UDC 519.6:519.8
DOI: 10.20535/SRIT.2308-8893.2024.2.09

ПРОЕКЦІЯ ГРАДІЄНТА: СПРОЩЕННЯ ОБЛАСТІ МІНІМІЗАЦІЇ АФІННИМ ПЕРЕТВОРЕННЯМ

І.Я. СПЕКТОРСЬКИЙ

Анотація. Розглянуто класичну задачу оптимізації у скінченновимірному просторі, тобто знаходження мінімуму функції на непорожній множині. Пошук точного розв'язку цієї задачі аналітичними методами потребує множинних обчислювальних ресурсів або взагалі неможливий. Для реальних задач частіше застосовують методи пошуку наближеного розв'язку, серед яких одним з найпростіших і найвідоміших для задач безумовної оптимізації є метод градієнтного спуску. Узагальненням методу градієнтного спуску на випадок умовної оптимізації є запропонований у 1964 р. метод проекції градієнта. Для деяких типів множини (відрізок, паралелепіпед, куля) проекцію точки на множину можна знайти простими явними формулами, проте для складніших (напр., еліпс) проектування стає окремою задачею мінімізації. Однак у деяких випадках обчислення проекції не можна спростити афінним перетворенням — напр., еліпс афінним (і навіть лінійним) перетворенням можна звести до кулі. Спрощено задачу мінімізації функції на множині застосуванням афінного перетворення $F(x) = Ax+b$, де A — невироджена квадратна матриця, b — фіксований вектор відповідної розмірності.

Ключові слова: проекція градієнта, мінімізація, афінне перетворення.

ВСТУП

Класичною задачею оптимізації у скінченновимірному просторі є знаходження мінімуму функції $f:\mathbb{R}^n \to \mathbb{R}$ на непорожній множині $D \subset \mathbb{R}^n$ (тут і надалі символ «$\subset$» позначає нестроге вкладення, тобто можливий випадок $D = \mathbb{R}^n$ — задача безумовної оптимізації). Як правило, пошук точного розв'язку цієї задачі аналітичними методами потребує надто багато обчислювальних ресурсів або взагалі неможливий. Тому для реальних задач найчастіше застосовують методи пошуку наближеного розв'язку, серед яких одним з найпростіших і найвідоміших для задач безумовної оптимізації є метод градієнтного спуску (див., напр., [1–6]). Узагальненням методу градієнтного спуску на випадок умовної оптимізації ($D \subset \mathbb{R}^n$) є запропонований у 1964 р. метод проекції градієнта (див. [7]). Для деяких типів множини D (відрізок, паралелепіпед, куля) проекцію точки $x \in \mathbb{R}^n$ на D можна знайти простими явними формулами, проте для дещо складніших областей (напр., якщо $D \subset \mathbb{R}^2$ — еліпс) проектування стає окремою задачею мінімізації. Однак у деяких випадках обчислення проекції на D можна спростити афінним перетворенням — напр., еліпс афінним (і навіть лінійним) перетворенням можна звести до кулі.

Мета роботи — звести задачу мінімізації функції $f:\mathbb{R}^n \to \mathbb{R}$ на множині $D \subset \mathbb{R}^n$ до мінімізації функції $\widetilde{f}(\widetilde{x}) = f(F^{-1}(\widetilde{x}))$ на множині

$\widetilde{D} = \{Ax + b : x \in D\} = F(D)$, де $F(x) = Ax + b$, A — невироджена матриця розмірності $n \times n$, $b \in \mathbb{R}^n$ — фіксований вектор.

ОПУКЛІ МНОЖИНИ ТА ФУНКЦІЇ

Означення 1. Множину $D \subset \mathbb{R}^n$ називають опуклою, якщо $\lambda x + (1-\lambda) y \in D$ для будь-яких $x \in D, y \in D, \lambda \in [0,1]$. Інакше кажучи, D є опуклою, якщо разом з точками $x \in D, y \in D$ містить відрізок $[x, y]$:

$$\left.\begin{array}{l} x \in D, \\ y \in D \end{array}\right\} \Rightarrow [x, y] = \{\lambda x + (1-\lambda) y : \lambda \in [0,1]\} \subset D .$$

Приклад 1. На площині (простір $\mathbb{R}^2$) будь-який круг є опуклою фігурою (множиною), будь-яке коло — ні (рис. 1). Інші приклади опуклих та неопуклих фігур зображено на рис. 2.

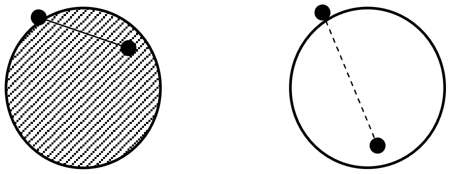

Рис. 1. Круг і коло — властивість опуклості

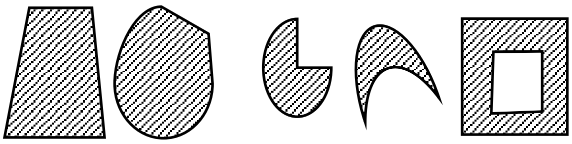

Рис. 2. Опуклі та неопуклі фігури

Зауваження 1. Надалі вважатимемо, що будь-який еліпс та будь-який багатокутник містять внутрішню ділянку, яку вони обмежують. Таким чином, еліпс та трикутник завжди є опуклими фігурами, однак чотирикутник може не бути опуклим (рис. 3).

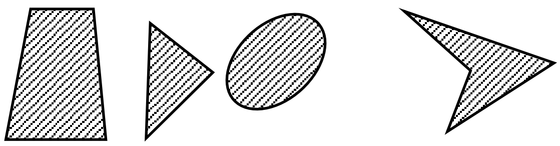

Рис. 3. Еліпс та багатокутники — властивість опуклості

Означення 2. Функцію $f : \mathbb{R}^n \to \mathbb{R}$ називають опуклою, якщо для будь-яких $x \in D, y \in D, \lambda \in [0,1]$ справджується нерівність

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y) . \tag{1}$$

Означення 3. Функцію $f : \mathbb{R}^n \to \mathbb{R}$ називають строго опуклою, якщо для будь-яких $x \in D, y \in D, \lambda \in (0,1)$, таких, що $x \neq y$, справджується нерівність

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y) .$$

Приклад 2. 1. Одновимірна ($n=1$) лінійна функція $f_1(x)=bx+c$ за будь-яких фіксованих констант $b\in\mathbb{R}$, $c\in\mathbb{R}$ є опуклою, але не строго опуклою, оскільки для будь-якої $\lambda\in[0,1]$ нестрога нерівність (1) виконується, проте перетворюється в рівність:

$$f_1(\lambda x+(1-\lambda)y)=b(\lambda x+(1-\lambda)y)+c=$$
$$=\lambda(bx+c)+(1-\lambda)(by+c)=\lambda f_1(x)+(1-\lambda)f_1(y)\,. \qquad (2)$$

2. Одновимірна квадратична функція $f_2(x)=ax^2$ ($a\in\mathbb{R}$ — фіксована константа) є опуклою тоді і тільки тоді, коли $a\geq 0$. Дійсно, для будь-якої $\lambda\in[0,1]$ маємо:

$$f_2(\lambda x+(1-\lambda)y)=a(\lambda x+(1-\lambda)y)^2=a\lambda^2x^2+a(1-\lambda)^2y^2+2a\lambda(1-\lambda)xy=$$
$$=a\lambda x^2+a(1-\lambda)y^2-a\lambda(1-\lambda)x^2-a\lambda(1-\lambda)y^2+2a\lambda(1-\lambda)xy= \qquad (3)$$
$$=\lambda f_2(x)+(1-\lambda)f_2(y)-a\lambda(1-\lambda)(x-y)^2.$$

Очевидно, у випадку $a>0$ функція $f_2(x)=ax^2$ є строго опуклою. Зазначимо, що отриманий результат з урахуванням рівностей (2) та (3) легко поширити на загальну одновимірну квадратичну функцію $f_3(x)=ax^2+bx+c$; функція опукла, якщо $a\geq 0$ (зокрема, строго опукла, якщо $a>0$); константи $b\in\mathbb{R}$, $c\in\mathbb{R}$ на опуклість (строгу опуклість) функції ax^2+bx+c не впливають.

Із означень 1 та 2 негайно випливає простий факт, що пов’язує обидва означення: функція $f:\mathbb{R}^n\to\mathbb{R}$ опукла тоді і тільки тоді, коли її надграфік $\operatorname{Epi} f=\{(x,y)\in\mathbb{R}^n\times\mathbb{R}: f(x)=y\}$ є опуклою множиною в просторі $\mathbb{R}^{n+1}=\mathbb{R}^n\times\mathbb{R}$. На рис. 4 схематично показано надграфіки функцій $f_1(x)=bx+c$ та $f_3(x)=ax^2+bx+c$.

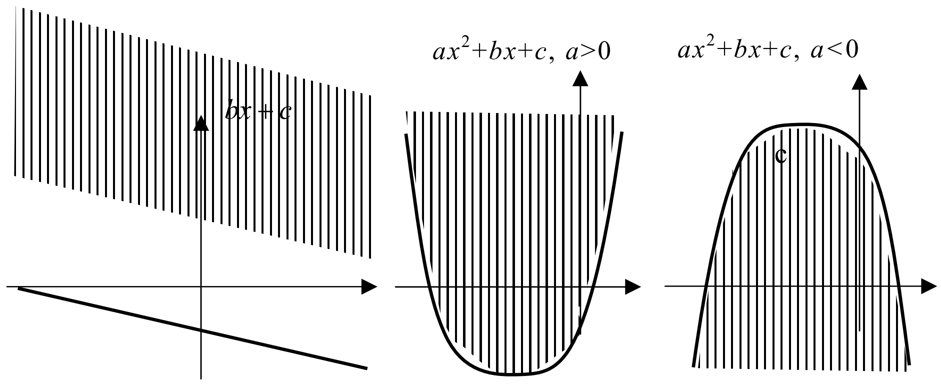

Рис.4. Надграфіки лінійної та квадратичної функцій

Означення 4. Функцію $f:\mathbb{R}^n\to\mathbb{R}$ називають сильно опуклою з параметром $\theta>0$, якщо для будь-яких $x\in D$, $y\in D$, $\lambda\in[0,1]$ справджується нерівність

$$f(\lambda x+(1-\lambda)y)\leq\lambda f(x)+(1-\lambda)f(y)-\tfrac{1}{2}\theta\lambda(1-\lambda)\,\|x-y\|^2\,.$$

Приклад 3. Квадрат евклідової норми $\| x \|^2$ як функція $\| \cdot \|^2 : \mathbb{R}^n \to \mathbb{R}$ є сильно опуклою з параметром $\theta = 2$. Дійсно, для будь-якої $\lambda \in [0,1]$ аналогічно (3) отримуємо:

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

Зауваження 2. Деякі автори (напр., [4, 5, 8]), визначаючи сильну опуклість функції, випускають коефіцієнт $\frac{1}{2}$ у доданку перед параметром опуклості, що зменшує параметр опуклості вдвічі. Так, згідно з таким визначенням, функція $\| x \|^2$ є сильно опуклою з параметром $\theta = 1$.

