TMA4220

Numerical Solution of PDEs using FEM

Compendium and Lecture Notes

Author
Trym Sæther
Email: trym.saether@ntnu.no

Semester
Autumn 2025
Last updated: September 19, 2025

Norwegian University of Science and Technology
Department of Mathematical Sciences

Contents

Chapter 1

Matematiske Forutsetninger

1.1 Linær Algebra

1.1.1 Indre Produkt og Hilbert-rom

La V være et vektorrom over $\mathbb{R}$ (eller $\mathbb{C}$). Et *indre produkt* på V er en funksjon $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ som oppfyller:

1. **Symmetri:** $\langle u, v \rangle = \langle v, u \rangle$ for alle $u, v \in V$
2. **Linearitet:** $\langle \alpha u + \beta w, v \rangle = \alpha \langle u, v \rangle + \beta \langle w, v \rangle$ for alle $u, v, w \in V$ og $\alpha, \beta \in \mathbb{R}$
3. **Positivitet:** $\langle v, v \rangle \geq 0$ for alle $v \in V$, og $\langle v, v \rangle = 0$ hvis og bare hvis $v = 0$

Det indre produktet induserer en norm gjennom:

$$\|v\| = \sqrt{\langle v, v \rangle} \tag{1.1}$$

Definition 1.1.1. Hilbert-rom

Et *Hilbert-rom* H er et vektorrom med indre produkt som er komplett med hensyn til normen indusert av det indre produktet.
Komplett betyr at enhver Cauchy-følge i H konvergerer til et element i H.

1.1.2 Ortogonalitet og Projeksjoner

Et sentralt konsept i Hilbert-rom er ortogonalitet, som generaliserer ideen om vinkelrette vektorer.

Definition 1.1.2. Ortogonalitet

To elementer u, v i et Hilbert-rom H er *ortogonale* hvis $\langle u, v \rangle = 0$. Vi skriver $u \perp v$.
For et delsett $S \subseteq H$, definerer vi det ortogonale komplementet:

$$S^{\perp} = \{v \in H : \langle v, s \rangle = 0 \text{ for alle } s \in S\} \tag{1.2}$$

Theorem 1.1.3. Projeksjonsteoremet

La H være et Hilbert-rom og $V \subseteq H$ være et lukket delrom. For ethvert element $u \in H$ eksisterer det en unik *ortogonal projeksjon* $P_V u \in V$ slik at:

$$\|u - P_V u\| = \min_{v \in V} \|u - v\| \tag{1.3}$$

Projeksjonen karakteriseres av:

$$u - P_V u \perp V \quad \Leftrightarrow \quad \langle u - P_V u, v \rangle = 0 \text{ for alle } v \in V \tag{1.4}$$

1.1.3 Riesz-representasjonsteorem

Riesz-representasjonsteoremet er et fundamentalt resultat som karakteriserer strukturen til Hilbert-rom og etablerer en fullstendig isomorfi mellom et Hilbert-rom og dets dualrom.

Theorem 1.1.4. Riesz-representasjonsteorem

La H være et Hilbert-rom med indre produkt $\langle x, y \rangle$ som er *sesquilineært*. For enhver kontinuerlig lineær funksjonell $\varphi \in H^*$ eksisterer det et unikt element $f_\varphi \in H$, kalt *Riesz-representasjonen* av φ, slik at:

$$\varphi(x) = \langle x, f_\varphi \rangle \quad \text{for alle } x \in H \tag{1.5}$$

Videre gjelder:

1. $\|\varphi\|_{H^*} = \|f_\varphi\|_H$ (isometrisk egenskap)
2. $f_\varphi \in (\ker \varphi)^\perp$ er unik vektor, bestemt ved $\varphi(f_\varphi) = \|\varphi\|^2$
3. Det er også det unike elementet med minimal norm i $C := \varphi^{-1}(\|\varphi\|^2)$; det vil si at f_φ er det unike elementet i C som tilfredsstiller $\|f_\varphi\| = \inf_{c \in C} \|c\|$.

Definition 1.1.5. Kanonisk Riesz-kart

Riesz-kartet $\Phi : H \to H^*$ er den antilineære isometrien definert ved:

$$\Phi(y)(x) = \langle x, y \rangle \quad \text{for alle } x \in H \tag{1.6}$$

Riesz-representasjonsteoremet fastslår at Φ er bijektiv når H er komplett, og det inverse kartet er:

$$\Phi^{-1} : H^* \to H, \quad \varphi \mapsto f_\varphi \tag{1.7}$$

Remark 1. Geometrisk tolkning

For en ikke-null lineær funksjonell φ, danner mengden $C = \varphi^{-1}(\|\varphi\|^2)$ et affint hyperplan parallelt med $\ker \varphi$. Siden $C = f_\varphi + \ker \varphi$, er Riesz-representasjonen f_φ det unike punktet i C som ligger nærmest origo.
For enhver $q \in (\ker \varphi)^\perp \setminus \{0\}$ gjelder:

$$q = \frac{\|q\|^2}{\varphi(q)} \cdot f_\varphi \tag{1.8}$$

Example 1. Riesz-representasjon i L^2

I Hilbert-rommet $L^2([0,1])$ med det vanlige indre produktet $\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$, betrakt den lineære funksjonellen $\varphi(f) = f(1/2)$ (punktevaluering i $x = 1/2$).

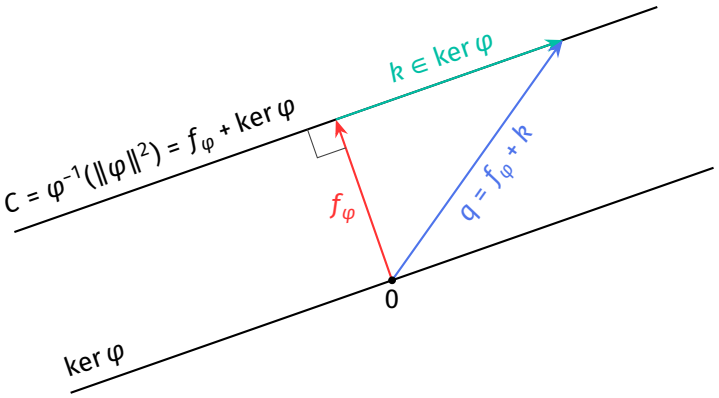


Figure 1.1: Geometrisk tolkning av Riesz-representasjonsteoremet. Det affine hyperplanet $C = \varphi^{-1}(\|\varphi\|^2)$ (blå linje) er parallelt med kjernekjernen $\ker \varphi$ (grå linje). Riesz-representasjonen f_φ (rød punkt) er det unike punktet i C med minimal avstand til origo.

Riesz-representasjonen er Dirac-deltafunksjonen $f_\varphi = \delta_{1/2}$, slik at:

$$\varphi(f) = f(1/2) = \int_0^1 f(x)\delta_{1/2}(x)\,dx = \langle f, \delta_{1/2}\rangle \tag{1.9}$$

1.1.4 Spesielle Matriser

Symmetriske matriser

Symmetriske matriser er sentrale i anvendelser grunnet deres egenskaper, spesielt innen lineær algebra og numeriske metoder som brukes i løsning av differensialligninger.

Definition 1.1.6. Symmetrisk matrise

En matrise $A \in \mathbb{R}^{n\times n}$ er *symmetrisk* hvis $A = A^T$, dvs. $a_{ij} = a_{ji}$ for alle i, j.

Symmetriske matriser har flere viktige egenskaper:

- Alle egenverdier er reelle: $\lambda_i \in \mathbb{R}$ for alle i.
- Egenvektorer tilhørende forskjellige egenverdier er ortogonale: $v_i \perp v_j$ hvis $\lambda_i \neq \lambda_j$.
- Matrisen er positivt definit hvis alle egenverdier er positive.

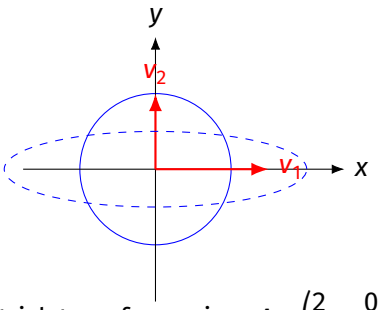
Theorem 1.1.7. Spektralteoremet for symmetriske matriser

Symmetriske matriser er ortogonalt diagonaliserbare: Det finnes en ortogonal matrise Q ($Q^TQ = I$) slik at

$$Q^TAQ = \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n),$$

hvor Λ er en diagonalmatrise med de reelle egenverdiene $\lambda_1, \dots, \lambda_n$ på diagonalen, og kolonnene i Q er de tilsvarende ortonormerte egenvektorene.

Dette teoremet er grunnleggende for mange numeriske algoritmer, da det tillater diagonaliserings-baserte løsninger.



Symmetrisk transformasjon: $A = \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix}$

Figure 1.2: Virkningen av en symmetrisk matrise på en sirkel. Transformasjonen strekker og komprimerer kun langs ortogonale akser (egenvektorene), uten rotasjon eller skjæring.

Example 2. Diagonaliserings av en symmetrisk matrise

La

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Denne matrisen er symmetrisk. Egenverdiene er $\lambda_1 = 3$ og $\lambda_2 = 1$, med egenvektorer $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ og $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Normalisering gir ortonormal basis, og diagonaliseringsmatrisen Q fører til $\Lambda = \text{diag}(3, 1)$.

1.1.5 Normer og spektrale egenskaper

Vektornormer

Normer gir oss en måte å måle *størrelsen* av vektorer og matriser.

Definition 1.1.8. Vektornorm

En vektornorm på $\mathbb{R}^n$ er en funksjon $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ som tilfredsstiller:

1. **Positivitet:** $\|x\| \geq 0$ for alle $x \in \mathbb{R}^n$, og $\|x\| = 0$ hvis og bare hvis $x = 0$.
2. **Homogenitet:** $\|\alpha x\| = |\alpha| \cdot \|x\|$ for alle $x \in \mathbb{R}^n$ og $\alpha \in \mathbb{R}$.
3. **Trekantulikheten:** $\|x + y\| \leq \|x\| + \|y\|$ for alle $x, y \in \mathbb{R}^n$ (trekantulikheten)

De mest vanlige vektornormene er p-normene:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \text{(taxicab)}$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \qquad \text{(euklidisk)}$$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \qquad \text{(maksnorm)}$$

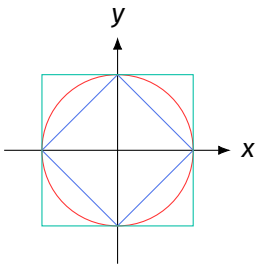


Figure 1.3: Blå: 1-norm (diamant). Rød: 2-norm (sirkel). Grønn: ∞-norm (kvadrat).

Matrisenormer

Definition 1.1.9. Matrisenorm

En matrisenorm på $\mathbb{R}^{m\times n}$ er en funksjon $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ som tilfredsstiller:

1. **Positivitet:** $\|A\| \geq 0$ for alle $A \in \mathbb{R}^{m\times n}$, og $\|A\| = 0$ hvis og bare hvis $A = 0$.
2. **Homogenitet:** $\|\alpha A\| = |\alpha| \cdot \|A\|$ for alle $A \in \mathbb{R}^{m\times n}$ og $\alpha \in \mathbb{R}$.
3. **Trekantulikheten:** $\|A + B\| \leq \|A\| + \|B\|$ for alle $A, B \in \mathbb{R}^{m\times n}$.

En vanlig matrisenorm er *Frobenius-normen*, som måler matrisen som om den var en "lang" vektor:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\text{tr}(A^TA)}$$

Definition 1.1.10. Konsistent matrisenorm

En matrisenorm $\|\cdot\|$ på $\mathbb{R}^{n\times n}$ kalles konsistent (eller sub-multiplikativ) hvis

$$\|AB\| \leq \|A\| \cdot \|B\|$$

for alle $A, B \in \mathbb{R}^{n\times n}$.

Denne egenskapen er viktig fordi den lar oss "ta ut" normen når vi har produkter av matriser, noe som er grunnleggende for mange konvergensbevis.

Definition 1.1.11. Underordnet matrisenorm

Gitt en vektornorm $\|\cdot\|_v$ på $\mathbb{R}^n$, defineres den underordnede matrisenormen $\|\cdot\|_M$ på $\mathbb{R}^{n\times n}$ som

$$\|A\|_M = \max_{\|x\|_v=1} \|Ax\|_v = \max_{x\neq 0} \frac{\|Ax\|_v}{\|x\|_v}$$

Intuitivt måler den underordnede normen den maksimale strekningen som matrisen A kan påføre enhver enhetsvektor.

De vanligste underordnede matrisenormene er:

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}| \quad \text{(maksimal kolonnesum)}$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^TA)} \quad \text{(største singulærverdi)}$$

$$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| \quad \text{(maksimal radsum)}$$

Visualisering av matrise-2-norm

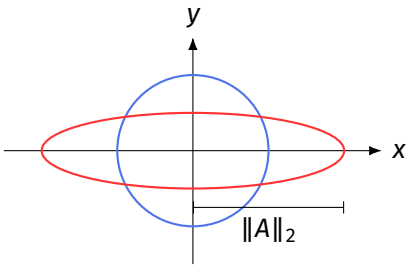


Figure 1.4: Den blå sirkelen er enhetskulen i 2-norm, og den røde ellipsen er resultatet av å anvende A på denne enhetskulen. Normen $\|A\|_2 = 2$ er den maksimale strekningen som A påfører.

Disse normene er praktiske fordi de er relativt enkle å beregne og har direkte tolkninger.

Theorem 1.1.12. Egenskaper ved underordnede normer

1. Enhver underordnet matrisenorm er konsistent.
2. For enhver matrise $A \in \mathbb{R}^{n\times n}$ og vektor $x \in \mathbb{R}^n$, gjelder

$$\|Ax\|_v \leq \|A\|_M \cdot \|x\|_v$$

Spektralradius

Definition 1.1.13. Spektralradius

Spektralradiusen til en matrise $A \in \mathbb{C}^{n\times n}$ er definert som

$$\rho(A) = \max\{|\lambda| : \lambda \text{ er en egenverdi av } A\}$$

altså den største absoluttverdien av egenverdiene til A.

Spektralradiusen er et fundamentalt mål på hvordan en matrise oppfører seg ved gjentatt anvendelse.

Theorem 1.1.14. Spektralradius og matrisenormer

La $A \in \mathbb{C}^{n\times n}$ og $\|\cdot\|$ være en konsistent matrisenorm. Da gjelder:

1. $\rho(A) \leq \|A\|$
2. For enhver $\varepsilon > 0$ finnes det en konsistent matrisenorm $\|\cdot\|_\varepsilon$ slik at $\|A\|_\varepsilon \leq \rho(A) + \varepsilon$

Dette betyr at spektralradiusen er den "best mulige nedre grense" for enhver konsistent matrisenorm.

Remark 2. Konvergens av matrisepotenser

La $A \in \mathbb{C}^{n\times n}$. Da gjelder følgende:

1. $\lim_{k\to\infty} A^k = 0$ hvis og bare hvis $\rho(A) < 1$
2. For enhver konsistent matrisenorm $\|\cdot\|$, hvis $\|A\| < 1$, så er $\lim_{k\to\infty} A^k = 0$

Gershgorins sirkelteorem

Gershgorins sirkelteorem er en praktisk metode for å finne egenverdiene til en matrise ved å konstruere sirkler i det komplekse planet.

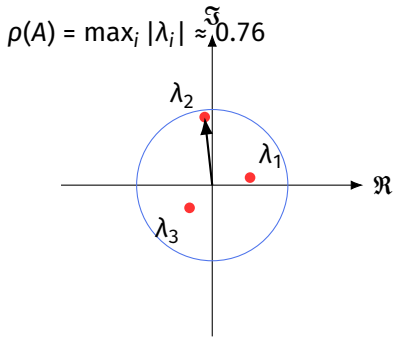


Figure 1.5: Spektralradius og konvergens. Når alle egenverdiene ligger innenfor enhetssirkelen ($\rho(A) < 1$), konvergerer sekvensen A^k mot nullmatrisen når $k \to \infty$. Dette er grunnleggende for stabilitetsanalyse av iterative metoder.

Theorem 1.1.15. Gershgorin's sirkelteorem

La $A = [a_{ij}] \in \mathbb{C}^{n\times n}$. For hver $i \in \{1, 2, \ldots, n\}$, definer Gershgorin-sirkelen

$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}$$

hvor $r_i = \sum_{j=1, j\neq i}^{n} |a_{ij}|$ er summen av absoluttverdiene til de ikke-diagonale elementene i rad i. Enhver egenverdi av A ligger i minst én av Gershgorin-sirklene, dvs. i unionen

$$\cup_{i=1}^{n} D_i$$

Videre, hvis k av sirklene danner et sammenhengende område som er separert fra de andre $n - k$ sirklene, da inneholder dette området nøyaktig k egenverdier av A (telt med multiplisitet).

Intuitivt forteller dette teoremet oss at diagonalelementene i en matrise gir en god indikasjon på hvor egenverdiene ligger, og at de ikke-diagonale elementene bestemmer hvor langt egenverdiene kan avvike fra diagonalelementene.

