

# Metody wariacyjne

J.JANUS, M.LUŚTYK

Niech  $V$  - będzie przestrzenią Hilberta z iloczynem skalarnym  $\langle \cdot, \cdot \rangle$  i normą  $\|u\| = \sqrt{\langle u, u \rangle}$ .

## Definicja 1.

Funkcjonał  $L : V \rightarrow \mathbb{R}$  jest liniowy, jeżeli

1.  $L(u + v) = L(u) + L(v), \quad \forall u, v \in V,$
2.  $L(\alpha \cdot v) = \alpha \cdot L(v), \quad \forall \alpha \in \mathbb{R}, \forall v \in V.$

Funkcjonał  $L$  jest ciągły jeżeli  $\exists C > 0, \forall u \in V |L(u)| \leq C\|u\|$ .

## Twierdzenie Riesz.

Jeżeli  $V$  jest przestrzenią Hilberta i  $L$  jest liniowym funkcyjonałem ciągłym na  $V$ , wtedy istnieje dokładnie jedno  $u \in V$  takie, że

$$\langle u, v \rangle = L(v), \quad \forall v \in V.$$

## Słaba forma równania różniczkowego.

Współczesne badania często opierają się na słabej formie równania różniczkowego, podobnie jak różne techniki numeryczne wyznaczania przybliżonych rozwiązań. Równania różniczkowe w słabej postaci można rozwiązywać wykorzystując aparat analizy funkcjonalnej. Słabe formy równania różniczkowego są często odnoszone do „postaci wariacyjne”, ale można je również stosować w celu rozwiązania problemów, które nie można sformułować jako problemu minimalizującego odpowiedni funkcyjonał. Słaba forma równania na ogół nie ułatwia rozwiązania równania analitycznie, ale zwykle jest bardziej odpowiednią formą dla matematycznej analiza (pozwalającą określić właściwości równania bez znajomości rozwiązania) i dla metod numerycznych. Wyprowadzenie słabej postaci równania różniczkowego przebiega zgodnie ze standardowym procesem:

1. Mnożymy równanie różniczkowe przez odpowiednią funkcję testową.
2. Otrzymaną równość całkujemy przez części.

Rozważmy następujący problem brzegowy Dirichleta:

$$(1) \quad \begin{cases} -(a(x) \cdot u'(x))' = f(x), & 0 < x < 1 \\ u(0) = 0, & u(1) = 0. \end{cases}$$

Wyznamy słabą formę równania (1). Mnożymy obustronnie równanie (1) przez funkcję  $v(x)$  różniczkowalną w sposób ciągły i spełniającą warunek  $v(0) = v(1) = 0$

$$(2) \quad -(a(x) \cdot u'(x))' \cdot v(x) = f(x) \cdot v(x).$$

Teraz całkujemy obustronnie równość (2)

$$(3) \quad -\int_0^1 (a(x) \cdot u'(x))' \cdot v(x) dx = \int_0^1 f(x) \cdot v(x) dx.$$

Ponieważ

$$(a(x) \cdot u'(x) \cdot v(x))' = (a(x) \cdot u'(x))' \cdot v(x) + a(x) \cdot u'(x) \cdot v'(x).$$

Stąd

$$(a(x) \cdot u'(x))' \cdot v(x) = (a(x) \cdot u'(x) \cdot v(x))' - a(x) \cdot u'(x) \cdot v'(x),$$

podstawiając powyższą zależność do lewej strony równania (3) otrzymamy:

$$-a(x) \cdot u'(x) \cdot v(x) \Big|_0^1 + \int_0^1 a(x) \cdot u'(x) \cdot v'(x) dx = \int_0^1 f(x) \cdot v(x) dx.$$

Uwzględniając, że  $v(0) = 0$ ,  $v(1) = 0$  dostajemy **słabą formę równania** (1)

$$(4) \quad \int_0^1 a(x) \cdot u'(x) \cdot v'(x) dx = \int_0^1 f(x) \cdot v(x) dx.$$

Rozwiązanie słabej formy jest równoznaczne z minimalizacją funkcjonału:

$$J(u) = \frac{1}{2} \int_0^1 a(x) \cdot (u'(x))^2 dx - \int_0^1 f(x) \cdot u(x) dx.$$

Isotnie, policzmy pochodną kierunkową funkcjonału  $J$  w kierunku wektora  $v(x)$  gdzie  $v(0) = v(1) = 0$ : w tym celu liczymy pochodną funkcji

$$i(t) = J(u + tv) = \frac{1}{2} \int_0^1 a \cdot (u' + t \cdot v')^2 dx - \int_0^1 f \cdot (u + t \cdot v) dx,$$

$$i'(t) = \int_0^1 a \cdot (u' + t \cdot v') \cdot v' dx - \int_0^1 f \cdot v dx.$$

Pochodna kierunkowa funkcjonału  $J$  w kierunku wektora  $v(x)$  jest równa  $i'(0)$ :

$$i'(0) = \int_0^1 a \cdot u' \cdot v' dx - \int_0^1 f \cdot v dx.$$

Funkcjonał  $J$  osiąga ekstremum jeżeli pochodna kierunkowa w kierunku dowolnego wektora  $v(x)$  jest równa zero, czyli

$$\int_0^1 a \cdot u' \cdot v' dx - \int_0^1 f \cdot v dx = 0.$$

To oznacza, że minimum funkcjonału  $J$  jest słabym rozwiązaniem i odwrotnie słabe rozwiązanie jest minimum funkcjonału  $J$ .

### Definicja 2.

Operator liniowy  $A : V \rightarrow V$  jest dodatnio określony jeżeli

$$\langle Au, u \rangle \geq 0 \quad \text{dla} \quad \forall u \in V.$$

Operator liniowy  $A : V \rightarrow V$  jest ujemnie określony jeżeli

$$\langle Au, u \rangle \leq 0 \quad \text{dla} \quad \forall u \in V.$$

Rozpatrzmy następujący problem Dirichleta

$$(5) \quad \begin{cases} Au(x) = f(x), & x \in \Omega, \\ u(x) = 0, & x \in \partial\Omega. \end{cases}$$

Prawdziwe jest następujące twierdzenie

### Twierdzenie 1.

Jeżeli operator liniowy  $A : V \rightarrow V$  jest dodatnio określony to funkcjonał

$$J(u) = \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle$$

przyjmuje dla  $u_0 \in D_A$  wartość minimalną wtedy i tylko wtedy jeżeli  $u_0$  jest rozwiązaniem równania (5).

