Monotonic diamond and DDFV type finite-volume schemes for 2D elliptic problems

Xavier Blanc, Francois Hermeline, Emmanuel Labourasse Null, Julie Patela

▶ **To cite this version:**

Xavier Blanc, Francois Hermeline, Emmanuel Labourasse Null, Julie Patela. Monotonic diamond and DDFV type finite-volume schemes for 2D elliptic problems. Communications in Computational Physics, 2023, 34 (2), pp.456-502. 10.4208/cicp.OA-2023-0081 . cea-04137599

HAL Id: cea-04137599

https://cea.hal.science/cea-04137599

Submitted on 22 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Monotonic diamond and DDFV type finite-volume schemes for 2D elliptic problems

Xavier Blanc[1], Francois Hermeline[2,3], Emmanuel Labourasse[2,3], and Julie Patela[1,2]

[1]Université Paris Cité, Sorbonne Université, CNRS, Laboratoire Jacques-Louis Lions, F-75013 Paris, France.
[2]CEA, DAM, DIF, F-91297 Arpajon, France.
[3]Université Paris-Saclay, CEA DAM DIF, Laboratoire en Informatique Haute Performance pour le Calcul et la Simulation, 91297 Arpajon, France.

June 22, 2023

Abstract

The DDFV (Discrete Duality Finite Volume) method is a finite volume scheme mainly dedicated to diffusion problems, with some outstanding properties. This scheme has been found to be one of the most accurate finite volume methods for diffusion problems. In the present paper, we propose a new monotonic extension of DDFV, which can handle discontinuous tensorial diffusion coefficient. Moreover, we compare its performance to a diamond type method with an original interpolation method relying on polynomial reconstructions. Monotonicity is achieved by adapting the method from [44, 19, 49, 18] to our schemes. Such a technique does not require the positiveness of the secondary unknowns. We show that the two new methods are second-order accurate and are indeed monotonic on some challenging benchmarks as a Fokker-Planck problem.

Contents

1 Introduction

Consider the model stationary diffusion problem

$$\begin{cases} -\boldsymbol{\nabla}\cdot(\boldsymbol{\kappa}\boldsymbol{\nabla}\bar{u}) + \lambda\bar{u} = f & \text{in} \quad \Omega, \\ \bar{u} = g & \text{on} \quad \Gamma_D, \\ \boldsymbol{\kappa}\boldsymbol{\nabla}\bar{u}\cdot\mathbf{n} = g & \text{on} \quad \Gamma_N, \end{cases} \tag{1}$$

where Ω is a bounded open domain of $\mathbb{R}^2$ with $\partial\Omega = \Gamma_D \cup \Gamma_N$ ($\Gamma_D \cap \Gamma_N = \emptyset$) and $\mathbf{n} \in \mathbb{R}^2$ the outgoing unit normal vector. The data are such that $f, \lambda \in L^2(\Omega)$, with $\lambda \geq 0$ (if $\lambda = 0$, then $|\Gamma_D| > 0$), $g \in H^{1/2}(\Gamma_D)$ and $g \in L^2(\Gamma_N)$. The tensor-valued diffusion coefficient $\boldsymbol{\kappa}$ is supposed to be bounded and to satisfy the uniform ellipticity condition

$$\forall \mathbf{x} \in \Omega, \quad \forall \mathbf{y} \in \mathbb{R}^2, \quad \alpha_{min}\|\mathbf{y}\|^2 \leq \mathbf{y}^t\boldsymbol{\kappa}(\mathbf{x})\mathbf{y} \leq \alpha_{max}\|\mathbf{y}\|^2,$$

where α_{min}, α_{max} are positive coefficients. Under the above conditions, and if either $\lambda > 0$ or Γ_D is of positive length, it is well known that system (1) has a unique solution in $H^1(\Omega)$. Such a solution satisfies a positiveness principle, i.e. if $f \geq 0$ and $g \geq 0$, then $\bar{u} \geq 0$ (see [15] for example).

Standard methods may be applied to the discretization of such diffusion equations with possibly discontinuous κ on arbitrary meshes. This proves to be an efficient strategy, as far as accuracy (or convergence) is concerned. However, it is well known that positiveness of the discrete solution does not hold. This lack of positiveness (also called monotonicity) can lead to serious difficulties, since $\bar{u}$ can account for a temperature or a concentration. A first attempt to solve the issue of monotonicity would be to truncate the discrete solution to zero. This is not satisfactory because conservation is lost in such a process, and conservation is an important property of the scheme. Some algorithms based on the repair technique introduced in [34] are employed to fix the conservation issue [8, 33, 43, 45]. However, these algorithms are only *globally* (and not locally) conservative, and the consistency is unclear. Some monotonic methods have been designed in the finite-element framework (see [9, 10, 24, 25, 41] among others), but they rely on restrictive conditions on the mesh, that we cannot afford. For fifteen years many original finite volume methods have been proposed to address the issue of monotonicity, while preserving conservation. Most of these schemes are nonlinear or have a larger stencil than standard methods. The finite volume framework is well suited to achieve monotonicity because it allows for an easy manipulation of the fluxes. The first works we know of are those of Le Potier [28] and Bertolazzi and Manzini [3]. In such methods, one uses a manipulation of the fluxes that leads to introduce a dependence on the discrete solution in the coefficients of the fluxes, making the scheme nonlinear, although (1) is linear. To this end, one usually introduces secondary unknowns (for instance vertex-located or face-located unknowns) in addition to the primary (cell-located) unknowns. Among others, important contributions to this field are [5, 18, 30, 39, 48], which propose efficient numerical schemes preserving the positiveness of the primary unknowns. In [38] the requirement of positive secondary unknowns is relaxed. The works [31, 50] explain how to build monotonic schemes without relying on secondary unknowns. In [29, 32, 37], maximum principle preserving schemes are proposed. Cancès and Guichard obtained moreover an entropy diminishing property in [7], introducing the nonlinearity directly at the continuous level via a change of variables. Some concepts and proofs about the existence of solutions for these types of scheme can be found in [11, 13, 36]. See also [42, 46] for recent advances in this field.

The DDFV (Discrete Duality Finite Volume [21], [12]) scheme relies on secondary (nodal) unknowns. However, in contrast with most above-mentioned methods, one considers an additional diffusion problem on a so-called *dual* mesh to calculate them. This scheme has been found to be one of the most accurate finite volume methods for diffusion problems [20], at the price of doubling the number of degrees of freedom compared for instance to the linear or bilinear finite element method or to cell centered methods such as MPFA (Multi Point Flux Approximation [1]) or SUSHI (Scheme Using Stabilization and Hybrid Interfaces [17]). However, none of latter methods are monotonic.
A monotonic extension of DDFV has been proposed in [6], but was not compatible with Neumann boundary conditions, and only first-order convergent for discontinuous tensor coefficients $\boldsymbol{\kappa}$. In the present paper, we propose a new monotonic extension of DDFV that remedies these flaws. Moreover, we compare its performance to a diamond type method with an original interpolation method relying on polynomial reconstructions. Monotonicity is achieved by adapting the method of [44, 19, 49, 18] to our schemes. Such a technique does not require the positiveness of the secondary unknowns.
The main steps of the proposed methods may be briefly summarized as follows.

1. Integration of the equation over each cell of the user's mesh that we will call *primal*.
2. Transformation of this surface integral into a sum of fluxes using the divergence theorem.
3. Approximation of the fluxes using the midpoint quadrature rule on each face of the cell.
4. Taylor expansion of the solution $\bar{u}$ in the neighborhood of the midpoint of each face along *two* independent privileged directions in order to obtain an approximation of $\boldsymbol{\nabla}\bar{u}$ involving the values of $\bar{u}$ and its derivatives at certain suitably chosen points, in this case the center and vertices of the cell.
5. Thanks to this Taylor expansion, estimation of $(\boldsymbol{\kappa}\boldsymbol{\nabla}\bar{u})\cdot\mathbf{n} = (\boldsymbol{\nabla}\bar{u})\cdot(\boldsymbol{\kappa}^t\mathbf{n})$.
6. Calculation of the values of $\bar{u}$ at vertices either by a polynomial interpolation formula in the neighborhood of the midpoint of each primal cell face or by integration of the equation over each cell of the dual mesh.
7. Calculation of the values of derivatives of $\bar{u}$ at centers and vertices of the neighboring cells by differentiating this polynomial interpolation.
8. Transformation of the scheme into a monotonic nonlinear two point flux approximation (or four point flux approximation if a DDFV type method is used).
9. Resolution of the nonlinear system by the Picard iteration method.

The integration over the primal mesh is common to the two monotonic schemes proposed here and is described in Sec. 3. The treatment of the vertex unknowns depends on the scheme and is addressed in Sec. 4. Monotonicity of both schemes is based on the same strategy, which is described in Sec. 5. It leads to a two point flux the coefficients of which depend on the unknown. The Picard iteration method to handle the nonlinearity is also described. The properties of the new DDFV schemes are listed in Sec. 6. Finally, both schemes are assessed in term of accuracy, monotonicity and computational efficiency, and compared with the non monotonic DDFV scheme in Sec. 7. It is shown that the interpolation-based scheme is more efficient for a given L^2 accuracy, but that the DDFV-based scheme achieves second-order accuracy in H^1 norm for the tests we ran. This outstanding feature has been already observed in [20, 23]. Our final test problem is a solution of a simplified Fokker-Planck equation. We show that our scheme is able to compute a correct monotonic solution while achieving the energy conservation.
In all that follows vectors and matrices will be noted with bold letters while $\mathbf{x} = (x, y)$ and $\mathbf{I}$ will stand for the position and 2×2 identity matrix, respectively.

2 Definitions and notations

In this section we gather most of the notations that will be used later. Consider an arbitrary primal mesh made of (possibly distorted, non-conformal, non convex...) *polygonal* cells that are denoted P_i $(1 \leq i \leq n)$. The center of a cell P_i is denoted by $\mathbf{x}_i$ (in general $\mathbf{x}_i$ is the *mass* center of P_i but other interior points for which P_i is starred could be chosen) and its faces are $F_\ell = \mathbf{x}_r\mathbf{x}_s$. The center of the face F_ℓ is $\mathbf{x}_\ell$, the unit vector orthogonal to the face F_ℓ (directed from cell P_i to cell P_j) is $\mathbf{n}_{sr}$ and $\mathbf{N}_{sr} = \|\mathbf{x}_s - \mathbf{x}_r\|\mathbf{n}_{sr}$.

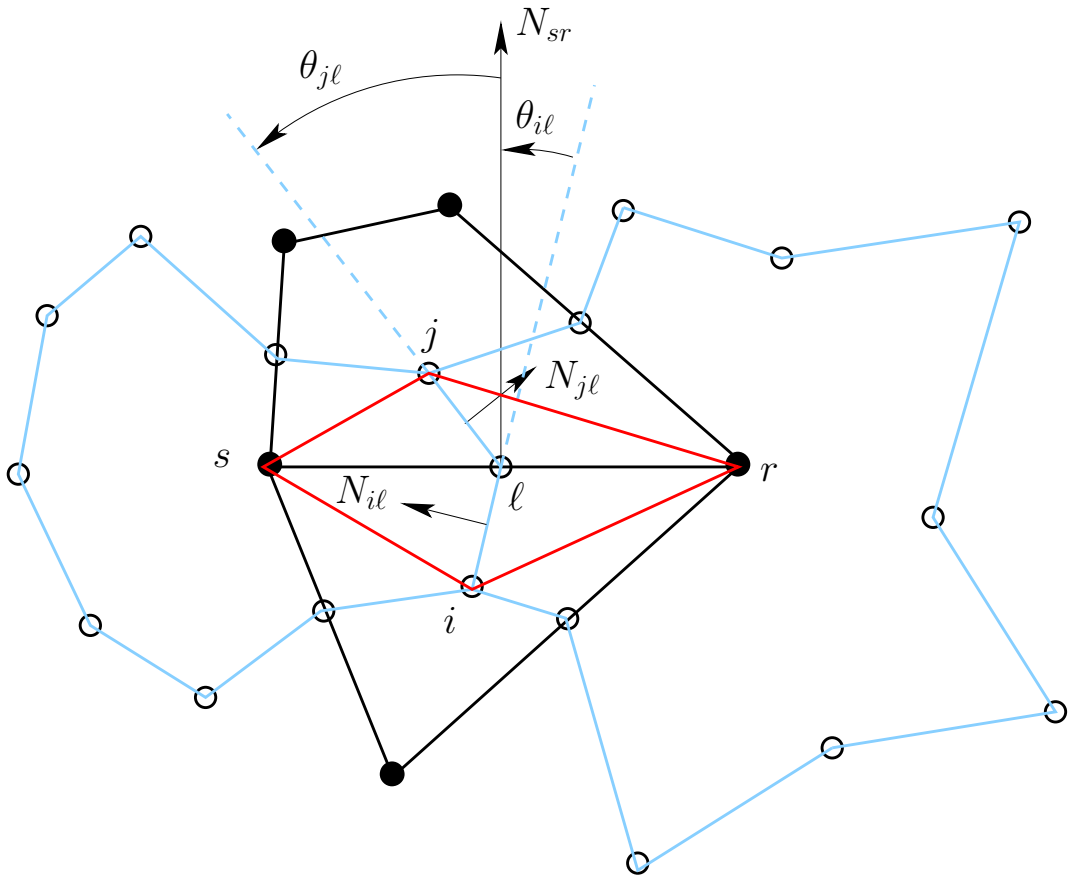


Figure 1: Two primal cells P_i, P_j (black lines) such that $P_i \cap P_j = F_\ell = \mathbf{x}_r\mathbf{x}_s$, two dual cells D_r, D_s (blue lines) such that $D_r \cap D_s = G_\ell = \mathbf{x}_i\mathbf{x}_\ell \cup \mathbf{x}_\ell\mathbf{x}_j$ and one intermediary cell $I_\ell = I_{i\ell} \cup I_{\ell j} = \mathbf{x}_i\mathbf{x}_r\mathbf{x}_\ell\mathbf{x}_s \cup \mathbf{x}_j\mathbf{x}_s\mathbf{x}_\ell\mathbf{x}_r$ (red lines).

In order to define DDFV type schemes we also need to define a dual mesh (often named barycentric or Donald dual mesh) obtained from the primal mesh by joining the center of each cell with the center of its neighbors and the middle of its boundary faces. The dual cells are denoted by D_r ($1 \leq r \leq m$), their faces are $G_{i\ell} = \mathbf{x}_i\mathbf{x}_\ell$, $G_{\ell j} = \mathbf{x}_\ell\mathbf{x}_j$ and we set $G_\ell = G_{i\ell} \cup G_{\ell j}$. The unit vector orthogonal to the face $G_{i\ell}$ (resp. $G_{\ell j}$) and directed from dual cell D_r to dual cell D_s is $\mathbf{n}_{i\ell}$ (resp. $\mathbf{n}_{\ell j}$) and $\mathbf{N}_{i\ell} = \|\mathbf{x}_\ell - \mathbf{x}_i\|\mathbf{n}_{i\ell}$ (resp. $\mathbf{N}_{\ell j} = \|\mathbf{x}_j - \mathbf{x}_\ell\|\mathbf{n}_{\ell j}$). Let $\theta_{i\ell}$ (resp. $\theta_{\ell j}$) be the (trigonometrically oriented) angle between vectors[1] $\mathbf{N}_{\ell i}^\perp$ (resp. $\mathbf{N}_{j\ell}^\perp$) and $\mathbf{N}_{sr}$.
If $F_\ell \not\subset \partial\Omega$ (resp. $F_\ell \subset \partial\Omega$) we denote by I_ℓ the quadrilateral $\mathbf{x}_i\mathbf{x}_r\mathbf{x}_j\mathbf{x}_s$ (resp. the degenerate quadrilateral $\mathbf{x}_i\mathbf{x}_r\mathbf{x}_\ell\mathbf{x}_s$). Note that all these cells, which we will call *diamond* or *intermediary*, also constitute a mesh of Ω. Each interior diamond cell I_ℓ can be divided into two degenerate quadrilaterals $I_{i\ell} = \mathbf{x}_i\mathbf{x}_r\mathbf{x}_\ell\mathbf{x}_s$ and $I_{\ell j} = \mathbf{x}_j\mathbf{x}_s\mathbf{x}_\ell\mathbf{x}_r$ that will be called diamond sub-cells.
Most of these notations are illustrated by Fig. 1. Finally, given a geometrical quantity X (face or cell), we will denote by $|X|$ its measure (length or area).
Define

$$h = \max_\ell(|F_\ell|, |G_{i\ell}|, |G_{\ell j}|),$$

we will assume that the primal and dual meshes satisfy the following assumptions.

