Artykuł

Stochastyczne algorytmy optymalizacji z perspektywy układów dynamicznych

Jerzy Ombach, Dawid Tarłowski

Streszczenie

W artykule tym prezentujemy nową metodologię, która może być używana do stwierdzania zbieżności szerokiej klasy algorytmów stochastycznej optymalizacji globalnej. Podstawową umiejętnością jest tutaj identyfikacja odpowiedniego układu dynamicznego na stosownie dobranej przestrzeni miar i skorzystanie z odpowiednich twierdzeń z teorii układów dynamicznych. Naszym celem jest nauczenie za pomocą – najpierw prostych – przykładów jak się to robi. Stosunkowo nowym wynikiem jest Wniosek 1.4.5 podający warunki dostateczne zbieżności algorytmu Simulated Annealing.

1.1. Wstęp

Powszechna dostępność komputerów oraz stale wzrastająca ich zdolność obliczeniowa spowodowały, że znane metody rozwiązywania niektórych klasycznych zagadnień zostają zastępowane przez nowe metody, które umieją wykorzystać nowe możliwości. W szczególności są to metody stochastyczne. Tak właśnie jest w zadaniach optymalizacyjnych [9], gdzie w ostatnich latach pojawiły się metody mające stochastyczny charakter. Przykłady takich metod czytelnik znajdzie w pozycjach takich jak [1], [2], [4], [5], [7], [8], [19]. W artykule tym nie zamierzamy wymieniać ich zalet ani wad. Natomiast zwracamy uwagę, że większość metod optymalizacji stochastycznej ma charakter heurystyczny: nasza intuicja oparta nieraz na wielu doświadczeniach zazwyczaj podpowiada, że metoda będzie w praktyce skuteczna i rzeczywiście często tak jest. Nie powinno to jednak zwalniać badaczy, zwłaszcza matematyków, od sprawdzania, czy używane heurystyki dają metody naprawdę zbieżne do rozwiązania optymalnego.

Tym bardziej, że zdarzają się pomyłki. Na przykład przez kilka lat utrzymywał się pogląd, że algorytm optymalizacji rojem cząstek (Particle Swarm Optimization, PSO) jest w zasadzie zbieżny. Tymczasem okazuje się, że przy pewnych dość naturalnych założeniach tak nie jest, patrz na przykład [4].

Niniejszy artykuł ma więc na celu zasygnalizowanie, że można z sukcesem dowodzić zbieżność wielu algorytmów stochastycznej optymalizacji. Chcemy pokazać, a nawet nauczyć jak to się robi. W tym celu przedstawimy kilka wybranych, najpierw prostych, potem bardziej złożonych algorytmów i na ich przykładzie zademonstrujemy jak działa nasza „maszynka dowodowa", rozwijana i wykorzystywana w serii prac [10], [11], [12], [13], [16], [17], [18].

Problemem optymalizacyjnym, który tutaj rozważamy jest znalezienie punktów w których dana funkcja osiąga swoje minimum globalne m. Dany jest więc zbiór $A \subset \mathbb{R}^n$ oraz funkcja $f : A \to \mathbb{R}$. Przez $A^\star$ oznaczamy zbiór rozwiązań problemu, czyli $A^\star = \operatorname{argmin} = \{x \in A : f(x) = m\}$. Będziemy zakładać, że jest to zbiór niepusty, co ma na przykład miejsce w przypadku, gdy A jest zbiorem zwartym, a f funkcją ciągłą. Będziemy rozważać te algorytmy, których matematycznym modelem jest formuła:

$$X_t = T_t(X_{t-1}, \mathbf{Y}_{t-1}) \text{ dla } t = 1, 2, 3, \ldots \tag{1.1}$$

gdzie T_t określają mechanizm algorytmu, a $\mathbf{Y}_t$ odpowiadają za losowość, czyli są zmiennymi (wektorami) losowymi. Chcemy odpowiedzieć na pytanie, jakie założenia gwarantują zbieżność ciągu X_t do zbioru $A^\star$. Naszym podstawowym pomysłem jest przeformułowanie związku (1.1) na odpowiedni związek na przestrzeni miar i skorzystanie z już dobrze rozwiniętej teorii układów dynamicznych celem udowodnienia zbieżności, a następnie zinterpretowanie wyniku w kontekście wektorów losowych. Bardzo ważną umiejętnością jest sprowadzanie konkretnego algorytmu do postaci (1.1).

Zaczniemy jednak od zademonstrowania możliwości jakie daje nam Lemat Borela-Cantellego, który oznacza czysto probabilistyczne podejście, a dopiero później pokażemy jak można zrealizować nasz pomysł w oparciu o teorię układów dymnamicznych. Zaczniemy od najprostszych przypadków, gdzie dość dokładnie omówimy podstawowe mechanizmy, a skończymy na zasygnalizowaniu sytuacji bardziej złożonych, w których jednak te mechanizmy dalej działają.

Z punktu widzenia praktyków, chyba ważniejszym zadaniem niż dowodzenie zbieżności, byłoby określenie tempa zbieżności poszczególnych algorytmów. Takie próby są podejmowane, na przykład [2] lub [15], lecz niestety jedynie dla bardzo wąskiej klasy metod lub przy bardzo restrykcyjnych założeniach dotyczących funkcji f. Autorzy niniejszego opracowania mają jednak nadzieję, że wykorzystanie narzędzi podobnych do tych stosowanych przez nas w dowodach zbieżności mogłoby pomóc także w ocenie tempa zbieżności. W szczególności, ciekawe wyniki może przynieść zastosowanie stochastycznych modyfikacji funkcji Lapunowa używanej w niniejszej pracy.

1.2. Pure Random Search

Najbardziej oczywistym stochastycznym algorytmem poszukiwania minimum globalnego wydaje się być algorytm zwany Pure Random Search, PRS. Dla ustalenia uwagi załóżmy, że $f : A \to \mathbb{R}$ jest funkcją ciągłą, zbiór $A = [0,1]^n \subset \mathbb{R}^n$.

Algorytm.

1. Losujemy punkt ze zbioru A zgodnie z rozkładem jednostajnym na A. Oznaczamy go jako x_0. Kolejne punkty $x_1, x_2, x_3 \ldots$, tworzymy w następujący sposób. Dla $t = 0, 1, 2, 3, \ldots$:
2. Losujemy punkt $\mathbf{y}_t \in A$ według rozkładu jednostajnego.
3. Jeżeli $f(\mathbf{y}_t) < f(x_t)$, kładziemy $x_{t+1} = \mathbf{y}_t$. W przeciwnym przypadku kładziemy $x_{t+1} = x_t$.
4. Idziemy do punktu 2 biorąc $t = t + 1$.

Możemy teraz stworzyć model matematyczny algorytmu PRS. Mianowicie, początkowy punkt x_0 jak i kolejno otrzymywane punkty x_t możemy traktować jako realizacje pewnych wektorów losowych $X_0, X_1, X_2, \ldots$. Tak samo, kolejne punkty $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \ldots$ są realizacjami ciągu wektorów losowych $\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2, \ldots$. Zakładamy, że wektor X_0 oraz wektory $\mathbf{Y}_t$ mają rozkłady jednostajne na zbiorze A oraz $X_0, \mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2, \ldots$ są niezależne. Natomiast dla $t > 0$ mamy:

$$X_t = \begin{cases} X_{t-1}, & \text{gdy } f(\mathbf{Y}_{t-1}) \geq f(X_{t-1}) \\ \mathbf{Y}_{t-1}, & \text{gdy } f(\mathbf{Y}_{t-1}) < f(X_{t-1}) \end{cases}$$

Możemy teraz wypowiedzieć twierdzenie o zbieżności.

Twierdzenie 1.2.1 *Przy powyższych założeniach:*

$$Prob\,(X_t \longrightarrow A^\star,\ gdy\ t \longrightarrow \infty) = 1,$$

to znaczy:

$$Prob\left(\{\omega \in \Omega\colon\ dist(X_t(\omega), A^\star) \longrightarrow 0,\ gdy\ t \longrightarrow \infty\}\right) = 1,$$

gdzie $dist(x, K)$ oznacza odległość punktu x od zbioru K.

Dowód. Przytoczymy dowód oparty na klasycznym Lemacie Borela-Cantellego.

Lemat 1.2.2 (Borel-Cantelli) *Niech* $(\Omega, \Sigma, Prob)$ *będzie przestrzenią probabilistyczną. Niech* $C_1, C_2, C_3, \cdots \in \Sigma$ *będzie ciągiem zdarzeń niezależnych. Wtedy:*

$$\sum_{i=1}^{\infty} Prob(C_i) = \infty \Longrightarrow Prob\left(\bigcap_{t=1}^{\infty}\bigcup_{i=t}^{\infty} C_i\right) = 1.$$

Przechodząc do naszego dowodu kładziemy $m = \min_A f$. Wtedy $A^\star = \{x \in A : f(x) = m\}$. Zauważmy kolejno, że:

1. Ciąg $f(X_t)$ jest nierosnący i ograniczony z dołu przez m, więc jest zbieżny do pewnej zmiennej losowej $\eta \geq m$.

Ustalmy $\varepsilon > m$.

2. Ponieważ funkcja f jest ciągła, zbiór $B = \{x \in A : f(x) < \varepsilon\}$ jest niepusty i otwarty, więc jego miara Lebesgue'a, $\nu(B) = \alpha > 0$.

3. Określmy: $C_t = \{\omega \in \Omega : f(\mathbf{Y}_t(\omega)) < \varepsilon\} = \mathbf{Y}_t^{-1}(B)$, dla $t = 1, 2, 3, \ldots$. Oczywiście C_t są zdarzeniami niezależnymi.

4. Ponieważ zmienne losowe $\mathbf{Y}_t$ mają rozkład ν, więc $\text{Prob}(C_t) = \text{Prob}_{\mathbf{Y}_t}(B) = \nu(B) = \alpha$ dla $t \geq 1$. Jest więc spełnione założenie Lematu Borela-Cantellego.

5. Na podstawie algorytmu PRS zachodzi implikacja:

$$\eta \geq \varepsilon \Longrightarrow \forall t \geq 1\ f(\mathbf{Y}_t) \geq \eta \geq \varepsilon,$$

czyli

$$\{\eta \geq \varepsilon\} \subset \bigcap_{t=1}^{\infty} (\Omega \setminus C_t) = \Omega \setminus \bigcup_{t=1}^{\infty} C_t.$$

$$\bigcap_{t=1}^{\infty}\bigcup_{i=t}^{\infty} C_i \subset \bigcup_{t=1}^{\infty} C_t \subset \{\eta < \varepsilon\}$$

6. Z Lematu Borela-Cantellego: $\text{Prob}(\{\eta < \varepsilon\}) = 1$. Ponieważ, $\varepsilon > m$ było ustalone dowolnie więc $\eta = m$. Tak więc: $\text{Prob}(f(X_t) \to m, \text{ gdy } t \longrightarrow \infty) = 1$.

7. Z ciągłości f oraz zwartości A, $\text{Prob}\,(X_t \longrightarrow A^\star, \text{ gdy } t \longrightarrow \infty) = 1$.

■

1.3. Algorytmy autonomiczne w czasie

Algorytm PRS jest najprostszym przykładem algorytmu autonomicznego, to znaczy jego mechanizm i parametry nie zmieniają się w czasie jego trwania. Zauważmy mianowicie, że można go zapisać w postaci:

$$X_t = T(X_{t-1}, \mathbf{Y}_{t-1}) \text{ dla } t = 1, 2, 3, \ldots \tag{1.2}$$

gdzie T jest operatorem $A \times B \longrightarrow A$ określonym równaniem:

$$T(x, \mathbf{y}) = \begin{cases} x, & \text{gdy } f(\mathbf{y}) \geq f(x) \\ \mathbf{y}, & \text{gdy.} f(\mathbf{y}) < f(x), \end{cases}$$

przy czym $A = B = [0,1]^n$, X_0 jest wektorem losowym o danym rozkładzie na A, $\mathbf{Y}_t$ są niezależnymi wektorami losowymi mającymi taki sam rozkład na B i są one niezależne od X_0. Innymi przykładami takich algorytmów są algorytm hybrydowy i jego uogólnienie, algorytm typu multistart.

Algorytm hybrydowy

Algorytmy hybrydowe, wykorzystujące metody stochastyczne jak i deterministyczne. Zakładamy, że dane są: zbiór $A \subset \mathbb{R}^n$ oraz funkcja $f : A \to \mathbb{R}^n$

Ustalmy miary probabilistyczną μ_0, ν_0 na zbiorze A. Zakładamy też, że mamy do dyspozycji jakąś metodę optymalizacji lokalnej, czyli funkcję $\varphi : A \to A$.

Algorytm.

1. Losujemy punkt ze zbioru A zgodnie z rozkładem μ_0. Oznaczamy go jako x_0. Kolejne punkty $x_1, x_2, x_3 \ldots$, tworzymy w następujący sposób. Dla $t = 0, 1, 2, 3, \ldots$:
2. Losujemy punkt $\mathbf{y}_t \in A$ według rozkładu ν_0.
3. Stosujemy φ do $\mathbf{y}_t$
4. Jeżeli $f(\varphi(\mathbf{y}_t)) < f(x_t)$, kładziemy $x_{t+1} = \varphi(\mathbf{y}_t)$. W przeciwnym przypadku kładziemy $x_{t+1} = x_t$.
5. Idziemy do punktu 2 biorąc $t = t + 1$.

Tutaj operator T ma postać:

$$T(x, \mathbf{y}) = \begin{cases} x, & \text{gdy } f(x) \geq f(\varphi(\mathbf{y})) \\ \varphi(\mathbf{y}), & \text{gdy } f(\varphi(\mathbf{y})) < f(x) \end{cases}$$

Zauważmy, że algorytm hybrydowy z identycznościową funkcją φ: $\varphi(\mathbf{y}) = \mathbf{y}$ oraz miarami Lebesgue'a $\mu_0 = \nu_0$ jest algorytmem PRS.

Multistart

Na ogół efektywniejszymi algorytmami są algorytmy typu Multistart. Podamy najprostszą wersję takiego algorytmu uogólniającą algorytm hybrydowy.

Ustalamy: μ_0 oraz ν_0 miary probabilistyczne na A, k, m liczby naturalne, metodę lokalną $\varphi : A \to A$.

Algorytm.

1. Losujemy niezależnie m punktów z A zgodnie z rozkładem μ_0:
   μ_0: $\mathbf{x} = (x_1, \ldots, x_m) \in A^m$ – populacja początkowa.
2. Losujemy niezależnie k punktów z A zgodnie z rozkładem ν_0:
   $\mathbf{y} = (y_1, \ldots, y_k) \in A^k$.
3. Stosujemy φ do każdego x_i oraz y_i i dostajemy:
   $$(\varphi(x_1), \ldots, \varphi(x_m), \varphi(y_1), \ldots, \varphi(y_k))$$
4. Sortujemy rosnąco używając funkcję f jako kryterium:
   $$(\bar{x}_1, \ldots, \bar{x}_{m+k}) \text{ gdzie } f(\bar{x}_1) \leq \cdots \leq f(\bar{x}_{m+k}).$$
5. Tworzymy nową populację biorąc m pierwszych punktów $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_m)$ i wracamy do punktu 2. biorąc $\mathbf{x} = \bar{\mathbf{x}}$ oraz $t = t + 1$.

Algorytm typu multistart należy do klasy określanej jako algorytmy populacyjne. W tym przypadku określimy $\hat{A} = A^m$, $\hat{B} = B^k$, natomiast operator $T : \hat{A} \times \hat{B} \to \hat{A}$ jest określony jako:

$$T(\mathbf{x}, \mathbf{y}) = \bar{\mathbf{x}},$$

gdzie $\bar{\mathbf{x}}$ było określone w ostatnim kroku algorytmu.

Zauważmy, że aby zachować ogólny schemat musi teraz ulec zmianie znaczenie samej minimalizowanej funkcji, tak aby była ona określona na $\hat{A}$. Można to zrobić kładąc dla $\mathbf{x} = (x_1, \ldots, x_m) \in \hat{A}$

$$\hat{f}(\mathbf{x}) = f(x_1).$$

Znajdując teraz argmin funkcji $\hat{f}$ otrzymujemy od razu argmin funkcji f.

Zbieżność algorytmu autonomicznego

Wypowiemy teraz twierdzenie gwarantujące zbieżność algorytmu autonomicznego (1.2), pochodzące z pracy [11], a następnie wniosek dotyczący algorytmu typu Multistart.

Twierdzenie 1.3.1 *Zakładamy, że A jest zbiorem zwartym oraz, że funkcja f jest ciągła. Ponadto zakładamy, że:*

(A) *Dla każdego $x_0 \in A$ oraz dowolnego ciągu $x_n \to x_0$:*

$$T(x_n, \mathbf{y}) \to T(x_0, \mathbf{y}) \quad p.w.\ \nu, \quad gdy\ n \to \infty.$$

(B) *Dla każdego* $x \in A^\star$ *oraz* $\mathbf{y} \in B$, $T(x, \mathbf{y}) \in A^\star$.

(C) *Dla każdego* $x \in A \setminus A^\star$:

$$\int_B f(T(x, \mathbf{y}))\, \nu(d\mathbf{y}) < f(x).$$

Wtedy, dla każdego $\varepsilon > 0$:

$$\lim_{t\to\infty} Prob\,(dist(X_t, A^\star) < \varepsilon) = 1.$$

Załóżmy dodatkowo warunek:

(D) *Dla każdego* $x \in A$ *oraz* $\mathbf{y} \in B$: $f(T(x, \mathbf{y})) \leq f(x)$.

Wtedy:

$$Prob\,(X_t \longrightarrow A^\star,\ gdy\ t \longrightarrow \infty) = 1.$$

Stosując powyższe twierdzenie do algorytmu Multistart, a więc także do algorytmu hybrydowego otrzymujemy:

Wniosek 1.3.2 *Zakładamy, że A jest zbiorem zwartym oraz, że funkcja f jest ciągła. Ponadto zakładamy, że:*

(a1) $\nu_0(l_c) = 0$ *gdzie,* , $l_c := \{x \in A : f(x) = c\}$.

(a2) *Metoda lokalna* φ *jest ciągła i niesingularna względem* ν_0.

(a3) *Jeżeli* $G \subset A$ *jest otoczeniem* $A^\star$, *to* $\nu_0(G) > 0$.

(a4) $f(\varphi(x)) \leq f(x)$, *dla* $x \in A$.

Wtedy:

$$Prob\,(X_{1,t} \longrightarrow A^\star,\ gdy\ t \longrightarrow \infty) = 1.$$

Dowód Wniosku 1.3.2 polega na stosunkowo łatwym sprawdzeniu, że jego założenia implikują założenia Twierdzenia 1.3.1. Warto zaznaczyć, że zbieżność algorytmu Multistart zachodzi w rzeczywistości przy założeniach łagodniejszych. Twierdzenie 1.3.1, oraz jego uogólnienia, dają ciekawsze rezultaty w przypadku metod bardziej zaawansowanych, które omówimy w dalszej części pracy.

Zanim przestawimy szkic dowodu Twierdzenia 1.3.1 musimy wprowadzić pewne oznaczenia i definicje.

Przez $\mathcal{B}(A)$ oznaczamy rodzinę wszystkich zbiorów borelowskich zawartych w A. Będziemy rozważać przestrzeń probabilistycznych miar borelowskich, $\mathcal{M} = \mathcal{M}(A)$, na zbiorze A. Wiadomo, że $\mathcal{M}$ jest przestrzenią metryczną zwartą, przy czym:

Ciąg miar $\mu_n \in \mathcal{M} \longrightarrow \mu \in \mathcal{M} \Longleftrightarrow$ dla każdej funkcji ciągłej h,

$$\int_A h\, d\mu_n \to \int_A h\, d\mu,$$

gdy $n \to \infty$.

Innym równoważnym warunkiem zbieżności jest:

$$\mu_n(C) \to \mu(C),$$

gdy $n \to \infty$, dla każdego $C \in \mathcal{B}(A)$ takiego, że $\mu(\delta C) = 0$, gdzie δC oznacza brzeg zbioru C.

Określamy tak zwany operator Foiasa:

$P : \mathcal{M} \to \mathcal{M}$: $P\mu(C) = (\mu \times \nu)(T^{-1}(C))$, dla $\mu \in \mathcal{M}$, $C \in \mathcal{B}(A)$. Łatwo stwierdzić, że:

$$P\mu(C) = \int_A \left(\int_B \mathrm{I}_C(T(x, \mathbf{y}))\nu(d\mathbf{y}) \right) \mu(dx), \text{ dla } \mu \in \mathcal{M},\ C \in \mathcal{B}(A),$$

gdzie I_C jest funkcją charakterystyczną C.

Oznaczamy prrzez μ_t^T rozkład X_t, to znaczy:

$$\mu_t^T(C) = \mathrm{Prob}(X_t \in C) = \mathrm{Prob}(X_t^{-1}(C)), \quad C \in \mathcal{B}(A), t = 1, 2, 3, \ldots$$

Zauważmy, że dla takich C:
$\mu_t^T(C) = \mathrm{Prob}(X_t \in C) = \mathrm{Prob}(T(X_{t-1}, \mathbf{Y}_t) \in C) = \mathrm{Prob}((X_{t-1}, \mathbf{Y}_t) \in T^{-1}(C)) = (\mu_{t-1}^T \times \nu)(C) = P\mu_{t-1}^T(C)$. Czyli:

$$\mu_1^T = P\mu_0, \quad \mu_2^T = P\mu_1^T, \quad \mu_3^T = P\mu_2^T, \quad \mu_4^T = P\mu_3^T, \ldots$$

i ogólnie:

$$\mu_t^T = P\mu_{t-1}^T, \quad t = 1, 2, 3, \ldots. \tag{1.3}$$

Oznaczmy dodatkowo: $\mu_0^T = \mu_0$.

Dowód Twierdzenia 1.3.1 będzie oparty na jednej z wielu wersji słynnego twierdzenia Lapunowa, które należy do teorii układów dynamicznych. Podajemy więc niezbędne pojęcia pojawiające się w tej teorii.

Definiujemy iteracje P^t jako: $P^0 = id_{\mathcal{M}}$, $P^{t+1} = P \circ P^t$, dla $t = 0, 1, 2, \ldots$. Operator Foiasa P dyktuje układ dynamiczny w następującym sensie: $\mathcal{M}$ jest przestrzeń fazowa, $\mu \in \mathcal{M}$ jest stanem) układu, a $t = 0, 1, 2, 3, ...$ interpretujemy jako czas (dyskretny).

W tym układzie orbitą punktu μ jest ciąg kolejnych iteracji tego punktu: $o(\mu) = \{P^t\mu : t = 0, 1, 2, 3, \dots\}$. Można pokazać, że założenie (A) implikuje, że $P : \mathcal{M} \to \mathcal{M}$ jest odwzorowaniem ciągłym, co pozwala na skorzystanie z szeregu dostępnych narzędzi stosowanych w teoii układów dynamicznych.

Zauważmy teraz ze z równości (1.3) wynika kluczowy dla nas związek wiążący ewolucję rozkładów wektorów losowych X_t z odpowiednimi orbitami układu dynamicznego dyktowanego przez operator Foiasa na przestrzeni miar:

$$P^t\mu_0 = \mu_t^T, t = 0, 1, 2, \dots, \tag{1.4}$$

czyli orbita stanu będącego rozkładem wektora losowego X_0 w układzie dynamicznym generowanym przez operator Foiasa, pokrywa się z ciągiem rozkładów wektorów X_t, μ_t^T, $t = 0, 1, 2, 3, \dots$.

Ponieważ przestrzeń miar probabilistycznych na przestrzeni zwartej jest przestrzenią zwartą, więc przypominamy poniżej podstawowe fakty z teorii układów dynamicznych na przestrzeniach zwartych, patrz na przykład [6]. Zbiór graniczny punktu μ jest to zbiór $\omega(\mu) = \{\lambda \in \mathcal{M} : \exists\, t_i \to \infty, P^{t_i}\mu \to \lambda\}$. Zbiór zwarty $\emptyset \neq \mathcal{K} \subset \mathcal{M}$ nazywamy zbiorem niezmienniczym, jeżeli $P(\mathcal{K}) \subset \mathcal{K}$. Niech ϱ będzie metryką na $\mathcal{M}$ zgodną z topologią, Wiadomo, że dla każdego $\mu \in \mathcal{M}$:

(1) $\omega(\mu) \neq \emptyset$,

(2) $\omega(\mu)$ jest niezmienniczy,

(3) Jeżeli $\mathcal{K} \subset \mathcal{M}$ jest zbiorem zbiorem niezmienniczym, to:
$\varrho(P^t\mu, \mathcal{K}) \to 0$ gdy $t \to \infty \iff \omega(\mu) \subset \mathcal{K}$.

Zapowiadane już twierdzenie Lapunowa wypowiemy, a później wykorzystamy w następującej postaci.

Twierdzenie 1.3.3 (Twierdzenie Lapunowa) *Niech (M, ϱ) będzie przestrzenią metryczną zwartą, $\emptyset \neq K \subset M$ będzie zbiorem niezmienniczym, $P : M \to M$ odwzorowaniem ciągłym. Niech $V : M \to \mathbb{R}$ będzie funkcją Lapunowa, to znaczy: (*) V jest ciągła, (*) $V(x) = 0$, dla $x \in K$, (*) $V(x) > 0$, dla $x \in M \setminus K$. (*) dla każdego $x \in M \setminus K$*

$$V(P(x)) < V(x).$$

Wtedy, dla każdego $x \in M$,

$$\varrho(P^t x, K) \to 0, \quad gdy\ t \to \infty.$$

Dowód powyższego twierdzenia można znaleźć w [10]. Natomiast dla nas najważniejszy jest następujący:

Wniosek 1.3.4 *Jeżeli zachodzą warunki (A), (B), (C) z Twierdzenia 1.3.1, to $\mathcal{M}^\star$ jest niezmienniczy oraz dla każdej miary $\mu \in \mathcal{M}$:*

$$\varrho(P^t\mu, \mathcal{M}^\star) \to 0,\ gdy\ t \to \infty.$$

Dowód Wniosku 1.3.4, szkic. Rozważamy układ dynamiczny dyktowany przez operator $P : \mathcal{M} \longrightarrow \mathcal{M}$. Naszym zbiorem niezmienniczym jest zbiór wszystkich miar skupionych na zbiorze $A^\star$, czyli

$$\mathcal{M}^\star := \{\mu \in \mathcal{M} : \operatorname{supp}\mu \subset A^\star\}.$$

Należy teraz sprawdzić, że funkcja $V : \mathcal{M} \to \mathbb{R}$ dana wzorem: $V(\mu) = \int_A (f - m)\, d\mu$ jest funkcją Lapunowa, co nie jest zadaniem zbyt trudnym. ■

Dowód Twierdzenia 1.3.1. Zauważmy kolejno, że:

1. Jeżeli miara $\mu^\star \in \mathcal{M}^\star$ oraz zbiór $C \in \mathcal{B}(A)$ są takie, że $A^\star \subset \operatorname{int} C$, to z określenia zbioru $\mathcal{M}^\star$:

$$\mu^\star(\delta C) = 0 \quad \text{oraz} \quad \mu^\star(C) = 1.$$

Wtedy też warunek zbieżności miar implikuje, że dla każdego ciągu miar $\mu_n \in \mathcal{M}$ takiego, że $\mu_n \to \mu^\star$ zachodzi: $\mu_n(C) \to 1$, gdy $n \to \infty$.

2. Ciąg X_t jest zbieżny stochastycznie do zbioru $A^\star$. Żeby to stwiedzić ustalamy $\varepsilon > 0$ i bierzemy $C = \{a \in A : \operatorname{dist}(a, A^\star) < \varepsilon\}$. Ustalamy $\mu_0 \in \mathcal{M}$. Własności zbiorów granicznych wspomnine powyżej gwarantują, że zbiór ω-graniczny, $\omega(\mu_0)$, jest zawarty w $\mathcal{M}^\star$. Dla każdego ciągu $t_n \to \infty$, można więc wybrać podciąg $t_{n_i} \to \infty$ oraz miarę $\mu^\star \in \mathcal{M}^\star$ takie, że $P^{t_{n_i}}\mu_0 \to \mu^\star$. Z udwodnionego przed chwilą punkt 1, $P^{t_{n_i}}\mu_0(C) \to \mu^\star(C) = 1$. Ale to oznacza, że $P^t\mu_0(C) \to 1$, as $t \to \infty$. Pamiętając, że $P^t\mu_0$ jest rozkładem wektora X_t, mamy: $\operatorname{Prob}(\operatorname{dist}(X_t, A^\star) < \varepsilon) = \operatorname{Prob}(X_t \in C) = P^t\mu_0(C)$. Stąd:

$$\lim_{t\to\infty} \operatorname{Prob}(\operatorname{dist}(X_t, A^\star) < \varepsilon) = 1.$$

3. Dla każdego $\varepsilon > 0$:

$$\lim_{t\to\infty} \operatorname{Prob}(f(X_t) < m + \varepsilon) = 1.$$

Rzeczywiście, ustalamy $\varepsilon > 0$. Zwartość $A^\star$ oraz ciągłość f pozwalają dobrać $\delta > 0$ takie, że $\operatorname{dist}(x, A^\star) < \delta \Longrightarrow f(x) < m + \varepsilon$. Możemy więc użyć udowodniony przed chwilą warunek biorąc w nim $\delta = \varepsilon$.

4. Przypominamy znany fakt. Niech $\xi_n : \Omega \to \mathbb{R}$ będzie nierosnącym ciągiem zmiennych losowych zbieżnym stochastycznie do m. Wtedy, ξ_n zmierza do m z prawdopodobieństwem 1.

5. Ciąg $\xi_t = f(X_t)$ jest stochastycznie zbieżny do m, a z założenia **(D)** jest nierosnący. Zmierza więc do m z prawdopodobieństwem 1. Zwartość A oraz $A^\star$, a także ciągłość f pociągają:

$$\{\omega : f(X_t(\omega)) \to m, \text{ gdy } \to \infty\} \subset \{\omega : X_t(\omega) \to A^\star, \text{ gdy } \to \infty\},$$

co oznacza, że X_t zmierzają z prawdopodobieństwem 1 do m.

■

Algorytm ewolucyjny $(\mu/\varrho + \lambda)$

Przedstawimy teraz dużo bardziej zaawansowaną metodę optymalizacyjną, stanowiącą przykład tzw. strategii ewolucyjnej $(\mu/\varrho+\lambda)$. Strategie ewolucyjne, zainspirowane procesem ewolucji biologicznej, wykorzystują mechanizmy rekombinacji, mutacji oraz selekcji, przetwarzając kolejne populacje osobników odpowiadające wektorom punktów z dziedziny optymalizowanej funkcji. W strategii ewolucyjnej $(\mu/\varrho + \lambda)$ mamy μ osobników w populacji, które w każdym kroku działania algorytmu (w każdej generacji) produkują λ potomków. ϱ jest liczbą losowo wybranych osobników rodzicielskich, które, poprzez proces rekombinacji oraz mutacji, są zaangażowane w produkcję jednego potomka. Algorytm ten wykorzystuje selekcję elitarną, tj. spośród łącznej liczby $\mu + \lambda$ osobników w nowo powstałej populacji przeżywa μ osobników najlepiej dostosowanych, stanowiących kolejną populację. Każdy osobnik ma przypisany parametr mutacji σ, który ewoluuje wraz z nim. Jest to mechanizm tzw. samoadaptacji, charakterystyczny dla omawianej klasy metod. Więcej informacji o strategiach ewolucyjnych, zainteresesowany czytelnik znajdzie w [3]. Skupimy się teraz na następującej wersji algorytmu.

Niech $0 < \sigma_1 < \sigma_2 \leq \infty$, $\tau > 0$ oraz niech μ, λ, ϱ będą liczbami naturalnymi. Niech $A = \mathbb{R}^n$ oraz $f\colon A \to \mathbb{R}$ niech będzie optymalizowaną funkcją. Definiujemy: $\bar{f}\colon \mathbb{R}^n \times [\sigma_1, \sigma_2]^n \to \mathbb{R}$ jako $\bar{f}(p, \sigma) = f(p)$.

Algorytm

1. Inicjalizacja: $x = \left(x_1, \ldots, x_\mu\right) \in (\mathbb{R}^n \times [\sigma_1, \sigma_2]) \times \ldots \times (\mathbb{R}^n \times [\sigma_1, \sigma_2])$, gdzie $x_l = (p_l, \sigma_l) \in (\mathbb{R}^n \times [\sigma_1, \sigma_2])$, $l = 1 \ldots, \varrho$

2. **For** $l = 1$ **To** λ powtarzamy następujące procedury:

   a) **Kojarzenie:** Wybieramy losowo ϱ różnych osobników $x_{l_1}, \ldots, x_{l_\varrho}$

   b) **Rekombinacja:** $x_{\nu+l} := \frac{1}{\varrho} \sum_{i=1}^{\varrho} x_{l_i}$

   c) **Mutacja:**

      i. $\sigma_{\mu+l} := \sigma_{\mu+l} \cdot e^{(\tau \cdot r)}$, gdzie r pochodzi z rozkładu normalnego $N(0, 1)$

      ii. If $\sigma_{\mu+l} < \sigma_1$, wtedy $\sigma_{\mu+l} := \sigma_1$

      iii. If $\sigma_{\mu+l} > \sigma_2$, wtedy $\sigma_{\mu+l} := \sigma_2$

      iv. $p_{\mu+l} := p_{\mu+l} + \sigma_{\mu+l} \cdot z$, gdzie z pochodzi z rozkładu normalnego $N_n(0, I_n)$, I_n jest n wymiarową macierzą identycznościową

   d) $x_{\mu+l} := (p_{\mu+l}, \sigma_{\mu+l})$

3. **Selekcja:**

   a) Sortujemy wektor $x_1 \ldots, x_{\mu+\lambda}$ by otrzymać

   $$\bar{x}_1, \ldots, \bar{x}_{\mu+\lambda} \text{ spełniający } \bar{f}(\bar{x}_1) \leq \ldots \leq \bar{f}(\bar{x}_{\mu+\lambda})$$

   b) Formujemy kolejną poppulację z pierwszych μ punktów,

   $$x := (\bar{x}_1, \ldots, \bar{x}_\mu)$$

4. Wracamy do kroku 2.

Model matematyczny powyższego algorytmu, spełniający nasze założenia, wymaga konstrukcji bardziej zaawansowanej od metod przedstawionych uprzednio.

Model matematyczny

Niech $\bar{A} = \mathbb{R}^n \times [\sigma_1, \sigma_2]$. Definiujemy

$$C = \{l = (l_1, \ldots, l_\varrho) \in \{1, 2 \ldots, \mu\}^\varrho : \ l_i \neq l_j \text{ dla } i \neq j\}$$

i

$$B = C^\lambda \times \left(\mathbb{R}^{n+1}\right)^\lambda.$$

Niech

$$T_1 : \bar{A}^\mu \times C^\lambda \ni (x, (l^1, \ldots, l^\lambda)) \longrightarrow \left(\frac{1}{\varrho} \sum_{i=1}^{\varrho} x_{l_i^1}, \ldots, \frac{1}{\varrho} \sum_{i=1}^{\varrho} x_{l_i^\lambda} \right) \in \bar{A}^\lambda$$

będzie operatorem rekombinacji. Definiujemy

$$mut_\sigma : [\sigma_1, \sigma_2] \times \mathbb{R} \ni (\sigma, r) \longrightarrow \min\{\max\{\sigma \cdot e^{\tau \cdot r}, \sigma_1\}, \sigma_2\} \in [\sigma_1, \sigma_2],$$

$$mut : \bar{A} \times \mathbb{R} \times \mathbb{R}^n \ni ((p, \sigma), r, z) \longrightarrow (p + mut_\sigma(\sigma, r) \cdot z, mut_\sigma(\sigma, r)) \in \bar{A}.$$

Niech

$$T_2 : \bar{A}^\lambda \times (\mathbb{R} \times \mathbb{R}^n)^\lambda \ni [(x_1, \ldots, x_\lambda), (r_1, z_1), \ldots, (r_\lambda, z_\lambda)] \longrightarrow (mut(x_i, r_i, z_i))_{i=1}^\lambda \in \bar{A}^\lambda$$

będzie operatorem mutacji oraz niech

$$T_3 : \bar{A}^{\mu+\lambda}(x_1, \ldots, x_{\mu+\lambda}) \longrightarrow (\bar{x}_1, \ldots, \bar{x}_\mu) \in \bar{A}^\mu$$

będzie operatorem selekcji który wybiera μ "'najlepszych" osobników z populacji zgodnie z relacją z Kroku 3. Definiujemy

$$T : \bar{A}^\mu \times C^\lambda \times (\mathbb{R} \times \mathbb{R}^n)^\lambda \ni (x, l, w) \longrightarrow T_3(x, T_2(T_1(x, l), w)) \in \bar{A}^\mu. \tag{1.5}$$

Niech $\mathbf{Y}_t \colon \Omega \to C^\lambda \times (\mathbb{R} \times \mathbb{R}^n)^\lambda$ będzie ciągiem niezależnych zmiennych losowych o rozkładzie $\Pi \times (N(0,1) \times N_n(0, I_n))^\lambda$, gdzie Π to rozkład jednostajny na zbiorze C^λ, zdefiniowany przez $\Pi(l) = \frac{1}{\#C^\lambda}$. Niech $X_0 \colon \Omega \to \bar{A}^\mu$ będzie mierzalne oraz niezależne od ciągu Y_t. Definiujemy

$$X_{t+1} = T(X_t, \mathbf{Y}_t).$$

Tak więc powyższy algorytm jest algorytmem autonomicznym (1.2). Można by się więc spodziewać, że dowód jego zbieżności będzie wnioskiem z Twierdzenia 1.3.1. Niestety założenia tamtego twierdzenia są zbyt restrykcyjne. Zbieżność powyższego algorytmu, Wniosek 1.4.3, wyniknie natomiast z Twierdzenia 1.4.2 przedstawionego w kolejnym punkcie, które jest istotnym wzmocnieniem Twierdzenia 1.3.1.

1.4. Algorytmy nieautonomiczne

Większość używanych obecnie algorytmów optymalizacji stochastycznej nie jest autonomiczna. Mianowicie, w trakcie działania algorytmu mogą się zmieniać jakieś jego parametry, także kolejne losowania mogą być dokonywane ze zmieniających się rozkładów. Warto więc rozważać ogólniejsze niż (1.2) algorytmy postaci:

$$X_t = T_t(X_{t-1}, \mathbf{Y}_{t-1}) \text{ dla } t = 1, 2, 3, \ldots \tag{1.6}$$

Tutaj T_t są mierzalnymi odwzorowaniami $A \times B \to A$.

Okazuje się, że wiele obecnie używanych algorytmów stochastycznej optymalizacji globalnej może być opisanych za pomocą schematu (1.6). Do badania ich zbieżności można użyć podobnej metodologii do tej zademonstrowanej przez na w poprzednim punkcie do zbadania zbieżności algorytmu multistart. W szczególności można skorzystać z twierdzeń, które uogólniają na przypadek nieautonomiczny i wzmacniają Twierdzenie 1.3.1. Uogólnienia te jednak nie są banalne. Poniżej prezentujemy kilka przykładów takich twierdzeń. Ich dowody też są oparte na teorii układów dynamicznych określonych na przestrzeni miar, jednak brak autonomiczności powoduje, że technicznie są bardziej złożone.

Podobnie jak poprzednio oznaczmy przez $\mathcal{B}(A)$, $\mathcal{B}(B)$ σ-algebry zbiorów borelowskich na A oraz B, natomiast $\mathcal{M}$ oraz $\mathcal{N}$ będą oznaczać zbiór wszystkich miar probabilistycznych, odpowiednio, na $\mathcal{B}(A)$ oraz $\mathcal{B}(B)$. Tak jak poprzednio w przestrzeniach tych rozpatrujemy topologię słabej zbieżności. Jako $\mathcal{T}$ oznaczamy zbiór operatorów $A \times B \to A$. Na $\mathcal{T}$ będziemy rozważać topologię zbieżności jednostajnej a na kartezjańskim produkcie $\mathcal{T} \times \mathcal{N}$ będziemy rozważać topologię produktową.

Zakładamy, że $X_0 : \Omega \longrightarrow \mathbb{R}^n$ ma rozkład $\mu_0 \in \mathcal{M}$. Natomiast $\mathbf{Y}_t :: \Omega \longrightarrow \mathbb{R}^m$ mogą mieć różne rozkłady $\nu_t \in \mathcal{N}$, dla $t = 0, 1, 2, \ldots$. Dalej zakładamy, że $X_0, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$ są niezależne.

Łatwo zauważyć, że rozkłady prawdopodobieństwa X_t są zdeterminowane przez rozkład μ_0 zmiennej X_0 oraz przez ciąg $\{(T_t, \nu_t)\}_{t=0}^{\infty}$.

Następujące twierdzenie istotnie wzmacnia Twierdzenie 1.3.1. Jego dowód, uogólniający omówioną wcześniej metodologię, można znaleźć w pracy [17].

Twierdzenie 1.4.1 *Załóżmy że A jest przestrzenią zwartą. Niech $U_0 \subset \mathcal{T} \times \mathcal{N}$ będzie zbiorem zwartym takim, że:*

(A) *dla każdej pary $(T, \nu) \in U_0$ oraz dla każdego $x \in A$, $f \circ T$ jest półciągła z góry w punkcie (x, y) dla $\nu-$ prawie każdego $y \in B$,*

(B) *dla wszystkich $x \in A$ oraz $t \in \mathbb{N}$,*

$$\int_B f(T_t(x, y))\nu_t(dy) \leq f(x), \tag{1.7}$$

(C) *dla każdej pary $(T, \nu) \in U_0$ oraz dla każedgo $x \in A \setminus A^*$*

$$\int_B f(T(x, y))\nu(dy) < f(x) \tag{1.8}$$

Jeśli (T_t, ν_t) zawiera podciąg $(T_{t_n}, \nu_{t_n}) \in U_0$, wtedy

$$Prob(X_t \longrightarrow A^\star) = 1, \ gdy\ t \longrightarrow \infty.$$

Poniższe twierdzenie, [17], jest pożytecznym i prostym wnioskiem z Twierdzenia 1.4.1.

Twierdzenie 1.4.2 *Załóżmy że zbiory „podpoziomicowe" $A_\delta = \{x \in A : \ f(x) \leq \delta\}$ są zwarte dla każdego $\delta > 0$. Niech $U_0 \subset \mathcal{T} \times \mathcal{N}$ będzie zbiorem zwartym takim, że że warunki **(A)** oraz **(C)** Twierdzenia 1.4.1 są spełnione. Załóżmy dodatkowo:*

(B') *dla każdego $t \in \mathbb{N}$ oraz $x \in A, \ y \in B$*

$$f(T_t(x, y)) \leq f(x).$$

Jeśli ciąg $u_t = (T_t, \nu_t)$zawiera $\{u_{t_n} : \ n = 0, 1, \ldots\} \subset U_0$, wtedy

$$Prob(X_t \longrightarrow A^\star) = 1, \ gdy\ t \longrightarrow \infty.$$

Szkic dowodu: Dla dowolnego $\delta > 0$ oraz $T\colon A \times B \to A$ niech

$$T_\delta = T|_{A_\delta \times B}\colon\ A_\delta \times B \longrightarrow A.$$

Dla dowolnego $U \subset \mathcal{T} \times \mathcal{N}$ oraz $\delta > 0$ niech

$$(U)_\delta = \{(\mathcal{T}_\delta, \nu)\colon (T, \nu) \in U\}.$$

Ustalmy $x_0 \in A$. Zbiór $(U_0)_{f(x_0)}$ jest zwarty. Jeżeli $\mu_0 = \delta_{x_0}$, wtedy $supp\ \mu_0 = \{x_0\} \subset A_{f(x_0)}$. Zbiór $A_{f(x_0)}$ jest zwarty, z założenia **(B')** mamy $T_t(A_{f(x_0)} \times B) \subset A_{f(x_0)}$ dla każdego $t \in \mathbb{N}$ oraz $A^\star \subset A_{f(x_0)}$. Zatem, przy założeniu $\mu_0 = \delta_{x_0}$, możemy zastosować Twierdzenie 1.4.1 z uwzględnieniem funkcji $f|_{A_{f(x_0)}}$. Stąd otrzymujemy, że algorytm jest zbieżny dla dowolnego punktu startowego $x_0 \in A$. Pozostaje skorzystać z twierdzenia Fubiniego. ■

Zbieżność algorytmu ewolucyjnego $(\mu/\varrho+\lambda)$ jest konsekwencją Twierdzenia 1.4.2. Mianowicie mamy następujący:

Wniosek 1.4.3 *Załóżmy że $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ jest funkcją ciągłą oraz że dla każdej liczby dodatniej $\delta > 0$ zbiory podpoziomicowe A_δ są ograniczone. Niech $\widehat{A} = [\mathbb{R}^n \times [\sigma_1, \sigma_2]]^\mu$ oraz $\widehat{A^\star} = (A^\star \times [\sigma_1, \sigma_2])^\mu$. Niech $X_t\colon \Omega \to \widehat{A}$ będzie ciągiem zmiennych losowych reprezentujących algorytm ewolucyjny $(\mu/\varrho + \lambda)$ przedstawiony w punkcie 1.3. Zachodzi:*

$$Prob(X_t \longrightarrow \widehat{A^\star}) = 1,\ gdy\ t \longrightarrow \infty.$$

Szkic dowodu: Definiujemy funkcję

$$\widehat{f}\colon \widehat{A} \ni \Big((p_1, \sigma_1), \ldots, (p_\mu, \sigma_\mu)\Big) \longrightarrow \sum_{i=1}^{\mu} f(p_i) \in \mathbb{R}.$$

Zbiór $\widehat{A^\star}$ jest zbiorem jej minimów globalnych. Definiujemy $U = U_0 = \{T\} \times \{\nu\}$ (algorytm jest autonomiczny), gdzie $\nu = \Pi \times (N(0, I_{n+1}))^\lambda$ jest rozkładem prawdopodobieństwa Y_t, $t = 0, 1, \ldots$. Pozostaje sprawdzić że spełnione są założenia Twierdzenia 1.4.2. ■

Alternatywna metodologia dowodzenia zbieżności szerokiej klasy algorytmów ewolucyjnych, bazująca na klasycznym rachunku prawdopodobieństwa, została wykorzystana w [14].

Przedstawiamy poniżej inne możliwe uogólnienie Twierdzenia 1.3.1. Jego dowód można znaleźć w [18]. Zaczniemy od przedstawienia poniższej definicji.

Dla dowolnej liczby $\delta > 0$, definiujemy zbiór $U_0(\delta)$:

$$\mathcal{T} \times \mathcal{N} \supset U_0(\delta) \ni (T, \nu) \overset{def}{\iff} \begin{cases} \int_B f(T(x,y))\nu(dy) < f(x) & \text{for } x \notin A(\delta) \\ \int_B f(T(x,y))\nu(dy) \le \delta & \text{for } x \in A(\delta). \end{cases}$$

Twierdzenie 1.4.4 *Niech A będzie zwartą przestrzenią metryczną. Niech $\{U_0^k\}_{k\in\mathbb{N}}$ będzie malejącą rodziną zbiorów zwartych spełniających $U_0^k \subset U_0(\frac{1}{k})$ oraz takich że następujące warunki są spełnione:*

(A1) *dla każdego $k \in \mathbb{N}$, $(T, \nu) \in U_0^k$ oraz $x \in A$, $f \circ T$ jest półciągła z góry w punkcie (x, y) dla ν-prawie każdego $y \in B$,*

(C2) *dla każdego $t \in \mathbb{N}$, $(T_t, \nu_t) \in U_0^{k_t}$, gdzie k_t jest ciągiem spełniającym $k_t \to \infty$.*

Wtedy

$$\forall \varepsilon > 0 \;\; Prob(dist(X_t, A^*) < \varepsilon) \overset{t\to\infty}{\longrightarrow} 1 \quad oraz \quad E(f(X_t)) \overset{t\to\infty}{\longrightarrow} \min_A f.$$

Dowód powyższego twierdzenia jest metodologicznie podobny do dowodu Twierdzenia 1.3.1, jakkolwiek dużo bardziej skomplikowany. W szczególności, musimy używać istotnie mocniejszej wersji twierdzenia Lapunowa. Powyższe twierdzenie prowadzi jednak do pokazania zbieżności szerokiej klasy algorytmów, takich jak algorytmy symulowanego wyżarzania.

Algorytm Simulated Annealing

Przedstawimy teraz przykład bardzo popularnej nieautonomicznej metody optymalizacyjnej, znanej jako algorytm Simulated Annealing, SA. Nazwa algorytmu nawiązuje do fizycznego procesu wyżarzania (annealing), stosowanego w celu usuwania defektów z pewnej klasy metali poprzez rozgrzanie oraz powolne ochładzanie materiału w celu obniżenia energii układu, patrz [8].

Niech $A \subset \mathbb{R}^n$ będzie zbiorem zwartym oraz niech B będzie ośrodkową przestrzenią metryczną. Niech $\bar{B} = B \times [0, 1]$. Niech $\xi_t \colon \Omega \to B$, $t \in \mathbb{N}$ będzie ciągiem niezależnych zmiennych losowych o wspólnym rozkładzie $\nu \in M(B)$ oraz niech $r_t \colon \Omega \to [0, 1]$, $t \in \mathbb{N}$ będą niezależnymi zmiennymi losowymi o rozkładach jednostajnych. Niech $M > 0$ i niech $[0, M] \ni \beta_t$ będzie ciągiem spełniającym $\lim_{t\to\infty} \beta_t = 0$ oraz niech $Q \colon A \times B \to A$ będzie odwzorowaniem mierzalnym. Określamy algorytm SA za pomocą równości:

$$X_t = T_t(X_{t-1}, \mathbf{Y}_{t-1}) \text{ dla } t = 1, 2, 3, \ldots \tag{1.9}$$

gdzie ciąg odwzorowań $T_t \colon A \times \bar{B} \to A$ definiujemy jako:

$$T_t(x, z, r) = \begin{cases} Q(x, z), \text{ jeżeli } f(Q(x, z)) \le f(x), \\ Q(x, z), \text{ jeżeli } f(Q(x, z)) > f(x) \wedge r \le \exp(-\frac{1}{\beta_t} \cdot |f(Q(x, z)) - f(x)|), \\ x, \text{ w przeciwnym przypadku,} \end{cases}$$

jeśli $\beta = 0$, w powyższej formule kładziemy $-\frac{1}{\beta} = -\infty$ oraz $\exp(-\infty) = 0$. Tutaj $\mathbf{Y}_t = (\xi_t, r_t)$, oraz X_0 jest zmienną losową niezależną od ciągu $\{\mathbf{Y}_t\}, t \in \mathbb{N}$. Istotą tego algorytmu jest stworzenie szansy na odrzucenie kolejnego przybliżenia nawet w przypadku, gdy jest ono lepsze, przy czym prawdopodobieństwo takiego działania maleje wraz ze wzrostem czasu, ale może być duże, gdy to przybliżenie jest tylko niewiele lepsze od poprzedniego.

Okazuje się, że przy stosunkowo skromnych założeniach algorytm ten jest stochastycznie zbieżny. Użycie Twierdzenia 1.4.4 prowadzi do dowodu zbieżności algorytmu SA bazującego na metodologii znacząco różnej od istniejących w literaturze metod dowodowych, [8], [20]. Wniosek 1.4.5 jest wnioskiem z Twierdzenia 1.4.4. Jego dowód polega na niebanalnym sprawdzeniu, że z założeń wniosku wynikają założenia Twierdzenia 1.4.4. Konstrukcja malejącej rodziny zbiorów U_0^k jest wyznaczona przez dowolny malejący podciąg β_{t_k} ciągu β_t. Występujący we wniosku warunek półciągłości implikuje założenie **(A1)**. Sprawdzenie warunku **(C2)** wymaga wykorzystania faktu, że że ciąg $\beta_t \in [0, \infty)$ jest zbieżny do zera oraz obu założeń wniosku.

Wniosek 1.4.5 *Załóżmy, że dla każdego $x \in A$, $\nu(D_{f\circ Q}(x)) = 0$, gdzie $D_{f\circ Q}(x)$ jest zbiorem $z \in B$ takich, że $f \circ Q$ nie jest ciągła w punkcje (x, z). Załóżmy dodatkowo że dla dowolnego $x \in A \setminus A^\star$,*

$$\nu(\{z \in B\colon f(Q(x,z)) < f(x)\}) > 0. \tag{1.10}$$

Wtedy,

$$\forall \varepsilon > 0 \; Prob(dist(X_t, A^*) < \varepsilon) \overset{t\to\infty}{\longrightarrow} 1 \quad oraz \quad E(f(X_t)) \overset{t\to\infty}{\longrightarrow} \min_A f.$$

Autorzy

Jerzy Ombach
Wydział Matematyki i Informatyki Uniwersytetu Jagiellońskiego,
Instytut Matematyki
ul. Prof. S. Łojasiewicza 6, 30-348 Kraków
jerzy.ombach@im.uj.edu.pl

Dawid Tarłowski
Wydział Matematyki i Informatyki Uniwersytetu Jagiellońskiego,
Instytut Matematyki
ul. Prof. S. Łojasiewicza 6, 30-348 Kraków
dawid.tarlowski@im.uj.edu.pl

Bibliografia

[1] A. Ahrari, A. A. Atai, *Grenade Explosion Method A novel tool for optimization of multimodal functions*, Appl. Soft Comput. 10 (2010), pp. 1132 - 1140.

[2] M. J. Appel, R. Labarre, D. Radulovic, *On Accelerated Randon Search*, SIAM J. Optim. Vol. 14(2003), 708-731,

[3] H.G. Beyer, H.P. Schwefel, *Evolution Strategies – A Comprehensive Introduction*, Nat. Comput. 1 (2002),pp. 3-52.

[4] M. Clerc and J. Kennedy, *The particle swarm - explosion, stability, and convergence in a multidimensional complex space*, Transactions on Evolutionary Computation 6(2002), 58-73.

[5] Y. Hani, L. Amodeo, F. Yalaoui, H. Chen, *Ant colony optimization for solving an industrial layout problem*, European Journal of Operational Research 183(2007) 633 Û 642.

[6] M. W. Hirsch, S. Smale, and R. L. Devaney *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, Elsevier Academic Press, 2013.

[7] D.Karaboga, B.Basturk , *A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm*, Journal of Global Optimization 39(2007). 459 Û 471.

[8] M. Locatelli, *Convergence of a Simulated Annealing Algorithm for Continuous Global Optimization*, J. Global Optim., 18 (2000), pp. 219-233

[9] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.

[10] J. Ombach, *Stability of evolutionary algorithms*, Journal Math Anal Appl. 342(2008), 326 - 333.

[11] J. Ombach, *A Proof of Convergence of General Stochastic Search for Global Minimum*, Journal of Difference Equations and Applications 13 (2007), pp. 795 Û 802.

[12] J. Ombach, D. Tarłowski, *Nonautonomous Stochastic Search in Global Optimization*, J NONLINEAR SCI vol. 22(2012) (2012), 169 - 185

[13] M. Radwański, *Convergence of nonautonomous evolutionary algorithm*, UIAM, 45(2007), 197 - 206.

[14] G. Rudolph: *Convergence Properties of Evolutionary Algorithms*, Kovac, Hamburg, 1997

[15] M. Semenov, D.A Terkel: *Analysis of convergence of an evolutionary algorithm with self-adaptation using a stochastic Lyapunov function*, Evolutionary Computation 2003, MIT Press

[16] D. Tarłowski, *Sufficient conditions for the convergence of non-autonomous stochastic search for a global minimum*, UIAM (2011), 73-83

[17] D. Tarłowski,*Nonautonomous stochastic search for global minimum in continuous optimization*, J MATH ANAL APPL vol. Volume 412 Issue 2 (2014), 631-645

[18] D. Tarłowski, *Nonautonomous Dynamical Systems in Stochastic Global Optimization*, rozprawa doktorska, IM UJ, 2014.

[19] D. Whitleya, S. Ranaa, J. Dzuberaa and K. E. Mathiasb, *Evaluating evolutionary algorithms*, Articicial Intelligence 85(1996) 245 - 276..

[20] R.L. Yang, *Convergence of the Simulated Annealing Algorithm for Continuous Global Optimization*, J. Optim. Theory Appl., 104 (2000),pp. 691–716