Agenda

- INTERPOLATION is a way of determining an expansion of a function u in terms of some ORTHONORMAL BASIS FUNCTIONS alternative to Galerkin spectral projections
- Assuming that $S_N = \operatorname{span}\{e^{i0k}, \dots, e^{\pm iNx}\}$, we can determine an INTERPOLANT $v \in S_N$ of u, such that v coincides with u at $2N+1$ points $\{x_j\}_{|j|\le N}$ defined by

$$x_j = jh, \quad |j| \le N, \quad \text{where} \quad h = \frac{2\pi}{2N+1}$$

- For the interpolant we set $v(x) = \sum_{|k|\le N} a_k e^{ikx}$ where the coefficients a_k, $k = 1, \dots, N$ can be determined by solving the algebraic system

$$\sum_{|k|\le N} e^{ikx_j} a_k = u(x_j), \quad |j| \le N$$

with the matrix $\mathbb{A}_{kj} = e^{ikx_j}, \quad k, j = -N, \dots, N$

- The system can be rewritten as

$$\sum_{|k|\le N} W^{jk}\, a_k = u(x_j), \quad |j| \le N$$

where $W = e^{ih} = e^{\frac{2i\pi}{2N+1}}$ is the principal root of order $(2N+1)$ of unity (since $W^{jk} = \left(e^{ih}\right)^{jk}$)

Theorem

The matrix $[\mathbb{W}]_{jk} = W^{jk}$ *is unitary , i.e.* $\mathbb{W}^T\,\overline{\mathbb{W}} = \mathbb{I}(2N+1)$

Proof.

Examine the expression

$$\frac{1}{2N+1}\mathbb{W}^T\,\overline{\mathbb{W}} = \mathbb{I} \quad \Longrightarrow \quad \frac{1}{2N+1}\sum_{|j|\le N} W^{jk}W^{-jl} = \delta_{kl}$$

- If $k = l$, then $W^{jk}W^{-jl} = W^{j(k-l)} = W^0 = 1$
- If $k \neq l$, define $\omega = W^{k-l}$, then

$$\frac{1}{2N+1}\sum_{|j|\le N} W^{jk}W^{-jl} = \frac{1}{2N+1}\sum_{|j|\le N}\omega^j = \frac{1}{M}\sum_{j'=0}^{M-1}\omega^{j'}$$

where $M = 2N+1$, $j' = j$ if $0 \le j \le N$ and $j' = j + M$ if $-N \le j < 0$, so that $\omega^{j+M} = \omega^{j'}$. The proof is completed by using the expression for the sum of a finite geometric series

$$(1-\omega)\sum_{j'=0}^{M-1}\omega^{j'} = 1 - \omega^M = 0.$$

- Since the matrix $\mathbb{W}$ is unitary and hence its INVERSE is given by its TRANSPOSE , the Fourier coefficients of the INTERPOLANT of u in S_N can be calculated as follows:

$$a_k = \frac{1}{2N+1} \sum_{|j| \le N} z_j W^{-jk}, \quad \text{where} \quad z_j = u(x_j)$$

- The mapping

$$\{z_j\}_{|j|\le N} \longrightarrow \{a_k\}_{|k|\le N}$$

is referred to as DISCRETE FOURIER TRANSFORM (DFT)

- Straightforward evaluation of the expressions for a_k, $k = -N, \dots, N$ (matrix–vector products) would result in the computational cost $\mathcal{O}(N^2)$; clever factorization of this operation, known as the FAST FOURIER TRANSFORMS (FFT) , reduces this cost down to $\mathcal{O}(N \log(N))$

- See `www.fftw.org` for one of the best publicly available implementations of the FFT.

- Let $P_C : C^0_p(I) \to S_N$ be the mapping which associates with u its interpolant $v \in S_N$. Let $(\cdot,\cdot)_N$ be the GAUSSIAN QUADRATURE approximation of the inner product $(\cdot,\cdot)$

$$(u,v) = \int_{-\pi}^{\pi} u\overline{v}\,dx \cong \frac{1}{2N+1}\sum_{|j|\le N} u(x_j)\overline{v(x_j)} \triangleq (u,v)_N$$

