Snadno se ověří, že pro $m \in \mathbb{N}$ je $A_m = -A_{-m} = -\mathrm{i}\, a_m/2$. Budeme předpokládat, že Fourierova řada (69) je absolutně konvergentní, tj.

$$\sum_{m=-\infty}^{\infty} |A_m| < \infty\,. \tag{70}$$

Jak víme, postačující podmínkou pro to je, aby funkce u^0 byla absolutně spojitá na $[0,1]$, $(u^0)' \in L^2(0,1)$ a $u^0(0) = u^0(1) = 0$.

Připomeňme, že funkce $\mathrm{e}^{\mathrm{i}\, m\,\pi\, x - (m\,\pi)^2\, t}$ jsou řešeními rovnice (59). V uzlech sítě platí

$$\mathrm{e}^{\mathrm{i}\, m\,\pi\, x_j - (m\,\pi)^2\, t_n} = \mathrm{e}^{\mathrm{i}\, k\, j\, h} \left[\mathrm{e}^{-k^2\,\tau}\right]^n,$$

kde jsem zavedli *vlnové číslo* $k = m\,\pi$. V analogii k tomu se můžeme ptát, kdy

$$U_j^n = \mathrm{e}^{\mathrm{i}\, k\, j\, h}\, \lambda^n$$

řeší diferenční rovnici (64). Dosazením do (64) získáme

$$\mathrm{e}^{\mathrm{i}\, k\, j\, h}\, \lambda^{n+1} = \mathrm{e}^{\mathrm{i}\, k\, j\, h}\, \lambda^n\, [1 + \mu\, (\mathrm{e}^{\mathrm{i}\, k\, h} - 2 + \mathrm{e}^{-\mathrm{i}\, k\, h})],$$

a tudíž

$$\lambda \equiv \lambda(k) = 1 - 2\,\mu\,[1 - \cos(k\,h)] = 1 - 4\,\mu\, \sin^2 \frac{k\,h}{2}\,. \tag{71}$$

Číslo $\lambda(k)$ se nazývá amplifikační (zesilující) faktor členu Fourierovy řady. Následující věta ukazuje, že přibližné řešení U_j^n můžeme zapsat ve tvaru podobném jako u.

Věta 6 *Nechť síťová funkce* U_j^n *splňuje* (64)–(66). *Pak*

$$U_j^n = \sum_{m=-\infty}^{\infty} A_m\, \mathrm{e}^{\mathrm{i}\, m\,\pi\, j\, h}\, [\lambda(m\,\pi)]^n\,, \qquad j = 0, 1, \ldots, J\,, \quad n \geq 0\,. \tag{72}$$

Důkaz. Jelikož amplifikační faktory jsou omezené a platí (70), konverguje řada (72) absolutně, a jelikož každý její člen řeší diferenční rovnici (64), řeší i součet řady (72) rovnici (64). Dále pro každé $m \in \mathbb{N}$ platí $A_m\,[\lambda(m\,\pi)]^n = -A_{-m}\,[\lambda(-m\,\pi)]^n$, a tudíž je pro $j = 0$ a $j = J$ součtem řady 0. Jsou tedy splněny okrajové podmínky (65). Konečně pro $n = 0$ se řada (72) redukuje na řadu (69) s $x = j\,h$, a součet řady tudíž splňuje i počáteční podmínku (66). Součet řady tedy splňuje (64)–(66), a jelikož je řešení úlohy (64)–(66) určeno jednoznačně, platí (72). □

Poznámka 5 Všimněme si, že z vlastností amplifikačního faktoru a koeficientů A_m plyne, že řadu (72) lze též přepsat do tvaru

$$U_j^n = \sum_{m=1}^{\infty} a_m\, \sin(m\,\pi\, j\, h)\, [\lambda(m\,\pi)]^n\,, \qquad j = 0, 1, \ldots, J\,, \quad n \geq 0\,. \tag{73}$$

Ze vztahu (68) plyne, že

$$u_j^n = \sum_{m=-\infty}^{\infty} A_m \, \mathrm{e}^{\mathrm{i}\, m\, \pi\, j\, h} \left[\mathrm{e}^{-(m\,\pi)^2\,\tau}\right]^n .$$

Jelikož chceme, aby U_j^n dobře aproximovalo u_j^n, vidíme ze srovnání uvedené řady s řadou (72), že je potřeba, aby hodnoty $\lambda(k)$ byly dobrou aproximací hodnot $\mathrm{e}^{-k^2\,\tau}$, alespoň pro nízké hodnoty vlnového čísla k. To je předmětem následujícího lemmatu.

