Math 6630: Numerical Solutions of Partial Differential Equations
Fourier spectral methods, I

See Canuto et al. 2011, Chapter 3,

Shen, Tang, and Wang 2011, Chapter 2

Akil Narayan[1]

[1] Department of Mathematics, and Scientific Computing and Imaging (SCI) Institute
University of Utah

March 22, 2023

Weighted residual methods

We can now discuss some basic approaches for constructing a particular class of weighted residual methods: spectral methods.

Recall: with $\mathcal{L}$ a (time-independent, stationary) linear operator and f a given function, we seek to compute the solution u to

$$\mathcal{L}(u) = f, \qquad \mathcal{R}(u) := \mathcal{L}(u) - f. \qquad + \text{BC's}$$

Strong notions of convergence are too restrictive, so we employ weak notions:

$$\langle \mathcal{R}(u), v \rangle = 0, \quad \forall\, v \in V,$$

where $U \ni u$ is a chosen trial space, and $V \ni v$ is a chosen test space.

These are generic weighted residual methods.

The inner product $\langle \cdot, \cdot \rangle$ corresponds to a particular Hilbert space (for us, almost always L^2).

Weighted residual methods

We can now discuss some basic approaches for constructing a particular class of weighted residual methods: spectral methods.

Recall: with $\mathcal{L}$ a (time-independent, stationary) linear operator and f a given function, we seek to compute the solution u to

$$\mathcal{L}(u) = f, \qquad \mathcal{R}(u) := \mathcal{L}(u) - f. \qquad + \text{BC's}$$

Strong notions of convergence are too restrictive, so we employ weak notions:

$$\langle \mathcal{R}(u), v \rangle = 0, \quad \forall\, v \in V,$$

where $U \ni u$ is a chosen trial space, and $V \ni v$ is a chosen test space.

These are generic weighted residual methods.

The inner product $\langle \cdot, \cdot \rangle$ corresponds to a particular Hilbert space (for us, almost always L^2).

We'll discuss methods where we choose U and/or V as a space of Fourier functions.

This will be a particular instance of a spectral method.

The Fourier-Galerkin method

The Fourier-Galerkin method:

- Is a Galerkin procedure: $U = V$
- Uses (global) Fourier functions: $V = \mathrm{span}\{e^{ikx} \mid k \in \mathbb{Z}\}$ (assuming one spatial dimension on $[0, 2\pi]$)

The Fourier-Galerkin method

The Fourier-Galerkin method:

- Is a Galerkin procedure: $U = V$
- Uses (global) Fourier functions: $V = \operatorname{span}\{e^{ikx} \mid k \in \mathbb{Z}\}$ (assuming one spatial dimension on $[0, 2\pi]$)

Of course, this choice only makes sense if the boundary conditions are periodic.

Recall that a proper strategy to address this problem is to identify the appropriate bilinear form a:

$$\mathcal{L}(u) = f \quad \xrightarrow{\text{IbP}} \quad a(u, v) = \langle f, v \rangle .$$

The Fourier-Galerkin method

The Fourier-Galerkin method:

- Is a Galerkin procedure: $U = V$
- Uses (global) Fourier functions: $V = \operatorname{span}\{e^{ikx} \mid k \in \mathbb{Z}\}$ (assuming one spatial dimension on $[0, 2\pi]$)

Of course, this choice only makes sense if the boundary conditions are periodic.

Recall that a proper strategy to address this problem is to identify the appropriate bilinear form a:

$$\mathcal{L}(u) = f \quad \xrightarrow{\text{IbP}} \quad a(u, v) = \langle f, v \rangle .$$

One can then choose finite-dimensional spaces U_N and V_N for discretization. I.e.,:

$$\text{Find } u_N \in U_N \text{ satisfying } a(u_N, v) = \langle f, v \rangle, \quad \text{for all } v \in V.$$

An elliptic equation

Let's construct a Fourier-Galerkin scheme for the following problem:

$$-u''(x) + u(x) = f(x), \qquad \text{periodic BC's}$$

Since this is a Galerkin scheme, $U = V$, and so we need only identify V. With periodic boundary conditions, it's sensible to choose:

$$u, v \in \operatorname{span}\{\phi_k(x)\}_{k\in\mathbb{Z}}, \qquad \phi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}.$$

