Lecture 2, Fourier and von Neumann analysis

1 The discrete Fourier transform

A *plane wave* (also called *Fourier mode*, or sine wave, ...) is

$$e^{ikx} = \cos(kx) + i\sin(kx) \ .$$

Fourier analysis: represent a general function as a sum or integral of plane waves

$$u(x) = \sum_k \widehat{u}_k e^{ikx}$$
$$u(x) = \int \widehat{u}_k e^{ikx}\, dx$$

Uses in numerical PDE:

- Theory/analysis:
  - von Neumann stability analysis: calculate the *symbol* of a time stepping method to determine whether the method is stable. Replaces guesswork with (often painful) calculation. Is particularly useful for high order methods.
  - Qualitative behavior of numerical solution: Does it do what the PDE does (smoothing, wave propagation, etc.)?
- Computing (often using the *FFT*):
  - Fast solvers: solve linear equation systems that arise from special PDE discretizations.
  - Fourier/spectral methods: highly accurate methods ("spectral accuracy") by using the exact PDE symbol on numerical Fourier modes.

Specifics:

$$k = \ \textit{wave number}, \text{ units are } 1/\text{length}$$
$$\lambda = \frac{2\pi}{k} = \ \textit{wavelength}, \text{ units are length}$$
$$e^{ikx} = \textit{Fourier mode}$$
$$\text{periodic with } \textit{period } \lambda$$
$$e^{ik(x+\lambda)} = e^{ikx}$$

Fourier representation of periodic functions: Suppose $u(x+L) = u(x)$ for all x ("u is *periodic* with *period* L"). A Fourier representation of u should use components ("basis functions", "Fourier modes", etc.) that also have period L, at least:

$$e^{ik(x+L)} = e^{ikx} \ .$$

This implies (you do "the math") that the wave number k must satisfy

$$kL = 2\pi m \ , \quad \text{for some integer } m \ .$$

Possibilities:

- $m = 0$, $k = 0$, the constant
- $m = \pm 1$, $e^{\pm 2\pi i x/L}$, the *fundamental*
  - $\cos(2\pi x/L) = 0$ and $\sin(2\pi x/L) = 0$ only twice in $[0, L]$, the smallest possible number of "nodes".
- $|m| > 1$, $|\lambda| = \left|\frac{L}{m}\right| \leq \frac{L}{2}$, "higher harmonics". These have more oscillations (go up and down more, the real and imaginary parts), more nodes, etc.

$$k_m = \frac{2\pi m}{L} \text{ gives a plane wave } e^{ik_m x/L} = e^{2\pi i m x/L} = e_m(x) \ .$$

The functions $e_m(x)$ form a "basis" (in the sense of linear algebra) for the space of "all" periodic functions with period L. There are *expansion coefficients*, $\widehat{u}_m$, so that

$$\begin{aligned} u(x) &= \sum_m (\text{expansion coefficient}) \cdot (\text{basis element}) \\ &= \sum_{m=-\infty}^{\infty} \widehat{u}_m \, e_m(x) \\ &= \sum_{-\infty}^{\infty} \widehat{u}_m \, e^{2\pi i m x/L} \ . \end{aligned}$$

The sum goes over all integers (positive and negative) m because the basis elements $e_m(x)$ are linearly independent, as we will see. The non-trivial thing is that these linearly independent modes are *complete*: if u is periodic with period L and u is orthogonal to all e_m, then $u = 0$. The Fourier representation is possible for "any" periodic function.

The Fourier modes $e_m(x)$ are *orthogonal* in the L^2 inner product. (Note: the period L and the "L" in L^2 have nothing to do with each other. One is the for "length" and the other is for the mathematician *Lebesgue*, pronounced "lee beg", who defined L^2, L^1, etc.) The L^2 inner product is

$$\langle u, v \rangle = \int_0^L \overline{u}(x) v(x) \, dx \ .$$

Here $\overline{u}$ is the complex conjugate of u, in case u is complex. Fourier analysis with complex basis functions $e_m(x)$ makes us use complex numbers even when the ultimate target functions are real. We also could do "real" Fourier analysis with real basis functions $c_m(x) = \cos(2\pi mx/L)$, $m \geq 0$, and $s_m(x) = \sin(2\pi mx/L)$, $m \geq 1$, but all the formulas take longer to write. Functions u and v are orthogonal if $\langle u, v\rangle = 0$. If $m \neq n$, the basis function e_m is orthogonal to the basis function e_n:

$$\begin{aligned}
\langle e_m, e_n\rangle &= \int_0^L \overline{e}_m(x) e_n(x)\, dx \\
&= \int_0^L \exp\left(\frac{2\pi i(n-m)}{L}\, x\right)\, dx \\
&= \frac{L}{2\pi i(n-m)} \left[\exp\left(\frac{2\pi i(n-m)}{L}\, x\right)\Bigg|_0^L\right] \\
&= 0\ .
\end{aligned}$$

We need $n \neq m$ so that the denominator on the next to last line is not zero. We need $n - m$ to be an integer so that

$$\exp\left(\frac{2\pi i(n-m)}{L}\, L\right) - \exp\left(\frac{2\pi i(n-m)}{L}\, 0\right) = 1 - 1 = 0\ .$$

If $n = m$, then

$$\langle e_m, e_n\rangle = L\ .$$

Therefore, if u has a Fourier representation, the Fourier coefficients may be found using orthogonality:

$$\begin{aligned}
\langle e_n u\rangle &= \langle e_n, \sum_k \widehat{u}_k e_m\rangle \\
&= \sum_k \widehat{u}_k \langle e_n, e_m\rangle \\
&= L\, \widehat{u}_n\ .
\end{aligned}$$

In the sum on the next to last line, the inner product is equal to zero for all k except $k = n$. Therefore the sum is equal to the term corresponding to $k = n$. We rewrite this as a formula for the Fourier coefficient, in two equivalent ways:

$$\widehat{u}_n = \frac{1}{L}\langle e_n, u\rangle = \frac{1}{L}\int_0^L e^{-2\pi inx/L} u(x)\, dx\ .$$

The minus in the exponent in the integral comes from the complex conjugate: $\overline{e^{i\theta}} = e^{-i\theta}$.

For stability (von Neumann analysis) we can tell whether a finite difference scheme increases the size of U by a Fourier calculation. By "size", we mean L^2

norm. Underlying this is the fact that you can tell the L^2 norm of a function from its Fourier coefficients. This is not true about the L^1 norm, or any other norm that isn't based on L^2. The algebra is (see explanations below)

$$\begin{aligned}
\|u\|_{L^2}^2 &= \langle u, u \rangle \\
&= \langle \left(\sum_{n=-\infty}^{\infty} \widehat{u}_n e_n \right), \left(\sum_{m=-\infty}^{\infty} \widehat{u}_m e_m \right) \rangle \\
&= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \langle \widehat{u}_n e_n, \widehat{u}_m e_m \rangle \\
&= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \overline{\widehat{u}_n} \widehat{u}_m \langle e_n, e_m \rangle \\
&= \sum_{n=-\infty}^{\infty} \overline{\widehat{u}_n} \widehat{u}_n \, L \\
&= L \sum_{n=-\infty}^{\infty} |\widehat{u}_n|^2 \ .
\end{aligned}$$

Here are the mathematical tricks in this:

- For the second line, we use n as the first dummy summation variable and m as the second one. This allows you to take the summations outside and make a double sum over both n and m.
- The inner product is "anti-linear" in the first variable, which means that $\langle au, v \rangle = \overline{a} \langle u, v \rangle$. That's why we have $\overline{\widehat{u}_n}$. Note that $\widehat{u}_n$ is a number while e_n is a function: $e_n(x) = e^{2\pi 2mx/L}$.
- The main step is orthogonality, $\langle e_n, e_m \rangle = 0$ unless $n = m$. That's how the double sum over n and m becomes a single sum over n. In the sum over m, only the term $m = n$ is different from zero.
- If z is any complex number, then $\overline{z} z = |z|^2$. Apply this to the complex number $\widehat{u}_n$.

Summary:

- Fourier representation, Fourier inversion formula

The *completeness theorem* says that if the coefficients are defined as above, then the Fourier sum converges to u. Summary:

- Fourier (series) representation/Fourier inversion formula

$$u(x) = \sum_{m=-\infty}^{\infty} \widehat{u}_m e^{2\pi i m x/L} \ .$$

- Fourier (series) transform:

$$\widehat{u}_m = \frac{1}{L} \int_0^L e^{-2\pi i m x/L} u(x)\, dx \ .$$

  Note: the "inversion formula" undoes the Fourier transform by calculating the original function u from the Fourier amplitudes $\widehat{u}$.

