Metody numeryczne

Wykład 2/3 - Układy równań liniowych

Janusz Szwabiński

Plan wykładu

1. Układy równań liniowych
2. Pojęcia podstawowe
3. Metody dokładne
4. Metody iteracyjne
5. Układy niedookreślone
6. Układy nadokreślone

Układy równań liniowych

$$\mathbf{A}\vec{x} = \vec{b}$$

- układ może mieć nieskończenie wiele rozwiązań, jedno rozwiązanie lub nie mieć ich wcale
- warunki istnienia rozwiązań układu są znane
- istnieją też gotowe wzory na wyliczenie $\vec{x}$ w wielu przypadkach
- numeryczne rozwiązanie może się okazać dość trudnym zadaniem

Układy równań liniowych

- jedno z ważniejszych zagadnień w ramach tego kursu
- wiele problemów fizycznych sprowadza się do rozwiązania układów równań liniowych
- w analizie numerycznej wiele algorytmów opartych jest o takie układy

Numeryczne metody rozwiązań

metody dokładne przy braku błędów zaokrągleń dają dokładne rozwiązanie po skończonej liczbie przekształceń układu wyjściowego

metody iteracyjne pozwalają na wyznaczenie zbieżnego ciągu rozwiązań przybliżonych

Normy

Definicja

Normą w przestrzeni $\mathbf{R}^n$ nazywamy funkcję

$$\|\cdot\| : \mathbf{R}^n \to \langle 0, +\infty)$$

o następujących własnościach:

1. $\|\vec{x}\| \geqslant 0$ dla każdego $x \in \mathbf{R}^n$,
2. $\|\alpha\vec{x}\| = |\alpha|\|\vec{x}\|$ dla każdego $\alpha \in \mathbf{R}$ i każdego $\vec{x} \in \mathbf{R}^n$,
3. $\|\vec{x}_1 - \vec{x}_2\| \leqslant \|\vec{x}_1\| + \|\vec{x}_2\|$ dla każdej pary $\vec{x}_1, \vec{x}_2 \in \mathbf{R}^n$ (nierówność trójkąta),
4. $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = 0$.

Normy wektorowe w $\mathbf{R}^n$

- dla $\vec{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}} \in \mathbf{R}^n$ można wprowadzić wiele norm
- najczęściej stosowane w obliczeniach numerycznych:

$$\begin{aligned} \|\vec{x}\|_1 &= |x_1| + |x_2| + \cdots + |x_n| \\ \|\vec{x}\|_2 &= \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \\ \|\vec{x}\|_\infty &= \max\{|x_1|, |x_2|, \ldots, |x_n|\} \end{aligned}$$

- równoważne w tym sensie, że jeśli ciąg wektorów $\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots$ dąży do wektora zerowego w jednej normie, to zbieżność zachodzi również w dowolnej innej

Normy macierzowe

Definicja

Normą macierzy $\mathbf{A}$ nazywamy

$$\|\mathbf{A}\|_{pq} = \max_{\vec{x}\in\mathbf{R}^n, \vec{x}\neq \mathrm{o}} \frac{\|\mathbf{A}\vec{x}\|_q}{\|\vec{x}\|_p}.$$

Przy tym, jeżeli $p = q$, będziemy pisać $\|\mathbf{A}\|_p$.

Normy macierzowe

$$
\begin{aligned}
\|\mathbf{A}\|_1 &= \max_{j=1,\ldots,n} \sum_{i=1}^{n} |a_{ij}| \\
\|\mathbf{A}\|_2 &= \sqrt{\lambda_{max}} \\
\|\mathbf{A}\|_\infty &= \max_{i=1,\ldots,n} \sum_{j=1}^{n} |a_{ij}|
\end{aligned}
$$

λ_{max} - największa wartość własna macierzy $\mathbf{A}^{\mathrm{T}}\mathbf{A}$

Normy macierzowe

Definicja

Euklidesową normą macierzy (normą Schura, normą Frobeniusza) nazywamy

$$\|\mathbf{A}\|_E = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}.$$

Norma Euklidesowa spełnia warunek zgodności z $\|\cdot\|_2$, tzn.:

$$\|\mathbf{A}\vec{x}\|_2 \leqslant \|\mathbf{A}\|_E \|\vec{x}\|_2.$$

Wyznaczniki

Definicja

Wyznacznikiem macierzy kwadratowej **A**,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{21} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

nazywamy liczbę

$$\det \mathbf{A} = \sum_f (-1)^{l_f} a_{1\alpha_1} a_{2\alpha_2} \ldots a_{n\alpha_n},$$

gdzie $\sum_f$ oznacza sumowanie po wszystkich permutacjach liczb naturalnych $1, 2, \ldots, n$, a l_f to liczba inwersji w permutacji f.

Wyznaczniki

- definicja ma niewielkie znaczenie praktyczne
- możemy próbować policzyć wyznacznik z rozwinięcia Laplace'a wzdłuż i-tego wiersza lub j-tej kolumny,

$$\det \mathbf{A} = \sum_{j=1}^{n} a_{ij} A_{ij}$$

$$\det \mathbf{A} = \sum_{j=1}^{n} a_{jk} A_{jk}$$

A_{ij} - dopełnienie algebraiczne elementu a_{ij} macierzy $\mathbf{A}$

- rozwinięcie Laplace'a wymaga $n!$ mnożeń
- można je stosować tylko dla bardzo małych n

Macierze trójkątne

$$\mathbf{L} = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{21} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix}.$$

- sumy, iloczyny i odwrotności macierzy trójkątnych tego samego rodzaju są znowu macierzami trójkątnymi
- łatwo wyliczyć ich wyznacznik

$$\det \mathbf{L} = l_{11} l_{22} \dots l_{nn}, \quad \det \mathbf{R} = r_{11} r_{22} \dots r_{nn}$$

Układy równań liniowych

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n &= b_3 \\
a_{41}x_1 + a_{42}x_2 + \cdots + a_{4n}x_n &= b_4 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
$$

Układy równań liniowych - twierdzenie Capellego

Macierz rozszerzona układu

$$\mathbf{D} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{21} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

Układy równań liniowych - twierdzenie Capellego

Twierdzenie

Warunkiem koniecznym i wystarczającym rozwiązywalności dowolnego układu równań liniowych jest, aby rząd r macierzy **A** *układu był równy rzędowi macierzy rozszerzonej* **D**

- jeśli warunek jest spełniony, układ ma rozwiązanie zależne od $n - r$ parametrów
- dla $n = r$ istnieje jednoznaczne rozwiązanie

Wzory Cramera

Macierz układu jest nieosobliwa i kwadratowa:

$$x_k = \frac{\det \mathbf{A_k}}{\det \mathbf{A}}, \quad k = 1, 2, \ldots, n$$

- $\mathbf{A}_k$ powstaje z macierzy $\mathbf{A}$ przez zastąpienie k–tej kolumny przez wektor b
- metoda wymaga bardzo dużego nakładu obliczeń (wyznaczniki)
- może prowadzić do dużych błędów w rozwiązaniu
- nieprzydatna w obliczeniach numerycznych

Analiza zaburzeń

- obliczenia na komputerach nie są dokładne
- rozwiązanie układu równań obarczone pewnym błędem
- wynik niedokładnego działania w arytmetyce zmiennopozycyjnej możemy przedstawić jako wynik działania nieobarczonego błędami wykonanego na zaburzonych argumentach (interpretacja Wilkinsona)

Analiza zaburzeń

- zastępujemy macierz **A** macierzą zaburzoną $\mathbf{A} + \delta\mathbf{A}$
- podobnie, $\vec{b} \to \vec{b} + \delta\vec{b}$
- zamiast rozwiązania $\vec{x}$ układu $\mathbf{A}\vec{x} = \vec{b}$ szukamy rozwiązania $\vec{x} + \delta\vec{x}$ układu

$$(\mathbf{A} + \delta\mathbf{A})(\vec{x} + \delta\vec{x}) = \vec{b} + \delta\vec{b}$$

- błąd $\delta\vec{x}$ zależeć będzie od zaburzeń danych wejściowych $\delta\mathbf{A}$ i $\delta\vec{b}$ oraz od uwarunkowania układu

$\delta\mathbf{A} = \mathrm{o}$ i $\delta\vec{b} \neq \mathrm{o}$

Z równania

$$(\mathbf{A} + \delta\mathbf{A})(\vec{x} + \delta\vec{x}) = \vec{b} + \delta\vec{b}$$

otrzymamy

$$\begin{aligned} \mathbf{A}(\vec{x} + \delta\vec{x}) &= \vec{b} + \delta\vec{b} \\ \mathbf{A}\vec{x} + \mathbf{A}\delta\vec{x} &= \vec{b} + \delta\vec{b} \\ \mathbf{A}\delta\vec{x} &= \delta\vec{b} \\ \delta\vec{x} &= \mathbf{A}^{-1}\delta\vec{b} \end{aligned}$$

$\delta\mathbf{A} = \mathrm{o}$ i $\delta\vec{b} \neq \mathrm{o}$

Dla dowolnych norm norm wektorów $\delta\vec{b}$ i $\delta\vec{x}$ oraz indukowanej przez nie normy macierzy $\mathbf{A}^{-1}$ mamy

$$\|\delta\vec{x}\|_p \leqslant \|\mathbf{A}^{-1}\|_{qp}\|\delta\vec{b}\|_q.$$

Jeśli $\vec{x} \neq \mathrm{o}$, to

$$\begin{aligned}\frac{\|\delta\vec{x}\|_p}{\|\vec{x}\|_p} &\leqslant \frac{\|\mathbf{A}^{-1}\|_{qp}}{\|\vec{x}\|_p}\|\delta\vec{b}\|_q = \frac{\|\mathbf{A}^{-1}\|_{qp}\|\vec{b}\|_q}{\|\vec{x}\|_p}\frac{\|\delta\vec{b}\|_q}{\|\vec{b}\|_q} \\ &= \frac{\|\mathbf{A}^{-1}\|_{qp}\|\mathbf{A}\vec{x}\|_q}{\|\vec{x}\|_p}\frac{\|\delta\vec{b}\|_q}{\|\vec{b}\|_q} = \underbrace{\|\mathbf{A}^{-1}\|_{qp}\|\mathbf{A}\|_{pq}}_{\text{wsk. uwarunkowania}}\frac{\|\delta\vec{b}\|_q}{\|\vec{b}\|_q} = K_{pq}\frac{\|\delta\vec{b}\|_q}{\|\vec{b}\|_q}\end{aligned}$$

$$\delta \mathbf{A} = \mathrm{o} \text{ i } \delta \vec{b} \neq \mathrm{o}$$

- wartość wskaźnika zależy od wyboru norm
- wskaźnik bliski jedności $\rightarrow$ zadanie dobrze uwarunkowane
- duży wskaźnik $\rightarrow$ zadanie źle uwarunkowane
  - nawet niewielkie zaburzenie w wektorze wyrazów wolnych jest wzmacniane i powoduje duży błąd w wyniku

$$\delta \mathbf{A} = 0 \text{ i } \delta \vec{b} \neq 0$$

Przykład

Rozważmy układ

$$\mathbf{A} = \begin{pmatrix} 1,2969 & 0,8648 \\ 0,2161 & 0,1441 \end{pmatrix}, \quad \mathbf{A}^{-1} = 10^8 \begin{pmatrix} 0,1441 & -0,8648 \\ -0,2161 & 1,2969 \end{pmatrix}$$

$\delta\mathbf{A} = 0$ i $\delta\vec{b} \neq 0$

Mamy

$$\|\mathbf{A}\|_\infty = 2,1617, \quad \|\mathbf{A}^{-1}\|_\infty = 1,513 * 10^8$$

oraz

$$K = \|\mathbf{A}^{-1}\|_\infty \|\mathbf{A}\|_\infty \approx 3,3 * 10^8$$

- wskaźnik uwarunkowania $\gg 1$
- przy rozwiązaniu układu w najgorszym wypadku możemy utracić 8 miejsc istotnych dokładności
- bardzo złe uwarunkowanie

