Lecture 3
Strictly coercive problems

In the previous lectures we have established that in the Banach space setting, a bounded linear map $A : X \to Y^*$ is surely invertible provided it, as well as its adjoint, are bounded below. Moreover, given a bilinear form $a : X \times Y \to \mathbb{R}$, one can (always) characterize A through

$$a(x, y) = \langle Ax, y\rangle \quad \forall x \in X \ \forall y \in Y.$$

For the problem $Ax_0 = b$, we seek an approximate solution, otherwise referred to as the Galerkin approximation, we considered closed linear spaces $\widehat{X} \subset X$ and $\widehat{Y} \subset Y$ and have defined the Petrov-Galerkin problem

$$\text{(PG)} \quad \widehat{x} \in \widehat{X} \ : \ \langle A\widehat{x}, y\rangle = b(y) \quad \forall y \in \widehat{Y}.$$

If (PG) admits a solution $\widehat{x}$, then we have an optimality result governing the Galerkin approximation error

$$\|x_0 - \widehat{x}\| \leq \left(1 + \frac{\|A\|}{\widehat{\alpha}}\right) \inf_{x \in \widehat{X}} \|x_0 - x\|.$$

We will now direct our focus to the case of *strictly coercive* problems in which we assume that $X = Y$ is reflexive and that $a(x, x) \geq \alpha\|x\|^2$ for all $x \in X$ for some positive α. As a result, we have the following:

- A is invertible; the inf-sup conditions are satisfied.
- If we take $\widehat{X} = \widehat{Y}$, then we obtain the Galerkin (or Ritz-Galerkin) formulation and

  $$\widehat{\alpha} = \inf_{x \in \widehat{X}} \sup_{y \in \widehat{X}} \frac{a(x, y)}{\|x\|\,\|y\|} \geq \alpha \implies \text{(PG) is solvable,}$$

  with

  $$\|x_0 - \widehat{x}\| \leq C(\alpha, \|A\|) \inf_{x \in \widehat{X}} \|x_0 - x\|.$$

Remark. The constant $\widehat{\alpha}$ depends on the subspace $\widehat{X}$. However since $\widehat{\alpha} \geq \alpha$, C does not depend on $\widehat{X}$.

§1 Fourier-Galerkin method: an example

In this example we introduce the Fourier-Galerkin method. Let $X = Y = H^1(\mathbb{T})$ and recall the problem in Example 4 in the previous lecture:

$$-u'' + u = f \quad \text{in } \mathbb{T} \quad \text{with} \quad a(u, v) = \int_0^{2\pi} u'v' + uv. \tag{3.1}$$

We consider $\{\phi_k\}_{|k|\leq N}$ comprising of sines and cosines up to frequency N and define $X_N = \text{span}\{\phi_k\}$. Let $u_N \in X_N$ be the Galerkin approximation of u from subspace X_N. We have

$$\|u - u_N\|_{H^1} \leq C \inf_{v \in X_N} \|u - v\|_{H^1}.$$

Suppose now that $u \in H^s$ with $s \geq 1$ and that $u = \sum_{k\in\mathbb{N}} a_k\phi_k$. We define an H^s-semi-norm

$$|u|_{H^s}^2 = \sum_{k\in\mathbb{N}} |k|^{2s}|a_k|^2, \tag{3.2}$$

and define the H^s-norm

$$\|u\|_{H^s}^2 = \|u\|_{L^2}^2 + |u|_{H^s}^2. \tag{3.3}$$

Remark. It is clear that $|u|_{H^1} \sim \|u'\|_{L^2}$.

Consider now $T_N u := \sum_{|k|\leq N} a_k\phi_k \in X_N$; this being the *Fourier truncation* of u. We have,

$$\begin{aligned}
\|u - T_N u\|_{H^1}^2 &= \sum_{|k|>N} \left(|a_k|^2 + |k|^2|a_k|^2\right) \\
&\leq 2 \sum_{|k|>N} |k|^2|a_k|^2 \\
&= 2 \sum_{|k|>N} |k|^{2s}|k|^{2-2s}|a_k|^2, \qquad (|k|^{2-2s} \leq N^{2-2s},\ 2-2s<0), \\
&\leq cN^{2-2s}|u|_{H^s}^2.
\end{aligned}$$

We obtain the Fourier-Galerkin error estimate for (3.1),

$$\|u - u_N\|_{H^1} \leq cN^{-(s-1)}|u|_{H^s}, \tag{3.4}$$

from which we conclude that convergence rate increases with the regularity of u. In particular, if $u \in H^s$ for all s, then the convergence is <u>at least</u> spectral; the convergence is faster than any power of N.

Suppose now we assume that

$$\|u - u_N\|_{H^1} \leq cN^{-\sigma} \quad \forall N,$$

Let $T_N : H^1 \to X_N$ denote the Fourier truncation operator and suppose that $w \in X_N$. Clearly, $T_N w = w$. We have

$$\begin{aligned}
\|u - T_N u\|_{H^1} &= \|u - w + T_N w - T_N u\|_{H^1} \\
&\leq \|u - w\|_{H^1} + \|T_N(w-u)\|_{H^1} \\
&\leq \left(1 + \|T_N\|_{H^1\to H^1}\right)\|u - w\|_{H^1},
\end{aligned} \tag{3.5}$$

while noting that $\|T_N\|_{H^1\to H^1} = 1$, we have

$$\|u - T_N u\|_{H^1} \leq c \inf_{w\in X_N} \|u - w\|_{H^1}. \tag{3.6}$$

In other words, $T_N u$ serves as a best approximation for u in X_N. Moreover,

$$cN^{-\sigma} \geq \|u - u_N\|_{H^1} \geq \|u - T_N u\|_{H^1},$$

which makes

$$|u - T_N u|^2_{H^1} = \sum_{|k|>N} |k|^2 |a_k|^2 \leq cN^{-2\sigma}. \tag{3.7}$$

