Numerical Methods II, Courant Institute, Spring 2024
http://www.math.nyu.edu/faculty/goodman/teaching/NumericalMethodsII2024.html
Jonathan Goodman, January, 2024

Assignment 1, Finite Element Methods,

due February 24, 5:15 pm.

Finite element methods are commonly used to solve elliptic PDEs. This assignment goes through some of finite element basics, as applied to the inhomogeneous Laplace equation on the unit square in 2D. Unfortunately, this obscures one of the main reasons finite element methods are used in practice (they are a billion dollar industry), their geometric flexibility and ability to handle PDEs other than the Laplace equation. Please pay attention to notes on individual exercises that explain where this generality comes from.

Take Ω to be the unit square in 2D and suppose $u(x,y)$ and $f(x,y)$ are defined in Ω. The boundary of Ω is Γ. Consider the problem of finding u so that

$$\triangle u = f \; , \quad \text{for } (x,y) \in \Omega \; , \;\; u = 0 \text{ on } \Gamma \; .$$

This is the *inhomogeneous Dirichlet* problem. *Inhomogeneous* means that f is not zero. *Dirichlet* means that the *boundary condition* is that $u = 0$ on Γ. The *variational principle* for this is that u is the minimizing function of

$$\min_{u \in H_0^1} \left\{ \tfrac{1}{2} \int_\Omega \left|\nabla u(x,y)\right|^2 dxdy \; + \; \int_\Omega f(x,y)\, u(x,y)\, dxdy \right\} \; . \tag{1}$$

The space H_0^1 is an example of *Sobolev space*. This one is defined by $u \in H_0^1$ if

$$\int_\Omega \left|\nabla u(x,y)\right|^2 dxdy < \infty \; , \quad \text{and } u = 0 \text{ for } (x,y) \in \Gamma \; .$$

This is a vector space (in the sense of linear algebra) because if u and v both have these properties (finite *Dirichlet integral* and *Dirichlet* boundary conditions), then so does any linear combination $au(x,t) + bv(x,y) = w(x,y)$. The superscript 1 means one derivative (H^2 requires second derivatives to be square integrable) and the subscript 0 refers to the boundary condition.

There are mathematical subtleties that we will ignore here but you can read about in a PDE book concerning H_0^1. One is that u can be in H_0^1 without being differentiable at every point. For example, the function

$$u(x,y) = \left(\tfrac{1}{2} - \left|x - \tfrac{1}{2}\right| \right) \left(\tfrac{1}{2} - \left|y - \tfrac{1}{2}\right| \right) \; . \tag{2}$$

This function has $|\nabla u(x,y)| < 1$ for all $(x,y) \in \Omega$ where the derivative is defined, but u is not differentiable when $x = \frac{1}{2}$ or $y = \frac{1}{2}$. It also has $u = 0$ on Γ, so maybe it should count as being in H_0^1. Consider, on the other hand, the function

$$u(x,y) = y(1-y) \cdot \begin{cases} x & \text{if } 0 \leq x < \frac{1}{2} \\ 0 & \text{if } x \geq \frac{1}{2} \; . \end{cases} \tag{3}$$

This function also satisfies the boundary conditions and has $|\nabla u| < 1$ for any (x,y) where u is differentiable. Nevertheless, this function is not in H^1_0.

Functions in H^1_0 have a *norm* (term from linear algebra)

$$\|u\|^2_{H^1_0} = \int_\Omega |\nabla u(x,y)|^2 \; dxdy \; . \tag{4}$$

The reason the u from (2) is in H^1_0 and the u from (3) is not is that first can be approximated by genuinely differentiable functions while the second cannot. For the first u, there is a family of smooth functions u^ϵ, defined as $\epsilon \to 0$ and a bound $M < \infty$ so that

$$\|u^\epsilon\|^2_{H^1_0} \leq M \; , \;\; \|u^\epsilon - u\|^2_{L^2} \leq \epsilon \; . \tag{5}$$

The L^2 norm is

$$\|v\|^2_{L^2} = \int_\Omega v(x,y)^2 dxdy \; .$$

You can create an approximating function u^ϵ by "smoothing out" the corners at $x = \frac{1}{2}$ and $y = \frac{1}{2}$. This is not possible for the discontinuous function of (3). If you try to smooth out the discontinuity at $x = \frac{1}{2}$, the gradient of the approximating u^ϵ will be so large that the first requirement of (5) cannot be satisfied. The Dirichlet integral of u^ϵ goes to infinity as $\epsilon \to 0$. This is the almost correct technical definition of H^1_0. The norm (4) should be finite, either in the literal sense sense that u is differentiable and the integral of the gradient is good, or in the indirect sense (5) that u can be approximated in the L^2 norm by functions smooth functions with good gradients. Informally, functions with discontinuous gradients like (2) are OK but discontinuous functions like (2) are not.

