Studijní materiál k přednášce

Numerické řešení parciálních diferenciálních rovnic

Petr Knobloch

1 Úvod

V mnoha vědeckých a technických oblastech se setkáváme s matematickými modely sestávajícími z parciálních diferenciálních rovnic (PDR). Obvykle tyto modely nelze vyřešit analyticky a jejich řešení je potřeba aproximovat pomocí numerických metod. Za tím účelem bylo vyvinuto velké množství různých postupů. Cílem přednášky *Numerické řešení parciálních diferenciálních rovnic* je seznámit studenty s metodou konečných diferencí, k jejíž analýze postačují základní vědomosti z přednášek věnovaných matematické analýze. Budou vysvětleny hlavní myšlenky a prostředky analýzy metody konečných diferencí a studovány vlastnosti různých diskretizací pro základní příklady PDR (transportní rovnice, rovnice vedení tepla, rovnice difúze). Získané znalosti budou užitečné i pro získání představy o tom, s jakými problémy je nutno se vypořádávat při vývoji a aplikaci jiných numerických metod pro řešení PDR. Navíc je metoda konečných diferencí stále běžným prostředkem pro aproximaci časových derivací v PDR, zatímco diskretizace vzhledem k prostorovým proměnným je často získávána jinými přístupy (např. metodou konečných prvků). Jelikož však každá numerická metoda vede na soustavu algebraických rovnic, lze některé poznatky z této přednášky využít i při studiu vlastností diskretizací získaných jinými postupy.

1.1 Diferenční kvocienty

Metoda konečných diferencí je založena na tom, že se diferenciální rovnice uvažuje pouze v izolovaných bodech a derivace v těchto bodech se nahradí diferenčními kvocienty. Zmíněné body se získají jako uzly sítě pokrývající výpočetní oblast, a proto jiný název této numerické metody je metoda sítí.

Mějme funkci $v : \mathbb{R} \to \mathbb{R}$. Pak její první derivace v bodě $x \in \mathbb{R}$ je definována vztahem

$$v'(x) = \lim_{h \to 0} \frac{v(x+h) - v(x)}{h}\,. \tag{1.1}$$

Zvolíme-li $h > 0$, pak můžeme derivaci $v'(x)$ aproximovat vztahem

$$v'(x) \approx \frac{v(x+h) - v(x)}{h}\,. \tag{1.2}$$

Je-li funkce v dostatečně hladká, lze očekávat, že diferenční kvocient v (1.2) aproximuje $v'(x)$ tím lépe, čím je h menší. Zanedlouho uvidíme, že toto očekávání je správné.

Místo (1.1) můžeme též psát

$$v'(x) = \lim_{h\to 0} \frac{v(x) - v(x-h)}{h}\,, \tag{1.3}$$

a tedy i (zprůměrujeme-li (1.1) a (1.3))

$$v'(x) = \lim_{h\to 0} \frac{v(x+h) - v(x-h)}{2\,h}\,.$$

Tomu odpovídají aproximace

$$v'(x) \approx \frac{v(x) - v(x-h)}{h}\,, \qquad \text{resp.} \qquad v'(x) \approx \frac{v(x+h) - v(x-h)}{2\,h}\,,$$

popř.

$$v'(x) \approx \frac{v(x+\frac{h}{2}) - v(x-\frac{h}{2})}{h}\,.$$

Abychom zápis uvažovaných aproximací zpřehlednili, zavádíme pro diference následující označení (vždy předpokládáme, že $h > 0$):

dopředná diference:	$\Delta_+ v(x) = v(x+h) - v(x),$
zpětná diference:	$\Delta_- v(x) = v(x) - v(x-h),$
centrální diference s dvojnásobnou délkou intervalu:	$\Delta_0 v(x) = \frac{1}{2}\,[v(x+h) - v(x-h)],$
centrální diference:	$\delta v(x) = v(x+\frac{h}{2}) - v(x-\frac{h}{2}).$

Pak uvedené aproximace můžeme psát ve tvaru

$$v' \approx \frac{\Delta_+ v}{h}\,, \qquad \text{resp.} \qquad v' \approx \frac{\Delta_- v}{h}\,, \qquad \text{resp.} \qquad v' \approx \frac{\Delta_0 v}{h}\,, \qquad \text{resp.} \qquad v' \approx \frac{\delta v}{h}\,.$$

Je-li $v \in C^2(\mathbb{R})$, zjišťujeme pomocí Taylorova vzorce, že všechny tyto aproximace vedou k chybě $O(h)$ (tj. velikost chyby lze v každém bodě x omezit hodnotou Ch s konstantou C nezávislou na h). Je-li $v \in C^3(\mathbb{R})$, dostáváme navíc

$$\frac{\Delta_0 v}{h} = v' + O(h^2)\,, \qquad \frac{\delta v}{h} = v' + O(h^2)\,.$$

Chceme-li aproximovat $v''(x)$, pak můžeme postupovat následovně:

$$v''(x) \approx \frac{\delta\, v'(x)}{h} \approx \frac{\delta^2\, v(x)}{h^2}\,.$$

Platí

$$\delta^2\, v(x) = v(x+h) - 2\,v(x) + v(x-h)$$

a tento výraz se nazývá centrální diference druhého řádu. Snadno ověříme, že

$$\delta^2 = \Delta_+\,\Delta_- = \Delta_-\,\Delta_+\,.$$

Použitím Taylorova vzorce dostaneme pro $v \in C^4(\mathbb{R})$

$$\frac{\delta^2 v}{h^2} = v'' + O(h^2)\,.$$

Podobně pro $v \in C^6(\mathbb{R})$ je

$$\frac{\delta^4 v}{h^4} = v^{(4)} + O(h^2)\,.$$

Přitom

$$\delta^4 v(x) = v(x+2h) - 4v(x+h) + 6v(x) - 4v(x-h) + v(x-2h)\,.$$

1.2 Příklady diskretizací Cauchyovy úlohy pro rovnici konvekce–difúze v jedné prostorové dimenzi

Ukažme si nyní, jak pomocí metody konečných diferencí lze získat diskretizaci Cauchyovy úlohy pro danou parciální diferenciální rovnici. Jako příklad budeme uvažovat rovnici konvekce–difúze v jedné prostorové dimenzi. Hledáme tedy funkci $u = u(x,t)$ závisející na prostorové proměnné $x \in \mathbb{R}$ a časové proměnné $t \in \mathbb{R}_0^+$ splňující parciální diferenciální rovnici

$$u_t = b\,u_{xx} - a\,u_x \qquad \text{v } \mathbb{R} \times \mathbb{R}^+ \tag{1.4}$$

s počáteční podmínkou

$$u(x,0) = u^0(x) \qquad \forall\, x \in \mathbb{R}\,, \tag{1.5}$$

kde u^0 je daná funkce. Parametr b představuje koeficient difúze, a proto předpokládáme, že $b > 0$. Parametr a určuje rychlost šíření veličiny u vlivem konvekce (unášení). Pro jednoduchost předpokládáme, že a a b jsou konstanty.

Jak již bylo zmíněno výše, základem diskretizace metodou konečných diferencí je síť pokrývající výpočetní oblast, v jejíchž uzlech aproximujeme parciální diferenciální rovnici rovnicemi algebraickými. Pro jednoduchost budeme nyní uvažovat rovnoměrnou síť s konstantním prostorovým krokem $h > 0$ a konstantním časovým krokem $\tau > 0$. Definujeme prostorové uzly $x_j = j\,h$, $j \in \mathbb{Z}$, a časové hladiny $t_n = n\,\tau$, $n \in \mathbb{N}_0$. Rovnoběžky se souřadnými osami protínající tyto osy v bodech x_j a t_n jsou takzvané síťové přímky, které vytvářejí síť pokrývající výpočetní oblast $\mathbb{R} \times \mathbb{R}_0^+$. Průsečíky síťových přímek (x_j, t_n), kde $j \in \mathbb{Z}$ a $n \in \mathbb{N}_0$, nazveme uzly sítě. Řešení u Cauchyovy úlohy (1.4), (1.5) aproximujeme v uzlech (x_j, t_n) hodnotami U_j^n. Síťová funkce $\{U_j^n\}$ je pak přibližným řešením uvažované Cauchyovy úlohy. Pro přehlednost zavádíme též označení $u_j^n := u(x_j, t_n)$.

Přibližné řešení $\{U_j^n\}$ získáme tak, že parciální diferenciální rovnici v uzlech sítě nahradíme diferenčním schématem. Uvažujme libovolný uzel (x_j, t_n) s $j \in \mathbb{Z}$ a $n \in \mathbb{N}_0$. Pak rovnici (1.4) můžeme s využitím vztahů z předešlého oddílu aproximovat například

následovně:

$$\begin{aligned}
0 &= (u_t - b\,u_{xx} + a\,u_x)(x_j, t_n) \\
&\approx \frac{u(x_j, t_n+\tau) - u(x_j, t_n)}{\tau} - b\,\frac{u(x_j+h, t_n) - 2\,u(x_j, t_n) + u(x_j-h, t_n)}{h^2} \\
&\quad + a\,\frac{u(x_j+h, t_n) - u(x_j-h, t_n)}{2\,h} \\
&= \frac{u_j^{n+1} - u_j^n}{\tau} - b\,\frac{u_{j+1}^n - 2\,u_j^n + u_{j-1}^n}{h^2} + a\,\frac{u_{j+1}^n - u_{j-1}^n}{2\,h} \\
&\approx \frac{U_j^{n+1} - U_j^n}{\tau} - b\,\frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2} + a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h}\,.
\end{aligned}$$

Použijeme-li symboly pro diference, můžeme uvedené aproximace zapsat v přehlednějším tvaru:

$$\begin{aligned}
0 = (u_t - b\,u_{xx} + a\,u_x)(x_j, t_n) &\approx \frac{\Delta_{+t}\,u_j^n}{\tau} - b\,\frac{\delta_x^2\,u_j^n}{h^2} + a\,\frac{\Delta_{0x}\,u_j^n}{h} \\
&\approx \frac{\Delta_{+t}\,U_j^n}{\tau} - b\,\frac{\delta_x^2\,U_j^n}{h^2} + a\,\frac{\Delta_{0x}\,U_j^n}{h}\,.
\end{aligned}$$

Nyní u symbolů pro diference vyznačujeme indexem x či t, vzhledem ke které proměnné je diference definována. Např.

$$\Delta_{+t}\,u(x,t) = u(x, t+\tau) - u(x,t)\,, \qquad \Delta_{+x}\,u(x,t) = u(x+h, t) - u(x,t)\,.$$

Rovnici (1.4) tedy aproximujeme numerickým schématem

$$\frac{U_j^{n+1} - U_j^n}{\tau} = b\,\frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2} - a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} \qquad \forall\ j \in \mathbb{Z},\ n \in \mathbb{N}_0\,, \tag{1.6}$$

které je nutno doplnit počáteční podmínkou

$$U_j^0 = u^0(x_j) \qquad \forall\ j \in \mathbb{Z}\,. \tag{1.7}$$

Jedná se o příklad tzv. explicitního schématu, neboť hodnotu přibližného řešení v libovolném uzlu (x_j, t_n) s $j \in \mathbb{Z}$ a $n \in \mathbb{N}$ lze ze schématu vyjádřit pomocí hodnot přibližného řešení na předcházející časové vrstvě t_{n-1}. Přibližné řešení tedy existuje a je určeno jednoznačně. Označíme-li

$$\mu = \frac{\tau}{h^2}\,b\,, \qquad \nu = \frac{\tau}{h}\,a\,, \tag{1.8}$$

můžeme schéma (1.6) psát ve tvaru

$$U_j^{n+1} = U_j^n + \mu\,(U_{j+1}^n - 2\,U_j^n + U_{j-1}^n) - \frac{\nu}{2}\,(U_{j+1}^n - U_{j-1}^n)\,, \tag{1.9}$$

který je v některých případech vhodnější pro analýzu.

Jelikož existuje řada způsobů, jak pomocí diferenčních kvocientů aproximovat derivace, získáme různými kombinacemi diferenčních kvocientů mnoho rozličných numerických schémat pro aproximaci dané parciální diferenciální rovnice. Jak uvidíme později, takto získaná schémata mohou mít velmi odlišné vlastnosti. Ukažme si ještě dvě schémata, která lze definovat pro rovnici (1.4). Uvažujeme-li tuto rovnici v uzlu (x_j, t_{n+1}) s $j \in \mathbb{Z}$ a $n \in \mathbb{N}_0$, můžeme provést následující aproximace:

$$\begin{aligned} 0 = (u_t - b\,u_{xx} + a\,u_x)(x_j, t_{n+1}) &\approx \frac{\Delta_{-t}\,u_j^{n+1}}{\tau} - b\,\frac{\delta_x^2\,u_j^{n+1}}{h^2} + a\,\frac{\Delta_{0x}\,u_j^{n+1}}{h} \\ &\approx \frac{\Delta_{-t}\,U_j^{n+1}}{\tau} - b\,\frac{\delta_x^2\,U_j^{n+1}}{h^2} + a\,\frac{\Delta_{0x}\,U_j^{n+1}}{h}\,, \end{aligned}$$

což vede ke schématu

$$\frac{U_j^{n+1} - U_j^n}{\tau} = b\,\frac{U_{j+1}^{n+1} - 2\,U_j^{n+1} + U_{j-1}^{n+1}}{h^2} - a\,\frac{U_{j+1}^{n+1} - U_{j-1}^{n+1}}{2\,h} \qquad \forall\, j \in \mathbb{Z},\, n \in \mathbb{N}_0\,, \tag{1.10}$$

k němuž opět musíme přidat počáteční podmínku (1.7). Nyní již v daném uzlu není možné jednoduchým způsobem vyjádřit hodnotu přibližného řešení pomocí hodnot z předchozí časové hladiny a hodnoty přibližného řešení na nové časové hladině je nutno získat vyřešením soustavy lineárních rovnic. Jedná se o příklad tzv. implicitního schématu.

Sečteme-li schémata (1.6) a (1.10), získáme po vydělení dvěma schéma

$$\frac{U_j^{n+1} - U_j^n}{\tau} = b\,\frac{\delta_x^2\,U_j^{n+1} + \delta_x^2\,U_j^n}{2\,h^2} - a\,\frac{U_{j+1}^{n+1} - U_{j-1}^{n+1} + U_{j+1}^n - U_{j-1}^n}{4\,h} \qquad \forall\, j \in \mathbb{Z},\, n \in \mathbb{N}_0\,.$$

Toto schéma je opět implicitní a nazývá se schéma Crankovo–Nicolsonové. Jelikož hodnoty přibližného řešení, které toto schéma kombinuje, vykazují symetrické uspořádání nejen vzhledem k x_j, ale i vzhledem k $t_{n+1/2} := t_n + \tau/2$, aproximuje toto schéma parciální diferenciální rovnici (1.4) při $h, \tau \to 0$ s menší chybou než předchozí dvě schémata.

1.3 Chyba diskretizace

Chyba diskretizace je chyba, které se dopouštíme, když derivace v diferenciální rovnici nahradíme diferenčními kvocienty. Získáme ji dosazením přesného řešení do numerického schématu ve tvaru odpovídajícím nahrazení derivací diferenčními kvocienty a odečtením pravé strany od levé.

Jako příklad uvažujme schéma (1.6) pro numerické řešení rovnice (1.4). Pro definici chyby diskretizace použijeme tvar schématu s diferenčními kvocienty (tj. (1.6)), nikoli ekvivalentní tvar (1.9). Chyba diskretizace je pak dána vztahem

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

kde $u_j^n = u(x_j, t_n)$. Tento vztah můžeme též psát ve tvaru

$$\varepsilon_j^n = \frac{\Delta_{+t}\,u_j^n}{\tau} - b\,\frac{\delta_x^2\,u_j^n}{h^2} + a\,\frac{\Delta_{0x}\,u_j^n}{h}\,.$$

Můžeme také zavést chybu diskretizace $\varepsilon_{h,\tau}(x,t)$, která je definována v každém bodě (x,t) výpočetní oblasti, pro nějž $(x \pm h, t)$ a $(x, t+\tau)$ též leží ve výpočetní oblasti, vztahem

$$\varepsilon_{h,\tau} = \frac{\Delta_{+t}\, u}{\tau} - b\,\frac{\delta_x^2\, u}{h^2} + a\,\frac{\Delta_{0x}\, u}{h}\,.$$

Zřejmě $\varepsilon_j^n = \varepsilon_{h,\tau}(x_j, t_n)$. Je-li u dvakrát spojitě diferencovatelné podle t a čtyřikrát spojitě diferencovatelné podle x, pak použitím Taylorova vzorce získáme

$$\varepsilon_{h,\tau}(x,t) = \frac{1}{2}\,u_{tt}(x,\eta)\,\tau - \frac{b}{12}\,u_{xxxx}(\xi,t)\,h^2 + \frac{a}{6}\,u_{xxx}(\zeta,t)\,h^2\,, \tag{1.11}$$

kde $\eta \in (t, t+\tau)$ a $\xi, \zeta \in (x-h, x+h)$. Říkáme, že schéma je prvního řádu přesnosti v čase a druhého řádu přesnosti v prostoru.

U každého schématu je základním požadavkem, aby chyba diskretizace konvergovala k nule pro $h, \tau \to 0$. Pak říkáme, že schéma je konzistentní s řešenou diferenciální rovnicí.

1.4 Chyba aproximace

Chyba aproximace je definována vztahem $e_j^n = U_j^n - u_j^n$, kde U_j^n je přibližné řešení a u_j^n je aproximované přesné řešení v uzlu (x_j, t_n).

Všimněme si, že chyba aproximace je řešením příslušného schématu, v němž se na pravé straně objeví dodatečný člen $-\varepsilon_j^n$ nebo násobek této hodnoty (v závislosti na použitém tvaru schématu). Např. v případě schématu (1.6) máme

$$\frac{e_j^{n+1} - e_j^n}{\tau} = b\,\frac{e_{j+1}^n - 2\,e_j^n + e_{j-1}^n}{h^2} - a\,\frac{e_{j+1}^n - e_{j-1}^n}{2\,h} - \varepsilon_j^n\,,$$

resp.

$$e_j^{n+1} = e_j^n + \mu\,(e_{j+1}^n - 2\,e_j^n + e_{j-1}^n) - \frac{\nu}{2}\,(e_{j+1}^n - e_{j-1}^n) - \tau\,\varepsilon_j^n\,.$$

Později uvidíme, že na základě vztahů tohoto typu lze za určitých podmínek chybu aproximace odhadnout pomocí odhadu chyby diskretizace.

2 Stabilita jednokrokových numerických schémat pro Cauchyovy úlohy

V této části budeme uvažovat Cauchyovu úlohu najít funkci $u = u(x,t)$ definovanou pro $x \in \mathbb{R}$ a $t \geq 0$, která splňuje

$$u_t + L\,u = 0 \quad \text{v } \mathbb{R} \times \mathbb{R}^+\,, \qquad u(x,0) = u^0(x) \quad \text{pro } x \in \mathbb{R}\,, \tag{2.1}$$

kde u^0 je zadaná počáteční podmínka a L je lineární diferenciální operátor s konstantními koeficienty obsahující pouze derivace podle x, tj. operátor tvaru

$$L = \sum_{k=0}^{m} a_k\,\frac{\partial^k}{\partial x^k}\,.$$

Obecné jednokrokové schéma pro úlohu (2.1) na stejnoměrné síti s uzly (x_j, t_n), kde $x_j = j\,h$, $t_n = n\,\tau$, $j \in \mathbb{Z}$ a $n \in \mathbb{N}_0$, má tvar

$$\sum_{s=-M}^{M} \alpha_s\, U_{j+s}^{n+1} = \sum_{s=-M}^{M} \beta_s\, U_{j+s}^{n} \qquad \forall\ j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,, \tag{2.2}$$

$$U_j^0 = u^0(x_j) \qquad \forall\ j \in \mathbb{Z}\,, \tag{2.3}$$

kde přirozené číslo M závisí na způsobu aproximace derivací podle x v operátoru L. Předpokládáme, že pokud je přibližné řešení omezené pro každé $n \in \mathbb{N}_0$, pak je uvedeným schématem určeno jednoznačně.

Pro schéma (2.2), (2.3) budeme analyzovat dva typy stability. Nejprve se budeme zabývat tzv. von Neumannovou analýzou stability, která je založena na Fourierově transformaci a udává podmínky pro stabilitu vzhledem k L^2 normě. Následně zformulujeme podmínky, za nichž uvažované numerické schéma splňuje diskrétní princip maxima a je tudíž stabilní vzhledem k maximové normě. Uvidíme též, jak princip maxima umožňuje odhadnout chybu aproximace pomocí chyby diskretizace.

2.1 Von Neumannova analýza stability

V teorii parciálních diferenciálních rovnic hraje důležitou roli Fourierova transformace. Pro funkci $u \in L^1(\mathbb{R})$ definujeme Fourierovu transformaci $\widehat{u}$ vztahem

$$\widehat{u}(\xi) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} u(x)\, \mathrm{e}^{-\mathrm{i}\,x\,\xi}\, \mathrm{d}x\,, \qquad \xi \in \mathbb{R}\,. \tag{2.4}$$

Za určitých předpokladů pak platí

$$u(x) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} \widehat{u}(\xi)\, \mathrm{e}^{\mathrm{i}\,x\,\xi}\, \mathrm{d}\xi \tag{2.5}$$

a je splněna Parsevalova rovnost

$$\|u\|_{L^2(\mathbb{R})} = \|\widehat{u}\|_{L^2(\mathbb{R})}\,. \tag{2.6}$$

Pravá strana vztahu (2.5) je inverzní Fourierova transformace. Vztah (2.5) vyjadřuje funkci u jako superpozici vln daných funkcemi $\mathrm{e}^{\mathrm{i}\,x\,\xi}$ s různými amplitudami $\widehat{u}(\xi)$. Funkce $\widehat{u}$ představuje alternativní reprezentaci funkce u a může být komplexní, i když je funkce u reálná.

Podobně jako výše lze postupovat též v diskrétním případě. Buď $l \in \mathbb{R}^+$ a označme

$$\varphi_j(\xi) = \frac{1}{\sqrt{l}}\, \mathrm{e}^{-\mathrm{i}\,(2\,\pi\,j/l)\,\xi}\,, \qquad \xi \in \mathbb{R}\,, \quad j \in \mathbb{Z}\,.$$

Pak pro libovolné reálné číslo a tvoří množina $\{\varphi_j\}_{j\in\mathbb{Z}}$ úplný ortonormální systém v prostoru $L^2(a, a+l)$. Je-li dána posloupnost $U = \{U_j\}_{j\in\mathbb{Z}}$ splňující $\sum_{j\in\mathbb{Z}} |U_j|^2 < \infty$, pak řada $\sum_{j\in\mathbb{Z}} U_j\,\varphi_j$ konverguje v prostoru $L^2(a, a+l)$ k funkci $\widetilde{U}$ a platí $U_j = \int_a^{a+l} \widetilde{U}\,\overline{\varphi_j}\,\mathrm{d}\xi$.

Navíc je splněna Parsevalova rovnost $\sum_{j\in\mathbb{Z}} |U_j|^2 = \|\widetilde{U}\|^2_{L^2(a,a+l)}$. Předpokládejme nyní, že hodnoty U_j představují hodnoty síťové funkce v uzlech $x_j = j\,h$. Zvolme $a = -l/2$, $l = 2\,\pi/h$ a definujme funkci $\widehat{U} \in L^2(-\pi/h, \pi/h)$ vztahem

$$\widehat{U}(\xi) = \frac{1}{\sqrt{2\,\pi}} \sum_{j=-\infty}^{\infty} h\,U_j\,\mathrm{e}^{-\mathrm{i}\,x_j\,\xi} \tag{2.7}$$

(tj. $\widehat{U} = \sqrt{h}\,\widetilde{U}$). Pak

$$U_j = \frac{1}{\sqrt{2\,\pi}} \int_{-\pi/h}^{\pi/h} \widehat{U}(\xi)\,\mathrm{e}^{\mathrm{i}\,x_j\,\xi}\,\mathrm{d}\xi \qquad \forall\ j \in \mathbb{Z}\,. \tag{2.8}$$

Vztahy (2.7) a (2.8) jsou diskrétními analogiemi vztahů (2.4) a (2.5). Vztah (2.8) opět vyjadřuje U jako superpozici vln. Označíme-li

$$\|U\|_2 = \sqrt{\sum_{j=-\infty}^{\infty} h\,|U_j|^2}$$

diskrétní analogii normy v prostoru $L^2(\mathbb{R})$, platí podobně jako v (2.6) Parsevalova rovnost

$$\|U\|_2 = \|\widehat{U}\|_{L^2(-\pi/h,\pi/h)}\,. \tag{2.9}$$

Uvažujme nyní obecné jednokrokové schéma (2.2), (2.3) pro numerické řešení Cauchyovy úlohy (2.1). Zapíšeme-li přibližné řešení $U^n = \{U_j^n\}_{j\in\mathbb{Z}}$ ve tvaru (2.8) pomocí Fourierovy transformace $\widehat{U}^n = \widehat{U}^n(\xi)$ definované vztahem (2.7), získáme dosazením do (2.2)

$$\int_{-\pi/h}^{\pi/h} \mathrm{e}^{\mathrm{i}\,x_j\,\xi} \left(\widehat{U}^{n+1}(\xi) \sum_{s=-M}^{M} \alpha_s\,\mathrm{e}^{\mathrm{i}\,s\,h\,\xi} - \widehat{U}^n(\xi) \sum_{s=-M}^{M} \beta_s\,\mathrm{e}^{\mathrm{i}\,s\,h\,\xi} \right) \mathrm{d}\xi = 0 \quad \forall\ j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,.$$

Jelikož funkce $\{\mathrm{e}^{\mathrm{i}\,x_j\,\xi}\}_{j\in\mathbb{Z}}$ tvoří úplný ortogonální systém v prostoru $L^2(-\pi/h, \pi/h)$, platí

$$\widehat{U}^{n+1}(\xi) \sum_{s=-M}^{M} \alpha_s\,\mathrm{e}^{\mathrm{i}\,s\,h\,\xi} = \widehat{U}^n(\xi) \sum_{s=-M}^{M} \beta_s\,\mathrm{e}^{\mathrm{i}\,s\,h\,\xi} \qquad \forall\ n \in \mathbb{N}_0\,. \tag{2.10}$$

Všimněme si, že tento vztah lze formálně získat tak, že diskrétní řešení U_j^n nahradíme v (2.2) výrazem $\widehat{U}^n(\xi)\,\mathrm{e}^{\mathrm{i}\,x_j\,\xi}$. Díky předpokladu o jednoznačné řešitelnosti schématu je součet na levé straně pro libovolné ξ nenulový (viz cvičení), a můžeme proto definovat funkci

$$\lambda(\xi) = \frac{\displaystyle\sum_{s=-M}^{M} \beta_s\,\mathrm{e}^{\mathrm{i}\,s\,h\,\xi}}{\displaystyle\sum_{s=-M}^{M} \alpha_s\,\mathrm{e}^{\mathrm{i}\,s\,h\,\xi}}\,.$$

Pak obdržíme

$$\widehat{U}^{n+1}(\xi) = \lambda(\xi)\,\widehat{U}^n(\xi) \qquad \forall\ n \in \mathbb{N}_0\,. \tag{2.11}$$

Tento vztah ukazuje, že provedení jednoho časového kroku schématu (2.2) je ekvivalentní přenásobení Fourierovy transformace diskrétního řešení amplifikačním faktorem $\lambda(\xi)$. Velikost amplifikačního faktoru $|\lambda(\xi)|$ představuje zesílení amplitudy $\widehat{U}^n(\xi)$ libovolné frekvence ξ při provedení jednoho časového kroku. Ze vztahu (2.11) získáme

$$\widehat{U}^n(\xi) = \lambda(\xi)^n\,\widehat{U}^0(\xi) \qquad \forall\ n \in \mathbb{N}_0\,. \tag{2.12}$$

Vidíme, že všechny informace o schématu jsou obsaženy v amplifikačním faktoru. Mimo jiné lze z amplifikačního faktoru snadno získat informace o stabilitě a přesnosti příslušného schématu. Fourierova transformace proto představuje standardní metodu pro studium vlastností diferenčních schémat.

Pro praktické výpočty amplifikačního faktoru můžeme využít výše uvedeného pozorování, že vztah (2.10) lze formálně získat nahrazením U_j^n v (2.2) výrazem $\widehat{U}^n(\xi)\,\mathrm{e}^{\mathrm{i}\,x_j\,\xi}$. Použijeme-li dále vztah (2.12), vidíme, že pro výpočet $\lambda(\xi)$ stačí do (2.2) dosadit místo U_j^n výraz $\lambda(\xi)^n\,\mathrm{e}^{\mathrm{i}\,x_j\,\xi}$ a následně získanou rovnost vydělit $\lambda(\xi)^n$. Takto budeme v budoucnu při výpočtu amplifikačního faktoru postupovat.

Diferenční schémata jsou často pouze podmíněně stabilní, což znamená, že jsou stabilní, jen pokud prostorový krok h a časový krok τ splňují jistou podmínku. Množinu $\Lambda \subset \mathbb{R}^+ \times \mathbb{R}^+$ takovou, že pro libovolnou dvojici $(h,\tau) \in \Lambda$ je příslušná podmínka stability splněna, nazveme oblastí stability diferenčního schématu. Vždy budeme předpokládat, že množina Λ je omezená a že dvojice $(0,0)$ je jejím hromadným bodem. Stabilitu schématu (2.2) lze definovat například následujícím způsobem.

