Metody Numeryczne

Wykład 7
Optymalizacja

Wojciech Kordecki

Collegium Witelona
Wydział Nauk Technicznych i Ekonomicznych
Zakład Informatyki, Energetyki i Technologii Cyfrowych

Semestr letni 2025/26

Minima i maksima

Niech f będzie funkcją n zmiennych, to znaczy

$$f(x) = f(x_1, x_2, \ldots, x_n),$$

czyli $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ jest wektorem.

Minima i maksima

Niech f będzie funkcją n zmiennych, to znaczy

$$f(x) = f(x_1, x_2, \ldots, x_n),$$

czyli $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ jest wektorem.

Minimum globalne w punkcie p: $f(p) \leq f(x)$ dla każdego x.

Minima i maksima

Niech f będzie funkcją n zmiennych, to znaczy

$$f(x) = f(x_1, x_2, \ldots, x_n),$$

czyli $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ jest wektorem.

Minimum globalne w punkcie p: $f(p) \leq f(x)$ dla każdego x.

Minimum lokalne w punkcie p: $f(p) \leq f(x)$ dla x należących do pewnego otwartego otoczenia punktu p.

Otoczenie otwarte punktu p

Założenie: $r > 0$ i oznaczenie $p = (p_1, p_2, \ldots, p_n)$.

Otoczenie otwarte punktu p

Założenie: $r > 0$ i oznaczenie $p = (p_1, p_2, \ldots, p_n)$.

Kule otwarte o środku p.

Otoczenie otwarte punktu p

Założenie: $r > 0$ i oznaczenie $p = (p_1, p_2, \ldots, p_n)$.

Kule otwarte o środku p.

Prosta $\mathbb{R}$: odcinek $(p - r, p + r)$.

Otoczenie otwarte punktu p

Założenie: $r > 0$ i oznaczenie $p = (p_1, p_2, \ldots, p_n)$.

Kule otwarte o środku p.

Prosta $\mathbb{R}$: odcinek $(p - r, p + r)$.

Płaszczyzna $\mathbb{R}^2$: koło

$$\left\{(x, y) : (x - p_x)^2 + (y - p_y)^2 < r^2\right\}.$$

Otoczenie otwarte punktu p

Założenie: $r > 0$ i oznaczenie $p = (p_1, p_2, \ldots, p_n)$.

Kule otwarte o środku p.

Prosta $\mathbb{R}$: odcinek $(p - r, p + r)$.

Płaszczyzna $\mathbb{R}^2$: koło

$$\left\{(x, y) : (x - p_x)^2 + (y - p_y)^2 < r^2\right\}.$$

Przestrzeń n-wymiarowa $\mathbb{R}^n$: kula n-wymiarowa

$$\left\{(x_1, x_2, \ldots, x_n) : (x_1 - p_1)^2 + (x_2 - p_2)\,2 + \cdots + (x_n - p_n)^2 < r^2\right\}.$$

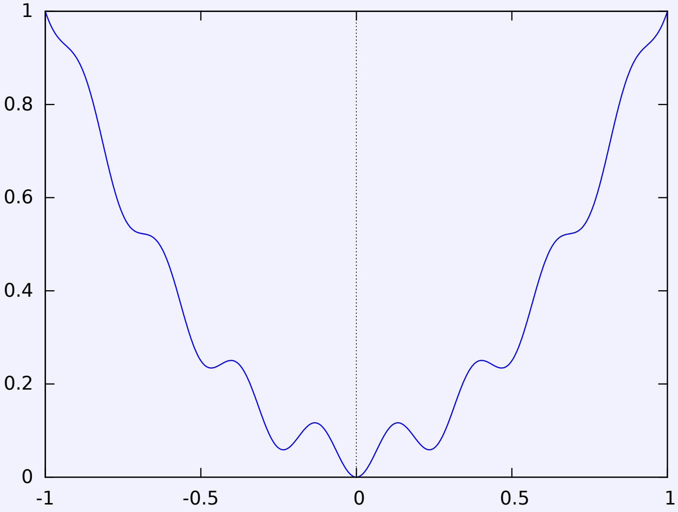
Przykład – funkcja jednej zmiennej

Funkcja

$$f(x) = x^2 - 0.1\left(\cos^2(4\pi x) - 1\right)$$

określona na przedziale $[-1, 1]$ ma jedno minimum globalne w punkcie $x = 0$ i ponadto jeszcze cztery minima lokalne.

