Přednáška 23 – nelineární LS & minimalizace

<u>opáčko</u> – nelineární nejmenší čtverce: $\min_{\vec{\theta}} \left\| \left[\frac{1}{\sigma_i} f(x_i, \vec{\theta}) - b_i \right]_{i=1}^{m} \right\|^2$

⇝ zadefinujeme $g_i(x_i, \vec{\theta}) := \frac{1}{\sigma_i} f(x_i, \vec{\theta}) - b_i$

⇝ $\min_{\vec{\theta}} \sqrt{g_1(x, \vec{\theta})^2 + \dots + g_m(x, \vec{\theta})^2}$

⇝ potřebujeme metody pro numerickou minimalizaci

→ chceme řešit $\boxed{\min_{\vec{\theta}} F(\vec{\theta})}$ (to se samozřejmě hodí i v jiných aplikacích)

My si tu víceméně představíme „jen hlavní myšlenky"

- všechny minimalizační metody jsou <u>iterativní</u>
  ⇒ produkují posloupnost $\vec{\theta}_0, \vec{\theta}_1, \dots, \vec{\theta}_N$
- myšlenky „jak získat $\vec{\theta}_n$ z $\vec{\theta}_{n-1}, \vec{\theta}_{n-2}, \dots$" se liší:

Derivative-based

⇝ založené na používání (parciálních) derivací funkce $F(\vec{\theta}_{n-1})$. Často založeno na hledání bodů $\vec{\theta}_{kandidát}$: $[\nabla F](\vec{\theta}_{kandidát}) = \vec{0}$. pomocí nějaké iterační metody

<u>first-order</u>

pouze derivace 1. řádu, tedy pouze gradient F

<u>second-order</u>

derivace 1. & 2. řádu, tedy gradient & Hessián

Derivative-free

⇝ odvozené pouze z vyhodnocování $F(\vec{\theta})$ v několika bodech & iterativním procesu založeném na jejich hodnotách. Používáme interpolace/heuristiky.

téměř všechny základní myšlenky „kopírují" hledání minima fce $\mathbb{R} \to \mathbb{R}$ z prváku

např.:
- $f'(x) < 0 \Rightarrow f$ je v bodě x klesající („ve směru $-f'(x)$")
- nutná podmínka $f'(x_{opt}) = 0$

V praxi můžeme mít i tzv. constrained minim. problémy kde parametry $\vec{\theta}$ jsou uvažovány pouze v nějaké množině, např.
$\vec{\theta} \in \{\|\vec{\theta}\|^2 \leq 1\}$ nebo $\{\max_{i=1,\dots,n} |\theta_i| < 5\}$ nebo $\{\vec{\theta} \in \mathbb{R}^n \mid C \cdot \vec{\theta} = \vec{d}\}$
pro dané $C \in \mathbb{R}^{m \times n}$, $\vec{d} \in \mathbb{R}^n$

Derivative-free metody (příklad: Nelder-Mead)

$\vec{\theta} \in \mathbb{R}^2 \to$ pracujeme s trojúhelníky (určeno 2+1 body v $\mathbb{R}^2$)
$\mathbb{R}^3 \to$ pracujeme se čtyřstěny (určeno 3+1 body v $\mathbb{R}^3$)
$\mathbb{R}^n \to$ pracujeme se „simplexy v $\mathbb{R}^n$":

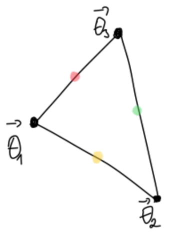
$\vec{x}_1, \dots, \vec{x}_{n+1} \in \mathbb{R}^n \implies S_{x_1,\dots,x_{n+1}} := \left\{ \sum_{i=1}^{n+1} d_i \cdot \vec{x}_i \;\middle|\; \begin{array}{l} d_i \in (0,1) \forall i \\ d_1 + \dots + d_{n+1} = 1 \end{array} \right\}$ je tzv. simplex
& $\vec{x}_2 - \vec{x}_1, \dots, \vec{x}_{n+1} - \vec{x}_1$ lin. nezáv.

Předpokládejme, že máme body $\vec{\theta}_1, \dots, \vec{\theta}_{n+1}$.
Chceme přidat bod $\vec{\theta}_{n+2}$.

