

## MOWNiT – Sprawozdanie 6

### Rozwiązywanie układów równań liniowych metodami bezpośrednimi

#### Problem 1:

Dany jest układ równań  $\mathbf{Ax}=\mathbf{b}$ .

Elementy macierzy  $\mathbf{A}$  o wymiarze  $n \times n$  są określone wzorem:

$$\begin{cases} a_{1j} = 1 \\ a_{ij} = \frac{1}{i+j-1} \quad \text{dla } i \neq 1 \end{cases} \quad i, j = 1, \dots, n$$

Przyjmij wektor  $\mathbf{x}$  jako dowolną  $n$ -elementową permutację ze zbioru  $\{1, -1\}$  i oblicz wektor  $\mathbf{b}$ .

Następnie metodą eliminacji Gaussa rozwiąż układ równań liniowych  $\mathbf{Ax}=\mathbf{b}$  (przyjmując jako niewiadomą wektor  $\mathbf{x}$ ). Przyjmij różną precyzję dla znanych wartości macierzy  $\mathbf{A}$  i wektora  $\mathbf{b}$ . Sprawdź, jak błędy zaokrągleń zaburzają rozwiązanie dla różnych rozmiarów układu (porównaj – zgodnie z wybraną normą – wektory  $\mathbf{x}$  obliczony z  $\mathbf{x}$  zadany). Przeprowadź eksperymenty dla różnych rozmiarów układu.

Eksperymenty przeprowadzono kolejno dla rozmiaru układów  $n=3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20$

Porównanie odbywa się przy użyciu normy wektorów (błąd maksymalny) wektora wejściowego  $\mathbf{X}$  oraz wektora z wyliczonymi wartościami  $\mathbf{X}$ :  $|x_{\text{zadany}} - x_{\text{wyliczone}}|$

Obliczenia wykonano z dwoma precyzjami: float32 oraz float64

<b>n</b>	<b>float32</b>	<b>float64</b>
3	3.3453E-06	0.0000E+00
4	6.6465E-15	3.0187E-13
5	3.3695E-04	9.2294E-12
6	5.4410E-11	3.6380E-10
7	2.2192E+00	1.3609E-08
8	1.9722E+01	1.2033E-07
9	4.9479E+01	5.4005E-07
10	1.8157E+01	1.6620E-04
11	1.3779E+01	1.2234E-02
12	4.6926E+00	1.2140E+00
13	8.3865E+00	2.1215E+01
14	5.7072E+00	2.1116E+01
15	1.3535E+01	1.5049E+01
16	4.8935E+01	2.5943E+01
17	8.2082E+01	2.2985E+01
18	7.3382E+01	2.2199E+01
19	9.4890E+01	8.5366E+01
20	2.6526E+02	8.7140E+02

Tabela 1 – Błąd maksymalny wektorów X zadanego oraz wyliczonego dla float32 oraz float64

Możemy zauważyć, że błędy zwiększają się wraz z rosnącym rozmiarem układu. Również można zaobserwować, że błąd jest różny dla różnych precyzji. Dla typów float32 błędy te są znacznie większe niż błędy dla typu float64

## Problem 2:

Powtórz eksperyment dla macierzy zadanej wzorem:

$$\begin{cases} a_{ij} = \frac{2i}{j} & \text{dla } j \geq i \\ a_{ij} = a_{ji} & \text{dla } j < i \end{cases} \quad i, j = 1, \dots, n$$

Porównaj wyniki z tym, co otrzymano w przypadku układu z punktu 1). Spróbuj uzasadnić, skąd biorą się różnice w wynikach. Sprawdź uwarunkowanie obu układów.

Ekspertymenty przeprowadzono kolejno dla rozmiaru układów  $n=3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 50, 100, 200$

Porównanie odbywa się przy użyciu normy wektorów (błąd maksymalny) wektora wejściowego  $X$  oraz wektora z wyliczonymi wartościami  $X$ :  $|x_{\text{zadany}} - x_{\text{wyliczone}}|$

Obliczenia wykonano z dwoma precyzjami: float32 oraz float64

n	float32	float64
3	6.8624E-08	3.1402E-16
4	4.4427E-08	2.4825E-16
5	1.0750E-07	4.1541E-16
6	1.5007E-07	9.7422E-16
7	2.0906E-07	1.6947E-15
8	1.3936E-06	4.6722E-15
9	1.3974E-06	3.3103E-15
10	1.4024E-06	3.0827E-15
11	1.5829E-06	4.4214E-15
12	4.8796E-06	1.9804E-14
13	5.4146E-06	2.2008E-14
14	5.4561E-06	2.2768E-14
15	5.6833E-06	2.8361E-14
16	8.6166E-06	3.8010E-14
17	8.6173E-06	3.7162E-14
18	8.7656E-06	3.6520E-14
19	8.9556E-06	3.8893E-14
20	1.2632E-05	3.8090E-14
50	1.1304E-04	3.4606E-13
100	9.0097E-04	2.9341E-12
200	5.4965E-03	2.5560E-11