Wskaźnik uwarunkowania w praktyce

- w przypadku dużych macierzy wyliczenie wskaźnika może być czasochłonne
- w praktyce często jako kryterium uwarunkowania stosuje się porównanie wartości wyznacznika macierzy **A** z jej elementami
- jeżeli jest on dużo mniejszy niż najmniejszy element macierzy, wówczas zadanie jest na ogół źle uwarunkowane

$\delta\mathbf{A} \neq \mathrm{o}$ i $\delta\vec{b} = \mathrm{o}$

Z równania macierzowego wynika

$$\delta\vec{x} = -\mathbf{A}^{-1}\delta\mathbf{A}(\vec{x} + \delta\vec{x})$$

Wówczas

$$\|\delta\vec{x}\| \leqslant \|\mathbf{A}^{-1}\|\|\delta\mathbf{A}\|\|\vec{x} + \delta\vec{x}\|$$

czyli

$$\frac{\|\delta\vec{x}\|}{\|\vec{x} + \delta\vec{x}\|} \leqslant \|\mathbf{A}^{-1}\|\|\delta\mathbf{A}\| = K\frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|}$$

$\delta\mathbf{A} \neq \mathrm{o}$ i $\delta\vec{b} \neq \mathrm{o}$

- nawet, jeżeli $\mathbf{A}$ i $\vec{b}$ są znane dokładnie, zwykle nie będą miały dokładnej reprezentacji maszynowej
- najczęściej będziemy mieli do czynienia z sytuacją $\delta\mathbf{A} \neq \mathrm{o}$ i $\delta\vec{b} \neq \mathrm{o}$

$\delta\mathbf{A} \neq \mathrm{o}$ i $\delta\vec{b} \neq \mathrm{o}$

Załóżmy, że zaburzenie $\delta\mathbf{A}$ jest na tyle małe, że macierz $\mathbf{A} + \delta\mathbf{A}$ pozostaje nieosobliwa. Wówczas otrzymamy

$$\delta\vec{x} = -\mathbf{A}^{-1}\left(\delta\vec{b} - \delta\mathbf{A}\vec{x} - \delta\mathbf{A}\delta\vec{x}\right)$$

$$\|\delta\vec{x}\| \leqslant \|\mathbf{A}^{-1}\|\left(\|\delta\vec{b}\| + \|\delta\mathbf{A}\|\|\vec{x}\| + \|\delta\mathbf{A}\|\|\delta\vec{x}\|\right)$$

czyli

$$\|\delta\vec{x}\| \leqslant \frac{1}{1 - \|\mathbf{A}^{-1}\|\|\delta\mathbf{A}\|}\|\mathbf{A}^{-1}\|\left(\|\delta\vec{b}\| + \|\delta\mathbf{A}\|\|\vec{x}\|\right)$$

$\delta\mathbf{A} \neq \mathrm{o}$ i $\delta\vec{b} \neq \mathrm{o}$

Z równości $\mathbf{A}\vec{x} = \vec{b}$ wynika

$$\frac{\|\vec{b}\|}{\|\vec{x}\|\|\mathbf{A}\|} \leqslant 1$$

Ostatecznie

$$\begin{aligned}
\frac{\|\delta\vec{x}\|}{\|\vec{x}\|} &\leqslant \frac{1}{1 - \|\mathbf{A}^{-1}\|\|\delta\mathbf{A}\|}\|\mathbf{A}^{-1}\|\|\mathbf{A}\|\left(\frac{\|\delta\vec{b}\|}{\|\vec{b}\|}\frac{\|\vec{b}\|}{\|\vec{x}\|\|\mathbf{A}\|} + \frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|}\right) \\
&\leqslant \frac{K}{1 - \|\mathbf{A}^{-1}\|\|\delta\mathbf{A}\|}\left(\frac{\|\delta\vec{b}\|}{\|\vec{b}\|} + \frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|}\right)
\end{aligned}$$

Układy z macierzami trójkątnymi

- szczególnie łatwe do rozwiązania
- aby istniało jednoznaczne rozwiązanie, macierz musi być nieosobliwa…
- …czyli wszystkie elementy na głównej przekątnej muszą być różne od zera

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{nn}x_n &= b_n \end{aligned}$$

Podstawianie w tył

- wstawiając x_n do przedostatniego równania obliczymy x_{n-1}
- procedurę kontynuujemy aż do wyliczenia x_1

$$\begin{aligned} x_n &= \frac{b_n}{a_{nn}}, \\ x_i &= \frac{b_i - \sum_{k=i+1}^{n} a_{ik} x_k}{a_{ii}}, \ i = n-1, n-2, \ldots, 1 \end{aligned}$$

- koszt obliczeń:

$$M = \frac{1}{2}n^2 + \frac{1}{2}n \ \text{ mnożeń i dzieleń}, \quad D = \frac{1}{2}n^2 - \frac{1}{2}n \ \text{ dodawań}$$

- niewiele większy od kosztu mnożenia wektora przez macierz trójkątną

Podstawianie w przód

$$\begin{aligned} a_{11}x_1 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

- wstawiając x_1 do drugiego równania obliczymy x_2 itd.

$$x_1 = \frac{b_1}{a_{11}}, \quad x_i = \frac{b_i - \sum_{k=1}^{i-1} a_{ik}x_k}{a_{ii}}, \ i = 2, 3, \ldots, n$$

- koszt obliczeń ten sam, co poprzednio

Jak rozwiązać dowolny układ?

1. Sprowadź układ wyjściowy do postaci trójkątnej
2. Zastosuj wzory na podstawianie w tył lub w przód

Eliminacja Gaussa

- jedna z metod sprowadzenia układu równań do postaci trójkątnej
- nazwana na cześć Carla Friedricha Gaussa
- po raz pierwszy zaprezentowana została dużo wcześniej, bo już około 150 roku p.n.e w słynnym chińskim podręczniku matematyki „Dziewięć rozdziałów sztuki matematycznej”

Eliminacja Gaussa - algorytm

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned}$$

Odejmujemy od drugiego wiersza pierwszy pomnożony przez a_{21}/a_{11}, a od trzeciego pierwszy pomnożony przez a_{31}/a_{11}

$$\begin{aligned} a_{11}^{(0)}x_1 + a_{12}^{(0)}x_2 + a_{13}^{(0)}x_3 &= b_1^{(0)} \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 &= b_2^{(1)} \\ a_{32}^{(1)}x_2 + a_{33}^{(1)}x_3 &= b_3^{(1)} \end{aligned}$$

Eliminacja Gaussa - algorytm

Przy tym

$$a_{ij}^{(0)} = a_{ij}, \quad b_i^{(0)} = b_i, \quad i, j = 1, 2, 3$$

oraz

$$a_{ij}^{(1)} = a_{ij}^{(0)} - \frac{a_{i1}^{(0)}}{a_{11}^{(0)}} a_{1j}^{(0)}, \quad b_i^{(1)} = b_i^{(0)} - \frac{a_{i1}^{(0)}}{a_{11}^{(0)}} b_1^{(0)}, \quad i, j = 2, 3$$

Eliminacja Gaussa - algorytm

Odejmujemy od trzeciego równania drugie pomnożone przez $a_{32}^{(1)}/a_{22}^{(1)}$

$$\begin{aligned} a_{11}^{(0)}x_1 + a_{12}^{(0)}x_2 + a_{13}^{(0)}x_3 &= b_1^{(0)} \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 &= b_2^{(1)} \\ a_{33}^{(2)}x_3 &= b_3^{(2)} \end{aligned}$$

$$a_{ij}^{(2)} = a_{ij}^{(1)} - \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}a_{2j}^{(1)}, \quad b_i^{(2)} = b_i^{(1)} - \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}b_2^{(1)}, \quad i,j = 3$$

Eliminacja Gaussa - przypadek ogólny

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} a_{kj}^{(k-1)}, \quad i,j = k+1, k+2, \ldots, n,$$

$$b_{i}^{(k)} = b_{i}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} b_{k}^{(k-1)}, \quad i,j = k+1, k+2, \ldots, n.$$

- otrzymaliśmy układ trójkątny
- jego rozwiązanie ma postać

$$x_i = \frac{b_i^{(i-1)} - \sum_{j=i+1}^{n} a_{ij}^{(i-1)} x_j}{a_{ii}^{(i-1)}}, \quad i = n, n-1, \ldots, 1$$

Eliminacja Gaussa - przypadek ogólny

- nakład obliczeń to

$$M = \frac{1}{3}n^3 + n^2 - \frac{1}{3} \text{ mnożeń i dzieleń}$$

$$D = \frac{1}{3}n^3 + \frac{1}{2}n^2 - \frac{5}{6}n \text{ dodawań}$$

- większa część przypada na sprowadzenie układu do postaci trójkątnej
- liczba operacji bez porównania mniejsza niż w przypadku wzorów Cramera

Niezawodność eliminacji Gaussa

Przykład

$$\begin{pmatrix} 0 & 2 & 2 \\ 3 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

- macierz jest nieosobliwa, a zatem istnieje jednoznaczne rozwiązanie
- mimo to eliminacja Gaussa zawodzi już w pierwszym kroku
- algorytm wymaga dzielenia przez a_{11}, które tutaj jest równe 0
  $\Rightarrow$ eliminacja Gaussa w formie przedstawionej powyżej nie jest niezawodna

Częściowy wybór elementu podstawowego

Definicja

Elementem podstawowym nazywamy ten element macierzy **A**, za pomocą którego dokonujemy eliminacji zmiennej z dalszych równań.

- rozwiązanie równania nie zmieni się, jeżeli zamienimy kolejność wierszy w układzie równań
- możemy to wykorzystać, aby uniknąć problemów związanych z dzieleniem przez zero

Częściowy wybór elementu podstawowego

$$\begin{pmatrix} 0 & 2 & 2 \\ 3 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

Rozważmy macierz rozszerzoną (z położeniem wierszy):

$$\left(\begin{array}{ccc|c} 0 & 2 & 2 & 1 \\ 3 & 3 & 0 & 3 \\ 1 & 0 & 1 & 2 \end{array}\right) \begin{array}{c} : \\ : \\ : \end{array} \begin{array}{c} w1 \\ w2 \\ w3 \end{array}$$

Częściowy wybór elementu podstawowego

Zamieniamy wiersze w macierzy układu, tak aby nowy element diagonalny w jej pierwszym wierszu był różny od zera:

$$\left(\begin{array}{ccc|c} 3 & 3 & 0 & 3 \\ 0 & 2 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{array}\right) \begin{array}{c} : \\ : \\ : \end{array} \begin{array}{c} w1^{(1)} \\ w2^{(1)} \\ w3^{(1)} \end{array}$$

Częściowy wybór elementu podstawowego

Po zamianie wierszy możemy wykonać pierwszy krok eliminacji Gaussa

$$
\begin{array}{lcl}
w1^{(1)} & \rightarrow & \left(\begin{array}{ccc|c} 3 & 3 & 0 & 3 \end{array}\right) : w1^{(2)} \\
w2^{(1)} - (a_{21}^{(1)}/a_{11}^{(1)}) \times w1^{(1)} & \rightarrow & \left(\begin{array}{ccc|c} 0 & 2 & 2 & 1 \end{array}\right) : w2^{(2)} \\
w3^{(1)} - (a_{31}^{(1)}/a_{11}^{(1)}) \times w1^{(1)} & \rightarrow & \left(\begin{array}{ccc|c} 0 & -1 & 1 & 1 \end{array}\right) : w3^{(2)}
\end{array}
$$

