Math 6008 Numerical PDEs–Lecture 4
Stability by Fourier analysis

Instructor: Lei Li, INS, Shanghai Jiao Tong University;
Email: leili2010@sjtu.edu.cn

The theory of stability region gives some intuition why the method is stable or unstable, but is not quite rigorous (though can be made rigorous). In this lecture, we focus on a rigorous analysis for **linear equations with constant coefficients**–the von Neumann analysis or the Fourier analysis.

1 Fourier transform

The Fourier transform decomposes functions into a series of planar waves (modes). The coefficients of the waves can be viewed as function of the frequency so that one can view the functions in the so-called "frequency domain"（频域）. The Fourier transform can usually give much insight for linear problems.

1.1 Continuous Fourier Transform

Consider 1D functions as examples. There are several conventions to define Fourier transforms. For example,

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-i\xi x}\,dx, \tag{1}$$

or

$$\hat{f}(\xi) = \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} f(x) e^{-i\xi x}\,dx,$$

or

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i\xi x}\,dx.$$

These are just different scalings. The motivation is that the planar waves (平面波) $\{e^{i\xi x}\}$ can form a basis for the integrable functions so that any reasonably good function $f(x)$ can be written as the superposition of them:

$$f(x) = \int c(\xi) e^{i\xi x} d\xi.$$

Using the orthogonality of these waves $\int e^{i(\xi - \xi')x} dx = 2\pi\delta(\xi - \xi')$, one can then find $c(\xi)$ as

$$c(\xi) = \frac{1}{2\pi} \int f(x) e^{-i\xi x} dx.$$

In analysis, the convention $e^{-2\pi i\xi x}$ is used more often. Here, we will use the traditional convention (1). With this the inverse Fourier transform is given by

$$f(x) = \frac{1}{2\pi} \int \hat{f}(\xi) e^{i\xi x} d\xi.$$

Then, $\hat{f}(\xi)/2\pi = c(\xi)$ is the expansion coefficient for f to be written as linear combination of the planar waves.

Remark 1. *In the signal analysis, the input $h(t)$ is a signal and $t \in (-\infty, \infty)$ is the time. Then, using the above convention, $\xi = \omega = 2\pi\nu$ is called the angular frequency. ν is the frequency. Hence, whether one has $2\pi\xi$ or ξ is to use frequency or angular frequency.*

Definition 1. *Convolution for two real functions: $g * h = \int_{-\infty}^{\infty} g(\tau)h(t - \tau)d\tau$. (For complex-valued functions, one sometimes define $g*h = \int_{-\infty}^{\infty} g(\tau)\bar{h}(t - \tau)d\tau$ but most people use the usual convention).*

It's easy to prove the property:

- $\widehat{g * h} = \bar{g}\bar{h}$.
- (Parsavel theorem)$\int_{-\infty}^{\infty} |h(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 d\xi$

1.2 Fourier series

If the function $f(x)$ is periodic (suppose the period is 2π), then the Fourier transform diverges. Instead, we can limit the integration onto one period:

$$\hat{f}_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx.$$

In this case, for e^{ikx} to have period 2π, we need k to be integers. Moreover, it can be shown that the basis $\{e^{ikx}\}_{k\in\mathbb{Z}}$ is enough to represent any reasonable functions with period 2π

$$f(x) = \sum_k \hat{f}_k e^{ikx}.$$

Hence we conclude that: a periodic function in spatial domain gives discrete Fourier coefficients (discrete frequency domain). In other words, bounded spatial domain gives discrete frequency domain.

2 Discrete Fourier Transform

Some detailed materials can be found in the book "spectral methods in MATLAB".

2.1 Unbounded grids and the semidiscrete Fourier transform

Consider $u(x)$, in most common situations, $u(x)$ is sampled (suppose) at evenly spaced intervals of size $h > 0$ so that we have a function defined on the grid $h\mathbb{Z}$. Then, we have some samples u_j.

The semidiscrete Fourier transform is defined by

$$\hat{u}(\xi) = h \sum_{j=-\infty}^{\infty} u_j e^{-i\xi x_j}.$$

It can be verified easily that

$$\hat{u}_\xi = \hat{u}_{\xi+2\pi/h}.$$

Hence, we can restrict ξ to be in $[-\pi/h, \pi/h)$.

