Notatki z Metod numerycznych

Jacek Olczyk

October 2018

Część I
Wykład

1 Rozwiązywanie układów równań liniowych

Znane metody

- $Ax = b, A \in \mathbb{R}^{n \times n}$
- Algorytm rokładem LU (elim. Gaussa) z wybraniem el. gł $O(\frac{2}{3}N^3)$.
- 1. złota myśl numeryka: Co zrobić jeśli zadanie jest za trudne? Zmienić zadanie
- Zamiast rozwiązywać układ równań, przybliżamy go
- Czy da się szybciej niż Gauss, który jest $O(n^3)$? To jak nisko da się zejść to problem otwarty, ale istnieją algorytmy lepsze niż sześcian.

2 Przybliżone rozwiązywanie układów równań

- Niech $A = M - Z$, wtedy $Ax = Mx - Zx = b$, zatem $Mx = Zx + b$
- TODO Metoda iteracji prostej Banacha $Mx_{n+1} = Zx_n + b$
- Jeśli wybierzemy M tak, by układ z macierzą M można było tanio rozwiązać, wtedy iteracja też będzie tania
- Chcemy, żeby M było dobrym przybliżeniem A, ale nie aż tak łatwo że
- Metoda Jacobiego
$$a_{kk}x_k^{n+1} = b - \sum_{j \neq k} a_{kj}x_j^n$$
- inny pomysł, metoda Gaussa-Seidela: $a_{kk}x_k^{(n+1)} = b_k - \sum_{j<k} a_{kj}x_j^{(n+1)} - \sum_{j>k} a_{kj}x_j^{(n)}$

- Uwaga: fakt życiowy. Gdy n jest bardzo duże, wówczas w A jest zazwyczaj bardzo dużo zer, o ile układ pochodzi z REAL LIFE^{TM}.
- To oznacza, że ilość elementów różnych od 0 jest rzędu $O(n)$. Mówimy wtedy że macierz jest rzadka.
- Wniosek: Jeśli A ma $O(n)$ niezerowych elementów, to mnożenie Ax kosztuje też $O(n)$. Ponadto, rozwiązanie układu z macierzą dolnotrójkątną też jest $O(n)$

3 Normy macierzowe i wektorowe

Normy wektorowe

$$||x||_p := (\sum_{i=1}^{N} |x_i|^p)^{\frac{1}{p}}$$

$$||x||_\infty := \max_i |x_i|$$

Norma macierzowa

$$||A||_p := \max_{x \neq 0} \frac{||Ax||_p}{||x||_p} = \max_{||x||_p} ||Ax||_p$$

Własności normy macierzowej

1. $$||Ax|| \leqslant ||A|| \cdots ||x|| \forall_{x \in \mathbb{R}^n}$$
2. $$||Ax||$$

   - nie dało się przeczytać tablicy
3. tu też coś było :(

4 Warunek wystarczający zbieżności klasycznej metody iteracyjnej ($A = M - Z$)

$$Mx_{k+1} = b + Zx_k \qquad (*)$$

Niech x^* będzie dokładnym rozwiązaniem $Ax^* = b$

$$x_{k+1} = M^{-1}(b - Zx_k)$$

$$x_{k+1} - x^* = M^{-1}(b - Zx_k) - x^*$$
$$= M^{-1}(Ax^* - Zx_k) - x^*$$

$$= M^{-1}(Ax^* - (M-A)x_k) - x^*$$
$$= M^{-1}Ax^* + (I - M^{-1}A)x_k - x^*$$
$$= -(I - M^{-1}A)x^* + (I - M^{-1}A)x_k$$
$$= (I - M^{-1}A)(x_k - x^*)$$

Czyli B pomnożony błąd k-ty.
Czyli $x_{n+1} - x^* = B(x_k - x^*) = B^2(x_{k-1} - x^*) \ldots = B^{k+1}(x_0 - x^*)$

Wniosek: Jeśli $||B|| < 1$, to $(*)$ zbieżna do $x*$ dla dow. $x_0 \in \mathbb{R}^N$

Twierdzenie: Metoda $(*)$ jest zbieżna do x^* z dowolnego x_0 wtw gdy $\rho(B) < 1$ gdzie $\rho(B) = \max\{|\lambda| : \lambda \text{ jest wartością własną } B\}$ - promień spektralny macierzy B Dowód pominięty

Twierdzenie: Jeśli macierz A jest ściśle diagonalnie dominująca, tzn zachodzi $|a_n| > \sum_{j \neq i} |a_{ij}|$ dla $i = 1..N$ to metoda Jacobiego jest zbieżna (dla dowolnych $x_n \in R^n$)

Dowód. Zbadajmy macierz iteracji.

