

Metody numeryczne – procedury

na podstawie [*Marciniak et. al. 1997*] oraz [*Bronsztejn et. al. 2004*]



dr inż. Paweł Zalewski
Akademia Morska w Szczecinie

Interpolacja wielomianowa:

Zadanie interpolacji polega na znalezieniu pewnej funkcji, która „przybliża” daną funkcję f . Dla funkcji f znane są przy tym wartości $f(x_i)$ lub wartości jej pochodnych w pewnej skończonej liczbie różnych punktów x_i , które nazywamy węzłami interpolacji.

Wartość wielomianu interpolacyjnego Lagrange’a L_n w punkcie x dla danej funkcji f oblicza się z wzoru:

$$L_n(x) = \sum_{i=0}^n f(x_i) \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}$$

gdzie wartości x_i oraz $f(x_i)$ ($i = 0, 1, \dots, n$) są dane.

Interpolacja wielomianowa:

Wartość **wielomianu interpolacyjnego** w punkcie x dla danej funkcji f można obliczyć również algorytmem Neville'a na podstawie zależności rekurencyjnych:

$$P_{i,0} = f(x_i), \quad i = 0, 1, \dots, n$$

$$P_{i,j} = \frac{(x - x_{i-j})P_{i,j-1} - (x - x_i)P_{i-1,j-1}}{(x_i - x_{i-j})}, \quad i = 0, 1, \dots, n, \quad j = 0, 1, \dots, i$$

gdzie wartości x_i oraz $f(x_i)$ ($i = 0, 1, \dots, n$) są dane. Wartość elementu $P_{n,n}$ jest poszukiwaną wartością wielomianu w punkcie x .

Aproksymacja wielomianowa:

Przykładem jest **aproksymacja średniokwadratowa wielomianem**, która dla danej funkcji $f(x)$ polega na określeniu wielomianu $p(x)$ stopnia m , który w przypadku danych dyskretnych minimalizuje sumę:

$$\sum_{i=1}^n [f(x_i) - p(x_i)]^2 w(x_i)$$

gdzie wartości x_i oraz $f(x_i)$ ($i = 0, 1, \dots, n$) są dane, $m < n$, a $w(x)$ oznacza funkcję wagową. Współczynniki wielomianu aproksymującego funkcję $f(x)$:

$$a_0 + a_1 x + \dots + a_m x^m$$

można wyznaczyć poszukując minimum poniższego funkcjonału przy danych wartościach x_i oraz $f(x_i)$:

$$H(a_0, a_1, \dots, a_m) = \sum_{j=1}^n \left[f(x_j) - \sum_{i=0}^m a_i (x_j^i) \right]^2$$

gdzie $m < n$.

Aproksymacja wielomianowa:

$$\frac{\partial H}{\partial a_k} = \sum_{j=1}^n \left[f(x_j) - \sum_{i=0}^m a_i (x_j^i) \right]^2 x_j^k = 0, \quad k = 0, 1, \dots, m$$

Wprowadzając oznaczenia:

$$y(k) = \sum_{j=1}^n f(x_j) x_j^k, \quad k = 0, 1, \dots, m$$

$$z(i, k) = \sum_{j=1}^n x_j^{i+k}, \quad i, k = 0, 1, \dots, m$$

otrzymujemy układ równań liniowych postaci:

$$\sum_{i=0}^m z(i, k) a_i = y_k, \quad k = 0, 1, \dots, m$$

Powyższy układ równań liniowych można rozwiązać np. metodą eliminacji Gaussa lub metodami iteracyjnymi (przybliżonymi).

Całkowanie numeryczne:

Wiele całek nie można przedstawić za pomocą funkcji elementarnych. Do ich obliczenia stosuje się **metody przybliżone** (przykładowo dla obliczenia pola powierzchni pod krzywą). W celu wyznaczenia przybliżonej wartości całki:

$$\int_a^b F(x)dx, \quad -\infty < a < b < \infty$$

funkcję podcałkową $F(x)$ przedstawia się w postaci iloczynu $w(x)f(x)$, gdzie funkcja $w(x)$, zwana funkcją wagową zazwyczaj wynosi 1, a funkcję $f(x)$ przybliża się wielomianem. Wówczas:

$$\int_a^b F(x)dx = \int_a^b w(x)f(x)dx \approx \sum_{k=0}^n A_k f(x_k)$$

gdzie współczynniki A_k zależą od zastosowanej metody numerycznej obliczeń.

