

Chapter 3

Linear Finite Element Methods

The finite element methods provide

- spaces V_n of functions that are piecewise smooth and “simple”, and
- locally supported basis function of these spaces

to achieve

- good approximations of the solution $u \in V$,
- an efficient assembly of the system matrix with desirable properties (*e. g.* sparse and/or well conditioned).

3.1 Meshes

- A **mesh** \mathcal{M} of the computational domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, is a collection of non-overlapping **cells** $\{K_i\}_{i=1}^M$, $M := |\mathcal{M}|$, that are open, simply connected subsets $K_i \subset \Omega$.
- Each cell K_i is a C^∞ -diffeomorphic image of a closed d -dimensional polytope \widehat{K} — the **reference cell** — which is a convex hull of at least $d + 1$ points in \mathbb{R}^d .
- Each cell K_i is an interval ($d = 1$), a Lipschitz polygon ($d = 2$), or a Lipschitz polyhedron ($d = 3$).
- The mesh \mathcal{M} contains a set of **nodes** (vertices) $\mathcal{N}(\mathcal{M})$, **edges** $\mathcal{E}(\mathcal{M})$ ($d > 1$), and **faces** $\mathcal{F}(\mathcal{M})$ ($d = 3$). (Sometimes the components of the mesh of dimension $d - 1$ are called *faces*).
- The topology of the mesh is described by the connectivity of cells, faces, edges and nodes, *e. g.*, does a face belong to a cell, or more precisely, is a face contained in the closure of a cell.
- A mesh is **conforming** if the intersection $\overline{K_i} \cap \overline{K_j}$ of (the closure of) any two adjacent cells K_i , K_j is a face of both cells. A node of a mesh \mathcal{M} that is located in the interior of a geometric face of one of its cells is known as **hanging** (dangling) **node**.
- **Grids** are meshes with **translation invariant** structure. These can be **tensor product grids**, that are meshes whose cells are quadrilaterals ($d = 2$) or hexahedra ($d = 3$) with parallel sides.

Meshes

- are a crucial building block in the design of the finite dimensional trial and test spaces used in the finite element method, and
- provide subdomains for integration to build the system matrix and vector of the right hand side.

The **orientation** of geometric objects of the mesh is important for defining high order basis functions.

- For an edge we have to define its direction.
- For a face ($d = 3$) we have to specify an ordering of the edges along its boundary.

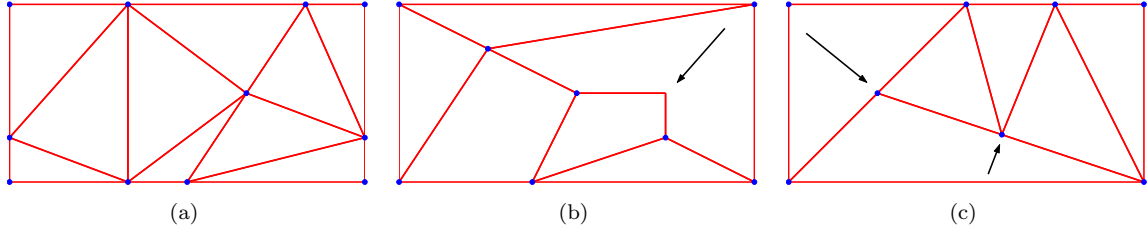


Figure 3.1: (a) Two-dimensional mesh and the sets of edges (red) and nodes (blue). (b) A geometrical node is not a topological node. (c) A triangulation that is not conforming and possesses two hanging nodes.

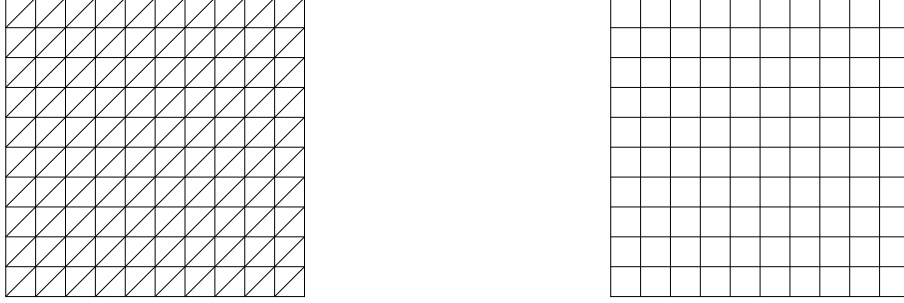


Figure 3.2: Examples of triangular and quadrilateral grids in two dimensions.

Reference cell \hat{K}

- for triangles: reference triangle with nodes $(0, 0)$, $(1, 0)$, $(1, 1)$
- for quadrilaterals: reference quad $[0, 1]^2$ or $[-1, 1]^2$
- for tetrahedra: reference tetrahedron with nodes $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, $(0, 1, 0)$
- for prism: reference prism = reference triangle $\times [0, 1]$
- for pyramid: reference pyramid with nodes $(0, 0, 0)$, $(1, 0, 0)$, $(1, 1, 0)$, $(0, 1, 0)$ and $(0.5, 0.5, 1)$
- for hexahedra: reference hexahedron $[0, 1]^3$

In meshes with different cell types, we have more than one reference cell.

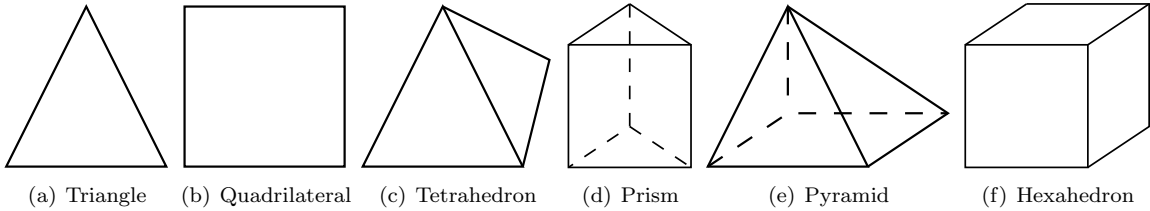


Figure 3.3: Cell types in 2D and 3D.

