Metody numeryczne

Uwarunkowanie problemu numerycznego

Eliminacja Gaussa

P. F. Góra

https://zfs.fais.uj.edu.pl/pawel_gora

14 października 2025

Uwarunkowanie zadania numerycznego

Niech $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ będzie pewną funkcją odpowiednio wiele razy różniczkowalną i niech $\mathbf{x} \in \mathbb{R}^n$.

Definicja: Mówimy, że zagadnienie obliczenia $\varphi(\mathbf{x})$ jest *numerycznie dobrze uwarunkowane*, jeżeli niewielkie względne zmiany danych dają niewielkie względne zmiany rozwiązania. Zagadnienia, które nie są numerycznie dobrze uwarunkowane, nazywamy źle uwarunkowanymi.

Przykład 1

Rozważmy problem znalezienia rozwiązań równania

$$x^2 + bx + c = 0\,, \tag{1}$$

przy czym zakładamy, że $b^2 - 4c > 0$. Wiadomo, że rozwiązania mają w tym wypadku postać

$$x_{1,2} = \tfrac{1}{2}\left(-b \pm \sqrt{b^2 - 4c}\right)\,. \tag{2}$$

Jak dobrze uwarunkowane jest zagadnienie obliczania (2)? *Danymi* są tu współczynniki trójmianu, b, c. Zaburzmy te współczynniki: $b \to b + \varepsilon_2$, $c \to c + \varepsilon_3$.

Rozwiązaniami są teraz

$$\begin{aligned} \bar{x}_{1,2} &= \tfrac{1}{2}\left(-b+\varepsilon_2 \pm \sqrt{(b+\varepsilon_2)^2-4(c+\varepsilon_3)}\right) \\ &\simeq \tfrac{1}{2}\left(-b \pm \sqrt{b^2-4c}+\varepsilon_2 \pm \frac{2b\varepsilon_2-4\varepsilon_3}{2\sqrt{b^2-4c}}\right), \end{aligned} \tag{3}$$

gdzie dokonaliśmy rozwinięcia Taylora do pierwszego rzędu w $\varepsilon_{1,2}$. Widzimy, że błąd względny

$$\left|\frac{\bar{x}_{1,2}-x_{1,2}}{x_{1,2}}\right| \tag{4}$$

rośnie nieograniczenie, gdy $b^2-4c \to 0^+$. Problem wyznaczania pierwiastków trójmianu (1) jest wówczas numerycznie źle uwarunkowany. Problem ten jest dobrze uwarunkowany, gdy $b^2-4c \gg 0$.

Potrzebujemy jakiejś *miary* uwarunkowania problemu numerycznego.

Norma wektora

Niech $\mathcal{V}$ będzie pewną przestrzenią wektorową nad ciałem $\mathbb{C}$ (lub $\mathbb{R}$). *Normą wektora* nazywam funkcję $\|\cdot\| : \mathcal{V} \to \mathbb{R}$, spełniającą następujące warunki ($\mathbf{x}, \mathbf{y} \in \mathcal{V}$):

1. $\|\mathbf{x}\| \geqslant 0 \quad \wedge \quad \|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = 0.$
2. $\|\alpha\,\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\| \quad \alpha \in \mathbb{C}.$
3. $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|.$

Mówiąc niezbyt prezycyjnie, norma jest uogólnieniem pojęcia wartości bezwzględnej na przypadek wektorów. Norma jest miarą długości wektora — ale można ją zdefiniować na wiele sposobów.

Przykłady norm wektorów

W naszych rozważaniach przestrzeń liniowa $\mathcal{V}$ najczęściej będzie przestrzenią $\mathbb{R}^n$. Można w niej definiować wiele (różnych) norm. Najczęściej używa się jednej z trzech:

- Norma taksówkowa:

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n| \tag{5a}$$

- Norma Euklidesowa:

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{5b}$$

- Norma maximum (*worst offender*):

$$\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} |x_i| \tag{5c}$$

Jeżeli nie zaznaczymy inaczej, przez normę wektorową będziemy rozumieć normę Euklidesową.

