria. – N. Y., Berlin, Springer, 1998. – 429 p.
5. *Ауман Р., Шепли Л.* Значений для неатомических игр. – М.: Мир, 1977. – 358 с.
6. *Ашманов С. А., Тимохов А. В.* Теория оптимизации в задачах и упражнениях. – М., Наука, 1981. – 304 с.
7. *Базара М., Шетти К.* Нелинейное программирование. Теория и алгоритмы. – М., Мир, 1982. – 584 с.
8. *Borwein J. M., Lewis A. S.* Convex analysis and nonlinear optimization. – Springer-Verlag, N.Y., 2000. – 273 p.
9. *Boyd S., Vandenberghe L.* Convex Optimization. – Cambridge University Press, 2004. – 740 p.
10. *Бутковский А. Г., Самойленко Ю. Н.* Управление квантовомеханическими процессами. – М., Наука, 1984. – 256 с.
11. *Гермейер Ю Б.* Введение в теорию исследований операций. – М., Наука, 1971. – 383 с.
12. *Гермейер Ю Б.* Игры с непротивоположными интересами. – М., Наука, 1976. – 328 с.
13. *Гольштейн Е. Г.* Теория двойственности в математическом программировании. – М., Наука, 1971. – 352 с.
14. *Горелик В. А., Кононенко А. Ф.* Теоретико-игровые модели принятия решений в эколого-экономических системах. – М., Радио и связь, 1982. – 144 с.
15. *Гупал А. М.* Стохастические методы решеня негладких экстремальных задач. – Киев, Наукова думка, 1979. – 152 с.

16. *Демьянов В. Ф.* Негладкие задачи оптимизации и управления. – Л., Из-во Ленингр. ун-та, 1982. – 324 с.

17. *Демьянов В. Ф., Васильев Л. В.* Недифференцируемая оптимизация. – М., Наука, 1981. – 384 с.

18. *Демьянов В. Ф., Рубинов А. М.* Основы негладкого анализа и квазидифференциальное исчисление. – М., Наука, 1990. – 432 с.

19. *Demyanov V. F., Dixon L. C. W.* Quasidifferential calculus// Math. Programming Study. – Vol. 29. – 221 p.

20. *Евтушенко Ю. Г.* Методы решения экстремальных задач и их приложение в системах оптимизации. – М., Наука, 1982. – 432 с.

21. *Ioffe A. D.* Nonsmooth analysis: differential calculus of nondifferentiable mappings// Trans. Amer. Math. Soc. – 1981. – V. 266. – No. 1. – P. 1–56.

22. *Иоффе А. Д., Тихомиров В. М.* Теория экстремальных задач. – М., Наука, 1974. – 480 с.

23. *Ириарт-Уррути Ж.–Б.* Оптимизация и выпуклый анализ. – Киев, КИТ, 2004. – 370 с.

24. *Канторович Л. В., Горстко А. Б.* Оптимальные решения в экономике. – М., Наука, 1982. – 232 с.

25. *Кларк Ф.* Оптимизация и негладкий анализ. – М., Наука, 1988. – 280 с.

26. *Красовский Н. Н.* Управление динамической системой. – М., Наука, 1985. – 520 с.

27. *Кругер А. Я.* Обобщенные дифференциалы негладких функций и необходимые условия экстремума// Сиб. мат. журн. – 1985ю – Т. 26.– № 3. – С. 78–90.

28. *Куржанский А. Б.* Управление и наблюдение в условиях неопределенности. – М., Наука, 1977. – 392 с.

29. *Лоран П. Ж.* Аппроксимация и оптимизация. – М., Мир, 1975. – 496 с.

30. *Льюис Р., Райфа Х.* Игры и решения. – М.: ИЛ, 1961. – 640 с.

31. *Макаров В. Л., Рубинов А. М.* Математическая теория экономической динамики и равновесия. – М., Наука, 1973. – 336 с.

