Numerical Solutions of Partial Differential Equations

Chapter 1: Finite difference methods for elliptic PDEs

Huazhong Tang

School of Mathematical Sciences, Peking University, P.R. China

September 18, 2020

前沿

偏微分方程(PDE)是一个包含一个以上自变量的未知函数(应变量)及其某些偏导数的方程. PDE的阶是PDE中出现的未知量/应变量的偏导数的最高阶数. 偏微分方程组是由几个方程组成, 其中包含一个未知的向量值函数u及其偏导数. 如果PDE对应于作用于未知量及其偏导数的线性算子, 则PDE是线性的; 否则是非线性的. 文献中还有半线性(最高阶导数项是线性的), 拟线性(最高阶导数项前系数依赖于低阶导数), 完全非线性的细分[1, P2].

注意, ODE的解包含任意常数, PDE的解包含任意函数. 同样，m阶的ODE有m个线性无关解, 而一个PDE有无穷多个(解中有任意函数!). 这是两个变量函数比只有一个变量的函数包含的信息量大得多的事实的结果[1].

导数的有限差分近似是求解微分方程的最简单和最老的方法之一. 它(一维空间情形)已经被欧拉(1707-1783)在1768年左右所知, 被龙格(1856-1927)在1908年左右推广到了二维. 有限差分技术在数值应用中的出现始于20世纪50年代初, 计算机的出现促进了其发展, 为处理复杂的科学技术问题提供了一个方便的框架. 在过去的几十年中, 关于PDE的有限差分方法的精度、稳定性和收敛性的理论研究已经取得了一些成果[2]. 有限差分法: FDM中的未知量为微分方程定解问题的解的离散点值/网格点值。差商近似微商；待定系数(Taylor展开确定)；插值法(插值出多项式后再求导)；积分法(后来的有限体积做法).

有限元方法是以PDE的变分形式为基础的. 对于PDE对应的变分问题, 有经典的Ritz和Galerkin 近似求解方法. 在Ritz法或者Galerkin法中, 取有限维空间为有限元空间就得到了FEM. 把变分原理作为有限元法的基础并不是一个过分的说法. 有限元方法包括连续有限元方法、不连续有限元方法(间断Galerkin 方法)；协调FE，非协调FE，混合(mixed, hybrid) FE等. 在FEM函数空间(有限维的具有局部紧致性的函数空间)中近似微分方程定解问题的解，FEM方程中的未知量是自由度或基函数展开系数.

有限体积方法: FVM方程中的未知量为微分方程定解问题的解的单元(控制体)平均值. 包括格点FVM和格心FVM等.

谱方法思想是将微分方程的解写成某些“基函数”的和(例如傅里叶级数和), 然后在和式中选择系数, 以便尽可能满足微分方程. 谱方法和有限元方法有着密切的联系和共同的思想, 它们的主要区别在于谱方法在整个域上使用非零的基函数, 而有限元方法在小的子域上使用非零的基函数. 换句话说, 谱方法采用全局方法, 而有限元方法采用局部方法. 部分由于这个原因, 谱方法具有极好的误差特性, 当解光滑时, 它具有可能的“指数收敛”性. 谱(spectral)方法: 包括tau方法(1938)[3]，拟谱方法/配置谱法(1970, 解在配置点处是精确的)，Galerkin 型方法. 与FEM类似，利用正交函数逼近微分方程定解问题的解，但是不同于FEM的是，是在整个区域中逼近.

无网格方法：在区域内撒一些散乱的点, 对这些点不连线, 利用这些点处的近似值和最小二乘等技术近似导数等.

边界元方法：把区域内的边值问题化为区域边界上的积分方程, 离散区域边界及该积分方程.

和有限元法一样, 有限差分法是基于函数的局部表示, 通常采用低阶多项式. 相比之下, 谱方法通常使用高阶多项式或傅立叶级数的全局表示. 在幸运的情况下, 结果是局部方法无法匹配的精度. 对于大型计算, 特别是在几个空间维度中, 这种更高的精度对于允许更粗的网格, 从而存储和操作的数据值的数量更少可能是最重要的.

[1] Viktor Grigoryan, Partial differential equations, Math 124A-Fall 2010, http://web.math.ucsb.edu/~grigoryan/124A.pdf

[2] https://www.ljll.math.upmc.fr/frey/cours/UdC/ma691/ma691_ch6.pdf

[3] tau方法：最初作为数学物理中特殊函数近似的一种方法，可以用简单的微分方程来表示. 它已发展成为求解复杂微分方程和泛函方程的一个强大而精确的工具.

Syllabus[教学大纲]

- FDM for elliptic PDEs (Chapter 1),
- FDM for parabolic PDEs (Chapter 2),
- FDM for hyperbolic PDEs (Chapter 3),
- Revisiting consistency, stability, and convergence for FDM (Chapter 4),
- Variational form for elliptic BVP (Chapter 5),
- FEM for elliptic PDEs (Chapter 6),
- Error estimates for FEM for elliptic PDEs (Chapter 7),
- Error control and adaptive FEM (Chapter 8).

Textbook: [5]

Prerequisites: A reasonable background in calculus, linear algebra, function of real variable, PDE, function analysis etc. Some programming experiences are helpful.

Grading: Homework (analytic part and computer projects) 40%+ Class participation 10% + Final 50%.

Contents

1 Introduction

在构造PDE的数值方法时, 了解PDE的类型和特性是重要的.

- The order of a PDE is the order of the highest derivative entering the equation.
- Linearity means that all instances of the unknown and its derivatives enter the equation linearly.

椭圆算子是推广拉普拉斯算子的微分算子. 它们是由PDE中最高阶导数的系数为正的条件定义的, 这意味着主符号是可逆的, 或者等价于没有实的特征方向.

椭圆算子是位势理论的典型代表, 在静电学和连续介质力学中经常出现. 椭圆正则性意味着它们的解趋向于光滑函数(如果算子中的系数是光滑的).

椭圆型偏微分方程是一类重要的PDE. 早在1900年D. 希尔伯特提的著名的23个问题中, 就有三个问题是关于椭圆型方程与变分法的. 八十多年来，椭圆型方程的研究获得了丰硕的成果. 椭圆型方程在流体力学、弹性力学、电磁学、几何学和变分法中都有应用. Laplace方程, Poisson方程, Helmholtz方程是椭圆型PDE最典型的特例。

Definition 1.1 *A 2nd order linear PDE with n independent variables* $\boldsymbol{x} = (x_1, ..., x_n)$

$$\pm\mathcal{L}(u) \triangleq \pm\left(\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i \frac{\partial}{\partial x_i} + c\right) u = f, \tag{1.1}$$

$$a_{ij} = a_{ij}(\boldsymbol{x}),\ b_i = b_i(\boldsymbol{x}),\ c = c(\boldsymbol{x}),\ f = f(\boldsymbol{x}),$$

or the operator $\mathcal{L}$ *in* (1.1) *is elliptic at the point* $\boldsymbol{x}_0 \in \Omega$*, if there exists* $\alpha(\boldsymbol{x}_0) > 0$ *such that*

$$\sum_{i,j=1}^{n} a_{ij}(\boldsymbol{x}_0)\xi_i\xi_j \geq \alpha(\boldsymbol{x}_0)\sum_{i=1}^{n}\xi_i^2,\ \forall\xi \in \mathbb{R}^n\backslash\{0\}. \tag{1.2}$$

If (1.1) *or* $\mathcal{L}$ *is elliptic at every point* $\boldsymbol{x} \in \Omega$*, then* (1.1) *or* $\mathcal{L}$ *is elliptic in* Ω. ■

Remark 1.1 (1.2) *implies that* $A := (a_{ij}(\boldsymbol{x}))$ *is positive definite.* ■

Definition 1.2 *The equation* (1.1) *or operator* $\mathcal{L}$ *is uniformly elliptic in* Ω*, if* (1.2) *holds with*

$$\inf_{\boldsymbol{x}\in\Omega} \alpha(\boldsymbol{x}) = \alpha_0 > 0, \tag{1.3}$$

i.e.

$$\sum_{i,j=1}^{n} a_{ij}(\boldsymbol{x})\xi_i\xi_j \geq \alpha_0 \sum_{i=1}^{n}\xi_i^2, \quad \forall\boldsymbol{\xi} \in \mathbb{R}^n\backslash\{0\}, \quad \forall\boldsymbol{x} \in \Omega. \tag{1.4}$$

■

Example 1.1 *Several 2nd-order, linear, uniformly elliptic PDEs:*

- $\Delta u = 0$. *Laplace equation.*
- $\Delta u + \Phi(\boldsymbol{x}) = 0$. *Poisson equation.*
- $\Delta u + \lambda u = -\Phi(\boldsymbol{x})$. *Helmholtz equation.*

■

Example 1.2 *The operator:* $\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$ *is is a linear 2nd order uniformly elliptic differential operator, since* $a_{ii} = 1$, $a_{ij} = 0$, $\forall i, j$, $i \neq j$. ▮

Example 1.3 *A 2nd-order PDE:* $u_{xx} + f(x)u_{yy} = 0$, *which is the generalized Tricomi equation.* ▮

Definition 1.3 (From WikiPedia) *A linear differential operator* L *of order* m *on a domain* Ω *in* $\mathbb{R}^n$ *given by*

$$Lu = \sum_{|\alpha|\leq m} a_\alpha(x)\partial^\alpha u$$

(where $\alpha = (\alpha_1, ..., \alpha_n)$ *is a multi-index, and* $\partial^\alpha u = \partial^{\alpha_1} \cdots \partial^{\alpha_n} u)$ *is called elliptic in* Ω *if for every* x *in* Ω *and every non-zero* ξ *in* $\mathbb{R}^n$,

$$\sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \neq 0,$$

where $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_n^{\alpha_n}$.

In many applications, this condition is not strong enough, and instead a uniform ellipticity condition may be imposed for operators of degree $m = 2k$:

$$(-1)^k \sum_{|\alpha|=2k} a_\alpha(x)\xi^\alpha > C|\xi|^{2k},$$

where C *is a positive constant. Note that ellipticity only depends on the highest-order terms*[4].

Remark 1.2 *The type depends only on the principal part of the differential operator. The type under coordinate transformations is invariant.* ▮

A nonlinear operator

$$L(u) = F(x, u, (\partial^\alpha u)_{|\alpha|\leq m})$$

is elliptic if its first-order Taylor expansion with respect to u *and its derivatives about any point is a linear elliptic operator.*

Example 1.4 (Poisson equation) *3D Poisson eq.*

$$\Delta\varphi = f, \quad x, y, z \in \mathbb{R},$$

has the solution[5]

$$\varphi(\mathbf{r}) = -\int_{\mathbb{R}} \frac{f(\mathbf{r}')}{4\pi|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}\mathbf{r}'.$$

▮

[4] Note that this is sometimes called strict ellipticity, with uniform ellipticity being used to mean that an upper bound exists on the symbol of the operator as well. It is important to check the definitions the author is using, as conventions may differ. See, e.g., Evans' book [1, Chapter 6], for a use of the 1st definition, and the book of Gilbarg and Trudinger [2, Chapter 3] for a use of the 2nd.

