

POSI – Laboratoria C2

Badanie metod skalowania

Opis zadania

Celem laboratorium jest zbadanie wpływu różnych metod skalowania liniowego na wartości funkcji ewaluacji w algorytmie genetycznym. W tym celu badane zostaną po dwa zestawy danych dla dwóch różnych funkcji.

Interaktywny zeszyt Jupyter Notebook z obliczeniami i analizą danych wykonanymi na potrzeby zadania jest dostępny pod adresem <https://github.com/SiwyKrzysiek/POSI-C2>

Badanie dla funkcji De Jonga

Funkcja De Jonga nazywana również paraboloidą jest zdefiniowana poniższym równaniem.

$$f(x) = \sum_{i=1}^3 x_i^2, \quad -5,12 \leq x_i \leq 5,12$$

Przy pomocy dwóch zbiorów danych testowych została zbadana metoda skalowania liniowego oraz skalowania σ -odcinającego.

Mały zbiór danych

Pierwszy badany zbiór składał się z 20 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.

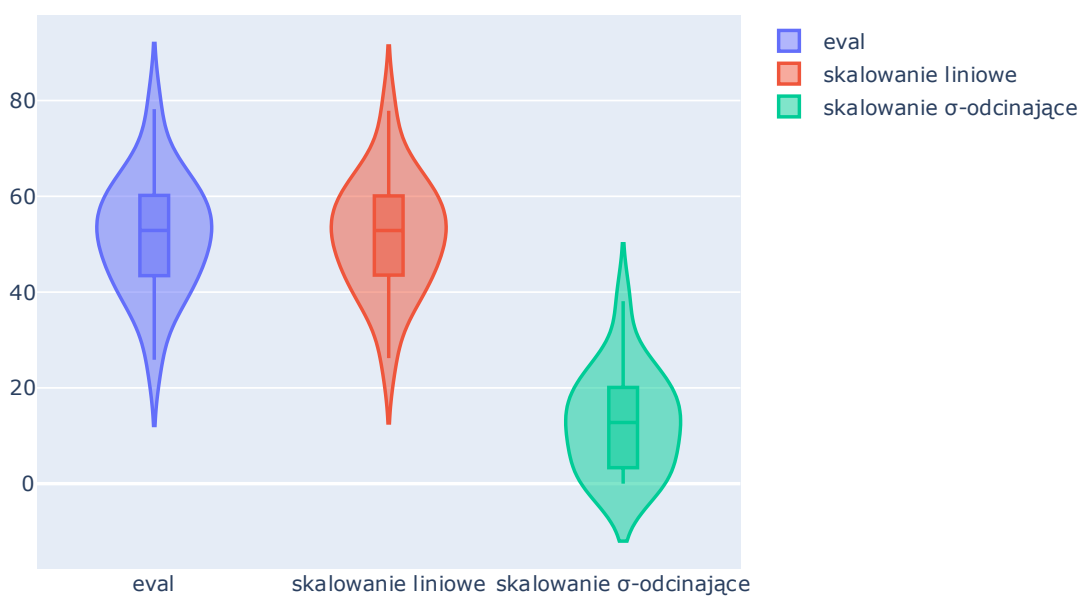
	x	y	z	f(x, y, z)	eval	skalowanie liniowe	skalowanie σ -odcinające
0	-0.760613	3.221338	2.410468	16.765904	63.234096	63.083761	23.133462
1	3.768353	-1.194181	4.909636	39.731075	40.268925	40.423250	0.168291
2	4.026310	-2.972517	2.476315	31.179165	48.820835	48.861709	8.720201
3	1.670588	3.960847	3.666050	31.919095	48.080905	48.131595	7.980271
4	2.552445	3.790282	-3.207620	31.170040	48.829960	48.870713	8.729326
5	-1.786197	-1.301121	3.007621	13.929200	66.070800	65.882832	25.970166
6	-3.573143	-3.379787	-4.288828	42.584358	37.415642	37.607820	0.000000
7	-1.995005	2.900888	-3.451841	24.310396	55.689604	55.639356	15.588971
8	-4.396633	2.058969	-3.266766	34.241496	45.758504	45.840004	5.657871
9	1.012913	-0.867700	0.138980	1.798211	78.201789	77.852890	38.101155
10	-2.860481	2.309868	3.578214	26.321453	53.678547	53.654978	13.577913
11	4.391863	2.416026	-0.239747	25.183117	54.816883	54.778213	14.716250
12	-0.072776	0.967888	-4.341381	19.789696	60.210304	60.100084	20.109671
13	-3.917036	4.776608	0.853220	38.887147	41.112853	41.255984	1.012220
14	-4.173600	-4.980434	3.450164	54.127289	25.872711	26.218020	0.000000
15	4.247428	2.094963	-1.157052	23.768287	56.231713	56.174273	16.131079
16	2.106306	4.334706	-2.169314	27.932126	52.067874	52.065672	11.967240
17	1.884363	0.207795	3.864226	18.526243	61.473757	61.346776	21.373124
18	4.340627	-0.110283	0.985231	19.823881	60.176119	60.066352	20.075485
19	-2.413820	5.035120	-2.965525	39.973293	40.026707	40.184246	0.000000

