Michał Szczurek
Informatyka II rok, grupa wtorek 12:50 B

Algorytmy geometryczne, ćwiczenie 1- Sprawozdanie

1. Wprowadzenie

Jednym z zagadnień szczególnie istotnych w geometrii obliczeniowej jest określenie po której stronie prostej znajduje się punkt. W tym celu można zastosować algorytm obliczający wartość jednego z dwóch wyznaczników:

a) Wyznacznik macierzy 3x3: $\begin{vmatrix} a_x & a_y & 1 \\ b_x & b_y & 1 \\ c_x & c_y & 1 \end{vmatrix}$

b) Wyznacznik macierzy 2x2: $\begin{vmatrix} a_x - c_x & a_y - c_y \\ b_x - c_x & b_y - c_y \end{vmatrix}$

Gdzie p_x jest współrzędną x punktu p, a p_y- współrzędną y. Powyższe wyznaczniki pozwalają określić położenie punktu c względem prostej wyznaczonej przez punkty a i b- punkt c znajduje się z prawej strony prostej, gdy wartość wyznacznika jest mniejsza od 0, z lewej gdy wartość jest większa od 0 i na prostej gdy wartość wyznacznika wynosi 0. Pomimo, że powyższe wyznaczniki są sobie równoważne to na skutek niedoskonałości reprezentacji liczb rzeczywistych w komputerze wyniki mogą się różnić w zależności od użytego wyznacznika. Celem ćwiczenia było wygenerowanie danych, a następnie wykrycie różnic w klasyfikacji w zależności od wyznacznika (zarówno zaimplementowanego samodzielnie jak i bibliotecznego).

2. Specyfikacja środowiska, w którym wykonano ćwiczenie

Ćwiczenie zostało wykonane przy pomocy jupyter notebook i Python 3 na komputerze z systemem Windows 10 (64 bitowym) i procesorem Intel Core i-5- 9300H (64 bitowy). Wykresy wygenerowano przy pomocy dostarczonego na laboratoria narzędzia graficznego i biblioteki matplotlib.

3. Sposób wykonania ćwiczenia

Pierwszym etapem wykonania ćwiczenia było wygenerowanie czterech zestawów danych:

- tab1- 10^5 losowych punktów o współrzędnych z przedziału [-1000, 1000],
- tab2- 10^5 losowych punktów o współrzędnych z przedziału [-10^{14}, 10^{14}],
- tab3- 1000 losowych punktów leżących na okręgu o środku (0,0) i promieniu R=100,
- tab4- 1000 losowych punktów o współrzędnych z przedziału [-1000, 1000] leżących na prostej wyznaczonej przez wektor (a, b), gdzie a = [-1.0, 0.0], b = [1.0, 0.1].

Do wygenerowania powyższych zestawów użyto funkcji uniform z biblioteki random, która generuje zmienne pseudolosowe typu float64.

Następnie zaimplementowano cztery funkcje obliczające wyznacznik, pozwalający zdeterminować po której stronie prostej wyznaczonej przez dwa punkty znajduje się trzeci:

- det3x3- Funkcja obliczająca wyznacznik 3x3, zaimplementowana samodzielnie
- det2x2- Funkcja obliczająca wyznacznik 2x2, zaimplementowana samodzielnie
- numpy_det3x3- Funkcja obliczająca wyznacznik 3x3, dostarczona przez bibliotekę numpy
- numpy_det2x2- Funkcja obliczająca wyznacznik 2x2, dostarczona przez bibliotekę numpy

Kolejnym krokiem było klasyfikowanie punktów z poszczególnych zestawów ze względu na ich położenie w stosunku do prostej wyznaczonej przez wektor (a = [-1.0, 0.0], b= [1.0, 0.1]) w celu porównania wyników dla różnych wyznaczników. Klasyfikację punktów powtórzono dla wartości tolerancji błędów wynoszących odpowiednio: $\varepsilon_1 = 10^{-14}$, $\varepsilon_2 = 10^{-12}$, $\varepsilon_3 = 10^{-16}$, $\varepsilon_4 = 0$. Na końcu sprawdzono również, działanie wyznaczników dla zmiennych o innej precyzji- float32.

4. Klasyfikacja poszczególnych zestawów punktów

Wyniki klasyfikacji zebrano w tabelach. Kolumny „lewo", „prawo" i „współ." oznaczają punkty zaklasyfikowane odpowiednio na lewo, prawo oraz punkty współliniowe z prostą wyznaczoną przez (a, b). W wierszach zamieszczono informacje o wyznacznikach użytych w danej klasyfikacji.

Tabela 1. Wyniki klasyfikacji zestawu tab1

	ε_1			ε_2			ε_3			ε_4		
	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo
det 3x3	50167	0	49833	50167	0	49833	50167	0	49833	50167	0	49833
det 2x2	50167	0	49833	50167	0	49833	50167	0	49833	50167	0	49833
numpy_det3x3	50167	0	49833	50167	0	49833	50167	0	49833	50167	0	49833
numpy_det2x2	50167	0	49833	50167	0	49833	50167	0	49833	50167	0	49833

Tabela 2. Wyniki klasyfikacji zestawu tab2

	ε_1			ε_2			ε_3			ε_4		
	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo
det 3x3	50031	0	49969	50031	0	49969	50031	0	49969	50031	0	49969
det 2x2	50028	8	49964	50028	8	49964	50028	8	49964	50028	8	49964
numpy_det3x3	50031	0	49969	50031	0	49969	50031	0	49969	50031	0	49969
numpy_det2x2	50031	0	49969	50031	0	49969	50031	0	49969	50031	0	49969

