Stochastyczne algorytmy optymalizacji globalnej

Ryszard Zieliński

Instytut Matematyczny PAN
ul. Śniadeckich 8, 00-950 Warszawa, Skr.poczt. 137
e-mail: rziel@impan.gov.pl

Streszczenie. W wykładzie przedstawiono różne podejścia do oceny dokładności rozwiązania, uzyskanego za pomocą podstawowych algorytmów optymalizacji globalnej. Wykład ma charakter przeglądowy i przypomina te różne podejścia w pewien usystematyzowany sposób. Oryginalne pomysły i nigdzie jeszcze nie publikowane twierdzenie podane jest tylko w ostaniej części wykładu. W pozycji LITERATURA podane jest tylko kilka prac, ale przez literaturę podaną z kolei w tych pracach można bez trudności dotrzeć do całej literatury na interesujący nas temat. Najnowsza chronogicznie pozycja Pintera (1996) zawiera aktualny wykład przedmiotu.

Wstęp

Problem jest następujący: dla danej funkcji $f : \mathcal{X} \to \mathcal{R}^1$ znaleźć liczbę $f^* = min_{x\in\mathcal{X}} f(x)$ oraz/lub punkt $x^* = argmin_{x\in\mathcal{X}} f(x)$, przyczym min jest tu rozumiane jako minimum globalne; chodzi więc nam o taką liczbę f^*, żeby $(\forall x \in \mathcal{X}) f(x) \geq f^*$ oraz o taki punkt, żeby $(\forall x \in \mathcal{X}) f(x) \geq f(x^*)$.

Będę zajmował się tylko niektórymi algorytmami stochastycznymi, a mianowicie takimi algorytmami, w których potrafimy oszacować dokładność uzyskanego rozwiązania lub nawet tak nimi sterować, żeby uzyskać z góry zaplanowaną dokładność. Przedstawię tylko trzy podejścia do oceny dokładności: prostą metodę Monte Carlo, statystyczną estymację w zadaniach wieloekstremalnych oraz zagadnienie szacowania f^* ze z góry zadaną dokładnością. Zgodnie z naturą algorytmów stochastycznych, wszystkie te oszacowania będą oczywiście sformułowane w terminach stochastycznych.

Prosta metoda Monte Carlo

Metoda pojawiła się po raz pierwszy w pracy Brooksa (1958) i była szerzej prezentowana i dyskutowana w dwóch, prawie jednocześnie opublikowanych, pracach Hooke'a i Jeevsa (1958) oraz Brooksa (1959); bardziej współczesną i łatwiej dostępną prezentację można znaleźć w minimonografii Zielińskiego i Neumanna (1986). Dokładny opis metody i sformułowanie twierdzenia o dokładności uzyskiwanych za jej pomocą wyników jest następujące.

Zbiór $\mathcal{X}$ argumentów minimalizowanej funkcji f traktujemy jako zbiór zdarzeń elementarnych pewnej przestrzeni probabilistycznej. Wymaga to albo wprowadzenia pewnych ograniczeń na ten zbiór, albo wprowadzenia odpowiedniej definicji miary. Wprowadzimy "odpowiednią" miarę. Nazwiemy ją λ, a jej "odpowiedniość" będzie polegała na tym, że liczba $\lambda(\mathcal{X})$ będzie dodatnia i skończona. Jeżeli zbiór $\mathcal{X}$ jest pewnym ograniczonym obszarem w (wielowymiarowej) przestrzeni $\mathcal{R}^m$, to za λ możemy na przykład wziąć zwykłą miarę Lebesgue'a (długość przedziału na prostej, pole obszaru na płaszczyźnie, itd). Jeżeli $\mathcal{X}$ nie jest ograniczonym obszarem, np. gdy jest całą przestrzenią $\mathcal{R}^m$, za λ możemy wziąć dowolny rozkład prawdopodobieństwa na tym zbiorze, np. rozkład normalny. Jeżeli $\mathcal{X}$ jest zbiorem skończonym (co jest typowe dla zadań kombinatorycznych, np. zadania komiwojażera), to może to być zwykła miara licząca: wtedy prawdopodobieństwo podzbioru $A \subset \mathcal{X}$ jest równe ilorazowi liczby punktów w zbiorze A i liczby punktów w zbiorze $\mathcal{X}$. Uogólniając pojęcie rozkładu równomiernego będziemy mówili, że losowy punkt X ma rozkład równomierny na zbiorze $\mathcal{X}$, jeżeli dla $A \subset \mathcal{X}$ prawdopodobieństwo zdarzenia $\{X \in A\}$ wyraża się wzorem