$$||B||_\infty = ||I - M^{-1}A||_\infty$$

$M^{-}1$ dla macierzy diagonalnej to podnoszenie wszystkich elementów do -1.
$M^{-1}A = I +$ macierz z zerami na diagonali i ułamkami na reszcie, pierwszy wiersz to $0, a_{12}/a_{11}, a_{13}/a_{11} \ldots$
Żeby uzyskać B odejmujemy I.
$||B||_\infty = \max_i w_i$
$w_i = \sum_j |b_{i,j}| = \sum_{j \neq i} |a_{ij}/a_{ji}| = \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| < 1$ zatem norma B jest mniejsza od 1 więc normy są zbieżne. □

5 Metody iteracyjne oparte na normalizacji w przestrzeni Kryłowa

k-ta przestrzeń Kryłowa

$$K_k = r_0, Ar_0, ..., A^{k-1}r_0$$

gdzie $r_k := b - Ax_k$ - reszta na k-tej iteracji

Metoda iteracyjna

- $x_k + 1 \in K_k$ przesunięta o x_0
- $x_k + 1$ normalizuje pewną miarę błędu na $x_0 + K_k$
- Na przykład:
$$||x_k - X^*||_C \leqslant ||y - x^*||_C \forall_{y \in x_0 + K_k}$$
lub
$$||r_k|| \leqslant ||b - Ay||_C \forall_{y \in x_0 + K_k}$$
gdzie $C = C^T > 0$

5.1 Metoda gradientów sprzężonych (CG - Conjugate Gradient) dla macierzy $A = A^T > 0$

5.1.1 Fakty o macierzach symetrycznych i dodatnio określonych

Niech $A = A^T > 0$ (symetryczna i dodatnio określona, a co za tym idzie $x^T Ax > 0$ dla $x \neq 0$). Wtedy:

1. Wartości własne są rzeczywiste a wektory własne są ortogonalne (czyli $A = Q\Lambda Q^T$, gdzie Q jest ortogonalna, a Λ jest diagonalna)
2. $||x||_A := \sqrt{x^T Ax}$ określa normę wektorową (norma energetyczna indukowana przez A)

Iterację metody gradientów sprzężonych definiujemy następująco:

$$x_{k+1} \in x_0 + K_k$$

$$||x_{k+1} - x^*||_A \leqslant ||y - x^*||_A \forall_{y \in x_0 + K_k}$$

Ale przecież potrzebujemy mieć rozwiązanie żeby to zrobić!

Fakt. Można stąd wyprowadzić algorytm iteracyjny, który na podstawie kilku poprzednio wyznaczonych wektorów wyznaczy x_{k+1} kosztem jednego mnożenia przez macierz A i $O(N)$

Twierdzenie. Po k iteracjach metody CG błąd $||x_k - x^*||_A \leqslant 2(\frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1})^k ||x_0 - x^*||_A$ gdzie $\alpha = \lambda_{max}(A)/\lambda_{min}(A)$.

6 Zagadnienia własne

Dla $A \in R^{NxN}$ znaleźć parę własną (λ, x), że $Ax = \lambda x$ oraz $x \neq 0$. λ pierwiastkiem wielomianu charakterystycznego: $det(A - \lambda I) = 0$ Gdy $A = A^T$ to wartości i wektory własne rzeczywiste, istnieje Q ortogonalna $A = Q * L * Q^T$ (L to tylko lambdy na przekątnej)

3 podstawowe klasy zadań obliczeniowych dla zagadnień własnych:

1. ekstremalne wartości własne (największa, najmniejsza, etc) i odp. wektory (PageRank)

2. wartości własne bliskie zadanej wartości (wieżowce w Japonii)

3. pełne zadanie własne

Wyznaczanie wektora odpowiadającego dominującej wartości własnej (zakładamy że istnieje dokładnie jedna wartość własna że jej moduł ostro większy od innych modułów)

$$||A_x|| = ||\lambda * x|| = |\lambda| * ||x|| = |\lambda|$$

(bo $||x|| = 1$)

Metoda potęgowa x_0 startowy o normie 1

$$x_{n+1} = A * x_n$$

$$x_{n+1} := x_{n+1}/||x_{n+1}||$$

skąd nazwa:

$$x_{n+1} = A * x_n = A * Ax_{n-1} = A^2 * x_{n-1} = ... = A^{n+1} * x_0$$

nie robić tego w ten sposób, bo A jest duże (ale rzadkie) i będzie coraz mniej rzadkie! Lepiej iteracyjnie, bo tanio mnożyć przez rzadką macierz

Twierdzenie o zbieżności tej metody: Załóżmy, że A diagonalizowalna - istnieje Y nieosobliwe że YAY^{-1} tworzy macierz diagonalną