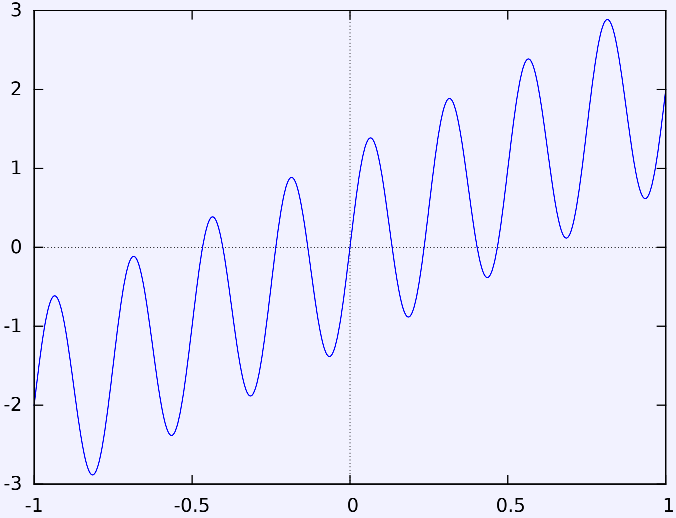
Przykład – wykres



Minima lokalne

Cztery minima tylko lokalne, co łatwo stwierdzić, patrząc na wykres pochodnej $f'(r)$.
Minima lokalne są w tych punktach r, w których $f'(r)$ zmienia znak z ujemnego na dodatni wraz ze wzrostem r.

Przykład – wykres pochodnej



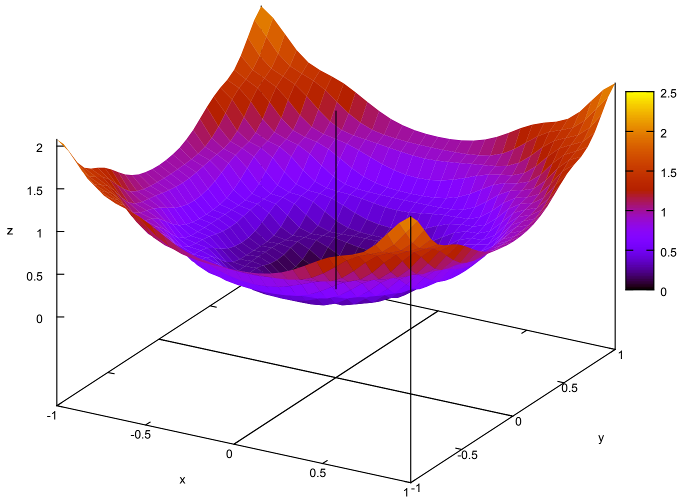
Przykład – funkcja dwóch zmiennych

Funkcja

$$f(x,y) = x^2 + y^2 - 0.1\left(\cos^2\left(4\pi\sqrt{x^2+y^2}\right) - 1\right)$$

określona na kwadracie $[-1,1]\times[-1,1]$ ma jedno minimum globalne w punkcie $p = (0,0)$ i ponadto jeszcze minima lokalne.

Przykład – wykres



Szczególny przypadek – funkcje jednej zmiennej

Algorytmy optymalizacji dla funkcji wielu zmiennych wymagają często jej badania wzdłuż pewnej prostej.

Sprowadza się to szukania minimów funkcji jedne zmiennej.

Szczególny przypadek – funkcje jednej zmiennej

Algorytmy optymalizacji dla funkcji wielu zmiennych wymagają często jej badania wzdłuż pewnej prostej.

Sprowadza się to szukania minimów funkcji jedne zmiennej.

Przykład. Ponieważ funkcja

$$f(x, y) = x^2 + y^2 - 0.1\left(\cos^2\left(4\pi\sqrt{x^2 + y^2}\right) - 1\right)$$

ma stałą wartość na okręgu $x^2 + y^2 = r^2$, więc szukanie jej minimów sprowadza się do szukania minimów funkcji

$$f(r) = r^2 - 0.1\left(\cos^2(4\pi r) - 1\right)$$

Pochodne i minima

Szukamy minimum funkcji $f(x)$.

Pochodne i minima

Szukamy minimum funkcji $f(x)$.

Dwa typy algorytmów:

- używamy pochodnej funkcji f (o ile istnieje),
- nie używamy pochodnej (nawet jeśli istnieje).

Jeśli pochodna nie jest znana, to można stosować procedurę przeszukiwania, obliczając wartości $f(kh)$ dla $k = 0, \pm 1, \pm 2, \ldots,$ co może dać informację o minimum.