Jeżeli operator  $A$  jest ujemnie określony to funkcjonał  $J$  przyjmuje dla  $u_0 \in D_A$  wartość maksymalną wtedy i tylko wtedy jeżeli  $u_0$  jest rozwiązaniem równania (5).

Dla problemu (1) operator  $A$  ma postać

$$A = -\frac{d}{dx} \left( a(x) \frac{d}{dx} \right), \quad a \in C^1(0, 1) \quad \text{i} \quad a(x) > 0,$$

a  $f \in C[0, 1]$ .

Dziedziną operatora  $A$  jest zbiór:

$$D_A = \{u \in C^2(0, 1), \quad u(0) = u(1) = 0\}.$$

Pokażemy, że operator  $A$  jest dodatnio określony

$$\begin{aligned} \langle Au, v \rangle &= \int_0^1 \frac{d}{dx} \left( -a \frac{du}{dx} \right) v \, dx = \left( -a \frac{du}{dx} v \right) \Big|_0^1 - \int_0^1 -a \frac{du}{dx} \frac{dv}{dx} \, dx = \\ &= \int_0^1 a \frac{du}{dx} \frac{dv}{dx} \, dx. \end{aligned}$$

Stąd wynika, że

$$\langle Au, u \rangle = \int_0^1 a \left( \frac{du}{dx} \right)^2 \, dx \geq 0,$$

czyli operator  $A$  jest dodatnio określony.

W tej części przedstawimy krótkie wprowadzenie do metod numerycznych rozwiązywania równań różniczkowych Rayleigha-Ritza i Galerkina, stanowią one podstawę metody elementów skończonych.

**Metoda Rayleigha-Ritza** jest stosowana bezpośrednio do wariacyjnej postaci równań, natomiast **metoda Galerkina** do słabej postaci równań.

**Metoda Rayleigha-Ritza.**

Etapy postępowania:

1. Dla równania

$$(6) \quad Au(x) = f(x), \quad x \in \Omega,$$

definiujemy funkcjonal

$$(7) \quad J(u) = \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle$$

dla funkcji  $u$  które na brzegu obszaru  $\Omega$  przyjmują wartość zero.

2. Wybieramy bazę liniowo niezależnych funkcji:

$$u_0, u_1(x), u_2(x), \dots, u_N(x).$$

Szukamy przybliżonego rozwiązania równania (6) w postaci kombinacji funkcji bazowych

$$(8) \quad \tilde{u}(x) = u_0(x) + c_1 u_1(x) + c_2 u_2(x) + \dots + c_N u_N(x),$$

funkcja  $u_0(x)$  musi spełniać warunek brzegowy lub początkowy w zależności jaki będzie rozpatrywany z równaniem (6) a funkcje  $u_1(x), \dots, u_N(x)$  na brzegu są równe zero.

3. Wstawiamy  $\tilde{u}(x)$  do funkcjonału  $J$  i definiujemy funkcję  $\tilde{J}$

$$\tilde{J}(c_1, \dots, c_N) := J(\tilde{u}).$$

4. Liczymy pochodne cząstkowe  $\frac{\partial \tilde{J}}{\partial c_i}$ ,  $i = 1, \dots, N$  i rozwiązujemy układ równań

$$\begin{cases} \frac{\partial \tilde{J}}{\partial c_1}(c_1, \dots, c_N) = 0 \\ \vdots \\ \frac{\partial \tilde{J}}{\partial c_N}(c_1, \dots, c_N) = 0 \end{cases}$$

ze względu na  $c_1, \dots, c_N$ .

### Przykład 1.

Wyznaczyć rozwiązanie problemu brzegowego:

$$(9) \quad \begin{cases} u''(x) + u(x) = x, & x \in [0, 2] \\ u(0) = 0, & u(2) = 5. \end{cases}$$

Rozwiązanie dokładne problemu (5) wyznaczone za pomocą wolframalpha <https://www.wolframalpha.com/input/?i=solve+u%27%27%2Bu%3Dx%2C+u%28%29%3D0%2C+u%282%29%3D5>

ma postać:

$$u(x) = x + \frac{3}{\sin 2} \sin x.$$

Definiujemy funkcjonał  $J$ . W naszym przypadku operator  $A$  ma postać

$$A = \frac{d^2}{dx^2} + 1 \text{ więc}$$

$$J(u) = \frac{1}{2} \langle Au, u \rangle - \langle x, u \rangle = \frac{1}{2} \int_0^2 \left( u \frac{d^2 u}{dx^2} + u^2 \right) dx - \int_0^2 x u dx.$$

Całkując przez części

$$\int_0^2 \left( u \frac{d^2 u}{dx^2} \right) dx = u \frac{du}{dx} \Big|_0^2 - \int_0^2 \left( \frac{du}{dx} \right)^2 dx = - \int_0^2 \left( \frac{du}{dx} \right)^2 dx.$$

Zatem

$$J(u) = -\frac{1}{2} \int_0^2 \left( (u')^2 - u^2 + 2xu \right) dx.$$

Rozwiązania przybliżonego problemu (9) będziemy szukać w postaci wielomianu stopnia trzeciego

$$(10) \quad \tilde{u}(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3.$$

Uwzględniając warunki brzegowe w (9) otrzymujemy:

$$u(0) = c_0 = 0, \quad u(2) = 2c_1 + 4c_2 + 8c_3 = 5,$$

stąd  $c_0 = 0$ ,  $c_3 = \frac{5}{8} - \frac{1}{4}c_1 - \frac{1}{2}c_2$  po podstawieniu do (10) dostajemy

$$(11) \quad \tilde{u}(x) = \frac{5}{8}x^3 + c_1 x \left(1 - \frac{1}{4}x^2\right) + c_2 x^2 \left(1 - \frac{1}{2}x\right).$$

