Math 6630: Numerical Solutions of Partial Differential Equations
Weighted residual methods

See Hesthaven, S. Gottlieb, and D. Gottlieb 2007, Chapters 3,

Shen, Tang, and Wang 2011, Chapter 1

Akil Narayan[1]

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

March 20, 2023

www.sci.utah.edu

Partial differential equations

Our discussion of Fourier Series suggests a natural strategy for solving PDE's: given an abstract PDE

$$\mathcal{L}(u) = f, \qquad\qquad \mathcal{R}(u) = \mathcal{L}(u) - f,$$

where f is given and we assume periodicity on the one-dimensional spatial domain $x \in [0, 2\pi]$, we'll make the ansatz,

$$u(x) \simeq u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k e^{ijx}.$$

Partial differential equations

Our discussion of Fourier Series suggests a natural strategy for solving PDE's: given an abstract PDE

$$\mathcal{L}(u) = f, \qquad\qquad \mathcal{R}(u) = \mathcal{L}(u) - f,$$

where f is given and we assume periodicity on the one-dimensional spatial domain $x \in [0, 2\pi]$, we'll make the ansatz,

$$u(x) \simeq u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k e^{ijx}.$$

So, plugging things in:

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

The function u_N has a *finite* number of degrees of freedom.

Unless we are *extraordinarily* lucky, we will never make the above statement true in, say, a pointwise sense.

Partial differential equations

Our discussion of Fourier Series suggests a natural strategy for solving PDE's: given an abstract PDE

$$\mathcal{L}(u) = f, \qquad\qquad \mathcal{R}(u) = \mathcal{L}(u) - f,$$

where f is given and we assume periodicity on the one-dimensional spatial domain $x \in [0, 2\pi]$, we'll make the ansatz,

$$u(x) \simeq u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k e^{ijx}.$$

So, plugging things in:

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

The function u_N has a *finite* number of degrees of freedom.

Unless we are *extraordinarily* lucky, we will never make the above statement true in, say, a pointwise sense.

The focus of what follows revolves around how we will enforce $\mathcal{R}(u_N) = 0$.

Strong enforcement, I

We will call $\mathcal{R}$ the PDE *residual*, and hence

$$\mathcal{R}(u) = 0$$

asks for zero residual.

To start, let's assume both $\mathcal{R}$ and u are independent of time t. (Stationary problems)

Requiring the above condition pointwise for every x is called strong enforcement of the PDE, and such a u is a strong solution.

Our one previous strategy, *finite difference* methods, asserted that the residual vanish at specified grid points.[1]

[1] Well, we approximated the residual with finite differences.

Strong enforcement, I

We will call $\mathcal{R}$ the PDE *residual*, and hence

$$\mathcal{R}(u) = 0$$

asks for zero residual.

To start, let's assume both $\mathcal{R}$ and u are independent of time t. (Stationary problems)

Requiring the above condition pointwise for every x is called strong enforcement of the PDE, and such a u is a strong solution.

Our one previous strategy, *finite difference* methods, asserted that the residual vanish at specified grid points.[1]

Hence, one strategy to move forward is strong enforcement at some selection of grid points.

While this is reasonable in many cases, there are some rather transparent situations when this enforcement is a poor choice.

[1] Well, we approximated the residual with finite differences.

Strong enforcement, II

Example

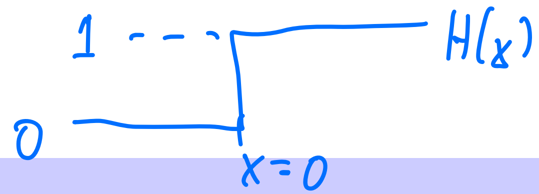
Consider the PDE

$$u_t + u_x = 0, \qquad\qquad u(x,0) = \sin x$$

The solution is $u(x,t) = \sin(x-t)$, and is valid pointwise for every (x,t).

Hence strong enforcement (everywhere) is fine here.

Strong enforcement, II



Example

Consider the PDE

$$u_t + u_x = 0, \qquad\qquad u(x,0) = H(x),$$

where $H(x)$ is the Heaviside (step) function centered at 0. The solution is $u(x,t) = H(x-t)$, and is valid pointwise for every (x,t) except where $x = t$.

Here, there is a single x (for each t) where it is not possible to enforce the PDE strongly.

But it's "just" one point for each t, so probably this is ok.

Strong enforcement, II

Example

Consider the PDE

$$u_t + u_x = 0, \qquad\qquad u(x,0) = H(x),$$

where $H(x)$ is the Heaviside (step) function centered at 0. The solution is $u(x,t) = H(x-t)$, and is valid pointwise for every (x,t) except where $x = t$.

Here, there is a single x (for each t) where it is not possible to enforce the PDE strongly.

But it's "just" one point for each t, so probably this is ok.

Example

Consider the PDE

$$u_t + u_x = 0, \qquad\qquad u(x,0) = H(x),$$

where $H(x)$ is the Heaviside (step) function centered at 0. The function $u(x,t) = H(x)$, is a strong solution pointwise for every (x,t) except where $x = 0$.

It's "just" one point for each t, so is this ok?

Strong enforcement, II

Example

Consider the PDE

$$u_t + u_x = 0, \qquad u(x,0) = H(x),$$

where $H(x)$ is the Heaviside (step) function centered at 0. The solution is $u(x,t) = H(x-t)$, and is valid pointwise for every (x,t) except where $x = t$.

Here, there is a single x (for each t) where it is not possible to enforce the PDE strongly.

But it's "just" one point for each t, so probably this is ok.