Відомі критерії опуклості для гладких функцій. Так, для двічі неперервно диференційовної функції $f : \mathbb{R}^n \to \mathbb{R}$ справджуються логічні еквівалентності (див., напр., [2, 4–6, 8]):

- Функція f опукла тоді й тільки тоді, коли $(f''h, h) \geq 0$ для будь-якого $h \in \mathbb{R}^n$ (гесіан f'' є невід'ємно визначеною матрицею).
- Функція f сильно опукла з параметром $\theta > 0$ тоді і тільки тоді, коли $(f''h, h) \geq \theta \| h \|^2$ для будь-якого $h \in \mathbb{R}^n$ (матриця $f'' - \theta I_n$ є невід'ємно визначеною матрицею).

Для строго опуклих функцій аналогічна еквівалентність неправильна, проте справджується наслідок: для строгої опуклості функції f достатньо, щоб $(f''h, h) > 0$ для будь-якого ненульового $h \in \mathbb{R}^n$ (гесіан f'' є додатно визначеною матрицею).

Приклад 4. Одновимірна функція $f(x) = x^4$ опукла, і навіть строго опукла. Дійсно, враховуючи строгу (навіть сильну) опуклість функції $g(x) = x^2$ (див. приклад 3) та строгу монотонність $g(x) = x^2$ на $[0, +\infty)$, отримуємо:

$$f(x) = g(g(x)) = (\lambda x + (1-\lambda) y)^4 = ((\lambda x + (1-\lambda) y)^2)^2 < (\lambda x^2 + (1-\lambda) y^2)^2 <$$
$$< \lambda^2 x^4 + (1-\lambda)^2 y^4 \leq \lambda x^4 + (1-\lambda) y^4 = \lambda f(x) + (1-\lambda) f(y)$$

для будь-якої $\lambda \in (0,1)$. Проте функція $f(x) = x^4$ не сильно опукла, оскільки для будь-якої константи $\theta > 0$ нерівність $f''(x) = 12x^2 \geq \theta$ не справджується для $| x | \leq \sqrt{\frac{\theta}{12}}$.

Сформулюємо у вигляді лем три відомі (див., напр., [2, 4, 5, 8]) властивості щодо мінімізації опуклої функції на опуклій множині.

Лема 1. Нехай $D \subset \mathbb{R}^n$ — опукла множина, . $f : \mathbb{R}^n \to \mathbb{R}$ — опукла функція, $x^* \in D$ — локальний мінімум f на D, тобто існує таке $\varepsilon > 0$, що $f(x^*) \leq f(x)$ для всіх $x \in D$, для яких $\| x - x^* \| < \varepsilon$. Тоді $x^* \in D$ — глобальний мінімум f на D, тобто $f(x^*) \leq f(x)$ для всіх $x \in D$.

Доведення. Нехай $x^* \in D$ — локальний мінімум f на D, $\varepsilon > 0$ — таке число, що $f(x^*) \le f(x)$ для всіх $x \in D$, для яких $\|x - x^*\| \le \varepsilon$. Зафіксуємо довільне $a \in D$. Тоді для $\lambda = \frac{\varepsilon}{2}\min(\frac{1}{\|x^*\|}, \frac{1}{\|a\|})$ маємо: $\|(\lambda a + (1-\lambda)x^*) - x^*\| = \|\lambda a - \lambda x^*\| \le \lambda(\|a\| + \|x^*\|) \le \varepsilon$, звідки, з урахуванням опуклості множини D та функції f, $f(x^*) \le f(\lambda a + (1-\lambda)x^*) \le \lambda f(a) + (1-\lambda) f(x^*)$. Отже, $f(x^*) \le \lambda f(a) + (1-\lambda) f(x^*)$, звідки $\lambda f(x^*) \le \lambda f(a)$ і, оскільки $\lambda > 0$, $f(x^*) \le f(a)$. Таким чином, ураховуючи довільність вибору $a \in D$, точка $x^* \in D$ дійсно є глобальним мінімумом функції f на D.□

Лема 2. Нехай $D \subset \mathbb{R}^n$ — опукла множина, . $f : \mathbb{R}^n \to \mathbb{R}$ — строго опукла функція. Тоді f досягає глобального мінімуму на D не більш ніж в одній точці.

Доведення. Припустимо, що f досягає глобального мінімуму y^* на D принаймні у двох різних точках $m_1 \in D$, $m_2 \in D$, тобто $f(m_1) = f(m_2) = y^*$, та $m_1 \ne m_2$. Тоді, враховуючи опуклість D та строгу опуклість f, отримуємо, що $\frac{m_1+m_2}{2} \in D$ та $f(\frac{m_1+m_2}{2}) < \frac{1}{2}(f(m_1) + f(m_2)) = y^*$, що суперечить глобальності мінімуму $f(m_1) = f(m_2) = y^*$ на множині D.□

Лема 3. Нехай $D \subset \mathbb{R}^n$ — опукла множина, . $f : \mathbb{R}^n \to \mathbb{R}$ — сильно опукла функція. Тоді f досягає глобального мінімуму на D у точності в одній точці.

Доведення (див., напр., [4, 5, 8]) спирається на обґрунтуванні обмеженості знизу множин Лебега $D_v = \{\forall x \in D : f(x) \le v\}$ та неперервності f на $\mathbb{R}^n$.

Детальніше про властивості опуклих множин і функцій див., [2, 4–6, 8].

АФІННЕ ПЕРЕТВОРЕННЯ ОПУКЛИХ МНОЖИН ТА АРГУМЕНТІВ ОПУКЛИХ ФУНКЦІЙ

Уведемо до розгляду афінне перетворення $F : \mathbb{R}^n \to \mathbb{R}^n$, тобто відображення $F(x) = Ax + b$, де A — фіксована невироджена матриця розмірності $n \times n$, $b \in \mathbb{R}^n$ — фіксований вектор. Зазначимо, що завдяки невиродженості A існує обернене (також афінне) перетворення $F^{-1}(x) = A^{-1}x - A^{-1}b$.

Лема 4. Нехай $D \subset \mathbb{R}^n$ — непорожня опукла множина. Тоді опуклою є також множина $\widetilde{D} = \{Ax + b : x \in D\} = F(D)$.

Доведення. Нехай $\widetilde{x} \in \widetilde{D}, \widetilde{y} \in \widetilde{D}$. За визначенням множини $\widetilde{D}$, маємо елементи $x \in D,\ y \in D$, такі, що $\widetilde{x} = Ax + b,\ \widetilde{y} = Ay + b$, звідки $x = A^{-1}\widetilde{x} - A^{-1}b \in D,\ y = A^{-1}\widetilde{y} - A^{-1}b \in D$. Тоді, зафіксувавши довільну $\lambda \in [0,1]$, отримуємо:

$$\lambda\widetilde{x} + (1-\lambda)\widetilde{y} = \lambda(Ax + b) + (1-\lambda)(Ay + b) =$$

$$= A(\lambda x + (1-\lambda)y + b) = F(\lambda x + (1-\lambda)y) \in \widetilde{D},$$

оскільки $\lambda x+(1-\lambda)y\in D$ завдяки опуклості D. Отже, $\lambda\widetilde{x}+(1-\lambda)\widetilde{y}\in\widetilde{D}$, що доводить опуклість множини $\widetilde{D}$.

Приклад 5. Розглянемо в $\mathbb{R}^2$ множину, обмежену прямокутником (сторони повернуті відносно координатних осей):

$$D=\left\{\begin{pmatrix}x_1\\x_2\end{pmatrix}:\begin{cases}2\le 3x_1+4x_2\le 5\\2\le -4x_1+3x_2\le 10\end{cases}\right\}.$$

Очевидно, область D можна отримати афінним (навіть лінійним) перетворенням із прямокутника $\widetilde{D}$, сторони якого паралельні координатним осям: $\begin{pmatrix}3x_1+4x_2\\-4x_1+3x_2\end{pmatrix}=\begin{pmatrix}3&4\\-4&3\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=5\begin{pmatrix}0.6&0.8\\-0.8&0.6\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}$. Таким чином, афінне перетворення $\begin{pmatrix}\widetilde{x}_1\\\widetilde{x}_2\end{pmatrix}=F\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\right)=\begin{pmatrix}0.6&0.8\\-0.8&0.6\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}$ (у даному випадку — поворот) переводить повернутий прямокутник $D=\left\{\begin{pmatrix}x_1\\x_2\end{pmatrix}:\begin{cases}2\le 3x_1+4x_2\le 5\\2\le -4x_1+3x_2\le 10\end{cases}\right\}$ у прямокутник $\widetilde{D}=\left\{\begin{pmatrix}\widetilde{x}_1\\\widetilde{x}_2\end{pmatrix}:\begin{cases}2\le 5\widetilde{x}_1\le 5\\2\le 5\widetilde{x}_2\le 10\end{cases}\right\}=\left\{\begin{pmatrix}\widetilde{x}_1\\\widetilde{x}_2\end{pmatrix}:\begin{cases}0.4\le \widetilde{x}_1\le 1\\0.4\le \widetilde{x}_2\le 2\end{cases}\right\}=F(D)$, сторони якого паралельні координатним осям (рис. 5).