- Plancharel theorem

$$\|u\|_{L^2}^2 = \int_0^L |u(x)|^2 \, dx = L \sum_{m=-\infty}^{\infty} |\widehat{u}_m|^2 \ .$$

Next, the *FFT*, which stands for *fast Fourier transform*, or *finite Fourier transform*. The finite Fourier transform is a linear operation on N component complex vectors

$$U \in \mathbb{C}^N \ \overset{\mathcal{F}}{\rightsquigarrow} \ \widehat{U} \in \mathbb{C}^N \ .$$

We will give the formula below. It is also called the *discrete* Fourier transform, or *DFT*, because it has all finite sums and no integrals. The direct calculation of $\widehat{U}$ from U takes $O(N^2)$ operations. The *fast* Fourier transform is an algorithm that computes all N components of $\widehat{U}$ from all N components of U in $O(N \log(N)$ operations. This makes FFT based algorithms practical for numerical computing.

The discrete "Fourier modes" are vectors $F_m \in \mathbb{C}^N$ with components

$$F_{mj} = e^{\frac{2\pi i m j}{N}} \ .$$

These resemble the Fourier modes we used before. The resemblance will get stronger soon. But first the algebra of the DFT. The new thing, the thing that makes the DFT different from the continuous Fourier transform or Fourier series, is *aliasing*. This is

$$E_{m+N} = E_m \ .$$

You verify this by using the component formulas above.

$$E_{m+N,j} = e^{\frac{2\pi i (m+N) j}{N}} = e^{\frac{2\pi i m j}{N} + 2\pi i j} = e^{\frac{2\pi i m j}{N}} e^{2\pi i j} = e^{\frac{2\pi i m j}{N}} = E_{mj} \ .$$

An *alias* is a different name for something. The labels m and $m+N$ are different labels for the same vector.

The labels $m = 0, 1, \ldots, N-1$ all correspond to distinct vectors. Any other label, for m outside the range $0, \ldots, N-1$, is "aliased" to one of these (it is the same vector with a different label). We show that these N vectors are orthogonal to each other. If $n \neq m$ and if both are in the range $0, \ldots, N-1$, then

$$E_n^* E_m = 0 \ .$$

For a column vector $U \in \mathbb{C}^N$, the notation U^* refers to the row vector whose elements are the complex conjugates:

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{pmatrix} \quad \Longleftrightarrow \quad U^* = (\overline{U}_1, \overline{U}_2, \ldots, \overline{U}_N) \; .$$

If U and V are two vectors, then

$$U^* V = \sum_{j=0}^{N-1} \overline{U}_j V_j \; .$$

This is the "algebraic" inner product for $\mathbb{C}^N$. We will soon define inner products with normalizing factors as we did in lecture 1.

For the actual inner product calculation we use geometric sum formulas that apply to any complex number z

$$\sum_{j=0}^{N-1} z^j = \begin{cases} \frac{z^N - 1}{z - 1} & \text{if } z \neq 1 \\ N & \text{if } z = 1 \end{cases}$$

Finally, the calculation. The inner product is

$$E_n^* E_m = E_n^* E_n = \sum_{j=0}^{N-1} e^{-2\pi i n j/N} e^{2\pi i m j/N} = \sum_{j=0}^{N-1} e^{2\pi i (m-n) j/N} \; .$$

If $n = m$, this is

$$E_n^* E_n = \sum_{j=0}^{N-1} \left| e^{2\pi 2 j n/N} \right| = \|E_n\|_2^2 = N \; .$$

If $n \neq m$, the general term in the sum is

$$e^{2\pi i (m-n) j/N} = z^j \quad , \quad z = e^{2\pi i (m-n)/N} \; .$$

Here's the new thing (not in continuous Fourier theory): if n and m are both in the range $0, \ldots, N-1$, then $|m - n| < N$. That means that if $n \neq m$, then $z \neq 1$. From the geometric series sum formula, we get

$$\sum_{j=0}^{N-1} e^{2\pi i (m-n) j/N} = \sum_{j=0}^{N-1} z_j = \frac{z^N - 1}{z - 1} \; .$$

But (this is the "punch line") the numerator is equal to zero because

$$z^N = e^{2\pi i (m-n) N/N} = 1 \; .$$

That shows that E_m and E_n are orthogonal if $n \neq m$ in the range.

