

PART IV

Spectral Methods

METHOD OF WEIGHTED RESIDUALS (I)

- **Spectral Methods** belong to the broader category of **Weighted Residual Methods**, for which approximations are defined in terms of series expansions, such that some quantity (**residual**, or error) is set to be zero in some approximate sense
- In general, an approximation $u_N(x)$ to $u(x)$ is constructed using a set of basis functions $\phi_k(x)$, $k = 0, \dots, N$ (note that $\phi_k(x)$ need not be **orthogonal**)

$$u_N(x) \triangleq \sum_{k \in I_N} \hat{u}_k \phi_k(x), \quad a \leq x \leq b$$

- **Residual** for
 - Problem of approximating a function u :

$$R_N(x) = u - u_N$$

- Approximate solution to a differential equation $\mathcal{L}u - f = 0$:

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

METHOD OF WEIGHTED RESIDUALS (II)

- Cancellation of the residual R_N in the following sense:

$$(R_N, \Psi_i)_{w_*} = \int_a^b w_* R_N \Psi_i dx = 0, \quad i \in I_N,$$

where $\Psi_i(x)$, $i \in I_N$ are the **trial (test) functions**

- **Spectral Method** is obtained by:
 - selecting the basis functions ϕ_k to form an orthogonal system under the weight w :

$$(\phi_i, \phi_k)_w = \delta_{ik}, \quad i, k \in I_N \text{ and}$$

- selecting the trial functions to coincide with the basis functions:

$$\Psi_k = \phi_k, \quad k \in I_N$$

with the weights $w_* = w$ (**Galerkin approach**), or

- selecting the trial functions as

$$\Psi_k = \delta(x - x_k), \quad x_k \in (a, b),$$

where x_k are chosen in a non-arbitrary manner, and the weights are

$w_* = 1$ (**Collocation**, “pseudo-spectral” approach)

METHOD OF WEIGHTED RESIDUALS (III)

- Note that the residual R_N vanishes
 - in the mean sense in the Galerkin approach
 - at the points x_k in the collocation approach

APPROXIMATION OF FUNCTIONS (I) — GALERKIN METHOD

- The residual

$$R_N(x) = u - u_N = u - \sum_{k=0}^N \hat{u}_k \phi_k$$

- Cancellation of the residual in the mean

$$(R_N, \phi_i)_w = \int_a^b \left(u - \sum_{k=0}^N \hat{u}_k \phi_k \right) \bar{\phi}_i w dx = 0, \quad i = 0, \dots, N$$

- Orthogonality of the basis / trial functions thus allows us to determine the coefficients \hat{u}_k by evaluating the expressions

$$\hat{u}_k = \int_a^b u \bar{\phi}_k w dx, \quad k = 0, \dots, N$$

- Note that, for this problem, the Galerkin approach is equivalent to the [Least Squares Method](#).

APPROXIMATION OF FUNCTIONS (II) — COLLOCATION METHOD

- The residual

$$R_N(x) = u - u_N = u - \sum_{k=0}^N \hat{u}_k \phi_k$$

- Pointwise cancellation of the residual

$$\sum_{k=0}^N \hat{u}_k \phi_k(x_i) = u(x_i), \quad i = 0, \dots, N$$

Determination of the coefficients \hat{u}_k thus requires solution of an algebraic system. Existence and uniqueness of solutions requires that $\det\{\phi_k(x_i)\} \neq 0$ (condition on the choice of the collocation points x_j)

- As will be shown later, for a judicious choice of the collocation points x_j the above system can be decoupled and therefore determination of \hat{u}_k may be reduced to evaluation of simple expressions
- For this problem the collocation method thus coincides with an interpolation technique based on the set $\{x_j\}$

APPROXIMATION OF PDES (I) — GALERKIN METHOD

- Consider a generic PDE problem

$$\begin{cases} \mathcal{L}u - f = 0 & a < x < b \\ \mathcal{B}_- u = g_- & x = a \\ \mathcal{B}_+ u = g_+ & x = b, \end{cases}$$

where \mathcal{L} is a linear, second-order differential operator, and \mathcal{B}_- and \mathcal{B}_+ represent appropriate boundary conditions (Dirichlet, Neumann, or Robin)