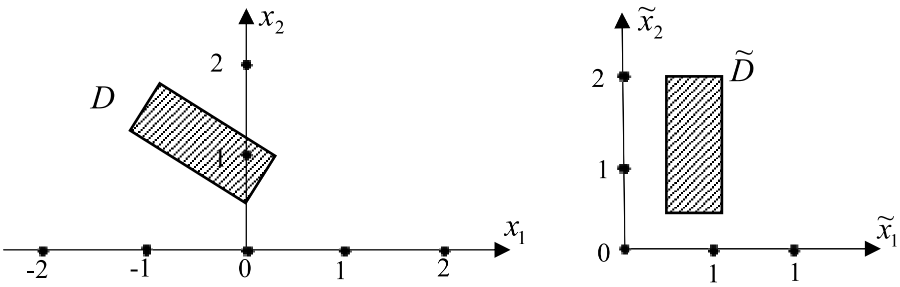

Рис. 5. Область $D=\left\{\begin{pmatrix}x_1\\x_2\end{pmatrix}:\begin{cases}2\le 3x_1+4x_2\le 5\\2\le -4x_1+3x_2\le 10\end{cases}\right\}$ — повернутий прямокутник, область $\widetilde{D}=\left\{\begin{pmatrix}\widetilde{x}_1\\\widetilde{x}_2\end{pmatrix}:\begin{cases}0.4\le \widetilde{x}_1\le 1\\0.4\le \widetilde{x}_2\le 2\end{cases}\right\}$ — прямокутник, сторони якого паралельні координатним осям

Приклад 6. Розглянемо в $\mathbb{R}^2$ множину, обмежену кривою другого порядку:

$$D=\left\{\begin{pmatrix}x_1\\x_2\end{pmatrix}:34(x_1-5)^2+41(x_2+10)^2+24(x_1-5)\ (x_2+10)\le 1\right\}=$$

$$=\left\{\begin{pmatrix}x_1\\x_2\end{pmatrix}:\left(\begin{pmatrix}34&12\\12&41\end{pmatrix}\begin{pmatrix}x_1-5\\x_2+10\end{pmatrix},\begin{pmatrix}x_1-5\\x_2+10\end{pmatrix}\right)\le 1\right\}.$$

Процедурою діагоналізації отримуємо:

$$\begin{pmatrix}34&12\\12&41\end{pmatrix}=\begin{pmatrix}0.6&-0.8\\0.8&0.6\end{pmatrix}\begin{pmatrix}50&0\\0&25\end{pmatrix}\begin{pmatrix}0.6&0.8\\-0.8&0.6\end{pmatrix}.$$

Оскільки матриця $\begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix}$, як і матриця $\begin{pmatrix} 50 & 0 \\ 0 & 25 \end{pmatrix}$, є додатно визначеною, область D є повернутим та зсунутим еліпсом і може бути отримана афінним перетворенням з одиничного круга $\widetilde{D} = \left\{ \begin{pmatrix} \widetilde{x}_1 \\ \widetilde{x}_2 \end{pmatrix} : \widetilde{x}_1^{\,2} + \widetilde{x}_2^{\,2} \le 1 \right\}$, рис. 6.

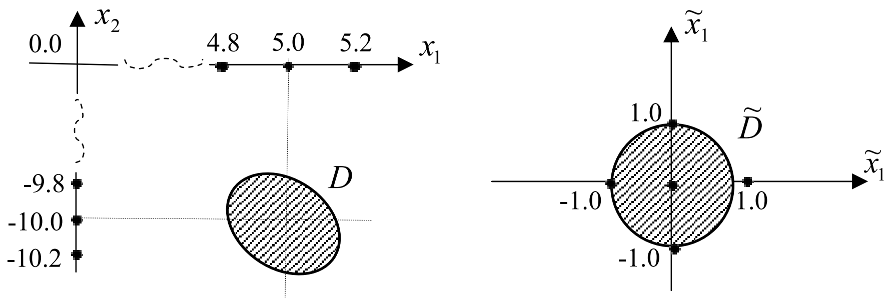

Рис. 6. Область D — еліпс $34(x_1-5)^2 + 41(x_2+10)^2 + 24(x_1-5)\,(x_2+10) \le 1$, область $\widetilde{D} = \left\{ \begin{pmatrix} \widetilde{x}_1 \\ \widetilde{x}_2 \end{pmatrix} : \widetilde{x}_1^{\,2} + \widetilde{x}_2^{\,2} \le 1 \right\}$ — одиничний круг

Еліптичну криву $34(x_1-5)^2 + 41(x_2+10)^2 + 24(x_1-5)\,(x_2+10) = 1$, або, у записі через квадратичну форму, $\left(\begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix}, \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} \right) = 1$, можна отримати з одиничного кола $x_1^{\,2} + x_2^{\,2} = 1$ послідовно зсувом, повертанням та розтягненням:

$$\left(\begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix}, \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} \right) = 1 \Leftrightarrow$$

$$\Leftrightarrow \left(\begin{pmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{pmatrix} \begin{pmatrix} 50 & 0 \\ 0 & 25 \end{pmatrix} \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix}, \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} \right) = 1 \Leftrightarrow$$

$$\Leftrightarrow \left(\begin{pmatrix} 50 & 0 \\ 0 & 25 \end{pmatrix} \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix}, \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} \right) = 1 \Leftrightarrow$$

$$\Leftrightarrow \left(\begin{pmatrix} 5\sqrt{2} & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 5\sqrt{2} & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix}, \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} \right) = 1 \Leftrightarrow$$

$$\Leftrightarrow \left(\begin{pmatrix} 5\sqrt{2} & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix}, \begin{pmatrix} 5\sqrt{2} & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} \right) = 1.$$

Отже, афінне перетворення

$$\begin{pmatrix} \widetilde{x}_1 \\ \widetilde{x}_2 \end{pmatrix} = F\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} 5\sqrt{2} & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} = \begin{pmatrix} 3\sqrt{2} & 4\sqrt{2} \\ -4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 25\sqrt{2} \\ 50 \end{pmatrix}$$

пов'язує одиничний круг $\widetilde{D}=\left\{\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}:\left(\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix},\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}\right)\le 1\right\}=\left\{\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}:\widetilde{x}_1{}^2+\widetilde{x}_2{}^2\le 1\right\}$ і зсунутий повернутий еліпс

$$D=\left\{\begin{pmatrix}x_1\\ x_2\end{pmatrix}:\left(\begin{pmatrix}34 & 12\\ 12 & 41\end{pmatrix}\begin{pmatrix}x_1-5\\ x_2+10\end{pmatrix},\begin{pmatrix}x_1-5\\ x_2+10\end{pmatrix}\right)\le 1\right\}=$$

$$=\left\{\begin{pmatrix}x_1\\ x_2\end{pmatrix}:34(x_1-5)^2+41(x_2+10)^2+24(x_1-5)\ (x_2+10)\le 1\right\}=$$

$$=\left\{\begin{pmatrix}x_1\\ x_2\end{pmatrix}:\left(F\left(\begin{pmatrix}x_1\\ x_2\end{pmatrix}\right),F\left(\begin{pmatrix}x_1\\ x_2\end{pmatrix}\right)\right)\le 1\right\}=\left\{F^{-1}\left(\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}\right):\widetilde{x}_1{}^2+\widetilde{x}_2{}^2\le 1\right\}=F^{-1}(\widetilde{D}).$$

Для матриці A уведемо позначення $\lambda_{\max}=\max\limits_{x\in\mathbb{R}^n:\|x\|=1}(AA^{\mathrm{T}}x,x)$ та $\lambda_{\min}=\min\limits_{x\in\mathbb{R}^n:\|x\|=1}(AA^{\mathrm{T}}x,x)$ відповідно максимальне та мінімальне власне число матриці AA^{T}; нагадаємо, що матриці AA^{T} та $A^{\mathrm{T}}A$ мають однаковий набір додатних (завдяки невиродженості A) власних чисел.

Лема 5. Нехай функція $f:\mathbb{R}^n\to\mathbb{R}$ — опукла (строго опукла, сильно опукла з параметром $\theta>0$). Тоді функція $\widetilde{f}:\mathbb{R}^n\to\mathbb{R}$, $\widetilde{f}(\widetilde{x})=$ $=f(F^{-1}(\widetilde{x}))=f(A^{-1}\widetilde{x}-A^{-1}b)$ теж опукла (відповідно строго опукла, сильно опукла з параметром $\theta\lambda_{\max}^{-1}>0$).