Here are the linear algebra consequences of these calculations:

- For any $U \in \mathbb{C}^N$, there is a representation of U as a sum of discrete Fourier modes

$$U = \sum_{m=0}^{N-1} \widehat{U}_m E_m \qquad \text{(vector form)}$$

$$U_j = \sum_{m=0}^{N-1} \widehat{U}_m e^{2\pi i m j/N} \qquad \text{(component form)} .$$

  We know expansion coefficients $\widehat{U}_m$ exist because the vectors E_m form a basis of $\mathbb{C}^N$. This is the discrete analogue of the Fourier representation/Fourier inversion formula.

- The Fourier expansion coefficients are given by

$$\widehat{U}_m = \frac{1}{N} E_m^* U = \frac{1}{N} \sum_{j=0}^{N-1} e^{-2\pi i j m/N} U_j .$$

  These formulas are the discrete Fourier transform. The vector $\widehat{U} \in \mathbb{C}^N$ is the discrete Fourier transform of U. Warning: other versions of the DFT formulas put the $\frac{1}{N}$ factor in a different place. There has to be a $\frac{1}{N}$ somewhere.

- Plancharel formulas in three equivalent forms, starting with the form we use to derive it

$$U^* U = N \, \widehat{U}^* \widehat{U}$$

$$\|U\|_2 = N \left\| \widehat{U} \right\|_2$$

$$\sum_{j=0}^{N-1} |U_j|^2 = N \sum_{m=0}^{N-1} \left| \widehat{U}_m \right|^2 .$$

- Mathematical remark: The algebra for discrete and continuous Fourier analysis is similar. But there is a sense in which the discrete version is much easier. Since $\mathbb{C}^N$ is an N dimensional vector space, once we have N linearly independent vectors, we know they form a basis. That means that any vector $U \in \mathbb{C}^N$ has a representation as a linear combination of the basis vectors. In the DFT, we show the E_m are linearly independent by showing they are orthogonal. For continuous Fourier representations, this "completeness" step is not easy. For Fourier series, there are famous proofs of completeness that use the Dirichlet kernel and the Fejer kernel.

2 von Neumann analysis

Suppose we solve the heat/diffusion equation with *periodic* boundary conditions. This mean

$$u(x + L, t) = u(x, t) ,$$

Assuming u is periodic isn't really the same thing as giving a boundary condition as we did before. There's no place you would call the "boundary". Also, there aren't many physical situations where the solution is periodic. Still, it's useful to ask what would happen if we were to apply the finite difference marching scheme to a periodic function. It's useful because periodic "boundary conditions" are easier. If the scheme does something bad in this case, it is likely to do the same bad thing with more realistic boundary conditions. (We will see some theorems of this kind in this class.)

We are motivated to consider periodic boundary conditions because we can study stability using Fourier analysis in that case. Suppose there are N grid points "in space". Then the unknowns at time t_k are the N numbers that form the components of a vector $U_k \in \mathbb{R}^N$:

$$U_k = (U_{k,0}, \ldots, U_{k,N-1})^t \ .$$

The t in $(\cdots)^t$ is the *transpose* that turns the row vector into a column vector. Sometimes we work with the elements of U_k and sometimes we treat them as components of a vector. It is helpful to be able to use different levels of abstraction at the same time.

The FFT is useful for stability because the finite difference operator $U_{k+1} = AU_k$ operates as a *multiplier* on the Fourier components $\widehat{U}_{k,m}$:

$$\widehat{U}_{k+1,m} = a_m \widehat{U}_{k,m} \ . \tag{1}$$

The a_m is the *multiplier*, also called the *symbol*, of the finite difference time-step operator A. We will write a simple formula for it soon. The formula is used together with the Plancharel identity that relates the L^2 norm of U_k to the L^2 norm of $\widehat{U}_k$. The *maximum amplification factor* is

$$\alpha = \max_m |a_m| \ . \tag{2}$$

von Neumann stability analysis is the theorem that

$$\|U_{k+1}\|_2 \leq \alpha \, \|U_k\|_2 \ . \tag{3}$$

This inequality is *sharp*, which means that there is no constant smaller than α that makes the inequality true for all U. The inequality, in general, is only for the L^2 norm. It may hold, for other reasons, for other norms, in specific cases.

