

6. prednáška

Simulované žihanie

Úvodné poznámky

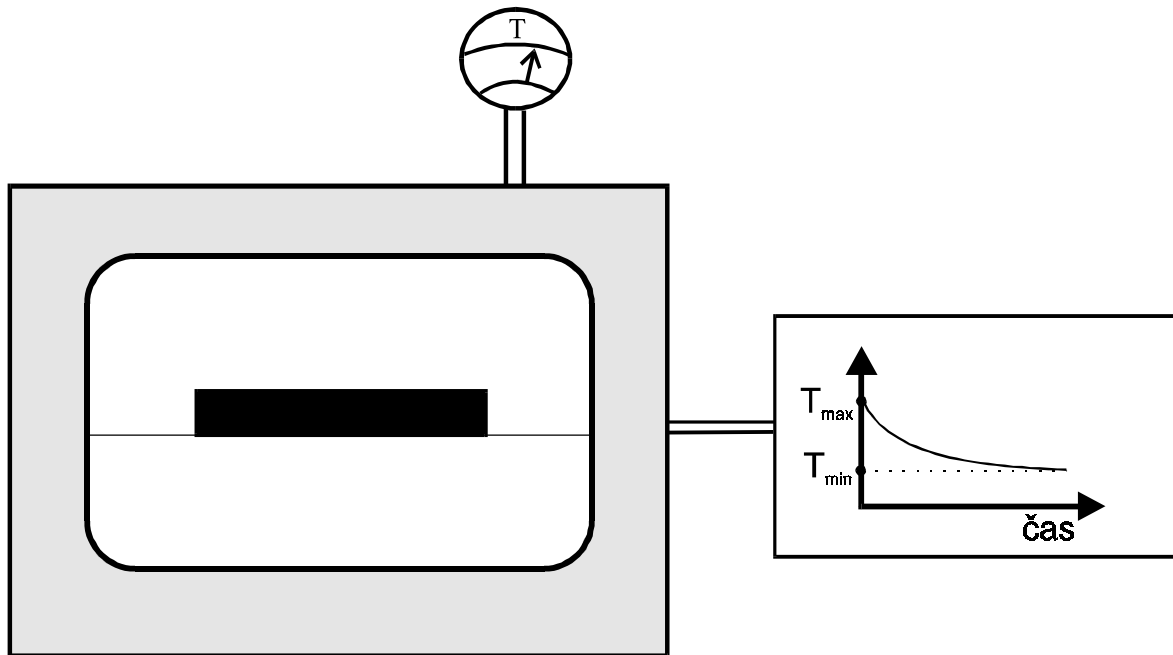
- Metóda simulovaného žihania (Simulated Annealing - SA) patrí medzi tie stochastické optimalizačné algoritmy, ktoré majú základ vo fyzike
- Algoritmus simulovaného žihania je založený na analógii medzi žiháním tuhých telies a optimalizačným problémom.

Pôvod

Počiatkom 80-tych rokov

1. Kirkpatrick, Gelatt a Vecchi [3] (Watson Research Center of the IBM, USA) a nezávisle
2. Doc. Vladimír Černý (Katedra teoretickej fyziky MFF UK v Bratislave)

dostali geniálny nápad, že problém hľadania globálneho minima môže byť realizovaný podobným spôsobom ako žíhanie tuhého telesa.



Znázornenie fyzikálnej realizácie žihania. Teleso sa vloží do pece (ľavý blok), ktorá je vyhriata na vysokú teplotu T_{\max} . Teplota sa programovacím zariadením (pravý blok) pomaly znižuje na teplotu T_{\min} . Týmto spôsobom sa odstránia štruktúrne defekty vyskytujúce sa v telese

Náznak teórie

Predpokladajme, že proces ochladzovania je dostatočne pomalý, potom pre každú teplotu T žíhané teleso je v tepelnej rovnováhe, ktorá je popísaná Boltzmannovským rozdelením pravdepodobnosti toho, že pri teplote T je teleso v stave i s energiou E_i

$$w_T(E_i) = \frac{1}{Q(T)} \exp\left(-\frac{E_i}{kT}\right)$$

kde k je Boltzmannova konštanta a $Q(T)$ je normalizačný faktor nazývaný *partičná funkcia*

$$Q(T) = \sum_i \exp\left(-\frac{E_i}{kT}\right)$$

kde sumácia obsahuje všetky stavy i telesa.

