

Optymalizacja globalna: sposoby wykorzystania próbek do budowy modelu funkcji celu

Michał Okulewicz

Wydział Matematyki i Nauk Informatycznych
Politechnika Warszawska

2020-10-14

Plan prezentacji I

- 1 Wstęp
 - Zadanie optymalizacji globalnej
 - Motywacja
- 2 Sposób weryfikacji
 - Funkcja testowa
- 3 Metody
- 4 Wyniki
 - Globalny model: zależność od sposobu próbkowania
 - Lokalny model: zależność od wyników próbkowania
 - Zbieżność
- 5 Wnioski

Zadanie optymalizacji globalnej

Mając:

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

Poszukujemy (wykonując jak najmniej obliczeń wartości funkcji):

$$X^* = \{x^* : \forall x \in \mathbb{R}^n f(x^*) \leq f(x)\}$$

W praktyce dziedzinę f upraszczamy do pewnej kostki:

$$a_i, b_i \in \mathbb{R}$$

$$f : [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n] \rightarrow \mathbb{R}$$

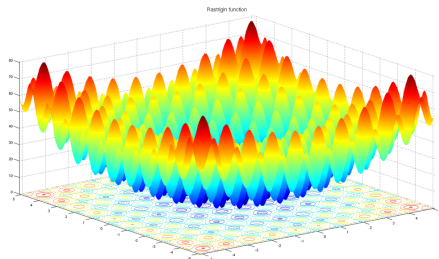
Motywacja

- Problem
 - W ogólnym przypadku znalezienie optimum globalnego wymagałoby sprawdzenia wszystkich przypadków
- Na szczęście większość rozważanych funkcji posiada jednak pewną strukturę
- **Cel:** Przyspieszenie zbieżności algorytmu poprzez:
 - Szybszą lokalizację potencjalnego optimum globalnego
 - Szybszą lokalizację najbliższego optimum lokalnego

Funkcja testowa

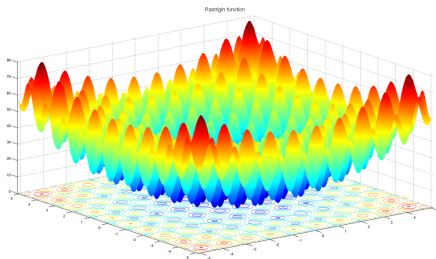
- Weźmy funkcję, której strukturę da się wykorzystać:
 - Funkcja Rastrigina

$$f_{Rast}(x) = n + \sum_{i=1}^n (x_i^2 - \cos(2\pi x_i))$$



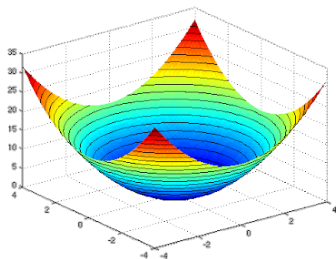
Ilu obliczeń wartości funkcji potrzeba?

$$f_{Rast}(x) = n + \sum_{i=1}^n (x_i^2 - \cos(2\pi x_i))$$



A tutaj?

$$f_{para}(x) = \sum_{i=1}^n x_i^2$$



Teoretycznie $2n+2$ - zbudowanie modelu n -wymiarowej funkcji kwadratowej i próbkowanie jej wierzchołka

Jak zatem podejść do funkcji Rastrigina?

Koncepcja:

- Jeden model powinien uzyskać zgrubną lokalizację optimum globalnego (szумы)
- Drugi model powinien uzyskać dokładną lokalizację optimum lokalnego

Środowisko testowe

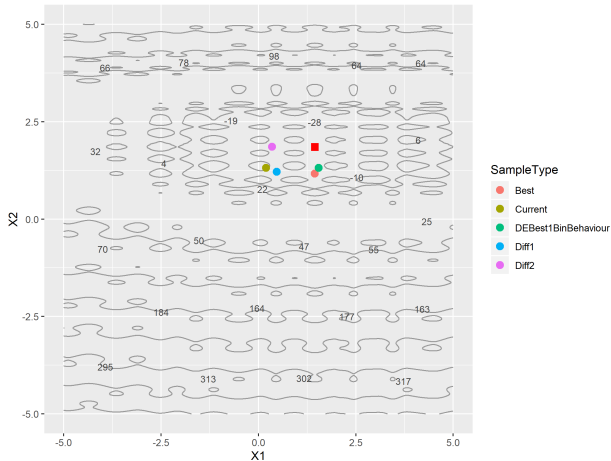
Algorytm:

- Próbki gromadzone przez hybrydowe podejście DE+PSO (populacyjny losowy algorytm optymalizacji ciągłej)

Funkcje:

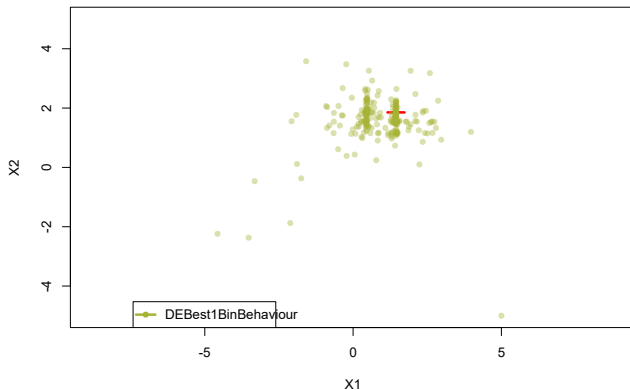
- 15 instancji testowych funkcji Rastrigina ze zbioru COCO BBOB

Kontekst: próbkowanie DE

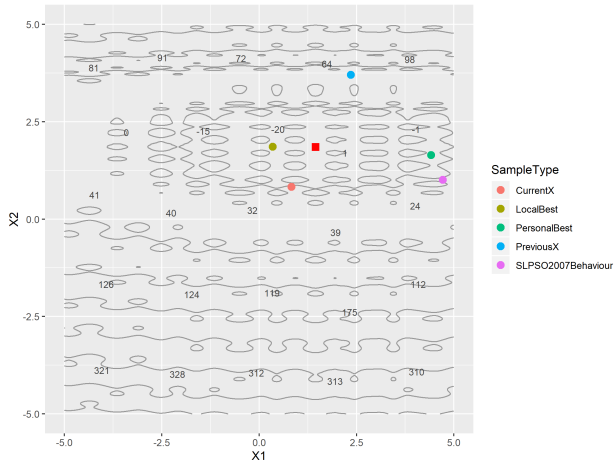


Kontekst: próbkowanie DE

PDLGBGBGGL_1000_1000_1_1_1_1_1_1_1_1_1-f003-i74-d02

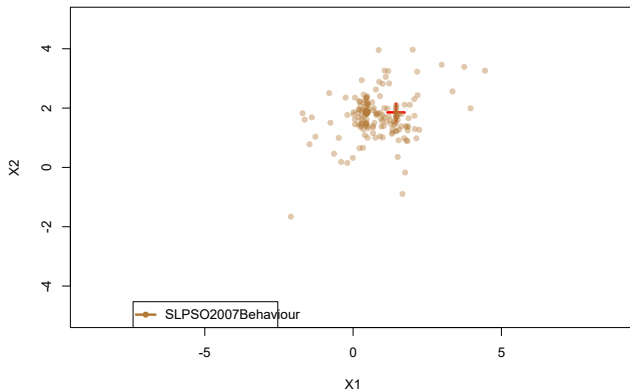


Kontekst: próbkowanie PSO



Kontekst: próbkowanie PSO

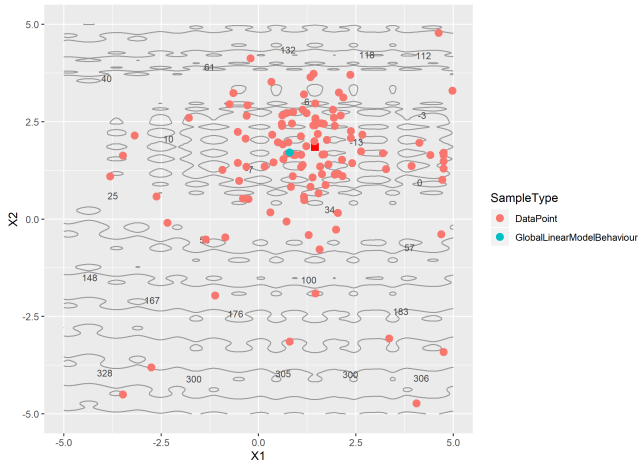
PDLGBGBGGL_1000_1000_1_1_1_1_1_1_1_1_1-f003-i74-d02



