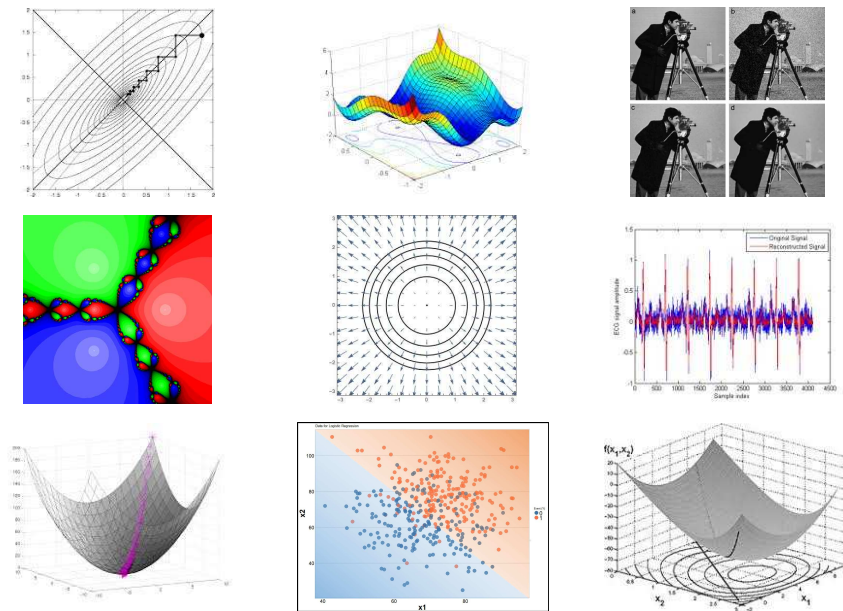
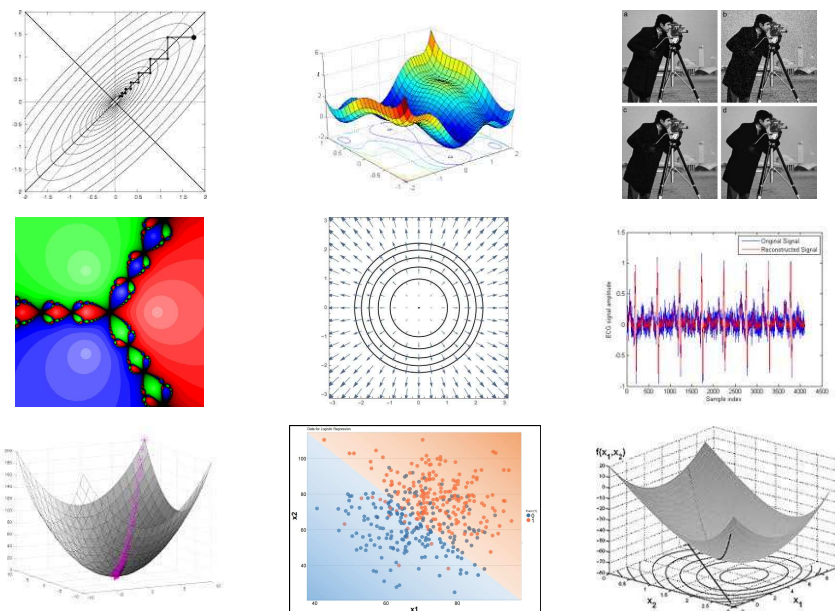




Metódy voľnej optimalizácie - letný semester 2018

M. Trnovská, KAMŠ, FMFI UK



Gradientná metóda

Gradientná metóda s konštantným krokom

Uvažujme úlohu na voľný extrém

$$\min_{x \in \mathbb{R}^n} f(x) \quad (U1)$$

s diferencovateľnou funkciou $f(x)$.

Gradient funkcie $f(x)$ budeme označovať

$$g(x) = \nabla f(x), \quad g_k = g(x^k).$$

Gradientná metóda s konštantným krokom

Ukážeme, že ak v metóde danej iteračnou formulou

$$x^{k+1} = x^k - cg_k,$$

(t. j. smer je záporný gradient funkcie $f(x)$) volíme za krok $\lambda_k = c$ dostatočne malú konštantu, tak príslušná postupnosť bodov $\{x^k\}$ konverguje k stacionárnemu bodu funkcie $f(x)$.

