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 Hessián 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 $h'(\mathbf{x}) = [e^{x_1} \ldots e^{x_n}] \in \mathbb{R}^{1\times n}$. Tak dostaneme

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

Označme si pro jednoduchost

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

Druhé parciální derivace jsou

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

To lze dále zjednodušit, pokud zavedeme značení

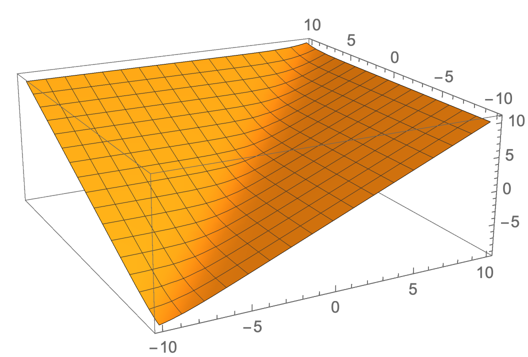
$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

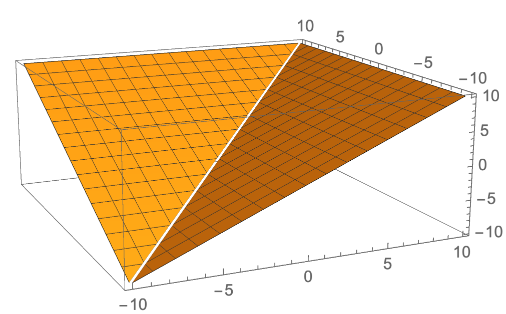
Potom totiž platí

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

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

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 je ekvivalentní konvexitě funkce $f$. 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:**

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

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

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

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

Předpokládejme dále, že má matice $\mathbf{g}'(x)$ plnou sloupcovou hodnost, neboli rank $\mathbf{g}'(\mathbf{x}) = 2$. Potom je iterace GN metody tvaru

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

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 $\mu_k$ tak dostaneme přibližně GN iteraci, což může pomoci urychlit konvergenci v blízkém okolí hledaného optima. Pro velká $\mu_k$ 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,

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

To nám umožnuje docílit robustnosti (spolehlivosti pro počáteční odhady daleko od optima). Newtonova metoda použitá na minimalizaci funkce $f$ vyžaduje Hessián funkce $f$, který 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 Hessiánu a není typicky robustní vůči počátečním iteracím daleko od hledaného optima.