Finite Element Analysis
Kazufumi Ito

1 INTRODUCTION

The finite element method (FEM), is a numerical method for solving partial differential equations that arise in engineering and mathematical physics. Typical problem areas of interest include structural analysis, heat transfer, fluid flow, mass transport, and electromagnetic potential. The analytical solution of these problems generally require the solution to boundary value problems for partial differential equations. The finite element method formulation of the problem results in a system of algebraic equations. The method yields approximate values of the unknowns at discrete number of points over the domain. To solve the problem, it subdivides a large system into smaller, simpler parts that are called finite elements. The simple equations that model these finite elements are then assembled into a larger system of equations that models the entire problem. FEM then uses variational methods from the calculus of variations to approximate a solution by minimizing an associated error function. Studying or analyzing a phenomenon with FEM is often referred to as finite element analysis (FEA). It is our intention that the lecture note is self-contained as much as possible and discusses the most of basic theory and implementation. In Appendix we discuss the sumplemental material and the algorithmic aspects.

We first formulate FEA for the one dimensional equations and motivate various aspects of FEA, i.e., weak form, stability and variational formulation (Euler-Lagrange). The function spaces and the theoretical foundation of FEM is then developed, i.e., Distributions, Weak derivatives, Sobolev spaces, Lax-Milgram and Babusika-Necas-Banach theory for the well-posedness of the weak form of equations.. Specifically, the mixed finite element formulation allows to use discontinuous basis elements and is demonstrated by concrete examples. The basis functions and the assembling of finite event system based on the iso-parametric method are discussed. The error estimate for FEM is analyzed, i.e., including Cea, Aubin-Nitche lemmas. Various applications of FEM is introduced for parabolic, hyperbolic and the related equations. A specific attention is given to the discontinuous Galerkin method.

1.1 What is the finite element method

The following is the major ingredients of the finite element method.

- The solution u is represented by

$$u(x) = \sum_{k=1}^{N} u_k \, \phi_k^N(x),$$

  where $\{\phi_k^N(x)\}$ is TRIAL BASIS (it is a generic expression, i.e., we need to specify the basis for a given example).

- The differential Equation whose unknowns are functions, is tested against TEST function Basis $\{\psi_k^N(x)\}$. i.e. Weak form

$$(E(x, u, u', u''), \psi_k^N) = 0 \text{ for all } 1 \le k \le N.$$

Moreover, one can relax is the smoothness requirement on u and $\{\phi_k^n\}$ by the integration by parts (Green) formula as will be shown.

Example (Two point Boundary value problem)

$$-(a(x)u')' = f(x) \text{ with } u(0) = 0, \quad u(1) = 0. \tag{1.1}$$

where the conductivity $a(x) > 0$. Let $x_k = \frac{k}{N}$ be the k-th node on $[0, 1]$.

$$u^N(x) = \text{the piecewise linear function} \Leftrightarrow \phi_k^N(x) = B_k^N(x) = \text{hat functions},$$

where

$$B_k^N(x) = \begin{cases} 1 - N|x - x_k| & \text{on } |x - x_k| \le h = \frac{1}{N} \\ \\ 0 & \text{otherwise} \end{cases}$$

Note that $u^N(x_k) = u_k,\ 0 \le k \le N$ and $u_0 = u_N = 0$.

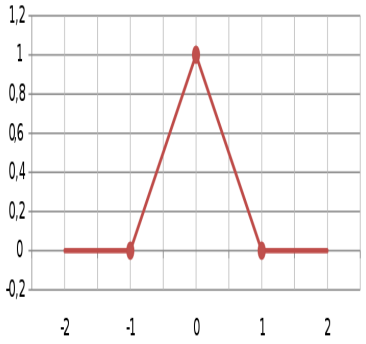

Figure 1: Hat function

<u>WEAK FORMULATION</u> Let $L^2(0, 1)$ be the space of square integrable functions. For all test function $\psi \in H_0^1(0, 1) = \{\psi \in L^2(0, 1) : \psi' \in L^2(0, 1),\ \psi(0) = \psi(1) = 0\}$,

$$\int_0^1 -(a(x)u')'\psi \, dx = \int_0^1 a(x)u'\psi' \, dx = \int f(x)\psi(x) \, dx. \tag{1.2}$$

is the weak formulation of the differential form (1.1).

For example, $a(x) = 1$ and $\psi = B_k^N(x)$ we have

$$\int_0^1 u'\psi' \, dx = N\Big(\int_{x_{k-1}}^{x_k} u' \, dx + \int_{x_k}^{x_{k+1}} u' \, dx = -N(u(x_{k+1}) - 2u(x_k) + u(x_{k-1})) = \int_0^1 f(x)\psi(x) \, dx.$$

That is the central difference approximation

$$-\frac{u_{k+1} - 2u_k + u_{k-1}}{h^2} = N \int_{x_{k-1}}^{x_{k+1}} f(x)B_k^N(x) \, dx.$$

is EXACT if we let $u_k = u(x_k)$ and evaluate

$$f_k = \int_{x_{k-1}}^{x_{k+1}} f(x) B_k^N(x)(x)\, dx.$$

Piecewise Linear Finite element method (Galerkin method) The standard Galerkin method $\phi_k^N(x) = \psi_k^N(x)$ reduces to the Galerkin system of equations for $\{u_k^N\},\ 1 \le k \le N-1$:

$$\int_0^1 a(x)(\sum u_k B_k^N(x)')B_j^N(x)'\, dx = \int_0^1 f(x) B_k^N(x)\, dx, \tag{1.3}$$

i.e.,

$$Hu^N = f^N$$

where the stiffness matrix $H \in R^{(N-1)\times(N-1)}$ tridiagonal and given by

$$H_{k,j} = \int_0^1 a(x) B_k^N(x)' B_j^N(x)'\, dx = N \begin{cases} a_k + a_{k-1} & k = j \\ -a_j & j = k \pm 1 \\ 0 & otherwise \end{cases} \tag{1.4}$$

where $a_k = \int_{x_k}^{x_{k+1}} a(x)\, dx$, and the right hand side is approximate by

$$f_k^N = \int_{x_{k-1}}^{x_{k+1}} f(x)\, dx \sim \frac{1}{N} f(x_k) \tag{1.5}$$

That is, for $h = 1/N$

$$-\frac{a_k \frac{u_{k+1}-u_k}{h} - a_{k-1}\frac{u_k - u_{k-1}}{h}}{h} = f_k$$

which is equivalent to the central difference approximation of $-(au')' = f$.

In general, the finite element method is characterized by the following process.

- One chooses a grid for domain Ω where the unknown u is defined In the preceding treatment, the grid consisted of triangles, but one can also use squares or curvilinear polygons.
- Then, one chooses basis functions. In our discussion, we used piecewise linear basis functions, but it is also common to use piecewise polynomial basis functions.

1.2 Where is it from?

Consider the variational problem

$$\min \quad J(u) = \int_0^1 \frac{1}{2} a(x)|u'(x)|^2 - f(x)u(x)\, dx$$

over $u \in H_0^1(0,1) = \{u \in L^2(0,1) : u \in L^2(0,1),\ u(0) = u(1) = 0\}$, where

$$L^2(0,1) = \text{the space of square integrable functions on } (0,1).$$

The necessary optimality condition (equation $J'(u) = 0$) is given by

$$\int_0^1 a(x)u'\psi'\,dx = \int f(x)\psi(x)\,dx.$$

for all $\psi \in H_0^1(0,1) = \{\psi' \in L^2(0,1),\ \psi(0) = \psi(1) = 0\}$. In fact,

$$J(u + t\,\psi) - J(u) = \int_0^1 t(a(s)u'(x)\psi'(x) - f(x)\psi(x))\,dx + \frac{t^2}{2}\int_0^1 a(x)|\psi'|^2\,dx$$

$$\lim_{t\to 0}\frac{J(u + t\,\psi) - J(u)}{t} = (J'(u), \psi) = \int_0^1 t(a(s)u'(x)\psi'(x) - f(x)\psi(x))\,dx = 0$$

which is the weak form (1.2). Since

$$(J'(u), \psi) = \int_0^1 (-(a(x)u'(x))' - f(x))\psi(x)\,dx = 0 \text{ for all } \psi \in H_0^1(0,1).$$

Thus, we obtain the strong (differential) form (1.1).

$$-(a(x)u'(x))' - f(x) = 0 \text{ with } u(0) = 0,\ \ u(1) = 0.$$

1.3 Ritz method

Consider the Ritz method:

$$\min \quad J(u) \text{ subject to } u = \sum_{k=1}^{N} u_k\phi_k^N(x).$$

Then, we have

$$\frac{\partial J}{\partial u_k} = (J'(u), \phi_k^N) = 0$$

which is equivalent to Galerkin system (1.3). That is, the the Ritz method is equivalent to the Galerkin method.

1.4 Examples

Consider the minimization

$$\min \quad J(u) = \int_0^1 \frac{1}{2}(a(x)|u'(x)|^2 + c(x)|u(x)|^2) - f(x)u(x)\,dx$$

over all $u \in H^1(0,1) = \{u \in L^2(0,1) : u' \in L^2(0,1)\}$. Then the minimizer $u \in H^1(0,1)$ satisfies

$$\int_0^1 (a(x)u'(x)\psi'(x) + c(x)u(x)\psi(x) - f(x))\,dx = 0$$

for all $\psi \in H^1(0,1)$. Since

$$\int_0^1 a(x)u'(s)\psi'(x)\,dx = a(x)u'(x)\psi(x)|_{x=0}^{x=1} - \int_0^1 (a(x)u')'\psi(x)\,dx,$$

we have

$$\int_0^1 (-(a(x)u'(x))' + c(x)u(x) - f(x))\psi(x)\,dx + a(x)u'(x)\psi(x)|_{x=0}^{x=1} = 0.$$

Thus,

$$-(a(x)u'(x))'(x) + c(x)u(x) - f(x) = 0$$

with $a(1)u'(1) = 0$ and $a(0)u'(0) = 0$.

Next, consider the minimization

$$\min \quad J(u) = \int_0^1 \frac{1}{2}(a(x)|u'(x)|^2 + c(x)|u(x)|^2) - f(x)u(x)\,dx + \frac{\alpha}{2}|u(0)|^2$$

over $u \in H_R^1(0,1) = \{u' \in L^2(0,1) : u(1) = 0\}$. Then the minimizer $u \in H_R^1(0,1)$ satisfies

$$\int_0^1 (a(x)u\prime(x)\psi'(x) + c(x)u(x)\psi(x) - f(x))\,dx + \alpha\, u(0)\psi(0) = 0$$

for all $\psi \in H_R^!(0,1)$. Since

$$\int_0^1 a(x)u'(x)\psi'(x)\,dx = a(x)u'(x)\psi(x)|_{x=0}^{x=1} - \int_0^1 (a(x)u')'\psi(x)\,dx$$

we have

$$\int_0^1 (-(a(x)u'(x))'(x) + c(x)u(x) - f(x))\psi(x)\,dx + (-a(0)u'(0) + \alpha\, u(0))\psi(0) = 0.$$

First, for all $\psi \in H_0^1(0,1)$

$$\int_0^1 (-(a(x)u'(x))' + c(x)u(x) - f(x))\psi(x)\,dx = 0$$

Since $H_0^1(0,1)$ is dense in $L^2(0,1)$, i.e., for given $u \in L^2(0,1)$ there exits a sequence in $u_n \in H_0^1(0,1)$ such that $\int_0^1 |u_n(x) - u(x)|^2\,dx \to 0$ as $n \to \infty$. we have

$$-(a(x)u'(x))' + c(x)u(x) - f(x) = 0$$

Next, by selecting $\psi(0)$ arbitrary we have

$$-a(0)u'(0) + \alpha\, u(0) = 0 \text{ and } u(1) = 0.$$

Exercise 1 In general, consider the minimization

$$\min \quad J(u) = \int_0^1 \frac{1}{2}(a(x)|u'(x)|^2 + c(x)|u(x)|^2) - f(x)u(x)\,dx + \frac{\alpha}{2}|u(0)|^2 - u(0)f_1 + \frac{\beta}{2}|u(1)|^2 - u(1)f_2$$

over all $u \in H^1(0,1) = \{u' \in L^2(0,1)$. Show that the minimizer $u \in H^1(0,1)$ satisfies

$$\int_0^1 (a(x)u\prime(x)\psi'(x) + c(x)u(x)\psi(x) - f(x))\,dx + (\alpha\, u(0) - f_1)\psi(0) + (\beta\, u(1) - f_2)\psi(1) = 0.$$

for all $\psi \in H^1(0,1)$. The differential (strong) form is given by

$$-(a(x)u'(x))' + c(x)u(x) - f(x) = 0$$

with $-a(0)u'(0) + \alpha\, u(0) = f_1$ and $a(1)u'(1) + \beta\, u(1) = f_2$.

Thus, Galerkin system is given by

$$u^N(x) = \sum_{k=0}^{N} u_k\, B_k^N(x)$$

and $Hu^N = f^N$ with $u^N,\ f^N \in R^{N+1}$. $H \in R^{(N+1)\times(N+1)}$ is given by

$$H_{jk} = \textstyle\int_{x_{k-1}}^{x_{k+1}} (a(x)B_k^N(x)'B_j^N(x)'\,dx + c(x)B_k^N(x)B_j^N(x))\,dx$$

$$H_{00} = \textstyle\int_0^{x_1} (a(x)B_0^N(x)'B_0^N(x)'\,dx + c(x)B_0^N(x)B_0^N(x))\,dx + \alpha$$

$$H_{NN} = \textstyle\int_{X_{N-1}}^1 (a(x)B_N^N(x)'B_N^N(x)'\,dx + c(x)B_N^N(x)B_N^N(x))\,dx + \beta$$

and $f^N \in R^{N+1}$ is given by

$$f_k^N = \textstyle\int_{x_{k-1}}^{x_{k+1}} f(x)B_k^N(x)\,dx$$

$$f_0^N = \textstyle\int_0^{x_1} f(x)B_0^N(x)\,dx + f_1$$

$$f_N^N = \textstyle\int_{x_{N-1}}^1 f(x)B_N^N(x)\,dx + f_2.$$

Remark: One can approximate

$$\int_0^1 c(x)B_k^N(x)B_j^N(x))\,dx \sim \frac{c(x_k)}{N}\delta_{k,j} \text{ for } 0 < k,\ j < N$$

without loosing accuracy. Or on can use the Gauss quadrature rule $(w_i, \xi_i),\ 1 \le i \le m$:

Gauss Quadrature rule:

$$\int_{x_k}^{x_{k-1}} F(x)\,dx \sim \frac{x_k - x_{k-1}}{2} \sum_{i=1}^{m} w_i F(x_{k-1} + \frac{x_k - x_{k-1}}{2}(\xi_i + 1))$$

which is exact if F is a polynomial of degree less than $2m-1$. For example

$$\begin{cases} m = 1,\ \xi_1 = 0,\ w_1 = 2 \\ \\ m = 2,\ \xi_1 = -\frac{1}{\sqrt{3}},\ \xi_2 = \frac{1}{\sqrt{3}},\ w_1 = w_2 = 1 \\ \\ m = 3;\ \xi_1 = -\sqrt{\frac{3}{5}},\ \xi_2 = 0,\ \xi_3 = \sqrt{\frac{3}{5}},\ w_1 = w_3 = \frac{5}{9},\ w_2 = \frac{8}{9}. \end{cases}$$

Exercise 2 (Discontinuos $a(x)$)

$$a(x) = 1 \text{ on } [0, \frac{1}{2}) \text{ and } 10 \text{ on } (\frac{1}{2}, 1].$$

Show that $a(x)u'$ is continuous since $(au')' \in L^2(0,1)$ and thus $[a(x)u'] = 0$ at $x = \frac{1}{2}$ and $u'(x)$ is discontinuous at $x = \frac{1}{2}$. Or, since

$$-\int_0^1 (a(x)u')'\psi \, dx = ((au')((\frac{1}{2})^+) - (au')((\frac{1}{2})^-))\psi(\frac{1}{2}) + \int_0^1 a(x)u'\psi' \text{ for } \psi \in H_0^1(0,1),$$

we have $(au')((\frac{1}{2})^+) = (au')((\frac{1}{2})^-)$.

