

1 Problemy optymalizacyjne

Problemy optymalizacyjne możemy podzielić na:

- problemy optymalizacyjne, których zmienne decyzyjne są ciągłe (zmienne decyzyjne o ciągłych dziedzinach) – *problemy optymalizacyjne ciągłe*,
- problemy optymalizacyjne, których zmienne decyzyjne są dyskretne (zmienne decyzyjne o dyskretnych dziedzinach) – *problemy optymalizacyjne dyskretne*.

Definicja 1. Egzemplarz problemu optymalizacyjnego, oznaczonego przez \mathcal{I} , jest parą (\mathbb{X}, f) , gdzie \mathbb{X} jest zbiorem rozwiązań dopuszczalnych, natomiast f jest funkcją celu (kosztu), $f : \mathbb{X} \rightarrow \mathbb{R}$.

Zatem, szukamy rozwiązania dopuszczalnego \mathbf{x} , $\mathbf{x} \in \mathbb{X}$, takiego, że $f(\mathbf{x}) \leq f(\mathbf{y})$ dla wszystkich $\mathbf{y} \in \mathbb{X}$, czyli:

$$f(\mathbf{x}) = \min_{\mathbf{y} \in \mathbb{X}} f(\mathbf{y}). \quad (1)$$

Rozwiązanie \mathbf{x} jest rozwiązaniem globalnie optymalnym dla egzemplarza \mathcal{I} lub po prostu rozwiązaniem optymalnym.

W definicji 1 rozważaliśmy problem (1), który jest problem minimalizacyjnym. Definicja ta jest poprawna również dla problemów, w których maksymalizujemy funkcję na zbiorze rozwiązań dopuszczalnych.

Definicja 2. Problem optymalizacyjny jest zbiorem egzemplarzy problemu optymalizacyjnego – jest kolekcją egzemplarzy.

Innymi słowy, w egzemplarzu mamy dane wejściowe i wystarczającą ilość informacji, aby otrzymać rozwiązanie. Natomiast problem optymalizacyjny jest zbiorem takich egzemplarzy.

Rozważmy poniżej przykład egzemplarza problemu *minimalnego drzewa rozpinającego*.

MINIMUM SPANNING TREE:

Dane: Graf nieskierowany spójny $G = (V, E)$, gdzie V jest zbiorem wierzchołków, $|V| = n$, E jest zbiorem krawędzi, $|E| = m$, koszty krawędzi $c_e \in \mathbb{Z}_+$, $e \in E$. Zbiór rozwiązań dopuszczalnych jest następującej postaci:

$$\mathbb{X} = \{T = (V, E') : E' \subseteq E, T \text{ jest spójny i acykliczny}\}.$$

$$\text{Funkcja celu (kosztu)} \quad f : \mathbb{X} \rightarrow \mathbb{Z}_+, \quad f(T) = \sum_{e \in T} c_e.$$

Wyjście: Drzewo rozpinające T^* takie, że $f(T^*) = \min_{T \in \mathbb{X}} f(T)$.

Przykład egzemplarza problemu *liniowego programowania LP*.

LINEAR PROGRAMMING:

Dane: liczba wierszy (ograniczeń) m , liczba kolumn (zmiennych decyzyjnych) n , wektor prawych stron $\mathbf{b} \in \mathbb{R}^m$, wektor współczynników funkcji celu $\mathbf{c} \in \mathbb{R}^n$, macierz ograniczeń $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m \times n}$. Zbiór rozwiązań dopuszczalnych jest następującej postaci:

$$\mathbb{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}.$$

$$\text{Funkcja celu (kosztu)} \quad f : \mathbb{X} \rightarrow \mathbb{R}, \quad f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} = \sum_{j \in [n]} c_j x_j,$$

Wyjście: Rozwiązanie dopuszczalne \mathbf{x}^* takie, że $f(\mathbf{x}^*) = \min_{\mathbf{x} \in \mathbb{X}} f(\mathbf{x})$.