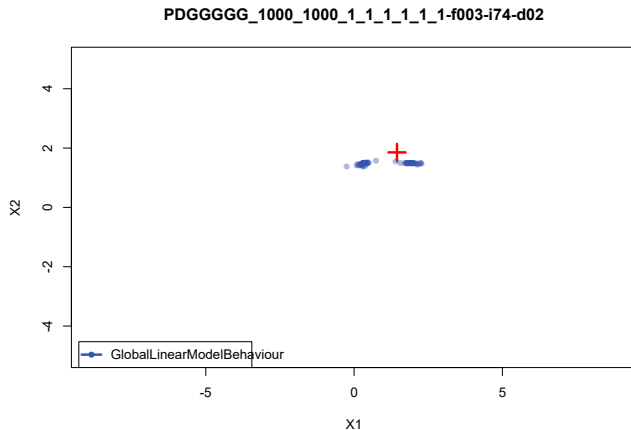
Próbki do modelu

- Model globalny
 - k próbek rozrzuconych w odległych liściach drzewa indeksu wszystkich próbek w danym przebiegu
 - próbki z dedykowanego archiwum blokującego zapisywanie bliskich sobie próbek
 - $100n$ losowych próbek spośród wszystkich próbek
 - próbki z usuniętymi obserwacjami odstającymi spośród $100n$ losowych próbek
- Model lokalny
 - najbliższych k -sąsiadów do p_{best} danej cząstki
 - największy klaster najbliższych k -sąsiadów do p_{best} danej cząstki
 - p_{best} wszystkich cząstek
 - największy klaster p_{best} wszystkich cząstek

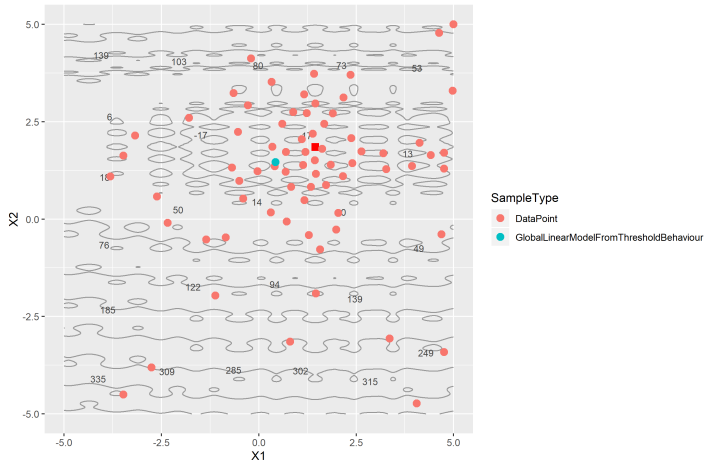
Archiwum standardowe (globalne R-Drzewo)



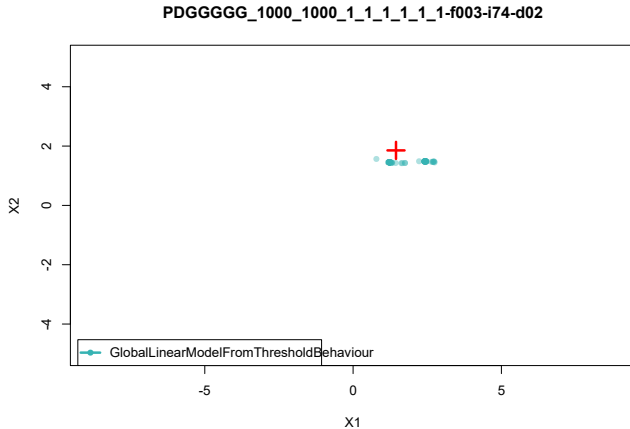
Archiwum standardowe (globalne R-Drzewo)



