

# PDE 数值解 笔记

原生生物

\* 张强《偏微分方程的有限差分方法》笔记

## 目录

一 全显/全隐格式	3
§1.1 基本记号	3
§1.2 离散格式	3
§1.3 古典格式推广	4
二 线性差分格式基本理论	5
§2.1 预备知识	5
§2.2 相容性	5
§2.3 稳定性	7
§2.4 收敛性	10
三 热传导方程	11
§3.1 相容性改进	11
§3.2 计算效率改进	14
§3.3 误差估计与收敛分析	16
§3.4 边界与初值条件离散	17
四 一维扩散方程	19
§4.1 光滑系数线性扩散	19
§4.2 间断系数线性扩散	21
§4.3 更复杂的扩散	22
五 高维扩散方程	23
§5.1 方程与边界条件的离散	24
§5.2 分数步长方法	26
六 线性常数对流方程	29
§6.1 基本格式与分析	29
§6.2 更多格式	32
七 线性双曲型方程	34
§7.1 更复杂的一阶方程	34
§7.2 高阶与高维推广	35

<b>八 非线性双曲守恒律</b>	<b>38</b>
§8.1 守恒型差分格式 . . . . .	39
§8.2 有限体积法 . . . . .	43
§8.3 稳定性与收敛性 . . . . .	45
<b>九 发展型偏微分方程</b>	<b>47</b>
§9.1 对流扩散方程 . . . . .	47
§9.2 修正方程与能量方法 . . . . .	49
<b>十 椭圆型方程</b>	<b>52</b>
§10.1 五点格式 . . . . .	52
§10.2 最大模估计 . . . . .	54
§10.3 提升数值精度 . . . . .	55

## 一 全显/全隐格式

### §1.1 基本记号

考虑 Dirichlet 边值问题 (HD), 其中扩散系数  $a > 0$ :

$$u_t = au_{xx} + f(x, t), x \in (0, 1), t \in (0, T]$$

$$u(x, 0) = u_0(x), u(0, t) = \phi_0(t), u(1, t) = \phi_1(t)$$

真解记为  $[u]$ , 数值解记为  $u$ , 考虑等距网格  $\mathcal{T}_{\Delta x, \Delta t}$ , 其中  $x_j$  代表  $j\Delta x, j = 0 : J$ ,  $t^n$  代表  $n\Delta t, n = 0 : N$ , 这里冒号连接整数表示其间的所有整数,  $\Delta x = \frac{1}{J}$  为空间步长,  $\Delta t = \frac{T}{N}$  为时间步长。

\* 非等距时将相邻最大的  $\Delta t, \Delta x$  称为时空步长。为方便讨论, 此后默认等距。

记  $[u](x_j, t^n) = [u]_j^n$ , 其即为逼近目标, 将对其的近似  $u_j^n$  称为数值解。

差分记号:

1.  $\mathbb{I}$  代表恒等算子;
2.  $\mathbb{E}$  代表正向移位算子,  $\mathbb{E}f_j = f_{j+1}$ , 其复合代表多次进行。
3.  $\mathbb{E}^{-1}$  为反向移位算子, 由此可定义  $\mathbb{E}$  的任意整数次方 [对实数次方, 由于  $f_\alpha^n = f(\alpha\Delta x, t^n)$ , 仍可类似定义];
4. 一阶向前差分算子  $\Delta = \mathbb{E} - \mathbb{I}$ ;
5. 一阶向后差分算子  $\Delta_- = \mathbb{I} - \mathbb{E}^{-1}$ ;
6. 一步中心差分算子  $\Delta_0 = \mathbb{E} - \mathbb{E}^{-1}$ ;
7. 半步中心差分算子  $\delta = \mathbb{E}^{1/2} - \mathbb{E}^{-1/2}$ ;
8. 二阶中心差分算子  $\delta^2 = \mathbb{E} - 2\mathbb{I} + \mathbb{E}^{-1}$ 。

\* 涉及操作变量时以下标标注, 如  $\mathbb{E}_x f_j^n = f_{j+1}^n, \mathbb{E}_t f_j^n = f_j^{n+1}$ 。

### §1.2 离散格式

利用 Taylor 展开可知

$$[u_t]_j^n = \frac{\Delta_t [u]_j^n}{\Delta t} + O(\Delta t)$$

$$[u_{xx}]_j^n = \frac{\delta_x^2 [u]_j^n}{(\Delta x)^2} + O((\Delta x)^2)$$

于是对 (HD) 忽略无穷小量  $O((\Delta x)^2 + \Delta t)$  可得到近似公式

$$\Delta_t u_j^n = \mu a \delta_x^2 u_j^n + \Delta t f_j^n$$

这里  $\mu = \frac{\Delta t}{(\Delta x)^2}$  称为网比。

另一方面, 若取近似  $[u_t]_j^n = \frac{\Delta_{-t} [u]_j^n}{\Delta t} + O(\Delta t)$  即得 (注意  $\Delta_{-t} [u]_j^n = \Delta_t [u]_j^{n-1}$ )

$$\Delta_t u_j^n = \mu a \delta_x^2 u_j^{n+1} + \Delta t f_j^{n+1}$$

将差分中出现的网格点集称为离散模板 [此后记  $(a, b)$  为  $u_{j+a}^{n+b}$  描述], 则前者涉及  $(0, 0), (-1, 0), (1, 0), (0, 1)$ , 只需要直接计算  $(0, 1)$  位置即可, 因此称为显式离散; 后者涉及  $(0, 0), (-1, 1), (1, 1), (0, 1)$ , 必须利用方程组求解, 因此称为隐式离散。

\* 事实上, 两者分别称为**全显格式**与**全隐格式**, 合称为**古典格式**。由于只涉及到两个时间层, 它们都是**双层格式**。

\* 此问题定解条件的离散非常简单, 即  $u_j^0 = u_0(x_j), u_0^n = \phi_0(t^n), u_j^n = \phi_1(t^n)$ 。

### 全隐格式可行性与求解

记  $u^n = (u_1^n, \dots, u_{J-1}^n)^T$ , 则全隐格式可以写为

$$\mathbb{B}_1 u^{n+1} = \mathbb{B}_0 u^n + \Delta t F^n$$

这里  $\mathbb{B}_1 = \text{tridiag}(-\mu a, 1 + 2\mu a, -\mu a)$  [ $\text{tridiag}(x, y, z)$  表示满足  $a_{i,i-1} = x, a_{ii} = y, a_{i,i+1} = z$ , 其他为 0 的三对角阵],  $\mathbb{B}_0 = \mathbb{I}$ , 而

$$F^n = f^{n+1} + \frac{\mu a}{\Delta t} (u_0^{n+1}, 0, \dots, 0, u_J^{n+1})^T$$

由于  $\mathbb{B}_1$  严格对角优, 其对应矩阵可逆, 因此必然可求解。

\* 对全显格式也可写成类似形式,  $\mathbb{B}_1 = \mathbb{I}, \mathbb{B}_0 = \text{tridiag}(-\mu a, 1 - 2\mu a, \mu a)$ , 而  $F_n$  为全隐格式的  $F_n$  上标的  $n+1$  全部更换为  $n$  的结果。

求解三对角线性方程组的常用方法是 **Thomas 算法**, 或称**追赶法**, 对三对角阵

$$A = \text{tridiag}(\alpha_{2:J-1}, \beta_{1:J-1}, \gamma_{1:J-2})$$

求解  $Ax = b$  的算法为:

1. 初始化  $p_0 = q_0 = 0$ ;

2. 计算

$$p_j = \frac{\gamma_j}{\beta_j - \alpha_j p_{j-1}}, q_j = \frac{b_j + \alpha_j q_{j-1}}{\beta_j - \alpha_j p_{j-1}}, j = 1 : J-1$$

3. 令  $x_{J-1} = q_{J-1}$ ;

4. 逆序计算  $x_j = p_j x_{j+1} + q_j, j = 1 : J-2$ 。

由此可见, 进行全隐格式的一步迭代需要约  $5J$  次乘除法, 而全显格式需要约  $2J$  次。然而, 由于全隐格式的时间步长可以设置得更大, 实际上全隐格式的计算效率更高, 关于此的分析将在之后讨论, 现在只陈述经验性的结果:

1. 在  $\mu \leq \frac{1}{2}$  时两种格式均有较理想的数值结果, 误差处于  $O((\Delta x)^2)$ ;

2. 在  $\mu > \frac{1}{2}$  时, 全显格式出现抖动, 且随时间越来越剧烈, 中心点趋于无穷, 全隐格式则仍然呈现  $O((\Delta x)^2)$  的误差。

### §1.3 古典格式推广

考虑整个数轴上的问题 (HI):

$$u_t = au_{xx} + f(x, t), x \in \mathbb{R}, t \in (0, T]$$

$$u(x, 0) = u_0(x)$$

这时定义的  $x_j$  下标  $j$  可取任何整数, 全显格式与全隐格式的迭代与 HD 相同, 而边界条件只有  $u_j^0 = u_0(x_j)$ 。

\* 对无穷区域问题, 理论表述可不考虑截断, 实际操作中则需要。为方便讨论, 我们假定适当的区域截断和人工边界条件设置后, 原有格式的数值结果不受影响。

考虑空间周期性问题 (HP):

$$u_t = au_{xx} + f(x, t), x \in \mathbb{R}, t \in (0, T]$$

$$u(x, 0) = u_0(x),$$

$$u(x, t) = u(x+1, t), u_0(x) = u_0(x+1), f(x, t) = f(x+1, t)$$

这时数值计算只需要截取一个完整的空间周期  $j = 1 : J$  即可，而边界满足  $u_0^n = u_J^n, u_{J+1}^n = u_1^n$ ，此时全隐格式的单步迭代矩阵  $\mathbb{B}_1$  即为  $J$  阶方阵

$$\text{tridiag}(-\mu a, 1 + 2\mu a, -\mu a) + \begin{pmatrix} & & & -\mu a \\ & & 0 & \\ & & \ddots & \\ 0 & & & \\ -\mu a & & & \end{pmatrix}$$

利用 **Sherman-Morrison** 公式，可改进 Thomas 算法使得能用  $11J$  次乘除法运算解出线性方程组  $[\mathbb{B}_1$  的四个角构成秩为 1 的方阵  $uv^T$ ，去除它们得三对角阵  $A$ ，只需求解  $A^{-1}b$  与  $A^{-1}u$  即可组合得  $\mathbb{B}_1^{-1}b$ 。

\* 本章对格式的探讨仅停留在求解可行性与求解效率上，而忽略了差分格式的可信度，即数值解是否能刻画真实解，这就是下一章的核心内容。

## 二 线性差分格式基本理论

### §2.1 预备知识

回顾之前讨论的时空网格  $\mathcal{T}_{\Delta x, \Delta t}$ ，我们仍然默认其等距性，初始点  $(x_0, t^0)$  成为参考网格点。

考虑一族空间网格上的函数  $u^n = \{u_n^j\}_{\forall j}$ ，我们用离散范数定义其整体度量：

$$\|u^n\|_\infty = \max_j |u_j^n|, \quad \|u^n\|_2 = \sqrt{\sum_j |u_j^n|^2 \Delta x}$$

\* 前者称为最大模，后者称为  $L^2$  模 (有时  $L^2$  模可以省略下标 2)

\* 无界区间在理论分析时无需截断，将  $u^n$  视为无穷维向量

差分格式一般有两种描述，一种是局部刻画每个点附近的方程，另一种则是写成如  $\mathbb{B}_1 u^{n+1} = \mathbb{B}_0 u^n + \Delta t G^n$  的整体形式。

对一般的双层格式，可以等价变形为类似上方的形式，我们出于可行性要求进一步规定  $\mathbb{B}_1$  可逆，则规范形式为

$$u^{n+1} = \mathbb{B} u^n + \Delta t H^n$$

\* 若差分格式具有相同规范形式，则认为它们是相同的

由于网格加密过程中  $\Delta x, \Delta t$  可以分别趋于 0，分析理论较为困难，因此我们假设趋于 0 时存在加密路径

$$\Delta x = \mathcal{G}(\Delta t)$$

这里  $\mathcal{G}$  为 0 处为 0 的连续函数。若某数值现象同加密路径无关，则称为无条件，否则称有条件。

### §2.2 相容性

\* 我们假设离散对象为偏微分方程  $\mathcal{L}[u] = g$ 。

#### 局部相容

考虑某个网格点处差分方程的局部描述

$$\mathcal{L}_{\Delta x, \Delta t} u_j^n = g_j^n$$

我们记代入真实解得到的差距为**局部截断误差**

$$\tau_j^n = \mathcal{L}_{\Delta x, \Delta t}[u]_j^n - g_j^n$$

若  $\Delta x, \Delta t$  趋于 0 时  $\tau_j^n \rightarrow 0$ , 则称其**逐点相容**于方程; 若存在不可改善的常数  $m_1, m_2$  使得

$$\tau_j^n = O((\Delta x)^{m_1} + (\Delta t)^{m_2})$$

则称  $(m_1, m_2)$  为其**局部截断误差阶**。

\* 几何含义: 差分方程与偏微分方程的**逼近程度**, 也即微分方程解通解满足差分方程的准确程度

\* 考虑第一章的全显格式, 其截断误差为

$$\tau_j^n = \frac{[u]_j^{n+1} - [u]_j^n}{\Delta t} - a \frac{[u]_{j+1}^n - 2[u]_j^n + [u]_{j-1}^n}{(\Delta x)^2} - f_j^n$$

以  $[u]_j^n$  为中心 Taylor 展开, 整理可得到

$$\tau_j^n = \frac{1}{2}[u_{tt}]_j^n \Delta t - \frac{a}{12}[u_{xxxx}]_j^n (\Delta x)^2 + o(\Delta t + (\Delta x)^2)$$

此即知其无条件具有  $(2, 1)$  阶局部截断误差。

\* 类似可知全隐格式也无条件具有  $(2, 1)$  阶局部截断误差。

\* 截断误差阶数与推导过程 [即 Taylor 展开的位置] 无关。

## 整体相容

考虑规范形式下, 真解满足

$$[u]^{n+1} = \mathbb{B}[u]^n + \Delta t H^n + \Delta t \Phi^n$$

其中  $\Phi^n$  为某网格函数, 根据定义可发现其即为**局部截断误差**。

若  $\Delta x, \Delta t$  趋于 0 时, 给定离散范数  $\|\cdot\|$ , 若  $\|\Phi^n\| \rightarrow 0, \forall n$ , 则称差分格式**整体相容**于方程; 若存在不可改善的常数  $m_1, m_2$  使得

$$\|\Phi^n\| = O((\Delta x)^{m_1} + (\Delta t)^{m_2}), \forall n$$

则称  $(m_1, m_2)$  为其  $\|\cdot\|$  **模相容阶** (如最大模相容阶等)。

\* 考虑一般形式  $\mathbb{B}_1 u^{n+1} = \mathbb{B}_0 u^n + \Delta t G^n$ , 将其与规范形式对比可知之前的局部截断误差  $\tau_j^n$  与此处整体相容的网格函数关系为

$$\Phi^n = \mathbb{B}_1^{-1} \tau^n$$

若已知局部截断误差阶, 还需要  $\mathbb{B}_1^{-1}$  在  $\Delta x$  充分小时在诱导范数意义下有界:

$$\|\mathbb{B}^{-1}\| = \sup_{\|v\|=1} \|\mathbb{B}_1^{-1} v\| \leq M$$

才能推出整体相容阶与局部截断误差阶相同。

\* 全显格式由于  $\mathbb{B}_1 = \mathbb{I}$ , 可直接得到最大模整体相容阶为  $(2, 1)$ 。对 HD、HP 与 HI,  $\mathbb{B}_1$  具体形式不同, 不过都可以写成  $\mu \mathbb{A} + \mathbb{I}$ , 由于  $\Delta x \rightarrow 0$  时  $\mu \rightarrow \infty$ , 而  $\mathbb{A}$  行列和有界, 即可以通过逆写成级数估算有界性, 因此都有  $(2, 1)$  阶整体相容性。

\* 绝大多数差分格式都满足要求的一致有界性, 因此无特殊说明时我们只考虑逐点相容性。

## 导数的离散

若函数  $p$  充分光滑,  $m$  阶导数记为  $\mathcal{D}^m p(x)$ , 则可以通过对应阶 Taylor 展开结合待定系数得到使用  $x_{j-l}$  到  $x_{j+r}$  的  $\sigma$  阶误差 [也称其**相容阶**为  $\sigma$ ] 离散公式:

$$\mathcal{D}^m p(x_*) = \sum_{s=-l}^r \alpha_s p(x_{j+s}) + O((\Delta x)^\sigma)$$

这里  $x_*$  为离散焦点  $x_j + \theta\Delta x$ , 记

$$\beta_k = \sum_{s=-l}^r \alpha_s \frac{(s-\theta)^k}{k!}$$

则  $\sigma$  阶误差的公式要求

$$\beta_k = \begin{cases} 0 & k = 0 : (m-1) \\ \frac{1}{(\Delta x)^m} & k = m \\ 0 & k = (m+1) : (m+\sigma-1) \end{cases}$$

\* 例: 计算可得  $x_{j-1}, x_j, x_{j+1}$  处对  $p_{xx}(x_j)$  的最好逼近就是二阶中心差商  $\delta_x^2 p_j$ , 其相容阶为 2。

\* 若  $p$  足够光滑, 通过插值等可以构造近似函数, 从而给出导函数的近似, 例如上例中取二次插值后计算  $x_j$  处导数亦可得到  $\delta_x^2 p_j$  的公式。而若以此三点估算  $p_{xx}(x_{j+1})$ , 则只能有一阶相容性。

回顾之前的差分记号, 也可从差分记号得到导数的离散方式: 假设空间步长  $h = \Delta x$ , 由 Taylor 展开

$$\mathbb{E} = \sum_{k=0}^{\infty} \frac{1}{k!} (h\mathcal{D})^k = e^{h\mathcal{D}}$$

于是有

$$\mathcal{D} = \frac{1}{h} \ln \mathbb{E} = \frac{1}{h} \ln(1 + \Delta)$$

由此展开即可得到  $\mathcal{D}$  或  $\mathcal{D}^k$  用  $\Delta$  的若干阶逼近结果。

\* 类似地,  $\delta = 2 \sinh \frac{h\mathcal{D}}{2}$ , 由此可得到  $\mathcal{D}$  用  $\delta$  的逼近。

\* 也可以对上述  $\ln$  或  $\sinh^{-1}$  采用有理逼近 [Páde 逼近], 这时可能涉及除法, 即对应逆运算。

\* 复指数函数  $e^{ikx}$  可以快速检验差商离散的相容阶。

### §2.3 稳定性

\* 由于计算机无法避免舍入误差, 为了使其引起的误差得到控制, 需要控制扰动的影响。

考虑双层格式的规范形式, 其中时间步长  $n = 0 : (N-1)$ ,  $T = N\Delta t$  代表终止时刻。我们先考虑齐次情况:

令线性差分格式中  $H^n = 0$ , 得到对应的齐次格式  $u^{n+1} = \mathbb{B}u^n, n = 0 : (N-1)$ , 给定离散范数后, 若  $\Delta x, \Delta t$  趋于 0 时存在与  $\Delta x, \Delta t, u^0$  无关的  $K$  使得其数值解满足

$$\|u^n\| \leq K \|u^0\|, \forall n = 0 : N$$

则称其按对应模具有初值稳定性 [若  $K$  与  $T$  有关, 则称为短时间初值稳定性, 否则称长时间初值稳定性]。

\* 由于齐次性, 初值进行扰动时解的差距关于扰动亦有界, 这就是稳定的几何含义。

\* 当且仅当  $\mu a \leq \frac{1}{2}$  时 HP 的全显格式有最大模稳定性: 注意到格点处方程为

$$u_j^{n+1} = \mu a (\mu_{j-1}^n + \mu_{j+1}^n) + (1 - 2\mu a) u_j^n$$

$\mu a \leq \frac{1}{2}$  时右端系数非负且和为 1, 从而  $|u_j^{n+1}| \leq \|u^n\|$ , 于是  $\|u^{n+1}\| \leq \|u^n\|$ , 即得证。否则, 不妨考虑  $\frac{1}{\Delta x}$  为偶数时, 初值  $u_j^0 = (-1)^j$ , 可归纳得  $u_j^n = (1 - 4\mu a)^n (-1)^j$  为数值解,

\* HP 的全隐格式无条件具有最大模稳定性: 此时类似上方化简可得到  $(1 + 2\mu a) |u_j^{n+1}| \leq \|u^n\| + 2\mu a \|u^{n+1}\|$ , 从而变形即得无条件最大模稳定。

\* 对 HD 与 HI, 基本类似可得与上方相同的结论, 不过全显格式  $\mu a > \frac{1}{2}$  的不稳定未必容易构造, 例如对 HI 的构造为

$$u_j^n = \sum_{k=0}^{\infty} \exp(-2^k) \left(1 - 4\mu a \sin^2(2^{k-1}\pi\Delta x)\right)^n \cos(2^k\pi j\Delta x)$$

再取  $\Delta x = 2^{-m}$ , 证明 0 处发散。

\* 事实上, 三种格式的  $L^2$  模稳定性结论也完全相同, 之后会给出证明。

非齐次情况下, 我们设差分格式满足  $u^0 = 0$ , 给定离散范数后, 若  $\Delta x, \Delta t$  趋于 0 时存在与  $\Delta x, \Delta t, u^0$  无关的  $M$  使得其数值解满足

$$\|u^n\| \leq M \sum_{m=0}^{n-1} \|H^m\| \Delta t, \forall n = 1 : N$$

则称其按对应模具有**右端项稳定性**。

\* 事实上, **初值稳定性蕴含右端项稳定性**, 因此一般只需考虑初值稳定性。证明: 类似线性微分方程的 Duhamel 原理, 我们将解  $u^n$  分裂为  $\sum_{m=0}^{n-1} v_{(m)}^n$ , 使得 (相当于以每个非齐次项作为新的初值形成的一系列齐次问题)

$$v_{(m)}^{l+1} = \begin{cases} 0 & l < m \\ \Delta t H^m & l = m \\ \mathbb{B} v_{(m)}^l & l > m \end{cases}$$

则直接根据初值稳定性可以得到取  $M = K$  记为要找的界。

### 稳定性分析

对稳定性的分析一般较为复杂, 常用的如上面已使用的**离散最大模原理**、Fourier 方法等。下面先介绍直接矩阵法与分离变量法, 再集中讨论 **Fourier 方法**。

**直接矩阵法:** 以分析 HD 全显格式的初值稳定性为例, 由于  $\mathbb{B} = \text{tridiag}(\mu a, 1 - 2\mu a, \mu a)$ , 齐次格式解为  $u^{n+1} = \mathbb{B}^{n+1} u^0$ , 分析相似标准形可知若  $\mathbb{B}$  最大特征值模长 [由对称, 此即为**谱范数**] 不超过 1,  $\mathbb{B}^n$  任何元素有界, 否则将存在无界元素。由此可知, 无论对最大模还是  $L^2$  模, 稳定性都等价于  $\mathbb{B}$  谱范数不超过 1。而其全部特征值为

$$\lambda_s = 1 - 4\mu a \sin^2 \frac{s\pi}{2J}, s = 1 : (J-1)$$

由此即得  $\mu a > \frac{1}{2}$  时只要  $J$  充分大就有谱范数大于 1, 否则谱范数恒小于 1, 因此全显格式最大模或  $L^2$  模稳定当且仅当  $\mu a \leq \frac{1}{2}$ 。

**分离变量法:** 以分析 HD 全隐格式的初值稳定性为例, 由于  $\mathbb{B}_1 u^{n+1} = u^n$  中,  $\mathbb{B}_1$  对称, 其特征向量  $\{v_k\}$  可构成单位正交基, 于是归纳得, 若  $u_0 = \sum_k \alpha_k v_k$ , 则全隐格式数值解为

$$u^n = \sum_{k=1}^{J-1} \alpha_k \lambda_k^{-n} v_k$$

计算可知  $\mathbb{B}_1$  全部特征值为  $1 + 4\mu a \sin^2 \frac{s\pi}{2J}, s = 1 : (J-1)$ , 均大于 1, 因此即知最大模或  $L^2$  模均无条件稳定。

### Fourier 方法

考虑两种情况:

1.  $u$  为整个空间数轴上的网格函数  $\{u_m\}_{-\infty}^{\infty}$ , 且其  $L^2$  模有限, 称为**速降网格函数**。

考虑将每点延拓到附近的区域, 构造  $\tilde{u}(x) = u_m, x \in ((m-1/2)\Delta x, (m+1/2)\Delta x)$ , 根据速降网格函数定义可知  $\|\tilde{u}\|_{L^2} = \|u\|_2$ , 于是其平方可积, 进行 Fourier 变换可得到存在  $\hat{u}$  使得

$$\hat{u}(k) = \mathcal{F}\tilde{u}(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{0ikx} \tilde{u}(x) dx, \quad \tilde{u}(x) = \mathcal{F}^{-1}\hat{u}(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ikx} \hat{u}(k) dk$$

利用 Parseval 恒等式可知  $\|\hat{u}\|_{L^2} = \|\tilde{u}\|_{L^2}$ 。

\* 此方法也可类似推广到高维网格或周期性网格。Fourier 理论中,  $\tilde{u}$  为时域函数,  $\hat{u}$  为频域函数, Parseval 恒等式表示时域函数能量等于所含简谐波的能量和。



2.  $u$  为周期性网格函数, 不妨设周期为  $[-\pi, \pi]$ , 考虑等距空间网格  $\mathcal{T}_{\Delta x}$ , 其中  $\Delta x = \frac{\pi}{J}$  为空间步长, 初始点  $-\pi$ , 则周期网格函数可以记为  $\{u_m\}_{-J}^J$ , 其  $L^2$  模视为一个周期上的  $L^2$  模 (即左右端点只计算一个), 且要求有限。

\* 其可以类似速降网格进行离散, 但往往不是最好的诠释方式, 因为它将有限网格点拆分为了无穷多简谐波。

由于  $2J$  个格点的网格最多能辨识出  $2J$  个不同波数的简谐波, 我们可以构造出 Fourier 级数 [这里求和号上的撇代表首项和末项只取  $\frac{1}{2}$ ]

$$\begin{aligned}\check{u}_k &= \frac{1}{\sqrt{2\pi}} \sum_{m=-J}^J{}' e^{-imk\Delta x} u_m \Delta x, k = -J : J \\ u_m &= \frac{1}{\sqrt{2\pi}} \sum_{k=-J}^J{}' e^{imk\Delta x} \check{u}_k \Delta x, m = -J : J\end{aligned}$$

这时对应的 Parseval 恒等式为

$$\|u\|_2^2 = \sum_{k=-J}^J{}' |\check{u}_k|^2$$

我们以纯初值问题 (即空间范围为整个  $x$  轴) 为例介绍双层格式的 Fourier 方法, 假设齐次情况下方程对所有点一致, 写为

$$\sum_{s=-l_1}^{r_1} a_s u_{j+s}^{n+1} = \sum_{s=-l_0}^{r_0} b_s u_{j+s}^n$$

**计算增长因子:** 按照之前第一类情况进行 Fourier 变换, 结合平移性质  $\mathcal{F}[f(x+a)] = e^{iak} \hat{f}(k)$  计算可得增长因子

$$\lambda(k) = \frac{\hat{u}^{n+1}(k)}{\hat{u}^n(k)} = \frac{\sum_{s=-l_0}^{r_0} b_s e^{isk\Delta x}}{\sum_{s=-l_1}^{r_1} a_s e^{isk\Delta x}}$$

\* 由解的形式可以发现, 任取波数  $k \in \mathbb{R}$ , 将  $u_j^n = \lambda^n e^{ikj\Delta x}$  代入即可算出完全相同结果的增长因子。此形式中,  $\lambda^n$  暗含振幅的增长。

