MATH578 - Numerical Analysis 1

Based on lectures from Fall 2025 by Prof. J.C. Nave.
Notes by Louis Meunier

Contents

§1 Polynomial Interpolation

In general, the goal of interpolation is, given a function $f(x)$ on $[a,b]$ and a series of distinct ordered points (often called *nodes* or *collocation points*) $\left\{x_j\right\}_{j=1}^{n} \subseteq [a,b]$, to find a polynomial $P(x)$ such that $f\left(x_j\right) = P\left(x_j\right)$ for each j.

↪**Theorem 1.1** (Existence and Uniqueness of Lagrange Polynomial): Let $f \in C[a,b]$ and $\left\{x_j\right\}$ a set of n distinct points. Then, there exists a unique $P(x) \in \mathbb{P}_{n-1}$, the space of $n-1$-degree polynomials, such that $P\left(x_j\right) = f\left(x_j\right)$ for each j.

We call such a P the *Lagrange polynomial* associated to the points $\left\{x_j\right\}$ for f.

Proof. We define the following $n-1$ degree "fundamental polynomials" associated to $\left\{x_j\right\}$,

$$\ell_j(x) \equiv \prod_{\substack{1 \leq i \leq n \\ i \neq j}} \frac{x - x_i}{x_j - x_i}, \qquad j = 1, \ldots, n.$$

Then, one readily verifes $\ell_j(x_i) = \delta_{ij}$, and that the distinctness of the nodes guarantees the denominator in each term of the product is nonzero. Define

$$P(x) = \sum_{j=1}^{n} f\left(x_j\right)\ell_j(x),$$

which, being a linear combination of $n-1$ degree polynomials is also in $\mathbb{P}_{n-1}$. Moreover,

$$P(x_i) = \sum f\left(x_j\right)\delta_{i,j} = f(x_i),$$

as desired.

For uniqueness, suppose $\overline{P}$ another $n-1$ degree polynomial satisfying the conditions of the theorem. Then, $q(x) \equiv P(x) - \overline{P}(x)$ is also a degree $n-1$ polynomial with $q(x_i) = 0$ for each $i = 1, \ldots n$. Hence, q a polynomial with more distinct roots than its degree, and thus it must be identically zero, hence $P = \overline{P}$, proving uniqueness. ■

↪**Theorem 1.2** (Interpolation Error): Suppose $f \in C^n[a,b]$, and let $P(x)$ be the Lagrange polynomial for a set of n points $\left\{x_j\right\}$, with $x_1 = a, x_n = b$. Then, for each $x \in [a,b]$, there is a $\xi \in [a,b]$ such that

$$f(x) - P(x) = \frac{f^{(n)}(\xi)}{n!}(x - x_1)\cdots(x - x_n).$$

Moreover, if we put $h := \max_i (x_{i+1} - x_i)$, then

$$\|f - P\|_\infty \leq \frac{h^n}{4n}\|f^{(n)}\|_\infty.$$

PROOF. We prove the first identity, and leave the second "Moreover" as a homework problem. Notice that it holds trivially for $x = x_j$ for any j, so assume $x \neq x_j$ for any j, and define the function

$$g(t) := f(t) - P(t) - \omega(t)\frac{f(x) - P(x)}{\omega(x)}, \qquad \omega(t) := (t - x_1)...(t - x_n) \in \mathbb{P}_n[t].$$

Then, we observe the following:

- $g \in C^n[a,b]$
- $g(x) = 0$
- $g\left(t = x_j\right) = 0$ for each j

Recall that by Rolle's Theorem, if a C^1 function has $\geq m$ roots, then its derivative has $\geq$ $m - 1$ roots. Thus, applying this principle inductively to $g(t)$, we conclude that $g^{(n)}(t)$ has at least one root. Take ξ to be such a root. Then, one readily verifies that $P^{(n)} \equiv 0$ and $\omega^{(n)} \equiv n!$ (using polynomial properties), from which we may use the fact that $g^{(n)}(\xi) = 0$ to simplify to the required identity. ■

Remark 1.1: In general, larger n leads to smaller maximum step size h. However, it is *not* true that $n \to \infty$ implies $P \to f$ in L^∞. From the previous theorem, one would need to guarantee $\left\|f^{(n)}\right\| \to 0$ (or at least, doesn't grow faster than $\frac{h^n}{4n}$), which certainly won't hold in general; we have no control on the nth-derivative of an arbitrarily given function. However, we can try to optimize our choice of points $\left\{x_j\right\}$ for a given j.