Example

Consider the PDE

$$u_t + u_x = 0, \qquad u(x,0) = H(x),$$

where $H(x)$ is the Heaviside (step) function centered at 0. The function $u(x,t) = H(x)$, is a strong solution pointwise for every (x,t) except where $x = 0$.

It's "just" one point for each t, so is this ok?
This should bother you: if you accept $u(x,t) = H(x-t)$ as a solution, you must also logically accept $u(x) = H(x)$ as a solution, and hence solutions are not unique.

Beyond strong solutions

Therefore, PDE enforcement pointwise on a grid *can* be useful, but we need an alternative strategy to weed out some undesirable behavior.

What is a reasonable alternative? A related (non-differential) problem of approximation with Fourier Series provides some motivation:

Example (Fourier approximation)

Consider $V_N := \mathrm{span}\left\{e^{ikx}, \;\middle|\; |k| \leqslant N\right\} \subset L^2([0, 2\pi]; \mathbb{C})$. Suppose we wish to construct u such that,

$$u(x) = \exp(\sin x), \qquad\qquad \mathcal{R}(u) := u(x) - \exp(\sin x)$$

You could consider $\mathcal{R}$ above our "PDE" residual.

Beyond strong solutions

Therefore, PDE enforcement pointwise on a grid *can* be useful, but we need an alternative strategy to weed out some undesirable behavior.

What is a reasonable alternative? A related (non-differential) problem of approximation with Fourier Series provides some motivation:

Example (Fourier approximation)

Consider $V_N := \operatorname{span}\left\{e^{ikx}, \;\middle|\; |k| \leqslant N\right\} \subset L^2([0, 2\pi]; \mathbb{C})$. Suppose we wish to construct u such that,

$$u(x) = \exp(\sin x), \qquad\qquad \mathcal{R}(u) := u(x) - \exp(\sin x)$$

You could consider $\mathcal{R}$ above our "PDE" residual.

If we make the ansatz $u(x) \in V_N$ via,

$$u(x) \simeq u_N(x) = \sum_{|k| \leqslant N} \widehat{u}_k \phi_k(x), \qquad\qquad \phi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx},$$

Then our "PDE" requires,

$$\mathcal{R}(u_N) = 0 \quad \Longrightarrow \quad u_N(x) = \exp(\sin x).$$

Like in the (real) PDE setting, we cannot make this true in general.

Beyond strong solutions

Therefore, PDE enforcement pointwise on a grid *can* be useful, but we need an alternative strategy to weed out some undesirable behavior.

What is a reasonable alternative? A related (non-differential) problem of approximation with Fourier Series provides some motivation:

Example

Our alternative strategy was to define u_N so that it is the (L^2) best possible approximation:

$$u_N = \underset{v \in V_N}{\arg\min} \|v(x) - \exp(\sin x)\|_{L^2} \implies \quad \widehat{u}_k = \langle u, \phi_k \rangle.$$

Beyond strong solutions

Therefore, PDE enforcement pointwise on a grid *can* be useful, but we need an alternative strategy to weed out some undesirable behavior.

What is a reasonable alternative? A related (non-differential) problem of approximation with Fourier Series provides some motivation:

Example

Our alternative strategy was to define u_N so that it is the (L^2) best possible approximation:

$$u_N = \arg\min_{v \in V_N} \|v(x) - \exp(\sin x)\| \quad \Longrightarrow \quad \widehat{u}_k = \langle u, \phi_k \rangle .$$

Here is an alternative computation in terms of $\mathcal{R}$ that accomplishes the same thing:

Instead of requiring say pointwise enforcement of $\mathcal{R}(u_N) = 0$, we require for every $|k| \leqslant N$:

$$\langle \mathcal{R}(u_N), \phi_k(x) \rangle = 0 \quad \Longrightarrow \quad \widehat{u}_k(x) = \langle \exp(\sin x), \phi_k(x) \rangle .$$

Beyond strong solutions

Therefore, PDE enforcement pointwise on a grid *can* be useful, but we need an alternative strategy to weed out some undesirable behavior.

What is a reasonable alternative? A related (non-differential) problem of approximation with Fourier Series provides some motivation:

Example

Our alternative strategy was to define u_N so that it is the (L^2) best possible approximation:

$$u_N = \underset{v \in V_N}{\arg\min} \|v(x) - \exp(\sin x)\| \quad \Longrightarrow \quad \widehat{u}_k = \langle u, \phi_k \rangle .$$

Here is an alternative computation in terms of $\mathcal{R}$ that accomplishes the same thing:

Instead of requiring say pointwise enforcement of $\mathcal{R}(u_N) = 0$, we require for every $|k| \leqslant N$:

$$\langle \mathcal{R}(u_N), \phi_k(x) \rangle = 0 \quad \Longrightarrow \quad u_k(x) = \langle \exp(\sin x), \phi_k(x) \rangle .$$

In particular, because $\{\phi_k\}_{|k| \leqslant N}$ is a basis for the subspace V_N, this is equivalent to,

$$\text{Find } u_N \in V_N \text{ such that } \langle \mathcal{R}(u_N), v \rangle = 0 \text{ for every } v \in V_N .$$

I.e., we do not enforce zero residual pointwise, but instead in some averaged sense.

It is this averaged sense that we will attempt to enforce zero PDE residuals.

Some functional analysis, I

Unlike the previous example, PDE residuals will involve derivatives, and in order to generalize our statements above, we need a little functional analysis notation. (Recall that we are starting with stationary problems.)

Let H be a Hilbert space, i.e., a Banach space with an inner product $\langle \cdot, \cdot \rangle$.

