Study Guide: Solving differential equations with finite elements

Hans Petter Langtangen[1,2]

Center for Biomedical Computing, Simula Research Laboratory[1]

Department of Informatics, University of Oslo[2]

Oct 23, 2013

Differential equation models

Our aim is to extend the ideas for approximating f by u, or solving

$$u = f$$

to real differential equations.
Three methods:

1. least squares
2. Galerkin/projection
3. collocation (interpolation)

Method 2 will be totally dominating!

Abstract differential equation

$$\mathcal{L}(u) = 0, \quad x \in \Omega \tag{1}$$

Examples:

$$\mathcal{L}(u) = \frac{d^2 u}{dx^2} - f(x), \tag{2}$$

$$\mathcal{L}(u) = \frac{d}{dx}\left(\alpha(x)\frac{du}{dx}\right) + f(x), \tag{3}$$

$$\mathcal{L}(u) = \frac{d}{dx}\left(\alpha(u)\frac{du}{dx}\right) - au + f(x), \tag{4}$$

$$\mathcal{L}(u) = \frac{d}{dx}\left(\alpha(u)\frac{du}{dx}\right) + f(u, x) \tag{5}$$

Abstract boundary conditions

$$\mathcal{B}_0(u) = 0,\ x = 0, \quad \mathcal{B}_1(u) = 0,\ x = L \tag{6}$$

Examples:

$$\mathcal{B}_i(u) = u - g, \qquad \text{Dirichlet condition} \tag{7}$$

$$\mathcal{B}_i(u) = -\alpha\frac{du}{dx} - g, \qquad \text{Neumann condition} \tag{8}$$

$$\mathcal{B}_i(u) = -\alpha\frac{du}{dx} - h(u - g), \qquad \text{Robin condition} \tag{9}$$

Reminder about notation

- $u_e(x)$ is the symbol for the *exact* solution of $\mathcal{L}(u_e) = 0$
- $u(x)$ denotes an *approximate* solution
- $V = \text{span}\{\psi_0(x), \ldots, \psi_N(x)\}$: we seek $u \in V$
- V has basis $\{\psi_i\}_{i\in I}$
- $I = \{0, \ldots, N\}$ is an index set
- $u(x) = \sum_{j\in I} c_j\psi_j(x)$
- Inner product: $(u, v) = \int_\Omega uv\,\mathrm{d}x$
- Norm: $||u|| = \sqrt{(u, u)}$

Residual-minimizing principles

- When solving $u = f$ we knew the error $e = f - u$ and could use principles for minimizing the error
- When solving $\mathcal{L}(u_\mathrm{e}) = 0$ we do not know u_e and cannot work with the error $e = u_\mathrm{e} - u$
- We only have the *error in the equation*: the residual R

Inserting $u = \sum_j c_j\psi_j$ in $\mathcal{L} = 0$ gives a residual

$$R = \mathcal{L}(u) = \mathcal{L}(\sum_j c_j\psi_j) \neq 0 \tag{10}$$

Goal: minimize R wrt $\{c_i\}_{i\in I}$ (and hope it makes a small e too)

$$R = R(c_0, \ldots, c_N; x)$$

The least squares method

Idea: minimize

$$E = ||R||^2 = (R, R) = \int_\Omega R^2 dx \qquad (11)$$

Minimization wrt $\{c_i\}_{i\in I}$ implies

$$\frac{\partial E}{\partial c_i} = \int_\Omega 2R\frac{\partial R}{\partial c_i} dx = 0 \quad \Leftrightarrow \quad (R, \frac{\partial R}{\partial c_i}) = 0, \quad i \in I \qquad (12)$$

$N+1$ equations for $N+1$ unknowns $\{c_i\}_{i\in I}$

The Galerkin method

Idea: make R orthogonal to V,

$$(R, v) = 0, \quad \forall v \in V \tag{13}$$

This implies

$$(R, \psi_i) = 0, \quad i \in I, \tag{14}$$

$N+1$ equations for $N+1$ unknowns $\{c_i\}_{i\in I}$

The Method of Weighted Residuals

Generalization of the Galerkin method: demand R orthogonal to some space W, possibly $W \neq V$:

$$(R, v) = 0, \quad \forall v \in W \tag{15}$$

If $\{w_0, \ldots, w_N\}$ is a basis for W:

$$(R, w_i) = 0, \quad i \in I \tag{16}$$

- $N + 1$ equations for $N + 1$ unknowns $\{c_i\}_{i \in I}$
- Weighted residual with $w_i = \partial R / \partial c_i$ gives least squares