We switch notation for convention's sake to $n + 1$ points x_j. Our goal is the optimization problem

$$\min_{x_j} \max_{x \in [a,b]} \left| \prod_j \left(x - x_j\right) \right|,$$

the only term in the error bound above that we have control over. Remark that we can expand the product term:

$$\prod_j \left(x - x_j\right) = x^{n+1} - r(x),$$

where $r(x) \in \mathbb{P}_n$. So, really, we equivalently want to solve the problem

$$\min_{r \in \mathbb{P}_n} \left\|x^{n+1} - r(x)\right\|_\infty,$$

namely, what n-degree polynomial minimizes the max difference between x^{n+1}?

↪**Theorem 1.3** (De la Vallé-Poussin Oscillation Theorem): Let $f \in C([a,b])$, and suppose $r \in \mathbb{P}_n$ for which there exists $n+2$ distinct points $\left\{x_j\right\}$ such that $a \leq x_0 < \cdots < x_{n+1} \leq b$ at which the error $f(x) - r(x)$ "oscillate" sign, i.e.

$$\operatorname{sign}\left(f\left(x_j\right) - r\left(x_j\right)\right) = -\operatorname{sign}\left(f\left(x_{j+1}\right) - r\left(x_{j+1}\right)\right).$$

Then,

$$\min_{P \in \mathbb{P}_n} \|f - P\|_\infty \geq \min_{0 \leq j \leq n+1} |f\left(x_j\right) - r\left(x_j\right)|.$$

↪**Definition 1.1** (Chebyshev Polynomial): The *degree n Chebyshev polynomial*, defined on $[-1,1]$, is defined by

$$T_n(x) := \cos\left(n \cos^{-1}(x)\right).$$

Remark 1.2: The fact that T_n actually is a polynomial follows from the double angle formula for cos, which says

$$\cos((n+1)\theta) = 2\cos(\theta)\cos(n\theta) - \cos((n-1)\theta).$$

In the context of T_n, this implies that for any $n \geq 1$, the recursive formula

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

This formula with a simple induction argument proves that each T_n a polynomial, with for instance $T_0(x) = 1, T_1(x) = x, T_2(x) = 2x^2 - 1$ and so on.

↪**Proposition 1.1**: $\{T_n\}$ are orthogonal with respect to the inner product given by

$$(f,g) := \int_{-1}^{1} f(x)g(x)\omega_2(x)\,\mathrm{d}x,$$

where $\omega_2(x) := \left(1 - x^2\right)^{1/2}$.

Remark 1.3: Defining similar *weight* functions by $\omega_n(x) := (1 - x^n)^{1/n}$, one can derive a more general class of polynomials called *Geigenbauer polynomials*, which are respectively orthogonal with respect to $\int \cdot \cdot \omega_n$.

↪**Proposition 1.2** (Some Properties of T_n):

- $|T_n(x)| \leq 1$ on $[-1, 1]$
- The roots of $T_n(x)$ are the n points
$$\xi_j := \cos\left(\frac{(2j-1)\pi}{2n}\right), \qquad j = 1, \ldots, n.$$
- For $n \geq 1$, $|T_n(x)|$ is maximal on $[-1, 1]$ at the $n + 1$ points
$$\eta_j := \cos\left(\frac{j\pi}{n}\right), \qquad j = 0, \ldots, n,$$
with $T_n\left(\eta_j\right) = (-1)^j$.