$A * y_i = \lambda * y_i$ gdzie y_i to kolumna Y

$$x_0 = \sum_1^n \alpha_i * y_i$$

$$x_n = A^n * x_0 = A^{n-1} * (A * x_0) =$$

$$= A^{n-1} * \sum_1^n \alpha_i * y_i =$$

$$= A^{n-1} * \sum_1^n \alpha_i * \lambda_i * y_i =$$

$$= \sum_1^n \alpha_i * \lambda_i^n * y_i =$$

$$= \lambda_1^n * \sum_1^n \alpha_i * ()\lambda_i/\lambda_1)^n * y_i$$

Jeżeli λ_1 dominujące, to $\lambda_i/\lambda_1^n \to 0$ ($\lambda_1 \neq 0$)

Odzyskanie wartości własnej na podstawie przybliżenia (znaleźć takie przybliżenie lambdy że norma przybliżenia $A*x-\lambda*x$ minimalna) - jest to zadanie najmniejszych kwadratów iloraz Rayleigh TODO Transformacje spektrum: 1. Jeżeli λ ww A to $\lambda-\mu$ ww $A-\mu*I$ 2. Jeżeli λ ww A nieosobliwego to $1/(\lambda)$ ww $A^(-1)$

Odwrotna metoda potęgowa na zadania typu 2:

Wartosci wlasne $(A-\mu*I)^{-1}$ to $1/(\lambda-\mu)$ Kiedy największe? Kiedy μ blisko λ_i to wtedy $1/(\lambda_i-\mu$ dominującą ww

RQI raileigh quotient iteration, bardzo szybko zbieżne ale niekoniecznie do najbliższego oryginałowi ww TODO

3. pełny problem - metoda QR TODO

Część II
Ćwiczenia

7 Układy nadokreślone - kontynuacja

7.1 Zadanie 1.

Macierz Hessenberga - to macierz trójkątna górna, z tym że niezerowe elementy mogą być jeden element pod diagonalą.

$$\begin{bmatrix} x & \dots & & & & x \\ x & x & \dots & & & x \\ & x & x & \dots & & x \\ & & \ddots & & & x \\ & & & x & x & x \\ & & & & x & x \end{bmatrix}$$

Jak najmniejszym kosztem znaleźć rozkład QR tej macierzy?
Metodą Householdera? Nie ma jak wykorzystać zer na dole.

7.1.1 Obroty Givensa - przypomnienie

1. G_{ij} - macierz Givensa
2. $b = G_{ij}a$
3. $b_j = 0$
4. $\cos\phi = \frac{a_i}{\sqrt{a_i^2+a_j^2}}$

5. $\sin\phi = \frac{a_j}{\sqrt{a_i^2+a_j^2}}$

7.1.2 Zamiana macierzy Hessenberga w górnotrójkątną obrotami Givensa

$$(G_{ij}a)_j = -\frac{a_i a_j}{\sqrt{a_i^2+a_j^2}} + \frac{a_i a_j}{\sqrt{a_i^2+a_j^2}} = 0$$

$$G_{n-1n}\dots G_{ii+1}\dots G_{12}A = R$$

7.1.3 Koszt

Robimy $n-1$ iteracji.
Dla $G_{i\ i+1}$ trzeba wykonać jeden pierwiastek, $w_i = cw_i + sw_{i+1}$ oraz $w_{i+1} = -sw_i + cw_{i+1}$, łącznie $4(n-1)$ mnożeń.
Wszystko razem: $4\sum_{i=1}^{n-1} n-i = 4\sum_{i=1}^{n-1} i = \frac{4n(n-1)}{2} \sim 2n^2$.

7.2 Zadanie 2.

Dane są punkty $(-1,-1),(0,2),(1,0),(2,1)$.
Znajdź prostą $y = ax+b$ najlepiej przybliżającą te punkty (w sensie LZNK).
Zadane punkty oznaczamy jako (x_i, y_i).
Zatem to, co chcemy zminimalizować to $y(x_i) - y_i$.
Policzmy normę: $\min_{a,b}\sum_{i=1}^{4}(y(x_i)-y_i)^2$
Niewiadome to a oraz b, więc niech:

$$z = \begin{bmatrix} a \\ b \end{bmatrix} \quad d = \begin{bmatrix} y_i \end{bmatrix}_{i=1,2,3,4} \quad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}$$

Teraz wystarczy użyć LZNK aby obliczyć $\min ||Az-d||_2$:

TODO: policzyć to

Uwaga: w ten sposób odległość między punktami a prostą liczymy w pionie, a nie najbliższą (to dobrze, tak działą LZNK).

Uwaga 2: LZNK nie działa dla równania $y = a + e^{bx}$, ale dla $y = a + be^x$ już tak!

