



Universidade do Minho

Escola de Ciências

Departamento de Matemática

Métodos Numéricos para Equações de Derivadas Parciais

M. Irene Falcão

Mestrado em Matemática e Computação

2023/2024



Equações Parabólicas

A descrição matemática de problemas de difusão ou de condução de calor unidimensionais, conduz a uma equação parabólica cuja solução é uma função u na variável espacial x e no tempo t . A equação parabólica

$$\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial x^2}, \quad \sigma > 0, \quad (1)$$

modela a condução de calor num filamento, termicamente isolado, sendo a constante σ o coeficiente de difusão térmica (dependente do material de que o filamento é feito).

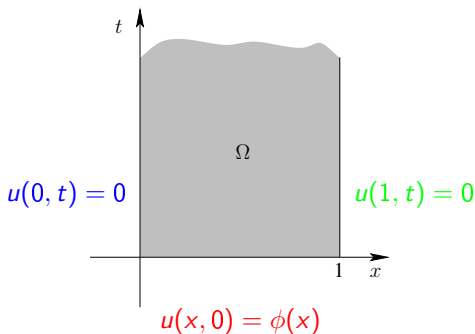
Consideremos, por exemplo, as condições iniciais e de fronteira

$$u(x, 0) = \phi(x), \quad 0 < x < 1, \quad (2)$$

$$u(0, t) = u(1, t) = 0, \quad t \geq 0. \quad (3)$$

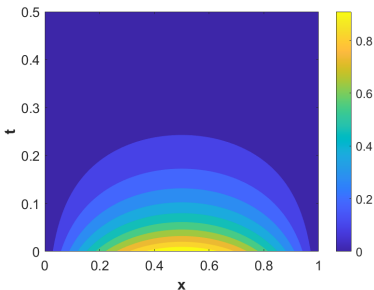
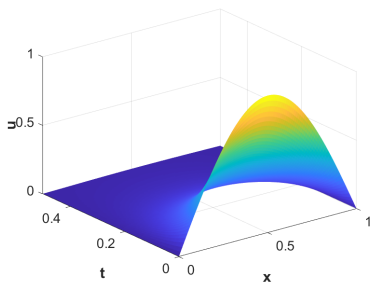


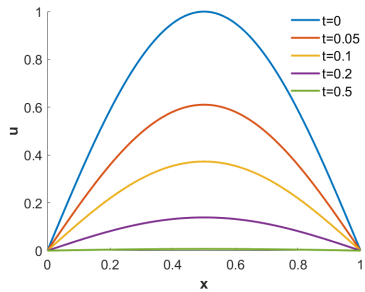
A solução u deste problema representa a temperatura u num ponto situado a uma distância x de uma das extremidades do filamento, ao fim de t unidades de tempo de condução de calor; neste caso, o filamento tem comprimento unitário e a temperatura nas suas extremidades é fixa (neste exemplo nula) em qualquer instante t ; no instante inicial $t = 0$, é conhecida a distribuição da temperatura $\phi(x)$ ao longo do filamento.



O domínio Ω onde se pretende encontrar a solução é a semifaixa

$\Omega = (0, 1) \times (0, \infty)$; em problemas típicos de difusão, a solução pode sofrer mudanças rápidas logo no início, mas a evolução de u vai-se tornando cada vez mais lenta, como se pode observar nas figuras seguintes, onde se representa graficamente a solução do problema, para $\sigma = 1$ e $\phi(x) = \sin(\pi x)$.





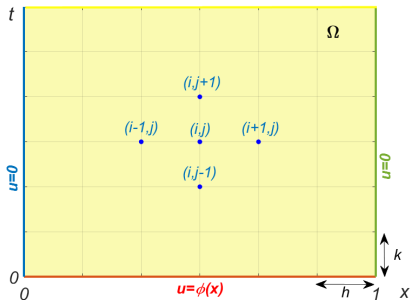
Objetivo

Estratégia

Objetivo: aproximar a solução do problema (1)-(3), na região

$$\Omega = (0, 1) \times (0, +\infty).$$

Estratégia: cobrir, a região $\bar{\Omega}$ por uma rede de malha retangular de dimensões h e k , nas direções de x e t , respetivamente e usar métodos de diferenças finitas para aproximar a solução u nos *nós* dessa rede.



