WSTĘP DO METOD NUMERYCZNYCH

Metodą numeryczną nazywa się każdą metodę obliczeniową sprowadzalną do operacji arytmetycznych dodawania, odejmowania, mnożenia i dzielenia. Są to podstawowe operacje matematyczne, znane od wieków przez człowieka a także rozpoznawalne przez każdy procesor komputerowy. Na fundamencie tych czterech działań liczbowych można zbudować całą bazę obliczeniową dla mniej lub bardziej skomplikowanych zagadnień (np. obliczanie pierwiastka kwadratowego z liczby nieujemnej, ale też operacje całkowania i różniczkowania numerycznego). Dlatego zazwyczaj przez *numerykę* rozumie się dziedzinę matematyki zajmującą się rozwiązywaniem przybliżonym zagadnień algebraicznych. I rzeczywiście, odkąd zjawiska przyrodnicze zaczęto opisywać przy użyciu formalizmu matematycznego, pojawiła się potrzeba rozwiązywania zadań analizy matematycznej czy algebry. Dopóki były one nieskomplikowane, dawały się rozwiązywać analitycznie, tzn. z użyciem pewnych przekształceń algebraicznych prowadzących do otrzymywania rozwiązań ścisłych danych problemów. Z czasem jednak, przy powstawaniu coraz to bardziej skomplikowanych teorii opisujących zjawiska, problemy te stawały się na tyle złożone, iż ich rozwiązywanie ścisłe było albo bardzo czasochłonne albo też zgoła niemożliwe. Numeryka pozwalała znajdywać przybliżone rozwiązania z żądaną dokładnością. Ich podstawową zaletą była ogólność tak formułowanych algorytmów, tzn. w ramach danego zagadnienia nie miało znaczenia czy było ono proste czy też bardzo skomplikowane (najwyżej wiązało się z większym nakładem pracy obliczeniowej). Natomiast wadą była czasochłonność. Stąd prawdziwy renesans metod numerycznych nastąpił wraz z powszechnym użyciem w pracy naukowej maszyn cyfrowych, a w szczególności mikrokomputerów (od lat siedemdziesiątych). Dziś złożoność metody numerycznej nie jest żadnym problemem – dziesiątki żmudnych dla człowieka operacji arytmetycznych wykonuje komputer – o wiele ważniejsza stała się analiza otrzymanego wyniku (gł. pod kątem jego dokładności) – tak, aby był on możliwie najbardziej wiarygodny.

Oczywiście metody numeryczne mogą służyć do rozwiązywania konkretnych zagadnień algebraicznych (takich jak np. równania nieliniowe czy problemy własne). Na ogół jednak są one ostatnim ogniwem w łańcuchu zwanym modelowaniem. W celu określenia zachowania się jakiegoś zjawiska w przyrodzie (tu uwaga będzie skierowana na zagadnienia fizyczne, czyli odwracalne), buduje się szereg jego przybliżeń zwanych *modelami*. Modele buduje się przyjmując coraz to nowe założenia i hipotezy upraszczające. Z rzeczywistego systemu fizycznego najpierw powstaje model mechaniczny, (czyli zbiór hipotez dotyczących np. materiału, środowiska, zachowania obciążenia itd.). Jego reprezentacją matematyczną jest model *matematyczny*, czyli opis jego zachowania się przy określonych warunkach mechanicznych w postaci układu równań różniczkowych cząstkowych (na ogół). Następny w kolejności *model numeryczny* polega na zamianie wielkości ciągłych na dyskretne – oznacza przejście do układu równań algebraicznych, do rozwiązania którego służy wybrana metoda numeryczna. Po otrzymaniu wyniku numerycznego (przybliżonego) należy przeprowadzić *analizę błędu*. Należy zauważyć, iż błąd końcowy będzie obarczony błędami ze wszystkich poprzednich etapów modelowania, a więc:

- **Błędem nieuniknionym (błędem modelu),**
- **Błędem metody,**
- **Błędem numerycznym.**

Błąd modelu zwykle wiąże się z przyjęciem złych parametrów początkowych lub brzegowych przy jego tworzeniu. Może się też okazać, iż przyjęto zbyt daleko idące uproszczenia nieoddające dobrze warunków rzeczywistych, w jakich odbywa się dane zjawisko. Mimo tego na ogół buduje się modele w miarę proste, a następnie przeprowadza *analizę wrażliwości*, tzn. sprawdza, jak duży wpływ ma dany pojedynczy czynnik na jego funkcjonowanie.
Błąd metody wiąże się z przyjęciem mało dokładnych parametrów dla tej metody (zbyt rzadki podział obszaru ciągłego na skończone odcinki) lub z zastosowaniem zbyt mało dokładnej metody (mimo dokładnych parametrów). Metod numerycznych dla danego zagadnienia jest na ogół bardzo dużo. Wybór powinien być dokonany z uwagi na przewidywaną postać rzeczywistego zachowania się zjawiska.
Błąd numeryczny wiąże się ściśle z precyzją wykonywanych obliczeń (ręcznych – przez człowieka, przez kalkulator, przez komputer). Wyróżnić można *błąd obcięcia* i *błąd zaokrągleń*. Błąd obcięcia wystąpi, gdy rozwijając daną funkcję w szereg odrzucamy nieskończoną liczbę wyrazów od pewnego miejsca, zachowując jedynie pewną początkową ich liczbę (w kalkulatorach działaniami pierwotnymi są operacje dodawania, odejmowania, mnożenia i dzielenia, natomiast wszystkie inne, np. obliczanie wartości funkcji trygonometrycznych wiąże się z rozwijaniem tychże funkcji w szeregi potęgowe z daną dokładnością obcięcia). Błąd zaokrągleń wiąże się z reprezentacją ułamków dziesiętnych nieskończonych (należy przy tym pamiętać, iż komputer prowadzi obliczenia z właściwą dla danego typu liczbowego precyzją, natomiast pokazywać graficznie wyniki może z dokładnością żądaną przez użytkownika – wtedy na potrzeby formatu prezentacji zaokrągla z daną dokładnością – tak samo jest zresztą w kalkulatorach).
Inna klasyfikacja błędu numerycznego (tu rozumianego jako dokładność) to:

- **Błąd względny (bezwymiarowy),**
- **Błąd bezwzględny.**

Przyjmując oznaczenia: $\overline{x}$ - wielkość przybliżona oraz x - wielkość ścisła, można zapisać błąd bezwzględny $\delta = |\overline{x} - x|$ i błąd względny $\varepsilon = \left|\frac{\overline{x} - x}{x}\right|$. Błąd względny jako bezwymiarowy często przedstawiany jest w procentach. Podanie samej wartości $\overline{x}$ w numeryce jest bezwartościowe – musi jej towarzyszyć jedna z powyższych dokładności, (co zapisuje się jako: $\overline{x} \pm \delta$ lub $\overline{x}(1 \pm \varepsilon)$).
Ważnym pojęciem w numeryce jest pojęcie cyfr znaczących. Pierwsza cyfra znacząca to pierwsza niezerowa cyfra licząc od lewej strony ułamka dziesiętnego. W praktyce jest to cyfra, do której można mieć „zaufanie”, iż nie pochodzi z zaokrąglenia, lecz znalazła się tam z rzeczywistych obliczeń. Np. 2345000 (4 cyfry znaczące), 2.345000 (7 cyfr znaczących), 0.023450 (5 cyfr znaczących), 0.02345 (4 cyfry znaczące) itd. Dokładność końcowa musi mieć tyle cyfr znaczących, ile mają warunki początkowe. Oznacza to w praktyce, iż nie można prowadzić obliczeń zachowując np. trzy miejsca po przecinku, a ostateczny wynik podawać bezkarnie z większą niż ta dokładnością. Będzie on wtedy bezwartościowy, gdyż błąd zaokrągleń może wkraść się nawet na pierwszą pozycję dziesiętną, zwłaszcza jeżeli w trakcie obliczeń przeprowadzano często działania dzielenia i odejmowania, które obniżają dokładność wyniku.

ROZWIĄZYWANIE NIELINIOWYCH RÓWNAŃ ALGEBRAICZNYCH

Najprostszym wykorzystaniem metod numerycznych jest numeryczne rozwiązywanie równań algebraicznych nieliniowych. Nieliniowość może być pochodzenia geometrycznego (np. w mechanice przyjęcie teorii dużych odkształceń czy przemieszczeń) lub fizycznego (nieliniowe związki konstytutywne, gdy materiał nie podlega liniowemu prawu sprężystości). Końcowym efektem takiego modelowania w przestrzeni jednowymiarowej przy jednej zmiennej niezależnej jest równanie postaci:

$$F(x) = 0$$

Tworząc w określony sposób równanie postaci:

$$x = f(x),$$

gdzie $f(x)$ jest dowolną, nieliniową funkcją zmiennej *x* można stworzyć ciąg liczbowy postaci

$$x_{n+1} = f(x_n) \tag{1}$$

rozpoczynając obliczenia od dowolnej (na ogół) liczby x_0, zwanej *punktem startowym*:

$$x_0, \quad x_1 = f(x_0), \quad x_2 = f(x_1), \quad x_3 = f(x_2), \quad ... \tag{2}$$

Graficznie proces ten polega na szukaniu punktu wspólnego dla prostej $y = x$ oraz krzywej $y = f(x)$.
Jeżeli wykona się odpowiednio dużo takich obliczeń, to przy odpowiednich warunkach, jakie musi spełniać funkcja $f(x)$, proces okaże się zbieżny (do określonej liczby $\hat{x}$). Równanie (1) nazywa się wtedy *schematem iteracyjnym*, a ciąg przybliżeń (2) *procesem iteracyjnym*. Liczby potrzebnych iteracji nie da się z góry określić (będzie ona funkcją punktu startowego oraz postaci schematu iteracyjnego). Dlatego o miejscu przerwania iteracji muszą świadczyć dodatkowe kryteria. Formułuje się je definiując następujące nieujemne wielkości skalarne:

- Tempo zbieżności: $\varepsilon^{(1)} = \left| \dfrac{x_{n+1} - x_n}{x_{n+1}} \right|$,
- Residuum: $\varepsilon^{(2)} = \left| \dfrac{F(x_{n+1})}{F(x_0)} \right|$,
- Ilość iteracji: $\varepsilon^{(3)}: \; n = ...$

Wtedy o zakończeniu obliczeń decydować będą warunki: $\varepsilon^{(1)} \le \varepsilon^{(1)}_{dop}$, $\varepsilon^{(2)} \le \varepsilon^{(2)}_{dop}$, $n \le n_{\max}$. Dwa pierwsze są niezależne od siebie i powinny być spełnione równocześnie. Trzeci jest dla nich alternatywą. Liczby (tu: bezwymiarowe) $\varepsilon^{(1)}_{dop}$, $\varepsilon^{(2)}_{dop}$, $n_{\max}$ są danymi z góry wielkościami dopuszczalnymi.

Przy formułowaniu powyższych kryteriów użyto wielkości względnych, (które mogą być łatwo porównywane między sobą). Czasami wskazane jest użycie wielkości wymiarowych, ale wtedy określenie czy liczba jest „mała" czy „duża" nie jest już takie oczywiste.
Malejące *tempo zbieżności* świadczy o zbieżności danego schematu iteracyjnego do jednej skończonej wartości (tu: $x_n \overset{n\to\infty}{\to} \hat{x}$). Schemat iteracyjny rozbieżny może dawać coraz większe liczby wraz ze wzrostem liczby iteracji (rozbieżność jako „zbieżność" do nieskończoności), może oscylować pomiędzy dwiema różnymi wartościami (tzw. proces niestabilny) lub po prostu okazać się osobliwym dla danego x_n. Takie sytuacje wychwytuje tempo zbieżności, które zamiast systematycznie maleć utrzymuje się na tym samym poziomie lub nieograniczenie rośnie do nieskończoności.
Natomiast małość kryterium residualnego (resztkowego) świadczy o spełnieniu wyjściowego równania algebraicznego (1). Może się bowiem zdarzyć, iż sama zbieżność procesu nie gwarantuje zbieżności schematu do właściwego rozwiązania $\overline{x}$,tj. takiego, że $F(\overline{x}) = 0$. Wtedy $\hat{x} \neq \overline{x}$ i wykaże ten fakt niezerowe residuum, natomiast tempo zbieżności będzie mimo to maleć. Dopiero spełnienie obydwu kryteriów gwarantuje uzyskanie przybliżenia właściwego rozwiązania wyjściowego równania (1).

Procesy iteracyjne mogą być zbieżne i rozbieżne jednostronnie (wtedy zbliżamy się lub oddalamy od właściwego rozwiązania z jednej strony – od dołu lub od góry) lub dwustronnie (wyniki iteracji „skaczą" z jednej strony wartości ścisłej na drugą cyklicznie, przybliżając się do niej lub oddalając). Przykłady takich procesów pokazują poniższe rysunki.

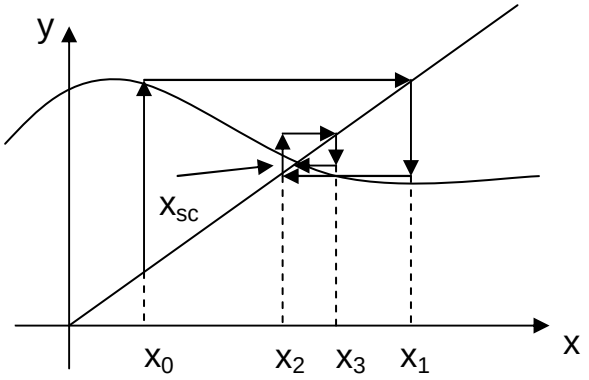
Można zauważyć pewną cechę wspólną dla funkcji prawej strony $f(x)$ w przypadku procesów zbieżnych i rozbieżnych. Wszystko zależy od nachylenia funkcji w pewnym otoczeniu (przedziale $[a,b]$), w którym poszukiwanie jest rozwiązanie. Funkcje „strome" powodują rozbieżność schematu. Tą „stromość" określa się przez kąt nachylenia wykresu do osi x, a kryterium zbieżności wynika z warunku Lipschitza.

Twierdzenie 1
Jeżeli $f(x)$ *spełnia warunek Lipschitza:*

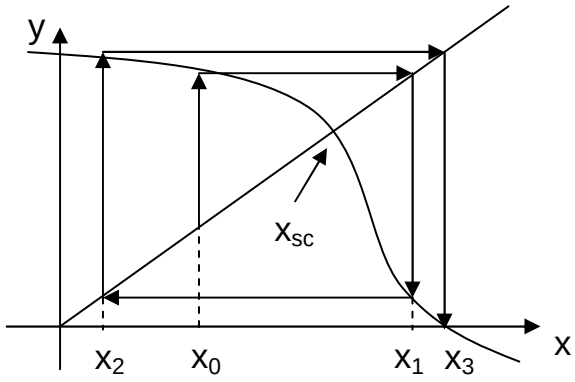
$$|f(x_1) - f(x_2)| \le L|x_1 - x_2| \quad , \quad 0 < L < 1, \quad x_1, x_2 \in [a,b]$$

to równanie algebraiczne $x = f(x)$ *posiada co najmniej jedno rozwiązanie w przedziale* $[a,b]$.

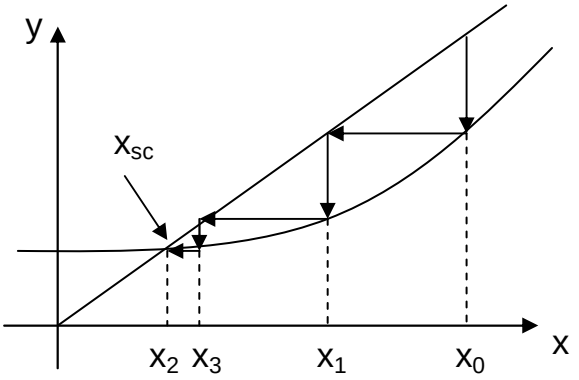
Twierdzenie 2
Jeżeli $f(x)$ *spełnia twierdzenie 1, to proces iteracyjny* $x_{n+1} = f(x_n)$ *jest zbieżny do rozwiązania ścisłego równania* $x = f(x)$ *dla* $x \in [a,b]$ *przy dowolnym punkcie startowym* x_0.



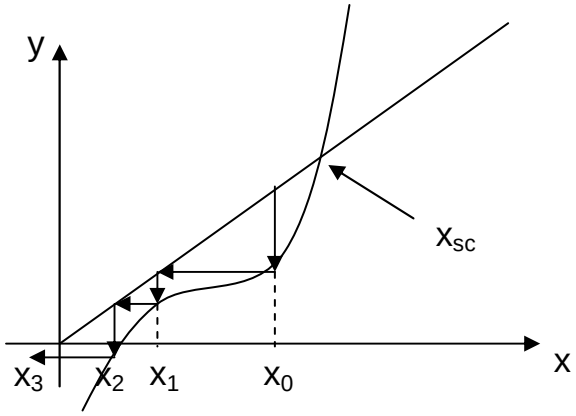
Proces zbieżny dwustronnie



Proces rozbieżny dwustronnie



Proces zbieżny jednostronnie



Proces rozbieżny jednostronnie

Konsekwencją powyższych twierdzeń jest następujący wniosek: jeżeli kąt nachylenia funkcji $f(x)$ na pewnym przedziale $x \in [x_1, x_2]$ jest mniejszy niż 45º, to schemat iteracyjny jest zbieżny przy dowolnym punkcie startowym. Tangens kąta nachylenia liczymy na podstawie ilorazu różnicowego funkcji $f(x)$.

O ewentualnej zbieżności lub rozbieżności decyduje schemat, a dokładniej: sposób znajdywania funkcji prawej strony $f(x)$. Sposób ten stanowi podstawę klasyfikacji dalszych metod iteracyjnych. Na ogół schemat powinien mieć zapewnioną bezwarunkową stabilność i zbieżność.

Równanie wyjściowe: $F(x) = 0$.

METODA ITERACJI PROSTEJ

$$\begin{cases} x_0 \\ x_{n+1} = f(x_n) \end{cases}.$$

Sposób znajdywania funkcji $f(x)$ nie jest z góry określony, może pochodzić z prostego przekształcenia: $F(x)=0 \rightarrow F(x)+x=x \rightarrow f(x)=x$. Taki schemat nie ma zagwarantowanej zbieżności ani stabilności.

METODA ITERACJI PROSTEJ Z RELAKSACJĄ

Pojęcie *relaksacji* w numeryce oznacza ingerencję w tempo zbieżności wyniku. W metodzie iteracji prostej można zrobić modyfikację polegającą na obróceniu wykresu funkcji $f(x)$ względem początku układu o taki kąt α, aby proces iteracyjny był optymalnie szybko zbieżny w okolicy danego punktu (punkt optymalnej zbieżności). Relaksacja nie tylko poprawia tempo zbieżności, ale również potrafi zamienić wyjściowy schemat rozbieżny na zbieżny.
Wartość kąta α należy wyznaczyć optymalizując nowo otrzymany schemat poprzez dodanie do starego czynnika liniowego odpowiadającego za obrót (punkt optymalnej zbieżności na ogół utożsamiany jest z punktem startowym x_0):

$$x = f(x)$$
$$x+\alpha x = f(x)+\alpha x$$
$$x(1+\alpha) = f(x)+\alpha x$$
$$x = \frac{f(x)}{1+\alpha}+\frac{\alpha x}{1+\alpha} = g(x)$$
$$x = g(x)$$

Optymalizujemy nowy schemat iteracyjny:

$$g'(x_0) = 0$$
$$\frac{f'(x_0)}{1+\alpha}+\frac{\alpha}{1+\alpha} = 0 \quad (\alpha \neq -1)$$
$$\alpha = -f'(x_0)$$

Tak policzone α wstawiamy do schematu $x = g(x)$:

$$x = \frac{f(x)}{1-f'(x_0)} - x\frac{f'(x_0)}{1-f'(x_0)}$$

Końcowa postać schematu iteracyjnego *metody iteracji prostej z relaksacją*:

$$\begin{cases} x_0 \\ x_{n+1} = \dfrac{f(x_n)}{1-f'(x_0)} - x_n\dfrac{f'(x_0)}{1-f'(x_0)} \end{cases}.$$

METODA STYCZNYCH (NEWTONA)

Dla pewnego otoczenia *h* punktu *x* rozwijamy wartość wyjściowej funkcji $F(x+h)$ w szereg Taylora:

$$F(x+h) = F(x) + F'(x)\cdot h + \frac{1}{2}F''(x)\cdot h^2 + \ldots \approx F(x) + F'(x)\cdot h$$

Ustalając x i podstawiając $F(x+h)=0$ można obliczyć przyrost h przy uprzednim odrzuceniu wyrazów rozwinięcia wyższych rzędem niż pierwszy (zlinearyzowany przyrost):

$$h = -\frac{F(x)}{F'(x)}.$$

Dla danej pary sąsiednich przybliżeń zachodzi: $x_{n+1} = x_n + h$.
Stąd po podstawieniu za *h* otrzymujemy schemat metody:

$$\begin{cases} x_0 \\ x_{n+1} = x_n - \dfrac{F(x_n)}{F'(x_n)} \end{cases}.$$

Graficznie metoda Newtona polega na budowaniu stycznych w kolejnych przybliżeniach x_n począwszy od punktu startowego oraz na szukaniu miejsc zerowych tych stycznych.
Wzór na metodę Newtona można też otrzymać stosując metodę relaksacji bezpośrednio do wyjściowego równania $F(x)=0$.
Znana jest też modyfikacja metody dla pierwiastków wielokrotnych (jeżeli równanie $F(x)=0$ takież posiada):

$$\begin{cases} x_0 \\ x_{n+1} = x_n - \dfrac{U(x_n)}{U'(x_n)}, \quad U(x) = \dfrac{F(x)}{F'(x)}, \quad U'(x) = 1 - \dfrac{F(x)\cdot F''(x)}{(F'(x))^2} \end{cases}.$$

METODA SIECZNYCH

W metodzie Newtona do schematu iteracyjnego potrzebna jest znajomość pochodnej funkcji $F(x)$. Aby uniknąć jej różniczkowania, można liczbową pochodną obliczać w sposób przybliżony korzystając z wartości ilorazu różnicowego. Wtedy potrzebne są zawsze dwa punkty wstecz, aby zbudować kolejne rozwiązanie (graficznie styczna przechodzi w sieczną), także na samym początku obliczeń.

$$\begin{cases} x_0, x_1 \\ x_{n+1} = x_n - F(x_n)\cdot \dfrac{x_n - x_{n-1}}{F(x_n) - F(x_{n-1})} \end{cases}.$$

METODA REGULA FALSI

Jeżeli zastosujemy metodę siecznych lub stycznych do funkcji nieregularnej, która w sposób gwałtowny przechodzi z wypukłej na wklęsłą lub z malejącej na rosnącą, jest

niebezpieczeństwo, iż kolejne przybliżenia rozwiązania „uciekną" daleko od początkowego przedziału bez żadnych szans na powrót i na znalezienie sensownego rozwiązania. Pomocna może się okazać pewna modyfikacja metody siecznych, gdzie jeden z punktów, na których buduje się kolejne sieczne, jest z góry ustalony (jest to jeden z punktów startowych), natomiast drugi z nich jest punktem zmiennym. W razie oddalenia się kolejnych przybliżeń od obszaru startowego, w ciągu następnych iteracji nastąpi powrót w jego okolice.

$$\begin{cases} x_0 \text{ (punkt stały)}, x_1 \\ x_{n+1} = x_n - F(x_n) \cdot \dfrac{x_n - x_0}{F(x_n) - F(x_0)} \end{cases}$$

METODA BISEKCJI (POŁOWIENIA)

Metoda bisekcji należy do najstarszych i najprostszych metod iteracyjnych, oprócz znajdywania pierwiastków równań również jest wykorzystywana przy zagadnieniach optymalizacji funkcji. Dla wyjściowego równania $F(x)=0$ szuka ona przybliżenia rozwiązania wewnątrz przedziału $x \in (a,b)$. Stąd, aby metoda zadziała, musi być gwarancja istnienia miejsca zerowego w tym przedziale: $F(a) \cdot F(b) < 0$.