1. (**H1**) There exists a constant θ_0 independent of h such that, for all ℓ,

$$|\theta_0| < \frac{\pi}{2}, \quad \cos(\theta_0) < \cos(\theta_{i\ell}), \quad \cos(\theta_0) < \cos(\theta_{\ell j}).$$

[1] Given $\mathbf{v} = (v_x, v_y)$ a vector in $\mathbb{R}^2$ we will use the common notation $\mathbf{v}^\perp = (-v_y, v_x)$.

2. (**H2**) Given N_i (resp. N_r) the number of faces of the primal (resp. dual) cell P_i (resp. D_r), there exists a constant $N_{\max}$ independent of h such that

$$\max(\max_i N_i, \max_r N_r) < N_{\max}.$$

3. (**H3**) There exists a constant ξ independent of h such that, for all ℓ,

$$|I_\ell| \leq \xi \min(|P_i|, |P_j|, |D_r|, |D_s|).$$

Given $\mathbf{v} = (v_i)$ a vector in $\mathbb{R}^n$ we will denote respectively its Euclidian, L^2 and L^∞ norms by

$$\|\mathbf{v}\| = \left(\sum_{i=1}^n v_i^2\right)^{1/2}, \quad \|\mathbf{v}\|_2 = \left(\sum_{i=1}^n |P_i| v_i^2\right)^{1/2}, \quad \|\mathbf{v}\|_\infty = \max_{1\leq i\leq n} |v_i|,$$

and we use the following notations

$$\mathbf{v} \geq 0 \quad \text{if} \quad \forall i,\ v_i \geq 0,$$
$$\mathbf{v} > 0 \quad \text{if} \quad \forall i,\ v_i > 0.$$

Given $\mathbf{x}_k$ any point and ϕ any function we will often note $\phi_k = \phi(\mathbf{x}_k)$.

3 Finite volume formulation on the primal mesh

We will assume that $\boldsymbol{\kappa}$ is continuous inside each cell but can be discontinuous along some primal faces F_ℓ. To simplify the presentation it will be assumed for now on that $\boldsymbol{\kappa}$ is scalar-valued, that is, $\boldsymbol{\kappa} = \kappa \boldsymbol{I}$ with $\alpha_{min} \leq \kappa \leq \alpha_{max}$. For a tensor-valued coefficient $\boldsymbol{\kappa} \in \mathbb{R}^{2,2}$ it is enough to replace $\kappa\mathbf{n}$ by $\boldsymbol{\kappa}^t\mathbf{n}$ in the following calculations.

The first step to design a finite volume scheme consists in integrating equation (1) on cell P_i

$$-\int_{P_i} \boldsymbol{\nabla} \cdot (\kappa \boldsymbol{\nabla} \bar{u}) + \int_{P_i} \lambda \bar{u} = \int_{P_i} f.$$

We can make use of the divergence formula to obtain

$$-\sum_{\ell \in i} \int_{F_\ell} \kappa \boldsymbol{\nabla} \bar{u} \cdot \mathbf{n} + \int_{P_i} \lambda \bar{u} = \int_{P_i} f, \tag{2}$$

where the compact notation $\sum_{\ell \in i}$ stands for the sum on all faces F_ℓ of the primal cell P_i. We need to approximate the flux

$$\bar{\mathcal{F}}_\ell = \int_{F_\ell} \kappa \boldsymbol{\nabla} \bar{u} \cdot \mathbf{n}_{sr}.$$

Suppose that $\kappa_{i\ell}$ (resp. $\kappa_{\ell j}$) is a first-order approximation of κ in the diamond sub-cell $I_{i\ell}$ (resp. $I_{\ell j}$), for example

$$\kappa_{i\ell} = \kappa(\mathbf{x}_i), \quad \kappa_{\ell j} = \kappa(\mathbf{x}_j),$$

or, if κ is continuous

$$\kappa_{i\ell} = \kappa_{\ell j} = \kappa(\mathbf{x}_\ell).$$

Denote by $(\boldsymbol{\nabla}\bar{u})_{i\ell}$ and $(\boldsymbol{\nabla}\bar{u})_{\ell j}$ expressions of $\boldsymbol{\nabla}\bar{u}$ in P_i and P_j. We have

$$-\sum_{\ell \in i} \int_{F_\ell} \kappa_{i\ell} \left(\boldsymbol{\nabla}\bar{u}\right)_{i\ell} \cdot \mathbf{n}_{sr} + \int_{P_i} \lambda \bar{u} = \int_{P_i} f + \mathcal{O}\left(h^3\right).$$

A Taylor expansion in the neighborhood of $\mathbf{x}_\ell$ gives

$$\bar{u}(\mathbf{x}) = \bar{u}(\mathbf{x}_\ell) + \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot(\mathbf{x}-\mathbf{x}_\ell) + \mathcal{O}\left(\|\mathbf{x}-\mathbf{x}_\ell\|^2\right).$$

Replacing $\mathbf{x}$ respectively by $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_r$, $\mathbf{x}_s$, we obtain

$$\bar{u}(\mathbf{x}_i) = \bar{u}(\mathbf{x}_\ell) + \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{i\ell}^{\perp} + \mathcal{O}\left(h^2\right),$$

$$\bar{u}(\mathbf{x}_j) = \bar{u}(\mathbf{x}_\ell) + \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{j\ell}^{\perp} + \mathcal{O}\left(h^2\right),$$

$$\bar{u}(\mathbf{x}_r) = \bar{u}(\mathbf{x}_\ell) + \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{\ell r}^{\perp} + \mathcal{O}\left(h^2\right),$$

$$\bar{u}(\mathbf{x}_s) = \bar{u}(\mathbf{x}_\ell) + \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{\ell s}^{\perp} + \mathcal{O}\left(h^2\right).$$

Subtracting the last two equations we have

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{sr}^{\perp} = \bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r) + \mathcal{O}\left(h^2\right).$$

Therefore we obtain

$$\begin{cases} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{\ell i}^{\perp} = \bar{u}(\mathbf{x}_\ell) - \bar{u}(\mathbf{x}_i) + \mathcal{O}\left(h^2\right), \\ \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{j\ell}^{\perp} = \bar{u}(\mathbf{x}_j) - \bar{u}(\mathbf{x}_\ell) + \mathcal{O}\left(h^2\right), \\ \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{sr}^{\perp} = \bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r) + \mathcal{O}\left(h^2\right). \end{cases} \tag{3}$$

Let us first consider an *interior* primal face F_ℓ. We can decompose the unit vector $\mathbf{n}_{sr}$ in the basis $(\mathbf{n}_{\ell i}^{\perp}, \mathbf{n}_{sr}^{\perp})$ or $(\mathbf{n}_{j\ell}^{\perp}, \mathbf{n}_{sr}^{\perp})$:

$$\mathbf{n}_{sr} = \alpha_{i\ell}\mathbf{n}_{\ell i}^{\perp} + \beta_{i\ell}\mathbf{n}_{sr}^{\perp} = \alpha_{\ell j}\mathbf{n}_{j\ell}^{\perp} + \beta_{\ell j}\mathbf{n}_{sr}^{\perp},$$

with

$$\alpha_{i\ell} = \frac{1}{\mathbf{n}_{sr}\cdot\mathbf{n}_{\ell i}^{\perp}}, \quad \beta_{i\ell} = \frac{\mathbf{n}_{sr}\cdot\mathbf{n}_{\ell i}}{\mathbf{n}_{sr}^{\perp}\cdot\mathbf{n}_{\ell i}}, \quad \alpha_{\ell j} = \frac{1}{\mathbf{n}_{sr}\cdot\mathbf{n}_{j\ell}^{\perp}}, \quad \beta_{\ell j} = \frac{\mathbf{n}_{sr}\cdot\mathbf{n}_{j\ell}}{\mathbf{n}_{sr}^{\perp}\cdot\mathbf{n}_{j\ell}},$$

that is, in view of Fig. 1

$$\alpha_{i\ell} = \frac{1}{\cos(\theta_{i\ell})}, \quad \beta_{i\ell} = \frac{\sin(\theta_{i\ell})}{\cos(\theta_{i\ell})}, \quad \alpha_{\ell j} = \frac{1}{\cos(\theta_{\ell j})}, \quad \beta_{\ell j} = \frac{\sin(\theta_{\ell j})}{\cos(\theta_{\ell j})}. \tag{4}$$

According to assumption **H1** these values are well defined. Note that $\alpha_{i\ell} > 0$, $\alpha_{\ell j} > 0$ as soon as the centers $\mathbf{x}_i$ and $\mathbf{x}_j$ of the primal cells P_i and P_j are separated by the line corresponding to their face $F_\ell = P_i \cap P_j$. It may happen that $\mathbf{x}_i$ and $\mathbf{x}_j$ are *not* separated by the face F_ℓ. This is the case for a non-convex cell P_i if its mass center $\mathbf{x}_i$ is not inside P_i (see the left-hand side of Fig. 2). In such a case we replace $\mathbf{x}_i$ by the midpoint of an inner diagonal of P_i or by any interior point for which P_i is star-shaped (right-hand side of Fig. 2). Doing so, the inequalities $\alpha_{i\ell} > 0$, $\alpha_{\ell j} > 0$, which are mandatory to enforce the positiveness of the scheme (see Section 5), are always satisfied.
The gradients in the direction $\mathbf{n}_{sr}$ in the cells P_i and P_j then write

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{sr} = \alpha_{i\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{\ell i}^{\perp} + \beta_{i\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{sr}^{\perp},$$

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{sr} = \alpha_{\ell j}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{j\ell}^{\perp} + \beta_{\ell j}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{sr}^{\perp},$$

that is to say, using Taylor expansions (3)

$$\begin{cases} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{sr} = \alpha_{i\ell}\dfrac{\bar{u}(\mathbf{x}_\ell)-\bar{u}(\mathbf{x}_i)}{|G_{i\ell}|} + \beta_{i\ell}\dfrac{\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r)}{|F_\ell|} + \mathcal{O}(h), \\ \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{sr} = \alpha_{\ell j}\dfrac{\bar{u}(\mathbf{x}_j)-\bar{u}(\mathbf{x}_\ell)}{|G_{\ell j}|} + \beta_{\ell j}\dfrac{\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r)}{|F_\ell|} + \mathcal{O}(h). \end{cases} \tag{5}$$

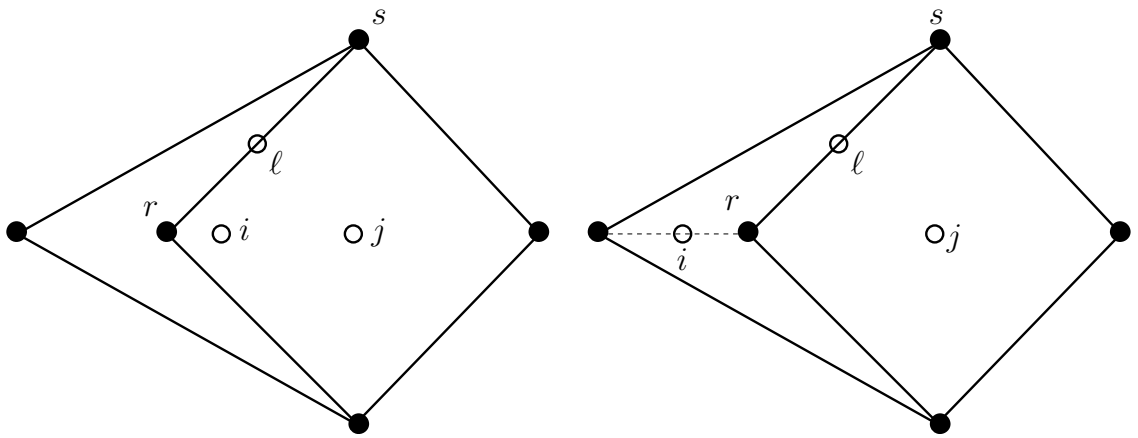


Figure 2: A non convex cell P_i and a convex cell P_j such that $\mathbf{x}_i$ and $\mathbf{x}_j$ are not separated by the line defined by face $F_\ell = \mathbf{x}_r\mathbf{x}_s$.

Note that these approximations can also be obtained by using the Green-Gauss formula applied to $\boldsymbol{\nabla}\bar{u}$ in diamond sub-cells $I_{i\ell}$ and $I_{\ell j}$

$$\begin{cases} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i = \dfrac{1}{|I_{i\ell}|}\displaystyle\int_{I_{i\ell}} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell) + \mathcal{O}(h) = \frac{1}{2}\frac{1}{|I_{i\ell}|}\left((\bar{u}(\mathbf{x}_\ell) - \bar{u}(\mathbf{x}_i))\mathbf{N}_{sr} + (\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r))\mathbf{N}_{i\ell}\right) + \mathcal{O}(h), \\ \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j = \dfrac{1}{|I_{\ell j}|}\displaystyle\int_{I_{\ell j}} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell) + \mathcal{O}(h) = \frac{1}{2}\frac{1}{|I_{\ell j}|}\left((\bar{u}(\mathbf{x}_j) - \bar{u}(\mathbf{x}_\ell))\mathbf{N}_{sr} + (\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r))\mathbf{N}_{\ell j}\right) + \mathcal{O}(h). \end{cases} \tag{6}$$

The fluxes can be *indifferently* estimated using one or the other of formulas (5), (6).

