Úvod do matematické optimalizace

Matematické metody pro ITS (11MAMY)

Jan Přikryl (volně dle M.T. Heathe)

12. přednáška 11MAMY

čtvrtek 14. dubna 2021

verze: 2022-04-11 19:59

Ústav aplikované matematiky
ČVUT v Praze, Fakulta dopravní

Obsah přednášky

Definice (Optimalizační problém)

Mějme dánu funkci $f : \mathbb{R}^n \to \mathbb{R}$ a množinu $\mathcal{S} \subseteq \mathbb{R}^n$. Nalezněte $\mathbf{x}^* \in \mathcal{S}$ takové, že $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pro všechna $\mathbf{x} \in \mathcal{S}$. Hodnotu $\mathbf{x}^*$ nazýváme optimum nebo minimum funkce f.

Stačí uvažovat minimalizaci, protože maximum f je minimum z $-f$.

Účelová (cílová) funkce f je obvykle diferencovatelná a může být lineární i nelinerání

Přípustná množina (často také množina omezení) $\mathcal{S}$ je definována soustavou lineárních nebo nelineárních rovnic a nerovnic. Body $\mathbf{x} \in \mathcal{S}$ nazýváme přípustné body.

Pokud platí $\mathcal{S} = \mathbb{R}^n$, jde o neomezený optimalizační problém. Většinou ale pracujeme s omezeními.

Základní spojitý optimalizační problém lze zapsat jako

$$\min f(\mathbf{x}) \text{ za podmínky } \boldsymbol{g}(\mathbf{x}) = \mathbf{0} \text{ a } \boldsymbol{h}(\mathbf{x}) \leq \mathbf{0},$$

kde $\mathbf{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ a $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^p$. Hodnoty m a p udávají počty omezujících podmínek, omezující podmínky definují $\mathbf{x} \in \mathcal{S}$ (přípustné body).

Definice (Lineární programování)

Funkce f, $\boldsymbol{g}$ a $\boldsymbol{h}$ jsou lineární.

Definice (Nelineární programování)

Alespoň jedna z funkcí f, $\boldsymbol{g}$ a $\boldsymbol{h}$ je nelineární.

Optimalizace – Příklady

- Minimalizujte hmotnost konstrukce při dodržení omezujících podmínek na její pevnost.
- Maximalizujte pevnost konstrukce při dodržení omezujících podmínek na její hmotnost.
- Minimalizujte náklady na potravu za podmínek minimálního příjmu určitých živin.
- Minimalizujte povrch válce daného objemu:

$$\min_{r,h} f(r,h) = 2\pi r(r+h)$$

$$\text{za podmínky } g(r,h) = \pi r^2 h - V = 0,$$

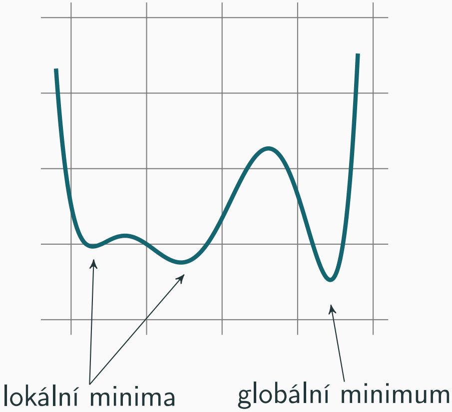
kde r a h jsou poloměr a výška válce a V je jeho požadovaný objem

Definice (Globální minimum)

Bod $\mathbf{x}^* \in \mathcal{S}$ je globálním minimem f, pokud $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pro všechna $\mathbf{x} \in \mathcal{S}$.

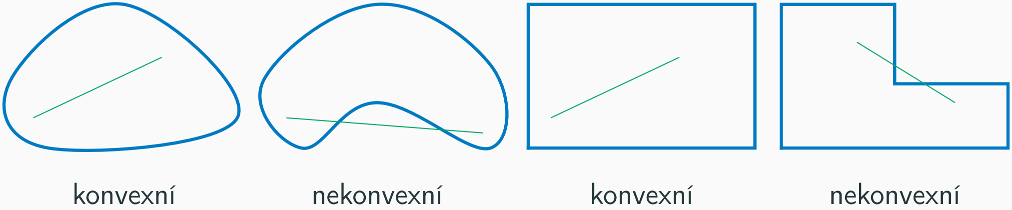
Definice (Lokální minimum)

Bod $\mathbf{x}^* \in \mathcal{S}$ je lokálním minimem f, pokud $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pro všechna přípustná $\mathbf{x}$ v okolí bodu $\mathbf{x}^*$.