**von Neumann 条件判定:** 利用 Parseval 恒等式, 递推可得

$$\|u^n\|_2 = \|\hat{u}^n\|_{L^2} \leq \left( \sup_k |\lambda(k)| \right)^n \|\hat{u}^0\|_{L^2} = \left( \sup_k |\lambda(k)| \right)^n \|u^0\|_2$$

由此可知, 给定终止时刻  $T$ , 记  $N = \lceil \frac{T}{\Delta t} \rceil$ , 若存在  $K$  使得

$$|\lambda(k)|^n \leq K, \forall k \in \mathbb{R}, n = 0 : N$$

则双层格式具有  $L^2$  模稳定性。反过来说, 由于  $u_0^n$  取  $e^{ikj\Delta x}$  实部即可取到倍数达到  $\lambda(k)^n$  的解, 上式的要求是充要的。

上述条件的判定往往是困难的, 不过, 若上式成立, 由于  $\Delta t$  充分小时有  $N\Delta t \geq \frac{T}{2}$ , 可得

$$|\lambda(k)| \leq K^{1/N} \leq 1 + \frac{K}{N} \leq 1 + \frac{2K\Delta t}{T}$$

也  $\Delta t$  充分小时存在  $C$  使得  $|\lambda(k)| \leq 1 + C\Delta t, \forall k \in \mathbb{R}$ , 这个必要条件称为 **von Neumann 条件**。

事实上 von Neumann 条件是充要的, 因为其被常数量控制, 可推出  $|\lambda(k)|^n \leq e^{CT}$ 。

\* 若  $\lambda(k)$  表达式中无时间步长, 则条件中可取  $C = 0$ , 即  $|\lambda(k)| \leq 1$ , 称为严格的 **von Neumann 条件**, 界定常数  $K = 1$ 。

例：分析 HI 与 HP 的全显/全隐格式  $L^2$  模稳定性。直接代入  $\lambda^n e^{ikj\Delta x}$  可知全显格式增长因子为

$$\lambda(k) = 1 - 4\mu a \sin^2 \frac{k\Delta x}{2}$$

于是 von Neumann 条件即等价于  $\mu a \leq \frac{1}{2}$ ，这就是  $L^2$  模稳定的充要条件。对全隐格式，有

$$\lambda(k) = \left(1 + 4\mu a \sin^2 \frac{k\Delta x}{2}\right)^{-1}$$

无条件  $L^2$  模稳定。

## §2.4 收敛性

讨论完相容性 (差分方程能否近似微分方程) 与稳定性 (误差下数值解能否接近准确数值解) 后，我们最终需要考虑收敛性 (数值解是否能逼近真解)。类似于相容性，我们可以做局部与全局的定义：

设  $[u]$  为真解，若当  $\Delta t^{(k)}$  趋于 0， $\Delta x^{(k)}$  沿某条加密路径趋于 0 时，对计算区域任何点  $(x_*, t^*)$ ，数值解序列  $u^{(k)}$  中可找到  $j^{(k)}, n^{(k)}$  使得

$$j^{(k)} \Delta x^{(k)} \rightarrow x_*, \quad n^{(k)} \Delta t^{(k)} \rightarrow t^*, \quad (u_j^n)^{(k)} \rightarrow [u](x_*, t^*)$$

则称差分格式 (逐点) 收敛于定解问题。

由真解连续性，事实上只需要保证格点处的收敛，因此可简化为  $e_j^n = [u]_j^n - u_j^n \rightarrow 0, \forall j, n$ ，给定离散范数  $\|\cdot\|$ ，若数值误差  $e^n$  满足  $\|e^n\| \rightarrow 0, \forall n$ ，则称差分格式按对应模收敛于定解问题，此即为整体收敛性。若存在不可改善的常数  $m_1, m_2$  使得

$$\|e^n\| = O((\Delta x)^{m_1} + (\Delta t)^{m_2})$$

则称差分格式按对应模具有  $(m_1, m_2)$  阶精度 (或误差)。

\* 以 HP 的全显格式为例：若其真解足够光滑，可直接利用差分格式的线性结构得到误差迭代的方程

$$e_j^{n+1} = \mu a (e_{j-1}^n + e_{j+1}^n) + (1 - 2\mu a) e_j^n + \Delta t \tau_j^n$$

这里  $\tau_j^n$  为局部截断误差。由于进行了光滑性假设，已有  $\tau_j^n = O((\Delta x)^2 + \Delta t)$ ，在  $\mu a \leq \frac{1}{2}$  的条件下，类似稳定性证明即有估计

$$\|e^{n+1}\|_\infty \leq \|e^n\|_\infty + \Delta t \|\tau^n\|_\infty$$

由  $\|e^0\|_\infty = 0$  即可知

$$\|e^n\|_\infty \leq \sum_{m=0}^{n-1} \Delta t \|\tau^m\|_\infty$$

由于  $n\Delta t \leq T$ ，即有

$$\max_{n|n\Delta t \leq T} \|e^n\|_\infty = O((\Delta x)^2 + \Delta t)$$

即误差阶亦为 (2, 1)。

\* 相容性、稳定性、收敛性与加密路径有关 [有条件、无条件定义区分]，与离散范数选取有关，而相容性最容易分析，收敛性最难分析。不过，在上例中，能感受到收敛性与稳定性论证相似性，这事实上暗示了它们有某种共通性。

**Lax-Richtmyer 等价定理：**假设线性微分方程定解问题是适定的 [存在唯一解，且连续依赖于初边值条件]。若线性差分格式是相容的，则稳定性与收敛性等价，且误差阶不低于相容阶，

\* 此定理充分性证明类似前例分析过程，但必要性则需要泛函分析的共鸣定理。

\* 由此，此后进行数值性态分析一般只关心相容性与稳定性，一般不进行繁琐的收敛性证明。

## 三 热传导方程

本节依然讨论线性常系数热传导方程

$$u_t = au_{xx}, a > 0$$

及各种定解问题的差分。未指出边界条件则默认纯初值问题或周期边值问题。

### §3.1 相容性改进

\* 古典格式相容阶仅为  $(2, 1)$ ，空间网格需要足够密集才能达到精度，对全显格式，时间网格还需要处于空间网格的平方量级。由此，我们希望能改善相容阶。

#### 加权平均格式

将两个古典格式组合起来得到加权平均格式：

$$\Delta_t u_j^n = \theta \mu a \delta_x^2 u_j^{n+1} + (1 - \theta) \mu a \delta_x^2 u_j^n$$

\* 这里  $\theta \in [0, 1]$ ，两头分别为全显格式与全隐格式，此格式也称为  $\theta$  格式、六点格式或 Rose 格式。

以  $(x_j, t^{n+1/2})$  为中心进行展开，直接计算可得

$$\tau_j^n = -\Delta t \left( \theta - \frac{1}{2} \right) [u_{tt}]_j^{n+1/2} - \frac{a(\Delta x)^2}{12} [u_{xxxx}]_j^{n+1/2} + O((\Delta x)^4 + (\Delta t)^2)$$

也即其至少具有  $(2, 1)$  阶截断误差，而取  $\theta = \frac{1}{2}$  所得到的 **CN 格式** [Crank-Nicolson 格式] 即无条件具有  $(2, 2)$  阶截断误差。

注意到  $[u_{tt}] = [u_{xxxx}]$ ，截断误差表达式可改写为

$$\tau_j^n = -\Delta t \left( \frac{1}{12\mu a} + \theta - \frac{1}{2} \right) [u_{tt}]_j^{n+1/2} + O((\Delta x)^4 + (\Delta t)^2)$$

若等距时空网格满足  $\mu a = \frac{1}{6-12\theta}$ ，则得到的 **Douglas 格式** 达到了  $O((\Delta x)^4)$ ，即整体四阶相容。

\* 由于 Douglas 格式网比固定， $\Delta t$  可合并到  $\Delta x$  中

下面考虑加权平均格式的**稳定性**。根据 Lax-Richtmyer 等价定理，只要具有稳定性，这些格式即能具有高阶精度。

代入  $\lambda^n e^{ikj\Delta x}$  可算出加权平均格式的增长因子为

$$\lambda(k) = \frac{1 - \mu a(1 - \theta) \sin^2(k\Delta x/2)}{1 + 4\mu a\theta \sin^2(k\Delta x/2)}$$

其绝对值不超过 1 即等价于

$$\mu a(1 - 2\theta) \leq \frac{1}{2}$$

因此，当  $\theta < \frac{1}{2}$  时，其称为**偏显格式**，有条件  $L^2$  模稳定，而  $\theta \geq \frac{1}{2}$  时称为**偏隐格式**，无条件  $L^2$  模稳定。

\* 由于 CN 格式属于偏隐格式，其应用非常广泛。

利用离散最大模原理进行拆分估算，可得网比满足

$$\mu a(1 - \theta) \leq \frac{1}{2}$$

时加权平均格式必然最大模稳定。也即直接估算得到的  $L^2$  模与最大模稳定性结论不同。

\* 但事实上，此例子中  $\mu a(1 - \theta) \leq \frac{1}{2}$  仅为充分条件，还是可以证明  $\mu a(1 - 2\theta) \leq \frac{1}{2}$  就足以最大模稳定。

\* 对问题 HD，若真解对  $t$  二阶导数符号不变，则可证明数值误差也有固定符号，形成**单侧逼近性质**。此时全显格式与全隐格式符号恰好相反，这也说明了加权平均格式能增加精度的原因。

\* 理论上高阶格式的数值误差在特定的时空网格下未必比低阶格式更小，这取决于实践中网格的密度与真解的光滑性。

### 三层格式分析

相比综合考虑各个导数离散的影响，直接扩张离散模板的思路是更加简单的，但是，对时间导数而言，扩张离散模板会导致出现更多时间层，从而引起复杂的问题。

利用一阶中心差商离散时间导数可以得到最基本的 Richardson 格式：

$$u_j^{n+1} = u_j^{n-1} + 2\mu a \delta_x^2 u_j^n$$

其为显式格式，且可验证其无条件具有 (2, 2) 阶局部截断误差。

\* 由离散模板，此形式也称为实心十字架格式，然而，其稳定性并不好，因此不能用于大规模数值计算，我们接下来将分析此算法的数值稳定表现。

考虑一般的三层格式

$$\sum_{s=-l_1}^{r_1} a_s^{(1)} u_{j+s}^{n+1} = \sum_{s=-l_0}^{r_0} a_s^{(0)} u_{j+s}^n + \sum_{s=-l_{-1}}^{r_{-1}} a_s^{(-1)} u_{j+s}^{n-1}$$

Fourier 方法分析一般分为三步。

#### 1. 转化为双层格式

引入辅助函数  $v^n$ ，常取如  $v^n = u^{n-1}$ ，将三层格式写为等价的双层格式

$$\sum_{s=-l'_1}^{r'_1} \mathbb{A}_s^{(1)} w_{j+s}^{n+1} = \sum_{s=-l'_0}^{r'_0} \mathbb{A}_s^{(0)} w_{j+s}^n, \quad w^n = (u^n, v^n)^T$$

这里  $\mathbb{A}_s$  均为二阶矩阵，与网格函数、网格点位置无关，与网格参数相关。

标量型三层格式  $L^2$  模稳定性定义：存在  $K_1$  使得  $\|u^n\| \leq K_1(\|u^0\| + \|u^1\|)$ ；

向量型双层格式  $L^2$  模稳定性定义：存在  $K_2$  使得  $\|w^n\| \leq K_2\|w^0\|^2$ ，这里  $\|w^n\|^2 = \|u^n\|^2 + \|v^n\|^2$ 。

由辅助网格函数的定义可知二者等价。

#### 2. 双层格式稳定必要条件

代入  $w_j^n = \hat{w}^n(k) e^{ikj\Delta x}$ ，这里  $\hat{w}^n$  为向量，可以解得

$$\hat{w}^{n+1} = \mathbb{G} \hat{w}^n$$

这里  $\mathbb{G}(k; \Delta t)$  为此格式的增长矩阵。

记  $\rho(\mathbb{G})$  为其谱半径 [最大特征值模长]， $\|\mathbb{G}\|$  为其二范数 [向量诱导的二范数，或谱范数]，给定终止时刻  $T$ ，记  $N = \lceil \frac{T}{\Delta t} \rceil$ ，则方法稳定等价于存在  $K$  使得

$$\|\mathbb{G}^n(k)\| \leq K, \forall k \in \mathbb{R}, n = 0 : N$$

注意到特征信息，可知  $\rho(\mathbb{G}) \leq \|\mathbb{G}\|$ ，因此必要条件为  $\rho(\mathbb{G})^n \leq K$ ，这即对应 von Neumann 条件， $\Delta t$  充分小时存在  $C$  使

$$\rho(\mathbb{G}) \leq 1 + C\Delta t$$

\* 类似地，若  $\mathbb{G}(k)$  表达式中无时间步长，则条件中可取  $C = 0$ ，即  $\rho(\mathbb{G}(k)) \leq 1$ ，称为严格的 von Neumann 条件，界定常数  $K = 1$ 。

\* 由此取  $v^n = u^{n-1}$  直接计算知 Richardson 格式地增长矩阵

$$\mathbb{G}(k) = \begin{pmatrix} -8\mu a \sin^2 \frac{k\Delta x}{2} & 1 \\ 1 & 0 \end{pmatrix}$$

直接计算可知其恒有大于  $1 + \mu a$  的特征值, 因此无条件  $L^2$  模不稳定。

### 3. 双层格式稳定充分条件

我们下面列举一些条件使得满足 von Neumann 条件时就能  $L^2$  模稳定 [Kreiss 定理]:

- $\Delta t$  充分小时  $\mathbb{G}$  正规, 即  $\mathbb{G}^H \mathbb{G} = \mathbb{G} \mathbb{G}^H$ ;
- $\Delta t$  充分小时  $\mathbb{G}$  所有元素关于  $k$  一致有界, 且模较小的特征值满足  $|\lambda_2| \leq \delta < 1$ ;
- $\Delta t$  充分小时  $\mathbb{G}$  具有完备的单位特征向量组, 拼成矩阵  $\mathbb{Q}$  后满足  $|\det \mathbb{Q}(k; \Delta t)| \geq \delta > 0$ 。
- $\mathbb{G}$  与  $\Delta t$  无关, 写为  $\mathbb{G} = \tilde{\mathbb{G}}(\xi)$ , 这里  $\xi = k\Delta x$ , 且抛物型方程网比  $\mu$  或双曲型方程网比  $\nu = \frac{\Delta t}{\Delta x}$  恒定。对任意  $\xi$ , 成立  $\tilde{\mathbb{G}}(\xi)$  特征值互异;
- 上述  $\tilde{\mathbb{G}}$  条件下, 对任意  $\xi$ , 存在正整数  $s$  使得  $\tilde{\mathbb{G}}(\xi)$  的  $0 : (s-1)$  阶导数为数量阵,  $s$  阶导数特征值互异。
- 上述  $\tilde{\mathbb{G}}$  条件下, 对任意  $\xi$ ,  $\rho(\tilde{\mathbb{G}}(\xi)) < 1$ 。

\* 若双层格式满足 von Neumann 条件但不稳定, 利用 Lax-Richtmyer 等价定理可知若收敛则无法收敛到真解, 此情况实践中容易误判为可信。

\* 直接将  $u_j^n = \lambda^n e^{ikj\Delta x}$  代入多层格式, 事实上可得到特征方程, 从而判定多层格式的 von Neumann 条件, 不过这不足以分析出收敛性。

### 实用的三层格式

**DF 格式** [Du Fort-Frankel 格式、空心十字架格式]:

$$u_j^{n+1} = u_j^{n-1} + 2\mu a(u_{j-1}^n - u_j^{n+1} - u_j^{n-1} + u_{j+1}^n)$$

其将中心点值换为时域两侧平均, 直接展开可知局部截断误差的阶数为

$$\tau_j^n = O\left((\Delta x)^2 + (\Delta t)^2 + \frac{(\Delta t)^2}{(\Delta x)^2}\right)$$

当网比固定时, 截断误差即为  $O((\Delta x)^2)$ , 其为有条件相容。

\* 由截断误差形式, DF 格式事实上更靠近方程  $u_t = au_{xx} - \mu a \Delta t u_{tt}$ , 这展现了一些修正方程的思想。

考虑  $v^n = u^{n-1}$ , 代入计算可得增长矩阵为

$$\mathbb{G}(k) = \frac{1}{1+2\mu a} \begin{pmatrix} 4\mu a \cos(k\Delta x) & 1-2\mu a \\ 1+2\mu a & 0 \end{pmatrix}$$

利用韦达定理即可推出其两特征值模都不超过 1, 严格 von Neumann 条件无条件成立, 再结合 Kreiss 定理第二条, 由

$$|\lambda_1 \lambda_2| = \frac{|1-2\mu a|}{|1+2\mu a|}$$

取  $\delta = \frac{|1-2\mu a|^{1/2}}{|1+2\mu a|^{1/2}}$  即知其无条件  $L^2$  模稳定。

\* 此格式展示了数值格式的微弱变化引起表现的明显差异

**基于双层平均的三层格式:** 设  $\theta \in [1/2, 1]$  为给定权重, 加权平均空间导数离散可以得到:

$$u_j^{n+1} = u_j^{n-1} + 2\mu a((1-\theta)\delta_x^2 u_j^n + \theta\delta_x^2 u_j^{n-1})$$

$$u_j^{n+1} = u_j^n + \mu a ((1 + \theta) \delta_x^2 u_j^n - \theta \delta_x^2 u_j^{n-1})$$

两者都无条件具有 (2, 1) 阶局部截断误差; 前者称为**内插格式**,  $L^2$  模稳定当且仅当  $2\theta\mu a > 1$  或  $\theta = \frac{1}{2\mu a} > \frac{1}{2}$  时  $L^2$  模稳定; 后者称为**外插格式**, 当且仅当  $2(1 + 2\theta)\mu a \leq 1$  时  $L^2$  模稳定。

\* 当  $\theta = \frac{1}{2}$  时后者称为**外插 CN 格式**, 无条件具有 (2, 2) 阶局部截断误差。

**基于三层平均的三层格式**: 设  $\theta \in [0, 1/2]$  为给定权重, 加权平均空间导数离散可以得到:

$$u_j^{n+1} - u_j^{n-1} = 2\mu a (\theta \delta_x^2 u_j^{n+1} + (1 - 2\theta) \delta_x^2 u_j^n + \theta \delta_x^2 u_j^{n-1})$$

其无条件具有 (2, 2) 阶局部截断误差, 无条件  $L^2$  模稳定当且仅当  $\theta \in [1/4, 1/2]$ 。

**BDF 格式**: 利用 ODE 的 BDF[backward difference formula] 技术离散时间导数得到, 取  $\theta > 0$  有:

$$(1 + \theta)(u_j^{n+1} - u_j^n) - \theta(u_j^n - u_j^{n-1}) = \mu a \delta_x^2 u_j^{n+1}$$

当  $\theta = \frac{1}{2}$  时称为 **Richtmyer 格式**, 无条件具有 (2, 2) 阶局部截断误差与  $L^2$  模稳定性。

\* 三层格式启动一般有两种思路, 假设真解充分光滑, 在**边界处利用空间导数估算时间导数**, 或**直接采用二层格式** [由于只需要一步, 此处只需时间一阶相容性, 也无需考虑稳定性]。

### §3.2 计算效率改进

\* 本部分即讨论各种格式中如何调整实现细节以改进计算效率

#### 时间步长轮替

均匀网格未必是最佳选择, 例如函数光滑性好的部分选择粗空间网格、光滑性差的部分选择细空间网格, 即可大幅改善效率, 这引出了某些**自适应**的思想。

对时间网格, 全显格式  $L^2$  模稳定时, 等距网格要求  $\Delta t \leq \frac{(\Delta x)^2}{2a}$ , 记右侧为  $\Delta t_m$ 。有如下结论:

循环交替使用时间步长  $\Delta t_1, \Delta t_2$ , 则保持稳定时平均时间步长可达到  $2\Delta t_m$ 。

注意到交替时  $L^2$  模稳定等价于 [设两个时间步长对应网比  $\mu_1, \mu_2$ ]

$$\forall k \in \mathbb{R}, \quad \left(1 - 2\mu_1 a \sin^2 \frac{k\Delta x}{2}\right) \left(1 - 2\mu_2 a \sin^2 \frac{k\Delta x}{2}\right) \leq 1$$

利用二次函数知识可分析得到, 取  $\mu_{1,2}a = 1 \pm \frac{\sqrt{2}}{2}$  即最佳, 平均时间步长  $2\Delta t_m$ 。

\* 时间步长序列越长, 平均时间步长即能越大。

#### 显隐格式交替

按空间标号进行分组 [**半隐格式**]:

$$u_{2m+1}^{n+1} = u_{2m+1}^n + \mu a \delta_x^2 u_{2m+1}^n$$

$$u_{2m}^{n+1} = u_{2m}^n + \mu a \delta_x^2 u_{2m}^{n+1}$$

\* 虽然对偶数点是隐式, 由于奇数点已显式计算出, 对偶数点计算可显式完成。

\* 将奇数点偶数点分别视为向量值函数, 代入

$$u_{2m+1}^n = \hat{w}_1^n e^{ik(2m+1)\Delta x}, \quad u_{2m}^n = \hat{w}_2^n e^{i2km\Delta x}$$

可利用二次函数知识解得  $L^2$  模稳定当且仅当  $\mu a \leq 1$ 。

按时空指标和进行分组 [**跳点格式**]:

$$u_j^{n+1} = \begin{cases} u_j^n + \mu a \delta_x^2 u_j^n & n + j \equiv 1 \pmod{2} \\ u_j^n + \mu a \delta_x^2 u_j^{n+1} & n + j \equiv 0 \pmod{2} \end{cases}$$

\* 相当于半隐格式交替先计算奇数点/偶数点

\* 对奇数点或偶数点等同于 DF 格式，无条件  $L^2$  模稳定。

按时间标号进行分组 [显隐交替格式]：先显再隐

$$u_j^{2m+1} = u_j^{2m} + \mu a \delta_x^2 u_j^{2m}, \quad u_j^{2m+2} = u_j^{2m+1} + \mu a \delta_x^2 u_j^{2m+1}$$

或先隐再显

$$u_j^{2m+1} = u_j^{2m} + \mu a \delta_x^2 u_j^{2m+1}, \quad u_j^{2m+2} = u_j^{2m+1} + \mu a \delta_x^2 u_j^{2m+1}$$

\* 两步看作一次整体迭代，计算得增长因子

$$\lambda(k) = \frac{1 - 4\mu a \sin^2(k\Delta x/2)}{1 + 4\mu a \sin^2(k\Delta x/2)} < 1$$

因此无条件  $L^2$  模稳定。

\* 看作整体迭代合并两式得到过程，可展开发现其有 (2,2) 阶局部相容性 [先显后隐在奇数时间层上事实上等价于 Richardson 格式，不过由于偶数层以 CN 格式迭代，增强了稳定性]。

### 非对称格式

**Saul'ev 格式**偏右版本：

$$u_j^{n+1} = u_j^n + \mu a (u_{j+1}^{n+1} - u_j^{n+1} - u_j^n + u_{j-1}^n)$$

\* 构造思路为用半步中心差商离散时空导数，局部截断误差  $O((\Delta t)^2 + (\Delta x)^2 + \frac{\Delta t}{\Delta x})$ ，网比固定时仅整体一阶相容性。

\* 直接计算增长因子得无条件  $L^2$  模稳定。

\* 偏左版本：

$$u_j^{n+1} = u_j^n + \mu a (u_{j-1}^{n+1} - u_j^{n+1} - u_j^n + u_{j+1}^n)$$

面对 Dirichlet 边值时可显式计算，偏左、偏右可对应方向的扫描。

**循环扫描** (交替使用两版本)：

$$u_j^{n+1} = u_j^n + \mu a (u_{j-1}^{n+1} - u_j^{n+1} - u_j^n + u_{j+1}^n)$$

$$u_j^{n+2} = u_j^{n+1} + \mu a (u_{j+1}^{n+2} - u_j^{n+2} - u_j^{n+1} + u_{j-1}^{n+1})$$

\* 两步看作一次迭代，局部截断误差为  $O((\Delta t)^2 + (\Delta x)^2 + \frac{(\Delta t)^2}{(\Delta x)^2})$ ，固定网比可二阶相容。

**平均策略**：

$$u_j^{n+1} - u_j^n = \mu a \delta_x^2 u_j^{n+1} + \mu^2 a^2 \delta_x^2 (u_j^{n+1} - u_j^n) - \frac{1}{2} \mu^2 a^2 \delta_x^4 u_j^n$$

\* 实质是每步分别用偏左偏右算出结果后再平均

\* 局部截断误差为  $O(\Delta t + (\Delta x)^2 + \frac{(\Delta t)^2}{(\Delta x)^2})$ ，固定网比亦为二阶相容。

**显式分组格式**：对目标时间层的  $u_{2m}^{n+1}$  与  $u_{2m+1}^{n+1}$  两网格点分别搭建偏左、偏右的 Saul'ev 格式：

$$\begin{cases} (1 + \mu a) u_{2m+1}^{n+1} - \mu a u_{2m}^{n+1} = (1 - \mu a) u_{2m+1}^n + \mu a u_{2m+2}^n \\ -\mu a u_{2m+1}^{n+1} + (1 + \mu a) u_{2m}^{n+1} = \mu a u_{2m-1}^n + (1 - \mu a) u_{2m}^n \end{cases}$$

联立求解得到

$$\begin{cases} u_{2m}^{n+1} = \kappa_1 u_{2m-1}^n + \kappa_2 u_{2m}^n + \kappa_3 u_{2m+1}^n + \kappa_4 u_{2m+2}^n \\ u_{2m+1}^{n+1} = \kappa_4 u_{2m-1}^n + \kappa_3 u_{2m}^n + \kappa_2 u_{2m+1}^n + \kappa_1 u_{2m+2}^n \end{cases}$$

$$\kappa = \left( \frac{\mu a(1 + \mu a)}{1 + 2\mu a}, \frac{1 - \mu^2 a^2}{1 + 2\mu a}, \frac{\mu a(1 - \mu a)}{1 + 2\mu a}, \frac{\mu^2 a^2}{1 + 2\mu a} \right)$$

\* 空间网格点两两分组，便于并行推进。

\* 与 Saul'ev 格式相容阶相容，当且仅当  $\mu a \leq 1$  时  $L^2$  模稳定。

\* 可将显式分组的思想推广到多个网格点，每组内部使用隐式迭代，边界采用 Saul'ev 的某个版本以降低耦合。

### §3.3 误差估计与收敛分析

**强正则性假设：**真解在整个计算区域上具有所需的高阶导数，使得差分格式处处具有清晰的局部截断误差阶。

**基本操作步骤：**得到局部误差阶  $\tau_j^n$  后，先利用线性叠加原理，可知  $\mathcal{L}u_j^n = f_j^n$  的误差方程应为  $\mathcal{L}e_j^n = \tau_j^n$  的像是，再以此建立误差函数在离散范数度量下的演化规律，最终不等式放缩得到估计 [类似差分格式的右端项稳定性分析过程]。

\* 此步骤即为 Lax-Richtmyer 等价定理充分性的证明。

\* 例：CN 格式最大模误差估计，简单计算得误差方程

$$(1 + \mu a)e_j^{n+1} = \frac{1}{2}\mu a \sum_{l \in \{-1, 1\}} \sum_{s \in \{0, 1\}} e_{j+l}^{n+s} - (1 - \mu a)e_j^n + \Delta t \tau_j^n$$

利用凸组合性质即可得到

$$\|e^{n+1}\|_\infty \leq \|e\|_\infty^n + \Delta t \|\tau^n\|_\infty$$

进一步组合，利用  $(n+1)\Delta t \leq T$  即可得到 CN 格式无条件具有与相容阶相容的最大模误差，即 (2, 2) 阶最大模误差。

\* 一般情况强正则未必满足，即只能满足某些弱正则性假设。

考虑热传导方程的周期边值问题，初值为以下两者之一

$$u_0^{(1)}(x) = \begin{cases} 1 & x \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ 0 & x \in [-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi] \end{cases}, \quad u_0^{(2)}(x) = \pi - |x|$$