Definice 2.1 Diferenční schéma (2.2) je stabilní v oblasti stability Λ, pokud pro každý pevný čas $T > 0$ existuje konstanta C_T taková, že pro libovolnou počáteční podmínku U^0 platí

$$\|U^n\|_2 \le C_T\,\|U^0\|_2 \qquad \forall\ n \in \mathbb{N},\ (h,\tau) \in \Lambda,\ n\tau \le T\,.$$

Věta 2.1 *Diferenční schéma* (2.2) *je stabilní v oblasti stability* Λ *právě tehdy, když existuje konstanta* K *nezávislá na* ξ, h *a* τ *taková, že*

$$|\lambda(\xi)| \le 1 + K\,\tau \qquad \forall\ \xi \in \mathbb{R}\,,\ (h,\tau) \in \Lambda\,. \tag{2.13}$$

Je-li funkce $\lambda(\xi/h)$ *na* Λ *nezávislá na* h *a* τ, *pak lze podmínku* (2.13) *nahradit podmínkou*

$$|\lambda(\xi)| \le 1 \qquad \forall\ \xi \in \mathbb{R}\,,\ (h,\tau) \in \Lambda\,. \tag{2.14}$$

Důkaz. Z Parsevalovy rovnosti (2.9) a vztahu (2.12) plyne

$$\|U^n\|_2 = \|\lambda^n\,\widehat{U}^0\|_{L^2(-\pi/h,\pi/h)}\,. \tag{2.15}$$

Poznamenejme, že λ^n zde značí n–tou mocninu λ. Platí-li (2.13), je pro $(h,\tau) \in \Lambda$ a $n \le T/\tau$

$$\|U^n\|_2 \le (1 + K\,\tau)^n\,\|\widehat{U}^0\|_{L^2(-\pi/h,\pi/h)} \le (1 + K\,\tau)^{T/\tau}\,\|U^0\|_2 \le \mathrm{e}^{K\,T}\,\|U^0\|_2\,,$$

tj. schéma (2.2) je stabilní v Λ. Předpokládejme nyní, že (2.13) neplatí pro žádnou konstantu K. Zvolme libovolné číslo $C > 0$. Jelikož funkce λ závisí na ξ spojitě a periodicky s periodou $2\,\pi/h$, existuje $(h,\tau) \in \Lambda$ a interval $(\xi_1,\xi_2) \subset (-\pi/h, \pi/h)$ tak, že $|\lambda(\xi)| > 1 + C\,\tau$ pro všechna $\xi \in (\xi_1,\xi_2)$. Nechť

$$\widehat{U}^0(\xi) = \frac{1}{\sqrt{\xi_2 - \xi_1}} \quad \text{pro } \xi \in (\xi_1,\xi_2) \qquad \text{a} \qquad \widehat{U}^0(\xi) = 0 \quad \text{pro } \xi \notin (\xi_1,\xi_2)\,.$$

Pak dle (2.9) a (2.15) je $\|U^0\|_2 = \|\widehat{U}^0\|_{L^2(-\pi/h,\pi/h)} = 1$ a

$$\|U^n\|_2 = \|\lambda^n\,\widehat{U}^0\|_{L^2(\xi_1,\xi_2)} > (1 + C\,\tau)^n \geq (1 + C\,\tau_{max})^{n\tau/\tau_{max}}\,\|U^0\|_2\,,$$

kde τ_{max} je takové, že $\Lambda \subset \mathbb{R}^+ \times (0,\tau_{max})$ (využili jsme, že funkce $(1+C\,\tau)^{1/\tau}$ je klesající). Pro libovolné $T > \tau_{max}$ a $n \in \mathbb{N}$ splňující $T/2 \leq n\tau \leq T$ je

$$\|U^n\|_2 > (1 + C\,\tau_{max})^{T/(2\,\tau_{max})}\,\|U^0\|_2\,.$$

Schéma (2.2) tedy není stabilní v Λ, neboť C lze volit libovolně velké. Je-li funkce $\lambda(\xi/h)$ na Λ nezávislá na h a τ, pak jsou zřejmě podmínky (2.13) a (2.14) ekvivalentní. □

Poznámka 2.1 Uvedená věta pochází od von Neumanna, a analýza diferenčních metod založená na Fourierově metodě se proto obvykle nazývá von Neumannova analýza. Nerovnost (2.13) se většinou nazývá von Neumannova podmínka.

Uvažujme rovnici (1.4), přičemž opět předpokládáme, že $b > 0$ a a jsou konstanty, a ukažme si provedení analýzy stability na příkladu explicitního schématu (1.6). Jak jsme viděli výše, pro výpočet amplifikačního faktoru $\lambda(\xi)$ stačí do schématu místo U_j^n dosadit výraz $\lambda(\xi)^n\,\mathrm{e}^{\mathrm{i}\,x_j\,\xi} = \lambda(\xi)^n\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}$ a následně získanou rovnost vydělit $\lambda(\xi)^n$. Použitím ekvivalentního tvaru (1.9) tak získáme

$$\lambda(\xi)\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h} = \mathrm{e}^{\mathrm{i}\,\xi\,j\,h} + \mu\left(\mathrm{e}^{\mathrm{i}\,\xi\,(j+1)\,h} - 2\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h} + \mathrm{e}^{\mathrm{i}\,\xi\,(j-1)\,h}\right) - \frac{\nu}{2}\left(\mathrm{e}^{\mathrm{i}\,\xi\,(j+1)\,h} - \mathrm{e}^{\mathrm{i}\,\xi\,(j-1)\,h}\right),$$

z čehož plyne

$$\begin{aligned}\lambda(\xi) &= 1 + \mu\left(\mathrm{e}^{\mathrm{i}\,\xi\,h} - 2 + \mathrm{e}^{-\mathrm{i}\,\xi\,h}\right) - \frac{\nu}{2}\left(\mathrm{e}^{\mathrm{i}\,\xi\,h} - \mathrm{e}^{-\mathrm{i}\,\xi\,h}\right)\\ &= 1 + 2\,\mu\,(\cos(\xi\,h) - 1) - \mathrm{i}\,\nu\,\sin(\xi\,h)\,.\end{aligned}$$

Použitím vztahů $\cos(\xi\,h) = \cos^2\frac{\xi\,h}{2} - \sin^2\frac{\xi\,h}{2}$ a $1 = \cos^2\frac{\xi\,h}{2} + \sin^2\frac{\xi\,h}{2}$ zjišťujeme, že amplifikační faktor je dán vztahem

$$\lambda(\xi) = 1 - 4\,\mu\,\sin^2\frac{\xi\,h}{2} - \mathrm{i}\,\nu\,\sin(\xi\,h)\,.$$

Pro nalezení nutné a postačující podmínky pro stabilitu v diskrétní L^2 normě je tedy třeba uvažovat podmínku (2.13). Snadno vypočítáme, že

$$|\lambda(\xi)|^2 = (1 - 4\,\mu\,s^2)^2 + 4\,\nu^2\,s^2\,(1 - s^2)\,, \quad \text{kde} \quad s = \sin\frac{\xi\,h}{2}\,.$$

Při $s^2 = 1$ stačí požadovat, aby $\mu \leq \frac{1}{2}$, neboť pak $|\lambda(\xi)| \leq 1$. Jelikož $\nu^2 = a^2\,\mu\,\tau/b$ a $4\,s^2\,(1-s^2) \leq 1$, dostáváme při $\mu \leq \frac{1}{2}$ nerovnost

$$|\lambda(\xi)| \leq \left(1 + \frac{a^2}{2\,b}\,\tau\right)^{1/2} \leq 1 + \frac{a^2}{4\,b}\,\tau\,.$$

Pro $\mu \leq \frac{1}{2}$ je tedy von Neumannova podmínka splněna a schéma je stabilní v diskrétní L^2 normě. Nicméně při $\mu \leq \frac{1}{2}$ může pro některé hodnoty ξ být $|\lambda(\xi)| > 1$, což vede k růstu příslušné amplitudy v diskrétním řešení, zatímco v řešení diferenciální rovnice jsou všechny amplitudy tlumeny. Pokud tedy exponenciální růst v čase, který von Neumannova podmínka umožňuje, neodpovídá vlastnostem aproximované parciální diferenciální rovnice, je von Neumannova podmínka v praxi příliš slabá. Zavádíme proto následující silnější definici stability.

Definice 2.2 Diferenční schéma se nazývá *silně stabilní*, jestliže platí: pokud Fourierova transformace řešení aproximované parciální diferenciální rovnice splňuje

$$|\widehat{u}(\xi, t+\tau)| \leq \mathrm{e}^{\alpha\,\tau}\,|\widehat{u}(\xi, t)| \qquad \forall\ \xi \in \mathbb{R} \tag{2.16}$$

pro nějaké $\alpha \geq 0$, pak amplifikační faktory diferenčního schématu splňují

$$|\lambda(\xi)| \leq \mathrm{e}^{\alpha\,\tau} \qquad \forall\ \xi \in \mathbb{R}\,.$$

Pro rovnici (1.4) platí (2.16) s $\alpha = 0$, a tedy požadujeme, aby $|\lambda(\xi)| \leq 1$ pro všechna $\xi \in \mathbb{R}$. Jelikož

$$|\lambda(\xi)|^2 = 1 - 4\,(2\,\mu - \nu^2)\,s^2 + 4\,(4\,\mu^2 - \nu^2)\,s^4\,,$$

dostáváme

$$|\lambda(\xi)| \leq 1 \ \ \forall\ \xi \in \mathbb{R} \quad \Leftrightarrow \quad (4\,\mu^2 - \nu^2)\,s^2 \leq 2\,\mu - \nu^2 \ \ \forall\ s \in [-1, 1] \quad \Leftrightarrow \quad \nu^2 \leq 2\,\mu \leq 1\,.$$

Kromě očekávané podmínky $\mu \leq \frac{1}{2}$ jsme tedy dostali ještě další podmínku, kterou lze zapsat ve tvaru $\tau \leq 2\,b/a^2$. To může být velmi vážné omezení, neboť v praxi je často $b \ll |a|$.

2.2 Diskrétní princip maxima a odhad chyby aproximace

Je známo, že řešení Cauchyovy úlohy (2.1) za určitých podmínek splňuje princip maxima, z něhož plyne, že pro $t_2 \geq t_1 \geq 0$ je $\|u(\cdot, t_2)\|_{L^\infty(\mathbb{R})} \leq \|u(\cdot, t_1)\|_{L^\infty(\mathbb{R})}$. Naším cílem nyní bude najít podmínky, za nichž vztahy tohoto typu platí pro řešení schématu (2.2). Je jistě rozumné požadovat, aby takovéto vztahy pro numerická schémata platily, platí-li pro aproximovanou Cauchyovu úlohu. Navíc uvidíme, že tím získáme nástroj pro odvození odhadu chyby aproximace pomocí chyby diskretizace.

Nejprve budeme předpokládat, že koeficient a_0 v definici operátoru L v Cauchyově úloze (2.1) je nulový, tj.,

$$L = \sum_{k=1}^{m} a_k\,\frac{\partial^k}{\partial x^k}\,.$$

Budeme předpokládat, že koeficienty schématu (2.2) splňují podmínky

$$\alpha_0 > 0\,, \qquad \alpha_s \le 0 \ \forall\ s \neq 0\,, \qquad \beta_s \ge 0 \ \forall\ s\,, \tag{2.17}$$

$$\sum_{s=-M}^{M} \alpha_s > 0\,, \tag{2.18}$$

$$\sum_{s=-M}^{M} \beta_s = \sum_{s=-M}^{M} \alpha_s\,. \tag{2.19}$$

Poznámka 2.2 *Podmínky* (2.17) *na znaménka koeficientů si lze snadno zapamatovat. Zapíšeme-li totiž schéma* (2.2) *v takovém tvaru, že na levé straně je pouze hodnota* U_j^{n+1} *a hodnoty přibližného řešení ve všech ostatních uzlech jsou na pravé straně, pak všechny koeficienty musí být nezáporné (a koeficient u* U_j^{n+1} *kladný).*

Poznámka 2.3 *Je-li schéma* (2.2) *ve tvaru s diferenčními kvocienty, pak*

$$\sum_{s=-M}^{M} \alpha_s\, u_{j+s}^{n+1} - \sum_{s=-M}^{M} \beta_s\, u_{j+s}^{n} \approx (u_t + L\,u)(x_j, t_{n+\theta})\,,$$

kde $\theta \in [0,1]$. *Je rozumné požadovat, aby tato aproximace byla přesná, pokud u je lineární funkce (polynom prvního stupně v x a t). Volbou* $u(x,t) = 1$ *pak dostáváme podmínku* (2.19). *Zvolíme-li* $u(x,t) = t$ *a použijeme-li* (2.19), *získáme*

$$\sum_{s=-M}^{M} \alpha_s = \frac{1}{\tau}\,, \tag{2.20}$$

z čehož plyne (2.18). *Poznamenejme, že všechna schémata, s nimiž se setkáme, podmínku* (2.20) *splňují.*

Pro libovolnou síťovou funkci $\{U_j^n\}_{j\in\mathbb{Z},n\in\mathbb{N}_0}$ zavedeme označení

$$U_{\min}^n = \inf_{j\in\mathbb{Z}} U_j^n\,, \qquad U_{\max}^n = \sup_{j\in\mathbb{Z}} U_j^n\,, \qquad n \in \mathbb{N}_0\,.$$

Věta 2.2 (Diskrétní princip maxima) *Nechť* $\{U_j^n\}_{j\in\mathbb{Z}}$ *je pro každé* $n \in \mathbb{N}_0$ *omezené a splňuje vztah* (2.2) *s koeficienty splňujícími* (2.17)–(2.19). *Pak*

$$U_{\min}^n \le U_j^{n+1} \le U_{\max}^n \qquad \forall\ j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,. \tag{2.21}$$

Pokud místo (2.2) *je pouze*

$$\sum_{s=-M}^{M} \alpha_s\, U_{j+s}^{n+1} \le \sum_{s=-M}^{M} \beta_s\, U_{j+s}^{n} \qquad \forall\ j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,, \tag{2.22}$$

platí

$$U_j^{n+1} \le U_{\max}^n \qquad \forall\ j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,. \tag{2.23}$$

Pokud

$$\sum_{s=-M}^{M} \alpha_s \, U_{j+s}^{n+1} \geq \sum_{s=-M}^{M} \beta_s \, U_{j+s}^{n} \qquad \forall \; j \in \mathbb{Z} \,, \; n \in \mathbb{N}_0 \,, \tag{2.24}$$

platí

$$U_{\min}^{n} \leq U_{j}^{n+1} \qquad \forall \; j \in \mathbb{Z} \,, \; n \in \mathbb{N}_0 \,. \tag{2.25}$$

Důkaz. Nechť platí (2.22). Pak díky (2.17) je pro libovolné $j \in \mathbb{Z}$ a $n \in \mathbb{N}_0$

$$\alpha_0 \, U_j^{n+1} \leq \sum_{\substack{s=-M \\ s \neq 0}}^{M} (-\alpha_s) \, U_{j+s}^{n+1} + \sum_{s=-M}^{M} \beta_s \, U_{j+s}^{n} \leq U_{\max}^{n+1} \sum_{\substack{s=-M \\ s \neq 0}}^{M} (-\alpha_s) + U_{\max}^{n} \sum_{s=-M}^{M} \beta_s \,.$$

Označíme-li $A = \sum_{s=-M}^{M} \alpha_s$, plyne z této nerovnosti díky (2.19)

$$\alpha_0 \, U_{\max}^{n+1} \leq (\alpha_0 - A) \, U_{\max}^{n+1} + A \, U_{\max}^{n} \,,$$

a tedy i (2.23), jelikož $A > 0$ podle (2.18). Je-li splněno (2.24), pak $-U_j^n$ splňuje (2.22), a tudíž pro všechna $j \in \mathbb{Z}$ a $n \in \mathbb{N}_0$ je $-U_j^{n+1} \leq \sup_{k \in \mathbb{Z}} (-U_k^n) = -U_{\min}^n$, tj. platí (2.25). Platí-li (2.2), platí dle předchozího (2.23) a (2.25), a tudíž i (2.21). □

Bude-li koeficient a_0 v definici operátoru L nenulový, pak nelze předpokládat podmínku (2.19). Při $a_0 \geq 0$ však dostaneme (srv. pozn. 2.3),

$$\sum_{s=-M}^{M} \beta_s \leq \sum_{s=-M}^{M} \alpha_s \,, \tag{2.26}$$

což je spolu s (2.17) a (2.18) postačující pro důkaz slabší varianty diskrétního principu maxima. Pro jeho formulaci zavedeme diskrétní analogii normy v prostoru $L^\infty(\mathbb{R})$ vztahem

$$\|U^n\|_\infty = \sup_{j \in \mathbb{Z}} |U_j^n| \,.$$

Věta 2.3 (Diskrétní princip maxima za slabších předpokladů) *Nechť $\{U_j^n\}_{j \in \mathbb{Z}}$ je pro každé $n \in \mathbb{N}_0$ omezené a splňuje vztah* (2.2) *s koeficienty splňujícími* (2.17), (2.18) *a* (2.26). *Pak*

$$\|U^{n+1}\|_\infty \leq \|U^n\|_\infty \qquad \forall \; n \in \mathbb{N}_0 \,. \tag{2.27}$$

Pokud místo (2.2) *je pouze*

$$\sum_{s=-M}^{M} \alpha_s \, U_{j+s}^{n+1} \leq \sum_{s=-M}^{M} \beta_s \, U_{j+s}^{n} \qquad \forall \; j \in \mathbb{Z} \,, \; n \in \mathbb{N}_0 \,, \tag{2.28}$$

platí

$$U_j^{n+1} \le (U_{\max}^n)^+ \qquad \forall\, j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,, \tag{2.29}$$

kde $(U_{\max}^n)^+ = \max\{0, U_{\max}^n\}$. *Pokud*

$$\sum_{s=-M}^{M} \alpha_s\, U_{j+s}^{n+1} \ge \sum_{s=-M}^{M} \beta_s\, U_{j+s}^n \qquad \forall\, j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,,$$

pak při označení $(U_{\min}^n)^- = \min\{0, U_{\min}^n\}$ *platí*

$$(U_{\min}^n)^- \le U_j^{n+1} \qquad \forall\, j \in \mathbb{Z}\,,\ n \in \mathbb{N}_0\,. \tag{2.30}$$

Důkaz. Analogickým způsobem jako ve větě 2.2 lze ukázat, že platí (2.29) a (2.30). Z toho pak plyne (2.27), využijeme-li nerovnosti $(U_{\max}^n)^+ \le \|U^n\|_\infty$ a $-(U_{\min}^n)^- \le \|U^n\|_\infty$. □

Pro ilustraci uvažujme opět rovnici (1.4) s konstantními koeficienty $b > 0$ a a diskretizovanou pomocí explicitního schématu (1.6). K vyšetření platnosti předpokladů pro diskrétní princip maxima použijeme ekvivalentní tvar (1.9). Pak

$$\alpha_0 = 1\,, \qquad \beta_1 = \mu - \frac{\nu}{2}\,, \qquad \beta_0 = 1 - 2\,\mu\,, \qquad \beta_{-1} = \mu + \frac{\nu}{2}\,.$$

Podmínky (2.18) a (2.19) (a tedy i (2.26)) jsou tudíž vždy splněné, avšak podmínky (2.17) platí právě tehdy, když $\mu \le \frac{1}{2}$ a $|\nu| \le 2\,\mu$. Druhá nerovnost je ekvivalentní podmínce $|a|\,h \le 2\,b$, což může být vážné omezení, neboť jak jsme již zmínili, v praxi je často $b \ll |a|$. Za uvedených podmínek je podle (2.27)

$$\|U^n\|_\infty \le \|U^0\|_\infty \qquad \forall\, n \in \mathbb{N}\,, \tag{2.31}$$

což znamená, že schéma (1.6) je stabilní vzhledem k diskrétní L^∞ normě. Příslušná oblast stability je

$$\Lambda = \{(h,\tau) \in \mathbb{R}^+ \times \mathbb{R}^+\,;\ 2\,\tau\,b \le h^2,\ |a|\,h \le 2\,b\}\,. \tag{2.32}$$

Podmínky na parametry h a τ, které jsme získali pro platnost stability (2.31) jsou silnější než v případě von Neumannovy analýzy, kde pro stabilitu v diskrétní L^2 normě stačila podmínka $\mu \le \frac{1}{2}$. Silná stabilita podle definice 2.2 byla získána za podmínky $\nu^2 \le 2\,\mu \le 1$, která je při $|\nu| \le 2\,\mu \le 1$ rovněž splněna.

Jak už jsme zmínili, diskrétní princip maxima je důležitý nejen pro stabilitu, ale též z toho důvodu, že umožňuje odhadnout chybu aproximace pomocí odhadu chyby diskretizace. Ukažme si to pro Cauchyovu úlohu (2.1) a obecné jednokrokové schéma (2.2), (2.3). Rovnici (2.1) budeme uvažovat pouze na časovém intervalu $(0,T]$, kde $T > 0$ je daný pevně zvolený čas, tj. uvažujeme úlohu

$$u_t + L\,u = 0 \quad \text{v } \mathbb{R} \times (0,T]\,, \qquad u(x,0) = u^0(x) \quad \text{pro } x \in \mathbb{R}\,. \tag{2.33}$$

Schéma (2.2) tedy uvažujeme pro $n = 0, \dots, N_T - 1$, kde N_T je dolní celá část čísla T/τ. Předpokládáme-li, že schéma (2.2) je ve tvaru s diferenčními kvocienty, pak chyba diskretizace splňuje

$$\varepsilon_j^n = \sum_{s=-M}^{M} \alpha_s \, u_{j+s}^{n+1} - \sum_{s=-M}^{M} \beta_s \, u_{j+s}^n \qquad \forall \; j \in \mathbb{Z} \,, \; n = 0, \dots, N_T - 1 \,,$$

kde opět $u_j^n = u(x_j, t_n)$. Pro chybu aproximace $e_j^n = U_j^n - u_j^n$ proto dostáváme

$$\sum_{s=-M}^{M} \alpha_s \, e_{j+s}^{n+1} - \sum_{s=-M}^{M} \beta_s \, e_{j+s}^n = -\varepsilon_j^n \qquad \forall \; j \in \mathbb{Z} \,, \; n = 0, \dots, N_T - 1 \,. \tag{2.34}$$

Pokud koeficienty schématu splňují pro dané h a τ podmínky (2.17), (2.20) a (2.26) a chyba aproximace je pro každé $n \in \{0, \dots, N_T\}$ omezená, pak můžeme chybu aproximace odhadnout podobným způsobem, jako v důkazech vět 2.2 a 2.3. Nejprve přepíšeme (2.34) pro libovolné $j \in \mathbb{Z}$ a $n \in \{0, \dots, N_T - 1\}$ do tvaru

$$\alpha_0 \, e_j^{n+1} = \sum_{\substack{s=-M \\ s \neq 0}}^{M} (-\alpha_s) \, e_{j+s}^{n+1} + \sum_{s=-M}^{M} \beta_s \, e_{j+s}^n - \varepsilon_j^n \,.$$

Díky (2.17) z toho plyne

$$\alpha_0 \, \|e^{n+1}\|_\infty \leq \|e^{n+1}\|_\infty \sum_{\substack{s=-M \\ s \neq 0}}^{M} (-\alpha_s) + \|e^n\|_\infty \sum_{s=-M}^{M} \beta_s + \|\varepsilon^n\|_\infty \,.$$

Použitím (2.26) a (2.20) odtud dostaneme

$$\|e^{n+1}\|_\infty \leq \|e^n\|_\infty + \tau \, \|\varepsilon^n\|_\infty \qquad \forall \; n = 0, \dots, N_T - 1 \,.$$

Pro $n = 0$ splňuje přibližné řešení vztah (2.3), a tudíž chyba aproximace splňuje $e_j^0 = 0$ pro $j \in \mathbb{Z}$. Je proto $\|e^0\|_\infty = 0$, a tedy pro $n = 1, \dots, N_T$ platí

$$\|e^n\|_\infty \leq \tau \sum_{m=0}^{n-1} \|\varepsilon^m\|_\infty \leq \tau \, n \max_{m=0,\dots,n-1} \|\varepsilon^m\|_\infty \leq T \max_{m=0,\dots,N_T-1} \|\varepsilon^m\|_\infty \,. \tag{2.35}$$

Chybu diskretizace můžeme zpravidla vyjádřit pomocí derivací u násobených mocninami h a τ, viz např. (1.11). Časový krok τ je přitom obvykle omezen násobkem nějaké mocniny prostorového kroku h, neboť platnost podmínek (2.17) vyžaduje, aby dvojice (h, τ) náležely do vhodné oblasti stability Λ, jako je např. (2.32) pro schéma (1.6). Jsou-li zmíněné derivace ve vyjádření chyby diskretizace v množině $\mathbb{R} \times [0, T]$ omezené, pak pro chybu diskretizace získáme odhad typu $\|\varepsilon^n\|_\infty \leq K \, h^p$, kde K je nezávislé na $(h, \tau) \in \Lambda$ a $n \in \{0, \dots, N_T - 1\}$ a p se nazývá řád přesnosti schématu (2.2). Z (2.35) pak dostáváme,

že existuje konstanta C taková, že pro libovolné $(h,\tau) \in \Lambda$ splňuje přibližné řešení určené schématem (2.2), (2.3) odhad

$$|U_j^n - u_j^n| \le C\,h^p \qquad \forall\, j \in \mathbb{Z}\,, \quad n = 0, \ldots, N_T\,. \tag{2.36}$$

To je hledaný odhad chyby aproximace, který ukazuje, že chyba přibližného řešení konverguje stejnoměrně (dokonce s řádem p) do nuly pro $h \to 0$.

Z uvedeného odvození není zřejmé, jak odhad chyby aproximace souvisí s diskrétním principem maxima. Ukažme si proto nyní formální odvození vztahu (2.35), které platnost diskrétního principu maxima explicitně využívá. Nejprve definujme tzv. *srovnávací funkci* $\Phi_j^n = t_n\,D$ pro $j \in \mathbb{Z}$ a $n = 0, \ldots, N_T$, kde $D = \max_{m=0,\ldots,N_T-1} \|\varepsilon^m\|_\infty$. Použitím (2.26) a (2.20) získáme

$$\sum_{s=-M}^{M} \alpha_s\,\Phi_{j+s}^{n+1} - \sum_{s=-M}^{M} \beta_s\,\Phi_{j+s}^{n} \ge D \qquad \forall\, j \in \mathbb{Z}\,,\ n = 0, \ldots, N_T - 1\,.$$

S přihlédnutím k (2.34) zjišťujeme, že síťová funkce $V_j^n = e_j^n - \Phi_j^n$ splňuje

$$\sum_{s=-M}^{M} \alpha_s\,V_{j+s}^{n+1} \le \sum_{s=-M}^{M} \beta_s\,V_{j+s}^{n} \qquad \forall\, j \in \mathbb{Z}\,,\ n = 0, \ldots, N_T - 1\,.$$

Podobně síťová funkce $W_j^n = -e_j^n - \Phi_j^n$ splňuje

$$\sum_{s=-M}^{M} \alpha_s\,W_{j+s}^{n+1} \le \sum_{s=-M}^{M} \beta_s\,W_{j+s}^{n} \qquad \forall\, j \in \mathbb{Z}\,,\ n = 0, \ldots, N_T - 1\,.$$

Jelikož předpokládáme, že koeficienty splňují (2.17), (2.18) a (2.26), a předchozí dvě nerovnosti odpovídají nerovnosti (2.28), platí podle věty 2.3 nerovnost (2.29), tj.

$$V_j^{n+1} \le (V_{\max}^n)^+\,, \quad W_j^{n+1} \le (W_{\max}^n)^+ \qquad \forall\, j \in \mathbb{Z}\,,\ n = 0, \ldots, N_T - 1\,.$$

První z těchto nerovností implikuje, že pro libovolné $j \in \mathbb{Z}$ a $n = 0, \ldots, N_T - 1$

$$e_j^{n+1} \le t_{n+1}\,D + (V_{\max}^n)^+ \le t_{n+1}\,D + (\|e^n\|_\infty - t_n\,D)^+\,.$$

Podobně druhá nerovnost implikuje, že pro libovolné $j \in \mathbb{Z}$ a $n = 0, \ldots, N_T - 1$

$$-e_j^{n+1} \le t_{n+1}\,D + (W_{\max}^n)^+ \le t_{n+1}\,D + (\|e^n\|_\infty - t_n\,D)^+\,.$$

Poslední dvě nerovnosti implikují, že

$$\|e^{n+1}\|_\infty \le t_{n+1}\,D + (\|e^n\|_\infty - t_n\,D)^+ \qquad \forall\, n = 0, \ldots, N_T - 1\,,$$

z čehož plyne, že $\|e^n\|_\infty \le t_n\,D \le T\,D$ pro $n = 1, \ldots, N_T$, což je nerovnost (2.35).

Je nutno zdůraznit, že odhad (2.36) byl získán za předpokladu existence a omezenosti příslušných derivací přesného řešení u v celé výpočetní oblasti. Tento předpoklad nemusí

být vždy splněn, např. u parciálních diferenciálních rovnic prvního řádu. Uvedený postup je však i v těchto případech užitečný pro lokální analýzu.

Z výše uvedeného by se mohlo zdát, že pro získání přibližných řešení aproximujících přesné řešení s vysokou přesností stačí konstruovat schémata vysokého řádu přesnosti splňující princip maxima. Bohužel se ukazuje, že to obecně není možné, což ilustruje i následující výsledek.

Lemma 2.1 *Uvažujme úlohu* (2.33) *s* $L\,u = a\,u_x$ *a* $u^0(x) = \sin(\pi\,x)$, *kde* $a > 0$ *je konstanta. Tuto úlohu diskretizujme na stejnoměrné síti s uzly* (x_j, t_n), *kde* $x_j = j\,h$, $t_n = n\,\tau$, $j \in \mathbb{Z}$ *a* $n = 0, \ldots, N_T$, N_T *je dolní celá část čísla* T/τ. *Předpokládejme, že použité numerické schéma pro* $2\,a\,\tau \le h$ *splňuje princip maxima, tj. přibližná řešení splňují*

$$U_j^{n+1} \le U_{\max}^n \qquad \forall\ j \in \mathbb{Z}\,,\ n = 0, \ldots, N_T - 1\,. \tag{2.37}$$

Dále předpokládejme, že za uvedené podmínky stability platí pro chybu aproximace odhad (2.36) *s konstantou* C *nezávislou na* h *a* τ. *Pak pro řád konvergence* p *musí platit podmínka* $p \le 2$.

