

Chapter 1

Course Overview

We had a stance that our workflow in numerical analysis goes in the following way.

1. We are given either a mathematical definition, or a theorem stating the existence.
 - (a) Example: The definition of *rank* of a matrix.
 - (b) Example: For an equation $u'(t) = f(t, u(t))$, $u(0) = \alpha$ with f smooth, there exists $T_{max} > 0$ and a solution in $[0, T_{max}]$.
2. Not all of them come with *methodology* to capture them.
3. If one comes with a methodology, we specify the algorithm and implement it.
4. If an implementation comes as an *approximation*, we should be able to estimate the size of error.

We will keep the same stance in the course.

Based on what we have learned, we could apply our study to two major fields of mathematics

1. many notions, definitions, and theorems in Linear Algebra,
2. existence theorem of a certain partial differential equation.

1. Each of the two would take one or two semesters.
2. So that students who studies PDE course along with this course can benefit from here,
3. We set our objective to be to implement what are studied in Chapter 6 and 7 in Evans textbook.
4. This means two things:
 - (a) We work with a pde that is either elliptic or parabolic, and the method we employ is the Galerkin method.
 - (b) The important class of pde of hyperbolic type will not be studied. Hyperbolic pde solver has to be implemented differently, where one needs to reflect knowledge of theory of hyperbolic pdes.

The objective of our course is from the following theorem.

We consider the Initial Boundary Value Problem of the following pde:

$$\begin{aligned}
 - \sum_{i,j=1}^n \partial_{x_i} (a^{ij}(x) \partial_{x_j} u(x)) &= \varphi(x), \quad x \in \Omega, \\
 - \sum_{i,j=1}^n (a^{ij}(x) \partial_{x_j} u(x)) \nu_i &= \psi(x), \quad x \in \partial\Omega
 \end{aligned} \tag{P}$$

with constraint $\int_{\partial\Omega} \psi = \int_{\Omega} \varphi$.

We will specify assumptions on Ω , coefficients $a^{ij}(x)$, and (r-h-s) $\varphi(x)$ and $\psi(x)$.

Theorem 1. *There exists a solution of the boundary value problem (P).*

We will make the course as parallel as possible to the previous one.

1. The problem, taken as a root-finding problem for an equation

$$F(x) = y \quad \text{for given } y,$$

is tackled in 2nd half of the course.

2. The first half of the course is to extend our knowledge on piecewise polynomial functions to the multi-dimensional settings.

The first half: pp functions in $\Omega \subset \mathbb{R}^n$.

In this course, unless otherwise specified, $\Omega \subset \mathbb{R}^n$ is a bounded open set with smooth boundary, that is simply connected. Also $n = 3$ in most cases.

This is to extend our far reaching 1d remainder theorem into multi-dimensional setting. Recall

Theorem 2 (1d remainder theorem). *Let $f \in C^{n+1}([a, b])$ and $x_0, x_1, \dots, x_n \in [a, b]$. Then, for $x \in [a, b]$,*

$$\begin{aligned} R(x) &= f[x_0, x_1, x_2, \dots, x_n, x](x - x_0)(x - x_1) \cdots (x - x_{n-1})(x - x_n) \\ &= f(x) - \left(f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \cdots \right. \\ &\quad \left. + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1)(x - x_2) \cdots (x - x_{n-1}) \right). \end{aligned}$$

The extension to multi-dimensional setting is not at all trivial.

Partitioning of Ω

To partition Ω into small pieces is not trivial. We will be speaking of the *Simplicial complex*, tetrahedrons in \mathbb{R}^3 for example, borrowing language of *Combinatorial Algebraic Topology*.

Kinds of functions on Ω

In \mathbb{R}^3 , not only the function $f : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ but also for instance the vector field

$$E : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3,$$

is relevant. We will be speaking of k -covector fields, for $k = 0, 1, 2, 3$, borrowing language of *Differential Geometry*.

