Institut für Mathematik
Universität Zürich

computational mathematics

Composite Finite Element Methods
Part 1: Finite Element Methods

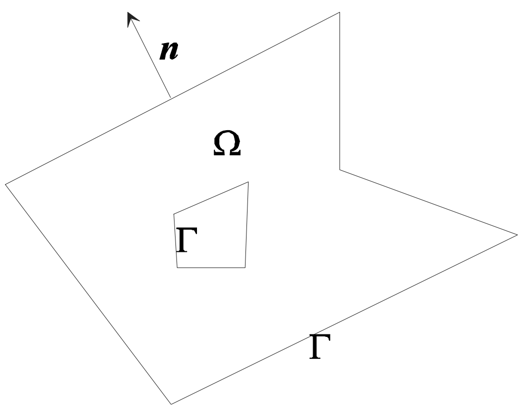
Stefan A. Sauter

Institut für Mathematik, Universität Zürich

June 2023

Model Problem:

Let $\Omega \subset \mathbb{R}^d$ be a bounded open interval ($d = 1$), a bounded polygonal domain ($d = 2$), a bounded polyhedral domain ($d = 3$), etc., with boundary $\Gamma := \partial\Omega$.



Polygonal domain Ω
with boundary $\Gamma := \partial\Omega$
and outward normal $\mathbf{n}$.

Notation

$$\nabla v := \left(\partial_j v\right)_{j=1}^d, \quad \Delta v := \sum_{j=1}^d \partial_j^2 v, \quad \operatorname{div} \mathbf{w} := \sum_{j=1}^d \partial_j w_j.$$

Poisson problem (classical formulation):

Let a "load function" $f : \Omega \longrightarrow \mathbb{R}$ be given. Find a function $u : \Omega \longrightarrow \mathbb{R}$ such that

$$\begin{aligned} -\Delta u(\mathbf{x}) &= f(\mathbf{x}) \quad && \forall \mathbf{x} \in \Omega, \\ u(\mathbf{x}) &= 0 && \forall \mathbf{x} \in \Gamma. \end{aligned}$$

Remark. This formulation requires the equality in every point $\mathbf{x} \in \Omega$. For existence, and uniqueness results this formulation has severe drawback.

Variational formulation of the Poisson model problem (I)

For the *weak* or *variational formulation* one replaces the pointwise conditions by *integral conditions*. Let $v \in C^\infty\left(\overline{\Omega}\right)$ be a *test function*. Multiplying the differential equation by v and integrating over the domain Ω leads to:
find $u : \Omega \to \mathbb{R}$ with $u|_\Gamma = 0$ such that

$$\int_\Omega (-\Delta u)\, v = \int_\Omega f v \quad \forall v \in C^\infty\left(\overline{\Omega}\right).$$

We integrate by parts to obtain

$$\int_\Omega \langle \nabla u, \nabla v \rangle - \int_\Gamma \frac{\partial u}{\partial \mathbf{n}} v = \int_\Omega f v \quad \forall v \in C^\infty\left(\overline{\Omega}\right).$$

Variational formulation of the Poisson model problem (II)

Since the function u is zero on Γ we may restrict to test functions v which are zero on the boundary. Hence, the boundary integral can be dropped and we have derived the weak formulation:

Find $u \in V_0$ such that

$$\int_\Omega \langle \nabla u, \nabla v \rangle = \int_\Omega f v \qquad \forall v \in C_0^\infty(\Omega)$$

with $C_0^\infty(\Omega) := \{v \in C^\infty(\Omega) : v|_\Gamma = 0\}$.

Variational formulation of the Poisson model problem (III)

The *energy space* V_0 must satisfy:
a) V_0 is a Hilbert space,
b) $a(u, v) := \int_\Omega \langle \nabla u, \nabla v \rangle$ defines a scalar product in V_0,
c) functions in V_0 are zero on Γ.