Example 3. Anvendelse av Gershgorin's teorem

For matrisen

$$A = \begin{pmatrix} 3 & 1 & 0 \\ 0.5 & 5 & 2 \\ 0 & 1 & 4 \end{pmatrix}$$

kan vi for hver rad i beregne radiusene r_i og sentrene a_{ii}:

$$\begin{aligned} r_1 &= |1| + |0| = 1, & a_{11} &= 3 \\ r_2 &= |0.5| + |2| = 2.5, & a_{22} &= 5 \\ r_3 &= |1| + |0| = 1, & a_{33} &= 4 \end{aligned}$$

slik at vi får de tre Gershgorin-sirklene:

$$\begin{aligned} D_1 &= \{z \in \mathbb{C} : |z - 3| \leq 1\} \\ D_2 &= \{z \in \mathbb{C} : |z - 5| \leq 2.5\} \\ D_3 &= \{z \in \mathbb{C} : |z - 4| \leq 1\} \end{aligned}$$

Dette betyr at enhver egenverdi av A må ligge i intervallet $[2, 7.5]$ på den reelle aksen.

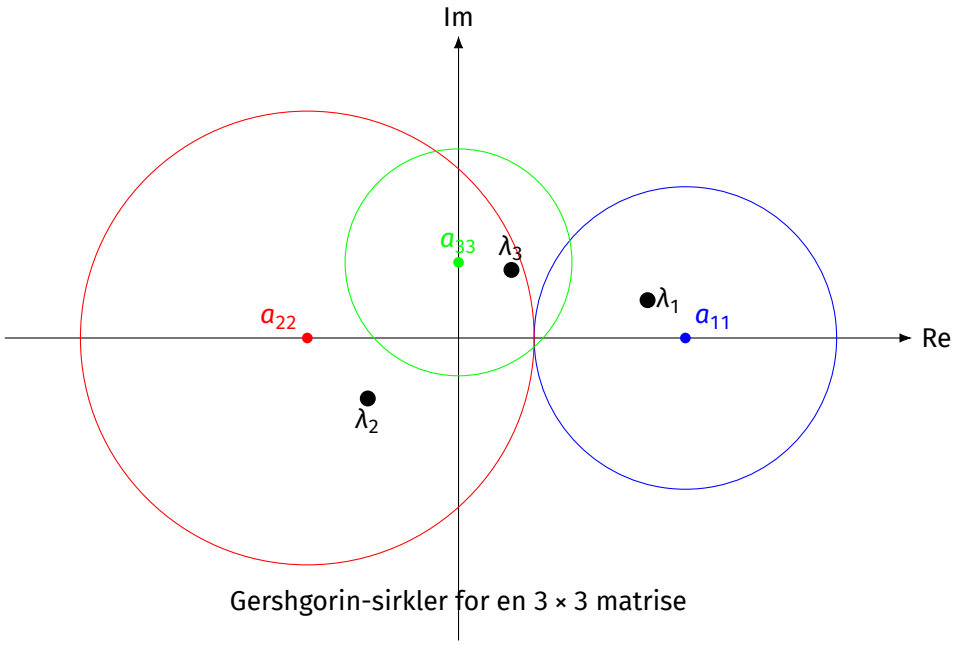


Figure 1.6: Gershgorin-sirkler for en 3 × 3 matrise. Hver sirkel er sentrert i et diagonalt element a_{ii} og har radius lik summen av absoluttverdiene til de ikke-diagonale elementene i rad i. Egenverdiene $\lambda_1, \lambda_2, \lambda_3$ må ligge innenfor minst én av sirklene.

1.2 Funksjonsrom

Funksjonsrom er samlinger av funksjoner med bestemte egenskaper. De danner grunnlaget for å analysere differensialligninger i en generalisert ramme.

Funksjonsrommet V Rommet av funksjoner som tilfredsstiller randbetingelsene.

$$V = \{v \in C^2(\Omega) \mid v(a) = \alpha, v(b) = \beta\}$$

Dette rommet representerer alle mulige funksjoner som kan være løsninger på problemet vårt.

Testfunksjoner $v(x)$ Funksjoner i V som brukes til å formulere den svake formen. Testfunksjoner lar oss "teste" en mulig løsning ved å integrere mot disse funksjonene.

Basisfunksjoner $\phi_i(x)$ Lokale funksjoner som spenner ut løsningsrommet.

- Har kompakt støtte (er null utenfor et lite område)
- Oppfyller $\phi_i(x_j) = \delta_{ij}$ (Kronecker delta)

Basisfunksjoner er byggesteinene vi bruker til å konstruere numeriske løsninger. I endelig element-metoden er disse typisk "hatt-funksjoner" eller polynomer som er positive i små områder og null ellers.

Lemma 1. Fundamental lemma of calculus of variations La $\Omega \subset \mathbb{R}^n$ være et åpent område med regulær rand og $f : \Omega \to \mathbb{R}$ være en kontinuerlig funksjon. Hvis

$$\int_\Omega \omega(x)\varphi(x)\,dx = 0 \tag{1.10}$$

holder for alle $\varphi \in \mathcal{C}_c^\infty(\Omega)$ (dvs. uendelig differensierbare funksjoner med kompakt støtte i Ω), da er $\omega(x) = 0$ for alle $x \in \Omega$.

Dette lemmaet er grunnleggende for variasjonelle formuleringer av differensialligninger. Det forteller oss at hvis integralet av et produkt er null for alle testfunksjoner, så må funksjonen selv være null nesten overalt.

1.2.1 Energifunksjoner

La $F : V \mapsto \mathbb{R}$ være en funksjon som tar inn en funksjon v og gir ut et reelt tall. Energien til en funksjon v er gitt ved:

$$F(v) = \frac{1}{2}\langle v, v\rangle - \langle f, v\rangle$$

I mange fysiske problemer representerer dette faktisk den fysiske energien i systemet. Minimering av denne energifunksjonen gir oss løsningen på det tilsvarende differensiallignings messige problemet, noe som er bakgrunnen for variasjonelle formuleringer.

1.3 Funksjonalanalyse

1.3.1 Svak derivasjon

Svak derivasjon er et konsept som lar oss definere derivasjon av funksjoner som ikke nødvendigvis er klassisk deriverbare.

Definition 1.3.1. Svak derivasjon

En funksjon $u \in L^2(\Omega)$ har en svak derivasjon $Du \in L^2(\Omega)$ hvis for alle testfunksjoner $\varphi \in C_c^\infty(\Omega)$ gjelder:

$$\int_\Omega u\frac{\partial \varphi}{\partial x}\,dx = -\int_\Omega Du\varphi\,dx.$$

1.3.2 Bilinær form

Definition 1.3.2. Bilinær form

En bilinær form $a : V \times V \to \mathbb{R}$ er en funksjon som er lineær i begge argumentene, dvs. for alle $u, v, w \in V$ og $\alpha, \beta \in \mathbb{R}$ gjelder:

$$a(\alpha u + \beta v, w) = \alpha a(u, w) + \beta a(v, w)$$
$$a(u, \alpha v + \beta w) = \alpha a(u, v) + \beta a(u, w).$$

For eksempel i Poisson-ligningen blir den bilinære formen:

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v\,dx \tag{1.11}$$

Bilinære former er byggeklossene i variasjonelle formuleringer og representerer typisk energien eller arbeidet i det fysiske systemet.

Lax-Milgram teoremet

Lax-Milgram-teoremet er et fundamentalt resultat som garanterer eksistens og entydighet av løsninger til bilinære variasjonsproblemer i Hilbert-rom.

Theorem 1.3.3. Lax-Milgram

La V være et Hilbert-rom, og la $a : V \times V \to \mathbb{R}$ være en bilinær form som er *kontinuerlig* og *koersiv*. Anta at $F : V \to \mathbb{R}$ er en begrenset lineær funksjonell. Da finnes det en entydig løsning $u \in V$ slik at

$$a(u, v) = F(v) \quad \forall v \in V.$$

For at teoremet skal gjelde, må bilinærformen $a(\cdot, \cdot)$ oppfylle to viktige egenskaper:

Kontinuitet: Det finnes en konstant $M > 0$ slik at

$$|a(u, v)| \leq M \|u\|_V \|v\|_V \quad \text{for alle } u, v \in V.$$

Dette betyr at små endringer i inngangsfunksjonene resulterer i små endringer i bilinærformen.

Koersivitet: Det finnes en konstant $\alpha > 0$ slik at

$$a(v, v) \geq \alpha \|v\|_V^2 \quad \text{for alle } v \in V.$$

Dette er en slags "positiv definitthet" for bilinærformen og sikrer at problemet er veldefinert.

Når F er en begrenset lineær funksjonell på V (altså $F \in V'$), garanterer teoremet at:

- Det finnes en entydig løsning $u \in V$ til variasjonsproblemet

$$a(u, v) = F(v) \quad \text{for alle } v \in V.$$

- Løsningen oppfyller den stabile estimeringen

$$\|u\|_V \leq \frac{1}{\alpha} \|F\|_{V'}.$$

  Dette betyr at løsningen avhenger kontinuerlig av dataene, en viktig egenskap for numeriske metoder.

Galerkins ortogonalitet

Definition 1.3.4. Galerkins Ortogonalitet

$$a(u - u_h, v) = 0 \quad \forall v \in V_h$$

hvor u_h er den approksimerte løsningen, u er den eksakte løsningen, og V_h er det endelige elementrommet.

Galerkins ortogonalitet betyr at feilen $u - u_h$ er "ortogonal" til hele approksiamasjonsrommet V_h med hensyn til bilinærformen $a(\cdot, \cdot)$. Dette er en fundamental egenskap ved Galerkins metode, og betyr at approksiamasjonen u_h er optimal i en viss forstand.

1.3.3 Céas lemma

Céas lemma sier at feilen mellom den eksakte løsningen u og den numeriske løsningen u_h kan estimeres ved hjelp av en konstant som avhenger av bilinærformen og koersivitetskonstanten.

Lemma 2. Céas lemma La $u \in V$ være den eksakte løsningen og $u_h \in V_h$ være den numeriske løsningen. Da gjelder:

$$\|u - u_h\| \leq \frac{M}{\alpha} \inf_{v \in V_h} \|u - v\|$$

hvor M er en konstant som avhenger av den kontinuerlige delen av bilinærformen, og α er koersivitetskonstanten for $a(\cdot, \cdot)$.

Dette lemmaet forteller oss at feilen i den numeriske løsningen u_h er begrenset av den best mulige approksimasjonen av den eksakte løsningen u i rommet V_h, multiplisert med forholdet $\frac{M}{\alpha}$.

Dette er et viktig resultat fordi det lar oss fokusere på approksiamasjonsegenskapene til rommet V_h når vi analyserer konvergensen av endelig element-metoder. For eksempel, hvis vi vet at V_h består av stykkevis lineære funksjoner, kan vi bruke klassisk approksimasjonsteori til å vise at $\inf_{v \in V_h} \|u - v\| = O(h^2)$ når u er glatt nok, og dermed at $\|u - u_h\| = O(h^2)$.

Chapter 2

Elementmetoden

Elementmetoden, eller *Finite Element Method* (FEM) er en numerisk metode for å løse partielle differensiallikninger (PDE) som beskriver fysiske fenomener som varmeoverføring, elastisitet, væskestrøm osv.

Kort forklart går metoden går ut på å dele opp domenet i små, enkle geometriske *elementer* (som en linje, trekant, firkant osv.). Innenfor hvert element tilnærmes løsningen med enkle funksjoner. Til slutt settes elementene sammen for å få en tilnærmet løsning over hele området.

2.1 Fremgangsmåte

1. Finn $u : \Omega \to \mathbb{R}$ slik at $\mathcal{L}u = f$ i Ω og $u\big|_{\partial\Omega} = g$ på randen $\partial\Omega$.
2. Finn $u \in V := \mathcal{H}_0^1(\Omega)$ slik at $\langle \mathcal{L}u, \mathcal{L}\phi \rangle_\omega = \langle f, \phi \rangle_\omega \ \forall \phi \in V$.
3. Finn $u_h \in V_h \subset V := \mathcal{H}_0^1(\Omega)$ slik at $\langle \mathcal{L}u_h, \mathcal{L}\phi_i \rangle_\omega = \langle f, \phi_i \rangle_\omega \ \forall 1 \le i \le n$.
4. Løs $A\mathbf{u} = \mathbf{f}$ with $A_{ij} := \left\langle \mathcal{L}\phi_i, \mathcal{L}\phi_j \right\rangle$.

1. **discretization 1:** Del domain Ω inn i enkle geometriske elementer Ω_e og tilnærme løsningen med en lineærkombinasjon av basisfunctioner:
$$u(x) \approx \sum_{i=1}^{N} c_i \phi_i(x)$$
2. **weakformulation:** Multipliser differensiallikningen med en testfunksjon $v(x)$ og integrer over domainet Ω:
$$\int_\Omega v(x)\mathcal{L}(u)\,dx = \int_\Omega v(x)f(x)\,dx$$
3. **Galerkin-prosedyre:** Tilnærme løsningen og testfunksjonen med basisfunctioner:
$$u(x) \approx \sum_{i=1}^{N} c_i \phi_i(x) \quad \text{og} \quad v(x) \approx \sum_{j=1}^{N} d_j \phi_j(x)$$
4. **discretization 2:** Sett inn tilnærmingene i den svake formen og diskretiser:
$$\sum_{j=1}^{N} d_j \int_\Omega \phi_j(x)\mathcal{L}\left(\sum_{i=1}^{N} c_i \phi_i(x)\right) dx = \sum_{j=1}^{N} d_j \int_\Omega \phi_j(x) f(x)\,dx$$

5. **Matriseform:** Skriv den diskretiserte formen som et ligningssett:

$$\mathbf{Kc} = \mathbf{F}$$

Og løs ligningssystemet for å finne koeffisientene c_i:

$$\mathbf{c} = \mathbf{K}^{-1}\mathbf{F}$$

1. **Definer problemet:** Finn pde, over hvilket domene og rand.

$$u : \Omega \to \mathbb{R}, \quad \mathcal{L}(u) = f(x) \quad \text{for } x \in \Omega \quad \text{og} \quad u|_{\partial\Omega} = g$$

2. **Diskretiser domenet:** Del opp Ω i enkle geometriske elementer.
3. **Lokal approksimering for hvert element:** Innenfor hvert element (ukjente området/løsning) tilnærmer vi løsningen med en lineærkombinasjon av basisfunksjoner.

$$u(x) \approx \sum_{i=1}^{N} c_i \phi_i(x)$$

4. **Fra Sterk til Svak formulering:** Formuler pde på weakformulation.

$$\int_\Omega v(x)\mathcal{L}(u(x))\,dx = \int_\Omega v(x) f(x)\,dx$$

$$u \in \mathcal{C}^2 \quad \text{s.t.} \quad \mathcal{L}(u) = f \text{ for } x \in \Omega,\ u|_{\partial\Omega} = g \implies u \in V \quad \text{s.t.} \quad \langle \mathcal{L}(u), \mathcal{L}(\phi)\rangle = \langle f, \phi\rangle\ \forall \phi \in V$$

5. **Formuler elementeligningene:** Sett sammen elementene til et system av ligninger, ved å bruke weakformulation av pde.

$$\mathbf{Ku} = \mathbf{F}$$

6. **Randbetingelser:** Sett opp ligningssystemet med randbetingelser.
7. **Løs ligningssystemet:** Løs ligningssystemet for å finne koeffisientene u_i.

$$\mathbf{u} = \mathbf{K}^{-1}\mathbf{F}$$

2.2 Elementrom og basisfunksjoner

2.2.1 Definisjon av elementrommet $\mathcal{X}_h^p \subset H^1(\Omega)$

Finite Element Space $\mathcal{X}_h^p$ er et rom av stykkvis polynomer av grad p begrenset på elementet $\mathcal{K}_i \in \mathcal{T}_h \subset \Omega$. Hvor $\mathcal{T}_h$ er en partisjon av domenet Ω i elementer K_i.

> **Example 4. Elementer**
>
> Eksempler på elementer er trekanter, firkanter, tetraedre osv.

Definition 2.2.1. Finite Element Space

$$\mathcal{X}_h^p = \{v \in \mathcal{C}^0(\Omega) \mid v\big|_K \in \mathrm{P}_p \text{ for alle } K_i \in \mathcal{T}_h\} \tag{2.1}$$

- $u \in \mathcal{C}^0(\Omega)$ betyr at u er kontinuerlig $\mathcal{C}^0$ i domenet Ω.
- $v\big|_K \in \mathrm{P}_p$ betyr at v er et polynom av grad p i elementet K.
- $\mathcal{T}_h$ er en partisjon av domenet Ω i elementer K.

2.2.2 Valg av basisfunksjoner

En sentral del av FEM-metoden er valget av basisfunksjoner for å representere løsningen. Disse funksjonene danner fundamentet for approksimeringen av løsningen i det diskrete rommet.

Egenskaper til basisfunksjoner

Basisfunksjoner er funksjoner som brukes til å representere den approksimerte løsningen u_h i det diskrete rommet $\mathcal{X}_h^p$. Gode basisfunksjoner for FEM har følgende egenskaper:

- **Lokal støtte**: $\phi_i(x) = 0$ utenfor elementene der node i inngår
- **Interpolasjon**: $\phi_i(x_j) = \delta_{ij}$ (Kronecker delta)
- **Partisjon av enheten**: $\sum_{i=1}^{n} \phi_i(x) = 1$ for alle $x \in \Omega$
- **Kontinuitet**: Basisfunksjonene er kontinuerlige på tvers av elementgrensene
- **Representasjon**: Gir en god approksimering $u_h(x) = \sum_{i=1}^{n} u_i \phi_i(x)$

Med disse egenskapene kan vi representere enhver funksjon $v \in X_h^p$ som:

$$v(x) = \sum_{i=0}^{M} v_i \phi_i(x) \in \mathbb{P}_p$$

Lagrange-basisfunksjoner

Lagrange-basisfunksjoner er den mest brukte typen polynomielle basisfunksjoner i FEM. De er spesielt praktiske fordi koeffisientene u_i direkte representerer funksjonsverdiene ved nodene.

