Math 6008 Numerical PDEs–Lecture 13

Finite element method

Instructor: Lei Li, INS, Shanghai Jiao Tong University;
Email: leili2010@sjtu.edu.cn

1 (Continuation of) variational problems for differential equations

In last lecture, we consider the Ritz and Galerkin variational problems for 1D problems. Here, we consider the poblems in multidimensional space.

1.1 Problems in multidimensional space

For problems defined in higher dimensions, the weak formulation or the Ritz variational formulation can be derived similarly.

Here, we look at some examples and consider deriving the Galerkin's weak formulation only. The Ritz formulation is left for your exercise.

The Poisson equation

Consider

$$-\Delta u = f, \quad u|_{\partial\Omega} = 0.$$

To obtain the weak formulation, we multiply the test function v and integrate:

$$-\int_{\Omega} (\Delta u) v\, dx = \int f v\, dx.$$

By parts,

$$-\int_{\partial\Omega} \frac{\partial u}{\partial n} v dS + \int_{\Omega} \nabla u \cdot \nabla v\, dx = \int_{\Omega} f v\, dx.$$

Same as before, we need to impose $v|_{\partial\Omega} = 0$. Otherwise, there is a boundary term $\frac{\partial u}{\partial n}$, and we need higher regularity of u to define $\partial u/\partial n$. Hence, the space is

$$H = \{v : \int v^2 + |\nabla v|^2 dx < \infty, v|_{\partial\Omega} = 0\}.$$

The space for u is the same here. Hence, the weak formulation is: Find $u \in H$ such that

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx, \quad \forall v \in H.$$

To find the corresponding Ritz variational formulation, we set $v = \delta u$ and do $\nabla u \cdot \nabla \delta u \to \delta \frac{1}{2} |\nabla u|^2$. We will get: find $u \in H$ to minimize

$$J(u) = \frac{1}{2} D(u,u) - F(u) = \frac{1}{2} \int_\Omega |\nabla u|^2 - \int f u \, dx.$$

The Poisson equation with mixed boundary condition

Consider

$$-\Delta u = f, \quad u|_{\partial\Omega_1} = \varphi, \quad (\frac{\partial u}{\partial n} + \alpha u)|_{\partial\Omega_2} = g.$$

To derive the weak formulation,

$$-\int (\Delta u) v \, dx = \int f v \, dx.$$

By parts,

$$-\int_{\partial\Omega} \frac{\partial u}{\partial n} v \, dS + \int \nabla u \cdot \nabla v = \int f v.$$

Clearly, one can replace $\partial u / \partial n$ with $g - \alpha u$ on $\partial\Omega_2$, yielding

$$-\int_{\partial\Omega_1} \frac{\partial u}{\partial n} v \, dS - \int_{\partial\Omega_2} g v \, dS + \int_{\partial\Omega_2} \alpha u v \, dS + \int \nabla u \cdot \nabla v = \int f v.$$

For $\partial\Omega_1$, we may impose $v(x) = 0$, so the term for $\partial u / \partial n = 0$ will vanish. Then, one has

$$\int_{\partial\Omega_2} \alpha u v \, dS + \int \nabla u \cdot \nabla v = \int f v + \int_{\partial\Omega_2} g v \, dS.$$

On $\partial\Omega_2$, there is a term $\int gv$ to balance, hence, we do not have to impose like $v = 0$. In fact, we should not impose this. Otherwise, the condition cannot be recovered, giving unsuitable problem.

Hence, the space for v is

$$H = \{v : \int v^2 + |\nabla v|^2 dx < \infty, \quad \int_{\partial\Omega_2} v^2 dS < \infty, v = 0, \partial\Omega_1\}.$$

The space for the solution is

$$E = \{u : \int u^2 + |\nabla u|^2 dx < \infty, \int_{\partial\Omega_2} u^2 dS < \infty, \quad u = \varphi, \partial\Omega_1\}.$$

Think about the Ritz variational formulation.

