



POLITECHNIKA WARSZAWSKA

WYDZIAŁ MATEMATYKI
I NAUK INFORMACYJNYCH



PRACA DYPLOMOWA MAGISTERSKA

**ANALIZA MOŻLIWOŚCI WYKORZYSTANIA
W ALGORYTMIE CMA-ES WIEDZY
O OGRANICZENIACH KOSZTOWYCH**

AUTOR:

INŻ. ROBERT JAKUBOWSKI

PROMOTOR:

DR HAB. INŻ. JAROSŁAW ARABAS
PROF. NZW. PW

WARSZAWA MAJ 2016

.....

podpis promotora

.....

podpis autora

Spis treści

1	Streszczenie	4
2	Wstęp	5
2.1	Cel pracy	5
3	Techniki uwzględniania ograniczeń	6
3.1	Transformacje rozwiązań	6
3.1.1	Powrót?	6
3.1.2	Rzutowanie	7
3.1.3	Reinicjalizacja	7
3.1.4	Odbicie	7
3.1.5	Próbkowanie	7
3.1.6	Zawijanie	8
3.2	Błądzenie przypadkowe	8
3.3	Metoda przeprowadzania testów	8
3.4	Wyniki testów	9
3.5	Test przesunięć	19
3.6	Wnioski	19
4	Zbiory benchmarkowe	20
4.1	CEC 2013	20
4.2	Porównanie wyników	20
4.3	Metody agregacji	20
4.4	Wnioski	20
5	Wpływ technik na efektywność CMA-ES	21
5.1	Algorytm CMA-ES	21
6	Podsumowanie	22
6.1	Wyniki	22
6.2	Możliwości rozwoju	22

1. Streszczenie

2. Wstęp

2.1. Cel pracy

3. Techniki uwzględniania ograniczeń

Niektóre problemy optymalizacyjne posiadają ograniczenia. Szukając rozwiązania należy zapewnić, że będzie ono dopuszczalne. Bazując na —przypis— techniki uwzględniania ograniczeń można podzielić w następujący sposób:

- definicja przestrzeni przeszukiwań - zapewnienie, że podczas krzyżowań, mutacji i innych zmian punktów, żaden z punktów nie wypadnie poza przestrzeń przeszukiwań,
- modyfikacja funkcji celu - zmienienie funkcji celu tak, aby funkcja celu dla punktów spoza ograniczeń zwracały gorsze wyniki,
- transformacja rozwiązań - punkty, które są poza ograniczeniami zostają zamieniane na punkty, które znajdują się w ograniczeniach.

W tej pracy skupiono się na transformacji rozwiązań.

3.1. Transformacje rozwiązań

Nie istnieje jedna technika transformacji rozwiązań spoza ograniczeń na dopuszczalne. W kolejnych podrozdziałach znajdują się metody transformacji rozwiązań, które były badane. Każda z technik jest opisana słownie oraz pseudokodem. Opis słowny zawiera wyjaśnienie, co się dzieje z punktem, który znalazł się poza ograniczeniem. W pseudokodzie zastosowane następujące oznaczenia:

- x - punkt, który poddajemy naprawie
- x' - rodzic punktu x , czyli z punktu x' z zadaniem rozkładem został wylosowany punkt x
- $x(i)$ - wartość i -tej współrzędnej punktu x
- lb - ograniczenie dolne
- ub - ograniczenie górne

3.1.1. Powrót?

Nowy punkt zostaje odrzucony i wraca do poprzedniej pozycji.

$$\mathbf{x} = \mathbf{x}'$$

3.1.2. Rzutowanie

Punkt jest transformowany do najbliższego punktu, który spełnia ograniczenie. Oznacza to, że dla każdej współrzędnej sprawdzany jest warunek zawierania się w ograniczeniach. Dla współrzędnych, dla których nie jest on spełniony, wartość jest zamieniana na wartość ograniczenia (dolnego lub górnego), które jest najbliższe.

```
dla każdej współrzędnej i
    jeżeli  $lb(i) > x(i)$ 
         $x(i) = lb(i)$ 
    jeżeli  $ub(i) < x(i)$ 
         $x(i) = ub(i)$ 
```

3.1.3. Reinicjalizacja

Punkt jest przenoszony do pozycji początkowej. W tej pracy był to jednocześnie środek układu współrzędnych oraz środek symetrii ograniczeń.

```
 $x = x_0$ 
```

3.1.4. Odbicie

Dla każdej współrzędnej sprawdzane są warunki na ograniczenie. W przypadku współrzędnych, na których punkt jest poza ograniczeniem, wartość punktu tej współrzędnej jest symetrycznie odbita względem ograniczenia, którego warunek został złamany.

```
dla każdej współrzędnej i
    jeżeli  $lb(i) > x(i)$ 
         $x(i) = x(i) + 2 * (lb(i) - x(i))$ 
    jeżeli  $ub(i) < x(i)$ 
         $x(i) = x(i) - 2 * (x(i) - ub(i))$ 
```

3.1.5. Próbkowanie

Punkt jest powtórnie losowany dopóty, dopóki spełnia ograniczenia kostkowe.

```
dopóki  $!w\_ograniczeniach(x)$ 
     $x = losuj(x')$ 
```