W naszym przypadku baza wygląda następująco:

$$u_0(x) = \frac{5}{8}x^3, \quad u_1(x) = x \left(1 - \frac{1}{4}x^2\right), \quad u_2(x) = x^2 \left(1 - \frac{1}{2}x\right).$$

Funkcja  $u_0$  spełnia warunki brzegowe problemu (9) a funkcje  $u_1$ ,  $u_2$ ,  $u_3$  mają dla  $x = 0$  i  $x = 2$  wartość zero. Podstawiamy funkcję  $\tilde{u}$  do funkcjonału  $J$ :

$$\begin{aligned} \tilde{J}(c_1, c_2) &= J(\tilde{u}) = \\ &= -\frac{1}{2} \int_0^2 \left( (u'_0 + c_1 u'_1 + c_2 u'_2)^2 - (u_0 + c_1 u_1 + c_2 u_2)^2 + 2x(u_0 + c_1 u_1 + c_2 u_2) \right) dx. \end{aligned}$$

Liczmy pochodne :

$$\frac{\partial \tilde{J}}{\partial c_i} = \int_0^2 \left( -(u'_0 + c_1 u'_1 + c_2 u'_2) u'_i + (u_0 + c_1 u_1 + c_2 u_2) u_i - x u_i \right) dx.$$

Przyrównując  $\frac{\partial \tilde{J}}{\partial c_1}$  i  $\frac{\partial \tilde{J}}{\partial c_2}$  do zera otrzymamy układ równań:

$$(12) \quad \begin{cases} a_{11}c_1 + a_{12}c_2 = b_1 \\ a_{21}c_1 + a_{22}c_2 = b_2 \end{cases},$$

gdzie

$$a_{ij} = \int_0^2 (u_i u_j - u'_i u'_j) dx, \quad b_i = \int_0^2 ((x - u_0) u_i + u'_0 u'_i) dx, \quad i, j = 1, 2.$$

Zauważmy, że  $a_{ij} = a_{ji}$ .

W naszym przypadku:

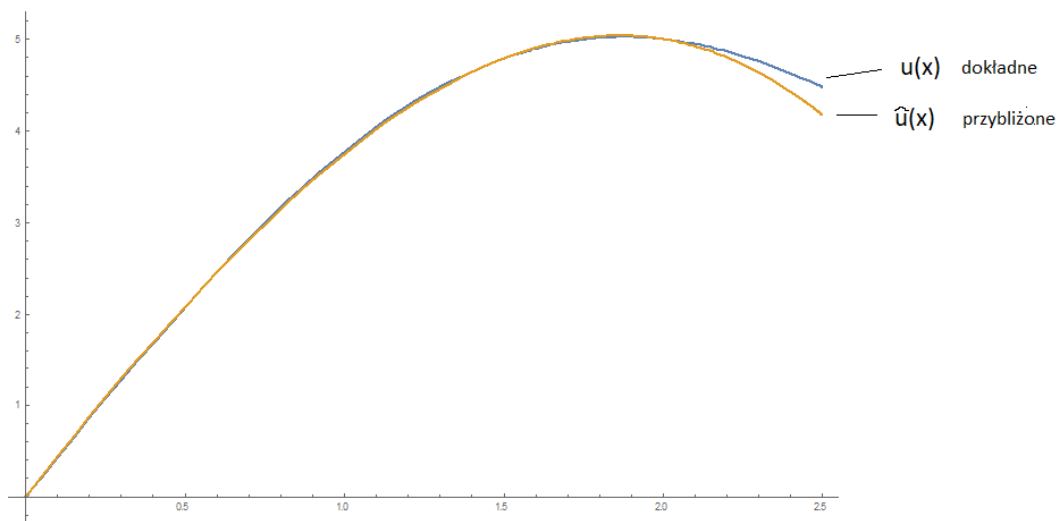
$$a_{11} = -\frac{104}{105}, \quad a_{22} = -\frac{16}{21}, \quad a_{12} = -\frac{82}{105}, \quad b_1 = -\frac{428}{105}, \quad b_2 = -\frac{331}{105}.$$

Rozwiązując układ (12) otrzymamy:

$$c_1 = \frac{169}{38}, \quad c_2 = -\frac{8}{19},$$

[https://www.wolframalpha.com/input/?i=solve+-104%2F105\\*c1-82%2F105\\*c2%3D-428%2F105%2C+-82%2F105\\*c1-16%2F21\\*c2%3D-331%2F105](https://www.wolframalpha.com/input/?i=solve+-104%2F105*c1-82%2F105*c2%3D-428%2F105%2C+-82%2F105*c1-16%2F21*c2%3D-331%2F105) zatem

$$\tilde{u}(x) = \frac{5}{8}x^3 + \frac{169}{38}x\left(1 - \frac{1}{4}x^2\right) - \frac{8}{19}x^2\left(1 - \frac{1}{2}x\right).$$



Rysunek 1:

Z wykresu widać, że na przedziale  $[0, 2]$  wykresy praktycznie się pokrywają.

### Metoda Galerkina (residuów ważonych).

Niech  $A : V \rightarrow V$  będzie operatorem różniczkowym, a  $V$  jest przestrzenią Hilberta z iloczynem skalarnym  $\langle \cdot, \cdot \rangle$ . Rozważmy równanie różniczkowe

$$(13) \quad Au = f, \quad u \in V,$$

zdefiniowane na obszarze  $\Omega$  wraz z warunkami brzegowymi określonymi na brzegu  $\partial\Omega$ . W metodzie residuów ważonych szukanie rozwiązania równania (13) sprowadzamy do słabej postaci wariacyjnej na przestrzeni  $V$ , to jest szukamy funkcji  $u \in V$  spełniającej warunek:

$$(14) \quad \langle Au, v \rangle = \langle f, v \rangle, \quad \forall v \in V.$$

Przybliżone rozwiązanie problemu (14) szukamy w przestrzeni skończenie wymiarowej  $V_n \subset V$ , to jest szukamy funkcji  $\tilde{u} \in V_n$  spełniającej warunek:

$$(15) \quad \langle A\tilde{u}, v_n \rangle = \langle f, v_n \rangle, \quad \forall v_n \in V_n.$$

Niech funkcje  $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$  stanowią bazę przestrzeni  $V_n$ . Rozwiązanie  $u$  jest aproksymowane przez funkcję postaci:

$$\tilde{u}(x) = \varphi_0(x) + \sum_{i=1}^n c_i \varphi_i(x),$$

gdzie funkcja  $\varphi_0$  spełnia wszystkie określone warunki brzegowe zagadnienia ( $\varphi_0(x) = 0$ , jeżeli warunki brzegowe są jednorodne).