The subspace of $L^2(\Omega)$ with derivatives in $L^2(\Omega)$ is the *Sobolev space*

$$H^1(\Omega) := \left\{ u \in L^2(\Omega) \mid \int_\Omega \|\nabla u\|^2 < \infty \right\}.$$

Trace theorem: Functions in $H^1(\Omega)$ have well-defined restrictions to the boundary $\partial\Omega$.

Definition (energy space) The *energy space* for the Poisson model problem is

$$V_0 := \left\{ v \in H^1(\Omega) \mid v|_\Gamma = 0 \right\}$$

The **weak (variational) formulation** of the Poisson problem reads: find $u \in V_0$ such that

$$a(u, v) = F(v) \quad \forall v \in V_0.$$

Here, the bilinear form $a : V_0 \times V_0 \longrightarrow \mathbb{R}$ and the functional $F : V_0 \longrightarrow \mathbb{R}$ is given by

$$a(u, v) := \int_\Omega \langle \nabla u, \nabla v \rangle \quad \text{and} \quad F(v) := \int_\Omega f v \quad \forall v \in V_0.$$

Analysis on the continuous problem

Definition. Let V be a Hilbert space with norm $\|\cdot\|_V$ and let a bilinearform $a : V \times V \longrightarrow \mathbb{R}$ be given.

$a(\cdot, \cdot)$ is *continuous* if there is some $C > 0$ such that

$$|a(v, w)| \leq C \|v\|_V \|w\|_V \quad \forall v, w \in V.$$

$a(\cdot, \cdot)$ is *coercive* if there is some $c > 0$ such that

$$|a(v, v)| \geq c \|v\|_V^2 \quad \forall v \in V.$$

Definition (cont'd)

$a(\cdot,\cdot)$ is symmetric if

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

A linear form $F \in V'$ is *continuous* if

$$\|F\|_{V'} := \sup_{v \in V \setminus \{0\}} \frac{|F(v)|}{\|v\|_V} < \infty.$$

Theorem (Lax-Milgram). Let V be a Hilbert space and $a : V \times V \longrightarrow \mathbb{R}$ be symmetric, continuous, and coercive. Then, the variational problem: for given continuous linear form $F \in V'$, find $u \in V$ such that

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

has a unique solution which satisfies

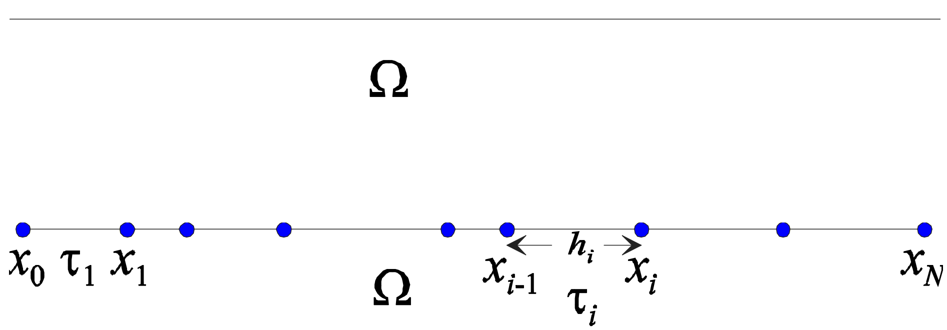
$$\|u\|_V \leq \frac{1}{c} \|F\|_{V'}.$$

Galerkin Finite Element Method:

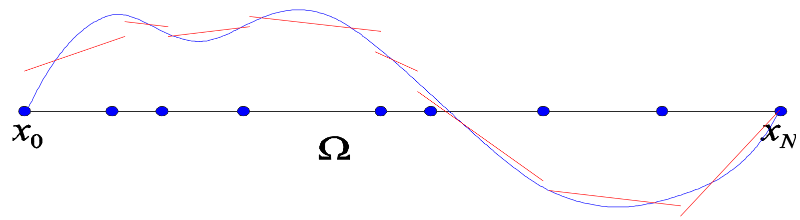
To approximate the continuous problem, a *finite-dimensional* function space $S \subset V_0$ has to be defined.

