

Metody Obliczeniowe w Nauce i Technice

Optymalizacja metodą simulated annealing

Marian Bubak, Katarzyna Rycerz

Department of Computer Science
AGH University of Science and Technology
Krakow, Poland
kzajac@agh.edu.pl
dice.cyfronet.pl

Contributors

Marcin Przewięźlikowski
Miłosz Błaszkwicz
Łukasz Janeczko



Outline

- 1 Wprowadzenie
- 2 Optymalizacja kombinatoryczna
- 3 Annealing & quenching
- 4 Fizyka statystyczna i optymalizacja - analogie
- 5 Algorytm Simulated Annealing
- 6 Pierwsze zastosowania Simulated Annealing
- 7 Odmiiany Simulated Annealing
- 8 Literatura
- 9 Wyżarzacz kwantowy

Wprowadzenie

Annealing

- wyżarzanie, odprężanie (powolne chłodzenie)
- przykład zastosowania **metody fizyki statystycznej w optymalizacji** [1]

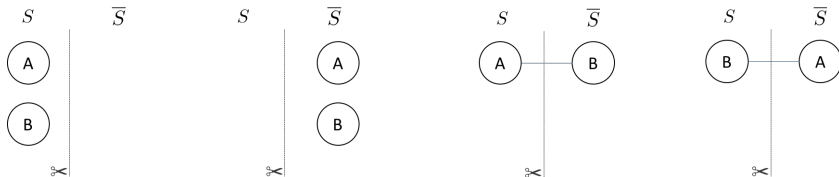
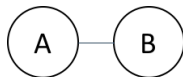
Simulated annealing

- rodzina heurystycznych technik optymalizacyjnych opartych na analogii z fizyką statystyczną układów losowych
- opis pochodzenia metody (ang.):
<https://pdfs.semanticscholar.org/beb2/1ee4a3721484b5d2c7ad04e6babd8d67af1d.pdf>

Optymalizacja kombinatoryczna

- rozwiązania o złożoności $\sim e^N$
- wiele stopni swobody
- dyskretne (wykluczone poszukiwanie w kierunku)
- funkcja celu łączy przeciwstawne cele cząstkowe

Przykład: problem Max-Cut



Funkcja kosztu:

$$g(x_1, x_2) = x_1 + x_2 - 2x_1x_2, \quad x_1, x_2 \in \{0, 1\}$$

$$g(00) = 0 \quad g(11) = 0$$

$$g(01) = 1 \quad g(10) = 1$$

source: <https://grove-docs.readthedocs.io/en/latest/qaoa.html>

Przykład - problemy BILP i QUBO

Funkcja kosztu i warunki (zmienne binarne)

$$f(x) = c^T x \text{ and } Ax = b \ x_i \in \{0, 1\}$$



C-diagonal, $\text{diag}(C)=c$

QUBO - quadratic and unconstraint cost function $f(x) = x^T Q x$

$$f(x) = x^T C x + P \underbrace{(Ax - b)^T (Ax - b)}$$

instead of solving $Ax=b$
we minimize inner product of $Ax-b$

Typowa funkcja celu

- skalar: wszystkie cele sprowadza do jednego
- wiele lokalnych minimów - rzędu e^N
- w praktyce - potrzebne dobre rozwiązanie - nie musi być to minimum globalne

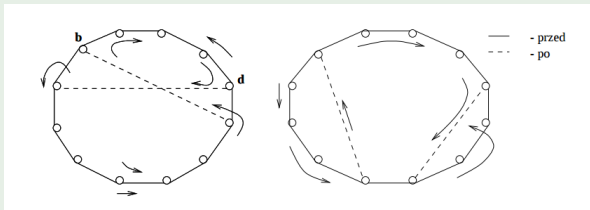
Przykład procedury generacji nowych konfiguracji

TSP [4]

Heurystyka: *iterative improvement* - akceptowalne zmiany zmniejszające funkcję celu

