Dlaczego algorytmy stochastyczne są zbieżne?

Jerzy Ombach

Instytut Matematyki, Uniwersytet Jagielloński

Będlewo, Sierpień 2014

Zadania optymalizacji globalnej

Problem

Dana jest funkcja $f : A \longrightarrow \mathbb{R}$
Wyznaczyć efektywnie minimum globalne, to znaczy punt $x_0 \in A$ taki, że

$$\forall x \in A \; f(x_0) \leqslant f(x).$$

lub jego możliwie najlepsze przybliżenie.

Oznaczamy zbiór rozwiązań:

$$A^\star := \{x_0 \in A : \text{ dla każdego } x \in A \;\; f(x_0) \leqslant f(x)\}.$$

Szukanie maximum globalnego funkcji f

$$\Longleftrightarrow$$

Szukanie minimum globalnego funkcji $-f$

Przykłady zadań optymalizacji globalnej

1. Metoda największej wiarygodności

Znajdź parametr $\theta \in \Theta \subset \mathbb{R}^k$ dla którego funkcja wiarygodności jest największa.

Na przykład: $l(\theta) = g_\theta(x^1) \cdot \ \ldots \ \cdot g_\theta(x^n)$, gdzie $x^1, \ldots, x^n$ jest zaobserwowaną próbką prostą z rozkładu o gęstości g_θ.

2. Rozwiązywanie układu równań

znajdź rozwiązanie $x_0 \in A \subset \mathbb{R}^n$ układu równań

$$\begin{cases} g^1(x) & = & 0 \\ \ldots & \ldots & \ldots \\ g^n(x) & = & 0 \end{cases}$$

Definiujemy funkcją: $f(x) = g^1(x)^2 + \ \ldots \ + g^n(x)^2$ i szukamy minimum globalnego $x_0 \in A$.

Przykłady zadań optymalizacji globalnej

3. Szukanie wystarczająco dobrych wartości funkcji

Dana jest funkcja $g : A \longrightarrow \mathbb{R}$ oraz liczba $\alpha > \min g$.
Wyznaczyć efektywnie wartości $x_0 \in A$ takie, że

$$g(x_0) \leqslant \alpha.$$

Definiujemy funkcją: $f(x) = \max(g(x), \alpha)$ i szukamy minimum globalnego f, $x_0 \in A$.

Klasyczne twierdzenia

Twierdzenie

A – zwarty, f – ciągła.
$\Longrightarrow$ *Istnieje minimum globalne.*

Twierdzenie (prowadzi do rozwiązywania układu n-równań)

A – otwarty w $\mathbb{R}^n$, f – różniczkowalna,
$x_0 \in A$ – minimum.
$\Longrightarrow f'(x_0) = 0.$

Klasyczne twierdzenia

Twierdzenie (prowadzi do rozwiązywania $n + k$ równań)

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $g : \mathbb{R}^n \longrightarrow \mathbb{R}^k$, $k < n$,
x_0 – *minimum* f *przy ograniczeniach*
$A = \{x : g(x) = 0\}$,
f, g *są klasy* C^1, $\operatorname{rząd} g'(x_0) = k$.
$\Longrightarrow$ *Istnieją* $\lambda \in \mathbb{R}^k$ *takie, że:*

$$f'(x_0) = \lambda^T g'(x_0).$$

Algorytm PRS – PURE RANDOM SEARCH

$f : A \to \mathbb{R}$ – ciągła, $\quad A = [0,1]^n \subset \mathbb{R}^n$.

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 $x_{t+1} = x_t$.

Niech X_t będzie ciągiem zmiennych losowych, których realizacje x_t określa algorytm.

Twierdzenie

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

Dowód.

Lemat Borela-Cantellego

Niech $(\Omega, \Sigma, \mathrm{Prob})$ będzie przestrzenią probabilistyczną.
Niech $C_1, C_2, C_3, \cdots \in \Omega$ będzie ciągiem zdarzeń niezależnych:

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

$m = \min_A f$. $A^\star = \{x \in A : f(x) = m\}$.

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

2. Niech $\mathbf{Y}_t$ będą niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na A, których realizacjami są punkty $\mathbf{y}_i$.

Ustalmy $\varepsilon > m$.

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

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

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

6. Na podstawie algorytmu PRS zachodzi implikacja:

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

czyli

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

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

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

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

Algorytm ARS – ACCELERATED RANDOM SEARCH

$f : A \to \mathbb{R}$ – ciągła, $\quad A = [0,1]^n \subset \mathbb{R}^n$
Ustalamy $c > 1$ oraz $\rho > 0$ – dokładność.

Algorytm

0. Zadajemy $t = 1$ oraz $r_1 = 1$. Losujemy x_1 według rozkładu jednostajnego na A.
1. Mając $x_t \in A$ oraz $r_t \in (0,1]$, losujemy y_t według rozkładu jednostajnego na $B(x_t, r_t) \cap A$, gdzie $B(x,r)$ jest kulą.
2. Jeżeli $f(y_t) < f(x_t)$, podstawiamy $x_{t+1} = y_t$ oraz $r_{t+1} = 1$.
3. Jeżeli $f(y_t) \geqslant f(x_t)$, to:
   1. Jeżeli $r_t \geqslant \rho$, kładziemy $x_{t+1} = x_t$ oraz $r_{t+1} = r_t/c$.
   2. Jeżeli $r_{t+1} < \rho$, kładziemy $r_{t+1} = 1$.
4. $t := t + 1$ i wracamy do kroku 2.

Twierdzenie

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

Algorytm hybrydowy

$f : A \to \mathbb{R}$.
Ustalmy miarę μ_0, $\nu_0 \in \mathcal{M}$. Ustalmy $\varphi : A \to A$ – „metodę lokalną".
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 $x_{t+1} = x_t$.

Twierdzenie

Założenia:

(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 $\nu_0(G) > 0$.

Wtedy:

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

Algorytm typu multistart

$f : A \to \mathbb{R}$.
Ustalamy μ_0, ν_0 – miary probabilistyczne na A,
k, m – liczby naturalne. $\varphi : A \to A$ – metoda lokalna

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:
$$(\bar{x}_1, \ldots, \bar{x}_{m+k}) \text{ gdzie } f(\bar{x}_1) \leqslant \cdots \leqslant f(\bar{x}_{m+k}).$$
5. Tworzymy nową populację biorąc m pierwszych punktów $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_m)$ i wracamy do punktu 2. biorąc $\mathbf{x} = \bar{\mathbf{x}}$.

Twierdzenie

Założenia:

(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) $f(\varphi(x)) \leqslant f(x)$, dla $x \in A$.

Wtedy:

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

Ogólna forma problemu

Postawienie problemu

$A \subset \mathbb{R}^n$ – zwarty, $f : A \to \mathbb{R}$ funkcja ciągła osiągająca $\min f$ na A.
Zakładamy, że $\min f = 0$.

$$A^\star = \{x \in A : f(x) = 0\}.$$

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

gdzie T_t są odpowiednimi operatorami, $\mathbf{Y}_t$ są zmiennymi (wektorami) losowymi.

Autonomiczność na początek: $T_t = T$, $T : A \times B \to A$.

X_0 – wektor losowy o danym rozkładzie na A

$\mathbf{Y}_t$ – wektory losowe mające wspólny rozkład na $B \subset \mathbb{R}^m$.

Problem

Czy X_t zmierzają do $A^\star$, dla $t \longrightarrow \infty$.

PRS, ARS, algorytm hybrydowy, Multistart

PRS

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

ARS

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

$\mathbf{y}_t$ NIE są losowane ze wspólnego rozkładu!!!

Algorytm hybrydowy

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

Multistart

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

Formalizacja

$\mathcal{B}(A)$, $\mathcal{B}(B)$ σ-algebry zbiorów borelowskich na A oraz B

$\mathcal{M}$ zbiór wszystkich miar probabilistycznych na $\mathcal{B}(A)$.

$(\Omega, \Sigma, \mathrm{Prob})$ przestrzeń probabilistyczna.

$X_0 : \Omega \longrightarrow \mathbb{R}^n$ ma rozkład $\mu_0 \in \mathcal{M}$.

$\mathbf{Y}_t :: \Omega \longrightarrow \mathbb{R}^m$ mają wspólny rozkład ν, dla $t = 0, 1, 2, \ldots$.

$X_0, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$ są niezależne.

T jest mierzalne $A \times B \to A$.(Wtedy X_t jest wektorem losowym.)

Twierdzenie A

Zakładamy, że A jest zwarty oraz są spełnione warunki

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

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

(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 $\varepsilon > 0$:

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

Załóżmy dodatkowo warunek:

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

Wtedy:

$$\mathrm{Prob}\left(X_t \longrightarrow A^\star, \text{ gdy } t \longrightarrow \infty\right) = 1.$$

Przestrzeń miar

$\mathcal{M}$ – przestrzeń probabilistycznych miar borelowskich.

Wiadomo, że $\mathcal{M}$ jest przestrzenią metryczną zwartą, przy czym:

Ciąg miar $\mu_n \in \mathcal{M} \longrightarrow \mu \in \mathcal{M} \iff$ 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.

Operator Foiasa

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

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

gdzie I_C jest funkcją charakterystyczną C.

Oznaczamy: 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: $\quad \mu_t^T = P\mu_{t-1}^T, \quad t = 1, 2, 3, \ldots$. Oznaczmy dodatkowo: $\mu_0^T = \mu_0$.

Operator Foiasa

$\{\mu_t^T : t = 0, 1, 2, 3, \dots\}$ – orbita (trajektoria) μ_0.

Iteracje P^t definiujemy jako $P^0 = id_{\mathcal{M}}$, $P^{t+1} = P \circ P^t$, dla $t = 0, 1, 2, \dots$.