[5] http://eqworld.ipmnet.ru/en/solutions/lpde/lpde302.pdf

Example 1.5 椭圆型方程的简单例子是*Laplace*方程(或调和方程). 三维笛卡尔坐标下的*Laplace*方程可以表示为

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0, \quad (x, y, z) \in \Omega, \tag{1.5}$$

这里假设Ω是单连通区域. 调和方程的解通常被称为调和函数, 它的性质在复变函数论中有较详细的讨论. 调和函数$u(x, y, z)$有如下基本积分公式

$$u(x, y, z) = -\frac{1}{4\pi} \int_{\partial\Omega} \left[u(\widetilde{x}, \widetilde{y}, \widetilde{z}) \frac{\partial}{\partial n} \left(\frac{1}{r}\right) + \frac{1}{r} \frac{\partial u(\widetilde{x}, \widetilde{y}, \widetilde{z})}{\partial n} \right] ds,$$

其中$\frac{\partial}{\partial n}$和$ds$分别表示$\partial\Omega$的外法向梯度算子和面元, ds依赖于空间点$(\widetilde{x}, \widetilde{y}, \widetilde{z})$, 而

$$r := r(x, y, z, \widetilde{x}, \widetilde{y}, \widetilde{z}) = \sqrt{(x - \widetilde{x})^2 + (y - \widetilde{y})^2 + (z - \widetilde{z})^2}.$$

由此可见, 调和函数$u(x, y, z)$完全由其在区域边界$\partial\Omega$上的取值决定, 这也表明椭圆型方程的适定的定解问题只能是边值问题. 调和函数的一个重要性质是极值原理: 不恒等于常数的调和函数只能在区域Ω的边界上取到最大值和最小值. ∎

Example 1.6 二阶线性双曲型方程的典型例子是波动方程

$$\frac{\partial^2 u}{\partial t^2} - a^2 \frac{\partial^2 u}{\partial x^2} = 0, \ x \in \mathbb{R}, \ 0 < t < T, \tag{1.6}$$

这里假设a是有限的正常数, $\mathbb{R}$是实数集合. 如果给定$u(x, t)$和$u_t(x, t)$在$t = 0$时刻的值

$$u(x, 0) = \varphi_1(x), \ \frac{\partial u}{\partial t}(x, 0) = \varphi_2(x), \ x \in \mathbb{R}, \tag{1.7}$$

其中$\varphi_1 \in C^2(\mathbb{R})$, $\varphi_2 \in C^1(\mathbb{R})$, 则达朗贝尔*(D' Alembert)*公式

$$u(x, t) = \frac{1}{2} \left(\varphi_1(x + at) + \varphi_1(x - at)\right) + \frac{1}{2a} \int_{x-at}^{x+at} \varphi_2(s) \ ds, \tag{1.8}$$

给出的$u(x, t)$是上述波动方程初值问题的解, $u(x, t) \in C^2(\mathbb{R})$. 从表达式(1.8)可以看出, 初值问题的解$u(x, t)$在点$(x, t)$处的值仅仅取决于初始函数$\varphi_1$和$\varphi_2$在有限区间$[x - at, x + at]$内的所有取值. 另一方面, 表达式(1.8)也可以写为如下形式

$$u(x, t) = F^+(x - at) + F^-(x + at),$$

而$F^+(x - at)$和$F^-(x + at)$分别是初值问题

$$\frac{\partial v}{\partial t} \pm a \frac{\partial v}{\partial x} = 0, \quad v(x, 0) = F^{\pm}(x), \tag{1.9}$$

的解. 这说明, 初始信号(或波)$F^{\pm}(x)$经过某一有限时刻t_0后的形状不变, 仅仅是向右和左分别平移了一段有限的距离at_0, 而信号的传播速度是$\pm a$. 波或波动是扰动或物理信息在空间上传播的一种物理现象, 扰动的形式任意, 传播路径上的其它介质也作同一形式振动. 波的传播速度总是有限的. 在数学上, 任何一个沿某一方向运动的函数形状都可以认为是一个波. ∎

Example 1.7 考虑一维热传导方程的初值问题

$$\begin{aligned} &\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2}, \ x \in \mathbb{R}, \ 0 < t < T, \ a > 0, \\ &u(x, 0) = \varphi(x), \ x \in \mathbb{R}. \end{aligned} \tag{1.10}$$

线性热传导方程是最简单的抛物型方程. 如果$\varphi(x)$是有界连续函数, 则问题(1.10)存在唯一解

$$u(x,t) = \frac{1}{2\sqrt{\pi a t}} \int_{-\infty}^{\infty} \varphi(s) e^{-(x-s)^2/(4at)}\, ds.$$

由此可见, 无论$t > 0$多么小, $u(x,t)$都依赖于初始函数$\varphi(x)$在区域$\{x : x \in \mathbb{R}\}$内的所有取值, 依赖程度随着距离$|x-s|$的增大而减小. 这说明, 某一点的扰动会很快地传播到整个区域内部的各点处, 换言之, 抛物型方程的信号传播速度不是有限的, 这一点不同于前面的波动方程. 问题(1.10)的解$u(x,t)$是连续可微函数, 并满足极值原理: 如果初始函数$\varphi(x)$满足$c \le \varphi(x) \le C$, 其中c和C是两个有限常数, 则对任意$t \in [0,T)$和$x \in \mathbb{R}$, 均成立$c \le u(x,t) \le C$. ∎

Definition 1.4 *A linear elliptic PDEs of order $2m$ with n independent variables:*

$$\pm\mathcal{L}(u) \triangleq \pm \left(\sum_{k=1}^{2m} \sum_{i_1,\ldots,i_k=1}^{n} a_{i_1,\ldots,i_k}(\boldsymbol{x}) \frac{\partial^k}{\partial x_{i_1} \ldots \partial x_{i_k}} + a_0(\boldsymbol{x}) \right) u = f(\boldsymbol{x}), \tag{1.11}$$

is elliptic in Ω, if for every $\boldsymbol{x} \in \Omega$ there exists $\alpha(\boldsymbol{x}) > 0$ s.t.

$$\sum_{i_1,\ldots,i_{2m}=1}^{n} a_{i_1,\ldots,i_{2m}}(\boldsymbol{x}) \xi_{i_1} \cdots \xi_{i_{2m}} \ge \alpha(\boldsymbol{x}) \sum_{i=1}^{n} \xi_i^{2m}, \ \forall \boldsymbol{\xi} \in \mathbb{R}^n \backslash \{0\}. \tag{1.12}$$

∎

Remark 1.3 *Eq.(1.12) implies that the $2m$ order tensor $A = (a_{i_1,\ldots,i_{2m}})$ is positive definite.* 这个定义基本同文献的. *The readers are referred to Antonio Tarsia's lecture note.*

The elliptic differential operators $L(\partial)$ of order $2m$ in the Euclidean space $\mathbb{R}^n$ with constant real coefficients is defined in the paper[6].

The elliptic operator of order $2m$ satisfies the Legendre-Hadamard ellipticity condition, see Definition 2.1 of the file "history2015-SM.pdf". ∎

Remark 1.4 *A elliptic PDE of order m is defined in Encyclopedia of Mathematics or by characteristic surface* [7], *see F. John, Higher-order elliptic equations with constant coefficients, In: Partial Differential Equations. Applied Mathematical Sciences, vol 1. Springer, 1978.*

Also see Section 7.2 of this note. ∎

Definition 1.5 *The operator $\mathcal{L}$ or the equation (1.14) is uniformly elliptic in Ω, if (1.12) holds with $\inf_{\boldsymbol{x}\in\Omega} \alpha(\boldsymbol{x}) = \alpha_0 > 0$, that is, there exists a positive constant $\alpha_0 > 0$ for every $\boldsymbol{x} \in \Omega$ s.t.*

$$\sum_{i_1,\ldots,i_{2m}=1}^{n} a_{i_1,\ldots,i_2}(\boldsymbol{x}) \xi_{i_1} \cdots \xi_{i_{2m}} \ge \alpha_0 \sum_{i=1}^{n} \xi_i^{2m}, \ \forall \boldsymbol{\xi} \in \mathbb{R}^n \backslash \{0\}. \tag{1.13}$$

∎

Example 1.8 *The $2m$th order harmonic equation*

$$(-\triangle)^m u = f,$$

[6] https://arxiv.org/pdf/math/0304395.pdf

[7] https://en.wikipedia.org/wiki/Partial_differential_equation

is a linear $2m$th order uniformly elliptic, since

$$\begin{aligned} a_{i_1,\ldots,i_{2m}}(x) &= 1, \text{ if the indexes appear in pairs;}\\ a_{i_1,\ldots,i_{2m}}(x) &= 0, \text{ otherwise.}\end{aligned}$$

∎

Example 1.9 *Two 4th-order PDEs*

- $u_{tt} + a^2 u_{xxxx} = \Phi(x,t)$. *Equation of transverse vibration of elastic rods (non-homogeneous).*
- $\Delta^2 u = \Delta\Delta u = \Phi(x,y)$. *Nonhomogeneous biharmonic equation.*

∎

Example 1.10 *For a non-negative number p, the p-Laplacian is a nonlinear elliptic operator defined by*

$$L(u) = -\sum_{i=1}^{d} \partial_i \left(|\nabla u|^{p-2} \partial_i u \right).$$

Definition 1.6 *System of linear PDEs with p unknows $\boldsymbol{u} = (u_1, \cdots, u_p)$*

$$\pm\mathcal{L}_i(\boldsymbol{u}) \triangleq \pm\sum_{j=1}^{p} \left(\sum_{k=1}^{m_j} \sum_{i_1,\ldots,i_k=1}^{n} a^{i,j}_{i_1,\ldots,i_k} \frac{\partial^k}{\partial x_{i_1} \ldots \partial x_{i_k}} + a^{i,j}_0 \right) u_j = f_i, \tag{1.14}$$

$$a^{i,j}_{i_1,\ldots,i_k} = a^{i,j}_{i_1,\ldots,i_k}(\boldsymbol{x}),\ a^{i,j}_0 = a^{i,j}_0(\boldsymbol{x}),\ f_i = f_i(\boldsymbol{x}), i,j = 1,2,\cdots,p,$$

is elliptic in Ω, if for every $\boldsymbol{x} \in \Omega$ there exists $\alpha(\boldsymbol{x}) > 0$ s.t.

$$det(A_{ij}(\boldsymbol{x},\boldsymbol{\xi})) \geq \alpha(\boldsymbol{x}) \left(\sum_{k=1}^{n} \xi_k^2 \right)^{\sum_{j=1}^{p} m_j/2} \tag{1.15}$$

where

$$A_{ij}(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{i_1,\ldots,i_{m_j}=1}^{n} a^{i,j}_{i_1,\ldots,i_{m_j}}(\boldsymbol{x}) \xi_{i_1} \cdots \xi_{i_{m_j}},$$

and m_j is even number, the order of the highest derivative of u_j entering the system, $m = \max_{1\leq j\leq n}\{m_j\}$. ∎

Example 1.11 *Buckling of plate*

$$\begin{pmatrix} \Delta^2 u & -[u,w] \\ \Delta^2 w & -\frac{1}{2}[u,u] \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}, \quad \Omega \subset \mathbb{R}^2,$$

with $[u,v] = u_{xx}v_{yy} - 2u_{xy}v_{xy} + u_{yy}v_{xx}$.

