Maciej Grzesiak

Optymalizacja

Oznaczenia. Część pojęć i twierdzeń jest formułowana dla ogólnej przestrzeni liniowej V. Jeśli jest ona skończenie wymiarowa, tzn. $V = \mathbb{R}^n$ dla pewnego n, to wektory traktujemy jak macierze kolumnowe.

1. Wstęp

Optymalizacja to ogólnie biorąc wyznaczanie najlepszego rozwiązania danego zagadnienia. Najlepsze rozwiązanie jest rozumiane w sensie maksymalizacji bądź minimalizacji pewnego kryterium liczbowego.
Zagadnienia optymalizacyjne w tak szerokim sensie występują w:
– "czystej" matematyce, np. zagadnienie brachistochrony, zagadnienie minimalizacji powierzchni bryły o ustalonej objętości
– problemach technicznych, np. optymalizacja powierzchni zajmowanej przez zaprojektowany układ elektroniczny, maksymalizacja oszczędzonego miejsca
– problemach zarządzania, np. szeregowanie zadań w celu minimalizacji czasu wykonania, minimalizacja czasu dotarcia (przejazdu) na miejsce
– ekonomii, np. ustalanie asortymentu produkcji w celu maksymalizacji zysku.
Bardziej szczegółowo opiszemy trzy problemy.

Zagadnienie transportowe. Producent musi wysłać pewną liczbę jednostek towaru z kilku składnic do określonych punktów sprzedaży. Każdy punkt sprzedaży zamawia określoną ilość towaru, a każda składnica może dostarczyć również określoną ilość towaru.
Wprowadzamy oznaczenia
— m - liczba składnic,
— n - liczba punktów sprzedaży,
— a_i - całkowita ilość towaru do wysyłki ze składnicy i,
— b_j - całkowite zapotrzebowanie towaru w punkcie sprzedaży j,
— x_{ij} - ilość towaru przesyłana ze składnicy i do punktu sprzedaży j
Załóżymy także, że zapotrzebowanie jest równe produkcji (choć nie jest to niezbędne, ale na początek będzie łatwiej).
Wartości x_{ij} należy wyliczyć.
Przykładowo dla $m = 2$ i $n = 3$ całkowita ilość przewożona ze składnicy 1 spełnia równanie

$$x_{11} + x_{12} + x_{13} = a_1,$$

a dla składnicy 2:

$$x_{21} + x_{22} + x_{23} = a_2.$$

Ponadto

$$\begin{aligned} x_{11} + x_{21} &= b_1 \\ x_{12} + x_{22} &= b_2 \\ x_{13} + x_{23} &= b_3 \end{aligned}$$

Przesłanie towaru ze składnicy i do punktu sprzedaży j kosztuje $c_{ij}x_{ij}$ (a więc c_{ij} jest kosztem przesłania jednostki towaru). Zatem koszt całkowity to

$$c_{11}x_{11} + c_{12}x_{21} + c_{13}x_{13} + c_{21}x_{21} + c_{22}x_{22} + c_{23}x_{23}.$$

Podsumowując, należy zminimalizować funkcję kosztu

$$c_{11}x_{11} + c_{12}x_{21} + c_{13}x_{13} + c_{21}x_{21} + c_{22}x_{22} + c_{23}x_{23},$$

przy ograniczeniach $x_{ij} \geqslant 0$ oraz

$$\begin{aligned} x_{11} + x_{12} + x_{13} &= a_1, \\ x_{21} + x_{22} + x_{23} &= a_2, \\ x_{11} + x_{21} &= b_1, \\ x_{12} + x_{22} &= b_2, \\ x_{13} + x_{23} &= b_3. \end{aligned}$$

Zagadnienie analizy działalności gospodarczej. Środki produkcji (surowce, praca załogi, sprzęt, technologia itd.) mogą służyć produkcji kilku różnych towarów. Wiadomo, jaka ilość środka i jest potrzebna do produkcji towaru j, i wiadomo także jaki dochód daje każda jednostka towaru j.
Wprowadzamy oznaczenia
— m - ilość środków,
— n - ilość towarów,
— a_{ij} - ilość jednostek środka i wymagana do wytworzenia jednostki towaru j,
— b_i - maksymalna dostępna ilość jednostek środka i,
— c_j - dochód jaki daje wyprodukowana jednostka towaru j
— x_j - produkowana ilość towaru j
Współczynniki a_{ij} nazywamy *współczynnikami nakładów i wyników.*
Całkowite zużycie i-tego środka jest określone wyrażeniem

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n.$$

Ponieważ mamy ograniczenia ilościowe, więc

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leqslant b_i.$$

Podsumowanie: należy zmaksymalizować funkcję zysku

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n,$$

przy ograniczeniach $x_j \geqslant 0$ dla $j = 1, 2, \ldots, n$ oraz

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = a_1 \leqslant b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = a_1 \leqslant b_2 \\
\\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = a_1 \leqslant b_m.
\end{aligned}
$$

Liczby $x_j \geqslant 0$, $j = 1, 2, \ldots, n$ wygodnie jest traktować jak współrzędne wektora $\mathbf{x}$, a liczby $a_{ij} \geqslant 0$ jak elementy macierzy.

Zagadnienie planowania produkcji. Firma pragnie ustalić plan swojej produkcji na pewien okres czasu. Dla tego okresu znana jest funcja popytu na produkt. Ewentualny nadmiar będzie zmagazynowany; koszt jest proporcjonalny do przechowywanej ilości dóbr. Oznaczamy:

$x(t)$ – zapasy posiadane w chwili t,
$r(t)$ – ilość produkowanego produktu w chwili t,
$d(t)$ – popyt w chwili t,

Wtedy

$$x'(t) = r(t) - d(t), \quad x(0) = x_0.$$

Szukamy funkcji r spełniającej warunki

$$
\begin{gathered}
r(t) \geqslant 0 \\
x_0 + \int_0^t (r(u) - d(u))du = x(t) \geqslant 0
\end{gathered}
$$

dla $0 \leqslant t \leqslant T$, i minimalizującej funkcjonał kosztów

$$I = \int_0^T \Big(c\big(r(t)\big) + h(t)\Big)dt,$$

gdzie $c(r)$ jest kosztem produkcji r, a hx jest kosztem magazynowania zapasów.
Funkcja r, której szukamy, jest elementem przestrzeni $C[0, T]$ i musi spełniać podane ograniczenia, a więc r musi należeć do obszaru wyznaczonego przez te ograniczenia.