Let us now recall that the properties of (1) imply that the normal component of the flux is continuous across the primal face F_ℓ. We therefore impose

$$\kappa_{i\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i \cdot \mathbf{n}_{sr} = \kappa_{\ell j}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j \cdot \mathbf{n}_{sr}.$$

This gives

$$\bar{u}(\mathbf{x}_\ell) = \frac{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}|\bar{u}(\mathbf{x}_i) + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|\bar{u}(\mathbf{x}_j)}{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|} + \frac{|G_{i\ell}||G_{\ell j}|(\kappa_{\ell j}\beta_{\ell j} - \kappa_{i\ell}\beta_{i\ell})}{|F_\ell|(\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|)}\left(\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r)\right) + \mathcal{O}(h^2). \tag{7}$$

Inserting this value into one of the two equations of (5) results in

$$\kappa_{i\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i \cdot \mathbf{n}_{sr} = \kappa_{\ell j}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j \cdot \mathbf{n}_{sr}$$

$$= \frac{\kappa_{i\ell}\kappa_{\ell j}\alpha_{i\ell}\alpha_{\ell j}}{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|}\big(\bar{u}(\mathbf{x}_j) - \bar{u}(\mathbf{x}_i)\big) + \frac{\kappa_{i\ell}\kappa_{\ell j}(\alpha_{i\ell}\beta_{\ell j}|G_{\ell j}| + \alpha_{\ell j}\beta_{i\ell}|G_{i\ell}|)}{|F_\ell|(\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|)}\big(\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r)\big) + \mathcal{O}(h).$$

In view of this relation the numerical flux $\mathcal{F}_\ell$ through the primal face F_ℓ is then given by

$$\mathcal{F}_\ell(\mathbf{u}) = \gamma_\ell(u_j - u_i) + \delta_\ell(u_s - u_r), \tag{8}$$

with

$$\gamma_\ell = \frac{\kappa_{i\ell}\kappa_{\ell j}\alpha_{i\ell}\alpha_{\ell j}|F_\ell|}{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|}, \quad \delta_\ell = \frac{\kappa_{i\ell}\kappa_{\ell j}\left(\alpha_{i\ell}\beta_{\ell j}|G_{\ell j}| + \alpha_{\ell j}\beta_{i\ell}|G_{i\ell}|\right)}{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|}. \tag{9}$$

Since $\alpha_{i\ell} > 0$, $\alpha_{\ell j} > 0$, $\kappa_{i\ell} > 0$, $\kappa_{\ell j} > 0$ we clearly have $\gamma_\ell > 0$.

Consider now a *boundary* face F_ℓ. If $F_\ell \subset \Gamma_D$ we have $\bar{u}_\ell = g(\mathbf{x}_\ell)$. From (5) we then obtain

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i \cdot \mathbf{n}_{sr} = \alpha_{i\ell}\frac{g(\mathbf{x}_\ell) - \bar{u}(\mathbf{x}_i)}{|G_{i\ell}|} + \beta_{i\ell}\frac{\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r)}{|F_\ell|} + \mathcal{O}(h),$$

so that the Dirichlet boundary flux is defined by

$$\mathcal{F}_\ell(\mathbf{u}) = \gamma_\ell \left(g(\mathbf{x}_\ell) - u_i\right) + \delta_\ell(u_s - u_r), \tag{10}$$

with

$$\gamma_\ell = \kappa_{i\ell}\alpha_{i\ell}\frac{|F_\ell|}{|G_{i\ell}|}, \quad \delta_\ell = \kappa_{i\ell}\beta_{i\ell}.$$

If $F_\ell \subset \Gamma_N$, we have

$$\bar{\mathcal{F}}_\ell = \int_{F_\ell} \kappa\boldsymbol{\nabla}\bar{u} \cdot \mathbf{n}_{sr} = \int_{F_\ell} g,$$

so that the exact flux $\bar{\mathcal{F}}_\ell$ and the approximated one $\mathcal{F}_\ell$ are

$$\bar{\mathcal{F}}_\ell = |F_\ell| g(\mathbf{x}_\ell) + \mathcal{O}(h^2), \quad \mathcal{F}_\ell(\mathbf{u}) = |F_\ell| g(\mathbf{x}_\ell).$$

4 Dealing with vertex unknows

In order to evaluate the numerical fluxes $\mathcal{F}_\ell(\mathbf{u})$, Equations (8) and (10) require the knowledge of the values of u at the vertices $\mathbf{x}_r$ of the primal mesh. To compute these values, we propose to use two different methods. For the first one, described in Section 4.1, vertex values are calculated by interpolation while for the second one, described in Section 4.2, they are calculated as the solution to the same diffusion problem (1) discretized on the dual mesh.

4.1 Diamond type scheme

The first way to calculate the vertex values u_r is to use a polynomial approximation calculated using the cell-centered values u_i.
For a polynomial of degree 1, we have 3 coefficients to calculate, so at least 6 (3×dimension) neighboring cells of the cell are required for stability reason: see [14, 26][2]. When it is possible, the stencil will be centered on the cell, but the closer the cell is to the boundary of the domain or to the discontinuity of κ, the more the stencil will be shifted in order not to cross the discontinuity.

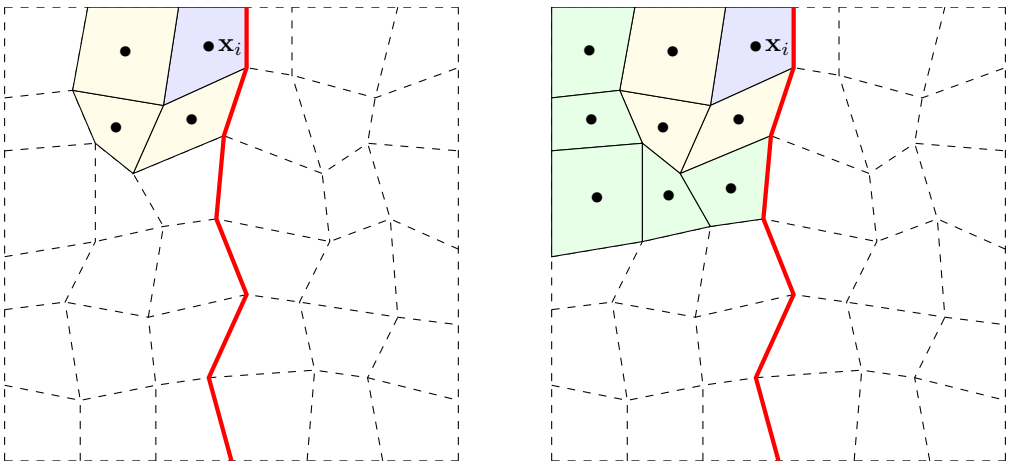


Figure 3: Construction of the stencil for the cell P_i with a discontinuity (in red).

[2]An example of the use of polynomials of degree 2 is also provided in the numerical experiments section.

To be more precise, the construction of the stencil of a cell P_i is illustrated on Fig. 3. We denote this stencil by $\mathcal{S}_i = \{P_0, ..., P_k\}$. For the sake of simplicity, we have assumed that the cells involved in the stencil have been renumbered. First the cell P_i itself (in blue) is added to the stencil and then we add the cells that share, at least, a vertex with the cell P_i (in yellow). If the number of cells we have already selected is not sufficient (in our case, 6 cells for a polynomial of order 1), we add the cells that have, at least, a vertex linked to the cells that we have just been added to the stencil (in green) and so on until we have enough cells. In all the above process, we impose that the stencil does not cross any discontinuity of κ (see Fig. 3).

Let $u_0, ..., u_k$ denote the $k+1$ values used for the calculation ($k \geq 5$). The polynomial is of the form

$$\mathcal{P}_i(\mathbf{x}) = a_{00}(u_0, ..., u_k) + a_{10}(u_0, ..., u_k)(x - x_i) + a_{01}(u_0, ..., u_k)(y - y_i),$$

and its coefficients a_{00}, a_{10}, a_{01} are chosen such that

$$\mathcal{P}_i(\mathbf{x}_0) = u_0, ..., \mathcal{P}_i(\mathbf{x}_k) = u_k.$$

This leads to the following system

$$\underbrace{\begin{pmatrix} 1 & x_0 - x_i & y_0 - y_i \\ \vdots & \vdots & \vdots \\ 1 & x_k - x_i & y_k - y_i \end{pmatrix}}_{=:\mathbf{M}} \underbrace{\begin{pmatrix} a_{00} \\ a_{10} \\ a_{01} \end{pmatrix}}_{=:\mathbf{a}} = \underbrace{\begin{pmatrix} u_0 \\ \vdots \\ u_k \end{pmatrix}}_{=:\mathbf{b}}.$$

Since matrix $\mathbf{M}$ has more rows than columns we have to use the least square method so that vector $\mathbf{a}$ is computed as a solution to the linear system: $\mathbf{M}^t\mathbf{M}\mathbf{a} = \mathbf{M}^t\mathbf{b}$.

In this process note that we do *not* enforce the continuity of u at the vertices. Indeed, a priori, $\mathcal{P}_i(\mathbf{x}_r) \neq \mathcal{P}_j(\mathbf{x}_r)$ for $i \neq j$.

We thus obtain expressions of the gradients in the direction $\mathbf{n}_{sr}$ in the cells P_i and P_j similar to (5)

$$\begin{cases} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i \cdot \mathbf{n}_{sr} = \alpha_{i\ell}\dfrac{\bar{u}(\mathbf{x}_\ell) - \bar{u}(\mathbf{x}_i)}{|G_{i\ell}|} + \beta_{i\ell}\dfrac{\mathcal{P}_i(\mathbf{x}_s) - \mathcal{P}_i(\mathbf{x}_r)}{|F_\ell|} + \mathcal{O}(h), \\ \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j \cdot \mathbf{n}_{sr} = \alpha_{\ell j}\dfrac{\bar{u}(\mathbf{x}_j) - \bar{u}(\mathbf{x}_\ell)}{|G_{\ell j}|} + \beta_{\ell j}\dfrac{\mathcal{P}_j(\mathbf{x}_s) - \mathcal{P}_j(\mathbf{x}_r)}{|F_\ell|} + \mathcal{O}(h). \end{cases} \tag{11}$$

Assuming the continuity of the flux $\mathcal{F}_\ell$ through the primal face F_ℓ

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i \cdot \mathbf{n}_{sr} = \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j \cdot \mathbf{n}_{sr},$$

provides

$$\bar{u}(\mathbf{x}_\ell) = \frac{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}|\bar{u}(\mathbf{x}_i) + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|\bar{u}(\mathbf{x}_j)}{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|} + \frac{|G_{i\ell}||G_{\ell j}|(\kappa_{\ell j}\beta_{\ell j}(\mathcal{P}_j(\mathbf{x}_s) - \mathcal{P}_j(\mathbf{x}_r)) - \kappa_{i\ell}\beta_{i\ell}(\mathcal{P}_i(\mathbf{x}_s) - \mathcal{P}_i(\mathbf{x}_r))}{|F_\ell|(\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|)} + \mathcal{O}(h^2).$$

In view of one of the two equations of (11), having inserting this value into it, the numerical flux $\mathcal{F}_\ell$ through the primal face F_ℓ results in

$$\mathcal{F}_\ell(\mathbf{u}) = \gamma_\ell(u_j - u_i) + \delta_{i\ell}(\mathcal{P}_i(\mathbf{x}_s) - \mathcal{P}_i(\mathbf{x}_r)) + \delta_{\ell j}(\mathcal{P}_j(\mathbf{x}_s) - \mathcal{P}_j(\mathbf{x}_r)),$$

with

$$\gamma_\ell = \frac{\kappa_{i\ell}\kappa_{\ell j}\alpha_{i\ell}\alpha_{\ell j}|F_\ell|}{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|},$$

$$\delta_{i\ell} = \frac{\kappa_{i\ell}\kappa_{\ell j}\alpha_{\ell j}\beta_{i\ell}|G_{i\ell}|}{|G_{\ell j}|\kappa_{i\ell}\alpha_{i\ell} + |G_{i\ell}|\kappa_{\ell j}\alpha_{\ell j}}, \quad \delta_{\ell j} = \frac{\kappa_{i\ell}\kappa_{\ell j}\alpha_{i\ell}\beta_{\ell j}|G_{\ell j}|}{|G_{\ell j}|\kappa_{i\ell}\alpha_{i\ell} + |G_{i\ell}|\kappa_{\ell j}\alpha_{\ell j})},$$

so that the diamond scheme writes

$$\begin{cases} -\sum_{\ell\in i,\ell\notin\partial\Omega} (\gamma_\ell(u_j - u_i) + \delta_{i\ell}(\mathcal{P}_i(\mathbf{x}_s) - \mathcal{P}_i(\mathbf{x}_r)) + \delta_{\ell j}(\mathcal{P}_j(\mathbf{x}_s) - \mathcal{P}_j(\mathbf{x}_r))) \\ -\sum_{\ell\in i,\ell\in\partial\Omega} (\gamma_\ell(u_\ell - u_i) + \delta_{i\ell}(\mathcal{P}_i(\mathbf{x}_s) - \mathcal{P}_i(\mathbf{x}_r))) + |P_i|\lambda_i u_i = |P_i|f_i, \\ u_\ell = g(\mathbf{x}_\ell) \qquad \mathbf{x}_\ell \in \Gamma_D, \\ \gamma_\ell(u_\ell - u_i) + \delta_{i\ell}(\mathcal{P}_i(\mathbf{x}_s) - \mathcal{P}_i(\mathbf{x}_r)) = |F_\ell| g(\mathbf{x}_\ell) \qquad \mathbf{x}_\ell \in \Gamma_N. \end{cases} \tag{12}$$

4.2 DDFV scheme

The second way to calculate the vertex values u_r is to consider them as additional unknowns that are solutions to problem (1) integrated on each cell of the dual mesh, thus following [21]. We have

$$-\int_{D_r} \boldsymbol{\nabla}\cdot(\kappa\boldsymbol{\nabla}\bar{u}) + \int_{D_r}\lambda\bar{u} = \int_{D_r} f,$$

that is, thanks to the divergence theorem

$$-\sum_{\ell\in r}\int_{G_\ell}\kappa\boldsymbol{\nabla}\bar{u}\cdot\mathbf{n} - \int_{D_r\cap\partial\Omega}\kappa\boldsymbol{\nabla}\bar{u}\cdot\mathbf{n} + \int_{D_r}\lambda\bar{u} = \int_{D_r} f,$$

where the compact notation $\sum_{\ell\in r}$ stands for the sum on all the faces $G_\ell = G_{i\ell}\cup G_{\ell j}$ (if $\mathbf{x}_\ell\notin\partial\Omega$) or $G_\ell = G_{i\ell}$ (if $\mathbf{x}_\ell\in\partial\Omega$) of the dual cell D_r. We obtain

$$-\sum_{\ell\in r}\int_{G_{i\ell}}\kappa_{i\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{i\ell} - \sum_{\ell\in r,\ell\notin\partial\Omega}\int_{G_{\ell j}}\kappa_{j\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{\ell j} - \int_{D_r\cap\partial\Omega}\kappa\boldsymbol{\nabla}\bar{u}\cdot\mathbf{n} + \int_{D_r}\lambda\bar{u} = \int_{D_r} f + \mathcal{O}\left(h^3\right).$$

We decompose the unit vector $\mathbf{n}_{i\ell}$ (resp. $\mathbf{n}_{\ell j}$) in the basis $(\mathbf{n}_{sr}^\perp, \mathbf{n}_{i\ell}^\perp)$ (resp. $(\mathbf{n}_{sr}^\perp, \mathbf{n}_{\ell j}^\perp)$)

$$\mathbf{n}_{i\ell} = \alpha_{i\ell}\mathbf{n}_{sr}^\perp - \beta_{i\ell}\mathbf{n}_{i\ell}^\perp, \quad \mathbf{n}_{\ell j} = \alpha_{\ell j}\mathbf{n}_{sr}^\perp - \beta_{\ell j}\mathbf{n}_{\ell j}^\perp,$$

where the values

$$\alpha_{i\ell} = \frac{1}{\mathbf{n}_{sr}^\perp\cdot\mathbf{n}_{i\ell}}, \quad \beta_{i\ell} = -\frac{\mathbf{n}_{sr}\cdot\mathbf{n}_{i\ell}}{\mathbf{n}_{sr}\cdot\mathbf{n}_{i\ell}^\perp}, \quad \alpha_{\ell j} = \frac{1}{\mathbf{n}_{sr}^\perp\cdot\mathbf{n}_{\ell j}}, \quad \beta_{\ell j} = -\frac{\mathbf{n}_{sr}\cdot\mathbf{n}_{\ell j}}{\mathbf{n}_{sr}\cdot\mathbf{n}_{\ell j}^\perp},$$

coincide with those of (4): see Fig. 1. We obtain

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{i\ell} = \alpha_{i\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{sr}^\perp - \beta_{i\ell}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{i\ell}^\perp,$$

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{\ell j} = \alpha_{\ell j}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{sr}^\perp - \beta_{\ell j}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{\ell j}^\perp,$$

that is to say, using Taylor expansions (3)

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_i\cdot\mathbf{n}_{i\ell} = \alpha_{i\ell}\frac{\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r)}{|F_\ell|} + \beta_{i\ell}\frac{\bar{u}(\mathbf{x}_\ell) - \bar{u}(\mathbf{x}_i)}{|G_{i\ell}|} + \mathcal{O}(h),$$

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)_j\cdot\mathbf{n}_{\ell j} = \alpha_{\ell j}\frac{\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r)}{|F_\ell|} + \beta_{\ell j}\frac{\bar{u}(\mathbf{x}_j) - \bar{u}(\mathbf{x}_\ell)}{|G_{\ell j}|} + \mathcal{O}(h).$$

Note that these approximations derive directly from the Green-Gauss formula applied to $\boldsymbol{\nabla}\bar{u}$ in diamond subcells $I_{i\ell}$ and $I_{\ell j}$ already given by (6).