- Reduce the problem to an equivalent [homogeneous](#) formulation via a “lifting” technique, i.e., substitute $u = \bar{u} + v$, where \bar{u} is an arbitrary function satisfying the boundary conditions above and the new (homogeneous) problem for v is

$$\begin{cases} \mathcal{L}v - h = 0 & a < x < b \\ \mathcal{B}_- v = 0 & x = a \\ \mathcal{B}_+ v = 0 & x = b, \end{cases}$$

where $h = f - \mathcal{L}\bar{u}$

- The reason for this transformation is that the basis functions ϕ_k (usually) satisfy homogeneous boundary conditions.

APPROXIMATION OF PDES (II) — GALERKIN METHOD

- The residual

$$R_N(x) = \mathcal{L}v_N - h, \quad \text{where } v_N = \sum_{k=0}^N \hat{v}_k \phi_k(x)$$

satisfies (“by construction”) the boundary conditions

- Cancellation of the residual in the mean

$$(R_N, \phi_i)_w = (\mathcal{L}v_N - h, \phi_i)_w, \quad i = 0, \dots, N$$

Thus

$$\sum_{k=0}^N \hat{v}_k (\mathcal{L}\phi_k, \phi_i)_w = (h, \phi_i)_w, \quad i = 0, \dots, N,$$

where the scalar product $(\mathcal{L}\phi_k, \phi_i)_w$ can be accurately evaluated using properties of the basis functions ϕ_i and $(h, \phi_i)_w = \hat{h}_i$

- An $(N+1) \times (N+1)$ algebraic system is obtained.

APPROXIMATION OF PDES (III) — COLLOCATION METHOD

- The residual (corresponding to the original inhomogeneous problem)

$$R_N(x) = \mathcal{L}u_N - f, \text{ where } u_N = \sum_{k=0}^N \hat{u}_k \phi_k(x)$$

- Pointwise cancellation of the residual, including the boundary nodes:

$$\begin{cases} \mathcal{L}u_N(x_i) = f(x_i) & i = 1, \dots, N-1 \\ \mathcal{B}_- u_N(x_0) = g_- \\ \mathcal{B}_+ u_N(x_N) = g_+, \end{cases}$$

which results in an $(N+1) \times (N+1)$ algebraic system. Note that depending on the properties of the basis $\{\phi_0, \dots, \phi_N\}$, this system may be **singular**.

- Sometimes an alternative formulation is useful, where the nodal values $u_N(x_j)$ $j = 0, \dots, N$, rather than the expansion coefficients \hat{u}_k $k = 0, \dots, N$ are unknown. The advantage is a convenient form of the expression for the derivative

$$u_N^{(p)}(x_i) = \sum_{j=0}^N d_{ij}^{(p)} u_N(x_j)$$

ORTHONORMAL SYSTEMS (I) — CONSTRUCTION

- Let \mathbf{H} be a separable Hilbert space and \mathcal{T} a compact Hermitian operator. Then, there exists a sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ and $\{W_n\}_{n \in \mathbb{N}}$ such that
 - $\lambda_n \in \mathbb{R}$,
 - the family $\{W_n\}_{n \in \mathbb{N}}$ forms a complete basis in \mathbf{H}
 - $\mathcal{T}W_n = \lambda_n W_n$ for all $n \in \mathbb{N}$
- Systems of orthogonal functions are therefore related to spectra of certain operators, hence the name **SPECTRAL METHODS**

ORTHONORMAL SYSTEMS (II) — EXAMPLE

- Let $\mathcal{T} : L_2(0, \pi) \rightarrow L_2(0, \pi)$ be defined for all $f \in L_2(0, \pi)$ by $\mathcal{T}f = u$, where u is the solution of the Dirichlet problem

$$\begin{cases} -u'' = f \\ u(0) = u(\pi) = 0 \end{cases}$$

Compactness of \mathcal{T} follows from the Lax–Milgram lemma and compact embeddedness of $H^1(0, \pi)$ in $L_2(0, \pi)$

- Eigenvalues and eigenvectors

$$\lambda_k = \frac{1}{k^2} \text{ and } W_k = \sqrt{2} \sin(kx) \text{ for } k \geq 1$$

- Thus, each function $u \in L_2(0, \pi)$ can be represented as

$$u(x) = \sqrt{2} \sum_{k \geq 1} \hat{u}_k W_k(x),$$

where $\hat{u}_k = (u, W_k)_{L_2} = \frac{\sqrt{2}}{\pi} \int_0^\pi u(x) \sin(kx) dx$

- Uniform (pointwise)** convergence is not guaranteed (only in L_2 sense)!