Norma Euklidesowa jest zadana przez iloczyn skalarny: Dla $\mathbf{x} \in \mathbb{R}^n$

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}\,. \tag{6}$$

Dla przypadku zespolonego $\mathbf{x} \in \mathbb{C}^n$ odpowiednik normy Euklidesowej *także* jest zadany przez iloczyn skalarny:

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\dagger \mathbf{x}} = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}\,. \tag{7}$$

Metryka i kula otwarta

Weźmy przestrzeń wektorową, w której zdefiniowano normę i weźmy dwa wektory, $\mathbf{x}, \mathbf{y}$, z tej przestrzeni. Ich odległością (metryką) jest*

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| . \tag{8}$$

Niech $\mathcal{V}$ będzie przestrzenią metryczną z metryką d. *Kulą otwartą* o środku w punkcie P i promieniu r nazywam zbiór punktów

$$\{X \in \mathcal{V} : d(P, X) < r\} \tag{9}$$

*Nie jest to najbardziej ogólna definicja metryki, ale na potrzeby tegoi wykładu wystarczy.

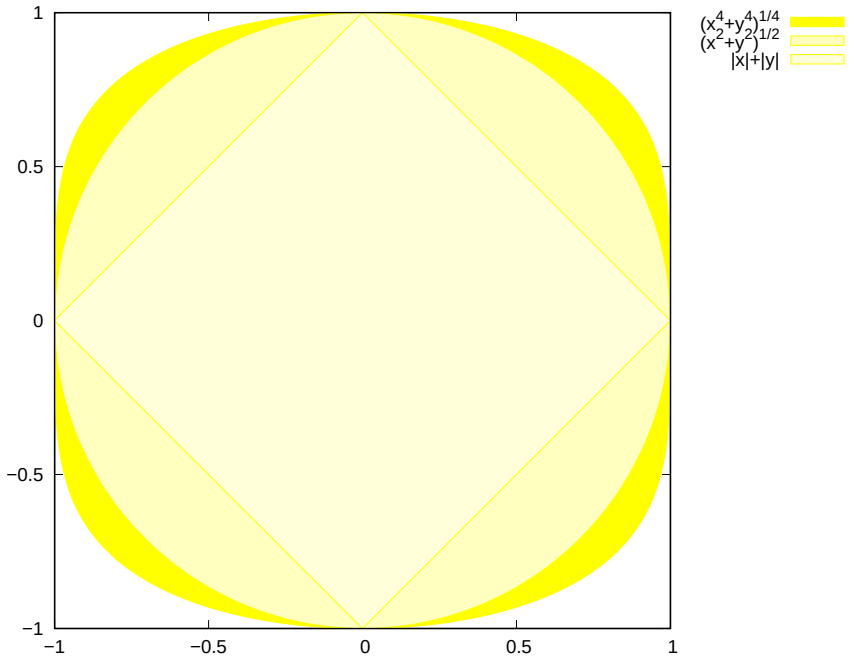
Przykład 2

W przestrzeni $\mathbb{R}^2$ kulą otwartą o środku w punkcie $P(x_0, y_0)$ i promieniu r jest wnętrze okręgu

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \tag{10}$$

W tej samej przestrzeni z metryką Manhattan kulą otwartą są punkty spełniające

$$|x - x_0| + |y - y_0| < r \tag{11}$$



Kule jednostkowe na płaszczyźnie w różnych metrykach: Manhattan, euklidesowej, $\sqrt[4]{x^4 + y^4}$. *Cały* kwadrat jednostkowy jest kulą w normie maksimum.

Współczynnik uwarunkowania

Niech $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ będzie pewną funkcją, $\mathbf{x} \in \mathbb{R}^n$ dokładną wartością argumentu, a $\bar{\mathbf{x}} \in \mathbb{R}^n$ znanym numerycznym przybliżeniem $\mathbf{x}$.

