Raport do zadania 2

Strategia ewolucyjna $\mu + \lambda$

Kacper Siemionek

Numer indeksu: 331430

1. Informacje o badanej funkcji oraz założeniach strategii

Funkcja użyta do testowania strategii ewolucyjnej $\mu + \lambda$ jest określona następującym wzorem:

$$f(x, y) = \frac{xy}{e^{x^2+0,5x+y^2}}$$

Obliczając ekstrema lokalne powyższej funkcji korzystając z metod analitycznych uzyskano następujące wyniki:

$$\max\left\{\frac{xy}{e^{x^2+0.5x+y^2}}\right\} = \left\{\left(-\frac{1}{8}-\frac{\sqrt{33}}{8}, -\frac{1}{\sqrt{2}}\right), \left(\frac{\sqrt{33}}{8}-\frac{1}{8}, \frac{1}{\sqrt{2}}\right)\right\} \approx$$

$$\approx \{(-0{,}8430703308, -0{,}7071067812), (0{,}5930703308, 0{,}7071067812)\}$$

$$\min\left\{\frac{xy}{e^{x^2+0.5x+y^2}}\right\} = \left\{\left(-\frac{1}{8}-\frac{\sqrt{33}}{8}, \frac{1}{\sqrt{2}}\right), \left(\frac{\sqrt{33}}{8}-\frac{1}{8}, -\frac{1}{\sqrt{2}}\right)\right\} \approx$$

$$\approx \{(-0{,}8430703308, 0{,}7071067812), (0{,}5930703308, -0{,}7071067812)\}$$

Mutacja osobników polega na dodaniu do ich wartości szumu gaussowskiego, natomiast wynikiem krzyżowania osobników o_1 i o_2 jest osobnik $a\, o_1 + (1-a)o_2$, gdzie a jest zmienną losową z rozkładu jednostajnego na przedziale $[0; 1]$. Strategia selekcji polega na losowym wyborze dwóch grup rodziców o rozmiarze λ z populacji μ. Strategia eliminacji natomiast odbywa się poprzez wybór μ najlepszych osobników z połączonej wcześniej populacji o rozmiarze $\mu + \lambda$. Algorytm z taką konfiguracją jest w stanie znaleźć ekstrema bardzo zbliżone do tych obliczonych analitycznie.

2. Przestrzeń poszukiwań, potrzeba gradientu

Podczas badania funkcji z poprzedniego punktu, każdy z osobników jest reprezentowany przez dwuwymiarowy wektor. Tym samym przestrzenią poszukiwań jest dwuwymiarowa płaszczyzna możliwych rozwiązań, które są rozważane przez algorytm podczas jego działania.