Przy każdej iteracji oblicza się środek przedziału $x = \dfrac{a+b}{2}$. Następnie sprawdza się, w którym z podprzedziałów (a,x) oraz (x,b) leży miejsce zerowe i ten przedział podlega dalszemu dzieleniu. Jeżeli $F(a) \cdot F(x) < 0$ to $b = x$, w przeciwnym przypadku $a = x$. Podział przedziału (a,b) niekoniecznie musi następować na dwie równe części, można go dzielić np. w tzw. złotym stosunku (czyli tak, aby $\dfrac{b-a}{b-x} = \dfrac{b-x}{a-x}$).

Przykład 1

Podać schematy iteracyjne rozwiązania równania $\sin(x) + x^2 = 2$ metodami: (i) iteracji prostej, (ii) iteracji prostej optymalnie szybko zbieżny, (iii) Newtona, (iv) siecznych, (v) regula falsi. Zastosować tak sformułowane schematy do znalezienia dwóch kolejnych przybliżeń rozwiązania startując z punktu $x_0 = -2$ (dla metody siecznych i regula falsi przyjąć drugi punkt startowy $x_1 = -0.5$). Po każdym kroku iteracyjnym określić tempo zbieżności oraz tempo zmiany residuum.

Wyjściowe równanie: $F(x) = \sin(x) + x^2 - 2, \quad F(x) = 0$

Pierwiastki ścisłe równania: $x_{sc_1} = -1.06155, \quad x_{sc_2} = 1.728466$

(i) metoda iteracji prostej

Z równania $F(x)=0$ wyznaczamy w dowolny prosty sposób zmienną *x*, np. $x = \sqrt{\sin(x) + 2}$.

Schemat iteracyjny: $\begin{cases} x_0 = -2 \\ x_{n+1} = \sqrt{\sin(x_n) + 2}, \quad n = 0,1,2,... \end{cases}$

Obliczenia:

Krok $n = 0$:

$$x_1 = \sqrt{\sin(x_0) + 2} = \sqrt{\sin(-2) + 2} = 1.044367$$

$$e_1^{(1)} = \left| \frac{x_1 - x_0}{x_1} \right| = \left| \frac{-2 - 1.044367}{1.044367} \right| = 2.915035$$

$$e_1^{(2)} = \left| \frac{F(x_1)}{F(x_0)} \right| = \left| \frac{\sin(1.044367) - 1.044367^2 - 2}{\sin(-2) - (-2)^2 - 2} \right| = 0.609736$$

Krok $n = 1$:

$$x_2 = \sqrt{\sin(x_1) + 2} = \sqrt{\sin(1.044367) + 2} = 1.692515$$

$$e_2^{(1)} = \left| \frac{x_2 - x_1}{x_2} \right| = \left| \frac{1.044367 - 1.692515}{1.692515} \right| = 0.382950$$

$$e_2^{(2)} = \left| \frac{F(x_2)}{F(x_0)} \right| = \left| \frac{\sin(1.692515) - 1.692515^2 - 2}{\sin(-2) - (-2)^2 - 2} \right| = 0.043995$$

Z dokładnością $e_1^{(1)} = 0.000002 < 10^{-6}$ otrzymano po $n = 16$ iteracjach wynik $x_6 = 1.728466$.

(ii) metoda iteracji prostej z relaksacją

Korzystając z poprzedniego schematu metody iteracji prostej $x = f(x)$: $x = \sqrt{\sin(x) + 2}$, znajdujemy nowy schemat optymalnie szybko zbieżny w okolicy punktu startowego $x_0 = -2$.

$$f(x) = \sqrt{\sin(x) + 2} \quad \rightarrow \quad f'(x) = \frac{1}{2} \frac{1}{\sqrt{\sin(x) + 2}} \cdot \cos(x)$$

$$f'(x_0) = \frac{1}{2} \frac{1}{\sqrt{\sin(x_0) + 2}} \cdot \cos(x_0) = \frac{1}{2} \frac{1}{\sqrt{\sin(-2) + 2}} \cdot \cos(-2) = -0.199234$$

$$1 - f'(x_0) = 1.199234, \quad \frac{1}{1 - f'(x_0)} = 0.833866, \quad \frac{f'(x_0)}{1 - f'(x_0)} = -0.166134$$

Schemat iteracyjny: $\begin{cases} x_0 = -2 \\ x_{n+1} = 0.833866 \cdot \sqrt{\sin(x_n) + 2} + 0.166134 \cdot x_n, \quad n = 0,1,2,... \end{cases}$

Obliczenia:

Krok $n = 0$:

$$x_1 = 0.833866 \cdot \sqrt{\sin(x_0) + 2} + 0.166134 \cdot x_0 =$$
$$= 0.833866 \cdot \sqrt{\sin(-2) + 2} + 0.166134 \cdot (-2) = 0.538593$$
$$e_1^{(1)} = \left| \frac{x_1 - x_0}{x_1} \right| = \left| \frac{-2 - 0.538593}{0.538593} \right| = 4.713379$$
$$e_1^{(2)} = \left| \frac{F(x_1)}{F(x_0)} \right| = \left| \frac{\sin(0.538593) - 0.538593^2 - 2}{\sin(-2) - (-2)^2 - 2} \right| = 0.764049$$

Krok $n = 1$:

$$x_2 = 0.833866 \cdot \sqrt{\sin(x_1) + 2} + 0.166134 \cdot x_1 = 1.411341$$
$$e_2^{(1)} = \left| \frac{x_2 - x_1}{x_2} \right| = \left| \frac{1.411341 - 0.538593}{1.411341} \right| = 0.618382$$
$$e_2^{(2)} = \left| \frac{F(x_2)}{F(x_0)} \right| = \left| \frac{\sin(1.411341) - 1.411341^2 - 2}{\sin(-2) - (-2)^2 - 2} \right| = 0.342155$$

Z dokładnością $e_1^{(1)} = 0.000002 < 10^{-6}$ otrzymano po $n = 8$ iteracjach wynik $x_8 = 1.728464$.

(iii) metoda Newtona

Postać wyjściowa równania: $F(x) = \sin(x) + x^2 - 2, \quad F(x) = 0$.
Obliczenie pochodnej funkcji $F(x)$: $F'(x) = \cos(x) + 2x$.

Schemat iteracyjny: $\begin{cases} x_0 = -2 \\ x_{n+1} = x_n - \dfrac{\sin(x_n) + x_n^2 - 2}{\cos(x_n) + 2x_n}, \quad n = 0,1,2,... \end{cases}$

Obliczenia:

$x_1 = -1.188221, \quad e_1^{(1)} = 0.683189 \quad e_1^{(2)} = 0.116721$
$x_2 = -1.064728, \quad e_1^{(1)} = 0.115985, \quad e_1^{(2)} = 0.002854$
...
$x_4 = -1.061550, \quad e_1^{(1)} < 10^{-6}, \quad e_1^{(2)} < 10^{-8}$

(iv) metoda siecznych

Postać wyjściowa równania: $F(x) = \sin(x) + x^2 - 2, \quad F(x) = 0$.

Schemat iteracyjny:

$$\begin{cases} x_0 = -2, \quad x_1 = -0.5 \\ x_{n+1} = x_n - (\sin(x_n) + x_n^2 - 2) \cdot \dfrac{x_{n-1} - x_n}{\sin(x_{n-1}) + x_{n-1}^2 - \sin(x_n) - x_n^2}, \quad n = 1,2,... \end{cases}$$

Obliczenia:

$x_1 = -0.955962, \quad e_1^{(1)} = 0.476967 \quad e_1^{(2)} = 0.092554$
$x_2 = -1.078578, \quad e_1^{(1)} = 0.113683, \quad e_1^{(2)} = 0.015336$
...
$x_5 = -1.061550, \quad e_1^{(1)} < 10^{-6}, \quad e_1^{(2)} < 10^{-8}$

(v) metoda regula falsi

Postać wyjściowa równania: $F(x) = \sin(x) + x^2 - 2, \quad F(x) = 0$.
Punkt stały: $x_0 = -2$.

Schemat iteracyjny:

$$\begin{cases} x_0 = -2, \quad x_1 = -0.5 \\ x_{n+1} = x_n - (\sin(x_n) + x_n^2 - 2) \cdot \dfrac{-2 - x_n}{3.090703 - \sin(x_n) - x_n^2}, \quad n = 1, 2, ... \end{cases}$$

Obliczenia:
$x_1 = -0.955962, \quad e_1^{(1)} = 0.476967 \quad e_1^{(2)} = 0.092554$
$x_2 = -1.044406, \quad e_1^{(1)} = 0.084684, \quad e_1^{(2)} = 0.015327$
...
$x_7 = -1.061548, \quad e_1^{(1)} < 10^{-6}, \quad e_1^{(2)} < 10^{-6}$

Przykład 2
Równanie z poprzedniego zadania rozwiązać w sposób przybliżony metodą bisekcji. Przyjąć przedział $(1,3)$. Rozwiązanie znaleźć z dokładnością $e_{dop} = 10^{-3}$.

Postać wyjściowa równania: $F(x) = \sin(x) + x^2 - 2, \quad F(x) = 0$.
Początek przedziału: $a_0 = 1$, koniec przedziału: $b_0 = 3$.

Obliczenia zestawiono w tabeli:

n	$x_n = \frac{a_{n-1} + b_{n-1}}{2}$	$F(x_n) \cdot F(a_{n-1})$	a_n	b_n	$e_n^{(1)} = \left\|\frac{x_{n-1} - x_n}{x_n}\right\|$	$\delta_n^{(2)} = \|F(x_n)\|$
1	2.000	-2.008497	1.000	2.000	0.500	1.090703
2	1.500	1.376490	1.500	2.000	0.333333	0.747495
3	1.750	-0.058689	1.500	1.750	0.142857	0.078514
4	1.625000	0.267533	1.625000	1.750	0.076923	0.357906
5	1.687500	0.052090	1.687500	1.750	0.037037	0.145542
6	1.718750	0.005090	1.718750	1.750	0.018182	0.034973
7	1.734375	-0.000749	1.718750	1.734375	0.009009	0.021406
8	1.726563	0.000240	1.726563	1.734375	0.004525	0.006875
9	1.730469	-0.000050	1.726563	1.730469	0.002257	0.007243
10	1.728516	-0.000001	1.726563	1.728516	0.001130	0.000178
11	**1.727539**	0.000023	1.727539	1.728516	**0.000565**	**0.003350**

UKŁADY RÓWNAŃ NIELINIOWYCH

Rozwiązywanie układów równań algebraicznych (liniowych lub nieliniowych) to najczęściej spotykany problem algebraiczny w zagadnieniach fizyki. Stąd potrzeba opracowania aparatu analizy takich układów, najczęściej w formie wektorowej i macierzowej. Ponieważ działania wykonywane będą już nie na pojedynczych liczbach tylko na wielkościach macierzowych, należy wprowadzić pojęcie normy (wektora, macierzy) – stanowiącej reprezentację tej wielkości w postaci pojedynczej liczby rzeczywistej dodatniej.

Definicja
Norma wektorowa $\|\boldsymbol{x}\|$ *z wektora* $\boldsymbol{x} \in V$, *gdzie* V *to liniowa n – wymiarowa przestrzeń wektorowa, jest skalarem spełniającym następujące warunki:*

1. $\|\boldsymbol{x}\| \geq 0 \quad \forall_{\boldsymbol{x}\in V} \quad , \quad \|\boldsymbol{x}\| = 0 \quad \Leftrightarrow \quad \boldsymbol{x} = \boldsymbol{0}$,
2. $\|\alpha \boldsymbol{x}\| = |\alpha| \cdot \|\boldsymbol{x}\| \quad \forall_{\boldsymbol{x}\in V} \quad , \quad \forall_{\alpha \in \Re}$,
3. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\| \quad \forall_{\boldsymbol{x},\boldsymbol{y}\in V}$.

Najczęściej używane normy wektorowe:

1. $\|\boldsymbol{x}\|_1 = \left[\sum_{i=1}^{n} |x_i|^2 \right]^{\frac{1}{2}}$, norma Euklidesa (średnio kwadratowa),
2. $\|\boldsymbol{x}\|_2 = \max_{(i)} |x_i|$, norma Czebyszewa (maksimum),
3. $\|\boldsymbol{x}\|_3 = \left[\sum_{i=1}^{n} |x_i|^p \right]^{\frac{1}{p}}$, $p \geq 1$, uogólnienie dwóch powyższych przypadków ($p = 2$ - norma Euklidesa, $p = \infty$ - norma Czebyszewa).

Definicja
Norma macierzowa $\|\boldsymbol{A}\|$ *z macierzy kwadratowej* $n \times n$ ($\underset{n\times n}{\boldsymbol{A}} = \left[a_{ij} \right]$) *jest skalarem spełniającym następujące warunki:*

1. $\|\boldsymbol{A}\| \geq 0 \quad , \quad \|\boldsymbol{A}\| = 0 \quad \Leftrightarrow \quad \boldsymbol{A} = \boldsymbol{0}$,
2. $\|\alpha \boldsymbol{A}\| = |\alpha| \cdot \|\boldsymbol{A}\| \quad , \quad \forall_{\alpha \in \Re}$,
3. $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$,
4. $\|\boldsymbol{A} \cdot \boldsymbol{B}\| \leq \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|$.

Najczęściej używane normy macierzowe:

1. $\|\boldsymbol{A}\|_1 = \left[\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right]^{\frac{1}{2}}$, norma Euklidesa (średnio kwadratowa),
2. $\|\boldsymbol{A}\|_2 = \max_{(i)} \sum_{j=1}^{n} |a_{ij}|$, norma Czebyszewa (maksimum).

Często używane jest też pojęcie średniej normy Euklidesa. Wtedy przed pierwiastkowaniem sumy kwadratów współrzędnych dzieli się dodatkowo tą sumę przez liczbę wyrazów *n*.

METODA NEWTONA – RAPHSONA

Metoda służy do rozwiązywania układów równań nieliniowych i stanowi uogólnienie metody iteracji prostej dla wielu równań jednocześnie.

Twierdzenie 1
Niech $F_i : x_i \in [a_i, b_i] \to \Re, \quad i = 1, 2, ..., n$ *należy do n – wymiarowej przestrzeni euklidesowej* $\Re^n$.
Niech $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x})$ *spełnia następujące warunki*

1. $\boldsymbol{f}$ *jest określone i ciągłe w* $\Re^n$,
2. *norma jakobianowa z* $\boldsymbol{f}$ *spełnia warunek* $\left\| \boldsymbol{J}_f(\boldsymbol{x}) \right\| \le L \le 1$,

$$\boldsymbol{J}_f = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial F_n}{\partial x_1} & \dots & \frac{\partial F_n}{\partial x_n} \end{bmatrix}$$

3. *dla każdego* $\boldsymbol{x} \in \Re^n$ *m* $\boldsymbol{f}(\boldsymbol{x})$ *również należy do* $\Re^n$.

Wtedy dla każdego $\boldsymbol{x_0} \in \Re^n$ *ciąg iteracyjny* $\boldsymbol{x}_{n+1} = \boldsymbol{f}(\boldsymbol{x}_n)$ *jest zbieżny do jednoznacznego rozwiązania* $\tilde{\boldsymbol{x}}$.

Rozważmy punkt $\boldsymbol{x}$ oraz jego bliskie otoczenie $\boldsymbol{x} + \boldsymbol{h}$. Wtedy $\boldsymbol{F}(\boldsymbol{x} + \boldsymbol{h}) = \boldsymbol{0}$. Rozwijając ostatnią wielkość wektorową w szereg Taylora otrzymuje się:

$$\boldsymbol{F}(\boldsymbol{x} + \boldsymbol{h}) = \boldsymbol{F}(\boldsymbol{x}) + \frac{\partial \boldsymbol{F}(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{h} + \frac{1}{2} \frac{\partial^2 \boldsymbol{F}(\boldsymbol{x})}{\partial^2 \boldsymbol{x}} \boldsymbol{h}^2 + ... = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{J}(\boldsymbol{x}) \cdot \boldsymbol{h} + \boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{0}$$

Linearyzując powyższy związek ze względu na $\boldsymbol{h}$ i wyliczając wektor $\boldsymbol{h}$ otrzymuje się:

$$\boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{J}(\boldsymbol{x}) \cdot \boldsymbol{h} = \boldsymbol{0} \quad \to \quad \boldsymbol{h} = -\boldsymbol{J}^{-1}(\boldsymbol{x}) \cdot \boldsymbol{F}(\boldsymbol{x})$$

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{h} \quad \to \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{J}^{-1}(\boldsymbol{x}_n) \cdot \boldsymbol{F}(\boldsymbol{x}_n)$$

Przy takim zapisie schematu konieczne byłoby odwracanie macierzy $\boldsymbol{J}^{-1}(\boldsymbol{x}_n)$ na każdym kroku. Aby tego uniknąć, mnoży się stronami przez $\boldsymbol{J}(\boldsymbol{x}_n)$, co prowadzi do sformułowania układu równań liniowych (rozwiązywanych analitycznie lub numerycznie).

$$\begin{cases} \boldsymbol{x_0} \\ \boldsymbol{J}(\boldsymbol{x}_n) \cdot \boldsymbol{x}_{n+1} = \boldsymbol{J}(\boldsymbol{x}_n) \cdot \boldsymbol{x}_n - \boldsymbol{F}(\boldsymbol{x}_n) \end{cases}$$

Kryteria przerwania iteracji w przypadku wielowymiarowym:

1. $\varepsilon^{(1)} = \dfrac{\left\| \boldsymbol{x}_{n+1} - \boldsymbol{x}_n \right\|}{\left\| \boldsymbol{x}_{n+1} \right\|} \le \varepsilon_{dop}^{(1)}$,
2. $\varepsilon^{(2)} = \dfrac{\left\| \boldsymbol{F}(\boldsymbol{x}_{n+1}) \right\|}{\left\| \boldsymbol{F}(\boldsymbol{x}_0) \right\|} \le \varepsilon_{dop}^{(2)}$.

Najczęściej sprawdza się rozwiązanie dla dwóch rodzajów norm: dla normy maksimum, która wychwytuje największy błąd w obszarze rozwiązania i średniej normy kwadratowej, która mówi o średniej jakości rozwiązania.

