Matematická analýza 2

Spádové metody

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

1. listopadu 2024
ZS 2024/2025

Hlavní body

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha \nabla f(\mathbf{x}_n)^T$$

Hlavní výsledky této přednášky

- Popis principu spádové metody pro hledání lokálních extrémů.

- Ukázka dvou konkrétních metod.

Hlavní body

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha \nabla f(\mathbf{x}_n)^T$$

Úloha

V této sekci si ukážeme obecný postup při *numerickém* hledání lokálních minim[1] funkcí $f : \mathbb{R}^n \to \mathbb{R}$ bez omezujících podmínek na proměnné x_i, $i = 1, 2, \ldots, n$.

Řešíme tedy úlohu

$$\text{minimalizuj } f(\mathbf{x}),$$

kde

- $f : \mathbb{R}^n \to \mathbb{R}$ a x probíhá všechny možné hodnoty $\mathbf{x} \in D_f$,
- definiční obor D_f je otevřená množina,
- funkce f je minimálně dvakrát spojitě diferencovatelná[2].

[1] To není žádné omezení, maxima f jsou minima $-f$.

[2] Tj. všechny její parciální derivace až do druhého řádu včetně jsou spojité na D_f.

Spádová metoda: princip

Řešení hledáme iterativně pomocí tzv. **spádové metody** (*descent methods*), jejichž princip je jednoduchý:

V každé iteraci se snažíme snížit hodnotu minimalizované funkce f.

Nechť je dáno $\mathbf{x}^{(1)} \in D_f$. Sestavíme posloupnost vektorů $\mathbf{x}^{(k)}$, $k = 1, 2, \ldots$, podle rekurentního předpisu

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)} \Delta \mathbf{x}^{(k)} ,$$

kde „$\Delta \mathbf{x}^{(k)}$“ je vhodně zvolený vektor (směr poklesu) a $t^{(k)}$ délka tzv. kroku.

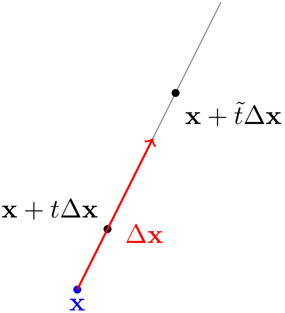
Následující člen vždy musí být zvolen tak, aby $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$, samozřejmě vyjma případu, kdy $\mathbf{x}^{(k)}$ je již optimální (je bodem lokálního minima).

Spádová metoda: obecný postup

Jeden krok typické spádové metody lze tedy popsat následovně (odpustíme si horní indexy, budeme mluvit o jednom kroku):

Nechť je dáno $\mathbf{x} \in D_f$.

1. Urči směr $\Delta\mathbf{x}$.
2. Zvol velikost kroku $t > 0$.
3. Napočti $\mathbf{x} + t\Delta\mathbf{x}$ a ulož ho do $\mathbf{x}$.



Tento postup opakuj dokud není splněno *kritérium pro zastavení iterační smyčky*. Praktické kritérium pro zastavení může být například dostatečná nulovost gradientu. Je také vhodné kontrolovat maximální povolený počet iterací.

Spádová metoda: obecná volba směru

Konkrétní volbě směru $\Delta\mathbf{x}$ se budeme věnovat podrobněji níže. V tento okamžik pouze uveďme, že pro t blízká nule přibližně platí (viz Taylor)

$$f(\mathbf{x} + t\Delta\mathbf{x}) \approx f(\mathbf{x}) + t\nabla f(\mathbf{x}) \cdot \Delta\mathbf{x}.$$

Pokud chceme být schopní i malým krokem funkční hodnotu zmenšit, musíme směr kroku, tj. $\Delta\mathbf{x}$, vždy volit tak, aby

$$\nabla f(\mathbf{x}) \cdot \Delta\mathbf{x} < 0.$$

Také ve zbytku popisu algoritmu předpokládáme, že v každém kroku $\nabla f(\mathbf{x}) \neq 0$. Pokud by gradient byl nulový, pak běh algoritmu již skončil.

Spádová metoda: volba t pomocí zpětného krokování

Nechť máme bod $\mathbf{x}$ a již jsme zvolili směr $\Delta\mathbf{x}$.

Ideálně bychom chtěli $t > 0$ zvolit tak, aby v tomto bodě t bylo nabyto minimum funkce jedné proměnné $s > 0$,

$$s \mapsto f(\mathbf{x} + s\Delta\mathbf{x})$$

Analytické řešení tohoto problému nemusí být obecně jednoduché, stejně bychom se museli uchýlit k numerickým metodám (ovšem s jednou proměnnou!).