$$P\{X \in A\} = \frac{\lambda(A)}{\lambda(\mathcal{X})}$$

Wszędzie dalej mówiąc o rozkładzie równomiernym będziemy mieli na myśli właśnie ten rozkład. Dla $P\{X \in A\}$, co jak zwykle w teorii prawdopodobieństwa będziemy zapisywali także w postaci $P\{A\}$, będziemy używali również nazwy "miara względna zbioru A".

Prosta metoda Monte Carlo polega na tym, że ze zbioru $\mathcal{X}$ losujemy według rozkładu równomiernego i w sposób niezależny, N punktów $X_1, X_2, \ldots, X_N$. W każdym z tych punktów liczymy wartość funkcji f. Oznaczmy

$$f_N^* = min\{f(X_1), f(X_2), \ldots, f(X_N)\}$$

Za rozwiązanie zadania przyjmujemy ten punkt X_N^*, dla którego

$$f(X_N^*) = f_N^*$$

Jak dokładne jest takie rozwiązanie?

Weźmy pod uwagę zbiór tych punktów $x \in \mathcal{X}$, w których minimalizowana funkcja f ma wartość nie większą niż znaleziona wartość f_N^*. Naturalną miarą dokładności rozwiązania jest miara względna tego zbioru, tzn.

$$\frac{\lambda\{x : f(x) \leq f(X_N^*)\}}{\lambda\{\mathcal{X}\}}$$

Rozwiązanie jest oczywiście tym lepsze, im mniejsza jest ta liczba, a gdy jest ona równa zeru, to rozwiązanie jest tak dobre, że prawdopodobieństwo znalezienia lepszego punktu w zbiorze $\mathcal{X}$ jest równe zeru.

Ustalmy dowolnie (małą) liczbę $\varepsilon > 0$, zdefiniujmy

$$f_\varepsilon = \sup\left\{t : \frac{\lambda\{x : f(x) \leq t\}}{\lambda\{\mathcal{X}\}} \leq \varepsilon\right\}$$

i oznaczmy

$$S_\varepsilon = \{x : f(x) \leq f_\varepsilon\}$$

Jeżeli $\{X_N^* \in S_\varepsilon\}$, to powiemy, że minimum globalne funkcji f zostało zlokalizowane z dokładnością do zbioru o mierze względnej nie większej niż ε.

Zauważmy jednak, że $\{X_N^* \in S_\varepsilon\}$ jest zdarzeniem losowym. Zadanie zaprojektowania algorytmu gwarantującego zadaną dokładność polega na tym, żeby dla danych liczb ε oraz γ wyznaczyć N w taki sposób, żeby minimum globalne funkcji zostało z prawdopodobieństwem równym co najmniej γ zlokalizowane z dokładnością do zbioru o mierze względnej równej co najwyżej ε. Jest to chyba jedyna możliwa interpretacja dokładności rozwiązania uzyskanego za pomocą prostej metody Monte Carlo.

Odpowiedź jest bardzo łatwa: N powinno być dowolną liczbą całkowitą, większą od $log(1-\gamma)/log(1-\varepsilon)$. Np. dla $\varepsilon = 0.001$ oraz $\gamma = 0.999$ otrzymujemy $N = 6905$.

Zalety prostej metody Monte Carlo i uzyskanego za jej pomocą rozwiązania polegają na tym, że metoda może być stosowana praktycznie dla dowolnych zbiorów $\mathcal{X}$ (skończonych lub nie), dla dowolnych (matematyk by się zastrzegł: dowolnych mierzalnych) funkcji f, i przy dowolnych wymiarach zadania.