Důkaz. Řešení úlohy (2.33) je dáno vztahem

$$u(x,t) = \sin\left(\pi\,(x - a\,t)\right)\,,$$

viz následující část (vztah (3.3)). Zvolme $N \in \mathbb{N}$ a položme $h = 1/(2N)$, $\tau = h/(2\,a)$. Pak

$$u_j^n = \sin\frac{(2\,j - n)\,\pi}{4\,N}\,.$$

Zvolme $n \in \{0, \ldots, N_T - 1\}$ liché a označme $J = N + (n-1)/2$. Pak

$$\begin{aligned}\sup_{j\in\mathbb{Z}} u_j^n &= \sup_{j\in\mathbb{Z}} u_{j+(n-1)/2}^n = \sup_{j\in\mathbb{Z}} \sin\frac{(2\,j-1)\,\pi}{4\,N} = \max_{j=1,\ldots,4\,N} \sin\frac{(2\,j-1)\,\pi}{4\,N} \\ &= \max_{j=1,\ldots,2\,N} \sin\frac{(2\,j-1)\,\pi}{4\,N} = \max_{j=1,\ldots,N} \sin\frac{(2\,j-1)\,\pi}{4\,N} = \sin\frac{(2\,N-1)\,\pi}{4\,N} = u_J^n\,.\end{aligned}$$

Uvedené rovnosti plynou po řadě z toho, že funkce sinus je periodická s periodou $2\,\pi$, záporná na intervalu $(\pi, 2\,\pi)$, sudá vzhledem k bodu $\frac{\pi}{2}$ a rostoucí na intervalu $(0, \frac{\pi}{2})$. Díky odhadu (2.36) platí tedy

$$U_j^n \le u_j^n + C\,h^p \le u_J^n + C\,h^p \qquad \forall\ j \in \mathbb{Z}\,,$$

z čehož pomocí (2.37) plyne

$$U_j^{n+1} \le u_J^n + C\,h^p \qquad \forall\ j \in \mathbb{Z}\,.$$

Opětovné použití odhadu (2.36) pak dává

$$C\,h^p \ge u_{J+1}^{n+1} - U_{J+1}^{n+1} \ge u_{J+1}^{n+1} - u_J^n - C\,h^p\,. \tag{2.38}$$

Použitím Taylorova vzorce dostáváme

$$u_{J+1}^{n+1} - u_J^n = \sin\frac{\pi}{2} - \sin\left(\frac{\pi}{2} - \frac{\pi\,h}{2}\right) = \frac{1}{2}\,\sin\xi\,\left(\frac{\pi\,h}{2}\right)^2 > \left(\frac{\pi\,h}{4}\right)^2\,,$$

kde $\xi \in (\frac{\pi}{2} - \frac{\pi\,h}{2}, \frac{\pi}{2})$. Dosazením do (2.38) získáme $\pi^2\,h^2 < 32\,C\,h^p$. Jelikož tento vztah platí pro $h \to 0$, musí být $p \leq 2$. □

3 Numerické řešení transportní rovnice

V této části se budeme zabývat numerickým řešením Cauchyovy úlohy pro transportní rovnici v jedné prostorové dimenzi, tj. úlohy

$$u_t + a\,u_x = 0 \qquad \text{v } \mathbb{R}\times\mathbb{R}^+\,, \tag{3.1}$$

$$u(x,0) = u^0(x) \qquad \forall\, x \in \mathbb{R} \tag{3.2}$$

pro neznámou funkci $u = u(x,t)$ závisející na prostorové proměnné $x \in \mathbb{R}$ a časové proměnné $t \in \mathbb{R}_0^+$. Funkce u^0 je daná počáteční podmínka a koeficient a je rychlost šíření veličiny u. Náš hlavní zájem bude patřit rovnici (3.1) s konstatním koeficientem a, avšak v některých případech budeme uvažovat též rovnici (3.1) s nenulovou pravou stranou či s nekonstatním koeficientem a.

Je známo, že řešení rovnice (3.1) jsou konstantní podél chrakteristik, což jsou křivky $x = x(t)$ splňující $x'(t) = a(x(t),t)$. Je-li funkce a lipschitzovská v x a spojitá v t, charakteristiky se neprotnou. Jejich sestrojením tedy získáme řešení dané vztahem $u(x(t),t) = u^0(x(0))$. Je-li a konstantní, pak charakteristiky jsou určeny rovnicí $x - at = konst.$ Jsou to tedy rovnoběžné přímky v rovině (x,t) se směrnicí $1/a$. Z toho plyne, že pro $u^0 \in C^1(\mathbb{R})$ a konstatní a má úloha (3.1), (3.2) právě jedno klasické řešení, které je dáno vztahem

$$u(x,t) = u^0(x - a\,t)\,. \tag{3.3}$$

Vidíme tedy, že řešení u v čase t je rovno počáteční podmínce u^0 posunuté o vzdálenost $|a|t$ (doprava pro $a > 0$, doleva pro $a < 0$). Parametr a se nazývá rychlost šíření podél charakteristik. Řešení u lze tak považovat za vlnu šířící se rychlostí a beze změny tvaru. Proto se rovnice (3.1) také nazývá *jednosměrná vlnová rovnice*.

Samozřejmě vyvstává otázka, proč se zabývat numerickým řešením úlohy (3.1), (3.2), když je její řešení známo. Důvodem je to, že u komplikovanějších úloh, v nichž hrají důležitou roli transportní mechanismy, již řešení zpravidla nejsme schopni v analytickém tvaru získat a musíme ho aproximovat pomocí numerických metod. Přitom navržení vhodné metody není vůbec triviální a řada potíží, s nimiž se setkáváme, se vyskytuje již při numerickém řešení rovnice (3.1). Abychom příslušné jevy lépe pochopili, je rozumné je studovat na co nejjednodušší úloze, a proto uvažujeme rovnici (3.1). U metod, které se ukáží jako nevhodné pro řešení rovnice (3.1), nelze samozřejmě očekávat, že by dávaly dobré výsledky pro komplikovanější úlohy.

3.1 Příklady numerických schémat

Opět budeme uvažovat rovnoměrnou síť s konstantním prostorovým krokem h a konstantním časovým krokem τ a definujeme uzly sítě (x_j, t_n), kde $x_j = j\,h$, $t_n = n\,\tau$, $j \in \mathbb{Z}$ a $n \in \mathbb{N}_0$. Řešení u Cauchyovy úlohy (3.1), (3.2) opět aproximujeme v uzlech (x_j, t_n) hodnotami U_j^n. Rovnici (3.1) aproximujeme v libovolném uzlu (x_j, t_n) diferenčním schématem získaným nahrazením parciálních derivací diferenčními kvocienty. Takto lze získat mnoho rozličných numerických schémat, z nichž několik nyní uvedeme. Pro jejich přehlednější zápis budeme užívat parametr $\nu = a\,\tau/h$, který jsme zavedli již v části 1.2.

Pro numerické řešení rovnice (3.1) lze uvažovat například schémata:

$$\frac{U_j^{n+1} - U_j^n}{\tau} + a\,\frac{U_{j+1}^n - U_j^n}{h} = 0\,, \qquad \text{tj.} \qquad U_j^{n+1} = (1+\nu)\,U_j^n - \nu\,U_{j+1}^n\,, \tag{3.4}$$

$$\frac{U_j^{n+1} - U_j^n}{\tau} + a\,\frac{U_j^n - U_{j-1}^n}{h} = 0\,, \qquad \text{tj.} \qquad U_j^{n+1} = \nu\,U_{j-1}^n + (1-\nu)\,U_j^n\,, \tag{3.5}$$

$$\frac{U_j^{n+1} - U_j^n}{\tau} + a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} = 0\,, \qquad \text{tj.} \qquad U_j^{n+1} = \frac{\nu}{2}\,U_{j-1}^n + U_j^n - \frac{\nu}{2}\,U_{j+1}^n\,, \tag{3.6}$$

$$\frac{U_j^{n+1} - U_j^{n-1}}{2\,\tau} + a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} = 0\,, \qquad \text{tj.} \qquad U_j^{n+1} = U_j^{n-1} + \nu\,U_{j-1}^n - \nu\,U_{j+1}^n\,, \tag{3.7}$$

$$\frac{U_j^{n+1} - \frac{1}{2}\,(U_{j+1}^n + U_{j-1}^n)}{\tau} + a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} = 0\,, \qquad \text{tj.} \tag{3.8}$$

$$U_j^{n+1} = \frac{1}{2}\,(1+\nu)\,U_{j-1}^n + \frac{1}{2}\,(1-\nu)\,U_{j+1}^n\,.$$

Schéma (3.7) je známo pod označením *leapfrog scheme* a 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. Je zřejmé, že kromě schématu (3.7) jsou všechna výše uvedená schémata jednokroková. Poznamenejme též, že všechna uvedená schémata jsou explicitní.

Schéma (3.8) se nazývá Laxovo–Friedrichsovo schéma. Všimněme si, že

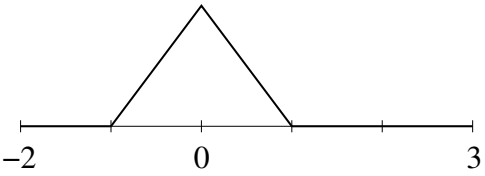
$$\frac{U_j^{n+1} - \frac{1}{2}\,(U_{j+1}^n + U_{j-1}^n)}{\tau} = \frac{U_j^{n+1} - U_j^n}{\tau} - \frac{h^2}{2\,\tau}\,\frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2}\,.$$

Jelikož diferenční kvocient $(U_{j+1}^n - 2\,U_j^n + U_{j-1}^n)/h^2$ reprezentuje aproximaci $u_{xx}(x_j, t_n)$, lze tedy speciální aproximaci časové derivace použitou v (3.8) interpretovat jako přidání umělé difúze o velikosti $h^2/(2\,\tau)$ ke schématu (3.6). Laxovo–Friedrichsovo schéma tedy odpovídá diskretizaci rovnice

$$u_t - \frac{h^2}{2\,\tau}\,u_{xx} + a\,u_x = 0$$

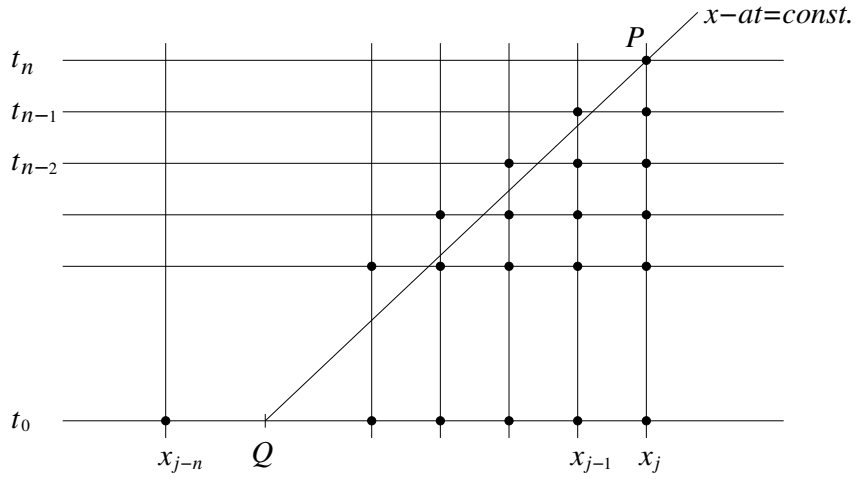
s použitím dopředné diference pro časovou derivaci a centrálních diferencí pro prostorové derivace.

Cvičení 3.1 *Uvažujme rovnici $u_t+u_x = 0$ v oblasti $(-2,3)\times\mathbb{R}^+$ s počáteční podmínkou u^0 znázorněnou v obr. 1. Definujme ekvidistantní prostorové uzly $-2 = x_0 < x_1 < \cdots < x_J = 3$ s prostorovým krokem $h = 5/J$. Dále uvažujme konstatní časový krok $\tau > 0$ a časové hladiny $t_n = n\tau$, $n \in \mathbb{N}_0$. V čase t_0 je přibližné řešení určeno počáteční podmínkou. Na následujících časových hladinách pak můžeme hodnoty přibližného řešení ve vnitřních uzlech určit pomocí libovolného z výše uvedených schémat, přičemž položíme $U_0^n = 0$ a $U_J^n = U_{J-1}^n$ pro $n \in \mathbb{N}$ (pro $n = 1$ však nelze použít schéma (3.7)). Naprogramujte výše uvedená schémata a proveďte výpočty pro $J = 50$ a $\tau = 0.08$. U Laxova-Friedrichsova schématu otestujte, jaký vliv má zmenšení h na polovinu při nezměněném τ, zmenšení τ na polovinu při nezměněném h a současné zmenšení h i τ na polovinu, popř. na čtvrtinu.*



Obrázek 1: Počáteční podmínka u^0 použitá ve cvičení 3.1.

Provedeme-li výpočty popsané ve cvičení 3.1, zjistíme, že v přibližných řešeních schémat (3.4) a (3.6) se již po několika málo časových krocích objeví velké oscilace a dostaneme zcela nepoužitelné výsledky. Na druhou stranu zbývající tři schémata vedou při $J = 50$ a $\tau = 0.08$ k přijatelným aproximacím řešení uvažované úlohy. Změníme-li však J či τ, objeví se v některých případech nestabilní chování i u těchto schémat. Naším cílem nyní bude vysvětlit, proč k uvedeným jevům dochází.



Obrázek 2: Oblast závislosti PDR a schématu (3.5).

Zvolme bod $P \in \mathbb{R} \times \mathbb{R}^+$ a uvažujme schéma (3.5) a takovou síť, že bod P je její uzel. Nechť např. $P = (x_j, t_n)$. Pak hodnota U_j^n je určena hodnotami U_{j-1}^{n-1} a U_j^{n-1}. Podobně tyto dvě hodnoty jsou určeny hodnotami U_{j-2}^{n-2}, U_{j-1}^{n-2} a U_j^{n-2}, viz obr. 2. Z toho plyne, že hodnota U_j^n závisí pouze na hodnotách počáteční podmínky u^0 v bodech

$x_{j-n}, x_{j-n+1}, \ldots, x_{j-1}, x_j$. Pokud bychom uvažovali v rovnici (3.1) nenulovou pravou stranu f, závisela by hodnota U_j^n též na hodnotách f v uzlech sítě ležících v trojúhelníku s vrcholy (x_j, t_n), $(x_{j-n}, 0)$ a $(x_j, 0)$. Tento trojúhelník se nazývá *oblast závislosti* schématu v uzlu (x_j, t_n).

Pro řešení u úlohy (3.1), (3.2) plyne ze vztahu (3.3), že $u(P) = u^0(Q)$, kde Q je průsečík přímky $t = 0$ a charakteristiky $x - a\,t = const.$ procházející bodem P, viz obr. 2. V případě rovnice (3.1) s nenulovou pravou stranou f závisí hodnota $u(P)$ též na hodnotách funkce f v bodech úsečky PQ. Úsečka PQ je *oblastí závislosti* uvažované parciální diferenciální rovnice v bodě P. Nyní můžeme zformulovat následující podmínku.

Věta 3.1 (CFL podmínka; Courant, Friedrichs, Lewy (1928)) *Nutná podmínka konvergence diferenčního schématu je, aby oblast závislosti parciální diferenciální rovnice ležela uvnitř oblasti závislosti diferenčního schématu.*

Pro schéma (3.5) je CFL podmínka splněna, leží-li bod Q na přímce $t = 0$ mezi body x_{j-n} a x_j. To je právě tehdy, když $a \geq 0$ a $a\,\tau \leq h$. Uvažujme posloupnost sítí s $\tau/h = const.$ a $h \to 0$ a nechť P je uzlem všech těchto sítí. Pak oblast závislosti schématu (3.5) v uzlu P je pro všechny tyto sítě stejná. Pokud $a < 0$ nebo $a\,\tau > h$, pak pro všechny tyto sítě hodnota přibližného řešení U v uzlu P nezávisí na hodnotách počáteční podmínky u^0 v pevně daném okolí bodu Q (neměnném při $h \to 0$). Přibližná řešení v uzlu P tudíž obecně nemohou konvergovat k hodnotě $u(P)$. Z uvedeného je zřejmé, že platí též následující věta.

Věta 3.2 *Nutnou podmínkou konvergence explicitního schématu typu* $U_j^{n+1} = \alpha\,U_{j-1}^n + \beta\,U_j^n + \gamma\,U_{j+1}^n$ *pro rovnici* (3.1) *při* $\tau/h = const.$ *je, aby platilo* $|a\,\tau/h| \leq 1$.

Nerovnost $|a\,\tau/h| \leq 1$ z věty 3.2 se také často nazývá CFL podmínka.

Všimněme si, že pro schéma tvaru uvažovaného ve větě 3.2 se informace během jednoho časové kroku rozšíří o prostorový krok h. Numerická rychlost šíření informace je tudíž h/τ. Uvedená nerovnost tedy říká, že numerická rychlost šíření musí být větší nebo rovna rychlosti šíření odpovídající uvažované parciální diferenciální rovnici. Pokud diferenční schéma nemůže šířit řešení alespoň tak rychle, jako se šíří řešení parciální diferenciální rovnice, nemůže řešení schématu konvergovat k řešení parciální diferenciální rovnice.

Na CFL podmínku se lze dívat též jako na nutnou podmínku stability diferenčního schématu. Obecně není pro stabilitu postačující, ale její velkou předností je její jednoduchost. Umožňuje tak vyřadit řadu diferenčních schémat s nepatrnou námahou věnovanou jejich vyšetřování. Teprve schémata, která splňují CFL podmínku je vhodné vyšetřovat podrobněji použitím kritérií, která jsou pro jejich stabilitu postačující.

Vrátíme-li se ke schématům (3.4)–(3.8), pak vidíme, že ve všech případech je nutno splnit podmínku $|a\,\tau/h| \leq 1$. To vysvětluje, proč schéma (3.8) vede pro úlohu ze cvičení 3.1 při $h = 0.05$ a $\tau = 0.08$ k oscilacím. Je též zřejmé, že schéma (3.4) nelze pro úlohu ze cvičení 3.1 použít, neboť oblast závislosti parciální diferenciální rovnice neleží v oblasti závislosti schématu. Oscilace lze pozorovat též v případě schématu (3.6) při splnění CFL podmínky. Vyšetřeme proto stabilitu tohoto schématu pomocí Fourierovy metody (tj. proveďme von Neumannovu analýzu stability).

Rigorózní postup určení amplifikačního faktoru schématu (3.6) spočívá ve vyjádření přibližného řešení pomocí jeho Fourierovy transformace a využití ortonormality funkcí $\{\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}\}_{j\in\mathbb{Z}}$. Viděli jsme však, že formálně lze amplifikační faktor jednokrokového schématu získat nahrazením přibližného řešení síťovou funkcí $\lambda(\xi)^n\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}$. Dosazením této funkce za U_j^n ve schématu (3.6) získáme

$$\lambda(\xi) = 1 + \frac{\nu}{2}\left(\mathrm{e}^{-\mathrm{i}\,\xi\,h} - \mathrm{e}^{\mathrm{i}\,\xi\,h}\right) = 1 - \mathrm{i}\,\nu\,\sin(\xi\,h)\,.$$

Pro libovolnou síť a každé ξ, pro které $\sin(\xi\,h) \neq 0$, je tedy $|\lambda(\xi)| > 1$. Pro každý proces zjemňování, při němž je ν pevné, je proto schéma nestabilní. Vidíme tedy, že CFL podmínka skutečně není pro stabilitu postačující.

Jak jsme viděli, při $a > 0$ lze k řešení rovnice (3.1) použít schéma (3.5). Bude-li $a < 0$, nelze toto schéma použít, neboť není splněna CFL podmínka, avšak můžeme použít schéma (3.4). To nás vede ke schématu

$$U_j^{n+1} = \begin{cases} (1+\nu)\,U_j^n - \nu\,U_{j+1}^n\,, & \text{je-li} \quad a < 0\,, \\ \nu\,U_{j-1}^n + (1-\nu)\,U_j^n\,, & \text{je-li} \quad a > 0\,. \end{cases} \tag{3.9}$$

Je-li a nekonstantní, pak a v (3.9) značí hodnotu $a(x_j, t_n)$. Všimněme si, že při $a > 0$ se informace šíří ve směru kladné poloosy x (tj. zleva doprava) a k diskretizaci v prostorovém bodě x_j je využita informace v bodě x_j a vlevo od něj. Na druhou stranu při $a < 0$ se informace šíří ve směru záporné poloosy x (tj. zprava doleva) a k diskretizaci v bodě x_j je použita informace v bodě x_j a vpravo od něj. Jedná se o tzv. *diskretizaci typu upwind*, kdy k diskretizaci v daném bodě využíváme informaci, která leží proti směru šíření (tj. která do daného bodu přichází).

Je-li $|a\,\tau/h| \leq 1$, splňuje schéma (3.9) CFL podmínku. Proveďme též von Neumannovu analýzu stability (při konstantním a). Dosazením $U_j^n := \lambda(\xi)^n\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}$ do schématu (3.9) dostaneme při $a < 0$

$$\lambda(\xi) = (1+\nu) - \nu\,\mathrm{e}^{\mathrm{i}\,\xi\,h} = 1 + \nu - \nu\cos(\xi\,h) - \mathrm{i}\,\nu\,\sin(\xi\,h)$$

a při $a > 0$

$$\lambda(\xi) = \nu\,\mathrm{e}^{-\mathrm{i}\,\xi\,h} + (1-\nu) = 1 - \nu + \nu\cos(\xi\,h) - \mathrm{i}\,\nu\,\sin(\xi\,h)\,.$$

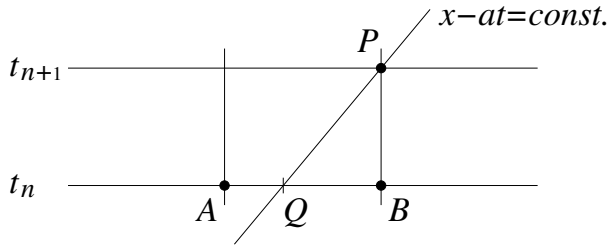
Tyto dva vztahy lze pro obě znaménka koeficientu a psát ve tvaru

$$\lambda(\xi) = 1 - |\nu| + |\nu|\cos(\xi\,h) - \mathrm{i}\,\nu\,\sin(\xi\,h)\,, \tag{3.10}$$

z čehož plyne, že

$$|\lambda(\xi)|^2 = 1 - 4\,|\nu|\,(1-|\nu|)\sin^2\frac{\xi\,h}{2}\,. \tag{3.11}$$

Zjišťujeme tedy, že $|\lambda(\xi)| \leq 1$ pro všechna $\xi \in \mathbb{R}$ právě tehdy, když $|\nu| \leq 1$. V tomto případě tedy dává CFL podmínka správné meze stability.



Obrázek 3: Odvození schématu (3.9) pomocí charakteristik.

Všimněme si, že schéma (3.9) lze přepsat do tvaru

$$\frac{U_j^{n+1} - U_j^n}{\tau} - \frac{|a|\,h}{2}\,\frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2} + a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} = 0\,.$$

Podobně jako Laxovo–Friedrichsovo schéma můžeme tedy i schéma (3.9) interpretovat jako přidání umělé difúze ke schématu (3.6). V tomto případě má umělá difúze velikost $|a|h/2$. Jelikož $|a|h/2 = |\nu|\,h^2/(2\,\tau)$, je při $|\nu| < 1$ umělá difúze přidávaná schématem (3.9) vždy menší než umělá difúze odpovídající Laxovu–Friedrichsovu schématu. Při $|\nu| = 1$ jsou obě schémata stejná a dávají přesné řešení, neboť uzly sítě leží na charakteristikách.

Ukažme si ještě jiný způsob odvození schématu (3.9). Pro jednoduchost uvažujme případ $a > 0$ a nechť $a\,\tau \le h$. Předpokládejme, že známe hodnoty U_A, U_B přibližného řešení v uzlech A, B a chceme určit hodnotu U_P přibližného řešení v uzlu P, viz obr. 3. Charakteristika procházející uzlem P protíná síťovou přímku $t = t_n$ v bodě Q ležícím mezi body A a B. Řešení u rovnice (3.1) splňuje $u(P) = u(Q)$ a je tedy přirozené definovat hodnotu U_P jako aproximaci u v bodě Q získanou pomocí hodnot U_A a U_B. Použijeme-li lineární interpolaci, dostaneme vztah

$$U_P = \frac{|BQ|}{|AB|}\,U_A + \frac{|AQ|}{|AB|}\,U_B\,.$$

To je přesně schéma (3.5), tj. schéma (3.9) pro $a > 0$, neboť $|AB| = h$, $|BQ| = a\,\tau = \nu\,h$ a $|AQ| = |AB| - |BQ| = (1 - \nu)\,h$.

Poznamenejme nakonec, že při splnění podmínky $|\nu| \le 1$ schéma (3.9) splňuje podmínky (2.17), (2.18) a (2.19), a platí tedy pro něj diskrétní princip maxima formulovaný ve větě 2.2. To znamená, že přibližné řešení zůstane omezeno minimem a maximem počáteční podmínky a neobjeví se v něm narůstající oscilace pozorované u některých metod ve cvičení 3.1. Diskrétní princip maxima splňuje za podmínky $|\nu| \le 1$ též Laxovo–Friedrichsovo schéma (3.8).

3.2 Fázová rychlost a disperze

Uvažujme úlohu (3.1), (3.2) s $a = const.$ Definujme Fourierovu transformaci řešení u vztahem

$$\widehat{u}(\xi, t) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} u(x,t)\,\mathrm{e}^{-\mathrm{i}\,\xi\,x}\,\mathrm{d}x\,, \qquad \xi \in \mathbb{R}\,.$$

Pak

$$\widehat{u}_t + \mathrm{i}\,\xi\,a\,\widehat{u} = 0 \quad \text{v } \mathbb{R}\times\mathbb{R}^+\,, \qquad \widehat{u}(\xi,0) = \widehat{u}^0(\xi) \quad \forall\,\xi\in\mathbb{R}\,,$$

kde

$$\widehat{u}^0(\xi) = \frac{1}{\sqrt{2\,\pi}}\int_{-\infty}^{\infty} u^0(x)\,\mathrm{e}^{-\mathrm{i}\,\xi\,x}\,\mathrm{d}x\,, \qquad \xi\in\mathbb{R}\,.$$

Je tedy

$$\widehat{u}(\xi,t) = \widehat{u}^0(\xi)\,\mathrm{e}^{-\mathrm{i}\,\xi\,a\,t}$$

a platí

$$u(x,t) = \frac{1}{\sqrt{2\,\pi}}\int_{-\infty}^{\infty}\widehat{u}(\xi,t)\,\mathrm{e}^{\mathrm{i}\,\xi\,x}\,\mathrm{d}\xi = \frac{1}{\sqrt{2\,\pi}}\int_{-\infty}^{\infty}\widehat{u}^0(\xi)\,\mathrm{e}^{\mathrm{i}\,\xi\,(x-a\,t)}\,\mathrm{d}\xi = u^0(x-a\,t)\,.$$

Z uvedeného plyne, že

$$\widehat{u}(\xi,t+\tau) = \widehat{u}(\xi,t)\,\mathrm{e}^{-\mathrm{i}\,\xi\,a\,\tau}\,. \tag{3.12}$$

Během jednoho časového kroku se tedy fáze Fourierovy transformace řešení úlohy (3.1), (3.2) změní o $-\xi\,a\,\tau$. Všimněme si též, že pro libovolné $\xi\in\mathbb{R}$ je $|\widehat{u}(\xi,t)| = |\widehat{u}^0(\xi)|$, tj. velikost amplitudy libovolného Fourierova členu je v čase konstatní.

V diskrétním případě je

$$U_j^n = \frac{1}{\sqrt{2\,\pi}}\int_{-\pi/h}^{\pi/h}\widehat{U}^n(\xi)\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}\,\mathrm{d}\xi\,, \qquad \text{kde} \qquad \widehat{U}^n(\xi) = \frac{1}{\sqrt{2\,\pi}}\sum_{j=-\infty}^{\infty} h\,U_j^n\,\mathrm{e}^{-\mathrm{i}\,\xi\,j\,h}\,.$$

Všimněme si, že interval $[-\pi/h,\pi/h]$ obsahuje všechny frekvence, které lze na síti s krokem h reprezentovat, neboť pro libovolné $k\in\mathbb{Z}$ je

$$\mathrm{e}^{\mathrm{i}\left(\xi+\frac{2\,\pi}{h}\,k\right)j\,h} = \mathrm{e}^{\mathrm{i}\,\xi\,j\,h}\,\mathrm{e}^{\mathrm{i}\,2\,\pi\,k\,j} = \mathrm{e}^{\mathrm{i}\,\xi\,j\,h}\,.$$

Pro libovolné jednokrokové diferenční schéma platí $\widehat{U}^{n+1}(\xi) = \lambda(\xi)\,\widehat{U}^n(\xi)$. Vzhledem ke vztahu (3.12) je tedy žádoucí, aby $\lambda(\xi)$ byla dobrá aproximace $\mathrm{e}^{-\mathrm{i}\,\xi\,a\,\tau}$. Abychom tato dvě čísla mohli lépe porovnat, zapíšeme amplifikační faktor ve tvaru $\lambda(\xi) = |\lambda(\xi)|\,\mathrm{e}^{-\mathrm{i}\,\xi\,\alpha(\xi)\,\tau}$. Hodnota $|\lambda(\xi)|$ vyjadřuje pokles amplitudy složky přibližného řešení o frekvenci ξ během jednoho časového kroku. Jelikož, jak jsme viděli, pro přesné řešení libovolný Fourierův člen není tlumený, požadujeme, aby $|\lambda(\xi)|$ bylo blízké hodnotě 1. Veličina $\alpha(\xi)$ se nazývá *fázová rychlost*. Je to rychlost, kterou diferenční schéma šíří vlny o frekvenci ξ. Pokud by bylo $\alpha(\xi) = a$ pro všechna ξ, pak by se vlny šířily správnou rychlostí, avšak s tím se obecně nesetkáme u žádného z diferenčních schémat. Rychlost $\alpha(\xi)$ je pouze aproximací a a je obecně různá pro různé hodnoty ξ, což se projevuje deformací tvaru řešení diferenčního schématu. Jev, kdy se vlny o různých frekvencích pohybují různými rychlostmi, se nazývá *disperze*. Rozdíl $a-\alpha(\xi)$ se nazývá *fázová chyba*.

