École Normale Supérieure de Rennes

2024

Research Diary

Study and implementation of an "a posteriori estimator" for Finite Volume Methods in CFD

Internship diary of
LECOQ Raphaël

Supervised by
STABILE Giovanni

This research diary is intended to sum up the framework, goals and research ideas that I have come across during my research internship.
It is not meant to be used for education purposes but contains the knowledge I found most important to consider for my research on CFD Reduced Order Models.

If you have any questions or find mistyping, please contact us:

raphael.lecoq@ens-rennes.fr
giovanni.stabile@santannapisa.it

This research project takes part in the DANTE ERC StG project.
I am highly grateful to **Giovani STABILE** for accepting me as the first (temporary) member of his AI and Reduced Order Methods (ROM) lab and giving me so much of his time.

My focus was on a ROM "a posteriori" error estimator that is presented p.15 which is based on the Finite Elements Method (FEM) framework. Namely, I tried to formulate the Finite Volumes Methods (FVM) in the FEM framework to make use of this estimator.

This report is divided as follow:
Part 1 focus on a quick introduction of the FEM without going deep into the theory.
Part 2 introduces the Certified Reduced Basis Method that is already established for FEM.
Part 3 studies a cell-centered FVM andd some error estimators that are used in CFD.
Part 4 links the two methods with the Discontinuous Galerkin Method and the Box Method.
Part 5 presents our work: the merging of a Weak FVM formulation with the RBM.
An **Appendix** and a **Bibliography** are given for standards results and references.

The main result of this report is the connection between a FVM formulation that blends in the FEM framework and the ROM framework.
For further research, the extension of the result may lies in the *Mathematical aspects of discontinuous Galerkin method* [PE12] or the *The Gradient Discretisation Method* [Dro+18], whose links with the problem will not be explicitly addressed in the report.

I created a shorter version of this report for my Internship Report, with more straightforward presentation of the research, see [Lec24b].

Contents

Chapter 4 Discontinous Galerkin Method

Chapter 5 Cell centered FVM Reduced Basis Method

Chapter 6 Appendix

Chapter Bibliography

Part 1
Finite Element Method

The goal is to approximate the solution of a PDE that lives in an infinite dimensional space by the solution of the same PDE restricted to a finite dimensional dimension.
The Finite Element Method is one of the most used methods for solving such problems efficiently.
See more in the following book: *The Mathematical Theory of Finite Element Methods,* [BS08]

I - Weak formulation

1) Parametrized Partial Differential Equation

Take any regular set open $\Omega \subset \mathbb{R}^d, d \in \{1,2,3\}$.
Define $\partial\Omega := \overline{\Omega}\backslash\Omega$.
We consider field variables $\omega : \Omega \longrightarrow \mathbb{R}^{d_v}$.
Set $(\Gamma_i^D)_{1\leq i\leq d_v}$ such that $\Gamma^D := \bigcup \Gamma_i^D \subset \partial\Omega$ (not necessary equal).
Define $\mathbb{V}_i := \left\{ v \in \mathbf{H}^1(\Omega,\mathbb{R}) \ / \ v|_{\Gamma_i^D} = 0 \right\}$ $(v : \Omega \to \mathbb{R})$.

$$\mathbb{V} := \prod_{i=1}^{d_v} \mathbb{V}_i$$

Remark :

$\mathbb{V}_i$ is the space of the $i-th$ coordinate in $\mathbb{R}^{d_v}$ of a solution.

$$v \in \mathbb{V} \Rightarrow v \cong \sum_{i=1}^{d_v} v_i \underbrace{\varphi_i}_{\in \mathbb{V}_i} \Rightarrow \mathbb{V} \cong \left\{ v \in \mathbf{H}^1(\Omega) \ / \ v : \Omega \to \mathbb{R}^{d_v}, v\big|_{\Gamma^D} = 0 \right\}$$

Note that $\mathbb{V} \subset \mathbf{H}^1$, s.t. if $\langle\cdot|\cdot\rangle_{\mathbb{V}}$ induces $\|\cdot\|_{\mathbb{V}} \sim \|\cdot\|_{\mathbf{H}^1}$, then $(\mathbb{V}, \langle\cdot|\cdot\rangle_{\mathbb{V}})$ is an Hilbert.
We focus on $\mathbb{P} \subset \mathbb{R}^P$ closed set of parameters.

Definition 1.1: Parametrized PDE

Let $f : \mathbb{V} \times \mathbb{P} \to \mathbb{R}$ continuous linear with respect to $\mathbb{V}$.
$\ell : \mathbb{V} \times \mathbb{P} \to \mathbb{R}$ linear with respect to $\mathbb{V}$.
$a : \mathbb{V} \times \mathbb{V} \times \mathbb{P} \to \mathbb{R}$ bilinear coercive continuous symmetric with respect to $\mathbb{V} \times \mathbb{V}$.
We consider

$$\begin{cases} \text{Solve for } u \in \mathbb{V} & a(u,v\ ;\mu) = f(v\ ;\mu) \quad \forall v \in \mathbb{V} \\ \text{Evaluate for } \mu \in \mathbb{P} & s(\mu) := \ell(u\ ;\mu) \end{cases}$$

Let $\alpha(\mu)$ be the coercive constant, $\gamma(\mu)$ the continuous one.
The symmetry and continuity ensure well-posedness of the PDE through Lax-Milgram.

Let $\mu \in \mathbb{P},\ \mu = (\mu_{[1]},\ \dots\ ,\mu_{[P]})$, then we define the solution of the PPDE $u(\mu) = (u_1,\ \dots\ ,u_{d_v})$.

Remark :

ℓ is any linear function to define depending on which output correlation we're looking for.

2) Dicretization

Take $\mu \in \mathbb{P}$.
Suppose there exists $\mathbb{V}_\delta \subset \mathbb{V}$ finite dimensional vector space, we search $u_\delta(\mu) \in \mathbb{V}_\delta$ solution of the PDE on $\mathbb{V}_\delta$.
Let $N_\delta = \dim(\mathbb{V}_\delta)$ such that $\mathbb{V}_\delta = Vect\left(\{\varphi_i\}_{i=1}^{N_\delta}\right)$.

Definition 1.2: Discretized PDE

Find $u_\delta(\mu)$ such that

$$\begin{cases} a(u_\delta(\mu), v_\delta; \mu) = f(v_\delta; \mu) \quad \forall\ v_\delta \in \mathbb{V}_\delta \\ s_\delta(\mu) = \ell(u_\delta(\mu); \mu) \end{cases}$$

Since $a(u_\delta(\mu), v_\delta; \mu) = f(v_\delta; \mu) = a(u(\mu), v_\delta; \mu)$ there holds

Proposition 1.3: Galerkin's orthogonality

For all v_δ in $\mathbb{V}_\delta$ the following orthogonality holds:

$$a(u_\delta(\mu) - u(\mu), v_\delta; \mu) = 0$$

One important lemma is the following one, which states that the error induced by the solution of the equation is proportional to the best estimation of u we could hope in the discrete space $\mathbb{V}_\delta$.

Lemma 1.4: Céa's lemma

For all $v_\delta \in \mathbb{V}_\delta$:

$$\|u(\mu) - u_\delta(\mu)\|_{\mathbb{V}} \leq \left(1 + \frac{\gamma(\mu)}{\alpha(\mu)}\right) \inf_{v_\delta \in \mathbb{V}_\delta} \|u(\mu) - v_\delta\|_{\mathbb{V}}$$

$\mathcal{D}$

First note

$$\begin{aligned} \alpha \|u(\mu) - u_\delta(\mu)\|_{\mathbb{V}}^2 \leq a(u(\mu) - u_\delta(\mu), u(\mu) - u_\delta(\mu)) &= a(u(\mu) - u_\delta(\mu), u(\mu)) \\ &= a(u(\mu) - u_\delta(\mu), u(\mu)) - v_\delta) \\ &\leq \gamma \|u(\mu) - u_\delta(\mu)\|_{\mathbb{V}} \|u(\mu) - v_\delta\|_{\mathbb{V}} \end{aligned}$$

Then

$$\begin{aligned}\|u(\mu) - u_\delta(\mu)\|_{\mathbb{V}} &\leq \|u(\mu) - v_\delta\|_{\mathbb{V}} + \|v_\delta - u_\delta(\mu)\|_{\mathbb{V}} \\ &\leq \|u(\mu) - v_\delta\|_{\mathbb{V}} + \frac{\gamma}{\alpha}\|u(\mu) - v_\delta\|_{\mathbb{V}} \\ &= (1 + \frac{\gamma}{\alpha})\|u(\mu) - v_\delta\|_{\mathbb{V}}\end{aligned}$$

□

The solution of the discrete equation is easy to write and we define the linear system as:

Definition 1.5: Truth Solver

We call the truth solver, the solution of the linear system $A_\delta^\mu u_\delta(\mu) = f_\delta^\mu$ where

$$\begin{cases}(M_\delta)_{i,j} &= \langle\varphi_i|\varphi_j\rangle_{\mathbb{V}} \\ (A_\delta^\mu)_{i,j} &= a(\varphi_i, \varphi_j\ ;\mu) \\ (f_\delta^\mu)_i &= f(\varphi_i\ ;\mu) \\ (\ell_\delta^\mu)_i &= \ell(\varphi_i\ ;\mu)\end{cases}$$

II - Space approximation

1) Finite Elements

We cut Ω in N_e disjoints subspaces $\Omega^{(e)}$ and we set N_i nodes that constitutes the nodal basis, we note $(x_i)_{i\leq N_i}$ their coordinates.
A subspace $\Omega^{(e)}$ is called a Finite Element the set of all the Finite Elements is called the mesh. To be have a well defined method, we need to do some assumptions on the geometry of a finite element:

- Each element is a star shapped open set
- Each element is polygonal

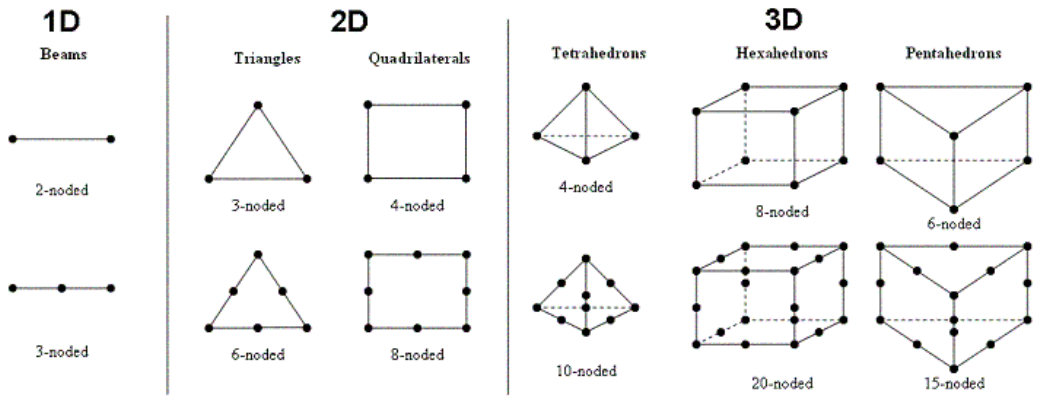


Figure 1: Polygonal elements in 1D, 2D, 3D

One can find conditions over the choice polygonal and how to place the nodes for best results. Most importantly, the diameter of an element controls the polynomial approximation thanks to Poincaré's inequality 6.10. Hence a thinner grid gives better approximation.

2) Polynomial Approximation

The Bramble Hilbert lemma 6.9 shows that a polynomial approximation is a good candidate for Sobolev spaces approximation. Moreover, the polynomials are the easiest functions to work with.
On each nodes, we define a piece-wise polynomial function φ_i of degree lesser than m which is such that

$$\varphi_i(x_j) = \delta_{ij}$$

Such polynomial exists and form an orthogonal basis of functions using Averaged Taylor Polynomials ([BS08] Chap 4). Define

$$\mathbb{V}_\delta := S^{N_e} = \text{Span}\,\{(\varphi_i)_i\} \subset \mathbb{V}$$

We will only consider the linear approximation :

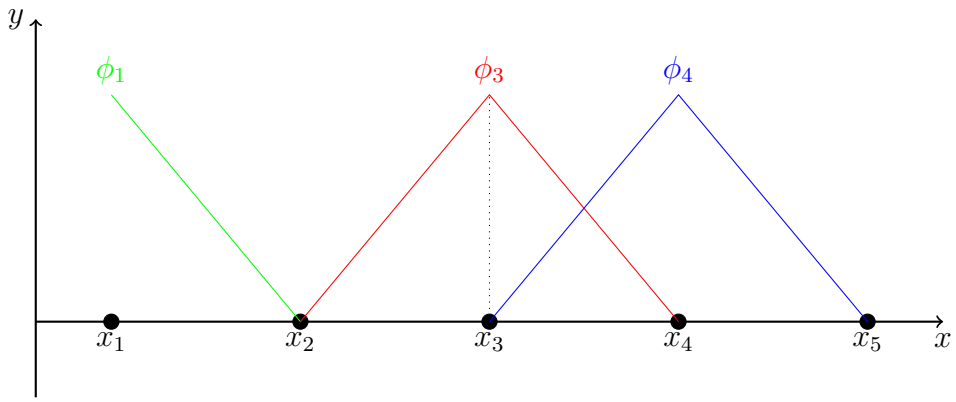


Figure 2: Some basis functions of the nodal basis of linear functions

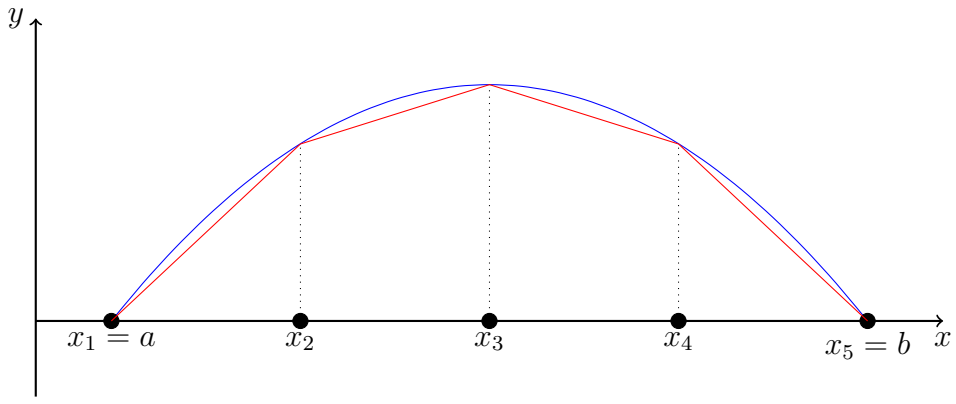


Figure 3: Finite Element approximation in 1D of the function $(x-a)(x-b)$

3) Error estimation

The errors estimators of the FEM will be not adressed since the only one that is of our interest will be studied in the following part.

Part 2
Certified Reduced Basis Method

This part is a rewriting of the *Certified Reduced Basis Methods for Parametrized Partial Differential Equations.* [HRS16] book's chapters that are most important in our research.
The goal is to create a reduced space from $\mathbb{V}_\delta$. The idea is that we are willing to lose time on an "Offline mode" that generate approximation space $\mathbb{V}_{\text{rb}}$ of $\mathbb{V}_\delta$ with a dimension $N_{\text{rb}} \ll N_\delta$ which will allow fast computation of a reduced solution u_{rb} during the called "Online mode".

I - Reduced Basis Method

1) Solution manifold and Reduced Basis Approximation

Definition 2.1: Solution manifold

If we are able to write $u(\mu)$ in analytic form, the solution manifold is :

$$\mathcal{M} = \{u(\mu) \ / \ \mu \in \mathbb{P}\} \subset \mathbb{V}$$

If we can't, consider $\mathbb{V}_\delta$ such as in 2) Discretization

$$\mathcal{M}_\delta = \{u_\delta(\mu) \ / \ \mu \in \mathbb{P}\} \subset \mathbb{V}_\delta$$

Admits there exists $\mathbb{V}_{rb} \subset \mathbb{V}_\delta$, $\dim(\mathbb{V}_{rb}) = N$ such that $N \ll N_\delta < \dim(\mathbb{V}) = +\infty$, there exists $\xi_1, \ \dots \ , \xi_N \in \mathbb{V}_\delta$, such that

$$\mathbb{V}_{rb} = \text{Span}(\xi_1, \ \dots \ , \xi_N)$$

and $\mathbb{V}_{rb}$ is such that for a certain $\varepsilon > 0$ tiny enough to be interesting,

$$\forall v_\delta \in \mathbb{V}_\delta, \ \inf_{v_{rb} \in \mathbb{V}_{rb}} \|v_\delta - v_{rb}\|_{\mathbb{V}} < \varepsilon$$

Definition 2.2: Reduced PDE

Find $u_{rb}(\mu) \in \mathbb{V}_{rb}$ such that

$$\begin{cases} a(u_{rb}(\mu), v_\delta; \mu) = f(v_{rb}; \mu) \ \ \forall \ v_{rb} \in \mathbb{V}_{rb} \\ s_{rb}(\mu) = \ell(u_{rb}(\mu); \mu) \end{cases}$$

For a given $\mathbb{V}_{rb}$ and $\mu \in \mathbb{P}$, Céa's lemma holds with same proof as before:

$$\|u(\mu) - u_{rb}(\mu)\|_{\mathbb{V}} \leq \left(1 + \frac{\gamma(\mu)}{\alpha(\mu)}\right) \inf_{v_{rb} \in \mathbb{V}_{rb}} \|u(\mu) - v_{rb}\|_{\mathbb{V}}$$

The goal is to get $\|u_\delta(\mu) - u_{rb}(\mu)\|$ as close to 0 as possible while keeping $N = dim(\mathbb{V}_{\text{rb}})$ small.

$$\inf_{v_{rb} \in \mathbb{V}_{rb}} \|u(\mu) - v_{rb}\|_{\mathbb{V}} \leq \|u(\mu) - u_{rb}(\mu)\|_{\mathbb{V}} \leq \underbrace{\|u_\delta(\mu) - u_{rb}(\mu)\|_{\mathbb{V}}}_{\text{to be controlled}} + \underbrace{\|u(\mu) - u_\delta(\mu)\|_{\mathbb{V}}}_{\text{controlled by 1.4}}$$

For that we will define some measure of the distance between the space of δ solutions and the reduced space.

$$E(\mathcal{M}_\delta, \mathbb{V}_{rb}) = \sup_{u_\delta \in \mathcal{M}_\delta} \inf_{v_{rb} \in \mathbb{V}_{rb}} \|u_\delta - v_{rb}\|_{\mathbb{V}}$$

We some measure of the distance between the space of δ solutions and the reduced space.

$$E(\mathcal{M}_\delta, \mathbb{V}_{rb}) = \sup_{u_\delta \in \mathcal{M}_\delta} \inf_{v_{rb} \in \mathbb{V}_{rb}} \|u_\delta - v_{rb}\|_{\mathbb{V}}$$

Definition 2.3: Kolmogorov N-Width

Assuming a reduced space exists, the Kolmogorov N-width measures the best distance we can hope with a N dimensional reduced basis and is defined as:

$$d_N(\mathcal{M}_\delta) := \inf_{\{\mathbb{V}_{rb} / \dim\ (\mathbb{V}_{rb}) = N\}} E(\mathcal{M}_\delta, \mathbb{V}_{rb})$$

Remark :

$d_N(\mathcal{M}_\delta)$ measures the best distance we can hope with a N dimensional reduced basis (as long as long as we can find it).

Instead of

$$\sup_{u_\delta \in \mathcal{M}_\delta} \inf_{v_{rb} \in \mathbb{V}_{rb}} \|u_\delta - v_{rb}\|_{\mathbb{V}}$$

we can consider a distance which has a faster computation and that gives the same amount of information with regard to the distance between two spaces :

Definition 2.4: Least squared distance

$$\sqrt{\int_{\mu \in \mathbb{P}} \inf_{v_{rb} \in \mathbb{V}_{rb}} \|u_\delta(\mu) - v_{rb}\|_{\mathbb{V}}^2}$$

Once we defined some quantities that estimate the distance between a supposed reduced basis and the discrete solution manifold, we present two algorithms that allow to compute the reduced basis.

2) Reduced basis generation by Proper Orthogonal Decomposition

For anyone familiar with the Principal Component Analysis, this is the exact same process, described in the functionnal analysis framework.

Let $\mathbb{P}_h = \{\mu_1, \ \dots \ , \mu_M\} \subset \mathbb{P}$ be a discrete and finite point-set.
Define :

$$\mathcal{M}_\delta(\mathbb{P}_h) = \{u_\delta(\mu) \ / \ \mu \in \mathbb{P}_h\}$$

of cardinality $\mathbf{M} = |\mathbb{P}_h|$.
We assume that M is big enough for $\mathcal{M}_\delta(\mathbb{P}_h)$ to efficiently approximate $\mathcal{M}_\delta$.