- By construction, the operator P_C satisfies:

$$(P_C u)(x_j) = u(x_j), \quad |j| \le N$$

and therefore also (orthogonality of the defect to S_N)

$$(u - P_C u, v_N)_N = 0, \quad \forall v_N \in S_N$$

- By the definition of P_N we have

$$(u - P_N u, v_N) = 0, \quad \forall v_N \in S_N$$

- Thus, $P_C u(x) = \sum_{k=-N}^{N}(u, e^{ikx})_N e^{ikx}$ can be obtained analogously to $P_N u(x) = \sum_{k=-N}^{N}(u, e^{ikx}) e^{ikx}$ by replacing the scalar product $(\cdot,\cdot)$ with the DISCRETE SCALAR PRODUCT $(\cdot,\cdot)_N$

Corollary

Thus, the INTERPOLATION COEFFICIENTS a_k *are equivalent to the* FOURIER SPECTRAL COEFFICIENTS $\hat{u}_k$ *when the latter are evaluated using the* GAUSSIAN QUADRATURES .

Theorem

The two scalar products coincide on S_N*, i.e.*

$$(u_N, v_N) = (u_N, v_N)_N, \quad \forall u_N, v_N \in S_N,$$

hence for $u \in S_N$*,* $\hat{u}_k = a_k$*,* $k = -N, \dots, N$*.*

Proof.

Examine the numerical integration formula $\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx \cong \frac{1}{2N+1}\sum_{|j|\le N} f(x_j)$; then for every $f = \sum_{k=-2N}^{2N} \hat{u}_k e^{ikx} \in S_{2N}$ we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ikx}\,dx = \frac{1}{2N+1}\sum_{|j|\le N} e^{ikx_j} = \frac{1}{2N+1}\sum_{|j|\le N} W^{jk} = \begin{cases} 1 & k=0 \\ 0 & \text{otherwise} \end{cases}, \quad k = 0, \dots, 2N$$

Thus, for the uniform distribution of x_j, the Gaussian (trapezoidal) formula is EXACT for $f \in S_{2N}$. $\square$

Relation between Fourier coefficients $\hat{u}_k$ of a function $u(x)$ and Fourier coefficients a_k of its interpolant; assume that $u(x) \notin S_N$

$$\hat{u}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} u \overline{W}_k \, dx, \qquad W_k(x) = e^{ikx}$$

$$a_k = \frac{1}{2N+1} \sum_{|j| \le N} u(x_j) \overline{W_k(x_j)}$$

Theorem

For $u \in C_p^0(I)$ we have the relation

$$a_k = \sum_{l \in \mathbb{Z}} \hat{u}_{k+lM}, \quad \textit{where } M = 2N+1$$

Proof.

Consider the set of basis functions (in $L_2(I)$) $U_k = e^{ikx}$. We have:

$$(U_k, U_n)_N = \frac{1}{2N+1} \sum_{|j| \le N} U_k(x_j)\overline{U_n(x_j)} = \frac{1}{2N+1} \sum_{|j| \le N} W^{j(k-n)} = \begin{cases} 1 & k = n \ (mod\ M) \\ 0 & \text{otherwise} \end{cases}$$

Since $P_C u = \sum_{|j| \le N} a_j W_j$, we infer from $(P_C u, W_k)_N = (u, W_k)_N$ that

$$a_k = (P_C u, W_k)_N = (u, W_k)_N = \left(\sum_{n \in \mathbb{Z}} \hat{u}_n W_n, W_k \right)_N = \sum_{n \in \mathbb{Z}} \hat{u}_n \left(W_n, W_k\right)_N = \sum_{l \in \mathbb{Z}} \hat{u}_{k+lM}$$

$\square$

- Thus

$$u(x_j) = v(x_j) = \sum_{k=-\infty}^{\infty} \hat{u}_k e^{ikx_j} = \sum_{|k|\le N} a_k e^{ikx_j} = \sum_{|k|\le N} \left(\hat{u}_k + \sum_{l\in\mathbb{Z}\setminus\{0\}} \hat{u}_{k+lM} \right) e^{ikx_j}$$

Corollary (Extremely Important Corollary Concerning Interpolation)

two trigonometric polynomials e^{ik_1x} and e^{ik_2x} with different frequencies k_1 and k_2 are equal at the collocation points x_j, $|j| \le N$ when

$$k_2 - k_1 = l(2N+1), \quad l = 0, \pm 1, \dots.$$

Therefore, given a set of values at the collocation points x_j, $|j| \le N$, it is impossible to distinguish between e^{ik_1x} and e^{ik_2x}. This phenomenon is referred to as ALIASING *.*

Note, however, that the modes appearing in the alias term correspond to frequencies larger than the cut–off frequency N.