Rysunek 1 - wartości małej grupy dla paraboloidy

	eval	skalowanie liniowe	skalowanie σ -odcinające
std	11.801293	11.644734	10.313839
min	25.872711	26.218020	0.000000
25%	44.597092	44.693999	4.496458
50%	52.873210	52.860325	12.772577
75%	60.184665	60.074785	20.084031
max	78.201789	77.852890	38.101155

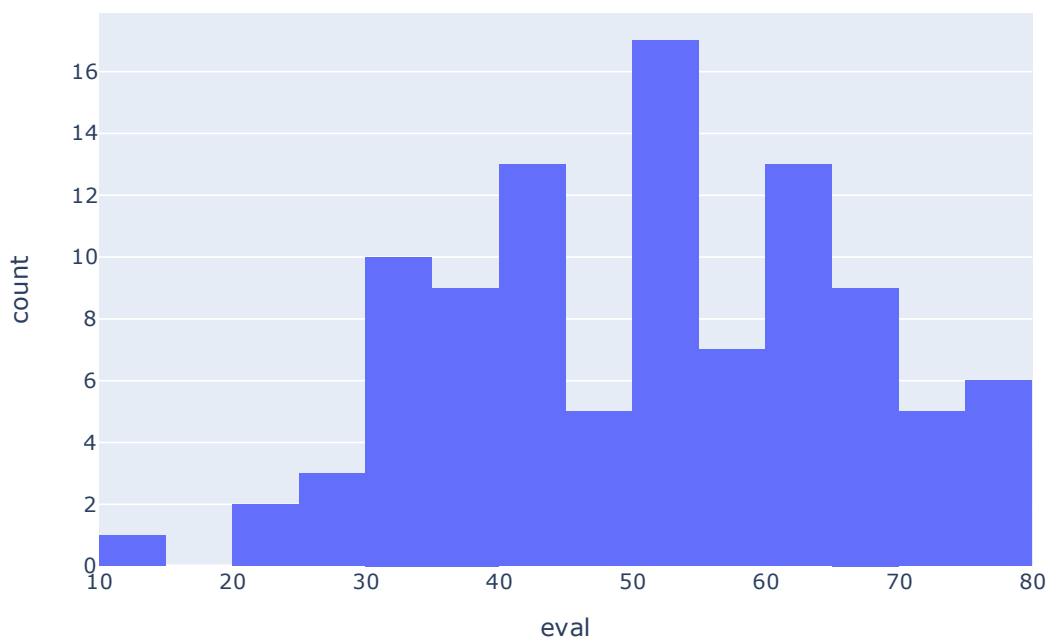
Rysunek 2 - Parametry statystyczne skalowania małej grupy dla paraboloidy

Wykres skrzypcowy skalowania małej grupy



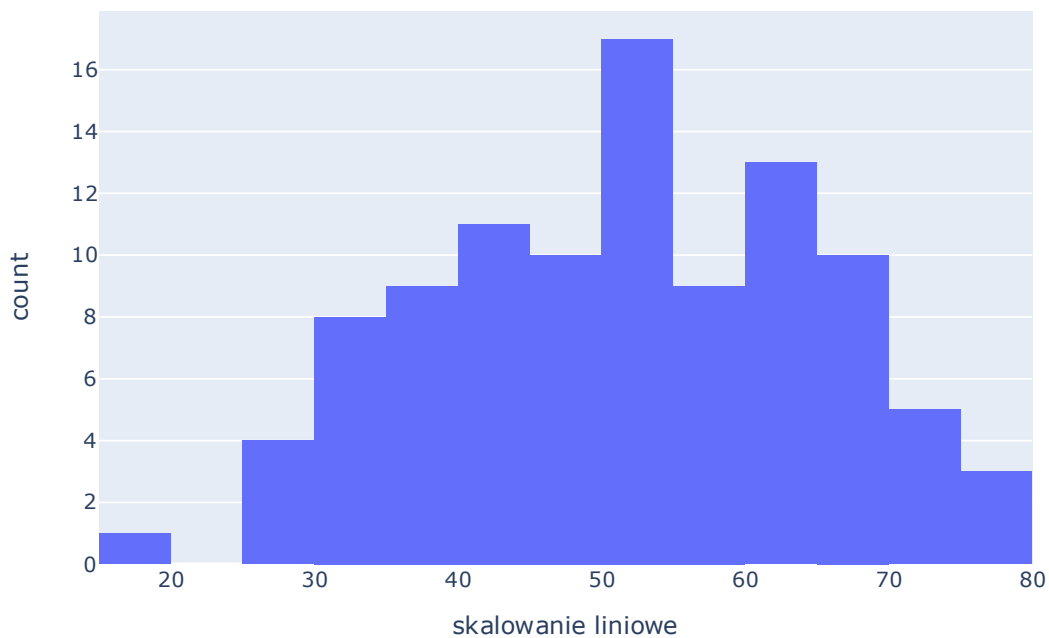
Rysunek 3 - Wykres skrzypcowy skalowania małej grupy dla paraboloidy

