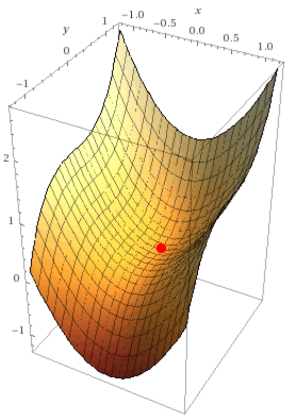
Komentář k úlohám probíraným na cvičeních dle plánu 08

Petr Olšák

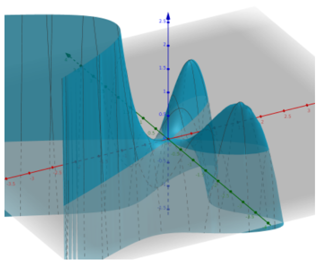
10.1 c) Hessián má kladná vlastní čísla, je tedy pozitivně definitní a funkce má v daném bodě minimum. **a)** Hessián je indefinitní, tj. v daném bodě je sedlo. Ve směrech prvních dvou vlastních vektorů se funkce v daném bodě chová řádově jako x^2 v nule a ve směru posledního vlastního vektoru jako $-x^2$ v nule. **b)** Nelze o extrému nic říci, protože je jedno vlastní číslo nula. Ve směru vlastního vektoru příslušného tomuto nulovému vlastnímu číslu se funkce může chovat jako x^3 v nule, pak tam je „skluzavka", ale ne minimum ani maximum. Nebo se tam může chovat řádově jako x^4, pak v daném bodě je minimum, nebo se tam může chovat jako $-x^4$ a pak je v daném bodě sedlo. To by se dalo vyšetřit tak, že bychom redukovali vyšetřování funkce jen v daném směru, měli bychom tím pádem funkci jedné proměnné, která má první i druhou derivaci v daném bodě nulovou. Při nenulové třetí derivaci máme „skluzavku", při nulové třetí derivaci a nenulové čtvrté máme minimum nebo sedlo. Je-li i tato derivace nulová, pak záleží na tom, který řád derivace je poprvé nenulový (lichý nebo sudý). Jak vypadá skluzavka zjistíte prohlídkou dětských hřišť nebo pohledem na obrázek vlevo. Bod, ve kterém jsou první derivace ve všech směrech nulové (tj. tečny tam jsou vodorovné), je vyznačen červeným puntíkem.



10.2 e) $f' = [6y^2 - 6x^2, 12xy - 12y^3]$. Stacionární body jsou body, kde $f' = [0\ 0]$, takže je třeba řešit soustavu $6y^2 - 6x^2 = 0,\ 12xy - 12y^3 = 0$. Jedno řešení je $(0,0)$. Dále z první rovnice vidíme, že $x = \pm y$. Dosadíme-li $x = y$ do druhé rovnice, máme řešení $(1,1)$. Dosadíme-li $x = -y$, máme řešení $(1,-1)$. Jiná řešení soustava nemá. Máme tedy tři stacionární body. Hessián:

$$f''(x,y) = \begin{bmatrix} -12x & 12y \\ 12y & 12x - 36y^2 \end{bmatrix},$$

$$f''(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad f''(1,1) = 12\begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix}, \quad f''(1,-1) = 12\begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix}.$$



Pro oba případy $f''(1,1)$ i $f''(1,-1)$ je příslušný Hessián negativně definitní (má záporná vlastní čísla, spočtě te si je), takže v bodech $(1,1)$ a $(1,-1)$ má funkce lokální maximum. V bodě (0,0) a ve směru osy x se zadaná funkce $f(x,y) = 6xy^2 - 2x^3 - 3y^4$ chová jako $-2x^3$ (dosaďte $y = 0$) a ve směru osy y se chová jako $-3y^4$ (dosaďte $x = 0$). Takže v tomto bodě to je „skluzavka" (ve směry osy x) navíc orientovaná „vzhůru dnem", protože ve směru osy y se chová jako $-3y^4$ (podstatné je, že ten koeficient u sudé mocniny je záporný).