Częściowy wybór elementu podstawowego

W kolejnym kroku nie musimy zamieniać wierszy ze sobą:

$$\begin{array}{lcl} w1^{(2)} & \rightarrow & \left(\begin{array}{cccc} 3 & 3 & 0 & 3 \\ 0 & 2 & 2 & 1 \\ 0 & 0 & 2 & 3/2 \end{array}\right) \begin{array}{c} : \\ : \\ : \end{array} \begin{array}{l} w1^{(3)} \\ w2^{(3)} \\ w3^{(3)} \end{array} \\ w2^{(2)} & \rightarrow & \\ w3^{(2)} - (a_{32}^{(2)}/a_{22}^{(2)}) \times w2^{(2)} & \rightarrow & \end{array}$$

Końcowe rozwiązanie znajdziemy podstawiając w tył

$$x_3 = \frac{b_3^{(3)}}{a_{33}^{(3)}} = \frac{3}{4}, \quad x_2 = \frac{b_2^{(3)} - a_{23}^{(3)} x_3}{a_{22}^{(3)}} = -\frac{1}{4}, \quad x_1 = \frac{b_1^{(3)} - a_{12}^{(3)} x_2 - a_{13}^{(3)} x_3}{a_{11}^{(3)}} = \frac{5}{4}$$

Częściowy wybór elementu podstawowego

- teoretycznie możemy dowolnie dobierać wiersze do zamiany
- ze względu na błędy zaokrągleń w *i*-tym kroku eliminacji powinniśmy wybierać wiersz, który ma największy element w *i*-tej kolumnie
- częściowy wybór elementu podstawowego zalecany jest również dla układów, których macierze nie mają zerowych elementów na głównej przekątnej, ponieważ w większości przypadków prowadzi do redukcji błędów zaokrągleń

Częściowy wybór elementu podstawowego - ograniczenia

- nie zawsze prowadzi do poprawy dokładności obliczeń
- odpowiedni wybór jest sprawą delikatną
- czasami warto przeprowadzić równoważenie układu
- ostatecznie można zmienić strategię wyboru z częściowego na całkowity
  - bierzemy pod uwagę wartości elementów w *i*-tej kolumnie i w *i*-tym wierszu
  - duży nakład obliczeń

Ograniczenia - przykład

$$\begin{pmatrix} 10^{-15} & 1 \\ 1 & 10^{11} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 + 10^{-15} \\ 10^{11} + 1 \end{pmatrix}$$

Przy dokładnych obliczeniach eliminacja Gaussa bez wyboru elementu podstawowego da poprawne rozwiązanie

$$\left(\begin{array}{cc|c} 10^{-15} & 1 & 1 + 10^{-15} \\ 1 & 10^{11} & 10^{11} + 1 \end{array}\right) \xrightarrow{el.} \left(\begin{array}{cc|c} 1 & 10^{15} & 1 + 10^{15} \\ 0 & 10^{11} - 10^{15} & 10^{11} - 10^{15} \end{array}\right)$$

$$\xrightarrow{podstawianie} \vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Ograniczenia - przykład

Błędy zaokrągleń spowodują, że wynik będzie znacznie odbiegał od idealnego:

$$\xrightarrow{el.} \left(\begin{array}{cc|c} 1 & 9.999999999999999e+14 & 1.0000000000000001e+015 \\ 0 & -9.9989999999999999e+014 & -9.9990000000000000e+014 \end{array}\right)$$

$$\xrightarrow{podstawianie} \vec{x} = \begin{pmatrix} 8.7500000000000000e-001 \\ 1.0000000000000000e+000 \end{pmatrix}$$

Ograniczenia - przykład

Lepszy wynik uzyskamy, dokonując częściowego wyboru elementu podstawowego:

$$\left(\begin{array}{cc|c} 10^{-15} & 1 & 1+10^{-15} \\ 1 & 10^{11} & 10^{11}+1 \end{array}\right) \xrightarrow{\textit{zamiana wierszy}} \left(\begin{array}{cc|c} 1 & 10^{11} & 10^{11}+1 \\ 10^{-15} & 1 & 1+10^{-15} \end{array}\right)$$

$$\xrightarrow{\textit{eliminacja}} \left(\begin{array}{cc|c} 1 & 1.000e+011 & 1.0000000000010000e+011 \\ 0 & 9.999e-001 & 9.9990000000000001e-001 \end{array}\right)$$

$$\xrightarrow{\textit{podstawianie}} \vec{x} = \begin{pmatrix} 9.9998474121093750e-001 \\ 1.0000000000000000e+000 \end{pmatrix}$$

Ograniczenia - przykład 2

Rozważmy układ

$$\begin{pmatrix} 10^{-14.6} & 1 \\ 1 & 10^{15} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 + 10^{-14.6} \\ 10^{15} + 1 \end{pmatrix}$$

Jego dokładne rozwiązanie wynosi $\vec{x} = (1, 1)^{\mathrm{T}}$. Eliminacja Gaussa da poprawny wynik

$$\xrightarrow{\textit{eliminacja}} \left(\begin{array}{cc|c} 1 & 3.9810717055349 69e+014 & 3.9810717055349 79e+014 \\ 0 & 6.0189282944650 30e+014 & 6.0189282944650 30e+014 \end{array}\right)$$

$$\xrightarrow{\textit{podstawianie}} \vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Ograniczenia - przykład 2

Częściowy wybór elementu podstawowego „zepsuje” wynik

$$\left(\begin{array}{cc|c} 10^{-14.6} & 1 & 1+10^{-14.6} \\ 1 & 10^{15} & 10^{15}+1 \end{array}\right)$$

$$\xrightarrow{\textit{zamiana wierszy}} \left(\begin{array}{cc|c} 1 & 10^{15} & 10^{15}+1 \\ 10^{-14.6} & 1 & 1+10^{-14.6} \end{array}\right)$$

$$\xrightarrow{\textit{eliminacja}} \left(\begin{array}{cc|c} 1 & 1.000e+015 & 1.0000000000000001e+015 \\ 0 & -1.5118864315095819 & -1.5118864315095821 \end{array}\right)$$

$$\xrightarrow{\textit{podstawianie}} \vec{x} = \begin{pmatrix} 0.750000000000000 \\ 1.000000000000002 \end{pmatrix}$$

Równoważenie układu - przykład

Rozważmy układ

$$\begin{pmatrix} 1 & 10000 \\ 1 & 0,0001 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 10000 \\ 1 \end{pmatrix}$$

Układ ten ma rozwiązanie $x_1 = x_2 = 0,9999$, poprawnie zaokrąglone do czterech cyfr dziesiętnych.

Przyjmijmy a_{11} jako element podstawowy i poszukajmy rozwiązań układu w trzycyfrowej arytmetyce zmiennopozycyjnej. Otrzymamy następujące, złe rozwiązanie

$$x_1 = 0.00, \quad x_2 = 1.00.$$

Równoważenie układu - przykład

Pomnóżmy teraz pierwsze równanie przez 10^{-4}

$$\begin{pmatrix} 0,0001 & 1 \\ 1 & 0,0001 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Wybierając a_{21} jako element podstawowy, otrzymamy

$$x_1 = 1.00, \quad x_2 = 1.00,$$

co w trzycyfrowej arytmetyce jest wynikiem dobrym

Eliminacja Gaussa i macierze osobliwe - przykład

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

Po kilku krokach dojdziemy do sytuacji (sprawdzić!):

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Eliminacja Gaussa i macierze osobliwe - przykład

- same zera w ostatnim wierszu sygnalizują, że wyjściowa macierz była osobliwa
- nie istnieje rozwiązanie jednoznaczne
- ponieważ ostatni element wektora wyrazów wolnych jest również równy zero, rozwiązań jest nieskończenie wiele

Macierze odwrotne

- wiele układów różniących się tylko wyrazem wolnym

$$\mathbf{A}\vec{x}_1 = \vec{b}_1, \quad \mathbf{A}\vec{x}_2 = \vec{b}_2, \quad \ldots, \mathbf{A}\vec{x}_N = \vec{b}_N$$

$$\mathbf{A}\left(\vec{x}_1\vec{x}_2\ldots\vec{x}_N\right) = \left(\vec{b}_1\vec{b}_2\ldots\vec{b}_N\right)$$

$$\mathbf{AX} = \mathbf{B}$$

- formalne rozwiązanie ostatniego równania macierzowego ma postać

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$$

- jeżeli **B** będzie macierzą jednostkową, znajdziemy w ten sposób macierz odwrotną do macierzy **A**

Eliminacja Jordana

$$
\begin{aligned}
a_{11}^{(1)}x_1 &+ a_{12}^{(1)}x_2 + \ldots + a_{1n}^{(1)}x_n &&= b_1^{(1)} \\
a_{21}^{(1)}x_1 &+ a_{22}^{(1)}x_2 + \ldots + a_{2n}^{(1)}x_n &&= b_2^{(1)} \\
&&&\vdots \\
a_{n1}^{(1)}x_1 &+ a_{n2}^{(1)}x_2 + \ldots + a_{nn}^{(1)}x_n &&= b_n^{(1)}
\end{aligned}
$$

Eliminacja Jordana

Dzielimy pierwsze równanie obustronnie przez $a_{11}^{(1)}$,a następnie od i–tego wiersza ($i = 2, 3, \ldots, n$) odejmujemy pierwszy pomnożony przez $a_{i1}^{(1)}$,

$$\begin{aligned} x_1 \quad + \quad a_{12}^{(2)} x_2 \quad + \quad \ldots \quad + \quad a_{1n}^{(2)} x_n \quad &= \quad b_1^{(2)} \\ a_{22}^{(2)} x_2 \quad + \quad \ldots \quad + \quad a_{2n}^{(2)} x_n \quad &= \quad b_2^{(2)} \\ &\vdots \\ a_{n2}^{(2)} x_2 \quad + \quad \ldots \quad + \quad a_{nn}^{(2)} x_n \quad &= \quad b_n^{(2)} \end{aligned}$$

Eliminacja Jordana

W kolejnym kroku dzielimy drugie równanie obustronnie przez $a_{22}^{(2)}$ i odejmujemy od i–tego wiersza ($i = 1, 3, 4, \ldots, n$) wiersz drugi pomnożony przez $a_{i2}^{(2)}$,

$$\begin{array}{rcccccl} x_1 & & + & \ldots & + & a_{1n}^{(3)} x_n & = & b_1^{(3)} \\ & x_2 & + & \ldots & + & a_{2n}^{(3)} x_n & = & b_2^{(3)} \\ & & & & & \vdots & & \\ & & & \ldots & + & a_{nn}^{(3)} x_n & = & b_n^{(3)} \end{array}$$

Eliminacja Jordana

Po $(n-1)$ eliminacjach otrzymujemy układ

$$\begin{aligned} x_1 &= b_1^{(n)} \\ x_2 &= b_2^{(n)} \\ \ddots &\ \vdots \\ x_n &= b_n^{(n)} \end{aligned}$$

Eliminacja Jordana

- koszt obliczeń
$$M = \frac{1}{2}n^3 + \frac{1}{2}n^2, \quad D = \frac{1}{2}n^3 - \frac{1}{2}$$
- potrzebujemy wyboru elementu podstawowego w celu zagwarantowania niezawodności
- zalety:
  - oszczędne gospodarowanie pamięcią
  - możliwość określenia rozwiązania „obciętego” układu równań
- wady:
  - duży nakład obliczeń (około 1,5 raza większy niż w eliminacji Gaussa)
  - brak odpowiednika rozkładu **LU** (o tym zaraz)

