Matematická analýza 2

Extrémy funkcí více proměnných

Pavel Hrabák[1], Tomáš Kalvoda[2], Ivo Petr[3]

[1]pavel.hrabak@fit.cvut.cz, [2]tomas.kalvoda@fit.cvut.cz, [3]ivo.petr@fit.cvut.cz

Katedra aplikované matematiky
Fakulta informačních technologií
České vysoké učení technické v Praze

22. prosince 2024
ZS 2024/2025

Hlavní body

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_2(\mathbf{x})$$

$$\nabla f(\mathbf{a}) = \theta$$

$\nabla^2 f(\mathbf{a})$ PD/ND

Hlavní výsledky této přednášky

- Motivační zmínka obecných optimalizačních úloh.
- Zavedení lokálních extrémů funkcí více proměnných.
- Nutné podmínky pro existenci lokálního extrému funkcí více proměnných.
- Analytický postup pro hledání extrému funkcí více proměnných.

Hlavní body

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_2(\mathbf{x})$$

1. Motivace: Optimalizační úlohy
2. Extrémy funkcí více proměnných
3. Nutné podmínky
4. Postačující podmínka
5. Příklady

$$\nabla f(\mathbf{a}) = \theta$$

$\nabla^2 f(\mathbf{a})$ PD/ND

Co je to optimalizační úloha?

Co mají následující úlohy společného?

- Jak rovnoměrně rozmístit objekty na tiskovou podložku 3D tiskárny?
- Jak „naučit“ neuronovou síť vykonávat jistý úkol?
- Jak z několika potravin s různým složením vytvořit optimální jídelníček splňující zadaná dietologická kritéria?
- Jak na desku optimálně rozmístit elektronické součástky při splnění zejména prostorových omezení a nákladnosti výroby?
- Jak nalézt co nejkratší (nebo nejlevnější) trasu mezi zadanými místy?

Pod každou z těchto úloh dříme optimalizační úloha, tedy úloha nalézt minimum/maximum jisté funkce (potenciálně opravdu mnoha proměnných) za jistých omezujících podmínek.

Obecná optimalizační úloha

V maximální obecnosti by matematická formulace „obecné optimalizační úlohy“ mohla znít následovně:

Obecná optimalizační úloha

Nalezněte maximum/minimum (existuje-li; alespoň lokální) funkce $f : D_f \to \mathbb{R}$, kde $D_f \subset \mathbb{R}^n$, na množině

$$M = \{\mathbf{x} \in D_f \mid g_i(\mathbf{x}) = 0,\ i \in \hat{I},\ h_j(\mathbf{x}) \leq 0,\ j \in \hat{J}\},$$

kde I a J jsou nezáporná celá čísla a g_i a h_j funkce s definičními obory ležícími v $\mathbb{R}^n$, $i \in \hat{I}$, $j \in \hat{J}$, připouštíme i I nebo J nulové a pak odpovídající podmínky nemáme ($\hat{0} = \emptyset$).

Podle různých situací rozlišujeme řadu různých typů této optimalizační úlohy (pouze lineární funkce, pouze kvadratické funkce, konvexní funkce, funkce definované na $\mathbb{Z}^n$ nebo dokonce jenom $\{0, 1\}^n$, …).

Obecná optimalizační úloha

- V BI-MA2 se budeme věnovat úlohám bez rovnostních i bez nerovnostních omezení a s reálnými proměnnými.
- Tj. úloze nalézt (lokální) minima/maxima reálné funkce více reálných proměnných.
- Koncepty, které zde zavedeme, hrají ale důležitou roli i v případě úloh s omezeními, nebo v případě implementace numerických algoritmů snažících se tyto úlohy řešit.
- Úlohy s omezeními potkáte v magisterském povinném předmětu NI-MPI nebo ve specializovaných předmětech NI-LOM, NI-NON, nebo NI-KOP, ale i jinde.

