Komentář k oddílu XI.2 – extrémy konkávních funkcí

O charakterizacích konkávních funkcí:

- Konvexní a konkávní funkce byly pro funkce jedné proměnné definovány v oddílu IV.8 a pro funkce více proměnných v oddílu V.8.

  Konvexní a konkávní funkce vždy uvažujeme definované na konvexní množině, jen v tom případě totiž definice dává smysl.

- Konvexní a konkávní funkce třídy C^1 lze charakterizovat pomocí parciálních derivací.

  Pro funkce jedné proměnné platí, že f je konkávní na otevřeném intervalu, právě když f' je na tomto intervalu nerostoucí (viz Věta IV.38 a následující poznámka).

  Jistou analogií této věty pro funkce více proměnných je Věta V.23.

- Konvexní a konkávní funkce třídy C^2 lze charakterizovat pomocí parciálních derivací druhého řádu.

  Pro funkce jedné proměnné Věta IV.39 říká, že f je konkávní na otevřeném intervalu, právě když $f'' \leq 0$ na tomto intervalu.

  Analogií pro funkce více proměnných je právě Věta XI.3, která tvoří hlavní obsah tohoto oddílu.

Důkaz Věty XI.3:

- Tato věta je, jak bylo řečeno výše, analogií Věty IV.39. Z této věty se také dokáže.

- Příprava:

  Připomeňme tvrzení Poznámky (2) za definicí konkávní funkce v oddílu V.8:

  L1: *Nechť $M \subset \mathbf{R}^n$ je konvexní množina a $f : M \to \mathbf{R}$ je funkce. Pak platí*

$$\begin{aligned} &f \text{ je konkávní na } M \Leftrightarrow \forall \boldsymbol{x}, \boldsymbol{y} \in M : f \text{ je konkávní na úsečce } \boldsymbol{xy} \\ &\Leftrightarrow \forall \boldsymbol{x}, \boldsymbol{y} \in M : \text{ funkce } t \mapsto f(\boldsymbol{x}+t(\boldsymbol{y}-\boldsymbol{x})) \text{ je konkávní na intervalu } \langle 0, 1\rangle. \end{aligned}$$

Pokud pracujeme na otevřené konvexní množině, můžeme tuto charakterizaci modifikovat následovně:

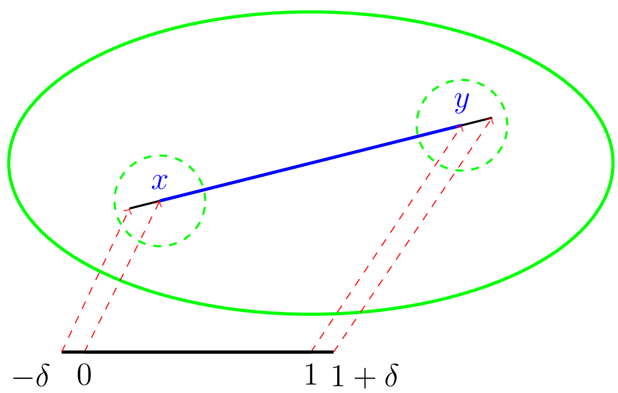
L2: *Nechť $G \subset \mathbf{R}^n$ je otevřená konvexní množina a $f : G \to \mathbf{R}$ je funkce. Pak platí*

$$f \text{ je konkávní na } G \Leftrightarrow \forall \boldsymbol{x}, \boldsymbol{y} \in G\, \exists \delta > 0 : \text{ funkce } t \mapsto f(\boldsymbol{x}+t(\boldsymbol{y}-\boldsymbol{x})) \\ \text{je konkávní na intervalu } (-\delta, 1+\delta).$$

To jsme vlastně již použili v důkazu Věty V.23. Vysvětleme to:

Implikace $\Leftarrow$ je zřejmá – plyne totiž z charakterizace L1 výše. Pokud je totiž uvedená funkce konkávní na na intervalu $(-\delta, 1+\delta)$ pro nějaké $\delta > 0$, je tím spíše konkávní na menším intervalu $\langle 0, 1 \rangle$.