Operator Foiasa T dyktuje układ dynamiczny w następującym sensie:

$\mathcal{M}$ – przestrzeń fazowa,

$\mu \in \mathcal{M}$ – stan (punkt) układu,

$t = 0, 1, 2, 3, \dots$ – czas (dyskretny)

Orbita punktu μ: $o(\mu) = \{P^t\mu : t = 0, 1, 2, 3, \dots\}$.

Wtedy

$$P^t\mu_0 = \mu_t^T, t = 0, 1, 2, \dots$$

Operator Foiasa

Podstawowe własności P

Definiujemy dla funkcji ciągłej $h : A \longrightarrow \mathbb{R}$ funkcję:

$$Uh(x) = \int_B h(T(x, \mathbf{y}))\nu(d\mathbf{y})$$

1. Jeżeli $\mu \in \mathcal{M}$. oraz h jest ciągła, to:

$$\int_A h\, d(P\mu) = \int_A Uh\, d\mu.$$

2. Założenie **(A)** implikuje, że $P : \mathcal{M} \to \mathcal{M}$ jest odwzorowaniem ciągłym.

Układ dynamiczny

$\mathcal{M}$ – przestrzeń metryczna zwarta, $P : \mathcal{M} \to \mathcal{M}$ – odwzorowanie ciągłe.

Orbita punktu μ: $o(\mu) = \{P^t\mu : t = 0, 1, 2, 3, \dots\}$.

Zbiór graniczny punktu μ: $\omega(\mu) = \{\lambda \in \mathcal{M} : \exists\ t_i \to \infty, P^{t_i}\mu \to \lambda\}$.

Zbiór zwarty $\emptyset \neq \mathcal{K} \subset \mathcal{M}$ jest 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:
$$\varrho(P^t\mu, \mathcal{K}) \to 0 \text{ gdy } t \to \infty \iff \omega(\mu) \subset \mathcal{K}.$$

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:

1. V jest ciągła,
2. $V(x) = 0$, dla $x \in K$,
3. $V(x) > 0$, dla $x \in M \setminus K$.
4. dla każdego $x \in M \setminus K$

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

Wtedy, dla każdego $x \in M$,

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

Wniosek z Twierdzenie Lapunowa

$P : \mathcal{M} \longrightarrow \mathcal{M}$ – operator Foiasa.

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

Twierdzenie

Zachodzą warunki Twierdzenia A:

(A) Dla każdego $x_0 \in A$ oraz dowolnego ciągu $x_n \to x_0$:
$T(x_n, \mathbf{y}) \to T(x_0, \mathbf{y})$ p.w. ν, 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, $\mathcal{M}^\star$ jest niezmienniczy oraz dla każdej miary $\mu \in \mathcal{M}$:

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

Dowód. Funkcja: $V : \mathcal{M} \to \mathbb{R} : \quad V\mu) = \int_A f\, d\mu$ jest funkcją Lapunowa.

Twierdzenie A – przypomnienie

Zakładamy, że A jest zwarty oraz są spełnione warunki

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

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

(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 $\varepsilon > 0$:

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

Załóżmy dodatkowo warunek:

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

Wtedy:

$$\mathrm{Prob}\left(X_t \longrightarrow A^\star, \text{ gdy } t \longrightarrow \infty\right) = 1.$$

Dowód Twierdzenia A

Fakt (*) Jeżeli $\mu^\star \in \mathcal{M}^\star$ oraz $C \in \mathcal{B}(A)$ taki, że $A^\star \subset \operatorname{int} C$, to

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

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

Ustalam $\varepsilon > 0$ i biorę $C = \{a \in A : \operatorname{dist}(a, A^\star) < \varepsilon\}$. Ustalam $\mu_0 \in \mathcal{M}$.

Twierdzenie poprzednie gwarantuje, że zbiór ω-graniczny, $\omega(\mu_0)$, jest zawarty w $\mathcal{M}^\star$. Dla każdego ciągu $t_n \to \infty$, można wybrać podciąg $t_{n_i} \to \infty$ oraz miarę $\mu^\star \in \mathcal{M}^\star$ takie, że $P^{t_{n_i}}\mu_0 \to \mu^\star$. Zauważmy, że ma miejsce Fakt(*). A więc $P^{t_{n_i}}\mu_0(C) \to \mu^\star(C) = 1$. 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:

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

Dowód Twierdzenia A

Fakt (**) Dla każdego $\varepsilon > 0$:

$$\lim_{t\to\infty} \mathrm{Prob}\,(f(X_t) < \varepsilon) = 1.$$

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

Fakt (***)(folklor). Niech $\xi_n : \Omega \to \mathbb{R}$ będzie nierosnącym ciągiem zmiennych losowych zbieżnym stochastycznie do 0. Wtedy, ξ_n zmierza do 0 z prawdopodobieństwem 1.

Ciąg $\xi_t = f(X_t)$ jest stochastycznie zbieżny do 0, a z założenia **(D)** jest nierosnący. Zmierza więc do 0 z prawdopodobieństwem 1. Zwartość A oraz $A^\star$, a także ciągłość f pociągają:

$$\{\omega : f(X_t(\omega)) \to 0,\ \text{gdy}\ \to \infty\} \subset \{\omega : X_t(\omega) \to A^\star,\ \text{gdy}\ \to \infty\},$$

co oznacza, że X_t zmierzają z prawdopodobieństwem 1 do 0.

$\square$

Stochastyczny algorytm ewolucyjny

Podstawowe pojęcia:

f – funkcja kosztu/funkcja przystowoania:
$(\Omega, \Sigma, \mathrm{Prob})$ – przestrzeń probabilistyczna.
Populacja zerowa – skończony, M-elementowy, podzbiór zbioru A.
$T_t : A^M \times \Omega \longrightarrow A^M$ – mechanizm ewolucyjny zależny od czasu.
Kolejne generacje:

$$P_0 \quad \text{– populacja zerowa}$$

$$P_{t+1} = T_t(P_t, Y_{t+1}), \ t = 0, 1, 2, 3, \ldots, \quad \text{– kolejna populacje,}$$

gdzie Y_t – obiekt losowy.

CEL

W kolejnych populacjach znajdować elementy minimalizujące koszt f.
W kolejnych populacjach znajdować elementy maksymalizujące przystosowanie f.