Nech teplota T klesá, potom Boltzmannovská distribúcia uprednostňuje stavy s menšou energiou. V prípade, že teplota sa blíži nule, stav s minimálnou energiou má nenulovú (jednotkovú) pravdepodobnosť.

Ak je ochladzovanie telesa príliš rýchle (telesu nie je umožnené získať tepelnú rovnováhu pre každú teplotu T), defekty v telese môžu zamrznúť za vzniku metastabilných štruktúr, ktoré sú vzdialené od mriežkovej štruktúry s najnižšou energiou.

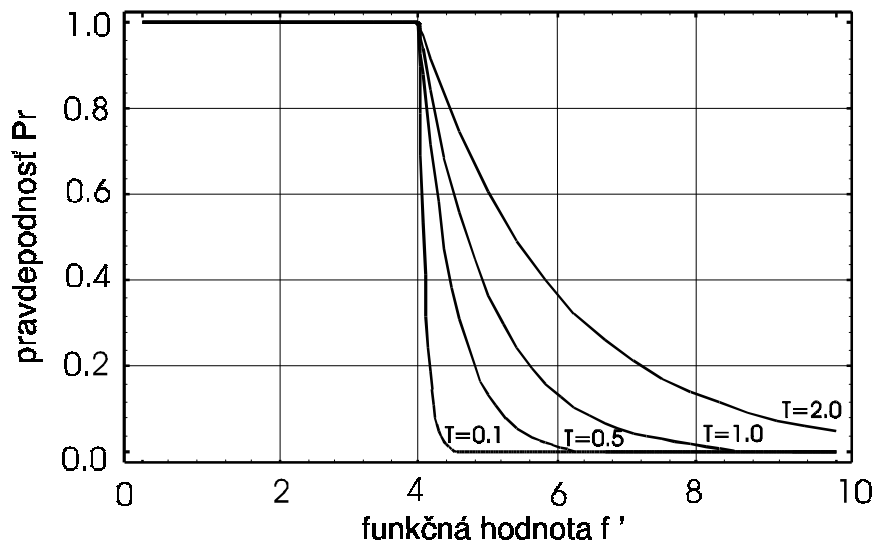
Simulácia fyzikálnej evolúcie

Metropolis a spol. navrhli Monte Carlo metódu, ktorá simuluje evolúciu systému tak, že generuje postupnosť stavov systému nasledujúcim spôsobom:

Nech je daný aktuálny stav systému (určený polohou častíc telesa), potom malá náhodná porucha je generovaná tak, že častice sú "jemne" posunuté. Ak rozdiel $\Delta E = E_{perturbed} - E_{current}$ medzi porušeným stavom a aktuálnym stavom je negatívny ($E_{perturbed} < E_{current}$), potom proces pokračuje s novým porušeným stavom. V opačnom prípade, ak $\Delta E \geq 0$, pravdepodobnosť akceptovania porušeného stavu,

$$Pr(perturbed \leftarrow current) = \min(1, \exp(-\Delta E / kT))$$

Toto pravidlo akceptovania porušeného stavu sa nazýva *Metropolisovo kritérium*.



Priebeh Meteropolisovho kritéria

$$Pr = \min(1, \exp(-(f' - f)/T))$$

pre rôzne teploty T , kde f je fixná funkčná hodnota ($f=4$) a f' je nezávislá premenná braná z intervalu $[0,10]$. Pre klesajúce hodnoty teploty T a pre $f' > f$, pravdepodobnosť $Pr \rightarrow 0$ ak $T \rightarrow 0$.

