Metoda iteracyjna Schulza do odwracania macierzy kwadratowej.

Kamil Michalak

Analiza Numeryczna (M) - Zadanie P3.12

1 Wstęp

1.1 Treść zadania

Niech A będzie macierzą nieosobliwą oraz niech $\{X_k\}$ dla $k = 0, 1, ...$ będzie ciągiem macierzy spełniających

$$X_{k+1} = X_k + X_k(I - AX_k),$$

gdzie I jest macierzą jednostkową. Udowodnij, że:

a) przy założeniu $||I - AX_0|| < 1$, gdzie $||\cdot||$ jest normą macierzową indukowaną przez normę wektorową, zachodzi zbieżność $\{X_k\}$ do A^{-1}. Ponadto, dla $E_k := I - AX_k$, pokaż że $Ek + 1 = E_kE_k$;

b) Powyższa metoda jest lokalnie zbieżna kwadratowo

c) Przy założeniu $AX_0 = X_0A$ zachodzi $AX_k = X_kA$ dla $k \geqslant 0$

Na podstawie wybranych macierzy A, sprawdź działanie powyższej metody iteracyjnej Schulza w praktyce. W jaki sposób wybrać X_0? Czy metoda jest szybsza niż znane metody bezpośrednie odwracania macierzy?

1.2 Dane techniczne

Wszystkie programy wykorzystywane w eksperymentach zaprogramowałem w języku Julia, w wersji 1.0.1. Do obliczeń wykorzystywałem liczby zmiennoprzecinkowe Float64. Obliczenia przeprowadziłem na komputerze z procesorem Intel Core i5 i 8 GB pamięci RAM.

2 Normy - definicje

2.1 Norma wektorowa

Definicja 1. ***Normą wektorową*** *nazywamy nieujemną funkcję rzeczywistą* $||\cdot||$, *określoną w przestrzeni* R^n, *o następujących własnościach:*

$$\forall_{x\in R^n\setminus\{\theta\}}||x|| > 0;$$

$$\forall_{x\in R^n}\forall_{\alpha\in R}||\alpha x|| = ||\alpha||||x||$$

$$\forall_{x,y\in R^n}||x+y|| \leqslant ||x|| + ||y||$$

Najpopularniejsze normy wektorowe są definiowane następująco ($x = [x_1, x_2, ..., x_n]^T$)

$$||x||_1 = |x_1| + ... + |x_n|$$

$$||x||_2 = \sqrt{x_1^2 + ... + x_n^2}$$

$$||x||_\infty = \max_{1\leqslant i\leqslant n} |x_i|$$

2.2 Norma macierzowa

Definicja 2. ***Normą macierzy*** *nazywamy nieujemną funkcję rzeczywistą* $||\cdot||$, *określoną w przestrzeni liniowej* $R^{n\times n}$ *wszystkich macierzy kwadratowych stopnia* n, *o następujących własnościach:*

$$\forall_{A\in R^{n\times n}\setminus\{\Theta\}}||A|| > 0$$

$$\forall_{A\in R^{n\times n}}\forall_{\alpha\in R}||\alpha A|| = ||\alpha||||A||$$

$$\forall_{A,B\in R^{n\times n}}||A+B|| \leqslant ||A|| + ||B||$$

$$\forall_{A,B\in R^{n\times n}}||AB|| \leqslant ||A||||B||$$

Norma macierzowa jest ***indukowana*** *przez normę wetkorową jeśli zachodzi równość*

$$||A|| = \sup_{x\in R^n\setminus\{\theta\}} \frac{||Ax||}{||x||}.$$

Najpopularniejsze normy indukowane to

$$||A||_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{n} |a_{i,j}|$$

$$||A||_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |a_{i,j}|$$

$$||A||_2 = \sqrt{\varrho(A^TA)}$$

Gdzie $\varrho(A^TA)$ jest największą wartością własną macierzy A^TA (λ jest wartością własną macierzy M jeśli istnieje wektor v taki, że $Mv = \lambda v$).

3 Dowody twierdzeń

Twierdzenie 1. *Jeśli* $||I - AX_0|| < 1$ *to metoda Schulza jest zbieżna do* A^{-1}

Dowód: Niech $E_k = I - AX_k$. Wtedy

$$\begin{aligned} E_{k+1} = I - AX_{k+1} = I - A[X_k + X_k(I - AX_k)] = \\ = I - AX_k - AX_k(I - AX_k) = I - AX_k - AX_kE_k = \\ = E_k - AX_kE_k = E_k(I - AX_k) = E_k^2 \end{aligned}$$

$E_{k+1} = E_k^2$ oraz $||E_0|| < 1$ Zatem

$$||E_0|| > ||E_1|| > ... > ||E_k|| > ||E_{k+1}|| > ...$$

Więc

$$\lim_{k\to\infty} ||E_k|| = 0$$

$$\lim_{k\to\infty} ||I - AX_k|| = 0$$

$$\lim_{k\to\infty} AX_k = I$$

$$\lim_{k\to\infty} X_k = A^{-1}$$

Co kończy dowód.