32. *Michel P., Penot J.–P.* Calculus sous-differential pour les fonctions lipschitizienness et non-lipschitizienness// C. R. Acad. Sci. Paris, Ser. I. – 1984.– v. 298. – P. 269–272.

33. *Михалевич В. С., Гупал А. М., Норкин В. И.* Методы невыпуклой оптимизации. – М., Наука, 1987.

34. *Моисеев Н. Н.* Современное состояние теории исследования операций. – М., Наука, 1979. – 464 с.

35. *Моисеев Н. Н.* Математические методы системного анализа. – М., Наука, 1981. – 488 с.

36. *Моклячук М. П.* Варіаційне числення. Екстремальні задачі. – Київ, 2004. – 384 с.

37. *Моклячук М. П.* Основи опуклого аналізу. – Київ, ТВіМС, 2004. – 240 с.

38. *Моклячук М. П., Ямненко Р. Є.* Лекції з теорії вибору та прийняття рішень. – Київ, ВПЦ "Київський університет", 2007. – 256 с.

39. *Moreau J.-J.* Fonctionelles sous-differentiables// C. R. Acad. Sci. Paris, Ser. A-B. – 1963.– v. 257. – p. 4117–4119.

40. *Мордухович В. С.* Методы аппроксимации в задачах оптимизации и управления. – М., Наука, 1988. – 360 с.

41. *Мулен Э.* Кооперативное принятие решений. Аксиомы и модели. – М.: Мир, 1991. – 464 с.

42. *Нейман Дж., Моргенштерн О.* Теория игр и экономическое поведение. – М., Наука, 1970. – 708 с.

43. *Никайдо Х.* Выпуклые структуры и математичемкая экономика. – М., Мир, 1972. – 516 с.

44. *Обен Ж.-П.* Нелинейный анализ. – М., Мир, 1988. – 264 с

45. *Обен Ж.–П., Экланд И.* Прикладной нелинейный анализ. – М., Мир, 1988. – 510 с.

46. *Оуэн Г.* Теория игр. – М., Мир, 1971. – 230 с.

47. *Пшеничный Б. Н.* Выпуклый анализ и экстремальные задачи. – М., Наука, 1980. – 320 с.

48. *Пшеничный Б. Н.* Необходимые условия экстремума. – М., Наука, 1982. – 144 с.

49. *Rockafellar R. T.* Theorems for convex functions// Bull. Amer. Math. Soc., Vol. 70. —- 1964. -— P. 189–192.

50. *Рокафеллар Р.* Выпуклый анализ. – М., Мир, 1973. – 472 с.

51. *Сакс С.* Теория интеграла. – М., ИЛ, 1949. – 494 с.

52. *Тихомиров В. М.* Выпуклый анализ. – М., Итоги науки и техники ВИНИТИ, 1987. – 102 с.

53. *Тихомиров В. М.* Некоторые вопросы теории приближений. – М., Из-во Моск. ун-та, 1976. – 324 с.

54. *Федоренко Р. П.* Приближенное решение задач оптимального управления. – М., Наука, 1978. – 488 с.

55. *Hiriart-Urruty J.-B.* New concepts of nondifferentiable programming// Bull. Soc. Math. France, Mem. 60. – 1979. – P. 57–85.

56. *Hiriart-Urruty J.-B., Lemarechal C.* Fundamentals of Convex Analysis. – N.Y., Springer-Verlag, 2002. – 260 p.

57. *Shapley L. S.* A value for n-person games// Ann. Math. Studies. – Princeton.– 1953. – Vol. 28.– P. 307–317.

58. *Цурков В. И.* Декомпозиция в задачах большой размерности. – М., Наука, 1981. – 362 с.

59. *Шор Н. З.* О классе почти-дифференцируемых функций и одном методе минимизации функций этого класса// Кибернетика. – 1972. – № 4. – С. 65–70.

60. *Шор Н. З.* Методы минимизации недифференцируемых функций и их приложения. – Киев, Наукова думка, 1979. – 200 с.

61. *Экланд И.* Элементы математичемкой экономики. – М., Мир, 1983. – 248 с.