Fázovou rychlost můžeme určit ze vztahu

$$\tan(\xi\,\alpha(\xi)\,\tau) = -\frac{\operatorname{Im}\lambda(\xi)}{\operatorname{Re}\lambda(\xi)}\,.$$

Je-li $|\lambda(\xi)| = 1$, platí $\sin(\xi\,\alpha(\xi)\,\tau) = -\mathrm{Im}\,\lambda(\xi)$. Fázovou chybu je vhodné vyšetřovat zejména pro malé hodnoty $|\xi\,h|$, neboť vlny s takovýmito frekvencemi lze na síti s krokem h dobře aproximovat. Přitom můžeme využít následující vztahy pro počítání s malými čísly:

$$\frac{1}{1+x} = 1 - x + O(x^2)\,, \quad \sin x = x\left(1 - \frac{x^2}{6} + O(x^4)\right), \quad \cos x = 1 - \frac{x^2}{2} + O(x^4)\,,$$
$$\tan x = x\left(1 + \frac{x^2}{3} + O(x^4)\right), \qquad \tan^{-1} y = y\left(1 - \frac{y^2}{3} + O(y^4)\right).$$

Příklad 3.1 *Vyšetřeme fázovou chybu pro schéma* (3.9).

Řešení: Podle (3.10) a výše uvedených vztahů je

$$\tan(\xi\,\alpha(\xi)\,\tau) = \frac{\nu\,\sin(\xi\,h)}{1 - |\nu| + |\nu|\cos(\xi\,h)} = \frac{\nu\,\xi\,h\left(1 - \frac{(\xi\,h)^2}{6} + O((\xi\,h)^4)\right)}{1 - |\nu|\,\frac{(\xi\,h)^2}{2} + O((\xi\,h)^4)}$$
$$= \nu\,\xi\,h\left[1 - \left(\frac{1}{6} - \frac{|\nu|}{2}\right)(\xi\,h)^2 + O((\xi\,h)^4)\right],$$

z čehož plyne (je $\nu\,\xi\,h = \xi\,a\,\tau$)

$$\alpha(\xi) = a\left[1 - \left(\frac{1}{6} - \frac{|\nu|}{2}\right)(\xi\,h)^2 + O((\xi\,h)^4)\right]\left(1 - \frac{\nu^2}{3}\,(\xi\,h)^2 + O((\xi\,h)^4)\right)$$
$$= a\left[1 - \frac{1}{6}\,(1 - |\nu|)\,(1 - 2\,|\nu|)\,(\xi\,h)^2 + O((\xi\,h)^4)\right].$$

Vidíme, že fázová chyba je řádu $(\xi\,h)^2$ a znaménko závisí na ν. Pro $|\nu| \in (\frac{1}{2}, 1)$ dochází k předbíhání a pro $|\nu| \in (0, \frac{1}{2})$ ke zpožďování fáze. Pro $|\nu| = \frac{1}{2}$ dostaneme z (3.10)

$$\lambda(\xi) = \frac{1}{2}\,[1 + \cos(\xi\,h) - \mathrm{i}\,2\,\nu\,\sin(\xi\,h)] = \frac{1}{2}\,[1 + \cos(2\,\nu\,\xi\,h) - \mathrm{i}\,\sin(2\,\nu\,\xi\,h)]$$
$$= \frac{1}{2}\left[1 + \cos^2(\nu\,\xi\,h) - \sin^2(\nu\,\xi\,h) - \mathrm{i}\,2\,\sin(\nu\,\xi\,h)\cos(\nu\,\xi\,h)\right]$$
$$= \cos(\nu\,\xi\,h)\,[\cos(\nu\,\xi\,h) - \mathrm{i}\,\sin(\nu\,\xi\,h)] = \cos\left(\frac{\xi\,h}{2}\right)\mathrm{e}^{-\mathrm{i}\,\xi\,a\,\tau}\,,$$

a tudíž chyba fáze je nulová. Pro $|\nu| = 1$ víme, že schéma dává přesné řešení, což plyne i z toho, že podle (3.10) je

$$\lambda(\xi) = \cos(\xi\,h) - \mathrm{i}\,\nu\,\sin(\xi\,h) = \cos(\nu\,\xi\,h) - \mathrm{i}\,\sin(\nu\,\xi\,h) = \mathrm{e}^{-\mathrm{i}\,\xi\,a\,\tau}.$$

Ukažme si to např. pro $\nu = 1$. Pak $\widehat{U}^n(\xi) = \lambda(\xi)^n\,\widehat{U}^0(\xi) = \mathrm{e}^{-\mathrm{i}\,\xi\,n\,h}\,\widehat{U}^0(\xi)$, a tudíž

$$U_j^n = \frac{1}{\sqrt{2\,\pi}}\int_{-\pi/h}^{\pi/h} \widehat{U}^0(\xi)\,\mathrm{e}^{\mathrm{i}\,\xi\,(j-n)\,h}\,\mathrm{d}\xi = U_{j-n}^0 = u^0(x_{j-n}) = u^0(x_j - n\,h)$$
$$= u^0(x_j - a\,t_n) = u(x_j, t_n)\,.$$

Ze vztahu (3.11) vidíme, že kromě případu $|\nu| = 1$ dochází u upwind schématu (3.9) vždy k chybě v amplitudě řádu $(\xi\, h)^2$ v jednom časovém kroku, což vede ke globální chybě řádu $\xi\, h$ (na daném časovém intervalu $[0, T]$). Pro většinu problémů je takovéto tlumení nepřijatelné.

V případě Laxova–Friedrichsova schématu (3.8) je

$$\lambda(\xi) = \cos(\xi\, h) - \mathrm{i}\,\nu\, \sin(\xi\, h)\,, \qquad |\lambda(\xi)|^2 = 1 - (1 - \nu^2)\, \sin^2(\xi\, h)\,.$$

Schéma je tedy stabilní pro $|\nu| \leq 1$, tj. opět pro ty hodnoty ν, pro něž je splněna CFL podmínka. Tlumení je stejného řádu jako u schématu (3.9), avšak pokles amplitudy je více než dvojnásobný (v jednom časovém kroku zhruba $(1 + |\nu|)(1 - |\nu|)(\xi\, h)^2$ namísto $|\nu|(1 - |\nu|)(\xi\, h)^2$ u schématu (3.9)). Analogicky jako výše dostaneme

$$\alpha(\xi) = a \left[1 + \frac{1}{3}\,(1 - \nu^2)\,(\xi\, h)^2 + O((\xi\, h)^4)\right].$$

Kromě případu $|\nu| = 1$, kdy Laxovo–Friedrichsovo schéma je totožné se schématem (3.9) a dává tedy přesné řešení, dochází podle uvedeného vztahu vždy k předbíhání fáze.

3.3 Numerické okrajové podmínky

Dosud se naše teoretické úvahy týkaly pouze numerického řešení Cauchyových úloh. V praxi ovšem obvykle řešíme parciální diferenciální rovnice na omezených prostorových oblastech. Na jejich hranicích (či jejich částech) pak obvykle předepisujeme okrajové podmínky. Může se však stát, že použité numerické schéma vyžaduje okrajovou podmínku na části hranice, kde parciální diferenciální rovnice žádnou okrajovou podmínku nevyžaduje. V takovém případě je nutno předepsat tzv. *numerickou okrajovou podmínku*.

Uvažujme rovnici (3.1) na omezeném prostorovém intervalu (A, B). Je-li $a > 0$ konstatní, pak charakteristiky vycházející z bodů ležících na úsečce $(A, B) \times \{0\}$ protínají polopřímku $\{B\} \times \mathbb{R}^+$, avšak nikoliv polopřímku $\{A\} \times \mathbb{R}^+$. Na polopřímce $\{A\} \times \mathbb{R}^+$ je proto nutno předepsat okrajovou podmínku, zatímco na polopřímce $\{B\} \times \mathbb{R}^+$ je řešení u určeno počáteční podmínkou a okrajovou podmínkou na $\{A\} \times \mathbb{R}^+$. Je-li $a < 0$ konstatní, pak je situace opačná: okrajovou podmínku je třeba předepsat na polopřímce $\{B\} \times \mathbb{R}^+$, zatímco na polopřímce $\{A\} \times \mathbb{R}^+$ je řešení u určeno počáteční podmínkou a okrajovou podmínkou na $\{B\} \times \mathbb{R}^+$. Pokud a není konstantní, může např. každý bod polopřímek $\{A\} \times \mathbb{R}^+$ a $\{B\} \times \mathbb{R}^+$ ležet na nějaké charakteristice vycházející z bodu na $(A, B) \times \{0\}$, a tudíž žádnou okrajovou podmínku nelze předepsat. Nebo naopak žádná charakteristika zmíněné polopřímky neprotíná a pak na obou polopřímkách je potřeba předepsat okrajové podmínky.

Výpočetní oblast $[A, B] \times \mathbb{R}_0^+$ pokryjme rovnoměrnou sítí s prostorovým krokem h a časovým krokem τ. Předpokládáme, že $h = |B - A|/J$, kde $J \in \mathbb{N}$. Pak síť obsahuje uzly (x_j, t_n), kde $x_j = A + j\, h$, $t_n = n\, \tau$, $j = 0, \ldots, J$, $n \in \mathbb{N}_0$. V uzlech (x_j, t_n) s $j = 1, \ldots, J-1$ a $n \in \mathbb{N}_0$ uvažujme Laxovo–Friedrichsovo schéma (3.8). Toto schéma vyžaduje hodnoty U_0^n a U_J^n, bez ohledu na to, zda rovnice (3.1) umožňuje tyto hodnoty předepsat. Je-li např. $a > 0$ konstatní, pak, jak jsme již uvedli, rovnice (3.1) neumožňuje předepsat

okrajovou podmínku podél polopřímky $\{B\} \times \mathbb{R}^+$, a je tedy nutno předepsat numerickou okrajovou podmínku v uzlech (x_J, t_n). Příklady numerických okrajových podmínek v uzlech (x_J, t_n) jsou následující vztahy:

$$U_J^n = U_{J-1}^n \,, \tag{3.13}$$

$$U_J^n = 2\,U_{J-1}^n - U_{J-2}^n \,, \tag{3.14}$$

$$U_J^n = U_{J-1}^{n-1} \,, \tag{3.15}$$

$$U_J^n = 2\,U_{J-1}^{n-1} - U_{J-2}^{n-2} \,. \tag{3.16}$$

Vztahy (3.13) a (3.14) jsou jednoduché extrapolace přibližného řešení z vnitřních uzlů na hranici. Vztahy (3.15) a (3.16) se někdy nazývají kvazicharakteristické extrapolace, neboť využívají hodnot přibližného řešení v uzlech ležících na přímkách procházejících uzly (x_{J-1}, t_{n-1}) a (x_J, t_n), jejichž směr přibližně odpovídá směru charakteristik.

Alternativou k výše uvedenému postupu je zavést fiktivní uzly (x_{-1}, t_n) a (x_{J+1}, t_n), do nichž extrapolujeme přibližné řešení z původních uzlů sítě, a ve všech původních uzlech sítě (tj. i hraničních) použít schéma (3.8). Použijeme-li k extrapolaci vztah (3.14), pak získáme

$$\frac{U_J^{n+1} - \frac{1}{2}\,(U_{J+1}^n + U_{J-1}^n)}{\tau} + a\,\frac{U_{J+1}^n - U_{J-1}^n}{2\,h} = 0 \,, \qquad U_{J+1}^n = 2\,U_J^n - U_{J-1}^n \,,$$

z čehož plyne

$$\frac{U_J^{n+1} - U_J^n}{\tau} + a\,\frac{U_J^n - U_{J-1}^n}{h} = 0 \,.$$

Vidíme tedy, že extrapolace do fiktivních uzlů je v tomto případě ekvivalentní aproximaci diferenciální rovnice v hraničních uzlech pomocí jednostranných diferencí, což je další typ numerické okrajové podmínky. Často je však jednodušší použít některý z extrapolačních vztahů (3.13)–(3.16) než zavádět fiktivní uzly nebo jednostranné diference.

Je důležité zdůraznit, že některé numerické okrajové podmínky mohou pro dané schéma způsobit nestabilní chování, které se projeví narůstajícími oscilacemi v přibližném řešení. Přitom numerická okrajová podmínka, kterou lze použít s daným numerickým schématem, může být s jiným schématem nestabilní a naopak. Pro každou zvolenou kombinaci schématu a numerické okrajové podmínky je proto potřeba stabilitu vyšetřovat zvlášť.

3.4 Laxovo–Wendroffovo schéma

Z jednokrokových schémat, která jsme dosud analyzovali, jsou pro řešení rovnice (3.1) použitelná (při libovolném znaménku a) pouze schémata (3.8) a (3.9). Obě tato schémata jsou pouze prvního řádu přesnosti (v případě schématu (3.8) za podmínky $h = O(\tau)$). Naším cílem je nyní odvodit schéma druhého řádu přesnosti v prostoru i v čase. Pro jednoduchost budeme opět uvažovat konstantní rychlost a. Bude však poučné schéma odvodit pro nenulovou pravou stranu $f = f(x,t)$. Řešíme tedy rovnici

$$u_t + a\,u_x = f \qquad \text{v } \mathbb{R} \times \mathbb{R}^+ \tag{3.17}$$

s počáteční podmínkou (3.2).

Předpokládejme, že $u \in C^4(\mathbb{R} \times \mathbb{R}_0^+)$. Pak pro $(x,t) \in \mathbb{R} \times \mathbb{R}_0^+$ je

$$u(x,t+\tau) = u(x,t) + u_t(x,t)\,\tau + \frac{1}{2}\,u_{tt}(x,t)\,\tau^2 + O(\tau^3)\,. \tag{3.18}$$

Derivace u_t a u_{tt} získáme z rovnice (3.17). V případě u_t je to triviální. Abychom získali u_{tt}, zderivujeme rovnici (3.17) podle t. Tím vznikne člen u_{xt}, který získáme z rovnice (3.17) zderivováním podle x. Máme tedy

$$u_{tt} = f_t - a\,u_{xt} = f_t - a\,(f_x - a\,u_{xx}) = a^2\,u_{xx} - a\,f_x + f_t\,.$$

Dosazením do (3.18) obdržíme

$$u(x,t+\tau) = u - a\,\tau\,u_x + \tau\,f + \frac{a^2\,\tau^2}{2}\,u_{xx} - \frac{a\,\tau^2}{2}\,f_x + \frac{\tau^2}{2}\,f_t + O(\tau^3)\,,$$

kde členy na pravé straně jsou uvažovány v bodě (x,t). Jelikož

$$u_x = \frac{\Delta_{0x}\,u}{h} + O(h^2)\,, \qquad u_{xx} = \frac{\delta_x^2\,u}{h^2} + O(h^2)\,,$$

dostáváme

$$\Delta_{+t}\,u = -\nu\,\Delta_{0x}\,u + O(\tau\,h^2) + \frac{\nu^2}{2}\,\delta_x^2\,u + \frac{\tau}{2}\,[f + f(x,t+\tau)] - \frac{a\,\tau^2}{2\,h}\,\Delta_{0x}\,f + O(\tau^3)\,.$$

To nás vede ke schématu

$$\frac{U_j^{n+1} - U_j^n}{\tau} + a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} - \frac{a^2\,\tau}{2}\,\frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2} = \frac{1}{2}\left(f_j^{n+1} + f_j^n\right) - \frac{a\,\tau}{4\,h}\left(f_{j+1}^n - f_{j-1}^n\right), \tag{3.19}$$

které můžeme též zapsat ve tvaru

$$U_j^{n+1} = U_j^n - \frac{\nu}{2}\,(U_{j+1}^n - U_{j-1}^n) + \frac{\nu^2}{2}\,(U_{j+1}^n - 2\,U_j^n + U_{j-1}^n) + \frac{\tau}{2}\left(f_j^{n+1} + f_j^n\right) - \frac{\nu\,\tau}{4}\left(f_{j+1}^n - f_{j-1}^n\right).$$

Z odvození je zřejmé, že chyba diskretizace splňuje $\varepsilon_{h,\tau} = O(h^2 + \tau^2)$, tj. skutečně jsme získali schéma druhého řádu přesnosti v prostoru i v čase. Pro $f \equiv 0$ je Laxovo–Wendroffovo schéma ekvivalentní přidání umělé difúze o velikosti $\frac{1}{2}\,a^2\,\tau$ k nestabilnímu schématu (3.6). V tomto případě můžeme Laxovo–Wendroffovo schéma odvodit též pomocí charakteristik, podobně jako schéma (3.9), viz obr. 3. Přidáme-li v obr. 3 na přímce $t = t_n$ uzel C tak, že B je střed úsečky AC, a definujeme-li v bodě Q aproximaci u pomocí kvadratické interpolace hodnot přibližného řešení v uzlech A, B a C, získáme Laxovo–Wendroffovo schéma.

Von Neumannovu analýzu provádíme pro $f \equiv 0$. Získáme

$$\lambda(\xi) = 1 - 2\,\nu^2\,\sin^2\frac{\xi\,h}{2} - \mathrm{i}\,\nu\,\sin(\xi\,h)\,, \qquad |\lambda(\xi)|^2 = 1 - 4\,\nu^2\,(1-\nu^2)\,\sin^4\frac{\xi\,h}{2}\,,$$

a schéma je tedy stabilní pro $|\nu| \le 1$, přičemž celý tento interval je povolen CFL podmínkou. Fázová rychlost splňuje

$$\alpha(\xi) = a\left[1 - \frac{1}{6}\,(1-\nu^2)\,(\xi\,h)^2 + O((\xi\,h)^4)\right]$$

a při $|\nu| < 1$ tedy dochází vždy ke zpožďování fáze (pro malé $|\xi h|$). Pro $|\nu| = 1$ je $\lambda(\xi) = \mathrm{e}^{-\mathrm{i}\,\xi\,a\,\tau}$, a schéma tudíž dává přesné řešení. Fázová chyba je stejného řádu jako u schémat (3.8) a (3.9), avšak tlumení je podstatně menší (v jednom časovém kroku je chyba v amplitudě řádu $(\xi\,h)^4$ místo $(\xi\,h)^2$). Pro nejvyšší frekvenci na síti ($\xi = \pi/h$) je $\lambda(\pi/h) = 1 - 2\,\nu^2$, a tedy $\alpha(\pi/h) = 0$, což znamená, že tato rychle oscilující složka řešení se nepohybuje pryč. Při $|\nu| < 1$ je však tato složka řešení tlumena: např. počáteční podmínce $U_j^0 = (-1)^j$ odpovídá řešení $U_j^n = (1 - 2\,\nu^2)^n (-1)^j$. Nedostatkem Laxova–Wendroffova schématu je, že na rozdíl od schémat (3.8) a (3.9) nesplňuje diskrétní princip maxima.

Cvičení 3.2 *Proveďte výpočet ze cvičení 3.1 pro Laxovo–Wendroffovo schéma a srovnejte výsledek s přibližným řešením získaným pomocí Laxova–Friedrichsova schématu* (3.8).

Cvičení 3.3 *Uvažujme rovnici $u_t + u_x = 0$ v oblasti $(-1,1) \times \mathbb{R}^+$ s počáteční podmínkou $u^0(x) = \sin(2\,\pi\,x)$. Definujme ekvidistantní prostorové uzly $-1 = x_0 < x_1 < \cdots < x_J = 1$ s prostorovým krokem $h = 2/J$. Dále definujme uzel $x_{J+1} = 1 + h$ ležící mimo interval $[-1,1]$. Zvolme $J = 20$ a uvažujme konstatní časový krok $\tau = 0.09$. V čase $t = 0$ je přibližné řešení určeno počáteční podmínkou. Na následujících časových hladinách $t_n = n\,\tau$, $n \in \mathbb{N}$ počítejte v uzlech $x_1, \ldots, x_J$ přibližná řešení pomocí schémat* (3.8), (3.9) *a* (3.19). *Pak položte $U_0^n = U_J^n$ a $U_{J+1}^n = U_1^n$. Uvažujeme tedy periodické okrajové podmínky. U zmíněných schémat porovnejte fázovou chybu a tlumení a srovnejte výsledky provedených numerických experimentů s teoretickými výsledky uvedenými výše.*

3.5 Schéma Crankovo–Nicolsonové

Schéma Crankovo–Nicolsonové jsme v úvodní části o metodě konečných diferencí definovali jako aritmetický průměr explicitního a implicitního schématu. Nyní si ukážeme jiný způsob, jak toto schéma odvodit. Naše motivace bude opět získat jednokrokové schéma pro rovnici (3.17), které je druhého řádu v prostoru i v čase.

Odvození bude založeno na tom, že na diferenci $u(x,t+\tau) - u(x,t)$ můžeme pohlížet jako na centrální diferenci vzhledem k času $t + \frac{\tau}{2}$. Je tedy

$$\frac{u(x,t+\tau) - u(x,t)}{\tau} = \frac{\delta_t\,u(x,t+\frac{\tau}{2})}{\tau} = u_t(x,t+\tfrac{\tau}{2}) + O(\tau^2)\,.$$

Dále platí

$$\frac{u(x,t+\tau) + u(x,t)}{2} = u(x,t+\tfrac{\tau}{2}) + \frac{u(x,t+\tau) - 2\,u(x,t+\frac{\tau}{2}) + u(x,t)}{2}\,.$$

Čitatel zlomku na pravé straně představuje centrální diferenci druhého řádu vzhledem k t v bodě $(x, t+\frac{\tau}{2})$ s krokem $\frac{\tau}{2}$. Jelikož $\delta_t^2\, u = O(\tau^2)$, dostáváme

$$\frac{u(x,t+\tau)+u(x,t)}{2} = u(x,t+\tfrac{\tau}{2}) + O(\tau^2)\,.$$

Z toho plyne, že

$$\begin{aligned} u_x(x,t+\tfrac{\tau}{2}) &= \frac{1}{2}\left[u_x(x,t+\tau)+u_x(x,t)\right] + O(\tau^2) \\ &= \frac{1}{2}\left[\frac{u(x+h,t+\tau)-u(x-h,t+\tau)}{2\,h} + \frac{u(x+h,t)-u(x-h,t)}{2\,h}\right] + O(h^2+\tau^2)\,. \end{aligned}$$

To ukazuje, že v bodě $(x, t+\frac{\tau}{2})$ může být vhodné rovnici (3.17) aproximovat schématem

$$\frac{U_j^{n+1}-U_j^n}{\tau} + a\,\frac{U_{j+1}^{n+1}-U_{j-1}^{n+1}+U_{j+1}^n-U_{j-1}^n}{4\,h} = \frac{f_j^{n+1}+f_j^n}{2}\,,$$

což lze psát též ve tvaru

$$-\frac{\nu}{4}\,U_{j-1}^{n+1} + U_j^{n+1} + \frac{\nu}{4}\,U_{j+1}^{n+1} = \frac{\nu}{4}\,U_{j-1}^n + U_j^n - \frac{\nu}{4}\,U_{j+1}^n + \frac{\tau}{2}\left(f_j^{n+1}+f_j^n\right).$$

Nyní již není možné hodnotu přibližného řešení v daném uzlu na nové časové vrstvě jednoduchým způsobem vyjádřit pomocí hodnot přibližného řešení na předcházející časové vrstvě a jedná se tedy o implicitní schéma. Hodnoty přibližného řešení na nové časové vrstvě jsou navzájem svázány a získají se vyřešením soustavy lineárních rovnic. Jelikož hodnota přibližného řešení v daném uzlu na nové časové vrstvě závisí na všech hodnotách přibližného řešení z předcházející časové vrstvy, nevede CFL podmínka k žádné podmínce na parametr ν.

Von Neumannova analýza (pro $f \equiv 0$) dává amplifikační faktor

$$\lambda(\xi) = \frac{1 - \mathrm{i}\,\frac{\nu}{2}\,\sin(\xi\,h)}{1 + \mathrm{i}\,\frac{\nu}{2}\,\sin(\xi\,h)}\,.$$

Je tedy $|\lambda(\xi)| = 1$ pro všechna $\xi \in \mathbb{R}$, a tudíž schéma je nepodmíněně stabilní. Nedochází však k žádnému tlumení, takže drobné poruchy v řešení zůstávají trvale přítomny. Speciálně pro vlnu s nejvyšší frekvencí π/h opět platí, že fázová rychlost je nulová, avšak na rozdíl od Laxova–Wendroffova schématu se její amplituda v čase nemění: počáteční podmínce $U_j^0 = (-1)^j$ odpovídá řešení $U_j^n = (-1)^j$. Stejně jako u Laxova–Wendroffova schématu neplatí ani pro schéma Crankovo–Nicolsonové diskrétní princip maxima.

3.6 Analýza leapfrog scheme

Názvem *leapfrog scheme* je označováno dvoukrokové schéma (3.7). Je zřejmé, že se jedná o schéma druhého řadu přesnosti v prostoru i v čase, které nesplňuje diskrétní princip maxima. CFL podmínka je splněna, jestliže $|\nu| \le 1$, což budeme nadále předpokládat.

Zabývejme se nyní stabilitou leapfrog scheme. Stejně jako v případě jednokrokových schémat vyjádříme přibližné řešení U_j^n v integrálním tvaru pomocí jeho Fourierovy transformace $\widehat{U}^n(\xi)$ a dosadíme do schématu. Tím získáme vztah

$$\int_{-\pi/h}^{\pi/h} \mathrm{e}^{\mathrm{i}\,\xi\,j\,h} \left(\widehat{U}^{n+1}(\xi) + \mathrm{i}\,2\,\nu\,\sin(\xi\,h)\,\widehat{U}^n(\xi) - \widehat{U}^{n-1}(\xi) \right) \mathrm{d}\xi = 0 \quad \forall\, j \in \mathbb{Z}\,,\ n \in \mathbb{N}\,,$$

z něhož díky tomu, že funkce $\{\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}\}_{j\in\mathbb{Z}}$ tvoří úplný ortogonální systém, plyne

$$\widehat{U}^{n+1}(\xi) + \mathrm{i}\,2\,\nu\,\sin(\xi\,h)\,\widehat{U}^n(\xi) - \widehat{U}^{n-1}(\xi) = 0 \quad \forall\, n \in \mathbb{N}\,,\ \xi \in [-\pi/h, \pi/h]\,. \tag{3.20}$$

Pro každé $\xi \in [-\pi/h, \pi/h]$ tedy máme soustavu lineárních diferenčních rovnic s konstantními koeficienty. Charakteristická rovnice této soustavy diferenčních rovnic je

$$\lambda^2 + \mathrm{i}\,2\,\nu\,\sin(\xi\,h)\,\lambda - 1 = 0\,.$$

Její kořeny jsou

$$\lambda_\pm(\xi) = -\mathrm{i}\,\nu\,\sin(\xi\,h) \pm \sqrt{1 - \nu^2\,\sin^2(\xi\,h)}$$

a platí $|\lambda_\pm(\xi)| = 1$. Pokud $\lambda_+(\xi) = \lambda_-(\xi) =: \lambda(\xi)$, pak obecné řešení soustavy diferenčních rovnic (3.20) má tvar

$$\widehat{U}^n(\xi) = \alpha(\xi)\,\lambda(\xi)^n + \beta(\xi)\,n\,\lambda(\xi)^{n-1}\,,$$

kde $\alpha(\xi)$ a $\beta(\xi)$ jsou konstanty určené hodnotami $\widehat{U}^0(\xi)$ a $\widehat{U}^1(\xi)$. Je-li $\beta(\xi) \neq 0$, bude $\widehat{U}^n(\xi)$ růst lineárně s n. Toto nestabilní chování je třeba vyloučit. Jelikož $\lambda_+(\xi) = \lambda_-(\xi)$ může nastat pouze při $|\nu| = 1$, budeme požadovat, aby $|\nu| < 1$. Pak $\lambda_+(\xi) \neq \lambda_-(\xi)$ a obecné řešení soustavy diferenčních rovnic (3.20) je

$$\widehat{U}^n(\xi) = \alpha_+(\xi)\,\lambda_+(\xi)^n + \alpha_-(\xi)\,\lambda_-(\xi)^n\,. \tag{3.21}$$

Snadno zjistíme, že schéma je v tomto případě stabilní. Nutná a postačující podmínka pro stabilitu leapfrog scheme tedy je $|\nu| < 1$.

Leapfrog scheme (3.7) nelze použít k určení přibližného řešení na časové vrstvě t_1 a schéma je tedy potřeba inicializovat pomocí vhodného jednokrokového schématu. Lze ukázat, že k inicializaci vícekrokových schémat lze použít libovolné jednokrokové schéma, které je konzistentní s řešenou parciální diferenciální rovnicí. Toto schéma může být i nestabilní, neboť aplikací schématu pouze v několika prvních krocích dojde jen k malému růstu řešení (malému díky konzistenci) a tento růst nebude při použití stabilního vícekrokového schématu dále zesilován. Jeli $\tau/h = const.$, pak inicializační schéma může mít o 1 nižší řád přesnosti než vícekrokové schéma, přičemž celková přesnost bude odpovídat vícekrokovému schématu.

Ze vztahu (3.21) plyne

$$\widehat{U}^n(\xi) = \widehat{U}^0(\xi)\,\lambda_+(\xi)^n + \gamma(\xi)\,\frac{\lambda_+(\xi)^n - \lambda_-(\xi)^n}{\lambda_+(\xi) - \lambda_-(\xi)}\,,$$

kde $\gamma(\xi) = \widehat{U}^1(\xi) - \widehat{U}^0(\xi)\,\lambda_+(\xi)$. Použijeme-li k inicializaci schéma (3.6), je $\widehat{U}^1(\xi) = (1 - \mathrm{i}\,\nu\,\sin(\xi\,h))\,\widehat{U}^0(\xi)$. Použitím vztahu $\sqrt{1-x} = 1 - \frac{1}{2}\,x + O(x^2)$ získáme

$$\lambda_\pm(\xi) = -\mathrm{i}\,\nu\,\sin(\xi\,h) \pm 1 \mp \frac{1}{2}\,\nu^2\,\sin^2(\xi\,h) + O((\xi\,h)^4)\,,$$

a tudíž

$$\gamma(\xi) = (1 - \mathrm{i}\,\nu\,\sin(\xi\,h) - \lambda_+(\xi))\,\widehat{U}^0(\xi) = \left(\frac{\nu^2}{2}\,\sin^2(\xi\,h) + O((\xi\,h)^4)\right)\widehat{U}^0(\xi)\,.$$

Vidíme, že pro malé hodnoty $|\xi\,h|$ je $\gamma(\xi)$ malé, a schéma se tudíž chová jako jednokrokové schéma s amplifikačním faktorem λ_+. Pro větší hodnoty $|\xi\,h|$ velikost $\gamma(\xi)$ být malá nemusí. Část řešení příslušná λ_- se nazývá parazitická složka řešení. Jelikož $\lambda_-(0) = -1$, tato složka řešení v průběhu času rychle osciluje. Později ukážeme, že se parazitická složka řešení pohybuje špatným směrem (pro $a > 0$ se pohybuje zprava doleva).