Lemma 1 *Platí*

$$|\mathrm{e}^{-k^2\tau} - \lambda(k)| \le C(\mu)\, k^4\, \tau^2 \qquad \forall\, k, \tau > 0\,, \tag{74}$$

kde $C(\mu)$ závisí pouze na μ. Pokud $\mu = \frac{1}{6}$, pak

$$|\mathrm{e}^{-k^2\tau} - \lambda(k)| \le \frac{4}{15}\, k^6\, \tau^3 \qquad \forall\, k, \tau > 0\,.$$

Důkaz. K důkazu využijeme Taylorův vzorec ve tvaru

$$f(x) = \sum_{k=0}^{l-1} \frac{1}{k!}\, f^{(k)}(0)\, x^k + \frac{1}{l!}\, f^{(l)}(\xi)\, x^l\,,$$

který platí pro libovolné $l \in \mathbb{N}$, $f \in C^l(\mathbb{R})$ a $x > 0$ s vhodným $\xi \in (0, x)$. Aplikujeme-li tento vzorec na $f(x) = \mathrm{e}^{-x}$ a $f(x) = \cos x$, získáme

$$\begin{aligned} \mathrm{e}^{-k^2\tau} &= 1 - k^2\,\tau + \tfrac{1}{2}\,\mathrm{e}^{-\xi}\, k^4\, \tau^2 \\ \cos(k\,h) &= 1 - \tfrac{1}{2}\,(k\,h)^2 + \tfrac{1}{24}\,(\cos\zeta)\,(k\,h)^4\,, \end{aligned}$$

kde $\xi \in (0, k^2\tau)$ a $\zeta \in (0, kh)$. Podle (71) je tedy

$$\lambda(k) = 1 - 2\,\mu\,[1 - \cos(k\,h)] = 1 - k^2\,\tau + \tfrac{1}{12}\,(\cos\zeta)\, k^4\, h^2\, \tau\,,$$

z čehož plyne

$$|\mathrm{e}^{-k^2\tau} - \lambda(k)| = \frac{1}{2}\left| k^4\,\tau^2\,\mathrm{e}^{-\xi} - \frac{1}{6}\, k^4\, h^2\, \tau \cos\zeta \right| \le \frac{1}{2}\left(1 + \frac{1}{6\,\mu}\right) k^4\, \tau^2\,,$$

což dokazuje (74). Podobně získáme

$$|\mathrm{e}^{-k^2\tau} - \lambda(k)| = \frac{1}{360}\, |60\, k^6\, \tau^3\, \mathrm{e}^{-\widetilde{\xi}} - k^6\, h^4\, \tau \cos\widetilde{\zeta}| \le \frac{4}{15}\, k^6\, \tau^3 \qquad \text{pro} \quad \mu = \frac{1}{6}\,,$$

kde opět $\widetilde{\xi} \in (0, k^2\tau)$ a $\widetilde{\zeta} \in (0, kh)$. □

Uvedené lemma můžeme využít jako alternativní prostředek pro vyšetřování chyby diskretizace. Na základě vztahů (68) a (70) platí

$$\varepsilon_j^n = \frac{u_j^{n+1} - u_j^n}{\tau} - \frac{u_{j+1}^n - 2\,u_j^n + u_{j-1}^n}{h^2}$$

$$= \sum_{m=-\infty}^{\infty} A_m \, \mathrm{e}^{\mathrm{i}\, m\, \pi\, x_j - (m\, \pi)^2\, t_n} \left[\frac{\mathrm{e}^{-(m\,\pi)^2\,\tau} - 1}{\tau} - \frac{\mathrm{e}^{\mathrm{i}\, m\, \pi\, h} - 2 + \mathrm{e}^{-\mathrm{i}\, m\, \pi\, h}}{h^2} \right]$$

$$= \sum_{m=-\infty}^{\infty} A_m \, \mathrm{e}^{\mathrm{i}\, m\, \pi\, x_j - (m\, \pi)^2\, t_n} \, \frac{1}{\tau} \left[\mathrm{e}^{-(m\,\pi)^2\,\tau} - \lambda(m\,\pi) \right] .$$

Použitím odhadu (74) dostáváme

$$|\varepsilon_j^n| \le \sum_{m=-\infty}^{\infty} |A_m| \, \frac{1}{\tau} \, |\mathrm{e}^{-(m\,\pi)^2\,\tau} - \lambda(m\,\pi)| \le \sum_{m=-\infty}^{\infty} |A_m| \, C(\mu) \, (m\,\pi)^4 \, \tau \, .$$

Abychom při konstantním μ získali stejnoměrný odhad chyby diskretizace prvního řádu přesnosti, je tedy nutné požadovat, aby $\pi^4 \sum_{m=-\infty}^{\infty} |A_m| \, m^4 < \infty$. Tato hodnota představuje odhad $|u_{xxxx}| = |u_{tt}|$ a dostáváme tedy odhad (67). Pro $h^2 = 6\,\tau$ lze pomocí lemmatu 1 získat přesnost druhého řádu (srv. pozn. 4).

