Advanced computational methods
X071521-Lecture 2

1 Elliptic equation and the 5-point scheme

Consider the 2D poisson equation

$$-\Delta u = f, \quad \Omega = [0,1] \times [0,1],$$
$$u = g, \quad \partial\Omega.$$

A first way to approximate the Laplacian: 5 point stencil.
For simplicity, we use uniform step size for both directions:

$$\Delta x = \Delta y = h = 1/(m+1).$$

u_{ij} represents the value at $x = x_i = ih, y = y_j = jh$. We approximate $\Delta_h = D_x^2 + D_y^2$:

$$-\Delta_h u_{ij} = -\frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{h^2} - \frac{u_{i,j+1} - 2u_{ij} + u_{i,j-1}}{h^2} = f_{ij} = f(x_i, y_j).$$

This is called 5-point stencil since we only used five points.

To set up the matrix, we must order the points. The most straightforward way is to order them as follows:

$$u = (u_{11}, u_{21}, \ldots, u_{m1}, u_{12}, u_{22}, \ldots, u_{m2}, \ldots, u_{mm}).$$

This is convenient in Matlab since this actually corresponds to reshaping by columns of matrices. Other ordering may be possible to make the matrix even sparser.

How do we set up the matrix? The most convenient way is to introduce

$$\vec{u}_j = (u_{1j}, u_{2j}, \ldots, u_{mj}).$$

Consider

```
e = ones(m,1);
A1 = spdiags( [e -2*e e], -1:1, m, m);
```

Then, we have

$$-\frac{1}{h^2} A_1 \vec{u}_j - \frac{1}{h^2}(\vec{u}_{j+1} - 2\vec{u}_j + \vec{u}_{j-1}) = \vec{f}_j$$

$f_{i1} = f(x_i, y_1) + \frac{1}{h^2} g(x_i, 0)$, $f_{im} = f(x_i, y_m) + \frac{1}{h^2} g(x_i, 1)$ and $f_{ij} = f(x_i, y_j)$ for others. From this equation, it's clear that the big matrix M has $m * m$ blocks. The (p, p) block is $-\frac{1}{h^2}(A_1 - 2I)$ and the $(p, p-1)$ and $(p, p+1)$ blocks are $-\frac{1}{h^2} I$. Note that $\vec{u}_0$ and $\vec{u}_{m+1}$ can be determined by the boundary values and can be moved to right hand side.

The big matrix can be constructed using

```
A1=spdiags(ones(m,1)*[1 -2 1], -1:1, m, m);
M=-(kron(A1, speye(m))+kron(speye(m), A1))/h^2;
```

1.1 Analysis of the scheme

Consistency: the LTE is defined to be

$$\tau_{ij} = -\frac{u(x_{i+1}, y_j) - 2u(x_i, y_j) + u(x_{i-1}, y_j)}{h^2} - \frac{u(x_i, y_{j+1}) - 2u(x_i, y_j) + u(x_i, y_{j-1})}{h^2} - f(x_i, y_j),$$

which measures how the true solution satisfies the numerical approximation. Direct Taylor expansion shows that $\tau_{ij} = O(h^2)$. As before, consistency means $\|\tau\| \to 0$ as $h \to 0$.

By the analysis here, we have

Theorem 1. *Suppose that the exact solution $u \in C^4(\Omega)$. Then, there exists $h_0 > 0$, $C_1 > 0, C_2 > 0$ such that for all $h < h_0$:*

$$\|\tau\|_{\ell^2} \leq C_1 h^2$$

and

$$\|\tau\|_{\ell^\infty} \leq C_2 h^2.$$

1.2 Analysis of the scheme: ℓ^2 stability and convergence

One way is to check the eigenvalues. Here we provide **another proof**, which is interesting itself. Consider the case $g = 0$ for simplicity. $g \neq 0$ case can be shown as well.

The method here is analogy to the continuous PDE:

$$-\Delta u = f, \ x \in \Omega, \quad u = 0, \ x \in \partial\Omega.$$

Multiplying u and taking integral, we have

$$\int |\nabla u|^2 \, dx = \int f u \, dx \leq \|f\|_2 \|u\|_2.$$

Then, we need the Poincare inequality:

$$\|u\|_2 \leq C(\Omega)\|\nabla u\|_2.$$

Then, we get

$$\frac{1}{C^2(\Omega)}\|u\|_2^2 \leq \|f\|_2\|u\|_2, \Rightarrow \|u\|_2 \leq C^2(\Omega)\|f\|_2.$$

Now, we move onto the discrete case following similar ideas:

Theorem 2. *Consider $g = 0$, and define $\|w\|_2 = \sqrt{h^2 \sum_{i,j} |w_{ij}|^2}$. Then,*

$$\|u\|_2 \leq \frac{1}{2}\|\Delta_h u\|_2 = \frac{1}{2}\|f\|_2.$$

In other words, the 5-point stencil is ℓ^2 stable.