eval



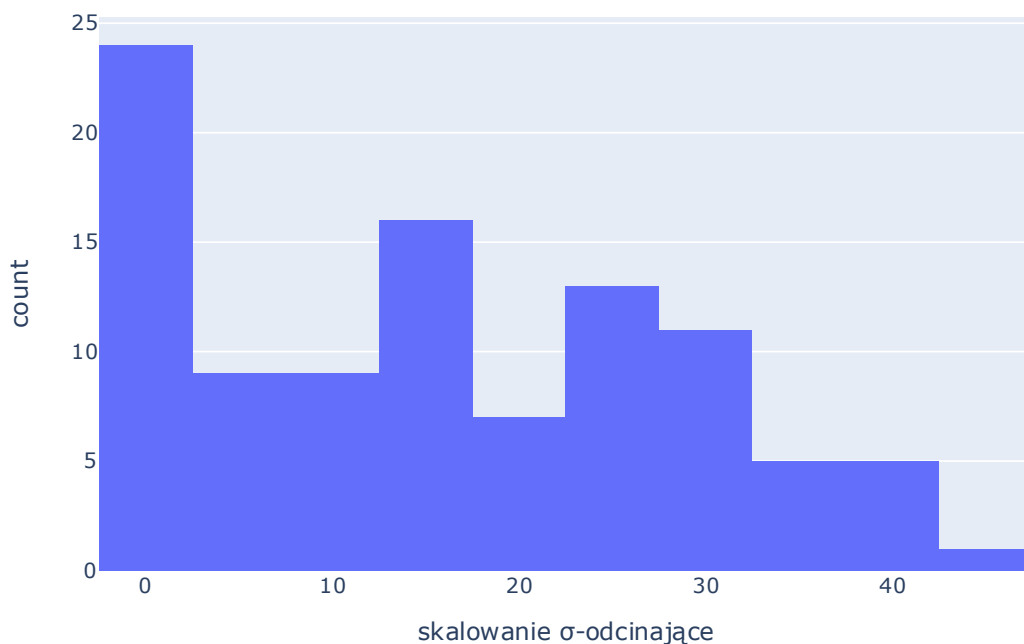
Rysunek 4 - histogram ewaluacji małej grupy dla paraboloidy

skalowanie liniowe



Rysunek 5 - histogram skalowania liniowego małej grupy dla paraboloidy

skalowanie σ -odcinające



Rysunek 6 - histogram skalowania σ -odcinającego małej grupy dla paraboloidy

Dużego zbiór danych

Pierwszy badany zbiór składał się z 100 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.

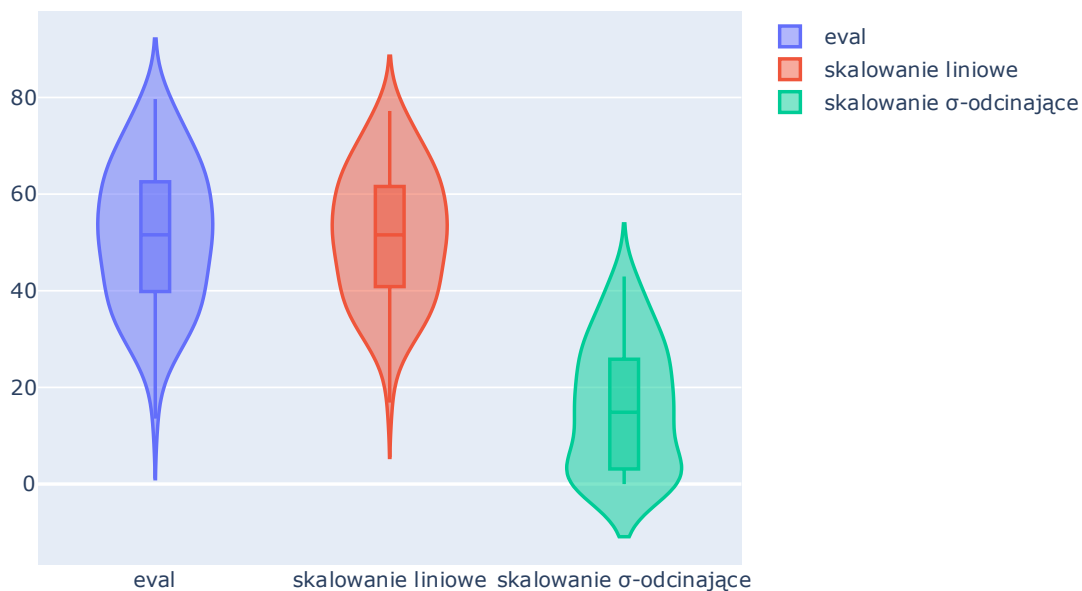
	x	y	z	f(x, y, z)	eval	skalowanie liniowe	skalowanie σ -odcinające
0	0.853710	1.524735	4.208345	20.763808	59.236192	58.557750	22.516371
1	-3.874223	1.800410	2.608654	25.056158	54.943842	54.640292	18.224021
2	4.203882	-2.645116	4.870098	48.387119	31.612881	33.347045	0.000000
3	-2.009808	4.102147	-4.048281	37.255517	42.744483	43.506419	6.024661
4	1.772747	3.784119	0.624922	17.852716	62.147284	61.214588	25.427462
...
95	4.348954	-0.340486	-1.250281	20.592539	59.407461	58.714061	22.687640
96	-2.522757	1.627496	2.307300	14.336678	65.663322	64.423537	28.943500
97	-4.735369	4.409538	0.026306	41.868433	38.131567	39.296393	1.411746
98	-0.696990	0.707975	1.879944	4.521213	75.478787	73.381725	38.758966
99	2.114185	-3.256954	-0.195645	15.115805	64.884195	63.712459	28.164374