Rozkład LU

- przypuśćmy, że macierz **A** układu da się przedstawić w postaci iloczynu macierzy trójkątnej dolnej **L** i trójkątnej górnej **U**

$$\mathbf{A} = \mathbf{LU}$$

- jeżeli macierz **A** jest nieosobliwa, zachodzi

$$\mathbf{A}^{-1} = (\mathbf{LU})^{-1} = \mathbf{U}^{-1}\mathbf{L}^{-1}$$

- rozwiązanie układu da się przedstawić w postaci

$$\vec{x} = \mathbf{A}^{-1}\vec{b} = \mathbf{U}^{-1}\left(\mathbf{L}^{-1}\vec{b}\right)$$

Rozkład LU

$\Rightarrow$ aby znaleźć rozwiązanie $\vec{x}$ układu dysponując rozkładem **LU** jego macierzy, wystarczy rozwiązać dwa układy trójkątne

$$\begin{aligned} \mathbf{L}\vec{y} &= \vec{b} \\ \mathbf{U}\vec{x} &= \vec{y} \end{aligned}$$

Eliminacja Gaussa a rozkład LU

Przekształcenie

$$\mathbf{A}^{(1)}x = b^{(1)} \quad \rightarrow \quad \mathbf{A}^{(2)}x = b^{(2)}$$

jest równoważne pomnożeniu obu stron układu $\mathbf{A}^{(1)}x = b^{(1)}$ przez macierz

$$\mathbf{L}^{(1)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -l_{21} & 1 & 0 & \dots & 0 \\ -l_{31} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -l_{n1} & 0 & 0 & \dots & 1 \end{pmatrix}, \quad l_{i1} = \frac{a_{i1}^{(1)}}{a_{11}^{(1)}}, \quad i = 2, 3, \dots, n$$

Eliminacja Gaussa a rozkład LU

W ten sposób otrzymujemy dwa równania:

$$\mathbf{L}^{(1)}\mathbf{A}^{(1)} = \mathbf{A}^{(2)}, \quad \mathbf{L}^{(1)}b^{(1)} = b^{(2)}$$

Podobnie

$$\mathbf{L}^{(2)}\mathbf{A}^{(2)} = \mathbf{A}^{(3)}, \quad \mathbf{L}^{(2)}b^{(2)} = b^{(3)}$$

$$\mathbf{L}^{(2)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & -l_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -l_{n2} & 0 & \dots & 1 \end{pmatrix}, \quad l_{i2} = \frac{a_{i2}^{(2)}}{a_{22}^{(2)}}, \quad i = 3, \dots, n$$

Eliminacja Gaussa a rozkład LU

Ostatecznie

$$\mathbf{L}^{(n-1)}\mathbf{L}^{(n-2)}\dots\mathbf{L}^{(1)}\mathbf{A}^{(1)} = \mathbf{A}^{(n)}$$

oraz

$$\mathbf{L}^{(n-1)}\mathbf{L}^{(n-2)}\dots\mathbf{L}^{(1)}b^{(1)} = b^{(n)}$$

Macierze $\mathbf{L}^{(i)}$, $i = 1, \dots, n-1$ są nieosobliwe, więc

$$\mathbf{A}^{(1)} = (\mathbf{L}^{(1)})^{-1}(\mathbf{L}^{(2)})^{-1}\dots(\mathbf{L}^{(n)})^{-1}\mathbf{A}^{(n)}$$

Eliminacja Gaussa a rozkład LU

Ponadto

$$(\mathbf{L}^{(1)})^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & 0 & 0 & \dots & 1 \end{pmatrix}, \quad (\mathbf{L}^{(2)})^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & l_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & l_{n2} & 0 & \dots & 1 \end{pmatrix} \dots$$

Eliminacja Gaussa a rozkład LU

Stąd

$$\mathbf{L} \equiv (\mathbf{L}^{(1)})^{-1}(\mathbf{L}^{(2)})^{-1}\dots(\mathbf{L}^{(n)})^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & 1 \end{pmatrix}.$$

Z drugiej strony wiemy, że $\mathbf{A}^{(n)} = \mathbf{U}$ jest macierzą trójkątną górną.

Eliminacja Gaussa a rozkład LU

- zapamiętując macierze **L** i **U**, możemy szybko rozwiązać wiele układów różniących się tylko kolumnami wyrazów wolnych
- w ramach oszczędności pamięci możemy zapisywać elementy tych macierzy w miejsce elementów macierzy **A**,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ l_{21} & u_{22} & u_{23} & \dots & u_{2n} \\ l_{31} & l_{32} & u_{33} & \dots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & u_{nn} \end{pmatrix}$$

Eliminacja Gaussa a rozkład LU

- nie każdą macierz nieosobliwą można przedstawić w postaci
- aby rozkład istniał, wszystkie minory główne macierzy muszą być różne od zera
- jeżeli eliminację Gaussa można przeprowadzić do końca, rozkład LU na pewno istnieje

Rozkład LU a wybór elementu podstawowego

Jeżeli eliminacja Gaussa wymaga zamiany wierszy, wówczas zamiast rozkładu LU macierzy **A** znajdziemy rozkład permutacji jej wierszy

$$\mathbf{PA} = \mathbf{LU}$$

Znaczenie macierzy permutacji **P** ilustruje następujący przykład:

$$\mathbf{PA} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

Rozkład LU a wybór elementu podstawowego

Macierz permutacji ma następującą własność:

$$\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I} \;\Rightarrow\; \mathbf{P}^{\mathrm{T}} = \mathbf{P}^{-1}$$

Stąd wynika

$$\mathbf{A} = \mathbf{P}^{\mathrm{T}}\mathbf{L}\mathbf{U}$$

Rozkład LU i metoda Doolittle'a

Potraktujmy równość

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

jako układ n^2 równań dla n^2 niewiadomych l_{ij} i u_{ij}

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ l_{21}u_{11} & l_{21}u_{12} + u_{22} & l_{21}u_{13} + u_{23} \\ l_{31}u_{11} & l_{31}u_{12} + l_{32}u_{22} & l_{31}u_{13} + l_{32}u_{23} + u_{33} \end{pmatrix}$$

Rozkład LU i metoda Doolittle'a

Stąd

$$
\begin{array}{lll}
u_{11} = a_{11}, & u_{12} = a_{12}, & u_{13} = a_{13} \\
l_{21} = \frac{a_{21}}{u_{11}}, & u_{22} = a_{21} - l_{21}u_{12}, & u_{23} = a_{23} - l_{21}u_{13} \\
l_{31} = \frac{a_{31}}{u_{11}}, & l_{32} = \frac{a_{32} - l_{31}u_{12}}{u_{22}} & u_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23}
\end{array}
$$

Rozkład LU i metoda Doolittle'a

- w przypadku ogólnym elementy macierzy **L** i **U** obliczamy dla $i = 1, 2, \ldots, n$ ze wzorów

$$
\begin{aligned}
u_{ij} &= a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj}, \quad j = i, i+1, \ldots, n \\
l_{ji} &= \frac{a_{ji} - \sum_{k=1}^{i-1} l_{jk} u_{ki}}{u_{ii}}, \quad j = i+1, i+2, \ldots, n
\end{aligned}
$$

Rozkład LU i metoda Doolittle'a

- koszt obliczeń (łącznie z rozw. układów trójkątnych)

$$M = \frac{1}{3}n^3 + n^2 - \frac{1}{3}n, \quad D = \frac{1}{3}n^3 + \frac{1}{3}n^2 - \frac{5}{6}n$$

- koszt taki sam, jak w eliminacji Gaussa
- niezawodna w połączeniu z wyborem elementu podstawowego
- wiersze zamieniamy ze sobą miejscami tak, aby element u_{ii} był jak największy

Metoda Doolitle'a - przykład

Chcemy wyznaczyć rozkład LU macierzy

$$\begin{pmatrix} 20 & 31 & 23 \\ 30 & 24 & 18 \\ 15 & 32 & 21 \end{pmatrix}$$

metodą Doolittle'a z częściowym wyborem elementu podstawowego. W tym celu wprowadzamy dodatkową kolumnę indeksującą wiersze

$$\begin{pmatrix} 20 & 31 & 23 \\ 30 & 24 & 18 \\ 15 & 32 & 21 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Metoda Doolitle'a - przykład

Element podstawowy wybieramy tak, aby element u_{ii} występujący we wzorach ogólnych miał jak największą wartość.

Dla $i = 1$ w zależności od tego, czy na pierwszym miejscu ustawimy wiersz pierwszy, drugi czy trzeci, uzyskamy odpowiednio $u_{11} = 20$, $u_{11} = 30$ oraz $u_{11} = 15$.

Zamieniamy miejscami wiersz pierwszy z drugim

$$\begin{pmatrix} 30 & 24 & 18 \\ 20 & 31 & 23 \\ 15 & 32 & 21 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

Metoda Doolitle'a - przykład

Otrzymamy

$$u_{11} = 30, \quad u_{12} = a_{21} = 24, \quad u_{13} = a_{13} = 18$$
$$l_{21} = \frac{2}{3}, \qquad l_{31} = \frac{1}{2}$$

Wartości te wpisujemy do macierzy **A**

$$\begin{pmatrix} 30 & 24 & 18 \\ \frac{2}{3} & 31 & 23 \\ \frac{1}{2} & 32 & 21 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

Metoda Doolitle'a - przykład

Dla $i = 2$ otrzymamy

$$u_{22} = a_{22} - a_{21}a_{12} = 31 - \frac{2}{3} * 24 = 15$$

lub

$$u_{22} = a_{32} - a_{31}a_{12} = 32 - \frac{1}{2} * 24 = 20$$

w zależności od tego, czy na drugim miejscu ustawimy wiersz drugi czy trzeci.

Metoda Doolitle'a - przykład

Zamieniamy wiersze miejscami

$$\begin{pmatrix} 30 & 24 & 18 \\ \frac{1}{2} & 32 & 21 \\ \frac{2}{3} & 31 & 23 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

Znajdujemy

$$u_{22} = 20, \quad u_{23} = a_{23} - a_{21}a_{13} = 12, \quad u_{32} = \frac{15}{20}$$

Metoda Doolitle'a - przykład

Uzyskane wartości wpisujemy do macierzy

$$\begin{pmatrix} 30 & 24 & 18 \\ \frac{1}{2} & 20 & 12 \\ \frac{2}{3} & \frac{3}{4} & 23 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

Dla $i = 3$ obliczamy

$$u_{33} = 2.$$

Metoda Doolitle'a - przykład

Stąd

$$\begin{pmatrix} 30 & 24 & 18 \\ \frac{1}{2} & 20 & 12 \\ \frac{2}{3} & \frac{3}{4} & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

W ten sposób w miejsce macierzy **A** otrzymaliśmy rozkład **LU** macierzy, która składa się z wierszy 2, 3 i 1 macierzy wyjściowej **A**.

Rozkład LU i metoda Crouta

Przyjmujemy dla odmiany, że **U** ma na głównej przekątnej same jedynki

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

i ponownie potraktujemy powyższe wyrażenie jak równanie na niewiadome elementy macierzy trójkątnych.

Rozkład LU i wyznaczniki

$$\det \mathbf{A} = \det(\mathbf{LU}) = \det \mathbf{L} \det \mathbf{U} = \begin{cases} u_{11}u_{22}\ldots u_{nn}, & l_{ii} = 1 \\ l_{11}l_{22}\ldots l_{nn}, & u_{ii} = 1 \end{cases}$$

Macierze dominujące diagonalnie

Definicja

Macierz kwadratową **A** nazywamy diagonalnie dominującą, jeżeli

$$|a_{ii}| \geqslant \sum_{\substack{k=1 \\ k \neq i}}^{n} |a_{ik}|, \quad i = 1, 2, \ldots, n$$

Jeżeli nierówności są ostre, mówimy o macierzy silnie diagonalnie dominującej.

