Numerické metody a programování

Lekce 8

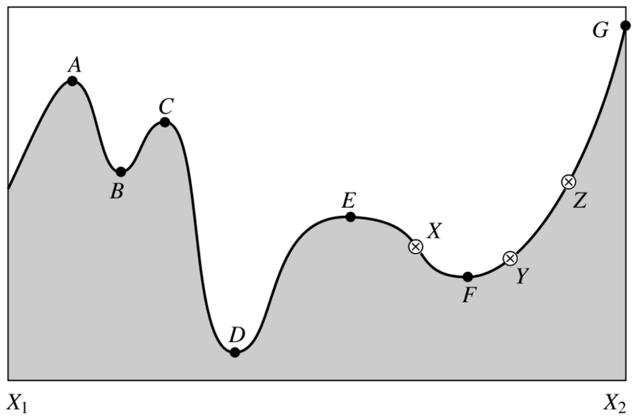
Optimalizace

hledáme bod $\boldsymbol{x}$, ve kterém funkce jedné nebo více proměnných $f(\boldsymbol{x})$ má minimum (maximum)

- maximalizace $f(\boldsymbol{x})$ je totéž jako minimalizace $-f(\boldsymbol{x})$

Minimum funkce

- lokální: minimum na konečném okolí vyjma hranice
- globální: skutečné minimum
  - obtížný problém
  - obvykle se opakuje hledání lokálních minim z různých počátečních bodů

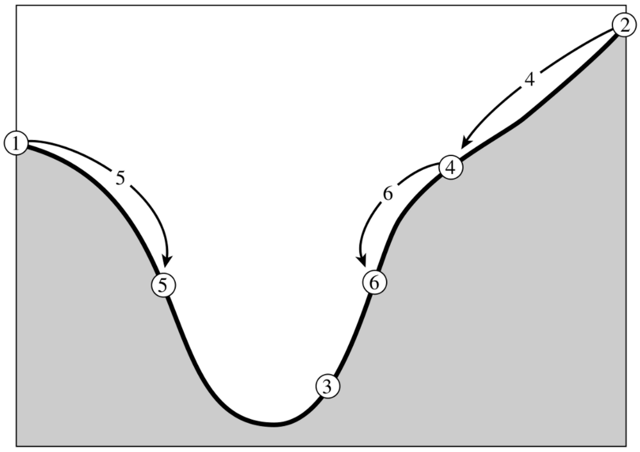


- minimalizace v přítomnosti/nepřítomnosti vazeb
- uzavření minima
  - obdoba uzavření kořene
  - body $a<b<c$, pro které zároveň platí $f(b)<f(a)$ a $f(b)<f(c)$

Metody v 1D nevyžadující derivace

dělení intervalu zlatým řezem

- obdoba bisekce
- na začátku je minimum uzavřené body $a<b<c$
- zvolíme bod x v jednom z intervalů např. $x \in (b,c)$
  - pokud je $f(b) < f(x)$ máme nový interval (a,b,x)
  - pokud je $f(b) > f(x)$ máme (b,x,c)
  - prostřední bod vždy představuje nejlepší odhad polohy minima
- toto dělení opakujeme dokud se interval dostatečně nezmenší

