# SMC-SA

Kamil Kisiel

Marzec 2023

# Plan prezentacji

- SA
- SMC-SA
- Zakres błędu
- Eksperymenty numeryczne

# SA - przypomnienie

## Co to było SA?

Algorytm szukający rozwiązania danego problemu, którego działanie zaprezentuję na następującym przykładzie:

## Problem optymalizacyjny

Znalezienie maksimum danej funkcji:

## Założenia

$\mathcal{X}$ - niepusty zbiór na $\mathbb{R}^n$
$H : \mathcal{X} \rightarrow \mathbb{R}$ (jest ograniczona i ciągła)

# SA - Algorytm

1. Losujemy startowy punkt $x_k$
2. Generujemy $y_k$ z rozkładu o gęstości $g_k(y|x_k)$
3. Obliczamy prawdopodobieństwo $\rho_k = \min\{\frac{H(y_k)-H(x_k)}{T_k}, 1\}$
4. Generujemy losowe $u \in [0,\ 1]$
5. Ustalamy
$$x_{k+1} = \begin{cases} y_k, & u \geqslant \rho_k \\ x_k, & u < \rho_k \end{cases} \tag{1}$$
6. Sprawdzamy kryterium stopu, ewentualnie zwiększamy $k$ i zmiejszamy temperaturę

# SMC-SA

## Co to SMC-SA?

W skrócie jest to połączenie SA oraz metody Monte Carlo, dzięki czemu jesteśmy w stanie pracować na wielu punktach na raz.

## Start

1. Dostarczamy ciąg wielkości próbek $\{N_k\}$ oraz ciąg temperatur $\{T_k\}$
2. Inicjalizacja: generujemy $x_0^i \overset{\text{iid}}{\sim} Unif(X), i = 1, 2, ..., N_0$
3. Ustawiamy $k = 1$

# SMC-SA - Algorytm

1. Importance updating: generujemy $w_k^i$ z rozkładu:

$$\begin{cases} exp\{\frac{H(x_0^i)}{T_1}\}, & k = 1 \\ exp\{H(x_{k-1}^i)(\frac{1}{T_k} - \frac{1}{T_{k-1}})\}, & k > 1 \end{cases} \tag{2}$$

2. Resampling: generujemy $\{\tilde{x}_k^i\}_{i=1}^{N_k}$ z $\{x_{k-1}^i, w_k^i\}_{i=1}^{N_{k-1}}$
3. SA Move: generujemy $x_k^i$ z $\tilde{x}_k^i$ dla $i = 1, .., N_k$, korzystając z SA
4. Sprawdzamy kryterium stopu oraz ewentualnie zwiększamy $k$

# SMC-SA - zakres błędu

Przy naszych założeniach, rozkład Boltzmanna słabo się zbiega do równomiernego rozkładu na zbiorze optymalnych rozwiązań.

**Propozycja 1**

Dla każdego $\xi > 0$ :

$$\lim_{T_k \to 0} \pi_k(\mathcal{X}_\xi) = 1,$$

gdzie $\mathcal{X}_\xi = \{x \in \mathcal{X} : H(x) > H^* - \xi\}$
$H^*$ - optymalna wartość funkcji $H$

# Oznaczenia pomocnicze

$\mathcal{F}$ - $\sigma$-ciało na $(\mathcal{X})$
$\mathcal{B}(\mathcal{X})$ - zbiór mierzalnych i ograniczonych funkcji $\phi : \mathcal{X} \to \mathbb{R}$
$\mathcal{B}_+(\mathcal{X})$ - zbiór mierzalnych i ograniczonych funkcji $\phi : \mathcal{X} \to \mathbb{R}_+$
$\langle \upsilon, \phi \rangle = \int \phi(x)\upsilon(dx), \quad \forall \phi(x) \in \mathcal{B}(\mathcal{X})$

# Definicje pomocnicze pt.1

## Supremum normy

$$||\phi|| = \sup_{x \in \mathcal{X}} |\phi(x)|$$

## Całkowity dystans zmienności

$\upsilon_1, \upsilon_2$ - miary probabilistyczne na $(\mathcal{X}, \mathcal{F})$

$$||\upsilon_1 - \upsilon_2||_{TV} = \sup_{A \in \mathcal{F}} ||\upsilon_1(A) - \upsilon_2(A)||$$

