Princip gradientních optimalizačních metod

Tomáš Kroupa

20. května 2014

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tento studijní materiál je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Obsah

Úkol a základní princip

Úkol

Nalézt

$$\min_{\mathbf{x}} \{f(\mathbf{x})\} \tag{1}$$

kde f je funkcí mnoha proměnných (parametrů)

$$f = f(x_1, x_2, ..., x_N), \tag{2}$$

na oblasti Ω, která je definovaná tak, že $x_1 \in \left\langle x_1^{lb}, x_1^{ub} \right\rangle$ atd.
Často se značí

$$\mathbf{x} = [x_1, x_2, ..., x_N]^{\mathrm{T}}. \tag{3}$$

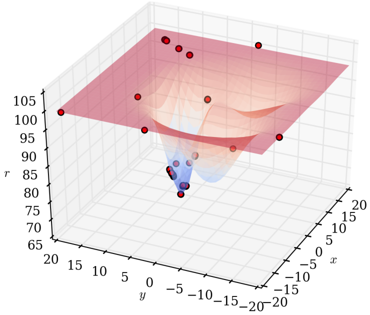
Často je nutné uvažovat ještě podmínky rovnosti a nebo nerovnosti[1]

$$\mathbf{0} \geq \mathbf{h}_{in}(\mathbf{x}) \quad \textit{a nebo} \quad \mathbf{0} = \mathbf{h}_{eq}(\mathbf{x}). \tag{4}$$

[1] Pak ovšem musí být algoritmy více propracované, princip těchto algoritmů snad bude jednou ukázán.

Úkol

Funkce f nemusí být popsaná jen analyticky, ale i tak, že pomocí parametrů $\mathbf{x}$ se postaví MKP model a výstup bude například maximální průhyb modelu, nebo první vlastní frekvence. A funkce f může být rozdíl mezi vlastní frekvencí, na kterou chceme model „naladit“ nebo ten samotný průhyb, který chceme minimalizovat apod.



Praktické příklady

1. Identifikace materiálových parametrů
   - f – Rozdíl mezi MKP a exp.
   - $\mathbf{x}$ – E_1, E_2, σ^y, ...
   - $0 \geq h$ – Relaxační časy u viskoelastického modelu větší než 0
2. Minimalizace hmotnosti lávky
   - f – Hmotnost lávky
   - $\mathbf{x}$ – Všechny možné rozměry, skladba vrstev laminátu,...
   - $0 \geq h$ – Průhyb, První vlastní frekvence > 6 Hz
3. Identifikace místa dopadu impaktoru
   - f – Rozdíl mezi signály z analýzy a exp.
   - $\mathbf{x}$ – Pozice razníku v analýze
   - $0 \geq h$ – Hodnoty rázové funkce musí být vždý vetší nebo rovny 0

Základní princip

Sestavení posloupnosti

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k \tag{5}$$

kde

- $\mathbf{x}^k$ je vektor hodnot parametrů v kroku k.
- α^k je délka kroku.
- $\mathbf{d}^k$ je směr kroku.

Rychlost konvergence a dost často i schopnost algoritmu nalézt minimum ovlivňuje volba $\mathbf{x}^0$.

Další obecné informace

- Při vhodné volbě $\mathbf{d}^k$ a vlastnostech funkce f lze dokázat, že výše zmíněná posloupnost konverguje k minimu (v okolí bodu minima).
- Volba $\mathbf{d}^k$ se liší metodu od metody.
- Parametr α^k se určuje jako řešení 1D optimalizace

$$\min_{\alpha^k} \left\{ f\left(\mathbf{x}^{k+1}\right) \right\} = \min_{\alpha^k} \left\{ f\left(\mathbf{x}^k + \alpha^k \mathbf{d}^k\right) \right\}. \tag{6}$$

- V praxi často nelze nalézt přesné minimum, proto se výpočet ukončuje podmínkou kdy

$$|f| \leq \epsilon, \tag{7}$$

kde ϵ je velmi malé číslo např. 10^{-6}.

Značení a co je co?