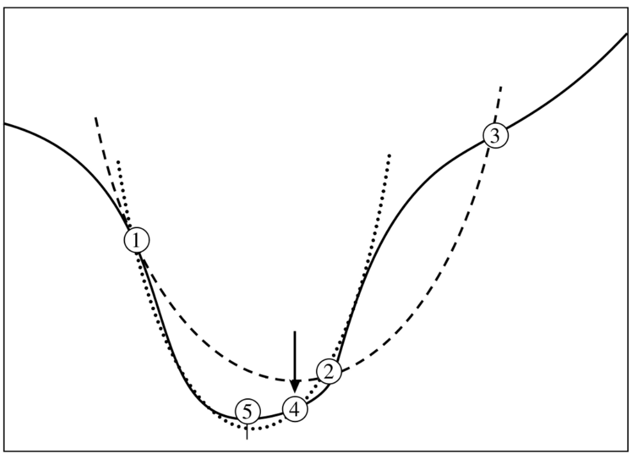


- optimální volba nového bodu x vede na zlatý řez
  - x symetricky položeno vzhledem k b v (a,c)
  - vždy rozdělen větší ze segmentů
  - x volíme ve vzdálenosti $\frac{3-\sqrt{(5)}}{2} \approx 0.38197$ délky delšího ze segmentů, měřeno od bodu b
  - rychlá konvergence ke zlatému řezu z původního poměru segmentů
  - $\epsilon_{n+1} = \frac{\sqrt{5}-1}{2}\,\epsilon_n \approx 0.61803\,\epsilon_n$ (o něco pomalejší než bisekce)

Brentova metoda (parabolická interpolace)

- předpoklad: blízko minima se funkce chová jako parabola
- jeden krok stačí k nalezení minima
- minimum paraboly procházející body a,b,c:

$$x = b - \frac{1}{2}\frac{(b-a)^2[f(b)-f(c)]-(b-c)^2[f(b)-f(a)]}{(b-a)[f(b)-f(c)]-(b-c)[f(b)-f(a)]}$$



- kontrola konvergence: krok je akceptován pokud
  - padne do intervalu uzavírajícího minimum
  - krok se dostatečně rychle zpomaluje
- pokud není krok akceptován $\rightarrow$ zlatý řez

Minimalizace s použitím derivace

- interval (a, b, c) : derivace v bodě b ukazuje, ve kterém segmentu volit další bod
- hledáme kořen derivace
  - derivace v dosavadních dvou nejlepších bodech
  - metoda sečen (superlineární s exponentem 1.618 – zlatý řez!)
- kontrola konvergence, viz. Brentova metoda
- pokud problém s konvergencí $\rightarrow$ bisekce

Minimalizace ve více dimenzích

metoda "downhill simplex"

- simplex: těleso s $N + 1$ vrcholy v dimenzi N
  - trojúhelník v 2D
  - tetrahedron v 3D
- počáteční simplex v okolí bodu $\boldsymbol{P}_0$

$\boldsymbol{P}_i = \boldsymbol{P}_0 + \lambda \boldsymbol{e}_i$

$\boldsymbol{e}_i$ - jednotkové vektory

λ - parametr (velikost)

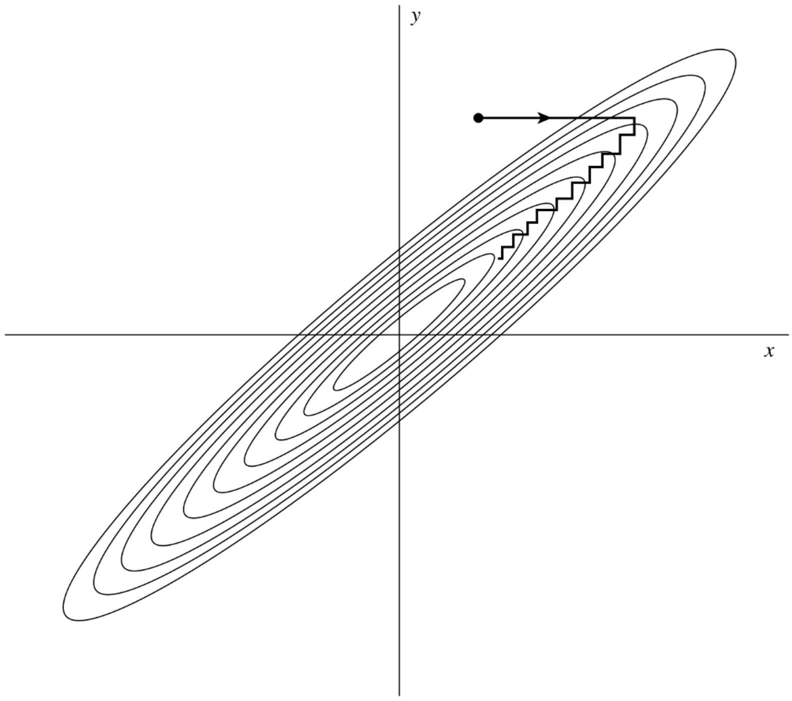
- hledání minima pomocí transformací simplexu
  - zrcadlení nejhoršího bodu přes protilehlou stranu, beze změny objemu
  - expanze simplexu (prodloužení kroku)
  - kontrakce simplexu (např. průchod úzkým místem)
- hledání ukončeno pokud poslední krok byl kratší než tolerance
- kontrola "restartem" na místě minima