Semi-linear reaction diffusion equations are typical models in nuclear reactor physics, and in chemical and biological reactions.

$$\frac{\partial \boldsymbol{u}}{\partial t} = D\Delta \boldsymbol{u} + f(\boldsymbol{u}, \alpha) = 0, \text{ in } \Omega \times (0,T)$$

http://www.internonlinearscience.org/upload/papers/20111207-Mishra(revised).pdf

The equations of 3D linear elasticity is found in (1.1.8) of textbook [5]. ∎

Remark 1.5 *Chapter 1 of Hackbusch's book [4] presents the classification of PDEs into types: classification of 2nd-order equations in n variables into types (§1.2), type classification for systems of 1st order (§1.3), and characteristic properties of the different types (§1.4). Definition 1.2.3 gives the definition of types for the general linear PDE of 2nd order. Definition 1.2.3 makes it clear that the 3 types mentioned by no means cover all cases [三种类型决不能涵盖所有情况].* ▮

Remark 1.6 *N个自变量的二阶PDE*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} a_{jk}\frac{\partial^2\phi}{\partial x_j \partial x_k} + H = 0, \quad H\text{表示低阶项},$$

的分类取决于二阶偏导数前的"系数矩阵"$A = (a_{jk})$的特征值，即特征方程$\det[A - \lambda I] = 0$的解.

分类准则是[8]：

- 如果有一个特征值$\lambda = 0$，而其它的或全正或全负，则方程为抛物型.
- 如果所有特征值或全正或全负，则椭圆型.
- 如果有一个特征值为正而其它的全负，或者有一个特征值为负而其它的全正，则双曲型.

▮

Example 1.12 (定常对流-扩散方程[9]) 对流-扩散方程是扩散和对流方程的组合，描述了粒子、能量或其它物理量由于扩散和对流两个过程在物理系统内传递的物理现象.

令c为感兴趣的变量(例如传质中的物种浓度，传热中的温度). 定常的对流-扩散方程描述对流-扩散系统的定常行为. 在一个定常状态，$\partial c/\partial t = 0$，所以数学模型为

$$0 = \nabla \cdot (D\nabla c) - \nabla \cdot (\boldsymbol{v}c) + R, \tag{1.16}$$

其中D是扩散系数，例如粒子运动的质量扩散系数或热传输的热扩散系数，$\boldsymbol{v}$是速度场，可是时间和空间的函数. 例如，在对流，c可是河流中的盐浓度，$\boldsymbol{v}$是水流动的速度，随时间和地点的变化. 另一个例子，c是平静的湖中小气泡的浓度，而$\boldsymbol{v}$ 是气泡通过浮力向表面上升的速度，取决于气泡的时间和位置. 对于多相流和多孔介质中的流动，$\boldsymbol{v}$是(假设的)表面速度. R 描述c的源或池. 例如，对于化学物种，$R > 0$意味着化学反应产生了更多的物种，而$R < 0$意味着化学反应破坏物种. 对于热传输，如果热能是由摩擦产生的，则$R > 0$就可能发生. ∇代表梯度，$\nabla\cdot$代表散度. 方程中，∇c代表浓度梯度.

(1.16)的右端项是三部分之和.

- 第*1*部分，$\nabla(D\nabla c)$，描述扩散. 想象一下，C 是一种化学物质的浓度. 与周围地区相比，当某个地方的浓度较低时(如局部最低浓度)，物质会从周围扩散进来，因此浓度会增加. 相反，如果与周围环境相比，浓度较高(例如浓度的局部最大值)，则物质将扩散出去，浓度将降低. 如果扩散系数D为常数，则净扩散与浓度的拉普拉斯(或二阶导数)成正比.
- 第*2*部分，$-\nabla(\boldsymbol{v}c)$，描述扩散. 想象一下，站在河岸上测量每秒海水的盐度(含盐量). 上游有人往河里倒了一桶盐. 过了一会儿，你会看到咸水带经过时，盐度突然上升，然后下降. 因此，给定位置的浓度会因流量而改变.

[8] C.R. Chester, Techniques in Partial Differential Equations, McGraw-Hill, 1971

- 第3部分，R，描述物理量的创建或销毁．例如，如果c是分子浓度，则R 描述分子是如何通过化学反应产生或破坏的．R可以是c和其它参数的函数．

通常有几个量，每个量都有自己的对流扩散方程，其中一个量的破坏意味着另一个量的产生．如，当甲烷燃烧时，不仅会破坏甲烷和氧气，还会产生二氧化碳和水蒸气．因此，虽然每种化学物质都有自己的对流扩散方程，但它们是耦合在一起的，必须作为一个联立微分方程组来求解．例如半导体物理中的*drift–diffusion*方程：

$$\begin{aligned}\frac{\partial n}{\partial t} &= -\nabla\cdot\frac{\mathbf{J}_n}{-q}+R, \quad \frac{\mathbf{J}_n}{-q} = -D_n\nabla n - n\mu_n\mathbf{E},\\ \frac{\partial p}{\partial t} &= -\nabla\cdot\frac{\mathbf{J}_p}{q}+R, \quad \frac{\mathbf{J}_p}{q} = -D_p\nabla p + p\mu_p\mathbf{E}.\end{aligned}$$

式中n, p分别是电子和空穴的浓度(密度)．$q > 0$是基本电荷，J_n, J_p分别由电子和空穴产生的电流，$J_n/-q$，J_p/q分别是电子和空穴对应的“粒子电流”，R 表示载流子产生和重组($R > 0$：电子空穴对的产生，$R < 0$电子空穴对的重组)，E是电场向量，μ_n，μ_p 电子和空穴迁移率．

流体力学中的不可压*Navier–Stokes*方程

$$\overbrace{\underbrace{\frac{\partial \mathbf{u}}{\partial t}}_{\textit{Variation}} + \underbrace{(\mathbf{u}\cdot\nabla)\mathbf{u}}_{\textit{Convection}}}^{\textit{Inertia (per volume)}} - \overbrace{\underbrace{\nu\,\nabla^2\mathbf{u}}_{\textit{Diffusion}} = \underbrace{-\nabla w}_{\substack{\textit{Internal}\\ \textit{source}}}}^{\textit{Divergence of stress}} + \underbrace{\mathbf{g}}_{\substack{\textit{External}\\ \textit{source}}} .$$

Millennium Problem on existence and smoothness of NS solutions [10] *asks for a proof of one of the following four statements:*

- *Existence and smoothness of NS solutions on* $\mathbb{R}^3$.
- *Existence and smoothness of NS solutions in* $\mathbb{R}^3/\mathbb{Z}^3$.
- *Breakdown of NS solutions on* $\mathbb{R}^3$.
- *Breakdown of NS Solutions on* $\mathbb{R}^3/\mathbb{Z}^3$.

在统计力学中，*Fokker–Planck*方程是一个*PDE*

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left[\mu(x,t)p(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[D(x,t)p(x,t)\right],$$

它描述了粒子速度的概率密度函数在阻力和随机力作用下的时间演化，如布朗运动．

在数学金融中，*Black–Scholes*方程是一个*PDE*，

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0,$$

它控制了欧洲看涨期权或*black-scholes*模型下的欧洲看涨期权的价格演变，式中期权价格V是股票价格S和时间t的函数，r这是无风险利率，$\sigma\sigma$是股票的波动性． ∎

Three types of the most commonly used boundary conditions:

$$u = u_D, \quad \forall x\in\partial\Omega \tag{1.17}$$

$$\frac{\partial u}{\partial \nu} = g, \quad \forall x\in\partial\Omega \tag{1.18}$$

$$\frac{\partial u}{\partial \nu} + \alpha u = g, \quad \forall x\in\partial\Omega \tag{1.19}$$

where $\alpha \geq 0$, and $\alpha > 0$ at least on some part of the boundary (physical meaning: higher density produces bigger outward diffusion flux).

[10] C.L. Fefferman, https://www.claymath.org/sites/default/files/navierstokes.pdf

Consistency	Stability	Convergence
Accuracy 精度	Efficiency 效率	Scalability 可扩展性

- 1st type BC —— Dirichlet BC,
- 2nd type BC —— Neumann BC,
- 3rd type BC —— Robin BC;
- Mixed-type BC —— Different types of BCs imposed on different parts of the boundary.

1.1 Numerical Approximation

A PDE and its numerical approximation[11]

$$Lu = f, \qquad L_h u_h = f_h.$$

where u belongs to some infinite dimensional function space, while u_h belongs to some finite dimensional space. Here h denotes the mesh/grid size, or $1/h$ indicates the number of mesh points or the dimension of the finite-dimensional space. It may also be referred to as number of degrees of freedom.

The main concept behind any finite difference scheme is related to the definition of the derivative of a smooth function u at a point $x \in \mathbb{R}$:

$$u'(x) = \lim_{h \to 0} \frac{u(x+h) - u(x)}{h}$$

and to the fact that when $h \to 0$ (without vanishing), the quotient on the right-hand side provides a "good" approximation of the derivative.

$$\begin{array}{rl} \text{derivative} & u'(x) \\ \text{difference} & u(x+h) - u(x) \\ \text{difference quotient} & \dfrac{u(x+h) - u(x)}{h} \end{array}$$

1.2 Finite difference coefficient

In mathematics, to approximate a derivative to an arbitrary order of accuracy, it is possible to use the finite difference. A finite difference can be central, forward or backward [12].

`Findiff` & `Finitediff` are two Python packages for finite difference numerical derivatives and partial differential equations.

Finite Difference Coefficients Calculator, Prof. Cameron Taylor (MIT, 12 December 2019).