Terminology: test and Trial Functions

- ψ_j used in $\sum_j c_j\psi_j$: *trial function*
- ψ_i or w_i used as weight in Galerkin's method: *test function*

The collocation method

Idea: demand $R = 0$ at $N + 1$ points

$$R(x_i; c_0, \ldots, c_N) = 0, \quad i \in I \tag{17}$$

Note: The collocation method is a weighted residual method with delta functions as weights

$$\text{property of } \delta(x): \quad \int_\Omega f(x)\delta(x - x_i)dx = f(x_i), \quad x_i \in \Omega \tag{18}$$

$$0 = \int_\Omega R(x; c_0, \ldots, c_N)\delta(x - x_i)\,\mathrm{d}x = R(x_i; c_0, \ldots, c_N)$$

Examples on using the principles

Goal.

Exemplify the least squares, Galerkin, and collocation methods in a simple 1D problem with global basis functions.

The first model problem

$$-u''(x) = f(x), \quad x \in \Omega = [0, L], \quad u(0) = 0,\ u(L) = 0 \tag{19}$$

Basis functions:

$$\psi_i(x) = \sin\left((i+1)\pi\frac{x}{L}\right), \quad i \in I \tag{20}$$

The residual:

$$\begin{aligned} R(x; c_0, \ldots, c_N) &= u''(x) + f(x), \\ &= \frac{d^2}{dx^2}\left(\sum_{j\in I} c_j\psi_j(x)\right) + f(x), \\ &= -\sum_{j\in I} c_j\psi_j''(x) + f(x) \end{aligned} \tag{21}$$

Boundary conditions

Since $u(0) = u(L) = 0$ we must ensure that all $\psi_i(0) = \psi_i(L) = 0$. Then

$$u(0) = \sum_j c_j \psi_j(0) = 0, \quad u(L) = \sum_j c_j \psi_j(L)$$

- u known: Dirichlet boundary condition
- u' known: Neumann boundary condition
- Must have $\psi_i = 0$ where Dirichlet conditions apply

The least squares method; principle

$$(R, \frac{\partial R}{\partial c_i}) = 0, \quad i \in I$$

$$\frac{\partial R}{\partial c_i} = \frac{\partial}{\partial c_i}\left(\sum_{j\in I} c_j\psi_j''(x) + f(x)\right) = \psi_i''(x) \tag{22}$$

Because:

$$\frac{\partial}{\partial c_i}\left(c_0\psi_0'' + c_1\psi_1'' + \cdots + c_{i-1}\psi_{i-1}'' + c_i\psi_i'' + c_{i+1}\psi_{i+1}'' + \cdots + c_N\psi_N''\right) =$$

The least squares method; equation system

$$(\sum_j c_j\psi_j'' + f, \psi_i'') = 0, \quad i \in I, \tag{23}$$

Rearrangement:

$$\sum_{j\in I}(\psi_i'', \psi_j'')c_j = -(f, \psi_i''), \quad i \in I \tag{24}$$

This is a linear system

$$\sum_{j\in I} A_{i,j}c_j = b_i, \quad i \in I,$$

with

$$\begin{aligned}
A_{i,j} &= (\psi_i'', \psi_j'')\\
&= \pi^4(i+1)^2(j+1)^2L^{-4}\int_0^L \sin\left((i+1)\pi\frac{x}{L}\right)\sin\left((j+1)\pi\frac{x}{L}\right)\, dx\\
&= \begin{cases} \frac{1}{2}L^{-3}\pi^4(i+1)^4 & i = j\\ 0, & i \neq j \end{cases}
\end{aligned} \tag{25}$$

Orthogonality of the basis functions gives diagonal matrix

Useful property:

$$\int_0^L \sin\left((i+1)\pi\frac{x}{L}\right)\sin\left((j+1)\pi\frac{x}{L}\right)\,dx = \delta ij, \quad \Rightarrow (\psi_i'', \psi_j'') = \delta_{ij}, \quad \delta_{ij} = \tag{27}$$

With diagonal $A_{i,j}$ we can easily solve for c_i:

$$c_i = \frac{2L}{\pi^2(i+1)^2}\int_0^L f(x)\sin\left((i+1)\pi\frac{x}{L}\right)\,dx \tag{28}$$

Least squares method; solution

Let's sympy do the work ($f(x) = 2$):

```
from sympy import *
import sys

i, j = symbols('i j', integer=True)
x, L = symbols('x L')
f = 2
a = 2*L/(pi**2*(i+1)**2)
c_i = a*integrate(f*sin((i+1)*pi*x/L), (x, 0, L))
c_i = simplify(c_i)
print c_i
```

$$c_i = 4\frac{L^2\left((-1)^i + 1\right)}{\pi^3\left(i^3 + 3i^2 + 3i + 1\right)}$$

$$u(x) = \sum_{k=0}^{N/2} \frac{8L^2}{\pi^3(2k+1)^3} \sin\left((2k+1)\pi\frac{x}{L}\right) . \tag{29}$$