Idea of finite elements:

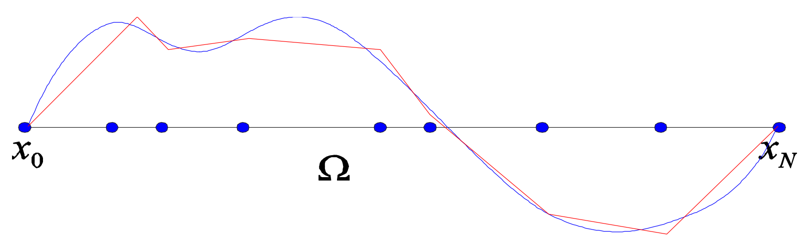
a) subdivide Ω in small *simplices* (intervals, triangles, tetrahedrons)



b) approximate on the simplices by piecewise *polynomials*



c) Enforce continuity across element boundaries and boundary conditions to ensure $S \subset V_0$

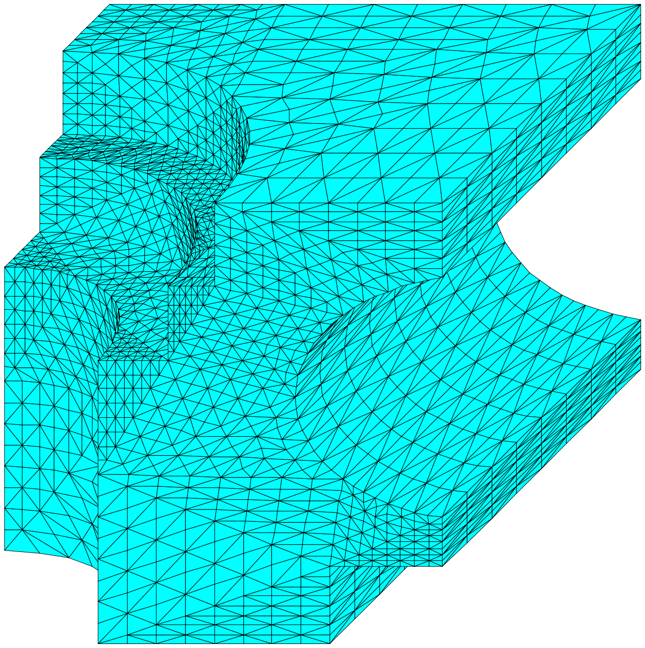


Definition (shape regularity, mesh width). Let $\mathcal{T} := \left\{K_j : 1 \leq j \leq N\right\}$ denote a conforming (no hanging nodes), simplicial finite element mesh for Ω. The (local) mesh width is given by

$$h_K := \operatorname{diam} K \quad \text{and} \quad h := \max\{h_K : K \in \mathcal{T}\}.$$

For the approximation quality, the *shape regularity constant* is important

$$\gamma_{\mathrm{sr}}(\mathcal{T}) := \max\left\{\frac{h_K^d}{|K|} : K \in \mathcal{T}\right\}.$$



Example of a tetrahedral mesh with good shape regularity constant.

Definition (Finite Element space). Let $\mathcal{T} := \left\{K_j : 1 \leq j \leq N\right\}$ denote a conforming, simplicial finite element mesh for Ω and $p \geq 1$. Then

$$S_{\mathcal{T}}^{p} := \left\{u \in C^{0}(\Omega) \mid \forall K \in \mathcal{T} : \quad u|_{K} \in \mathbb{P}_p(K)\right\}$$
$$S := S_{\mathcal{T},0}^{p} := \left\{u \in S_{\mathcal{T}}^{p} \mid u|_{\partial\Omega} = 0\right\}.$$

Definition (Galerkin method). The *Galerkin discretization* of a variational problem is characterized by a finite-dimensional subspace $S \subset V_0$, $N := \dim S < \infty$:

Find $u_S \in S$ such that

$$a(u_S, v) = F(v) \quad \forall v \in S.$$

Stability and convergence analysis

Theorem (Céa). Let V be a Hilbert space and $a : V \times V \rightarrow \mathbb{R}$ be symmetric, continuous, and coercive. Let $S \subset V$ with $\dim S < \infty$. Then, the Galerkin method has a *unique solution* which satisfies the quasi-optimal error estimate

$$\|u - u_S\|_V \leq \frac{C}{c} \inf_{v \in S} \|u - v\|_V \, .$$

The *Galerkin orthogonality* holds

$$a\,(u - u_S, v) = 0 \quad \forall v \in S.$$

Theorem. If the exact solution of the Poisson model problem is *regular*, i.e., $u \in H_0^1(\Omega) \cap H^{p+1}(\Omega)$ then the *energy error* satisfies

$$\|u - u_S\|_{H^1(\Omega)} \leq \frac{C}{c} C_{\mathsf{sr}} h_{\mathcal{T}}^p \|u\|_{H^{p+1}(\Omega)}.$$

Computational aspects:

For the numerical solution, a basis for S is needed

$$S = \text{span}\,\{B_i : i \in \mathcal{I}\} \quad \text{with} \quad |\mathcal{I}| = \dim S = N.$$

Basis representation of Galerkin discretization

The *stiffness (system) matrix* $\mathbf{A} = \left(a_{i,j}\right)_{i,j=1}^{N} \in \mathbb{R}^{N\times N}$ and the *load vector (right-hand side)* $\mathbf{r} := (r_i)_{i=1}^{N}$ are given by

$$a_{i,j} := a\left(B_j, B_i\right) = \int_\Omega \left\langle \nabla B_j, \nabla B_i \right\rangle,$$
$$r_i = F\left(B_i\right) = \int_\Omega f B_i.$$

The Galerkin solution u_S has a unique basis representation

$$u_S = \sum_{i \in \mathcal{I}} u_i B_i$$

and the coefficient vector $\mathbf{u} = (u_i)_{i \in \mathcal{I}}$ is the unique solution of the system of linear equations

$$\mathbf{Au} = \mathbf{r}$$

Remark. The coercivity and symmetry of $a(\cdot, \cdot)$ implies that the matrix $\mathbf{A}$ is *symmetric, positive definite (spd)*.

Affine equivalence

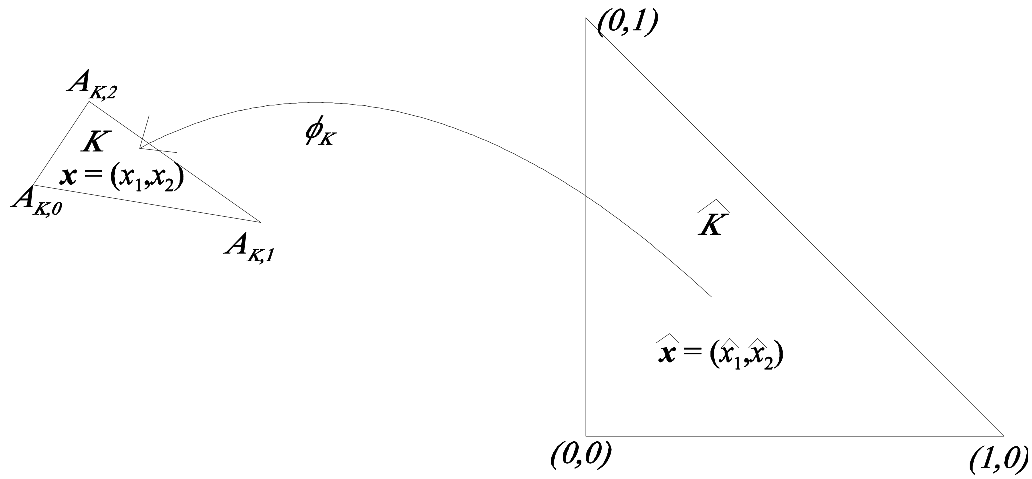
All finite element computations should be transformed to the *affine equivalent* reference element:

$$\widehat{K} := \left\{ \hat{\mathbf{x}} = (\hat{x}_i)_{i=1}^d \in \mathbb{R}^d_{\geq 0} \mid \sum_{i=1}^d \hat{x}_i \leq 1 \right\}.$$

