A p-adaptative finite volume method for the convection-diffusion problem

S. Clain [a,b], G. Machado [a], R. M. S. Pereira [a]

[a] *centro de Matemática*
Campus de Azurèm - 4800-058 Guimarães, Portugal
[b] *Institut de Mathématiques de Toulouse,*
Université Paul Sabatier, 31062 Toulouse, France

Abstract

Key words: Finite Volume, high-order, convection-diffusion, polynomial reconstruction, MOOD, p-adaptative.

Email addresses: clain@math.uminho.pt (S. Clain), gjm@math.umimho.pt (G. Machado), rmp@math.uminho.pt (R. M. S. Pereira).

1 Finite volume scheme for the convection Diffusion problem

1.1 Introduction

Second-order scheme, Patankar. Stability, unstructured.

1.2 The steady-state problems

Let Ω be a bounded polygonal domain of $\mathbb{R}^2$ and $\partial\Omega$ the boundary. We denote by a a positive function (the diffusive coefficient) and $V = (u, v)$ a vectorial function (the velocity) defined on $\overline{\Omega}$. We shall consider two types of problem: the pure convective steady-state problem and the convection-diffusion steady-state problem.

In the first case, we introduce the inflow and boundary given by

$$\Gamma_D = \{x \in \partial\Omega;\ V(x).n(x) < 0\},$$

where n denote the outward normal vector. The steady-state convective problem write:

$$\nabla.(V\phi) = f, \quad \text{in } \Omega, \tag{1a}$$
$$\phi = \phi_D, \quad \text{on } \Gamma_D, \tag{1b}$$

where ϕ is the unkwon function, f the source term and ϕ_D a prescribed Dirichlet boundary condition at the inflow boundary. Note that in the case we have $\Gamma_D = \partial\Omega$, function ϕ_D has to satisfy the compatibility condition $\displaystyle\int_{\partial\Omega} V.n\phi_D \, ds = \int_\Omega f(x) \, dx$.

For the convection-diffusion problem we divide the boundary in three parts, namely Γ_D, Γ_P and Γ_T corresponding to the Dirichlet condition, the Neumann condition for the diffusive flux and the Neumann condition for the total flux respectively. The steady-state problem writes

$$\nabla.(V\phi - a\nabla\phi) = f, \quad \text{in } \Omega, \tag{2a}$$
$$\phi = \phi_D, \quad \text{on } \Gamma_D, \tag{2b}$$
$$-a\nabla\phi.n = g_P, \quad \text{on } \Gamma_P, \tag{2c}$$
$$V.n\phi - a\nabla\phi.n = g_T, \quad \text{on } \Gamma_T, \tag{2d}$$

where ϕ_D, g_P, and g_T are prescribed functions for the Partial or Total flux.

1.3 Mesh and discretization

We consider a partition $\mathcal{T}$ of Ω in convex polygonal cells and denote by $\mathcal{C}$ the cell index set. To design the p-adaptative finite volume method, we adopt the following conventions.

- Inner interface situtation: two cells c_i and c_j shares a common edge e_{ij} and n_{ij} is the normal vector from c_i toward c_j. Note that $n_{ij} = -n_{ji}$.
- Interface on the boundary: a cell c_i shares a common edge with the boundary we shall denote e_{iD} (resp. e_{iP} or e_{iT}) if the edge belongs to Γ_D (resp: Γ_P or Γ_T). Note that the convention implies that n_{iD} (resp. n_{iP} or n_{iT}) corresponds to the outward normal vector.
- To deal with the interface on the boundary, we augment the cell index $\mathcal{C}$ with the indexes D, P and T and set $\mathcal{C}_b = \mathcal{C} \cup \{D, P, T\}$. Note that for the pure convection problem we shall only need index $\{D\}$ to characterize the edges on Γ_D.
- For any cell $i \in \mathcal{C}$, we associate the index set $\nu(i) \subset \mathcal{C}_b$ such that $j \in \nu(i)$ if e_{ij} is a common side between c_i and c_j or with the boundary if $j = D, P, T$.
- We denote by $|c_i|$ and $|e_{ij}|$, $i \in \mathcal{C}$, $j \in \nu(i)$ the area and length of c_i and e_{ij} respectively.
- For any interface e_{ij}, $i \in \mathcal{C}$, $j \in \nu(i)$, we denote by $q_{ij,r}$ the Gauss points with weight ζ_r, $r = 1, ..., R$ while M_{ij} represent the midpoint of the interface. Note that M_{iD}, M_{iP}, M_{iT} correspond to the midpoint of interface lying on Γ_D, Γ_P, Γ_T respectively.