The (topological) dual H^* of H is the collection of continuous (=bounded) linear functionals from H to $\mathbb{C}$ (or $\mathbb{R}$).

An example of such a functional is $h \mapsto \langle h, h_\phi \rangle$ for some $h_\phi \in H$.

Some functional analysis, I

Unlike the previous example, PDE residuals will involve derivatives, and in order to generalize our statements above, we need a little functional analysis notation. (Recall that we are starting with stationary problems.)

Let H be a Hilbert space, i.e., a Banach space with an inner product $\langle \cdot, \cdot \rangle$.

The (topological) dual H^* of H is the collection of continuous (=bounded) linear functionals from H to $\mathbb{C}$ (or $\mathbb{R}$).

An example of such a functional is $h \mapsto \left\langle h, h_\phi \right\rangle$ for some $h_\phi \in H$.

The Riesz Representation Theorem essentially implies that this example is generic, i.e., $H = H^*$.
In particular, for any $\phi \in H^*$, there exists an $h_\phi \in H$ such that,

$$\phi(h) = \left\langle h, h_\phi \right\rangle,$$

and vice versa.

Some functional analysis, I

Unlike the previous example, PDE residuals will involve derivatives, and in order to generalize our statements above, we need a little functional analysis notation. (Recall that we are starting with stationary problems.)

Let H be a Hilbert space, i.e., a Banach space with an inner product $\langle \cdot, \cdot \rangle$.

The (topological) dual H^* of H is the collection of continuous (=bounded) linear functionals from H to $\mathbb{C}$ (or $\mathbb{R}$).

An example of such a functional is $h \mapsto \langle h, h_\phi \rangle$ for some $h_\phi \in H$.

The Riesz Representation Theorem essentially implies that this example is generic, i.e., $H = H^*$.
In particular, for any $\phi \in H^*$, there exists an $h_\phi \in H$ such that,

$$\phi(h) = \langle h, h_\phi \rangle,$$

and vice versa.

Now let V be a subspace of H: $V \subseteq H$. In practice, V will contain elements of H with extra smoothness conditions.

$\tilde{w} \in H^* \Rightarrow v \mapsto \langle v, \tilde{w} \rangle$ bdd $\forall\, v \in H$

The dual V^* of V with respect to the inner product on H is the collection of objects w such that $v \mapsto \langle v, w \rangle$ is continuous for every $v \in V$. Note that this condition for $v \in V$ is *looser* than asking for continuity for every $h \in H$. Hence:

$$H^* \subseteq V^*.$$

Some functional analysis, II

$$V \subseteq H \qquad\qquad H^* \subseteq V^*$$

By the Riesz Representation Theorem, we have,

$$V \subseteq H \subseteq V^*$$

$H^1_\rho \subseteq L^2 \subseteq ?$

Some functional analysis, II

$$V \subseteq H \qquad\qquad H^* \subseteq V^*$$

By the Riesz Representation Theorem, we have,

$$V \subseteq H \subseteq V^*$$

In this setup, the triple (V, H, V^*) is called a Gelfand triple, or a rigged Hilbert space.

A notable consequence of such a setup is that there is a natural pairing between elements v of V and w of V^*:

$H = L^2 \rightarrow \; // \; \int v(x) w(x) dx$

$$(v, w)_{V \times V^*} := \langle v, w \rangle_H.$$

The inner product above might not seem sensible because w can be too "rough" to belong to H.

However, the basic utility of this construction is that since v has extra smoothness, we can use integration by parts to transfer smoothness from v to w, which can yield a sensible inner product.

An example, I

recall: $\langle f, cg \rangle = \bar{c} \langle f, g \rangle$
$\langle cf, g \rangle = c \langle f, g \rangle$

Example

Consider $H = L^2([0, 2\pi]; \mathbb{C})$, with the standard inner product $\langle \cdot, \cdot \rangle$. Define:

$$V := H^1_p([0, 2\pi]; \mathbb{C}) = \left\{ f = \sum_{k \in \mathbb{Z}} c_k \phi_k(x) \in H \;\middle|\; f' \in H, \;\; f(0) = f(2\pi) \right\},$$

which is a subspace of H.

Then $H^{-1}_p := V^*$ is the space of functions whose first "Fourier" antiderivative is in L^2.

An example, I

Example

Consider $H = L^2([0, 2\pi]; \mathbb{C})$, with the standard inner product $\langle \cdot, \cdot \rangle$. Define:

$$V := H_p^1([0, 2\pi]; \mathbb{C}) = \left\{ f = \sum_{k \in \mathbb{Z}} c_k \phi_k(x) \in H \;\middle|\; f' \in H, \;\; f(0) = f(2\pi) \right\},$$

which is a subspace of H.

Then $H_p^{-1} := V^*$ is the space of functions whose first "Fourier" antiderivative is in L^2.

To see why, note that if $v \in V$ and we take some w satisfying $\langle w, \phi_0 \rangle = 0$ with antiderivative W,

$$(v, w) = \langle v, w \rangle \overset{\text{(IbP)}}{=} -\langle v', W \rangle,$$

$W v \Big|_0^{2\pi} = 0$

Hence, if $W \in L^2$, then,

$$|\,(v, w)\,| = \left|\langle v', W \rangle\right| \leqslant \|v'\|_{L^2} \|W\|_{L^2} \leqslant C(w) \|v\|_{H_p^1},$$

and hence $v \mapsto (v, w)$ is a bounded map, thus $w \in V^*$.