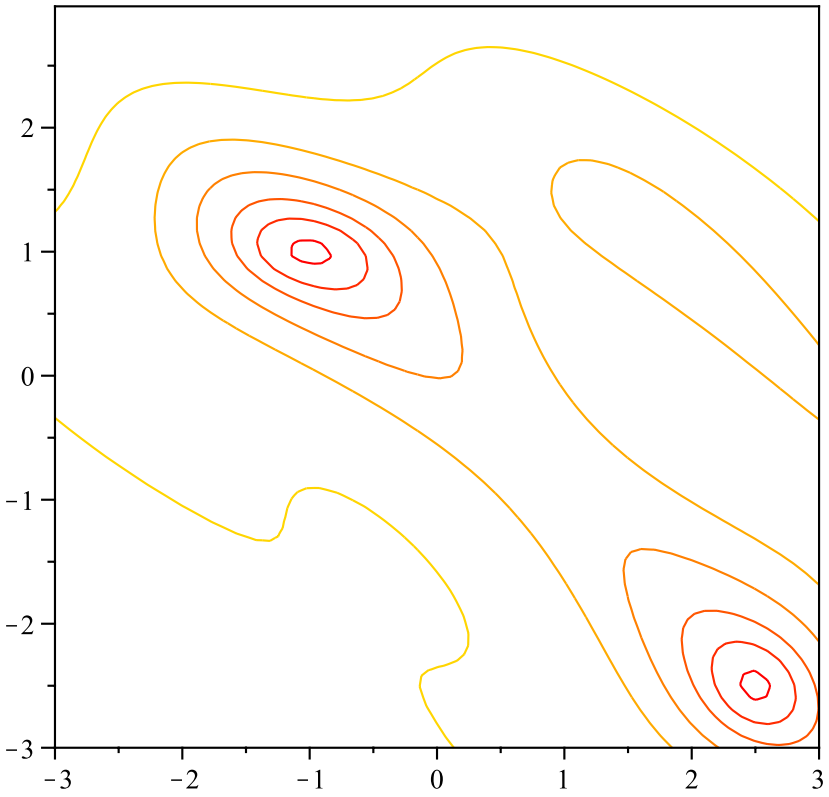
Mechanizmy ewolucyjne

1. Selekcja
2. Krzyżowanie
3. Mutacje
4. Dojrzewanie

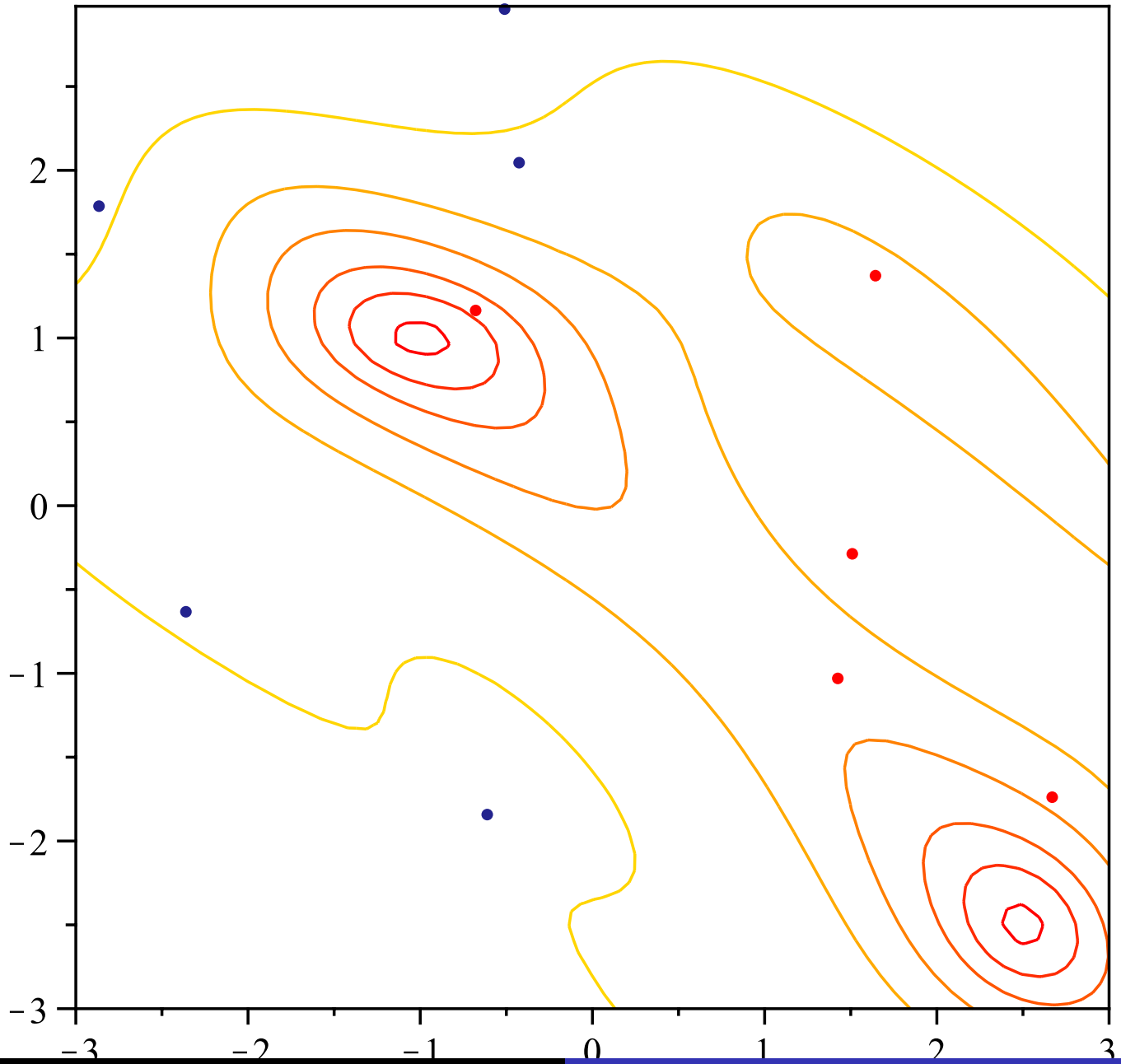
Przykład

$$f : [-3,3] \times [-3,3] \longrightarrow \mathbb{R}$$

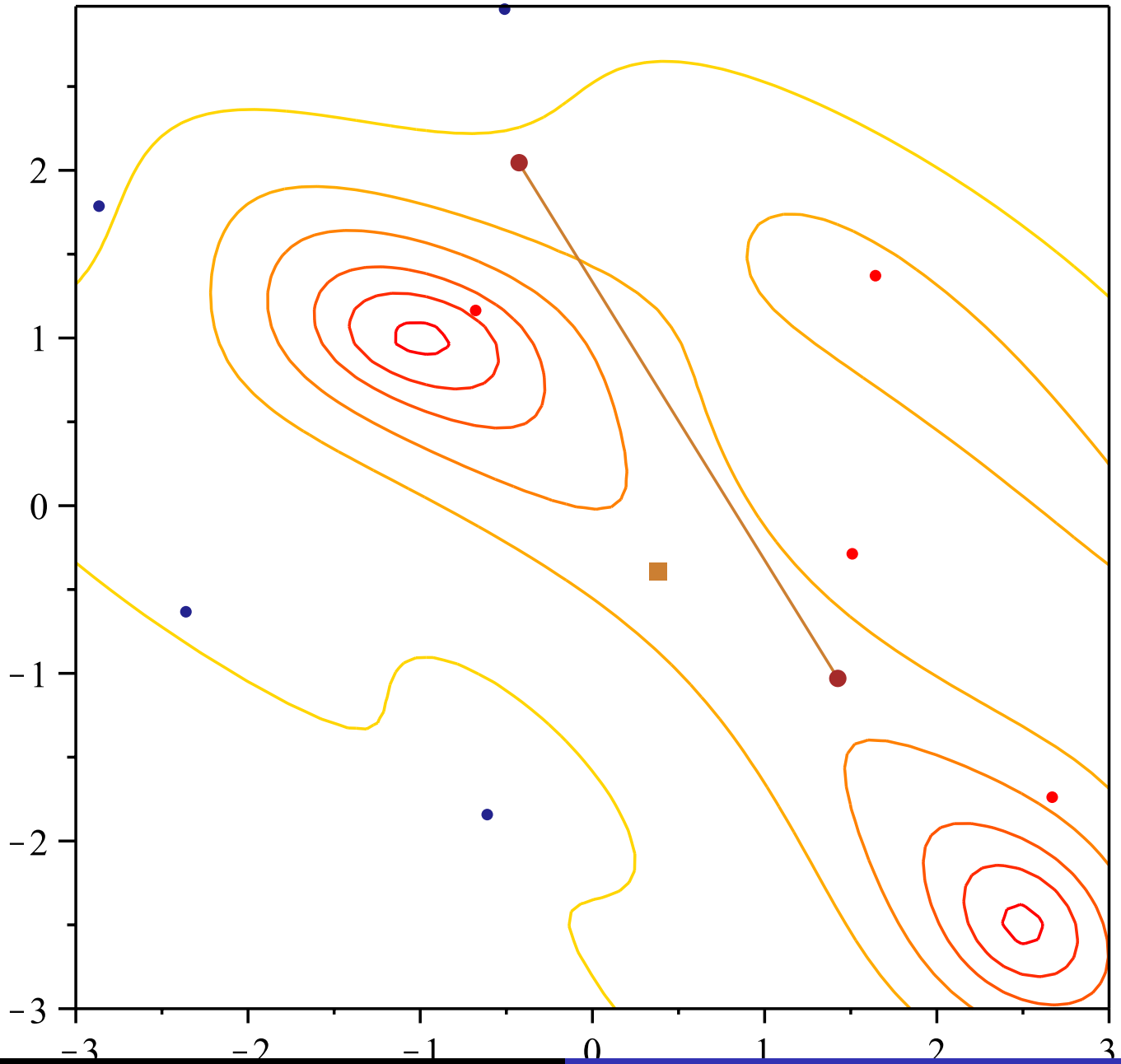
$$f(a,b) = \sqrt{\left((-2.5+a)^2 + (2.5+b)^2\right)\left((1+a)^2 + 4\,(-1+b)^2\right)} + 6\,(\sin(a+b))^2$$