2. Główne twierdzenia

Teoria optymalizacji wyrosła z kilku prostych i intuicyjnych związków geometrycznych. Metody analizy funkcjonalnej umożliwiają rozwinięcie intuicji geometrycznej w kierunku złożonych problemów, nawet nieskończenie wymiarowych.

Twierdzenie o rzucie ortogonalnym

W najprostszym trójwymiarowym przypadku mówi ono, że najkrótsza odległość punktu od płaszczyzny jest wyznaczona przez odcinek prostej prostopadłej do tej płaszczyzny. W uogólnionej postaci, dla przestrzeni o wyższym wymiarze i nieskończenie wymiarowych przestrzeni Hilberta, jest ono podstawą procesów estymacyjnych i aproksymacji metodą najmniejszych kwadratów.

Twierdzenie Hahna-Banacha

Istnieje sporo wersji tego twierdzenia. Następująca nie jest najbardziej ogólna.

Twierdzenie 1. *Niech f będzie ograniczonym funkcjonałem liniowym określonym na podprzestrzeni M pewnej rzeczywistej i unormowanej przestrzeni liniowej X. Wówczas istnieje ograniczony funkcjonał liniowy F będący rozszerzeniem funkcjonału f na całą przestrzeń X oraz $\|F\|_X = \|f\|_M$.*

Można to traktować jako twierdzenie o istnieniu rozwiązania w zagadnieniach minimalizacji. Mając dany ograniczony funkcjonał liniowy f na podprzestrzeni M przestrzeni unormowanej X nie jest trudno rozszerzyć go na całą przestrzeń, ale uzyskane rozszerzenie nie musi być ograniczone. Powstaje zatem problem znalezienia rozszerzenia o minimalnej normie. To właśnie gwarantuje twierdzenie Hahna-Banacha: istnienie rozszerzenia o minimalnej normie. Jednocześnie mówi nam ile wynosi ta norma.

Dualność

Najprostszym twierdzeniem o dualności jest konstatacja, że w przestrzeni problemy dotyczące wektorów są połączone z problemami wyrażonymi za pomocą płaszczyzn. Np. problem odległości punktu od zbioru, dla którego w ogólniejszej wersji mamy stwierdzenie: *najkrótszy odstęp pomiędzy punktem a zbiorem wypukłym jest równy maksimum odległości pomiędzy tym punktem a hiperpłaszczyzną oddzielającą go od zbioru wypukłego.* A zatem początkowa minimalizacja po wektorach może być zastąpiona maksymalizacją po hiperpłaszczyznach.

Różniczki

Nawet na najskromniejszych kursach rachunku różniczkowego prezentuje się metodę optymalizacji polegającą na przyrównaniu do zera pochodnej danej funkcji. Ta metoda ma bezpośrednie rozszerzenie nawet na przestrzenie nieskończenie wymiarowe.
W jednowymiarowym przypadku interpretacja geometryczna jest znana: w maksimum lub minimum styczna do wykresu funkcji przyjmuje położenie poziome. W przypadku wyższych wymiarów jest podobnie: w maksimum lub minimum hiperpłaszczyzna styczna do wykresu funkcji przyjmuje położenie poziome.

3. Podstawowe typy problemów optymalizacji

Programowanie liniowe. Jest to najbardziej naturalny sposób formułowania różnorodnych problemów w stosunkowo nietrudny sposób. Mamy tu liniowe funkcje wielkości szukanych i liniowe równania bądź nierówności precyzujące ograniczenia. Np. problem budżetu: jeśli mamy go podzielić na dwa różne przedsięwzięcia, to ograniczenie ma postać $x_1 + x_2 \leqslant B$, gdzie x_i są kwotami przeznaczonymi na te przedsięwzięcia, a B jest budżetem.
Problemy optymalizacji bez ograniczeń. Takie problemy warto rozważać, gdyż są prostsze w analizie i stanowią kamień węgielny metod rozwiązywania problemów optymalizacji z ograniczeniami. Ponadto wiele problemów z ograniczeniami można sprowadzić do tych bez ograniczeń, przez wyrażenie pewnych zmiennych jako funkcji innych. Np. ograniczenie $x_1 + x_2 = B$ może zostać wyeliminowane, gdy będziemy podstawiać $x_2 = B - x_1$ wszędzie tam, gdzie x_2 występuje. Z drugiej strony w problemie budżetu można pominąć ograniczenia, jeśli są możliwości pozyskania dodatkowych funduszy. Należałoby wtedy raczej do problemu wprowadzić składnik odzwierciedlający koszt pozyskania funduszy.

Problemy optymalizacji z ograniczeniami. Mimo wszystko takie problemy pojawiają się najczęściej, bo często problemy ogólne (np. szczegółowy plan produkcji dużej korporacji) jest dzielony na podproblemy, z których każdy będzie miał ograniczenia, bo muszą się one złożyć w całość.
Ogólne sformułowanie ma postać

$$\begin{cases} f(\mathbf{x}) \to \min \\ h_i(\mathbf{x}) = 0, \quad i = 1, 2, \dots m \\ g_j(\mathbf{x}) \leqslant 0, \quad j = 1, 2, \dots p \\ \mathbf{x} \in S \subset \mathbb{R}^n \end{cases} \tag{1}$$

Ogólnie będziemy zakładać, że problem jest "gładki", tzn. że funkcje są ciągłe (lub mają ciągłe pochodne), co zapewni, że małe zmiany w $\mathbf{x}$ skutkują małymi zmianami w funkcjach związanych z tym problemem. Również zbiór S nie jest dowolny; zakładamy przynajmniej, że jest to obszar spójny.
Skala problemów. Naturalną miarą złożoności problemów jest ich rozmiar, mierzony liczbą występujących zmiennych i funkcji ograniczających. Przyjmuje się podział na na trzy klasy: *problemy o małej skali*, zawierające do pięciu zmiennych, *problemy o średniej skali*, zawierające do stu (lub tysiąca) zmiennych, i *problemy o dużej skali*, które mogą mieć nawet miliony zmiennych.