Suppose that $\mathbf{x}_\ell \notin \partial\Omega$. In view of the previous equations and of (7), the numerical flux $\mathcal{G}_\ell$ through the common boundary $G_\ell = G_{i\ell} \cup G_{\ell j}$ of the dual cells D_r and D_s is given by

$$\mathcal{G}_\ell(\mathbf{u}) = \Gamma_\ell(u_j - u_i) + \Delta_\ell(u_s - u_r),$$

with

$$\left\{\begin{array}{l} \Gamma_\ell = \dfrac{\kappa_{i\ell}\kappa_{\ell j}(\alpha_{i\ell}\beta_{\ell j} + \alpha_{\ell j}\beta_{i\ell})}{\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|}, \\ \\ \Delta_\ell = \dfrac{(\kappa_{i\ell}\alpha_{i\ell}^2 - \kappa_{i\ell}\beta_{i\ell}^2 + \kappa_{\ell j}\alpha_{i\ell}\alpha_{\ell j} + \kappa_{\ell j}\beta_{i\ell}\beta_{\ell j})\kappa_{i\ell}|G_{\ell j}|}{|F_\ell|\,(\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|)} \\ \\ \qquad + \dfrac{(\kappa_{\ell j}\alpha_{\ell j}^2 - \kappa_{\ell j}\beta_{\ell j}^2 + \kappa_{i\ell}\alpha_{i\ell}\alpha_{\ell j} + \kappa_{i\ell}\beta_{i\ell}\beta_{\ell j})\kappa_{\ell j}|G_{i\ell}|}{|F_\ell|\,(\kappa_{i\ell}\alpha_{i\ell}|G_{\ell j}| + \kappa_{\ell j}\alpha_{\ell j}|G_{i\ell}|)}. \end{array}\right. \tag{13}$$

If $\mathbf{x}_\ell \in \partial\Omega$, the common boundary of the dual cells D_r and D_s is $G_\ell = G_{i\ell}$ and the numerical flux $\mathcal{G}_\ell$ through G_ℓ is given by

$$\mathcal{G}_\ell(\mathbf{u}) = \Gamma_\ell(u_\ell - u_i) + \Delta_\ell(u_s - u_r),$$

with

$$\Gamma_\ell = \frac{\kappa_{i\ell}\beta_{i\ell}}{|G_{i\ell}|}, \quad \Delta_\ell = \frac{\kappa_{i\ell}\alpha_{i\ell}}{|F_\ell|}. \tag{14}$$

If $\mathbf{x}_r \in \Gamma_N$, the boundary dual flux is

$$\int_{D_r\cap\partial\Omega} \kappa\nabla\bar{u}.\bar{n} = \int_{D_r\cap\partial\Omega} g,$$

the right hand side of which is approximated by (see Fig. 4)

$$|D_r \cap \partial\Omega| g(\mathbf{x}_r) = \frac{1}{2}(|F_\ell| + |F_k|)g(\mathbf{x}_r).$$

Recalling that δ_ℓ, γ_ℓ are defined by (9) while Γ_ℓ, Δ_ℓ are defined by (13) or (14), the DDFV scheme thus writes

$$\left\{\begin{array}{ll} -\displaystyle\sum_{\ell\in i,\ell\notin\partial\Omega} \gamma_\ell(u_j - u_i) - \sum_{\ell\in i,\ell\in\partial\Omega} \gamma_\ell(u_\ell - u_i) - \sum_{\ell\in i} \delta_\ell(u_s - u_r) + |P_i|\lambda_i u_i = |P_i| f_i & \\ \\ -\displaystyle\sum_{\ell\in r,\ell\notin\partial\Omega} \Gamma_\ell(u_j - u_i) - \sum_{\ell\in r,\ell\in\partial\Omega} \Gamma_\ell(u_\ell - u_i) - \sum_{\ell\in r} \Delta_\ell(u_s - u_r) & \\ \qquad + |D_r|\lambda_r u_r = |D_r| f_r + |D_r\cap\partial\Omega| g(\mathbf{x}_r) & \mathbf{x}_r \notin \Gamma_D, \\ u_\ell = g(\mathbf{x}_\ell) & \mathbf{x}_\ell \in \Gamma_D, \\ u_r = g(\mathbf{x}_r) & \mathbf{x}_r \in \Gamma_D, \\ \gamma_\ell(u_\ell - u_i) + \delta_\ell(u_s - u_r) = |F_\ell| g(\mathbf{x}_\ell) & \mathbf{x}_\ell \in \Gamma_N. \end{array}\right. \tag{15}$$

Note that $\Delta_\ell > 0$: the proof is given in [22] in the general case where $\boldsymbol{\kappa}$ is a positive definite matrix.

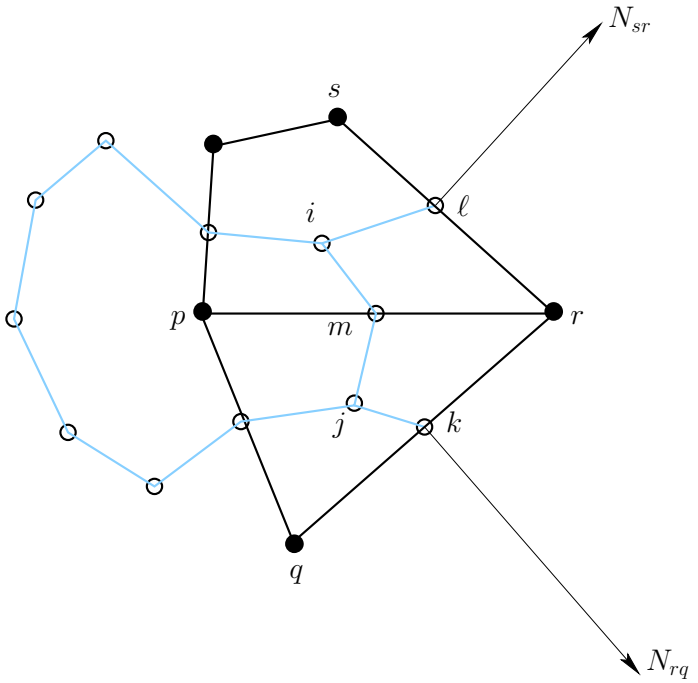


Figure 4: Two primal cells P_i, P_j (black lines), one *interior* dual cell D_p (blue lines) and one *boundary* dual cell $D_r = \mathbf{x}_r\mathbf{x}_\ell\mathbf{x}_i\mathbf{x}_m\mathbf{x}_j\mathbf{x}_k$ (blue lines) such that $D_r \cap \partial\Omega = \mathbf{x}_k\mathbf{x}_r \cup \mathbf{x}_r\mathbf{x}_\ell$.

5 A method to make the schemes monotonic

In this section we propose to find a method for the previous methods to be made monotonic. A method borrowed from [44, 19, 49, 18] and developed in the framework of 2D diffusion on arbitrary meshes can be used. For any value r we will use the common notation $r = r^+ - r^-$ with

$$r^+ = \frac{1}{2}(|r| + r) \geq 0, \quad r^- = \frac{1}{2}(|r| - r) \geq 0.$$

The numerical primal and dual fluxes $\mathcal{F}_\ell$ and $\mathcal{G}_\ell$ read as

$$\mathcal{F}_\ell(\mathbf{u}) = \gamma_\ell(u_j - u_i) + r_\ell(\mathbf{u}), \quad \mathcal{G}_\ell(\mathbf{u}) = \Delta_\ell(u_s - u_r) + R_\ell(\mathbf{u}),$$

where

$$r_\ell(\mathbf{u}) = \delta_\ell(u_s - u_r), \quad R_\ell(\mathbf{u}) = \Gamma_\ell(u_j - u_i).$$

Suppose that, for all i and r, $u_i > 0$ and $u_r > 0$, we can set

$$\mathcal{F}_\ell(\mathbf{u}) = \left(\gamma_\ell + \frac{r_\ell^+(\mathbf{u})}{u_j}\right) u_j - \left(\gamma_\ell + \frac{r_\ell^-(\mathbf{u})}{u_i}\right) u_i, \quad \mathcal{G}_\ell(\mathbf{u}) = \left(\Delta_\ell + \frac{R_\ell^+(\mathbf{u})}{u_s}\right) u_s - \left(\Delta_\ell + \frac{R_\ell^-(\mathbf{u})}{u_r}\right) u_r.$$

As $\gamma_\ell > 0$, $\Delta_\ell > 0$ we end up with two points primal and dual flux approximations with *positive* coefficients. The diamond scheme (12) then rewrites

$$
\left\{
\begin{array}{ll}
-\displaystyle\sum_{\ell\in i,\ell\notin\partial\Omega}\left(\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u})}{u_j}\right)u_j-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u})}{u_i}\right)u_i\right) & \\
-\displaystyle\sum_{\ell\in i,\ell\in\partial\Omega}\left(\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u})}{u_\ell}\right)u_\ell-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u})}{u_i}\right)u_i\right)+|P_i|\lambda_i u_i=|P_i|f_i, & \\
u_\ell=g(\mathbf{x}_\ell) & \mathbf{x}_\ell\in\Gamma_D,\\
\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u})}{u_\ell}\right)u_\ell-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u})}{u_i}\right)u_i=|F_\ell|g(\mathbf{x}_\ell) & \mathbf{x}_\ell\in\Gamma_N,
\end{array}
\right.
$$

while the DDFV scheme (15) rewrites

$$
\left\{
\begin{array}{ll}
-\displaystyle\sum_{\ell\in i,\ell\notin\partial\Omega}\left(\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u})}{u_j}\right)u_j-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u})}{u_i}\right)u_i\right) & \\
-\displaystyle\sum_{\ell\in i,\ell\in\partial\Omega}\left(\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u})}{u_\ell}\right)u_\ell-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u})}{u_i}\right)u_i\right)+|P_i|\lambda_i u_i=|P_i|f_i, & \\
-\displaystyle\sum_{\ell\in r}\left(\left(\Delta_\ell+\frac{R_\ell^+(\mathbf{u})}{u_s}\right)u_s-\left(\Delta_\ell+\frac{R_\ell^-(\mathbf{u})}{u_r}\right)u_r\right)+|D_r|\lambda_r u_r & \\
\qquad=|D_r|f_r+|D_r\cap\partial\Omega|g(\mathbf{x}_r), & \mathbf{x}_r\notin\Gamma_D,\\
u_\ell=g(\mathbf{x}_\ell) & \mathbf{x}_\ell\in\Gamma_D,\\
u_r=g(\mathbf{x}_r) & \mathbf{x}_r\in\Gamma_D,\\
\left(\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u})}{u_\ell}\right)u_\ell-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u})}{u_i}\right)u_i\right)=|F_\ell|g(\mathbf{x}_\ell) & \mathbf{x}_\ell\in\Gamma_N.
\end{array}
\right.
\tag{16}
$$

The matrices associated with these systems are not symmetric and depend respectively on u_i, u_ℓ ($\ell\in\partial\Omega$) and u_r. More details about this are given in the following section.

5.1 Matrix form

Denoting

$$
\begin{array}{lll}
\mathbf{u}^{primal}=(u_i)_{1\le i\le n}, & \mathbf{u}^{dual}=(u_r)_{1\le r\le m}, & \mathbf{u}=(\mathbf{u}^{primal},\mathbf{u}^{dual}),\\
\mathbf{b}^{primal}=(b_i)_{1\le i\le n}, & \mathbf{b}^{dual}=(b_r)_{1\le r\le m}, & \mathbf{b}=(\mathbf{b}^{primal},\mathbf{b}^{dual}),
\end{array}
\tag{17}
$$

and

$$
\left\{
\begin{array}{ll}
b_i^{primal}=|P_i|f_i+\displaystyle\sum_{\ell\in i,\ell\in\Gamma_D}\left(r_\ell(\mathbf{u}^{dual})^+\gamma_\ell g(\mathbf{x}_\ell)\right)+\sum_{\ell\in i,\ell\in\Gamma_N}|F_\ell|g(\mathbf{x}_\ell), & \\
b_r^{dual}=|D_r|f_r+|D_r\cap\partial\Omega|g(\mathbf{x}_\ell) & \mathbf{x}_r\notin\Gamma_D,\\
b_r^{dual}=|D_r|f_r+\zeta g(\mathbf{x}_\ell) & \mathbf{x}_r\in\Gamma_D,
\end{array}
\right.
\tag{18}
$$

where ζ is a large value dedicated to taking into account of the Dirichlet boundary conditions by penalization (for example $\zeta=10^{12}$), system (16) then rewrites under the more compact form

$$
\mathbf{A}(\mathbf{u})\mathbf{u}=\begin{pmatrix}\mathbf{A}^{primal}(\mathbf{u}^{primal},\mathbf{u}^{dual}) & \mathbf{0}\\ \mathbf{0} & \mathbf{A}^{dual}(\mathbf{u}^{primal},\mathbf{u}^{dual})\end{pmatrix}\begin{pmatrix}\mathbf{u}^{primal}\\ \mathbf{u}^{dual}\end{pmatrix}=\begin{pmatrix}\mathbf{b}^{primal}\\ \mathbf{b}^{dual}\end{pmatrix}=\mathbf{b},
\tag{19}
$$

with

$$\begin{cases} A_{ii}^{primal}(\mathbf{u}^{primal},\mathbf{u}^{dual}) = \displaystyle\sum_{\ell\in i,\ell\notin\Gamma_N}\left(\gamma_\ell + \frac{r_\ell(\mathbf{u}^{dual})^-}{u_i}\right) + |P_i|\lambda_i, & \\ A_{ij}^{primal}(\mathbf{u}^{primal},\mathbf{u}^{dual}) = -\left(\gamma_\ell + \frac{r_\ell(\mathbf{u}^{dual})^+}{u_j}\right) & \mathbf{x}_i \neq \mathbf{x}_j, \\ A_{rr}^{dual}(\mathbf{u}^{primal},\mathbf{u}^{dual}) = \displaystyle\sum_{\ell\in r,\ell\notin\Gamma_N}\left(\Delta_\ell + \frac{R_\ell(\mathbf{u}^{primal})^-}{u_r}\right) + |D_r|\lambda_r & \mathbf{x}_r \notin \Gamma_D, \\ A_{rr}^{dual}(\mathbf{u}^{primal},\mathbf{u}^{dual}) = \displaystyle\sum_{\ell\in r,\ell\notin\Gamma_N}\left(\Delta_\ell + \frac{R_\ell(\mathbf{u}^{primal})^-}{u_r}\right) + |D_r|\lambda_r + \zeta & \mathbf{x}_r \in \Gamma_D, \\ A_{rs}^{dual}(\mathbf{u}^{primal},\mathbf{u}^{dual}) = -\left(\Delta_\ell + \frac{R_\ell(\mathbf{u}^{primal})^+}{u_s}\right) & \mathbf{x}_r \neq \mathbf{x}_s, \mathbf{x}_\ell \notin \Gamma_N. \end{cases} \tag{20}$$