Using an L^2 inner product on $[0, 2\pi]$, then

$$\left\langle -u'' + u, v \right\rangle \xrightarrow{\text{IbP}} \left\langle u', v' \right\rangle + \left\langle u, v \right\rangle.$$

An elliptic equation

Let's construct a Fourier-Galerkin scheme for the following problem:

$$-u''(x) + u(x) = f(x), \qquad \text{periodic BC's}$$

Since this is a Galerkin scheme, $U = V$, and so we need only identify V. With periodic boundary conditions, it's sensible to choose:

$$u, v \in \operatorname{span}\{\phi_k(x)\}_{k\in\mathbb{Z}}, \qquad \phi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}.$$

Using an L^2 inner product on $[0, 2\pi]$, then

$$\left\langle -u'' + u, v \right\rangle \xrightarrow{\text{IbP}} \left\langle u', v' \right\rangle + \left\langle u, v \right\rangle.$$

This immediately suggests an appropriate sesquilinear form:

$$a(u, v) := \left\langle u', v' \right\rangle + \left\langle u, v \right\rangle,$$

$|a(v,v)| \geq \|v\|_V^2$

$|a(u,v)| \leq \sqrt{2}\,\|u\|_V \|v\|_V$

which we know is continuous and coercive with respect to H^1_p.

Thus, we choose $V = H^1_p$, and the dual of V with respect to L^2 is $V^* = H^{-1}_p$.

Thus (V, L^2, V^*) is our Gelfand triple, and for any $f \in V^*$,

$$\text{Find } u \in V \text{ satisfying } a(u, v) = \left\langle f, v \right\rangle \text{ for all } v \in V$$

is well-posed by Lax-Milgram.

A Fourier-Galerkin scheme

$$\text{Find } u \in V \text{ satisfying } a(u,v) = \langle f, v \rangle \text{ for all } v \in V$$

We computed last time:

$$\sup_{u,v \in V} \frac{|a(u,v)|}{\|u\|_V \|v\|_V} \leqslant \sqrt{2} =: C, \qquad \inf_{v \in V} \frac{|a(v,v)|}{\|v\|_V^2} = 1 =: c.$$

which will be useful when computing the error in our scheme.

A Fourier-Galerkin scheme

$$\text{Find } u \in V \text{ satisfying } a(u,v) = \langle f, v\rangle \text{ for all } v \in V$$

We computed last time:

$$\sup_{u,v\in V} \frac{|a(u,v)|}{\|u\|_V \|v\|_V} \leqslant \sqrt{2} =: C, \qquad \inf_{v\in V} \frac{|a(v,v)|}{\|v\|_V^2} = 1 =: c.$$

which will be useful when computing the error in our scheme.

To turn all this into a scheme, we need to discretize. Let's choose,

$$V_N = \operatorname{span}\left\{\phi_k, \;\middle|\; |k| \leqslant N\right\} \subset V. \qquad \dim V_N = 2N+1$$

Then our discrete scheme is

$$\text{Find } u_N \in V_N \text{ satisfying } a(u_N, v) = \langle f, v\rangle \text{ for all } v \in V_N$$

A Fourier-Galerkin scheme

$$\text{Find } u \in V \text{ satisfying } a(u,v) = \langle f, v\rangle \text{ for all } v \in V$$

We computed last time:

$$\sup_{u,v\in V} \frac{|a(u,v)|}{\|u\|_V \|v\|_V} \leqslant \sqrt{2} =: C, \qquad \inf_{v\in V} \frac{|a(v,v)|}{\|v\|_V^2} = 1 =: c.$$

which will be useful when computing the error in our scheme.

To turn all this into a scheme, we need to discretize. Let's choose,

$$V_N = \operatorname{span}\left\{\phi_k, \;\middle|\; |k| \leqslant N\right\} \subset V.$$

Then our discrete scheme is

$$\text{Find } u_N \in V_N \text{ satisfying } a(u_N,v) = \langle f, v\rangle \text{ for all } v \in V_N$$

The above problem is equivalent to making the ansatz,

$$u_N(x) = \sum_{|k|\leqslant N} \widehat{u}_k \phi_k(x),$$

and using $v \leftarrow \phi_k$ for every k satisfying $|k| \leqslant N$.

