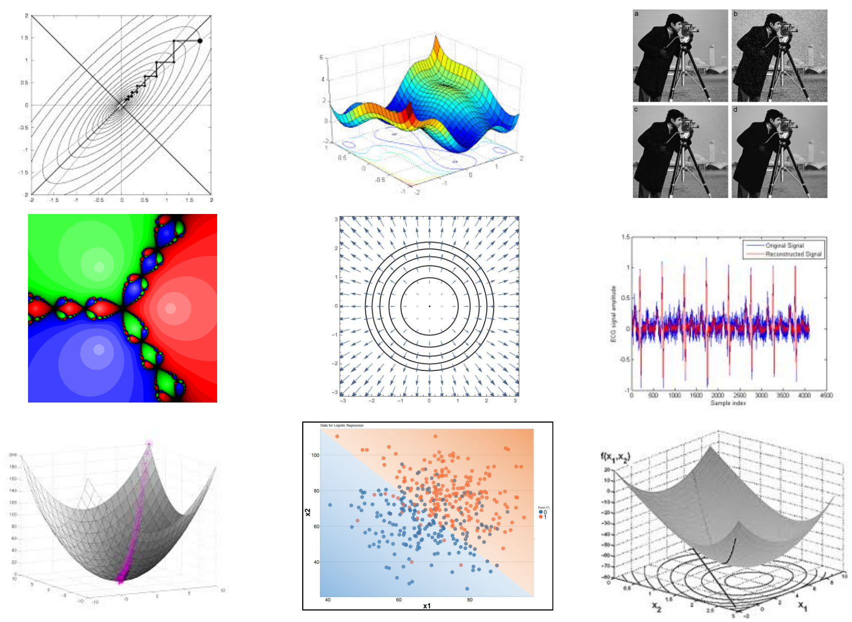
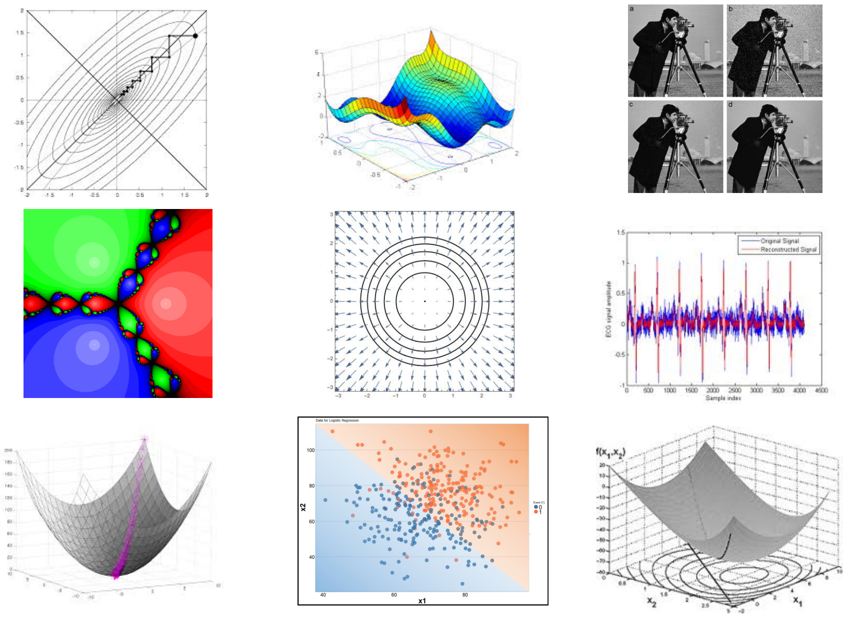


Metódy voľnej optimalizácie - letný semester 2018

M. Trnovská, KAMŠ, FMFI UK



Metóda združených gradientov

Združené smery a ich vlastnosti

Metóda CSR - pre separovaťeľné funkce konverguje za n krokov.

