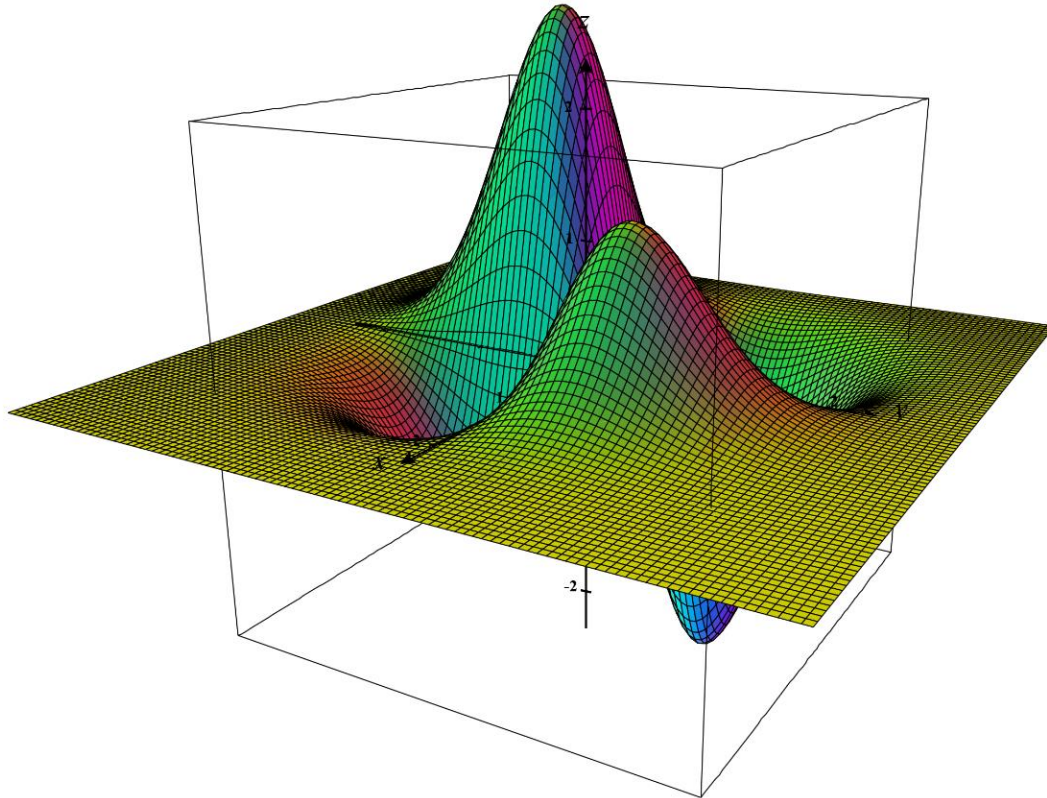


Nasza funkcja, $f(x, y) = \frac{10xy}{e^{x^2+0.5x+y^2}}$, przedstawiona za pomocą strony internetowej <https://c3d.libretexts.org/CalcPlot3D/index.html> wygląda następująco:



Minima funkcji: -1.330163 oraz -2.707631.

Maksima funkcji: 1.330163 oraz 2.707631.

- Co jest przestrzenią poszukiwań?
Przestrzenią przeszukiwań jest dwuwymiarowa płaszczyzna (jeden wymiar to oś x, drugi to oś y).
- Czy jest potrzebny gradient?
Nie, do ES nie jest potrzebny gradient.
- Jak odchylenie standardowe w szumie mutacji wpływa na wyniki (wartości 0.1, 1, 10, 100)?
Jego działanie przypomina parametr learning_rate z algorytmu SGD – zbyt małe wartości powodują wolne przeszukiwanie przestrzeni, co może skutkować zatrzymaniem algorytmu daleko od docelowego punktu, natomiast zbyt duże spowodują, że będziemy je „przeskakiwać” co również znacząco utrudni osiągnięcie pożądanej przez nas wartości.
Zostało to sprawdzone w sposób eksperymentalny, co zostało udokumentowane w postaci następującej tabeli:

Wartość odchylenia standardowego	Wynik otrzymany w:		
	5 iteracjach	15 iteracjach	30 iteracjach
0.01	-0.0549172	-0.3335592	-0.9269302
0.1	-2.5114302	-2.7061500	-2.7061501
1	-2.5026035	-2.6000943	-2.6656818
10	-0.0591743	-1.8079515	-1.8079515
100	-0.0000001	-1.2533152	-1.2533152