Istnieją różne modyfikacje metody Newtona. Najprostsza polega na nie zmienianiu wyjściowej macierzy jakobianowej, co pociąga większą liczbę kroków, niż przy oryginalnej metodzie, ale tylko dla jednego obliczania macierzy (pomocne może być omawiane w dalszych rozdziałach opracowania rozbicie na czynniki trójkątne). Możliwe jest też uaktualnianie macierzy co pewną liczbę kroków, a więc tam, gdzie macierz mogła ulec istotnej zmianie.
Inna metoda polega na dokonaniu *relaksacji*.

$$\begin{cases} \boldsymbol{x_0} \\ \boldsymbol{J}(\boldsymbol{x_n}) \cdot \boldsymbol{x_{n+1}} = \boldsymbol{J}(\boldsymbol{x_n}) \cdot \boldsymbol{x_n} - \alpha \cdot \boldsymbol{F}(\boldsymbol{x_n}) \end{cases}.$$

(najczęściej $\alpha = 1.2, 1.3, 1.4$ - tzw. *nadrelaksacja*)

W przypadku wyraźnej oscylacji rozwiązania (np. wynik przechodzi z jednej na drugą stronę osi „x") możliwe jest wprowadzenie przyśpieszenia zbieżności iteracji *metodą Aitkena*. Wprowadzając oznaczenia: x_j - wartość rozwiązania na j-tym kroku, x - rozwiązanie ścisłe można zapisać liniowy związek:

$$x - x_n = \alpha(x - x_{n-1})$$

Przy założeniu, że współczynnik α jest stały dla dwóch sąsiednich iteracji, można zapisać układ równań dla trzech kolejnych przybliżeń rozwiązania:

$$\begin{cases} x - x_n = \alpha(x - x_{n-1}) \\ x - x_{n-1} = \alpha(x - x_{n-2}) \end{cases}$$

Rozwiązując go ze względu na x otrzymuje się związek:

$$x = \frac{x_{n-2} \cdot x_n - x_{n-1}^2}{x_n - 2x_{n-1} + x_{n-2}}.$$

Wzór należy używać osobno dla każdej zmiennej niezależnej poprawiając wartość otrzymaną na n-tym kroku iteracji.

Przykład 1

Rozwiązać następujący układ równań nieliniowych $\begin{cases} y^2 = 2x \\ x^2 + y^2 = 8 \end{cases}$ metodą Newtona – Raphsona. Przyjąć wektor startowy $\boldsymbol{x}_0 = \left\{0, 2\sqrt{2}\right\}$. Po każdym kroku iteracyjnym przeprowadzać analizę błędu. Przyjąć następujące poziomy błędu: $\varepsilon_{dop}^{(1)} = 10^{-6}$, $\varepsilon_{dop}^{(2)} = 10^{-8}$.

Wektor funkcyjny: $\boldsymbol{F}(x,y) = \begin{cases} F_1(x,y) = y^2 - 2x \\ F_2(x,y) = x^2 + y^2 - 8 \end{cases}$.

Macierz jakobianowa: $\boldsymbol{J}(x,y) = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{bmatrix}(x,y) = \begin{bmatrix} -2 & 2y \\ 2x & 2y \end{bmatrix}$.

Wektor startowy: $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 2\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0.0 \\ 2.8284 \end{bmatrix}$.

Schemat iteracyjny: $\boldsymbol{J}(x_n, y_n) \cdot \boldsymbol{x}_{n+1} = \boldsymbol{J}(x_n, y_n) \cdot \boldsymbol{x}_n - \boldsymbol{F}(x_n, y_n) \quad \rightarrow \quad \boldsymbol{x}_{n+1} = \ldots$

$$\begin{bmatrix} -2 & 2y_n \\ 2x_n & 2y_n \end{bmatrix} \cdot \begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} -2 & 2y_n \\ 2x_n & 2y_n \end{bmatrix} \cdot \begin{bmatrix} x_n \\ y_n \end{bmatrix} - \begin{bmatrix} y_n^2 - 2x_n \\ x_n^2 + y_n^2 - 8 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \ldots$$

Analiza błędów:

$$\varepsilon^{(1)} = \frac{\left\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\right\|}{\left\|\boldsymbol{x}_{n+1}\right\|} = \frac{\left\|\begin{bmatrix} x_{n+1} - x_n \\ y_{n+1} - y_n \end{bmatrix}\right\|}{\left\|\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix}\right\|} \overset{?}{<} \varepsilon_{dop}^{(1)}, \quad \varepsilon^{(2)} = \frac{\left\|\boldsymbol{F}(\boldsymbol{x}_{n+1})\right\|}{\left\|\boldsymbol{F}(\boldsymbol{x}_0)\right\|} = \frac{\left\|\begin{bmatrix} y_n^2 - 2x_n \\ x_n^2 + y_n^2 - 8 \end{bmatrix}\right\|}{\left\|\begin{bmatrix} y_0^2 - 2x_0 \\ x_0^2 + y_0^2 - 8 \end{bmatrix}\right\|} \overset{?}{<} \varepsilon_{dop}^{(2)}$$

Krok $n = 0$:

$$\begin{bmatrix} -2 & 2y_0 \\ 2x_0 & 2y_0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} -2 & 2y_0 \\ 2x_0 & 2y_0 \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - \begin{bmatrix} y_0^2 - 2x_0 \\ x_0^2 + y_0^2 - 8 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \ldots$$

$$\begin{bmatrix} -2.0 & 5.6569 \\ 0.0 & 5.6569 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} -2.0 & 5.6569 \\ 0.0 & 5.6569 \end{bmatrix} \cdot \begin{bmatrix} 0.0 \\ 2.8284 \end{bmatrix} - \begin{bmatrix} 8.0 \\ 0.0 \end{bmatrix}$$

$$\begin{bmatrix} -2.0 & 5.6569 \\ 0.0 & 5.6569 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 8.0 \\ 16.0 \end{bmatrix} \rightarrow \quad \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{y_1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{4.0} \\ \boldsymbol{2.8284} \end{bmatrix}$$

Błędy w normie euklidesowej:

$$e_e^{(1)} = \frac{\left\|\boldsymbol{x}_1 - \boldsymbol{x}_0\right\|_e}{\left\|\boldsymbol{x}_1\right\|_e} = \frac{\left\|\begin{bmatrix}4.0\\2.8284\end{bmatrix} - \begin{bmatrix}0.0\\2.8284\end{bmatrix}\right\|_e}{\left\|\begin{bmatrix}4.0\\2.8284\end{bmatrix}\right\|_e} = \frac{\left\|\begin{bmatrix}4.0\\0.0\end{bmatrix}\right\|_e}{\left\|\begin{bmatrix}4.0\\2.8284\end{bmatrix}\right\|_e} = \frac{\sqrt{\frac{1}{2}\left(4.0^2 + 0.0^2\right)}}{\sqrt{\frac{1}{2}\left(4.0^2 + 2.8284^2\right)}} = 0.8165$$

$$e_e^{(2)} = \frac{\left\|\boldsymbol{F}(\boldsymbol{x}_1)\right\|_e}{\left\|\boldsymbol{F}(\boldsymbol{x}_0)\right\|_e} = \frac{\left\|\begin{bmatrix}2.8284^2 - 2\cdot 4.0\\4.0^2 + 2.8284^2 - 8.0\end{bmatrix}\right\|_e}{\left\|\begin{bmatrix}2.8284^2 - 2\cdot 0.0\\0.0^2 + 2.8284^2 - 8.0\end{bmatrix}\right\|_e} = \frac{\left\|\begin{bmatrix}0.0\\16.0\end{bmatrix}\right\|_e}{\left\|\begin{bmatrix}8.0\\0.0\end{bmatrix}\right\|_e} = \frac{\sqrt{\frac{1}{2}\left(0.0^2 + 16.0^2\right)}}{\sqrt{\frac{1}{2}\left(8.0^2 + 0.0^2\right)}} = 2.0$$

Błędy w normie maksimum:

$$e_m^{(1)} = \frac{\left\|\boldsymbol{x}_1 - \boldsymbol{x}_0\right\|_m}{\left\|\boldsymbol{x}_1\right\|_m} = \frac{\left\|\begin{bmatrix}4.0\\2.8284\end{bmatrix} - \begin{bmatrix}0.0\\2.8284\end{bmatrix}\right\|_m}{\left\|\begin{bmatrix}4.0\\2.8284\end{bmatrix}\right\|_m} = \frac{\left\|\begin{bmatrix}4.0\\0.0\end{bmatrix}\right\|_m}{\left\|\begin{bmatrix}4.0\\2.8284\end{bmatrix}\right\|_m} = \frac{4.0}{4.0} = 1.0$$

$$e_m^{(2)} = \frac{\left\|\boldsymbol{F}(\boldsymbol{x}_1)\right\|_m}{\left\|\boldsymbol{F}(\boldsymbol{x}_0)\right\|_m} = \frac{\left\|\begin{bmatrix}2.8284^2 - 2\cdot 4.0\\4.0^2 + 2.8284^2 - 8.0\end{bmatrix}\right\|_m}{\left\|\begin{bmatrix}2.8284^2 - 2\cdot 0.0\\0.0^2 + 2.8284^2 - 8.0\end{bmatrix}\right\|_m} = \frac{\left\|\begin{bmatrix}0.0\\16.0\end{bmatrix}\right\|_m}{\left\|\begin{bmatrix}8.0\\0.0\end{bmatrix}\right\|_m} = \frac{16.0}{8.0} = 2.0$$

Sprawdzenie kryterium zbieżności:

$$\varepsilon_e^{(1)} = 0.8165 > \varepsilon_{dop}^{(1)} = 10^{-6}$$
$$\varepsilon_m^{(1)} = 1.0000 > \varepsilon_{dop}^{(1)} = 10^{-6}$$
$$\varepsilon_e^{(2)} = 2.0000 > \varepsilon_{dop}^{(2)} = 10^{-8}$$
$$\varepsilon_m^{(2)} = 2.0000 > \varepsilon_{dop}^{(2)} = 10^{-8}$$

Krok $n = 1$:

$$\begin{bmatrix}-2 & 2y_1\\2x_1 & 2y_1\end{bmatrix}\cdot\begin{bmatrix}x_2\\y_2\end{bmatrix} = \begin{bmatrix}-2 & 2y_1\\2x_1 & 2y_1\end{bmatrix}\cdot\begin{bmatrix}x_1\\y_1\end{bmatrix} - \begin{bmatrix}y_1^2 - 2x_1\\x_1^2 + y_1^2 - 8\end{bmatrix} \rightarrow \begin{bmatrix}x_2\\y_2\end{bmatrix} = \ldots$$

$$\begin{bmatrix}-2.0 & 5.6569\\8.0 & 5.6569\end{bmatrix}\cdot\begin{bmatrix}x_1\\y_1\end{bmatrix} = \begin{bmatrix}-2.0 & 5.6569\\8.0 & 5.6569\end{bmatrix}\cdot\begin{bmatrix}4.0\\2.8284\end{bmatrix} - \begin{bmatrix}0.0\\16.0\end{bmatrix} = \begin{bmatrix}8.0\\16.0\end{bmatrix} \rightarrow \begin{bmatrix}\boldsymbol{x_2}\\\boldsymbol{y_2}\end{bmatrix} = \begin{bmatrix}\boldsymbol{2.40}\\\boldsymbol{2.2627}\end{bmatrix}.$$

Błędy w normie euklidesowej:

$$e_e^{(1)}=\frac{\|\boldsymbol{x}_2-\boldsymbol{x}_1\|_e}{\|\boldsymbol{x}_2\|_e}=\frac{\left\|\begin{bmatrix}2.40\\2.2627\end{bmatrix}-\begin{bmatrix}4.0\\2.8284\end{bmatrix}\right\|_e}{\left\|\begin{bmatrix}2.40\\2.2627\end{bmatrix}\right\|_e}=0.5145,\quad e_e^{(2)}=\frac{\|\boldsymbol{F}(\boldsymbol{x}_2)\|_e}{\|\boldsymbol{F}(\boldsymbol{x}_0)\|_e}=\frac{\left\|\begin{bmatrix}2.2627^2-2\cdot 2.40\\2.40^2+2.2627^2-8.0\end{bmatrix}\right\|_e}{\left\|\begin{bmatrix}8.0\\0.0\end{bmatrix}\right\|_e}=0.6667$$

Błędy w normie maksimum:

$$e_m^{(1)}=\frac{\|\boldsymbol{x}_2-\boldsymbol{x}_1\|_m}{\|\boldsymbol{x}_2\|_m}=\frac{\left\|\begin{bmatrix}2.40\\2.2627\end{bmatrix}-\begin{bmatrix}4.0\\2.8284\end{bmatrix}\right\|_m}{\left\|\begin{bmatrix}2.40\\2.2627\end{bmatrix}\right\|_m}=0.3622,\quad e_m^{(2)}=\frac{\|\boldsymbol{F}(\boldsymbol{x}_2)\|_m}{\|\boldsymbol{F}(\boldsymbol{x}_0)\|_m}=\frac{\left\|\begin{bmatrix}2.2627^2-2\cdot 2.40\\2.40^2+2.2627^2-8.0\end{bmatrix}\right\|_m}{\left\|\begin{bmatrix}8.0\\0.0\end{bmatrix}\right\|_m}=0.3600$$

Sprawdzenie kryterium zbieżności:

$$\varepsilon_e^{(1)}=0.5145>\varepsilon_{dop}^{(1)}=10^{-6}$$
$$\varepsilon_m^{(1)}=0.6667>\varepsilon_{dop}^{(1)}=10^{-6}$$
$$\varepsilon_e^{(2)}=0.3622>\varepsilon_{dop}^{(2)}=10^{-8}$$
$$\varepsilon_m^{(2)}=0.3600>\varepsilon_{dop}^{(2)}=10^{-8}$$

Krok $n=2$:

$$\begin{bmatrix}-2.0 & 4.5255\\4.80 & 4.5255\end{bmatrix}\cdot\begin{bmatrix}x_3\\y_3\end{bmatrix}=\begin{bmatrix}-2.0 & 4.5255\\4.80 & 4.5255\end{bmatrix}\cdot\begin{bmatrix}2.40\\2.2627\end{bmatrix}-\begin{bmatrix}0.320\\2.880\end{bmatrix}=\begin{bmatrix}5.120\\18.88\end{bmatrix}\rightarrow\begin{bmatrix}\boldsymbol{x_3}\\\boldsymbol{y_3}\end{bmatrix}=\begin{bmatrix}\boldsymbol{2.0235}\\\boldsymbol{2.0257}\end{bmatrix}$$

Błędy w normie euklidesowej: $e_e^{(1)}=\frac{\|\boldsymbol{x}_3-\boldsymbol{x}_2\|_e}{\|\boldsymbol{x}_3\|_e}=0.1554,\quad e_e^{(2)}=\frac{\|\boldsymbol{F}(\boldsymbol{x}_3)\|_e}{\|\boldsymbol{F}(\boldsymbol{x}_0)\|_e}=0.1859$

Błędy w normie maksimum: $e_m^{(1)}=\frac{\|\boldsymbol{x}_3-\boldsymbol{x}_2\|_m}{\|\boldsymbol{x}_3\|_m}=0.0257,\quad e_m^{(2)}=\frac{\|\boldsymbol{F}(\boldsymbol{x}_3)\|_m}{\|\boldsymbol{F}(\boldsymbol{x}_0)\|_m}=0.0247$

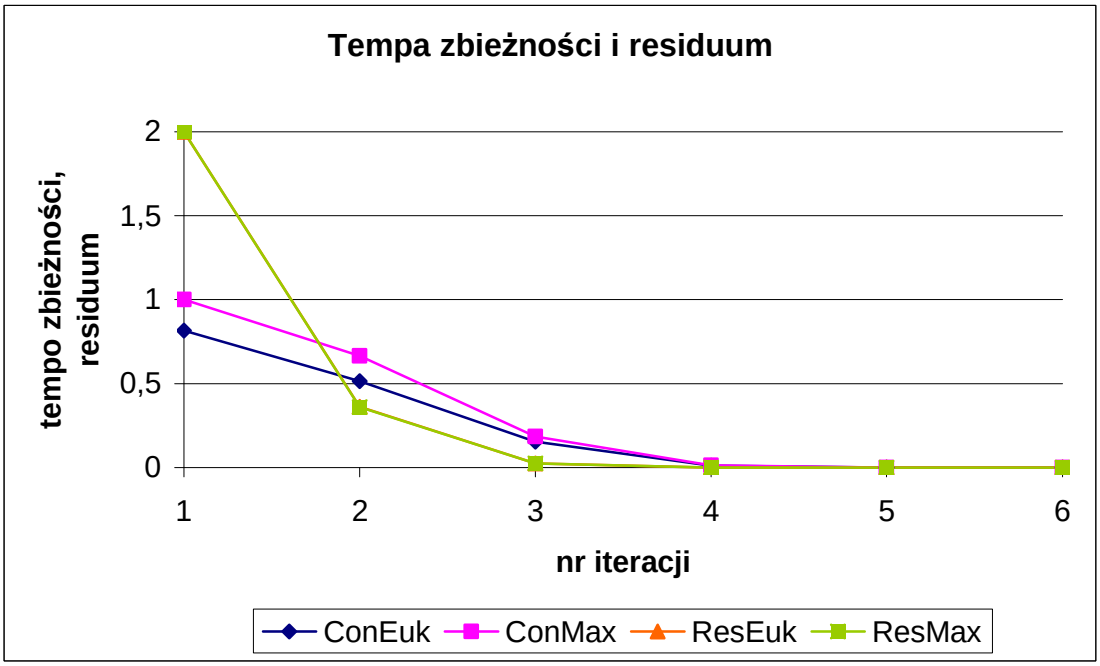
Sprawdzenie kryterium zbieżności:

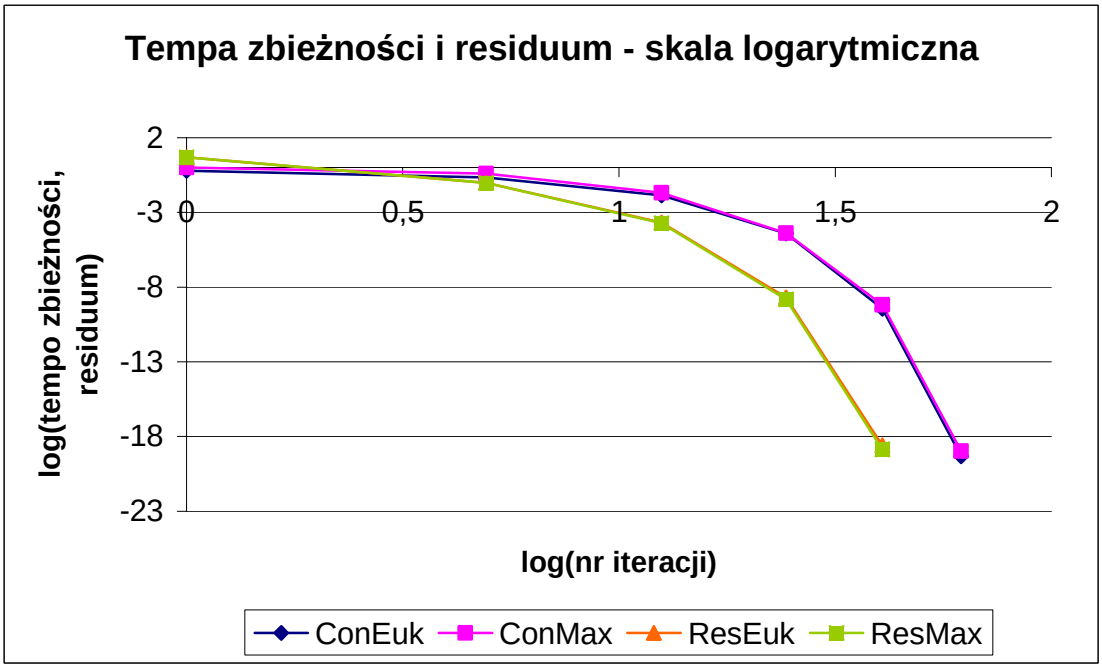
$$\varepsilon_e^{(1)}=0.1554>\varepsilon_{dop}^{(1)}=10^{-6}$$
$$\varepsilon_m^{(1)}=0.1859>\varepsilon_{dop}^{(1)}=10^{-6}$$
$$\varepsilon_e^{(2)}=0.0257>\varepsilon_{dop}^{(2)}=10^{-8}$$
$$\varepsilon_m^{(2)}=0.0247>\varepsilon_{dop}^{(2)}=10^{-8}$$

itd.

Wyniki spełniające kryteria zbieżności otrzymano po szóstej iteracji $\boldsymbol{x}_6 = \{x_6 = 2.000, y_6 = 2.0000\}$.

Poniższe wykresy przedstawiają tempo zbieżności rozwiązania i residuum: w skali dziesiętnej i logarytmicznej liczone dla obydwu powyżej zastosowanych norm.





Przyśpieszenie zbieżności metodą Aitkena ma sens wtedy, gdy rozwiązanie wyraźnie „skacze", przechodząc cyklicznie z jednej strony na drugą pewnej ustalonej wartości. W przypadku powyższym wyraźnie obserwowana jest zbieżność „od góry", a więc włączenie algorytmu Aitkena nie jest uzasadnione i może popsuć dobre już rozwiązania. Od strony formalnej jego zastosowanie będzie polegało na obliczeniu nowej, poprawionej wartości rozwiązania po trzecim kroku iteracyjnym.

Rozwiązanie uzyskane po trzecim kroku: $\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2.0235 \\ 2.0257 \end{bmatrix}$

Poprawienie współrzędnej x_3: $\overline{x}_3 = \dfrac{x_1 \cdot x_3 - x_2^2}{x_3 - 2x_2 + x_1} = \dfrac{4.0 \cdot 2.0235 - 2.40^2}{2.0235 - 2 \cdot 2.40 + 4.0} = 1.9985$

Poprawienie współrzędnej y_3: $\overline{y}_3 = \dfrac{y_1 \cdot y_3 - y_2^2}{y_3 - 2y_2 + y_1} = \dfrac{2.2627 \cdot 2.0257 - 2.8284^2}{2.0257 - 2 \cdot 2.8284 + 2.2627} = 1.9971$

Następny krok iteracyjny ($n = 3$) uwzględniałby oczywiście już poprawione powyżej wartości rozwiązania:

$$\begin{bmatrix} -2.0 & 3.9943 \\ 3.9971 & 3.9943 \end{bmatrix} \cdot \begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} -2.0 & 3.9943 \\ 3.9971 & 3.9943 \end{bmatrix} \cdot \begin{bmatrix} 1.9985 \\ 1.9971 \end{bmatrix} + \begin{bmatrix} 0.0085 \\ 0.0173 \end{bmatrix} = \begin{bmatrix} 3.9886 \\ 15.9827 \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{x_4} \\ \boldsymbol{y_4} \end{bmatrix} = \begin{bmatrix} \boldsymbol{2.0000} \\ \boldsymbol{2.0000} \end{bmatrix}$$

Wartości rozwiązania po tym kroku z dokładnością do sześciu miejsc po przecinku równają się wynikowi ścisłemu.

