Oscar Teeninga

Badanie efektywności programu równoległego

1. Wstęp

A. Pomiary

Zrobiłem ~30 pomiarów, natomiast zdarzały się przypadki pomiarów bardzo zaburzonych, które odrzucałem oraz przypadki nieprawidłowego importu bibliotek, przez co zadanie w pliku test.sh nie zostawało poprawnie ewaluowane. Ostatecznie pozostało 15 stosunkowo poprawnie zarejestrowanych prób.

B. Sposób obliczenia przyśpieszenia

Korzystałem z przyśpieszenia względnego, ponieważ w naszym przypadku algorytm sekwencyjny i równoległy na 1 procesorze są tożsame z dokładnością do redukcji, której czas wykonania nie powinien zaburzać wyników. Natomiast w kontekście naszego problemu moglibyśmy znaleźć znacznie lepszy algorytm sekwencyjny obliczający liczbę pi za pomocą szeregu (który z resztą również dałoby zmodyfikować do algorytmu równoległego na podobnej zasadzie co metoda Monte Carlo, jednak nie narzuca się to aż tak wyraźnie).

$$\pi = 4 \cdot \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2 \cdot n - 1} = 4 \cdot \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots \right)$$

Również dostępna jest Metoda Gauss-Legendre, który algorytm sekwencyjny bazujący na ciągach jest trudny do obliczenia współbieżnego, ze względu na to, że kolejne iteracje bazują na poprzednich, natomiast po 25 iteracjach pozwala wyznaczyć 45 milionów cyfr po przecinku, co oznacza, że znacząco efektywniej oblicza liczbę pi.

C. Problemy

Nie napotkałem wiekszych problemów w dalszej cześci realizacji zadania.

2. Skalowanie silne

Zdecydowałem się na trzy rozmiary problemu:

- o 20000000000
- o 447213595
- o 10000000

A. Zależność czasu od liczby procesorów

- W przypadku tych wykresów połączenie ich w jeden byłoby bezsensowne ze względu różnice o wielkości kilku rzędów.
- Wszystkie wykresy są niemal idealnie odzwierciedlają teoretyczną krzywą T(n) = T(1)/n.
- Wydaje się, że dla różnych rozmiarów problemu względna odległość teoretycznej krzywej od pomiarów jest podobna.

B. Zależność przyśpieszenia od liczby procesorów

- W tym przypadku przedstawiłem wszystkie pomiary na jednym wykresie.
- Wraz ze wzrostem liczby procesorów rośnie przyśpieszenie w oczekiwany sposób.
- Wyniki pokrywają się z oczekiwaniami dla każdego rozmiaru problemu.
- Osiągnąłem przyśpieszenia nieco gorsze, niż gdyby przyśpieszenie rosło liniowo względem liczby procesorów.
- Wydaje się, że rozmiar pośredni osiąga najwyższe przyśpieszenie dla dwunastu procesorów, natomiast zakładam, że jest to przypadek.
- Odchylenie standardowe zgodnie z oczekiwaniami największe dla najmniejszego rozmiaru problemu.

C. Zależność efektywności od liczby procesorów

- Zgodnie z oczekiwaniami, efektywność spada ze wzrostem liczby procesorów dla każdego rozmiaru problemu.
- Najmniejsze odchylenie osiągamy zgodnie z intuicją dla problemu największego, a największe dla najmniejszego, a średnie do średniego.

D. Zależność części sekwencyjnej od liczby procesorów

- Część sekwencyjna wraz ze wzrostem procesorów spada dla każdego rozmiaru problemu, co jest oczekiwane
- · Najmniejsze odchylenie osiągamy zgodnie z intuicją dla problemu największego, a największe dla najmniejszego, a średnie do średniego.
- · Ciekawą tendencją jest fakt, że dla rozmiaru średniej wielkości otrzymujemy wolniejsze zmniejszenie się części sekwencyjnej.