Cvičení 3.4 *Uvažujte ve cvičení 3.1 okrajovou podmínku $U_J^n = 0$, která není konzistentní s řešenou rovnicí, a proveďte výpočet pro leapfrog scheme. Měli byste pozorovat, že v důsledku zmíněné okrajové podmínky vznikne v bodě $x = 3$ parazitická složka řešení, která je silně oscilující a pohybuje se zprava doleva. Okrajová podmínka v bodě $x = -2$ přemění parazitickou složku na neparazitickou.*

Parazitická složka řešení se objeví při každém výpočtu s vícekrokovými schématy. Obvykle nezpůsobuje podstatné potíže, avšak v některých případech je ji potřeba redukovat nebo odstranit. Vliv parazitických složek lze redukovat pomocí disipace.

Definice 3.1 *Schéma je disipativní řádu $2\,r$, jestliže amplifikační faktory splňují podmínku*

$$|\lambda_j(\xi)| \le 1 - C\left(\sin\frac{\xi\,h}{2}\right)^{2\,r} \qquad \forall\ \xi \in \mathbb{R}\,,$$

kde C je kladná konstanta nezávislá na h a τ.

Pro leapfrog scheme a schéma Crankovo–Nicolsonové je velikost amplifikačních faktorů rovna 1 pro všechny frekvence. Říkáme, že tato schémata jsou *ostře nedisipativní*. Pro Laxovo–Friedrichsovo schéma je $|\lambda(\pi/h)| = 1$, avšak $|\lambda(\xi)| < 1$ pro $0 < |\xi| < \pi/h$ a $|\nu| < 1$. Toto schéma je proto nedisipativní, ale ne ostře. Redukuje amplitudu u většiny frekvencí, nikoli však u nejvyšší frekvence na síti.

Všimněme si, že síťová funkce $U_j^n = (-1)^{n+j}\,\eta$ je pro libovolné $\eta \in \mathbb{R}$ řešením leapfrog scheme. Vidíme tedy, že počáteční poruchy se šíří, aniž by byly tlumeny. O přítomnosti této šachovnicové složky řešení rozhodují pouze počáteční a okrajové podmínky.

Vlastnosti nedisipativního schématu lze zlepšit přidáním disipace. Je však třeba dát pozor, abychom tím nezhoršili řád přesnosti schématu. Např. leapfrog scheme lze modifikovat takto:

$$\frac{U_j^{n+1} - U_j^{n-1}}{2\,\tau} + a\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} + \frac{\varepsilon}{2\,\tau}\left(\frac{1}{2}\,\delta_x\right)^4 U_j^{n-1} = 0\,,$$

kde ε je vhodná kladná konstanta. Jelikož $\delta_x^4\, u = O(h^4)$, schéma je nadále druhého řádu přesnosti, pokud $\tau \geq C\, h^2$. Tato podmínka nepředstavuje žádné podstatné omezení, neboť časový krok volíme obvykle srovnatelný s prostorovým krokem. Pro vyšetření stability přepíšeme schéma do tvaru

$$U_j^{n+1} - U_j^{n-1} + \nu\, U_{j+1}^n - \nu\, U_{j-1}^n + \varepsilon \left(\frac{1}{2}\,\delta_x\right)^4 U_j^{n-1} = 0$$

a přibližné řešení vyjádříme v integrálním tvaru pomocí jeho Fourierovy transformace $\widehat{U}^n(\xi)$. Oproti případu bez disipace se navíc objeví člen $\varepsilon\,(\frac{1}{2}\,\delta_x)^4\,\widehat{U}^{n-1}(\xi)\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}$. Jak víme, $\delta_x^2\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h} = -4\,\sin^2\frac{\xi\,h}{2}\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}$, a tudíž $(\frac{1}{2}\,\delta_x)^4\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h} = \sin^4\frac{\xi\,h}{2}\,\mathrm{e}^{\mathrm{i}\,\xi\,j\,h}$. Místo soustavy diferenčních rovnic (3.20), proto nyní získáme

$$\widehat{U}^{n+1}(\xi) + \mathrm{i}\,2\,\nu\,\sin(\xi\,h)\,\widehat{U}^n(\xi) - \left(1 - \varepsilon\,\sin^4\frac{\xi\,h}{2}\right)\widehat{U}^{n-1}(\xi) = 0\,.$$

Kořeny odpovídající charakteristické rovnice jsou

$$\lambda_\pm(\xi) = -\mathrm{i}\,\nu\,\sin(\xi\,h) \pm \sqrt{1 - \varepsilon\,\sin^4\frac{\xi\,h}{2} - \nu^2\,\sin^2(\xi\,h)}\,.$$

Je-li

$$\nu^2\,\sin^2(\xi\,h) + \varepsilon\,\sin^4\frac{\xi\,h}{2} \leq 1\,, \tag{3.22}$$

je

$$|\lambda_\pm(\xi)| = \sqrt{1 - \varepsilon\,\sin^4\frac{\xi\,h}{2}} \leq 1 - \frac{\varepsilon}{2}\,\sin^4\frac{\xi\,h}{2}\,,$$

a schéma je tedy stabilní a disipativní řádu 4 (k tomu je opět nutné, aby $|\nu| < 1$, neboť jinak (3.22) nemůže platit). Podmínku (3.22) můžeme přepsat do tvaru

$$4\,\nu^2\,\sin^2\frac{\xi\,h}{2}\left(1 - \sin^2\frac{\xi\,h}{2}\right) + \varepsilon\,\sin^4\frac{\xi\,h}{2} \leq 1\,.$$

Stačí tedy nalézt takové hodnoty ε, že

$$4\,\nu^2\,s\,(1-s) + \varepsilon\,s^2 \leq 1 \qquad \forall\ s \in [0,1]\,.$$

Je-li $\nu^2 \in (0, \frac{1}{2}]$, můžeme volit $\varepsilon \in (0,1]$, neboť pak $4\,\nu^2\,s\,(1-s) + \varepsilon\,s^2 \leq 2\,s\,(1-s) + s^2 = s\,(2-s) \leq 1$. Je-li $\nu^2 \in [\frac{1}{2}, 1)$, volíme $\varepsilon \in (0, 4\,\nu^2\,(1-\nu^2)]$. Pak $4\,\nu^2\,s\,(1-s) + \varepsilon\,s^2 = 4\,\nu^2\,s + (\varepsilon - 4\,\nu^2)\,s^2 \leq 4\,\nu^2\,s\,(1-\nu^2\,s) \leq 1$.

Podobně jako výše můžeme modifikovat i metodu Crankovu–Nicolsonové pro řešení rovnice (3.17). Obdržíme schéma

$$\frac{U_j^{n+1} - U_j^n}{\tau} + a\,\frac{U_{j+1}^{n+1} - U_{j-1}^{n+1} + U_{j+1}^n - U_{j-1}^n}{4\,h} + \frac{\varepsilon}{\tau}\left(\frac{1}{2}\,\delta_x\right)^4 U_j^n = \frac{f_j^{n+1} + f_j^n}{2}\,,$$

které je druhého řádu přesnosti (pokud $\tau \geq C\,h^2$) a disipativní řádu 4 pro malé hodnoty ε.

Cvičení 3.5 *Zopakujte výpočet ze cvičení 3.4 pro leapfrog scheme s disipací (v prostorových uzlech x_1 a x_{J-1} uvažujte leapfrog scheme bez disipace).*

Pro vícekroková schémata vždy existuje právě jeden amplifikační faktor $\lambda_0(\xi)$ takový, že $\lambda_0(0) = 1$. Tento amplifikační faktor použijeme k definici fázové rychlosti. Pro leapfrog scheme je to amplifikační faktor λ_+. Platí

$$\sin(\xi\,\alpha(\xi)\,\tau) = \nu\,\sin(\xi\,h) = \nu\,\xi\,h\left(1 - \frac{(\xi\,h)^2}{6} + O((\xi\,h)^4)\right),$$

z čehož plyne

$$\begin{aligned}\alpha(\xi) &= a\left(1 - \frac{(\xi\,h)^2}{6} + O((\xi\,h)^4)\right)\left(1 + \frac{(\nu\,\xi\,h)^2}{6} + O((\xi\,h)^4)\right)\\ &= a\left(1 - \frac{1}{6}\,(1-\nu^2)\,(\xi\,h)^2 + O((\xi\,h)^4)\right).\end{aligned}$$

Získali jsme tedy totéž jako pro Laxovo-Wendroffovo schéma a rovněž platí $\alpha(\pi/h) = 0$.

Pro parazitickou složku řešení leapfrog scheme odpovídající λ_- zavedeme fázovou rychlost vztahem $\lambda_-(\xi) = -|\lambda_-(\xi)|\,\mathrm{e}^{-\mathrm{i}\,\xi\,\alpha_-(\xi)\,\tau}$, neboť $\lambda_-(0) = -1$. Pak $\sin(\xi\,\alpha_-(\xi)\,\tau) = -\nu\,\sin(\xi\,h)$, a tudíž $\alpha_-(\xi) = -\alpha(\xi)$. Parazitická složka řešení se tedy pohybuje opačným směrem než je správný směr řešení. Navíc platí $\alpha_-(\pi/h) = 0$, a tudíž nejvyšší frekvence, které neobsahují přesnou informaci, se nešíří pryč.

3.7 Volba parametru ν

Vlastnosti schémat vyšetřovaných v předchozích odstavcích závisí na volbě parametru ν, tj. na poměru prostorového a časového kroku. V řadě případů jsme viděli, že pro $|\nu| = 1$ schémata dávají přesné řešení, což je ovšem dáno jednoduchostí uvažované modelové úlohy. Pro úlohy s nekonstantními koeficienty nebo komplikovanější problémy k tomu již nedochází. Nicméně naše teoretické úvahy ukazují, že je obecně vhodné volit $|\nu|$ blízko hranice stability, neboť tím dosáhneme malé disipace i disperze. Zajímá-li nás pouze určitá frekvence ξ_0, měli bychom volit h tak, aby $|\xi_0\,h| \ll \pi$ a vlna o frekvenci ξ_0 mohla být tudíž na použité síti dobře aproximována.

4 Numerické řešení rovnice vedení tepla

V této části se budeme zabývat numerickým řešením smíšené úlohy pro rovnici vedení tepla v jedné prostorové dimenzi. Hledáme funkci $u = u(x,t)$ definovanou pro $x \in [0,1]$ a $t \geq 0$ takovou, že

$$u_t - u_{xx} = 0 \qquad \text{v } (0,1)\times\mathbb{R}^+, \tag{4.1}$$

$$u(0,t) = u(1,t) = 0 \qquad \forall\ t > 0, \tag{4.2}$$

$$u(x,0) = u^0(x) \qquad \forall\ x \in [0,1]. \tag{4.3}$$

Budeme předpokládat, že úloha (4.1)–(4.3) má klasické řešení. To mimo jiné vyžaduje, aby počáteční a okrajové podmínky splňovaly podmínky kompatibility $u^0(0) = u^0(1) = 0$.

Úlohy uvedeného typu popisují šíření tepla v tenké izolované homogenní tyči konečné délky bez přítomnosti tepelných zdrojů. Úloha typu (4.1)–(4.3) též popisuje šíření tepla napříč nekonečnou deskou, přičemž x je souřadnice kolmá ke stěnám desky, na každé z nichž je předepsána konstantní teplota.

Poznámka 4.1 Transformací $v(x,t) = u(x,\kappa\,t) + \alpha\,(1-x) + \beta\,x$ získáme úlohu s tepelnou difuzivitou κ a nehomogenními Dirichletovými okrajovými podmínkami:

$$\begin{aligned} &v_t - \kappa\,v_{xx} = 0 \qquad \text{v } (0,1)\times\mathbb{R}^+\,,\\ &v(0,t) = \alpha, \quad v(1,t) = \beta \qquad \forall\ t > 0\,,\\ &v(x,0) = v^0(x) \equiv u^0(x) + \alpha\,(1-x) + \beta\,x \qquad \forall\ x \in [0,1]\,. \end{aligned}$$

Uvažování $\kappa = 1$ a homogenních okrajových podmínek v (4.1)–(4.3) tedy nepředstavuje újmu na obecnosti. Podobně lze též jednoduchou transformací proměnné x přejít k úloze definované na libovolném zadaném prostorovém intervalu.

4.1 Řešení úlohy (4.1)–(4.3) Fourierovou metodou

Hledejme řešení úlohy (4.1)–(4.3) ve tvaru $u(x,t) = X(x)\,T(t)$. Dosazením do (4.1) a (4.2) snadno zjistíme, že

$$u(x,t) = a_m\,\mathrm{e}^{-(m\,\pi)^2\,t}\,\sin(m\,\pi\,x)$$

pro nějaké $m \in \mathbb{N}$ a $a_m \in \mathbb{R}$. Rovnice (4.1) a (4.2) splňuje zřejmě i libovolná lineární kombinace těchto funkcí. Nabízí se tedy hledat řešení u ve tvaru

$$u(x,t) = \sum_{m=1}^{\infty} a_m\,\mathrm{e}^{-(m\,\pi)^2\,t}\,\sin(m\,\pi\,x)\,. \tag{4.4}$$

Z počáteční podmínky (4.3) plyne

$$u^0(x) = \sum_{m=1}^{\infty} a_m\,\sin(m\,\pi\,x)\,, \tag{4.5}$$

a tudíž koeficienty a_m jsou Fourierovy koeficienty funkce u^0 při rozvoji do sinové řady. Platí

$$a_m = 2\int_0^1 u^0(x)\,\sin(m\,\pi\,x)\,\mathrm{d}x\,.$$

Je známo, že pro libovolné $u^0 \in L^2(0,1)$ řada (4.5) konverguje v $L^2(0,1)$. Součet řady (4.4) je pak nekonečně hladká funkce v $[0,1]\times\mathbb{R}^+$, která řeší parciální diferenciální rovnici (4.1) a splňuje okrajové podmínky (4.2). Pokud řada Fourierových koeficientů a_m konverguje absolutně, je součet řady (4.4) spojitá funkce na $[0,1]\times\mathbb{R}_0^+$, která splňuje počáteční podmínku (4.3). K tomu stačí, aby funkce u^0 byla absolutně spojitá na $[0,1]$, $(u^0)' \in L^2(0,1)$ a $u^0(0) = u^0(1) = 0$ (konzistence s okrajovou podmínkou).

V praxi výsledek (4.4) umožňuje získat pouze numerickou aproximaci řešení u, neboť koeficienty a_m jsme obecně schopni určit pouze přibližně a navíc jsme schopni sečíst pouze konečně mnoho členů řady. Skutečným omezením uvedené metody však je, že ji nelze snadno zobecnit na komplikovanější úlohy. Je proto nutné hledat jiné způsoby výpočtu přibližného řešení úlohy (4.1)–(4.3) a jedním z možných postupů je aplikace metody konečných diferencí.

4.2 Explicitní schéma pro úlohu (4.1)–(4.3)

Podobně jako dříve zavedeme rovnoměrnou síť, kterou pokryjeme výpočetní oblast. Interval $[0,1]$ nejprve rozdělíme na $J \in \mathbb{N}$ intervalů stejné délky $h = 1/J$, čímž vzniknou body $x_j = j\,h$, $j = 0, 1, 2, \ldots, J$. Kromě prostorového kroku sítě h zavedeme též časový krok sítě $\tau > 0$ a definujeme časové hladiny $t_n = n\,\tau$, $n = 0, 1, 2, \ldots$. Řešení u úlohy (4.1)–(4.3) aproximujeme v uzlech sítě (x_j, t_n) hodnotami U_j^n, kde $j = 0, 1, 2, \ldots, J$ a $n = 0, 1, 2, \ldots$. Opět položíme $u_j^n = u(x_j, t_n)$.

Uvažujeme-li rovnici (4.1) v uzlu (x_j, t_n) s $j \in \{1, \ldots, J-1\}$ a $n \in \mathbb{N}_0$, můžeme provést následující aproximace:

$$0 = (u_t - u_{xx})(x_j, t_n) \approx \frac{\Delta_{+t}\, u_j^n}{\tau} - \frac{\delta_x^2\, u_j^n}{h^2} \approx \frac{\Delta_{+t}\, U_j^n}{\tau} - \frac{\delta_x^2\, U_j^n}{h^2}\,.$$

To vede k diferenčním rovnicím

$$U_j^{n+1} = U_j^n + \mu\,(U_{j+1}^n - 2\,U_j^n + U_{j-1}^n)\,, \qquad j = 1, 2, \ldots, J-1\,, \tag{4.6}$$

kde

$$\mu = \frac{\tau}{h^2}\,.$$

Každou hodnotu na časové hladině t_{n+1} lze nezávisle spočítat z hodnot na časové hladině t_n a jedná se tedy o explicitní diferenční schéma.

K rovnicím (4.6) musíme ještě přidat okrajové a počáteční podmínky

$$U_0^n = U_J^n = 0\,, \qquad n = 1, 2, \ldots\,, \tag{4.7}$$

$$U_j^0 = u^0(x_j)\,, \qquad j = 0, 1, \ldots, J\,. \tag{4.8}$$

Hodnoty přibližného řešení získáme tak, že nejprve definujeme hodnoty U_j^0 pomocí (4.8) a pak evolučně počítáme hodnoty přibližného řešení na následujících časových hladinách z rovnic (4.6) a okrajových podmínek (4.7). Přibližné řešení U_j^n je tedy vztahy (4.6)–(4.8) jednoznačně určeno.

Cvičení 4.1 *Uvažujte schéma* (4.6)–(4.8) *pro úlohu* (4.1)–(4.3) *s počáteční podmínkou*

$$u^0(x) = \begin{cases} 2\,x & \text{pro} \quad x \in [0, \frac{1}{2}]\,, \\ 2 - 2\,x & \text{pro} \quad x \in [\frac{1}{2}, 1]\,. \end{cases}$$

Zvolte $J = 20$ *a proveďte výpočet pro* $\tau = 0.0012$ *a* $\tau = 0.0013$.

Provedeme-li výpočty z cvičení 4.1, zjistíme, že pro $\tau = 0.0012$ získáme dobrou aproximaci řešení úlohy (4.1)–(4.3), zatímco pro $\tau = 0.0013$ se v přibližném řešení objeví oscilace, které se zvětšujícím se n rychle rostou. Jedná se o typický příklad stability či nestability numerického schématu. Jak uvidíme později, numerické výsledky zásadně závisejí na hodnotě μ, tj. na vztahu mezi prostorovým a časovým krokem.

Chyba diskretizace schématu (4.6) je dána vztahem

$$\varepsilon_j^n = \frac{\Delta_{+t}\, u_j^n}{\tau} - \frac{\delta_x^2\, u_j^n}{h^2}$$

pro $j \in \{1, \ldots, J-1\}$ a $n \in \mathbb{N}_0$. Je-li $x \in [h, 1-h]$ a $t \geq 0$, můžeme též položit

$$\varepsilon_{h,\tau}(x,t) = \frac{\Delta_{+t}\, u(x,t)}{\tau} - \frac{\delta_x^2\, u(x,t)}{h^2}\,.$$

Pak $\varepsilon_j^n = \varepsilon_{h,\tau}(x_j, t_n)$. Použitím Taylorova vzorce získáme

$$\varepsilon_{h,\tau}(x,t) = \tfrac{1}{2}\, u_{tt}(x,\eta)\, \tau - \tfrac{1}{12}\, u_{xxxx}(\xi,t)\, h^2\,, \qquad \xi \in (x-h, x+h),\ \eta \in (t, t+\tau)\,.$$

Předpokládáme-li, že existují konstanty M_1 a M_2 takové, že $|u_{tt}| \leq M_1$, $|u_{xxxx}| \leq M_2$ na $[0,1] \times [0,T]$, kde $T > 0$ je pevně zvolený čas, pak

$$|\varepsilon_{h,\tau}(x,t)| \leq \tfrac{1}{2}\, M_1\, \tau + \tfrac{1}{12}\, M_2\, h^2 = \tfrac{1}{2}\, \tau \left(M_1 + \tfrac{1}{6\,\mu}\, M_2 \right) \quad \forall\ x \in [h, 1-h] \times [0, T-\tau]\,. \tag{4.9}$$

Vidíme tedy, že schéma je prvního řádu přesnosti v čase a druhého řádu přesnosti v prostoru. Pro pevný poměr μ (který je vhodné uvažovat pro zajištění stability, jak uvidíme později) se $\varepsilon_{h,\tau}$ chová jako $O(\tau)$ pro $\tau \to 0$ a v tomto smyslu je schéma prvního řádu přesnosti.

Poznámka 4.2 Jelikož $u_t = u_{xx}$, je $u_{tt} = u_{xxt} = (u_t)_{xx} = u_{xxxx}$, a tudíž

$$\varepsilon_{h,\tau}(x,t) = \tfrac{1}{2} \left(1 - \tfrac{1}{6\,\mu} \right) u_{xxxx}(x,t)\, \tau + O(\tau^2)\,.$$

Pro $\mu = \frac{1}{6}$ je tedy schéma druhého řádu přesnosti. Jedná se ale o velmi speciální případ, se kterým se v obecnějších situacích nesetkáme.

4.3 Konvergence explicitního schématu

Věta 4.1 *Uvažujme posloupnost $(h_i, \tau_i) \to (0,0)$ pro $i \to \infty$ a předpokládejme, že $\mu_i \equiv \tau_i / h_i^2 \leq \frac{1}{2}$. Nechť $T > 0$ a $|u_{tt}| \leq M_1$, $|u_{xxxx}| \leq M_2$ v $[0,1] \times [0,T]$. Pak pro libovolný bod $(x,t) \in [0,1] \times [0,T]$ a libovolnou posloupnost $(j_i, n_i) \in \mathbb{N} \times \mathbb{N}$ takovou, že $j_i\, h_i \to x$, $n_i\, \tau_i \to t$, konvergují aproximace $U_{j_i}^{n_i}$ generované explicitním schématem (4.6)–(4.7) k řešení $u(x,t)$, přičemž tato konvergence je stejnoměrná v $[0,1] \times [0,T]$.*

Důkaz. Uvažujme libovolnou dvojici h, τ ($\tau < T$) a libovolný bod $(x_j, t_n) \in (0,1) \times (0,T)$. Pro chybu aproximace $e_j^n = U_j^n - u_j^n$ platí

$$e_j^{n+1} = e_j^n + \mu\,[e_{j+1}^n - 2\,e_j^n + e_{j-1}^n] - \tau\,\varepsilon_j^n\,, \qquad j = 1, \ldots, J-1\,.$$

Definujeme-li $\|e^n\|_\infty = \max_{l=1,\ldots,J-1} |e_l^n|$, pak (díky $e_0^n = e_J^n = 0$)

$$\|e^{n+1}\|_\infty \le (|1 - 2\,\mu| + 2\,\mu)\,\|e^n\|_\infty + \tau\,\|\varepsilon^n\|_\infty \le \|e^n\|_\infty + \tau\,\|\varepsilon^n\|_\infty\,.$$

Jelikož $e_l^0 = U_l^0 - u^0(x_l) = 0$, $l = 0, \ldots, J$, dostáváme

$$\|e^n\|_\infty \le \tau \sum_{k=0}^{n-1} \|\varepsilon^k\|_\infty \le t_n \max_{k=0,\ldots,n-1} \|\varepsilon^k\|_\infty\,.$$

Použitím (4.9) získáváme

$$|U_j^n - u(x_j, t_n)| \le T\,(\tfrac{1}{2}\,M_1\,\tau + \tfrac{1}{12}\,M_2\,h^2)\,.$$

Tvrzení nyní plyne ze spojitosti u na $[0,1] \times [0,T]$. □

4.4 Fourierova analýza chyby

Víme, že řešení úlohy (4.1)–(4.3) lze zapsat ve tvaru Fourierovy řady (4.4). Ukážeme, že v podobném tvaru lze zapsat i přibližné řešení splňující (4.6)–(4.8). Za tím účelem zapíšeme řady (4.4) a (4.5) pomocí komplexních exponenciel:

$$u(x,t) = \sum_{m=-\infty}^{\infty} A_m\,\mathrm{e}^{\mathrm{i}\,m\,\pi\,x - (m\,\pi)^2\,t}\,, \tag{4.10}$$

$$u^0(x) = \sum_{m=-\infty}^{\infty} A_m\,\mathrm{e}^{\mathrm{i}\,m\,\pi\,x}\,. \tag{4.11}$$

Položíme-li $u^0(x) = -u^0(-x)$ pro $x \in [-1, 0)$, pak

$$A_m = \frac{1}{2} \int_{-1}^{1} u^0(x)\,\mathrm{e}^{-\mathrm{i}\,m\,\pi\,x}\,\mathrm{d}x\,.$$

Snadno se ověří, že pro $m \in \mathbb{N}$ je $A_m = -A_{-m} = -\mathrm{i}\,a_m/2$. Budeme předpokládat, že Fourierova řada (4.11) je absolutně konvergentní, tj.

$$\sum_{m=-\infty}^{\infty} |A_m| < \infty\,. \tag{4.12}$$

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 (4.1). 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 jsme 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 (4.6). Dosazením do (4.6) 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

$$\lambda \equiv \lambda(k) = 1-2\,\mu\,[1-\cos(k\,h)] = 1-4\,\mu\,\sin^2\frac{k\,h}{2}\,. \tag{4.13}$$

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 4.2 *Nechť síťová funkce U_j^n splňuje* (4.6)–(4.8). *Pak*

$$U_j^n = \sum_{m=-\infty}^{\infty} A_m\,\mathrm{e}^{\mathrm{i}\,m\,\pi\,j\,h}\,[\lambda(m\,\pi)]^n\,, \qquad j=0,1,\ldots,J\,, \quad n\geq 0\,. \tag{4.14}$$

Důkaz. Jelikož amplifikační faktory jsou omezené a platí (4.12), konverguje řada (4.14) absolutně, a jelikož každý její člen řeší diferenční rovnici (4.6), řeší i součet řady (4.14) rovnici (4.6). 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 (4.7). Konečně pro $n=0$ se řada (4.14) redukuje na řadu (4.11) s $x=j\,h$, a součet řady tudíž splňuje i počáteční podmínku (4.8). Součet řady tedy splňuje (4.6)–(4.8), a jelikož je řešení úlohy (4.6)–(4.8) určeno jednoznačně, platí (4.14). □

Poznámka 4.3 Všimněme si, že z vlastností amplifikačního faktoru a koeficientů A_m plyne, že řadu (4.14) lze též přepsat do tvaru

$$U_j^n = \sum_{m=1}^{\infty} a_m\,\sin(m\,\pi\,j\,h)\,[\lambda(m\,\pi)]^n\,, \qquad j=0,1,\ldots,J\,, \quad n\geq 0\,. \tag{4.15}$$

Ze vztahu (4.10) 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 (4.14), 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 4.1 *Platí*

$$|\mathrm{e}^{-k^2\tau} - \lambda(k)| \le C(\mu)\, k^4\, \tau^2 \qquad \forall\ k, \tau > 0\,, \tag{4.16}$$

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 (4.13) 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 (4.16). 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 (4.10) a (4.12) platí

$$\begin{aligned} \varepsilon_j^n &= \frac{u_j^{n+1} - u_j^n}{\tau} - \frac{u_{j+1}^n - 2\,u_j^n + u_{j-1}^n}{h^2} \\ &= \sum_{m=-\infty}^{\infty} A_m\, \mathrm{e}^{\mathrm{i}\,m\,\pi\,x_j - (m\,\pi)^2\,t_n} \left[\frac{\mathrm{e}^{-(m\,\pi)^2\,\tau} - 1}{\tau} - \frac{\mathrm{e}^{\mathrm{i}\,m\,\pi\,h} - 2 + \mathrm{e}^{-\mathrm{i}\,m\,\pi\,h}}{h^2} \right] \\ &= \sum_{m=-\infty}^{\infty} A_m\, \mathrm{e}^{\mathrm{i}\,m\,\pi\,x_j - (m\,\pi)^2\,t_n}\, \frac{1}{\tau}\, \left[\mathrm{e}^{-(m\,\pi)^2\,\tau} - \lambda(m\,\pi) \right]. \end{aligned}$$

Použitím odhadu (4.16) 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 (4.9). Pro $h^2 = 6\,\tau$ lze pomocí lemmatu 4.1 získat přesnost druhého řádu (srv. pozn. 4.2).