Macierze dominujące diagonalnie

Definicja

Macierz **A** jest diagonalnie dominująca kolumnowo, jeżeli $\mathbf{A}^{\mathrm{T}}$ jest diagonalnie dominująca, tzn.

$$|a_{ii}| \geqslant \sum_{\substack{k=1 \\ k \neq i}}^{n} |a_{ki}|, \quad i = 1, 2, \ldots, n$$

Macierze dominujące diagonalnie

Twierdzenie

Jeżeli macierz **A** *jest nieosobliwa i diagonalnie dominująca kolumnowo, to przy eliminacji metodą Gaussa nie ma potrzeby przestawiania wierszy.*

Macierze trójdiagonalne

$$\mathbf{T} = \begin{pmatrix} b_1 & c_1 & & & & \\ a_2 & b_2 & c_2 & & \mathrm{o} & \\ & a_3 & b_3 & c_3 & & \\ & & a_4 & b_4 & \ddots & \\ & & & \ddots & \ddots & \ddots \\ & \mathrm{o} & & & \ddots & b_{n-1} & c_{n-1} \\ & & & & & a_n & b_n \end{pmatrix}$$

Rozkład LU macierzy trójdiagonalnej

$$\mathbf{L} = \begin{pmatrix} 1 & & & 0 \\ l_2 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & l_n & 1 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} u_1 & c_1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & c_{n-1} \\ 0 & & & u_n \end{pmatrix}$$

$$u_1 = b_1, \quad l_i = \frac{a_i}{u_{i-1}}, \quad u_i = b_i - l_i c_{i-1}, \quad i = 2, 3, \ldots, n$$

- rozkład wymaga $O(n)$ operacji
- metoda niezawodna, jeśli $\mathbf{T}$ jest diagonalnie dominująca kolumnowo

Błędy zaokrągleń

- macierze **L** i **U** spełniają warunek

$$\mathbf{LU} = \mathbf{A} + \mathbf{E}$$

$\mathbf{E}$ - błąd rozkładu

- $\vec{y}$ i $\vec{x}$ możemy potraktować jako dokładne rozwiązania układów

$$\begin{aligned}(\mathbf{L} + \delta\mathbf{L})\vec{y} &= \vec{b} \\ (\mathbf{U} + \delta\mathbf{U})\vec{x} &= \vec{y}\end{aligned}$$

Błędy zaokrągleń

Stąd

$$(\mathbf{A} + \mathbf{E} + \mathbf{L}\delta\mathbf{U} + \delta\mathbf{L}\mathbf{U} + \delta\mathbf{L}\delta\mathbf{U})\vec{x} = \vec{b}$$

Można pokazać, że zaburzenie

$$\delta\mathbf{A} = \mathbf{E} + \mathbf{L}\delta\mathbf{U} + \delta\mathbf{L}\mathbf{U} + \delta\mathbf{L}\delta\mathbf{U}$$

ma oszacowanie

$$\frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|} \leqslant \epsilon\left(\frac{9}{2}n^3 + \frac{61}{2}n^2 - 18n - 16\right) + O(\epsilon)$$

gdzie ϵ to dokładność maszynowa. Stąd wynika

$$\frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \leqslant \frac{\alpha}{1-\alpha}, \quad \alpha = \epsilon KO(\frac{9}{2}n^3)$$

Inne rozkłady macierzy

- rozkład LU nie jest jedynym przydatnym rozkładem macierzy
- do innych często stosowanych rozkładów należą
  - rozkład Cholesky'ego (Banachiewicza)
  - rozkład SVD
  - rozkład QR

Rozkład Cholesky'ego (Banachiewicza)

Jeżeli macierz układu jest macierzą symetryczną, tzn.

$$a_{ij} = a_{ji}, \quad i, j = 1, \ldots, n$$

i dodatnio określoną

$$\vec{x}^{\mathrm{T}} \mathbf{A} \vec{x} > 0 \text{ dla każdego } \vec{x}$$

to istnieje dla niej bardziej wydajny od LU rozkład na macierze trójkątne

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$$

gdzie **L** to macierz trójkątna dolna

Rozkład Cholesky'ego (Banachiewicza)

Traktując ostatnie równanie jako układ równań ze względu na elementy macierzy **L**, znajdziemy:

$$
\begin{aligned}
l_{ii} &= \left(a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2 \right)^{1/2} \\
l_{ji} &= \frac{1}{l_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} l_{ik} l_{jk} \right), \quad j = i+1, i+2, \ldots, n
\end{aligned}
$$

- liczba operacji o połowę mniejsza od LU
- niezawodność (metoda nie wymaga wyboru elementu podstawowego)
- stabilność numeryczna

Rozkład SVD (ang. *Singular Value Decomposition*)

Twierdzenie

Każdą macierz $\mathbf{A} \in \mathbf{R}^{m \times n}$ *rzędu r możemy przedstawić jako*

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}, \quad \Sigma = \begin{pmatrix} \Sigma_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \in \mathbf{R}^{m \times n}, \quad \Sigma_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_r),$$

gdzie $\mathbf{U} \in \mathbf{R}^{m \times m}$ *i* $\mathbf{V} \in \mathbf{R}^{n \times n}$ *są macierzami ortogonalnymi, a* $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$. *Elementy* σ_i *macierzy* Σ *nazywane są wartościami osobliwymi macierzy* $\mathbf{A}$.

Szkic algorytmu rozkładu SVD

Krok 1 przekształcamy **A** do postaci

$$\mathbf{A} = \mathbf{QCH}^{\mathrm{T}}$$

gdzie **C** to macierz dwudiagonalna, a **Q** i **H** są iloczynami macierzy odpowiadających transformacji Householdera

Krok 2 nadajemy macierzy **C** postać diagonalną,

$$\mathbf{C} = \mathbf{U}'\Sigma'\mathbf{V}'^{\mathrm{T}}$$

gdzie $\mathbf{U}'$ i $\mathbf{V}'$ opisują transformację Givensa

Krok 3 porządkujemy elementy diagonalne macierzami ortogonalnymi $\mathbf{U}''$ i $\mathbf{V}''$, wyrażającymi się poprzez iloczyny macierzy permutacji

$$\Sigma = \mathbf{U}''^{\mathrm{T}}\Sigma'\mathbf{V}''$$

Szkic algorytmu rozkładu SVD

Macierze **U** i **V** rozkładu SVD to po prostu

$$\mathbf{U} = \mathbf{Q}\mathbf{U}'\mathbf{U}'', \quad \mathbf{V} = \mathbf{H}\mathbf{V}'\mathbf{V}''$$

Zastosowania

- do przybliżonych rozwiązań układów z macierzami osobliwymi albo prawie osobliwymi
- do układów niedookreślonych i nadokreślonych
- numeryczny rząd macierzy
- wskaźnik uwarunkowania macierzy

Rozkład QR

$$\mathbf{A} = \mathbf{QR}$$

$$\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{1}, \quad \mathbf{R} \text{ - macierz trójkątna górna}$$

Do wyznaczenia tego rozkładu stosuje się zmodyfikowaną metodę Grama-Schmidta. Polega ona na obliczeniu ciągu macierzy

$$\mathbf{A} = \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(n+1)} = \mathbf{Q},$$

gdzie $\mathbf{A}^{(k)}$ ma postać

$$\mathbf{A}^{(k)} = \left(\vec{q}_1, \ldots, \vec{q}_{k-1}, \vec{a}_k^{(k)}, \ldots, \vec{a}_n^{(k)}\right).$$

Kolumny $\vec{q}_1,\ldots,\vec{q}_{k-1}$ są $k-1$ początkowymi kolumnami macierzy $\mathbf{Q}$, a kolumny $\vec{a}_k^{(k)}, \ldots, \vec{a}_n^{(k)}$ powinny być ortogonalne do $\vec{q}_1,\ldots,\vec{q}_{k-1}$.

Rozkład QR

Ortogonalność w k-tym kroku kolumn od $k+1$ do n względem $\vec{q}_k$ zapewnia się w następujący sposób:

$$\vec{q}_k = \vec{a}_k^{(k)}, \quad d_k = \vec{q}_k^{\mathrm{T}} \vec{q}_k, \quad r_{kk} = 1, \quad \vec{a}_j^{k+1} = \vec{a}_j^{(k)} - r_{jk}\vec{q}_k$$

$$r_{jk} = \frac{\vec{q}_k^{\mathrm{T}} \vec{a}_j^{(k)}}{d_k}, \quad j = k+1, \ldots, n$$

Po n krokach ($k = 1, \ldots, n$) otrzymamy macierze $\mathbf{Q} = \left(\vec{q}_1, \ldots, \vec{q}_n\right)$ i $\mathbf{R} = \left(r_{kj}\right)$ o pożądanych własnościach.

Iteracyjne poprawianie rozwiązań

- rozwiązanie układu równań $\mathbf{A}\vec{x} = \vec{b}$ dowolną metodą bezpośrednią będzie zwykle obarczone pewnym błędem
- błąd ten możemy wykryć, sprawdzając, jak bardzo tzw. wektor reszt

$$\vec{r} = \vec{b} - \mathbf{A}\vec{x}$$

  różni się od zera
- powinniśmy przy tym liczyć $\vec{r}$ z dokładnością większą niż dokładność uzyskanego rozwiązania

Przykład

Układ

$$\begin{pmatrix} 0,99 & 0,70 \\ 0,70 & 0,50 \end{pmatrix} \vec{x} = \begin{pmatrix} 0,54 \\ 0,36 \end{pmatrix}$$

ma rozwiązanie dokładne

$$\vec{x}_{dok} = \begin{pmatrix} 0,80 \\ -0,36 \end{pmatrix}$$

Przykład

Obliczmy najpierw $\vec{r}$ w arytmetyce zmiennopozycyjnej o dwóch miejscach dziesiętnych w mantysie, dokonując zaokrągleń

$$\begin{aligned}\vec{r}(\vec{x}_{dok}) &= \begin{pmatrix} 0,54 \\ 0,36 \end{pmatrix} - \begin{pmatrix} 0,99 & 0,70 \\ 0,70 & 0,50 \end{pmatrix} \begin{pmatrix} 0,80 \\ -0,36 \end{pmatrix} \\ &= \begin{pmatrix} 0,54 - 0,79 + 0,25 \\ 0,38 - 0,56 + 0,18 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\end{aligned}$$

Nie możemy jednak wnioskować stąd, że $\vec{x}_{dok}$ jest dokładnym rozwiązaniem równania.

Przykład

Dla

$$\vec{x}_1 = \begin{pmatrix} 0,02 \\ 0,74 \end{pmatrix}$$

mamy również

$$\begin{aligned} \vec{r}(\vec{x}_1) &= \begin{pmatrix} 0,54 \\ 0,36 \end{pmatrix} - \begin{pmatrix} 0,99 & 0,70 \\ 0,70 & 0,50 \end{pmatrix} \begin{pmatrix} 0,02 \\ 0,74 \end{pmatrix} \\ &= \begin{pmatrix} 0,54 - 0,02 - 0,52 \\ 0,38 - 0,01 - 0,37 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

Przykład

$\vec{x}_1$ rozwiązaniem równania nie jest i różni się dość sporo od rozwiązania dokładnego,

$$\|\vec{x}_{dok} - \vec{x}_1\|_\infty = 1,1$$

Przykład

Policzmy teraz wektory reszt z większą liczbą miejsc dziesiętnych w mantysie. Otrzymamy

$$\begin{aligned}\vec{r}(\vec{x}_{dok}) &= \begin{pmatrix} 0,54 \\ 0,36 \end{pmatrix} - \begin{pmatrix} 0,99 & 0,70 \\ 0,70 & 0,50 \end{pmatrix} \begin{pmatrix} 0,80 \\ -0,36 \end{pmatrix} \\ &= \begin{pmatrix} 0,54 - 0,792 + 0,252 \\ 0,38 - 0,56 + 0,18 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\end{aligned}$$

Przykład

oraz

$$\begin{aligned}\vec{r}(\vec{x}_1) &= \begin{pmatrix} 0,54 \\ 0,36 \end{pmatrix} - \begin{pmatrix} 0,99 & 0,70 \\ 0,70 & 0,50 \end{pmatrix} \begin{pmatrix} 0,02 \\ 0,74 \end{pmatrix} \\ &= \begin{pmatrix} 0,54 - 0,0198 - 0,518 \\ 0,38 - 0,014 - 0,37 \end{pmatrix} = \begin{pmatrix} 0,0022 \\ -0,004 \end{pmatrix}\end{aligned}$$

Dopiero teraz widać, że $\vec{x}_{dok}$ jest rozwiązaniem naszego układu równań, natomiast $\vec{x}_1$ nim nie jest.

