Numerical Analysis of PDE I, Spring 2024, HW#3.

Assignment day: February 11th 2024.

Ignacio Tomas[1], Department of Mathematics and Statistics, Texas Tech University.

Problem #1. *PDE problem: continuous dependence on coefficients.* Let u_i for $i = 1, 2$ be the solution of

$$-\mathrm{div}\,(a_i \nabla u_i) = f \text{ in } \Omega \text{ with } u_i = 0 \text{ on } \partial\Omega,$$

where $\Omega \subset \mathbb{R}^d$ is bounded, $f \in L^2(\Omega)$ and the coefficients satisfy $a_i \in \mathcal{C}(\overline{\Omega})$ and

$$0 < \alpha_0 \leq a_i(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \Omega.$$

Prove the stability bound

$$\|\nabla(u_1 - u_2)\|_{L^2(\Omega)} \leq \frac{const}{\alpha_0^2} \|a_1 - a_2\|_{L^\infty(\Omega)} \|f\|_{L^2(\Omega)}.$$

Problem #2. *Change of variables and scaling arguments.* Let h be a positive number and consider the coordinate transformation $\boldsymbol{x} = h\widehat{\boldsymbol{x}}$ from the reference domain $\widehat{\Omega}$ onto Ω. For the sake of concreteness: assume that $\mathrm{diam}(\widehat{\Omega}) = 1$ and $0 < h << 1$, therefore Ω is much smaller than $\widehat{\Omega}$. Assume that the function v (defined on Ω) and the function $\widehat{v}$ (defined on $\widehat{\Omega}$) are related by the rule $\widehat{v}(\widehat{\boldsymbol{x}}) := v(h\widehat{x})$, or equivalently $v(\boldsymbol{x}) := \widehat{v}(h^{-1}\boldsymbol{x})$.

a. Prove the scaling identities

$$\|v\|_{L^2(\Omega)} = h^{\frac{d}{2}} \|\widehat{v}\|_{L^2(\widehat{\Omega})}\,, \ \|\nabla v\|_{L^2(\Omega)} = h^{\frac{d}{2}-1} \|\widehat{\nabla}\widehat{v}\|_{L^2(\widehat{\Omega})}\,, \ \|v\|_{L^2(\partial\Omega)} = h^{\frac{d-1}{2}} \|\widehat{v}\|_{L^2(\partial\widehat{\Omega})}\,.$$

b. Prove the scaled trace-inequality

$$\|u\|_{L^2(\partial\Omega)} \leq c \left(h^{-1}\|u\|^2_{L^2(\Omega)} + h\|\nabla u\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}$$

Hint: Start from the trace inequality $\|\widehat{u}\|_{L^2(\partial\widehat{\Omega})} \leq c\|\widehat{v}\|_{H^1(\widehat{\Omega})}$ in the reference domain $\widehat{\Omega}$ and use scaling arguments.

Note: in this exercise we are using the simplified diagonal map $h\mathbb{I}$ with $\mathbb{I} \in \mathbb{R}^{d\times d}$ in order to relate the reference and mapped domains. In general, we will be interested in affine maps $\boldsymbol{x} = \boldsymbol{A}\widehat{\boldsymbol{x}} + \boldsymbol{b}$ where $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ encodes contractions, dilations and rotations, while $\boldsymbol{b} \in \mathbb{R}^d$ represents a translation.

Problem #3. *Proving L^2-interpolation estimate using elementary arguments.* Let $u \in H^2(0,1)$, that is, u admits two weak derivatives in $L^2(0,1)$. Let $I_h u$ be the continuous piecewise linear interpolant of u over a partition $\mathcal{T}_h$ of $(0,1)$ of size h, i.e. $h = \max_{K\in\mathcal{T}_h} h_K$. The following (global) error estimates holds true:

$$\|u - I_h u\|_{L^2(0,1)} \leq c\,h^2 \|u''\|^2_{L^2(0,1)} \quad \text{and} \quad \|(u - I_h u)'\|_{L^2(0,1)} \leq c\,h \|u''\|^2_{L^2(0,1)}\,.$$