Hlavní body

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_2(\mathbf{x})$$

1 Motivace: Optimalizační úlohy

2 Extrémy funkcí více proměnných

$$\nabla f(\mathbf{a}) = \theta$$

3 Nutné podmínky

4 Postačující podmínka

5 Příklady

$\nabla^2 f(\mathbf{a})$ PD/ND

Extrémy funkcí více proměnných

Následující definice jako by z oka vypadla té z BI-MA1:

Definice (Extrémy funkcí více proměnných):

Mějme funkci $f : D_f \to \mathbb{R}$, $D_f \subset \mathbb{R}^n$, a bod $\mathbf{a} \in D_f$. Funkce f má v bodě $\mathbf{a}$

- **ostré lokální minimum**, právě když existuje okolí $U_\mathbf{a}$ bodu $\mathbf{a}$ takové, že pro všechna $\mathbf{x} \in U_\mathbf{a} \cap D_f$ různá od $\mathbf{a}$ platí $f(\mathbf{x}) > f(\mathbf{a})$.
- **ostré lokální maximum**, právě když existuje okolí $U_\mathbf{a}$ bodu $\mathbf{a}$ takové, že pro všechna $\mathbf{x} \in U_\mathbf{a} \cap D_f$ různá od $\mathbf{a}$ platí $f(\mathbf{x}) < f(\mathbf{a})$.
- **lokální minimum**, právě když existuje okolí $U_\mathbf{a}$ bodu $\mathbf{a}$ takové, že pro všechna $\mathbf{x} \in U_\mathbf{a} \cap D_f$ platí $f(\mathbf{x}) \geq f(\mathbf{a})$.
- **lokální maximum**, právě když existuje okolí $U_\mathbf{a}$ bodu $\mathbf{a}$ takové, že pro všechna $\mathbf{x} \in U_\mathbf{a} \cap D_f$ platí $f(\mathbf{x}) \leq f(\mathbf{a})$.

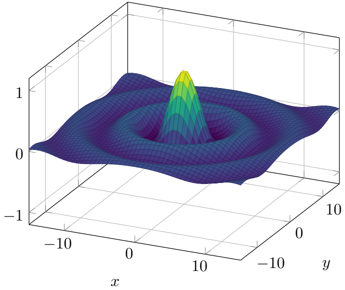
Hodnota tohoto extrému je ve všech případech $f(\mathbf{a})$.

Souhrnně budeme mluvit o (ostrém) lokálním extrému.

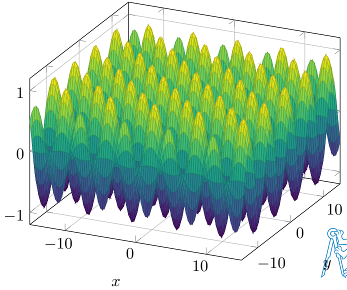
Extrémy funkcí více proměnných

- Nejprve se budeme soustředit na analytické hledání lokálních extrémů.
- Odvodíme nutné i postačující podmínky pro jejich existenci.
- V pozdější části přednášky si ukážeme i jeden z možných numerických přístupů k řešení tohoto problému (gradientní sestup).
- Je dobré si uvědomit, že struktura množiny bodů, kde je nabýván extrém, může být poměrně komplikovaná.

$f(x,y) = \frac{\sin\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}$



$f(x,y) = \sin(x)\cos(y)$



Hlavní body

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_2(\mathbf{x})$$

$$\nabla f(\mathbf{a}) = \theta$$

$\nabla^2 f(\mathbf{a})$ PD/ND

Nutná podmínka prvního řádu

Vzpomeňte si na známé tvrzení z BI-MA1: pokud má reálná funkce reálné proměnné lokální extrém v bodě a, pak její derivace v bodě a buď neexistuje, nebo je rovna nule.

Věta (Nutná podmínka existence lokálního extrému I):

Mějme funkci $f : D_f \to \mathbb{R}$, $D_f \subset \mathbb{R}^n$, mající v bodě $\mathbf{a}$ (ostrý) lokální extrém. Potom parciální derivace funkce f v bodě $\mathbf{a}$ podle j-té proměnné je rovna nule nebo neexistuje.

Důkaz.

Stačí použít výše zmíněné tvrzení na funkci $g(x) = f(\mathbf{a} + x\mathbf{e}_j)$ a vztahu mezi $g'(0)$ a $\partial_j f(\mathbf{a})$. □

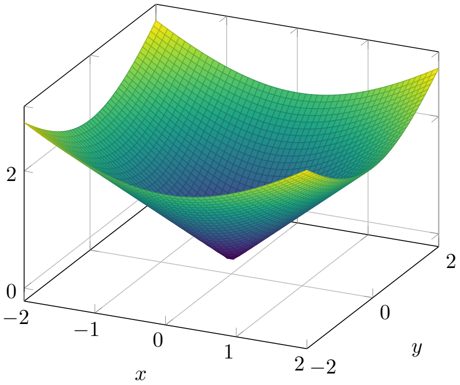
Důsledek:

Mějme funkci $f : D_f \to \mathbb{R}$, $D_f \subset \mathbb{R}^n$, mající v bodě $\mathbf{a}$ (ostrý) lokální extrém a mající parciální derivace v bodě $\mathbf{a}$ podle všech proměnných. Potom $\nabla f(\mathbf{a}) = \theta$.