Formalizácia Metropolisovho algoritmu

- Stav systému nech je určený *stavovou premenou* \mathbf{x} (vo všeobecnosti vektor obsahujúci mnoho nezávislých reálnych premenných) a energiou $f(\mathbf{x})$.
- Proces poruchy stavu \mathbf{x} na iný stav \mathbf{x}' je formálne reprezentovaný stochastickým operátorom, $\mathbf{x}' = Q_{\text{pertur}}(\mathbf{x})$. Stochastický charakter tohto operátora spočíva v náhodnej zmene stavu \mathbf{x} na stav \mathbf{x}' .


```

procedure Metropolis_algorithm
begin k:=0; x:=xini;
      while k<kmax do
      begin k:=k+1;
            x' :=Opertur(x);
            Pr:=
            min(1, exp(-(f(x')-f(x))/T));

            if random<Pr then x:=x';
      end;
      xout:=x;
end;

```

Implementácia Metropolisovho algoritmu. Procedúra je inicializovaná tak, že počiatočný stav sa položí rovný vstupnému stavu \mathbf{x}_{ini} , opakuje sa k_{max} krát (toto číslo musíme byť dostatočne veľké ktomu, aby sa dosiahla tepelná rovnováha), symbol O_{pertur} modifikuje aktuálny stav \mathbf{x} na \mathbf{x}' . Akceptovanie nového stavu je riešené pomocou Metropolisovho kritéria realizovaného pre teplotu T . Po skončení Metropolisovho algoritmu, výstupný stav \mathbf{x}_{out} je posledný stav \mathbf{x} .

```

procedure Simulated_annealing;
begin  $x_{ini}$  := randomly generated;
       $T := T_{max}$ ;
      while  $T > T_{min}$  do
      begin Metropolis_algorithm
              ( $x_{ini}, x_{out}, k_{max}, T$ );
               $x_{ini} := x_{out}$ ;
               $T := \alpha * T$ ;
      end;
       $x_{opt} := x_{out}$ ;
end;

```

Implementácia simulovaného žihania , vstupné parametre sú T_{min} , T_{max} , k_{max} , α , výstupný parameter je x_{opt} . Algorimus je inicializovaný náhodným generovaním poèiatoèného stavu x_{ini} a maximálnou teplotou T_{max} . While-cyklus sa opakuje tak dlho, pokiaľ platí $T > T_{min}$, , teplota T sa znižuje pomocou parametra α vzťahom $T := \alpha * T$. Po ukončení while-cyklu výsledný stav x_{out} je považovaný za výsledné riešenie oznaèené x_{opt} .

- Jeden zo základných problémov simulovaného žihania je špecifikácia teplôt T_{\max} , T_{\min} a pravidla pre znižovanie teploty T .
- V literatúre boli navrhnuté rôzne schémy, ako určiť tieto parametre simulovaného žihania.
- V praktických aplikáciách sa dobre osvedčil najjednoduchší prístup pre ochladzovanie (multiplikatívny prístup pomocou parametra α) a počiatočná teplota je zvolená tak, aby približne 50% stavov porušených stavov bolo akceptovaných Metropolisovým algoritmom.

Vzťah štatistickej fyziky

1. Základný predpoklad štatistickej fyziky je, že fyzika mnohočasticových systémov je kompatibilná so štatistickými súbormi a že časové stredné hodnoty mechanických veličín systému v rámci mikroskopickej rovnováhy sú rovné odpovedajúcim stredným hodnotám nad súbormi.
2. V štatistickej fyzike boli definované vhodné veličiny popisujúce fyzikálne vlastnosti systému, a tiež bolo ukázané, ako počítať tieto veličiny pomocou rovnovážneho rozdelenia stavov systému.
3. Ukazuje sa, že v tepelnej rovnováhe rozdelenie pravdepodobnosti je určené Boltzmannovým rozdelením. Vzťah medzi štatistickou fyzikou a simulovaným žíhaním môže byť študovaný pomocou formalizmu štatistickej fyziky.

Uvažujem funkciu f definovanú nad diskretnou a konečnou doménou D

$$f : D \rightarrow A \subset R_+$$

Rozdelenie pravdepodobnosti stavu \mathbf{x} (t.j. premennej \mathbf{x} z D), ktoré je výsledkom Metropolisovho algoritmu realizovaného pre teplotu T , je určené vzťahmi

$$w_T(E_i) = \frac{1}{Q(T)} \exp\left(-\frac{E_i}{kT}\right)$$

$$Q(T) = \sum_i \exp\left(-\frac{E_i}{kT}\right)$$

Pre naše ďalšie úvahy bude vhodné poznať hustotu funkčných hodnôt $y=f(x)$, ktorá je určená takto

$$w_T(y) = \frac{(D(y)) \left(\exp\left(-\frac{y}{T}\right) \right)}{Q(T)}$$

$$Q(T) = \sum_{y \in A} (D(y)) \left(\exp\left(-\frac{y}{T}\right) \right)$$

kde $D(y) \subset D$ je množina obsahujúca všetky $x \in D$ pre ktoré platí $y=f(x)$

$$D(y) = \{x \in D; y = f(x)\}$$

"Makroskopické veličiny"

