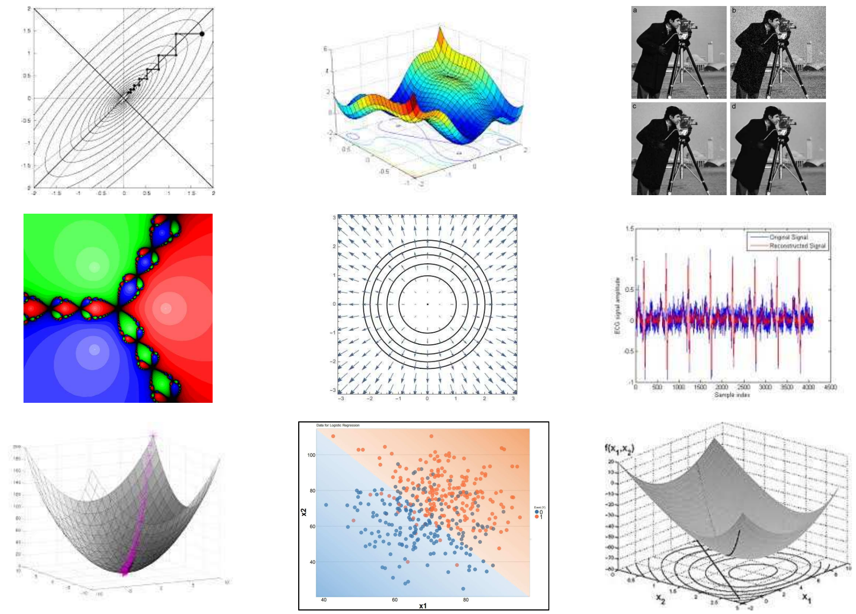
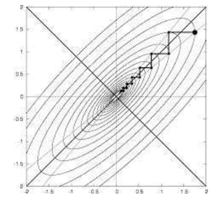
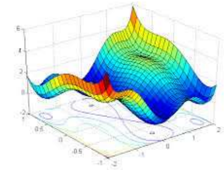
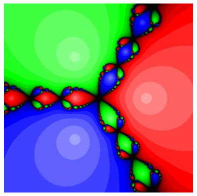
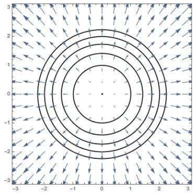
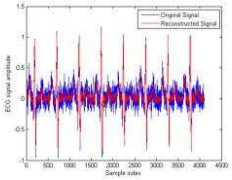
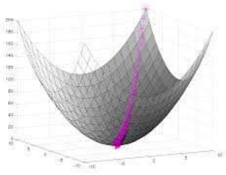
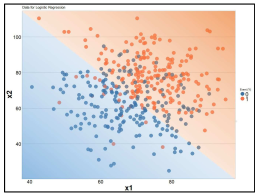
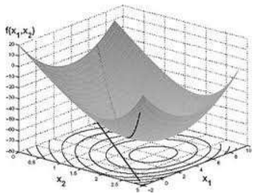


Metódy voľnej optimalizácie - letný semester 2018

M. Trnovská, KAMŠ, FMFI UK

Newtonova metóda

Newtonova metóda

Uvažujme úlohu na voľný extrém

$$\begin{array}{ll} \min & f(x) \\ & x \in \mathbb{R}^n \end{array} \qquad\qquad (U1)$$

kde funkcia $f(x)$ má spojité druhé parciálne derivácie.

Gradient a Hessovu maticu funkcie $f(x)$ budeme označovať

$$g(x) = \nabla f(x), \quad g_k = g(x^k), \quad G(x) = \nabla^2 f(x), \quad G_k = G(x^k).$$

V okolí bodu x^k aproximujeme funkciu $f(x)$ **Taylorovym polynómom II. rádu**

$$f(x) \approx f_k + g_k^T(x - x_k) + \frac{1}{2}(x - x_k)^T G_k (x - x_k) =: Q(x).$$

Newtonova metóda

V **Newtonovej metóde** sa nové približné riešenie $x^{k+1} \in \mathbb{R}^n$ úlohy (U1) definuje ako **minimum kvadratickej funkcie** $Q(x)$.

Preto musíme predpokladať, že matica G_k je kladne definitná a obmedzíme sa na prípad **rýdzo konvexnej** účelovej funkcie $f(x)$.