Matlab code with $N = 50$

```
d=[2*ones(24,1);11;20*ones(24,1)]; dd=[ones(24,1); 10*ones(24,1)];
h=spdiags(-dd,-1,49,49);h=h+h'+spdiags(d,0,49,49); h=50^2*h; u=h\ones(49,1);
plot(u)
```

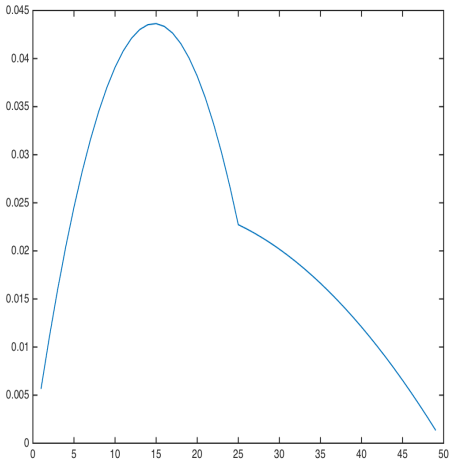

Figure 2: Numerical Example 1

Exercise 3 Consider the nonlinear case

$$\min \quad J(u) = \int_0^! \Psi(x, u(x), u'(x)) \text{ subject to } u(0) = 0, \tag{1.6}$$

where $p = u'$ defines the momentum (strain) of u. For example

$$\Psi(x, u, p) = (\frac{1}{q}|p|^q + \frac{c(x)}{4}(u(x)^2 - 1)^2 - f(x)u(x))\, dx \tag{1.7}$$

Then,

$$(J'(u), \psi) = \lim_{t\to 0} \frac{J(u + t\,\psi) - J(u)}{t} = \int_0^1 (\Psi_p(x, u, u')\psi'(x) + \Psi_u(x, u, u')\psi(x) - f(x)\psi(x))\, dx$$

where

$$\Psi_p = \frac{\partial}{\partial p}\Psi \text{ and } \Psi_u = \frac{\partial}{\partial u}\Psi$$

are the partial derivatives of the function $(x, u, p) \to \Psi(x, u, p) \in R$. Thus, a minimizer of (1.6) satisfies

$$(J'(u^*), \psi) = 0 \quad \text{for all } \psi \in C^1(0, 1) \text{ satisfying } \psi(0) = 0$$

The strong form is given by

$$-(\Psi_p(x, u, u'))' + \Psi_u(x, u, u') = 0$$

with

$$u(0) = 0, \quad \Psi_p(1, u(1), u'(1)) = 0.$$

For the example (1.7), we have

$$\Psi_p = |p|^{q-2}p, \;\; \Psi_u = c(x)\, u(u^2 - 1) - f(x)u$$

and

$$-(|u'(x)|^{q-2}u')' + c(x)\, u(u^2 - 1) - f(x) = 0,\; u(0) = 0 \text{ and } u'(1) = 0$$

The finite element method based on the linear element is given by

$$-\frac{\Psi_p(x_{k+\frac{1}{2}}, \frac{u_{k+1}-u_k}{h}) - \Psi_p(x_{k-\frac{1}{2}}, \frac{u_k-u_{k-1}}{h})}{h} + \Psi_u(x_k, u_k) = 0.$$

where we used the midpoint rule

$$\int_{x_{k-1}}^{x_{k+!}} F(x)B_k^N(x)\, dx \sim \frac{1}{N}F(x_k).$$

1.5 4th order equation (Beam equation)

Consider the 4th order equation

$$u'''' + c(x)u = f(x)$$

with various boundary conditions. Let us start with

$$u(0) = u'(0) = 0, \;\; u(1) = u'(1) = 0$$

Since

$$\int_0^1 u''''\psi\, dx = (u'''\psi - u''\psi')|_{x=0}^{x=1} + \int_0^1 u''\psi''\, dx. \tag{1.8}$$

The weak form is given by

$$\int_0^1 (u''(x)\psi''(x) + c(x)u(x) - f\psi)\, dx = 0$$

for all $\psi \in H_0^2(0,1) = \{u \in L^2(0,1) : u,\; u' \in L^2(0,1),\;\; u(0) = u'(0) = 0,\;\; u(1) = u'(1) = 0\}$. The corresponding variational problem is

$$\min \quad J(u) = \int_0^1 \frac{1}{2}(|u''(x)|^2 + c(x)u(x)|^2) - f(x)u(x)\, dx$$

over $u \in H_0^2(0,1)$.

Exercise 4 Consider the general case

$$u'''(0) + \alpha\, u(0) = f_1, \;\; -u''(0) + \beta\, u'(0) = f_2, \;\; u(1) = u'(1) = 0$$

The weak form is given by

$$\int_0^1 (u''(x)\psi''(x) + c(x)u(x) - f\psi)\,dx + (\alpha u(0) - f_1)\psi(0) + (\beta u'(0) - f_2)\psi'(0) = 0$$

for all $\psi \in H_R^2(0,1) = \{u \in L^2(0,1) : u,\ u',\ u'' \in L^2(0,1),\quad u(1) = u'(1) = 0\}$. The corresponding variational problem is

$$\min \quad J(u) = \int_0^1 \frac{1}{2}(|u''(x)|^2 + c(x)|u(x)|^2) - f(x)u(x)\,dx + \frac{\alpha}{2}|u(0)|^2 - u(0)f_1 + \frac{\beta}{2}|u'(0)|^2 - u'(0)f_2 = 0$$

over $u \in H_R^2(0,1)$. Set up the finite element system based on cubic B spline elements by computing the local stiffness matrix $\Phi \in R^{4\times 4}$:

$$\Phi_{i,j} = \int_0^1 \phi_i''\phi_j''\,dx$$

of local elements

$\phi_1 = x^3,\ \phi_2 = 1+3x+3x^2-3x^3,\ \phi_3 = 1+3(1-x)+3(1-x)^2-3(1-x)^3,\ \phi_4 = (1-x)^3$ on $(0,1)$.

1.6 Algorithm and Implement

The most attractive feature of the FEM is its ability to handle complicated geometries (and boundaries) with relative ease. One can use quadrature rules to evaluate H and f (1.4)-(1.5). H is block-diagonal and sparse. But the linear system for $u = u^N$:

$$u = H^{-1}f$$

is possibly large scale (2-3 Dim and system). In general the condition number of H:

$$cond(H) = \frac{\text{max of eigenvalues of } H}{\text{min of eigenvalues of } H}$$

is large. Thus, we use a pre-conditioner P and solve the pre-conditioned equation

$$PHu = Pf$$

by CG (conjugate gradient method) and GMRES (generalized method for Residual method).

1.7 Cubic B spline

Define the space

$$S_3 = \{u \in C^2(0,1) \cap H^3(0,1) : u \in P_3(\text{ cubic polynomial on })[x_{k-1}, x_k]\}$$

where $C^2(0,1)$ is the space of twice continuously differentiable functions on $[0,1]$ and

$$H^3(0,1) = \{u,\ ,u',\ u''.\ u''' \in L^2(0,1)\}$$

The cubic B-spline is defined by

$$B_k(x) = \begin{cases} h^{-3}g_1(x - x_{k-1} & [x_{k-2}, x_{k-1}] \\ g_2(\frac{x-x_{k-1}}{h}) & [x_{k-1}, x_k] \\ g_2(\frac{x_{k+1}-x}{h}) & [x_k, x_{k+1}] \\ h^{-3}g_1(x_{k+2} - x) & [x_{k+1}, x_{k+2}] \\ 0 & \text{otherwise} \end{cases}$$

where $g_1(x) = x^3,\ g_2(x) = 1 + 3x + 3x^2 - 3x^3$. The stiffness matrix H is pentadiagonal. For H_0^1 conformal element we use

$$B_0^D = B_0 - 4B_{-1},\ \ B_1^D = B_1 - B_{-1},\ \ B_N^D = B_N - 4B_{N+1},\ \ B_{N-1}^D = B_{N-1} - B_{N+1}$$

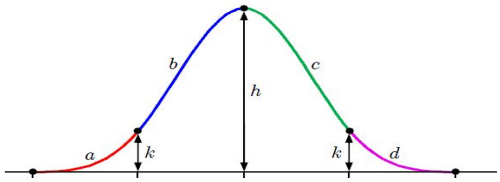

Figure 3: Cubic B spline

1.8 Cubic Hermite spline

Define the space

$$S_2 = \{u \in C^1(0,1) \cap H^2(0,1) : u \in P_3(\text{ cubic polynomial on }) [x_{k-1}, x_k]\}$$

$$u^N(x) = \sum_{k=1}^{N-1} \alpha_k\, \phi_k^N(x) + \sum_{k=1}^{N-1} \beta_k\, \bar{\phi}_k^N(x) \tag{1.9}$$

where we have the nodal property

$$u^N(x_k) = \alpha_k \text{ and } (u^N)'(x_k) = \beta_k.$$

and the two Hermite cubic elements are given by

$$\phi_k^N(x) = \begin{cases} \dfrac{(x - x_{k-1})^2}{(x_k - x_{k-1})^2}\left(\dfrac{2}{x_{k-1} - x_k}(x - x_k) + 1\right) & x \in [x_{k-1}, x_k] \\ \\ \dfrac{(x - x_{k+1})^2}{(x_k - x_{k+1})^2}\left(\dfrac{2}{x_{k+1} - x_k}(x - x_k) + 1\right) & z \in [x_k, x_{k+1}] \end{cases} \tag{1.10}$$

and

$$\bar{\phi}_k^N(x) = \begin{cases} \dfrac{(x-x_k)(x-x_{k-1})^2}{(x_k-x_{k-1})^2} & x \in [x_{k-1}, x_k] \\ \dfrac{(x-x_k)(x-x_{k+1})^2}{(x_k-x_{k+1})^2} & z \in [x_k, x_{k+1}]. \end{cases} \tag{1.11}$$

Recall that $\{\phi_k^N(x)\}$ and $\{\bar{\phi}_k^N(x)\}$ are independent basis functions and $\{\alpha_k\}$ and β_k satisfy the system of finite element equation.

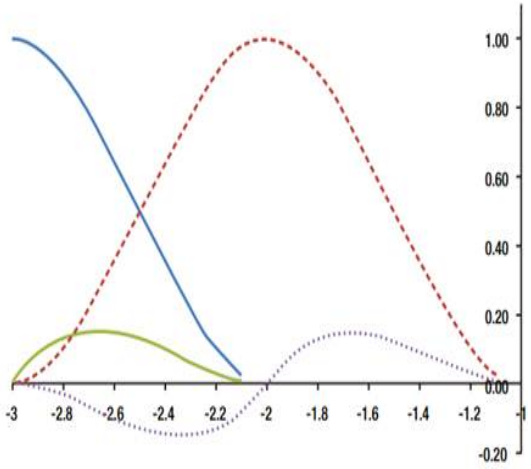

Figure 5-6. Hermite cubic basis functions at nodes (-3, -2)

Figure 4: Hermite Cubic element

1.9 Piecewise Quadratic elements

$$Q = \{u \in C(0,1) \cap H^1(0,1) : u \in P_2(\text{ quadratic polynomial on }) [x_{k-1}, x_k]\}$$

Let z_{2i+1} is the midpoint of the interval

$$z_{2k+1} = \frac{x_k + x_{k+1}}{2} \text{ and } z_{2k} = x_k$$

Then,

$$\phi_{2k}^N(z) = \begin{cases} \dfrac{(z-z_{2k-1})(z-z_{2k-2})}{(z_{2k}-z_{2k-1})(z_{2k}-z_{2k-2)})} & z \in [x_{k-1}, x_k] \\ \dfrac{(z-z_{2k+1})(z-z_{2k+2})}{(z_{2k}-z_{2k+1})(z_{2k}-z_{2k+2)})} & z \in [x_k, x_{k+1}] \end{cases}$$

and

$$\phi_{2k+1}^N(z) = \frac{(z-z_{2k})(z-z_{2(k+1)})}{(z_{2k+1}-z_{2k})(z_{2k+1}-z_{2(k+1)}} \quad z \in [x_k, x_{k+1}]$$

They are zero, otherwise. Then,

$$u^N(z) = \sum_{k=1}^{2N-1} \alpha_k \, \phi_k^N(z)$$

where

$$u^N(z_k) = \alpha_k.$$

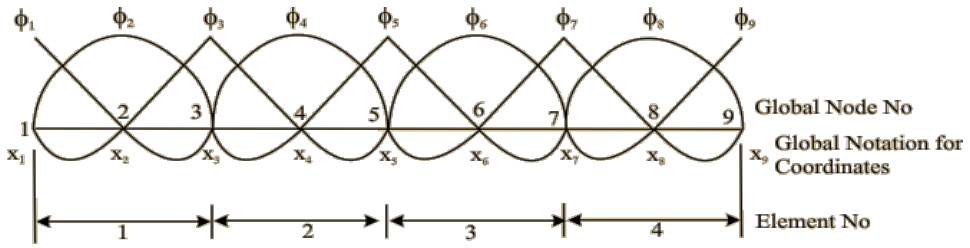

Figure 5: Quadratic spline

1.10 Variational method for constructing finite elements

The variational method can be used to construct best elements. For example consider the following constrained minimizations:

$$\min \quad J(u) = \int_0^1 \frac{1}{2} \, |u'|^2 \, dx \text{ subject to } u(0) = a_0, \;\; u(1) = a_1 \tag{1.12}$$

$$\min \quad J(u) = \int_0^1 \frac{1}{2} \, |u''|^2 \, dx \text{ subject to } u(0) = a_0, \; u'(0) = b_0, \;\; u(1) = a_1, \;\; u'(1) = b_1. \tag{1.13}$$

$$\min \quad J(u) = \int_0^1 \frac{1}{2} \, |u''|^2 \, dx \text{ subject to } u(0) = a_0, \;\; u(\frac{1}{3}) = a_1, \;\; u(\frac{2}{3}) = a_2, \;\; u(1) = a_3 \tag{1.14}$$

It can be shown that

$$u(x) = a_0\, B_0^1(x) + a_1\, B_1^1(x),$$

is the solution to (1.12), where $\{B_k^1(x)\},\ ; k = 0,\ 1$ $(N = 1)$ is the linear elements and

$$u(x) = \sum_{k=0}^{1} a_k\, \phi_k^1(x) + b_k\, \bar{\phi}_k^1(x)$$

is the solution to (1.13) where $\{\phi_k^1(x)\}$ and $\{\bar{\phi}_k^1(x)\}$ for $k = 0,\ 1$ $(N = 1)$ are the Hermite cubic elements.

Exercise 5 For the second problem (1.13) define the Lagrangian functional

$$L(u, \lambda, \mu) = J(u) + (\lambda_0, u(0) - a_0) + (\lambda_1, u(0) - a_1) + (\mu_0, u'(0) - b_0) + (\mu_1, u'(1) - b_1).$$

where λ_k and μ_k for $k = 0,\ 1$ is the Lagrange multiplier associated with the constraints. Then, it follows from the Lagrange multiplier theory [Ito] that for all $\psi \in H^2(0, 1)$

$$\frac{\partial}{\partial u} L(u, \lambda, \mu)(\psi) = \int_0^1 u''\psi''\, dx + \lambda_0\psi(0) + \lambda_1\psi(1) + \mu_0\psi'(0) + \mu_1\psi'(1) = 0.$$

It follows from (1.8) that the strong form is given by

$$u'''' = 0$$

with

$$u'''(0) = -\lambda_0,\ \ u'''(1) = \lambda_1,\ \ u''(0) = \mu_0,\ \ u''(1) = -\mu_1.$$

Thus, $u(x)$ is the cubic polynomial and given by the cubic Hermite polynomials.

For the third problem (1.14) for $x_k = \frac{k}{3},\ 0 \le k \le 3$

$$L(u, \lambda) = J(u) + \sum_{k=0}^{3} \lambda_k\, (u(x_k) - a_k).$$

The Lagrange theory gives

$$\frac{\partial}{\partial u} L(u, \lambda, \mu)(\psi) = \int_0^1 u''\psi''\, dx + \lambda_0\psi(0) + \lambda_1\psi(\frac{1}{3}) + \lambda_2\psi(\frac{2}{3}) + \lambda_3\psi(1) = 0.$$

By the integration by parts on each subinterval (x_{k-1}, x_k)

$$\int_{x_{k-1}}^{x_k} u''\psi''\, dx = u''\psi' - u'''\psi|_{x=x_{k-1}}^{x=x_k} + \int_{x_{k-1}}^{x_k} u''''\psi\, dx$$

we have

$$(u'''(0) + \lambda_0)\psi(0) + (-u'''((\tfrac{1}{3})^-) - u'''((\tfrac{1}{3})^+) + \lambda_1)\psi(\tfrac{1}{3})$$

$$+(-u'''((\tfrac{2}{3})^-) - u'''((\tfrac{2}{3})^+) + \lambda_2)\psi(\tfrac{2}{3}) + (-u'''(1) + \lambda_3)\psi(1) + \textstyle\int_0^1 u''''\psi\, dx = 0$$

for all $\psi \in H^2(0,1) \cup C^1(0,1)$ satisfying $\psi'(x_k) = 0$. Thus, we obtain the strong form

$$u'''' = 0$$

with

$$u'''(0) = -\lambda_0, \;\; [u'''(\frac{1}{3})] = \lambda_1, \;\; [u'''(\frac{2}{3})] = \lambda_2, \;\; u'''(1) = \lambda_3$$

Thus, $u(x)$ is the piecewise cubic polynomial on $(0,1)$ and

$$u(x) = \sum_{k=0}^{N} u_k \, \phi_k^N(x) \;\; (N = 3)$$

where $\{\phi_k^N\}$ is the cubic B-spline and the coefficients $\{u_k\}$ are determined by the constraints.

Lagrange multiplier theory [Ito] Consider the constrained minimization

$$\min J(u) \text{ subject to } E(u) = 0$$

Define the Lagrangian functional

$$L(u, \lambda) = J(u) + (\lambda, E(u))$$

Under an appropriate condition there exists a Lagrange multiplier λ such that

$$\frac{\partial}{\partial u} L(u, \lambda)(\psi) = (J'(u)(\psi) + (\lambda, E'(u)(\psi)) = 0$$

for all directions $\psi \in X$.

For the case of only one constraint and only two variables $(x, y) \in R^2 = X$ as depicted in Figure 6, consider the optimization problem

$$\min \quad f(x,y) \text{ subject to } g(x,y) = c \Rightarrow L(x, y, \lambda) = f(x,y) + \lambda \, (g(x,y) - c).$$

2 Theoretical foundation

In this section we develop the variational principle for the finite element analysis.