Доведення. 1. Нехай функція $f:\mathbb{R}^n\to\mathbb{R}$ опукла. Зафіксувавши $\lambda\in[0,1]$, дістаємо:

$$\widetilde{f}(\lambda\widetilde{x}+(1-\lambda)\widetilde{y})=f(\lambda(A^{-1}\widetilde{x}-A^{-1}b)+(1-\lambda)(A^{-1}\widetilde{y}-A^{-1}b))\le$$
$$\le\lambda f(A^{-1}\widetilde{x}-A^{-1}b)+(1-\lambda)f(A^{-1}\widetilde{y}-A^{-1}b)=\lambda\widetilde{f}(\widetilde{x})+(1-\lambda)\widetilde{f}(\widetilde{y})$$

для довільних $\widetilde{x}\in\mathbb{R}^n,\widetilde{y}\in\mathbb{R}^n$, що доводить опуклість функції g. □

2. Нехай функція $f:\mathbb{R}^n\to\mathbb{R}$ строго опукла; $\widetilde{x}\in\mathbb{R}^n,\widetilde{y}\in\mathbb{R}^n$ і $\widetilde{x}\ne\widetilde{y}$. За умовою, матриця A невироджена, а отже, матриця A^{-1} існує і невироджена, звідки $A^{-1}\widetilde{x}-A^{-1}b\ne A^{-1}\widetilde{y}-A^{-1}b$. Зафіксувавши $\lambda\in(0,1)$, маємо

$$\widetilde{f}(\lambda\widetilde{x}+(1-\lambda)\widetilde{y})=f(\lambda(A^{-1}\widetilde{x}-A^{-1}b)+(1-\lambda)(A^{-1}\widetilde{y}-A^{-1}b))<$$
$$<\lambda f(A^{-1}\widetilde{x}-A^{-1}b)+(1-\lambda)f(A^{-1}\widetilde{y}-A^{-1}b)=\lambda\widetilde{f}(\widetilde{x})+(1-\lambda)\widetilde{f}(\widetilde{y})$$

для довільних $\widetilde{x}\in\mathbb{R}^n$, $\widetilde{y}\in\mathbb{R}^n$, що доводить строгу опуклість функції $\widetilde{f}$. □

3. Нехай функція $f:\mathbb{R}^n\to\mathbb{R}$ сильно опукла з параметром $\theta>0$. Зазначимо, що матриця AA^T симетрична і, завдяки невиродженості матриці A, додатно визначена. Це означає, що всі власні числа матриці AA^T дійсні до-

датні числа. Очевидно, $\lambda_{\max}^{-1} = \min\limits_{x\in\mathbb{R}^n:\|x\|=1} ((AA^T)^{-1}x, x)$ — мінімальне власне число матриці $(AA^T)^{-1}$. Отже, зафіксувавши $\lambda \in [0,1]$, отримуємо:

$$\begin{gathered}\widetilde{f}(\lambda\widetilde{x} + (1-\lambda)\widetilde{y}) = f(\lambda(A^{-1}\widetilde{x} - A^{-1}b) + (1-\lambda)(A^{-1}\widetilde{y} - A^{-1}b)) \leq \\ \leq \lambda f(A^{-1}\widetilde{x} - A^{-1}b) + (1-\lambda) f(A^{-1}\widetilde{y} - A^{-1}b) - \lambda(1-\lambda)\tfrac{\theta}{2} \| A^{-1}(\widetilde{x} - \widetilde{y}) \|^2 = \\ = \lambda\widetilde{f}(\widetilde{x}) + (1-\lambda)\widetilde{f}(\widetilde{y}) - \lambda(1-\lambda)\tfrac{\theta}{2}(A^{-1}(\widetilde{x} - \widetilde{y}), A^{-1}(\widetilde{x} - \widetilde{y})) = \\ = \lambda\widetilde{f}(\widetilde{x}) + (1-\lambda)\widetilde{f}(\widetilde{y}) - \lambda(1-\lambda)\tfrac{\theta}{2}((AA^T)^{-1}(\widetilde{x} - \widetilde{y}), (\widetilde{x} - \widetilde{y})) \leq \\ \leq \lambda\widetilde{f}(\widetilde{x}) + (1-\lambda)\widetilde{f}(\widetilde{y}) - \lambda(1-\lambda)\tfrac{\theta}{2}\lambda_{\max}^{-1}\end{gathered}$$

для довільних $\widetilde{x} \in \mathbb{R}^n$, $\widetilde{y} \in \mathbb{R}^n$, що доводить сильну опуклість функції $\widetilde{f}$ з параметром $\theta\lambda_{\max}^{-1}$.

Приклад 7. Нехай $f: \mathbb{R}^n \to \mathbb{R}$, $f(x) = \| x \|^2$. Функція $f(x) = \| x \|^2$ сильно опукла з параметром $\theta = 2$ (див. приклад 3), а отже, функція $\widetilde{f}(\widetilde{x}) = \| F^{-1}(\widetilde{x}) \|^2 = \| A^{-1}\widetilde{x} - A^{-1}b \|^2$ є сильно опуклою з параметром $\theta\lambda_{\max}^{-1} > 0$. Цікаво, що у випадку $b = 0$ білінійний функціонал $\langle \widetilde{x}, \widetilde{y} \rangle = ((AA^{\mathrm{T}})^{-1}\widetilde{x}, \widetilde{y})$, завдяки симетричності та додатній визначеності матриці $(AA^{\mathrm{T}})^{-1}$, визначає інший скалярний добуток, а функція $\langle\langle \widetilde{x}, \widetilde{x} \rangle\rangle = g(\widetilde{x}) = \| F^{-1}(\widetilde{x}) \|^2 = \| A^{-1}\widetilde{x} \|^2$ іншу норму в $\mathbb{R}^n$.

У [4] наведено теорему про збереження опуклості функції (без умови строгої чи сильної опуклості) за застосування до аргумента довільного перетворення $Ax + b$.

ПРОЕКЦІЯ ТОЧКИ НА МНОЖИНУ

Означення 5. Проекцією точки $a \in \mathbb{R}^n$ на непорожню множину $D \subset \mathbb{R}^n$ називають точку $\mathrm{Pr}_D\, a \in D$, найближчу до a серед усіх точок $x \in D$:

$$\| \mathrm{Pr}_D\, a - a \| = \inf_{x\in D} \| x - a \|. \tag{4}$$

Оскільки норма $\| \cdot \|$ невід'ємна, рівність (4) можна подати через квадрат норми $\| \cdot \|^2$:

$$\| \mathrm{Pr}_D\, a - a \|^2 = \inf_{x\in D} \| x - a \|^2. \tag{5}$$

Очевидно, $\mathrm{Pr}_D\, a = a$ тоді й тільки тоді, коли $a \in D$.

Приклад 8. Для кола $D = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : (x_1 - x_{0,1})^2 + (x_2 - x_{0,2})^2 = r^2 \right\} \subset \mathbb{R}^2$ радіуса $r > 0$ із центром у $x_0 = \begin{pmatrix} x_{0,1} \\ x_{0,2} \end{pmatrix}$ проекцію точки $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \neq \begin{pmatrix} x_{0,1} \\ x_{0,2} \end{pmatrix}$

можна обчислити за формулою $\mathrm{Pr}_D\, a = x_0 + \dfrac{r}{\| a - x_0 \|}(a - x_0)$, тобто $\mathrm{Pr}_D\, a$ міститься на перетині відрізка $[x_0, a]$ з колом D (див. рис.7, *а*, *б*).; проекцією центра кола $a = x_0$ є кожна точка кола (рис. 7, *в*).

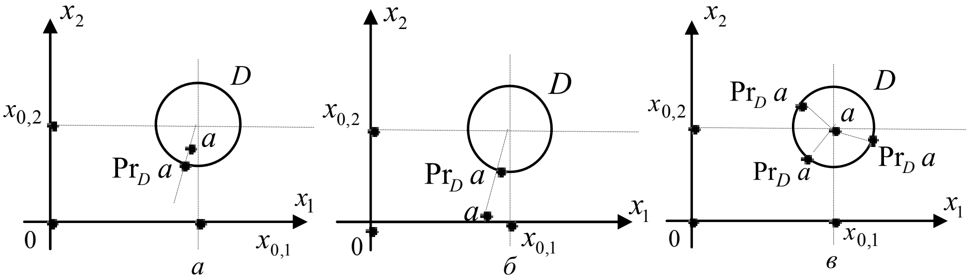

Рис. 7. Проекція точки на коло

Приклад 9. Для відкритого круга

$$D = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : (x_1 - x_{0,1})^2 + (x_2 - x_{0,2})^2 < r^2 \right\} \subset \mathbb{R}^2$$

радіуса $r > 0$ із центром в $x_0 = \begin{pmatrix} x_{0,1} \\ x_{0,2} \end{pmatrix}$ проекції точки $a \notin D$ ($\| a \| \geq r$) не існує. Дійсно, $\inf\limits_{x \in D} \| x - a \| = | \| x_0 - a \| + r$ міг би досягатись лише в точці $x_0 + \frac{r}{\|a - x_0\|}(a - x_0)$ (рис. 8), але $x_0 + \frac{r}{\|a - x_0\|}(a - x_0) \notin D$, оскільки $\| (x_0 + \frac{r}{\|a - x_0\|}(a - x_0)) - x_0 \| = \| (\frac{r}{\|a - x_0\|}(a - x_0) \| = $ $= \frac{r}{\|a - x_0\|} \| a - x_0 \| = r$.

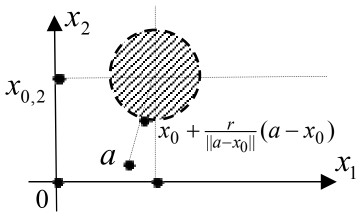

Рис. 8. Проекція точки на відкритий круг не існує

У загальному випадку точка $a \in \mathbb{R}^n$ може мати кілька проекцій на множину $D \subset \mathbb{R}^n$ (приклад 8) або не мати жодної (приклад 9). Проте відомий класичний факт (див., напр., [1, 4, 5]), який наводимо як лему.

Лема 6. 1. Нехай множина $D \subset \mathbb{R}^n$ непорожня і замкнена. Тоді будь-яка точка $a \in \mathbb{R}^n$ має принаймні одну проекцію на D.

2. Нехай множина $D \subset \mathbb{R}^n$ опукла. Тоді будь-яка точка $a \in \mathbb{R}^n$ має не більше однієї проекції на D.

Доведення. 1. Проекція $\mathrm{Pr}_D\, a$ на замкнену множину D існує за теоремою Веєрштраса: неперервна функція $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ досягає свого найменшого значення $\inf\limits_{x \in D} \| x - a \| = \min\limits_{x \in D} \| x - a \|$ на непорожній замкненій множині $D \subset \mathbb{R}^n$.

2. Функція $\| \cdot \|^2 : \mathbb{R}^n \to \mathbb{R}$ є строго (і навіть сильно) опуклою (див. приклад 3) і, за лемою 2, має не більше одного глобального мінімуму на опуклій множині $D \subset \mathbb{R}^n$. □

Приклад 10. 1. Нехай $D=\{x:\|x-x_0\|\le r\}\subset\mathbb{R}^2$ замкнений круг радіуса $r>0$ із центром в $x_0\in\mathbb{R}^2$. Очевидно, D опукла і замкнена і, за лемою 6, для будь-якого $a\in\mathbb{R}^2$ існує єдина проекція $\mathrm{Pr}_D\, a=$ $=\begin{cases} a,\ a\in D; \\ x_0+\frac{r}{\|a-x_0\|}(a-x_0),\ a\notin D \end{cases}$ точки $a\in\mathbb{R}^2$ на множину D (див. приклад 8).