Definition 2.2.2. Lagrange-basisfunksjoner

Lagrange-basisfunksjoner av grad p er definert ved interpolasjon over nodepunkter $\{x_i\}_{i=0}^{p}$. For hvert nodepunkt i, er den tilsvarende Lagrange-basisfunksjonen $\ell_i(x)$ gitt ved:

$$\ell_i(x) = \prod_{j=0, j\neq i}^{p} \frac{x - x_j}{x_i - x_j} \tag{2.2}$$

med egenskapene:
1. $\ell_i(x_j) = \delta_{ij}$ (Kronecker-delta)
2. $\ell_i(x)$ er et polynom av grad p
3. $\sum_{i=0}^{p} \ell_i(x) = 1$ for alle x (partisjon av enheten)

Lagrange-basisfunksjoner har flere fordeler i FEM:

- De gir en direkte tolkning av koeffisientene som funksjonsverdier ved nodene
- De forenkler håndteringen av randbetingelser
- De har kompakt støtte, noe som gir sparse matriser i implementeringen

Lagrange-basisfunksjoner i 1D I én dimensjon er Lagrange-basisfunksjonen for node i definert som:

$$\phi_i(x) = \prod_{j=0, j\neq i}^{M} \frac{x - x_j}{x_i - x_j}$$

Example 5. Lineære basisfunksjoner i 1D ($p = 1$)

For rommet $\mathcal{X}_h^1$ av stykkvis lineære funksjoner er basisfunksjonene gitt ved:

$$\phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & x_{i-1} < x \leq x_i, \\ \frac{x_{i+1} - x}{x_{i+1} - x_i}, & x_i < x < x_{i+1}, \\ 0, & \text{ellers.} \end{cases}$$

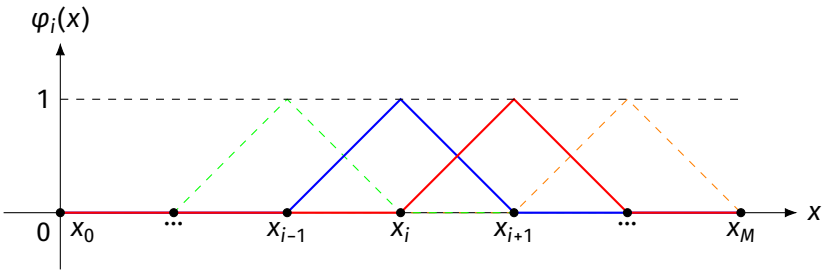


Figure 2.1: Basisfunksjoner $\phi_i(x)$ for $i = 0, 1, \ldots, M$ i rommet X_h^1

På et enkelt element $K_i = [x_i, x_{i+1}]$ er bare to basisfunksjoner ulik null:

$$\phi_i(x) = \frac{x_{i+1} - x}{x_{i+1} - x_i}, \quad \phi_{i+1}(x) = \frac{x - x_i}{x_{i+1} - x_i}, \quad \text{for } x \in [x_i, x_{i+1}].$$

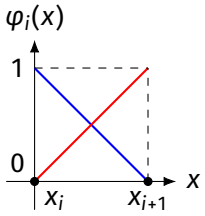


Figure 2.2: Basisfunksjoner $\phi_i(x)$ (blå) og $\phi_{i+1}(x)$ (rød) på elementet $K_i = [x_i, x_{i+1}]$

Nøkkelobservasjoner:

- **Lokal støtte:** Hver $\phi_i(x)$ er ulik null kun på de to elementene K_{i-1} og K_i
- **Partisjon av enheten:** $\sum_{i=0}^{M} \phi_i(x) = 1$ for alle $x \in [x_0, x_M]$
- **Interpolerende:** For en funksjon $v(x)$, gir $v_h(x) = \sum_{i=0}^{M} v(x_i)\phi_i(x)$ den stykkevis lineære interpolasjonen av v på nodepunktene

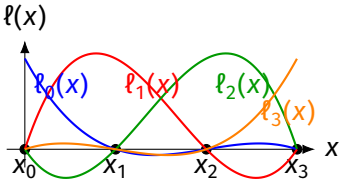


Figure 2.3: Kubiske Lagrange-basisfunksjoner ($p = 3$) på intervallet $[0, 3]$

Dette kan generaliseres til polynomer av høyere grad ved å bruke flere nodepunkter innenfor hvert

element, noe som gir bedre nøyaktighet i løsningen.

Lagrange-basisfunksjoner i 2D I to dimensjoner kan Lagrange-basisfunksjoner konstrueres ved bruk av tensorprodukt:

Definition 2.2.3. Lagrange-basisfunksjoner i 2D

$$\phi_i(x, y) = \prod_{j=0, j\neq i}^{M} \frac{x - x_j}{x_i - x_j} \cdot \prod_{k=0, k\neq i}^{M} \frac{y - y_k}{y_i - y_k}$$

hvor (x_i, y_i) er nodepunktet i elementet, med egenskapene:
- $\phi_i(x_j, y_j) = \delta_{ij}$ (Kronecker delta)
- $\phi_i(x, y)$ er lik null utenfor elementene som inneholder node i
- $\phi_i(x, y)$ er et polynom av grad p i x og y innenfor hvert element

For triangulære elementer med lineære basisfunksjoner ($p = 1$) brukes ofte barysentriske koordinater, som gir spesielt elegante uttrykk for basisfunksjonene.

Remark 3. FEM funksjonsrom

Vi må også kreve noe av rommet vi jobber med.

$$V := \{v : v \in \mathcal{C}[0, 1], v' \text{ er stykkvis kont. og bundet på } [0, 1] \text{ hvor } v(0) = v(1) = 0\}$$

2.3 fem-betingelser

Betingelser For å bruke fem til å løse et pde-problem, må problemet oppfylle følgende betingelser:

- **Linearitet:** Problemet må være lineært, dvs. ligningene kan uttrykkes som $\mathcal{L}(u) = f$, hvor $\mathcal{L}$ er en lineær operator.
- **Kontinuerlig Differensierbar:** Løsningen $u(x)$ må være kontinuerlig differensierbar i Ω.
- **Geometrisk Enkelhet:** domainet Ω bør kunne deles opp i enkle geometriske elementer (f.eks. trekanter, firkanter i 2D, tetraedre i 3D):

$$\Omega = \bigcup_{e=1}^{E} \Omega_e$$

- **Kvantiserbarhet:** Problemet må være kvantiserbart, dvs. løsningen kan tilnærmes godt ved hjelp av en endelig basisfunction:

$$u_h(x) = \sum_{i=1}^{N} c_i \phi_i(x)$$

- **Randbetingelser:** Randbetingelsene må være kompatible med valg av funksjonsrom:

$$u|_{\partial\Omega} = g \quad \text{eller} \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = h$$

For å anvende fem på en pde, må følgende betingelser være oppfylt:

2.3.1 Linearitet

Problemet må kunne uttrykkes i formen $\mathcal{L}(u) = f$, der $\mathcal{L}$ er en lineær differensialoperator.

2.3.2 Regularitet

Løsningen $u(x)$ må ha tilstrekkelig regualritet (kontinuitet og differensierbarhet) innenfor Ω.

$$u \in C^2(\Omega) \quad \text{og} \quad u \in C^1(\partial\Omega)$$

2.3.3 Diskretisering av domenet

Domenet Ω må kunne dekomponeres i ikke-overlappende elementer:

$$\Omega = \bigcup_{e=1}^{E} \Omega_e, \quad \Omega_i \cap \Omega_j = \emptyset \text{ for } i \neq j$$

2.3.4 Approksimering av funksjonsrommet

Løsningen må kunne representeres tilfredsstillende ved hjelp av basisfunksjoner:

$$u(x) \approx u_h(x) = \sum_{i=1}^{N} c_i \phi_i(x)$$

2.3.5 Veldefinerte randbetingelser

Problemet må ha klart definerte randbetingelser:

- Dirichlet-betingelser: $u = g_D$ på $\partial\Omega_D$
- Neumann-betingelser: $\nabla u \cdot \mathbf{n} = g_N$ på $\partial\Omega_N$
- Robin-betingelser: $\alpha u + \beta \nabla u \cdot \mathbf{n} = g_R$ på $\partial\Omega_R$

2.4 Sterk formulering

Gitt $f(x)$, finn $u(x)$ slik at

$$\begin{aligned} u''(x) &= f(x) \quad \text{for alle } 0 \leq x \leq 1, \\ u(0) &= 0, \quad u'(1) = 0. \end{aligned}$$

Introduserer testfunksjonene $v(x)$.

$$u''(x)v(x) = f(x)v(x)$$

Så integrerer vi over intervallet $[0, 1]$.

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

Alt vi har gjort til nå er å multiplisere med $v(x)$ og integrere over intervallet $[0, 1]$, som er generelt lov så lenge $v(x)$ er kontinuerlig og $u(x)$ er to ganger kontinuerlig deriverbar. Noe den er fra antagelsen om at $u(x)$ er to ganger kontinuerlig deriverbar.

Hvis vi sammenligner integralene på hver side, sier vi egentlig at arealet for $u''(x)v(x)$ er lik arealet for $f(x)v(x)$.

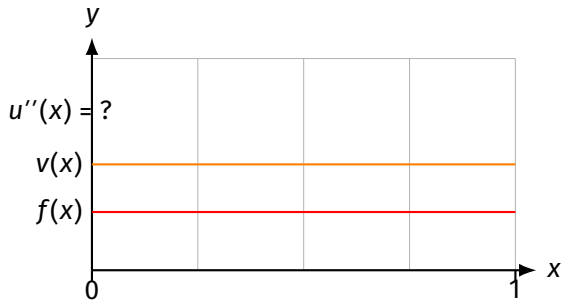


Figure 2.4: Vilkårlig testfunksjon $v(x)$, og kjent $f(x)$

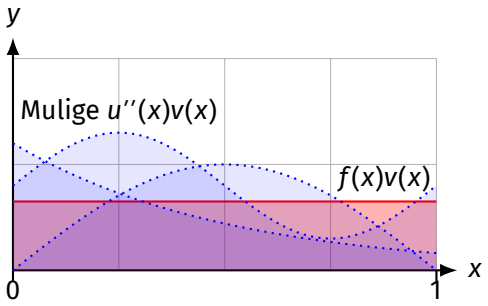


Figure 2.5: Flere mulige løsninger for $u''(x)v(x)$, som har samme areal som $f(x)v(x)$

Figure 2.6: Flere mulige løsninger for $u''(x)v(x)$, som har samme areal som $f(x)v(x)$

I figur 2.6 er det flere mulige løsninger for $u''(x)v(x)$, som har samme areal som $f(x)v(x)$.

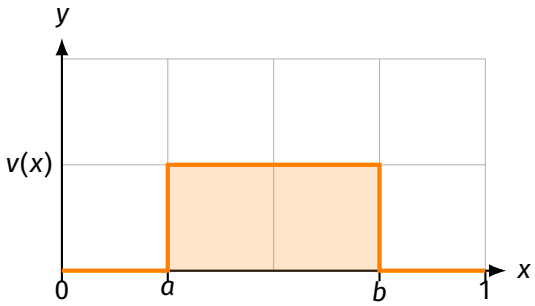
I dette tilfellet er testfunksjonen vår $v(x) = 1$ forteller ikke dette oss noe mer om hvordan $u''(x)$ ser ut. Altså det gir oss ikke noe mer informasjon om $u''(x)$.

Målet vårt nå er å finne/lage en testfunksjon $v(x)$ som faktisk vil kunne gi oss mer informasjon om $u''(x)$.

La oss se på testfunksjonen

$$v(x) = \begin{cases} 0 & \text{for } x < a, \\ 1 & \text{for } a \leq x \leq b, \\ 0 & \text{for } b < x. \end{cases}$$

Denne testfunksjonen er lik 1 bare i intervallet $[a, b]$ og 0 ellers, som vist i figuren under.



Vi vet at $u''(x) = f(x)$ fra den sterke formuleringen. Med vår valgte testfunksjon $v(x)$ får vi:

$$\int_0^1 u''(x) \cdot v(x)\, dx = \int_0^1 f(x) \cdot v(x)\, dx$$
$$\int_0^1 u''(x) \cdot v(x)\, dx = \int_a^b f(x) \cdot 1\, dx$$
$$\int_0^1 u''(x) \cdot v(x)\, dx = \int_a^b f(x)\, dx$$

Merk at integralet på høyre side er begrenset til intervallet $[a, b]$ siden $v(x) = 0$ utenfor dette intervallet. På samme måte, på venstre side, bidrar $u''(x)$ kun til integralet når $x \in [a, b]$.

En nyttig egenskap ved denne tilnærmingen er at hvis vi lar intervallet $[a, b]$ bli veldig lite, slik at $b \to a$, kan vi intuitivt tenke at:

$$\int_a^b u''(x)\,dx \approx u''(a) \cdot (b - a)$$

Tilsvarende for høyresiden:

$$\int_a^b f(x)\,dx \approx f(a) \cdot (b - a)$$

Dette gir oss $u''(a) \approx f(a)$ når $b \to a$, som stemmer med vår opprinnelige differensialligning.

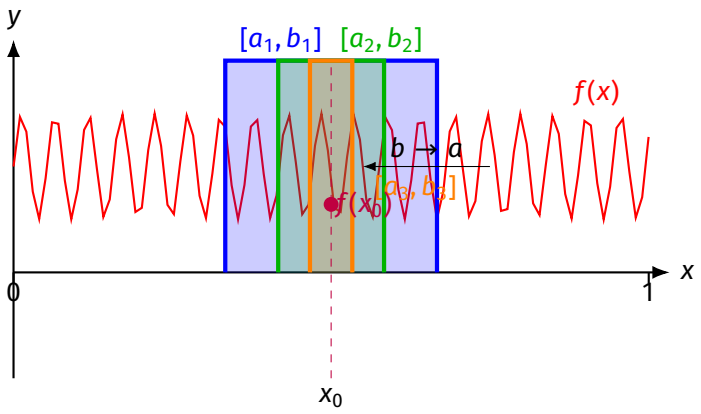


Figure 2.7: Når vi gjør intervallet $[a, b]$ stadig mindre, får vi til slutt punktverdien $u''(x_0) = f(x_0)$

Men den virkelige styrken med svak formulering kommer når vi bruker delvis integrasjon for å redusere derivasjonsordenene.

2.4.1 Delvis integrasjon og svakere krav til regularitet

La oss nå anvende delvis integrasjon på den svake formuleringen. Vi har generelt at:

$$\int_0^1 u''(x)v(x)\,dx = u'(x)v(x)|_0^1 - \int_0^1 u'(x)v'(x)\,dx$$

Med randbetingelsene våre $u(0) = 0$ og $u'(1) = 0$, og hvis vi krever at $v(x)$ oppfyller $v(0) = 0$ (siden $u(0) = 0$), får vi:

$$\begin{aligned}\int_0^1 u''(x)v(x)\,dx &= u'(x)v(x)|_0^1 - \int_0^1 u'(x)v'(x)\,dx \\ &= u'(1)v(1) - u'(0)v(0) - \int_0^1 u'(x)v'(x)\,dx \\ &= 0 \cdot v(1) - u'(0) \cdot 0 - \int_0^1 u'(x)v'(x)\,dx \\ &= -\int_0^1 u'(x)v'(x)\,dx\end{aligned}$$

Dermed blir den svake formuleringen:

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

Dette er en viktig omformulering fordi:

1. Vi trenger nå bare at $u(x)$ er én gang deriverbar, ikke to ganger som i den sterke formuleringen.

2. Randbetingelsen $u'(1) = 0$ er naturlig inkorporert i formuleringen.

3. Vi kan nå bruke stykkevis lineære funksjoner som har veldefinert første deriverte (nesten overalt), men ikke andre deriverte.