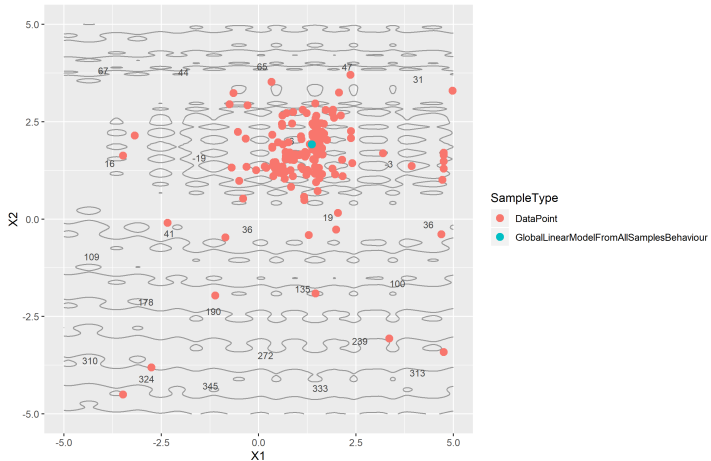
Archiwum rozproszone



Archiwum rozproszone

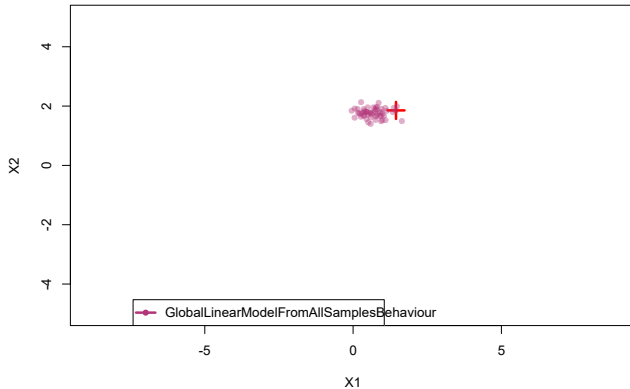


Archiwum całościowe

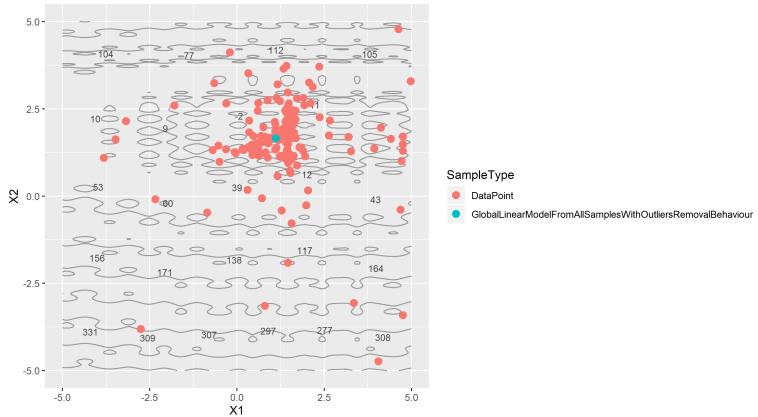


Archiwum całościowe

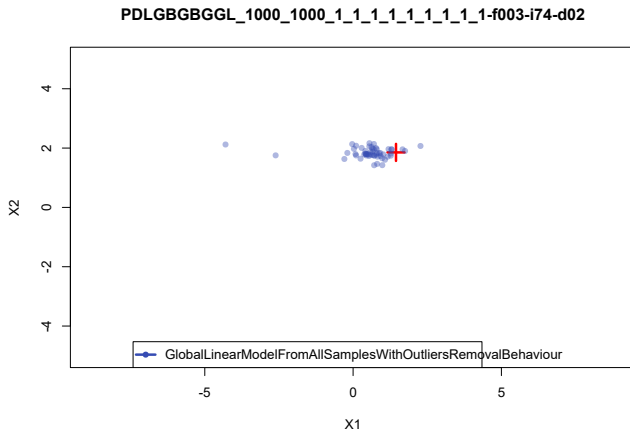
PDLGBGBGGL_1000_1000_1_1_1_1_1_1_1_1_1-f003-i74-d02



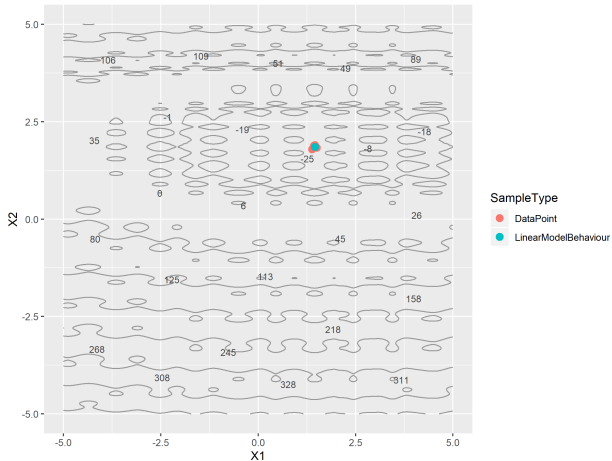
Archiwum całościowe z usuwaniem obserwacji odstających



