Maciej Grzesiak

# Pewne własności zbiorów i funkcji wypukłych w przestrzeniach unormowanych

24.05.2018

## 1. Pochodna funkcji o argumencie wektorowym

Niech $f : W \to \mathbb{R}$, gdzie $W \subset \mathbb{R}^n$ jest zbiorem otwartym.
Oznaczenia:
— $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ – wektor kolumnowy
— $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)^T$
— $Df(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$ – gradient funkcji $f$
— $D^2 f(\mathbf{x})$ – *macierz Hessego*[1] funkcji $f$:

$$D^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Wyznacznik macierzy Hessego nazywamy *hesjanem.*
Macierz Hessego funkcji $f$ o argumencie wektorowym będziemy też oznaczać $\mathbf{F}$.
Gradient i macierz Hessego pozwalają uprościć zapis wielu zależności. W obliczeniach wektory traktujemy jak macierze, np. $\mathbf{x}^T\mathbf{x}$ oznacza iloczyn macierzowy wektora wierszowego przez wektor kolumnowy.

**Przykład.** Obliczyć $\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\mathbf{x}(t))$ dla $f(\mathbf{x}) = f(x_1, x_2) = x_1^2 x_2^3$ jeśli $x_1 = x_1(t)$, $x_2 = x_2(t)$.

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}(t)) = 2x_1 x_2^3 x_1' + 3x_1^2 x_2^2 x_2',$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\mathbf{x}(t)) = 2x_2^3 (x_1')^2 + 6x_1 x_2^2 x_2' x_1' + 2x_1 x_2^3 x_1'' + 6x_1 x_2^2 x_1' x_2' + 6x_1^2 x_2^2 (x_2')^2 + 3x_1^2 x_2^2 x_2''.$$

Strukturę wyniku lepiej widać, gdy uogólnimy zadanie:
**Przykład.** Obliczyć $\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\mathbf{x}(t))$ dla $f(\mathbf{x}) = f(x_1, x_2)$ jeśli $x_1 = x_1(t)$, $x_2 = x_2(t)$.

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}(t)) = f_1' x_1' + f_2' x_2',$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\mathbf{x}(t)) = f_{11}''(x_1')^2 + 2f_{12}'' x_2' x_1' + f_1' x_1'' + f_{22}''(x_2')^2 + f_2' x_2''.$$

Te wyniki można zapisać w postaci:

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}(t)) = \nabla f(\mathbf{x}(t)) \cdot \mathbf{x}'$$

[1] Ludwig Otto Hesse, 22.04.1811–4.08.1874

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\mathbf{x}(t)) = \mathbf{x}'(t)^T \mathbf{F}(\mathbf{x}(t))\mathbf{x}'(t) + \nabla f(\mathbf{x}(t)) \cdot \mathbf{x}''.$$

**Definicja 1.** Funkcja $f$ jest *różniczkowalna* w punkcie $\mathbf{x}_0 \in W$ jeśli istnieje wektor $\mathbf{a} \in \mathbb{R}^n$ taki, że

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|)$$

dla $\mathbf{x} \in W$.
Funkcja $f$ jest *dwukrotnie różniczkowalna* w punkcie $\mathbf{x}_0 \in W$ jeśli istnieje wektor $\mathbf{a} \in \mathbb{R}^n$ oraz macierz $\mathbf{H} \in \mathbb{R}^{n\times n}$ takie, że

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2)$$

dla $\mathbf{x} \in W$.

**Uwaga**. Wykażemy, że macierz $\mathbf{H}$ w powyższej równości można zastąpić macierzą symetryczną $\frac{\mathbf{H}+\mathbf{H}^T}{2}$.
Ponieważ $(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}(\mathbf{x} - \mathbf{x}_0)$ jest liczbą, więc

$$(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}(\mathbf{x} - \mathbf{x}_0) = [(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}(\mathbf{x} - \mathbf{x}_0)]^T = (\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}^T(\mathbf{x} - \mathbf{x}_0),$$

a zatem

$$(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}(\mathbf{x} - \mathbf{x}_0) = (\mathbf{x} - \mathbf{x}_0)^T \frac{\mathbf{H} + \mathbf{H}^T}{2}(\mathbf{x} - \mathbf{x}_0).$$

W dalszym ciągu można więc zakładać, że macierz $\mathbf{H}$ jest symetryczna.
Ważne jest następujące twierdzenie.

**Twierdzenie 1.** *a) Jeśli funkcja $f$ jest różniczkowalna w punkcie $\mathbf{x}_0$, to $Df(\mathbf{x}_0)$ istnieje i $Df(\mathbf{x}_0) = \mathbf{a}^T$. Na odwrót, jeśli $Df(\mathbf{x})$ istnieje w pewnym otoczeniu $\mathbf{x}_0$ i jest ciągłe w $\mathbf{x}_0$, to $f$ jest różniczkowalna w $\mathbf{x}_0$.*
*b) Jeśli macierz Hessego $D^2 f(\mathbf{x}_0)$ istnieje w pewnym otoczeniu $\mathbf{x}_0$ i jest ciągła w $\mathbf{x}_0$, to $f$ jest dwukrotnie różniczkowalna w $\mathbf{x}_0$, $D^2 f(\mathbf{x}_0)$ jest macierzą symetryczną oraz $D^2 f(\mathbf{x}_0) = \mathbf{H}$.*

Do badania ekstremów funkcji przydatne jest rozwinięcie Taylora rzędu 2.

**Twierdzenie 2.** *Niech $W \subset \mathbb{R}^n$ otwarty. Dla funkcji $f : W \to \mathbb{R}$ klasy $C^2$ i punktów $\mathbf{x}, \mathbf{x}_0 \in W$ takich, że odcinek $[\mathbf{x}, \mathbf{x}_0] \subset W$ zachodzi*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T D^2 f(\bar{\mathbf{x}})(\mathbf{x} - \mathbf{x}_0), \tag{1}$$

*gdzie $\bar{\mathbf{x}}$ jest pewnym punktem wewnętrznym odcinka $[\mathbf{x}, \mathbf{x}_0]$.*

Jeśli zbiór $W$ jest wypukły, to założenie dotyczące odcinka można opuścić, bo jest automatycznie spełnione.
Warunki konieczne i dostateczne istnienia ekstremum funkcji jednej zmiennej mają swoje analogi dla funkcji określonej na podzbiorze $\mathbb{R}^n$.

**Twierdzenie 3.** *Jeśli funkcja $f : W \to \mathbb{R}$ jest różniczkowalna w punkcie $\mathbf{x}_0 \in \operatorname{int} W$ oraz $\mathbf{x}_0$ jest lokalnym minimum (maksimum) funkcji $f$, to $Df(\mathbf{x}_0) = 0$.*

D o w ó d. Niech $\mathbf{e}_i$ będzie $i$-tym wersorem. Funkcja $g(t) = f(\mathbf{x}_0 + t\mathbf{e}_i)$ jest określona na pewnym otoczeniu 0 i ma w tym punkcie ekstremum. Zatem $g'(0) = 0$. Ale

$$g'(t) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{x}_0 + t\mathbf{e}_i)\frac{\mathrm{d}(\mathbf{x}_0 + t\mathbf{e}_i)_k}{\mathrm{d}\,t} = \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + t\mathbf{e}_i),$$

gdzie $(\mathbf{x}_0 + t\mathbf{e}_i)_k$ oznacza $k$-tą współrzędną wektora $\mathbf{x}_0 + t\mathbf{e}_i$. A więc $g'(0) = 0$ oznacza, że $\frac{\partial f}{\partial x_i}(\mathbf{x}_0) = 0$. Stąd teza. $\square$
Punkt $\mathbf{x}_0 \in \operatorname{int} W$ nazywamy *punktem krytycznym* funkcji $f : W \to \mathbb{R}$ jeśli $f$ jest różniczkowalna w $\mathbf{x}_0$ oraz $Df(\mathbf{x}_0) = 0$. Istnieje także analog warunku o znaku drugiej pochodnej, ale mówi on o określoności macierzy Hessego (czyli macierzy drugich pochodnych).