UKŁADY RÓWNAŃ LINIOWYCH

W metodzie Newtona – Rhapsona po linearyzacji równań należy rozwiązać układ równań liniowych. Sposób algebraiczny (metoda Cramera) wymaga liczenia wyznaczników i jest dość kłopotliwy. Dlatego wprowadzono szereg metod numerycznych do rozwiązywania takich układów równań.

Rozważany będzie następujący problem algebry:

$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \det(\boldsymbol{A}) \neq 0$

$\boldsymbol{A}$ - macierz współczynników układu ($n \times n$),
$\boldsymbol{x}$ – wektor rozwiązań ($n \times 1$),
$\boldsymbol{b}$ – wektor prawej strony (wyrazy wolne) ($n \times 1$).

Klasyfikacja metod do rozwiązywania powyższego zagadnienia może opierać się na własnościach macierzy współczynników. Wtedy można rozróżnić:

1. macierz symetryczną: $\boldsymbol{A}^T = \boldsymbol{A}$,
2. macierz dodatnio określoną: $\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} > 0, \quad \forall_{\boldsymbol{x} \in \Re^n}$,
3. macierz o dużym rozmiarze: $n >> 1$,
4. macierz o specjalnej strukturze (np. pasmowej).

Metody rozwiązywania można podzielić wtedy na:

- metody eliminacji (polegają na odpowiednim rozkładzie macierzy **A** na czynniki a następnie na wyliczeniu jednego po drugim wszystkich rozwiązań) – są uciążliwe obliczeniowo, ale za to dają wynik ścisły, np. metoda Gaussa – Jordana, metoda Choleskiego;
- metody iteracyjne (polegają na zastosowaniu prostych metod iteracyjnych do każdego z równań algebraicznych z osobna, co daje w rezultacie ciąg wektorów przybliżeń rozwiązania ścisłego), np. metoda Jacobiego, metoda Gaussa – Seidela, metoda Richardsona;
- metody kombinowane (eliminacyjno – iteracyjne);
- metody specjalne, np. metody analizy frontalnej czy metody macierzy rzadkich (macierz ma wiele zer, mało współczynników niezerowych, np. metoda Thomasa).

Metody eliminacji, które zostaną omówione poniżej, polegają na rozkładzie wyjściowej macierzy **A** na czynniki, tzw. czynniki trójkątne **L** i **U**: $\boldsymbol{A} = \boldsymbol{L} \times \boldsymbol{U}$. Macierz dolnotrójkątna $\boldsymbol{L}$ ma następującą własność: współczynniki niezerowe występują jedynie poniżej wyrazów na przekątnej głównej, tj. $\underset{(n\times n)}{\boldsymbol{L}} : l_{ij} = \begin{cases} \neq 0, & j \le i \\ 0, & j > i \end{cases}$, macierz górnotrójkątna $\boldsymbol{U}$ ma własność

odwrotną: współczynniki niezerowe położone są powyżej przekątnej głównej, tj. $\underset{(n\times n)}{\boldsymbol{U}} : u_{ij} = \begin{cases} 0, & j < i \\ \neq 0, & j \le i \end{cases}$. Po znalezieniu tego rozkładu rozwiązuje się tzw. „pozorne” układy równań: krok wprzód: $\boldsymbol{Ly} = \boldsymbol{b}$ oraz krok wstecz: $\boldsymbol{Ux} = \boldsymbol{y}$. Układy, dzięki swojej trójkątnej strukturze, pozwalają na uzyskanie kolejnych rozwiązań rekurencyjnie wiersz po wierszu zaczynając liczenie od góry (przy macierzy dolnotrójkątnej) lub od dołu (przy macierzy górnotrójkątnej).

METODA GAUSSA – JORDANA

$$\boldsymbol{Ax} = \boldsymbol{b}, \quad \det(\boldsymbol{A}) \neq 0$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, 2, ..., n$$

Wzory dla wersji eliminacji elementów pod przekątną główną i krokiem wstecz:
$\boldsymbol{Ab} \rightarrow \boldsymbol{Uy} \rightarrow \boldsymbol{Ux} = \boldsymbol{y} \rightarrow \boldsymbol{x}$

$$\begin{aligned} a_{ij}^{(k)} &= a_{ij}^{(k-1)} - m_{ik} \cdot a_{kj}^{(k-1)} \\ b_i^{(k)} &= b_i^{(k-1)} - m_{ik} \cdot b_k^{(k-1)} \end{aligned}$$, gdzie: $m_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad k = 1, 2, ..., n-1; \quad i = k+1, ..., n; \quad j = 1, ..., n$

$$x_i = \left[b_i - \sum_{j=i+1}^{n} a_{ij} x_j \right] \cdot \frac{1}{a_{ii}}, \quad i = n, n-1, ..., 2, 1$$

Wzory dla wersji eliminacji elementów nad przekątną główną i krokiem wprzód: ***Ab*** $\rightarrow$ ***Ly*** $\rightarrow$ ***Lx = y*** $\rightarrow$ ***x***

$$\begin{aligned} a_{ij}^{(k)} &= a_{ij}^{(k-1)} - m_{ik} \cdot a_{kj}^{(k-1)} \\ b_i^{(k)} &= b_i^{(k-1)} - m_{ik} \cdot b_k^{(k-1)} \end{aligned}, \text{ gdzie: } m_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad k = n, n-1, ..., 2; \quad i = k-1, ..., 1; \quad j = 1, ..., n$$

$$x_i = \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j \right] \cdot \frac{1}{a_{ii}}, \quad i = 1, 2, ..., n$$

Wzory dla wersji pełnej eliminacji elementów macierzy: ***Ab*** $\rightarrow$ ***Uy*** $\rightarrow$ ***Lx*** $\rightarrow$ ***x***

$$\begin{aligned} a_{ij}^{(k)} &= a_{ij}^{(k-1)} - m_{ik} \cdot a_{kj}^{(k-1)} \\ b_i^{(k)} &= b_i^{(k-1)} - m_{ik} \cdot b_k^{(k-1)} \end{aligned}, \text{ gdzie: } m_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad k = 1, 2, ..., n-1; \quad i = k+1, ..., n; \quad j = 1, ..., n$$

$$\begin{aligned} a_{ij}^{(k)} &= a_{ij}^{(k-1)} - m_{ik} \cdot a_{kj}^{(k-1)} \\ b_i^{(k)} &= b_i^{(k-1)} - m_{ik} \cdot b_k^{(k-1)} \end{aligned}, \text{ gdzie: } m_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad k = n, n-1, ..., 2; \quad i = k-1, ..., 1; \quad j = 1, ..., n$$

$$x_i = \frac{b_i}{a_{ii}}, \quad i = 1, 2, ..., n$$

W przypadku, gdy przy $\det(\boldsymbol{A}) \neq 0$, a mimo to przy obliczaniu współczynnika m_{ik} wyraz $a_{kk}^{(k-1)} = 0$ należy odwrócić kolejność wierszy (o numerach ”i” oraz „k”) tablicy złożonej z wyrazów macierzy współczynników oraz wyrazów wektora prawej strony. Można też rozwiązywać układy równań z wieloma prawymi stronami, wtedy całą macierz prawych stron ($\underset{(n \times m)}{\boldsymbol{B}} = [b_{ij}]$, gdzie *m* jest liczbą prawych stron) przetwarza się równocześnie.

Przykład 1

Rozwiązać metodą eliminacji Gaussa – Jordana układ równań $\boldsymbol{Ax} = \boldsymbol{b}$, gdzie $\boldsymbol{A} = \begin{bmatrix} 1 & -2 & -1 \\ -2 & 6 & 3 \\ -1 & 3 & 10 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -6 \\ 19 \\ 35 \end{bmatrix}$.

Zastosowana zostanie wersja pełnej eliminacji wyrazów macierzy do postaci diagonalnej. Z wyrazów macierzy ***A*** oraz wyrazów wektora ***b*** budujemy tablicę liczb:

$$\left| \begin{array}{cccc} 1 & -2 & -1 & -6 \\ -2 & 6 & 3 & 19 \\ -1 & 3 & 10 & 35 \end{array} \right|.$$

W pierwszym kroku eliminacji podlegają elementy pod przekątną główną (z macierzy ***A*** powstanie macierz górnotrójkątna ***U***), kolejno „-2”, „”-1” oraz „3”. Do zerowania „-2”

używamy czynnika eliminacji $m_{21}=\frac{-2}{1}=-2$. Jest on równy ilorazowi wyrazu, który ma się wyzerować („-2") przez odpowiadający mu wyraz stojący w pierwszym wierszu od góry, który nie ulega zmianie (tu: wiersz pierwszy, wyraz „1"). Następnie zmianie podlega każdy wyraz w wierszu drugim (łącznie z ostatnią kolumną wyrazów wolnych) wg przepisu: nowy wyraz równa się różnicy starego wyrazu i iloczynu współczynnika „m" przez wyraz z tej samej kolumny z wiersza górnego niezmiennego dla tego kroku (znowu wiersz pierwszy).
Stąd nowa postać wiersza drugiego:
$a_{21}=-2-(-2)\cdot 1=-2+2=0$, $a_{22}=6-(-2)\cdot(-2)=6-4=2$, $a_{23}=3-(-2)\cdot(-1)=3-2=1$,
$b_2=19-(-2)\cdot(-6)=19-12=7$.

Podobnie dla wyzerowania wyrazu $a_{31}=-1$ współczynnik $m_{31}=\frac{-1}{1}=-1$ a nowy zestaw wyrazów:

$a_{31}=-1-(-1)\cdot 1=-1+1=0$, $a_{32}=3-(-1)\cdot(-2)=3-2=1$, $a_{33}=10-(-1)\cdot(-1)=10-1=9$,
$b_3=35-(-1)\cdot(-6)=35-6=29$.

Tablica wyrazów po tym kroku wygląda następująco:

$$\begin{vmatrix} 1 & -2 & -1 & -6 \\ 0 & 2 & 1 & 7 \\ 0 & 1 & 9 & 29 \end{vmatrix}.$$

Cały proces sprowadza się tak naprawdę do pomnożenia pierwszego równania najpierw przez „-2" i dodaniu go do drugiego a następnie przez „-1" i dodaniu go do trzeciego.

W następnym, ostatnim „górnotrójkątnym" kroku eliminacji podlega „1" (dawne „3"). Współczynnik $m_{32}=\frac{1}{2}$. Teraz wierszem, którym się nie zmienia jest wiersz drugi!. Dlatego w mianowniku jest „2" a nie „-2". Postać nowego wiersza po eliminacji (wyraz pierwszy nie ulega zmianie – można to łatwo sprawdzić, bo stoi nad nim „0"):

$$a_{32}=1-\frac{1}{2}\cdot 2=1-1=0,\ a_{33}=9-\frac{1}{2}\cdot 1=9-\frac{1}{2}=\frac{17}{2},\ b_3=29-\frac{1}{2}\cdot 7=29-\frac{7}{2}=\frac{51}{2}.$$

Tablica wyrazów wygląda teraz następująco:

$$\begin{vmatrix} 1 & -2 & -1 & -6 \\ 0 & 2 & 1 & 7 \\ 0 & 0 & \frac{17}{2} & \frac{51}{2} \end{vmatrix}.$$

Postać macierzy górnotrójkątnej: $\boldsymbol{U} = \begin{bmatrix} 1 & -2 & -1 \\ 0 & 2 & 1 \\ 0 & 0 & \frac{17}{2} \end{bmatrix}$. Macierz dolnotrójkątną tworzy się następująco: $\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & \frac{1}{2} & 1 \end{bmatrix}$. Łatwo sprawdzić, iż $\boldsymbol{LU} = \boldsymbol{A}$.

Teraz eliminacji podlegają wyrazy nad przekątną, kolejno „1”, „-1” oraz „-2”. Do eliminacji „1” współczynnik $m_{23} = \frac{1}{17/2} = \frac{2}{17}$ a do eliminacji „-1”: $m_{13} = \frac{-1}{17/2} = -\frac{2}{17}$.

Postać tablicy po przekształceniach:

$$\begin{vmatrix} 1 & -2 & 0 & -3 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & \frac{17}{2} & \frac{51}{2} \end{vmatrix}.$$

Ostatni krok wymaga wyzerowania „-2”. Ostatnie $m_{12} = \frac{-2}{2} = -1$.
Końcowa postać tablicy (macierz $\boldsymbol{A}$ jest teraz diagonalna):

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & \frac{17}{2} & \frac{51}{2} \end{vmatrix}.$$

Ostatnie przekształcenie polega na podzieleniu ostatniej kolumny wyrazów wolnych przez odpowiednie wyrazy diagonalne ($b_1 = \frac{1}{1} = 1, \quad b_2 = \frac{4}{2} = 2, \quad b_3 = \frac{51}{2} \cdot \frac{2}{17} = 3$). Z własności macierzy jednostkowej ($\begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{vmatrix}$, $\boldsymbol{Ix} = \boldsymbol{b} \rightarrow \boldsymbol{x} = \boldsymbol{b}$) wynika, iż wyrazy wolne są poszukiwanymi rozwiązaniami wyjściowego układu, czyli:

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Przykład 2

Rozwiązać metodą eliminacji Gaussa – Jordana układ równań $\begin{bmatrix} 6 & 2 & 2 & 4 \\ -1 & 2 & 2 & -3 \\ 0 & 1 & 1 & 4 \\ 1 & 0 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}$.

Na podstawie układu budujemy tablicę:

$$\left|\begin{array}{ccccc} 6 & 2 & 2 & 4 & 1 \\ -1 & 2 & 2 & -3 & -1 \\ 0 & 1 & 1 & 4 & 2 \\ 1 & 0 & 2 & 3 & 1 \end{array}\right| \rightarrow \left|\begin{array}{ccccc} 6 & 2 & 2 & 4 & 1 \\ 0 & \frac{7}{3} & \frac{7}{3} & -\frac{7}{3} & -\frac{5}{6} \\ 0 & 1 & 1 & 4 & 2 \\ 0 & -\frac{1}{3} & \frac{5}{3} & \frac{7}{3} & \frac{5}{6} \end{array}\right| \rightarrow \left|\begin{array}{ccccc} 6 & 2 & 2 & 4 & 1 \\ 0 & \frac{7}{3} & \frac{7}{3} & -\frac{7}{3} & -\frac{5}{6} \\ 0 & 0 & 0 & 5 & \frac{33}{14} \\ 0 & 0 & 2 & 2 & \frac{5}{7} \end{array}\right|.$$

Współczynniki do wyzerowania pierwszej kolumny: $m_{21} = -\frac{1}{6},\ m_{31} = 0,\ m_{41} = \frac{1}{6}$, do drugiej: $m_{32} = \frac{3}{7},\ m_{42} = -\frac{1}{7}$. Dalej konieczne jest wyzerowanie wyrazu $a_{43} = 2$. Jednak obliczenie współczynnika m_{43} wymagałoby dzielenia przez zero ($m_{43} = \frac{2}{"0"}$). Czy to oznacza, że wyjściowa macierz była osobliwa? Nie, po prostu wyjściowa kolejność równań powoduje takie ułożenie współczynników macierzy – w takim wypadku należy zamienić kolejność wierszy – w powyższym przykładzie ulegną zamianie wiersze trzeci i czwarty. Wtedy zero wskoczy na właściwe sobie miejsce. Natomiast osobliwość macierzy skutkowałaby w postaci późniejszego dzielenia przez zero w czasie obliczania wyrazów wektora rozwiązań.

Dalsze przekształcenia tablicy:

$$\left|\begin{array}{ccccc} 6 & 2 & 2 & 4 & 1 \\ 0 & \frac{7}{3} & \frac{7}{3} & -\frac{7}{3} & -\frac{5}{6} \\ 0 & 0 & 2 & 2 & \frac{5}{7} \\ 0 & 0 & 0 & 5 & \frac{33}{14} \end{array}\right| \rightarrow \left|\begin{array}{ccccc} 6 & 2 & 2 & 0 & -\frac{31}{35} \\ 0 & \frac{7}{3} & \frac{7}{3} & 0 & \frac{8}{15} \\ 0 & 0 & 2 & 0 & -\frac{8}{35} \\ 0 & 0 & 0 & 5 & \frac{33}{14} \end{array}\right| \rightarrow \left|\begin{array}{ccccc} 6 & 2 & 0 & 0 & -\frac{23}{35} \\ 0 & \frac{7}{3} & 0 & 0 & \frac{8}{15} \\ 0 & 0 & 2 & 0 & -\frac{8}{35} \\ 0 & 0 & 0 & 5 & \frac{33}{14} \end{array}\right|$$

$$\left|\begin{array}{cccc|c} 6 & 0 & 0 & 0 & -\frac{39}{35} \\ 0 & \frac{7}{3} & 0 & 0 & \frac{8}{15} \\ 0 & 0 & 2 & 0 & -\frac{8}{35} \\ 0 & 0 & 0 & 5 & \frac{33}{14} \end{array}\right| \rightarrow \left|\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -\frac{13}{70} \\ 0 & 1 & 0 & 0 & \frac{8}{35} \\ 0 & 0 & 1 & 0 & -\frac{4}{35} \\ 0 & 0 & 0 & 1 & \frac{33}{70} \end{array}\right| \rightarrow \left|\begin{array}{l} x_1 = -\frac{13}{70} \\ x_2 = \frac{8}{35} \\ x_3 = -\frac{4}{35} \\ x_4 = \frac{33}{70} \end{array}\right| .$$

METODA CHOLESKIEGO

Metoda opracowana jest dla macierzy współczynników symetrycznych dodatnio określonych. Dzięki takim własnością macierzy $\boldsymbol{A}$ jest możliwy następujący jej rozkład: $\boldsymbol{A} = \boldsymbol{L} \cdot \boldsymbol{L}^{\boldsymbol{T}}$.
Ze sprawdzeniem symetrii macierzy nie ma na ogół problemów, musi być spełniony warunek: $\boldsymbol{A} = \boldsymbol{A}^T \quad \rightarrow \quad a_{ij} = a_{ji}, \quad i, j = 1, 2, ..., n$. Natomiast badanie dodatniej określoności jest na ogół kłopotliwe, dlatego pomocne może okazać się następujące twierdzenie:

Twierdzenie 1
Jeżeli macierz $\boldsymbol{A}$ o współczynnikach rzeczywistych jest symetryczna i ściśle dominująca na przekątnej głównej i dodatkowo posiada wszystkie elementy diagonalne dodatnie, to macierz $\boldsymbol{A}$ jest dodatnio określona.

Macierz nazywamy ściśle dominującą na przekątnej głównej, jeżeli: $|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad i = 1, 2, 3, ..., n$.

Wzór na rozkład macierzy w metodzie Choleskiego oraz wzory na niewiadome wektory $\boldsymbol{x}$ i $\boldsymbol{y}$:

$$\begin{cases} l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2} \\ l_{ij} = \frac{1}{l_{jj}} (a_{ij} - \sum_{k=1}^{j-1} l_{ik} \cdot l_{jk}) \end{cases} , \quad j = 1, 2, ..., n; \quad i = j+1, ..., n$$

$$y_i = \frac{1}{l_{ii}} (b_i - \sum_{j=1}^{i-1} l_{ij} \cdot y_j), \quad i = 1, ..., n \qquad .$$

$$x_i = \left[y_i - \sum_{j=i+1}^{n} l_{ji} \cdot x_j \right] \cdot \frac{1}{l_{ii}}, \quad i = n, ..., 1$$

Przykład 3

Rozwiązać układ równań $\boldsymbol{Ax}=\boldsymbol{b}$, gdzie $\boldsymbol{A}=\begin{bmatrix}4 & -2 & 0\\ -2 & 5 & -2\\ 0 & -2 & 5\end{bmatrix}$, $\boldsymbol{b}=\begin{bmatrix}-6\\ 3\\ 8\end{bmatrix}$ metodą eliminacji Choleskiego.

Aby zastosować metodę Choleskiego do nieosobliwego układu równań, należy sprawdzić warunek symetrii i dodatniej określoności macierzy $\boldsymbol{A}$. Symetria jest spełniona, gdyż $a_{12}=a_{21}=-2,\quad a_{13}=a_{31}=-1,\quad a_{23}=a_{32}=3.$ Do zbadania dodatniej określoności wykorzystamy tw.1: macierz symetryczna jest dominująca na przekątnej głównej, gdyż: $4>|-2|+|0|=2,\quad 5>|-2|+|-2|=4,\quad 5>|0|+|-2|=2$.