Vyjádření řešení úlohy (59)–(61) ve tvaru řady (68) ukazuje, že pro libovolné vlnové číslo $k = m\,\pi$ zůstává amplituda příslušné složky řešení v čase omezená. Je přirozené tuto vlastnost požadovat i od přibližného řešení. Řekneme proto, že numerická metoda pro řešení úlohy (59)–(61) je stabilní, pokud existuje konstanta K nezávislá na k taková, že $|[\lambda(k)]^n| \le K$ pro všechna k a n. To je ekvivalentní požadavku, aby $|\lambda(k)| \le 1$ pro všechna k.

Věta 7 *Nechť pro dané hodnoty h a τ je $|\lambda(m\,\pi)| \le 1$ pro všechna $m \in \mathbb{N}$. Pak řešení schématu* (64)–(66) *zůstává pro libovolnou počáteční podmínku u^0 splňující* (70) *omezené pro $n \to \infty$.*

Důkaz. Jelikož $|\lambda(m\,\pi)| \le 1 \; \forall \; m \in \mathbb{N}$, je dle (72) a (70)

$$|U_j^n| \le \sum_{m=-\infty}^{\infty} |A_m| \, |\lambda(m\,\pi)|^n \le \sum_{m=-\infty}^{\infty} |A_m| < \infty \, .$$

□

Poznámka 6 Nechť existuje $m_1 \in \mathbb{N}$ tak, že $|\lambda(m_1\,\pi)| > 1$. Buď $u^0(x) = \sin(m_1 \pi x)$. Pak podle (73) je $U_j^n = \sin(m_1 \pi j h) \, [\lambda(m_1 \pi)]^n$, a tudíž $|U_j^n| \to \infty$ pro $n \to \infty$ a pro každé $j \in \{1, \ldots, J-1\}$ takové, že $\sin(m_1 \pi j h) \neq 0$.

Nyní vidíme význam podmínky $\mu \le \frac{1}{2}$. Je-li tato podmínka splněna, pak $|\lambda(m\,\pi)| \le 1$ $\forall \; m \in \mathbb{N}$, a řešení tudíž zůstává omezené. Je-li $\mu > \frac{1}{2}$, pak pro některá $m \in \mathbb{N}$ je $\lambda(m\,\pi) < -1$ a velikost příslušných členů v (72) s postupujícím časem roste nade všechny meze. Teoreticky je možné zvolit počáteční podmínku tak, aby $A_m = 0$, kdykoli $\lambda(m\,\pi) < -1$. Avšak to je velmi speciální situace a vpraxi by vlivem zaokrouhlovacích chyb vznikly malé nenulové koeficienty u všech takovýchto členů a s postupujícím časem by tyto členy opět neomezeně rostly. Schéma (64) je tedy stabilní pouze při splnění podmínky $\mu \le \frac{1}{2}$.

Fourierovu metodu můžeme použít též k důkazu konvergence. Její výhoda spočívá v tom, že nemusíme předpokládat dostatečnou hladkost řešení u a stejnoměrnou omezenost u_{xxxx} a u_{tt}. Naším jediným předpokladem o úloze (59)–(61) nyní bude absolutní konvergence řady (69) pro počáteční podmínku u^0. Počáteční podmínka tedy nemusí být hladká. Budeme předpokládat, že μ je pevné a $\mu \le \frac{1}{2}$. Chybu aproximace můžeme vyjádřit ve tvaru

$$e_j^n = U_j^n - u(x_j, t_n) = \sum_{m=-\infty}^{\infty} A_m \, \mathrm{e}^{\mathrm{i}\, m\, \pi\, x_j} \left\{ [\lambda(m\,\pi)]^n - [\mathrm{e}^{-(m\,\pi)^2\,\tau}]^n \right\}.$$

Při označení $\lambda_1 = \lambda(m\,\pi)$ a $\lambda_2 = \mathrm{e}^{-(m\,\pi)^2\,\tau}$ je $\lambda_1, \lambda_2 \in [-1, 1]$ a platí

$$\lambda_1^n - \lambda_2^n = \sum_{k=0}^{n-1} \lambda_1^{n-k}\, \lambda_2^k - \sum_{k=1}^{n} \lambda_1^{n-k}\, \lambda_2^k = (\lambda_1 - \lambda_2) \sum_{k=0}^{n-1} \lambda_1^{n-1-k}\, \lambda_2^k\,,$$

z čehož plyne

$$|\lambda_1^n - \lambda_2^n| \le n\, |\lambda_1 - \lambda_2| \qquad \forall\ \lambda_1, \lambda_2 \in [-1, 1]\,. \tag{75}$$