3.1.6. Zawijanie

Dla każdej współrzędnej sprawdzane są warunki na ograniczenie. W przypadku współrzędnych, na których punkt jest poza ograniczeniem, różnica, pomiędzy ograniczeniem a wartością współrzędnej punktu, jest zapamiętywana. Tę różnicę odkładamy na przeciwległym ograniczeniu po stronie, która jest wewnątrz ograniczenia. W tym miejscu znajduje się nowa wartość współrzędnej punktu. W intuicyjny sposób można to wyjaśnić tak, że dla punktów nie ma ograniczeń, a przestrzeń przeszukiwań po każdym wymiarze jest jakby "zawinięta".

```
dla każdej współrzędnej i
    jeżeli lb(i) > x(i)
        x(i) = ub(i) - (lb(i) - x(i))
    jeżeli ub(i) < x(i)
        x(i) = lb(i) + (x(i) - ub(i))
```

3.2. Błądzenie przypadkowe

Można się spodziewać, że algorytm CMA-ES dla funkcji stałej będzie zachowywał się analogicznie do błądzenia przypadkowego. Takie założenie skłoniło autorów, żeby przyjąć się błądzeniu przypadkowemu z ograniczeniami. Błądzenie przypadkowe jest algorytmem dużo prostrzym, niż CMA-ES, więc umożliwia szybsze testowanie i wyciąganie wniosków.

Niech X_1, X_2, \dots będą niezależnymi n -wymiarowymi zmiennymi losowymi o wartości oczekiwanej równej $\{0\}^n$. Błądzeniem przypadkowym nazywamy sekwencję zmiennych losowych:

$$S_0 = 0, S_i = X_1 + X_2 + \dots + X_i \quad (1)$$

3.3. Metoda przeprowadzania testów

W celu przeprowadzenia testów napisano szereg skryptów w języku MATLAB. Testy te obserwowały wpływ metod uwzględniania ograniczeń na ruch punktu. W rezultacie miały one pokazać rozkład prawdopodobieństwa punktu dla danej metody. Metody wybrane do testowania są takie jak w podrozdziale 3.1. Ponadto badano 2 różne metody losowania punktów: z rozkładem normalnym oraz jednostajnym. W przypadku

rozkładu jednostajnego losowano liczby z przedziału $[-0.5; 0.5]$ (przedział co najmniej kilkukrotnie krótszy od ograniczeń kostkowych).

Skrypty

Skrypty zostały zbudowane zgodnie z poniższym pseudokodem.

```
x - błądzący punkt
punkty - tablica wszystkich położenia punktu x
iteracje - liczba iteracji podana jako parametr
i = 0
dopóki i < iteracje
    wylosuj nowe położenie punktu x
    jeżeli x jest poza ograniczeniem
        popraw x
    dodaj x do tablicy punkty
    i = i + 1
```

3.4. Wyniki testów

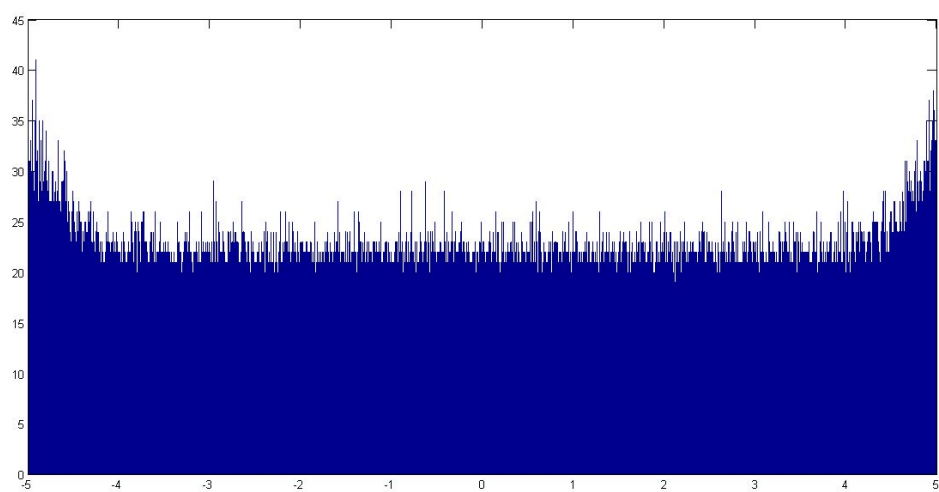
Wykresy zamieszczone w tym podrozdziale są histogramami wystąpień punktu. Symulacje skryptów były uruchamiane z liczbą iteracji wynoszącą 2-100 milionów. Szerokość przedziałów na wykresach jest różna i była dobierana tak, aby jak najlepiej przedstawić interesujące fakty.

Wszystkie wykresy, które pokazują wyniki błędzenia przypadkowego w dwóch wymiarach zostały przetworzone w następujący sposób: otrzymane wyniki symulacji powielono poprzez symetryczne odbicie ich względem osi X, osi Y oraz początku układu współrzędnych. Wykonano ten zabieg, aby zwiększyć dokładność wyników. Nie wpływają one na badania, ponieważ punkty 0 są środkami ograniczeń obu osi, a z perspektywy algorytmu nie ma znaczenia, w której ćwiartce znajduje się aktualnie punkt.