Wady polegają na tym, że o bliskości rozwiązania X_N^* do punktu, w którym funkcja f ma minimum globalne, możemy mówić tylko w terminach wartości funkcji f ("semimetryki generowanej przez funkcję f"). Ponadto to, czy "małe" ε jest rzeczywiście małe w olbrzymim stopniu zależy od wymiaru rozważanego zadania; bardziej szczegółowe komentarze na ten temat, sformułowane jeszcze przez Hooksa i Jeeves (1958), można znaleźć w cytowanej już książce Zielińskiego i Neumanna (1986).

Zadanie wieloekstremalne

Wybierzmy sobie jakiś lokalny algorytm (deterministyczny lub stochastyczny) minimalizacji funkcji metodą kolejnych przybliżeń. Rozumiemy to w ten sposób, że jeżeli w pewien sposób wybierzemy punkt startowy, to kolejno poprawiając ten punkt za pomocą wybranego algorytmu lokalnego, osiągniemy minimum lokalne funkcji. Startując z innego punktu, osiągniemy być może to samo, a być może inne minimum lokalne. Jeżeli wybierzemy losowo dużo punktów startowych, to możemy mieć nadzieję, że odkryjemy wszystkie minimam lokalne; wtedy minimum lokalne o najmniejszej wartości funkcji będzie oczywiście minimum globalnym funkcji.

Przypuśćmy, że funkcja f ma (nieznaną liczbę) m minimów lokalnych x_1^*, $x_2^*, \ldots, x_m^*$. Przy ustalonym algorytmie lokalnym, cały zbiór $\mathcal{X}$ zostaje podzielony na m podzbiorów przyciągania poszczególnych minimów lokalnych: ilekroć wystartujemy z pewnego punktu obszaru przyciągania k-tego minimum lokalnego

x_k^*, tylekroć algorytm lokalny doprowadzi nas do tego samego punku x_k^*. Oznaczmy przez $\theta_1, \theta_2, \ldots, \ldots \theta_m$ miary względne obszarów przyciagania poszczególnych minimów lokalnych. Umówmy się, że losujemy N niezależnych punktów startowych o jednakowym rozkładzie równomiernym (w wyżej opisanym sensie) na zbiorze $\mathcal{X}$. Rozumiemy to w ten sposób, że prawdopodobieństwo wylosowania punktu z k-tego obszaru przyciągania, a więc i również prawdopodobieństwo wykrycia k-tego minimum lokalnego, jest równe θ_k. Powstają naturalne pytania: 1) jeżeli losujemy w opisany wyżej sposób N punktów startowych, to ile wynosi prawdopodobieństwo odkrycia wszystkich minimów lokalnych oraz 2) jak duże powinno być to N, żeby prawdopodobieństwo wykrycia wszystkich minimów lokalnych (a więc również dokładnego znalezienia minimum globalnego) było równe co najmniej pewnej z góry zadanej, bliskiej jedności liczbie γ?

Odpowiedź na pierwsze pytanie uzyskuje się za pomocą elementarnego rachunku prawdopodobieństwa, ale zależy ona oczywiście od $(\theta_1, \theta_2, \ldots, \theta_m)$. Z powodu tej właśnie zależności, odpowiedź na drugie pytanie nie jest łatwa. Łatwo jest natomiast udowodnić, że dla każdego, dowolnie małego ε i dla każdego, dowolnie dużego N, istnieje takie $(\theta_1, \theta_2, \ldots, \theta_m)$, że prawdopodobieństwo wykrycia wszystkich minimów lokalnych jest mniejsze od tego ε. Wynika to stąd, że jeżeli obszar przyciągania jakieś minimum lokalnego jest bardzo mały, to nawet przy dużym N możemy mieć znikome szanse na trafienie weń swoim losowym punktem startowym. Tę trudność można ominąć w akceptowalny z punktu widzenia zastosowań sposób, ograniczając zainteresowania tylko do minimów lokalnych o niezbyt małych obszarach przyciągania. Jeżeli ustalimy dowolnie małą liczbę δ i przez κ oznaczymy najmniejszą liczbę całkowitą nie mniejszą od $1/\delta$, to najmniejsze N, z prawdopodobieństwem γ gwarantujące wykrycie wszystkich minimów lokalnych o obszarach przyciągania większych niż δ, otrzymuje się ze wzoru

$$\sum_{j=0}^{\kappa} (-1)^j \binom{\kappa}{j} \left(1 - \frac{j}{\kappa}\right)^N \geq \gamma, \quad N \geq \kappa$$