Finite element methods, as opposed to finite difference methods from Class 1, work by approximating the solution by a *trial function* from a finite dimensional *trial space*. This Assignment describes the C_0 method with trial functions that are *piecewise linear on triangles* often written PLT. The C in C_0 means "continuous", the 0 means no derivatives. Thus, C_0 just means that the trial function is continuous. Fancier methods might be C_1, which means that the trial function and its first partial derivatives are continuous, or C_2 (second derivatives), etc. Triangles are *triangular elements*. A triangular element is a triangle with vertices a, b, and c. There is an edge $e = (a,b)$ connecting each pair of vertices. The edge does not contain the vertices themselves, which we indicate by writing (a,b) instead of $[a,b]$. The "closed" interval $[a,b]$ includes its endpoints while the "open" interval (a,b) does not. This convention has the consequence that if $d \in e$ is any point on the edge, then d is not a vertex of the triangle. The *face* of the triangle is its "interior", which is all the triangle except the vertices and the edges. The *closure* of the triangle consists of the interior, the edges, and the vertices.

A *triangulation* of Ω, denoted by $\mathcal{T}$, is a set of triangles so that the interiors are disjoint and every point is in the closure of some triangle. A *valid* triangulation is one without *slave nodes*. A slave node is a vertex of one triangle $T \in \mathcal{T}$

that is on an edge of another triangle $T' \in \mathcal{T}$. A valid triangulation has the property that every triangle has exactly three *edge neighbors*[1] (neighbors that share an edge) and at least three more *vertex neighbors* (triangles that share a vertex but not an edge). You can draw triangulations in which an element has many vertex neighbors. Do a web search on "finite element triangulation", look for images, and you will get an idea of the geometric flexibility of triangulations.

A *piecewise linear* C_0 function is a continuous function that is affine ("linear" is the often used but slightly incorrect term) when restricted to any $T \in \mathcal{T}$. A function u is *affine* if there are constants so that $u(x,y) = \alpha x + \beta y + \gamma$. It is piecewise affine (piecewise linear) on $\mathcal{T}$ if every $T \in \mathcal{T}$ has α_T, β_T and γ_T so that if $(x,y) \in T$ then $u(x,y) = \alpha_T x + \beta_T y + \gamma_T$. If u is affine in T, then u is determined by its values at the three vertices of T. The set of affine functions on T is three dimensional (parameters α, β, and γ) and the set of vertex values is also three dimensional. The PLT trial space, denoted $\mathcal{S}_{\mathcal{T}}$, consists of all piecewise linear (affine) functions on $\mathcal{T}$ that satisfy the Dirichlet boundary conditions.

We use $\mathcal{V}$ to denote the *interior vertex set* or $\mathcal{T}$. Thus, $a \in \mathcal{V}$ if a is a vertex of some $T \in \mathcal{T}$ and $a \notin \Gamma$. If $u \in \mathcal{S}_{\mathcal{T}}$, then $u(a) = 0$ for any vertex $a \in \Gamma$. The values $u(a)$ for the interior vertices are arbitrary (why?). We use U to denote the vector of these vertex values: $U = (u(a),\, a \in \mathcal{V})$. More precisely, suppose the interior vertices are numbered in some way

$$\mathcal{V} = \{a_1, \cdots, a_n\} \ .$$

Then

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix} = \begin{pmatrix} u(a_1) \\ \vdots \\ u(a_n) \end{pmatrix} \ .$$

This shows that $\mathcal{S}_{\mathcal{T}}$ is a vector space of dimension n, where n is the number of interior vertices of $\mathcal{T}$.

Exercise 1. Show that $\mathcal{S}_{\mathcal{T}} \subset H_0^1$. Show that the *nodal basis functions* $b_j(x,y)$ are a basis of $\mathcal{S}_{\mathcal{T}}$. These are defined by $b_j \in \mathcal{S}_{\mathcal{T}}$ and $b_j(a_k) = 0$ if $j \neq k$ and $b_j(a_j) = 1$. Show that there is a symmetric positive definite matrix A so that if U represents the nodal values of $u \in \mathcal{S}_{\mathcal{T}}$, then

$$\int_\Omega \left|\nabla u(x,y)\right|^2 dxdy = U^T A U \ .$$

Show that the elements of A are

$$A_{jk} = \int_\Omega \nabla b_j(x,y) \cdot \nabla b_k(x,y) \, dxdy \ . \tag{6}$$

The *support* of a function is the set of (x,y) with $f(x,y) \neq 0$ (slightly inaccurate, but works for this exercise). Show that the support of b_j is the set of triangles

[1] This is slightly false. If an edge of T is part of Γ, then T has less than three edge neighbors and possibly less than three vertex neighbors.

with a_j as a vertex. Show that $A_{jk} = 0$ unless $j = k$ or (a_j, a_k) is an edge of some $T \in \mathcal{T}$.