二者都不具有高阶光滑性 (下称为问题 HP-WEAK)，但实验可发现全显格式仍有一阶到二阶的精度。

结论：若  $u_0(x)$  平方可积且有界 [即本例的弱正则性要求]，当  $\mu a \leq \frac{1}{2}$  时，HP-WEAK 的全显格式收敛。

证明：利用分离变量法，真解可精确表示为

$$[u]_j^n = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} a_k e^{ijk\Delta x} (e^{-ak^2\Delta t})^n$$

而数值解为

$$u_j^n = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} A_k e^{ijk\Delta x} \lambda(k)^n$$

$\lambda(k)$  即为增长因子，由之前分析，满足  $\mu a \leq \frac{1}{2}$  时模长不超过 1，展开系数  $a_k$  与  $A_k$  均为初值 Fourier 级数，但  $a_k$  对应真实初值的级数， $A_k$  对应初值在格点处的值常值延拓拼成整个区域后得到的系数。

将误差分解为  $a_k, A_k$  引起的误差、后端  $|k| > M$  项引起的误差与  $|k| \leq M$  项引起的误差三部分，第一部分可由 Parseval 恒等式控制，第二部分  $M$  充分大时可控制，第三部分即反映局部误差的累计，估算可得其

$$\leq \left( \sum_{k \leq |M|} |a_k|^2 k^8 \right)^{1/2} CT \Delta t$$

因此在  $\Delta t$  充分小时仍可控制，于是对任何  $\varepsilon$  都可选取充分大的  $M$  与足够密的网格使得总误差不超过  $\varepsilon$ ，即得证。

\* 事实上边界条件离散方式也会影响数值误差，将在之后讨论。



### §3.4 边界与初值条件离散

为研究导数边界条件的离散, 我们考虑混合边值问题 (HX):

$$u(x, 0) = u_0(x), x \in [0, 1], \quad -au_x(0, t) + \sigma u(0, t) = \phi_0(t), u(1, t) = \phi_1(t), t \in (0, T]$$

这里参数  $\sigma > 0$ , 保证问题适定性。

#### 单侧离散

由于其他离散  $u_j^0 = u_0(x_j), u_j^n = \phi_1(t^n)$  是自然的, 我们只考虑  $\phi_0(t)$  的离散。采用单侧差商离散边界导数可以得到

$$-a \frac{u_1^n - u_0^n}{\Delta x} + \sigma u_0^n = \phi_0(t^n)$$

每次迭代时联立边界条件与内部网格点的差分方程即可得到完整过程。

例如, 记  $\tilde{\sigma} = \frac{a}{a + \sigma \Delta x}$ , 考虑

$$\Phi^n = \left( \frac{\tilde{\sigma}}{\Delta x} \phi_0(t^n), 0, \dots, 0, \frac{a}{(\Delta x)^2} \phi_1(t^n) \right)^T$$

则 HX 的全显格式可写为

$$u^{n+1} = \mathbb{A}_e u^n + \Delta t \Phi^n$$

这里  $\mathbb{A}_e$  左上角为  $1 - (2 - \tilde{\sigma})\mu a$ ,  $a_{12} = a_{21} = \mu a$ , 右下  $n - 1$  阶子方阵为  $\text{tridiag}(\mu a, 1 - 2\mu a, \mu a)$ , 其余为 0。

HX 的全隐格式可写为

$$\mathbb{A}_i u^{n+1} = u^n + \Delta t \Phi^{n+1}$$

这里  $\mathbb{A}_e$  左上角为  $1 + (2 - \tilde{\sigma})\mu a$ ,  $a_{12} = a_{21} = -\mu a$ , 右下  $n - 1$  阶子方阵为  $\text{tridiag}(\mu a, 1 - 2\mu a, \mu a)$ , 其余为 0。

\* 注意上方  $\Phi$  的上标, 显式格式中数值边界条件一般滞后于时间层推进。

**相容性分析:** 注意  $j = 1$  时离散既与边界条件离散有关又与偏微分方程离散有关, 这时需要重新分析相容性。与之前类似, 从局部相容到整体相容一致有界性要求一般仍满足, 因此只考虑**局部相容性**。

\* 例: 对上述全显格式而言,  $x_1$  处的误差为

$$\tau_1^n = \frac{[u]_1^{n+1} - [u]_1^n}{\Delta t} - a \frac{[u]_2^n - 2[u]_1^n}{(\Delta x)^2} - \frac{\tilde{\sigma}}{\Delta x} \left( \frac{a[u]_1^n}{\Delta x} + \phi_0(t^n) \right)$$

此误差来自差分方程与边界条件误差组合, 而边界条件离散的误差  $\tau_b = O(\Delta x)$ , 因此第二部分误差为  $\frac{\tilde{\sigma}}{\Delta x} O(\Delta x) = O(1)$ , 整体误差即为  $O(1)$ , **并不相容**。

\* 由此我们需要改善边界的相容性, 可以考虑扩大离散模板宽度以建立更高阶相容的单侧离散, 或如下节使用双侧离散。

\* 直接估计矩阵每行元素和与谱半径可以得到**古典格式稳定性结论**与**不考虑边界条件离散时完全相同**。

#### 双侧离散

\* 以全显格式为例说明使用方式

**虚拟点方法:** 考虑  $x_0$  处, 加入虚拟点  $x_{-1}$ , 则

$$-a \frac{u_1^n - u_{-1}^n}{2\Delta x} + \sigma u_0^n = \phi_0(t^n)$$

此具有二阶空间相容性。结合差分方程

$$u_0^{n+1} = u_0^n + \mu a \delta_x^2 u_0^n$$

两式结合可消去  $u_{-1}^n$  得到

$$u_0^{n+1} = (1 - 2\mu(a + \sigma\Delta x))u_0^n + 2\mu au_1^n + 2\mu\Delta x\phi_0(t^n)$$

\* 非边界位置的差分方程直接使用全显格式即可。

\* 根据  $u_0^n$  处的迭代, 差分方程必须满足  $2\mu(a + \sigma\Delta x) \leq 1$  才能保证最大模稳定, 实践中一般使  $2\mu a < 1$  且  $\Delta x$  充分小。

**半网格方法:** 考虑所有  $x_{j-1/2}, j = 0, \dots, J+1$  点, 这里空间步长  $\Delta x = \frac{2}{2J+1}$ , 保证  $j = 0$  到  $J+1$  可以经过  $x = 1$  处, 本质边界条件为  $u_{j-1/2}^n = \phi_1(t^n)$ 。对  $x = 0$  处, 用半步中心插上进行离散, 并用算术平均离散点值, 可得二阶空间相容的差分方程

$$-\frac{a}{\Delta x}(u_{1/2}^n - u_{-1/2}^n) + \frac{\sigma}{2}(u_{1/2}^n + u_{-1/2}^n) = \phi_0(t^n)$$

这里  $x_{-1/2}$  事实上是虚拟点, 在全显离散差分方程

$$u_{j+1/2}^{n+1} = u_{j+1/2}^n + \mu a \delta_x^2 u_{j+1/2}^n, \quad j = 0 : (J-1)$$

中可以消去。

\* 其具有与虚拟点方法相同的空间误差阶, 而回到原方程, 空间误差从  $O((\Delta x)^2)$  提升至  $O(\Delta x)$  [理由与单侧离散时的分析相同, 乘  $\frac{\tilde{\sigma}}{\Delta x}$  项], 因此最终局部截断误差  $O(\Delta x + \Delta t)$ 。

\* 直接写出  $u_{1/2}^n$  消去虚拟点的表达式可分析得到  $2\mu a \leq 1$  足以保证半网格方法的最大模稳定, 这比虚拟点方法略好。

\* 虽然自然边界条件会导致局部出现相容阶的损失, 但事实上, 无论采取何种数值边界条件, 相应的 **CN 格式** 均可达到整体二阶的最大模误差, 这比 Lax-Richtmyer 等价定理直接使用得到的结论更强, 需要利用椭圆型差分格式的强最大值原理证明。

### 初值条件的离散

考虑特殊的 Neumann 边界条件定解问题 (HC):

$$u_x(0, t) = u_x(1, t) = g(t), \quad t \in [0, T]$$

这里  $g(t)$  为某已知函数。对  $u_t$  进行积分, 利用  $u_x$  首尾相同可知为 0, 交换积分、求导次序也即

$$\frac{d}{dt} \int_0^1 u(x, t) dx = 0$$

这说明问题 HC 中总热量保持恒定。我们希望数值格式能够保持此总热量守恒性质。

考虑全显格式利用单侧离散方法进行离散, 则计算差分方程可发现

$$\sum_{j=1}^{J-1} u_j^n \Delta x$$

为常数, 而利用半网格方法则可计算发现

$$\sum_{j=1}^J u_{j-1/2}^n \Delta x$$

为常数。这两个守恒量称为**数值守恒量**。

\* 即使  $u_0$  为常值, 单侧离散的数值守恒量也未必等于真实守恒量, 仅有  $O(\Delta x)$  的逼近效果。

\* 半网格方法数值守恒量具有  $O((\Delta x)^2)$  的逼近效果, 可视为矩形积分公式。

更进一步地, 半网格方法中若定义

$$u_{j-1/2}^0 = \frac{1}{\Delta x} \int_{x_{j-1}}^{x_j} u_0(x) dx$$

则可以使数值守恒量等于真实守恒量, 这称为**均值定义方式**, 在真解光滑度差时 [如  $u_0(x) = 1-x^2, g(t) = 0$ ] 可以改善结果。

## 四 一维扩散方程

非均匀介质的热传导一般可以描述为下面两种扩散方程，前者称为非守恒型，后者称为守恒型：

$$u_t = a(x, t)u_{xx}, \quad u_t = (a(x, t)u_x)_x$$

为保证方程抛物性，扩散系数  $a$  在计算区域内有正下确界。

### §4.1 光滑系数线性扩散

\* 本部分假设  $a, u$  都足够光滑。

#### 非守恒型扩散方程

最简单的想法即采用局部冻结技术在离散焦点冻结扩散系数，得到全显、全隐格式

$$\Delta_t u_j^n = \mu a_j^n \delta_x^2 u_j^n, \quad \Delta_t u_j^n = \mu a_j^{n+1} \delta_x^2 u_j^{n+1}$$

直接 Taylor 展开可知它们无条件具有  $(2, 1)$  阶局部截断误差。

加权平均格式对全隐、全显作线性组合，可以考虑多焦点冻结策略

$$\Delta_t u_j^n = \mu(\theta a_j^{n+1} \delta_x^2 u_j^{n+1} + (1 - \theta) a_j^n \delta_x^2 u_j^n)$$

或单焦点冻结策略：

$$\Delta_t u_j^n = \mu a_j^* (\theta \delta_x^2 u_j^{n+1} + (1 - \theta) \delta_x^2 u_j^n)$$

\* 多焦点策略无条件具有  $(2, 1)$  阶局部截断误差， $\theta = \frac{1}{2}$  时称为 CN 格式，无条件具有  $(2, 2)$  阶局部截断误差。

\* 单焦点策略可取  $a_j^* = a(x_j, t^*)$ ， $t^*$  在  $[t^n, t^{n+1}]$  中任取，即可有  $(2, 1)$  阶局部截断误差。当  $\theta = \frac{1}{2}$  时，需要取  $a_j^* = \frac{1}{2}(a_j^n + a_j^{n+1})$  或  $a_j^* = a(x_j, t^{n+1/2})$ ，都称为 CN 格式，无条件具有  $(2, 2)$  阶局部截断误差 [但事实上两者对  $a$  的最低光滑性要求不同]。

最后考察一般扩散方程达到对空间四阶的 Douglas 格式构造，假设  $a(x, t) = a(x)$  与时间无关：

从 CN 格式的构造过程中，唯一不符合四阶的项为  $-\frac{(\Delta x)^2}{12} [u_{xxxx}]_j^{n+1/2}$ ，利用  $u_{xxxx} = (a^{-1}u)_{xxt}$  对其进行二阶相容离散后再加入原 CN 格式的迭代得到

$$\frac{\Delta_t u_{j+1}^n}{12a_{j+1}} + \frac{5\Delta_t u_j^n}{6a_j} + \frac{\Delta_t u_{j-1}^n}{12a_{j-1}} = \frac{1}{2}\mu(\delta_x^2 u_j^{n+1} + \delta_x^2 u_j^n)$$

#### 守恒型扩散方程

最容易的想法即展开空间导数，记  $b = a_x$  离散得到

$$\Delta_t u_j^n = \mu a_j^n \delta_x^2 u_j^n + \frac{\Delta x}{2} \mu b_j^n \Delta_{0x} u_j^n$$

其无条件有  $(2, 1)$  阶局部截断误差，但  $a(x, t)$  剧烈变化时稳定性事实上很差，这是因为其守恒性

$$\int_{z_1}^{z_2} u_t(x, t) dx = W(z_2, t) - W(z_1, t), \quad W = au_x$$

并没有得到保持，这里  $W$  称为热流量。

为保证守恒，一般有两种方法：

##### 1. 积分插值方法

\* 此方法事实上是某种有限体积方法。

将左侧  $u$  写为  $u_t$  积分后交换积分次序可得积分恒等式

$$\int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^{n+1}) dx - \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^n) dx = \int_{t^n}^{t^{n+1}} W(x_{j+1/2}, t) dt - \int_{t^n}^{t^{n+1}} W(x_{j-1/2}, t) dt$$

对左右积分作常值近似可得

$$([u]_j^{n+1} - [u]_j^n) \Delta x \approx ([W]_{j+1/2}^n - [W]_{j-1/2}^n) \Delta t$$

另一方面, 利用一阶中心差分与冻结系数可离散得

$$[W]_{j+1/2}^n \approx a_{j+1/2}^n \frac{[u]_{j+1}^n - [u]_j^n}{\Delta x}$$

这里  $a_{j+1/2}^n$  为扩散系数的局部冻结, 可取  $a(x_{j+1/2}, t^n)$  或  $\frac{1}{2}(a_j^n + a_{j+1}^n)$ , 综合可得守恒型扩散方程的全显格式

$$\Delta_t u_j^n = \mu \delta_x (a_j^n \delta_x u_j^n)$$

\* 一般可忽略两种  $a_{j+1/2}^n$  实现的差别, 统称为**算术平均方式**。

\* 对随时间发展的偏微分方程, **线方法**是常用思路, 即仅离散空间变量, 保持对时间变量的连续性的**半离散格式**。在此例子中, 在某个时刻  $t_0$ , 记  $Q(x) = (a(x, t_0)u_x)_x$ , 则积分近似可得到  $[Q]_j \Delta x \approx [W]_{j+1/2} - W_{j-1/2}$ , 从而与上方完全类似得到离散结果

$$\frac{d}{dt} u_j(t) = \frac{1}{(\Delta x)^2} \delta_x (a_j(t) \delta_x u_j(t))$$

这是一个常微分方程组。进一步数值推进时间变量即可推出**全离散格式**。

\* 并不是所有全离散格式都有对应的半离散格式, 反例如 DF 格式。

## 2. 盒子格式

考虑原方程转化为方程组的降阶处理:

$$v = au_x, \quad v_x = u_t$$

盒子格式 [Keller] 格式考虑时空网格中相邻的四个格点, 在中心点利用中心差商离散第二个方程, 再在水平边中点利用中心差商与系数冻结离散第一个方程, 最终得到

$$\frac{\Pi_x u_{j+1/2}^{n+1} - \Pi_x u_{j+1/2}^n}{\Delta t} = \frac{\Pi_t v_{j+1}^{n+1/2} - \Pi_t v_j^{n+1/2}}{\Delta x}, \quad \Pi_x v_{j+1/2}^n = a_{j+1/2}^n \frac{u_{j+1}^n - u_j^n}{\Delta x}$$

这里  $\Pi_x f_j^n = \frac{1}{2}(f_{j-1/2}^n + f_{j+1/2}^n)$ ,  $\Pi_t f_j^n = \frac{1}{2}(f_j^{n-1/2} + f_j^{n+1/2})$  为算术平均算子,  $a_{j+1/2}^n$  的取法同前。

\* 直接计算可知其具有 (2, 2) 阶局部截断误差。

\* 扩散系数为常数时, 设  $u_j^n = \hat{u}^n e^{ikj\Delta x}$ ,  $v_j^n = \hat{v}^n e^{ikj\Delta x}$  代入计算可知  $\hat{u}^n$  对应增长因子为

$$\lambda(k) = \frac{\cos^2 \sigma - 2\mu a \sin^2 \sigma}{\cos^2 \sigma - 2\mu a \sin^2 \sigma}, \quad \sigma = \frac{k\Delta x}{2}$$

于是其无条件  $L^2$  模稳定。

\* 一般情况下也可说明其无条件  $L^2$  模稳定, 下面给出分析。

## 稳定性分析方法

\* 一般扩散方程中 Lax-Richtmyer 等价定理仍然成立, 因此一般只讨论相容性与稳定性, 相容性的分析与之前是类似的, 下面着重考虑稳定性。

**冻结系数方法：**考虑  $a$  为常数的情况，并对所有稳定范围取交集。例如，非守恒型扩散方程全显格式的最大模稳定条件与  $L^2$  模稳定条件都可以视为

$$\sup_{x,t} a(x,t)\mu \leq \frac{1}{2}$$

此方法并不严格，只是找到了一个必要条件，且可能在  $a$  的频率发生数值共振时线性不稳定。但由其易于分析而常用，一般会将上界进一步减少至 60% 至 80% 以保证稳定。

**能量方法：**更一般的分析方法，利用能量范数进行推导。以周期性边值条件， $a$  与时间无关的守恒型扩散方程为例。

\* 根据偏微分方程知识可知真解满足  $\int_0^1 u^2(x,t)dx \leq \int_0^1 u_0^2(x)$ 。

考虑差分方程  $\Delta_t u_j^n = \mu \delta_x (a_j^n \delta_x u_j^n)$ ，两端同乘  $u_j^{n+1} + u_j^n$  后对  $j = 0 : (J-1)$  累加，利用周期性调整右侧求和次序得到

$$\sum_{j=0}^{J-1} (u_j^{n+1})^2 \Delta x - \sum_{j=0}^{J-1} (u_j^n)^2 \Delta x = -\mu \sum_{j=0}^{J-1} a_{j+1/2} \Delta x u_j^n \Delta x (u_j^{n+1} + u_j^n) \Delta x$$

记 (此  $\mathcal{E}(u^n)$  一般称为**能量范数**，但只有适当的时空条件下才成为离散范数)

$$\mathcal{E}(u^n) = \sum_{j=0}^{J-1} (u_j^n)^2 \Delta x - \frac{1}{2} \mu \sum_{j=0}^{J-1} a_{j+1/2} (\Delta x u_j^n)^2 \Delta x$$

则有

$$\mathcal{E}(u^{n+1}) - \mathcal{E}(u^n) = -\frac{1}{2} \mu \sum_{j=0}^{J-1} a_{j+1/2} (\Delta x (u_j^{n+1} + u_j^n))^2 \Delta x \leq 0$$

记  $\sup_{x \in (0,1)} a(x) = A$ ，利用算术平均不等式可知

$$(1 - 2\mu A) \sum_{j=0}^{J-1} (u_j^n)^2 \Delta x \leq \mathcal{E}(u^n)$$

于是若存在正常数  $\delta > 0$  使得  $1 - 2\mu A \geq \delta$ ，即有  $L^2$  模稳定性

$$\sum_{j=0}^{J-1} (u_j^n)^2 \Delta x \leq \frac{1}{\delta} \sum_{j=0}^{J-1} u_0^2(x_j) \Delta x$$

\* 由上方推导  $\delta < 1$ ，因此此结论弱于真解满足的结论。

\* 对最大模稳定性，要使系数有显式凸组合结构，上述积分插值得到的格式只需满足

$$2a(x_j, t^n) \Delta t \leq (\Delta x)^2, \forall j, n$$

而对最基础的格式  $\Delta_t u_j^n = \mu a_j^n \delta_x^2 u_j^n + \frac{\Delta x}{2} \mu b_j^n \Delta_{0x} u_j^n$ ，还要求

$$|a_x(x_j, t^n)| \Delta x \leq 2a(x_j, t^n), \forall j, n$$

这也就意味着它在扩散系数空间变化剧烈时效果更差。扩散系数间断时，基础格式的问题就更加严重了。

## §4.2 间断系数线性扩散

\* 现实情况中，考虑两种材料焊接在一起时的扩散，就会出现扩散系数间断。

假设守恒型扩散方程中

$$a(x, t) = \begin{cases} a_L & x < x_* \\ a_R & x > x_* \end{cases}$$

期中  $a_L \neq a_R$ , 真解应满足  $x_*$  左右都光滑, 连接处  $u$  连续且  $a_R u_x(x_*^+, t) = a_L u_x(x_*^-, t)$  [此时解事实上为弱解, 第二个条件即为热流通量  $W$  连续]。

由连续性要求, 热量局部守恒性质仍满足。但由于  $u_x$  左右不再连续, 数值计算难度大幅增加。可以说明积分插值方法可以保持局部守恒性质, 从而收敛到真解, 下面研究此方法的误差与系数冻结的影响。

若直接采取算术平均方式冻结  $a_{j+1/2}^n$ , 收敛效果并不好, 因此需要考察更佳的局部冻结方式。利用积分中值定理有

$$[u]_{j+1}^n - [u]_j^n = \int_{x_j}^{x_{j+1}} \frac{[W](x, t^n)}{a(x, t^n)} dx \approx [W]_{j+1/2}^n \int_{x_j}^{x_{j+1}} \frac{dx}{a(x, t^n)}$$

由此  $a$  可以冻结为

$$a_{j+1/2}^n = \Delta x \left( \int_{x_j}^{x_{j+1}} \frac{dx}{a(x, t^n)} \right)^{-1}$$

此积分可直接用数值积分近似, 也可利用两端调和平均近似, 即

$$a_{j+1/2}^n = \left( \frac{\theta_{j+1/2}^n}{a_j^n} + \frac{1 - \theta_{j+1/2}^n}{a_{j+1}^n} \right), \quad \theta_{j+1/2}^n = \begin{cases} \frac{x_* - x_j}{\Delta x} & x_* \in [x_j, x_{j+1}] \\ \frac{1}{2} & x_* \notin [x_j, x_{j+1}] \end{cases}$$

这样的冻结方法统称为调和平均方式。

\* 当扩散系数二阶导数连续有界时, 调和平均与算术平均的差在  $O((\Delta x)^2)$  量级, 但存在间断点时, 调和平均更加准确保持了热流通量在两侧的连续性, 因此效果更好。

\* 事实上, 第一类间断点附近采用算术平均方式局部截断误差阶为  $O((\Delta x)^{-1})$ , 会趋于无穷, 而调和平均方式为  $O(1)$ 。由于两侧 Taylor 展开不同, 分析主要依赖连接处的条件。

### §4.3 更复杂的扩散

#### 极坐标下的扩散

设  $\alpha$  为 1 或 2, 考虑  $\alpha + 1$  维方程

$$u_t = \Delta u, \quad x \in \mathbb{R}^{\alpha+1}, t > 0$$

若真解具有中心对称性  $u(\vec{x}, t) = u(r, t)$ , 这里  $r = \|\vec{x}\|$ , 则可以利用极坐标或球坐标变换得到

$$u_t = r^{-\alpha} (r^\alpha u_r)_r, \quad r \geq 0, t > 0$$

记  $u_j^t$  代表  $r = j\Delta r, t = n\Delta t$  处的值, 且  $j \geq 0$ 。当  $j \geq 1$  时, 可直接利用积分插值方法离散得到

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = \frac{1}{r_j^\alpha (\Delta r)^2} \delta_r (r_j^\alpha \delta_r u_j^n)$$

但  $j = 0$  时出现奇性, 需要单独处理。若  $u$  充分光滑, 由中心对称性必有  $u_r(0, t) = 0$ , 考虑  $u$  在 0 处的 Taylor 展开可以得到方程离散为

$$u_t(0, t^n) = (\alpha + 1) u_{rr}(0, t^n) \approx \frac{2(\alpha + 1)}{(\Delta r)^2} ([u]_1^n - [u]_0^n)$$

再将左侧改写为  $\frac{u_0^{n+1} - u_0^n}{\Delta t}$  即可得到 0 处的迭代公式。

\* 记  $R = \frac{\Delta t}{(\Delta r)^2}$ , 则根据差分方程形式可知  $(\alpha + 1)R \leq \frac{1}{2}$  时系数具有凸组合形式, 可保证最大模稳定性。直接从高维热传导方程出发构造数值格式: 记半径  $r_{j+1/2}$  的球面为  $S_{j+1/2}$ , 其测度为  $r_{j+1/2}^\alpha \pi(1)$ , 这里  $\pi(1)$  代表对应维度单位球面测度。在  $S_{j-1/2}$  与  $S_{j+1/2}$  间的部分  $V_j$  进行类似积分插值的考虑可得

$$\int_{V_j} u_t(x, t) dx = u_r(r_{j+1/2}, t) r_{j+1/2}^\alpha \pi(1) - u_r(r_{j-1/2}, t) r_{j-1/2}^\alpha \pi(1)$$

利用中点近似积分得到

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = \frac{\alpha + 1}{r_{j+1/2}^{\alpha+1} - r_{j-1/2}^{\alpha+1}} \frac{1}{\Delta r} \delta_r(r_j^\alpha \delta_r u_j^n)$$

\* 其在  $\alpha = 1$  时与看作一维后积分插值的结果相同，而  $\alpha = 2$  时不同，更好保持热量局部守恒性。

### 非线性扩散方程

\* 之前的离散方式一般仍然有效，但计算效率和理论分析面临挑战。

考虑方程  $u_t = b(u)u_{xx}$  的纯初值问题或周期边值问题，可直接构造

$$u_j^{n+1} = u_j^n + \mu b(u_j^n) \delta_x^2 u_j^n, \quad u_j^{n+1} = u_j^n + \mu b(u_j^{n+1}) \delta_x^2 u_j^{n+1}$$

的全显、全隐格式，或将两者增量平均得到 CN 格式。

\* 仍通过 Taylor 展开可知古典格式 (2, 1) 阶局部截断误差，CN 格式 (2, 2) 阶局部截断误差。

**Strang 定理：**若非线性差分格式相容于某个适定的非线性问题，则稳定性与收敛性等价。

\* 通过参数冻结方法仍可分析稳定性，如全显格式  $L^2$  模稳定性条件可取为  $\mu \max_{j,n} b(u_j^n) \leq \frac{1}{2}$ ，而 CN 格式无条件  $L^2$  模稳定。

注意到全隐格式与 CN 格式可能涉及非线性方程组，需要通过**局部线性化**技术避免求解，以 CN 格式为例。

**时间延迟技术：**将  $b(u_j^{n+1})$  替换为  $b(u_j^n)$ ，得到  $u_j^{n+1} = u_j^n + \frac{1}{2} \mu b(u_j^n) \delta_x^2 (u_j^n + u_j^{n+1})$ ，其具有 (2, 1) 阶局部截断误差。

**预估校正方法：**每次迭代中先对  $u_j^{n+1}$  进行预估求解

$$\tilde{u}_j^{n+1} = u_j^n + \frac{1}{2} \mu (b(u_j^n) \delta_x^2 u_j^n + b(u_j^n) \delta_x^2 \tilde{u}_j^{n+1})$$

再进行校正得到最终值

$$u_j^{n+1} = u_j^n + \frac{1}{2} \mu (b(u_j^n) \delta_x^2 u_j^n + b(\tilde{u}_j^{n+1}) \delta_x^2 u_j^{n+1})$$

**Richtmyer 方法：**利用时间方向的 Taylor 公式逼近非线性部分，以  $u_t = (u^m)_{xx}, m > 1$  为例，直接使用 CN 格式得到

$$u_j^{n+1} = u_j^n + \frac{1}{2} \mu \delta_x^2 (u_j^n)^m + \frac{1}{2} \mu \delta_x^2 (u_j^{n+1})^m$$