Theorem (Error Estimates in $H_p^s(I)$)

Suppose $s \leq r$, $r > \frac{1}{2}$ are given, then there exists a constant C such that if $u \in H_p^r(I)$, we have

$$\|u - P_C u\|_s \leq C(1 + N^2)^{\frac{s-r}{2}} \|u\|_r$$

Outline of the proof.

Note that P_C leaves S_N invariant, therefore $P_C P_N = P_N$ and we may thus write

$$u - P_C u = u - P_N u + P_C(P_N - I)u$$

Setting $w = (I - P_N)u$ and using the "triangle inequality" we obtain

$$\|u - P_C u\|_s \leq \|u - P_N u\|_s + \|P_C w\|_s$$

- The term $\|u - P_N u\|_s$ is upper–bounded using an earlier theorem
- Need to estimate $\|P_C w\|_s$ — straightforward, but tedious ...

$\square$

- Until now, we defined the Discrete Fourier Transform for an ODD number $(2N+1)$ of grid points
- FFT algorithms generally require an EVEN number of grid points
- We can define the discrete transform for an EVEN number of grid points by constructing the interpolant in the space $\tilde{S}_N$ for which we have $\dim(\tilde{S}_N) = 2N$. To do this we choose:

$$\tilde{x}_j = j\tilde{h}, \qquad -N+1 \le j \le N, \qquad \tilde{h} = \frac{\pi}{N}$$

- All results presented before can be established in the case with $2N$ grid points with only minor modifications
- However, now the N-th Fourier mode $\hat{u}_N$ does not have its complex conjugate! This coefficient is usually set to zero $(\hat{u}_N = 0)$ to avoid an uncompensated imaginary contribution resulting from differentiation
- ODD or EVEN collocation depending on whether $M = 2N+1$ or $M = 2N$

- Before we focused on representing the INTERPOLANT as a Fourier series $v(x_j) = \sum_{k=-N}^{N} a_k e^{ikx_j}$

- Alternatively, we can represent the INTERPOLANT using the nodal values as (assuming, for the moment, infinite domain $x \in \mathbb{R}$)

$$v(x) = \sum_{j=-\infty}^{\infty} u(x_j)\, C_j(x),$$

where $C_j(x)$ is a CARDINAL FUNCTION with the property that $C_j(x_i) = \delta_{ij}$ (i.e., generalization of the LAGRANGE POLYNOMIAL for infinite domain)

- In an infinite domain we have the WHITTAKER CARDINAL or SINC function

$$C_k(x) = \frac{\sin[\pi(x - kh)/h]}{\pi(x - kh)/h} = \operatorname{sinc}[(x - kh)/h],$$

where $\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$

Proof.

The Fourier transform of δ_{j0} is $\hat{\delta}(k) = h$ for all $k \in [-\pi/h, \pi/h]$; hence, the interpolant of δ_{j0} is $v(x) = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} e^{ikx}\, dk = \frac{\sin(\pi x/h)}{\pi x/h}$ $\square$

- Thus, the spectral interpolant of a function in an INFINITE domain is a linear combination of WHITTAKER CARDINAL functions
- In a PERIODIC DOMAIN we still have the representation

$$v(x) = \sum_{j=0}^{N-1} u(x_j) S_j(x),$$

but now the CARDINAL FUNCTIONS have the form

$$S_j(x) = \frac{1}{N} \sin\left[\frac{N(x - x_j)}{2}\right] \cot\left[\frac{(x - x_j)}{2}\right]$$

- Proof — similar to the previous (unbounded) case, except that now the interpolant in given by a DISCRETE Fourier Transform
- The relationship between the Cardinal Functions corresponding to the PERIODIC and UNBOUNDED domains

$$S_0(x) = \frac{1}{2N} \sin(Nx) \cot(x/2) = \sum_{m=-\infty}^{\infty} \operatorname{sinc}\left(\frac{x - 2\pi m}{h}\right)$$

- Two ways to calculate the derivative $w(x_j) = u'(x_j)$ based on the values $u(x_j)$, where $0 \le j \le 2N+1$; denote $U = [u_0, \dots, u_{2N+1}]^T$ and $U' = [u'_0, \dots, u'_{2N+1}]^T$

- Method One — approach based on differentiation in Fourier space:

  - calculate the vector of Fourier coefficients $\hat{U} = \mathbb{T}U$
  - apply the diagonal differentiation matrix $\hat{U}' = \hat{\mathbb{D}}\hat{U}$
  - return to real space via inverse Fourier transform $U = \mathbb{T}^T\hat{U}$

- Remark — formally we can write

$$U' = \mathbb{T}^T\hat{\mathbb{D}}\mathbb{T}\,U,$$

however in practice matrix operations are replaced by FFTs

- METHOD TWO — approach based on differentiation (in real space) of the interpolant $u'(x) = v'(x) = \sum_{j=0}^{N-1} u(x_j) S_j'(x)$, where the cardinal function has the following derivatives

$$S_j'(x_k) = \begin{cases} 0, & j = k \\ \frac{1}{2}(-1)^{k+j} \cot\left[\frac{(k-j)h}{2}\right], & j \neq k \end{cases}$$

- Thus, since the interpolant is a linear combination of shifted Cardinal Functions, the differentiation matrix has the form of a TOEPLITZ CIRCULANT matrix

$$\mathbb{D} = \begin{bmatrix} 0 & & & -\frac{1}{2}\cot[(1h)/2] \\ -\frac{1}{2}\cot[(1h)/2] & \ddots & \ddots & \frac{1}{2}\cot[(2h)/2] \\ \frac{1}{2}\cot[(2h)/2] & & \ddots & -\frac{1}{2}\cot[(3h)/2] \\ -\frac{1}{2}\cot[(3h)/2] & & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2}\cot[(1h)/2] \\ \frac{1}{2}\cot[(1h)/2] & & & 0 \end{bmatrix}$$

- Higher–order derivatives obtained calculating $S^{(p)}(x_j)$

- We are interested in a PARTIAL DIFFERENTIAL EQUATION (a boundary value problem) of the general form $\mathcal{L}u = f$
- We will look for solutions in the form:

$$
\begin{aligned}
u_N(x) &= \sum_{|k|\leq N} \hat{u}_k e^{ikx}, \\
&= \sum_{j=1}^{2N+1} u(x_j) S_j(x),
\end{aligned}
$$

  where $S_j(x)$ is the periodic cardinal function centered at x_j
- For the above model problem we will analyze:
  - spectral Galerkin method
  - spectral Collocation method
    - variant with the FOURIER COEFFICIENTS $\hat{u}_k$ as the unknowns
    - variant with the NODAL VALUES $u(x_j)$ as the unknowns

- Consider the following 1D second–order elliptic problem in a periodic domain $\Omega = [0, 2\pi]$

$$\mathcal{L}u \triangleq \nu u'' - au' + bu = f,$$

  where ν, a and b are constant and $f = f(x)$ is a smooth 2π–periodic function.

- For $\nu = 10$, $a = 1$, $b = 5$ and the RHS function

$$f(x) = e^{\sin(x)}\left[\nu(\cos^2(x) - \sin(x)) - a\cos(x) + b\right]$$

  the solution is

$$u(x) = e^{\sin(x)}$$

- For the GALERKIN approach we are interested in 2π–periodic solutions in the form

$$u_N(x) = \sum_{|k| \le N} \hat{u}_k e^{ikx}$$

- RESIDUAL

$$R_N(x) = \mathcal{L}u_N - f = \sum_{|k|\le N} \hat{u}_k \mathcal{L}e^{ikx} - f$$

- Cancellation of the residual in the mean (setting the projections on the basis functions $W_n(x) = e^{inx}$ equal to zero)

$$(R_N, W_n) = \sum_{k=-N}^{N} \hat{u}_k(\mathcal{L}e^{ikx}, e^{inx}) - (f, e^{inx}) = 0, \quad n = -N, \dots, N$$

- Noting that $\mathcal{L}e^{ikx} = (-\nu k^2 - iak + b)e^{ikx} \triangleq \mathcal{G}_k e^{ikx}$ we obtain

$$\sum_{k=-N}^{N} \mathcal{G}_k \hat{u}_k \int_0^{2\pi} e^{i(k-n)}\, dx = \hat{f}_n, \quad n = -N, \dots, N$$

- Assuming $\mathcal{G}_k \neq 0$, we obtain the GALERKIN EQUATIONS for $\hat{u}_k$

$$\mathcal{G}_k \hat{u}_k = \hat{f}_k, \qquad\qquad k = -N, \dots, N$$

  - The Galerkin equations are DECOUPLED
  - Since u is real, it is necessary to calculate $\hat{u}_k$ for $k \geq 0$ only