Let $\mathbb{V}_{\mathcal{M}} = \text{Span}\{u_\delta(\mu) \ / \ \mu \in \mathbb{P}_h\}$. The POD minimizes the least-squared distance for $\mathbb{P}_h$ on all N-dimensional subspaces of $\mathbb{V}_{\mathcal{M}}$:

$$\sqrt{\frac{1}{M}\sum_{\mu\in\mathbb{P}_h}\inf_{v_{rb}\in\mathbb{V}_{rb}}\|u_\delta(\mu)-v_{rb}\|_{\mathbb{V}}^2}$$

Let $\psi_m = u_\delta(\mu_m)$ for $m \in \{1, ..., M\}$ (ψ_m is well-defined by unicity of Lax-Milgram). We project any $v_\delta \in \mathbb{V}_{\mathcal{M}}$ on the space generated by the ψ_m :

$$C(v_\delta) = \frac{1}{M}\sum_{m=1}^{M}\langle v_\delta|\psi_m\rangle_{\mathbb{V}}\,\psi_m \in \mathbb{V}_{\mathcal{M}}$$

This operator is linear and symmetric (OK). This operator is positive :

$$\langle C(v_\delta)|v_\delta\rangle = \frac{1}{M}\sum_{m=1}^{M}\langle v_\delta|\psi_m\rangle\langle\psi_m|v_\delta\rangle = \frac{1}{M}\sum_{m=1}^{M}\langle v_\delta|\psi_m\rangle^2 \geq 0$$

Since it is symmetric and $\mathbb{P}_{\mathcal{M}}$ is finite dimensional, there exists an orthonormal basis of eigenvectors and real eigenvalues $(\lambda_n, \xi_n) \in \mathbb{R}_+ \times \mathbb{V}_{\mathcal{M}}$ such that

$$\langle C(\xi_n)|\psi_m\rangle_{\mathbb{V}} = \lambda_n\langle\xi_n|\psi_m\rangle_{\mathbb{V}}$$

and we can chose the numerating (permutation matrices are orthogonals) such that $\lambda_1 \geq \ ... \ \geq \lambda_M \geq 0$.

Remark :

C being SPD is a consequence of an algebra point of view $C = SS^T$ where S is a snapshot of solutions, i.e. it is the SVD matrix. See [Vol12].

Check that

$$\langle C(\xi_n)|\psi_m\rangle_{\mathbb{V}} = \lambda_n\langle\xi_n|\psi_m\rangle_{\mathbb{V}} \iff C(\xi_n) = \lambda_n\xi_n$$

$\mathcal{D}$

$\Rightarrow$ Suppose $\langle C(\xi_n)|\psi_m\rangle_{\mathbb{V}} = \lambda_n\langle\xi_n|\psi_m\rangle_{\mathbb{V}}$

$$\begin{aligned}
\text{For any } v_\delta \in \mathbb{V}_{\mathcal{M}}, v_\delta &= \sum_{i=1}^{M}\langle v_\delta|\psi_i\rangle_{\mathbb{V}}\,\psi_i \\
\text{Then } C(\xi_m) &= \sum_{i=1}^{M}\langle C(\xi_m)|\psi_i\rangle_{\mathbb{V}}\,\psi_i \\
&= \sum_{i=1}^{M}\lambda_m\langle\xi_m|\psi_i\rangle_{\mathbb{V}}\,\psi_i \\
&= \lambda_m\sum_{i=1}^{M}\langle\xi_m|\psi_i\rangle_{\mathbb{V}}\,\psi_i \\
&= \lambda_m\xi_m
\end{aligned}$$

$\Leftarrow$ OK $\square$

Proposition 2.5: Proper Orthogonal Projection

$\mathbb{V}_{\text{POD}} = \text{Span}\left(\{\xi_m\}_{1\leq m\leq N}\right) \subset \mathbb{V}$ of dimension N (or less).

$\mathcal{D}$

See the lecture notes "Proper Orthogonal Decomposition: Theory and Reduced-Order Modelling" [Vol12]. □

Figure 4: A manifold and a possible plan POD representation

We can define the (orthogonal) projection on the subspace

$$P_N[f] = \sum_{i=1}^{N} \langle f|\xi_i\rangle_{\mathbb{V}}\, \xi_i$$

If the projection is applied to all element of $\mathcal{M}_\delta(\mathbb{P}_h)$

$$\frac{1}{M}\sum_{m=1}^{M} \|\psi_m - P_N(\psi_m)\|_{\mathbb{V}}^2 = \sum_{m=N+1}^{M} \lambda_m$$

$\mathcal{D}$

First, [Vol12] proves that $\inf\limits_{v_{rb}\in\mathbb{V}_{rb}} \|u_\delta(\mu_m) - v_{rb}\|_{\mathbb{V}}^2 = \|u_\delta(\mu_m) - \psi_m\|_{\mathbb{V}}^2$.
Then, we note that:

$$\begin{aligned}
\|\psi_i - P_N[\psi_i]\|_{\mathbb{V}}^2 &= \frac{1}{M}\left\|\sum_{m=1}^{M} \langle\psi_i|\xi_m\rangle\, \mathbb{V}\xi_m - \sum_{m=1}^{N} \langle\psi_i|\xi_m\rangle_{\mathbb{V}}\right\|_{\mathbb{V}}^2 \\
&= \frac{1}{M}\sum_{m=N+1}^{M} \|\langle\psi_i|\xi_m\rangle_{\mathbb{V}}\, \xi_m\|_{\mathbb{V}}^2 \\
&= \frac{1}{M}\sum_{m=N+1}^{M} \langle\psi_i|\xi_m\rangle_{\mathbb{V}}^2 \quad \text{orthonormality and Pythagore}
\end{aligned}$$

Then

$$C(\xi_m) = \frac{1}{M}\sum_{i=1}^{M} \langle \xi_m | \psi_i \rangle_{\mathbb{V}} \, \psi_i = \lambda_m \xi_m$$

Applying the inner product against ξ_m, we recall that $\|\xi_m\|_{\mathbb{V}} = 1$:

$$\langle C(\xi_m) | \xi_m \rangle = \frac{1}{M}\sum_{i=1}^{M} \langle \xi_m | \psi_i \rangle_{\mathbb{V}} \langle \psi_i | \xi_m \rangle_{\mathbb{V}} = \frac{1}{M}\sum_{i=1}^{M} \langle \xi_m | \psi_i \rangle_{\mathbb{V}}^2 = \lambda_m$$

$$\begin{aligned}
\text{Hence } \frac{1}{M}\sum_{i=1}^{M} \|\psi_i - P_N(\psi_i)\|_{\mathbb{V}}^2 &= \frac{1}{M}\sum_{i=1}^{M}\sum_{m=N+1}^{M} \langle \psi_i | \xi_m \rangle_{\mathbb{V}}^2 \\
&= \frac{1}{M}\sum_{m=N+1}^{M} \underbrace{\sum_{i=1}^{M} \langle \psi_i | \xi_m \rangle_{\mathbb{V}}^2}_{=M\lambda_m} \\
&= \sum_{m=N+1}^{M} \lambda_m
\end{aligned}$$

□

Notice that when the projection space grows, this error estimation tend to 0 which is exactly the expected behavior of the reduced basis.

Remark :

One first major flaw of this algorithm is that we have no control over the approximation other than the first non considered eigenvalue of the SVD matrix.

Second major flaw : we need to produce M times the truth solver that solves in dimension N_δ which costs at best MN_δ^2 to obtain all the solutions to compute $\mathbb{V}_{\mathcal{M}}$ and since $M \gg N$ the complexity scales as $\mathcal{O}(NN_\delta^2)$ complexity.

Hence we seek an alternative, less precise approach that will allows faster computing with an error estimator.

3) Reduced basis generation by Greedy algorithm

The goal is to construct a basis thanks to an error estimator that should behaves as the error behaves. Its expected behavior is described in the next section.
This estimator has been the main focus of our research in the framework of Finite Volume Methods, that will also be studied later.

Assume there exists be an upper bound of the error approximation $\eta(\mu)$ such that

$$\|u_\delta(\mu) - u_{rb}(\mu)\|_\mu \leq \eta(\mu) \quad \forall \mu \in \mathbb{P}$$

At dimensionality n, choose $\psi_{n+1} = u_\delta(\mu_{n+1})$ such that

$$\mu_{n+1} = \arg\max_{\mu \in \mathbb{P}} \eta_n(\mu)$$

i.e. we add the parametrized solution that the current space worst approximates.

Remark :

Note η_n depends on the iteration (otherwise we take the same μ each time).

As usual, we introduce $\mathbb{P}_h$ discrete and finite set-point of parameter to compute the arg max.

Remark :

We only need to control from above $\eta(\mu)$ to estimate the arg max. If we find a good η the computation is much faster, but there still need to find such one.

General form (without considering PDE) :
Let $F = \{f(\mu) \backslash \mu \in \mathbb{P}\}$ where $f(\mu) : \Omega \to \mathbb{R}$.
Take $f_1 \in F$ such that $f_1 = \arg\max_{\mu \in \mathbb{P}} \|f(\mu)\|_{\mathbb{V}}$. Suppose we chose $f_1, \ \dots \ , f_n$. Let f_{n+1} such that

$$f_{n+1} = \arg\max_{\mu \in \mathbb{P}} \|f(\mu) - P_n[f(\mu)]\|_{\mathbb{V}}$$

where P_n is the projection onto $F_n = \text{Span}\,(f_1, \ \dots \ , f_n)$.

Theorema 2.6

Assume that F has exponentially small Kolmogorov N-width, i.e. $d_N(F) \leq ce^{-aN}$ with $a > log(2)$.
Then there exists $\beta > 0$ such that

$$\|f - P_N[f]\|_{\mathbb{V}} \leq Ce^{-\beta N}$$

This result can be applied to the Greedy Algorithm and its generic quantities:

Theorema 2.7

Assume that $\mathcal{M}$ has exponentially small Kolmogorov N-width, i.e. $d_N(F) \leq ce^{-aN}$ with $a > log(1 + \sqrt{\frac{\gamma}{\alpha}})$.
Then there exists $\beta > 0$ such that

$$\forall \mu \in \mathbb{P}, \|u_\delta(\mu) - u_{rb}(\mu)\|_{\mathbb{V}} \leq Ce^{-\beta N}$$

Remark :

$\alpha(\mu) \|u(\mu)\|_{\mathbb{V}}^2 \leq a(u(\mu), u(\mu); \mu) \leq \gamma(\mu) \|u(\mu)\|_{\mathbb{V}}^2$
$\Rightarrow \alpha \leq \gamma$
$\Rightarrow 1 \leq \frac{\gamma}{\alpha}$
$\Rightarrow a > log(1 + \frac{\gamma}{\alpha}) \geq log(2)$
Hence the condition on a is stronger than in the previous theorem.

$\mathcal{D}$

"Apriori convergence of the greedy algorithm for the parametrized reduced basis method."

[Buf+21] □

Remark :

The reduced basis generation using a Greedy method realizes the same asymptotic rate of decay as the Kolmogorov N-width [Bin+11].

Example 1: Some Kolmogorov N-width for several PDEs [Sta23]

$INS(f)$,	$f \in \mathcal{K}_\gamma^{\bar{\omega},s}$,	$d_n(\mathcal{M})_{L^2} < \exp -n^{1/3}$	[Schwab and Suri 1999]
$-\mathrm{div}(\mu\nabla u) = f$,	$\mu \in \mathbb{P} \subset \mathbb{R}^m$,	$d_n(\mathcal{M})_{L^2} < \exp -n$	[Babuška et al. 2007]
$u^3 - \nabla\cdot(\exp\mu)\nabla u = f$,	$\mu \in K \subset W^{s,\infty}(\Omega)$,	$d_n(\mathcal{M})_{L^2} < n^{-\frac{s}{d}}$	[Cohen and DeVore, 2016]
$\partial_t u - \mu\partial_x u = 0$,	$(\mu, t) \in [0,1]^2$,	$d_n(\mathcal{M})_{L^2} > n^{-\frac{1}{2}}$	[Ohlberger and Rave, 2015]
$\partial_{tt}^2 u - \mu\partial_{xx}^2 u = 0$,	$(\mu, t) \in [0,1]^2$,	$d_n(\mathcal{M})_{L^2} > n^{-\frac{1}{2}}$	[Greif and Urban, 2019]

4) Reduced solution computation

Suppose there exists an **affine decomposition** of a, f, ℓ i.e. there exists :

$$\begin{aligned} \mathcal{Q}_a \in \mathbb{N}, (a_q(v,w))_{1\le q\le \mathcal{Q}_a} \quad & a_q : \mathbb{V}\times\mathbb{V} \to \mathbb{R} \\ \mathcal{Q}_f \in \mathbb{N}, (f_q)_{1\le q\le \mathcal{Q}_f} \quad & f_q : \mathbb{V} \to \mathbb{R} \\ \mathcal{Q}_\ell \in \mathbb{N}, (\ell_q)_{1\le q\le \mathcal{Q}_\ell} \quad & \ell_q : \mathbb{V} \to \mathbb{R} \end{aligned}$$

such that

$$\begin{cases} a(v,w\,;\mu) &= \displaystyle\sum_{q=1}^{\mathcal{Q}_a} \theta_a^q(\mu) a_q(v,w) \\ f(v\,;\mu) &= \displaystyle\sum_{q=1}^{\mathcal{Q}_f} \theta_f^q(\mu) f_q(v) \\ \ell(v\,;\mu) &= \displaystyle\sum_{q=1}^{\mathcal{Q}_\ell} \theta_q^\ell(\mu) \ell_q(v) \end{cases}$$

With: $\theta_a^q : \mathbb{P} \to \mathbb{R}$ $\theta_f^q : \mathbb{P} \to \mathbb{R}$ $\theta_\ell^q : \mathbb{P} \to \mathbb{R}$

i.e. it is supposed that the equation is described by linear functions independents of μ multiplied by a scalar dependent of μ. It is called the **affine assumption**.

Example 2: Affine assumption example

The heat equation admits an affine decomposition, see 2.3.1 and 3.4.1 [HRS16]. It can also be forced through the Empirical Interpolation Method, see Part 5 of the same reference.

Compute for each $1 \le q \le \mathcal{Q}_a, \mathcal{Q}_f, \mathcal{Q}_\ell$ the quantities

$$\mathbf{A}_\delta^q \mid f_\delta^q \mid \ell_\delta^q$$

which are the representation of these functions in the basis of discretization (as for the Truth Solver 1.5). Then compute for each q

$$\begin{cases} \mathbf{A}_{rb}^q &= \mathbf{B}\mathbf{A}_\delta^q\mathbf{B}^T \\ f_{rb}^q &= \mathbf{B}^T f_\delta^q \\ \ell_{rb}^q &= \mathbf{B}^T \ell_\delta^q \end{cases}$$

where $\mathbf{B}$ is the projection matrix from Span $(\varphi_1, \ \dots \ , \varphi_{N_\delta})$ to Span $(\xi_1, \ \dots \ , \xi_N)$ which is the orthonormed reduced basis (by Gram-Schmidt) for stability.

Then for each $\mu \in \mathbf{P}$, considering the dependency in μ being only on the θ^q, we can rapidly compute the X^μ quantities such that

$$\begin{cases} \mathbf{A}_{rb}^\mu &= \displaystyle\sum_{q=1}^{\mathcal{Q}_a} \theta_a^q(\mu)\mathbf{A}_{rb}^q \\ f_{rb}^q &= \displaystyle\sum_{q=1}^{\mathcal{Q}_f} \theta_f^q(\mu) f_\delta \\ \ell_{rb}^\mu &= \displaystyle\sum_{q=1}^{\mathcal{Q}_\ell} \theta_\ell^q(\mu)\ell_\delta^q \end{cases}$$

We can finally solve

$$\boxed{\mathbf{A}_{rb}^\mu u_{rb}^\mu = f_{rb}^\mu}$$

that gives us the reduced basis solution, and ℓ_{rb}^μ gives us the output.
The advantage after having created all the X_{rb}^q (offline procedure), we only have to compute the X^μ for each μ by computing $\theta_x^q(\mu)$ (online procedure).

II - Error estimation

Lets introduce the discrete coercivity and continuous constants such that

Definition 2.8: Discrete constants

$$\alpha_\delta(\mu) = \inf_{\substack{v_\delta \in \mathbb{V}_\delta \\ \|v_\delta\|_{\mathbb{V}} = 1}} |a(v_\delta, v_\delta \ ; \mu)|, \quad \text{and} \quad \gamma_\delta(\mu) = \sup_{\substack{v_\delta, w_\delta \in \mathbb{V}_\delta \\ \|v_\delta\|_{\mathbb{V}} = \|w_\delta\|_{\mathbb{V}} = 1}} |a(v_\delta, w_\delta \ ; \mu)|$$

Since the supremum and the infimum are taken on a subset of $\mathbb{V}$, there holds :

$$\alpha \leq \alpha_\delta \text{ and } \gamma_\delta \leq \gamma$$

1) Expected behavior of an error estimate

Following "Error Analysis and Estimation for the Finite Volume Method With Applications to Fluid Flows" [Jas96], we expect the following behavior from an error estimate :

- Give reliable informations about the distribution of the error
- Work well on coarse mesh

- Scale corresponding to mesh refinement
- Scale corresponding to discretisation
- Based on local solution and mesh information, cell-by-cell
- Asymptotically correct
- Over-estimate of the actual error

Definition 2.9: Asymptotically correct

Let N be the number of computation points.
Let E_N the exact error of the approximation solution u_N with respect to the exact solution u for a prescribed PDE.

$$E_N = \|u_N - u_h\|$$

Let e_N be an error estimate of E_N.
e_N is asymptotically correct if

$$\frac{e_N - E_N}{E_N} \underset{N\to\infty}{\longrightarrow} 0$$

or equivalently

$$\xi_N := \frac{e_N}{E_N} \underset{N\to\infty}{\longrightarrow} 1$$

where $\xi_N \geq 1$ is the effectivity of the error estimate.
It means the error estimate tends to the exact error faster than the estimated solution tends to the exact solution.

2) Error estimator

We define naturally the error and its classic error residual that quantifies how much the reduced solution satisfies the discrete equation :

Definition 2.10: Error and classic error equation

For $\mu \in \mathbb{P}$, we define the error of the discrete space by the reduced basis such that

$$e(\mu) = u_\delta(\mu) - u_{rb}(\mu)$$

which satisfies the equation

$$a(e(\mu), v_\delta\ ;\mu) = r(v_\delta\ ;\mu) \quad \forall v_\delta \in \mathbb{V}_\delta$$

where $r(\cdot\ ;\mu) \in \mathbb{V}'_\delta$ (the topological dual),

$$r(v_\delta\ ;\mu) = f(v_\delta\ ;\mu) - a(u_{rb}, v_\delta\ ;\mu)$$

Note that $r(\cdot\ ;\mu)$ being in the dual of $\mathbb{V}_\delta$, we can apply Riesz (see Theorem 6.5) hence it exists $\hat{r}_\delta$ satisfying

$$\langle \hat{r}_\delta(\mu) | v_\delta \rangle_{\mathbb{V}} = r(v_\delta\ ;\mu)$$

We recall that

$$\|\hat{r}_\delta(\mu)\|_{\mathbb{V}} = \|r(\cdot\,;\mu)\|_{\mathbb{V}'_\delta} = \sup_{\substack{v_\delta \in \mathbb{V}_\delta \\ \|v_\delta\|_{\mathbb{V}}=1}} |r(v_\delta\,;\mu)|$$

Proposition 2.11

For a compliant problem, it holds for all $\mu \in \mathbb{P}$

$$s_\delta(\mu) - s_{rb}(\mu) = \|u_\delta(\mu) - u_{rb}(\mu)\|^2_\mu$$

Hence

$$s_\delta(\mu) \geq s_{rb}(\mu)$$

$\mathcal{D}$

Set $\mu \in \mathbb{P}$. By Definition 1.2 and Definition 2.2, Galerkin's Orthogonality 1.3 holds :

$$a(u_\delta(\mu) - u_{rb}(\mu), v_{rb}\,;\mu) = 0 \quad \forall v_{rb} \in \mathbb{V}_{rb}$$

Then

$$\begin{aligned}
s_\delta(\mu) - s_{rb}(\mu) &= \ell(u_\delta\,;\mu) - \ell(u_{rb}\,;\mu) \\
&= \ell(u_\delta(\mu) - u_{rb}(\mu)\,;\mu) \\
&= \ell(e(\mu)\,;\mu) \text{ the problem is compliant} \\
&= f(e(\mu)\,;\mu) \text{ note that } e(\mu) \in \mathbb{V}_\delta \\
&= a(u_\delta, e(\mu)\,;\mu) \\
&= a(e(\mu), e(\mu)\,;\mu) + a(u_{rb}, e(\mu)\,;\mu) \quad a \text{ is symmetric} \\
&= a(e(\mu), e(\mu)\,;\mu) + \underbrace{a(e(\mu), u_{rb}\,;\mu)}_{=0} \text{ by Galerkin's orthogonality} \\
&= a(e(\mu), e(\mu)\,;\mu) \\
&= \|e(\mu)\|^2_\mu \geq 0
\end{aligned}$$

□

Assume there is a known lower bound α_{LB} of α_δ in a way that's independent of N_δ.
The construction of such lower bound is adressed in [HRS16] and in the following parts of the report.