For a sufficient function f, the Taylor series expansion reads

$$\begin{aligned} f(x+h) =& f(x) + \frac{h}{1!} f'(x) + \frac{h^2}{2!} f''(x) + \cdots = \sum_{i=0}^{\infty} \frac{h^i}{i!} f^{(i)}(x) \\ =& \left(\sum_{i=0}^{\infty} \frac{h^i}{i!} \frac{d^i}{dx^i} \right) f(x) = \exp\left(h \frac{d}{dx} \right) f(x), \end{aligned}$$

[11] http://cpraveen.github.io/teaching/index.html

[12] B. Fornberg, Generation of finite difference formulas on arbitrarily spaced grids, *Math. Comput.*, 51(184): 699-706, 1988.

which can be considered as the “function” of h. For example, replacing h with $-h$ or $2h$ gives

$$f(x-h)=\left(\sum_{i=0}^{\infty}\frac{(-h)^i}{i!}\frac{d^i}{dx^i}\right)f(x)=\exp\left(-h\frac{d}{dx}\right)f(x),$$

$$f(x+2h)=\left(\sum_{i=0}^{\infty}\frac{(2h)^i}{i!}\frac{d^i}{dx^i}\right)f(x)=\exp\left(2h\frac{d}{dx}\right)f(x).$$

1.2.1 Central finite difference

The following table contains the coefficients of the central differences, for several orders of accuracy and with uniform grid spacing.

导数	精度	-5	-4	-3	-2	-1	0	1	2	3	4
1	2					-1/2	0	1/2			
	4				1/12	-2/3	0	2/3	-1/12		
	6			-1/60	3/20	-3/4	0	3/4	-3/20	1/60	
	8		1/280	-4/105	1/5	-4/5	0	4/5	-1/5	4/105	-1/
2	2					1	-2	1			
	4				-1/12	4/3	-5/2	4/3	-1/12		
	6			1/90	-3/20	3/2	-49/18	3/2	-3/20	1/90	
	8		-1/560	8/315	-1/5	8/5	-205/72	8/5	-1/5	8/315	-1/
3	2				-1/2	1	0	-1	1/2		
	4			1/8	-1	13/8	0	-13/8	1	-1/8	
	6		-7/240	3/10	-169/120	61/30	0	-61/30	169/120	-3/10	7/
4	2				1	-4	6	-4	1		
	4			-1/6	2	-13/2	28/3	-13/2	2	-1/6	
	6		7/240	-2/5	169/60	-122/15	91/8	-122/15	169/60	-2/5	7/
5	2			-1/2	2	-5/2	0	5/2	-2	1/2	
	4		1/6	-3/2	13/3	-29/6	0	29/6	-13/3	3/2	-1
	6	-13/288	19/36	-87/32	13/2	-323/48	0	323/48	-13/2	87/32	-19
6	2			1	-6	15	-20	15	-6	1	
	4		-1/4	3	-13	29	-75/2	29	-13	3	-1
	6	13/240	-19/24	87/16	-39/2	323/8	-1023/20	323/8	-39/2	87/16	-19

1.2.2 Forward finite difference

The following table contains the coefficients of the forward differences, for several orders of accuracy and with uniform grid spacing.

导数	精度	0	1	2	3	4	5	6	7
1	1	-1	1						
	2	-3/2	2	-1/2					
	3	-11/6	3	-3/2	1/3				
	4	-25/12	4	-3	4/3	-1/4			
	5	-137/60	5	-5	10/3	-5/4	1/5		
	6	-49/20	6	-15/2	20/3	-15/4	6/5	-1/6	
2	1	1	-2	1					
	2	2	-5	4	-1				
	3	35/12	-26/3	19/2	-14/3	11/12			
	4	15/4	-77/6	107/6	-13	61/12	-5/6		
	5	203/45	-87/5	117/4	-254/9	33/2	-27/5	137/180	
	6	469/90	-223/10	879/20	-949/18	41	-201/10	1019/180	-7/10
3	1	-1	3	-3	1				
	2	-5/2	9	-12	7	-3/2			
	3	-17/4	71/4	-59/2	49/2	-41/4	7/4		
	4	-49/8	29	-461/8	62	-307/8	13	-15/8	
	5	-967/120	638/15	-3929/40	389/3	-2545/24	268/5	-1849/120	29/15
	6	-801/80	349/6	-18353/120	2391/10	-1457/6	4891/30	-561/8	527/30
4	1	1	-4	6	-4	1			
	2	3	-14	26	-24	11	-2		
	3	35/6	-31	137/2	-242/3	107/2	-19	17/6	
	4	28/3	-111/2	142	-1219/6	176	-185/2	82/3	-7/2
	5	1069/80	-1316/15	15289/60	-2144/5	10993/24	-4772/15	2803/20	-536/15

1.2.3 Backward finite difference

The following table contains the coefficients of the backward differences, for several orders of accuracy and with uniform grid spacing.

导数	精度	-8	-7	-6	-5	-4	-3	-2	-1	0
1	1								-1	1
	2							1/2	-2	3/2
	3						$-1/3$	3/2	-3	11/6
2	1							1	-2	1
	2						-1	4	-5	2
3	1						-1	3	-3	1
	2					3/2	-7	12	-9	5/2
4	1					1	-4	6	-4	1
	2				-2	11	-24	26	-14	3

1.3 Consistency

Does the numerical scheme approximate the PDE?[13]

Local truncation error: $\tau_h := L_h u - f_h \neq 0$, where u is exact solution of PDE, which implies $L_h u \neq f_h$.

The truncation error is discretization error[14].

The term truncation error reflects the fact that a finite part of a Taylor series is used in the approximation.

The scheme is consistent if

$$\tau_h = \mathcal{O}(h^p),$$

for some $p > 0$, since $\tau_h \to 0$ as $h \to 0$.

Easier to check consistency property. Use of Taylor formula.

Remark 1.7 *We will use the order symbol $\mathcal{O}(\#)$ frequently. By $\tau_h = \mathcal{O}(h^p)$ we mean that $\tau_h = Ch^p$ for some C which does not depend on h. Usually we show that*

$$\tau_h \leq C(u)h^p,$$

where C independent of h.

1.4 Stability

For IVP, one interpretation of stability of a difference scheme is that for a stable difference scheme small errors in the initial conditions cause small errors in the solution. see §2.4 (Page 73) of book. 此外[15] [16]

- 相容性只是收敛的一个必要条件, 而不是一个充分条件.
- 计算过程中产生的舍入误差可能导致解的爆炸, 或侵蚀整个计算.
- 如果计算中舍入误差不被放大, 则格式是稳定的.
- Fourier 方法等可以用于检查格式是否稳定.

[13] http://cpraveen.github.io/teaching/index.html

[14] https://www.ljll.math.upmc.fr/frey/cours/UdC/ma691/ma691_ch6.pdf

[15] J.W. Thomas, *Numerical Partial Differential Equations: Finite Difference Methods*, Springer, 1995.

[16] http://wwwf.imperial.ac.uk/~mdavis/FDM11/LECTURE_SLIDES2.PDF

Is the numerical solution bounded independently of h?[17]

$$\|u_h\| \leq C\|f_h\| \quad \Longrightarrow \quad \|L_h^{-1}\| = \sup_{f_h} \frac{\|L_h^{-1} f_h\|}{\|f_h\|} \leq C,$$

式中C不依赖于h或f_h或u_h. Stability is in the sense of Lax-Richtmyer[18].

- 稳定性的需要是确保在迭代过程中数值计算不会爆破(blow up); 舍入误差不能被放大.
- 希望数值格式能继承精确解的稳定性: 有界性、正性、单调性等.
- 稳定性需要用于显示收敛性

$$u_h \to u, \qquad h \to 0.$$

1.5 Convergence

Does the numerical solution approach the exact solution

$$u_h \to u \quad \text{as} \quad h \to 0?$$

Define the solution error or global error:

$$e_h := u - u_h.$$

Then

$$L_h e_h = L_h u - L_h u_h = L_h u - f_h = \tau_h \quad \text{i.e.} \quad \boxed{L_h e_h = \tau_h},$$

so that for a stable scheme

$$\|e_h\| \leq C\|\tau_h\|.$$

If the scheme is consistent, then

$$\|e_h\| \leq C\|\tau_h\| = Ch^p \to 0 \quad \text{as} \quad h \to 0.$$

Lax equivalence theorem[see Page 159]:
For a linear scheme approximating a well-posed IVP of linear PDE,

consistency + stability = convergence

1.6 Accuracy

- Truncation error for a consistent scheme

$$\tau_h = \mathcal{O}(h^p), \quad p > 0$$

- Global error of a consistent and stable scheme

$$\|e_h\| \leq C\|\tau_h\| = \mathcal{O}(h^p)$$

where p measures the rate of convergence w.r.t. h.

[17] http://cpraveen.github.io/teaching/index.html
[18] P465

- We would like to have p as large as possible, since this leads to a smaller error, as $h < 1$.
- $p = 1$: 1st order accurate. They have very high errors and require very large grids (small h) to reduce error to acceptable levels.
- Most schemes used in practice have $p = 2$: 2nd order accurate.

$$h \to h/2, \quad e_{h/2} = \frac{1}{4} e_h.$$

- Schemes with $p > 2$ are referred to as high-order accurate schemes.

1.7 Building confidence

How can you trust a numerical solution?

- Construct schemes which have as many qualitative properties of the PDE model as possible: stability, invariance, conserved quantities, positivity, monotonicity, etc.
- Do as much theoretical analysis as possible: stability, error estimates
- For problems that really matter, it is usually not possible to make a complete theoretical analysis;
  - do lots of numerical experiments.
  - compare with analytical results
  - compare with experimental results
  - compare with results from other numerical simulations
- Studying linear models and linear schemes is a first step, and a necessary condition.
- Linearize non-linear models and schemes; study their properties
- Many numerical schemes constructed for linear problems and/or scalar problems; they are also applied to non-linear and systems of PDEs without full theoretical justification (Hyperbolic conservation laws).

1.8 Verification and Validation

- Verification [验证]
  - Are you solving the PDE correctly? Consistency analysis
  - Purely a mathematical step (you can prove theorems) without relation to the real world.
  - Compare numerical solution with analytical solutions
  - Method of manufactured solutions[19][20]: Assume R be some differential operator and solve (numerically) problem: $R(u) = 0$.

[19] P.J. Roache, Code verification by the method of manufactured solutions, *Trans. ASME, J. Fluids Engrg.*, 124(2002), 4-10.

[20] K. Salari, P. Knupp, Code verification by the method of manufactured solutions, Sandia Report, Sandia National Laboratories, 2000.