Then any simplex K with vertices $\mathbf{A}_{K,j}$, $0 \le j \le d$, has an affine pullback $\phi_K : \widehat{K} \to K$ given by

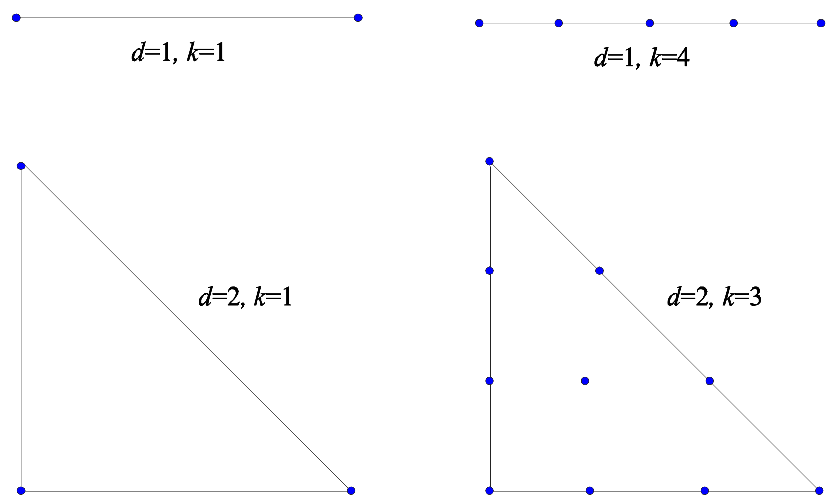
$$\phi_K(\hat{\mathbf{x}}) = \mathbf{A}_{K,0} + \mathbf{m}_K \hat{\mathbf{x}}$$

with the $d \times d$ matrix $\mathbf{m}_K$ having column vectors $\mathbf{A}_{K,j} - \mathbf{A}_{K,0}$.



Basis functions of $\mathbb{P}_p\left(\widehat{K}\right)$ are defined by using nodal points $\widehat{\mathcal{N}}_k$:

$$\widehat{\mathcal{N}}_k := \left\{\frac{i}{p} : 0 \leq i \leq p\right\}^d \cap \widehat{K}$$



For all $\mathbf{z} \in \mathcal{N}_k$ the *Lagrange* basis function $\hat{B}_{\mathbf{z}} \in \mathbb{P}_k\left(\widehat{K}\right)$ is characterized by

$$\hat{B}_{\hat{\mathbf{z}}}\left(\hat{\mathbf{y}}\right) := \begin{cases} 1 & \hat{\mathbf{y}} = \hat{\mathbf{z}}, \\ 0 & \hat{\mathbf{y}} \in \widehat{\mathcal{N}_k} \backslash \left\{\hat{\mathbf{z}}\right\}. \end{cases}$$

$$\left.\begin{aligned} \hat{B}_{(0)}\left(\hat{x}\right) &= 1 - \hat{x} \\ \hat{B}_{(1)}\left(\hat{x}\right) &= \hat{x} \end{aligned}\right\} \quad d = 1, k = 1$$

$$\left.\begin{aligned} \hat{B}_{(0,0)}\left(\hat{\mathbf{x}}\right) &= 1 - \hat{x}_1 - \hat{x}_2 \\ \hat{B}_{(1,0)}\left(\hat{\mathbf{x}}\right) &= \hat{x}_1 \\ \hat{B}_{(0,1)}\left(\hat{\mathbf{x}}\right) &= \hat{x}_2 \end{aligned}\right\} \quad d = 2, k = 1$$