Vyjádření řešení úlohy (4.1)–(4.3) ve tvaru řady (4.10) 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 (4.1)–(4.3) 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 4.3 *Nechť pro dané hodnoty h a τ je $|\lambda(m\,\pi)| \le 1$ pro všechna $m \in \mathbb{N}$. Pak řešení schématu* (4.6)–(4.8) *zůstává pro libovolnou počáteční podmínku u^0 splňující* (4.12) *omezené pro $n \to \infty$.*

Důkaz. Jelikož $|\lambda(m\,\pi)| \le 1 \;\forall\; m \in \mathbb{N}$, je dle (4.14) a (4.12)

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

□

Poznámka 4.4 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 (4.15) 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 (4.14) 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 (4.6) 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 (4.1)–(4.3) nyní bude absolutní konvergence řady (4.11) 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

$$|\lambda_1^n - \lambda_2^n| \le n\,|\lambda_1 - \lambda_2| \qquad \forall\ \lambda_1, \lambda_2 \in [-1,1]\,. \tag{4.17}$$

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 (4.17) 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 (4.16) 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 (4.14), 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

$$\mathrm{e}^{\mathrm{i}\,m\,\pi\,j\,h}\,[\lambda(m\,\pi)]^n\,. \tag{4.18}$$

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 (4.18) odpovídajících

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

4.5 Implicitní metoda pro úlohu (4.1)–(4.3)

Podmínka stability $\tau \le h^2/2$ pro explicitní schéma (4.6) 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 (4.1), 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

$$-\mu\, U_{j-1}^{n+1} + (1+2\,\mu)\, U_j^{n+1} - \mu\, U_{j+1}^{n+1} = U_j^n\,, \qquad j = 1, 2, \ldots, J-1\,, \quad n \geq 0\,. \tag{4.19}$$

Na rozdíl od schématu (4.6) 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 (4.19) s okrajovými a počátečními podmínkami (4.7) a (4.8) můžeme vyšetřovat Fourierovou metodou analogicky jako pro schéma (4.6). Snadno zjistíme, že funkce $U_j^n = \mathrm{e}^{\mathrm{i}\,k\,j\,h}\,\lambda^n$ splňuje (4.19) 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 (4.19) 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

$$a_j > 0\,, \qquad b_j > 0\,, \qquad c_j > 0\,, \qquad b_j > a_j + c_j\,, \tag{4.20}$$

což splňuje (4.19). Soustavu rovnic nejprve převedeme na soustavu s horní trojúhelníkovou maticí tvaru

$$U_j - e_j\, U_{j+1} = f_j\,, \qquad j = 1, 2, \ldots, J-1\,. \tag{4.21}$$

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 (4.21) 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