- $*$ 加一源项和选一“编造的解”(光滑)使得 $f(u^*) = R(u^*)$.
- $*$ 数值求解带源项的问题$R(u) = f(u^*)$. 注: u is unknown; u^* is some assumed function, $f(u^*)$ is a known source term[21].
- $*$ Perform numerical study of convergence: Plot error $||e_h||$ w.r.t. h and find convergence rate. Compare with theoretical analysis.

$$\frac{\|e_h\|}{\|e_{h/2}\|} = 2^p \implies p = \frac{\log\left(\frac{\|e_h\|}{\|e_{h/2}\|}\right)}{\log(2)}.$$

- Validation [确认]
  - Are you solving the correct (PDE) model?
  - This is a modeling/physics issue.
  - All PDE models are approximations to reality.
  - How well does the PDE model + numerical solution represent reality?
  - Compare numerical solution with (lots of) experiments.

2 FD approximation for model problem

教材的第5-8页.

A finite difference is a mathematical expression of the form $f(x+b)-f(x+a)$. If a finite difference is divided by $b-a$, one gets a difference quotient. The approximation of derivatives by finite differences plays a central role in finite difference methods for the numerical solution of differential equations, especially boundary value problems.

Consider Dirichlet BVP of 2D Poisson equation

$$\begin{cases} -\Delta u(x,y) &= f(x,y), \qquad \forall (x,y) \in \Omega, \\ \quad u(x,y) &= u_D(x,y), \quad \forall (x,y) \in \partial\Omega, \end{cases}$$

where $\Omega = (0,1) \times (0,1)$.

2.1 Summation by part

Summation by part is a discrete version of Green theorem or integration by part.

If $v, w \in L(\Omega_h)$ and $w = 0$ on $\partial\Omega_h$, then

$$-\langle \Delta_h v, w\rangle_h = \langle \partial_x v, \partial_x w\rangle_h + \langle \partial_y v, \partial_y w\rangle_h.$$

Proof: If assuming that $v_i, w_i \in \mathbb{R}$, $i = 0, 1, \cdots, N$ and $w_0 = w_N = 0$, then

$$\sum_{i=1}^{N}(v_i - v_{i-1})(w_i - w_{i-1}) = \sum_{i=1}^{N}(v_i - v_{i-1})w_i - \sum_{i=1}^{N}(v_i - v_{i-1})w_{i-1}$$

$$= \sum_{i=1}^{N}(v_i - v_{i-1})w_i - (v_1 - v_0)w_0 - \sum_{i=1}^{N}(v_{i+1} - v_i)w_i + (v_{N+1} - v_N)w_N$$

$$= -\sum_{i=1}^{N}(v_{i+1} - 2v_i + v_{i-1})w_i,$$

[21] 加源项的方式: 选择一个光滑函数u^*, 例如$\sin(x)$, 把u^*看成是加了源项后的问题的解(即manufactured solution), 将u^*代入加了源项后的问题, 由此获得此时的源项表达式$f(u^*)$.

i.e.

$$-\langle D_x^2 v, w\rangle_h = \langle \partial_x v, \partial_x w\rangle_h.$$

Similarly,

$$-\langle D_y^2 v, w\rangle_h = \langle \partial_y v, \partial_y w\rangle_h.$$

∎

Discrete inner product:

$$(V, W)_h := \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} h^2 V_{i,j} W_{i,j}.$$

The forward/backward difference operators are defined by

$$\begin{aligned} D_x^+ V_{i,j} &= V_{i+1,j} - V_{i,j},\ D_x^- V_{i,j} = V_{i,j} - V_{i-1,j},\\ D_y^+ V_{i,j} &= V_{i,j+1} - V_{i,j},\ D_y^- V_{i,j} = V_{i,j} - V_{i,j-1}. \end{aligned}$$

Summation by part: $V_i : \overline{\Omega}_h \to \mathbb{R}$ and $V_i = 0$ on $\partial\Omega_h$, then

$$-\left(D_x^+ D_x^- V + D_y^+ D_y^- V, V\right) = \sum_{i=1}^{N}\sum_{j=1}^{N-1} h^2 |D_x^- V_{i,j}|^2 + \sum_{i=1}^{N-1}\sum_{j=1}^{N} h^2 |D_y^- V_{i,j}|^2. \qquad (2.1)$$

Proof: Using the fact that $V_{0,j} = V_{N,j} = 0$ gives

$$\begin{aligned}
&-\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}(D_x^+ D_x^- V_{i,j})V_{i,j} = -\sum_{j=1}^{N-1}\sum_{i=1}^{N-1}\big((V_{i+1,j} - V_{i,j}) - (V_{i,j} - V_{i-1,j})\big)V_{i,j}\\
&= \sum_{j=1}^{N-1}\sum_{i=1}^{N-1} -(V_{i+1,j} - V_{i,j})V_{i,j} + \sum_{j=1}^{N-1}\sum_{i=1}^{N-1}(V_{i,j} - V_{i-1,j})V_{i,j}\\
&= -(V_{N,j} - V_{N-1,j})V_{N-1,j} + \sum_{j=1}^{N-1}\sum_{i=0}^{N-2} -(V_{i+1,j} - V_{i,j})V_{i,j}\\
&+ (V_{1,j} - V_{0,j})V_{0,j} + \sum_{j=1}^{N-1}\sum_{i=1}^{N-1}(V_{i,j} - V_{i-1,j})V_{i,j}\\
&= (V_{N,j} - V_{N-1,j})^2 - \sum_{j=1}^{N-1}\sum_{i=1}^{N-1}(V_{i,j} - V_{i-1,j})V_{i-1,j} + \sum_{j=1}^{N-1}\sum_{i=1}^{N-1}(V_{i,j} - V_{i-1,j})V_{i,j}\\
&= (V_{N,j} - V_{N-1,j})^2 + \sum_{j=1}^{N-1}\sum_{i=1}^{N-1}(V_{i,j} - V_{i-1,j})^2 = \sum_{j=1}^{N-1}\sum_{i=1}^{N}(V_{i,j} - V_{i-1,j})^2
\end{aligned}$$

∎

Operator $-L_h = -D_x^+ D_x^- - D_y^+ D_y^-$ is invertible.

Using the summation by part (2.1) gives that the quadratic form $(-L_h U, U)$ satisfies

$$(-L_h U, U) \geq 0.$$

If $-L_h U = 0$, then the summation by part (2.1) tells us that $D_x^- V_{i,j} = 0$ and $D_y^- V_{i,j} = 0$.

Because $U = 0$ on $\partial\Omega_h$, $U_{i,j} = 0$ in Ω_h. Thus $-L_h U = 0$ is equivalent to $U \equiv 0$ so that $-L_h$ is invertible and there exists a unique solution to $-L_h U_{i,j} = f_{i,j}$. ∎

3.5 Finite Differences and Fast Poisson Solvers(Gilbert Strang, 2006)

2.2 Computational complexity

The Complexity of Solving the Discrete Poisson Equation using Jacobi, SOR, Conjugate Gradients, and the FFT (James Demmel, CS267).

Solving the Discrete Poisson Equation using Multigrid (James Demmel, CS267).

Discuss Poisson's equation, which arises in heat flow, electrostatics, gravity, and other situations. In 1-dimension the Poisson equation is

$$-u'' = f(x)$$

for x in a region $\Omega = [0,1]$ with zero Dirichlet boundary conditions. It can be discretized by using finite differences

$$-U_{i+1} + 2U_i - U_{i-1} = h^2 f_i, \quad i = 1, 2, \cdots, N-1,$$

or

$$\mathbf{P}\boldsymbol{U} = \boldsymbol{b}, \ \boldsymbol{U} = (U_1, \cdots, U_{N-1})^T, \ \boldsymbol{b} = h^2(f_1, \cdots, f_{N-1})^T.$$

and

$$\mathbf{P} = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}$$

The matrix $\mathbf{P}$ is diagonally dominant [22] irreducible [23] (see Page 22 of LZP textbook) symmetric positive definite [对称正定]. For any $0 \neq \boldsymbol{\xi} \in \mathbb{R}^{N-1}$, we have

$$\boldsymbol{\xi}^T \mathbb{P} \boldsymbol{\xi} =$$

In 2-dimensions the Poisson equation is

$$-u_{xx} - u_{yy} = f(x,y)$$

for (x,y) in a region Ω in the (x,y) plane, say the unit square $0 < x, y < 1$ with zero Dirichlet boundary conditions.

Discretize this equation by using finite differences: use an $(n+1)$-by-$(n+1)$ grid on Ω = the unit square, where $h = 1/(n+1)$ is the grid spacing. Let $U_{i,j}$ be the approximate solution at $x = ih$ and $y = jh$. This is shown below for $n = 7$, including the system of linear equations it leads to:

$$4U_{i,j} - U_{i+1,j} - U_{i-1,j} - U_{i,j+1} - U_{i,j-1} = b_{i,j}.$$

The above linear equation relating $U_{i,j}$ and the value at its neighbors (indicated by the blue stencil) must hold for $1 \leq i, j \leq n$, giving us $N = n^2$ equations in N unknowns. When (i,j) is adjacent to a boundary ($i = 1$ or $j = 1$ or $i = n$ or $j = n$), one or more of the $U_{i\pm1,j\pm1}$ values is on the boundary and therefore 0. $b_{i,j} = h^2 f(ih, jh)$, the scaled value of the right-hand-side function $f(x,y)$ at the corresponding grid point (i,j).

[22] 如果A的每个对角元的绝对值都比所在行的非对角元的绝对值的和要大, 那么A是(行)严格对角占优的. 弱对角占优: $|a_{i,i}| \geq \sum_{i\neq j=1}^{n} |a_{i,j}|$, $i = 1, 2, \cdots, n$, 且该式中至少有一个不等式严格成立.

[23] 对n阶方阵, 如果存在排列阵P, s.t. PAP^T 为一个分块上三角阵, $PAP^T = \begin{pmatrix} B & C \\ O & D \end{pmatrix}$, O为零矩阵, B, D是阶数不小于1的方阵, 则称A是可约的. 否则就称A是不可约的.