Rysunek 7 - Przykładowe wartości dużej grupy dla paraboloidy

	eval	skalowanie liniowe	skalowanie σ -odcinające
std	11.801293	11.644734	10.313839
min	25.872711	26.218020	0.000000
25%	44.597092	44.693999	4.496458
50%	52.873210	52.860325	12.772577
75%	60.184665	60.074785	20.084031
max	78.201789	77.852890	38.101155

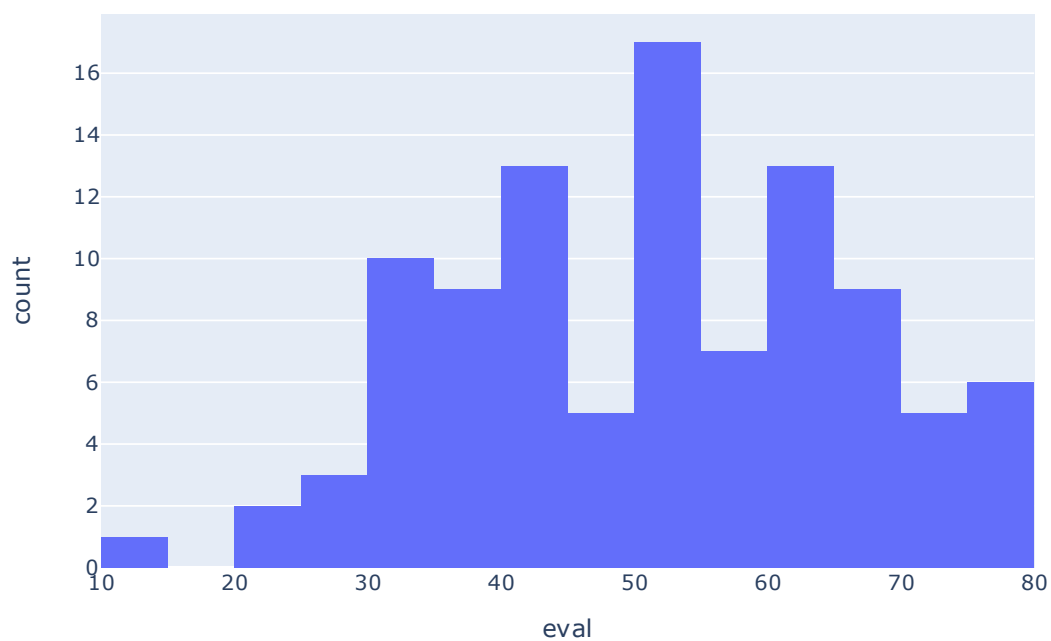
Rysunek 8 - Parametry statystyczne skalowania dużej grupy dla paraboloidy

Wykres skrzypcowy skalowania dużej grupy



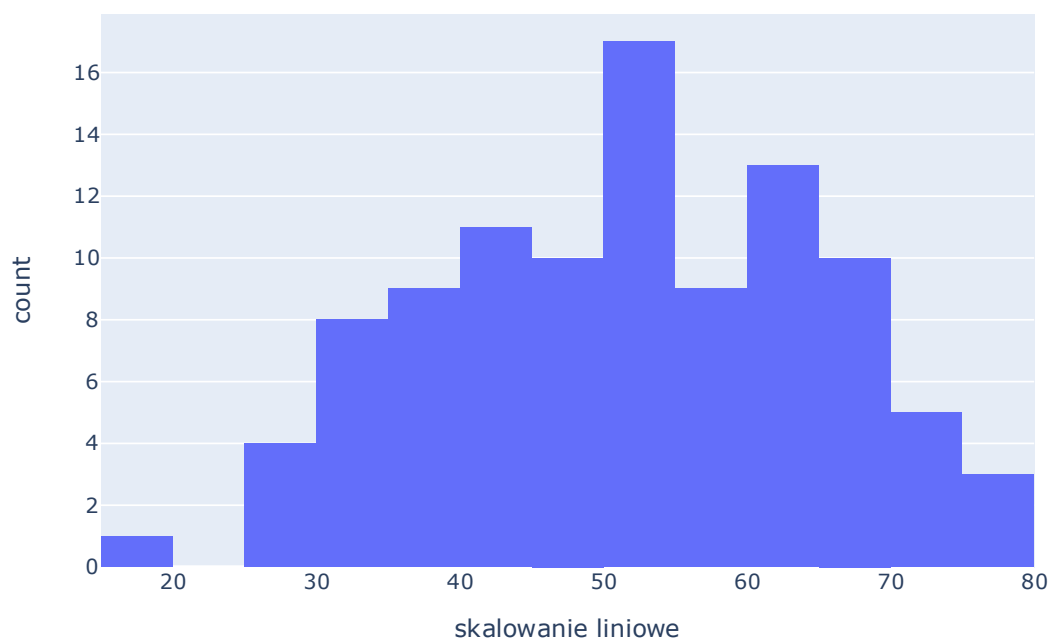
Rysunek 9 - Wykres skrzypcowy skalowania małej dużej grupy dla paraboloidy