## 2. Określoność macierzy

Niech $\mathbf{A} = [a_{ij}]$ będzie macierzą symetryczną stopnia $n$. Określa ona formę kwadratową

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i,j=1}^{n} a_{ij}x_ix_j.$$

**Definicja 2.** Mówimy, że macierz $\mathbf{A}$
— jest nieujemnie określona (oznaczenie: $\mathbf{A} \geqslant 0$), jeśli

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \geqslant 0 \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

— jest dodatnio określona (oznaczenie: $\mathbf{A} > 0$), jeśli

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\},$$

Mówimy też o określoności formy kwadratowej odpowiadającej tej macierzy.
Analogicznie definiujemy niedodatnią określoność i ujemną określoność.
Jeśli istnieją wektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ takie, że

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0, \quad \mathbf{y}^T\mathbf{A}\mathbf{y} < 0$$

to mówimy, że $\mathbf{A}$ jest nieokreślona.
Ponieważ dla $i$-tego wersora $\mathbf{e}_i$ mamy $\mathbf{e}_i^T\mathbf{A}\mathbf{e}_i = a_{ii}$, więc np. dla macierzy dodatnio określonej musi być $a_{ii} > 0$ dla każdego $i$. Analogiczne warunki konieczne można wywnioskować dla pozostałych przypadków.
Warunki konieczne i dostateczne podaje następujące *kryterium Sylvestera*, w którym wykorzystuje się *minory główne* macierzy, tj.

$$D_1 = |a_{11}|,\ D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \ldots, D_n = \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}.$$

**Twierdzenie 4. (kryterium Sylvestera)** [2] *Prawdziwe są równoważności:*

1. *Forma kwadratowa* $\mathbf{x}^T\mathbf{A}\mathbf{x}$ *jest dodatnio określona wtedy i tylko wtedy, gdy*

$$D_1 > 0,\ D_2 > 0,\ \ldots,\ D_n > 0.$$

2. *Forma kwadratowa* $\mathbf{x}^T\mathbf{A}\mathbf{x}$ *jest ujemnie określona wtedy i tylko wtedy, gdy* $\mathbf{x}^T(-\mathbf{A})\mathbf{x}$ *jest dodatnio określona, czyli gdy*

$$-D_1 > 0,\ D_2 > 0,\ \ldots,\ (-1)^nD_n > 0.$$

3. *Forma kwadratowa* $\mathbf{x}^T\mathbf{A}\mathbf{x}$ *jest nieujemnie określona wtedy i tylko wtedy, gdy dla dowolnych* $1 \leqslant k \leqslant n$ *oraz* $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ *zachodzi*

$$\begin{vmatrix} a_{i_1i_1} & a_{i_1i_2} & \ldots & a_{i_1i_k} \\ a_{i_2i_1} & a_{i_2i_2} & \ldots & a_{i_2i_k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_ki_1} & a_{i_ki_2} & \ldots & a_{i_ki_k} \end{vmatrix} \geqslant 0$$

Ponieważ wyznaczniki macierzy podobnych są równe, więc konsekwencją kryterium Sylvestera jest fakt, że macierze podobne mają tę samą określoność. Jednocześnie każda macierz symetryczna jest podobna do macierzy diagonalnej, z wartościami własnymi na przekątnej.

[2] James Joseph Sylvester, 3.09.1814–15.03.1897

**Wniosek 1.** *Macierz* $\mathbf{A}$ *jest*
– *nieujemnie określona wtedy i tylko wtedy, gdy wszystkie jej wartości własne są nieujemne,*
– *dodatnio określona wtedy i tylko wtedy, gdy wszystkie jej wartości własne są dodatnie,*
– *niedodatnio określona wtedy i tylko wtedy, gdy wszystkie jej wartości własne są niedodatnie,*
– *ujemnie określona wtedy i tylko wtedy, gdy wszystkie jej wartości własne są ujemne.*

Sformułujemy warunki konieczne i dostateczne dla istnienia ekstremum lokalnego.

**Twierdzenie 5.** *Jeżeli* $f$ *jest klasy* $C^2$ *na zbiorze otwartym i wypukłym* $W \subset \mathbb{R}^n$ *oraz w punkcie* $\mathbf{x}_0 \in W$ *jest minimum lokalne, to macierz* $D^2 f(\mathbf{x}_0)$ *jest nieujemnie określona. Gdy w punkcie* $\mathbf{x}_0 \in W$ *jest maksimum lokalne, to macierz* $D^2 f(\mathbf{x}_0)$ *jest niedodatnio określona.*

**Twierdzenie 6.** *Jeśli* $f$ *jest klasy* $C^2$ *na zbiorze otwartym* $W \subset \mathbb{R}^n$, $Df(\mathbf{x}_0) = 0$ *i macierz* $D^2 f(\mathbf{x}_0)$ *jest dodatnio określona (ujemnie określona), to w* $\mathbf{x}_0$ *jest minimum (maksimum) lokalne właściwe.*

Twierdzenia te wynikają z analizy wzoru Taylora (1).
Gdy zbiór $W$ jest wypukły, to ekstremum jest globalne.

**Twierdzenie 7.** *Jeśli* $f$ *jest klasy* $C^2$ *na zbiorze otwartym* $W \subset \mathbb{R}^n$, $\mathbf{x}_0 \in \operatorname{int} W$ *jest punktem krytycznym* $f$, *to*
1. $D^2 f(\mathbf{x}) \geqslant 0$ *dla każdego* $\mathbf{x} \in \operatorname{int} W \Rightarrow$ *w* $\mathbf{x}_0$ *jest minimum globalne,*
2. $D^2 f(\mathbf{x}) \leqslant 0$ *dla każdego* $\mathbf{x} \in \operatorname{int} W \Rightarrow$ *w* $\mathbf{x}_0$ *jest maksimum globalne.*

*Jeśli dodatkowo* $D^2 f(\mathbf{x}_0) > 0$ *w punkcie 1 (*$D^2 f(\mathbf{x}_0) < 0$ *w punkcie 2), to odpowiednie ekstrema są ścisłe.*

D o w ó d. Niech $\mathbf{x} \in W$. Ponieważ $W$ jest wypukły, więc odcinek $[\mathbf{x}_0, \mathbf{x}) \in \operatorname{int} W$ i na mocy wzoru Taylora istnieje $\bar{\mathbf{x}} \in (\mathbf{x}_0, \mathbf{x})$ takie, że

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T D^2 f(\bar{\mathbf{x}})(\mathbf{x} - \mathbf{x}_0).$$

Jeśli $D^2 f \geqslant 0$, więc $f(\mathbf{x}) \geqslant f(\mathbf{x})$, a stąd wynika teza 1. Analogicznie dowodzimy pozostałych części twierdzenia.

## 3. Funkcjonały

Najogólniej, funkcjonał oznacza odwzorowanie z przestrzeni liniowej do jej ciała skalarów. Np. dla dowolnej macierzy $\mathbf{A}$ stopnia $n$ odwzorowanie $f : \mathbb{R}^n \to \mathbb{R}$:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

jest funkcjonałem.
Ciało można traktować jak 1-wymiarową przestrzeń liniową, więc dołączając do pojęcia funkcjonału warunek liniowości uzyskujemy definicję.

**Definicja 3.** Niech $V$ będzie przestrzenią liniową nad ciałem $F$. Odwzorowanie $f : V \to F$ spełniające warunek

$$f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) \quad \text{dla } \alpha, \beta \in F,\ \mathbf{x}, \mathbf{y} \in V$$

nazywamy *funkcjonałem liniowym.*

Jeżeli $\dim V < \infty$, to funkcjonał nazywamy też *formą liniową*. Wtedy, wybierając w $V$ bazę $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ możemy wartość formy na wektorze $\mathbf{x} = \sum_{i=1}^n \lambda_i \mathbf{e}_i$ zapisać w postaci

$$f(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \beta_i = \lambda_1 \beta_1 + \lambda_2 \beta_2 + \cdots + \lambda_n \beta_n, \tag{2}$$

gdzie $\beta_i = f(\mathbf{e}_i)$ są skalarami, zależnymi od wyboru bazy. I na odwrót, układ skalarów $\beta_1, \beta_2, \ldots, \beta_n$ wyznacza jednoznacznie formę.
W zbiorze funkcjonałów określonych na przestrzeni $V$ mamy naturalne działania dodawania i mnożenia przez skalar.

**Definicja 4.** Zbiór $V^* = L(V, F)$ wszystkich funkcjonałów na $V$ z działaniami dodawania i mnożenia przez skalar nazywamy przestrzenią liniową dualną (sprzężoną) do $V$.