# Definicje pomocnicze pt.2

## Rozkłady prawdopodobieństwa w k-tej iteracji SMC-SA

$$\pi_k^d = \frac{exp(H(x)/T_k)}{\int exp(H(x)/T_k)dx}$$

$$\tilde{\mu}_k = \sum_{i=1}^{N_{k-1}} \omega_k^i \delta_{x_{k-1}^i}$$

$$\tilde{\mu}_k^{N_k} = \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{\tilde{x}_k^i}$$

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{x_k^i}$$

# Definicje pomocnicze pt.3

$$\Psi_k = \frac{\pi_k^d}{\pi_{k-1}^d}$$

## Związki między rozkładami

$$\mu_{k-1} \longrightarrow \tilde{\mu}_k = \frac{\mu_{k-1}\Psi_k}{\langle \mu_{k-1}, \Psi_k \rangle} \longrightarrow \tilde{\mu}_k^{N_k} \longrightarrow \mu_k = \tilde{\mu}_k^{N_k} P_k$$

# SMC-SA - zakres błędu c.d.

**Założenie 1**

Gęstość zaproponowana w SA Move musi spełniać następujący warunek:

$$g_k(y|x) \geqslant \epsilon_k > 0, \quad \forall x, y \in \mathcal{X}$$

# Jednostajna ergodyczność

## Twierdzenie 1

Rozważmy Łańcuch Markowa o kernelu przejściowym $P(x, dy)$ dla $x, y \in \mathcal{X}$ i stacjonarny rozkład prawdopodobieństwa $\pi$ Wtedy przestrzeń $\mathcal{X}$ nazywamy małą jeśli istnieją:

- $n_0 \in \mathbb{Z}_+$
- Stała $\epsilon \in (0, 1)$
- Miara probabilistyczna $\upsilon$ na $\mathcal{X}$

takie, że spełniony jest warunek minoryzacji:

$$P^{n_0}(x, A) \leqslant \epsilon \upsilon(A), \quad \forall x \in \mathcal{X}, \forall A \in \mathcal{F}$$

Wtedy Łańcuch jest ergodyczny oraz:

$$||P^n(x) - \pi||_{TV} \leqslant (1 - \epsilon)^{\lfloor n/n_0 \rfloor}, \quad \forall x \in \mathcal{X}$$

# Wnioski z Twierdzenia 1

## Wniosek 1.1

Przy założeniu 1 Łańcuch Markowa zgodny z krokiem SA Move po każdej iteracji jest jednostajnie ergodyczny oraz:

$$\exists \epsilon_k \in (0,1) \quad ||P_k^n(x) - \pi_k||_{TV} \leqslant (1-\epsilon_k)^n,$$

$$\epsilon_k = \varepsilon_k exp\{\frac{H_l - H_u}{T_k}\}\lambda(\mathcal{X})$$

# Wnioski z Twierdzenia 1 c.d.

## Wniosek 1.2

Rozważmy Łańcuch Markowa z początkowym rozkładem $\mu$, kernelem przejść $P$ oraz stacjonarnym rozkładem prawdopodobieństwa $\pi$. Załóżmy, że $\forall \phi \in \mathcal{B}(\mathcal{X}) \quad |\langle \mu - \pi, \phi \rangle| \leqslant c||\phi||$, gdzie $c$ jest dodatnią stałą. Wtedy jeśli łańuch jest jednostajnie ergodyczny, to:

$$|\langle \mu P^n - \pi, \phi \rangle| \leqslant (1-\epsilon)^{\lfloor n/n_0 \rfloor} c||\phi||, \quad \forall \phi \in \mathcal{B}_+(\mathcal{X})$$

# Lematy

## Lemat 1

Weźmy zmienne losowe $x^1, ..., x^N$, które są i.i.d. i mają (warunkowy) rozkład $\upsilon$. Oznaczając $\upsilon^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}$ mamy:

$$E[|\langle \upsilon - \upsilon^N, \phi \rangle||\mathcal{F}] \leqslant \frac{||\phi||}{\sqrt{N}}, \quad \forall \phi \in \mathcal{B}(\mathcal{X})$$