- 1 odwrócenie kolejności obiegu 5-ciu górnych
- 2 wstawienie 5-ciu górnych między 2 dolne

→ pewne ugrzęźnięcie w lokalnym minimum



Annealing & quenching

Czyli powolne i szybkie schładzanie

Annealing

Stopniowe, powolne zmniejszanie temperatury

- topnienie (ciecz)
- w każdej temperaturze trwa długo, do uzyskania równowagi termicznej

⇒ uporządkowanie w kryształ - struktura regularna, symetryczna o minimalnej energii

Annealing & quenching

Quenching

Szybkie zmniejszanie temperatury

⇒ uzyskanie stanu metastabilnego

- polikryształ
- kryształ z defektami

Odpowiednik *iterative improvement*

Wniosek: zamiast zawsze odrzucać konfigurację zwiększającą funkcję celu, niekiedy należy ją akceptować z odpowiednim prawdopodobieństwem (uphill)

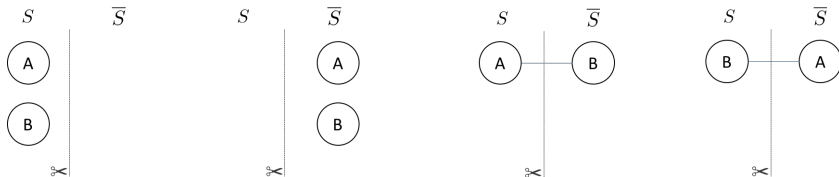
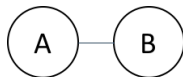
Fizyka statystyczna i optymalizacja - analogie

Fizyka statystyczna

uśrednione wartości dla zespołów o dużej ilości molekuł

| Mechanika statystyczna | Optymalizacja |
|--|---|
| wiele oddziałujących molekuł | wiele parametrów |
| układ, zbiór położeń molekuł | konfiguracja |
| schłodzenie do stabilnego stanu niskoenergetycznego | znalezienie konfiguracji prawie optymalnej |
| temperatura | parametr sterujący przebiegiem optymalizacji |
| hamiltonian (operator energii) | funkcja celu |
| w hamiltonianie człony wynikające z różnych oddziaływań | współzawodniczące człony w funkcji celu |

Przykład: problem Max-Cut



Funkcja kosztu:

$$g(x_1, x_2) = x_1 + x_2 - 2x_1x_2, \quad x_1, x_2 \in \{0, 1\}$$

$$g(00) = 0 \quad g(11) = 0$$

$$g(01) = 1 \quad g(10) = 1$$

source: <https://grove-docs.readthedocs.io/en/latest/qaoa.html>

Hamiltonian dla MaxCut

Mamy efektywną symulację dla modelu Isinga:

<http://www.bdhammel.com/ising-model/>

<https://stanford.edu/~jeffjar/statmech/intro4.html>

Funkcja kosztu naszego problemu:

$$g(x_1, x_2) = x_1 + x_2 - 2x_1x_2, \quad x_1, x_2 \in \{0, 1\}$$

W modelu Isinga

- $x=0$ - spin do góry ($z=1$)
- $x=1$ - spin w dół ($z=-1$)

Funkcja energii zależna od konfiguracji spinów

$$x \rightarrow \frac{1}{2}(1 - z)$$

$$g(z_1, z_2) = \frac{1}{2}(1 - z_1z_2) \quad z_1, z_2 \in \{-1, 1\}$$

Model Isinga

$$H_{\text{maxcut}} = \frac{1}{2} - \frac{1}{2} \sum_{\langle ij \rangle} z_i z_j$$

$$H_{\text{Ising}} = - \sum_i h_i z_i - \sum_{\langle ij \rangle} J_{i,j} z_i z_j$$

$$H_{\text{maxcut}} = \frac{1}{2} - \frac{1}{2} H_{\text{Ising}}$$

dla $h = 0$, $J = -1$

Typowa funkcja celu

Niezbędne:

- C_i - reprezentacja konfiguracji układu
- $g(C_i)$ - funkcja celu
- procedura generacji kolejnych konfiguracji

Algorytm Simulated Annealing

| | | |
|--|--|---|
| ustal $\{T_1, T_2, \dots, T_l, \dots, T_n\}$ | | |
| $l := 1;$ | | |
| $T := T_l;$ | | |
| $i := 1;$ | | |
| wygeneruj konfigurację C_i | | |
| | | $C_i \rightarrow C_{i+1}$ (zgodnie ze schematem Metropolis) |
| | | oblicz $g(C_{i+1}), \Delta g_i$ |
| | | until równowaga dla $T = T_l$ (min. lok.) |
| | | $l := l + 1;$ |
| $T := T_l;$ | | |
| until $l < n$ | | |

$\{T_l\}$ - temperature schedule: $T_l > T_{l+1}$

np. $T_{l+1} = 0.9 \cdot T_l$

Fundamentalny fakt z mechaniki statystycznej

Założmy, że układ znajduje się w równowadze termicznej (w temperaturze T)

Prawdopodobieństwo, że układ znajduje się w mikrostanie α jest proporcjonalne do **czynnika Boltzmann**

$$e^{\frac{-E_{\alpha}}{k_B T}}$$

gdzie: E_{α} - energia stanu

Schemat Metropolisisa

Podstawa metody Monte Carlo symulacji molekularnej

| | |
|--|--|
| wybierz losowo molekułę m | |
| przenieś losowo m : $\vec{r}_m := \vec{r}_m + \underbrace{\Delta \vec{r}}_{\text{losowe}}$ (inaczej: zmien konfigurację) | |
| wyznacz zmianę energii (funkcji celu) układu Δg | |
| Δg | |
| < 0 ≥ 0 | |
| zaakceptuj nową konfig. $\vec{r}_m := \vec{r}_m$; | wylosuj $R \in (0, 1)$, rozkład równomierny |
| | $e^{-\frac{\Delta g}{kT}}$ |
| | $> R$ $\leq R$ |
| zaakceptuj $\vec{r}_m := \vec{r}_m$; | odrzuć $\vec{r}_m := \vec{r}_m$ |
| until zadana liczba kroków | |



Nicolas Metropolis, Arianna W. Rosenbluth, Marshal M. Rosenbluth, Augusta H. Teller, Edward Teller
J. Chem. Phys. 21 (1953) 1087

Sprawdzanie, czy układ jest w równowadze

Sposób A

utrzymywać T_I przez $\begin{cases} 100 \cdot N \text{ prób} \\ 10 \cdot N \text{ prób udanych } (\Delta g < 0) \end{cases}$

Sposób B

n - ustalona liczba prób (\sim epoka)

- ① wykonać n prób ($C_i \rightarrow C_{i+1}$)
- ② zachować $g(C_n)$
- ③ porównać $g_i(C_n)$ dla kilku ostatnich zestawów po n próbach
brak istotnej zmiany $g(C_n) \rightarrow$ nowe T .

Wartość początkowa T

- wystarczająco wysoka by zapewnić akceptację niemal wszystkich przejść
- wartość początkowa parametru T zależy od postawionego problemu
- np. przyjąć jakiś wstępną wartość prawdopodobieństwa $P \approx 1$ i dla losowej próby wyliczyć średnią różnicę pomiędzy
$$\Delta g = g(C_i) - g(C_j)$$
- wybrać $T_0 = -\frac{\Delta g}{\ln(P)}$

Charakterystyka schematu Metropolis'a

- $T \nearrow$ - łatwiej akceptowalne kroki z $g(C_i) \nearrow (E)$
 \Rightarrow możliwość opuszczenia stanu metastabilnego (lokalnego minimum).
- zmiany $g(C_i) \searrow$ są akceptowane zawsze