Note too that $T_{n+1}(x)$ has leading coefficient 2^n, which can be seen by the recursive formula above; define the *normalized* Chebyshev polynomials by $\hat{T}_{n+1}(x) := 2^{-n}T_{n+1}(x)$. Thus, we may write

$$\hat{T}_{n+1}(x) = x^{n+1} - r_n(x),$$

with $r_n(x) \in \mathbb{P}_n$. It follows for one that

$$\max_{x\in[-1,1]} |x^{n+1} - r_n(x)| = 2^{-n}.$$

Moreover, we know that at the $n + 2$ points η_j, we have

$$\hat{T}_{n+1}\left(\eta_j\right) = 2^{-j}(-1)^j = \eta_j^{n+1} - r_n\left(\eta_j\right).$$

Namely, because of the inclusion of $(-1)^j$ term, this means that $\hat{T}_{n+1}(x)$ oscillates sign between the η_j points, which fulfils the condition stated in the Oscillation Theorem. Thus, these observations readily imply the following result, settling our original question on optimizing locations of interpolation points for Lagrange interpolation:

↪**Theorem 1.4** (Optimal Approximation of x^{n+1} in $\mathbb{P}_n$): The optimal approximation of x^{n+1} in $\mathbb{P}_n$ on $[-1, 1]$ with respect to the L^∞ norm is given by

$$r_n(x) := x^{n+1} - 2^{-n}T_{n+1}(x).$$

Thus, the optimal Lagrange interpolation points are the $n + 1$ roots of $x^{n+1} - r_n(x)$, namely $\xi_j = \cos\left(\frac{(2j+1)\pi}{2n+2}\right)$ for $j = 0, \ldots, n$.

Remark 1.4: This, and previous results, were stated over $[-1, 1]$. A linear change of coordinates transforming any closed interval to $[-1, 1]$ readily leads to analgous results.

§2 Fourier Transform

Recall that the Fourier transform of a (Lebesgue) measurable function $u(x)$ on $\mathbb{R}$ is defined

$$(\mathcal{F}u)(\xi) = \hat{u}(\xi) = \int_{\mathbb{R}} e^{-i\xi x}u(x)\,\mathrm{d}x.$$

↪**Theorem 2.1**: Let $u \in L^2(\mathbb{R})$. Then,

1. $\hat{u} \in L^2$
2. the *inversion* formula holds, ie $u(x) = \int_{\mathbb{R}} \hat{u}(\xi) e^{i\xi x} \,\mathrm{d}x = \left(\mathcal{F}^{\{-1\}} u\right)(x)$
3. $\|\hat{u}\|_2 = \sqrt{2\pi} \|u\|_2$
4. for $u \in L^2, v \in L^1$, $u * v \in L^2$, and $\widehat{u * v} = \hat{u}\hat{v}$.

↪**Theorem 2.2** (Further Properties of Fourier Transform): Let $u, v \in L^2$. Then,

1. $\mathcal{F}$ is linear over $\mathbb{R}$
2. $\mathcal{F}(u(\cdot + x_0))(\xi) = e^{i\xi x_0} \hat{u}(x_0)$
3. $\mathcal{F}\left(e^{i\xi_0 x} u(x)\right)(\xi) = \hat{u}(\xi - x_0)$
4. If $c \neq 0$, $\mathcal{F}(u(c\,\cdot))(\xi) = \frac{\hat{u}\left(\frac{\xi}{c}\right)}{c}$
5. $\mathcal{F}(\overline{u})(\xi) = \overline{\hat{u}(-\xi)}$
6. if u_x exists and is in L^2, then
$$\mathcal{F}(u_x)(\xi) = i\xi \hat{u}(\xi).$$
By extension, if $\partial_\alpha u \in L^2$, then $\widehat{\partial_\alpha u}(\xi) = (i\xi)^\alpha \hat{u}(\xi)$
7. $\left(\mathcal{F}^{-1} u\right)(\xi) = \frac{1}{2\pi} \hat{u}(-\xi)$.

In a sense, 6. implies a duality between the smoothness of $u(x)$ and rapid decay (as $|\xi| \to \infty$) of $\hat{u}(x)$; 7. indicates that the same analogy holds switching the roles of u and $\hat{u}$. We make this more precise.

↪**Definition 2.1** (Bounded Variation): We say a function u on $\mathbb{R}$ is of *bounded variation* or write $\in$ BV if there exists a constant M such that for any finite integer m and collection of points $x_0 < x_1 < \ldots < x_m$,

$$\sum_{j=1}^{m} |u\left(x_j\right) - u\left(x_{j-1}\right)| \leq M.$$

In a sense, this notion of BV captures a notion of "limited oscillation".

