Wypukłość funkcji i nierówność Jensena.

Zbiór na płaszczyźnie[1] nazywamy wypukłym, jeżeli wraz z każdymi dwoma punktami zawiera odcinek o końcach w tych punktach.

DEFINICJA: Funkcję f określoną na przedziale[2] I nazywamy **wypukłą**[3], jeżeli obszar nad jej wykresem[4] jest zbiorem wypukłym na płaszczyźnie. Obszarem tym jest zbiór

$$\{(x,y):\ x\in I\ \wedge\ y\geqslant f(x)\}\ .$$

Nietrudno zauważyć, że specyfika postaci powyższego zbioru powoduje, że jego wypukłość jest równoważna temu, że wraz z każdymi dwoma punktami wykresu funkcji f zawiera on cięciwę o końcach w tych punktach. Możemy więc przeformułować definicję wypukłości:

DEFINICJA: Funkcję f określoną na przedziale I nazywamy **wypukłą**, jeżeli każda cięciwa jej wykresu leży na lub nad jej wykresem.

Ponieważ cięciwa o końcach $(x_1, f(x_1))$ i $(x_2, f(x_2))$ składa się z punktów postaci

$$(a\cdot x_1+(1-a)\cdot x_2,\ a\cdot f(x_1)+(1-a)\cdot f(x_2))\,, \qquad a\in[0,1]$$

definicję wypukłości można przepisać używając wzoru, który powie, że powyższy punkt leży nie niżej niż odpowiedni punkt wykresu funkcji f, czyli punkt

$$(a\cdot x_1+(1-a)\cdot x_2,\ f(a\cdot x_1+(1-a)\cdot x_2))\,.$$

DEFINICJA: Funkcję f określoną na przedziale I nazywamy **wypukłą**, jeżeli dla każdych $x_1, x_2\in I$ oraz każdej liczby $a\in[0,1]$ zachodzi nierówność

$$a\cdot f(x_1)+(1-a)\cdot f(x_2)\geqslant f(a\cdot x_1+(1-a)\cdot x_2)\,.$$

Nieco inne oznaczenia[5] pozwalają przepisać powyższą definicję jako:

DEFINICJA: Funkcję f określoną na przedziale I nazywamy **wypukłą**, jeżeli dla każdych $x_1, x_2\in I$ oraz każdych liczb nieujemnych a_1 i a_2 spełniających warunek $a_1+a_2=1$ zachodzi nierówność

$$a_1\cdot f(x_1)+a_2\cdot f(x_2)\geqslant f(a_1\cdot x_1+a_2\cdot x_2)\,.$$

[1] Lub w euklidesowej przestrzeni trójwymiarowej lub nawet w $\mathbb{R}^n$.

[2] Dziedzina funkcji wypukłej sama musi być wypukła, co w przypadku podzbioru zbioru liczb rzeczywistych oznacza, że musi być w jednym kawałku (bez dziur), czyli musi być przedziałem. Domkniętym lub otwartym lub z jednej strony takim, a z drugiej siakim. Ograniczonym lub nieograniczonym.

[3] W domyśle: słabo wypukłą.