Integrating equation (2) over cell c_i and applying the divergence theorem yields

$$\int_{\partial c_i} (V.n\phi - a\nabla\phi.n)ds - \int_{c_i} f\, dx = 0.$$

Using the Gauss quadrature formula on the interfaces provide the approximation

$$\sum_{j\in\nu(i)} |e_{ij}| \sum_{r=1}^{R} \zeta_r \Big[V(q_{ij,r}).n(q_{ij,r})\phi(q_{ij,r}) - a(q_{ij,r})\nabla\phi(q_{ij,r}).n(q_{ij,r}) \Big] - |c_i| f_i = 0(h^{2R})$$

where f_i is the mean value of f over c_i and $h = \max_{k\in\mathcal{E}} |e_k|$.
Base on the previous approximation, we consider a numerical scheme under the residual form

$$G_i = \sum_{j\in\nu(i)} \frac{|e_{ij}|}{|c_i|} \sum_{r=1}^{R} \zeta_r \mathcal{F}_{ij,r} - f_i$$

where $\mathcal{F}_{ij,r}$ will be an accurate approximation of the flux function at the Gauss points.

2 The Polynomial Reconstruction Operator (PRO)

To provide very high accurate approximations of the flux across the interface, we introduce the Polynomial Reconstruction Operator (PRO) in order to recorver a local polynomial aproximation associated to cells or interfaces.

2.1 Stencil

Basically a stencil is a collection of cells in the vicinity of a reference geometrical entity. We shall consider two situations whether we deal with a cell or an edge. For a given cell c_i and a

polynomial degree d, we associate the stencil $S(c_i, d)$ while $S(e_{ij}, d)$, $j \in \nu(i)$ denote the stencil associated to edge e_{ij}.

Remark 2.1 *A strategy to obtain a relevant stencil consist to choose the N_d cells closest to c_i or to e_{ij}. Since the polynomial of degree d has $(d+1)(d+2)/2$ coeficients, on ha to take $N_d \geq (d+1)(d+2)/2$. In pratice we choose $N_d \in](d+1)(d+2)/2, (d+1)(d+2)]$ for the sake of stability.*

2.2 Conservative reconstruction on cell

Let c_i be a cell, d the degree of the polynomial reconstruction we intend to build and $S(c_i, d)$ the associated stencil. We assume that we known the mean values ϕ_i on c_i and ϕ_l on cells c_l, $l \in S(c_i, d)$. We consider the polynomial function of degree d

$$\widehat{\phi}_i(x; d) = \phi_i + \sum_{1 \leq |\alpha| \leq d} \mathfrak{R}_i^{d,\alpha} \Big\{ (x - b_i)^\alpha - M_i^\alpha \Big\} \tag{3}$$

with $\alpha = (\alpha_1, \alpha_2)$, $|\alpha| = \alpha_1 + \alpha_2$ and b_i the centroid of cell c_i. We set $M_i^\alpha = \dfrac{1}{|c_i|} \displaystyle\int_{c_i} (x - b_i)^\alpha \, dx$ such that the conservative property $\dfrac{1}{|c_i|} \displaystyle\int_{c_i} \widehat{\phi}_i(x; d) \, dx = \phi_i$ holds. To fixe the coefficients $\mathfrak{R}_i^{d,\alpha}$ of relation (3) we introduce the functional

$$E_i(\mathfrak{R}_i^d; d) = \sum_{l \in S(c_i, d)} \omega_{i,l} \Big[\frac{1}{|c_l|} \int_{c_l} \widehat{\phi}_i(x; d) \, dx - \phi_l \Big]^2, \tag{4}$$

where $\omega_{i,l}$ are positive weights and $\mathfrak{R}_i^d$ is the vector gathering all the coefficients $\mathfrak{R}_i^{d,\alpha}$. The quadratique functional has a unique solution $\widehat{\mathfrak{R}}_i^d$ which minimizes the functional and provides the best approximation.