Tabela 2 – Błąd maksymalny wektorów  $X$  zadanego oraz wyliczonego dla float32 oraz float64

Ponownie możemy zauważyć, że błędy zwiększają się wraz z rosnącym rozmiarem układu oraz dla typów float32 błędy są większe niż błędy dla typu float64

**Porównanie błędów z problemu 1 oraz problemu 2:**

<b>n</b>	<b>Punkt 1 - float32</b>	<b>Punkt 2 - float32</b>	<b>Punkt 1 - float64</b>	<b>Punkt 2 - float64</b>
3	3.3453E-06	6.8624E-08	0.0000E+00	3.1402E-16
4	6.6465E-15	4.4427E-08	3.0187E-13	2.4825E-16
5	3.3695E-04	1.0750E-07	9.2294E-12	4.1541E-16
6	5.4410E-11	1.5007E-07	3.6380E-10	9.7422E-16
7	2.2192E+00	2.0906E-07	1.3609E-08	1.6947E-15
8	1.9722E+01	1.3936E-06	1.2033E-07	4.6722E-15
9	4.9479E+01	1.3974E-06	5.4005E-07	3.3103E-15
10	1.8157E+01	1.4024E-06	1.6620E-04	3.0827E-15
11	1.3779E+01	1.5829E-06	1.2234E-02	4.4214E-15
12	4.6926E+00	4.8796E-06	1.2140E+00	1.9804E-14
13	8.3865E+00	5.4146E-06	2.1215E+01	2.2008E-14
14	5.7072E+00	5.4561E-06	2.1116E+01	2.2768E-14
15	1.3535E+01	5.6833E-06	1.5049E+01	2.8361E-14
16	4.8935E+01	8.6166E-06	2.5943E+01	3.8010E-14
17	8.2082E+01	8.6173E-06	2.2985E+01	3.7162E-14
18	7.3382E+01	8.7656E-06	2.2199E+01	3.6520E-14
19	9.4890E+01	8.9556E-06	8.5366E+01	3.8893E-14
20	2.6526E+02	1.2632E-05	8.7140E+02	3.8090E-14

Tabela 3 - Porównanie błędów z problemu pierwszego oraz drugiego dla precyzji float32 oraz float64

### Uwarunkowanie układów z problemu 1 i 2:

Wskaźnik uwarunkowania macierzy jest miarą jak bardzo zmieni się rozwiązanie  $x$  układu równań w stosunku do zmiany  $b$ . Jeżeli wskaźnik macierzy jest duży, to nawet mały błąd  $b$  może spowodować duże błędy w  $x$ . Wskaźnik uwarunkowania macierzy obliczamy ze wzoru:

$$\kappa = ||A^{-1}|| \cdot ||A||$$

n	Problem 1 wskaźnik uwarunkowania	Problem 2 wskaźnik uwarunkowania
3	2.1600E+02	1.4444
4	2.8800E+03	1.8333
5	2.8000E+04	2.2333
6	2.2680E+05	2.6444
7	1.6299E+06	3.0317
8	1.2862E+07	3.4484
9	1.1200E+08	3.8492
10	8.8414E+08	4.2492
11	6.4738E+09	4.6594
12	4.4079E+10	5.0552
13	1.3477E+11	5.4655
14	2.4592E+11	5.8689
15	1.7333E+11	6.2689
20	4.0039E+11	8.2896

Tabela 4 – wskaźnik uwarunkowania macierzy

Z powyższej tabeli możemy zauważyć, że wskaźnik uwarunkowania dla macierzy z problemu 1 jest znacznie większy od problemu 2. Oznacza to, że niewielki błąd znacznie wpływa na wynik.

### Wnioski:

W implementacji z Problemu 1 algorytm eliminacji Gaussa zawsze jako element wiodący wybiera kolejne elementy przekątnej macierzy. Jest to jedna z przyczyn złych wyników.

Możemy zauważyć, że wskaźniki uwarunkowania dla macierzy w problemie 1 są znacznie większe od wskaźników uwarunkowania w problemie 2. Oznacza to, że niewielki błąd znacznie wpływa na wyniki.