$$m = 1, \quad n = 3, \quad \mathbf{A} = [1, 1, 1], \quad \mathbf{b} = [2], \quad \mathbf{c}^T = [c_1, c_2, c_3]^T,$$

$$\begin{aligned} \mathbf{c}^T \mathbf{x} &= c_1 x_1 + c_2 x_2 + c_3 x_3 \rightarrow \min \\ \mathbb{X} &= \begin{cases} x_1 + x_2 + x_3 & = 2 \\ x_1, x_2, x_3 & \geq 0 \end{cases} \end{aligned}$$

Rozważmy egzemplarz minimalnego drzewa rozpinającego, tzn. dane są: graf pełny o trzech wierzchołkach $n = 3$, $m = 3$, koszty krawędzi c_{12} , c_{13} , c_{23} . Egzemplarz tego problemu optymalizacyjnego można wyrazić za pomocą następującego programowania liniowego:

$$\begin{aligned} c_1 x_1 + c_2 x_2 + c_3 x_3 &\rightarrow \min \\ \mathbb{X} &= \begin{cases} x_1 + x_2 + x_3 & = 2 \\ x_1 \leq 1, x_2 \leq 1, x_3 \leq 1 \\ x_1, x_2, x_3 & \geq 0 \end{cases} \end{aligned}$$

gdzie $c_1 = c_{12}$, $c_2 = c_{13}$, $c_3 = c_{23}$.

2 Otoczenie

Dla danego rozwiązania $\mathbf{x} \in \mathbb{X}$ otoczenie jest zbiorem rozwiązań dopuszczalnych, które leżą "blisko" \mathbf{x} .

Definicja 3. Dla dowolnego problemu optymalizacyjnego, będącego zbiorem egzemplarzy (\mathbb{X}, f) , otoczenie jest odwzorowaniem $N : \mathbb{X} \rightarrow 2^{\mathbb{X}}$ zdefiniowanym dla każdego egzemplarza.

Jeżeli $\mathbb{X} = \mathbb{R}^n$, wówczas otoczenie można definiować np. za pomocą normy euklidesowej. Zatem dla programowania liniowego możemy podać następujące otoczenie

$$N_\epsilon(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{Ay} = \mathbf{b}, \mathbf{y} \geq \mathbf{0}, \|\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}.$$

W przypadku problemów optymalizacji dyskretnej wybór otoczenia zależy od struktury zbioru rozwiązań dopuszczalnych \mathbb{X} . Dla minimalnego drzewa rozpinającego otoczenie można zdefiniować w następujący sposób (jest to najpopularniejsze otoczenie):

$$N(T) = \{T' \in \mathbb{X} : T' \text{ jest otrzymany z } T \text{ następująco} \quad (2)$$

dodaj krawędź e do T
powstaje cykl C
usuń dowolną krawędź z C
różną od e).

3 Lokalne i globalne optimum

Definicja 4. Dla danego egzemplarza (\mathbb{X}, f) problemu optymalizacyjnego i otoczenia N , rozwiązanie $\mathbf{x} \in \mathbb{X}$ jest nazywane *optimum lokalnym* (minimum lokalnym) względem otoczenia N , jeśli

$$f(\mathbf{x}) \leq f(\mathbf{y}) \text{ dla wszystkich } \mathbf{y} \in N(\mathbf{x}).$$

Definicja 5. Dany jest problem optymalizacyjny, zbiór rozwiązań dopuszczalnych \mathbb{X} i otoczenie N . Mówimy, że otoczenie N jest *dokładne* (ang. exact), jeśli optimum lokalne $\mathbf{x} \in \mathbb{X}$ względem N jest również optimum globalnym.

Otoczenie (2) dla minimalnego drzewa rozpinającego jest dokładne.

4 Zbiory wypukłe i funkcje wypukłe

Definicja 6. D dowolnych punktów $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, punkt postaci

$$\mathbf{z} = \lambda \mathbf{x} + (1 - \lambda) \mathbf{y}, \quad \lambda \in [0, 1]$$

nazywamy *wypukłą kombinacją* \mathbf{x} i \mathbf{y} .