- Fast decay: $c_2 = c_0/27$, $c_4 = c_0/125$
- Only one term might be good enough

$$u(x) \approx \frac{8L^2}{\pi^3} \sin\left(\pi\frac{x}{L}\right) .$$

The Galerkin method; principle

$$(u'' + f, v) = 0, \quad \forall v \in V,$$

or

$$(u'', v) = -(f, v), \quad \forall v \in V \tag{30}$$

This is a *variational formulation* of the differential equation problem.
$\forall v \in V$ means for all basis functions:

$$(\sum_{j \in I} c_j \psi_j'', \psi_i) = -(f, \psi_i), \quad i \in I \tag{31}$$

The Galerkin method; solution

Since $\psi_i'' \propto \psi_i$, Galerkin's method gives the same linear system and the same solution as the least squares method (in this particular example).

The collocation method

$R = 0$ or the differential equation must be satisfied at $N+1$ points:

$$-\sum_{j\in I} c_j \psi_j''(x_i) = f(x_i), \quad i \in I \tag{32}$$

This is a linear system $\sum_j A_{i,j} = b_i$ with entries

$$A_{i,j} = -\psi_j''(x_i) = (j+1)^2\pi^2 L^{-2} \sin\left((j+1)\pi\frac{x_i}{L}\right), \quad b_i = 2$$

Choose: $N = 0$, $x_0 = L/2$

$$c_0 = 2L^2/\pi^2$$

Comparison of the methods

- Exact solution: $u(x) = x(L - x)$
- Galerkin or least squares ($N = 0$): $u(x) = 8L^2\pi^{-3}\sin(\pi x/L)$
- Collocation method ($N = 0$): $u(x) = 2L^2\pi^{-2}\sin(\pi x/L)$.
- Max error in Galerkin/least sq.: $-0.008L^2$
- Max error in collocation: $0.047L^2$

Integration by parts

Second-order derivatives will hereafter be integrated by parts

$$\begin{aligned} \int_0^L u''(x)v(x)dx &= -\int_0^L u'(x)v'(x)dx + [vu']_0^L \\ &= -\int_0^L u'(x)v'(x)dx + u'(L)v(L) - u'(0)v(0) \end{aligned} \tag{33}$$

Motivation:

- Lowers the order of derivatives
- Gives more symmetric forms (incl. matrices)
- Enables easy handling of Neumann boundary conditions
- Finite element basis functions φ_i have discontinuous derivatives (at cell boundaries) and are not suited for terms with φ_i''

Boundary function; principles

- What about nonzero Dirichlet conditions?
- E.g. $u(L) = D$

Boundary function; example

$u(0) = C$ and $u(L) = D$. Choose

$$B(x) = L^{-1}(C(L-x) + Dx): \qquad B(0) = C,\ B(L) = D$$

$$u(x) = L^{-1}(C(L-x) + Dx) + \sum_{j\in I} c_j\psi_j(x), \tag{34}$$

$$u(0) = C, \quad u(L) = 0$$

Abstract notation for variational formulations

The finite element literature (and much FEniCS documentation) applies an abstract notation for the variational formulation:

*Find $(u - B) \in V$ such that

$$a(u, v) = L(v) \quad \forall v \in V$$

Example on abstract notation

Given a variational formulation for $-u'' = f$:

$$\int_\Omega u'v'dx = \int_\Omega fvdx \quad \text{or} \quad (u', v') = (f, v) \quad \forall v \in V$$

Abstract formulation: finn $(u - B) \in V$ such that

$$a(u, v) = L(v) \quad \forall v \in V$$

We identify

$$a(u, v) = (u', v'), \quad L(v) = (f, v)$$

Bilinear and linear forms

- $a(u, v)$ is a *bilinear form*
- $L(v)$ is a *linear form*

Linear form means

$$L(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 L(v_1) + \alpha_2 L(v_2),$$

Bilinear form means

$$\begin{aligned} a(\alpha_1 u_1 + \alpha_2 u_2, v) &= \alpha_1 a(u_1, v) + \alpha_2 a(u_2, v), \\ a(u, \alpha_1 v_1 + \alpha_2 v_2) &= \alpha_1 a(u, v_1) + \alpha_2 a(u, v_2) \end{aligned}$$