Scheme details

$$\text{Find } u_N \in V_N \text{ satisfying } a(u_N, v) = \langle f, v \rangle \text{ for all } v \in V_N$$

$$u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k \phi_k(x),$$

With $v = \phi_k$ for a fixed k, the weak form reads,

$$\left\langle \sum_{|\ell| \leqslant N} (i\ell) \widehat{u}_\ell \phi_\ell, (ik)\phi_k \right\rangle + \langle u_N, \phi_k \rangle = \langle f, \phi_k \rangle .$$

Scheme details

$$\text{Find } u_N \in V_N \text{ satisfying } a(u_N, v) = \langle f, v \rangle \text{ for all } v \in V_N$$

$$u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k \phi_k(x),$$

With $v = \phi_k$ for a fixed k, the weak form reads,

$$\left\langle \sum_{|\ell| \leqslant N} (i\ell) \widehat{u}_\ell \phi_\ell, (ik)\phi_k \right\rangle + \langle u_N, \phi_k \rangle = \langle f, \phi_k \rangle .$$

Then defining,

$$\widehat{f}_k = \langle f, \phi_k \rangle, \qquad\qquad d_k = k^2 + 1,$$

our scheme is,

$$\widehat{\boldsymbol{D}}_2 \widehat{\boldsymbol{u}} = \widehat{\boldsymbol{f}}, \quad \widehat{\boldsymbol{D}}_2 = \operatorname{diag}\left((-N)^2 + 1, \ (-N+1)^2 + 1, \ \cdots \ (N-1)^2 + 1, \ (N^2) + 1 \right),$$

and so $\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{D}}^{-1} \widehat{\boldsymbol{f}}$ is the solution, prescribing u_N.

$\widehat{D}_2$ is sometimes called a "modal" differentiation matrix.

In practice, we cannot compute $\widehat{f}_k$ exactly, so these coefficients are approximated with quadrature/interpolation (injecting aliasing error into the right-hand side).

Error estimates

One of the significant advantages of all this is that we have already laid all the groundwork we need to compute a direct error estimate.

First, we have that,

$$\|u - u_N\|_V \leqslant \frac{C}{c} \inf_{v \in V_N} \|u - v\|_V, \qquad \frac{C}{c} = \sqrt{2},$$

due to Céa's Lemma, which in turn relies on Lax-Milgram.

Error estimates

One of the significant advantages of all this is that we have already laid all the groundwork we need to compute a direct error estimate.

First, we have that,

$$\|u - u_N\|_V \leqslant \frac{C}{c} \inf_{v \in V_N} \|u - v\|_V, \qquad \frac{C}{c} = \sqrt{2},$$

due to Céa's Lemma, which in turn relies on Lax-Milgram.

We also have,

$$\inf_{v \in v_N} \|u - v\|_V \leqslant \|u - P_N u\|_V,$$

where P_N is the L^2-orthogonal projection onto $W = V_N$.

Error estimates

One of the significant advantages of all this is that we have already laid all the groundwork we need to compute a direct error estimate.

First, we have that,

$$\|u - u_N\|_V \leqslant \frac{C}{c} \inf_{v \in V_N} \|u - v\|_V, \qquad \frac{C}{c} = \sqrt{2},$$

due to Céa's Lemma, which in turn relies on Lax-Milgram.

We also have,

$$\inf_{v \in v_N} \|u - v\|_V \leqslant \|u - P_N u\|_V,$$

where P_N is the L^2-orthogonal projection onto $V = H^1_p$.

Finally, we have our basic approximation estimate for Fourier series:

$$\|u - P_N u\|_{H^1} \leqslant N^{1-s}. \qquad u \in H^s_p \Rightarrow \|u - P_N u\|_{H^r_p} \leqslant N^{r-s}$$

Error estimates

One of the significant advantages of all this is that we have already laid all the groundwork we need to compute a direct error estimate.