An example, II

$\phi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}$

Example

Consider $H = L^2([0,1];\mathbb{C})$, with the standard inner product $\langle \cdot, \cdot \rangle$. Define:

$$H = L^2, \qquad\qquad V := H^1_p([0,1];\mathbb{C})$$

What kinds of "functions" are in V^*? Consider the expression,

$$w(x) = \sum_{k \in \mathbb{Z}} \overline{\phi_k(0)} \phi_k(x) = \overline{\phi_0(0)} \phi_0(x) + \underbrace{\sum_{|k|>0} \overline{\phi_k(0)} \phi_k(x)}_{w_1(x)} = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{ikx}$$

An example, II

Example

Consider $H = L^2([0,1];\mathbb{C})$, with the standard inner product $\langle \cdot, \cdot \rangle$. Define:

$$H = L^2, \qquad V := H^1_p([0,1];\mathbb{C})$$

2π

What kinds of "functions" are in V^*? Consider the expression,

$$w(x) = \sum_{k \in \mathbb{Z}} \overline{\phi_k(0)}\phi_k(x) = \overline{\phi_0(0)}\phi_0(x) + \underbrace{\sum_{|k|>0} \overline{\phi_k(0)}\phi_k(x)}_{w_1(x)} = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} e^{ikx}$$

Note that this series is convergent nowhere since $|e^{ikx}| = 1$ for all k.

The (formally computed) antiderivative of w_1 is,

$$W_1(x) = \sum_{|k|>0} \frac{\overline{\phi_k(0)}}{ik}\phi_k(x),$$

$\|W_1\|^2_{L^2} \simeq \sum \frac{1}{k^2}$

which is an element of L^2 since the terms decay like $1/k$.

An example, III

$v = \sum_k \hat{v}_k \phi_k$

Example

This yields for arbitrary $v \in V$,

$$\begin{aligned}
(v, w) &= \left(v, \overline{\phi_0(0)}\phi_0(x)\right) + (v, w_1) \\
&= \left\langle v, \overline{\phi_0(0)}\phi_0(x) \right\rangle + \langle v, w_1 \rangle \\
&= \widehat{v}_0 \phi_0(0) - \langle v', W_1 \rangle.
\end{aligned}$$

Hence,

$$\begin{aligned}
(v, w) = \widehat{v}_0 \phi_0(0) - \left\langle v', W_1 \right\rangle &= \widehat{v}_0 \phi_0(0) - \left\langle \sum_{k \in Z} ik \widehat{v}_k \phi_k, \sum_{|\ell|>0} \frac{\overline{\phi_\ell(0)}}{i\ell} \phi_\ell \right\rangle \\
&= \widehat{v}_0 \phi_0(0) + \sum_{|k|,|\ell|>0} \frac{k}{\ell} \phi_\ell(0) \widehat{v}_k \langle \phi_k, \phi_\ell \rangle \\
&= \sum_{k \in \mathbb{Z}} \widehat{v}_k \phi_\ell(0) = v(0)
\end{aligned}$$

$T(v) : v \mapsto \langle v, w \rangle$

An example, III

Example

This yields for arbitrary $v \in V$,

$$
\begin{aligned}
(v, w) &= \left(v, \overline{\phi_0(0)}\phi_0(x)\right) + (v, w_1) \\
&= \left\langle v, \overline{\phi_0(0)}\phi_0(x)\right\rangle + \langle v, w_1\rangle \\
&= \widehat{v}_0\phi_0(0) - \langle v', W_1\rangle.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
(v, w) = \widehat{v}_0\phi_0(0) - \left\langle v', W_1\right\rangle &= \widehat{v}_0\phi_0(0) - \left\langle \sum_{k \in Z} ik\widehat{v}_k\phi_k, \sum_{|\ell|>0} \frac{\overline{\phi_\ell(0)}}{i\ell}\phi_\ell \right\rangle \\
&= \widehat{v}_0\phi_0(0) + \sum_{|k|,|\ell|>0} \frac{k}{\ell}\phi_\ell(0)\widehat{v}_k \langle \phi_k, \phi_\ell\rangle \\
&= \sum_{k \in \mathbb{Z}} \widehat{v}_k\phi_\ell(0) = v(0)
\end{aligned}
$$

(V, L^2, V^*)

Hence $w = \delta_0$, the Dirac delta centered at 0, is an element of $H_p^{-1} = V^*$.

$h \in L^2$ $\quad \langle h, w\rangle \neq h(0)$ $\quad L^2 \supset C \supset H_p^1$

Partial differential equations

Here is how all this scaffolding helps us with PDEs. Let's take a particular, simple example:

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

with $f(x)$ given.

Partial differential equations

Here is how all this scaffolding helps us with PDEs. Let's take a particular, simple example:

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

with $f(x)$ given.

From our discussion of Fourier approximation, a plausible strategy is to consider the problem:

$$\text{Find } u \in V \text{ such that } \langle \mathcal{R}(u), v\rangle = \langle f, v\rangle \quad \text{for every } v \in V,$$

where V is a subspace of L^2-periodic functions (e.g., frequency-truncated Fourier modes), and

$$\mathcal{R}(u) := -u''(x) + u(x).$$

Partial differential equations

Here is how all this scaffolding helps us with PDEs. Let's take a particular, simple example:

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

with $f(x)$ given.