Remark 1. *There is an example for biharmonic equation. Read by yourself. Note that in the biharmonic equation, the Neumann boundary condition is no longer the natural boundary condition.*

2 Approximation of the variational problems

Previously, we have the Galerkin and Ritz variational problems for the PDEs. However, the space H and E are of inifinite dimension, which brings difficult for us to solve. For approximation, one will use a finite dimensional space to approximate the spaces. This approach will then give some approximation methods.

2.1 Ritz and Galerkin

Consider the space H, which is a linear vector space. We may choose a finite dimensional subspace of H, denoted by S_N with a basis $\phi_i, 1 \le i \le N$:

$$S_N = \mathrm{span}\{\phi_1, \cdots, \phi_N\} \subset H.$$

Noting that $E = u_0 + H$, we can thus naturally approximate E by

$$U_N := u_0 + S_N.$$

Of course, U_N is a vector space only if u has a homogeneous boundary condition, in which case $U_N = S_N$.

The Ritz formulation can thus be approximated by:

Definition 1. *The Ritz method is to solve the following problem: find $u_N \in U_N$ such that*

$$J(u) = \frac{1}{2}D(u,u) - F(u)$$

is minimized on U_N. In other words, for all $w \in U_N$, $J(u_N) \leq J(w)$.

Similarly, the Galerkin's method is given as follows.

Definition 2. *The Galerkin's method is to solve the following problem. Find $u_N \in U_N$ such that*

$$D(u_N, w) = F(u_N), \forall w \in S_N.$$

In these two methods, we may write

$$u_N = \sum_{i=1}^{N} c_i \phi_i(x).$$

Hence, the problems are reduced to problems for $c = (c_1, \cdots, c_N)$. In other words, we have an optimization problem in $\mathbb{R}^N$ and a linear system in $\mathbb{R}^N$. For example, for the Galerkin's method, one has

$$D(u_0, \phi_i) + \sum_{j=1}^{N} c_j D(\phi_j, \phi_i) = F(\phi_i), \forall i = 1, \cdots, N.$$

This naturally gives a linear system for c.

Note that these approximation methods are not only computational methods. They can be used for theoretical analysis as well, where the basis can be taken as the polynomials (like the Legendre polynomials) or trigonometric polynomials (the sine and cosine modes).

2.2 Various other methods

We take the problem $Lu = f$ as the example.

Recall that the Galerkin's formulation corresponds to the virtual work principle

$$\langle J'(u), v \rangle = 0.$$

Here, $J'(u) = Lu - f$ is the operator acting on u. The differential equation is

$$Lu - f = 0.$$

For a general u, $Lu - f \neq 0$. Hence, it is natural to introduce the **defect** or residue (余量):

$$R(u) = Lu - f.$$

One may do least square method (最小二乘) to minize R and solve for c_i.

Here, we focus on the so-called weighted residue method (权余量方法). One seeks the solution u_h in a subspace $U_h \subset E$ as well. However, the space for the test functions may be another V_h that is different from U_h. Often, one let them to have the same dimension so that the system is a square system.

Let

$$U_h = u_0 + \mathrm{span}(\phi_1, \cdots, \phi_N)$$

and

$$V_h = \mathrm{span}(\psi_1, \cdots, \psi_N).$$

Definition 3. *The weighted residue method is: find $u_h \in U_h$ such that*

$$\langle R(u_h), \psi_i \rangle = 0, \forall 1 \le i \le N.$$

This means the weigthed integral of the residue is zero. Of course, if $\psi_i = \phi_i$ and ψ_i has zero bounday value, this essentially is the Galerkin's method.

There are some special cases:

- If one chooses $\psi_i = \delta(x - x_i)$ for some discrete points x_i, called the collocation points (配置点), then one has the collocation method (配置法):

  $$L(u_h(x_i)) - f(x_i) = 0.$$

  The collocation method is a generalization of the finite difference method, and the difference is that the operator here can be the real differential operator.

- If one takes $\psi_i = 1_{\Omega_i}$, then one has the subregion collocation method (子区域配置法)

In the weighted residue method, when ϕ (and thus u_h) has enough regularity, ψ does not have to have zero boundary conditions. If ψ has zero boundary conditions, one may move some derivatives onto ψ and get

$$D(u_h, \psi_i) = 0.$$

This is called the **Petrov-Galerkin's method**, which could be beneficial for some stationary advection-diffusion equations. Clearly, if $\psi_i = \varphi_i$, it reduces to Galerkin's method.