We see another feature of Fourier transform:**the discrete spatial domain gives bounded frequency domain and unbounded spatial domain gives continuous frequency domain**. The inverse transform is

$$u_j = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{i\xi x_j} \hat{u}(\xi) d\xi.$$

Here, once we do sampling using discrete points, we can have the issue of "aliasing"(混叠). This means that we can only see frequencies with width at most $2\pi/h$. For example, if we take $h = 1/4$ and do sampling for $\sin(\pi x)$ and $\sin(9\pi x)$. Then, the two functions are identical on this grid $h\mathbb{Z}$ and we will count the waves from $\sin(9\pi x)$ as waves with frequency $\xi = \pi$. This is why the semi-discrete transform only has $\xi \in [-\pi/h, \pi/h)$ because other signals with higher frequencies are projected down to this range!

2.2 Discrete Fourier Transform (DFT)

Consider N (an even number) points $v_0, v_2, \ldots, v_{N-1}$. This can be viewed as the samples from a periodic function. Comparing to the Fourier series and the semidiscrete Fourier transform, we can then define the DFT (离散傅里叶变换) as

$$\hat{v}_k = \sum_{n=0}^{N-1} e^{-ikx_n} v_n, \quad k \in \mathbb{Z}$$

where $x_n = \frac{2\pi}{N} n$. (You can regard v as a function on $[0, 2\pi)$ and x_n is a sample point on the interval.)

- *In Matlab, the summation is $\sum_{n=1}^{N} e^{-ikx_{n-1}} v_n$, which will not cause a trouble, since the index in MATLAB starts with* 1.
- *Other people also use the definition as $\sum_{n=0}^{N-1} e^{-ikx_n} v_n$. Of course, this is equivalent to ours if we regard x_N as x_0.*
- *Different definitions (shifting the sequence) may result in a uniform factor phase for the DFT. As long as we have consistent inverse formula, it is just fine.*

Introducing $\omega = e^{-2\pi i/N}$, one can see that the DFT matrix is given by

$$M = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega & \dots & \omega^{(N-1)} \\ 1 & \omega^2 & \dots & \omega^{2(N-1)} \\ \dots & \dots & \dots & \dots \\ 1 & \omega^{N-1} & \dots & \omega^{(N-1)^2} \end{pmatrix} \tag{2}$$

Clearly, M is a dense matrix, and direct computation of DFT is $O(N^2)$.

Properties

- $\hat{v}_k = \hat{v}_{k+mN}$. This implies that we can choose k to be in $k = 0, 1, 2, \cdots, N-1$ or $-N/2+1, \cdots, N/2$. In Matlab, the output is for $k = 0, 1, 2, \dots, N-1$ or $k = 0, 1, 2, \dots, N/2, -N/2+1, \dots, -1$.
- If v is real valued, then $\hat{v}_k = \bar{\hat{v}}_{-k}$, where the bar means complex conjugate. If follows that $\hat{v}_{N/2}$ is real.
- Define the discrete convolution:
$$w_n = (u * v)_n = \sum_{j=1}^{N} u_j v_{n-j}$$
where $v_{n-j} = v_{N+n-j}$ if $n - j < 0$. Then,
$$\hat{w}_k = \hat{u}_k \hat{v}_k.$$
- The discrete Parseval equality holds:
$$\sum_n u_n \overline{v}_n = \frac{1}{N} \sum_{k=-N/2+1}^{N/2} \hat{u}_k \overline{\hat{v}_k}$$

Remark 2. *Some people want to define* $\hat{v}_k = \frac{1}{N} \sum_{n=1}^{N} e^{-ikx_n} v_n$, *then the Parseval is*

$$h \sum_n u_n \overline{v}_n = 2\pi \sum_{k=-N/2+1}^{N/2} \hat{u}_k \overline{\hat{v}_k},$$

where $h = 2\pi/N$.