- Najít globální minimum či ověřit jeho existenci je obecně velmi obtížné
- Většina optimizačních metod je navržena na hledání lokálních minim, která mohou, ale nemusí být i globálním minimem
- Pokud hledáme globální minimum, můžeme zkusit několik široce rozprostřených starovacích bodů a ověřit, že konvergují k tomu samému výsledku
- Pro určité problémy (například pro lineární programování) je nalezení globálního optima výpočetně schůdné

Množina $\mathcal{S} \subseteq \mathbb{R}^n$ je konvexní, pokud obsahuje i úsečku spojující libovolné dva body z této množiny.



Funkce $f : \mathcal{S} \subseteq \mathbb{R}^n \to \mathbb{R}$ je konvexní na konvexní množině $\mathcal{S}$, pokud její graf mezi dvěma libovolnými hodnotami z $\mathcal{S}$ leží na nebo pod úsečkou, spojující funkční hodnoty na koncových bodech úsečky.

Jakékoliv lokální minimum konvexní funkce f na konvexní množině $\mathcal{S} \subseteq \mathbb{R}^n$ je zároveň globálním minimem f na $\mathcal{S}$

Jakékoliv lokální minimum striktně konvexní funkce f na konvexní množině $\mathcal{S} \subseteq \mathbb{R}^n$ je zároveň jednoznačným globálním minimem f na $\mathcal{S}$

Podmínky optimality prvního řádu – Derivace, gradient

Pro funkci jedné proměnné hledáme extrémy jako nulové body první derivace

Obdobně u funkcí n proměnných hledáme kritické body, tedy řešení nelineární úlohy

$$\nabla f(\mathbf{x}) = \mathbf{0},$$

kde $\nabla f(\mathbf{x})$ je gradientní vektor (prvky jsou $\partial f(\mathbf{x})/\partial x_i$)

Pokud je $f : \mathcal{S} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ spojitě diferencovatelná, každý vnitřní bod $\mathbf{x}^*$ množiny $\mathcal{S}$, na němž nabývá f lokálního minima, musí být kritickým bodem f

Ne každý kritický bod je ale minimem: může jít i o maxima či sedlové body.

Podmínky optimality druhého řádu

Definice (Hessián)

Hessián funkce $f : \mathbb{R}^n \to \mathbb{R}$ je symetrická $n \times n$ matice

$$\mathbf{H}_f(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right).$$

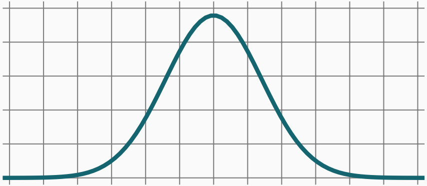
Hessián umožňuje určit typ kritického bodu funkce: Je-li $\mathbf{H}_f(\mathbf{x}^*)$)

- pozitivně definitní, pak $\mathbf{x}^*$ je minimem f,
- negativně definitní, pak $\mathbf{x}^*$ je maximem f,
- indefinitní, pak $\mathbf{x}^*$ je sedlový bod f,
- singulární, pak nastávají různé patologické situace.

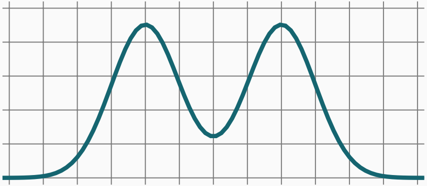
Unimodalita funkce – Co je to „mód“?

Pro minimalizaci funkce jedné proměnné potřebujeme „uzávorkovat“ interval řešení analogicky k uzávorkování intervalu změny znaménka při řešení nelineárních rovnic.

S existiencí maxima funkce spojujeme pojem mód.



unimodální



bimodální

Víme, že $\min f = \max -f$, unimodalitu tedy můžeme definovat i v kontextu minimalizace:

Definice (Unimodalita)

Reálná funkce f je unimodální na intervalu $\langle a, b\rangle$ v případě, že existuje jedinečné $x^* \in \langle a, b\rangle$ takové, že $f(x^*)$ je minimum f na $\langle a, b\rangle$ a f je striktně klesající pro $x \leq x^*$ a striktně rostoucí pro $x \geq x^*$.