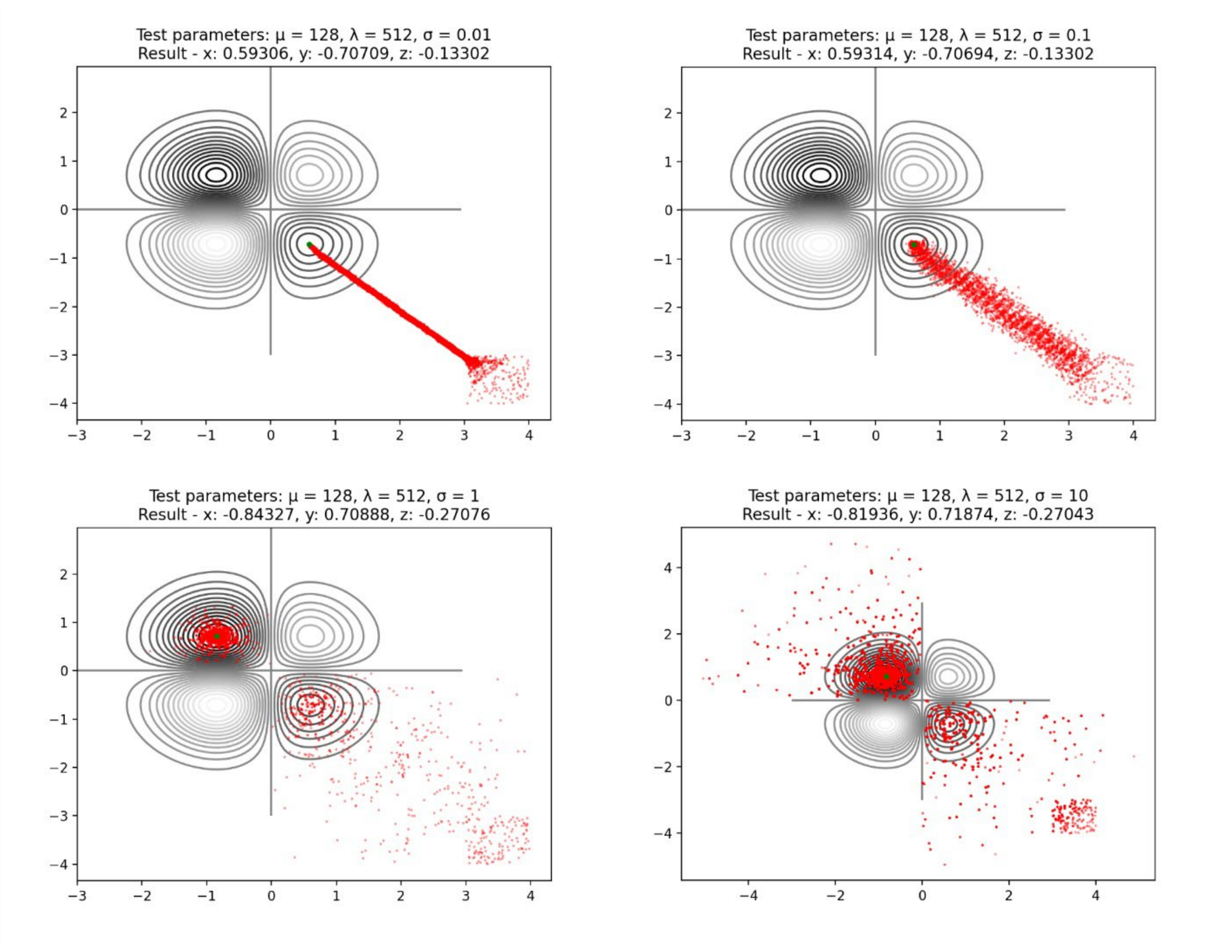
W strategii ewolucyjnej $\mu + \lambda$ nie ma potrzeby korzystania z gradientu. Proces poszukiwania ekstremów lokalnych przebiega za pomocą mutacji, selekcji, krzyżowania i eliminacji. Do wymienionych zabiegów nie potrzebujemy pochodnej funkcji. Umożliwia to badanie takich funkcji, które są nieróżniczkowalne lub posiadają wiele ekstremów lokalnych, z czym nie poradziłby sobie algorytm spadku wzdłuż gradientu.

3. Wpływ odchylenia standardowego

Odchylenie standardowe jest kluczowym elementem strategii ewolucyjnej. To, jaką wartość wybierzemy, ma ogromny wpływ na wyniki działania algorytmu. Zbyt duże odchylenie powoduje większe zmiany w mutacjach osobników, a to z kolei może doprowadzić do oddalonych i mniej precyzyjnych wyników. Natomiast przy małym odchyleniu mutacja nie wprowadza znaczących zmian, cała populacja wolniej się rozwija i może utknąć w ekstremum lokalnym, jednak sam wynik może być dokładniejszy.