Pochodne i minima

Szukamy minimum funkcji $f(x)$.

Dwa typy algorytmów:

- używamy pochodnej funkcji f (o ile istnieje),
- nie używamy pochodnej (nawet jeśli istnieje).

Jeśli pochodna nie jest znana, to można stosować procedurę przeszukiwania, obliczając wartości $f(kh)$ dla $k = 0, \pm 1, \pm 2, \ldots,$ co może dać informację o minimum.

Uwaga. Dla każdego h można znaleźć taką funkcję, dla której ta procedura zawodzi.

Przykład

$$f(x) = x/2 + 0.1 \sin(10x + 0.2\pi)$$

Szukamy maksimum w przedziale $[-1, 1]$ co $h = 0.1$. Otrzymujemy w punktach $-0.1k$ dla $k = -10, -9, \ldots, 9, 10$ wartości

-0.5053071546401395	-0.5368959789109834	-0.4885930409124076
-0.3588380696667052	-0.2209574179282806	-0.1557481516230514
-0.1771936000509204	-0.219607147414871	-0.1980242044553964
-0.08631825998141032	0.05877852522924731	0.1498346054151921
0.149103209793281	0.1032265495384814	0.1003531839311835
0.1890946414727188	0.3338322047647849	0.4474645950213138
0.4714884860962557	0.429786192740345	0.4066683805582128

Maksima są zaś osiągane w punktach

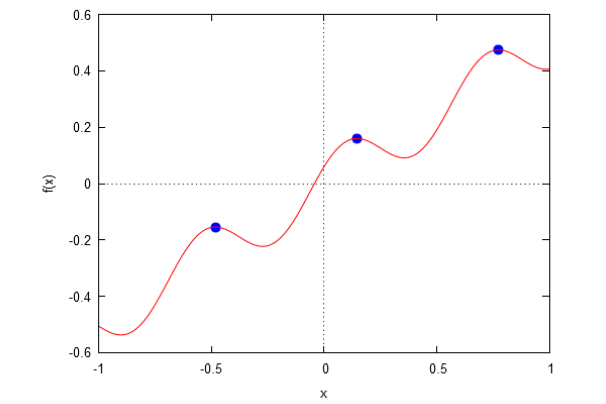
-0.481710873550435, 0.1466076571675237, 0.7749261878854823

i wynoszą odpowiednio

-0.1542528963967736, 0.1599063689622057, 0.4740656343211851.

Przykład – wykres

x	-0.481710873550435	$0.146607657167523$7	$0.774926187885482$3
$h(x)$	-0.1542528963967736	0.1599063689622057	$0.474065634321185$1

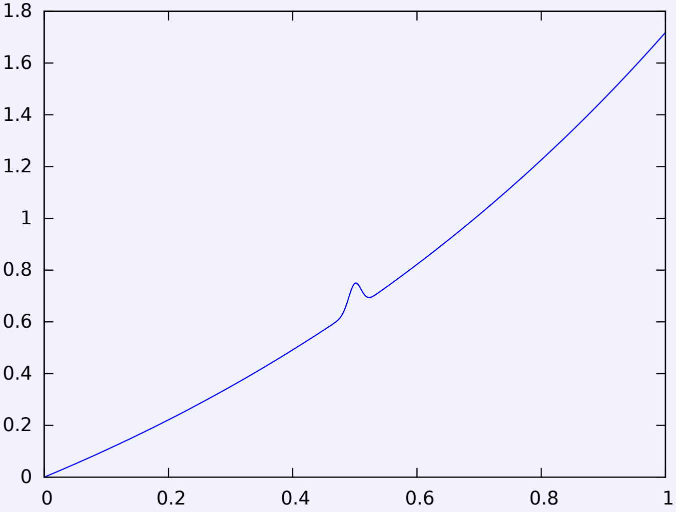


Przykład

$$f(x) = e^x + 0.1e^{-5000(x-0.5)^2} - 1.$$

Jak widać z wykresu funkcji tej funkcji, jej przeszukiwanie dla $h = 0.1$, począwszy od $x = 0$, nie znajdzie maksimum, które znajduje się w punkcie $x_0 = 0.5017$ i ma wartość $f(x_0) = 0.7501$. Wartości $f(x)$ w punktach $0.4, 0.5, 0.6$ wynoszą odpowiednio $0.4918, 0.7487, 0.8221$.
Inaczej mówiąc, przeszukiwanie co $h = 0.1$ „przeskoczy" punkt, w którym funkcja ma maksimum.