Pro implikaci $\Rightarrow$ připomeňme argument z oddílu V.8:



Zelený ovál ohraničuje množinu G (je otevřená a konvexní). Máme v ní body $\boldsymbol{x}$ a $\boldsymbol{y}$, celá úsečka je obsažena v G (protože je konvexní). Tuto úsečku lze parametrizovat jako

$$\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}), \quad t \in \langle 0, 1 \rangle.$$

Protože G je otevřená, jsou body $\boldsymbol{x}, \boldsymbol{y}$ vnitřní body, tedy existují nějaká jejich okolí, která jsou celá obsažena v G (na obrázku ohraničena zelenými čárkovanými kružnicemi).

Proto, když úsečku $\boldsymbol{xy}$ prodloužíme na přímku, na obou stranách bude ještě kousek ležet v množině G. Tuto prodlouženou úsečku můžeme parametrizovat

$$\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}), \quad t \in (-\delta, 1+\delta)$$

pro vhodné $\delta > 0$. Na tomto intervalu je pak konkávní funkce $t \mapsto f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$.

- Převedení na použití Věty IV.39 pro funkce jedné proměnné:

  Nechť $\boldsymbol{x}, \boldsymbol{y} \in G$. Definujme funkci

  $$\varphi(t) = f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})), \quad t \in (-\delta, 1 + \delta),$$

  kde $\delta > 0$ je zvolené tak, aby úsečka $\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})$, $t \in (-\delta, 1 + \delta)$ byla celá obsažena v G (viz předchozí bod).

  Pak φ je třídy C^2 na intervalu $(-\delta, \delta)$ (podle Věty o derivaci složené funkce, viz Věta V.14).

  Tedy podle Věty IV.39 platí

  $$\varphi \text{ je konkávní na } (-\delta, 1 + \delta) \Leftrightarrow \forall t \in (-\delta, 1 + \delta)\colon \varphi''(t) \leq 0. \qquad (*)$$

- Výpočet φ''.

  Tento výpočet jsme v podstatě už několikrát provedli. V tomto případě vypadá takto (rovnosti platí pro $t \in (-\delta, 1 + \delta)$).

  $$\begin{aligned}
  \varphi(t) &= f(x_1 + t(y_1 - x_1), \dots, x_n + t(y_n - x_n)),\\
  \varphi'(t) &= \sum_{i=1}^{n} \tfrac{\partial f}{\partial x_i}(x_1 + t(y_1 - x_1), \dots, x_n + t(y_n - x_n)) \cdot (y_i - x_i),\\
  \varphi''(t) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \tfrac{\partial^2 f}{\partial x_i \partial x_j}(x_1 + t(y_1 - x_1), \dots, x_n + t(y_n - x_n)) \cdot (y_j - x_j) \cdot (y_i - x_i)\\
  &= (\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}).
  \end{aligned}$$

- Důkaz implikace $\Leftarrow$ z Věty XI.3:

  Předpokládejme, že Hessova matice je v každém bodě množiny G negativně semidefinitní.

  Abychom ukázali, že f je konkávní, použijeme charakterizaci L2. Vezměme tedy $\boldsymbol{x}, \boldsymbol{y} \in G$ a definujme funkci φ jako výše.

  Stačí, když ukážeme, že φ je konkávní na $(-\delta, 1 + \delta)$.

Víme, že pro $t \in (-\delta, 1+\delta)$ platí

$$\varphi''(t) = (\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}) \leq 0,$$

protože matice $\nabla^2 f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$ je negativně semidefinitní.

Tedy podle $(*)$ je φ konkávní a důkaz je hotov.

- Důkaz implikace $\Rightarrow$ z Věty XI.3:

  Předpokládejme, že f je konkávní. Pak podle charakterizace L2 je pro každé $\boldsymbol{x}, \boldsymbol{y} \in G$ příslušná funkce φ konkávní na $(-\delta, 1+\delta)$. Tedy podle $(*)$ je $\varphi'' \leq 0$ na $(-\delta, 1+\delta)$. Protože φ'' jsme spočítali výše, dostáváme

  $$\forall \boldsymbol{x}, \boldsymbol{y} \in G \, \exists \delta > 0 \, \forall t \in (-\delta, 1+\delta) : (\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}) \leq 0.$$

  Speciálně, dosazením $t = 0$, dostaneme

  $$\forall \boldsymbol{x}, \boldsymbol{y} \in G : (\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) \leq 0. \tag{$**$}$$

  Z toho nyní odvoďme, že Hessova matice je v každém bodě negativně semidefinitní.