Remainder formula or remainder estimates

We present Poincare inequality, Bramble-Hilbert inequality, etc.

The second half: a rootfinding problem framework

We recall the set up for a rootfinding problem for $F : X \rightarrow Y$, solving for a given $y \in Y$ such that

$$F(x) = y.$$

1. We will specify X and Y for the pde.
2. We will recall the design of the solver. This includes
3. Consistency of the parametrized approximation (\tilde{F}_h) .
4. Consistency of the parametrized approximate solver (\tilde{R}_h) .
5. Continuity properties of solver.
6. A priori error estimate.
7. A posteriari error estimate.

Chapter 2

Partitioning of Domain

- In 1-d, the domain $[a, b]$ is partitioned simply by specifying points in ascending order

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b,$$

producing n small intervals

$$I_j = [x_{j-1}, x_j] \quad j = 1, 2, \cdots, n.$$

- The role of the interval in 1-d is taken by the oriented triangle in 2-d, by the oriented tetrahedron in 3-d, and by the oriented n -simplex in n -d.
- We first introduce a few notions to speak of partitioning a domain.

k -cells

- A polyhedral convex set is a finite intersection of closed half spaces of \mathbb{R}^n .
- A nonempty polyhedral convex set that is k -dimensional and compact is called a k -cell.
 1. We make use of the fact that the dimension of any convex set is well-defined.
 2. The k -dimensional area of k -cell τ is thus

$$0 < \mathcal{H}^k(\tau) < \infty.$$

An open set $\Omega \subset \mathbb{R}^n$ equipped with the partition by n -cells

We say (Ω, \mathcal{P}) of an open set $\Omega \subset \mathbb{R}^n$ and a set of n -cells \mathcal{P} is an n -cell partition of Ω if

1. If $K_1, K_2 \in \mathcal{P}$ and $K_1 \neq K_2$ then $\text{int}(K_1) \cap \text{int}(K_2) = \emptyset$.
2. $\bar{\Omega} = \bigcup_{K \in \mathcal{P}} K$.

Examples of n -cell partition

CF. For an open set with smooth curved boundary, we in general consider a partition of Ω by a homeomorphic image of n -cell, not n -cell itself. For simplicity, we assume Ω is just a union of n -cells, omitting the flattening procedure.

1. A partition of Ω by n -cells works fine. But we may want more structured partitioning of Ω .
2. (Ω, \mathcal{P}) may be said to be inconvenient in the following sense.
 - (a) n -cells in \mathcal{P} are not conformal to each other. For example in 3-d, some may be tetrahedrons, some may be cubes, octahedrons, and so on.
 - (b) It is not suitable to define the boundary faces.

Examples of n -cell partition

In our course, we work with n -simplices. We consider (Ω, \mathcal{S}) an open set $\Omega \subset \mathbb{R}^n$ equipped with a *simplicial complex*. We borrow terminology from combinatorial topology, which describes every (topological) detail of Ω in precise manner.

k -simplex

1. As a generalization of an interval in 1-d, we define the oriented k -simplex.
2. (Warning: Let us not bother too much about definitions arising one after another here.)
3. Let $k \in \{0, 1, \dots, n\}$. A k -simplex is a juxtaposition of $k + 1$ points

$$[x_0 x_1 x_2 \cdots x_k],$$

where x_0, x_1, \dots, x_k are affine independent.

4. x_0, x_1, \dots, x_k are affine independent if

$$\lambda_0 + \lambda_1 + \cdots + \lambda_k = 0 \quad \text{and} \quad \lambda_0 x_0 + \lambda_1 x_1 + \cdots + \lambda_k x_k = 0 \quad \implies \lambda_0 = \lambda_1 = \cdots = \lambda_k = 0.$$

This notion is to say that any point of them are not lied in the plane that rest make up.