The following error estimator is the main interest of this report.

Definition 2.12: Energy norm, output, relative output error estimators

We define computable upper bound of the energy norm, output and relative output :

$$\eta_{en}(\mu) = \frac{\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}}{\alpha_{\mathrm{LB}}(\mu)^{1/2}}$$
$$\eta_s(\mu) = \frac{\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}^2}{\alpha_{\mathrm{LB}}(\mu)} = (\eta_{en}(\mu))^2$$
$$\eta_{s,rel} = \frac{\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}^2}{\alpha_{\mathrm{LB}}(\mu) s_{rb}(\mu)} = \frac{\eta_s(\mu)}{s_{rb}(\mu)}$$

Remark :

η_{en} is a natural upper bound :

Recalling the definition of $\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}$ p.14

$$\begin{aligned}
\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}^2 &\geq \left(\frac{|r(e(\mu)\ ;\mu)|}{\|e(\mu)\|_{\mathbb{V}}}\right)^2 \\
&= \left(\frac{a(e(\mu), e(\mu)\ ;\mu)}{\|e(\mu)\|_{\mathbb{V}}}\right)^2 \\
&\geq \frac{a(e(\mu), e(\mu)\ ;\mu)}{\|e(\mu)\|_{\mathbb{V}}^2} \alpha_{\mathrm{LB}}(\mu) \|e(\mu)\|_{\mathbb{V}}^2 \\
&= \alpha_{\mathrm{LB}}(\mu) \|e(\mu)\|_{\mu}^2 \\
\text{Hence } \frac{\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}}{\sqrt{\alpha_{\mathrm{LB}}(\mu)}} &\geq \|e(\mu)\|_{\mu}
\end{aligned}$$

And $\|\cdot\|_\mu$ is called the energy norm induced by the PDE (thus $a(\cdot,\cdot\ ;\mu)$) since it's the natural norm defined by the PDE.

Proposition 2.13: Upper bound control

$$\|u_\delta(\mu) - u_{rb}(\mu)\|_\mu \leq \eta_{en}(\mu)$$
$$s_\delta(\mu) - s_{rb}(\mu) \leq \eta_s(\mu)$$
Suppose $s_\delta > 0$,
$$\frac{s_\delta(\mu) - s_{rb}(\mu)}{s_\delta(\mu)} \leq \eta_{s,rel}$$

$\mathcal{D}$

Recall $\langle \hat{r}_\delta(\mu) | v_\delta \rangle_{\mathbb{V}} = r(v_\delta\ ;\mu) = a(e(\mu), v_\delta\ ;\mu)$.

With $e(\mu) \in \mathbb{V}_\delta$ and Cauchy-Schwarz, we deduce

$$\|e(\mu)\|_\mu^2 = a(e(\mu), e(\mu)\ ;\mu) = \langle \hat{r}_\delta(\mu) | e(\mu) \rangle_{\mathbb{V}} \leq \|\hat{r}_\delta(\mu)\|_{\mathbb{V}} \|e(\mu)\|_{\mathbb{V}}$$

By hypothesis, $\alpha_{\text{LB}} \leq \alpha$ hence

$$\alpha_{\text{LB}} \|e(\mu)\|_{\mathbb{V}}^2 \leq \alpha \|e(\mu)\|_{\mathbb{V}}^2 \leq a(e(\mu), e(\mu)\ ;\mu) = \|e(\mu)\|_\mu^2 \leq \|\hat{r}_\delta(\mu)\|_{\mathbb{V}} \|e(\mu)\|_{\mathbb{V}}$$

Thus

$$\|e(\mu)\|_{\mathbb{V}} \leq \frac{1}{\alpha_{LB}} \|\hat{r}_\delta(\mu)\|_{\mathbb{V}}$$

And

$$\|e(\mu)\|_\mu^2 \leq \frac{1}{\alpha_{LB}} \|\hat{r}_\delta(\mu)\|_{\mathbb{V}}^2$$

It follows the first inequality.
Then by Proposition 2.11

$$s_\delta(\mu) - s_{rb}(\mu) = \|e(\mu)\|_\mu^2$$

Since

$$\eta_s(\mu) = \eta_{en}(\mu)^2 \text{ and } u_\delta(\mu) - u_{rb}(\mu) = e(\mu)$$

By taking the square

$$s_\delta(\mu) - s_{rb}(\mu) = \|u_\delta(\mu) - u_{rb}\|_\mu^2 \leq \eta_{en}(\mu)^2 = \eta_s(\mu)$$

hence second inequality holds.
Last, by Proposition 2.11 $s_\delta(\mu) \geq s_{rb}(\mu)$, hence $\dfrac{1}{s_\delta(\mu)} \leq \dfrac{1}{s_{rb}(\mu)}$.
Then

$$\frac{e(\mu)}{s_\delta(\mu)} \leq \frac{e(\mu)}{s_{rb}(\mu)} \leq \frac{\eta_s(\mu)}{s_{rb}(\mu)} = \eta_{s,rel}(\mu)$$

□

We proved that η_{en} is a error estimator that we can use for the Greedy Algorithm ! Following Definition 2.9, we will then define the effectivity of theses estimates :

Definition 2.14: Effectivity

We define the effectivity of the computable estimators :

$$\text{eff}_{en}(\mu) = \frac{\eta_{en}(\mu)}{\|e(\mu)\|_\mu}$$

$$\text{eff}_s(\mu) = \frac{\eta_s(\mu)}{s_\delta(\mu) - s_{rb}(\mu)}$$

$$\text{eff}_{s,rel}(\mu) = \frac{n_{s,rel}(\mu)}{(s_\delta(\mu) - s_{rb}(\mu))/s_\delta(\mu)}$$

It mesures the sharpness of the estimators.

These effectivities are ≥ 1 by Proposition 2.13.
We require them to be as close as possible to 1 according to Def 2.9.

Defined like that, there is no fast way to compute the effectivities. The following property gives computables upper bounds:

Proposition 2.15: Effectivity control

For all $\mu \in \mathbb{P}$

$$1 \leq \text{eff}_{en} \leq \sqrt{\gamma_\delta/\alpha_{\text{LB}}}$$
$$1 \leq \text{eff}_s \leq \gamma_\delta/\alpha_{\text{LB}}$$
$$\text{Suppose } s_\delta > 0$$
$$1 \leq \text{eff}_{s,rel} \leq (1+\eta_{s,rel})\gamma_\delta/\alpha_{\text{LB}}$$

$\mathcal{D}$

Inequality 1 : by definition p.14 and Cauchy-Schwarz with the dot product $a(\cdot,\cdot\,;\mu)$: $\|\hat{r}_\delta(\mu)\|^2_{\mathbb{V}} = a(e(\mu), \underbrace{\hat{r}_\delta(\mu)}_{\in \mathbb{V}_\delta}) \leq \|e(\mu)\|_\mu \|\hat{r}_\delta(\mu)\|_\mu$.

And $\|\hat{r}_\delta(\mu)\|^2_\mu = a(\hat{r}_\delta(\mu), \hat{r}_\delta(\mu)) \leq \gamma_\delta(\mu)\|\hat{r}_\delta(\mu)\|^2_{\mathbb{V}} \leq \gamma_\delta(\mu)\|e(\mu)\|_\mu \|\hat{r}_\delta(\mu)\|_\mu$.

Hence $\eta_{\text{en}}(\mu)^2 \leq \dfrac{\gamma_\delta(\mu)}{\alpha_{LB}(\mu)}$.

Inequality 3 :

$$\eta_{s,rel} = \eta_s/s_{rb} \;\Rightarrow \text{eff}_{s,rel} = \frac{s_\delta}{s_{rb}}\text{eff}_s$$

Yet

$$\frac{s_\delta}{s_{rb}} = \frac{s_\delta - s_{rb}}{s_{rb}} + 1 \leq \frac{\eta_s}{s_{rb}} + 1 = \eta_{s,rel} + 1$$

Hence

$$\text{eff}_{s,rel} \leq (\eta_{s,rel}+1)\frac{\gamma_\delta}{\alpha_{\text{LB}}}$$

□

The estimator effectivity of the energy norm and output error is bounded from above by independent of N.
Thus we will only have to compute this quantities once for each parameter considered.
We then try to provide error bounds with respect to the $\mathbb{V}$-norm for $e(\mu)$.

Definition 2.16: $\mathbb{V}$-norm error estimator

The $\mathbb{V}$-norm can be replaced by any $\mathbb{V}_\delta$ norm adapting the definition of α_{LB}.

$$\eta_{\mathbb{V}}(\mu) = \frac{\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}}{\alpha_{\text{LB}}(\mu)}$$
$$\eta_{\mathbb{V},rel}(\mu) = \frac{2}{\|u_{rb}(\mu)\|_{\mathbb{V}}}\eta_{\mathbb{V}}(\mu)$$

As for the previous estimator, the next property proves that this define real error estimator. What's most important is that these are fast to compute.

Proposition 2.17: $\mathbb{V}$-norm error control

$$\|e(\mu)\|_{\mathbb{V}} \leq \eta_{\mathbb{V}}(\mu)$$

Furthermore, if $\eta_{\mathbb{V},rel} \leq 1$ for a $\mu \in \mathbb{P}$, then the relative error verifies

$$\frac{\|e(\mu)\|_{\mathbb{V}}}{\|u_\delta(\mu)\|_{\mathbb{V}}} \leq \eta_{\mathbb{V},rel}(\mu)$$

$\mathcal{D}$

First inequality : $\|\hat{r}_\delta(\mu)\|_{\mathbb{V}} \geq r\left(\frac{e(\mu)}{\|e(\mu)\|_{\mathbb{V}}}\,;\mu\right) = a\left(e(\mu), \frac{e(\mu)}{\|e(\mu)\|_{\mathbb{V}}}\,;\mu\right) \geq \alpha_{LB}(\mu)\,\|e(\mu)\|_{\mathbb{V}}$.

□

Definition 2.18: $\mathbb{V}$-effectivity

As per the μ-control case, we define the effictivity of the estimators such that

$$\text{eff}_{\mathbb{V}}(\mu) = \frac{\eta_{\mathbb{V}}(\mu)}{\|e(\mu)\|_{\mathbb{V}}}$$

$$\text{eff}_{\mathbb{V},rel} = \frac{\eta_{\mathbb{V},rel}(\mu)}{\|e(\mu)\|_{\mathbb{V}} \,/\, \|u_\delta(\mu)\|_{\mathbb{V}}}$$

Proposition 2.19: $\mathbb{V}$-effectivity control

It holds that

$$1 \leq \text{eff}_{\mathbb{V}} \leq \frac{\gamma_\delta}{\alpha_{\text{LB}}}$$

Furthermore, if $\eta_{\mathbb{V},rel}(\mu) \leq 1$ for a $\mu \in \mathbb{P}$ then

$$\text{eff}_{\mathbb{V},rel}(\mu) \leq 3\frac{\gamma_\delta(\mu)}{\alpha_{\text{LB}}(\mu)}$$

$\mathcal{D}$

The first inequality follows from 2.15 and $\|e(\mu)\|_\mu \leq \sqrt{\gamma_\delta(\mu)}\,\|e(\mu)\|_{\mathbb{V}}$ as

$$\text{eff}_{\mathbb{V}}(\mu) = \frac{\text{eff}_{en}(\mu)}{\sqrt{\alpha_{\text{LB}}(\mu)}}\frac{\|e(\mu)\|_\mu}{\|e(\mu)\|_{\mathbb{V}}} \leq \frac{\sqrt{\gamma_\delta(\mu)}}{\alpha_{\text{LB}}(\mu)}\frac{\|e(\mu)\|_\mu}{\|e(\mu)\|_{\mathbb{V}}} \leq \frac{\gamma_\delta(\mu)}{\alpha_{LB}(\mu)}$$

□

3) Computation of the estimator

Goal :

$$\text{Compute } \|\hat{r}_\delta(\mu)\|_{\mathbb{V}}$$

Recall the affine assumption

$$\begin{aligned} r(v_\delta\ ;\mu) &= a(e(\mu), v_\delta\ ;\mu) \\ &= a(u_\delta(\mu), v_\delta\ ;\mu) - a(u_{rb}(\mu), v_\delta\ ;\mu) \\ &= f(v_\delta;\mu) - a(u_{rb}(\mu), v_\delta\ ;\mu) \\ &= \sum_{q=1}^{\mathcal{Q}_f} \theta_f^q(\mu) f_q(v_\delta) - \sum_{q=1}^{\mathcal{Q}_a} \theta_a^q(\mu) a_q(u_{rb}(\mu), v_\delta) \text{ affine assumption} \quad 4) \end{aligned}$$

Then we know

$$u_{rb}(\mu) = \sum_{n=1}^{N} (u_{rb}^\mu)^n \xi_n$$

Hence

$$\begin{aligned} r(v_\delta\ ;\mu) &= a(e(\mu), v_\delta\ ;\mu) \\ &= \sum_{q=1}^{\mathcal{Q}_f} \theta_f^q(\mu) f_q(v_\delta) - \sum_{q=1}^{\mathcal{Q}_a} \sum_{n=1}^{N} (u_{rb}^\mu)^n \theta_a^q(\mu) a_q(\xi_n, v_\delta) \end{aligned}$$

Let $\mathcal{Q}_r := \mathcal{Q}_f + N\mathcal{Q}_a$ and define

$$\begin{aligned} r(\mu) &:= \left(\theta_f^1(\mu),\ \dots\ , \theta_f^{\mathcal{Q}_f}(\mu), -(u_{rb}^\mu)_1 \theta_a^1(\mu),\ \dots\ , -(u_{rb}^\mu)_N \theta_a^1(\mu), -(u_{rb}^\mu)_1 \theta_a^2(\mu),\ \dots\ , -(u_{rb}^\mu)_N \theta_a^{\mathcal{Q}_a}(\mu)\right)^T \\ &= \left(\theta_f^1(\mu),\ \dots\ , \theta_f^{\mathcal{Q}_f}(\mu), -(u_{rb}^\mu)^T \theta_a^1(\mu),\ \dots\ , -(u_{rb}^\mu)^T \theta_a^{\mathcal{Q}_a}(\mu)\right)^T \in \mathbb{R}^{\mathcal{Q}_r} \end{aligned}$$

Then consider the vectors of forms $F \in (\mathbb{V}'_\delta)^{\mathcal{Q}_f}$ and $A_q \in (\mathbb{V}'_\delta)^N$ for $1 \le q \le \mathcal{Q}_a$ such that

$$F = \left(f_1,\ \dots\ , f_{\mathcal{Q}_f}\right), \quad \text{and} \quad A_q = (A_1,\ \dots\ , A_{\mathcal{Q}_a}, a_q(\xi_1, \cdot),\ \dots\ , a_q(\xi_N, \cdot))$$

and define the vector of forms $R \in (\mathbb{V}'_\delta)^{\mathcal{Q}_r}$ as

$$R := (F, A_1,\ \dots\ , A_{\mathcal{Q}_a})$$

It allows us to write the inner product representation of $\hat{r}_\delta$:

$$\langle \hat{r}_\delta(\mu) | v_\delta \rangle_{\mathbb{V}} = r(v_\delta\ ;\mu) = \sum_{q=1}^{\mathcal{Q}_r} r_q(\mu) R_q(v_\delta) \quad \forall v_\delta \in \mathbb{V}_\delta$$

Since R_q is a form on $\mathbb{V}_\delta$, Riesz 6.5 ensures it exists $\hat{r}_\delta^q$ for each $1 \le q \le \mathcal{Q}_r$ such that

$$R_q(v_\delta) = \langle \hat{r}_\delta^q | v_\delta \rangle$$

Hence

$$\hat{r}_\delta = \sum_{q=1}^{\mathcal{Q}_r} r_q(\mu) \hat{r}_\delta^q$$

and

$$\|\hat{r}_\delta\|_{\mathbb{V}}^2 = \sum_{q,q'=1}^{\mathcal{Q}_r} r_q(\mu) r_{q'}(\mu) \left\langle \hat{r}_\delta^q \middle| \hat{r}_\delta^{q'} \right\rangle_{\mathbb{V}}$$

4) Computation of a lower bound of the Stability Constant

The goal here is to provide a computable lower bound of the stability constant $\alpha(\mu)$.
Recall the definition of the discrete coercivity constant:

$$\alpha_\delta(\mu) = \inf_{v_\delta \in \mathbb{V}_\delta} \frac{a(v_\delta, v_\delta\,;\mu)}{\|v_\delta\|_{\mathbb{V}}^2}$$

Theorema 2.20: Representation of SP(D) bilinear forms

$a : \mathbb{V}_\delta \times \mathbb{V}_\delta \times \mathbb{P} \to \mathbb{R}$ being a symmetric positive (definite) bilinear form, there exists a SP(D) matrix A such that

$$a(x, y) = \langle Ax | y \rangle_{\mathbb{V}}$$

$\mathcal{D}$

We know $\mathcal{L}(E, \mathcal{L}(F, \mathbb{R})) \equiv \mathcal{L}(E \times F, \mathbb{R})$.
Claim : A is SPD.

$$A_{i,j} = \langle AE_i | E_j \rangle = a(e_i, e_j)$$

where e_i is a basis of $\mathbb{V}_\delta$ and E_i the basis vectors.
Since a is SPD, A is also SPD. □

Since A is SPD, there exists $\lambda_1, \dots, \lambda_{N_\delta} \in \mathbb{R}_+$,$w_i \in \mathbb{V}_\delta$ orthonormal basis, such that $Aw_i = \lambda_i w_i$.
One can verify that λ eigenvalue of A if and only if :

$$a(w_i, v_\delta) = \lambda_i \langle w_i | v_\delta \rangle_{\mathbb{V}} \, \forall v_\delta \in \mathbb{V}_\delta$$

Proposition 2.21: Eugenvalue problem

The coercive constant is such that $\alpha_\delta = \inf \{\lambda \;/\; \exists w,\ a(w, v_\delta) = \lambda \langle w | v_\delta \rangle_{\mathbb{V}}, \forall v_\delta \in \mathbb{V}_\delta\}$

$\mathcal{D}$

Define $\lambda_{\min} = \inf \{\lambda \;/\; \exists w,\ a(w, v_\delta) = \lambda \langle w | v_\delta \rangle_{\mathbb{V}}, \forall v_\delta \in \mathbb{V}_\delta\}$ the smallest eigenvalue of A.
Recall $\lambda_{\min} \geq 0$ because A is positive.
Take $v \in \mathbb{V}_\delta$. Then $v = \sum_{i=1}^{N_\delta} v_i w_i$.

$$a(v, v) = \sum_{i=1}^{N_\delta} \lambda_i v_i^2 \geq \lambda_{\min} \|v\|_{\mathbb{V}}^2$$

Hence

$$\forall v \in \mathbb{V}_\delta, \ \frac{a(v, v)}{\|v\|_{\mathbb{V}}^2} \geq \lambda_{\min}$$

The right quantity is independant of v, we can take the infimum

$$\alpha_\delta \geq \lambda_{\min}$$

Then take w such that $Aw = \lambda_{\min} w$, by the very definition of the coercive constant :

$$\frac{a(w,w)}{\|w\|_{\mathbb{V}}^2} = \lambda_{\min} \geq \alpha_\delta$$

Hence the equality. □

For computation, one can rewrite the equation p.20 as

$$\text{Find smallest } \lambda > 0 \text{ such that } \mathbf{A}_\delta^\mu w_\delta = \lambda \mathbf{M}_\delta w_\delta$$

where $\mathbf{A}_\delta^\mu = (a(\varphi_i, \varphi_j\ ;\mu)_{i,j}$ and $\mathbf{M}_\delta = (\langle \varphi_i | \varphi_j \rangle_{\mathbb{V}})_{i,j}$.
It allows fast computation of the constant.

a) Min-θ-approach

Recall the affine assumption :

$$a(u,v\ ;\mu) = \sum_{q=1}^{\mathcal{Q}_a} \theta_a^q(\mu) a_q(u,v)$$

Definition 2.22: Parametrically coercive problem

- $\theta_a^q(\mu) > 0$
- a_q is positive semi-definite

Assume there exists μ' such that we computed $\alpha_\delta(\mu')$. Then

$$\begin{aligned}
a_\delta(\mu) &= \inf_{v \in \mathbb{V}_\delta} \frac{a(v,v)}{\|v\|_{\mathbb{V}}^2} \\
&= \inf_{v \in \mathbb{V}_\delta} \sum_{q=1}^{\mathcal{Q}_a} \theta_a^q(\mu) \frac{a_q(v,v)}{\|v\|_{\mathbb{V}}^2} \\
&= \inf_{v \in \mathbb{V}_\delta} \sum_{q=1}^{\mathcal{Q}_a} \frac{\theta_a^q(\mu)}{\theta_a^q(\mu')} \theta_a^q(\mu') \frac{a_q(v,v)}{\|v\|_{\mathbb{V}}^2} \\
&\geq \inf_{v \in \mathbb{V}_\delta} \min_{q=1,\dots,\mathcal{Q}_a} \frac{\theta_a^q(\mu)}{\theta_a^q(\mu')} \sum_{q=1}^{\mathcal{Q}_a} \theta_a^q(\mu') \frac{a_q(v,v)}{\|v\|_{\mathbb{V}}^2} \\
&= \underbrace{\alpha_\delta(\mu') \min_{q=1,\dots,\mathcal{Q}_a} \frac{\theta_a^q(\mu)}{\theta_a^q(\mu')}}_{\alpha_{\mathrm{LB}}}
\end{aligned}$$