Unimodalita umožňuje vyřazení části intervalu na základě vzorkování funkčních hodnot, analogicky s metodou půlení intervalu.

Unimodalita funkce – Praktické využití

Mějme unimodální funkci f na $\langle a, b\rangle$ a v tomto intervalu dva body x_1 a x_2 takové, že $x_1 < x_2$:

- Porovnáním funkčních hodnot $f(x_1)$ a $f(x_2)$ můžeme vyřadit buď interval $\langle a, x_1)$ nebo $(x_2, b\rangle$, minimum bude ležet ve zbylém podintervalu.
- Každý nový pár bodů se ale musí nacházet na relativních pozicích, které jsou vždy totožné vzhledem k relativní délce celého intervalu.
- Při vhodné volbě rozložení bodů stačí pro další iteraci jen jedno vyhodnocení funkce f, podobně jako u metody půlení itervalu.

Obsah přednášky

Hledání zlatým řezem

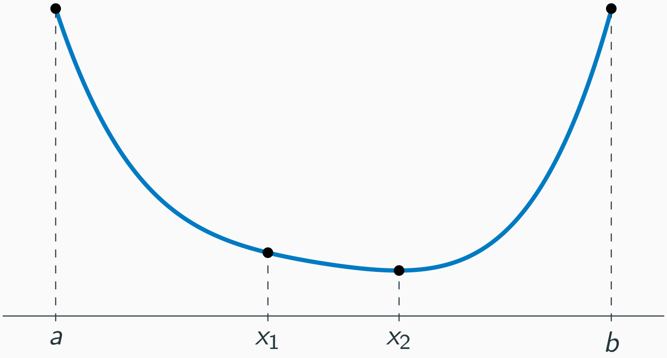
Jednou z voleb je vybírat relativní pozice bodů takové, že body se nacházejí ve vzdálenostech τ a $1-\tau$ od počátku, kde $\tau^2 = 1-\tau$ a tedy $\tau = (\sqrt{5}-1)/2 \approx 0{,}618$ a $1-\tau \approx 0{,}382$

Podintervaly budou mít relativní délku τ původního intervalu a zbylý vnitřní bod bude na pozici buď τ nebo $1-\tau$ vzhledem k délce podintervalu

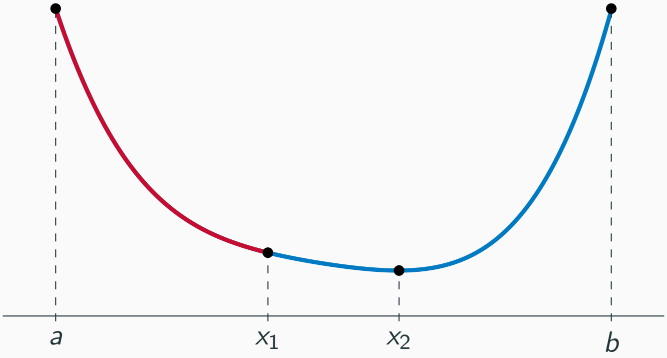
Pro další iteraci tedy stačí vyhodnotit pouze jednu funkční hodnotu f

Hledání zlatým řezem je bezpečný způsob minimalizace, jeho rychlost konvergence je ale pouze lineární s $C \approx 0{,}618$