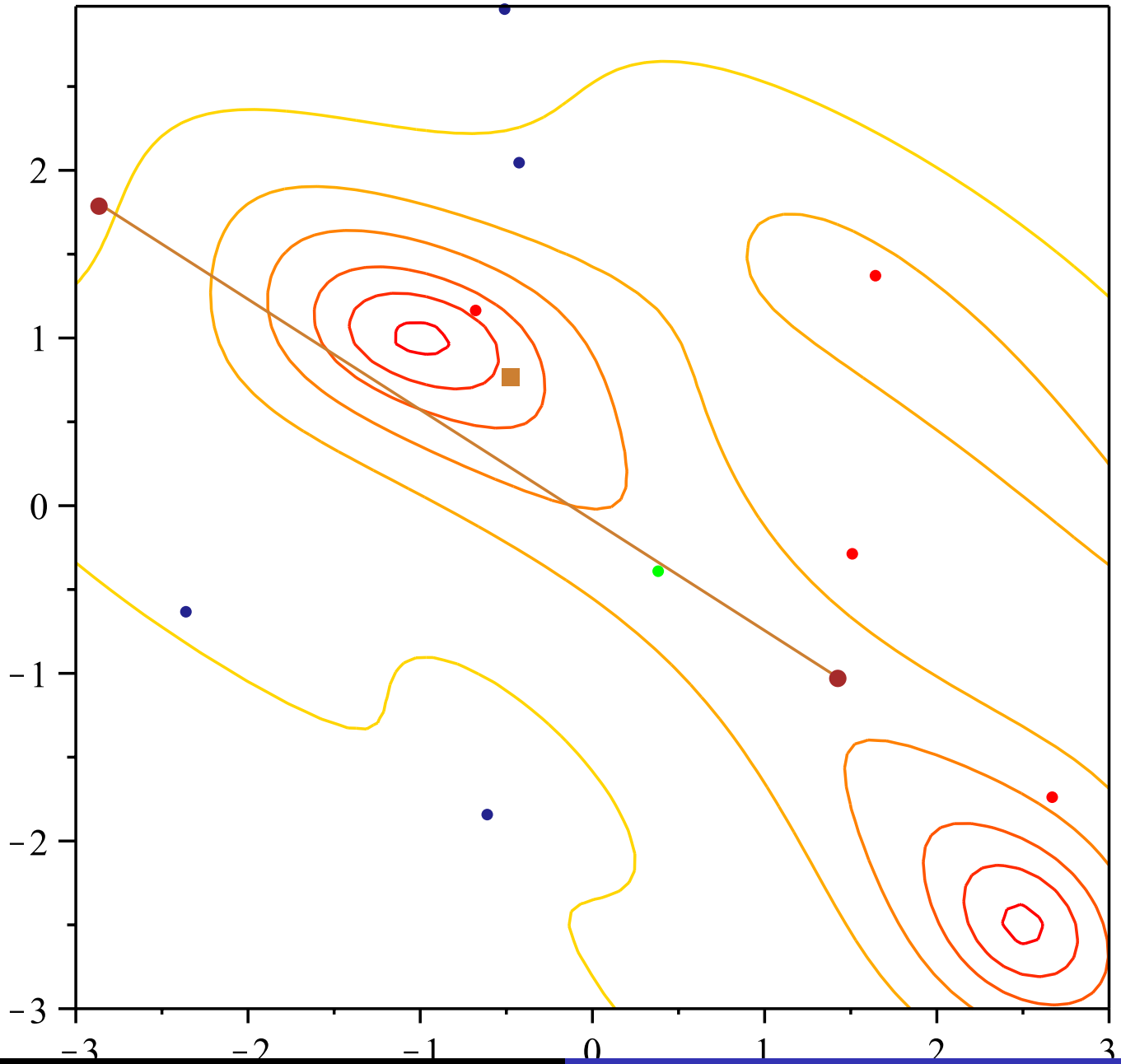
Populacja zerowa



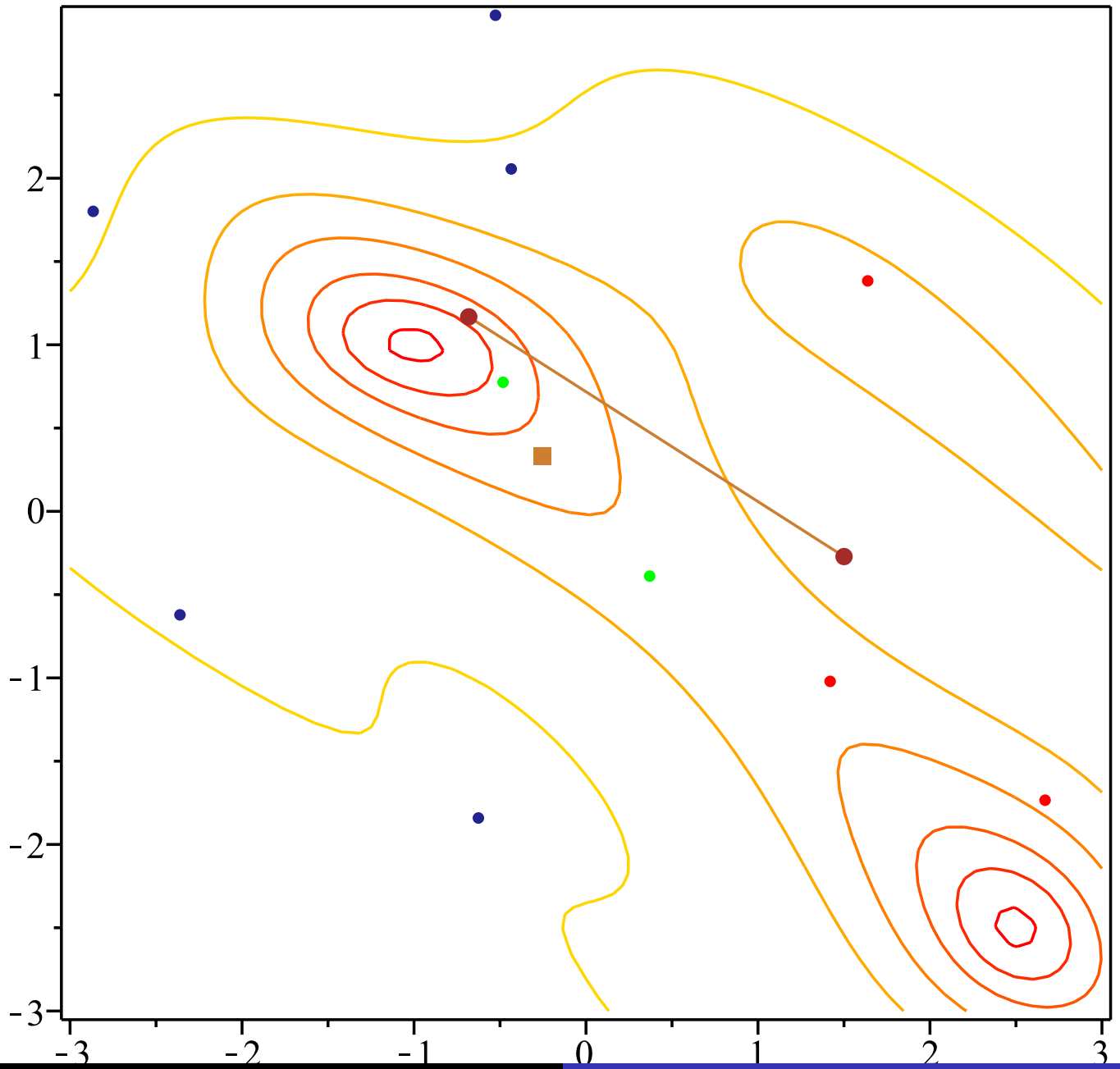
Tworzenie nowej gneracji

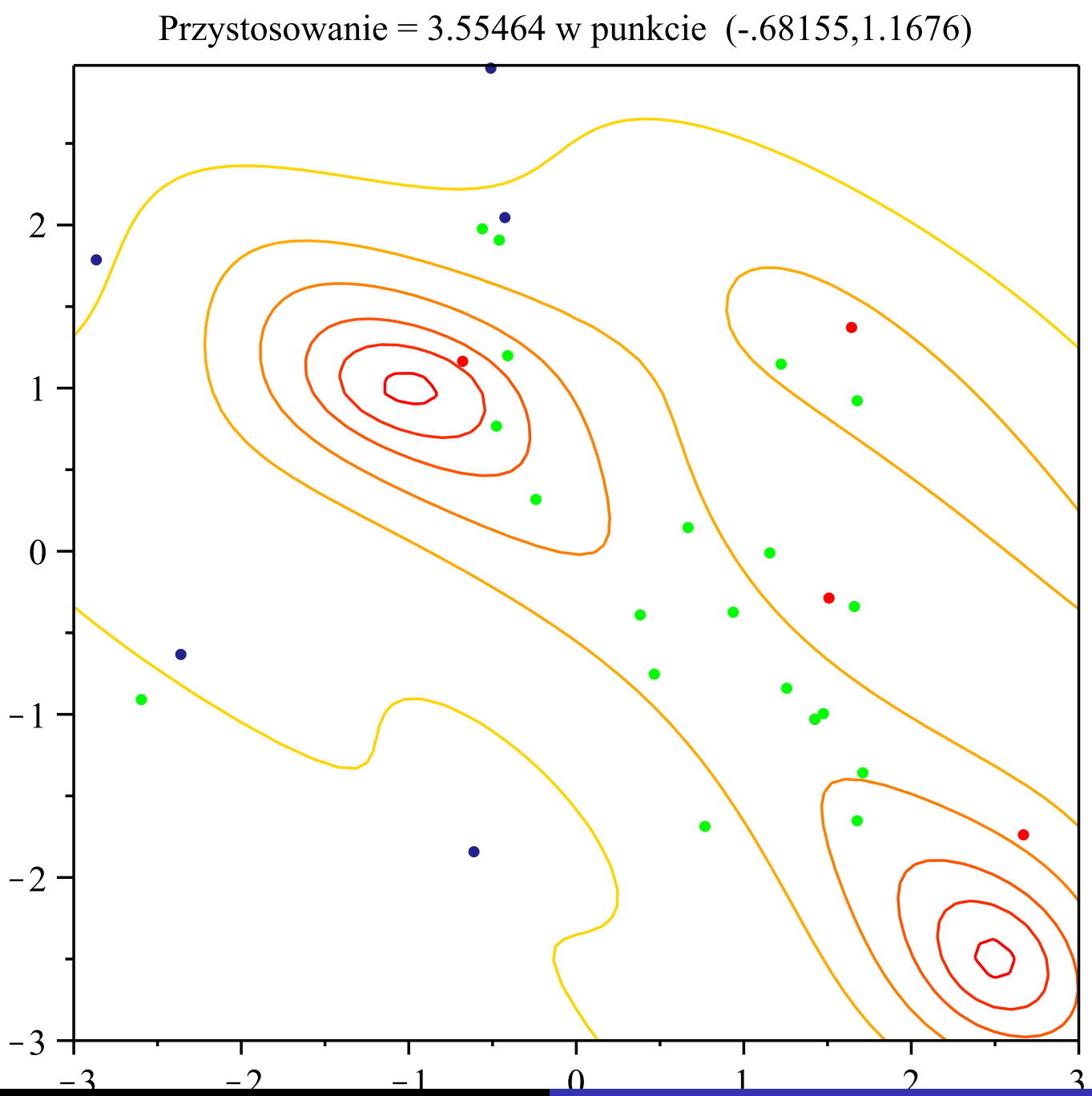


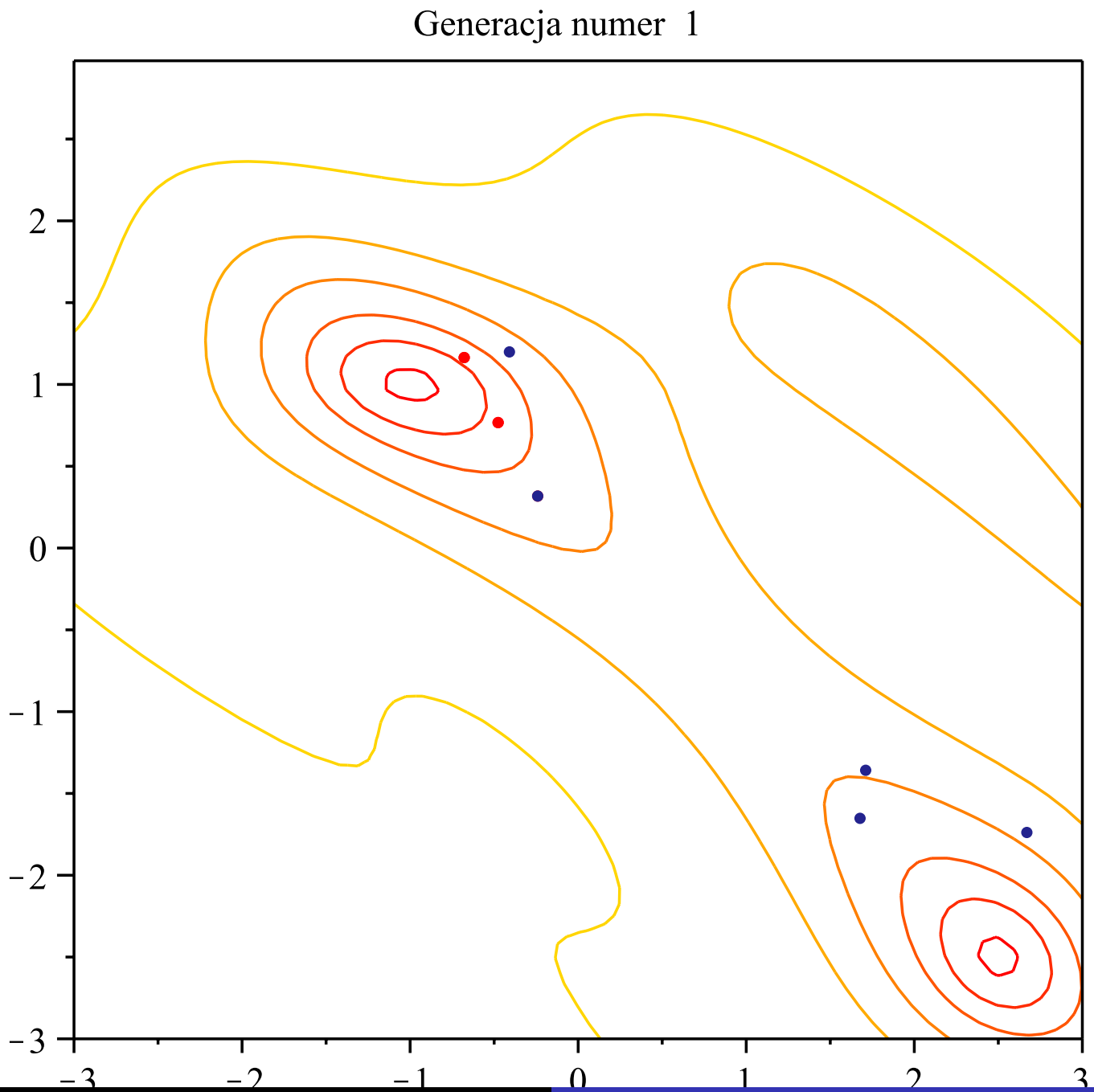
Tworzenie nowej gneracji



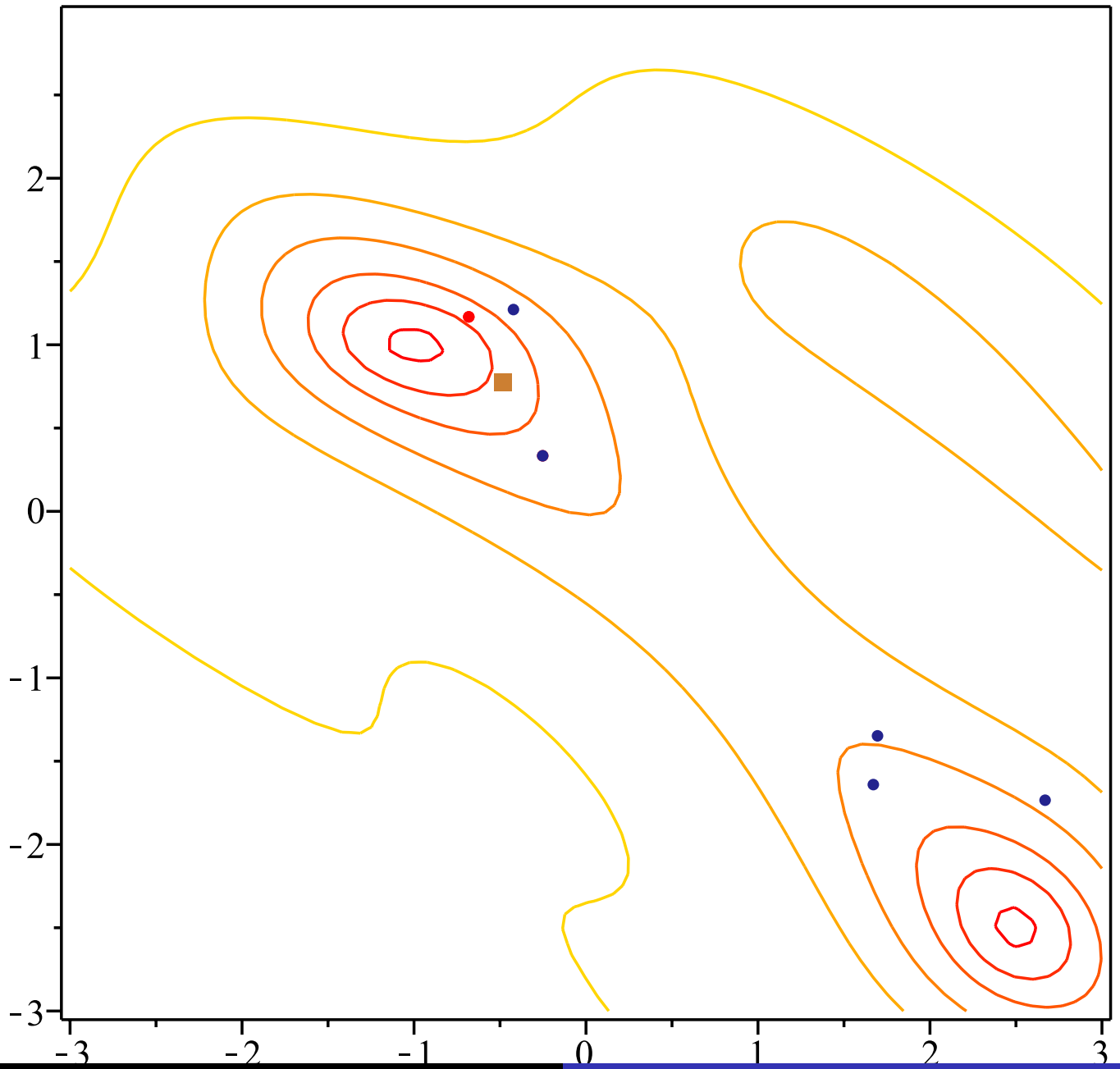
Tworzenie nowej gneracji

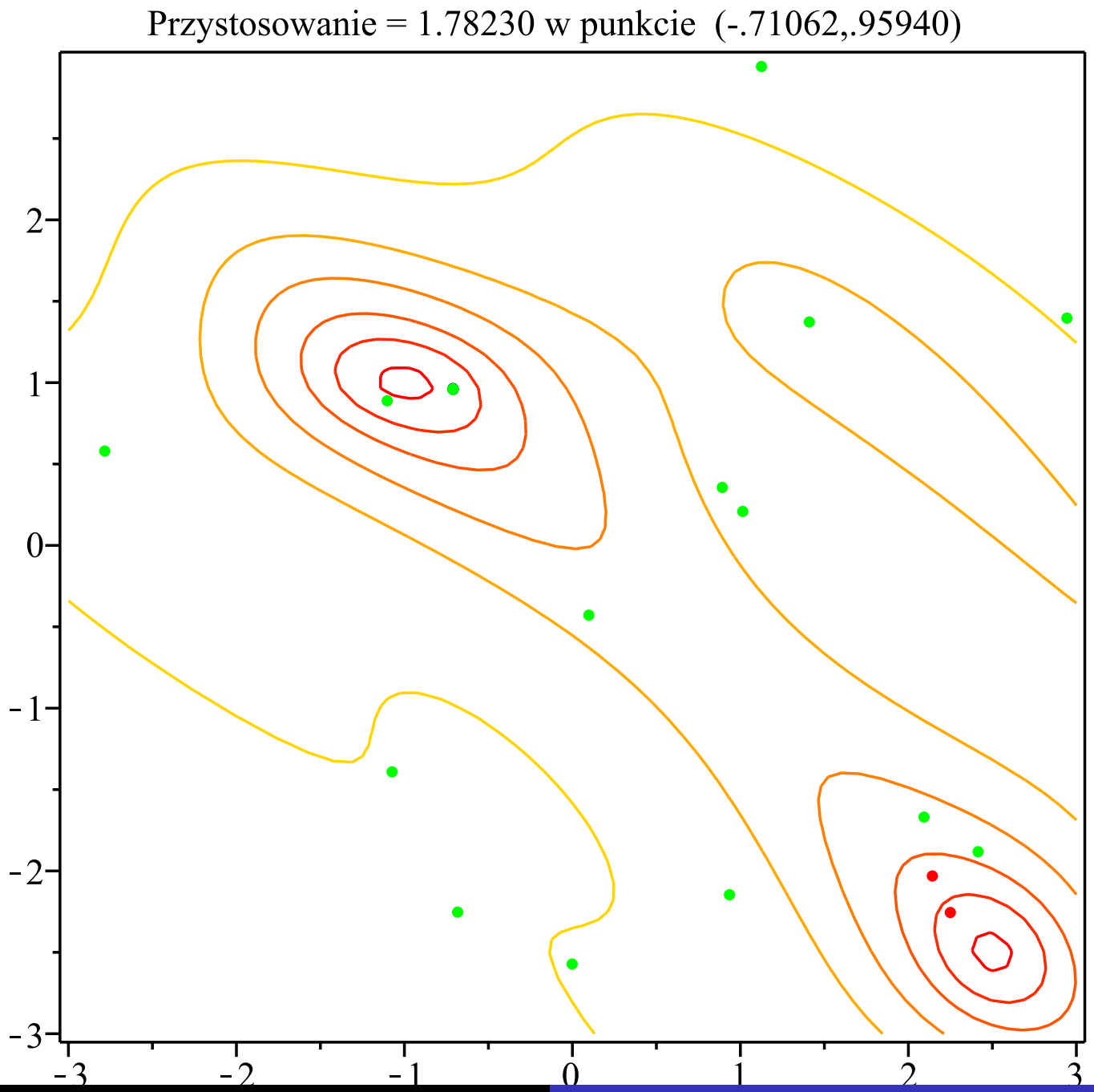


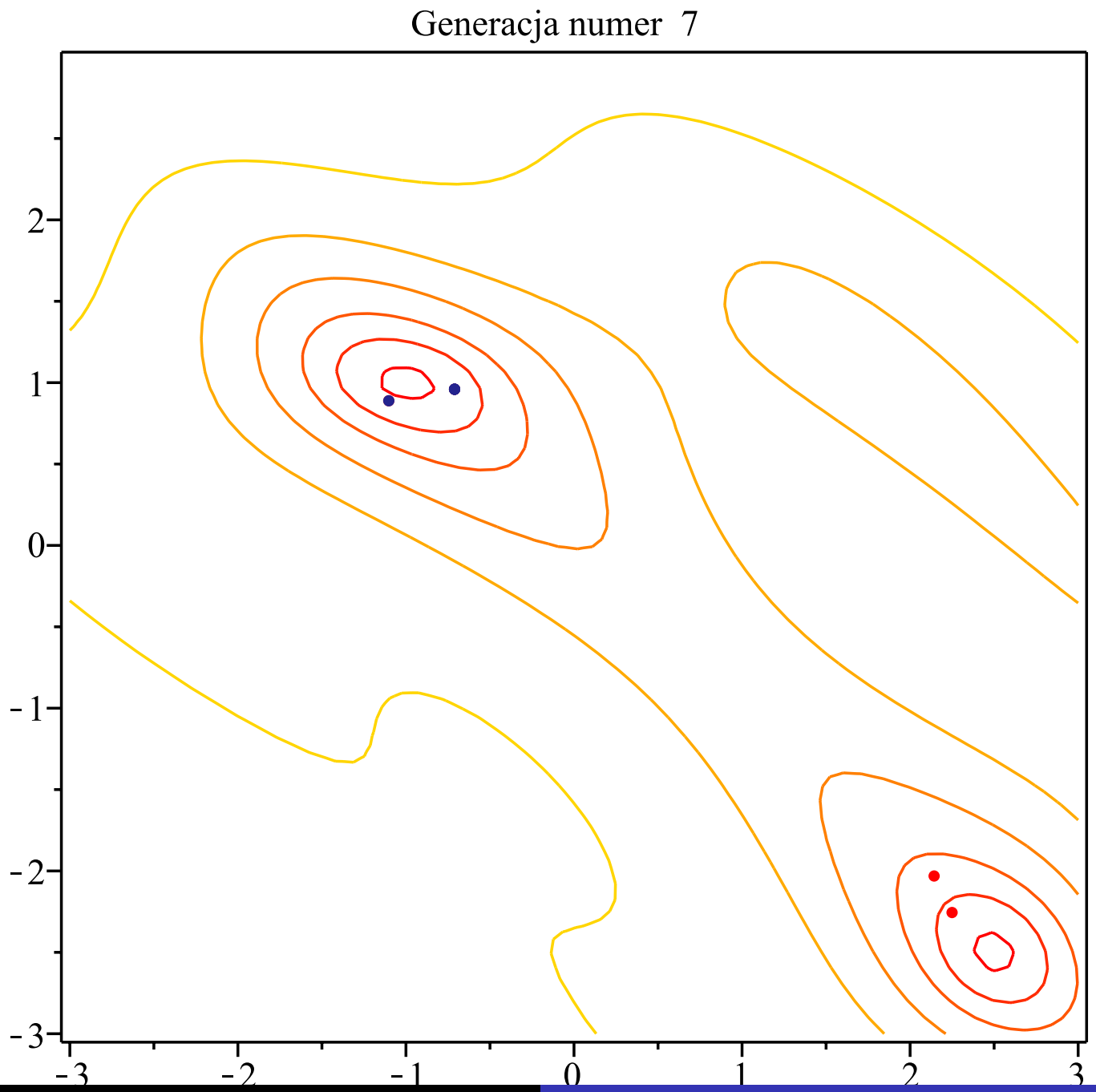

Przystosowanie = 3.55464 w punkcie (-.68155,1.1676)
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

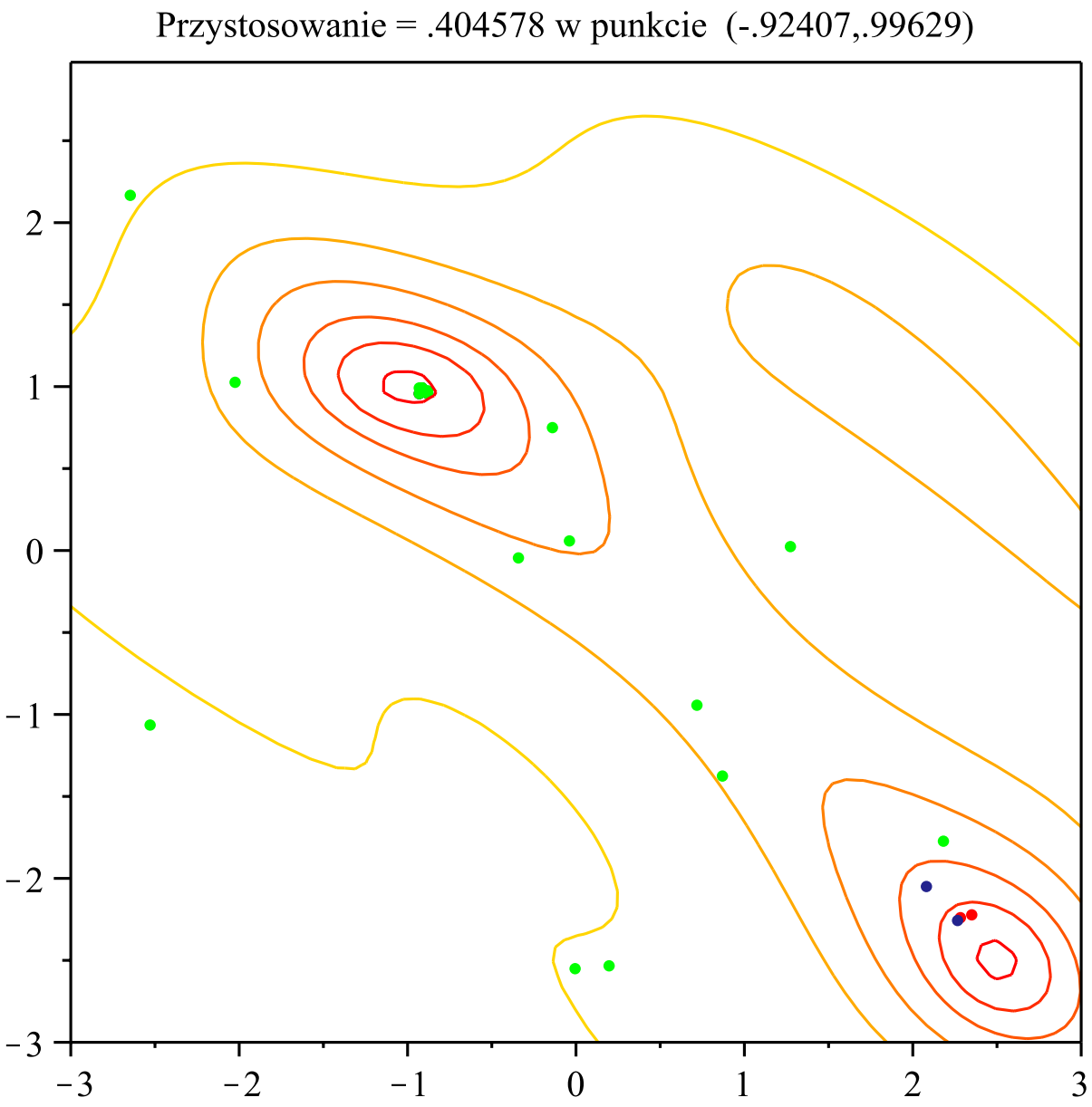

Generacja numer 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

Tworzenie nowej gneracji




Przystosowanie = 1.78230 w punkcie (-.71062,.95940)
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


Generacja numer 7