$\mathrm{e}_1 := \mathrm{c}_1/\mathrm{b}_1,\ \ \mathrm{f}_1 := \mathrm{d}_1/\mathrm{b}_1$
`for` $\mathrm{j} = 2, \ldots, \mathrm{J}-2$ `do`

```
    e_j := c_j/(b_j − a_j e_{j−1})
    f_j := (d_j + a_j f_{j−1})/(b_j − a_j e_{j−1})
enddo
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 (4.21).

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 (4.19) na jeden uzel tři sčítání, tři násobení a dvě dělení, zatímco explicitní schéma (4.6) 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 (4.1)–(4.3) 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 (3.7). To vede v nejjednodušším případě ke schématu

$$\frac{U_j^{n+1} - U_j^{n-1}}{2\,\tau} = \frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2}\,, \qquad j = 1, 2, \ldots, J-1\,, \quad n \geq 1\,. \tag{4.22}$$

Snadno zjistíme, že chyba diskretizace je v tomto případě druhého řádu v čase i v prostoru. Jelikož k určení přibližného řešení na nové časové vrstvě využívá toto schéma hodnot přibližného řešení z předcházejících dvou časových vrstev, jedná se o dvoukrokové schéma.

Hodnoty přibližného řešení v čase t_1, které je nutno předepsat nebo vypočítat pomocí nějakého jednokrokového schématu, zapíšeme ve tvaru

$$U_j^1 = \sum_{m=-\infty}^{\infty} B_m\, \mathrm{e}^{\mathrm{i}\,m\,\pi\,j\,h}\,, \qquad j = 0, 1, \ldots, J\,, \tag{4.23}$$

a opět předpokládáme, že řada konverguje absolutně a že $B_m = -B_{-m}\ \forall\ m \in \mathbb{Z}$. Podobně jako při Fourierově analýze jednokrokových metod výše hledejme nejprve řešení diferenčního schématu ve tvaru se "separovanými proměnnými", tj. položme

$$U_j^n = \mathrm{e}^{\mathrm{i}\,k\,j\,h}\,\widehat{U}^n(k)\,,$$

kde $k = m\,\pi$, $m \in \mathbb{Z}$. Dosazením do (4.22) získáme

$$\widehat{U}^{n+1}(k) + 2\,q(k)\,\widehat{U}^n(k) - \widehat{U}^{n-1}(k) = 0\,, \quad n \geq 1\,, \quad \text{kde } q(k) = \frac{4\,\tau}{h^2}\,\sin^2\frac{k\,h}{2}\,. \tag{4.24}$$

Charakteristický polynom $\lambda^2+2\,q(k)\,\lambda-1$ této soustavy diferenčních rovnic má dva různé reálné kořeny $\lambda_\pm = -q(k) \pm \sqrt{q(k)^2+1}$, a tudíž obecné řešení soustavy (4.24) je

$$\widehat{U}^n(k) = \alpha_+(k)\,\lambda_+(k)^n + \alpha_-(k)\,\lambda_-(k)^n\,, \tag{4.25}$$

kde koeficienty $\alpha_\pm$ jsou libovolná komplexní čísla. Tato čísla určíme tak, aby platilo $\widehat{U}^0(m\,\pi) = A_m$ a $\widehat{U}^1(m\,\pi) = B_m$. Z toho plyne

$$\alpha_+(m\,\pi) = \frac{B_m - A_m\,\lambda_-(m\,\pi)}{\lambda_+(m\,\pi) - \lambda_-(m\,\pi)}\,, \qquad \alpha_-(m\,\pi) = \frac{A_m\,\lambda_+(m\,\pi) - B_m}{\lambda_+(m\,\pi) - \lambda_-(m\,\pi)}\,. \tag{4.26}$$

Řešení diskrétního problému (4.22), (4.7), (4.8), (4.23) je pak dáno řadou

$$U_j^n = \sum_{m=-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\,m\,\pi\,j\,h}\,\widehat{U}^n(m\,\pi)\,, \qquad j = 0, 1, \ldots, J\,, \quad n \geq 0\,. \tag{4.27}$$

Bohužel složky řešení odpovídající kořenu λ_- jsou nestabilní, neboť $\lambda_-(k) < -1$, kdykoli $\sin(\frac{1}{2}\,k\,h) \neq 0$. Schéma (4.22) je tedy bezcenné, neboť je pro libovolnou volbu h a τ nestabilní.

Poznamenejme, že pokud charakteristický polynom soustavy diferenčních rovnic odpovídajících dvoukrokovému schématu má pro dané k jeden dvojnásobný kořen, tj. $\lambda_+(k) = \lambda_-(k) \equiv \lambda(k)$, má obecné řešení soustavy diferenčních rovnic tvar

$$\widehat{U}^n(k) = \alpha(k)\,\lambda(k)^n + \beta(k)\,n\,\lambda(k)^{n-1}\,,$$

kde $\alpha(k), \beta(k) \in \mathbb{C}$. V tomto případě tedy je

$$\widehat{U}^n(m\,\pi) = A_m\,\lambda(m\,\pi)^n + [B_m - A_m\,\lambda(m\,\pi)]\,n\,\lambda(m\,\pi)^{n-1}\,. \tag{4.28}$$

Pokud je koeficient u druhého členu nenulový, je ke stabilitě nutné, aby $|\lambda(m\,\pi)| < 1$, neboť jinak $\widehat{U}^n$ poroste lineárně v n.

Uvedená analýza schématu (4.22) samozřejmě neznamená, že každé dvoukrokové explicitní schéma je vždy nestabilní. Uvažujme např. schéma

$$\frac{U_j^{n+1} - U_j^{n-1}}{2\,\tau} = \frac{\delta_x^2\,U_j^n + \delta_x^2\,U_j^{n-1}}{2\,h^2}\,, \qquad j = 1, 2, \ldots, J-1\,, \quad n \geq 1\,. \tag{4.29}$$

V tomto případě získáme soustavu diferenčních rovnic

$$\widehat{U}^{n+1}(k) + q(k)\,\widehat{U}^n(k) + (q(k)-1)\,\widehat{U}^{n-1}(k) = 0\,, \qquad n \geq 1\,, \tag{4.30}$$

jejíž charakteristický polynom má kořeny $\lambda_+(k) = 1 - q(k)$ a $\lambda_-(k) = -1$. Je-li $q(k) \neq 2$, jsou oba kořeny různé a $\widehat{U}^n(k)$ je opět dáno vztahy (4.25) a (4.26). Zřejmě je $|\lambda_\pm(k)| \leq 1$ právě tehdy, když $q(k) \in [0, 2)$, z čehož plyne podmínka stability $\tau \leq h^2/2$. Je-li $q(k) = 2$, je $\lambda_+(k) = \lambda_-(k)$ a $\widehat{U}^n(k)$ je dáno vztahem (4.28) s $\lambda(k) = -1$. Případ $q(m\,\pi) = 2$ za uvedené podmínky stability může nastat pouze tehdy, je-li $m = (2\,l+1)J$, kde $l \in \mathbb{Z}$. Členy

s $m = (2\,l + 1)J$ se však v řešení neprojeví, neboť řadu (4.27) můžeme díky vlastnosti $\widehat{U}^n(m\,\pi) = -\widehat{U}^n(-m\,\pi)\ \forall\ m \in \mathbb{Z}$ zapsat ve tvaru

$$U_j^n = 2\,\mathrm{i}\sum_{m=1}^{\infty} \sin(m\,\pi\,j\,h)\,\widehat{U}^n(m\,\pi)\,, \qquad j = 0, 1, \ldots, J\,, \quad n \geq 0\,.$$

Schéma (4.29) je tedy pro $\tau \leq h^2/2$ stabilní.

Stabilita metody však nemusí znamenat, že diskrétní řešení bude bez oscilací. Jak jsme viděli, platí v případě $q(k) \neq 2$

$$\widehat{U}^n(k) = \alpha_+(k)\,(1 - q(k))^n + \alpha_-(k)\,(-1)^n\,.$$

Označíme-li

$$V_j^n = \sum_{\substack{m=-\infty \\ q(m\pi)<2}}^{\infty} \mathrm{e}^{\mathrm{i}\,m\,\pi\,j\,h}\,\alpha_+(m\,\pi)\,(1 - q(m\,\pi))^n\,, \qquad W^j = \sum_{\substack{m=-\infty \\ q(m\pi)<2}}^{\infty} \mathrm{e}^{\mathrm{i}\,m\,\pi\,j\,h}\,\alpha_-(m\,\pi)\,,$$

můžeme psát

$$U_j^n = V_j^n + (-1)^n\,W_j\,, \qquad j = 0, 1, \ldots, J\,, \quad n \geq 0\,.$$

Síťová funkce W_j nezávisí na čase, a pokud je nenulová, bude pro velké n představovat dominantní složku řešení, neboť $V_j^n \to 0$ pro $n \to \infty$. Obvykle se projeví ve formě oscilací, jejichž velikost může být i větší než jsou hodnoty počáteční podmínky, jelikož velikost $\alpha_+(m\,\pi)$ a $\alpha_-(m\,\pi)$ může být podstatně větší než $|A_m|$. Ze stability metody však plyne, že tyto oscilace nebudou pro $n \to \infty$ narůstat. Navíc se jejich velikost zmenší, pokud zjemníme síť.

Časově nezávislá oscilující složka nebude v řešení schématu (4.29) přítomna, pokud řešení v čase t_1 určíme pomocí schématu (4.6). Podle (4.14) a (4.13) je pak totiž $B_m = A_m\,(1 - q(m\,\pi)) = A_m\,\lambda_+(m\,\pi)$, a tudíž $\alpha_-(m\,\pi) = 0$ pro každé $m \in \mathbb{Z}$.

4.8 Disipace

Z diskuse v předchozím odstavci plyne, že je žádoucí, aby diferenční schéma vedlo v každém časovém kroku k poklesu amplitud vysokofrekvenčních složek řešení, tj. aby velikost příslušných amplifikačních faktorů byla menší než 1. Jak již víme z části 3.6, tento pokles vysokofrekvenčních oscilací se nazývá disipace. Analogicky jako v definici 3.1 na str. 32 řekneme, že diferenční schéma je disipativní řádu $2\,r$ (má disipaci řádu $2\,r$), jestliže existuje kladná konstanta C nezávislá na h a τ taková, že každý amplifikační faktor $\lambda_j(k)$ splňuje

$$|\lambda_j(k)| \leq 1 - C\left(\sin\frac{k\,h}{2}\right)^{2\,r} \tag{4.31}$$

pro všechna vlnová čísla k. Poznamenejme, že pro každé k uvažujeme obecně více amplifikačních faktorů, abychom do našich úvah zahrnuli i vícekroková schémata. Pro ověřování

platnosti podmínky (4.31) může být užitečné si všimnout, že nerovnost (4.31) je ekvivalentní podmínce

$$|\lambda_j(k)|^2 \le 1 - C' \left(\sin \frac{k\,h}{2} \right)^{2\,r}$$

s konstantou C' nezávislou na h a τ. Disipativnost schématu je v případě parabolických rovnic velmi přirozeným požadavkem, neboť pak dochází v průběhu času ke zhlazování diskrétního řešení, stejně jako je tomu u řešení aproximované diferenciální rovnice.

Použitím vztahu (4.13) snadno zjistíme, že explicitní schéma (4.6) je disipativní řádu 2 pro $\mu \in [\mu_0, \mu_1] \subset (0, \frac{1}{2})$, kde μ_0 a μ_1 jsou konstanty nezávislé na h a τ. Proto se obvykle nepoužívá volba $\mu = \frac{1}{2}$, při níž je schéma (4.6) ještě stabilní. Implicitní schéma (4.19) je disipativní řádu 2 , je-li $\mu \ge \mu_0 > 0$, kde μ_0 je konstanta nezávislá na h a τ. Schéma (4.29) samozřejmě disipativní není.

4.9 θ–metoda (θ–schéma, metoda váženého průměru)

Uvažujeme-li vážený průměr explicitního schématu (4.6) a implicitního schématu (4.19), získáme šestibodové schéma

$$U_j^{n+1} - U_j^n = \mu \left[\theta\, \delta_x^2\, U_j^{n+1} + (1-\theta)\, \delta_x^2\, U_j^n \right], \qquad j = 1, 2, \ldots, J-1\,, \;\; n \ge 0\,. \tag{4.32}$$

Budeme předpokládat, že $\theta \in [0, 1]$. Pro $\theta = 0$ získáváme explicitní schéma (4.6) a pro $\theta = 1$ plně implicitní schéma (4.19). Pro libovolné $\theta \in (0, 1]$ je k určení hodnot přibližného řešení v čase t_{n+1} nutno vyřešit tridiagonální soustavu lineárních rovnic

$$-\theta\,\mu\, U_{j-1}^{n+1} + (1 + 2\,\theta\,\mu)\, U_j^{n+1} - \theta\,\mu\, U_{j+1}^{n+1} = \left[1 + (1-\theta)\,\mu\,\delta_x^2 \right] U_j^n\,, \qquad j = 1, 2, \ldots, J-1\,.$$

Koeficienty splňují nerovnosti (4.20), a můžeme tedy použít Thomasův algoritmus.

Stabilitu vyšetříme opět pomocí Fourierovy metody. Dosazením $U_j^n = \mathrm{e}^{\mathrm{i}\,k\,j\,h}\,\lambda^n$ do (4.32) získáme

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

Zřejmě $\lambda \le 1$. Nestabilita se může objevit pouze pokud $\lambda < -1$, což nastane právě tehdy, když

$$4\,(1 - 2\,\theta)\,\mu\,\sin^2 \frac{k\,h}{2} > 2\,.$$

Z toho plyne, že

$$\begin{cases} \text{Je-li } \theta \in [0, \frac{1}{2}), \text{ pak (4.32) je stabilní } \iff \mu \le \dfrac{1}{2\,(1-2\,\theta)}\,. \\ \\ \text{Je-li } \theta \in [\frac{1}{2}, 1], \text{ pak (4.32) je stabilní } \forall\ \mu > 0\,. \end{cases} \tag{4.33}$$

V prvním případě je tedy schéma podmíněně stabilní, v druhém nepodmíněně stabilní.

Chybu diskretizace schématu (4.32) je vhodné počítat v čase $t_{n+1/2} \equiv (n + \frac{1}{2})\,\tau$, tj. definujeme

$$\varepsilon_j^{n+1/2} = \varepsilon_{h,\tau}(x_j, t_{n+1/2}) = \frac{u(x_j, t_{n+1}) - u(x_j, t_n)}{\tau} - \frac{\theta\,\delta_x^2\, u(x_j, t_{n+1}) + (1-\theta)\,\delta_x^2\, u(x_j, t_n)}{h^2}\,.$$

Tedy

$$\begin{aligned}\varepsilon_{h,\tau}(x,t) &= \frac{\delta_t\, u(x,t)}{\tau} - \frac{\theta\,\delta_x^2\, u(x,t+\frac{\tau}{2}) + (1-\theta)\,\delta_x^2\, u(x,t-\frac{\tau}{2})}{h^2} \\ &= \frac{\delta_t\, u(x,t)}{\tau} - (\theta-\tfrac{1}{2})\,\frac{\delta_t\,\delta_x^2\, u(x,t)}{h^2} - \frac{\delta_x^2\, u(x,t+\frac{\tau}{2}) + \delta_x^2\, u(x,t-\frac{\tau}{2})}{2\,h^2}\,.\end{aligned}$$

Dosazením Taylorových rozvojů získáme

$$\begin{aligned}\varepsilon_{h,\tau}(x,t) &= [u_t + \tfrac{1}{24}\,u_{ttt}\,\tau^2 + \dots] - (\theta-\tfrac{1}{2})\,[u_{xxt}\,\tau + \tfrac{1}{12}\,u_{xxxxt}\,h^2\,\tau + \dots] \\ &\quad -[u_{xx} + \tfrac{1}{12}\,u_{xxxx}\,h^2 + \tfrac{2}{6!}\,u_{xxxxxx}\,h^4 + \tfrac{1}{8}\,u_{xxtt}\,\tau^2 + \dots]\,.\end{aligned}$$

Použitím (4.1) zjišťujeme, že obecně $\varepsilon_{h,\tau} = O(\tau+h^2)$, avšak pro $\theta = \frac{1}{2}$ je $\varepsilon_{h,\tau} = O(\tau^2+h^2)$. Pro $\theta = \frac{1}{2}$ je tedy schéma (4.32) druhého řádu přesnosti v prostoru i v čase a, jak víme, nazývá se schéma Crankovo–Nicolsonové. Jelikož je nepodmíněně stabilní, můžeme uvažovat $h = O(\tau)$. Pak $\varepsilon_{h,\tau} = O(\tau^2)$ a jsme tedy schopni dosáhnout dobrou přesnost při malé výpočetní náročnosti. Při volbě $h = O(\tau)$ však schéma Crankovo–Nicolsonové není disipativní, což způsobuje, že při nehladké počáteční podmínce může být méně přesné než plně implicitní schéma (4.19), které je disipativní řádu 2.

Metodu druhého řádu přesnosti v čase lze získat též pro

$$\theta = \frac{1}{2} - \frac{h^2}{12\,\tau}\,, \quad \text{tj.} \qquad \mu = \frac{1}{6\,(1-2\,\theta)}\,.$$

(Musí být $h^2 \le 6\,\tau$, aby bylo $\theta \ge 0$.) Při této volbě je dle (4.33) schéma (4.32) stabilní a platí $\varepsilon_{h,\tau} = O(\tau^2 + h^4)$. Opět tedy můžeme používat velké časové kroky a metoda bude přitom pro hladké počáteční podmínky přesná a stabilní. Při volbě $h = O(\tau)$ však schéma opět není disipativní a pro malé h je blízké schématu Crankovu–Nicolsonové.

I když lze odvodit řadu dalších schémat pro řešení úlohy (4.1)–(4.3), nejpoužívanější je v praxi schéma (4.32). Nejlepší volba parametru θ však závisí na řešeném problému a často není jasné, které schéma je opravdu nejlepší.

4.10 Princip maxima a konvergence

Věta 4.4 *θ–schéma* (4.32) *s $\theta \in [0,1]$ a $\mu\,(1-\theta) \le \frac{1}{2}$ dává přibližné řešení $\{U_j^n\}$ splňující*

$$U_{\min}^n \le U_j^n \le U_{\max}^n\,,$$

kde

$$\begin{aligned}U_{\min}^n &= \min\{U_0^m,\ 0 \le m \le n\,;\ U_j^0,\ 0 \le j \le J\,;\ U_J^m,\ 0 \le m \le n\}\,, \\ U_{\max}^n &= \max\{U_0^m,\ 0 \le m \le n\,;\ U_j^0,\ 0 \le j \le J\,;\ U_J^m,\ 0 \le m \le n\}\,.\end{aligned}$$

Důkaz. Schéma (4.32) zapíšeme ve tvaru

$$\begin{aligned}(4.34)\qquad &(1+2\,\theta\,\mu)\,U_j^{n+1} \\ &\quad = \theta\,\mu\,(U_{j-1}^{n+1} + U_{j+1}^{n+1}) + (1-\theta)\,\mu\,(U_{j-1}^n + U_{j+1}^n) + [1-2\,(1-\theta)\,\mu]\,U_j^n\,.\end{aligned}$$

Koeficienty na pravé straně jsou nezáporné a jejich součet je $(1+2\,\theta\,\mu)$ (koeficienty před dvojčleny počítáme dvakrát). Předpokládejme, že U nabývá svého maxima na množině $[0,1]\times[0,t_{n+1}]$ v uzlu (x_j,t_{n+1}), kde $j\in\{1,\dots,J-1\}$. Pak hodnoty U na pravé straně vztahu (4.34) jsou menší nebo rovny U_j^{n+1}, a jelikož součet koeficientů je $(1+2\,\theta\,\mu)$, musí být $U=U_j^{n+1}$ v každém z pěti sousedních uzlů v (4.34), pokud příslušný koeficient je nenulový. Je-li tedy $\theta\neq 0$, dostáváme $U_j^{n+1}=U_0^{n+1}=U_J^{n+1}$. Je-li $\theta=0$, můžeme zkonstruovat posloupnost bodů, až dosáhneme hranice. Tedy $U_j^{n+1}=U_{\max}^{n+1}$. Stejným způsobem lze postupovat pro minimum. □

Věta 4.5 *Uvažujme posloupnost $(h_i,\tau_i)\to(0,0)$ pro $i\to\infty$ a nechť $\mu_i\,(1-\theta)\le\frac12$. Nechť chyba diskretizace odpovídající schématu* (4.32) *konverguje k nule stejnoměrně v množině $[0,1]\times[0,T]$. Nechť chyby v okrajových a počátečních podmínkách rovněž konvergují stejnoměrně k nule pro $i\to\infty$. Pak aproximace dané schématem* (4.32) *konvergují stejnoměrně v $[0,1]\times[0,T]$ k řešení rovnice* (4.1) *s konzistentními okrajovými a počátečními podmínkami.*

Důkaz. Dle definice chyby diskretizace je pro $e_j^n=U_j^n-u(x_j,t_n)$

$$(4.35)\quad \begin{aligned}(1+2\,\theta\,\mu)\,e_j^{n+1}&=\theta\,\mu\,(e_{j-1}^{n+1}+e_{j+1}^{n+1})+(1-\theta)\,\mu\,(e_{j-1}^n+e_{j+1}^n)\\&\quad+[1-2\,(1-\theta)\,\mu]\,e_j^n-\tau\,\varepsilon_j^{n+1/2}\,,\qquad j=1,2,\dots,J-1,\ \ n=0,1,2,\dots\,.\end{aligned}$$

Předpokládejme nejprve, že $e_j^0=0$, $j=0,\dots,J$, $e_0^n=e_J^n=0$, $n=0,1,2,\dots$ a označme

$$\|e^n\|_\infty=\max_{j=0,\dots,J}|e_j^n|\,,\qquad \|\varepsilon^{n+1/2}\|_\infty=\max_{j=1,\dots,J-1}|\varepsilon_j^{n+1/2}|\,.$$

Pak

$$(1+2\,\theta\,\mu)\,\|e^{n+1}\|_\infty\le 2\,\theta\,\mu\,\|e^{n+1}\|_\infty+\|e^n\|_\infty+\tau\,\|\varepsilon^{n+1/2}\|_\infty\,,$$

a tudíž $\|e^{n+1}\|_\infty\le\|e^n\|_\infty+\tau\,\|\varepsilon^{n+1/2}\|_\infty$, z čehož plyne

$$\|e^n\|_\infty\le\tau\sum_{m=0}^{n-1}\|\varepsilon^{m+1/2}\|_\infty\le n\,\tau\max_{m=0,\dots,n-1}\|\varepsilon^{m+1/2}\|_\infty\to 0\quad\text{pro}\quad i\to\infty\,.$$

Předpokládejme nyní, že chyby v okrajových a počátečních podmínkách jsou nenulové, tj.

$$(4.36)\qquad e_j^0=\eta_j^0\,,\quad j=0,\dots,J\,,\qquad e_0^n=\eta_0^n\,,\ \ e_J^n=\eta_J^n\,,\quad n=0,1,2,\dots\,.$$

Pak $e_j^n=\bar e_j^n+\tilde e_j^n$, kde $\bar e_j^n$ splňuje (4.35) s homogenními počátečními a okrajovými podmínkami a $\tilde e_j^n$ splňuje (4.34) a (4.36). Pak $\|\bar e^n\|_\infty$ splňuje předchozí odhad a $\|\tilde e^n\|_\infty\le\max\{|\eta_0^m|,\ 0\le m\le n\,;\ \ |\eta_j^0|,\ 0\le j\le J\,;\ \ |\eta_J^m|,\ 0\le m\le n\}$ dle předchozí věty. □

Podmínka pro platnost principu maxima je mnohem více omezující než podmínka stability plynoucí z Fourierovy analýzy. Například pro $\theta=\frac12$ dostáváme $\mu\le 1$.

Princip maxima představuje alternativní prostředek pro získání podmínek stability. Oproti Fourierově analýze má tu výhodu, že ho lze snadno aplikovat i na úlohy s nekonstantními koeficienty. Avšak snadné je odvodit pouze postačující podmínky stability.

4.11 Obecnější okrajové podmínky

Nahraďme Dirichletovu okrajovou podmínku v bodě $x = 0$ okrajovou podmínkou

$$u_x(0,t) = \alpha(t)\, u(0,t) + g(t) \qquad \forall\ t > 0\,, \tag{4.37}$$

kde $\alpha(t) \geq 0$.

Nejjednodušší aproximace okrajové podmínky (4.37) v čase $t = t_n$ je

$$\frac{U_1^n - U_0^n}{h} = \alpha^n\, U_0^n + g^n\,, \qquad \alpha^n \equiv \alpha(t_n), \quad g^n \equiv g(t_n)\,,$$

z čehož plyne

$$U_0^n = \beta^n\, U_1^n - \beta^n\, g^n\, h\,, \qquad \text{kde} \qquad \beta^n = \frac{1}{1+\alpha^n\, h}\,. \tag{4.38}$$

Nyní můžeme definovat θ–schéma stejným způsobem jako výše. Soustava je opět tridiagonální a má nyní J rovnic. Rovnice (4.38) je první rovnicí této soustavy.

Z (4.38) plyne, že v prvním vnitřním bodě je

$$\delta_x^2\, U_1^n = U_2^n - 2\, U_1^n + U_0^n = U_2^n - (2-\beta^n)\, U_1^n - \beta^n\, g^n\, h\,.$$

Dostáváme tedy

$$\begin{aligned} U_1^{n+1} - U_1^n = {} & \mu\,\theta\,\{U_2^{n+1} - (2-\beta^{n+1})\, U_1^{n+1} - \beta^{n+1}\, g^{n+1}\, h\} \\ & + \mu\,(1-\theta)\,\{U_2^n - (2-\beta^n)\, U_1^n - \beta^n\, g^n\, h\}\,, \end{aligned}$$

z čehož plyne

$$\begin{aligned} [1 + \theta\,\mu\,(2-\beta^{n+1})]\, U_1^{n+1} = {} & [1 - (1-\theta)\,\mu\,(2-\beta^n)]\, U_1^n + \theta\,\mu\, U_2^{n+1} + (1-\theta)\,\mu\, U_2^n \\ & - \mu\, h\,[\theta\,\beta^{n+1}\, g^{n+1} + (1-\theta)\,\beta^n\, g^n]\,. \end{aligned}$$

Definujeme-li obvyklým způsobem chybu diskretizace $\varepsilon_1^{n+1/2}$, platí pro chybu aproximace $e_1^n = U_1^n - u(x_1, t_n)$

$$[1 + \theta\,\mu\,(2-\beta^{n+1})]\, e_1^{n+1} = [1 - (1-\theta)\,\mu\,(2-\beta^n)]\, e_1^n + \theta\,\mu\, e_2^{n+1} + (1-\theta)\,\mu\, e_2^n - \tau\,\varepsilon_1^{n+1/2}\,.$$

Jelikož se tato rovnice liší od rovnic v ostatních uzlech sítě, nemůžeme použít Fourierovu analýzu chyby. Lze však využít princip maxima, neboť pro $\mu\,(1-\theta) \leq \frac{1}{2}$ jsou všechny koeficienty nezáporné a součet koeficientů napravo není větší než koeficient nalevo. Můžeme proto odhadnout chybu aproximace pomocí chyby diskretizace stejně jako výše.

Zbývá odhadnout $\varepsilon_1^{n+1/2}$. Uvažujme případ $\theta = 0$ (explicitní metoda). Pak

$$\begin{aligned} \frac{U_1^{n+1} - U_1^n}{\tau} &= \frac{\delta_x^2\, U_1^n}{h^2} + \frac{1}{h^2}\,[-U_0^n + \beta^n\, U_1^n - \beta^n\, g^n\, h] \\ &= \frac{\delta_x^2\, U_1^n}{h^2} + \frac{\beta^n}{h}\left[\frac{U_1^n - U_0^n}{h} - \alpha^n\, U_0^n - g^n\right]. \end{aligned}$$

Tedy (při označení $u_j^n = u(x_j, t_n)$)

$$\begin{aligned}
\varepsilon_1^{n+1/2} &= \frac{u_1^{n+1} - u_1^n}{\tau} - \frac{\delta_x^2\, u_1^n}{h^2} - \frac{\beta^n}{h}\left[\frac{\Delta_{-x}\, u_1^n}{h} - \alpha^n\, u_0^n - g^n\right] \\
&= [u_t + \tfrac{1}{2}\, u_{tt}\, \tau + \ldots](x_1, t_n) - [u_{xx} + \tfrac{1}{12}\, u_{xxxx}\, h^2 + \ldots](x_1, t_n) \\
&\quad - \frac{\beta^n}{h}\, [u_x + \tfrac{1}{2}\, u_{xx}\, h + \ldots - \alpha\, u - g](x_0, t_n) \\
&\approx -\tfrac{1}{2}\, \beta^n\, u_{xx}(x_0, t_n)\,.
\end{aligned}$$

Chyba diskretizace tudíž nekonverguje k nule. Důkaz konvergence chyby aproximace lze sice zachránit, avšak diskrétní řešení je zatíženo poměrně velkou chybou.

Zkusme jiný postup. Zavedeme fiktivní hodnotu U_{-1}^n vně $[0, 1]$, takže okrajovou podmínku (4.37) můžeme aproximovat vztahem

$$\frac{U_1^n - U_{-1}^n}{2\, h} = \alpha^n\, U_0^n + g^n\,. \tag{4.39}$$

V bodě $x = 0$ aproximujeme rovnici jako ve vnitřních bodech a za U_{-1}^n dosadíme z (4.39). Pak

$$\begin{aligned}
U_0^{n+1} - U_0^n = \mu\, \theta\, [U_1^{n+1} - 2\, U_0^{n+1} + (U_1^{n+1} - 2\, h\, \alpha^{n+1}\, U_0^{n+1} - 2\, h\, g^{n+1})] \\
+ \mu\, (1 - \theta)\, [U_1^n - 2\, U_0^n + (U_1^n - 2\, h\, \alpha^n\, U_0^n - 2\, h\, g^n)]\,,
\end{aligned}$$

z čehož plyne

$$\begin{aligned}
[1 + 2\, \theta\, \mu\, (1 + \alpha^{n+1}\, h)]\, U_0^{n+1} = [1 - 2\, (1 - \theta)\, \mu\, (1 + \alpha^n\, h)]\, U_0^n \\
+ 2\, \theta\, \mu\, U_1^{n+1} + 2\, (1 - \theta)\, \mu\, U_1^n - 2\, \mu\, h\, [\theta\, g^{n+1} + (1 - \theta)\, g^n]\,.
\end{aligned}$$

Pokud $\mu\, (1 - \theta)\, (1 + \alpha^n\, h) \le \frac{1}{2}$, je možno chybu aproximace opět odhadnout pomocí chyby diskretizace postupem založeným na principu maxima. Chyba diskretizace je v tomto případě $\varepsilon_0^{n+1/2} = O(\tau + h)$.

4.12 Obecnější lineární rovnice

Uvažujme nejdříve rovnici

$$u_t = b\, u_{xx} \qquad \forall\ t > 0, \quad x \in (0, 1)\,,$$

kde $b = b(x, t) > 0$. Explicitnímu schématu (4.6) pak odpovídá diskretizace

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

kde $b_j^n = b(x_j, t_n)$. Stejně jako dříve získáme

$$\varepsilon_{h,\tau}(x, t) = \tfrac{1}{2}\, u_{tt}\, \tau - \tfrac{1}{12}\, b(x, t)\, u_{xxxx}\, h^2 + \ldots\,.$$

Konvergenci lze dokázat stejným způsobem jako pro $b = 1$, ale podmínku stability je třeba nahradit podmínkou

$$\frac{\tau}{h^2}\, b(x,t) \le \frac{1}{2}\,.$$

Odhad chyby pak je

$$|U_j^n - u(x_j,t_n)| \le T\,(\tfrac{1}{2}\,M_1\,\tau + \tfrac{1}{12}\,B\,M_2\,h^2)\,,$$

kde $B \ge b(x,t)\ \forall\ (x,t) \in [0,1]\times[0,T]$.

θ–schéma lze definovat různými způsoby. Jednou možností je uvažovat

$$U_j^{n+1} - U_j^n = \frac{\tau}{h^2}\, b^*\,[\theta\,\delta_x^2\,U_j^{n+1} + (1-\theta)\,\delta_x^2\,U_j^n]\,,$$

kde b^* je nějaká vhodná hodnota. Nabízí se položit $b^* = b_j^{n+1/2}$. Rozvoj chyby diskretizace je pak stejný jako dříve až na přenásobení faktorem b. Rovněž konvergenci lze dokázat jako dříve pomocí principu maxima, avšak potřebujeme, aby

$$\frac{\tau}{h^2}\,(1-\theta)\,b(x,t) \le \frac{1}{2}\,.$$

Je též možné položit $b^* = \frac{1}{2}(b_j^{n+1} + b_j^n)$, což nezhorší odhad chyby diskretizace, neboť $b^* = [b + \frac{1}{4}\,b_{tt}\,\tau^2 + \dots](x_j, t_{n+1/2})$.

Nejobecnější tvar lineární parabolické rovnice druhého řádu je

$$u_t = b\,u_{xx} - a\,u_x + c\,u + d \qquad \forall\ t > 0, \quad x \in (0,1)\,, \tag{4.40}$$

kde $a = a(x,t)$, $b = b(x,t)$, $c = c(x,t)$, $d = d(x,t)$ jsou dané funkce, přičemž $b > 0$. Explicitní schéma je přirozené uvažovat ve tvaru

$$\frac{U_j^{n+1} - U_j^n}{\tau} = b_j^n\,\frac{U_{j+1}^n - 2\,U_j^n + U_{j-1}^n}{h^2} - a_j^n\,\frac{U_{j+1}^n - U_{j-1}^n}{2\,h} + c_j^n\,U_j^n + d_j^n\,. \tag{4.41}$$

Označíme-li

$$\mu_j^n = \frac{\tau}{h^2}\,b_j^n\,, \qquad \nu_j^n = \frac{\tau}{h}\,a_j^n\,,$$

zjistíme, že chyba aproximace splňuje

$$e_j^{n+1} = (1 - 2\,\mu_j^n + \tau\,c_j^n)\,e_j^n + (\mu_j^n - \tfrac{1}{2}\,\nu_j^n)\,e_{j+1}^n + (\mu_j^n + \tfrac{1}{2}\,\nu_j^n)\,e_{j-1}^n - \tau\,\varepsilon_j^n\,.$$

Abychom při odhadu chyby mohli postupovat jako dříve, musíme zajistit, že koeficienty jsou nezáporné a jejich součet není větší než 1. To vyžaduje

$$\tfrac{1}{2}\,|\nu_j^n| \le \mu_j^n\,, \qquad 2\,\mu_j^n - \tau\,c_j^n \le 1\,, \qquad c_j^n \le 0\,. \tag{4.42}$$

Speciálně (dle první podmínky) musí být

$$h\,\frac{|a_j^n|}{2\,b_j^n} \le 1$$

a toto omezení implikuje omezení τ prostřednictvím druhé podmínky:

$$\tau \le \frac{h^2}{2\,b_j^n - h^2\,c_j^n}\,.$$

V mnoha úlohách z praxe je $|a_j^n| \gg b_j^n$, což vyžaduje velmi malé prostorové a časové kroky.

Jednoduchý způsob, jak tento problém napravit, je použít aproximace

$$u_x(x_j,t_n) \approx \begin{cases} \dfrac{U_j^n - U_{j-1}^n}{h} & \text{je-li } a(x_j,t_n) \ge 0, \\ \dfrac{U_{j+1}^n - U_j^n}{h} & \text{je-li } a(x_j,t_n) < 0. \end{cases}$$

Funkci a můžeme interpretovat jako rychlost látky, v níž sledujeme rozložení veličiny u, ve směru kladné x–ové poloosy. K diskretizaci $u_x(x_j,t_n)$ tedy využíváme hodnoty u z té strany, odkud se do bodu x_j v čase t_n látka pohybujeme. Hovoříme proto o diskretizaci typu *upwind*.

Předpokládejme pro jednoduchost, že $a(x,t) \ge 0$ a $c(x,t) = 0$. Explicitní schéma má pak tvar

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

což dává

$$e_j^{n+1} = (1 - 2\,\mu_j^n - \nu_j^n)\,e_j^n + \mu_j^n\,e_{j+1}^n + (\mu_j^n + \nu_j^n)\,e_{j-1}^n - \tau\,\varepsilon_j^n\,.$$

Aby všechny koeficienty na pravé straně byly nezáporné, potřebujeme nyní pouze podmínku $2\,\mu_j^n + \nu_j^n \le 1$. Stabilita tedy nevyžaduje žádné omezení prostorového kroku h. Cenou za to je, že nyní máme pouze $\varepsilon_j^n = O(h+\tau)$.

Někdy se můžeme setkat s parabolickou rovnicí v samoadjungovaném tvaru

$$u_t = (p(x,t)\,u_x)_x \qquad \forall\ t > 0, \quad x \in (0,1)\,,$$

kde $p > 0$. Rozderivováním můžeme tuto rovnici převést do tvaru (4.40), ale obvykle je výhodnější zkonstruovat diferenční aproximaci původního samoadjungovaného tvaru:

$$\begin{aligned} [(p\,u_x)_x](x_j,t_n) &\approx \frac{1}{h}\,[(p\,u_x)(x_{j+1/2},t_n) - (p\,u_x)(x_{j-1/2},t_n)] \\ &\approx \frac{1}{h^2}\,[p_{j+1/2}^n\,(u_{j+1}^n - u_j^n) - p_{j-1/2}^n\,(u_j^n - u_{j-1}^n)]\,. \end{aligned}$$

Explicitní diferenční schéma má tedy tvar

$$\frac{U_j^{n+1} - U_j^n}{\tau} = \frac{p_{j+1/2}^n\,(U_{j+1}^n - U_j^n) - p_{j-1/2}^n\,(U_j^n - U_{j-1}^n)}{h^2}\,,$$

a tudíž

$$U_j^{n+1} = [1 - \mu\,(p_{j+1/2}^n + p_{j-1/2}^n)]\,U_j^n + \mu\,p_{j+1/2}^n\,U_{j+1}^n + \mu\,p_{j-1/2}^n\,U_{j-1}^n\,, \qquad \mu = \frac{\tau}{h^2}\,.$$

Chybu aproximace můžeme tudíž vyšetřovat stejným způsobem jako dříve, budou-li všechny koeficienty nezáporné, což je splněno, pokud $\mu\,P \le \frac{1}{2}$, kde P splňuje $p(x,t) \le P$ v uvažované oblasti. Máme tedy omezení stejného typu jako dříve.

Zřejmým způsobem lze na výše uvažované obecnější rovnice zobecnit θ–schéma vedoucí k implicitnímu schématu.

5 Parabolické rovnice ve dvou prostorových dimenzích

Nechť $\Omega \subset \mathbb{R}^2$ je omezená oblast. Hledáme funkci $u = u(x, y, t)$ definovanou pro $(x, y) \in \overline{\Omega}$ a $t \geq 0$ takovou, že

$$u_t = \Delta u \equiv u_{xx} + u_{yy} \qquad \forall\, t > 0, \quad (x, y) \in \Omega\,, \tag{5.1}$$

$$u(x, y, t) = u^b(x, y, t) \qquad \forall\, t > 0, \quad (x, y) \in \partial\Omega\,, \tag{5.2}$$

$$u(x, y, 0) = u^0(x, y) \qquad \forall\, (x, y) \in \overline{\Omega}\,, \tag{5.3}$$

kde u^b a u^0 jsou zadané funkce.

Předpokládejme nejdříve, že $\Omega = (0, X) \times (0, Y)$, zvolme $J_x, J_y \in \mathbb{N}$ a položme $h_x = X/J_x$, $h_y = Y/J_y$. Oblast Ω pokryjeme rovnoměrnou pravoúhlou sítí s krokem dělení h_x v x–ovém směru a h_y v y–ovém směru. Uzly prostorové sítě jsou $(x_r, y_s) = (r\,h_x, s\,h_y)$, kde $r = 0, 1, \ldots, J_x$ a $s = 0, 1, \ldots, J_y$. Přibližné řešení značíme

$$U^n_{r,s} \approx u(x_r, y_s, t_n)\,, \qquad r = 0, 1, \ldots, J_x, \quad s = 0, 1, \ldots, J_y, \quad n = 0, 1, 2, \ldots\,.$$

Nejjednodušší explicitní diferenční schéma je

$$\frac{U^{n+1}_{r,s} - U^n_{r,s}}{\tau} = \frac{\delta^2_x\, U^n_{r,s}}{h^2_x} + \frac{\delta^2_y\, U^n_{r,s}}{h^2_y}\,, \qquad r = 1, \ldots, J_x - 1, \quad s = 1, \ldots, J_y - 1\,.$$

Hodnota $U^{n+1}_{r,s}$ je určena hodnotami $U^n_{r,s}$, $U^n_{r+1,s}$, $U^n_{r-1,s}$, $U^n_{r,s+1}$, $U^n_{r,s-1}$; hovoříme o tzv. pětibodovém schématu.

Vlastnosti schématu lze analyzovat analogicky jako v jedné dimenzi. Chyba diskretizace je

$$\varepsilon(x, t) = \tfrac{1}{2}\, u_{tt}\, \tau - \tfrac{1}{12}\, [u_{xxxx}\, h^2_x + u_{yyyy}\, h^2_y] + \ldots\,.$$

Za předpokladu omezenosti uvedených derivací a při

$$\mu_x + \mu_y \leq \tfrac{1}{2}\,, \qquad \text{kde} \quad \mu_x = \frac{\tau}{h^2_x}, \;\; \mu_y = \frac{\tau}{h^2_y}\,,$$

dostaneme stejně jako v jedné dimenzi pro chybu aproximace

$$\|e^n\|_\infty \leq T\, (\tfrac{1}{2}\, M_1\, \tau + \tfrac{1}{12}\, M^x_2\, h^2_x + \tfrac{1}{12}\, M^y_2\, h^2_y)\,.$$

Lze rovněž aplikovat Fourierovu analýzu stability. Pro $U^n_{r,s} = \lambda^n\, \mathrm{e}^{\mathrm{i}[k_x\, r\, h_x + k_y\, s\, h_y]}$ dostaneme amplifikační faktor

$$\lambda = \lambda(\boldsymbol{k}) = 1 - 4 \left[\mu_x\, \sin^2 \frac{k_x\, h_x}{2} + \mu_y\, \sin^2 \frac{k_y\, h_y}{2}\right],$$

kde $\boldsymbol{k} = (k_x, k_y)$. Vidíme, že $|\lambda(\boldsymbol{k})| \leq 1\ \forall\ \boldsymbol{k}$ právě tehdy, když $\mu_x + \mu_y \leq \frac{1}{2}$.

Zřejmým způsobem můžeme též rozšířit do dvou dimenzí θ–metodu. Speciálně metoda Crankova–Nicolsonové bude

$$(1 - \tfrac{1}{2}\, \mu_x\, \delta^2_x - \tfrac{1}{2}\, \mu_y\, \delta^2_y)\, U^{n+1}_{r,s} = (1 + \tfrac{1}{2}\, \mu_x\, \delta^2_x + \tfrac{1}{2}\, \mu_y\, \delta^2_y)\, U^n_{r,s}\,.$$

Soustava rovnic již není tridiagonální a její vyřešení je podstatně dražší než výpočet hodnot $U^{n+1}_{r,s}$ u explicitní metody. Hledáme proto jiné možnosti, jak získat nepodmíněně stabilní metodu.

5.1 Metoda střídavých směrů

Uvažujme následující modifikaci schématu Crankova–Nicolsonové:

$$(1-\tfrac{1}{2}\,\mu_x\,\delta_x^2)(1-\tfrac{1}{2}\,\mu_y\,\delta_y^2)\,U_{r,s}^{n+1}=(1+\tfrac{1}{2}\,\mu_x\,\delta_x^2)(1+\tfrac{1}{2}\,\mu_y\,\delta_y^2)\,U_{r,s}^{n}\,. \tag{5.4}$$

Dodatečné členy ve schématu nezhorší chybu diskretizace, neboť

$$\begin{aligned}\frac{1}{4}\,\mu_x\,\mu_y\,\delta_x^2\,\delta_y^2\,\frac{u(x,y,t+\frac{1}{2}\,\tau)-u(x,y,t-\frac{1}{2}\,\tau)}{\tau} &\approx \frac{\tau^2}{4\,h_x^2\,h_y^2}\,\delta_x^2\,\delta_y^2\,u_t(x,y,t)\\ &\approx \frac{1}{4}\,u_{xxyyt}(x,y,t)\,\tau^2\,,\end{aligned}$$