Poznámka Protože jen v ojedinělých případech se daří najít analyticky řešení příslušné nelineární soustavy rovnic sestavené z požadavku na stacionární body $f' = \mathbf{o}$, máme za úkol se v tomto týdnu zabývat numerickými metodami, které umožní najít řešení takové soustavy rovnic aspoň přibližně a za předpokladu, že metodě nabídneme „dostatečně rozumnou" výchozí aproximaci. Metody pak vezmou tuto hodnotu, udělají jeden *iterační krok* a vytvoří další aproximaci. Tu pak mohou vzít znovu a vytvořit (obvykle stejným) iteračním krokem další aproximaci. Za ideálního stavu věcí se tyto jednotlivé aproximace stále zlepšují a výsledek se (rozumnou rychlostí) zpřesňuje.

V tomto textu značím výchozí aproximaci (i postupnou další) jako $\mathbf{a}$ (ve skriptech značeno obvykle $\mathbf{x}_k$) a nově počítanou aproximaci značím jako $\mathbf{x}$ (ve skriptech značeno $\mathbf{x}_{k+1}$).

Přehled numerických metod zmíněných v tomto kurzu je v následující tabulce

Tabulka shrnující numerické metody tohoto týdne

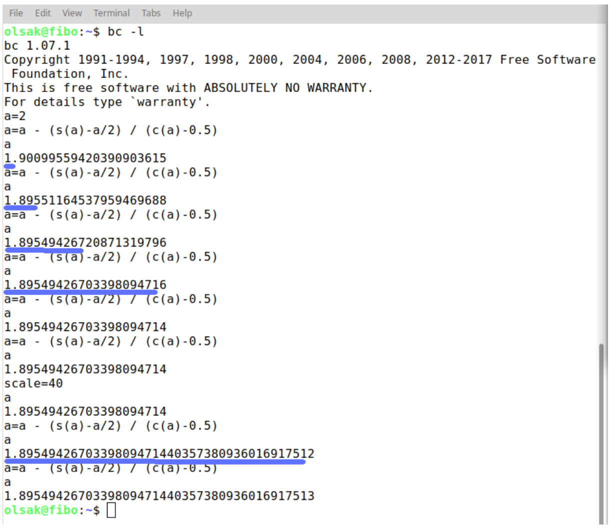
Název metody	co počítá	iterační vzorec	poznámka
Gradientní	$\min f(\mathbf{x})$ pro $f\colon \mathbb{R}^n \to \mathbb{R}$	$\mathbf{x} = \mathbf{a} - f'(\mathbf{a})^T$	směr největšího spádu
Newtonova	$\mathbf{g}(\mathbf{x}) = \mathbf{o}$ pro $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^n$	$\mathbf{x} = \mathbf{a} - (\mathbf{g}'(\mathbf{a}))^{-1}\,\mathbf{g}(\mathbf{a})$	$\mathbf{g}'(\mathbf{a}) \in \mathbb{R}^{n\times n}$ regul.
Newtonova	$\min f(\mathbf{x})$ pro $f\colon \mathbb{R}^n \to \mathbb{R}$	$\mathbf{x} = \mathbf{a} - (f''(\mathbf{a}))^{-1}\,(f'(\mathbf{a}))^T$	$(f')^T = \mathbf{g}$
Gauss-Newtonova	$\min \lVert\mathbf{g}(\mathbf{x})\rVert^2$ pro $g\colon \mathbb{R}^n \to \mathbb{R}^m$	$\mathbf{x} = \mathbf{a} - (\mathbf{g}'(\mathbf{a}))^{+}\,\mathbf{g}(\mathbf{a})$	$(\mathbf{g}')^+ = ((\mathbf{g}')^T\mathbf{g}')^{-1}(\mathbf{g}')^T$
Levenberg-Marq.	$\min \lVert\mathbf{g}(\mathbf{x})\rVert^2$ pro $g\colon \mathbb{R}^n \to \mathbb{R}^m$	$\mathbf{x} = \mathbf{a} - (\mathbf{g}'(\mathbf{a}))^{*}\,\mathbf{g}(\mathbf{a})$	$(\mathbf{g}')^* = ((\mathbf{g}')^T\mathbf{g}' + \mu_k\mathbf{I})^{-1}(\mathbf{g}')^T$

Newtonova metoda najde obecně stacionární bod, ne nutně minimum.