Then assuming we computed $\alpha_\delta(\mu_1), \dots, \alpha_\delta(\mu_M)$ stability constants.
A sharper lower bound is

$$\alpha_{\mathrm{LB}}(\mu) := \max_{m=1,..,M} \alpha_{\mathrm{LB}}(\mu_m)$$

This is an expensive approach that is used in practice because very easy to compute, and we are willing to lose time during the offline mode (when we compute the reduced basis) in order to have the best online mode (when we compute the solution of the reduced basis for a new given parameter).

b) Successive Constraint Method

We present this other method that is faster but more complex. It is based on a functional analysis approach. The goal is to minimise the following functional :

$$\begin{array}{rccc} \mathbf{S}: & \mathbf{P} \times \mathbf{R}^{\mathcal{Q}_a} & \longrightarrow & \mathbf{R} \\ & (\mu, y) & \longmapsto & \sum_{q=1}^{\mathcal{Q}_a} \theta_a^q(\mu) y_q \end{array}$$

over the set of admissible solutions

$$\mathcal{Y} := \left\{ y = (y_1, \dots, y_{\mathcal{Q}_a}) \in \mathbf{R}^{\mathcal{Q}_a} \;\middle/\; \exists v_\delta, \forall q,\ y_q = \frac{a_q(v_\delta, v_\delta)}{\|v_\delta\|_{\mathbb{V}}^2} \right\}$$

Remark :

It is equivalent to define:

$$\mathcal{Y} := \left\{ y = \left(\frac{a_1(v_\delta, v_\delta)}{\|v_\delta\|_{\mathbb{V}}^2}, \dots, \frac{a_{\mathcal{Q}_a}(v_\delta, v_\delta)}{\|v_\delta\|_{\mathbb{V}}^2} \right) \in \mathbf{R}^{\mathcal{Q}_a} \;\middle/\; v_\delta \in \mathbb{V}_\delta \right\}$$

Then there holds

$$\alpha_\delta(\mu) = \min_{y \in \mathcal{Y}} \mathbf{S}(\mu, y)$$

To find α_{LB} and α_{UB} we search for $\mathcal{Y}_{\text{UB}} \subset \mathcal{Y} \subset \mathcal{Y}_{\text{LB}}$.
Then

$$\alpha_{\text{LB}}(\mu) := \min_{y \in \mathcal{Y}_{\text{LB}}} \mathbf{S}(\mu, y), \text{ and } \alpha_{\text{UB}}(\mu) := \min_{y \in \mathcal{Y}_{\text{UB}}} \mathbf{S}(\mu, y)$$

Let $\mathbb{P}_a \subset \mathbb{P}$ such that $\mu_1, \mu_2 \in \mathbb{P}_a, \mu_1 \neq \mu_2 \Rightarrow a(\cdot, \cdot\ ; \mu_1) \neq a(\cdot, \cdot\ ; \mu_2)$. Set $\Theta_a \subset \mathbb{P}_a$ be representative discrete point-set of $\mathbb{P}_a$.
Define $\mathcal{B} = \prod_{q=1}^{\mathcal{Q}_a} [\sigma_q^-; \sigma_q^+]$ where

$$\sigma_q^- = \inf_{v_\delta \in \mathbb{V}_\delta} \frac{a_q(v_\delta, v_\delta)}{\|v_\delta\|_{\mathbb{V}}^2}, \text{ and } \sigma_q^+ = \sup_{v_\delta \in \mathbb{V}_\delta} \frac{a_q(v_\delta, v_\delta)}{\|v_\delta\|_{\mathbb{V}}^2}$$

We get directly $\mathcal{Y} \subset \mathcal{B}$.
Define

$$\mathbb{P}_M(\mu\ ; E) := \begin{cases} M \text{ closest points to } \mu \text{ in } E & \text{if Card } (E) > M \\ E & \text{if Card } (E) \leq M \end{cases}$$

For $n = 1$:
Set $\alpha_{\text{LB}}^0(\mu) = 0\ \forall \mu \in \Theta_a$. Take $\mu_1 \in \mathbb{P}_a$. Denote

$$\mathbb{P}_1 = \{\mu_1\}$$

Solve the Eigenvalue problem p.20 for μ_1 which gives $(\alpha_\delta(\mu), w_\delta^1)$.
Define y^1 such that

$$(y^1)_q = \frac{a_q(w_\delta^1, w_\delta^1)}{\|w_\delta^1\|_{\mathbb{V}}^2}$$

We then define

$$\mathcal{Y}_{\mathrm{UB}}^1 := \{y^1\}$$

and

$$\mathcal{Y}_{\mathrm{LB}}^1(\mu) = \left\{ y \in \mathcal{B} \;\middle/\; \begin{array}{ll} \mathbf{S}(\mu', y) \geq \alpha_\delta(\mu'), & \forall \mu' \in \mathbb{P}_{M_e}(\mu\,; \mathbb{P}_1) \\ \mathbf{S}(\mu', y) \geq \alpha_{\mathrm{LB}}^0(\mu') = 0, & \forall \mu' \in \mathbb{P}_{M_p}(\mu\,; \Theta_a \backslash \mathbb{P}_1) \end{array} \right\}$$

Finally we define

$$\alpha_{\mathrm{LB}}^1(\mu) = \min_{y \in \mathcal{Y}_{\mathrm{LB}}^1(\mu)} \mathbf{S}(\mu, y), \text{ and } \alpha_{\mathrm{UB}}^1(\mu) = \min_{y \in \mathcal{Y}_{\mathrm{UB}}^1} \mathbf{S}(\mu, y)$$

Set $n \geq 1$.
Suppose we constructed $\mathbb{P}_{n-1} = \{\mu_1, \ldots, \mu_{n-1}\} \subset \mathbb{P}_a$ and $\alpha_{\mathrm{LB}}^{n-1}(\mu) > 0 \; \forall \mu \in \Theta_a$.
Define for each $\mu \in \mathbb{P}$

$$\eta(\mu, \mathbb{P}_{n-1}) = 1 - \frac{\alpha_{\mathrm{LB}}^{n-1}}{\alpha_{\mathrm{UB}}^{n-1}}$$

Choose μ_n such that

$$\mu_n := \arg\max_{\mu \in \mathbb{P}} \eta(\mu, \mathbb{P}_{n-1})$$

Then set

$$\mathbb{P}_n := \mathbb{P}_{n-1} \cup \{\mu_n\}$$

Solve the Eigenvalue problem p.20 which gives $\alpha_\delta(\mu^n)$ and y^n. As before, define

$$\mathcal{Y}_{\mathrm{UB}}^n := \{y^1, \ldots, y^n\}, \text{ and } \mathcal{Y}_{\mathrm{LB}}^n(\mu) = \left\{ y \in \mathcal{B} \;\middle/\; \begin{array}{ll} \mathbf{S}(\mu', y) \geq \alpha_\delta(\mu'), & \forall \mu' \in \mathbb{P}_{M_e}(\mu\,; \mathbb{P}_n) \\ \mathbf{S}(\mu', y) \geq \alpha_{\mathrm{LB}}^{n-1}(\mu'), & \forall \mu' \in \mathbb{P}_{M_p}(\mu\,; \Theta_a \backslash \mathbb{P}_n) \end{array} \right\}$$

We can prove "A successive constraint linear optimization method for lower bounds of parametric coercivity and inf–sup stability constants" [Huy+07] :

$$\mathcal{Y}_{\mathrm{UB}} \subset \mathcal{Y} \subset \mathcal{Y}_{\mathrm{LB}}(\mu) \; \forall \mu$$

and the construction of such sets make them naturally increasing for inclusion (resp. decreasing).

Once n_0 fixed by a tol over the estimator η, the Online Procedure goes as follow :

$$\alpha_{\mathrm{LB}}(\mu) = \min_{y \in \mathcal{Y}_{\mathrm{LB}}(\mu)} \mathbf{S}(\mu, y)$$

where $\mathcal{Y}_{\mathrm{LB}}(\mu) := \mathcal{Y}_{\mathrm{LB}}^{n_0}(\mu)$.

5) Online and Offline computation

We know how to :
Compute $\alpha_{\text{LB}}(\mu)$.
Compute $\|\hat{r}_\delta(\mu)\|_{\mathbb{V}}$.
To use the ROM with the Greedy Algorithm we proceed as follow:

- **Offline mode**:
  Estimate $\eta_{\text{en}}(\mu)$ $\forall \mu \in \mathbb{P}_h$. Following p.9, add the worst estimate solution to the basis.
  Do it until $\eta_{\text{en}}(\mu)$ is lesser than a set tolerance.
- **Online mode**:
  For a new $\mu \in \mathbb{P}$, compute all the $\theta_a^q(\mu), \theta_f^q(\mu), \theta_\ell^q(\mu)$.
  Solve the reduced basis linear system.

The hope is that the Online Mode will be significantly faster.

Part 3
Finite Volumes Method

The Finite Volume Method in Fluid Dynamics [MMD16]
I will only consider the cell centered FVM.

I - Integral of Finite Volume

We consider the General Conservation Equation for a scalar quantity ϕ in a fluid

$$\partial_t(\rho\phi) + \nabla \cdot (\rho v\phi) = \nabla \cdot (\Gamma \nabla \phi) + Q$$

Suppose the steady-state

$$\nabla \cdot (\rho v\phi) = \nabla \cdot (\Gamma \nabla \phi) + Q$$

Integrate in a Control Volume

$$\int_{V_C} \nabla \cdot (\rho v\phi) = \int_{V_C} \nabla \cdot (\Gamma \nabla \phi) + \int_{V_C} Q$$

Using Green-Ostrogradsky on the gradients

$$\int_{\partial V_C} (\rho v\phi - \Gamma \nabla \phi) \cdot dS = \int_{V_C} Q$$

Consider the integrand as the sum of the integrand on each faces

$$\int_{\partial V_C} (\rho v\phi - \Gamma \nabla \phi) \cdot dS = \sum_f \int_f (\rho v\phi - \Gamma \nabla \phi) \cdot dS_f$$

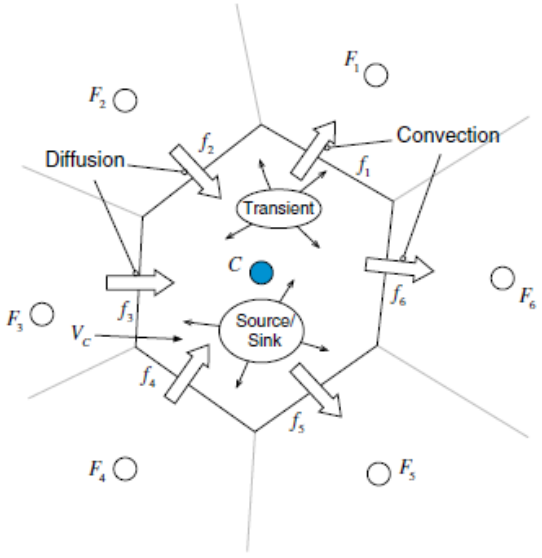


Figure 5: Conservation in a controle volume, from [MMD16]

From here, the Gauss-Legendre quadrature can be used to prove mathematically that we can approximate efficiently the integrand on the faces by a sum. Using one point, it is equivalent to the Mean Value Theorem for Integrals.

Theorema 3.1: Mean Value Theorem for Integrals

Consider $\Omega \subset \mathbb{R}^n$, $n \geq 1$, measurable with finite measure, connected set.
Consider $f : \Omega \to \mathbb{R}$ a function which is

- Continuous
- Integrable
- Bounded

Then it exists $c \in \Omega$ such that

$$f(c) = \frac{1}{\text{mes}(\Omega)} \int_\Omega f(x)dx$$

$\mathcal{D}$

Since f is bounded on the set, it exists m,M such that

$$m \leq f \leq M$$

One can integrate this inequality

$$\text{mes}(\Omega)m \leq \int_\Omega f \leq \text{mes}(\Omega)M \iff m \leq \frac{1}{\text{mes}(\Omega)} \int_\Omega f \leq M$$

Then by Intermediate Value Theorem for real valued functions it exists $c \in \mathring{\Omega}$ such that

$$f(c) = \frac{1}{\text{mes}(\Omega)} \int_\Omega f$$

□

One can interpret this equation such that the mean of f over the set is attained by one point c in the set.

Proposition 3.2: Centroid

Suppose f is linear and Ω convex.
Then the centroid of the considered set $c = \frac{\int_\Omega x dx}{\text{mes}(\Omega)} \in \Omega$ is such that

$$f(c) = \frac{1}{\text{mes}(\Omega)} \int_\Omega f$$

$\mathcal{D}$

$\frac{\int_\Omega x dx}{\text{mes}(\Omega)} \in \Omega$ comes by convexity.

Search $c_0 \in \Omega$ such that

$$\begin{aligned} f(c_0) &= \frac{1}{\text{mes}(\Omega)} \int_\Omega f(x)dx \\ &= \frac{1}{\text{mes}(\Omega)} f\left(\int_\Omega x dx\right) \text{ by linearity} \\ &= f\left(\frac{1}{\text{mes}(\Omega)} \int_\Omega x dx\right) \end{aligned}$$

Then $c_0 = \frac{1}{\text{mes}(\Omega)} \int_\Omega x dx$ is a solution to the equation.
Uniqueness is not true in general. □

Coming back to

$$\int_{\partial V_C} (\rho v \phi - \Gamma \nabla \phi) \cdot dS = \sum_f \int_f (\rho v \phi - \Gamma \nabla \phi)_f \cdot dS_f$$

We suppose the grid thin enough to approximate linearity, hence using the centroid approximation 3.2

$$\int_{\partial V_C} (\rho v \phi - \Gamma \nabla \phi) \cdot dS \simeq \sum_f (\rho v \phi - \Gamma \nabla \phi)_f \cdot S_f$$

and similarly

$$\int_{V_C} Q \simeq Q_C V_C$$

Hence the discretized equation for each cell :

Definition 3.3: Discretized Conservation Equation

$$\sum_{f \sim \text{faces}(C)} (\rho v \phi - \Gamma \nabla \phi)_f \cdot S_f = Q_C V_C$$

Then suppose that we can linearise the flux

$$(\rho v \phi - \Gamma \nabla \phi)_f \cdot S_f = \text{Flux} C_f \phi_C + \text{Flux} F_f \phi_f + \text{Flux} V_f$$

and the source

$$Q_C V_C = \text{Flux} C \phi_C + \text{Flux} V$$

Then it is easy to write the equation such that

$$a_C \phi_C + \sum_{f \sim \text{faces}(C)} a_F \phi_F = b_C$$

with a_C,a_F,b_C depending on the $\text{Flux} X_{f/C}$.

II - Linearisation of the discretised equation

1) Linearisation of the diffusion flux

In the general case, the grid and the centroids don't have any reason to have create an orthogonal mesh.

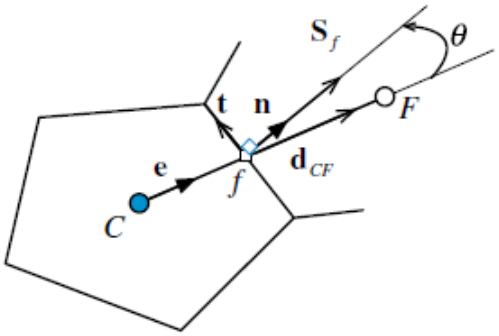


Figure 6: Non orthogonal mesh, from [MMD16]

We want to approximate $J_D = -\Gamma\nabla\phi$ as a linear function of ϕ_C and ϕ_F. The orthogonal situation would be

$$\nabla\phi = \frac{\partial\phi}{\partial n} \simeq \frac{\phi_F - \phi_C}{\|d\|}\mathbf{n}$$

and

$$(\nabla\phi)_f \cdot \mathbf{S_f} \simeq \frac{\phi_F - \phi_C}{\|d\|}S_f$$

with $S_f = S_f\mathbf{n}$.
In the non-orthogonal case :

$$\nabla\phi = \frac{\partial\phi}{\partial e} \simeq \frac{\phi_F - \phi_C}{\|d\|}\mathbf{e} + ((\nabla\phi)\cdot\mathbf{t})\mathbf{t}$$

$$\mathbf{S}_f = E_f\mathbf{e} + T_f\mathbf{t} = \mathbf{E}_f + \mathbf{T}_f$$

Hence

$$(\nabla\phi)_f \cdot \mathbf{S}_f = \frac{\phi_F - \phi_C}{\|d\|}E_f + (\nabla\phi)_f \cdot \mathbf{T}_f$$

The choice of $\mathbf{E}_f$ and $\mathbf{T}_f$ is not discussed.
We need to compute $(\nabla\phi)_f$:

$$\nabla\phi_f = g_C\nabla\phi_C + g_F\nabla\phi_F$$

where $g_C + g_F = 1$ are geometric interpolation factors with respect to F and C (coefficients of the barycenter).

2) Implicit computation of $(\nabla\phi)_f$

a) Green-Gauss gradient

$$\nabla\phi_C = \frac{1}{V_C}\sum_f \phi_f \mathbf{S}_f$$

It is still needed to compute ϕ_f.
A simple and natural way is

$$\phi_f = g_c\phi_C + g_F\phi_F$$

F being the centroid of the neighbour cell that shares the face. Another more accurate way is to compute a mean based on the vertices and F.
Both way are just using convex combination of neighbour cells.

b) Least-square distance

Considering the 1st order approxmiation

$$\phi_F = \phi_C + (\nabla\phi)_C \cdot \mathbf{r}_{\mathrm{CF}}$$

We want to minimize the quantity

$$\begin{aligned} G_C &= \sum_{k=1}^{\mathrm{NB}(C)} \left(w_k \left[\phi_F - (\phi_C + \nabla\phi_C \cdot \mathbf{r}_{\mathrm{CF}})\right]^2\right) \\ &= \sum_{k=1}^{\mathrm{NB}(C)} \left(w_k \left[\Delta\phi_k - \Delta x_k \left(\frac{\partial\phi}{\partial x}\right)_C - \Delta y_k \left(\frac{\partial\phi}{\partial y}\right)_C - \Delta z_k \left(\frac{\partial\phi}{\partial z}\right)_C\right]^2\right) \end{aligned}$$

where $\Delta X = X_F - X_C$.

Remark :

> It's the squared error on each faces that we minimze with regard to the unknown coefficients of the gradient.

This quantity is minimised when

$$\begin{aligned} \frac{\partial G_C}{\partial\left(\frac{\partial\phi}{\partial x}\right)} &= 0 \\ \frac{\partial G_C}{\partial\left(\frac{\partial\phi}{\partial y}\right)} &= 0 \\ \frac{\partial G_C}{\partial\left(\frac{\partial\phi}{\partial z}\right)} &= 0 \end{aligned}$$

which is equivalent to

$$\sum_{k=1}^{\mathrm{NB}(C)} \left(2\Delta x_k w_k \left[-\Delta\phi_k + \Delta x_k \left(\frac{\partial\phi}{\partial x}\right)_C + \Delta y_k \left(\frac{\partial\phi}{\partial y}\right)_C + \Delta z_k \left(\frac{\partial\phi}{\partial z}\right)_C \right] \right) = 0$$
$$\sum_{k=1}^{\mathrm{NB}(C)} \left(2\Delta y_k w_k \left[-\Delta\phi_k + \Delta x_k \left(\frac{\partial\phi}{\partial x}\right)_C + \Delta y_k \left(\frac{\partial\phi}{\partial y}\right)_C + \Delta z_k \left(\frac{\partial\phi}{\partial z}\right)_C \right] \right) = 0$$
$$\sum_{k=1}^{\mathrm{NB}(C)} \left(2\Delta z_k w_k \left[-\Delta\phi_k + \Delta x_k \left(\frac{\partial\phi}{\partial x}\right)_C + \Delta y_k \left(\frac{\partial\phi}{\partial y}\right)_C + \Delta z_k \left(\frac{\partial\phi}{\partial z}\right)_C \right] \right) = 0$$

that can also be written under the follozing form :

$$A(\nabla\phi)_C = b$$

The choice of w_k has to be discussed. It can be a constant or depend on the inverse to the distance $\frac{1}{r_{CF}^n}$ on any power $n \geq 1$.
This equation also gives the usual solution of 1st order gradient for the cartesian grid

$$\partial_x\phi \simeq \frac{\phi_F - \phi_C}{x_F - x_C}$$

We also can write the linear equation first and realize the G_C quantity appears :

$$\phi_F = \phi_C + (\nabla\phi)_C \cdot \mathbf{r}_{\mathrm{CF}} \quad \forall F \iff \mathbf{R}\nabla\phi_C = [\phi_N - \phi_C]$$

where $\mathbf{R}$ is a $N \times 3$ matrix of the $r_{CF_i,j}$, $j \in \{x, y, z\}$. This equation is overdetermined, one can use the least square quantity writing

$$\nabla\phi_C = (d^T d)^{-1} d^T [\phi_N - \phi_C]$$

and the least-square method makes G_C appears naturally.