Przykład – wykres



Funkcje unimodalne

Funkcja $f(x)$ jest *unimodalna*, gdy ma tylko jedno minimum (maksimum) lokalne.

Przykład.

$$f(x) = 2 - \frac{1}{1 + x^2}$$

jest fukcją unimodalną.

Funkcje unimodalne

Funkcja $f(x)$ jest *unimodalna*, gdy ma tylko jedno minimum (maksimum) lokalne.

Przykład.

$$f(x) = 2 - \frac{1}{1+x^2}$$

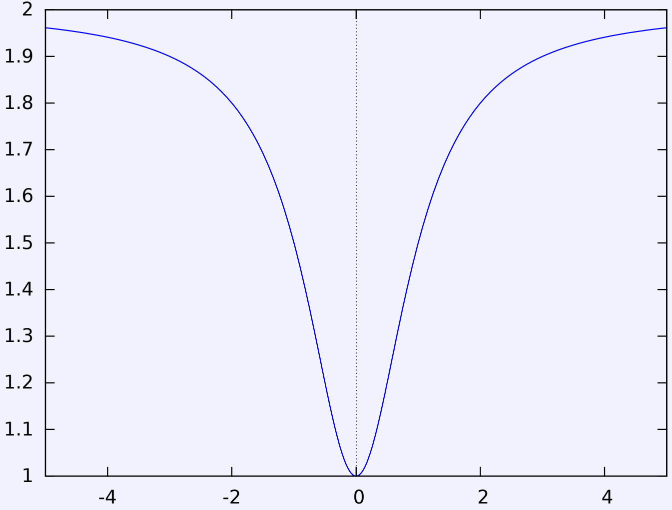
jest fukcją unimodalną.

Funkcje

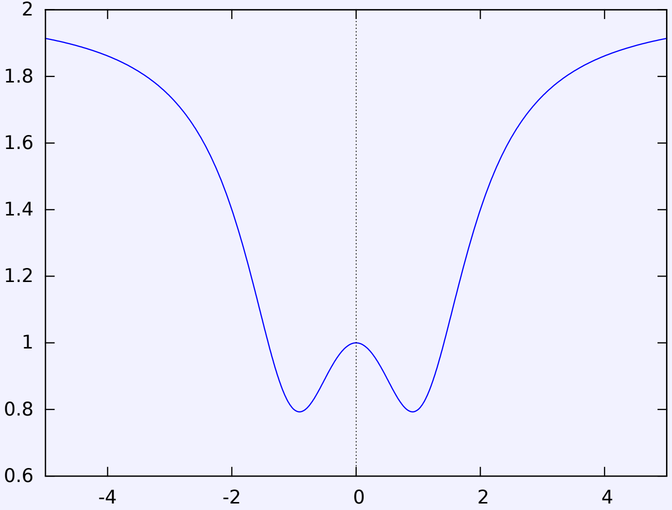
$$f_1(x) = 2 - \frac{1}{1+(x-1)^2} - \frac{1}{1+(x+1)^2},$$
$$f_2(x) = 2\frac{1}{1+(x-1)^2} - \frac{0.7}{1+(x+1)^2}$$

nie są funkcjami unimodalnymi.

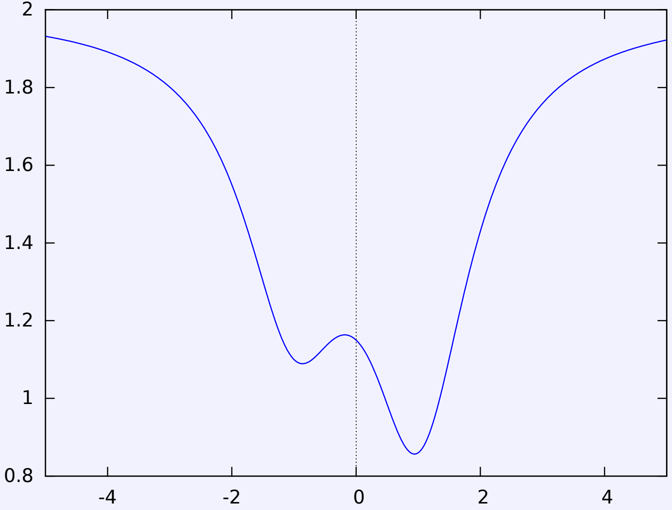
Funkcja $f(x)$ – wykres



Funkcja $f_1(x)$ – wykres



Funkcja $f_2(x)$ – wykres



Metoda złotego podziału odcinka

Taką metodę można stosować w przypadku funkcji unimodalnej.

$$r = \frac{1}{2}\left(\sqrt{5} - 1\right) = 0.6180339887498949\ldots,$$

jest pierwiastkiem równania

$$r^2 + r - 1 = 0.$$

Szukamy minimum funkcji $f(x)$ w przedziale $[a, b]$.

$$\begin{aligned} x_R &= a + r(b - a), \\ x_L &= b - r(b - a), \end{aligned}$$

$$\begin{aligned} u &= f(x_R), \\ v &= f(x_L). \end{aligned}$$

Metoda złotego podziału – algorytm

Jeśli $u > v$, to minimum jest w $[a, x_R]$.