The proof follows from the following two lemmas. The first is the discrete Poincare inequality:

Lemma 1. *If $g = 0$, then*

$$\|u\|_2 \leq \frac{1}{2}(\|D_{+,x}u\|_2 + \|D_{+,y}u\|_2)$$

where $D_{+,x}u_{ij} = (u_{i+1,j} - u_{ij})/h$ and $D_{+,x}u_{m+1,j} = 0$. $D_{+,y}$ is defined similarly

The second is the discrete Green's identity:

Lemma 2. *Suppose both v and w vanish on the boundary, then*

$$-\langle \Delta_h v, w\rangle = \langle D_{+,x}v, D_{+x}w\rangle + \langle D_{+,y}v, D_{+y}w\rangle,$$

where

$$\langle v, w\rangle = h^2 \sum_{i=0}^{m+1}\sum_{j=0}^{m+1} v_{ij}w_{ij}.$$

Proof. For the Poincare:

$$\begin{aligned}|u_{ij}|^2 = |\sum_{p=i}^{m}(u_{p+1,j} - u_{pj})|^2 &= (\sum_{p=i}^{m} h|D_{+,x}u_{pi}|)^2 \\ &\leq (\sum_{p=i}^{m} h|D_{+,x}u_{pj}|^2)(\sum_{p=i}^{m} h) \leq \sum_{p=0}^{m} h|D_{+,x}u_{pj}|^2.\end{aligned}$$

Hence,

$$\|u\|_2^2 = \sum_{i,j} h^2 |u_{ij}|^2 \le \sum_{i,j} h^3 \sum_{p=0}^{m} |D_{+,x} u_{pj}|^2 = \sum_{p,j} h^2 |D_{+,x} u_{pj}|^2.$$

or

$$\|u\|_2 \le \|D_{+,x} u\|_2.$$

Similarly,

$$\|u\|_2 \le \|D_{+,y} u\|_2.$$

Adding these two yields the desired inequality.

For the discrete Green's identity:

$$\begin{aligned} -\langle \Delta_h v, w\rangle &= -\sum_{i=0,j=0}^{m+1} h^2 \Delta_h v_{ij} w_{ij} = -\sum_{i=1}^{m}\sum_{j=1}^{m} h^2 \Delta_h v_{ij} w_{ij} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} (2v_{ij} - v_{i+1,j} - v_{i-1,j}) w_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{m} (2v_{ij} - v_{i,j+1} - v_{i,j-1}) w_{ij}. \end{aligned}$$

The first term equals

$$\begin{aligned} &\sum_{i=1}^{m}\sum_{j=1}^{m} (v_{ij} - v_{i-1,j}) w_{ij} - \sum_{i=1}^{m}\sum_{j=1}^{m} (v_{i+1,j} - v_{i,j}) w_{ij} \\ &\quad = \sum_{i=0}^{m-1}\sum_{j=1}^{m} (v_{i+1,j} - v_{ij}) w_{i+1,j} - \sum_{i=1}^{m}\sum_{j=1}^{m} (v_{i+1,j} - v_{i,j}) w_{ij} \\ &\quad = \sum_{i=0}^{m}\sum_{j=0}^{m} (v_{i+1,j} - v_{ij}) w_{i+1,j} - \sum_{i=0}^{m}\sum_{j=0}^{m} (v_{i+1,j} - v_{i,j}) w_{ij} \\ &\quad = \sum_{i=0}^{m}\sum_{j=0}^{m} h^2 D_{+,x} v_{ij} D_{+,x} w_{ij} = \langle D_{+,x} v, D_{+x} w\rangle. \end{aligned}$$

The second term is similarly computed. Hence, the claim discrete Green's identity holds. □

The theorem is easy to prove now using these two lemmas:

$$\|u\|_2^2 \le \frac{1}{4}(\|D_{+,x} u\|_2 + \|D_{+,y} u\|_2)^2 \le \frac{1}{2}(\|D_{+,x} u\|_2^2 + \|D_{+,y} u\|_2^2) = \frac{1}{2}\langle -\Delta u, u\rangle \le \frac{1}{2}\|\Delta_h u\|_2 \|u\|_2.$$

Corollary 1. $\|E\|_2 = \|u - \hat{u}\|_2 \to 0$ *as* $h \to 0$ *where* $\hat{u}$ *consists of the true values at the grid points.*