2.1 Function spaces

In this section we introduce function spaces defined on a domain=subset (sufficiently smooth) in R^n Let $X = C(\Omega)$ is the space of continuous functions on domain. Then, X is a vector space since X is a vector space if

$$(\alpha \, f_1 + \beta \, f_2)(x) = \alpha \, f_1(x) + \beta \, f_2(x) \in X \tag{2.1}$$

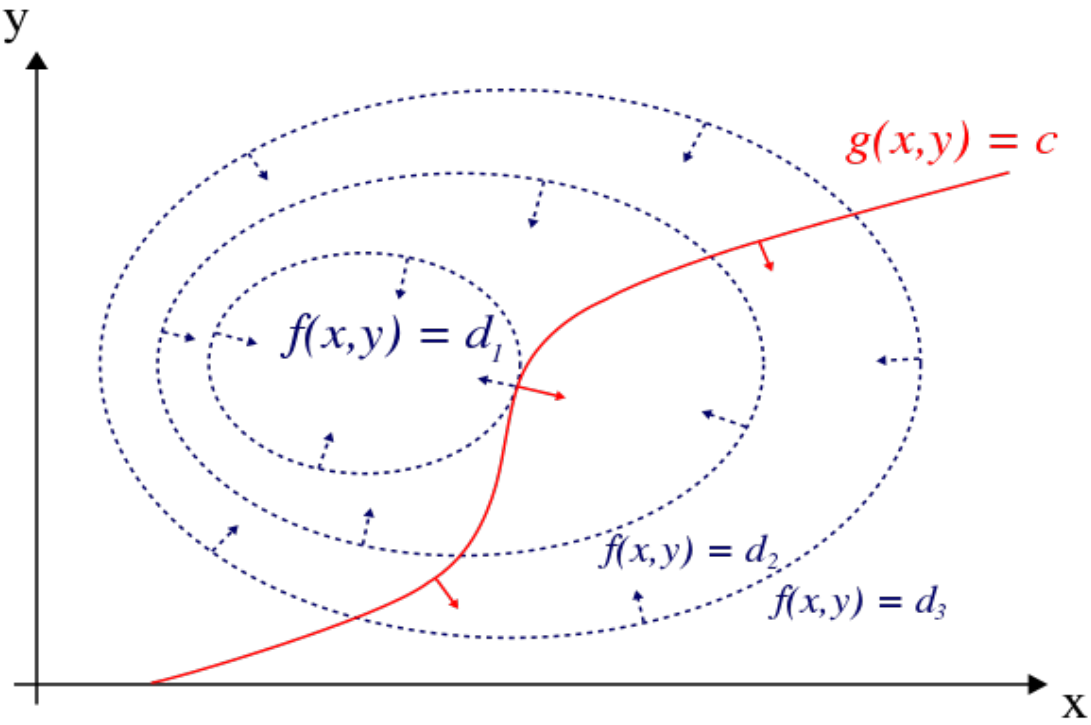

Figure 6: Geometrical proof of the Lagrange theory: The red curve shows the constraint $g(x,y) = c$ The blue curves are contours of $f(x,y)$. The point where the red constraint tangentially touches a blue contour is the maximum of $f(x,y)$ along the constraint, since $d_1 < d_2$. That is $\nabla L = \nabla f + \lambda \nabla g = 0$ at the point.

for all all $f_1,\ f_2 \in X$ and $\alpha,\ \beta \in R$. Define L^p norm on $X = C(\Omega)$ by

$$|f|_{L^p} = (\int_\Omega |f(x)|^p\, dx)^{\frac{1}{p}}. \tag{2.2}$$

In general $(X, ||_X)$ is the vector space X with equipped by norm $||_X)$.

Definition (Normed space $(X, ||_X)$

(1) A set X is a vector space if for all all $f_1,\ f_2 \in X$ and $\alpha,\ \beta \in R$.

$$\alpha\, f_1 + \beta\, f_2 \in X. \tag{2.3}$$

(2) A normed space is a vector space equipped with norm $x \in X \to |x|_X \in R^+$ satisfying

- $|f|_X = 0$ if and only if $f = 0$,
- $|cf|_X = |c||f|_X$ for all $c \in R$ and $f \in X$
- the triangle inequality
$$|f_1 + f_2|_X \leq |f_1|_X + |f_2|_X$$
holds for all $f_1,\ f_2 \in X$

(3) A vector space X is pre-Hilbert (Inner product) space equipped with an inner product $(x_1, x_2) \in X \times X \to R$ satisfying

$$(\alpha\, x_1 + \beta\, x_2, y) = \alpha\,(x_1, y) + \beta\,(x_2, y)$$

$$(y, \alpha\, x_1 + \beta\, x_2) = \alpha\,(x_1, y) + \beta\,(x_2, y)$$

$$(x_1, x_2) = (x_2, x_1)$$

$$(x, x) \geq 0 \text{ and } (x, x) = 0 \text{ if and only if } x = 0$$

Define $|x|_X = \sqrt{(x, x)}$. Then, X is a normed space. In fact,

$$|f + g|_X^2 = (f, f)_X + 2(f, g)_X + (g, g)_X \leq |f|_X^2 + 2|f|_X|g|_X + |g|_X^2 = (|f|_X + |g|_X)^2$$

where we used Cauchy-Schwarz inequality $|(f, g)|_X| \leq |f|_X|g|_X$. Since

$$|f + tg|_X^2 = |f|^2 + 2t\,(f, g)_X + t^2\,|g|_X^2 \text{ for all } t \in R,$$

thus $|(f, g)_X|^2 \leq |f|_X^2|g|_X^2$.

Let $f(x) = f(x_1, \cdot, x_n)$ is an n-dimensional function and Ω is a subset of R^n. Let $C(\Omega)$ =the space of continuous functions on Ω. For $p \geq 1$ define $p-$norm on X by

$$|f|_{L^p} = (\int_\Omega |f(x)|^p\, dx|^{\frac{1}{p}}$$

If $p = 2$, $X = C(\Omega)$ is a pre-Hilbert space with the inner product

$$(f, g)_{L^2} = \int_\Omega f(x)g(x)\, dx$$

Definition ((Cauchy sequence and Banach space If a sequence f_n is a normed space X is called a Cauchy sequence if $|f_n - f_m|_X \to 0$ as $m \geq n \to \infty$. The if every Coach sequence has the limit f in X, the X is complete and X is a Banach space.

Example The rational numbers $\{Q, ||\}$ is not complete. Consider $\sqrt{2} = 1.41..$ Let $a_k \leq b_k$ be rational number pairs with $|a_k - b_k| \leq 10^{-k}$. for example $a_0 = 1,\ b_0 = 2,\ a_1 = 1.4,\ b_1 = 1.5,$ $a_0 = 1.41,\ b_0 = 1.42$, Then $\{a_k\}$ and $\{b_k\}$ are Cauchy sequence in Q (i.e., $|a_n - a_m| \leq 10^{-n}$ and $|b_n - b_m| \leq 10^{-n}$). But the candidate limit $\sqrt{2} \notin Q$ and $(Q, ||)$ is not complete. Also, $|a_n - b_n| \leq i^{-n} \to 0$ and the Cauchy sequences $\{a_n\}$ and $\{b_n\}$ are in an equivalent class, say $\sqrt{2}$ of Cauchy sequences in Q. If we add the all irrational numbers to Q, then it becomes the real line R, which is complete. Thus, R is the completion of $(Q, ||)$.

In general two Cauchy sequences $\{f_k\}$ and $\{g_k\}$ in $(X, ||_X)$ are said to be in a equivalent class if $|f_n - g_n|_X \to 0$ as $n \to \infty$. Let $\bar{X}$ =the completion of X be all equivalent classes of Cauchy sequences in $(X, ||_X)$ with norm

$$|f|_{\bar{X}} = \lim_{k\to\infty} |f_k|$$

for $f \in \bar{X}$. Then $\bar{X}$ is a Banach space [Ito].

Then, for $p \geq 1$

$$L^p(\Omega) = \{ \int_\Omega |f(x)|^p \, dx < \infty \}$$

is the completion of $C(\Omega)$ with respect to L^p norm by the equivalent classes of Cauchy sequences in $C(\Omega)$ with $L^p(\Omega)$ norm:

$$|f|_{L^p} = (\int_\Omega |f|^p \, dx)^{\frac{1}{p}}.$$

The complete pre-Hilbert space is called Hilbert space, i.e., $L^2(\Omega)$ is a Hilbert space.
Example Let $X = C(0,1)$ with L^1 norm. Consider function sequences

$$f_n(x) = \begin{cases} 0 & \text{on } (0, \frac{1}{2}) \\ n(x - \frac{1}{2}) & \text{on } (\frac{1}{2}, \frac{1}{2} + \frac{1}{n}) \\ 1 & \text{on } (\frac{1}{2} + \frac{1}{n}, 1)) \end{cases} \qquad g_n(x) = \begin{cases} 0 & \text{on } (0, \frac{1}{2} - \frac{1}{n}) \\ n(x - \frac{1}{2} + \frac{1}{n}) & \text{on } (\frac{1}{2} - \frac{1}{n}, \frac{1}{2}) \\ 1 & \text{on } (\frac{1}{2}, 1)) \end{cases}$$

Then, $g_n \leq f_n$ are Cauchy sequences in L^1 norm and

$$|f_n - g_n|_1 \leq \frac{1}{n} \to 0$$

Thus, $\{f_n\}$ and $\{g_n\}$ belong to an equivalent class of Cauchy sequences ($\{f_n\} \sim \{g_n\}$). Thus, the candidate limits f and g of $\{f_n\}$ and $\{g_n\}$ defined by

$$f(x) = \begin{cases} 0 & \text{on } (0, \frac{1}{2}] \\ 1 & \text{on } (\frac{1}{2}, 1) \end{cases} \qquad g(x) = \begin{cases} 0 & \text{on } (0, \frac{1}{2}) \\ 1 & \text{on } [\frac{1}{2}, 1)) \end{cases}$$

belong to the same equivalent class. Note that f and g differ at a a single point $x = \frac{1}{2}$. In general if $f \sim g$ (i.e., f and g belong to the same equivalent class) differ at a countable many points.

Exercise 6 Consider

$$\int_0^1 (u'(x)\psi'(x) - f_\epsilon(x)\psi(x)) \, dx, \quad u(0) = u(1) = 0$$

with

$$f_\epsilon = \begin{cases} 0 & |x - \frac{1}{2}| \geq \frac{\epsilon}{2} \\ \frac{1}{\epsilon} & |x - \frac{1}{2}| \leq \frac{\epsilon}{2}. \end{cases}$$

Since $-u_\epsilon'' = f_\epsilon$, we have for $\epsilon = \frac{1}{n}$

$$u_\epsilon(x) = \begin{cases} \frac{1}{2}(\frac{1}{2} - |x - \frac{1}{2}|) & |x - \frac{1}{2}| \geq \frac{\epsilon}{2} \\ -\frac{1}{2\epsilon}(x - \frac{1}{2})^2 + \frac{1}{4} - \frac{\epsilon}{8} & |x - \frac{1}{2}| \leq \frac{\epsilon}{2} \end{cases}$$

That is, $u'_\epsilon(x) \in C(0,1)$ is a Cauchy sequence in $L^1(0,1)$-norm and if let $L^1(0,1)$ is the completion of $C(0,1)$ with $L^1(0,1)$-norm, then $u'_\epsilon(x)$ converges to $u' \in L^1(0,1)$ with $u = \frac{1}{2}(\frac{1}{2} - |x - \frac{1}{2}|)$, i.e.,

$$u'(x) = \begin{cases} \frac{1}{2} & x < \frac{1}{2} \\ 0 & x = \frac{1}{2} \\ -\frac{1}{2} & x > \frac{1}{2} \end{cases}$$

Also, note that

$$\lim_{\epsilon \to 0^+} \int_0^1 f_\epsilon(x)\psi(x)\,dx = \psi(\frac{1}{2}) \text{ for all } \psi \in C(0,1)$$

and thus u satisfies

$$\int_0^1 u'(x)\psi'(x)\,dx = \psi(\frac{1}{2}) \text{ for all } \psi \in H_0^1(0,1)$$

2.2 Distribution and Weak derivative

In this section we introduce the distribution (generalized function). The concept of distribution is very essential for defining a generalized solution to PDEs and provides the foundation of PDE theory. Let $\mathcal{D}(\Omega)$ be a vector space of all infinitely many continuously differentiable functions $C_0^\infty(\Omega)$ with compact support in Ω. That is,

$$\alpha\, f_1 + \beta\, f_2 \in \mathcal{D}(\Omega)$$

for all $f_1,\ f_2 \in \mathcal{D}(\Omega)$ and $\alpha,\ \beta \in R$. For example $\Omega = R^n$. For any compact set K (closed and bounded) of Ω, let $\mathcal{D}_K(\Omega)$ be the set of all functions $f \in C_0^\infty(\Omega)$ whose support are in K. Define the derivatives for $\alpha = (\alpha_1, \cdots \alpha_n)$

$$D^\alpha f = \frac{\partial^{\alpha_1} \cdots \partial^{\alpha_n} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

for $f \in \mathcal{D}(R^n)$.

A linear functional T defined on $\mathcal{D}(\Omega)$ satisfies

$$T(\alpha\,\phi_1 + \beta\,\phi_2) = \alpha\, T(\phi_1) + \beta\, T(\phi_2)$$

for all $\alpha,\ \beta \in R$ and $\phi_1,\ \phi_2 \in \mathcal{D}(\Omega)$.

Definition (Distribution) A linear functional T defined on $C_0^\infty(\Omega)$ is a distribution if for every compact subset K of Ω, there exists a positive constant C and a positive integer k such that

$$|T(\phi)| \le C\, sup_{|\alpha| \le k,\, x \in K} |D^\alpha \phi(x)| \text{ for all } \phi \in \mathcal{D}_K(\Omega).$$

Example (Distribution) (1) For f is a locally integrable function on Ω, one defines the corresponding distribution by

$$T_f(\phi) = \int_\Omega f\phi\,dx \text{ for all } \phi \in C_0^\infty(\Omega).$$

since

$$|T_f(\phi)| \le \int_K |f|\,dx \sup_{x\in K} |\phi(x)|.$$

(2) The point evaluation $T(\phi) = \phi(0)$ defines the Dirac delta δ_0 at $x = 0$, i.e.,

$$|\delta_0(\phi)| \le \sup_{x\in K} |\phi(x)|.$$

(3) The line integral $T(\phi) = \int_\Gamma \frac{\partial\phi}{\partial n}\,ds$ at the helper surface, i.e.,

$$|\int_\Gamma \frac{\partial\phi}{\partial n}\,ds| \le \sup_{x\in K} |\nabla\phi(x)|$$

.

Definition (Weak Derivative) A distribution S defined by

$$S(\phi) = -T(D_{x_k}\phi) \text{ for all } \phi \in C_0^\infty(\Omega)$$

is called the distributional derivative of T with respect to x_k and we denote the distribution $S = D_{x_k}T$.

That is, the weak derivative of a distribution T always exits as a distribution In general we have

$$S(\phi) = D^s T(\phi) = (-1)^{|s|}\, T(D^s\phi) \text{ for all } \phi \in C_0^\infty(\Omega).$$

This definition is naturally followed from that for f is continuously differentiable

$$\int_\Omega D_{x_k} f\phi\,dx = -\int_\Omega f\frac{\partial}{\partial x_k}\phi\,dx$$

and thus $D_{x_k}f = D_{x_k}T_f = T_{\frac{\partial f}{\partial x_k}}f$ if f is continuously differentiable. Thus, we let $D^\alpha f$ denote the distributional derivative of T_f if f is a locally integrable function on Ω, i.e.

$$D^\alpha f(\phi) = \int_\Omega (-1)^{|\alpha|} f(x) D^\alpha\phi(x)\,dx.$$

Also, note that

$$\lim_{h\to 0}\frac{\delta_{x+h}(\psi) - \delta_x(\psi)}{h} = \lim_{h\to 0}\frac{\psi(x+h)-\psi(x)}{h} = \psi'(x) = \delta_x(\psi')$$

and thus $D\delta_x = \delta'(x)$.

Example (Weak derivative) Let H be the Heaviside function defined by

$$H(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \ge 0 \end{cases}$$

Then,

$$D_{T_H}(\phi) = -\int_{-\infty}^{\infty} H(x)\phi'(x)\,dx = \phi(0)$$

and thus $DT_H = DH = \delta_0$ is the Dirac delta function at $x = 0$. Moreover for H_ϵ defined by

$$H_\epsilon(x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{\epsilon} & 0 \le \epsilon \\ 1 & x \ge \epsilon \end{cases}$$

we have

$$H'_\epsilon(\phi) = \frac{1}{\epsilon}\int_0^\epsilon \phi(x)\,dx \to \phi(0)$$

as $\epsilon \to 0^+$ and thus H'_ϵ converges to δ_0 in the sense of distribution.