Thus the monotonicity enforcing procedure leads to two decoupled sparse matrices of size $m \times m$ and $n \times n$ depending on $\mathbf{u}$. This is a significant difference with the usual DDFV scheme for which all degrees of freedom are coupled, leading to a single $(m+n)\times(m+n)$ matrix independent of $\mathbf{u}$.
In the case of the monotonic diamond method, we obtain the system

$$\mathbf{A}^{diamond}(\mathbf{u}^{primal})\mathbf{u}^{primal} = \mathbf{b}^{diamond}, \tag{21}$$

with

$$\begin{cases} A_{ii}^{diamond}(\mathbf{u}^{primal}) = \displaystyle\sum_{\ell\in i,\ell\notin\Gamma_N}\left(\gamma_\ell + \frac{r_\ell(\mathbf{u}^{primal})^-}{u_i}\right) + |P_i|\lambda_i, & \\ A_{ij}^{diamond}(\mathbf{u}^{primal}) = -\displaystyle\sum_{\ell\in i\cap j}\left(\gamma_\ell + \frac{r_\ell(\mathbf{u}^{primal})^+}{u_j}\right) & \mathbf{x}_i \neq \mathbf{x}_j, \end{cases} \tag{22}$$

and

$$b_i^{diamond} = |P_i|f_i + \sum_{\ell\in i,\ell\in\Gamma_D}\left(r_\ell(\mathbf{u}^{primal})^+ + \gamma_\ell g(\mathbf{x}_\ell)\right) + \sum_{\ell\in i,\ell\in\Gamma_N}|F_\ell|g(\mathbf{x}_\ell). \tag{23}$$

Remark 5.1. *Assuming that $f \geq 0$ and $g \geq 0$, all the components of the right hand side $\mathbf{b}$ are non-negative. Assuming moreover that f and g are not zero, then at least one component of $\mathbf{b}$ is positive.*

5.2 Picard iteration method

Both systems (19) and (21) are of the form $\mathbf{A}(\mathbf{u})\mathbf{u} = \mathbf{b}$. In order to solve them, we use a Picard iteration method. We start with an initial guess $\mathbf{u}^0 > 0$, compute the matrix $\mathbf{A}(\mathbf{u}^0)$ and solve $\mathbf{A}(\mathbf{u}^0)\mathbf{u}^1 = \mathbf{b}$. Repeating this process, we build a sequence $(\mathbf{u}^\nu)$ that, if it converges to a positive vector, tends to a solution of the scheme. We stop the algorithm when the difference $\mathbf{u}^{\nu+1} - \mathbf{u}^\nu$ between two successive iterates is small enough. To summarize, the following algorithm is used

$\nu = 0$
$\mathbf{A}(\mathbf{u}^0)\mathbf{u}^1 = \mathbf{b}$
do while $\epsilon\|\mathbf{u}^\nu\|_2 < \|\mathbf{u}^{\nu+1} - \mathbf{u}^\nu\|_2$
  $\mathbf{A}(\mathbf{u}^\nu)\mathbf{u}^{\nu+1} = \mathbf{b}$
  $\nu = \nu + 1$
enddo

For the monotonic DDFV scheme (16), for example, the linear system $\mathbf{A}(\mathbf{u}^\nu)\mathbf{u}^{\nu+1} = \mathbf{b}$ writes

$$\begin{cases} -\displaystyle\sum_{\ell\in i,\ell\notin\partial\Omega}\left(\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u}^\nu)}{u_j^\nu}\right)u_j^{\nu+1}-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u}^\nu)}{u_i^\nu}\right)u_i^{\nu+1}\right) & \\ -\displaystyle\sum_{\ell\in i,\ell\in\partial\Omega}\left(\left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u}^\nu)}{u_\ell^\nu}\right)u_\ell^{\nu+1}-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u}^\nu)}{u_i^\nu}\right)u_i^{\nu+1}\right)+|P_i|\lambda_i u_i^{\nu+1}=|P_i|f_i, & \\ -\displaystyle\sum_{\ell\in r}\left(\left(\Delta_\ell+\frac{R_\ell^+(\mathbf{u}^\nu)}{u_s^\nu}\right)u_s^{\nu+1}-\left(\Delta_\ell+\frac{R_\ell^-(\mathbf{u}^\nu)}{u_r^\nu}\right)u_r^{\nu+1}\right)+|D_r|\lambda_r u_r^{\nu+1} & \\ \qquad =|D_r|f_r+|D_r\cap\partial\Omega|g(\mathbf{x}_r) & \mathbf{x}_r\notin\Gamma_D, \\ u_\ell^{\nu+1}=g(\mathbf{x}_\ell) & \mathbf{x}_\ell\in\Gamma_D, \\ u_r^{\nu+1}=g(\mathbf{x}_r) & \mathbf{x}_r\in\Gamma_D, \\ \left(\gamma_\ell+\frac{r_\ell^+(\mathbf{u}^\nu)}{u_\ell^\nu}\right)u_\ell^{\nu+1}-\left(\gamma_\ell+\frac{r_\ell^-(\mathbf{u}^\nu)}{u_i^\nu}\right)u_i^{\nu+1}=|F_\ell|g(\mathbf{x}_\ell) & \mathbf{x}_\ell\in\Gamma_N. \end{cases} \tag{24}$$

Unfortunately, we are unable to prove that the above algorithm converges. Nevertheless, we prove in Section 6.2 below that the scheme is well defined at each iteration of the algorithm, as soon as the initial guess $\mathbf{u}^0$ is positive. Furthermore we prove in section 6.3 that the solution of the *usual* DDFV scheme (15) is *close* (in some sense) to the solution of the *monotonic* DDFV scheme (16).

6 Properties

6.1 Monotonicity

Consider the definition of an M-matrix (see for instance [35])

Definition 6.1. *An $n\times n$ matrix $\mathbf{A}$ that can be expressed in the forme $\mathbf{A}=s\mathbf{I}-\mathbf{B}$, where $\mathbf{B}=(b_{ij})_{1\le i,j\le n}$ with $b_{ij}\ge 0$, $1\le i,j\le n$, and $s\ge\rho(\mathbf{B})$, the maximum of the moduli of the eigenvalues of $\boldsymbol{B}$, is called an M-matrix.*

We use the following lemma

Lemma 6.2. *A matrix $\mathbf{A}=(A_{ij})_{1\le i,j\le n}$ is an M-matrix if it satisfies the following inequalities*

$$\forall i\neq j,\quad A_{ij}\le 0,\quad \text{and}\quad \forall i,\quad \sum_{j=1}^n A_{ij}\ge 0.$$

Moreover, if the last inequality is strict, we say that $\mathbf{A}$ is a strict M-matrix.

Proposition 6.3. *Assume that $\mathbf{u}>\mathbf{0}$. Then the matrices $\mathbf{A}^{primal}$ and $\mathbf{A}^{dual}$ defined by (20) and the matrix $\mathbf{A}^{diamond}$ defined by (22) are such that $(\mathbf{A}^{primal})^t$, $(\mathbf{A}^{dual})^t$ and $(\mathbf{A}^{diamond})^t$ are strict M-matrices.*

Proof. The matrix $\mathbf{A}^{primal}$ satisfies

$$\forall i\neq j,\quad A_{ij}^{primal}\le 0\quad\text{and}\quad \forall j,\quad \sum_{i=1}^n A_{ij}^{primal}>0.$$

Indeed we have, for all j

$$\sum_{i=1}^n A_{ij}^{primal}=\sum_{i=1}^n\left(\sum_{\ell\in i,\ell\notin\Gamma_N}\left(\gamma_\ell+\frac{r_\ell(\mathbf{u}^{dual})^-}{u_i}\right)-\sum_{\ell\in i\cap j}\left(\gamma_\ell+\frac{r_\ell(\mathbf{u}^{dual})^+}{u_j}\right)\right)+\lambda_j|P_j|.$$

Thanks to conservativity, only the boundary terms and the mass term remain, for all j

$$\sum_{i=1}^{n} A_{ij}^{primal} = \sum_{i=1}^{n} \sum_{\ell \in (i \cap \Gamma_D)} \left(\gamma_\ell + \frac{r_\ell(\mathbf{u}^{dual})^-}{u_i} \right) + \lambda_j |P_j| > 0.$$

The above argument has been carried out on $\mathbf{A}^{primal}$ but the proof applies *mutatis mutandis* for $\mathbf{A}^{dual}$ or $\mathbf{A}^{diamond}$. □

Remark 6.4. *According to* (19)*, it is sufficient to prove that* $\mathbf{A}^{primal}$ *and* $\mathbf{A}^{dual}$ *are both strict M-matrices to prove that* $\mathbf{A}$ *is a strict M-matrix.*

Theorem 6.5. *Assume that* $f > 0$ *and* $g > 0$*. Let* $\mathbf{A}$ *and* $\mathbf{b}$ *be defined by* (18)-(20) *or* (22)-(23)*. Then* $\mathbf{A}^{-1}\mathbf{b} = \mathbf{u} \geq \mathbf{0}$.

Proof. As $\mathbf{A}^t$ is a strict M-matrix $\mathbf{A}$ is invertible and its inverse has only non-negative entries (see for example [40], Corollary 3.20). In view of Remark 5.1, the right hand side is non-negative, hence $\mathbf{u} = \mathbf{A}^{-1}\mathbf{b} \geq \mathbf{0}$. □

6.2 Well-posedness of the Picard iteration method

Proposition 6.6. *Assume that* $f \geq 0$*,* $g \geq 0$*, and either* $\|f\|_{L^2(\Omega)} > 0$ *or* $\|g\|_{L^2(\partial\Omega)} > 0$*. Assume moreover that* $\mathbf{u}^0 > \mathbf{0}$*. Then for all* ν*,* $\mathbf{u}^\nu > \mathbf{0}$.

To prove this property, we need to introduce the concept of irreducible matrix. We quote here [40, Definition 1.15].

Definition 6.7. *An* $n \times n$ *matrix* $\mathbf{A}$ *is* ***reducible*** *if there exits an* $n \times n$ *permutation matrix* $\mathbf{P}$ *such that*

$$\mathbf{P}\mathbf{A}\mathbf{P}^t = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A}_{11}$*,* $\mathbf{A}_{12}$*,* $\mathbf{A}_{22}$ *are respectively* $r \times r$*,* $r \times (n-r)$ *and* $(n-r) \times (n-r)$ *sub-matrices with* $1 \leq r < n$*. If no such permutation matrix exists, then* $\mathbf{A}$ *is* ***irreducible****.*

The matrices $\mathbf{A}^{primal}$, $\mathbf{A}^{dual}$ defined by (20) and the matrix $\mathbf{A}^{diamond}$ defined by (22) are irreducible thanks to the following Lemma (see [40, Theorem 1.17]).

Lemma 6.8. *To any* $n \times n$ *matrix* $\mathbf{A}$ *we associate the graph of nodes* $1, 2, ..., n$ *and of directed edges connecting* $\mathbf{x}_i$ *to* $\mathbf{x}_j$ *if* $A_{ij} \neq 0$*. Then* $\mathbf{A}$ *is irreducible if and only if for any pair* $i \neq j$ *there exists a chain of edges that allows to go from* $\mathbf{x}_i$ *to* $\mathbf{x}_j$*,*

$$A_{i,k1} \neq 0 \to A_{k1,k2} \neq 0 \to \cdots \to A_{km,j} \neq 0.$$

With these definitions we can make use of the following theorem (see [40], Corollary 3.20).

Theorem 6.9. *If* $\mathbf{A}$ *is an irreducible strict M-matrix, then it is invertible and, for all* i*,* j *(*$1 \leq i, j \leq n$*),* $(A^{-1})_{ij} > 0$.

We are now in position to prove Proposition 6.6.

Proof of Proposition 6.6. We argue by induction on the index ν. We assume that $\mathbf{u}^\nu > 0$. Hence $(\mathbf{A}^{primal}(\mathbf{u}^\nu))^t$ is a strict M-matrix (see Proposition 6.3). It is easy to check that $(\mathbf{A}^{primal}(\mathbf{u}^\nu))^t$ is also irreducible. Thus, applying Theorem 6.9, all the entries of $(\mathbf{A}^{primal}(\mathbf{u}^\nu))^{-t}$ are positive. Consequently, all the entries of $(\mathbf{A}^{primal}(\mathbf{u}^\nu))^{-1}$ are positive. Using Remark 5.1, we know that all components of $\mathbf{b}$ are non-negative. Moreover, because of the assumption that either $\|f\|_{L^2(\Omega)} > 0$ or $\|g\|_{L^2(\partial\Omega)} > 0$, at least one component of $\mathbf{b}$ is positive. We thus have, for all i ($1 \leq i \leq n$)

$$u_i^{\nu+1} = \sum_{j=1}^{n} (\mathbf{A}^{primal}(\mathbf{u}^\nu))_{ij}^{-1} b_j > 0,$$

since all terms of this sum are non-negative, with one at least that does not vanish.
The above proof has been carried out on $\mathbf{A}^{primal}$ but the same argument applies for $\mathbf{A}^{dual}$ or $\mathbf{A}^{diamond}$. □

Proposition 6.6 shows that the condition $\mathbf{u}^\nu > \mathbf{0}$ remains satisfied during the Picard iteration method, which allows to define $\mathbf{A}^{primal}(\mathbf{u}^\nu)$ for all $\nu \geq 0$.

6.3 About the convergence of the fixed-point for the monotonic DDFV scheme

Recall that

- $\bar{\mathbf{u}} = ((\bar{u}_i)_{1\le i\le n}, (\bar{u}_r)_{1\le r\le m})$ is the *exact* solution of (1),
- $\mathbf{u} = ((u_i)_{1\le i\le n}, (u_r)_{1\le r\le m})$ is the *DDFV* solution defined by (15),
- $\mathbf{u}^\nu = ((u_i^\nu)_{1\le i\le n}, (u_r^\nu)_{1\le r\le m}))$ is the ν-th iterate associated with the *monotonic DDFV* scheme, that is, the solution to (24).

We will make use of the following theorem, the proof of which is postponed to Appendix A.

Theorem 6.10. *Under assumptions* **H1**, **H2**, **H3** *the DDFV scheme defined by* (15) *is first-order accurate in the discrete L^2 norm, that is, there exists a constant C_1 independent of h such that*

$$\|\bar{\mathbf{u}} - \mathbf{u}\|_2 = \left(\sum_i |P_i|(\bar{u}(\mathbf{x}_i) - u_i)^2 + \sum_r |D_r|(\bar{u}(\mathbf{x}_r) - u_r)^2\right)^{1/2} \le C_1 h.$$

We will need the following lemma to prove Theorem 6.12.