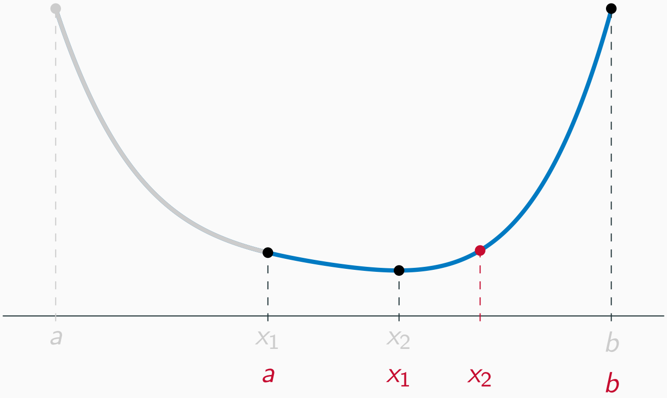
Hledání zlatým řezem



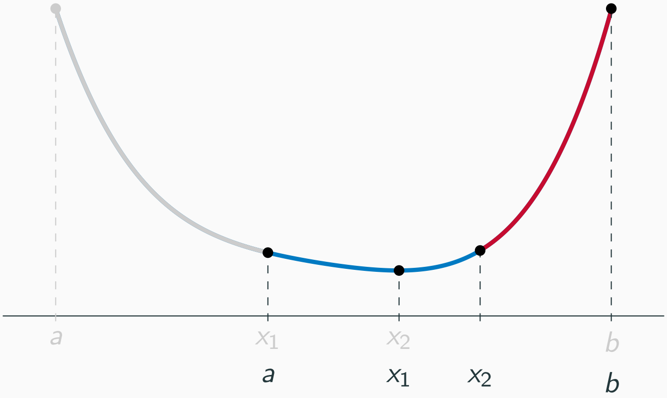
Hledání zlatým řezem



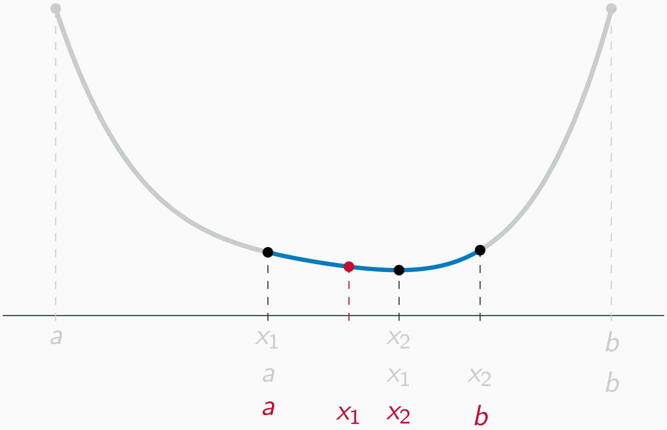
Hledání zlatým řezem



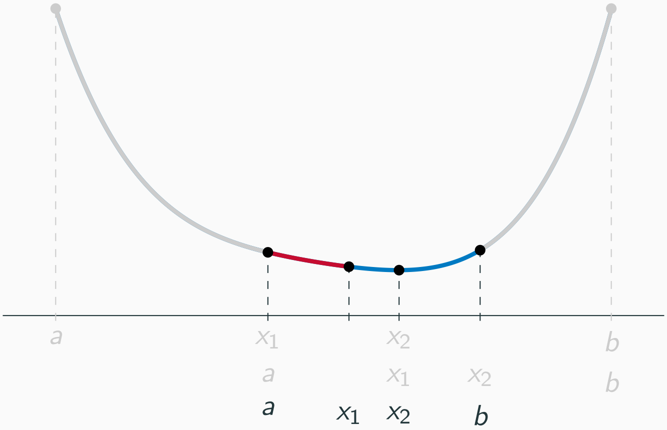
Hledání zlatým řezem



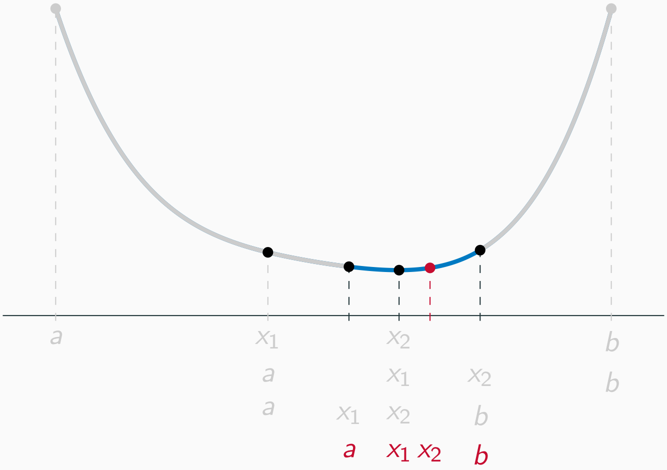
Hledání zlatým řezem



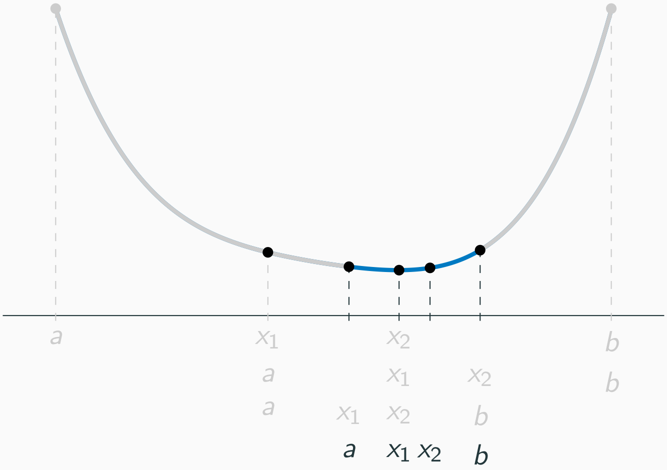
Hledání zlatým řezem



Hledání zlatým řezem