Remark 2. *In the weighted residue method, it seems that one needs $u_h \in H^2$ with second order derivatives to be defined while in $D(u_h, \psi_i)$, it is not needed. In fact, this is not necessarily the case; if one allows $R(u)$ to be distributions in some negative Sobolev spaces, u is allowed not to be so smooth; of course the test functions need to have better regularity and certain zero boundary condition–these are beyond our course so we omit.*

3 The finite element method

In the finite element method (有限元法), one constructs the finite dimensional subspace using the triangulation of the domains. The coefficinets of the basis functions can be interpretated as the function values or the vlaues of the derivatives at the vertices.

Below, we consider the **Galerkin's FEM** only.

We will only give a brief introduction. If you would like to know more, read the book "有限单元法王勖成" for basics and applications. If you would like to know more about the mathematical theory, read the book by Johnson as in the reference list in our syllabus.

3.1 Galerkin's FEM for the 1D problems

Consider generally the problem

$$
\begin{aligned}
&-(p(x)u')' + \mu(x)u' + q(x)u = f, \quad a < x < b, \\
&u(a) = A, \quad p(b)u'(b) + \alpha u(b) = g.
\end{aligned}
$$

Here, $p(x) \geq p_0 > 0$.

It is hard to construct an energy functional so that the Ritz formulation is hard for this equation due to $\mu(x)u'$. Anyhow, the Galerkin's formulation is easy to obtain:

$$\text{(W)} \quad \text{Find } u \in u_0 + H \text{ such that}$$

$$\int_a^b p(x)u'(x)v'(x)dx + \int_a^b \mu(x)u'vdx + \int_a^b q(x)uvdx + \alpha u(b)v(b) = \int_a^b f(x)v(x)dx + gv(b).$$

for all $v \in H$, where u_0 is a particular function with $u(a) = A$.

$$H = \{v : \int_a^b v^2 + (v')^2 dx < \infty, v(a) = 0\}$$

Below, we consider $A = 0$ so that we can choose $u_0 = 0$ and thus both u, v are chosen from H for convenience of discussion.

We will denote the bilinear form:

$$D(u, v) = \int_a^b p(x)u'(x)v'(x)dx + \int_a^b \mu(x)u'vdx + \int_a^b q(x)uvdx + \alpha u(b)v(b),$$

and the linear form

$$F(v) = \int_a^b f(x)v(x)dx + gv(b).$$

3.2 The triangulation and the construction of spaces

Now, we investigate how to construct the subspace using the triangulation.

Consider a triangulation of $[a, b]$ with grid points

$$a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b.$$

$h_j = x_j - x_{j-1}$. Then, each interval $e_j = [x_{j-1}, x_j]$ is called an **element** (单元).

Below, we construct a subspace V_h of H mentioned above using this triangulation. A common way is to let $\varphi \in V_h$ be a polynomial on each element (on the whole region, it is not necessarily a polynomial), because polynomials are simple functions. To start with, we consider the piecewise linear functions because linear functions are the simplest polynomials.

For $V_h \subset H$, we need to require that every $\varphi \in V_h$ to be continuous. In fact, if it is not continuous, the derivative is not a (measureable) function (it is a generalized function, or distribution), and thus not square integrable. Reversely, if it is continuous, the derivative is then defined almost everywhere and is piecewise constant. Hence, it is square integrable.