**Przykład.** Jak już wiadomo, funkcjonał na $\mathbb{R}^n$ określony jest układem liczb $\beta_1, \beta_2, \ldots, \beta_n$, czyli wektorem $\mathbf{b} = (\beta_1, \beta_2, \ldots, \beta_n)$:

$$f(\mathbf{x}) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n. \tag{3}$$

Można wykazać, że przestrzeń dualna $(\mathbb{R}^n)^*$ jest izomorficzna z $\mathbb{R}^n$.
Wykorzystując pojęcie iloczynu skalarnego $\langle \cdot, \cdot \rangle$ można warunek (3) zapisać w postaci:

$$f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{b} \rangle. \tag{4}$$

Można też wektory $\mathbf{b}, \mathbf{x}$ traktować jak macierze kolumnowe i wtedy

$$f(\mathbf{x}) = \mathbf{b}^T \mathbf{x}. \tag{5}$$

W $\mathbb{R}^n$ mamy naturalną normę euklidesową $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n x_i^2}$ i zrozumiałe jest oczekiwanie, by funkcjonał był ciągły.

**Definicja 5.** Niech $V$ będzie unormowaną przestrzenią liniową. Przestrzeń wszystkich ciągłych funkcjonałów liniowych na $V$ nazywamy przestrzenią sprzężoną do $V$ i oznaczamy $V^*$.
Norma funkcjonału $f \in V^*$ jest równa

$$\|f\| = \sup_{\|\mathbf{x}\| \leqslant 1} |f(\mathbf{x})|.$$

## 4. Zbiory wypukłe

Niech $V$ będzie przestrzenią wektorową (w zastosowaniach to będzie przestrzeń euklidesowa $\mathbb{R}^n$), a $\mathbf{x}, \mathbf{y} \in V$ będą dwoma punktami. Zbiór

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

nazywamy odcinkiem łączącym punkty $\mathbf{x}, \mathbf{y}$.

**Definicja 6.** Zbiór $K$ zawarty w przestrzeni wektorowej $V$ nazywamy zbiorem wypukłym, jeśli wraz z każdymi dwoma punktami $\mathbf{x}, \mathbf{y} \in K$ łączący je odcinek $[\mathbf{x}, \mathbf{y}]$ jest także zawarty w $K$.

Odcinkami niewłaściwymi w $\mathbb{R}$ nazywane są półproste i cała prosta $\mathbb{R}$.

**Lemat 1.** *$K \subseteq \mathbb{R}$ jest wypukły wtedy i tylko wtedy, gdy jest odcinkiem, być może niewłaściwym.*

**Lemat 2.** *Dla dowolnej rodziny $\{K_j\}_{j \in J}$ zbiorów wypukłych w $V$ ich przekrój*

$$K = \bigcap_{j \in J} K_j$$

*jest zbiorem wypukłym.*

**Wniosek 2.** *Dla dowolnego zbioru $F \subset V$ najmniejszy zbiór wypukły zawierający zbiór $F$ jest przekrojem wszystkich zbiorów wypukłych zawierających $F$. Nazywamy go* powłoką wypukłą *zbioru $F$ i oznaczamy* $\operatorname{conv} F$.

**Przykłady.**
1. W $\mathbb{R}^n$ zbiorem wypukłym jest kula jednostkowa względem normy euklidesowej

$$K^n = \{(x_1, x_2, \ldots, x_n) : \sum_{j=1}^{n} x_j^2 \leqslant 1\}$$

Również kula otwarta

$$K_0^n = \{(x_1, x_2, \ldots, x_n) : \sum_{j=1}^{n} x_j^2 < 1\}$$

a także każdy zbiór $S$ taki, że $K_0^n \subset S \subset K^n$, jest wypukły.
2. *Sympleks jednostkowy* $\Delta^n$ w $\mathbb{R}^n$ zdefiniowany jako

$$\Delta^n = \{(x_1, x_2, \ldots, x_n) : 0 \leqslant x_j \leqslant 1, \sum_{j=1}^{n} x_j \leqslant 1\}$$

jest wypukły.
Sympleks w $\mathbb{R}^2$ to trójkąt, a w $\mathbb{R}^3$ to czworościan.
3. Dla macierzy $\mathbf{A}$ typu $m \times n$ i wektora $\mathbf{b} \in \mathbb{R}^m$ zbiór $X \subseteq \mathbb{R}^n$ zdefiniowany wzorem (nierówności dla wektorów interpretujemy „po współrzędnych”):

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

jest wypukły, jako przekrój skończonej liczby półprzestrzeni. (Może to być zbiór pusty)
Wprowadzimy teraz pojęcie kombinacji wypukłej.

**Definicja 7.** Niech $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in V$. Jeśli skalary $\lambda_1, \lambda_2, \ldots, \lambda_k$ spełniają warunki $\lambda_j \geqslant 0$ i $\sum_{j=1}^{k} \lambda_j = 1$, to wektor

$$\mathbf{v} = \sum_{j=1}^{k} \lambda_j \mathbf{v}_j$$

nazywamy *kombinacją wypukłą* wektorów $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ o współczynnikach $\lambda_1, \lambda_2, \ldots, \lambda_k$.

W szczególności kombinacje wypukłe dwóch punktów (wektorów) $\mathbf{v}_1, \mathbf{v}_2$ można zapisać w postaci $\lambda \mathbf{v}_1 + (1 - \lambda)\mathbf{v}_2$, gdzie $0 \leqslant \lambda \leqslant 1$. Zatem kombinacjami wypukłymi tych punktów są wszystkie punkty odcinka łączącego $\mathbf{v}_1$ z $\mathbf{v}_2$ i tylko one.
Z definicji zbioru wypukłego wynika więc, że każdy zbiór wypukły ma tę własność, że jeśli jakieś dwa punkty do niego należą, to zawiera on także wszystkie kombinacje wypukłe tych punktów. Przez indukcję można dość łatwo wykazać następujące twierdzenie.

**Twierdzenie 8.** *Niech $K$ będzie zbiorem wypukłym. Każda kombinacja wypukła punktów $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in K$ należy do $K$.*

Jeżeli zbiór $K$ jest skończony, to jego powłokę wypukłą nazywamy *wielościanem wypukłym.*
Np. powłoką wypukłą zbioru 2-punktowego w $\mathbb{R}$ jest odcinek, zbioru 3-punktowego w $\mathbb{R}^2$ jest trójkąt (lub odcinek), zbioru 4-punktowego w $\mathbb{R}^3$ jest czworościan (lub trójkąt, lub odcinek).
Następujące twierdzenie pokazuje, że jest także na odwrót: każdy punkt powłoki wypukłej zbioru $K$ jest kombinacją wypukłą skończonego zbioru punktów ze zbioru $K$.

**Twierdzenie 9. (Carathéodory'ego)** [3] *Jeśli $K \subset \mathbb{R}^n$, to dowolny punkt $\mathbf{v} \in \operatorname{conv} K$ można zapisać jako kombinację wypukłą co najwyżej $n + 1$ elementów z $K$.*

[3] Constantin Carathéodory, 13.09.1873–2.02.1950

D o w ó d.
Niech $\mathbf{v} \in \operatorname{conv} K$. Wtedy $\mathbf{v} = \sum_{j=1}^{k} \lambda_j \mathbf{v}_j$ dla pewnych $\mathbf{v}_j \in K$, $\lambda_j \geqslant 0$, $\sum_{j=1}^{k} \lambda_j = 1$. Przypuśćmy, że $k > n+1$. Wtedy wektory $\mathbf{v}_2 - \mathbf{v}_1, \mathbf{v}_3 - \mathbf{v}_1, \ldots, \mathbf{v}_k - \mathbf{v}_1$ muszą być liniowo zależne, więc istnieją skalary $\mu_j$ nie wszystkie równe 0, że $\sum_{j=2}^{k} \mu_j (\mathbf{v}_j - \mathbf{v}_1) = 0$. Niech $\mu_1 := -\sum_{j=2}^{k} \mu_j$. Wtedy $\sum_{j=1}^{k} \mu_j \mathbf{v}_j = 0$, $\sum_{j=1}^{k} \mu_j = 0$ i przynajmniej jedno $\mu_j > 0$. Zatem dla dowolnego $\alpha \in \mathbb{R}$:

$$\mathbf{v} = \sum_{j=1}^{k} \lambda_j \mathbf{v}_j - \alpha \sum_{j=1}^{k} \mu_j \mathbf{v}_j = \sum_{j=1}^{k} (\lambda_j - \alpha\mu_j) \mathbf{v}_j.$$