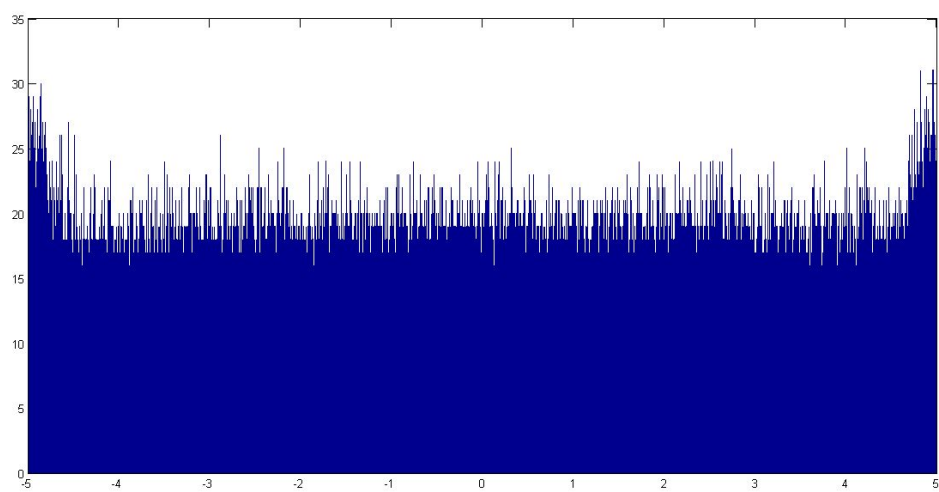
Powrót?

Histogramy pokazują, że punkt z większym prawdopodobieństwem występuje przy ograniczeniach. Takie zachowanie można uznać za intuicyjne, ponieważ potomkowie punktów blisko granicy mogą "wpaść" poza ograniczenie. Po naprawie dziecko będzie tym samym punktem. Symulacja dwuwymiarowa nie pokazuje tego tak dokładnie,

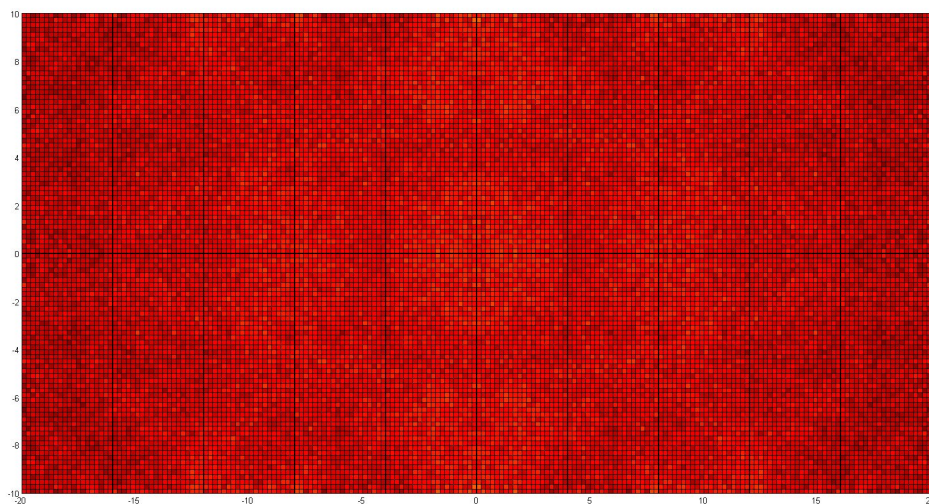
natomiast również można dostrzec ten efekt.



Rysunek 1: Rozkład normalny



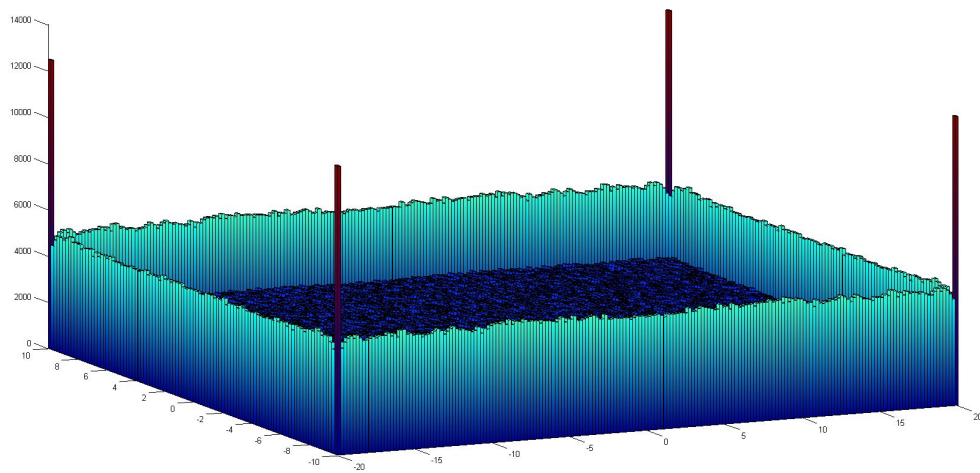
Rysunek 2: Rozkład jednostajny



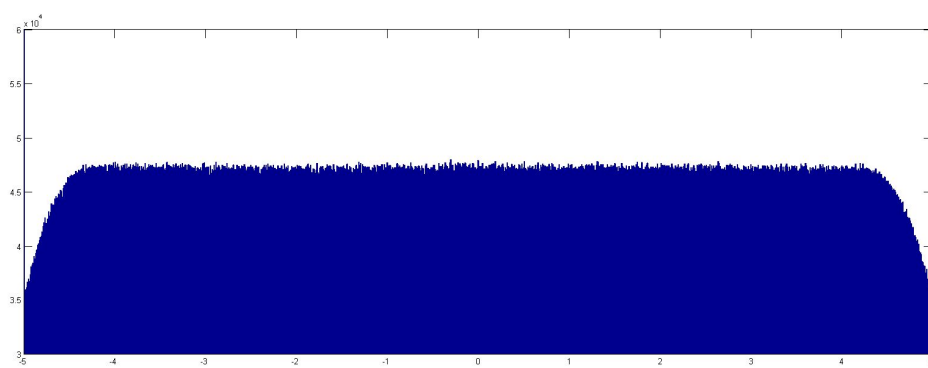
Rysunek 3: Rozkład normalny na dwóch wymiarach