ORTHONORMAL SYSTEMS (III) — EXAMPLE

- Let $\mathcal{T} : L_2(0, \pi) \rightarrow L_2(0, \pi)$ be defined for all $f \in L_2(0, \pi)$ by $\mathcal{T}f = u$, where u is the solution of the Neumann problem

$$\begin{cases} -u'' + u = f \\ u'(0) = u'(\pi) = 0 \end{cases}$$

Compactness of \mathcal{T} follows from the Lax–Milgram lemma and compact embeddedness of $H^1(0, \pi)$ in $L_2(0, \pi)$

- Eigenvalues and eigenvectors

$$\lambda_k = \frac{1}{1+k^2} \text{ and } W_0(x) = 1, W_k = \sqrt{2} \cos(kx) \text{ for } k > 1$$

- Thus, each function $u \in L_2(0, \pi)$ can be represented as

$$u(x) = \sqrt{2} \sum_{k \geq 0} \hat{u}_k W_k(x),$$

where $\hat{u}_k = (u, W_k)_{L_2} = \frac{\sqrt{2}}{\pi} \int_0^\pi u(x) \cos(kx) dx$

- Uniform (pointwise)** convergence is not guaranteed (only in L_2 sense)!

ORTHONORMAL SYSTEMS (IV) — EXAMPLE

- Expansion in **sine series** good for functions vanishing on the boundaries
- Expansion in **cosine series** good for functions with first derivatives vanishing on the boundaries
- Combining sine and cosine expansions we obtain the **Fourier series expansion** with the basis functions (in $L_2(-\pi, \pi)$)

$$W_k(x) = e^{ikx}, \text{ for } k \geq 0$$

W_k form a Hilbert basis with better properties than sine or cosine series alone.

- Fourier series vs. Fourier transform —the Fourier transform of $u(x)$ vanishing outside the interval $(-\pi, \pi)$ takes the values $\sqrt{2\pi} \hat{u}$ at the points $k = 0, 1, 2, \dots$

ORTHONORMAL SYSTEMS (V) — POLYNOMIAL APPROXIMATION

- **Weierstrass Approximation Theorem** —To any function $f(x)$ that is continuous in $[a, b]$ and to any real number $\epsilon > 0$ there corresponds a polynomial $P(x)$ such that $\|P(x) - f(x)\|_{C(a,b)} < \epsilon$, i.e. the set of polynomials is **dense** in the Banach space $C(a, b)$ ($C(a, b)$ is the Banach space with the norm $\|f\|_{C(a,b)} = \max_{x \in [a,b]} |f(x)|$)
- Thus the power functions $x^k, k = 0, 1, \dots$ represent a natural basis in $C(a, b)$
- **Question** —Is this set of basis functions useful?

ORTHONORMAL SYSTEMS (VI) — EXAMPLE

- Find the polynomial \bar{P}_N (of order N) that best approximates a function $f \in L_2(a, b)$ [note that we will need the structure of a Hilbert space, hence we go to $L_2(a, b)$, but $C(a, b) \subset L_2(a, b)$], i.e.

$$\int_a^b [f(x) - \bar{P}_N(x)]^2 dx = \int_a^b [f(x) - P_N(x)]^2 dx$$

where

$$\bar{P}_N(x) = \bar{a}_0 + \bar{a}_1 x + \bar{a}_2 x^2 + \dots + \bar{a}_N x^N$$

- Using the formula $\bar{q}(e_j, e_k) = (f, e_k)$, $j = 0, \dots, N$, where $e_k = x^k$

$$\sum_{k=0}^N \bar{a}_k \int_a^b x^{k+j} dx = \int_a^b x^j f(x) dx$$

$$\sum_{k=0}^N \bar{a}_k \frac{b^{k+j+1} - a^{k+j+1}}{k+j+1} = \int_a^b x^j f(x) dx$$