Affine mapping: linear, bijective mapping from a reference cell (triangle/parallelogram) to a physical cell (triangle/parallelogram) in the mesh

$$\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad \xi \mapsto \mathbf{F}\xi + \boldsymbol{\tau}, \quad \mathbf{F} \in \mathbb{R}^{d,d} \text{ regular}, \quad \boldsymbol{\tau} \in \mathbb{R}^d. \quad (3.1)$$

A mesh is called **affine equivalent** if all its cells arise as affine images of a single d -dimensional (reference) cell.

Automatic mesh generation: algorithms that create a mesh starting from a description of Ω which can be given

- in terms of geometric primitives (ball, brick, etc.) whose unions or intersections constitute Ω ,
- by means of a parametrization of the faces of Ω ,
- through a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, whose sign indicates whether a point is located inside Ω or outside,
- by a mesh covering the surface of Ω and a direction of the exterior unit normal.

Various strategies can be employed for automatic grid generation:

- advancing front methods, that build cells starting from the boundary,
- Delaunay refinement techniques, that can create a mesh starting from a mesh for $\partial\Omega$ or a “cloud” of points covering Ω ,
- the quadtree ($d = 2$) or octree ($d = 3$) approach, which fills Ω with squares/cubes of different sizes supplemented by special measures for resolving the boundary, and
- mapping techniques, that split Ω into subdomains of “simple” shape (curved triangles, parallelograms, bricks), endow those with parametric grids and glue these together.

Remark 3.1. *Traditional codes for the solution of boundary value problems based on the finite element method usually read the geometry from a file describing the topology and geometry of the underlying mesh. Then an approximate solution is computed and written to a file in order to be read by post-processing tools like visualization software, see Figure 3.4.*

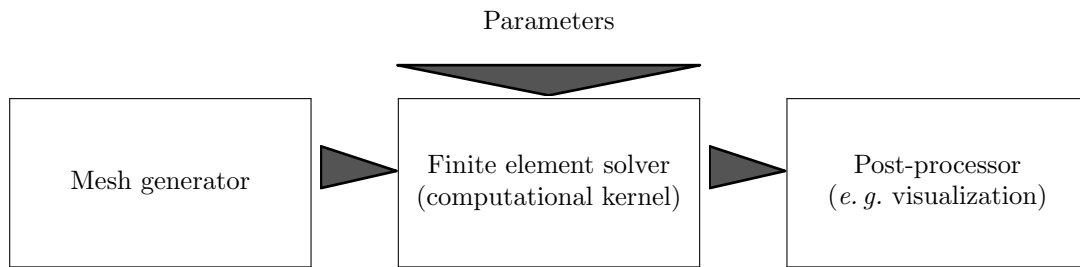


Figure 3.4: Flow of data in traditional finite element simulations

Remark 3.2. *A typical file format for a mesh of a simplicial conforming triangulation of a two-dimensional polygonal domain is the following:*

```

# Two-dimensional simplicial mesh
N ∈ ℕ           # Number of nodes
ξ1 η1         # Coordinates of first node
ξ2 η2         # Coordinates of second node
⋮
ξN ηN         # Coordinates of N-th node
M ∈ ℕ           # Number of triangles
n11 n21 n31 X1  # Indices of nodes of first triangle
n12 n22 n32 X2  # Indices of nodes of second triangle
⋮
n1M n2M n3M XM # Indices of nodes of M-th triangle
  
```

Here, X_i , $i = 1, \dots, M$, is an additional piece of information that may, for instance, describe what kind of material properties prevail in triangle $\#i$. In this case X_i may be an integer index into a look-up table of material properties or the actual value of a coefficient function inside the triangle.

Additional information about edges located on $\partial\Omega$ may be provided in the following form:

$K \in \mathbb{N}$		# Number of edges on $\partial\Omega$
$n_1^1 \ n_2^1$	Y_1	# Indices of endpoints of first edge
$n_1^2 \ n_2^2$	Y_2	# Indices of endpoints of second edge
\vdots		
$n_1^K \ n_2^K$	Y_K	# Indices of endpoints of K -th edge

where Y_k , $k = 1, \dots, K$, provides extra information about the type of boundary condition to be imposed on edge k .

Note that the ordering of the nodes in the above file formats implies an orientation of triangles and edges.

For a comprehensive account on mesh generation see [FG00]. An interesting algorithm for Delaunay meshing is described in [Rup95, She96]. Free mesh generation software is also available, just to name some, netgen, gmsh, triangle, emc2. However, the most sophisticated mesh generation tools are commercial products and their algorithmic details are classified.

3.2 Linear finite elements on triangular meshes

3.2.1 Basis functions

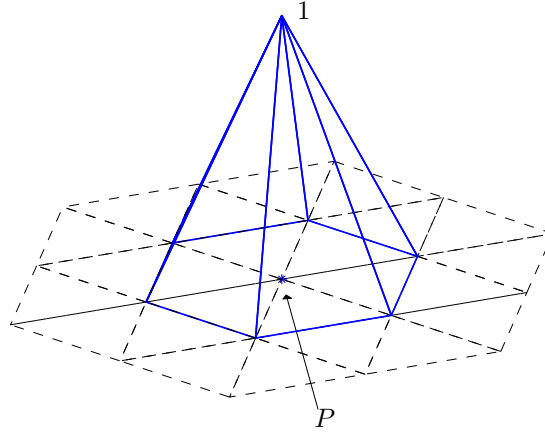


Figure 3.5: Finite element basis function $b_P(\mathbf{x})$.