The key difficulty in showing that these estimates are true lies in proving the (local) error estimate at each element $K = [x_j, x_{j+1}]$:

$$\|u - I_h u\|_{L^2(K)} \leq c h_K^2 \|u''\|_{L^2(K)} \quad \text{and} \quad \|(u - I_h u)'\|_{L^2(K)} \leq c h_K \|u''\|_{L^2(K)}$$

Provide two alternative proofs for these error estimates:

---

[1] `https://www.math.ttu.edu/~igtomas/`, igtomas@ttu.edu

a. Strategy #1 (Using scaling arguments): start from the Poincare inequality

$$\|\widehat{v}\|_{L^2(0,1)} \leq c\|\widehat{v}'\|_{L^2(0,1)},$$

which holds true for any $\widehat{v} \in H(0,1)$ provided $\widehat{v}$ vanishes for least at one point in $[0,1]$. You have already proved this inequality in Homework #1. Consider $\widehat{v} := \widehat{u} - \widehat{I_h u}$. Note that this choice of $\widehat{v}$ satisfies $\widehat{v}(0) = 0$ and $\widehat{v}(1) = 0$ because of the definition of the nodal interpolator. Similarly, note that, from the mean value theorem, there exists at least one $\widehat{x}_0 \in [0,1]$ such that $\widehat{v}'(\widehat{x}_0) = 0$. This allows us to use Poincare's inequality twice! Finish the proof using scaling arguments. Thoroughly justify each step.

b. Strategy #2 (Via $\mathcal{C}$-norm estimates and no scaling arguments): let $Q_1 v$ be the polynomial of degree 1 obtained from Taylor's formula for u at x_j (the left-point of the element K). Notice the invariance $I_h[Q_1 u] = Q_1 v$, and the bound $\|I_h w\|_{\mathcal{C}(K)} \leq \|w\|_{\mathcal{C}(K)}$ for all $w \in \mathcal{C}(K)$, so that

$$\|I_h v - v\|_{\mathcal{C}(K)} = \|I_h(v - Q_1 v) + (Q_1 v - v)\|_{\mathcal{C}(K)} \leq 2\|v - Q_1 v\|_{\mathcal{C}(K)}.$$

Then estimate the remainder

$$\|v - Q_1 v\|_{\mathcal{C}(K_j)} \leq \max_{x \in K} \int_{K_j} |x - y||v''(y)|dy.$$

This should lead to the conclusion that $\|I_h v - v\|_{C(K)} \leq 2h_K \int_K |v''(y)|\mathrm{d}y$. The estimate for $\|u - I_h u\|_{L^2(K)}$ follows by some very simple, but delicate, Cauchy-Schwarz trickery.

Regarding the semi-norm error, start by noting that $(I_h v)'(x) - v'(x) = \frac{1}{|K|}\int_K (v'(y) - v'(x))\mathrm{d}y$ for all $x \in K$, since $(I_h v)'(x)$ is a constant at each element.

Note: these proofs are "artisanal", since they only use very elementary tools. We will see later during the class that there exist more powerful and general tools to prove error estimates in arbitrary space dimensions.

Problem #4. *Symmetric vs non-symmetric bilinear form.* Let $a(u,v)$ and $F(v)$ satisfy the assumptions of the Lax-Milgram lemma, meaning

$$\begin{aligned} |a(u,v)| &\leq c_1\|u\|_{\mathbb{V}}\|v\|_{\mathbb{V}} && \text{for all } u, v \in \mathbb{V}, \\ a(u,u) &\geq c_2\|u\|_{\mathbb{V}}^2 && \text{for all } u \in \mathbb{V}, \\ |F(v)| &\leq c_3\|v\|_{\mathbb{V}} && \text{for all } v \in \mathbb{V}. \end{aligned}$$

Let $\mathbb{V}_h \subset \mathbb{V}$ be a finite dimensional subspace of $\mathbb{V}$ and let $u_h \in \mathbb{V}_h$ be the solution of

$$a(u_h, v_h) = F(v_h) \text{ for all } v_h \in \mathbb{V}_h.$$

Prove that

$$\|u_h - u\|_{\mathbb{V}} \leq \tfrac{c_1}{c_2} \min_{w_h \in \mathbb{V}_h} \|w - u\|_{\mathbb{V}}$$

for the general case when $a(u_h, v_h)$ may be non-symmetric. Prove that if $a(u_h, v_h)$ is symmetric we can define the norm $\|u\|_a := a(u,u)^{\frac{1}{2}}$ and we have that:

$$\|u_h - u\|_a = \min_{w_h \in \mathbb{V}_h} \|w_h - u\|_a \ \text{ and } \ \|u_h - u\|_{\mathbb{V}} \leq \left(\tfrac{c_1}{c_2}\right)^{\frac{1}{2}} \min_{w_h \in \mathbb{V}_h} \|w_h - u\|_{\mathbb{V}}.$$