Wprowadzimy pojęcie błędu lub residuum,  $R$ , aproksymacji (przez zastąpienie w równaniu (13) szukanej funkcji jej przybliżeniem), które definiujemy jako

$$R = A\tilde{u} - f.$$

Im wartość residuum jest mniejsze, tym aproksymacja rozwiązania jest lepsza. Najlepsze oszacowanie dostaniemy jeżeli spełniony będzie warunek:

$$(16) \quad \langle R, \varphi_j \rangle = 0, \quad j = 1, \dots, n.$$

Powyższą zależność można zapisać w postaci układu równań:

$$(17) \quad \begin{cases} \langle A\varphi_1, \varphi_1 \rangle c_1 + \dots + \langle A\varphi_n, \varphi_1 \rangle c_n = \langle f, \varphi_1 \rangle - \langle A\varphi_0, \varphi_1 \rangle \\ \langle A\varphi_1, \varphi_2 \rangle c_1 + \dots + \langle A\varphi_n, \varphi_2 \rangle c_n = \langle f, \varphi_2 \rangle - \langle A\varphi_0, \varphi_2 \rangle \\ \vdots \\ \langle A\varphi_1, \varphi_n \rangle c_1 + \dots + \langle A\varphi_n, \varphi_n \rangle c_n = \langle f, \varphi_n \rangle - \langle A\varphi_0, \varphi_n \rangle \end{cases}$$



z którego wyznaczamy  $c_1, \dots, c_n$ . Taki sposób wyznaczania rozwiązania przybliżonego nazywamy **metodą Galerkina**.

Można też w przestrzeni  $V_n$  wybrać inną bazę  $\{\psi_1, \psi_2, \dots, \psi_n\}$  i zamiast układu (16) rozważać układ

$$\langle R, \psi_j \rangle = 0, \quad j = 1, \dots, n,$$

taka metoda nazywa się metodą Petrova-Galerkina.

Zakres stosowania metody Galerkina jest o wiele szerszy niż metody Ritz'a. Do jej stosowania nie jest konieczne aby operator  $A$  był dodatnio określony. W metodzie Galerkina punktem wyjścia jest równanie (13), zaś w metodzie Ritz'a- minimalizacja funkcjonału  $J(u)$ .

## Przykład 2.

Rozważmy następujący problem początkowy:

$$(18) \quad u' - u = 0, \quad u(0) = 1, \quad 0 \leq x \leq 1.$$

Rozwiązaniem tego równania jest funkcja  $u(x) = e^x$ .

Wyznamy teraz rozwiązanie przybliżone stosując metodę Galerkina.

Szukamy przybliżonego rozwiązania w postaci:

$$\tilde{u}(x) = \varphi_0(x) + \sum_{i=1}^N c_i \varphi_i(x),$$

gdzie  $\varphi_0(0) = u(0) = 1$ , zatem przyjmujemy, że  $\varphi_0(x) = 1$ .

Natomiast funkcje bazowe  $\varphi_i(x) = x^i$ ,  $i = 1, \dots, N$ .

W naszym przypadku operator  $A$  ma postać

$$A = \frac{d}{dx} - 1,$$

a iloczyn skalarny określony jest następująco:

$$\langle f, g \rangle = \int_0^1 f(x)g(x) dx.$$

Wprowadzimy następujące oznaczenia:

$$a_{ij} = \langle A\varphi_j, \varphi_i \rangle = \int_0^1 (\varphi_j'(x) - \varphi_j(x))\varphi_i(x) dx = \int_0^1 (jx^{j-1} - x^j)x^i dx = \int_0^1 (jx^{j+i-1} - x^{j+i}) dx = j \frac{x^{j+i}}{j+i} - \frac{x^{j+i+1}}{j+i+1} \Big|_0^1 = \frac{j}{j+i} - \frac{1}{j+i+1},$$

$$b_i = - \langle A\varphi_0, \varphi_i \rangle = - \int_0^1 (1-x)x^i dx = \int_0^1 x^i dx = \frac{x^{i+1}}{i+1} \Big|_0^1 = \frac{1}{i+1}.$$

Układ równań (17) można zapisać w postaci macierzowej:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}.$$

Wykonamy obliczenia dla  $N = 3$ .

$$\begin{pmatrix} \frac{1}{6} & \frac{5}{12} & \frac{11}{20} \\ \frac{1}{12} & \frac{1}{10} & \frac{3}{30} \\ \frac{1}{20} & \frac{7}{30} & \frac{5}{14} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{4} \end{pmatrix}.$$

Rozwiązując ten układ otrzymamy

<https://www.wolframalpha.com/input/?i=%28Inverse%7B%7B1%2F6%2C5%2F12%2C11%2F20%7D%2C%7B1%2F12%2C3%2F10%2C13%2F30%7D%2C%7B1%2F20%2C7%2F30%2C5%2F14%7D%7D%29.%7B%7B1%2F2%7D%2C%7B1%2F3%7D%2C%7B1%2F4%7D%7D>

$$c_1 = \frac{120}{116}, \quad c_2 = \frac{45}{116}, \quad c_3 = \frac{35}{116}.$$

Przybliżone rozwiązanie jest równe

$$\tilde{u}(x) = 1 + \frac{120}{116}x + \frac{45}{116}x^2 + \frac{35}{116}x^3.$$

Na poniższym wykresie mamy rozwiązanie dokładne i przybliżone

[https://www.wolframalpha.com/input/?i=plot%28%7Be%5Ex%2C1%2B120%2F116\\*x%2B45%2F116\\*x%5E2%2B35%2F116\\*x%5E3%7D+from+x%3D0+to+%29](https://www.wolframalpha.com/input/?i=plot%28%7Be%5Ex%2C1%2B120%2F116*x%2B45%2F116*x%5E2%2B35%2F116*x%5E3%7D+from+x%3D0+to+%29) z wykresu widać, że rozwiązania praktycznie się pokrywają na przedziale  $[0, 1]$ .