Let the domain $\Omega \in \mathbb{R}^2$ be a polygon with a triangular conforming mesh \mathcal{M} . We define the finite element space of piecewise linear, continuous functions as

$$S^1(\Omega, \mathcal{M}) := \{u \in C^0(\Omega) : u(\mathbf{x})|_K = a + bx_1 + cx_2, \quad \forall K \in \mathcal{M}\}.$$

Proposition 3.3 (Properties of $S^1(\Omega, \mathcal{M})$).

- It holds $S^1(\Omega, \mathcal{M}) \subset H^1(\Omega)$.
- $u \in S^1(\Omega, \mathcal{M})$ is uniquely defined by the values of $u(P)$ on the nodes $P \in \mathcal{N}(\mathcal{M})$.
- $N = \dim S^1(\Omega, \mathcal{M}) = |\mathcal{N}| < \infty$.
- $S^1(\Omega, \mathcal{M}) = \text{span}\{b_P(\mathbf{x}) : P \in \mathcal{N}(\mathcal{M})\}$, with so called **hat functions** defined as

$$b_P \in S^1(\Omega, \mathcal{M}), \quad b_P(P') = \delta_{P=P'}.$$

Let \mathbf{b} be the vector of the basis functions. Then, an arbitrary FE function $v \in S^1(\Omega, \mathcal{M})$ can be written as

$$v(\mathbf{x}) = \sum_{P \in \mathcal{N}(\mathcal{M})} v(P) b_P(\mathbf{x}) = \mathbf{v}^\top \mathbf{b}(\mathbf{x}),$$

so also the solution

$$u_n(\mathbf{x}) = \sum_{P \in \mathcal{N}(\mathcal{M})} u_n(P) b_P(\mathbf{x}) = \mathbf{u}^\top \mathbf{b}(\mathbf{x}).$$

3.2.2 Assembling of system matrix and load vector

The support of the basis functions $b_P(\mathbf{x})$ consists of only a few triangles, such that the integrals in the (bi-)linear forms reduce to smaller sets and vanish for many pairs of basis functions.

Changing the point of view, we consider the **shape functions**, which are the restrictions of basis functions to one cell $K \in \mathcal{M}$. For $S^1(\Omega, \mathcal{M})$ these are exactly three, each for one node of K .

The shape functions can be defined on a single cell K as

$$N_{K,P_j(K)}(\mathbf{x}) = \hat{N}_j(\Phi_K^{-1}\mathbf{x}),$$

where \hat{N}_j are the **element shape functions** defined by

$$\hat{N}_0(\boldsymbol{\xi}) = 1 - \xi_1 - \xi_2, \quad \hat{N}_j(\boldsymbol{\xi}) = \xi_j, j = 1, 2,$$

on the reference element \hat{K} , which is the triangle with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$. $P_j(K)$ is the j -th node of the triangle K , $j = 0, 1, 2$, with the coordinate \mathbf{p}_j . Then, the (affine) element mapping reads

$$\mathbf{x} = \Phi_K(\boldsymbol{\xi}) = \mathbf{p}_0 + \xi_1(\mathbf{p}_1 - \mathbf{p}_0) + \xi_2(\mathbf{p}_2 - \mathbf{p}_0) = \boldsymbol{\tau} + \mathbf{F}_K \boldsymbol{\xi},$$

with $\mathbf{F}_K = (\mathbf{p}_1 - \mathbf{p}_0, \mathbf{p}_2 - \mathbf{p}_0)$ and $\boldsymbol{\tau} = \mathbf{p}_0$.

As the element mapping is linear, the shape functions are in fact linear as well. The basis functions result from the shape functions by glueing, *i. e.*,

$$b_P(\mathbf{x}) = \begin{cases} N_{K,P_j(K)}(\mathbf{x}) & \mathbf{x} \in K, P \in \mathcal{N}(K), P = P_j(K), \\ 0 & \text{otherwise.} \end{cases}$$

This can be expressed by the connectivity matrices or **T**-matrices

$$\mathbf{T}_K \in \mathbb{R}^{3,N}, \quad (\mathbf{T}_K)_{ij} = \begin{cases} 1, & P_j(K) = P_i(\mathcal{M}), \\ 0, & \text{otherwise.} \end{cases}$$

These **T**-matrices give a relation between the local number of (element) shape functions and global numbering of basis functions. They have the form

$$\mathbf{T}_K = \begin{pmatrix} \cdot & \dots & \cdot & 1 & \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot & \dots & \cdot \\ \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot & 1 & \cdot & \dots & \cdot & \cdot & \dots \\ \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \dots & \cdot & 1 & \cdot & \dots \end{pmatrix}.$$

For the **T**-matrices an extremely sparse format is used. For linear finite elements on triangles we have only to store the three indices P_{i_0} , P_{i_1} and P_{i_2} of the three nodes in the triangle.

Now, we can express a basis function in one cell K as

$$b_{P_i}(\mathbf{x})|_K = \sum_{k=1}^3 (\mathbf{T})_{ki} N_{K,P_k(K)}(\mathbf{x}),$$

and the system matrix is given with

$$\begin{aligned} (\mathbf{B})_{ij} &= \mathbf{b}(b_{P_j}, b_{P_i}) = \sum_K \mathbf{b}_K(b_{P_j}|_K, b_{P_i}|_K) \\ &= \sum_K \mathbf{b}_K\left(\sum_{k=1}^3 (\mathbf{T}_K)_{kj} N_{K,P_k(K)}, \sum_{\ell=1}^3 (\mathbf{T}_K)_{\ell i} N_{K,P_\ell(K)}\right) \\ &= \sum_K \sum_{k=1}^3 \sum_{\ell=1}^3 (\mathbf{T}_K)_{kj} (\mathbf{T}_K)_{\ell i} \mathbf{b}_K(N_{K,P_k(K)}, N_{K,P_\ell(K)}) \\ &= \sum_K (\mathbf{T}_K)_{\cdot, j}^\top \mathbf{B}_K (\mathbf{T}_K)_{\cdot, i}, \end{aligned}$$

i. e., as sum of “weighted” element matrices

$$\mathbf{B} = \sum_K \mathbf{T}_K^\top \mathbf{B}_K \mathbf{T}_K.$$

This means we have to integrate only over the shape functions, and to sum up the contributions over all the cells.