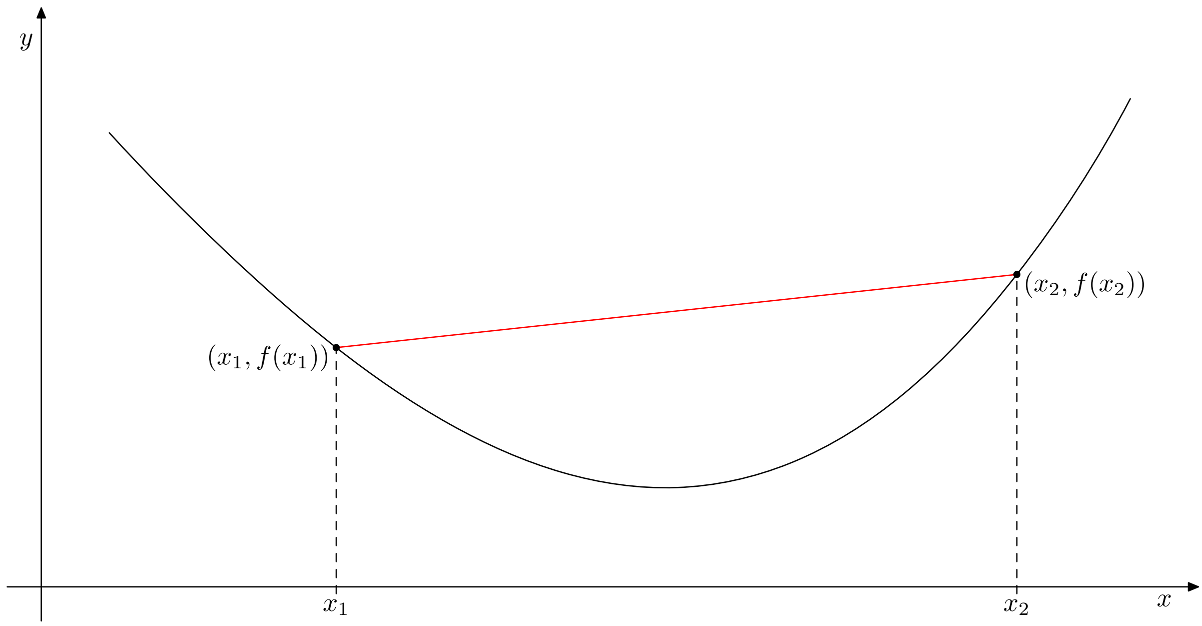
[4] Dla wygody przyjmijmy, że myślimy o obszarze nad wykresem wraz z tym wykresem. Równie dobrze można byłoby konsekwentnie przyjąć w definicji, że myślimy o obszarze nad wykresem (bez samego wykresu), gdyż wypukłość tego obszaru nie zależy od tego, czy zaliczamy do niego wykres funkcji czy nie.

[5] Które w tym momencie mogą wydawać się nieco dziwne, ale wkrótce staną się naturalne i zrozumiałe.

Definicja: Funkcję f określoną na przedziale I nazywamy **ściśle wypukłą,** jeżeli dla każdych różnych $x_1, x_2 \in I$ oraz każdych liczb dodatnich a_1 i a_2 spełniających warunek $a_1 + a_2 = 1$ zachodzi nierówność

$$a_1 \cdot f(x_1) + a_2 \cdot f(x_2) > f(a_1 \cdot x_1 + a_2 \cdot x_2)\,.$$

Innymi słowy: funkcja jest ściśle wypukła, jeżeli każda cięciwa jej wykresu leży **nad wykresem**[6] (jak na rysunku 1).



rys. 1

Definicja: Funkcję f określoną na przedziale I nazywamy **wklęsłą**[7], jeżeli dla każdych $x_1, x_2 \in I$ oraz każdych liczb nieujemnych a_1 i a_2 spełniających warunek $a_1 + a_2 = 1$ zachodzi nierówność

$$a_1 \cdot f(x_1) + a_2 \cdot f(x_2) \leqslant f(a_1 \cdot x_1 + a_2 \cdot x_2)\,.$$

Definicja: Funkcję f określoną na przedziale I nazywamy **ściśle wklęsłą,** jeżeli dla każdych różnych $x_1, x_2 \in I$ oraz każdych liczb dodatnich a_1 i a_2 spełniających warunek $a_1 + a_2 = 1$ zachodzi nierówność

$$a_1 \cdot f(x_1) + a_2 \cdot f(x_2) < f(a_1 \cdot x_1 + a_2 \cdot x_2)\,.$$

Innymi słowy: funkcja jest ściśle wklęsła, jeżeli każda cięciwa jej wykresu leży **pod wykresem**.

[6] Wyjąwszy rzecz jasna punkty końcowe cięciwy, które z definicji cięciwy leżą na wykresie.

[7] W domyśle: słabo wklęsłą.

Jeżeli na wykresie funkcji f wybierzemy trzy różne punkty, odpowiadające argumentom x_1, x_2 i x_3, to trójkąt o wierzchołkach w tych punktach (rys. 2) składa się z punktów postaci

$$(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3,\ a_1\cdot f(x_1)+a_2\cdot f(x_2)+a_3\cdot f(x_3))\ ,$$

gdzie a_1, a_2, a_3 są liczbami dodatnimi[8] o sumie 1. Jeśli ponadto funkcja jest ściśle wypukła[9], to punkt