5. For a $\tau = [x_0 x_1 x_2 \cdots x_k]$, we define its closed point set

$$\overline{\text{pts}}(\tau) = \overline{\text{conv}}\{x_0, x_1, \dots, x_k\}.$$

1d example, 2d example

We can consider a partition whose n -cells are all the closed point sets of n -simplices.

- But we can simply do better by storing all the information by the notion of simplicial complex we soon introduce, and more importantly
- n -simplices partitioning still does not resolve the problem of defining the boundary face.

A simplicial complex S satisfying further assumptions.

We consider a set S of closed point sets of k -simplices for $k \in \{0, 1, 2, \dots, n\}$ that satisfies the following conditions:

1. If $A, B \in S$ then either $A \cap B = \emptyset$ or the intersection is a common face of both A and B .
2. If $A \in S$ then every face of A is also included in S .
3. Every $A \in S$ whose dimension is less than n is a face of some $\sigma \in S$ of dimension n .

We observe from the definition following:

1. Up to item 1, the hanging node problem is resolved.
2. We are able to speak of boundary faces now, which are all stored in S by the condition item 2. A set S satisfying item 1 and 2 are called a simplicial complex.

Example

3. S satisfying further the item 3 is suitable for our purpose.

Example

Eventually, we consider (Ω, \mathcal{S}) where \mathcal{S} is a simplicial complex satisfying additionally the condition item 3, such that

$$\begin{aligned}
 \bar{\Omega} &= |\mathcal{S}| \quad \text{that is} \\
 &= \bigcup_{A \in \mathcal{S}} A \\
 &= \bigcup_{\sigma \in S_n} \sigma, \quad S_k = \{A \in \mathcal{S} \mid \dim(A) = k\} \quad \text{for } k = 0, 1, 2, \dots, n.
 \end{aligned}$$

Inspite of all efforts borrowing terminology from the combinatorial topology, from now on we assume (Ω, \mathcal{S}) is given such that Ω is simply connected and bounded.

Implementing \mathcal{S}

We can store simplices of \mathcal{S} in the following manner.

- We store \mathcal{S}_0 and \mathcal{S}_n legitimately.
- We identify the set \mathcal{S}_0 as the set of coordinates $x = (x^0, x^1, x^2, \dots, x^n)$ and store it.
 1. Let $n_0 =$ number of elements in \mathcal{S}_0 .
 2. Consider enumeration of \mathcal{S}_0 by $i = 1, 2, \dots, n_0$.

- We identify the set \mathcal{S}_n as the $(n + 1)$ -tuples

$$[i_0 i_1 i_2 \dots i_n], \quad i_0, i_1, \dots, i_n \in \{1, 2, \dots, n_0\}.$$

- Now from $k = n - 1$ to $k = 1$ we can store \mathcal{S}_k by the following manner.

$$\mathcal{S}_{k-1} = \{\text{boundary faces of } a \mid a \in \mathcal{S}_k\}.$$

- Along with this, one can store for each $f \in \mathcal{S}_{k-1}$

$$\{a \in \mathcal{S}_k \mid f \text{ is a boundary face of } a.\}$$

Chapter 3

Polynomials on n -simplex

- Started from the preceding chapter, we are in the program of implementing an approximation of a given function v defined in Ω .
- On Ω , one thinks of real-valued functions, vector fields, and so on.
- For a while, we first consider a set of smooth real-valued functions defined on Ω ,

$$\Lambda_0(\Omega) = C^\infty(\bar{\Omega}).$$

- We recall the thumb rules in making approximation from data:
 1. Under the limited number of available (sampling) data, do the piecewise low order polynomial approximation rather than one high order polynomial approximation.
 2. If we go for the piecewise approximation, in one such a small domain, choose the preferable sampling points and the preferable basis whenever possible.
- Following the thumb rule, we did the partitioning of Ω into small nice n -simplices.
- Now we discuss polynomials on a n -simplex.