3. Skalowanie słabe

Zdecydowałem się na trzy rozmiary problemu (dla jednego procesora):

- 200000000
- 447213595
- 10000000

A. Zależność czasu od liczby procesorów

- W przypadku tych wykresów połączenie ich w jeden byłoby bezsensowne ze względu różnice o wielkości kilku rzędów.
- Wszystkie wykresy są niemal idealnie odzwierciedlają teoretyczną krzywą T(n) = T(1)/n.
- Wydaje się, że dla różnych wielkości względna odległość teoretycznej krzywej od pomiarów jest podobna.

B. Zależność przyśpieszenia od liczby procesorów

- Mniejszy rozmiar problemu wykazuje się mniejszym odchyleniem standardowym, co jest przeciwną tendencją względem skalowania silnego.
- Wydaje się, że dla różnych rozmiarów problemu względna odległość teoretycznej krzywej od pomiarów jest podobna.
- Wszystkie obliczony speedup jest wiec niezależny od rozmiaru problemu, co oznacza, że cześć sekwencyjna nie zależy od rozmiaru problemu.

C. Zależność efektywności od liczby procesorów

- Mniejszy rozmiar problemu wykazuje się mniejszym odchyleniem standardowym, co jest przeciwną tendencją względem skalowania silnego.
- Najlepszą efektywność otrzymałem dla rozmiaru pośredniego, natomiast nie jestem przekonany, że jest to nieprzypadkowy rezultat.
- Nieintuicyjny jest fakt, że efektywność wykazuje się największym odchyleniem standardowym dla największego rozmiaru problemu.

D. Zależność części sekwencyjnej od liczby procesorów

- Wykres jest znacznie mniej stromy względem skalowania silnego.
- Nieintuicyjny jest fakt, że rozmiar części sekwencyjnej wykazuje się największym odchyleniem standardowym dla największego rozmiaru problemu.

4. Porównanie skalowania silnego do słabego

A. Przyśpieszenia od liczby procesorów

• Zdaje się, że dla obu skalowań otrzymaliśmy podobne przyśpieszenie ~11 dla 12 procesorów.