Rozłożenie macierzy $\boldsymbol{A}$ na czynniki trójkątne $\boldsymbol{L}\cdot\boldsymbol{L}^{\boldsymbol{T}}$:

$$\begin{bmatrix}4 & -2 & 0\\ -2 & 5 & -2\\ 0 & -2 & 5\end{bmatrix}=\begin{bmatrix}l_{11} & 0 & 0\\ l_{21} & l_{22} & 0\\ l_{31} & l_{32} & l_{33}\end{bmatrix}\cdot\begin{bmatrix}l_{11} & l_{21} & l_{31}\\ 0 & l_{22} & l_{32}\\ 0 & 0 & l_{33}\end{bmatrix}$$

Dokonując odpowiednich mnożeń wierszy macierzy $\boldsymbol{L}$ i kolumn macierzy $\boldsymbol{L}^{\boldsymbol{T}}$ i porównując wynik z odpowiednim wyrazem macierzy $\boldsymbol{A}$ wyznaczamy nieznane wyrazy l_{ij}:

$$l_{11}\cdot l_{11}=a_{11}=4\quad\rightarrow\quad l_{11}=\sqrt{4}=2$$

$$l_{11}\cdot l_{21}=a_{21}=-2\quad\rightarrow\quad l_{21}=\frac{-2}{l_{11}}=\frac{-2}{2}=-1$$

$$l_{11}\cdot l_{31}=a_{31}=0\quad\rightarrow\quad l_{31}=\frac{0}{l_{11}}=0$$

$$l_{21}^2+l_{22}^2=a_{22}=5\quad\rightarrow\quad l_{22}=\sqrt{5-l_{21}^2}=\sqrt{5-1}=2$$

$$l_{31}\cdot l_{21}+l_{32}\cdot l_{22}=a_{32}=-2\quad\rightarrow\quad l_{32}=\frac{-2-l_{31}\cdot l_{21}}{l_{22}}=\frac{-2-0}{2}=-1$$

$$l_{21}^2+l_{22}^2+l_{33}^2=a_{33}=5\quad\rightarrow\quad l_{33}=\sqrt{5-l_{31}^2-l_{32}^2}=\sqrt{5-0-1}=2$$

Macierz dolnotrójkątna: $\boldsymbol{L}=\begin{bmatrix}l_{11} & 0 & 0\\ l_{21} & l_{22} & 0\\ l_{31} & l_{32} & l_{33}\end{bmatrix}=\begin{bmatrix}2 & 0 & 0\\ -1 & 2 & 0\\ 0 & -1 & 2\end{bmatrix}$.

Macierz górnotrójkątna: $\boldsymbol{U}=\boldsymbol{L}^{T}==\begin{bmatrix}l_{11} & l_{21} & l_{31}\\ 0 & l_{22} & l_{32}\\ 0 & 0 & l_{33}\end{bmatrix}=\begin{bmatrix}2 & -1 & 0\\ 0 & 2 & -1\\ 0 & 0 & 2\end{bmatrix}$.

Krok wprzód $\boldsymbol{Ly}=\boldsymbol{b}$:

$$\begin{bmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -6 \\ 3 \\ 8 \end{bmatrix} \quad \rightarrow \quad \begin{array}{c} y_1 = \dfrac{-6}{2} = -3 \\ y_2 = \dfrac{1}{2}(3 + y_1) = 0 \\ y_3 = \dfrac{1}{2}(8 - y_2) = 4 \end{array} \quad \rightarrow \quad \boldsymbol{y} = \begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix}$$

Krok wstecz $\boldsymbol{L}^T \boldsymbol{x} = \boldsymbol{y}$:

$$\begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix} \quad \rightarrow \quad \begin{array}{c} x_3 = \dfrac{4}{2} = 2 \\ x_2 = \dfrac{1}{2}(0 + x_3) = 1 \\ x_3 = \dfrac{1}{2}(-3 + x_2) = -1 \end{array} \quad \rightarrow \quad \boldsymbol{x} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}.$$

Ostatecznym wektorem rozwiązań jest wektor $\boldsymbol{x}$.

Wymaganie związane z dodatnią określonością macierzy $\boldsymbol{A}$ może być w wyjątkowych sytuacjach niespełnione. Wtedy macierze trójkątne $\boldsymbol{L} \cdot \boldsymbol{L}^T$ istnieją w dziedzinie liczb zespolonych, ale końcowe rozwiązanie jest rzeczywiste o ile tylko układ nie jest osobliwy.

Metody iteracyjne, w odróżnieniu od metod eliminacyjnych, dostarczają w wyniku metod iteracji prostej (z relaksacją) całego zbioru przybliżeń wektora rozwiązań, który przy odpowiedniej liczbie iteracji będzie zbieżny do rozwiązania ścisłego $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$.
W metodach iteracyjnych z każdego z równań wyznaczamy niewiadomą (z „i"-tego równania pochodzi „i-ta" niewiadoma) za pomocą wszystkich pozostałych. Niewiadome te podlegają obliczeniu na podstawie znajomości poprzedniego przybliżenia, na samym początku na znajomości wektora startowego. Tak działa metoda Jacobiego, która zawsze korzysta z wyników z poprzedniej iteracji, natomiast metoda Gaussa – Seidela korzysta z informacji z aktualnej iteracji, jeżeli jest to już możliwe. Metody te są zbieżne, jeżeli macierz $\boldsymbol{A}$ jest dodatnio określona (jest to warunek wystarczający zbieżności).

Sformułowanie problemu: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \det(\boldsymbol{A}) \neq 0 \quad \rightarrow \quad \sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, 2, ..., n$.

METODA JACOBIEGO

$$\begin{cases} \boldsymbol{x}^{(0)} = \{x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}\} \\ x_i^{(k+1)} = \dfrac{1}{a_{ii}}(b_i - \sum\limits_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} \cdot x_j^{(k)}), \quad i = 1, 2, ..., n \end{cases}$$

Po rozłożeniu macierzy $\boldsymbol{A}$ na składniki: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \quad \rightarrow \quad \boldsymbol{L}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{x} + \boldsymbol{U}\boldsymbol{x} = \boldsymbol{b}$ można algorytm sformułować w zapisie macierzowym:

$$\boldsymbol{x}^{(k+1)} = -\boldsymbol{D}^{-1} \cdot (\boldsymbol{L} + \boldsymbol{U})\boldsymbol{x}^{(k)} + \boldsymbol{D}^{-1} \cdot \boldsymbol{b}$$

METODA GAUSSA - SEIDELA

$$\begin{cases} \boldsymbol{x}^{(0)} = \{x_1^{(0)}, x_2^{(0)}, ..., x_n^{(0)}\} \\ x_i^{(k+1)} = \dfrac{1}{a_{ii}}(b_i - \sum\limits_{j=1}^{i-1} a_{ij} \cdot x_j^{(k+1)} - \sum\limits_{j=i+1}^{n} a_{ij} \cdot x_j^{(k)}), \quad i = 1, 2, ..., n \end{cases}$$

W zapisie macierzowym:

$$\boldsymbol{x}^{(k+1)} = -\boldsymbol{D}^{-1} \cdot \boldsymbol{L} \cdot \boldsymbol{x}^{(k+1)} - \boldsymbol{D}^{-1} \cdot \boldsymbol{U} \cdot \boldsymbol{x}^{(k)} + \boldsymbol{D}^{-1} \cdot \boldsymbol{b}$$

Kryteria przerwania procesu iteracyjnego są takie same dla obydwu powyższych metod:

1. kontrola tempa zbieżności: $\varepsilon^{(1)} = \dfrac{\left\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|}{\left\| \boldsymbol{x}^{(k+1)} \right\|} \le \varepsilon_{dop}^{(1)}$,
2. kontrola wielkości residuum: $\varepsilon^{(2)} = \dfrac{\left\| \boldsymbol{A} \cdot \boldsymbol{x}^{(k+1)} - \boldsymbol{b} \right\|}{\left\| \boldsymbol{A} \cdot \boldsymbol{x}_0 - \boldsymbol{b} \right\|} \le \varepsilon_{dop}^{(2)}$.

Zastosowanie relaksacji polega na poprawieniu wektora rozwiązań po każdym kroku iteracyjnym wg wzoru:

$$\tilde{x}_i^{(k+1)} = x_i^{(k)} + \lambda \cdot (x_i^{(k+1)} - x_i^{(k)}),$$

gdzie λ jest parametrem relaksacji (przyjmowanym arbitralnie na początku lub ustalanym dynamicznie po każdym kroku).

Przykład 4

Rozwiązać układ równań $\boldsymbol{Ax} = \boldsymbol{b}$, gdzie $\boldsymbol{A} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 5 & -2 \\ 0 & -2 & 5 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -6 \\ 3 \\ 8 \end{bmatrix}$ metodą iteracji Jacobiego. Przyjąć wektor startowy $\boldsymbol{x}^{(0)} = \{0, 0, 0\}$.

Po zapisaniu tradycyjnym powyższego układu i wyliczeniu z każdego z równań kolejnych niewiadomych, otrzymujemy schemat iteracyjny metody Jacobiego.

$$\begin{cases} 4x_1 - 2x_2 = -6 \\ -2x_1 + 5x_2 - 2x_3 = 3 \\ -2x_2 + 5x_3 = 8 \end{cases} \rightarrow \begin{cases} x_1^{(k+1)} = \dfrac{1}{4}(-6 + 2x_2^{(k)}) \\ x_2^{(k+1)} = \dfrac{1}{5}(3 + 2x_1^{(k)} + 2x_3^{(k)}) \\ x_3^{(k+1)} = \dfrac{1}{5}(8 + 2x_2^{(k)}) \end{cases}.$$

Rozpoczynając obliczenia od wektora startowego $\boldsymbol{x}^{(0)} = \{0,0,0\}$ otrzymujemy ciąg przybliżeń wektora rozwiązań, po każdym kroku kontrolując błąd obliczeń (tempo zbieżności i residuum liczone dla dwóch rodzajów norm: euklidesowej i maksimum):

Iteracja pierwsza ($k = 0$):

$$\begin{cases} x_1^{(1)} = \frac{1}{4}(-6 + 2x_2^{(0)}) = \frac{1}{4}(-6+0) = -1.5 \\ x_2^{(1)} = \frac{1}{5}(3 + 2x_1^{(0)} + 2x_3^{(0)}) = \frac{1}{5}(3+0+0) = 0.6 \\ x_3^{(1)} = \frac{1}{5}(8 + 2x_2^{(0)}) = \frac{1}{5}(8+0) = 1.6 \end{cases}$$

$$\boldsymbol{\varepsilon}^{(1)} = \frac{\left\|\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(0)}\right\|}{\left\|\boldsymbol{x}^{(1)}\right\|}, \quad \begin{cases} \varepsilon_e^{(1)} = 1.0 \\ \varepsilon_m^{(1)} = 1.0 \end{cases}, \quad \boldsymbol{\varepsilon}^{(2)} = \frac{\left\|\boldsymbol{A}\boldsymbol{x}^{(1)} - \boldsymbol{b}\right\|}{\left\|\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b}\right\|}, \quad \begin{cases} \varepsilon_e^{(2)} = 0.163673 \\ \varepsilon_m^{(2)} = 0.150 \end{cases}$$

Iteracja druga ($k = 1$):

$$\begin{cases} x_1^{(2)} = \frac{1}{4}(-6 + 2x_2^{(1)}) = \frac{1}{4}(-6 + 2\cdot 0.6) = -1.2 \\ x_2^{(2)} = \frac{1}{5}(3 + 2x_1^{(1)} + 2x_3^{(1)}) = \frac{1}{5}(3 + 2\cdot(-1.5) + 2\cdot 1.6) = 0.64 \\ x_3^{(2)} = \frac{1}{5}(8 + 2x_2^{(1)}) = \frac{1}{5}(8 + 2\cdot 0.6) = 1.84 \end{cases}$$

$$\boldsymbol{\varepsilon}^{(1)} = \frac{\left\|\boldsymbol{x}_2 - \boldsymbol{x}_1\right\|}{\left\|\boldsymbol{x}_2\right\|}, \quad \begin{cases} \varepsilon_e^{(1)} = 0.1019 \\ \varepsilon_m^{(1)} = 0.1630 \end{cases}, \quad \boldsymbol{\varepsilon}^{(2)} = \frac{\left\|\boldsymbol{A}\boldsymbol{x}_2 - \boldsymbol{b}\right\|}{\left\|\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}\right\|}, \quad \begin{cases} \varepsilon_e^{(2)} = 0.1040 \\ \varepsilon_m^{(2)} = 0.1350 \end{cases}$$

Iteracja trzecia ($k = 2$):

$$\begin{cases} x_1^{(3)} = \frac{1}{4}(-6 + 2x_2^{(2)}) = \frac{1}{4}(-6 + 2\cdot 0.64) = -1.18 \\ x_2^{(3)} = \frac{1}{5}(3 + 2x_1^{(2)} + 2x_3^{(2)}) = \frac{1}{5}(3 + 2\cdot(-1.2) + 2\cdot 1.84) = 0.856 \\ x_3^{(3)} = \frac{1}{5}(8 + 2x_2^{(2)}) = \frac{1}{5}(8 + 2\cdot 0.64) = 1.856 \end{cases}$$

$$\boldsymbol{\varepsilon}^{(1)} = \frac{\left\|\boldsymbol{x}_3 - \boldsymbol{x}_2\right\|}{\left\|\boldsymbol{x}_3\right\|}, \quad \begin{cases} \varepsilon_e^{(1)} = 0.0922 \\ \varepsilon_m^{(1)} = 0.1164 \end{cases}, \quad \boldsymbol{\varepsilon}^{(2)} = \frac{\left\|\boldsymbol{A}\boldsymbol{x}_3 - \boldsymbol{b}\right\|}{\left\|\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}\right\|}, \quad \begin{cases} \varepsilon_e^{(2)} = 0.0589 \\ \varepsilon_m^{(2)} = 0.0540 \end{cases}$$

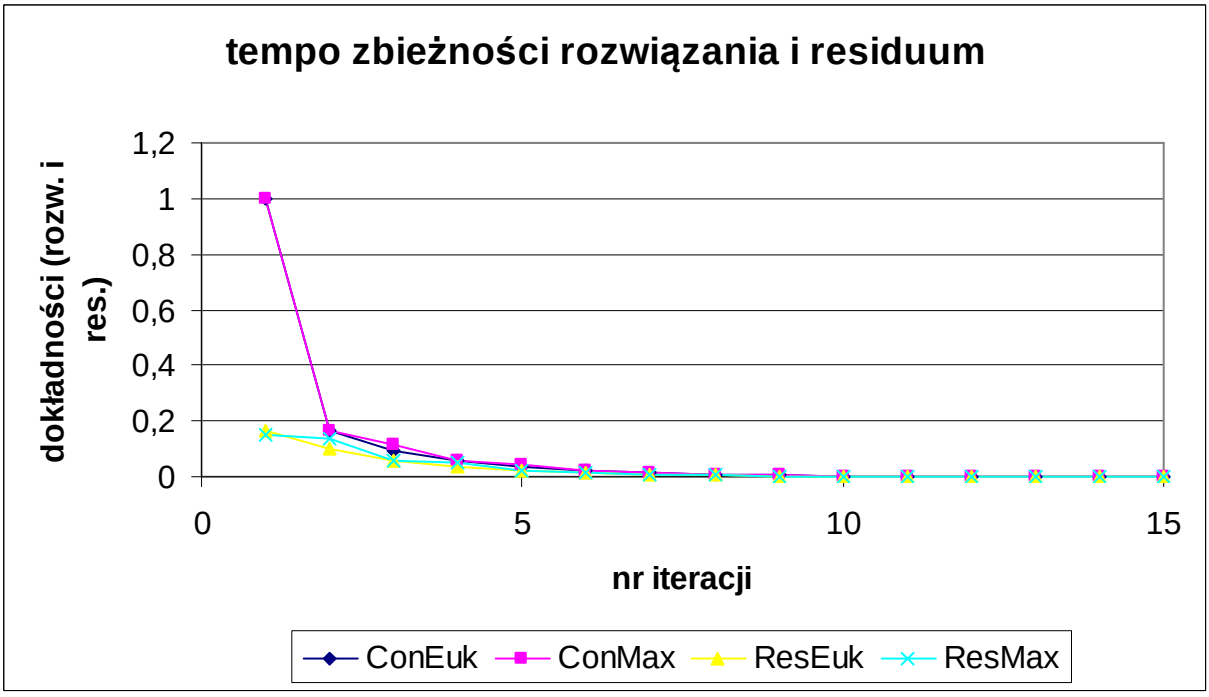
Proces jest bardzo wolno zbieżny do rozwiązania ścisłego $\bar{\boldsymbol{x}} = \{-1,1,2\}$. Po piętnastu iteracjach otrzymano wynik $\boldsymbol{x}^{(15)} = \{$-1.000392, 0.999687, 1.999687$\}$. Aby przyśpieszyć obliczenia, można zastosować technikę, np. nadrelaksacji z parametrem $\lambda = 1.6$. Wtedy poprawione rozwiązania po drugim kroku iteracji wynosić będą:

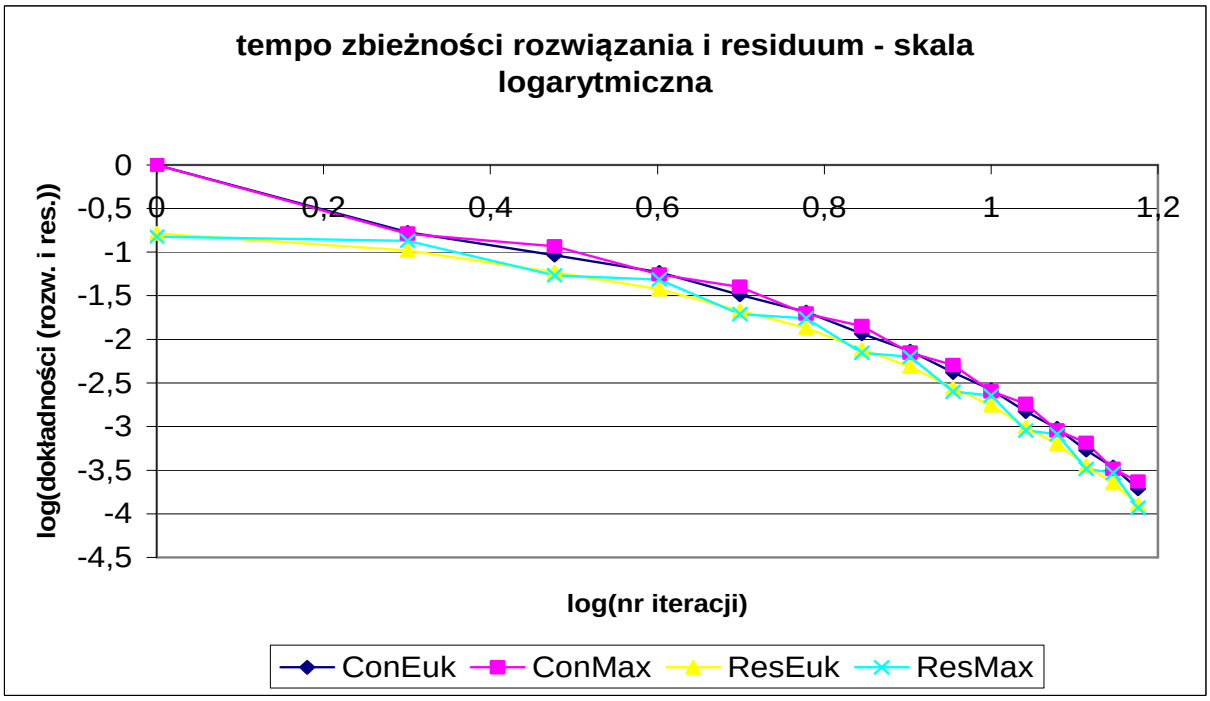
$$\begin{cases} \bar{x}_1^{(2)} = x_1^{(1)} + \lambda \cdot (x_1^{(2)} - x_1^{(1)}) = -1.5 + 1.6 \cdot (-1.2 + 1.5) = -1.02 \\ \bar{x}_2^{(2)} = x_2^{(1)} + \lambda \cdot (x_2^{(2)} - x_2^{(1)}) = 0.6 + 1.6 \cdot (0.64 - 0.6) = 0.664 \\ \bar{x}_3^{(2)} = x_3^{(1)} + \lambda \cdot (x_3^{(2)} - x_3^{(1)}) = 1.6 + 1.6 \cdot (1.84 - 1.6) = 1.984 \end{cases}$$

Dopiero dla tych wyników policzone błędy wynoszą:

$$\varepsilon^{(1)} = \frac{\|\bar{\boldsymbol{x}}_2 - \boldsymbol{x}_1\|}{\|\bar{\boldsymbol{x}}_2\|}, \quad \begin{cases} \varepsilon_e^{(1)} = 0.1019 \\ \varepsilon_m^{(1)} = 0.2419 \end{cases}, \quad \varepsilon^{(2)} = \frac{\|\boldsymbol{A}\bar{\boldsymbol{x}}_2 - \boldsymbol{b}\|}{\|\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}\|}, \quad \begin{cases} \varepsilon_e^{(2)} = 0.1736 \\ \varepsilon_m^{(2)} = 0.2010 \end{cases}$$

Poniższe wykresy przedstawiają tempa zbieżności rozwiązania i residuum równania dla opcji metody bez relaksacji w normach: dziesiętnej i logarytmicznej.





Można spodziewać się większego przyśpieszenia zbieżności po zastosowaniu metody iteracyjnej Gaussa – Seidela.

Przykład 5
Rozwiązać powyższe zadanie metodą iteracji Gaussa – Seidela.

Wyjściowy układ równań: $\boldsymbol{Ax}=\boldsymbol{b}$, gdzie $\boldsymbol{A}=\begin{bmatrix}4 & -2 & 0\\ -2 & 5 & -2\\ 0 & -2 & 5\end{bmatrix}$, $\boldsymbol{b}=\begin{bmatrix}-6\\ 3\\ 8\end{bmatrix}$.