Global nodal points and basis functions are defined via affine pullbacks:

$$\mathcal{N}_k(\mathcal{T}) := \left\{ \phi_K(\hat{\mathbf{z}}) : K \in \mathcal{T}, \hat{\mathbf{z}} \in \widehat{\mathcal{N}_k} \right\},$$
$$\mathcal{N}_{k,0}(\mathcal{T}) := \{\mathbf{z} \in \mathcal{N}_k(\mathcal{T}) \mid \mathbf{z} \notin \partial\Omega\}.$$

For $\mathbf{z} \in \mathcal{N}_k(\mathcal{T})$ the *nodal patch* is given by

$$\mathcal{T}_{\mathbf{z}} := \{K \in \mathcal{T} \mid \mathbf{z} \in K\}$$
$$\omega_{\mathbf{z}} := \bigcup_{K \in \mathcal{T}_{\mathbf{z}}} K.$$

For $\mathbf{z} \in \mathcal{N}_{k,0}$, the associated Lagrange basis is given by

$$B_{\mathbf{z}}|_K := \begin{cases} \hat{B}_{\hat{\mathbf{z}}} \circ \phi_K^{-1} & \text{if } \mathbf{z} \in K \in \mathcal{T}_{\mathbf{z}}, \\ 0 & \text{otherwise,} \end{cases}$$

for $\hat{\mathbf{z}} := \phi_K^{-1}(\mathbf{z})$

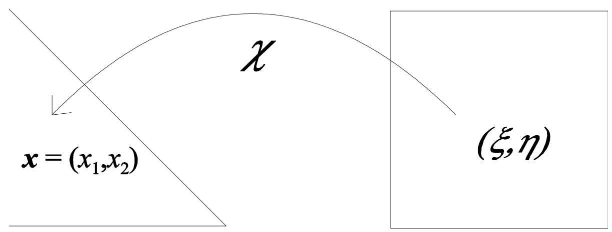
Computation of the right-hand side

$$r_{\mathbf{z}} = \int_\Omega f B_{\mathbf{z}} = \sum_{K \in \mathcal{T}_{\mathbf{z}}} \frac{|K|}{|\widehat{K}|} \int_{\widehat{K}} \widehat{f}_K \widehat{B}_{\hat{\mathbf{z}}}$$

$$\widehat{f}_K := f \circ \phi_K$$

Employ quadrature formula for the evaluation of the integrals.

Duffy transform



$$\chi(\xi,\boldsymbol{\eta}) := \begin{cases} \xi & d=1, \\ \xi\,(1-\eta_1,\eta_1) & d=2, \\ \xi\,(1-\eta_1,\eta_1\,(1-\eta_2)\,,\eta_1\eta_2) & d=3, \end{cases}$$

$$\det\chi'(\xi,\boldsymbol{\eta}) = \begin{cases} 1 & d=1, \\ \xi & d=2, \\ \xi^2\eta_1 & d=3. \end{cases}$$

Hence,

$$\int_{\widehat{K}} \widehat{f}_K \widehat{B}_{\hat{\mathbf{z}}} = \begin{cases} \int_0^1 \widehat{f}_K \widehat{B}_{\hat{\mathbf{z}}} & d = 1, \\ \int_0^1 \int_0^1 \xi \left(\widehat{f}_K \widehat{B}_{\hat{\mathbf{z}}}\right) \circ \chi\left(\xi, \eta_1\right) d\eta_1 d\xi & d = 2, \\ \int_0^1 \int_0^1 \int_0^1 \xi^2 \eta_1 \left(\widehat{f}_K \widehat{B}_{\hat{\mathbf{z}}}\right) \circ \chi\left(\xi, \eta_1, \eta_2\right) d\eta_2 d\eta_1 d\xi & d = 3 \end{cases}$$

and this can be approximated by Gauss quadrature to high accuracy.

