Optimalizace

Pokud máme nelineární problém nejmenších čtverců pak hledáme $\vec{\theta}_* \in \mathbb{R}^n$ které řeší minimalizaci $\left\| \left[\frac{1}{\sigma_i} f(x_i, \vec{\theta}) - b_i \right]_{i=1}^m \right\|^2$

tj. řešíme

$$\min_{\vec{\theta} \in \mathcal{C} \subseteq \mathbb{R}^n} F(\vec{\theta}) \quad \text{pro} \quad \begin{array}{l} F(\vec{\theta}) := \sum_{i=1}^m \left(\frac{1}{\sigma_i} f(x_i, \vec{\theta}) - b_i \right)^2 \\ \mathcal{C} \equiv \mathbb{R}^n \quad (\text{tzv. „constraint set"}) \end{array}$$

Úlohy tohoto typu jsou obecně označovány za tzv. <u>optimalizační</u>

(úlohy hledání maxima lze převést na hledání minima změnou znaménka)

Drtivá většina numerických metod pro aproximaci řešení $\vec{\theta}_*$ funguje opět <u>iteračně</u> $\rightarrow$ na základě aproximace $\vec{\theta}_n$ získáme (snad lepší) aproximaci $\vec{\theta}_{n+1}$ – a tedy je zapotřebí nějaký počáteční odhad $\vec{\theta}_0$ k nastartování.

Pokud $\mathcal{C} = \mathbb{R}^n$ mluvíme o „unconstrained optimalizaci"

$\mathcal{C} \subsetneq \mathbb{R}^n$ mluvíme o „constrained optimalizaci"

Zjevně musíme používat „jiné nástroje/metody" podle $\mathcal{C}$

$\rightarrow$ my se budeme bavit pouze o „unconstrained" tj. $\mathcal{C} = \mathbb{R}^n$.

Metody, které si ukážeme zastupují 3 základní kategorie.

optimalizační metody

derivative-free	1^{st} order	2^{nd} order
metoda si vystačí s vyhodnocováním $F(\cdot)$	myšlenka je založena na vyhodnocování $F(\cdot)$ & $\nabla F(\cdot)$	myšlenka je založena na vyhodnocování $F, \nabla F$ & $[H_F]$

Derivative-free

- idea je klasicky založena na „porovnávání hodnot a postupu ve směru nejmenších funkčních hodnot"
- nejdražší operace je většinou vyhodnocování $F(\cdot)$
- konvergenci většinou nelze dokázat, ale v praxi jsou pro $\underline{n}$ malé velmi robustní

<u>Zástupce</u>: Nelder-Mead algoritmus

→ pracujeme s tzv. <u>simplexy</u> a v každém kroku jeden vrchol nahradíme novým – konkrétně ten s nejvyšší funkční hodnotou.

(simplex v $\mathbb{R}^n$ je dán $n+1$ body $\vec{\theta}_0, \dots, \vec{\theta}_n$ a odpovídá všem konvexním kombinacím $\vec{\theta}_0, \dots, \vec{\theta}_n$

simplex v $\mathbb{R}$ ≡ úsečka – dána 1+1 body
simplex v $\mathbb{R}^2$ ≡ trojúhelník – dán 2+1 body
simplex v $\mathbb{R}^3$ ≡ čtyřstěn – dán 3+1 body)

V každém kroku tedy vyměníme 1 bod ≡ změníme simplex

To jak volíme ten nový bod je dáno heuristicky a existuje více variant – my si ilustrujeme jednu konkrétní

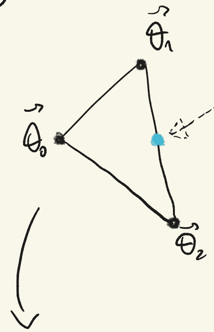
(shoduje se s Wiki)