(2) The distributional solution for $-D^2u = \delta_{x_0}$ satisfies

$$-\int_{-\infty}^{\infty} u\phi''\,dx = \phi(x_0)$$

for all $\phi \in C_0^\infty(R)$. That is, $u = \frac{1}{2}|x - x_0|$ is the fundamental solution, i.e.,

$$-\int_{-\infty}^{\infty} |x - x_0|\phi''\,dx = \int_{\infty}^{x_0} \phi'(x)\,dx - \int_{x_0}^{\infty} \phi'(x)\,dx = 2\phi(x_0).$$

In general for $d \ge 2$ let

$$G(x, x_0) = \begin{cases} \frac{1}{4\pi}\log|x - x_0| & d = 2 \\ c_d\,|x - x_0|^{2-d} & d \ge 3. \end{cases}$$

Then

$$\Delta G(x, x_0) = 0, \quad x \ne x_0.$$

and $u = G(x, x_0)$ is the fundamental solution to to $-\Delta$ in R^d,

$$-\Delta u = \delta_{x_0}.$$

In fact, let $B_\epsilon = \{|x - x_0| \le \epsilon\}$ and $\Gamma = \{|x - x_0| = \epsilon\}$ be the surface. By the divergence theorem

$$\int_{R^d \setminus B_\epsilon(x_0)} G(x, x_0)\Delta\phi(x)\,dx = \int_\Gamma \frac{\partial}{\partial\nu}\phi(G(x, x_0) - \frac{\partial}{\partial\nu}G(x, x_0)\phi(s))\,ds$$

$$= \int_\Gamma (\epsilon^{2-d}\frac{\partial\phi}{\partial\nu} - (2 - d)\epsilon^{1-d}\phi(s))\,ds \to \frac{1}{c_d}\phi(x_0)$$

That is, $G(x, x_0)$ satisfies

$$-\int_{R^d} G(x, x_0)\Delta\phi\,dx = \phi(x_0).$$

In general let $\mathcal{L}$ be a linear differential operator and $\mathcal{L}^*$ denote the formal adjoint operator of $\mathcal{L}$, i.e.

$$(\mathcal{L}\phi, \psi) = (\phi, \mathcal{L}^*\psi)$$

$\psi \in C_0^\infty(\Omega)$. An locally integrable function u is said to be a distributional solution to $\mathcal{L}u = T$ where $\mathcal{L}$ is a differential operator and T is distribution if

$$\int_\Omega u(x)(\mathcal{L}^*\psi)(x)\,dx = T(\psi)$$

for all $\psi \in C_0^\infty(\Omega)$.

Exercise 7 (1) Let $\{B_k^N\}$ be the linear spline function. Then,

$$u^N = \sum_{k=1}^{N-1} u_k\, B_k^N(x)$$

satisfies

$$D^2 u^N = \frac{u_{k+1} - 2u_k + u_{k-1}}{h^2}\delta_{x_k}.$$

(2) Let $\{B_k^N\}$ be the cubic spline function. Then,

$$D^3 B_k^N = 6N^3\,[1, 3, -3, -1]$$

is picewise constant on (x_{k-2}, x_{k+2}) and

$$D^4 B_k^N(u) = 6N^3\,(u_{k-2} - 4u_{k-1} + 6u_k - 4u_{k+1} + u_{k+2}).$$

2.3 Sobolev space

Define the the Sobolev space

$$W^{m,p}(\Omega) = \{f \in L^p(\Omega) : D^\alpha f \in L^p(\Omega),\ |\alpha| = \alpha_1 + \cdots \alpha_n \le m\ |s| \le m\}$$

with norm

$$|f|_{W^{m,p}(\Omega)} = \left(\int_\Omega \sum_{|\alpha|\le m} |D^\alpha f|^p\,dx\right)^{\frac{1}{p}}.$$

Here, $D^\alpha f \in L^2(\Omega)$ is equivalent to there exists a $g \in L^2(\Omega)$ such that

$$\int_\Omega g(x)\psi(x)\,dx = D^\alpha(\psi) = \int_\Omega (-1)^\alpha f(x) D^\alpha \psi\,dx.$$

If f belongs to the completion of $C^m(\Omega)$ with $W^{m,p}(\Omega)$ norm, then $f \in W^{m,p}(\Omega)$.

$X = W^{m,p}(\Omega)$ is complete. In fact If $\{f_n\}$ is Cauchy in X, then $\{D^\alpha f_n\}$ is Cauchy in $L^p(\overline{\Omega})$ for all $|\alpha| \le m$. Since $\overline{L}^p(\Omega)$ is complete, $D^s f_n \to g^\alpha$ in $L^p(\Omega)$. But since

$$\lim_{n\to\infty}\int_\Omega f_n D^\alpha\phi\,dx = \int_\Omega f D^s\phi\,dx = \int g^\alpha\phi\,dx,$$

we have $D^s f = g^s$ for all $|s| \le m$ and $|f_n - f|_X \to 0$ as $n \to \infty$.

Let $H^m(\Omega) = W^{m,2}(\Omega)$ be the Hilbert space with the inner product

$$(f,g)_{H^m} = \int_\Omega \sum_{|\alpha|\le m} D^\alpha f D^\alpha g \, dx.$$

Define the gradient of f by

$$\nabla f = \text{grad} f = (\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n})$$

and the divergence of vector function $\vec{\psi} = (\psi_1, \cdots, \psi_n)$ by

$$\nabla \cdot \vec{\psi} = \text{div}\vec{\psi} = \frac{\partial \psi_1}{\partial x_1} + \cdots + \frac{\partial \psi_n}{\partial x_n}$$

Let n be the outward normal vector at the boundary $\partial\Omega$ of Ω. We have the divergence formula

$$\int_\Omega \text{div}\vec{\psi}\, dx = \int_{\partial\Omega} n \cdot \vec{\psi}\, ds. \tag{2.4}$$

Note that for a rectangular domain$R = (a,b) \times (c,d)$

$$\int_a^b \int_c^d (\frac{\partial \psi_1}{\partial x} + \frac{\partial \psi_2}{\partial y})\, dxdy =$$

$$= \int_c^d (\psi_1(b,y) - \psi_1(a,y))\, dy + \int_c^d (\psi_1(b,y) - \psi_1(a,y))\, dy = \int_{\partial R} n \cdot \vec{\psi}\, ds,$$

Since Ω is the limit of partitioned domain Ω_h by sub-rectangular domains $(a_i, b_i)\, times (c_i, d_i)$ and the Riemann sum satisfies

$$\sum_i \int_{a_i}^{b_i} \int_{c_i}^{d_i} (\frac{\partial \psi_1}{\partial x} + \frac{\partial \psi_2}{\partial y})\, dxdy = \int_{\partial R_h} n \cdot \vec{\psi}\, ds,$$

where the inner path integrals are cancelling out, the divergence theory follows from taking the limit $h \to 0^+$. Since $\text{div}(\phi\,\psi) = \nabla\phi \cdot \psi + \phi \text{div}\psi$ we have the Green formula

$$\int_\Omega \phi\, \text{div}\psi = \int_{\partial\Omega} \phi\, n \cdot \psi\, ds - \int_\Omega \nabla\phi \cdot \psi\, dx \tag{2.5}$$

and thus for $\psi = \nabla v$ and $u = \phi$

$$\int_\Omega u\Delta v = \int_{\partial\Omega} u \frac{\partial}{\partial n} v\, ds - \int \nabla u \cdot \nabla v\, dx \tag{2.6}$$

where $\Delta = \text{divgrad} = \nabla \cdot \nabla$ is the Laplace operator

$$\Delta u = \sum_k \frac{\partial^2 u}{\partial x_k^2} = \nabla \cdot \nabla u.$$

Note that for the one dimensional (2.5) reduces to the integration by parts

$$\int_a^b \phi(x)\psi'(x)\,dx = \phi(x)\psi(x)|_{x=a}^{x=b} - \int_a^b \phi'(x)\psi(x)\,dx$$

since

$$\int_{\partial\Omega} \phi\, n\cdot\psi\, ds = \phi(x)\psi(x)|_{x=a}^{x=b}.$$

In particular

$$H^1(\Omega) = \{u \in L^2(\Omega) : \nabla u \in L^2(\Omega)^n\}.$$

and

$$H_0^1(\Omega) = \{u \in H^1(\Omega) : u|_{\partial\Omega} = 0\}.$$

2.4 Lax-Milgram Theorem and Banach-Necas-Babuska Theorem

In this section we discuss the existence and uniqueness of solution s to a linear equation: $Ax = f$. Let X be a Hilbert space=the complete inner product space with

$$(x,y)_X \text{ is a bounded bilinear form on } X\times X.$$

Let σ be a (complex-valued) sesquilinear form on $X\times X$ satisfying

$$\sigma(\alpha\, x_1 + \beta\, x_2, y) = \alpha\,\sigma(x_1,y) + \beta\,\sigma(x_2,y)$$

$$\sigma(x, \alpha\, y_1 + \beta\, y_2) = \bar\alpha\,\sigma(x,y_1) + \bar\beta\,\sigma(x,y_2),$$

$$|\sigma(x,y)| \le M\,|x||y| \quad \text{for all } x,y\in X \quad \text{(Bounded)} \tag{2.7}$$

and

$$Re\,\sigma(x,x) \ge \delta\,|x|^2 \text{ for all } x\in X \text{ and } \delta>0 \quad \text{(Coercive)}. \tag{2.8}$$

If X is real, then $\sigma : (x,y)\in X\times X \to R$ is a bounded bilinear form.

Then for each $f\in X^*$ =the dual space of X, i.e.,

$$X^* = \text{ the space of bounded linear functionals } f \text{ on } X,$$

there exist a unique solution $x\in X$ to

$$\sigma(x,y) = f(y) = \langle f, y\rangle_{X^*\times X} \quad \text{for all } y\in X \tag{2.9}$$

and

$$|x|_X \le \delta^{-1}\,|f|_{X^*}.$$

Proof: Let us define the linear operator S from X^* into X by

$$Sf = x, \quad f\in X^*$$

where $x\in X$ satisfies

$$\sigma(x,y) = \langle f, y\rangle \quad \text{for all } y\in X.$$

The operator S is well defined since if $x_1,\ x_2 \in X$ satisfy the above, then $\sigma(x_1 - x_2, y) = 0$ for all $y \in X$ and thus $\delta\,|x_1 - x_2|_X^2 \le Re\,\sigma(x_1 - x_2, x_1 - x_2) = 0$.

Next we show that dom(S) is closed in X^*. Suppose $f_n \in dom(S)$, i.e., there exists $x_n \in X$ satisfying $\sigma(x_n, y) = \langle f_n, y\rangle$ for all $y \in X$ and $f_n \to f$ in X^* as $n \to \infty$. Then

$$\sigma(x_n - x_m, y) = \langle f_n - f_m, y\rangle \quad \text{for all } \ y \in X$$

Setting $y = x_n - x_m$ in this we obtain

$$\delta\,|x_n - x_m|_X^2 \le Re\,\sigma(x_n - x_m, x_n - x_m) \le |f_n - f_m|_{X^*}|x_n - x_m|_X.$$

Thus $\{x_n\}$ is a Cauchy sequence in X and so $x_n \to x$ for some $x \in X$ as $n \to \infty$. Since σ and the dual product are continuous, thus $x = Sf$.

Now we prove that dom$(S) = X^*$. Suppose dom$(S) \ne X^*$. Since dom(S) is closed there exists a nontrivial $x_0 \in X$ such that $\langle f, x_0\rangle = 0$ for all $f \in$dom(S). Consider the linear functional $F(y) = \sigma(x_0, y),\ y \in X$. Then since σ is bounded $F \in X^*$ and $x_0 = SF$. Thus $F(x_0) = 0$. But since $\sigma(x_0, x_0) = \langle F, x_0\rangle = 0$, by the coercivity of σ $x_0 = 0$, which is a contradiction. Hence dom$(S) = X^*$. $\square$

Remark (1) If $X = R^N$ and $A \in R^{N\times N}$, define a bilinear form σ by

$$\sigma(x, y) = (Ax, y)$$

and $f(y) = (b, y)$. The Lax-Milgram theory states if A is positive definite, i.e., $(Ax, x) > 0$ for all $x \in R^N$, there exists a unique solution to x to $Ax = b$ and $x = A^{-1}b$.
(2) In general there exists $A \in \mathcal{L}(X, X^*)$ such that

$$\sigma(x, y) = \langle Ax, y\rangle$$

and (2.9) is equivalent to the linear equation $Ax = f$, where

$$\langle f, \psi\rangle_{X^*\times X} = f(\psi)$$

is the dual product of $X^* \times X$.
(3) if σ is symmetric ($\sigma(x, y) = \sigma(y, x)$), then the minimization

$$\min \quad J(u) = \frac{1}{2}\sigma(u, u) - f(u)$$

has a unique solution u and $Au = f$. In fact

$$\frac{d}{dt}J(u + t\,\psi)|_{t=0} = \sigma(u, \psi) - f(\psi).$$

Also, $\sigma(x, y)$ defines an inner-product on X. and $u = SF$ coincides with the Riesz representation of $F \in X^*$, i.e., given $f \in X^*$ there exists a unique $u = u_f \in X$ satisfying

$$f(\phi) = (u, \phi)_X \quad \text{for all } \phi \in X,$$

and $|u_f|_X = |f|_{X^*}$. Exercise (1) X^* with graph norm

$$|f|_{X^*} = \inf\{M : |f(x)| \le M\,|x|\}$$

is a Banach space. Suppose $f_n \in X^*$ is Cauchy sequence. Then, for all $x \in X$

$$|f_n(x) - f_m(x)| \le |f_n - f_m|_{X^*}|x|_X \to 0 \text{ as } m \to \infty$$

and thus, $\lim f_n(x)$ exist since $(R, ||)$ is complete. Define a functional f by

$$f(x) = \lim_{n\to\infty} f_n(x)$$

Then, f is linear and bounded. Given $\epsilon > 0$ there exists N such that

$$|f_n(x) - f(x)| = |f_n(x) - f(x)| \le \epsilon|x|$$

for $n \ge N$. Thus, $|f_n - f|_{X^*} \to 0$ as $n \to \infty$.
(2) Since $|(f(\phi)| = |(u_f, \phi)| \le |u_f|_X|\phi|_X$, $|f|_{X^*} \le |u_f|_X$ and thus $|f|_{X^*} = |u_f|_X$ for the Riesz representation.

Example (Laplace operator) Consider $X = H_0^1(\Omega)$, $H = L^2(\Omega)$ and

$$\sigma(u, \phi) = (u, \phi)_X = \int_\Omega \nabla u \cdot \nabla \phi \, dx.$$

Then, by Green formula

$$\int_\Omega -\Delta u \phi \, dx = -\int_{\partial\Omega} \frac{\partial u}{\partial \nu}\phi + \int_\Omega \nabla u \cdot \nabla \phi \, dx,$$

$$Au = -\Delta u = -(\frac{\partial^2}{\partial x_1^2}u + \frac{\partial^2}{\partial x_2^2}u)$$

For $\Omega = (0, 1)$ and $f \in L^2(0, 1)$

$$\int_0^1 \frac{d}{dx} y \frac{d}{dx} u \, dt = \int_0^1 f(x)y(x) \, dx$$

is equivalent to

$$\int_0^1 \frac{d}{dx} y \, (\frac{d}{dx}u + \int_x^1 f(s) \, ds) \, dx = 0$$

for all $y \in H_0^1(0, 1)$. Thus,

$$\frac{d}{dx}u + \int_x^1 f(s) \, ds = c \text{ (a constant)}$$

and therefore $\frac{d}{dx}u \in H^1(0, 1)$ and

$$Au = -\frac{d^2}{dx^2}u = f \text{ in } L^2(0, 1).$$

Transport equation Consider the transport equation for $u = u(x, t)$:

$$u_t + b(x) \cdot \nabla u = 0, \quad u(x, 0) = u_0(x). \tag{2.10}$$

For $U(t) = u(t, x(t))$ by the chain rule

$$\frac{d}{dt}U(t) = u_t + \frac{d}{dt}x(t) \cdot \nabla u.$$

Let $x(t)$ satisfies the (backward) ordinary differential equation (characteristic)

$$\frac{d}{dt}x(t) = b(x(t)), \quad x(t) = x.$$

Then, $\frac{d}{dt}U(t) = 0$ along the characteristic curve $x(t)$ and thus

$$u(x,t) = u(x(0),0) = u_0(x_0), \quad x_0 = x(0),$$

is the solution to the transport equation. Thus we consider the diffusive transport equation

$$\mathcal{A}u = -\nabla \cdot (a(x)\nabla u) + b(x) \cdot \nabla u + c(x)u(x) = f(x).$$

Exercise 8 For the one dimensional non-symmetric form

$$a(u,\psi) = \int_0^1 a(x)u'\psi' + b(x)u'\psi + c(x)u(x)\psi = f(\psi).$$

we have

$$a(u,u) = \int_0^1 (a(x_|u'(x)|^2 + (-\frac{1}{2}b' + c(x))|u|^2\,dx \geq \delta \int_0^1 |u'|^2\,dx$$

with $X = H_0^1(0,1)$. If assume $-\frac{1}{2}b' + c(x)) \geq 0$. the coercivity (2.8) holds. For boundedness (2.7) we use $|\psi|_\infty \leq |\psi|_{H_(^1 0,1)}$.

One can prove that if a is symmetric, then the inf-sup condition (2.12) is satisfied provided that $a(u,u) \neq 0$ for $u \in H_0^1(0,1)$ and the BNB theory applies.