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


Przystosowanie = .404578 w punkcie (-.92407,.99629)
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

Problem zbieżności

Uwaga

Jeżeli

$$\int \bar{f}(T_t(P, y))\, dy \leqslant \bar{f}(P), \quad P \in A^M,$$

to przy pewnych dodatkowych założeniach ciąg $\{p_t^0\}$ zmierza prawie wszędzie do zbioru $A^\star = \{a \in A : \forall x \in A\ f(x) \geqslant f(a)\}$.
Tutaj: $p_t^0 \in P_t$.
$\bar{f}(p_1, \ldots, p_M) := \min(f(p_1), \ldots, f(p_M)) = f(p_t^0)$.

Układ nieautonomiczny

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

gdzie T_t są mierzalne $A \times B \to A$.

$\mathcal{B}(A)$, $\mathcal{B}(B)$ σ-algebry zbiorów borelowskich na A oraz B
$\mathcal{M}$, $\mathcal{N}$ zbiór wszystkich miar probabilistycznych na $\mathcal{B}(A)$, $\mathcal{B}(B)$.
$X_0 : \Omega \longrightarrow \mathbb{R}^n$ ma rozkład $\mu_0 \in \mathcal{M}$.
$\mathbf{Y}_t :: \Omega \longrightarrow \mathbb{R}^m$ mają rozkład $\nu_t \in \mathcal{N}$, dla $t = 0, 1, 2, \ldots$.
$X_0, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$ są niezależne.

$\mathcal{T}$ - rodzina operatorów $T : A \times B \to A$ z topologią zbieżności jednostajnej.

Układ nieautonomiczny

Niech $U \subset \mathcal{T} \times \mathcal{N}$ będzie zbiorem zwartym.
Zakładamy, że dla każdego $u = (T, \nu) \in U$:

(A) Dla każdego $x_0 \in A$ odwzorowanie $f \circ T$ jest dolnie półciągłe w punktach (x_0, y), takich, że $y \in B \setminus D_T(x_0)$, przy czym $\nu(D_T(x_0)) = 0$.