In nonlinear problems: Find $(u - B) \in V$ such that $F(u; v) = 0 \ \forall v \in V$

The linear system associated with abstract form

$$a(u, v) = L(v) \quad \forall v \in V$$

is equivalent to

$$a(u, \psi_i) = L(\psi_i) \quad i \in I$$

Insert $u = \sum_j c_j\psi_j$ and use linearity:

$$\sum_{j \in I} a(\psi_j, \psi_i)c_j = L(\psi_i) \quad i \in I$$

This is a linear system

$$\sum_{j \in I} A_{i,j}c_j = b_i, \quad i \in I$$

with

$$\begin{aligned} A_{i,j} &= a(\psi_j, \psi_i) \\ b_i &= L(\psi_i) \end{aligned}$$

Equivalence with minimization problem

If $a(u, v) = a(v, u)$,

$$a(u, v) = L(v) \quad \forall v \in V,$$

is equivalent to minimizing the functional

$$F(v) = \frac{1}{2} a(v, v) - L(v)$$

over all functions $v \in V$. That is,

$$F(u) \leq F(v) \quad \forall v \in V.$$

- Much used in the early days of finite elements
- Still much used in structural analysis and elasticity
- Not as general as Galerkin's method (since $a(u, v) = a(v, u)$)

Examples on variational formulations

Goal.

Derive variational formulations for many prototype differential equations in 1D that include

- variable coefficints
- mixed Dirichlet and Neumann conditions
- nonlinear coefficients

Variable coefficient; problem

$$-\frac{d}{dx}\left(\alpha(x)\frac{du}{dx}\right) = f(x), \quad x \in \Omega = [0, L],\ u(0) = C,\ u(L) = D\,. \tag{35}$$

- Variable coefficient $\alpha(x)$
- *Nonzero* Dirichlet conditions at $x = 0$ *and* $x = L$
- Must have $\psi_i(0) = \psi_i(L) = 0$
- $V = \text{span}\{\psi_0, \ldots, \psi_N\}$
- $v \in V$: $v(0) = v(L) = 0$

$$u(x) = B(x) + \sum_{j\in I} c_j\psi_i(x)$$

$$B(x) = C + \frac{1}{L}(D - C)x$$

Variable coefficient; variational formulation (1)

$$R = -\frac{d}{dx}\left(a\frac{du}{dx}\right) - f$$

Galerkin's method:

$$(R, v) = 0, \quad \forall v \in V,$$

or with integrals:

$$\int_\Omega \left(\frac{d}{dx}\left(\alpha\frac{du}{dx}\right) - f\right) v \,\mathrm{d}x = 0, \quad \forall v \in V\,.$$

Variable coefficient; variational formulation (2)

Integration by parts:

$$-\int_\Omega \frac{d}{dx}\left(\alpha(x)\frac{du}{dx}\right) v \,\mathrm{d}x = \int_\Omega \alpha(x)\frac{du}{dx}\frac{dv}{dx}\,\mathrm{d}x - \left[\alpha\frac{du}{dx}v\right]_0^L .$$

Boundary terms vanish since $v(0) = v(L) = 0$

Variational formulation.

Find $(u - B) \in V$ such that

$$\int_\Omega \alpha(x)\frac{du}{dx}\frac{dv}{dx}dx = \int_\Omega f(x)vdx, \quad \forall v \in V,$$

Compact notation:

$$(\alpha u', v') = (f, v), \quad \forall v \in V$$

Variable coefficient; abstract notation

$$a(u, v) = L(v) \quad \forall v \in V,$$

$$a(u, v) = (\alpha u', v'), \quad L(v) = (f, v)$$

Variable coefficient; linear system

$v = \psi_i$ and $u = B + \sum_j c_j\psi_j$:

$$(\alpha B' + \alpha \sum_{j\in I} c_j\psi_j', \psi_i') = (f, \psi_i), \quad i \in I\,.$$

Reorder to form linear system:

$$\sum_{j\in I} (\alpha\psi_j', \psi_i')c_j = (f, \psi_i) + (a(D - C)L^{-1}, \psi_i'), \quad i \in I\,.$$

This is $\sum_j A_{i,j}c_j = b_i$ with

$$A_{i,j} = (a\psi_j', \psi_i') = \int_\Omega \alpha(x)\psi_j'(x), \psi_i'(x)\,\mathrm{d}x,$$

$$b_i = (f, \psi_i) + (a(D - C)L^{-1}, \psi_i') = \int_\Omega \left(f(x)\psi_i(x) + \alpha(x)\frac{D - C}{L}\psi_i'(x)\right.$$

First-order derivative in the equation and boundary condition; problem

[[[