3) Gradient on faces

Once we computed the gradient on centroids, we can approximate the gradient on faces :

$$\overline{\nabla\phi_f} = g_C\nabla\phi_C + g_F\nabla\phi_F$$

and consider

$$\nabla\phi_f = \overline{\nabla\phi_f} + \underbrace{\left(\frac{\phi_F - \phi_C}{d_{CF}} - \overline{\nabla\phi_f} \cdot \mathbf{e}_{CF} \right) \mathbf{e}_{CF}}_{\text{Correction interpolated face gradient}}$$

where

$$d_{CF} = |r_F - r_C|$$
$$\mathbf{e}_{CF} = \mathbf{r}_F - \mathbf{r}_C$$

4) Convection flux and source term

We admit the linearisation process of these terms.
Source term will be admitted to be constant or at least independent of the solution.

III - Error estimation of full order FV

"Error Analysis and Estimation for the Finite Volume Method With Applications to Fluid Flows" "Error Analysis and Estimation for the Finite Volume Method With Applications to Fluid Flows"

1) Taylor Extension estimates

Consider u the exact solution, hoping u is smooth.
Then one could write the Taylor Expansion

$$u(x) = \sum_{n=0}^{\infty} \frac{1}{n!}(x - x_C)^n \otimes^n (\nabla^n u)_C$$

A discretisation is of p-order if one approximate u such that

$$\phi(x) = \sum_{n=0}^{p-1} \frac{1}{n!}(x - x_C)^n \otimes^n (\nabla^n \phi)_C$$

And put that into the equation.
The discretisation error is then

$$e(x) = \sum_{n=p}^{\infty} \frac{1}{n!}(x - x_P)^n \otimes^n (\nabla^n \phi)_C$$

The error on the control volume is introduced as

$$e_t(\phi) = \left\| \frac{1}{V_C} \int_{V_C} e(x)dx \right\| \leq \frac{1}{V_C} \sum_{n=p}^{\infty} \left\| \int_{V_C} \frac{1}{n!}(x - x_C)^n \otimes^n (\nabla^n \phi)_C \right\|$$

The error estimate can be then defined

$$E(\phi) = \frac{1}{V_C} \sum_{n=p}^{\infty} \left\| \int_{V_C} \frac{1}{n!}(x - x_C)^n \otimes^n (\nabla^n \phi)_C \right\| = \frac{1}{V_C} \sum_{n=p}^{\infty} \frac{1}{n!} \left\| \left[\int_{V_C} (x - x_C)^n \right] \otimes^n (\nabla^n \phi)_C \right\|$$

which can't be computed. One use a more computable estimate

Definition 3.4: Taylor error estimate

$$e_t(\phi) = \frac{1}{V_C} \frac{1}{p!} \left\| \left[\int_{V_C} (x - x_C)^n \right] \otimes^n (\nabla^n \phi)_C \right\|$$

This error estimate really surprise me since it's not greater than the exact error.
For the FVM, the discretisation is of order 2.

$$e_t(\phi) = \frac{1}{V_C} \cdot \frac{1}{2} \left\| \mathbf{M} \otimes (\nabla^2 \phi)_C \right\|$$

where $\mathbf{M} = \int_{V_C} (x - x_C)^2 dV$ is the order 2 geometric moment of the control volume.
The hope is that when the function is smooth, the n-th value tend quickly to 0.
One have to add the diffusion numerical error to the truncated error, where T is the caracteristic time length.

$$e_{\text{num}} = \frac{\left\| \int_{V_C} \nabla(\cdot \Gamma_{\text{num}} \nabla \phi) \right\| T}{V_C}$$

2) Moment Estimates

We consider the steady-state scalar transport equation

$$\nabla\cdot(\rho\mathbf{U}\phi) - \nabla\cdot(\rho\Gamma_\phi\nabla\phi) = S_\phi(\phi) = \underbrace{\mathrm{Sp}}_{\text{non linear part}} + \underbrace{\mathrm{Su}\phi}_{\text{linear part}}$$

The solution verifies all the higher moments equations.
We consider the 2nd moment :

$$m = \frac{1}{2}\phi^2$$

It holds

$$\begin{aligned}
\underbrace{\nabla\cdot(\rho\mathbf{U}m)}_{=\frac{1}{2}\nabla(\rho\mathbf{U}\phi)+\frac{1}{2}\rho\mathbf{U}\cdot\nabla\phi)} - \nabla\cdot(\rho\Gamma_\phi\underbrace{\nabla m}_{=\phi\nabla\phi}) &= \frac{1}{2}\nabla\cdot(\rho\mathbf{U}\phi)\phi + \frac{1}{2}\rho\mathbf{U}\phi\cdot\nabla\phi - \underbrace{\nabla\cdot(\rho\Gamma_\phi\phi\nabla\phi)}\\
&= \underbrace{\frac{1}{2}\nabla\cdot(\rho\mathbf{U}\phi)\phi}_{=-\frac{1}{2}\nabla(\rho\mathbf{U}\phi)\phi+\nabla(\rho\mathbf{U}\phi)\phi} + \frac{1}{2}\rho\mathbf{U}\cdot\phi\nabla\phi - \nabla\cdot(\rho\Gamma_\phi\nabla\phi)\phi - \rho\Gamma_\phi(\nabla\phi\cdot\nabla\phi)\\
&= -\frac{1}{2}\nabla\cdot(\rho\mathbf{U}\phi)\phi + \frac{1}{2}\rho\mathbf{U}\phi\cdot\nabla\phi + \underbrace{(\nabla\cdot(\rho\mathbf{U}\phi) - \nabla\cdot(\rho\Gamma_\phi\nabla\phi))}_{=S_\phi(\phi)}\phi\\
&\quad - \rho\Gamma_\phi(\nabla\phi\cdot\nabla\phi)\\
&= -\frac{1}{2}\left(\nabla\cdot(\rho\mathbf{U}\phi) - \rho\mathbf{U}\cdot\nabla\phi\right)\phi + S_\phi(\phi)\phi - \rho\Gamma_\phi(\nabla\phi\cdot\nabla\phi)\\
&= -\frac{1}{2}\underbrace{\nabla\cdot(\rho\mathbf{U})}_{=0}\phi + S_\phi(\phi)\phi - \rho\Gamma_\phi(\nabla\phi\cdot\nabla\phi) \qquad (\star)\\
&= S_\phi(\phi)\phi - \rho\Gamma_\phi(\nabla\phi\cdot\nabla\phi)
\end{aligned}$$

($\star$) holds for a steady-state equation with no source term.

Since ϕ is an approximation, one can define the *local imbalance* denoted *res* such that

$$res_m(m_\phi) = \int_{\Omega_C}\left[\nabla(\rho\mathbf{U}m) - \nabla(\rho\Gamma_\phi\nabla m) - (S_\phi(\phi)\phi - \rho\Gamma_\phi(\nabla\phi\cdot\nabla\phi))\right]dV$$

Then considering the steady-state transport equation for a general vector property $\mathbf{a}$:

$$\nabla\cdot(\rho\mathbf{U}\mathbf{a}) - \nabla(\rho\Gamma_a\nabla\mathbf{a}) = \mathbf{Su} + Sp\ \mathbf{a}$$

and define

$$m_a = \frac{1}{2}\mathbf{a}\cdot\mathbf{a}$$

In the same way, we can obtain the equation

$$\nabla(\rho\mathbf{U}m_a) - \nabla(\rho\Gamma_a\nabla m_a) = \mathbf{Su}\cdot\mathbf{a} + 2Sp\ m_a - \rho\Gamma_a(\nabla\mathbf{a}\otimes\nabla\mathbf{a})$$

and the *local imbalance*

$$res_m(m_a) = \int_{\Omega_C}\left[\nabla(\rho\mathbf{U}m_a) - \nabla(\rho\Gamma_a\nabla m_a) - (\mathbf{Su}\cdot\mathbf{a} + 2Sp\ m_a - \rho\Gamma_a(\nabla\mathbf{a}\otimes\nabla\mathbf{a}))\right]dV$$

Note that everything is scalar.
res has a defined dimension $[\phi]^2[L]^3/[T]$ and we need to normalise it : define the caracteristic time T such that

$$T = \frac{h}{U_{\text{transport}}}$$

where

$$U_{\text{transport}} = \|\mathbf{U}\| + \frac{\Gamma}{h}$$

hence we define the error on the cell

$$e_m(\phi) = 2\sqrt{\frac{\|res_m(m_\phi)\|\, T}{V_C}}$$

3) Residual Estimates

"The residual is a function that measures how well the local solution satisfies the original governing equations. It is therefore natural to associate the level of residual with the local solution error."
The following choice is made in the definition of the residual estimates :

$$\begin{aligned} res_C &= \int_{V_C} [\nabla\cdot(\rho\mathbf{U}\phi) - \nabla\cdot(\rho\gamma_\phi\nabla\phi) - Su - Sp\phi_C]dV \\ &= \sum_f [(\rho\mathbf{U}\phi)_f - (\rho\gamma_\phi)_f(\nabla\phi)_f]\cdot\mathbf{S}_f - \mathrm{Su}V_C - \mathrm{Sp}\phi_C V_C \end{aligned}$$

One could have worked on the estimate error for the faces.
We recall that

$$\begin{aligned} \phi_f &= \phi_C + (x_f - x_C)\cdot\nabla\phi_C \\ (\nabla\phi)_f &= \nabla\phi_C \end{aligned}$$

This expression dimension is $[\phi][L]^3/[T]$ thus need to be normalised.
It is built on the caracteristic diffusion and convection.

$$\begin{aligned} F_{\text{diff}} &= \frac{1}{V_C}\sum_f \left[\|\mathbf{S}\|\frac{(\rho\gamma_\phi)_f}{\|\mathbf{d}\|}\right] \\ F_{\text{conv}} &= \frac{1}{V_C}\sum_f \max(F,0) \\ F_{\text{norm}} &= F_{\text{diff}} + F_{\text{conv}} + \mathrm{Sp} \end{aligned}$$

Hence

$$e_r(\phi) = \frac{res[\phi)]}{V_C F_{\text{norm}}}$$

IV - Issue of the FVM and its estimates

The main issue of the FVM is that there is no such thing as "basis functions" : we do not create the solution by using discrete spaces. There is not much mathematical framework other

than classical approximations.
The only basis function that create the method is piecewise constant fonctions on each control volume.

The second flaw is the estimates that are not standards nor mathematically well established.

That is why we try to see the FVM as a derivation of FEM with piecewise constant fonctions approximation instead of piecewise linear.

Part 4
Discontinous Galerkin Method

I - Theoretical aspects

Following *Mathematical aspects of discontinuous Galerkin method* [PE12]

1) Definitions

Definition 4.1: Petrov-Galerkin approximation

$a(\cdot,\cdot) : V \times W \to \mathbb{R}$, V, W Hilberts.
$\mathbf{f} \in \mathcal{L}(W, \mathbf{R})$.

$$(\star\star) : \quad \text{find } u \in V, \quad a(u,v) = \mathbf{f}(v) \quad \forall v \in W$$

One can invoke Banach–Nečas–Babuška theorem 6.7 to ensure well posedness.

Definition 4.2: Jump and average

Consider a finite element T_1 of and its interface ∂T of dimension $dim(T) - 1$ with another finite element T_2.
We define the componentwise average of v on ∂T as

$$\{\!\{v\}\!\}(x) := \frac{1}{2}\left(v|_{T_1}(x) + v|_{T_2}(x)\right)$$

and the componentwise jump of v on ∂T as

$$[\![v]\!](x) = v|_{T_1}(x) \cdot n_1 + v|_{T_2}(x) \cdot n_2$$

where n_i is the normal vector defined by the borders.

The average is used to approximate the function on interfaces.

Definition 4.3: Discontinuous Galerkin Method

Trial basis of the Petrov-Galerkin approximation made of piecewise polynomials of certain degree : P is defined on each element T such that $P|_T$ is polynomial but $[\![P(x)]\!] \neq 0$ in general.
The test basis is also made of piecewise polynomials of certain degree.

Using the DGM, the solution $u \in V$ lives in $W^{s,p}(\mathcal{T}) := \{v \in \mathbf{L}^2(\Omega) \ / \ \forall T \in \mathcal{T}, \ u|_T \in W^{s,p}(T)\}$, where $\mathcal{T}$ is the set of finite elements.
We define the the solutions, gradients, spaces the trial basis and the tests basis as piecewise spaces that are called "broken whose objects are defined on each elements.

2) Equivalence with FEM

This equivalence is direct since one just have to add a continuity condition on the interfaces i.e. $[\![v]\!] \equiv 0$ for all considered functions and chose trial and test functions in $\mathbb{V}_{\mathcal{T}}$ polynomials functions.

3) Equivalence with FVM

We note that FVM can be written as DGM with $\mathbf{1}_C$ as basis function. This equivalence is harder to write since you have to define the gradient reconstruction depending on which FVM method is considered.
This is a case by case equivalence that we need to investigate further.

4) Local Problem Error Estimate for FVM

[Jas96]
Let $\mathcal{L}u = -\nabla\cdot(a\nabla u) + cu$

$$\mathcal{L}u = f \qquad \in \Omega$$

with boundary conditions

$$\begin{cases} \phi = \phi_D(x) & x \in \Gamma_D \\ a\ \mathbf{n}_f \cdot \nabla\phi_f = g(x) & x \in \Gamma_N \end{cases}$$

where

$$\Gamma_D \cup \Gamma_N = \partial\Omega$$
$$\Gamma_N \cap \Gamma_D = \emptyset$$

Let

$$a(u,v) = \int_V -\nabla\cdot(a\nabla u)v + cuv = \int_V a\nabla u\cdot\nabla v + cuv$$

Define the error

$$e = u - u_h$$

$$\|e\|_V^2 = a(e,e) = \int_V a\nabla e\cdot\nabla e + ce^2$$

The error has the following convergence property

$$\|e\|_V \leq Ch^k$$

where k denotes the order of approximation and C is independant of k and h.

Theorema 4.4: Local problem estimate

For every control, a local error problem

$$-\nabla\cdot\nabla\psi_C = r_C \quad x \in \Omega_C$$

with boundary conditions

$$\begin{cases} \mathbf{n}_f \cdot \nabla\psi_C = R_C & x \in \partial\Omega_C\Gamma_D \\ \psi_C = 0 & x \in \partial\Omega_C \cap \Gamma_D \end{cases}$$

where

$$R_C := \begin{cases} g - a\mathbf{n}_f \nabla \cdot u_h & \text{on } \partial\Omega_C \cap \Gamma_N \\ -\alpha_C [\![a\mathbf{n}_f \nabla \cdot u_h]\!] & \text{on } \partial\Omega_C \backslash \Gamma_N \end{cases}$$

and

$$r_C = f - \mathcal{L}u$$

produce an upper bound of the error energy norm

$$\|e\|_V^2 \leq \sum_C^N \varepsilon_C^2(\nabla \psi_C)$$

where N is the number of subdomains.

It can be extended on Diffusion-Convection problems.

II - Box Method

Following "Some Error Estimates for the Box Method" [BR87] and "On the convergence of the Rhie–Chow stabilized Box method for the Stokes problem" [NPV24]

This method could be used to create an error estimator without having to compute the Box Method solution if $u_{\text{FVM}} = u_B$.

Construct a triangulation $\mathcal{T}$ and suppose there exists $\delta_0 > 0$ such that

$$\forall t \in \mathcal{T}, \delta_0 \leq \frac{k_t}{h_t}$$

where h_t, k_t are respectively the diameter of the circumscribing (resp. inscribing) circle of the triangle.
Denote by E the set of all edges of triangles in $\mathcal{T}$.

For each vertices v_i of $\mathcal{T}$, we define Ω_i the union of each triangles (and their border) that have v_i as vertex.
For each triangle t in Ω_i, one can chose a point p that will be the vertex of the control volume and define b_i as the polygonal with vertices $(p_{i,t})_t$.
One can construct the dual mesh such that there exists $\alpha > 0$

$$\forall (\Omega_i, b_i), \alpha \leq \frac{|b_i|}{|\Omega_i|} (\star)$$

The Delaunay and Volonoï dual meshes can satisfy such conditions. See Fig 1 of [NPV24]

Denote the piecewise constant polynomials space on the dual mesh

$$\mathbb{P}^0(\mathcal{B}) = \left\{ v \in \mathbf{H}^1(\Omega) \;/\; \forall b \in \mathcal{B}, \quad v\big|_b \in \mathbb{P}^0(b) \right\}$$

and the piecewise linear polynomials space on the triangulation mesh

$$\mathbb{P}^1(\mathcal{T}) = \left\{ v \in \mathbf{H}^1(\Omega) \;/\; \forall t \in \mathcal{T}, \quad v\big|_t \in \mathbb{P}^1(t) \right\}$$

Recall the broken Sobolev space

$$\mathbf{H}^1(\mathcal{B}) = \left\{ v \in \mathbf{L}^2(\Omega) \ / \ \forall b \in \mathcal{B}, \quad v\big|_b \in \mathbf{H}^1(b) \right\}$$

As for $\mathbf{H}^1_0$, for any functional space $\mathbf{X}$ we denote by $\mathbf{X}_0$ the subset of $\mathbf{X}$ whose functions are zero on $\partial\Omega$.

1) Duality map

There is a natural map between $\mathbb{P}^1(\mathcal{T})$ and $\mathbb{P}^0(\mathcal{B})$.
$\mathbb{P}^1(\mathcal{T})$ is the usual vector space of the finite elements method of nodal basis $\{\phi_i)$ where, for each node vertex v_j :

$$\phi_i(v_j) = \delta_{ij}$$

The construction of $\mathbb{P}^1(\mathcal{B})$ admits the control volume basis χ_i such that for each control volume b_i :

$$\chi_i = \mathbf{1}_{b_i}$$

The construction of $\mathcal{B}$ ensures that for each vertex v_i you can map a control volume b_i hence dim Span $\{\phi_i\}$ = dim Span $\{\chi_i\}$.
We can then define the invertible mapping :

$$\begin{array}{ccc} G : \left(\mathbb{P}^1(\mathcal{T}), \|\cdot\|_{\mathbf{H}^1(\Omega)}\right) & \longrightarrow & \left(\mathbb{P}^0(\mathcal{B}), \|\cdot\|_{\mathbf{H}^1(\mathcal{B})}\right) \\ u = \sum_i u_i \phi_i & \longmapsto & \overline{u} = \sum_i u_i \chi_i \end{array}$$

G is an isomorphism because it is surjective between two vector spaces of same dimension. Note that $\left(\mathbb{P}^1(\mathcal{T}), \|\cdot\|_{\mathbf{H}^1(\Omega)}\right)$ and $\left(\mathbb{P}^0(\mathcal{B}), \|\cdot\|_{\mathbf{H}^1(\mathcal{B})}\right)$ are Hilbert spaces.

2) Property of the map

This lemma proves that the jumps can be effectively used to approximate the gradient and also to behaves as the gradient norm. It can allow to create a piecewise constant H^1 norm.

Lemma 4.5: Semi-norm control

For $u \in \mathbb{P}^1(\mathcal{T})$, denote $\overline{u} = G(u) \in \mathbb{P}^0(\mathcal{B})$.
There exists $C_0 = C_0(\delta_0) > 0$ s.t.

$$C_0^{-1} \|\nabla u\|_{\mathbf{L}^2(\Omega)} \leq \left(\sum_{e \in E} \|[\![\overline{u}]\!]_e\|^2 \right)^{1/2} \leq C_0 \|\nabla u\|_{\mathbf{L}^2(\Omega)}$$

i.e. the Broken Sobolev semi-norm and the Sobolev semi-norm are equivalent.
Note that the norm $\|\cdot\|$ can be any norm on $\mathbb{R}^d$ where d is the space dimension.

We can also find equivalence between the piecewise norm and the $\mathbf{L}^2$ norm.