Metody založené na opakovaných 1D minimalizacích (Direction set methods)

obecná strategie

- bod $\boldsymbol{P}$ a směr $\boldsymbol{n}$
- hledám minimum na přímce: $\min_{\lambda} f(\boldsymbol{P} + \lambda \boldsymbol{n})$

- multidimenzionální problém: opakovaná minimalizace na přímce
- metody se liší výběrem směrů pro 1D minimalizaci



naivní metoda

- ortonormální vektory $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_N$ definují N směrů hledání
- postupná minimalizace v každém směru, cyklicky se opakuje
- málo efektivní pro některé funkce, viz. obrázek

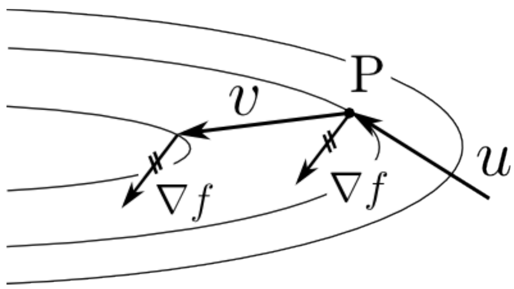
konjugované směry

rozvoj $f(\boldsymbol{x})$ v okolí bodu $\boldsymbol{P}$, počátek souřadnic volíme v bodě $\boldsymbol{P}$

$$f(\boldsymbol{P}+\boldsymbol{x}) \approx f(\boldsymbol{P}) + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x}$$

$\boldsymbol{b} = \nabla f(\boldsymbol{P})$ (gradient)

$\boldsymbol{A}_{ij} = \dfrac{\partial^2 f(\boldsymbol{P})}{\partial x_i \partial x_j}$ (Hessian)



gradient v okolí bodu $\boldsymbol{P}$

$$\nabla f = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{b}$$

změna gradientu posunem v novém směru $\boldsymbol{v}$

$$\delta(\nabla f) = \boldsymbol{A}\,\boldsymbol{v}$$

po posunu požaduji zachování vlastnosti $\nabla f \perp \boldsymbol{u}$, která je v bodě $\boldsymbol{P}$ splněna pro starý směr $\boldsymbol{u}$

$0 = \boldsymbol{u}^{\mathrm{T}}\delta(\nabla f) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{v}$ (vektory $\boldsymbol{u}, \boldsymbol{v}$ jsou konjugované)

kvadratická forma: přesné řešení minimalizací podél N nezávislých konjugovaných směrů

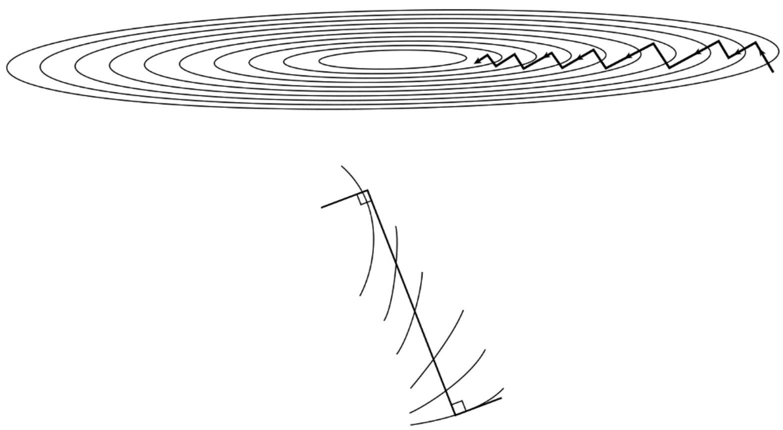
obecná funkce – kvadratická konvergence v blízkosti minima