To rozwiązanie uzyskano jednak kosztem nałożenia pewnych warunków na liczbę wykrywanych minimów lokalnych, a więc pośrednio pewnych warunków na liczbę takich minimów w oryginalnym zadaniu. Stąd już tylko mały krok do sformułowania tego typu ograniczeń w terminach rozkładu apriori liczby minimów i rozkładu apriori parametru $(\theta_1, \theta_2, \ldots, \theta_m)$. Paradygmat statystyki bayesowskiej pozwala tu na wprowadzenie do rozważań kosztów związanych z losowaniem punktów i obliczaniem wartości minimalizowanej funkcji f w tych punktach. Szczególnie to ostatnie jest bardzo ważne, bo w praktycznych zastosowaniach pojawiają się problemy minimalizacji funkcji, która nie jest zadana za pomocą formuły matematycznej, ale której wartości można obliczać na przykład za pomocą odpowiednich eksperymentów w skali laboratoryjnyj lub przemysłowej. Znane są również rozwiązania odpowiednich problemów także w sytuacji, gdy te oszacowania wartości minimalizowanej funkcji są obarczone błędem losowym. Nie będę tutaj rozwijał tej problematyki, a zainteresowanych odsyłam do wspominanej już minimonografii Zielińskiego i Neumanna (1986).

Algorytmy hybrydowe

Algorytmy globalne typu prostej metody Monte Carlo pozwalają na oszacowanie błędu rozwiązania, przynajmniej w takim sensie, o jakim mówiliśmy w pierwszej części. Algorytmy lokalne (wszelkiego rodzaju metody kolejnych przybliżeń, np. typu *simulated annealing*, zwykle nie dają takich możliwości, a wszystko, co się czasami daje tu uzyskać to twierdzenia o zbieżności, co z punktu widzenia praktycznych zastosowań jest jednak mało użyteczne. Wysiłkom skierowanym na uzyskanie oszacowań dokładności rozwiązania towarzyszy zazwyczaj wprowadzanie różnych założeń o funkcji f. Okazuje się jednak, że gdy tych założeń jest zbyt dużo, efektywniejsze są metody deterministyczne, patrz np. Nowak (1988). Na tym tle obiecująco prezentują się różnego rodzaju algorytmy hybrydowe, będące pewną kombinacją algorytmów lokalnych z algorytmami czysto globalnymi. Najprostszy schemat takiego algorytmu ma postać ciągu kolejnych przybliżeń postaci

$$x_{n+1} = \begin{cases} \xi_n, & \text{jeżeli } f(\xi_n) < f(x_n), \\ x_n, & \text{w przeciwnym przypadku} \end{cases}$$

gdzie "kandydat" ξ_n na kolejne, lepsze przybliżenie, jest punktem losowym w zbiorze $\mathcal{X}$ o rozkładzie prawdopodobieństwa $Q_n(\cdot)$ postaci

$$Q_n(\cdot) = \alpha_n Q_{n,1}(\cdot) + (1 - \alpha_n) Q_{n,2}(\cdot | \mathcal{A}_n)$$

Tutaj $Q_{n,1}(\cdot)$ jest ciągiem rozkładów prawdopodobieństwa odpowiedzialnym za globalne przeszukiwanie zbioru $\mathcal{X}$, natomiast $Q_{n,2}(\cdot|\mathcal{A}_n)$ jest ciągiem rozkładów warunkowych, warunkowanych całą dotychczasową historią procesu kolejnych przybliżeń, opisaną przez σ-ciało $\mathcal{A}_n$ (w szczególnym przypadku deterministycznych algorytmów lokalnych, tzn. rozkładów zdegenerowanych). Tego typu algorytmy były już analizowane w książce Zielińskiego i Neumanna (1986), a zachęcające eksperymenty z taką globalizacją algorytmu *simulated annealing* są opisane w pracy Zielińskiego (1993).