Twierdzenie 2. *Niech* $E_k = I - AX_k$. *Wtedy* $E_{k+1} = E_kE_k$.

Dowód:

$$\begin{aligned} E_kE_k = (I - AX_k)(I - AX_k) = I - AX_k - AX_k(I - AX_k) = \\ = I - A(X_k + X_k(I - AX_k)) = I - AX_{k+1} = E_{k+1} \end{aligned}$$

Twierdzenie 3. *Metoda Schulza jest lokalnie zbieżna kwadratowo.*

Dowód: Niech $\varepsilon_k = A^{-1} - X_k$. Wtedy zachodzi

$$E_{k+1} = E_k^2$$

$$I - AX_{k+1} = (I - AX_k)^2$$

$$A\varepsilon_{k+1} = (A\varepsilon_k)^2$$

$$\varepsilon_{k+1} = \varepsilon_k A \varepsilon_k$$

$$||\varepsilon_{k+1}|| \leqslant ||A||||\varepsilon_k||^2$$

Zatem

$$\lim_{k\to\infty} \frac{||X_{k+1} - A^{-1}||}{||X_k - A^{-1}||} = C,$$

gdzie $C \in (0, ||A||]$, co kończy dowód.

Twierdzenie 4. *Przy założeniu* $AX_0 = X_0A$ *zachodzi* $AX_k = X_kA$ *dla* $k \geqslant 0$

Dowód: Przeprowadźmy dowód indukcyjny. Podstawa indukcji zawiera się w założeniu twierdzenia. Załóżmy, że dla pewnego k twierdzenie zachodzi. Wtedy dla $k+1$ mamy

$$\begin{aligned}AX_{k+1} &= A(X_k + X_k(I - AX_k)) = AX_k + AX_k(I - AX_k) = \\ &= X_kA + X_kA(I - AX_k) = X_kA + X_kA + X_kAAX_k = \\ &= X_kA + X_kA + X_kAX_kA = X_kA + X_k(A - AX_kA) = \\ &\qquad = (X_k + X_k(I - AX_k))A = X_{k+1}A\end{aligned}$$

Co kończy dowód.

4 Implementacja metody Schulza w Julii

Implementacja metody wygląda następująco (wersja dla określonej liczby iteracji).

```
function Schulz(A,X0,n)
    I := one(A)
    X := X0
    for i := 1:n
        X := X + X*(I - A*X)
    return X
```

Gdzie `A` - macierz, której odwrotność znajdujemy, `I`-macierz jednostkowa.

Poniższa wersja oblicza macierz odwrotną z zadaną dokładnością

```
function Schulz2(A,X0,epsilon)
    I := one(A)
        X = X0
        while norm_inf(I-A*X)>epsilon
            X := X + X*(I - A*X)
        return X
```

4.1 Wybór X_0

W [1] Mamy podane różne sposoby wyboru X0. Przykładowo

$$X_0 = \frac{A}{||A||_\infty^2}$$

Z czego wynika

$$\frac{||X_0||_\infty}{||A^{-1}||_\infty} = \frac{||A||_\infty}{||A||_\infty^2 ||A^{-1}||_\infty}$$

Ponieważ $||A||||A^{-1}|| \geqslant ||I|| = 1$, więc

$$\frac{||X_0||_\infty}{||A^{-1}||_\infty} = \frac{1}{||A||_\infty ||A^{-1}||_\infty} \leqslant 1$$

Co sugeruje, że dla $X_0 = \frac{A}{||A||_\infty^2}$ metoda będzie prawie zawsze zbieżna.

Inne możliwe sposoby wyboru X_0:

- $\frac{A}{||A||_\infty^2}$
- $\frac{A^T}{n||A||_1||A||_\infty}$
- $\frac{A^T}{||A||_2^2}$
- $diag(\frac{1}{a_{1,1}}, \frac{1}{a_{2,2}}, ..., \frac{1}{a_{n,n}})$

4.2 Działanie metody na przykładowej macierzy.

Weźmy macierz A.

$$A = \begin{bmatrix} 1.0 & 2.0 \\ 3.0 & 4.0 \end{bmatrix}$$

Niech

$$X_0 = \frac{A^T}{||A||_2^2} = \begin{bmatrix} 0.0334828 & 0.100448 \\ 0.0669656 & 0.133931 \end{bmatrix}$$

Macierz X_k w kolejnych iteracjach prezentuje się następująco.

$$X_2 = \begin{bmatrix} 0.00624849 & 0.112496 \\ 0.0861582 & 0.125441 \end{bmatrix}$$

$$X_5 = \begin{bmatrix} -0.230988 & 0.217442 \\ 0.253343 & 0.0514833 \end{bmatrix}$$

$$X_{10} = \begin{bmatrix} -1.97952 & 0.990938 \\ 1.48556 & -0.493614 \end{bmatrix}$$

$$X_{15} = \begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix} = A^{-1}$$

