Lección 1.7: Métodos de Galerkin y mínimos cuadrados.

Ritz $\Rightarrow$ se trabaja en un espacio de dim $< \infty$

En el método de Galerkin aplicamos esta idea al caso en que el operador no es simétrico (Lax-Milgram).

Motivación: H de Hilbert real, $D_A \subset H$ $\bar{D_A} = H$, $A : D_A \subset H \to H$ operador lineal y acotado. Nota: no suponemos A simétrico.

Consideramos el problema

$$\text{hallar } u \in D_A \text{ tal que } Au = f \quad \cdots (1)$$

con $f \in H$ dado. Supongamos que $\exists\ u_0 \in D_A$ solución de (1). Entonces,

$$\langle Au_0 - f, v \rangle = 0 \qquad \forall\ v \in H.$$

Consideremos el espacio finito-dimensional

$$E_n = \text{span}\{\varphi_j\}_{j=1}^{n} \subset D_A, \qquad n \in \mathbb{N}.$$

Por ser $u_0 \in D_A$ solución

$$\langle Au_0 - f, \varphi_j \rangle = 0 \qquad \forall\ 1 \leq j \leq n \quad \cdots (2)$$

Idea: aproximar u_0 mediante

$$u_n = \sum_{j=1}^{n} a_j \varphi_j \qquad \cdots \qquad (3)$$

donde las a_j's se calculan resolviendo las ecuaciones (2).

Sistema de n ecuaciones con n incógnitas:

$$0 = \langle A u_n - f, \varphi_j \rangle = \left\langle \sum_{k=1}^{n} a_k A\varphi_k - f, \varphi_j \right\rangle$$
$$= \sum_{k=1}^{n} a_k \langle A\varphi_k, \varphi_j \rangle - \langle f, \varphi_j \rangle \qquad \forall\ 1 \leq j \leq n$$

es decir,

$$(4)\quad \begin{cases} \langle A\varphi_1, \varphi_1 \rangle a_1 + \cdots + \langle A\varphi_n, \varphi_1 \rangle a_n = \langle f, \varphi_1 \rangle \\ \quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ \langle A\varphi_1, \varphi_n \rangle a_1 + \cdots + \langle A\varphi_n, \varphi_n \rangle a_n = \langle f, \varphi_n \rangle \end{cases}$$

Nota: si A es simétrico este esquema coincide con el método de Ritz: $\langle A\varphi_k, \varphi_j \rangle = \langle \varphi_k, \varphi_j \rangle_A$ y la matriz es simétrica.

En el caso general, (4) se escribe como:

$$Ma = b \qquad \cdots \qquad (5)$$

con $M \in \mathbb{R}^{n\times n}$, $M_{kj} = \langle A\varphi_j, \varphi_k \rangle$, $1 \leq j, k \leq n$

$b \in \mathbb{R}^{n\times 1}$, $b_j = \langle f, \varphi_j \rangle$, $\forall\ 1 \leq j \leq n$

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Hay que analizar:

- solvabilidad de (S)
- si hay solución, ¿converge a u_0?
- → ¿qué pasa si $u_0 \notin D_A$?

Ventaja del método de Galerkin: no se asume la simetría del operador. Se va a plantear con mayor generalidad: forma bilineal no simétrica.

Caso importante: Lax-Milgram:

$$\begin{cases} D_A = H \\ A \text{ continuo y "}H\text{-elíptico"} \end{cases}$$

Definición Sea X un espacio de Banach. Un esquema de aproximación de Galerkin es una sucesión de subespacios de dimensión finita,

$$\{V_n\}_{n\in\mathbb{N}}, \quad V_n \subset X, \quad \dim V_n < \infty$$

tales que: $\forall u \in X \quad \exists \{v_n\}_{n\in\mathbb{N}} \subset X$ que satisface

(a) $v_n \in V_n, \quad \forall n \in \mathbb{N}$

(b) $\|u - v_n\| \to 0$ si $n \to \infty$.

Nota: (a) y (b) son equivalentes a

$$\forall u \in X \quad \lim_{n\to\infty} \operatorname{dist}(u, V_n) = 0$$

donde $\operatorname{dist}(u, W) = \inf_{w \in W} \|u - w\|$

$\forall\, W \subset X$ subespacio

Ejercicio: probar esta equivalencia.

<u>Lema</u> Sea X un espacio de Banach <u>separable</u>. Entonces existe un esquema de Galerkin, el cual se define mediante el siguiente método constructivo:

(i) Tómese cualquier conjunto numerable denso en X, $\{u_j\}_{j\in\mathbb{N}} \subset X$.

(ii) Defínase $V_n := \operatorname{span}\{u_1, \dots, u_n\}$, $n \in \mathbb{N}$

Entonces $\{V_n\}_{n\in\mathbb{N}}$ es un esquema de Galerkin.

<u>Demostración</u>: Sea $u \in X$. Por densidad de $\{u_j\}_{j\in\mathbb{N}}$, $\forall \epsilon > 0 \;\; \exists\, N = N(\epsilon) \in \mathbb{N}$ tal <u>que</u>

$$\| u_n - u \| < \epsilon \quad \text{si} \quad n \geq N$$

Pero por definición $u_n \in V_n$. Por lo tanto

$$\operatorname{dist}(u_n, V_n) < \epsilon.$$

Dado que $V_n \subset V_m$ si $m > n$ entonces : $\forall \epsilon > 0$ $\exists$ $N \in \mathbb{N}$ tal que $\text{dist}(u, V_n) < \epsilon$ $\forall n \geq N$. Es decir, $\text{dist}(u, V_n) \to 0$ si $n \to \infty$

$\square$

Observaciones :

(a) Los vectores $u_1, \dots, u_n$ no son necesariamente linealmente independientes.

(b) La sucesión de espacios es monótona creciente : $V_1 \subset V_2 \subset \dots \subset V_n \subset \dots$

(c) $X = \overline{\bigcup_{n \in \mathbb{N}} V_n}$ (por densidad).

Aproximación de Galerkin a la solución de Lax-Milgram.

$(H, \langle , \rangle)$ de Hilbert real

$a(\cdot, \cdot) : H \times H \to \mathbb{R}$ forma bilineal continua y H-elíptica : $\exists$ $\alpha, \beta > 0$ tales que

- $|a(u,v)| \leq \alpha \|u\| \|v\|$
- $a(u,u) \geq \beta \|u\|^2$ $\quad \forall u, v \in H$

Sea $\ell \in H^*$ (funcional lineal continuo en H).

(Si $D_A = H$, $A : H \to H$, continuo tal que $\langle Au, u \rangle \geq \beta \|u\|^2$ $\forall u \in H$, tenemos estas hipótesis con $a(u,v) := \langle Au, v \rangle$; A no es simétrico).

Lax-Milgram : $\exists!$ $u_0 \in H$ tal que

$$a(u_0, v) = \ell(v), \qquad \forall v \in H \quad \cdots (6)$$

¿Cómo aproximamos u_0?

Supongamos que H es <u>separable</u>. $\therefore \exists$ un esquema de Galerkin

$$\begin{cases} \exists \{V_n\}_{n\in\mathbb{N}}, \quad V_1 \subset V_2 \subset \cdots \subset V_n \subset \cdots \\ \text{subespacios de dim} < \infty \text{ tales que} \\ V_n \subset H, \quad \text{dist}(u, V_n) \to 0 \text{ si } n \to \infty \quad \forall u \in H \end{cases}$$

Aproximación de Galerkin: problema,

$$(7) \cdots \begin{cases} \text{Hallar } u_n \in V_n \text{ tal que} \\ a(u_n, v) = \ell(v), \quad \forall v \in V_n, \quad n \in \mathbb{N} \end{cases}$$

Nota: el problema (7) siempre tiene solución $\forall n \in \mathbb{N}$: $(V_n, \langle , \rangle)$ es un espacio de Hilbert de dim $< \infty$, $a_n(\cdot,\cdot) := a(\cdot,\cdot)|_{V_n \times V_n}$ es una forma bilineal continua y V_n-elíptica $\ell_n := \ell|_{V_n}$ es un funcional lineal continuo en V_n. Por Lax-Milgram : $\exists!$ $u_n \in V_n$ solución de (7).

¿Cómo calcular u_n? Fijemos $n \in \mathbb{N}$; supongamos $\dim V_n = m \le n$. Sea

$$\{\varphi_1, \ldots, \varphi_m\} \text{ base de } V_n$$

$\therefore \quad u_n = \sum_{j=1}^{m} a_j \varphi_j \quad$ para ciertos a_j's

Sustituyendo,

$$a\left(\sum_{j=1}^{m} a_j \varphi_j, v\right) = \ell(v), \quad \forall v \in V_n$$

$$\Rightarrow \sum_{j=1}^{m} a_j\, a(\varphi_j, v) = \ell(v), \qquad \text{ⅱ}$$

Tomando $v = \varphi_i$, $1 \le i \le n$:

$$\sum_{j=1}^{m} a(\varphi_j, \varphi_i)\, a_j = \ell(\varphi_i), \quad \forall i$$

obtenemos el sistema $M \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = b$

con $M \in \mathbb{R}^{m \times m}$, $M_{ij} = a(\varphi_j, \varphi_i) \quad \forall\ 1 \le i, j \le m$

M no es simétrica

$b \in \mathbb{R}^{m \times 1}$, $b_j = \ell(\varphi_i)$, $1 \le j \le m$

La solución u_n se calcula resolviendo este sistema. Galerkin consiste en incrementar n para aproximar u_0 por u_n.

Convergencia : tomando $v = u_n \in V_n$, tenemos

$$\beta \|u_n\|^2 \le a(u_n, u_n) = \ell(u_n) \le \|\ell\|_* \|u_n\|$$

$$\therefore \quad \|u_n\| \le \frac{\|\ell\|_*}{\beta} < \infty$$

La sucesión $\{u_n\}_{n \in \mathbb{N}}$ es acotada en H.

$\therefore\ \exists\ \{u_{n_j}\}_{j\in\mathbb{N}}$ que converge débilmente a $u\in H$:

$$\langle u_{n_j}, v\rangle \longrightarrow \langle u, v\rangle \quad \forall v\in H$$
si $j\to\infty$

Así, dado $v\in V_m$, $m\in\mathbb{N}$, para $j\gg 1$ suf. grande $V_m\subset V_{n_j}$ y por ende,

$$a(u_{n_j}, v) = l(v), \qquad j\gg 1$$

Entonces, $u_{n_j}\rightharpoonup u$ en H implica que

$$a(u_{n_j}, v)\to a(u,v) \quad \text{si } j\to\infty$$

En efecto: por la demostración de Lax-Milgram, $\exists\ A\in\mathcal{B}(H,H)$ tal que $a(u,v) = \langle Au, v\rangle$ $\forall v\in H$. Entonces,

$$a(u_{n_j}, v) = \langle Au_{n_j}, v\rangle = \langle u_{n_j}, A^*v\rangle$$
$$\downarrow \quad j\to\infty$$
$$\langle u, A^*v\rangle = \langle Au, v\rangle = a(u,v)$$

Así, $a(u,v) = l(v)$, $\forall v\in V_m$, $\forall m\in\mathbb{N}$.

$\therefore\ a(u,v) = l(v) \quad \forall v\in \bigcup_{m\in\mathbb{N}} V_m$

Dado que $H = \overline{\bigcup_{m\in\mathbb{N}} V_m}$ obtenemos

$$a(u,v) = l(v) \qquad \forall v\in H.$$

Por unicidad de la solución de Lax-Milgram:

$$u = u_0 \quad \in H.$$

$u_{n_j} \rightharpoonup u_0 \in H$ De hecho: $u_n \to u$ en H.

Teorema (lema de Cea)

Sean H de Hilbert separable y $\{V_n\}_{n\in\mathbb{N}}$ un esquema de Galerkin. Suponiendo que

$$(8)\cdots \begin{cases} a(u_n, v) = \ell(v) & \forall v \in V_n \\ u_n \in V_n \end{cases}$$

y que

$$(9)\cdots \begin{cases} a(u, v) = \ell(v) & \forall v \in H \\ u \in H \end{cases}$$

(por ejemplo, si a y ℓ satisfacen las hipótesis de Lax-Milgram) entonces

$$(10)\cdots \quad \|u_n - u\| \leq \frac{\alpha}{\beta} \underbrace{\operatorname{dist}(u, V_n)}_{\xrightarrow{n\to\infty} 0} \quad \forall n \in \mathbb{N}$$

Demostración Restando (8) y (9):

$$a(u - u_n, v) = 0 \qquad \forall v \in V_n \subset H$$

En particular, $a(u - u_n, u_n) = 0$.

Por lo tanto, $\forall v \in V_n$:

$$a(u-u_n, u-u_n) = a(u-u_n, u-v) + \underbrace{a(u-u_n, \underbrace{v-u_n}_{\in V_n})}_{=0}$$

$$= a(u-u_n, u-v)$$

Por continuidad y H-elipticidad de $a(\cdot,\cdot)$:

$$\beta \|u-u_n\|^2 \leq a(u-u_n, u-u_n)$$
$$= a(u-u_n, u-v) \leq \alpha \|u-u_n\| \|u-v\|$$

$$\therefore \quad \|u-u_n\| \leq \frac{\alpha}{\beta} \|u-v\|, \quad \forall v \in V_n$$

Así,

$$\|u-u_n\| \leq \frac{\alpha}{\beta} \inf_{v \in V_n} \|u-v\|$$
$$= \frac{\alpha}{\beta} \operatorname{dist}(u, V_n) \xrightarrow[n \to \infty]{} 0$$

por ser esquema de Galerkin

$\square$

Observaciones :

(1) La estimación (10) implica que para estimar el error en la aproximación de Galerkin, es suficiente con estimar $\inf_{v \in V_n} \|u-v\|$

(2) Cuando V_n se genera usando una base de funciones lineales, por pedazos :
Galerkin $\Rightarrow$ método de elemento finito.

Ejemplo: Consideremos el problema

$$(11) \quad \cdots \begin{cases} -u'' = f, & x \in (0,1) \\ u(0) = u(1) = 0 \end{cases}$$

dada $f \in C([0,1])$. El espacio de Hilbert $H = \cancel{H_0^1(0,1)}$ $L^2(0,1)$ (espacio de Sobolev). (Sección 3)

Esquema de Galerkin: idea natural, considerar polinomios que se anulan en $x=0$ y $x=1$

$$\tilde{V}_n := \text{span}\{ x^j(1-x) : 1 \leq j \leq n \}$$

Aprox. de Galerkin: $u_n(x) = \sum_{j=1}^{n} a_j x^j (1-x)$

Forma bilineal de (11):

$$a(u,v) = \int_0^1 u'(x)\, v'(x)\, dx$$

Problema: $a(u,v) = \langle f, v \rangle_{L^2} \quad \forall\, v \in L^2(0,1)$

$$\Leftrightarrow \int_0^1 u'(x) v'(x)\, dx = \int_0^1 f(x) v(x)\, dx$$

$$b_j := \langle f, \varphi_j \rangle = \int_0^1 f(x)\, x^j (1-x)\, dx$$

$b \in \mathbb{R}^{n \times 1}$

$M \in \mathbb{R}^{n \times n}$, $\quad M_{ij} = a(\varphi_i, \varphi_j)$

$$M_{ij} = \int_0^1 \frac{d}{dx}\left(x^i(1-x)\right)\frac{d}{dx}\left(x^j(1-x)\right) dx$$

$$= \frac{(i+1)(j+1)}{i+j+1} + \frac{(i+2)(j+2)}{i+j+3} + $$

$$- \frac{(i+1)(j+2) + (i+2)(j+1)}{i+j+2}$$

M es invertible y la aproximación de Galerkin es convergente, pero la matriz M está mal condicionada.

El no. de condición,

$$\mathcal{K}(M) = \|M\|_2 \|M^{-1}\|_2$$

aumenta con $n \in \mathbb{N}$. $\|M\|_2 := \sqrt{\lambda_{max}(M^T M)}$

Si $n = 3$, $\mathcal{K}(M) \approx 8.9 \times 10^2$

$n = 4$, $\mathcal{K}(M) \approx 2.4 \times 10^4$

$\vdots$

$n = 10$, $\mathcal{K}(M) \approx 1.1 \times 10^{13}$

calcular M^{-1} es muy costoso numéricamente y el error aumenta, conforme $n \to \infty$.
$\mathcal{K}$ depende de M no del algoritmo que se usa para calcular M^{-1}.

Consideremos mismo problema (11) con el siguiente esquema:

$$V_n = \text{span}\left\{ \sin(j\pi x) \; : \; 1 \leq j \leq n \right\}$$

En este caso :

$$M_{ij} = a(\varphi_i, \varphi_j) = \int_0^1 \frac{d}{dx}\left(\sin(j\pi x)\right)\frac{d}{dx}\left(\sin(i\pi x)\right)$$

$$= ij\,\pi^2 \int_0^1 \cos(j\pi x)\cos(i\pi x)\; dx$$

$$= ij\,\frac{\pi^2}{2}\,\delta_i^j \qquad \forall\; 1 \leq i,j \leq n$$

matriz diagonal

$\Rightarrow$ aprox. de Galerkin $\quad u_n = \sum_{j=1}^{n} a_j \sin(j\pi x)$

con

$$a_j = \frac{2}{\pi^2 j^2}\int_0^1 f(x)\sin(j\pi x)\; dx \qquad \forall\, j$$

$$b_j = \int_0^1 f(x)\sin(j\pi x)\; dx$$

$$\Rightarrow \quad u_n = \frac{2}{\pi^2}\int_0^1 f(y)\sum_{j=1}^{n}\frac{\sin(j\pi y)\sin(j\pi x)}{j^2}\; dy$$

Notas :

(4) $\exists$ generalizaciones de Galerkin :

- Petrov - Galerkin : teorema de Nečas
  (tarea 1) : $a(\cdot,\cdot) : H \times V \to \mathbb{R}$

- Galerkin no conforme o generalizado: no necesariamente $V_n \subset V$. Como $V_n \not\subset V$, $a(\cdot,\cdot)$ se aproxima por $a_n(\cdot,\cdot): V_n \times V_n$. Fórmula para el error.
- Galerkin discontinuo: funciones que generan V_n son continuas por pedazos.

(B) Elemento finito basado en Galerkin:
  - discretización del dominio Ω
  - elección de la base de V_n: funciones lineales por pedazos. (sección 4).

Método de mínimos cuadrados (Courant)

H de Hilbert separable

$A: D_A \subset H \longrightarrow H$ simétrico, positivo definido

$\overline{D_A} = H$

Hipótesis: existe una A-base de H:

conjunto numerable y completo $\{\varphi_j\}_{j\in\mathbb{N}} \subset D_A$ tal que

$$\{A\varphi_j\}_{j\in\mathbb{N}} \text{ es una base de } H$$

Bajo esta hipótesis: $\forall f \in H$ dado y $\forall \varepsilon > 0$ es posible hallar $m \in \mathbb{N}$ y constantes $c_1, \dots, c_m$ tales que

$$\left\| \sum_{j=1}^{m} c_j A\varphi_j - f \right\| < \epsilon$$

$$\iff \left\| \sum_{j=1}^{m} A(c_j\varphi_j) - f \right\| < \epsilon$$

Mínimos cuadrados: aproximar la solución generalizada $u \in H_A$, de $Au = f$, por la sucesión

$$u_n = \sum_{j=1}^{n} a_j \varphi_j$$

donde las constantes a_j's se determinan mediante

$$\min_{u_n \in \operatorname{span}\{\varphi_1, \dots, \varphi_n\}} \| Au_n - f \|^2$$

Para cualquier $v \in \operatorname{span}\{\varphi_1, \dots, \varphi_n\}$,

$v = \sum_{j=1}^{n} b_j \varphi_j$ y $\| Av - f \|^2 = G(b_1, \dots, b_n)$
forma cuadrática

Al resolver $\left.\dfrac{\partial G}{\partial b_j}\right|_{(a_1, \dots, a_n)} = 0$, $\forall j$

obtenemos

$$\begin{cases} \langle A\varphi_1, A\varphi_1 \rangle a_1 + \dots + \langle A\varphi_1, A\varphi_n \rangle a_n = \langle f, A\varphi_1 \rangle \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ \langle A\varphi_1, A\varphi_n \rangle a_1 + \dots + \langle A\varphi_n, A\varphi_n \rangle a_n = \langle f, A\varphi_n \rangle \end{cases}$$

$M_{ij} = \langle A\varphi_i, A\varphi_j \rangle$ $\quad$ M invertible pues $\{A\varphi_j\}_{j\in\mathbb{N}}$ es base

$\therefore \ \exists !$ solución $\quad a = M^{-1} b$

coeficientes de $\quad u_n = \sum_{j=1}^{n} a_j \varphi_j$

Convergencia: sabemos que $\exists\, u_0 \in H_A$ solución generalizada de $Au_0 = f$.

A definido positivo: $\quad \|u\|_A \geq c\,\|u\| \quad \forall u \in H_A$

sea $\epsilon > 0$ arbitrario. Verificamos que

$$(12) \cdots \quad \|u_n - u_0\|_A \leq \frac{1}{c} \|Au_n - f\|$$

Esto se sigue de:

$$\begin{aligned} \|u_n - u_0\|_A^2 &= \langle A(u_n - u_0), u_n - u_0 \rangle \\ &= \langle Au_n - f, u_n - u_0 \rangle \\ &\leq \|Au_n - f\| \, \|u_n - u_0\| \\ &\leq \frac{1}{c} \|Au_n - f\| \, \|u_n - u_0\|_A \end{aligned}$$

$\Rightarrow (12)$.

Basta con demostrar que $\|Au_n - f\| < \epsilon$ para $n \gg 1$.

Pero u_n se escogio tal que minimiza la $\|Av - f\|^2$ en V_n. Se puede verificar (ejercicio), para cada $\epsilon > 0$ podemos escoger $n \gg 1$ suf. grande tal que

$$\| A u_n - f \| < \epsilon$$

$$y \therefore \quad \| u_n - u_0 \|_A < \frac{\epsilon}{C}$$

Es decir, $\| u_n - u_0 \|_A \to 0$ si $n \to \infty$

$$\Rightarrow \| u_n - u_0 \| \to 0 \quad \text{si } n \to \infty$$

$\therefore$ convergencia.

<u>Teorema</u> Bajo las hipótesis, supongamos que $\{A\varphi_j\}_{j\in\mathbb{N}}$ es base de H. Entonces la aproximación de mínimos cuadrados, u_n, converge en H_A y en H, a la solución generalizada $u_0 \in H_A$ de $Au_0 = f$.

Nota: Método de Courant: combinar el método de Ritz con mínimos cuadrados:

$$J[u] := \underbrace{\tilde{F}[u]}_{\text{funcional cuadrática}} + \underbrace{\| Au - f \|^2}_{\text{mínimos cuadrados}}, \quad u \in D_A$$

u_n minimiza $J[\cdot]$ en V_n.

<u>Puede</u> resultar en una convergencia más rápida.