Nelder-Mead: jeden krok

- seřadíme body $\vec{\theta}_0, \dots, \vec{\theta}_n$ tak, aby platilo $F(\vec{\theta}_0) \le F(\vec{\theta}_1) \le \dots \le F(\vec{\theta}_n)$

→ chceme tedy vyměnit $\vec{\theta}_n$ za nový bod $\theta^{(new)}$

- spočítáme $\vec{\theta}^{(c)}$ (= centroid naproti bodu $\vec{\theta}_n$)



$\vec{\theta}^{(c)} = \frac{1}{2}(\vec{\theta}_1 + \vec{\theta}_2)$

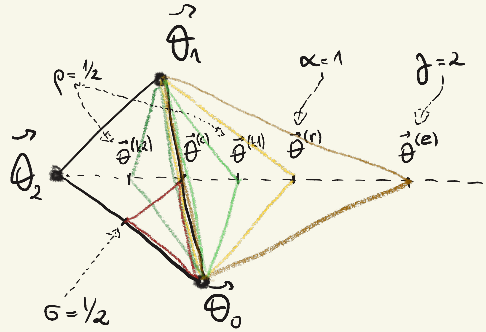
... centrum stěny simplexu, která je „naproti" $\vec{\theta}_n$.

- zkusíme najít $\vec{\theta}^{(new)}$ na přímce $\vec{\theta}_0, \vec{\theta}^{(c)}$ tak, aby $F(\vec{\theta}^{(new)}) < F(\vec{\theta}_n)$:

(i) zkusíme bod $\vec{\theta}^{(r)} := \vec{\theta}^{(c)} + \alpha \cdot (\vec{\theta}^{(c)} - \vec{\theta}_n)$

($\alpha > 0$ je tzv. koeficient reflexe, parametr alg.)

$\boxed{\text{Pokud } F(\vec{\theta}_0) \le F(\vec{\theta}^{(r)}) < F(\vec{\theta}_{n-1}), \text{ pak } \vec{\theta}^{(new)} := \vec{\theta}^{(r)}}$



(ii) pokud $F(\vec{\theta}^{(r)}) < F(\vec{\theta}_0)$, nabýváme dojmu, že $F(\cdot)$ výrazně klesá ve směru $\vec{\theta}^{(c)} - \vec{\theta}_n \equiv \vec{\theta}^{(r)} - \vec{\theta}^{(c)}$ a vydáme se (expandujeme) ještě o trochu dále:

$\vec{\theta}^{(e)} := \vec{\theta}^{(c)} + \gamma \cdot (\vec{\theta}^{(r)} - \vec{\theta}^{(c)})$ ($\gamma > 0$ je tzv. parametr expanze) & $\boxed{\vec{\theta}^{(new)} := \begin{cases} \vec{\theta}^{(e)} & \text{pokud } F(\vec{\theta}^{(e)}) < F(\vec{\theta}^{(r)}) \\ \vec{\theta}^{(r)} & \text{pokud } F(\vec{\theta}^{(e)}) \ge F(\vec{\theta}^{(r)}) \end{cases}}$

(iii) pokud $F(\vec{\theta}^{(r)}) \ge F(\vec{\theta}_{n-1})$, nabýváme dojmu, že jsme se ve směru $\vec{\theta}^{(r)} - \vec{\theta}^{(c)}$ vydali nejspíše moc daleko a tedy délku kroku/posunu zmenšíme (kontrahujeme), podle toho „jak špatný je $\vec{\theta}^{(r)}$":

$\boxed{\text{pokud } F(\vec{\theta}^{(r)}) < F(\vec{\theta}_{n+1}), \text{ pak } \vec{\theta}^{(k)} := \vec{\theta}^{(c)} + \rho \cdot (\vec{\theta}^{(r)} - \vec{\theta}^{(c)})}$

$\boxed{\text{pokud } F(\vec{\theta}^{(r)}) \ge F(\vec{\theta}_{n+1}), \text{ pak } \vec{\theta}^{(k)} := \vec{\theta}^{(c)} + \rho \cdot (\vec{\theta}_{n+1} - \vec{\theta}^{(c)})}$