Element stiffness matrix

The bilinear form in the variational problem related to (1.4) consists of two parts, which contribute to the system matrix. The first part

$$\mathbf{a}(u, v) = \int_{\Omega} \langle a \nabla u, \nabla v \rangle \, d\mathbf{x}$$

is constituting the **stiffness matrix** \mathbf{A} .

We transform the derivatives to \hat{K} . By chain rule of differentiation and with

$$\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2} \right)^\top, \quad \hat{\nabla} = \left(\frac{\partial}{\partial \xi_1}, \frac{\partial}{\partial \xi_2} \right)^\top,$$

we have

$$\begin{aligned} \partial_{\xi_1} N_{K,P_j}(\mathbf{x}(\boldsymbol{\xi})) &= \partial_{x_1} N_{K,P_j}(\mathbf{x}) \frac{\partial x_1}{\partial \xi_1} + \partial_{x_2} N_{K,P_j}(\mathbf{x}) \frac{\partial x_2}{\partial \xi_1} \\ \partial_{\xi_2} N_{K,P_j}(\mathbf{x}(\boldsymbol{\xi})) &= \partial_{x_1} N_{K,P_j}(\mathbf{x}) \frac{\partial x_1}{\partial \xi_2} + \partial_{x_2} N_{K,P_j}(\mathbf{x}) \frac{\partial x_2}{\partial \xi_2} \end{aligned}$$

and thus

$$\hat{\nabla} N_{K,P_j}(\mathbf{x}(\boldsymbol{\xi})) = \underbrace{\left(\frac{\partial \Phi_K}{\partial \boldsymbol{\xi}} \right)^\top}_{\mathbf{D} \Phi_K^\top = \mathbf{F}_K^\top} \nabla N_{K,P_j}(\mathbf{x}).$$

Hence, for triangular cells we obtain

$$\nabla N_{K,P_j}(\mathbf{x}) = \mathbf{F}_K^{-\top} \hat{\nabla} N_{K,P_j}(\mathbf{x}(\boldsymbol{\xi})) = \mathbf{F}_K^{-\top} \hat{\nabla} \hat{N}_j(\boldsymbol{\xi}).$$

For a single cell K the entries of the element stiffness matrix \mathbf{A}_K are given by

$$\begin{aligned} \mathbf{a}_K(N_{K,P_j}, N_{K,P_i}) &= \int_K a(\mathbf{x}) (\nabla N_{K,P_j}(\mathbf{x}))^\top \nabla N_{K,P_i}(\mathbf{x}) \, d\mathbf{x} \\ &= \int_{\hat{K}} a(\Phi_K \boldsymbol{\xi}) (\hat{\nabla} \hat{N}_j(\boldsymbol{\xi}))^\top \mathbf{F}_K^{-1} \mathbf{F}_K^{-\top} \hat{\nabla} \hat{N}_i(\boldsymbol{\xi}) |\mathbf{F}_K| \, d\boldsymbol{\xi} \\ &= \int_{\hat{K}} a(\Phi_K \boldsymbol{\xi}) (\hat{\nabla} \hat{N}_j(\boldsymbol{\xi}))^\top \text{adj}(\mathbf{F}_K) \text{adj}(\mathbf{F}_K)^\top \hat{\nabla} \hat{N}_i(\boldsymbol{\xi}) |\mathbf{F}_K|^{-1} \, d\boldsymbol{\xi}, \end{aligned}$$

where we used the relation of the inverse and the adjoint matrix

$$\mathbf{F}_K^{-1} = |\mathbf{F}_K|^{-1} \text{adj}(\mathbf{F}_K).$$

The adjoint matrix of a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Note, that the gradient of the element shape functions are constant vectors

$$\hat{\nabla} \hat{N}_0 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \hat{\nabla} \hat{N}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \hat{\nabla} \hat{N}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

So we can simplify

$$(\mathbf{A}_K)_{ij} = \mathbf{a}_K(N_{K,P_j}, N_{K,P_i}) = \frac{a_K}{2} (\hat{\nabla} \hat{N}_j)^\top \text{adj}(\mathbf{F}_K) \text{adj}(\mathbf{F}_K)^\top \hat{\nabla} \hat{N}_i |\mathbf{F}_K|^{-1}$$

with the average heat conduction $a_K = \frac{1}{|K|} \int_K a(\mathbf{x}) \, d\mathbf{x}$ in K . (Note that $|\mathbf{F}_K| = 2|K|$). The latter simplifies for constant material or may be obtained by numerical quadrature for more general smooth functions.