8 Normy

Przypomnienie definicji: Norma $||\cdot|| : V \to \mathbb{R}^+$ spełnia następujące warunki:

1. $||u+v|| \leqslant ||u|| + ||v||$
2. $||\alpha v|| = ||av||$
3. $||v|| = 0 \implies v = 0$ - wektor zerowy

p-te normy wektorowe:

$$||x||_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$$

$$||x||_\infty = \max_i |x_i|$$

Normy macierzowe Niech $A \in \mathbb{R}^{n\times n}$.
Macierzowe normy indukowane są postaci

$$||A|| = \sup_{x\neq 0} \frac{||Ax||}{||x||} = \sup_{||x||=1} ||Ax||$$

p-te normy macierzowe

$$||A||_p = \sup_{||x||_p=1} ||Ax||_p, p = 1, 2, \ldots, \infty$$

wszystkie poza $1, 2, \infty$ zwykle się pomija

8.1 Własności norm indukowanych macierzy

1. $||Ax|| \leqslant ||A||||x||$ - z definicji mamy $||A|| \geqslant \frac{||Ax||}{||x||}$
2. $||AB|| \leqslant ||A||||B||, A, B \in \mathbb{R}^{n\times n}$ - bo

$$||ABx|| \leqslant ||A||||Bx|| \leqslant ||A||||B||||x||$$

oraz

$$||AB|| = \sup_{x\neq 0} \frac{||ABx||}{||x||} \leqslant ||A||||B||$$

Fakt. W przestrzeniach skończonego wymiaru wszystkie normy spełniają równanie: $\exists_{c_1,c_2>0}\forall_x c_1||x||_1 \leqslant ||x||2 \leqslant c_2||x||_1$, gdzie normy są dowolne (niekoniecznie pierwsza i druga)

8.2 Zależności między normami

Niech $x \in \mathbb{R}^n$, a normy będą p-te.

$$||x||_1^2 = (\sum_i |x_i|)^2 \geqslant ||x||_2^2$$

$$||x||_1 \leqslant \alpha ||x||_2$$

Jakie α wybrać?

$$||x||_1 \geqslant ||x||_\infty$$

$$n||x||_\infty \geqslant ||x||_1$$

$$||x||_\infty \leqslant ||x||_2$$

$$\sqrt{n}||x||_\infty \geqslant ||x||_2$$

$$||x||_1 \leqslant n||x||_\infty \leqslant n||x||_2$$

Zatem $\alpha = n$.

Nierówności $\frac{1}{n}||A||_2 \leqslant \frac{1}{\sqrt{n}}||A||_\infty \leqslant ||A||_2 \leqslant \sqrt{n}||A||_1 \leqslant n||A||_2$

8.3 Wzory na normy macierzowe

$$||A||_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$$

$$||A||_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

Zatem $||A^T||_1 = ||A||_\infty$.

Norma druga (spektralna)

$$||A||_2 = \max_{\lambda \in \delta(A^T A)} \sqrt{\lambda}$$

Gdzie $\delta(M)$ jest zbiorem wartości własnych macierzy M.
Jeśli Q ortogonalna: $||Q||_2 = 1$, co za tym idzie $||I||_2 = 1$ λ- wartość własna oraz v - wektor własny spełnieją $Av = \lambda v$

Norma Frobeniusa (Euklidesowa)

$$||A||_F = \sqrt{\sum_{i,j} |a_{ij}|}$$

Nie jest normą indukowaną, bo dla wszystkich norm indukowanych zachodzi $||I|| = \sup_{x \neq 0} \frac{||Ix||}{||x||} = 1$, a $||I||_F = \sqrt{n}$ - zatem nie pochodzi od drugiej normy wektorowej!

9 Ćwiczenia 26/10

9.1 Dalsze własności norm

$$||A||_2 = \max_{\lambda \in \delta(A^T A)} \sqrt{\lambda}$$

Jeśli $A = A^T$ to $||A||_2 = \max_{\lambda \in \delta(A)} |\lambda|$.

Twierdzenie. Jeśli λ jest wartością własną A, to λ^2 jest wartością własną A^2.

Dowód. $Av = \lambda v \implies A^2 v = \lambda A v = \lambda^2 v$ □

$$||A||_2 = \sup_{||x||=1} ||Ax||_2$$

$$||Ax||_2 = \sqrt{(Ax)^T Ax} = \sqrt{x^T A^T Ax} = (*)$$

$A^T A$ jest macierzą symetryczną, oraz $A^T A = Q^T \Lambda Q$, gdzie Λ jest macierzą diagonalną $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ oraz Q jest macierzą ortogonalną wektorów własnych.

UZUPEŁNIĆ TODO

10 Ćwiczenia 9/11

10.1 Metoda Richardsona - kontynuacja zadania

Metoda Richardsona - $x_{k+1} = x_k + \tau(b - Ax_k)$ Szukaliśmy parametru τ t. że metoda Richardsona jest zbieżna. A ma wartości własne $\lambda_1, \ldots, \lambda_n > 0$. Jest zbieżna dla $\tau \in (0, \frac{2}{\lambda_{max}})$. Dla jakich τ jest zbieżna najszybciej?