利用近似  $(u_j^{n+1})^m \approx (u_j^n)^m + m(u_j^n)^{m-1}$ ，记  $w_j^n = u_j^{n+1} - u_j^n$ ，可得线性化格式

$$w_j^n = \mu \delta_x^2 (u_j^n)^m + \frac{1}{2} m \mu \delta_x^2 ((u_j^n)^{m-1} w_j^n)$$

求解  $w_j$  即可。

\* 有时多层格式对非线性扩散方程有优势，如  $u_j^{n+1} = u_j^n + \frac{1}{2} \mu b(u_j^n) \delta_x^2 (u_j^n + u_j^{n+1})$  中将  $b(u_j^n)$  替换为  $b(u_j^{n+1/2})$ ，此处  $u_j^{n+1/2} = \frac{3}{2} u_j^n - \frac{1}{2} u_j^{n-1}$ ，此线性化格式具有 (2, 2) 阶局部截断误差且无条件  $L^2$  模稳定。

## 五 高维扩散方程

本节考虑二维常系数方程

$$u_t = a u_{xx} + b u_{yy}$$

这里  $a > 0, b > 0$ 。

## §5.1 方程与边界条件的离散

### 方程的离散

仍考虑等距时空网格  $\mathcal{T}_{\Delta x, \Delta y, \Delta t}$ , 两方向网比记作  $\mu_x = \frac{\Delta t}{(\Delta x)^2}, \mu_y = \frac{\Delta t}{(\Delta y)^2}$ ,  $f_{jk}^n$  表示  $f(x_j, y_k, t^n)$ 。微分方程的离散可以类似一维时构造, 例如双层加权平均格式

$$u_{jk}^{n+1} = u_{jk}^n + \theta(\mu_x a \delta_x^2 u_{jk}^{n+1} + \mu_y b \delta_y^2 u_{jk}^{n+1}) + (1 - \theta)(\mu_x a \delta_x^2 u_{jk}^n + \mu_y b \delta_y^2 u_{jk}^n)$$

与三层 **Du Fort-Frankel** 格式:

$$u_{jk}^{n+1} = u_{jk}^{n-1} + 2\mu_x a(u_{j-1,k}^n - u_{jk}^{n-1} - u_{jk}^{n+1} + u_{j+1,k}^n) + 2\mu_y b(u_{j,k-1}^n - u_{jk}^{n-1} - u_{jk}^{n+1} + u_{j,k+1}^n)$$

由于 Lax-Richtmyer 等价定理成立, 仍只需考虑相容性与稳定性。这里的稳定性依赖于二维的离散范数:

$$\|u^n\|_\infty = \max_{j,k} |u_{jk}^n|, \quad \|u^n\|_2 = \sqrt{\sum_{j,k} |u_{jk}^n|^2 \Delta x \Delta y}$$

\* 以加权平均格式为例, 直接 Taylor 展开可得误差阶数为

$$\tau_{jk}^n = O((\Delta x)^2 + (\Delta y)^2 + (2\theta - 1)\Delta t + (\Delta t)^2)$$

当  $\theta \neq \frac{1}{2}$  无条件具有 (2, 2, 1) 阶局部截断误差, 而 CN 格式具有 (2, 2, 2) 阶误差。对最大模稳定性而言, 由于满足系数凸组合形式即可, 一个充分条件是

$$\mu_x a + \mu_y b \leq \frac{1}{2(1 - \theta)}$$

对  $L^2$  模稳定性, 仍可采用 Fourier 方法, 代入  $u_{jk}^n = \lambda^n e^{i(jl_1 \Delta x + kl_2 \Delta y)}$  可得

$$\lambda(l_1, l_2) = \frac{1 - 4(1 - \theta)s}{1 + 4\theta s}, \quad s = \mu_x a \sin^2 \frac{l_1 \Delta x}{2} + \mu_y b \sin^2 \frac{l_2 \Delta y}{2}$$

类似考虑 von Neumann 条件可知稳定性等价于

$$(1 - 2\theta)(\mu_x a + \mu_y b) \leq \frac{1}{2}$$

类似一维时可区分偏显格式与偏隐格式。

\* 高维扩散方程有一异于一维情形的技术难点, 如混合导数  $u_{xy}$  的离散可能导致原本最大模原理成立的方程离散后不再成立。

\* 注意到二维中的稳定性要求式常关于  $\mu_x a + \mu_y b$ , 若  $a = b, \mu_x = \mu_y$ , 可发现时间推进速度与维数反比, 事实上维数危机还有更严重的形式, 将在之后讨论。

下面讨论**边界条件的离散**: 为了考虑边界条件的处理, 本部分考虑二维有界区域  $\Omega$  中的方程  $u_t = \Delta u = u_{xx} + u_{yy}$  初边值问题 (HIBVP)。

### 矩形区域

设  $\Omega = (0, 1) \times (0, 1)$ , 且边界条件为

$$u(0, y, t) = u(x, 0, t) = u_y(x, 1, t) = u_x(1, y, t) + u(1, y, t) = 0$$

假设网格  $\Delta x = \Delta y$ , 记为  $h$ , 令  $J = \frac{1}{h}, \mu = \frac{\Delta t}{h^2}$ 。若采用全隐格式, 分类考虑离散:

1. 对内部网格, 即  $j, k$  均非 0 或  $J$  时, 直接使用基本格式

$$(1 + 4\mu)u_{jk}^{n+1} - \mu(u_{j-1,k}^{n+1} + u_{j+1,k}^{n+1} + u_{j,k-1}^{n+1} + u_{j,k+1}^{n+1}) = u_{jk}^n$$



2. 对  $x = 0$  或  $y = 0$  的情况, 取  $u_{0k}^n = u_{j1}^n = 0$  即可。

3. 对  $y = 1, x \neq 1$  的情况, 考虑虚拟点方法, 边界条件离散为  $u_{j,J+1}^{n+1} = u_{j,J-1}^{n+1}$ , 代入基本格式可得迭代

$$(1 + 4\mu)u_{jJ}^{n+1} - \mu(u_{j-1,J}^{n+1} + u_{j+1,J}^{n+1} + 2u_{j,J-1}^{n+1}) = u_{jJ}^n$$

4. 对  $x = 1, y \neq 1$  的情况, 同样使用虚拟点, 边界条件离散为  $u_{J+1,k}^{n+1} = u_{J-1,k}^{n+1} - 2hu_{Jk}^{n+1}$ , 代入基本格式得

$$(1 + 4\mu + 2\mu h)u_{Jk}^{n+1} - \mu(2u_{J-1,k}^{n+1} + u_{J,k-1}^{n+1} + u_{J,k+1}^{n+1}) = u_{Jk}^n$$

5. 对  $x = 1, y = 1$  的点  $u_{JJ}$ , 分别对两个自然边界条件采用虚拟点离散, 代入即得

$$(1 + 4\mu + 2\mu h)u_{JJ}^{n+1} - \mu(2u_{J-1,J}^{n+1} + 2u_{J,J-1}^{n+1}) = u_{JJ}^n$$

\* 所有涉及自然边界条件的迭代式两端乘  $\frac{1}{2}$ , 即能将迭代写为对称的系数矩阵  $\mathbf{A}$  的方程组  $\mathbf{A}\mathbf{u}^{n+1} = \mathbf{F}^n$  的形式, 这里  $\mathbf{u}^{n+1}$  为  $u^{n+1}$  展平形成的向量, 具体  $\mathbf{A}, \mathbf{F}$  的形式较为复杂。

### 一般区域

\* 空间情况复杂时相对完美的网格是难以构造的, 且是较为独立的研究方向。这里只考虑正方形网格逼近策略。

考虑某正方形网格  $\mathcal{M}_h$ , 若  $\mathcal{M}_h$  的网格点落在  $\Omega$  内, 称为**网格内点**, 集合为  $\Omega_h$ , 若内点周围四个位置还是内点, 则称为**规则内点**, 否则称为**非规则内点**。

将  $\mathcal{M}_h$  的网格线与  $\partial\Omega$  的交点称为**网格边界点** [未必为格点], 集合为  $\Gamma_h$ , 离散时只需考虑  $\Omega_h \cup \Gamma_h$ 。

考虑网格使得点  $W$  为非规则内点, 上方的  $A$  为边界点, 也恰为网格点, 左侧的  $G$ 、下方的  $T$ 、右侧的  $C$  均为内点; 此外,  $C$  上方的  $P$  与右侧的  $Q$  为非网格点的边界点,  $P$  上方的网格点为  $N$ ,  $Q$  右侧的网格点为  $E$ ;  $C$  下方的  $S$  为内点,  $S$  右侧的非网格点的边界点为  $B$ 。以此网格为例介绍边界的离散技术。

**本质边界条件处理:** 对本质边界条件,  $APQB$  的值都已知 (记对应点的值为  $g(A)$  等, 不妨设与时间有关, 相关时类似), 由于  $A$  恰好为网格点, 对  $W$  直接有

$$-\mu(u_G^{n+1} + u_C^{n+1}) + (1 + 4\mu)u_W^{n+1} = u_W^n + \mu g(A)$$

对  $C$  处, 有两种思路, 通过**插值**可考虑直接将最近的边界点信息。设

$$|CQ| = s_1 h, \quad |CP| = s_2 h, \quad |CW| = s_3 h, \quad |CS| = s_4 h$$

不妨设  $s_1 < s_2$ , 可直接将  $g(Q)$  作为  $C$  的值, 或进行插值

$$u_C^{n+1} = \frac{s_1}{s_1 + s_3} u_W^{n+1} + \frac{s_3}{s_1 + s_3} g(Q)$$

\* 直接将  $g(Q)$  作为  $C$  值的局部截断误差  $O(h)$ , 插值则为  $O(h^2)$ 。

也可通过 Taylor 展开推导**非等臂长差分方程**得到

$$u_C^{n+1} - u_C^n = \mu \sum_{i \in \{Q, P, W, S\}} \beta_i (u_i^{n+1} - u_C^{n+1})$$

$$\beta_Q = \frac{2}{s_1(s_1 + s_3)}, \quad \beta_P = \frac{2}{s_2(s_2 + s_4)}, \quad \beta_W = \frac{2}{s_3(s_1 + s_3)}, \quad \beta_S = \frac{2}{s_4(s_2 + s_4)}$$

\* 此方法局部截断误差为  $(1, 1, 1)$  阶, 且会导致差分方程中的系数矩阵不对称。若希望对称, 可以强行修正四个差分系数为  $\frac{1}{s_1 s_3}, \frac{1}{s_2 s_4}, \frac{1}{s_3^2}, \frac{1}{s_4^2}$ , 此式空间局部截断误差为  $O(1)$ , 不相容。不过, 事实上, 利用椭圆型差分格式强最大值原理可以说明这两种方法均**不破坏整体二阶精度**。

**自然边界条件处理：**考虑 Neumann 边界条件，已知边界处每点外法向的方向导数值，仍记为  $g(A)$  等。由于 A 恰好为网格点，仍可写出 W 附近的基本全隐格式，而设 A 处外法向量为  $(\gamma_1, \gamma_2)$ ，由单侧差商可知（设 A 左侧为网格内点 D）

$$\gamma_1(u_A^{n+1} - u_D^{n+1}) + \gamma_2(u_A^{n+1} - u_W^{n+1}) = hg(A)$$

联立即可消去  $u_A$  得到

$$-\mu(u_G^{n+1} + u_C^{n+1} + u_T^{n+1} - \gamma_1 u_D^{n+1}) + \left(1 + 4\mu - \frac{\mu\gamma_2}{\gamma_1 + \gamma_2}\right)u_W^{n+1} = u_W^n + \frac{\mu hg(A)}{\gamma_1 + \gamma_2}$$

对 C，亦有直接插值方法：设 C 对弧 PQ 的垂线（或近似为对线段 PQ 的垂线）垂足为 U，CU 反向延长线同内部网格线的交点为 V，则可通过插值逼近近似

$$g(U) = \frac{u_C^{n+1} - u_V^{n+1}}{|VC|}$$

或使用积分插值的思路，步骤为：

1. 确定 C 的控制区域  $\Omega_C$ ，一般可取为曲边三角形 HKG，其中 K 是正方形 WCST 的中心，向上竖直延长交边界为 H，向右水平延长交边界为 G；
2. 考虑热传导方程在控制区域的积分，利用散度定理可得（这里  $u_\gamma$  代表随曲线法向的导数）

$$\int_{\Omega_C} [u_t]^{n+1} dx dy = \oint_{\partial\Omega_C} [u_\gamma]^{n+1} ds$$

3. 记  $S_C$  为  $\Omega_C$  面积，差商离散导数得到数值公式

$$S_C \frac{u_C^{n+1} - u_C^n}{\Delta t} = -\frac{u_C^{n+1} - u_W^{n+1}}{\Delta x} |HK| - \frac{u_C^{n+1} - u_S^{n+1}}{\Delta y} |KG| + \int_{HG} g(x, y) ds$$

\* 可采用数值积分近似右侧的积分

\* 高维扩散方程的计算效率存在不足：为使非等臂长的差分方程满足凸组合结构，计算可发现需

$$1 - \mu(\beta_Q + \beta_P + \beta_W + \beta_S) \geq 0$$

当空间网格质量很差时， $\beta_C$  会极大，导致时间步长必须很小。于是，保持一族网格线与边界曲线平行的贴体网格成为重要的技术。

\* 此外，由于隐式格式线性方程组规模较大且不再三对角，计算复杂度会升高很多，极大抵消了时间步长的优势。对此一般有两类解决方法，利用共轭梯度 [CG]、超松弛迭代 [SOR] 等方法改进数值代数求解过程，或利用分数步长、多网格、区域分解等方法改进离散。第二类的几种解决办法将在后文叙述。

## §5.2 分数步长方法

\* 先介绍三种二维交替方向隐式方法 [ADI 方法]

**Peaceman-Rachford 格式 [PR 格式]：**考虑二维扩散方程的基本 CN 格式

$$u^{n+1} = u^n + \frac{1}{2}\mu_x a \delta_x^2 (u^n + u^{n+1}) + \frac{1}{2}\mu_y b \delta_y^2 (u^n + u^{n+1})$$

添加修正项  $\frac{1}{4}\mu_x \mu_y a b \delta_x^2 \delta_y^2 (u^n + u^{n+1})$ ，保持 (2, 2, 2) 阶相容性，且可进行算子分解

$$\left(\mathbb{I} - \frac{1}{2}\mu_x a \delta_x^2\right) \left(\mathbb{I} - \frac{1}{2}\mu_y b \delta_y^2\right) u^{n+1} = \left(\mathbb{I} + \frac{1}{2}\mu_x a \delta_x^2\right) \left(\mathbb{I} + \frac{1}{2}\mu_y b \delta_y^2\right) u^n$$

引进辅助网格函数  $u^{n+1/2}$ ，求解迭代可以分为两步

$$\left(\mathbb{I} - \frac{1}{2}\mu_x a \delta_x^2\right) u^{n+1/2} = \left(\mathbb{I} + \frac{1}{2}\mu_y b \delta_y^2\right) u^n, \quad \left(\mathbb{I} - \frac{1}{2}\mu_y b \delta_y^2\right) u^{n+1} = \left(\mathbb{I} + \frac{1}{2}\mu_x a \delta_x^2\right) u^{n+1/2}$$

\* 这里  $u^{n+1/2}$  与真解  $[u]^{n+1/2}$  并没有关系，不过可以视为逼近，从而看作两个半步时间的推进。

\* 由于 PR 格式中  $x$  方向导数显式离散， $y$  方向导数隐式离散，其可以称为交替方向隐式方法。

\* 代入  $u_{jk}^m = \hat{u}^m e^{i(l_1 j \Delta x + l_2 k \Delta y)}$  可以得到其无条件  $L^2$  模稳定。因此其相容性与稳定性都不弱于 CN 格式。为估计效率，考虑  $\Omega = (0, 1)^2$ ，以  $x, y$  方向步长均  $h$  进行离散，给定 Dirichlet 边值条件，记  $J = \frac{1}{h}$ ，构造分块三对角矩阵

$$\mathbb{A}_r = \text{tridiag}(-rb\mathbb{I}, (2 + 2rb)\mathbb{I} + ra\mathbb{C}, -rb\mathbb{I}), \quad \mathbb{C} = \text{tridiag}(-1, 2, -1)$$

这里  $\mathbb{C}$  为  $J-1$  阶方阵， $\mathbb{A}_r$  为  $(J-1)^2$  阶方阵。

考虑先行后列、从左到右、从上到下将  $u$  展平为向量  $\mathbf{u}$ ，CN 格式的迭代可以写为

$$\mathbb{A}_\mu \mathbf{u}^{n+1} = \mathbb{A}_{-\mu} \mathbf{u}^n$$

消元求解的时间复杂度是  $O(J^4)$ 。

而对 PR 格式，第一步相当于对每行进行求解 ( $\mathbf{u}_j^n = (u_{1,j}^n, \dots, u_{J-1,j}^n)^T$ )

$$(2\mathbb{I} + \mu a \mathbb{C}) \mathbf{u}_j^{n+1/2} = \mu b (\mathbf{u}_{j-1}^n + \mathbf{u}_{j+1}^n) + (2 - 2\mu b) \mathbf{u}_j^n$$

而第二步相当于对每列求解 ( $\mathbf{v}_j^n = (u_{j,1}^n, \dots, u_{j,J-1}^n)^T$ )

$$(2\mathbb{I} + \mu b \mathbb{C}) \mathbf{v}_j^{n+1} = \mu a (\mathbf{v}_{j-1}^{n+1/2} + \mathbf{v}_{j+1}^{n+1/2}) + (2 - 2\mu a) \mathbf{v}_j^{n+1/2}$$

由于系数矩阵三对角，两步时间推进总共只需  $12J^2$  量级的乘法运算，正比未知数总数。

### Douglas-Rachford 格式 [DR 格式]

在二维全隐格式右端添加修正项  $ab\mu_x\mu_y\delta_x^2\delta_y^2(u^n - u^{n+1})$  得到

$$(\mathbb{I} - \mu_x a \delta_x^2)(\mathbb{I} - \mu_y b \delta_y^2) u^{n+1} = u^n + ab\mu_x\mu_y\delta_x^2\delta_y^2 u^n$$

其可以写为

$$u^{n+1/2} - u^n = \mu_x a \delta_x^2 u^{n+1/2} + \mu_y b \delta_y^2 u^n, \quad u^{n+1} - u^n = \mu_x a \delta_x^2 u^{n+1/2} + \mu_y b \delta_y^2 u^{n+1/2}$$

\* 其也属于预估校正方法，无条件  $L^2$  模稳定，但只有  $(2, 2, 1)$  阶相容。

### Douglas 格式

直接写出其迭代方式

$$u^{n+1/2} - u^n = \mu_x a \delta_x^2 \frac{u^{n+1/2} + u^n}{2} + \mu_y b \delta_y^2 u^n, \quad u^{n+1} - u^n = \mu_x a \delta_x^2 \frac{u^{n+1/2} + u^n}{2} + \mu_y b \delta_y^2 \frac{u^n + u^{n+1}}{2}$$

\* 其思路来自预估校正方法，合并计算可知其与 PR 格式的迭代完全等价，也有  $(2, 2, 2)$  阶相容，可以视为对 DR 格式的时间相容改进。不过，三维情况下其与 PR 格式并不等价：三维时 PR 格式为

$$\begin{aligned} u^{n+1/3} - u^n &= \frac{1}{3}\mu_x a \delta_x^2 u^{n+1/3} + \frac{1}{3}\mu_y b \delta_y^2 u^n + \frac{1}{3}\mu_z c \delta_z^2 u^n \\ u^{n+2/3} - u^{n+1/3} &= \frac{1}{3}\mu_x a \delta_x^2 u^{n+1/3} + \frac{1}{3}\mu_y b \delta_y^2 u^{n+2/3} + \frac{1}{3}\mu_z c \delta_z^2 u^{n+1/3} \\ u^{n+1} - u^{n+2/3} &= \frac{1}{3}\mu_x a \delta_x^2 u^{n+2/3} + \frac{1}{3}\mu_y b \delta_y^2 u^{n+2/3} + \frac{1}{3}\mu_z c \delta_z^2 u^{n+1} \end{aligned}$$

记  $\tilde{u}^\kappa = \frac{1}{2}(u^\kappa + u^n)$ , 则 Douglas 格式为

$$\begin{aligned} u^{n+1/3} - u^n &= \mu_x a \delta_x^2 \tilde{u}^{n+1/3} + \mu_y b \delta_y^2 u^n + \mu_z c \delta_z^2 u^n \\ u^{n+2/3} - u^n &= \mu_x a \delta_x^2 \tilde{u}^{n+1/3} + \mu_y b \delta_y^2 \tilde{u}^{n+2/3} + \mu_z c \delta_z^2 u^n \\ u^{n+1} - u^n &= \mu_x a \delta_x^2 \tilde{u}^{n+1/3} + \mu_y b \delta_y^2 \tilde{u}^{n+2/3} + \mu_z c \delta_z^2 \tilde{u}^{n+1} \end{aligned}$$

利用 Fourier 方法可说明 Douglas 格式无条件  $L^2$  模稳定, 而 PR 格式有条件  $L^2$  模稳定。

### 局部一维化方法 [LOD 方法]

考虑时间半离散问题, 函数  $u^{\Delta t}$  满足  $t = 0$  时与  $u$  相同, 且迭代规律为

$$\frac{1}{2}u_t^{\Delta t} = \begin{cases} au_{xx}^{\Delta t} & t \in (t^n, t^{n+1/2}] \\ bu_{yy}^{\Delta t} & t \in (t^{n+1/2}, t^{n+1}] \end{cases}$$

若时间步长恒定, 计算可知  $u^{\Delta t}$  在格点处与  $u$  相同, 而时间步长不等时则有  $\Delta t \rightarrow 0$  时  $u^{\Delta t} \rightarrow u$ 。

\* 可以视为高维传导的逐维分解过程, 符合物理规律。

将上方的半离散问题采用 CN 迭代, 可得经典 LOD 格式

$$\left(\mathbb{I} - \frac{1}{2}\mu_x a \delta_x^2\right)u^{n+1/2} = \left(\mathbb{I} + \frac{1}{2}\mu_x a \delta_x^2\right)u^n, \quad \left(\mathbb{I} - \frac{1}{2}\mu_y b \delta_y^2\right)u^{n+1} = \left(\mathbb{I} + \frac{1}{2}\mu_y b \delta_y^2\right)u^{n+1/2}$$

\* 对纯初值与周期边值问题  $\delta_x^2 \delta_y^2 = \delta_y^2 \delta_x^2$ , 因此 LOD 格式与 PR 格式等价, 具有相应的误差。

LOD 方法单步时间推进效率较好, 常用于预估校正格式的预估值计算, 例如 **Yanenko 格式**:

$$\begin{aligned} u^{n+1/2} &= u^n + \frac{1}{2}\mu_x a \delta_x^2 u^{n+1/4}, \quad u^{n+1/2} = u^{n+1/4} + \frac{1}{2}\mu_y b \delta_y^2 u^{n+1/2} \\ u^{n+1} &= u^n + \mu_x a \delta_x^2 u^{n+1/2} + \mu_y b \delta_y^2 u^{n+1/2} \end{aligned}$$

\* 若  $\delta_x^2 \delta_y^2 = \delta_y^2 \delta_x^2$ , Yanenko 格式也与 PR 格式等价。

\* 半离散问题也可以采取其他方式求解, 如全隐格式, 这时无条件  $L^2$  模稳定, 但相容阶为  $(2, 2, 1)$ 。

\* 上述的 ADI 方法与 LOD 方法统称**经济性方法**, ADI 方法的每个差分方程都是相容于整体的, 但 LOD 只有作为整体才相容, 单个并不相容。ADI 格式可以利用辅助变量转化为格式。

### Strang 格式

考虑 L 型区域等  $\delta_x^2$  与  $\delta_y^2$  不可交换的情况, 这时二维 LOD 格式不再等价于二维 PR 格式, 时间方向没有二阶相容阶, 为解决此问题可采用**双重循环策略**, 构造 Strang 格式:

$$\begin{aligned} \left(\mathbb{I} - \frac{1}{2}\mu_x a \delta_x^2\right)u^{n+1/2} &= \left(\mathbb{I} + \frac{1}{2}\mu_x a \delta_x^2\right)u^n, \quad \left(\mathbb{I} - \frac{1}{2}\mu_y b \delta_y^2\right)u^{n+1} = \left(\mathbb{I} + \frac{1}{2}\mu_y b \delta_y^2\right)u^{n+1/2} \\ \left(\mathbb{I} - \frac{1}{2}\mu_y b \delta_y^2\right)u^{n+3/2} &= \left(\mathbb{I} + \frac{1}{2}\mu_y b \delta_y^2\right)u^{n+1}, \quad \left(\mathbb{I} - \frac{1}{2}\mu_x a \delta_x^2\right)u^{n+2} = \left(\mathbb{I} + \frac{1}{2}\mu_x a \delta_x^2\right)u^{n+3/2} \end{aligned}$$

### 过渡时间层的边界条件

考虑单位正方形上的 Dirichlet 边值条件, 边界函数为  $g$ 。

对整数时间层可直接设置  $u_{jk}^n = g_{jk}^n$  这里  $j \in \{0, J\}$  或  $k \in \{0, J\}$ , 但对过渡时间, 不能直接采用  $u_{jk}^{n+1/2} = g_{jk}^{n+1/2}$  的方案, 而应通过差分方程局部求解实现:

1. 对 ADI 方法, 考虑 PR 格式差分方程相减得到竖直边界的条件为

$$u^{n+1/2} = \frac{1}{2}(g^{n+1} + g^n) - \frac{1}{4}\mu_y b \delta_y^2 (g^{n+1} - g^n)$$

2. 对 LOD 方法, 作用算子  $\mathbb{I} - \frac{1}{2}\mu_y b \delta_y^2$  作用到第二个差分方程, 忽略高阶差商得到

$$u^{n+1/2} = (\mathbb{I} - \mu_y b \delta_y^2) g^{n+1}$$

\* 同 ADI 方法相比, LOD 方法需要更多边界条件, 例如还需要水平边界的过渡时间层边界信息, 这时需要适当使用多项式外插近似。

\* 对其他数值格式与其他类型的边界条件, 设置方式是类似的, 而对三维问题或算子分裂个数较多时, 边界设置将更加复杂。

## 六 线性常系数对流方程

本节考虑方程

$$u_t + au_x = 0$$

的纯初值问题或周期边值问题, 这里  $a \neq 0$  为常数, 网比  $\nu = \frac{\Delta t}{\Delta x}$ 。

### §6.1 基本格式与分析

迎风格式

\* 容易构造离散格式  $u_j^{n+1} = u_j^n - \frac{1}{2}\nu a(u_{j+1}^n - u_{j-1}^n)$ , 但代入  $u_j^n = \lambda^n e^{ikj\Delta x}$  可得其无条件  $L^2$  模不稳定, 因此 (2,1) 阶相容性没有实际意义。

考虑单侧的离散方式, 采用向上游方向的离散:

$$u_j^{n+1} = u_j^n - \nu a \Delta_{-x} u_j^n, \quad a > 0$$

$$u_j^{n+1} = u_j^n - \nu a \Delta_x u_j^n, \quad a < 0$$

即称为迎风格式, 仅具有 (1,1) 阶局部截断误差。

\* 利用 von Neuman 条件可知第一式当且仅当  $0 < \nu a \leq 1$  时  $L^2$  模稳定, 第二式当且仅当  $-1 \leq \nu a < 0$  时  $L^2$  模稳定。也即迎风情况下两格式均有条件稳定, 逆风格式无条件不稳定。