Tabela 3. Wyniki klasyfikacji zestawu tab3

	ε_1			ε_2			ε_3			ε_4		
	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo
det 3x3	518	0	482	518	0	482	518	0	482	518	0	482
det 2x2	518	0	482	518	0	482	518	0	482	518	0	482
numpy_det3x3	518	0	482	518	0	482	518	0	482	518	0	482
numpy_det2x2	518	0	482	518	0	482	518	0	482	518	0	482

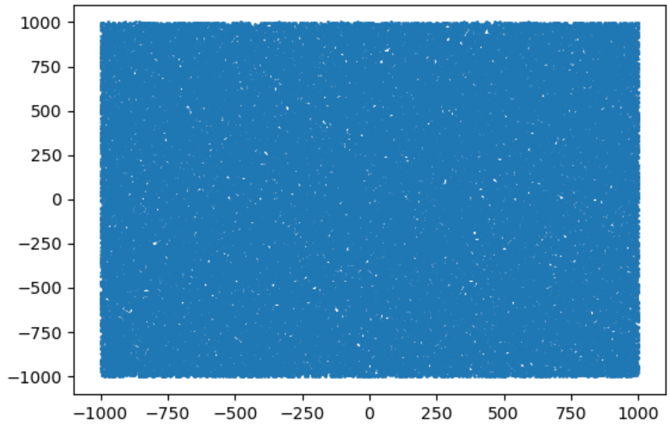
Tabela 4. Wyniki klasyfikacji zestawu tab4

	ε_1			ε_2			ε_3			ε_4		
	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo
det 3x3	0	1000	0	0	1000	0	171	462	367	171	460	369
det 2x2	137	743	120	82	854	64	142	733	125	142	733	125
numpy_det3x3	7	907	86	0	1000	0	469	36	495	483	1	516
numpy_det2x2	460	77	463	136	734	130	489	5	506	492	0	508