Definicja: Jeżeli istnieje $\kappa \in \mathbb{R}$ taka, że

$$\forall \mathbf{x}, \bar{\mathbf{x}}: \qquad \frac{\|\varphi(\mathbf{x}) - \varphi(\bar{\mathbf{x}})\|_{\mathbb{R}^m}}{\|\varphi(\mathbf{x})\|_{\mathbb{R}^m}} \leqslant \kappa \cdot \frac{\|\mathbf{x} - \bar{\mathbf{x}}\|_{\mathbb{R}^n}}{\|\mathbf{x}\|_{\mathbb{R}^n}} \tag{12}$$

nazywamy ją *współczynnikiem uwarunkowania* zagadnienia wyliczenia wartości $\varphi(\cdot)$ (względem zadanych norm).

Współczynnik uwarunkowania mówi jak bardzo błąd względny wyniku obliczeń “przekracza” błąd względny samej różnicy przybliżenia i wartości dokładnej. Spodziewamy się, że jeżeli przybliżenie znacznie różni się od wartości dokładnej, także wyniki obliczeń będą się znacznie różnić. W zagadnieniach numerycznie źle uwarunkowanych *może* się zdarzyć, że nawet niewielkie odchylenie przybliżenia od wartości dokładnej doprowadzi do znacznej różnicy wyników.

Układy równań liniowych

Niech $\mathbf{A} \in \mathbb{R}^{n\times n}$ będzie macierzą, $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$. Rozpatrujemy równanie

$$\mathbf{A}\mathbf{x} = \mathbf{b}\,, \tag{13}$$

Zakładamy, że macierz $\mathbf{A}$ oraz wektor wyrazów wolnych $\mathbf{b}$ są znane. Poszukujemy wektora $\mathbf{x}$. Równanie (13) jest równoważne następującemu układowi równań liniowych:

$$\left\{\begin{array}{ccccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \ldots & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \ldots & + & a_{2n}x_n & = & b_2 \\
a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \ldots & + & a_{3n}x_n & = & b_3 \\
\vdots & & \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\
a_{n1}x_1 & + & a_{n2}x_2 & + & a_{n3}x_3 & + & \ldots & + & a_{nn}x_n & = & b_n
\end{array}\right. \tag{14}$$

gdzie a_{ij} są elementami macierzy $\mathbf{A}$, natomiast x_j, b_j są elementami wektorów, odpowiednio, $\mathbf{x}, \mathbf{b}$.

Rozwiązywanie układów równan liniowych rzadko stanowi “samoistny” problem numeryczny. Zagadnienie to występuje jednak **bardzo często** jako pośredni etap wielu problemów obliczeniowych. Dlatego też dogłębna znajomość algorytmów numerycznego rozwiązywania układów równań liniowych jest niezwykle ważna.

Rozwiązywalność układów równań liniowych

Układ równań (13) ma jednoznaczne rozwiązanie wtedy i tylko wtedy, gdy

$$\det \mathbf{A} \neq 0\,. \tag{15}$$

Z elementarnej algebry wiadomo, że rozwiązania można wówczas skonstruować posługując się *wzorami Cramera*. Uwaga: Numeryczne korzystanie ze wzorów Cramera jest **koszmarnie drogie** i dlatego w praktyce korzystamy z innych algorytmów.

Jak dobrze uwarunkowane jest zagadnienie rozwiązania równania (13)?

Przykład 3

Rozważmy następujące układy równań:

$$\begin{cases} 2x + 6y & = 8 \\ 2x + 6.00001y & = 8.00001 \end{cases} \qquad \begin{cases} 2x + 6y & = 8 \\ 2x + 5.99999y & = 8.00002 \end{cases}$$

Współczynniki tych układów równan różnią się co najwyżej o $0.00002 = 2 \cdot 10^{-5}$. Rozwiązaniem pierwszego są liczby $(1, 1)$, drugiego — liczby $(10, -2)$. Widzimy, że mała zmiana współczynników powoduje, że różnica rozwiązań jest $\sim 10^6$ razy większa, niż zaburzenie współczynników. Powyższe układy równań są źle uwarunkowane.