& $\boxed{\vec{\theta}^{(new)} := \vec{\theta}^{(k)} \text{ pokud } F(\vec{\theta}^{(k)}) < F(\vec{\theta}_{n+1})}$

- pokud jsme $\vec{\theta}^{(new)}$ takto nenašli (tj. dostali jsme se do bodu (iii) a podmínka není splněna) pak nevyměníme pouze $\vec{\theta}_n$, ale rovnou $\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_n$ (= smrskneme simplex)

$\forall i > 0$ předefinujeme $\boxed{\vec{\theta}_i := \vec{\theta}_0 + \sigma \cdot (\vec{\theta}_i - \vec{\theta}_0)}$

Obecně o konvergenci nelze mnoho dokázat, ale v praxi pro malé n ($n \leq 5$) je Nelder-Mead robustní algoritmus. Jeho další předností jsou výpočetní náklady – nejvýše 2 vyhodnocení $F(\cdot)$ během každé iterace.

1st order (zástupce: Steepest descent method ≡ metoda největšího spádu)

Idea je založena na iterativním procesu: mám aproximaci $\vec{\theta}_n$ minima funkce $F(\cdot)$ a chci ji vylepšit $\rightarrow$ $\vec{\theta}_{n+1} := \vec{\theta}_n + \underbrace{\alpha_{n+1}}_{\geq 0} \underbrace{\vec{v}_{n+1}}_{\in \mathbb{R}^n \ \& \ \|\vec{v}_{n+1}\| = 1}$ ≡ posuneme se z $\vec{\theta}_n$ ve směru $\vec{v}_{n+1}$ o délku α_{n+1}.

Jak volit $\vec{v}_{n+1} \in \mathbb{R}^n$ & $\alpha_{n+1} \geq 0$?

<u>Krok 1</u>: předpokládejme, že $\vec{\theta}_n$ je dobrá aproximace, tj. $0 < \alpha_{n+1} \ll 1$.

Pak $F(\vec{\theta}_{n+1}) = F(\vec{\theta}_n + \alpha_{n+1}\vec{v}_{n+1}) = F(\vec{\theta}_n) + \nabla F(\vec{\theta}_n) \cdot \alpha_{n+1}\vec{v}_{n+1} + \mathcal{O}((\alpha_{n+1})^2)$

V jakém směru se máme vydat? Tak abychom minimalizovali $F(\vec{\theta}_{n+1})$.

Člen $\alpha_{n+1} \nabla F(\vec{\theta}_n) \cdot \vec{v}_{n+1}$ je skalární součin 2 vektorů $\rightarrow$ minimální pokud mají tyto vektory stejný směr a opačnou orientaci $\rightarrow$ $v_{n+1} := -\nabla F(\vec{\theta}_n)$

Existuje mnoho variant & adaptací, ale my nepůjdeme do detailů.

tzv. metody největšího spádu (protože F klesá nejrychleji ve směru $-\nabla F$)

<u>Krok 2</u>: jak volit α_{n+1}?

Teoreticky jasné $\rightarrow$ pokud zafixuju $\vec{v}_{n+1}$, chci $\alpha \geq 0$ volit abych minimalizoval $F(\cdot)$.

$$\rightarrow \alpha_{n+1} := \underset{\alpha \geq 0}{\operatorname{argmin}} F(\vec{\theta}_n - \alpha \cdot \nabla F(\vec{\theta}_n))$$ ... tzv. (exact) line-search

≡ součástí optimalizace v $\mathbb{R}^n$ je pak sekvence optimalizací v $\mathbb{R}$

V praxi se dá použít, ale často je považováno za overkill $\rightarrow$ místo toho se používají heuristiky a přibližné aproximace

Vyhodnocení/aproximace $\nabla F(\cdot)$ může být výpočetně náročné („odpovídá $\mathcal{O}(n)$ vyhodnocení $F(\cdot)$")

Ani zde není zaručená konvergence.