- The resulting algebraic problem is **ill-conditioned**, e.g. for $a = 0$ and $b = 1$

$$[A]_{kj} = \frac{1}{k+j+1}$$

ORTHONORMAL SYSTEMS (VII) — POLYNOMIAL APPROXIMATION

- Much better behaved approximation problems are obtained with the use of **orthogonal basis functions**
- Such systems of **orthogonal basis functions** are derived by applying **Schmidt orthogonalization procedure** to the system $\{1, x, \dots, x^N\}$
- Various families of **ORTHOGONAL POLYNOMIALS** are obtained depending on the choice of:
 - the domain $[a, b]$ over which the polynomials are defined, and
 - the weight w characterizing the inner product (\cdot, \cdot) used for orthogonalization

ORTHONORMAL SYSTEMS (VIII) — ORTHOGONAL POLYNOMIALS

- Polynomials defined on the interval $[-1, 1]$
 - Legendre polynomials** ($w = 1$)

$$P_k(x) = \sqrt{\frac{2k+1}{2}} \frac{1}{2^k k!} \frac{d^k}{dx^k} (x^2 - 1)^k, \quad k = 0, 1, 2, \dots$$

- Jacobi polynomials** ($w = (1-x)^\alpha(1+x)^\beta$)

$$J_k^{(\alpha, \beta)}(x) = C_k (1-x)^{-\alpha} (1+x)^{-\beta} \frac{d^k}{dx^k} [(1-x)^{\alpha+k} (1+x)^{\beta+k}], \quad k = 0, 1, 2, \dots,$$

where C_k is a very complicated constant

- Chebyshev polynomials** ($w = \frac{1}{\sqrt{1-x^2}}$)

$$T_n(x) = \cos(k \arccos(x)), \quad k = 0, 1, 2, \dots,$$

Note that Chebyshev polynomials are obtained from Jacobi polynomials for $\alpha = \beta = -1/2$

ORTHONORMAL SYSTEMS (IX) — ORTHOGONAL POLYNOMIALS

- Polynomials defined on the interval $[0, +\infty]$
 - Laguerre polynomials** ($w = e^{-x}$)

$$L_k(x) = \frac{1}{k!} e^x \frac{d^k}{dx^k} (e^{-x} x^k), \quad k = 0, 1, 2, \dots$$

- Polynomials defined on the interval $[-\infty, +\infty]$
 - Hermite polynomials** ($w = 1$)

$$H_k(x) = \frac{(-1)^k}{(2^k k! \sqrt{\pi})^{1/2}} e^{x^2} \frac{d^k}{dx^k} e^{-x^2}, \quad k = 0, 1, 2, \dots$$

ORTHONORMAL SYSTEMS (X) — ORTHOGONAL POLYNOMIALS

- What is the relationship between **orthogonal polynomials** and eigenfunctions of a **compact operator Hermitian operator** (cf. Theorem on page 55)?
- Each of the aforementioned families of **orthogonal polynomials** forms the set of eigenvectors for the following **Sturm-Liouville problem**

$$\frac{d}{dx} \left[p(x) \frac{dy}{dx} \right] + [q(x) + \lambda r(x)] y = 0$$

$$a_1 y(a) + a_2 y'(a) = 0$$

$$b_1 y(b) + b_2 y'(b) = 0$$

for appropriately selected domain $[a, b]$ and coefficients p, q, r, a_1, a_2, b_1 and b_2 .

FOURIER SERIES (I) — CALCULATION OF FOURIER COEFFICIENTS

- Truncated Fourier series:

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

- The series involves $2N + 1$ complex coefficients of the form (with $w \equiv 1$):

$$\hat{u}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} u e^{-ikx} dx, \quad k = -N, \dots, N$$

- The expansion is redundant for real-values u — the property of **conjugate symmetry** $\hat{u}_k = \bar{\hat{u}}_{-k}$, which reduces the number of complex coefficients to $N + 1$; furthermore, $\Im(\hat{u}_k) \equiv 0$ for real u , thus one has $2N + 1$ **real** coefficients; in the real case one can work with positive frequencies only.
- Equivalent real representation:

$$u_N(x) = a_0 + \sum_{k=1}^N [a_k \cos(kx) + b_k \sin(kx)],$$

where $a_0 = \hat{u}_0$, $a_k = 2\Re(\hat{u}_k)$ and $b_k = 2\Im(\hat{u}_k)$.

