Zastosowanie metod Monte Carlo w zagadnieniach optymalizacji

- Wstęp.
- Metoda „orzeł–reszka”.
- Metody sekwencyjne.
  - ▷ Algorytm podstawowy i jego zbieżność.
  - ▷ Algorytm Charłamowa.
- Inne metody.
- Rozwiązywanie układów równań liniowych metodą minimalizacji formy kwadratowej.

Zagadnienie optymalizacji

Dany jest zbiór $X \subset \mathbb{R}^m$ i funkcja $F : X \to \mathbb{R}$.

Niech $x = (x^1, x^2, \ldots, x^m) \in X$.

▶ Zadanie: **Znaleźć taki punkt**

$$x_{\text{opt}} \in X : \forall_{x \in X}\ F(x) \geq F(x_{\text{opt}})\,.$$

▷ Zagadnienie to różni się od zwykłego zagadnienia szukania minimum funkcji F tym, że poszukiwania wartości minimalnej ograniczone są do zbioru X, co ma zasadnicze znaczenie przy konstrukcji odpowiednich algorytmów obliczeniowych.

▶ **Metody Monte Carlo szukania x_{opt}:**

- **Metoda „orzeł–reszka”** – najprostsza, ale zarazem najmniej wydajna.
- **Metody sekwencyjne** – probabilistyczny odpowiednik metod kolejnych przybliżeń.
- **Inne metody:** optymalizacja statystyczna, algorytmy genetyczne.

Metoda „orzeł–reszka”

Schemat obliczeń:

1. Losujemy niezależnie N punktów $x_1, x_2, \ldots, x_N \in X$ według rozkładu P równomiernego na X (zakładamy, że jest to możliwe).

2. Obliczamy wartość funkcji F w każdym punkcie $x_1, x_2, \ldots, x_N$:

$$F_1 = F(x_1), \quad F_2 = F(x_2), \quad \ldots, \quad F_N = F(x_N)\,.$$

3. Znajdujemy: $F^* = \min\{F_1, F_2, \ldots, F_N\}$

► **Za rozwiązanie przyjmujemy punkt:** $\boxed{x_j:\ F(x_j) = F^*}$

▷ Przeanalizujemy dokładność i wydajność tego algorytmu.

Jeżeli $\{x_k\}$, $k = 1, 2, \ldots, N$ – ciąg niezależnych zmiennych losowych o wartościach w zbiorze X i jednakowym rozkładzie P, to ciąg $\{F_k = F(x_k)\}$, $k = 1, 2, \ldots, N$, jest również ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie z pewną dystrybuantą:

$$G(f) = \mathcal{P}\{F_j < f\}\,,$$

która może być wyrażona za pomocą rozkładu P:

$$G(f) = P\{x_j \in \mathcal{F}(f)\}\,, \qquad \text{gdzie} \quad \mathcal{F}(f) \equiv \{x \in X:\ F(x) < f\}\,.$$

Metoda „orzeł–reszka”

Ozn. g – gęstość zmiennej losowej F_j, tzn. $g(f) = dG(f)/df$.

Niech

$$F^* = \min_{1 \le j \le N} F_j \, .$$

Dla zmiennej losowej $G(F^*)$ mamy:

$$\mathcal{P}\{G(F^*) < \gamma\} = \mathcal{P}\{F^* < G^{-1}(\gamma)\} \, , \qquad 0 < \gamma \le 1 \, .$$

▷ Ponieważ F^* jest statystyką pozycyjną próbki losowej $F_1, F_2, \ldots, F_N$, to (patrz podręczniki rachunku prawdopodobieństwa i statystyki matematycznej):

$$\begin{aligned} \mathcal{P}\{F^* < G^{-1}(\gamma)\} &= N \int_{-\infty}^{G^{-1}(\gamma)} [1 - G(x)]^{N-1} g(x) dx \\ &= N \int_0^{\gamma} (1-u)^{N-1} du = 1 - (1-\gamma)^N \, . \end{aligned}$$

▶ Na podstawie powyższych wzorów otrzymujemy:

$$\mathcal{P}\{\, P\{x \in \mathcal{F}(F^*)\} < \gamma \,\} = 1 - (1-\gamma)^N \, .$$

Metoda „orzeł–reszka”

Ponieważ P jest rozkładem równomiernym na X, to

$$P\{x \in \mathcal{F}(F^*)\} \underset{\text{ozn.}}{=} \mathcal{V}^*(x) \ \leq 1$$

może być interpretowane jako objętość (miara) zbioru takich punktów x, dla których $F(x) < F^*$.