Rzutowanie

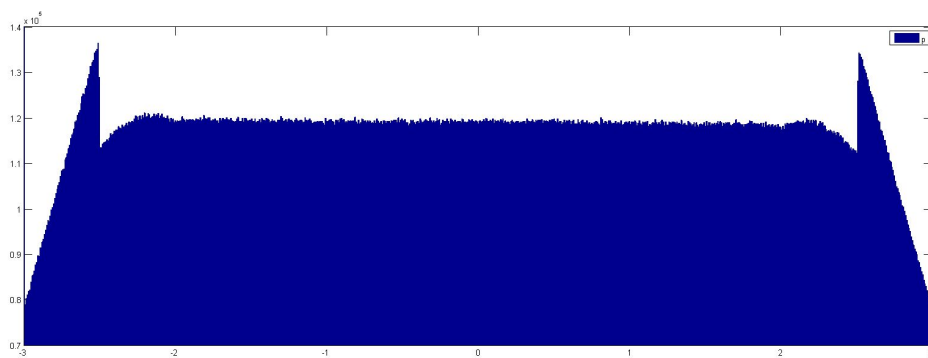
Zgodnie z przewidywaniami największe prawdopodobieństwo wystąpienia punktu jest na ograniczeniu, co bardzo dobrze obrazuje rysunek 4. Na rysunkach 5 i 6 zakres osi Y został zmniejszony do odpowiednio $[30000; 60000]$ oraz $[70000; 140000]$, aby uwypuklić interesujące efekty. W związku z tym, zostały "ucięte" wartości przedziałów brzegowych. Warto zwrócić uwagę na niespodziewane zjawisko - wartości przedziałów blisko ograniczeń są mniejsze, niż pozostałe. Oznacza to, że punkty w tych przedziałach występują z mniejszym prawdopodobieństwem, niż pozostałe. Wydawać by się mogło, że powinno być inaczej, ponieważ podczas symulacji punkty często są rzutowane na ograniczenie. Dodatkowo, w rozkładzie jednostajnym widać, że istnieją przedziały, w których punkty występują z większym prawdopodobieństwem. Autorom nie udało się formalnie uzasadnić tego zjawiska. Intuicja prowadzi do hipotezy, że punkty, które znalazły się poza ograniczeniem, bez naprawy po kilku iteracjach zapełniałyby "dołek". Naprawa natomiast sprawia, że cała dalsza symulacja zostaje niejako przesunięta. W ten sposób powstaje "górką" w rozkładzie jednostajnym. Z charakterystyki rozkładu normalnego może brać się fakt, iż owa górką w nim nie występuje.



Rysunek 4: Rozkład normalny na dwóch wymiarach



Rysunek 5: Rozkład normalny



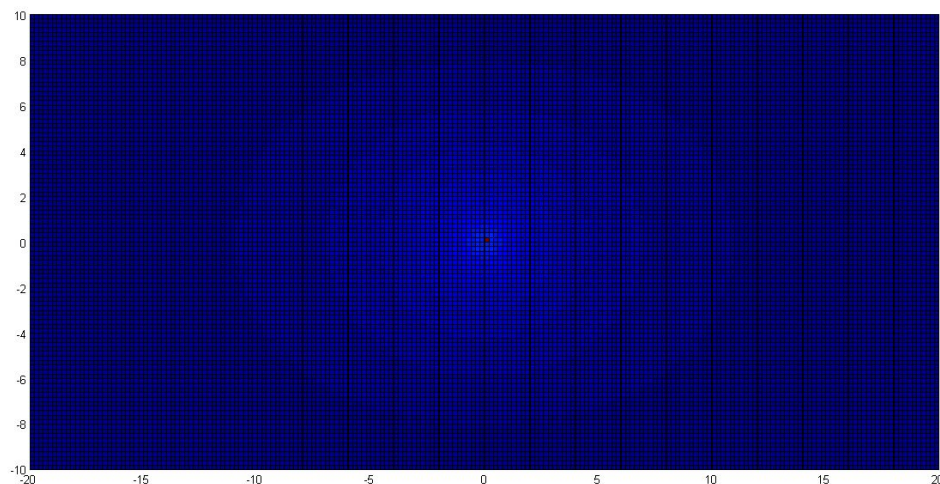
Rysunek 6: Rozkład jednostajny

Reinicjalizacja

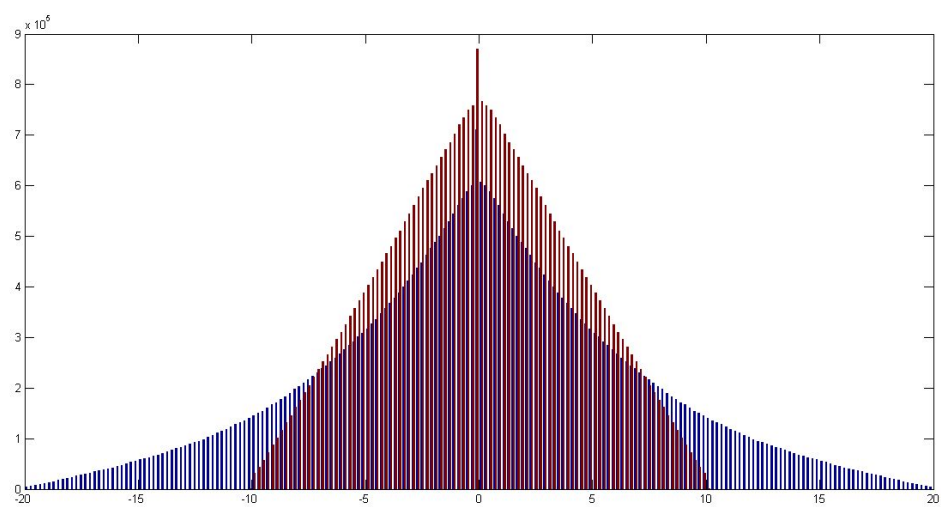
Na pierwszy rzut oka wykresy nie ukazują nic nadzwyczajnego, ponieważ łatwo zauważyć pik związany z reinicjalizacją oraz wartości histogramu malejące wraz ze zbliżaniem się do ograniczeń. Autorzy pragną zwrócić uwagę na 2 fakty.