Całkowanie numeryczne:

W metodzie Simpsona dany przedział $[a, b]$ dzieli się na równe podprzedziały. W przypadku trzech podprzedziałów:

$$[a, x_1] = [x_0, x_1], \quad [x_1, x_2], \quad [x_2, x_3] = [x_2, b]$$

w każdym podprzedziale stosuje się wzór Simpsona:

$$\int_{x_i}^{x_{i+1}} f(x) dx \approx \frac{h}{3} [f(x_i) + 4f(x_i + h) + f(x_{i+1})] = I_i$$

$$h = \frac{x_{i+1} - x_i}{2}, \quad i = 0, 1, 2$$

Następnie każdy z przedziałów dzieli się znowu na trzy części i otrzymuje się drugie przybliżenie całki. Procedura ta jest kontynuowana do chwili, gdy różnica pomiędzy dwoma kolejnymi przybliżeniami całki jest dostatecznie mała lub, gdy długości podprzedziałów są mniejsze od:

$$\frac{b-a}{3^n}, \quad n \in N$$

Układy równań liniowych:

Metody dokładne rozwiązywania układu równań liniowych:

$$Ax = b,$$

gdzie macierz kwadratowa A stopnia n i wektor $b \in \mathbb{R}^n$ są dane, teoretycznie pozwalają uzyskać rozwiązanie dokładne:

$$x = A^{-1}b,$$

gdzie $x \in \mathbb{R}^n$, o ile macierz kwadratowa A jest nieosobliwa. Ponieważ wartości elementów macierzy A i wektora b nie są reprezentowane dokładnie (rozdzielczość typów liczbowych w operacjach zmiennoprzecinkowych w danym środowisku programowym), i podczas obliczeń występują błędy zaokrągleń, więc rozwiązanie obliczone metodą „dokładną” może różnić się od teoretycznego rozwiązania dokładnego.

Układy równań liniowych:

W przypadku układu równań liniowych z macierzą trójkątną górną:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1,$$

$$a_{22}x_2 + \dots + a_{2n}x_n = b_2,$$

...

$$a_{nn}x_n = b_n$$

jeżeli macierz układu nie jest osobliwa, tj. gdy $a_{ii} \neq 0$ dla każdego $i = 1, 2, \dots, n$ to niewiadome x_i są obliczane na podstawie zależności:

$$x_i = \frac{b_i - \sum_{k=i+1}^n a_{ik}x_k}{a_{ii}}, \quad i = n, n-1, \dots, 1$$

Układy równań liniowych:

W przypadku układu równań liniowych z macierzą trójkątną dolną:

$$a_{11}x_1 = b_1,$$

$$a_{21}x_1 + a_{22}x_2 = b_2,$$

...

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n$$

jeżeli macierz układu nie jest osobliwa, tj. gdy $a_{ii} \neq 0$ dla każdego $i = 1, 2, \dots, n$ to niewiadome x_i są obliczane na podstawie zależności:

$$x_i = \frac{b_i - \sum_{k=1}^{i-1} a_{ik}x_k}{a_{ii}}, \quad i = 1, 2, \dots, n$$

Układy równań liniowych:

Rozwiązywanie układu równań liniowych metodą eliminacji Gaussa z częściowym wyborem elementu podstawowego:

Macierz A jest rozkładana na iloczyn macierzy LU , gdzie L oznacza macierz trójkątną dolną z jedynkami na głównej przekątnej, a U macierz trójkątną górną. Krok k ($k = 1, 2, \dots, n-1$) operacji rozkładu macierzy A można opisać następująco:

Znajdujemy element $a_{i_k k}^{(k)}$ taki, że:

$$\left| a_{i_k k}^{(k)} \right| = \max_{k \leq i \leq n} \left| a_{ik}^{(k)} \right|$$