Computation of the system matrix

$$a_{\mathbf{x},\mathbf{y}} = \int_\Omega \langle \nabla B_{\mathbf{z}}, \nabla B_{\mathbf{y}} \rangle = \sum_{K \in \mathcal{T}_{\mathbf{z}} \cap \mathcal{T}_{\mathbf{y}}} \frac{|K|}{\left|\widehat{K}\right|} \int_{\widehat{K}} \left\langle \mathbf{g}_K \nabla \widehat{B}_{\hat{\mathbf{z}}}, \nabla \widehat{B}_{\hat{\mathbf{y}}} \right\rangle$$

with $\mathbf{g}_K := \left(\mathbf{m}_K^{-1}\right)^T \mathbf{m}_K^{-1}.$

Sparsity of the system matrix

Theorem. The system matrix is *sparse*: for any $\mathbf{x}, \mathbf{y} \in \mathcal{N}_{k,0}(\mathcal{T})$ it holds

$$\mathcal{T}_{\mathbf{x}} \cap \mathcal{T}_{\mathbf{y}} = \emptyset \Longrightarrow a_{\mathbf{x},\mathbf{y}} = 0.$$

The number of the non-zero entries per matrix row is bounded by C where C only depends on shape regularity and the polynomial degree p.

Corollary. A matrix-vector multiplication has **linear complexity** $O(N)$ while a general direct solver (Gauss elimination, QR decomposition, ...) of the linear system has a cost of $O\left(N^3\right)$. Details $\rightarrow$ *talk of W. Hackbusch*.

Adaptive finite element methods (AFEM)

Idea: Start with a very coarse approximation and set up an algorithm with the structure:

Compute $\longrightarrow$ **Estimate** $\longrightarrow$ **Mark** $\longrightarrow$ **Refine**

In contrast to *a priori* estimates, *AFEM* uses the computed Galerkin solution to enrich/adapt the space.

Error versus residual:

$$\|u - u_S\|_V = \sup_{v \in V \setminus \{0\}} \frac{a(u-u_S,v)}{\|v\|_V} = \|\mathfrak{R}(u_S)\|_{V'}$$

with *residual* $\mathfrak{R}(u_S) : V \to V' \quad \mathfrak{R}(u_S)(v) := a(u - u_S, v)\,.$

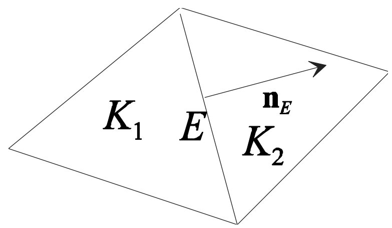
Estimate of residual via local integration by parts (assume $f \in L^2(\Omega)$):

$$\begin{aligned} a(u - u_S, v) &= \int_\Omega \langle \nabla (u - u_S), \nabla v \rangle \\ &= -\int_\Omega \boldsymbol{\Delta}_{\mathcal{T}} (u - u_S)\, v + \sum_{K \in \mathcal{T}} \int_{\partial K} \left(\frac{\partial u}{\partial \mathbf{n}_K} - \frac{\partial u_S}{\partial \mathbf{n}_K} \right) v, \end{aligned}$$

where $\mathbf{n}_K$ is the unit outward normal vector for K and $\boldsymbol{\Delta}_{\mathcal{T}}$ denotes the piecewise gradient

$$\boldsymbol{\Delta}_{\mathcal{T}} v|_{\overset{\circ}{K}} = \boldsymbol{\Delta} \left(v|_{\overset{\circ}{K}} \right) \qquad \forall K \in \mathcal{T}.$$

Let $\mathcal{E}_\Omega$ denote the set of inner element facets (inner mesh points, $d = 1$; inner edges, $d = 2$; inner triangular facets, $d = 3$).