W szczególności, jeśli przyjmiemy, że

$$\alpha = \min_{1 \leqslant j \leqslant k} \left\{ \frac{\lambda_j}{\mu_j} : \ \mu_j > 0 \right\} = \frac{\lambda_i}{\mu_i},$$

to $\alpha > 0$, $\lambda_j - \alpha\mu_j \geqslant 0$ dla $1 \leqslant j \leqslant k$ oraz $\lambda_i - \alpha\mu_i = 0$. Zatem $\sum_{j=1}^{k} (\lambda_j - \alpha\mu_j)\mathbf{v}_j$ jest kombinacją wypukłą co najwyżej $k-1$ punktów zbioru $K$. To kończy dowód, ponieważ możemy kontynuować opisane postępowanie aż otrzymamy $k \leqslant n+1$. □

**Wniosek 3.** *Powłoka wypukła zbioru $K$ jest zbiorem wszystkich możliwych kombinacji wypukłych $\mathbf{v} = \sum_{j=1}^{k} \lambda_j \mathbf{v}_j$ o dowolnej liczbie elementów ze zbioru $K$.*

Wielościan jest jednoznacznie określony przez swoje wierzchołki. Ogólniej, zbiór wypukły jest określony przez swoje *punkty ekstremalne*, które teraz zdefiniujemy.

**Definicja 8.** Punkt $\mathbf{v}$ należący do zbioru wypukłego $K$ nazywamy punktem ekstremalnym tego zbioru, jeśli nie jest on punktem wewnętrznym żadnego odcinka całkowicie zawartego w $K$.
Inaczej mówiąc, punkt $\mathbf{v} \in K$ jest punktem ekstremalnym zbioru $K$, jeśli nie jest możliwe przedstawienie go w postaci

$$\mathbf{v} = \lambda \mathbf{v}_1 + (1-\lambda)\mathbf{v}_2,$$

gdzie $\mathbf{v}_1, \mathbf{v}_2 \in K$, $\mathbf{v}_1, \mathbf{v}_2 \neq \mathbf{v}$, $0 \leqslant \lambda \leqslant 1$.
Albo jeszcze trochę inaczej:

$$\Big(\mathbf{v} = \lambda \mathbf{v}_1 + (1-\lambda)\mathbf{v}_2\Big) \Rightarrow \Big(\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}\Big)$$

Dla zbioru wypukłego można też zdefiniować ściany i krawędzie.

**Definicja 9.** *Ścianą* zbioru wypukłego $K$ nazywamy wypukły podzbiór $F \subset K$ taki, że żaden punkt zbioru $F$ nie jest punktem wewnętrznym odcinka o końcach należących do $K$ i nie należących do $F$. Inaczej:

$$\big(\mathbf{v} \in F \text{ i } v = \lambda \mathbf{v}_1 + (1-\lambda)\mathbf{v}_2\big) \Rightarrow \big(\mathbf{v}_1, \mathbf{v}_2 \in F\big)$$

*Krawędzią* zbioru wypukłego K nazywa się jego jednowymiarową ścianę (ścianę będącą odcinkiem, prostą lub półprostą).

**Przykłady.**
1. Punktami ekstremalnymi wielościanu są wierzchołki, a krawędzie i ściany to figury w zwykłym sensie geometrii.
2. W kuli jednostkowej w $\mathbb{R}^n$ punktami ekstremalnymi są wszystkie punkty sfery

$$S^n = \{(x_1, x_2, \ldots, x_n) : \ \sum_{j=1}^{n} x_j^2 = 1\}$$

Kula nie ma ścian ani krawędzi.
Z twierdzenia Carathéodory'ego i definicji punktu ekstremalnego wynika ważny wniosek.

**Wniosek 4. (tw. Minkowskiego)** [4] *Jeśli $K \subset \mathbb{R}^n$ jest zbiorem wypukłym, domkniętym i ograniczonym, to dowolny punkt $\mathbf{v} \in \operatorname{conv} K$ można zapisać jako kombinację wypukłą punktów ekstremalnych zbioru $K$.*

**Przykład.** Niech $K \subset \mathbb{R}^2$ będzie trójkątem o wierzchołkach $(2,1), (6,3), (1,4)$. Przedstawić punkt $(3,3)$ jako kombinację wypukłą punktów ekstremalnych zbioru $K$.
Odp. $(3,3) = \frac{3}{14}(2,1) + \frac{5}{14}(6,3) + \frac{3}{7}(1,4)$.
W dalszych zastosowaniach ważne będzie twierdzenie o osiąganiu ekstremum funkcjonału w punkcie ekstremalnym.

**Twierdzenie 10.** *Niech $K \subset \mathbb{R}^n$ będzie zbiorem wypukłym domkniętym i ograniczonym, a $f : \mathbb{R}^n \to \mathbb{R}$ funkcjonałem liniowym. Jeśli istnieje ekstremum $z_0$ funkcjonału $f$, to istnieje punkt ekstremalny $\mathbf{v} \in K$ taki, że $f(\mathbf{v}) = z_0$.*

**Przykład.** Niech $K \subset \mathbb{R}^2$ będzie czworokątem $ABCD$, gdzie $A = (1,2)$, $B = (5,1)$, $C = (7,3)$, $D = (3,5)$. Niech $\mathbf{v} = (x,y)$.
Funkcjonał $f(x,y) = x + y$ osiąga maksimum w wierzchołku $C$, a minimum w wierzchołku $A$.
Natomiast funkcjonał $g(x,y) = -x+y$ osiąga maksimum w wierzchołku $D$, a minimum w wierzchołkach $B$ i $C$ (tę samą minimalną wartość ma na całym odcinku $BC$).
D o w ó d twierdzenia.
Załóżmy, że funkcjonał ma maksimum (dowód dla minimum jest analogiczny). Niech $\mathbf{v} \in K$ będzie takim punktem, że $f(\mathbf{v}) = z_0$. Niech $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ będą wszystkimi punktami ekstremalnymi, i przypuśćmy, że $f(\mathbf{v}_i) < z_0$ dla $1 \leqslant i \leqslant p$. Z twierdzenia Minkowskiego mamy $\mathbf{v} = \sum_{i=1}^p \lambda_i \mathbf{v}_i$ dla pewnych $\lambda_i \geqslant 0$, $\sum_{i=1}^p \lambda_i = 1$, więc

$$f(\mathbf{v}) = f\big(\sum_{i=1}^p \lambda_i \mathbf{v}_i\big) = \sum_{i=1}^p \lambda_i f(\mathbf{v}_i) < \sum_{i=1}^p \lambda_i z_0 = z_0,$$

sprzeczność. $\square$
Twierdzenie można uogólnić: jeżeli $f : K \to \mathbb{R}$ jest funkcją wypukłą i ciągłą, to istnieje punkt ekstremalny $\mathbf{v} \in K$ taki, że $f(\mathbf{v}) = z_0$.

## 5. Stożki.

Wśród zbiorów wypukłych szczególną rolę w optymalizacji pełnią stożki.

**Definicja 10.** Zbiór $C \subset \mathbb{R}^n$ nazywamy *stożkiem*, jeżeli dla każdego $\mathbf{x} \in C$ i $\lambda \geqslant 0$ punkt $\lambda\mathbf{x}$ także należy do $C$. Jeżeli dodatkowo jest to zbiór wypukły, to nazywamy go *stożkiem wypukłym.*

Zatem $C$ jest stożkiem, jeśli wraz z każdym punktem $\mathbf{x} \in C$ zawiera otwartą półprostą $\{\lambda \cdot \mathbf{x} : \lambda \in \mathbb{R}, \lambda > 0\}$ przechodzącą przez $\mathbf{x}$.
**Przykład.** Podać interpretację geometryczną stożków:
a) $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \geqslant 0\}$,
b) $A = \{\mathbf{x} \in \mathbb{R}^2 : \langle \mathbf{x}, \mathbf{e}_1 \rangle \leqslant 0, \langle \mathbf{x}, \mathbf{e}_1 + \mathbf{e}_2 \rangle \geqslant 0\}$, $B = (\mathbf{e}_1 + \mathbf{e}_2)$, $A + B$, $-A = \{-\mathbf{x} : \mathbf{x} \in A\}$,
c) $D = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \alpha \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}, \alpha \geqslant 0\}$.