From our discussion of Fourier approximation, a plausible strategy is to consider the problem:

$$\text{Find } u \in V \text{ such that } \langle \mathcal{R}(u), v \rangle = \langle f, v \rangle \quad \text{for every } v \in V,$$

where V is a subspace of L^2-periodic functions (e.g., frequency-truncated Fourier modes), and

$$\mathcal{R}(u) := -u''(x) + u(x).$$

Note that if $u \in V$, it need not be true that $u'' \in V$. Hence, it is useful to consider a Gelfand triple (V, L^2, V^*) to properly interpret $\langle \mathcal{R}(u), v \rangle$:

$$v \in V, \;\; \mathcal{R}(u) \in V^* \quad \Longrightarrow \quad \langle \mathcal{R}(u), v \rangle = (-u'', v) + \langle u, v \rangle \overset{\text{(IbP)}}{=} \left\langle u', v' \right\rangle + \langle u, v \rangle .$$

where we've used the boundary conditions for integration by parts.

Partial differential equations

Here is how all this scaffolding helps us with PDEs. Let's take a particular, simple example:

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

with $f(x)$ given.

From our discussion of Fourier approximation, a plausible strategy is to consider the problem:

$$\text{Find } u \in V \text{ such that } \langle \mathcal{R}(u), v \rangle = \langle f, v \rangle \quad \text{for every } v \in V,$$

where V is a subspace of L^2-periodic functions (e.g., frequency-truncated Fourier modes), and

$$\mathcal{R}(u) := -u''(x) + u(x).$$

Note that if $u \in V$, it need not be true that $u'' \in V$. Hence, it is useful to consider a Gelfand triple (V, L^2, V^*) to properly interpret $\langle \mathcal{R}(u), v \rangle$:

$$v \in V, \;\; \mathcal{R}(u) \in V^* \quad \Longrightarrow \quad \langle \mathcal{R}(u), v \rangle = (-u'', v) + \langle u, v \rangle \overset{\text{(IbP)}}{=} \left\langle u', v' \right\rangle + \langle u, v \rangle.$$

where we've used the boundary conditions for integration by parts. Hence, our new PDE statement could be instead:

$$\text{Find } u \in V \text{ such that } \left\langle u', v' \right\rangle + \langle u, v \rangle = \langle f, v \rangle \quad \text{for every } v \in V.$$

Weak enforcement

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

$$\text{Find } u \in V \text{ such that } \left\langle u', v' \right\rangle + \langle u, v \rangle = \langle f, v \rangle \quad \text{for every } v \in V.$$

The first statement above is the *strong form* of the PDE.

The second statement is called a weak form (or variational form) for the PDE, and the corresponding u (if it exists) is a weak solution.

Weak enforcement

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

$$\text{Find } u \in V \text{ such that } \langle u', v' \rangle + \langle u, v \rangle = \langle f, v \rangle \quad \text{for every } v \in V.$$

The first statement above is the *strong form* of the PDE.

The second statement is called a weak form (or variational form) for the PDE, and the corresponding u (if it exists) is a weak solution.

Note that we've "fixed" one issue that cropped up with strong solutions: It's ok if a weak solution u doesn't have two strong derivatives, it need only have a single *weak* derivative.

In addition, if u is a *bona fide* strong solution to the PDE (for every x), then it must also be a weak solution. The converse need not be true.

Weak enforcement

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

$$\text{Find } u \in V \text{ such that } \langle u', v' \rangle + \langle u, v \rangle = \langle f, v \rangle \quad \text{for every } v \in V.$$

The first statement above is the *strong form* of the PDE.

The second statement is called a weak form (or variational form) for the PDE, and the corresponding u (if it exists) is a weak solution.

Note that we've "fixed" one issue that cropped up with strong solutions: It's ok if a weak solution u doesn't have two strong derivatives, it need only have a single *weak* derivative.

In addition, if u is a *bona fide* strong solution to the PDE (for every x), then it must also be a weak solution. The converse need not be true.

This also gives us some notion of what kind of function f is allowed to be for weak solutions: the term $\langle f, v \rangle$ only makes sense if $f \in V^*$ through the duality pairing on (V, L^2, V^*). But recall that elements of V^* can be quite "rough", which is a considerable relaxation from what would be required for strong solutions.

Weak enforcement

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

$$\text{Find } u \in V \text{ such that } \langle u', v' \rangle + \langle u, v \rangle = \langle f, v \rangle \quad \text{for every } v \in V.$$

The first statement above is the *strong form* of the PDE.

The second statement is called a weak form (or variational form) for the PDE, and the corresponding u (if it exists) is a weak solution.

Note that we've "fixed" one issue that cropped up with strong solutions: It's ok if a weak solution u doesn't have two strong derivatives, it need only have a single *weak* derivative.

In addition, if u is a *bona fide* strong solution to the PDE (for every x), then it must also be a weak solution. The converse need not be true.

This also gives us some notion of what kind of function f is allowed to be for weak solutions: the term $\langle f, v \rangle$ only makes sense if $f \in V^*$ through the duality pairing on (V, L^2, V^*). But recall that elements of V^* can be quite "rough", which is a considerable relaxation from what would be required for strong solutions.

However there are some nontrivial questions that arise here. Perhaps the foremost questions are: Do weak solutions exist in general? Is the weak form well-posed?

A little extra notation/terminology is required:

Let $a(\cdot,\cdot) : V \times V \to \mathbb{C}$ be a sesquilinear form[2]. A sesquilinear/bilinear form $a(\cdot,\cdot)$ is (strongly) coercive or elliptic, if there exists a constant $c > 0$ such that,

$$|a(v,v)| \geqslant c\|v\|_V^2, \quad \text{for every } v \in V,$$

It is bounded (or continuous) if there is a $C > 0$ such that $|a(u,v)| \leqslant C\|u\|\|v\|$ for every $v \in V$.