Powellova metoda

- inicializace směrů: $\boldsymbol{u}_i = \boldsymbol{e}_i$
- počáteční bod $\boldsymbol{P}_0$
- nyní opakujeme
  - postupně N minimalizací ve směrech $\boldsymbol{u}_i$, získáme $\boldsymbol{P}$
  - aktualizace směrů $\boldsymbol{u}_{i+1} \to \boldsymbol{u}_i, \quad i = 1, 2, \ldots, N-1$
  - $\boldsymbol{u}_N = \boldsymbol{P} - \boldsymbol{P}_0$
  - vlastní krok: posun $\boldsymbol{P}$ do minima ve směru $\boldsymbol{u}_N \to \boldsymbol{P}_0$
- každé opakování generuje jeden konjugovaný směr, t.j. N opakování vede k přesné minimalizaci kvadratické formy

Metoda konjugovaného gradientu

- znalost gradientu může zkrátit výpočetní čas
- naivní použití gradientu vede k neefektivním metodám, např. "steepest descent"



- konstrukce konjugovaných směrů pomocí gradientu
- stačí N 1D minimalizací pro přesnou lokalizaci minima kvadratické formy

kvazi-Newtonovy metody

v okolí bodu $\boldsymbol{P}$ jsme měli

$$\nabla f(\boldsymbol{P} + \boldsymbol{x}) = \nabla f(\boldsymbol{P}) + \boldsymbol{A}\boldsymbol{x}$$

hledáme místo, kde gradient vymizí

$$\boldsymbol{0} = \boldsymbol{A}\,\boldsymbol{x} + \nabla f(\boldsymbol{P})$$

tedy

$$\boldsymbol{x} = -\boldsymbol{A}^{-1}\,\nabla f(\boldsymbol{P})$$

neznáme ovšem $\boldsymbol{A}^{-1}$

BGFS (Broyden-Fletcher-Goldfarb-Shanno) algoritmus

- postupná aproximace (aktualizace) $\boldsymbol{A}^{-1}$ v každém kroku na základě funkčních hodnot a gradientů
- daleko od minima – pohyb ve směru poklesu funkce
- blízko minima
  - dobrá aproximace hessianovské matice
  - kvadratická konvergence Newtonovy metody

Lineární programování

obecný problém:

maximalizovat $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ (minimalizovat $-\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$)

s M vazbami $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ (zahrnuje i nerovnosti)

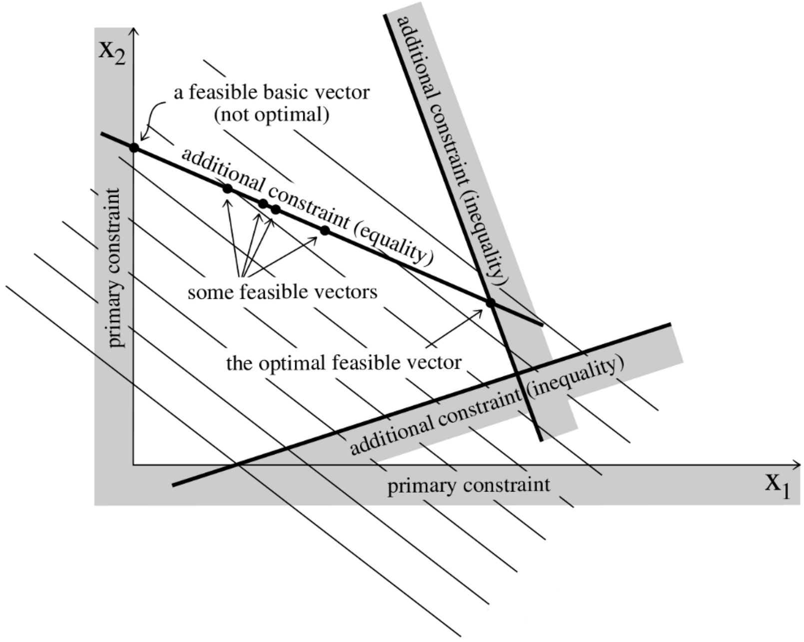
a $\boldsymbol{x} \geq \boldsymbol{0}$ (primární vazby)

kde $x_1, x_2, \cdots x_N$ jsou proměnné a $\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{A}$ jsou známé vektory (matice) koeficientů