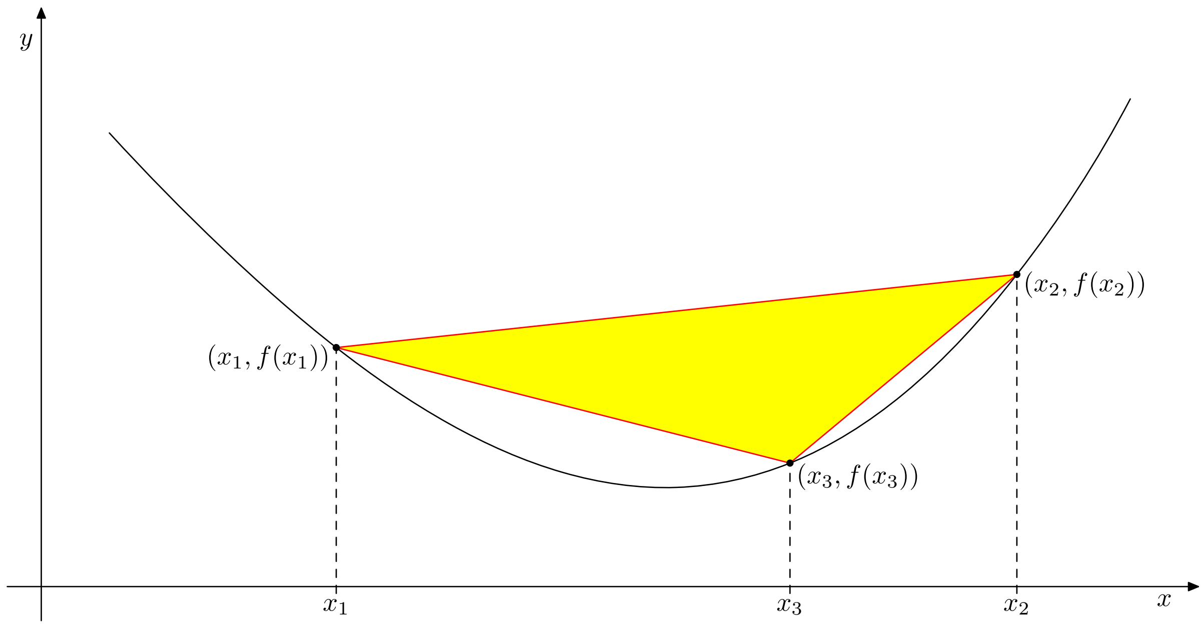
$$(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3,\ a_1\cdot f(x_1)+a_2\cdot f(x_2)+a_3\cdot f(x_3))$$

leży powyżej[10] odpowiadającego mu punktu wykresu

$$(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3,\ f(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3))\ .$$

To oznacza, że zachodzi nierówność[11]

$$a_1\cdot f(x_1)+a_2\cdot f(x_2)+a_3\cdot f(x_3)>f(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3)\,.$$



rys. 2

Analogiczne nierówności zachodzą, gdy mamy nie trzy, a więcej punktów. Formalny dowód można przeprowadzić indukcyjnie ze względu na liczbę punktów, ale ja go pominę.

Otrzymaną nierówność nazywamy nierównością Jensena. Sformułuję jej cztery warianty (dla funkcji wypukłej/wklęsłej w wersji słabej/ostrej).

---

[8] Dodatnimi, jeśli interesuje nas tylko wnętrze trójkąta. A jeśli chodzi nam o trójkąt wraz z brzegiem, to nieujemnymi.

[9] Ewentualnie: słabo wypukła.

[10] W przypadku funkcji słabo wypukłej: nie niżej.

[11] Ewentualnie słaba wersja tej nierówności, jeśli funkcja jest słabo wypukła.

Nierówność Jensena: Niech f będzie funkcją wypukłą na przedziale I. Wówczas dla każdych[12] $x_1, x_2, x_3, \ldots, x_n \in I$ oraz każdych liczb nieujemnych $a_1, a_2, a_3, \ldots, a_n$ spełniających warunek

$$a_1+a_2+a_3+\ldots+a_n=1$$

zachodzi nierówność

$$a_1\cdot f(x_1)+a_2\cdot f(x_2)+a_3\cdot f(x_3)+\ldots+a_n\cdot f(x_n)\geqslant f(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3+\ldots+a_n\cdot x_n)\,.$$