Gradient

$$\mathbf{c}\left(\mathbf{x}\right) = \bigtriangledown f\left(\mathbf{x}\right) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_N}\right]^{\mathrm{T}} \tag{8}$$

Hessova matice

$$\mathbf{H}\left(\mathbf{x}\right) = \bigtriangledown^2 f\left(\mathbf{x}\right) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \frac{\partial^2 f}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_N \partial x_N} \end{bmatrix} \tag{9}$$

Výpočet gradientu

- Analyticky – to snad nebudeme probírat, ne? ;-)
- Numericky je nutné definovat interval, na kterém bude gradient vypočten Δx_i.
  - Single sided

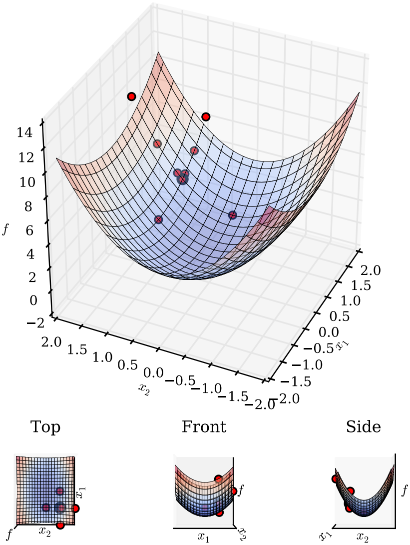
$$\frac{\partial f}{\partial x_1} = \frac{f(x_1 + \Delta x_1, x_2, ...) - f(x_1, x_2, ...)}{\Delta x_1} \tag{10}$$

  - Central differences

$$\frac{\partial f}{\partial x_1} = \frac{f(x_1 + \Delta x_1, x_2, ...) - f(x_1 - \Delta x_1, x_2, ...)}{2\Delta x_1} \tag{11}$$

  - A mnoho dalších ještě fikanějších způsobů.

Výpočet gradientu - ukázka na paraboloidu



$f = x_1^2 + 2x_2^2$

$\bigtriangledown f(x_1, x_2) = [2x_1, 4x_2]^T$

Analyticky

$\bigtriangledown f(1.0,1.0) = [2.0,4.0]^T$

Numericky -- single sided

$\bigtriangledown f(1.0,1.0) = [3.0,6.0]^T$, $[\Delta x_1, \Delta x_2]^T = [1,1]^T$

$\bigtriangledown f(1.0,1.0) = [2.5,5.0]^T$, $[\Delta x_1, \Delta x_2]^T = [0.5,0.5]^T$

$\bigtriangledown f(1.0,1.0) = [2.1,4.2]^T$, $[\Delta x_1, \Delta x_2]^T = [0.1,0.1]^T$

$\bigtriangledown f(1.0,1.0) = [2.01,4.02]^T$, $[\Delta x_1, \Delta x_2]^T = [0.01,0.01]^T$

Numericky -- central differences

$\bigtriangledown f(1.0,1.0) = [2.0,4.0]^T$, $[\Delta x_1, \Delta x_2]^T = [1,1]^T$

$\bigtriangledown f(1.0,1.0) = [2.0,4.0]^T$, $[\Delta x_1, \Delta x_2]^T = [0.01,0.01]^T$

Vybrané metody neomezené optimalizace

Metoda nejvetšího spádu

Připomeňme, že

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k.$$

Základní myšlenkou této metody je, že

$$\mathbf{d}^k = -\mathbf{c}^k = -\mathbf{c}\left(\mathbf{x}^k\right). \tag{12}$$

Metoda sdružených gradientů

Připomeňme, že

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k.$$

Základní myšlenkou této metody je, že

$$\mathbf{d}^k = -\mathbf{c}^k + \left(\frac{||\mathbf{c}^k||}{||\mathbf{c}^{k-1}||}\right)^2 \mathbf{d}^{k-1}. \tag{13}$$

Platí, že $\mathbf{d}^0 = -c^0$.

Newtonova metoda

Připomeňme, že

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k.$$

Základní myšlenkou této metody je, že

$$\mathbf{d}^k = -(\mathbf{H}^k)^{-1}\mathbf{c}^k, \tag{14}$$

kde se v praxi $\mathbf{d}^k$ získá jako řešení soustavy lineárních algebraických rovnic $-\mathbf{H}^k\mathbf{d}^k = \mathbf{c}^k$.

Newtonova metoda

- Typicky dochází k problémům pokud je v $\mathbf{H}^k$ singulární.
- Pokud ovšem je funkce f šikovná, má tato metoda nejrychlejší konvergenci.
- Pokud je $\alpha = 1$ jedná se o klasickou Newtonovou metodu, pokud se určuje délka kroku α jde o modifikovanou Newtonovu metodu.