Pierwsza poprawka rozwiązania

Szukamy poprawki $\delta\vec{x}$ takiej, że

$$\vec{x} + \delta\vec{x} = \vec{x}_{dok}$$

Ponieważ zachodzi

$$\vec{r} = \vec{b} - \mathbf{A}\vec{x} = \mathbf{A}\vec{x}_{dok} - \mathbf{A}\vec{x} = \mathbf{A}(\vec{x}_{dok} - \vec{x}) = \mathbf{A}\delta\vec{x}$$

wystarczy, że rozwiążemy układ

$$\vec{r} = \mathbf{A}\delta\vec{x}$$

- łatwe, jeżeli dysponujemy już rozkładem LU macierzy **A**
- wymaga n^2 mnożeń i $n^2 - n$ dodawań

Dalsze poprawki

- w rzeczywistych obliczeniach numerycznych nie potrafimy liczyć dokładnie
- zamiast poprawki $\delta\vec{x}$ znajdziemy tylko poprawkę przybliżoną

$$\delta\vec{x} + \delta(\delta\vec{x})$$

- do ulepszonego rozwiązania $\vec{x} + \delta\vec{x}$ możemy znaleźć kolejną poprawkę

Przepis praktyczny

1. rozwiąż układ równań $\mathbf{A}\vec{x}^{(1)} = \vec{b}$ stosując rozkład LU macierzy $\mathbf{A}$
2. oblicz wektor reszt $\vec{r}^{(1)} = \vec{b} - \mathbf{A}\vec{x}^{(1)}$ (w podwójnej precyzji)
3. jeśli $\|\vec{r}^{(1)}\|_\infty \leqslant \|\mathbf{A}\vec{x}^{(1)}\|_\infty u$ (lub $\|\vec{r}^{(1)}\|_\infty \leqslant \|\vec{b}\|_\infty u$), gdzie u to jednostka maszynowa, przerwij obliczenia. Jeżeli nie , to …
4. oblicz $\delta\vec{x}^{(1)}$ i wyznacz $\vec{x}^{(2)} = \vec{x}^{(1)} + \delta\vec{x}^{(1)}$,
5. oblicz $\vec{r}^{(2)} = \vec{b} - \mathbf{A}\vec{x}^{(2)}$ i przejdź ponownie do punktu 3

Przepis praktyczny

- jeżeli macierz układu jest źle uwarunkowana, może się zdarzyć, że metoda ta nie doprowadzi do rozwiązania bliższego dokładnemu
- wtedy należy jest spróbować liczyć wszystkie wielkości w podwójnej precyzji
- w pozostałych przypadkach metoda pozwala na wyznaczenie rozwiązania, którego wektor reszt jest rzędu $u\|\vec{b}\|_\infty$

Metody iteracyjne

- przybliżone metody rozwiązywania układów równań
- startują z pewnego przybliżenia początkowego, które jest stopniowo ulepszane aż do uzyskania dostatecznie dokładnego rozwiązania
- najczęściej stosowane do dużych układów rzadkich, tzn. takich, których macierze zawierają w większości zera

Pojęcia podstawowe

Definicja

Promieniem spektralnym $\rho(\mathbf{A})$ macierzy $\mathbf{A}$ nazywamy liczbę

$$\rho(\mathbf{A}) = \max_{i=1,\ldots,n} |\lambda_i|,$$

przy czym λ_i są wartościami własnymi macierzy $\mathbf{A}$.
Dla dowolnej normy macierzowej zgodnej z normą wektorów obowiązuje

$$|\lambda_i| \leqslant \|\mathbf{A}\|, \quad \text{dla każdego } i = 1, \ldots, n$$

Zatem

$$\rho(\mathbf{A}) \leqslant \|\mathbf{A}\|_p, \quad p = 1, 2, \infty, E$$

Pojęcia podstawowe

Rozważmy ciąg wektorów $\vec{x}^{(0)}, \vec{x}^{(1)}, \ldots, \vec{x}^{(i)}$, określony dla dowolnego wektora $\vec{x}^{(0)}$ w następujący sposób:

$$\vec{x}^{(i+1)} = \mathbf{M}\vec{x}^{(i)} + \vec{w}, \quad i = 0, 1, \ldots,$$

gdzie $\mathbf{M}$ jest pewną macierzą kwadratową, a $\vec{w}$ wektorem

Twierdzenie

Ciąg określony powyższym wzorem przy dowolnym wektorze $x^{(0)}$ jest zbieżny do jedynego punktu granicznego wtedy i tylko wtedy, gdy

$$\rho(\mathbf{M}) < 1$$

Jak konstruować metody iteracyjne?

Należy tak dobrać macierz **M**, aby

- ciąg

$$\vec{x}^{(i+1)} = \mathbf{M}\vec{x}^{(i)} + \vec{w}, \quad i = 0, 1, \ldots$$

  był zbieżny, tzn. $\rho(\mathbf{M}) < 1$

- spełniony był warunek zgodności

$$\vec{x}_{dok} = \mathbf{M}\vec{x}_{dok} + \vec{w}$$

Jak konstruować metody iteracyjne?

Teoretycznie wystarczy wziąć dowolną macierz $\mathbf{M}$ o promieniu spektralnym mniejszym od 1, a następnie wyliczyć $\vec{w}$ ze warunku zgodności

$$\vec{w} = (\mathbf{I} - \mathbf{M})\mathbf{A}^{-1}\vec{b}$$

jednak wymagałoby to wyliczenia macierzy $\mathbf{A}^{-1}$

Jak konstruować metody iteracyjne? - inny sposób

Załóżmy, że

$$\vec{w} = \mathbf{N}\vec{b}, \quad \mathbf{N} \text{ - macierz kwadratowa}$$

Z warunku zgodności mamy

$$\vec{x}_{dok} = \mathbf{M}\vec{x}_{dok} + \vec{w} \;\Rightarrow (\mathbf{A}^{-1} - \mathbf{N} - \mathbf{M}\mathbf{A}^{-1})\vec{b} = \mathrm{o} \;\Rightarrow \mathbf{M} = \mathbf{I} - \mathbf{N}\mathbf{A},$$

co prowadzi do

$$\vec{x}^{(i+1)} = (\mathbf{I} - \mathbf{N}\mathbf{A})\vec{x}^{(i)} + \mathbf{N}\vec{b}$$

Jak konstruować metody iteracyjne?

- rodzina iteracyjna zbieżna dla

$$\rho(\mathbf{I} - \mathbf{NA}) < 1$$

- przy pewnych szczególnych własnościach macierzy układu **A** stosunkowo proste metody wyboru macierzy **N**

Kryteria przydatności metody iteracyjnej

- liczba działań niezbędnych do wykonania
- potrzebna pamięć
- wielkość błędów zaokrągleń
- szybkość zmian wektora błędu

$$\vec{e}^{(i)} = x^{(i)} - \vec{x}_{dok}$$

Może się okazać, że mimo spełnionego warunku zbieżności zagadnienie jest na tyle źle uwarunkowane, że osiągnięcie zadowalającej dokładności w rozsądnym czasie jest niemożliwe.

Przykład

$$\mathbf{M} = \begin{pmatrix} \frac{1}{2} & 1 & & & \\ & \frac{1}{2} & 1 & & \\ & & \frac{1}{2} & \ddots & \\ & & & \ddots & 1 \\ & & & & \frac{1}{2} \end{pmatrix}, \vec{w} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \vdots \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

Mamy tutaj $\rho(\mathbf{M}) = \frac{1}{2}$, a więc dla dowolnego $\vec{x}^{(0)}$ rodzina iteracyjna dąży do $\vec{x}_{dok} = (1, \ldots, 1)^{\mathrm{T}}$.

Przykład

Przyjmijmy $\vec{x}^{(0)} = 0$:

$$\|\vec{e}^{(0)}\|_\infty = 1, \quad \|\vec{e}^{(1)}\|_\infty = \frac{3}{2}, \quad \|\vec{e}^{(2)}\|_\infty = \frac{9}{4}, \ldots$$

Wzrost błędu w początkowych krokach iteracji może uniemożliwić numeryczne wyznaczenie rozwiązania.

Rola błędów zaokrągleń

- w skrajnym przypadku mogą doprowadzić do uzyskania

$$\vec{x}^{(i+1)} = \vec{x}^{(0)}$$

- powstanie ciąg wektorów, który nie jest zbieżny do rozwiązania
- przed taką sytuacją trudno się ustrzec

Metoda Jacobiego

Zapiszmy macierz układu w postaci

$$\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U},$$

gdzie macierze **L**, **D** i **U** to odpowiednio macierz poddiagonalna, diagonalna i naddiagonalna, np.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 4 & 0 & 0 \\ 7 & 8 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix} + \begin{pmatrix} 0 & 2 & 3 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix}$$

Jako macierz **N** wybierzemy

$$\mathbf{N} = \mathbf{D}^{-1}$$

Metoda Jacobiego

Wówczas

$$\begin{aligned} \mathbf{M_J} &= \mathbf{I} - \mathbf{NA} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A} \\ &= \mathbf{I} - \mathbf{D}^{-1}(\mathbf{L} + \mathbf{D} + \mathbf{U}) = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{U}) \end{aligned}$$

Wzór Jacobiego na rodzinę iteracyjną wektorów będzie miał postać

$$\mathbf{D}\vec{x}^{(i+1)} = -(\mathbf{L} + \mathbf{U})\vec{x}^{(i)} + \vec{b}, \quad i = 0, 1, 2, \ldots$$

Metoda Jacobiego

Aby wzór Jacobiego był niezawodny, należy wcześniej (w razie konieczności) tak pozmieniać kolejność równań w układzie $\mathbf{A}\vec{x} = \vec{b}$, aby na diagonali macierzy układu były tylko elementy niezerowe:

1. spośród kolumn z elementem zerowym na diagonali wybieramy tę, w której jest największa liczba zer
2. w kolumnie tej wybieramy element o największym module i tak przestawiamy wiersze, aby znalazł się on na głównej przekątnej; wiersz ustalamy i pomijamy go w dalszych rozważaniach
3. spośród pozostałych kolumn z elementem zerowym na diagonali wybieramy tę o największej liczbie zer i wracamy do punktu 2 aż do usunięcia wszystkich zer z głównej przekątnej

Przykład

Rozważmy macierz

$$\begin{pmatrix} 0 & 0 & 1 & 2 \\ 2 & 1 & 0 & 2 \\ 7 & 3 & 0 & 1 \\ 0 & 5 & 0 & 0 \end{pmatrix}$$

Najwięcej zer znajduje się w kolumnie trzeciej, a element o największym module w tej kolumnie to element a_{13}.

Przykład

Zamieniamy miejscami wiersze 1 i 3, tak, aby element ten znalazł się na diagonali,

$$\begin{pmatrix} 7 & 3 & 0 & 1 \\ 2 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 5 & 0 & 0 \end{pmatrix}$$

Przykład

Zero na diagonali znajduje się jeszcze w kolumnie czwartej, a element o największym module w niej to a_{24}. Zamieniamy więc miejscami wiersze 2 i 4,

$$\begin{pmatrix} 7 & 3 & 0 & 1 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 2 & 1 & 0 & 2 \end{pmatrix}$$

W ten sposób otrzymaliśmy macierz, dla której można zastosować metodę Jacobiego.