(1) *Středná hodnota* funkcie $f(\mathbf{x})$

$$\langle f \rangle_T = \sum_{y \in A} y w_T(y)$$

(2) *Středná hodnota* funkcie $f^2(\mathbf{x})$

$$\langle f^2 \rangle_T = \sum_{y \in A} y^2 w_T(y)$$

(3) *Disperzia* funkcie $f(\mathbf{x})$

$$\sigma^2(T) = \sum_{y \in A} w_T(y) (\langle f \rangle_T - y)^2 = \langle f^2 \rangle_T - \langle f \rangle_T^2$$

(4) *Entropia*

$$S(T) = - \sum_{y \in A} w_T(y) \ln \left(\frac{w_T(y)}{|D(y)|} \right)$$

(4) *Špecifické teplo*

$$C(T) = \frac{\partial}{\partial T} \langle f \rangle_T = \frac{\sigma^2(T)}{T^2} = T \frac{\partial S(T)}{\partial T}$$

Hustota funkčných hodnôt (6a-b) vyhovuje týmto asymptotickým podmienkam

$$\lim_{T \rightarrow \infty} w_T(y) = w_\infty(y) = \frac{1}{|D|} |D(y)|$$

$$\lim_{T \rightarrow 0} w_T(y) = w_0(y) = \delta(y, y_{opt}) = \begin{cases} 1 & (\text{pre } y = y_{opt}) \\ 0 & (\text{pre } y \neq y_{opt}) \end{cases}$$

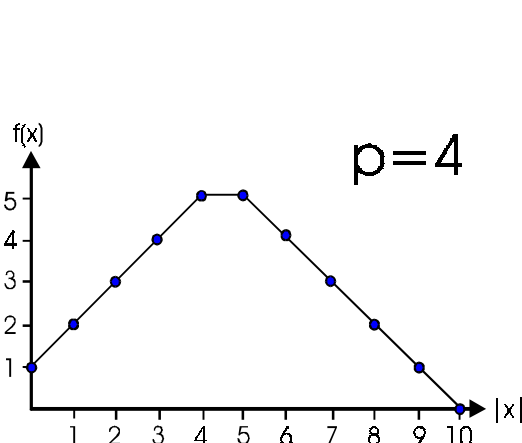
kde $y_{opt} = f(x_{opt})$ je optimálna hodnota funkcie $f(x)$ odpovedajúca globálnemu minimumu. Druhý asymptotický vzťah je veľmi dôležitý pre metódu simulovaného žihania; podľa tejto formule ak teplota T sa blíži k nule (cez rovnovážne stavy), potom *pravdepodobnosť výskytu stavu odpovedajúceho globálnemu minimumu je jednotková.*

Modelové výpočty

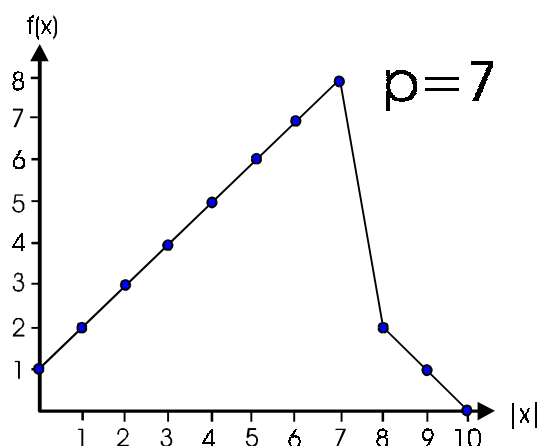
Nech zobrazenie $f : D = \{0,1\}^N \rightarrow A \subset R_+$, s definičným oborom obsahujúcim binárne vektory dĺžky N , je definované takto

$$f(x) = f(x_1, x_2, \dots, x_N) = \begin{cases} |x| + 1 & (\text{pre } |x| \leq p) \\ N - |x| & (\text{pre } |x| > p) \end{cases}$$

kde $|x| = \sum x_i$ je L_1 norma vektora $x \in \{0,1\}^N$, $0 \leq |x| \leq N$. Parameter p je ohraňovaný podmienkou $0 \leq p \leq N$. Pre $p < N$ táto funkcia má dve minima, prvé minimum je $f(x) = 1$, pre $|x| = 0$, druhé (globálne) minimum je $f(x) = 0$, pre $|x| = N$. Parameter p určuje stupeň *deceptívnosti* (falošnosti) tejto funkcie