Algorithm	Type	Serial Time	PRAM Time	Storage	#Procs
Dense LU	D	N^3	N	N^2	N^2
Band LU	D	N^2	N	$N^{(3/2)}$	N
Inv(P)*bhat	D	N^2	$\log N$	N^2	N^2
Jacobi	I	N^2	N	N	N
Sparse LU	D	$N^{(3/2)}$	$N^{(1/2)}$	$N \log N$	N
CG	I	$N^{(3/2)}$	$N^{(1/2)} \log N$	N	N
SOR	I	$N^{(3/2)}$	$N^{(1/2)}$	N	N
FFT	D	$N \log N$	$\log N$	N	N
Multigrid	I	N	$(\log N)^2$	N	N
Lower Bound		N	$\log N$	N	

Table 1: Also see CS267 "Applications of Parallel Computers" (2020 Lecture 20 on FFT, Page 11 of PPT) of Prof. James Demmel, U.C. Berkeley

To write this as a linear system in the more standard form

$$\mathbf{P}\boldsymbol{U} = \boldsymbol{b},$$

where $\boldsymbol{U}$ and $\boldsymbol{b}$ are column vectors, we need to choose a linear ordering of the unknowns $U_{i,j}$. For example, the natural row ordering and the Red-Black ordering can be considered.

$$\left(\begin{array}{cccc|cccc|cccc|cccc}
4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 4 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
-1 & 0 & 0 & 0 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 4 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & 0 & 0 & 0 & -1 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 4 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4
\end{array}\right)
\begin{pmatrix} U_{1,1} \\ U_{2,1} \\ U_{3,1} \\ U_{4,1} \\ U_{1,2} \\ U_{2,2} \\ U_{3,2} \\ U_{4,2} \\ U_{1,3} \\ U_{2,3} \\ U_{3,3} \\ U_{4,3} \\ U_{1,4} \\ U_{2,4} \\ U_{3,4} \\ U_{4,4} \end{pmatrix}
=
\begin{pmatrix} b_{1,1} \\ b_{2,1} \\ b_{3,1} \\ b_{4,1} \\ b_{1,2} \\ b_{2,2} \\ b_{3,2} \\ b_{4,2} \\ b_{1,3} \\ b_{2,3} \\ b_{3,3} \\ b_{4,3} \\ b_{1,4} \\ b_{2,4} \\ b_{3,4} \\ b_{4,4} \end{pmatrix}.$$

The grid points is in the Red-Black order (named after the resemblance to a chess board).

The 1st column in the table identifies the algorithm, except the last entry, which gives a simple Lower Bound on the running time for any algorithm. The Lower Bound is obtained as follows. For the serial time, the time required simply to print each of the N solution components is N. For the PRAM time, which assumes as many processors as we like and assumes communication is free, we note that the inverse $\mathbf{P}^{-1}$ of the discrete Poisson matrix $\mathbf{P}$ is dense, so that each component of the solution $\boldsymbol{U} = \mathbf{P}^{-1}\boldsymbol{b}$ is a nontrivial function of each of the N components of $\boldsymbol{b}$. The time required to compute any nontrivial function of N values in parallel is $\log N$.

The 2nd column says whether the algorithm is of Type D=Direct, which means that after a finite number of steps it produces the exact answer (modulo roundoff error), or of Type I=Indirect, which means that one step of the algorithm decreases the error by a constant factor $\rho < 1$, where ρ depends on the algorithm and N. This means that if one wants the final error to be epsilon times smaller than the initial error, one must take m steps where $\rho^m \leq \epsilon$. To compute the complexities in the table, we choose m so that ρ^m is about as small as the discretization error , i.e. the difference between the true solution $u(ih, jh)$ and the exact discrete solution $U_{i,j}$. There is no point in making the error in the computed $U_{i,j}$ any smaller than this, since this could only decrease the more significant error measure, the difference between the true solution $u(ih, jh)$ and the computed $U_{i,j}$, by a factor of 2.

The 2nd and 3rd columns give the running time for the algorithm on a serial machine and a PRAM, respectively. Recall that on a PRAM we can have as many processors as we need (shown in the last column), and communication is free. Thus, the PRAM time is a lower bound for any implementation on a real parallel machine. Finally, the 5th column indicated how much storage is needed. LU decomposition requires significantly more storage than the other methods, which require just a constant amount of storage per grid point.

All table entries are meant in the $\mathcal{O}(\cdot)$ sense. PRAM is an idealized parallel model with zero cost communication.

Key to abbreviations:

Dense LU : Gaussian elimination (GE), treating $\mathbf{P}$ as dense
Band LU : GE, treating $\mathbf{P}$ as zero outside a band of half-width $n-1$ near diagonal
Sparse LU : GE, exploiting entire zero-structure of $\mathbf{P}$
Inv(P)*bhat : precompute and store inverse of $\mathbf{P}$, multiply it by RHS $\boldsymbol{b}$
CG : Conjugate Gradient method
SOR : Successive Overrelaxation
FFT : Fast Fourier Transform based method

3 FD approximation for general problem

教材的第8-18页. TBA

4 Error analysis based on maximum principle

教材的第19-24页. TBA

5 Asymptotic Error Analysis and extrapolation

教材的第25-27页. TBA

6 Supplement and Notes

教材的第28-29页. TBA

7 Appendix: Some concepts

7.1 Overview of types of abstract spaces

见图1.

7.2 Elliptic operator

In the theory of PDEs, elliptic operators are differential operators that generalize the Laplace operator. They are defined by the condition that the coefficients of the highest-order derivatives be positive, which implies the key property that the principal symbol is invertible, or equivalently that there are no real characteristic directions.

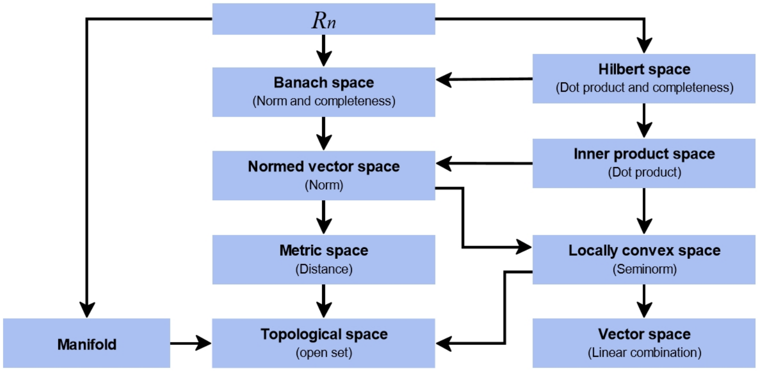

Figure 1: Overview of types of abstract spaces. An arrow from space A to space B implies that space A is also a kind of space B. That means, for instance, that a normed vector space is also a metric space. [Boris Tsirelson et al., Spaces in mathematics, WikiJournal of Science, 2018, 1(1):2]

Elliptic operators are typical of potential theory, and they appear frequently in electrostatics and continuum mechanics. Elliptic regularity implies that their solutions tend to be smooth functions (if the coefficients in the operator are smooth). Steady-state solutions to parabolic equations generally solve elliptic equations.

7.2.1 Principal symbol of operators on Euclidean space

The symbol of a linear differential operator is obtained from a differential operator of a polynomial by, roughly speaking, replacing each partial derivative by a new variable. The symbol of a differential operator has broad applications to Fourier analysis.

Let P be a linear differential operator of order k on the Euclidean space $\mathbb{R}^d$. Then P is a polynomial in the derivative D, which in multi-index notation can be written

$$P = p(x, D) = \sum_{|\alpha| \le k} a_\alpha(x) D^\alpha.$$

The total symbol of P is the polynomial p:

$$p(x, \xi) = \sum_{|\alpha| \le k} a_\alpha(x) \xi^\alpha.$$

The leading symbol, also known as the principal symbol, is the highest-degree component of p:

$$\sigma_P(\xi) = \sum_{|\alpha| = k} a_\alpha \xi^\alpha$$

and is of importance later because it is the only part of the symbol that transforms as a tensor under changes to the coordinate system.

The symbol of P appears naturally in connection with the Fourier transform as follows. Let f be a Schwartz function. Then by the inverse Fourier transform,

$$Pf(x) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} e^{ix\cdot\xi} p(x, i\xi) \hat{f}(\xi)\, d\xi.$$

This exhibits P as a Fourier multiplier. A more general class of functions $p(x, \xi)$ which satisfy at most polynomial growth conditions in ξ under which this integral is well-behaved comprises the pseudo-differential operators.

7.2.2 Definition of elliptic operator

A linear differential operator L of order m on a domain Ω in $\mathbb{R}^n$ given by

$$Lu = \sum_{|\alpha| \leq m} a_\alpha(x) \partial^\alpha u$$

where $\alpha = (\alpha_1, ..., \alpha_n)$ is a multi-index, and $\partial^\alpha u = \partial^{\alpha_1} \cdots \partial^{\alpha_n} u)$ is called elliptic if for every x in Ω and every non-zero ξ in $\mathbb{R}^n$,

$$\sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \neq 0,$$

where $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_n^{\alpha_n}$.

In many applications, this condition is not strong enough, and instead a uniform ellipticity condition may be imposed for operators of order $m = 2k$:

$$(-1)^k \sum_{|\alpha|=2k} a_\alpha(x)\xi^\alpha > C|\xi|^{2k},$$

where C is a positive constant. Note that ellipticity only depends on the highest-order terms. [24]

A nonlinear operator

$$L(u) = F(x, u, (\partial^\alpha u)_{|\alpha| \leq m})$$

is elliptic if its first-order Taylor expansion with respect to u and its derivatives about any point is a linear elliptic operator.

7.2.3 Examples

Example 7.1 *The negative of the Laplacian in $\mathbb{R}^d$ given by*

$$-\Delta u = -\sum_{i=1}^{d} \partial_i^2 u$$

is a uniformly elliptic operator. The Laplace operator occurs frequently in electrostatics. If ρ is the charge density within some region Ω, the potential Φ must satisfy the equation

$$-\Delta\Phi = 4\pi\rho.$$

Example 7.2 *Given a matrix-valued function $A(x)$ which is symmetric and positive definite for every x, having components a_{ij}, the operator*

$$Lu = -\partial_i \left(a^{ij}(x)\partial_j u\right) + b^j(x)\partial_j u + cu$$

is elliptic. This is the most general form of a second-order divergence form linear elliptic differential operator. The Laplace operator is obtained by taking $A = I$. These operators also occur in electrostatics in polarized media.

[24] Note that this is sometimes called strict ellipticity, with uniform ellipticity being used to mean that an upper bound exists on the symbol of the operator as well. It is important to check the definitions the author is using, as conventions may differ. See, e.g., Chapter 6 of Evans' "PDEs" (2nd ed.) for a use of the first definition with $k = 1$, and Chapter 3 of Gilbarg and Trudinger's "Elliptic PDES of 2nd order" (2nd ed.) for a use of the second.

Example 7.3 *For p a non-negative number, the p-Laplacian is a nonlinear elliptic operator defined by*

$$L(u) = -\sum_{i=1}^{d} \partial_i \left(|\nabla u|^{p-2} \partial_i u \right).$$

A similar nonlinear operator occurs in glacier mechanics. The Cauchy stress tensor of ice, according to Glen's flow law, is given by

$$\tau_{ij} = B \left(\sum_{k,l=1}^{3} (\partial_l u_k)^2 \right)^{-\frac{1}{3}} \cdot \frac{1}{2} (\partial_j u_i + \partial_i u_j)$$

for some constant B. The velocity of an ice sheet in steady state will then solve the nonlinear elliptic system

$$\sum_{j=1}^{3} \partial_j \tau_{ij} + \rho g_i - \partial_i p = Q,$$

where ρ is the ice density, g is the gravitational acceleration vector, p is the pressure and Q is a forcing term.