Remark 2.2 *The minimizing point of functional (4) corresponds to an over-determined linear system where one has to check that the rank is equal to the number of coefficients $\mathfrak{R}_i^{d,\alpha}$, i.e. $(d+1)(d+2)/2 - 1$ for the two-dimension case. The determination of vector $\mathfrak{R}_i^d$ is obtained using the Moore-Penrose pseudo-inverse matrix. Preconditioning technique proposed in [] is also employed to reduce the conditioning number of the linear system.*

2.3 Conservative reconstruction on edge

We turn now to the polynomial reconstruction associated to an edge. We first consider the situation where e_{iD} corresponds to an edge situated on the boundary Γ_D. We assume that we know $\phi_{iD} = \dfrac{1}{|e_{iD}|} \displaystyle\int_{e_{iD}} \phi_D(s) \, ds$ and ϕ_l on cells c_l, $l \in S(e_{iD}, d)$. We consider the polynomial

function of degree d

$$\widehat{\phi}_{iD}(x;d) = \phi_{iD} + \sum_{1\leq|\alpha|\leq d} \mathfrak{R}_{iD}^{d,\alpha}\Big\{(x-b_{iD})^{\alpha} - M_{iD}^{\alpha}\Big\} \tag{5}$$

where b_{iD} the centroid of edge e_{iD}. We set $M_{iD}^{\alpha} = \dfrac{1}{|e_{iD}|}\displaystyle\int_{e_{iD}} (x-b_{iD})^{\alpha}\, ds$ such that the conservative property $\dfrac{1}{|e_{iD}|}\displaystyle\int_{e_{iD}} \widehat{\phi}_{iD}(x;d)\, ds = \phi_{iD}$ holds. As in the previous case, we introduce the functional

$$E_{iD}(\mathfrak{R}_{iD}^{d};d) = \sum_{l\in S(e_{iD},d)} \omega_{iD,l}\Big[\frac{1}{|c_l|}\int_{c_l} \widehat{\phi}_{iD}(x;d)\, dx - \phi_l\Big]^2, \tag{6}$$

where $\omega_{iD,l}$ are positive weights and $\mathfrak{R}_{iD}^{d}$ is the vector gathering all the coefficients $\mathfrak{R}_{iD}^{d,\alpha}$. Vector $\widehat{\mathfrak{R}}_{iD}^{d}$ stands for the unique vector minimizing the functional and providing the best approximation.

2.4 Non Conservative reconstruction on edge

If the edge does not belong to Γ_D, we do not have any information about the mean value so a non-conservative reconstruction is required. Let e_{ij} be an interface with $j \neq D$, then we consider the polynomial reconstruction of the form

$$\widetilde{\phi}_{ij}(x;d) = \sum_{0\leq|\alpha|\leq d} \mathfrak{R}_{ij}^{d,\alpha}(x-b_{ij})^{\alpha} \tag{7}$$

where b_{ij} is the centroid of interface e_{ij}. Note that the index summation start with $|\alpha| = 0$. Assume that we know ϕ_l on cells c_l, $l \in S(e_{ij},d)$ we introduce the functional

$$E_{ij}(\mathfrak{R}_{ij}^{d};d) = \sum_{l\in S(e_{ij},d)} \omega_{ij,l}\Big[\frac{1}{|c_l|}\int_{c_l} \widetilde{\phi}_{ij}(x;d)\, dx - \phi_l\Big]^2, \tag{8}$$

where $\omega_{ij,l}$ are positive weights and $\mathfrak{R}_{ij}^{d}$ is the vector gathering all the coefficients $\mathfrak{R}_{ij}^{d,\alpha}$. Vector $\widetilde{\mathfrak{R}}_{ij}^{d}$ stands for the unique vector minimizing the functional and providing the best approximation.