Buď $\varepsilon > 0$ libovolné a nechť $m_0 \in \mathbb{N}$ je takové, že

$$\sum_{|m|>m_0} |A_m| \le \frac{\varepsilon}{4}\,.$$

Pak použitím (75) získáme

$$|e_j^n| \le \frac{\varepsilon}{2} + \sum_{|m|\le m_0} n\, |A_m|\, |\lambda(m\,\pi) - \mathrm{e}^{-(m\,\pi)^2\,\tau}|\,.$$

Z nerovnosti (74) plyne

$$|e_j^n| \le \frac{\varepsilon}{2} + n\, \tau^2\, C(\mu)\, \pi^4 \sum_{|m|\le m_0} |A_m|\, m^4$$

a vidíme tedy, že pro τ dostatečně malé je $|e_j^n| \le \varepsilon\ \forall\ (x_j, t_n) \in [0, 1] \times [0, T]$, neboť $n\,\tau \le T$.

Při aplikaci Fourierovy metody jsme reprezentovali přibližné řešení pomocí nekonečné řady (72), neboť ji bylo možné snadno srovnávat s řadou pro přesné řešení. Jelikož však na uvažované prostorové síti s $J+1$ uzly lze reprezentovat jen konečně mnoho různých frekvencí, lze přibližné řešení vyjádřit jako lineární kombinaci $2\,J$ po sobě jdoucích funkcí

$$\mathrm{e}^{\mathrm{i}\, m\, \pi\, j\, h}\, [\lambda(m\,\pi)]^n\,. \tag{76}$$

Snadno ověříme, že tyto funkce se skutečně nezmění, nahradíme-li m hodnotou $m + 2\,J$. Můžeme tedy např. uvažovat U_j^n jakožto lineární kombinaci funkcí (76) odpovídajících

$$m = -(J-1), -(J-2), \ldots, -1, 0, 1, \ldots, J\,.$$

4.5 Implicitní metoda pro úlohu (59)–(61)

Podmínka stability $\tau \le h^2/2$ pro explicitní schéma (64) je velmi vážné omezení, které implikuje, že bude třeba velmi mnoho časových kroků. Navíc, budeme-li muset zmenšit h pro zvýšení přesnosti, velmi se zvýší celková výpočetní náročnost, neboť budeme muset též podstatně zmenšit τ. Ukážeme nyní, že zmíněných omezení se můžeme zbavit použitím zpětné diference pro diskretizaci časové derivace v (59), tj. použitím schématu

$$\frac{U_j^{n+1} - U_j^n}{\tau} = \frac{U_{j+1}^{n+1} - 2\,U_j^{n+1} + U_{j-1}^{n+1}}{h^2}\,.$$

Toto schéma lze přepsat do tvaru

$$(77) \qquad -\mu\,U_{j-1}^{n+1} + (1 + 2\,\mu)\,U_j^{n+1} - \mu\,U_{j+1}^{n+1} = U_j^n\,, \qquad j = 1, 2, \ldots, J-1\,, \quad n \ge 0\,.$$

Na rozdíl od schématu (64) nyní danou hodnotu U_j^{n+1} nelze určit nezávisle na ostatních hodnotách na časové hladině t_{n+1}, nýbrž všechny tyto hodnoty je nutno vypočítat současně vyřešením soustavy $J-1$ lineárních rovnic pro $J-1$ neznámých. Jedná se tedy o implicitní metodu.

Stabilitu schématu (77) s okrajovými a počátečními podmínkami (65) a (66) můžeme vyšetřovat Fourierovou metodou analogicky jako pro schéma (64). Snadno zjistíme, že funkce $U_j^n = \mathrm{e}^{\mathrm{i}\,k\,j\,h}\,\lambda^n$ splňuje (77) právě tehdy, když

$$\lambda \equiv \lambda(k) = \frac{1}{1 + 4\,\mu\,\sin^2\frac{k\,h}{2}}\,.$$

Tedy $\lambda(k) \in (0, 1]$ pro libovolné $\mu > 0$ a libovolné $k \in \mathbb{R}$, což znamená, že metoda je nepodmíněně stabilní.