2nd order (zástupce: Newtonova metoda & Levenberg-Marquardt)

Idea je založena na pozorování, že pokud $\vec{\theta}_* \in \mathbb{R}^n$ je lokální min. pak jistě $\boxed{\nabla F(\vec{\theta}_*) = \vec{0}}$. Stačí tedy najít takové body, pro které platí ●.
Pokud mám aproximaci $\vec{\theta}_n \approx \vec{\theta}_*$, pak můžu použít Newtonovu metodu pro řešení soustavy rovnic ● s počátečním odhadem $\vec{\theta}_n$:

$$\vec{\theta}_{n+1} \overset{\text{Newton pro }●}{=} \vec{\theta}_n - \vec{v}_{n+1}, \quad \text{kde } \vec{v}_{n+1} \text{ je řešení soustavy rovnic } [H_F(\vec{\theta}_n)]\,\vec{v}_{n+1} = \nabla F(\vec{\theta}_n)$$

$$\text{"} = \vec{\theta}_n - \|[H_F(\vec{\theta}_n)]^{-1} \nabla F(\vec{\theta}_n)\| \cdot \frac{[H_F(\vec{\theta}_n)]^{-1} \nabla F(\vec{\theta}_n)}{\|[H_F(\vec{\theta}_n)]^{-1} \nabla F(\vec{\theta}_n)\|}$$

Hessián ≡ matice 2. parciálních derivací ≡ Jacobiho matice gradientu

Odpovídá lokální aproximaci F kvadratickou formou z Taylora a minimal. této kvadratické formy v každém kroku.

Newtonova metoda s line-search: $\vec{\theta}_{n+1} = \vec{\theta}_n + \alpha_{n+1} \vec{v}_{n+1}$

beru pomocí line-search; beru jako řešení

tj. $\alpha_{n+1} := \operatorname{argmin}_\alpha F(\vec{\theta}_{n+1})$

Opět není obecně zaručená konvergence (viz konvergence Newtona)

Vyžaduje vyhodnocení/aproximaci $[H_F]$ a ∇F ⇒ opět drahé „odpovídá $\mathcal{O}(n^2+n)$ vyhod. $F(\cdot)$"

Ale pokud jsme „blízko" stacionárního bodu, konvergence je kvadratická (viz Newton).

Adaptace „do praxe":

- derivace nahradíme „diferencema - škrtneme limity" → aproximace $\nabla F, [H_F]$ pomocí F
- použijeme $\nabla F(\vec{\theta}_n)$ a/nebo $[H_F(\vec{\theta}_n)]$ pro několik kroků za sebou
- $\nabla F(\vec{\theta}_{n+1})$ a/nebo $[H_F(\vec{\theta}_{n+1})]$ nepočítáme na novo ale jako levný update $\nabla F(\vec{\theta}_n)$ a/nebo $[H_F(\vec{\theta}_n)]$

Metoda Levenberg - Marquardt

„Kompromis mezi metodou největšího spádu & Newtonovou metodou"

$\vec{\theta}_{n+1} := \vec{\theta}_n - \alpha_{n+1} \vec{v}_{n+1}$, kde $\vec{v}_{n+1}$ řeší $\left[H_F(\vec{\theta}_n) + \beta_{n+1} I\right] \vec{v} = \nabla F(\vec{\theta}_n)$

& β_{n+1} je parametr „kompromisu" ⟶

⟶ $\beta_{n+1} = +\infty \Rightarrow \vec{v}_{n+1}$ odpovídá metodě největšího spádu

„ $\beta_{n+1} = 0 \Rightarrow \vec{v}_{n+1}$ odpovídá Newtonovy "

Parametr β volíme adaptivně, podle vývoje F ⟶ pokud jsme daleko od minima, častěji používáme „levnější" největší spád a pokud jsme blízko, chceme využít Newtonovy kvadratické konv.