**Twierdzenie 11.** *Zbiór $C \subset \mathbb{R}^n$ jest stożkiem wypukłym wtedy i tylko wtedy, gdy*

$$\forall_{\mathbf{x},\mathbf{y} \in C}\ \forall_{\lambda,\mu \geqslant 0} \quad \lambda\mathbf{x} + \mu\mathbf{y} \in C.$$

[4] Hermann Minkowski, 22.06.1864–12.01.1909

D o w ó d.
($\Rightarrow$) Załóżmy, że $C$ jest stożkiem wypukłym. Przynależność $\lambda\mathbf{x}+\mu\mathbf{y}$ do $C$ jest oczywista, gdy $\lambda=\mu=0$. Natomiast gdy $\lambda^2+\mu^2>0$, to punkt

$$\mathbf{z}=\frac{\lambda}{\lambda+\mu}\mathbf{x}+\frac{\mu}{\lambda+\mu}\mathbf{y}$$

należy do $C$ (bo $C$ jest wypukły). Ale $C$ jest stożkiem, więc również punkt $(\lambda+\mu)\mathbf{z}=\lambda\mathbf{x}+\mu\mathbf{y}$ też należy do C.
($\Leftarrow$) Oczywiste. $\square$

**Definicja 11.** *Stożkiem wypukłym* generowanym przez zbiór wypukły $K$ nazywamy zbiór

$$\operatorname{cone}(K)=\{\mathbf{x}\in\mathbb{R}^n:\ \mathbf{x}=\lambda\mathbf{a}\text{ dla pewnych }\lambda\geqslant 0,\mathbf{a}\in K\}$$

Zbiór $\operatorname{cone}(K)$ jest najmniejszym stożkiem wypukłym zawierającym zbiór K.

**Lemat 3.** *Każda podprzestrzeń wektorowa $W\subseteq V$ jest zbiorem wypukłym i stożkiem. Każda podprzestrzeń afiniczna postaci*

$$\mathbf{a}+W=\{\mathbf{v}=\mathbf{a}+\mathbf{w}:\ \mathbf{w}\in W\},$$

*jest zbiorem wypukłym, ale jest stożkiem tylko wtedy, gdy $\mathbf{a}\in W$.*

Niech $K$ będzie zbiorem wypukłym, $\mathbf{a}\in K$ oraz $K-\mathbf{a}=\{\mathbf{x}-\mathbf{a}:\ \mathbf{x}\in K\}$.

**Definicja 12.** Stożek

$$F_K(\mathbf{a})=\operatorname{cone}(K-\mathbf{a})$$

nazywamy *stożkiem kierunków osiągalnych (dopuszczalnych)* w punkcie $\mathbf{a}$.
Jeżeli $\mathbf{v}\in F_K(\mathbf{a})$ i $\mathbf{v}\neq\mathbf{0}$, to wektor $\mathbf{v}$ nazywamy wektorem osiągalnym dla $K$ w punkcie $\mathbf{a}$. Każdy taki wektor wyznacza kierunek wzdłuż którego można się przez pewien czas poruszać nie opuszczając zbioru $K$.
Domknięcie stożka $F_K(\mathbf{a})$ nazywamy *stożkiem kierunków stycznych* do $K$ w punkcie $\mathbf{a}$ i oznaczamy $T_K(\mathbf{a})$.

Jeśli zbiór $K$ nie jest wypukły, to stożek kierunków stycznych definiuje się nieco inaczej, ale to nie będzie nam potrzebne.

## 6. Hiperpłaszczyzny

W tym rozdziale $V$ jest przestrzenią unormowaną.

**Definicja 13.** *Rozmaitością liniową* lub *przestrzenią afiniczną* w przestrzeni $V$ nazywamy zbiór $A$ postaci $\mathbf{v}+W$, gdzie $\mathbf{v}\in V$, a $W$ jest podprzestrzenią.

Inaczej, rozmaitość jest to przesunięta podprzestrzeń.
Np. w $\mathbb{R}^2$ rozmaitościami są proste, a w $\mathbb{R}^3$ rozmaitościami są proste i płaszczyzny.

**Lemat 4.** *Zbiór $A$ jest przestrzenią afiniczną wtedy i tylko wtedy, gdy $\sum_{i=1}^{k}\lambda_i\mathbf{v}_i\in A$ dla dowolnych $(\mathbf{v}_i)\subset A$ i skalarów $(\lambda_i)\subset\mathbb{R}$ takich, że $\sum_{i=1}^{k}\lambda_i=1$.*

Wymiarem rozmaitości liniowej nazywamy wymiar związanej z nią podprzestrzeni liniowej.
Mając dany podzbiór $S\subset V$ możemy zbudować najmniejszą rozmaitość liniową zawierającą $S$. Jest to przekrój wszystkich rozmaitości liniowych zawierających $S$. Nazywamy go też otoczką afiniczną zbioru $U$.

**Definicja 14.** *Otoczką afiniczną* zbioru $U \subset V$ nazywamy podprzestrzeń afiniczną generowaną przez $U$, tzn.:

$$\operatorname{aff} U = \Big\{ \sum_{i=1}^{k} \lambda_i \mathbf{v}_i : \ \mathbf{v}_1, \ldots, \mathbf{v}_k \in U, \ \sum_{i=1}^{k} \lambda_i = 1 \Big\}$$

Kombinacje $\sum\limits_{i=1}^{k} \lambda_i \mathbf{v}_i$ nazywamy *kombinacjami afinicznymi.* W odróżnieniu od kombinacji wypukłych nie ma tu warunku $\lambda_i \geqslant 0$.

**Definicja 15.** *Hiperpłaszczyzna* $H$ w przestrzeni $V$ jest to największa właściwa rozmaitość liniowa, tzn. taka rozmaitość $H \neq V$, że jeśli istnieje rozmaitość liniowa $G$ taka, że $H \subset G$, to $G = V$ lub $G = H$.

Hiperpłaszczyzny są ściśle związane z funkcjonałami liniowymi.

**Twierdzenie 12.** *Niech $H$ będzie hiperpłaszczyzną w przestrzeni $V$. Istnieje wtedy taki funkcjonał liniowy $f$ określony na $V$ i taka stała $c$, że $H = \{\mathbf{v} : \ f(\mathbf{v}) = c\}$. I na odwrót, jeżeli $f$ jest niezerowym funkcjonałem liniowym określonym na $V$, to zbiór $\{\mathbf{v} : \ f(\mathbf{v}) = c\}$ jest hiperpłaszczyzną w $V$.*

Np. $H = \{\mathbf{x} : \ 3x_1 - 5x_2 + x_3 = 5\}$ jest hiperpłaszczyzną w przestrzeni $\mathbb{R}^3$. Odpowiadający jej funkcjonał to

$$f(\mathbf{x}) = 3x_1 - 5x_2 + x_3 = [3 \ \ -5 \ \ \ 1]\mathbf{x}.$$

Powyższe twierdzenie umożliwia interpretację funkcjonałów jako hiperpłaszczyzn znajdujących się w przestrzeni pierwotnej $V$, a tym samym połączenie elementów przestrzeni $V$ i $V^*$ w jeden twór geometryczny.
Gdy $V = \mathbb{R}^n$, to każdy funkcjonał liniowy określony jest przez pewien wektor. Dokładniej:

**Lemat 5.** *$f$ jest funkcjonałem liniowym wtedy i tylko wtedy, gdy istnieje $\mathbf{a} \in \mathbb{R}^n$ takie, że $f(\mathbf{v}) = \mathbf{a}^T \mathbf{v}$ dla dowolnego $\mathbf{v} \in V$.*

**Twierdzenie 13.** *Niech $H$ będzie hiperpłaszczyzną w przestrzeni $V$. Jeżeli $H$ nie zawiera wektora zerowego, to istnieje dokładnie jeden funkcjonał liniowy $f$ określony na $V$ taki, że $H = \{\mathbf{v} : \ f(\mathbf{v}) = 1\}$.*

W przestrzeni $\mathbb{R}^2$ twierdzenie to oznacza, że każda prosta nieprzechodząca przez początek układu daje się opisać równaniem postaci $ax + by = 1$. Analogicznie, w $\mathbb{R}^3$: każda płaszczyzna nieprzechodząca przez początek układu daje się opisać równaniem $ax + by + cz = 1$.