2.5 Linearity and Consitency

We have define three polynomial reconstruction $\widehat{\phi}_i$, $\widehat{\phi}_{iD}$ and $\widetilde{\phi}_{ij}$ which correspond to linear operators with respect to the mean values. Let us denote by $\Phi = (\phi_i)_{i\in\mathcal{C}}$, we then have defined linear operators with respect to Φ. Note that the mean values of ϕ_D on Γ_D are not considered as unknows since function ϕ_{iD} is well-defined.

We say that the reconstruction is d-consistent in the following sense. For any polynomial function ϕ of degree d, let denote by $\overline{\phi}_i$ and $\overline{\phi}_{iD}$ the mean values on the cell and interfaces. The reconstruction is d-consistent if

$$\widehat{\phi}_i(x;d) = \widehat{\phi}_{iD}(x;d) = \widetilde{\phi}_{ij}(x;d) = \phi(x).$$

If such a property holds, we say that the finite volume method assoicated to the polynomial reconstruction (in short PRO-d) is a $d+1^{th}$-order.

3 High-order finite volume scheme

3.1 The Patankar scheme

We shortly present the traditional Patankar method cast in the residual formulation. The flux ares given by the following rules

(1) Assume that e_{ij} is an inner edge, then we define the flux at the quadrature point with

$$\mathcal{F}_{ij}^{PAT} = [V(M_{ij}).n(M_{ij})]^+\phi_i + [V(M_{ij}).n(M_{ij})]^-\phi_j - a(M_{ij})\frac{\phi_j - \phi_i}{|b_i b_j|}.$$

(2) Assume that e_{iD} belongs to Γ_D,

$$\mathcal{F}_{iD}^{PAT} = [V(M_{iD}).n(M_{iD})]^+\phi_i + [V(M_{iD}).n(M_{iD})]^-\phi_D(M_{iD}) - a(M_{iD})\frac{\phi_D(M_{iD}) - \phi_i}{|b_i M_{iD}|}.$$

(3) Assume that e_{iP} is on Γ_P,

$$\mathcal{F}_{iP,r}^{PAT} = [V(M_{iP}).n(M_{iP})]^+\phi_i + g_P(M_{iP}).$$

(4) Assume that e_{iT} is on Γ_T,

$$\mathcal{F}_{iT,r}^{PAT} = g_T(M_{iT}).$$

For any vector $\Phi \in \mathbb{R}^I$ with $I = \#\mathcal{C}$, we then define the residual quantity

$$G_i^{PAT}(\Phi) = \sum_{j\in\nu(i)} \frac{|e_{ij}|}{|c_i|}\mathcal{F}_{ij}^{PAT}(\Phi) - f_i \tag{9}$$

and setting $G^{PAT}(\Phi) = (G_1^{PAT}(\Phi), \ldots, G_I^{PAT}(\Phi))^t$. We then obtain an afine operator from $\mathbb{R}^I$ into $\mathbb{R}^I$. The solution is the vector $\widehat{\Phi}$ such that $G^{PAT}(\widehat{\Phi}) = 0$ while the consistency error is given by $G^{PAT}(\overline{\Phi}) = 0$ with $\overline{\phi}_i$ the mean values of the exact solution.

3.2 Unlimited PRO-d schemes

Based on the polynomial reconstruction, we now design the finite volume scheme. We assume that d is fixed and that we have provided stencils which guarantee a d-consistent reconstruction. We introduce the numerical flux in function of the interface. One has to distinguish four situations.

(1) Assume that e_{ij} is an inner edge, then we define the flux at the quadrature point with

$$\mathcal{F}_{ij,r} = [V(q_{ij,r}).n(q_{ij,r})]^+ \widehat{\phi}_i(q_{ij,r};d) + [V(q_{ij,r}).n(q_{ij,r})]^- \widehat{\phi}_j(q_{ij,r};d) - a(q_{ij,r}) \nabla \widetilde{\phi}_{ij}(q_{ij,r};d).n(q_{ij,r}).$$

(2) Assume that e_{iD} belongs to Γ_D,

$$\mathcal{F}_{iD,r} = [V(q_{iD,r}).n(q_{iD,r})]^+ \widehat{\phi}_i(q_{iD,r};d) + [V(q_{iD,r}).n(q_{iD,r})]^- \phi_D(q_{iD,r}) - a(q_{iD,r}) \nabla \widehat{\phi}_{iD}(q_{iD,r};d).n(q_{iD,r}).$$