Lemma 6.11. *Assume that there exists $\nu > 0$ and $\epsilon > 0$ such that*

$$\max\left(\max_i \left|\frac{u_i^{\nu+1} - u_i^\nu}{u_i^\nu}\right|, \max_r \left|\frac{u_r^{\nu+1} - u_r^\nu}{u_r^\nu}\right|\right) \le \epsilon. \tag{25}$$

Then the monotonic DDFV scheme (24) *writes*

$$\begin{cases} -\displaystyle\sum_{\ell\in i} \left(\gamma_\ell \left(u_j^{\nu+1} - u_i^{\nu+1}\right) + \delta_\ell \left(u_s^{\nu+1} - u_r^{\nu+1}\right)\right) + |P_i|\lambda_i u_i^{\nu+1} = |P_i| f_i + \rho_i^\nu, \\ -\displaystyle\sum_{\ell\in r} \left(\Gamma_\ell \left(u_j^{\nu+1} - u_i^{\nu+1}\right) + \Delta_\ell \left(u_s^{\nu+1} - u_r^{\nu+1}\right)\right) + |D_r|\lambda_r u_r^{\nu+1} = |D_r| f_r + \rho_r^\nu, \end{cases} \tag{26}$$

with

$$|\rho_i^\nu| \le C\epsilon, \quad |\rho_r^\nu| \le C\epsilon, \tag{27}$$

where C is a constant independant of h and ϵ.

Proof. Recall that, for all i, r, ν, $u_i^\nu > 0$ and $u_r^\nu > 0$. Suppose, for example, that

$$r_\ell(\mathbf{u}^\nu) = \delta_\ell(u_s^\nu - u_r^\nu) \ge 0, \quad R_\ell(\mathbf{u}^\nu) = \Gamma_\ell(u_j^\nu - u_i^\nu) \ge 0,$$

then $r_\ell^-(\mathbf{u}^\nu) = R_\ell^-(\mathbf{u}^\nu) = 0$ and the scheme (24) rewrites

$$\begin{cases} -\displaystyle\sum_{\ell\in i} \left(\gamma_\ell \left(u_j^{\nu+1} - u_i^{\nu+1}\right) + \delta_\ell(u_s^\nu - u_r^\nu)\frac{u_j^{\nu+1}}{u_j^\nu}\right) + |P_i|\lambda_i u_i^{\nu+1} = |P_i| f_i, \\ -\displaystyle\sum_{\ell\in r} \left(\Gamma_\ell(u_j^\nu - u_i^\nu)\frac{u_s^{\nu+1}}{u_s^\nu} + \Delta_\ell \left(u_s^{\nu+1} - u_r^{\nu+1}\right)\right) + |D_r|\lambda_r u_r^{\nu+1} = |D_r| f_r. \end{cases} \tag{28}$$

From assumption (25) we deduce that, for all i, r, there exists ϵ_i ($|\epsilon_i| \le \epsilon$) and ϵ_r ($|\epsilon_r| \le \epsilon$) such that

$$u_i^{\nu+1} = u_i^\nu + \epsilon_i u_i^\nu, \quad u_r^{\nu+1} = u_r^\nu + \epsilon_r u_r^\nu.$$

Inserting these values into (28) gives (26) with

$$\rho_i^\nu = \sum_{\ell\in i} \delta_\ell \left(\epsilon_r u_r^\nu - \epsilon_s u_s^\nu - \epsilon_j u_r^\nu + \epsilon_j u_s^\nu\right), \quad \rho_r^\nu = \sum_{\ell\in r} \Gamma_\ell \left(\epsilon_i u_i^\nu - \epsilon_j u_j^\nu - \epsilon_s u_i^\nu + \epsilon_s u_j^\nu\right).$$

As a consequence,

$$|\rho_i^\nu| \leq 4N_{max} \left(\max_\ell |\delta_\ell| \right) \left(\max_r u_r^\nu \right) \epsilon, \quad |\rho_r^\nu| \leq 4N_{max} \left(\max_\ell |\Gamma_\ell| \right) \left(\max_i u_i^\nu \right) \epsilon,$$

where we recall that N_{max} is the maximum number of faces of primal and dual cells. This concludes the proof. □

Theorem 6.12. *Assume that* **H1**, **H2**, **H3** *hold, and that the assumptions of Lemma 6.11 are satisfied. Then, there exists a constant C_4, independent of h and ϵ, such that*

$$\|\bar{\mathbf{u}} - \mathbf{u}^{\nu+1}\|_2 \leq C_1 h + C_4 \epsilon,$$

with C_1 the constant defined by Theorem 6.10.

Proof. System (15) writes

$$\mathbf{A}\mathbf{u} = \mathbf{f}$$

with

$$\mathbf{f} = \left((|P_i| f_i)_{1 \leq i \leq n}, (|D_r| f_r)_{1 \leq r \leq m} \right),$$

while system (26) writes

$$\mathbf{A}\mathbf{u}^{\nu+1} = \mathbf{f} + \mathbf{f}_\epsilon$$

with

$$\mathbf{f}_\epsilon = \left((\rho_i^\nu)_{1 \leq i \leq n}, (\rho_r^\nu)_{1 \leq r \leq m} \right).$$

By difference and thanks to the stability Lemma A.5, there exists a constant C_2 such that

$$\|\mathbf{u} - \mathbf{u}^{\nu+1}\|_2 \leq C_2 \|\mathbf{f}_\epsilon\|_2.$$

Thanks to Lemma 6.11 there exists a constant C_3 such that

$$\|\mathbf{f}_\epsilon\|_2 \leq C_3 \epsilon.$$

Then choosing $C_4 = C_2 C_3$ and applying the triangle inequality and Theorem 6.10 we obtain

$$\|\bar{\mathbf{u}} - \mathbf{u}^{\nu+1}\|_2 \leq \|\bar{\mathbf{u}} - \mathbf{u}\|_2 + \|\mathbf{u} - \mathbf{u}^{\nu+1}\|_2 \leq C_1 h + C_4 \epsilon,$$

which concludes the proof. □

Note that Theorem 6.12 is *not* a convergence theorem. Indeed if we make both h and ϵ tend to zero, the positive solution $\mathbf{u}^{\nu+1}$ tends to the DDFV numerical solution $\mathbf{u}$ which is only possible if $\mathbf{u}$ *itself* is non negative. Roughly speaking one can say that the (positive) numerical solution $\mathbf{u}^{\nu+1}$ obtained at the end of the iterative process is *close* to the (non necessarily positive) DDFV numerical solution $\mathbf{u}$ that itself is close to the exact solution $\bar{\mathbf{u}}$. Note also that condition (25) is constraining: in practice we rather use the condition $\|\mathbf{u}^{\nu+1} - \mathbf{u}^\nu\|_\infty \leq \epsilon \|\mathbf{u}^\nu\|_\infty$ or $\|\mathbf{u}^{\nu+1} - \mathbf{u}^\nu\|_2 \leq \epsilon \|\mathbf{u}^\nu\|_2$ as a stopping criterion.

7 Numerical experiments

Given $\Omega =]0,1[^2$, κ a diffusion coefficient and g a function defined on $\partial\Omega$, consider Problem (1) with $\lambda = 0$ and $\Gamma_N = \emptyset$

$$\begin{cases} -\boldsymbol{\nabla} \cdot (\kappa \boldsymbol{\nabla} \bar{u}) = f & \text{in} \quad \Omega, \\ \bar{u} = g & \text{on} \quad \partial\Omega. \end{cases} \tag{29}$$

In addition to Cartesian meshes we will use the two following types of meshes (see Fig. 5):

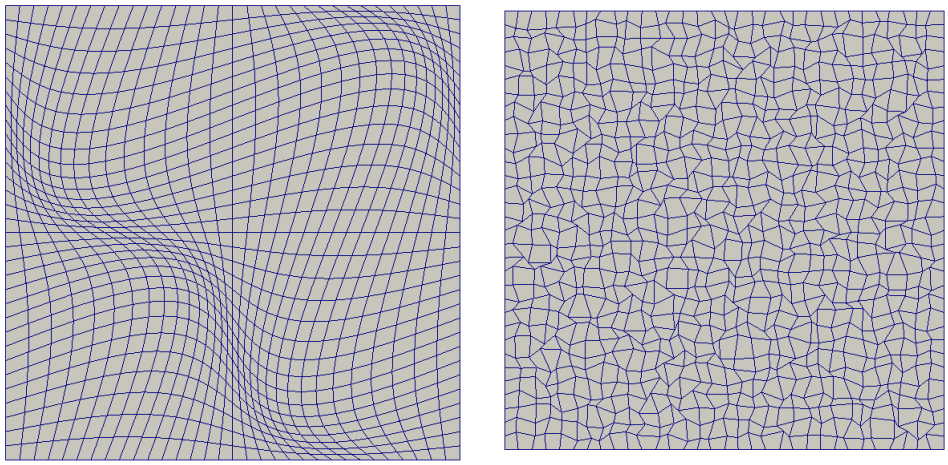
1. deformed meshes, the deformation of which from the Cartesian mesh is given by

$$(x, y) \to (x + 0.1 \sin(2\pi x) \sin(2\pi y), y + 0.1 \sin(2\pi x) \sin(2\pi y)),$$

2. randomly deformed meshes, the deformation of which from the unit Cartesian mesh with cells of size Δx is given by

$$(x, y) \to 0.1(x, y) + 0.9(x + 0.45a\Delta x, y + 0.45b\Delta x),$$

where a, b are random numbers distributed according to the uniform law on $[-1, 1]$.



(a) A deformed mesh (b) A random mesh

Figure 5: Examples of deformed meshes.

The L^2 and H^1-errors used in the following tests are respectively given by

$$\frac{\|\mathbf{u} - \bar{\mathbf{u}}\|_2}{\|\bar{\mathbf{u}}\|_2} \quad \text{and} \quad \frac{\|\boldsymbol{\nabla}_h \mathbf{u} - \boldsymbol{\nabla} \bar{u}\|_2}{\|\boldsymbol{\nabla} \bar{u}\|_2},$$

where

$$\|\boldsymbol{\nabla} \bar{u}\|_2 = \left(\sum_\ell |I_\ell| \, \|\boldsymbol{\nabla} \bar{u}(\mathbf{x}_\ell)\|^2 \right)^{1/2},$$

$$\|\boldsymbol{\nabla}_h \mathbf{u} - \boldsymbol{\nabla} \bar{u}\|_2 = \left(\sum_\ell |I_\ell| \left\| \frac{1}{2} \frac{1}{|I_\ell|} \left((u_j - u_i) \mathbf{N}_{sr} + (u_s - u_r) \mathbf{N}_{ij} \right) - \boldsymbol{\nabla} \bar{u}(\mathbf{x}_\ell) \right\|^2 \right)^{1/2}.$$

For DDFV type schemes we plot on figures 8, 9, 11, 12, the *primal* numerical values while on tables 2, 3, 4, the maxima and minima are computed over *both* primal *and* dual values.

7.1 Accuracy

Three simple benchmarks are proposed to assess the accuracy of our monotonic schemes in comparison with the usual (non monotonic) DDFV scheme. For these three tests, we choose $\epsilon = 10^{-12}$ as the stopping criterion of the fixed point algorithm.

7.1.1 Checking the preservation of linear solutions

Given $\kappa(\mathbf{x}) = 1$ $f(\mathbf{x}) = 0$ and $g(\mathbf{x}) = -x - y + 2$, the positive linear function $\bar{u}(\mathbf{x}) = -x - y + 2$ is solution to (29). We perform a study of this problem on the deformed mesh (see Fig. 5a) with 32×32 cells for each of the three schemes. The L^2-error between the exact solution $\bar{\mathbf{u}}$ and the approximated one $\mathbf{u}$ are reported in Table 1. The error is zero, to machine precision, when $\bar{u}$ is a polynomial of degree 1.

Scheme	L^2-error	H^1-error
DDFV	$2.58e-15$	$4.46e-14$
Monotonic DDFV	$9.42e-15$	$6.30e-13$
Monotonic diamond (degree 1)	$1.05e-14$	$1.02e-13$

Table 1: Comparison between the different schemes for the positive linear solution to problem of Section 7.1.1.

7.1.2 Anisotropic diffusion coefficient

Given

$$\boldsymbol{\kappa}(\mathbf{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad f(\mathbf{x}) = 3\pi^2 \sin(\pi x)\sin(\pi y), \quad g(\mathbf{x}) = 0,$$

the function $\bar{u}(\mathbf{x}) = \sin(\pi x)\sin(\pi y)$ is solution to (29). We perform a convergence study for this problem with a sequence of successively refined deformed meshes like the one of Fig. 5a. Results are summarized in Fig. 6 which shows that all schemes are second-order accurate in the L^2 norm. Of course, similar results may be obtained for a scalar-valued diffusion coefficient κ. We see that the error in H^1-norm is second-order convergent for DDFV methods (which is a nice feature already observed [20, 23]). The diamond scheme is only first-order accurate in the H^1 norm. However, we show that we are able to achieve the second-order accuracy for the H^1 norm for this scheme. To do that, we reconstruct the gradient with polynomials of degree two instead of one.

7.1.3 Discontinuous diffusion coefficient

Recall that we have assumed that possible discontinuities of the diffusion coefficient κ occur only along the primal cell faces. Given

$$\kappa(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad x \leq \dfrac{1}{2} \\ 2 & \text{if} \quad x > \dfrac{1}{2} \end{cases}, \quad f(\mathbf{x}) = 2\pi^2 \cos(\pi x)\cos(\pi y) + 20, \quad g(\mathbf{x}) = 0,$$

the function

$$\bar{u}(\mathbf{x}) = \begin{cases} \cos(\pi x)\cos(\pi y) - 10x^2 + 12 & \text{if} \quad x \leq \dfrac{1}{2}, \\ \dfrac{1}{2}\cos(\pi x)\cos(\pi y) - 5x^2 + \dfrac{43}{4} & \text{if} \quad x > \dfrac{1}{2}, \end{cases}$$

is solution to (29). We perform a convergence study for this problem with a sequence of successively refined deformed meshes as shown in Fig. 5a.
Fig. 7 shows that, in the present case of a discontinuous κ, the results are similar to those of the continuous case, that is to say, the scheme is second-order accurate. However, both schemes are only first-order accurate in H^1 norm in this case.

7.2 Monotonicity test problems

We propose two benchmarks to compare the usual DDFV scheme, which can give nonpositive solutions, with our monotonic diamond and DDFV schemes which always give nonnegative solutions.

7.2.1 Tensor-valued coefficient κ and square domain with a square hole

Consider the square domain with a square hole $\Omega =]0,1[^2 \setminus \left[\frac{4}{9}, \frac{5}{9}\right]^2$, $f(\mathbf{x}) = 0$ in Ω and $g(\mathbf{x}) = 0$ (resp. $g(\mathbf{x}) = 2$) on the external (resp. internal) boundary. We have chosen

$$\boldsymbol{\kappa} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 10^4 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \theta = \frac{\pi}{6}.$$

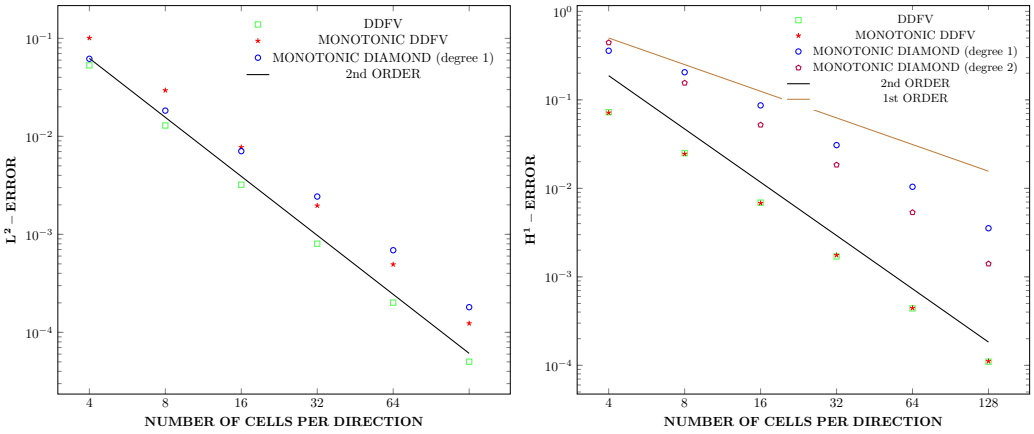


Figure 6: L^2 (on the left) and H^1 (on the right) errors for problem of Section 7.1.2.