A



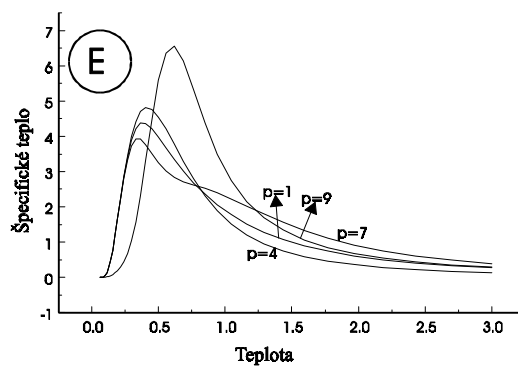
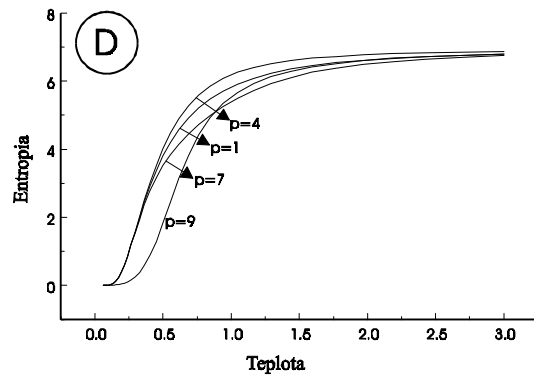
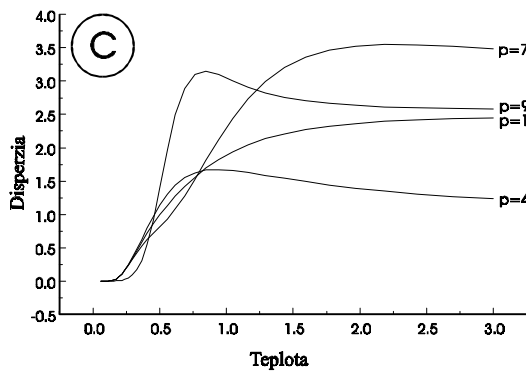
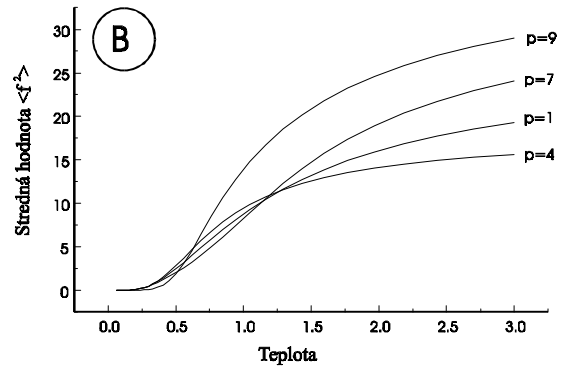
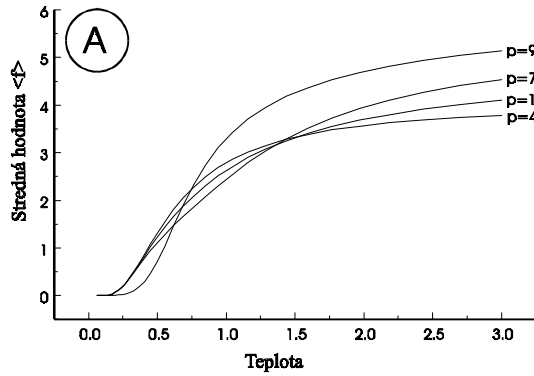
B