Lemma 4.6: L^2 norm control

For $u \in \mathbb{P}^1(\mathcal{T})$, denote $\overline{u} = G(u) \in \mathbb{P}^0(\mathcal{B})$.
There exists $C_1 = C_1(\delta_0) > 0$ s.t.

$$\|\overline{u}\|_{\mathbf{L}^2(\Omega)} \leq C_1(\delta_0) \|u\|_{\mathbf{L}^2}$$

Moreover suppose $(\star)$ p.37 then there exists $C_2(\delta_0, \Omega) > 0$ s.t.

$$\|u\|_{\mathbf{L}^2} \leq C_2(\delta_0, \Omega) \|\overline{u}\|_{\mathbf{L}^2(\Omega)}$$

$\mathcal{D}$

See paper for details.
See Generalized Rayleigh Quotient for the construction of C_0. We get a computable expression of C_0 which I will write later.
Same for C_1 and C_2 but the computation are more mysterious in the paper. □

Proposition 4.7: Norm equivalence

Suppose $(\star)$ p.37. Then there exists $\alpha, \beta > 0$ s.t.

$$\alpha \|u\|_{\mathbf{H}^1(\Omega)} \leq \|\overline{u}\|_{\mathbf{H}^1(\mathcal{B})} \leq \beta \|u\|_{\mathbf{H}^1(\Omega)}$$

where

$$\|\overline{u}\|_{\mathbf{H}^1(\mathcal{B})} := \|\overline{u}\|_{\mathbf{L}^2(\Omega)} + \left(\sum_{e \in E} \|[\![\overline{u}]\!]_e\|^2 \right)^{1/2}$$

$\mathcal{D}$

Applying Lemma 4.5 and Lemma 4.6 directly gives the result. □

Proposition 4.8: Integration equivalence

Let $u, v \in \mathbb{P}^1(\mathcal{T})$. Then it holds :

$$\int_\Omega \nabla u \cdot \nabla v = -\sum_{b \in \mathcal{B}} \int_{\partial b} \frac{\partial u}{\partial n} \overline{v}$$

where n is the outward pointing normal with respect to the interior of b.

3) The Poisson equation

For $f \in \mathbf{L}^2(\Omega)$ consider the equation :

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

Define $a(u,v) = \int_\Omega \nabla u \cdot \nabla v$, the weak form is written:

$$\text{Find } u \in \mathbf{H}^1_0(\Omega), \quad a(u,v) = \langle f | v \rangle_{\mathbf{L}^2}, \quad \forall v \in \mathbf{H}^1_0(\Omega)$$

Recall $\|u\|^2_\mu = a(u,u)$ and define $\|\overline{v}\|_\mu := \|G^{-1}(\overline{v})\|_\mu$.

One can adapt this equation to $\mathbf{H}^1_0(\mathcal{B})$ with Lemma 4.8 :

$$\int_\Omega \nabla u \cdot \nabla v = -\sum_{b\in\mathcal{B}} \int_{\partial b} \frac{\partial u}{\partial n}\overline{v} = -\sum_{b\in\mathcal{B}} \int_{\partial b} \frac{\partial u}{\partial n} G(v) := \overline{a}(u,v) \underbrace{\iff}_{G \text{ isomorphism}} \overline{a}(u,\overline{v}) = -\sum_{b\in\mathcal{B}} \int_{\partial b} \frac{\partial u}{\partial n}\overline{v}$$

Define the box weak form

$$\text{Find } u_B \in \mathbb{P}^1(\mathcal{T}), \quad \overline{a}(u_B,\overline{v}) = \langle f | \overline{v} \rangle_{\mathbf{L}^2}, \quad \forall v \in \mathbb{P}^0(\mathcal{B})$$

Denote by u the analytical solution and by u_L the FEM solution.

Theorema 4.9: Box error control for Poisson Problem

There exists $C = C(\delta_0, \Omega) > 0$

$$\|u - u_L\|_\mu \leq \|u - u_B\|_\mu \leq C \|u - u_L\|_\mu$$

4) Self-adjoint problem

Let $f \in \mathbf{L}^2(\Omega)$. Let Γ be a bounded real valued function. Let σ be such that $0 \leq \sigma(x) \leq \sigma^+$. Consider the equation

$$\begin{cases} -\nabla\cdot(\Gamma\nabla u) + \sigma u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

The box method weak form is

$$\text{Find } u_B \in \mathbb{P}^1(\mathcal{T}) \overline{a}(u_B,\overline{v}) + \langle \sigma \overline{u}_B | \overline{v} \rangle_{\mathbf{L}^2} = \langle f | \overline{v} \rangle_{\mathbf{L}^2} \quad \forall \overline{v} \in \mathbb{P}^0(\mathcal{B})$$

A natural generalization of the Galerkin formulation of the box method would use u_B but using $\overline{u}_B$ allows to keep the discretisation of σu diagonal and symmetric.

Theorema 4.10: Box error control for the self-adjoint problem

There exists $C = C(\Gamma, \sigma, \alpha, \delta_0) > 0$ such that :

$$\|u - u_L\|_\mu \leq \|u - u_B\|_\mu \leq C \left(\|u - u_L\|_\mu + \|u - \overline{u}_L\|_{\mathbf{L}^2(\Omega)} \right)$$

Part 5
Cell centered FVM Reduced Basis Method

The goal is to write the Finite Volume Method in the same framework as the Finite Element Method i.e. with a bilinear form. The idea is to apply the well established Certified Reduced Basis Method error estimators to a Weak Finite Volumes formulation.

It is possible to interpret FVM as a Discontinuous Galerkin Method of lowest order with the test function to be basis $V = \text{Span}(\mathbf{1}_C)$.
Since FVM searches the solution in the piecewise constant functions, we want to find a bilinear form $a(\cdot,\cdot) : \mathbb{P}^0 \times \mathbb{P}^0 \to \mathbb{R}$ with $\mathbb{P}^0$ the piecewise constant functions.
Wrote under this bilinear form, we hope to use the Dual Error Estimate 2.12 for faster reduced basis generation.

It leads naturally first to *Mathematical aspects of discontinuous Galerkin method* [PE12].
This book proposes a DGM for the Cell Centered Finite Volume Method.
Unfortunately, to construct the FVM, the book consider a reconstruction of the Gradient then look for a solution in $\mathbb{P}_1$ the piecewise linear functions.
This work still could be seen as looking for a solution in $\mathbb{P}_0$ then mapping in $\mathbb{P}^1$, but it didn't seemed direct to us.

A second book treats the problem : *The Gradient Discretisation Method* [Dro+18].
It handles the two-points flux approximation Finite Volumes on cartesian meshes and the multi-point flux approximation.
Unfortunately, reserves are given in the book about such a bilinear form.
From the same authors, the article "Analysis tools for finite volume schemes" [Eym+07] that writes explicitly the bilinear form.

This two books are very dense and hard to manipulate in a short amount of time. We decided to follow the most explicit [Eym+07] for the numerical trials. Nonetheless, both books will be considered in future research.

We also searched in the direction of the Box Method, that was introduced in [NPV24] that follow Bank and Rose (1987) [BR87]. The main issue is that we can't really define any coercivity, and we are unsure whether the estimate is interesting.

As of today, no simulations were run, and we don't know whether or not it works as expected. From a theoretical point of view, the proposed framework seems solid enough, and we firmly believe that we may have found a good starting point.

I - Box Method

See "Some Error Estimates for the Box Method" [BR87]
We consider $\mathcal{B}$ a control volume mesh.
We consider $\mathcal{T}$ a triangulation such that $\mathcal{B}$ is the dual mesh of $\mathcal{T}$.
It is possible to have two admissible primal (FEM) and dual (FVM) meshes regarding Delaunay and Volonoï.
Let $u_{\text{FVM}} \in \mathbb{P}^0(\mathcal{B})$ be the FVM approximation of the solution.

$$\textit{Suppose } u_B = G(u_{\text{FVM}}).$$

Still to find proof.

Then for $S_{\text{FVM}} = [u_{\text{FVM}}(\mu_1), \dots, u_{\text{FVM}}(\mu_n)]$ a snapshot of full-order solutions that we will consider to be the RB, consider the primal snapshot :

$$S_B := [G(u_{\text{FVM}}(\mu_1)), \dots, G(u_{\text{FVM}}(\mu_n))] = [u_B(\mu_1), \dots, u_B(\mu_n)]$$

$$u_{\text{FVM,rb}} := P_S(u) = \sum_{i=1}^{n} \langle u_{\text{FVM}}(\mu_i) | u \rangle_{\mathbb{P}^0(\mathcal{B})} \, u_{\text{FVM}}(\mu_i)$$

and

$$u_{B,\text{rb}} = G(u_{\text{FVM},rb})$$

Then we might expect somehow

$$\|u_{\text{FEM}} - u_{\text{FEM},rb}\|_\mu \leq \|u_B - u_{B,rb}\|_\mu \leq C \, \|u_{\text{FEM}} - u_{\text{FEM},rb}\|_\mu$$

recalling

$$\|u_B - u_{B,rb}\|_\mu = \|u_{\text{FVM}} - u_{\text{FVM},rb}\|_\mu$$

Then defining

$$r(v) = \langle f | v \rangle_{\mathbf{L}^2} - \overline{a}(u_{B,rb}, v) = \overline{a}(u_B, v) - \overline{a}(u_{B,rb}, v) = \overline{a}(u_B - u_{B,rb}, v)$$

if we know a to be coercive we should have $\overline{a}$ Inf-Sup stable with constant β hence if we define $\hat{r}$ the Riesz representation on $\mathbb{V}_{\text{Box}}$ we should have

$$\|u_B - u_{B,rb}\|_\mu \leq \frac{\|\hat{r}\|}{\beta}$$

The Box Method is not really satisfying because we only have inf-sup constant, because the considered energy norm is one from FEM method that does not take into account the conservation of fluxes and lastly we are not sure that the FVM solution is mapped to the Box Solution.

II - Finite Volume Weak Formulation

1) Construction of the weak formulation

Raphaele Herbin presentation that is also described in "Analysis tools for finite volume schemes" [Eym+07]

Problem : let $b \geq 0$, $Q \in \mathbf{L}^2(\Omega)$ and $v \in \mathbb{R}^d$

$$\begin{cases} -\Delta u + \nabla\cdot(vu) + bu = Q & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

Define $\mathcal{B}$ the set of control volumes and $\mathcal{E}$ the set of its faces. We suppose the mesh to be admissible in the sense that orthogonality conditions holds.
Recall :

$$\mathbb{P}^0(\mathcal{B}) = \left\{ u \in \mathbf{L}^2(\Omega) \ / \ \forall C \in \mathcal{B}, u\big|_b \in \mathbb{P}^0(b) \right\}$$

Bilan on a control volume $C \in \mathcal{B}$ with faces f :

$$\sum_{f\in \mathrm{F}(C)} \int_f -\nabla u \cdot \mathbf{n}_f d\gamma(x) + \sum_{f\in \mathrm{F}(C)} \int_f (vu)\cdot \mathbf{n}_f d\gamma(x) + \int_{V_C} bu = \int_{V_C} Q$$

Remark :

$\gamma(x)$ is the $(d-1)$-Lesbegue measure.

We then use the FVM approximation with the Upwind scheme :

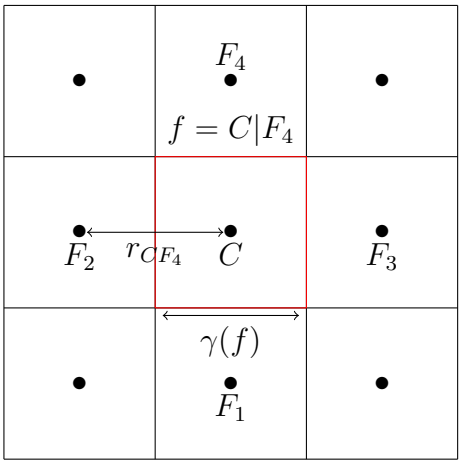
$$\sum_{f\in \mathrm{F}(C)} F_{C,f}(u + \sum_{f\in \mathrm{F}(C)} (v_f^+ u_C + v_f^- u_F) + b_C \left|V_C\right| u_C = \left|V_C\right| f_C$$

where

$$v^+ := \max(0, v) \quad \& \quad v^- := \max(0, -v)$$

Define

$$\mathcal{E}_{\text{int}} := \mathcal{E} \cap \Omega \quad \& \quad \mathcal{E}_{\text{ext}} := \mathcal{E} \cap \partial\Omega$$



Then we define $F_{C,f}(u)$ such that :

$$F_{C,f}(u) := \begin{cases} -\dfrac{\gamma(f)}{r_{CF}}(u_F - u_C) & \text{if } f \in \mathcal{E}_{\text{int}} \\ -\dfrac{\gamma(f)}{r_{Cf}}(-u_C) & \text{if } f \in \mathcal{E}_{\text{ext}} \end{cases}$$

Note r_{Cf} is the distance between C and the border.
We do a sumation over $\mathcal{B}$:

$$\sum_{C\in\mathcal{B}}\left[\sum_{f\in\mathrm{F}(C)}F_{C,f}(u)+\sum_{f\in\mathrm{F}(C)}(v_f^+u_C+v_f^-u_F)+\int_{V_C}bu\right]=\sum_{C\in\mathcal{B}}\int_{V_C}Q$$

$$\iff\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}}-\frac{\gamma(f)}{r_{CF}}(u_F-u_C)+\sum_{f\in\mathcal{E}_{\text{ext}}}\frac{\gamma(f)}{r_{Cf}}u_C+\sum_{C\in\mathcal{B}}\left[\sum_{f\in\mathrm{F}(C)}(v_f^+u_C+v_f^-u_F)\right]+\int_\Omega bu=\int_\Omega Q$$

We want to describe this equation only with bilinear form. Remark:

$$-\frac{\gamma(f)}{r_{CF}}(u_F-u_C)=\frac{\gamma(f)}{r_{CF}}(u_F-u_C)(0-1)=\frac{\gamma(f)}{r_{CF}}(u_F-u_C)(\mathbf{1}_C(x_F)-\mathbf{1}_C(x_C))$$

and

$$\frac{\gamma(f)}{r_{Cf}}u_C=\frac{\gamma(f)}{r_{Cf}}u_C\mathbf{1}_C(x_C)$$

Then we can define :

Definition 5.1: Discrete inner product

$$\langle u|\phi\rangle_{\mathcal{B}}=\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}}\frac{\gamma(f)}{2r_{CF}}(u_F-u_C)(\phi_F-\phi_C)+\sum_{f\in\mathcal{E}_{\text{ext}}}\frac{\gamma(f)}{r_{Cf}}u_C\phi_C$$

Note that we divided by 2. The following property explains why we decided to do so.

Proposition 5.2: Control volume

Let $u\in\mathbb{P}^0(\mathcal{B})$:

$$\langle u|\mathbf{1}_C\rangle_{\mathcal{B}}=\sum_{f\in\mathrm{F}(C)}F_{C,f}(u)$$

Choosing $\phi=\mathbf{1}_C$ or $u=\mathbf{1}_C$ is equivalent to only consider the corresponding cell and its fluxes.

$\mathcal{D}$

We split the following sum between the set neighbour cells of K and the set of cells that has K as neighbour cell:

$$\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}}\frac{\gamma(f)}{2r_{CF}}(u_F-u_C)(\mathbf{1}_K(x_F)-\mathbf{1}_K(x_C))=\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=K|F}}\frac{\gamma(f)}{2r_{KF}}(u_F-u_K)(\mathbf{1}_K(x_F)-\mathbf{1}_K(x_K))$$
$$+\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|K}}\frac{\gamma(f)}{2r_{CK}}(u_K-u_C)(\mathbf{1}_K(x_K)-\mathbf{1}_K(x_C))$$

$$
\begin{aligned}
&= \sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=K|F}} \frac{\gamma(f)}{2r_{KF}}(u_F-u_K)(0-1)\\
&+ \sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|K}} \frac{\gamma(f)}{2r_{CK}}(u_K-u_C)(1-0)\\
&= \sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=K|F}} -\frac{\gamma(f)}{2r_{KF}}(u_F-u_K) + \sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|K}} \frac{\gamma(f)}{2r_{CK}}(u_K-u_C)\\
&= 2\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=K|F}} -\frac{\gamma(f)}{2r_{KF}}(u_F-u_K) \quad (\star)\\
&= \sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=K|F}} -\frac{\gamma(f)}{r_{KF}}(u_F-u_K)\\
&= \sum_{f\in(F(K))_{\text{int}}} F_{K,f}(u)
\end{aligned}
$$

$(\star)$ the set of Control Volumes that has K as a neighbour is exactly $(F(K))_{\text{int}}$.

$$
\sum_{f\in\mathcal{E}_{\text{ext}}} \frac{\gamma(f)}{r_{Cf}} u_C \mathbf{1}_K(x_C) = \sum_{f\in(F(K))_{\text{ext}}} \frac{\gamma(f)}{r_{Cf}} u_K = \sum_{f\in(F(K))_{\text{ext}}} F_{K,f}(u)
$$

The sum of both terms gives the result. □

Remark :

The original paper does not divide by 2. Maybe we did not understand well the sum.

Proposition 5.3

Applying $\mathbf{1}_{V_j}$ and $\mathbf{1}_{V_i}$ where V_i is the i-th cell and V_j is the j-th gives the following sum.

$$
\left\langle \mathbf{1}_{V_i} \middle| \mathbf{1}_{V_j} \right\rangle_{\mathcal{B}} = \begin{cases} \displaystyle\sum_{f\in F(V_i)} \frac{\gamma(f)}{r_{if}} & \text{if } i=j \\ \displaystyle -\frac{\gamma(f_{ij})}{r_{ij}} & \text{if } i,j \text{ are neighbour} \\ 0 & \text{otherwise} \end{cases}
$$

It also proves the following result:

$$
\langle u | \mathbf{1}_C \rangle_{\mathcal{B}} = -\sum_{f\in F(C)} F_{f,C}(u)
$$

where $F_{f,C}$ is the flux from the unique cell F that shares the face f to C. It should be interpreted as the conservation of the flux accross all cells. Which is exactly what we would expect from the inner product defined by the flux for a Weak FVM formulation.

The same way we want to define a bilinear form which gives the convection term of C when we apply $\mathbf{1}_C$, thus we define :

$$c_{\mathcal{B}}(u,\phi) = \sum_{C\in\mathcal{B}} \phi_C \left[\sum_{f\in \mathrm{F}(C)} (v_f^+ u_C + v_f^- u_F) \right]$$

And the weak formulation of integrals is natural since for linear functions:

$$b_C u_C = \frac{1}{|V_C|}\int_{V_C} bu = \frac{1}{|V_C|}\int_\Omega bu\mathbf{1}_C \text{ and } Q_C = \frac{1}{|V_C|}\int_{V_C} Q = \frac{1}{|V_C|}\int_\Omega Q\mathbf{1}_C$$

Hence we define the following weak FVM :

$$\text{Find } u\in\mathbb{P}^0(\mathcal{B}) \text{ s.t. } \quad \langle u|\phi\rangle_{\mathcal{B}} + c_{\mathcal{B}}(u,\phi) + \int_\Omega bu\phi = \int_\Omega Q\phi \quad \forall \phi\in\mathbb{P}^0(\mathcal{B})$$

$\mathcal{D}$: Weak FVM $\iff$ FVM

$\Longrightarrow$: $\phi = \mathbf{1}_K$
$\Longleftarrow$: Take $\phi\in\mathbb{P}^0(\mathcal{B})$, multiply the strong form on a control volume and sum over $C\in\mathcal{B}$.

□

2) Inequalities, norms, structure

Proposition 5.4: WFVM Poincaré inequality

Take $u\in\mathbb{P}^0(\mathcal{B})$.

$$\|u\|_{\mathbf{L}^2} \leq \mathrm{diam}(\Omega)\,\|u\|_{1,\mathcal{B}}$$

$\mathcal{D}$

"Finite Volume Methods" [EGH00] Lemma 9.1

□

Then by analogy with Poincaré inequality on $\mathbf{H}_0^1$ we can define :

Definition 5.5: Discrete $\mathbf{H}_0^1(\mathcal{B})$ norm

$$\|u\|_{1,\mathcal{B}} := \left(\langle u|u\rangle_{\mathcal{B}}\right)^{1/2}$$

From here the most important point is to have the Hilbertian structure to make all the developed theory beforehand works. Thanksfully, the following property holds:

Proposition 5.6: Hilbert structure

$(\mathbb{P}^0(\mathcal{B}), \langle\cdot|\cdot\rangle_{\mathcal{B}})$ is an Hilbert.