↪**Theorem 2.3**: Let $u \in L^2$. Then:

1. If u has $p - 1$ continuous derivatives in L^2 and its pth derivative is in BV, then
$$\hat{u}(\xi) = O\left(|\xi|^{-p-1}\right).$$
2. If u has infinitely many derivatives all in L^2, then
$$\hat{u}(\xi) = O\left(|\xi|^{-M}\right), \qquad \forall M \geq 1.$$

§2.1 Discrete Fourier Transform

Let $h > 0$ be a *step size*. Let $x_j = jh$ for $j \in \mathbb{Z}$. We write $v = \left\{v_j\right\}_{j \in \mathbb{Z}}$ for discrete approximations of a function u on the grid $\left\{x_j\right\}_{j \in \mathbb{Z}}$, i.e. $v_j \approx u\left(x_j\right)$.

The ℓ_h^2 norm is defined for such v by

$$\|v\|_2 := \left[h \sum_{j \in \mathbb{Z}} |v_j|^2 \right]^{1/2}.$$

Then, ℓ_h^2 is defined as the space of such sequences v such that this norm is finite. analogous definitions hold for other ℓ_h^p spaces and norms.

↪**Proposition 2.1** (Nesting): $\ell_h^p \subset \ell_h^q$ for each $q \geq p$.

Remark 2.1: Note that the analogous result to this does *not* hold for L^p spaces (unless restricted to a compact domain).

We define the convolution of two sequences v, w by the new sequence $v * w$ with entries

$$(v * w)_m = h \sum_{j \in \mathbb{Z}} v_j w_{m-j} = h \sum_{j \in \mathbb{Z}} v_{m-j} w_j.$$

For any $v \in \ell_h^2$, we define too the *semi-discrete Fourier transform* of v by

$$\hat{v}(\xi) = (\mathcal{F}_h v)(\xi) = h \sum_{j \in \mathbb{Z}} e^{-i\xi x_j} v_j, \qquad \xi \in \left[-\frac{\pi}{h}, \frac{\pi}{h} \right],$$

where we remark that $\hat{v}(\xi)$ $\frac{2\pi}{h}$-periodic (hence the domain restriction) and continuous.

We define the norm of $\hat{v}$ by the usual L^2-norm, restricted to the appropriate domain:

$$\|\hat{v}\|_2 := \left(\int_{-\pi/h}^{\pi/h} |\hat{v}(\xi)|^2 \, \mathrm{d}\xi \right)^{1/2},$$

and L_h^2 the space of such functions with finite norm.

↪**Theorem 2.4**: If $v \in \ell_h^2$, then $\hat{v} \in L_h^2$, and we can recover v from $\hat{v}$ by the "inverse semi-discrete Fourier transform", i.e.

$$v_j = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{i\xi x_j} \hat{v}(\xi) \, \mathrm{d}\xi.$$

Also, Parseval's identity holds, i.e. $\|\hat{v}\|_2 = \sqrt{2\pi} \|v\|_2$, as does the expected convolution identity (for $v \in \ell_h^2, w \in \ell_h^1$ for instance).

Remark 2.2: Note that each wave number ξ is indistinguishable from $\xi + \frac{2\pi j}{h}$ for $j \in \mathbb{Z}$ on $h\mathbb{Z}$; this is called *aliasing*. The cutoff $\frac{\pi}{h}$ is called the *Nyquist Wave Number*.

↪**Theorem 2.5**: Let $u \in L^2$, sufficiently smooth, with $v \in \ell_h^2$ be a restriction of u to $h\mathbb{Z}$. Then,

$$\hat{v}(\xi) = \sum_{j \in \mathbb{Z}} \hat{u}\left(\xi + \frac{2\pi j}{h} \right), \qquad \xi \in \left[-\frac{\pi}{h}, \frac{\pi}{h} \right].$$

§2.2 Wavelet Transform

The heuristic idea of the wavelet transform is to construct a basis of functions which effectively compromise between localization in space and frequency; indeed, the issues related to aliasing in the discrete case are linked to the fact that localization of a function simultaneously in physical and fourier space is impossible except for the zero function.