**Twierdzenie 14.** *Niech $f$ będzie niezerowym funkcjonałem liniowym określonym na unormowanej przestrzeni $V$. Wówczas dla dowolnego $c$, hiperpłaszczyzna $H = \{\mathbf{v} : f(\mathbf{v}) = c\}$ jest domknięta wtedy i tylko wtedy, gdy funkcjonał $f$ jest ciągły.*

Jeżeli $f$ jest niezerowym funkcjonałem liniowym określonym na liniowej przestrzeni $V$, to z hiperpłaszczyzną $H = \{\mathbf{v} : \ f(\mathbf{v}) = c\}$ związane są cztery zbiory:

$$\{\mathbf{v} : \ f(\mathbf{v}) \leqslant c\}, \quad \{\mathbf{v} : \ f(\mathbf{v}) < c\}, \quad \{\mathbf{v} : \ f(\mathbf{v}) \geqslant c\}, \quad \{\mathbf{v} : \ f(\mathbf{v}) > c\},$$

które nazywamy *półprzestrzeniami* wyznaczonymi przez $H$. Pierwsze dwa zbiory nazywamy ujemnymi półprzestrzeniami określonymi przez $f$, a następne dwa dodatnimi półprzestrzeniami.

**Lemat 6.** *Niech $H = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}^T \mathbf{x} = \alpha\}$ będzie hiperpłaszczyzną w $\mathbb{R}^n$ i niech*

$$H_+ = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}^T \mathbf{x} \geqslant \alpha\}, \quad H_- = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}^T \mathbf{x} \leqslant \alpha\}$$

*będą półprzestrzeniami domkniętymi, na które $H$ dzieli przestrzeń $\mathbb{R}^n$. Każdy ze zbiorów $H, H_+, H_-$ jest zbiorem wypukłym w $\mathbb{R}^n$.*

## 7. Hiperpłaszczyzny i zbiory wypukłe

Jeżeli $V$ jest przestrzenią unormowaną, to $V^*$ oznacza *przestrzeń sprzężoną*, tj. przestrzeń funkcjonałów liniowych ciągłych określonych na przestrzeni $V$.

**Definicja 16.** Niech $K$ będzie zbiorem wypukłym w unormowanej, liniowej, rzeczywistej przestrzeni $V$ i załóżmy, że $\mathbf{0}$ jest punktem wewnętrznym $K$ (ten warunek zapewnia, że $K$ jest *zbiorem pochłaniającym*, tzn. dla każdego elementu $\mathbf{v}$ przestrzeni $V$ istnieje taka liczba dodatnia $\alpha$, że $\mathbf{v} \in \alpha K$).
*Funkcjonał Minkowskiego* (ang.: gauge function) $p$ zbioru $K$ określamy wzorem

$$p(\mathbf{v}) = \inf\big\{r \geqslant 0 : \ \mathbf{v} \in rK\big\}.$$

W szczególnym przypadku, gdy $K$ jest kulą jednostkową w $V$, mamy $p(\mathbf{v}) = \|\mathbf{v}\|$.

Zatem $p(\mathbf{v})$ określa, jak należy rozszerzyć zbiór $K$, aby to rozszerzenie zawierało $\mathbf{v}$.
Własności funkcjonału Minkowskiego:
1. $0 \leqslant p(\mathbf{v}) < \infty$;
2. $p(\alpha\mathbf{v}) = \alpha p(\mathbf{v})$ dla $\alpha \geqslant 0$;
3. $p(\mathbf{v}_1 + \mathbf{v}_2) \leqslant p(\mathbf{v}_1) + p(\mathbf{v}_2)$;
4. $p$ jest ciągły;
5. $\overline{K} = \{\mathbf{v} : \ p(\mathbf{v}) \leqslant 1\}$ oraz $\operatorname{int} K = \{\mathbf{v} : \ p(\mathbf{v}) < 1\}$

**Twierdzenie 15. (Mazura)** [5] *Niech $K$ będzie zbiorem wypukłym, posiadającym niepuste wnętrze, zawartym w unormowanej, liniowej, rzeczywistej przestrzeni $V$. Przypuśćmy, że $W$ jest podprzestrzenią afiniczną nie zawierającą punktów wewnętrznych zbioru $K$. Wówczas istnieje w $V$ domknięta hiperpłaszczyzna zawierająca $W$, lecz nie zawierająca punktów wewnętrznych zbioru $K$, tzn. istnieje taki element $v^* \in V^*$ i taka stała $c$, że $v^*(\mathbf{v}) = c$ dla każdego $\mathbf{v} \in W$ i $v^*(\mathbf{k}) < c$ dla każdego $\mathbf{k} \in \operatorname{int} K$.*

D o w ó d. Można założyć, że $\mathbf{0} \in K$ (gdyby tak nie było, to $K$ można przesunąć). Niech $M$ będzie podprzestrzenią $V$ generowaną przez $W$. Wówczas $W$ jest hiperpłaszczyzną w $M$ i nie zawiera $\mathbf{0}$, istnieje zatem funkcjonał liniowy $f$ określony na $M$ i taki, że $W = \{\mathbf{v} : \ f(\mathbf{v}) = 1\}$. Jeżeli $p$ jest funkcjonałem Minkowskiego zbioru $K$, to ponieważ $\operatorname{int} K = \{\mathbf{v} : \ p(\mathbf{v}) < 1\}$ oraz $W \cap \operatorname{int} K = \emptyset$, więc

$$1 = f(\mathbf{v}) \leqslant p(\mathbf{v}), \quad \mathbf{v} \in W.$$

Stąd z jednorodności funkcjonału $f$ wynika:

$$f(\alpha\mathbf{v}) = \alpha \leqslant p(\alpha\mathbf{v}), \quad \mathbf{v} \in W, \ \alpha > 0,$$

oraz

$$f(\alpha\mathbf{v}) \leqslant 0 \leqslant p(\alpha\mathbf{v}), \quad \mathbf{v} \in W, \ \alpha < 0.$$

Zatem $f(\mathbf{v}) \leqslant p(\mathbf{v})$ dla $\mathbf{v} \in M$. Na podstawie twierdzenia Hahna-Banacha istnieje funkcjonał $F$ będący rozszerzeniem $f$ na całą przestrzeń $V$ i taki, że $F(\mathbf{v}) \leqslant p(\mathbf{v})$. Niech $H = \{\mathbf{v} : \ F(\mathbf{v}) = 1\}$. $F$ jest funkcjonałem ciągłym oraz $F(\mathbf{v}) < 1$ dla $\mathbf{v} \in K$, zatem $H$ jest szukaną domkniętą hiperpłaszczyzną, co kończy dowód. □

**Definicja 17.** Domkniętą hiperpłaszczyznę $H$ w unormowanej przestrzeni $V$ nazywamy *hiperpłaszczyzną podpierającą* zbiór wypukły $K$, jeżeli $K$ jest zawarte w jednej z domkniętych półprzestrzeni określonych przez $H$ oraz $H$ zawiera punkt z $K$.

**Twierdzenie 16. (o hiperpłaszczyźnie podpierającej)** *Jeżeli $\mathbf{v}$ nie jest punktem wewnętrznym wypukłego zbioru $K$ o niepustym wnętrzu, to istnieje domknięta hiperpłaszczyzna $H$ zawierająca $\mathbf{v}$ i taka, że $K$ leży po jednej stronie $H$.*

Wynika stąd, że mając dany zbiór wypukły $K$ o niepustym wnętrzu można zbudować hiperpłaszczyznę podpierającą przechodzącą przez dowolny punkt zbioru $\overline{K}$.