Excursus 9 (Elliptic operator) Consider a second order elliptic equation

$$\mathcal{A}u = -\nabla \cdot (a(x)\nabla u) + b(x) \cdot \nabla u + c(x)u(x) = f(x), \quad \frac{\partial u}{\partial \nu} = g \text{ at } \Gamma_1 \quad u = 0 \text{ at } \Gamma_0$$

where Γ_0 and Γ_1 are disjoint and $\Gamma_0 \cup \Gamma_1 = \Gamma$. Integrating this against a test function ϕ, we have

$$\int_\Omega \mathcal{A}u\phi\,dx = \int_\Omega (a(x)\nabla u \cdot \nabla\phi + b(x) \cdot \nabla u\phi + c(x)u\phi)\,dx - \int_{\Gamma_1} g\phi\,ds_x = \int_\Omega f(x)\phi(x)\,dx,$$

for all $\phi \in C^1(\Omega)$ vanishing at Γ_0. Let $X = H^1_{\Gamma_0}(\Omega)$ is the completion of $C^1(\Omega)$ vanishing at Γ_0 with inner product

$$(u,\phi) = \int_\Omega \nabla u \cdot \nabla\phi\,dx$$

i.e.,

$$H^1_{\Gamma_0}(\Omega) = \{u \in H^1(\Omega) : u|_{\Gamma_0} = 0\}$$

Define the bilinear form σ on $X \times X$ by

$$\sigma(u,\phi) = \int_\Omega (a(x)\nabla u \cdot \nabla\phi + b(x)\cdot\nabla u\phi + c(x)u\phi.$$

Then, by the Green's formula

$$\sigma(u,u) = \int_\Omega (a(x)|\nabla u|^2 + b(x)\cdot\nabla(\frac{1}{2}|u|^2) + c(x)|u|^2)\,dx$$

$$= \int_\Omega (a(x)|\nabla u|^2 + (c(x) - \frac{1}{2}\nabla\cdot b)\,|u|^2)\,dx + \int_{\Gamma_1} \frac{1}{2} n\cdot b|u|^2\,ds_x.$$

If we assume

$$0 < \underline{a} \le a(x) \le \bar{a}, \quad c(x) - \frac{1}{2}\nabla\cdot b \ge 0, \quad n\cdot b \ge 0 \text{ at } \Gamma_1,$$

then σ is coercive with $\delta = \underline{a}$. For boundedness (2.7) we use the Poincare inequality

$$\int_\Omega |u|^2\,dx \le \tilde{M} \int_\Omega |\nabla u|^2\,dx$$

for some $\tilde{M} > 0$ and all $u \in H^1_{\Gamma_0}$.

Remark $\Gamma_0 = \{n\cdot\vec{b} < 0\}$ is inflow boundary and we specify u and $\Gamma_1 = \{n\cdot\vec{b} > 0\}$ is outflow boundary and the flux $\frac{\partial}{\partial n}$ is specified.

Exercise 10 (Bi-Harmonic equation) Consider the bi-harmonic equation

$$\Delta^2 u + c(x)u = f$$

with various boundary conditions at $\partial\Omega$. For example

$$u = \frac{\partial u}{\partial n} = 0$$

In this case $X = \{u \in H^2(\Omega) : u = \frac{\partial u}{\partial n} = 0\}$ and

$$\sigma(u,v) = \int_\Omega (\Delta u\Delta v + c(x)u(x)v(x))\,dx$$

since by Green formula

$$\int_\Omega \Delta^2 uv\,dx = \int_{\partial\Omega} (\frac{\partial\Delta u}{\partial n} - \Delta u\frac{\partial v}{\partial n})\,ds + \int_\Omega \Delta u\Delta v\,dx.$$

Exercise 11 Consider the boundary condition

$$\frac{\partial\Delta u}{\partial n} - \alpha u = f_1, \quad \Delta u + \beta\frac{\partial u}{\partial n} = f_2,$$

Then, we have $X = H^2(\Omega)$ and

$$\sigma(u,v) = \int_\Omega (\Delta u \Delta v + c(x)u(x)v(x))\,dx + \int_{\partial\Omega} \alpha uv + \beta \frac{\partial u}{\partial n}\frac{\partial v}{\partial n}$$

$$= \int_\Omega fv\,dx + \int_{\partial\Omega} (f_1 v - f_2 \frac{\partial v}{\partial n})\,ds.$$

The Banach space version of Lax-Milgram theorem is as follows.

Banach-Necas-Babuska (BNB) Theorem Let V and W be Banach spaces. Consider the linear equation for $u \in W$

$$a(u,v) = f(v) \quad \text{for all } v \in V \tag{2.11}$$

for given $f \in V^*$, where a is a bounded bilinear form on $W \times V$. The problem is well-posed if and only if the following conditions hold:

$$\begin{gathered} \inf_{u\in W} \sup_{v\in V} \frac{a(u,v)}{|u|_W |v|_V} \geq \delta > 0 \\ a(u,v) = 0 \text{ for all } u \in W \text{ implies } v = 0. \end{gathered} \tag{2.12}$$

Under conditions we have the unique solution $u \in W$ to (2.11) satisfies

$$|u|_W \leq \frac{1}{\delta}\,|f|_{V^*}.$$

Proof: Let A be a bounded linear operator from W to V^* defined by

$$\langle Au, v\rangle_{V^*\times V} = a(u,v) \text{ for all } u \in W,\ v \in V.$$

The inf-sup condition is equivalent to for any $u \in W$:

$$|Au|_{V*} \geq \delta |u|_W,$$

and thus the range of A, $R(A)$ is closed in V^* and $N(A) = 0$. But since V is reflexive and

$$\langle Au, v\rangle_{V^*\times V} = \langle u, A^* v\rangle_{W\times W^*}$$

from the second condition $N(A^*) = \{0\}$. It thus follows from the closed range that $R(A) = N(A^*)^\perp = V^*$. Thus, A is bijective and from the open mapping theorems that A^{-1} is bounded. □

2.5 Mixed finite element

In this section we discuss the applications of the Banach-Necas-Babuska thorem. An elliptic equation $-\nabla\cdot(a\,\nabla u) = f,\ u = 0$ at $\partial\Omega$ is equivalent to

$$\nabla \cdot p = f, \quad a\,\nabla u + p = 0.$$

(1) Note that

$$\int_\Omega \nabla \cdot p\,\phi_1\,dx = \int_{\partial\Omega} n \cdot \cdot p\,\phi_1\,ds + \int_\Omega p \cdot \nabla\phi_1\,dx = \int_\Omega p \cdot \nabla\phi_1\,dx$$

for $\phi_1 \in H_0^1(\Omega)$. Thus, we define a and f (2.11) by

$$a((u,p),(\phi_2,\phi_1)) = (a\,\nabla u + p, \phi_2) - (p, \nabla\phi_1) = (f, \phi_2)$$

for $(u,p) \in W = H_0^1(\Omega) \times L^2(\Omega)^2$ and $(\phi_2,\phi_1) \in V = L^2(\Omega)^2 \times H_0^1(\Omega)$ We use the linear element for u and the piecewise constant element for p.

(2) Note that

$$\int_\Omega a\nabla u \cdot \phi_2\,dx = \int_{\partial\Omega} n \cdot (a\phi_2)\,u\,ds + \int_\Omega \nabla \cdot (a\phi_2)\,u\,dx = \int_\Omega \nabla \cdot (a\phi_2)\,u\,dx$$

Thus, we have the second formulation:

$$a((u,p),((\phi_2,\phi_1)) = -(u, \mathrm{div}\,(a\phi_2) + (\mathrm{div}p, \phi_1) = (f, \phi_1)$$

for $(u,p) \in L^2(\Omega) \times H^1_{div}(\Omega)$ and $(\phi_2,\phi_1) \in V = H^1_{div}(\Omega) \times L^2(\Omega)$, where

$$H_{div}(\Omega) = \{\psi \in L^2(\Omega)^n : \mathrm{div}\psi \in L^2(\Omega)\}.$$

We can use the piecewise constant for u and the linear elements for p.

(3) A bi-harmonic equation $\Delta^2 u = f,\ u = 0,\ \frac{\partial u}{\partial n} = 0$ at $\partial\Omega$ is equivalent to

$$\Delta v = f, \quad \Delta u = v$$

Note that

$$\int_\Omega \Delta v\,\phi_1\,dx = \int_{\partial\Omega} \frac{\partial v}{\partial n}, \phi_1\,dx + \int_\Omega \nabla v \cdot \nabla\phi_1\,dx = \int_\Omega \nabla v \cdot \nabla\phi_1\,dx,$$

for $\phi_1 \in H_0^1(\Omega)$ and

$$\int_\Omega \Delta u\,\phi_2\,dx = \int_{\partial\Omega} \frac{\partial u}{\partial n}, \phi_2\,dx + \int_\Omega \nabla u \cdot \nabla\phi_2\,dx = \int_\Omega \nabla u \cdot \nabla\phi_2\,dx.$$

Thus, we define the bilinear form a and f (2.11) by

$$a((u,v),(\phi_2,\phi_1)) = (\nabla u, \nabla\phi_2) + (v, \phi_2) - (\nabla v, \nabla\phi_1) = (f, \phi_1)$$

for $(u,v) \in W = H_0^1(\Omega) \times H^1(\Omega)$ and $(\phi_2,\phi_1) \in V = H^1(\Omega) \times H_0^1(\Omega)$. Thus, one can use the linear finite element for (u,v).

Now we consider the Galerkin method for (2.11). Let $V_h \subset V$ and $W_h \subset W$ be finite dimensional. The Ritz-Galerkin method is

$$u_h \in W_h \text{ satisfying } a(u_h, v_h) = f(v_h) \text{ for all } v_h \in V_h. \tag{2.13}$$

We assume the discrete inf-sup condition:

$$\sup_{v\in V_h} \frac{a(u_h,v)}{|v|_W} \geq \delta |u_h| \text{ for } u_h \in W_h. \tag{2.14}$$

Then, we have the error estimate:
Lemma 1. (Cea′ lemma) Assume the discrete inf-sup condition (2.14) is satisfied. Then the Galerkin method (2.13) has the unique solution $u_h \in V_h$ and

$$|u_h - u|_W \leq (1 + \frac{c}{\delta}) \inf_{v\in V_h} |v-u|_W$$

where $a(u,v) \leq c\,|u|_W |v|_V$.
Proof: Let u is the solution to $a(u,v) = f(v)$ for all $v \in V$. Then, we have the Galerkin orthogonality

$$a(u_h - u, v) = 0 \text{ for all } v \in\in V_h.$$

Thus, it follows from (2.14) that for all $\tilde{u} \in W_h$

$$\delta\,|u_h - \tilde{u}|_W \leq \sup_{v\in V_h} \frac{a(u_h-\tilde{u},v)}{|v|_V} = \sup_{v\in V_h} \frac{a(u-\tilde{u},v_h)}{|v|_V} \leq c\,|u-\tilde{u}|_W.$$

Now, by triangle inequality for any $\tilde{u} \in W_h$ we have

$$|u_h - u|_W \leq |u_h - \tilde{u}|_W + |\tilde{u} - u|_W \leq (1 + \frac{c}{\delta})|u - \tilde{u}|_W. \square$$

Example (BNB) Consider $\mathrm{div}(\vec{b}(x)u) = f$. Let $W = L^2(\Omega)$ and $V = H^1(\Omega)$. Since

$$\int_\Omega div(\vec{b}u)\psi\,dx = \int_{\partial\Omega} u\psi\, n\cdot\vec{b}\,ds - \int_\Omega \vec{b}(x)u(x)\psi(x)\,dx.$$

and assume that $u = 0$ at $n\cdot\vec{b} < 0$, define on $W \times V \cap \{\psi = 0 \text{ at } n\cdot b > 0\}$, define

$$a(u,\psi) = -\int_\Omega b(x)u(x)\psi'(x)\,dx.$$

Let $\Omega = [-1,1]\times[-1,1]$ and

$$V_h = \text{piecewise constant and } W_h = span\{\psi_{i,j} = B_i(x_1)B_j(x_2)\} \cap \{\psi = 0 \text{ at } n\cdot b > 0\}$$

Then, we obtain for $\{u_{i+1/2,j+1/2}\}$

$$\tfrac{1}{2}((b^1u)_{i+1/2,j+1/2} - (b^1u)_{i-1/2,j+1/2} + (b^1u)_{i+1/2,j-1/2} - (b^1u)_{i-1/2,j-1/2})$$

$$+\frac{1}{2}((b^2u)_{i+1/2,j+1/2} - (b^2u)_{i+1/2,j-1/2} + (b^2u)_{i-1/2,j+1/2} - (b^2u)_{i-1/2,j-1/2}) = f_{i,j}$$

Take Home Exam I

Problem 1 Consider the biharmonic equation

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

with boundary conditions

$$u(0) = 0, \;\; u'(0) = 0, \;\; u'''(1) = 0 \; u''(1) = \alpha\, u(1) + 1,$$

Formulate the weak form and find a sufficient condition on $c(x)$ and α for the existence and uniqueness of solutions. Hint: Use the Lax-Milgram theory and the inequality $|u|_\infty \leq \sqrt{\int_0^1 |u'|^2\, dx}$ for all $u \in H^1(0,1)$ satisfying $u(0) = 0$.

Problem 2 Formulate the mixed finite element for

$$(-a(x)u' + b(x)\, u)' + c(x)u = f(x), \;\; u(0) = 0, \;\; u(1) = 0$$

with the linear finite element for u and the piecewise constant for p. Hint: Use the mixed finite element formulation.

Problem 3 Formulate the mixed finite element method of

$$u''''(x) + u(x) = f, \quad 0 < x < 1,$$

based on the linear finite elements for the following boundary conditions:

$$(a) \quad u(0) = 0, \;\; u'(0) = 1, \;\; u(1) = 0, \; u'(1) = 2.$$

$$(b) \quad u(0) = 0, \;\; u''(0) = 0, \;\; u(1) = 0, \; u''(1) = 0.$$

Take Home Exam II

Problem 1 Consider the periodic boudady condition $u(0) = u(1)$ and $u'(0) = u'(1)$ for

$$-u'' + c(x)\, u = f, ; 0 < x < 1$$

Drive the weak form and develop the finite element method based on the linear basis elements.

Problem 2 Derive the mixed finite element formulation (u, v) by defining $v = u''$ for the biharmonic equation

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

with boundary conditions

$$u(0) = 0, \;\; u'(0) = 0, \;\; u'''(1) = 0, \; u'''(1) = \alpha\, u(1) + 1.$$

Problem 3 Consider the boundary value problem:

$$-u'' + u = f, \;\; 0 < x < 1$$

with $u(0) = u(1) = 0$. Using the basis function $\{\phi_k(x) = \sin(k\pi x)\}$, i.e.

$$u_N(x) = \sum_{k=0}^{N} u_k \, \sin(k\pi \, x)$$

develop the finite element method. Hint: one can use the othogonality

$$\int_0^1 \sin(k\pi \, x) \sin(j\pi \, x) \, dx = \frac{1}{2}\delta_{k,j}, \quad \int_0^1 \cos(k\pi \, x) \cos(j\pi \, x) \, dx = \frac{1}{2}\delta_{k,j}.$$

Problem 4 Consider the two-dimensional (stationary) Navier Stokes equations

$$-\Delta \vec{u} + \vec{u} \cdot \nabla \vec{u} + \nabla p = \vec{f}, \;\; \vec{u} = 0 \text{(non-slip) at boundary } \partial\Omega.$$

$$\nabla \cdot \vec{u} = 0$$

Define the stream function ψ and vorticity ω by

$$\vec{u} = \operatorname{curl} \psi = (\frac{\partial \psi}{\partial x_1}, -\frac{\partial \psi}{\partial x_2})$$

$$\omega = \operatorname{curl} \vec{u} = \frac{\partial u^2}{\partial x_1} - \frac{\partial u^1}{\partial x_2}$$

Show that (ω, ψ) satisfies (Hint: div curl $\psi = 0$ and curl $\nabla p = 0$)

$$\Delta \psi = \omega, \quad -\Delta \omega + \vec{u} \cdot \nabla \omega = \operatorname{curl} \vec{f} \tag{2.15}$$

with $\omega = 0$ and $\psi = 0$ at boundary $\partial\Omega$. Formulate the mixed finite element method for (2.15), assuming $\vec{u}$ is a given vector field.

Problem 5 Consider the Crank-Nicolson and weak form (Lax-Milgram) for the parabolic equation

$$(\frac{u_h^n - u_h^{n-1}}{\Delta t}, \psi_h) + a(\frac{u_h^n + u_h^{n-1}}{2}, \psi_h) = (f, \psi_h) \text{ for all } \psi_h \in X_h$$

Show that u_h^n satisfies

$$u_h^n = (I + \frac{1}{2}\Delta t A_h)^{-1}((I - \frac{1}{2}\Delta t A_h)u_h^{n-1} + \Delta t \, f_h^{n+\frac{1}{2}})$$

$$= (I + \frac{1}{2}\Delta t A_h)^{-1}(2u_h^{n-1} + \Delta t \, f_h^{n+\frac{1}{2}}) - u_h^{n-1}.$$

Letting $\psi_h = \frac{u_h^n + u_h^{n-1}}{2}$ we have

$$|u_h^n|_H^2 - |u_h^{n-1}|_H^2 + a(\frac{u_h^n + u_h^{n-1}}{2}, \frac{u_h^n + u_h^{n-1}}{2}) = (f_h^{n+\frac{1}{2}}, \frac{u_h^n + u_h^{n-1}}{2}) \leq \frac{\delta}{2}|\frac{u_h^n + u_h^{n-1}}{2}|_X^2 + \frac{1}{2\delta}|f_h^{n+\frac{1}{2}}|_{X^*}^2.$$

Using the coercivity $a(u, u) \geq \delta \, |u|_X^2$, we have the estimate

$$|u_h^n|_H^2 + \sum_{k=1}^{n} \frac{\delta}{2}|\frac{u_h^n + u_h^{n-1}}{2}|_X^2 \, \Delta t \leq |u_h^0|_H^2 + \sum_{k=1}^{n} \frac{1}{2\delta}|f_h^{k+\frac{1}{2}}|_{X^*}^2 \Delta t.$$

2.6 Saddle point problem and the Mixed finite element method

Next, we consider the generalized Stokes system. The abstract form of a saddle point problem can be expressed as follows. Let V and Q be Hilbert spaces and consider the mixed variational problem for $(u, p) \in V \times Q$ of the form

$$a(u, v) + b(p, v) = f(v), \quad b(u, q) = g(q) \tag{2.16}$$

for all $v \in V$ and $q \in Q$, where a and b is bounded bilinear form on $V \times V$ and $V \times Q$. If we define the linear operators $A \in \mathcal{L}(V, V^*)$ and $B \in \mathcal{L}(V, Q^*)$ by

$$\langle Au, v\rangle = a(u, v) \quad \text{and} \quad \langle Bu, q\rangle = b(u, q)$$

then it is equivalent to the operator form:

$$\begin{pmatrix} A & B^* \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}.$$

Assume the coercivity on a

$$a(u, u) \geq \delta \, |u|_V^2 \tag{2.17}$$

and the inf-sup condition (Ladyzhenskaya?Babuska?Brezzi condition) on b

$$\inf_{q \in P} \sup_{u \in V} \frac{b(u, q)}{|u|_V |q|_Q} \geq \beta > 0 \tag{2.18}$$

Note that inf-sup condition that for all q there exists $u \in V$ such that $Bu = q$ and $|u|_V \leq \frac{1}{\beta}|q|_Q$. Also, it is equivalent to $|B^* p|_{V^*} \geq \beta \, |p|_Q$ for all $p \in Q$.