Pierwsze spostrzeżenie, to kształt histogramu. Zarówno dla rozkładu jednostajnego, jak i normalnego w jednym wymiarze jest on stożkowy. Łuk można zauważyć tylko blisko punktu środkowego, w pozostałej części spadek jest liniowy. Brakuje charakterystycznego, gaussowskiego przegięcia. Sytuacja jest ciekawsza, gdy występuje więcej wymiarów. Widać wówczas przegięcie. Dokładniejsze badania pokazały, że przegięcie nie występuje tylko na jednym wymiarze. Tym, który jest relatywnie najkrótszy. Celowo jest użyte słowo relatywnie, ponieważ z perspektywy błędzenia przypadkowego i rozkładu normalnego trzeba brać pod uwagę parametr σ . Wymiary o małym σ będą relatywnie dłuższe od tych z dużym σ , ponieważ błędzenie będzie wykonywało mniejsze kroki.

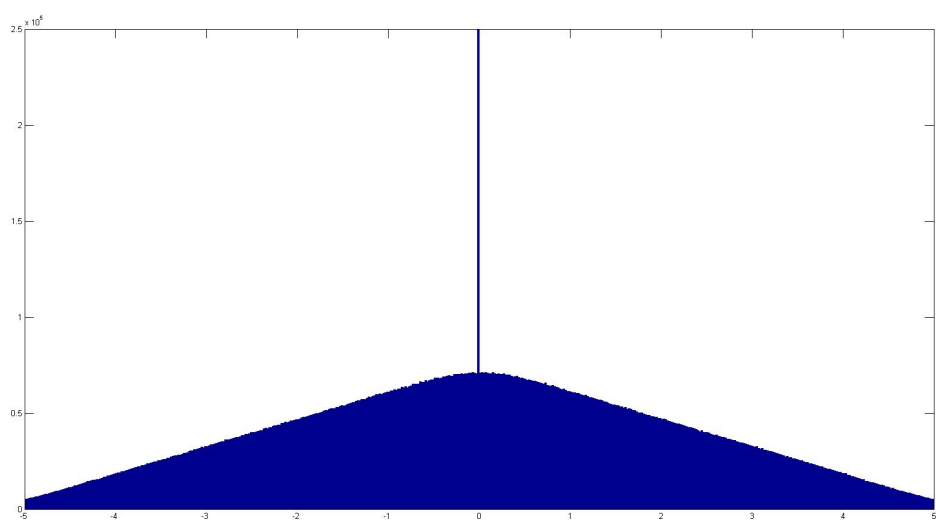
Na rysunku 13 można dostrzec też dwa uskoki, które związane są z pikiem w punkcie 0 oraz charakterystyką rozkładu.



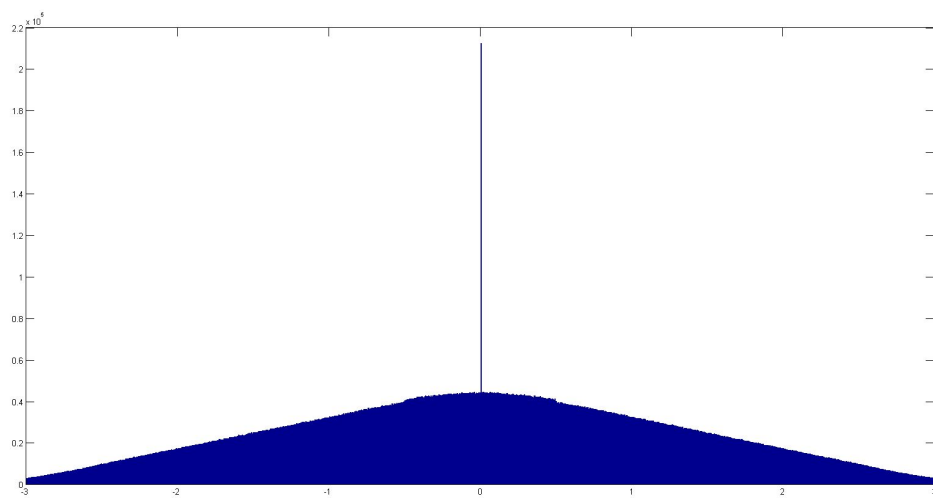
Rysunek 7: Rozkład normalny na dwóch wymiarach



Rysunek 8: Rozkład normalny na dwóch wymiarach; oddzielne histogramy dla obu wymiarów