Teoreticky sa dá ľubovoľná kvadratická funkcia

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

miminmalizovať za n krokov nasledovným postupom:

1. Najskôr maticu G transformujeme na diagonálny tvar, t. j. funkciu $\mathcal{Q}(x)$ transformujeme na separovateľnú funkciu - umožňuje nám to spektrálny rozklad matice G:

$$G = Q\Lambda Q^T,\ \Lambda = diag(\lambda_1, \dots, \lambda_n),\ QQ^T = Q^T Q = I.$$

Združené smery a ich vlastnosti

Transformáciou premenných $z = Q^T x$ dostávame separovateľnú kvadratickú funkciu

$$\tilde{\mathcal{Q}}(z) = \sum_{i=1}^{n} \lambda_i z_i^2 + h^T Q z.$$

2. Takto získanú funkciu minimalizujeme metódou cyklickej súradnicovej redukcie.

3. Nakoniec sa spätnou transformáciou vrátime k pôvodným premenným.

Motivácia pre tvorbu nových metód využívajúcich združené (konjugované) smery vzhľadom na nejakú kladne definitnú maticu.

Združené smery a ich vlastnosti

Definícia:
Nech $G = G^T$ je kladne definitná matica typu $n \times n$. Smery $s_1, s_2, \ldots, s_m$ sa nazývajú G-**združené** (resp. G-konjugované, G-ortogonálne), ak pre každé $i \neq j$ platí

$$s_i^T G s_j = 0.$$

Tvrdenie 1: G-združené smery sú lineárne nezávislé.

Tvrdenie 2: Nech $G = G^T$ je kladne definitná matica typu $n \times n$ a vektory $s_1, \ldots, s_n$ tvoria úplný systém G-združených smerov. Potom pre inverznú maticu G^{-1} platí

$$G^{-1} = H = \sum_{i=1}^{n} \frac{s_i s_i^T}{s_i^T G s_i}.$$

Združené smery a ich vlastnosti

Veta:
Nech $G = G^T$ je kladne definitná matica typu $n \times n$, smery $s_0, \ldots, s_{n-1}$ sú G-združené, bod x^* je bodom minima kvadratickej funkcie

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

a x^0 je ľubovoľný štartovací bod iteračného procesu

$$x^{k+1} = x^k + \lambda_k s_k,$$

kde λ_k je **optimálny krok.** Potom

$$x^n = x^*,$$

t. j. iteračný proces určí optimálne riešenie úlohy (U1) s kvadratickou účelovou funkciou **za nanajvýš n krokov.**

Metóda združených gradientov Fletchera-Reevesa

Rekapitulácia klasických metód:

Newtonova metóda - pre kvadratickú funkciu konverguje za 1 krok, výpočet smeru je náročnejší

Cauchyho metóda - pre kvadratickú funkciu konverguje len v limite, výpočet smeru nie je náročný

Cieľ - "vylepšiť Cauchyho smer tak aby sa "kvalitou" priblížil Newtonovskému smeru ale bez dodatočného výpočtu druhých parciálnych derivácii.
Využije sa informácia z predchádzajúcich iterácií - optimalizácia dĺžky kroku sa prevedie v dvoch smeroch (Cauchyho a smer z predchádzajúcej iterácie) - je to **neautonómna metóda**.
Ako dôsledok získame metódu, ktorá generuje **združené smery** a teda pre kvadratickú funkciu konverguje za n krokov.

Metóda združených gradientov Fletchera-Reevesa

Úloha minimalizácie kvadratickej funkcie

Riešme úlohu:

$$\min_x \quad Q(x) = \frac{1}{2}x^T G x + h^T x \qquad (U1)$$

kde $G = G^T$ je kladne definitná matica typu $n \times n$ a $h \in \mathbb{R}^n$.