Na tej podstawie możemy stwierdzić, że 0.01 jest przykładem wartości zbyt małej – algorytm dążył, lecz nie udało mu się dotrzeć do końca. Wartość 0.1 osiągnęła jako jedyna oczekiwany wynik, dlatego też do eksperymentów w dalszej części zadania będziemy używać tej liczby. Pozostałe wartości są już natomiast za duże – wszystkie w mniejszym lub większym stopniu omijały prawdziwe minimum funkcji. Wszystkie eksperymenty zostały przeprowadzone dla pięciu początkowych osobników oraz $\lambda = 15$.

- Jak liczba rodziców μ i dzieci λ wpływa na proces optymalizacji (wartości $\{1, 1\}$, $\{1, 16\}$, $\{16, 1\}$, $\{16, 16\}$, $\{128, 512\}$ i 8 innych dowolnych).

Aby dobrze przedstawić odpowiedź na to pytanie posłużę się tabelką:

μ, λ	Ile razy zostało znalezione minimum:					
	lokalne przy limicie iteracji:			globalne przy limicie iteracji:		
	5	15	50	5	15	50
1, 1	1	1	1	0	0	0
1, 16	0	0	0	0	1	0
16, 1	4	4	4	2	5	5
16, 16	2	4	0	4	3	6
128, 512	10	7	9	15	16	13
3, 5	0	0	2	1	0	2
5, 3	1	2	0	0	1	1
5, 25	0	2	1	0	0	1
25, 5	6	6	4	4	5	4
64, 256	12	7	5	11	12	13
256, 64	4	5	11	23	22	18
256, 1024	8	6	7	20	22	18
1024, 256	1	1	2	29	29	28
1024, 1024	0	1	0	29	29	30

Powyższe dane pokazują, że większa liczba rodziców i dzieci dają większą szansę na odnalezienie maksimum globalnego. Jest to spowodowane tym, że większa liczba dzieci znacznie rozszerza możliwości eksploracyjne przestrzeni rozwiązań, jednak również jest bardziej kosztowna obliczeniowo.

- Proszę punkt startowy ES (dla $\mu, \lambda=(128, 512)$) oraz algorytmu SGD z poprzedniego zadania ustawić na punkt (10,10) i porównać wyniki. Co i dlaczego można zaobserwować?
W przypadku ES po 100 iteracjach algorytm zakończył się w punkcie (19.15325464, 18.41754936) – obliczona wartość funkcji: 0.0, natomiast w przypadku algorytmu SGD otrzymany punkt końcowy jest również punktem startowym – (10, 10) z wartością $9.324621449370703e-87$. Jest ona na tyle mała, że możemy uznać, iż obydwa algorytmy otrzymały wynik równy 0. Powodem tego był zbyt bardzo oddalony punkt startowy od minimum funkcji oraz delikatne „nachylenie”, które w przypadku ES sugerowało, aby zwiększać wartości x oraz y .
- Kiedy ES jest lepszym rozwiązaniem niż SGD?
ES jest lepszym rozwiązaniem niż SGD w przypadku, kiedy funkcja celu ma wiele minimów lokalnych – daje on większą szansę na uniknięcie „utknięcia” w najbliższym optimum. Również powinna się lepiej sprawdzić, gdy mamy dużą przestrzeń przeszukiwań.
Dodatkową korzyścią dla ES jest to, że nie wymaga od nas wiedzy na temat gradientu, co umożliwia nam użycie tego algorytmu w większej ilości przypadków niż SGD.

- Który proces optymalizacji jest bardziej złożony obliczeniowo?
ES jest bardziej złożona obliczeniowo gdyż:
 - w każdej iteracji generuje się wiele dzieci (a w SGD generujemy tylko jeden punkt),
 - rozmiar populacji do oceny jest znacznie większy (analogicznie jak powyżej, w SGD oceniamy tylko jeden punkt),
 - wprowadzony jest element losowości dla każdego osobnika - podczas krzyżowania oraz mutacji,
 - potrzebny jest sposób selekcji oraz eliminacji, co również pochłania moc obliczeniową.