Schemat iteracyjny metody Gaussa – Seidela (zmodyfikowany schemat metody Jacobiego):

$$\begin{cases}4x_1-2x_2=-6\\ -2x_1+5x_2-2x_3=3\\ -2x_2+5x_3=8\end{cases}\rightarrow\begin{cases}x_1^{(k+1)}=\frac{1}{4}(-6+2x_2^{(k)})\\ x_2^{(k+1)}=\frac{1}{5}(3+2x_1^{(k+1)}+2x_3^{(k)})\\ x_3^{(k+1)}=\frac{1}{5}(8+2x_2^{(k+1)})\end{cases}.$$

Tam gdzie to jest możliwe wykorzystuje się już informację „najświeższą” z aktualnego kroku iteracyjnego $k+1$.

Iteracja pierwsza ($k=0$):

$$\begin{cases} x_1^{(1)} = \frac{1}{4}(-6 + 2x_2^{(0)}) = \frac{1}{4}(-6+0) = -1.5 \\ x_2^{(1)} = \frac{1}{5}(3 + 2x_1^{(1)} + 2x_3^{(0)}) = \frac{1}{5}(3 + 2\cdot(-1.5) + 0) = 0.0 \\ x_3^{(1)} = \frac{1}{5}(8 + 2x_2^{(1)}) = \frac{1}{5}(8+0) = 1.6 \end{cases}$$

$$\varepsilon^{(1)} = \frac{\left\| \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(0)} \right\|}{\left\| \boldsymbol{x}^{(1)} \right\|}, \quad \begin{cases} \varepsilon_e^{(1)} = 1.0 \\ \varepsilon_m^{(1)} = 1.0 \end{cases}, \quad \varepsilon^{(2)} = \frac{\left\| \boldsymbol{A}\boldsymbol{x}^{(1)} - \boldsymbol{b} \right\|}{\left\| \boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} \right\|}, \quad \begin{cases} \varepsilon_e^{(2)} = 0.3065 \\ \varepsilon_m^{(2)} = 0.4000 \end{cases}$$

Iteracja druga ($k = 1$):

$$\begin{cases} x_1^{(2)} = \frac{1}{4}(-6 + 2x_2^{(1)}) = \frac{1}{4}(-6 + 2\cdot 0.0) = -1.50 \\ x_2^{(2)} = \frac{1}{5}(3 + 2x_1^{(2)} + 2x_3^{(1)}) = \frac{1}{5}(3 + 2\cdot(-1.50) + 2\cdot 1.6) = 0.640 \\ x_3^{(2)} = \frac{1}{5}(8 + 2x_2^{(2)}) = \frac{1}{5}(8 + 2\cdot 0.64) = 1.8560 \end{cases}$$

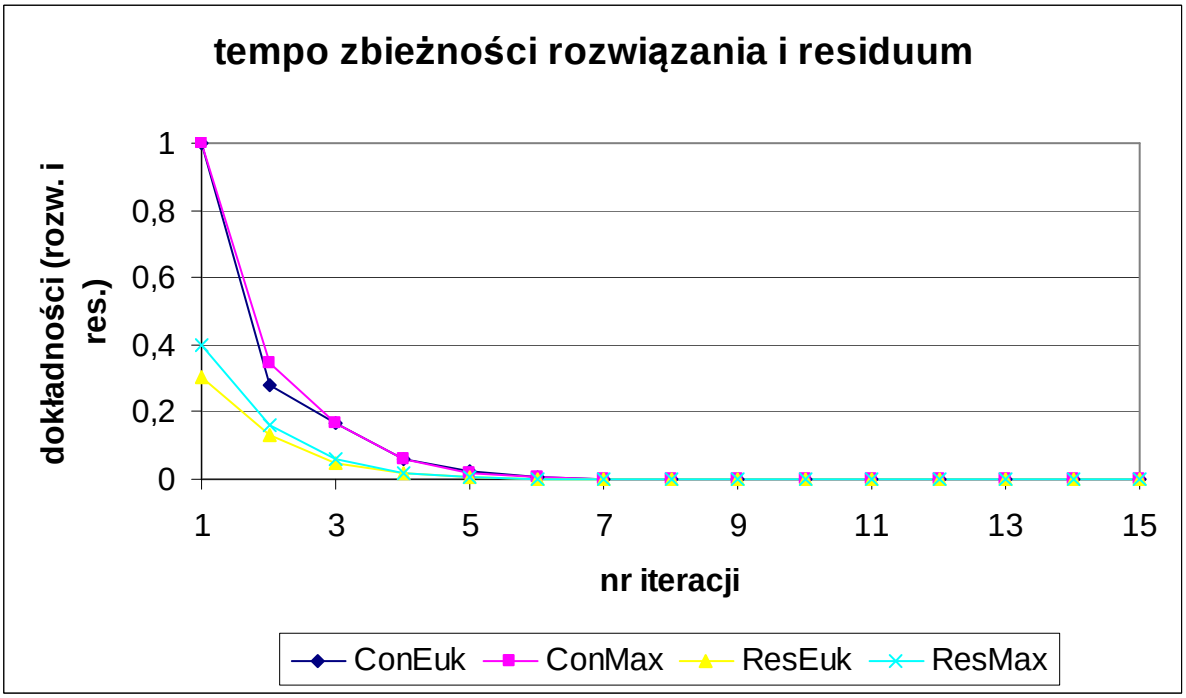
$$\varepsilon^{(1)} = \frac{\left\| \boldsymbol{x}_2 - \boldsymbol{x}_1 \right\|}{\left\| \boldsymbol{x}_2 \right\|}, \quad \begin{cases} \varepsilon_e^{(1)} = 0.2790 \\ \varepsilon_m^{(1)} = 0.3448 \end{cases}, \quad \varepsilon^{(2)} = \frac{\left\| \boldsymbol{A}\boldsymbol{x}_2 - \boldsymbol{b} \right\|}{\left\| \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b} \right\|}, \quad \begin{cases} \varepsilon_e^{(2)} = 0.1320 \\ \varepsilon_m^{(2)} = 0.1600 \end{cases}$$

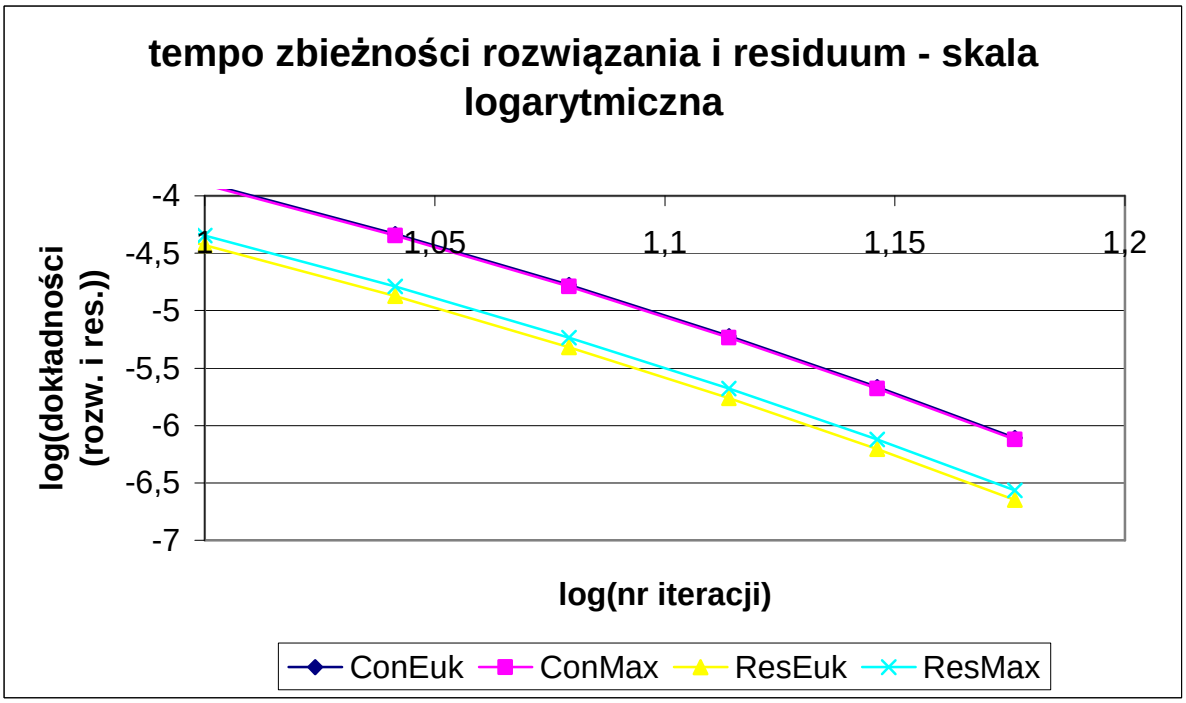
Iteracja trzecia ($k = 2$):

$$\begin{cases} x_1^{(3)} = \frac{1}{4}(-6 + 2x_2^{(2)}) = \frac{1}{4}(-6 + 2\cdot 0.640) = -1.18 \\ x_2^{(3)} = \frac{1}{5}(3 + 2x_1^{(3)} + 2x_3^{(2)}) = \frac{1}{5}(3 + 2\cdot(-1.18) + 2\cdot 1.856) = 0.8704 \\ x_3^{(3)} = \frac{1}{5}(8 + 2x_2^{(3)}) = \frac{1}{5}(8 + 2\cdot 0.8704) = 1.9482 \end{cases}$$

$$\varepsilon^{(1)} = \frac{\left\| \boldsymbol{x}_3 - \boldsymbol{x}_2 \right\|}{\left\| \boldsymbol{x}_3 \right\|}, \quad \begin{cases} \varepsilon_e^{(1)} = 0.1661 \\ \varepsilon_m^{(1)} = 0.1643 \end{cases}, \quad \varepsilon^{(2)} = \frac{\left\| \boldsymbol{A}\boldsymbol{x}_3 - \boldsymbol{b} \right\|}{\left\| \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b} \right\|}, \quad \begin{cases} \varepsilon_e^{(2)} = 0.0475 \\ \varepsilon_m^{(2)} = 0.0576 \end{cases}$$

Po piętnastu iteracjach otrzymano rozwiązanie $\boldsymbol{x}^{(15)} = \{-1.0000,\ 1.0000,\ 2.0000\}$ z dokładnością do sześciu miejsc po przecinku. Wykresy zbieżności przedstawiono poniżej.


tempo zbieżności rozwiązania i residuum
dokładności (rozw. i res.)
1
0,8
0,6
0,4
0,2
0
1
3
5
7
9
11
13
15
nr iteracji
ConEuk
ConMax
ResEuk
ResMax


tempo zbieżności rozwiązania i residuum - skala logarytmiczna
log(dokładności (rozw. i res.))
-4
-4,5
-5
-5,5
-6
-6,5
-7
1
1,05
1,1
1,15
1,2
log(nr iteracji)
ConEuk
ConMax
ResEuk
ResMax

ODWRACANIE MACIERZY

Odwracanie macierzy dolnotrójkątnej

Dana jest macierz dolnotrójkątna $\boldsymbol{L}$ o wymiarze *n*, szukana jest macierz $\boldsymbol{C}$ taka, że $\boldsymbol{L}\cdot\boldsymbol{C}=\boldsymbol{I}$. Macierz $\boldsymbol{C}$, odwrotna do macierzy $\boldsymbol{L}$ jest również macierzą dolnotrójkątna.

Wzory ogólne:

$$\begin{cases} c_{ii}=\dfrac{1}{l_{ii}} \quad i=1,2,...,n \\ c_{ij}=-\dfrac{1}{l_{ii}}\sum_{k=j}^{i-1} l_{ik}\cdot c_{kj} \quad i=1,2,...,n \qquad j=1,2,...,i-1 \end{cases}$$

Przykład 13
Odwrócić macierz dolnotrójkątna.

$$\boldsymbol{L}=\begin{bmatrix}1&0&0\\2&4&0\\3&5&6\end{bmatrix} \qquad \boldsymbol{C}=\begin{bmatrix}c_{11}&0&0\\c_{21}&c_{22}&0\\c_{31}&c_{32}&c_{33}\end{bmatrix}$$

$$\boldsymbol{L}\cdot\boldsymbol{C}=\boldsymbol{I} \;\Rightarrow\; \begin{bmatrix}1&0&0\\2&4&0\\3&5&6\end{bmatrix}\cdot\begin{bmatrix}c_{11}&0&0\\c_{21}&c_{22}&0\\c_{31}&c_{32}&c_{33}\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

Dokonujemy mnożeń odpowiednich wierszy macierzy $\boldsymbol{L}$ i kolumn macierzy $\boldsymbol{C}$ tak, aby wyznaczyć wyrazy macierzy $\boldsymbol{C}$ za każdym razem porównując wyniki tych mnożeń z odpowiednim wyrazem macierzy jednostkowej.

$c_{11}\cdot 1+c_{21}\cdot 0+c_{31}\cdot 0=1 \;\rightarrow\; c_{11}=1$

$c_{11}\cdot 2+c_{21}\cdot 4+c_{31}\cdot 6=0 \;\rightarrow\; c_{21}=-\dfrac{1}{2}$

$c_{11}\cdot 3+c_{21}\cdot 5+c_{31}\cdot 0=0 \;\rightarrow\; c_{31}=-\dfrac{1}{12}$

$0\cdot 2+c_{22}\cdot 4+c_{32}\cdot 6=1 \;\rightarrow\; c_{22}=\dfrac{1}{4}$

$3\cdot 0+c_{22}\cdot 5+c_{32}\cdot 6=0 \;\rightarrow\; c_{32}=-\dfrac{5}{24}$

$0\cdot 3+0\cdot 5+c_{33}\cdot 0=1 \;\rightarrow\; c_{33}=\dfrac{1}{6}$

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{4} & 0 \\ -\frac{1}{12} & -\frac{5}{24} & \frac{1}{6} \end{bmatrix}$$

Odwracanie macierzy górnotrójkątnej

Dana jest macierz górnotrójkątna $\boldsymbol{U}$ o wymiarze n, szukana jest macierz $\boldsymbol{C}$ taka, że $\boldsymbol{U} \cdot \boldsymbol{C} = \boldsymbol{I}$. Macierz $\boldsymbol{C}$, odwrotna do macierzy $\boldsymbol{U}$ jest również macierzą górnotrójkątna.

Wzory ogólne:

$$\begin{cases} c_{ii} = \dfrac{1}{u_{ii}} \quad i = n,...,1 \\ c_{ij} = -\dfrac{1}{u_{ii}} \sum_{k=i+1}^{j} u_{ik} \cdot c_{kj} \quad i = n,...,1 \qquad j = n,...,i+1 \end{cases}$$

Przykład 14
Odwrócić macierz górnotrójkątna.

$$\boldsymbol{U} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \qquad \boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ 0 & c_{22} & c_{23} \\ 0 & 0 & c_{33} \end{bmatrix}$$

$$\boldsymbol{U} \cdot \boldsymbol{C} = \boldsymbol{I} \quad \Rightarrow \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \cdot \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ 0 & c_{22} & c_{23} \\ 0 & 0 & c_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Postępowanie jest identyczne jak w przypadku macierzy dolnotrójkątnej.

$$c_{11}\cdot 1+2\cdot 0+3\cdot 0=1 \quad \rightarrow \quad c_{11}=1$$

$$c_{12}\cdot 0+c_{22}\cdot 4+5\cdot 0=0 \quad \rightarrow \quad c_{22}=\frac{1}{4}$$

$$c_{13}\cdot 0+c_{23}\cdot 0+c_{33}\cdot 6=0 \quad \rightarrow \quad c_{33}=\frac{1}{6}$$

$$c_{12}\cdot 1+c_{22}\cdot 2+3\cdot 0=1 \quad \rightarrow \quad c_{12}=-\frac{1}{2}$$

$$c_{13}\cdot 0+c_{23}\cdot 4+c_{33}\cdot 5=0 \quad \rightarrow \quad c_{23}=-\frac{5}{24}$$

$$c_{13}\cdot 1+c_{23}\cdot 2+c_{33}\cdot 3=1 \quad \rightarrow \quad c_{13}=-\frac{1}{12}$$

$$\boldsymbol{C}=\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{12} \\ 0 & \frac{1}{4} & -\frac{5}{24} \\ 0 & 0 & \frac{1}{6} \end{bmatrix}$$

Metoda Choleskiego

Jest to metoda odwracania macierzy symetrycznych, dodatnio określonych. Polega ona na rozłożeniu wyjściowej macierzy na czynniki trójkątne: $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{L}^T$ a następnie na odwróceniu każdego z nich osobno i wymnożeniu tak, że: $\boldsymbol{A}^{-1} = \boldsymbol{L}^{-T}\,\boldsymbol{L}^{-1}$.

Wzory na rozkład macierzy $\boldsymbol{A}$ na czynniki trójkątne:

$$\begin{cases} l_{jj}=\sqrt{a_{jj}-\sum_{k=1}^{j-1} l_{jk}^2} & j=1,\ldots,n \\ l_{ij}=\frac{1}{l_{jj}}\left(a_{ij}-\sum_{k=1}^{j-1} l_{ik}\cdot l_{jk}\right) & i=j+1,\ldots,n \end{cases}$$

Po uzyskaniu macierzy dolnotrójkątnej $\boldsymbol{L}$ i górnotrójkątnej $\boldsymbol{L}^T$ odwraca się je korzystając ze wzorów zaprezentowanych w poprzednich podrozdziałach, a następnie mnoży obydwie macierze odwrotne, ale w odwrotnej kolejności.

Metody powiązane z rozwiązywaniem układów równań

Z definicji macierzy odwrotnej do macierzy $\boldsymbol{A}$ wynika następująca zależność:

$$\boldsymbol{A}\cdot\boldsymbol{A}^{-1}=\boldsymbol{I} \quad \Rightarrow \quad \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Powyższy zapis można rozbić na *n* układów równań, z których każdy służy do obliczenia kolejnej, *k-tej* kolumny macierzy $\boldsymbol{A}^{-1}$.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} c_{1k} \\ c_{2k} \\ \dots \\ c_{nk} \end{bmatrix} = \begin{bmatrix} b_{1k} \\ b_{2k} \\ \dots \\ b_{nk} \end{bmatrix} \quad k=1,2,...,n\,,$$

gdzie wyrazy wektora prawej strony: $b_{jk}=\begin{cases} 0 & dla \quad j\neq k \\ 1 & dla \quad j=k \end{cases}$.

W zależności od metody rozwiązywania tych układów równań można mówić o metodach eliminacji (np. *metoda eliminacji Gaussa* – wtedy rozwiązuje się jeden układ, ale z *n* prawymi stronami) lub metodach iteracyjnych (np. *metoda Jacobiego* lub *metoda Gaussa-Seidla*).

Metoda eliminacji Gaussa

Transformacji podlegają wyjściowa macierz nieosobliwa A oraz macierz C, która na początku obliczeń jest macierzą jednostkową, tzn. $C_{ij}=\begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases}$.

$$\begin{aligned} a_{ij}^{(k)} &= a_{ij}^{(k-1)} - m_{ik}\cdot a_{kj}^{(k-1)} \\ c_{ij}^{(k)} &= c_{ij}^{(k-1)} - m_{ik}\cdot c_{kj}^{(k-1)} \end{aligned}, \text{ gdzie: } m_{ik}=\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad k=1,2,...,n-1; \quad i=k+1,...,n; \quad j=1,...,n$$

$$\begin{aligned} a_{ij}^{(k)} &= a_{ij}^{(k-1)} - m_{ik}\cdot a_{kj}^{(k-1)} \\ c_{ij}^{(k)} &= c_{ij}^{(k-1)} - m_{ik}\cdot c_{kj}^{(k-1)} \end{aligned}, \text{ gdzie: } m_{ik}=\frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad k=n,n-1,...,2; \quad i=k-1,...,1; \quad j=1,...,n$$

$$c_{ij}=\frac{c_{ij}}{a_{ii}}, \quad i,j=1,2,...,n$$

NADOKREŚLONY UKŁAD RÓWNAŃ

Jeżeli w danym układzie równań liniowych $\boldsymbol{A}x = b$ jest więcej równań niż niewiadomych zmiennych, to taki układ nazywa się *nadokreślonym.* Jeżeli wszystkie równania są liniowo niezależne, to układ nie ma jednego wspólnego rozwiązania, tj. punktu, w którym wszystkie proste przecinają się.
W takim wypadku szuka się tzw. *pseudorozwiązania*, czyli punktu, który nie leży na żadnej prostej, ale jego odległości od każdej z prostych są minimalne w sensie jakieś normy.

Niech $\boldsymbol{A}$ będzie macierzą $n \times m$, gdzie *n* (liczba wierszy) oznacza liczbę równań, natomiast *m* (liczba kolumn) oznacza liczbę niewiadomych. W *układzie nadokreślonym: n>m*, w *układzie niedookreślonym: n<m*. Niech *b* będzie wektorem wyrazów wolnych o wymiarze *n*.
W pierwszym kroku buduje się wektor $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ zawierający odległości prostych od *pseudorozwiązania*. Następnie szuka się $\min \|\varepsilon\|$. Jeżeli zastosujemy normę średniokwadratową: $\|\varepsilon\| = \sqrt{\varepsilon_1^2 + \varepsilon_2^2 + ... + \varepsilon_n^2}$, to dalsze postępowanie nazywa się *metodą najmniejszych kwadratów*. Można też stosować normę maksimum.