2. Нехай $D=\{x: m_i\le x_i\le M_i\,(i=1{,}2)\}\subset\mathbb{R}^2$ прямокутник (рис. 9), визначений точками $m\in\mathbb{R}^2$, $M\in\mathbb{R}^2$, $m_i\le M_i\,(i=1{,}2)$. Очевидно, D опукла і замкнена, і, за лемою 6, для будь-якого $a\in\mathbb{R}^2$ існує єдина проекція $\mathrm{Pr}_D\, a$ $(i=1{,}2)$:

$$(\mathrm{Pr}_D\, a)_i=\begin{cases} m_i; \\ a_i, m_i<a_i<M_i; \\ M_i, a_i\ge M_i, \end{cases}$$

Формули для проекцій на круг і прямокутник (сторони якого паралельні координатним осям), наведені у прикладі 10, природним чином узагальнюються на кулю та паралелепіпед у $\mathbb{R}^n$. Детальніше про проекції на деякі прості множини $D\subset\mathbb{R}^n$ (куля, паралелепіпед, півпростір та ін.) див., напр. [5]. Однак для проекцій на більш складні множини (зокрема, на еліпс) надати прості явні формули неможливо.

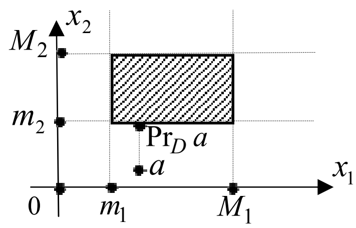

Рис. 9. Проекція точки на прямокутник

Приклад 11. Нехай

$$D=\left\{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}: 34(x_1-5)^2+41(x_2+10)^2+24(x_1-5)\,(x_2+10)\le 1\right\}$$

повернутий еліпс (див. приклад 6), $a=\begin{pmatrix} 5 \\ -9.7 \end{pmatrix}$. Оскільки $a\notin D$, проекція $\mathrm{Pr}_D\, a$ лежить на еліптичній кривій $34(x_1-5)^2+41(x_2+10)^2+$ $+24(x_1-5)\,(x_2+10)=1$, яка обмежує область D. Застосовуючи до пошуку екстремуму у рівності (5) метод множників Лагранжа, запишемо відповідну функцію Лагранжа:

$$L(x_1,x_2,\lambda)=(x_1-5)^2+(x_2+9.7)^2-\lambda(34(x_1-5)^2+$$
$$+41(x_2+10)^2+24(x_1-5)\,(x_2+10)-1).$$

Мінімум функції $L(x_1,x_2,\lambda)$ можливий лише в нулях часткових похідних:

$$L'_{x_1}, L'_{x_2}, L'_\lambda:\begin{cases} 2(x_1-5)-\lambda(68(x_1-5)+24(x_2+10))=0 \\ 2(x_2+9.7)-\lambda(24(x_1-5)+82(10+x_2))=0 \\ 34(x_1-5)^2+41(x_2+10)^2+24(x_1-5)\,(x_2+10)-1=0 \end{cases} \qquad (6)$$

Легко перевірити, що аналітичне розв'язання системи (6) зводиться до рівняння 4-го степеня відносно λ — такі рівняння, хоч і допускають явний розв'язок, але відповідні формули складні і містять радикали до 4-го степеня включно. Тому раціональним бачиться чисельне розв'язання системи (6), у результаті чого, після відкидання «зайвих» коренів (які не можуть бути мінімумом правої частини рівності (5)), отримуємо: $\mathrm{Pr}_D\, a \approx \begin{pmatrix} 4.9757 \\ -9.8381 \end{pmatrix}$ (множник $\lambda \approx -0.0218$, але це значення безпосередньо не входить до $\mathrm{Pr}_D\, a$), рис. 10.

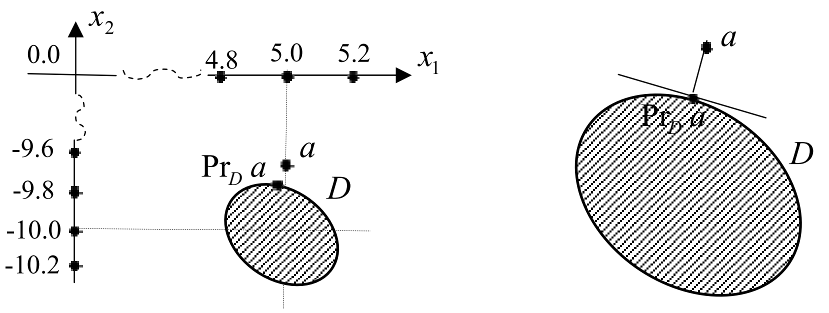

Рис. 10. Проекція точки на еліпс $34(x_1-5)^2 + 41(x_2+10)^2 + 24(x_1-5)\,(x_2+10) \le 1$

Приклад 12. Нехай $D = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \begin{cases} 2 \le 3x_1 + 4x_2 \le 5 \\ 2 \le -4x_1 + 3x_2 \le 10 \end{cases} \right\}$ повернутий прямокутник (див. приклад 5), $a = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$.

Для обчислення $\mathrm{Pr}_D\, a$ проаналізуємо розташування точки a відносно пар паралельних прямих $3x_1 + 4x_2 = 2$ і $3x_1 + 4x_2 = 5$ та $-4x_1 + 3x_2 = 2$ і $-4x_1 + 3x_2 = 10$. Оскільки $3 \cdot 1 + 4 \cdot 4 = 19 \ge 5$, проекція $\mathrm{Pr}_D\, a$ лежить на прямій $3x_1 + 4x_2 = 5$. Оскільки $2 < -4 \cdot 1 + 3 \cdot 4 = 8 < 10$, проекція $\mathrm{Pr}_D\, a$ лежить на прямій, яка проходить через точку a та паралельна прямим $-4x_1 + 3x_2 = 2$ і $-4x_1 + 3x_2 = 10$, тобто на прямій $-4(x_1 - 1) + 3(x_2 - 4) = 0$, або, в іншому вигляді, на прямій $-4x_1 + 3x_2 = 8$. Розв'язуючи систему $\begin{cases} 3x_1 + 4x_2 = 5 \\ -4x_1 + 3x_2 = 8 \end{cases}$, дістаємо: $\mathrm{Pr}_D\, a = \begin{pmatrix} -\frac{17}{25} \\ \frac{44}{25} \end{pmatrix} = \begin{pmatrix} -0.68 \\ 1.76 \end{pmatrix}$, рис. 11.

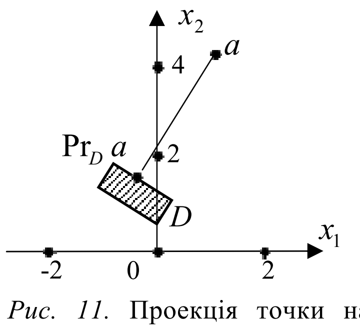

Рис. 11. Проекція точки на прямокутник $\begin{cases} 2 \le 3x_1 + 4x_2 \le 5 \\ 2 \le -4x_1 + 3x_2 \le 10 \end{cases}$

Зауваження 3. У прикладах 11 і 12 вектор $a - \mathrm{Pr}_D\, a$ ортогональний дотичній у точці a до кривої, що обмежує область D (рис. 10 та 11 відповідно). Така властивість є загальною і поширюється на множини D з негладкою межею (див., напр., [1, 4, 5]).

АФІННЕ ПЕРЕТВОРЕННЯ В ОПЕРАТОРІ ПРОЕКТУВАННЯ

Нехай $D \subset \mathbb{R}^n$ — непорожня опукла та замкнена множина, a довільна точка в $\mathbb{R}^n$. Згідно з лемою 6, проекція $\mathrm{Pr}_D\, a$ існує і єдина, тобто інфімум у рівності (4) досягається у єдиній точці $\mathrm{Pr}_D\, a = \arg\min\limits_{x \in D} \| x - a \|$.

Розглянемо заміну змінних $\widetilde{x} = F(x)$ за афінним перетворенням $F: \mathbb{R}^n \to \mathbb{R}^n$, $F(x) = Ax + b$ (A —невироджена матриця розмірності $n \times n$, $b \in \mathbb{R}^n$). Позначивши, як і в п. 0, $\widetilde{D} = \{F(x) : x \in D\} = F(D)$, $\widetilde{x} = F(x)$, $\widetilde{a} = F(a)$, уведемо оператор $\mathrm{Pr}_D^F\, a$, пов'язаний з обчисленням проекції $\mathrm{Pr}_{\widetilde{D}}\, \widetilde{a}$ з подальшою заміною змінних $\widetilde{x} = F(x)$:

$$\mathrm{Pr}_D^F\, a = F^{-1}(\mathrm{Pr}_{\widetilde{D}}\, \widetilde{a}) \tag{7}$$

Очевидно, множина $\widetilde{D} = F(D)$ є непорожньою, за лемою 4 $\widetilde{D} = F(D)$ є опуклою і, завдяки замкненості множини D та неперервності відображення F — замкненою. Таким чином, проекція $\mathrm{Pr}_{\widetilde{D}}\, \widetilde{a}$ за лемою 6 існує і єдина.