Consider

$$U_h = \{\varphi : \varphi \text{ is a piecewise linear continuous function.}\}$$

It is not hard to verify that the dimension is $N+1$, where the basis function is $\phi_i(x)$ such that $\phi_i(x_j) = \delta_{ij}$. Since it is piecewise linear, ϕ_i is determined uniquely by the values at the vertices. The formula of ϕ_i is easily written out

$$\phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{h}, & x_{i-1} \le x < x_i, \\ \frac{x_{i+1} - x}{h_{i+1}} & x_i \le x < x_{i+1}, \\ 0 & \text{otherwise.} \end{cases}$$

With the requirement that $v(a) = 0$, we see clearly that the subspace V_h we need is given by

$$V_h = \{v_h : v_h \in U_h, \quad v_h(a) = 0\} = \sum_{j=1}^{N} v_j \phi_i(x).$$

Clearly, for any $v_h \in V_h$,

$$v_h = \sum_{j=1}^{N} v_j \phi_i(x),$$

the coefficients v_j are the **values** at the vertices.

Since we assumed $A = 0$, so the approximation solution we are seeking for is also from V_h. The Galerkin's method can therefore be written as

Find $u_h \in V_h$ such that

$$D(u_h, v_h) = F(v_h), \quad \forall v_h \in V_h.$$

u_h is then the numerical solution obtained by Galerkin’ s FEM. It is then equivalent to find the coefficients u_j.

Formulating the linear system

Let us plug in the form of u_h and v_h:

$$u_h = \sum_j u_j \phi_j(x), \quad v_h = \sum_i v_i \phi_i(x).$$

Then, the problem is reduced to

$$\sum_{j,i} u_j v_i D(\phi_j, \phi_i) = \sum_i v_i F(\phi_i).$$

Since v_i is arbitrary, one in fact has

$$\sum_{j=1}^{N} D(\phi_j, \phi_i) u_j = F(\phi_i), \forall i = 1, \cdots, N.$$

Note that due to $\mu u' v$ term, D is not symmetric, $D(\phi_j, \phi_i) \neq D(\phi_i, \phi_j)$. If there is no $\mu u'$ term as in the book, then it is symmetric.

Define K_{ij} to be the coefficient in the front of u_j and $F_i = F(\phi_i)$. Then, one has a linear system

$$Ku = F.$$

K is called the (total) **stiff matrix**(刚度矩阵) and F is called the (total) **load vector**(荷载向量).

3.3 Computation of the stiff matrix

How do we construct the matrix? If we compute K_{ij}'s by looping over i and j since it will cost $O(N^2)$ iterations. For each iteration, if you compute the integral on $[a, b]$, it will cost another $O(N)$ and the total cost is $O(N^3)$. If you observe that ϕ_i is only nonzero on $[x_{i-1}, x_{i+1}]$, then you may reduce the cost per iteration to $O(1)$, but the total cost is still $O(N^2)$. This is not very satisfactory.

The observation is that for many i, j, K_{ij} is zero. Hence, there should be some other indices for us to loop over to reduce the cost. The idea is to break the integrals $\int_0^1$ into $\sum_i \int_{e_i}$. We then loop over e_i. In particular, one writes

$$\sum_i \int_{e_i} (p(x)u_h'(x)v_h'(x)dx + \mu(x)u_h' v_h dx + q(x)u_h v_h)dx + \alpha u(b)v(b)$$
$$= \sum_i \int_{e_i} f(x)v_h(x)dx + gv(b)$$

Each element e_i will only contribute to four components in the stiff matrix: $A_{\ell m}, i-1 \leq \ell, m \leq i$ and b_{i-1} and b_i because only ϕ_{i-1} and ϕ_i will be nonzero on e_i. In particular, such a procedure can be divided into two steps:

- For each e_i, compute the element stiffness matrix (单元刚度矩阵) and the element load vector (单元荷载向量)
- Assembling procedure. The element stiffness matices and load vectors will assembled into the total stiffness matrix and the total load vector.
- Treating the constraints and the boundary conditions.

Read the book for more details.

In the actual code, one does not compute the element matrices first. In fact, one directly adds the entries into the total stiffness matrix and the load vector.

For the integrals, you can compute the entries by hand, or you may use subroutine to compute the integrals numerically.

Code presentation We'll solve the equation

$$-(a(x)u')' = f(x), \quad u(0) = u(1) = 0.$$

In the code, we construct the stiff matrix and load vector by code instead of by formulas obtained by hand.

For the mathematical proof of the convergence, see the book by Johnson. We omit here.