Derivovaním funkcie $Q(x)$ dostávame

$$\nabla Q(x) = g_k + G_k(x - x_k) = 0$$

a teda

$$(x - x_k) = -G_k^{-1} g_k,$$

resp.

$$x^{k+1} = x^k - G_k^{-1} g_k.$$

Newtonova metóda

Algoritmus Newtonovej metódy

Vstup:	- účelová funkcia $f(x)$ úlohy $(U1)$, - štartovací bod $x^0 \in \mathbb{R}^n$, - tolerančná konštanta $\varepsilon > 0$, - nastavenie počítadla iterácií $k = 0$.
k-ta iterácia:	1) Výpočet gradientu $\nabla f(x^k) = g_k$, ak $\|g_k\| < \varepsilon$, koniec. 2) Výpočet Hessovej matice $\nabla^2 f(x^k) = G_k$ a riešenie sústavy rovníc $G_k p = -g_k$, ktorej riešenie označíme p_k. 3) Výpočet novej iterácie $x^{k+1} = x^k + p_k$, 4) $k = k + 1$, choď na 1).

Newtonova metóda

Smer v Newtonovej metóde je teda **spádový** smer daný predpisom

$$s_k = p_k = -G_k^{-1} g_k,$$

dĺžka kroku je konštantná $\lambda_k = 1$.

Modifikácie Newtonovej metódy:

$$x^{k+1} = x^k + \lambda_k p_k = x^k - \lambda_k G_k^{-1} g_k,$$

kde krok λ_k je optimálny, resp. približne optimálny.

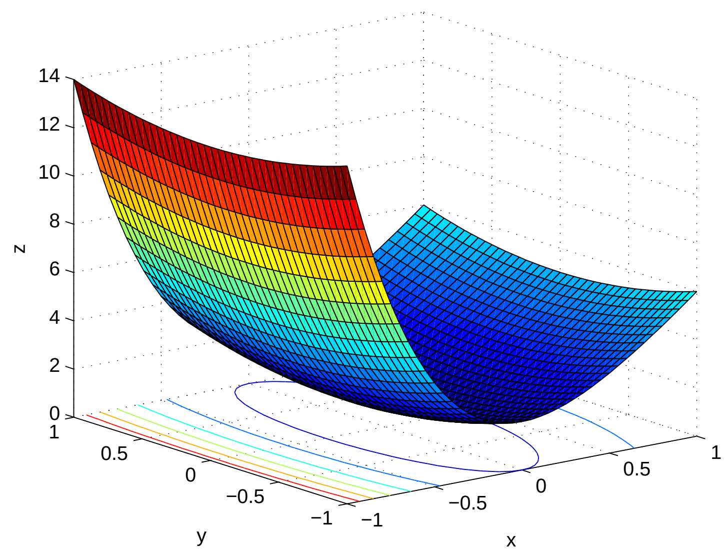
Koľko iterácií potrebuje Newtonova metóda na nájdenie minima kvadratickej funkcie?

Newtonova metóda

Príklad: Uvažujme minimalizáciu konvexnej funkcie

$$f(x_1, x_2) = (x_1 - 1)^4 + x_2^4 + (x_1 - 1)^2 + (x_2 - 2)^2 + 6x_1 + 4x_2 - 5,$$

so štartovacím bodom $[-1, -1]^T$.



Newtonova metóda

Newtonova metóda s presnosťou 10^{-8} skonverguje za 6 iterácií;

Gradientná metóda s približne optimálnou dĺžkou kroku s rovnakou presnosťou skonvergovala za 68 iterácií.

Newtonova metóda

k	x1	x2	f(x1,x2)
0	-1.000000	-1.000000	15.000000
1	-0.400000	-0.571429	2.834753
2	-0.095298	-0.252217	1.134776
3	-0.007069	-0.046449	1.002513
4	-0.000043	-0.000396	1.000000
5	-0.000000	-0.000000	1.000000
6	-0.000000	-0.000000	1.000000