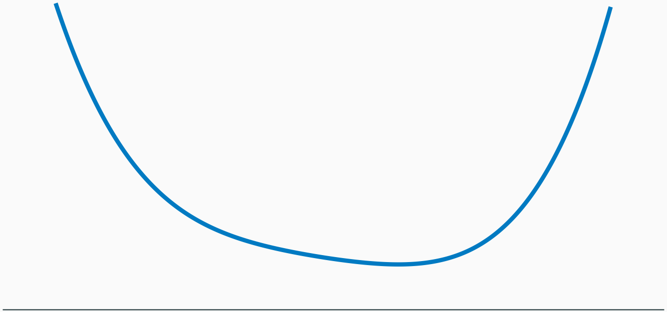
Hledání zlatým řezem



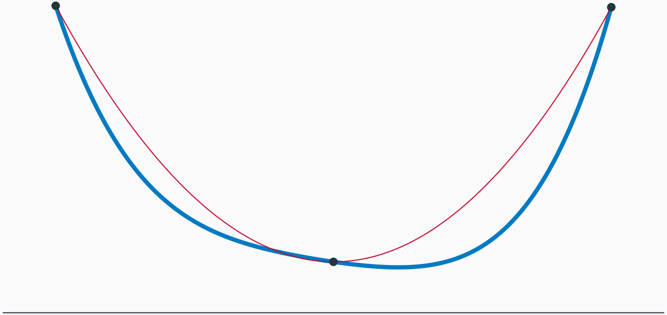
Postupná parabolická interpolace

- Proložíme kvadratický polynom třemi funkčními hodnotami
- Minimum této kvadratické interpolace je aproximací k minima funkce
- Nový bod nahradí nejstarší ze tří předchozích bodů a proces se opakuje až do konvergence
- Míra konvergence postupné parabolické interpolace je superlineární s $r \approx 1{,}324$