V praxi se často používá tzv. **zpětné krokování** (*backtracking*), které popíšeme na následujícím slide.

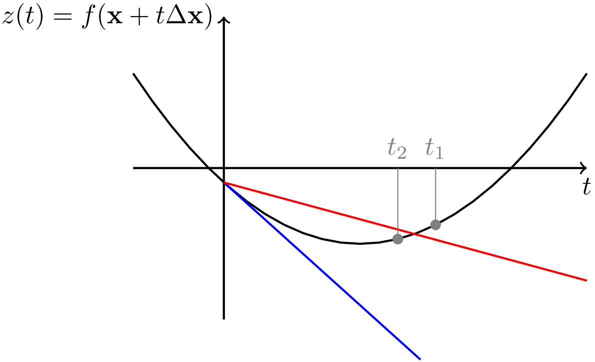
Spádová metoda: volba t pomocí zpětného krokování

Algoritmus **zpětného krokování** probíhá následovně:

- Jsou dány f, směr $\Delta\mathbf{x}$, bod $\mathbf{x} \in D_f$ a parametry $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.
- Polož $t := 1$. Dokud $f(\mathbf{x} + t\Delta\mathbf{x}) > f(\mathbf{x}) + \alpha t \nabla f(\mathbf{x}) \cdot \Delta\mathbf{x}$ změň t na βt.
- Vrať t.

Parametr α kontroluje, jaký pokles jsme ochotni akceptovat. Parametr β pak udává, jak rychle zmenšujeme prvotní t (proto zpětné krokování).

Spádová metoda: volba t pomocí zpětného krokování



Na obrázku je znázorněn řez grafu funkce nad přímkou $\mathbf{x} + t\Delta\mathbf{x}$ (černě), tečna $f(\mathbf{x}) + t\nabla f(\mathbf{x}) \cdot \Delta\mathbf{x}$ a přípustná mez poklesu $f(\mathbf{x}) + \alpha t\nabla f(\mathbf{x}) \cdot \Delta\mathbf{x}$.

- Bod t_1 by nebyl akceptován (i když k poklesu došlo).
- Bod t_2 by již akceptován byl (i když nedává nejmenší možnou hodnotu).

Hlavní body

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha \nabla f(\mathbf{x}_n)^T$$

Volby směru

Předpokládejme, že jsme v bodě $\mathbf{x}$ a vydáme se malým krokem směrem $\mathbf{s}$. Potom lineární aproximace změny funkční hodnoty je

$$f(\mathbf{x}+\mathbf{s}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{s}\,.$$

Směry, ve kterých dojde k poklesu funkční hodnoty, splňují $\nabla f(\mathbf{x}) \cdot \mathbf{s} < 0$.

Volby směru

Předpokládejme, že jsme v bodě $\mathbf{x}$ a vydáme se malým krokem směrem $\mathbf{s}$. Potom lineární aproximace změny funkční hodnoty je

$$f(\mathbf{x}+\mathbf{s}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{s}\,.$$

Směry, ve kterých dojde k poklesu funkční hodnoty, splňují $\nabla f(\mathbf{x}) \cdot \mathbf{s} < 0$.
Zvolme normu $\|\cdot\|$ na $\mathbb{R}^n$. Směr nejprudšího poklesu je pak dán řešením úlohy

$$\Delta\mathbf{x} = \operatorname{argmin}\left\{\nabla f(\mathbf{x}) \cdot \mathbf{s} \mid \|\mathbf{s}\| \leq 1\right\}.$$

Volby směru

Předpokládejme, že jsme v bodě $\mathbf{x}$ a vydáme se malým krokem směrem $\mathbf{s}$. Potom lineární aproximace změny funkční hodnoty je

$$f(\mathbf{x}+\mathbf{s}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{s}\,.$$

Směry, ve kterých dojde k poklesu funkční hodnoty, splňují $\nabla f(\mathbf{x}) \cdot \mathbf{s} < 0$.
Zvolme normu $\|\cdot\|$ na $\mathbb{R}^n$. Směr nejprudšího poklesu je pak dán řešením úlohy

$$\Delta \mathbf{x} = \operatorname{argmin} \{\nabla f(\mathbf{x}) \cdot \mathbf{s} \mid \|\mathbf{s}\| \leq 1\}\,.$$