[5] Stanisław Mazur, 1.01.1905–5.11.1981

**Twierdzenie 17. (Eidelheita o oddzielaniu)** [6] *Niech $K_1$ i $K_2$ będą zbiorami wypukłymi zawartymi w przestrzeni $V$ i takimi, że $K_1$ ma niepuste wnętrze, a $K_2$ nie zawiera punktów wewnętrznych zbioru $K_1$. Wówczas istnieje domknięta hiperpłaszczyzna $H$ oddzielająca zbiory $K_1$ i $K_2$, tzn. istnieje element $v^* \in V^*$ taki, że*

$$\sup_{\mathbf{v}\in K_1} v^*(\mathbf{v}) \leqslant \inf_{\mathbf{v}\in K_2} v^*(\mathbf{v}).$$

*Innymi słowami, $K_1$ i $K_2$ leżą po przeciwnych stronach hiperpłaszczyzny $H$.*

D o w ó d. Niech $K = K_1 - K_2$. Z założeń wynika, że $\operatorname{int} K \neq \emptyset$ oraz $\mathbf{0} \neq \operatorname{int} K$. Na podstawie twierdzenia 16 istnieje $v^* \in V$, $v^* \neq 0$ takie, że $v^*(\mathbf{v}) \leqslant 0$ dla $\mathbf{v} \in K$. Zatem dla $\mathbf{v}_1 \in K_1$, $\mathbf{v}_2 \in K_2$ jest $v^*(\mathbf{v}_1 - \mathbf{v}_2) \leqslant 0$, czyli $v^*(\mathbf{v}_1) \leqslant v^*(\mathbf{v}_2)$. A więc istnieje liczba $c$ taka, że

$$\sup_{K_1} v^*(\mathbf{k}_1) \leqslant c \leqslant \inf_{K_2} v^*(\mathbf{k}_2).$$

Szukaną hiperpłaszczyzną jest więc

$$H = \{\mathbf{v} :\ v^*(\mathbf{v}) = c\}. \ \square$$

**Wniosek 5.** *Jeżeli $K$ jest domkniętym zbiorem wypukłym i $\mathbf{v} \notin K$, to istnieje domknięta półprzestrzeń zawierająca $K$, lecz nie zawierająca $\mathbf{v}$.*

D o w ó d. Niech $d = \inf\limits_{k\in K} \|\mathbf{v} - \mathbf{k}\|$. Z domkniętości zbioru $K$ wynika, że $d > 0$. Niech $S$ będzie kulą otwartą o środku $\mathbf{v}$ i promieniu $\frac{1}{2}d$. Teza wynika z twierdzenia Eidelheita zastosowanego do zbiorów $K$ i $S$. $\square$

## 8. Funkcje wypukłe

**Definicja 18.** Funkcję $f : W \to \mathbb{R}$ określoną na zbiorze wypukłym $W \subset \mathbb{R}^n$ nazywamy *funkcją wypukłą* na $W$, jeśli dla każdych $\mathbf{x}, \mathbf{y} \in W$ i każdego $\lambda \in (0, 1)$ zachodzi

$$f\big(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\big) \leqslant \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}). \tag{6}$$

Jeżeli nierówność (6) jest ostra dla $\mathbf{x} \neq \mathbf{y}$, to funkcję nazywamy *ściśle wypukłą*.

Funkcja $f$ jest (ściśle) wklęsła, jeśli funkcja $-f$ jest (ściśle) wypukła.
Można wykazać, że jeżeli $f$ jest funkcją ciągłą, to nierówność (6) jest równoważna nierówności:

$$f\Big(\frac{\mathbf{x}+\mathbf{y}}{2}\Big) \leqslant \frac{f(\mathbf{x}) + f(\mathbf{y})}{2}. \tag{7}$$

**Przykłady.**
— Funkcja afiniczna $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$ jest wypukła i wklęsła.
— Norma w $\mathbb{R}^n$ jest wypukła.
— Odległość punktu od zbioru wypukłego $W$, tj. funkcja $f(\mathbf{x}) = \inf_{\mathbf{y}\in W} \|\mathbf{x} - \mathbf{y}\|$, jest wypukła.
Gdy $W \subset \mathbb{R}^2$, to przyjmujemy $\mathbf{x} = (x_1, x_2)$, $\mathbf{a} = (a_1, a_2)$. Wtedy powyższe funkcje mają postać:
— $f(x_1, x_2) = a_1x_1 + a_2x_2 + b$;.
— $\sqrt{x_1^2 + x_2^2}$;
— $f(x_1, x_2) = \inf_{(y_1,y_2)\in W} \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$.
Zdefiniujemy teraz dwa zbiory związane z funkcją .

[6] Meier (Maks) Eidelheit, 6.07.1910– marzec 1943

**Definicja 19.** *Epigrafem (nadwykresem)* funkcji $f : W \to \mathbb{R}$ określonej na zbiorze wypukłym $W \subset \mathbb{R}^n$ nazywamy zbiór

$$\mathrm{epi}(f) = \{(\mathbf{x}, z) \in W \times \mathbb{R} : \ z \geqslant f(\mathbf{x})\}.$$

**Definicja 20.** *Zbiorem poziomicowym* funkcji $f : W \to \mathbb{R}$ określonej na zbiorze wypukłym $W \subset \mathbb{R}^n$ nazywamy zbiór

$$W_\alpha(f) = \{\mathbf{x} \in W : \ f(\mathbf{x}) \leqslant \alpha\}, \quad \alpha \in \mathbb{R}.$$

**Twierdzenie 18.** *Funkcja $f$ jest wypukła wtedy i tylko wtedy, gdy jej epigraf jest wypukłym podzbiorem $\mathbb{R}^{n+1}$.*

**Twierdzenie 19.** *Jeżeli funkcja $f$ jest wypukła, to dla dowolnego $\alpha \in \mathbb{R}$ zbiór poziomicowy $W_\alpha(f)$ jest wypukły.*

Stosując twierdzenie 16 o hiperpłaszczyźnie podpierającej udowodnimy ważną własność funkcji wypukłej.

**Twierdzenie 20.** *Jeżeli funkcja $f$ jest wypukła, to dla dowolnego $\mathbf{x}_0 \in \mathrm{int}\, W$ istnieje $\mathbf{a} \in \mathbb{R}^n$ takie, że*

$$f(\mathbf{x}) \geqslant f(\mathbf{x}_0) + \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0), \quad \mathbf{x} \in W. \tag{8}$$

*Jeśli funkcja $f$ jest ściśle wypukła, to*

$$f(\mathbf{x}) > f(\mathbf{x}_0) + \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0), \quad \mathbf{x} \in W \setminus \{\mathbf{x_0}\}.$$

*Jeżeli funkcja $f$ jest różniczkowalna w $\mathbf{x}_0$, to w obu powyższych nierównościach można przyjąć $\mathbf{a} = Df(\mathbf{x}_0)^T$.*

Dowód. Zastosujemy twierdzenie 16 dla $V = \mathbb{R}^n \times \mathbb{R}$, $K = \mathrm{epi}(f)$ i $\mathbf{v} = (\mathbf{x}_0, f(\mathbf{x}_0))$. Hiperpłaszczyzna w $\mathbb{R}^{n+1}$ jest określona niezerowym wektorem $\mathbf{b} = (\mathbf{a}, \alpha) \in \mathbb{R}^n \times \mathbb{R}$, więc z twierdzenia 16 mamy:

$$\mathbf{a}^T\mathbf{x} + \alpha y \leqslant \mathbf{a}^T\mathbf{x}_0 + \alpha f(\mathbf{x}_0) \tag{9}$$

dla $(\mathbf{x}, y) \in \mathrm{epi}(f)$. Ponieważ nierówność jest dla wszystkich $y \geqslant f(\mathbf{x})$, więc $\alpha \leqslant 0$. Wykażemy, że $\alpha < 0$. Gdyby $\alpha = 0$, to z (9) byłoby $\mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) \leqslant 0$. Przyjmując $\mathbf{x} = \mathbf{x}_0 + \varepsilon\mathbf{a}$ (taki wektor należy do $W$ dla małych $\varepsilon > 0$) otrzymamy

$$0 \geqslant \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) = \varepsilon\mathbf{a}^T\mathbf{a} = \varepsilon\|\mathbf{a}\|^2,$$

a zatem $\mathbf{a} = \mathbf{0}$. Wektor $\mathbf{b} = (\mathbf{a}, \alpha)$ byłby więc równy zero, co jest niemożliwe.
Skoro $\alpha < 0$, to możemy przyjąć $\alpha = -1$ i nierówność (9) przyjmuje postać

$$\mathbf{a}^T\mathbf{x} - y \leqslant \mathbf{a}^T\mathbf{x}_0 - f(\mathbf{x}_0), \quad \mathbf{x} \in W,$$

czyli

$$y \geqslant f(\mathbf{x}_0) + \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0)$$

dla wszystkich $y \geqslant f(\mathbf{x})$. Zatem

$$f(\mathbf{x}) \geqslant f(\mathbf{x}_0) + \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0).$$

Nieskomplikowane uzasadnienie, że nierówność jest ścisła dla funkcji ściśle wypukłej pominiemy.
Załóżmy, że funkcja $f$ jest różniczkowalna w $\mathbf{x}_0$. Wtedy z wypukłości mamy