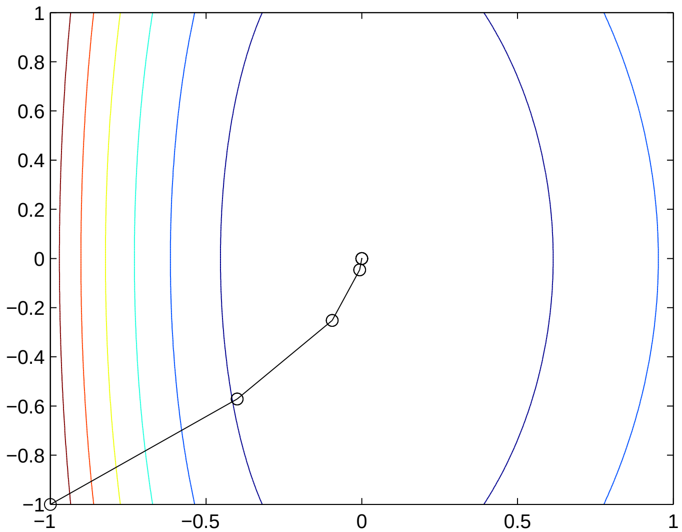
Gradientná metóda

k	x1	x2	f(x1,x2)
0	-1.000000	-1.000000	15.000000
1	0.729440	-0.654112	4.066132
2	-0.174719	-0.246088	1.300179
3	0.158665	-0.181171	1.194779
4	-0.069008	-0.135743	1.053438
5	0.051532	-0.102626	1.028692
6	-0.029661	-0.077970	1.012380
7	0.020448	-0.059400	1.006434
8	-0.012645	-0.045325	1.003186
....			

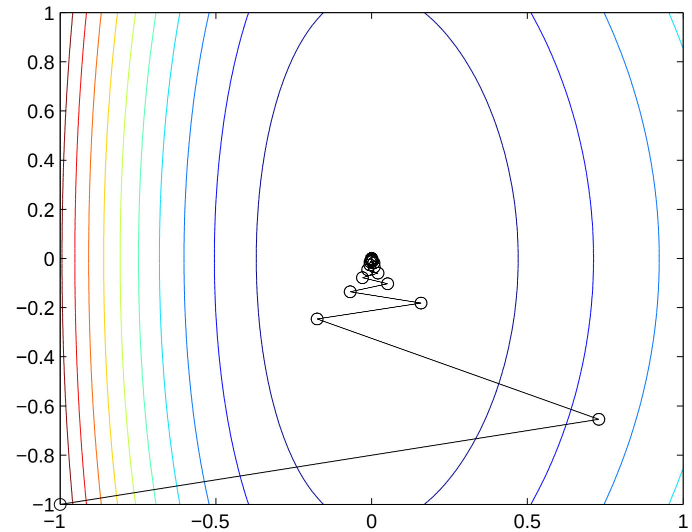
Newtonova metóda

Grafické porovnanie konvergencie metód:

Newtonova metóda



Gradientná metóda



Newtonova metóda

Konvergencia Newtonovej metódy

Ukážeme, že za istých rozumných predpokladov **Newtonova metóda konverguje kvadraticky** k lokálnemu riešeniu $\hat{x}$ úlohy (U1) - t. j. existuje postupnosť bodov $\{x^k\}$ generovaná iteračným predpisom

$$x^{k+1} = x^k - G_k^{-1} g_k,$$

a konštanta $c > 0$ tak, že platí

$$\|x^{k+1} - \hat{x}\| \le c\|x^k - \hat{x}\|^2.$$

V nasledujúcom budeme uvažovať **spektrálnu maticovú normu**

$$\|A\| = \max\{\sqrt{\lambda} \mid \lambda \text{ je vlastná hodnota matice } A^T A\}.$$

a **Euklidovskú vektorovú normu** $\|x\|_2 = \sqrt{x^T x}$.

Newtonova metóda

Konvergenčná veta:

Nech funkcia $f : \mathbb{R}^n \to \mathbb{R}$ spĺňa nasledujúce predpoklady:

1. Funkcia $f(x)$ má **Lipschitzovsky spojitú Hessovu maticu** $G(x)$, t. j. existuje konštanta $L > 0$ tak, že pre každú dvojicu $x, y \in \mathbb{R}^n$ platí

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

2. **Existuje ostré lokálne minimum** $\hat{x}$ funkcie $f(x)$, t. j. bod $\hat{x}$ taký, že $g(\hat{x}) = 0$ a $G(\hat{x}) \succ 0$.

Newtonova metóda

(pokračovanie Konvergenčnej vety)

Potom pre ľubovoľný štartovací bod x^0 **"dostatočne blízko" bodu** $\hat{x}$ možno generovať postupnosť bodov $\{x^k\}$ pomocou iteračného vzorca

$$x^{k+1} = x^k - G_k^{-1} g_k,$$

(teda $\forall k$ existuje G_k^{-1}) a táto postupnosť konverguje k $\hat{x}$ **kvadraticky**, t. j.

$$\|x^{k+1} - \hat{x}\|_2 \leq c \|x^k - \hat{x}\|_2^2.$$

Newtonova metóda

Čo znamená "dostatočne blízko" bodu $\hat{x}$?