Metoda złotego podziału – algorytm

Jeśli $u > v$, to minimum jest w $[a, x_R]$.

$b \leftarrow x_R$, $x_R \leftarrow x_L$,
$x_L \leftarrow b - r\,(b - a)$.

Metoda złotego podziału – algorytm

Jeśli $u > v$, to minimum jest w $[a, x_R]$.

$b \leftarrow x_R$, $x_R \leftarrow x_L$,
$x_L \leftarrow b - r\,(b - a)$.

Jeśli $u \leq v$, to minimum jest w $[x_L, b]$.

Metoda złotego podziału – algorytm

Jeśli $u > v$, to minimum jest w $[a, x_R]$.

$b \leftarrow x_R$, $x_R \leftarrow x_L$,
$x_L \leftarrow b - r\,(b - a)$.

Jeśli $u \leq v$, to minimum jest w $[x_L, b]$.

$a \leftarrow x_L$, $x_L \leftarrow x_R$,
$x_R \leftarrow a + r\,(b - a)$.

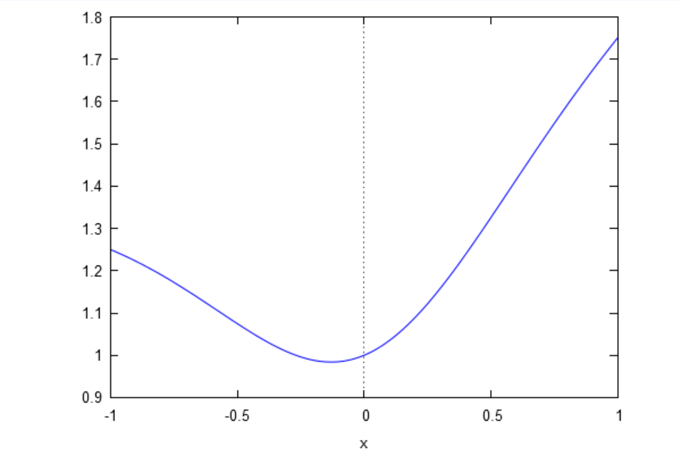
Ilustracja algorytmu (1)

Ilustracja algorytmu (2)

Funkcja ta jest unimodalna i ma minimum w punkcie x_0.
Ponieważ $v = f(x_L) > f(x_R) = u$, to minimum znajduje się w przedziale $[x_L, b]$.
Podstawiamy $a \leftarrow x_L$, $x_L \leftarrow x_R$, wyznaczamy nowe x_R i ponawiamy szukanie minimum według tej samej procedury.
Powtarzamy procedurę, aż $\delta = |a - b|$ będzie mniejsza niż założona dokładność.

Przykład (1)

$$f(x) = 2 + \frac{x}{4} - \frac{1}{1+x^2}.$$



Przykład (2)

Przedział $[-1, 1]$, czyli $a = -1$, $b = 1$.
Stąd $x = 0.2360679774997898$, $y = -0.2360679774997896$,
$u = 1.111803398874\,99$, $v = 0.9937694101250946$, $u > v$.

Przykład (2)

Przedział $[-1, 1]$, czyli $a = -1$, $b = 1$.
Stąd $x = 0.2360679774997898$, $y = -0.2360679774997896$,
$u = 1.111803398874990$, $v = 0.9937694101250946$, $u > v$.

Minimum jest w przedziale $[-1, 0.2360679774997898]$.

Przykład (2)

Przedział $[-1, 1]$, czyli $a = -1$, $b = 1$.
Stąd $x = 0.2360679774997898$, $y = -0.2360679774997896$,
$u = 1.111803398874990$, $v = 0.9937694101250946$, $u > v$.

Minimum jest w przedziale $[-1, 0.2360679774997898]$.