(B) Dla każdych $x \in A^\star$, $y \in B$: $T(x, y) \in A^\star$.

(C1) Dla każdego $x \in A \setminus A^\star$:

$$\int_B f(T(x, y))\, \nu(dy) \leqslant f(x). \tag{1}$$

Istnieje domknięty podzbiór $U_0 \subset U$ taki, że:

(C2) dla każdych $(T, \nu) \in U_0$, $x \in A \setminus A^\star$:

$$\int_B f(T(x, y))\, \nu(dy) < f(x). \tag{2}$$

Twierdzenie B (Dawid Tarłowski)

Niech ciąg $(T_t, \nu_t) : t \geqslant 1\} \subset U$ zawiera podciąg $\{(T_{t_n}, \nu_{t_n})\} \subset U_0$. Wtedy:

$$\mathrm{Prob}\,(X_t \longrightarrow A^\star, \text{ gdy } \ t \longrightarrow \infty) = 1, \tag{3}$$

to znaczy:

$$\mathrm{Prob}\,(\{\omega \in \Omega\colon \ d(X_t(\omega), A^\star) \longrightarrow 0, \text{ gdy } t \longrightarrow \infty\}) = 1.$$

Optymalizacja oparta na MCMC

Dana jest funkcja $f : A \to \mathbb{R}$. Chcemy znaleźć maksimum globalne f. Ze względu na dużą liczność A nie warto przeszukiwać tego zbioru element po elemencie. Czasami jedynym możliwym podejściem jest losowanie.

Zamiast losować zgodnie z rozkładem jednostajnym na A chcemy losować z takiego rozkładu π, że:

$$\pi(x) \geqslant \pi(y) \iff f(x) \geqslant f(y).$$

Przykład: $\beta > 0$

$$\pi_\beta(x) = \frac{\exp(\beta f(x))}{C}, \quad C = \sum_{y \in A} \exp(\beta f(y))$$

Podstawowa zasada:
Wyznacz próbkę z rozkładu π i znajdź wartość największą f na tej próbce. Punkt x_0^π, w którym ta wartość jest osiągnięta jest „statystycznie" blisko zbioru $\{a \in A : \forall x \in A \ f(x) \leqslant f(a)\}$.

Metoda: Markov Chain Monte Carlo

A – skończony, $\#A$ (bardzo duże)
π – rozkład prawdopodobieństwa na A, $\pi_i > 0$ dla $i \in A$.
Chcemy losować punkty ze zbioru A zgodnie z rozkładem π.

Metoda MCMC

Określa się na zbiorze A pewien ergodyczny łańcuch Markowa $\mathbf{P}$, którego stanem stacjonarnym jest π.
Wybierając teraz dowolny $x_0 \in A$ i generując kolejne punkty $x_1, x_2, \ldots$ mamy gwarancję, że dla dużych n, x_n są losowane faktycznie zgodnie z rozkładem π.

Twierdzenie

Algorytm Metropolisa (opisany poniżej w szczególnym przypadku – problem plecakowy) generuje punkty będące realizacjami pewnego ergodycznego łańcucha Markowa o stanie stacjonarnym π.

Problem plecakowy

Maksymalizuj wartość

$$f(x) = c^T x$$

przy ograniczeniu

$$w^T x \leqslant W, \quad x^i \in \{0, 1\},$$

$c = (c^1, \ldots, c^k)$ wektor cen,
$w = (w^1, \ldots, w^k)$ wektor wag,
W – maksymalna waga dopuszczalne.

$$A = \{x \in \{0,1\}^k : w^T x \leqslant W\}$$

Przykładowe dane: $W = 20$,
[cena,waga] = [50,10], [20,5], [20,4], [10,1], [5,3], [5,5], [4,4], [3,3], [3,3], [3,3], [2,2], [3,1], [3,1], [2,1], [2,1], [2,1], [2,1], [1,1], [1,1], [1,1]
Rozwiązanie:

$$x^* = (1, 1, 1, 1, 0, 0, \ldots, 0), \quad f(x^*) = 100.$$

Algorytm Metropolisa w przypadku problemu plecakowego

Ustalmy $\beta > 0$ oraz $T > 0$.
Wybierzmy punkt początkowy $x_1 \in A$. Podstawmy $x \leftarrow x_1$
Od t $= 1$ do $T - 1$ wykonuj:

1. Wylosuj J spośród liczb $1, \ldots, k$.
2. Niech $y = (x^1, \ldots, x^{J-1}, 1 - x^J, x^{J+1}, \ldots, x^k)$.
3. Jeżeli $y \notin A$, to $x_{t+1} \leftarrow x_t$.
   W przeciwnym przypadku:
4. Niech $\alpha = \begin{cases} \exp(-\beta\, c^J), & \text{gdy} \quad x^J = 1 \\ 1 & \text{gdy} \quad x^J = 0 \end{cases}$.
5. Wylosuj $u \in \{0, 1\}$ zgodnie z rozkładem

$$\text{Prob}(0) = 1 - \alpha, \quad \text{Prob}(1) = \alpha.$$

   Jeżeli $u = 0$, wtedy $x_{t+1} \leftarrow x_t$.
   Jeżeli $u = 1$, wtedy $x_{t+1} \leftarrow y$.

Dla dużych T, X_T ma w przybliżeniu rozkład π_β.

Przykład: $T = 100$, $\beta = 0.5$, $x_1 = (0, \ldots, 0)$.

$$\begin{bmatrix}
[1,0,1,1,1,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0] & [88,20] \\
[0,1,1,1,0,0,1,1,0,0,0,1,1,1,0,0,0,0,0,0] & [65,20] \\
[0,0,1,1,1,1,0,0,0,0,1,1,1,1,1,0,1,0,0,0] & [54,20] \\
[0,1,0,1,1,0,1,0,0,0,1,1,0,1,0,1,1,0,0,1] & [51,20] \\
[0,1,1,1,0,0,0,1,0,0,1,0,1,1,0,0,1,0,1,1] & [64,20] \\
[0,0,1,1,0,0,1,0,0,0,1,0,1,1,0,1,0,1,1,1] & [46,17] \\
[0,1,0,1,0,1,0,0,1,1,1,1,0,0,0,0,0,0,0,0] & [46,20] \\
[1,0,0,1,1,0,0,0,1,0,0,0,1,0,1,1,0,0,0,0] & [75,20] \\
[1,0,0,0,0,0,0,1,0,0,0,1,1,1,1,1,1,1,0,0] & [68,20] \\
[0,0,1,0,1,1,0,0,0,0,1,1,1,0,0,1,1,0,1,1] & [44,20]]
\end{bmatrix}$$

Metoda: Simulated Annealing

W algorytmie Metropolisa wolno zwiększamy β:
Na przykład:

$$\beta(t) = \log t,$$

$$\beta(t) = 1.0001^t,$$

$$\beta(t) = b_0 \left(1 + \frac{1}{T}\right)^{at},$$

gdzie T jest liczbą korków, b_0 oraz a są dodatnie.

Przykład: $T = 100$, $\beta = \beta(t) = 0.5(1 + \frac{1}{T})^t$, $x_1 = (0, \ldots, 0)$.

$$\begin{bmatrix} [1,0,0,1,0,0,0,0,1,1,0,0,1,0,0,1,0,1,0,0] & [72,20] \\ [0,1,0,1,1,0,1,1,0,0,0,1,0,0,0,1,1,0,0,1] & [50,20] \\ [0,0,1,1,0,0,1,1,0,0,1,1,0,1,1,0,1,1,1,0] & [50,20] \\ [0,0,1,1,1,1,0,0,0,1,0,0,0,0,0,1,1,1,1,0] & [49,20] \\ [0,1,1,1,0,0,0,0,1,0,0,1,0,1,1,1,1,1,0,0] & [65,19] \\ [0,1,0,1,1,0,1,0,0,0,1,1,1,0,1,1,0,1,0,0] & [52,20] \\ [0,1,1,1,0,0,0,0,1,1,0,1,1,0,0,1,1,0,0,0] & [66,20] \\ [0,0,1,1,0,0,0,1,0,1,1,1,1,0,1,1,1,0,1,1] & [52,20] \\ [0,1,1,0,0,0,0,1,1,0,1,0,1,1,0,0,1,0,0,0] & [55,20] \\ [0,0,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,0,0,1] & [55,20] \end{bmatrix}$$

Dalsze modyfikacje algorytmu Metropolisa

Oryginalny algorytm

Ustalmy $\beta > 0$ oraz $T > 0$.
Wybierzmy punkt początkowy $x_1 \in A$. Podstawmy $x \leftarrow x_1$
Od t $= 1$ do $T - 1$ wykonuj:

1. Wylosuj J spośród liczb $1, \ldots, k$.
2. Niech $y = (x^1, \ldots, x^{J-1}, 1 - x^J, x^{J+1}, \ldots, x^k)$.
3. Jeżeli $y \notin A$, to $x_{t+1} \leftarrow x_t$.
   W przeciwnym przypadku:
4. Niech $\alpha = \begin{cases} \exp(-\beta\, c^J), & \text{gdy} \quad x^J = 1 \\ 1 & \text{gdy} \quad x^J = 0 \end{cases}$.
5. Wylosuj $u \in \{0, 1\}$ zgodnie z rozkładem

$$\text{Prob}(0) = 1 - \alpha, \quad \text{Prob}(1) = \alpha.$$

   Jeżeli $u = 0$, wtedy $x_{t+1} \leftarrow x_t$.
   Jeżeli $u = 1$, wtedy $x_{t+1} \leftarrow y$.