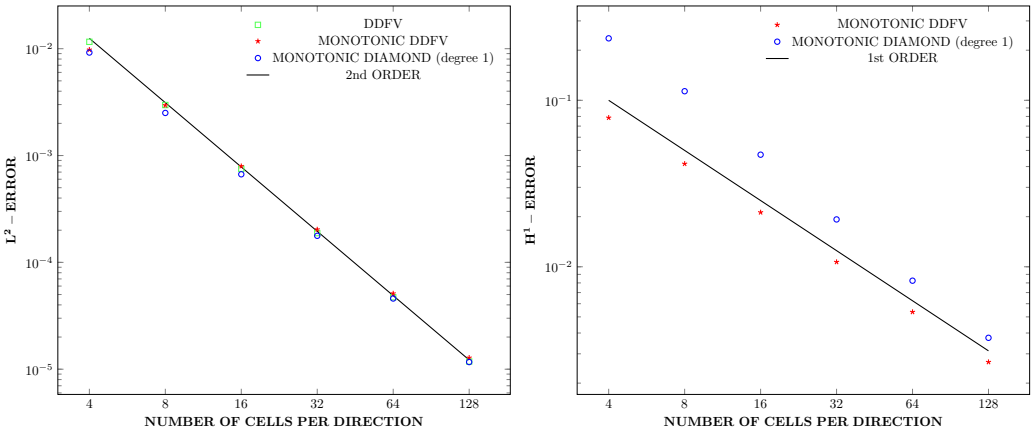


Figure 7: L^2 (on the left) and H^1 (on the right) errors for problem of Section 7.1.3.

We compare the results obtained with the monotonic diamond and DDFV schemes on a Cartesian mesh with 36 cells per direction. We use a quite low number of degrees of freedom for this test to exhibit the non-monotonicity of the DDFV scheme (which tends to cancel in refining the mesh, see also section 7.2.2). The stopping criterion of the fixed point algorithm is $\epsilon = 10^{-12}$. Figure 8 shows the mesh, the DDFV solution and its negative and positive parts. Fig. 9 displays the monotonic DDFV and diamond solutions while Table 2 gives the minimum and the maximum of each solution.
While the solution obtained with the usual DDFV scheme has a negative minimum we can see that the solutions obtained with the monotonic methods are always positive, as expected.

Scheme	Minimum of the solution	Maximum of the solution
DDFV	-4.59×10^{-1}	2.05
Monotonic DDFV	1.65×10^{-17}	2.01
Monotonic diamond (degree 1)	2.46×10^{-32}	1.95

Table 2: Minimum and maximum of the numerical solution to the problem of section 7.2.1 for the Cartesian mesh with 36 cells by direction.

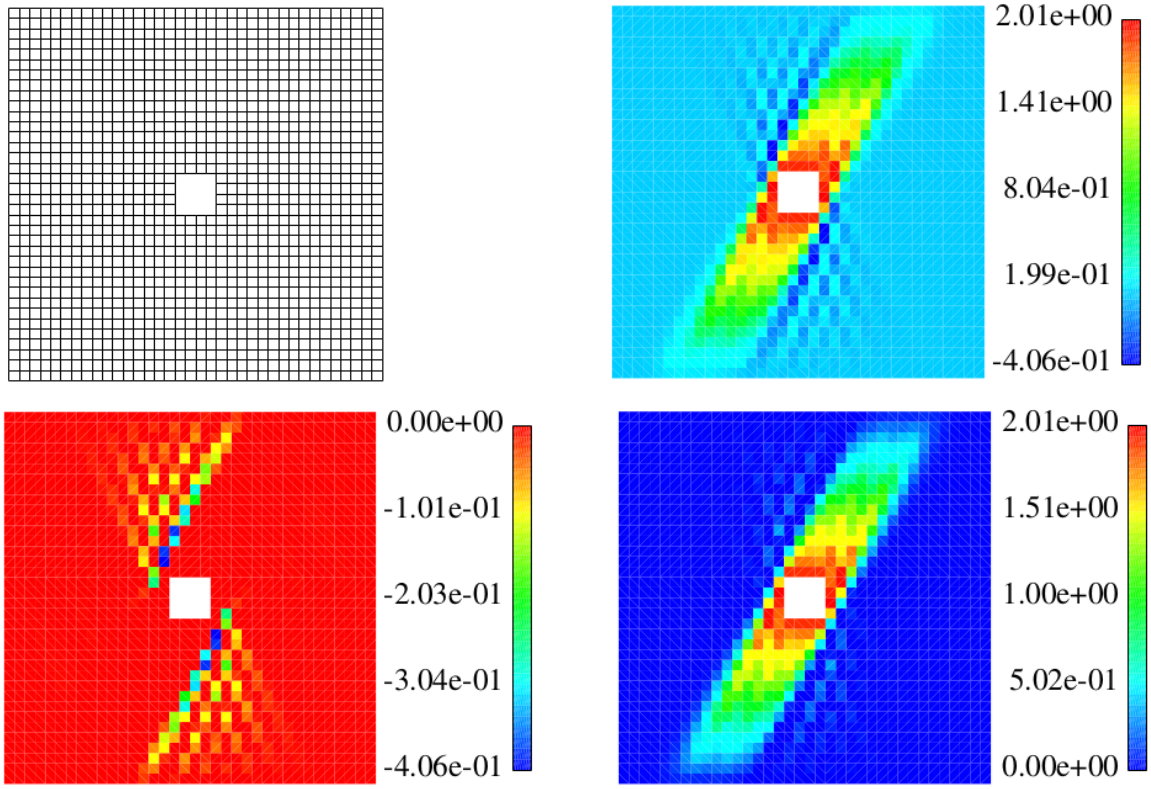


Figure 8: Mesh (top, left), DDFV solution to problem of section 7.2.1 (top, right) and its negative (bottom, left) and positive (bottom, right) parts.

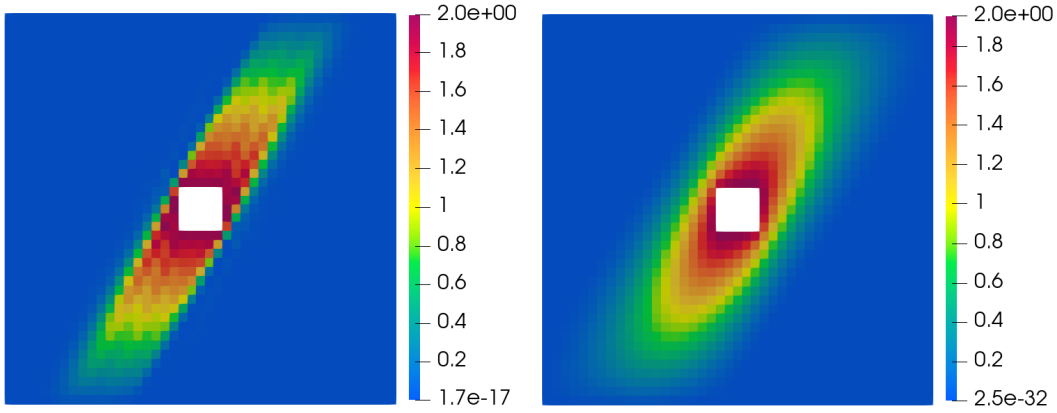


Figure 9: Monotonic DDFV (on the left) and diamond (degree 1, on the right) solutions to problem of section 7.2.1.

7.2.2 Fokker-Planck type diffusion equation

This benchmark is a simplified version of the one from [27]. Given $\Omega =]-50, 50[^2$, $T = 250$, $\mathbf{v} = (v_x, v_y)$ the velocity variable and $\mathbf{V} = (-20, 20)$ the averaged velocity, we are looking for the distribution function

$\bar{u} = \bar{u}(\mathbf{v}, t)$, solution to the simplified Fokker-Planck equation

$$\begin{cases} \dfrac{\partial \bar{u}}{\partial t} - \boldsymbol{\nabla}_{\mathbf{v}} \left(\boldsymbol{\kappa} \boldsymbol{\nabla}_{\mathbf{v}} \bar{u} \right) = 0 & \text{in} \quad \Omega \times [0, T], \\ \boldsymbol{\kappa} \boldsymbol{\nabla}_{\mathbf{v}} \bar{u} \cdot \mathbf{n} = 0 & \text{on} \quad \partial\Omega \times [0, T], \\ \bar{u}(0) = \bar{u}^0 & \text{in} \quad \Omega, \end{cases} \tag{30}$$

where the diffusion coefficient $\boldsymbol{\kappa} = \boldsymbol{\kappa}(\mathbf{v})$ and the initial condition $\bar{u}^0$ are given by

$$\boldsymbol{\kappa}(\mathbf{v}) = \mathbf{I} - \frac{1}{\|\mathbf{v}\|^2} \mathbf{v} \otimes \mathbf{v}, \quad \bar{u}^0(\mathbf{v}) = \frac{1}{\pi} \exp(-\|\mathbf{v} - \mathbf{V}\|^2).$$

Note that the full Fokker-Planck equation would read as

$$\frac{\partial \bar{u}}{\partial t} + \boldsymbol{\nabla}_{\mathbf{v}} \cdot (\mathbf{v} \bar{u}) - \boldsymbol{\nabla}_{\mathbf{v}} \left(\boldsymbol{\kappa} \boldsymbol{\nabla}_{\mathbf{v}} \bar{u} \right) = 0.$$

It is well known that the n-order moments of $\bar{u}$ $(0 \leq n \leq 2)$ are preserved over the time

$$\frac{d}{dt} \left(\int_\Omega \bar{u} \right) = 0, \quad \frac{d}{dt} \left(\int_\Omega \mathbf{v} \bar{u} \right) = \mathbf{0}, \quad \frac{d}{dt} \left(\int_\Omega \|\mathbf{v}\|^2 \bar{u} \right) = 0.$$

The backward Euler scheme is used for time discretization.
To limit the calculation time, the stopping criterion of the fixed point algorithm is $\epsilon = 10^{-5}$. Fig. 11 (resp. 12) displays the DDFV (resp. monotonic DDFV and diamond) numerical solutions obtained with the Cartesian, deformed and random meshes of 200^2 cells. Table 3 gives the minima and maxima of the DDFV solution for a sequence of refined Cartesian meshes and Table 4 gives the minima and the maxima of the numerical solution obtained with the DDFV, monotonic DDFV and diamond schemes. We observe that the minima of the DDFV solution are negative but converge to zero as h tends to zero while the minima of the solutions to monotonic schemes always remain non negative, as expected. Compared to both the non monotonic and monotonic DDFV schemes the monotonic diamond scheme is more diffusive. This could be explained by the use of a larger stencil required for polynomial reconstruction.
The conservation of the zero-order moment of $\bar{u}$ at the discrete level is a property of ours schemes. It is more challenging to obtain a conservation of a discrete equivalent of the second-order moment.
Thanks to the identity

$$\mathbf{v} = \frac{1}{2} \boldsymbol{\nabla}_{\mathbf{v}} (\|\mathbf{v}\|^2),$$

one can introduce an approximation $\mathbf{v}_\ell$ of $\mathbf{v}$ in the diamond cell I_ℓ by using the Green-Gauss formula

$$\begin{cases} \mathbf{v}_\ell = \dfrac{1}{4} \dfrac{1}{|I_\ell|} \left((\|\mathbf{v}_j\|^2 - \|\mathbf{v}_i\|^2) \mathbf{N}_{sr} + (\|\mathbf{v}_s\|^2 - \|\mathbf{v}_r\|^2) \mathbf{N}_{ij} \right) & \ell \notin \partial\Omega, \\ \mathbf{v}_\ell = \dfrac{1}{4} \dfrac{1}{|I_\ell|} \left((\|\mathbf{v}_\ell\|^2 - \|\mathbf{v}_i\|^2) \mathbf{N}_{sr} + (\|\mathbf{v}_s\|^2 - \|\mathbf{v}_r\|^2) \mathbf{N}_{i\ell} \right) & \ell \in \partial\Omega. \end{cases} \tag{31}$$

We then prove the following proposition.

Proposition 7.1. *Consider the DDFV solution to* (30)*, that is,*

$$\left\{\begin{array}{ll}
|P_i|\dfrac{u_i^{n+1}-u_i^n}{\Delta t} - \dfrac{1}{2}\displaystyle\sum_{\ell\in i,\ell\notin\partial\Omega}\dfrac{1}{|I_\ell|}\left((u_j^{n+1}-u_i^{n+1})\mathbf{N}_{sr}\boldsymbol{\kappa}_\ell\mathbf{N}_{sr}+(u_s^{n+1}-u_r^{n+1})\mathbf{N}_{ij}\boldsymbol{\kappa}_\ell\mathbf{N}_{sr}\right) & \\
\qquad -\dfrac{1}{2}\displaystyle\sum_{\ell\in i,\ell\in\partial\Omega}\dfrac{1}{|I_\ell|}\left((u_\ell^{n+1}-u_i^{n+1})\mathbf{N}_{sr}\boldsymbol{\kappa}_\ell\mathbf{N}_{sr}+(u_s^{n+1}-u_r^{n+1})\mathbf{N}_{i\ell}\boldsymbol{\kappa}_\ell\mathbf{N}_{sr}\right)=0, & \\
|D_r|\dfrac{u_r^{n+1}-u_r^n}{\Delta t} - \dfrac{1}{2}\displaystyle\sum_{\ell\in r,\ell\notin\partial\Omega}\dfrac{1}{|I_\ell|}\left((u_j^{n+1}-u_i^{n+1})\mathbf{N}_{sr}\boldsymbol{\kappa}_\ell\mathbf{N}_{ij}+(u_s^{n+1}-u_r^{n+1})\mathbf{N}_{ij}\boldsymbol{\kappa}_\ell\mathbf{N}_{ij}\right) & \\
\qquad -\dfrac{1}{2}\displaystyle\sum_{\ell\in r,\ell\in\partial\Omega}\dfrac{1}{|I_\ell|}\left((u_\ell^{n+1}-u_i^{n+1})\mathbf{N}_{sr}\boldsymbol{\kappa}_\ell\mathbf{N}_{i\ell}+(u_s^{n+1}-u_r^{n+1})\mathbf{N}_{i\ell}\boldsymbol{\kappa}_\ell\mathbf{N}_{i\ell}\right)=0, & \\
\dfrac{1}{2}\dfrac{1}{|I_\ell|}\left((u_\ell^{n+1}-u_i^{n+1})\mathbf{N}_{sr}\boldsymbol{\kappa}_\ell\mathbf{N}_{ij}+(u_s^{n+1}-u_r^{n+1})\mathbf{N}_{ij}\boldsymbol{\kappa}_\ell\mathbf{N}_{ij}\right)=0 & \mathbf{x}_\ell\in\partial\Omega,
\end{array}\right. \tag{32}$$

with the following approximations of $\boldsymbol{\kappa}$ *in a diamond cell* I_ℓ *such that* $\mathbf{v}_\ell\notin\partial\Omega$

$$\boldsymbol{\kappa}_\ell=\mathbf{I}-\frac{1}{\|\mathbf{v}_\ell\|^2}\mathbf{v}_\ell\otimes\mathbf{v}_\ell,$$

with $\mathbf{v}_\ell$ *calculated by* (31)*.*

Let E^n *be the following discrete equivalent of the second-order moment*

$$E^n=\frac{1}{2}\left(\sum_i|P_i|\|\mathbf{v}_i\|^2u_i^n+\sum_r|D_r|\|\mathbf{v}_r\|^2u_r^n\right).$$