Veta 1

Nech $f(x)$ je zdola ohraničená na $\mathbb{R}^n$, t. j. existuje konštanta $\bar{f} \in \mathbb{R}$ taká, že

$$f(x) \geq \bar{f}.$$

Predpokladajme, že $f(x)$ je diferencovateľná s Lipschitzovským gradientom $g(x)$, t. j. existuje konštanta $L > 0$ tak, že pre ľubovoľnú dvojicu bodov x, y platí

$$\|g(x) - g(y)\| \leq L\|x - y\|.$$

Potom iteračný proces $x^{k+1} = x^k - cg_k$, s krokom $c > 0$ vyhovujúcim podmienke $0 < cL < 2$ má vlastnosť

$$\lim_{k \rightarrow \infty} g(x^k) = 0,$$

pričom tento iteračný proces je spádový.

Lema 1: Nech $f(x)$ je diferencovateľná na $\mathbb{R}^n$. Potom $\forall y \in \mathbb{R}^n$ platí

$$f(x + y) - f(x) = y^T g(x) + \int_0^1 y^T (g(x + ty) - g(x)) dt.$$

Dôkaz Vety 1:

- Aplikuje sa Lema 1 na $x = x^k$ a $y = -cg_k$.
- Cauchy Schwarzova nerovnosť + Lipschitzovkosť gradientu
- Dostaneme

$$f_{k+1} - f_k \leq -c_0 \|g_k\|_2^2,$$

kde $c_0 = \frac{1}{2}c(2 - Lc) > 0$.

- Z ohraničenosti funkcie potom máme

$$0 < \sum_{k=0}^m \|g_k\|_2^2 \leq \frac{1}{c_0}(f_0 - f_{m+1}) \leq \frac{1}{c_0}(f_0 - \bar{f}),$$

- Dôsledok: $\lim_{k \rightarrow \infty} g(x^k) = 0$,

Cauchyho metóda najväčšieho spádu

Smer: $s_k = -\nabla f(x^k) = -g_k$, krok λ_k - **optimálny**

Za predpokladov **Vety 1** je metóda konvergentná a dáva aspoň taký dobrý výsledok ako metóda s konštantným krokom.

Vo všeobecnosti nie je rýchlosť konverencie dobrá.

Smery v Cauchyho metóde sú na seba kolmé, t. j.

$$s_{k+1}^T s_k = 0.$$

Ukážeme, že rýchlosť konverencie Cauchyho metódy je v prípade kvadratickej funkcie **lineárna**.

Cauchyho metóda najväčšieho spádu

Budeme uvažovať **kvadratickú funkciu**

$$Q(x) = \frac{1}{2}x^T Gx + h^T x,$$

kde $G \in S_{++}^n$ (kladne definitná matica rozmeru $n \times n$) s vlastnými hodnotami $0 < \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$. a $h \in \mathbb{R}^m$.

Optimálna dĺžka kroku je daná ako

$$\lambda_k = \frac{g_k^T g_k}{g_k^T G g_k}.$$

Cauchyho metóda najväčšieho spádu Vlastnosti kvadratickej funkcie:

a) Pre každý bod $x \in \mathbb{R}^n$ a smer $p \in \mathbb{R}^n$ platí

$$Q(x + p) - Q(x) = p^T g(x) + \frac{1}{2} p^T G p,$$

b) Pre minimum $\hat{x}$ funkcie $Q(x)$ platí

$$Q(y) - Q(\hat{x}) = \frac{1}{2} (y - \hat{x})^T G (y - \hat{x}) = \frac{1}{2} g(y)^T G^{-1} g(y), \quad \forall y \in \mathbb{R}^n.$$

Cauchyho metóda najväčšieho spádu

Konvergenca Cauchyho metódy

Veta 2:

Uvažujme kvadratickú funkciu $Q(x) = \frac{1}{2}x^T Gx + h^T x$, s kladne definitnou maticou G a gradientom

$$\nabla Q(x) = g(x) = Gx + h.$$

Nech $0 < \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$ sú vlastné čísla matice G . Potom pre Cauchyho iteráciu platí

$$Q_{k+1} - \hat{Q} \leq \left(\frac{\mu_1 - \mu_n}{\mu_1 + \mu_n} \right)^2 (Q_k - \hat{Q}),$$

kde $\hat{Q} = Q(\hat{x})$ je minimálna hodnota a $Q_k = Q(x_k)$.