} přechod od simplexu $S_{\theta_1,\dots,\theta_{n+1}}$ k novému, lepšímu $S^{(new)}$ výměnou jednoho bodu $\vec{\theta}_1, \dots, \vec{\theta}_{n+1}$

Krok 1: spočíst $F(\vec{\theta}_1), \dots, F(\vec{\theta}_{n+1})$
$\to$ vyměníme $\vec{\theta}_{max} \equiv \vec{\theta}_{\hat{i}} := \operatorname{argmax}_i |F(\vec{\theta}_i)|$
$\to$ zachováme $\vec{\theta}_1, \dots, \vec{\theta}_{\hat{i}-1}, \vec{\theta}_{\hat{i}+1}, \dots, \vec{\theta}_{n+1} \to$ $\underline{n}$ bodů, které určují „protější stěnu"
(tj. protější stěnu simplexu $S_{\theta_1,\dots,\theta_{n+1}}$ k bodu $\vec{\theta}_{\hat{i}}$) $\rightsquigarrow$ označíme $\vec{\theta}_c$ jako centrum této stěny

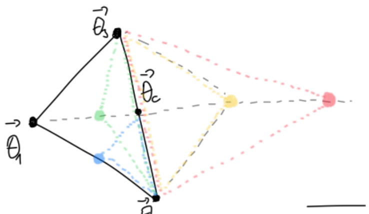
v $\mathbb{R}^2$:



$\hat{i} = \begin{cases} 1 \to \vec{\theta}_c = \\ 2 \to \vec{\theta}_c = \\ 3 \to \vec{\theta}_c = \end{cases}$

Krok 2: za co $\vec{\theta}_{\hat{i}}$ vyměnit?
$\to$ heuristika: $\vec{\theta}_{\hat{i}}$ bylo nejhorší $\rightsquigarrow$ vezmeme $\vec{\theta}_{n+2}$ jako nějaký bod na přímce dané body $\vec{\theta}_{\hat{i}}, \vec{\theta}_c$ (pokud nejsou „všechny horší")

v $\mathbb{R}^2$:



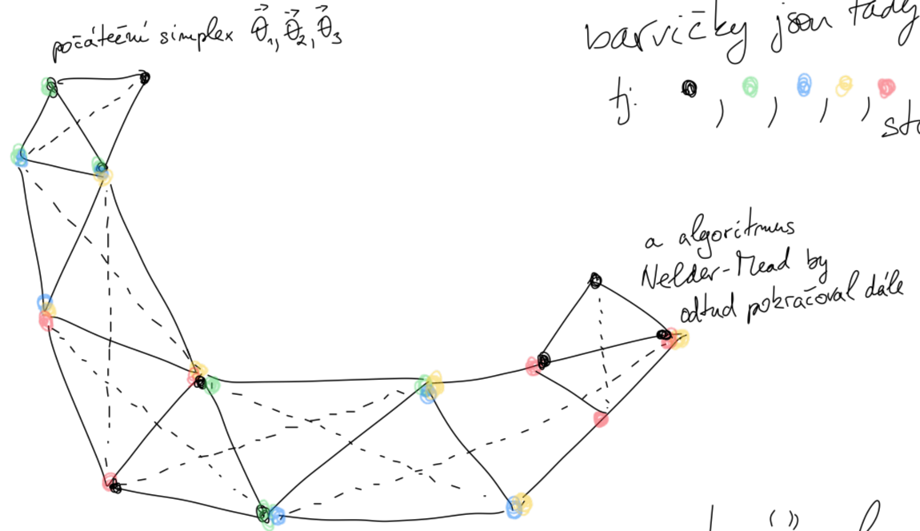
$\vec{\theta}_{n+2} =$
- : pokud $F(\cdot) < F(\vec{\theta}_2)$ nebo $F(\cdot) < F(\vec{\theta}_3)$
- : pokud $F(\cdot) < F(\vec{\theta}_2)$ & $F(\cdot) < F(\vec{\theta}_3)$
- : pokud $F(\vec{\theta}_2), F(\vec{\theta}_3) \leq F(\cdot) < F(\vec{\theta}_1)$
- : pokud $F(\vec{\theta}_1), F(\vec{\theta}_2), F(\vec{\theta}_3) \leq F(\cdot)$ & & $F(\cdot) \geq F(\vec{\theta}_1)$

je reflexe $\vec{\theta}_1$ podél $\overline{\vec{\theta}_2\vec{\theta}_3}$

Obdobné heuristické pokyny jsou dány i pro obecné $\mathbb{R}^n$.