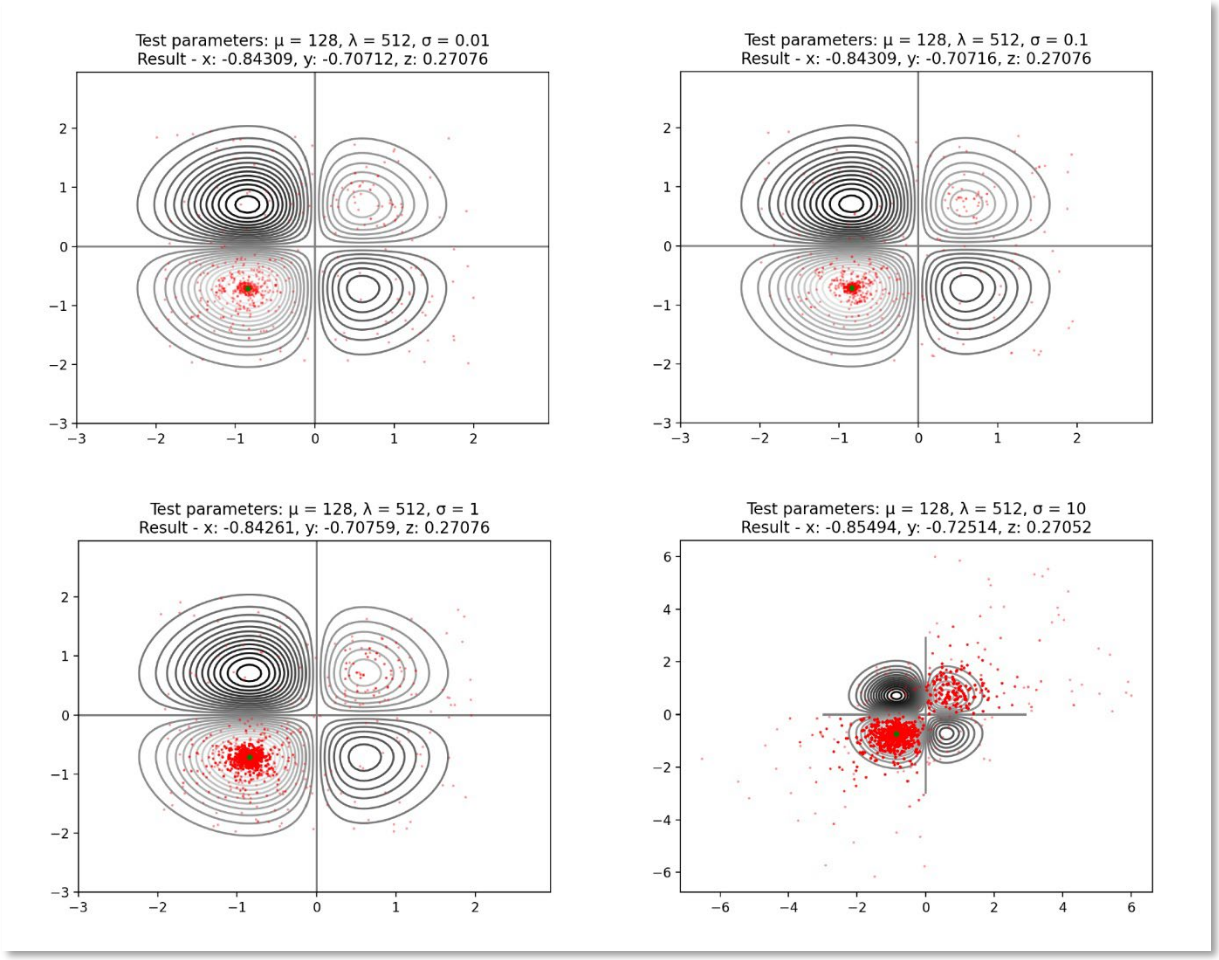
W poniższych eksperymentach ustawiono wartości $\mu, \lambda = (128, 512)$, ilość generacji wynosiła 1000 oraz początkowa populacja została zainicjowana na przedziale:

- $[3, 4] \times [-3, -4]$ dla szukanego minimum.
- $[-2, 2] \times [-2, 2]$ dla szukanego maksimum.



Rysunek 3.1 Wyniki dla różnych wartości odchylenia standardowego (minimum)

Analizując wyniki pierwszego eksperymentu, możemy zauważyć, że dla małych wartości odchylenia standardowego ($\sigma \in \{0.01, 0.1\}$) populacja ostatecznie zatrzymuje się w minimum lokalnym, nie jest w stanie dotrzeć do globalnego, mimo to otrzymaliśmy wynik z dużą dokładnością. Dopiero przy $\sigma = 1$ kolejne mutacje sprawiły, że populacja odnalazła minimum globalne również bliskie prawidłowemu wynikowi. Jednak wartość $\sigma = 10$ doprowadziła do nieoczekiwanych zachowań, kolejne populacje rozproszyły się w różnych kierunkach, otrzymany wynik nieco odbiega od tego wyliczonego analitycznie.