Cauchyho metóda najväčšieho spádu

Lema: (**Kantorovičova nerovnosť**)

Nech C je kladne definitná a $0 < \mu_1 \leq \dots \leq \mu_n$ sú jej vlastné hodnoty, potom

$$1 - \frac{(x^T x)^2}{(x^T C x)(x^T C^{-1} x)} \leq \left[\frac{\mu_1 - \mu_n}{\mu_1 + \mu_n} \right]^2.$$

Dôkaz:

S využitím vlastností kvadratickej funkcie vyššie dostaneme:

$$(Q_{k+1} - \hat{Q}) = \left(1 - \frac{(g_k^T g_k)^2}{(g_k^T G g_k)(g_k^T G^{-1} g_k)} \right) (Q_k - \hat{Q}).$$

Cauchyho metóda najväčšieho spádu

Ak označíme

$$q = \left(\frac{\mu_1 - \mu_n}{\mu_1 + \mu_n} \right)^2 \in [0, 1),$$

tak po m krokoch minimalizácie dostávame odhad

$$Q_m - \hat{Q} \leq q^m(Q_0 - \hat{Q}),$$

čo implikuje

$$\lim_{m \rightarrow \infty} Q_m = \hat{Q}.$$

Ak matica G funkcie $Q(x)$ má rozptýlené vlastné čísla (je zle podmienená), tak zrejme

$$q \approx 1$$

a rýchlosť konvergenzie je pomalá.

Cauchyho metóda najväčšieho spádu

Veta 3 (Lineárna konvergencia Cauchyho metódy):

Nech $\{x^k\}$ je postupnosť bodov generovaná Cauchyho gradientnou metódou pre minimalizáciu kvadratickej funkcie $Q(x) = \frac{1}{2}x^T Gx + h^T x$, ($G \succ 0$), konvergujúca k $\hat{x}$. Potom platí

G-norma:

$$\frac{\|x^{k+1} - \hat{x}\|_G}{\|x^k - \hat{x}\|_G} \leq \frac{\kappa - 1}{\kappa + 1}$$

l_2 -norma:

$$\frac{\|x^{k+1} - \hat{x}\|_2}{\|x^k - \hat{x}\|_2} \leq \sqrt{\kappa} \frac{\kappa - 1}{\kappa + 1}$$

kde $\kappa = \mu_n / \mu_1 \geq 1$.

Cauchyho metóda najväčšieho spádu

Príklad.

Uvažujme minimalizáciu kvadratickej funkcie

$$Q(x) = \frac{1}{2}x^T \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} x,$$

kde $a > 0$ je parameter, Cauchyho gradientnou metódou so štartovacím bodom $x^0 = [a, 1]^T$. Matematickou indukciou sa dá ukázať, že pre každé $k = 0, 1, 2, \dots$ platí

$$\lambda_k = \frac{2}{1+a}, \quad x^k = \left(\frac{a-1}{a+1} \right)^k \begin{bmatrix} a \\ (-1)^k \end{bmatrix}$$

Cauchyho metóda najväčšieho spádu

1988 - **Barzilai & Borwein** - *Two-point step size gradient methods*, IMA Journal of Numerical Analysis 8.

Špeciálnou voľbou kroku bola dosiahnutá **superlineárna konvergen-
cia**.

Výpočet dĺžky kroku vyžaduje informáciu z dvoch predchádzajúcich iterácií
- **neautonómna** iteračná metóda.

Označme

$$p_k = x^{k+1} - x^k, \quad y_k = g_{k+1} - g_k.$$

B & B metóda používa smer $s_{k+1} = -g_{k+1}$ a kroky (kratší a dlhší krok)

$$\lambda_{k+1} = \frac{p_k^T y_k}{y_k^T y_k}, \quad \lambda_{k+1} = \frac{p_k^T p_k}{p_k^T y_k}$$

Cauchyho metóda najväčšieho spádu

Zhrnutie

Cauchyho gradientná metóda vo všeobecnosti často vykazuje lineárnu konvergenciu (pre kvadratické úlohy vždy) - chybový člen $f(x^k) - \hat{f}$ konverguje k nule približne tak rýchlo ako geometrická postupnosť.

Konvergenca veľmi závisí od čísla podmienenosti Hessovej matice - môže byť pomalá aj v prípade keď je číslo podmienenosti rádovo 100.

Hlavnou výhodou je jednoduchosť.