Następnie w macierzy $A^{(k)}$ zamieniamy wiersze o numerach i_k oraz k , a w wektorze $b^{(k)}$ składowe o takich samych numerach. Odpowiada to lewostronnemu przemnożeniu macierzy $A^{(k)}$ i wektora $b^{(k)}$ przez macierz permutacji $P^{(k)}$. W otrzymanym układzie równań:

$$\bar{A}^{(k)} x = \bar{b}^{(k)}$$

Układy równań liniowych:

Rozwiązywanie układu równań liniowych metodą eliminacji Gaussa z częściowym wyborem elementu podstawowego:

eliminujemy niewiadomą x_k z wierszy o numerach $k + 1, k + 2, \dots, n$ korzystając z wzorów:

$$a_{ij}^{(k+1)} = \bar{a}_{ij}^{(k)} - l_{ik} \bar{a}_{kj}^{(k)},$$

$$b_i^{(k+1)} = \bar{b}_i^{(k)} - l_{ik} \bar{b}_k^{(k)},$$

$$i = k + 1, k + 2, \dots, n, \quad j = k + 1, k + 2, \dots, n,$$

gdzie:

$$l_{ik} = \frac{\bar{a}_{ik}^{(k)}}{\bar{a}_{kk}^{(k)}}$$

Układy równań liniowych:

Rozwiązanie układu równań liniowych metodą eliminacji Gaussa z częściowym wyborem elementu podstawowego:

po $n-1$ krokach otrzymamy układ równań postaci:

$$Ux = L^{-1}Pb,$$

w którym $P = P^{(n-1)}P^{(n-2)} \dots P^1$, $U = A^{(n)}$ oraz:

$$L = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & l_{n3} & \dots & 1 \end{bmatrix}$$

Powyższy układ równań jest rozwiązywany za pomocą zależności:

$$x_i = \frac{\hat{b}_i - \sum_{j=i+1}^n u_{ij}x_j}{u_{ii}}, \quad i = n, n-1, \dots, 1; \quad \hat{b} = L^{-1}Pb$$

Układy równań liniowych:

Wyznaczenie macierzy odwrotnej do macierzy rzeczywistej:

W celu wyznaczenia macierzy odwrotnej do macierzy rzeczywistej A :

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix},$$

macierz A jest najpierw rozkładana na iloczyn LU metodą eliminacji Gaussa. Po wykonaniu tej operacji otrzymujemy:

$$PA = LU$$

$$A^{-1} = U^{-1}L^{-1}P$$

gdzie P oznacza macierz permutacji.

Układy równań liniowych:

Wyznaczenie macierzy odwrotnej do macierzy rzeczywistej:

Elementy macierzy L^{-1} oraz elementy macierzy U^{-1} są określone wzorami:

$$\bar{l}_{ij} = \frac{\delta_{ij} - \sum_{k=j}^{i-1} l_{ik} \bar{l}_{kj}}{l_{ij}}, \quad j=1, 2, \dots, n, \quad i=j, j+1, \dots, n,$$

$$\bar{u}_{ij} = \frac{\delta_{ij} - \sum_{k=j}^{i-1} u_{ik} \bar{u}_{kj}}{u_{ij}}, \quad j=1, 2, \dots, n, \quad i=j, j+1, \dots, n,$$

gdzie:

$$\delta_{ij} = \begin{cases} 0, & \text{gdy } i \neq j, \\ 1, & \text{gdy } i = j. \end{cases}$$

Układy równań liniowych:

Metody iteracyjne rozwiązywania układu równań liniowych:

$$Ax = b \quad (1),$$

gdzie macierz kwadratowa A stopnia n i wektor $b \in \mathbb{R}^n$ są dane, generują ciąg kolejnych przybliżeń wektora x .

Rozpoczęcie procesu iteracyjnego wymaga podania początkowych przybliżeń wszystkich składowych tego wektora.

Jedną z metod iteracyjnych rozwiązywania układu równań liniowych z nieosobliwą macierzą kwadratową A jest metoda Jacobiego. W metodzie tej macierz A jest przekształcana na sumę trzech macierzy:

$$A = L + D + U,$$

gdzie L oznacza macierz trójkątną dolną, D – macierz diagonalną, a U macierz trójkątną górną.