Linear elliptic equations
Laplace equation $\Delta w = 0$.
Poisson equation $\Delta w + \Phi(x) = 0$.
Helmholtz equation $\Delta w + \lambda w = \Phi(x)$
Nonlinear elliptic equations

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = f(w)$$

$$\frac{\partial}{\partial x}\left[f(x) \frac{\partial w}{\partial x} \right] + \frac{\partial}{\partial y}\left[g(y) \frac{\partial w}{\partial y} \right] = f(w)$$

$$\frac{\partial}{\partial x}\left[f(w) \frac{\partial w}{\partial x} \right] + \frac{\partial}{\partial y}\left[g(w) \frac{\partial w}{\partial y} \right] = f(w)$$

More PDE examples can be found in [1, P3-6].

7.2.4 Elliptic regularity theorem

Let L be an elliptic operator of order $2k$ with coefficients having $2k$ continuous derivatives. The Dirichlet problem for L is to find a function u, given a function f and some appropriate boundary values, such that $Lu = f$ and such that u has the appropriate boundary values and normal derivatives. The existence theory for elliptic operators, using Gårding's inequality and the Lax-Milgram lemma, only guarantees that a weak solution u exists in the Sobolev space H^k.

This situation is ultimately unsatisfactory, as the weak solution u might not have enough derivatives for the expression Lu to even make sense.

The elliptic regularity theorem guarantees that, provided f is square-integrable, u will in fact have $2k$ square-integrable weak derivatives. In particular, if f is infinitely-often differentiable, then so is u.

Any differential operator exhibiting this property is called a hypoelliptic operator; thus, every elliptic operator is hypoelliptic. The property also means that every fundamental solution of an elliptic operator is infinitely differentiable in any neighborhood not containing 0.

As an application, suppose a function f satisfies the Cauchy-Riemann equations. Since the Cauchy-Riemann equations form an elliptic operator, it follows that f is smooth.

7.3 Comparison principle

The comparison principle refers to the general concept that a subsolution to an elliptic equation stays below a supersolution of the same equation. It known to hold under a great generality of assumptions.

The comparison principle can also be understood as the fact that the difference between a subsolution and a supersolution satisfies the maximum principle. The uniqueness of the solution of the equation is an immediate consequence.

The two statements below (see Sections 7.3.1-7.3.2) correspond to the comparison principle for elliptic and parabolic equations with Dirichlet boundary conditions. The main difference with the local case, is that for nonlocal equations the Dirichlet condition has to be taken in the whole complement of the domain Ω instead of only the boundary.

Other boundary conditions require appropriate modifications.

$\Omega \subset \mathbb{R}^n$是一个有界开集, $u, v \in C^2(\Omega) \cap C^0(\overline{\Omega})$, 则不等式

$$\Delta u \geq \Delta v \text{ in } \Omega; \;\; u \leq v \text{ on } \partial\Omega$$

隐含着不等式$u \leq v$ even on $\overline{\Omega}$.

If u has a local maximum at point $M \in \Omega$, then at this point

$$u_x(M) = 0, \;\; u_y(M) = 0,$$
$$u_{xx}(M) \leq 0, \;\; u_{yy}(M) \leq 0.$$

Therefore, if $\Delta u > 0$ at each point of Ω, u cannot attain maximum inside Ω.

In general, ff $b_1(x_1, ..., x_n)$, $b_2(x_1, x_2, ..., x_n)$,...,$b_n(x_1, x_2, ..., x_n)$ are any bounded functions in Ω and

$$\Delta u + \sum_{i=1}^{n} b_i \frac{\partial u}{\partial x_i} > 0 \text{ in } \Omega,$$

then u cannot attain its maximum inside Ω.

Maximum Theorem: Let $\Delta u \geq 0$ in Ω. If u attains its maximum M at any point of Ω, then $u \equiv M$ in Ω.

Definition A function u satisfying $\Delta u \geq 0$ in Ω is called subharmonic, while if $\Delta u \leq 0$ in Ω, u is called superharmonic.

Maximum principles, a start, collected by G. Sweers 2000 (rev.)

A positive operator on a Hilbert space is a linear operator A for which the corresponding quadratic form $(A\boldsymbol{x}, \boldsymbol{x})$ is nonnegative.

7.3.1 Elliptic case

We say that an elliptic equation $Iu = 0$ satisfies the comparison principle if the following statement is true.

Given two functions $u : \mathbb{R}^n \to \mathbb{R}$ and $v : \mathbb{R}^n \to \mathbb{R}$ such that u and v are upper and lower semicontinuous in $\overline{\Omega}$ respectively, where Ω is an open domain, $Iu \geq 0$ and $Iv \leq 0$ in the viscosity sense in Ω, and $u \leq v$ in $\mathbb{R}^n \setminus \Omega$, then $u \leq v$ in Ω as well.

7.3.2 Parabolic case

We say that a parabolic equation $u_t - Iu = 0$ satisfies the comparison principle if the following statement is true.

Given two functions $u : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ and $v : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ such that u and v are upper and lower semicontinuous in $[0, T] \times \overline{\Omega}$ respectively, $Iu \leq 0$ and $Iv \geq 0$ in the viscosity sense in $(0, T] \times \Omega$, and $u \leq v$ in $(\{0\} \times \mathbb{R}^n) \cup ([0, T] \times (\mathbb{R}^n \setminus \Omega))$, then $u \leq v$ in $[0, T] \times \Omega$ as well.

7.4 Lamé常数

教材的第3页:

- 在连续力学中, Lamé常数(也称为Lamé系数或Lamé参数)是由应变-应力关系中出现的λ和μ表示的两个材料相关量. 通常, λ 和μ 分别被分别称为Lamé的第一个参数和Lamé 的第二个参数.

- λ表示材料的压塑性, 等价于体弹性模量或杨氏模量. 不同环境下, 参数μ和λ的意义不同. 例如, μ在流体动力学中被称为流体的动力学粘度; 而在与弹性相关的环境中, μ称为剪切模量.

- 应力应变关系取决于材料的物理性质, 即屋子的本构特性, 故统称本构方程或本构关系.

- 正交曲线坐标系(α,β,γ)的Lamé 系数. 由坐标变换$\boldsymbol{x}=\boldsymbol{x}(\alpha,\beta,\gamma)$, 计算得曲线坐标$\alpha,\beta,\gamma$的切向的线元长度平方分别为

$$ds_1^2=h_1^2d\alpha^2,\ ds_1^2=h_2^2d\beta^2,\ ds_1^2=h_3^2d\gamma^2,$$

  其中

$$\begin{aligned}h_1&=\left((\frac{\partial x}{\partial\alpha})^2+(\frac{\partial y}{\partial\alpha})^2+(\frac{\partial z}{\partial\alpha})^2\right)^{1/2},\\ h_2&=\left((\frac{\partial x}{\partial\beta})^2+(\frac{\partial y}{\partial\beta})^2+(\frac{\partial z}{\partial\beta})^2\right)^{1/2},\\ h_3&=\left((\frac{\partial x}{\partial\gamma})^2+(\frac{\partial y}{\partial\gamma})^2+(\frac{\partial z}{\partial\gamma})^2\right)^{1/2}.\end{aligned}$$

7.5 对角占优矩阵

教材的第22页[14, Page 123]:

- 矩阵中每个主对角元素的模都大于与它同行的其他元素的模的总和, 这种矩阵就叫严格对角占优的; 对列同样成立.

- 设$A=(a_{ij})$. 如果$|a_{ii}|>\sum_{j=1,j\neq i}^n|a_{ij}|$, $i=1,2,\cdots,n$, 则称A为严格对角占优矩阵.

- 设$A=(a_{ij})$. 如果$|a_{ii}|\geq\sum_{j=1,j\neq i}^n|a_{ij}|$, $i=1,2,\cdots,n$, 且其中至少有一个式子取严格不等号, 则称A为弱对角占优矩阵.

- 若A是严格对角占优矩阵, 则线性代数方程组$A\boldsymbol{x}=\boldsymbol{b}$有解.

- 如果A为严格对角占优矩阵, 则A为非奇异矩阵.

- 若A为严格对角占优矩阵, 则雅克比迭代法、高斯-赛德尔迭代法和$0<\omega\leq 1$的超松弛迭代法均收敛.

7.6 不可约矩阵

教材的第22页:

- 定义1: 设$A \in \mathbb{R}^{n\times n}$, 若存在置换矩阵$P$, 使得

$$PAP^T = \begin{pmatrix} B & C \\ O & D \end{pmatrix},$$

其中B, D是阶数不小于1的方阵, O是零矩阵, 则称A 是可约矩阵(reducible matrix); 否则称矩阵A是不可约的(irreducible).

- 定义2: 对于n阶方阵$A = (a_{ij})$而言, 如果指标集$\{1, 2, \cdots, n\}$能够被划分成两个不相交的非空指标集J 和K, 使得对任意的$j \in J$和任意的$k \in K$都有$a_{jk} = 0$, 则称矩阵A是可约的; 否则是不可约的.

- A可约意即A可经过若干**行列重排**化为$\begin{pmatrix} B & C \\ O & D \end{pmatrix}$. 如果$A$经过两行交换的同时进行两列交换, 则称对$A$进行一次**行列重排**.

- $A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 3 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ 是可约的, 因为存在置换矩阵$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, 使得$PAP^T = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 3 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$.

- $A = \begin{pmatrix} 4 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix}$ 是不可约的.

- 如果A的所有元素都非零, 则A是不可约的.

- 如果A是可约的, 则$A\boldsymbol{x} = \boldsymbol{b}$可以化为两个低阶的方程组求解.

- 对角占优定理: 如果$A \in \mathbb{R}^{n\times n}$ 为严格对角占优阵或为不可约对角占优阵, 则A为非奇异阵.