Remark. A is the *stiffness* matrix. *Assembling* the stiffness matrix means computing and storing (most likely in sparse matrix format) all its elements. Among other things, requires you to order the interior nodes and compute the integrals (6).

Suppose u is any continuous function defined in Ω. The piecewise linear *interpolant* for triangulation $\mathcal{T}$, which we denote by $u_{\mathcal{T}}$, is the function the piecewise linear function $u_{\mathcal{T}} \in \mathcal{S}_{\mathcal{T}}$ that interpolates u in the sense that $u(a) = u_{\mathcal{T}}(a)$ for any $a \in \mathcal{V}$. The *diameter* of an element T is the maximum distance between points in T:

$$\operatorname{diam}(T) = \max \{ \, |(x,y) - (x',y')| \ , \ \ (x,y) \in T \, , \ (x',y') \in T \, \}$$

The *mesh size* is the maximum diameter of any element in the mesh. It is often called h:

$$h = \max_{T \in \mathcal{T}} \ \operatorname{diam}(T) \ .$$

Exercise 2. Show that if f is smooth in Ω, then there is a C so that

$$\max_{(x,y)\in\Omega} |f_{\mathcal{T}}(x,y) - f(x,y)| \leq Ch^2 \ .$$

Suppose T is a triangular element with vertices (a,b,c). The *standard triangular element*, called T_0, is the unit right triangle with vertices $\alpha = (0,0)$, $\beta = (1,0)$, and $\gamma = (0,1)$. The *condition number* of T is the condition number of the linear (affine, actually) map with $a \to \alpha$, $b \to \beta$, and $c \to \gamma$. The map is affine in the sense that it takes (x,y) to $M\begin{pmatrix} x \\ y \end{pmatrix} + r$ with $r \neq 0$ in general. The map is linear if $r = 0$. The condition number is the condition number of M: $\kappa(T) = \kappa(M)$. Note a little sloppiness here: The condition number $\kappa(T)$ defined here depends a little bit on which vertex of T goes to which vertex of T_0. You could fix that by making T_0 a unit isosceles triangle (all edges have length 1) or by noting that knowing the condition number to within a factor of 2 is good enough for this exercise. Making T_0 an isosceles triangle makes gradient estimation more complicated.

Exercise 3. Show that if u is smooth and there is a K with $\kappa(T) \leq K$, then

$$\max_{(x,y)\in T} |\nabla u(x,y) - \nabla u_{\mathcal{T}}(x,y)| \leq C_K \operatorname{diam}(T) \ .$$

Show that $C_K \to \infty$ as $K \to \infty$. *Hint* The bad elements with C large have two small (and possibly equal) angles ϵ and one angle $\pi - 2\epsilon$, and small diameter. Warning: this might not be easy if you don't see the tricks. If you can't get the general problem, you may assume that one of the angles of T is a right angle, and that this one is mapped to the right angle of T_0.

Remark. A triangulation $\mathcal{T}$ is *shape regular* with constant K if the condition numbers of all the elements is bounded by K. Being well conditioned is equivalent to there being a minimum angle θ_{min} so that all the angles in $\mathcal{T}$ are at least θ_{min}. The condition number definition of "shape regular" makes sense also in 3D and it is what is used in analysis. Finite element convergence theorems say that the approximate solutions u_k on a family of triangulations $\mathcal{T}_k$ converge to the exact solution if $h_k \to 0$ and all the triangles of all the triangulations have $\kappa(T) \leq K$. When making fine meshes (triangulations with small h) shape regularity is a big issue. If you just put points where you think you need them and connect them to make triangles, it is likely that the triangulation will not be shape regular. *Mesh smoothing* is the process of making shape regular triangulations from bad ones.

Fancy finite elements face the issue of *quadrature*. The elements of the stiffness matrix are integrals of polynomials over elements, which are evaluated numerically by quadrature or by possibly complicated formulas. The quadratures of Exercise 1 for the stiffness matrix are easy because ∇u is constant in an element for PLT methods.