Segue-se imediatamente das fórmulas de diferenças e da equação (1) que, em cada ponto $P = (ih, jk)$; $i = 1, \dots, M - 1$, $j = 0, 1, 2, \dots$ onde $M = \frac{1}{h}$, se tem

$$\frac{u_{i,j+1} - u_{i,j}}{k} = \sigma \left(\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} \right) + \tau_{i,j}, \quad (4)$$

onde

$$\tau_{i,j} = \frac{k}{2} \frac{\partial^2}{\partial t^2} u(x_i, t_j + \nu_j k) - \sigma \frac{h^2}{12} \frac{\partial^4}{\partial x^4} u(x_i + \xi_i, t_j) = \mathcal{O}(k + h^2).$$

Aproximações $U_{i,j}$ para $u_{i,j}$ podem obter-se ignorando os termos de ordem $\mathcal{O}(k + h^2)$ em (4), isto é, de

$$\frac{U_{i,j+1} - U_{i,j}}{k} = \sigma \left(\frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{h^2} \right). \quad (5)$$

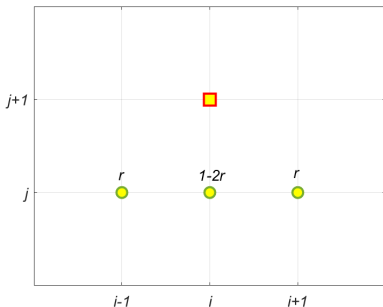


Fazendo

$$r := \sigma \frac{k}{h^2}, \quad (6)$$

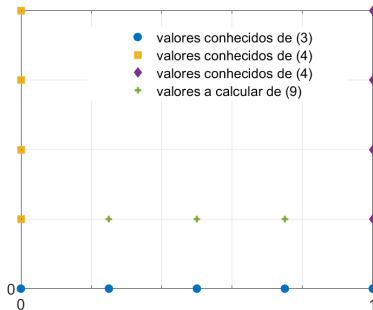
a equação (5) pode ser reescrita como

$$U_{i,j+1} = rU_{i+1,j} + (1 - 2r)U_{i,j} + rU_{i-1,j} \quad (7)$$
$$i = 1, \dots, M - 1; \quad j = 0, 1, 2, \dots$$



A condição inicial (2) permite obter os valores $U_{i,0} = \phi(ih)$ e das condições de fronteira (3) resulta $U_{0,j} = U_{M,j} = 0$.

Os valores $U_{i,1}$ podem então ser calculados através de (7), usando os valores $U_{i,0}$ conhecidos; em seguida, os valores $U_{i,1}$ podem ser usados para calcular os valores $U_{i,2}$ e assim sucessivamente.



A fórmula (7) exprime um valor desconhecido explicitamente em termos de valores conhecidos. Por essa razão, o método definido por (7) diz-se um **método explícito**.

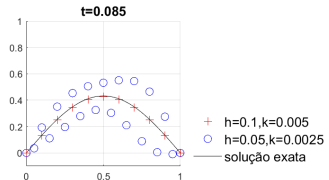
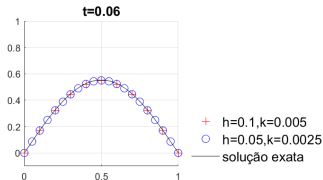
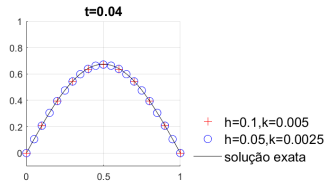
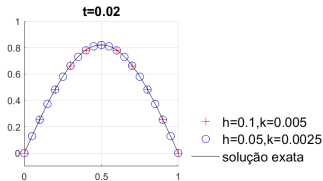
O valor de $r = \sigma \frac{k}{h^2}$ é bastante importante; de facto, como veremos mais adiante, só podemos garantir a **validade deste método se $0 < r \leq \frac{1}{2}$** , ou seja, se **$k \leq \frac{1}{2\sigma} h^2$** . Para se obter uma precisão razoável, o valor de h deve ser pequeno o que, por sua vez, exige um valor de k bastante pequeno. Isto implica uma grande quantidade de cálculos para se “progredir” com a solução ao longo de t .