## Lemat 2

Załóżmy, że $\forall \phi \in \mathcal{B}(\mathcal{X}) |\langle \mu - \upsilon, \phi \rangle| \leqslant c||\phi||$ gdzie c to dodatnia stała oraz $\mu' = \frac{\mu\Psi}{\langle \mu, \Psi \rangle}$. Wtedy:

$$|\langle \mu' - \upsilon', \phi \rangle| \leqslant c||\Psi||||\phi||, \quad \forall \phi \in \mathcal{B}_+(\mathcal{X})$$

## Twierdzenie 2

Bez straty ogólności zakładamy, że $\forall x \in \mathcal{X}\ \ H(x) > 0$. Ustalmy, że wstępny rozkład to $\upsilon$, a jego gęstość to $\upsilon^d$, to przy założeniu 1:

$$E[|\langle \mu_k - \pi_k, \phi\rangle||\mathcal{F}] \leqslant c_k||\phi||, \quad \forall \phi \in \mathcal{B}_+(\mathcal{X})$$

, gdzie:

$$c_k = \begin{cases} \frac{||\pi_0^d/\upsilon^d||^2}{N_0}, & k = 0 \\ (1 - \epsilon_k)(\frac{1}{\sqrt{N_k}}) + exp(H^*\Delta_k)ck_1, & k > 0 \end{cases} \tag{3}$$

# Następstwo twierdzenia 2

Jeśli $T_k = T_0/log(k+1), \quad \varepsilon_k \lambda(\mathcal{X}) = \varepsilon < 1$, gdzie, $\varepsilon > (1/2)^{1-\frac{H_u - H_l}{T_0}}$ oraz $\frac{H_u - H_l}{T_0} < 1$ i $\{N_k\}$ wzrasta wystarczająco szybko wraz z wzrostem $k$, to:

$$k \to \infty \implies \{c_k\} \to 0$$

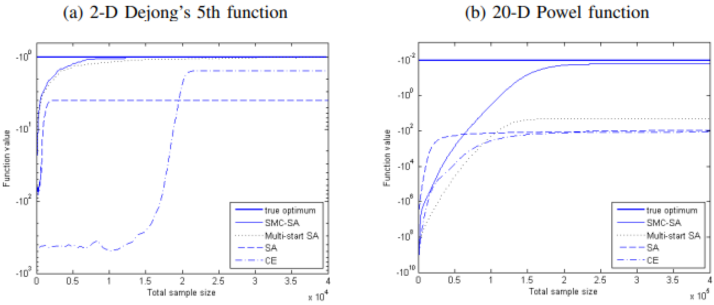
# Ekperymenty numeryczne

## Problemy optymalizacyjne, których użyjemy do porównania

- 5. Funkcja Dejong'a ($H_a$)
- 20-wymiarowa funkcja Powel'a ($H_b$)

| | | SMC-SA | | multi-start SA | | standard SA | |
|---|---|---|---|---|---|---|---|
| | $H^*$ | $\bar{H}^*(std\_err)$ | $M_\varepsilon$ | $\bar{H}^*(std\_err)$ | $M_\varepsilon$ | $\bar{H}^*(std\_err)$ | $M_\varepsilon$ |
| $H_a$ | -0.998 | -0.998(1.34E-7) | 100 | -1.0024(0.0014) | 19 | -3.999(0.2117) | 4 |
| $H_b$ | -0.01 | -0.0164(4.95E-4) | 81 | -20.46(4.26) | 0 | -89.63(1.277) | 0 |

# Wykresy



(a) 2-D Dejong's 5th function
Function value
Total sample size
x 10^4
true optimum
SMC-SA
Multi-start SA
SA
CE
(b) 20-D Powel function
Function value
Total sample size
x 10^5
true optimum
SMC-SA
Multi-start SA
SA
CE

# Bibliografia


- Enlu Zhou i Xi Chen (2011) *Sequential Monte Carlo Simulated Annealing*, Springer Sciensce+Business Media
- H. E. Romeijn i R. L. Smith *Simulated annealing for constrained global optimization*, Journal of Global Optimization