Všechny metody mají iterační vzorec ve tvaru $\mathbf{a}$ – „směr“. Je-li ten „směr“ použit bez další multiplikativní konstanty, mluví se o *čisté* metodě. Někdy se „směr“ nejprve násobí nějakým koeficientem α_k, pak mluvíme o zobecněné metodě.

Doporučuji se neučit vzorečky metod nazpaměť, ale spíše je umět odvodit. Např. Newtonovu metodu odvodíme z obrázku ze strany 114 skript s nakreslenou tečnou, Gauss-Newtonovu metodu odvodíme z normální rovnice přeurčené soustavy $\mathbf{g}^T\mathbf{g} = 0$ (kde $\mathbf{g}$ lokálně aproximujeme jako $\mathbf{A}\mathbf{x} - \mathbf{b}$, takže $\mathbf{A} = \mathbf{g}'$) atd. Toto vše bylo jistě podrobně ukázáno na přednášce.

10.5 V úloze máme $g(x) = \sin x - \frac{1}{2}x$, tedy $g\colon \mathbb{R}^1 \to \mathbb{R}^1$ a hledáme x tak, že $g(x) = 0$. Iterační předpis Newtonovy metody je $x = a - (g'(a))^{-1}g(a) = a - \frac{1}{\cos a - 1/2}(\sin a - \frac{1}{2}a)$. V zadání úlohy byl požadavek použít kalkulačku a zkusit na ni najít výsledné x s největší přesností, jakou dokážeme. Moje oblíbená kalkulačka se jmenuje `bc` (pradávný unixový program „big calculator“). Ta umožňuje výpočty s libovolnou přesností. Z toho plyne, že nemohu výsledek napsat s největší přesností, jakou dokážu, protože tato přesnost není omezena (až na můj čas a strojovou paměť, což nejsou jednoznačně dané veličiny). Modře jsem podtrhnul, kolik cifer už je v dané iteraci vyhodnoceno přesně. Pro Newtonovu metodu jedné proměnné platí, že když se funkce chová „normálně“ a je nalezena dobrá výchozí iterace, pak každou další iterací se zhruba zdvojnásobí počet cifer s přesným výsledkem. To je ve výpočtu (který byl původně počítán na 20 míst) krásně vidět. Při přechodu na vyšší přesnost výpočtu stačila jedna další iterace, a měl jsem přesnost na 39 míst.



10.6 $f'(x,y) = [2x - 2\cos(y^2 - 2x),\ 2y\cos(y^2 - 2x) - 1]$, Hessián:

$$f''(x,y) = \begin{bmatrix} 2 - 4\sin(y^2 - 2x) & 4y\sin(y^2 - 2x) \\ 4y\sin(y^2 - 2x) & -4y^2\sin(y^2 - 2x) + 2\cos(y^2 - 2x) \end{bmatrix}$$

Iterační krok:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} 2 - 4\sin(b^2 - 2a) & 4b\sin(b^2 - 2a) \\ 4b\sin(b^2 - 2a) & -4b^2\sin(b^2 - 2a) + 2\cos(b^2 - 2a) \end{bmatrix}^{-1} \begin{bmatrix} 2a - 2\cos(b^2 - 2a) \\ 2b\cos(b^2 - 2a) - 1 \end{bmatrix}$$