Układy równań liniowych:

Rozwiązywanie układu równań liniowych metodą iteracyjną (przybliżoną) Jacobiego:

Uwzględniając rozkład macierzy A , układ równań (1) można zapisać w postaci:

$$\begin{aligned}(L + D + U)x &= b, \\ Dx &= -(L + U)x + b,\end{aligned}$$

z czego wynika następujący proces iteracyjny:

$$\begin{aligned}Dx^{k+1} &= -(L + U)x^k + b, \\ x^{k+1} &= -D^{-1}(L + U)x^k + D^{-1}b\end{aligned}\quad (2)$$

jeżeli promień spektralny macierzy $-D^{-1}(L+U)$ jest mniejszy od 1, to powyższy proces iteracyjny jest zbieżny.

Układy równań liniowych:

Rozwiązywanie układu równań liniowych metodą iteracyjną (przybliżoną) Jacobiego:

Z zależności (2) wynika, że $(k+1)$ -sze przybliżenie i -tej składowej rozwiązania jest określone wzorem:

$$x_i^{k+1} = \frac{- \sum_{j=1, j \neq i}^n a_{ij} x_j^k + b_i}{a_{ii}}, \quad i = 1, 2, \dots, n \quad (3),$$

przy czym $a_{ii} \neq 0$. Proces iteracyjny kończy się, gdy:

$$\frac{\|x^{k+1} - x^k\|}{\max(\|x^{k+1}\|, \|x^k\|)} \leq \varepsilon, \quad x^{k+1} \neq 0 \quad \text{lub} \quad x^k \neq 0,$$

gdzie:

$$\|x\| = \max_{1 \leq i \leq n} |x_i|,$$

a ε oznacza zadaną dokładność, lub gdy $x^{k+1} = x^k = 0$ lub też, gdy liczba iteracji w procesie (3) jest większa od przyjętej wartości maksymalnej.

Równania i układy równań nieliniowych:

Równanie skalarne z jedną niewiadomą może być zapisane w postaci:

$$f(x) = 0,$$

gdzie $f: X \rightarrow Y$, przy czym $X, Y \in R$. Jeśli równanie to nie może być rozwiązane analitycznie, to do znalezienia jego pierwiastków stosujemy zwykle metodę iteracyjną, za pomocą której można otrzymać przybliżoną wartość pierwiastków, w przypadku, gdy rozwiązanie równania istnieje.

Stosując metodę *regula falsi* znajduje się wartość pierwiastka równania skalarnego z jedną niewiadomą leżącego w przedziale $[a, b]$ takim, że:

$$f(a)f(b) \leq 0$$

Równania i układy równań nieliniowych:

Jeden krok metody *regula falsi* polega na zastąpieniu bieżącego przedziału $[a, b]$ przedziałem mniejszym, którym jest jeden z przedziałów $[a, z]$ lub $[z, b]$, gdzie:

$$z = b - f(b) \frac{b - a}{f(b) - f(a)}$$

Wybór odpowiedniego przedziału zależy od spełnienia jednego z warunków $f(a)f(z) \leq 0$ lub $f(z)f(b) \leq 0$.

Obliczenia kontynuuje się dopóty, dopóki $a < z < b$.

Równania i układy równań nieliniowych:

Wartość pierwiastka równania skalarne z jedną niewiadomą leżącego w przedziale $[a, b]$ można również znaleźć za pomocą metody siecznych (interpolacji liniowej), która polega na zastosowaniu zależności:

$$x_k = x_{k-1} - f(x_{k-1}) \frac{x_{k-1} - x_{k-2}}{f(x_{k-1}) - f(x_{k-2})}, \quad k = 2, 3, \dots$$

Za początkowe przybliżenia x_1 i x_2 w procesie iteracyjnym przyjąć można np. wartości (zależnie od rozdzielczości podziału przedziału $[a, b]$):

$$a + 0,1(b - a) \quad \text{ i } \quad b - 0,1(b - a)$$

uporządkowane w taki sposób, że:

$$|f(x_1)| \geq |f(x_2)|$$

Różniczkowanie numeryczne:

Przybliżone wartości **pochodnej funkcji** $f(x)$ mogą być wyznaczone z wielomianu aproksymującego $p(x)$. Ponieważ różnica $f(x) - p(x)$ może być bardzo mała, a różnica $f'(x) - p'(x)$ bardzo duża, podejście takie może spowodować powstanie zbyt dużego błędu. Gdy jest wymagana duża dokładność obliczania pochodnej zaleca się stosować **metodę Romberga**. W metodzie tej dla danego początkowego kroku h_0 wyznacza się wartości:

$$T_{i0} = \frac{f(x+h_i) - f(x-h_i)}{2h_i},$$

gdzie:

$$h_i = \frac{h_0}{2^i}, \quad i = 0, 1, \dots$$

następnie korzystając z wzoru:

$$T_{i,k} = T_{i,k-1} \frac{T_{i,k-1} - T_{i-1,k-1}}{4^k - 1}, \quad k = 1, 2, \dots, i$$

oblicza się kolejne elementy następującej tablicy:

Różniczkowanie numeryczne:

$$T_{00}$$

$$T_{10} \rightarrow T_{11}$$

$$T_{20} \rightarrow T_{21} \rightarrow T_{22}$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$T_{i,k} \rightarrow f'(x), \text{ gdy } i \rightarrow \infty \text{ dla danego } k$$

Obliczenia kończą się, gdy:

$$|T_{i,k} - T_{i-1,k-1}| < \varepsilon$$

gdzie ε oznacza zadaną z góry dokładność.

Jeżeli dla założonego i oraz k powyższy warunek nie będzie spełniony to dany krok h_0 dzieli się na połowę i cały proces obliczeń powtarza się. Gdy po $m \in \mathbb{N}$ iteracjach zadana dokładność ε nie jest osiągnięta to za rozwiązanie przyjmuje się ostatnio obliczone przybliżenie $T_{i,k}$.

Różniczkowanie numeryczne:

Rozwiązanie zagadnienia początkowego dla pojedynczego równania różniczkowego to inaczej obliczenie przybliżonej wartości $y(x_1)$ rozwiązania skalarne równania:

$$y' = \frac{dy}{dx} = f(x, y)$$

z warunkiem początkowym:

$$y(x_0) = y_0$$

Jedną z metod numerycznych prowadzących do rozwiązania tego zagadnienia jest **metoda Rungego-Kutty** z automatycznym doбором kroku całkowania.

Różniczkowanie numeryczne:

Wartość $y(x_1)$ oblicza się stosując zależność:

$$y(x+h) = y(x) + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4)$$

gdzie:

$$K_1 = f(x, y),$$

$$K_2 = f\left(x + \frac{h}{2}, y + \frac{h}{2}K_1\right),$$

$$K_3 = f\left(x + \frac{h}{2}, y + \frac{h}{2}K_2\right),$$

$$K_4 = f(x+h, y+hK_3)$$

Różniczkowanie numeryczne:

Na podstawie przedstawionej zależności wyznacza się dwie wartości przybliżonego rozwiązania w punkcie $x+h$, gdzie h oznacza krok całkowania, a mianowicie: $y(x+h, h)$ przy użyciu kroku h i $y(x+h, h/2)$ stosując dwukrotnie krok $h/2$. Jeżeli:

$$\frac{\left| y(x+h, h) - y\left(x+h, \frac{h}{2}\right) \right|}{\max \left\{ \left| y(x+h, h) \right|, \left| y\left(x+h, \frac{h}{2}\right) \right|, \eta \right\}} \leq \varepsilon \quad (4)$$

gdzie $\varepsilon, \eta > 0$ oznaczają wartości charakteryzujące dokładność rozwiązania, to końcową przybliżoną wartość rozwiązania w punkcie $x+h$ oblicza się z wzoru:

$$y(x+h) = y\left(x+h, \frac{h}{2}\right) + \frac{y\left(x+h, \frac{h}{2}\right) - y(x+h, h)}{15}$$

Różniczkowanie numeryczne:

W przypadku, gdy nierówność (4) nie zostanie spełniona, to zmienia się krok całkowania h .

Ponieważ całkowanie od punktu x_0 do punktu x_1 wykonuje się w pewnej liczbie kroków, w praktyce błąd względny otrzymanego przybliżonego rozwiązania $y(x_1)$ może być większy od ε .