4. Nieliniowe zagadnienia optymalizacji (NZO)

W ZPL zarówno funkcja celu jak i funkcje ograniczające są liniowe. Ogólniej, rozważmy zbiór wypukły $W \subset \mathbb{R}^n$, funkcję wypukłą $f : W \to \mathbb{R}$ i *problem minimalizacji*

$$\begin{cases} f(\mathbf{x}) \to \min \\ \mathbf{x} \in W \end{cases} \tag{2}$$

Definicja 1. a) Funkcja $f(\mathbf{x})$ ma *minimum globalne* w punkcie $\mathbf{x}_0 \in W$ jeśli $f(\mathbf{x}_0) \leqslant f(\mathbf{x})$ dla każdego $\mathbf{x} \in W$.
b) Funkcja $f(\mathbf{x})$ posiada *minimum lokalne* w punkcie $\mathbf{x}_0 \in W$ jeśli istnieje $\varepsilon > 0$ takie, że $f(\mathbf{x}_0) \leqslant f(\mathbf{x})$ dla każdego $\mathbf{x} \in W \cap B(\mathbf{x}_0, \varepsilon)$.
Minimum nazywamy ścisłym jeśli $f(\mathbf{x}_0) < f(\mathbf{x})$ dla każdego $\mathbf{x} \neq \mathbf{x}_0$.

W NZO celem jest wyznaczenie minimum globalnego, lecz metody obliczeniowe na ogół prowadzą do rozwiązania, które jest tylko minimum lokalnym. Pozostaje problem wyjaśnienia, czy znalezione minimum jest także rozwiązaniem globalnym, a jeśli nie, to jak przejść do minimum globalnego.
Następujące twierdzenie pokazuje, że kluczowym założeniem jest założenie wypukłości funkcji celu.

Twierdzenie 2. *Niech $W \subset \mathbb{R}^n$ będzie zbiorem wypukłym, $f : W \to \mathbb{R}$ funkcją wypukłą. Jeżeli $\mathbf{x}_0$ jest rozwiązaniem lokalnym problemu (1), to*
1. *$\mathbf{x}_0$ jest rozwiązaniem globalnym;*
2. *zbiór rozwiązań globalnych jest wypukły;*
3. *jeśli f jest ściśle wypukła, to $\mathbf{x}_0$ jest ścisłym rozwiązaniem lokalnym;*

4. jeśli $\mathbf{x}_0$ jest ścisłym rozwiązaniem lokalnym, to $\mathbf{x}_0$ jest jedynym rozwiązaniem globalnym.

D o w ó d. Aby wykazać 1 przypuśćmy, że $\mathbf{x}_0$ nie jest rozwiązaniem globalnym, tzn. istnieje $\mathbf{x}^* \in W$ takie, że $f(\mathbf{x}^*) < f(\mathbf{x}_0)$. Jako że $\mathbf{x}_0$ jest rozwiązaniem lokalnym, więc $f(\mathbf{x}_0) \leqslant f(\mathbf{x})$ dla każdego $\mathbf{x} \in W \cap B(\mathbf{x}_0, \varepsilon)$ dla pewnego $\varepsilon > 0$. Odcinek $[\mathbf{x}_0, \mathbf{x}^*]$ jest zawarty w W oraz $[\mathbf{x}_0, \mathbf{x}^*] \cap B(\mathbf{x}_0, \varepsilon) \neq \emptyset$, tzn. istnieje $\lambda \in (0, 1)$ takie, że $\lambda\mathbf{x}_0 + (1-\lambda)\mathbf{x}^* \in B(\mathbf{x}_0, \varepsilon)$. Ale wtedy $f(\lambda\mathbf{x}_0 + (1-\lambda)\mathbf{x}^*) \leqslant \lambda f(\mathbf{x}_0) + (1-\lambda)f(\mathbf{x}^*) < f(\mathbf{x}_0)$. Sprzeczność, bo $\mathbf{x}_0$ jest minimum lokalnym.
Łatwiejsze dowody pozostałych tez twierdzenia pomijamy. □
Jeżeli funkcja f jest różniczkowalna, to mamy dość proste kryterium.

Twierdzenie 3. *Niech $W \subset \mathbb{R}^n$ będzie zbiorem wypukłym, $f : W \to \mathbb{R}$ funkcją wypukłą. Jeżeli f jest różniczkowalna w punkcie $\mathbf{x}_0 \in W$, to $\mathbf{x}_0$ jest rozwiązaniem problemu (1) wtedy i tylko wtedy, gdy $Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \geqslant 0$ dla $\mathbf{x} \in W$.*

D o w ó d. ($\Leftarrow$). Załóżmy, że $Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \geqslant 0$ dla $\mathbf{x} \in W$. Na mocy własności funkcji wypukłej (wniosek z twierdzenia o hiperpłaszczyźnie podpierającej) mamy:

$$f(\mathbf{x}) \geqslant f(\mathbf{x}_0) + Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \geqslant f(\mathbf{x}_0).$$

Ponieważ $\mathbf{x} \in \mathbf{W}$ jest dowolne, więc otrzymujemy tezę.
($\Rightarrow$). Jeżeli $\mathbf{x}_0$ jest rozwiązaniem problemu (1) i $\mathbf{x} \in W$ jest ustalonym elementem, to dla $\lambda \in [0, 1]$ wektor $\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0) = (1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x} \in W$, bo W jest wypukły. Zatem $f(\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0)) \geqslant f(\mathbf{x}_0)$, więc

$$Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \lim_{\lambda \to 0} \frac{f(\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0)) - f(\mathbf{x}_0)}{\lambda} \geqslant 0 \quad \square$$

Do dowodu drugiej części twierdzenia nie było potrzebne założenie wypukłości funkcji f. Zatem mamy wniosek.

Wniosek 1. *Jeżeli $\mathbf{x}_0 \in W$, gdzie $W \subset \mathbb{R}^n$ jest zbiorem wypukłym, jest rozwiązaniem lokalnym problemu (1) dla funkcji $f : W \to \mathbb{R}$ różniczkowalnej w punkcie $\mathbf{x}_0 \in W$, to $Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \geqslant 0$ dla $\mathbf{x} \in W$.*

5. Stożek kierunków stycznych

Dla zbioru $W \subset \mathbb{R}^n$, niekoniecznie wypukłego, rozważmy NZO postaci (1)

$$\begin{cases} f(\mathbf{x}) \to \min \\ \mathbf{x} \in W \end{cases}$$

gdzie $f : W \to \mathbb{R}$. Jeśli mamy rozwiązanie lokalne na W, to kierunki spadku wartości funkcji f nie mogą należeć do zbioru kierunków, w których możemy się poruszać nie opuszczając W. Te kierunki wyznaczają pewien stożek, zdefiniowany niżej. [1]

[1] Jest to uogólnienie znanej definicji dla zbiorów wypukłych. Wtedy dla zbioru wypukłego K definiowaliśmy stożek kierunków osiągalnych (dopuszczalnych) w punkcie $\mathbf{a}$ jako $F_K(\mathbf{a}) = \mathrm{cone}(K - \mathbf{a})$.