Here is the proof of the von Neumann analysis inequality (3). In the *Fourier*

domain we have

$$\begin{aligned}
\left\|\widehat{U}_{k+1}\right\|_2^2 &= \sum_{m=0}^{N-1} \left|\widehat{U}_{k+1,m}\right|^2 \\
&= \sum_{m=0}^{N-1} \left|a_m \widehat{U}_{k,m}\right|^2 \\
&= \sum_{m=0}^{N-1} |a_m|^2 \left|\widehat{U}_{k,m}\right|^2 \\
&\leq \max_m |a_m|^2 \sum_{m=0}^{N-1} \left|\widehat{U}_{k,m}\right|^2 \\
&= \alpha^2 \sum_{m=0}^{N-1} \left|\widehat{U}_{k,m}\right|^2 \\
&= \alpha^2 \left\|\widehat{U}_k\right\|^2
\end{aligned}$$

This shows that $\left\|\widehat{U}_{k+1}\right\| \leq \alpha \left\|\widehat{U}_k\right\|$. The Plancharel identity is, for some specific constant that depends on your conventions of norm and Fourier transform,

$$\|U\|_2 = C_P \left\|\widehat{U}\right\|_2 .$$

This identity is true only for L^2 norms, which is why von Neumann analysis is only for L^2. We get the inequality (3) by applying the Plancharel identity once at level $k+1$ and again at level k:

$$\|U_{k+1}\|_2 = C_P \left\|\widehat{U}_{k+1}\right\|_2 \leq C_P \alpha \left\|\widehat{U}_k\right\|_2 = \alpha \|U_k\|_2 .$$

To see that the inequality is sharp, let m_* be the mode that achieves the maximum in (2). Then $\alpha = |a_{m_*}|$. Suppose U_k is the corresponding Fourier mode

$$U_{k,j} = e^{2\pi i m_* j/n} .$$

Then $U_{k+1} = a_{m_*} U_k$, and $\|U_{k+1}\| = \alpha \|U_k\|$.

The explicit three point finite difference scheme we used last week is

$$U_{k+1,j} = b_{-1} U_{k,j-1} + b_0 U_{kj} + b_1 U_{k,j+1} .$$

The coefficients were

$$b_{-1} = b_1 = \frac{D\Delta t}{\Delta x^2} \quad , \quad b_0 = 1 - \frac{2D\Delta t}{\Delta x^2} .$$

The basis of von Neumann analysis is the fact that we can plug Fourier modes in here and get a simple result:

$$U_{kj} = e^{2\pi i m j/n} \implies U_{k+1,j} = a_m e^{2\pi i m j/n} .$$

For example,

$$U_{kj} = e^{2\pi imj/n} \implies U_{k,j+1} = e^{2\pi im(j+1)/n} = e^{2\pi im/n} e^{2\pi imj/n} \ .$$

The first factor on the right, $e^{2\pi im/n}$, doesn't depend on j. That makes $e^{2\pi im/n}$ a *Fourier multiplier*. The rest of the calculation is

$$b_{-1}U_{k,j-1} + b_0 U_{kj} + b_1 U_{k,j+1} = \left(b_{-1} e^{-2\pi im/n} + b_0 + b_1 e^{2\pi im/n} \right) e^{2\pi imj/n} \ .$$

This gives the symbol (Fourier multiplier) as

$$\begin{aligned} a_m &= 2\cos(2\pi m/n) b_1 + b_0 \\ &= 2\cos\left(\frac{2\pi m}{n}\right) \frac{\Delta t D}{\Delta x^2} + 1 - 2\frac{\Delta t D}{\Delta x^2} \\ a_m &= 1 + 2\frac{\Delta t D}{\Delta x^2} \left[\cos\left(\frac{2\pi m}{n}\right) - 1 \right] \ . \end{aligned}$$