Remark The general form of the saddle point problem is given by

$$\begin{pmatrix} A & B^* \\ -B & C \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}.$$

where C is a Q-coercive form.

Corollary (Error estimate) Let $V_h \times Q_h$ be a finite event subspace of $V \times Q$ and $(u_h, p_h) \in V_h \times Q_h$ be the solution to the mixed finite element system

$$\begin{aligned} &a(u_h, v_h) + b(v_h, p_h) = f(v_h) \text{ for all } v_h \in V_h, \\ &b(u_h, q_h) = g(q_h) \text{ for all } q_h \in Q_h. \end{aligned} \tag{2.19}$$

Assume the inf-sup condition

$$\inf_{q_h \in Q_h} \sup_{u_h \in V_h} \frac{b(u_h, p_h)}{|u_h|_V |p_h|_Q} \geq \delta > 0 \tag{2.20}$$

uniformly in $h > 0$. Note that

$$a(u_h - u, v_h) + b(p_h - p, v_h) = 0 \text{ for all } v_h \in V_h$$

and thus

$$|p_h - \tilde{p}| \leq \frac{1}{\delta}(|u - u_h|_V + c\,|p - \tilde{p}|) \text{ for all } \tilde{p} \in Q_h.$$

Since

$$a(u_h - \tilde{u}, v_h) + a(u - \tilde{u}, v_h) + b(p_h - p, v_h) + b(p - \tilde{p}, v_h) = 0 \text{ for all } \tilde{v} \in V_h, \;\; \tilde{p} \in Q_h,$$

it follows that there exits a constant C such that

$$|u_h - u|_V + |p_h - p|_Q \leq C\,(\inf_{\tilde{u} \in V_h} |u - \tilde{u}|_V + \inf_{\tilde{p} \in Q_h} |p - \tilde{p}|_V)$$

Exercise 13 Consider the mixed finite element approximation (2.19):

$$u_h = \textstyle\sum_{k=1}^{N} u_k\,\phi_k(x) \in V_h$$

$$p_h = \textstyle\sum_{j=1}^{M} p_j\,\psi_j(x) \in Q_h.$$

Let A_h and B_h be defined by

$$A_{k,\ell} = a(\phi_\ell, \phi_k), \;\; B_{j,\ell} = b(\phi_\ell, \psi_j).$$

Then, we have for (u_h, p_h)

$$\begin{pmatrix} A_h & B_h^* \\ B_h & 0 \end{pmatrix} \begin{pmatrix} u_h \\ p_h \end{pmatrix} = \begin{pmatrix} f_h \\ g_h \end{pmatrix}.$$

Example: Stokes equations Consider the minimization

$$\min \quad J(\vec{u}) = \int_\Omega \frac{1}{2}|\nabla \vec{u}|^2 - (\vec{f}, \vec{u})\,dx \text{ subject to } \nabla \cdot \vec{u} = 0$$

over $\vec{u} \in H_0^1(\Omega)$. Define the Lagrangian

$$L(\vec{u}, p) = J(\vec{u}) - (\nabla \cdot \vec{u}, p).$$

The -Lagrange equation is given by the Stokes equation

$$\tfrac{\partial J}{\partial \vec{u}} = -\Delta \vec{u} + \nabla p - \vec{f} = 0$$

$$\tfrac{\partial J}{\partial p} = -\nabla \cdot \vec{u} = 0,$$

where p denotes the pressure. In this case

$$B = -\text{div}, \quad B^* = \text{grad}$$

and the inf-sup condition is given by

$$\inf_{q\in L^2(\Omega)} \sup_{u\in H_0^1(\Omega)} \frac{(\nabla\cdot u, q)}{|u|_{H_0^1}|q|_{L^2}} \geq \beta > 0 \tag{2.21}$$

The discrete ind-sup condition (2.21) holds for the Taylor-Hood elements, i.e. quadratic elements for the velocities and linear elements for the pressure. But most obvious choice, use linear elements for the both is failed, i.e., $\delta_h \to 0$ as $h \to 0$.

Theorem (Mixed problem) Under conditions (2.17)-(2.18) there exits a unique solution $(u,p) \in V \times Q$ to (2.16) and

$$|u|_V + |p|_Q \leq c\,(|f|_{V^*} + |g|_{Q^*})$$

Proof: For $\epsilon > 0$ consider the penalized problem

$$\begin{aligned} a(u_\epsilon, v) + b(v, p_\epsilon) &= f(v), \quad \text{for all } v \in V \\ -b(u_\epsilon, q) + \epsilon(p_\epsilon, q)_Q &= -g(q), \quad \text{for all } q \in Q. \end{aligned} \tag{2.22}$$

By the Lax-Milgram theorem for every $\epsilon > 0$ there exists a unique solution $(u_\epsilon, p_\epsilon) \in V \times Q$. From the first equation and (2.18),

$$\beta\,|p_\epsilon|_Q \leq |f - Au_\epsilon|_{V^*} \leq |f|_{V^*} + M\,|u_\epsilon|_V.$$

Letting $v = u_\epsilon$ and $q = p_\epsilon$ in the first and second equation and (2.18), we have

$$\delta\,|u_\epsilon|_V^2 + \epsilon\,|p_\epsilon|_Q^2 \leq |f|_{V^*}||u_\epsilon|_V + |p_\epsilon|_Q|g|_{Q^*} \leq C\,(|f|_{V^*} + |g|_{Q^*})|u_\epsilon|_V),$$

and thus $|u_\epsilon|_V$ and thus $|p_\epsilon|_Q$ are bounded uniformly in $\epsilon > 0$. Thus, (u_ϵ, p_ϵ) has a weakly convergent subspace to (u,p) in $V \times Q$ and (u,p) satisfies (2.16). □

2.7 Error analysis

In this section we discuss the convergence and error analysis of elliptic equation $\sigma(u,\psi) = f(\psi)$ for all $\psi \in X$ in the framework of Lax-Milgram theorem.

2.7.1 Conformal case

First, we assume $X_h \subset X$. Assume that $u_h \in X_h$ satisfies

$$\sigma(u_h, \psi_h) = f(\psi_h) \text{ for all } \psi_h \in X_h$$

Since the solution $u \in V$ satisfies$\sigma(u, \psi_h) = f(\psi_h)$ we have

$$\sigma(u - u_h, \psi_h) = 0.$$

This means that the finite element solution u_h is the projection of u onto the space X_h when σ is symmetric form and X is equipped with the inner product $\sigma(\cdot,\cdot)$. Thus, It is the best solution in X_h in the energy norm $\sqrt{\sigma(u,u)}$. In fact, for all $v_h \in X_h$

$$\sigma(u - v_h, u - v_h) = \sigma(u - u_h, u - u_h) + 2\,\sigma(u - u_h, u_- v_h) + \sigma(u_h - v_h, u_h - v_h)$$

$$= \sigma(u - u_h, u - u_h) + \sigma(u_h - v_h, u_h - v_h) \geq \sigma(u - u_h, u - u_h)$$

Moreover, we have

$$\sigma(u-u_h,u-u_h)=\sigma(u-u_h,u-u_h)+\sigma(u-u_h,u_h-\hat{u}_h)=\sigma(u-u_h,u_h-\hat{u}_h)$$

for all $\hat{u}_h \in X_h$. Thus,

$$|u-u_h|_X \le \frac{M}{\delta}\inf_{\hat{u}_h\in X_h}|u-\hat{u}_h|_X,$$

which provides the error estimate of u_h in X.

Aubin-Nitche lemma (L^2 error estimate) Let $w \in X$ is the adjoint system for $w \in X$: for all $v \in X$,

$$a(v,w)=(e_h,v), \quad e_h=u-u_h.$$

We use the elliptic regularity

$$|w|_{H^2(\Omega)} \le c\,|e_h|_{L^2(\Omega)}$$

Then, for the interpolation function $I_h w \in X_h$ of w

$$(e_h,e_h)_{L^2}=(e_h,u-u_h)=a(u-u_h,w)=a(u-u_h,w-I_hw)$$

$$\le M\,|u-u_h|_X|w-I_hw|_X \le M\,|u-u_h|_X h\,|w|_{H^2} \le \tilde{M}\,h^2|e_h|_{L^2}.$$

where we used the Galerkin orthogonality. Thus, we obtain

$$|u-u_h|_{L^2} \le C\,h^2\,|u|_{H^2}|w|_{H^2}.$$

2.7.2 Non-conformal case

Consider the non-confomal finite element system $u_h \in X_h \notin X$

$$a_h(u_h,\psi_h)=f(\psi_h), \text{ for all } \psi_h \in X_h, \tag{2.23}$$

where a_h is a uniformly bounded bilinear on $X_h \times X_h$ with

$$a_h(v_h,v_h) \ge \delta\,|v_h|^2_{X_h} \text{ for all } v_h \in V_h.$$

Here, $X_h \notin X$ but $X \in X_h$. Then, there exits a unique solution $u_h \in V_h$ to (2.23). Assume that

$$a_h(u-u_h,\psi_h)=0.$$

Thus, for all $\hat{u}_h \in X_h$

$$a_h(u-u_h,u-u_h)=a_h(u-u_h,u-\hat{u}_h) \le M|u-u_h|_{X_h}|u-\hat{u}_h|_{X_h}$$

and thus

$$|(u-u_h|_{X_h} \le \frac{M}{\delta}\inf_{\hat{u}_h\in V_h}|u-\hat{u}_h|_{X_h}.$$

3 Finite Elements

Local linear element on the right triangle

$$\phi_1 = 1 - \xi - \eta., \quad \phi_2 = \xi, \quad \phi_3 = \eta.$$

Global element in Figure 7

$$\phi_1 = 1 - x_1 - x_2, \ \nabla\phi_1 = (-1,-1), \quad \phi_2 = 1 - x_2, \quad \nabla\phi_2 = (0,-1), \quad \phi_3 = 1 + x_1, \ \nabla\phi_3 = (1,0)$$

$$\phi_4 = 1 + x_1 + x_2, \ \nabla\phi_4 = (1,1), \quad \phi_5 = 1 + x_2, \quad \nabla\phi_2 = (0,1), \quad \phi_6 = 1 - x_1, \ \nabla\phi_3 = (-1,0) \tag{3.1}$$

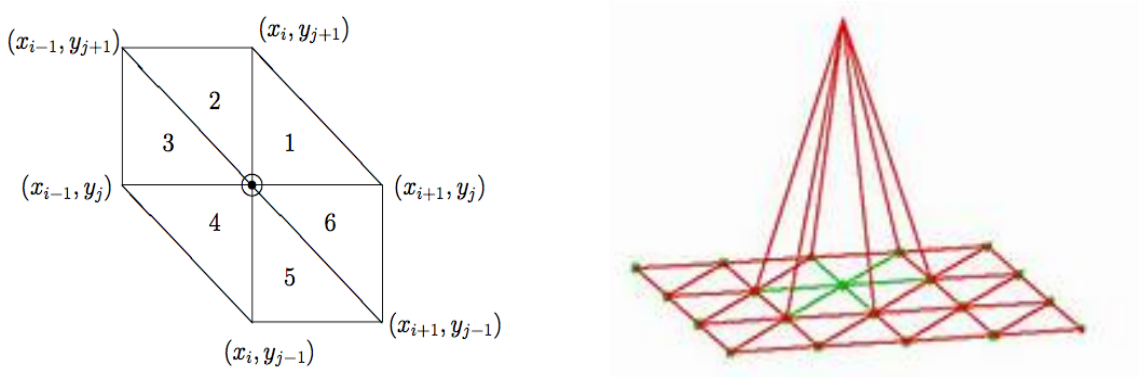

Figure 7: Triangle element

Let $\phi_{i,j}(x_1, x_2)$ be the global event at node $(i/N, j/N)$ and

$$u^N(x_1, x_2) = \sum_{1 \le \le N-1} \sum_{1 \le j \le N-1} u_{i,j}\, \phi_{i,j}(x_1, x_j) \in H_0^1((0,1) \times (0,1)$$

Then, from the weak form for Poisson equation $-\Delta u = f$ in $\Omega = (0,1) \times (0,1)$

$$\int_\Omega \nabla u^N \cdot \nabla \phi_{i,j}\, dx = \int f(x)\phi_{i,j}(x)\, dx$$

Exercise 12 Show that for the uniform triangulation on the square $(0,1) \times (0,1)$ ((3.1) and Figure 7) with mesh size $h = \frac{1}{N}$ we obtain

$$-(\Delta_h u)_{i,j} = \frac{4u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1}}{h^2} = f_{i,j} \sim f(x_i, y_j). \tag{3.2}$$

Thus, the stiffens matrix H in the column-wise order of $u_{i,j}$ is a tri-diagonal block matrix of

the form

$$H = \begin{bmatrix} H_0 & -I & 0 & \cdots & \\ -I & H_0 & -I & 0 & \cdots \\ & \ddots & \ddots & \ddots & \\ \cdots & 0 & -I & H_0 & -I \\ & \cdots & 0 & -I & H_0 \end{bmatrix}, \quad H_0 = \begin{bmatrix} 4 & -1 & 0 & \cdots & \\ -1 & 4 & -1 & 0 & \cdots \\ & \ddots & \ddots & \ddots & \\ \cdots & 0 & -1 & 4 & -1 \\ & \cdots & 0 & -1 & 4 \end{bmatrix}.$$

Local Quadratic element on the right triangle Let consider local triangle element as in Figure 8 and the six local quadratic element are given by

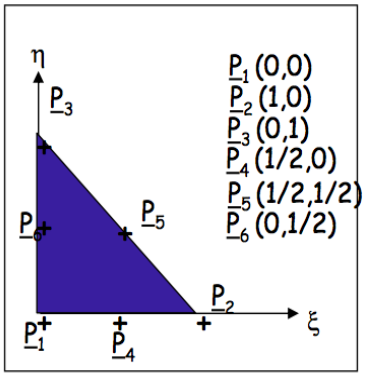

Figure 8: Quadratic element

$$N_1 = (1-\xi-\eta), \quad N_2 = \xi(2\xi-1), \quad N_3 = \eta(2\eta-1)$$

$$N_4 = 4\xi(1-\xi-\eta), \quad N_5 = 4\xi\eta, \quad N_6 = 4\eta(1-\xi-\eta)$$

Local element on the square (quadrilateral)

$$\phi_1 = (1-\xi)(1-\eta), \quad \phi_2 = (1-\xi)\eta, \quad \phi_3 = \xi(1-\eta), \quad \phi_4 = \xi\eta$$

Tensor products

$$\phi_{i,j}(x_1, x_2) = B_i^N(x_1) B_j^N(x_2).$$

3.1 Iso-parametric Curved element

Consider the coordinate transformation form $x, yto(\xi, \eta)$

$$\xi = T_1(x, y), \quad \eta = T_2(x, y)$$

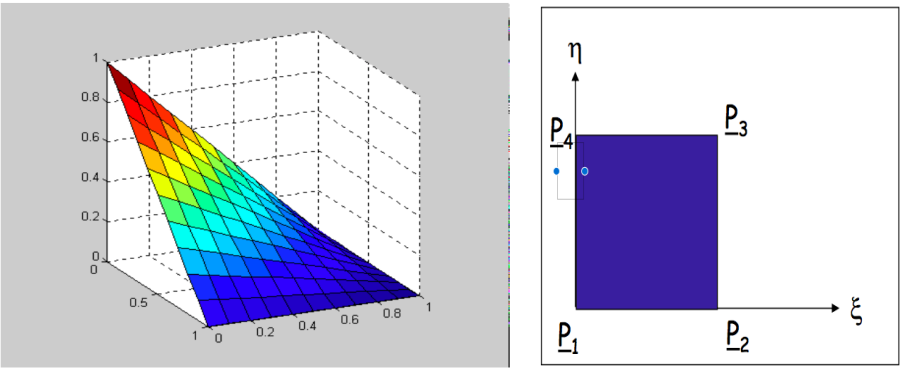

Figure 9: Quadrateral element

and define the local basis function:

$$\phi_i(x,y) = \psi_i(T_1(x,y), T_2(x,y)),$$

where $\{\psi_i(\xi,\eta)\}$ is the master basis on the master element. For example, the linear transform from a triangle with vertices (x_1,y_1), (x_2,y_2) and (x_3,y_3) in the counter-clockwise direction to the master triangle is given by

$$\xi = \tfrac{1}{2A}\left((y_3-y_1)(x-x_1) - (x_3-x_1)(y-y_1)\right.$$

$$\eta = \tfrac{1}{2A}\left((y_2-y_1)(x-x_1) - (x_2-x_1)(y-y_1)\right)$$

where $A = \dfrac{(y_3-y_1)(x_2-x_1)-(x_3-x_1)(y_2-y_1)}{2}$ is the area of the triangle

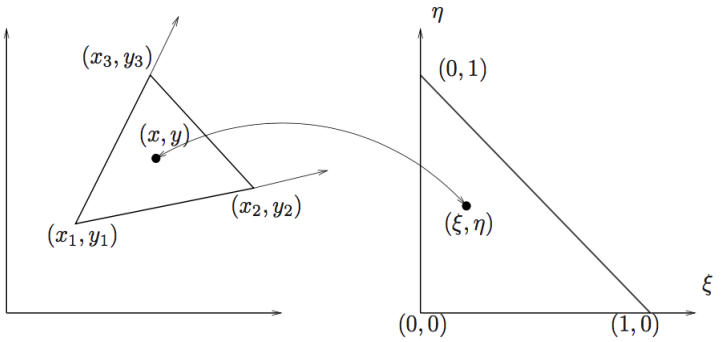

Figure 10: The linear transform from an arbitrary triangle to the standard triangle (master element) and the inverse map.