  Nechť $\boldsymbol{x} \in G$ je libovolné. Protože G je otevřená, existuje $r > 0$, že $B(\boldsymbol{x}, r) \subset G$. Pak pro každé $\boldsymbol{h} \in \mathbf{R}^n \setminus \{\boldsymbol{o}\}$ platí

  $$\begin{aligned} \boldsymbol{h}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{h} &= \tfrac{4}{r}^2 \|\boldsymbol{h}\|^2 \cdot \tfrac{r\boldsymbol{h}^T}{2\|h\|} \nabla^2 f(\boldsymbol{x}) \tfrac{r\boldsymbol{h}}{2\|h\|} \\ &= \tfrac{4}{r}^2 \|\boldsymbol{h}\|^2 \cdot \left(\underbrace{\boldsymbol{x} + \tfrac{r\boldsymbol{h}}{2\|h\|}}_{\in B(\boldsymbol{x}, r) \subset G} - \boldsymbol{x} \right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{x} + \tfrac{r\boldsymbol{h}}{2\|h\|} - \boldsymbol{x} \right) \overset{(**)}{\leq} 0 \end{aligned}$$

  Protože $\boldsymbol{o}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{o} = 0$, dostáváme, že $\nabla^2 f(\boldsymbol{x})$ je negativně semidefintní a důkaz je hotov.

- Alternativní důkaz implikace $\Rightarrow$:

  Ukažme si ještě krátký poněkud trikový důkaz implikace $\Rightarrow$:

  Nechť f je konkávní funkce třídy C^2 na otevřené konvexní množině G.

  Nechť $\boldsymbol{x} \in G$ je libovolný prvek.

Definujme pomocnou funkci $g : G \to \mathbf{R}$ předpisem

$$g(\boldsymbol{y}) = f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}) \cdot \boldsymbol{y}.$$

Pak g je také konkávní funkce na G, protože funkce $\boldsymbol{y} \mapsto \nabla f(\boldsymbol{x}) \cdot \boldsymbol{y}$ je lineární zobrazení $\mathbf{R}^n$ do $\mathbf{R}$, a tedy je zároveň konvexní i konkávní.

Navíc g je také třídy C^2 na G (lineární zobrazení jsou třídy C^∞).

Dále lze snadno spočítat, že

$$\nabla g(\boldsymbol{y}) = \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}), \quad \boldsymbol{y} \in G,$$

speciálně

$$\nabla g(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x}) = \boldsymbol{o},$$

tedy g nabývá v bodě $\boldsymbol{x}$ maxima na G (viz důsledek Věty V.23).

Protože G je otevřená, je to i lokální maximum, a tedy $\nabla^2 g(\boldsymbol{x})$ je negativně semidefinitní podle Věty XI.1(i).

Protože ovšem parciální derivace druhého řádu lineární funkce jsou nulové, je

$$\nabla^2 g(\boldsymbol{y}) = \nabla^2 f(\boldsymbol{y}), \quad \boldsymbol{y} \in G.$$

Speciálně $\nabla^2 f(\boldsymbol{x}) = \nabla^2 g(\boldsymbol{x})$ je negativně semidefinitní.

- Analogicky se dokážou další tvrzení:

  Nechť f je funkce třídy C^2 na otevřené konvexní množině G. Pak platí:

  - f je konvexní na $G \iff \forall \boldsymbol{x} \in G\colon \nabla^2 f(\boldsymbol{x})$ je pozitivně semidefinitní.
  - $\forall \boldsymbol{x} \in G\colon \nabla^2 f(\boldsymbol{x})$ je negativně definitní $\implies f$ je ryze konkávní na G.
  - $\forall \boldsymbol{x} \in G\colon \nabla^2 f(\boldsymbol{x})$ je pozitivně definitní $\implies f$ je ryze konvexní na G.

K Větě XI.4: Tato věta je důsledkem již známých vět. Z bodu (i) plyne podle Věty XI.3. že f je konkávní na G. Z bodu (ii) a důsledku Věty V.23 pak plyne, že v bodě $\boldsymbol{a}$ je maximum f na G.