eval



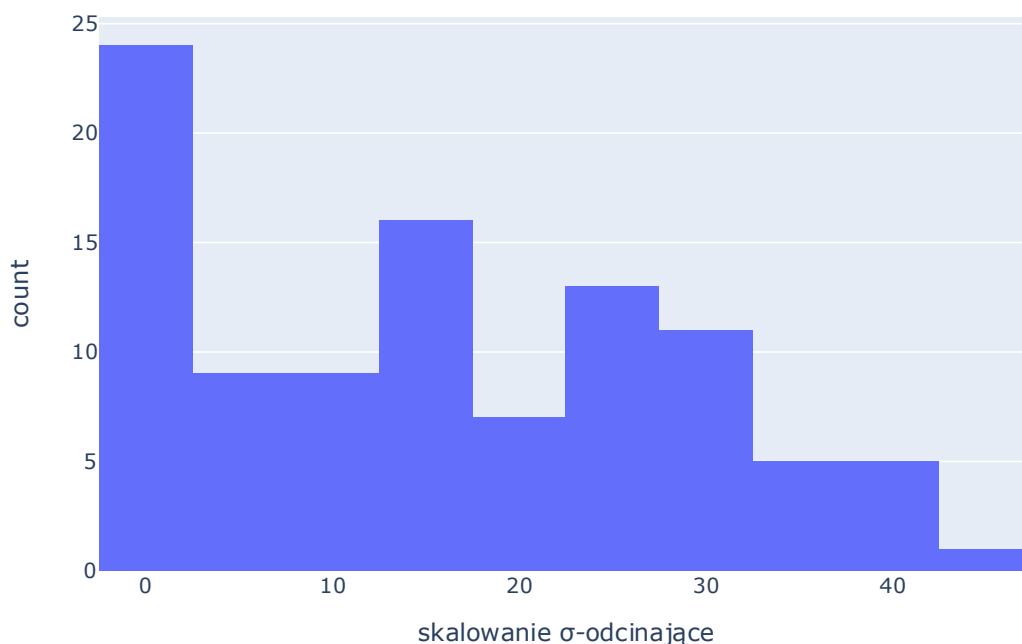
Rysunek 10 - histogram ewaluacji dużej grupy dla paraboloidy

skalowanie liniowe



Rysunek 11 - histogram skalowania liniowego dużej grupy dla paraboloidy

skalowanie σ -odcinające



Rysunek 12 - histogram skalowania σ -odcinającego dużej grupy dla paraboloidy

Wnioski

Skalowanie liniowe w niewielkim stopniu modyfikuje parametry statystyczne danych.

Na podstawie histogramów można zaobserwować, że skalowanie liniowe niejako wypełnia luki wygładzając histogram wartości.

Skalowanie σ -odcinające zmniejsza wartości każdego elementu. Jednak nie wpływa zbyt mocno na odchylenie standardowe danych.

Na wykresach skrzypcowych oraz na histogramach można zaobserwować, że skalowanie σ -odcinające grupuje sporą część danych w okolicy początkowych wartości. Może to wynikać z mapowania wszystkich ujemnych wartości na 0.

Zmniejszenie wartości wszystkich wartości w skalowaniu σ -odcinającym wynika bezpośrednio ze wzoru, który zakłada odejmowanie stałej od każdej wartości danych.

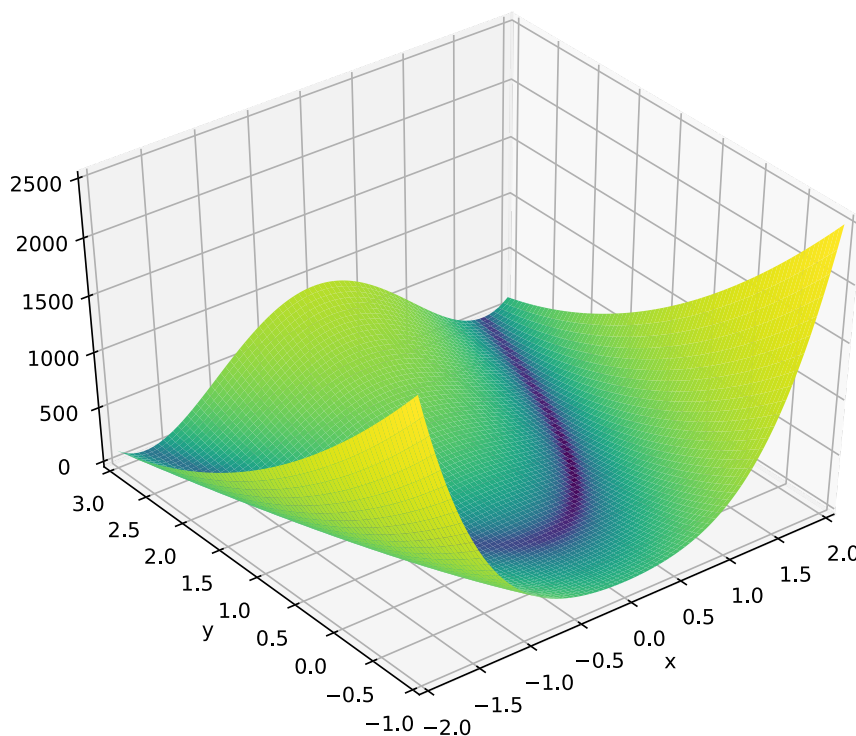
Badanie funkcji siodła Rosenbrocka

Funkcja Rosenbrocka jest parametryzowaną funkcją dwóch zmiennych. Jej wykres przypomina kanion wyżłobiony przez górski strumień. Funkcja ta jest często wykorzystywana przy testowaniu algorytmów optymalizacyjnych. Bardzo łatwo znaleźć jest jej dolinę, jednak dojście do minimum globalnego jest nietrywialne.