Nierówność Jensena (wersja ostra): Niech f będzie funkcją ściśle wypukłą na przedziale I. Wówczas dla każdych różnych[13] $x_1, x_2, x_3, \ldots, x_n \in I$ oraz każdych liczb dodatnich $a_1, a_2, a_3, \ldots, a_n$ spełniających warunek

$$a_1+a_2+a_3+\ldots+a_n=1$$

zachodzi nierówność

$$a_1\cdot f(x_1)+a_2\cdot f(x_2)+a_3\cdot f(x_3)+\ldots+a_n\cdot f(x_n)> f(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3+\ldots+a_n\cdot x_n)\,.$$

Nierówność Jensena: Niech f będzie funkcją wklęsłą na przedziale I. Wówczas dla każdych $x_1, x_2, x_3, \ldots, x_n \in I$ oraz każdych liczb nieujemnych $a_1, a_2, a_3, \ldots, a_n$ spełniających warunek

$$a_1+a_2+a_3+\ldots+a_n=1$$

zachodzi nierówność

$$a_1\cdot f(x_1)+a_2\cdot f(x_2)+a_3\cdot f(x_3)+\ldots+a_n\cdot f(x_n)\leqslant f(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3+\ldots+a_n\cdot x_n)\,.$$

Nierówność Jensena (wersja ostra): Niech f będzie funkcją ściśle wklęsłą na przedziale I. Wówczas dla każdych różnych $x_1, x_2, x_3, \ldots, x_n \in I$ oraz każdych liczb dodatnich $a_1, a_2, a_3, \ldots, a_n$ spełniających warunek

$$a_1+a_2+a_3+\ldots+a_n=1$$

zachodzi nierówność

$$a_1\cdot f(x_1)+a_2\cdot f(x_2)+a_3\cdot f(x_3)+\ldots+a_n\cdot f(x_n)< f(a_1\cdot x_1+a_2\cdot x_2+a_3\cdot x_3+\ldots+a_n\cdot x_n)\,.$$

Wykres funkcji wypukłej leży nad styczną: Niech f będzie funkcją ściśle wypukłą na przedziale I, różniczkowalną w punkcie $x_0 \in I$.
Wówczas dla każdego $x_1 \in I\setminus\{x_0\}$ zachodzi nierówność (rys. 3)

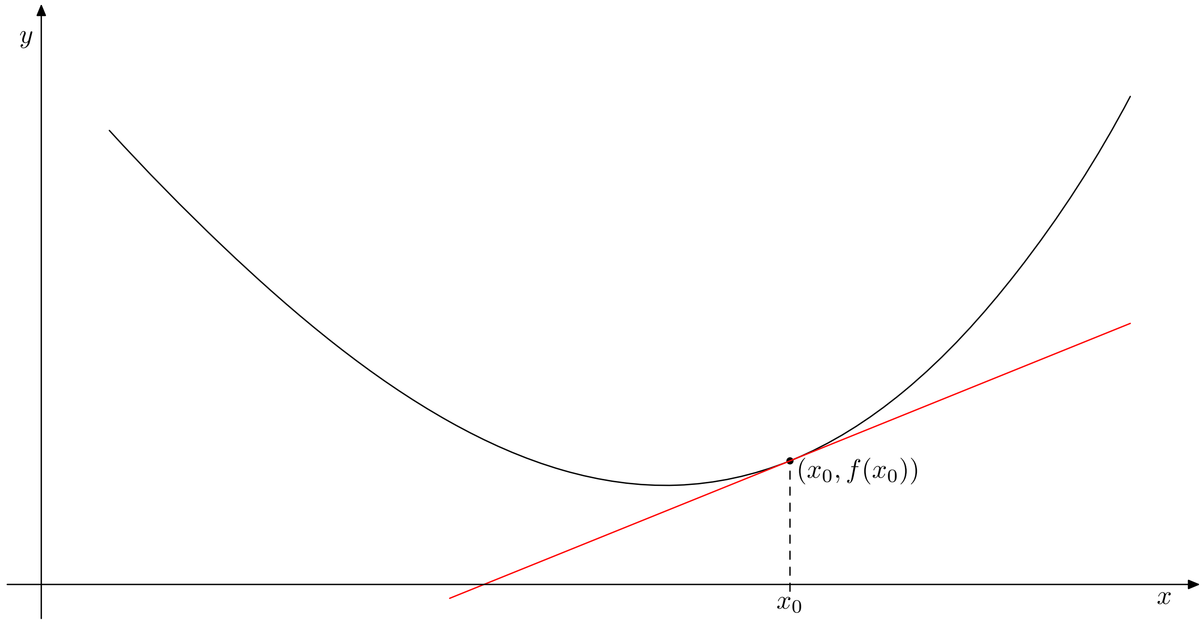
$$f(x_1)>f(x_0)+(x_1-x_0)\cdot f'(x_0)\,.$$