Exemplo 1

Consideremos novamente o problema (1)-(3), com $\sigma = 1$ e $\phi(x) = \sin(\pi x)$.

Os resultados obtidos usando o método explícito estão representados nas figuras seguintes.



Para $h = 0.1$ e $k = 0.005$ a solução calculada concorda bastante bem com a solução exata $e^{-\pi^2 t} \sin(\pi x)$. Contudo, quando os passos h e k são divididos por dois, produz-se uma solução de diferenças que cresce rapidamente e oscila, sendo claramente instável. A tabela seguinte mostra o valor das soluções nos pontos $(x, 0.085)$, com $x=0:0.1:0.5$ (a solução é simétrica).

x	solução 1	solução 2	solução exata
0	0	0	0
0.1	0.13167	0.19302	0.13355
0.2	0.25045	0.35092	0.25403
0.3	0.34472	0.45355	0.34964
0.4	0.40524	0.50586	0.41103
0.5	0.42610	0.53228	0.43218



Crank e Nicolson propuseram um método que reduz o volume total de cálculos e que é válido para qualquer valor de r . Neste método, a derivada $(\frac{\partial^2 u}{\partial x^2})_{i,j}$ é substituída pela média das suas aproximações de diferenças centrais nas linhas $(j+1)$ e j , isto é, a equação

$$\frac{\partial u}{\partial y} = \sigma \frac{\partial^2 u}{\partial x^2}$$

é aproximada, no ponto $P = (ih, jk)$ por

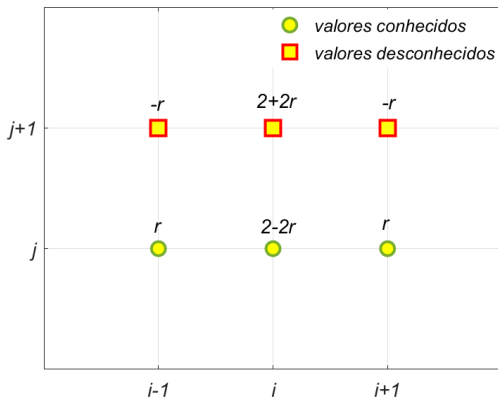
$$\frac{U_{i,j+1} - U_{i,j}}{k} = \frac{\sigma}{2} \left(\frac{U_{i+1,j+1} - 2U_{i,j+1} + U_{i-1,j+1}}{h^2} + \frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{h^2} \right);$$
$$i = 1, 2, \dots, M-1, j > 0.$$



Definindo novamente $r := \sigma \frac{k}{h^2}$, a equação anterior pode ser reescrita como

$$-rU_{i-1,j+1} + (2+2r)U_{i,j+1} - rU_{i+1,j+1} = rU_{i-1,j} + (2-2r)U_{i,j} + rU_{i+1,j}; \quad (8)$$

$$i = 1, 2, \dots, M-1, \quad j > 0.$$



Trata-se de um **método implícito**, uma vez que envolve mais do que um ponto da linha $(j + 1)$ de cada vez; assim, contrariamente ao método explícito, não é possível determinar cada valor $U_{i,j+1}$ isoladamente.