▶ **Wniosek:**

Jeżeli $F^* = \min_{1\leq j\leq N}\{F(x_j)\}$, gdzie x_j, $j = 1, 2, \ldots, N$, są punktami wylosowanymi według rozkładu równomiernego na zbiorze X, to z prawdopodobieństwem $1 - (1-\gamma)^N$ objętość zbioru tych x, dla których $F(x) < F^*$, jest mniejsza of γ.

▷ Mówimy, że z prawdopodobieństwem $1 - (1-\gamma)^N$ punkt x_{opt} został zlokalizowany z dokładnością do zbioru o objętości mniejszej od γ.

$\Rightarrow$ Im mniejsza ta objętość, tym lepsza lokalizacja x_{opt}!

$\longrightarrow$ W szczególności dla $F^* = F(x_{\text{opt}})$ mamy $\mathcal{V}^*(x) = 0$.

● Ile punktów x należy wylosować, aby z prawdopodobieństwem co najmniej $1 - \epsilon$ punkt x_{opt} został zlokalizowany z dokładnością do zbioru o objętości mniejszej od γ?

▶ Na podstawie powyższego mamy:

$$N_{\min} = \min\{N\colon 1 - (1-\gamma)^N \geq 1 - \epsilon\}\,.$$

Najmniejsze liczby N spełniające nierówność: $1-(1-\gamma)^N \geq 1-\epsilon$

γ	$1-\epsilon$					
	0.5	0.9	0.95	0.99	0.999	0.9999
0.5	1	4	5	7	10	14
0.1	7	22	29	44	66	88
0.05	14	45	59	90	135	180
0.01	69	230	299	459	688	917
0.001	693	2302	2995	4603	6905	9206
0.0001	6932	23025	29956	46050	69075	92099

▷ Przykład: Niech $X = [0,1]^m$ i $F : X \to \mathbb{R}$.

Ile trzeba wylosować punktów, aby z prawdopodobieństwem 0.9 zlokalizować minimum tej funkcji z dokładnością do połowy zakresu zmienności każdego z argumentów?

- Dla $m = 1$: $\gamma = 1/2 \;\Rightarrow\; N = 4$;
- Dla $m = 2$: $\gamma = 1/4 \;\Rightarrow\; N = 9$;
- Dla $m = 14$: $\gamma = 2^{-14} < 10^{-4} \;\Rightarrow\; N > 23\,000$.

► **Metoda niewydajna w wielu wymiarach!**

Ogólny schemat algorytmów sekwencyjnych:

1. Ustalić punkt początkowy $x_1 \in X$, np. wylosować z pewnego rozkładu prawdopodobieństwa na zbiorze X.
2. Jeżeli już wylosowano punkty $x_1, x_2, \ldots, x_n$, sprawdzić czy spełnione są określone warunki zwane **regułą stopu (RS)**.
   - Jeżeli TAK – zakończyć obliczenia i punkt x_n przyjąć za rozwiązanie zadania.
   - Jeżeli NIE – wylosować punkt x_{n+1} według rozkładu prawdopodobieństwa zależnego od tego, jakie punkty został już wylosowane i od wartości funkcji F w tych punktach.

▶ **Podstawowy algorytm sekwencyjny:**

1. Ustalamy punkt początkowy x_1.
2. Jeżeli wyznaczyliśmy już punkty $x_1, x_2, \ldots, x_n$, to losujemy pomocniczy punkt ξ_n według rozkładu P_n i obliczamy:

$$x_{n+1} = \begin{cases} x_n, & \text{jeżeli } F(x_n + \xi_n) \geq F(x_n) - \epsilon\,, \\ x_n + \xi_n, & \text{jeżeli } F(x_n + \xi_n) < F(x_n) - \epsilon\,, \end{cases}$$

gdzie ϵ jest pewną stałą dodatnią.