First, we have that,

$$\|u - u_N\|_V \leqslant \frac{C}{c} \inf_{v \in V_N} \|u - v\|_V, \qquad \frac{C}{c} = \sqrt{2},$$

due to Céa's Lemma, which in turn relies on Lax-Milgram.

We also have,

$$\inf_{v \in v_N} \|u - v\|_V \leqslant \|u - P_N u\|_V,$$

where P_N is the L^2-orthogonal projection onto $V = H_p^1$.

Finally, we have our basic approximation estimate for Fourier series:

$$\|u - P_N u\|_{H^1} \leqslant N^{1-s}.$$

One final piece of information: while Lax-Milgram establishes $f \in H_p^{-1} \Rightarrow u \in H_p^1$, one can also show that $f \in H_p^r \Rightarrow u \in H_p^{r+2}$.

Error estimates

One of the significant advantages of all this is that we have already laid all the groundwork we need to compute a direct error estimate.

First, we have that,

$$\|u - u_N\|_V \leqslant \frac{C}{c} \inf_{v \in V_N} \|u - v\|_V, \qquad \frac{C}{c} = \sqrt{2},$$

due to Céa's Lemma, which in turn relies on Lax-Milgram.

We also have,

$$\inf_{v \in v_N} \|u - v\|_V \leqslant \|u - P_N u\|_V,$$

where P_N is the L^2-orthogonal projection onto $V = H_p^1$.

Finally, we have our basic approximation estimate for Fourier series:

$$\|u - P_N u\|_{H^1} \leqslant N^{1-s}.$$

One final piece of information: while Lax-Milgram establishes $f \in H_p^{-1} \Rightarrow u \in H_p^1$, one can also show that $f \in H_p^r \Rightarrow u \in H_p^{r+2}$.

Thus, overall, we have that for any $r > -1$,

$$f \in H_p^r \quad \Longrightarrow \quad u \in H_p^{r+2} \quad \Longrightarrow \quad \|u - u_N\|_{H^1} \leqslant \sqrt{2} N^{1-(r+2)} \lesssim N^{-(r+1)}.$$

This convergence rate is optimal (since it is the optimal rate of approximation).

A "strong" form Fourier-Galerkin method

Sometimes in practice a "strong" version of the Fourier-Galerkin method will be implemented. This approach directly imposes the Galerkin condition on the strong PDE form:

$$\left\langle -u_N'' + u, v \right\rangle = \langle f, v \rangle, \qquad v \in V.$$

A "strong" form Fourier-Galerkin method

Sometimes in practice a "strong" version of the Fourier-Galerkin method will be implemented. This approach directly imposes the Galerkin condition on the strong PDE form:

$$\left\langle -u_N'' + u, v \right\rangle = \langle f, v \rangle, \qquad v \in V.$$

Note that this is not unreasonable since if u_N is smooth enough, we have,

$$\left\langle -u_N'', v \right\rangle = \left\langle u_N', v' \right\rangle,$$

which implies that the strong form above is exactly equivalent to our first weak formulation that employed bilinear forms.

A "strong" form Fourier-Galerkin method

Sometimes in practice a "strong" version of the Fourier-Galerkin method will be implemented. This approach directly imposes the Galerkin condition on the strong PDE form:

$$\left\langle -u_N'' + u, v \right\rangle = \langle f, v \rangle, \qquad v \in V.$$

Note that this is not unreasonable since if u_N is smooth enough, we have,

$$\left\langle -u_N'', v \right\rangle = \left\langle u_N', v' \right\rangle,$$

which implies that the strong form above is exactly equivalent to our first weak formulation that employed bilinear forms.

On an implementation level, the strong form can be implemented by taking $v = \phi_k$ for every $|k| \leqslant N$. For a fixed k, this reads,

$$\left\langle \sum_{|\ell| \leqslant N} -(i\ell)^2 \widehat{u}_\ell \phi_\ell, \phi_k \right\rangle + \langle u_N, \phi_k \rangle = \langle f, \phi_k \rangle.$$

This is *exactly*,

$$\widehat{\boldsymbol{D}}_2 \widehat{\boldsymbol{u}} = \widehat{\boldsymbol{f}},$$

that is equivalent to the previous Fourier-Galerkin scheme we derived.

Caution: It is not always the case that the strong variational form and weak variational form coincide.