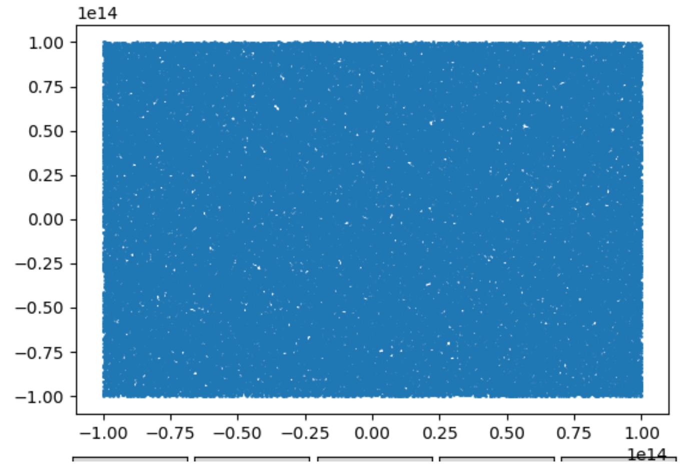
5. Wybrane wykresy

5.1 Wykresy zbiorów danych (bez klasyfikacji)

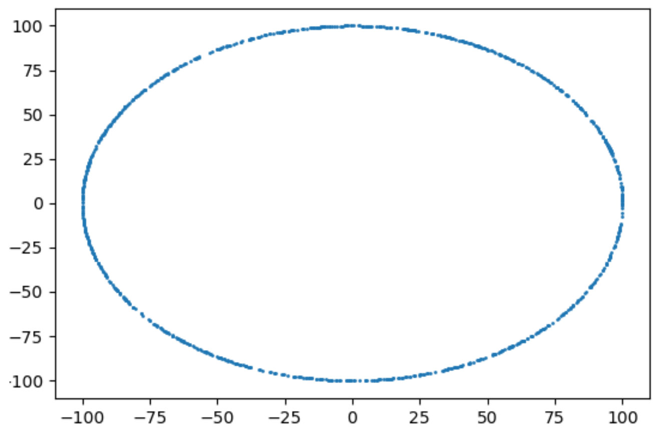
Wykres 1. Zbiór danych tab1



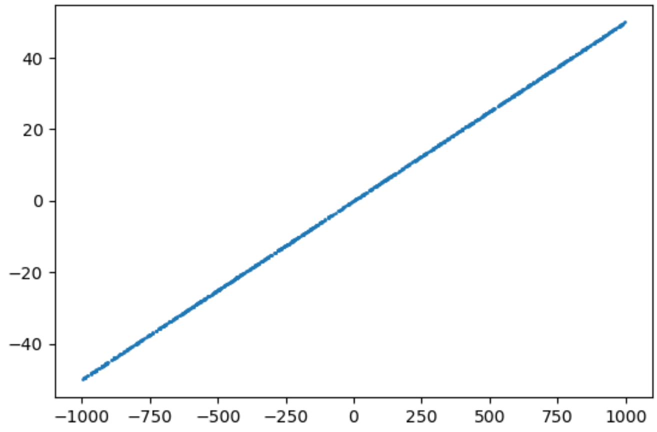
Wykres 2. Zbiór danych tab2



Wykres 3. Zbiór danych tab3



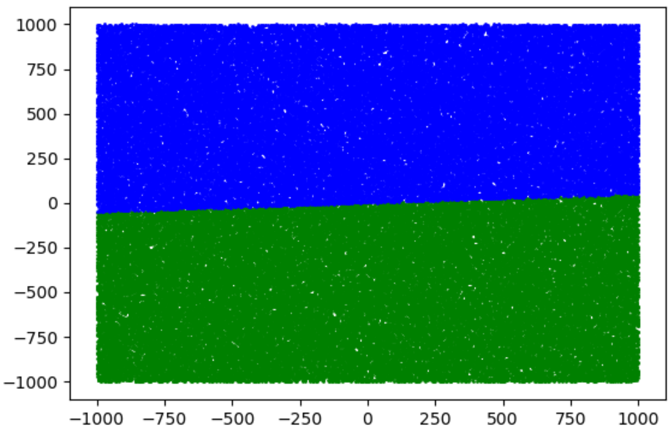
Wykres 4. Zbiór danych tab4



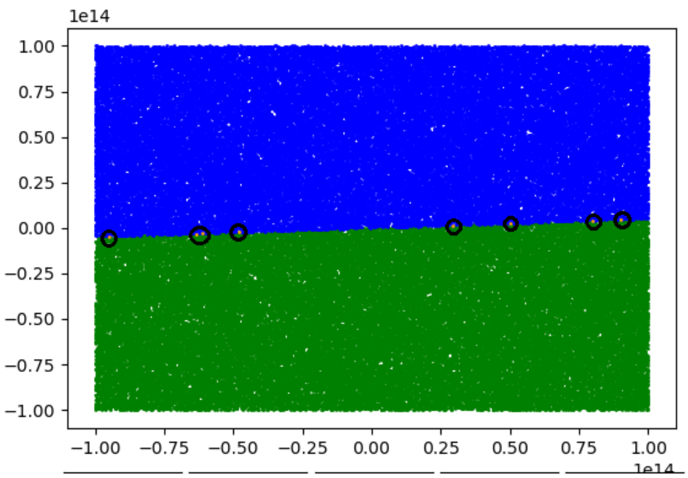
5.2 Wykresy klasyfikacji punktów z poszczególnych zestawów

Kolory na wykresie oznaczają odpowiednio: zielony – punkty na prawo od prostej, niebieski- punkty na lewo od prostej, pomarańczowy- punkty współliniowe.

Wykres 5. Klasyfikacja punktów zestawu tab1 (identyczna dla wszystkich tolerancji ε i wyznaczników)

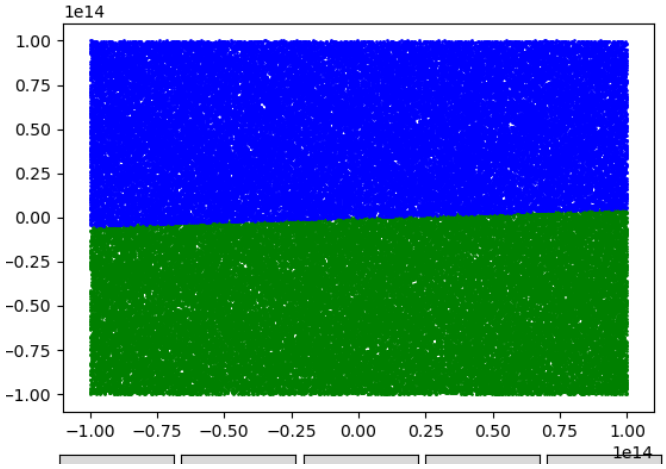


Wykres 6. Klasyfikacja punktów zestawu tab2 dla wyznacznika det2x2 (identyczna dla wszystkich tolerancji ε)

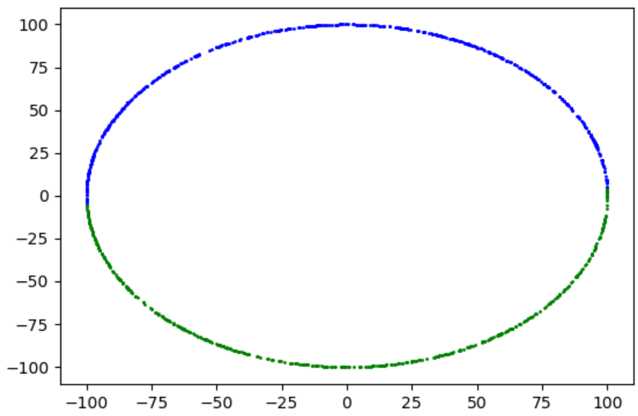


Na powyższym wykresie otoczono punkty współliniowe czarnymi okręgami w celu poprawy ich widoczności. (W 2. okręgu od lewej strony znajdują się 2 punkty).

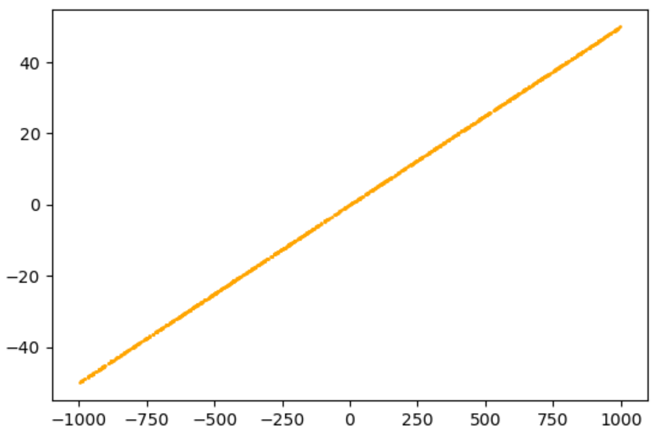
Wykres 7. Klasyfikacja punktów zestawu tab2 dla wyznaczników innych niż det2x2 (identyczna dla wszystkich tolerancji ε



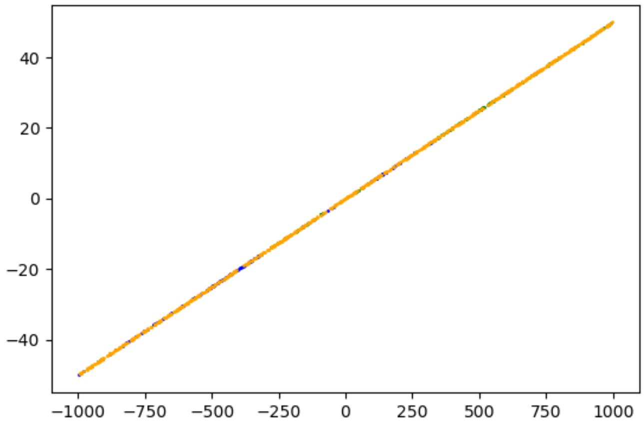
Wykres 8. Klasyfikacja punktów zestawu tab3 (identyczna dla wszystkich tolerancji ε i wyznaczników)