Em vez disso, é formado o sistema de $M - 1$ equações dadas por (8), o qual, uma vez que $U_{0,j+1} + U_{M,j+1} = 0$, pode ser escrito na seguinte forma matricial

$$A_1 U_{j+1} = A_2 U_j, \quad (9)$$

onde

$$U_j = \begin{pmatrix} U_{1,j} \\ U_{2,j} \\ \vdots \\ U_{M-2,j} \\ U_{M-1,j} \end{pmatrix}$$



$$A_1 = \begin{pmatrix} 1+r & -r/2 & & & \\ -r/2 & 1+r & -r/2 & & \\ & \ddots & \ddots & \ddots & \\ & & -r/2 & 1+r & -r/2 \\ & & & -r/2 & 1+r \end{pmatrix}.$$

e

$$A_2 = \begin{pmatrix} 1-r & r/2 & & & \\ r/2 & 1-r & r/2 & & \\ & \ddots & \ddots & \ddots & \\ & & r/2 & 1-r & r/2 \\ & & & r/2 & 1-r \end{pmatrix}.$$



Para determinar a solução (de diferenças) no nível $(j + 1)$ da rede, isto é, para calcular $(U_{1,j+1}, \dots, U_{M-1,j+1})^T$, é necessário resolver o sistema (9). Trata-se de um sistema cuja matriz de coeficientes é tridiagonal. Como a matriz A_1 é sempre a mesma para cada nível da rede, é conveniente efetuar a decomposição¹ LU de A_1 e resolver os diversos sistemas por substituição direta e inversa.

¹Note-se que A_1 é uma matriz estritamente de diagonal dominante por linhas e colunas, pelo que não há necessidade de efetuar escolha de pivot na sua decomposição LU.



O método definido por (8) é um caso particular de uma classe de métodos definidos por

$$\frac{U_{i,j+1} - U_{i,j}}{k} = \frac{\sigma}{h^2} (\theta \delta^2 U_{i,j+1} + (1 - \theta) \delta^2 U_{i,j}) ;$$
$$i = 1, \dots, M - 1, j > 0, 0 \leq \theta \leq 1.$$

- ▶ $\theta = 1/2$ obtém-se o método de Crank-Nicolson;
- ▶ $\theta = 0$ corresponde ao método explícito;
- ▶ $\theta = 1$ resulta num método *totalmente implícito*.



Que condições devem ser satisfeitas para que a solução do método de diferenças seja uma aproximação razoável para a solução da equação diferencial correspondente?

Estas condições estão associadas a dois problemas diferentes, mas relacionados:

1. **convergência** da solução exata das equações de diferenças para a solução exata da equação diferencial;
2. **estabilidade** das equações de diferenças, ou seja, crescimento ilimitado ou não de quaisquer erros associados com a resolução destas equações.



A equação diferencial (1) pode ser reescrita na forma $Lu = 0$, com L o operador diferencial $L := \frac{\partial}{\partial t} - \sigma \frac{\partial^2}{\partial x^2}$. Em cada ponto $P = (ih, jk)$, tem-se

$$[Lu]_{i,j} = 0.$$

Nos MDF, a equação anterior é substituída por uma equação de diferenças

$$F(u_{i,j}) = \tau_{i,j}, \quad (10)$$

obtendo-se uma solução aproximada $U_{i,j}$ para $u_{i,j}$ ignorando os erros $\tau_{i,j}$, i.e., resolvendo

$$F(U_{i,j}) = 0. \quad (11)$$



Por exemplo, no caso do método explícito simples,

$$F(u_{i,j}) = \frac{u_{i,j+1} - u_{i,j}}{k} - \sigma \frac{u_{i+1,j} - 2u_{i,j} + u_{i,j-1}}{h^2}$$

e

$$\tau_{i,j} = \frac{k}{2} \frac{\partial^2}{\partial t^2} u(x_i, t_j + \nu_j k) - \sigma \frac{h^2}{12} \frac{\partial^4}{\partial x^4} u(x_i + \xi_i h, t_j).$$

O valor $\tau_{i,j} = F(u_{i,j})$ chama-se **erro de truncatura local** no ponto (ih, jk) .

Se $\tau_{i,j} \rightarrow 0$ quando $h, k \rightarrow 0$, o método de diferenças diz-se **consistente** com a equação diferencial.