Pro *reálnou* symetrickou pozitivně definitní matici $\mathbf{P} \in \mathbb{R}^{n,n}$ můžeme definovat normu (pro Euklidovskou máme $\mathbf{P} = \mathbf{E}$)

$$\|\mathbf{s}\|_{\mathbf{P}} := \sqrt{\mathbf{s}^T \mathbf{P} \mathbf{s}}\,, \quad \mathbf{s} \in \mathbb{R}^n\,.$$

Volby směru

Předpokládejme, že jsme v bodě $\mathbf{x}$ a vydáme se malým krokem směrem $\mathbf{s}$. Potom lineární aproximace změny funkční hodnoty je

$$f(\mathbf{x}+\mathbf{s}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{s}\,.$$

Směry, ve kterých dojde k poklesu funkční hodnoty, splňují $\nabla f(\mathbf{x}) \cdot \mathbf{s} < 0$. Zvolme normu $\|\cdot\|$ na $\mathbb{R}^n$. Směr nejprudšího poklesu je pak dán řešením úlohy

$$\Delta\mathbf{x} = \operatorname{argmin}\left\{\nabla f(\mathbf{x}) \cdot \mathbf{s} \mid \|\mathbf{s}\| \leq 1\right\}.$$

Pro *reálnou* symetrickou pozitivně definitní matici $\mathbf{P} \in \mathbb{R}^{n,n}$ můžeme definovat normu (pro Euklidovskou máme $\mathbf{P} = \mathbf{E}$)

$$\|\mathbf{s}\|_{\mathbf{P}} := \sqrt{\mathbf{s}^T\mathbf{P}\mathbf{s}}\,, \quad \mathbf{s} \in \mathbb{R}^n\,.$$

Směr největšího poklesu vzhledem k této normě je pak dán vektorem

$$\Delta\mathbf{x} = -\mathbf{P}^{-1}\nabla f(\mathbf{x})^T.$$

Důkaz pro případ $\mathbf{P} = \boldsymbol{E}$

Uvažme tedy obecně úlohu minimalizovat lineární reálnou funkci

$$g(\mathbf{s}) = \mathbf{c}^T \mathbf{s}, \quad D_g = \mathbb{R}^n,$$

kde $\mathbf{c} \in \mathbb{R}^n$ je zadaný vektor, za podmínky

$$\|\mathbf{s}\|^2 = \mathbf{s}^T \mathbf{s} = 1.$$

Vzpomeňme si, že ze Schwarzovy nerovnosti plyne nerovnost $|g(\mathbf{s})| \leq \|\mathbf{c}\| \|\mathbf{s}\|$, což za uvedené podmínky na vektory $\mathbf{s}$ implikuje $|g(\mathbf{s})| \leq \|\mathbf{c}\|$. Této maximální meze ale snadno dosáhneme právě volbami

$$\mathbf{s}_+ = \frac{c}{\|c\|} \quad \text{a} \quad \mathbf{s}_- = -\frac{c}{\|c\|}.$$

První vektor dává maximální hodnotu a druhý minimální hodnotu funkce g za uvedených podmínek. To jsme přesně očekávali: v kontextu předcházejících úvah je $\mathbf{c}$ gradient.

Důkaz pro obecný případ: náčrt

Uvažme funkci $g(\mathbf{s}) = \mathbf{c}^T\mathbf{s}$, $\mathbf{s} \in D_g = \mathbb{R}^n$, pro zadaný vektor $\mathbf{c} \in \mathbb{R}^n$. Hledáme řešení úlohy

$$\begin{cases} \text{minimalizuj} & g(\mathbf{s}) = \mathbf{c}^T\mathbf{s}\,, \\ \text{za podmínky} & \mathbf{s}^T\mathbf{P}\mathbf{s} = 1\,, \\ & \mathbf{s} \in \mathbb{R}^n\,. \end{cases}$$

kde $\mathbf{P}$ je (reálná) symetrická pozitivně definitní matice.

Důkaz pro obecný případ: náčrt

Uvažme funkci $g(\mathbf{s}) = \mathbf{c}^T\mathbf{s}$, $\mathbf{s} \in D_g = \mathbb{R}^n$, pro zadaný vektor $\mathbf{c} \in \mathbb{R}^n$. Hledáme řešení úlohy

$$\begin{cases} \text{minimalizuj} & g(\mathbf{s}) = \mathbf{c}^T\mathbf{s}\,, \\ \text{za podmínky} & \mathbf{s}^T\mathbf{P}\mathbf{s} = 1\,, \\ & \mathbf{s} \in \mathbb{R}^n\,. \end{cases}$$

kde $\mathbf{P}$ je (reálná) symetrická pozitivně definitní matice.