Norma macierzy

Niech $\mathbf{A} \in \mathbb{R}^{N \times N}$. *Normą macierzy* (indukowaną) nazywam

$$\|\mathbf{A}\| = \max\left\{\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} : \mathbf{x} \in \mathbb{R}^N, \mathbf{x} \neq 0\right\} = \max_{\|\mathbf{x}\|=1} \{\|\mathbf{A}\mathbf{x}\|\} \tag{16}$$

Promieniem spektralnym macierzy $\mathbf{A} \in \mathbb{R}^{N \times N}$ nazywam

$$\rho = \sqrt{\left\|\mathbf{A}\mathbf{A}^T\right\|} \tag{17}$$

W przestrzeniach skończeniewymiarowych promień spektralny macierzy jest równy jej normie.

Współczynnik uwarunkowania układu równań liniowych

Rozwiązujemy układ rownań ($\det \mathbf{A} \neq 0$)

$$\mathbf{A}\mathbf{y} = \mathbf{b} \tag{18a}$$

Przypuśćmy, że wyraz wolny $\mathbf{b}$ jest obarczony jakimś błędem $\boldsymbol{\Delta}\mathbf{b}$, czyli rozwiązujemy

$$\mathbf{A}\tilde{\mathbf{y}} = \mathbf{b} + \boldsymbol{\Delta}\mathbf{b} \tag{18b}$$

Zauważmy, że $\tilde{\mathbf{y}} - \mathbf{y} = \mathbf{A}^{-1}\left(\mathbf{b} + \boldsymbol{\Delta}\mathbf{b}\right) - \mathbf{A}^{-1}\mathbf{b} = \mathbf{A}^{-1}\boldsymbol{\Delta}\mathbf{b}$.

Jak błąd wyrazu wolnego wpływa na rozwiązanie? Obliczamy

$$\frac{\|\tilde{\mathbf{y}}-\mathbf{y}\|}{\|\mathbf{y}\|}=\frac{\left\|\mathbf{A}^{-1}\boldsymbol{\Delta}\mathbf{b}\right\|}{\|\mathbf{y}\|}\leqslant\frac{\left\|\mathbf{A}^{-1}\right\|\cdot\|\boldsymbol{\Delta}\mathbf{b}\|}{\|\mathbf{y}\|} \tag{19a}$$

Z drugiej strony

$$\|\mathbf{b}\|=\|\mathbf{A}\mathbf{y}\|\leqslant\|\mathbf{A}\|\cdot\|\mathbf{y}\|$$

skąd wynika, że

$$\frac{1}{\|\mathbf{y}\|}\leqslant\frac{\|\mathbf{A}\|}{\|\mathbf{b}\|} \tag{19b}$$

Ostatecznie

$$\frac{\|\tilde{\mathbf{y}}-\mathbf{y}\|}{\|\mathbf{y}\|}\leqslant\underbrace{\left\|\mathbf{A}\right\|\cdot\left\|\mathbf{A}^{-1}\right\|}_{\kappa}\cdot\frac{\|\boldsymbol{\Delta}\mathbf{b}\|}{\|\mathbf{b}\|} \tag{19c}$$

Współczynnik uwarunkowania macierzy symetrycznej, rzeczywistej

Niech $\mathbf{A} \in \mathbb{R}^{N\times N}$ będzie odwracalną macierzą symetryczną, rzeczywistą. W takim wypadku jej wartości własne są rzeczywiste a jej unormowane wektory własne $\{\mathbf{e}_i\}_{i=1}^N$ stanowią bazę ortogonalną w $\mathbb{R}^N$. Oznaczmy wartości własne tej macierzy przez $\{\lambda_i\}_{i=1}^N$. Weźmy dowolny $\mathbf{x} \in \mathbb{R}^N$ taki, że $\|\mathbf{x}\| = 1$. Wówczas

$$\mathbf{x} = \sum_{i=1}^N \alpha_i \mathbf{e}_i \,, \quad \sum_{i=1}^N \alpha_i^2 = 1 \,. \tag{20}$$