The jump of the normal derivative is given by:

$$\left[\frac{\partial v}{\partial \mathbf{n}_E}\right]_E := \frac{\partial}{\partial \mathbf{n}_E}\left(u|_{K_2}\right) - \frac{\partial}{\partial \mathbf{n}_E}\left(u|_{K_1}\right).$$

Lemma: If $f \in L^2(\Omega)$ the exact solution u of the Poisson model problem satisfies

$$-\Delta_{\mathcal{T}} u = f \quad \text{and} \quad \left[\frac{\partial u}{\partial \mathbf{n}_E}\right]_E = 0.$$

This leads to

$$a(u - u_S, v) = \int_\Omega \text{res}(u_S)\, v + \sum_{E \in \mathcal{E}_\Omega} \int_E \text{Res}(u_S)\, v.$$

with $\quad \text{res}(u_S) := f + \Delta_{\mathcal{T}} u_S \quad \text{Res}(u_S)|_E := \left[\frac{\partial u_S}{\partial \mathbf{n}_E}\right]_E.$

Remark. The *volume residual* $\text{res}(u_S)$ and the *edge residual* $\text{Res}(u_S)$ are computable.

Employ *Galerkin's orthogonaliy*: $a\left(u-u_S, v_S\right)=0$ for all $v_S \in S$:

$$a\left(u-u_S, v-v_S\right) \quad = \int_\Omega \mathsf{res}\left(u_S\right)\left(v-v_S\right) + \sum_{E \in \mathcal{E}_\Omega} \int_E \mathsf{Res}\left(u_S\right)\left(v-v_S\right).$$

Standard trace estimates with *mesh size function* h lead to

$$\begin{aligned} a\left(u-u_S, v\right) \leq \quad & \left(\left\| h\, \mathsf{res}\left(u_S\right)\right\|_{L^2(\Omega)} + C_{\mathsf{trace}} \left\| h^{1/2}\, \mathsf{Res}\left(u_S\right)\right\|_{L^2(\cup \mathcal{E}_\Omega)}\right) \times \\ & \times \left\| h^{-1}\left(v-v_S\right)\right\|_{L^2(\Omega)}. \end{aligned}$$

Quasi-interpolation

Choose v_S as a "quasi-interpolation" of v with an estimate

$$\left\|h^{-1}(v - v_S)\right\|_{L^2(\Omega)} \leq \; C_{\text{int}}\,\|v\|_{H^1(\Omega)}$$

Theorem. Let $f \in L^2(\Omega)$ and let u be the exact solution of the Poisson model problem. Let u_S be the Galerkin solution. Then

$$\|u - u_S\|_{H^1(\Omega)} \leq C_{\text{int}}\left(\|h\,\text{res}\,(u_S)\|_{L^2(\Omega)} + C_{\text{trace}}\left\|h^{1/2}\,\text{Res}\,(u_S)\right\|_{L^2(\cup\mathcal{E}_\Omega)}\right).$$

Local error estimator: Define the local error estimator:

$$\eta_K := \sqrt{\|h\,\mathsf{res}\,(u_S)\|^2_{L^2(K)} + \frac{1}{2}\left\|h^{1/2}\,\mathsf{Res}\,(u_S)\right\|^2_{L^2(\partial K)}}.$$

Then

$$\|u - u_S\|_{H^1(\Omega)} \lesssim \sqrt{\sum_{K \in \mathcal{T}} \eta_K^2}.$$

Marking strategy

Let

$$\eta_{\max} := \max_{K \in \mathcal{T}} \eta_K^2.$$

Choose $\alpha \in\,]0, 1[$, e.g., $\alpha = 0.7$. Then, mark all simplices for refinement with

$$\eta_K \geq \alpha \eta_{\max}.$$

Remark. AFEM is a strategy to enrich a finite element space by *mesh refinement*.

Outlook

Composite Finite Element spaces are used as a strategy to improve the finite element space *without* increasing the number of unknowns. See next talk.

Thank you for your attention