Ak označíme $\lambda_1 > 0$ najmenšiu vlastnú hodnotu Hessovej matice $G(\hat{x})$, t. j. platí

$$\|G(\hat{x})^{-1}\| = \lambda_1^{-1}.$$

a označíme

$$c = \frac{3}{2}\frac{L}{\lambda_1},$$

tak **dostatočne blízko znamená z nasledujúceho okolia bodu** $\hat{x}$:

$$\mathcal{O}(\hat{x}, c^{-1}) = \left\{x \mid \|x - \hat{x}\| < \frac{1}{c}\right\},$$

pričom c je zároveň konštanta zo vzťahu

$$\|x^{k+1} - \hat{x}\|_2 \leq c\|x^k - \hat{x}\|_2^2.$$

Newtonova metóda

Lema 1: Nech A je ľubovolná $n \times n$ matica a $z \in \mathbb{R}^n$. Potom

$$|z^T Az| \leq \|A\| \|z\|_2^2.$$

Lema 2: Pre spektrálnu normu symetrickej kladne definitnej matice C s vlastnými hodnotami $0 < \lambda_1 \leq \cdots \leq \lambda_n$ a $\forall\ z \neq 0$ platí

$$\lambda_1 = \frac{1}{\|C^{-1}\|} \leq \frac{z^T Cz}{z^T z} \leq \|C\| = \lambda_n.$$

Newtonova metóda

Lema 3: Nech funkcia $f(x)$ definovaná na $\mathbb{R}^n$ má spojité druhé parciálne derivácie. Potom pre ľubovolné dva body x, y platí

$$g(x) - g(y) = \int_0^1 G\left(y + t(x-y)\right)(x-y)\mathrm{d}t.$$

Lema 4: Nech funkcia $f(x)$ má Lipschitzovsky spojitú Hessovu maticu $G(x)$. Potom

$$\|g(x) - g(y) - G(y)(x-y)\|_2 \leq \frac{1}{2}L\|x-y\|_2^2.$$

A sme pripravení na **Dôkaz konvergenčnej vety**

...

Newtonova metóda

Vlastnosti potrebné v dôkaze konvergenčnej vety:

(V1) Nech C je kladne definitná matica s vlastnými hodnotami $0 < \lambda_1 \leq \cdots \leq \lambda_n$. Potom $\forall\ z : \|z\|_2 = 1$ platí

$$\lambda_1 = \|C^{-1}\|^{-1} \leq z^T C z \leq \|C\| = \lambda_n.$$

(V2) Pre $n \times n$ maticu A a $z \in \mathbb{R}^n$ platí

$$\|Az\|_2 \leq \|A\| \|z\|_2.$$

(V3) Pre $n \times n$ maticu A a $z \in \mathbb{R}^n$ platí

$$|z^T A z| \leq \|A\| \|z\|_2^2.$$

(V4) Nech funkcia $f(x)$ má Lipschitzovsky spojitú Hessovu maticu $G(x)$. Potom

$$\|g(x) - g(y) - G(y)(x-y)\|_2 \leq \tfrac{1}{2} L \|x-y\|_2^2.$$

Newtonova metóda

Newtonova metóda nemusí konvergovať ani v prípade jednoduchých konvexných funkcií.

Príklad. a) Uvažujme funkciu

$$f(x) = |x|^{\frac{3}{2}},\ f'(x) = \frac{3}{2}|x|^{\frac{1}{2}}\mathrm{sgn(x)},\ \mathrm{f''(x)} = \frac{3}{4}|\mathrm{x}|^{-\frac{1}{2}} > 0, \quad \text{pre } \mathrm{x} \neq 0$$

Teda

$$\frac{f'(x)}{f''(x)} = 2|x|\mathrm{sgn(x)} = 2\mathrm{x}.$$

a

$$x^{k+1} = x^k - 2x^k = -x^k.$$

pre ľubovolný štartovací bod $x^0 \neq 0$ Newtonova metóda **osciluje**.

Newtonova metóda

Príklad.

b) Uvažujme funkciu $f(x) = |x|^{\frac{4}{3}}$. Potom

$$\frac{f'(x)}{f''(x)} = 3|x|\mathrm{sgn(x)} = 3\mathrm{x}.$$

Pre ľubovolný štartovací bod $x^0 \neq 0$ Newtonova metóda **diverguje**.

c) Uvažujme funkciu $f(x) = |x|^p,\ p > 2$. Potom

$$\frac{f'(x)}{f''(x)} = \frac{1}{p-1}|x|\mathrm{sgn(x)} = \frac{1}{\mathrm{p}-1}\mathrm{x}.$$

Pre ľubovolný štartovací bod $x^0 \neq 0$ Newtonova metóda **konverguje lineárne**.