$$\begin{aligned} \|\mathbf{A}\mathbf{x}\| &= \left\|\mathbf{A}\sum_{i=1}^N \alpha_i \mathbf{e}_i\right\| = \left\|\sum_{i=1}^N \alpha_i \mathbf{A}\mathbf{e}_i\right\| = \left\|\sum_{i=1}^N \alpha_i \lambda_i \mathbf{e}_i\right\| = \sqrt{\sum_{i=1}^N \alpha_i^2 \lambda_i^2} \\ &\leqslant \sqrt{\sum_{i=1}^N \alpha_i^2 \max_j(\lambda_j^2)} = \max_j |\lambda_j| \cdot \sqrt{\sum_{i=1}^N \alpha_i^2} = \max_j |\lambda_j| \end{aligned} \tag{21}$$

Uwzględniając (16), widzimy, że $\|\mathbf{A}\| = \max_j |\lambda_j|$: norma odwracalnej macierzy symetrycznej, rzeczywistej jest równa największemu modułowi spośród jej wartości własnych.

Rozważmy teraz macierz $\mathbf{A}^{-1}$. Ma ona te same wektory własne, co $\mathbf{A}$, natomiast jej wartości własne są odwrotnościami wartości własnej macierzy nieodwróconej, $\mathbf{A}^{-1}\mathbf{e}_i = \frac{1}{\lambda_i}\mathbf{e}_i$. Postępując jak powyżej, łatwo możemy pokazać, że

$$\left\|\mathbf{A}^{-1}\right\| = \max_j \frac{1}{\left|\lambda_j\right|} = \frac{1}{\min_j \left|\lambda_j\right|}\,. \tag{22}$$

Widzimy zatem, że

Współczynnik uwarunkowania macierzy $\mathbf{A} \in \mathbb{R}^{n \times n}$ wynosi

$$\kappa = \|\mathbf{A}\| \cdot \left\|\mathbf{A}^{-1}\right\| \tag{23}$$

Dla macierzy symetrycznych, rzeczywistych sprowadza się to do

$$\kappa = \frac{\max\limits_i |\lambda_i|}{\min\limits_i |\lambda_i|}, \tag{24}$$

gdzie λ_i oznaczają wartości własne macierzy.

Co można zrobić z układem równań

. . . tak, aby jego rozwiązania się nie zmieniły?

Rozważam układ równań (przykład 3×3 dla oszczędności miejsca):

$$\left\{\begin{array}{rcrcrcl} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & = & b_3 \end{array}\right. \tag{25}$$

1. Równania można zapisać w innej kolejności:

$$\left\{\begin{array}{rcrcrcl} a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = & b_2 \\ a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & b_1 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & = & b_3 \end{array}\right. \tag{26}$$

Odpowiada to **permutacji wierszy macierzy układu równań, z jednoczesną permutacją kolumny wyrazów wolnych**.

2. Równania można dodać stronami, po pomnożeniu przez dowolną stałą różną od zera:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ (z \cdot a_{11} + a_{31})x_1 + (z \cdot a_{12} + a_{32})x_2 + (z \cdot a_{13} + a_{33})x_3 = z \cdot b_1 + b_3 \end{cases} \tag{27}$$

Odpowiada to **zastąpieniu jednego wiersza macierzy układu równań przez dowolną kombinację liniową tego wiersza z innymi, z jednoczesną analogiczną operacją na kolumnie wyrazów wolnych**.

3. We wszystkich równaniach można przestawić kolejność, w jakiej pojawiają się zmienne:

$$\begin{cases} a_{11}x_1 + a_{13}x_3 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{23}x_3 + a_{22}x_2 = b_2 \\ a_{31}x_1 + a_{33}x_3 + a_{32}x_2 = b_3 \end{cases} \tag{28}$$

Odpowiada to **permutacji *kolumn* macierzy układu równań, z jednoczesną permutacją kolumny niewiadomych**.