Wersja słaba: Niech f będzie funkcją wypukłą na przedziale I, różniczkowalną w punkcie $x_0 \in I$. Wówczas dla każdego $x_1 \in I$ zachodzi nierówność

$$f(x_1)\geqslant f(x_0)+(x_1-x_0)\cdot f'(x_0)\,.$$

---

[12] Formalista dodałby najpierw: "dla każdej liczby naturalnej n".
[13] Nie muszą być parami różne, wystarczy, że nie są wszystkie jednakowe.



rys. 3

WYKRES FUNKCJI WKLĘSŁEJ LEŻY POD STYCZNĄ: Niech f będzie funkcją ściśle wklęsłą na przedziale I, różniczkowalną w punkcie $x_0 \in I$.
Wówczas dla każdego $x_1 \in I \setminus \{x_0\}$ zachodzi nierówność

$$f(x_1) < f(x_0) + (x_1 - x_0) \cdot f'(x_0)\,.$$

WERSJA SŁABA: Niech f będzie funkcją wklęsłą na przedziale I, różniczkowalną w punkcie $x_0 \in I$. Wówczas dla każdego $x_1 \in I$ zachodzi nierówność

$$f(x_1) \leqslant f(x_0) + (x_1 - x_0) \cdot f'(x_0)\,.$$

Funkcja wypukła nie musi być różniczkowalna. Na przykład funkcja $f : \mathbb{R} \to \mathbb{R}$ określona wzorem $f(x) = |x|$ jest wypukła, ale nie jest różniczkowalna w zerze.

Jeśli jednak funkcja jest różniczkowalna lub dwukrotnie różniczkowalna, to wypukłość/wklęsłość daje się scharakteryzować w języku pochodnych:

Funkcja f różniczkowalna na przedziale jest wypukła wtedy i tylko wtedy, gdy jej pochodna f' jest niemalejąca.

Funkcja f różniczkowalna na przedziale jest wklęsła wtedy i tylko wtedy, gdy jej pochodna f' jest nierosnąca.

Funkcja f różniczkowalna na przedziale jest ściśle wypukła wtedy i tylko wtedy, gdy jej pochodna f' jest rosnąca.

Funkcja f różniczkowalna na przedziale jest ściśle wklęsła wtedy i tylko wtedy, gdy jej pochodna f' jest malejąca.

Funkcja f dwukrotnie różniczkowalna na przedziale jest wypukła wtedy i tylko wtedy, gdy jej druga pochodna f'' jest nieujemna.

Funkcja f dwukrotnie różniczkowalna na przedziale jest wklęsła wtedy i tylko wtedy, gdy jej druga pochodna f'' jest niedodatnia.

Funkcja f dwukrotnie różniczkowalna na przedziale jest ściśle wypukła wtedy i tylko wtedy, gdy jej druga pochodna f'' jest nieujemna oraz nie zeruje się na żadnym przedziale.

Funkcja f dwukrotnie różniczkowalna na przedziale jest ściśle wklęsła wtedy i tylko wtedy, gdy jej druga pochodna f'' jest niedodatnia oraz nie zeruje się na żadnym przedziale.

* * * * * * * * * * * * * * * * * * * *

Jako przykładowy dowód wykażę ścisłą wypukłość funkcji o dodatniej drugiej pochodnej na podstawie wzoru Taylora. Nie jest to sytuacja ograniczająca, gdyż w praktyce mamy do czynienia z funkcjami dwukrotnie różniczkowalnymi, które przedziałami są ściśle wypukłe/wklęsłe, ponieważ ich druga pochodna jest przedziałami dodatnia/ujemna.