Nutná podmínka prvního řádu

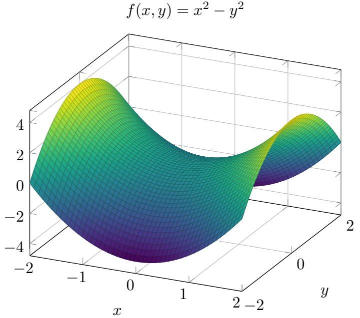
Ukázkou extrému (ostrého lokálního minima; snadno ukážeme z definice) v bodě s neexistujícím gradientem je například kužel:

$$f(x,y) = \sqrt{x^2 + y^2}$$



Nutná podmínka druhého řádu

Opět upozorněme, že jde pouze o nutnou podmínku! Z nulovosti gradientu neplyne existence extrému. Například: pro $f(x,y) = x^2 - y^2$ sice platí $\nabla f(0,0) = \theta$, ale v $(0,0)$ očividně (viz definice) extrém nenastává!



Definice (Stacionární bod / *stationary point*):

Mějme funkci $f : D_f \to \mathbb{R}$, $D_f \subset \mathbb{R}^n$. Bod $\mathbf{a} \in D_f$ splňující $\nabla f(\mathbf{a}) = \theta$ nazýváme **stacionárním bodem** (kritickým bodem nazýváme bod, kde neexistuje gradient nebo je stacionární).

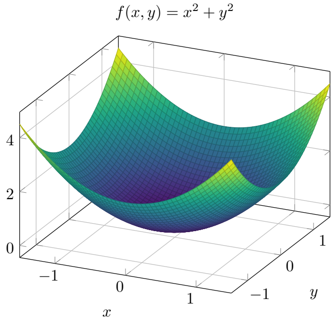
Nutná podmínka druhého řádu

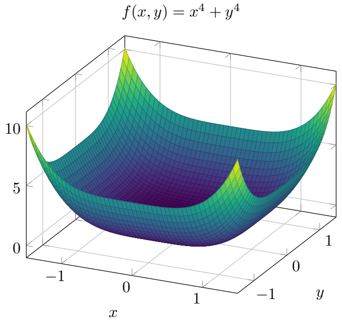
- V případě funkce jedné proměnné jsme často při hledání extrémů hledali nulové body první derivace a poté zkoumali znaménko první derivace na okolí takovéhoto bodu.
- U funkcí více proměnných tento přístup použitelný není. Musíme se podívat na analog druhé derivace.

Věta (Nutná podmínka existence lokálního extrému II):

Nechť funkce $f : D_f \to \mathbb{R}$, $D_f \subset \mathbb{R}^n$, má spojité všechny druhé parciální derivace na okolí bodu $\mathbf{a}$ a nechť má v tomto bodě lokální minimum (resp. maximum), potom je Hesseova matice $\nabla^2 f(\mathbf{a})$ PSD (resp. NSD).

Nutná podmínka druhého řádu





$$\nabla f(0,0) = \theta$$
$$\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

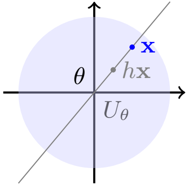
Ostré lokální minimum a PD Hesseova matice.

$$\nabla f(0,0) = \theta$$
$$\nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Ostré lokální minimum a pouze PSD Hesseova matice.

Podmínka druhého řádu: důkaz 1/2

BÚNO $\mathbf{a} = \theta$. Buď U_θ okolí bodu θ kde má funkce f spojité všechny parciání derivace až do druhého řádu (včetně) a platí $f(\mathbf{x}) \geq f(\theta)$ pro každé $\mathbf{x} \in U_\theta$.



Mějme $\mathbf{x} \in U_\theta$ libovolné nenulové a uvažme funkci $g(h) = f(h\mathbf{x})$ definovanou určitě na nějakém okolí $U_0 \subset \mathbb{R}$ obsahujícím 1 ($\mathbf{x} \in U_\theta$).
Tato funkce má v bodě 0 lokální minimum $(g(h) \geq f(\theta))$ a pro její derivace podle pravidla o derivaci složené funkce platí

$$g'(h) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(h\mathbf{x}) \cdot x_j,$$

$$g''(h) = \sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_k \partial x_j}(h\mathbf{x}) \cdot x_j x_k.$$

g má proto spojitou druhou derivaci a navíc dle předpokladu platí $g'(0) = 0$.