Definicja 2. *Stożkiem kierunków stycznych* $T(\mathbf{x}_0)$ do zbioru W w punkcie $\mathbf{x}_0 \in \overline{W}$ nazywamy zbiór wektorów $\mathbf{d} \in \mathbb{R}^n$ takich, że

$$\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \mathbf{x}_0)$$

dla pewnych $\lambda_k > 0, \mathbf{x}_k \in W$ oraz $\mathbf{x}_k \to \mathbf{x}_0$.

Równoważnie

$$T(\mathbf{x}_0) = \left\{\mathbf{d} \in \mathbb{R}^n :\ \mathbf{d} = \lambda \lim_{k\to\infty} \frac{\mathbf{x}_k - \mathbf{x}_0}{\|\mathbf{x}_k - \mathbf{x}_0\|}\right\}$$

dla pewnych $\lambda \geqslant 0$, $(\mathbf{x}_k) \subset W$, $\mathbf{x}_k \to \mathbf{x}_0$, $\mathbf{x}_k \neq \mathbf{x}_0$.
Łatwo sprawdzić, że jeśli $\mathbf{d} \in T(\mathbf{x}_0)$, to także $\lambda\mathbf{d} \in T(\mathbf{x}_0)$ dla każdego $\lambda \geqslant 0$, a więc zbiór $T(\mathbf{x}_0)$ jest rzeczywiście stożkiem. Dla punktów wewnętrznych zbioru W jest $T(\mathbf{x}_0) = \mathbb{R}^n$. Własności gradientu uzasadniają następującą definicję.

Definicja 3. Załóżmy, że funkcja wypukła $f : W \to \mathbb{R}^n$ jest różniczkowalna w punkcie $\mathbf{x}_0 \in W$. *Zbiorem kierunków spadku* funkcji f w punkcie $\mathbf{x}_0$ nazywamy zbiór

$$D(\mathbf{x}_0) = \{\mathbf{d} \in \mathbb{R}^n :\ Df(\mathbf{x}_0)\mathbf{d} < 0\}$$

Twierdzenie 4. *Niech $\mathbf{x}_0$ będzie rozwiązaniem lokalnym problemu (1). Jeśli f jest różniczkowalna w punkcie $\mathbf{x}_0 \in W$, to*

$$D(\mathbf{x}_0) \cap T(\mathbf{x}_0) = \emptyset.$$

Przykład. Rozważmy problem

$$\begin{cases} 2x_1^2 + x_2^2 \to \min \\ 3x_1 + x_2 \geqslant 6 \end{cases}$$

Zatem $f(x_1, x_2) = 2x_1^2 + x_2^2$, $Df(x_1, x_2) = \begin{bmatrix} 4x_1 \\ 2x_2 \end{bmatrix}$,
$W = \{\mathbf{x} \in \mathbb{R}^2 :\ 3x_1 + x_2 \geqslant 6\}$.

Rozważymy trzy przykładowo wybrane punkty.
1) Punkt $\mathbf{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ leży wewnątrz W, więc $T(\mathbf{x}) = \mathbb{R}^2$. Gradient $Df(2,2) = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$, więc $D(\mathbf{x}) = \{\mathbf{d} :\ 2d_1 + d_2 < 0\}$. Zatem $T(\mathbf{x}) \cap D(\mathbf{x}) \neq \emptyset$. Punkt $\mathbf{x}$ nie może być rozwiązaniem.

2) Punkt $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ leży na brzegu W. $T(\mathbf{x}) = \{\mathbf{d} : \ 3d_1 + d_2 \geqslant 0\}$. Gradient $Df(1,3) = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$, więc $D(\mathbf{x}) = \{\mathbf{d} : \ 2d_1 + 3d_2 < 0\}$. Zatem i teraz $T(\mathbf{x}) \cap D(\mathbf{x}) \neq \emptyset$. Punkt $\mathbf{x}$ nie może być rozwiązaniem.
3) Punkt $\mathbf{x} = \begin{bmatrix} \frac{18}{11} \\ \frac{12}{11} \end{bmatrix}$ leży na brzegu W, więc $T(\mathbf{x}) = \{\mathbf{d} : \ 3d_1 + d_2 \geqslant 0\}$ jest taki jak poprzednio. Gradient $Df(\frac{18}{11}, \frac{12}{11}) = \begin{bmatrix} \frac{72}{11} \\ \frac{24}{11} \end{bmatrix} = \frac{24}{11} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, więc $D(\mathbf{x}) = \{\mathbf{d} : \ 3d_1 + d_2 < 0\}$. Tym razem $T(\mathbf{x}) \cap D(\mathbf{x}) = \emptyset$. W punkcie $\mathbf{x}$ może być minimum.

6. Zagadnienie z ograniczeniami nierównościowymi

Rozważmy problem

$$\begin{cases} f(\mathbf{x}) \to \min \\ g_i(\mathbf{x}) \leqslant 0, \quad i = 1, 2, \ldots, m, \\ \mathbf{x} \in W \end{cases} \tag{3}$$

gdzie $W \subset \mathbb{R}^n$ jest zbiorem otwartym, oraz $f, g_1, g_2, \ldots, g_m : X \to \mathbb{R}^n$.
Funkcje g_i nazywamy *ograniczeniami nierównościowymi.* Będziemy zakładać, że są one ciągłe.
Dla ustalonego $\mathbf{x} \in W$, jeśli $g_i(\mathbf{x}) < 0$, to istnieje otoczenie na którym także $g_i < 0$. Zatem kierunki zmian wokół $\mathbf{x}$ ograniczają tylko te funkcje g_i, dla których $g_i(\mathbf{x}) = 0$.

Definicja 4. Zbiorem ograniczeń aktywnych w punkcie $\mathbf{x} \in W$ nazywamy zbiór

$$I(\mathbf{x}) = \{i : \ g_i(\mathbf{x}) = 0\}.$$

Pojęcie zbioru ograniczeń aktywnych jest fundamentalne, gdyż pozwala nam sprowadzić ograniczenia nierównościowe do równościowych. Zatem teorię można w zasadzie rozwijać dla problemu z ograniczeniami równościowymi, uzupełniając ją potem o rozważania pozwalające na określenie, które ograniczenia są aktywne.
Ze zbiorem ograniczeń aktywnych związany jest pewien stożek.