$$\rho(I - \tau A) = \max\{|1 - \tau\lambda_{min}|, |1 - \tau\lambda_{max}|\}$$

Powiedzieliśmy, że

$$||x_{k+1} = x^*|| \leqslant ||I - Q^{-1}A||||x_k - x^*||$$

Zatem szukamy τ realizującego:

$$\arg\min_{\tau \in (0, \frac{2}{\lambda_{max}})} \rho(I - \tau A)$$

Czyli:

$$\arg\min_{\tau \in (0, \frac{2}{\lambda_{max}})} \max\{|1 - \tau\lambda_{min}|, |1 - \tau\lambda_{max}|\}$$

Pierwsz funkcja ma miejsce zerowe w $\frac{1}{\lambda_{min}}$, a druga w $\frac{1}{\lambda_{max}}$. Można narysować obie funkcje, one przecinają się w zwykle dwóch punktach, czyli $|1 - \tau\lambda_{min}| = |1 - \tau\lambda_{max}|$. Zatem albo $\tau = 0$, co daje nam rozbieżność, albo $\tau = \frac{2}{\lambda_{max} + \lambda_{min}}$ Żeby wystartować z metodą, to wystarczyłoby ograniczenie górne na λ_{max}

10.2 Metoda Gaussa-Seidela

$$A = \begin{bmatrix} 2 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 2 \end{bmatrix}$$

Wykaż, że metoda Gaussa-Seidela jest zbieżna dla macierzy A.

Fakt. Metoda Gaussa-Seidela jest zbieżna dla macierzy diagonalnie dominującej ($\forall_i |a_{ii}| > \sum_{j \neq i} |a_{ij}|$). Metoda iteracyjna jest zbieżna, jeśli promień spektralny macierzy jest mniejszy od 1.

Promień spektralny

- $$\rho(I - Q^{-1}A)$$
- kres dolny po wszystkich normach indukowanych:

$$\inf_{||\cdot|| \text{jest indukowana}} ||I - Q^{-1}A||$$

Wystarczy pokazać, że dla pewnej normy indukowanej $||\cdot||$(jakiej?) zachodzi:

$$||I - Q^{-1}A|| < 1$$

Dowód. Tutaj,

$$Q = \begin{bmatrix} 2 & & & & & & \\ -1 & 2 & & & & & \\ & -1 & 2 & & & & \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & -1 & 2 & \\ & & & & & -1 & 2 \end{bmatrix}$$

Korzystamy z tego, że $Q^{-1}Q = I$, i prostym spostrzeżeniem jest:

$$Q = \begin{bmatrix} 2^{-1} & & & & & & \\ 2^{-2} & 2^{-1} & & & & & \\ \vdots & & 2^{-1} & & & & \\ \vdots & & \ddots & \ddots & & & \\ \vdots & & & \ddots & \ddots & & \\ \vdots & & & & & 2^{-1} & \\ 2^{-n} & & & & & 2^{-2} & 2^{-1} \end{bmatrix}$$

Teraz

$$Q^{-1}A = \begin{bmatrix} 1 & -\frac{1}{2} & & & & & \\ 0 & \frac{3}{4} & -\frac{1}{2} & & & & \\ & -2^{-3} & \frac{3}{4} & -\frac{1}{2} & & & \\ & \vdots & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \frac{3}{4} & -\frac{1}{2} \\ 0 & -2^{-n} & \dots & \dots & & -2^{-3} & \frac{3}{4} \end{bmatrix}$$

$$G = I - Q^{-1}A = \begin{bmatrix} 1 & \frac{1}{2} & & & & & \\ 0 & \frac{1}{4} & \frac{1}{2} & & & & \\ & 2^{-3} & \frac{1}{4} & \frac{1}{2} & & & \\ & \vdots & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \frac{1}{4} & \frac{1}{2} \\ 0 & 2^{-n} & \dots & \dots & & 2^{-3} & \frac{1}{4} \end{bmatrix}$$

$||G||_1 = \sum_{i=1}^{n} (\frac{1}{2})^i = 1 - \frac{1}{2^n} < 1$ □

10.3 Błędy numerycznych rozwiązań układów równań

Rozwiązujemy układ $Ax^* = b$, x^* jest dokładnym rozwiązaniem, x - wynik obliczeń numerycznych. Wtedy $x^* = x + A^{-1}(b - Ax) = x + A^{-1}r = x + e$, gdzie r jest wektorem residualnym ($r = b - Ax$), a e - błędem. Rozwiązując układ równań $Ae = r$ otrzymamy poprawkę rozwiązania. Tylko czy to ma sens? Układy z macierzą A już umiemy łatwo rozwiązywać, tylko to wciąż nie będzie idealne, bo znów mamy przybliżenie.

Iteracyjne poprawianie rozwiązań

$$\begin{aligned} x^0 &= x \\ x^{(k+1)} &= x^{(k)} + A^{-1}r^{(k)}, \quad k = 0, 1, \dots \end{aligned}$$

Żeby poprawianie poprawiało, obliczanie wektora residualnego musi być wykonane w jak największej precyzji.