This symbol formula is all you need, so you could stop here. But it can be done more simply. First, the big constant is the dimensionless CFL number

$$\lambda = \frac{\Delta t D}{\Delta x^2} \ .$$

This makes a shorter formula

$$a_m = 1 - 2\lambda \left[1 - \cos\left(\frac{2\pi m}{n}\right) \right] \ .$$

Second, the arguments of cos are numbers

$$\theta_m = \frac{2\pi m}{n}$$

which are uniformly spaced in the interval $0 = \theta_0$ to $2\pi - \frac{2\pi}{n} = \theta_{n-1}$. As $n \to \infty$, these numbers become "dense" in the range $0 \le \theta < 2\pi$. The symbol is

$$a_m = 1 - 2\lambda \left[\cos(\theta_m) - 1 \right] \ .$$

We're interested in the maximum modulus (maximum or minimum, whichever is "larger") of the numbers a_m when n is large. Since $\cos(\theta)$ is a continuous function of θ, as $n \to \infty$, this converges to

$$\max_{0 \le \theta < 2\pi} |a(\theta)| \ .$$

We can solve this simple calculus problem without asking precisely where the numbers θ_m fall. If the maximizer is

$$\theta_* = \arg \max_{0 \le \theta < 2\pi} |a(\theta| \ ,$$

then as $n \to \infty$, the θ_m closest to θ_* converges to θ_*. Therefore, the minimum over the finite (but large) set $\{\theta_m\}$ converges to the continuous minimum. In the present example, we need to solve the minimization problem

$$\alpha = \min_{0 \le \theta < 2\pi} \left| 1 - 2\lambda \left[1 - \cos(\theta) \right] \right| \ .$$

This is easy. The cosine has range between 1 and -1. The maximum is attained at one of these. When $\theta = 0$ and $\cos(\theta) = 1$, we get $a(0) = 1$. This is independent of λ. When $\theta = \pi$ and $cos(\pi) = -1$, the symbol is

$$a(\pi) = 1 - 2\lambda(\cos(\pi) - 1) = 1 - 4\lambda \ .$$

Clearly (think about this) $|a(\pi)| \le 1$ if $a(\pi) \ge -1$, which is

$$\lambda \le \frac{1}{2} \ .$$

That is the CFL restriction we had last week. One difference here is that it isn't guesswork, but a calculation. Also, now we know that if the condition is violated the time stepping scheme definitely is unstable.

The symbol and Fourier modes are eigenvalues and eigenvectors of the one step matrix A. Let $V_m \in \mathbb{C}^n$ be the vector whose components are the complex exponentials

$$V_{m,j} = e^{2\pi i m j/n} \ .$$

The calculation above is just

$$AV_m = a_m V_m \ .$$

We calculated above that the Fourier mode vectors V_m are orthogonal. If $m' \neq m$ (in the range $0 \le m < n$ and $0 \le m' < n$), then

$$\langle V_m, V_{m'} \rangle = 0 \ .$$

Therefore, the norm of A, in L^2, is equal to the largest eigenvalue (in modulus)[1].

There are other ways, and possibly more physical, to think about the discrete Fourier modes V_m. These differ in the way the "Fourier" variable (m or θ) is *scaled*. However you scale the Fourier variable, you have to scale the space variable (j or y or x, see below) accordingly. Suppose you use θ, with discrete values $\theta_m = \frac{2\pi}{n}$. We call the corresponding space variable y (the name is irrelevant). Its discrete values are $y_j = j$. The distance between neighboring θ values is $\Delta\theta = \frac{2\pi}{n}$. The distance between neighboring y values is $\Delta y = 1$. The product is $\Delta\theta \cdot \Delta y = \frac{2\pi}{n}$. This product will be the same for all of the scalings we consider. The values of the discrete Fourier modes are

$$V_{m,j} = e^{i\theta_m y_j} \ .$$

[1]Warning: This is for periodic boundary conditions. Eigenvalue analysis can be wrong with other boundary conditions. There are unstable schemes with $|\lambda_m| \le 1$ for all m. We will see an example, for a different PDE, with Dirichlet boundary conditions.