Due to the special values of $\hat{\nabla} \hat{N}_j$ we can even simplify

$$\mathbf{A}_K = \frac{a_K}{4|K|} \mathbf{D}_K^\top \mathbf{D}_K$$

with a matrix \mathbf{D}_K with coordinate differences

$$\mathbf{D}_K = \begin{pmatrix} y_1 - y_2 & y_2 - y_0 & y_0 - y_1 \\ x_2 - x_1 & x_0 - x_2 & x_1 - x_0 \end{pmatrix}.$$

Element mass matrix

The mass matrix is related to the bilinear form

$$\mathbf{m}(u, v) = \int_{\Omega} c uv \, d\mathbf{x}.$$

The element mass matrix \mathbf{M}_K of the cell K can be computed as

$$\mathbf{m}_K(N_{K,P_j}, N_{K,P_i}) = \int_K c(\mathbf{x}) N_{K,P_j}(\mathbf{x}) N_{K,P_i}(\mathbf{x}) \, d\mathbf{x} = \int_{\hat{K}} c(\Phi_K \boldsymbol{\xi}) \hat{N}_j(\boldsymbol{\xi}) \hat{N}_i(\boldsymbol{\xi}) |F_K| \, d\boldsymbol{\xi}.$$

In case of a constant function c_K in the cell K we can write

$$(\mathbf{M}_K)_{ij} = \begin{cases} \frac{1}{6} c_K |K| & i = j, \\ \frac{1}{12} c_K |K| & i \neq j. \end{cases}$$

Element load vector

The element load vector is related to the linear form

$$\ell_K(v) = \int_K f v \, d\mathbf{x}.$$

For general, smooth function f we use numerical quadrature to evaluate the integrals. The simplest quadrature rule is

$$\int_K f v \, d\mathbf{x} \approx |K| f(\mathbf{x}_K) v(\mathbf{x}_K)$$

where \mathbf{x}_K is the barycenter of the cell K . As this quadrature rule is only exact for linear functions, and the shape functions are already linear, it will provide only reasonable results if f is (almost) constant in K .

3.2.3 Numerical quadrature

Numerical quadrature for triangular cells are defined on the triangle with nodes $(-1, -1)$, $(1, -1)$, $(-1, 1)$, where integrals over \hat{K} can easily transformed to

$$\int_0^1 \int_0^{1-\xi_1} f(\boldsymbol{\xi}) \, d\xi_2 \, d\xi_1 = \frac{1}{4} \int_{-1}^1 \int_{-1}^{-\xi_1} f\left(\frac{\xi_1+1}{2}, \frac{\xi_2+1}{2}\right) \, d\xi_2 \, d\xi_1 \approx \frac{1}{4} \sum_{j=1}^n w_j f\left(\frac{\xi_{1,j}+1}{2}, \frac{\xi_{2,j}+1}{2}\right).$$

There are Gauß quadrature rules for triangles which are exact for polynomials of maximal total degree $1, 2, \dots, 5$ and which use only $1, 3, 4, 6$ and 7 points, respectively. See, *e. g.*, page 141 in Šolín [Šol06].

To get higher accuracies in a systematic matter there is, for example, the **Duffy transformation**, which transforms an integral over the reference triangle \hat{K} by an integral over a reference square and employs a tensor-product quadrature rule. After applying the Duffy transformation, we integrate over

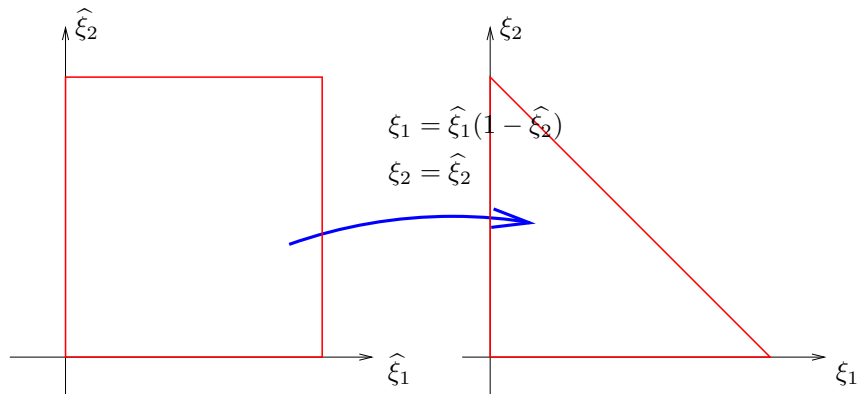


Figure 3.6: Duffy transformation of a square into a triangle.

the reference cell $[0, 1]^2$, so a tensor product of 1D quadrature rule can be applied. The quadrature rules in 1D are, however, usually given in the interval $[-1, 1]$. Hence, we have to transform 1D integrals via

$$\int_0^1 f(\xi) d\xi = \frac{1}{2} \int_{-1}^1 f\left(\frac{\xi+1}{2}\right) d\xi \approx \frac{1}{2} \sum_{j=1}^n w_j f\left(\frac{\xi_j+1}{2}\right),$$

where w_j are the weights and ξ_j the abscissas of the quadrature rule.

Accurate quadrature rule for smooth functions are variants of the Gauß quadrature. The most well-known is the Gauß-Legendre rule for which the abscissas are the zeros of the n -th Legendre polynomial $P_n(\xi)$ and the weights are given by (see [AS64, page 887])

$$w_j = \frac{2}{(1 - \xi_j^2)[P_n'(\xi_j)]^2}.$$

The Gauß-Legendre rule is exact for polynomials of degree $2n - 1$ and the remainder (for the interval $[0, 1]$) is

$$R_n = \frac{(n!)^4}{(2n+1)[(2n)!]^3} f^{(2n)}(\xi), \quad 0 < \xi < 1.$$

The zeros of the Legendre polynomials are tabulated (see [AS64, page 921ff]) and there is an algorithm (see Matlab version `gauleg.m` on the webpage of the lecture) in the numerical recipes [PTFV07].

The **Gauß-Lobatto** rule is only accurate for polynomials of degree $2n - 3$. However, both end-points are included in the set of abscissas.