Funkcja dana jest poniższym wzorem:

$$f(x, y) = (a - x)^2 + b(y - x^2)^2, \quad (-2,048 \leq x \leq 2,048)$$

Typowo funkcja ta jest stosowana z parametrami $a = 1$ i $b = 100$. Takie parametry zostały zastosowane w badaniach laboratoryjnych.



*Rysunek 13 - Wykres funkcji Rosenbrocka dla standardowych parametrów
Źródło: Wikipedia*

Przy pomocy dwóch zbiorów danych testowych została zbadana metoda skalowania rankingowa linowa oraz skalowania logarytmicznego.

Mały zbiór danych

Pierwszy badany zbiór składał się z 20 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.



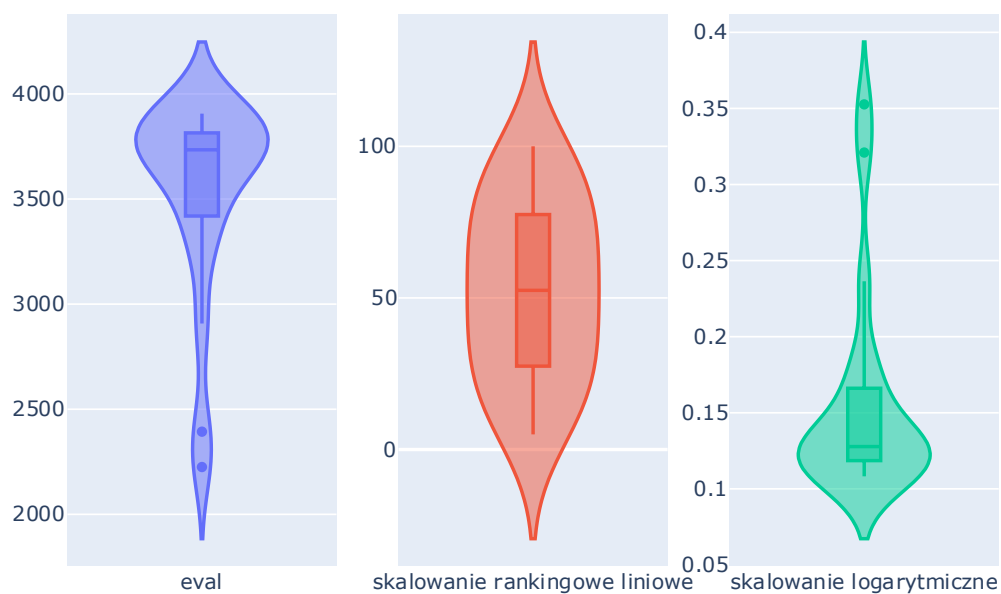
	x	y	f(x, y)	eval	skalowanie rankingowe liniowe	skalowanie logarytmiczne
0	0.634713	-1.906510	533.452622	3376.547378	25	0.171527
1	-0.073416	1.870197	348.902599	3561.097401	35	0.148416
2	0.132510	-1.074342	119.977267	3790.022733	75	0.121358
3	1.717462	0.412097	644.446334	3265.553666	20	0.186043
4	0.511012	-0.869054	127.971576	3782.028424	70	0.122275
5	-0.286525	1.434662	184.598369	3725.401631	45	0.128827
6	-1.882055	0.389668	1002.109127	2907.890873	15	0.236422
7	1.564549	-1.656557	1684.905320	2225.094680	5	0.352652
8	-1.761620	1.909416	150.163583	3759.836417	60	0.124831
9	-0.472850	-1.888448	448.238224	3461.761776	30	0.160703
10	-0.740152	1.677538	130.653133	3779.346867	65	0.122583
11	-0.408050	0.807220	43.034147	3866.965853	85	0.112630
12	1.030318	1.903448	70.879144	3839.120856	80	0.115768
13	0.275119	-1.262220	179.525834	3730.474166	50	0.128236
14	1.131032	1.080018	3.985875	3906.014125	95	0.108266
15	-0.560624	-1.183449	226.760404	3683.239596	40	0.133770
16	-0.469056	0.352583	3.915598	3906.084402	100	0.108258
17	-1.653960	-1.149915	1516.753944	2393.246056	10	0.321013
18	-1.157124	1.290828	4.884627	3905.115373	90	0.108366
19	-1.680944	1.541746	172.008969	3737.991031	55	0.127362

Rysunek 14 - wartości małej grupy dla siódła Rosenbrocka

	eval	skalowanie rankingowe liniowe	skalowanie logarytmiczne
std	486.934142	29.580399	0.069263
min	2225.094680	5.000000	0.108258
25%	3440.458177	28.750000	0.119961
50%	3734.232598	52.500000	0.127799
75%	3802.297264	76.250000	0.163409
max	3906.084402	100.000000	0.352652