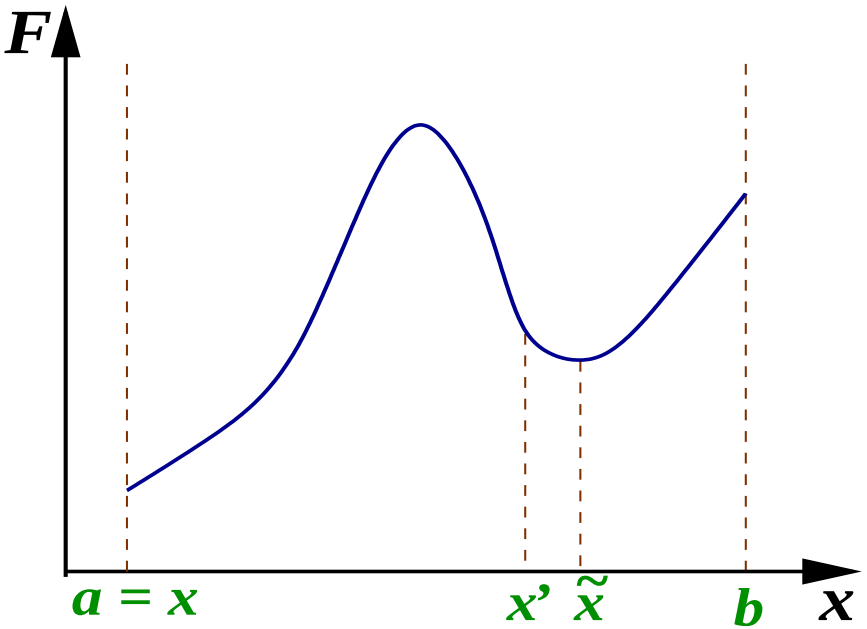
Zakładamy, że $\{\xi_n\}$ jest ciągiem niezależnych zmiennych losowych oraz, że $x_n + \xi_n \in X$.

Stosując powyższy algorytm otrzymujemy ciąg punktów $\{x_n\}$ takich, że:

$$F(x_1) \geq F(x_2) \geq F(x_3) \geq \ldots \geq F(x_n) \geq F(x_{n+1}) \geq \ldots .$$

▷ Jeżeli funkcja F jest ograniczona od dołu, to ciąg powyższy jest zbieżny.

▶ Czy ciąg $\{x_n\}$ jest również zbieżny, a jeżeli tak, to czy jego granicą jest punkt $x_{\rm opt}$?



Przykład funkcji posiadającej minimum lokalne w punkcie $\tilde{x}$.

▷ Przy odpowiednim wyborze rozkładów P_n, każdy ciąg zaczynający się np. w punkcie $x_1 = x'$ będzie zbieżny do punktu $\tilde{x}$, w którym funkcja F ma **minimum lokalne**, a nie do punktu $x_{\rm opt}$, w którym funkcja ta osiąga **minimum globalne** na zbiorze $X = [a, b]$!

Metody sekwencyjne: algorytm podstawowy

Algorytmy dające ciągi $\{x_n\}$ zbieżne do $x_{\rm opt}$, w którym funkcja F osiąga wartość najmniejszą nazywają się **algorytmami globalnymi**, natomiast algorytmy generujące ciągi $\{x_n\}$, z których niektóre są zbieżne do $x_{\rm opt}$, a inne do jednego z minimów lokalnych funkcji F zwane są **algorytmami lokalnymi**.

▷ Jeżeli w ciągu $\{x_n\}$ pojawi się punkt x', dla którego

$$F(x_{\rm opt}) < F(x') < F(x_{\rm opt}) + \epsilon\,,$$

to zgodnie w powyższym algorytmem będzie powtarzał się w ciągu $\{x_n\}$ nieskończenie wiele razy, tzn. dla takiego ciągu

$$\lim_{k\to\infty} x_k = x'\,.$$

$\longrightarrow$ Oczywiście, zmieniając wartość ϵ możemy dostawać różne wartości x'.