Beyond Fourier-Galerkin

It's worthwhile to mention how this error analysis might be generalized to another problem or setup:

- The crucial components were Céa's Lemma and fundamental approximation estimates
- There is little theoretical difference if the spatial domain is multidimensional, or if the basis basis functions are different.
- For variable coefficient problems, e.g.,

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left((2+\sin x)\frac{\mathrm{d}}{\mathrm{d}x}u(x)\right)+u(x)=f(x),$$

  the main difficulty is computational: the term $(2+\sin x)$ results in the need to compute more complicted inner products, e.g.,

$$\left\langle (2+\sin x)\phi_k'(x), \phi_\ell'(x)\right\rangle.$$

  This is a computational issue, but does not adversely affect theory.
- For non-periodic problems: the theoretical strategies are quite similar, and much of the marginal extra work involves fundamental approximation estimates for non-periodic basis functions
- Time-dependent problems require somewhat different theoretical strategies, but *stability* is a big piece of the puzzle.

Fourier-Collocation

The Fourier collocation strategy uses a very similar approach to the Galerkin case.

$$-u''(x) + u(x) = f(x), \qquad \text{periodic BC's}$$

As with the Galerkin method, our trial space U is the *Fourier* space:

$$u \in \operatorname{span}\{\phi_k(x)\}_{k\in\mathbb{Z}}, \qquad \phi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}.$$

However, our test space V is formally defined by Dirac functions centered at the grid points:

$$V = \operatorname{span}\left\{\delta_{x_m}\right\}_{m\in[2N+1]}, \qquad x_m = \frac{2\pi(m-1)}{2N+1}.$$

Fourier-Collocation

The Fourier collocation strategy uses a very similar approach to the Galerkin case.

$$-u''(x) + u(x) = f(x), \qquad \text{periodic BC's}$$

As with the Galerkin method, our trial space U is the *Fourier* space:

$$u \in \operatorname{span}\{\phi_k(x)\}_{k\in\mathbb{Z}}, \qquad \phi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}.$$

However, our test space V is formally defined by Dirac functions centered at the grid points:

$$V = \operatorname{span}\left\{\delta_{x_m}\right\}_{m\in[2N+1]}, \qquad x_m = \frac{2\pi(m-1)}{2N+1}.$$

I.e., our weighted residual method in the collocation setup is quite simple:

$$-u_N''(x_m) + u_N(x_m) = f(x_m), \qquad m \in [2N+1].$$

Equivalently, we enforce the condition,

$$I_N \mathcal{R}(u_N) = 0,$$

which is an exact equality since zero residual on interpolation nodes implies that the interpolant is 0.

This method is quite informal as we simply enforce pointwise zero residual.

Implementation

Recall some Fourier notation:

$$u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k \phi_k(x), \qquad \widehat{\boldsymbol{u}} = \begin{pmatrix} \widehat{u}_{-N} \\ \vdots \\ \widehat{u}_N \end{pmatrix}, \qquad \boldsymbol{u} = \begin{pmatrix} u(x_1) \\ \vdots \\ u(x_M) \end{pmatrix}.$$

$$\widetilde{\boldsymbol{u}} = \widetilde{\boldsymbol{V}}^* \boldsymbol{u}, \qquad \boldsymbol{u} = \frac{M}{2\pi} \widetilde{\boldsymbol{V}} \widetilde{\boldsymbol{u}}.$$

There are two computational options for implementing collocation:

- Store/solve for $\widehat{\boldsymbol{u}}$ as degrees of freedom
- Store/solve for $\boldsymbol{u}$ as degrees of freedom.

The latter, being more consistent with the idea of collocation, is more standard.

Implementation

Recall some Fourier notation:

$$u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k \phi_k(x), \qquad \widehat{\boldsymbol{u}} = \begin{pmatrix} \widehat{u}_{-N} \\ \vdots \\ \widehat{u}_N \end{pmatrix}, \qquad \boldsymbol{u} = \begin{pmatrix} u(x_1) \\ \vdots \\ u(x_M) \end{pmatrix}.$$

$$\widetilde{\boldsymbol{u}} = \widetilde{\boldsymbol{V}}^* \boldsymbol{u}, \qquad \boldsymbol{u} = \frac{M}{2\pi} \widetilde{\boldsymbol{V}} \widetilde{\boldsymbol{u}}.$$

There are two computational options for implementing collocation:

- Store/solve for $\widehat{\boldsymbol{u}}$ as degrees of freedom
- Store/solve for $\boldsymbol{u}$ as degrees of freedom.