[2] I.e., linear in the first argument, conjugate linear in the second. If the field is $\mathbb{R}$ instead of $\mathbb{C}$, linearity in the second argument is enough, and such an a is bilinear.

A little extra notation/terminology is required:

Let $a(\cdot,\cdot) : V \times V \to \mathbb{C}$ be a sesquilinear form[2]. A sesquilinear/bilinear form $a(\cdot,\cdot)$ is (strongly) coercive or elliptic, if there exists a constant $c > 0$ such that,

$$|a(v,v)| \geqslant c\|v\|_V^2, \quad \text{for every } v \in V,$$

It is bounded (or continuous) if there is a $C > 0$ such that $|a(u,v)| \leqslant C\|u\|\|v\|$ for every $v \in V$.

The following is one of the foundational results in modern PDE theory:

Theorem (Lax-Milgram)

a Hilbert space

Let $a(\cdot,\cdot)$ be a sesquilinear/bilinear form on V, and let (V, H, V^*) be a Gelfand triple with and H-inner product $\langle\cdot,\cdot\rangle$. Assume a is coercive (with constant c) and bounded, and let $f \in V^*$.

Then there exists a unique solution $u \in V$ to the problem,

$$\textit{Find } u \in V \textit{ such that } a(u,v) = \langle f, v\rangle \quad \textit{for every } v \in V.$$

Moreover, the solution is well-posed with respect to f:

$$\|u\|_V \leqslant \frac{1}{c}\|f\|_{V^*}.$$

[2] I.e., linear in the first argument, conjugate linear in the second. If the field is $\mathbb{R}$ instead of $\mathbb{C}$, linearity in the second argument is enough, and such an a is bilinear.

An interpretation of Lax-Milgram

The Lax-Milgram theorem is an existence/uniqueness statement for an infinite-dimensional analogue of,

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{f}, \qquad \boldsymbol{A} = \boldsymbol{A}^*$$

– $a(u, v)$ is the infinite-dimensional version of $\boldsymbol{v}^*\boldsymbol{A}\boldsymbol{u}$ for vectors $\boldsymbol{u}, \boldsymbol{v}$.
– $a(u, v)$ being coercive and bounded is analogous to statements about the singular values of $\boldsymbol{A}$: $\sigma_{\min}(\boldsymbol{A}) > 0$ and $\sigma_{\max}(\boldsymbol{A}) < \infty$.
– $f \in V^*$ is equivalent to the condition that $\boldsymbol{v} \mapsto \boldsymbol{f}^*\boldsymbol{v}$ is a bounded map, i.e., $\boldsymbol{f}$ has to be a finite vector.
– In finite dimensions, V, H, and V^* are all the same space $\mathbb{C}^N$ since all norms in finite dimensions are equivalent.
– The statement $a(u, v) = \langle f, v\rangle$ for every $\boldsymbol{v} \in V$ is analogous to $\boldsymbol{v}^*\boldsymbol{A}\boldsymbol{u} = \boldsymbol{v}^*\boldsymbol{f}$ for every $\boldsymbol{v} \in \mathbb{C}^N$.

An interpretation of Lax-Milgram

The Lax-Milgram theorem is an existence/uniqueness statement for an infinite-dimensional analogue of,

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{f}, \qquad \boldsymbol{A} = \boldsymbol{A}^*$$

- $a(u, v)$ is the infinite-dimensional version of $\boldsymbol{v}^* \boldsymbol{A}\boldsymbol{u}$ for vectors $\boldsymbol{u}, \boldsymbol{v}$.
- $a(u, v)$ being coercive and bounded is analogous to statements about the singular values of $\boldsymbol{A}$: $\sigma_{\min}(\boldsymbol{A}) > 0$ and $\sigma_{\max}(\boldsymbol{A}) < \infty$.
- $f \in V^*$ is equivalent to the condition that $\boldsymbol{v} \mapsto \boldsymbol{f}^* \boldsymbol{v}$ is a bounded map, i.e., $\boldsymbol{f}$ has to be a finite vector.
- In finite dimensions, V, H, and V^* are all the same space $\mathbb{C}^N$ since all norms in finite dimensions are equivalent.
- The statement $a(u, v) = \langle f, v \rangle$ for every $\boldsymbol{v} \in V$ is analogous to $\boldsymbol{v}^* \boldsymbol{A}\boldsymbol{u} = \boldsymbol{v}^* \boldsymbol{f}$ for every $\boldsymbol{v} \in \mathbb{C}^N$.

Requiring $|a(v, v)| \geqslant c\|v\|^2$ is analogous to $\sigma_{\min}(\boldsymbol{A}) > 0$ (at least for Hermitian $\boldsymbol{A}$). I.e., $c = \sigma_{\min}(\boldsymbol{A})$.
If $\sigma_{\min}(\boldsymbol{A}) > 0$, then $\boldsymbol{A}$ is invertible, hence there is a unique solution $\boldsymbol{u}$. In particular:

$$\|\boldsymbol{u}\|_2 \leqslant \|\boldsymbol{A}^{-1}\boldsymbol{f}\|_2 \leqslant \|\boldsymbol{A}^{-1}\|_2 \|\boldsymbol{f}\|_2 = \sigma_{\max}(\boldsymbol{A}^{-1})\|\boldsymbol{f}\|_2 = \frac{1}{\sigma_{\min}(\boldsymbol{A})}\|\boldsymbol{f}\|_2 = \frac{1}{c}\|\boldsymbol{f}\|_2,$$

which is precisely the well-posedness statement of Lax-Milgram.