Archiwum całościowe z usuwaniem obserwacji odstających

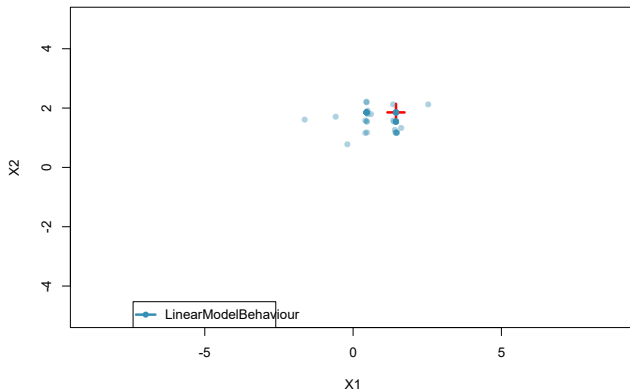


k-sasiadow

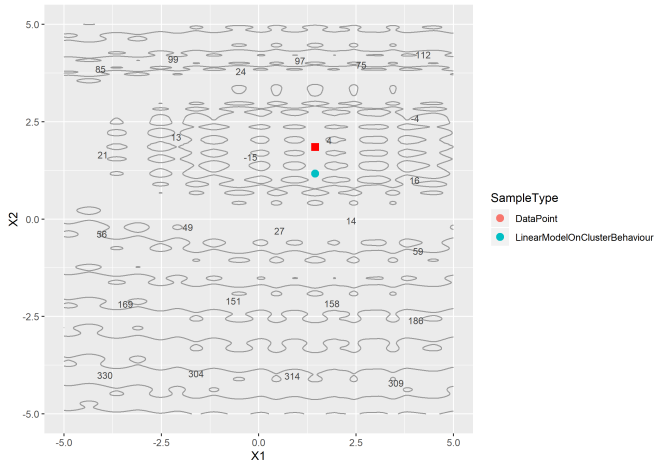


k-sasiadów

PDLGBGBGGL_1000_1000_1_1_1_1_1_1_1_1_1-f003-i74-d02

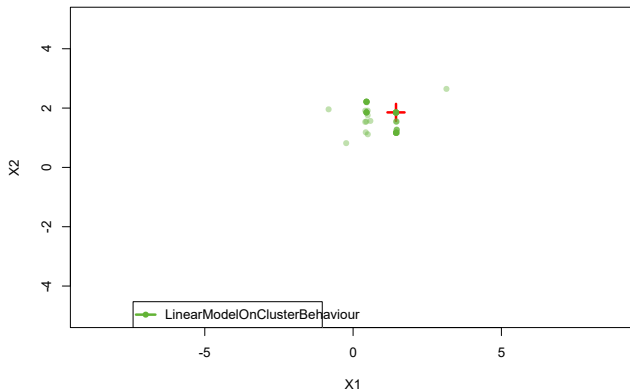


Klaster k-sasiadów

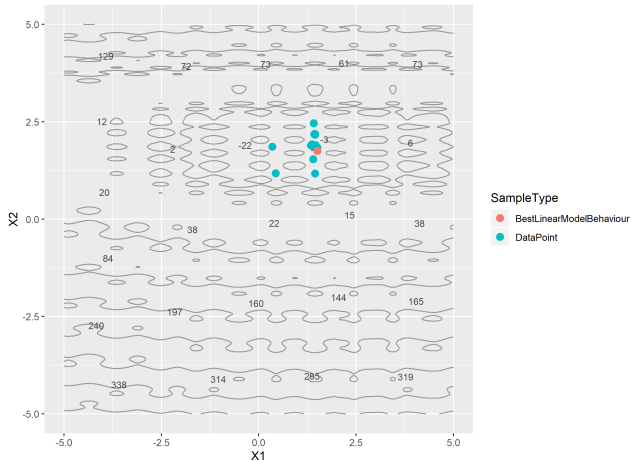


Klaster k-sasiadów

PDLGBGBGGL_1000_1000_1_1_1_1_1_1_1_1_1_1-f003-i74-d02

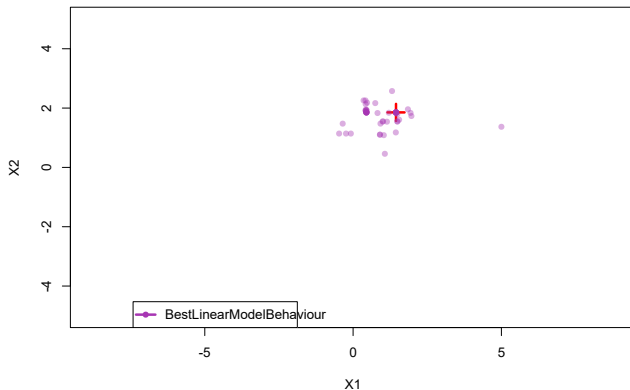


pBest

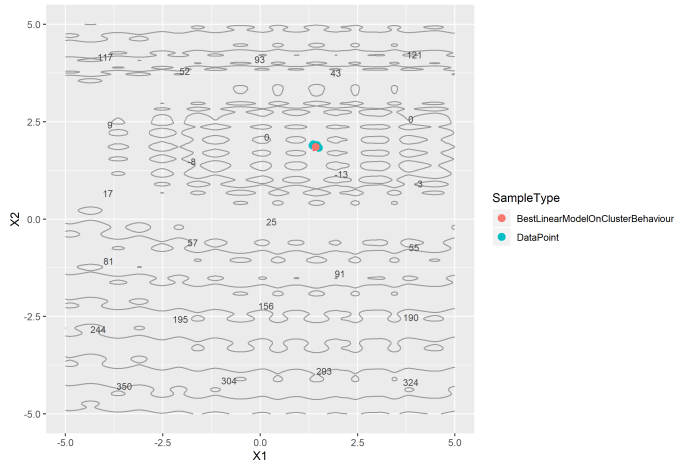


pBest

PDLGBGBGGL_1000_1000_1_1_1_1_1_1_1_1_1-f003-i74-d02

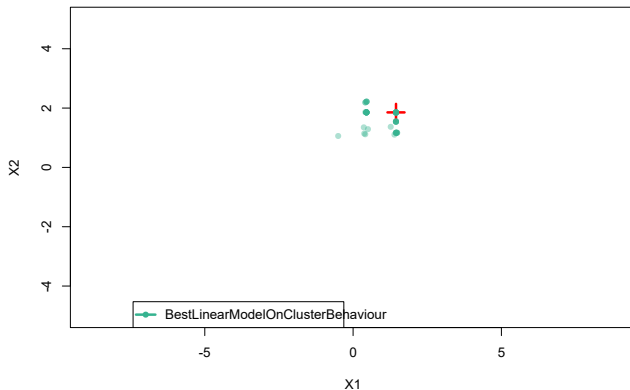


Klaster pBest