\* 直接利用凸组合性质可以得到  $|\nu a| \leq 1$  时迎风格式最大模稳定。

**Lax-Wendroff 格式** [LW 格式]

对时间 Taylor 展开, 将时间导数转化为空间导数后利用中心差商离散可得格式

$$u_j^{n+1} = u_j^n - \frac{1}{2}\nu a(u_{j+1}^n - u_{j-1}^n) + \frac{1}{2}\nu^2 a^2 \delta_x^2 u_j^n$$

其无条件具有 (2,2) 阶局部截断误差。

\* 分析可知当且仅当  $|\nu a| \leq 1$  时 LW 格式  $L^2$  模稳定, 但实验表明其在  $|\nu a| \neq 1$  时不具有最大模稳定性, 不过最大模满足  $\|u^n\|_\infty \leq C n^{1/12} \|u^0\|_\infty$ , 增长速率相对可控。

\* 为了分析双曲差分格式的稳定性, 除了利用之前的思路外, 还有其他的分析方法, 虽然得到的结果不够严谨, 但有充分的指导价值。

**数值黏性方法** [本质为修正方程方法]: 利用绝对值可将迎风格式写为

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = \frac{|a|\Delta x}{2} \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{(\Delta x)^2}$$

同中心差商格式相比, 右侧即为数值黏性产生的修正项,  $\frac{|a|\Delta x}{2}$  称为数值黏性系数, 随网格加密而消失。类似地, LW 格式的数值黏性系数为  $\frac{1}{2}|a|^2 \Delta t$ , 于是它们分别可以看作对  $u_t + au_x = \frac{1}{2}|a|\Delta x u_{xx}$  与  $u_t + au_x = \frac{1}{2}|a|^2 \Delta t u_{xx}$  进行的离散, 这两个方程即称为格式对应的修正方程。

\* 根据微分方程理论可知对流扩散方程  $u_t + au_x = bu_{xx}, b > 0$  在  $b$  越大时能量衰减越快, 适定性表现越好, 因此数值黏性系数越大, 稳定性表现越好。

\* 当  $|\nu a| \leq 1$  时, LW 的数值黏性系数弱于迎风格式, 稳定性表现相对变差, 但数值黏性小的优势在于高相容性。更多细节论证见后文。

**CFL 条件:** 设差分方程与微分方程相容, 若其稳定, 则微分方程的依赖区域在  $\Delta x, \Delta t \rightarrow 0$  时必须包含于差分方程的依赖区域。

\* 可从流动或特征线观点得到。直观理解为, 若微分方程依赖区域不被差分方程依赖区域包含, 则扰动时某点真解变化而数值解不变, 即无法收敛。

\* 考虑偏心格式的数值依赖区域, 如左偏心格式中  $(x_j, t^{n+1})$  前一刻依赖  $[x_j - \Delta x, x_j]$ , 而真解依赖区域边界  $x_j - a\Delta t$ , 由此即可知稳定等价于  $0 \leq a\Delta t \leq \Delta x$ , 这即与迎风格式的 Fourier 方法稳定性条件一致, 对右偏心格式与 LW 格式同理。

\* 此条件仅为必要条件 (例如中心差商格式也满足 CFL 条件), 也并没有明确指出何种范数下稳定, 因此效果是模糊的。

两格式在实验中的数值表现:

1. 真解相对光滑的区域两格式表现均理想, LW 相容阶更高, 因此误差更小;
2. 间断界面附近迎风格式相对完美, LW 格式会出现虚假振荡 [Gibbs 现象] 与上下溢出, 无法通过网格加密改善;
3. 在真解相对光滑区域, LW 格式事实上也无法避免数值振荡, 但可以通过网格加密改善。

综合两者优点, 在光滑区域有高阶相容性与良好稳定性、间断界面能够捕捉位置并控制数值振荡的格式称为高精度高分辨率格式。

### 线性常系数差分格式

对线性常系数对流方程的一般线性常系数差分格式可以写成

$$u_j^{n+1} = \sum_{s=-l}^r \alpha_s u_{j+s}^n$$

直接对局部截断误差 [上式左减右除以  $\Delta t$ ] 进行 Taylor 展开可知相容的充要条件是  $\sum_s \alpha_s = 1, \sum_s s\alpha_s = -a\nu$ 。另一方面, 由于  $u = u_0(x - at)$  为解 (称为行波解), 事实上可代入  $u_j^n = (x_j - at^n)^l$ , 若  $l = 0, 1, \dots, k$  都精确成立, 则局部截断误差至少  $k$  阶。

\* 于是相容的充要条件也可写为对  $1, x - at$  均精确成立。

\* 若  $a < 0$ , 线性差分方程  $L^2$  模稳定, 则其能达到的最高局部截断误差阶为  $p = \min(l + r, 2l + 2, 2r)$ , 且离散模板  $r - l \in \{0, 1, 2\}$ 。对  $a > 0$  时有对称的结论。

基于行波解结构, 若初值单调, 此后将一直保持相同的单调性, 这称为**单调保持性质**。若差分格式不保持单调, 则很可能出现虚假的数值振荡。直接计算、构造验证可发现, 线性常系数差分格式能保持单调性当且仅当  $\alpha_s \geq 0$  对所有  $s$  成立。

\* 此时称为**单调格式或正格式**, CFL 条件成立时迎风格式是单调格式, 但 LW 格式不是。

\* 结合相容的充要条件, 相容的单调格式必然满足系数凸组合结构, 因此**最大模稳定**, 事实上通过 Jensen 不等式可证明其亦  $L^2$  模稳定。

单调格式至多具有一阶局部截断误差 [Godunov 定理]: 计算可知二阶局部截断误差除一阶外还需满足条件  $a^2\nu^2 = \sum_s s^2\alpha_s$ , 但根据 Cauchy 不等式可知此条件与  $\sum_s \alpha_s = 1, \sum_s s\alpha_s = -a\nu$  同时成立当且仅当  $\alpha_s$  只有一个非零, 无实际意义, 因此只能具有一阶局部截断误差。

\* 这意味着高阶相容的线性格式**必定存在负系数**, 数值振荡不可避免。

### 色散分析

定义**波函数**为形如  $u(x, t) = \frac{A}{\sqrt{2\pi}} e^{i\phi(x, t)}$  的函数，实函数  $\phi$  称为**相位函数**，最简单的形式是  $kx + \omega t$ ，此时波函数称为**简谐波**， $A$  为振幅， $k = \phi_x$  为波数， $\omega = \phi_t$  为相位速度， $c = -\frac{\omega}{k}$  称为**波速**，其为正代表从左向右传播，为负则反之。

\* 其他相关定义：波长  $l = \frac{2\pi}{|k|}$ ，频率  $f = \frac{|\omega|}{2\pi}$ ，能量  $E = |A|^2$ 。

将不同波数的简谐波 [指标集合记为  $\mathcal{O}$ ] 叠加，相应的波包可以表示为：

$$u(x, t) = \sum_{k \in \mathcal{O}} \frac{A_k}{\sqrt{2\pi}} e^{i(kx + \omega t)}$$

$A_k$  蕴含振幅与初始相位， $\omega = \omega(k)$  称为**色散关系**，通常约定  $\omega(0) = 0$ 。

\* 分析可发现  $\omega(k)$  为线性函数时各个简谐波波速相同，整体波形不变，而其非线性时整体波形可能变换，称为**色散**。

\* 波包函数的包络线也是一个波函数，若波数都集中在  $k$  附近，其传播速度应为  $C(k) = -\omega'(k)$ ，称为**群速度**，也称能量的传播速度，常规介质中低于所有包含的简谐波的波速。

\* 波函数的峰、谷或间断界面称为**波面**，沿着波传播方向，波面前、后的位置称为**波前**、**波后**。

考虑对流方程中不同波数的单位简谐波  $u(x, 0) = e^{ikx}$  的传播，其真解为  $u(x, t) = e^{i(kx + \omega t)}$ ，这里  $\omega = -ak$  为线性，不存在色散。

代入一般的线性常系数差分格式，可得到数值解

$$u_j^n = e^{i(kj\Delta x + \omega^* n\Delta t)} = \lambda(k)^n e^{ikj\Delta x}$$

这里  $\omega^* = \omega^*(k)$  称为**广义数值色散关系**， $\lambda(k) = e^{i\omega^*(k)\Delta t}$  为增长因子。

按照实部与虚部展开，可以得到

$$\lambda(k) = e^{-\omega_{Im}^*(k)\Delta t} e^{i\omega_{Re}^*(k)\Delta t} = |\lambda(k)| e^{i \arg \lambda(k)}$$

于是  $|\lambda(k)|$  可表示推进  $\Delta t$  后的振幅变化率， $\arg \lambda(k)$  则为相位改变量，两者均代表数值波动信息。

\* 具体来说， $\omega^*(k)$  的虚部为正则简谐波振幅会衰减，称为**差分格式有耗散**，一般对高波数简谐波更强； $\omega^*(k)$  的虚部为负会引起简谐波振幅膨胀，形成**反数值耗散**，严重时破坏  $L^2$  模稳定性；若  $\omega^*(k)$  的虚部为 0，振幅保持不变，称为**无耗散**的差分格式。

\* 增长因子的辐角决定色散性质，记相位速度相对误差为

$$\frac{\omega_{Re}^*(k)}{\omega(k)} - 1 = -\frac{\arg \lambda(k)}{ka\Delta t} - 1$$

若其为正，数值简谐波超前于真实简谐波，否则滞后于真实简谐波。一般无耗散格式的数值色散现象严重，可能产生波形变化 [事实上数值耗散与数值色散为数值误差产生的根本原因，因此这样的分析方法适用任何 PDE 数值差分算法]。

根据数值波动信息，发生数值振荡时可以判断位置：

1. 计算相位速度相对误差，其恒正则振荡出现在波前，否则在波后；
2. 计算数值群速度

$$C_{\Delta x}(k) = -(\omega_{Re}^*)'(k) = -\frac{1}{\nu} \frac{d \arg \lambda(\xi)}{d\xi}, \quad \xi = k\Delta x$$

要保证数值解传播方向正确，数值群速度应与  $a$  同号，若  $|C_{\Delta x}| > |a|$  恒成立，则数值振荡出现在波前，若小于恒成立则在波后。

\* 对 LW 格式利用 Taylor 展开估算相位速度相对误差

$$\arg \lambda(k) = -\nu a \xi \left( 1 - \frac{1}{6}(1 - \nu^2 a^2) \xi^2 + \dots \right)$$

括号内第二项即为相对误差，由此可得时其数值振荡必然出现在波后 [若用数值群速度估算，可得相同结论]。利用色散理论也可以解释其数值现象：右上式可知波数越高时相对误差越大，而 Fourier 理论可知函数越光滑，高频成分越少，因此间断函数的数值振荡严重。对光滑函数，波包函数基本形状没有受到严重破坏，数值振荡不明显，但注意到增长因子满足

$$|\lambda(k)| = 1 - \frac{1}{8} \nu^2 a^2 (1 - \nu^2 a^2) \xi^4 + \dots$$

其数值耗散速度远低于色散速度，可以不妨认为振幅保持不变，于是长时间发展后微弱的高频波终将明显改变整体形状，产生肉眼可见的数值振荡。也即数值色散是数值振荡的根本原因，数值耗散同数值色散的平衡决定数值振荡的具体表现。

\* 对迎风格式，类似分析可得

$$\arg \lambda(k) = -\nu a \xi \left( 1 - \frac{1}{6}(1 - \nu a)(1 - 2\nu a) \xi^2 + \dots \right)$$

$$|\lambda(k)| = 1 - \frac{1}{2} \nu a (1 - \nu a) \xi^2 + \dots$$

相位速度比 LW 更精准，捕捉波面位置能力更强，而数值耗散比 LW 格式更严重，稳定性表现更好。

\* 从数值黏性的角度，数值黏性系数越大，耗散越强，抑制数值振荡的能力就越强，但其过大会导致相容阶受损。

## §6.2 更多格式

**Lax-Friedrichs 格式** [LF 格式、Lax 格式]

由于真解满足  $[u]_j^{n+1} = u(x_j - a\Delta t, t^n)$ ，当  $|\nu a| \leq 1$  时  $x_j - a\Delta t$  在  $x_{j-1}$  与  $x_{j+1}$  之间，用这两点线性插值近似即可得到 Lax 格式 [即 Lax 格式可看作特征线方法与插值的结合]：

$$u_j^{n+1} = \frac{1}{2}(u_{j-1}^n + u_{j+1}^n) - \frac{1}{2} \nu a (u_{j+1}^n - u_{j-1}^n)$$

\* 直接 Taylor 展开可知其有条件相容，在  $\nu$  固定时有一阶局部相容性。

\* 其  $L^2$  模稳定当且仅当  $|\nu a| \leq 1$ ，同 CFL 条件一致，且当 CFL 条件成立时，其为单调格式。

\* 由于无需判断流场方向，其容易推广到变系数与方程组的情况。

将其化为修正方程的形式可发现其数值黏性系数为  $\frac{(\Delta x)^2}{2\Delta t}$ ，CFL 条件成立时比迎风格式的数值黏性系数  $\frac{|a|\Delta x}{2}$  更大，数值耗散更强，间断界面倾向于被抹平。

### 蛙跳格式

直接中心差商离散时空导数得到

$$\frac{u_j^{n+1} - u_j^{n-1}}{2\Delta t} + a \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = 0$$

\* 显式三层格式，无条件具有 (2,2) 阶局部间断误差。利用对三层格式的分析方式，取  $v^n = u^{n-1}$  可知其增长矩阵的两个特征值为

$$\lambda_{\pm}(k) = -i\nu a \sin \xi \pm \sqrt{1 - \nu^2 a^2 \sin^2 \xi}, \quad \xi = k\Delta x$$

于是，其  $L^2$  模稳定等价于  $|\nu a| < 1$  [与之前的格式相比， $|\nu a| = 1$  时其可能不稳定，此稳定要求高于 CFL 条件]。



直接计算可知  $|\lambda_{\pm}(k)| = 1$ ，因此蛙跳格式**无耗散**，而辐角为

$$\arg \lambda_{\pm}(k) = -\nu a \xi \left( \pm 1 \mp \frac{1}{6}(1 - \nu^2 a^2) \xi^2 + \dots \right)$$

括号内的  $\pm$  表明  $\lambda_{-}(k)$  相应的数值简谐波反向传播，这称为**衍生波**现象，而  $\lambda_{+}(k)$  占主导地位，于是根据符号可知数值振荡必然出现在波后。

\* 多层格式普遍存在衍生波，意味着只有一个增长因子成为定解问题真实的增长因子。一般来说，数值耗散可以减弱衍生波的影响。

\* 由于蛙跳格式离散模板的空心十字架结构，其可以拆解成两个互不干扰的独立计算系统（相邻格点分属于不同系统）。若初值不够恰当，有可能产生两不相关的波形，形成**组间振荡**现象。解决方案为，利用初值定义  $u^0$  后利用迎风格式或 LW 格式计算  $u^1$ ，以降低反向波比重。

### 盒子格式

与扩散方程的盒子格式类似，选取四个格点，利用半步中心差商离散导数后均值近似半点值可得到

$$\frac{u_{j+1}^{n+1} - u_{j+1}^n + u_j^{n+1} - u_j^n}{2\Delta t} + a \frac{u_{j+1}^n - u_j^n}{2\Delta x} = 0$$

或等价写为

$$u_{j+1}^{n+1} = u_j^n + \frac{1 - \nu a}{1 + \nu a} (u_{j+1}^n - u_j^n)$$

在  $a > 0$  时，通过后一式从左到右扫描网格点即可显式计算，相应的盒子格式为**半隐**的。

\* 记  $\xi = k\Delta x$ ，计算可发现盒子格式  $|\lambda(k)| = 1$ ，无条件  $L^2$  模稳定，而

$$\arg \lambda(k) = -\nu a \xi \left( 1 + \frac{1}{12}(1 - \nu^2 a^2) \xi^2 + \dots \right)$$

其相位速度的相对误差是蛙跳格式或 LW 格式的一半，当  $|\nu a| < 1$  时出现在波前， $|\nu a| > 1$  时出现在波后。

\* 与蛙跳格式类似，盒子格式也存在**棋盘解**  $u_j^n = (-1)^{j+n}$ ，这意味着其也分为两个相对独立的计算系统。通过初边值条件的合理设置可降低此解的比重。

\* 根据 Taylor 展开的规律与数值群速度定义，数值群速度的 Taylor 展开可直接看作辐角的 Taylor 展开对  $\xi$  求导并乘比例，因此群速度出发的分析结果与辐角出发完全一致。

总结上述的格式：

名称	局部误差阶	数值耗散	数值色散
迎风格式	一阶	强	很弱 (单调格式)
Lax 格式	一阶	强	很弱 (单调格式)
LW 格式	二阶	弱	波后
蛙跳格式	二阶	无	波后，有衍生波
盒子格式	二阶	无	波前，无衍生波

### 人工边界条件

对于双曲方程的初边值问题，边界条件不能随意设置。例如对流方程

$$u_t = u_x, \quad x \in (0, 1), t > 0, \quad u(x, 0) = u_0(x)$$

只能在  $x = 1$  处提供入流边值条件 (以  $u(t, 1) = 0, t > 0$  为例)，而不能在  $x = 0$  处提供出流边值条件。

\* 利用能量方法可知上述问题是适定的，真解  $L^2$  模不增。

下面构造其差分格式，考虑网格  $x_j = j\Delta x, j = 0 : J$ ，其中  $J = \frac{1}{\Delta x}$ 。0 与  $J$  之外的点可以使用之前的任何差分格式， $u_j^n$  直接恒定为 0 即可，但  $u_0^n$  则可能遇到困难。

对单边离散模板的迎风格式与盒子格式，边界处可以直接计算，但对 Lax 格式、LW 格式或蛙跳格式，需要采用其他途径进行人工定义，一般有两种常见的设置方式：

1. 利用内部数值解进行外插多项式逼近, 例如具有一阶局部截断误差的**常值**外插  $u_0^n = u_1^n$  或具有二阶局部截断误差的**线性**外插  $u_0^n = 2u_1^n - u_2^n$ ;
2. 考虑从特征线回溯理论进行内插多项式逼近, 即  $u_0^{n+1} = u_0^n + \nu(u_1^n - u_0^n)$ , 这事实上就是局部的迎风格式。

\* 实验可得到前一种人工边界条件对蛙跳格式不稳定, 而两种人工边界条件对 LW 格式都可行。严谨的理论分析一般难以实现, 目前较成功的方法有分离变量、能量方法与 GKS 理论等。

\* 虚拟网格方法也可用于出流边界条件的设置, 可根据不同格式推导。

## 七 线性双曲型方程

### §7.1 更复杂的一阶方程

#### 变系数对流方程

考虑方程 (设  $a$  为已知的连续函数)

$$u_t + a(x, t)u_x = 0, \quad t > 0$$

\* 利用特征线法可知其特征方程为  $\frac{dx}{dt} = a(x, t)$ , 不同  $x(0)$  出发的特征线互不相交, 且每条特征线上  $u$  保持不变, 故给定  $u_0$  后真解存在唯一, 即使初值是间断函数。

其迎风格式为

$$u_j^{n+1} = \begin{cases} u_j^n - \nu a_j^n (u_j^n - u_{j-1}^n) & a_j^n \geq 0 \\ u_j^n - \nu a_j^n (u_{j+1}^n - u_j^n) & a_j^n < 0 \end{cases}$$

而对应的 **Lax** 格式为

$$u_j^{n+1} = \frac{1}{2}(u_{j-1}^n + u_{j+1}^n) - \frac{1}{2}\nu a_j^n (u_{j+1}^n - u_{j-1}^n)$$

\* 网比固定时两者均一阶相容, 但迎风格式要在所有网格点确定迎风方向, Lax 格式则不必。

记  $b = aa_x - a_t$ , Taylor 展开后将时空导数互相转换至二阶, 得到 **LW** 格式

$$u_j^{n+1} = u_j^n - \frac{\nu}{2} \left( a_j^n - \frac{1}{2} b_j^n \Delta t \right) \Delta_{0x} u_j^n + \frac{1}{2} \nu^2 (a_j^n)^2 \delta_x^2 u_j^n$$

\* 其 LW 格式仍然无条件具有 (2, 2) 阶局部截断误差。

\* 冻结系数方法与 **CFL** 方法仍然可以使用, 均可得到三个格式的稳定性条件为 (对 LW 格式, 需要将  $a, b$  均冻结, 两种方法得到的稳定性条件相同)

$$\max_{x,t} |a(x, t)| \nu \leq 1$$

而利用离散最大模原理可知此时迎风格和 Lax 格式有最大模稳定性; 当  $a(x, t)$  足够光滑时, 利用能量方法可以证明此时两者也具有  $L^2$  模稳定性。

\* 考虑原问题的纯初值或周期边值问题, 其中  $a(x, t) = a(t)$  连续且只与时间有关, 此时可验证真解的  $L^2$  模保持恒定, 但 LW 格式则可能导致离散  $L^2$  模出现指数增长, 完全偏离了解的适定性, 尤其当  $a(t)$  不够光滑或有零点时, 数值格式可能是不稳定的。

\* 此外, 冻结系数方法无法判别线性不稳定现象, 考虑  $a(x, t) = a(x)$ , 在四个网格点  $x_0, x_1, x_2, x_3$  上满足  $a_0 = a_3 = 0, a_1 > 0 > a_2$ , 此时利用蛙跳格式递推有

$$\begin{pmatrix} u_1^{n+1} \\ u_2^{n+1} \end{pmatrix} = \begin{pmatrix} u_1^n \\ u_2^n \end{pmatrix} + \nu \begin{pmatrix} a_1 \\ -a_2 \end{pmatrix} \begin{pmatrix} u_1^n \\ u_2^n \end{pmatrix}$$

利用分离变量方法, 假设  $\sqrt{|a_1 a_2|}$  对应的特征向量为  $\Theta$ , 则有

$$\begin{pmatrix} u_1^n \\ u_2^n \end{pmatrix} = O(e^{n\nu\sqrt{|a_1 a_2|}})\Theta$$

若  $a(x)$  局部连续可导, 确保  $a'(x)$  局部有界, 则指数增长有界, 否则考虑类似  $a(x) = x^{1/3} \sin \frac{1}{x^2}$  在原点附近, 加密时  $\sqrt{|a_1 a_2|} = O((\Delta x)^{1/3})$ , 数值解趋于无穷, 产生明显不稳定。(反之,  $a(x)$  符号不变时蛙跳格式一定线性稳定。)

### 一阶双曲型方程组

设  $\mathbf{u}(x, t)$  是未知的  $m$  维向量值函数, 考虑一节偏微分方程组

$$\mathbf{u}_t = \mathbb{A} \mathbf{u}_x$$

这里  $\mathbb{A}$  是给定的  $m$  阶方程组。若  $\mathbb{A}$  可实相似对角化为  $\mathbb{R} \mathbb{D} \mathbb{R}^{-1}$ , 则称此方程组为双曲型。记对角阵  $\mathbb{D}$  的对角元 (即  $\mathbb{A}$  特征值) 为  $d_1, \dots, d_m$ 。

记  $\mathbb{D}^{\oplus} = \text{diag}\{\max(d_1, 0), \dots, \max(d_m, 0)\}$ ,  $\mathbb{D}^{\ominus} = \text{diag}\{\min(d_1, 0), \dots, \min(d_m, 0)\}$ , 即将特征值分为正负部分, 并对对应记

$$\mathbb{A}^{\oplus} = \mathbb{R} \mathbb{D}^{\oplus} \mathbb{R}^{-1}, \quad \mathbb{A}^{\ominus} = \mathbb{R} \mathbb{D}^{\ominus} \mathbb{R}^{-1}, \quad |\mathbb{A}| = \mathbb{A}^{\oplus} - \mathbb{A}^{\ominus}$$

由于此时记  $\mathbf{v} = \mathbb{R}^{-1} \mathbf{u}$  有

$$\mathbf{v}_t + \mathbb{D} \mathbf{v}_x = 0$$

可直接将迎风格式的离散逐项写出, 合并为

$$\mathbf{v}_j^{n+1} = \mathbf{v}_j^n - \nu \mathbb{D}^{\oplus} (\mathbf{v}_j^n - \mathbf{v}_{j-1}^n) - \nu \mathbb{D}^{\ominus} (\mathbf{v}_{j+1}^n - \mathbf{v}_j^n)$$

于是有

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \nu \mathbb{D}^{\oplus} (\mathbf{u}_j^n - \mathbf{u}_{j-1}^n) - \nu \mathbb{D}^{\ominus} (\mathbf{u}_{j+1}^n - \mathbf{u}_j^n)$$

\* 注意到  $\mathbb{A}^{\oplus}$  特征值均非负,  $\mathbb{A}^{\ominus}$  特征值均非正, 这样的分解事实上隐含着**通量分裂技术**, 使得  $\mathbb{A}^{\oplus} \mathbf{u}$  与  $\mathbb{A}^{\ominus} \mathbf{u}$  的上游方向均确定。

\* 转化为  $\mathbf{v}$  后方程独立, 由于  $\rho(\mathbb{A}) = \max_i |d_i|$ ,  $\mathbf{v}$  的稳定性条件即为  $\rho(\mathbb{A})\nu \leq 1$ 。由于  $\mathbf{v}$  与  $\mathbf{u}$  有线性关系, 二者必然同时稳定/不稳定, 于是此稳定条件即为  $\mathbf{u}$  的  $L^2$  模稳定充要条件。

\* 对线性变系数问题, 特征分解需要在每个网格点执行, 迎风格式就会消耗大量时间。这时无须特征分解的格式更受欢迎, 如 Lax 格式

$$\mathbf{u}_j^{n+1} = \frac{1}{2}(\mathbf{u}_{j-1}^n + \mathbf{u}_{j+1}^n) - \frac{1}{2}\nu \mathbb{A} \Delta_{0,x} \mathbf{u}_j^n$$

或 LW 格式

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{1}{2}\nu \mathbb{A} \Delta_{0,x} \mathbf{u}_j^n + \frac{1}{2}\nu^2 \mathbb{A}^2 \delta_x^2 \mathbf{u}_j^n$$

它们都相当于方程情况的  $a$  换为  $\mathbb{A}$ , 稳定性条件均与迎风格式相同。但 Lax 格式数值耗散过多, LW 格式则容易振荡。

## §7.2 高阶与高维推广

### 二阶声波方程

声音的传播或弦的振动可以简化为一维二阶声波方程

$$u_{tt} = a^2 u_{xx}, \quad x \in \mathbb{R}, t > 0$$

其中  $a > 0$  为波在给定介质的传播速度,  $u$  则代表波。额外给定初值条件

$$u(x, 0) = f(x), \quad u_t(x, 0) = g(x)$$

\* 真解可以用 D'Alembert 公式表示, 即

$$u(x, t) = \frac{1}{2}(f(x - at) + f(x + at)) + \frac{1}{2a} \int_{x-at}^{x+at} g(s) ds$$

由此可见  $[x - at, x + at]$  是初始时刻的真实依赖区域, 与一阶对流方程不同, 其真解双向传播。

**直接离散方式:** 直接使用二阶中心差商离散导数可得三层格式

$$\delta_t^2 u_j^n = \nu^2 a^2 \delta_x^2 u_j^n$$

无条件具有 (2, 2) 阶局部截断误差。作为三层格式, 其数值启动可用

$$u_j^0 = f_j, \quad u_j^1 = f_j + \frac{1}{2} \nu^2 a^2 \delta_x^2 f_j + \Delta t g_j$$

\* 数值启动来源为  $u_j^1 - u_j^{-1} = 2\Delta t g_j$  的虚拟点法, 结合  $t = 0$  时的声波方程消去虚拟点。

之前我们将  $p + 1$  层格式的  $L^2$  模稳定性定义为

$$\exists C, \quad \|u^n\|_2 \leq C \sum_{m=0}^{p-1} \|u^m\|_2, \quad \forall n \leq \frac{T}{\Delta t}$$