Zanim przejdziemy do metody elementów skończonych wprowadzimy pewne definicje.

$$L^2(\Omega) = \{u : \Omega \rightarrow \mathbb{R}, \int_{\Omega} u^2(x) d\Omega < +\infty, \quad \Omega \subset \mathbb{R}^n\}.$$

Zbiór  $L^2(\Omega)$  jest przestrzenią Hilberta z iloczynem skalarnym

$$\langle u, v \rangle = \int_{\Omega} u(x)v(x) d\Omega.$$

**Definicja słabej pochodnej.** Mówimy, że funkcja  $u(x) \in L^2(\Omega)$  ma słabą pochodną ze względu na zmienną  $x_i$  jeżeli istnieje funkcja  $w(x) \in L^2(\Omega)$  taka, że dla każdej funkcji  $\psi(x)$  klasy  $C^1$  na  $\Omega \subset \mathbb{R}^n$  i mającej wartość zero na brzegu obszaru  $\partial\Omega$  spełnione jest równanie:

$$\int_{\Omega} u(x) \frac{\partial \psi}{\partial x_i}(x) d\Omega = - \int_{\Omega} w(x) \psi(x) d\Omega.$$

Funkcję  $w$  oznaczają będziemy  $\frac{\partial u}{\partial x_i}$  i nazywać pochodną słabą funkcji  $u$ .

**Przykład 3.** Funkcja  $u(x) = |x|$ ,  $x \in \mathbb{R}$  nie jest różniczkowalna w normalnym sensie ponieważ nie istnieje pochodna w  $x = 0$ , natomiast jest różniczkowalna w słabym sensie  $\frac{du}{dx} = \text{sgn}(x)$ .

Istotnie, jeżeli weźmiemy dowolny przedział  $[a, b]$  zawierający 0 i dowolną funkcję  $\psi(x)$  klasy  $C^1$  na  $(a, b)$  i przyjmującą wartość zero w  $a$  i  $b$  to

$$\int_a^b |x| \frac{d\psi}{dx} dx = \int_a^0 -x \frac{d\psi}{dx} + \int_0^b x \frac{d\psi}{dx}.$$

Wyrażenia po prawej stronie równości całkujemy przez części:

$$\int_a^0 -x \frac{d\psi}{dx} dx = -x\psi(x)|_a^0 - \int_a^0 \frac{d(-x)}{dx} \psi(x) dx = - \int_a^0 \text{sgn}(x) \psi(x) dx,$$

$$\int_0^b x \frac{d\psi}{dx} dx = x\psi(x)|_0^b - \int_0^b \frac{d(x)}{dx} \psi(x) dx = - \int_0^b \text{sgn}(x) \psi(x) dx.$$

Zatem

$$\int_a^b |x| \frac{d\psi}{dx} dx = - \int_a^b \text{sgn}(x) \psi(x) dx,$$

co kończy dowód, że funkcja  $u(x) = |x|$  jest różniczkowalna w słabym sensie i jej słaba pochodna jest równa  $\frac{du}{dx} = \text{sgn}(x)$ .

**Definicja przestrzeni  $H^1(\Omega)$**

$$H^1(\Omega) = \left\{ v(x) : v \in L^2(\Omega) \text{ i słabe pochodne } \frac{\partial v}{\partial x_i} \in L^2(\Omega), i : 1, \dots, n \right\}.$$

Przestrzeń  $H^1(\Omega)$  jest przestrzenią Hilberta z iloczynem skalarnym:

$$\langle u, v \rangle = \int_{\Omega} u(x)v(x) d\Omega + \sum_{i=1}^n \int_{\Omega} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} d\Omega.$$

### Metoda elementów skończonych MES

Przedstawione powyżej metody wariacyjne, mogą być efektywnie wykorzystywane tylko do rozwiązywania problemów z małą liczbą stopni swobody. Spowodowane jest to przede wszystkim trudnościami w doborze funkcji aproksymacyjnych. Jeśli nawet takie funkcje zostałyby dobrane, to z kolei obliczanie współczynników macierzy równań algebraicznych rozwiązania przybliżonego nie może być zautomatyzowane, ponieważ funkcje aproksymacyjne zmieniają się zależnie od rozważanego problemu.

Metoda elementów skończonych jest procedurą wariacyjną, w której funkcje aproksymacyjne są wyznaczane w obszarze zastąpionym przez zbiór prostych podobszarów, na jakie obszar ten został podzielony. Układ równań algebraicznych problemu jest w tej metodzie generowany w odmienny sposób niż to ma miejsce w tradycyjnych metodach wariacyjnych, ponieważ funkcje aproksymacyjne są wielomianami algebraicznymi, wyznaczonymi według zasad aproksymacji interpolacyjnej. Podobszary, nazywane elementami skończonymi, mają zwykle geometrycznie proste kształty, co ułatwia budowanie funkcji aproksymacyjnych. Ponieważ funkcje aproksymacyjne są wielomianami algebraicznymi, to możliwe jest zautomatyzowanie na komputerze obliczenia współczynników macierzy równań algebraicznych. W metodzie elementów skończonych definiowanie funkcji aproksymacyjnych jest niezależne od danych warunków brzegowych i innych danych definiujących rozwiązywany problem.

**Przykład 4.** Rozważmy następujący problem brzegowy:

$$(19) \quad \begin{cases} -(p(x) \cdot u'(x))' + q(x)u(x) = f(x), & 0 < x < 1 \\ u(0) = 0, & u(1) = 0. \end{cases},$$

gdzie  $p$ ,  $q$ ,  $f$  dane funkcje.