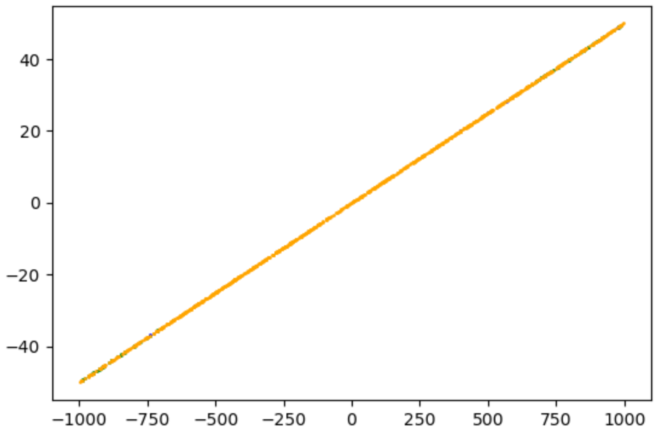
Wykres 8. Klasyfikacja punktów zestawu tab4 dla ε_1 i wyznacznika det3x3



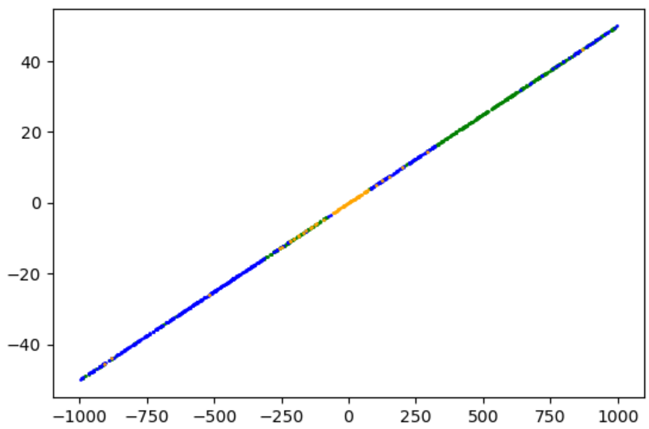
Wykres 9. Klasyfikacja punktów zestawu tab4 dla ε_1 i wyznacznika det2x2



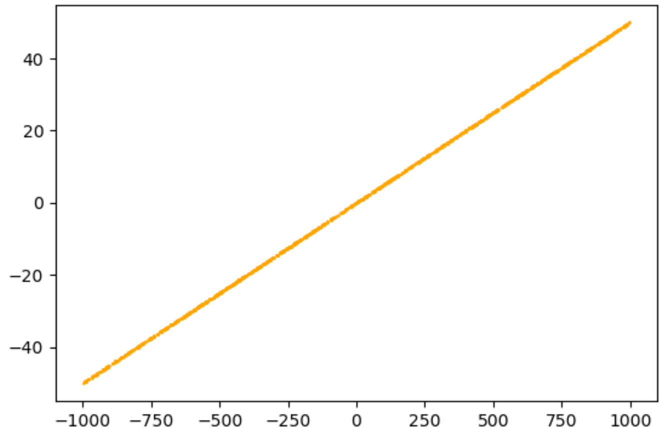
Wykres 10. Klasyfikacja punktów zestawu tab4 dla ε_1 i wyznacznika numpy_det3x3



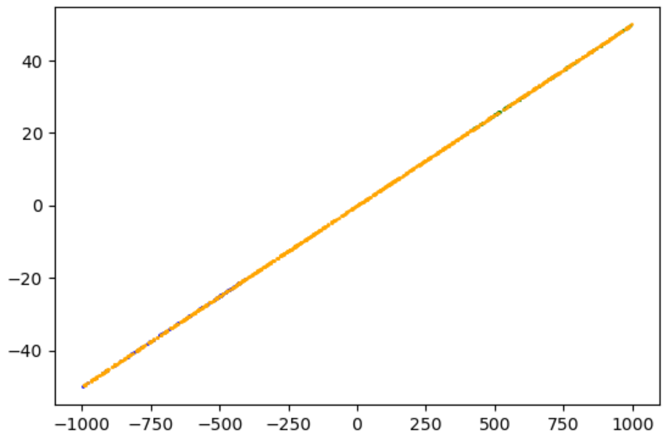
Wykres 11. Klasyfikacja punktów zestawu tab4 dla ε_1 i wyznacznika numpy_det2x2



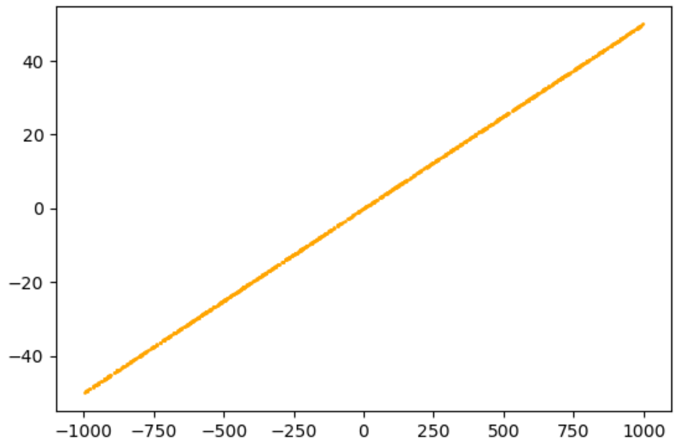
Wykres 12. Klasyfikacja punktów zestawu tab4 dla ε_2 i wyznacznika det3x3