Po wielu krokach system \rightarrow stan równowagi termodynamicznej z parametrami oscylującymi wokół wartości średnich zgodnie z rozkładem Boltzmanna

Pierwsze zastosowania Simulated Annealing

- **Kirkpatrick**

TSP - 3000 random cities (dokładne rozwiązanie ≤ 318)
miasta w klastrach:

- duże T - optymalna droga między klastrami
- małe T - optymalna droga wewnątrz klastrów

\Rightarrow "divide and conquer" behaviour \rightarrow podział zagadnienia na różne skale

- **Kirkpatrick, Gelatt**

Optymalizacja rozłożenia mikroukładów elektronicznych na 1 lub więcej chipach, łączenie chipów.

Pierwsze zastosowania Simulated Annealing

- **Vecchi, Kirkpatrick**

Optimal wiring (między VLSI)

- min length
- min bends
- no crowding

- **Kenneth Wilson, Dean Jacobs, Jan Prins - Cornell University**

Algorytm Metropolisa do optymalizacji kodu komputerowego
(\approx iteracyjne przestawianie komend)

Dalsze zastosowania Simulated Annealing

- **Biologia i chemia molekularna**

Przykłady:

- Optymalizacja struktury cząsteczki białka zbudowanej metodami modelowania molekularnego
- Uściślanie struktur rozwiązanych metodą dyfrakcji rentgenowskiej
- Przejście "od więzów do struktury" przy wyznaczaniu struktury przestrzennej białka metodą NMR
- Proces selekcji potencjalnych leków przez dokowanie

Na podstawie <https://bioinfo.mol.uj.edu.pl/courses/AppliedModelling/LectureSchedule?action=AttachFile&do=get&target=w3.pdf>

Odmiany Simulated Annealing

- COSA - Cooperative Simulated Annealing
- **ASA - Adaptive Simulated Annealing**
Strona prof. Lester Ingbera poświęcona ASA:
<http://alumnus.caltech.edu/~ingber/>
- PARSA - Parallel Simulated Annealing
- EBSA - Ensemble Based Simulated Annealing

Literatura I



Scott Kirkpatrick, Daniel Gelatt, Mario Venchi
IBM T.J. Watson Research Center
Science 220 (1983) 671-670



S. Kirkpatrick
J. Stat. Phys. 34 (1984) 975 (No 5/6)



M.R. Garey, D.S. Johnson
Computers and Intractability: A Guide to the Theory of NP
Completeness, Freeman, San Francisco, 1979



S. Lin, B.W. Kernighan
Oper. Res. 21 (1973) 498

Literatura II



Nicolas Metropolis, Arianna W. Rosenbluth, Marshal M.
Rosenbluth, Augusta H. Teller, Edward Teller
J. Chem. Phys. 21 (1953) 1087

Wyżarzacz kwantowy D-Wave

- metaheurystyka do znajdowania globalnego minimum
- tunelowanie kwantowe (wyjście z minimum lokalnego)

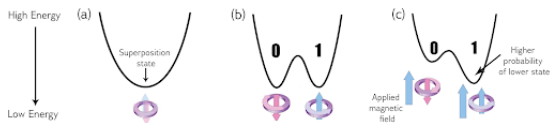


Figure: Zmiany energii podczas kwantowego wyżarzania

source: https://docs.dwavesys.com/docs/latest/c_gs_2.html

Notacja Diraca, iloczyn tensorowy

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$|00\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad |01\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad |10\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad |11\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$|01\rangle = |0\rangle \otimes |1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 0 \\ 1 \cdot 1 \\ 0 \cdot 0 \\ 0 \cdot 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Kwantowy Hamiltonian dla MaxCut