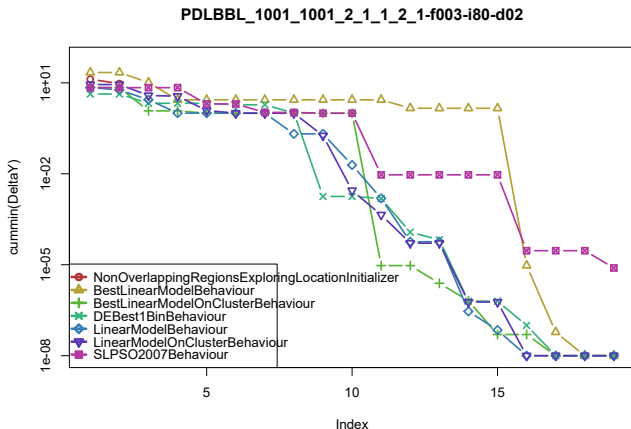


Klaster pBest

PDLGBGBGGL_1000_1000_1_1_1_1_1_1_1_1_1-f003-i74-d02



Przykładowy przebieg



Szybkość zbieżności w iteracjach (limit 1000)

Konfiguracja	Średnia liczba iteracji
PD-1000-1000-1	56
PDL-1000-1000-6-1	35
PDLBBL-1000-1000-2-1-1-2-1	32
PDLBBGL-1000-1000-2-1-1-1-2-1	32
PDLBBLg-1000-1000-2-1-1-2-1	28

Table: Funkcja Rastrigina 2D (15 instancji)

Szybkość zbieżności w iteracjach (limit 1000)

Konfiguracja	Średnia liczba iteracji
PD-1000-1000-1	227
PDL-1000-1000-6-1	87
PDLBBL-1000-1000-2-1-1-2-1	83
PDLBBGL-1000-1000-2-1-1-1-2-1	79
PDLBBLg-1000-1000-2-1-1-2-1	74