- We consider an n -simplex $\sigma = [x_0 x_1 x_2 \cdots x_n]$ and its point set

$$M = \overline{\text{pts}}(\sigma).$$

- We first consider real-valued functions defined on M ,

$$v \in \Lambda_0(M) = C^\infty(M).$$

- We consider the subspace of Λ_0 that consists of polynomials of order at most m ,

$$\mathbb{P}_m(M) \subset \Lambda_0(M).$$

- We consider a problem of choosing an element $p \in \mathbb{P}_m(M)$ for an approximation of $v \in \Lambda_0(M)$, out of suitable number of sampling data

$$(x_i, v(x_i)), \quad x_i \in M, \quad i = 1, 2, \dots, d.$$

Polynomials in M and Multi index notation

As an example, let us consider a second order polynomial in \mathbb{R}^2 that is written as

$$p(x, y) = ax^2 + bxy + cy^2 + dx + ey + f, \quad a, b, c, d, e, f \in \mathbb{R}$$

of three quadratic terms, two linear terms, and one constant term.

To study polynomials in \mathbb{R}^n in a systematic way, we introduce the multi index.

Multi Index

We introduce a convenient notation for a polynomial in multi dimensions.

- A multi index α is an n -tuple of nonnegative integers $\alpha_1, \alpha_2, \dots, \alpha_n$,

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in (\mathbb{N} \cup \{0\})^n.$$

- We let

$$x^\alpha = (x_1, x_2, \dots, x_n)^{(\alpha_1, \alpha_2, \dots, \alpha_n)} = x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} \cdots x_n^{\alpha_n} \in \mathbb{R}.$$

- The degree or the order of α is

$$|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n \in \mathbb{N} \cup \{0\}.$$

- A homogeneous r -th order polynomial is thus a linear combination of

$$\{x^\alpha \mid |\alpha| = r\}.$$

- For a fixed $r \in \mathbb{N} \cup \{0\}$, how many distinct multi indices with degree r are there ?

This is to choose r numbers out of $\{1, 2, \dots, n\}$ with repetition allowed,

$$d_{n,r} = \binom{n+r-1}{r}.$$

- An element $p \in \mathbb{P}_m(M)$ of polynomials of order at most m is thus written as

$$p(x) = \sum_{0 \leq |\alpha| \leq m} c_\alpha x^\alpha, \quad c_\alpha \in \mathbb{R} \text{ is a coefficient.}$$

- We note that

$$\mathbb{P}_m(M) \simeq \mathbb{R}^d, \quad d = \sum_{r=0}^m d_{n,r}.$$

In fact, d is to choose m numbers for the exponents of $\{1, x_1, x_2, \dots, x_n\}$ of $n+1$ elements with repetition allowed, and thus

$$d = \sum_{r=0}^m d_{n,r} = \binom{n+m}{m}.$$

- For later purposes, we also define the factorial

$$\alpha! = \alpha_1! \alpha_2! \cdots \alpha_n! \in \mathbb{R}, \quad \text{having in mind that } 0! = 1.$$

Examples: $n = 2$.

- $\mathbb{P}_0(M)$ is of dimension 1,

$$d_{2,0} = 1.$$

- $\mathbb{P}_1(M)$ is of dimensions

$$1 + d_{2,1} = 1 + 2 = 3.$$

- $\mathbb{P}_2(M)$ is of dimensions

$$1 + 2 + d_{2,2} = 1 + 2 + 3 = 6.$$

Examples: $n = 3$.

- $\mathbb{P}_0(M)$ is of dimension 1,

$$d_{3,0} = 1.$$

- $\mathbb{P}_1(M)$ is of dimensions

$$1 + d_{3,1} = 1 + 3 = 4.$$

- $\mathbb{P}_2(M)$ is of dimensions

$$1 + 3 + d_{3,2} = 1 + 3 + 6 = 10.$$

Hence, for $M \subset \mathbb{R}^2$, to fix an element in $\mathbb{P}_0(M)$, $\mathbb{P}_1(M)$, and $\mathbb{P}_2(M)$, we need to provide sampling data $(x_i, v(x_i))$ respectively as many as 1, 3, and 6.