Well-posedness for elliptic equations, I

We can immediately demonstrate the utility of Lax-Milgram:

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

with

u(0) = u(2π)
V

$$H = \left\{ u(x) = \sum_{k \in \mathbb{Z}} c_k \phi_k(x) \;\middle|\; \|u\|_{L^2} < \infty \right\}, \qquad V = \left\{ u \in H \;\middle|;\; u' \in H \right\}.$$

With $\langle \cdot, \cdot \rangle$ the standard L^2 inner product on $[0, 2\pi]$, the norms on H and V are:

$$\|u\|_H^2 = \langle u, u \rangle, \qquad\qquad \|u\|_V^2 = \langle u', u' \rangle + \langle u, u \rangle.$$

Well-posedness for elliptic equations, I

We can immediately demonstrate the utility of Lax-Milgram:

$$-u''(x) + u(x) = f(x), \qquad\qquad u(0) = u(2\pi),$$

with

$$H = \left\{ u(x) = \sum_{k \in \mathbb{Z}} c_k \phi_k(x) \;\middle|\; \|u\|_{L^2} < \infty \right\}, \qquad V = \left\{ u \in H \;\middle|;\; u' \in H \right\}.$$

With $\langle \cdot, \cdot \rangle$ the standard L^2 inner product on $[0, 2\pi]$, the norms on H and V are:

$$\|u\|_H^2 = \langle u, u \rangle, \qquad\qquad \|u\|_V^2 = \langle u', u' \rangle + \langle u, u \rangle.$$

Define the bilinear form,

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

which satisfies:

$$\begin{aligned} |a(v,v)| &= |\langle v', v' \rangle + \langle v, v \rangle| = \|v'\|_H^2 + \|v\|_H^2 = \|v\|_V^2, \\ |a(u,v)| &\leqslant |\langle u', v' \rangle| + |\langle v, v \rangle| = \|u'\|_H \|v'\|_H + \|u\|_H \|v\|_H \\ &\leqslant (\|u\|_H + \|u'\|_H)(\|v\|_H + \|v'\|_H) \leqslant \sqrt{2} \|u\|_V \|v\|_V \end{aligned}$$

and hence a is coercive (with constant 1) and is bounded.

Well-posedness for elliptic equations, II

Hence, Lax-Milgram implies that the variational problem,

$$\text{Find } u \in V \text{ such that } a(u,v) = \langle f, v \rangle \quad \text{for every } v \in V.$$

has a unique solution if $f \in V^* = H_p^{-1}$, and $\|u\|_V \leqslant \|f\|_{V^*}$.

Well-posedness for elliptic equations, II

Hence, Lax-Milgram implies that the variational problem,

$$\text{Find } u \in V \text{ such that } a(u,v) = \langle f, v \rangle \quad \text{for every } v \in V.$$

has a unique solution if $f \in V^* = H_p^{-1}$, and $\|u\|_V \leqslant \|f\|_{V^*}$.

Note that this statement is an abstract existence/uniqueness statement, and does *not* depend on discretizations.

Of course, there is nothing stopping us from taking V as a finite-dimensional subspace....

Discrete schemes

With the same setup as before....

If V_N is a finite-dimensional (say N-dimensional) subspace of V, then $a(\cdot, \cdot)$ is still a continuous, coercive operator on $V_N \times V_N$, and therefore,

$$\text{Find } u_N \in V_N \text{ such that } a(u_N, v) = \langle f, v \rangle \quad \text{for every } v \in V_N,$$

specifies a unique solution u_N, by Lax-Milgram.

However, *how accurate* is this solution? I.e., what is $\|u - u_N\|_V$?

Discrete schemes

With the same setup as before....

If V_N is a finite-dimensional (say N-dimensional) subspace of V, then $a(\cdot,\cdot)$ is still a continuous, coercive operator on $V_N \times V_N$, and therefore,

$$\text{Find } u_N \in V_N \text{ such that } a(u_N, v) = \langle f, v \rangle \quad \text{for every } v \in V_N,$$

specifies a unique solution u_N, by Lax-Milgram.

$a(u_N, v) = \int_0^{2\pi} u_N' v' dx + \int_0^{2\pi} u_N v dx$

However, *how accurate* is this solution? I.e., what is $\|u - u_N\|_V$?

Lemma (Céa)

Consider the setup as above: (V, H, V^*) is a Gelfand triple, $a(\cdot,\cdot)$ a continuous and coercive bilinear form on $V \times V$, $f \in V^*$ is given, and u is the unique weak solution to $a(u,v) = \langle f, v\rangle$ for every $v \in V$.

With V_N some finite-dimensional subspace of V, let u_N solve the above finite-dimensional weak form. Then,

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

where C and c are the continuity and coercivity constants, respectively, of a.

Discrete schemes

With the same setup as before....

If V_N is a finite-dimensional (say N-dimensional) subspace of V, then $a(\cdot,\cdot)$ is still a continuous, coercive operator on $V_N \times V_N$, and therefore,

$$\text{Find } u_N \in V_N \text{ such that } a(u_N, v) = \langle f, v\rangle \quad \text{for every } v \in V_N,$$

specifies a unique solution u_N, by Lax-Milgram.

However, *how accurate* is this solution? I.e., what is $\|u - u_N\|_V$?

Lemma (Céa)

Consider the setup as above: (V, H, V^*) is a Gelfand triple, $a(\cdot,\cdot)$ a continuous and coercive bilinear form on $V \times V$, $f \in V^*$ is given, and u is the unique weak solution to $a(u,v) = \langle f, v\rangle$ for every $v \in V$.

With V_N some finite-dimensional subspace of V, let u_N solve the above finite-dimensional weak form. Then,

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

where C and c are the continuity and coercivity constants, respectively, of a.