For the square $[0, 1]^2$ we have the product quadrature rule

$$\int_0^1 \int_0^1 f(\xi) d\xi_1 d\xi_2 = \frac{1}{4} \int_{-1}^1 \int_{-1}^1 f\left(\frac{\xi_1+1}{2}, \frac{\xi_2+1}{2}\right) d\xi_1 d\xi_2 \approx \frac{1}{4} \sum_{i=1}^{n_1} w_i \sum_{j=1}^{n_2} w_j f\left(\frac{\xi_i+1}{2}, \frac{\xi_j+1}{2}\right),$$

with n_1 and n_2 quadrature points in the two directions.

It is important to note that the Duffy transformation changes the total polynomial degree. To see this point, let us start with a definition

Definition 3.4. Given a domain $K \subset \mathbb{R}^d$, $d \in \mathbb{N}$, we write

$$\mathcal{P}_m(K) := \text{span}\{\xi \in K \mapsto \xi^\alpha := \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}, \alpha \in \mathbb{N}_0^d, |\alpha| \leq m\}$$

for the vector space of *d-variate polynomials* of (total) degree m , $m \in \mathbb{N}_0$.

If $\mathbf{m} = (m_1, \dots, m_d)^T \in \mathbb{N}_0^d$ we designate by

$$\mathcal{Q}_{\mathbf{m}}(K) := \text{span}\{\xi \in K \mapsto \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}, 0 \leq \alpha_k \leq m_k, 1 \leq k \leq d\}$$

the space of *tensor product polynomials* of maximal degree m_k in the k -th coordinate direction.

Let $\widehat{K} = \text{convex}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$ and $\widetilde{K} =]0, 1]^2$. Then

$$\int_{\widehat{K}} f(\xi_1, \xi_2) d\xi = \int_{\widetilde{K}} f(\widetilde{\xi}_1(1 - \widetilde{\xi}_2), \widetilde{\xi}_2) (1 - \widetilde{\xi}_2) d\widetilde{\xi}.$$

If $f \in \mathcal{P}_m(\widehat{K})$, then the integrand on the right hand side will belong to $\mathcal{Q}_{m, m+1}(\widetilde{K})$, i. e. the degree of polynomial is increased by one in the second variable.

The usage of numerical quadrature inevitably introduces another approximation, which will contribute to the overall discretization error. The general rule is that

The error due to numerical quadrature must not dominate the total discretization error in the relevant norms.

Remark 3.5. An alternative to numerical quadrature is polynomial interpolation followed by analytical evaluation of the localized integrals.

3.2.4 A-priori estimate of the discretization error

If the variational problem is elliptic (see Céa's lemma 2.28) the discretization error is bounded “quasi optimally” by the best approximation error

$$\|u - u_n\|_V \leq \frac{\|b\|}{\gamma_n} \inf_{w_n \in W_n} \|u - w_n\|_V. \quad (3.2)$$

In case of more general bilinear forms that satisfy (IS1), (IS2) and (DIS) we have

$$\|u - u_n\|_V \leq \left(1 + \frac{\|b\|}{\gamma_n}\right) \inf_{w_n \in W_n} \|u - w_n\|_V.$$

- For any $u \in V$ we define some projection $\mathbf{l}_n : V \rightarrow W_n$, that we can estimate. Trivially, we have

$$\inf_{w_n \in W_n} \|u - w_n\|_V \leq \|u - \mathbf{l}_n u\|_V. \quad (3.3)$$

- For example, for bounded functions u , *e. g.*, for $u \in H^2(\Omega)$ we can project onto the space of linear finite elements by

$$\mathbf{l}_n u = \sum_{P \in \mathcal{N}(\mathcal{M})} u(P) b_P(\mathbf{x}),$$

which is an evaluation of the values on the nodes.

- The interpolation operator of Raviart-Thomas is defined as

$$(\mathbf{l}_n v)(\mathbf{x}) := \sum_{j=1}^N l_j(v) b_j(\mathbf{x})$$

where $l_j : V \rightarrow \mathbb{R}/\mathbb{C}$ are functionals, called *global degrees of freedom*. Number of global degrees of freedom equals the number of basis functions.

Lemma 3.6 (Bramble-Hilbert lemma). *If $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz-domain and $m \in \mathbb{N}$, then*

$$\exists \gamma = \gamma(m, \Omega) > 0 : \inf_{p \in \mathcal{P}_{m-1}(\Omega)} \|v - p\|_{H^m(\Omega)} \leq \gamma |v|_{H^m(\Omega)} \quad \forall v \in H^m(\Omega).$$

- Generalization of Poincaré's inequality (Lemma 2.26) and a best approximation of polynomials in Sobolev spaces which is similar to the pointwise statement of Taylor's theorem.
- The constant γ depends on the domain. Using reference element \hat{K} is beneficial as the constant is then always the same.
- The global degrees of freedom, *e. g.*, point evaluations, edge integrals or cell integrals, define on \hat{K} *local degrees of freedom* and the *local interpolation operator*

$$\hat{I}\hat{u} = (\mathbf{l}_n u)(\mathbf{x}(\boldsymbol{\xi})),$$

which acts on the *pull-back* $\hat{u}(\boldsymbol{\xi}) = u(\Phi_K(\boldsymbol{\xi}))$ of u .