BFGS – Broyden-Fletcher-Goldfarb-Shanno

Připomeňme, že

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k.$$

Základní myšlenkou této metody je, že

$$\mathbf{d}^k = -(\mathbf{B}^k)^{-1}\mathbf{c}^k, \tag{15}$$

kde

$$\begin{aligned}
\mathbf{s}^k &= \mathbf{x}^{k+1} - \mathbf{x}^k \text{ rozdíl parametrů,} && (16)\\
\mathbf{y}^k &= \mathbf{c}^{k+1} - \mathbf{c}^k \text{ rozdíl gradientu,} && (17)\\
\mathbf{B}^{k+1} &= \mathbf{B}^k - \frac{\mathbf{B}^k\mathbf{s}^k{\mathbf{s}^k}^{\mathrm{T}}\mathbf{B}^k}{{\mathbf{s}^k}^{\mathrm{T}}\mathbf{B}^k\mathbf{s}^k} + \frac{\mathbf{y}^k{\mathbf{y}^k}^{\mathrm{T}}}{{\mathbf{y}^k}^{\mathrm{T}}\mathbf{s}^k}, && (18)
\end{aligned}$$

přičemž opět lze řešit $-\mathbf{B}^k\mathbf{d}^k = \mathbf{c}^k$.

BFGS – Broyden-Fletcher-Goldfarb-Shanno

Případně lze inverzi matice $\mathbf{B}^{k+1}$ vypočítat i jiným způsobem a to pomocí Sherman–Morisson vztahu [2]

$$(\mathbf{B}^{k+1})^{-1} = (\mathbf{B}^k)^{-1} + \frac{\left(\mathbf{s}^{k\,\mathrm{T}}\mathbf{y}^k + \mathbf{y}^{k\,\mathrm{T}}(\mathbf{B}^k)^{-1}\mathbf{y}^k\right)\left(\mathbf{s}^k\mathbf{s}^{k\,\mathrm{T}}\right)}{\left(\mathbf{s}^{k\,\mathrm{T}}\mathbf{y}^k\right)^2} + \frac{(\mathbf{B}^k)^{-1}\mathbf{y}^k\mathbf{s}^{k\,\mathrm{T}} + \mathbf{s}^k\mathbf{y}^{k\,\mathrm{T}}(\mathbf{B}^k)^{-1}}{\mathbf{s}^{k\,\mathrm{T}}\mathbf{y}^k}. \tag{19}$$

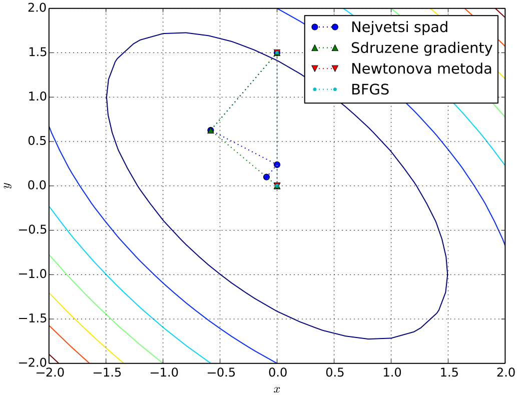
Často se bere $\mathbf{B}^0 = \mathbf{I}$ nebo $\mathbf{B}^0 = \mathbf{H}(\mathbf{x}^0)$.

[2] k a $k+1$ – jde jen o to z jakého kroku se na daný vztah koukáme.

Porovnání metod

Porovnání

$$f = x^2 + xy + \frac{3}{4}y^2$$



- Metoda nejvetsiho spadu

#	x	y	f
0	0.00	1.50	1.688
1	−0.58	0.63	0.269
2	−0.00	0.24	0.042
3	−0.09	0.10	0.006

- Metoda sdruzenych gradientu

#	x	y	f
0	0.00	1.50	1.688
1	−0.58	0.63	0.269
2	−0.00	0.00	0.000

- Newtonova metoda

#	x	y	f
0	0.00	1.50	1.688
1	−0.00	0.00	0.000

- BFGS

#	x	y	f
0	0.00	1.50	1.688
1	−0.00	0.00	0.000

Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky v rámci projektu č. **CZ.1.07/2.2.00/28.0206**

„Inovace výuky podpořená praxí“.

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tento studijní materiál je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.