直接使用 Fourier 方法分析可得其无条件线性不稳定, 但实验可发现  $\nu a \leq 1$  时中心差商格式数值解收敛到真解, 与 Lax-Richtmyer 等价定理得到的稳定矛盾。由此, 我们需要调整稳定性概念: 对于带有二阶时间导数的发展型偏微分方程  $u_{tt} = P(x, u_x, u_{xx}, \dots)$ , 由于解增添随时间线性增长的部分  $at + b$  仍为解, 应当允许离散  $L^2$  模为  $O(n)$ , 即修正为

$$\exists C, \quad \|u^n\|_2 \leq C(1 + n) \sum_{m=0}^{p-1} \|u^m\|_2, \quad \forall n \leq \frac{T}{\Delta t}$$

此定义下其  $L^2$  模稳定充要条件即为  $\nu a \leq 1$ , 符合 CFL 条件推导的结果。

\* 对初边值问题, 边界条件的离散与热传导方程介绍的方式类似。

**间接离散方式:** 记  $v = u_t, w = -au_x$ , 可将方程改写为一阶双曲型方程组

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} + \begin{pmatrix} & a \\ a & \end{pmatrix} \begin{pmatrix} v_x \\ w_x \end{pmatrix} = 0$$

可直接用方程组构造对应的迎风格式、Lax 格式或 LW 格式。不过, 由于三者均为耗散格式, 无法保持数值声波方程的能量守恒, 下面给出三个无耗散格式:

### 1. 非交错网格盒子格式

直接利用盒子格式离散两方程得到

$$\Pi_x v_{j+1/2}^{n+1} - \Pi_x v_{j+1/2}^n + \nu a (\Pi_t w_{j+1}^{n+1/2} - \Pi_t w_j^{n+1/2}) = 0$$

$$\Pi_x w_{j+1/2}^{n+1} - \Pi_x w_{j+1/2}^n + \nu a (\Pi_t v_{j+1}^{n+1/2} - \Pi_t v_j^{n+1/2}) = 0$$

这里  $\Pi_x, \Pi_t$  仍表示算术平均算子, 初值  $w_j^0 = -af_x(x_j), v_j^0 = g(x_j)$ 。

### 2. 非交错网格蛙跳格式

直接使用蛙跳格式离散两方程得到

$$v_j^{n+1} - v_j^{n-1} + \nu a (w_{j+1}^n - w_{j-1}^n) = 0$$

$$w_j^{n+1} - w_j^{n-1} + \nu a (v_{j+1}^n - v_{j-1}^n) = 0$$

初值  $w^0, v^0$  与盒子格式一致, 由于其为三层格式,  $w^1, v^1$  须通过其他格式计算。

### 3. 交错网格蛙跳格式

此处交错网格指将  $v$  定义在时间半整数、空间整数的网格上，而将  $w$  定义在时间整数、空间半整数的网格上，再用蛙跳格式离散得到

$$v_j^{n+1/2} - v_j^{n-1/2} + \nu a(w_{j+1/2}^n - w_{j-1/2}^n) = 0$$

$$w_{j+1/2}^{n+1} - w_{j+1/2}^n + \nu a(v_{j+1}^{n+1/2} - v_j^{n+1/2}) = 0$$

其事实上是双层格式，于是直接设置  $w_{j+1/2}^0 = -af_x(x_{j+1/2})$ ,  $v_j^{1/2} = g(x_j) + \frac{1}{2}\Delta t a^2 f_{xx}(x_j)$  即可递推。

\* 三种格式都有  $(2, 2)$  阶局部截断误差。

\* 直接从 CFL 稳定性条件可得  $\nu a \leq 1$ ，考虑交错网格蛙跳格式，在  $\nu a = 1$  时可取  $w^0 = 0$ ,  $v_j^{1/2} = (-1)^j$ ，递推可得 $\nu a < 1$ 。[由于离散对象仅带有一阶时间导数，这里的稳定性条件并不允许线性增长。]

### 哈密顿系统与辛格式

本部分用哈密顿问题的视角解释交错网格蛙跳格式的数值优势。

哈密顿问题是指，假设某系统总能量可以写为**哈密顿泛函**

$$\mathcal{H}(u, v) = \int E(x, t) dt$$

积分区域为给定的封闭空间， $E(x, t)$  代表**能量密度**， $u(x, t)$  为位移， $v(x, t)$  为速度。

则可以定义相应的哈密顿微分系统：定义  $\delta_u \mathcal{H}, \delta_v \mathcal{H}$  为  $\mathcal{H}$  的变分，即其满足

$$\int \delta_u \mathcal{H}(u, v) \delta u dx = \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}(u + \varepsilon \delta u, v) - \mathcal{H}(u, v)}{\varepsilon}, \quad \int \delta_v \mathcal{H}(u, v) \delta v dx = \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{H}(u, v + \varepsilon \delta v) - \mathcal{H}(u, v)}{\varepsilon},$$

对任意不影响物理约束的扰动函数  $\delta u, \delta v$  成立，则有哈密顿微分系统

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \delta_u \mathcal{H} \\ \delta_v \mathcal{H} \end{pmatrix}$$

\* 这里右侧的二阶矩阵具有**辛结构**，即其平方为负单位阵。

下面考虑一个具体的问题，小振幅钟摆系统的哈密顿泛函可写为

$$\mathcal{H}(u, v) = \int \left( f(u) + g(u_x) + \frac{1}{2}v^2 \right) dx$$

直接利用分部积分公式计算 [由空间方向周期性，两侧可消去] 可得

$$\delta_u \mathcal{H} = f'(u) - \frac{\partial g'(u_x)}{\partial x}, \quad \delta_v \mathcal{H} = v$$

于是

$$u_t = v, \quad v_t + \frac{\partial g'(u_x)}{\partial x} = f'(u)$$

定义**能量通量函数**  $F = -vg'(u_x)$ ，利用上式代入有  $E_t + F_x = 0$ ，代表某种**能量守恒**，再次使用分部积分可知**哈密顿泛函守恒**。

\* 哈密顿微分系统与哈密顿泛函的守恒事实上存在一般性的结论。

要长时间模拟哈密顿系统而不失准确，必须数值保持能量守恒定律，这样的数值格式最早出现在常微分方程中，称为**单辛格式**，推广到偏微分方程中称为**多辛格式**。

取  $f = 0, g(u_x) = \frac{1}{2}(-au_x)^2$ ，则相应的哈密顿微分系统即为  $u_t = v, v_t = a^2 u_{xx}$ ，与声波方程等价，于是能量守恒定律成立，记  $w = au_x$  有

$$E = \frac{1}{2}(v^2 + w^2), \quad F = avw$$

另一方面, 交错网格蛙跳格式的两迭代式可以推出

$$\Delta E \Delta x + \Delta F \Delta t = 0$$

这里

$$\Delta E = \frac{1}{2}((v_j^{n+1/2})^2 + (w_{j+1/2}^{n+1})^2) - \frac{1}{2}((v_j^{n-1/2})^2 + (w_{j+1/2}^n)^2)$$

$$\Delta F = \frac{a}{2}(v_j^{n-1/2}w_{j+1/2}^n + v_{j+1}^{n+1/2}w_{j+1/2}^n + v_{j+1}^{n+1/2}w_{j+1/2}^{n+1}) - \frac{a}{2}(v_j^{n+1/2}w_{j+1/2}^{n+1} + v_j^{n+1/2}w_{j-1/2}^n + v_j^{n-1/2}w_{j-1/2}^n)$$

此式事实上可以看作  $F dt - E dx$  沿着与网格线成  $45^\circ$  角的斜线构成的矩形积分为 0 的数值近似, 此积分恒为 0 利用 Green 公式即得与能量守恒等价。

\* 因此, 交错网格蛙跳格式的数值优势在于保持局部能量守恒, 事实上非交错网格盒子格式也有类似的性质。

### 高维对流方程

考虑  $u_t + au_x + bu_y = 0$ , 不妨设  $a, b$  均为正, 两方向的网比为  $\nu_x, \nu_y$ 。可直接构造其迎风格式:

$$u_{jk}^{n+1} = u_{jk}^n - \nu_x a (u_{jk}^n - u_{j-1,k}^n) - \nu_y b (u_{jk}^n - u_{j,k-1}^n)$$

\* 为方便稳定性分析, 假设  $a = b = 1$ ,  $\Delta x = \Delta y = h$ , 并记网比为  $\nu$  的特殊情况。此时, 空间点  $(x_j, y_k)$  在  $\Delta t$  时间的真实依赖区域边界点  $(x_j - \Delta t, y_k - \Delta t)$ , 而数值依赖区域是以此点为直角顶点, 两直角边长为  $h$  的向左下延伸的直角三角形, 直接计算可知 CFL 条件 [边界点包含在直角三角形中] 为  $\nu \leq \frac{1}{2}$ 。

\* 另一方面, 直接利用 Fourier 方法代入二维模态解可得到  $L^2$  稳定性的充要条件即为  $\nu \leq \frac{1}{2}$ 。

此外, 利用中心差商离散空间导数可得 LW 格式

$$u_{jk}^{n+1} = u_{jk}^n - \frac{1}{2}(\nu_x a \Delta_{0x} u_{jk}^n + \nu_y b \Delta_{0y} u_{jk}^n) + \frac{1}{2}(\nu_x^2 a^2 \delta_x^2 u_{jk}^n + \nu_y^2 b^2 \delta_y^2 u_{jk}^n) + \frac{1}{4} \nu_x \nu_y ab \Delta_{0x} \Delta_{0y} u_{jk}^n$$

\* 其离散模板含有 9 个空间网格点, 利用 Fourier 方法可知  $L^2$  模稳定性条件为

$$|\nu_x a| \leq \frac{1}{2\sqrt{2}}, \quad |\nu_y b| \leq \frac{1}{2\sqrt{2}}$$

\* 二维 LW 格式并不是一维 LW 格式的直接推广, 直接采用逐维离散会导致时间不再二阶相容且线性无条件  $L^2$  模不稳定。

算子分裂方法可改善高维双曲型方程计算效率, 如 LOD 格式

$$u^{n+1/2} = u^n - \frac{1}{2} \nu_x a \Delta_{0x} u^n + \frac{1}{2} \nu_x^2 a^2 \delta_x^2 u^n, \quad u^{n+1} = u^{n+1/2} - \frac{1}{2} \nu_y b \Delta_{0y} u^{n+1/2} + \frac{1}{2} \nu_y^2 b^2 \delta_y^2 u^{n+1/2}$$

有更宽松的时空约束, 若采用类似 Strang 格式的双重循环策略, 效果更加理想。

## 八 非线性双曲守恒律

考虑一维标量双曲守恒律

$$u_t + f(u)_x = 0, \quad x \in \mathbb{R}, t > 0$$

的纯初值或周期边值问题, 初值  $u(x, 0) = u_0(x)$ ,  $f$  是连续可微的已知通量函数。

与线性双曲型方程相比, 非线性双曲守恒律有很多不同的性质。

### §8.1 守恒型差分格式

#### 弱解

首先, 即使初值充分光滑, 非线性双曲守恒律的古典解也可能不存在。

设  $a = f'$ , 则连续可微的古典解  $u$  必然满足  $u_t + a(u)u_x = 0$ , 根据特征线理论可知特征线为  $\frac{dx}{dt} = a(u)$ , 均为直线段 [由于  $u$  在特征线上不变, 右端可视为常数], 即古典解必然满足方程

$$u(x, t) = u_0(x - a(u(x, t))t)$$

利用隐函数定理可知适当条件下古典解可以在短时间内存在唯一, 此时同一点出发特征线不交。然而, 考虑 Burgers 方程  $f(u) = \frac{u^2}{2}$ , 对严格单调减的初值, 由于  $\frac{dx}{dt} = u$ ,  $u_0(x)$  上越靠左的点对应特征线斜率越小, 必然相交, 交点处函数值不再唯一。由此, 我们必须推广解的概念。

考虑  $x \in \mathbb{R}, t \in \mathbb{R}^+$  上具有紧支集 [非零点闭包有界] 的函数构成检验函数空间  $\mathcal{H}$ , 双曲守恒律的弱解  $u(x, t)$  需满足

$$\forall \phi \in \mathcal{H}, \quad \iint_{t \geq 0} (u\phi_t + f(u)\phi_x) + \int_{t=0} u_0(x)\phi(x, 0) dx = 0$$

简单起见, 我们只考虑分片古典解, 且任意时刻间断点个数有限。设一条连续可微的时空界面曲线  $\Gamma: x = x(t), t \geq 0$  将上半平面分为左右两段, 解为

$$u(x, t) = \begin{cases} u_1(x, t) & x < x(t) \\ u_2(x, t) & x > x(t) \end{cases}$$

其中  $u_1, u_2$  均为古典解, 则其必须满足 **Rankine Hugoniot 跳跃条件** [RH 跳跃条件]:

$$\forall t, (u_+ - u_-)s = f(u_+) - f(u_-), \quad s(t) = x'(t), u_{\pm}(t) = \lim_{x \rightarrow x(t)^{\pm}} u(x, t)$$

这里上标  $\pm$  表示左极限与右极限。

\* 此条件可由弱解定义导出, 事实上对任何弱解与任何光滑曲线  $\Gamma$  成立, 本质是局部守恒性质。

\* 弱解常不唯一, 需要筛选出唯一的物理解。

#### 熵解

对双曲守恒律的某可如上分为  $u_1, u_2$  的弱解还满足 **Oleinik 熵条件**

$$\forall t, \frac{f(u_-) - f(v)}{u_- - v} \geq s \geq \frac{f(u_+) - f(v)}{u_+ - v}, \quad \forall v \in [\min(u_-, u_+), \max(u_-, u_+)]$$

则其称为熵解。

标量双曲守恒律的熵解存在唯一, 若  $f$  是凸可微的, Oleinik 熵条件可以简化为 **Osher 熵条件**

$$\forall t, f'(u_-) \geq s \geq f'(u_+)$$

事实上也即熵解的特征线不会远离时空界面曲线。

下面给出一些几何直观, 假设  $(x_*, t^*)$  处为熵解间断点, 则:

1. 若后续时刻的时空区域处处有特征线穿过, 则  $(x_*, t^*)$  将演化为间断界面  $x = x(t)$ , 满足之前的熵条件。若熵条件处处成立严格不等号, 两侧特征线交汇到间断界面, 称为**激波**, 移动速度  $s = x'(t)$  称为**激波速度**。
2. 若上述情况下熵条件局部成立等号, 则特征线与间断界面局部平行, 此时退化为线性双曲型方程, 间断结构称为**接触间断**, 类似之前线性常系数对流方程的间断初值引起的间断界面。

3. 若后续时刻的某个扇形时空区域某处无特征线穿过, 则间断点将消失, 相应的熵解局部结构为**稀疏波**, 扇形区域内部存在自相似结构  $u(x, t) = \tilde{u}(\frac{x-x_*}{t-t_*})$ , 将外侧的两状态连接。除去起始点外后续时刻稀疏波局部连续, 但边界的空间导数间断。

\* 以 Burgers 方程  $u_t + (u^2/2)_x = 0$  为例, 假设初值为

$$u_0(x) = \begin{cases} u_- & x < 0 \\ u_+ & x > 0 \end{cases}$$

当  $u_- = u_+$  时熵解为  $u(x, t) = u_-$  常值, 属于古典解。当  $u_- > u_+$  时, 熵解为**激波**, 记  $s = \frac{u_- + u_+}{2}$  有

$$u(x, t) = \begin{cases} u_- & x - st < 0 \\ u_+ & x - st > 0 \end{cases}$$

而当  $u_- < u_+$  时熵解为**稀疏波**

$$u(x, t) = \begin{cases} u_- & x < u_- t \\ \frac{x}{t} & x \in [u_- t, u_+ t] \\ u_+ & x > u_+ t \end{cases}$$

[事实上激波解一定为弱解, 但  $u_- < u_+$  时不为熵解。]

由此, 非线性双曲守恒律的数值格式应该做到:

1. 激波速度刻画与间断界面捕获准确;
2. 真解相对光滑区域, 相容阶与计算效率较高;
3. 间断界面附近数值振荡得到控制;
4. 数值解收敛到熵解 [或至少是弱解]。

实现上述目标的格式称为**高精度高分辨率格式**, 常用构造方法为激波装配技术 (先确定间断界面再分别计算光滑解) 与激波捕捉技术 (直接采用统一数值操作过程模拟), 我们主要介绍后者。

\* 基于激波捕捉技术的**守恒型差分格式**构造一般分为两步: 假设真解足够光滑得到数值性质、存在间断结构时讨论健壮性。

记  $a = f'$ , 对  $u_t + a(u)u_x = 0$  的离散称为基于**非守恒形式**的构造:

在离散焦点进行简单系数冻结可得**迎风格式**

$$u_j^{n+1} = \begin{cases} u_j^n - \nu a(u_j^n)(u_j^n - u_{j-1}^n) & a(u_j^n) \geq 0 \\ u_j^n - \nu a(u_j^n)(u_{j+1}^n - u_j^n) & a(u_j^n) < 0 \end{cases}$$

\* 无条件具有 (2, 2) 阶局部截断误差, 利用 CFL 或系数冻结可得模糊的稳定性条件

$$\max_{j,n} |a(u_j^n)|\nu \leq 1$$

利用算术平均虚化离散焦点可得 **Lax 格式**

$$u_j^{n+1} = \frac{1}{2}(u_{j-1}^n + u_{j+1}^n) - \frac{1}{2}\nu a(u_j^n)(u_{j+1}^n - u_{j-1}^n)$$

\* 网比固定时有整体一阶局部截断误差, 模糊的稳定性条件与迎风格式相同。根据离散最大模原理, 上述条件成立时两格式均**最大模稳定**。



对时间 Taylor 展开后转换为空间导数可得 **LW 格式**, 记  $a_j^n = a(u_j^n), b_j^n = a(u_j^n)a'(u_j^n)$ , 则有

$$u_j^{n+1} = u_j^n - \frac{1}{2}\nu a_j^n \Delta_{0x} u_j^n + \frac{1}{2}(\nu a_j^n)^2 \delta_x^2 u_j^n + \frac{1}{4}\nu^2 b_j^n (\Delta_{0x} u_j^n)^2$$

\* 无条件具有 (1, 1) 阶局部截断误差, 模糊的稳定性条件同上。

\* 非线性问题的差分格式可能存在非线性不稳定现象, 如对 Burgers 方程的非守恒性质  $u_t + uu_x = 0$  可定义蛙跳格式为

$$u_j^{n+1} = u_j^{n-1} - \nu u_j^n (u_j^{n+1} - u_j^{n-1})$$

考虑空间网格函数  $r_j$  为  $0, \varepsilon, -\varepsilon$  循环, 时间网格函数满足  $c^{n+1} - c^{n-1} = \nu \epsilon (c^n)^2$ , 则可验证  $u_j^n = c^n r_j$  精确满足蛙跳格式。令  $c^0 = 1, c^1 = \alpha$  可归纳得到  $c^n \geq (1 + \gamma)^{\lfloor n/2 \rfloor}$ , 即其指数趋于无穷。于是对任何网比, 蛙跳格式不稳定。

反之, 若将 Burgers 方程写为  $u_t + \frac{1}{3}uu_x + \frac{1}{3}(u^2)_x = 0$ , 则蛙跳格式为

$$u_j^{n+1} = u_j^{n-1} - \frac{\nu}{3}(u_{j+1}^n + u_j^n + u_{j-1}^n)(u_{j+1}^n - u_{j-1}^n)$$

可改善非线性因素的不稳定。

记  $f_j^n = f(u_j^n)$ , 直接基于守恒形式也可构造:

迎风格式可写为

$$u_j^{n+1} = \begin{cases} u_j^n - \nu(f_j^n - f_{j-1}^n) & a(u_j^n) \geq 0 \\ u_j^n - \nu(f_{j+1}^n - f_j^n) & a(u_j^n) < 0 \end{cases}$$

**Lax 格式**即

$$u_j^{n+1} = \frac{1}{2}(u_{j-1}^n + u_{j+1}^n) - \frac{1}{2}\nu(f_{j+1}^n - f_{j-1}^n)$$

而利用积分插值的思路可构造 **LW 格式**, 记  $a_{j+1/2}^n = a(u_{j+1/2}^n) = a(\frac{u_j^n + u_{j+1}^n}{2})$ , 则

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}(f_{j+1}^n - f_{j-1}^n) + \frac{\nu^2}{2}(a_{j+1/2}^n(f_{j+1}^n - f_j^n) - a_{j-1/2}^n(f_j^n - f_{j-1}^n))$$

\* 三者相容性、局部截断误差、模糊的稳定性条件与基于非守恒形式的结果完全相同。

\* 若  $f(u) = au$ , 无论是基于非守恒形式还是守恒形式的三种格式都能退化为线性常系数的对应情况。

### 健壮性分析

考虑 Burgers 方程, 初值为  $u_- = 1, u_+ = 0$ , 对应真解为  $s = \frac{1}{2}$  的激波, 数值初值  $u_j^0$  当  $j \leq 1$  时为 1, 否则为 0。

基于守恒与非守恒的迎风格式数值解始终为初值, 失效。[由于流动方向的改变, 迎风格式丧失了整体结构的统一性。]

\* 基于非守恒形式构造的差分无法刻画激波的**局部守恒**, 必然失效。

由此需要**守恒型格式**的概念, 定义为可以表述为

$$u_j^{n+1} = u_j^n - \nu(\hat{f}_{j+1/2}^n - \hat{f}_{j-1/2}^n)$$

的格式,  $\hat{f}_{j+1/2}^n$  为数值通量, 为  $\hat{f}(u_{j-l+1}^n, u_{j-l+2}^n, \dots, u_{j+r}^n)$ ,  $l, r$  为给定正整数,  $\hat{f}$  为给定的数值通量函数。由定义可知守恒格式满足**局部守恒性质**

$$\sum_{j=p}^q u_j^{n+1} = \sum_{j=p}^q u_j^n - \nu(\hat{f}_{q+1/2}^n - \hat{f}_{p-1/2}^n)$$

\* 数值通量函数一般要求对于每个边缘 Lipschitz 连续, 且有相容性条件  $\hat{f}(p, p, \dots, p) = f(p)$ , 前者用于控制舍入误差, 后者保证相容性。

基于守恒形式构造的 Lax 格式数值通量为

$$\hat{f}_{j+1/2}^n = \frac{1}{2}(f_j^n + f_{j+1}^n) - \frac{1}{2\nu}(u_{j+1}^n - u_j^n)$$

而基于守恒形式的 LW 格式数值通量则为

$$\hat{f}_{j+1/2}^n = \frac{1}{2}(f_j^n + f_{j+1}^n) - \frac{\nu}{2}a_{j+1/2}^n(f_{j+1}^n - f_j^n)$$

验证数值通量函数条件即知二者均为守恒型。

\* 两格式的具有不同的性质, 当 CFL 条件成立时, Lax 的数值通量对第一个变元不减, 对第二个变元不减, 称为**熵数值通量**或**单调数值通量**, LW 格式则无此性质。

守恒格式的优点包括: 数值误差  $u_j^n - [u]_j^n$  在空间方向的求和与时间无关, 满足整体**质量守恒**; **RH 跳跃条件**近似满足, 间断界面可以被捕获; **Lax-Wendroff 定理** [LW 定理]: 守恒型差分格式相容时, 数值解若在网格尺度趋于 0 时几乎处处有界且收敛, 则收敛结果必为**弱解**。

### 双曲守恒律组

设通量函数  $\mathbf{f}: \mathbb{R}^m \rightarrow \mathbb{R}^m$  的 Jacobi 阵处处可实相似对角化为  $\mathbb{S}(\mathbf{u})^{-1}\mathbb{D}(\mathbf{u})\mathbb{S}(\mathbf{u})$ , 则称

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = 0$$

为双曲守恒律组。若特征值处处互不相同, 则称为严格双曲守恒律组。

若真解连续可微, 原方程可化为**特征形式**  $\mathbb{S}(\mathbf{u})\mathbf{u}_t + \mathbb{D}(\mathbf{u})\mathbb{S}(\mathbf{u})\mathbf{u}_x = 0$ 。若向量函数  $\mathbf{r}$  满足

$$\mathbf{r}_t = \mathbb{S}(\mathbf{u})\mathbf{u}_t, \quad \mathbf{r}_x = \mathbb{S}(\mathbf{u})\mathbf{u}_x$$

则称为方程组的 **Riemman 不变量**, 代入可得  $\mathbf{r}_t + \mathbb{D}(\mathbf{u})\mathbf{r}_x = 0$ 。

\* 由于  $\mathbb{D}$  对角元与  $\mathbf{u}$  相关, 事实上并没有解耦。此外, 黎曼不变量在  $m = 2$  时一定存在, 但  $m > 2$  时未必存在。

下面考虑 Riemman 不变量存在时的求解。若  $t^n$  时刻  $\mathbf{u}^n$  已知, 可局部近似为  $\mathbf{v}_t + D(\mathbf{u}_j^n)\mathbf{v}_x = 0$ , 于是**迎风离散**的迭代过程为:

1. 计算  $\mathbf{v}_j^n = \mathbb{S}(\mathbf{u}_j^n)\mathbf{u}_j^n$ ;
2. 局部离散近似方程得到  $[\mathbb{D}^\oplus$  与  $\mathbb{D}^\ominus$  定义同上章]

$$\mathbf{v}_j^{n+1} = \mathbf{v}_j^n - \nu\mathbb{D}^\oplus(\mathbf{u}_j^n)(\mathbf{v}_j^n - \mathbf{v}_{j-1}^n) - \nu\mathbb{D}^\ominus(\mathbf{u}_j^n)(\mathbf{v}_{j+1}^n - \mathbf{v}_j^n)$$

3. 还原得到  $\mathbf{u}_j^{n+1} = \mathbb{S}(\mathbf{u}_j^n)^{-1}\mathbf{v}_j^{n+1}$ 。

由于计算效率很低, 无需解耦的格式更受欢迎, 如 Lax 格式与 LW 格式, 为对应的基于守恒形式构造的格式将  $u, f$  替换为向量, 其中 LW 格式  $a_{j+1/2}^n$  替换为  $\frac{\mathbf{u}_j^n + \mathbf{u}_{j+1}^n}{2}$  处的 Jacobi 矩阵。

计算 Jacobi 矩阵仍需大量时间, 因此对其优化得到两步 LW 格式, 如先后使用 Lax 与蛙跳的 **Richtmyer** 格式

$$\begin{aligned} \mathbf{u}_{j+1/2}^{n+1/2} &= \frac{1}{2}(\mathbf{u}_j^n + \mathbf{u}_j^{n+1}) - \frac{\nu}{2}(\mathbf{f}(\mathbf{u}_{j+1}^n) - \mathbf{f}(\mathbf{u}_j^n)) \\ \mathbf{u}_j^{n+1} &= \mathbf{u}_j^n - \nu(\mathbf{f}(\mathbf{u}_{j+1/2}^{n+1/2}) - \mathbf{f}(\mathbf{u}_{j-1/2}^{n+1/2})) \end{aligned}$$

或先后使用蛙跳与 Lax 的 **MacCormack** 格式

$$\begin{aligned} \tilde{\mathbf{u}}_j^n &= \mathbf{u}_j^n - \nu(\mathbf{f}(\mathbf{u}_{j+1}^n) - \mathbf{f}(\mathbf{u}_j^n)) \\ \mathbf{u}_j^{n+1} &= \frac{1}{2}(\mathbf{u}_j^n + \tilde{\mathbf{u}}_j^n) - \frac{\nu}{2}(\mathbf{f}(\tilde{\mathbf{u}}_j^n) - \mathbf{f}(\tilde{\mathbf{u}}_{j-1}^n)) \end{aligned}$$

二者均有整体二阶的局部截断误差。

## §8.2 有限体积分法

对于发展型 PDE, 有限体积格式对时空变量采取不同的离散方式, 先将空间区域进行单元剖分,  $I_j = (x_{j-1/2}, x_{j+1/2})$ , 记  $\Delta x_j$  为  $I_j$  长度, 其最大值  $\Delta x$  为剖分的参数。