Matice $\mathbf{P}$ má všechna vlastní čísla kladná a je diagonalizovatelná, tj.

$$\mathbf{P} = \boldsymbol{U}\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{U}^T,$$

kde $\lambda_i > 0$, $i \in \hat{n}$, jsou vlastní čísla matice $\mathbf{P}$ a matice U je (reálná) a ortogonální.

Důkaz pro obecný případ: náčrt

Uvažme funkci $g(\mathbf{s}) = \mathbf{c}^T\mathbf{s}$, $\mathbf{s} \in D_g = \mathbb{R}^n$, pro zadaný vektor $\mathbf{c} \in \mathbb{R}^n$. Hledáme řešení úlohy

$$\begin{cases} \text{minimalizuj} & g(\mathbf{s}) = \mathbf{c}^T\mathbf{s}\,, \\ \text{za podmínky} & \mathbf{s}^T\mathbf{P}\mathbf{s} = 1\,, \\ & \mathbf{s} \in \mathbb{R}^n\,. \end{cases}$$

kde $\mathbf{P}$ je (reálná) symetrická pozitivně definitní matice.

Matice $\mathbf{P}$ má všechna vlastní čísla kladná a je diagonalizovatelná, tj.

$$\mathbf{P} = \boldsymbol{U}\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{U}^T,$$

kde $\lambda_i > 0$, $i \in \hat{n}$, jsou vlastní čísla matice $\mathbf{P}$ a matice U je (reálná) a ortogonální.

Definujme matici $\mathbf{P}^{1/2}$ předpisem

$$\mathbf{P}^{1/2} := \boldsymbol{U}\operatorname{diag}\left(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2}\right)\boldsymbol{U}^T.$$

Tato matice splňuje $\mathbf{P}^{1/2}\mathbf{P}^{1/2} = \mathbf{P}$, je (reálná) symetrická a regulární.

Důkaz pro obecný případ: náčrt

Naší minimalizační úlohu proto můžeme přepsat do následujícího tvaru,

$$\begin{cases} \text{minimalizuj} & g(\mathbf{s}) = \left((\mathbf{P}^{1/2})^{-1}\mathbf{c}\right)^T \mathbf{P}^{1/2}\mathbf{s}\,, \\ \text{za podmínky} & (\mathbf{P}^{1/2}\mathbf{s})^T(\mathbf{P}^{1/2}\mathbf{s}) = 1\,,\ \mathbf{s} \in \mathbb{R}^n\,. \end{cases}$$

Důkaz pro obecný případ: náčrt

Naší minimalizační úlohu proto můžeme přepsat do následujícího tvaru,

$$\begin{cases} \text{minimalizuj} & g(\mathbf{s}) = \left((\mathbf{P}^{1/2})^{-1}\mathbf{c}\right)^T \mathbf{P}^{1/2}\mathbf{s}\,, \\ \text{za podmínky} & (\mathbf{P}^{1/2}\mathbf{s})^T(\mathbf{P}^{1/2}\mathbf{s}) = 1\,,\ \mathbf{s} \in \mathbb{R}^n\,. \end{cases}$$

Pokud provedeme změnu souřadnic $\mathbf{z} := \mathbf{P}^{1/2}\mathbf{s}$, tak dostaneme minimalizační úlohu tvaru

$$\begin{cases} \text{minimalizuj} & \tilde{g}(\mathbf{z}) = \left((\mathbf{P}^{1/2})^{-1}\mathbf{c}\right)^T \mathbf{z}\,, \\ \text{za podmínky} & \mathbf{z}^T\mathbf{z} = 1 \ \ \mathbf{z} \in \mathbb{R}^n\,. \end{cases}$$

O té z předchozího případu víme, že nabývá maxima a minima v

$$\mathbf{z}_+ = \frac{(\mathbf{P}^{1/2})^{-1}\mathbf{c}}{\|(\mathbf{P}^{1/2})^{-1}\mathbf{c}\|} \quad \text{a} \quad \mathbf{z}_- = -\frac{(\mathbf{P}^{1/2})^{-1}\mathbf{c}}{\|(\mathbf{P}^{1/2})^{-1}\mathbf{c}\|}.$$

V původních proměnných pak v

$$\mathbf{s}_+ = \frac{\mathbf{P}^{-1}\mathbf{c}}{\|(\mathbf{P}^{1/2})^{-1}\mathbf{c}\|} \quad \text{a} \quad \mathbf{s}_- = -\frac{\mathbf{P}^{-1}\mathbf{c}}{\|(\mathbf{P}^{1/2})^{-1}\mathbf{c}\|}.$$