Inverse DFT

The inverse DFT is given by

$$v_n = \frac{1}{N}\sum_{k=0}^{N-1} e^{ikx_n}\hat{v}_k = \frac{1}{N}\sum_{k=-N/2+1}^{N/2} e^{ikx_n}\hat{v}_k.$$

(In Matlab convention, we should have $v_n = \frac{1}{N}\sum_{k=0}^{N-1} e^{ikx_{n-1}}\hat{v}_k$, as v_n is associated with x_{n-1}.)

证明. First of all, $e^{-ik(x_n - x_j)} = e^{-ikx_{n-j}}$,

$$\begin{aligned}\hat{w}_k = \sum_n \sum_j u_j v_{n-j} e^{-ikx_n} &= \sum_j u_j \sum_n v_{n-j} e^{-ikx_n} \\ &= \sum_j u_j e^{-ikx_j} \sum_n v_{n-j} e^{-ikx_{n-j}} = \sum_j u_j e^{-ikx_j}\hat{v}_k = \hat{u}_k \hat{v}_k.\end{aligned}$$

For the second,

$$\frac{1}{N}\sum_{k=0}^{N-1} e^{ikx_n}\hat{v}_k \;=\; \frac{1}{N}\sum_{k=0}^{N-1} e^{ikx_n}\sum_m v_m e^{-ikx_m} \;=\; \frac{1}{N}\sum_m v_m \sum_k e^{ik(x_n - x_m)}$$

Clearly, $\sum_k e^{ik(x_n - x_m)} = \sum_k e^{ik(n-m)h}$. If $n = m$, the sum is N. If $n \neq m$, the sum equals zero by the geometric sum since $[e^{ik(n-m)h}]^N = 1$. Hence,

$$\frac{1}{N}\sum_m v_m \sum_k e^{ik(x_n - x_m)} = \frac{1}{N}\sum_m v_m N \delta_{n-m} = v_n.$$

$\square$

2.3 FFT: fast algorithms to compute DFT

The fast Fourier tranform (快速傅里叶变换, FFT) is an algorithm to compute DFT eifcfiently. The most frequently used one is the Cooley-Tukey algorithm (the algorithm was independently discovered also by Gauss).

The idea is based on the simple fact. Let N be even, then

$$\sum_{n=1}^{N} u_n e^{-ikn\frac{2\pi}{N}} = \sum_{m=1}^{N/2} u_{2m} e^{-ikm\frac{2\pi}{(N/2)}} + e^{ik\frac{2\pi}{N}} \sum_{m=1}^{N/2} u_{2m-1} e^{-ikm\frac{2\pi}{(N/2)}}$$

The DFT of an array of size N is reduced to 2 DFT of arrays with size $N/2$ plus extra N operations. By this way, the whole complexity is $O(N \log N)$.

3 Von Neumann-analysis for linear PDEs (Fourier analysis)

[The content in this section is a combination of the textbook and the book by Leveque. In particular, we make use of the semi-discrete Fourier transform instead of the continuous Fourier transform of the extended function. Of course, there is no intrinsic difference.]

We take the norm to be the ℓ^2 norm and discuss the ℓ^2 stability. Here, we mainly consider an unbounded domain (无界区域). If the domain is with periodic boundary condition （周期边界条件） or other boundary conditions, the discussion can be performed similarly.

For illustration, we only conisder $x \in \mathbb{R}$. The grid points are given by

$$x_j = jh : j : -\infty \to \infty.$$

We have u_j defined at x_j. The semi-discrete Fourier transform is then given by

$$\hat{u}(\xi) = h \sum_{j=-\infty}^{\infty} u_j e^{-ix_j \xi} = h \sum_{j=-\infty}^{\infty} u_j e^{-ijh\xi}, \xi \in [-\pi/h, \pi/h].$$

Then, we can recover u_j by

$$u_j = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} \hat{u}(\xi) e^{ijh\xi} d\xi.$$

We have the Parseval's equality:

$$\|u\|_2 = \sqrt{h \sum_j u_j^2} = \frac{1}{2\pi} (\int_{-\pi/h}^{\pi/h} |\hat{u}(\xi)|^2 d\xi)^{1/2} = \frac{1}{2\pi} \|\hat{u}\|_2.$$

The ℓ^2 stability requires that $\|u^n\|_2$ is bounded. Hence, it is enough to check the L^2 norm of $\hat{u}(\xi)$.