The latter, being more consistent with the idea of collocation, is more standard.

To implement such a scheme, the difficult part is really to compute $u_N''(x_m)$. I.e., we must differentiate twice and evaluate at x_m. This can be accomplished by:

1. Transform $\boldsymbol{u}$ to $\widehat{\boldsymbol{u}}$. (Apply $\widetilde{\boldsymbol{V}}^*$.)
2. Twice differentiate $u_N = \sum \widehat{u}_k \phi_k$. (Apply $\operatorname{diag}((-N)^2, \cdots, N^2)$.)
3. Transform $\widehat{\boldsymbol{u}}$ back to $\boldsymbol{u}$. (Apply $\frac{M}{2\pi} \widetilde{\boldsymbol{V}}$.)

Differentiation matrices

This allows us to create a *nodal* equivalent of the continuous operation,

$$u_N(x) \mapsto u_N''(x)$$

This operation, with nodal degrees of freedom, is equivalent to applying the matrix,

$$\widetilde{\boldsymbol{D}}_2 := \frac{M}{2\pi} \widetilde{\boldsymbol{V}} \begin{pmatrix} (-N)^2 & & & \\ & (-N+1)^2 & & \\ & & \ddots & \\ & & & (N)^2 \end{pmatrix} \widetilde{\boldsymbol{V}}^* .$$

Differentiation matrices

This allows us to create a *nodal* equivalent of the continuous operation,

$$u_N(x) \mapsto u_N''(x)$$

This operation, with nodal degrees of freedom, is equivalent to applying the matrix,

$$\widetilde{\boldsymbol{D}}_2 := \frac{M}{2\pi} \widetilde{\boldsymbol{V}} \begin{pmatrix} (-N)^2 & & & \\ & (-N+1)^2 & & \\ & & \ddots & \\ & & & (N)^2 \end{pmatrix} \widetilde{\boldsymbol{V}}^*.$$

This matrix is dense and is equivalent to a finite difference matrix whose stencil spans the entire grid. (Except we approximate with Fourier Series and not polynomials.)

With some abuse of notation, let $I_N \boldsymbol{u}$ denote the interpolant (an element of V_N) corresponding to the unique element whose grid values are $\boldsymbol{u}$. Then,

$$I_N \widetilde{\boldsymbol{D}}_2 \boldsymbol{u} = (I_N \boldsymbol{u})'',$$

i.e., $\widetilde{\boldsymbol{D}}_2$ accomplishes *exact* second differentiation for elements from V_N.

Fourier collocation scheme

The Fourier collocation scheme associated to

$$-u''(x) + u(x) = f(x), \qquad \text{periodic BC's},$$

then reads,

$$\left(-\widetilde{\boldsymbol{D}}_2 + \boldsymbol{I}\right)\boldsymbol{u} = \boldsymbol{f}, \qquad \boldsymbol{f} = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_M) \end{pmatrix},$$

which is a linear system that can be solved for $\boldsymbol{u}$, which uniquely identifies $I_N \boldsymbol{u} \in V_N$.

Note here that $\left(-\widetilde{\boldsymbol{D}}_2 + \boldsymbol{I}\right)$ is a dense matrix, so some advantages from the Fourier-Galerkin setting are missing.

Fourier collocation scheme

The Fourier collocation scheme associated to

$$-u''(x) + u(x) = f(x), \qquad \text{periodic BC's},$$

then reads,

$$\left(-\widetilde{\boldsymbol{D}}_2 + \boldsymbol{I}\right)\boldsymbol{u} = \boldsymbol{f}, \qquad \boldsymbol{f} = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_M) \end{pmatrix},$$

which is a linear system that can be solved for $\boldsymbol{u}$, which uniquely identifies $I_N \boldsymbol{u} \in V_N$.