4.6 Thomasův algoritmus

Soustava (77) je tridiagonální a můžeme ji zapsat ve tvaru

$$-a_j\,U_{j-1} + b_j\,U_j - c_j\,U_{j+1} = d_j\,, \qquad j = 1, 2, \ldots, J-1\,,$$

kde

$$U_0 = U_J = 0\,.$$

Budeme předpokládat, že

$$(78) \qquad a_j > 0\,, \qquad b_j > 0\,, \qquad c_j > 0\,, \qquad b_j > a_j + c_j\,,$$

což splňuje (77). Soustavu rovnic nejprve převedeme na soustavu s horní trojúhelníkovou maticí tvaru

$$(79) \qquad U_j - e_j\,U_{j+1} = f_j\,, \qquad j = 1, 2, \ldots, J-1\,.$$

Máme-li v tomto tvaru k–tou rovnici a chceme-li upravit $k+1$–vou rovnici, tj.

$$-a_{k+1}\,U_k + b_{k+1}\,U_{k+1} - c_{k+1}\,U_{k+2} = d_{k+1}\,,$$

pak k této rovnici přičteme rovnici (79) s $j = k$ přenásobenou a_{k+1}, čímž získáme

$$(b_{k+1} - a_{k+1}\, e_k)\, U_{k+1} - c_{k+1}\, U_{k+2} = d_{k+1} + a_{k+1}\, f_k\,.$$

Algoritmus je tedy následující

$\mathrm{e_1 := c_1/b_1,\;\; f_1 := d_1/b_1}$
$\texttt{for}\ \mathrm{j = 2, \ldots, J-2}\ \texttt{do}$
$\quad \mathrm{e_j := c_j/(b_j - a_j\, e_{j-1})}$
$\quad \mathrm{f_j := (d_j + a_j\, f_{j-1})/(b_j - a_j\, e_{j-1})}$
$\texttt{enddo}$
$\mathrm{f_{J-1} := (d_{J-1} + a_{J-1}\, f_{J-2})/(b_{J-1} - a_{J-1}\, e_{J-2})}$

Řešení U_j pak jednoduše určíme z rovnic (79).

Snadno lze ověřit, že $e_j \in (0,1)$, $j = 1, 2, \ldots, J-2$. Algoritmus lze tedy vždy provést a je numericky stabilní (nevede ke vzrůstajícím chybám).

Uvedený algoritmus potřebuje pro vyřešení soustavy (77) na jeden uzel tři sčítání, tři násobení a dvě dělení, zatímco explicitní schéma (64) vyžaduje tři sčítání a dvě násobení (popř. čtyři sčítání a jedno násobení). Výpočetní náročnost implicitního schématu je tedy asi dvojnásobná oproti explicitnímu schématu. Důležité však je, že lze volit mnohem delší časové kroky (aniž by se zhoršila přesnost), neboť nyní není žádná podmínka stability omezující volbu τ. Proto je celková výpočetní náročnost implicitní metody pro dosažení zvoleného času T mnohem menší než u explicitní metody.

4.7 Dvoukrokové metody

Výše uvažované metody pro řešení úlohy (59)–(61) jsou druhého řádu přesnosti v prostoru, avšak obecně pouze prvního řádu přesnosti v čase. Nabízí se proto uvažovat pro diskretizaci časové derivace místo jednostranné diference centrální diferenci, podobně jako u leapfrog scheme (32). To vede v nejjednodušším případě ke schématu

$$(80)\qquad \frac{U_j^{n+1} - U_j^{n-1}}{2\,\tau} = \frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2}\,, \qquad j = 1, 2, \ldots, J-1\,, \quad n \geq 1\,.$$

Snadno zjistíme, že chyba diskretizace je v tomto případě druhého řádu v čase i v prostoru.

Uvedené schéma je příkladem dvoukrokového schématu, neboť k určení hodnot U_j^{n+1} nestačí znát hodnoty U_j^n, ale potřebujeme též hodnoty U_j^{n-1}. Přibližné řešení tedy závisí nejen na počáteční podmínce, ale též na hodnotách v čase t_1. Tyto hodnoty buď musíme předepsat, a nebo musíme stanovit postup, jak je určit. Obvykle se pro určení hodnot U_j^1 použije nějaké jednokrokové schéma. Metody, které zahrnují více než dvě časové hladiny, se souhrnně nazývají vícekroková schémata.

Hodnoty přibližného řešení v čase t_1 zapíšeme ve tvaru

$$(81)\qquad U_j^1 = \sum_{m=-\infty}^{\infty} B_m\, \mathrm{e}^{\mathrm{i}\, m\, \pi\, j\, h}\,, \qquad j = 0, 1, \ldots, J\,,$$