Usually, for the FDM of a linear PDE, the Fourier modes $e^{ix_j\xi}$ are decoupled. Just like the dispersion relation for linear PDE, for the discrete case, we'll have

$$\hat{u}^{n+1}(\xi) = G(\xi, \tau, h)\hat{u}^n(\xi).$$

To figure out $G(\xi, \tau, h)$, we can simply assume that $u_j^n = v_n(\xi)e^{ix_j\xi}$ and then compute $u_j^{n+1} = v_{n+1}(\xi)e^{ix_j\xi}$ so that $v_{n+1} = G(\xi, \tau, h)v_n$. This works since u_j^n is just the superposition of these modes. We can check what happens for each mode.

Below, we look at some examples.

Example 1. *Consider the advection equation $u_t + au_x = 0$ and its scheme*

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

Now, for the modes $e^{ix_j\xi}$, one can easily find that

$$v_{n+1} = v_n - a\lambda(v_n - v_n e^{-ih\xi}) \Rightarrow G(\xi, \tau, h) = 1 - a\lambda(1 - e^{-ih\xi}), \quad \xi \in (-\pi/h, \pi/h].$$

Of course, for system of equations, G can be a matrix.

Example 2. *Consider the wave equation*

$$\partial_{tt}u = c_0^2\partial_{xx}u.$$

This corresponds to the system of equations

$$\partial_t u + \frac{c_0^2}{\rho_0}\partial_x\rho = 0, \quad \partial_t\rho + \rho_0\partial_x u = 0.$$

If we do the FDM

$$\frac{u_j^{n+1} - u_j^n}{\tau} + \frac{c_0^2}{\rho_0}\frac{\rho_{j+1}^n - \rho_{j-1}^n}{2h} = 0,$$
$$\frac{\rho_j^{n+1} - \rho_j^n}{\tau} + \rho_0\frac{u_{j+1}^n - u_{j-1}^n}{2h} = 0.$$

Define $U_j^n = [u_j^n, \rho_j^n]^T$. We then have

$$U_j^{n+1} = U_j^n + \frac{\lambda}{2}\left\{\begin{bmatrix} 0 & -c_0^2/\rho_0 \\ -\rho_0 & 0 \end{bmatrix} U_{j+1}^n + \begin{bmatrix} 0 & c_0^2/\rho_0 \\ \rho_0 & 0 \end{bmatrix} U_{j-1}^n\right\}.$$

Taking the (semidiscrete) Fourier transform:

$$V_{n+1} = V_n + \frac{\lambda}{2}\left\{ \begin{bmatrix} 0 & -c_0^2/\rho_0 \\ -\rho_0 & 0 \end{bmatrix} e^{i\xi h} + \begin{bmatrix} 0 & c_0^2/\rho_0 \\ \rho_0 & 0 \end{bmatrix} e^{-i\xi h} \right\} V_n.$$

Hence,

$$G(\xi,\tau,h) = I - i\lambda \sin(\xi h) \begin{bmatrix} 0 & c_0^2/\rho_0 \\ \rho_0 & 0 \end{bmatrix}.$$

Hence, the growth factor now becomes a growth matrix (增长矩阵).

3.1 von Neumann condition

The following is direct observation:

Proposition 1. *For the linear evolutionary equation with constant coefficients, a scheme is stable if and only if there exists τ_0, K such that for wall $\tau \le \tau_0$, $n\tau \le T$, $\xi \in [-\pi/h, \pi/h)$ such that*

$$\|G(\xi,\tau,h)^n\| \le K.$$

Using a fact from linear algebra, which says that

$$\rho(A) \le \|A\|$$

for any matrix A and operator norm $\|\cdot\|$ where $\rho(A) = \sup_i |\lambda_i|$, the largest magnitude of the eigenvalues, is called the spectral radius. [For general operators, the spectrum may include other points than eigenvalues]. Then,

$$(\rho(G))^n = \rho(G^n) \le \|G^n\| \le K.$$

Hence,

$$\rho(G(\xi,\tau,h)) \le K^{1/n} = K^{\tau/T} \le 1 + M\tau, \forall \tau \le \tau_0.$$

for some constant M that depends on K and τ_0. Hence,

Theorem 1. *A necessary condition for the stability is that*

$$|\lambda_j| \le 1 + M\tau, \forall j, \tau \le \tau_0.$$

This condition is called the **von Neumann condition**.