### Problem 3

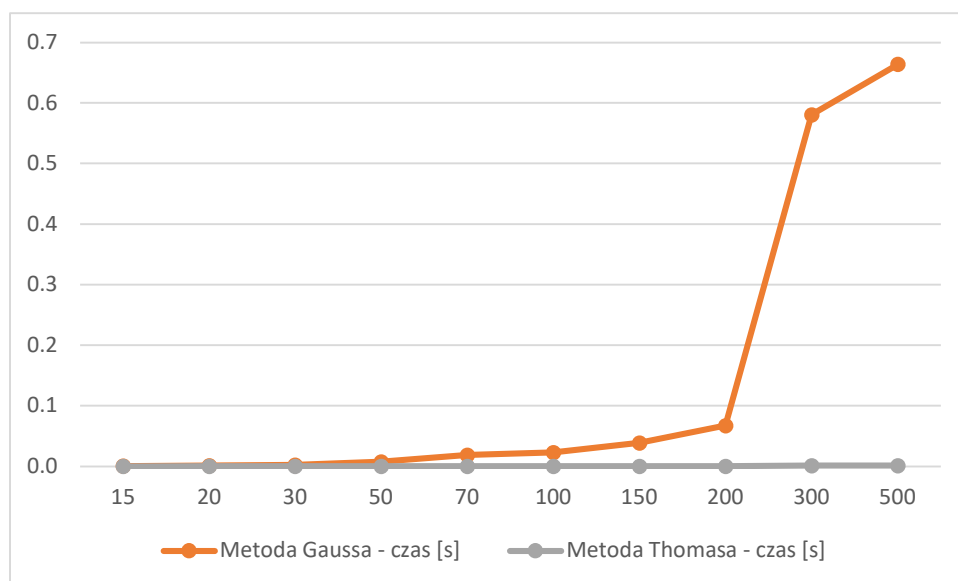
Powtórz eksperyment dla jednej z macierzy zadanej wzorem poniżej (macierz i parametry podane w zadaniu indywidualnym). Następnie rozwiąż układ metodą przeznaczoną do rozwiązywania układów z macierzą trójdagonalną. Porównaj wyniki otrzymane dwoma metodami (czas, dokładność obliczeń i zajętość pamięci) dla różnych rozmiarów układu. Przy porównywaniu czasów należy pominąć czas tworzenia układu. Opisz, jak w metodzie dla układów z macierzą trójdagonalną przechowywano i wykorzystywano macierz A.

$$\begin{cases} a_{i,i} = -2 * i - 4 \\ a_{i,i+1} = i \\ a_{i,i-1} = \frac{2}{i} \quad , \quad i > 1 \\ a_{i,j} = 0 \text{ , dla } j < i - 1 \text{ oraz } j > i + 1 \end{cases}$$

Dla  $i, j = 1, \dots, n$

n	Metoda Gaussa - błąd	Metoda Thomasa - błąd	Metoda Gaussa - czas [s]	Metoda Thomasa - czas [s]
3	3.1402E-16	3.1402E-16	0.000365	0.000021
4	3.1402E-16	3.1402E-16	0.000152	0.000025
5	4.0030E-16	4.0030E-16	0.000181	0.000025
6	3.5108E-16	3.5108E-16	0.000658	0.000028
7	3.5108E-16	3.5108E-16	0.000200	0.000030
8	3.5108E-16	3.5108E-16	0.000304	0.000033
9	3.5108E-16	3.5108E-16	0.000274	0.000033
10	3.5108E-16	3.5108E-16	0.000537	0.000037
11	4.7103E-16	4.7103E-16	0.000423	0.000043
12	4.7103E-16	4.7103E-16	0.000541	0.000049
13	4.7103E-16	4.7103E-16	0.000552	0.000044
14	5.6610E-16	5.6610E-16	0.000572	0.000046
15	4.8394E-16	4.8394E-16	0.000668	0.000050
20	4.8394E-16	4.8394E-16	0.001184	0.000080
30	6.1815E-16	6.1815E-16	0.002588	0.000113
50	9.2888E-16	9.2888E-16	0.007533	0.000185
70	1.0115E-15	1.0115E-15	0.019005	0.000144
100	1.3688E-15	1.3688E-15	0.022827	0.000194
150	1.9004E-15	1.9004E-15	0.038782	0.000284
200	2.1812E-15	2.1812E-15	0.067442	0.000452
300	2.5847E-15	2.5847E-15	0.580591	0.001191
500	3.3766E-15	3.3766E-15	0.664062	0.001415

Tabela 5 – Błędy oraz czasy otrzymane w problemie trzecim dla algorytmu Gaussa oraz Thomasa



Wykres 1 Czas w zależności od liczby punktów dla konkretnej metody

## Wnioski

Z Tabeli 5 możemy zaobserwować, że niezależnie której metody użyjemy to błędy (norm) są takie same

Metoda Thomasa ogranicza się tylko do działania na elementach trójdzielności. Porównując czasy działania to metoda Thomasa jest znacznie szybsza od metody eliminacji Gaussa. Nie jest to jednak zaskakujące, ponieważ wykonuje ona znacznie mniej operacji, ograniczając się tylko do trójdzielności macierzy.

Metoda Gaussa jest skutecznym i prostym sposobem na rozwiązywanie układów równań liniowych. W niektórych przypadkach jest ona jednak wolna z powodu błędów dokładności, które są spowodowane słabym uwarunkowaniem układów wejściowych lub sposobu wybierania elementu wejściowego.

W przypadku, gdy macierz jest trójdzielna, to warto stosować zamiast metody eliminacji Gaussa metodę Thomasa. Jest to uproszczona wersja algorytmu Gaussa, która daje wyniki o tej samej dokładności jednak działa znacznie szybciej.