1.3 Analysis of the scheme: ℓ^∞ stability and convergence

Here, we focus on the l^∞-stability (not in the book). The goal is then to show that there exists C independent of h such that

$$\|u\|_\infty \leq C(\|f\|_\infty + \|g\|_\infty).$$

To prove this, we first show the **discrete maximum principle**:

Theorem 3. *Let Ω_h be the set of all interior points, i.e. $\Omega_h = \{(x_i, y_j)\}\backslash\partial\Omega$. Let Γ_h be the grid points on $\partial\Omega$.*

Suppose $\Delta_h u_{i,j} \geq 0$ for all $(x_i, y_j) \in \Omega_h$. Then $\max_{\Omega_h} u \leq \max_{\Gamma_h} u$. Further, if $\max_{\Omega_h} u = \max_{\Gamma_h}$, then u is a constant.

Proof. The first condition implies

$$u_{ij} \leq \frac{1}{4}(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}).$$

Suppose $\max_{\Omega_h} u \geq \max_{\Gamma_h} u$ and the maximum is achieved at (x_{i^*}, y_{j^*}) which is inside. The values at all the neighbors are no bigger than u_{i^*,j^*}. Since it is the maximum, by the inequality, they must be equal. For the interior point in the neighbors, the same argument applies. Then, the values at all interior points and their neighbors are equal. This means u is a constant. Hence, the bigger sign never holds and the claim follows. □

Theorem 4. *The 5-point stencil has a unique solution for any f, g and it is ℓ^∞-stable.*

Proof. For the uniqueness, suppose there are two solutions u_1 and u_2. Then, $\Delta_h^2(u_1 - u_2) = 0$ and the boundary values of $u_1 - u_2$ are zero. The discrete maximum principle implies that $u_1 - u_2 \leq 0$ for all interior points. Then, switching the roles of u_1 and u_2, we have $u_2 - u_1 \leq 0$. Hence, $u_1 = u_2$.

For the l^∞ stability, consider an auxiliary function ϕ such that $\Delta_h \phi = 1$. Then,

$$\Delta_h(u + \phi\|f\|_\infty) = -f + \|f\|_\infty \geq 0.$$

The discrete maximum principle implies that

$$u + \phi\|f\|_\infty \leq \max_{\Gamma_h}(g + \phi\|f\|_\infty) \Rightarrow u \leq \|g\|_\infty + 2\|\phi\|_\infty\|f\|_\infty.$$

Then, one applies the same argument for $-u$. The claims follows. To finish the proof, we must show that ϕ exists. One example is

$$\phi = \frac{1}{4}((x - \frac{1}{2})^2 + (y - \frac{1}{2})^2).$$

□

Corollary 2. $\|E\|_\infty = \|u - \hat{u}\|_\infty \to 0$ *as* $h \to 0$ *where* $\hat{u}$ *consists of the true values at the grid points.*

2 Schemes for ODEs

The discretization of ODEs is very important for time discretization of evolutionary PDEs.

Consider the ODE

$$u'(t) = f(t, u(t)), \quad u(0) = u_0,$$

where u could be a vector valued function. Any ODE can be reduced to a first order system, so this is general enough.

The ODE solvers are all approximations to

$$u(t_{n+1}) = u(t_n) + \int_{t_n}^{t_{n+1}} f(s, u(s))ds.$$

$\int_{t_n}^{t_{n+1}} f(s, u(s))$ will be approximated by data $u_0, u_1, \ldots, u_{n+1}$.

(Sections 5.3-5.9 in Leveque's book.)

2.1 One-step method

- If we approximate $f(s, u(s)) \approx f(t_n, u^n)$, then we have the forward Euler:
$$u^{n+1} = u^n + kf(t_n, u^n)$$

- $f(s, u(s)) \approx f(t_{n+1}, u^{n+1})$, we have the backward Euler:
$$u^{n+1} = u^n + kf(t_{n+1}, u^{n+1})$$

- $f(s, u(s)) \approx \frac{1}{2}(f(t_n, u^n) + f(t_{n+1}, u^{n+1}))$, then we have the trapezoidal method:
$$u^{n+1} = u^n + \frac{k}{2}(f(t_n, u^n) + f(t_{n+1}, u^{n+1}))$$

- Runga-Kutta methods (multi-stage, one-step methods)

  The idea of RK is to approximate the integral with more grid points so that it is more accurate:
$$\int \approx k \sum_{j=1}^{r} b_j f(t_n + \lambda_j, y(t_n + \lambda_j))$$