  证明: 这里只证明严格对角占优阵情形. 利用反证法, 反设A为奇异阵, 则$A\boldsymbol{x} = \boldsymbol{0}$有非零解, 记为$\boldsymbol{x} = (x_1, \cdots, x_n)^T$. 则

$$|x_k| = \max_{1\le i\le n} \{|x_i|\} \neq 0.$$

  由齐次方程组$A\boldsymbol{x} = \boldsymbol{0}$的第$k$个方程$\sum_{j=1}^n a_{kj}x_j = 0$, 得

$$|a_{kk}x_k| = |\sum_{j=1, j\neq k}^n a_{kj}x_j| \le \sum_{j=1, j\neq k}^n |a_{kj}||x_j| \le \sum_{j=1, j\neq k}^n |a_{kj}||x_k|.$$

  由于$|x_k| \neq 0$, $|a_{kk}| \le \sum_{j=1, j\neq k}^n |a_{kj}|$, 这与假设矛盾. ∎

7.7 M矩阵

教材的第22页[14, Page 121]: Here, for any real matrices A, B of size $m \times n$, we write $A \ge B$ (or $A > B$) if $a_{ij} \ge b_{ij}$ (or $a_{ij} > b_{ij}$) for all i, j.

- 设$A = (a_{ij}) \in \mathbb{R}^{n\times n}$. 若$A$可表示为$A = sI - B$, 其中$B = (b_{ij}) \ge 0$即$b_{ij} \ge 0$, $\forall ij$, 则当$s > \rho(B)$ 时, 称A为非奇异的M矩阵, 简称M 矩阵.

- 若A满足

  $$a_{ij} \le 0,\ 1 \le i \ne j \le n;\quad a_{ii} > 0,\ i = 1, 2, \cdots, n,$$

  则称A为L矩阵.

- $A = (a_{ij}) \in \mathbb{R}^{n\times n}$是M矩阵的充要条件是$A$ 是L矩阵, 且$A^{-1} \ge 0$.

7.8 Positive matrix

A positive matrix is a real or integer matrix $(a)_(ij)$ for which each matrix element is a positive number, i.e., $a_(ij) > 0$ for all i, j.

Positive matrices are therefore a subset of nonnegative matrices.

Note that a positive matrix is not the same as a positive definite matrix.

7.9 适定性

教材的第33页:

适定性问题(Well-posed problem)来自于哈达玛(Hadamard)所给出的定义, 他认为物理现象中的数学模型应该具备下述性质: 存在解; 解是惟一的; 解连续地取决于初边值条件.

Problems that are not well-posed in the sense of Hadamard are termed ill-posed. Inverse problems are often ill-posed. For example, the inverse heat equation, deducing a previous distribution of temperature from final data, is not well-posed in that the solution is highly sensitive to changes in the final data.

7.10 采样定理

教材的第42页倒数第4行: "步长$1/N$的网格能分辨的最高频率为$k = N$": 类似的见[9, Page109].

采样定理, 又称奈奎斯特-香农采样定理(Nyquist-Shannon sampling theorem): 只要采样频率大于或等于被模拟的信号的最高频率的两倍, 采样值就可以包含原始信号的所有信息, 被采样的信号就可以不失真地还原成原始信号.

- 采样指的是理想采样, 即直接记录信号在某时间点的精确取值, 所以采样定理只涉及到了从连续信号到离散信号的理想采样过程, 而未涉及到对测量值的量化过程.

- 采样频率指单位时间内的采样点数, 采样是一种周期性的操作, 非周期性采样不在采样定理的范围之内.

Aliasing [混淆]: Nyquist – Shannon sampling theorem,
Aliasing

7.11 Spectral Accuracy

教材的第43页倒数第13行: "指数型增长": $y = ae^{kt}$ 指数型增长(exponential growth), 当$k > 0$时; 而y关于k指数型衰减(decay), 当$k < 0$时.

以下摘自[12, Chap4]

We are ready to discuss the accuracy of spectral methods. As stated in Chapter 1, the typical convergence rate is $O(N^{-m})$ for every m for functions that are smooth (fast!) and $O(c^N)$ $(0 < c < 1)$ for functions that are analytic (faster!). Such behavior is known as spectral accuracy.

To derive these relationships we shall make use of the Fourier transform in an argument consisting of two steps. First, a smooth function has a rapidly decaying transform. The reason is that a smooth function changes slowly, and since high wavenumbers correspond to rapidly oscillating waves, such a function contains little energy at high wavenumbers. Second, if the Fourier transform of a function decays rapidly, then the errors introduced by discretization are small. The reason is that these errors are caused by aliasing of high wavenumbers to low wavenumbers.

7.12 追赶法(Thomas algorithm)

教材的第50页第7行:

追赶法[25]的基本原理是矩阵的LU分解，即将三对角矩阵A分解为$A = LU$, 其中, L为一个对角线上元素为1的下三角矩阵，U为一个上三角矩阵, 容易验证, 一个三对角矩阵作LU分解以后, 得到一个下二对角矩阵与一个上二对角矩阵的乘积.

In numerical linear algebra, the tridiagonal matrix algorithm, also known as the Thomas algorithm (named after Llewellyn Thomas), is a simplified form of Gaussian elimination that can be used to solve tridiagonal systems of equations.

A tridiagonal system for n unknowns may be written as

$$a_i x_{i-1} + b_i x_i + c_i x_{i+1} = d_i,$$

where $a_1 = 0$ and $c_n = 0$.

$$\begin{pmatrix} b_1 & c_1 & & & 0 \\ a_2 & b_2 & c_2 & & \\ & a_3 & b_3 & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ 0 & & & a_n & b_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_n \end{pmatrix}.$$

For such systems, the solution can be obtained in $O(n)$ operations instead of $O(n^3)$ required by Gaussian elimination.

A first sweep eliminates the a_i's, and then an (abbreviated) backward substitution produces the solution.

Thomas' algorithm is not stable in general, but is so in several special cases, such as when the matrix is diagonally dominant (either by rows or columns) or symmetric positive definite; for a more precise characterization of stability of Thomas' algorithm, see Higham Theorem 9.12 [26]. If stability is required in the general case, Gaussian elimination with partial pivoting (GEPP) is recommended instead [27].

`Thomas' algorithm`:

- Forward sweep:

$$\left.\begin{aligned} w_i &= \frac{a_i}{b_{i-1}}, \\ b_i &:= b_i - w_i c_{i-1}, \\ d_i &:= d_i - w_i d_{i-1}, \end{aligned}\right. \qquad i = 2, 3, \ldots, n.$$

- Back substitution

$$\begin{aligned} x_n &= \frac{d_n}{b_n}, \\ x_i &= \frac{d_i - c_i x_{i+1}}{b_i}, \quad i = n-1, n-2, \ldots, 1. \end{aligned}$$

http://web.mit.edu/18.06/www/Course-Info/Mfiles/tridiag.m

[25] L.H. Thomas, Elliptic problems in linear differential equations over a network, Watson Science Computer Laboratory Report, 1949.

[26] N.J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd Edition, SIAM, 2002, p. 175.

[27] B.N. Datta, *Numerical Linear Algebra and Applications*, 2nd Edition, SIAM, 2010, p. 162

7.13 Softwares

- Gnuplot is a portable command-line driven graphing utility for Linux, OS/2, MS Windows, OSX, VMS, and many other platforms.
- Gmsh is a free 3D finite element mesh generator with a built-in CAD engine and post-processor. Its design goal is to provide a fast, light and user-friendly meshing tool with parametric input and advanced visualization capabilities.
- PETSc is a library for large scale numerical linear algebra but it also has support for finite elements, time stepping schemes, mesh management, etc. The three main methods for solving equations are:
  - KSP: Methods for solving matrix equations, i.e., linear solvers
  - SNES: Newton method for solving non-linear equations
  - TS: Time stepping schemes

  http://cpraveen.github.io/teaching/petsc.html
- FEniCS is a popular open-source (LGPLv3) computing platform for solving PDEs.
- FreeFEM offers a fast interpolation algorithm and a language for the manipulation of data on multiple meshes.
- FreeFem++ is a PDE solver. It has its own language. FreeFem scripts can solve multiphysics non linear systems in 2D and 3D.

References

[1] L.C. Evans, *Partial Differential Equations*, 2nd ed., American Mathematical Society, 2010. 1st ed.

[2] D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd ed., Springer-Verlag, 1983.

[3] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, 2011. PDF

[4] W. Hackbusch, *Eliptic Differential Equations: Theory and Numerical Treatment*, Translated from the German by Regine Fadiman and Patrick D.F. Ion, Springer, 2003.

[5] Z.P. Li, *Lecture on Numerical Solutions of Partial Differential Equations*, PKU Publisher, 2010. (in Chinese)

[6] J.L. Lions, *Lectures on Elliptic Partial Differential Equations*, Tata Institute of Fundamental Research, Bombay, 1957 http://www.math.tifr.res.in/~publ/ln/tifr10.pdf

[7] E. Miersemann, *Partial Differential Equations*, https://math.libretexts.org/Bookshelves/Differential_Equations/Book%3A_Partial_Differential_Equations_(Miersemann) http://www.math.uni-leipzig.de/~miersemann/pde2book.pdf

[8] A.D. Polyanin et al., *Partial Differential Equation*, Scholarpedia, 3(10), 2008, 4605.

[9] G. Söderlind, *Numerical Methods for Differential Equations: An Introduction to Scientific Computing*, Springer, December 5, 2017.

[10] J.W. Thomas, *Numerical Partial Differential Equations: Finite Difference Methods*, Springer, 1995.

[11] D.H. Yu and H.Z. Tang, *Numerical Solutions of Differential Equations* (2nd edition), Science Press, 2018.

[12] L.N. Trefethen, *Spectral Methods in Matlab*, SIAM, Philadelphia, 2000.

[13] L.N. Trefethen, *Finite Difference and Spectral Methods for Ordinary and Partial Differential Equations*, unpublished text, 1996, available at `http://people.maths.ox.ac.uk/trefethen/pdetext.html`

[14] 徐树方, 矩阵计算的理论与方法, 北京大学出版社, 1995.

[15] 余德浩和汤华中, 微分方程数值解法, 第二版, 科学出版社, 2018.

[16] J. Nordström, Numerical Solution of Initial Boundary Value Problems (MAI0122), Linköping University. Lecture Notes

[17] J. Nordström, A roadmap to well posed and stable problems in computational physics, *J. Sci. Comput.*, 71(2017), 365-385.

[18] B. Gustafsson, *High Order Difference Methods for Time Dependent PDE*, Springer-Verlag, 2008.

[19] B. Gustafsson, H.-O. Kreiss, and J. Oliger, *Time Dependent Problems and Difference Methods*, 2nd ed., John Wiley & Sons, 2013 (第8-10章PDE情形, pp.249; 第11 章差分近似, P339). (§5.3 in 1st ed., 1995, pp.182-194).

[20] P.D. Lax, The scope of the energy method, *Bulletin of the American Math. Soc.*, 66(1960), 32-35.

[21] B. Gustafsson, *Fundamentals of Scientific Computing*, Springer-Verlag, 2011.

[22] B. Gustafsson, *Scientific Computing: A Historical Perspective*, Springer-Verlag, 2018.