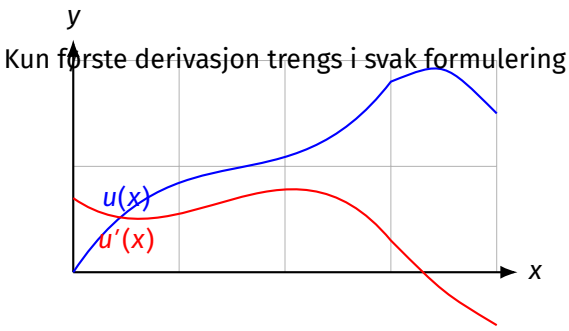


Figure 2.8: I svak formulering trenger vi bare første deriverte av $u(x)$

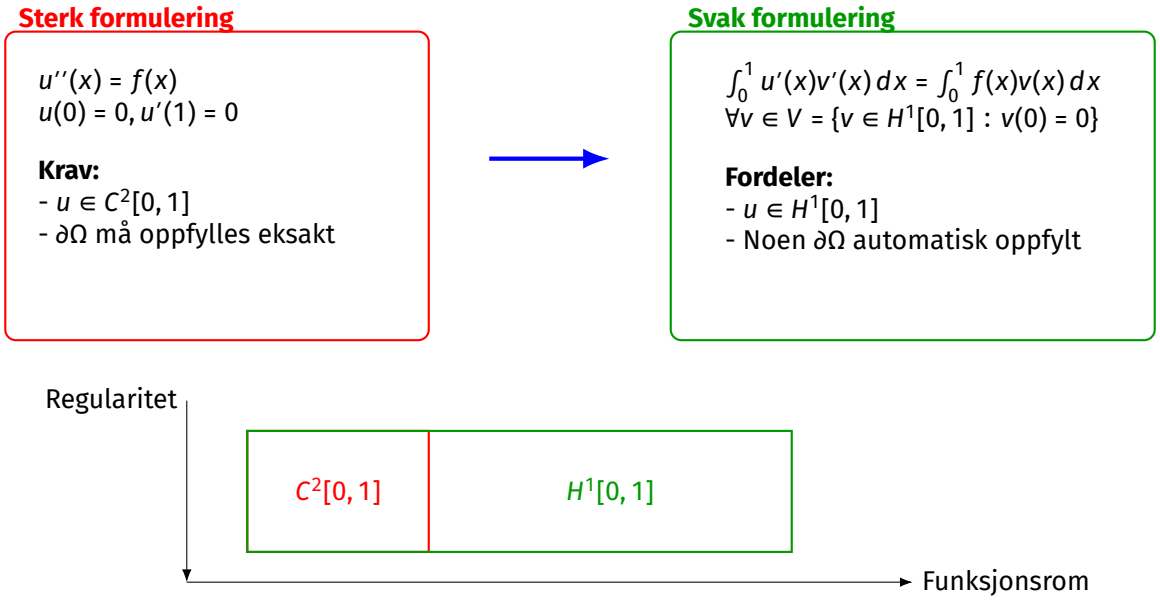


Figure 2.9: Sammenligning av sterk og svak formulering av differensialligninger

2.4.2 Sammenheng med FEM

Den svake formuleringen er grunnlaget for endelig element-metoden (FEM). I FEM antar vi at:

$$u(x) \approx \sum_{i=1}^{n} c_i \varphi_i(x)$$
$$v(x) = \varphi_j(x) \quad \text{for } j = 1, 2, ..., n$$

hvor $\varphi_i(x)$ er basisfunksjoner (typisk stykkevis lineære funksjoner). Ved å sette disse uttrykkene inn i den svake formuleringen, får vi et lineært ligningssystem for koeffisientene c_i:

$$\sum_{i=1}^{n} c_i \int_0^1 \varphi_i'(x)\varphi_j'(x)\,dx = \int_0^1 f(x)\varphi_j(x)\,dx \quad \text{for } j = 1, 2, ..., n$$

Dette kan skrives på matriseform som:

$$\mathbf{Kc} = \mathbf{F}$$

hvor:

$$K_{ji} = \int_0^1 \varphi_i'(x)\varphi_j'(x)\,dx$$
$$F_j = \int_0^1 f(x)\varphi_j(x)\,dx$$

Denne formuleringen er grunnlaget for numerisk løsning av differensialligninger ved hjelp av endelig element-metoden.

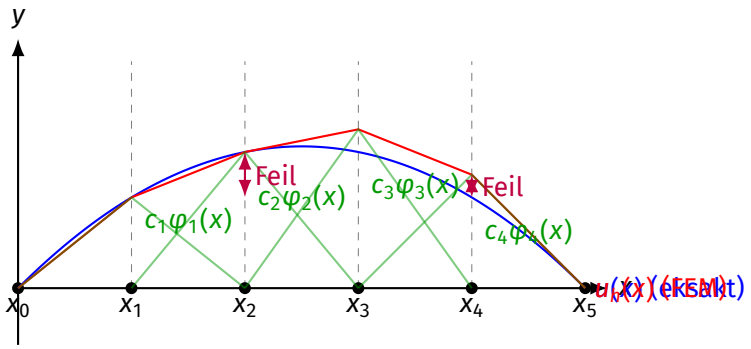


Figure 2.10: FEM-løsning ($u_h(x)$) som en sum av vektede basisfunksjoner, sammenlignet med den eksakte løsningen ($u(x)$)

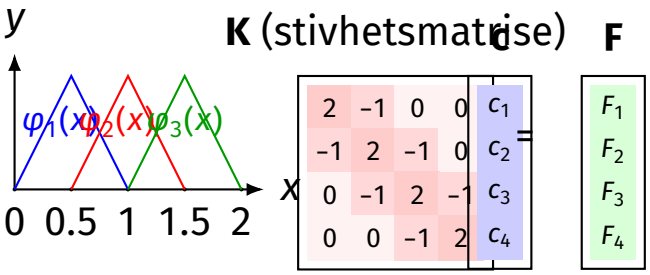


Figure 2.11: FEM diskretisering: basisfunksjoner $\varphi_i(x)$ og tilhørende ligningssystem **Kc** = **F**

Her skriver vi pde direkte i differensialform. Det betyr at vi forutsetter at løsningen $u(x)$ er glatt nok til at alle deriverte eksisterer. Da har vi:

$$\mathcal{L}(u) = f,$$

samt presise krav på grenseverdier, for eksempel:

$$u|_{\partial\Omega} = g \quad \text{eller} \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = h.$$

Hvis løsningen u ikke er glatt nok, bruker vi en weakformulation av problemet.

2.5 Svak form

Her tester vi u med en testfunksjon $v(x)$ over hele domenet:

$$\int_\Omega v^{(k)}(x)\,\mathcal{L}(u(x))\,dx = \int_\Omega v(x)\,f(x)\,dx.$$

Denne metoden gjør det mulig å finne løsninger i et bredere funksjonsrom.

testfunksjon er en vilkårlig funksjon som tilfredsstiller randbetingelsene, og velges ofte til å være det samme som basisfunction:

$$v(x) = \sum_{j=1}^{n} v_j\phi_j(x) = \mathbf{v}^T\boldsymbol{\phi}(x)$$

2.5.1 Fra sterk form

Her skriver vi pde-en direkte i differensialform. Det betyr at vi forutsetter at løsningen u er glatt nok til at alle deriverte eksisterer. Da har vi:

$$\mathcal{L}(u) = f,$$

samt presise krav på grenseverdier, for eksempel:

$$u|_{\partial\Omega} = g \quad \text{eller} \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = h.$$

2.5.2 Til svak form

For løsninger som ikke er tilstrekkelig glatte, bruker vi en weakformulation:

1. Multipliser PDE-en med en testfunksjon $v(x)$
2. Integrer over domenet Ω:

$$\int_\Omega v^{(k)}(x)\,\mathcal{L}(u(x))\,dx = \int_\Omega v(x)\,f(x)\,dx$$

2.5.3 Fordeler med svak form

- Tillater løsninger i bredere funksjonsrom
- Reduserer glatthetskrav
- Gir mer fleksible randbetingelser

En testfunksjon $v(x)$ tilfredsstiller randbetingelsene og uttrykkes i Galerkin-metoden ved:

$$v(x) = \sum_{j=1}^{n} v_j \phi_j(x) = \mathbf{v}^T \boldsymbol{\phi}(x)$$

2.6 Basisfunksjoner (Formfunksjoner)

Basisfunksjoner er byggesteinene i fem-metoden. De er enkle funksjoner som gjør at vi kan representere en komplisert funksjon ved hjelp av enkle byggesteiner.

En basisfunction $\phi_i(x)$ er en lokal funksjon som:

- Er null overalt unntatt i nærheten av node i (lokalt definert).
- Har verdien 1 i node i og 0 i alle andre noder
- Til sammen kan bygge opp løsningen vår $u(x)$ som en sum:

$$u(x) = \sum_{i=1}^{n} u_i \phi_i(x) = \mathbf{u}^T \boldsymbol{\phi}(x)$$

- u_i er koeffisientene som bestemmer hvor mye av hver basisfunction som skal brukes.

2.6.1 Stykkvis lineære basisfunksjoner

La $u(x)$ være en tilnærming til løsningen av et pde, og la $u(x)$ være gitt ved en lineærkombinasjon av basisfunctioner:

$$u(x) = \sum_{i=1}^{6} u_i N_i(x)$$

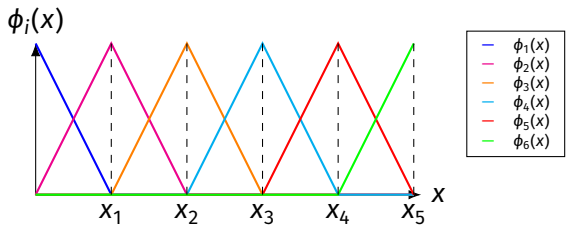


Figure 2.12: Piecewise Linear Basis Functions $\phi_i(x)$

FEM for Poisson-ligningen

$$\begin{cases} -\frac{d^{u(x)}2}{dx^{u(x)}} = f(x), & x \in (0,1) \\ u(0) = \alpha, \quad u(1) = \beta & \text{(randbetingelser)} \end{cases} \tag{2.3}$$

$$\mathbf{F} = -\int_0^1 f(x)\mathbf{N}(x)dx \tag{2.4}$$

La $f(x) = \bar{f} = C$ være konstant.

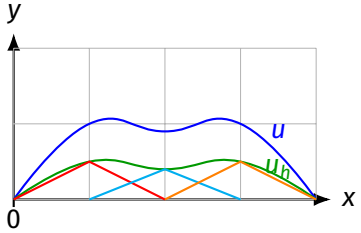
Vi antar at $u(x_m)$ er ukjent for $m = 1, \dots, M$ punkter (noder) i det diskrete domainet Ω_h.

I mellom nodene definerer vi *elementene* $\implies$ *formfunksjoner* $N_i(x)$.

Example 6

- **Eksakt løsning:** u
- **Numerisk løsning:** u_h
- **Feil:** $e = \|u - u_h\|$

La $\|u - v\|_{\mathcal{H}^1}$ for en tilfeldig $v \in \mathcal{X}_h^1$.



2.7 Interpolasjonsoperator

Vi definerer en interpolasjonsoperator $\Pi_h^1 : \mathbb{H}(\Omega)^1 \to \mathbb{X}_h^1$.

$$\Pi_h^1 v(x_i) = v(x_i) \quad i = 0, 1, \dots, M$$

$$\Pi_h^1 v(x) = \sum_{i=0}^{M} v(x_i)\varphi_i(x) \in \mathbb{X}_h^1 = V_h$$

Med interpolasjonsfeilen:

$$e(x) = v(x) - \Pi_h^1 v(x)$$

som vi ønsker å finne en øvre grense for $\|e(x)\|_{\mathbb{H}^1}$.

La nå $H^2(\Omega) = \{v \in H^1(\Omega) : v_{xx} \in L^2(\Omega)\}$ være et Hilbert-rom med norm $|v|_{H^2}^2 = \int_\Omega v_{xx}^2 \, dx$.

$$\begin{aligned}\|e(x)\|_{\mathbb{H}^1} &= \int_0^1 e^2 \, dx + \int_0^1 e_x^2 \, dx \\ &= \sum_{k \in \mathcal{T}_h} \int_k e^2 \, dx + \sum_{k \in \mathcal{T}_h} \int_k e_x^2 \, dx\end{aligned}$$

Vi vet at $e(x_k) = e(x_{k+1}) = 0$.

For $x > \xi_k$:

$$e_x(x) = \int_{\xi_k}^x e_{xx} \, ds = \int_{\xi_k}^x v_{xx} \, ds$$

Fordi $(\Pi_h^1 v)_{xx} = 0$.

For $x < \xi_k$:

$$\begin{aligned}e_x(x) &= -\int_x^{\xi_k} v_{xx} \, ds \\ |e_x|_K \le \int_{K_k} |v_{xx}| \, dx &= \int_{K_k} 1 \cdot |v_{xx}| \, dx \\ &= \langle 1, |v_{xx}| \rangle_{L^2(K_k)} \le \|1\|_{L^2(K_k)} \| |v_{xx}| \|_{L^2(K_k)} \qquad \text{(Cauchy-Schwarz)} \\ |e_x|_{K_k} &\le \underbrace{\sqrt{\int_{K_k} 1^2 \, ds}}_{h_k} \times \sqrt{\int_{K_k} v_{xx}^2 \, ds} \quad \text{hvor } h_k = x_{k+1} - x_k \\ |e_x|_{K_k}^2 &\le h_k \int_{K_k} v_{xx}^2 \, ds \\ h &= \max_{k \in \mathcal{T}_h} h_k \\ |e|_{H^1(K)}^2 &= \int_{K_k} e_x^2 \, dx \le h_k^2 \int_{K_k} v_{xx}^2 \, dx \\ |e|_{H^1(\Omega)}^2 &= \sum_{k \in \mathcal{T}_h} |e_x|_{H^1(K)}^2 \le h^2 \sum_{k \in \mathcal{T}_h} \int_{K_k} v_{xx}^2 \, dx = h^2 \int_\Omega v_{xx}^2 \, dx = h^2 |v|_{H^2(\Omega)}^2 \\ \|e\|_{\mathcal{L}^2(\Omega)} &\le h^4 |v|_{H^2(\Omega)}^2 \\ \|e\|_{H^1(\Omega)} &= \|e\|_{\mathcal{L}^2(\Omega)} + |e|_{H^1(\Omega)} \le (h^4 + h^2) |v|_{H^2(\Omega)}^2 \le K^2 h^2 \|v\|_{H^2(\Omega)}^2\end{aligned}$$

Hvis vi bruker Ceas lemma med $v = \Pi_h^1(u)$, får vi:

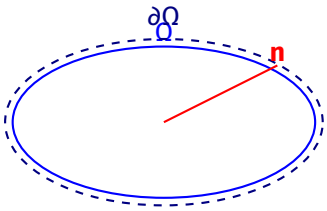
$$\|u - u_h\|_{H^1(\Omega)} \le \frac{M}{\alpha} |u|_{H^2(\Omega)} h \le C |u|_{H^2(\Omega)} h$$

Er en første ordens metode. Fungerer kun hvis $u \in H^2(\Omega)$.

2.7.1 Modell problem

$$-\Delta u(x) = f(x) \quad x \in \Omega$$
$$u(x) = 0 \quad x \in \partial\Omega$$

- Hva er den svake formuleringen av dette problemet?
- Er LM-teoremet oppfylt?
- Hva er X_h^1 i dette tilfellet?



- Ω er en åpen, begrenset og sammenhengende delmengde.
- $\partial\Omega$ er den lukkede mengden av alle punkter i Ω.
- $\overline{\Omega} = \Omega \cup \partial\Omega$ er den lukkede mengden av alle punkter i Ω.
- $v : \Omega \to \mathbb{R}$
- $\mathbf{F} : \Omega \to \mathbb{R}^2$

Vektor notasjon

Gradient

$$\nabla u(x) = \begin{pmatrix} \frac{\partial u}{\partial x_1} \\ \frac{\partial u}{\partial x_2} \end{pmatrix} = \begin{pmatrix} u_{x_1} \\ u_{x_2} \end{pmatrix} = (u_{x_1}, u_{x_2})^T$$

Divergens

$$\nabla \cdot \mathbf{F} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} = F_{x_1} + F_{x_2}$$

Laplace operator

$$\Delta u(x) = \nabla^2 u(x) = \nabla \cdot \nabla u(x) = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = u_{x_1 x_1} + u_{x_2 x_2}$$

Theorem 2.7.1. Divergensteoremet

$$\int_\Omega \nabla \cdot \mathbf{F}\, dx = \int_{\partial\Omega} \mathbf{F} \cdot \mathbf{n}\, ds$$

- $\mathbf{F}$ er en vektorfelt.
- $\mathbf{n}$ er en normalvektor til grensen $\partial\Omega$.
- ds er et infinitesimalt areal på grensen $\partial\Omega$.
- dx er et infinitesimalt areal i domenet Ω.
- $\int_{\partial\Omega} \mathbf{F} \cdot \mathbf{n}\, ds$ er fluks gjennom grensen $\partial\Omega$.
- $\int_\Omega \nabla \cdot \mathbf{F}\, dx$ er divergensen av vektorfeltet $\mathbf{F}$ i domenet Ω.

Laplace operator

$$\Delta u(x) = \nabla^2 u(x) = \nabla \cdot \nabla u(x) = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = u_{x_1 x_1} + u_{x_2 x_2}$$

Vektor Kalkulus

$$\nabla \cdot (v\mathbf{F}) = \nabla v \cdot \mathbf{F} + v \nabla \cdot \mathbf{F}$$

Green's teorem La $u : \Omega \to \mathbb{R}$ og $\mathbf{F} = \nabla u$ være en vektorfelt. Da har vi at:

$$\int_\Omega \nabla v \cdot \nabla u \, d\Omega + \int_\Omega v \Delta u \, d\Omega = \oint_{\partial\Omega} v \nabla u \cdot \mathbf{n} \, d\gamma$$

$-\Delta u = f$ på Ω og $u = 0$ på $\partial\Omega$ gir oss:

$$\int_\Omega \nabla v \cdot \nabla u \, d\Omega = \oint_{\partial\Omega} v \nabla u \cdot \mathbf{n} \, d\gamma$$

Krever at $v = 0$ på $\partial\Omega$.