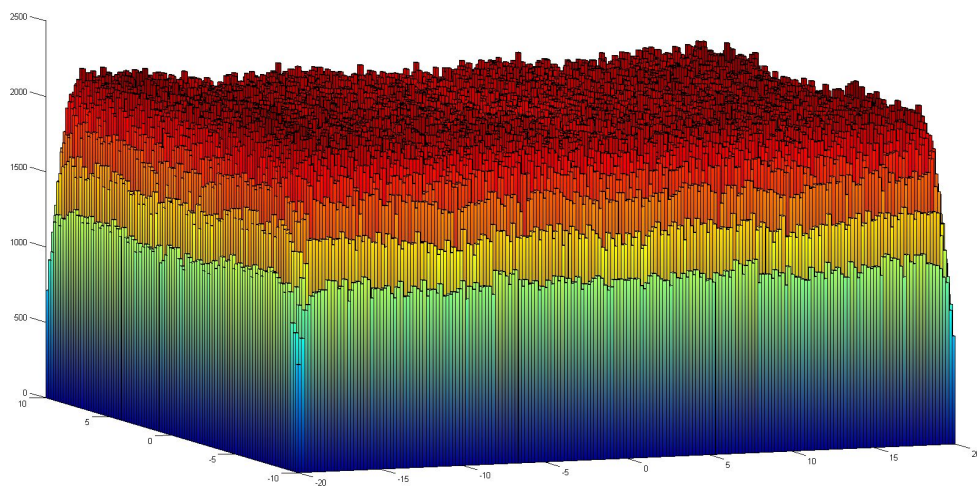
Rysunek 9: Rozkład normalny



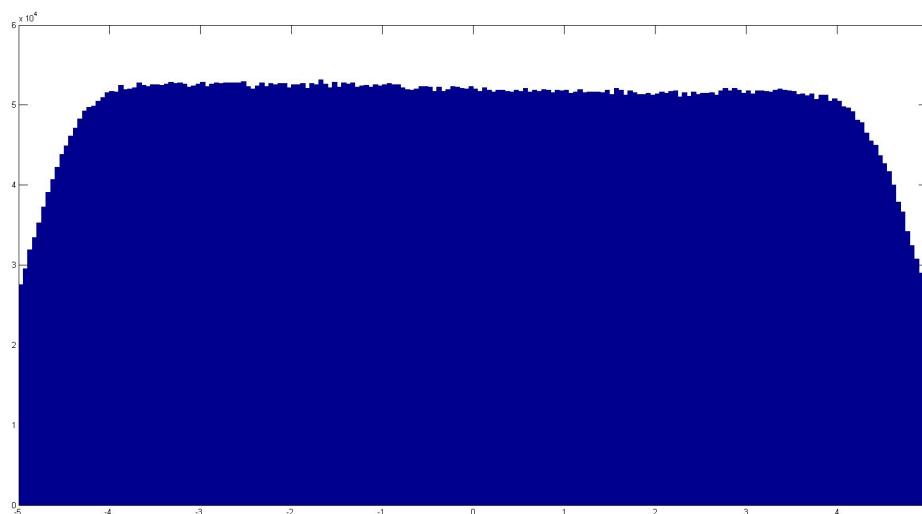
Rysunek 10: Rozkład jednostajny

Próbkowanie

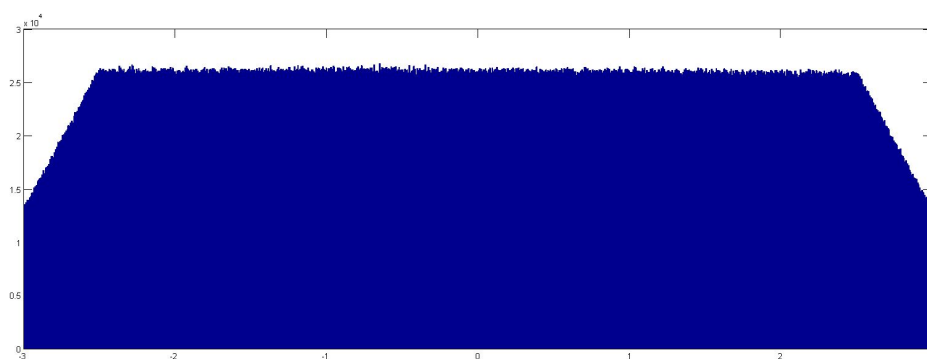
Ten rozkład charakteryzuje się spadkiem wartości prawdopodobieństwa wraz ze zbliżaniem się do ograniczenia. Punkty, które wypadłyby poza ograniczenia oraz ich potomkowie są przesuwane w kierunku środka przedziału.