V ďalšom budeme používať štandardnú symboliku:

$$\begin{array}{rcl} x^0 & - & \text{štartovací bod,} \\ x^k & - & \text{bod } k\text{-tej iterácie,} \\ g_k = g(x^k) = Gx^k + h & - & \text{gradient účelovej funkcie } Q(x). \end{array}$$

Metóda združených gradientov Fletchera-Reevesa

Schéma algoritmu Fletchera-Reevesa:

Vstup:	Matica $G = G^T$, vektor $h \in \mathbb{R}^n$, Štartovací bod iteračného procesu x^0, Tolerančná konštanta ε pre kritérium presnosti, Nastavenie počítadla iterácií: $k = 1$.
Algoritmus:	1) Cauchyovská iterácia: $s_0 = -g_0, x^1 = x^0 + \lambda_0 s_0$ 2) Výpočet gradientu $g_k = Gx^k + h$, ak $\|g_k\| < \varepsilon$ - koniec. 3) Výpočet smeru $s_k = -g_k + \mu_k s_{k-1}$, $\mu_k = \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$ 4) Výpočet novej aproximácie $x^{k+1} = x^k + \lambda_k s^k$, pričom $\lambda_k = -\dfrac{g_k^T s_k}{s_k^T G s_k}$. 5) Opakovanie cyklu. $k = k + 1$ choď na 2).

Metóda združených gradientov Fletchera-Reevesa

Ukážeme, že ak riešime úlohu minimalizácie kvadratickej funkcie algoritmom Fletchera-Reevesa, tak

- dvojica (λ_k, μ_k), ktorej zložky vystupujú v 3. a 4. kroku algoritmu je bodom minima funkcie dvoch premenných

$$\psi_k(\lambda, \mu) = Q\left(x^k + \lambda(-g_k + \mu s_{k-1})\right);$$

- gradienty sú na seba kolmé, t.j.

$$g_k^T g_j = 0, \quad \forall\, k \neq j;$$

- smery sú G-združené, t.j.

$$s_k^T G s_j = 0, \quad \forall\, k \neq j.$$

(Mat. indukcia vzhľadom na k.)

Metóda združených gradientov Fletchera-Reevesa

Ukázali sme vlastnosti:

(1)	$g_{k+1}^T s_k = 0$	(5)	$g_{k+1}^T g_k = 0$
(2)	$g_k^T s_k = -g_k^T g_k$	(6)	$s_k^T G s_{s-1} = 0$
(3)	$\lambda_k = -\frac{g_k^T s_k}{s_k^T G s_k}$	(7)	$\mu_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$
(4)	$g_{k+1}^T s_{k-1} = 0$	(8)	$g_{k+1} = g_k + \lambda_k G s_k$

Metóda združených gradientov Fletchera-Reevesa

Algoritmus Fletchera-Reevesa pre nekvadratickú funkciu:

Vstup:	Účelová funkcia $f(x)$ úlohy (U1). Štartovací bod iteračného procesu x^0. Tolerančná konštanta ε pre kritérium presnosti.
1. krok	Cyklus pre $k = 0, 1, \ldots, n-1$. a) Výpočet gradientu $g_k = g(x^k) = \nabla f(x^k)$. b) Testovanie presnosti - ak $\|g_k\| < \varepsilon$ - koniec. c) Výpočet smeru s_k: Ak $k = 0$, tak $s_0 = -g_0$, inak $s_k = -g_k + \mu_k s_{k-1}$, kde $\mu_k = \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$. d) výpočet novej aproximácie $x^{k+1} = x^k + \lambda_k s_k$, kde λ_k je optimálny krok.
2. krok	Položme $x^0 = x^n$, späť na 1. krok.

Metóda združených gradientov Fletchera-Reevesa

Alternatívne výrazy pre výpočet veličiny μ_k:

1. Fletcher-Reeves (1964):

$$\mu_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}};$$

2. Polak-Ribiére (1969):

$$\mu_k = \frac{y_{k-1}^T g_k}{g_{k-1}^T g_{k-1}}, \quad y_{k-1} = g_k - g_{k-1};$$

3. Sorenson (1969):

$$\mu_k = \frac{y_{k-1}^T g_k}{y_{k-1}^T s_{k-1}}, \quad y_{k-1} = g_k - g_{k-1}.$$