将时间离散为有限个点,  $\Delta t^n = t^{n+1} - t^n$  为局部时间步长, 其最大值  $\Delta t$  为时间步长。

有限体积分法数值求解的目标定义为

$$[\bar{u}]_j^n = [\bar{u}]_j(t^n) = \frac{1}{\Delta x_j} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^n) dx$$

即真解在不同单元的均值。

利用 Green 公式对矩形时空区域积分可得, 记  $F_{j+1/2}^n = \frac{1}{\Delta t^n} \int_{t^n}^{t^{n+1}} f(u(x_{j+1/2}, t)) dt$ , 则有精确等式

$$[\bar{u}]_j^{n+1} = [\bar{u}]_j^n - \frac{\Delta t^n}{\Delta x_j} (F_{j+1/2}^n - F_{j-1/2}^n)$$

这里  $F$  为真实平均通量, 若有数值通量函数  $\hat{f}([\bar{u}]_{j-l-1}^n, \dots, [\bar{u}]_{j+r}^n)$  近似为  $F_{j+1/2}^n$ , 将其记为  $\hat{f}_{j+1/2}^n$  即有递推

$$\bar{u}_j^{n+1} = \bar{u}_j^n - \frac{\Delta t^n}{\Delta x_j} (\hat{f}_{j+1/2}^n - \hat{f}_{j-1/2}^n)$$

\* 通常  $\hat{f}_{j+1/2}^n = \hat{f}(\bar{u}_j^n, \bar{u}_{j+1}^n)$  只与两点有关, 数值操作更加灵活, 也可推广到高维。

\* 有限体积方法可类似有限差分定义相容、稳定与收敛。

### 线性问题

考虑  $f(u) = au$ ,  $a$  为给定常数, 回到线性问题。为方便, 假设离散网格均匀, 时间步长与单元长度恒定为  $\Delta t, \Delta x$ 。

定义数值通量

$$\hat{f}_{j+1/2}^n = \begin{cases} a\bar{u}_j^n & a \geq 0 \\ a\bar{u}_{j+1}^n & a < 0 \end{cases}$$

其只依赖上游单元, 因此称为迎风数值通量, 对应的迎风有限体积格式

$$\frac{\bar{u}_j^{n+1} - \bar{u}_j^n}{\Delta t} + \frac{a + |a|}{2\Delta x} (\bar{u}_j^n - \bar{u}_{j-1}^n) + \frac{a - |a|}{2\Delta x} (\bar{u}_{j+1}^n - \bar{u}_j^n)$$

\* 事实上迎风数值通量可视为中心型数值通量  $\frac{a}{2}(\bar{u}_j^n + \bar{u}_{j+1}^n)$  加上数值黏性修正项  $-\frac{|a|}{2}(\bar{u}_{j+1}^n - \bar{u}_j^n)$ ,  $\frac{|a|}{2}$  称为修正强度, 括号内为界面位置的数值跳跃。

记修正强度为  $\frac{2}{\nu}$  得到 Lax 数值通量, 对应 Lax 有限体积格式

$$\bar{u}_{j+1}^n = \frac{1}{2}(\bar{u}_{j-1}^n + \bar{u}_{j+1}^n) - \frac{\nu a}{2}(\bar{u}_{j+1}^n - \bar{u}_j^n)$$

\* 若函数足够光滑, 单元均值与中心点值差距  $O((\Delta x)^2)$ , 相容阶不超过 2 的有限差分与有限体积可转化, 反之亦然。

\* 有限体积法的收敛性需要运用有限元方法的分析技术, 尤其单元剖分非均匀时。

\* 有限体积与积分插值的想法类似, 但积分插值依靠周边点值进行数值积分; 有限体积方法通过周边均值进行数值积分。事实上迎风格式、Lax 格式等也可通过积分插值构造。

\* 由此, 有限体积与有限差分关联密切, 如无特殊说明可省略均值符号  $\bar{u}$ , 仍记为  $u$ , 格式可作两种理解。

### 非线性问题

回到一般的非线性双曲守恒律。记  $\bar{u}_{j+1/2}^n = \frac{1}{2}(\bar{u}_j^n + \bar{u}_{j+1}^n)$ , 并记  $A_{j+1/2}^n = f'(\bar{u}_{j+1/2}^n)$ ,  $f_j^n = f(\bar{u}_j^n)$ , 则仍可定义迎风数值通量

$$\hat{f}_{j+1/2}^n = \begin{cases} f_j^n & A_{j+1/2}^n \geq 0 \\ f_{j+1}^n & A_{j+1/2}^n < 0 \end{cases}$$

对应得到**守恒型迎风格式** [省略均值符号]

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}((1 - \operatorname{sgn} A_{j+1/2}^n)\Delta_x f_j^n + (1 + \operatorname{sgn} A_{j-1/2}^n)\Delta_{-x} f_j^n)$$

\* 之前的简单迎风格式将此处的  $A_{j+1/2}^n, A_{j-1/2}^n$  都替换为  $A_j^n$ , 即上游方向判断位置有差异。

类似地, **Lax 有限体积格式**为

$$u_j^{n+1} = \frac{1}{2}(u_{j-1}^n + u_{j+1}^n) - \frac{\nu}{2}(f_{j+1}^n - f_{j-1}^n)$$

\* 注意到基于守恒形式构造的 Lax 格式已具有局部守恒性, 其形式上 [即省略均值符号时] 与 Lax 有限体积格式完全相同。

而 **LW 有限体积格式**形式上也与基于守恒形式构造的 LW 格式完全相同, 只是将  $a_{j+1/2}^n$  替换为  $A_{j+1/2}^n$ , 而  $a_{j+1/2}^n$  与  $A_{j+1/2}^n$  形式上亦相同。

### Godunov 方法

\* 公认的首个成功模拟非线性双曲守恒律的有限体积格式, 已发展为主流数值方法, 以下均省略均值符号。

#### EA 过程

若  $u_j^n$  已知, 计算  $u_j^{n+1}$  可以分为两步:

1. **局部推进**: 在  $x_{j+1/2}$  处构造局部 Riemman 问题, 初值为

$$u(x, t^n) = \begin{cases} u_j^n & x < x_{j+1/2} \\ u_{j+1}^n & x > x_{j+1/2} \end{cases}$$

其真解可能是古典解、激波、接触间断或稀疏波等, 当  $u_{j+1}^n \neq u_j^n$  时记  $s_{j+1/2}^n = \frac{f_{j+1}^n - f_j^n}{u_{j+1}^n - u_j^n}$  表示 (可能存在的) 激波速度, 若时间步长满足  $\max_j \{|s_{j+1/2}^n|, |f'(u_j^n)|\} \nu \leq \frac{1}{2}$ , 相邻的局部 Riemann 解在  $t^{n+1}$  前不会冲突, 拼接成的函数记为  $\tilde{u}(x, t^{n+1})$ 。

2. **单元平均**: 定义

$$u_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \tilde{u}(x, t^{n+1}) dx$$

\* 由于单元平均的逼近效果, 局部截断误差为整体一阶。

\* 以线性情况  $f(u) = au$  为例, 解得局部

$$\tilde{u}(x, t^{n+1}) = \begin{cases} u_j^n & x < x_{j+1/2} + a\Delta t \\ u_{j+1}^n & x > x_{j+1/2} + a\Delta t \end{cases}$$

不妨设  $a > 0$ , 积分即有 [按上方推导, 这里时间步长应满足  $\nu a \leq \frac{1}{2}$ , 但事实上可以放宽到  $\nu a \leq 1$ ]

$$u_j^{n+1} = u_{j-1}^n a\nu + (1 - a\nu)u_j^n$$

这事实上就是线性双曲方程的迎风格式。此外, 数值通量  $au_j^n$  恰为局部 Riemman 解  $\tilde{u}$  在  $x_{j+1/2}$  的通量取值  $f(\tilde{u}(x_{j+1/2}, t^{n+1}))$ 。

关于数值通量的结论事实上普遍成立, 因为其实现过程等同于双曲守恒律在  $I_j \times (t^n, t^{n+1})$  的积分, 即其满足

$$u_j^{n+1} = u_j^n - \nu(\hat{f}_{j+1/2}^n - \hat{f}_{j-1/2}^n), \quad \hat{f}_{j+1/2}^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(\tilde{u}(x_{j+1/2}, t)) dt$$

另一方面, 根据 PDE 知识,  $\tilde{u}$  可表示为  $\mathcal{R}(\frac{x-x_{j+1/2}}{t-t^n}; u_j^n, u_{j+1}^n)$ , 这里  $\mathcal{R}$  称为**局部 Riemman 解算子**, 于是  $f(\tilde{u}(x_{j+1/2}, t))$  在积分区域恒定, 即有

$$\hat{f}_{j+1/2}^n = f(\mathcal{R}(0; u_j^n, u_{j+1}^n))$$

于是, 数值通量构造方式的核心工作为局部 Riemann 解的计算。对一般的非线性双曲守恒律, 其往往难以准确计算, 可以将其局部近似为线性双曲型方程  $u_t + A_{j+1/2}^n u_x = 0$ 。这样得到的格式即为一般的**守恒型迎风格式**。

\* 更有效的局部线性化方式为定义系数为单元界面的 **Roe 平均值**

$$A_{j+1/2} = \begin{cases} \frac{f_{j+1}^n - f_j^n}{u_{j+1}^n - u_j^n} & u_{j+1}^n \neq u_j^n \\ f'(u_j^n) & u_{j+1}^n = u_j^n \end{cases}$$

其满足离散版本 RH 跳跃条件

$$A_{j+1/2}^n (u_{j+1}^n - u_j^n) = f_{j+1}^n - f_j^n$$

可以更准确刻画激波条件, 这样得到的迎风格式称为 **Roe 型迎风格式**, 真解光滑时与守恒型迎风格式差异为二阶, 但间断真解时效果更好。

### REA 过程

REA 过程为局部推进前增加**高阶重构** [Reconstruction] 操作, 即局部将初值定义为  $u(x, t^n) = u_j^n + \sigma_j^n \frac{x - x_j}{\Delta x}, x \in I_j$ 。这里  $\sigma_j^n$  称为**广义斜率**, 可定义为  $u_{j+1}^n - u_j^n$ 、 $u_j^n - u_{j-1}^n$ 、 $\frac{1}{2}u_{j+1}^n - u_{j-1}^n$  等。

\* 若广义斜率恒为 0, 则退化为 EA 过程。

\* 考虑  $\sigma_j^n = u_{j+1}^n - u_j^n$  的线性问题  $f(u) = au$ , 直接计算可得局部解为

$$\tilde{u}(x, t^{n+1}) = \begin{cases} u_j^n + \frac{1}{2}(u_{j+1}^n - u_j^n) & x < x_{j+1/2} + a\Delta t \\ u_{j+1}^n - \frac{1}{2}(u_{j+2}^n - u_{j+1}^n) & x > x_{j+1/2} + a\Delta t \end{cases}$$

当  $\nu|a| \leq 1$  时相邻的解不存在冲突, 计算单元均值可发现其即为线性双曲方程的 LW 格式。

\* 类似地, 对非线性情况, 取  $\sigma_j^n = u_{j+1}^n - u_j^n$  后线性近似可得到一般的**守恒型 LW 格式**, 将  $A_{j+1/2}^n$  用 Roe 平均替代则得到 **Roe 型 LW 格式**。

## §8.3 稳定性与收敛性

\* 虽然守恒型格式可以某种程度上保证收敛到弱解, 但高阶相容与数值振荡的矛盾仍存在, 对 Burgers 方程实验可发现 LW 格式有二阶局部截断误差, 但间断界面附近数值振荡, 网格加密也无法减弱振荡; 迎风格式只有一阶局部截断误差, 但数值间断界面更平坦, 没有数值振荡。因此, 需要改善数值解在间断界面的光滑度与陡峭度。

### 单调格式

由 PDE 理论可知, 双曲守恒律**若初值单调**, 则**任何时刻熵解有相同单调性**, 若数值格式也满足此性质则称为**单调保持格式**。此性质一般难以验证, 因此我们关注一更强的性质:

双曲守恒律的熵解满足若初值  $v_0(x) \geq u_0(x)$  处处成立, 则此后一定都有  $v(x, t) \geq u(x, t)$ 。若数值格式满足此性质, 即

$$v_j^n \geq u_j^n, \forall j \implies v_j^{n+1} \geq u_j^{n+1}, \forall j$$

则称为**单调格式**, 其等价定义为, 对给定的臂长  $l, r$ , 格式

$$u_j^{n+1} = H(u_{j-l}^n, u_{j-l+1}^n, \dots, u_{j+r}^n)$$

对  $H$  的每个分量非减, 可微时即偏导非负。

\* 由等价定义可验证**单调格式一定是单调保持格式**, 而对线性的情况, 单调格式与单调保持格式等价。

\* 验证可知守恒型 Lax 格式在 CFL 条件下为单调格式, 而守恒型 LW 格式并非单调格式 [事实上可举反例说明 LW 格式并非单调保持格式]。

对单调格式，有结论其数值解一致有界，且收敛到熵解，但 Godunov 定理依然成立，即其至多一阶截断误差。

\* 构造  $u_{\pm} = \pm 1$ ，对应的 Burgers 方程的 Riemann 问题利用 Roe 型迎风格式可发现数值解保持不变，因此只收敛到弱解而非熵解，Roe 型迎风格式并非单调格式。

### TVD 格式

\* 由于 Godunov 定理，高阶的单调保持格式不可能单调，也即会存在微弱的数值振荡。[且我们希望振荡能随网格加密消失。]

我们用全变差描述某区间 Lebesgue 可测函数的振荡强度：

$$TV(v) = \limsup_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int |v(x) - v(x - \varepsilon)| dx$$

\* 若考虑弱导数，全变差也可以记为对  $|v'(x)|$  的积分。此外，其可以视为某种弱化定义的度量，作为数值不稳定的衡量标准。

PDE 理论可以证明熵解满足时间增加时全变差不增，而数值全变差定义为

$$TV(w) = \sum_j |w_{j+1} - w_j|$$

若数值全变差随时间增加不增，则称为 **TVD 格式** [Total Variation Diminishing]。

**Harten 引理**：设数值格式具有增量形式 [这里  $C, D$  可能与  $u$  相关]

$$u_j^{n+1} = u_j^n - C_{j-1/2}(u_j^n - u_{j-1}^n) + D_{j+1/2}(u_{j+1}^n - u_j^n)$$

且满足  $C_{j+1/2} \geq 0, D_{j+1/2} \geq 0, C_{j+1/2} + D_{j+1/2} \leq 1$  则直接利用三角不等式可得到其是 TVD 格式。

\* 在 CFL 条件下，对 Roe 型迎风格式可取

$$C_{j+1/2} = \frac{\nu}{2} (1 + \operatorname{sgn} A_{j+1/2}^n) \frac{\Delta_x f_j^n}{\Delta_x u_j^n}, \quad D_{j+1/2} = -\frac{\nu}{2} (1 - \operatorname{sgn} A_{j+1/2}^n) \frac{\Delta_x f_j^n}{\Delta_x u_j^n}$$

可从 Harten 引理验证其 TVD 性。

\* **单调格式是 TVD 格式，TVD 格式是单调保持格式**，且对线性情况三者等价。

\* 由于 Roe 型迎风格式是守恒型 TVD 格式，其可以避免数值振荡，但未必收敛到熵解，这是由于计算过程中只会出现激波，完全忽略了稀疏波。区别这两种结构需要使用熵修正，如数值通量改为

$$\hat{f}_{j+1/2}^n = \frac{1}{2} (1 + \operatorname{sgn} f'(u_j^n)) f_j^n + \frac{1}{2} (1 - \operatorname{sgn} f'(u_{j+1}^n)) f_{j+1}^n + \frac{1}{2} (\operatorname{sgn} f'(u_{j+1}^n) - \operatorname{sgn} f'(u_j^n)) f(u_s)$$

这里  $u_s$  满足  $f'(u_s) = 0$ ，称为**声波点**。在  $f$  凸可微时，熵修正的 Roe 型迎风格式是单调格式，保证数值解收敛到熵解。

### TVD 修正技术

\* 我们希望能构造高阶的 TVD 格式，一般有数值通量修正与数值斜率修正两种技术实现。

**数值通量修正技术**：

\* 原始思想为人工黏性方法，即数值解局部间断时局部增加数值黏性，压制可能产生的数值振荡。由于数值黏性会影响相容阶，这相当于高阶格式到低阶格式的自动跳转。

考虑两个数值通量分别为  $\hat{f}_H$  与  $\hat{f}_L$  的守恒型差分格式，前者高阶后者低阶，定义数值通量

$$(\hat{f}_M)_{j+1/2}^n = \theta_{j+1/2}^n (\hat{f}_H)_{j+1/2}^n + (1 - \theta_{j+1/2}^n) (\hat{f}_L)_{j+1/2}^n$$

这里  $\theta_{j+1/2}^n$  称为**开关函数或通量限制器**，常见定义为

$$\theta_{j+1/2}^n = \max(0, \min(1, s_{j+1/2}^n)), \quad s_{j+1/2}^n = \begin{cases} \frac{u_j^n - u_{j-1}^n}{u_{j+1}^n - u_j^n} & A_{j+1/2}^n \geq 0 \\ \frac{u_{j+2}^n - u_{j+1}^n}{u_{j+1}^n - u_j^n} & A_{j+1/2}^n < 0 \end{cases}$$

\* 利用 Harten 引理可证明高阶格式是 Roe 型 LW 格式, 低阶格式是 Roe 型迎风格式时, 由此数值通量构造结果是二阶 TVD 格式。

**数值斜率修正技术:**

我们介绍源自 LW 格式 REA 过程的 **MUSCL**[Monotone Upwind Scheme for Conservation Law] 格式, 回顾 REA 过程第一步的定义 [省略均值符号]

$$u(x, t^n) = u_j^n + \sigma_j^n \frac{x - x_j}{\Delta x}, \quad x \in I_j$$

LW 格式中广义斜率定义为  $u_{j+1}^n - u_j^n$ , 而我们修正定义为

$$\sigma_j^n = \minmod\left(u_{j+1}^n - u_j^n, u_j^n - u_{j-1}^n, \frac{1}{2}(u_{j+1}^n - u_{j-1}^n)\right)$$

这里 minmod 表示后方三个数符号相同时取其中绝对值最小的数, 否则为 0, 称为**斜率限制器**。

\* 利用 Harten 引理可证明 CFL 条件下 MUSCL 格式是 TVD 格式。

\* 三个数符号不同时, 相邻三个单元的均值有增有减, 可能产生振荡, 因此抹平广义斜率, 退化为迎风格式; 三个数符号相同时, 校正解使得斜率最小, 压制整体数值振荡。然而, 此限制器可能将光滑解的极值判断为间断解, 需要结合其他技术改善。

\* 一维 TVD 格式至多有二阶局部截断误差, 高维 TVD 格式至多一阶局部截断误差, 需要提出更多格式的限制。

## 九 发展型偏微分方程

本章是关于发展型偏微分方程 (含  $u_t$  项的 PDE) 的最后一章, 分为一般的对流扩散方程离散与一般的稳定性分析技术两部分。

### §9.1 对流扩散方程

多数情况下对流与扩散同时出现, 此时方程可写为

$$u_t + cu_x = au_{xx}$$

简单起见考虑流动速度  $c$  与扩散速度  $a \geq 0$  为常数。记对流网比  $\nu = \frac{\Delta t}{\Delta x}$ , 扩散网比  $\mu = \frac{\Delta t}{(\Delta x)^2}$ 。

**中心差商显格式**

最基本的格式为

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{c(u_{j+1}^n - u_{j-1}^n)}{2\Delta x} = \frac{a\delta_x^2 u_j^n}{(\Delta x)^2}$$

其无条件具有 (2, 1) 阶局部截断误差。

代入  $u_j^n = \lambda^n e^{ikj\Delta x}$  可以得到

$$\lambda(k) = 1 - i\nu c \sin(k\Delta x) - 4\mu a \sin^2 \frac{k\Delta x}{2}$$

直接计算可知  $\mu a \leq \frac{1}{2}$  时有

$$|\lambda(k)| \leq 1 + \frac{c^2}{4a} \Delta t$$

符合 von Neumann 条件,  $L^2$  模稳定。

然而, 当  $a \ll |c|$  的对流占优情况, 误差可能会变得非常巨大, 因此需要引入**强稳定性**: 数值解的稳定性表现同真解的适定性一致。

对流扩散方程真解  $L^2$  模不增, 因此  $L^2$  模强稳定性等价于  $|\lambda(k)| \leq 1$ , 可得到对中心差商显格式即

$$(\nu c)^2 \leq 2\mu a \leq 1$$

\* 第一个式子要求  $\Delta t \leq \frac{2a}{c^2}$ , 当  $a \ll |c|$  时需要很小。

利用离散最大模原理, 最大模强稳定性即等价于

$$\nu|c| \leq 2\mu a \leq 1$$

\* 第一个式子要求  $\Delta x \leq \frac{2a}{|c|}$ , 当  $a \ll |c|$  时亦需很小。

\* 以下为方便设  $c > 0$ , 讨论如何解决  $a \ll |c|$  时的离散问题。

### 数值黏性修正

直接采用一阶导数的偏心迎风离散, 定义

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{c(u_j^n - u_{j-1}^n)}{\Delta x} = \frac{a\delta_x^2 u_j^n}{(\Delta x)^2}$$

\* 具有 (1,1) 阶局部截断误差, 验证得  $L^2$  模强稳定性充要条件为  $2\mu a + \nu c \leq 1$ 。

也可类似 LW 格式离散, 建立双曲部分的 LW 格式后加入扩散部分, 得到修正中心差商显格式

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{c(u_{j+1}^n - u_{j-1}^n)}{2\Delta x} = \left(a + \frac{c^2}{2}\Delta t\right) \frac{\delta_x^2 u_j^n}{(\Delta x)^2}$$

\* 具有 (2,1) 阶局部截断误差, 验证得  $L^2$  模强稳定性充要条件为  $2\mu a + (\nu c)^2 \leq 1$ 。

**指数格式:** 从稳态方程  $d + cu_x = au_{xx}$  出发设计, 我们希望差分方程  $\alpha u_{j-1} + \beta u_j + \gamma u_{j+1}$  在所有网格点为 0, 这里  $\alpha, \beta, \gamma$  与  $d$  无关为待定系数, 推导出它们的形式后再将  $d$  替换为  $u_t$  并差分, 得到

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{c(u_{j+1}^n - u_{j-1}^n)}{2\Delta x} = a\sigma \frac{\delta_x^2 u_j^n}{(\Delta x)^2}$$

这里拟合因子  $\sigma$  满足

$$\sigma = R \coth R, \quad R = \frac{c\Delta x}{2a}$$

$R$  称为**网格 Péclet 数**。其具有 (2,1) 阶局部截断误差,  $L^2$  模强稳定性条件为  $\sigma\mu a \leq \frac{1}{2}$ 。

**Samapckii 格式:** 将指数格式中的一阶导数改为迎风差商离散, 并对  $\sigma$  Taylor 展开简化, 得到

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{c(u_j^n - u_{j-1}^n)}{\Delta x} = \frac{a}{1+R} \frac{\delta_x^2 u_j^n}{(\Delta x)^2}$$

\* 具有 (2,1) 阶局部截断误差, 回避了指数函数的计算。

### 隐式格式

考虑隐式时间离散, 如**中心全隐格式**:

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{c(u_{j+1}^{n+1} - u_{j-1}^{n+1})}{2\Delta x} = \frac{a\delta_x^2 u_j^{n+1}}{(\Delta x)^2}$$

\* 无条件具有  $L^2$  模强稳定性, 结合数值黏性修正可获得更好效果。

\* 具有  $L^2$  模强稳定性未必满足最大模稳定性, 仍然可能发生数值振荡。

### 算子分裂



类似二维热传导方程的 LOD 方法, 先用 LW 格式离散前一半时间的纯对流问题  $\frac{1}{2}u_t + u_{xx} = 0$ , 再用全显格式离散后一半时间的纯扩散问题  $\frac{1}{2}u_t = u_{xx}$ , 得到

$$u_j^{n+1/2} = u_j^n - \frac{1}{2}\nu c(u_{j+1}^n - u_{j-1}^n) + \frac{1}{2}(\nu c)^2 \delta_x^2 u_j^n$$

$$u_j^{n+1} = u_j^{n+1/2} + \mu a \delta_x^2 u_j^{n+1/2}$$

\* 其  $L^2$  模强稳定性条件为  $\max(\nu|c|, 2\mu a) \leq 1$ 。

### 特征差分

考虑  $a = 0$  时的特征线方向, 定义时间全导数

$$\bar{D}_t u = \frac{1}{\sqrt{1+c^2}} u_t + \frac{c}{\sqrt{1+c^2}} u_x$$

则方程可以看作斜坐标系下纯扩散问题

$$\sqrt{1+c^2} \bar{D}_t u = a u_{xx}$$

设  $\tilde{x}_j = x_j - c\Delta t$  落在网格点  $x_{j,L}$  与  $x_{j,R} = x_{j,L} + \Delta x$  之间, 利用线性插值离散  $\bar{D}_t u$  得到

$$u_j^{n+1} = \tilde{u}_j^n + \mu a \delta_x^2 u_j^{n+1}, \quad \tilde{u}_j^n = \left(1 - \frac{\tilde{x}_j - x_{j,L}}{\Delta x}\right) u_{j,L}^n + \frac{\tilde{x}_j - x_{j,L}}{\Delta x} u_{j,R}^n$$

\* 无条件具有  $L^2$  模稳定性, 数值误差达到整体一阶, 界定常数与沿特征线方向的变化率相关。

## §9.2 修正方程与能量方法

### 修正方程方法

分为两步:

#### 1. 从差分方程出发导出含有网格参数的修正方程

设网格函数可由光滑函数  $w(x, t)$  限制而成, 逐点 Taylor 展开得到微分恒等式, 反复利用得到偏微分方程

$$w_t = \sum_{l=0}^m \alpha_l D_x^l w$$

#### 2. 利用修正方程的性质解释差分方程的数值表现

若修正方程不适定, 差分格式不稳定, 若适定则基本可以说明稳定。此外, 修正方程的色散、耗散与数值格式相近。

对一般的线性常系数偏微分方程  $w_t = \sum_{l=0}^m \alpha_l D_x^l w$ , Fourier 理论分析是高效的, 空间导数中奇数阶描述波动, 一阶代表对流, 更高阶代表色散; 偶数阶描述扩散或反扩散, 扩散时真解衰减, 方程适定, 反扩散 [如负系数二阶导数或正系数四阶导数] 则不适定。

\* 考虑对流方程  $u_t + u_x = 0$  的中心差商显格式

$$\frac{u_{j+1}^n - u_j^n}{\Delta t} + \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = 0$$

Taylor 展开得到

$$w_t + w_x + \frac{\Delta t}{2} w_{tt} + O((\Delta t)^2 + (\Delta x)^2) = 0$$

另一方面, 对  $x, t$  分别求导后代换  $t$  的二阶导数有修正方程

$$w_t + w_x = -\frac{\Delta t}{2} w_{xx}$$

扩散系数为负，因此不适定。

\* 对迎风格式

$$\frac{u_{j+1}^n - u_j^n}{\Delta t} + \frac{u_j^n - u_{j-1}^n}{\Delta x} = 0$$

Taylor 展开得到

$$w_t + w_x + \frac{\Delta t}{2} w_{tt} - \frac{\Delta x}{2} w_{xx} + O((\Delta t)^2 + (\Delta x)^2) = 0$$

求导后进一步代换得到

$$w_t + w_x = \frac{\Delta x - \Delta t}{2} w_{tt}$$

当  $\nu < 1$  时适定， $\nu > 1$  时不适定。

\* 过程中 Taylor 展开需要技巧性与针对性，双曲型方程差分格式中低频简谐波更重要，且其光滑性好，因此可用此类方法，抛物型数值不稳定则主要来源于高频，光滑性差，无法 Taylor 展开。