Carrying the necessary change in the index of summation we have

$$\sum_{N\geq 1} N^{2s} \sum_{|k|>N} |k|^2 |a_k|^2 < \infty,$$

provided that $2s - 2\sigma < -1$. Rearranging the previous summation,

$$\begin{aligned} \sum_{N=1}^{\infty} \sum_{|k|>N} N^{2s} |k|^2 |a_k| 62 &= \sum_{k=-\infty}^{\infty} \sum_{N=1}^{k} N^{2s} |k|^2 |a_k|^2 \\ &\geq \sum_{k\in\mathbb{Z}} |k|^{2s+1} |k|^2 |a_k|^2 = |u|^2_{H^{s+3/2}}, \end{aligned}$$

which is finite provided that $s + \frac{3}{2} < \sigma + 1$. We conclude the following:

$$\|u - u_N\|_{H^1} \leq cN^{-\sigma} \implies u \in H^s \ \forall s < \sigma + 1.$$

To sum up,

$$u \in H^s \implies \|u - u_N\|_{H^1} \leq cN^{-(s-1)} \implies u \in H^{s-\epsilon} \ \forall \epsilon > 0.$$

Remark. Note that argument (3.5) is not sharp but general. Furthermore, it can be shown that the optimality constant c in (3.6) is equal to 1, meaning that the Fourier truncation is indeed an optimal approximation in X_N.

§2 Linear finite element method in 1D: an example

We now consider the approximation space X_N generated by piecewise linear functions. Let $X = Y = H^1_0(\mathcal{I})$ where $\mathcal{I} = (0,1)$ and consider the 1-dimensional Poisson equation with Dirichlet boundary conditions

$$\begin{cases} -u'' = f & x \in (0,1) \\ u(0) = u(1) = 0 \end{cases} \tag{3.8}$$

Consider now a partition $\mathcal{P}$ of $(0,1)$ given by

$$\mathcal{P} = \{x_0 = 0 < x_1 < \cdots < x_n = 1\},$$

and define the functions space

$$X_{\mathcal{P}} = \{u \in C(\mathcal{I}) : u(0) = u(1) = 0, u|_{[x_k, x_{k+1}]} \in \mathbb{P}_1\}.$$

We need to compute $\overline{A}_{jk} = \int_{\mathcal{I}} \phi_j' \phi_k'$ and $\bar{b}_k = \int_{\mathcal{I}} f \phi_k$. Notice that the stiffness matrix $\overline{A}$ is tridiagonal; i.e. $\overline{A}_{jk} = 0$ if $|j-k| \geq 2$.

Let now $u_{\mathcal{P}} \in X_{\mathcal{P}}$ be the Galerkin approximation of u from $X_{\mathcal{P}}$. We have

$$\|u - u_{\mathcal{P}}\|_{H^1} \leq c \inf_{v \in X_{\mathcal{P}}} \|u - v\|_{H^1}.$$

Recall that $H^1(\mathcal{I}) \hookrightarrow C(\mathcal{I})$ so take $v = I_{\mathcal{P}} u$ where $I_{\mathcal{P}} u$ denotes the piecewise linear interpolation of u on $\mathcal{I}$ i.e. $v(x_i) = u(x_i)$ with $v \in X_{\mathcal{P}}$ and suppose that u is smooth. The error function $e = u - I_{\mathcal{P}} u$ satisfies $e(x_i) = 0$ for all i. Moreover, if $a = x_k$ and $b = x_{k+1}$,

$$|e(x)| = \left| \int_a^x e'(t)\, dt \right| \leq \int_a^b |e'(t)|\, dt = (b-a)^{1/2} \|e'\|_{L^2(a,b)},$$

which makes $\|e\|^2_{L^2(a,b)} \leq (b-a)^2 \|e'\|^2_{L^2(a,b)}$. By Rolle's theorem we obtain an analogous bound on $\|e'\|^2_{L^2}$ which together with the previous,

$$\|e\|^2_{L^2(a,b)} \leq (b-a)^2 \|e'\|^2_{L^2} \leq (b-a)^4 \|e''\|^2_{L^2},$$

which implies

$$\|e\|^2_{H^1(a,b)} \leq (b-a)^4 \|e''\|^2_{L^2} + (b-a)^2 \|e''\|^2_{L^2},$$

so we may write

$$\begin{aligned}
\|e\|^2_{H^1(\mathcal{I})} = \sum_{k=0}^{n-1} \|e\|^2_{H^1(x_k, x_{k+1})} &\leq 2 \sum_{k=0}^{n-1} (x_{k+1} - x_k)^2 \|e''\|^2_{L^2(x_k, x_{k+1})} \\
&= 2 \sum_{k=0}^{n-1} (x_{k+1} - x_k)^2 |u|^2_{H^2(x_k, x_{k+1})} \leq 2h^2 |u|_{H^1(\mathcal{I})},
\end{aligned}$$

where $h = \max_k (x_{k+1} - x_k)$. We conclude that

$$\|u - u_{\mathcal{P}}\|_{H^1} \leq C \|u - v\|_{H^1} \leq ch |u|_{H^2(\mathcal{I})}. \tag{3.9}$$

We have assumed that u is smooth, but the result holds for $u \in H^2(\mathcal{I})$ by density.

Remark. If the grid size is uniform, then $n = 1/h$ which makes $\|u - u_{\mathcal{P}}\|_{H^1} \leq cn^{-1} |u|_{H^2}$.