This may be written

$$(V_{m,0}, V_{m,1}, V_{m,2}, \ldots) = \left(1, e^{i\theta}, e^{2i\theta}, \ldots\right) \ .$$

As we saw last week, the time step formula (forward Euler in time, centered difference in space) may be written in terms of left and right shifts. The left shift moves components one spot to the left. If $W = S_L U$, then $W_j = U_{j+1}$. If U is one of the Fourier modes, then the components of $W = S_L U$ are

$$W_j = U_{j+1} = e^{i(j+1)\theta} = e^{i\theta} e^{ij\theta} \ .$$

This may be written as $S_L V_m = e^{i\theta} V_m$. Similarly, the right shift satisfies $S_R V_m = e^{-i\theta} V_m$. The scheme is

$$U_{k+1} = AU_k = \left(b_1 S_L + b_0 I b_{-1} S_R\right) U_k \ .$$

Acting on a Fourier mode, this becomes

$$AV_m = b_1 e^{i\theta} + b_0 + b_{-1} e^{-i\theta} \ .$$

Here is the advantage of this scaling. It makes the symbol calculation very simple. Just put in $e^{ir\theta}$ for a shift by r units to the left in the difference scheme.

I present one more scaling. This one has the advantage that the space variable, called x here, is given in the physical units of the problem. If the problem is on a "periodic interval" of length L (a bit of an oxymoron, $u(x + L, t) = u(x, t)$, then the space variable runs from 0 to (almost) L. The Fourier variable will be called[2] p. If $x_j = j\Delta x$ with $\Delta x = L/n$, then the relation

$$e^{ip_m x_j} = e^{2\pi i m j/n} \ ,$$

gives

$$p_m x_j = \frac{2\pi m j}{n} \ , \quad x_j = \frac{jL}{n} \quad \Longrightarrow \quad p_m = \frac{2\pi m}{L} \ .$$

The spacings are $\Delta x = x_{j+1} - x_j = \frac{L}{n}$, and $\Delta p = p_{m+1} - p_m = \frac{2\pi}{L}$. As we said, the product is $\Delta x \Delta p = \frac{2\pi}{n}$.

This scaling is good for seeing the relation between the discrete Fourier transform and Fourier series. This representation is easier to interpret if we use a different set of labels for the Fourier modes. Instead of $0 \le m < n$, we use a set of modes that is as symmetric as possible around $m = 0$. If n is odd, then we use $m = \frac{-n+1}{2}, \ldots, 0, \ldots, \frac{n-1}{2}$. For example, for $n = 5$, the mode labels would be $(-2, -1, 0, 1, 2)$. This is exactly symmetric. For n even, we have the small asymmetry of having one more positive mode than negative. For example, for $n = 6$, we take $m = (-2, -3, 0, 1, 2, 3)$. In either case, we will write $|m| \le \frac{n}{2}$,

[2] Physicists use p for momentum. In quantum mechanics, a "state" with momentum p has "wave function $e^{ipx/\hbar}$, where $\hbar$ (pronounced "h bar") is a physical constant that may be set equal to 1.

though it isn't exactly true in either case. If U_j is a set of grid values, the representation is

$$U_j = \sum_{|m|<\frac{n}{2}} \widehat{U}_m e^{ip_m x_j} \ .$$

If $u(x)$ is periodic with period L and $U_j = u(x_j)$, then the finite Fourier sum is equal to u at the grid points. As the grid points become more densely spaced (as $n \to \infty$ and $\Delta x \to 0$), we have the limiting formula

$$u(x) = \sum_{m=-\infty}^{\infty} \widehat{u}_m e^{ip_m x} = \sum_{m=-\infty}^{\infty} \widehat{u}_m e^{2\pi i m x/L} \ .$$

Now recall that $\Delta x = \frac{L}{n}$, so

$$\begin{aligned}
\widehat{u}_m &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{-ip_m x_j} u(x_j) \\
&= \frac{1}{L} \lim_{n\to\infty} \Delta x \sum_{j=0}^{n-1} e^{-ip_m x_j} u(x_j) \\
&= \frac{1}{L} \int_{x=0}^{L} e^{-ip_m x} u(x) dx \\
&= \frac{1}{L} \int_{x=0}^{L} e^{-2\pi i m x/L} u(x) dx \ .
\end{aligned}$$

These are the usual Fourier series formulas you can find in a book.