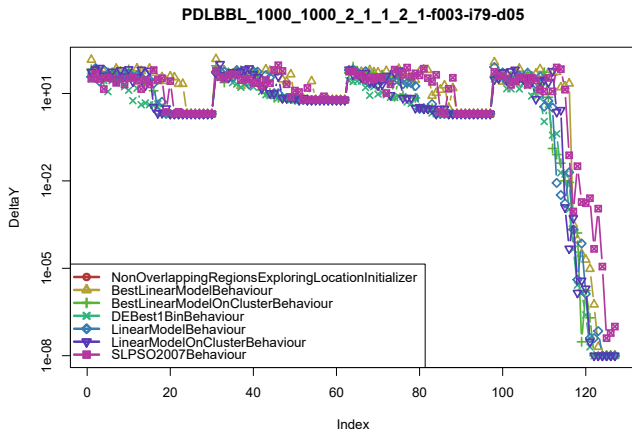
Table: Funkcja Rastrigina 3D (15 instancji)

Szybkość zbieżności w iteracjach (limit 1000)

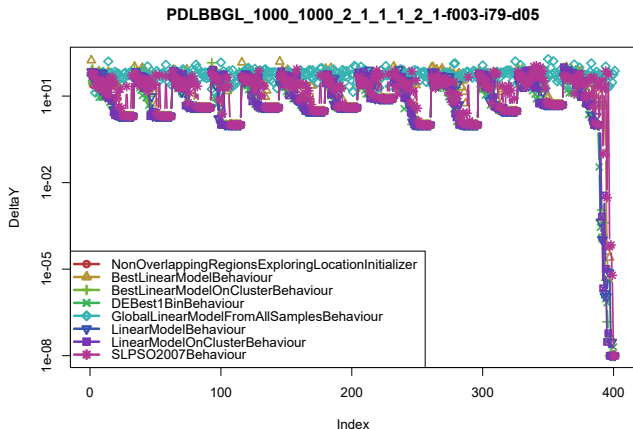
Konfiguracja	Średnia liczba iteracji
PD-1000-1000-1	961
PDL-1000-1000-6-1	673
PDLBBL-1000-1000-2-1-1-2-1	635
PDLBBGL-1000-1000-2-1-1-1-2-1	371
PDLBBLg-1000-1000-2-1-1-2-1	285

Table: Funkcja Rastrigina 5D (15 instancji)

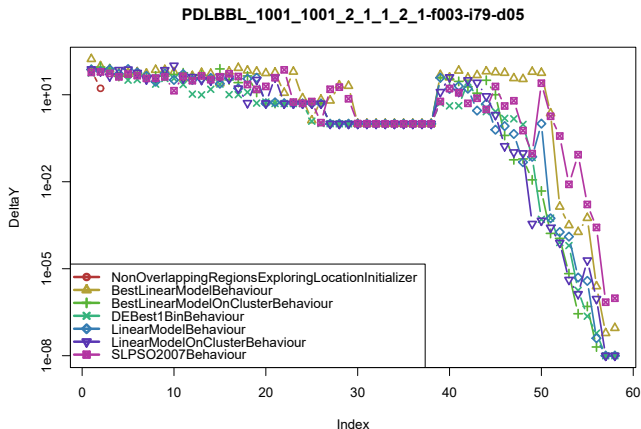
Dodatkowe obserwacje PDLBBL na 5D



Dodatkowe obserwacje PDLBBGL na 5D



Dodatkowe obserwacje PDLBBLg na 5D



Wnioski

- Do globalnego modelu najlepiej sprawdziły się losowe próbki spośród nieuporządkowanej listy (ew. z usunięciem outlierów na podstawie odległości Cooka)
- Globalny model najlepiej działa jako inicjalizator roju (zapewne to zachowanie na funkcjach bez globalnej struktury będzie niepożądane)
- Lokalne modele (p_{best} i k -sąsiadów) wydają się działać lepiej na klastrze

Dalsze pytania

- Jak lokalne modele przeniosą się na inne funkcje?
- Czy globalny model nie wywoła efektu przeuczenia przy kolejnych restartach?
- Jak poradzić sobie z zależnościami między zmiennymi (obrotu)?
- Jak poradzić sobie z asymetrycznym optimum?

Jak poradzić sobie z funkcjami o globalnej charakterystyce innej niż parabola (np. funkcja Ackleya)?