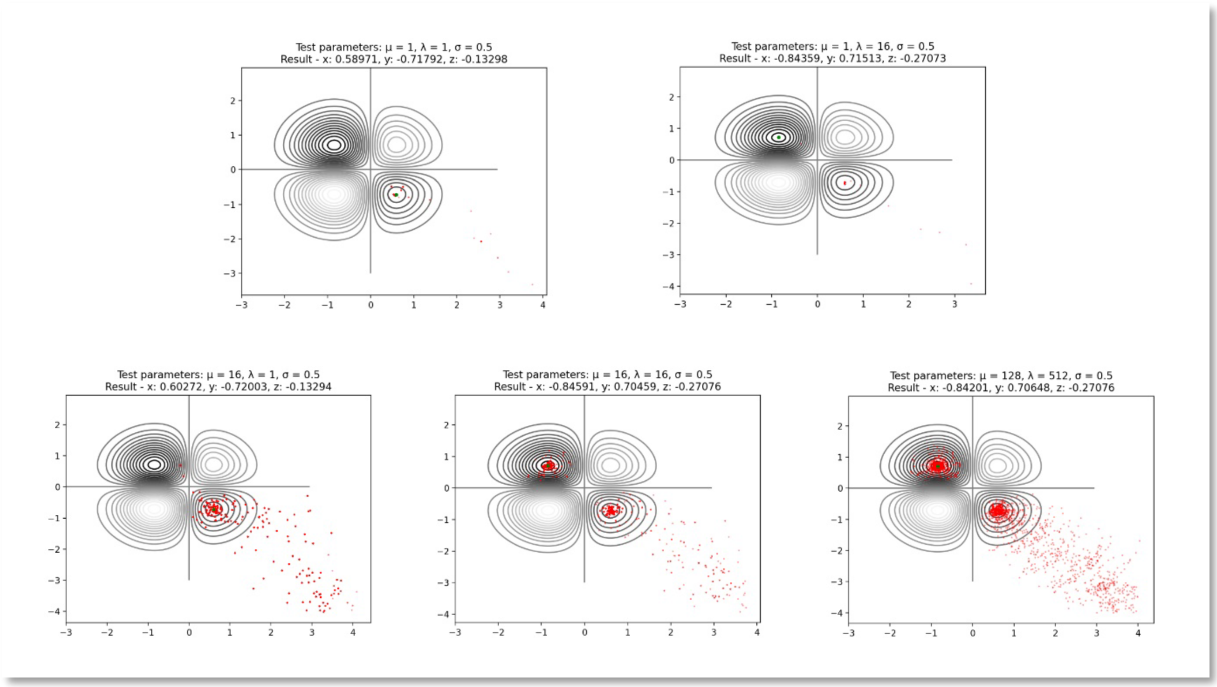
Rysunek 3.2 Wyniki dla różnych wartości odchylenia standardowego (maksimum)

W drugim eksperymencie wybrany został inny przedział do zainicjowania początkowej populacji. Poskutkowało to w znalezieniu maksimum globalnego w każdym przypadku. Mimo to otrzymano różniące się wyniki dla każdej z wartości, najbardziej dokładny został znaleziony podczas testu z wartością $\sigma = 0.01$.

4. Wpływ liczby rodziców μ i dzieci λ

Wybór odpowiednich wartości μ oraz λ ma znaczący wpływ na proces optymalizacji algorytmu. Małe populacje mają ograniczone możliwości eksploracji przestrzeni poszukiwań, co wpływa na wydłużenie całego procesu oraz może powodować utknięcie w ekstremum lokalnym. Natomiast większe populacje są w stanie intensywniej eksplorować daną przestrzeń, co pozytywnie wpływa na wyniki poszukiwań, mimo to czas eksploracji w takiej sytuacji może się wydłużyć.

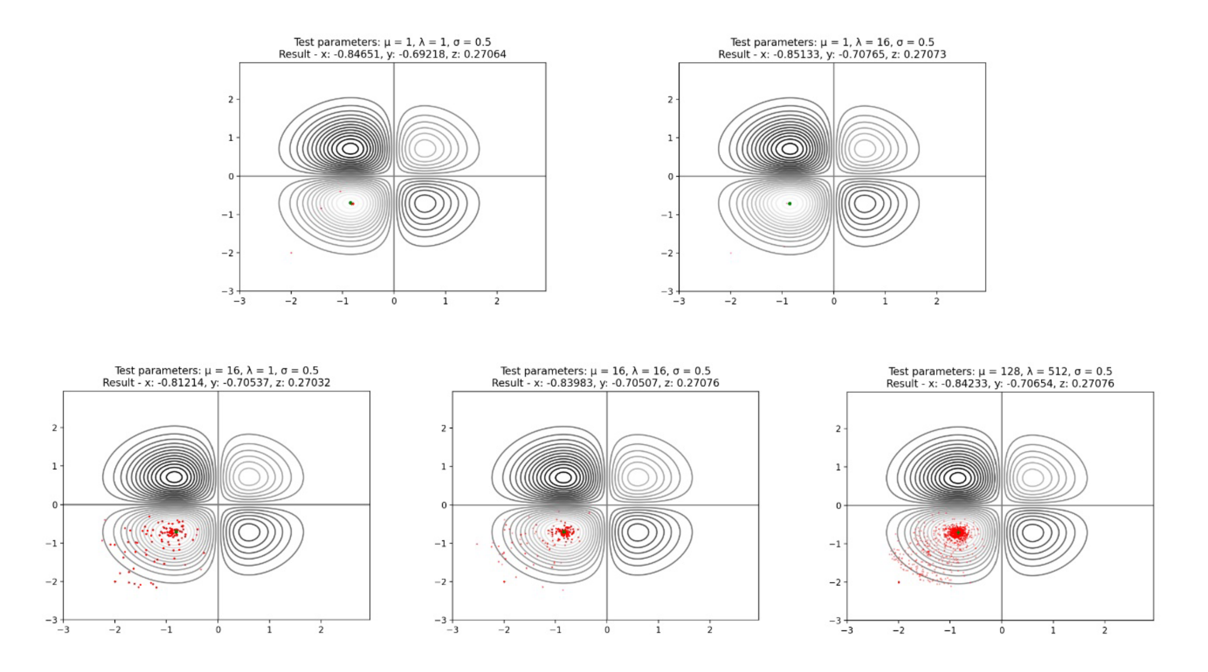
Podobnie jak w poprzednim eksperymencie, w przypadku poszukiwania minimum zainicjowano populację w przedziale $[3, 4] \times [-3, -4]$, a dla maksimum - $[-2, 2] \times [-2, 2]$. Wartość odchylenia standardowego ustawiono na 0.5, a ilość generacji wynosiła 1000.