Powtarzamy procedurę szukania minimum w przedziale $[-1, 0.2360679774997898]$.
Teraz $u = -0.5278640450004205 < v = 0.2360679774997898$, więc minimum jest w przedziale $[-0.5278640450004205, 0.2360679774997898]$.

Przykład (3)

Pochodna

$$f'(x) = \frac{2x}{x^2+1}^{2} + 0.25 = \frac{x^4 + 2x^2 + 8x + 1}{4x^4 + 8x^2 + 4}$$

zeruje się w punkcie $x_0 = -0.1292085522452846$.

Przykład (3)

Pochodna

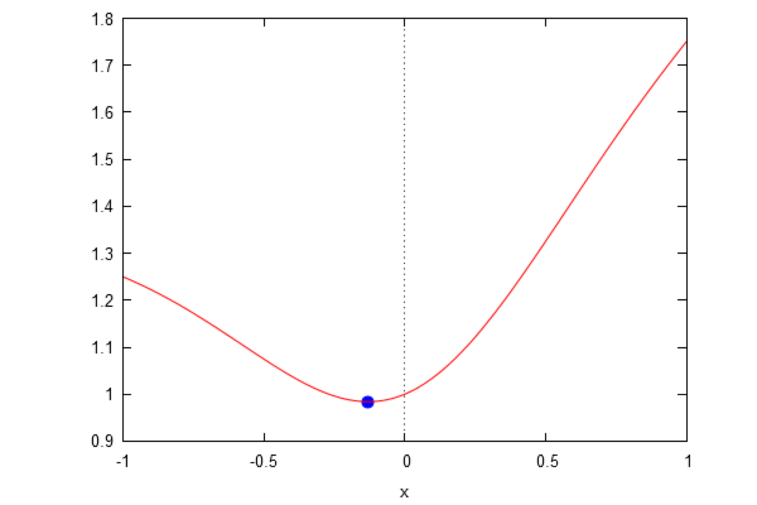
$$f'(x) = {\frac{2x}{x^2+1}}^2 + 0.25 = \frac{x^4 + 2x^2 + 8x + 1}{4x^4 + 8x^2 + 4}$$

zeruje się w punkcie $x_0 = -0.1292085522452846.$

W tym punkcie funkcja $f(x)$ ma minimum.

$$f(x_0) = 0.9841185706437151.$$

Przykład (4)



Przykład (5) – kolejne przybliżenia

k	a	b
0	-1	1
1	-1	0.2361
2	-0.5279	0.2361
3	-0.2361	0.2361
...	...	...
19	-0.1294	-0.1290
20	-0.1294	-0.1291

Gradient i hesjan

Niech $x \in \mathbb{R}^n$.

Gradient i hesjan

Niech $x \in \mathbb{R}^n$.

Gradient $G = \nabla f$ – wektor o składowych

$$\frac{\partial f(x)}{\partial x_i}, \qquad 1 \leq i \leq n.$$

Hesjan – macierz H o elementach

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \qquad 1 \leq i, j \leq n.$$

Wzór Taylora

Początkowe wyrazy wzoru Taylora dla funkcji f:

$$f(x+h) = f(x) + G(x)^T h + \frac{1}{2} h^T H(x) h + \dots$$

Wzór Taylora

Początkowe wyrazy wzoru Taylora dla funkcji f:

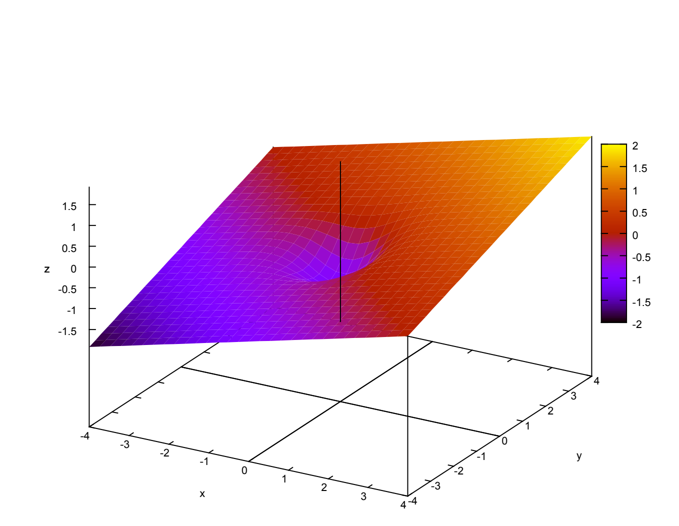
$$f(x+h) = f(x) + G(x)^T h + \frac{1}{2} h^T H(x) h + \ldots$$