Niech:

$$A_\epsilon = \{x\colon F(x) < F(x_{\rm opt}) + \epsilon\}\,.$$

▶ Będziemy mówić, że ciąg $\{x_n\}$ jest zbieżny do $x_{\rm opt}$, jeżeli jest on zbieżny do jakiegokolwiek punktu zbioru A_ϵ.

Metody sekwencyjne: algorytm podstawowy

- **Twierdzenie Drimla–Hanša:**

Jeżeli funkcja F jest ograniczona od dołu i jeżeli

$$\forall_{x\in X} \text{ i } \forall_{\delta>0}: \mathcal{P}\{\, F(x+\xi_n) \le F(x_{\rm opt}) + \delta \,\} > 0\,,$$

to ciąg $\{x_n\}$ z prawdopodobieństwem 1 osiągnie zbiór A_ϵ.

▷ Tezę powyższego twierdzenia można też sformułować w następujący sposób:

Ciąg $\{x_n\}$ jest z prawdopodobieństwem 1 zbieżny do zmiennej losowej $\bar{x}$ takiej, że:

$$\mathcal{P}\{\, F(\bar{x}) < F(x_{\rm opt}) + \epsilon \,\} = 1\,.$$

▶ **Problem praktyczny:**

Założenie tw. Drimla–Hanša jest spełnione, gdy nośnikiem każdego rozkładu P_n jest cały zbiór X (skonstruowanie innego rozkładu jest trudne w praktyce).

$\longrightarrow$ Słaba wydajność algorytmu, ponieważ za każdym razem losujemy punkt z całego zbioru X!

▷ Lepszą wydajność uzyskalibyśmy losując punkt x_{n+1} w bliskim otoczeniu najlepszego z dotychczas wylosowanych punktów x_n, co jest równoważne losowaniu ξ_n blisko zera (w praktyce ξ_n losuje się często z rozkładu równomiernego na sferze o promieniu r_n).

Metody sekwencyjne: algorytm podstawowy

- **Twierdzenie Zielińskiego:**

Jeżeli dla każdego punktu $x \in X \setminus A_\epsilon$ (tzn. dla każdego punktu spoza zbioru A_ϵ) spełniony jest warunek:

$$\mathcal{P}\{\, F(x+\xi_n) < F(x) - \epsilon \,\} > 0\,,$$

to ciąg $\{x_n\}$ z prawdopodobieństwem 1 osiągnie zbiór A_ϵ.

▶ **Wnioski:**

- Rozkłady P_n nie mogą być zbyt ,,skupione" wokół zera.
- Jeżeli są na tyle ,,rozmyte", że dla każdego punktu x istotnie różnego od $x_{\rm opt}$ (tzn. spoza zbioru A_ϵ) istnieje dodatnie prawdopodbobieństwo wylosowania punktu x' takiego, że

$$F(x') < F(x) - \epsilon\,,$$

to warunek powyższego twierdzenia jest spełniony i generowane przez opisany algorytm ciągi $\{x_n\}$ są zbieżne do $x_{\rm opt}$ (w sensie sformułowanym powyżej).

▷ Porównamy pod względem **efektywności** powyższy sekwencyjny algorytm Monte Carlo ze standardową metodą gradientową.

Metoda gradientowa:

Niech

$$dF_j(x) = F(x^1, x^2, \ldots, x^{j-1}, x^j + h, x^{j+1}, \ldots, x^m) - F(x^1, x^2, \ldots, x^{j-1}, x^j, x^{j+1}, \ldots, x^m)$$

będzie przyrostem funkcji F przy zmianie jej j-tej współrzędnej z x^j ma $x^j + h$.

Różnicowy odpowiednik gradientu funkcji F definiujemy jako:

$$dF(x) = \beta \, [\, dF_1(x), dF_2(x), \ldots, dF_m(x) \,] \, ,$$

gdzie współczynnik β ustalamy tak, aby wektor $dF(x)$ miał jednostkową długość.