Podmínka druhého řádu: důkaz 2/2

Podle Taylorovy věty pro $h \in U_0$ platí

$$0 \leq g(h) - g(0) = \frac{1}{2} g''(\xi_h) h^2, \qquad (\star)$$

kde ξ_h leží mezi 0 a h (a tedy $\xi_h \to 0$ když $h \to 0$).

Dále podle předchozího slide platí

$$g''(\xi_h) = \mathbf{x}^T \cdot \nabla^2 f(\xi_h \mathbf{x}) \cdot \mathbf{x}.$$

Dosazením tohoto vztahu do $(\star)$, podělením nerovnosti h^2 a provedením limity $h \to 0$ získáváme nerovnost

$$\mathbf{x}^T \cdot \nabla^2 f(\theta) \cdot \mathbf{x} \geq 0,$$

kde $\mathbf{x}$ bylo libovolné z U_θ.

Libovolné $\mathbf{x} \neq \theta$ ale snadno přenásobením vhodnou kladnou konstantou lze převést do U_θ, kde nerovnost platí a poté opět vytknout škálovací faktor.

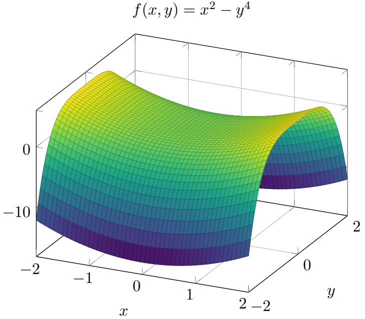
Tím je důkaz dokončen.

Nutná podmínka prvního řádu

Opět upozorněme, že jde pouze o nutnou podmínku! Z nulovosti gradientu a PSD (resp. NSD) Hessovy matice v daném bodě neplyne existence lokálního extrému. Například: pro $f(x,y) = x^2 - y^4$ sice platí $\nabla f(0,0) = \theta$ a Hesseova matice

$$\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

je PSD, ale v $(0,0)$ očividně (viz definice) extrém nenastává!



Hlavní body

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}) + R_2(\mathbf{x})$$

1. Motivace: Optimalizační úlohy
2. Extrémy funkcí více proměnných
3. Nutné podmínky
4. Postačující podmínka
5. Příklady

$$\nabla f(\mathbf{a}) = \theta$$

$\nabla^2 f(\mathbf{a})$ PD/ND

Postačující podmínka

Po studiu nutných podmínek konečně přistupme k postačující podmínce. Stačí zpřísnit podmínku definitnosti (ze semidefinitnosti na definitnost).

Navíc dále ukážeme, že ID existenci extrému vylučuje (tzv. **sedlový bod**).

Věta (Postačující podmínka existence lokálního extrému):

Mějme funkci $f : D_f \to \mathbb{R}$, $D_f \subset \mathbb{R}^n$, mající spojité všechny třetí parciální derivace na okolí bodu $\mathbf{a}$ a nechť jsou splněny následující dvě podmínky

1. $\nabla f(\mathbf{a}) = \theta$,
2. $\nabla^2 f(\mathbf{a})$ je PD (resp. ND).

Potom má funkce f v bodě $\mathbf{a}$ ostré lokální minimum (resp. maximum).

Pokud platí první podmínka a Hesseova matice $\nabla^2 f(\mathbf{a})$ je ID, pak tato funkce v bodě $\mathbf{a}$ lokální extrém nemá.

Postačující podmínka

Než se pustíme do důkazu, tak opět upozorníme na záludnosti předchozí věty.
Studenti mají tendenci v předchozí větu rozšiřovat i na případ PSD (resp. NSD) a z toho pak vyvozovat existenci *neostrého* lokálního extrému.
Následující příklad ukazuje, že nic takového neplatí.

Příklad.