Por outras palavras, a consistência de uma fórmula de diferenças com uma equação diferencial tem a ver com quanto a fórmula de diferenças se “aproxima” da equação diferencial quando a rede é tornada cada vez mais fina.



Erro de discretização local

Ao valor $e_{i,j} = u_{i,j} - U_{i,j}$, isto é, à diferença entre a solução *exata* da equação diferencial no ponto $(x, t) = (ih, jk)$ e a solução *exata* da equação de diferenças (11) nesse ponto, chamamos **erro de discretização local** no ponto (x, t) .

Se $e_{i,j} \rightarrow 0$ quando $h, k \rightarrow 0$ (e $i, j \rightarrow \infty$, de modo que $ih = x$, $jk = t$, permaneça fixo) dizemos que o método de diferenças **converge** no ponto (x, t) .

Segue-se imediatamente de (10) e (11) que $F(e_{i,j}) = \tau_{i,j}$.

Note-se que a **consistência de uma fórmula de diferenças não implica a sua convergência**. Na realidade, as equações de diferenças não são resolvidas exatamente, sendo a sua solução afetada de erros inevitáveis (por exemplo, fruto da representação no computador, dos seus coeficientes).



Erro de arredondamento global e erro total

Ao resolver as equações de diferenças, a introdução de um pequeno erro afeta ou não de forma significativa a solução obtida?

Relembre Exemplo 1!

Se $U_{i,j}$ representa a solução exata da equação de diferenças e $V_{i,j}$ a solução realmente calculada, a diferença

$$R_{i,j} = U_{i,j} - V_{i,j} \quad (12)$$

chamamos **erro de arredondamento global** no ponto $(x, t) = (ih, jk)$.

O **erro total** $E_{i,j}$ no ponto $(x, t) = (ih, jk)$ é dado por

$$\begin{aligned} E_{i,j} = u_{i,j} - V_{i,j} &= u_{i,j} - U_{i,j} + U_{i,j} - V_{i,j} \\ &= \text{Erro discretização} + \text{Erro arredon. global} . \quad (13) \end{aligned}$$



A forma mais usual de estudar a estabilidade é fazer a **análise da propagação dos erros iniciais**.

Um MDF é **estável** se uma perturbação limitada dos dados iniciais não “cresce” de forma ilimitada, quando j tende para infinito (mantendo-se os passos h e k constantes).

Existem dois processos usuais de investigar a estabilidade de um MDF:

1. **análise matricial**;
2. método de Neumann ou **método de Fourier**



O método explícito (7) para resolver o problema (1)-(3) pode escrever-se matricialmente como

$$\mathbf{U}_{j+1} = \mathbf{A}\mathbf{U}_j, \quad (14)$$

onde

$$\mathbf{A} = \begin{pmatrix} 1-2r & r & & & \\ r & 1-2r & r & & \\ & \ddots & \ddots & \ddots & \\ & & r & 1-2r & r \\ & & & r & 1-2r \end{pmatrix}.$$

e

$$\mathbf{U}_j = \begin{pmatrix} U_{1,j} & U_{2,j} & \dots & U_{M-2,j} & U_{M-1,j} \end{pmatrix}^T$$



De (14) resulta

$$\mathbf{U}_{j+1} = A\mathbf{U}_j = A^2\mathbf{U}_{j-1} = \cdots = A^{j+1}\mathbf{U}_0,$$

onde \mathbf{U}_0 é o vetor dos valores iniciais. Suponhamos que foi introduzido um erro inicial ε_0 em \mathbf{U}_0 .

► o erro ε_1 em \mathbf{U}_1 é $\varepsilon_1 = A(\mathbf{U}_0 + \varepsilon_0) - A\mathbf{U}_0 = A\varepsilon_0$

► o erro ε_2 em \mathbf{U}_2 é $\varepsilon_2 = A(\mathbf{U}_1 + \varepsilon_1) - A\mathbf{U}_1 = A\varepsilon_1 = A^2\varepsilon_0$

...

► o erro ε_{j+1} em \mathbf{U}_{j+1} é $\varepsilon_{j+1} = A^{j+1}\varepsilon_0$.