10.4 Wartości i wektory własne

λ, v - para własna dla A, spełnia $Av = \lambda v$

$$||Av|| = ||\lambda v||$$

$$||Av|| = |\lambda|||v||$$

$$\frac{||Av||}{||v||} = |\lambda|$$

Czyli dla dowolnej wartości własnej i dowolnej normy indukowanej zachodzi:

$$\sup_{v \neq 0} \frac{||Av||}{||v||} \geqslant |\lambda| \implies |\lambda| \leqslant ||A||$$

To się przydaje w metodzie Richardsona.

Twierdzenie Gerszgorina - o lokalizacji wartości własnych. Każda wartość własna macierzy A leży co najmniej w jednym z kół na płaszczyźnie zespolonej:

$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leqslant \sum_{j=1, j \neq i}^{n} |a_{ij}|\} \text{ dla } i = 1, 2, \ldots, n$$

Dowód. Weźmy dowolną wartość własną λ z wektorem v. Pokażemy, że istnieje wiersz i macierzy A t. że $\lambda \in D_i$. Niech $||v||_\infty = 1$ co implikuje że $|v_i| = 1$. Teraz $(Av)_i = \lambda v_i = \sum_{j=1}^{n} a_{ij} v_j$

$$(\lambda - a_{ii}) v_i = \sum_{j=1, j \neq i}^{n} a_{ij} v_j$$

$$|(\lambda - a_{ii}) v_i| = |\sum_{j=1, j \neq i}^{n} a_{ij} v_j|$$

$$|(\lambda - a_{ii}) v_i| \leqslant \sum_{j=1, j \neq i}^{n} |a_{ij}||v_j| \leqslant \sum_{j \neq i} |a_{ij}|$$

□

Wniosek. Macierz diagonalnie dominująca nie ma zerowych wartości własnych, czyli jest nieosobliwa.

11 Ćwiczenia 16/11

11.1 Wyznaczanie wektorów i wartości własnych - c.d.

Metody wyznaczania - dla $Av_i = \lambda_i v_i$

- Metoda potęgowa - Zaczynamy od wektora x_0 i mnożymy z lewej przez A. Zakładamy $|\lambda_1| > |\lambda_2| \geqslant \ldots \geqslant |\lambda_n|$ oraz A ma n wektorów własnych. Robimy:

$$x_0, \|x_0\|_2 = 1$$

$$y_{k+1} = Ax_k, x_{k+1} = \frac{y_{k+1}}{\|y_{k+1}\|_2}$$

Na koniec wartość własną dostajemy:

$$\sigma_k = \frac{x_k^T A x_k}{x_k^t x_k} = x_k^T y_{k+1}$$

- Odwrotna metoda potęgowa.

$$x_0, \|x_0\|_2 = 1$$
$$(A - \mu I) y_{k+1} = x_k$$
$$x_{k+1} = \frac{y_{k+1}}{\|y_{k+1}\|_2}$$
$$\sigma_k = x_k^T y_{k+1}$$

  Jeśli $\mu = 0$ to $|\lambda_n|$ musi być ostro mniejsze, bo w normalnej $|\lambda_0|$ musiało być ostro większe, a $v_i = \lambda_i A^{-1} v_i \implies A^{-1} v_i = \frac{1}{\lambda_i} v_i$ Wartości własne $(A - \mu I)^{-1}$ to $\frac{1}{\lambda_i - \mu}$

- Co jeśli $|\lambda_1| = |\lambda_2| > |\lambda_3| \geqslant \ldots$? Dostaniemy wektor będący kombinacją liniową wektorów własnych odpowiadających λ_1 i λ_2.

- Mamy $|\lambda_1| > |\lambda_2| > |\lambda_3| \geqslant \ldots \geqslant |\lambda_n|$. Wyznaczyliśmy λ_1 z metody potęgowej. Jak wyznaczyć $|\lambda_2|$? Przyjmujemy $x_0 = \sum_{i=2}^{n} \alpha_i v_i$ jeśli $\{v_i\}$ baza ortogonalna: wybieramy $x_0 \perp v_1$. Wtedy co kilka kroków trzeba x_k ortogonalizować, żeby zachować ortogonalność utraconą ze względu na błędy zmiennoprzecinkowe.

Dane są:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, x_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Czy metoda potęgowa dla A jest zbieżna dla x_0? Policzmy ręcznie i zobaczmy XDD
Wartości własne: rozwiążmy $\det(A - \lambda I) = 0$

$$\det(A - \lambda I) = (2 - \lambda)(2 - \lambda) - 1 = 0$$
$$\lambda_1 = 3, \lambda_2 = 1$$

Czyli wektory własne:

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$
$$v_1 = \begin{bmatrix} 1 \\ -1 + \varepsilon \end{bmatrix}$$
$$Ax_0 = \begin{bmatrix} 1 + \varepsilon \\ -1 + 2\varepsilon \end{bmatrix}$$

Z epsilona zrobiły się dwa epsilony, w następnym 4 i 5, pottem 13 i 14. Iloraz współczynników przy epsilonach zbiega do jedynki, więc wynikowy wektor będzie w dobrym kierunku, ale możemy zbiec do wektora do którego mieliśmy dalej.