Ураховуючи визначення оператора $\mathrm{Pr}_D\, a$, подамо рівність (7) через $\arg\min$:

$$\mathrm{Pr}_D^F\, a = F^{-1}(\arg\min_{\widetilde{x} \in \widetilde{D}} \| \widetilde{x} - \widetilde{a} \|) = \arg\min_{\widetilde{x} \in \mathbb{R}^n : F^{-1}(\widetilde{x}) \in F^{-1}(\widetilde{D})} \| F(x) - F(a) \| =$$

$$= \arg\min_{x \in D} \| (Ax + b) - (Aa + b) \| = \arg\min_{x \in D} \| A(x - a) \| = \arg\min_{x \in D} \| x - a \|_F .$$

Отже, оператор $\mathrm{Pr}_D^F\, a$, як і оператор $\mathrm{Pr}_D\, a$, визначає на D точку, найближчу до a, але найближчу — у сенсі метрики, визначеної нормою $\| x \|_F = \| A x \|$ (див. також приклад 7). Зазначимо, що норма $\| x \|_F = \| A x \|$ (а отже, і оператор $\mathrm{Pr}_D^F\, a$) визначається матрицею A і не залежить від b. Очевидно, що у випадку ортогональної матриці A маємо рівність норм:

$$\| x \|_F{}^2 = \| A x \|^2 = (Ax, Ax) = (x, x) = \| x \|^2 ,$$

а отже, $\mathrm{Pr}_D^F\, a = \mathrm{Pr}_D\, a$.

Приклад 13. Розглянемо в $\mathbb{R}^2$ еліпс (див. приклади 6 і 11):

$$D = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : 34(x_1 - 5)^2 + 41(x_2 + 10)^2 + 24(x_1 - 5)\ (x_2 + 10) \le 1 \right\} =$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \left(\begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix} \begin{pmatrix} x_1 - 5 \\ x_2 + 10 \end{pmatrix}, \begin{pmatrix} x_1 - 5 \\ x_2 + 10 \end{pmatrix} \right) \le 1 \right\} =$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \left(\begin{pmatrix} 3\sqrt{2} & 4\sqrt{2} \\ -4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 25\sqrt{2} \\ 50 \end{pmatrix}, \begin{pmatrix} 3\sqrt{2} & 4\sqrt{2} \\ -4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 25\sqrt{2} \\ 50 \end{pmatrix} \le 1 \right) \right\} .$$

Очевидно (див. також приклад 6), еліпс D пов'язаний з одиничним кругом $\widetilde{D}=\left\{\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}: \widetilde{x}_1^{\,2}+\widetilde{x}_2^{\,2}\le 1\right\}$ афінним перетворенням $\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}=F\begin{pmatrix}x_1\\ x_2\end{pmatrix}=$ $=\begin{pmatrix}3\sqrt{2} & 4\sqrt{2}\\ -4 & 3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}+\begin{pmatrix}25\sqrt{2}\\ 50\end{pmatrix}$: $D=F^{-1}(\widetilde{D})$, або $\widetilde{D}=F(D)$.

Обчислимо $\mathrm{Pr}_D^F\, a$ для $a=\begin{pmatrix}5\\ -9.7\end{pmatrix}$. Оскільки $\widetilde{a}=F(a)\approx\begin{pmatrix}1{,}70\\ 0.90\end{pmatrix}\notin\widetilde{D}$, отримуємо (див. приклад 10, п. 1):

$$\mathrm{Pr}_{\widetilde{D}}\,\widetilde{a}=\begin{cases}\widetilde{a},\ \widetilde{a}\in\widetilde{D};\\ 0+\frac{1}{\|\widetilde{a}-0\|}(\widetilde{a}-0),\ \widetilde{a}\notin\widetilde{D}\end{cases}=\frac{\widetilde{a}}{\|\widetilde{a}\|}\approx\begin{pmatrix}0.46\\ 0.89\end{pmatrix},$$

звідки остаточно маємо:

$$\mathrm{Pr}_D^F a=F^{-1}(\mathrm{Pr}_{\widetilde{D}}\,\widetilde{a})=\begin{pmatrix}3\sqrt{2} & 4\sqrt{2}\\ -4 & 3\end{pmatrix}^{-1}\left(\frac{\widetilde{a}}{\|\widetilde{a}\|}-\begin{pmatrix}25\sqrt{2}\\ 50\end{pmatrix}\right)=$$

$$=\frac{1}{50}\begin{pmatrix}3\sqrt{2} & -8\\ 4\sqrt{2} & 6\end{pmatrix}\left(\frac{\widetilde{a}}{\|\widetilde{a}\|}-\begin{pmatrix}25\sqrt{2}\\ 50\end{pmatrix}\right)\approx\begin{pmatrix}5\\ -9.84\end{pmatrix}.$$

Зауважимо, що $\mathrm{Pr}_D^F\, a\ne\mathrm{Pr}_D\, a$ (рис. 12).

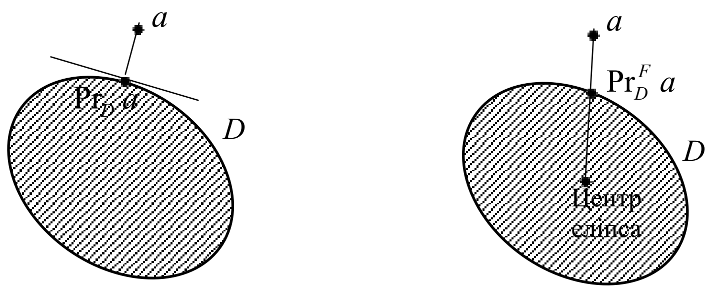

Рис. 12. Оператори $\mathrm{Pr}_D\, a$ і $\mathrm{Pr}_D^F\, a$ для еліпса $34(x_1-5)^2+41(x_2+10)^2+24(x_1-5)\times$ $\times(x_2+10)\le 1$

Приклад 14. Розглянемо в $\mathbb{R}^2$ повернутий прямокутник (див. також приклади 5 та 12) $D=\left\{\begin{pmatrix}x_1\\ x_2\end{pmatrix}:\begin{cases}2\le 3x_1+4x_2\le 5\\ 2\le -4x_1+3x_2\le 10\end{cases}\right\}$. Оскільки $\begin{pmatrix}3x_1+4x_2\\ -4x_1+3x_2\end{pmatrix}=5\begin{pmatrix}0.6 & 0.8\\ -0.8 & 0.6\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}$, область D пов'язана афінним перетворенням $\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}=F\left(\begin{pmatrix}x_1\\ x_2\end{pmatrix}\right)=\begin{pmatrix}0.6 & 0.8\\ -0.8 & 0.6\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}$ із прямокутником $\widetilde{D}=$ $=\left\{\begin{pmatrix}\widetilde{x}_1\\ \widetilde{x}_2\end{pmatrix}:\begin{cases}0.4\le\widetilde{x}_1\le 1\\ 0.4\le\widetilde{x}_2\le 2\end{cases}\right\}=F(D)$, сторони якого паралельні координатним осям (детальніше див. приклад 5). У цьому випадку, оскільки матриця $A = \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix}$ є ортогональною (матрицею повороту), оператор $\mathrm{Pr}_D^F\, a$ збігається із проекцією: $\mathrm{Pr}_D^F\, a = \mathrm{Pr}_D\, a$ для довільного $a \in \mathbb{R}^n$ (див. приклад 12, рис. 11).

МЕТОД ПРОЕКЦІЇ ГРАДІЄНТА З АФІННИМ ПЕРЕТВОРЕННЯМ

Нехай $D \subset \mathbb{R}^n$ — непорожня опукла та замкнена множина, $f : \mathbb{R}^n \to \mathbb{R}$ диференційовна строго опукла функція на $\mathbb{R}^n$ За сформульованих умов, згідно з лемою 3, існує єдиний мінімум функції f на D: $x^* = \arg\min\limits_{x \in D} f(x)$. Аналітичне обчислення x^* у практичних випадках є складним і почасти неможливим. Для чисельного розв'язання задачі $x^* = \arg\min\limits_{x \in D} f(x)$ у 1964 р. американським вченим А. Голдштейном був запропонований (див. [7]) ітераційний метод, нині відомий як метод проекції градієнта:

- $x^0 \in \mathbb{R}^n$ — обрана початкова точка;
- $x^{k+1} = \mathrm{Pr}_D(x^k + \alpha^k h^k)$ ($k \ge 0$), де $h^k = -f'(x^k)$ напрямок кроку (антиградієнт), $a^k \in \mathbb{R}^n$ — довжина кроку (обирається залежно від модифікації методу).

Одним з найпоширеніших методів вибору довжини кроку є метод найшвидшого спуску: $\alpha^k = \arg\min\limits_{\alpha > 0} f(x^k + \alpha h^k)$, який, разом з додатковими умовами на f і D, забезпечує збіжність $x^k \xrightarrow[k \to +\infty]{} x^*$; сформулюємо відомий (див., напр. [4]) результат у вигляді теореми.

Теорема 1. Нехай $D \subset \mathbb{R}^n$ — непорожня опукла та замкнена множина, $f : \mathbb{R}^n \to \mathbb{R}$ функція на $\mathbb{R}^n$, яка задовольняє умови:

- неперервна диференційовність на $\mathbb{R}^n$;
- сильна опуклість з параметром $\theta > 0$;
- градієнт є ліпшицевим з константою L, тобто $\| f'(x) - f'(y) \| \le \le L \| x - y \|$.

Тоді рекурентно задана послідовність

- $x^0 \in \mathbb{R}^n$ довільно обрана початкова точка,
- $x^{k+1} = \mathrm{Pr}_D(x^k + \alpha^k h^k)$ ($k \ge 0$), де $h^k = -f'(x^k)$, $\alpha^k = = \arg\min\limits_{\alpha > 0} f(x^k + \alpha h^k)$ збігається до x^* з геометричною швидкістю: $x^k \xrightarrow[k \to +\infty]{} x^*$, $\| x^{k+1} - x^* \| \le q \| x^k - x^* \|$ ($k \ge k_0$), де $q \in [0,1)$ — фіксована константа, k_0 — фіксований номер.

Доведення див., напр. у [4].

Зауваження 4. Очевидно, умова геометричної швидкості (або лінійної швидкості) $\| x^{k+1} - x^* \| \le q \| x^k - x^* \|$ ($k \ge k_0$) еквівалентна умові $\| x^k - x^* \| \le$ $\le Cq^k$ ($k \ge 0$), де $q \in [0,1)$, $C \ge 0$ — фіксовані константи.

Метод проекції градієнта пов'язаний з обчисленням проекції точки x^k на множину D, що нескладно у простих випадках (наприклад, якщо D круг або прямокутник, див. приклад 10), проте ускладнюється до окремої задачі мінімізації для складніших випадків (наприклад, якщо D еліпс, див. приклад 11).