Metoda Jacobiego - niezawodności ciąg dalszy

- zamiana wierszy w macierzy gwarantuje jedynie, że będzie istniała macierz odwrotna do macierzy **D**
- spełnienie warunku zbieżności metody Jacobiego, tzn.

$$\rho(-\mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})) < 1$$

  nie jest gwarantowane w każdym przypadku
- można jedynie pokazać, że jest tak zawsze, jeżeli macierz **A** jest silnie diagonalnie dominująca lub silnie diagonalnie dominująca kolumnowo

Metoda Gaussa–Seidla

Rozkładamy macierz układu na sumę macierzy poddiagonalnej, diagonalnej i naddiagonalnej (w razie konieczności odpowiednio przestawiając wiersze)

$$\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$$

Przyjmujemy

$$\mathbf{N} = (\mathbf{D} + \mathbf{L})^{-1}$$

co prowadzi do

$$\mathbf{M}_{GS} = -(\mathbf{D} + \mathbf{L})^{-1}\mathbf{U}$$

Metoda Gaussa–Seidla

Stąd

$$\mathbf{D}\vec{x}^{(i+1)} = -\mathbf{L}\vec{x}^{(i+1)} - \mathbf{U}\vec{x}^{(i)} + b, \quad i = 0, 1, 2, \ldots$$

- na pierwszy rzut oka powyższe równanie wygląda tak, jakby niewiadome występowały po obu stronach jednocześnie
- jednak przy obliczaniu pierwszej współrzędnej szukanego wektora po prawej stronie równania nie wystąpi żadna współrzędna wektora $\vec{x}^{(i+1)}$
- przy obliczaniu $x_2^{(i+1)}$ prawa strona równości będzie zależała tylko od $x_1^{(i+1)}$
- ogólnie, przy obliczaniu kolejnej składowej szukanego wektora będziemy korzystali z wyznaczonych już poprzednio składowych

Niezawodność metody Gaussa–Seidla

- odpowiednie przestawienie wierszy nie gwarantuje w ogólnym przypadku spełnienia warunku zbieżności

$$\rho(\mathbf{M}_{GS}) = \rho(-(\mathbf{D}+\mathbf{L})^{-1}\mathbf{U}) < 1$$

- jeżeli potrafimy uzasadnić, że dla danej macierzy $\mathbf{A}$ metoda Jacobiego jest zbieżna oraz macierz $\mathbf{M}_J$ ma nieujemne elementy, to zbieżna jest również metoda Gaussa–Seidla
- zachodzi przy tym

$$\rho(\mathbf{M}_{GS}) < \rho(\mathbf{M}_J) < 1$$

Niezawodność metody Gaussa–Seidla

- zbieżność metody jest gwarantowana, jeśli macierz **A** układu równań jest:
  - symetryczna, dodatnio określona
  - silnie diagonalnie dominująca
  - silnie diagonalnie dominująca kolumnowo

Analiza błędów zaokrągleń

Jeżeli w każdej iteracji zamiast wartości $\mathbf{M}\vec{x}^{(i)} + \vec{w}$ obliczamy

$$\mathbf{M}\vec{x}^{(i)} + \vec{w} + \vec{\delta}^{(i)}, \quad \delta^{(i)} \text{ - błąd zaokrągleń}$$

to

$$\vec{x}^{(i+1)} = \mathbf{M}^{i+1}\vec{x}^{(0)} + \mathbf{M}^{i}\vec{w} + \ldots + \vec{w} + \vec{\delta}^{(i)} + \mathbf{M}\vec{\delta}(i-1) + \ldots + \mathbf{M}^{i}\vec{\delta}^{(0)}$$

Łączny błąd zaokrągleń wynosi

$$\vec{x}_{dok}^{(i+1)} - \vec{x}^{(i+1)} = \vec{\delta}^{(i)} + \mathbf{M}\vec{\delta}^{(i-1)} + \ldots + \mathbf{M}^{i}\vec{\delta}^{(0)}.$$

Jeżeli algorytm iteracyjny jest zbieżny i indeks iteracji jest dostatecznie duży, możemy przyjąć

$$\frac{1}{2}\|\vec{x}_{dok}\| < \|\vec{x}^{(j)}\| < 2\|\vec{x}_{dok}\|$$

Analiza błędów zaokrągleń

Stąd wynika, że jeżeli $\vec{x}_{dok} \neq 0$, to

$$\frac{\|\vec{x}_{dok}^{(i+1)} - \vec{x}^{(i+1)}\|}{\|\vec{x}^{(i+1)}\|} \leqslant \frac{2}{\|\vec{x}_{dok}\|}(\|\vec{\delta}^{(i)}\| + \|\mathbf{M}\| \cdot \|\vec{\delta}^{(i-1)}\| + \ldots + \|\mathbf{M}\|^i \cdot \|\vec{\delta}^{(0)}\|)$$

czyli

$$\frac{\|\vec{x}_{dok}^{(i+1)} - \vec{x}^{(i+1)}\|}{\|\vec{x}^{(i+1)}\|} \leqslant \frac{2\kappa}{\|\vec{x}_{dok}\|}(1 + \|\mathbf{M}\| + \ldots + \|\mathbf{M}\|^i)$$

gdzie κ to wspólne oszacowanie błędów $\vec{\delta}^{(j)}$, tzn.

$$\|\vec{\delta}^{(j)}\| < \kappa, \quad j = 0, 1, 2, \ldots, i$$

Analiza błędów zaokrągleń

Jeśli $\|\mathbf{M}\| < 1$, to

$$\frac{\|\vec{x}_{dok}^{(i+1)} - \vec{x}^{(i+1)}\|}{\|\vec{x}^{(i+1)}\|} < \frac{1}{1 - \|\mathbf{M}\|} \frac{2\kappa}{\|\vec{x}_{dok}\|}$$

Analiza błędów zaokrągleń

Gdy macierz układu jest macierzą silnie diagonalnie dominującą, można pokazać, że

$$\frac{\|\vec{x}_{dok}^{(i+1)} - \vec{x}^{(i+1)}\|_\infty}{\|\vec{x}^{(i+1)}\|_\infty} \leqslant \frac{1}{1 - \|\mathbf{M}_{GS}\|_\infty} \frac{12\alpha}{1 - \alpha}$$

gdzie

$$\alpha = \epsilon O(2n^2) \|\mathbf{D}\|_\infty \|\mathbf{D}^{-1}\|_\infty$$

$\Rightarrow$ z porównania powyższego oszacowania z błędem eliminacji Gaussa wynika, że stosując metodę Gaussa–Seidla można zyskać na dokładności, jeżeli tylko wskaźnik $\|\mathbf{D}\|_\infty \|\mathbf{D}^{-1}\|_\infty$ jest mały w porównaniu ze wskaźnikiem uwarunkowania K_∞

Nakłady obliczeń

- w każdej iteracji wykonujemy około n^2 mnożeń (jeżeli macierz układu nie jest rzadka)
- dla porównania, metody dokładne wymagają około $\frac{1}{3}n^3$ mnożeń do uzyskania rozwiązania
- aby metody dokładne i iteracyjne były porównywalne pod względem nakładu obliczeń, powinniśmy wykonać tylko około n iteracji
- proste przykłady pokazują, że liczba iteracji musi być dużo większa niż n, aby dokładność była zadowalająca

Metody iteracyjne w przypadku ogólnym są nieefektywne!

Przykład

Układ

$$\begin{pmatrix} 1 & \frac{3}{4} \\ \frac{3}{4} & 1 \end{pmatrix} \vec{x} = \begin{pmatrix} 448 \\ 448 \end{pmatrix}$$

ma rozwiązanie $x = (256, 256)^{\mathrm{T}}$. Stosując np. eliminację Gaussa, musimy wykonać 6 mnożeń, aby otrzymać wynik dokładny. Jeżeli zastosujemy metodę Gaussa–Seidla, po ośmiu iteracjach (32 mnożenia) mamy

$$\|x^{(8)} - \vec{x}_{dok}\| > 0,1$$

Warunki przerwania obliczeń

- niezbędną do uzyskania zaplanowanej dokładności liczbę iteracji trudno jest przewidzieć
- w praktyce nie zakłada się konkretnej liczby iteracji z góry
- zamiast tego stosuje się testy na przerwanie obliczeń (tzw. testy stopu):

$$\|\vec{x}^{(i+1)} - \vec{x}^{(i)}\| < \Delta$$

$$\frac{1}{\|\vec{b}\|}\|\mathbf{A}\vec{x}^{(i+1)} - \vec{b}\| < \Delta$$

gdzie Δ to żądana dokładność

„Niedoskonałości” testów stopu

- jeżeli norma macierzy **A** jest mała, to wartość reszty

$$\|\mathbf{A}\vec{x}^{(i+1)} - \vec{b}\| = \|\mathbf{A}(\vec{x}^{(i+1)} - \vec{x}_{dok})\| \leqslant \|\mathbf{A}\| \cdot \|\vec{x}^{(i+1)} - \vec{x}_{dok}\|$$

  może być mała, mimo dużego odchylenia wektora $\vec{x}^{(i+1)}$ od rozwiązania dokładnego $\vec{x}_{dok}$
- ponieważ

$$\|\vec{x}^{(i+1)} - \vec{x}^{(i)}\| = \|\vec{e}^{(i+1)} - \vec{e}^{(i)}\| = \|\mathbf{M}\vec{e}^{(i)} - \vec{e}^{(i)}\| = \|(\mathbf{M} - \mathbf{I})\vec{e}^{(i)}\|$$

  gdy norma macierzy $\mathbf{M} - \mathbf{I}$ jest mała, wektory $\vec{x}^{(i+1)}$ i $\vec{x}^{(i)}$ mogą się mało różnić, mimo że błąd $\vec{e}^{(i)}$ jest duży
- testy mogą się okazać mało przydatne z powodu błędów zaokrągleń (wektory reszt należy zawsze liczyć z dużą dokładnością)

Niedookreślone układy równań ($m < n$)

- liczba równań m jest mniejsza od liczby niewiadomych n
- dość często spotykane w praktyce (np. w zagadnieniach optymalizacji)
- nie są one często dyskutowane w literaturze poświęconej metodom numerycznym
- nigdy nie są rozwiązywalne jednoznacznie
  - jeżeli wektor wyrazów wolnych $\vec{b}$ należy do przestrzeni rozpinanej przez kolumny macierzy **A**, wówczas układ

$$\mathbf{A}\vec{x} = \vec{b}$$

    będzie miał nieskończenie wiele rozwiązań
  - w przeciwnym wypadku rozwiązań nie będzie wcale

Niedookreślone układy równań ($m < n$)

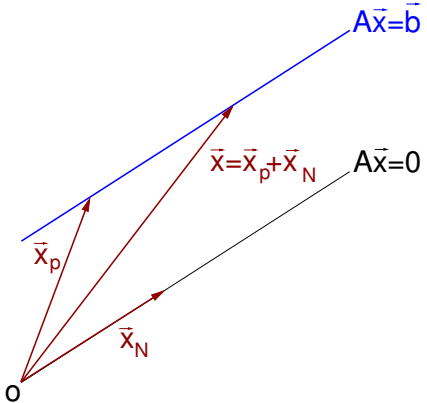
- jeżeli rząd macierzy **A** jest równy liczbie równań, wówczas $\vec{b}$ zawsze będzie należał do przestrzeni rozpinanej przez **A** (układ będzie rozwiązywalny)
- ogólne rozwiązanie takiego układu zapisze się w postaci:

$$\vec{x} = \vec{x}_p + \vec{x}_N$$

gdzie $\vec{x}_p$ jest specjalnym rozwiązaniem równania, a $\vec{x}_N$ należy do jądra przekształcenia liniowego **A**

$$\mathbf{A}\vec{x}_N = \mathrm{o}$$

Graficzna interpretacja rozwiązania dla $m = 1$ i $n = 2$



Rozwiązanie szczególne $\vec{x}_p$

Twierdzenie

Jeżeli macierz $\mathbf{A} \in \mathbf{R}^{m \times n}$ ma rząd m, układ $\mathbf{A}\vec{x} = \vec{b}$ jest zawsze rozwiązywalny. Dla każdego $\vec{b}$ istnieje wówczas nieskończenie wiele rozwiązań, z których

$$\vec{x}_p = \mathbf{A}^{\mathrm{T}} \left(\mathbf{A}\mathbf{A}^{\mathrm{T}} \right)^{-1} \vec{b}$$

jest tym o najmniejszej normie. Macierz $\mathbf{A}^{\mathrm{T}} \left(\mathbf{A}\mathbf{A}^{\mathrm{T}} \right)^{-1}$ nazywana jest przy tym macierzą pseudoodwrotną macierzy $\mathbf{A}$.