Rysunek 4.1 Wyniki dla różnych wartości rodziców μ i dzieci λ (minimum)

Wyniki powyższego eksperymentu bardzo dobrze pokazują nam wpływ wartości μ oraz λ. W pierwszym teście liczba rodziców oraz dzieci była zbyt mała, aby populacja prawidłowo się rozwijała, dlatego wynik przedstawia minimum lokalne. Przy większej ilości dzieci istnieje szansa na otrzymanie potomka bliższego ekstremum globalnego, co możemy zauważyć w drugim teście. Kolejny test przedstawia sytuację, w której populacja mimo większej liczbie rodziców ma mniejsze szanse na rozwój w prawidłowym kierunku ze względu na małą liczbę dzieci, co spowodowało utknięcie w minimum lokalnym. Z kolei wartości $\mu, \lambda = (16, 16)$ wprowadzają balans pomiędzy ilością rodziców a dzieci, dzięki czemu możliwe było znalezienie minimum globalnego.

Podobnie stało się w ostatnim teście, jednak z uwagi na rozmiar populacji, poszukiwanie trwało nieco dłużej.



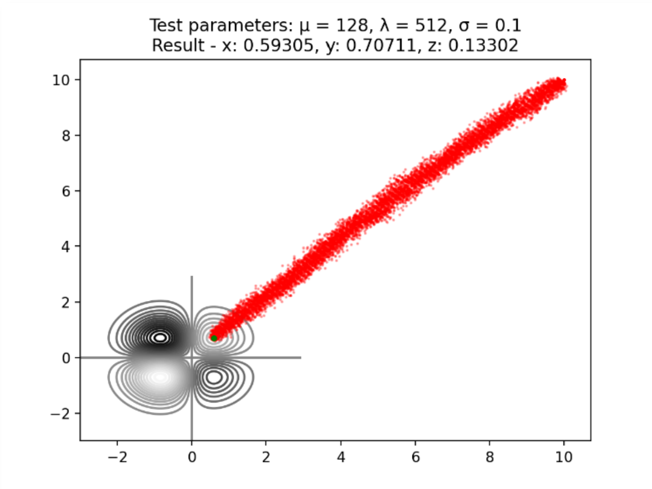
Rysunek 4.2 Wyniki dla różnych wartości rodziców μ i dzieci λ (maksimum)

W drugim eksperymencie początkowa populacja została zainicjowana w innym przedziale, co poskutkowało w znalezieniu maksimum globalnego w każdym teście, jednak z różną dokładnością. Tak jak w poprzednim eksperymencie, najlepiej poradziła sobie największa populacja z dużą ilością potomków.