Motivated by this formula, we can write the general r-stage Runga-Kutta method is

$$u^{n+1} = u^n + k\sum_{j=1}^{r} b_j f(t_n + c_j k, Y_j),$$

$$Y_i = u^n + k\sum_{j=1}^{r} a_{ij} f(t_n + c_j k, Y_j), i = 1, 2, \ldots, r$$

Some necessary conditions for the r-th order accuracy:

- Y_i is an approximation of the value at $t_n + c_i k$. Hence, $\sum_{j=1}^{r} a_{ij} = c_i$.
- We apply the method to the model problem with $f(t, u) = \lambda u$. Then,

  $$Y_i = u^n + k\sum_{j=1}^{r} a_{ij}\lambda Y_j, \qquad u^{n+1} = u^n + k\sum_{j=1}^{r} b_j \lambda Y_j$$

  However, we know that $u(t_{n+1}) = e^{\lambda k} u(t_n)$. Note

  $$e^{\lambda k} = \sum_{n\geq 0} \frac{(\lambda k)^n}{n!}$$

  We can therefore solve Y_j out in the first equation and determine the coefficients in

  $$\sum_{n\geq 0} \frac{(\lambda k)^n}{n!} u_n = u^n + k\sum_{j=1}^{r} b_j \lambda Y_j,$$

  by comparing the powers of k.

The most frequently used schemes are $RK2, RK3, RK4$. In general, they are not unique. For example, there are two typical $RK2$ schemes:

$$u_{n+1} = u_n + \frac{1}{2}(f(t_n, u_n) + K_2),$$
$$K_2 = f(t_n + k, u_n + kK_1)$$

and

$$u_{n+1} = u_n + kK_2,$$
$$K_1 = f(t_n, u_n), \quad K_2 = f(t_n + k/2, u_n + kK_1/2)$$

All the above methods are the one-step method because we only use u^n to compute u^{n+1}.

2.2 LMM

Linear Multistep Methods (LMM) are another class of ODE solvers. The solvers involve the values at several steps. The most frequently used are the Adams methods

$$u^{n+r} = u^{n+r-1} + k\sum_{j=0}^{r} \beta_j f(t_{n+j}, u^{n+j}).$$

If $\beta_r = 0$, we have the Adams-Bashforth methods.

3 Local truncation error and consistency for schemes of ODEs

3.1 Basic concepts

The consistency is measured by the local truncation error (LTE) where u^n is replaced by $u(t_n)$,:

$$LTE = \frac{1}{k}(LHS - RHS)$$

note that we have divided k here because $\frac{u^{n+1}-u^n}{k}$ is in the same order as the derivative. This is different from the so-called one-step error.

Example: For the forward Euler (FE):

$$\tau_n = \frac{1}{k}(u(t_{n+1}) - u(t_n) - kf(t_n, u(t_n))) = \frac{1}{k}(u(t_{n+1}) - u(t_n) - ku'(t_n)) = O(k).$$

The ODE solvers are said to be consistent if the local truncation error goes to zero as $k \to 0$.

The **order** of the method is the order of the LTE. Direct Taylor expansion shows that the two Euler methods are first order while the trapezoidal method is a second order method.

An ODE solver is convergent if for a problem $u' = f(t, u)$ where f is continuous and Lipschitz continuous in u on $[0, T]$, we have

$$\lim_{k\to 0, nk=T} |u^n - u(T)| = 0,$$

where T is in the largest interval of existence.

f is Lipschitz in u means

$$\sup_{0\le t\le T} |f(t, u_1) - f(t, u_2)| \le L(T)|u_1 - u_2|$$

3.2 Convergence for one-step method

Claim:

For one step solvers, $u^{n+1} = u^n + k\Psi(u^n, t_n, k)$, as long as Ψ is continuous and Lipschitz continuous in u, the solver is stable. Here, 'stable' means that the global error introduced by the m-th step error will not be amplified too much. If further it is consistent, then it is convergent.

We take the forward Euler as the example. Again f is assumed to be Lipschitz.

Let τ_j be the local truncation error. Then, the one step error is

$$u(t_{j+1}) - u(t_j) - kf(t_j, u(t_j)) = k\tau_j = O(k^2).$$

Let $E^j = |u(t_j) - u_j|$. Then, we have

$$E^{j+1} \leq E^j + k|f(t_j, u(t_j)) - f(t_j, u_j)| + k|\tau_j| \leq E^j + kLE^j + Ck^2.$$

Then,

$$E^n \leq E^0 e^{nkL} + \sum_{j=0}^{n-1} k|\tau_{n-1-j}|(1+kL)^j \leq Ce^{TL}k$$

In this proof, we have implicitly use the stability. The error term $k\tau_j$ is amplified by $(1+kL)^{n-j}$ which is uniformly bounded by e^{TL}, so we have stability.

This verifies the claim.