Wyznamy słabą formę równania (19). Mnożymy obustronnie równanie (19) przez funkcję  $v(x) \in H^1(0, 1)$  spełniającą warunek  $v(0) = v(1) = 0$

$$(20) \quad -(p(x) \cdot u'(x))' \cdot v(x) + q(x)u(x)v(x) = f(x) \cdot v(x).$$

Teraz całkujemy obustronnie równość (20) w przedziale  $[0, 1]$ :

$$(21) \quad - \int_0^1 [(p(x) \cdot u'(x))' - q(x) \cdot u(x)] \cdot v(x) dx = \int_0^1 f(x) \cdot v(x) dx.$$

Ponieważ

$$(p(x) \cdot u'(x) \cdot v(x))' = (p(x) \cdot u'(x))' \cdot v(x) + p(x) \cdot u'(x) \cdot v'(x).$$

Stąd

$$(p(x) \cdot u'(x))' \cdot v(x) = (p(x) \cdot u'(x) \cdot v(x))' - p(x) \cdot u'(x) \cdot v'(x),$$

podstawiając tą zależność do poniższej całki

$$- \int_0^1 (p(x) \cdot u'(x))' \cdot v(x) dx$$

i całkując przez części dostajemy

$$- \int_0^1 (p(x) \cdot u'(x))' \cdot v(x) dx = -p(x) \cdot u'(x) \cdot v(x) \Big|_0^1 + \int_0^1 p(x) \cdot u'(x) \cdot v'(x) dx.$$

Uwzględniając, że  $v(0) = v(1) = 0$  mamy

$$- \int_0^1 (p(x) \cdot u'(x))' \cdot v(x) dx = \int_0^1 p(x) \cdot u'(x) \cdot v'(x) dx.$$

Po podstawieniu tej zależności do równości (21) otrzymamy słabą formę równania 19

$$(22) \quad \int_0^1 [p(x) \cdot u'(x) \cdot v'(x) + q(x) \cdot u(x) \cdot v(x)] dx = \int_0^1 f(x) \cdot v(x) dx.$$

Szukamy funkcji  $u \in H^1(0, 1)$  która spełnia równanie (22) i warunki brzegowe  $u(0) = u(1) = 0$  dla każdej funkcji  $v \in H^1(0, 1)$  takiej, że  $v(0) = v(1) = 0$ .  
Dzielimy przedział  $[0, 1]$  na  $n$  podprzedziałów:

$$[x_i, x_{i+1}], \quad i = 0, \dots, n-1, \quad \text{gdzie } x_i = ih, \quad i = 0, \dots, n, \quad h = \frac{1}{n}, \quad n \geq 2.$$

W przypadku ogólnym długości przedziałów nie muszą być równe. Przyjmujemy równą długość przedziałów w celu uproszczenia rachunków.

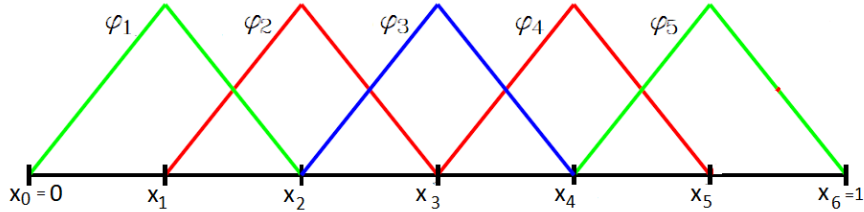
Niech

$$V = \{v \in C[0, 1] : v \text{--liniowa na każdym podprzedziale } [x_i, x_{i+1}], v(0) = v(1) = 0\}.$$

Baza przestrzeni  $V$  składa się z funkcji daszkowych  $\varphi_1, \dots, \varphi_{n-1}$ , gdzie

$$\varphi_i(x) = \max\left\{1 - \frac{|x - x_i|}{h}, 0\right\}.$$

Przykład funkcji bazowych przy podziale przedziału  $[0, 1]$  na 6-podprzedziałów.



Rysunek 2: Przykład funkcji bazowych dla  $n=6$

Niech

$$\text{sup } \varphi_i := \{x \in [0, 1] : \varphi_i(x) > 0\}.$$

Dla  $1 < i < n - 1$  mamy

$$\text{sup } \varphi_i \cap \text{sup } \varphi_{i-1} \neq \emptyset \quad \text{i} \quad \text{sup } \varphi_i \cap \text{sup } \varphi_{i+1} \neq \emptyset.$$

Dla  $i = 1$  i  $i = n - 1$  mamy

$$\text{sup } \varphi_1 \cap \text{sup } \varphi_2 \neq \emptyset \quad \text{i} \quad \text{sup } \varphi_{n-2} \cap \text{sup } \varphi_{n-1} \neq \emptyset.$$

Dla  $|i - j| > 1$  mamy

$$\text{sup } \varphi_i \cap \text{sup } \varphi_j = \emptyset.$$

Szukamy przybliżonego rozwiązania równania (22) w postaci kombinacji funkcji bazowych:

$$(23) \quad \tilde{u}(x) = \sum_{i=1}^{n-1} c_i \cdot \varphi_i(x).$$

Podstawiamy  $\tilde{u}$  do równania (22) w miejsce  $u$  i  $\varphi_j$  w miejsce  $v$  i dostajemy

$$(24) \quad \sum_{i=1}^{n-1} c_i \int_0^1 [p(x) \cdot \varphi_i'(x) \cdot \varphi_j'(x) + q(x) \cdot \varphi_i(x) \cdot \varphi_j(x)] dx = \int_0^1 f(x) \varphi_j(x) dx.$$

Ta zależność ma zachodzić dla każdego  $j = 1, \dots, n - 1$ . Z tych  $n - 1$  równań wyznaczymy współczynniki  $c_1, \dots, c_{n-1}$ .