Problem #5. *Two-point boundary value problem: boundary conditions and algebraic constraints.* Consider the following boundary value problem in $(0,1)$:

$$-u'' + u = f\,, u'(0) = 1\,,\ u(1) = 0$$

(a) Write the variational formulation: clearly identify the solution and test space, show that the weak formulation satisfies the conditions of the Lax-Milgram theorem. Questions: What boundary conditions become incorporated naturally, without any additional intellectual overhead, into the resulting formulation $a(u,v) = F(v)$? What boundary condition will have to be enforced into the solution/test space?

(b) Let $0 = x_0 < x_1 < x_2 < ... < x_{N-1} < x_N = 1$ be a partition of $(0,1)$ with $h_i = |K_i| = x_i - x_{i-1}$. Let $\{\phi_i\}_{i=0}^N$ be the corresponding basis of hat functions. Let $u_h = \sum_{i=0}^N U_i\phi_i$ be the continuous piecewise linear finite element solution. Write down the matrix equation satisfied by $U \in \mathbb{R}^{N+1}$ and $F \in \mathbb{R}^{N+1}$

$$(\boldsymbol{M} + \boldsymbol{K})U = F \tag{1}$$

for suitable vector F with entries F_i. Compute the entries of the $(N+1)\times(N+1)$ stiffness matrix $\boldsymbol{K}$, mass matrix $\boldsymbol{M}$, and right hand side vector F, by assembling elementary contributions of each element $K = (x_i, x_{i+1})$. Note however, that system (1) has too many rows/columns: we have to do something with it before we try to solve it using direct or iterative solvers, see next bullet.

(c) Now, assume general Dirichlet data on the right point, that is $u(1) = \alpha \neq 0$. Just like in bullet (b), one row of system (1) corresponds with a degree of freedom that we already know. In order to obtain an actually meaningful linear algebra system, we could consider removing some entries from the system. Similarly, regarding the vector F, some specific entries will require some manual modifications in order to incorporate the boundary data (known degrees of freedom). However, the most widely accepted strategy is not to remove any row/column of the system, but to modify the original $(N+1)\times(N+1)$ system accordingly without modifying its dimensionality in order to incorporate Dirichlet boundary conditions. For instance, it is a common practice to set the entire row of the known degree of freedom to zero with the exception of its diagonal value, which is set equal to one. Other additional modifications are necessary as well. Make precise what modifications are required and where. Show that the resulting modified system actually yields the expected behaviour. This is not mandatory, but if you feel adventurous enough: provide a brief argument explaining why the modified system is invertible (no need to get too rigorous).

(d) Consider instead the following set of boundary conditions: $u'(0) = 1$ and $u'(1) = 1$. Write a variational formulation, define a proper choice of solution and test space, clearly identify the bilinear form $a(u,v)$ and right hand side functional $F(v)$. Prove that the conditions of the Lax-Milgram theorem hold. Again, write a finite element discretization. For this case, given the "unmodified" linear system $(\boldsymbol{M} + \boldsymbol{K})U = F$. Question: Do we have to introduce any algebraic modification into specific rows/columns or right hand side vector? Or the resulting linear algebra system is good as is? Why?

The process of picking-up contributions from each cell K in order to compute the matrix of the system and right hand side vector is commonly known as "assembly". The process of introducing manual modifications into that system does not have a particular name. Perhaps, we could call it "introduction of algebraic constraints into the linear system" or "enforcement of boundary conditions". In fact, in large scale codes it could be found under the name of `set_affine_algebraic_constraints` , `set_algebraic_constraints`, etc. The question for you here is: What kind of boundary conditions appear to require the introduction of algebraic modifications into the system after its assembly?