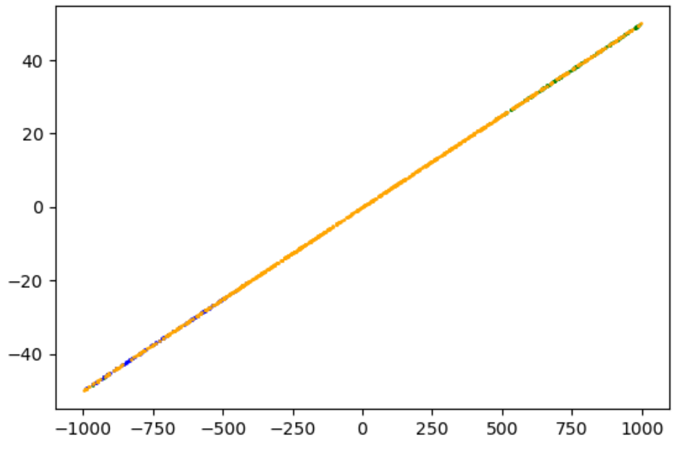
Wykres 13. Klasyfikacja punktów zestawu tab4 dla ε_2 i wyznacznika det2x2



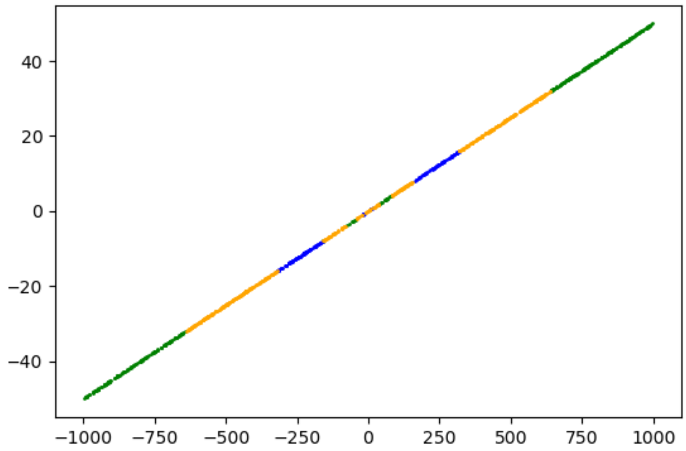
Wykres 14. Klasyfikacja punktów zestawu tab4 dla ε_2 wyznacznika numpy_det3x3



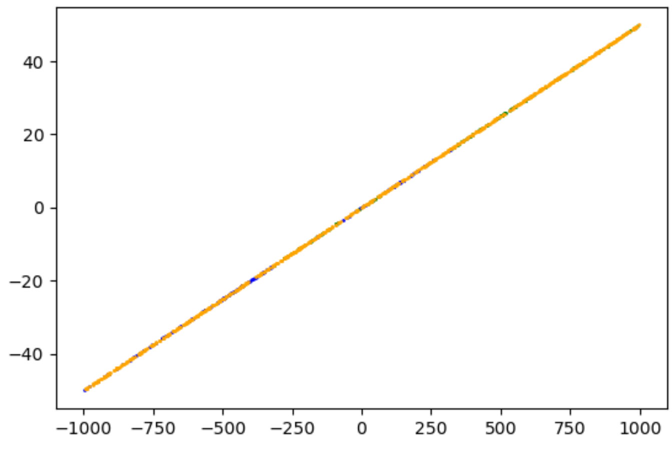
Wykres 15. Klasyfikacja punktów zestawu tab4 dla ε_2 i wyznacznika numpy_det2x2



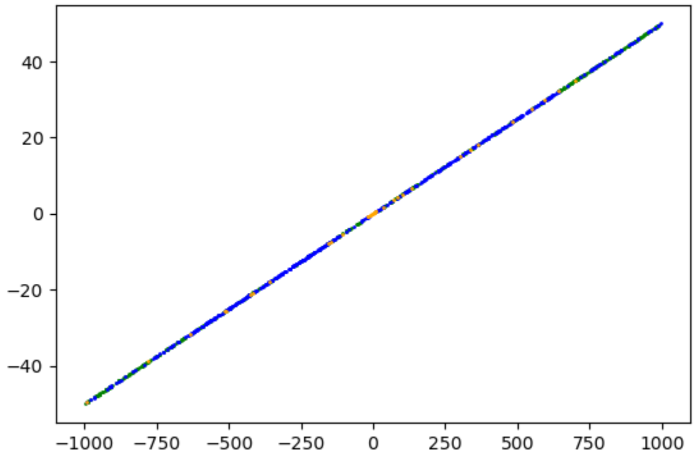
Wykres 16. Klasyfikacja punktów zestawu tab4 dla ε_3 i wyznacznika det3x3



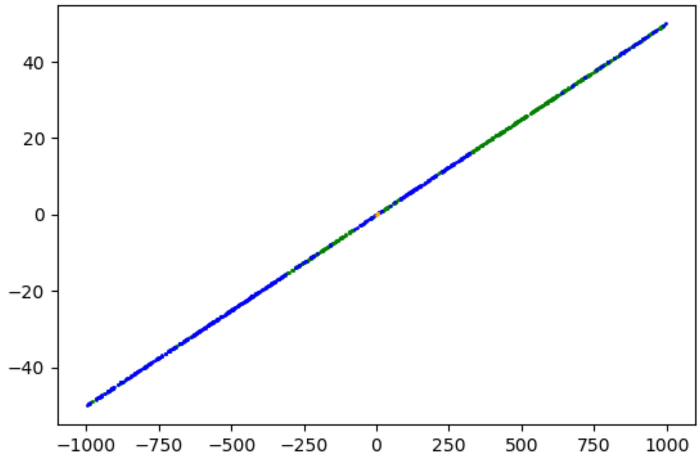
Wykres 17. Klasyfikacja punktów zestawu tab4 dla ε_3 i wyznacznika det2x2