11.2 Metoda QR

$$A_0 = A$$

$$A_k = Q_k R_k - \text{rozkład QR}$$

$$A_{k+1} = R_k Q_k - \text{mnożymy na odwrót}$$

Wyznaczanie rozkładów jest kosztowne, ale da się łatwiej używając macierz Hessenberga która jest podobna do macierzy A (czyli ma te same wartości własne), a metoda QR zachowuje hessenbergowość.

Dowód. Mamy pokazać, że jeśli macierz Hessenberga $A = QR$ to macierz RQ też jest Hessenberga.

1. Jeśli A - Hessenberga to Q też:

$$\begin{bmatrix} x & \dots & & & x \\ x & x & \dots & & x \\ & x & x & \dots & x \\ & & \ddots & & x \\ & & & x & x \end{bmatrix} = \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \\ & & & \end{bmatrix} \begin{bmatrix} x & \dots & & & x \\ & x & \dots & & x \\ & & x & \dots & x \\ & & & \ddots & x \\ & & & & x \end{bmatrix}$$

   Każda kolejna kolumna Q to kolumna A odpowiednio przemnożona.

2. Iloczyn RQ tj, trójkątna górna razy Hessenberga daje macierz Hessenberga. Rozpisać tak samo.

□

Jak sprowadzić macierz do postaci Hessenberga przy pomocy podobieństw? Używamy Householdera. Dlaczego by nie do trójkątnej? Pomnożymy z lewej strony i dostaniemy zera w pierwszej kolumnie, ale potem pomnożymy z prawej żeby było podobieństwo i rozwali nam to pierwszą kolumnę. Jeśli zrobimy tak, żeby nie tykać pierwszego wiersza, to z prawej ten sam householder nie zmieni nam pierwszej kolumny. Dowód poprawności:

$$A_k = \begin{bmatrix} B & F^T \\ D & E \end{bmatrix}$$

Gdzie B jest Hessenberga $k \times k$, a D ma zera we wszystkich kolumnach poza ostatnią, w której jest wektor d. Dobieramy $\tilde{H}_k$ tak aby $\tilde{H}_k d = \alpha \tilde{e_1}$ Wtedy $H_k = \begin{bmatrix} I & 0 \\ 0 & \tilde{H}_k \end{bmatrix}$ Teraz

$$A_{k+1} = H_k A_k H_k = A_k = \begin{bmatrix} B & F^T \\ \tilde{H}_k D & \tilde{H}_k E \end{bmatrix} H_k = \begin{bmatrix} B & F^T \tilde{H}_k \\ \tilde{H}_k D \tilde{H}_k & \tilde{H}_k E \tilde{H}_k \end{bmatrix}$$

W ten sposób otrzymujemy $H_{n-2} \dots H_1 A H_i \dots H_{n-2} = T$, T jest postaci Hessenberga i jest podobna do A.

12 Ćwiczenia n+2

12.1 Arytmetyka fl

W arytmetyce fl nie ma łączności w działaniach, np. $1+\gamma+\gamma, \gamma = \frac{3}{2}10^{-16}$ Precyzja arytmetyki to około $2.2 \cdot 10^{-16}$. Policzmy:

$$fl(1+\gamma+\gamma) = fl((1+\gamma)+\gamma) = fl(fl(1+\gamma)+\gamma)$$

$$fl(1+\gamma) = 1$$

, bo najmniejsza reprezentowalna liczba większa od 1 to $1+\varepsilon$. Zatem wynikiem jest 1. Ale co jeśli inaczej znawiasujemy? Mnożenie przez 2 jest dokładne.

$fl(1+\gamma+\gamma) = fl(1+(\gamma+\gamma)) = fl(1+fl(\gamma+\gamma)) = fl(1+2\gamma) \neq 1$, bo $2\gamma > \varepsilon$

12.2 Utrata cyfr znaczących przy odejmowaniu

$$x = 0.3721478693,\ y = 0.3720230572,\ x-y = ?$$

Mamy system który obsługuje tylko 5 cyfr znaczących.

$$rd(x) = 0.37215,\ rd(y) = 0.37202$$

W naszym systemie uzyskamy wynik $fl(x-y) = 0.00013$, podczas gdy w idealnej arytm. $x-y = 0.0001248121$. Błąd bezwględny nie jest duży, ale względny: $\frac{fl(x-y)-(x-y)}{x-y} \simeq 4 \cdot 10^{-2}$. Arytmetyka ma dokładność rzędu 10^{-5}, a nasz błąd jest rzędu aż 10^{-2}! Jest to związane z tym, że liczby które od siebie odejmujemy są bliskie sobie.