▶ Schemat obliczeń:

Jeżeli wyznaczyliśmy już punkty $x_1, x_2, \ldots, x_n$, to punkt x_{n+1} znajdujemy według wzoru:

$$x_{n+1} = x_n + dF(x_n) \, .$$

▷ Załóżmy, że w danym otoczeniu punktu x_n funkcja F może być dostatecznie dobrze aproksymowana pewną płaszczyzną (dla prostoty rozważań).

$\longrightarrow$ Punkt x_{n+1} znajduje się o jednostkę bliżej punktu $x_{\rm opt}$ niż punkt x_n.

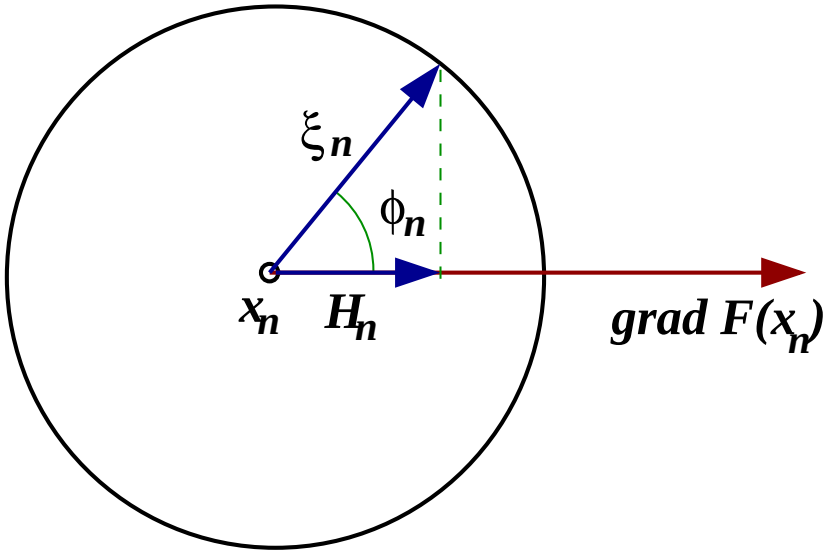
▶ W wyniku wykonania jednego kroku w metodzie gradientowej uzyskujemy przesunięcie w kierunku minimum funkcji F o jedną jednostkę.

▷ Jakie przesunięcie uzyskamy w sekwencyjnym algorytmie Monte Carlo?

Rozważmy wariant tego algorytmu, w którym $\epsilon = 0$, a punkt ξ_n losujemy według rozkładu równomiernego na sferze jednostkowej o środku w początku układu współrzędnych.

▶ Punkt x_{n+1} wyznaczamy według wzoru:

$$x_{n+1} = \begin{cases} x_n, & \text{jeżeli } F(x_n + \xi_n) \geq F(x_n)\,, \\ x_n + \xi_n, & \text{jeżeli } F(x_n + \xi_n) < F(x_n)\,. \end{cases}$$



▷ Przesunięcie H_n w kierunku minimum funkcji F jest zmienną losową:

$$H_n = \begin{cases} \cos\phi_n, & -\pi/2 \leq \phi_n \leq \pi/2\,, \\ 0, & \text{poza tym,} \end{cases}$$

czyli jest rzutem wektora ξ_n na kierunek gradientu funkcji F, jeżeli kąt między tymi wektorami jest bezwzględnie mniejszy od $\pi/2$.

▶ Powtarzając losowanie $(m+1)$ razy otrzymamy ciąg punktów:

$$x_n,\, x_{n+1},\, x_{n+2},\, \ldots,\, x_{n+m}$$

i odpowiadający mu ciąg przesunięć:

$$H_n,\, H_{n+1},\, H_{n+2},\, \ldots,\, H_{n+m}\,.$$

● **Przesunięciem jednostkowym w algorytmie stochastycznym** nazwiemy wielkość:

$$H(m) = \sum_{j=0}^{m} H_{n+j}\,.$$

▷ Wymaga takiej samej liczby obliczeń wartości funkcji F jak przesunięcie jednostkowe w metodzie gradientowej.