Výchozí iterační krok pro $(a,b) = (1,1)$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 - 4\sin(-1) & 4\sin(-1) \\ 4\sin(-1) & -4\sin(-1) + 2\cos(-1) \end{bmatrix}^{-1} \begin{bmatrix} 2 - 2\cos(-1) \\ 2\cos(-1) - 1 \end{bmatrix} = \begin{bmatrix} 0.6521 \\ 0.7185 \end{bmatrix}$$

Nyní bychom dosadili $(a,b) = (0.6521, 0.7185)$ a iterační krok opakovali. Stroj typicky necháme opakovaně počítat inverzní matici (pokaždé s jinými konstantami). Druhá možnost je nechat stroj vypočítat inverzní matici symbolicky a pak jen v iteračních krocích dosazovat konkrétní (a,b). V tomto příkladě jednotlivé iterace vypadají takto: $(1,1) \to (0.6521, 0.7185) \to (0.6778, 0.7272) \to (0.6794, 0.7313) \to (0.6802, 0.7331) \to 15\times \to (0.68074537, 0.73448902)$. Po 20 iteracích máme výsledek s přesností na 8 desetinných míst.

10.8 Dostatečně podrobné řešení je ve skriptech.

10.9 Soustava není pochopitelně lineární (jsou tam monomy druhého stupně). Proč soustava nemá řešení je vysvětleno ve skriptech. Dále počítáme iterační vzorce, což bude *dřina*. Hledáme minimum funkce $f\colon \mathbb{R}^2 \to \mathbb{R}$:

$$f(x,y) = (x+y-2xy-1)^2 + (-x+y+xy+3)^2 + (x-y+xy-1)^2.$$

$$(f'(x,y))^T = \begin{bmatrix} 2(x+y-2xy-1)(1-2y) + 2(-x+y+xy+3)(-1+y) + 2(x-y+xy-1)(1+y) \\ 2(x+y-2xy-1)(1-2x) + 2(-x+y+xy+3)(1+x) + 2(x-y+xy-1)(-1+x) \end{bmatrix} =$$

$$= \begin{bmatrix} 12xy^2 - 8xy + 6x - 4y^2 + 6y - 10 \\ 12x^2y - 4x^2 + 6x - 8xy + 6y + 6 \end{bmatrix}.$$

$$f''(x,y) = \begin{bmatrix} 12y^2 - 8y + 6 & 24xy - 8x - 8y + 6 \\ 24xy - 8x - 8y + 6 & 12x^2 - 8x + 6 \end{bmatrix}.$$

Gradientní metoda:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} 12ab^2 - 8ab + 6a - 4b^2 + 6b - 10 \\ 12a^2b - 4a^2 + 6a - 8ab + 6b + 6 \end{bmatrix}.$$

Newtonova metoda:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} 12b^2 - 8b + 6 & 24ab - 8a - 8b + 6 \\ 24ab - 8a - 8b + 6 & 12a^2 - 8a + 6 \end{bmatrix}^{-1} \begin{bmatrix} 12ab^2 - 8ab + 6a - 4b^2 + 6b - 10 \\ 12a^2b - 4a^2 + 6a - 8ab + 6b + 6 \end{bmatrix}.$$

Gauss-Newtonova metoda:

$$\mathbf{g}(x,y) = \begin{bmatrix} x+y-2xy-1 \\ -x+y+xy+3 \\ x-y+xy-1 \end{bmatrix}, \quad \mathbf{g}'(x,y) = \begin{bmatrix} 1-2y & 1-2x \\ -1+y & 1+x \\ 1+y & -1+x \end{bmatrix},$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} - \left(\begin{bmatrix} 1-2b & -1+b & 1+b \\ 1-2a & 1+a & -1+a \end{bmatrix} \begin{bmatrix} 1-2b & 1-2a \\ -1+b & 1+a \\ 1+b & -1+a \end{bmatrix} \right)^{-1}$$
$$\begin{bmatrix} 1-2b & -1+b & 1+b \\ 1-2a & 1+a & -1+a \end{bmatrix} \begin{bmatrix} a+b-2ab-1 \\ -a+b+ab+3 \\ a-b+ab-1 \end{bmatrix}.$$