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}_0) &= \tfrac{(1-\lambda)f(\mathbf{x}_0)+\lambda f(\mathbf{x})-f(\mathbf{x}_0)}{\lambda} \geqslant \tfrac{f((1-\lambda)\mathbf{x}_0+\lambda\mathbf{x})-f(\mathbf{x}_0)}{\lambda} = \\ &= \tfrac{f(\mathbf{x}_0+\lambda(\mathbf{x}-\mathbf{x}_0))-f(\mathbf{x}_0)}{\lambda}. \end{aligned}$$

Przy $\lambda \to 0$ prawa strona dąży do pochodnej, więc

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

więc w nierówności (8) możemy przyjąć $\mathbf{a} = Df(\mathbf{x}_0)^T$. $\square$
W najprostszym (jednowymiarowym) przypadku powyższe twierdzenie mówi, ze wykres funkcji wypukłej leży powyżej stycznej.
**Przykład.** Niech $f : \mathbb{R}^2 \to \mathbb{R}$, $f(\mathbf{x}) = 2x_1^2 + x_2^2 - 3$. Biorąc np. $\mathbf{x}_0 = (1, 2)$ mamy $f(\mathbf{x}_0) = 3$, $Df(\mathbf{x}_0)^T = (4, 4)$, i nierówność (8) ma postać

$$2x_1^2 + x_2^2 - 3 \geqslant 3 + (4, 4) \cdot (x_1 - 1, x_2 - 2) = 4x_1 + 4x_2 - 9,$$

czyli równoważnie $2(x_1 - 1)^2 + (x_2 - 2)^2 \geqslant 0$.
Płaszczyzna $x_3 = 4x_1 + 4x_2 - 9$ jest hiperpłaszczyzną podpierającą epigraf funkcji $f$ w punkcie $(1, 2, 3)$. $\square$

**Wniosek 6.** *Jeśli $f : W \to \mathbb{R}$ jest wypukła i różniczkowalna, to w punkcie $\mathbf{x}_0 \in \operatorname{int} W$ jest minimum globalne wtedy i tylko wtedy, gdy $Df(\mathbf{x}_0) = 0$.*

D o w ó d. ($\Rightarrow$) jest znane. ($\Leftarrow$) wynika z tego, że jeśli $Df(\mathbf{x}_0) = 0$, to nierówność (8) ma postać

$$f(\mathbf{x}) \geqslant f(\mathbf{x}_0) \quad \text{dla } x \in W. \square$$

## 9. Funkcje quasiwypukłe i pseudowypukłe

Niech $f : W \to \mathbb{R}$, gdzie $W \subset \mathbb{R}^n$ jest zbiorem otwartym.

**Definicja 21.** Funkcję $f : W \to \mathbb{R}$ określoną na zbiorze wypukłym $W \subset \mathbb{R}^n$ będziemy nazywać *funkcją quasiwypukłą* na $W$, jeśli dla każdych $\mathbf{x}, \mathbf{y} \in W$ i każdego $\lambda \in [0, 1]$ zachodzi

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leqslant \max\{f(\mathbf{x}), f(\mathbf{y})\},$$

i odpowiednio *funkcją quasiwklęsłą* na $W$, jeśli przy tych samych założeniach spełniona jest nierówność $f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \geqslant \min\{f(\mathbf{x}), f(\mathbf{y})\}$

Innymi słowy, funkcja jest quasiwklęsła, jeśli na odcinku łączącym punkty $\mathbf{x}, \mathbf{y}$ przyjmuje wartości nie mniejsze od mniejszej z wartości na krańcach tego odcinka (tj. minimum funkcji jest osiągane na jednym z końców odcinka), a quasiwypukła, jeśli na odcinku łączącym punkty $\mathbf{x}, \mathbf{y}$ przyjmuje wartości nie większe od większej z wartości na krańcach tego odcinka.
Funkcje wypukłe (wklęsłe) są quasiwypukłe (quasiwklęsłe).

**Twierdzenie 21.** *Funkcja $f : W \to \mathbb{R}$ określona na zbiorze wypukłym $W \subset \mathbb{R}^n$ jest quasiwypukła na $X$ wtedy i tylko wtedy, gdy dla każdej liczby $\alpha \in \mathbb{R}$ zbiór $\{\mathbf{x} \in \mathbb{R} : f(\mathbf{x}) \leqslant \alpha\}$ jest wypukły. Analogicznie, $f$ jest funkcją quasiwklęsłą na $W$ wtedy i tylko wtedy, gdy zbiór $\{\mathbf{x}) \in \mathbb{R} : f(\mathbf{x} \geqslant \alpha\}$ jest wypukły dla każdej liczby $\alpha \in \mathbb{R}$.*

D o w ó d. Dowiedziemy charakteryzacji quasiwklęsłości funkcji — pozostały przypadek jest w pełni analogiczny. Niech $\alpha \in \mathbb{R}$ będzie dowolne. Załóżmy najpierw, że $f$ jest quasiwklęsła i niech $\mathbf{x}, \mathbf{y} \in G(\alpha) = \{\mathbf{x} \in \mathbb{R} : f(\mathbf{x}) \geqslant \alpha\}$ . Dla dowolnego $\lambda \in [0, 1]$ mamy zatem $f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \geqslant \min\{f(\mathbf{x}), f(\mathbf{y})\}$, gdyż obie wartości $f(\mathbf{x}, f(\mathbf{y})$ funkcji $f$ są nie mniejsze niż $\alpha$. A zatem $\lambda\mathbf{x}+(1-\lambda)\mathbf{y} \in G(\alpha)$. Odwrotnie, jeśli zbiór $G(\alpha)$ jest wypukły dla każdego $\alpha \in \mathbb{R}$, to dla dowolnie obranych dowolnie $\mathbf{x}, \mathbf{y} \in W$ określamy $\alpha = \min\{f(\mathbf{x}), f(\mathbf{y})\}$. Dla tej wartości zbiór $G(\alpha)$ zawiera $\mathbf{x}, \mathbf{y}$, więc także $\lambda\mathbf{x}+(1-\lambda)\mathbf{y}$, czyli $f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \geqslant \min\{f(\mathbf{x}), f(\mathbf{y})\}$, co trzeba było wykazać.
Podamy jeszcze jedną definicję.

**Definicja 22.** (funkcja pseudowypukła) Funkcja różniczkowalna $f : W \to \mathbb{R}$ nazywa się funkcją pseudowypukłą, gdy

$$\forall \mathbf{y} \in W : \ Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geqslant 0 \Rightarrow \mathbf{y} \geqslant \mathbf{x}$$

Jeśli

$$\forall \mathbf{y} \in W : \ Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geqslant 0 \Rightarrow \mathbf{y} > \mathbf{x}$$

to funkcję nazywamy ściśle pseudowypukłą. Jeśli funkcja $-f$ jest pseudowypukła, to mówimy, że funkcja $f$ jest pseudowklęsła.

Inaczej, funkcja jest pseudowklęsła gdy spełniona jest implikacja z odwróconymi nierównościami w poprzedniku i następniku.
Podobnie jak poprzednia, powyższa definicja jest rozszerzeniem definicji wypukłości, gdyż różniczkowalne funkcje wypukłe są pseudowypukłe (ale nie na odwrót), a nadto funkcje pseudowypukłe są quasiwypukłe.
Wiadomo, że dla funkcji wypukłej $f : W \to \mathbb{R}$ określonej na zbiorze wypukłym $W \subset \mathbb{R}^n$ prawdziwy jest warunek

$$Df(\mathbf{x}) = 0 \Leftrightarrow \text{ w } \mathbf{x} \text{ jest minimum globalne.}$$

Wykażemy, że jest on prawdziwy także dla funkcji pseudowypukłych.

**Twierdzenie 22.** *Niech $f : W \to \mathbb{R}$, gdzie $W \subset \mathbb{R}^n$ wypukły, otwarty i niepusty, będzie funkcją pseudowypukłą w punkcie $\mathbf{x}_0 \in W$. Wtedy*

$$Df(\mathbf{x}_0) = 0 \Leftrightarrow \ \textit{w } \mathbf{x}_0 \textit{ jest minimum globalne.}$$

D o w ó d. Załóżmy, że $Df(\mathbf{x}_0) = 0$. Z definicji funkcji pseudowypukłej mamy

$$\forall \mathbf{y} \in W : \ \underbrace{Df(\mathbf{x}_0)}_{=0}(\mathbf{y} - \mathbf{x}_0) \geqslant 0 \Rightarrow \mathbf{y} \geqslant \mathbf{x}_0.$$

Zatem następnik implikacji jest prawdziwy. W drugą stronę twierdzenie oczywiste. □