In designing finite difference methods we often want to compare a finite difference operator to the corresponding differential operator. Symbol calculus (calculating with symbols of operators) can be helpful for that. A simple example is the first derivative operator ∂_x and the second order centered finite difference approximation

$$\partial_x f(x) \approx D_0(\Delta x) f(x) = \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} \ .$$

We set $f(x) = e^{ipx}$ and calculate the *symbol* of the differential operator

$$\partial_x e^{ipx} = ipe^{ipx} \ .$$

This says that the symbol of the *differentiation operator* is $a(p) = ip$. Similarly,

we can calculate the action of the finite difference operator on the plane wave:

$$\begin{aligned} D_0(\Delta x)e^{ipx} &= \frac{e^{ip(x+\Delta x)} - e^{ip(x-\Delta x)}}{2\Delta x} \\ &= \frac{e^{ipx}e^{ip\Delta x} - e^{ipx}e^{-ip\Delta x}}{2\Delta x} \\ &= \frac{e^{ip\Delta x} - e^{-ip\Delta x}}{2\Delta x}\, e^{ipx} \\ &= \frac{i}{\Delta x}\frac{e^{ip\Delta x} - e^{-ip\Delta x}}{2i}\, e^{ipx} \\ &= \frac{i\sin(p\Delta x)}{\Delta x}\, e^{ipx} \end{aligned}$$

This says that the symbol of $D_0(\Delta x)$ is $a(p) = \frac{i\sin(p\Delta x)}{\Delta x}$. What is the relation between these two symbols? For each fixed p, we have

$$a(p, \Delta x) \to ip\ , \quad \text{as} \ \ \Delta x \to 0\ .$$

In fact, it's second order accurate

$$\begin{aligned} a(p, \Delta x) &= i\,\frac{\sin(p\Delta x)}{\Delta x} + O(\Delta x^2) \\ &= i\,\frac{p\Delta x - \frac{1}{6}p^2\Delta x^3 + p^5 O(\Delta x^5)}{\Delta x} \\ &= ip - \frac{ip^3}{6}\Delta x^2 + O(\Delta x^4)\ . \end{aligned}$$

Here is the interpretation of this formula. The operator D_0 is supposed to estimate the derivative of a function. So how well does it do on a plane wave. For any fixed plane wave, which means a fixed p, the result is second order accurate as $\Delta x \to 0$. That't the second order accuracy of the three point centered difference formula. Recall that the wavelength of a plane wave is

$$\lambda = \frac{2\pi}{p}\ .$$

The argument of the sine function is

$$p\Delta x = \frac{1}{2\pi}\,\frac{\Delta x}{\lambda} = \frac{1}{2\pi}\,\frac{1}{N_w}\ .$$

Here, N_w is the number of points per wavelength, how many grid points there are in one full cycle of the plane wave. This is a measure of how well resolved the plane wave is. Engineers calculating waves talk about resolution in this way: "My method gives 5% accuracy using only ten points per wavelength!"

Summary:

- Stability in L^2 is determined by the symbol, a_m or $a(\theta)$.
- The scheme is stable if $|a_m| \leq 1$ for all m.
- The numbers a_m are the eigenvalues of the one step matrix A, which is defined by $U_{k+1} = AU_k$.
- The corresponding eigenvectors are the discrete Fourier modes V_m with components $V_{m,j} = e^{2\pi i m j/n}$. The eigenvector/eigenvalue relation is $AV_m = a_m V_m$.
- The eigenvectors V_m are orthogonal, which implies that the norm of A is given by the "largest" eigenvalue (maximum modulus): $\|A\|_2 = \max_m |a_m|$.
- With a different convention, this is $|a(\theta)| \leq 1$ for all θ.
- The relation between mode m and θ is $\theta_m = 2\pi m/n$.
- As $n \to \infty$, the discrete values θ_m become "dense" on the circle $0 \leq \theta < 2\pi$, so
$$\max_{0\leq m<n} |a_m| \;=\; \max_{0\leq m<n} |a(\theta_m)| \;\to\; \max_{0\leq\theta\leq 2\pi} |a(\theta)| \;.$$
- We may take $0 \leq \theta \leq 2\pi$ or $-\pi \leq \theta \leq \pi$ because $a(\theta)$ is periodic in θ with period 2π.