Gradient pokazuje kierunek najszybszego wzrostu funkcji f, a w przeciwnym kierunku – najszybszego spadku.
Przybliżenie:

$$f(\mathrm{x}) \approx f(\mathrm{x}_0) + G(\mathrm{x}_0)^T (\mathrm{x}-\mathrm{x}_0) + \frac{1}{2} (\mathrm{x}-\mathrm{x}_0)^T H(\mathrm{x}_0)(\mathrm{x}-\mathrm{x}_0). \quad (1)$$

Przykład funkcji

$$f(x,y) = -e^{-\left(x^2+y^2\right)} + 0.24\,(x+y),$$



Obliczenia (1)

$$
\begin{aligned}
\frac{\partial f}{\partial x} &= 2xe^{-y^2-x^2} + 0.24,\\
\frac{\partial f}{\partial y} &= 2ye^{-y^2-x^2} + 0.24,\\
\frac{\partial^2 f}{\partial x^2} &= -\left(4x^2 - 2\right) e^{-y^2-x^2},\\
\frac{\partial^2 f}{\partial y^2} &= -\left(4y^2 - 2\right) e^{-y^2-x^2},\\
\frac{\partial^2 f}{\partial x \partial y} &= -4xye^{-y^2-x^2}.
\end{aligned}
$$

Obliczenia (2)

Stąd gradient

$$G(x,y) = \begin{bmatrix} 2xe^{-y^2-x^2} + 0.24 \\ 2ye^{-y^2-x^2} + 0.24 \end{bmatrix}$$

i hesjan

$$H(x,y) = \begin{bmatrix} -\left(4x^2-2\right)e^{-y^2-x^2} & -4xye^{-y^2-x^2} \\ -4xye^{-y^2-x^2} & -\left(4y^2-2\right)e^{-y^2-x^2} \end{bmatrix}.$$

W punkcie $(0,0)$ mamy

$$G(0,0) = \begin{bmatrix} 0.24 \\ 0.24 \end{bmatrix}$$

oraz

$$H(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Obliczenia (3) – przybliżenie

$$f(x,y) \approx t(x,y) = -1 + 0.24\,(x+y) + \left(x^2+y^2\right)$$

dla (x, y) w pobliżu punktu $(0, 0)$.

Obliczenia (3) – przybliżenie

$$f(x,y) \approx t(x,y) = -1 + 0.24(x+y) + \left(x^2 + y^2\right)$$

dla (x, y) w pobliżu punktu $(0, 0)$.

Istotnie $t(x,y)$ jest tylko przybliżeniem $f(x,y)$. Na przykład $f(0.1, 0.1) = -0.9322$, natomiast $t(0.1, 0.1) = -1.068$.

Szukamy minimum

$$f(x, y) = -e^{-\left(x^2+y^2\right)} + 0.24\,(x+y)\,,$$

Szukamy minimum

$$f(x,y) = -e^{-\left(x^2+y^2\right)} + 0.24\,(x+y)\,,$$

$f(x,y) = f(y,x)$ dla dowolnej pary (x,y).
Jeśli $f(x,y)$ ma minimum, to będzie ono leżeć na prostej o równaniu $y = x$. Wtedy wystarczy znaleźć minimum funkcji

$$g(x) = f(x,x) = -e^{-2x^2} + 0.48x.$$

Dowolną metodą znajdujemy punkt, w którym $g(x)$ ma minimum.
Jest nim $x = -0.1237$: $f(x,y)$ ma minimum w punkcie $(-0.1237, -0.1237)$, a jego wartość wynosi -1.0292.
Można też obliczyć pochodną funkcji $g(x)$:

$$g'(x) = 4xe^{-2x^2} + 0.48.$$

Metoda gradientu

Jest to metoda największego spadku.

1. Startujemy z dowolnego punktu x^0 i idziemy w kierunku $-\nabla f\left(\mathrm{x}^0\right)$.
2. Na półprostej $\left\{\mathrm{x}^0 - t\nabla f\left(\mathrm{x}^0\right) : t \geq 0\right\}$ znajdujemy punkt x^1, w którym $f\left(\mathrm{x}\right)$ ma minimum, a z niego powtarzamy algorytm, otrzymując dalsze punkty x^i.