Rysunek 11: Rozkład normalny na dwóch wymiarach



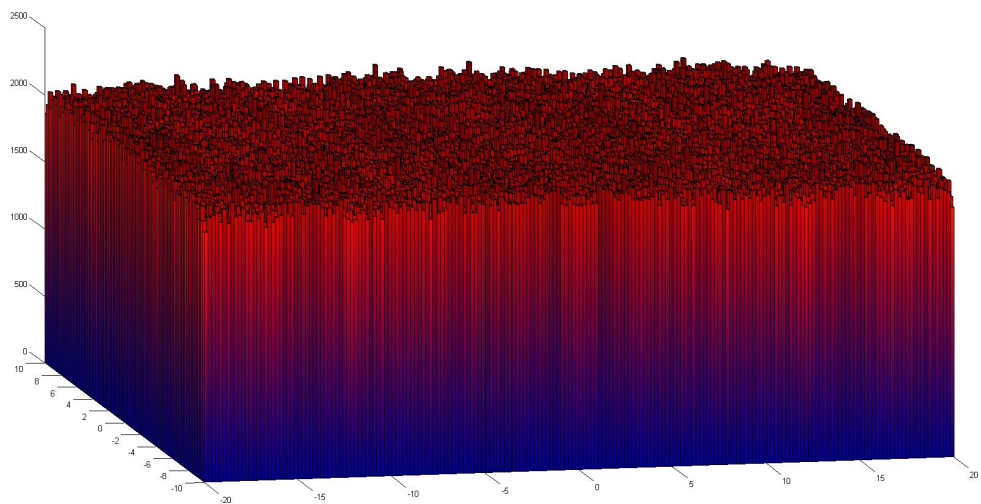
Rysunek 12: Rozkład normalny



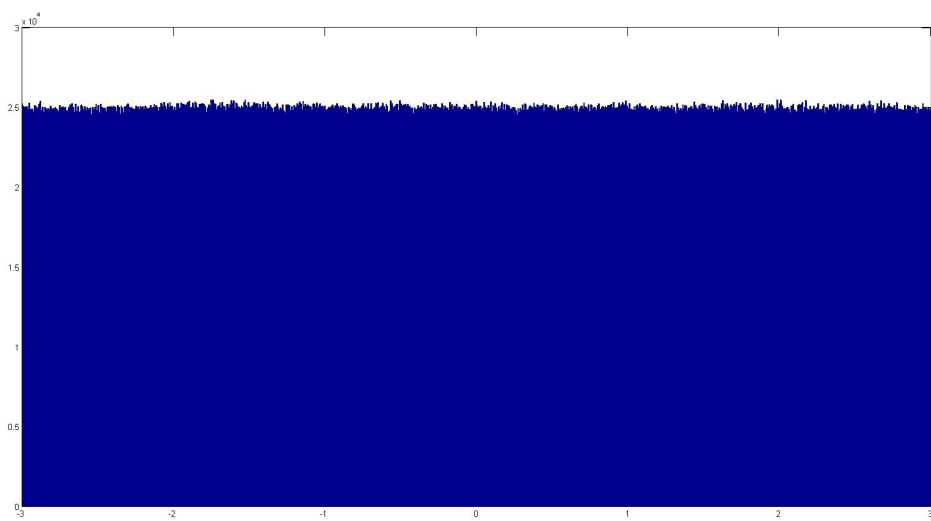
Rysunek 13: Rozkład jednostajny

Zawijanie

Wyobraźmy sobie, że naszą symulację przeprowadzamy bez ograniczeń, następnie wykres z wynikami tniemy w równych odstępach wzdłuż każdej osi. Na koniec tak pocięte części łączymy w jeden wykres. Takie zachowanie symuluje zawijanie. W związku z tym nie dziwi fakt, że poniższe histogramy przedstawiają rozkład jednostajny z szumem.



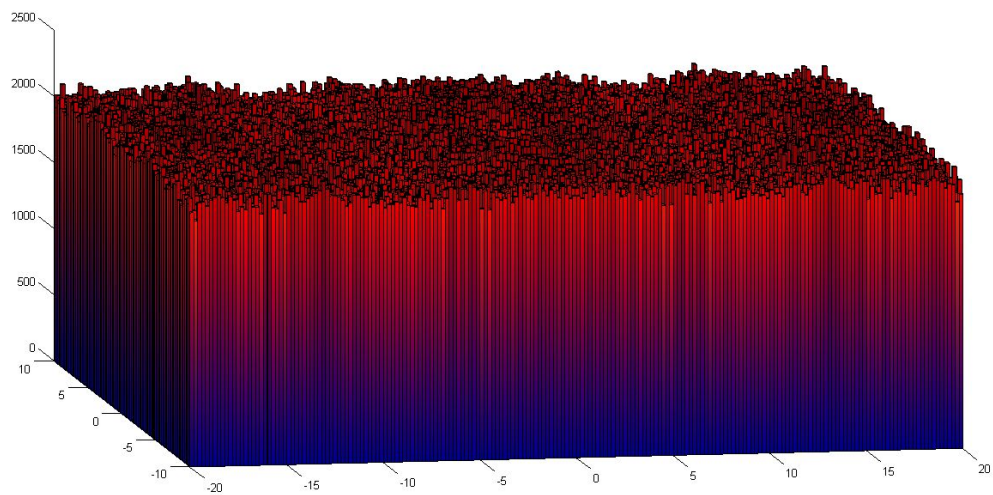
Rysunek 14: Rozkład normalny na dwóch wymiarach