Hence, for $M \subset \mathbb{R}^3$, to fix an element in $\mathbb{P}_0(M)$, $\mathbb{P}_1(M)$, and $\mathbb{P}_2(M)$, we need to provide sampling data $(x_i, v(x_i))$ respectively as many as 1, 4, and 10.

Choosing a Basis of $\mathbb{P}_m(M)$

- We do not use *power basis*

$$\{x^\alpha \mid 0 \leq |\alpha| \leq m.\}$$

- We use *Lagrange basis* (nodal basis): for each $i = 1, 2, \dots, d$, we choose sampling points $(x_i)_{i=1}^d$ and basis functions $\theta_i \in \mathbb{P}_m(M)$ so that

$$\theta_i(x_j) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

1. This is because, of course, if we are given sampling data $(x_i, v(x_i))$, $i = 1, 2, \dots, d$, we immediately pick up an element in $\mathbb{P}_m(M)$ that is

$$x \mapsto \sum_{i=1}^d v(x_i) \theta_i(x),$$

compatible with the sampling data.

2. Then it matters that how we choose the sampling points x_i , $i = 1, 2, \dots, d$. In our course, we will not be bothered too much on the choice of preferable sampling points unlike in 1-d.

Examples of sampling points in triangle and in tetrahedron

$$\mathbb{P}_0(M), \quad \mathbb{P}_1(M), \quad \mathbb{P}_2(M).$$

- Such nodal basis, as well as other basis in many cases, are better expressed in the *barycentric coordinate system* rather than the given \mathbb{R}^n -coordinate system. We specify the barycentric coordinate system for a given k -simplex now.

Barycentric coordinate system for k -simplex

Example: line passing x_0 and x_1

$$\ell : \{(1 - \lambda)x_0 + \lambda x_1 \mid \lambda \in \mathbb{R}\}.$$

For a given k -simplex $\tau = [x_0 x_1 x_2 \cdots x_k]$, there is the unique k -dimensional plane where τ is lied. It is a set of points expressed by a combination

$$P(\tau) = \{\lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k \in \mathbb{R}^n \mid \lambda_i \in \mathbb{R}, \quad \lambda_0 + \lambda_1 + \lambda_2 + \cdots + \lambda_k = 1.\}$$

Now,

1. Consider a hyperplane \hat{P}_k in \mathbb{R}^{k+1} constrained by one equation:

$$\hat{P}_k = \{\lambda \in \mathbb{R}^{k+1} \mid \lambda_0 + \lambda_1 + \lambda_2 + \cdots + \lambda_k = 1\} \subset \mathbb{R}^{k+1}$$

2. The barycentric coordinate system is a parametrization from \hat{P}_k to $P(\tau)$:

$$\chi : \hat{P}_k \rightarrow P(\tau), \quad \lambda \mapsto \lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k.$$

3. In particular, the parametrization of $\overline{\mathbf{pts}}(\sigma)$ is from the set

$$\chi^{-1}(\overline{\mathbf{pts}}(\sigma)) = \{(\lambda_0, \lambda_1, \cdots, \lambda_k) \in \hat{P}_k \mid \forall i \quad \lambda_i \geq 0\} =: L_k \subset \hat{P}_k.$$

Before we specify basis functions in barycentric coordinate system, we get familiar with them by a few observations:

We let $n = 3$ and consider a tetrahedron $\sigma = [x_0x_1x_2x_3]$.

1. Faces parametrized by $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3) \in L_3$.

2. Level sets of λ_e , $e = 0, 1, 2, 3$.

3. The point $x_c = \frac{1}{4}(x_0 + x_1 + x_2 + x_3)$ corresponds to $\lambda = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. We will show that x_c is the center of mass.