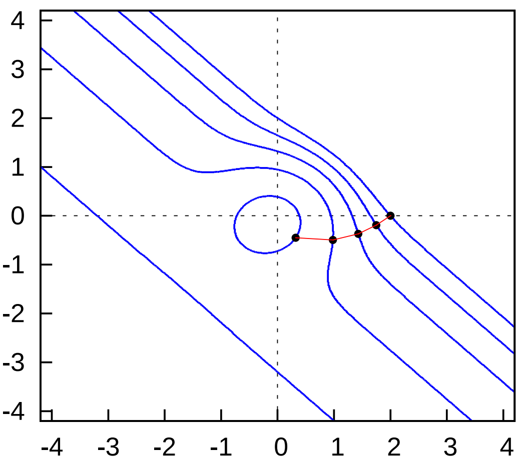
Metoda gradientu prostego

Prostsza wersja tego algorytmu zastępuje szukanie minimum na półprostej z kroku 2 przez wybór następnego punktu $\mathrm{x}^{i+1} = \mathrm{x}^{i} - h\nabla f\left(\mathrm{x}^{i}\right)$, gdzie $h > 0$ jest ustalonym, małym parametrem.

Przykład

$$f(x, y) = -e^{-\left(x^2+y^2\right)} + 0.24\,(x+y)\,,$$

Korzystamy z prostszej wersji algorytmu i przyjmujemy $h = 0.8$.



Przykład c.d.

Jako początkowy punkt wybrano $(2, 0)$. Ponieważ

$$\nabla f\,(2,0) = (0.3133, 0.24)\,,$$

to następnym punktem jest

$$(2,0) - h\nabla f\,(2,0) = (1.7494, -0.1920)\,.$$

W nim ponownie obliczamy gradient i idziemy do następnego punktu.

Kolejne punkty

i	x_i	y_i	$f(x_i, y_i)$
0	2.0000	0.0000	0.4617
1	1.7494	−0.1920	0.3286
2	1.4309	−0.3701	0.1421
3	0.9813	−0.4955	−0.1820
4	0.3204	−0.4507	−0.4507

Wielomian kwadratowy

Zakłada się, że funkcję f można lokalnie przybliżyć wielomianem kwadratowym

$$p(x) = a - b^T x + \frac{1}{2} x^T A x,$$

gdzie

a – skalar,
b – wektor o n składowych,
A – macierz kwadratowa, symetryczna, stopnia n.

Wielomian kwadratowy

Zakłada się, że funkcję f można lokalnie przybliżyć wielomianem kwadratowym

$$p(x) = a - b^T x + \frac{1}{2} x^T A x,$$

gdzie

a – skalar,
b – wektor o n składowych,
A – macierz kwadratowa, symetryczna, stopnia n.

Funkcja f ma minimum, gdy jest nieujemnie określona.

Wielomian kwadratowy c.d.

$$f(x) = a - \sum_{i=1}^{n} b_i x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \, .$$

Wielomian kwadratowy c.d.

$$f(x) = a - \sum_{i=1}^{n} b_i x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \,.$$

Gradient ∇f ma składowe

$$\frac{\partial f}{\partial x_k} = b_k + \sum_{i=1}^{n} A_{ik} x_i \,,$$

Wielomian kwadratowy c.d.

$$f(x) = a - \sum_{i=1}^{n} b_i x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j .$$

Gradient ∇f ma składowe

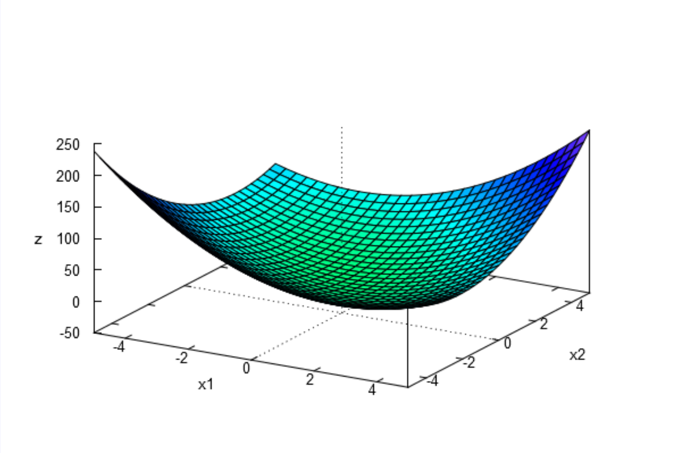
$$\frac{\partial f}{\partial x_k} = b_k + \sum_{i=1}^{n} A_{ik} x_i ,$$

czyli

$$\nabla f(x) = Ax - b.$$

Przykład – jest minimum

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad a = -1, \quad \det A > 0.$$



Przykład – nie ma minimum

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad a = -1, \quad \det A < 0.$$