Levenberg-Marquardtova metoda:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} - \left(\begin{bmatrix} 1-2b & -1+b & 1+b \\ 1-2a & 1+a & -1+a \end{bmatrix} \begin{bmatrix} 1-2b & 1-2a \\ -1+b & 1+a \\ 1+b & -1+a \end{bmatrix} + \mu_k \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1}$$
$$\begin{bmatrix} 1-2b & -1+b & 1+b \\ 1-2a & 1+a & -1+a \end{bmatrix} \begin{bmatrix} a+b-2ab-1 \\ -a+b+ab+3 \\ a-b+ab-1 \end{bmatrix}.$$

10.10 Obecně minimální vzdálenost dvou křivek řeší optimalizační úloha

$$\min\{\|(x,y)-(u,v)\|^2;\ (x,y) \text{ leží na první křivce a } (u,v) \text{ leží na druhé křivce}\} =$$
$$= \min\{(x-u)^2 - (y-v)^2;\ x^2 = y,\ (u-2)^2 + v^2 = 1\},$$

což s metodou nejmenších čtverců $\min\{(x^2-y)^2 + ((x-2)^2 + y - 1)^2\}$ (viz úlohu zmíněnou v zadání) nemá nic společného. Protože poloha bodu na parabole minimalizující vzdálenost od kružnice je stejná, jako poloha bodu na parabole minimalizující vzdálenost od středu této kružnice, stačí řešit úlohu $\min\{(x-2)^2 + (y-0)^2, y = x^2\}$, což po převedení na jednu proměnou vyřešíme jako $\min\{(x-2)^2 + (x^2)^2\}$. To se dá zderivovat a najít minimum. Řešení kubické rovnice $4x^3 + 2x - 4 = 0$ je třeba spočítat strojem (třeba Newtonovou metodou), výsledek je přibližně 0.83512. Přímý vzorec pro kořeny této kubické rovnice dává *obludný* výsledek se čtyřmi do sebe vzájemně vnořenými druhými a třetími odmocninami a konstantami 330, 18 atd.

10.11 Máme odvodit vzorec Newtonovy metody $\mathbf{x} = \mathbf{a} - (f''(\mathbf{a}))^{-1}(f'(\mathbf{a}))^T$. Taylorův polynom druhého řádu funkce f v bodě $\mathbf{a}$ je $T(\mathbf{x}) = f(\mathbf{a}) + f'(\mathbf{a})\,(\mathbf{x}-\mathbf{a}) + (\mathbf{x}-\mathbf{a})^T\,(f''(\mathbf{a})/2)\,(\mathbf{x}-\mathbf{a})$. Minimum polynomu druhého řádu je v bodě, kde má polynom první derivaci nulovou, tedy v bodě $\mathbf{x}$, pro který je $T'(\mathbf{x}) = 0$. Funkci T derivujeme podle $\mathbf{x}$, tj. na objekty $f'(\mathbf{a})$ a $f''(\mathbf{a})$ se díváme jako na konstantní matice. Při derivování nám pomohou první a třetí vzorec v tabulce na straně 117 skript. $T'(\mathbf{x}) = f'(\mathbf{a}) + (\mathbf{x}-\mathbf{a})^T(f''(\mathbf{a})/2 + (f''(\mathbf{a}))^T/2) = f'(\mathbf{a}) + (\mathbf{x}-\mathbf{a})^T f''(\mathbf{a}) = 0$. Přejdeme k transponované rovnici: $f''(\mathbf{a})(\mathbf{x}-\mathbf{a}) = -(f'(\mathbf{a}))^T$. Vynásobíme zleva maticí inverzní: $\mathbf{x} - \mathbf{a} = -(f''(\mathbf{a}))^{-1}(f'(\mathbf{a}))^T$. a požadovaný vzorec je odvozen.