Załóżmy więc, że funkcja f ma dodatnią drugą pochodną w pewnym przedziale i niech x_1 oraz x_2 będą różnymi punktami tego przedziału. Niech a_1 i a_2 będą dodatnimi liczbami rzeczywistymi o sumie 1. Przyjmijmy

$$x_0 = a_1 \cdot x_1 + a_2 \cdot x_2 .$$

Wówczas na podstawie wzoru Taylora otrzymujemy

$$f(x_1) = f(x_0) + (x_1 - x_0) \cdot f'(x_0) + \frac{(x_1 - x_0)^2 \cdot f''(x_0 + t_{x_1}(x_1 - x_0))}{2}$$

oraz

$$f(x_2) = f(x_0) + (x_2 - x_0) \cdot f'(x_0) + \frac{(x_2 - x_0)^2 \cdot f''(x_0 + t_{x_2}(x_2 - x_0))}{2} ,$$

co po przemnożeniu odpowiednio przez a_1 i a_2, a następnie dodaniu stronami daje:

$$a_1 \cdot f(x_1) + a_2 \cdot f(x_2) = a_1 \cdot f(x_0) + a_2 \cdot f(x_0) + a_1 \cdot (x_1 - x_0) \cdot f'(x_0) + a_2 \cdot (x_2 - x_0) \cdot f'(x_0) +$$
$$+ a_1 \cdot \frac{(x_1 - x_0)^2 \cdot f''(x_0 + t_{x_1}(x_1 - x_0))}{2} + a_2 \cdot \frac{(x_2 - x_0)^2 \cdot f''(x_0 + t_{x_2}(x_2 - x_0))}{2} .$$

Po uwzględnieniu równości $a_1 + a_2 = 1$ oraz[14]

$$a_1 \cdot (x_1 - x_0) + a_2 \cdot (x_2 - x_0) = 0$$

otrzymujemy

$$a_1 \cdot f(x_1) + a_2 \cdot f(x_2) = f(x_0) +$$
$$+ a_1 \cdot \frac{(x_1 - x_0)^2 \cdot f''(x_0 + t_{x_1}(x_1 - x_0))}{2} + a_2 \cdot \frac{(x_2 - x_0)^2 \cdot f''(x_0 + t_{x_2}(x_2 - x_0))}{2} ,$$

[14] To wynika z $x_0 = a_1 \cdot x_1 + a_2 \cdot x_2$.

skąd wobec dodatniości dwóch ostatnich składników otrzymujemy nierówność występującą w warunku ścisłej wypukłości:

$$a_1\cdot f(x_1)+a_2\cdot f(x_2)>f(x_0)=f\left(a_1\cdot x_1+a_2\cdot x_2\right).$$

Jeszcze prościej dowodzimy, że wykres funkcji ściśle wypukłej leży nad styczną do wykresu. Wzór Taylora mówi bowiem

$$f(x_1)=f(x_0)+(x_1-x_0)\cdot f'(x_0)+\frac{(x_1-x_0)^2\cdot f''(x_0+t_{x_1}(x_1-x_0))}{2},$$

skąd wobec dodatniości ostatniego składnika

$$f(x_1)>f(x_0)+(x_1-x_0)\cdot f'(x_0)$$

dla $x_1\neq x_0$.

* * * * * * * * * * * * *
* * * * *

Przykład 1: Niech $f(x)=\ln x$ dla $x>0$. Wówczas

$$f''(x)=-\frac{1}{x^2}<0,$$

więc funkcja f jest wklęsła. Zatem nierówność Jensena zastosowana do liczb dodatnich x_1, x_2, x_3, $\ldots$, x_n oraz współczynników $a_1=a_2=a_3=\ldots=a_n=1/n$ daje

$$f\left(\frac{x_1+x_2+x_3+\ldots+x_n}{n}\right)\geqslant\frac{f(x_1)+f(x_2)+f(x_3)+\ldots+f(x_n)}{n},$$

czyli

$$\ln\frac{x_1+x_2+x_3+\ldots+x_n}{n}\geqslant\frac{\ln x_1+\ln x_2+\ln x_3+\ldots+\ln x_n}{n}.$$

Po przekształceniu otrzymujemy kolejno nierówności równoważne:

$$\ln\frac{x_1+x_2+x_3+\ldots+x_n}{n}\geqslant\frac{\ln\left(x_1\cdot x_2\cdot x_3\cdot\ldots\cdot x_n\right)}{n},$$

$$\ln\frac{x_1+x_2+x_3+\ldots+x_n}{n}\geqslant\ln\sqrt[n]{x_1\cdot x_2\cdot x_3\cdot\ldots\cdot x_n},$$

$$\frac{x_1+x_2+x_3+\ldots+x_n}{n}\geqslant\sqrt[n]{x_1\cdot x_2\cdot x_3\cdot\ldots\cdot x_n},$$

a to jest znana nam nierówność między średnią arytmetyczną i geometryczną.