Definerer:

$$\begin{aligned} H^1(\Omega) &= \{v : \Omega \to \mathbb{R}, v, v_{x_1}, v_{x_2} \in L^2(\Omega)\} \\ L^2(\Omega) &= \{v : \int_\Omega v^2 \, dx < \infty\} \\ \langle u, v \rangle_{L^2(\Omega)} &= \int_\Omega uv \, d\Omega, \quad \|v\|^2_{L^2(\Omega)} = \int_\Omega v^2 \, d\Omega \\ H_0^1(\Omega) &= \{v \in H^1(\Omega), v = 0 \text{ på } \partial\Omega\} \end{aligned}$$

Variasjonsformulering (Variational form) Finn $u \in H_0^1(\Omega)$ slik at $\int_\Omega \nabla u \cdot \nabla v \, d\Omega = \int_\Omega fv \, d\Omega$ for alle $v \in H_0^1(\Omega)$.

$$\begin{aligned} V &= H_0^1(\Omega) \\ a(u,v) = \int_\Omega \nabla u \cdot \nabla v \, d\Omega, \quad F(v) &= \int_\Omega fv \, d\Omega \end{aligned}$$

Example 7. Poisson

Sterk form Finn $u : \Omega \to \mathbb{R}$ slik at

$$\begin{aligned} -\Delta u &= f \quad \text{i } \Omega, \\ u &= 0 \quad \text{på } \partial\Omega, \end{aligned}$$

Svak form Finn $u \in V$ slik at

$$\langle \nabla u, \nabla \phi \rangle = \langle f, \phi \rangle \quad \forall \phi \in V := H_0^1(\Omega) = \{u \in H^1(\Omega) : u = 0 \text{ på } \partial\Omega\}$$

Diskret svak form Finn $u_h \in V_h \subset V$ slik at

$$\langle \nabla u_h, \nabla \phi_i \rangle = \langle f, \phi_i \rangle \quad \forall i = 1, \dots, n,$$

Løsning Løs det lineære systemet $A\vec{u} = \vec{f}$, der

$$\langle \nabla u_h, \nabla \phi_i \rangle = \langle f, \phi_i \rangle, \quad \Leftrightarrow \quad A\vec{u} = \vec{f}$$
$$A_{ij} = \left\langle \nabla \phi_j, \nabla \phi_i \right\rangle, \quad \vec{u}_i, \quad \vec{f}_i = \langle f, \phi_i \rangle$$

2.7.2 Ekvivalente former

Sterk form

Finn $u \in \mathcal{C}^2(\Omega)$ slik at

$$\begin{aligned} -u'' &= f \quad \text{i } \Omega, \\ u &= 0 \quad \text{på } \partial\Omega, \end{aligned}$$

Svak form

Finn $u \in V = H_0^1(\Omega)$ slik at

$$\langle u', \phi' \rangle = \langle f, \phi \rangle \quad \forall \phi \in V := H_0^1(\Omega)$$

Energiminimeringsform

La $F : V \to \mathbb{R}$ være en energifunksjon definert som

$$F(u) = \frac{1}{2}\langle u', u' \rangle - \langle f, u \rangle$$

Der $F(u)$ representerer energien i systemet. Den første termen representerer den kinetiske energien, mens den andre termen representerer den potensielle energien.

Finn $u \in V$ slik at

$$F(u) \leq F(\phi) \quad \forall \phi \in V$$

Dette er en variabel minimeringsoppgave der vi ønsker å finne den funksjonen u som minimerer energien i systemet.[1]

2.8 Eksempler

2.8.1 Poisson-ligningen

Example 8. Poisson-ligningen

Poisson-ligningen er en vanlig pde som beskriver mange fysiske fenomener, inkludert varmeledning og elektrisk potensial. Den kan skrives som:

$$\begin{cases} -\Delta u(x) = f(x), & x \in (0,1) \\ u(0) = \alpha, \quad u(1) = \beta & \text{(randbetingelser)} \end{cases} \tag{2.5}$$

$$\mathbf{F} = -\int_0^1 f(x)\mathbf{N}(x)dx \tag{2.6}$$

[1] Det er ikke alltid at denne formen eksisterer for problemet vårt.

$$\begin{cases} -\frac{d^2u}{dx^2} = f(x), & x \in (0,1) \\ u(0) = \alpha, \quad u(1) = \beta & \text{(randbetingelser)} \end{cases} \tag{2.7}$$

La $f(x) = \bar{f}$ være konstant. Vi antar at $u(x_m)$ er ukjent for $m = 1, \ldots, M$ punkter (noder) i det diskrete domenet Ω_h.

I mellom nodene definerer vi *elementene* $\implies$ *formfunksjoner* $N_i(x)$.

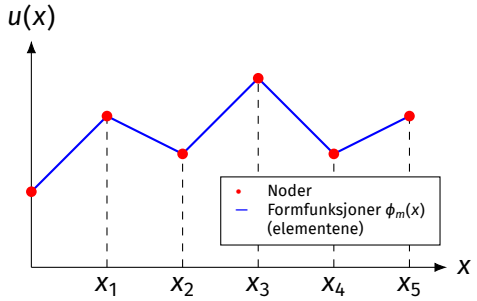


Figure 2.13: Tilfeldig valgt formfunksjoner $N_i(x)$ mellom nodene x_i.

Definerer formfunksjonene som:

$$\phi_i(x) = \begin{cases} 1 - 2|x - x_i|, & |x - x_i| < 0.5 \\ 0, & \text{ellers} \end{cases}$$

Og testfunksjonene som:

$$v(x) = \sum_{j=1}^{n} v_j \phi_j(x) = \mathbf{v}^T \boldsymbol{\phi}(x)$$

Den svake formuleringen blir:

$$\int_0^1 \left(\sum_{j=1}^{n} v_j \phi_j'(x) \right) \left(\sum_{i=1}^{n} u_i \phi_i'(x) \right) dx = \int_0^1 \left(\sum_{j=1}^{n} v_j \phi_j(x) \right) f(x)\, dx$$

$$\mathbf{v}^T \int_0^1 \boldsymbol{\phi}' \boldsymbol{\phi}'^T dx\, \mathbf{u} = \mathbf{v}^T \int_0^1 -f(x) \boldsymbol{\phi}\, dx$$

$$\mathbf{v}^T \mathbf{K} \mathbf{u} = \mathbf{v}^T \mathbf{F}$$

$$\mathbf{K}\mathbf{u} = \mathbf{F}$$

Hvor:

$$\mathbf{K} = \int_0^1 \boldsymbol{\phi}'\boldsymbol{\phi}'^T dx = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$$\mathbf{F} = -\bar{f} \int_0^1 \mathbf{N}(x) dx = -\bar{f} \begin{bmatrix} 0.1 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.1 \end{bmatrix}$$

Løser ligningssystemet for å finne koeffisientene u_i:

$$\mathbf{u} = \mathbf{K}^{-1}\mathbf{F}$$

2.8.2 Svak formulering av Poisson-ligningen

Vi kan skrive den svake formuleringen som:

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

Velger basisfunctioner til å være:

$$\phi_i(x) = \begin{cases} 1 - 2|x - x_i|, & |x - x_i| < 0.5 \\ 0, & \text{ellers} \end{cases}$$

Med testfunksjonene $v(x) = \sum_{j=1}^{n} v_j\phi_j(x)$.

Da får vi:

$$\int_0^1 \left(\sum_{j=1}^{n} v_j\phi_j'(x)\right)\left(\sum_{i=1}^{n} u_i\phi_i'(x)\right) dx = \int_0^1 \left(\sum_{j=1}^{n} v_j\phi_j(x)\right) f(x)\,dx$$

$$\mathbf{v}^T \int_0^1 \boldsymbol{\phi}'\boldsymbol{\phi}'^T dx\,\mathbf{u} = \mathbf{v}^T \int_0^1 -f(x)\boldsymbol{\phi}\,dx$$

$$\mathbf{v}^T\mathbf{K}\mathbf{u} = \mathbf{v}^T\mathbf{F}$$

$$\mathbf{K}\mathbf{u} = \mathbf{F}$$

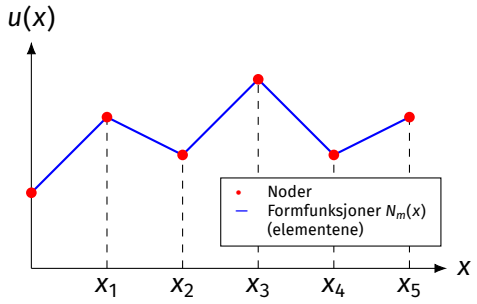


Figure 2.14: Tilfeldig valgt formfunksjoner $N_i(x)$ mellom nodene x_i.

$$\mathbf{F} = -\bar{f}\int_0^1 \mathbf{N}(x)dx = -\bar{f}\begin{bmatrix}0.1\\0.2\\0.2\\0.2\\0.2\\0.1\end{bmatrix}$$

Example 9. Eksamen Vår 2024

Vi har ett en-dimesjonalt elliptisk randverdi-problem.

$$-\mathcal{L}u = -\frac{d}{dx}\left((1+x)\frac{du}{dx}\right) + 2u = 2x \quad x \in (0,1) \quad u(0) = \sqrt{2}, \quad u(1) = \sqrt{3}$$

La nå $M \in \mathbb{N}$ og $\mathsf{T}_h = \bigcup_{i=1}^{M} K_i$ være en triangulering av $(0,1)$ med $K_i = (x_{i-1}, x_i)$, $x_i = ih$ og $h = \frac{1}{M}$.

$$\mathsf{T}_h = \{K_1, K_2, \dots, K_M\} = \{(x_0, x_1), (x_1, x_2), \dots, (x_{M-1}, x_M)\}$$

Finn **elementmatrisen** A^{K_i} og **lastvektoren** $\mathbf{F}^{K_i}$ for Lagrange FEM i $\mathbb{P}_1$ på trianguleringen T_h for problemet i (2.8.2).
På $\mathbb{P}_1$ er standard Lagrange-FEM-basisfunksjonene gitt ved:

$$\phi_1(x) = \frac{x_i - x}{h} \qquad \phi_1(x_{i-1}) = 1, \phi_1(x_i) = 0$$
$$\phi_2(x) = \frac{x - x_{i-1}}{h} \qquad \phi_2(x_{i-1}) = 0, \phi_2(x_i) = 1$$

Først finner vi testfunksjonene $v \in \mathcal{H}_0^1(0,1)$ og basisfunksjonene $\phi_i(x)$ for $i = 1, \dots, M$.

$$v(x) = \sum_{i=1}^{M} v_i \phi_i(x)$$

2.8.3 Svak formulering

Den svake formuleringen av problemet i (2.8.2) finner man ved å gange PDE med testfunksjonen v på begge sidene og integrere over $(0, 1)$:

$$
\begin{aligned}
-\int_0^1 \mathcal{L}u(x)v(x)\,dx &= 0 \\
\int_0^1 \left[-\frac{d}{dx}\left((1+x)\frac{du}{dx}\right) + 2u\right] v(x)\,dx &= 2\int_0^1 xv(x)\,dx \\
-\int_0^1 \frac{d}{dx}\left((1+x)u'(x)\right)v(x)\,dx + 2\int_0^1 u(x)v(x)\,dx &= 2\int_0^1 xv(x)\,dx
\end{aligned}
$$

For det første leddet i LHS bruker vi delvis integrasjon:

$$
\begin{aligned}
-\int_0^1 \overbrace{\frac{d}{dx}\left((1+x)u'(x)\right)}^{w'(x)} v(x)\,dx &= -\left[w(x)v(x)\right]_0^1 + \int_0^1 w(x)v'(x)\,dx \\
&= -\left[(1+x)u'(x)v(x)\right]_0^1 + \int_0^1 (1+x)u'(x)v'(x)\,dx \\
&= -\left[(1+1)u'(1)v(1) - (1+0)u'(0)v(0)\right] + \int_0^1 (1+x)u'(x)v'(x)\,dx \\
&= -\left[2u'(1)v(1) - u'(0)v(0)\right] + \int_0^1 (1+x)u'(x)v'(x)\,dx \\
&= 0 + \int_0^1 (1+x)u'(x)v'(x)\,dx
\end{aligned}
$$

Siden $v \in \mathcal{H}_0^1(0,1)$, så er $v(0) = v(1) = 0$ får vi at $[w(x)v(x)]_0^1 = 0$.

Dermed kan vi skrive om den svake formuleringen til:

$$
\begin{aligned}
\int_0^1 (1+x)u'(x)v'(x)\,dx + 2\int_0^1 u(x)v(x)\,dx &= 2\int_0^1 xv(x)\,dx \\
\int_0^1 (1+x)u'(x)v'(x) + 2u(x)v(x)\,dx &= 2\int_0^1 xv(x)\,dx
\end{aligned}
$$

Vi approksimerer u med Lagrange-FEM-basisfunksjoner $\phi_i(x)$ i $\mathbb{P}_1$:

$$
\begin{aligned}
u(x) &\approx \sum_{i=1}^{M} u_i\phi_i(x) = \mathbf{u}^T\boldsymbol{\phi}(x) \\
v(x) &\approx \sum_{i=1}^{M} v_i\phi_i(x) = \mathbf{v}^T\boldsymbol{\phi}(x)
\end{aligned}
$$

Setter inn i den svake formuleringen:

$$\int_0^1 (1+x)\left(\sum_{i=1}^M u_i\phi_i'(x)\right)\left(\sum_{j=1}^M v_j\phi_j'(x)\right) + 2\left(\sum_{i=1}^M u_i\phi_i(x)\right)\left(\sum_{j=1}^M v_j\phi_j(x)\right) dx = 2\int_0^1 x\left(\sum_{i=1}^M u_i\phi_i(x)\right)\left(\sum_{j=1}^M v_j\phi_j(x)\right) dx$$

$$\int_0^1 (1+x)(\mathbf{u}^\top\boldsymbol{\phi}'(x))(\mathbf{v}^\top\boldsymbol{\phi}'(x))\,dx + \int_0^1 2(\mathbf{u}^\top\boldsymbol{\phi}(x))(\mathbf{v}^\top\boldsymbol{\phi}(x))\,dx = 2\int_0^1 x(\mathbf{u}^\top\boldsymbol{\phi}(x))(\mathbf{v}^\top\boldsymbol{\phi}(x))\,dx$$

$$\int_0^1 (1+x)(\mathbf{u}^\top\boldsymbol{\phi}'(x))(\boldsymbol{\phi}'(x)^\top\mathbf{v})\,dx + \int_0^1 2(\mathbf{u}^\top\boldsymbol{\phi}(x))(\boldsymbol{\phi}(x)^\top\mathbf{v})\,dx = 2\int_0^1 x(\mathbf{u}^\top\boldsymbol{\phi}(x))(\boldsymbol{\phi}(x)^\top\mathbf{v})\,dx$$

$$\mathbf{u}^\top\int_0^1 (1+x)\boldsymbol{\phi}'(x)\boldsymbol{\phi}'(x)^\top\,dx\,\mathbf{v} + 2\mathbf{u}^\top\int_0^1 \boldsymbol{\phi}(x)\boldsymbol{\phi}(x)^\top\,dx\,\mathbf{v} = 2\mathbf{u}^\top\int_0^1 x\boldsymbol{\phi}(x)\boldsymbol{\phi}(x)^\top\,dx\,\mathbf{v}$$

$$\mathbf{u}^\top\left[\int_0^1 (1+x)\boldsymbol{\phi}'(x)\boldsymbol{\phi}'(x)^\top + 2\boldsymbol{\phi}(x)\boldsymbol{\phi}(x)^\top\,dx\right]\mathbf{v} = 2\mathbf{u}^\top\left[\int_0^1 x\boldsymbol{\phi}(x)\boldsymbol{\phi}(x)^\top\,dx\right]\mathbf{v}$$

Vi definerer nå elementmatrisen A^{K_i} og lastvektoren $\mathbf{F}^{K_i}$ som:

$$A^{K_i} = \int_0^1 (1+x)\boldsymbol{\phi}'(x)\boldsymbol{\phi}'(x)^\top + 2\boldsymbol{\phi}(x)\boldsymbol{\phi}(x)^\top\,dx$$

$$\mathbf{F}^{K_i} = 2\int_0^1 x\boldsymbol{\phi}(x)\boldsymbol{\phi}(x)^\top\,dx$$

$$\mathbf{u}^\top A^{K_i}\mathbf{v} = \mathbf{u}^\top\mathbf{F}^{K_i}$$

$$\begin{aligned}
A^{K_i} &= \int_0^1 (1+x)\boldsymbol{\phi}'(x)\boldsymbol{\phi}'(x)^\top + 2\boldsymbol{\phi}(x)\boldsymbol{\phi}(x)^\top\,dx \\
&= \int_0^1 (1+x)\begin{bmatrix}\phi_1'(x) & \phi_2'(x)\end{bmatrix}\begin{bmatrix}\phi_1'(x)\\ \phi_2'(x)\end{bmatrix} + 2\begin{bmatrix}\phi_1(x) & \phi_2(x)\end{bmatrix}\begin{bmatrix}\phi_1(x)\\ \phi_2(x)\end{bmatrix} dx \\
&= \int_0^1 (1+x)\begin{bmatrix}\phi_1'(x)\phi_1'(x) & \phi_1'(x)\phi_2'(x)\\ \phi_2'(x)\phi_1'(x) & \phi_2'(x)\phi_2'(x)\end{bmatrix} + 2\begin{bmatrix}\phi_1(x)\phi_1(x) & \phi_1(x)\phi_2(x)\\ \phi_2(x)\phi_1(x) & \phi_2(x)\phi_2(x)\end{bmatrix} dx \\
&= \int_0^1 (1+x)\begin{bmatrix}\phi_1'(x)^2 & \phi_1'(x)\phi_2'(x)\\ \phi_2'(x)\phi_1'(x) & \phi_2'(x)^2\end{bmatrix} dx + 2\int_0^1 \begin{bmatrix}\phi_1(x)^2\,dx & \phi_1(x)\phi_2(x)\,dx\\ \phi_2(x)\phi_1(x)\,dx & \phi_2(x)^2\,dx\end{bmatrix} dx
\end{aligned}$$

Hvor de deriverte av Lagrange-basisfunksjonene er:

$$\phi_1'(x) = -\frac{1}{h}, \quad \phi_1'(x)^2 = \frac{1}{h^2},$$

$$\phi_2'(x) = \frac{1}{h}, \quad \phi_2'(x)^2 = \frac{1}{h^2},$$

$$\phi_1'(x)\phi_2'(x) = \phi_2'(x)\phi_1'(x) = \frac{-1}{h^2} = -\frac{1}{h^2}$$

Lectures

.1 Lecture 6: 04.09.2025

Now let's consider more complex BCs

$$\begin{aligned} -\Delta \mathbf{u} &= \mathbf{f} \quad \text{in } \Omega \\ \mathbf{u} &= \mathbf{g} \quad \text{on } \Gamma_D \\ \nabla \mathbf{u} \cdot \mathbf{n} &= l \quad \text{on } \Gamma_N \end{aligned}$$

where $\Gamma_D \cup \Gamma_N = \partial\Omega$ and $\Gamma_D \cap \Gamma_N = \emptyset$.