Then, the stiffness is

$$\int_\Omega \nabla\phi_i \nabla\phi_j = \int_{\bar{\Omega}} (J\nabla\psi_i)\cdot(J\nabla\psi_j)\,\left|\frac{\partial(x,y)}{\partial(\xi,\eta)}\right| d\xi d\eta. \tag{3.3}$$

where the Jacobean J is defined by

$$J = \begin{pmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \eta}{\partial x} \\ \\ \frac{\partial \xi}{\partial y} & \frac{\partial \eta}{\partial y} \end{pmatrix}$$

3.2 Quadrature rules

A quadrature formula has the form

$$\int_{\bar{\Omega}} g(\xi, \eta)\, d\xi d\eta = \sum_{k=1}^{L} w_k g(\xi_k, \eta_k)$$

where $\bar{\Omega}$ is the standard right triangle and L is the number of points involved in the quadrature. Below we list some commonly used quadrature formulas in 2D using one, three and four points. The geometry of the points are illustrated in Fig. 11, and the coordinates of the points and the weights are given in Table. It is noted that only the three-point quadrature formula is closed, since the three points are on the boundary of the triangle, and the other quadrature formulas are open.

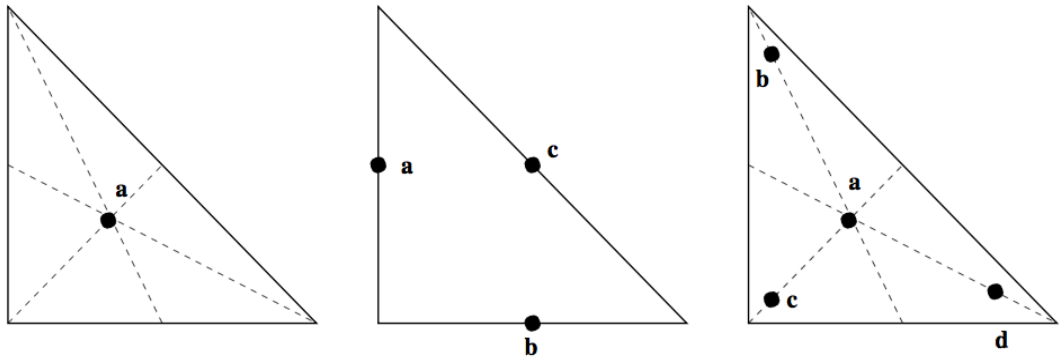

Figure 11: A diagram of the quadrature formulas in 2D with one, three and four quadrature points, respectively.

A triangulation is determined by its elements and nodal points. We use the following notation:

- Triangular elements K_j is sorted and the nodal points (x_i, y_i) is ordered.
- Coordinate of nodal points $Node(1:3, j)$ for each jth triangle, i.e. A 2D array nodes A: The first index is the index of nodal point in an element j, (in the counter-clockwise) and the second index is the index of the element.

Element wise assembling: Sequentially, in the order of element K_j we evaluate

$$s_{k,i} = \int_{K_j} \nabla \phi_{k,j} \cdot \nabla \phi_{l,j}\, dxdy \text{ for } (k,i) \in Node(:, j)$$

L	Points	(ξ_k, η_k)	w_k
1	a	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\frac{1}{2}$
3	a	$\left(0, \frac{1}{2}\right)$	$\frac{1}{6}$
	b	$\left(\frac{1}{2}, 0\right)$	$\frac{1}{6}$
	c	$\left(\frac{1}{2}, \frac{1}{2}\right)$	$\frac{1}{6}$
4	a	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$-\frac{27}{96}$
	b	$\left(\frac{2}{15}, \frac{11}{15}\right)$	$\frac{25}{96}$
	c	$\left(\frac{2}{15}, \frac{2}{15}\right)$	$\frac{25}{96}$
	d	$\left(\frac{11}{15}, \frac{2}{15}\right)$	$\frac{25}{96}$

Figure 12: A diagram of the quadrature formulas in 2D with one, three and four quadrature points, respectively.

based on the iso-parametric formula (3.3). Then accumulate

$$H_{k,i} = H_{k,i} + s_{k,i}$$

3.3 Interpolation Error Analysis

A global interpolation function is defined as

$$v_I(x, y) = \sum_{i:allnodes} v(a_i)\, \phi_i(x, y), \quad a_i = (x_i, y_i)$$

where $\phi_i(x, y)$ is the piecewise linear function that satisfies $\phi_i(a_j) = \delta_{ij}$.

Theorem If $u \in C^2(K)$, then we have an error estimate for the interpolation

$$|v - v_I|_\infty \leq 2h^2 \max_{|\alpha|=2} |D^\alpha v|_\infty$$

Furthermore, we have

$$\max |\nabla(v - v_I)|_\infty \leq \frac{4h^2}{\rho} \max_{|\alpha|=2} |D^\alpha v|_\infty$$

where $\rho > 0$ is the radius of the largest ball contained in the triangle element K.

$$
\begin{aligned}
nodes(1,1) &= 5\,, \quad (x_5, y_5) = (0, h)\,,\\
nodes(2,1) &= 1\,, \quad (x_1, y_1) = (0, 0)\,,\\
nodes(3,1) &= 6\,, \quad (x_6, y_6) = (h, h)\,,\\
\\
nodes(1,10) &= 7\,, \quad (x_7, y_7) = (2h, h)\,,\\
nodes(2,10) &= 11\,, \quad (x_{11}, y_{11}) = (2h, 2h)\,,\\
nodes(3,10) &= 6\,, \quad (x_6, y_6) = (h, h)\,.
\end{aligned}
$$

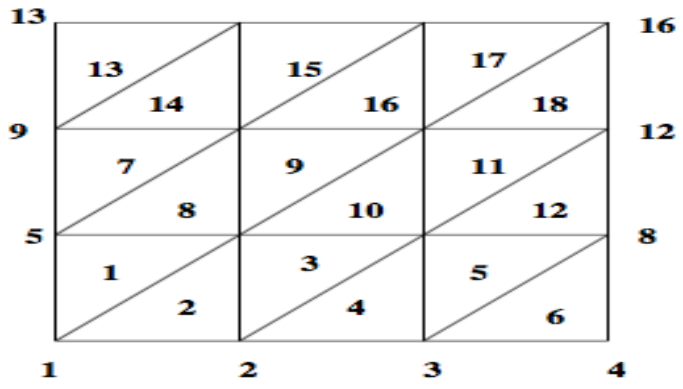

Figure 13: A simple triangulation with the row-wise natural ordering.

Proof: From the definition of the interpolation function and the Taylor expansion

$$v_I = \sum_{i=1} u(x_i)\phi_i(x,y) = \sum_{i=1}^{3} \phi_{x,y}(u(x,y) + \nabla u(x,y)\cdot(x_i - x, y_i - y) + \frac{1}{2}(x_i - x, y_i - y)^t D^2(\xi,\eta)(x_i - x)$$

where (ξ, η) is a point in the triangle K. The claim thus follows from the fact that $\phi(x,y) \geq 0$ and $\sum \phi_i(x,y) = 1$. $\square$

4 Parabolic system

Conservation law

$$\frac{d}{dt}\int_\Omega q(x,t)\,dx = -\int_{\partial\Omega} \vec{f}(x,t)\,ds = \text{ net flux throgh the boundary,}$$

where

$$Q(t) = \int_\Omega q(x,t)\,dx$$

is the total energy on am arbitrary volume and $\vec{f}(x,t)$ is a flux function. By divergence theorem

$$\int_\omega (\frac{\partial}{\partial t} q(x,t) + \text{div}\vec{f}(x,t))\,dx = 0$$

and we have the conservation

$$\frac{\partial}{\partial t}q(x,t) + \mathrm{div}\vec{f}(x,t)) = 0$$

If $q = u$, then the mass conservation is

$$\frac{\partial}{\partial t}u(x,t) + \mathrm{div}\vec{f}(x,t)) = 0$$

with specified flux $f = \vec{b}(x)u(1-u)$ (traffic flow) $f = \vec{b}\frac{u^2}{2}$ (Burgers flow) and $f = (\vec{c}(x,t)u)$ (continuity). If $q(x,t) = \rho c_f u(x,t)$ is the energy density where ρ is the mass density and c_f is the specific heat and assume the Fourier law $\vec{f} = K(x)\nabla u(x,t)$ we obtain the heat conduction equation

$$\frac{\partial \rho c_f\, u(x,t)}{\partial t} = \nabla(K(x)\nabla u(x,t)$$

where $K(x)$ is a matrix conductivity. It includes the heat equation of the form

$$\frac{\partial}{\partial t}u(x,t) = \Delta u_(a,t), \quad u(0,x) = u_0(x) \in L^2(\Omega) = H.$$

In general we can formulate the parabolic equation for $u = u(x,t),\ x \in \Omega,\ t > 0$

$$(u_t, \psi) + a(u(t), \psi) = (f(t), \psi), \quad u(0) = u_0 \in H, \tag{4.1}$$

where $u(t) = u(:.t) \in X$ and the bounded bilinear form a on $X \times X$ as in the Lax-Mailgram theorem satisfies

$$a(u,u) \geq \delta\,|u|_X^2 - \rho\,|u|_H^2.$$

Letting $\psi = u(t)$ in (4.1), we have

$$\int_\Omega \frac{\partial u(t,x)}{\partial t}u(t,x)\,dx + a(u(t),u(t)) = (f(t),u(t))$$

where

$$\int_\Omega \frac{\partial u(t,x)}{\partial t}u(t,x)\,dx = \frac{d}{dt}\int_\Omega \frac{1}{2}|u(t,x)|^2\,dx$$

and

$$a(u(t),u(t)) + (f(t),u(t)) \geq \delta\,|u(t)|_X^2 - \rho\,|u(t)|_H^2 - |f(t)|_{X^*}|u(t)|_X$$

$$\geq \frac{\delta}{2}\,|u(t)|_X^2 - \rho\,|u(t)|_H^2 - \frac{1}{2\delta}|f(t)|_{X^*}^2,$$

where we used the Young inequality

$$ab \leq \frac{\delta}{2}|a|^2 + \frac{1}{2\delta}|b|^2.$$

By integrating this in time, we obtain

$$|u(t)|_H^2 - |u(0,x)|_H^2 + \int \delta\,|u(s)|^2 x\,ds \leq \int_0^t (2\rho\,|u(s)|_H^2 + \frac{1}{\delta}\,|f(s)|_{X^*}^2)\,ds.$$

By Gronwall?s inequality we have the estimate

$$|u(t)|_H^2 + \int_0^t \delta\,|u(s)|_X^2\,ds \le e^{2\rho t}\,|u_0|_H^2 + \int_0^t e^{2\rho(t-s)}\frac{1}{\delta}\,|f(s)|_{X^*}^2)\,ds.$$

Gronwall's inequality If $t \to e(t) \in R$ satisfies

$$0 \le e(t) \le F + \int_0^t c(s)e(s)\,ds,$$

then

$$e(t) \le F\exp(\int_0^t c(s)\,ds).$$

Proof: Let

$$0 \le e(t) \le f(t) = F + \int_0^t c(s)e(s)\,ds$$

Then,

$$f'(t) = c(t)e(t) \le c(t)f(t)$$

and

$$(f(t)e^{-\int_0^t c(s)\,ds})' \le 0$$

Integrating it in time, we obtain

$$e(t) \le f(t) \le F\,e^{\int_0^t c(s)\,ds}.\square$$

Example (Fokker-Planck equation)

$$u_t = \nabla\cdot(A(x)\nabla u + \vec{b}(x)u) - q(x)u(t.x), \quad u(0,x) = u_0,$$

where $u(t,x) \ge 0$ represents the probability density function, where $A \in R^{n\times n}$ is a symmetric positive definite matrix, $\vec{b}(x)$ is the drift vector and $q(x)$ is the potential function. One can formulate it as (4.1) with

$$a(u,\psi) = \int_\Omega (A(x)\nabla u(x) + \vec{b}(x)u(x), \nabla\psi) + (q(x)u(x), \psi(x))\,dx$$

Exercise 14 Consider a system of diffusion equations for $u(t,x), v(t,x),\ x \in \Omega$:

$$\begin{pmatrix} \frac{\partial u}{\partial t} \\ \\ \frac{\partial v}{\partial t} \end{pmatrix} = \mathrm{div}D(x)\begin{pmatrix} \nabla u \\ \\ \nabla v \end{pmatrix} + \mathrm{div}B(x)\begin{pmatrix} u \\ \\ v \end{pmatrix} + C(x)\begin{pmatrix} u \\ \\ v \end{pmatrix} = \vec{f}(t,x)$$

$$\begin{pmatrix} n \\ \\ n \end{pmatrix}\cdot\left(D(x)\begin{pmatrix} \nabla u \\ \\ \nabla v \end{pmatrix}\right) + E(x)\begin{pmatrix} u \\ \\ v \end{pmatrix} = \vec{g}(x) \text{ at boundary } \partial\Omega$$

where $D(x)$ is 4 by 4 matrix, $D = D^t > 0$ and $B(x),\ C(x),\ E(x)$ are 2 by 2 matrices. Setup it as (4.1).

4.1 Time integration

Consider an ordinary differential equation

$$\frac{d}{dt}x(t) = f(x(t), t), \quad x(0) = x_0.$$

Integrating this in time

$$x(t_n) - x(t_{n-1}) = \int_{t_{n-1}}^{t_n} f(x(s), s)\, ds.$$

We apply various quadrature rule to integrate the RHS, i.e. including

$$\int_{t_{n-1}}^{t_n} f(x(s), s)\, ds \sim \begin{cases} f(x(t_{n-1}), t_{n-1})\, \Delta t \text{ (explicit)}. \\ f(x(t_n), t_n)\, \Delta t \text{ (implicit)}. \\ \frac{1}{2}(f(x(t_{n-1}), t_{n-1}) + f(x(t_n), t_n))\, \Delta t \text{ (Crank-Nicolson)}. \end{cases}$$

Implicit in time (First order): $u^n \in X_h$ satisfies

$$(\frac{u_h^n - u_h^{n-1}}{\Delta t}, \psi_h) + a(u_h^n, \psi_h) = (f, \psi_h) \text{ for all } \psi_h \in X_h \tag{4.2}$$

Crank-Nicolson scheme: (Second order) is given by

$$(\frac{u_h^n - u_h^{n-1}}{\Delta t}, \psi_h) + a(\frac{u_h^n + u_h^{n-1}}{2}, \psi_h) = (f^{n+\frac{1}{2}}, \psi_h) \text{ for all } \psi_h \in X_h$$

Explicit time-integration is conditionally stable, i.e., the time wise step $\Delta t > 0$ depends on space-wise mehsize $h > 0$ and $\Delta t > 0$ need be chosen very small. Implicit scheme is unconditionally stable but it is of the first order. In general one can use the Runge-Kutta method and Backward difference schemes and Adams-Burshforth methods for ordinary differential equations for the space discretized problem.