Estymacja z zadaną precyzją

Jeżeli $X_1, X_2, \ldots, X_n, \ldots$ jest ciągiem niezależnych punktów losowych o rozkładzie równomiernym na zbiorze $\mathcal{X}$, to ciąg $Y_1 = f(X_1)$, $Y_2 = f(X_2), \ldots, Y_n = f(X_n), \ldots$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie, przy czym f^* jest dolną granicą nośnika tego rozkładu. Ciąg

$$Y_{(n)} = min\{Y_1, Y_2, \ldots, Y_n\}$$

zbiega z prawdopodobieństwem 1 do f^*. Oznacza to, że z prawdopodobieństwem 1 dla każdego ε istnieje taka liczba N_ε, że jeżeli $n > N_\varepsilon$, to $Y_{(n)} < f^* + \varepsilon$.

Wyobraźmy teraz sobie, że na równoległych procesorach możemy zrealizować pewną liczbę, powiedzmy k, niezależnych kopii ciągu $X_1, X_2, \ldots, X_n, \ldots$, a w konsekwencji k niezależnych kopii ciągu $Y_{(n)}$, $n = 1, 2, \ldots$ Oznaczmy te kopie przez

$Y^j_{(n)}$, $n = 1, 2, \ldots$, $j = 1, 2, \ldots, k$. Każda kopia $Y^j_{(n)}, j = 1, 2, \ldots, k$, z prawdopodobieństwem 1 zbiega do f^*, a więc wszystkie te kopie muszą się kiedyś spotkać. Mówiąc dokładniej: oznaczmy przez ρ_n jakąś miarę rozrzutu liczb $Y^j_{(n)}$, np. niech

$$\rho_n = max\{Y^j_{(n)}, j = 1, 2, \ldots, k\} - min\{Y^j_{(n)}, j = 1, 2, \ldots, k\}$$

Oczywiście $\rho_n \to 0$ z prawdopodobieństwem 1, i jeżeli wszystkie kopie znajdą się w końcu w ε-otoczeniu punktu f^*, to wtedy również ρ_n będzie mniejsze od ε.

Liczby f^* nie znamy, więc nie potrafimy powiedzieć, jak jeszcze daleko są od niej procesy kolejnych przybliżeń $Y^j_{(n)}, j = 1, 2, \ldots, k$. Ale liczby ρ_n możemy obserwować. Czy z faktu, że ρ_n jest już bardzo małe wynika, że procesy $Y^j_{(n)}, j = 1, 2, \ldots, k$, zbliżyły się już do swojej granicy f^*, a więc że estymator

$$f^*_n = min\{Y^j_{(n)}, j = 1, 2, \ldots, k\}$$

już dobrze przybliża poszukiwaną liczbę f^* ? Okazuje się, że czasami można na to odpowiedzieć pozytywnie.

Zdefiniujmy nastepującą regułę zatrzymania:

$$N. = min\{n \geq n_0 : \rho_n \leq \delta\}$$

gdzie $n_0 \in \{1, 2, \ldots, \}$ oraz $\delta > 0$ są pewnymi ustalonymi liczbami.

Oznaczmy przez F_Y dystrybuantę rozkładu zmiennej losowej $Y = f(X)$, gdzie X jest punktem losowym o rozkładzie równomiernym na zbiorze $\mathcal{X}$.

Można udowonić następujące lemat i twierdzenie:

Lemat. Dla ustalonych $n_0 \in \{1, 2, \ldots, \}$ oraz $\delta > 0$, prawdopodobieństwo $P\{N < \infty\} = 1$.

