# 5 Newtonova metóda

**5.1 Bikvadratickú funkcia.** Pre $x \in \mathbb{R}^n$ a symetrické matice $A$ a $Q$ definujeme bikvadratickú funkciu ako

$$f(x) = \frac{1}{4}\left(x^T Q x\right)^2 + \frac{1}{2}x^T A x + b^T x.$$

Vyjadrite gradient $\nabla f(x)$ a Hessovu maticu $\nabla^2 f(x)$.

**5.2 Newtonova metóda pre bikvadratickú funkciu.** V súbore `newton.m` naprogramujte Newtonovu metódu pre bikvadratickú funkciu. Využite vzorce pre gradient a Hessovu maticu z predchádzajúceho príkladu. Otestujte na funkcii s

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad A = \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ -5 \end{pmatrix}$$

a so štartovacím bodom $x^0 = (-2, -5)^T$.

**5.3 Experimenty s podmienenosťou matíc.** Otestujte program z predchádzajúceho príkladu na bikvadratickej funkcii s parametrami

(a) $Q = \begin{pmatrix} 2 & 2 \\ 2 & 1.9 \end{pmatrix}$, $A = \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ -5 \end{pmatrix}$,

(b) $Q = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $A = \begin{pmatrix} 7 & 6.9 \\ 6.9 & 7 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ -5 \end{pmatrix}$,

(c) $Q = \begin{pmatrix} 2 & 2 \\ 2 & 1.9 \end{pmatrix}$, $A = \begin{pmatrix} 7 & 6.9 \\ 6.9 & 7 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ -5 \end{pmatrix}$.

**5.4 Newtonova metóda pre všeobecnú funkciu.** Upravte program z predchádzajúceho príkladu pre všeobecnú konvexnú funkciu s explicitným zadaním potrebných derivácií. Otestujte na niektorej z funkcií

$$\begin{aligned} f_1(x_1, x_2) &= x_2^4 + (x_1 - 2)^2(5 + x_2^2) - 2x_1x_2 + 9x_1 - 5x_2, & x^0 &= (-3, 3)^T, \\ f_2(x_1, x_2) &= e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}, & x^0 &= (-1, 1)^T. \end{aligned}$$

**5.5 Numerické derivácie.** Upravte program z predchádzajúceho príkladu tak, aby miesto parciálnych derivácií používal ich numerické aproximácie

$$\frac{\partial f(x)}{\partial x_i} \approx \frac{f(x + \delta e_i) - f(x - \delta e_i)}{2\delta},$$

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \approx \frac{f(x + \delta e_i + \delta e_j) - f(x + \delta e_i - \delta e_j) - f(x - \delta e_i + \delta e_j) + f(x - \delta e_i - \delta e_j)}{4\delta^2},$$

kde $\delta > 0$ je parameter a $e_i$ je vektor, ktorého $i$-ta zložka je jednotka a ostatné sú nuly. Porovnajte nájdené riešenia s riešeniami z predchádzajúceho príkladu.

**5.6 Zložitejšia funkcia.** Otestujte program z predchádzajúceho príkladu na funkcii

$$f_3(x_1, x_2) = e^{x_1^2 - 2x_2 + 5} + e^{-2x_1 + x_2^2 + 2} + e^{10 \arctan(x_2)}$$

so štartovacím bodom $x^0 = (-2, 2)^T$.

# Pravidlá vektorového derivovania

Vektorovému derivovaniu je venovaný Dodatok B v knihe (strana 273). Tu uvádzame len vybrané pravidlá derivovania.

Nech $f : \mathbb{R}^n \to \mathbb{R}$. Symbolom $\frac{df(x)}{dx}$ označujeme **riadkový** vektor

$$\frac{df(x)}{dx} = \left( \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n} \right)$$

a symbolom $\nabla f(x) = \left(\frac{df(x)}{dx}\right)^T$ označujeme **stĺpcový** vektor gradientu funkcie $f$.

## Derivácia skalárneho súčinu

Nech

$$u : \mathbb{R}^n \to \mathbb{R}^m, \;\; v : \mathbb{R}^n \to \mathbb{R}^m, \;\; f : \mathbb{R}^n \to \mathbb{R},$$

pričom $f(x) = u(x)^T v(x)$. Potom

$$\frac{df(x)}{dx} = \frac{d}{dx}\Big[u(x)^T v(x)\Big] = u(x)^T \frac{dv(x)}{dx} + v(x)^T \frac{du(x)}{dx},$$

$$\nabla f(x) = \left(\frac{df(x)}{dx}\right)^T = \nabla v(x)\; u(x) + \nabla u(x)\; v(x).$$

## Derivácia zloženej funkcie

Nech

$$u : \mathbb{R}^m \to \mathbb{R}, \;\; v : \mathbb{R}^n \to \mathbb{R}^m, \;\; f : \mathbb{R}^n \to \mathbb{R},$$

pričom $f(x) = u(v(x))$. Potom

$$\frac{df(x)}{dx} = \frac{du\big(v(x)\big)}{dx} = \frac{du\big(v(x)\big)}{dv} \frac{dv(x)}{dx},$$

$$\nabla f(x) = \left(\frac{df(x)}{dx}\right)^T = \nabla v(x)\; \nabla u\big(v(x)\big).$$