(3) Assume that e_{iP} is on Γ_P,

$$\mathcal{F}_{iP,r} = [V(q_{iP,r}).n(q_{iP,r})]^+ \widehat{\phi}_i(q_{iP,r};d) + g_P(q_{iP,r}).$$

(4) Assume that e_{iT} is on Γ_T,

$$\mathcal{F}_{iT,r} = g_T(q_{iT,r}).$$

For any vector $\Phi \in \mathbb{R}^I$ with $I = \#\mathcal{C}$, we then define the residual quantity

$$G_i(\Phi) = \sum_{j \in \nu(i)} \frac{|e_{ij}|}{|c_i|} \sum_{r=1}^{R} \zeta_r \mathcal{F}_{ij,r}(\Phi) - f_i \tag{10}$$

As in the previous case, we set $G(\Phi) = (G_1(\Phi), \ldots, G_I(\Phi))^T$ and then obtain an afine operator from $\mathbb{R}^I$ into $\mathbb{R}^I$. The solution is the vector $\widehat{\Phi}$ such that $G(\widehat{\Phi}) = 0$ while the consistency error is given by $G(\overline{\Phi}) = 0$ with $\overline{\phi}_i$ the mean values of the exact solution.

3.3 The p-adaptative finite volume scheme

The previous scheme assumes that the solution is regular enough such that the polynomial reconstruction is valid. Unfortunately, there exist situations where the solution is not smooth and the regularity may depend on a, V but also f (the source term may be a Dirac distribution for example). To take the local regularity into account, we associate a specific degree d_i (name the Cell Polynomial Degree or CellPD in short) for each cell c_i. We gather all the CellPD in a vector D called the CellPD Map which provides the picture of the function regularity.
We modify our previous scheme introducing the following rule:

$$d_{ij} = \min(d_i, d_j),\ i \in \mathcal{C},\ j \in \nu(i) \cap \mathcal{C},$$
$$d_{ij} = d_i,\ i \in \mathcal{C},\ j = D, P, T.$$

The flux for the p-adaptative finite volume scheme are then given by

(1) inner edge

$$\mathcal{F}_{ij,r} = [V(q_{ij,r}).n(q_{ij,r})]^+ \widehat{\phi}_i(q_{ij,r}; d_{ij}) + [V(q_{ij,r}).n(q_{ij,r})]^- \widehat{\phi}_j(q_{ij,r}; d_{ij}) - a(q_{ij,r}) \nabla \widetilde{\phi}_{ij}(q_{ij,r}; d_{ij}).n(q_{ij,r}).$$

(2) Assume that e_{iD} belongs to Γ_D,

$$\mathcal{F}_{ij,r} = [V(q_{iD,r}).n(q_{iD,r})]^+ \widehat{\phi}_i(q_{iD,r}; d_i) + [V(q_{iD,r}).n(q_{iD,r})]^- \widehat{\phi}_D(q_{iD,r}) - a(q_{iD,r}) \nabla \widehat{\phi}_{iD}(q_{iD,r}; d_i).n(q_{iD,r}).$$

(3) Assume that e_{iP} is on Γ_P,

$$\mathcal{F}_{ij,r} = [V(q_{iP,r}).n(q_{iP,r})]^+ \widehat{\phi}_i(q_{iP,r}; d_i) + g_P(q_{iP,r}).$$

(4) Assume that e_{iT} is on Γ_T,

$$\mathcal{F}_{ij,r} = g_T(q_{iT,r}).$$

It result that for a given CellPD map D, we have an linear operator $\Phi \to G(\Phi; D)$ and we denote by $\widehat{\Phi}(D)$ the solution of $G(\Phi; D) = 0$.