- popisuje mnoho problémů ekonomie, inženýrství, plánování, dopravy apod.
- nezápornost je vlastnost množství různých komodit
- lineární vazby zahrnují přirozené požadavky na maximální cenu, existující zdroje apod.

terminologie

- maximalizovaná veličina – *objective function* (cílová funkce)
- vektor $\boldsymbol{x}$ splňující vazby – *feasible vector* (přípustný vektor)
- *feasible vector* maximalizující objective function – *optimal feasible vector*



příklad (podobný problému na obrázku):

maximalizovat $\boldsymbol{x_1 + 2x_2}$

$2x_1 + x_2 \leq 2$

$x_1 - x_2 \leq 1/2$

$2x_1 + 10x_2 = 11$

řešení:

$x_1 = 1/2, \quad x_2 = 1$

přípustný prostor je vymezen vazbami

- rovnost určuje přípustnou hyperplochu
- nerovnost rozděluje N- dimenzionální prostor na přípustnou a nepřípustnou část

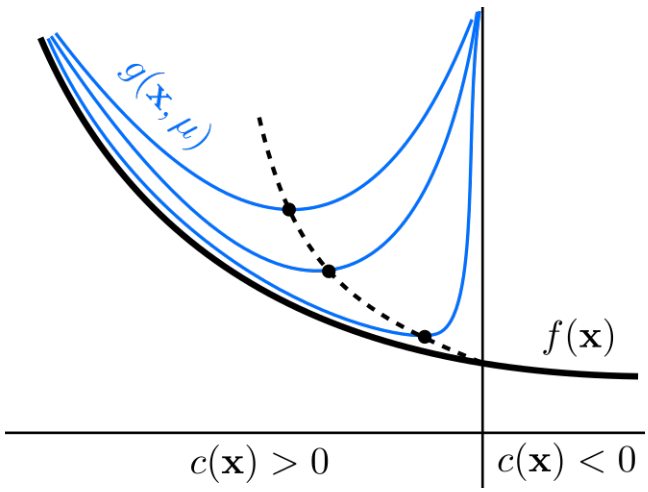
optimalizace linearní cílové funkce

- maximum je na hranici v jednom z vrcholů
- každý vrchol je určen N rovnostmi obsaženými v $M + N$ vazbách (včetně primárních)
  - řešení se redukuje na hledání této podmnožiny
  - pokud $M < N$, je $N - M$ souřadnic vrcholu rovno nula (nutno použít primární vazby)

- numerické algoritmy
  - simplexový algoritmus – prohledávání vrcholů (kombinatorika)
  - metoda vnitřních bodů – iterační postup konvergující k optimálnímu bodu zevnitř přípustné oblasti

Nelineární programování

minimalizovat $f(\boldsymbol{x})$ za podmínek $c_i(\boldsymbol{x}) \geq 0,\ i=1,\ldots,m$



logaritmická bariéra

$g(\boldsymbol{x},\mu) = f(\boldsymbol{x}) - \mu \sum_i \ln[c_i(\boldsymbol{x})]$

- bariéra drží aproximace uvnitř přípustné oblasti
- postupným snižováním výšky bariéry, $\mu \to 0$, konverguje minimum $g(\boldsymbol{x},\mu)$ k minimu $f(\boldsymbol{x})$

Metody simulovaného temperování (simulated annealing)

simulovaný fyzikální proces, kdy systém při chladnutí dosahuje minima energie

- krok je generován náhodně
- někdy jsou akceptovány i kroky opačným směrem (do kopce)
  - možnost uniknout z blízkosti lokálního minima
- postupně je snižována “teplota”
- $T = 0$ odpovídá metodě “downhill simplex” a konvergenci k lokálnímu minimu
- použito na složité optimalizační problémy: obchodní cestující, návrhy složitých integrovaných obvodů atd.