Wprowadzamy następujące oznaczenia:

$$a_{i,j} = \int_0^1 [p(x) \cdot \varphi_i'(x) \cdot \varphi_j'(x) + q(x) \cdot \varphi_i(x) \cdot \varphi_j(x)] dx, \quad i, j = 1, \dots, n-1$$

$$b_j = \int_0^1 f(x) \varphi_j(x) dx, \quad j = 1, \dots, n-1.$$

Definiujemy macierze:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n-1} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n-1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & \dots & a_{n-1,n-1} \end{pmatrix}, \quad C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \end{pmatrix}.$$

Układ równań (24) można zapisać w postaci macierzowej:

$$A \cdot C = B$$

Wyliczone z tego układu współczynniki  $c_i$ ,  $i = 1, \dots, n-1$  podstawiamy do (23) i dostajemy rozwiązanie przybliżone.

Pokażemy, że w każdym wierszu macierzy  $A$  co najwyżej tylko trzy elementy są różne od zera. Istotnie

1.  $i = 1$

$$a_{1,j} = 0 \text{ dla } j > 2 \text{ ponieważ } \text{sup } \varphi_1 \cap \text{sup } \varphi_j = \emptyset,$$

$$a_{1,1} = \int_0^{2h} [p(x) \cdot \varphi_1'(x) \cdot \varphi_1'(x) + q(x) \cdot \varphi_1(x) \cdot \varphi_1(x)] dx,$$

$$a_{1,2} = \int_h^{2h} [p(x) \cdot \varphi_1'(x) \cdot \varphi_2'(x) + q(x) \cdot \varphi_1(x) \cdot \varphi_2(x)] dx.$$

2.  $1 < i < n-1$

$$a_{i,j} = 0 \text{ gdy } |i-j| > 1 \text{ ponieważ } \text{sup } \varphi_i \cap \text{sup } \varphi_j = \emptyset,$$

$$a_{i,i-1} = \int_{(i-1)h}^{ih} [p(x) \cdot \varphi_i'(x) \cdot \varphi_{i-1}'(x) + q(x) \cdot \varphi_i(x) \cdot \varphi_{i-1}(x)] dx,$$

$$a_{i,i} = \int_{(i-1)h}^{(i+1)h} [p(x) \cdot \varphi_i'(x) \cdot \varphi_i'(x) + q(x) \cdot \varphi_i(x) \cdot \varphi_i(x)] dx,$$

$$a_{i,i+1} = \int_{ih}^{(i+1)h} [p(x) \cdot \varphi_i'(x) \cdot \varphi_{i+1}'(x) + q(x) \cdot \varphi_i(x) \cdot \varphi_{i+1}(x)] dx.$$

3.  $i = n-1$

$$a_{n-1,j} = 0 \text{ dla } j < n-2 \text{ ponieważ } \text{sup } \varphi_{n-1} \cap \text{sup } \varphi_j = \emptyset,$$

$$a_{n-1,n-2} = \int_{(n-2)h}^{(n-1)h} [p(x) \cdot \varphi'_{n-1}(x) \cdot \varphi'_{n-2}(x) + q(x) \cdot \varphi_{n-1}(x) \cdot \varphi_{n-2}(x)] dx,$$

$$a_{n-1,n-1} = \int_{(n-2)h}^{nh} [p(x) \cdot \varphi'_{n-1}(x) \cdot \varphi'_{n-1}(x) + q(x) \cdot \varphi_{n-1}(x) \cdot \varphi_{n-1}(x)] dx.$$

W przypadku kiedy  $p$  i  $q$  są funkcjami stałymi to

$$a_{i,j} = \begin{cases} \frac{2p}{h} + \frac{2hq}{3}, & i = j \\ \frac{-p}{h} + \frac{hq}{6}, & |i - j| = 1 \\ 0, & |i - j| > 1 \end{cases}$$

### Przykład 5.

Rozważmy następujący problem brzegowy:

$$(24) \quad -u''(x) + 4u(x) = 4, \quad u(0) = u(1) = 0, \quad 0 < x < 1.$$

Poniżej podaje kod do obliczeń w programie "Mathematica".

Remove[n,h,r,p,y,ode,sol,exactSolution,coef,s,yapprox,error,tab,coe] // **usu-**

**wamy następujące symbole z pamięci**

ode[y\_,x\_.]:=D[y[x],{x,2}]+4\*y[x]-4;//  $-y''(x) + 4y(x) - 4$

sol=Flatten[DSolve[{ode[y,x]==0,y[0]==0,y[1]==0},y[x],x]);// **wyznaczamy rozwiązanie dokładne problemu (24)**

exactSolution[x\_.]:=y[x]/.sol;// **wyznaczone rozwiązanie dokładne**

n=10;// **ilość podprzedziałów na którą dzielimy przedział [0, 1]**

h=N[1/n];// **długość każdego z podprzedziału**

coe=Array[c,n-1];// **definiujemy wektor nieznanych współczynników**

$c_1, \dots, c_{n-1}$

Do[

p[j\_,x\_.]:=Piecewise[{{x/h-j+1,(j-1)\*h<=x<=j\*h},{j+1-x/h,j\*h<x<=(j+1)\*h}},{j,n-1}];// **definicja funkcji bazowych**

r[1]=Integrate[D[p[1,x],x]\*(coe[[1]]\*D[p[1,x],x]+coe[[2]]\*D[p[2,x],x])+

4\*p[1,x]\*(coe[[1]]\*p[1,x]+coe[[2]]\*p[2,x]-1),{x,0,2\*h}];// **obliczamy  $a_{1,j} - b_1$**

r[n-1]=Integrate[D[p[n-1,x],x]\*(coe[[n-2]]\*D[p[n-2,x],x]+coe[[n-1]]\*D[p[n-1,x],x])+

4\*p[n-1,x]\*(coe[[n-2]]\*p[n-2,x]+coe[[n-1]]\*p[n-1,x]-1),{x,(n-2)\*h,n\*h}];// **ob-**

**liczamy  $a_{n-1,j} - b_{n-1}$**

Do[

r[i]=Integrate[D[p[i,x],x]\*(coe[[i-1]]\*D[p[i-1,x],x]+coe[[i]]\*D[p[i,x],x]