Definicja 5. Niech $\mathbf{x} \in W$, niech $i \in I(\mathbf{x})$, i załóżmy, że g_i jest różniczkowalne w punkcie $\mathbf{x}$. *Stożkiem kierunków stycznych dla ograniczeń aktywnych* nazywamy zbiór

$$T_{\text{akt}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n : \ \forall i \in I(\mathbf{x}) \quad Dg_i(\mathbf{x})\mathbf{d} \leqslant 0\}.$$

Ponieważ nierówność $Dg_i(\mathbf{x})\mathbf{d} \leqslant 0$ określa półprzestrzeń w $\mathbb{R}^n$, więc stożek $T_{\text{akt}}(\mathbf{x})$ jest zbiorem wielościennym, a zatem wypukłym i domkniętym.

Lemat 1. *$T(\mathbf{x}) \subset T_{\text{akt}}(\mathbf{x})$ dla $\mathbf{x} \in W$.*

Dowód. Jeżeli $\mathbf{d} \in T(\mathbf{x})$, to $\mathbf{d} = \lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \mathbf{x})$ dla pewnego ciągu $(\mathbf{x}_k) \subset W$, $\mathbf{x}_k \to \mathbf{x}$, i ciągu (λ_k) liczb dodatnich.
Z definicji pochodnej mamy

$$g_i(\mathbf{x}_k) = g_i(\mathbf{x}) + Dg_i(\mathbf{x})(\mathbf{x}_k - \mathbf{x}) + o(\|\mathbf{x}_k - \mathbf{x}\|),$$

ale jeśli $i \in I(\mathbf{x})$, to $g_i(\mathbf{x}) = 0$ oraz $g_i(\mathbf{x}_k) \leqslant 0$, więc

$$Dg_i(\mathbf{x})(\mathbf{x}_k - \mathbf{x}) + o(\|\mathbf{x}_k - \mathbf{x}\|) \leqslant 0.$$

Stąd dla $\lambda_k > 0$:

$$Dg_i(\mathbf{x})\Big(\lambda_k(\mathbf{x}_k - \mathbf{x})\Big) + \lambda_k\|\mathbf{x}_k - \mathbf{x}\|\frac{o(\|\mathbf{x}_k - \mathbf{x}\|)}{\|\mathbf{x}_k - \mathbf{x}\|} \leqslant 0.$$

Ponieważ $\lambda_k(\mathbf{x}_k - \mathbf{x}) \to \mathbf{d}$, $\frac{o(\|\mathbf{x}_k - \mathbf{x}\|)}{\|\mathbf{x}_k - \mathbf{x}\|} \to 0$, więc po przejściu do granicy mamy $Dg_i(\mathbf{x})\mathbf{d} \leqslant 0$, czyli $\mathbf{d} \in T_{\text{akt}}(\mathbf{x})$. □

Przykład. Pokażemy, że stożki $T(\mathbf{x})$ i $T_{\text{akt}}(\mathbf{x})$ mogą być naprawdę różne. Niech

$$W = \{\mathbf{x} : \; x_1^2 + x_2^2 \leqslant 1, \; x_2^3 \geqslant 0\} \subset \mathbb{R}^2$$

będzie określony funkcjami

$$g_1(x_1, x_2) = x_1^2 + x_2^2 - 1, \quad g_2(x_1, x_2) = -x_2^3.$$

Weźmy $\mathbf{x} = (0, 1)$. Wtedy $T(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^2 : \; d_2 \leqslant 0\}$. Ponieważ $Dg_1(0, 1) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $Dg_2(0, 1) = \begin{bmatrix} 0 \\ -3 \end{bmatrix}$ oraz ograniczeniem aktywnym jest tylko g_1, więc w tym przypadku

$$T_{\text{akt}}(0, 1) = \{\mathbf{d} \in \mathbb{R}^2 : \; 2d_2 \leqslant 0\} = T(0, 1).$$

Ale dla $\mathbf{x} = (1, 0)$ jest inaczej. $T(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^2 : \; d_1 \leqslant 0, d_2 \geqslant 0\}$, $Dg_1(1, 0) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $Dg_2(1, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, ograniczeniami aktywnymi są g_1 i g_2, więc teraz

$$T_{\text{akt}}(1, 0) = \{\mathbf{d} \in \mathbb{R}^2 : \; 2d_1 \leqslant 0\} \supset T(1, 0).$$

Ważne są te punkty, w których mamy równość stożków: $T(\mathbf{x}) = T_{\text{akt}}(\mathbf{x})$.
Przed sformułowaniem głównego twierdzenia wykażemy lemat.

Lemat 2. (Farkasa) *Niech* $\mathbf{A}$ *będzie macierzą* $m \times n$ *oraz* $\mathbf{d} \in \mathbb{R}^n$. *Wówczas dokładnie jeden z układów ma rozwiązanie:*

$$(1) \begin{cases} \mathbf{A}\mathbf{x} \leqslant 0 \\ \mathbf{d}^T\mathbf{x} > 0 \\ \mathbf{x} \in \mathbb{R}^n \end{cases} \qquad (2) \begin{cases} \mathbf{A}^T\mathbf{y} = \mathbf{d} \\ \mathbf{y} \geqslant 0 \\ \mathbf{y} \in \mathbb{R}^m \end{cases}$$

D o w ó d. Załóżmy, że układ (2) ma rozwiązanie $\mathbf{y}$, tj. $\mathbf{d} = \mathbf{A}^T\mathbf{y}$. Wtedy układ (1) ma postać

$$\begin{cases} \mathbf{A}\mathbf{x} \leqslant 0 \\ \mathbf{y}^T\mathbf{A}\mathbf{x} \geqslant 0 \end{cases}$$

i jest sprzeczny, bo wektor $\mathbf{Ax}$ ma współrzędne niedodatnie, a $\mathbf{y}$ nieujemne. Załóżmy teraz, że układ (2) nie ma rozwiązania. Rozważmy zbiór

$$V = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{x} = \mathbf{A}^T\mathbf{y} \ \text{ dla pewnego } \mathbf{y} \geqslant 0\}.$$

Zbiór V jest wypukły i domknięty i z założenia $\mathbf{d} \notin V$. Z twierdzenia Eidelheita istnieje hiperpłaszczyzna oddzielająca $\mathbf{d}$ od zbioru V, tzn. istnieje wektor $\mathbf{a} \in \mathbb{R}^n$ taki, że

$$\sup_{\mathbf{x} \in V} \mathbf{a}^T\mathbf{x} < \mathbf{a}^T\mathbf{d}.$$

Wykażemy, że wektor $\mathbf{a}$ spełnia układ (1). Ponieważ $\mathbf{0} \in V$, więc $\sup\limits_{\mathbf{x} \in V} \mathbf{a}^T\mathbf{x} \geqslant 0$. Stąd $\mathbf{d}^T\mathbf{a} = (\mathbf{a}^T\mathbf{d})^T > \mathbf{0}$.

Należy jeszcze wykazać, że $\mathbf{Aa} \leqslant 0$. Przypuśćmy, że jest przeciwnie, i że $\mathbf{Aa}$ ma np. i-tą współrzędną dodatnią. Dla $\mathbf{y} \geqslant \mathbf{0}$:

$$\mathbf{y}^T\mathbf{Aa} = \left[\mathbf{a}^T(\mathbf{A}^T\mathbf{y})\right]^T = \mathbf{a}^T(\mathbf{A}^T\mathbf{y}) \leqslant \sup_{\mathbf{x} \in V} \mathbf{a}^T\mathbf{x}$$

Weźmy ciąg $\mathbf{y}_n = [0, \ldots, 0, n, 0 \ldots, 0]^T$, gdzie i-ta współrzędna jest równa n. Wtedy

$$\lim_{n \to \infty} \mathbf{y}^T\mathbf{Aa} = \lim_{n \to \infty} n(\mathbf{Aa})_i = \infty,$$

bo założyliśmy, że $(\mathbf{Aa})_i > 0$. Otrzymaliśmy sprzeczność, bo $\mathbf{y}^T\mathbf{Aa} \leqslant \sup\limits_{\mathbf{x} \in V} \mathbf{a}^T\mathbf{x}$.

Zatem $\mathbf{Aa} \leqslant \mathbf{0}$. $\square$

Interpretacja geometryczna. Ponieważ $\mathbf{y} \geqslant 0$, więc zbiór $\{\mathbf{A}^T\mathbf{y}\}$ jest stożkiem wyznaczonym przez wiersze macierzy $\mathbf{A}$ i warunek $\mathbf{A}^T\mathbf{y} = \mathbf{d}$ oznacza, że wektor $\mathbf{d}$ należy do tego stożka. Natomiast nierówności (1) oznaczają, że istnieje wektor $\mathbf{x}$, który z wierszami macierzy $\mathbf{A}$ tworzy kąty nieostre, a z wektorem $\mathbf{d}$ kąt ostry. Taki wektor $\mathbf{x}$ określa hiperpłaszczyznę oddzielającą wektor $\mathbf{d}$ od stożka. Podsumowując, lemat stwierdza, że dla danego stożka i danego wektora: albo wektor należy do stożka, albo można go oddzielić od stożka hiperpłaszczyzną.

Ilustracja na przykładzie. Weźmy $m = 3$, $n = 2$ oraz

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 2 & 5 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Wtedy

$$(1) \left\{ \begin{array}{l} \begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leqslant 0 \\ [2\ 3] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} > 0 \end{array} \right. \qquad (2) \left\{ \begin{array}{l} \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \geqslant 0 \end{array} \right.$$

W tym przypadku układ (2) nie ma rozwiązania, co można łatwo zobaczyć rysując stożek, lub rachunkiem (trudniej). Można też wyliczyć, że układ (1) ma rozwiązanie $x_1 > 0, -\frac{2}{3}x_1 < x_2 \leqslant -\frac{1}{2}x_1$ (np. $x_1 = 2, x_2 = -1$).

Twierdzenie 5. (Karusha-Kuhna-Tuckera) *Niech* $\mathbf{x}_0$ *będzie rozwiązaniem lokalnym zagadnienia z ograniczeniami nierównościowymi (3). Jeżeli funkcje* f *oraz* g_i, $i \in I(\mathbf{x}_0)$ *są różniczkowalne w* $\mathbf{x}_0$ *oraz* $T(\mathbf{x}_0) = T_{\text{akt}}(\mathbf{x}_0)$, *to istnieją liczby* $\mu_i \geqslant 0$, $i = 1, \ldots, m$ *takie że*

$$\begin{cases} Df(\mathbf{x}_0) + \sum\limits_{i \in I(\mathbf{x}_0)} \mu_i Dg_i(\mathbf{x}_0) = \mathbf{0}^T \\ \mu_i g_i(\mathbf{x}_0) = 0, \quad i = 1, \ldots, m \end{cases} \tag{4}$$

D o w ó d. Z twierdzenia 4 mamy $D(\mathbf{x}_0) \cap T(\mathbf{x}_0) = \emptyset$, czyli (z założenia) $D(\mathbf{x}_0) \cap T_{\text{akt}}(\mathbf{x}_0) = \emptyset$, więc układ

$$\begin{cases} Df(\mathbf{x}_0)\mathbf{x} < \mathbf{0} \\ Dg_i(\mathbf{x}_0)\mathbf{x} \leqslant \mathbf{0}, \quad i \in I(\mathbf{x}_0) \end{cases} \tag{5}$$

nie ma rozwiązania. Zastosujemy lemat Farkasa do macierzy $\mathbf{A}$ zbudowanej wierszowo z gradientów ograniczeń aktywnych i wektora $\mathbf{d} = -Df(\mathbf{x}_0)^T$. Skoro układ (5) nie ma rozwiązania, tj. układ (1) z lematu Farkasa nie ma rozwiązania, to istnieje $\mathbf{y} \geqslant \mathbf{0}$ takie, że

$$\mathbf{y}^T \mathbf{A} = -Df(\mathbf{x}_0)$$

czyli

$$Df(\mathbf{x}_0) + \mathbf{y}^T \mathbf{A} = \mathbf{0}^T.$$

Określimy liczby μ_i następująco: $(\mu_i)_{i \in I(\mathbf{x}_0)} = \mathbf{y}$, $(\mu_i)_{i \notin I(\mathbf{x}_0)} = \mathbf{0}$. Wtedy powyższą równość można zapisać

$$Df(\mathbf{x}_0) + \sum_{i \in I(\mathbf{x}_0)} \mu_i Dg_i(\mathbf{x}_0) = \mathbf{0}^T$$

Oczywiście $\mu_i g_i(\mathbf{x}_0) = 0$, bo albo μ_i, albo $g_i(\mathbf{x}_0)$ jest równe 0. □

Liczby μ_i nazywamy *mnożnikami Lagrange'a*. Natomiast warunki (4) nazywane są *warunkami koniecznymi pierwszego rzędu.*

Wniosek 2. *Dla zagadnienia*

$$\begin{cases} f(\mathbf{x}) \to \min \\ g_i(\mathbf{x}) \leqslant 0, \quad i = 1, 2, \ldots, m, \\ \mathbf{x} \geqslant 0 \end{cases} \tag{6}$$

warunki konieczne pierwszego rzędu (KKT) można zapisać w postaci:

$$\begin{cases} Df(\mathbf{x}) + \sum\limits_{i \in I(\mathbf{x})} \mu_i Dg_i(\mathbf{x}) \geqslant \mathbf{0}^T \\ \Big(Df(\mathbf{x}) + \sum\limits_{i \in I(\mathbf{x})} \mu_i Dg_i(\mathbf{x})\Big)\mathbf{x} = 0 \\ \mu_i g_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m \end{cases} \tag{7}$$

D o w ó d. Jest to szczególny przypadek zagadnienia (3), w którym $W = \{\mathbf{x} : \ \mathbf{x} \geqslant 0\}$. Więzy są postaci

$$g_i(\mathbf{x}) \leqslant 0, \quad i = 1, 2, \ldots, m, \quad h_j(\mathbf{x}) = -x_j \leqslant 0 \quad j = 1, 2, \ldots, n.$$

Wobec tego istnieją mnożniki Lagrange'a $\mu_i \geqslant 0, \lambda_j \geqslant 0$ takie, że

$$Df(\mathbf{x}) + \sum_{i \in I(\mathbf{x})} \mu_i Dg_i(\mathbf{x}) + \sum_{j=1}^{n} \lambda_j Dh_j(\mathbf{x}) = \mathbf{0}^T \tag{8}$$

Niech $\mathbf{m} = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$. Wtedy $\sum\limits_{j=1}^{n} \lambda_j Dh_j(\mathbf{x}) = -(\lambda_1, \lambda_2, \ldots, \lambda_n) = -\mathbf{m}^T$, więc

$$Df(\mathbf{x}) + \sum_{i \in I(\mathbf{x})} \mu_i Dg_i(\mathbf{x}) = \mathbf{m}, \tag{9}$$

a ponieważ $\mathbf{m} \geqslant 0$, więc

$$Df(\mathbf{x}) + \sum_{i \in I(\mathbf{x})} \mu_i Dg_i(\mathbf{x}) \geqslant \mathbf{0}^T.$$

Natomiast drugi z warunków (4) ma teraz postać

$$\mu_i g_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m, \quad \lambda_j h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, n.$$

Drugą grupę tych równości można zapisać krótko jako $\mathbf{m}^T\mathbf{x} = 0$. Wtedy po uwzględnieniu równości (9) otrzymujemy

$$\Big(Df(\mathbf{x}) + \sum_{i \in I(\mathbf{x})} \mu_i Dg_i(\mathbf{x})\Big)\mathbf{x} = 0.$$

7. Warunki regularności

Ponieważ równość $T(\mathbf{x}) = T_{\text{akt}}(\mathbf{x})$ stożka kierunków stycznych i stożka kierunków stycznych dla ograniczeń aktywnych jest założeniem twierdzenia Karusha-Kuhna-Tuckera, to istotną rzeczą jest rozpoznanie sytuacji, w których te stożki są równe.
Warunki dostateczne równości $T(\mathbf{x}) = T_{\text{akt}}(\mathbf{x})$ nazywamy warunkami regularności.

Definicja 6. W punkcie $\mathbf{x} \in W$ spełniony jest
— *warunek liniowej niezależności*, jeśli funkcje g_i, $i \notin I(\mathbf{x})$ są ciągłe w $\mathbf{x}$ oraz wektory $Dg_i\mathbf{x}$, $i \in I(\mathbf{x})$, są liniowo niezależne;
— *warunek afiniczności*, jeśli funkcje g_i, $i \in I(\mathbf{x})$ są afiniczne oraz funkcje g_i, $i \notin I(\mathbf{x})$, są ciągłe w $\mathbf{x}$;
— *warunek Slatera*, jeśli funkcje g_i, $i \in I(\mathbf{x})$ są pseudowypukłe oraz funkcje g_i, $i \notin I(\mathbf{x})$, są ciągłe w $\mathbf{x}$ oraz istnieje $\mathbf{x} \in X$ dla którego $g_i(\mathbf{x}) < 0$ dla $i \in I(\mathbf{x})$.

Każdy z tych warunków implikuje równość $T(\mathbf{x}) = T_{\text{akt}}(\mathbf{x})$. Przykładowo udowodnimy to dla warunku afiniczności.

Twierdzenie 6. *Jeśli w punkcie $\mathbf{x} \in W$ spełniony jest warunek afiniczności, to*

$$T(\mathbf{x}) = T_{\text{akt}}(\mathbf{x}).$$

D o w ó d. Ponieważ na mocy lematu 1 jest $T(\mathbf{x}) \subset T_{\text{akt}}(\mathbf{x})$, więc należy wykazać zawieranie przeciwne.
Niech $\mathbf{d} \in T_{\text{akt}}(\mathbf{x})$. Wykażemy najpierw, że istnieje $\lambda^* > 0$ takie, że odcinek $\mathbf{x} + \lambda\mathbf{d}$, $0 \leqslant \lambda < \lambda^*$ zawiera się w zbiorze W. Jeśli ograniczenie g_i jest nieaktywne, to $g_i(\mathbf{x}) < 0$,

a więc z ciągłości g_i wynika, że istnieje $\lambda^* > 0$ takie, że $g_i(\mathbf{x} + \lambda\mathbf{d}) \leqslant 0$ dla $i \notin I(\mathbf{x})$, $0 \leqslant \lambda < \lambda^*$.
Natomiast jeśli ograniczenie g_i jest aktywne, to rozumujemy tak. Ponieważ g_i jest afiniczne, więc jest postaci $g_i(\mathbf{x}) = \mathbf{a}_i^T\mathbf{x} + b_i$ dla pewnych $\mathbf{a}_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$. Stąd $Dg_i(\mathbf{x})\mathbf{d} = \mathbf{a}_i^T\mathbf{d}$ oraz $Dg_i(\mathbf{x})\mathbf{d} \leqslant 0$ (ponieważ $\mathbf{d} \in T_{\text{akt}}(\mathbf{x})$). Zatem

$$g_i(\mathbf{x} + \lambda\mathbf{d}) = \mathbf{a}_i^T(\mathbf{x} + \lambda\mathbf{d}) + b_i = \underbrace{\mathbf{a}_i^T\mathbf{x} + b_i}_{=0} + \lambda\underbrace{\mathbf{a}_i^T\mathbf{d}}_{\leqslant 0} \leqslant 0.$$

Mamy wykazać, że $\mathbf{d} \in T(\mathbf{x})$, a więc należy wskazać ciągi $(\mathbf{x}_k) \subset W$ i $(\lambda_k \subset (0, \infty)$ takie, że $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$ oraz $\lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \mathbf{x}) = \mathbf{d}$. Weźmy

$$\mathbf{x}_k = \mathbf{x} + \frac{\lambda^*}{k}\mathbf{d}, \quad \lambda_k = \frac{k}{\lambda^*}.$$

Wtedy $\mathbf{x}_k \in W$, $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$ oraz

$$\lim_{k\to\infty} \lambda_k(\mathbf{x}_k - \mathbf{x}) = \lim_{k\to\infty} \lambda_k \cdot \frac{\lambda^*}{k}\mathbf{d} = \lim_{k\to\infty} \mathbf{d} = \mathbf{d}.$$

Przykład. Niech

$$W = \{\mathbf{x} \in \mathbb{R}^2 : \ x_1^2 + 2x_2^2 \leqslant 1, \ 2x_1 - x_2 \leqslant 2, \ x_1 + 2x_2 \leqslant 1\}.$$

Zatem

$$g_1(x_1, x_2) = x_1^2 + 2x_2^2 - 1, \quad g_2(x_1, x_2) = 2x_1 - x_2 - 2, \quad g_3(x_1, x_2) = x_1 + 2x_2 - 1.$$

Punktem, w którym wszystkie ograniczenia sa aktywne jest $(1, 0)$. Spełniony jest tylko warunek Slatera.

8. Warunki Karusha-Kuhna-Tuckera i analiza ekonomiczna

Warunki Karusha-Kuhna-Tuckera mogą być pomocne przy rozwiązywaniu niektórych problemów numerycznych. W ekonomii ich użyteczność polega bardziej na wyprowadzaniu wniosków jakościowych, bez konieczności podawania liczbowych współczynników problemu. Celem jest wskazanie optymalnej strategii.
Ustalanie cen.
Wiele firm staje przed problemem zmiennego zapotrzebowania na produkt w różnych godzinach dnia. W godzinach szczytu możliwości firmy są w pełni wykorzystane, ale poza szczytem pozostają wolne moce. Przykładem mogą być firmy telekomunikacyjne.

Twierdzenie 7. *Firma osiągnie maksymalny zysk, jeśli ceny poza godzinami szczytu będą pokrywały tylko krańcowe koszty operacyjne, a w godzinach szczytu będą je przekraczać. Suma nadwyżek cen nad krańcowymi kosztami operacyjnymi we wszystkich godzinach szczytu jest równa krańcowemu kosztowi kapitału.*

D o w ó d. Niech $x_1, x_2, \ldots, x_{24}$ będzie zapotrzebowaniem w kolejnych godzinach doby, a p_1, p_2,$\ldots, p_{24}$ odpowiadającymi im cenami. Zakładamy, że $x_i > 0$. Niech y będzie

godzinową produktywnością firmy. Funkcja $C(x_1, x_2, \ldots, x_{24})$ określa dobowy koszt całkowity, a $g(y)$ dobowy koszt kapitału.
Celem firmy jest maksymalizacja zysku

$$\pi = \sum_{i=1}^{24} p_i x_i - C(x_1, x_2, \ldots, x_{24}) - g(y),$$

przy ograniczeniach

$$x_i - y \leqslant 0 \quad (i = 1, 2, \ldots, 24) \tag{10}$$

$$x_i \geqslant 0 \quad (i = 1, 2, \ldots, 24) \tag{11}$$

$$y \geqslant 0. \tag{12}$$

Funkcja π jest funkcją wektora $(x_1, x_2, \ldots, x_{24}, y)$. Pierwsza równość warunków (4) w przypadku maksimum ma postać

$$D\pi(\mathbf{x}) - \sum_{i \in I(\mathbf{x})} \mu_i D g_i(\mathbf{x}) = \mathbf{0}^T,$$

gdzie $g_i(y) = x_i - y$. Zatem po rozpisaniu

$$\frac{\partial \pi}{\partial x_i} = p_i - \frac{\partial C}{\partial x_i} - \mu_i = 0, \tag{13}$$

$$\frac{\partial \pi}{\partial y} = -\frac{\mathrm{d}\, g}{\mathrm{d}\, y} + \sum_{i=1}^{24} \mu_i = 0, \tag{14}$$

$$\mu_i (x_i - y) = 0, \tag{15}$$

$$\mu_i \geqslant 0, \tag{16}$$

$$x_i > 0. \tag{17}$$

W dowolnym okresie t poza godzinami szczytu mamy $y > x_t$, a więc musi być $\mu_t = 0$. Wtedy pierwsza równość oznacza, że

$$p_t = \frac{\partial C}{\partial x_t},$$

a więc należy ustalić cenę równą krańcowemu kosztowi operacyjnemu $\frac{\partial C}{\partial x_t}$.
Natomiast dla okresu s w szczycie jest $x_s = y$. Ponieważ $\frac{\mathrm{d}\, g}{\mathrm{d}\, y} > 0$ (zwiększenie produktywności wymaga dodatkowego kapitału), więc

$$\frac{\mathrm{d}\, g}{\mathrm{d}\, y} = \sum_s \mu_s > 0.$$

Wtedy także

$$p_s = \frac{\partial C}{\partial x_s} + \mu_s, \quad \text{dla każdego okresu } s \text{ w szczycie.}$$

Cena przewyższa koszt krańcowy o wartość mnożnika Lagrange'a, i co więcej, suma tych nadwyżek jest równa $\frac{\mathrm{d}\, g}{\mathrm{d}\, y}$, czyli krańcowemu kosztowi produktywności. Zapotrzebowanie w okresie szczytu skłania do zwiększenia mocy produkcyjnych, ale to wymaga dodatkowego kapitału, a zatem cena musi pokryć koszt krańcowy.