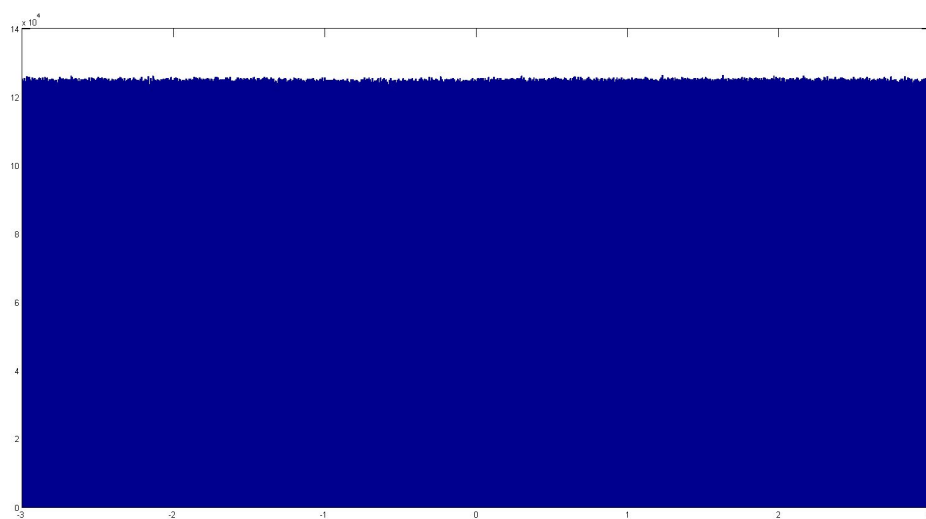
Rysunek 15: Rozkład jednostajny

Odbicie

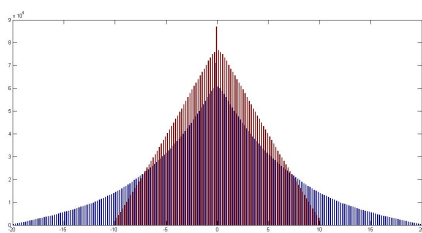
Podobnie jak zawijanie, odbicie zwraca histogram rozkładu normalnego z szumem. W tej sytuacji nieco trudniej o analogię, lecz po chwili zastanowienia nie dziwi kształt poniższych wykresów.



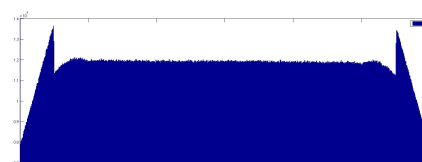
Rysunek 16: Rozkład normalny na dwóch wymiarach



Rysunek 17: Rozkład jednostajny



(a) test 1

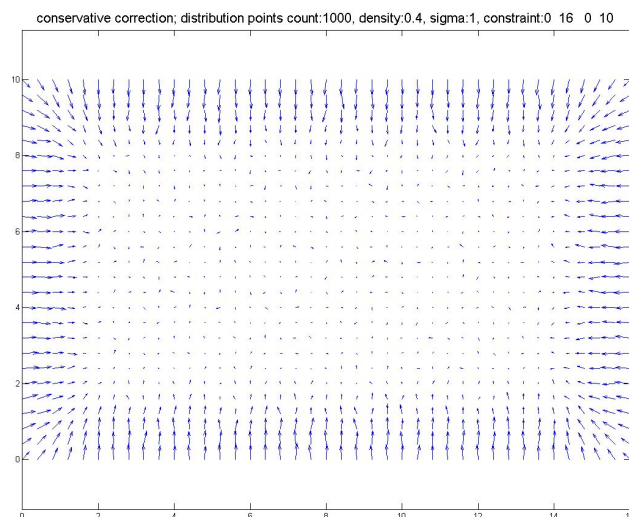


(b) test 2

Rysunek 18: TEST Obrazy obok siebie

3.5. Test przesunięć

Dodatkowo został wykonany test, w którym badano wartość oczekiwaną przesunięcia punktu w kolejnej iteracji (razem z naprawą). Dla każdej z metod wyglądał on podobnie:



Rysunek 19: Test przesunięć - powrót?

Jedynie dla zawijania oraz reinicjalizacji długości wektorów były znacząco większe. W pozostałych zmieniał się kierunek oraz długość wektora. Były to jednak zmiany wynikające pośrednio z pokazanych wcześniej histogramów.

3.6. Wnioski

Każda z technik ma swoje charakterystyczne cechy.

4. Zbiory benchmarkowe

4.1. CEC 2013

- wybrane kilka funkcji
- kilkukrotne powtórzenie testów
- porównanie dokładności wyników i liczby iteracji

4.2. Porównanie wyników

Zgodnie z założeniami poczynionymi w rozdziale 3.2 testy algorytmu CMA-ES powinny przynieść rezultaty zbliżone do testów błędzenia przypadkowego. Do przeprowadzania testów została użyta biblioteka przygotowana przez Nikolausa Hansena, współautora algorytmu CMA-ES. Podobnie, jak w przypadku błędzenia przypadkowego, wykorzystano implementację w języku MATLAB —przypis—.

4.3. Metody agregacji

4.4. Wnioski

5. Wpływ technik na efektywność CMA-ES

5.1. Algorytm CMA-ES

Klasyczne algorytmy ewolucyjne nie dostosowują się do charakterystyki optymalizowanej funkcji. W większości z nich rozkład prawdopodobieństwa losowanych punktów jest stały. Z tego faktu wynika problem doboru parametrów przeszukiwania. Na przeciw tym problemom wychodzi algorytm CMA-ES, który w swej idei ma dopasowywać się do badanej funkcji.

Rozwinięcie akronimu CME-ES podpowiada, w jaki sposób jest to realizowane: Covariance Matrix Adaptation - Evolution Strategy (adaptacja macierzy kowariancji - strategia ewolucyjna). Punkty losowane są na podstawie macierzy kowariancji, która jest w każdej iteracji dostosowywana do aktualnej sytuacji przeszukiwań.

Szczegóły

Rola dotychczas wygenerowanych punktów

6. Podsumowanie

6.1. Wyniki

6.2. Możliwości rozwoju

Literatura

Warszawa, dnia 19 maja 2016

Oświadczenie

Oświadczam, że pracę magisterską pod tytułem „Analiza możliwości wykorzystania w algorytmie CMA-ES wiedzy o ograniczeniach kosztowych”, której promotorem jest dr hab. inż. Jarosław Arabas prof. nzw. PW, wykonałem samodzielnie, co poświadczam własnoręcznym podpisem.

.....