Rozwiązanie szczególne $\vec{x}_p$

Dowód.
Dla każdego $\vec{x}$ zachodzi

$$\vec{x}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\vec{x} = \left(\mathbf{A}^{\mathrm{T}}\vec{x}\right)^{\mathrm{T}}\left(\mathbf{A}^{\mathrm{T}}\vec{x}\right) = \|\mathbf{A}^{\mathrm{T}}\vec{x}\|^2 \geqslant 0.$$

Ponadto, jeśli $\mathrm{rank}\mathbf{A} = m$, to $\|\mathbf{A}^{\mathrm{T}}\vec{x}\| = 0$ wtedy i tylko wtedy, gdy $\vec{x} = 0$. Czyli macierz $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ jest dodatnio określona i nieosobliwa. Ponieważ

$$\mathbf{A}\vec{x}_p = \mathbf{A}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{-1}\vec{b} = \vec{b},$$

więc x_p rzeczywiście jest rozwiązaniem równania $\mathbf{A}\vec{x} = \vec{b}$. Pozostaje nam pokazać, że każde inne rozwiązanie ma normę większą od $\|\vec{x}_p\|$. □

Rozwiązanie szczególne $\vec{x}_p$

Niech $\vec{x}$ będzie innym rozwiązaniem naszego układu. Wówczas

$$\|\vec{x}\|^2 = \|\vec{x}_p + (\vec{x} - \vec{x}_p)\|^2 = \|\vec{x}_p\|^2 + \|\vec{x} - \vec{x}_p\|^2 + 2\vec{x}_p^{\mathrm{T}}(\vec{x} - \vec{x}_p).$$

Ponieważ z założenia $\mathbf{A}\vec{x}_p = \mathbf{A}\vec{x}$, trzeci wyraz w powyższym równaniu jest równy zero:

$$\vec{x}_p^{\mathrm{T}}(\vec{x} - \vec{x}_p) = \left[\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\vec{b}\right]^{\mathrm{T}}(\vec{x} - \vec{x}_p) = \vec{b}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{A}(\vec{x} - \vec{x}_p) = 0.$$

Stąd

$$\|\vec{x}\|^2 = \|\vec{x}_p\|^2 + \|\vec{x} - \vec{x}_p\|^2 \geqslant \|\vec{x}_p\|^2,$$

przy czym równość zachodzi tylko dla $\vec{x} = \vec{x}_p$. □

Układy niedookreślone - przykład

Rozważmy układ

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3$$

czyli

$$x_1 + 2x_2 = 3, \quad x_2 = -\frac{1}{2}x_1 + \frac{3}{2}$$

Dowolne z tych dwóch wyrażeń jest rozwiązaniem układu. Rozwiązaniem o najmniejszej normie będzie

$$\vec{x}_p = \mathbf{A}^{\mathrm{T}} \left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{-1} \vec{b} = \begin{pmatrix} 0,6 \\ 1,2 \end{pmatrix}$$

Układy niedookreślone - przykład

Wektory należące do jądra przekształcenia **A** będą miały postać

$$\mathbf{A}\vec{x}_N = 0 \quad \rightarrow \quad x_{N2} = -\frac{1}{2}x_{N1}$$

więc ogólne rozwiązanie jest następujące:

$$\vec{x} = \begin{pmatrix} 0,6 \\ 1,2 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -0,5 \end{pmatrix}$$

gdzie α jest dowolną liczbą rzeczywistą.

Macierz pseudoodwrotna i rozkład Cholesky'ego

Ponieważ $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ jest macierzą symetryczną i dodatnio określoną, możemy rozłożyć ją na iloczyn dwóch macierzy trójkątnych

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$$

Teraz wystarczy rozwiązać układy równań

$$\begin{aligned} \mathbf{L}\vec{w} &= \vec{b} \\ \mathbf{L}^{\mathrm{T}}\vec{z} &= \vec{w} \end{aligned}$$

i na tej podstawie wyliczyć $\vec{x}_p$

$$\vec{x}_p = \mathbf{A}^{\mathrm{T}}\vec{z}$$

Macierz pseudoodwrotna i rozkład SVD

Z równości

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}$$

wynika

$$\begin{aligned}
\vec{x}_p &= \mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{-1}\vec{b} = \mathbf{V}\Sigma\mathbf{U}^{\mathrm{T}}\left(\mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}\mathbf{V}\Sigma\mathbf{U}^{\mathrm{T}}\right)^{-1}\vec{b} \\
&= \mathbf{V}\Sigma\mathbf{U}^{\mathrm{T}}\left(\mathbf{U}\Sigma\Sigma\mathbf{U}^{\mathrm{T}}\right)^{-1}\vec{b} = \mathbf{V}\Sigma\mathbf{U}^{\mathrm{T}}\left(\mathbf{U}^{\mathrm{T}}\right)^{-1}\Sigma^{-1}\Sigma^{-1}\mathbf{U}^{-1}\vec{b} \\
&= \mathbf{V}\Sigma^{-1}\mathbf{U}^{\mathrm{T}}\vec{b}
\end{aligned}$$

Macierz pseudoodwrotna i rozkład SVD

Zapisując rozkład SVD w postaci

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^{\mathrm{T}}, \quad r = \mathrm{rank}\mathbf{A},$$

gdzie $\vec{u}_i$ i $\vec{v}_i$ to kolumny macierzy **U** i **V**, otrzymamy

$$\vec{x}_p = \sum_{i=1}^{r} \frac{\vec{u}_i^{\mathrm{T}} \vec{b}}{\sigma_i} \vec{v}_i.$$

Macierz pseudoodwrotna i rozkład QR

Jeżeli dysponujemy rozkładem QR macierzy $\mathbf{A}^{\mathrm{T}}$

$$\mathbf{A}^{\mathrm{T}} = \mathbf{QR},$$

wówczas

$$\begin{aligned} \vec{x}_p &= \mathbf{A}^{\mathrm{T}} \left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{-1} \vec{b} = \mathbf{QR} \left(\mathbf{R}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{QR}\right)^{-1} \vec{b} \\ &= \mathbf{QR} \left(\mathbf{R}^{\mathrm{T}}\mathbf{R}\right)^{-1} \vec{b} = \mathbf{QRR}^{-1} \left(\mathbf{R}^{\mathrm{T}}\right)^{-1} \vec{b} \\ &= \mathbf{Q} \left(\mathbf{R}^{\mathrm{T}}\right)^{-1} \vec{b} \end{aligned}$$

Aby wyliczyć $\vec{x}_p$, musimy wyznaczyć $\left(\mathbf{R}^{\mathrm{T}}\right)^{-1} \vec{b}$. Ale to nic innego, jak rozwiązanie równania trójkątnego

$$\mathbf{R}^{\mathrm{T}}\vec{z} = \vec{b}$$

Nadokreślone układy równań ($m > n$)

- równań (m) jest więcej niż niewiadomych (n)
- w zależności od wektora wyrazów wolnych nie ma rozwiązań, jest ich nieskończona liczba lub tylko jedno rozwiązanie jednoznaczne
- w praktyce najczęściej dokładne rozwiązanie układu nie istnieje, ale możliwe jest na ogół znalezienie rozwiązania przybliżonego (np. regresja liniowa)

Nadokreślone układy równań ($m > n$)

- równania $\mathbf{A}\vec{x} = \vec{b}$ dla macierzy $\mathbf{A} \in \mathbf{R}^{m \times n}$ przy $m > n$ nie można rozwiązać uniwersalnie, ponieważ rząd tej macierzy jest mniejszy od m
- rozwiązanie dokładne nie istnieje w ogóle, gdy wektor $\vec{b}$ nie należy do przestrzeni rozpinanej przez kolumny macierzy układu
- w tym przypadku zadany układ można potraktować jak zadanie aproksymacyjne i poszukać takiego $\vec{x}$, który zminimalizuje kwadrat normy wektora błędu

$$\vec{e} = \mathbf{A}\vec{x} - \vec{b}.$$

- takie przybliżone rozwiązanie może okazać się bardzo użyteczne w wielu praktycznych zagadnieniach
- to nic innego jak metoda najmniejszych kwadratów

Rozwiązanie układu nadokreślonego

Szukamy minimum wyrażenia

$$J = \frac{1}{2}\|\vec{e}\|_2^2 = \frac{1}{2}\|\mathbf{A}\vec{x} - \vec{b}\|_2^2 = \frac{1}{2}\left(\mathbf{A}\vec{x} - \vec{b}\right)^{\mathrm{T}}\left(\mathbf{A}\vec{x} - \vec{b}\right)$$

Z warunku na istnienie minimum,

$$\frac{\partial}{\partial \vec{x}} J = \mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\vec{x} - \vec{b}\right) = \mathrm{o}$$

znajdziemy

$$\vec{x}_p = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\vec{b}$$

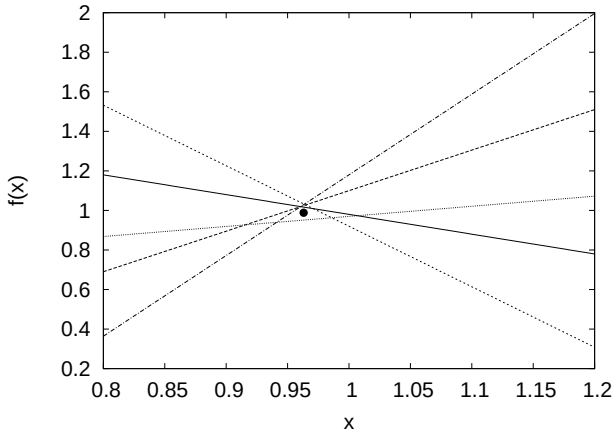
- do obliczenia macierzy pseudoodwrotnej do macierzy **A** możemy znowu wykorzystać rozkłady SVD lub QR macierzy **A**

Układ nadokreślony - przykład

Rozważmy układ

$$\begin{aligned} x + y &= 1,98 \\ 2,05 * x - y &= 0,95 \\ 3,06 * x + y &= 3,98 \\ -1,02 * x + 2 * y &= 0,92 \\ 4,08 * x - y &= 2,90 \end{aligned}$$

Układ nadokreślony - przykład



Układ nadokreślony - przykład

- rozwiązanie ma prostą interpretację geometryczną - to punkt przecięcia prostych zdefiniowanych poszczególnymi równaniami
- dokładne rozwiązanie układu nie istnieje
- rozwiązanie przybliżone wynosi

$$\vec{x}_p = \begin{pmatrix} 0,963101 \\ 0,988543 \end{pmatrix}$$

- błąd przybliżenia

$$\|\mathbf{A}\vec{x}_p - \vec{b}\|_2 = 0,10636$$