▶ **Wartość oczekiwana** tego przesunięcia wynosi:

$$E[H(m)] = (m+1)\,E[H_n] = (m+1)\int_{-\pi/2}^{\pi/2} \cos\phi\; w_m(\phi)\, d\phi\,,$$

gdzie $w_m(\phi)$ jest gęstością rozkładu kąta ϕ w przestrzeni m-wymiarowej.

▶ Ponieważ:

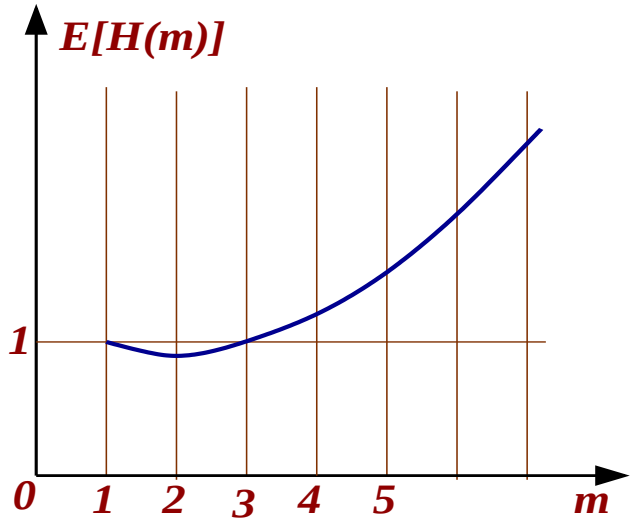
$$w_m(\phi) = C_m \, \sin^{m-2} \phi \, ,$$

to ostatecznie na **wartość oczekiwaną** otrzymujemy wyrażenie:

$$E[H(m)] = \frac{(m+1)\,\Gamma(m)}{2^m \, \Gamma^2\left(\frac{m+1}{2}\right)} \, .$$

▷ Podstawiając w powyższym wzorze kolejne liczby naturalne otrzymujemy:

$$E[H(1)] = 1, \, E[H(2)] = 3/\pi, \, E[H(3)] = 1, \, E[H(4)] = 10/3\pi, \, E[H(5)] = 9/8, \ldots .$$