Jeśli $\lambda \in (0, 1)$, \mathbf{z} nazywamy *ściłą wypukłą kombinacją* \mathbf{x} i \mathbf{y} .

Definicja 7. Zbiór $\mathbb{U} \subseteq \mathbb{R}^n$ jest *wypukły*, jeśli zawiera wszystkie wypukłe kombinacje par punktów \mathbf{x} i \mathbf{y} należących do \mathbb{U} .

Zbiór \mathbb{R}^n jest wypukły. Każdy przedział jest wypukły.

Lemat 1. Przekrój zbiorów wypukłych jest zbiorem wypukłym.

Dowód. Oczywisty. □

Definicja 8. Niech $U \subseteq \mathbb{R}^n$ będzie zbiorem wypukłym. Funkcja $f : U \rightarrow \mathbb{R}$ jest wypukła w zbiorze U , jeśli dla dowolnych dwóch punktów $\mathbf{x}, \mathbf{y} \in U$ zachodzi nierówność:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}), \quad \lambda \in [0, 1].$$

Jeśli $U = \mathbb{R}^n$, to mówimy, że f jest wypukła.

Dowolna liniowa funkcja jest wypukła w dowolnym zbiorze wypukłym U .

Lemat 2. Niech $f, f : U \rightarrow \mathbb{R}$, będzie funkcją wypukłą w zbiorze U . Wtedy dla dowolnego $t \in \mathbb{R}$ zbiór

$$U_t = \{\mathbf{x} \in U : f(\mathbf{x}) \leq t\}$$

jest wypukły.

Dowód. Niech \mathbf{x} i \mathbf{y} będą dowolnymi punktami $\mathbf{x}, \mathbf{y} \in U_t$. Zatem $\mathbf{x}, \mathbf{y} \in U$. Z wypukłości U mamy, że $\lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in U$, $\lambda \in [0, 1]$. Ponieważ f jest wypukła w U następujące nierówności są prawdziwe:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) \leq \lambda t + (1 - \lambda)t = t.$$

Stąd $\lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in U_t$. □

Definicja 9. Niech $U \subseteq \mathbb{R}^n$ będzie zbiorem wypukłym. Funkcja $f : U \rightarrow \mathbb{R}$ jest wklęsła w zbiorze U , jeśli $-f$ jest wypukła w zbiorze U .

Dowolna liniowa funkcja jest wklęsła w dowolnym zbiorze wypukłym U .

Twierdzenie 1 (Charakteryzacja funkcji wypukłej - gradient). Niech $f \in C^1$ (f ma ciągłe pochodne cząstkowe) w otwartym zbiorze zawierającym $U \subseteq \mathbb{R}^n$. Wtedy f jest wypukła w zbiorze U wtedy i tylko wtedy, gdy

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$$

dla wszystkich $\mathbf{x}, \mathbf{y} \in U$, gdzie $\nabla f(\mathbf{x})$ jest gradientem w punkcie \mathbf{x} .

Dowód. (\Rightarrow) Niech $\mathbf{x}, \mathbf{y} \in U$, $\mathbf{x} \neq \mathbf{y}$. Wtedy dla każdego $\lambda \in (0, 1]$

$$f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) = f(\lambda \mathbf{y} + (1 - \lambda)\mathbf{x}) \leq \lambda f(\mathbf{y}) + (1 - \lambda)f(\mathbf{x}).$$

Rozwijamy lewą stronę w szereg Taylor'a w \mathbf{x}

$$f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) = f(\mathbf{x}) + \nabla f(\boldsymbol{\eta})^T(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x}) - \mathbf{x}) = f(\mathbf{x}) + \nabla f(\boldsymbol{\eta})^T(\mathbf{y} - \mathbf{x})\lambda,$$

gdzie $\boldsymbol{\eta}$ jest kombinacją wypukłą \mathbf{x} i $\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})$. Zatem

$$\begin{aligned} f(\mathbf{x}) + \nabla f(\boldsymbol{\eta})^T(\mathbf{y} - \mathbf{x})\lambda &\leq \lambda f(\mathbf{y}) + (1 - \lambda)f(\mathbf{x}) = f(\mathbf{x}) + \lambda(f(\mathbf{y}) - f(\mathbf{x})) \\ \nabla f(\boldsymbol{\eta})^T(\mathbf{y} - \mathbf{x}) &\leq f(\mathbf{y}) - f(\mathbf{x}). \end{aligned}$$

Przechodząc do granicy $\lambda \rightarrow 0$, $\nabla f(\boldsymbol{\eta}) \rightarrow \nabla f(\boldsymbol{x})$.