Note here that $\left(-\widetilde{\boldsymbol{D}}_2 + \boldsymbol{I}\right)$ is a dense matrix, so some advantages from the Fourier-Galerkin setting are missing.

What about error estimates? For this particularly simple equation and for *smooth* solutions u_N[1], note that the Fourier-Galerkin method for $u_{N,G}$ and the Fourier-Collocation method for $u_{N,C}$ are given by, respectively, is given by:

$$P_N\left(-u''_{N,G} + u_{N,G}\right) = P_N f, \qquad I_N\left(-u''_{N,C} + u_{N,C}\right) = I_N f,$$

where P_N is the L^2-orthogonal projector onto V_N.

[1] I.e., u_N smooth enough so that $\langle -u''_N, v\rangle = \langle u'_N, v'\rangle$, which is always true if $u_N, v \in V_N$.

Collocation error estimates, I

$$P_N\left(-u''_{N,G} + u_{N,G}\right) = P_N f, \qquad I_N\left(-u''_{N,C} + u_{N,C}\right) = I_N f,$$

Since $-u''_N + u \in V_N$, then

$$I_N\left(-u''_{N,C} + u_{N,C}\right) = P_N\left(-u''_{N,C} + u_{N,C}\right),$$

and therefore the difference $\Delta u_N := u_{N,G} - u_{N,C}$ satisfies,

$$P_N\left(-\Delta u''_N + \Delta u_N\right) = A_N f, \quad = P_N A_N f$$

where $A_N f = P_N f - I_N f$ is the aliasing error for f.

Collocation error estimates, I

$$P_N\left(-u''_{N,G} + u_{N,G}\right) = P_N f, \qquad I_N\left(-u''_{N,C} + u_{N,C}\right) = I_N f,$$

Since $-u''_N + u \in V_N$, then

$$I_N\left(-u''_{N,C} + u_{N,C}\right) = P_N\left(-u''_{N,C} + u_{N,C}\right),$$

and therefore the difference $\Delta u_N := u_{N,G} - u_{N,C}$ satisfies,

$$P_N\left(-\Delta u''_N + \Delta u_N\right) = A_N f,$$

where $A_N f = P_N f - I_N f$ is the aliasing error for f.

Hence, the error between $u_{N,C}$ and $u_{N,G}$ is given by the solution to $-u'' + u = g$, where g is the aliasing error.

At the discrete level, this is,

$$\left(\widehat{\boldsymbol{D}}_2 + \boldsymbol{I}\right)\widehat{\Delta u_N} = \left(\widehat{\boldsymbol{f}} - \widetilde{\boldsymbol{f}}\right),$$

where $\widetilde{\boldsymbol{f}}$ are the expansion coefficients for f computed using interpolation I_N.

Collocation error estimates, II

Therefore, the norm of Δu_N satisfies,

$$\begin{aligned}
\|\Delta u_N\|_2 &\leqslant \left\| \left(\widehat{\boldsymbol{D}}_2 + \boldsymbol{I} \right)^{-1} \right\|_2 \|\widehat{\boldsymbol{f}} - \widetilde{\boldsymbol{f}}\|_2 \\
&= \left\| \left(\widehat{\boldsymbol{D}}_2 + \boldsymbol{I} \right)^{-1} \right\|_2 \|A_N f\|_{L^2} \\
&\leqslant \|A_N f\|_{L^2} \lesssim N^{-r}.
\end{aligned}$$

where we have assumed $f \in H^r_p$.

Collocation error estimates, II

Therefore, the norm of Δu_N satisfies,

$$\begin{aligned}\|\Delta u_N\|_2 &\leqslant \left\|\left(\widehat{\boldsymbol{D}}_2 + \boldsymbol{I}\right)^{-1}\right\|_2 \|\widehat{\boldsymbol{f}} - \widetilde{\boldsymbol{f}}\|_2 \\ &= \left\|\left(\widehat{\boldsymbol{D}}_2 + \boldsymbol{I}\right)^{-1}\right\|_2 \|A_N f\|_{L^2} \\ &\leqslant \|A_N f\|_{L^2} \lesssim N^{-r}.\end{aligned}$$

where we have assumed $f \in H^r_p$.