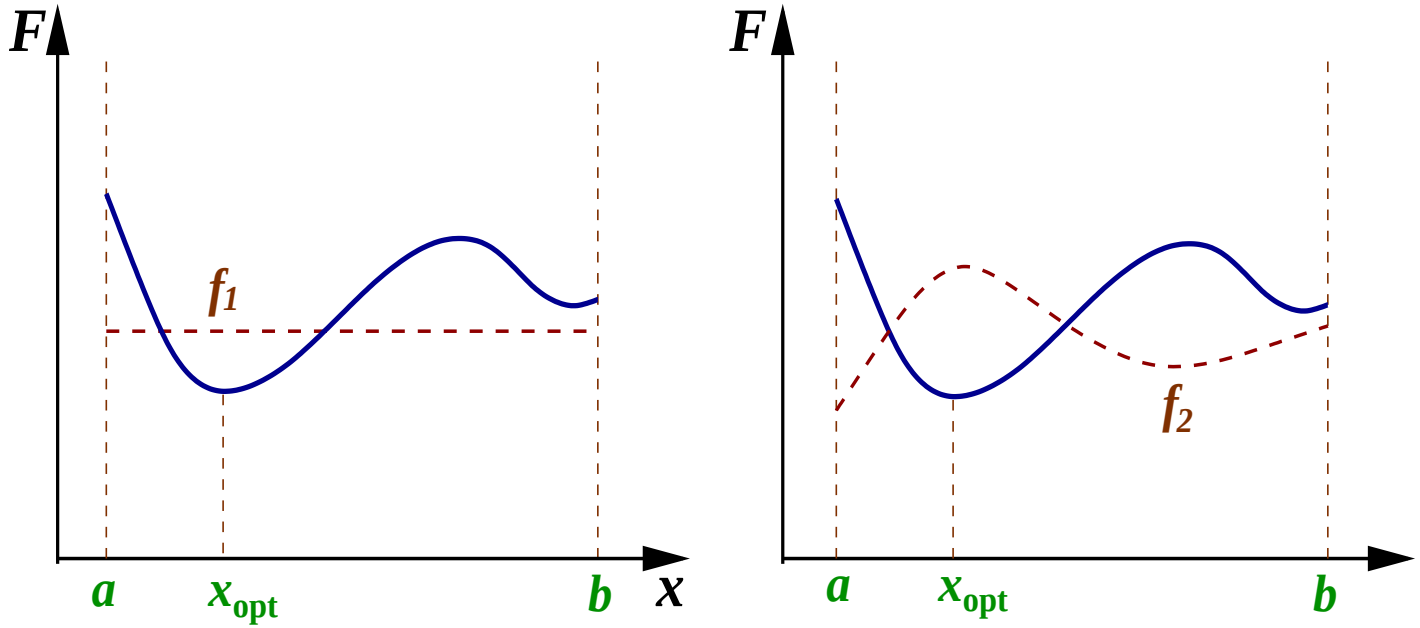


⇒ **Wartość oczekiwana przesunięcia w kierunku minumim funkcji F w powyższym algorytmie stochastycznym jest dla $m > 3$ większa niż przesunięcie w metodzie gradientowej!**

Metody sekwencyjne: algorytm Charłamowa

▷ Prosty przykład: Szukamy minimum funkcji F na przedziale $[a, b]$ używając dwóch metod:

1. Losujemy punkty na przedziale $[a, b]$ według rozkładu równomiernego o gęstości f_1.
2. Losujemy punkty na przedziale $[a, b]$ według rozkładu o gęstości f_2 skupionego wokół mniejszych wartości funkcji F.



▷ Intuicyjnie spodziewamy się, że losowanie według rozkładu f_2 daje większe szanse trafienia w pobliże punktu $x_{\rm opt}$, a zatem metoda nr 2 powinna być wydajniejsza.

▶ Formalnie wyraża się to poprzez wartości oczekiwane:

$$E_1 = \int_a^b F(x)\, f_1(x)\, dx \quad > \quad E_2 = \int_a^b F(x)\, f_2(x)\, dx\,.$$

Metody sekwencyjne: algorytm Charłamowa

Niech r będzie malejącą funkcją odwzorowującą zbiór wartości funkcji F w przedział $(0, 1)$.

Schemat obliczeń:

1. Położenie początkowe x_1 cząsteczki C wybieramy dowolnie.
2. Jeżeli w pewnej chwili n cząsteczka znajduje się w punkcie x_n, to jej położenie x_{n+1} w chwili $(n+1)$ wybieramy w następujący sposób:
   - Z prawdopodobieństwem $r(F(x_n))$ cząsteczka pozostaje w punkcie x_n, tzn. $x_{n+1} = x_n$.
   - Z prawdopodobieństwem $1 - r(F(x_n))$ cząsteczka przechodzi do nowej pozycji x_{n+1} w zbiorze X, którą losujemy według pewnego rozkładu prawdopodobieństwa.

▶ Bardziej formalnie:

Jeżeli w pewnej chwili cząsteczka C jest w punkcie $x \in X$, to z prawdopodobieństwem

$$P(S|x) = \begin{cases} r(F(x)) + [\,1 - r(F(x))\,]\,Q(S), & x \in S\,, \\ [\,1 - r(F(x))\,]\,Q(S), & x \in X \setminus S \end{cases}$$

znajdzie się w pewnym punkcie zbioru S.

Q jest ustalonym rozkładem prawdopodobieństwa (np. rozkładem równomiernym) na X.

Metody sekwencyjne: algorytm Charłamowa

Niech $x_1,\ x_2, \ldots$ będzie ciągiem kolejnych położeń cząsteczki C w zbiorze X, a $f_1(x),\ f_2(x), \ldots$ ciągiem rozkładów zmiennych losowych $x_1,\ x_2, \ldots$.

▶ Ciąg $\{x_n\}$ jest zbieżny do $x_{\rm opt}$ w tym sensie, że ciąg $\{f_n\}$ jest zbieżny do pewnego rozkładu granicznego f skupionego wokół punktu $x_{\rm opt}$.