Newtonova metóda

Príklad.

Daná je funkcia $f : \mathbb{R}^n \to \mathbb{R}$ a rýdzo konvexná, rastúca funkcia $\phi : \mathbb{R} \to \mathbb{R}$. Definujme zloženú funkciu $F : \mathbb{R}^n \to \mathbb{R}$, $F(x) = \phi(f(x))$.

Newtonovský smer pre funkciu f: $p_k^f := -G_k^{-1} g_k$
Newtonovský smer pre funkciu F: $p_k^F := -[\nabla^2 F(x^k)]^{-1} \nabla F(x^k)$

Dá sa ukázať, že $p_k^f = \alpha_k p_k^F$, kde

$$\alpha_k = \left(1 - \frac{\phi_k'' \gamma_k}{\phi_k' + \phi_k'' \gamma_k}\right), \ \gamma_k = g_k^T G_k^{-1} g_k$$

Newtonova metóda

Afínna invariantnosť Newtonovej metódy

Nech A je regulárna $n \times n$ matica a $b \in \mathbb{R}^n$. Definujme afínnu transformáciu premenných

$$x \mapsto Ax + b := y$$

a funkciu

$$h(x) = f(Ax + b) = f(y).$$

Potom zrejme

$$\nabla h(x) = A^T \nabla f(Ax + b) = A^T \nabla f(y),$$

a

$$\nabla^2 h(x) = A^T \nabla^2 f(Ax + b) A = A^T \nabla^2 f(y) A.$$

Newtonova metóda

Newtonovský krok pre funkciu $h(x)$ je daný:

$$\triangle x = -\left[\nabla^2 h(x)\right]^{-1} \nabla h(x) = -A^{-1}\left[\nabla^2 f(y)\right]^{-1} \nabla f(y) = A^{-1}\triangle y.$$

Nech $\{y^k\}$ je postupnosť bodov daných iteračným predpisom Newtonovej metódy aplikovanej na funkciu f. Ak aplikujeme Newtonovu metódu na funkciu $h(x) = f(Ax+b) = f(y)$, so štartovacím bodom x^0, ktorý spĺňa $y^0 = Ax^0 + b$, tak zrejme aj pre všetky iterácie bude platiť

$$y^k = Ax^k + b,$$

teda postupnosť $\{y^k\}$ vznikne z $\{x^k\}$ po afínnej transformácii súradníc.

Newtonova metóda

Aproximácie Newtonovej metódy

1. Vypočet novej Hessovej matice iba každých $N > 1$ iterácií.

2. Hessova matica sa aproximuje diagonálnou maticou - vypočítajú sa len $\frac{\partial^2 f(x)}{(\partial x_i)^2},\ i = 1, \ldots, n$.

3. **Gauss-Newtonova metóda** - pre funkcie typu $f(x) = \frac{1}{2} \Sigma_{i=1}^m f_i(x)^2$

$$s^k = -\left(\sum_{i=1}^{m} \nabla f_i(x^k) \nabla f_i(x^k)^T\right)^{-1} \left(\sum_{i=1}^{m} f_i(x^k) \nabla f_i(x^k)\right)$$

Newtonova metóda

Newtonova metóda - zhrnutie

• Konvergencia Newtonovej metódy je rýchla a v dostatočne malom okolí optimálneho riešenia je **kvadratická**. Akonáhle sa dosiahne kvadratická fáza konvergencie, je treba len veľmi málo iterácií na dosiahnutie optimálneho riešenia (s veľkou presnosťou).

• Newtonova metóda je **afínne invariantná** a teda nie je citlivá na zmenu súradníc a podmienenosť úlohy.

• Newtonova metóda sa "dobre škáluje" s rozmerom úlohy - pre úlohy na $\mathbb{R}^{10000}$ funguje podobne ako pre úlohy na $\mathbb{R}^{10}$.

• Hlavnou nevýhodou je potreba počítania a ukladania Hessovej matice do pamäte a tiež riešenia systému rovníc pri hľadaní Newtonovského smeru. Avšak v niektorých prípadoch sa dá využiť špeciálna štruktúra Hessovej matice, čo sa využíva najmä v **moderných metódach vnútorného bodu**.

• Vyššie uvedené nedostatky Newtonovej metódy boli motiváciou k vzniku tzv. **kvázinewtonovských metód**.