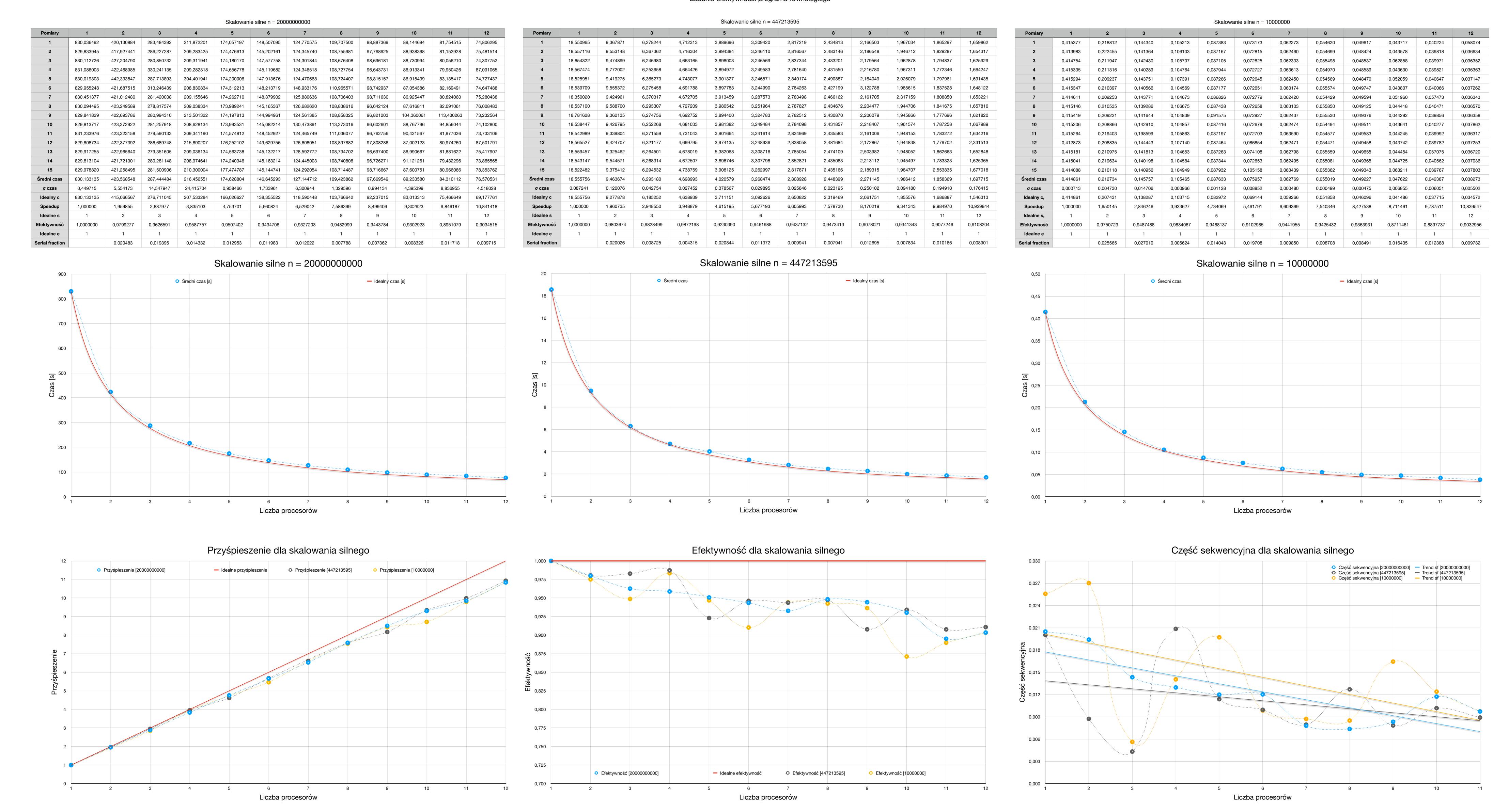
B. Efektywności od liczby procesorów

Zdaje się, że dla obu skalowań otrzymaliśmy podobną efektywność ~90% dla 12 procesorów.

C. Części sekwencyjnej od liczby procesorów

- Można dostrzec, że w przypadku skalowania słabego spadek części sekwencyjnej jest nieco mniej gwałtowny.
- Aproksymowana funkcja liniowa jest bardziej pozioma dla skalowania słabego, co jest oczekiwane.
- Oznacza to, że część sekwencyjna jest zależna w większym stopniu od wielkości problemu, a jeżeli od liczby procesorów, bądź otrzymałem taki rezultat przypadkowo. Wydaje mi się, że wraz ze zwiększeniem problemu rośnie prawdopodobieństwo tego, że jeden węzeł napotka trudności losowe, przez co rootowy będzie musiał oczekiwać na jego zakończenie i stąd takie wyniki (gdyż takie oczekiwanie jest wówczas interpretowane jako część sekwencyjna), a sama część sekwencyjna jest tak naprawdę zaniedbywalna (sprowadza się do redukcji).
- Teoretycznie części sekwencyjnej w naszym programie nie uświadczymy (prócz redukcji), co oznacza, że wykres wraz ze wzrostem rozmiaru problemu będzie zbliżać się do osi x.
- · Gdyby istniała jakaś część sekwencyjna byłoby to obserwowane poprzez stałą funkcję części sekwencyjnej, a więc otrzymany rezultat jest oczekiwany.
- Dla rozmiaru średniego, skalowanie silne oraz skalowanie słabe wydaje się mieć odmienną tendencję wzrostową względem dwóch innych rozmiarów problemu. Natomiast dla skalowania silnego tendencja jest taka, że część sekwencyjna jest zależna od liczby procesorów słabiej, a jeżeli w skalowaniu słabym, gdzie widzimy szybszą tendencję spadkową względem innych rozmiarów problemu. Różnica zdaje się być wynikiem przypadku.

Oscar Teeninga Badanie efektywności programu równoległego



Oscar Teeninga Badanie efektywności programu równoległego