(\Leftarrow) $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x})$ zachodzi dla wszystkich $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{U}$.
Niech $\boldsymbol{z} = \lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}$. Stąd

$$\begin{aligned} f(\boldsymbol{x}) &\geq f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{x} - \boldsymbol{z}) = f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{x} - \lambda \boldsymbol{x} - (1 - \lambda)\boldsymbol{y}), \\ f(\boldsymbol{y}) &\geq f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{y} - \boldsymbol{z}) = f(\boldsymbol{z}) + \nabla f(\boldsymbol{z})^T(\boldsymbol{y} - \lambda \boldsymbol{x} - (1 - \lambda)\boldsymbol{y}). \end{aligned}$$

Mnożąc pierwszą nierówność przez λ a drugą przez $1 - \lambda$ i dodając stronami otrzymujemy

$$\lambda f(\boldsymbol{x}) + (1 - \lambda)f(\boldsymbol{y}) \geq f(\boldsymbol{z}) = f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}).$$

□

Twierdzenie 2 (Charakteryzacja funkcji wypukłej - Hesjan). *Niech $f \in C^2$ (f ma ciągłe drugie pochodne cząstkowe) w otwartym zbiorze zawierającym $\mathbb{U} \subseteq \mathbb{R}^n$. Wtedy f jest wypukła w zbiorze \mathbb{U} wtedy i tylko wtedy, gdy*

$$\boldsymbol{s}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{s} \geq 0$$

dla wszystkich $\boldsymbol{s} \in \mathbb{R}^n$ i dla wszystkich $\boldsymbol{x} \in \mathbb{U}$ gdzie $\nabla^2 f(\boldsymbol{x})$ jest Hesjanem w punkcie \boldsymbol{x} (macierzą pochodnych cząstkowych).

Dowód. (\Leftarrow) Niech $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{U}$. Z twierdzenia Taylor'a

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{\eta})(\boldsymbol{y} - \boldsymbol{x}),$$

gdzie $\boldsymbol{\eta}$ jest kombinacją wypukłą \boldsymbol{x} i \boldsymbol{y} , stąd $\boldsymbol{\eta} \in \mathbb{U}$.

Z założenia $\boldsymbol{s}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{s} \geq 0$, co daje

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{\eta}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}).$$

Stąd i twierdzenia 1 mamy, że f jest wypukła.

(\Rightarrow) Niech $\boldsymbol{x} \in \mathbb{U}$, $\boldsymbol{s} \in \mathbb{R}^n$ i niech $t > 0$ (małe), takie, że $\boldsymbol{x} + t\boldsymbol{s} \in \mathbb{U}$.
Rozwijając w szereg Taylora mamy

$$f(\boldsymbol{x} + t\boldsymbol{s}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T t\boldsymbol{s} + \frac{1}{2}t^2 \boldsymbol{s}^T \nabla^2 f(\boldsymbol{\eta}) \boldsymbol{s},$$

gdzie $\boldsymbol{\eta}$ jest kombinacją wypukłą $\boldsymbol{x} + t\boldsymbol{s}$ i \boldsymbol{x} .