Operátor $O_{perturb}$ z Metropolisovho algoritmu transformujúci binárny vektor $x \in \{0,1\}^N$ na iný binárny vektor $x' \in \{0,1\}^N$ je v tomto prípade realizovaný pomocou operátora mutácie $O_{mut}^{(P_{mut})}$ (terminológia je prevzatá z genetického algoritmu)

K tomu, aby sme mohli vypočítať makroskopické veličiny definované v predchádzajúcej kapitole 6.2 musíme poznať kardinality množín $D(y)$ pre funkciu (6.19). Tak napr. pre $N=10$ a pre $p=4$ a $p=7$, tieto kardinality sú určené takto

$$p=4 : |D(0)|=1, \quad |D(1)|=11, \quad |D(2)|=55, \\ |D(3)|=165, \quad |D(4)|=330, \quad |D(5)|=462$$

$$p=7 : |D(0)|=1, \quad |D(1)|=11, \quad |D(2)|=55, \\ |D(3)|=45, \quad |D(4)|=120, \quad |D(5)|=210, \\ |D(6)|=252, \quad |D(7)|=210, \quad |D(8)|=120$$



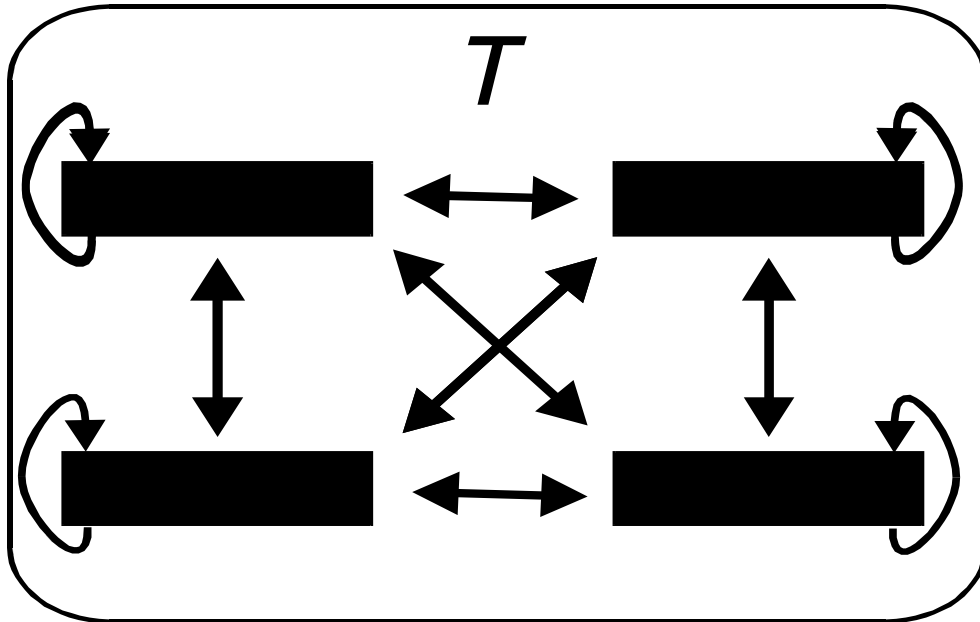
Priebehy makroskopických veličin vypočítaných teoreticky pomocou formúl pre parametre $p=1,4,7,9$ (diagramy A - E).

Paralelné simulované žihanie

Táto verzia simulovaného žihania bola s úspechom použitá pre riešenie komplikovaných kombinatoriálnych grafovo-teoretických problémov, kde štandardná verzia simulovaného žihania nie je schopná poskytnúť správne globálne riešenie.

Základná idea paralelného simulovaného žihania spočíva v súčasnom aplikovaní simulovaného žihania na množinu P stavov x, x', x'' , ... ktoré sú "synchronizované" tak, že Metropolisové algoritmy bežiace nad nimi majú rovnakú teplotu.

K interakcii stavov je použitá operácia *kríženia*.



Schematické znázornenie paralelného simulovaného žihania. Na jednotlivé stavy (znázornená čiernymi blokmi) sú aplikované nezávislé Metropolisove algoritmy s rovnakou teplotou T . Šípky znázorňujú, že stavy medzi sebou interagujú s malou pravdepodobnosťou P_{cross} , kde r je náhodne vybraný bod kríženia, $1 < r < N$.