$\mathcal{D}$

Let $\varepsilon > 0$. Let $(f_n)_n \in \mathbb{P}^0(\mathcal{B})^{\mathbb{N}}$ a Cauchy sequence :

$$\exists n_0 \text{ s.t.} \quad \|f_p - f_q\|_{1,\mathcal{B}} \leq \varepsilon \quad \forall p, q \geq n_0$$

Then by WFVM Poincaré 5.4 $\|f_p - f_q\|_{\mathbf{L}^2(\Omega)} \underset{p,q\to\infty}{\longrightarrow} 0$. Yet $\mathbf{L}^2$ is complete hence $f_n \underset{n\to\infty}{\longrightarrow} f$ in $\mathbf{L}^2(\Omega)$.
It is easy to verify that f is piecewise constant on $\mathcal{B}$ and that $f_n|_b \longrightarrow f|_b$.
Hence $F_{C,\sigma}(f_n) \longrightarrow F_{C,\sigma}(f)$.
$\star$ This property was not in the paper $\star$ □

Regarding the jump norm equivalence 4.5 we can hope that if $u \in \mathbf{H}^1(\Omega)$:

$$\|u\|_{1,\mathcal{B}} \sim \|\nabla u\|_{\mathbf{L}^2(\Omega)}$$

We can define the discrete $\mathbb{P}^0(\mathcal{B})$ norm defined by analogy with the $\mathbf{H}^1(\Omega)$ norm:

$$\|u\|_{\mathcal{B}} := \|u\|_{\mathbf{L}^2(\Omega)} + \|u\|_{1,\mathcal{B}} \underset{5.4}{\sim} \|u\|_{\mathcal{B}}$$

The following lemma allows a control over the oscillations which is used to prove the 5.8 theorem.

Lemma 5.7: Oscillations

For any $v \in \mathbb{P}^0(\mathcal{B})$:

$$\forall \eta \in \mathbb{R}^d, \quad \|v(\cdot + \eta) - v\|_{\mathbf{L}^2}^2 \leq \|\eta\|_1 (\|\eta\|_1 + 4h_{\mathcal{B}}) \|v\|_{1,\mathcal{B}}^2$$

The paper shows the existence of an unique solution that has the following properties.

Theorema 5.8: Discrete Rellich

Let $(\mathcal{B}_n)_n$ be a sequence of FVM mesh satisfying the orthogonality conditions, s.t. $h_{\mathcal{B}_n} \underset{n\to\infty}{\longrightarrow} 0$.
Let $(u_n)_n \in (\mathbb{P}^0(\mathcal{B}))^{\mathbb{N}}$ s.t. $\|u_n\|_{1,\mathcal{B}} \leq C$.

Then there exists a subsequence $(u_{\varphi(n)})_n$ with $\varphi : \mathbb{N} \to \mathbb{N}$ strictly increasing and $\overline{u} \in \mathbf{H}_0^1(\Omega)$ such that

$$u_{\varphi(n)} \underset{n\to\infty}{\overset{\mathbf{L}^2}{\longrightarrow}} \overline{u}$$

III - Elliptic case : Pure diffusion

1) Model

We will consider a diffusion problem in a simple square divided in 9 smaller squares with different diffusion constants.

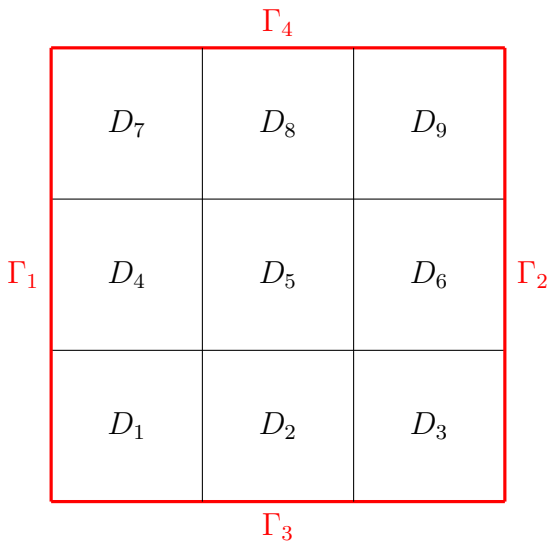


Figure 7: Parametrized diffusion problem

Let S_i be the i-th square associated to D_i. We define:

$$D(x) = \sum_{i=1}^{9} D_{[i]} \mathbf{1}_{S_i}$$

Hence the parameter $\mu = \begin{bmatrix} D_{[1]} \\ D_{[2]} \\ \vdots \\ D_{[9]} \end{bmatrix}$ lives in $\mathbb{P} =]0;1]^9 \simeq [\varepsilon;1]^9$ where $\varepsilon \simeq 0$ (to have closed set of parameters).
The FVM solve for $u \in \mathbb{P}^0(\mathcal{B})$ in each V_C :

$$-\nabla\cdot(D\nabla u) = f \iff \int_{\partial V_C} D\nabla u = D(x_C) \int_{\partial V_C} \nabla u = \int_{V_C} f$$

Hence, defining the **symmetric bilinear form**:

$$a(u,\phi\ ;\mu) = \sum_{\substack{f\in\mathcal{E}_{\text{int}} \\ f=C|F}} D(x_C)\frac{\gamma(f)}{2r_{CF}}(u_F - u_C)(\phi_F - \phi_C) + \sum_{f\in\mathcal{E}_{\text{ext}}} \frac{\gamma(f)}{r_{Cf}} D(x_C) u_C \phi_C$$

is enough to decribe the weak formulation for the FVM.
Then we will solve for u:

$$\begin{gathered} a(u,v\ ;\mu) = \langle f|v\rangle_{\mathbf{L}^2} \ \ \forall v \in \mathbb{P}^0(\mathcal{B}) \\ u = 0 \ \ \text{on } \Gamma \end{gathered}$$

2) Theoretical properties

We still have **coercivity**:

$$a(u,u) \geq \min_{i=1,\dots,9} D_i \left\|u\right\|_{1,\mathcal{B}}^2$$

Note that we also proved $\alpha \geq \min_{i=1,\dots,9} D_i$ i.e. wan can define $\alpha_{LB}(\mu) = \min_{i=1,\dots,9} D_i$.

a is **continuous**:

$$\begin{aligned}
|a(u,\phi)| &= \left| \sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}} D(x_C)\frac{\gamma(f)}{2r_{CF}}(u_F-u_C)(\phi_F-\phi_C) + \sum_{f\in\mathcal{E}_{\text{ext}}} \frac{\gamma(f)}{r_{Cf}} D(x_C) u_C\phi_C \right| \\
&\leq \sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}} \left| D(x_C)\frac{\gamma(f)}{2r_{CF}}(u_F-u_C)(\phi_F-\phi_C)\right| + \sum_{f\in\mathcal{E}_{\text{ext}}} \left|\frac{\gamma(f)}{r_{Cf}} D(x_C) u_C\phi_C\right| \\
&\leq \underbrace{\max_{i=1,\dots,9} D_i}_{=C} \left(\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}} \sqrt{\frac{\gamma(f)}{2r_{CF}}}|u_F-u_C| \times \sqrt{\frac{\gamma(f)}{2r_{CF}}}\,|\phi_F-\phi_C| \right. \\
&\quad \left. + \sum_{f\in\mathcal{E}_{\text{ext}}} \sqrt{\frac{\gamma(f)}{r_{CF}}}|u_C| \times \sqrt{\frac{\gamma(f)}{r_{CF}}}\,|\phi_C| \right) \\
\text{Cauchy-Schwarz} &\leq C \left(\sqrt{\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}} \frac{\gamma(f)}{2r_{CF}}(u_F-u_C)^2} \times \sqrt{\sum_{\substack{f\in\mathcal{E}_{\text{int}}\\ f=C|F}} \frac{\gamma(f)}{2r_{CF}}(\phi_F-\phi_C)^2} \right. \\
&\quad \left. + \sqrt{\sum_{f\in\mathcal{E}_{\text{ext}}} \frac{\gamma(f)}{r_{CF}} u_C^2} \times \sqrt{\sum_{f\in\mathcal{E}_{\text{ext}}} \frac{\gamma(f)}{r_{CF}} \phi_C^2} \right) \\
&\leq C\left(\left\|u\right\|_{1,\mathcal{B}} \left\|\phi\right\|_{1,\mathcal{B}} + \left\|u\right\|_{1,\mathcal{B}} \left\|\phi\right\|_{1,\mathcal{B}} \right) \\
&\leq 2C \left\|u\right\|_{1,\mathcal{B}} \left\|\phi\right\|_{1,\mathcal{B}}
\end{aligned}$$

We can also write the **affine assumption**:

$$\begin{aligned}
a(u,\phi\,;\mu) &= \sum_{i=1}^{9} D_{[i]} \left(\sum_{\substack{f\in\mathcal{S}_i\cap\mathcal{E}_{\text{int}}\\ f=C|F}} \frac{\gamma(f)}{2r_{CF}}(u_F-u_C)(\phi_F-\phi_C) + \sum_{f\in S_i\cap\mathcal{E}_{\text{ext}}} \frac{\gamma(f)}{r_{Cf}} u_C\phi_C \right) \\
&= \sum_{i=1}^{9} D_{[i]} a_i(u,\phi)
\end{aligned}$$

where

- $a_i(u,u) \geq 0$ ($= 0$ if $u = 0$ on S_i) i.e. a_i is a semi-definite bilinear form.
- $D_{[i]} > 0$.

That fills the conditions for the affine assumption 4.3 [HRS16].

Suppose $\mathbb{V}_{rb} \subset \mathbb{V}_\delta = \mathbb{P}^0(\mathcal{B})$ already exists. Define u_{rb} as the solution in $\mathbb{V}_{rb}$ of:

$$\begin{aligned} a(u, v\ ;\mu) &= \langle f|v\rangle_{\mathbf{L}^2} \ \ \forall v \in \mathbb{V}_{rb}(\mathcal{B}) \\ u_{rb} &= 0 \ \ \text{on } \Gamma \end{aligned}$$

Then define the error and residual error:

$$e(\mu) = u_{\text{FVM}}(\mu) - u_{rb}(\mu)$$

$$r : \phi \in \mathbb{V}_\delta \mapsto a(e(\mu), \phi\ ;\mu)$$

By Riesz, there exists $\hat{r}(\mu)$ such that:

$$r(\phi\ ;\mu) = \langle \hat{r}(\mu)|\phi\rangle_{\mathcal{B}}$$

Define the **energy norm error estimator**:

$$\eta_{\text{en}}(\mu) = \frac{\|\hat{r}(\mu)\|_{\mathcal{B}}}{\sqrt{\alpha_{LB}(\mu)}} \geq \|e(\mu)\|_\mu = \sqrt{a(e(\mu), e(\mu)\ ;\mu)}$$

3) Computational methodology

Recall that we want to solve for $u \in \mathbb{V}_\delta = \mathbb{P}^0(\mathcal{B})$:

$$\begin{aligned} -\nabla\cdot(D\nabla u) &= f \ \ \text{in } \Omega \\ u &= 0 \ \ \text{on } \Gamma \end{aligned}$$

where for $(D_{[i]})_i \in [\varepsilon; 1]^9$ we define

$$D(x) = \sum_{i=1}^{9} D_{[i]} \mathbf{1}_{S_i}$$

Set $\mathcal{B}$ an orthogonal mesh. $\mathcal{E}$ be the set of edges of the mesh.

Sidenote :
- ξ_i is the $i-th$ vector of the reduced basis and $\mathbf{B}$ the matrix of ξ coordinates in the $\mathbf{1}_{V_i}$ basis.
- $\mathbf{M}_\delta = \left\langle \mathbf{1}_{V_i} \middle| \mathbf{1}_{V_j} \right\rangle_{\mathcal{B}}$ for all i, j.
- We will suppose that $f(\cdot\ ;\mu) = f(\cdot)$.

a) Precomputation

Compute $\mathbf{M}_\delta$:

$$\left\langle \mathbf{1}_{V_i} \middle| \mathbf{1}_{V_j} \right\rangle_{\mathcal{B}} = \begin{cases} \displaystyle\sum_{f \in F(V_i)} \frac{\gamma(f)}{r_{if}} & \text{if } i = j \\ -\dfrac{\gamma(f_{ij})}{r_{ij}} & \text{if } i, j \text{ are neighbour} \\ 0 & \text{otherwise} \end{cases}$$

Compute once for all $q = 1, ..., 9$ the matrices $\mathbf{A}_\delta^q = \left(a_q(\mathbf{1}_{V_i}, \mathbf{1}_{V_j})\right)_{i,j}$:

$$a_q(\mathbf{1}_{V_i}, \mathbf{1}_{V_j}) = \begin{cases} \displaystyle\sum_{f \in F(V_i)} \frac{\gamma(f)}{r_{if}} & \text{if } i = j \text{ and } i \in S_q \\ -\dfrac{\gamma(f_{ij})}{r_{ij}} & \text{if } i, j \text{ are neighbour and } i \in S_q \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathbf{f}_\delta = \left[(f(\mathbf{1}_{V_i}))_i\right]^T$.

b) Computation of α_{LB}

Set $\mathbb{P}_M \subset \mathbb{P}$ be a set of M arbitrary chosen parameters.
For each $\mu = \begin{bmatrix} D_{[1]} & \dots & D_{[9]} \end{bmatrix}^T \in \mathbb{P}_M$, compute:

$$\mathbf{A}_\delta^\mu = \sum_{i=1}^{9} D_{[q]} \mathbf{A}_\delta^q$$

Solve for $(\lambda, w_\delta) \in \mathbb{R}^+ \times \mathbb{R}^{N_\delta}$ the eigenvalue problem:

$$\mathbf{A}_\delta^\mu w_\delta = \lambda \mathbf{M}_\delta w_\delta$$

The smallest eigenvalues gives you M coercive constant $\alpha_\delta(\mu_m)$.
Then define the function:

$$\alpha_{LB}(\mu) = \max_{m=1,\dots,M} \left(\alpha_\delta(\mu_m) \min_{q=1,\dots,9} \frac{D_{[q]}(\mu)}{D_{[q]}(\mu_m)} \right)$$

c) Step by step offline generation

Set $\mathbb{P}_h = \begin{bmatrix} \mu_{[1]} & \dots & \mu_{[p]} \end{bmatrix}$ all the chosen trial parameters.
Chose any $\mu_1 \in \mathbb{P}_h$.

Loop at n:

- Compute the FOM solution $u_{\text{FVM}}(\mu_n)$ for $\mu_n \in \mathbb{P}_h$ computed in the last iteration or μ_1 if it's the first loop iteration.
  Define $\mathbf{B} = \begin{bmatrix} u_{\text{FVM}}(\mu_1) & \dots & u_{\text{FVM}}(\mu_n) \end{bmatrix} = \begin{bmatrix} \xi_1 & \dots & \xi_n \end{bmatrix}$.
  Compute for $q = 1, ..., 9$:
  $$\mathbf{A}_{rb}^q = \mathbf{B}^T A_\delta^q \mathbf{B}$$
  and
  $$\mathbf{f}_{rb} = \mathbf{B}^T \mathbf{f}_\delta$$
- For each $\mu \in \mathbb{P}_h$:
  Compute $\mathbf{A}_{rb}^\mu = \sum_{i=1}^{9} D_{[i]} A_{rb}^i$. Compute u_{rb}^μ s.t.
  $$\mathbf{A}_{rb}^\mu u_{rb}^\mu = \mathbf{f}_{rb}$$

We then compute $\eta(\mu)$. First compute

$$\mathbf{R} = \left(\mathbf{f}_\delta, A_\delta^1\mathbf{B}, \dots, A_\delta^9\mathbf{B}\right)^T$$

We then need to focus on

$$\mathbf{G} = \mathbf{R}^T\mathbf{M}_\delta^{-1}\mathbf{R}$$

Solve the linear system $\mathbf{M}_\delta y = \mathbf{R}$ then compute $\mathbf{G} = \mathbf{R}^T y$.
Then compute

$$r(\mu) = \left[1, -(u_{rb}^\mu)^T D_{[1]}, \dots, -(u_{rb}^\mu)^T D_{[9]}\right]$$

and finally compute

$$\|\hat{r}_\delta(\mu)\|_{1,\mathcal{B}} = \sqrt{r(\mu)^T\mathbf{G}r(\mu)}$$

Note that $\alpha_{LB}(\mu) = \min\limits_{i=1,\dots 9} D_{[i]}$ is a lower bound of $\alpha(\mu)$. Hence we define:

$$\eta(\mu) = \|\hat{r}_\delta(\mu)\|_{1,\mathcal{B}} / \sqrt{\alpha_{LB}(\mu)}$$

Or we can apply the min-θ approach p.51:

$$\alpha_{LB}(\mu) = \max_{m=1,\dots,M} \left(\alpha_\delta(\mu_m) \min_{q=1,\dots,9} \frac{D_{[q]}(\mu)}{D_{[q]}(\mu_m)}\right)$$

- Choose $\mu_{n+1} = \arg\max\limits_{\mu\in\mathbb{P}_h} \eta(\mu)$.
- If $\eta(\mu_{n+1}) >$ tol (that was previously computed) then **go back** to the beginning of the loop, otherwise **terminate**.

Ensuring stability: Apply the Gram-Schmidt algorithm and redefine

$$\mathbf{B} = \text{Gram-Schmidt}(\mathbf{B})$$

d) Online procedure

Recall $\mathbf{B} = \begin{bmatrix}\xi_1 & \dots & \xi_N\end{bmatrix}$.
For a new $\mu = \begin{bmatrix}D_{[1]} & \dots & D_{[9]}\end{bmatrix} \in \mathbb{P}$, compute:

$$\mathbf{A}_{rb}^\mu = \sum_{i=1}^{9} D_{[i]} A_{rb}^i$$

Compute u_{rb}^μ s.t.

$$\mathbf{A}_{rb}^\mu u_{rb}^\mu = \mathbf{f}_{rb}$$

IV - Parabolic case: Heat equation

1) Model

We don't change the geometry of the model:

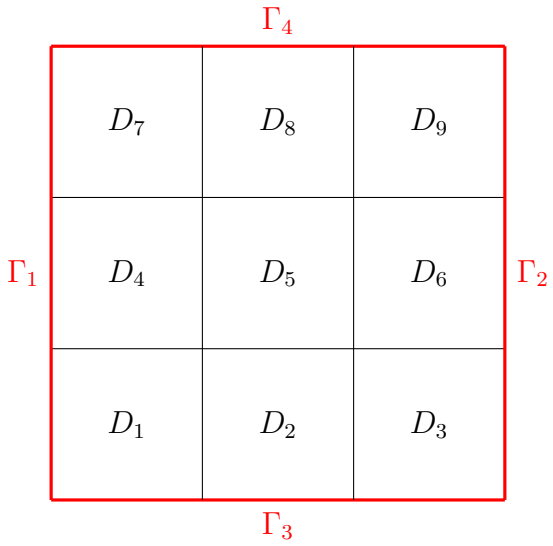


Figure 8: Parametrized heat problem

Let S_i be the i-th square associated to D_i. We define:

$$D(x) = \sum_{i=1}^{9} D_{[i]} \mathbf{1}_{S_i}$$

The parameter $\mu = \begin{bmatrix} D_{[1]} \\ D_{[2]} \\ \vdots \\ D_{[9]} \end{bmatrix}$ lives in $\mathbb{P} = [\varepsilon; 1]^9$ where $\varepsilon \simeq 0$.

Here, we try to solve:

$$\begin{cases} \partial_t u - \nabla \cdot (D \nabla u) = g(t) f(v) & \text{in } \Omega \times \mathbb{R}_+^* \\ u(x,t) = 0 & \text{in } \partial\Omega \times \mathbb{R}_+ \\ u(x,0) = u_0 \in \mathbf{L}^2 & \text{in } \Omega \end{cases}$$

A weak formulation would be:
Find $u \in \mathbf{L}^2(\mathbf{H}_0^1(\Omega) \times \mathbb{R}_+, \mathbb{R})$ such that $u_t \in \mathbf{L}^2(\mathbf{H}^{-1}(\Omega), \mathbb{R}_+)$ and:

$$\begin{cases} \langle \partial_t u | \varphi \rangle_{\mathbf{H}^{-1}, \mathbf{H}_0^1} + \displaystyle\int_\Omega D \nabla u \cdot \nabla v dx = g(t) \langle f(x) | \varphi \rangle_{\mathbf{H}_0^1} & \forall \varphi \in \mathbf{H}_0^1(\Omega) \\ u(x,0) = u_0 \in \mathbf{L}^2 & \text{in } \Omega \end{cases}$$

We discretise the time steps by finite difference, i.e.

$$\partial_t u = \frac{u^{n+1} - u^n}{t^{n+1} - t^n}$$

Consider $T > 0$ a final time, M the number of time steps and $k = T/M$ the uniform time step.
The Weak FVM formulation can be written as:
Find $(u^n)_{0 \leq n \leq M-1}$ such that $u^n \in \mathbb{P}^0(\Omega)$ and:

$$\begin{cases} \frac{1}{k} \langle u^{n+1} - u^n | \phi \rangle_{\mathbf{L}^2} + a(u^{n+1}, \phi) = g(t^{n+1}) \langle f | \phi \rangle_{\mathbf{L}^2} & \forall \phi \in \mathbb{P}^0, \ 0 \leq n \leq M-1 \\ \langle u^0 | \phi \rangle_{\mathbf{L}^2} = \langle u_0 | \phi \rangle_{\mathbf{L}^2} & \forall \phi \in \mathbb{P}^0 \end{cases}$$

where $a(u, v)$ is defined as the diffusion symmetric continuous coercive bilinear form p.48.

Remark :

When applying cell control, we get $\frac{1}{k} \langle u^{n+1} - u^n | \mathbf{1}_C \rangle_{\mathbf{L}^2} = V_C \frac{u_C^{n+1} - u_C^n}{\Delta t}$ which is the transient approximation in FVM.