Z założenia f jest wypukła. Zatem z twierdzenia 1 dostajemy:

$$f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T t\boldsymbol{s} \leq f(\boldsymbol{x} + t\boldsymbol{s}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T t\boldsymbol{s} + \frac{1}{2}t^2 \boldsymbol{s}^T \nabla^2 f(\boldsymbol{\eta}) \boldsymbol{s}.$$

Stąd $\frac{1}{2}t^2 \boldsymbol{s}^T \nabla^2 f(\boldsymbol{\eta}) \boldsymbol{s} \geq 0$. Dzieląc przez $t^2/2$ i przechodząc do granicy $t \rightarrow 0$ dostajemy $\nabla^2 f(\boldsymbol{\eta}) \rightarrow \nabla^2 f(\boldsymbol{x})$. □

Twierdzenie 3. Niech (f, \mathbb{X}) będzie egzemplarzem problemu optymalizacyjnego, $\mathbb{X} \subseteq \mathbb{R}^n$ będzie zbiorem wypukłym a f będzie funkcją wypukłą. Wtedy otoczenie

$$N_\epsilon(\mathbf{x}) = \{\mathbf{y} : \mathbf{y} \in \mathbb{X}, \|\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}$$

jest dokładne dla dowolnego $\epsilon > 0$.

Dowód. Niech \mathbf{x} będzie lokalnym optimum względem $N_\epsilon(\mathbf{x})$, $\epsilon > 0$, tj.

$$f(\mathbf{x}) \leq f(\mathbf{z}) \text{ dla wszystkich } \mathbf{z} \in N_\epsilon(\mathbf{x}).$$

Pokażemy, że $f(\mathbf{x}) \leq f(\mathbf{y})$ dla wszystkich $\mathbf{y} \in \mathbb{X}$. Oczywiście $N_\epsilon(\mathbf{x}) \subset \mathbb{X}$. Jeśli $N_\epsilon(\mathbf{x}) = \mathbb{X}$, to \mathbf{x} jest optimum globalnym w \mathbb{X} .

Zatem wybierzmy dowolny $\mathbf{y} \in \mathbb{X} \setminus N_\epsilon(\mathbf{x})$. Z wypukłości \mathbb{X} i z faktu, że $\epsilon > 0$ istnieje $\mathbf{z} \in N_\epsilon(\mathbf{x})$, $\mathbf{z} \neq \mathbf{x}$, taki, że $\mathbf{z} = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$ dla pewnego $\lambda \in (0, 1)$. Z wypukłości f i lokalnej optymalności \mathbf{x} mamy

$$f(\mathbf{x}) \leq f(\mathbf{z}) = f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

Porządkując dostajemy:

$$(1 - \lambda)f(\mathbf{x}) \leq (1 - \lambda)f(\mathbf{y}), \text{ gdzie } \lambda \in (0, 1) \\ f(\mathbf{x}) \leq f(\mathbf{y}).$$

□

Definicja 10. Egzemplarz (f, \mathbb{X}) problemu optymalizacyjnego jest egzemplarzem problemu programowania wypukłego, jeśli f jest funkcją wypukłą, zbiór rozwiązań dopuszczalnych $\mathbb{X} \subseteq \mathbb{R}^n$ jest zdefiniowany następująco:

$$g_i(\mathbf{x}) \geq 0, \quad i \in [m],$$

gdzie $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ są funkcjami wklęsłymi.

Lemat 3. Zbiór rozwiązań dopuszczalnych \mathbb{X} w problemie programowania wypukłego jest wypukły.

Dowód. Wynika z lematu 2 i lematu 1. □

Z lematu 3 i twierdzenia 3 dostajemy bardzo ważną własność programowania wypukłego.

Twierdzenie 4. W problemie programowania wypukłego każde rozwiązanie \mathbf{x} lokalnie optymalne względem otoczenia $N_\epsilon(\mathbf{x})$ jest globalnie optymalne.

Twierdzenie 5. Niech $f : \mathbb{X} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1$ w zbiorze wypukłym \mathbb{X} i f jest wypukła w \mathbb{X} . Jeśli dla $\mathbf{x}^* \in \mathbb{X}$,

$$\nabla f(\mathbf{x}^*) = \mathbf{0},$$

to \mathbf{x}^* jest optymalny w \mathbb{X} .

Uwagi na temat treści wykładu

Treść wykładu w całości została przygotowana na podstawie książki [1].

Literatura

- [1] Christos H. Papadimitriou, Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Dover Publications Inc., 1998.