FOURIER SERIES (II) — UNIFORM CONVERGENCE

- Consider a function u that is continuous, periodic (with the period 2π) and differentiable; note the following two facts:

– The Fourier coefficients are always less than the average of u

$$|\hat{u}_k| = \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} u(x) e^{ikx} dx \right| \leq M(u) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} |u(x)| dx$$

– If $v = u^{(\alpha)}$

$$\hat{u}_k = \frac{\hat{v}_k}{(ik)^\alpha}$$

- Then, using integration by parts, we have

$$\hat{u}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(x) e^{-ikx} dx = \frac{1}{2\pi} \left[u(x) \frac{e^{-ikx}}{-ik} \right]_{-\pi}^{\pi} - \frac{1}{2\pi} \int_{-\pi}^{\pi} u'(x) \frac{e^{-ikx}}{-ik} dx$$

- Repeating integration by parts p times

$$\hat{u}_k = (-1)^p \frac{1}{2\pi} \int_{-\pi}^{\pi} u^{(p)}(x) \frac{e^{-ikx}}{(-ik)^p} dx \implies |\hat{u}_k| \leq \frac{M(u^{(p)})}{|k|^p}$$

Therefore, the more regular is the function u , the more rapidly its Fourier coefficients tend to zero as $|n| \rightarrow \infty$

FOURIER SERIES (III) — UNIFORM CONVERGENCE

- We have

$$|\hat{u}_k| \leq \frac{M(u'')}{|k|^2} \implies \sum_{k \in \mathbb{Z}} |\hat{u}_k e^{ikx}| \leq \hat{u}_0 \sum_{n \neq 0} \frac{M(u'')}{n^2}$$

The latter series converges **absolutely**

- Thus, if u is twice continuously differentiable and its first derivative is continuous and periodic with period 2π , then its Fourier series $u_N = P_N u$ **converges uniformly** to u
- Spectral convergence** – if $\phi \in C_p^\infty(-\pi, \pi)$, then for all $\alpha > 0$ there exists a positive constant C_α such that $|\hat{\phi}_k| \leq \frac{C_\alpha}{|k|^\alpha}$, i.e., for a function with an infinite number of smooth derivatives, the Fourier coefficients vanish faster than algebraically

FOURIER SERIES (IV) — DISTRIBUTIONS

- Let $D'_p(I)$ be the dual space of $C_p^\infty(I)$, i.e., the space of periodic distributions with period 2π [$I = (-\pi, \pi)$]. The duality between $D'_p(I)$ and $C_p^\infty(I)$ is denoted by $\langle \cdot, \cdot \rangle$, i.e. for $f \in L_2(I)$ and $\phi \in C_p^\infty(I)$ we have $\langle f, \phi \rangle = (f, \phi)$

- Using $W_k = e^{ikx}$ (note that $W_k \in C_p^\infty(I)$), the Fourier series for $f \in D'_p(I)$ can be written as

$$\hat{f}_k = \langle f, W_k \rangle$$

We have for any $\phi \in C_p^\infty(I)$

$$\langle f, \phi \rangle = \langle f, \sum_{k \in \mathbb{Z}} \hat{\phi}_k W_k \rangle = \sum_{k \in \mathbb{Z}} \langle f, W_k \rangle \hat{\phi}_k \implies \langle f, \phi \rangle = \sum_{k \in \mathbb{Z}} \hat{f}_k \hat{\phi}_k$$

- This, given rapid decrease of $\hat{\phi}_k$, the Fourier coefficient of f may increase slowly — $f \in D'_p(I)$ iff there exists $q > 0$ such that $\lim_{|k| \rightarrow \infty} \frac{\hat{f}_k}{(1+k^2)^q} = 0$
- The Fourier series of a distribution $f \in D'_p(I)$ converges to f in $D'_p(I)$