We again test with an arbitrary function $\mathbf{v}$:

$$\langle \mathbf{f}, \mathbf{v} \rangle = \langle -\Delta \mathbf{u}, \mathbf{v} \rangle = \langle \nabla \mathbf{u}, \nabla \mathbf{v} \rangle + \int_{\partial\Omega} \nabla \mathbf{u} \cdot \mathbf{n}\, \mathbf{v}\, ds$$

As we have multiple BCs, on different subdomains we can consider them independently.

$$\int_{\partial\Omega} \nabla \mathbf{u} \cdot \mathbf{n}\, \mathbf{v}\, ds = \int_{\Gamma_D} \nabla \mathbf{u} \cdot \mathbf{n}\, \mathbf{v}\, ds + \int_{\Gamma_N} \nabla \mathbf{u} \cdot \mathbf{n}\, \mathbf{v}\, ds = 0 + \int_{\Gamma_N} l\, \mathbf{v}\, ds$$

- **Neumann:** The natural BC becomes:

  $$\int_{\Gamma_N} \nabla \mathbf{u} \cdot \mathbf{n}\, \mathbf{v}\, ds = \int_{\Gamma_N} l\, \mathbf{v}\, ds$$

  and remains in our formulation.

- **Dirichlet:** By specifying $\mathbf{u} \in \mathcal{H}^1_{\Gamma_D}(\Omega)$ where $\mathcal{H}^1_{\Gamma_D}(\Omega) = \{\mathbf{v} \in \mathcal{H}^1(\Omega) : \mathbf{v}|_{\Gamma_D} = \mathbf{g}\}$ we obtain the weak form:

  $$\text{Find } \mathbf{u} \in \mathcal{H}^1_{\Gamma_D}(\Omega) \text{ s.t. } \langle \nabla \mathbf{u}, \nabla \mathbf{v} \rangle = \langle \mathbf{f}, \mathbf{v} \rangle + \int_{\Gamma_N} l\, \mathbf{v}\, ds \quad \forall \mathbf{v} \in \mathcal{H}^1_0(\Omega)$$

Warning: Trial and test spaces dont match! And $\mathcal{H}^1_{\Gamma_D}(\Omega) \not\subset \mathcal{H}^1(\Omega)$. **Solution:** Lift the solution $\mathbf{u}$ so we solve a homogeneous problem. Suppose we have operator $R_g \in \mathcal{H}^1(\Omega)$ s.t. $R_g|_{\Gamma_D} = \mathbf{g}$, then we set $\odot\mathbf{u} = \mathbf{u} - R_g$ and solve for $\odot\mathbf{u}$:

$$\text{Find } \odot\, \mathbf{u} \in \mathcal{H}^1_D(\Omega) \text{ s.t. } \langle \nabla \odot \mathbf{u}, \nabla \mathbf{v} \rangle = \langle \mathbf{f}, \mathbf{v} \rangle + \int_{\Gamma_N} l\, \mathbf{v}\, ds - \langle \nabla R_g, \nabla \mathbf{v} \rangle \quad \forall \mathbf{v} \in \mathcal{H}^1_D(\Omega)$$

where

$$\mathcal{H}^1_D(\Omega) = \{\mathbf{v} \in \mathcal{H}^1(\Omega) : \mathbf{v}|_{\Gamma_D} = 0\}$$

Thus the problem is *symmetric* again!

Exercise: More general elliptic problems:

Find the weak form for the PDE

$$\begin{aligned} -\operatorname{div}(\mu\nabla\mathbf{u}) + \sigma\mathbf{u} &= \mathbf{f} \quad \text{in } \Omega \\ \mathbf{u} &= \mathbf{g} \quad \text{on } \Gamma_D \\ \mu\nabla\mathbf{u}\cdot\mathbf{n} &= l \quad \text{on } \Gamma_N \end{aligned}$$

Functional analysis recap (also see Q2)

A functional F on V is an operator mapping $F : V \to \mathbb{R}$. The functional we have in mind is :

$$F = \langle \mathbf{f}, \mathbf{v} \rangle$$

A linear functional is said to be bounded if $\exists C > 0$ s.t.

$$|F(\mathbf{v})| \leq C\|\mathbf{v}\|_V \quad \forall \mathbf{v} \in V$$

If a functional is linear and bounded on some Banach space, then it is continuous.

Banach space:

- A vector space V over $\mathbb{R}$ or $\mathbb{C}$
- Equipped with a norm $\|\cdot\|_V$
- Complete w.r.t. the metric induced by the norm

A bilinear form $a : V \times V \to \mathbb{R}$ is continuous if $\exists M > 0$ s.t.

$$|a(\mathbf{u}, \mathbf{v})| \leq M\|\mathbf{u}\|_V\|\mathbf{v}\|_V \quad \forall \mathbf{u}, \mathbf{v} \in V$$

and coercive if $\exists \alpha > 0$ s.t.

$$a(\mathbf{v}, \mathbf{v}) \geq \alpha\|\mathbf{v}\|_V^2 \quad \forall \mathbf{v} \in V$$

Theorem .1.1. Lax-Milgram

Let V be a Hilbert space, $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ a continuous and coercive bilinear form, and $F : V \to \mathbb{R}$ a continuous linear functional. Then there exists a unique solution to the problem:

$$\text{Find } \mathbf{u} \in V \text{ s.t. } a(\mathbf{u}, \mathbf{v}) = F(\mathbf{v}) \quad \forall \mathbf{v} \in V$$

Moreover, $\|\mathbf{u}\|_V \leq \frac{1}{\alpha}\|F\|_{V'}$ where $\|F\|_{V'} = \sup_{\mathbf{v}\in V\setminus\{0\}} \frac{|F(\mathbf{v})|}{\|\mathbf{v}\|_V}$.

- Based on *Riesz representation theorem* and the *Banach fixed point theorem*.
- To prove unique solutions exists we need to satisfy the conditions on the linear F and bilinear form $a(\cdot,\cdot)$.

Example 10

Consider seeking $u \in H_0^1([0,1])$ s.t.

$$\begin{aligned} \int_0^1 u'(x)v'(x)\,dx &= \int_0^1 \sin(2\pi x)v(x)\,dx \quad \forall v \in H_0^1([0,1]) \\ a(u,v) &= \int_0^1 u'(x)v'(x)\,dx \\ F(v) &= \int_0^1 \sin(2\pi x)v(x)\,dx \end{aligned}$$

Clearly

$$|F(v)| \le \|\sin(2\pi x)\|_{L^2}\|v\|_{L^2} \le \|v\|_{L^2} \le \|v\|_{H_0^1}$$
$$\|v\|_{H^1}^2 = \|v\|_{L^2}^2 + \|\nabla v\|_{L^2}^2 \ge \|v\|_{L^2}^2$$

so F is bounded and linear. Also

$$|a(u,v)| = |\langle u,v\rangle_{H_0^1}| \le \|u\|_{H_0^1}\|v\|_{H_0^1} \quad \text{(Cauchy-Schwarz)}$$

thus continuous. For coercivity we must use the Poincaré inequality:

$$\|v\|_{L^2} \le C\|\nabla v\|_{L^2} \quad \forall v \in H_0^1(\Omega)$$

so

$$a(v,v) = \|\nabla v\|_{L^2}^2 \ge \frac{1}{C^2}\|v\|_{L^2}^2 \ge \frac{1}{C^2+1}\|v\|_{H_0^1}^2$$

and we have coercivity. Thus by Lax-Milgram a unique solution exists.

Elliptic finite elements in 1D

We have discussed the infinitely dimensional case, now we want to move into the finite dimensional case.

Definition .1.2. Ritz-Galerkin approximation

Let $a : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ be a bilinear form, and $\mathbb{V}_h \subset \mathbb{V}$ a finite dimensional subspace. Consider the weak form restricted to $\mathbb{V}_h$:

$$\text{Find } \mathbf{u}_h \in \mathbb{V}_h \quad \text{s.t.} \quad a(\mathbf{u}_h, \mathbf{v}) = F(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbb{V}_h$$

This is called a Ritz-Galerkin approximation of the weak solution $\mathbf{u} \in \mathbb{V}$.

Example 11. Polynomial subspace for the 1D problem

Consider a (naive) approx. of the form:

$$\hat{u}(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

to our 1D problem:

$$-u''(x) = f(x) \quad x \in (0,1)$$
$$u(0) = 0$$
$$u'(1) = 0$$

with the weak form:

$$\text{Find } u \in H_0^1(0,1) \quad \text{s.t.} \quad a(u,v) = F(v) \quad \forall v \in H_0^1(0,1)$$

where

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

We observe that $\hat{u}(0) = a_0 = 0$ so we can rewrite:

$$\hat{u}(x) = a_1 x + a_2 x^2 + a_3 x^3$$

The monomials form a polynomial basis of a subspace of $\mathbb{V}_h \subset \mathbb{V}$ where we consider an approx. over a single element. The coefficients of $\hat{u}$ are determined by the constraints:

$$\langle \hat{u}', 1 \rangle = \langle f, x \rangle$$
$$\langle \hat{u}', 2x \rangle = \langle f, x^2 \rangle$$
$$\langle \hat{u}', 3x^2 \rangle = \langle f, x^3 \rangle$$

This is because we want to satisfy the weak form for all $v \in \mathbb{V}_h$ and we have three basis functions. This gives us a linear system of equations for the coefficients a_1, a_2, a_3.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & \frac{4}{3} & \frac{3}{2} \\ 1 & \frac{3}{2} & \frac{9}{5} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \langle f, x \rangle \\ \langle f, x^2 \rangle \\ \langle f, x^3 \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{4}{\pi} + 2 \\ \frac{\pi^2}{4}\left(-\frac{96}{\pi^4} + \frac{12}{\pi^2}\right) \end{bmatrix}$$

which we can solve for:

$$\hat{u}(x) \approx 1.6x - 0.16x^2 - 0.44x^3$$

which remains close to the true solution

$$u(x) = \sin(\frac{\pi x}{2}) \text{ over } [0, 1]$$

Questions:

- Does a discrete solution always exist?
- Is it unique?
- How accurate is it?

Theorem .1.3. 1D existence and uniqueness for $-u''(x) = f(x)$ with $u(0) = u'(1) = 0$

If $f \in L^2([0,1])$ then $\exists! u_h$ to the Ritz-Galerkin approx. when $\mathbb{V}_h \subset \mathbb{V}$ is finite dimensional.

Since $\mathbb{V}_h$ is finite dimensional it has a finite dimensional basis:

$$\mathbb{V}_h = \text{span}\{\phi_1, \phi_2, \dots, \phi_N\}$$

So, for any $u_h \in \mathbb{V}_h$ where:

$$u_h(x) = \sum_{i=1}^{n} \alpha_i \phi_i(x)$$
$$\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^n$$

The Ritz-Galerkin approx. seeks $\boldsymbol{\alpha}$ s.t.

$$a(u_h, v) = F(v) \quad \forall v \in \mathbb{V}_h$$
$$a\left(\sum_{i=1}^{n} \alpha_i \phi_i, v\right) = \langle f, \phi_j \rangle \quad j = 1, 2, \dots, n$$

Since any $v \in \mathbb{V}_h$ is also a linear combination of basis functions this is equivalent to seeking $\boldsymbol{\alpha}$ s.t.

$$a\left(\sum_{i=1}^{n} \alpha_i \phi_i, \phi_j\right) = \langle f, \phi_j \rangle \quad j = 1, 2, \dots, n$$

where the rhs is finite as $f \in L^2([0,1])$.

Through bilinearity:

$$\sum_{i=1}^{n} \alpha_i a(\phi_i, \phi_j) = \langle f, \phi_j \rangle, \quad j = 1, 2, \dots, n$$

Defining the matrix A by $A_{ij} = a(\phi_i, \phi_j)$ and the vector $\mathbf{b}$ by $b_j = \langle f, \phi_j \rangle$, our approximation is equivalent to solving the linear system:

$$A\boldsymbol{\alpha} = \mathbf{b}$$

Existence and uniqueness is equivalent to solving $A\boldsymbol{\alpha} = \mathbf{b}$. Assume A is singular, i.e.

$$\exists \boldsymbol{\beta} \neq 0 \text{ s.t. } A\boldsymbol{\beta} = 0$$

and

$$\boldsymbol{\beta}^\top A \boldsymbol{\beta} = 0$$

We now define:

$$\tilde{v}(x) = \sum_{i=1}^{n} \beta_i \phi_i(x) \in \mathbb{V}_h$$

We see that:

$$\begin{aligned} \boldsymbol{\beta}^\top A \boldsymbol{\beta} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i a(\phi_i, \phi_j) \beta_j \\ &= a(\sum_{i=1}^{n} \beta_i \phi_i, \sum_{j=1}^{n} \beta_j \phi_j) \\ &= a(\tilde{v}, \tilde{v}) = \int_0^1 \tilde{v}'(x)^2 \, dx = 0 \end{aligned}$$

Now, if:

$$\boldsymbol{\beta}^\top A \boldsymbol{\beta} = 0 \implies \tilde{v}'(x) = 0 \, \forall x, implies \tilde{v}(x) = C$$

As $\tilde{v} \in \mathbb{V}_h \subset \mathbb{V} = H_0^1([0,1])$ it must satisfy the BC $\tilde{v}(0) = 0$ so $C = 0$ and $\boldsymbol{\beta} = 0$ which is a contradiction. As A must be non-singular, $A\boldsymbol{\alpha} = \mathbf{b}$ has a unique solution $\boldsymbol{\alpha}$ and thus a unique solution.