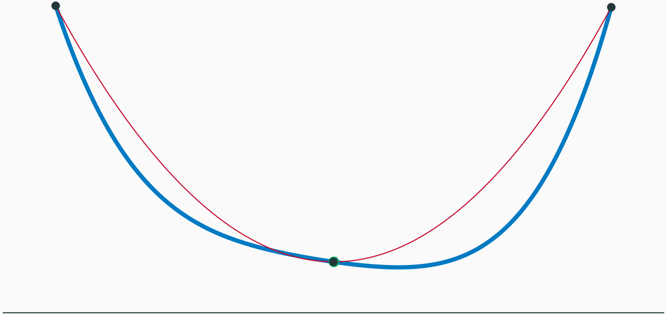
Postupná parabolická interpolace



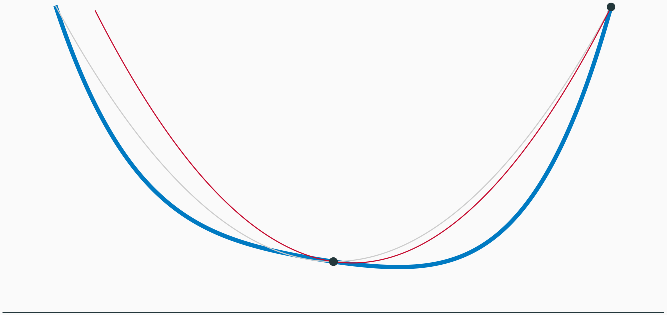
Postupná parabolická interpolace



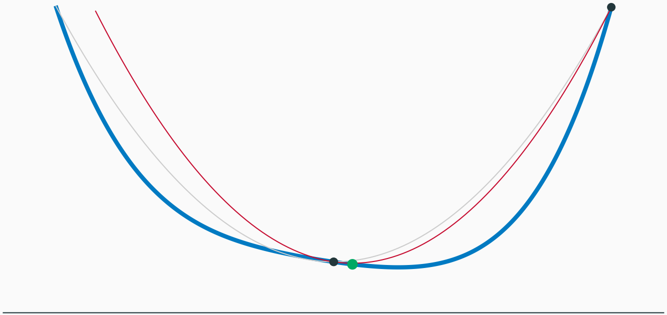
Postupná parabolická interpolace



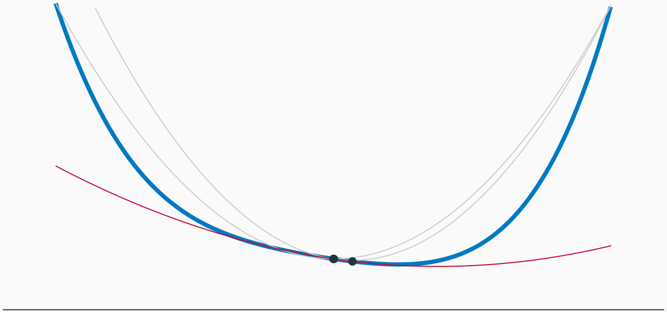
Postupná parabolická interpolace

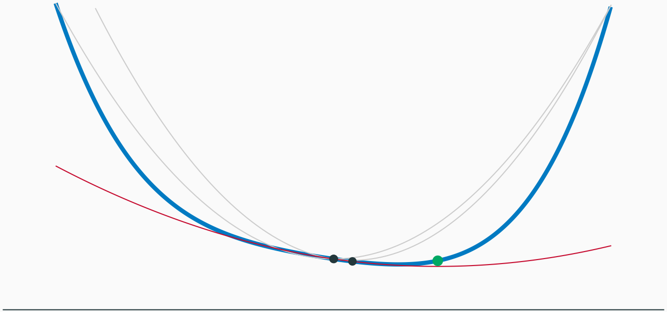


Postupná parabolická interpolace

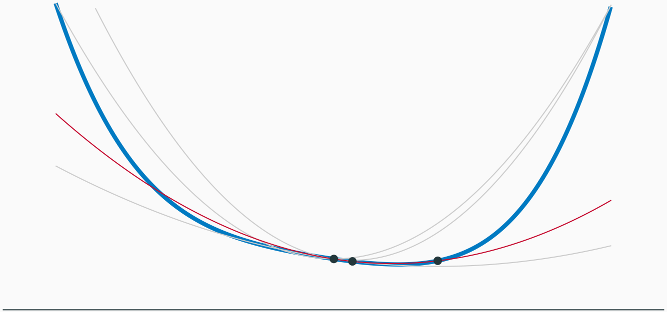


Postupná parabolická interpolace

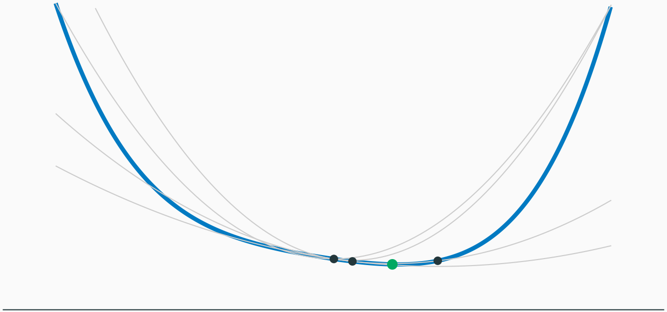




Postupná parabolická interpolace



Postupná parabolická interpolace



Přímé metody vícerozměrné optimalizace

Přímé metody optimalizace využívají pouze informace funkčních hodnotách účelové funkce a to pouze k jejich porovnání. Jde to obdoba hledání minima pomocí zlatého řezu.

Nelder-Meadova metoda pro minimalizaci funkce f o n proměnných začíná s $n+1$ počátečními body $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1}$ a vytváří tak simplex v $\mathbb{R}^n$.

Následuje posun po přímce spojující bod s nejvyšší funkční hodnotou $\mathbf{x}^+ = \arg\max_i f(\mathbf{x}_i)$ s těžištěm o krok délky α. Parametr α zadá uživatel.

Nový bod $\mathbf{x}'$ nahradí původní nejhorší bod $\mathbf{x}^+$ a celý proces se opakuje.

Přímé metody jsou vhodné pro nehladké funkce a pro malá n, ale drahé pro velká n.