Jak policzyć $a^2 - b^2$? Wersja 1:

```
s := a * a;
t := b * b;
w := s - t;
```

Wersja 2:

```
u := a + b;
v := a - b;
w := u * v;
```

Mamy zagwarantowane, że wszystkie działania spełniają

$$fl(x \cdot y) = (x \cdot y)(1+\nu), |\nu| \leqslant \varepsilon, \cdot \in \{+,-,*,/\}$$

Ale to zakłada, że liczby są dokładnie reprezentowane! Jakie są błędy naszych "algorytmów"?

1. $fl(a^2-b^2) = [a^2(1+\delta_1) - b^2(1+\delta_2)](1+\delta_3) = a^2(1+\nu_1) - b^2(1+\nu_2)$, gdzie $\nu_1 = \delta_1 + \delta_3 + \delta_1\delta_3, \nu_2 = \delta_2 + \delta_3 + \delta_2\delta_3$. $\frac{fl(a^2-b^2)-(a^2-b^2)}{a^2-b^2} = \frac{a^2\nu_1 - b^2\nu_2}{a^2-b^2}$ Błąd względny zależy od danych! Jeśli a^2, b^2 są bliskie i duże i dodatkowo błędy ν_1, ν_2 są przeciwnego znaku, to błąd względny jest DUŻY!

2. $fl(a^2-b^2) = [(a+b)(1+\delta_1)(a-b)(1+\delta_2)](1+\delta_3)$
$= (a^2-b^2)(1+\delta_1)(1+\delta_2)(1+\delta_3) = (a^2-b^2)(1+E)$
$E = \delta_1 + \delta_2 + \delta_3 + \delta_1\delta_2 + \delta_2\delta_3 + \delta_1\delta_3 + \delta_1\delta_2\delta_3$ Poza pierwszymi trzema składnikami, reszta jest grubo poniżej dokładności arytmetyki, zatem $|E| \tilde{\leqslant} 3\varepsilon$!

Zatem licząc różnicę kwadratów, zawsze liczmy wzorem skróconego mnożenia.

12.3 Uwarunkowanie zadania

Wcześniej były dokładne dane i niedokładne obliczenia, teraz mamy dokładne obliczenia na niedokładnych danych. Policzmy uwarunkowanie zadania różnicy kwadratów.
Zaburzone dane:

$$\tilde{a} = a(1+\delta_1),\ \tilde{b} = b(1+\delta_2),\ |\delta_1| \leqslant \varepsilon$$

$$\left|\frac{\tilde{a}^2 - \tilde{b}^2 - (a^2-b^2)}{a^2-b^2}\right| = \left|\frac{a^2(1+\delta_1)^2 - b^2(1+\delta_2)^2 - (a^2-b^2)}{a^2-b^2}\right| \tilde{\leqslant} 2\frac{a^2+b^2}{|a^2+b^2|}$$ - wskaźnik uwarunkowania

Wskaźnik uwarunkowania jest wysoki, co oznacza że niedokładne dane mają duży wpływ na błąd, czyli zadanie jest źle uwarunkowane. Uwaga, tutaj nie miało znaczenia którego algorytmu użyliśmy! Dla niedokładnych danych oba algorytmy dadzą niedokładne wyniki.

12.4 Numeryczna poprawność algorytmu

Jak obliczyć $f(x) = 1 - \cos x$? Dla $x \approx 0$ mamy $\cos x \approx 1$. Jak przekształcić?

$$\cos x = \cos 2\frac{x}{2} = \cos^2\frac{x}{2} - \sin^2\frac{x}{2}$$

Wtedy

$$1 - \cos x = \cos^2\frac{x}{2} + \sin^2\frac{x}{2} - \cos^2\frac{x}{2} + \sin^2\frac{x}{2} = 2\sin^2\frac{x}{2}$$

Dzielenie przez 2 jest dość dokładne, a f. tryg. i tak musimy policzyć.

$$g(x) = \sqrt{x^2+1} - x$$

Mamy utratę precyzji gdy $x >> 1$

$$g(x) = \sqrt{x^2+1} - x = \frac{1}{\sqrt{x^2+1} + x}$$

Jak policzyć iloczyn skalarny $x^T y$?

$$x^T y = \sum_{i=1}^{n} x_i y_i$$

Sprawdźmy uwarunkowanie:

$$\tilde{x}_i = x_i(1 + \delta_i)$$

$$\tilde{y}_i = y_i(1 + \gamma_i)$$

$$|\frac{\sum \tilde{x}_i \tilde{y}_i - \sum x_i y_i}{\sum x_i y_i}| \approx |\frac{\sum \tilde{x}_i \tilde{y}_i (\delta_i + \gamma_i) - \sum x_i y_i}{\sum x_i y_i}| \leqslant 2\frac{\sum |x_i||y_i|}{|\sum x_i y_i|}\varepsilon$$

Zadanie jest niestety źle uwarunkowane.