\* 考虑热传导方程  $u_t = u_{xx}$  的全显格式

$$\frac{u_{j+1}^n - u_j^n}{\Delta t} = \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{(\Delta x)^2}$$

分裂为高低频成分  $u_j^n = v_j^n + (-1)^{j+n} \phi_j^n$ ，两者均满足全显格式。代入高频成分成立的差分方程可得到修正方程

$$\phi_t = \frac{2(2\mu - 1)}{\Delta t} \phi$$

直接利用 ODE 知识可知  $\mu > \frac{1}{2}$  时解不稳定。

### 能量方法

对网格函数  $u = \{u_j\}_{j=0}^J, v = \{v_j\}_{j=0}^J$ ，可定义各种类型的离散内积与诱导范数：

$$\langle u, v \rangle = \sum_{j=1}^{J-1} u_j v_j \Delta x, \quad \|u\|_2 = \langle u, v \rangle^{1/2}$$

$$\langle u, v \rangle = \sum_{j=1}^J u_j v_j \Delta x, \quad \|u\|_2 = \langle u, v \rangle^{1/2}$$

$$[u, v] = \sum_{j=0}^{J-1} u_j v_j \Delta x, \quad \|u\|_2 = [u, v]^{1/2}$$

$$[u, v] = \sum_{j=0}^J u_j v_j \Delta x, \quad \|u\|_2 = [u, v]^{1/2}$$

均可以看作  $L^2$  范数的离散表示，平行于分部积分公式即可得到分部求和公式

$$\langle \Delta u, v \rangle = -\langle u, \Delta_- v \rangle + u_J v_J - u_1 v_0$$

从而有 Green 公式离散版本

$$\langle \Delta(a\Delta_- u), v \rangle = -\langle a\Delta_- u, \Delta_- v \rangle + a_J \Delta_- u_J v_J - a_1 \Delta_- u_1 v_0$$

$$\langle \Delta(a\Delta_- u), v \rangle - \langle \Delta(a\Delta_- v), u \rangle = a_J(v\Delta_- u_J - u\Delta_- v_J) - a_1(v\Delta u_0 - \Delta v_0)$$

### 常用不等式

$\varepsilon$ - $ab$  不等式

$$|ab| \leq \varepsilon a^2 + \frac{1}{4\varepsilon} b^2, \quad \varepsilon > 0$$

Cauchy 不等式

$$\langle u, v \rangle^2 \leq \|u\|_2 \|v\|_2$$

离散范数控制关系 [右端为 Poincaré 不等式的离散描述]

$$\frac{1}{2} \|[\Delta u]\|_2 \leq \|u\|_2 \leq \frac{L}{2\sqrt{2}\Delta x} \|[\Delta u]\|_2^2$$

Gronwall 不等式:  $f_n, n \geq 0$  与  $g_n, n \geq 0$  为非负序列, 且  $g_n$  单调增, 若存在给定正数  $C$  使得

$$f_{n+1} \leq C \sum_{m=0}^n f_m \Delta t + g_{n+1}, \quad \forall n \geq 0$$

则  $\Delta t$  充分小时有

$$f_n \leq e^{Cn\Delta t} g_n$$

\* 若  $a(x, t) > 0$  且  $a_x(x, t)$  有界, 考虑纯初值问题的迎风格式

$$u_j^{n+1} = \nu a_j^n u_{j-1}^n + (1 - \nu a_j^n) u_j^n$$

记  $A = \max_{x,t} a(x, t), B = \max_{x,t} |a_x(x, t)|$ , CFL 条件为  $A\nu \leq 1$ , 此时右端系数非负, 两边平方后由 Jessen 不等式可得

$$(u_j^{n+1})^2 \leq \nu a_j^n (u_{j-1}^n)^2 + (1 - \nu a_j^n) (u_j^n)^2$$

由于  $|a_j^n - a_{j-1}^n| \leq B\Delta x$ , 进一步放缩为

$$(u_j^{n+1})^2 \leq \nu a_{j-1}^n (u_{j-1}^n)^2 + (1 - \nu a_j^n) (u_j^n)^2 + B(u_{j-1}^n)^2 \Delta t$$

所有网格点上求和化简得到

$$\|u^{n+1}\|^2 \leq (1 + B\Delta t) \|u^n\|^2$$

满足  $L^2$  模稳定性, 当  $T$  为终止时刻时有  $\|u^n\|^2 \leq e^{BT} \|u^0\|^2$ 。

能量方法也可用于数值边界条件的设置, 以确保  $L^2$  模稳定性。考虑对流方程初边值问题

$$u_t + u_x = 0, \quad u(x, 0) = u_0(x), u(0, t) = 1$$

计算得真解满足

$$\int_0^1 u^2(x, t) dx \leq \int_0^1 u^2(x, 0) dx$$

我们希望数值解也有相同性质。

考虑蛙跳格式

$$u_j^{n+1} = u_j^{n-1} - \nu(u_{j+1}^n - u_{j-1}^n), \quad j = 1 : (J-1)$$

入流条件  $u_0^n = 0$ , 下面建立出流条件。

蛙跳格式两端同乘  $u_j^{n+1} + u_j^{n-1}$  后叠加, 化简得到 [内积上的撇表示剔除出流边界点信息, 合并到  $\Pi$  中]

$$\|u^{n+1}\|^2 - \|u^{n-1}\|^2 = -\nu \langle u^{n+1}, \Delta_{0x} u^n \rangle' + \nu \langle u^n, \Delta_{0x} u^{n-1} \rangle' + \Pi$$

$$\Pi = -\Delta t (u_{J-1}^{n-1} + u_{J-1}^{n+1}) u_J^n$$

定义人工出流边界条件  $u_J^n = \frac{1}{2}(u_{J-1}^{n-1} + u_{J+1}^{n-1})$  使  $\Pi \leq 0$ , 则记

$$S^n = \|u^{n+1}\|^2 + \|u^n\|^2 + \nu \langle u^{n+1}, \Delta_{0x} u_J^n \rangle'$$

有  $S^n$  不增, 另一方面

$$S^n \geq (1 - \nu)(\|u^{n+1}\|^2 + \|u^n\|^2)$$

于是即有  $\nu < 1$  时的  $L^2$  模稳定性。

联立蛙跳格式求解可知人工出流边界条件等价于

$$u_J^n = \frac{2}{2+\nu} u_{J-1}^{n-1} + \frac{\nu}{2+\nu} u_{J-2}^n$$

## 十 椭圆型方程

与之前发展型方程不同, 本章介绍椭圆型方程的差分, 考虑二维 Poisson 方程

$$-\Delta u = f(x, y), \quad (x, y) \in \Omega$$

这里  $f$  已知,  $\Omega$  是边界逐段光滑的二维有界区域。

\* 此问题与二维扩散方程  $u_t = \Delta u + f(x, y)$  密切相关, 其解即为扩散方程稳态解。

### §10.1 五点格式

回顾第五章对一般区域的分析, 考虑  $\Delta x = \Delta y = h$  的离散网格, 将离散网格分为内点  $\Omega_h$  与边界点  $\Gamma_h$  两部分, 并希望在网格内点上建立相应的差分方程。简便起见, 我们记离散中心点为  $c$ , 上下左右分别为  $n$ 、 $s$ 、 $w$ 、 $e$ 。

直接利用二阶中心差商离散可得**五点差分方程**:

$$\frac{1}{h^2}(4u_c - u_e - u_n - u_s - u_w) = f_c$$

以其为主体的差分格式称为**五点差分格式**。

\* 代入真解后两端的差距就是局部截断误差, 这里误差即为  $O(h^2)$ , 二阶相容。

\* 稳定性概念描述与定解条件的连续依赖关系, 收敛性描述与真解的逼近程度, 二者仍满足 Lax-Richtmyer 等价定理。

下面回顾边界条件的离散。若离散焦点  $c$  是紧邻 Dirichlet 边界的非规则内点, 则可能需要构造**非等臂长差分方程**, 设

$$ce = s_1 h, \quad cn = s_2 h, \quad cw = s_3 h, \quad cs = s_4 h$$

$$\beta_e = \frac{2}{s_1(s_1 + s_2)}, \quad \beta_s = \frac{2}{s_2(s_2 + s_4)}, \quad \beta_w = \frac{2}{s_3(s_1 + s_3)}, \quad \beta_n = \frac{2}{s_4(s_2 + s_4)}, \quad \beta_c = \beta_e + \beta_s + \beta_w + \beta_n$$

则可构造

$$\frac{1}{h^2}(\beta_c u_c - \beta_e u_e - \beta_n u_n - \beta_s u_s - \beta_w u_w) = f_c$$

\* 也可对  $\beta$  近似使得求解的方程组对称, 与第五章相同, 局部误差结论亦相同。

五点格式的差分事实上会形成一个大規模线性方程组, 例如  $\Omega = (0, 1) \times (0, 1)$  下考虑 Dirichlet 问题, 边值恒为 0, 记  $h = \frac{1}{J}$ , 构造分块三对角矩阵 [刚度矩阵]

$$\mathbb{A}_h = \frac{1}{h^2} \text{tridiag}(-\mathbb{I}_h, 2\mathbb{I}_h + \mathbb{C}_h, -\mathbb{I}_h), \quad \mathbb{C}_h = \text{tridiag}(-1, 2, -1)$$

这里  $\mathbb{C}_h, \mathbb{I}_h$  为  $J-1$  阶方阵,  $\mathbb{A}_h$  为  $(J-1)^2$  阶方阵。

\* 利用矩阵张量积的定义, 也可写为  $\mathbb{A}_h = \frac{1}{h^2}(\mathbb{C}_h \otimes \mathbb{I}_h + \mathbb{I}_h \otimes \mathbb{C}_h)$ 。

考虑先行后列、从左到右、从上到下将  $u$  展平为向量  $\mathbf{u}_h$ ,  $\mathbf{f}_h$  同理, 称为**荷载向量**, 则线性方程组可写为

$$\mathbb{A}_h \mathbf{u}_h = \mathbf{f}_h$$

\* 利用线性代数知识,  $\mathbb{A}_h$  的特征值为所有  $\lambda_p + \lambda_q$ ,  $p, q$  分别从 1 取到  $J-1$ , 其中  $\lambda_i = \frac{4}{h^2} \sin^2 \frac{sh\pi}{2}$ 。于是刚度矩阵谱条件数为  $\cot^2 \frac{\pi h}{2}$ , 为  $O(h^{-2})$  量级, 意味着网格越小, 计算规模越大, 方程组也越病态, 因此五点格式主要难点在于方程组求解效率。

### 数值求解

\* 这里不介绍纯代数的迭代方法, 以利用微分方程数值离散求解的方法为重点。

**交替方向法**: 由于 Poisson 方程的差分格式可以看作二维扩散方程的半离散格式的稳态解计算过程, 二维扩散方程的迭代即可看作迭代求解方法。为提升效率, 可采取 ADI、LOD 等经济型格式。

例如, 基于  $u_t = u_{xx} + u_{yy} + f$  的 PR 格式, 上述方程组的迭代可以定义为

$$\begin{aligned}\mathbf{u}_h^{k+1/2} &= \mathbf{u}_h^k - \tau_k (\mathbb{L}_1 \mathbf{u}_h^{k+1/2} + \mathbb{L}_2 \mathbf{u}_h^k - \mathbf{f}_h) \\ \mathbf{u}_h^{k+1} &= \mathbf{u}_h^k - \tau_k (\mathbb{L}_1 \mathbf{u}_h^{k+1/2} + \mathbb{L}_2 \mathbf{u}_h^{k+1} - \mathbf{f}_h)\end{aligned}$$

这里  $\mathbb{L}_1 = \frac{1}{h^2} \mathbb{C}_h \otimes \mathbb{I}_h$ ,  $\mathbb{L}_2 = \frac{1}{h^2} \mathbb{I}_h \otimes \mathbb{C}_h$ ,  $\tau_k$  称为虚拟时间步长, 代表迭代步长。

\* 注意到  $\mathbb{L}_1 + \mathbb{L}_2 = \mathbb{A}_h$ , 迭代若收敛则能收敛到真解。实质上的迭代矩阵为

$$\mathbb{T}_k = \tau_k^2 (\mathbb{I}_h + \tau_k \mathbb{L}_2)^{-1} \mathbb{L}_1 (\mathbb{I}_h + \tau_k \mathbb{L}_1)^{-1} \mathbb{L}_2$$

若  $\tau_k$  恒定为  $\tau$ , 取最优值  $\frac{1}{2 \sin(\pi h)}$ ,  $\mathbb{T}$  也恒定, 且具有最小谱半径  $\left(\frac{1 - \tan(\pi h/2)}{1 + \tan(\pi h/2)}\right)^2$ , 收敛速度接近带有最佳因子的超松弛迭代。

**多重网格法:** 利用 Fourier 理论分析可发现高频简谐波迭代误差衰减快, 而低频缓慢, 由于细网格上的低频简谐波可视为粗网格上的高频波, 于是先粗后细的迭代可能有较好效果。考虑  $\Delta x = \Delta y = 2h$  的粗网格与其加密后的  $\Delta x = \Delta y = h$  的细网格, 二重网格的一步迭代为:

1. 以  $\mathbf{u}_h^k$  为起点, 在细网格进行  $m$  次简单迭代 [如 Jacobi 或 GS 迭代], 得到  $\bar{\mathbf{u}}_h^k$ ;

2. 计算残量  $\mathbf{r}_h^k = \mathbf{f}_h - \mathbb{A}_h \bar{\mathbf{u}}_h^k$ ;

3. 将残量限制在粗网格  $\mathbf{r}_{2h}^k = \mathbb{I}_{2h}^k \mathbf{r}_h^k$ , 这里  $\mathbb{I}_{2h}^k$  为**限制算子**, 定义为加权平均  $\frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$ , 也即每个

粗网格点用周围九个细网格点平均得到;

4. 粗网格上求解  $\mathbb{A}_{2h} \mathbf{e}_{2h}^k = -\mathbf{r}_{2h}^k$ ;

5. 延拓到细网格  $\bar{\mathbf{e}}_h^k = \mathbb{I}_{2h}^h \mathbf{e}_{2h}^k$ , 这里  $\mathbb{I}_{2h}^h$  为**延拓算子**, 定义为加权  $\frac{1}{4} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$ , 也即每个粗网格点按

此权重分配到周围九个细网格点;

6. **校正**  $\mathbf{u}_h^{k+1} = \bar{\mathbf{u}}_h^k - \bar{\mathbf{e}}_h^k$ 。

\* 要使精度达到要求, 迭代步数与计算规模或空间步长  $h$  无关, 计算复杂度为  $O(J^2 \ln J)$ , 几乎是最优的线性方程组计算效率。

\* 可以利用更粗的网格进行校正、回溯, 由此从二重网格成为多重网格方法。

\* 限制算子与延拓算子定义不唯一。

**区域分解方法:** 基本思想为将大规模问题划分为小规模问题, 这里仅介绍重叠型 Schwarz 方法。

分解  $\Omega = \Omega_1 \cup \Omega_2$ , 且交集测度大于 0, 假设  $\Omega_\kappa$  的内部边界  $\gamma_\kappa = \partial\Omega_\kappa \cap \omega$  都在网格线上, 则迭代解计算方式为:

基于整体网格  $\bar{\Omega}_h$  在  $\bar{\Omega}_1$  限制成的子网格, 利用五点差分格式计算  $\Omega_1$  内的二维 Poisson 方程, 边值条件为

$$u|_{\gamma_1} = \mathbf{u}_h^k, \quad u|_{\partial\Omega \cap \partial\Omega_1} = 0$$

这里  $\mathbf{u}_h^k$  为猜测初值或已知迭代解, 第二条对应原本的 Dirichlet 零边值条件, 将计算结果称为  $\tilde{\mathbf{u}}_h^{k+1}$ 。

基于整体网格  $\bar{\Omega}_h$  在  $\bar{\Omega}_2$  限制成的子网格, 利用五点差分格式计算  $\Omega_2$  内的二维 Poisson 方程, 边值条件为

$$u|_{\gamma_1} = \tilde{\mathbf{u}}_h^{k+1}, \quad u|_{\partial\Omega \cap \partial\Omega_2} = 0$$

将计算结果称为  $\mathbf{u}_h^{k+1}$ 。

\* 若  $\tilde{\mathbf{u}}_h^{k+1}$  与  $\mathbf{u}_h^{k+1}$  充分接近, 迭代可以停止, 可证明收敛性, 且重叠面积越大收敛越快。

\* 这里的计算流程称为**异步并行**, 边界信息交换称为 Dirichlet-Dirichlet 方式, 也可采用其他策略。

## §10.2 最大模估计

\* 由于  $L^2$  模稳定性与误差估计可通过能量方法或直接矩阵方法给出, 这里主要分析最大模稳定性。为行文简便, 用单下标  $l$  标注离散点, 对应数值解为  $u_l$ 。

仍考虑内点集和边界点集, 给定 Dirichlet 边界条件, 记为函数  $g$ 。若  $u$  满足差分格式 [记  $\mathcal{D}_h u_j = d_{jj} u_j - \sum_{k \in O(j)} d_{jk} u_k$ , 这里  $O(j)$  表示其空心邻域, 所有  $d$  为给定正数]

$$\mathcal{D}_h u_j = f_j, \quad j \in \Omega_h$$

$$u_j = g_j, \quad j \in \Gamma_h$$

若  $d_{jj} \geq \sum_{k \in O(j)} d_{jk}$  处处成立, 则称  $\mathcal{D}_h$  为椭圆型差分算子, 对应椭圆型差分格式。下面默认其为椭圆型差分算子。

\* 不妨假设  $\bar{\Omega}_h = \Omega_h \cup \Gamma_h$  是连通的, 即对任两网格点存在一个网格内点形成的路径连接, 使得每点在前一点的空间邻域内。

**强最大值原理:** 若  $\bar{\Omega}_h$  上的  $u$  不恒为常值, 且  $\mathcal{D}_h u_j \leq 0$  处处成立, 则  $u$  不可能在  $\Omega_h$  上取到正的最大值。

\* 由椭圆型差分格式的定义, 一点处取到正的最大值可以得到其周围亦取到, 从而由连通可得所有网格点取到。

于是可以引入**优函数**: 若  $|f_j| \leq F_j, |g_j| \leq G_j$  在定义域上成立, 则称根据

$$\mathcal{D}_h U_j = F_j, \quad j \in \Omega_h$$

$$U_j = G_j, \quad j \in \Gamma_h$$

解出的  $U_j$  为  $u_j$  的优函数, 作差并对原方程取相反数作差, 利用强最大值原理可知  $|u_j| \leq U_j$  处处成立。

### 简单估计

假设差分方程均具有  $\frac{1}{h^2}(\beta_c u_c - \beta_e u_e - \beta_n u_n - \beta_s u_s - \beta_w u_w) = f_c$  形式, 规则内点的五点差分等臂长, 非规则内点可能不等臂长。

椭圆型差分格式**必然最大模稳定**, 即存在  $C$  使得

$$\|u_h\|_{\bar{\Omega}_h, \infty} \leq C(\|f\|_{\Omega_h, \infty} + \|g\|_{\Gamma_h, \infty})$$

证明: 假设计算中心包含在某圆中, 中心  $(x_*, y_*)$ , 半径  $\rho$ , 定义

$$U(x, y) = K(\rho^2 - (x - x_*)^2 - (y - y_*)^2)$$

计算得  $\mathcal{D}_h U_j = \theta K$ , 其中规则内点  $\theta = 4$ , 非规则内点  $\theta > 2$ , 取  $K = \frac{1}{2} \max_{\Omega_h} |f_j|$ , 则可保证  $U$  是对应区域 Dirichlet 零边值问题的解  $u^{(1)}$  的优函数。

另一方面, 考虑  $f$  恒为 0 时, 设其解为  $u^{(2)}$ , 考虑常值函数  $V_j = \max_{\Gamma_h} |g_j|$ , 其是  $u^{(2)}$  的优函数, 因此由  $u = u^{(1)} + u^{(2)}$  利用三角不等式可得

$$\|u_h\|_{\bar{\Omega}_h, \infty} \leq \frac{1}{2} \rho^2 \|f\|_{\Omega_h, \infty} + \|g\|_{\Gamma_h, \infty}$$

从而取  $C$  为较大的系数即得证。

\* 由于数值误差  $e_j$  只在内部出现, 最大模误差即为  $\mathcal{D}_h e_j = \tau_j$  的 Dirichlet 零边值问题, 对矩形边界, 由于五点差分格式有  $\tau_j = O((\Delta x)^2)$ , 利用最大模稳定即知  $\|e\|_{\bar{\Omega}_h, \infty}$  亦为  $O((\Delta x)^2)$ 。

### 精细估计

下面证明, 对一般的区域, 最大模误差**仍能达到二阶**。

设规则内点集  $\Omega_h^*$ , 非规则内点集  $\Omega_h^{**}$ , 进一步将  $u^{(1)}$  分裂为  $u^* + u^{**}$ , 其满足

$$\begin{cases} \mathcal{D}_h u_j^* = f_j, & \mathcal{D}_h u_j^{**} = 0 & j \in \Omega_h^* \\ \mathcal{D}_h u_j^* = 0, & \mathcal{D}_h u_j^{**} = f_j & j \in \Omega_h^{**} \\ u_j^* = 0, & u_j^{**} = 0 & j \in \Gamma_h^* \end{cases}$$

与之前利用相同的优函数可知存在  $C_1$  使得  $\|u^*\|_{\bar{\Omega}, \infty} \leq C_1 \|f\|_{\Omega_h^*, \infty}$ , 接下来要处理  $u^{**}$ 。仿照强最大值原理的证明, 可证明  $u^{**}$  不是常值函数, 则正最大值与负最小值均不会出现在  $\Omega_h^*$  上。

注意到, 由于非规则内点会涉及边界点, 记

$$\tilde{d}_{jj} = d_{jj} - \sum_{k \in O(j), k \notin \Gamma_h} d_{jk}$$

则由于所有  $d$  大于 0,  $\tilde{d}_{jj}$  一定大于 0, 根据差分格式设计方式可找到与  $h$  无关的  $C_2$  使得  $\tilde{d}_{jj} \geq \frac{C_2}{h^2}$ 。

将  $u^{**}$  在  $\Omega_h^{**}$  的差分方程写为

$$\tilde{d}_{jj} u_j + \sum_{k \in O(j), k \notin \Gamma_h} d_{jk} (u_j - u_k) = f_j$$

若正最大值或负最小值在某个非规则内点取到, 考虑相应位置的差分方程

$$\|u^{**}\|_{\Omega^{**}, \infty} \leq \frac{1}{C_2} h^2 \|f\|_{\Omega^{**}, \infty}$$

于是更精细的最大模稳定结论为存在  $C$  使得

$$\|u_h\|_{\bar{\Omega}_h, \infty} \leq C (\|f\|_{\Omega^*, \infty} + h^2 \|f\|_{\Omega^{**}, \infty} + \|g\|_{\Gamma_h, \infty})$$

于是, 即使 Dirichlet 边界数值处理有  $O(1)$  误差, 整体误差仍然能达到二阶。

\* 类似地, 适当时空约束下, 二维扩散方程 Dirichlet 问题的偏隐格式也对任何区域具有最优的最大模误差估计, 例如对全隐格式误差为  $O(h^2 + \Delta t)$ 。

\* 若带有自然边界条件, 需要利用能量方法与不同范数等价关系得到最大模误差估计。

### §10.3 提升数值精度

#### Richardson 外推

在正方形区域, 考虑辅助方程  $-\Delta w = [u_{xxxx}] + [u_{yyyy}]$  的 Dirichlet 零边值问题, 解记为  $[w]$ 。对数值解泰勒展开可得到

$$u_{jk} = [u]_{jk} + \frac{1}{12} [w]_{jk} + O(h^4)$$

用多重网格思路, 考虑粗网格  $\bar{\Omega}_{2h}$  与加密得到的细网格  $\bar{\Omega}_h$ , 得到数值解  $u^{2h}$  与  $u^h$ , 作粗网格上的外推

$$\tilde{u}^{2h} = \frac{4}{3} u^h - \frac{1}{3} u^{2h}$$

则其为粗网格真解的四阶逼近。

\* 对本质边界条件, 外推较为有效, 对自然边界条件则表现不够理想。

#### 九点格式

设  $u_c$  右上、左上、右下、左下分别为  $u_{ne}, u_{nw}, u_{se}, u_{sw}$ , 记斜五点差分

$$\mathcal{S}_h u_c = \frac{1}{2h^2} (4u_c - u_{ne} - u_{nw} - u_{se} - u_{sw})$$

而之前的正五点差分记作

$$\mathcal{L}_h u_c = \frac{1}{h^2} (4u_c - u_e - u_n - u_s - u_w)$$

四阶九点格式定义为

$$\left(\frac{2}{3}\mathcal{L}_h + \frac{1}{3}\mathcal{S}_h\right)u_c = f_c - \frac{1}{12}\mathcal{L}_h f_c$$

六阶九点格式定义为

$$\left(\frac{2}{3}\mathcal{L}_h + \frac{1}{3}\mathcal{S}_h\right)u_c = f_c - \frac{1}{12}\mathcal{L}_h f_c - \frac{h^4}{240}(\delta_x^4 + \delta_y^4)f_c + \frac{h^4}{90}\delta_x^2\delta_y^2 f_c$$

\* 可直接泰勒展开得到误差阶数。

对非正方形网格，四阶九点格式可定义为

$$-\left(\frac{\delta_x^2}{(\Delta x)^2} + \frac{\delta_y^2}{(\Delta y)^2} + \frac{(\Delta x)^2 + (\Delta y)^2}{12(\Delta x)^2(\Delta y)^2}\delta_x^2\delta_y^2\right)u_c = \left(1 + \frac{\delta_x^2 + \delta_y^2}{12}\right)f_c$$

\* 为确保其为椭圆型差分格式，须  $\frac{\Delta x}{\Delta y} \in [\frac{1}{\sqrt{5}}, \sqrt{5}]$ 。

\* 离散模板扩张可以提升相容阶，但导致系数矩阵更稠密，数值求解更困难。

### Kreiss 差分格式

回顾第二章的导数离散方式，可得到二阶导数的四阶逼近

$$D^2 = \frac{1}{h^2} \frac{\delta^2}{\mathbb{I} + \delta^2/12} + O(h^4)$$

引进网格函数  $u$  逼近  $[u]$ ， $p$  逼近  $[u_{xx}]$ ， $q$  逼近  $[u_{yy}]$ ，利用上式离散空间导数可得到

$$\begin{aligned} \frac{1}{12}p_{j+1,k} + \frac{5}{6}p_{jk} + \frac{1}{12}p_{j-1,k} &= \frac{1}{h^2}\delta_x^2 u_{jk} \\ \frac{1}{12}q_{j,k+1} + \frac{5}{6}q_{jk} + \frac{1}{12}q_{j,k-1} &= \frac{1}{h^2}\delta_y^2 u_{jk} \end{aligned}$$

再结合方程  $-p_{jk} - q_{jk} = f_{jk}$ ，即可联立得到未知量为网格点数三倍的方程组。

其可迭代求解， $u^n$  到  $u^{n+1}$  的迭代方法为：

1. 利用空间导数的离散逐行分别求解得到  $p^n, q^n$ ，每行对应的方程组都三对角，乘除法次数同未知数正比；
2. 利用  $\frac{u_{jk}^{n+1} - u_{jk}^n}{\tau^n} - p_{jk}^n - q_{jk}^n = f_{jk}$  迭代得到  $u^{n+1}$ ， $\tau^n$  称为**虚拟时间步长**， $u^{n+1}$  与  $u^n$  差距足够小时可停止迭代。

\* **有限元方法**对形状复杂且边界条件出现导数时的情况更为灵活，展现出数值优势。