I.e., to within the factor C/c, u_N is the best possible approximation to the weak solution u.

Generalizations of Lax-Milgram

The Lax-Milgram theorem has some quite useful generalizations:

- Hilbertian structure is not needed; V can be a Banach space with dual V^*.
- $a(\cdot, \cdot)$ can operate on different spaces $U \times V$. One requires appropriate generalizations of continuity and coercivity.

This theory is limited to linear PDE's, and is typically used for elliptic-type or parabolic-type PDEs.

Weighted residual methods

The idea of weak formulations is at the heart of numerous numerical methods for PDEs.

For our prototypical PDE,

$$\mathcal{L}(u) = f, \qquad\qquad \mathcal{R}(u) = \mathcal{L}(u) - f,$$

a generic formulation for a weak solution is given by,

$$\text{Find } u \in V \text{ satisfying } \langle \mathcal{R}(u), v \rangle = 0 \text{ for all } v \in V,$$

where V is some Banach or Hilbert space, and $\langle \cdot, \cdot \rangle$ is some inner product (or duality pairing).

Weighted residual methods

The idea of weak formulations is at the heart of numerous numerical methods for PDEs.

For our prototypical PDE,

$$\mathcal{L}(u) = f, \qquad\qquad \mathcal{R}(u) = \mathcal{L}(u) - f,$$

a generic formulation for a weak solution is given by,

$$\text{Find } u \in V \text{ satisfying } \langle \mathcal{R}(u), v \rangle = 0 \text{ for all } v \in V,$$

where V is some Banach or Hilbert space, and $\langle \cdot, \cdot \rangle$ is some inner product (or duality pairing).

Taking V as some finite-dimensional space makes the above a numerical scheme (implementable in principle).
Such a scheme is called a weighted residual method, as one forms the scheme by enforcing that the residual vanish in some weighted sense. (Inner products can be weighted.)

Weighted residual methods

The idea of weak formulations is at the heart of numerous numerical methods for PDEs.

For our prototypical PDE,

$$\mathcal{L}(u) = f, \qquad\qquad \mathcal{R}(u) = \mathcal{L}(u) - f,$$

a generic formulation for a weak solution is given by,

$$\text{Find } u \in V \text{ satisfying } \langle \mathcal{R}(u), v \rangle = 0 \text{ for all } v \in V,$$

where V is some Banach or Hilbert space, and $\langle \cdot, \cdot \rangle$ is some inner product (or duality pairing).

Taking V as some finite-dimensional space makes the above a numerical scheme (implementable in principle).
Such a scheme is called a weighted residual method, as one forms the scheme by enforcing that the residual vanish in some weighted sense. (Inner products can be weighted.)

Some additional terminology is useful to know:

- The space of functions from which we select u is the trial space. (It's V above.)
- The space of functions that we use to weakly enforce zero residual is the test space. (It's also V above.)

A weighted residual method for which trial and test spaces coincide is called a Galerkin scheme or approximation.

Other weighted residual methods

In general, we can have different trial and test spaces:

$$\text{Find } u \in U \text{ satisfying } \langle \mathcal{R}(u), v \rangle = 0 \text{ for all } v \in V.$$

If $U \neq V$, the weighted residual method above is generically a Petrov-Galerkin method.

Other weighted residual methods

In general, we can have different trial and test spaces:

$$\text{Find } u \in U \text{ satisfying } \langle \mathcal{R}(u), v \rangle = 0 \text{ for all } v \in V.$$

If $U \neq V$, the weighted residual method above is generically a Petrov-Galerkin method.

A particularly salient specialization of Petrov-Galerkin methods occurs when the test space V is chosen as,

$$V = \text{span}\left\{\delta_{x_1}, \ldots, \delta_{x_M}\right\},$$

where δ_x is the Dirac delta centered at a spatial location x.

Other weighted residual methods

In general, we can have different trial and test spaces:

$$\text{Find } u \in U \text{ satisfying } \langle \mathcal{R}(u), v \rangle = 0 \text{ for all } v \in V.$$

If $U \neq V$, the weighted residual method above is generically a Petrov-Galerkin method.

A particularly salient specialization of Petrov-Galerkin methods occurs when the test space V is chosen as,

$$V = \operatorname{span}\left\{\delta_{x_1}, \ldots, \delta_{x_M}\right\},$$

where δ_x is the Dirac delta centered at a spatial location x.

In this case,

$$\langle \mathcal{R}(u), v \rangle = 0 \text{ for all } v \in V,$$

is equivalent to

$$\mathcal{R}(u)\Big|_{x=x_m} = 0, \qquad\qquad m \in [M].$$

This particular Petrov-Galerkin method is a collocation scheme.
(This is *not* a finite difference scheme since u is a function, unlike a finite difference solution.)

Trial and test spaces

Weighted residual methods are a somewhat different philosophy compared to finite difference methods.

In weighted residual methods, we *weakly* enforce the PDE, and must make some choices:

- How do we represent the solution u? (The trial space U)
- How do we satisfy the PDE? (The test space V)

These freedoms allow significant flexibility in designing and analyzing numerical schemes.

References I

Cea, Jean (1964). "Approximation Variationnelle Des Problèmes Aux Limites". In: *Annales de l'Institut Fourier* 14.2, pp. 345–444. ISSN: 1777-5310. DOI: 10.5802/aif.181.

Hesthaven, Jan S., Sigal Gottlieb, and David Gottlieb (2007). *Spectral Methods for Time-Dependent Problems*. Cambridge University Press. ISBN: 0-521-79211-8.

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.