Funkce $f(x,y) = x^2 + y^3$ má v bodě θ nulový gradient, skutečně:

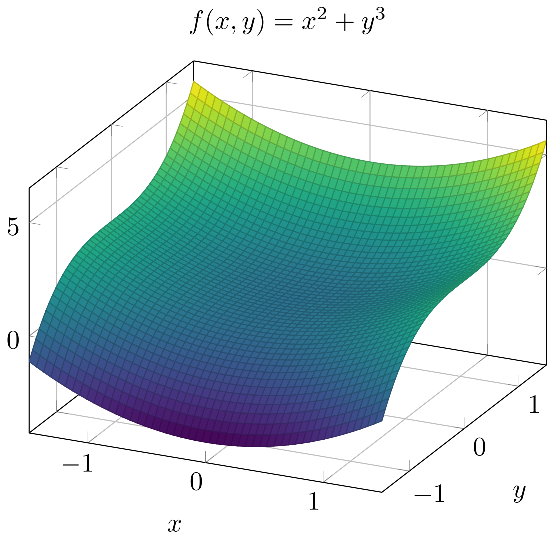
$$\nabla f(x,y) = \big(2x,\, 3y^2\big) \;\Rightarrow\; \nabla f(\theta) = \theta.$$

Pro Hessovu matici platí

$$\nabla^2 f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 6y \end{pmatrix}$$

a proto $\nabla^2 f(\theta) = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Odpovídající kvadratická forma je rovna $(x_1, x_2) \mapsto 2x_1^2$ a je PSD.
Funkce f v bodě θ extrém nemá: $f(0,y) = y^3$ je kladná pro $y > 0$ a záporná pro $y < 0$, $f(\theta) = 0$.

Postačující podmínka



Postačující podmínka

Klíčem k důkazu je v podstatě Taylorova věta, kterou zde zformulujeme jenom v jednodušší verzi využívající pouze kvadratické členy.

Lemma (Taylorova věta do kvadratických členů s odhadem chyby):

Mějme funkci $f : D_f \to \mathbb{R}$, $D_f \subset \mathbb{R}^n$, mající spojité všechny parciální derivace do třetího řádu včetně na okolí $U_\mathbf{a}$ bodu $\mathbf{a} \in D_f$. Potom existuje konstanta $M > 0$ taková, že pro každé $\mathbf{x} \in U_\mathbf{a}$ platí

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_2(\mathbf{x}),$$

kde $|R_2(\mathbf{x})| \leq M\|\mathbf{x} - \mathbf{a}\|^3$.

Postačující podmínka: důkaz 1/3

Náčrt důkazu Lemmatu.

Uvažme $\mathbf{x} \in U_{\mathbf{a}}$ a funkci $g(t) := f(\mathbf{a} + (\mathbf{x} - \mathbf{a})t)$, definovanou pro všechna přípustná t (zcela jistě pro nějaký otevřený interval obsahující interval $\langle 0, 1\rangle$). Funkce g má na tomto intervalu spojité derivace do řádu 3 včetně.
Podle Taylorovy věty pro funkci g a $t \in \langle 0, 1\rangle$ platí

$$g(t) = g(0) + g'(0)t + \frac{1}{2}g''(0)t^2 + \frac{1}{3!}g'''(\xi)t^3, \quad \text{kde } \xi \in \langle 0, t).$$

Použitím věty o derivaci složené funkce opět zjistíme, že

$$g'(0) = \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) \quad \text{a} \quad g''(0) = (\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}).$$

Konečně, pro g''' platí

$$g'''(t) = \sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}\big(\mathbf{a} + (\mathbf{x} - \mathbf{a})t\big)(\mathbf{x} - \mathbf{a})_i(\mathbf{x} - \mathbf{a})_j(\mathbf{x} - \mathbf{a})_k.$$

Díky spojitosti parciálních derivací a uzavřenosti intervalu $\langle 0, 1\rangle$ můžeme v absolutní hodnotě tento výraz odhadnout výrazem tvaru $C \cdot \left(\sum_{j=1}^{n} |(\mathbf{x} - \mathbf{a})_i|\right)^3$. Lze ukázat, že tento výraz je dále menší než $M\|\mathbf{x} - \mathbf{a}\|^3$ a konstanta M nezávisí na $\mathbf{x}$. Nyní stačí položit $t = 1$. □