Для спрощення задачі проектування застосуємо у задачі пошуку $x^* = \arg\min_{x \in D} f(x)$ заміну змінних, пов'язану з афінним перетворенням $F: \mathbb{R}^n \to \mathbb{R}^n$, $F(x) = Ax + b$ (A —невироджена матриця розмірності $n \times n$, $b \in \mathbb{R}^n$). Позначивши, як і в п. 2, $\widetilde{D} = \{F(x) : x \in D\} = F(D)$, $\widetilde{x} = F(x)$, $\widetilde{f}(\widetilde{x}) = f(F^{-1}(\widetilde{x})) = f(A^{-1}\widetilde{x} - A^{-1}b)$, отримуємо область $\widetilde{D} \subset \mathbb{R}^n$ і функцію $\widetilde{f}: \mathbb{R}^n \to \mathbb{R}$. Таким чином, пошук $x^* = \arg\min_{x \in D} f(x)$ можна звести до пошуку $\widetilde{x}^* = \arg\min_{\widetilde{x} \in \widetilde{D}} \widetilde{f}(\widetilde{x})$, $\widetilde{x}^* = F^{-1}(x^*)$, з подальшим «поверненням» до $x = F^{-1}(\widetilde{x})$.

Теорема 2. Нехай функція $f: \mathbb{R}^n \to \mathbb{R}$ і область $D \subset \mathbb{R}^n$ задовольняють умови теореми 1, $x^{F,0} \in \mathbb{R}^n$ — довільна початкова точка, $x^{F,k+1} == \mathrm{Pr}_D^F(F^{-1}(F(x^{F,k}) + \widetilde{\alpha}^k \widetilde{h}^k))$ ($k \ge 0$), де $\widetilde{h}^k = -(A^T)^{-1} f'(x^{F,k})$, $\widetilde{\alpha}^k = \arg\min_{\widetilde{\alpha} > 0} f(x^{F,k} + \widetilde{\alpha}\widetilde{h}^k)$. Тоді послідовність $x^{F,k}$ ($k \ge 0$) збігається до $x^* = \arg\min_{x \in D} f(x)$ зі швидкістю геометричної прогресії:

$$x^{F,k} \xrightarrow[k \to +\infty]{} x^*, \ \| x^{F,k} - x^* \| \le C(q^F)^k \ (k \ge 0),$$

де $q^F \in [0,1)$, $C \ge 0$ — фіксовані константи.

Доведення. Розглянемо (див. також п. 2 і 4) отримані заміною змінних $\widetilde{x} = F(x)$ область $\widetilde{D} = \{\widetilde{x} = F(x) : x \in D\} = F(D)$ та функцію $\widetilde{f}(\widetilde{x}) =$ $= f(F^{-1}(\widetilde{x})) = f(A^{-1}\widetilde{x} - A^{-1}b)$.

Оскільки область D за умовою непорожня, опукла та замкнена, область $\widetilde{D} = F(D)$ за побудовою також є непорожньою, за лемою 4 опуклою, за неперервністю відображення F замкненою. Оскільки f за умовою є сильно опуклою з параметром $\theta > 0$, функція $\widetilde{f}$ за лемою 5 є сильно опуклою з параметром $\theta\lambda_{\max}^{-1} > 0$. Функція f за умовою є неперервно диференційовною, тоді функція $\widetilde{f}$ теж є диференційовною як композиція диференційовних відображень f і F^{-1}. Для обчислення $\widetilde{f}'$ знайдемо $(\widetilde{f}', d)$ для довільного фіксованого $d \in \mathbb{R}^n$:

$$(\widetilde{f}'(\widetilde{x}),d)=(f'(A^{-1}\widetilde{x}-A^{-1}b),A^{-1}d)=((A^T)^{-1}f'(A^{-1}\widetilde{x}-A^{-1}b),d),$$

звідки, враховуючи довільність $d\in\mathbb{R}^n$, отримуємо, що $\widetilde{f}'(\widetilde{x})=$ $=((A^T)^{-1}f'(F^{-1}(\widetilde{x}))$ є неперервним, і функція $\widetilde{f}$ є неперервно диференційовною. Нарешті, із ліпшицевості f' з константою L, для $\widetilde{f}'$ маємо:

$$\|\widetilde{f}'(\widetilde{x})-\widetilde{f}'(\widetilde{y})\|^2=\|((A^T)^{-1}f'(A^{-1}\widetilde{x}-A^{-1}b)-((A^T)^{-1}f'(A^{-1}\widetilde{y}-A^{-1}b)\|^2=$$
$$=((A^TA)^{-1}(f'(A^{-1}\widetilde{x}-A^{-1}b)-f'(A^{-1}\widetilde{y}-A^{-1}b)),$$
$$f'(A^{-1}\widetilde{x}-A^{-1}b)-f'(A^{-1}\widetilde{y}-A^{-1}b))\le$$
$$\le\lambda_{\min}\|f'(A^{-1}\widetilde{x}-A^{-1}b)-f'(A^{-1}\widetilde{y}-A^{-1}b))\|^2\le$$
$$\le\lambda_{\min}L^2\|A^{-1}(\widetilde{x}-\widetilde{y})\|^2\le\lambda_{\min}^2L^2\|\widetilde{x}-\widetilde{y}\|^2,$$

де $\lambda_{\min}=\min\limits_{x\in\mathbb{R}^n:\|x\|=1}(A^TAx,x)$ — мінімальне власне число матриці AA^T. Таким чином, $\widetilde{f}'$ ліпшицевий з константою $\lambda_{\min}L$. Отже, область $\widetilde{D}$ і функція $\widetilde{f}$ задовольняють умови теореми 1, тобто рекурентно задана послідовність $\widetilde{x}^k$ ($k\ge0$), де $\widetilde{x}^0\in\mathbb{R}^n$ довільно обрана початкова точка, $\widetilde{x}^{k+1}=\mathrm{Pr}_{\widetilde{D}}(\widetilde{x}^k+\widetilde{\alpha}^k\widetilde{h}^k)$, $\widetilde{h}^k=-\widetilde{f}'(\widetilde{x}^k)=-(A^T)^{-1}f'(F^{-1}(\widetilde{x}^k))$, $\widetilde{\alpha}^k=\arg\min\limits_{\widetilde{\alpha}>0}\widetilde{f}(\widetilde{x}^k+\widetilde{\alpha}\widetilde{h}^k)$ збігається до x^* з геометричною швидкістю: $\widetilde{x}^k\xrightarrow[k\to+\infty]{}\widetilde{x}^*$, $\|\widetilde{x}^{k+1}-\widetilde{x}^*\|\le\widetilde{q}\,\|\widetilde{x}^k-\widetilde{x}^*\|$ ($k\ge0$), де $\widetilde{q}\in[0,1)$ фіксована константа. Увівши позначення $x^{F,k}=$ $=F^{-1}(\widetilde{x}^k)$, дістаємо: $x^{F,0}=F^{-1}(\widetilde{x}^0)$ (початкова точка),

$$x^{F,k+1}=F^{-1}(\widetilde{x}^{k+1})=F^{-1}(\mathrm{Pr}_{\widetilde{D}}(\widetilde{x}^k+\widetilde{\alpha}^k\widetilde{h}^k)=\mathrm{Pr}_D^F(F^{-1}(F(x^{F,k})+\widetilde{\alpha}^k\widetilde{h}^k)).$$

Оскільки за лемою 3 існує єдиний $\widetilde{x}^*=\arg\min\limits_{\widetilde{x}\in\widetilde{D}}\widetilde{f}(\widetilde{x})$, за побудовою маємо зв'язок $\widetilde{x}^*=F(x^*)$, звідки $x^{F,k}\xrightarrow[k\to+\infty]{}x^*$. Нарешті, оцінимо швидкість збіжності $x^{F,k}\xrightarrow[k\to+\infty]{}x^*$:

$$\|x^{F,k+1}-x^*\|\le\|F^{-1}(\widetilde{x}^{k+1})-F^{-1}(\widetilde{x}^*)\|=\|A^{-1}(\widetilde{x}^{k+1}-\widetilde{x}^*)\|=$$
$$=((AA^T)^{-1}(\widetilde{x}^{k+1}-\widetilde{x}^k),\widetilde{x}^{k+1}-\widetilde{x}^*)\le$$
$$\le\frac{1}{\lambda_{\min}}\|\widetilde{x}^{k+1}-\widetilde{x}^*\|\le\frac{\widetilde{q}}{\lambda_{\min}}\|\widetilde{x}^k-\widetilde{x}^*\|=C\widetilde{q}^k,$$

тобто $x^{F,k}\xrightarrow[k\to+\infty]{}x^*$ із лінійною швидкістю, що завершує доведення теореми. □

Зауваження 5. Послідовності x^k ($k\ge0$) та $x^{F,k}$ ($k\ge0$) мають однакову границю x^*, однак, взагалі кажучи, поточково розрізняються, тобто найчастіше $x^k\ne x^{F,k}$.

Приклад 15. Нехай $f:\mathbb{R}^2 \to \mathbb{R}$, $f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = (x_1-6)^2 + (x_2+11)^2 + (x_1-6)^4$. Функція f сильно опукла з параметром $\theta = 2$, оскільки $f''\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} 2+12(x_1-6)^2 & 0 \\ 0 & 2 \end{pmatrix}$.

Розглянемо задачу мінімізації функції f на множині $D \subset \mathbb{R}^2$ в еліпсі (див. приклади 6, 11 і 13):

$$D = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : 34(x_1-5)^2 + 41(x_2+10)^2 + 24(x_1-5)\ (x_2+10) \le 1 \right\} =$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \left(\begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix} \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix}, \begin{pmatrix} x_1-5 \\ x_2+10 \end{pmatrix} \right) \le 1 \right\} =$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \left(\begin{pmatrix} 3\sqrt{2} & 4\sqrt{2} \\ -4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 25\sqrt{2} \\ 50 \end{pmatrix}, \begin{pmatrix} 3\sqrt{2} & 4\sqrt{2} \\ -4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 25\sqrt{2} \\ 50 \end{pmatrix} \le 1 \right) \right\},$$

який пов'язаний афінним перетворенням

$$\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} = F\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} 3\sqrt{2} & 4\sqrt{2} \\ -4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 25\sqrt{2} \\ 50 \end{pmatrix}$$

з кругом $\tilde{D} = \left\{ \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} : \tilde{x}_1^{\,2} + \tilde{x}_2^{\,2} \le 1 \right\} = F(D)$.