Eliminacja Gaussa

Rozpatrzmy jeszcze raz układ równań

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & = & b_3 \end{cases} \tag{29}$$

Podzielmy pierwsze równanie stronami przez a_{11}

$$\begin{cases} x_1 & + & \frac{a_{12}}{a_{11}}x_2 & + & \frac{a_{13}}{a_{11}}x_3 & = & \frac{b_1}{a_{11}} \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & = & b_3 \end{cases} \tag{30}$$

Teraz mnożymy pierwsze z równań (30) przez a_{21} i odejmijmy stronami od drugiego, a następnie mnożymy pierwsze z równań (30) przez a_{31} i odejmijmy stronami od trzeciego. Otrzymujemy

$$\begin{cases} x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3 = \frac{b_1}{a_{11}} \\ \left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right)x_2 + \left(a_{23} - \frac{a_{21}a_{13}}{a_{11}}\right)x_3 = b_2 - \frac{a_{21}}{a_{11}}b_1 \\ \left(a_{32} - \frac{a_{31}a_{12}}{a_{11}}\right)x_2 + \left(a_{33} - \frac{a_{31}a_{13}}{a_{11}}\right)x_3 = b_3 - \frac{a_{31}}{a_{11}}b_1 \end{cases} \tag{31a}$$

Przepiszmy to w postaci (tylko zmiana oznaczeń!)

$$\begin{cases} x_1 + a'_{12}x_2 + a'_{13}x_3 = b'_1 \\ a'_{22}x_2 + a'_{23}x_3 = b'_2 \\ a'_{32}x_2 + a'_{33}x_3 = b'_3 \end{cases} \tag{31b}$$

W układzie równań (31b) pierwsza zmienna, x_1, występuje wyłącznie w pierwszym równaniu. Tego równania już nie przekształcamy, natomiast z pozostałymi równaniami postępujemy analogicznie: dzielimy drugie stronami

przez a'_{22} i odpowiednio mnożąc, odejmujemy od trzeciego. Otrzymujemy

$$\begin{cases} x_1 + a'_{12}x_2 + a'_{13}x_3 = b'_1 \\ \qquad\quad x_2 + a''_{23}x_3 = b''_2 \\ \qquad\qquad\quad a''_{33}x_3 = b''_3 \end{cases} \tag{32}$$

Teraz pierwsza zmienna występuje wyłącznie w pierwszym równaniu, druga — w pierwszym i w drugim. Gdyby równań było więcej, moglibyśmy to postępowanie kontynuować.

Ostatecznie otrzymalibyśmy równanie postaci

$$\begin{cases} x_1 + \bullet x_2 + \bullet x_3 + \ldots + \bullet x_N = \tilde{b}_1 \\ \qquad x_2 + \bullet x_3 + \ldots + \bullet x_N = \tilde{b}_2 \\ \qquad\qquad x_3 + \ldots + \bullet x_N = \tilde{b}_3 \\ \qquad\qquad\quad \ldots \qquad\quad \ldots \quad\; \ldots \\ \qquad\qquad\qquad\qquad x_N = \tilde{b}_N \end{cases} \tag{33}$$

gdzie symbole $\bullet$ oznaczają *jakieś* współczynniki, dające się wyliczyć z pierwotnych współczynników równania, $\tilde{b}_i$ są przekształconymi w toku całej procedury wyrazami wolnymi.

Równanie w postaci (33) nazywamy układem równań *z macierzą trójkątną górną*. Algorytm prowadzący od (29) do (33) nazywamy *eliminacją Gaussa*.

Dygresja: Złożoność obliczeniowa

Niech N oznacza liczbę danych wejściowych pewnego algorytmu. Niech $\mathcal{M}(N)$ oznacza liczbę operacji, jaką algorytm ten wykonuje dla N danych. Mówimy, że algorytm ma złożoność obliczeniową $O\left(\mathcal{P}(N)\right)$ jeżeli

$$\exists N_0 \in \mathbb{N},\ A_1, A_2 > 0\ \forall N > N_0 :\ \ A_1 \cdot \mathcal{P}(N) \leqslant \mathcal{M}(N) \leqslant A_2 \cdot \mathcal{P}(N) \tag{34}$$

Dlaczego złożoność obliczeniowa jest ważna?