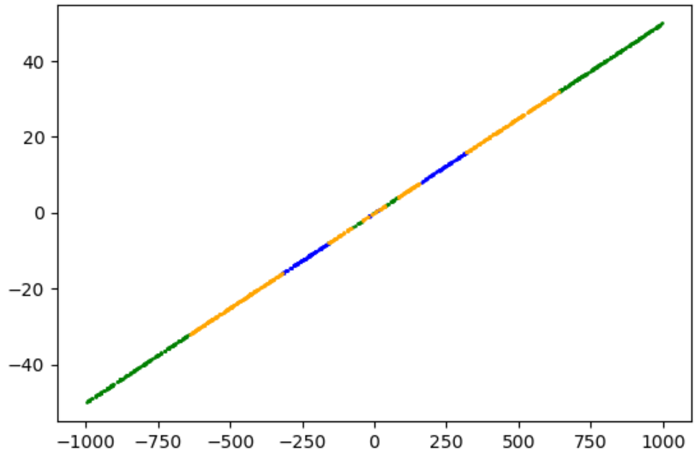
Wykres 18. Klasyfikacja punktów zestawu tab4 dla ε_3 i wyznacznika numpy_det3x3



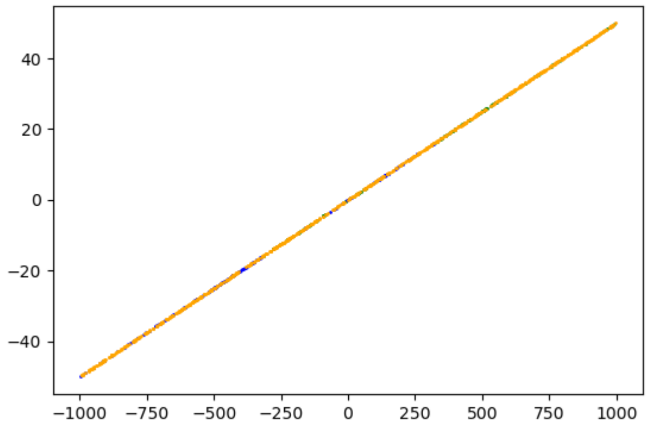
Wykres 19. Klasyfikacja punktów zestawu tab4 dla ε_3 i wyznacznika numpy_det2x2



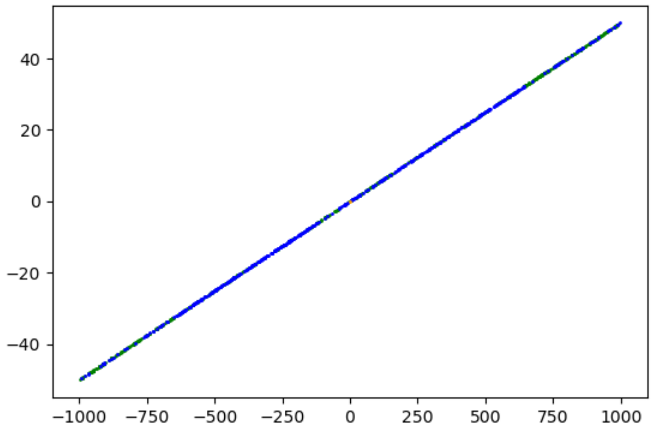
Wykres 20. Klasyfikacja punktów zestawu tab4 dla ε_4 i wyznacznika det3x3



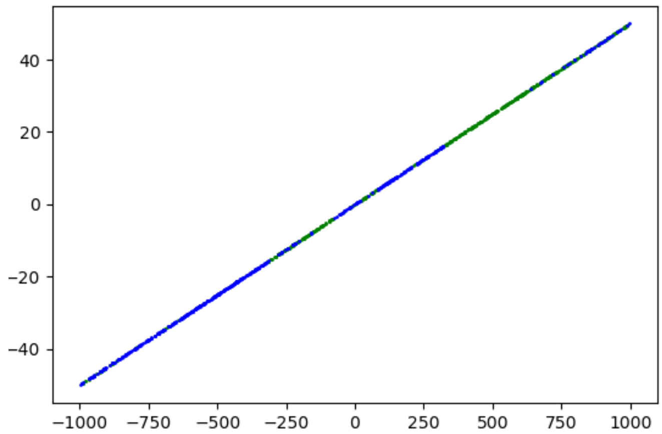
Wykres 21. Klasyfikacja punktów zestawu tab4 dla ε_4 i wyznacznika det2x2



Wykres 22. Klasyfikacja punktów zestawu tab4 dla ε_4 i wyznacznika numpy_det3x3



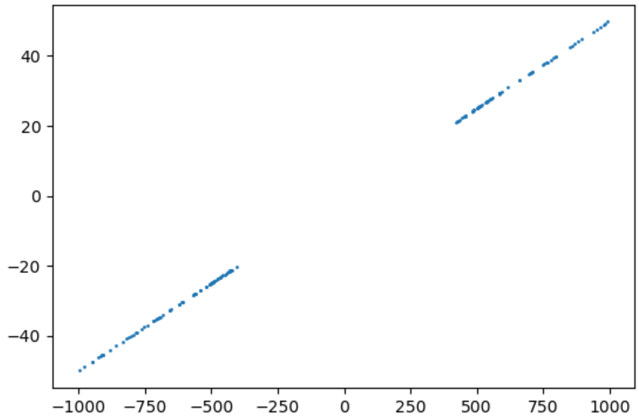
Wykres 23. Klasyfikacja punktów zestawu tab4 dla ε_4 i wyznacznika numpy_det2x2