3.4 Some comments on the problem resolution

Since all the stage are linear operations, the problem leads to an afine operator given by $G(\Phi; D) = A_D \Phi - b_D$ where A_D is a $I \times I$ matrix and $b_D \in \mathbb{R}^I$. We are dealing with a free-matrix method in the sense that we do no construct explicitely the matrix and iterative method such that the GMRES method are commonly used to determine an approximation.
For numerical purpose, it can also be interesting to determine the associated matrix (determination of the eigenvalues, characterizetion of an M-matrix). We can easily recover matrix A_D and vector b_D noting that $b_D = G(0; D)$ and A_D is constituted of the column vectors $G(e^{(k)}; D)$ with $e^{(k)}$ the canonical vectors, $k = 1, \ldots, I$.

4 Determination of the CellPD map

The MOOD method was previously introduce in [] to determine the optimal degree for the polynomial reconstruction still preserving the stability. We shall use a similar technique inspired from the original method. The idea is to determine the CellPD map such that the polynomial reconstructions for ϕ remains stable (no Gibbs phenomena) still providing an high accuracy. To this end, we proceed in two step. In a first stage we determine an initial CellPD map D^{ini} in function of the coefficient regularity, namely a, V but also the source term (a Dirac distribution provides a local rougth solution). The second stage implements the MOOD method where an iterative procedure to obtain the optimal degree on each cell. The MOOD procedure is constituted of two main ingredients: a set of detectors which determine if the solution on a

cell is locally eligible and a correction procedure which reduces the polynomial degree when not eligible.

As mention above, function a or V are not necessary smooth enough to achieve a very high order approximation of the solution. Since the function are known, a first idea could be to define the initial CellPD map "by hand" fixing the degree in function of the local regularities.11 Our objective is to propose an automatics procedure which determines the polynomial degree with respect to the stencils we shall use in the procedure reconstruction for ϕ. We present the method with the diffusion coefficient case, the other cases are treated in the same way.

We compute the mean values a_i, u_i and v_i over the cells c_i, $i \in \mathcal{C}$ and denote by $\widehat{a}_i(x;d)$, $\widehat{u}_i(x;d)$, $\widehat{v}_i(x;d)$ the respective polynomial reconstruction of degree d based on the stencil $S(c_i,d)$ for function ϕ . The point that one has to reduce the polynomial degree where function a is locally rough. To detect if a reconstruction on cell c_i of degree d is available, we adapt the smoothness detector proposed in the ENO/WENO method [**?**]. The basic idea is that for C^s functions the derivatives of order $d > s$ of the polynomial function shall behave as $O(h_i)^{s-d+1}$ where h_i is a characteristic lenght of the cell while the derivatives of order $d \leq s$ behave as $O(1)$. Moreover, if the function is localy discontinuous, we espect a behavior of type $O(h_i^d)$ for the derivatives of order d.
Based on this assumption, we introduce the coefficients

$$\beta_i^d = \sum_{|\alpha|=d} \left|\mathfrak{R}_i^{d,\alpha}\right|^2, \quad a_i^{d,m} = \min_{j\in S(c_i;d)}(a_i,a_j),\ a_i^{d,M} = \max_{j\in S(c_i;d)}(a_i,a_j).$$

The reconstruction of degree d is eligible if one of the two following situations holds

(1) $a_i^{d,M} - a_i^{d,m} \leq \varepsilon |c_i|^{d/2}$, (plateau case)
(2) $|c_i|\beta_i^d < \varepsilon (a_i^{d,M} - a_i^{d,m})^2$ (d-regularity),

where ε is a positive coefficient lower than one which characterizes the level of smoothness.

To build the CellPD map associated to function a, we addopt the following procedure. We loop over all the cell c_i, $i \in \mathcal{C}$. For each cell, we initialise $d = d_{max}$ where d_{max} is the maximum degree for the reconstruction. If the reconstruction of degree d is eligible we set $d_i = d$ else we decrement $d := d-1$. We repeat the procedure till we reach an eligible reconstruction or $d = 0$.

At this end, we obtain the CellPD map D^a associated to the a function. In the same we we conpute D^u and D^v. Finally, we set the initial CellPD map D^{ini} with

$$D_i^{ini} = \min(D_i^a + 1, D_i^u, D_i^v).$$

We prescribe $D_i^a + 1$ because a acts with the gradient of ϕ while V acts with the ϕ function itself.