Twierdzenie. Jeżeli spełnione jest następujące założenie:

$$(Z) \qquad \exists (c > 0) \quad \eta(c) = \sup_{-\infty < y < \infty} (F_Y(y + c) - F_Y(y)) < \tfrac{1}{2}$$

to dla każdego $\varepsilon > 0$ i dla każdego $\gamma \in (0, 1)$ istnieje reguła zatrzymania N (tzn. istnieją n_0, k, δ) taka, że

$$P\{f^*_N - f^* \leq \varepsilon\} > 1 - \gamma$$

Lenat orzeka, że proponowa procedura równoległych obliczeń z prawdopodobieństwem 1 zakończy się w skończonym czasie, a twierdzenie, że z zadanym z góry prawdopodobieństwem błąd oszacowania minimum globalnego f^* funkcji f nie przekroczy zadanej z góry liczby ε. Ponieważ ani lemat ani twierdzenie nie były jeszcze nigdzie publikowane, podamy krótki szkic dowodu.

Dowód lematu (*Szkic*)

$$\begin{aligned} P\{N=\infty\} &= P\left(\bigcap_{n=n_0}^{\infty}\{\rho_n > \delta\}\right) \\ &= \lim_{M\to\infty} P\{\rho_{n_0} > \delta, \rho_{n_0+1} > \delta, \ldots, \rho_M > \delta\} \\ &\leq \lim_{M\to\infty} P\{\rho_M > \delta\} \\ &\leq \lim_{M\to\infty} P\left(\max\{Y_{(M)}^{(1)}, \ldots, Y_{(M)}^{(k)}\} > f^* + \delta\right) \\ &= \lim_{M\to\infty}\left[1 - \left(1 - \{1 - F_Y(f^*+\delta)\}^M\right)^k\right] = 0 \end{aligned}$$

□

Dowód twierdzenia (*Szkic*)

Ustalmy $\delta > 0$ takie, żeby $\eta(\delta) < \frac{1}{2}$. Wybierzmy n_0 oraz t (naturalne) w taki sposób, żeby

$$2 \leq n_0 \leq t < \frac{1}{\eta(\delta)}$$

Oznaczmy

$$P_1 = \sum_{n=n_0}^{t} P\{f_n^* - f^* > \varepsilon, N = n\}, \quad P_1 = \sum_{n=t}^{\infty} P\{f_n^* - f^* > \varepsilon, N = n\}$$

i zauważmy, że

$$P\{f_N^* - f^* > \varepsilon\} = \sum_{n=n_0}^{\infty} P\{f_n^* - f^* > \varepsilon, N = n\} = P_1 + P_2$$

Idea dowodu polega na tym, żeby wykazać, iż przy założeniach twierdzenia, $P_1 < \gamma/2$ oraz $P_2 < \gamma/2$. Fakt, że $P_1 < \gamma/2$ oznacza, iż proces nie zatrzyma się zbyt szybko, a fakt $P_2 < \gamma/2$, że proces nie zatrzyma się zbyt daleko od f^*. Do udowodnienia tych faktów wystarczą bardzo grube oszacowania.

Dla P_1 mamy oszacowanie

$$\begin{aligned} P_1 &\leq \sum_{n=n_0}^{t} P\{N = n\} = P\{N \leq t\} = \\ &= P\left\{\bigcup_{n=n_0}^{t}\{\rho_n \leq \delta\}\right\} \leq \sum_{n=n_0}^{t} P\{\rho_n \leq \delta\} \end{aligned}$$

Z definicji ρ_n jest rozstępem w "próbie losowej" $Y_{(n)}^{(1)}, Y_{(n)}^{(2)}, \ldots, \ldots Y_{(n)}^{(k)}$, w której niezależne i jednakowo rozłożone obserwacje $Y_{(n)}^{(j)}$ mają rozkład o dystrybuancie określonej wzorem

$$G_n(y) = P\{Y_{(n)}^{(1)} \leq y\} = 1 - [1 - F_Y(y)]^n$$

Zatem

$$P\{\rho_n \leq \delta\} = k \int_0^\infty [G_n(x+\delta) - G_n(x)]^{k-1}\, dG_n(x)$$