Diferenciální (gradientní) metody vyžadují určování hodnot účelové funkce a její první, respektive druhé derivace

Nechť je $f : \mathbb{R}^n \to \mathbb{R}$ reálná funkce n proměnných: V jakémkoliv bodě $\mathbf{x}$ s nenulovým gradientem ukazuje $-\nabla f(\mathbf{x})$ směrem k nižším hodnotám f

Hodnota $-\nabla f(\mathbf{x})$ vlastně udává spádnici: funkční hodnoty f klesají ve směru negativního gradientu rychleji, než jakýmkoliv jiným směrem

Definice (Metoda největšího spádu)

Od nástřelu $\mathbf{x}_0$ pokračujeme iteracemi

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k),$$

kde α_k udává délku kroku získanou pomocí metody spádových směrů (angl. *line-search*).

Definice (Metoda spádových směrů)

Známe-li směr sestupu $-\nabla f(\mathbf{x}_k)$, hodnotu α_k určíme jako

$$\alpha_k = \arg\min_{\alpha} f(\mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k))$$

což je jednorozměrná optimalizační úloha.

Metoda největšího spádu je velmi spolehlivá (dokud je gradient nenulový)

Je ale „krátkozraká“: zkoumá pouze nejbližší okolí bodu $\mathbf{x}_k$, iterace proto často oscilují a metoda konverguje pomalu.

Nejběžnější a jeden z nejdůležitějších optimalizačních postupů:

Definice (Lineární programování)

Pro $\mathbf{c} \in \mathbb{R}^n$ a $\mathbf{x} \in \mathbb{R}^n$ najdi

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \mathbf{c}^\top \mathbf{x}$$

za podmínky

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \qquad \mathbf{A} \in \mathbb{R}^{m \times n}, m < n, \mathbf{b} \in \mathbb{R}^m$$
$$\mathbf{x} \geq \mathbf{o}$$

Přípustná oblast je konvexní polyhedron v $\mathbb{R}^n$ a minimum je v nějakém jeho vrcholu

Řeší se pomocí simplexové metody: procházíme postupně všechny vrcholy, až najdeme minimum

Simplexová metoda je spolehlivá a obvykle efektivní: je schopna řešit problémy s tisíci proměnných, v nejhorším případě ale může vyžadovat dobu exponenciálně úměrnou velikosti řešeného problému

Metody vnitřního bodu vyvinuté pro LP v posledních letech mají v nejhorším případě polynomiální dobu řešení

Tyto metody se pohybují přes vnitřek přípustné oblasti, neomezují se na vyšetřování pouze vrcholů polyhedronu

Ačkoli metody vnitřního bodu mají značný praktický význam, simplexová metoda stále ve standardních balíččích pro LP převládá – její praktická účinnost je vynikající

Příklad (Lineární program)

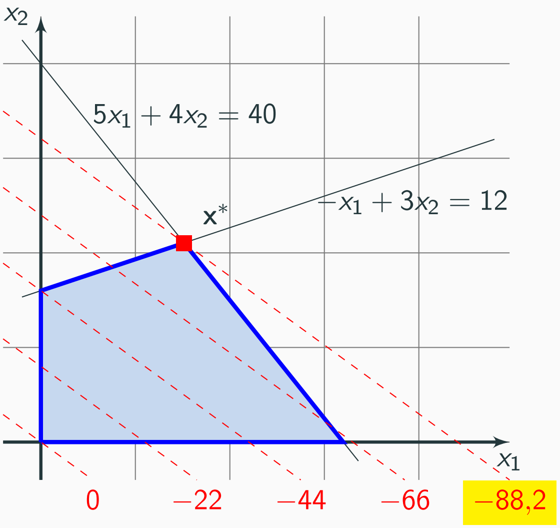
Uvažujme

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x} = -8x_1 - 11x_2$$

za podmínek

$$5x_1 + 4x_2 \leq 40, \quad -x_1 + 3x_2 \leq 12, \quad x_1 \geq 0, x_2 \geq 0$$

Minimum $\mathbf{x}^*$ se musí nacházet v jednom z vrcholů přípustné oblasti, v tomto případě v bodě $x_1 = 3{,}79$, $x_2 = 5{,}26$, v němž má cílová funkce hodnotu $f(\mathbf{x}^*) = -88{,}2$.


x_2
$5x_1 + 4x_2 = 40$
$-x_1 + 3x_2 = 12$
$\mathbf{x}^*$
x_1
0
−22
−44
−66
−88,2

"You know, we're just not reaching that guy."