Оскільки область $D \subset \mathbb{R}$ непорожня, опукла і замкнена, можемо застосувати результат теореми 2; формулу для обчислення проекції точки на круг $\tilde{D}$ наведено у прикладі 10. У табл. 1 подано результати роботи класичного методу проекції градієнта (послідовність x^k, $k \ge 0$) та методу проекції градієнта з афінним перетворенням (послідовність $x^{F,k}$, $k \ge 0$); в обох випадках для збіжності з точністю до 10^{-5}. За початкову точку взято центр еліпса: $x^0 = x^{F,0} = \begin{pmatrix} 5 \\ -10 \end{pmatrix}$ (те ж саме спостерігається і за інших початкових точок).

Таблиця 1

k	**0**	**1**	**2**	**3**	**4**	**5**
x^k	$\begin{pmatrix} 5.00000 \\ -10.00000 \end{pmatrix}$	$\begin{pmatrix} 5.10119 \\ -10.15950 \end{pmatrix}$	$\begin{pmatrix} 5.11600 \\ -10.15380 \end{pmatrix}$	$\begin{pmatrix} 5.11628 \\ -10.15380 \end{pmatrix}$	$\begin{pmatrix} 5.11629 \\ -10.15380 \end{pmatrix}$	$\begin{pmatrix} 5.11628 \\ -10.15380 \end{pmatrix}$
$x^{F,k}$	$\begin{pmatrix} 5.00000 \\ -10.00000 \end{pmatrix}$	$\begin{pmatrix} 5.10640 \\ -10.15750 \end{pmatrix}$	$\begin{pmatrix} 5.11568 \\ -10.15400 \end{pmatrix}$	$\begin{pmatrix} 5.11623 \\ -10.15380 \end{pmatrix}$	$\begin{pmatrix} 5.11626 \\ -10.15380 \end{pmatrix}$	$\begin{pmatrix} 5.11626 \\ -10.15380 \end{pmatrix}$

Отже, комп'ютерне моделювання підтверджує, що $\lim\limits_{k \ge 0} x^k = \lim\limits_{k \ge 0} x^{F,k} \approx \begin{pmatrix} 5.1627 \\ -10.15380 \end{pmatrix}$, що збігається зі значенням $x^* = \arg\min\limits_{x \in D} f(x)$ з точністю до 10^{-5}, проте траєкторії x^k і $x^{F,k}$ розрізняються.

Приклад 16. Нехай $f:\mathbb{R}^2 \to \mathbb{R}$, $f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = (x_1+1)^2 + (x_2-1)^2 + (x_1+1)^4$.

Функція f сильно опукла з параметром $\theta = 2$, оскільки $f''\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) =$ $= \begin{pmatrix} 2+12(x_1+1)^2 & 0 \\ 0 & 2 \end{pmatrix}$.

Розглянемо задачу мінімізації функції f на множині $D \subset \mathbb{R}^2$ у повернутому прямокутнику (див. приклади 5, 10 і 12) $D = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \begin{cases} 2 \le 3x_1 + 4x_2 \le 5 \\ 2 \le -4x_1 + 3x_2 \le 10 \end{cases} \right\}$, який пов'язаний афінним перетворенням $\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} = F\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ із прямокутником $\tilde{D} = \left\{ \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} : \begin{cases} 2 \le 5\tilde{x}_1 \le 5 \\ 2 \le 5\tilde{x}_2 \le 10 \end{cases} \right\} = \left\{ \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} : \begin{cases} 0.4 \le \tilde{x}_1 \le 1 \\ 0.4 \le \tilde{x}_2 \le 2 \end{cases} \right\} = F(D)$, сторони якого паралельні координатним осям.

Оскільки область $D \subset \mathbb{R}$ непорожня, опукла і замкнена, можемо застосувати результат теореми 2, формула для обчислення проекції точки на прямокутник $\tilde{D}$ наведена у прикладі 10. У табл. 2 наведено результати роботи класичного методу проекції градієнта (послідовність x^k, $k \ge 0$) та методу проекції градієнта з афінним перетворенням (послідовність $x^{F,k}$, $k \ge 0$); в обох випадках для збіжності з точністю до 10^{-5}. За початкову точку взято центр прямокутника: $x^0 = x^{F,0} = \begin{pmatrix} -0.54 \\ 1.28 \end{pmatrix} = F^{-1}\left(\begin{pmatrix} 0.7 \\ 1.2 \end{pmatrix}\right)$ (схоже спостерігається і за інших початкових точок).

Таблиця 2

k	**0**	**1**	**2**	**3**	**4**	**5**
x^k	$\begin{pmatrix} -0.54000 \\ 1.28000 \end{pmatrix}$	$\begin{pmatrix} -0.85268 \\ 1.13951 \end{pmatrix}$	$\begin{pmatrix} -0.87694 \\ 1.15770 \end{pmatrix}$	$\begin{pmatrix} -0.87626 \\ 1.15719 \end{pmatrix}$	$\begin{pmatrix} -0.87627 \\ 1.15720 \end{pmatrix}$	$\begin{pmatrix} -0.87627 \\ 1.15720 \end{pmatrix}$
$x^{F,k}$	$\begin{pmatrix} -0.54000 \\ 1.28000 \end{pmatrix}$	$\begin{pmatrix} -0.85268 \\ 1.13951 \end{pmatrix}$	$\begin{pmatrix} -0.87694 \\ 1.15770 \end{pmatrix}$	$\begin{pmatrix} -0.87626 \\ 1.15719 \end{pmatrix}$	$\begin{pmatrix} -0.87627 \\ 1.15720 \end{pmatrix}$	$\begin{pmatrix} -0.87627 \\ 1.15720 \end{pmatrix}$

Отже, комп'ютерне моделювання підтверджує, що $\lim_{k\ge 0} x^k = \lim_{k \ge 0} x^{F,k} \approx$ $\approx \begin{pmatrix} 5.1627 \\ -10.15380 \end{pmatrix}$, що збігається зі значенням $x^* = \arg\min_{x \in D} f(x)$ з точністю до 10^{-5}, і траєкторії x^k і $x^{F,k}$ також збігаються, що пояснюється ортогональністю матриці $\begin{pmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{pmatrix}$, яка визначає $F\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$.

ВИСНОВКИ

1. Афінне перетворення координат у деяких випадках суттєво спрощує проектування точки на множину.

2. Афінне перетворення у методі проекції градієнта принаймні не погіршує швидкість збіжності методу.

3. Напрямом подальших досліджень може бути узагальнення отриманого результату на ширший клас перетворень; принциповим має бути зберігання опуклості множин та сильної опуклості функцій.

ЛІТЕРАТУРА

1. Andersen Ang, "Projected Gradient Algorithm," *Mathématique et recherche opérationnelle*, 2021. Available: https://angms.science/doc/CVX/CVX_PGD.pdf
2. Sébastien Bubeck, "Convex Optimization: Algorithms and Complexity," *Foundations and Trends in Machine Learning*, vol. 8, no. 3-4, pp. 232–357, 2015. doi: 10.1561/2200000050
3. H. Jongen, K. Meer, and E. Triesch, *Optimization Theory*. New York, Boston, Dordrechr, London, Moscow: Kluwer Academic Publisher, 2007, 443 p.
4. A. Sukharev, A. Timokhov, and V. Fedorov, *Course Optimization Methods*. M.: Nauka, 2005, 368 p.
5. F.P. Vasil'ev, *Numerical Methods for Solving Extremal Problems.* M.: Nauka, 1988, 552 p.
6. Y. Nesterov, *Introductory Lectures on Convex Optimization*. Boston, Dordrechr, London: Kluwer Academic Publisher, 2004, 236 p.
7. A. Goldstein, "Convex programming in Hilbert space," *Bulletin of the American Mathematical Society*, no. 70, pp. 709–710, 1964.
8. M.P. Moklyachuk, *Foundations of Convex Analysis: Textbook.* Kyiv: TViMS, 2004, 240 p.
9. A. Ahmadi, *Characterizations of convex functions, strict and strong convexity, optimality conditions for convex problems*. 2021. Available: http://www.princeton.edu/~aaa/Public/Teaching/ORF523/ORF523_Lec7.pdf
10. R. Rockafellar, *Convex Analysis*. Princeton: Princeton University Press, 2015, 472 p.

Received 02.08.2023

INFORMATION ON THE ARTICLE

Igor Ya. Spectorsky, ORCID: 0000-0003-4863-7986, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: i.spectorsky@gmail.com

GRADIENT PROJECTION: SIMPLIFYING MINIMIZATION AREA BY AFFINE TRANSFORM / I.Ya. Spectorsky

Abstract. One of the classical problems of optimization theory in a finite-dimensional space is to find a minimum of a function on a nonempty set. Usually, finding the precise solution to this task analytically requires a lot of computational resources or is even impossible at all. So, approximate methods are used most often in practical cases. One of the simplest and the most well-known among such approximate methods for unconditional optimization is the method of gradient descent; its generalization for conditional optimization was found in 1964, the method of projected gradient. For some simple sets (line segment, parallelepiped, ball), the projection of the point on the set can be easily found by an explicit formula. However, for more complicated sets (e.g., an ellipse), projecting becomes a separate task. Nevertheless, sometimes computing projection can be simplified by affine transform; e.g., an ellipse can be transformed into a ball by affine (moreover, by linear) transformation. The paper aims to simplify the problem of minimizing function on the set by changing the condition set by affine transform $F(x)= Ax+b$, where A is a non-degenerated square matrix, and b is a fixed vector of proper dimension.

Keywords: gradient projection, minimization, affine transformation.