5 Porównanie metody Schulza z bezpośrednimi metodami odwracania macierzy, testy

5.1 Eliminacja Gaussa - opis metody

Do bezpośredniego odwracania macierzy możemy wykorzystać zmodyfikowaną eliminację Gaussa. Standardowo służy ona do rozwiązywania układów równań $Ax = b$. Problem odwrócenia macierzy $n \times n$ możemy sprowadzić do problemu rozwiązania n układów równań $Ax_i = v_i$ gdzie x_i to i-ta kolumna macierzy A^{-1} a v_i to i-ta kolumna macierzy jednostkowej (z jedynką na i-tej współrzędnej i zerami w pozostałych miejscach).

Definicja 3 (Macierz trójkątna górna (dolna))**.** *Macierz $A = [a_{i,j}] \in R^{n\times n}$ nazywamy trójkątną górną (dolną), jeśli $a_{i,j}$=0 dla $i > j$ $(i < j)$*

Definicja 4 (Macierz schodkowa)**.** *Macierz schodkowa – macierz, której pierwsze niezerowe elementy kolejnych niezerowych wierszy znajdują się w coraz dalszych kolumnach, a wiersze zerowe umieszczone są najniżej.*

Definicja 5 (Operacje elementarne)**.** *Następujące operacje elementarne przekształcają dany układ w układ do niego równoważny, czyli układ o tym samym zbiorze rozwiązań co wyjściowy:*

- *dodanie do równania innego równania pomnożonego przez liczbę,*
- *zamiana dwóch równań miejscami,*
- *pomnożenie równania przez liczbę różną od zera (w ogólności: odwracalną).*

Żeby obliczyć układ równań $Ax = b$ sprowadzamy macierz $[A|b]$ do postaci górnotrójkątnej za pomocą operacji elementarnych na wierszach i rozwiązujemy kolejne równania od dołu macierzy. Żeby obliczyć macierz A^{-1} sprowadzamy do postaci schodkowej macierz $[A|I]$ i rozwiązujemy n prostych układów równań.

5.2 Implementacja metody eliminacji Gaussa

```
function Gauss(M)
    rows := size(M,1)
    factor := 1
    A := [M one(M)] //one(M) - macierz jednostkowa o wymiarach M
    for j := 1,...,rows
        for i := j+1,...,rows
            factor = -A[i,j]/A[j,j]
            for k:=1,...,2*rows
```

```
            A[i,k] := A[i,k] + A[j,k] * factor
U := A[:,1:rows]
I := A[:,(rows+1):end]
res = zero(I) //zerowa macierz n x n
for j := 1...rows
    for i := rows,rows-1,...,1
        res[i,j] = I[i,j]
        for k = (i+1),...,rows
            res[i,j] -= U[i,k]*res[k,j]
        res[i,j] /=U[i,i]
return res
```

5.3 Testy

Losowa macierz M rozmiaru 1000x1000 z wartościami z przedziału (0,1], liczba iteracji i dokładność w funkcjach `Schulz()` i `Schulz2` dobrane tak, żeby ich dokładność była podobna do dokładności funkcji `gauss()` ($||AX_{end}||_\infty \in (1-10^{-6}, 1+10^{-6})$, $X_0 = \frac{M^T}{||M||_2^2}$

- Czas wykonania procedury `Gauss()` - 21,17s
- Czas wykonania procedury `Schulz()` - 3,71s
- Czas wykonania procedury `Schulz2()` - 5,57s

Macierz 100 x 100 z wartościami z przedziału (0,100).$X_0 = \frac{M^T}{n||M||_1||M||_\infty}$

- Czas wykonania procedury `Gauss(M)` - 0.023s
- Czas wykonania procedury `Schulz(M,X0,50)` - 0.019s
- Czas wykonania procedury `Schulz2()` - 0.032s

Dla małych macierzy metoda Schulza jest porównywalnie szybka do metody Gaussa.

6 Wnioski

Metoda Schulza to bardzo ciekawa metoda odwracania macierzy. Jak widać po testach umożliwia szybkie obliczenie macierzy odwrotnej z dużą dokładnością. Dla dużych macierzy jest zdecydowania szybsza niż metoda eliminacji Gaussa, więc prawdopodobnie jest szybsza od zdecydowanej większości metod bezpośredniego odwracania macierzy. Jest również łatwa do zaimplementowania. Jedynym problemem w tej metodzie może być wyznaczenie macierzy X_0, ale dla większości macierzy powinien się sprawdzić jeden z podanych wyżej sposobów. Warto stosować metodę Schulza do odwracania dużych macierzy.

7 Źródła

1. F. Soleymani, Predrag S. Stanimirović "A Higher Order Iterative Method for Computing the Drazin Inverse", The Scientific World Journal, Volume 2013, Article ID 708647, 11 pages