Logo, a fórmula para a propagação dos erros é igual à do cálculo de \mathbf{U}_j .



O método explícito será estável, se ε_j se mantiver limitado, quando $j \rightarrow \infty$.

Segue-se imediatamente que uma condição para que tal se verifique é que

$$\|A\| < 1. \quad (15)$$

Como $\rho(A) \leq \|A\|_2$ (relembre Exercício 1.5) resulta imediatamente que se (15) é satisfeita, então $\rho(A) < 1$. Neste caso particular, a matriz A é real e simétrica, pelo que $\|A\|_2 = \rho(A)$, i.e.

Uma condição para a estabilidade do método explícito é que os valores próprios da matriz A tenham módulo não superior a 1.



Além disso, a matriz A é uma matriz tridiagonal de ordem $M - 1$, sendo os seus valores próprios da forma

$$\lambda_j = 1 - 2r + 2r \cos \frac{j\pi}{M} = 1 - 4r \sin^2 \frac{j\pi}{2M}; \quad j = 1, \dots, M - 1.$$

(relembre Exercício 1.9).

Logo, a condição de estabilidade do método explícito é

$$|1 - 4r \sin^2 \frac{j\pi}{2M}| \leq 1 \Leftrightarrow r \sin^2 \frac{j\pi}{2M} \leq \frac{1}{2}.$$

O método explícito é estável se

$$r \leq \frac{1}{2}. \quad (16)$$



Consideremos o método explícito (7) para resolução da equação do calor

$$\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial x^2}, \quad \sigma > 0,$$

com condições iniciais

$$u(x, 0) = \phi(x), \quad 0 \leq x \leq L.$$

O método de Fourier para análise da propagação dos erros iniciais exprime estes erros em termos de uma série de Fourier finita e considera o crescimento de uma função que se reduz a esta série para $t = 0$, por um método de “separação de variáveis” análogo ao usualmente usado para obter a solução analítica da EDP.



Denotemos então os erros nos pontos $x_i = ih$ ao longo de $t = 0$ por $\varepsilon_{i,0}$; $i = 0, \dots, M$, os quais podem ser representados por uma série de Fourier finita da forma

$$\varepsilon_{i,0} = \sum_{m=0}^M A_m e^{i\beta_m x_i}, \quad (17)$$

onde i é a unidade imaginária, $\beta_m = \frac{m\pi}{Mh}$ e $Mh = L$.

Como o esquema (7) (e todos os outros que estudamos) é linear, basta estudar o efeito da propagação dos erros num termo da forma $e^{i\beta_m x_i}$; os coeficientes A_m são constantes e podem, por isso, ser omitidos na análise.



Para descrever o efeito da propagação deste erro à medida que t aumenta, é necessário encontrar a solução da equação de diferenças (relembre que a fórmula para a propagação dos erros é igual à do cálculo de U_j .)

$$\varepsilon_{i,j+1} = r\varepsilon_{i-1,j} + (1 - 2r)\varepsilon_{i,j} + r\varepsilon_{i+1,j} \quad (18)$$

que se reduz a $e^{i\beta x_i}$, quando $t = jk = 0$.

Assumamos que

$$\varepsilon_{i,j} = e^{i\beta x_i} e^{\alpha t_j} = e^{i\beta i h} e^{\alpha j k} = e^{i\beta i h} \zeta^j, \quad (19)$$

onde $\zeta = e^{\alpha k}$ e α é (em geral) uma constante complexa. Então, (19) reduz-se a $e^{i\beta i h}$, quando $j = 0$. Logo o erro não aumenta de forma descontrolada com t se

$$|\zeta| \leq 1 \quad (20)$$



Substituindo (19) em (18) vem

$$e^{i\beta ih}\zeta^{j+1} = re^{i\beta(i-1)h}\zeta^j + (1 - 2r)e^{i\beta ih}\zeta^j + re^{i\beta(i+1)h}\zeta^j.$$