Przykład 2: Niech $f(x)=1/x$ dla $x>0$. Wówczas

$$f''(x)=\frac{2}{x^3}>0,$$

więc funkcja f jest wypukła. Zatem nierówność Jensena zastosowana do liczb dodatnich x_1, x_2, x_3, $\ldots$, x_n oraz współczynników $a_1=a_2=a_3=\ldots=a_n=1/n$ daje

$$f\left(\frac{x_1+x_2+x_3+\ldots+x_n}{n}\right)\leqslant\frac{f(x_1)+f(x_2)+f(x_3)+\ldots+f(x_n)}{n},$$

czyli

$$\frac{1}{\frac{x_1+x_2+x_3+\ldots+x_n}{n}} \leqslant \frac{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_n}}{n}\,.$$

Po odwróceniu stronami otrzymujemy nierówność równoważną:

$$\frac{x_1+x_2+x_3+\ldots+x_n}{n} \geqslant \frac{1}{\frac{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_n}}{n}}\,,$$

która jest nierównością między średnią arytmetyczną i harmoniczną.

Przykład 3: Niech $f(x)=x^2$ dla $x\geqslant 0$. Wówczas

$$f''(x)=2>0\,,$$

więc funkcja f jest wypukła. Zatem nierówność Jensena zastosowana do liczb nieujemnych x_1, x_2, x_3, ..., x_n oraz współczynników $a_1=a_2=a_3=\ldots=a_n=1/n$ daje

$$f\left(\frac{x_1+x_2+x_3+\ldots+x_n}{n}\right) \leqslant \frac{f(x_1)+f(x_2)+f(x_3)+\ldots+f(x_n)}{n}\,,$$

czyli

$$\left(\frac{x_1+x_2+x_3+\ldots+x_n}{n}\right)^2 \leqslant \frac{x_1^2+x_2^2+x_3^2+\ldots+x_n^2}{n}\,.$$

Po spierwiastkowaniu stronami otrzymujemy nierówność równoważną:

$$\frac{x_1+x_2+x_3+\ldots+x_n}{n} \leqslant \sqrt{\frac{x_1^2+x_2^2+x_3^2+\ldots+x_n^2}{n}}\,,$$

która jest nierównością między średnią arytmetyczną i kwadratową[15].

Punkty przegięcia.

Punktem przegięcia wykresu[16] funkcji nazywamy punkt, w którym wykres zmienia swoją ścisłą wypukłość. Oznacza to, że po jednej stronie tego punktu funkcja jest ściśle wypukła (choćby w małym otoczeniu), a po drugiej ściśle wklęsła.

Jeżeli funkcja jest dwukrotnie rózniczkowalna, to punkt przegięcia funkcji jest punktem, w którym jej druga pochodna zmienia znak.

Jeżeli funkcja jest różniczkowalna, to punkt przegięcia funkcji jest punktem, w którym jej pochodna zmienia monotoniczność, czyli jest rosnąca po jednej stronie, a malejąca po drugiej.

Styczna do wykresu w punkcie przegięcia przecina ten wykres.

Modelowy przykład: Funkcja f określona wzorem $f(x)=x^3$ ma punkt przegięcia w punkcie 0 (czyli punkt $(0,\,0)$ jest punktem przegięcia krzywej będącej wykresem funkcji).

[15] Ta nierówność zachodzi dla dowolnych liczb rzeczywistych x_1, x_2, x_3, ..., x_n, niekoniecznie nieujemnych, mamy bowiem

$$\frac{x_1+x_2+x_3+\ldots+x_n}{n} \leqslant \left|\frac{x_1+x_2+x_3+\ldots+x_n}{n}\right| \leqslant \sqrt{\frac{x_1^2+x_2^2+x_3^2+\ldots+x_n^2}{n}}\,,$$

ale dla ujemnych liczb nie można nazywać prawej strony nierówności średnią kwadratową.

[16] Często dla uproszczenia mówi się o punkcie przegięcia funkcji.