More precisely, we dictate that a wavelet ψ should have:

1. non-negligible values in a limited range of space *and* frequency;
2. finite energy, by which we mean

$$\int_0^\infty |\hat{\psi}(\omega)|^2 \frac{\mathrm{d}\omega}{|\omega|} < \infty.$$

3. zero mean, i.e. $\int_{-\infty}^{\infty} \psi(t)\,\mathrm{d}t = 0$.

Note that 2., 3., imply that ψ has actual "frequency content" and zero mean, so ψ satisfying these properties must oscillate.

We call such a ψ a the *model wavelet*, from which we will generate our desired basis by translating and scaling:

$$\psi_{s,\tau}(t) := \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right).$$

From this, we define

$$\gamma(s,\tau) = \int f(t)\psi^*_{s,\tau}(t)\,\mathrm{d}t.$$

Then, one can retrieve f (with appropriate properties) by the *inverse wavelet transform*

$$f(t) = \int_{\mathbb{R}^+}\int_{\mathbb{R}} \gamma(s,\tau)\psi_{s,\tau}(t)\,\mathrm{d}\tau\,\mathrm{d}s.$$

In a sense, $\gamma(s,t)$ provides a *compromise* between space (i.e. τ) and frequency/scale (i.e. s) and energy localization.

More precisely, we'd like a quantitative decay of $\gamma(s,t)$ for small s, i.e. small frequencies, which are the problematic range. If we Taylor expand f in the definition of $\gamma(s,\tau)$ about $s=0$ (and taking $\tau=0$ for convenience), one notices that

$$\gamma(s,0) = \frac{1}{\sqrt{s}}\left[\sum_{p=0}^{n} f^{(p)}(0)\int \frac{t^p}{p!}\psi(t/s)\,\mathrm{d}t + \mathcal{O}(n+1)\right].$$

If we define $M_p := \int t^p\psi(t)\,\mathrm{d}t$ to be the pth moment of ψ, then one clearly sees that if the first n moments of ψ are identically 0, then

$$\psi(s,\tau) = \mathcal{O}(s^{n+2}),$$

those providing a qualitative decay rate for these coefficients. Thus, we generally want such vanishing moments in designing "good" wavelets.

§3 Finite Difference (FD) Approximation

Given $u \in C^\ell$, our goal is to approximate derivatives of u by a combination of finitely many function values, i.e.

$$\frac{\partial^k u}{\partial x^k}|_{x_0} = \sum_{i=0}^{m} \alpha_i u(x_i), \qquad k \leq \ell.$$

The vector $\alpha = (\alpha_i)$ is called the *finite difference stencil*. Such schemes are found by Taylor expanding about x_0:

$$u(x) = u(x_0) + u_x(x_0)(x - x_0) + \frac{1}{2}u_{xx}(x_0)(x - x_0)^2 + \mathcal{O}(|x - x_0|^3).$$

So assuming we are given a grid of points $x_i, i = 0, ..., m$, put $\overline{x}_i := x_i - x_0$; summing the above line over i with x evaluated on each x_i gives

$$\sum_{i=0}^{m} \alpha_i u(x_i) = u(x_0)\left(\sum_{i=0}^{m} \alpha_i\right) + u_x(x_0)\left(\sum_{i=0}^{m} \alpha_i \overline{x}_i\right) + \frac{1}{2}u_{xx}(x_0)\left(\sum_{i=0}^{m} \alpha_i \overline{x}_i^2\right) + \mathcal{O}\left(\sum \overline{x}_i^2\right).$$

So, suppose we want an approximation of $u_x(x_0)$; then, we need to cancel the first and third paranthesed terms and set the second to 1;

$$\sum \alpha_i = 0, \qquad \sum \alpha_i \overline{x}_i = 1, \qquad \sum \alpha_i \overline{x}_i^2 = 0.$$