5.3 Wykres punktów zaklasyfikowanych inaczej przez wyznaczniki

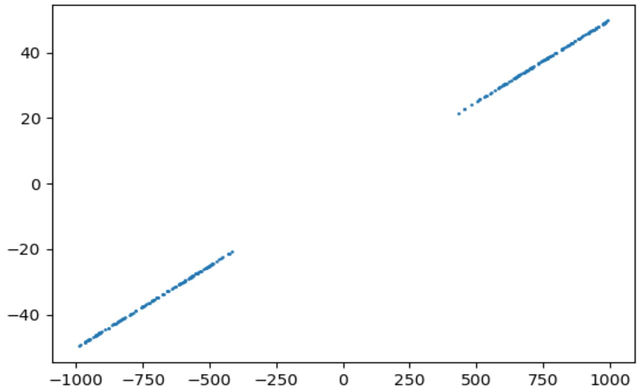
Poniższe wykresy przedstawiają punkty, które zostały sklasyfikowane inaczej w zależności od wyznacznika. Klasyfikacje dotyczące tych wykresów przeprowadzono z tolerancją ε_2.

Wykres 24. Różnice dla wyznaczników det3x3 i det2x2



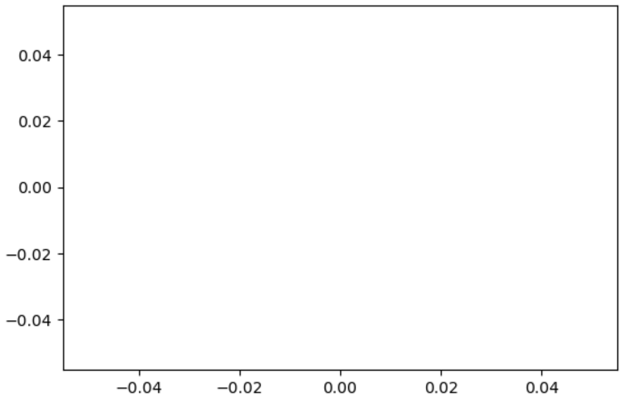
Ilość różnic: 146

Wykres 25. Różnice dla wyznaczników numpy_det3x3 i numpy_det2x2



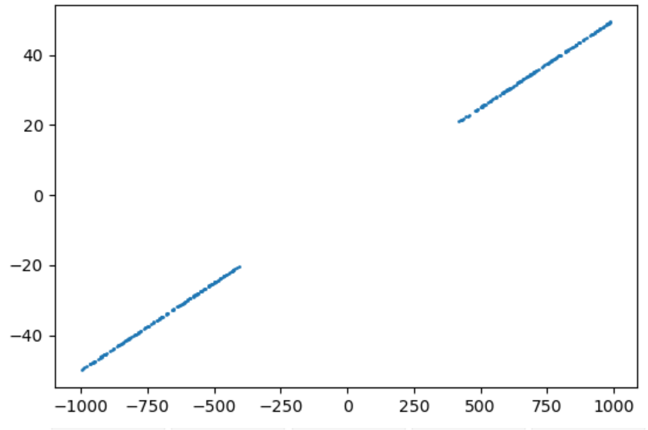
Ilość różnic: 266

Wykres 26. Różnice dla wyznaczników numpy_det3x3 i det3x3



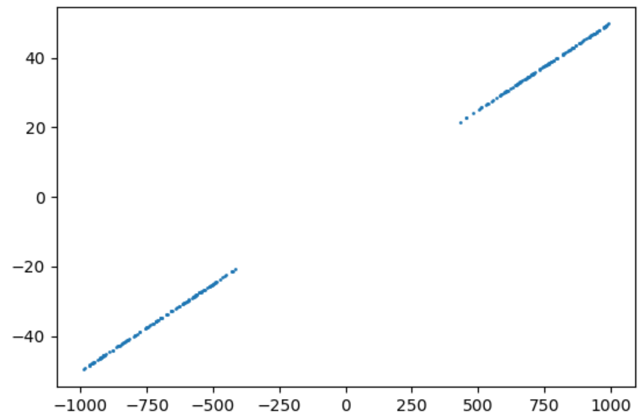
Ilość różnic: 0

Wykres 27. Różnice dla wyznaczników numpy_det2x2 i det2x2



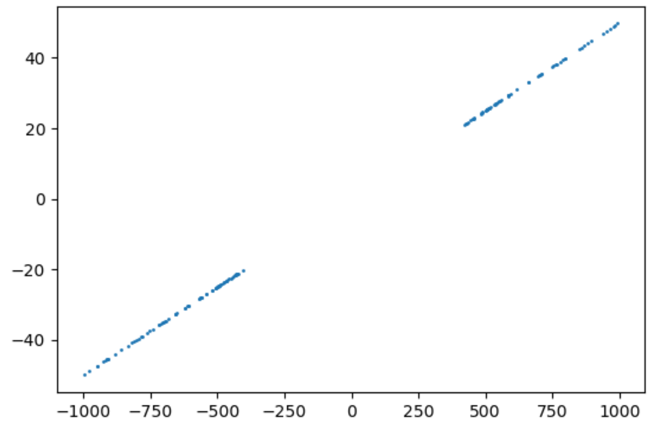
Ilość różnic: 303

Wykres 28. Różnice dla wyznaczników numpy_det2x2 i det3x3



Ilość różnic: 266

Wykres 29. Różnice dla wyznaczników numpy_det3x3 i det2x2



Ilość różnic: 146

6. Przeprowadzenie testów dla zmiennych typu float32

W celu sprawdzenia jaki wpływ na wartość wyznacznika ma typ zmiennych przechowujących danych stworzono piąty zestaw danych tab4_32. Z racji, że biblioteka random nie dostarcza możliwości generowania zmiennych typu float32, do utworzenia tego zbioru danych użyto punktów z zestawu tab4 przekonwertowanych przy pomocy funkcji numpy.float32() z biblioteki numpy. Wyniki klasyfikacji przedstawiono w tabeli poniżej