+coe[[i+1]]\*D[p[i+1,x],x])+4\*p[i,x]\*(coe[[i-1]]\*p[i-1,x]+coe[[i]]\*p[i,x]

+coe[[i+1]]\*p[i+1,x]-1),{x,(i-1)\*h,(i+1)\*h},

{i,2,n-2,1}];// **obliczamy  $a_{i,j} - b_i, i = 2, \dots, n - 2$**



```

s=Table[r[i]==0,{i,1,n-1}];// układ równań z niewiadomymi  $c_1, \dots, c_{n-1}$ 
coef=Flatten[Solve[s,coef]];// wyznaczamy  $c_1, \dots, c_{n-1}$ 
yapprox[x_]:=Sum[coef[[i,2]]*p[i,x],{i,1,n-1}];// rozwiązanie przybliżone  $\tilde{u}(x)$ 
error[x_]:=N[Abs[exactSolution[x]-yapprox[x]]];// wyliczamy różnice  $|u(x) - \tilde{u}(x)|$ 
tab=Table[{x,exactSolution[x],yapprox[x],error[x]},{x,0,1,h}];// określamy tabelę wyników
TableForm[tab, TableHeadings -> None, "x", "exact", "approx", "error"];// wyniki obliczeń

```

x	exact	approx	error
0.	0.	0.	0.
0.1	0.13327	0.133604	0.000334289
0.2	0.231754	0.232319	0.000565078
0.3	0.299406	0.300121	0.00071495
0.4	0.338941	0.33974	0.000799071
0.5	0.351946	0.352772	0.000826163
0.6	0.338941	0.33974	0.000799071
0.7	0.299406	0.300121	0.00071495
0.8	0.231754	0.232319	0.000565078
0.9	0.13327	0.133604	0.000334289
1.	$2.86568 \times 10^{-17}$	0.	$2.86568 \times 10^{-17}$

Rysunek 3: n=10

x	exact	approx	error
0.	0.	0.	0.
0.05	0.0712822	0.0713275	0.0000452983
0.1	0.13327	0.133353	0.0000832878
0.15	0.186582	0.186697	0.000114867
0.2	0.231754	0.231895	0.000140798
0.25	0.269237	0.269399	0.000161718
0.3	0.299406	0.299585	0.00017815
0.35	0.322564	0.322754	0.000190511
0.4	0.338941	0.33914	0.000199117
0.45	0.348703	0.348907	0.000204193
0.5	0.351946	0.352152	0.00020587
0.55	0.348703	0.348907	0.000204193
0.6	0.338941	0.33914	0.000199117
0.65	0.322564	0.322754	0.000190511
0.7	0.299406	0.299585	0.00017815
0.75	0.269237	0.269399	0.000161718
0.8	0.231754	0.231895	0.000140798
0.85	0.186582	0.186697	0.000114867
0.9	0.13327	0.133353	0.0000832878
0.95	0.0712822	0.0713275	0.0000452983
1.	$2.86568 \times 10^{-17}$	0.	$2.86568 \times 10^{-17}$

Rysunek 4: n=20

```

Remove[n, h, r, p, y, ode, sol, exactSolution, coef, s, yapprox, error, tab, coef]
|usuń symbole
ode[y_, x_] := -D[y[x], {x, 2}] + 4*y[x] - 4;
|oblicz pochodną
sol = Flatten[DSolve[{ode[y, x] == 0, y[0] == 0, y[1] == 0}, y[x], x]];
|spłaszcz |rozwiązywanie równań różniczkowych
exactSolution[x_] := y[x] /. sol;
n = 20;
h = N[1/n];
|przybliżenie numeryczne
coef = Array[c, n - 1];
|tablica wielowymiarowa
Do[
|rób
  p[j_, x_] := Piecewise[{{x/h - j + 1, (j - 1)*h <= x <= j*h}, {j + 1 - x/h, j*h < x <= (j + 1)*h}}, {j, n - 1}];
|funkcja odcinkowa
r[1] = Integrate[D[p[1, x], x] * (coef[[1]] * D[p[1, x], x] + coef[[2]] * D[p[2, x], x]) + 4*p[1, x] * (coef[[1]] * p[1, x] +
|całka |oblicz pochodną |oblicz pochodną |oblicz pochodną
  coef[[2]] * p[2, x] - 1), {x, 0, 2*h}];
r[n - 1] = Integrate[D[p[n - 1, x], x] * (coef[[n - 2]] * D[p[n - 2, x], x] + coef[[n - 1]] * D[p[n - 1, x], x]) +
|całka |oblicz pochodną |oblicz pochodną |oblicz pochodną
  4*p[n - 1, x] * (coef[[n - 2]] * p[n - 2, x] + coef[[n - 1]] * p[n - 1, x] - 1), {x, (n - 2)*h, n*h}];
Do[
|rób
  r[i] = Integrate[D[p[i, x], x] * (coef[[i - 1]] * D[p[i - 1, x], x] + coef[[i]] * D[p[i, x], x] + coef[[i + 1]] * D[p[i + 1, x], x]) +
|całka |oblicz pochodną |oblicz pochodną |oblicz pochodną |oblicz pochodną
  4*p[i, x] * (coef[[i - 1]] * p[i - 1, x] + coef[[i]] * p[i, x] + coef[[i + 1]] * p[i + 1, x] - 1), {x, (i - 1)*h, (i + 1)*h}], {i, 2, n - 2, 1}];

s = Table[r[i] == 0, {i, 1, n - 1}];
|tabela
s
coef = Flatten[Solve[s, coef]];
|spłaszcz |rozwiąż równanie
yapprox[x_] := Sum[coef[[i, 2]] * p[i, x], {i, 1, n - 1}];
|sumowanie
error[x_] := N[Abs[exactSolution[x] - yapprox[x]]];
|wartość bezwzględna
tab = Table[{x, exactSolution[x], yapprox[x], error[x]}, {x, 0, 1, h}];
|tabela
TableForm[tab, TableHeadings -> {None, {"x", "exact", "approx", "error"}}]
|w formie tabeli |nagłówki tabeli |zaden

```

Rysunek 5: