1. *Měkké maximum* je funkce $f(\mathbf{x}) = \ln(e^{x_1} + \cdots + e^{x_n})$, kde $\mathbf{x} \in \mathbb{R}^n$.
   (a) Spočtěte derivaci a Hessovu matici funkce f.
   (b) Pro $n = 2$ rozhodněte, zda je funkce f konvexní.

Řešení:

Derivaci spočítáme pomocí řetízkového pravidla. Pišme $f(\mathbf{x}) = g(h(\mathbf{x}))$, kde

$$h(\mathbf{x}) = e^{x_1} + \cdots + e^{x_n}$$

a $g(y) = \ln y$. Platí

$$f'(\mathbf{x}) = g'(h(\mathbf{x}))h'(\mathbf{x}).$$

Dostaneme $g'(y) = \frac{1}{y}$ a $\frac{\partial h}{\partial x_i} = e^{x_i}$, tedy $h'(\mathbf{x}) = [e^{x_1} \dots e^{x_n}] \in \mathbb{R}^{1\times n}$. Tak dostaneme

$$f'(\mathbf{x}) = \frac{1}{e^{x_1} + \cdots + e^{x_n}}[e^{x_1} \dots e^{x_n}].$$

Označme si pro jednoduchost

$$f_i \coloneqq \frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{e^{x_1} + \cdots + e^{x_n}}.$$

Druhé parciální derivace jsou

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial f_i}{\partial x_j} = \begin{cases} -\frac{e^{x_i}e^{x_j}}{(e^{x_1}+\cdots+e^{x_n})^2} & i \neq j, \\ \frac{e^{x_i}}{e^{x_1}+\cdots+e^{x_n}} - \frac{e^{2x_i}}{(e^{x_1}+\cdots+e^{x_n})^2} & i = j. \end{cases}$$

To lze dále zjednodušit, pokud zavedeme značení

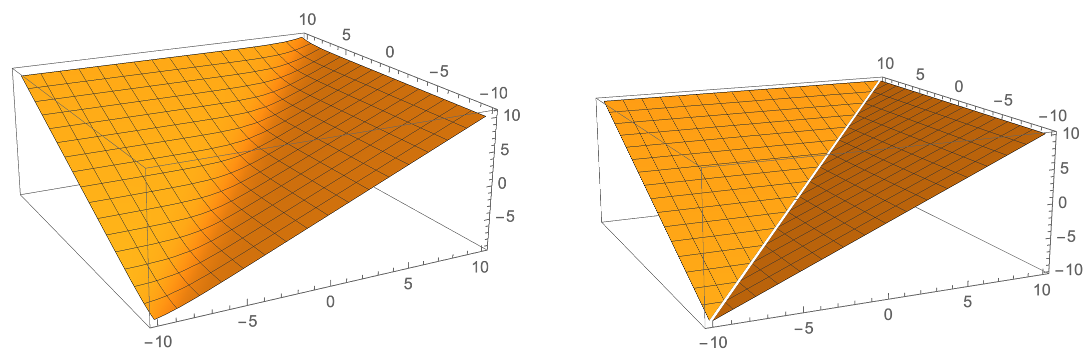
$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Potom totiž platí

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = f_i\delta_{ij} - f_i f_j, \qquad i, j = 1, \dots, n,$$

což jsou právě složky Hessovy matice $f''(\mathbf{x}) \in \mathbb{R}^{n\times n}$.

Předpokládejme, že $n = 2$. Na obrázku jsou funkce f a funkce $\max(x_1, x_2)$:



Ukážeme, že $f''(\mathbf{x})$ je pozitivně semidefinitní matice, což implikuje, že funkce f je konvexní. Chceme tedy ověřit, že platí $\mathbf{x}^\mathsf{T} f''(\mathbf{x})\mathbf{x} \geq 0$ pro každé $\mathbf{x} \in \mathbb{R}^2$. Platí

$$\mathbf{x}^\mathsf{T} f''(\mathbf{x})\mathbf{x} = [x_1\ x_2] \begin{bmatrix} f_1 - f_1^2 & -f_1 f_2 \\ -f_1 f_2 & f_2 - f_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = f_1 x_1^2 + f_2 x_2^2 - (f_1 x_1 + f_2 x_2)^2.$$

Protože platí $f_1 + f_2 = 1$, $f_1, f_2 > 0$ a funkce $\varphi(z) = z^2$ je konvexní, dostaneme

$$f_1 x_1^2 + f_2 x_2^2 \geq (f_1 x_1 + f_2 x_2)^2.$$

Tudíž $\mathbf{x}^\mathsf{T} f''(\mathbf{x})\mathbf{x} \geq 0$.

2. Hledáme neznámé parametry $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ nelineárního regresního modelu

$$p(t) = \frac{x_1 t}{x_2 + t}, \qquad t \in \mathbb{R},$$

na základě dat $(t_1, y_1), \ldots, (t_7, y_7) \in \mathbb{R}^2$.

(a) Formulujte odpovídající úlohu nelineárních nejmenších čtverců.
(b) Jak vypadá krok Gauss-Newtonovy (GN) metody?
(c) Jak vypadá krok Levenberg-Marquardtovy (LM) metody?
(d) Jak vypadá krok Newtonovy metody?

Řešení:

(a) Definujeme nejprve jednotlivá rezidua modelu pro $i = 1, \ldots, 7$,

$$g_i(\mathbf{x}) = y_i - p(t_i) = y_i - \frac{x_1 t}{x_2 + t}$$

a dále klademe $\mathbf{g} = (g_1, \ldots, g_7)\colon\ \mathbb{R}^2 \to \mathbb{R}^7$. Úloha nelineárních nejmenších čtverců je

$$\text{Minimalizuj} \qquad f(\mathbf{x}) \coloneqq \|\mathbf{g}(\mathbf{x})\|^2 = \sum_{i=1}^{7} g_i(\mathbf{x})^2, \qquad \mathbf{x} \in \mathbb{R}^2.$$

(b) Zobrazení $\mathbf{g}$ v okolí bodu $\mathbf{x}_k$ aproximujeme jeho Taylorovým rozvojem,

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k),$$

kde Jacobiho matice (derivace) zobrazení $\mathbf{g}$ v bodě $\mathbf{x}$ je matice $\mathbf{g}'(x) \in \mathbb{R}^{7\times 2}$, která má v i-tém řádku parciální derivace

$$\frac{\partial g_i}{\partial x_1} = -\frac{t_i}{x_2 + t_i}, \qquad \frac{\partial g_i}{\partial x_2} = \frac{x_1 t_i}{(x_2 + t_i)^2}.$$

Tedy minimalizujeme

$$\|\mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)\|^2 = \|\mathbf{g}'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) - (-\mathbf{g}(\mathbf{x}_k))\|^2$$

v proměnné $\mathbf{x}$. To je ovšem úloha lineárních nejmenších čtverců, jejíž řešení odpovídají řešením soustavy normálních rovnic

$$\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}'(\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k) = -\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}(\mathbf{x}_k).$$

Předpokládejme dále, že má matice $\mathbf{g}'(x)$ plnou sloupcovou hodnost. Tudíž můžeme vyjádřit

$$\mathbf{x}-\mathbf{x}_k = -\left[\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}'(\mathbf{x}_k)\right]^{-1}\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}(\mathbf{x}_k),$$

z čehož plyne iterace GN metody tvaru

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}'(\mathbf{x}_k)\right]^{-1}\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}(\mathbf{x}_k).$$

(c) Krok LM metody je

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}'(\mathbf{x}_k)+\mu_k\mathbf{I}\right]^{-1}\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}(\mathbf{x}_k),$$

kde $\mu_k > 0$ je regularizační parametr. Pro malá μ_k tak dostaneme přibližně GN iteraci, což může pomoci urychlit konvergenci v blízkém okolí hledaného optima. Pro velká μ_k je naopak krok LM iterace přibližně iterací gradientní metody, neboť derivace funkce f je $f'(\mathbf{x}) = 2\mathbf{g}(\mathbf{x})^\mathsf{T}\mathbf{g}'(\mathbf{x})$ a platí,

$$\mathbf{x}_{k+1} \approx \mathbf{x}_k - \mu_k^{-1}\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - \frac{1}{2}\mu_k^{-1}f'(\mathbf{x}_k)^\mathsf{T}.$$

To umožnuje docílit robustnosti (spolehlivosti pro počáteční odhady daleko od optima).

(d) Newtonova metoda použitá na minimalizaci funkce f vyžaduje Hessovu matici funkce f, kterou lze vyjádřit jako

$$f''(\mathbf{x}) = 2\mathbf{g}'(\mathbf{x})^\mathsf{T}\mathbf{g}'(\mathbf{x}) + 2\sum_{i=1}^{7} g_i(\mathbf{x})g_i''(\mathbf{x}).$$

Iterace Newtonovy metody je

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1}f'(\mathbf{x}_k)^\mathsf{T}.$$

Newtonova metoda konverguje kvadraticky k lokálnímu minimu v jeho blízkém okolí, ovšem vyžaduje navíc výpočet Hessovy matice a není typicky robustní vůči počátečním iteracím daleko od hledaného optima.

3. Spočtěte vzdálenost bodu $\mathbf{z} \in \mathbb{R}^n$ od nadroviny popsané rovnicí $\mathbf{a}^\mathsf{T}\mathbf{x} = b$ v $\mathbb{R}^n$. Využijte výsledku k výpočtu šířky pásu mezi nadrovinami $\mathbf{a}^\mathsf{T}\mathbf{x} - b = -1$ a $\mathbf{a}^\mathsf{T}\mathbf{x} - b = 1$.

Řešení:

Označme nadrovinu $A = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^\mathsf{T}\mathbf{x} = b\}$. Vzdálenost $\mathbf{z}$ od A je podle definice kladnou odmocninou z optimální hodnoty úlohy

$$\min \{\|\mathbf{z} - \mathbf{x}\|^2 \mid \mathbf{x} \in A\}.$$

Volme libovolný vektor $\mathbf{x}_0 \in A$. Potom je množina $A - \mathbf{x}_0$ lineární prostor, neboť obsahuje počátek. Původní úloha je tak ekvivalentní úloze

$$\min \{\|(\mathbf{z} - \mathbf{x}_0) - \mathbf{x}\|^2 \mid \mathbf{x} \in A - \mathbf{x}_0\}.$$

Z lineární algebry víme, že hledaná vzdálenost je délkou ortogonální projekce vektoru $\mathbf{z}-\mathbf{x}_0$ na ortogonální doplněk $(A-\mathbf{x}_0)^\perp$ prostoru $A-\mathbf{x}_0$. Ovšem $(A-\mathbf{x}_0)^\perp$ je přímka se směrovým vektorem $\mathbf{a}$. Proto je hledaná projekce rovna

$$\mathbf{x}^* = \frac{\mathbf{a}^\mathsf{T}(\mathbf{z} - \mathbf{x}_0)}{\|\mathbf{a}\|} \frac{\mathbf{a}}{\|\mathbf{a}\|} = \frac{(\mathbf{a}^\mathsf{T}\mathbf{z} - b)\mathbf{a}}{\|\mathbf{a}\|^2}$$

a její délka

$$\|\mathbf{x}^*\| = \frac{|\mathbf{a}^\mathsf{T}\mathbf{z} - b|}{\|\mathbf{a}\|^2}\|\mathbf{a}\| = \frac{|\mathbf{a}^\mathsf{T}\mathbf{z} - b|}{\|\mathbf{a}\|}.$$

Šířka hledaného pásu odpovídá vzdálenosti libovolného bodu $\mathbf{z}$ splňujícího $\mathbf{a}^\mathsf{T}\mathbf{z} - b = -1$ od nadroviny $\mathbf{a}^\mathsf{T}\mathbf{x} - b = 1$. Víme, že to je číslo

$$\frac{|\mathbf{a}^\mathsf{T}\mathbf{z} - (b+1)|}{\|\mathbf{a}\|} = \frac{|b - 1 - (b+1)|}{\|\mathbf{a}\|} = \frac{2}{\|\mathbf{a}\|}.$$

4. Hledáme tři jednotkové vektory $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^3$ tak, aby byla společná hodnota skalárního součinu $\mathbf{x}_1^\mathsf{T}\mathbf{x}_2 = \mathbf{x}_1^\mathsf{T}\mathbf{x}_3 = \mathbf{x}_2^\mathsf{T}\mathbf{x}_3$ minimální možná.

Řešení:

Povšimněme si, že podobná úloha pro dva jednotkové vektory $\mathbf{x}_1, \mathbf{x}_2$ má triviální řešení. Stačí volit $\mathbf{x}_1$ jako libovolný jednotkový vektor a položit $\mathbf{x}_2 = -\mathbf{x}_1$. Potom $\mathbf{x}_1^\mathsf{T}\mathbf{x}_2 = -1$, což je minimální hodnota, neboť podle Cauchyho–Schwartzovy nerovnosti platí

$$|\mathbf{x}_1^\mathsf{T}\mathbf{x}_2| \le \|\mathbf{x}_1\|\|\mathbf{x}_2\| = 1.$$

Pto tři vektory řešíme tuto úlohu: minimalizuj α za podmínek $\alpha = \mathbf{x}_1^\mathsf{T}\mathbf{x}_2 = \mathbf{x}_1^\mathsf{T}\mathbf{x}_3 = \mathbf{x}_2^\mathsf{T}\mathbf{x}_3$, kde $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^3$ a $\alpha \in \mathbb{R}$. Definujme matici

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^\mathsf{T} \\ \mathbf{x}_2^\mathsf{T} \\ \mathbf{x}_3^\mathsf{T} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix} = \begin{pmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{pmatrix},$$

která je z definice pozitivně semidefinitní. Naopak víme, že každá reálná symetrická matice na pravé straně rovnosti výše už je nutně tvaru součinu na levé straně. Podle kritéria hlavních minorů je to dále ekvivalentní s podmínkami

$$1-\alpha^2 \geq 0 \qquad \text{a} \qquad 2\alpha^3-3\alpha^2+1 \geq 0.$$

Snadno uhodneme, že hodnota $\alpha = 1$ vyhoví oběma nerovnostem a tak lze postupně rozložit

$$2\alpha^3-3\alpha^2+1 = (\alpha-1)^2(2\alpha+1).$$

Tedy druhá nerovnost $(\alpha-1)^2(2\alpha+1) \geq 0$ platí jen pro $-0.5 \leq \alpha \leq 1$ a proto je optimální hodnota $\alpha^* = -0.5$. Ze vztahu

$$-0.5 = \mathbf{x}_1^\mathsf{T}\mathbf{x}_2 = \cos\varphi,$$

dostaneme $\varphi = \frac{2}{3}\pi$. Z toho plyne, že optimálním řešení úlohy jsou např. takové konfigurace tří jednotkových vektorů, které mají poslední souřadnici nulovou a dělí jednotkovou kružnici v rovině $x_3 = 0$ na třetiny.

5. Najděte extrémy funkce

$$f(x,y) = x^2 + 2\,y^2$$

za podmínky $g(x,y) = 0$, kde

$$g(x,y) = x^2 - 2\,x\,y + y^2 - 6\,x + 6\,y + 9\,.$$

Řešení:

Lagrangeova funkce je

$$L(x,y,\lambda) = f(x,y) + \lambda\,g(x,y) = x^2 + 2\,y^2 + \lambda\,(x^2 - 2\,x\,y + y^2 - 6\,x + 6\,y + 9)\,,$$

gradienty

$$\begin{aligned} f'(x.y) &= (2\,x, 4\,y)\,,\\ g'(x.y) &= (2\,x - 2\,y - 6, -2\,x + 2\,y + 6)\,. \end{aligned}$$

Podmínky pro stacionární body Lagrangeovy funkce jsou

$$\begin{aligned} 0 = \mathrm{D}_1 L(x,y,\lambda) &= \frac{\partial}{\partial x} L(x,y,\lambda) = 2\,(1+\lambda)\,x - 2\,\lambda\,y - 6\,\lambda\,,\\ 0 = \mathrm{D}_2 L(x,y,\lambda) &= \frac{\partial}{\partial y} L(x,y,\lambda) = -2\,\lambda\,x + (4 + 2\,\lambda)\,y + 6\,\lambda\,. \end{aligned}$$

Můžeme si všimnout, že součet těchto rovnic je

$$0 = 2\,x + 4\,y\,,$$

tedy $x = -2\,y$; dosazením např. do první rovnice dostaneme

$$\begin{aligned} 0 &= 2\,(1+\lambda)\,x + \lambda\,x - 6\,\lambda\,, \\ x &= \frac{6\,\lambda}{3\,\lambda+2}\,, \\ y &= \frac{-3\,\lambda}{3\,\lambda+2}\,, \end{aligned}$$

pokud jmenovatel je nenulový. Dosazením do podmínky $g(x,y) = 0$ vyjde

$$0 = x - y - 3 = \frac{6\,\lambda + 3\,\lambda - 9\,\lambda - 6}{3\,\lambda+2} = \frac{-6}{3\,\lambda+2}\,,$$

což nemá řešení. Důvodem je, že nulové body funkce g nejsou regulární.

Povšimneme si, že

$$x^2 - 2\,x\,y + y^2 - 6\,x + 6\,y + 9 = (x-y-3)^2\,,$$

takže ekvivalentní je omezující podmínka (zde označená stejně $g(x,y) = 0$) pro

$$g(x,y) = x - y - 3\,.$$

Lagrangeova funkce je

$$L(x,y,\lambda) = f(x,y) + \lambda\,g(x,y) = x^2 + 2\,y^2 + \lambda\,(x-y-3)\,,$$

nový gradient

$$g'(x.y) = (1,-1)\,.$$

Podmínky pro stacionární body Lagrangeovy funkce jsou

$$\begin{aligned} 0 &= \mathrm{D}_1 L(x,y,\lambda) = \frac{\partial}{\partial x} L(x,y,\lambda) = 2\,x + \lambda\,, \\ 0 &= \mathrm{D}_2 L(x,y,\lambda) = \frac{\partial}{\partial y} L(x,y,\lambda) = 4\,y - \lambda\,. \end{aligned}$$

Řešením této soustavy a podmínky $g(x,y) = 0$ dostaneme

$$x = \frac{-\lambda}{2}\,, \qquad y = \frac{\lambda}{4} = \frac{-x}{2}\,,$$

$$0 = x - y - 3 = \frac{3}{2}\,x - 3\,,$$

$$x = 2\,, \qquad y = -1\,.$$

Jedná se o extrém pozitivně definitní kvadratické formy na přímce, tedy globální minimum. Můžeme to ověřit i Hessovou maticí Lagrangeovy funkce

$$L''(x,y,\lambda) = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 0 \end{bmatrix}\,,$$

z níž nás ale zajímá jen blok odpovídající argumentům x, y, což je pozitivně definitní matice

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\,.$$