Metoda najmniejszych kwadratów

Zapis wskaźnikowy (korzystny przy obliczeniach ręcznych):

$B = \sum_{i=1}^{n} (\sum_{j=1}^{m} a_{ij} x_j - b_i)^2$ - funkcjonał błędu

$\frac{\partial B}{\partial x_k} = 2\sum_{i=1}^{n} a_{ik} (\sum_{j=1}^{m} a_{ij} x_j - b_i) = 0$ - minimalizacja funkcjonału

Nowy układ równań liniowych (wymiar: $m \times m$):

$$\sum_{i=1}^{n} a_{ik} \sum_{j=1}^{m} a_{ij} x_j = \sum_{i=1}^{n} a_{ik} b_i, \quad k = 1, 2, ..., m$$

Zapis macierzowy (korzystny przy implementacji komputerowej):

$$\boldsymbol{B} = (\boldsymbol{A}x - b) \cdot (\boldsymbol{A}x - b)^T$$
$$\frac{\partial \boldsymbol{B}}{\partial x} = 2\boldsymbol{A}^{\boldsymbol{T}} (\boldsymbol{A}x - b) = 0$$
$$\boldsymbol{A}^T \boldsymbol{A}\, x = \boldsymbol{A}^T b$$

Przykład 11
Rozwiązać nadokreślony układ równań.

$$\begin{cases} x + y = 2 \\ x - y = 0 \\ x - 2y = -2 \end{cases} \Rightarrow \begin{cases} x + y - 2 = 0 \\ x - y = 0 \\ x - 2y + 2 = 0 \end{cases}$$

$$B(x,y) = \left\|\varepsilon(x,y)\right\|^2 = (x+y-2)^2 + (x-y)^2 + (x-2y+2)^2$$

$$\begin{cases} \dfrac{\partial B}{\partial x} = 2\cdot(x+y-2) + 2\cdot(x-y) + 2\cdot(x-2y+2) = 0 \\ \dfrac{\partial B}{\partial y} = 2\cdot(x+y-2) - 2\cdot(x-y) - 4\cdot(x-2y+2) = 0 \end{cases}$$

$$\begin{cases} 3x-2y=0 \\ -2x+6y=6 \end{cases} \Rightarrow \begin{cases} \boldsymbol{x_0 = \frac{6}{7} \approx 0.857143} \\ \boldsymbol{y_0 = \frac{9}{7} \approx 1.285714} \end{cases} \qquad \textbf{\textit{pseudorozwiązanie.}}$$

Można też policzyć maksymalny błąd tego wyniku: $B_{\max} = B(x_0, y_0) = 0.285714$

Czasami stosuje się też tzw. *ważoną metodę najmniejszych kwadratów*. Aby zwiększyć lub zmniejszyć wpływ jednego z równań na wynik końcowy, można przypisać każdemu z równań wagę (funkcję lub liczbę) – im większą tym bliżej tej prostej będzie leżało pseudorozwiązanie.
Wprowadza się diagonalną macierz wagową: $W = diag\{w_{ii}\},\ i = 1,2,...,n$ zbierającą wagi przypisane wszystkim równaniom. Odpowiednie modyfikacje ostatecznych układów równań są następujące:

w zapisie wskaźnikowym: $\sum_{i=1}^{n} a_{ik} \sum_{j=1}^{m} w_{ii} a_{ij} x_j = \sum_{i=1}^{n} w_{ii} a_{ik} b_i, \quad k = 1,2,...,m$

w zapisie macierzowym: $\boldsymbol{A^T W A}\, x = \boldsymbol{A^T W}\, b$

Przykład 12
Rozwiązań nadokreślony układ równań z przykładu 11, przypisując każdemu z równań wagę będącą jego numerem kolejnym.

Wagi: $w_{11} = 1, \quad w_{22} = 2, \quad w_{33} = 3$
Funkcjonał błędu: $B(x,y) = 1\cdot(x+y-2)^2 + 2\cdot(x-y)^2 + 3\cdot(x-2y+2)^2$

$$\begin{cases} \dfrac{\partial B}{\partial x} = 2\cdot 1\cdot(x+y-2) + 2\cdot 2\cdot(x-y) + 2\cdot 3\cdot(x-2y+2) = 0 \\ \dfrac{\partial B}{\partial y} = 2\cdot 1\cdot(x+y-2) - 2\cdot 2\cdot(x-y) - 4\cdot 3\cdot(x-2y+2) = 0 \end{cases}$$

$$\begin{cases} 12x-14y=-8 \\ -14x+30y=28 \end{cases} \Rightarrow \begin{cases} \boldsymbol{x_0 = 0.926829} \\ \boldsymbol{y_0 = 1.365854} \end{cases}, \quad B_{\max} = 0.585366$$

WARTOŚCI WŁASNE I WEKTORY WŁASNE MACIERZY

Wartościami własnymi macierzy ***A*** stopnia *n* nazywamy takie wartości $\lambda_1, \lambda_2, ..., \lambda_n$ parametru λ, dla których układ równań

$$\boldsymbol{A}x = \lambda x \tag{1}$$

ma niezerowe rozwiązanie.
Wektor x_r, spełniający przy $\lambda = \lambda_r$ układ równań (1), nazywamy *wektorem własnym macierzy* ***A***. Układ (1) ma niezerowe rozwiązanie wtedy, gdy jego wyznacznik jest równy zero, tzn.

$$(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$$

Po rozwinięciu powyższego wyznacznika otrzymamy równanie algebraiczne stopnia *n* :

$$a_0 + a_1\lambda + a_2\lambda^2 + ... + (-1)^n \lambda^n = 0$$

zwane *równaniem charakterystycznym macierzy* ***A***. Pierwiastki tego równania są oczywiście wartościami własnymi macierzy ***A***

Przykład 1
Niech

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 3 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}$$

Znajdziemy teraz równanie charakterystyczne macierzy ***A***

$$\left|\boldsymbol{A} - \lambda \boldsymbol{I}\right| = \begin{vmatrix} 1-\lambda & 0 & 0 & 4 \\ 0 & 3-\lambda & 2 & 0 \\ 1 & 0 & -\lambda & 0 \\ 1 & 1 & 0 & 2-\lambda \end{vmatrix}$$

Rozwijając ten wyznacznik według elementów pierwszego wiersza, otrzymujemy

$$(1-\lambda)\begin{vmatrix} 3-\lambda & 2 & 0 \\ 0 & -\lambda & 0 \\ 1 & 0 & 2-\lambda \end{vmatrix} - 4\begin{vmatrix} 0 & 3-\lambda & 2 \\ 1 & 0 & \lambda \\ 1 & 1 & 0 \end{vmatrix} =$$

$$= (1-\lambda)(3-\lambda)(-\lambda)(2-\lambda) - 4[(3-\lambda)(-\lambda) + 2] = (\lambda-4)(\lambda-2)(\lambda-1)(\lambda+1)$$

Wartości własne macierzy $\boldsymbol{A}$ są równe $\lambda_1 = 4, \quad \lambda_2 = 2, \quad \lambda_3 = 1, \quad \lambda_4 = -1$.
Aby otrzymać wektory własne, należy rozwiązać układ równań $\boldsymbol{A}x = \lambda x$, gdzie zamiast λ będziemy podstawiać kolejne obliczone wartości własne.
Podstawiając $\lambda = 4$, oraz oznaczając współrzędne wektora własnego przez v_1, v_2, v_3, v_4, otrzymujemy następujący układ

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 3 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = 4 \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

lub po rozpisaniu

$$\begin{cases} v_1 + 4v_4 = 4v_1 \\ 3v_2 + 2v_3 = 4v_2 \\ v_1 = 4v_3 \\ v_1 + v_2 + 2v_4 = 4v_4 \end{cases}$$

skąd obliczamy $v_1 = 4v_3, \quad v_2 = 2v_3, \quad v_4 = 3v_3$.
Oczywiście wektor własny nie jest określony jednoznacznie. Jeżeli dodatkowo dokonać jego normalizacji, np. zażądać, aby jego największa współrzędna była równa jedności to wtedy otrzymamy

$$x_1 = (1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4})$$

Podobnie otrzymamy pozostałe wektory własne

$$x_2 = (1, -1, \frac{1}{2}, \frac{1}{4}), \quad x_3 = (1, -1, 1, 0), \quad x_4 = (-1, -\frac{1}{2}, 1, \frac{1}{2})$$

Można oczywiście inaczej znormalizować dany wektor *x*, np. tak, aby jego długość była równa jedności, tzn.

$$\|x\| = \sqrt{v_1^2 + v_2^2 + ... + v_n^2} = 1$$

Podana w powyższym przykładzie metoda znajdywania wartości własnych oraz wektorów własnych jest bardzo pracochłonna, szczególnie w przypadku macierzy wysokiego stopnia.
Dlatego też rzadko rozwiązuje się problem własny macierzy na podstawie definicji. Szczególnie kłopotliwe może być wyznaczenie samych wartości własnych, gdy wielomian występujący w równaniu charakterystycznym nie ma pierwiastków wymiernych.

Przykład 2
Niech

$$\boldsymbol{A}=\begin{bmatrix} 1 & 3 & -1 \\ 1 & 2 & 4 \\ -1 & 2 & 3 \end{bmatrix}.$$

Równanie charakterystyczne

$$(\boldsymbol{A}-\lambda\boldsymbol{I})=\begin{vmatrix} 1-\lambda & 3 & -1 \\ 1 & 2-\lambda & 4 \\ -1 & 2 & 3-\lambda \end{vmatrix}$$

po rozwinięciu (np. względem pierwszego wiersza) ma postać następującego wielomianu

$$\lambda^3-6\lambda^2-\lambda+27=0$$

Wielomian ten nie posiada pierwiastków wymiernych, (co łatwo sprawdzić, gdyż mogłyby one wynosić odpowiednio $\lambda_i=1,3,9,27$ ale żadna z tych liczb nie spełnia równania). Równanie trzeciego stopnia posiada odpowiednie wzory na swoje pierwiastki rzeczywiste, (jeżeli istnieją) – tzw. *wzory Cardana*, ale są one dość uciążliwe w użyciu. Dlatego posłużymy się w tym przypadku *metodami numerycznymi* dla określenia jednego z pierwiastków, aby pozostałe dwa wyznaczyć już w sposób analityczny. Budując z powyższego wielomianu schemat iteracyjny dla *metody Newtona*

$$F(\lambda)=\lambda^3-6\lambda^2-\lambda+27$$
$$\lambda_{n+1}=\lambda_n-\frac{F(\lambda_n)}{F'(\lambda_n)}=\lambda_n-\frac{\lambda_n^3-6\lambda_n^2-\lambda_n+27}{3\lambda_n^2-12\lambda_n-1}$$

oraz startując np. z $\lambda_0=1$ otrzymujemy dla czterech kolejnych iteracji

$$\lambda_1=3.1\quad \lambda_2=2.676414\quad \lambda_3=2.720801\quad \lambda_4=2.721158$$

Ostatni wynik można uznać już za satysfakcjonujący gdyż odpowiadające mu tempo zbieżności $\frac{|\lambda_3-\lambda_4|}{|\lambda_4|}=0.000131$ jest relatywnie małą liczba.

Zatem przyjmujemy do dalszych obliczeń $\lambda=\lambda_4=2.721158$. W celu wyznaczenia pozostałych pierwiastków równania dzielimy wyjściowy wielomian przez $(\lambda-2.721158)$ otrzymując w rezultacie

$$\lambda^3-6\lambda^2-\lambda+27=(\lambda-2.721158)(\lambda^2-3.278842\lambda-9.922247)$$

Równanie kwadratowe rozwiązujemy w znany analityczny sposób wyznaczając pozostałe dwa pierwiastki. Ostatecznie wartości własne macierzy $\boldsymbol{A}$ wynoszą (w kolejności rosnącej)

$$\lambda_1=-1.911628,\quad \lambda_2=2.721158,\quad \lambda_3=5.190470$$

Dla porównania analitycznie policzone wartości własne wynoszą $-1.911629,\ 2.721159,\ 5.190470$. Zatem powyższe wielkości numeryczne są bardzo dobrym przybliżeniem ścisłych wyników analitycznych.

Dalej postępujemy podobnie jak w przykładzie 1 w celu wyznaczenia wektorów własnych. Dla $\lambda_1 = -1.911628$ i odpowiadającego jej wektora $v_1 = (x_1, y_1, z_1)$ budujemy układ równań

$$\begin{bmatrix} 1-\lambda_1 & 3 & -1 \\ 1 & 2-\lambda_1 & 4 \\ -1 & 2 & 3-\lambda_1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 2.911628 & 3 & -1 \\ 1 & 3.911628 & 4 \\ -1 & 2 & 4.911628 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Do dwóch pierwszych równań (trzecie to tożsamość w stosunku do nich) dołączamy warunek na jednostkową długość wektora własnego.

$$\begin{cases} 2.911628\, x_1 + 3\, y_1 - z_1 = 0 \\ x_1 + 3.911628\, y_1 + 4\, z_1 = 0 \\ x_1^2 + y_1^2 + z_1^2 = 1 \end{cases}$$

Rozwiązanie tego układu daje współrzędne wektora v_1

$$x_1 = 0.723635 \quad y_1 = -0.575143 \quad z_1 = 0.381527$$

Analogiczne obliczenia można przeprowadzić dla pozostałych wartości własnych. Odpowiadające im wektory własne wynoszą

$$v_2 = (x_2, y_2, z_2) \quad x_2 = -0.878231 \quad y_2 = -0.458209 \quad z_2 = 0.136948$$
$$v_3 = (x_3, y_3, z_3) \quad x_3 = 0.423079 \quad y_3 = 0.756967 \quad z_3 = 0.498000$$

W ogólności dla dowolnej macierzy może okazać się, iż dana macierz nie posiada wartości własnych rzeczywistych lub posiada wartości własne wielokrotne. W drugim przypadku nie istnieje jeden unormowany wektor własny, ale cały ich zbiór leżący na konkretnej płaszczyźnie.
Bardzo często występującymi macierzami w naukach technicznych są macierze symetryczne, np. w mechanice ciała odkształcalnego takimi macierzami są macierz naprężeń i macierz odkształceń dla materiału izotropowego. Można wykazać następujące twierdzenie:

Twierdzenie 1
Każda macierz symetryczna dodatnio określona posiada wszystkie wartości własne rzeczywiste dodatnie i różne od siebie.

Przykład 3
Macierz naprężeń dla płaskiego stanu naprężenia opisana jest w każdym punkcie ciała

$$\boldsymbol{A} = \begin{bmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{bmatrix}$$

Znaleźć postać macierzy w układzie własnym oraz jej kierunki główne.

Układamy równanie charakterystyczne (tu również nazywane *równaniem wiekowym* lub *sekularnym*)

$$(\boldsymbol{A}-\lambda\boldsymbol{I})=\begin{vmatrix}3-\lambda & \sqrt{2}\\ \sqrt{2} & 2-\lambda\end{vmatrix}=0 \quad\Rightarrow\quad \lambda^2-5\lambda+4=0$$

Wartości własne wynoszą $\lambda_1=1,\quad \lambda_2=4$.
Wektory własne:

- dla $\lambda_1=1$

$$\begin{bmatrix}3-1 & \sqrt{2}\\ \sqrt{2} & 2-1\end{bmatrix}\begin{bmatrix}x_1\\ y_1\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix} \quad\Rightarrow\quad \begin{cases}2x_1+\sqrt{2}y_1=0\\ x_1^2+y_1^2=1\end{cases}$$

  stąd $x_1=0.816497,\quad y_1=0.577350$.

- dla $\lambda_2=4$

$$\begin{bmatrix}3-4 & \sqrt{2}\\ \sqrt{2} & 2-4\end{bmatrix}\begin{bmatrix}x_2\\ y_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix} \quad\Rightarrow\quad \begin{cases}-x_2+\sqrt{2}y_2=0\\ x_2^2+y_2^2=1\end{cases}$$

  stąd $x_2=-0.577350,\quad y_2=0.816497,.$

Dla wektorów własnych (tu: wersorów wyznaczających osie główne) macierzy symetrycznych istnieje warunek ich wzajemnej ortogonalności

$$\begin{bmatrix}x_1\\ y_1\end{bmatrix}^T\cdot\begin{bmatrix}x_2\\ y_2\end{bmatrix}=\begin{bmatrix}0.816497\\ 0.577350\end{bmatrix}^T\cdot\begin{bmatrix}-0.577350\\ 0.816497\end{bmatrix}=0$$

co sprowadza się do obliczenia iloczynu skalarnego wektorów (dla wektorów prostopadłych iloczyn skalarny jest równy zeru).

W analitycznej i numerycznej analizie problemów własnych macierzy pomocnicze są następujące twierdzenia:

Twierdzenie 2
Jeżeli macierz posiada różne wartości własne to istnieje zbiór liniowo niezależnych wektorów własnych, z dokładnością do stałej, co oznacza istnienie jednoznacznych kierunków tych wektorów.

Twierdzenie 3 *(Cayley – Hamiltona)*
Macierz symetryczna drugiej walencji ($A=\left[a_{ij}\right]$) spełnia swoje własne równanie charakterystyczne.

$$\boldsymbol{A}^3 - I_1^A \boldsymbol{A}^2 + I_2^A \boldsymbol{A} - I_3^A \boldsymbol{I},$$

gdzie I_1^A, I_2^A, I_3^A *są jej niezmiennikami.*

Twierdzenie 4
Jeżeli $g(x)$ *jest wielomianem, a* λ *jest wartością własną macierzy* $\boldsymbol{A}$*, to* $g(\lambda)$ *jest wartością własną macierzy* $g(\boldsymbol{A})$.

Przykład 4
Wartości własne macierzy $\boldsymbol{A}$ wynoszą $\lambda_i = \{-2, 0, 1, 3\}$. Obliczyć wartości własne macierzy $\boldsymbol{B} = \boldsymbol{A}^3 - 2\boldsymbol{A}^2 + \boldsymbol{A} - 10\boldsymbol{I}$

Konsekwencją twierdzenia 4 jest przeniesienie zależności miedzy macierzami $\boldsymbol{A}$ i $\boldsymbol{B}$ na zależność między ich wartościami własnymi, czyli:

$$\lambda_B = \lambda_A^3 - 2\lambda_A^2 + \lambda_A - 10$$

co pozwala bardzo łatwo obliczyć wartości własne macierzy $\boldsymbol{B}$

$$\lambda_1 = (-2)^3 - 2(-2)^2 + (-2) - 10 = -28$$
$$\lambda_2 = 0^3 - 2 \cdot 0^2 + 0 - 10 = -10$$
$$\lambda_3 = 1^3 - 2 \cdot 1^2 + 1 - 10 = -10$$
$$\lambda_4 = 3^3 - 2 \cdot 3^2 + 3 - 10 = 2$$

Twierdzenie 5
Transformacja macierzy $\boldsymbol{A}$ *przez podobieństwo nie zmienia jej wartości własnych.*
Jeżeli $\boldsymbol{R}$ *jest macierzą nieosobliwą to transformacją przez podobieństwo nazywamy przekształcenie* $\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R}$. *Wartości własne tej nowej macierzy są takie same jak wartości własne macierzy wyjściowej* $\boldsymbol{A}$.

Twierdzenie 6
Transformacja ortogonalna macierzy $\boldsymbol{A}$ *nie zmienia ani jej wartości własnych ani jej ewentualnej symetrii.*
Jeżeli $\boldsymbol{Q}$ *jest macierzą nieosobliwą i taką, że* $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ *to transformacją ortogonalna nazywamy przekształcenie* $\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}$. *Wartości własne tej nowej macierzy są takie same jak wartości własne macierzy wyjściowej* $\boldsymbol{A}$.

Twierdzenie 7 (*Gerszgorina*)
Niech $\boldsymbol{A}$ *będzie macierzą kwadratową o wymiarze n i wyrazach* a_{ij} $(i, j = 1, 2, ..., n)$. *Jeżeli określimy dyski* $u_i, i = 1, 2, ..., n$ *o środkach odpowiadającym wyrazom* a_{ii} *na przekątnej*

głównej macierzy i promieniach R_i, *gdzie* $R_i = \sum_{\substack{k=1 \\ k \neq i}}^{n} |a_{ik}|$ *to widmo macierzy* $\boldsymbol{A}$ *(zbiór wartości własnych) można oszacować poprzez wzory:*

$$\lambda \in \langle \lambda_{\min}, \lambda_{\max} \rangle$$
$$\lambda_{\min} > \min_i (a_{ii} - R_i)$$
$$\lambda_{\max} < \max_i (a_{ii} + R_i)$$

Oszacowania powyższe stają się rzeczywistymi wartościami $\lambda_{\min}$ i $\lambda_{\max}$ dla macierzy ściśle dominującej na przekątnej głównej.
Macierz nazywamy *macierzą ściśle dominującą na przekątnej głównej*, jeżeli:
$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$, $i = 1, 2, ..., n$.

Przykład 5
Oszacować widmo wartości własnych korzystając z *twierdzenia Gerszgorina* dla macierzy:

$$\boldsymbol{A} = \begin{bmatrix} -2 & 1 & 3 \\ -1 & 4 & 2 \\ 3 & -2 & 3 \end{bmatrix}$$

Wyrazy na przekątnej głównej: $a_{11} = -2, \quad a_{22} = 4, \quad a_{33} = 3$.
Promienie dysków: $R_1 = 1 + 3 = 4, \quad R_2 = |-1| + 2 = 3, \quad R_3 = 3 + |-2| = 5$.
Oszacowanie wartości własnych:

$$\lambda_{\min} > \min \begin{bmatrix} -2-4 \\ 4-3 \\ 3-5 \end{bmatrix} = \min \begin{bmatrix} -6 \\ 1 \\ -2 \end{bmatrix} = -6, \quad \lambda_{\min} > -6$$

$$\lambda_{\max} < \max \begin{bmatrix} -2+4 \\ 4+3 \\ 3+5 \end{bmatrix} = \max \begin{bmatrix} 2 \\ 7 \\ 8 \end{bmatrix} = 8, \quad \lambda_{\max} < 8$$

czyli $\lambda \in \langle -6, 8 \rangle$.
W rzeczywistości macierz $\boldsymbol{A}$ ma jedną wartość własną rzeczywistą równą: -2.980286 mieszczącą się w powyższym przedziale.