.2 Lecture 7: 10.09.2025

Lemma 3. Galerkin Orthogonality Let $u \in V$ be the solution to the weak form and $u_h \in V_h \subset V$ the Ritz-Galerkin approximation.

$$a(u - u_h, v) = 0 \quad \forall v \in V_h$$

Recall the weak form:

$$a(u, v) = F(v) \quad \forall v \in V$$

and the Ritz-Galerkin approx.:

$$a(u_h, v) = F(v) \quad \forall v \in V_h$$

As $V_h \subset V$ through restricting the weak form and:

$$a(u - u_h, v) = \langle f, v \rangle - \langle f, v \rangle = 0 \quad \forall v \in V_h$$

.2.1 Cea's Lemma Vol. 1 (Optimality of Ritz-Galerkin approximation)

Lemma 4. Cea's Lemma If $u \in V$ solves the weak form and $u_h \in V_h$ the Ritz-Galerkin approximation, then:

$$\|u - u_h\|_V = \min_{v \in V_h} \|u - v\|_V$$

Recall that:

$$\|u - u_h\|_V^2 = a(u - u_h, u - u_h)$$

For any $v \in V_h$ we have:

$$a(u - u_h, u - u_h) = a(u - u_h, u - v) + a(u - u_h, v - u_h)$$

Noting that $v - u_h \in V_h$ we use Galerkin orthogonality to eliminate the second term:

$$\|u - u_h\|_V^2 = a(u - u_h, u - v) \quad \forall v \in V_h$$

As $a(\cdot, \cdot)$ is an inner product (continuous and coercive) we can use the Cauchy-Schwarz inequality:

$$\|u - u_h\|_V^2 \leq \|u - u_h\|_V \|u - v\|_V \quad \forall v \in V_h$$

as the statement holds for all $v \in V_h$ it also holds for the minimiser:

$$\|u - u_h\|_V^2 \leq \|u - u_h\|_V \min_{v \in V_h} \|u - v\|_V \quad \square$$

Before proving error estimates in 1D we need some assumptions and results:

- **Approximation assumption:** Given $V_h \subset V$, assume for V_h there $\exists \varepsilon > 0$ s.t. $\forall u \in C^2([0,1]) \cap V$ where:

$$\min_{v \in V_h} \|u - v\|_V \leq \varepsilon \|u''\|_{L^2}$$

- **Theorem: Aubin-Nitsche duality argument:** Let $f \in L^2([0,1])$, $u \in V$ solves the weak form and $u_h \in V$ the Ritz-Galerkin approx. If the approx. assumption holds then:

$$\|u - u_h\|_{L^2} \leq \varepsilon \|u - u_h\|_V$$

  Proof: Let $w \in C^2([0,1]) \cap V$ solve the dual problem:

$$\begin{aligned} -w'' &= u - u_h, \quad x \in (0,1) \\ w(0) &= 0 \\ w'(1) &= 0 \end{aligned}$$

  Then w solves the weak form:

$$a(w, v) = \langle u - u_h, v \rangle \quad \forall v \in V$$

  Since $u - u_h \in V$ we have:

$$\begin{aligned} \|u - u_h\|_{L^2}^2 &= \langle u - u_h, u - u_h \rangle = a(w, u - u_h) \\ &= a(w, u - u_h) - a(u - u_h, v) \quad \forall v \in V_h \text{ (Galerkin orthogonality)} \end{aligned}$$

  Through Galerkin orthogonality. Through Cauchy-Schwarz (CS):

$$\|u - u_h\|_{L^2}^2 \leq \|u - u_h\|_V \|w - v\|_V \quad \forall v \in V_h$$

  Choosing v to be the minimiser we can use the approx. assumption:

$$\|u - u_h\|_{L^2}^2 \leq \varepsilon \|u - u_h\|_V \|w''\|_{L^2} = \varepsilon \|u - u_h\|_V \|u - u_h\|_{L^2}$$

  by construction of the dual problem, thus:

$$\|u - u_h\|_{L^2} \leq \varepsilon \|u - u_h\|_V \quad \square$$

Corollary 1. Let the assumptions of *Aubin-Nitsche* hold. If $f \in C^0([0,1])$ and $u \in C^2([0,1])$, then:

$$\|u - u_h\|_{L^2} \leq \varepsilon \|u - u_h\|_V \leq \varepsilon^2 \|f\|_{L^2}$$

From 4 we have:

$$\|u - u_h\|_V = \min_{v \in V_h} \|u - v\|_V$$

and the approx. assumption:

$$\min_{v \in V_h} \|u - v\|_V \leq \varepsilon \|u''\|_{L^2}$$

we see that:

$$\varepsilon \|u - u_h\|_V \leq \varepsilon \min_{v \in V_h} \|u - v\|_V \leq \varepsilon^2 \|u''\|_{L^2}.$$

Since $u \in C^2([0,1])$ it is a strong solution of the PDE and $\|u''\|_{L^2} = \|f\|_{L^2}$ allowing us to conclude:

$$\|u - u_h\|_{L^2} \leq \varepsilon \|u - u_h\|_V \leq \varepsilon^2 \|f\|_{L^2} \quad \square$$

.2.2 Meshes

Definition .2.1. Mesh

A mesh on $[0,1]$ is given by a set of nodes satisfying:

$$0 = x_0 < x_1 < x_2 < \ldots < x_n = 1$$

which divide $[0,1]$ into n elements:

$$I_i = [x_i, x_{i+1}], \quad i = 0, 1, \ldots, n-1$$

The mesh spacing is defined by $h_i = x_{i+1} - x_i$.

Definition .2.2. Finite Element Space

For $k \geq 1$ we define:

$$\begin{aligned} V_h^k &= \{v \in C^0([0,1]) : v(x) \text{ is polynomial of degree } \leq k \text{ on each element } I_i \text{ and } v(0) = 0\} \\ &= \{v \in C^0([0,1]) : v(x)|_{I_i} \in \mathbb{P}_k,\ i = 0, 1, \ldots, n-1,\ v(0) = 0\} \end{aligned}$$

.3 Lecture 8: 11.09.2025

Recall if $k = 1$ we may use nodal basis functions:

$$\phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & x \in [x_{i-1}, x_i] = I_{i-1} \\ \frac{x_{i+1} - x}{x_{i+1} - x_i}, & x \in [x_i, x_{i+1}] = I_i \\ 0, & \text{otherwise} \end{cases} \qquad i = 1, 2, \ldots, n-1$$

With the last basis function given by:

$$\phi_n(x) = \begin{cases} \frac{x - x_{n-1}}{x_n - x_{n-1}}, & x \in [x_{n-1}, x_n] = I_{n-1} \\ 0, & \text{otherwise} \end{cases}$$

Now we have defined V_h, we can study the error.

Theorem .3.1

Let $h = \max_i h_i$, $w \in C^2([0,1]) \cap V_h^1$ and the interpolant Iw. Then:

$$\|w - Iw\|_V \leq \frac{h}{\sqrt{2}} \|w''\|_{L^2}$$

Proof. We begin by restricting to an arbitrary element I_i noting:

$$\|w - Iw\|_V^2 = \sum_{i=0}^{n-1} \int_{I_i} ([w(x) - Iw(x)]')^2 \, dx$$

and:

$$\|w''\|_{L^2}^2 = \sum_{i=0}^{n-1} \int_{I_i} [w'']^2 \, dx = \sum_{i=0}^{n-1} \int_{I_i} [w'' - (Iw)'']^2 \, dx$$

as $(Iw)'' = 0$ on I_i.
Let/write $e(x) = w(x) - Iw(x)$, then we want to show:

$$\int_{I_i} (e'(x))^2 \, dx \leq \frac{(x_{i+1} - x_i)^2}{2} \int_{I_i} (e''(x))^2 \, dx$$

To simplify we remap our element $I_i \mapsto [0,1]$.
Let $x = x_i + s(x_{i+1} - x_i)$ for $s \in (0,1)$ and define $\tilde{e}(s) = e(x_i + s(x_{i+1} - x_i))$. Then:

$$\frac{d\tilde{e}}{ds} = h_i e'(x_i + s(x_{i+1} - x_i))$$
$$\frac{d^2\tilde{e}}{ds^2} = h_i^2 e''(x_i + s(x_{i+1} - x_i))$$

Allowing us to express what we want to show as:

$$\int_0^1 \left(\frac{d\tilde{e}}{ds}\right)^2 ds \leq \frac{1}{2} \int_0^1 \left(\frac{d^2\tilde{e}}{ds^2}\right)^2 ds$$

Note: we have defined $\tilde{e}(s)$ so it matches $e(x)$ at the nodes:

$$\tilde{e}(x_i) = e(x_{i+1}) = 0 \implies \tilde{e}(0) = \tilde{e}(1) = 0$$

As w and Iw are continuous and differentiable, so is $\tilde{e}$.
Using *Rolle's theorem*:

$$\exists \xi \in (0,1) \text{ s.t. } \tilde{e}'(\xi) = \int_\xi^s \tilde{e}''(t) \, dt \quad \forall s \in [0,1]$$

Using Cauchy-Schwarz (CS):

$$|\tilde{e}'(s)| = \left| \int_\xi^s \tilde{e}''(s) \, ds \right| \leq \left(\int_\xi^s ds \right)^{1/2} \left(\int_\xi^s \left[\tilde{e}''(s)\right]^2 ds \right)^{1/2}$$
$$\leq |s - \xi|^{1/2} \left(\int_0^s \left[\tilde{e}''(s)\right]^2 ds \right)^{1/2}$$

Squaring and integrating over $s \in [0,1]$:

$$\int_0^1 \left[\tilde{e}'(s)\right]^2 ds \leq \left(\int_0^1 |s - \xi| \, ds \right) \left(\int_0^1 \left[\tilde{e}''(s)\right]^2 ds \right)$$

Noting that:

$$\int_0^1 |s-\xi|\,ds = \xi^2 - \xi + \frac{1}{2} \le \frac{1}{2} \quad \forall \xi \in (0,1)$$

We obtain:

$$\int_0^1 \left[\tilde{e}'(s)\right]^2 ds \le \frac{1}{2}\int_0^1 \left[\tilde{e}''(s)\right]^2 ds$$

Allowing us to conclude. □

Corollary 2 ($\mathbb{P}_1$ error estimate for the Ritz-Galerkin approx. $u_h \in V_h^1$)**.** Let $f \in C^0([0,1])$ and $u \in C^2([0,1]) \cap V$. Let u_h be the Ritz-Galerkin approx. of $u \in V$ over V_h^1 i.e.

$$a(u_h, v) = \langle f, v\rangle \quad \forall v \in V_h^1$$

Then:

$$\|u - u_h\|_{L^2} \le \frac{h}{\sqrt{2}}\|u - u_h\|_V \le \frac{h^2}{2}\|f\|_{L^2}$$

Corollary 3 (Approximation property for V_h^k)**.** Approximation property for V_h^k: Let $h = \max_i h_i$, $w \in C^2([0,1]) \cap V$, then:

$$\min_{v \in V_h^k} \|w - v\|_V \le \frac{h}{\sqrt{2}}\|w''\|_{L^2}$$

This follows immediately from $V_h^1 \subset V_h^k$. This is not an optimal error bound and in practice we expect to gain one order (power of h) per polynomial degree. Higher order error bounds will require higher regularity on the weak solution u.[2]

.3.1 Existence and uniqueness

Functional analysis recap: Recall the function space:

$$C^m(\Omega) = \{f \in C^0(\Omega) : D^{\mathbf{k}} f \in C^0(\Omega) \quad \forall |\mathbf{k}| \le m\}$$

where:

$$|\mathbf{k}| = k_1 + k_2 + \ldots + k_d$$
$$D^{\mathbf{k}} = \frac{\partial^{|\mathbf{k}|}}{\partial x_1^{k_1} \partial x_2^{k_2} \ldots \partial x_d^{k_d}}$$

Now define:

$$C^\infty(\Omega) = \{f \in C^0(\Omega) : D^{\mathbf{k}} f \in C^0(\Omega) \quad \forall \mathbf{k}\} \qquad \text{(infinitly smooth)}$$
$$C_0^m(\Omega) = \{f \in C^m(\Omega) : \operatorname{supp}(f) \subset\subset \Omega\} \qquad \text{(compact support)}$$

Where the support of $f(x)$ is the closure of the set of points where $f(x) \neq 0$, and $f(x)$ is compactly supported when the support of $f(x)$ is a compact set in $\mathbb{R}^n$.

[2] Password for lecture notes: FiniteElement314159

Example 12
Let $\Omega = \mathbb{R}$, then:

$$f(x) = \begin{cases} 4(x - x^2), & x \in (0,1) \\ 0, & \text{otherwise} \end{cases}$$

is compact in $\mathbb{R}$, as $4(x - x^2)$ takes the values 0 at $x = 0, 1$. The function is $C_0^0(\Omega)$, but as

$$\frac{d}{dx}\left(4(x - x^2)\right)|_{x=1} = -4 \neq 0$$

it fails to be $C_0^1(\Omega)$.

To design $g(x) \in C_0^k(\Omega)$ for $k < \infty$ we can choose $g(x) = f(x)^{k+1}$. Given a normed vector space V, a sequence of functions $\{f_i\}_{i=1}^{\infty} \subset V$ or $f_i \in V$ is said to converge to a limit if:

$$\lim_{i \to \infty} \|f_i - f\|_V = 0$$

The sequence $\{f_i\}_{i=1}^{\infty}$ is a Cauchy sequence if:

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \text{ s.t. } \|f_i - f_j\|_V < \varepsilon \quad \forall i, j > N$$

Definition .3.2. Banach space
Is a complete normed vector space for which every Cauchy sequence $\{f_i\}_{i=1}^{\infty}\}$ has a limit $f \in V$.

Definition .3.3. L^p spaces
For $1 \leq p < \infty$ we define:

$$L^p(\Omega) = \{u : u \text{ is real and measurable }, \int |u(x)|^p \, dx < \infty\}$$

$$\|u\|_p = \left(\int_\Omega |u(x)|^p \, dx\right)^{\frac{1}{p}} \qquad \text{(norm)}$$

For $p = \infty$ we define:

$$L^\infty(\Omega) = \{u : |u(x)| < \infty \text{ a.e. in } \Omega\}$$
$$\|u\|_\infty = \inf\{C : |u(x)| \leq C \text{ a.e. in } \Omega\}$$

Inequalities in L^p spaces:

- **Hölder's inequality:** For $1 \leq p, q \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$ we have:

$$\|uv\|_1 \leq \|u\|_p \|v\|_q$$

- **Minkowski's inequality:** For $1 \leq\leq \infty$ we have:

$$\|u + v\|_p \leq \|u\|_p + \|v\|_p$$

Hilbert spaces is an inner product space which is complete w.r.t. the norm induced by the inner product $\langle \cdot, \cdot \rangle^{\frac{1}{2}} = \| \cdot \|$.

Theorem .3.4. Hilbert projection

Let $\mathcal{M} \subset V$ be a closed subspace of a Hilbert space V. Then:

$$\forall u \in V \, \exists! v \in \mathcal{M} \text{ and } w \in \mathcal{M}^\perp \text{ s.t. } u = v + w$$

where

$$\mathcal{M}^\perp = \{w \in V : \langle v, w \rangle_V = 0 \quad \forall v \in \mathcal{M}\}$$
$$V = \mathcal{M} \oplus \mathcal{M}^\perp$$

We define the space of locally integrable functions:

$$L^1_{\text{loc}}(\Omega) = \{u(x) : \int_\Omega |u(x)\phi(x)| \, dx < \infty \quad \forall \phi \in C_0^\infty(\Omega)\}$$

Note that:

$$L^1(\Omega) \subset L^1_{\text{loc}}(\Omega)$$

as it enforces compact support for solutions which may blow up at the boundary (e.g. $u(x) = \frac{1}{x}$ on $\Omega = [0, 1]$).

Definition .3.5. Weak derivative

$v \in L^1_{\text{loc}}(\Omega)$ is the weak partial derivative of $u \in L^1_{\text{loc}}(\Omega)$ in the k-th component if:

$$\int_\Omega v\phi \, dx = -\int_\Omega u \frac{\partial \phi}{\partial x_k} \, dx \quad \forall \phi \in C_0^\infty(\Omega)$$

For higher order derivative let $\boldsymbol{\alpha}$ be a multi-index:

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_d) \quad \alpha_i \in \mathbb{N}_0$$

then the $|\boldsymbol{\alpha}|$ order partial derivative v is:

$$\int_\Omega v\phi \, dx = (-1)^{|\boldsymbol{\alpha}|} \int_\Omega u \frac{\partial^{|\boldsymbol{\alpha}|}\phi}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_d^{\alpha_d}} \, dx \quad \forall \phi \in C_0^\infty(\Omega)$$

We write this weak derivative as $v = D^{\boldsymbol{\alpha}} u$.

.4 Lecture 9: 17.09.2025

Example 13. Weak derivatives

Consider the function:

$$f(x) = \begin{cases} 2x, & 0 \le x \le \frac{1}{2} \\ 2 - 2x, & \frac{1}{2} < x \le 1 \end{cases}$$

This function isnt differentiable at $x = \frac{1}{2}$, but weakly differentiable:

$$D_w f(x) = \begin{cases} 2, & 0 \le x < \frac{1}{2} \\ \alpha, & x = \frac{1}{2} \\ -2, & \frac{1}{2} < x \le 1 \end{cases}$$

Note that at $x = \frac{1}{2}$ the pointwise value is *immaterial* (meaning arbitrary) due to the integral in the

definition of weak derivatives **??**.

.4.1 Sobolev spaces

Definition .4.1. Sobolev space

Let $\Omega \subset \mathbb{R}^n$, $k \in \mathbb{N}_0$ and $1 \leq p < \infty$. The (k,p)-Sobolev space is defined as:

$$W^{k,p}(\Omega) = \{u : \Omega \to \mathbb{R} \mid \|u\|_{k,p} < \infty\}$$

where:

$$\|u\|_{k,p}^p = \sum_{|\boldsymbol{\alpha}| \leq k} \|D_w^{\boldsymbol{\alpha}} u\|_{L^p}^p = \sum_{|\boldsymbol{\alpha}| \leq k} \int_\Omega |D_w^{\boldsymbol{\alpha}} u|^p \, dx$$

For $W^{0,2}(\Omega) = L^2(\Omega)$. Typically one chooses $p = 2$, so we introduce the following notation L^2-Sobolev spaces.

Definition .4.2. L^2-Sobolev space

Let $\Omega \subset \mathbb{R}^n$, $k \in \mathbb{Z}^+$, then:

$$H^k(\Omega) = W^{k,2}(\Omega)$$

which is a *Hilbert space*.

Consider the (new) elliptic Model Problem:

$$\begin{aligned} -\nabla \cdot \nabla u(x) + u(x) &= f(x), \quad x \in \Omega \subset \mathbb{R}^n,\ f \in L^2(\Omega) \\ u(x) &= 0, \quad x \in \partial\Omega \end{aligned}$$

The weak formulation is given by seeking $u \in V = H_0^1(\Omega)$ s.t.

$$\int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega uv \, dx = \int_\Omega fv \, dx \quad \forall v \in V$$

Equivalently we have:

$$\begin{aligned} a(u,v) &= \int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega uv \, dx \\ g(v) &= \int_\Omega fv \, dx \end{aligned}$$

.4.2 Existence and uniqueness of weak solutions

To show existence and uniqueness we need to show that the bilinear form $a(\cdot,\cdot)$ is *non-degenerate (coercive)*.