(Alternatively, we can just restrict this last term to be $\mathcal{O}(|x|^2)$, or some similar consistency result.) To discuss existence/uniqueness of such schemes, we define first the $k \times m$-*Vandermonde matrix* associated to a set of points $\{x_0, ..., x_m\}$,

$$V(x_0, ..., x_m) := \begin{pmatrix} 1 & \dots & 1 \\ \overline{x}_0 & \dots & \overline{x}_m \\ \overline{x}_0^2 & \dots & \overline{x}_m^2 \\ \vdots & & \vdots \\ \overline{x}_0^k & \dots & \overline{x}_m^k \end{pmatrix}.$$

For $k = 1$, notice that

$$V\alpha = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

has solution identical to our first scheme for the first derivative. Similarly, for $k = 2$,

$$V\alpha = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

gives a stencil for the second derivative.

If $m = k$, V is square and so has unique solution (assuming invertibility). If $m > k$, there are multiple solutions necessarily; we can specify a solution by adding more constraints to make it square.

Extending to higher-dimensions is similar, where the Taylor-expansion logic leads to extra cross terms being involved. So, for instance, in 2 dimensions,

$$V = \begin{pmatrix} 1 & \dots & 1 \\ \overline{x}_0 & \dots & \overline{x}_m \\ \overline{y}_0 & \dots & \overline{y}_m \\ \overline{x}_0^2 & \dots & \overline{x}_m^2 \\ \overline{y}_0^2 & \dots & \overline{y}_m^2 \\ \overline{x}_0\overline{y}_0 & \dots & \overline{x}_m\overline{y}_m \\ \vdots & & \vdots \end{pmatrix}.$$

Of course, we aren't just restricting to approximation single derivatives in this way; for instance with the above V and RHS set to $(0,0,0,2,2,0)^t$, we obtain an approximation of the Laplacian.

To apply this to numerically solving and ODE, we consider the 1-dimensional Poisson's equation with mixed boundary conditions,

$$-u_{xx} = f(x), x \in (0,1) \qquad u(0) = a, u_x(1) = c.$$

We discretize with a uniform grid $\overline{x} = (0, h, \dots, nh, 1)$. On the interior points, we'll use a centered difference scheme;

$$f(x_i) = -u_{xx}(x_i) = -\frac{1}{h^2}(u_{i-1} - 2u_i + u_{i+1}) + O(h^2), \qquad 1 \leq i \leq n.$$

We know $u(x_0) = u_0 = a$ from the Dirichlet boundary condition, so on the left-hand-most-point we obtain

$$-\frac{a - 2u_1 + u_2}{h^2} = f(x_1).$$

For the Neumann boundary condition, we don't have access to information about "u_{n+2}", so we can't directly use the centered scheme above. One idea would be to naively use a backward difference scheme, putting

$$c = u_x(1) = \frac{u_{n+1} - u_n}{h},$$

but this is of order $O(h)$, so not amazing. However, suppose we use a centered difference approximation of the first derivative here, then we'd have

$$O(h^2) + \frac{u_{n+2} - u_n}{2h} = c.$$

Again, u_{n+2} is not defined. However, we can employ what is called a "ghost" point; assume there is some point x_{n+2}, and solve for what it should be using the interior scheme. Namely, if we put $u_{n+2} = 2hc + u_n$, then the interior scheme would say

$$-\frac{u_n - 2u_{n+1} + u_{n+2}}{h^2} = f(x_{n+1}),$$

which implies

$$-2\frac{u_n - 2(u_n + 1)}{h^2} = f(x_{n+1}) + \frac{2c}{h},$$

which gives now a second-order approximation.

§3.1 Error, Consistency and Stability

We'll discuss the results here for the specific instance of the Poisson equation, $u'' = f$, for sake of concreteness, but they hold in a more general setting. Let U be a discrete approximation of u (i.e., $U_i \approx u(x_i)$) and A a matrix for which $AU = F \approx f$.

↪**Definition 3.1** (Local Truncation Error (LTE)): Plug the "true" solution $\hat{u} = (u_i)$ into the FD scheme, and put τ for the difference vector, i.e.

$$\tau := A\hat{u} - F.$$

↪**Definition 3.2** (Global Truncation Error (GTE)): Put $E := U - \hat{u}$, called the GTE. Note that $AE = -\tau$.