Estimate on the reference element Using the triangle inequality, the fact that there exists a constant $C(t, m, \hat{K})$, $1 \leq t \leq m+1$ if $d = 1$ and $2 \leq t \leq m+1$ if $d = 2, 3$, such that

$$\|\hat{I}\hat{u}\|_{H^t(\hat{K})} \leq C(t, m, \hat{K}) \|\hat{u}\|_{H^t(\hat{K})}$$

for all $\hat{u} \in H^t(\hat{K})$, which follows from the continuous embedding of $H^t(\hat{K})$ in $C^0(\hat{K})$ and, finally, using the Bramble-Hilbert lemma we obtain an estimate for the interpolation error of \hat{I} . Let us measure the error in the H^r -norm with $r \leq t \leq m+1$ ($t \geq 1$ for $d = 1$ and $t \geq 2$ for $d = 2, 3$). Then

$$\begin{aligned} \|\hat{u} - \hat{I}\hat{u}\|_{H^r(\hat{K})} &\leq \|\hat{u} - \hat{I}\hat{u}\|_{H^t(\hat{K})} \\ &= \inf_{p \in \mathcal{P}_m(\hat{K})} \|(\hat{u} - p) - \hat{I}(\hat{u} - p)\|_{H^t(\hat{K})} \\ &\leq \inf_{p \in \mathcal{P}_m(\hat{K})} \left(\|\hat{u} - p\|_{H^t(\hat{K})} + \|\hat{I}(\hat{u} - p)\|_{H^t(\hat{K})} \right) \\ &\leq (1 + C(t, m, \hat{K})) \inf_{p \in \mathcal{P}_m(\hat{K})} \|\hat{u} - p\|_{H^t(\hat{K})} \\ &\leq \underbrace{\gamma(m, \hat{K})(1 + C(t, m, \hat{K}))}_{=C(t, m), \text{ since } \hat{K} \text{ is fixed}} \|\hat{u}\|_{H^t(\hat{K})}. \end{aligned} \quad (3.4)$$

Transformation techniques To apply the Bramble-Hilbert lemma on \widehat{K} and transforming back to K we transform Sobolev norms.

Lemma 3.7. *If $\Phi_K : \widehat{K} \rightarrow K$ is an affine mapping $\xi \mapsto \mathbf{F}_K \xi + \boldsymbol{\tau}$, then, for all $m \in \mathbb{N}_0$,*

$$\begin{aligned} |\widehat{u}|_{H^m(\widehat{K})} &\leq \binom{m+d}{d} d^m \|\mathbf{F}_K\|^m |\det(\mathbf{F}_K)|^{-1/2} |u|_{H^m(K)} \quad \forall u \in H^m(K), \\ |u|_{H^m(K)} &\leq \binom{m+d}{d} d^m \|\mathbf{F}_K^{-1}\|^m |\det(\mathbf{F}_K)|^{1/2} |\widehat{u}|_{H^m(\widehat{K})} \quad \forall u \in H^m(\widehat{K}). \end{aligned}$$

with $\|\mathbf{F}\|_K$ denoting the matrix norm of \mathbf{F}_K associated with the Euclidean vector norm.

Estimate on a cell K With the transformation techniques presented above we can relate the interpolation error on K with that one of the pull-back \widehat{u}

$$\begin{aligned} \|u - \mathbf{I}_n u\|_{H^r(K)}^2 &= \sum_{\ell=0}^r |u - \mathbf{I}_n u|_{H^\ell(K)}^2 \\ &\leq \sum_{\ell=0}^r \binom{\ell+d}{d}^2 d^{2\ell} \|\mathbf{F}_K^{-1}\|^{2\ell} |\det(\mathbf{F}_K)| \left\| \widehat{u} - \underbrace{\mathbf{I}_n u}_{\widehat{I}\widehat{u}} \right\|_{H^\ell(\widehat{K})}^2 \\ &\leq C(r) \binom{r+d}{d}^2 d^{2r} \|\mathbf{F}_K^{-1}\|^{2r} |\det(\mathbf{F}_K)| \left\| \widehat{u} - \widehat{I}\widehat{u} \right\|_{H^r(\widehat{K})}^2. \end{aligned}$$

Using the estimate (3.4) on the reference element we get

$$\|u - \mathbf{I}_n u\|_{H^r(K)} \leq C(r, t, m) \binom{r+d}{d} d^r \|\mathbf{F}_K^{-1}\|^r |\det(\mathbf{F}_K)|^{1/2} |\widehat{u}|_{H^t(\widehat{K})},$$

and transformed back to K

$$\|u - \mathbf{I}_n u\|_{H^r(K)} \leq \underbrace{C(r, t, m) \binom{r+d}{d} \binom{t+d}{d} d^{r+t}}_{C(d, t, r, m)} \|\mathbf{F}_K^{-1}\|^r \|\mathbf{F}_K\|^t |u|_{H^t(K)}, \quad (3.5)$$

with $r \leq t \leq m+1$, $t \geq 2$. The estimate depends on the size and shape of the triangle through \mathbf{F}_K .

Note that convergence in terms of h_K is only expected if the solution has higher regularity, *i. e.*, $u \in H^2(K)$ if we measure in $H^1(K)$.

Definition 3.8. *Given a cell K of a mesh \mathcal{M} we define its **diameter***

$$h_K := \sup\{|\mathbf{x} - \mathbf{y}|, \mathbf{x}, \mathbf{y} \in K\},$$

and the **maximum radius of an inscribed ball**

$$r_K := \sup\{r > 0 : \exists \mathbf{x} \in K : |\mathbf{x} - \mathbf{y}| < r \Rightarrow \mathbf{y} \in K\}.$$

The ratio $\rho_K = h_K/r_K$ is called the **shape regularity measure** of K .

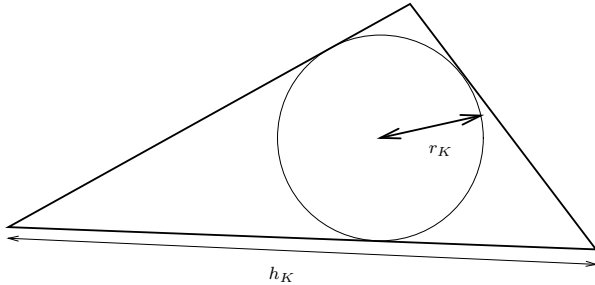


Figure 3.7: Diameter h_K and maximum radius r_K of an inscribed ball for a triangular cell.