▷ Jeżeli funkcja r zostanie tak dobrana, że $r(F(x_{\rm opt})) = 1$, to cały rozkład graniczny f skupiony jest w punkcie $x_{\rm opt}$.

Załóżmy, że mamy dwa algorytmy tego typu:

Niech $f^{(1)}$ – rozkład graniczny ciągów rozkładów $\{f_n\}$ w algorytmie pierwszym,
a $f^{(2)}$ – rozkład graniczny w algorytmie drugim.

▶ Algorytm pierwszy uznamy za bardziej efektywny, jeżeli:

$$\int_X F(x)\, f^{(1)}(x)\, dx \quad < \quad \int_X F(x)\, f^{(2)}(x)\,.$$

Algorytm Charłamowa jest procedurą sekwencyjnego budowania rozkładu granicznego skupionego wokół minimum funkcji F w oparciu o losowe błądzenie cząsteczki C po zbiorze X.

▷ Problemem może być odpowiedni wybór funkcji r.

Inne metody

- **Optymalizacja statystyczna:**

  W niektórych zagadnieniach optymalizacyjnych wartość funkcji może być obliczona w każdym punkcie, ale formuła opisująca funkcję jest nieznana.

  Np. wartości funkcji F są wartościami zaobserwowanymi empirycznie dla odpowiednio dobranych parametrów produkcyjnych $x^1, x^2, \ldots, x^m$.

  ▷ Patrz np.: R. Zieliński, *„Metody Monte Carlo”,* WNT 1970.

- **Algorytmy genetyczne** – stanowią przypadek szczególny stochastycznych algorytmów optymalizacji globalnej.

  Stosują one mechanizmy adaptacji rozkładu prawdopodobieństwa w kolejnych krokach iteracji zapożyczone z biologii – oparte na zjawiskach mutacji i krzyżowania kodów genetycznych oraz selekcji naturalnej organizmów żywych. Ukierunkowują one proces poszukiwania, czyniąc go bardziej efektywnym niż poszukiwanie całkowicie przypadkowe.

  ▶ Sprawdzają się w zastosowaniu do funkcji wielu zmiennych, skomplikowanych, nieregularnych itd.

  ▷ Więcej szczegółów np. w książce:

  R. Schaefer, *„Podstawy genetycznej optymalizacji globalnej”,* Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2002.

Minimalizacja formy kwadratowej

Zadanie rozwiązania układu równań liniowych:

$$\mathbf{A}\,\vec{x} = \vec{b}\,, \qquad \vec{x},\,\vec{b} \in \mathbb{R}^n, \quad \mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^n,$$

można traktować jako zadanie znalezienia **minimum** funkcji n zmiennych (**formy kwadratowej**):

$$f(\vec{x}) = f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij}\, x_j - b_i \right)^2 .$$

▶ **Najprostsza metoda Monte Carlo:**

1. Wybrać dowolnie punkt początkowy $\vec{x}^{(1)}$.
2. Jeżeli wyznaczone zostały już punkty $\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots, \vec{x}^{(k)}$, to wylosować punkt $\vec{\xi}$ według rozkładu równomiernego na sferze o środku w punkcie $\vec{x}^{(k)}$ i promieniu $r^{(k)}$ ($r^{(k)} \underset{k\to\infty}{\longrightarrow} 0$).
3. Wyznaczyć:

$$\vec{x}^{(k+1)} = \begin{cases} \vec{\xi}, & \text{gdy } f(\vec{\xi}) < f(\vec{x}^{(k)})\,, \\ \vec{x}^{(k)}, & \text{gdy } f(\vec{\xi}) \geq f(\vec{x}^{(k)})\,. \end{cases}$$

▶ Jeżeli spełnione są warunki tw. Drimla–Hanša lub tw. Zielińskiego, to tak skonstruowany ciąg punktów $\{\vec{x}^{(k)}\}$ jest z prawdopodobieństwem 1 zbieżny do rozwiązania $\vec{x}^{0}$.