Příklad vývoje v $\mathbb{R}^2$:

barvičky jsou tady „pouze" pořadí iterací, tj: •, •, •, •, • nesouvisí s předchozí stránkou a tím jak z jednoho simplexu dostat následující



Záruky konvergence „skoro neexistují", ale v praxi často dobře funguje (speciálně v nižších dimenzích), je robustní & výpočetně „nenáročný" (1 iterace ≤ 2× vyhodnocení funkce F)

Derivative-based metody

Analýza/Kalkulus: • $F: \mathbb{R}^n \to \mathbb{R}$, $\vec\theta_0 \in \mathbb{R}^n$ $\Longrightarrow$ funkce F klesá nejrychleji ve směru $(-\nabla F)(\theta_0)$

• $\vec\theta_{opt} \in \mathbb{R}^d$ je lokální extrém F & $(\nabla_\theta F)(\vec\theta_{opt})$ existuje $\Rightarrow$ $(\nabla_\theta F)(\vec\theta_{opt}) = \vec 0$

1st order metody (příklad: Steepest descent metody ⇝ založeno na •)

idea: mám $\vec\theta_k$ a chci $\vec\theta_k = \vec\theta_k + \overrightarrow{update}$ ⇝ $\overrightarrow{update}$ = nějaký vektor $\equiv \alpha_k \cdot \vec d_k$

pouze směr, tj. $\|\vec d_k\| = 1$

jak daleko půjdu ve směru $\vec d_k$

→ jak volím $\vec d_k$? Jako směr ve kterém okolo $\vec\theta_k$ klesá F nejrychleji ⇝ podle • $\vec d_k = (-\nabla F)(\vec\theta_k) / \|(\nabla F)(\vec\theta_k)\|$

V praxi buď musí user poskytnout rutinu pro výpočet $\vec\theta \longrightarrow \nabla F(\vec\theta)$ nebo se používají aproximace → $[\nabla F(\vec\theta)]_i = \lim_{h\to 0} \frac{F(\vec\theta + h\vec e_i) - F(\vec\theta)}{h} \approx \frac{F(\vec\theta + h\vec e_i) - F(\vec\theta)}{h}$

→ jak volím α_k? Záleží na konkrétním typu metody. Nalezení optimálního α_k už je samo o sobě pouze 1D minimalizace ($\arg\min_{\alpha\in\mathbb{R}} F(\vec\theta_k + \alpha_k \vec d_k)$) → tzv. line-search. V praxi používáme spoustu různých heuristik (Armijo rule, trust region, ...). Ale v principu lze použít i např. Nelder-Mead (v $\mathbb{R}^1$) na nalezení optimálního α_k

2^{nd} order metody (příklad: quasi-Newton metody → založeno na ●)

idea: chci najít body $\vec{\theta}^{(1)}_{kand}, \vec{\theta}^{(2)}_{kand}, \dots, \vec{\theta}^{(d)}_{kand}$ pro které $\nabla F(\vec{\theta}^{(i)}_{kand}) = \vec{0}$,

tj: chci vyřešit (nelineární) soustavu rovnic → Newtonova metoda.

⇒ budu potřebovat „aproximovat moji nelineární funkci $\vec{\nabla F}$ pomocí lineární → Jakobián funkce $\vec{\nabla F}$

⇝ tzv. Hessián funkce F:

$$\vec{\theta} \longmapsto \begin{bmatrix} \frac{\partial^2 F_1}{\partial \theta_1^2}(\vec{\theta}) & \cdots & \frac{\partial^2 F}{\partial \theta_1 \partial \theta_n}(\vec{\theta}) \\ \vdots & & \vdots \\ \frac{\partial^2 F_n}{\partial \theta_n \partial \theta_1}(\vec{\theta}) & \cdots & \frac{\partial^2 F}{\partial \theta_n^2}(\vec{\theta}) \end{bmatrix} =: H_{\vec{\theta}}$$

→ z Newtonovy metody:

$$\vec{\theta}_{k+1} = \vec{\theta}_k - \vec{V}_k \quad \text{kde} \quad [H_{\vec{\theta}_k}]\vec{V}_k = \nabla_{\vec{\theta}} \vec{F}(\vec{\theta}_k)$$

→ Jak získat $\nabla F(\vec{\theta}_k)$ & $[H_{\vec{\theta}_k}]$? ⇝ buď nám user poskytne routiny $\vec{\theta} \longmapsto \nabla F(\vec{\theta})$ & $\vec{\theta} \longmapsto [H_{\vec{\theta}}]$ nebo aproximace.

- Aproximace ve stylu „diferencí":

$$(\nabla F(\vec{\theta}))_i = \lim_{h \to 0} \frac{F(\vec{\theta} + h \cdot \vec{e}_i) - F(\vec{\theta})}{h} \approx \frac{F(\vec{\theta} + h \cdot \vec{e}_i) - F(\vec{\theta})}{h}$$

- Aproximace ve stylu „update":

→ tzv. quasi-Newton metody, protože nepočítáme „správný Hessián", ale místo toho položíme

$$[H_{\vec{\theta}_{k+1}}] = [H_{\vec{\theta}_k}] + \text{update}$$

např. rank-1: $\beta_k \cdot \vec{w}_k \vec{w}_k^T$

& volíme $\beta_k, \vec{w}_k$ „optimálně"