Dla różnicy w nawiasie klamrowym pod znakiem całki mamy

$$\begin{aligned} G_n(x+\delta) - G_n(x) &\leq n\,(F_Y(x+\delta) - F_Y(x))\,[1 - F_Y(x)]^{n-1} \\ &\leq n \cdot \eta(\delta)\,[1 - F_Y(x)]^{n-1} \end{aligned}$$

Zatem

$$\begin{aligned} P\{\rho_n \leq \delta\} &\leq k n^k \eta^{k-1}(\delta) \int_0^\infty [1 - F_Y(x)]^{k(n-1)}\, dF_Y(x) \\ &= \frac{k n^k \eta^{k-1}(\delta)}{k(n-1)+1} \end{aligned}$$

Stąd

$$P_1 \leq \sum_{n=n_0}^{t} \frac{k n^k \eta^{k-1}(\delta)}{k(n-1)+1} \leq [t\eta(\delta)]^{k-1} \sum_{n=n_0}^{t} \frac{n}{n-1+\frac{1}{k}}$$

a szacując ostatnią sumę z góry przez $2t$ otrzymujemy

$$P_1 \leq 2t\,[t\eta(\delta)]^{k-1}$$

Korzystając z tego, że

$$f_n^* = \min_{1 \leq j \leq k} \min_{1 \leq i \leq n} Y_i^{(j)}$$

otrzymujemy

$$P\{f_n^* - f^* > \varepsilon\} = [1 - F_Y(f^* + \varepsilon)]^{kn}$$

i stąd oszacowanie

$$\begin{aligned} P_2 &= \sum_{n=t+1}^{\infty} P\{f_n^* - f^* > \varepsilon, N = n\} \leq \sum_{n=t+1}^{\infty} P\{f_n^* - f^* > \varepsilon\} \\ &\leq \sum_{n=t+1}^{\infty} \left([1 - F_Y(f^* + \varepsilon)]^k\right)^n = \frac{[1 - F_Y(f^* + \varepsilon)]^{k(t+1)}}{1 - [1 - F_Y(f^* + \varepsilon)]^k} \end{aligned}$$

Ostatecznie

$$P\{f_N^* - f^* > \varepsilon\} \leq 2t\,[t\eta(\delta)]^{k-1} + \frac{[1 - F_Y(f^* + \varepsilon)]^{k(t+1)}}{1 - [1 - F_Y(f^* + \varepsilon)]^k}$$

Dla ustalonych δ oraz t, takich że $0 < t\eta(\delta) < 1$, zawsze można znaleźć k takie, żeby

$$P\{f_n^* - f^* > \varepsilon\} < \gamma$$

co kończy dowód twierdzenia. □

Przedstawiona tu idea wykorzystania równoległych obliczeń do uzyskania zadanej z góry dokładności rozwiązania jest bardzo atrakcyjna, ale niestety na razie wynik ma charakter tylko teoretyczny (twierdzenie egzystencjalne).

Literatura

Brooks, S.H. (1958): *A discussion of random methods for seeking maxima.* Opns. Res. 6, 244–251

Brooks, S.H. (1959): *A comparison of maxima–seeking methods.* Opns. Res. 7, 430–457

Hooke, R., Jeeves, T.A. (1958): *Comments on Brooks' discussion of random methods.* Opns. Res. 6, 881–882

Novak, E. (1988): *Deterministic and Stochastic Error Bounds in Numerical Analysis.* Springer-Verlag Lecture Notes in Mathematics 1349

Pinter, J.D. (1996): *Global Optimization in Action.* Kluwer Academic Press

Zieliński, R. i Neumann, P. (1986): *Stochastyczne metody poszukiwania minimum funkcji.* WNT, Warszawa

Zieliński, R. (1993): *Records of simulated annealing.* In: *Model-Oriented Data Analysis*, W.G.Müller, H.P.Wynn, A.A.Zhigljavsky (Eds.). Physika-Verlag, Springer-Verlag Company, Contributions to Statistics, pp.241–247