Jednym z zastosowań powyższego twierdzenia jest jego wykorzystanie do zbadania dodatniej określoności danej macierzy kwadratowej $\boldsymbol{A}$.
Macierz $\boldsymbol{A}$ o wymiarze n nazywamy *macierzą dodatnio określoną*, jeśli jest nieosobliwa ($\det(\boldsymbol{A}) \neq 0$) oraz dla dowolnego wektora $x \in \Re^n$ spełniona jest nierówność $x^T \boldsymbol{A}\, x > 0$.

Ponieważ badanie dodatniej określoności macierzy z definicji jest kłopotliwe, stosuje się to tego różne twierdzenia, oprócz *twierdzenia 1-szego* także:

Twierdzenie 8
*Jeżeli macierz kwadratowa **A** o wyrazach rzeczywistych jest ściśle dominująca na przekątnej głównej i ma dodatnie wyrazy na przekątnej głównej to **A** jest dodatnio określona.*

Często również wykorzystuje się do badania dodatniej określoności macierzy pojęcie podwyznacznika macierzy: jeśli znaki podwyznaczników macierzy (od rzędu *1-szego* aż do rzędu *n-tego*) tworzą naprzemienny ciąg lub są takie same to macierz jest dodatnio określona.

Według *twierdzenia 1-szego*, aby wykazać, że macierz jest dodatnio określona, należy udowodnić, iż jej wartości własne są dodatnie i różne od siebie. Ponieważ *twierdzenie Gerszgorina* oszacowuje widmo macierzy, można go zastosować w celu zbadania pierwszej tezy. Natomiast zbadanie, czy wartości własne są od siebie różne, wymaga zastosowania tzw. *ciągów Sturma* i nie będzie rozważane w tym opracowaniu.

Przykład 6
Wykorzystać *twierdzenie Gerszgorina* do zbadania dodatniej określoności następujących macierzy:

$$\boldsymbol{A}=\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{bmatrix} \qquad \boldsymbol{B}=\begin{bmatrix}3 & -2 & 1\\ -2 & 3 & 2\\ 1 & 2 & 3\end{bmatrix}$$

Dla macierzy ***A***:

$$\lambda_{\min} > \min\begin{bmatrix}2-2\\ 2-2\\ 2-2\end{bmatrix} = \min\begin{bmatrix}0\\ 0\\ 0\end{bmatrix} = 0, \quad \lambda_{\min} > 0$$

$$\lambda_{\max} < \max\begin{bmatrix}2+2\\ 2+2\\ 2+2\end{bmatrix} = \max\begin{bmatrix}4\\ 4\\ 4\end{bmatrix} = 4, \quad \lambda_{\max} < 4$$

$$\Rightarrow \quad \lambda \in (0,4)$$

Wniosek: macierz ***A*** może być dodatnio określona.

Dla macierzy ***B***:

$$\lambda_{\min} > \min\begin{bmatrix}3-3\\ 3-4\\ 3-4\end{bmatrix} = \min\begin{bmatrix}0\\ -1\\ -1\end{bmatrix} = -1, \quad \lambda_{\min} > -1$$

$$\lambda_{\max} < \max\begin{bmatrix}3+3\\ 3+4\\ 3+4\end{bmatrix} = \max\begin{bmatrix}6\\ 7\\ 7\end{bmatrix} = 7, \quad \lambda_{\max} < 7$$

$$\Rightarrow \quad \lambda \in (-1,7)$$

Wniosek: macierz ***B*** nie jest dodatnio określona.

Metody numeryczne do znajdywania wartości i wektorów własnych można podzielić na :

- metody obliczania wszystkich wartości i wektorów własnych (np. *metoda Jacobiego*),
- metody obliczania wartości własnych i odpowiadających im wektorów własnych w z góry określonych pasmach widma wartości własnych,
- metody obliczania pojedynczej wartości własnej i odpowiadającego jej wektora własnego.

W opracowaniu zostaną przedstawione jedynie metody z ostatniej grupy. Większość z nich to metody iteracyjne.

Metoda potęgowa

Jedną z najprostszych metod jednoczesnego obliczania wartości własnych oraz wektorów własnych macierzy $\boldsymbol{A}$ jest następująca metoda iteracyjna.
Przypuśćmy, że wartości własne $\lambda_1, \lambda_2, ..., \lambda_n$ są rzeczywiste i spełniają nierówności $|\lambda_1| > |\lambda_2| > ... > |\lambda_n|$. Wybiera się dowolny wektor y_0, a następnie za pomocą wzoru iteracyjnego $y_{n+1} = A\, y_n$ buduje się ciąg wektorów $y_1, y_2, ...$ Okazuje się, że dla dostatecznie dużych *n*, wektor y_n jest bliski wektorowi własnemu macierzy $\boldsymbol{A}$, odpowiadającemu największej, co do modułu wartości własnej. Wartość własną otrzymamy dzieląc dowolną współrzędną wektora y_{n+1} przez tą samą współrzędną wektora y_n .

Przykład 7
Niech macierz $\boldsymbol{A}$ będzie postaci:

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

Przyjmijmy wektor początkowy $y_0 = (1, -1, 1, -1)$. Kolejne iteracje dają następujący ciąg wektorów:

$$y_1 = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \\ 4 \\ -3 \end{bmatrix} \qquad y_2 = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -4 \\ 4 \\ -3 \end{bmatrix} = \begin{bmatrix} 10 \\ -15 \\ 15 \\ -10 \end{bmatrix} \quad itd.$$

$y_3 = (35, -55, 55, -35)$ $y_4 = (125, -200, 200, -125)$ $y_5 = (450, -725, 725, -405)$
$y_6 = (1625, -2625, 2625, -1625)$ $y_7 = (5875, -9500, 9500, -5875)$
$y_8 = (21250, -34375, 34375, -21250)$
Stosunki odpowiednich współrzędnych wektorów y_8 i y_7 są równe:

$$\frac{5875}{21250}=3.61702,\quad \frac{-9500}{-34375}=3.61842,\quad \frac{9500}{34375}=3.61842,\quad \frac{-5875}{-21250}=3.61702$$

Widzimy, że wszystkie cztery liczby są dość bliskie sobie, stąd wnioskujemy, że każda z nich jest bliska największej, co do modułu wartości własnej macierzy $\boldsymbol{A}$.
Bardziej dokładną wartość własną otrzymamy, jeżeli podzielimy skalarny kwadrat wektora y_8 przez iloczyn skalarny $y_7 \cdot y_8$. Otrzymamy wówczas

$$\lambda_1^* = \frac{y_8 \cdot y_8}{y_7 \cdot y_8} = \frac{21250\cdot 21250+(-34375)\cdot(-34375)+34375\cdot 34375+(-21250)\cdot(-21250)}{5875\cdot 21250+(-9500)\cdot(-34375)+9500\cdot 34375+(-5875)\cdot(-21250)} =$$
$$= 3.61804$$

Odpowiedni wektor własny jest równy $x_1^* = (0.61818, -1, 1, -0.61818)$. Wektor x_1^* jest tak znormalizowany, że jego największa współrzędna jest równa jedności. Gdyby przyjąć kryterium jednostkowej długości wektora własnego, to wynosiłby on wtedy:
$x_1^* = (0.37118, -0.60146, 0.60146, -0.37181)$.
Dokładna wartość największej, co do modułu wartości własnej jest równa $\lambda_1 = 3.618034$.

Metoda Rayleigha

Jest to najpopularniejsza metoda wśród metod iteracyjnych znajdywania wartości własnej macierzy $\boldsymbol{A}$ o wymiarze *n*, największej, co do modułu. Wywodzi się ona z omawianej wyżej *metody potęgowej*, wykorzystuje m.in. własności *twierdzenia 6*, oraz wyrażenie postaci:

$\boldsymbol{A}x = \lambda x \qquad \lambda = \dfrac{x^T \boldsymbol{A} x}{x^T x}$, zwane w literaturze *ilorazem Rayleigha*.

Algorytm metody wygląda następująco:
Poszukiwana jest wartość własna λ największa, co do modułu oraz odpowiadający jej wektor własny *x* (lub unormowany *v*): $\boldsymbol{A}x = \lambda x$
Przyjmujemy na starcie wektor x_0. Przypisujemy $x_{k=0} = x_0$, gdzie *k* oznacza *k-tą* iterację.

- Normalizujemy wektor x_k (dzielimy go przez jego długość):

  $$v_k = \frac{x_k}{\|x_k\|} = \frac{x_k}{\sqrt{x_k^T x_k}}$$

- Obliczamy kolejne przybliżenie wektora własnego $x_{k+1} = \boldsymbol{A} \cdot v_k$.
- Obliczamy *iloraz Rayleigha*:

  $\lambda_{k+1} = \dfrac{v_k^T \boldsymbol{A} v_k}{v_k^T v_k} = v_k^T x_{k+1}$ (jest to po prostu iloczyn skalarny dwóch wektorów będący liczbą – kolejnym przybliżeniem wartości własnej λ_1.

- Obliczamy poziom błędów (począwszy od drugiej iteracji – dla *k = 1*):

  $$\varepsilon_{k+1}^{\lambda} = \left|\frac{\lambda_{k+1}-\lambda_k}{\lambda_{k+1}}\right|$$, norma błędu przy obliczaniu wartości własnej

  $\varepsilon_{k+1}^{v} = \|v_{k+1} - v_k\|$, norma błędu przy obliczaniu wektora własnego.

- Sprawdzamy kryterium przerwania iteracji:

$\varepsilon_{k+1}^{\lambda} \le B_1, \quad \varepsilon_{k+1}^{v} \le B_2$, gdzie B_1, B_2 - zadane poziomy dokładności obydwu wielkości.

Jeżeli powyższe kryterium jest spełnione to: $\lambda_1 \approx \lambda_{k+1}, \quad v_1 \approx v_{k+1}$

Przykład 8
Dana jest macierz:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Znaleźć jej wartość własną największą, co do modułu i odpowiadający jej wektor własny korzystając z *metody Rayleigha.*

Wartości własne macierzy wynoszą: $\lambda_1 = 4, \quad \lambda_2 = \lambda_3 = 1$
Przyjmujemy wektor startowy $x_0 = (1,0,0)$

Pierwsza iteracja $k = 0$:

$$\|x_0\| = 1 \quad \Rightarrow \quad v_0 = \frac{x_0}{1} = (1,0,0)$$

$$x_1 = \boldsymbol{A}v_0 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

$$\lambda_1 = v_0^T x_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = 2$$

Druga iteracja $k = 1$:

$$\|x_1\| = 2.499490 \quad \Rightarrow \quad v_1 = \frac{x_1}{2.499490} = (0.816497, 0.408248, 0.408248)$$

$$x_2 = \boldsymbol{A}v_1 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.816497 \\ 0.408248 \\ 0.408248 \end{bmatrix} = \begin{bmatrix} 2.449490 \\ 2.041241 \\ 2.041241 \end{bmatrix}$$

$$\lambda_2 = v_1^T x_2 = \begin{bmatrix} 0.816497, 0.408248, 0.408248 \end{bmatrix} \begin{bmatrix} 2.449490 \\ 2.041241 \\ 2.041241 \end{bmatrix} = 3.666667$$

$$\|x_2\| = 3.785939 \quad \Rightarrow \quad v_2 = \frac{x_2}{3.785939} = (0.649997 \quad 0.539164 \quad 0.539164)$$

$$\varepsilon_2^{\lambda} = \left| \frac{\lambda_2 - \lambda_1}{\lambda_2} \right| = \left| \frac{3.666667 - 2}{2} \right| = 0.454545,$$

$$\varepsilon_2^{v} = \|v_2 - v_1\| = \left\| \begin{bmatrix} 0.649997 \\ 0.539164 \\ 0.539164 \end{bmatrix} - \begin{bmatrix} 0.816497 \\ 0.408248 \\ 0.408248 \end{bmatrix} \right\| = 0.251014$$

Trzecia iteracja $k = 2$:

$$x_3 = \boldsymbol{A}v_2 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.649997 \\ 0.539164 \\ 0.539164 \end{bmatrix} = \begin{bmatrix} 2.372321 \\ 2.264488 \\ 2.264488 \end{bmatrix}$$

$$\lambda_3 = v_2^T x_3 = \begin{bmatrix} 0.649997, 0.539164, 0.539164 \end{bmatrix} \begin{bmatrix} 2.372321 \\ 2.264488 \\ 2.264488 \end{bmatrix} = 3.976744$$

$$\|x_3\| = 3.985439 \quad \Rightarrow \quad v_3 = \frac{x_3}{3.985439} = (0.595247 \quad 0.568190 \quad 0.568190)$$

$$\varepsilon_3^{\lambda} = \left| \frac{\lambda_3 - \lambda_2}{\lambda_3} \right| = \left| \frac{3.976744 - 3.666667}{3.666667} \right| = 0.07797,$$

$$\varepsilon_3^{v} = \|v_3 - v_2\| = \left\| \begin{bmatrix} 0.595247 \\ 0.568190 \\ 0.568190 \end{bmatrix} - \begin{bmatrix} 0.649997 \\ 0.539164 \\ 0.539164 \end{bmatrix} \right\| = 0.066054$$

Już po trzech iteracjach widoczne jest, na jakim poziomie stabilizują się wyniki. Wartość własną z precyzją do sześciu miejsc otrzymano po $k = 7$ iteracjach:

$$\lambda \approx \lambda_7 = 4.0, \quad \varepsilon_7^{\lambda} = 0.000001$$

$$v \approx v_7 = (0.577421, 0.577315, 0.577315), \quad \varepsilon_7^{v} = 0.000259$$

Zaobserwować można szybszą zbieżność samej wartości własnej niż wektora własnego.

Zarówno w przypadku *metody potęgowej* jak i *metody Rayleigha* pozostałe wartości i wektory własne można znaleźć stosując różne modyfikacje tych metod jak np. *metodę iteracji odwrotnej* zbieżną do wartości własnej najbliższej zeru czy *przesunięcie widma macierzy* o zadaną wartość. Stosuje się też zabiegi mające na celu przyśpieszenie zbieżności metod iteracyjnych.

UKŁADY RÓWNAŃ ŹLE UWARUNKOWANYCH

Przy rozwiązywaniu układów równań liniowych postaci $\boldsymbol{A}x = b$ można mieć do czynienia z przypadkiem, gdy

- ***det(A) = 0*** - osobliwość macierzy współczynników powoduje brak rozwiązań przy dowolnym niezerowym wektorze wyrazów wolnych *b* lub tożsamość dla zerowego wektora *b*,
- ***det(A)*** $\neq$ ***0*** - zapewnia istnienie jednoznacznego rozwiązania postaci $x = \boldsymbol{A}^{-1} b$
- ***det(A)*** $\approx$ ***0*** - *układ źle uwarunkowany*. W takiej sytuacji bardzo małe zmiany w wyrazach macierzy współczynników mogą spowodować ogromne zmiany w rozwiązaniu.

W celu zbadania stopnia uwarunkowania układu równań oblicza się tzw. *wskaźnik uwarunkowania k* – liczbę o takiej własności, że

- $k = 1$ - idealne uwarunkowanie,
- $k = \infty$ - układ osobliwy.

Sposoby obliczania *wskaźnika uwarunkowania k* dla macierzy współczynników $\boldsymbol{A}$:

- $k_A = \|\boldsymbol{A}\| \cdot \|\boldsymbol{A}^{-1}\|, \quad \|\boldsymbol{A}\| = \sqrt{\sum_{i=1}^{n} a_{ij}^2}$
- $k_A = \left| \frac{\lambda_{\max}}{\lambda_{\min}} \right|$.

Posługując się ostatnim wzorem można obliczać wartości własne analitycznie (wtedy wzór ma słuszność gł. dla macierzy symetrycznych) lub numerycznie (np. z *twierdzenia Gerszgorina* dla macierzy ściśle dominujących na przekątnej głównej)

Przykład 9

Wykazać, która z macierzy $\boldsymbol{H} = \begin{bmatrix} -1 & -\frac{1}{3} \\ 1 & 1 \end{bmatrix}$ oraz $\boldsymbol{J} = \begin{bmatrix} -1 & -4 \\ -1 & -3 \end{bmatrix}$ jest lepiej uwarunkowana.

Wynik uzasadnić liczbowo.

W zadaniu należy obliczyć osobno wskaźniki uwarunkowania dla każdej z macierzy i sprawdzić, który z nich jest bliższy jedności. Posłużymy się przy obliczaniu wskaźnika kryterium normowym.

Macierze $\boldsymbol{H}^{-1}$ oraz $\boldsymbol{J}^{-1}$ można obliczyć analitycznie (ze *wzoru Gaussa*) lub stosując odpowiedni algorytm numeryczny (*eliminacja Gaussa*, rozkład na czynniki trójkątne, metody iteracyjne). Ponieważ wymiary macierzy są małe, ich odwrotności obliczono analitycznie.

$$\det(\boldsymbol{H}) = -\frac{2}{3} \quad \Rightarrow \quad \boldsymbol{H}^{-1} = -\frac{3}{2}\begin{bmatrix} 1 & \frac{1}{3} \\ -1 & -1 \end{bmatrix}$$

$$\det(\boldsymbol{J}) = -1 \quad \Rightarrow \quad \boldsymbol{J}^{-1} = -\begin{bmatrix} -3 & 4 \\ 1 & -1 \end{bmatrix}$$

Odpowiednie normy średniokwadratowe wynoszą:

$$\|\boldsymbol{H}\| = \sqrt{1+\frac{1}{9}+1+1} = \sqrt{\frac{28}{9}} = \frac{2}{3}\sqrt{7}, \qquad \|\boldsymbol{H}^{-1}\| = \frac{3}{2}\sqrt{1+\frac{1}{9}+1+1} = \frac{3}{2}\sqrt{\frac{28}{9}} = \sqrt{7}$$

$$\|\boldsymbol{J}\| = \sqrt{1+16+1+9} = \sqrt{27} = 3\sqrt{3}, \qquad \|\boldsymbol{J}^{-1}\| = \sqrt{9+16+1+1} = 3\sqrt{3}$$

zaś wskaźniki uwarunkowania :

$$k_H = \|\boldsymbol{H}\| \cdot \|\boldsymbol{H}^{-1}\| = \frac{2}{3}\sqrt{7} \cdot \sqrt{7} = \frac{14}{3} \approx\sim 4.666667$$

$$k_J = \|\boldsymbol{J}\| \cdot \|\boldsymbol{J}^{-1}\| = 3\sqrt{3} \cdot 3\sqrt{3} = 27$$

Ponieważ $|k_H - 1| < |k_J - 1|$ to lepiej uwarunkowana jest macierz $\boldsymbol{H}$.

Kryterium związanego ze ścisłym wyznaczeniem wartości własnych nie można zastosować, gdyż macierz $\boldsymbol{J}$ nie posiada rzeczywistych rozwiązań problemu własnego. Oszacowanie widm macierzy z twierdzenia Gerszgorina daje w rezultacie:

- dla macierzy $\boldsymbol{H}$:

$$\lambda_{\min} \approx \min(\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ 1 \end{bmatrix}) = -\frac{4}{3}, \qquad \lambda_{\max} \approx \max(\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{3} \\ 1 \end{bmatrix}) = 2$$

$$k_H = \left|\frac{2}{\frac{-4}{3}}\right| = \frac{3}{2} = 1.5$$

- dla macierzy $\boldsymbol{J}$:

$$\lambda_{\min} \approx \min(\begin{bmatrix} -1 \\ -3 \end{bmatrix} - \begin{bmatrix} 4 \\ 1 \end{bmatrix}) = -5, \qquad \lambda_{\max} \approx \max(\begin{bmatrix} -1 \\ -3 \end{bmatrix} + \begin{bmatrix} 4 \\ 1 \end{bmatrix}) = 3$$

$$k_J = \left|\frac{3}{-5}\right| = \frac{3}{5} = 0.6$$

Oszacowanie okazało się fałszywe (macierze nie są ściśle dominujące na przekątnej głównej).

Przykład 10

Zbadać uwarunkowanie macierzy

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

Macierz spełnia wymagania kryteria do stosowania wzoru opartego na wartościach własnych. Równanie charakterystyczne wynosi: $\lambda^2 - 4\lambda + 3$ a wartości własne: $\lambda_{\max} = 3, \quad \lambda_{\min} = 1$

Wskaźnik uwarunkowania: $k_A = \left| \frac{\lambda_{\max}}{\lambda_{\min}} \right| = 3$.

Na podstawie wskaźnika można ustalić, z jaką precyzją należy podać elementy macierzy $\boldsymbol{A}$ aby uzyskać żądaną dokładność rozwiązania. Służy do tego wzór :

$q \approx p - \log(k)$,

gdzie : q – liczba cyfr znaczących elementów macierzy, p – dokładność rozwiązania.

Np. dla $p = 6$ mamy $p = q + \log(k) = 6 + \log(3) = 6.47 \approx 7$ miejsc znaczących współczynników macierzy $\boldsymbol{A}$.

Uwarunkowanie macierzy można poprawić stosując większą precyzję obliczeń lub tzw. *metody regularyzacji*.