Exercise 4. Suppose that $u \in \mathcal{S}_{\mathcal{T}}$ and replace f with its piecewise linear interpolant $f_{\mathcal{T}}$. Explain how to find the components of a vector $F \in \mathbb{R}^n$ so that

$$\int_\Omega u(x,y) f_{\mathcal{T}}(x,y)\, dxdy = U^T F \ .$$

Hint. The core of this is to find an algorithm for evaulating

$$\int_T (\gamma_1 x + \gamma_2 y + \gamma_3)(\delta_1 x + \delta_2 y + \delta_3)\, dxdy \ .$$

when $u(x,y) = (\gamma_1 x + \gamma_2 y + \gamma_3)$ and $f(x,y) = (\delta_1 x + \delta_2 y + \delta_3)$, T is the standard element, and and we given the values $u(a)$, $u(b)$, $u(c)$, $f(a)$, $f(b)$, and $f(c)$, where a, b, and c are the vertices of T. If T is a general triangular element, you also need the area of T, which is a "simple" formula involving the coordinates of its vertices.

Remark. One takeaway from this exercise is that the basic operations of finite element methods are conceptually simple but can be complicated to implement. That explains that fact that many finite element calculations use large finite element packages such as `FEniCS`. Writing all the code for a finite element calculation yourself can take a long time, is error prone, and leads to inefficient code.

Exercise 5. Suppose you solve $\triangle v_{\mathcal{T}} = f_{\mathcal{T}}$ with Dirichlet boundary conditions, and f is smooth. Show that

$$\|u - v_{\mathcal{T}}\|_{L^2} \leq Ch^2 \ .$$

Hint. The error $u - v$ satisfies $\triangle(u-v) = f - f_{\mathcal{T}}$, and $u - v$ also satisfies Dirichlet boundary contiions. Use this, Exercise 2, and the Poincaré inequality.

Remark. The simplicity of this illustrates the power of the finite element point of view: that you are working with functions rather than grid values. Inequalities for functions such as the Poincaré inequality tell you things about the numerical method. Exercise 6 is more elegance. Exercise 4 hints at the messy side.

Exercise 6. The finite element solution $u_{\mathcal{T}}$ is the solution of

$$\min_{u \in \mathcal{S}_{\mathcal{T}}} \|u\|_{H_0^1}^2 - \int_\Omega u(x,y) f_{\mathcal{T}}(x,y)\, dxdy \tag{7}$$

Show that $u_{\mathcal{T}}$ also is the minimizing w in

$$\mathrm{dist}_{H_0^1}^2(\, u,\, \mathcal{S}_{\mathcal{T}}\,) = \min_{w \in \mathcal{S}_{\mathcal{T}}} \|w - u\|_{H_0^1}^2 \ . \tag{8}$$

Hint. This is an orthogonality property. Draw a picture, imagining H_0^1 is just $\mathbb{R}^2$ and $\|u\|_{H_0^1}^2$ is $x^2 + y^2$.

Table of notation and terminolgy

Engineering papers often include a table of terminology and notation. Finite element methods are often used by engineers and there seems to be a lot of terminology and notation, so . . .

Ω	the domain on which f and u are defined, $[0,1]\times[0,1]$
Γ	the boundary of Ω
$\mathcal{T}$	a valid triangulation of Ω
T	a triangular element of the triangulation $\mathcal{T}$
a, b, c	the three vertices of an element $T \in \mathcal{T}$
$\mathcal{S}_{\mathcal{T}}$	piecewise linear functions on Ω that are continuous and affine on each element $T \in \mathcal{T}$
α, β, γ	parameters defining an affine function, as $\alpha x + \beta y + \gamma$, or $\gamma_1 x + \gamma_2 y + \gamma_3$
H_0^1	the Sobolev space of functions with $\displaystyle\int_\Omega \left|\nabla u(x,y)\right|^2 dxdy < \infty\, , u = 0$ on Γ
u	the exact solution to $\triangle u = f$ on $\Omega\, , u = 0$ on $\Gamma\,$, or a generic element of H_0^1
$u_{\mathcal{T}}$	the piecewise linear interpolant of the exact solution on the triangulation $\mathcal{T}$
U	a vector of the values of u (and also $u_{\mathcal{T}}$ on the vertices of $\mathcal{T}$
$f_{\mathcal{T}}$	piecewise linear interpolant of f on $\mathcal{T}\,$, can have $f_{\mathcal{T}} \notin \ \mathcal{S}_{\mathcal{T}}$ if $f \neq 0$ on Γ
$v_{\mathcal{T}}$	solution with piecewise linear $f\,$, solution of $\triangle v_{\mathcal{T}} = f_{\mathcal{T}}$ on $\Omega\, , v_{\mathcal{T}} = 0$ on Γ warning: $v_{\mathcal{T}} \notin \mathcal{S}_{\mathcal{T}}\,$.
$w_{\mathcal{T}}$	the closest piecewise linear function to the true solution, in the H_0^1 norm: $\displaystyle\min_{w \in \mathcal{S}_{\mathcal{T}}} \|w - u\|_{H_0^1}$, u being the solution of the original problem $\triangle u = f$