2) Computational methodology: POD-Greedy algorithm

Set T the final time of the simulation.
Set K the number of time steps.
Set N_1 the number of chosen principal temporal modes.
Set $N_2 \leq N_1$ for the POD compression.

a) Precompute

Compute $\mathbf{M}_\delta$ such that

$$[\mathbf{M}_\delta]_{i,j} = \left\langle \mathbf{1}_{V_i} \middle| \mathbf{1}_{V_j} \right\rangle_{\mathcal{B}} = \begin{cases} \sum_{f \in F(V_i)} \frac{\gamma(f)}{r_{if}} & \text{if } i = j \\ -\frac{\gamma(f_{ij})}{r_{ij}} & \text{if } i, j \text{ are neighbour} \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathbf{L}_\delta = \left\langle \mathbf{1}_{V_i} \middle| \mathbf{1}_{V_j} \right\rangle_{\mathbf{L}^2} = |V_i| \delta_{ij}$.
Compute once for all $q = 1, ..., 9$ the matrices $\mathbf{A}_\delta^q = \left(a_q(\mathbf{1}_{V_i}, \mathbf{1}_{V_j}) \right)_{i,j}$:

$$a_q(\mathbf{1}_{V_i}, \mathbf{1}_{V_j}) = \begin{cases} \sum_{f \in F(V_i)} \frac{\gamma(f)}{r_{if}} & \text{if } i = j \text{ and } i \in S_q \\ -\frac{\gamma(f_{ij})}{r_{ij}} & \text{if } i, j \text{ are neighbour and } i \in S_q \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathbf{f}_\delta = \left[(f(\mathbf{1}_{V_i}))_i \right]^T$.
Compute $\Delta t = \frac{T}{K}$.

b) Step by step offline generation

Set $\mathbb{P}_h = \begin{bmatrix} \mu_{[1]} & \dots & \mu_{[p]} \end{bmatrix}$ all the chosen trial parameters.
Set $\mathcal{Z} = 0$.
Choose any $\mu_1 \in \mathbb{P}_h$.

Loop at n:

- Compute the full order solution $(u^k_{\text{FVM}}(\mu_n))_{1\leq k\leq K} = (u_{\text{FVM}}(k\Delta t\ ;\mu)_{1\leq k\leq K}$ for $\mu_n \in \mathbb{P}_h$ computed in the last iteration or μ_1 if it's the first loop iteration.
  Apply the POD algorithm for the temporal reduction:

$$\{\zeta_1,\dots,\zeta_{N_1}\} = \text{POD}(\{u^1_{\text{FVM}}(\mu_n),\dots,u^K_{\text{FVM}}(\mu_n)\},N_1)$$

  Set $\mathcal{Z} = \{\mathcal{Z},\{\zeta_1,\dots,\zeta_{N_1}\}\}$.
  Set $N = N + N_2$ and compute $\{\xi_1,\dots,\xi_N\} = \text{POD}(\mathcal{Z},N)$.
  Define $\mathbf{B} = \begin{bmatrix}\xi_1 & \dots & \xi_n\end{bmatrix}$.
  Compute for $q = 1,...,9$:

$$\mathbf{A}^q_{rb} = \mathbf{B}^T A^q_\delta \mathbf{B}$$

  and

$$\mathbf{f}_{rb} = \mathbf{B}^T\mathbf{f}_\delta$$

- For each $\mu = \left[D_{[1]},...,D_{[9]}\right] \in \mathbb{P}_h$:
  Compute $\mathbf{A}^\mu_{rb} = \sum_{i=1}^{9} D_{[i]}A^i_{rb}$ and $\mathbf{L}_{rb} = \mathbf{B}^T\mathbf{L}_\delta\mathbf{B}$.
  Compute $(u^{\mu,k}_{rb})_{1\leq k\leq K}$ s.t.

$$\left(\frac{1}{\Delta t}\mathbf{L}_{rb} + \mathbf{A}^\mu_{rb}\right)u^{\mu,k+1}_{rb} = \frac{1}{\Delta t}\mathbf{L}_{rb}u^{\mu,k}_{rb} + g(t^k)\mathbf{f}_{rb}$$

  We then compute $\eta(t^K,\mu)$. First compute

$$\mathbf{R} = \left(\mathbf{f}_\delta, \mathbf{L}_\delta\mathbf{B}, A^1_\delta\mathbf{B},\dots,A^9_\delta\mathbf{B}\right)^T$$

  Define $\Delta^{\mu,k}_{rb} = u^{\mu,k}_{rb} - u^{\mu,k-1}_{rb}$.
  Then compute

$$r^k(\mu) = \left[g(t^k), -\tfrac{1}{\Delta t}(\Delta^{\mu,k}_{rb})^T, -(u^\mu_{rb})^T D_{[1]},\dots,-(u^\mu_{rb})^T D_{[9]}\right]$$

  We then need to focus on

$$\mathbf{G} = \mathbf{R}^T\mathbf{M}^{-1}_\delta\mathbf{R}$$

  Solve the linear system $\mathbf{M}_\delta y = \mathbf{R}$ then compute $\mathbf{G} = \mathbf{R}^T y$.
  and finally compute

$$\left\|\hat{r}^k_\delta(\mu)\right\|^2_{1,\mathcal{B}} = r^k(\mu)^T\mathbf{G}r^k(\mu)$$

  Note that $\min_{i=1,\dots 9} D_{[i]}$ is a lower bound of $\alpha(\mu)$. Hence we define:

$$\eta(t^K,\mu) = \sqrt{\frac{\Delta t}{\alpha_{LB}(\mu)}\sum_{k=1}^{K}\left\|\hat{r}^k_\delta(\mu)\right\|^2_{1,\mathcal{B}}}$$

  Or we can apply the min-θ approach p.51:

$$\alpha_{LB}(\mu) = \max_{m=1,\dots,M}\left(\alpha_\delta(\mu_m)\min_{q=1,\dots,9}\frac{D_{[q]}(\mu)}{D_{[q]}(\mu_m)}\right)$$

- Choose $\mu_{n+1} = \underset{\mu \in \mathbb{P}_h}{\arg\max}\ \eta(\mu)$.
- If $\eta(\mu_{n+1}) >$ tol (that was previously computed) then **go back** to the beginning of the loop, otherwise **terminate**.

Ensuring stability: Apply the Gram-Schmidt algorithm and redefine

$$\mathbf{B} = \text{Gram-Schmidt}(\mathbf{B})$$

Store $\mathbf{B}$ and $\mathbf{L}_{rb} = \mathbf{B}^T \mathbf{L}_\delta \mathbf{B}$.

c) Online procedure

Recall $\mathbf{B} = \begin{bmatrix} \xi_1 & \dots & \xi_N \end{bmatrix}$ and $\mathbf{L}_{rb} = \mathbf{B}^T \mathbf{L}_\delta \mathbf{B}$.
For a new $\mu = \begin{bmatrix} D_{[1]} & \dots & D_{[9]} \end{bmatrix} \in \mathbb{P}$, compute:

$$\mathbf{A}^\mu_{rb} = \sum_{i=1}^{9} D_{[i]} A^i_{rb}$$

Compute $(u^{\mu,k}_{rb})_k$ such that:

$$\left(\frac{1}{\Delta t} \mathbf{L}_{rb} + \mathbf{A}^\mu_{rb} \right) u^{\mu,k+1}_{rb} = \frac{1}{\Delta t} \mathbf{L}_{rb} u^{\mu,k}_{rb} + g(t^k) \mathbf{f}_{rb}$$

V - Results and conclusion

Numerical results involve programming knowledge that I did not have the time to learn. Giovanni is taking care of this part, but it is a time consuming part of the research and we will be able to present in September 2024.

The work presented gives insights on a method enhance CFD with a given estimator, which can be adapted to Inf-Sup problems. If the numerical results look promising, we could hope to use this theory for libraries such that OpenFOAM (see ISTHACA-FV project).

The next step is probably to write the methodology for Inf-Sup problems, and also to create an estimator that takes into account the Implict part of the gradient estimation for non-orthogonal meshes.

Part 6
Appendix

I - Standards definitions

Definition 6.1: Hilbert space

A space $(H, \langle\cdot|\cdot\rangle)$ is a Hilbert if H is a vector space and $\langle\cdot|\cdot\rangle$ is a inner product that induces a complete norm.

Definition 6.2: Coercive bilinear form

A bilinear form a is coercive over a Hilbert V if

$$\exists \alpha > 0 \ \forall v \in V : : \ \frac{a(v,v)}{\|v\|_V^2} \geq \alpha$$

Definition 6.3: Inf-Sup form

A bilinear form $a : V \times W \to \mathbb{R}$ is Inf-Sup stable over a Hilbert $V \times W$ if

$$\exists \beta_0 > 0 \ : \ \inf_{w \in W} \sup_{v \in V} \frac{a(v,w)}{\|v\|_V \|w\|_W} \geq \beta_0$$

Definition 6.4: Sobolev space

For $\Omega \subset \mathbb{R}^d$, we define the Sobolev space $H^1(\Omega)$ as

$$H^1(\Omega) := \left\{ f \in \mathbb{L}^2(\Omega) \ / \ \forall i, \ \partial_i f \in \mathbb{L}^2(\Omega) \right\}$$

$H^1(\Omega)$ is a Hilbert space for the norm $\|f\| := \|f\|_{\mathbf{L}^2} + \|\nabla f\|_{\mathbf{L}^2}$.
We also define $H_0^1(\Omega) := \overline{\mathcal{D}(\Omega)}^{H^1(\Omega)}$ the closure of $\mathcal{D}(\Omega) := \mathcal{C}_0^\infty(\Omega)$ in $H^1(\Omega)$.

II - Representation theorems

Theorema 6.5: Riesz-Fréchet

Let $(H, \langle\cdot|\cdot\rangle_H)$ be a Hilbert space on $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.
If $\phi \in H'$, then there exists a unique $x_0 \in H$ such as

$$\forall x \in H, \phi(x) = \langle x | x_0 \rangle$$

Moreover the map

$$x_0 \in H \longmapsto \varphi_{x_0}; \left| \begin{array}{ccc} H & \to & \mathbb{K} \\ x & \mapsto & \langle x | x_0 \rangle \end{array} \right. \in H'$$

is bijective, antilinear and an isometry between H and H'.

Theorema 6.6: Lax-Milgram

Let H be a Hilbert space, $a : H \times H \longrightarrow \mathbb{R}$ a continous coercive bilinear form on H. Given any $\varphi \in H'$, there exists a unique $u \in H$ such that

$$a(u, v) = \langle \varphi | v \rangle = \varphi(v) \quad \forall v \in H$$

Moreover if a is symmetric then u is characterized by the property

$$\boxed{u \in H \text{ and } \frac{1}{2} a(u, u) - \varphi(u) = \min_{v \in H} \left\{ \frac{1}{2} a(v, v) - \varphi(v) \right\}}$$

Theorema 6.7: Banach–Nečas–Babuška

Let V, W be respectively Banach and reflexive Banach space.

$a : V \times W$ continuous bilinear form.
$f : W \to \mathbb{R}$ continuous linear form.

$$(\star\star) : \quad \text{find } u \in V, \quad a(u, w) = f(w) \quad \forall w \in W$$

$(\star\star)$ is well-posed if and only if

$$(i) \qquad \exists C_{\text{sta}} > 0, \forall v \in V, \quad \sup_{w \in W \setminus \{0\}} \frac{a(v, w)}{\|w\|_W} \geq C_{\text{sta}} \|v\|_V$$

$$(ii) \qquad \forall w \in W, \quad (\forall v \in V, a(v, w) = 0) \Longrightarrow w = 0$$

(i) is equivalent to the Inf Sup 6.3 condition.
The following control holds true :

$$\|u\|_V \leq \frac{1}{C_{\text{sta}}} \|f\|_{W'}$$

$\mathcal{D}$: Proof of Banach–Nečas–Babuška

See "Banach-Nečas-Babuška theorem and proof." [Lec24a]. □

III - Standards inequalities

Lemma 6.8: Bramble Hilbert 1D

If u has m derivates on (a,b) and $\mathbb{P}_k$ is the space of polynomials of degree lesser than $m-1$.

$$\inf_{v\in\mathbb{P}_{m-1}} \left\|u^{(k)}-v^{(k)}\right\|_{\mathbf{L}^p} \leq C(m)(b-a)^{m-k}\left\|u^{(m)}\right\|_{\mathbf{L}^p}$$

Hence for $p=\infty$ and $m=2$ we have the linear interpolation :

$$\inf_{v\in\mathbb{P}_1} \|u-v\|_\infty \leq C(b-a)^2\|u''\|_\infty$$

Lemma 6.9: Bramble Hilbert

If Ω is regular enough and satisfies the strong cone property, $u\in W^{m,p}(\Omega)$ and $\mathbb{P}_k$ is the space of polynomials of degree lesser than $m-1$.

$$\forall k\leq m,\quad \inf_{v\in\mathbb{P}_{m-1}} \|u-v\|_{W^{k,p}} \leq Cd^{m-k}\|u\|_{W^{m,p}}$$

Proposition 6.10: Poincaré's inequality

Let K be a convex polygon, h its diameter, $\varphi\in\mathbf{H}^1(K)$.
Let $\varphi_K=\dfrac{\langle\varphi|1\rangle_K}{|K|}$ be the average estimation of φ. Then

$$\|\varphi-\varphi_K\|_K^2 \leq Ch_k^2\|\nabla\varphi_K\|_K^2$$

where C is independent of K.

Proposition 6.11: Generalised Friedrichs' inequality

Let K be a convex polygon, h its diameter, $\varphi\in\mathbf{H}^1(K)$.
Let $\varphi_\sigma=\dfrac{\langle\varphi|1\rangle_\sigma}{|\sigma|}$ be the average estimation of φ on the K border. Then

$$\|\varphi-\varphi_\sigma\|_K^2 \leq Ch_k^2\|\nabla\varphi_K\|_K^2$$

where C depends on K geometry, dimension and its border.

References

[AO93] Mark Ainsworth and J. Tinsley Oden. "A unified approach to a posteriori error estimation using element residual methods". In: *Numerische Mathematik* 65 (1993). ISSN: 0945-3245. DOI: https://doi.org/10.1007/BF01385738.

[BR87] Randolph E. Bank and Donald J. Rose. "Some Error Estimates for the Box Method". In: *SIAM Journal on Numerical Analysis* 24.4 (1987), pp. 777–787. DOI: 10.1137/0724050.

[BMO96] Jacques Baranger, Jean-François Maitre, and Fabienne Oudin. "Connection between finite volume and mixed finite element methods". en. In: *ESAIM: Modélisation mathématique et analyse numérique* 30.4 (1996), pp. 445–465. URL: http://www.numdam.org/item/M2AN_1996__30_4_445_0/.

[Bin+11] Peter Binev et al. "Convergence Rates for Greedy Algorithms in Reduced Basis Methods". In: *SIAM J. Math. Analysis* 43 (Jan. 2011), pp. 1457–1472. DOI: 10.1137/100795772.

[BS08] Susanne C. Brenner and L. Ridgway Scott. *The Mathematical Theory of Finite Element Methods,* Springer New York, NY, 2008. DOI: https://doi.org/10.1007/978-0-387-75934-0.

[Bre10] Haïm Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York, NY, 2010. DOI: https://doi.org/10.1007/978-0-387-70914-7.

[Buf+21] A. Buffa et al. "Apriori convergence of the greedy algorithm for the parametrized reduced basis method." In: *ESAIM Math. Model. Numer. Anal.* 46 (2021), pp. 595–603. DOI: https://doi.org/10.1051/m2an/2011056.

[Che10] Long Chen. "A New Class of High Order Finite Volume Methods for Second Order Elliptic Equations". In: *SIAM Journal on Numerical Analysis* 47.6 (2010), pp. 4021–4043. DOI: 10.1137/080720164.

[Dro+18] Jérôme Droniou et al. *The Gradient Discretisation Method*. Springer Cham, 2018. DOI: https://doi.org/10.1007/978-3-319-79042-8.

[EG10] Alexandre Ern and Jean-Luc Guermond. *Theory and Practice of Finite Elements*. Springer New York, NY, 2010. DOI: https://doi.org/10.1007/978-1-4757-4355-5.

[EGH00] Robert Eymard, Thierry Gallouët, and Raphaèle Herbin. "Finite Volume Methods". In: *Solution of Equation in $\mathbb{R}^n$ (Part 3), Techniques of Scientific Computing (Part 3)*. Ed. by J. L. Lions and Philippe Ciarlet. Vol. 7. Handbook of Numerical Analysis. Elsevier, 2000, pp. 713–1020. DOI: 10.1016/S1570-8659(00)07005-8. URL: https://hal.science/hal-02100732.

[Eym+07] Robert Eymard et al. "Analysis tools for finite volume schemes". In: *Acta Mathematica Universitatis Comenianae* 76 (May 2007). URL: https://hal.science/hal-03085383.

[GM21] Grosjean, Elise and Maday, Yvon. "Error estimate of the non-intrusive reduced basis method with finite volume schemes". In: *ESAIM: M2AN* 55.5 (2021), pp. 1941–1961. DOI: 10.1051/m2an/2021044.

[HRS16] Jan S. Hesthaven, Gianluigi Rozza, and Benjamin Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer Cham, 2016. DOI: https://doi.org/10.1007/978-3-319-22470-1.

[Huy+07] D.B.P. Huynh et al. "A successive constraint linear optimization method for lower bounds of parametric coercivity and inf–sup stability constants". In: *Comptes Rendus Mathematique* 345.8 (2007), pp. 473–478. ISSN: 1631-073X. DOI: https://doi.org/10.1016/j.crma.2007.09.019.

[Jas96] Hrvoje Jasak. "Error Analysis and Estimation for the Finite Volume Method With Applications to Fluid Flows". PhD thesis. Imperial College, 1996.

[Lec24a] Raphaël Lecoq. "Banach-Nečas-Babuška theorem and proof." 2024. URL: https://perso.eleves.ens-rennes.fr/people/raphael.lecoq/Banach_Necas_Babuska_proof.pdf.

[Lec24b] Raphaël Lecoq. "Internship Report: Study and implementation of a posteriori estimator for Finite Volume Methods in CFD". 2024. URL: https://perso.eleves.ens-rennes.fr/people/raphael.lecoq/projects.

[MMD16] F. Moukalled, L. Mangani, and M. Darwish. *The Finite Volume Method in Fluid Dynamics*. Springer, 2016. DOI: 10.1007/978-3-319-16874-6.

[NPV24] G. Negrini, N. Parolini, and M. Verani. "On the convergence of the Rhie–Chow stabilized Box method for the Stokes problem". In: *International Journal for Numerical Methods in Fluids* 96.8 (May 2024), pp. 1489–1516. ISSN: 1097-0363. DOI: 10.1002/fld.5295.

[PE12] Daniele Antonio Di Pietro and Alexandre Ern. *Mathematical aspects of discontinuous Galerkin method*. Vol. 69. Springer, 2012. DOI: https://doi.org/10.1007/978-3-642-22980-0.

[RSB22] Gianluigi Rozza, Giovanni Stabile, and Francesco Ballarin. *Advanced Reduced Order Methods and Applications in Computational Fluid Dynamics*. English. SIAM PUBLICATIONS, 2022. ISBN: 978-1-61197-724-0. DOI: 10.1137/1.9781611977257.

[Sta23] Giovanni Stabile. "From linear to nonlinear model order reduction: some results and perspectives." Slides. 2023. URL: https://docs.google.com/presentation/d/1t3AdQeNKUvfUpP9LI5lZK6Lmllh8cRBFe3ltGxoPFE4/edit#slide=id.g275ceeab28c_0_23.

[SR18] Giovanni Stabile and Gianluigi Rozza. "Finite volume POD-Galerkin stabilised reduced order methods for the parametrised incompressible Navier-Stokes equations". In: *Computers and Fluids* (2018). DOI: 10.1016/j.compfluid.2018.01.035.

[Sta+17] Giovanni Stabile et al. "POD-Galerkin reduced order methods for CFD using Finite Volume Discretisation: vortex shedding around a circular cylinder". In: *Communications in Applied and Industrial Mathematics* 8.1 ((2017)), pp. 210–236. DOI: 10.1515/caim-2017-0011.

[Vol12] Stefan Volkwein. "Proper Orthogonal Decomposition: Theory and Reduced-Order Modelling". In: *Lecture Notes, University of Konstanz* (Jan. 2012), pp. 1–10. URL: https://www.math.uni-konstanz.de/numerik/personen/volkwein/teaching/POD-Book.pdf.

[Wu+22] Cheng-Chieh Wu et al. "The finite volume method in the context of the finite element method". In: *Materials Today: Proceedings* 62 (2022). 37th Danubia Adria Symposium on Advances in Experimental Mechanics, pp. 2679–2683. ISSN: 2214-7853. DOI: https://doi.org/10.1016/j.matpr.2022.05.460.