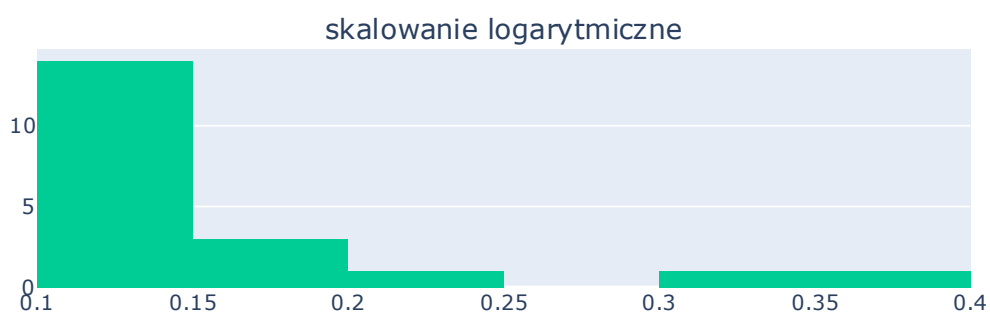
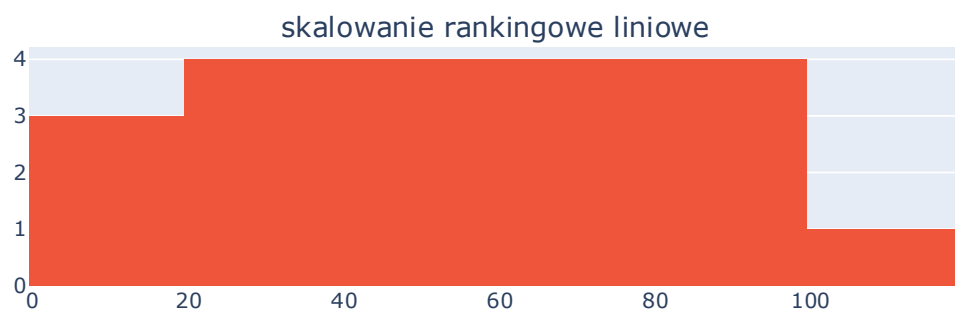
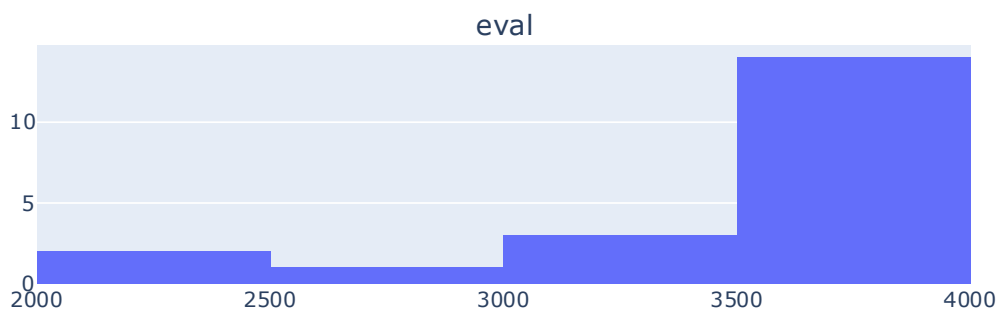
Rysunek 15 - Parametry statystyczne skalowania małej grupy dla siódła Rosenbrocka

Wykresy skrzypcowe skalowania małej grupy



Rysunek 16 - Wykres skrzypcowy małej grupy dla siódma Rosenbrocka

Histogramy skalowania małej grupy



Rysunek 17 - Histogramy małej grupy dla siódła Rosenbrocka

Duży zbiór danych

Pierwszy badany zbiór składał się z 100 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.

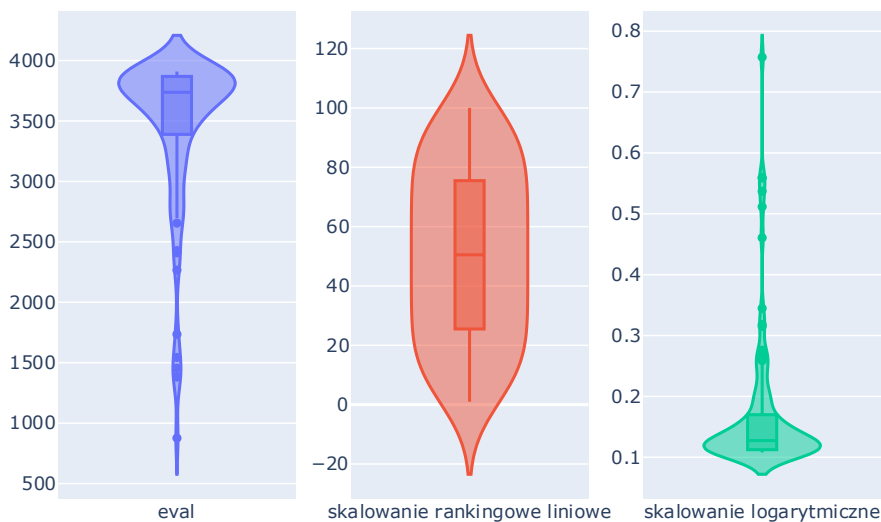
	x	y	f(x, y)	eval	skalowanie rankingowe liniowe	skalowanie logarytmiczne
0	-1.815931	-0.173172	1212.557586	2697.442414	11	0.269048
1	-0.151906	-0.819539	72.326844	3837.673156	67	0.115932
2	-0.009495	-1.072893	116.148282	3793.851718	60	0.120920
3	1.113961	-0.244542	220.669149	3689.330851	46	0.133052
4	0.136136	-1.764445	318.647227	3591.352773	39	0.144742
...
95	-0.378027	0.228658	2.634320	3907.365680	96	0.108116
96	-1.858741	-1.564371	2527.498851	1382.501149	2	0.559334
97	-0.088662	0.569586	32.738715	3877.261285	81	0.111475
98	-0.834157	-1.430108	455.320392	3454.679608	27	0.161592
99	-1.833806	-1.291686	2174.494675	1735.505325	6	0.460574