↪**Definition 3.3** (Stability): Suppose we have a parametrized family of discretizations by some (maximum, say) grid size h, so A^h, E^h, τ^h are all given. Following from the above, we know that

$$\left\|E^h\right\| \leq \left\|\left(A^h\right)^{-1}\right\|\left\|\tau^h\right\|.$$

we say the parametrized scheme is stable if norm((A^h)^(-1)) if bounded from above uniformly in h for sufficiently small h.

↪**Definition 3.4** (Consistency): The scheme above is consistent if $\left\|\tau^h\right\| \to 0$ as $h \to 0$.

↪**Definition 3.5** (Convergence): The scheme above converges if $\left\|E^h\right\| \to 0$ as $h \to 0$.

↪**Theorem 3.1** (Lax Equivalence): A scheme is Consistent and stable ⇔ it is convergent.

§4 Spectral Methods

The previous section lead to schemes that were of $\mathcal{O}(h^p)$ error for some fixed p. Our schemes here lead to $\mathcal{O}(h^p)$ error for all p *if* $u \in C^\infty$. Such *spectral* methods have limited domain of application (namely linear equations, simple boundary conditions, and smooth functions), but for such problems they are very good.

Suppose we have a discretization of a periodic domain, $x_1 = h, x_2 = 2h, \ldots, x_N = 2\pi$. Using finite difference, i.e. setting ansatz $u'(x_i) \approx \sum_j \alpha_{ij} u_j$, we obtain the following classes of order of convergence:

$$\begin{aligned}
&\mathcal{O}(h^2) : u'(x_i) \approx \frac{1}{2h}\left(u_{j+1} - u_{j-1}\right) && \text{3 pts}\\
&\mathcal{O}(h^4) : u'(x_i) \approx \frac{1}{12h}\left(-u_{j+2} + 8u_{j+1} - 8u_{j-1} + u_{j-2}\right) && \text{5 pts}\\
&\mathcal{O}(h^6) : u'(x_i) \approx \frac{1}{60h}\left(u_{j+3} - 9u_{j+2} + 45u_{j+1} - 45u_{j-1} + 9u_{j-2} - u_{j-3}\right) && \text{7 pts}\\
&\qquad\qquad\ddots
\end{aligned}$$

In the limit, we would like to use all $N = \frac{2\pi}{h}$ points, which would give us *spectral* (higher than any polynomial) order, $O\left(h^N\right) = O\left(h^{1/h}\right)$.

Recall that

$$\cot\left(\frac{nh}{2}\right) = \frac{2}{nh} - \frac{nh}{6} - \cdots,$$

so that

$$\begin{aligned} u'(x_i) \approx & \frac{1}{2}\cot\left(\frac{h}{2}\right)\left[u_{j+1} - u_{j-1}\right] \\ & -\frac{1}{2}\cot\left(\frac{2h}{2}\right)\left[u_{j+2} - u_{j-2}\right] \\ & +\frac{1}{2}\cot\left(\frac{3h}{2}\right)\left[u_{j+3} - u_{j-3}\right] \\ & -\cdots, \end{aligned}$$

as $N \to \infty$.

Now, just as we could write $u = \left(u_j\right), w = \left(w_j\right) \approx \left(u'\left(x_j\right)\right)$ as vectors and then, for some fixed stencil, find a banded, sparse, circulant, Toeplitz matrix D such that our FD scheme may be written

$$w = Du,$$

we can find a similar matrix for a spectral-type method; however, it will no longer be sparse. For, say, $N = 6$ points, we get the matrix

$$D_6 = \begin{pmatrix} 0 & \alpha_1 & -\alpha_2 & \alpha_3 & -\alpha_4 & \alpha_5 \\ -\alpha_1 & \cdots & & & & \\ \alpha_2 & \cdots & & & & \\ -\alpha_3 & \cdots & & & & \\ \alpha_4 & \cdots & & & & \\ -\alpha_5 & \cdots & & & & 0 \end{pmatrix}$$

where $\alpha_j = \frac{1}{2}\cot\left(\frac{jh}{2}\right)$. How can we actually use such a D_6? It kind of sucks because of its far from sparse structure.

§5 Some Background on PDEs