Chaining this with our H^1_p-norm estimate form $u - P_N u$ yields,

$$f \in H^r_p \quad \Longrightarrow \quad \|u - u_{N,C}\|_{L^2} \lesssim N^{-r},$$

where we have used the fact that $f \in H^r_p$ implies that $u \in H^{r+2}_p$.

A finer estimate can be constructed with sharper treatement of the last inequality in the estimate for $\|\Delta u_N\|_2$.
(I.e., the estimate we've derived is not optimal, but one can derive an optimal estimate.)

One can also construct an estimate in $V = H^1_p$.

Galerkin vs collocation, I

For the particular example we've considered:

In practice, both the Galerkin and collocation schemes behave quite similarly for smooth u.

Both schemes can utilize the FFT: For the Galerkin method this is straightforward. For collocation, the application $\boldsymbol{v} \mapsto \widehat{\boldsymbol{D}}_2$ can be accomplished by sandwiching application of the (diagonal) matrix $\widehat{\boldsymbol{D}}_2$ with the forward/inverse FFT.

If quadrature/interpolation is used to approximate $\widehat{\boldsymbol{f}}$ in the Galerkin scheme, then the schemes produce identical solutions.

Galerkin vs collocation, I

For the particular example we've considered:

In practice, both the Galerkin and collocation schemes behave quite similarly for smooth u.

Both schemes can utilize the FFT: For the Galerkin method this is straightforward. For collocation, the application $\boldsymbol{v} \mapsto \widehat{\boldsymbol{D}}_2$ can be accomplished by sandwiching application of the (diagonal) matrix $\widehat{\boldsymbol{D}}_2$ with the forward/inverse FFT.

If quadrature/interpolation is used to approximate $\widehat{\boldsymbol{f}}$ in the Galerkin scheme, then the schemes produce identical solutions.

The property is due to the equation we have considered. For example, if we instead consider,

$$-u''(x) + (2 + \sin x)u(x) = f(x),$$

then the Galerkin/collocation discretizations for the left hand side are *not* the same, and hence one must augment the previous analysis to estimate the discrepancy.

In general, it's easier to prove estimates for Galerkin methods.

Galerkin vs collocation, II

Collocation is more conceptually easy to implement, although constructing the differentiation matrix is somewhat nontrivial, even if one uses the FFT.

Nevertheless, an example such as,

$$-u''(x) + \frac{u(x)}{2+\sin x} = f(x),$$

shows when collocation is more straightforward than Galerkin.

$P_N\left[\frac{u_N(x)}{2+\sin x}\right]$

$P_N\left[\frac{\phi_k(x)}{2+\sin x}\right] = ?$

Consider, in particular, when there are nonlinear terms in the equation: Galerkin methods can be difficult to implement.

In constrats, the collocation discretization of the above corresponds to

$$-u_N''(x_m) + \frac{u_N(x_m)}{2+\sin x_m} = f(x_m), \qquad m \in [M],$$

which is straightforward to implement as it is linear in u_N.

Galerkin vs collocation, III

One can borrow the interpolation idea from collocation to aid in approximating Galerkin quantities. I.e., the Galerkin scheme requires us to enforce,

$$P_N\left(-u_N''(x) + \frac{u_N(x)}{2+\sin x}\right) = P_N f(x),$$

and computing $P_N(u_N/(2+\sin x))$ is the difficult part. However, we could approximate this by:

$$P_N\left(-u_N''(x)\right) + I_N\left(\frac{u_N(x)}{2+\sin x}\right) = P_N f(x),$$

which is easy to implement, as $I_N u_N/(2+\sin x)$ is computed by evaluating at the gridpoints x_m and then taking a DFT.

This type of approximation of projections (in particular of nonlinear terms) is called a pseudospectral approximation.

References I


Canuto, Claudio et al. (2011). *Spectral Methods: Fundamentals in Single Domains*. 1st ed. 2006. Corr. 4th printing 2010 edition. Berlin ; New York: Springer. ISBN: 978-3-540-30725-9.


Shen, Jie, Tao Tang, and Li-Lian Wang (2011). *Spectral Methods: Algorithms, Analysis and Applications*. Springer Science & Business Media. ISBN: 978-3-540-71041-7.