FOURIER SERIES (V) — PERIODIC SOBOLEV SPACES

- Let $H_p^r(I)$ be a **periodic Sobolev space**, i.e.,

$$H_p^r(I) = \{u : u^{(\alpha)} \in L_2(I), \alpha = 0, \dots, r\}$$

The space $C_p^\infty(I)$ is dense in $H_p^r(I)$

- The following two norms can be shown to be **equivalent** in H_p^r :

$$\|u\|_r = \left[\sum_{k \in \mathbb{Z}} (1+k^2)^r |\hat{u}_k|^2 \right]^{1/2}$$

$$\|u\|_r = \left[\sum_{\alpha=0}^r C_r^\alpha \|u^{(\alpha)}\|^2 \right]^{1/2}$$

Note that the first definition is naturally generalized for the case when r is non-integer!

- The projection operator P_N commutes with the derivative in the distribution sense:

$$(P_N u)^{(\alpha)} = \sum_{|k| \leq N} (ik)^\alpha \hat{u}_k W_k = P_N u^{(\alpha)}$$

FOURIER SERIES (VI) — APPROXIMATION ERROR ESTIMATES IN $H_p^s(I)$

- Let $r, s \in \mathbb{R}$ with $0 \leq s \leq r$; then we have:

$$\|u - P_N u\|_s \leq (1 + N^2)^{\frac{s-r}{2}} \|u\|_r, \text{ for } u \in H_p^r(I)$$

Proof:

$$\begin{aligned} \|u - P_N u\|_s^2 &= \sum_{|k| > N} (1 + k^2)^{s-r+r} |\hat{u}_k|^2 \leq (1 + N^2)^{s-r} \sum_{|k| > N} (1 + k^2)^r |\hat{u}_k|^2 \\ &\leq (1 + N^2)^{s-r} \|u\|_r^2 \end{aligned}$$

- Thus, accuracy of the approximation $P_N u$ is better when u is smoother; More precisely, for $u \in H_p^r(I)$ the L_2 leading order error is $O(N^{-r})$ which improves when r increases.

FOURIER SERIES (VII) — APPROXIMATION ERROR ESTIMATES IN $L_\infty(I)$

- First, a useful lemma (**Sobolev inequality**) —let $u \in H_p^1(I)$, then there exists a constant C such that

$$\|u\|_{L_\infty(I)}^2 \leq C \|u\|_0 \|u\|_1$$

Proof: Suppose $u \in C_p^\infty(I)$; note the following facts

- \hat{u} is the average of u
- From the mean value theorem: $\exists x_0 \in I$ such that $\hat{u} = u(x_0)$

Let $v(x) = u(x) - \hat{u}$, then

$$\frac{1}{2} |v(x)|^2 = \int_{x_0}^x v(y) v'(y) dy \leq \left(\int_{x_0}^x |v(y)|^2 dy \right)^{1/2} \left(\int_{x_0}^x |v'(y)|^2 dy \right)^{1/2} \leq 2\pi \|v\| \|v'\|$$

$$|u(x)| \leq |\hat{u}_0| + |v(x)| \leq |\hat{u}_0| + 2\pi^{1/2} \|v\|^{1/2} \|v'\|^{1/2} \leq C \|u\|_0^{1/2} \|u\|_1^{1/2},$$

since $v' = u'$, $\|v\| \leq \|u\|$ and $|\hat{u}| \leq \|u\|$.

As $C_p^\infty(I)$ is dense in $H_p^1(I)$, the inequality also holds for $u \in H_p^1(I)$.

FOURIER SERIES (VIII) — APPROXIMATION ERROR ESTIMATES IN $L_\infty(I)$

- An estimate in the norm $L_\infty(I)$ follows immediately from the previous lemma and estimates in the $H_p^s(I)$ norm

$$\|u - P_N u\|_{L_\infty(I)}^2 \leq C(1 + N^2)^{-\frac{r}{2}} (1 + N^2)^{\frac{1-r}{2}},$$

where $u \in H_p^r(I)$

- Thus for $r \geq 1$

$$\|u - P_N u\|_{L_\infty(I)}^2 = O(N^{\frac{1}{2}-r})$$

- Uniform convergence** for all $u \in H_p^1(I)$
(Note that u need only to be **continuous**, therefore this result is stronger than the one given on page 67)