Hlavní body

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha \nabla f(\mathbf{x}_n)^T$$

Gradientová metoda / Gradientní sestup

Očividným kandidátem pro volbu směru kroku při minimalizaci je krok ve směru záporně vzatého gradientu (je-li nenulový), tj. volíme $\mathbf{P} = \mathbf{E}$,

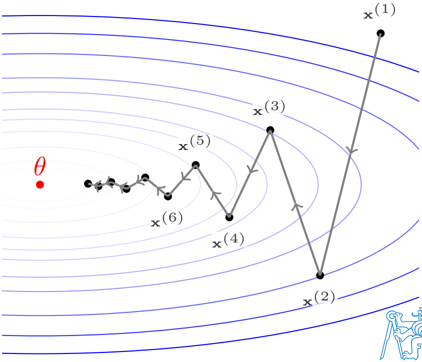
$$\Delta \mathbf{x} := -\nabla f(\mathbf{x})\,.$$

Potom skutečně

$$\nabla f(\mathbf{x}) \cdot \Delta \mathbf{x} = -\|\Delta \mathbf{x}\|^2 < 0$$

Poznámky:

- Relativně snadno spočítatelný.
- Výsledný směr nemusí být „nejvhodnější“.
- Nevyužívá informaci obsaženou ve vyšších derivacích.



Vpravo je ukázka pro $f(x, y) = \frac{1}{9}x^2 + y^2$.

Hlavní body

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha \nabla f(\mathbf{x}_n)^T$$

Newtonova metoda

Newtonova metoda volí směr v bodě $\mathbf{x}$ vzhledem k Hesseově matici $\mathbf{P} = \nabla^2 f(\mathbf{x})$, tedy dle výpočtů výše

$$\mathbf{s}_- = -\frac{\mathbf{P}^{-1}\mathbf{c}}{\|(\mathbf{P}^{1/2})^{-1}\mathbf{c}\|},$$

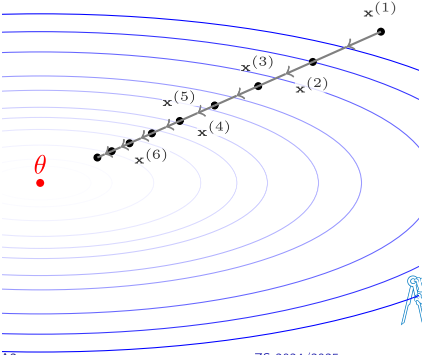
kde $\mathbf{c} = \nabla f(\mathbf{x})$. Tj. bez pro nás nepodstatného normalizačního faktoru dostáváme

$$\Delta\mathbf{x} = -\left(\nabla^2 f(\mathbf{x})\right)^{-1} \cdot \nabla f(\mathbf{x})^T .$$

- Vpravo uvádíme opět ukázku pro

$$f(x,y) = \frac{1}{9}x^2 + y^2.$$

- Pro kvadratickou funkci nepřekvapivě dostáváme směr přímo k θ.



Newtonova metoda: poznámky

- Neplést s Newtonovou metodou v BI-MA1.
- Výpočetně je tato metoda náročnější. Není ale potřeba počítat maticovou inverzi, směr $\Delta \mathbf{x}$ je řešením nehomogenní lineární soustavy

$$\nabla^2 f(\mathbf{x}) \cdot \Delta \mathbf{x} = \nabla f(\mathbf{x})^T.$$

- Lze očekávat podstatně lepší míru konvergence než u obyčejné gradientové metody.
- Existuje mnoho modifikací této metody, které zjednodušují výpočet ale stále se snaží získat informace o derivacích druhého řádu (Hesseově matici).

Newtonova metoda: poznámky

- Neplést s Newtonovou metodou v BI-MA1.
- Výpočetně je tato metoda náročnější. Není ale potřeba počítat maticovou inverzi, směr $\Delta\mathbf{x}$ je řešením nehomogenní lineární soustavy

$$\nabla^2 f(\mathbf{x}) \cdot \Delta\mathbf{x} = \nabla f(\mathbf{x})^T.$$

- Lze očekávat podstatně lepší míru konvergence než u obyčejné gradientové metody.
- Existuje mnoho modifikací této metody, které zjednodušují výpočet ale stále se snaží získat informace o derivacích druhého řádu (Hesseově matici).

Hlavní body

5 Dodatek

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha \nabla f(\mathbf{x}_n)^T$$

Komentář

TBA