The map from x to λ .

1. We should be able to obtain for a given $x \in \overline{\text{pts}}(\tau)$ the $(\lambda_0, \lambda_1, \dots, \lambda_k)$.
2. This can be simply done as below, which is numerically not preferable.

For a given $(\lambda_0, \lambda_1, \dots, \lambda_k) \in L_k$,

$$\begin{aligned} x(\lambda) &= \lambda_0 x_0 + \lambda_1 x_1 + \dots + \lambda_k x_k \\ &= x_0 + \lambda_1(x_1 - x_0) + \lambda_2(x_2 - x_0) + \dots + \lambda_k(x_k - x_0) \\ &= x_0 + \begin{pmatrix} | & | & \dots & | \\ x_1 - x_0 & x_2 - x_0 & \dots & x_k - x_0 \\ | & | & \dots & | \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} \end{aligned}$$

By the affine independence assumption, the matrix must be invertible. This gives that

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} = \begin{pmatrix} | & | & \dots & | \\ x_1 - x_0 & x_2 - x_0 & \dots & x_k - x_0 \\ | & | & \dots & | \end{pmatrix}^{-1} (x(\lambda) - x_0),$$

$$\lambda_0 = 1 - \lambda_1 - \lambda_2 - \dots - \lambda_k.$$

3. Importantly, we record here that $x \mapsto \lambda$ map is just linear.
4. The inverse map as generalization of internal dividing point in a line segment.

(a) We denote the k -volume of $[x_0 x_1 x_2 \dots x_k]$ by

$$|[x_0 x_1 x_2 \dots x_k]| = \mathcal{H}^k(\overline{\text{pts}}([x_0 x_1 x_2 \dots x_k]))$$

(b) Then, the i -th barycentric coordinate is computed by the volume ratio

$$\lambda_i = \frac{|[x_0 x_1 \dots x_{i-1} \ x \ x_{i+1} x_{i+2} \dots x_k]|}{|[x_0 x_1 x_2 \dots x_k]|}.$$

We will prove this in the next lecture.

Nodal basis functions for $\mathbb{P}_0(M)$, $\mathbb{P}_1(M)$, $\mathbb{P}_2(M)$

- Here, we make use of $x \mapsto \lambda_e(x)$ for $e = 0, 1, 2, \dots, n$, the barycentric coordinates, to express the nodal basis functions.
- It is nice to know that, for given degree m , there are exactly d elements in the set

$$\{\lambda_0(x)^{m_0} \lambda_1(x)^{m_1} \cdots \lambda_n(x)^{m_n} \mid m_0 + m_1 + \cdots + m_n = m\}$$

and that each of $x \mapsto \lambda_e(x)$ is linear in x .

- We proceed with $n = 3$ for $\mathbb{P}_0(M)$, $\mathbb{P}_1(M)$, and $\mathbb{P}_2(M)$.

$\mathbb{P}_0(M)$

The only basis function of $\mathbb{P}_0(M)$ is simply a constant function

$$\theta(x) \equiv 1.$$

$\mathbb{P}_1(M)$

For $\mathbb{P}_1(M)$, we define 4 basis functions to be

$$\theta_e(\lambda) = \lambda_e, \quad e = 0, 1, 2, 3.$$

- Note that

$$\theta_e(\hat{e}') = \begin{cases} 0 & \text{if } e \neq e' \\ 1 & \text{if } e = e'. \end{cases}$$

where \hat{e} is the e -th coordinate basis.

$\mathbb{P}_2(M)$

For $\mathbb{P}_2(M)$, we define the 10 basis functions to be

$$\begin{aligned} &\lambda_e(2\lambda_e - 1), \quad e = 0, 1, 2, 3, \\ &4\lambda_e\lambda_{e'}, \quad e, e' = 0, 1, 2, 3, \quad e \neq e'. \end{aligned}$$