Rysunek 18 - Przykładowe wartości dużej grupy dla siódła Rosenbrocka

	eval	skalowanie rankingowe liniowe	skalowanie logarytmiczne
std	651.212412	29.011492	0.114233
min	876.492468	1.000000	0.107946
25%	3393.954078	25.750000	0.112487
50%	3737.516794	50.500000	0.127418
75%	3868.234584	75.250000	0.169295
max	3908.891637	100.000000	0.757252

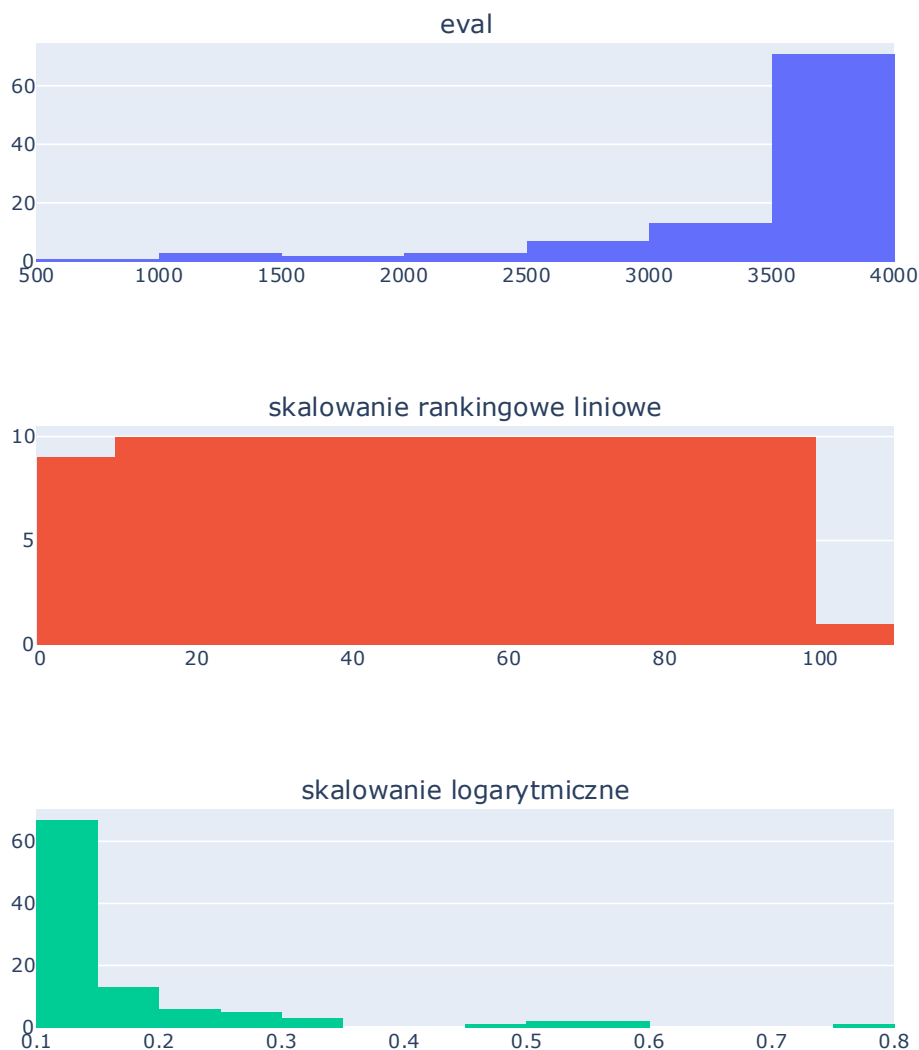
Rysunek 19 - Parametry statystyczne skalowania dużej grupy dla siódła Rosenbrocka

Wykresy skrzypcowe skalowania dużej grupy



Rysunek 20 - Wykresy skrzypcowe dużej grupy dla siódła Rosenbrocka

Histogramy skalowania dużej grupy



Rysunek 21 - Histogramy dużej grupy dla siodła Rosenbrocka

Wnioski

Wykresy skrzypcowe oraz histogramy funkcji ewaluacji i skalowania logarytmicznego są symetryczne odbite względem osi y. Świadczy to o odwróceniu kolejności danych poprzez skalowanie liniowe.

Wykres skrzypcowy skalowania rankingowego liniowego jest równo podzielony na cztery kwartyle. Świadczy to o równym rozkładzie punktów. Potwierdza to również praktycznie idealnie płaski histogram tego skalowania.

Bardzo mały odchyleniu standardowym oraz skali wykresu skrzypcowego można zaobserwować, że skalowanie logarytmiczne mocno zbliża punkty do siebie oraz do początku układu współrzędnych. Widać to również po zmianie wartości minimalnej i maksymalnej przeskalowanych wartości.

Na wykresie pudełkowym wartości funkcji ewaluacji można zaobserwować punkty bardzo oddalone od reszty. Jest to szczególnie widoczne na wykresie dużej grupy testowej. Dane te świadczą o gwałtownym spadku wartości funkcji w pobliżu jej ekstremum. Po rozkładzie kwartyli oraz jądrowego estymatora gęstości można zaobserwować, że obszar o znacznym spadku wartości funkcji jest stosunkowo niewielki.