Intuicyjnie, na skutek doświadczenia codziennego, spodziewamy się, że *skutek jest proporcjonalny do przyczyny*[†]. W kontekście obliczeń numerycznych spodziewamy się zatem, że jeżeli zwiększymy rozmiar problemu p-krotnie, czas działania wydłuży się także p-krotnie. Tak wcale być nie musi.

Przykład 4

Przypuśćmy, że pewien program na danych testowych wykonuje się 1 minutę, a jego złożoność wynosi $O(N^3)$. Zwiększamy rozmiar danych 10-krotnie. Program, zamiast 10 minut, wykonuje się w ciągu $10^3 = 1000$ minut, czyli przez ok. 17 godzin.

[†]Fizyk rozpozna w tym stwierdzeniu prawo Hooke'a.

Złożoność obliczeniowa eliminacji Gaussa

Aby usunąć zmienną x_1 z jednego wiersza, należy wykonać $O(N)$ operacji. Ponieważ zmienną x_1 musimy usunąć z $N-1$ wierszy, musimy łącznie wykonać $O(N^2)$ operacji. Ponieważ musimy to samo zrobić ze zmiennymi $x_2, x_3, \ldots$, ostatecznie musimy wykonać $O(N^3)$ operacji.

Złożoność obliczeniowa eliminacji Gaussa wynosi $O(N^3)$.

Backsubstitution

Rozpatrzmy układ równań w postaci (33). Ostatnie równanie jest rozwiązane ze względu na x_N. Podstawiamy to rozwiązanie do wszystkich poprzednich równań. Teraz drugie od dołu równanie ma tylko jedną nieznaną zmienną — x_{N-1}, a coś takiego umiemy rozwiązać. Podstawiamy to rozwiązanie do równania trzeciego od dołu i do poprzednich. Teraz trzecie od dołu równanie zawiera tylko jedną zmienną, x_{N-2}. Rozwiązujemy, podstawiamy do poprzednich i tak dalej...

Ponieważ wyeliminowanie jednej zmiennej wymaga $O(N)$ operacji, a musimy wyeliminować N zmiennych, cały koszt rozwiązania układu z macierzą trójkątną górną za pomocą algorytmu *backsubstitution* wynosi $O(N^2)$. Jest to *niewiele* w porównaniu z kosztem eliminacji Gaussa.

Całkowity koszt rozwiązania układu N równań liniowych za pomocą eliminacji Gaussa z następującym *backsubstitution* wynosi $O(N^3)$.

Czy coś może pójść źle?

Cały algorytm zawali się, **jeżeli w którymś momencie trzeba będzie wykonać dzielenie przez zero**

$$a_{11} = 0 \text{ lub } a'_{22} = 0, \text{ lub } a''_{33} = 0 \text{ itd.}$$

Przykład 5

Układu równań

$$\begin{cases} & & y & + & z & = & 1 \\ x & + & y & + & z & = & 2 \\ 2x & & & - & z & = & 0 \end{cases} \tag{35}$$

nie da się doprowadzić do postaci trójkątnej górnej za pomocą eliminacji Gaussa. Jeśli jednak przestawimy pierwszy wiersz z drugim lub z trzecim, eliminacja Gaussa powiedzie się.

Ze względów numerycznych staramy się także unikać dzielenia przez liczby bardzo małe co do wartości bezwzględnej. *Formalnie*, w arytmetyce dokładnej, jest to wykonalne, ale *w praktyce* może to doprowadzić do bardzo znacznej utraty dokładnosci, tak, że ostateczny wynik będzie numerycznie bezwartościowy.