Definition .4.3. Bounded linear functional

A linear functional $g : V \to \mathbb{R}$ is bounded if $\exists C > 0$ s.t.

$$|g(v)| \leq C\|v\|_V \quad \forall v \in V.$$

.4.3 Dual space

Let V be a Hilbert space with norm $\|\cdot\|_V$. The dual space $V^\star$ is the space of bounded linear functionals on V:

$$V^\star := \{g : V \to \mathbb{R} : g \text{ is a bounded linear functional}\}$$

with the operator norm defined by:

$$\|g\|_{V^\star} := \sup_{\substack{u\in V\\ u\neq 0}} \frac{|g(u)|}{\|u\|_V}$$

As $f \in L^2$ for (MP) $g(v) = \int_\Omega fv\,dx$. This assumption can be relaxed as for $g(v)$.

$$|g(v)| = \left|\int_\Omega f(x)v\,dx\right| \quad \forall v \in H_0^1(\Omega)$$

which can be bounded even if $f \notin L^2(\Omega)$. Formally we introduce the (negative) Sobolev space $w \in H^{-1}(\Omega)$:[3]

$$\|w\|_{-1} = \sup_{v\in H^1(\Omega)} \frac{|\langle w, v\rangle_0|}{\|v\|_1} < \infty$$

.4.4 Riesz representation theorem

Theorem .4.4. Riesz representation theorem

Let V be a Hilbert space and $g \in V^\star$. Then $\exists! u \in V$ s.t.

$$g(v) = \langle u, v\rangle_V \quad \forall v \in V.$$

Proof. Consider $g \in V^\star$ and denote the kernel of g by:

$$\ker(g) = \{v \in V : g(v) = 0\}$$

Denote $\mathcal{K} := \ker(g)$ which is a closed linear subspace. Let's split our argument into two cases:

1. If $\mathcal{K} = V$. Here for each $v \in V, g(v) = 0 \implies u = 0$.
2. If $\mathcal{K} \neq V$. Consider $w \in \mathcal{K}^\perp$ where $w \neq 0$. Now let $y = g(w) \neq 0$. We observe that:

$$g(\frac{v}{y}w) = \frac{g(v)}{y}g(w) = g(v) \quad \forall v \in V$$

Through the linearity of $g(\cdot)$. Consequently:

$$\begin{aligned} g(v - \frac{g(v)w}{y}) = g(v) - g(v) = 0 \\ \implies v - \frac{g(v)w}{y} \in \mathcal{K} \\ \implies \langle v - \frac{g(v)w}{y}, w\rangle_V = 0 \quad \forall v \in V \end{aligned}$$

as $w \in \mathcal{K}^\perp$ by Hilbert projection theorem, so we have:

$$\langle \frac{g(v)}{y}w, w\rangle_V = \langle v, w\rangle_V \quad \forall v \in V$$

leading to:

$$g(v) = \frac{\langle v, w\rangle_V}{\langle w, w\rangle_V} y = \langle v, u\rangle_V$$

[3]The negative Sobolev space is the dual space of $H_0^1(\Omega)$, i.e. $H^{-1}(\Omega) = (H_0^1(\Omega))^\star$, similarly $\|v\|_1 = \|v\|_{H^1}$ and $\langle f, v\rangle_0 = \langle f, v\rangle_{L^2}$.

Through defining u as:

$$u = \frac{wy}{\langle w, w\rangle_V} = \frac{wg(w)}{\langle w, w\rangle_V}$$

Now we have found u. We must show that its unique.
Suppose $\exists \hat{u} \in V$ s.t. $g(v) = \langle \hat{u}, v\rangle_V$ for all $v \in V$. Then:

$$g(v) = \langle u, v\rangle = \langle \hat{u}, v\rangle \quad \forall v \in V,$$

showing that:

$$\langle u - \hat{u}, v\rangle_V = 0 \quad \forall v \in V.$$

If $v \in u - \hat{u}$ then:

$$\langle u - \hat{u}, u - \hat{u}\rangle = 0 \implies u - \hat{u} = 0 \implies u = \hat{u}.$$

Thus u is unique. □

Corollary 4. For (MP) the bilinear form is the H^1 inner product:

$$a(u, v) = \langle \nabla u, \nabla v\rangle + \langle u, v\rangle_{L^2} \quad \forall u, v \in H_0^1(\Omega),$$

so our weak formulation is:

$$g(v) = \langle u, v\rangle_{H^1} \quad \forall v \in H_0^1(\Omega).$$

Through seeking $u \in H_0^1(\Omega)$ we satisfy the conditions of the Riesz representation theorem **??** and so there exists a unique weak solution to (MP).

In general, not all weak forms are expressible with inner products on Hilbert spaces.

Consider:

$$-\nabla \cdot \nabla u(x) = f(x), \quad x \in \Omega \subset \mathbb{R}^n u(x) \qquad = 0, \quad x \in \partial\Omega \text{ with } f \in L^2(\Omega)$$

The weak form is given by seeking $u \in V$ s.t.

$$\begin{aligned} a(u, v) &= \int_\Omega \nabla u \cdot \nabla v \, dx \\ g(v) &= \int_\Omega f v \, dx \\ a(u, v) &= g(v) \quad \forall v \in V \end{aligned} \tag{MP2}$$

.4.5 V-ellipticity

Given a Hilbert space V consider a bilinear form $a : V \times V \to \mathbb{R}$. Then:

- $a(\cdot, \cdot)$ is **coercive** if $\exists C_0 > 0$ s.t.

$$C_0 \|u\|_V^2 \le a(u, u) \quad \forall u \in V$$

- $a(\cdot, \cdot)$ is **continuous** if $\exists C_1 > 0$ s.t.

$$|a(u, v)| \le C_1 \|u\|_V \|v\|_V \quad \forall u, v \in V$$

$\implies$ if $a(\cdot, \cdot)$ is **coercive** and **continuous** on V then $a(\cdot, \cdot)$ is called V-**elliptic**.

.4.6 Lax-Milgram (symmetric case)

Theorem .4.5. Lax-Milgram (symmetric case)

Let V be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_V$. Assume $a(\cdot, \cdot)$ is **symmetric** (i.e. $a(u,v) = a(v,u) \forall v, u \in V$) and V-**elliptic**, and that g is a BLF (Bounded Linear Functional) on V. Then $\exists! u \in V$ s.t.

$$a(u,v) = g(v) \quad \forall v \in V$$

Proof.

1. $a(u,v) = a(v,u)$ by symmetry.
2. $a(cu + w, v) = ca(u,v) + a(w,v)$ by bilinearity.
3. $a(u,u) \geq C_0 \|u\|_V^2$ by coercivity.
4. if $u = 0$ (a.e.), then $0 \leq a(u,u) \leq C_1 \|u\|_V^2 = 0$ and $a(u,u) = 0$ by **continuity**.
5. if $a(u,u) = 0$ then $0 \leq C_1 \|u\|_V^2 \leq a(u,u) = 0 \implies u = 0$ (a.e.) by **coercivity**.

Together these show $a(\cdot, \cdot)$ defines an inner prioduct on V and we can conclude by the Riesz representation theorem **??**. □

.5 Lecture 10: 18.09.2025

Note that if $a(\cdot, \cdot)$ is symmetric and V-elliptic it's norm is equivalent to the V-norm as:

$$c_0 \|u\|_V^2 \leq a(u,u) \leq C_1 \|u\|_V^2 \quad \forall u \in V$$

Theorem .5.1. Lax-Milgram

Let V be a Hilbert space, $a(\cdot, \cdot)$ be V-elliptic and $g(\cdot)$ be a BLF on V. Then $\exists! u \in V$ s.t.

$$a(u,v) = g(v) \quad \forall v \in V$$

Proof. For $u \in V$ note that $a_u(v) := a(u,v)$ is a BLF due to the continuity:

$$|a_u(v)| = |a(u,v)| \leq C_1 \|u\|_V \|v\|_V \quad \forall v \in V$$

So by the RRT $\exists! t_u \in V$ s.t.

$$a_u(v) = \langle v, t_u \rangle_V \quad \forall v \in V$$

Note that $a_u(v) = \langle v, t_u \rangle_V$ defines a mapping $T : V \to V$.
For a given u, $t_u = Tu$ is defined by RRT.
Now we show that T is linear and bounded on V:

- T **is linear**: Consider $\alpha u + w \in V$ and $\alpha \in \mathbb{R}$, then for all $v \in V$:

$$\begin{aligned} \langle v, T(\alpha u + w) \rangle_V &= a_{\alpha u + w}(v) = a(\alpha u + w, v) = \alpha a(u,v) + a(w,v) \\ &= \alpha a_u(v) + a_w(v) = \alpha \langle v, Tu \rangle_V + \langle v, Tw \rangle_V \\ &= \langle v, \alpha Tu + Tw \rangle_V \end{aligned}$$

  So by the definiton of inner products $T(\alpha u + w) =$
  $alphaTu + Tw$.
- T **is bounded**: For $u \in V$ we have:

$$\|Tu\|_V^2 = \langle Tu, Tu \rangle_V = a_u(Tu) \leq C_1 \|Tu\|_V \|u\|_V$$

  by continuity and:

$$\|Tu\|_V \leq C_1 \|u\|_V \quad \forall u \in V \quad \implies \|T\| \leq C_1 < \infty$$

  Thus T is bounded.

- range(T) **is closed**: Let $z_n = Tu_n$ be a sequence in range(T), then:

$$a(u_n, v) = \langle v, Tu_n\rangle_V = \langle v, z_n\rangle_V \quad \forall v \in V$$

by choosing $v = u_n - u_m$ we have:

$$a(u_n - u_m, v) = \langle v, z_n - z_m\rangle_V \quad \forall v \in V, \text{ and } C_0\|u_n - u_m\|_V \leq \|z_n - z_m\|_V$$

If: $z_n \to z \in V$ the sequence u_n must be Cauchy. Since V is a Hilbert space, u_n must converge to some $u \in V$. By continuity $a(u_n, v) \to a(u, v)$. Note we have that:

$$|\langle Tu_n - z, v\rangle_V| \to 0$$

so $\langle v, Tu_n\rangle_V \to \langle v, z\rangle_V$. Since $a(u_n, v) = \langle v, Tu_n\rangle_V$ we see that $z = Tu$. Thus since $z_n \to z \in$ range(T), range(T) is closed.
- T **is onto** V**:** Suppose (for contradiction) that $\mathcal{M}$ = range(T) with $\mathcal{M} \neq V$. By the Hilbert projection theorem **??** $\exists w \in \mathcal{M}^\perp$ with $w \neq 0$ and $\langle w, \overline{v}\rangle_V = 0$ for all $\overline{v} \in \mathcal{M}$. Morover:

$$\langle w, Tu\rangle_V = 0 \quad \forall u \in V$$

For $u = w$:

$$0 = \langle w, Tw\rangle_V = a(w, w) \neq 0$$

so $w = 0$ which is a contradiction. Thus range(T) = V.
- **Existence of** u**:** By RRT $\exists! t_g \in V$ s.t.

$$g(v) = \langle v, t_g\rangle_V \quad \forall v \in V$$

Since $T : V \to V$ is onto there must exist $u \in V$ s.t. $t_g = Tu \in V$. Through construction:

$$g(v) = \langle v, Tu\rangle_V = a(u, v) \quad \forall v \in V$$

showing existence of u.
- **Uniqueness of** u**:** Suppose u is not unique, and that $t_g = T\hat{u}$ for some $\hat{u} \in V$. Then:

$$a(u - \hat{u}, v) = 0 \quad \forall v \in V, \text{ and } a(u - \hat{u}, u - \hat{u}) = 0 \implies \hat{u} = u.$$

Thus u is unique.

.5.1 Poincaré inequality (BEVIS KOMMER PÅ EKSAMEN)

To apply Lax-Milgram we need to show continuity and coercivity. To show coercivity we need the following result:

Lemma 5. Poincaré inequality Suppose $\Omega \subset \mathbb{R}^n$ is bounded and $u \in H_0^1(\Omega)$. Then $\exists C > 0$ s.t.

$$\|u\|_0 \leq C\|\nabla u\|_0$$

Proof. We will first assume that $u \in C_0^\infty(\Omega)$ (and we relax/extend this assumption later). For $u \in C_0^\infty(\Omega)$ we have:

$$\begin{aligned}
\nabla \cdot (u^2\mathbf{x}) &= \partial_{x_1}(u^2 x_1) + \partial_{x_2}(u^2 x_2) + \ldots + \partial_{x_n}(u^2 x_n) \\
&= u^2 + 2u\partial_{x_1} u x_1 + u^2 + 2u\partial_{x_2} u x_2 + \ldots + u^2 + 2u\partial_{x_n} u x_n \\
&= nu^2 + 2u\nabla u \cdot \mathbf{x} \\
u^2 &= \frac{1}{n}\nabla \cdot (u^2\mathbf{x}) - \frac{2}{n}u(\nabla u \cdot \mathbf{x})
\end{aligned}$$

Through Hölder's inequality:

$$\begin{aligned}\|u\|_0^2 &= \int_\Omega u^2\,dx = \frac{1}{n}\int_\Omega \nabla\cdot(u^2\mathbf{x})\,dx - \frac{2}{n}\int_\Omega u(\nabla u\cdot\mathbf{x})\,dx \\ &= \underbrace{\frac{1}{n}\int_{\partial\Omega}\mathbf{n}\cdot(u^2\mathbf{x})\,ds}_{=0} - \frac{2}{n}\int_\Omega u(\nabla u\cdot\mathbf{x})\,dx \\ &\overset{\text{Hölder}}{\leq} \frac{2}{n}\max_{\mathbf{x}\in\Omega}|\mathbf{x}|\int_\Omega |u\nabla u|\,dx \\ &\leq \frac{2}{n}\max_{\mathbf{x}\in\Omega}\sqrt{\int_\Omega u^2\,dx}\sqrt{\int_\Omega |\nabla u|^2\,dx}\end{aligned}$$

Giving us:

$$\|u\|_0 \leq \underbrace{\frac{2}{n}\max_{\mathbf{x}\in\Omega}|\mathbf{x}|}_{C}\,\|\nabla u\|_0$$

If $u\in H_0^1(\Omega)$, as $C_0^\infty(\Omega)$ is dense in $H_0^1(\Omega)$ we can define $\{u_k\}$ as a sequence congverging to u for each $u_k\in C_0^\infty(\Omega)$. Then the inequality holds for each k and $\|u_k\|_0\to\|u\|_0$ and $\|\nabla u_k\|_0\to\|\nabla u\|_0$ as $k\to\infty$ completing the proof. □

.5.2 Application of Lax-Milgram to (MP2)

For (MP2) we need that:

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

is V-elliptic (we have already shown that $g(v)$ is a BLF).

For **continuity** we have:

$$\begin{aligned}|a(u,v)| &= \left|\int_\Omega \nabla u\cdot\nabla v\,dx\right| \\ &\leq \sqrt{\int_\Omega |\nabla u|^2\,dx}\sqrt{\int_\Omega |\nabla v|^2\,dx} \\ &\leq \sqrt{\int_\Omega u^2\,dx + \int_\Omega |\nabla u|^2\,dx}\sqrt{\int_\Omega v^2\,dx + \int_\Omega |\nabla v|^2\,dx} \\ &= \|u\|_1\|v\|_1\end{aligned}$$

where $C_1 = 1$ in this case.

For **coercivity** we use the Poincaré inequality 5 $C^2\|\nabla u\|_0^2 > \|u\|_0^2$, and adding $\|\nabla u\|_0^2$ to both sides gives:

$$a(u,u) = \|\nabla u\|_0^2 \geq \frac{1}{C^2+1}\left(\|u\|_0^2 + \|\nabla u\|_0^2\right) = \frac{\|u\|_1^2}{C^2+1}$$

We can employ *Lax-Milgram* **??** and conclude there's a unique solution to (MP2).

We have existence/uniqueness, now we want to know how accurate a discrete approximation $u_h\in V_h\subset V$ is, which leads to *Céa's lemma*.

.5.3 Céa's lemma Vol. 2 (Approximation error)

Theorem .5.2. Céa's lemma

Let $a(\cdot,\cdot)$ be V-elliptic and $g(\cdot)$ be a BLF on V. Let $u \in V$ satisfy:

$$a(u,v) = g(v) \quad \forall v \in V.$$

Consider the finite dimensional subspace $V_h \subset V$, and let $u_h \in V_h$ s.t.

$$a(u_h, v_h) = g(v_h) \quad \forall v_h \in V_h.$$

Then:

$$\|u - u_h\|_V \leq \frac{C_1}{C_0} \min_{v_h \in V_h} \|u - v_h\|_V$$

Proof. Since $V_h \subset V$ we have:

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h$$

So we have *Galerkin orthogonality*:

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h$$

For any $v_h \in V_h$ we have:

$$\begin{aligned} C_0\|u - u_h\|_V^2 \leq a(u - u_h, u - u_h) &= a(u - u_h, u - v_h) + \overbrace{a(u - u_h, v_h - u_h)}^{=0, \text{ since } v_h - u_h \in V_h} && \text{(coercivity)} \\ &\leq C_1\|u - u_h\|_V\|u - v_h\|_V && \text{(continuity)} \end{aligned}$$

Thus:

$$\|u - u_h\|_V \leq \frac{C_1}{C_0}\|u - v_h\|_V \quad \forall v_h \in V_h$$

Allowing us to conclude. □