Postačující podmínka: důkaz 2/3

Důkaz případu PD: Mějme okolí $\mathcal{U} := U_{\mathbf{a}}$ bodu $\mathbf{a}$, na kterém má funkce f spojité všechny třetí parciální derivace a nechť platí uvedené podmínky: $\nabla f(\mathbf{a}) = \theta$ a $\nabla^2 f(\mathbf{a})$ je PD. Uvažme libovolný bod $\mathbf{x} \in U_{\mathbf{a}}$ různý od $\mathbf{a}$. Podle předchozího Lemmatu existuje konstanta $M > 0$ nezávislá na $\mathbf{x}$ taková, že

$$f(\mathbf{x}) - f(\mathbf{a}) = 0 + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}) + R_2(\mathbf{x}),$$

kde $|R_2(\mathbf{x})| < M\|\mathbf{x}-\mathbf{a}\|^3$.
Z PD kvadratické formy $q(\mathbf{y}) = \mathbf{y}^T \nabla^2 f(\mathbf{a})\mathbf{y}$ plyne existence ortogonální regulární matice $\mathbf{P}$ splňující

$$q(\mathbf{y}) = (\mathbf{P}\mathbf{y})^T \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{P}\mathbf{y} = \sum_{j=1}^{n} \lambda_j (\mathbf{P}\mathbf{y})_j^2 \geq \lambda_* \sum_{j=1}^{n} (\mathbf{P}\mathbf{y})_j^2 = \lambda_* \|\mathbf{y}\|^2,$$

kde λ_* je nejmenší z kladných vlastních čísel $\nabla^2 f(\mathbf{a})$.
Je-li nyní $\mathbf{x} \in U_{\mathbf{a}}(\varepsilon) \subset \mathcal{U}$, kde $M\varepsilon < \frac{\lambda_*}{4}$, pak

$$f(\mathbf{x}) - f(\mathbf{a}) \geq \frac{\lambda_*}{2}\|\mathbf{x}-\mathbf{a}\|^2 - M\|\mathbf{x}-\mathbf{a}\|^3 > \|\mathbf{x}-\mathbf{a}\|^2 \left(\frac{\lambda_*}{2} - M\varepsilon\right) > \frac{\lambda_*}{4}\|\mathbf{x}-\mathbf{a}\|^2.$$

Postačující podmínka: důkaz 3/3

Případ ND ihned plyne z PD (jaký je vztah mezi typem extrému funkce f a $-f$, definitností formy q a $-q$?)

Případ ID se ošetří analogicky, využijeme existence $\mathbf{x}$ a $\mathbf{y}$ takových, že

$$(\mathbf{x}-\mathbf{a})^T \nabla^2 f(\mathbf{a})(\mathbf{x}-\mathbf{a}) > 0,$$
$$(\mathbf{y}-\mathbf{a})^T \nabla^2 f(\mathbf{a})(\mathbf{y}-\mathbf{a}) < 0,$$

k tomu, abychom v odpovídajícím směru od $\mathbf{a}$ nalezli vhodným škálováním (dost blízko k $\mathbf{a}$) kladná h a t splňující

$$f(\mathbf{a}+h\mathbf{x}) > f(\mathbf{a}) \quad \text{a} \quad f(\mathbf{a}+t\mathbf{y}) < f(\mathbf{a}).$$

Hlavní body

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_2(\mathbf{x})$$

1. Motivace: Optimalizační úlohy
2. Extrémy funkcí více proměnných
3. Nutné podmínky
4. Postačující podmínka
5. Příklady

$$\nabla f(\mathbf{a}) = \theta$$

$\nabla^2 f(\mathbf{a})$ PD/ND

Příklad

Příklad.

Nalezněte extrémy a sedlové body funkce

$$f(x,y) = x^4 + y^4 - x^2 - 2xy - y^2.$$

Definičním oborem této funkce je $D_f = \mathbb{R}^2$. Pro gradient platí

$$\nabla f(x,y) = \big(4x^3 - 2x - 2y, 4y^3 - 2x - 2y\big), \quad (x,y)^T \in \mathbb{R}^2.$$

Jeho nulovost je ekvivalentní podmínkám

$$2x^3 = x + y \ \wedge \ 2y^3 = x + y,$$

které nutně vyžadují $x^3 = y^3$, tj. $x = y$. Dosazením do původních rovnic dostaneme podmínku $x^3 - x = x(x^2 - 1) = 0$.
Stacionární body existují proto celkem 3:

$$\mathbf{a} = (-1,-1)^T, \quad \mathbf{b} = (0,0)^T, \quad \mathbf{c} = (1,1)^T.$$

Příklad: pokračování

Pro Hessovu matici funkce f dále platí

$$\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 - 2 & -2 \\ -2 & 12y^2 - 2 \end{pmatrix}, \quad (x,y)^T \in \mathbb{R}^2.$$

Ve stacionárních bodech pak máme

$$\nabla^2 f(\mathbf{a}) = \begin{pmatrix} 10 & -2 \\ -2 & 10 \end{pmatrix} = \nabla^2 f(\mathbf{c}), \quad \nabla^2 f(\mathbf{b}) = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix}.$$

První matice je PD (Sylvester: $10 > 0$ a $100 - 4 = 96 > 0$). V bodech $\mathbf{a}$ a $\mathbf{c}$ má naše funkce proto ostrá lokální minima.