a tudíž $\varepsilon_{h_x,h_y,\tau}=O(h_x^2+h_y^2+\tau^2)$. Přednost schématu tkví v tom, že výpočet $U_{r,s}^{n+1}$ lze rozložit na řešení problémů s tridiagonálními maticemi. Zavedeme mezivýsledek $U_{r,s}^{n+1/2}$ a (5.4) zapíšeme v ekvivalentním tvaru

$$(1-\tfrac{1}{2}\,\mu_x\,\delta_x^2)\,U_{r,s}^{n+1/2}=(1+\tfrac{1}{2}\,\mu_y\,\delta_y^2)\,U_{r,s}^{n}\,,\qquad r=1,\dots,J_x-1,\ \ s=1,\dots,J_y-1\,, \tag{5.5}$$

$$(1-\tfrac{1}{2}\,\mu_y\,\delta_y^2)\,U_{r,s}^{n+1}=(1+\tfrac{1}{2}\,\mu_x\,\delta_x^2)\,U_{r,s}^{n+1/2}\,,\qquad r=1,\dots,J_x-1,\ \ s=1,\dots,J_y-1\,. \tag{5.6}$$

Výpočet $U_{r,s}^{n+1/2}$ z (5.5) představuje vyřešení J_y-1 soustav lineárních rovnic řádu J_x-1. Podobně (5.6) sestává z J_x-1 soustav lineárních rovnic řádu J_y-1. Každá ze soustav má tridiagonální matici a provedení jednoho časového kroku je tak mnohem rychlejší než řešení soustavy lineárních rovnic odpovídající Crankově–Nicolsonové metodě. Výpočetní náročnost je asi trojnásobná ve srovnání s jedním krokem explicitního schématu. Dosazením Fourierova členu $\lambda^n\,\mathrm{e}^{\mathrm{i}[k_x\,r\,h_x+k_y\,s\,h_y]}$ do (5.4) získáme

$$\lambda(\boldsymbol{k})=\frac{\left(1-2\,\mu_x\,\sin^2\dfrac{k_x\,h_x}{2}\right)\left(1-2\,\mu_y\,\sin^2\dfrac{k_y\,h_y}{2}\right)}{\left(1+2\,\mu_x\,\sin^2\dfrac{k_x\,h_x}{2}\right)\left(1+2\,\mu_y\,\sin^2\dfrac{k_y\,h_y}{2}\right)}$$

z čehož plyne, že schéma (5.4) je nepodmíněně stabilní.

6 Numerické řešení eliptických parciálních diferenciálních rovnic 2. řádu

6.1 Diskretizace Poissonovy rovnice

Uvažujme modelovou úlohu

$$-\Delta\,u=f\quad\text{v }\Omega:=(0,1)^2\,,\qquad u=0\quad\text{na }\partial\Omega\,. \tag{6.1}$$

Množinu $\overline{\Omega}$ pokryjeme rovnoměrnou čtvercovou sítí s J intervaly v každém směru, čímž vzniknou uzly $(x_r,y_s):=(r\,h,s\,h)$, $r,s=0,\dots,J$, kde $h=1/J$. Označíme

$$\begin{aligned}N_\Omega &=\{(x_r,y_s)\,;\ r,s\in\{1,\dots,J-1\}\}\,,\\ N_{\partial\Omega} &=\{(x_r,0),(x_r,1),(0,y_s),(1,y_s)\,;\ r,s\in\{0,\dots,J\}\}\,.\end{aligned}$$

Pak $N_\Omega \subset \Omega$ a $N_{\partial\Omega} \subset \partial\Omega$, tj. N_Ω je množina vnitřních uzlů a $N_{\partial\Omega}$ je množina hraničních uzlů. Řešení u úlohy (6.1) budeme v uzlech $(x_r, y_s) \in N_\Omega \cup N_{\partial\Omega}$ aproximovat hodnotami $U_{r,s}$. Dále zavedeme označení $u_{r,s} := u(x_r, y_s)$ a $f_{r,s} := f(x_r, y_s)$.

K diskretizaci úlohy (6.1) použijeme v každém vnitřním uzlu $(x_r, y_s) \in N_\Omega$ aproximaci

$$\begin{aligned}(\Delta\, u)(x_r, y_s) = u_{xx}(x_r, y_s) + u_{yy}(x_r, y_s) &\approx \frac{\delta_x^2\, u_{r,s}}{h^2} + \frac{\delta_y^2\, u_{r,s}}{h^2} \\ &= \frac{u_{r+1,s} - 2\,u_{r,s} + u_{r-1,s}}{h^2} + \frac{u_{r,s+1} - 2\,u_{r,s} + u_{r,s-1}}{h^2}\,,\end{aligned}$$

což nás vede k následující definici přibližného řešení $\{U_{r,s}\}_{r,s=0}^{J}$:

$$\frac{4\,U_{r,s} - U_{r+1,s} - U_{r-1,s} - U_{r,s+1} - U_{r,s-1}}{h^2} = f_{r,s}\,, \qquad r, s = 1, \ldots, J-1\,, \tag{6.2}$$

$$U_{r,0} = U_{r,J} = U_{0,s} = U_{J,s} = 0\,, \qquad r, s = 0, \ldots, J\,. \tag{6.3}$$

K určení přibližného řešení je tedy třeba vyřešit soustavu $(J-1)^2$ lineárních algebraických rovnic.

6.2 Konvergence přibližných řešení

Nechť $U = \{U_{r,s}\}_{r,s=0}^{J}$ je libovolná síťová funkce a definujme operátor L_h vztahem

$$(L_h\, U)_{r,s} := \frac{4\,U_{r,s} - U_{r+1,s} - U_{r-1,s} - U_{r,s+1} - U_{r,s-1}}{h^2}\,, \qquad r, s = 1, \ldots, J-1\,.$$

Místo $(L_h\, U)_{r,s}$ budeme užívat jednodušší značení $L_h\, U_{r,s}$. Pak úlohu (6.2)–(6.3) můžeme ekvivalentně zapsat ve tvaru

$$L_h\, U_P = f_P \qquad \forall\ P \in N_\Omega\,, \qquad\qquad U_P = 0 \qquad \forall\ P \in N_{\partial\Omega}\,, \tag{6.4}$$

kde pro uzly sítě nyní používáme značení P místo (x_r, y_s). Později ukážeme (viz důsledek 6.1 na str. 64), že operátor L_h splňuje diskrétní princip maxima

$$L_h\, U_P \le 0 \quad \forall\ P \in N_\Omega \qquad \Rightarrow \qquad \max_{P \in N_\Omega} U_P \le \max_{Q \in N_{\partial\Omega}} U_Q\,. \tag{6.5}$$

Podobně jako u evolučních úloh můžeme princip maxima využít k odhadu chyby aproximace. Nejprve definujeme chybu diskretizace

$$\varepsilon_{r,s} := L_h\, u_{r,s} - f_{r,s}\,, \qquad r, s = 1, \ldots, J-1\,,$$

což můžeme ekvivalentně zapsat ve tvaru

$$\varepsilon_P := L_h\, u_P - f_P \qquad \forall\ P \in N_\Omega\,. \tag{6.6}$$

Použitím Taylorova rozvoje získáme pro libovolné $P \in N_\Omega$

$$|\varepsilon_P| \le \frac{1}{12}\,(M_1 + M_2)\,h^2\,, \qquad \text{kde} \qquad M_1 = \max_{\overline{\Omega}} |u_{xxxx}|\,, \quad M_2 = \max_{\overline{\Omega}} |u_{yyyy}|\,. \tag{6.7}$$

Pomocí (6.6), (6.4) a (6.1) zjišťujeme, že chyba aproximace $e_P := U_P - u_P$ splňuje

$$L_h\, e_P = -\varepsilon_P \qquad \forall\ P \in N_\Omega\,, \qquad\qquad e_P = 0 \qquad \forall\ P \in N_{\partial\Omega}\,. \tag{6.8}$$

Vztah (6.8) nám umožňuje pomocí diskrétního principu maxima (6.5) a odhadu chyby diskretizace (6.7) odvodit odhad chyby aproximace. Za tím účelem je potřeba definovat vhodnou funkci splňující levou stranu implikace (6.5), k čemuž podobně jako u odvození odhadu chyby aproximace pro diskretizaci obecné Cauchyovy úlohy v části 2.2 využijeme tzv. *srovnávací funkci* Φ. V tomto případě ji můžeme definovat vztahem

$$\Phi(x,y) := (x - \tfrac{1}{2})^2 + (y - \tfrac{1}{2})^2\,.$$

Funkci Φ přiřadíme síťovou funkci $\{\Phi_P\}_{P\in N_\Omega\cup N_{\partial\Omega}}$, kde $\Phi_P = \Phi(P)$. Jelikož funkce Φ má nulové čtvrté derivace, plyne ze (6.7), že

$$L_h\, \Phi_P = (-\Delta\Phi)(P) = -4 \qquad \forall\ P \in N_\Omega\,.$$

Označme

$$\Psi_P := e_P + \frac{1}{4}\,\frac{h^2}{12}\,(M_1 + M_2)\,\Phi_P\,.$$

Pak

$$L_h\,\Psi_P = L_h\, e_P - \frac{h^2}{12}\,(M_1 + M_2) = -\varepsilon_P - \frac{h^2}{12}\,(M_1 + M_2) \le 0 \qquad \forall\ P \in N_\Omega$$

a podle (6.5) je

$$\Psi_P \le \frac{1}{4}\,\frac{h^2}{12}\,(M_1 + M_2)\,\max_{Q\in N_{\partial\Omega}}\,\Phi_Q = \frac{1}{8}\,\frac{h^2}{12}\,(M_1 + M_2) \qquad \forall\ P \in N_\Omega\,.$$

Jelikož $e_P \le \Psi_P$, dostáváme

$$U_P - u_P \le \frac{1}{96}\,(M_1 + M_2)\,h^2\,.$$

Označíme-li $\Psi_P := -e_P + \frac{1}{4}\,\frac{h^2}{12}\,(M_1 + M_2)\,\Phi_P$, získáme stejný odhad pro $-(U_P - u_P)$. Je tedy

$$|U_{r,s} - u(x_r, y_s)| \le \frac{1}{96}\,(M_1 + M_2)\,h^2 \qquad \forall\ r,s \in \{0,\dots,J\}\,.$$

6.3 Obecnější rovnice difúze

Uvažujme úlohu

$$-\mathrm{div}(a\,\nabla u) = f \quad \text{v } \Omega\,, \qquad \alpha_0\, u + \alpha_1\,\frac{\partial u}{\partial n} = g \quad \text{na } \partial\Omega\,, \tag{6.9}$$

kde Ω je omezená oblast v $\mathbb{R}^2$, n je vnější jednotková normála k $\partial\Omega$, a je hladká funkce splňující $a(x,y) \ge a_0 > 0$ a α_0, α_1 jsou konstanty splňující

$$\alpha_0 \ge 0\,, \qquad \alpha_1 \ge 0\,, \qquad \alpha_0 + \alpha_1 > 0\,.$$

Rovnice (6.9) popisuje difúzi veličiny u v nehomogenním izotropním prostředí. Pokud je $a(x,y) = \varepsilon$, jedná se o difúzi v homogenním izotropním prostředí a rovnice (6.9) se redukuje na rovnici (6.1).

Oblast Ω pokryjeme pravidelnou obdélníkovou sítí s krokem h_x ve směru osy x a s krokem h_y ve směru osy y. Uzly sítě, které leží vedle sebe na některé ze síťových přímek nazýváme sousední. Každý uzel má tedy čtyři sousední uzly. Uzly ležící v Ω, jejichž všichni čtyři sousedé leží v $\overline{\Omega}$ nazýváme *regulární uzly*. Uzly ležící v Ω, jejichž některý soused neleží v $\overline{\Omega}$ nazýváme *neregulární uzly*. Pokud Ω je mnohoúhelník, jehož hrany jsou rovnoběžné se souřadnými osami, můžeme zkonstruovat síť tak, aby neobsahovala neregulární uzly. U oblastí s komplikovanější hranicí to však obecně nelze.

Pro diskretizaci úlohy (6.9) se nabízí levou stranu parciální diferenciální rovnice rozderivovat a derivace funkce u aproximovat diferenčními kvocienty uvažovanými dříve. Označíme-li $b = -a_x$, $c = -a_y$, můžeme rovnici (6.9) zapsat ve tvaru

$$-a\,\Delta\,u + b\,u_x + c\,u_y = f \quad \text{v } \Omega \tag{6.10}$$

a v regulárních uzlech (x_r, y_s) aproximovat použitím centrálních diferencí, čímž získáme

$$-a_{r,s}\left(\frac{\delta_x^2\,U_{r,s}}{h_x^2} + \frac{\delta_y^2\,U_{r,s}}{h_y^2}\right) + b_{r,s}\,\frac{\Delta_{0x}\,U_{r,s}}{h_x} + c_{r,s}\,\frac{\Delta_{0y}\,U_{r,s}}{h_y} = f_{r,s}\,. \tag{6.11}$$

Snadno lze ukázat, že chyba diskretizace je druhého řádu v h_x a h_y. Později uvidíme, že k platnosti diskrétního principu maxima je potřeba, aby koeficienty schématu (6.11) v uzlech sousedících s uzlem (x_r, y_s) byly nekladné. Avšak koeficient u $U_{r-1,s}$ je roven $-(2\,a_{r,s} + b_{r,s}\,h_x)/(2\,h_x^2)$ a koeficient u $U_{r+1,s}$ je $-(2\,a_{r,s} - b_{r,s}\,h_x)/(2\,h_x^2)$. Musí tedy být $|b_{r,s}|\,h_x \le 2\,a_{r,s}$ a podobně $|c_{r,s}|\,h_y \le 2\,a_{r,s}$. Z toho tedy plyne, že tam, kde a je malé, ale rychle se mění (a má tudíž velké první derivace), musíme použít dostatečně jemnou síť.

Je proto výhodnější diskretizovat přímo rovnici (6.9), která má podle definice divergence tvar

$$-(a\,u_x)_x - (a\,u_y)_y = f \quad \text{v } \Omega\,.$$

Ukažme si podrobně aproximaci výrazu $(a\,u_x)_x$ v regulárním uzlu (x_r, y_s). Označíme-li $x_{r\pm\frac12} := x_r \pm \frac12\,h_x$, pak

$$\begin{aligned}
(a\,u_x)_x(x_r, y_s) &\approx \frac{(a\,u_x)(x_{r+\frac12}, y_s) - (a\,u_x)(x_{r-\frac12}, y_s)}{h_x} \\
&\approx \frac{1}{h_x}\left(a_{r+\frac12,s}\,\frac{u_{r+1,s} - u_{r,s}}{h_x} - a_{r-\frac12,s}\,\frac{u_{r,s} - u_{r-1,s}}{h_x}\right) \\
&\approx \frac{1}{h_x^2}\left(a_{r+\frac12,s}\,(U_{r+1,s} - U_{r,s}) - a_{r-\frac12,s}\,(U_{r,s} - U_{r-1,s})\right).
\end{aligned}$$

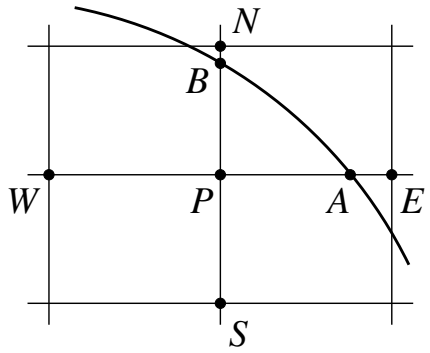
Analogicky postupujeme u výrazu $(a\,u_y)_y$, čímž získáme pro aproximaci rovnice (6.9) schéma

$$\begin{aligned}
&-\frac{1}{h_x^2}\left(a_{r+\frac12,s}\,(U_{r+1,s} - U_{r,s}) - a_{r-\frac12,s}\,(U_{r,s} - U_{r-1,s})\right) \\
&-\frac{1}{h_y^2}\left(a_{r,s+\frac12}\,(U_{r,s+1} - U_{r,s}) - a_{r,s-\frac12}\,(U_{r,s} - U_{r,s-1})\right) = f_{r,s}\,.
\end{aligned} \tag{6.12}$$

Toto schéma lze zapsat v kompaktním tvaru

$$-\left(\frac{\delta_x(a\,\delta_x U)}{h_x^2}+\frac{\delta_y(a\,\delta_y U)}{h_y^2}\right)_{r,s}=f_{r,s}\,.$$

Je vidět, že koeficienty schématu (6.12) jsou v uzlech sousedících s uzlem (x_r, y_s) nekladné bez jakéhokoli omezení na síť.



Obrázek 4: Neregulární uzel v blízkosti zakřivené hranice.

6.4 Diskretizace úlohy (6.9) v neregulárních uzlech

V této části si ukážeme, jakým způsobem lze diferenční schémata (6.11) a (6.12) modifikovat v neregulárních uzlech.

Uvažujme nejprve situaci znázorněnou v obr. 4. Uzel P leží v oblasti Ω v blízkosti zakřivené hranice, přičemž jeho sousední uzly S a W leží v $\overline{\Omega}$, zatímco zbývající dva sousední uzly E a N v $\overline{\Omega}$ neleží. Průsečík hranice s úsečkou PE je označen A a průsečík hranice s úsečkou PN je označen B. Předpokládejme, že na uvažované části hranice je předepsána Dirichletova okrajová podmínka (tj. v (6.9) je $\alpha_0 > 0$ a $\alpha_1 = 0$). V bodech A a B tudíž známe hodnoty řešení u a tedy i přibližného řešení U.

Nechť $|PA| = \theta\, h_x$, $\theta \in (0,1)$. Pak hodnoty funkce u v bodech A a W můžeme vyjádřit pomocí následujících Taylorových rozvojů:

$$u_A=\sum_{k\ge 0}\frac{1}{k!}\frac{\partial^k u}{\partial x^k}(P)\,(\theta\,h_x)^k\,,\qquad u_W=\sum_{k\ge 0}\frac{1}{k!}\frac{\partial^k u}{\partial x^k}(P)\,(-h_x)^k\,.$$

Tedy

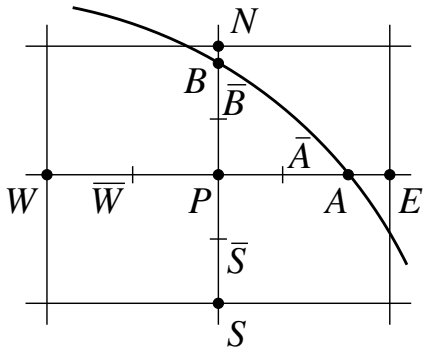
$$\begin{aligned}u_A+\theta\,u_W&=(1+\theta)\,u_P+\frac{1}{2}\,\theta\,(1+\theta)\,u_{xx}(P)\,h_x^2+O(\theta\,h_x^3)\,,\\ u_A-\theta^2\,u_W&=(1-\theta^2)\,u_P+\theta\,(1+\theta)\,u_x(P)\,h_x+O(\theta^2\,h_x^3)\,,\end{aligned}$$

z čehož plyne

$$u_x(P)=\frac{u_A-\theta^2\,u_W-(1-\theta^2)\,u_P}{\theta\,(1+\theta)\,h_x}+O(h_x^2)\,, \tag{6.13}$$

$$u_{xx}(P)=\frac{u_A+\theta\,u_W-(1+\theta)\,u_P}{\frac{1}{2}\,\theta\,(1+\theta)\,h_x^2}+O(h_x)\,. \tag{6.14}$$

Podobně můžeme postupovat pro body S, P, B a odvodit aproximace derivací $u_y(P)$ a $u_{yy}(P)$. Dosazením do (6.10) získáme modifikaci schématu (6.11) v nerelulárním uzlu P. Pro platnost principu maxima musíme samozřejmě obecně opět požadovat omezení kroku sítě, jako tomu bylo v regulárních uzlech.



Obrázek 5: Další značení v okolí neregulárního uzlu.

Aproximujeme-li přímo rovnici (6.9), můžeme i v neregulárním uzlu postupovat analogicky jako při odvození schématu (6.12). Označíme $\overline{A}$, $\overline{B}$, $\overline{W}$ a $\overline{S}$ středy úseček PA, PB, PW a PS, viz obr. 5, a uvažujeme aproximace

$$\begin{aligned}(a\,u_x)_x(P) &\approx \frac{(a\,u_x)(\overline{A}) - (a\,u_x)(\overline{W})}{|\overline{A}-\overline{W}|}\\ &\approx \frac{1}{\frac{1}{2}\,|AW|}\left(a_{\overline{A}}\,\frac{u_A-u_P}{|AP|} - a_{\overline{W}}\,\frac{u_P-u_W}{|PW|}\right)\\ &\approx \frac{1}{\frac{1}{2}\,(1+\theta)\,h_x}\left(a_{\overline{A}}\,\frac{U_A-U_P}{\theta\,h_x} - a_{\overline{W}}\,\frac{U_P-U_W}{h_x}\right).\end{aligned}$$

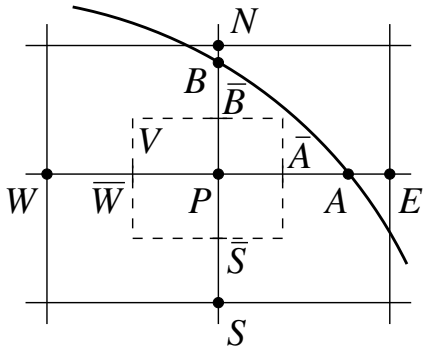
Analogicky lze aproximovat $(a\,u_y)_y$. Označíme-li $\phi := |PB|/h_y$, dostáváme v bodě P síťovou rovnici

$$\begin{aligned}&-\frac{1}{\frac{1}{2}\,\theta\,(1+\theta)\,h_x^2}\Big(a_{\overline{A}}\,(U_A-U_P) - \theta\,a_{\overline{W}}\,(U_P-U_W)\Big)\\ &-\frac{1}{\frac{1}{2}\,\phi\,(1+\phi)\,h_y^2}\Big(a_{\overline{B}}\,(U_B-U_P) - \phi\,a_{\overline{S}}\,(U_P-U_S)\Big) = f_P\,.\end{aligned} \tag{6.15}$$

Pro $\theta = \phi = 1$ získáváme rovnici (6.12). Je-li $a = const.$, je síťová rovnice shodná se síťovou rovnicí získanou pomocí (6.14). Výhodou (6.15) však je, že v obecném případě koeficienty splňují podmínky požadované pro princip maxima, aniž by bylo nutno omezit velikost kroku sítě.

Rovnici (6.15) lze též odvodit pomocí integrálního tvaru rovnice (6.9). Nechť V je obdélník (tzv. *kontrolní objem*), jehož hrany jsou rovnoběžné se síťovými přímkami a procházejí body $\overline{A}$, $\overline{B}$, $\overline{W}$ a $\overline{S}$, viz obr. 6. Integrací rovnice (6.9) přes kontrolní objem V a aplikací Gaussovy věty o integraci získáme

$$\int_V f\,\mathrm{d}x = -\int_V \mathrm{div}(a\,\nabla u)\,\mathrm{d}x = -\int_{\partial V} a\,\frac{\partial u}{\partial n}\,\mathrm{d}\sigma\,, \tag{6.16}$$



Obrázek 6: Konstrukce obdélníku V.

kde n je vnější jednotková normála k ∂V. Pro levou stranu identity (6.16) použijeme aproximaci

$$\int_V f \,\mathrm{d}x \approx |V|\, f_P = \frac{1}{4}\,(1+\theta)(1+\phi)\, h_x\, h_y\, f_P\,.$$

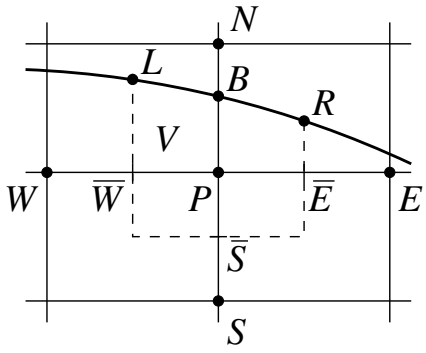
Integrál na pravé straně identity (6.16) rozdělíme na integrály přes jednotlivé hrany obdélníku V a každý z nich aproximujeme zvlášť. Např. integrál přes hranu V procházející bodem $\overline{S}$ můžeme aproximovat výrazem

$$\frac{1}{2}\,(1+\theta)\, h_x\, a_{\overline{S}}\, \frac{\partial u}{\partial n}(\overline{S}) \approx \frac{1}{2}\,(1+\theta)\, h_x\, a_{\overline{S}}\, \frac{U_S - U_P}{h_y}\,.$$

Aproximujeme-li obdobně integrály přes ostatní hrany V, získáme síťovou rovnici (6.15).

Použití integrálního tvaru (6.16) pro odvození síťové rovnice v neregulárním uzlu je zejména výhodné, pokud je na části hranice v blízkosti tohoto uzlu předepsána okrajová podmínka obsahující derivaci, neboť pak je odvození vhodné diskretizace podstatně komplikovanější než v dirichletovském případě. Pro jednoduchost předpokládejme, že hranice oblasti Ω v blízkosti neregulárního uzlu P protíná pouze svislé síťové přímky (viz obr. 7) a že je na této části hranice předepsána Neumannova okrajová podmínka

$$\frac{\partial u}{\partial n} = g$$



Obrázek 7: Konstrukce kontrolního objemu V v blízkosti neumannovské hranice.

(tj. v (6.9) je $\alpha_0 = 0$ a $\alpha_1 = 1$). Kontrolní objem V nyní definujeme tak, že část jeho hranice je tvořena částí LBR hranice oblasti Ω a zbývající část jeho hranice sestává z úseček kolmých k síťovým přímkám a protínajících úsečky PE, PW a PS v jejich středech $\overline{E}$, $\overline{W}$ a $\overline{S}$. Označíme $|\overline{W}L| = \phi_1\, h_y$, $|\overline{E}R| = \phi_2\, h_y$ a $|LR| = \psi\, h_x$. V integrálním tvaru (6.16) aproximujeme levou stranu vztahem

$$\int_V f \,\mathrm{d}x \approx \frac{1}{2}\,(1 + \phi_1 + \phi_2)\, h_x\, h_y\, f_P\,,$$

integrály přes rovné části ∂V aproximujeme analogicky jako výše a pro integrál přes oblouk LBR použijeme vztah

$$\int_{LBR} a\,\frac{\partial u}{\partial n}\,\mathrm{d}\sigma \approx a_B\, g_B\, \psi\, h_x\,.$$

To vede k diferenční rovnici

$$\begin{aligned}(6.17)\quad &-\frac{2\,\phi_2 + 1}{(1 + \phi_1 + \phi_2)\, h_x^2}\, a_{\overline{E}}\,(U_E - U_P) - \frac{2\,\phi_1 + 1}{(1 + \phi_1 + \phi_2)\, h_x^2}\, a_{\overline{W}}\,(U_W - U_P)\\ &\qquad - \frac{2}{(1 + \phi_1 + \phi_2)\, h_y^2}\, a_{\overline{S}}\,(U_S - U_P) = f_P + \frac{2\,\psi\, a_B\, g_B}{(1 + \phi_1 + \phi_2)\, h_y}\,.\end{aligned}$$

Pokud hranice oblasti Ω protíná síťové přímky v blízkosti uzlu P tak jako na obr. 4, je základní myšlenka odvození síťové rovnice v uzlu P stejná jako výše, avšak konstrukce kontrolního objemu V a aproximace integrálu přes jeho hranici je komplikovanější.

6.5 Princip maxima pro diskretizace eliptických úloh

Předpokládejme, že jsme diskretizací úlohy (6.1) či jiné lineární eliptické úlohy získali v každém bodě $P \in N_\Omega$ síťovou rovnici

$$(6.18)\qquad L_h\, U_P = f_P + q_P\,,$$

kde q_P je určeno okrajovými podmínkami jinými než dirichletovskými. Příkladem síťové rovnice s nenulovým členem q_P je rovnice (6.17). Množina N_Ω obsahuje všechny uzly ležící v oblasti Ω, ale může obsahovat i některé body na hranici Ω, v nichž je předepsána jiná okrajová podmínka než dirichletovská. Dále zavádíme množinu $N_{\partial\Omega}$ sestávající z bodů na hranici Ω, v nichž je předepsána Dirichletova okrajová podmínka. Body množiny $N_{\partial\Omega}$ mohou náležet mezi uzly sítě pokrývající oblast Ω nebo mohou být průsečíky síťových přímek s hranicí $\partial\Omega$ jako např. body A a B v obr. 4. Přibližné řešení $\{U_P\}$ v rovnici (6.18) představuje množinu hodnot v bodech množiny $N_\Omega \cup N_{\partial\Omega}$.

Činíme následující předpoklady o operátoru L_h a množinách N_Ω a $N_{\partial\Omega}$:

(P1) Pro každé $P \in N_\Omega$ definujeme množinu $N_P \subset (N_\Omega \cup N_{\partial\Omega}) \setminus \{P\}$ a předpokládáme, že

$$(6.19)\qquad L_h\, U_P = c_P\, U_P - \sum_{Q \in N_P} c_{PQ}\, U_Q\,,$$

kde

$$c_{PQ} > 0 \quad \forall\, Q \in N_P\,, \qquad c_P \geq \sum_{Q \in N_P} c_{PQ}\,. \tag{6.20}$$

(P2) Množina N_Ω je souvislá v tom smyslu, že

$$\forall\, P, Q \in N_\Omega \quad \exists\, P_1, \ldots, P_m \in N_\Omega\, : \\ P_1 \in N_P,\, P_2 \in N_{P_1}, \ldots, P_m \in N_{P_{m-1}},\, Q \in N_{P_m}\,.$$

(P3) $\exists\, P \in N_\Omega,\, Q \in N_{\partial\Omega}\, : \quad Q \in N_P\,.$

Lze snadno ověřit, že rovnice (6.2), (6.12), (6.15) a (6.17) splňují (6.19) a (6.20) (místo druhé nerovnosti v (6.20) platí ve všech čtyřech případech rovnost). Pro diskretizaci (6.2) Poissonovy rovnice je splněn též přepoklad (P2) o souvislosti množiny N_Ω. Předpoklad (P3) implikuje, že na části hranice oblasti Ω je předepsána Dirichletova okrajová podmínka. Pro diskretizace (6.2) i (6.15) je přepoklad (P3) zřejmě splněn.

Nyní zformulujeme a dokážeme diskrétní princip maxima pro výše uvedený operátor L_h.

Věta 6.1 *Nechť jsou splněny předpoklady (P1)–(P3). Pak pro libovolnou síťovou funkci* $\{U_P\}_{P \in N_\Omega \cup N_{\partial\Omega}}$ *platí*

$$L_h U_P \leq 0 \quad \forall\, P \in N_\Omega \qquad \Rightarrow \qquad \max_{P \in N_\Omega} U_P \leq \max_{Q \in N_{\partial\Omega}} U_Q^+\,, \tag{6.21}$$

$$L_h U_P \geq 0 \quad \forall\, P \in N_\Omega \qquad \Rightarrow \qquad \min_{P \in N_\Omega} U_P \geq \min_{Q \in N_{\partial\Omega}} U_Q^-\,, \tag{6.22}$$

kde $U_Q^+ = \max\{U_Q, 0\}$, $U_Q^- = \min\{U_Q, 0\}$. *Speciálně*

$$L_h U_P = 0 \quad \forall\, P \in N_\Omega \qquad \Rightarrow \qquad \max_{P \in N_\Omega} |U_P| \leq \max_{Q \in N_{\partial\Omega}} |U_Q|\,. \tag{6.23}$$

Důkaz. Nechť platí levá strana implikace (6.21) a označme $M_\Omega := \max_{P \in N_\Omega} U_P$ a $M_{\partial\Omega} := \max_{Q \in N_{\partial\Omega}} U_Q$. Je-li $M_\Omega \leq 0$, pak (6.21) triviálně platí. Nechť tedy $M_\Omega > 0$ a předpokládejme, že $M_\Omega > M_{\partial\Omega}$. Nechť $P^* \in N_\Omega$ je takový, že $M_\Omega = U_{P^*}$. Pak

$$M_\Omega = U_{P^*} \leq \frac{1}{c_{P^*}} \sum_{Q \in N_{P^*}} c_{P^*Q}\, U_Q \leq M_\Omega\,, \tag{6.24}$$

kde první nerovnost plyne z toho, že $L_h U_{P^*} \leq 0$, a druhá z (6.20). Vztah (6.24) může platit pouze tehdy, pokud v něm platí rovnosti, a z (6.20) pak plyne, že $U_Q = M_\Omega$ pro všechna $Q \in N_{P^*}$. Ze souvislosti množiny N_Ω formulované v předpokladu (P2) pak plyne, že $U_P = M_\Omega$ pro všechna $P \in N_\Omega$. Konečně z předpokladu (P3) plyne, že existuje $Q \in N_{\partial\Omega}$ takové, že $U_Q = M_\Omega$, což je spor s předpokladem, že $M_\Omega > M_{\partial\Omega}$. Je tedy $M_\Omega \leq M_{\partial\Omega}$, tj. platí (6.21).

Platí-li levá strana implikace (6.22), pak $\{-U_P\}_{P \in N_\Omega \cup N_{\partial\Omega}}$ splňuje levou stranu implikace (6.21), a tudíž

$$-\min_{P\in N_\Omega} U_P = \max_{P\in N_\Omega}(-U_P) \le \max_{Q\in N_{\partial\Omega}} (-U_Q)^+ = \max_{Q\in N_{\partial\Omega}} (-U_Q^-) = -\min_{Q\in N_{\partial\Omega}} U_Q^-\,,$$

tj. platí (6.22).

Platí-li levá strana implikace (6.23), pak $\{U_P\}_{P\in N_\Omega\cup N_{\partial\Omega}}$ i $\{-U_P\}_{P\in N_\Omega\cup N_{\partial\Omega}}$ splňují levou stranu implikace (6.21), a tudíž pro libovolné $P \in N_\Omega$ platí

$$U_P \le \max_{Q\in N_{\partial\Omega}} U_Q^+\,, \quad -U_P \le \max_{Q\in N_{\partial\Omega}} (-U_Q)^+ \qquad \Rightarrow \qquad |U_P| \le \max_{Q\in N_{\partial\Omega}} |U_Q|\,,$$

což dává (6.23). □

Důsledek 6.1 *Nechť jsou splněny předpoklady (P1)–(P3) a navíc*

$$c_P = \sum_{Q\in N_P} c_{PQ} \qquad \forall\ P \in N_\Omega\,. \tag{6.25}$$

Pak pro libovolnou síťovou funkci $\{U_P\}_{P\in N_\Omega\cup N_{\partial\Omega}}$ platí

$$L_h U_P \le 0 \quad \forall\ P \in N_\Omega \qquad \Rightarrow \qquad \max_{P\in N_\Omega} U_P \le \max_{Q\in N_{\partial\Omega}} U_Q\,, \tag{6.26}$$

$$L_h U_P \ge 0 \quad \forall\ P \in N_\Omega \qquad \Rightarrow \qquad \min_{P\in N_\Omega} U_P \ge \min_{Q\in N_{\partial\Omega}} U_Q\,. \tag{6.27}$$

Důkaz. Nechť platí levá strana implikace (6.26) a označme $M := \min_{Q\in N_\Omega\cup N_{\partial\Omega}} U_Q$. Buď $\tilde U_P := U_P - M$ pro všechna $P \in N_\Omega \cup N_{\partial\Omega}$. Pak $\tilde U_P \ge 0$ pro všechna $P \in N_\Omega \cup N_{\partial\Omega}$ a díky (6.25) je $L_h \tilde U_P = L_h U_P \le 0$ pro všechna $P \in N_\Omega$. Použitím (6.21) získáme

$$\Big(\max_{P\in N_\Omega} U_P\Big) - M = \max_{P\in N_\Omega} \tilde U_P \le \max_{Q\in N_{\partial\Omega}} \tilde U_Q^+ = \max_{Q\in N_{\partial\Omega}} \tilde U_Q = \Big(\max_{Q\in N_{\partial\Omega}} U_Q\Big) - M\,,$$

z čehož plyne (6.26). Implikace (6.27) se získá přechodem k $\{-U_P\}_{P\in N_\Omega\cup N_{\partial\Omega}}$ a aplikací implikace (6.26), analogicky jako bylo získáno (6.22) z (6.21) v důkazu věty 6.1. □

Stejně jako ve spojitém případě plyne z diskrétního principu maxima formulovaného ve větě 6.1 jednoznačnost řešení úlohy

$$L_h U_P = f_P + q_P \qquad \forall\ P \in N_\Omega\,, \tag{6.28}$$

$$U_P = g_P \qquad \forall\ P \in N_{\partial\Omega}\,, \tag{6.29}$$

kde $g_P := g(P)$ udává Dirichletovu okrajovou podmínku v bodě P. Jelikož úloha (6.28)–(6.29) je lineární a konečně rozměrná, plyne z jednoznačnosti též existence řešení, jak ukážeme v následující větě.

Věta 6.2 *Nechť jsou splněny předpoklady (P1)–(P3). Pak pro libovolnou pravou stranu má úloha* (6.28)–(6.29) *právě jedno řešení.*

Důkaz. Úloze (6.28)–(6.29) odpovídá soustava lineárních algebraických rovnic se čtvercovou maticí pro neznámé hodnoty $\{U_P\}_{P\in N_\Omega}$. Úloha je tedy jednoznačně řešitelná právě tehdy, když homogenní soustava má pouze triviální řešení. Je-li však pravá strana v (6.28)–(6.29) nulová, dostáváme z (6.23) ihned, že $U_P = 0$ pro všechna $P \in N_\Omega$. □

6.6 Odhad chyby aproximace

V této části odvodíme obecné odhady chyby aproximace pro úlohu (6.28)–(6.29), přičemž zobecníme odhad chyby aproximace provedený na začátku této kapitoly pro diskretizaci Poissonovy rovnice.

Nechť u je řešení spojité úlohy, kterou aproximujeme pomocí diskretizace (6.28)–(6.29). Jako obvykle zavádíme chybu diskretizace

$$\varepsilon_P := L_h\, u_P - f_P - q_P \qquad \forall\ P \in N_\Omega\,,$$

kde $u_P = u(P)$. Předpokládáme, že přibližné řešení splňuje stejné Dirichletovy okrajové podmínky jako řešení přesné, tj. $U_P = u(P)$ pro všechna $P \in N_{\partial\Omega}$. Pak chyba aproximace $e_P := U_P - u_P$ opět splňuje

$$L_h\, e_P = -\varepsilon_P \qquad \forall\ P \in N_\Omega\,, \qquad\qquad e_P = 0 \qquad \forall\ P \in N_{\partial\Omega}\,.$$

Věta 6.3 *Nechť jsou splněny předpoklady (P1)–(P3) a nechť Φ je nezáporná síťová funkce definovaná na $N_\Omega \cup N_{\partial\Omega}$ splňující*

$$L_h\, \Phi_P \le -1 \qquad \forall\ P \in N_\Omega\,.$$

Pak pro chybu aproximace odpovídající diskretizaci (6.28)–(6.29) *platí odhad*

$$\max_{P\in N_\Omega} |e_P| \le \Big(\max_{Q\in N_{\partial\Omega}} \Phi_Q\Big)\Big(\max_{P\in N_\Omega} |\varepsilon_P|\Big).$$

Důkaz. Buď $D := \max_{P\in N_\Omega} |\varepsilon_P|$. Pak

$$L_h(D\,\Phi_P \pm e_P) = D\,L_h\,\Phi_P \mp \varepsilon_P \le -D \mp \varepsilon_P \le 0 \qquad \forall\ P \in N_\Omega\,,$$

a tudíž podle (6.21) platí pro libovolné $P \in N_\Omega$

$$\pm e_P \le D\,\Phi_P \pm e_P \le \max_{Q\in N_{\partial\Omega}} (D\,\Phi_Q \pm e_Q)^+ = D \max_{Q\in N_{\partial\Omega}} \Phi_Q\,,$$

z čehož plyne dokazovaný odhad. □

Zatímco odhad chyby diskretizace je poměrně snadný, nalezení srovnávací funkce Φ nemusí být vždy jednoduché. Tato funkce není samozřejmě určena jednoznačně a cílem je nalézt takové Φ, aby $\max_{Q\in N_{\partial\Omega}} \Phi_Q$ bylo co nejmenší.

Pro úlohu (6.1) a diskretizaci (6.2)–(6.3) plyne z věty 6.3 odhad $|U_P - u_P| \le C\,h^2$ pro libovolné $P \in N_\Omega$. Bude-li však mít Ω zakřivenou hranici, získáme odhad $|\varepsilon_P| \le C\,h^2$ pouze v regulárních uzlech, zatímco v uzlech u hranice budeme mít obecně jen $|\varepsilon_P| \le C\,h$ (viz vztah (6.14)). Aplikací věty 6.3 pak dostaneme pouze $|U_P - u_P| \le C\,h$ pro $P \in N_\Omega$. Tento odhad však není optimální a lze ho zlepšit použitím následující věty.

Věta 6.4 *Nechť $N_\Omega = N_1 \cup N_2$, kde $N_1 \cap N_2 = \emptyset$. Nechť jsou splněny předpoklady (P1)–(P3) a nechť Φ je nezáporná síťová funkce definovaná na $N_\Omega \cup N_{\partial\Omega}$ splňující*

$$L_h\,\Phi_P \le -C_1 < 0 \quad \forall\ P \in N_1\,, \qquad L_h\,\Phi_P \le -C_2 < 0 \quad \forall\ P \in N_2\,.$$

Nechť chyba diskretizace odpovídající úloze (6.28)–(6.29) *splňuje*

$$|\varepsilon_P| \le D_1 \quad \forall\ P \in N_1\,, \qquad |\varepsilon_P| \le D_2 \quad \forall\ P \in N_2\,.$$

Pak pro chybu aproximace platí odhad

$$\max_{P\in N_\Omega} |e_P| \le \Big(\max_{Q\in N_{\partial\Omega}} \Phi_Q\Big) \max\left\{\frac{D_1}{C_1}, \frac{D_2}{C_2}\right\}.$$

Důkaz. Označíme-li

$$D := \max\left\{\frac{D_1}{C_1}, \frac{D_2}{C_2}\right\},$$

můžeme důkaz provést analogicky jako pro větu 6.3. □

Příklad 6.1 *Uvažujme úlohu* (6.1) *s* $\Omega := \{(x,y) \in \mathbb{R}^2;\ x^2+y^2 < 1\}$. *Oblast* Ω *pokryjeme rovnoměrnou čtvercovou sítí s krokem* h *a v regulárních uzlech použijeme schéma* (6.2)*, zatímco v nereguláních uzlech použijeme pro druhé derivace aproximace typu* (6.14)*. Aplikací věty 6.4 odvoďte odhad chyby aproximace, který je druhého řádu v* h.

Řešení: Nechť N_1 je množina uzlů z N_Ω, jejichž všechny sousední uzly leží v Ω a $N_2 := N_\Omega \setminus N_1$. Podle (6.7) a (6.14) platí

$$|\varepsilon_P| \le K_1\, h^2 \quad \forall\ P \in N_1\,, \qquad |\varepsilon_P| \le K_2\, h \quad \forall\ P \in N_2\,.$$

Nechť $\Psi(x,y) = x^2+y^2$ a položme

$$\Phi_P = M_1\, \Psi(P) \quad \forall\ P \in N_\Omega\,, \qquad \Phi_P = M_1\, \Psi(P) + M_2 \quad \forall\ P \in N_{\partial\Omega}\,,$$

kde M_1, M_2 jsou kladné konstanty, které budou určeny později. Pak

$$L_h\, \Phi_P = M_1\, L_h\, \Psi_P = M_1\, (-\Delta\Psi)_P = -4\, M_1 \qquad \forall\ P \in N_1\,.$$

Je-li $P \in N_2$, uvažujme např. situaci z obr. 4 a aproximaci $u_{xx}(P)$ pomocí výrazu v (6.14). Jelikož chyba této aproximace závisí na třetích derivacích u, je tato aproximace pro kvadratické funkce přesná, a proto získáme

$$\frac{\Phi_A + \theta\,\Phi_W - (1+\theta)\,\Phi_P}{\frac{1}{2}\,\theta\,(1+\theta)\,h^2} = M_1\, \Psi_{xx}(P) + \frac{M_2}{\frac{1}{2}\,\theta\,(1+\theta)\,h^2} \ge 2\, M_1 + \frac{M_2}{h^2}\,.$$

Obdobně postupujeme v ostatních situacích, které mohou nastat, což vede ke vztahu

$$L_h\, \Phi_P \le -4\, M_1 - \frac{M_2}{h^2} \le -\frac{M_2}{h^2} \qquad \forall\ P \in N_2\,.$$

Aplikací věty 6.4 dostáváme

$$\max_{P\in N_\Omega} |e_P| \le (M_1+M_2) \max\left\{\frac{K_1\, h^2}{4\, M_1}, \frac{K_2\, h^3}{M_2}\right\}.$$

Položíme-li $M_1 = \frac{1}{4} K_1\, h^2$ a $M_2 = K_2\, h^3$, získáme

$$\max_{P\in N_\Omega} |e_P| \le \frac{1}{4}\, K_1\, h^2 + K_2\, h^3\,,$$

tj. ukázali jsme, že chyba aproximace je skutečně druhého řádu přesnosti.