5. Porównanie algorytmów ES oraz SGD

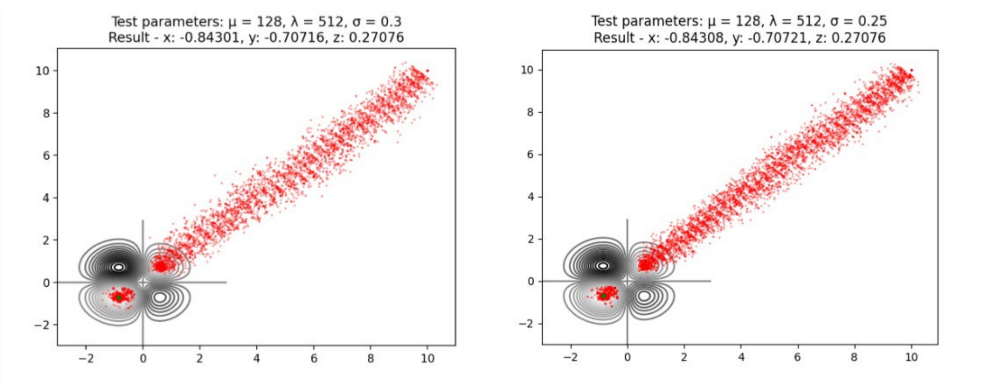
W celu porównania strategii ewolucyjnej $\mu + \lambda$ oraz algorytmu spadku wzdłuż gradientu w obu przypadkach ustawiono punkt startu na $(10, 10)$.

W przypadku algorytmu SGD nie otrzymano żadnego ekstremum globalnego, algorytm zatrzymał się w punkcie startowym. Powodem takiego rezultatu jest wartość gradientu w puncie $(10, 10)$, która była bliska zeru. Algorytm w takiej sytuacji nie mógł poruszyć się w żadną stronę.



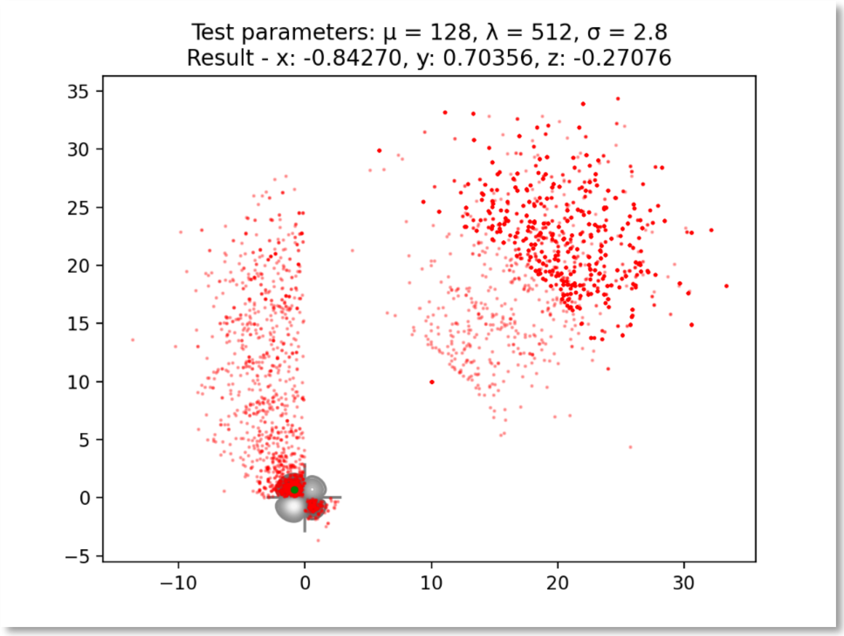
Rysunek 5.1 Wyniki algorytmu ES z ustalonymi wartościami

Algorytm ES bez problemu znalazł maksimum, startując z populacji składającej się wyłącznie z punktów $(10, 10)$. Ze względu na małe odchylenie standardowe otrzymany wynik wskazuje na maksimum lokalne. Jakiekolwiek minimum nie jest możliwe do znalezienia przez algorytm o zadanych parametrach, populacja zmierza w przeciwną stronę.



Rysunek 5.2 Wyniki poszukiwania maksimum globalnego przez algorytm ES

Do znalezienia maksimum globalnego potrzebna była zdecydowanie większa liczba iteracji oraz zwiększenie odchylenia standardowego. Po wielu testach populacja rozwijająca się przez 100 000 generacji była w stanie dotrzeć do maksimum globalnego przy $\sigma = 0.25$. Wyższe wartości σ otrzymywały zbliżone wyniki przy znacznie mniejszej ilości iteracji, np. $10\ 000$ w przypadku $\sigma = 0.3$.



Rysunek 5.2 Wyniki poszukiwania minimum globalnego przez algorytm ES

Znalezienie minimum globalnego wymagało zmiany odchylenia standardowego na wartość 2.8, w takiej sytuacji mutacje były w stanie nakierować populację na prawidłowe rozwiązanie.