Druhá matice je NSD: $-2x^2 - 4xy - 2y^2 = -2(x+y)^2$. Naše kritéria proto ohledně ne/existence extrému v bodě $\mathbf{b}$ nic neimplikují. Musíme provést podrobnější inspekci.

Příklad: pokračování

Pro naší funkci platí $f(x,y) = x^4 + y^4 - x^2 - 2xy - y^2 = x^4 + y^4 - (x+y)^2$.

Využitím u $\mathbf{b} = \theta$ podstatnějšího (ale vynulovatelného) záporného členu se zkusme k bodu θ blížit následujícími dvěma způsoby:

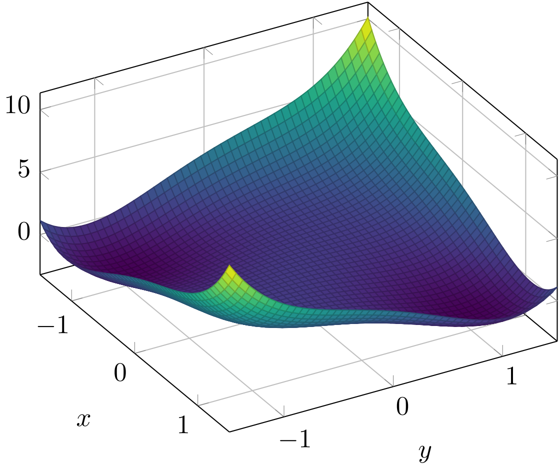
$$\begin{aligned} f(t,-t) &= 2t^4 > 0, && \text{pro } t \neq 0, \\ f(t,0) &= t^4 - t^2 = t^2(t^2-1) < 0, && \text{pro } t \in (-1,1) \smallsetminus \{0\}. \end{aligned}$$

V bodě $\mathbf{b}$ proto extrém nenastává. Vzhledem k výše popsanému chování bychom ho také mohli označit za sedlový bod.

Na následujícím slide pro ukázku uvádíme graf této funkce.

Příklad: pokračování

$$f(x, y) = x^4 + y^4 - x^2 - 2xy - y^2$$



Příklad: Regrese pomocí metody nejmenších čtverců

Předpokládejme, že máme k dispozici sadu dat $\{(x_i, y_i)\}_{i=1}^n$ a chceme nalézt lineární kombinaci daných funkcí $f_1, f_2, \ldots, f_m$, tj. funkci

$$f_{\mathbf{c}} = \sum_{i=1}^{m} c_i f_i \,,$$

tak, aby funkční hodnoty funkce f v bodech x_i co **nejlépe** odpovídaly hodnotám y_i pro každé $i \in \hat{n}$. Alternativně: snažíme se dobře vystihnout hypotetickou závislost $y \approx f(x)$.

Úkolem je určit neznámé koeficienty lineární kombinace: $c_1, c_2, \ldots, c_m$.

Předpokládejme, že $m \leq n$. To není na závadu, neboť typicky je množství dat n mnohem větší než počet funkcí m.

Funkce f_i mohou být např. voleny tak, že $f_i(x) = x^i$. V tomto případě prokládáme data polynomiální křivkou. Funkce $f_1, f_2, \ldots, f_m$ nelze volit úplně libovolně, jedna podmínka svazující funkce $f_1, f_2, \ldots, f_m$ a data $x_1, x_2, \ldots, x_n$ nám vypadne z výpočtu dále.