LAGRANGE INTERPOLATION (I)

- In practice, for any arbitrary $u \in C_p^0(I)$ it is not possible to calculate **exactly** the Fourier coefficients \hat{u} (need to evaluate quadratures numerically); therefore, in general we do not know $P_N u$, i.e., the optimal projection on $S_N = \text{span}\{e^{i0k}, \dots, e^{iNk}\}$
- Can determine an **interpolant** $v \in S_N$ of u , such that v **coincides** with u at $2N + 1$ points $\{x_j\}_{|j| \leq N}$ defined by

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

- For the interpolant we set

$$v(x) = \sum_{|k| \leq N} a_k e^{ikx}$$

where the coefficients a_k can be determined by solving the algebraic system (cf. page 51)

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

LAGRANGE INTERPOLATION (II)

- The system can be rewritten as

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

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

- 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

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

- 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

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

where $M = 2N+1$, $j' = j$ if $0 \leq j \leq N$ and $j' = j+M$ if $-N \leq j < 0$, so that $\omega^{j+M} = \omega^j$. Using the expression for the sum of a finite geometric series completes the proof: $(1-\omega) \sum_{|j| \leq N}^{M-1} \omega^j = 1 - \omega^M = 0$

LAGRANGE INTERPOLATION (III)

- Consequently, the Fourier coefficients of the **interpolant** of u in S_N can be calculated as follows:

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

- The mapping

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

is referred to as **Discrete Fourier Transform (DFT)**

- Straightforward evaluation of the expression for a_k (matrix-vector product) would result in the computational cost $O(N^2)$. Algorithms known as **Fast Fourier Transforms (FFT)** reduce this cost down to $O(N \log(N))$ via a suitable factorization of the matrix \mathbb{W}^T . See www.fftw.org for one of the best publicly available implementations of the FFT.

LAGRANGE INTERPOLATION (IV)

- Let $P_C : C_p^0(I) \rightarrow S_N$ be the mapping which associates with u its interpolant $v \in S_N$. Let $(\cdot, \cdot)_N$ be the following form on $C_p^0(I)$:

$$(u, v)_N \triangleq \frac{1}{2N+1} \sum_{|j| \leq N} u(x_j) \overline{v(x_j)}$$

- By construction, the operator P_C satisfies:

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

and therefore also

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

- 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 can be obtained by replacing the scalar product (\cdot, \cdot) with the “discrete scalar product” $(\cdot, \cdot)_N$

LAGRANGE INTERPOLATION (V)

- The two scalar products coincide on S_N

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

- Proof—examine the numerical integration formula

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx \cong \frac{1}{2N+1} \sum_{|j| \leq N} f(x_j)$$

for $f \in S_N$

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

Thus for uniform distribution of x_j , the trapezoidal formula is **exact** for $u \in S_N$

LAGRANGE INTERPOLATION (VI)

- Relation between Fourier coefficients of a function and Fourier coefficients of its interpolant ($W_k(x) = e^{ikx}$)

$$\hat{u}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} u \overline{W_k} dx$$

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

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

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

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

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

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

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

LAGRANGE INTERPOLATION (VII)

- Thus

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

- A Very Important Corollary concerning **Discretization** — two trigonometric functions with different frequencies, $e^{ik_1 x}$ and $e^{ik_2 x}$, are equal on collocation points x_j , $j \leq N$ when $k_2 - k_1 = l(2N+1)$, $l = 0, \pm 1, \dots$. Therefore, the same set of values at the collocation points may represent $e^{ik_1 x}$ as well as $e^{ik_2 x}$. 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 .

LAGRANGE INTERPOLATION (VIII) — 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

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

- The term $\|u - P_N u\|_s$ is upper-bounded using theorem from page 70
- Need to estimate $\|P_C w\|_s$ —straightforward, but tedious ...

LAGRANGE INTERPOLATION (IX)

- 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:

$$\tilde{x} = j\tilde{h}, \quad -N+1 \leq j \leq N$$

$$\tilde{h} = \frac{\pi}{N}$$

- 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 $M = 2N$