# Domáca úloha

**5.7 Gradient a Hessova matica.** ***[3 body]*** Vyjadrite prvky gradientu $\nabla J(x)$ a Hessovej matice $\nabla^2 J(x)$ účelovej funkcie

$$J(x) = \sum_{i=1}^{m} (1 - v^i)\, x^T u^i + \ln\left(1 + e^{-x^T u^i}\right).$$

Vykreslite účelovú funkciu $J(x)$ na oblasti $x \in [-100, 0] \times [0, 50]$.

**5.8 Newtonova metóda.** ***[3 body]*** Riešte úlohu

$$\text{Min}\left\{ J(x) = \sum_{i=1}^{m} (1 - v^i)\, x^T u^i + \ln\left(1 + e^{-x^T u^i}\right) \;\middle|\; x \in \mathbb{R}^2 \right\}.$$

pomocou Newtonovej metódy. Ako kritérium optimality použite $\left\|\nabla J(x^k)\right\| \le \varepsilon = 10^{-5}$. Vo vrstevnicovom grafe vykreslite iteračnú postupnosť bodov $x^k$.

**5.9 Približná Newtonova metóda.** ***[4 body]*** Upravte Newtonovu metódu z predchádzajúceho príkladu tak, aby sa Hessova matica vyčísľovala len v každej $N$-tej iterácii. V iteráciách medzitým sa použije posledná známa Hessova matica. Experimentujte s hodnotou $N$.
Pri výpočte Newtonovského smeru $p$, t.j. pri riešení sústavy $G_k p = g_k$

(a) počítajte priamo s Hessovou maticou $G_k$,

(b) využite Choleského rozklad $G_k = LL^T$, kde $L$ je dolná trojuholníková matica.

Pre nejaké vhodné $N > 1$ vykreslite vo vrstevnicovom grafe iteračnú postupnosť bodov $x^k$ približnej Newtonovej metódy spolu s iteračnou postupnosťou klasickej Newtonovej metódy. Čo pozorujete?

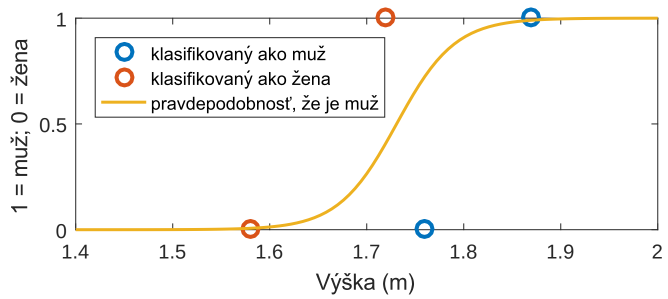
**5.10 Porovnanie.** ***[2 body]*** V tabuľke porovnajte pre rôzne verzie Newtonovej metódy nájdené riešenie, počet potrebných iterácií a trvanie výpočtu. Zhodnoťte výsledky porovnania. Zhoduje sa nájdené riešenie s riešením, ktoré ste našli v prvej domácej úlohe?

**5.11 Grafické porovnanie konvergencie.** ***[2 body]*** Označme $J^*$ nájdené $\varepsilon$-presné riešenie. Pre rôzne verzie Newtonovej metódy vykreslite do jedného obrázku grafy znázorňujúce vývoj hodnoty $J\left(x^k\right) - J^*$ s rastúcim číslom iterácie $k$. Na osi $y$ použite logaritmickú mierku.

**5.12 Binárna klasifikácia.** Použite model odhadnutý na pôvodných dátach *data.csv* na binárnu klasifikáciu, t.j. ak pravdepodobnosť toho, že študent je muž, je aspoň 50 %, tak ho klasifikujeme ako muža, inak ako ženu. Na dátach zo súboru *data2.csv* (majú rovnakú štruktúru ako pôvodné dáta *data.csv*) otestujte, ako tento model klasifikuje pohlavie študenta na základe výšky.

(a) *[1 bod]* Zistite, akú časť študentov klasifikuje model správne, koľko mužov klasifikuje ako ženy a koľko žien klasifikuje ako mužov.

(b) *[2 body]* Vykreslite obrázok, ktorý graficky ukáže úspešnosť klasifikácie. Pre jednotlivých študentov zakreslite ich výšku a skutočné pohlavie; klasifikované pohlavie vyznačte farebne. Okrem toho v obrázku vykreslite aj odhadnutú funkciu pravdepodobnosti Ukážka výstupu so štyrmi študentami:




(c) *[1 bod]* Zistite, aká je *kritická výška*, od ktorej už model klasifikuje študenta ako muža.

(d) *[2 body]* Ako dopadne binárna klasifikácia pomocou modelu z druhej domácej úlohy?