Příklad: Regrese pomocí metody nejmenších čtverců

Metoda nejmenších čtverců spočívá v myšlence minimalizovat kvadrát celkové chyby mezi y_i a $f(x_i)$. Přesněji, hledáme hodnoty $c_1, c_2, \ldots, c_m$ tak, aby hodnota

$$F(\mathbf{c}) := \sum_{i=1}^{n} \big(y_i - f_{\mathbf{c}}(x_i)\big)^2 = \sum_{i=1}^{n} \big(y_i - (\mathbf{A}\mathbf{c})_i\big)^2 = \|\mathbf{y} - \mathbf{A}\mathbf{c}\|_2^2$$

byla co nejmenší. Zde jsme označili

$$(\mathbf{A}\mathbf{c})_i = \sum_{j=1}^{m} f_j(x_i) c_j \,.$$

Matice $\mathbf{A} \in \mathbb{R}^{n,m}$ je dána hodnotami jednotlivých funkcí $f_1, f_2, \ldots, f_m$ v bodech $x_1, x_2 \ldots, x_n$:

$$\mathbf{A} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & f_3(x_1) & \cdots & f_m(x_1) \\ f_1(x_2) & f_2(x_2) & f_3(x_2) & \cdots & f_m(x_2) \\ f_1(x_3) & f_2(x_3) & f_3(x_3) & \cdots & f_m(x_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & f_3(x_n) & \cdots & f_m(x_n) \end{pmatrix}. \tag{1}$$

Příklad: Regrese pomocí metody nejmenších čtverců

Metodu nejmenší čtverců tedy můžeme shrnout takto

$$\begin{cases} \text{minimalizuj } \|\mathbf{y} - \mathbf{A}\mathbf{c}\|_2^2\,, \\ \text{vůči } \mathbf{c} = (c_1, c_2, \ldots, c_m)^T \in \mathbb{R}^m\,, \\ \text{kde } \mathbf{y} = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n \text{ a } \mathbf{A}_{ij} = f_j(x_i) \text{ jsou dány}\,. \end{cases} \tag{2}$$

Aplikujme analytický postup pro hledání extrémů funkcí více proměnných. Předpokládejme navíc, že matice $\mathbf{A}$ má plnou hodnost, tzn. $h(\mathbf{A}) = m$ (předpokládáme $m \leq n$).

Nejprve hledejme gradient funkce F. Pro její parciální derivaci platí

$$(\partial_{c_j} F)(\mathbf{c}) = -2\sum_{i=1}^{n} \big(y_i - (\mathbf{A}\mathbf{c})_i\big)\mathbf{A}_{ij} = -2\sum_{i=1}^{n} (\mathbf{A}^T)_{ji} y_i + 2\sum_{i=1}^{n} (\mathbf{A}^T)_{ji} (\mathbf{A}\mathbf{c})_i =$$
$$= -2(\mathbf{A}^T\mathbf{y})_j + 2(\mathbf{A}^T\mathbf{A}\mathbf{c})_j\,.$$

Příklad: Regrese pomocí metody nejmenších čtverců

Pro její gradient proto platí

$$(\nabla_{\mathbf{c}} F)(\mathbf{c})^T = -2\mathbf{A}^T\mathbf{y} + 2\mathbf{A}^T\mathbf{A}\mathbf{c}\,, \tag{3}$$

kde spodní index u symbolu $\nabla_{\mathbf{c}}$ nám připomíná, vůči kterým proměnným derivujeme.

Hledáme bod $\mathbf{c}$, kde $\nabla_{\mathbf{c}} F(\mathbf{c}) = \theta$. Protože $\mathbf{A}$ má dle předpokladu plnou hodnost, je matice $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{m,m}$ regulární. Z rovnosti (3) potom dostáváme řešení

$$\mathbf{c} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}\,, \tag{4}$$

resp. umíme vždy vyřešit příslušnou nehomogenní lineární soustavu.

Příklad: Regrese pomocí metody nejmenších čtverců

Hessovou maticí funkce F je shodou okolností právě matice $\mathbf{A}^T\mathbf{A}$ (až na multiplikativní číselný faktor). Skutečně, pro prvek Hessovy matice funkce F platí

$$\partial^2_{c_i c_j} F(c) = 2(\mathbf{A}^T\mathbf{A})_{ji}\,.$$

Matice $\mathbf{A}^T\mathbf{A}$ je pozitivně definitní, neboť pro $\mathbf{x} \in \mathbb{R}^m$ je

$$x^T\big(\mathbf{A}^T\mathbf{A}\big)\mathbf{x} = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$$

a rovnost nastává pouze pro $\mathbf{x} = 0$, protože je $\mathbf{A}^T\mathbf{A}$ regulární.

Shrnutí: prakticky tedy stačí sestavit příslušnou matici a vyřešit uvedenou soustavu lineárních rovnic.

Hlavní body

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \cdot \nabla^2 f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_2(\mathbf{x})$$

6 Dodatek

$$\nabla f(\mathbf{a}) = \theta$$

$\nabla^2 f(\mathbf{a})$ PD/ND

Komentář

TBA