Dividindo ambos os membros da relação anterior por $e^{i\beta ih}\zeta^j$, obtém-se

$$\begin{aligned}\zeta &= re^{-i\beta h} + (1 - 2r) + re^{i\beta h} \\ &= 2r(\cos \beta h - 1) + 1 \\ &= 1 - 4r \operatorname{sen}^2 \frac{\beta h}{2}.\end{aligned}$$

Logo

$$|\zeta| \leq 1 \Leftrightarrow |1 - 4r \operatorname{sen}^2 \frac{\beta h}{2}| \leq 1 \Leftrightarrow r \leq \frac{1}{2}.$$

(Relembre resultado página 27)



Na análise precedente da estabilidade, considerámos o passo k como constante, quando j tende para infinito, o que é de facto o que acontece na prática.

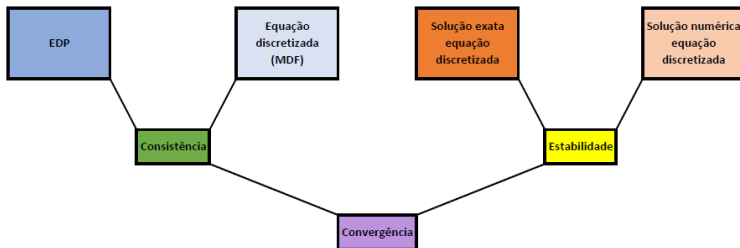
Para relacionar estabilidade com convergência torna-se necessário, no entanto, formular um conceito de estabilidade que seja independente dos passos h e k .

Suponhamos, então, que $t = jk$ é um número fixo. Um esquema de diferenças finitas é **estável** quando existir um crescimento limitado de qualquer forma de erro, quando j tende para infinito (e k tende para 0, de forma a que $t = jk$ permaneça constante).



Teorema da equivalência de Lax

Dado um problema de valores iniciais bem posto e um esquema de diferenças finitas que seja consistente com ele, a estabilidade é condição necessária e suficiente de convergência.



Teorema da equivalência de Lax

Método explícito simples

Para ilustrar (parcialmente) o Teorema da equivalência de Lax, vamos mostrar que, se $r \leq \frac{1}{2}$ (a qual é também a condição de estabilidade no novo sentido), então o método é convergente.

Relembremos que a solução da equação (1) satisfaz (cf. (4))

$$u_{i,j+1} = u_{i,j} + r(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}) + k\tau_{i,j},$$

onde $\tau_{i,j} = \mathcal{O}(k + h^2)$ e escrevamos as equações de diferenças (7) como

$$U_{i,j+1} = U_{i,j} + r(U_{i+1,j} - 2U_{i,j} + U_{i-1,j}).$$

Temos então que

$$e_{i,j+1} = u_{i,j+1} - U_{i,j+1} = e_{i,j} + r(e_{i+1,j} - 2e_{i,j} + e_{i-1,j}) + k\tau_{i,j}.$$



Se $E_j = \max_i |e_{i,j}|$ e $E = \sup_{i,j} |\tau_{i,j}|$, então

$$\begin{aligned} |e_{i,j+1}| &\leq r|e_{i+1,j}| + |1 - 2r||e_{i,j}| + r|e_{i-1,j}| + kE \\ &\leq rE_j + |1 - 2r|E_j + rE_j + kE = (2r + |1 - 2r|)E_j + kE. \end{aligned}$$

Mas $r \leq \frac{1}{2}$, logo

$$|e_{i,j+1}| \leq E_j + kE \Rightarrow E_{j+1} \leq E_j + kE.$$

Aplicando este resultado repetidamente, obtém-se

$$E_j \leq E_{j-1} + kE \leq E_{j-2} + 2kE \leq \dots \leq E_0 + jkE = E_0 + tE.$$

Como, para $t = 0$, $U = u$, segue-se que $E_0 = 0$. Além disso, como t é finito e o método é consistente, $tE \rightarrow 0$, quando $h, k \rightarrow 0$, ou seja, o método converge.