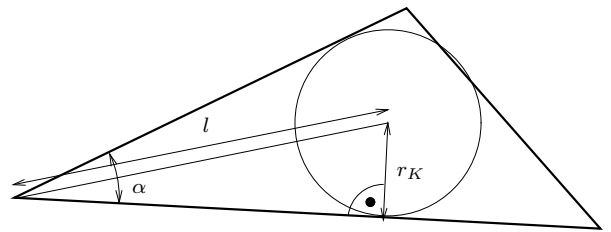


Figure 3.8: Angle condition for shape regularity of a triangle.

- The shape regularity measure of a simplex can be calculated from bounds for the smallest and largest angles enclosed by edge/face normals.
- For triangles with the smallest angle $\alpha > 0$ we have

$$\sin(\alpha/2)^{-1} \leq \rho_K \leq 2 \sin(\alpha/2)^{-1}.$$

Estimates for affine elements maps

$$\frac{\rho_{\mathcal{M}}^{1-d}}{\mu_{\mathcal{M}}^d} h_{\mathcal{M}}^d \leq |\det(\mathbf{F}_K)| \leq h_{\mathcal{M}}^d, \quad \|\mathbf{F}_K\| \leq h_{\mathcal{M}}, \quad \|\mathbf{F}_K^{-1}\| \leq \rho_{\mathcal{M}} \mu_{\mathcal{M}} h_{\mathcal{M}}^{-1}, \quad (3.6)$$

with the **meshwidth**

$$h_{\mathcal{M}} := \max\{h_K, K \in \mathcal{M}\},$$

the **shape regularity measure**

$$\rho_{\mathcal{M}} := \max\{\rho_K, K \in \mathcal{M}\},$$

and the **quasi-uniformity measure**

$$\mu_{\mathcal{M}} := \max\{h_K/h_{K'}, K, K' \in \mathcal{M}\} = \max\{h_K, K \in \mathcal{M}\} \cdot \min\{h_K^{-1}, K \in \mathcal{M}\}.$$

Interpolation error estimate in the mesh The interpolation error can be decomposed into the contributions from all triangles of \mathcal{M} , using (3.5) and (3.6), and summing the errors from each triangle we get

Theorem 3.9. *Let \mathbf{l}_n stand for the finite element interpolation operator belonging to the finite element space $S^m(\mathcal{M})$ on a simplicial mesh \mathcal{M} . Then, for $2 \leq t \leq m+1$, $0 \leq r \leq t$*

$$\exists \gamma = \gamma(t, r, m, \rho_{\mathcal{M}}, \mu_{\mathcal{M}}) : \quad \|u - \mathbf{l}_n u\|_{H^r(\Omega)} \leq \gamma h_{\mathcal{M}}^{t-r} |u|_{H^t(\Omega)} \quad \forall u \in H^t(\Omega).$$

Discretization error in the mesh

From (3.2) and (3.3), *i. e.*, from the fact that

$$\text{Discretization error} \leq \text{best-approximation error} \leq \text{interpolation error}$$

and Theorem 3.9 we can deduce that the discretization error satisfies

$$\|u - u_n\|_{H^1(\Omega)} \leq \gamma h_{\mathcal{M}}^{t-1} |u|_{H^t(\Omega)} \quad \text{for } 2 \leq t \leq m+1 \quad \text{and } u \in H^t(\Omega), \quad (3.7)$$

where $\gamma = \gamma(\Omega, \gamma_n, \|b\|, \rho_{\mathcal{M}}, \mu_{\mathcal{M}}, m, t)$.

Discussion

- To expect convergence we need u to be of higher regularity. For certain problems higher regularity $u \in H^{r+2}(\Omega)$ is guaranteed by **elliptic shift theorems** if $f \in H^r(\Omega)$, and the boundary and the boundary data are smooth.
- The bound (3.7) can be interpreted as **asymptotic a-priori error estimate**

$$\|u - u_n\|_{H^1(\Omega)} = O(h_{\mathcal{M}_n}^{t-1}) \quad \text{for } n \rightarrow \infty,$$

where \mathcal{M}_n , $n \in \mathbb{N}$, is a sequence of simplicial, *uniformly* shape-regular ($\rho_{\mathcal{M}_n} < \rho_{\max}$ for all $n \in \mathbb{N}$) meshes of Ω , that become infinitely fine ($h_{\mathcal{M}_n} \rightarrow 0$ as $n \rightarrow \infty$). In other words, we obtain a convergence with the meshwidth of order $t-1$, which is equal to the polynomial degree if the solution u is sufficiently regular.

- If Ω is convex or has C^1 -boundary, and $a \in C^1(\overline{\Omega})$, then

$$\|u - u_n\|_{L^2(\Omega)} \leq \gamma h_{\mathcal{M}} \|u - u_n\|_{H^1(\Omega)}$$

with $\gamma = \gamma(\Omega, a, m, \rho_{\mathcal{M}}, \mu_{\mathcal{M}}) > 0$. In other words, the weaker the norm of the discretization error the faster it converges to zero as $h_{\mathcal{M}} \rightarrow 0$.

- A more sophisticated best approximation estimates for **increasing polynomial degree**, *i. e.*, $m, t \geq 1$, reads

$$\inf_{v_n \in S^m(\mathcal{M})} \|u - v_n\|_{H^1(\Omega)} \leq \gamma(\rho_{\mathcal{M}}, \mu_{\mathcal{M}}) \left(\frac{h_{\mathcal{M}}}{m} \right)^{\min\{m+1, t\}-1} \|u\|_{H^t(\Omega)} \quad (3.8)$$

which is an a-priori error estimates for the p -version of H^1 -conforming finite elements.