Theorem 2. *If there exists $M > 0$ such that*

$$\|G(\xi, \tau, h)\| \leq 1 + M\tau$$

for any $\xi \in [-\pi/h, \pi/h)$ and all $n\tau \leq T$, then the scheme is stable.

证明. The proof is straightforward:

$$\|G^n\| \leq \|G\|^n \leq (1 + M\tau)^n \leq \exp(Mn\tau) \leq \exp(MT) =: K.$$

Then,

$$\|\hat{u}^n\|_2 \leq \sup_{\xi} \|G^n(\xi)\| \|\hat{u}_0\|_2 \leq K\|\hat{u}_0\|_2.$$

Using the Parseval equality, the conclusion follows. □

Hence, if there are some conditions such that

$$\rho(G) = \|G\|,$$

the the von Neumann condition is both sufficient and necessary. One possible condition is that G is a normal matrix (正规矩阵, $GG^* = G^*G$ where $G^* = \bar{G}^T$). In fact, if G is a scalar, this holds.

Corollary 1. *If $G(\xi, \tau, h) = g(\xi, \tau, h)$ is a scalar (i.e., u is a scalar), then a sufficient and necessary condition for the ℓ^2-stability is that there exists $M \geq 0$ such that the amplification factor g satisfies*

$$|g(\xi, \tau, h)| \leq 1 + M\tau.$$

3.2 More examples

Example 3. *For the method,*

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

we find

$$G(\xi) = 1 - a\lambda i \sin(\xi h).$$

Then,

$$|G| = \sqrt{1 + a^2\lambda^2 \sin^2(\xi h)}.$$

Clearly, if in the limit $\tau, h \to 0$, the ratio $\lambda = \tau/h$ is fixed, the method cannot be stable.

However, since

$$\sqrt{1 + a^2 \frac{\tau^2 \sin^2(\xi h)}{h^2}} \lesssim 1 + a^2 \frac{\tau^2}{2h^2}$$

Then, if τ/h^2 is fixed as $\tau, h \to 0$. The method is stable.

Example 4. *Consider again the scheme*

$$\frac{u_j^{n+1} - u_j^n}{\tau} = \frac{1}{h^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n).$$

We have

$$v_{n+1} = [1 + \frac{\tau}{h^2}(e^{ih\xi} - 2 + e^{-ih\xi})]v_n e^{ijh\xi}.$$

Hence, we find

$$G(\xi) = 1 + \frac{\tau}{h^2} 2(\cos(\xi h) - 1).$$

We find $2(\cos(\xi h) - 1) \in [-4, 0]$. Then, if $-4\tau/h^2 \geq -2$, $|G| \leq 1$, the method is stable. (Due to $1/h^2$, it's not possible to expect positive α.) We obtain the same requirement.

Example 5. *Consider the Richardson scheme (applying the leapfrog in time and centered difference in space) for the diffusion equation $u_t = au_{xx}$,*

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

This is a three-level scheme.

Introducing $v_j^{n+1} = u_j^n$ and defining $U_j^n = [u_j^n, v_j^n]$, we can get the growth matrix

$$G = \begin{bmatrix} -8a\mu \sin^2(\xi h/2) & 1 \\ 1 & 0 \end{bmatrix}$$

For this matrix, one eigenvalue is

$$\mu_1 = -4a\mu \sin^2(\xi h/2) - (1 + 16a^2\mu^2 \sin^4(\xi h/2))^{1/2}.$$

Then, one sees that

$$\rho \geq 1 + 4a\mu \sin^2(\xi h/2) = 1 + \tau \frac{4a \sin^2(\xi h/2)}{h^2}.$$

This violates the von Neumann condition and it cannot be stable. In fact, for $\tau \to 0, h \to 0$, *the condition is violated regardless of the relative speed of convergence to zero.*