- RESIDUAL (with the expansion coefficients $\hat{u}_k$ as unknowns)

$$R_N(x) = \mathcal{L}u_N - f = \sum_{|k|\le N} \hat{u}_k \mathcal{L} e^{ikx} - f$$

- Cancelling the residual pointwise at the collocation points x_j, $j = 1, \dots, M$

$$\sum_{k=-N}^{N} (\mathcal{G}_k \hat{u}_k - \tilde{f}_k) e^{ikx_j} = 0, \quad j = 1, \dots, M$$

where (note the ALIASING ERROR) $\tilde{f}_k = \hat{f}_k + \sum_{l \in \mathbb{Z} \setminus \{0\}} \hat{f}_{k+lM}$

- Thus, the COLLOCATION EQUATIONS for the Fourier coefficients

$$\mathcal{G}_k \hat{u}_k = \tilde{f}_k = \hat{f}_k + \sum_{l \in \mathbb{Z} \setminus \{0\}} \hat{f}_{k+lM}, \quad k = -N, \dots, N$$

  - Formally, the GALERKIN and COLLOCATION methods are DISTINCT
  - In practice, the projection (f, e^{ikx}) is evaluated using FFT and therefore also involves aliasing errors. Therefore, for the present problem, the two approaches are NUMERICALLY EQUIVALENT .

- RESIDUAL (with the nodal values $u_N(x_j)$, $j = 1, \dots, M$, as unknowns)

$$R_N(x) = \mathcal{L}u_N - f$$

- Cancelling the residual pointwise at the collocation points x_j, $j = 1, \dots, M$

$$[R_N(x_1), \dots, R_N(x_M)]^T = \mathbb{L}U_N - F = (\nu\mathbb{D}_2 - a\mathbb{D}_1 + b\mathbb{I})U_N - F = 0,$$

where $U_N = [u_N(x_1), \dots, u_N(x_M)]^T$ and $\mathbb{D}_1$ and $\mathbb{D}_2$ are the differentiation matrices.

- Derivation of the DIFFERENTIATION MATRICES

$$\left.\begin{aligned} u_N^{(p)}(x_j) &= \sum_k (ik)^p \hat{u}_k e^{ikx_j} \\ \hat{u}_k &= \frac{1}{M}\sum_{j=1}^{M} u_N(x_j) e^{-ikx_j} \end{aligned}\right\} \Longrightarrow \quad u_N^{(p)}(x_i) = \sum_{j=1}^{M} d_{ij}^{(p)} u_N(x_j)$$

- Differentiation Matrices (for even collocation, i.e., $I_N = -N+1, \dots, N$ and $M = 2N$) with $h_{ij} = (i-j)\pi/N$

$$d_{ij}^{(1)} = \begin{cases} \frac{1}{2}(-1)^{i+j}\cot(h_{ij}) & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases}, \quad d_{ij}^{(2)} = \begin{cases} \frac{1}{4}(-1)^{i+j}N + \frac{(-1)^{i+j+1}}{2\sin^2(h_{ij})} & \text{if} \quad i \neq j \\ -\frac{(N-1)(N-2)}{12} & \text{if} \quad i = j \end{cases}$$

- Remarks:
  - The differentiation matrices are full (and not so well–conditioned ...), so the system of equations for $u_N(x_j)$ is now COUPLED
  - For constant coefficient PDEs the present approach is therefore inferior to the first collocation approach with the Fourier coefficients used as unknowns
  - Note the relationship to the banded matrices obtained when approximating differential operators using finite differences
- QUESTION — Derive the above differentiation matrices, also for the case of odd collocation

Nyquist-Shannon Sampling Theorem

- If a periodic function $f(x)$ has a Fourier transform $\hat{f}_k = 0$ for $|k| > M$, then it is completely determined by providing the function values at a series of points spaced $\Delta x = \frac{1}{2M}$ apart. The values $f_n = f(\frac{n}{2M})$ are called the SAMPLES OF $f(x)$.
- The minimum sampling frequency that allows for reconstruction of the original signal, that is $2M$ samples per unit distance, is known as the NYQUIST FREQUENCY . The time in between samples is called the NYQUIST INTERVAL .
- The NYQUIST–SHANNON SAMPLING THEOREM is a fundamental tenet in the field of INFORMATION THEORY (originally formulated by Nyquist in 1928, but formally proved by Shannon only in 1949)