Stability of Implicit scheme Consider the Galerkin approximation of $u(t) \in X$:

$$u_h(t) = \sum_k^N u_k(t)\, \phi_k(x) \in X_h$$

and let A_h is the stiffness matrix defined by

$$(A_h)_{ji} = a(\phi_i, \phi_j)$$

The, implicit-explicit scheme (4.2) is written as

$$\frac{u_h^n - u^{n-1}}{\Delta t} + A_h u_h^n = f_h^n$$

$u_h^n \in X_h$ is well-defend by the Laxi-Milgram applied to

$$a_{\Delta t}(u,v) = \frac{1}{\Delta t}(u,v)_h + a(u,v).$$

Thus, u_h^n is given by

$$u_h^n = (I + \Delta\, A_h)^{-1})u_h^{n-1} + \Delta t\, f_h^n).$$

Note that

$$(u_h^n - u_h^{n-1}, u_h^n) = \tfrac{1}{2}(|u_h^n|_H^2 - |u_h^{n-1}|_H^2 + |u_h^n - u_h^{n-1}|_H^2)$$

$$a(u_h^n, u_h^n) \geq \delta\, |u_h^n|_X^2 - \rho\, |u_h^n|_H^2.$$

Letting $\psi_h = u_h^n$, we obtain

$$|u^n|_H^2 + \sum_{k=1}^{n} \delta\, |u^k|^2 \Delta t + |u_h^k - u_h^{k-1}|^2 \leq |u_h^0|_H^2 + \sum_{k}^{n} (2\rho\, |u^k|_H^2 + \frac{1}{\delta}|f^k|_{X^*}^2)\Delta t.$$

By Gronwall?s inequality we have the stability estimate

$$|u^n|_H^2 + \sum_{k=1}^{n} \delta\, |u^k|^2 \Delta t + |u_h^k - u_h^{k-1}|^2 \leq M\, (|u_h^0|_H^2 + \sum_{k}^{n} + \frac{1}{\delta}|f^k|_{X^*}^2)\Delta t).$$

for $M \geq 0$ independent of $\Delta t,\ h$.

4.2 Semilinear Equation

Consider the semilinear equation (e.g., Navier-Stokes equation)

$$(u_t, \psi) + a_0(u(t), \psi) + a_1(u(t), \psi) = (f, \psi)$$

where a_0 is $X-$elliptic and $u \in X \to a_1(u, \psi)$ is nonlinear, the Explicit-Implicit second order method is given by

$$(\frac{u^n - u^{n-1}}{\Delta t}, \psi_h) + a_0(\frac{u^n + u^{n-1}}{2}, \psi_h) + \frac{1}{2}(a_1(u^{n-1}, \psi_h) - a_1(u^{n-2}, \psi_h)) = (f, \psi_h)$$

That is, the linear part is treated by the Crank-Nicolson scheme and the nonlinear part is evaluated by the explicit second order scheme in time.

Example (Navier Stokes equations The incompressible Navier Stokes equation is

$$\vec{u}_t + \vec{u} \cdot \nabla \vec{u} + \nabla p = \mu \Delta \vec{u} + \vec{f}(t.x)$$

$$\nabla \cdot \vec{u}(t) = 0$$

where $\vec{u}(t,x)$ is the velocity field and $p(t,x)$ is the pressure and

$$\vec{u} \cdot \nabla \vec{u} = \begin{pmatrix} u_1 \frac{\partial u_1}{\partial x_1} + u_2 \frac{\partial u_1}{\partial x_2} \\ \\ u_1 \frac{\partial u_2}{\partial x_1} + u_2 \frac{\partial u_2}{\partial x_2} \end{pmatrix}.$$

For $\vec{\psi} \in X = \{H_0^1(\Omega)^d : \nabla \cdot \vec{\psi} = 0\}$ $\vec{u} \in X$ satisfies

$$(\vec{u}_t, \vec{\psi}) + a(\vec{u}, \vec{\psi}) + b(\vec{u}(t), \vec{u}(t), \vec{\psi}) = (\vec{f}(t), \vec{\psi})$$

where the bilinear form $a : X \times X \to R$ is given by

$$a(\vec{u}, \vec{\psi}) = \int_\Omega \mu(\nabla \vec{u}, \nabla \vec{\psi})\, dx$$

and the tri-linear form $b : X \times X \times X \to R$ is given by

$$b(\vec{u}, \vec{v}, \vec{\psi}) = \int_\Omega (\vec{u} \cdot \nabla \vec{v}, \vec{\psi})\, dx.$$

Here, we used

$$(\nabla p, \vec{\psi}) = -(p, \nabla \cdot \vec{\psi}) = 0 \text{ for } \vec{\psi} \in X.$$

In the case of non confomal $u_h \notin X$ we use the mixed finite element formulation (2.19) augmenting the pressure for the Stokes equation based on the Taylor-Hood elements.

Time-Splitting mehod One can apply the time-splitting method for mixed operators case

$$u_t = A_0 u + A_1(u),$$

for example

$$u_t = k\,\Delta u + f(x, u) \text{ (nonlinear reaction)}$$

$$u_t = \epsilon\,\Delta u - \vec{b}(x) \cdot \nabla u \text{ (convection dominated).}$$

by the operator splitting

$u_t = k\,\Delta u$ and then $\frac{d}{dt}u(x,t) = f(x, u(x,t))$ pointwise ODEs foe each $x \in \Omega$

$u_t = k\,\Delta u$ and then $u_t + \vec{b}(x) \cdot \nabla u$ via the characteristic method.

5 (Abstract) Wave equation

By the Newton?s law of the motion

$$\rho(x) u_{tt}(x,t) = \text{div}\,(\sigma(x,t))$$

where ρ is the mass density and $\sigma(x,t)$ is the stress. For example,

$$\sigma(x,t) = C(x)\nabla u(x,t) + D(x)\nabla u_t(x,t)$$

for Kelvin-Vogt model, where $C(x)$ is the compliance and $D(x)$ is the damping rate. Let $D = 0$ $C(x) = 1$, we have the linear wave equation

$$\rho(x)\, u_t = \Delta u(x,t).$$

Consider the damped wave equation of the form

$$(\rho(x)u_{tt}, \psi) + d(u_t, \psi) + a(u(t), \psi) = (f, \psi) \text{ for all } \psi \in X, \tag{5.1}$$

where $\rho(x) > 0$ is the mass density, a is a bilinear symmetric stiffness form and d is a bilinear damping form on $X \times X$.

Example

$$\rho(x)\, u_{tt} = \Delta u \text{ in } \Omega, \quad \frac{\partial u}{\partial n} = \alpha \frac{\partial u}{\partial t} \text{ at boundary } \partial\Omega$$

Since

$$\int_\Omega \Delta u \psi \, dx = \int_{\partial\Omega} -\alpha \, u_t \psi \, ds - \int_\Omega \nabla u \cdot \nabla \psi \, dx$$

we have

$$a(u, \psi) = \int_\Omega \nabla u \cdot \nabla \psi \, dx, \quad d(u_t, \psi) = \int_{\partial\Omega} \alpha \, u_t \psi \, ds.$$

Assume a is $X-$coercive and $d(u, u) \geq 0$. Letting $\psi = u_t$ in (5.1)

$$\int_\Omega \rho(x) u_{tt}, u_t \, dx + d(u_t, u_u) + a(u(t) U_t) = (f, u_t)$$

where

$$\int_\Omega \rho(x) u_{tt}, u_t \, dx + a(u(t), u_t) = \frac{d}{dt} E(t)$$

and the total energy E is given by

$$E(t) = \frac{1}{2}(\int_\Omega \rho(x) |u_t(t, x)|^2 \, dx + a(u(t), u(t))).$$

Thus, we have the energy conservation:

$$E(t) + \int_0^t d(u_t(s), u_t(s)) \, ds = E(0) + \int_0^t (d(t), u_t) \, dt.$$

For the wave equation based on the linear triangular element we obtain the fully explicit discretization of the form

$$\frac{u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1}}{\Delta t^2} = c_{i,j}^2 \, (\Delta_h u^n)_{i,j}$$

where Δ_h is the central difference approximation of Δ.

Von-Neumann stability analysis Let c be a constant and $\Omega = (0, 1) \times (0, 1)$. It can be shown that

$$e_{k,\ell} = \sin(\frac{k\pi}{N} x_1) \sin(\frac{\ell\pi}{N} x_2)$$

is an eigenvector of Δ_h corresponding to the eigenvale

$$\mu_{k,\ell} = \frac{2(\cos(\frac{k\pi}{N}) - 1) + 2(\cos(\frac{\ell\pi}{N}) - 1)}{h^2}, \quad h = \frac{1}{N}$$

In fact, it follows from

$$h^2\Delta_h e^{i\frac{k\pi}{N}x_1}e^{i\frac{\ell\pi}{N}x_2} = (e^{i\frac{k\pi}{N}} + e^{-i\frac{k\pi}{N}} - 2 + e^{i\frac{\ell\pi}{N}} + e^{-i\frac{\ell\pi}{N}} - 2)\, e^{i\frac{k\pi}{N}x_1}e^{i\frac{\ell\pi}{N}x_2}.$$

For the (k,ℓ) mode let λ be the magnification factor in time, $u^n = \lambda u^n - 1$. Then

$$\lambda - 2 + \frac{1}{\lambda} = c^2\frac{\Delta t^2}{h^2}(2(\cos(\frac{k\pi}{N}) - 1) + 2(\cos(\frac{\ell\pi}{N}) - 1)).$$

Thus, $|\lambda| \le 1$ (stable) provided that $4\,c^2\frac{\Delta t^2}{h^2} \le 1$.

6 Concrete Examples

6.1 Scalar elliptic equation

A scalar function $u(x) \in H^1(\Omega)$ satisfies

$$-\nabla\cdot(A(x)\nabla u) + b(x)\cdot\nabla u + c(x)u = f.$$

6.2 Elastic equation

A vector function $u(x) \in H^1(\Omega)^n$ satisfies

$$-Div\,\sigma = \vec{f}$$

where the stress tensor σ is given by

6.3 Maxwell equation

6.4 Hyperbolic systems and Conservation law

Consider the scalar conservation law

$$u_t + (f(x,u))_x = 0$$

The discontinuous Galelkin method is given as follows. We define a finite element space consisting of piecewise polynomials

$$V_h^k = \{v|_{I_i} \in P^k(I_i),\ 1 \le i \le B\},$$

where $P^k(I_i)$ denotes the set of polynomials of degree up to k defined on the cell $I_i = (x_{i-1/2}, x_{i+1/2})$. Testing against $v_h \in V_h^k$ we have

$$\int_{I_i}(u_h)_t v_h\,dx - \int_{I_i} f(u_h)(v_h)_x\,dx + \hat{f}_{i+1/2}v_h(x_{i+1/2}) - \hat{f}_{i-1/2}v_h(x_{i-1/2}) = 0.$$

where $\hat{f}_{i+1/2}$ is the numerical flux, which is a single valued function defined at the cell interfaces and in general depends on the values of the numerical solution uh from both sides of the interface

$$\hat{f}_{i+1/2} = f(u_h(x^-_{i-1/2}, t), u_h(x^-_{i-1/2}, t))$$

We use the so-called monotone fluxes from finite difference and finite volume schemes for solving conservation laws, which satisfy the following conditions:

- Consistency: $\hat{f}(u,u) = f(u)$.
- Continuity: $\hat{f}(u^-,u^+)$ is at least Lipschitz continuous with respect to both arguments u^- and u^+.
- Monotonicity: $\hat{f}(u^-,u^+)$ is a non-decreasing function of its first argument u^- and non-increasing function of its second argument u^+.

Well known monotone fluxes include the Lax-Friedrichs flux

$$\hat{f}^{LF}(u^-,u^+) = \frac{1}{2}(f(u^-)+f(u^+)) - \alpha\,(u^+-u^-), \quad \alpha = \max_u |f'(u)|$$

the Godunov flux

$$\hat{f}^{LF}(u^-,u^+) = \begin{cases} \min_{u^-\le u\le u^+} f(u) & \text{if } u^- < u^+ \\ \\ \max_{u^+\le u\le u^-} f(u) & \text{if } u^+ \le u^- \end{cases}$$

and the Engquist-Osher flux

$$\hat{f}^{EO}(u^-,u^+) = \int_0^{u^-} \max(f'(u),0)\,du + \int_0^{u^+} \in (f'(u),0)\,du = f(0)$$

6.5 Hamilton-Jacobi-Bellman equation

Consider the Hamilton-Jacobi-Bellman equation

$$v_t + H(x,v_x) = \epsilon\,\Delta v.$$

7 Discontinuous Galerkin method foe Elliptic equation

We present the Discontinuous Galerkin method for

$$-\nabla\cdot(a(x)\nabla u) = f \tag{7.1}$$

Let Ω_h is a trigulation of Ω and let X_h be the piecewise $H^1(\Omega)$, i.e.

$$X_h = \{u \in H^1(E) \text{ on each element } E \text{ of } \Omega_h\}.$$

If u belongs to X_h, then the trace of u on any side of one element E is well defined. If two elements E_1 and E_2 are neighbors and share one common side Γ, there are two traces of u on Γ from the both side. We define average and jump for u. We assume that the normal vector n is oriented from E_1 to E_2.

$$\{u\} = \frac{1}{2}(u|_{\Gamma^-} + u|_{\Gamma^+}), \quad [u] = u|_{\Gamma^+} - u|_{\Gamma^-}.$$

Note that for $\psi \in X_h$

$$\int_E -\nabla\cdot(a(x)\nabla u\psi\,dx = \int_\Gamma a(x)\frac{\partial u}{\partial n}[\psi] + \int_E a(x)\nabla u\cdot\nabla\psi\,dx$$

where we have continuity of the flux at Γ for the solution u

We now define the DG bilinear forms a_h on $X_h = H^1(\Omega_h)$ by

$$a_h(u,v) = \sum_i \int_{\Omega_i} a(x)\nabla u \cdot \nabla v \, dx - \int_{\Gamma_i} \{an \cdot \nabla u\}[v] \pm \{an \cdot \nabla v\}[u] \, ds$$

$$+\alpha_1 \sum_i \int_{\Gamma_i} [u][v] \, ds + \alpha_2 \sum_i \int_{\Gamma_i} [an \cdot \nabla u][(an \cdot \nabla v] \, ds$$

where the last two terms are the penalizing jumps.

7.1 Immersed finite elements method (IFEM)

Consider (7.1) with a discontinuous $a(x)$, i.e.,

$$a(x) = \begin{cases} \beta^+ & x \in \Omega^+ \\ \\ \beta^- & x \in \Omega^-. \end{cases}$$

We have the continuity condition at interface Γ

$$[u]_\Gamma = 0 \text{ and } [\beta \frac{\partial u}{\partial n}]_\Gamma = 0.$$

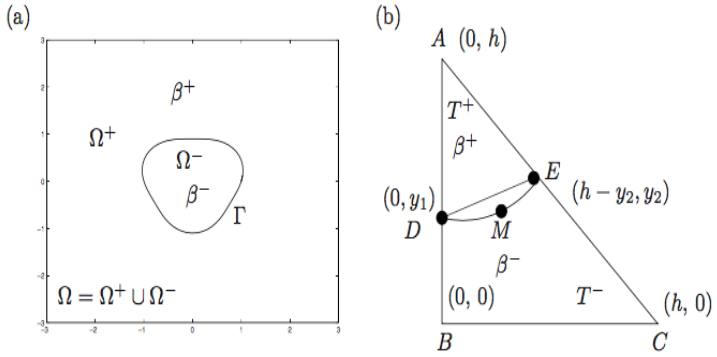

Figure 14: (a) a configuration of a rectangular domain $\Omega = \Omega^+ \cup \Omega^-$ with an interface Γ from an IFEM test. The coefficient $a(x)$ may have a finite jump across the interface Γ. (b) an interface triangle and the geometry after transformed to the standard right triangle.

We construct a piecewise linear function on the triangle based on given the values at the three vertices and that satisfies the natural jump condition. Assume that the values at vertices A, B, and C of the element T are specified, we construct the following piecewise linear function as in Figure 14:

$$u^{\pm}(x) = a_0^{\pm} + a_1^{\pm} \, x_1 + a_3^{\pm} \, x_2$$

satisfying

$$u^+(D) = u^-(D), \quad u^+(E) = u^-(E), \quad \beta^+ \frac{\partial}{\partial n} u^+(M) = \beta^- \frac{\partial}{\partial n} u^-(M)$$

where n is the unit normal direction of the line segment DE. Thus, there are six constraints and six coefficients, so we show that the solution exists and is unique.

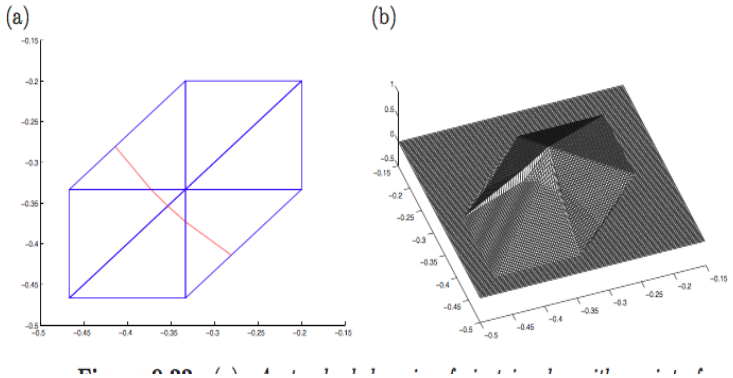

Figure 15: (a). A standard domain of six triangles with an interface cutting through. (b). A global basis function on its support of the non-conforming immersed finite element space. The basis function has small jump across some edges.

8 Finite difference and Finite volume and Spectral methods in view of FEA

One can derive the standard and more advanced highbrid methods of finite difference and finite volume method in terms of the weak formulation. The methods use the nodal values of solution and thus one can incoorpolate the local basis function in terms of the finite element interpolation. The error analysis can be formulated for Finite difference and Finite volume methods in terms of Lax equivalence, e.g., stability and consistency imply the convergence. If we use the operator splitting method, each step of the integration can be performed independently via nodal interpolation between FEM and finite difference (volume) methods.

8.1 Elliptic System

8.2 Hyprbolic System

8.3 Hamilton-Jacobi-Beelman equation

8.4 Spectral element method via discontinuous Galerkin

9 Appendix