Dalsze modyfikacje algorytmu Metropolisa

Modyfikacje

Ustalmy funkcję β (wolno) rosnącą, $\beta(t) > 0$ oraz $T > 0$.
Wybierzmy punkt początkowy $x_1 \in A$. Podstawmy $x \leftarrow x_1$
Od t $= 1$ do $T - 1$ wykonuj

1. Wylosuj J spośród liczb $1, \ldots, k$ według rozkładu $\mathrm{Prob}(J = i) = \frac{c^i}{\sum_{s=1}^{k} c^s}$.
2. Niech $y = (x^1, \ldots, x^{J-1}, 1 - x^J, x^{J+1}, \ldots, x^k)$.
3. Jeżeli $y \notin A$, to $x_{t+1} \leftarrow x_t$.
   W przeciwnym przypadku:
4. Niech $\alpha = \begin{cases} \exp(-\beta(t)c^J), & \text{gdy} \quad x^J = 1 \\ 1 & \text{gdy} \quad x^J = 0 \end{cases}$.
5. Wylosuj $u \in \{0, 1\}$ zgodnie z rozkładem

$$\mathrm{Prob}(0) = 1 - \alpha, \quad \mathrm{Prob}(1) = \alpha.$$

   Jeżeli $u = 0$, wtedy $x_{t+1} \leftarrow x_t$.
   Jeżeli $u = 1$, wtedy $x_{t+1} \leftarrow y$.

Przykład: $T = 100$, $\beta = \beta(t) = 0.5(1 + \frac{1}{T})^t$, $x_1 = (0, \ldots, 0)$.

$$\begin{bmatrix} [1,1,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0] & [93,20] \\ [0,1,1,1,0,1,1,0,0,0,0,0,0,0,0,0,0,1,0,0] & [60,20] \\ [1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] & [100,20] \\ [1,1,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0] & [93,20] \\ [1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] & [100,20] \\ [1,1,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0] & [86,20] \\ [1,0,1,1,0,0,0,0,0,0,1,1,0,0,1,0,0,1,0,0] & [88,20] \\ [1,0,1,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0] & [76,20] \\ [0,1,1,1,1,0,0,0,1,0,1,0,0,0,0,1,1,0,0,0] & [64,20] \\ [1,0,1,1,0,0,0,0,0,0,1,0,1,0,1,1,0,0,0,0] & [89,20] \end{bmatrix}$$

SIMULATED ANNEALING

$f : A \to \mathbb{R}$ – ciągła, A – zwarty, $B \subset \mathbb{R}^m$

Dane:
ξ_t – ciąg i.i.d o rozkładzie ν na B.
r_t – ciąg i.i.d o rozkładzie jednostajnym na $(0, 1)$.
$\beta_t > 0$ – ciąg liczb, taki, że $\lim_{t\to\infty} \beta_t = \infty$.
$Q : A \times B \to A$ – odwzorowanie mierzalne.
X_0 zmienna losowa na A niezależna od zmiennych ξ_t, r_t,

$$X_{t+1} = \begin{cases} Q(X_t, \xi_{t+1}) & \text{gdy } f(Q(X_t, \xi_{t+1})) < f(X_t) \\ Q(X_t, \xi_{t+1}) & \text{gdy } f(Q(X_t, \xi_{t+1})) \geqslant f(X_t) \text{ oraz} \\ & r_t \leqslant \exp(-\beta_t) \cdot |f(Q(X_t, \xi_{t+1})) - f(X_t)|) \\ X_t & \text{ w przeciwnym przypadku} \end{cases}$$

SIMULATED ANNEALING

Twierdzenie, Dawid Tarłowski

Załóżmy, że:

(E) dla każdego $x \in A$ miara ν zbioru punktów nieciągłości odwzorowania $y \longrightarrow Q(x, y)$ równa się zeru.

(F) dla każdego $x \in A \setminus A^\star$, $\nu(\{y \in B : f(Q(x, y)) < f(x)\}) > 0$.

Wtedy dla każdego $\varepsilon > 0$:

$$\mathrm{Prob}(d(X_t, A^\star) < \varepsilon) \longrightarrow 0, \text{ dla } t \to \infty.$$

oraz

$$E(f(X_t)) \longrightarrow \min_A f, \text{ dla } t \to \infty.$$

Algorytmu Gibbsa

Niech $A = \mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_k$
Często $\mathcal{X}$ to $\mathbb{R}^k$ lub jej podzbiór. Albo $\mathcal{X} \subset \{0,1\}^k$.

Niech π będzie rozkładem na A (niekoniecznie w pełni znanym)
Załóżmy, że dla dowolnego $x \in \mathcal{X}$ potrafimy losować ze wszystkich rozkładów warunkowych postaci

$$\pi(\cdot|x_{-i}),$$

gdzie $i = 1, \ldots, k$ oraz $x_{-i} = (x_1 \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$.

Zadanie: wylosować próbkę z rozkładu π.

Algorytmu Gibbsa

Problem plecakowy c.d.

$$A = \{x \in \{0,1\}^k : w^T x \leqslant W\}$$

$$\pi_\beta(x) = \frac{\exp(\beta f(x))}{C}, \quad C = \sum_{y \in A} \exp(\beta f(y)),$$

gdzie $f(x) = c^T x$, c – wektor cen, $\beta > 0$ – parametr algorytmu.

$\pi_\beta(\cdot|x_{-i})$ – rozkład dwupunktowy
Gdy $w_{-i}^T x_{-i} + c_i \leqslant W$, wtedy $\pi_\beta(0|x_{-i}) = \frac{1}{1+e^{c_i}}$, $\pi_\beta(1|x_{-i}) = \frac{e^{c_i}}{1+e^{c_i}}$.
Gdy $w_{-i}^T x_{-i} + c_i > W$, wtedy $\pi_\beta(0|x_{-i}) = 1$, $\pi_\beta(1|x_{-i}) = 0$.

Algorytmu Gibbsa

Losowy algorytm Gibbsa

0. $X := x_0 \in \mathcal{X}$ (ustalony)
1. $i \sim U(\{1, \ldots, k\})$
2. $x_i' \sim \pi_0(\cdot | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$
3. $X = (x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_k)$
4. przejdź do kroku 1.

Przykład: $T = 100$, $\beta = 0.5$, $x_0 = (0, \ldots, 0)$.

$$\begin{bmatrix}
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,1,1,1,0,1,0] & [63,19] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,1,0,1,0,1,0] & [61,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,1,0] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,1,0] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,1,0] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,1,0] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,0,1] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,0,1] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,0,1] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,1,0,1] & [60,18] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,0,0,1] & [59,17] \\
[1,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,1,0,0,1] & [59,17]
\end{bmatrix}$$

Algorytmu Gibbsa

Losowemu Algorytmowi Gibbsa odpowiada pewien łańcuch Markowa $(X_n)_{n\in\mathbb{N}}$ o rozkładzie stacjonarnym π.

Jądro przejścia algorytmu ma postać:

$$P(x,\cdot) = \frac{1}{k}\sum_{i=1}^{k} P_i(x,\cdot),$$

gdzie $P_i(x,\cdot)$ jest rozkładem o funkcji gęstości:

$$p_i(x,y) = \mathbf{1}_{\{x_{-i}\}}(y_{-i}) \cdot \pi(y_i|x_{-i}).$$

$\mathbf{1}_{\{x_{-i}\}}$ oznacza tutaj funkcję charakterystyczną w punkcie x_{-i}, tzn. wszystkie współrzędne x i y są równe poza indeksem i-tym

Algorytmu Gibbsa

Twierdzenie, Iwona Żerda

Jeżeli rozkład π na $\mathbb{R}^k$ ma podwójnie różniczkowalną i dodatnią funkcję gęstości p taką, że

$$\forall_{i=1,\ldots,k} \quad \limsup_{|x|\to\infty} \frac{\partial^2}{\partial x_i^2} \log p(x) < 0,$$

$$\exists_{\delta>0} \quad \liminf_{|x|\to\infty} \min_i \sup_{|t|<\delta} sgn(x_i) \cdot \frac{\partial}{\partial x_i} \log p(x + sgn(x_i) t e_i) = -\infty,$$

to losowy algorytm Gibbsa o rozkładzie stacjonarnym π jest geometrycznie zbieżny. To znaczy, istnieje stała $\rho < 1$ oraz funkcja $M : \mathcal{X} \to [0, \infty)$ takie, że

$$\forall_{x\in\mathcal{X} \ \pi-p.w.} \quad ||P^n(x,\cdot) - \pi(\cdot)||_{tv} \leqslant M(x)\rho^n,$$

gdzie $||\mu - \nu||_{tv} = \sup_{B\in\mathcal{B}(\mathcal{A})} |\mu(B) - \nu(B)|$.

Niektóre podręczniki

1. J. Arabas, Wykłady z algorytmów ewolucyjnych, WNT, 2004.
2. j. Kusiak, A. Danielewska-Tułecka, Piotr Oprocha, Optymalizacja, wybrane metody z przykładami zastosowań. PWN, 2009.
3. S. Luke, Essentials of Metaheuristics, preprint.
4. N. Madras, Lectures on Monte Carlo Methods, AMS, 2002.
5. A.P. Meyn and R.L. Tweedie, Markov Chains and Stochastic Stability, Springer Verlag 1993.
6. W. Niemiro, Symulacje stochastyczne i metody Monte Carlo, UW 2013.
7. R. Schaefer, Podstawy genetycznej optymalizacji globalnej, UJ, 2002

Ali Ahrari, Ali Asghar Atai, *Grenade Explosion Method-A novel tool for optimization of multimodal functions*, preprint, 2009.

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

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.

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 – 642.

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 – 471.

Jorge Nocedal and Stephen J. Wright, *Numerical Optimization*, Springer, 1999.

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

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

J. Ombach, D. Tarłowski, *Nonautonomous Stochastic Search in Global Optimization*, J. Nonlinear Sci., DOI: 10.1007/s00332-011-9112-3

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

Marcin Radwańsli, *Convergence of nonautonomous evolutionary algorithm*, UIAM, 45(2007), 197 - 206.

Darrell Whitleya, Soraya Ranaa, John Dzuberaa and Keith E. Mathiasb, *Evaluating evolutionary algorithms*, Articicial Intelligence 85(1996) 245 - 276..

D Z I Ę K U J Ę
Z A U W A G Ę