Wybór elementu podstawowego

Przypuśćmy, że na pewnym etapie eliminacji Gaussa mamy układ równań

$$\begin{cases} x_1 + \ldots \ldots \ldots \ldots \ldots = b_1 \\ x_2 + \ldots \ldots \ldots \ldots \ldots = b_2 \\ \ldots \ldots \ldots \ldots \ldots \ldots \\ a_{kk}x_k + a_{k,k+1}x_{k+1} + \ldots = b_k \\ a_{k+1,k}x_k + a_{k+1,k+1}x_{k+1} + \ldots = b_{k+1} \\ \ldots \ldots \ldots \ldots \\ a_{Nk}x_k + a_{N,k+1}x_{k+1} + \ldots = b_N \end{cases} \tag{36}$$

"Powinniśmy" teraz dzielić przez a_{kk}. Zamiast tego wśród współczynników $a_{kk}, a_{k+1,k}, a_{k+2,k}, \ldots, a_{Nk}$ wyszukujemy największy co do modułu, permutujemy wiersze tak, aby ten największy co do modułu znalazł się w pozycji diagonalnej i dzielimy przez niego. Współczynnik wypromowany do pozycji diagonalnej nazywa się elementem podstawowym (ang. pivot). Ten krok algorytmu nazywa się *częściowym wyborem elementu podstawowego*. Dalej postępujemy jak poprzednio.

Koszt wyszukania jednego elementu podstawowego wynosi $O(N)$. Jeżeli robimy to w każdym kroku, całkowity koszt jest rzędu $O(N^2)$, a więc jest mały w porównaniu ze złożonością obliczeniową samej eliminacji Gaussa. Wynika z tego, iż częściowego wyboru elementu podstawowego należy zawsze dokonywać, gdyż nie zwiększa to znacznie kosztu całej procedury, może natomiast zapewnić numeryczną stabilność algorytmu.

Zamiast szukać elementu podstawowego wyłącznie w jednej kolumnie, można szukać największego co do modułu współczynnika wśród wszystkich $a_{i,j}, k \leqslant i, j \leqslant N$. Po znalezieniu, należy tak spermutować wiersze i kolumny układu równań, aby element podstawowy znalazł się w pozycji diagonalnej. Nazywa się to *pełnym wyborem elementu podstawowego*. Zauważmy, że koszt numeryczny wynosi $O(N^3)$, a więc staje się porównywalny z kosztem całej eliminacji Gaussa, ponadto zaś permutacja kolumn wymaga późniejszego odwikłania permutacji elementów rozwiązania, co

jest kłopotliwe. Pełny wybór elementu podstawowego zapewnia większą stabilność numeryczną, niż wybór częściowy, ale w praktyce jest rzadko używany, ze wskazanych wyżej powodów.

Do skutecznego przeprowadzenia eliminacji Gaussa potrzebna jest znajomość kolumny wyrazów wolnych, gdyż wyrazy wolne także są przekształcane i permutowane w czasie eliminacji.

Uwagi o eliminacji Gaussa

Przypuśćmy, że mamy rozwiązać kilka układów równań z tą samą lewą stroną, a różnymi wyrazami wolnymi:

$$\mathbf{A}\mathbf{x}^{(i)} = \mathbf{b}^{(i)}, \quad i = 1, 2, \ldots, M \tag{37}$$

gdzie $\mathbf{A} \in \mathbb{R}^{N\times N}$, $\mathbf{x}^{(i)}, \mathbf{b}^{(i)} \in \mathbb{R}^N$. Eliminacja Gaussa (z wyborem elementu podstawowego!) jest efektywna, jeżeli z góry znamy *wszystkie* prawe strony $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \ldots, \mathbf{b}^{(M)}$, gdyż w tym wypadku przeprowadzając eliminację Gaussa, możemy przekształcać wszystkie prawe strony *jednocześnie*. Całkowity koszt rozwiązania (37) wynosi wówczas $O(N^3)+O(MN^2)$.

Jeżeli jednak wszystkie prawe strony nie są z góry znane — co jest sytuacją typową w obliczeniach iteracyjnych — eliminacja Gaussa jest nieefektywna, gdyż trzebaby ją niepotrzebnie przeprowadzać dla każdej prawej strony z osobna, co podnosiłoby koszt numeryczny do $O(MN^3) + O(MN^2)$.