Tabela 5. Wyniki klasyfikacji zestawu tab4_32

	ε_1			ε_2			ε_3			ε_4		
	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo	lewo	współ.	prawo
det 3x3	401	177	422	401	177	422	456	55	489	456	55	489
det 2x2	410	162	428	401	177	422	410	162	428	410	162	428
numpy_det3x3	405	168	427	401	177	422	497	6	497	498	0	502
numpy_det2x2	483	12	505	427	124	449	491	2	507	491	0	509

7. Obserwacje i wnioski

Klasyfikacja danych z zestawów tab1, tab2, tab3 w większości dawała identyczne rezultaty niezależnie od użytego wyznacznika oraz dobranej tolerancji. Jest to spowodowane tym, że istnieje niewielkie prawdopodobieństwo tego, że losowo wygenerowany punkt będzie znajdował się na prostej (z dokładnością do tolerancji). Wyjątek stanowi tutaj klasyfikacja zestawu tab2 przy użyciu wyznacznika det2x2, która wskazała 8 punktów jako punkty współliniowe. Jest to najprawdopodobniej spowodowane tym, że współrzędne tych punktów mają duże bezwzględnie wartości (rzędu 10^{12}, 10^{13}). Operacja klasyfikacji wymaga z kolei znacznej precyzji (największa testowana tolerancja wynosiła 10^{-12}), która jest tracona na skutek operowania takimi wartościami- z racji, że mantysa liczby zmiennoprzecinkowej jest skończona to dla „dużej" liczby będzie przechowywana tylko informacja o pewnej liczbie cyfr, pomijając tym samym, ważne dla tego ćwiczenia, cyfry znajdujące się na odległych miejscach dziesiętnych.

Dla tolerancji $\varepsilon \geq 10^{-14}$ najlepsze wyniki dawały wyznaczniki typu 3x3 (zarówno wersja zaimplementowana samodzielnie jak i biblioteczna). W szczególności dla tolerancji 10^{-12} oba wyznaczniki zaklasyfikowały wszystkie punkty zestawu tab4 poprawnie.

Funkcje biblioteczne sprawdziły się wyraźnie gorzej od ich odpowiedników napisanych samodzielnie. Różnica pomiędzy funkcjami bibliotecznymi, a zaimplementowanymi samodzielnie jest szczególnie widoczna dla $\varepsilon = 10^{-16}$ w zestawie tab4. Dla tej tolerancji widoczne było wyraźne pogorszenie się względem większych wartości ε. Podczas gdy dla tolerancji $\varepsilon = 10^{-12}$ większość punktów była klasyfikowana przez funkcje bibliotecznie poprawnie, to dla tolerancji $\varepsilon \leq 10^{-14}$ dla wyznacznika numpy_det2x2 i $\varepsilon \leq 10^{-16}$ dla numpy_det3x3 znaczna część punktów została błędnie sklasyfikowana. Najprawdopodobniej jest to przyczyną zastosowania optymalizacji- macierz była przekształcana najpierw do postaci schodkowej, co w przypadku małych macierzy (3x3 i 2x2) zwiększa liczbę elementarnych operacji- a więc zmniejsza precyzję wyniku. Innym możliwym powodem gorszych wyników funkcji dostarczonych przez numpy może być konwersja typów natywnych python'a do typów, używanych przez bibliotekę.

Analizując rezultaty osiągnięte poprzez wyznacznik det3x3 i det2x2 można zauważyć, że pomimo, iż funkcja używająca wyznacznika det2x2 nigdy nie sklasyfikowała poprawnie wszystkich punktów z zestawu tab4, to była znacznie stabilniejsza od alternatywy- nawet dla $\varepsilon = 0$ poprawnie sklasyfikowała aż ok. 73% punktów. Niestety stabilność ta związana jest również z faktem, że zwiększenie tolerancji nieznacznie zwiększyło liczbę poprawnie sklasyfikowanych punktów (do ok. 85% dla $\varepsilon = 10^{-12}$), co czyni ten wyznacznik gorszym w większości praktycznych zastosowań od det3x3. Wspomnianej stabilności nie zaobserwowano dla numpy_det2x2- biblioteczna wersja wyznacznika 2x2 dla wszystkich sprawdzanych tolerancji błędu poza $\varepsilon = 10^{-12}$ dawał w większości błędne wyniki.

Na wykresach ilustrujących różnice w klasyfikacji przy użyciu różnych wyznaczników (Wykresy 24- 29) można zaobserwować, że ilość różnic zwiększa się wraz ze zwiększeniem wartości bezwzględnej współrzędnych punktów. Pozwala to postawić tezę, że niezależnie od tolerancji wszystkie wyznaczniki działają najlepiej dla punktów o mniejszych (bezwzględnie) współrzędnych. Tezę tą potwierdzają wykresy 11, 16, 18, 19 i 20.

Wyniki uzyskane dla danych reprezentowanych przez float32 były zdecydowanie gorsze, od tych uzyskanych dla danych reprezentowanych przez float64. Mimo tego dla zestawu tab4_32 prawdziwe były następujące zależności prawdziwe również dla zestawu tab4:

- Wyznaczniki biblioteczne sprawdzały się gorzej od napisanych samodzielnie
- Wyznacznik det3x3 nadal był najlepszy dla tolerancji $\varepsilon \geq 10^{-14}$ (Co wcale nie oznacza, że był dobry).
- Na ogół najgorzej spisywał się wyznacznik numpy_det2x2
- Wyznacznik det2x2 był stabilny (na tle pozostałych).