Funkcja kosztu:

$$g(x_1, x_2) = x_1 + x_2 - 2x_1x_2, \quad x_1, x_2 \in \{0, 1\}$$

$$x \rightarrow \frac{1}{2}(1 - z)$$

$$g(z_1, z_2) = \frac{1}{2}(1 - z_1z_2) \quad z_1, z_2 \in \{-1, 1\}$$

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\sigma_z |0\rangle = |0\rangle, \quad \lambda = 1$$

$$\sigma_z |1\rangle = -|1\rangle, \quad \lambda = -1$$

aby zbudować Hamiltonian: $z \rightarrow \sigma_z$

$$H_{i,j} = \frac{1}{2}(I - \sigma_z^i \otimes \sigma_z^j) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{matrix}$$

Wartości własne

- 0 dla $|00\rangle$ i $|11\rangle$; $g(00) = 0$ $g(11) = 0$

- 1 dla $|01\rangle$ i $|10\rangle$; $g(01) = 1$ $g(10) = 1$

Quantum Annealing

- startuje w stanie o najniższej energii początkowego H
- powoli zmienia początkowy H w H dla naszego problemu
- zmiana dokonuje się poprzez wprowadzenie tzw couplers (J) oraz biases (h)
- w idealnej sytuacji system pozostaje w stanie o minimalnej energii przez cały ten proces
- działanie kończy się, gdy system znajduje się w stanie o minimalnej energii dla naszego problemu
- wynik zwracany jest jako klasyczna wartość

$$\underbrace{-\frac{A(s)}{2} \left(\sum_i \sigma_x^{(i)} \right)}_{\text{Initial Hamiltonian}} + \underbrace{\frac{B(s)}{2} \left(\sum_i h_i \sigma_z^{(i)} + \sum_{i>j} J_{i,j} \sigma_z^{(i)} \sigma_z^{(j)} \right)}_{\text{Final Hamiltonian}}$$

Co należy zrobić?

- sformułuj problem jako QUBO
- dopasuj QUBO do architektury komputera
- otrzymaj rezultaty
- dokonaj odwrotnego dopasowania wyników do początkowego QUBO

QUBO i Hamiltonian

QUBO - quadratic and unconstrained cost function $f(x) = x^T Qx$

$$f(x) = x^T Cx + P \underbrace{(Ax - b)^T (Ax - b)}_{\substack{\text{instead of solving } Ax=b \\ \text{we minimize inner product of } Ax-b}}$$

$$x_i \rightarrow \frac{1 - \sigma_z^i}{2} \quad \updownarrow \quad \sigma_z^i = I \otimes I \cdots \otimes \underbrace{\sigma_z}_{\text{i-th position}} \cdots \otimes I$$

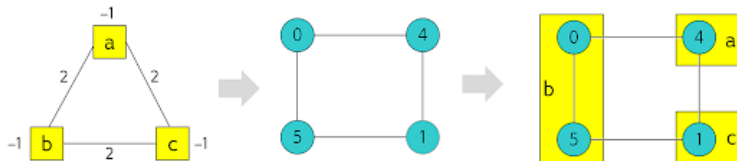
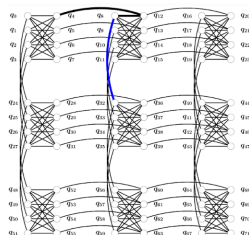
Hamiltonian

$$H = \sum_i c_i \frac{1 - \sigma_z^i}{2} + P \sum_j \left(\sum_i a_{j,i} \left(I - \frac{\sigma_z^i}{2} \right) - b_j I \right)^2$$

Obydwie formy mogą być wejściem do wyżarzacza kwantowego

Minor Embedding

- Architektura kwantowego wyżarzacza nie jest grafem pełnym
- konieczna jest transformacja grafu naszego problemu do tej architektury
- zwykle konieczne jest reprezentowanie jednej zmiennej problemu przez wiele qbitów (łańcuchy)



Embedowanie pełnego grafu

- przykład jak osadzić graf K_9 w grafie Chimera 2×2
- trzy qbity na jedną zmienną (poza 8)
- ta metoda działa dla grafów do 9 wierzchołków