Then, for all $n\geq 0$*,* $E^n=E^0$*.*

Proof. We multiply the first (resp. second) equation of (32) by $\|\mathbf{v}_i\|^2$ (resp. $\|\mathbf{v}_r\|^2$) and sum over primal (resp. dual) cells P_i (resp. D_r). Adding these two sums we get

$$\frac{1}{\Delta t}\left(\sum_i|P_i|\|\mathbf{v}_i\|^2u_i^{n+1}+\sum_r|D_r|\|\mathbf{v}_r\|^2u_r^{n+1}-\sum_i|P_i|\|\mathbf{v}_i\|^2u_i^n-\sum_r|D_r|\|\mathbf{v}_r\|^2u_r^n\right)$$
$$-\frac{1}{2}\sum_\ell\frac{1}{|I_\ell|}\left((\|\mathbf{v}_i\|^2-\|\mathbf{v}_j\|^2)\mathbf{N}_{sr}+(\|\mathbf{v}_r\|^2-\|\mathbf{v}_s\|^2)\mathbf{N}_{ij}\right)\boldsymbol{\kappa}_\ell\left((u_j^{n+1}-u_i^{n+1})\mathbf{N}_{sr}+(u_s^{n+1}-u_r^{n+1})\mathbf{N}_{ij}\right)=0.$$

Then, noting that $\boldsymbol{\kappa}_\ell\mathbf{v}_\ell=\mathbf{0}$, we obtain thanks to (31)

$$\sum_i|P_i|\|\mathbf{v}_i\|^2u_i^{n+1}+\sum_r|D_r|\|\mathbf{v}_r\|^2u_r^{n+1}=\sum_i|P_i|\|\mathbf{v}_i\|^2u_i^n+\sum_r|D_r|\|\mathbf{v}_r\|^2u_r^n,$$

that is, $E^{n+1}=E^n$. □

The numerical results displayed in Fig. 10 show that the second order moment is conserved over time for the non-monotonic DDFV scheme, as it has been proved. However, it is not exactly conserved with monotonic DDFV scheme because we do not exactly solve the DDFV system. However, the conservation error is far lower than for the positive diamond scheme.

	Number of cells	Cartesian mesh	Deformed mesh	Random mesh
Minima	100×100	-1.89×10^{-3}	-2.38×10^{-3}	-3.15×10^{-3}
	200×200	-2.48×10^{-4}	-1.25×10^{-3}	-2.41×10^{-3}
	400×400	-6.32×10^{-13}	-2.14×10^{-4}	-9.92×10^{-4}
	800×800	-1.66×10^{-13}	-7.95×10^{-7}	-7.63×10^{-4}
	1600×1600	-8.53×10^{-14}	-1.97×10^{-7}	-4.58×10^{-4}
Maxima	100×100	1.19×10^{-2}	1.16×10^{-2}	1.65×10^{-2}
	200×200	1.04×10^{-2}	1.04×10^{-2}	1.14×10^{-2}
	400×400	1.01×10^{-2}	1.01×10^{-2}	1.09×10^{-2}
	800×800	1.01×10^{-2}	1.01×10^{-2}	1.09×10^{-2}
	1600×1600	1.01×10^{-2}	1.01×10^{-2}	1.08×10^{-2}

Table 3: Minima and maxima of the DDFV solution of (30) at time $T = 250$ on refined Cartesian meshes.

	Scheme	Cartesian mesh	Deformed mesh	Random mesh
Minima	DDFV	-2.48×10^{-4}	-1.25×10^{-3}	-2.41×10^{-3}
	Monotonic DDFV	5.46×10^{-30}	2.53×10^{-30}	4.63×10^{-40}
	Monotonic diamond (degree 1)	1.86×10^{-29}	1.42×10^{-22}	1.58×10^{-23}
Maxima	DDFV	1.04×10^{-2}	1.04×10^{-2}	1.14×10^{-2}
	Monotonic DDFV	1.04×10^{-2}	0.97×10^{-2}	1.02×10^{-2}
	Monotonic diamond (degree 1)	0.29×10^{-2}	0.32×10^{-2}	0.31×10^{-2}

Table 4: Minima and maxima of the numerical solutions to (30) at time $T = 250$ on the three types of 200×200 cells meshes.

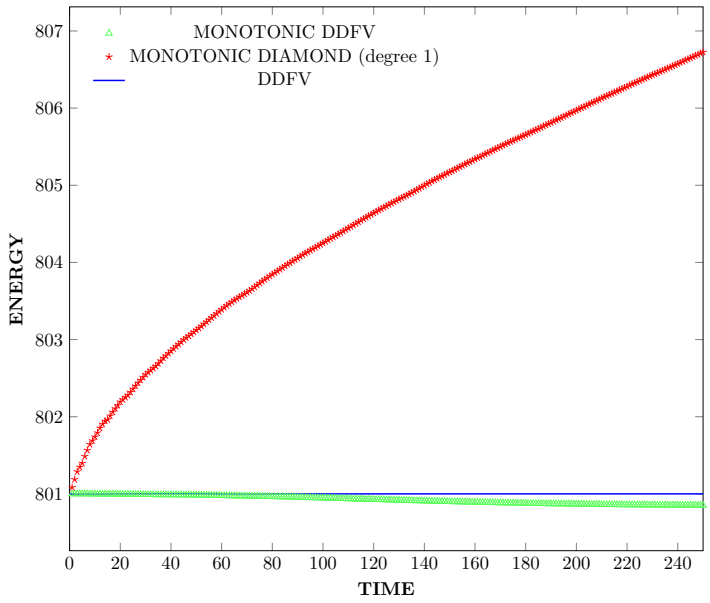


Figure 10: Variation of energy over time for the 3 schemes on cartesian mesh of 200×200 cells.

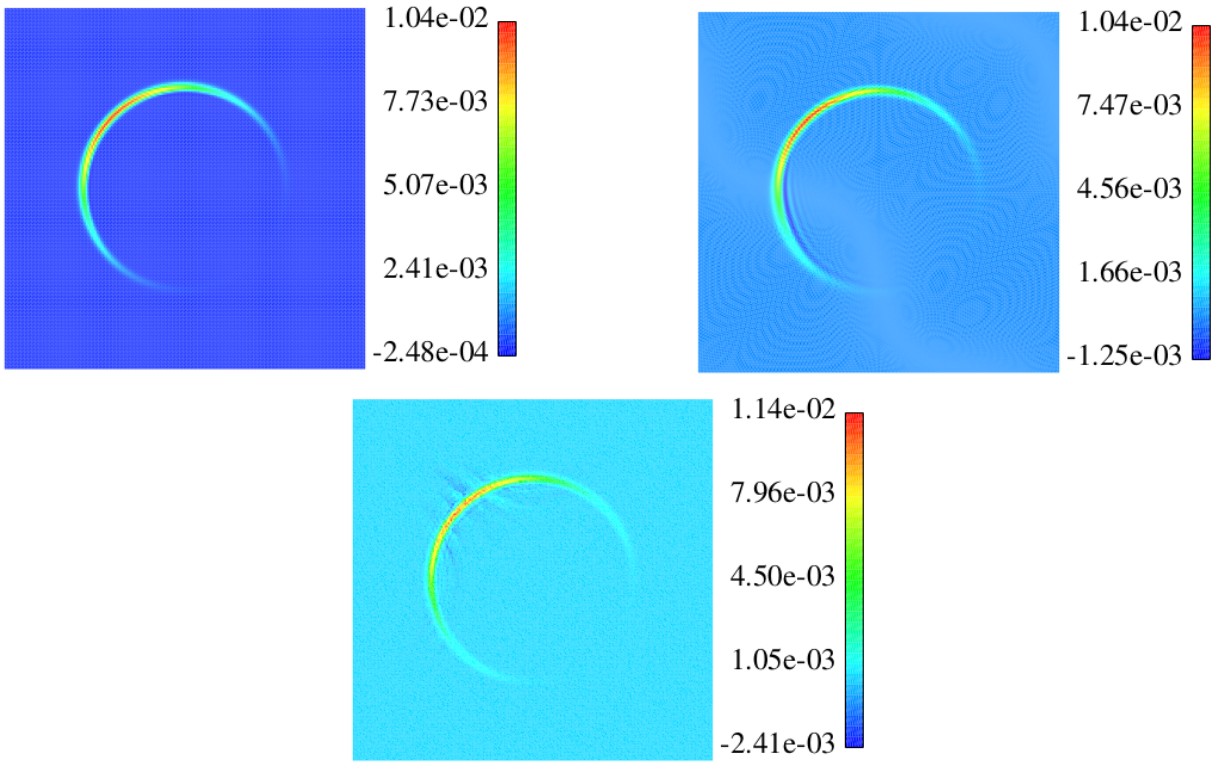


Figure 11: DDFV solution to (30) at time $T = 250$ on the Cartesian (top left), deformed (top right) and random (bottom) mesh of 200×200 cells.

8 Concluding remarks

In this paper, we propose two new monotonic schemes for the diffusion equation, which are based on the same cell-centered discretization. This first step is called primal scheme, and the consistency of the primal fluxes relies on a correct evaluation of dual (node-centered) unknowns. The difference between the two schemes lies

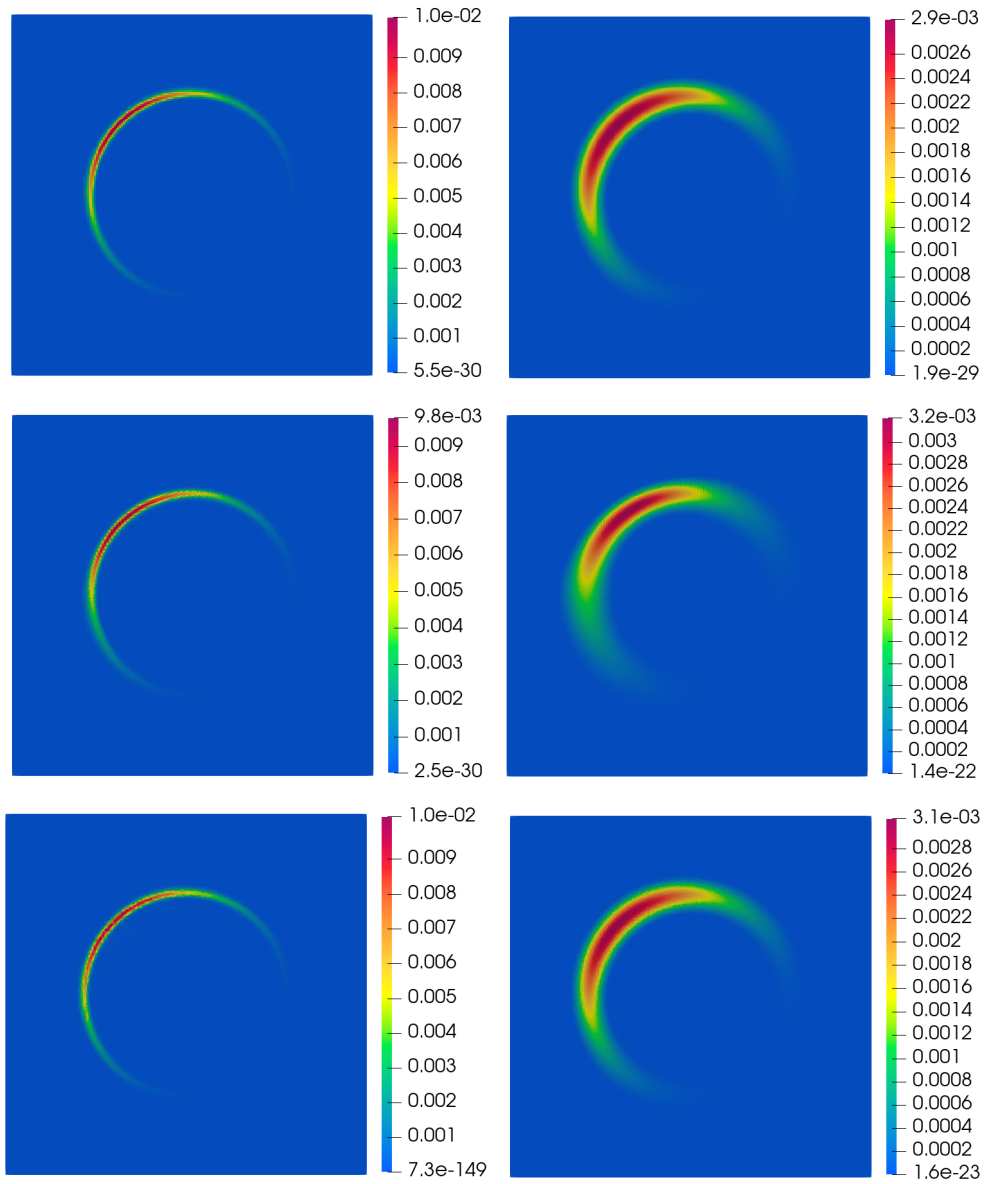


Figure 12: Monotonic DDFV (on the left) and diamond (degree 1, on the right) solutions to (30) at time $T = 250$ on the Cartesian (top), deformed (middle) and random (bottom) mesh of 200×200 cells.

in the evaluation of these dual quantities. For the first one, which is called diamond type, the dual unknowns are evaluated, using a polynomial reconstruction involving values in neighbouring (primal) cells. For the second one, called DDFV type, the evaluation of the dual unknown is obtained by solving a diffusion problem discretized on the dual mesh. This second scheme is an improvement with respect to the nonlinear monotonic DDFV method of [6]. Indeed, the new nonlinear method we have proposed here makes it possible to deal with all types of boundary conditions (Dirichlet, Neumann) and is second-order convergent even for discontinuous diffusion coefficients. For both methods, we adapt the same non-linear process borrowed from [44, 19, 49, 18], we assess their monotonicity and accuracy on several test cases and compare the results with the classical (non-monotonic) DDFV scheme. Moreover, the DDFV type monotonic scheme takes advantage of very nice features of the DDFV scheme, such as second-order accuracy in H^1 norm, while providing non-negative solutions.
In the future, we plan to extend theses schemes to arbitrary order, using the techniques developed in the 1D setting in [4].

A Proof of convergence for the DDFV scheme

For simplicity we will restrict ourselves to the case $\kappa = 1$, $\lambda = 0$, $g = 0$ and $\Gamma_N = \emptyset$ in (1), that is,

$$\begin{cases} -\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \bar{u}) = f & \text{in} \quad \Omega, \\ \bar{u} = g & \text{on} \quad \partial\Omega. \end{cases} \tag{33}$$

To simplify the proof, we suppose further that the dual mesh is made of cells obtained by joining the center of each primal cell with the center of each of its neightbors and with the middle of its boundary faces (but it extends to the barycentric dual mesh used in this paper). In this case we observe that the dual boundary $G_\ell = D_r \cap D_s$ coincides with the segment $\mathbf{x}_i\mathbf{x}_j$. Denote by $\mathbf{n}_{ij}$ the unit vector orthogonal to G_ℓ directed from D_r to D_s, $\mathbf{N}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|\mathbf{n}_{ij}$, and by θ_ℓ the angle between vectors $\mathbf{n}_{ji}^\perp$ and $\mathbf{n}_{sr}$ (see Fig. 13).

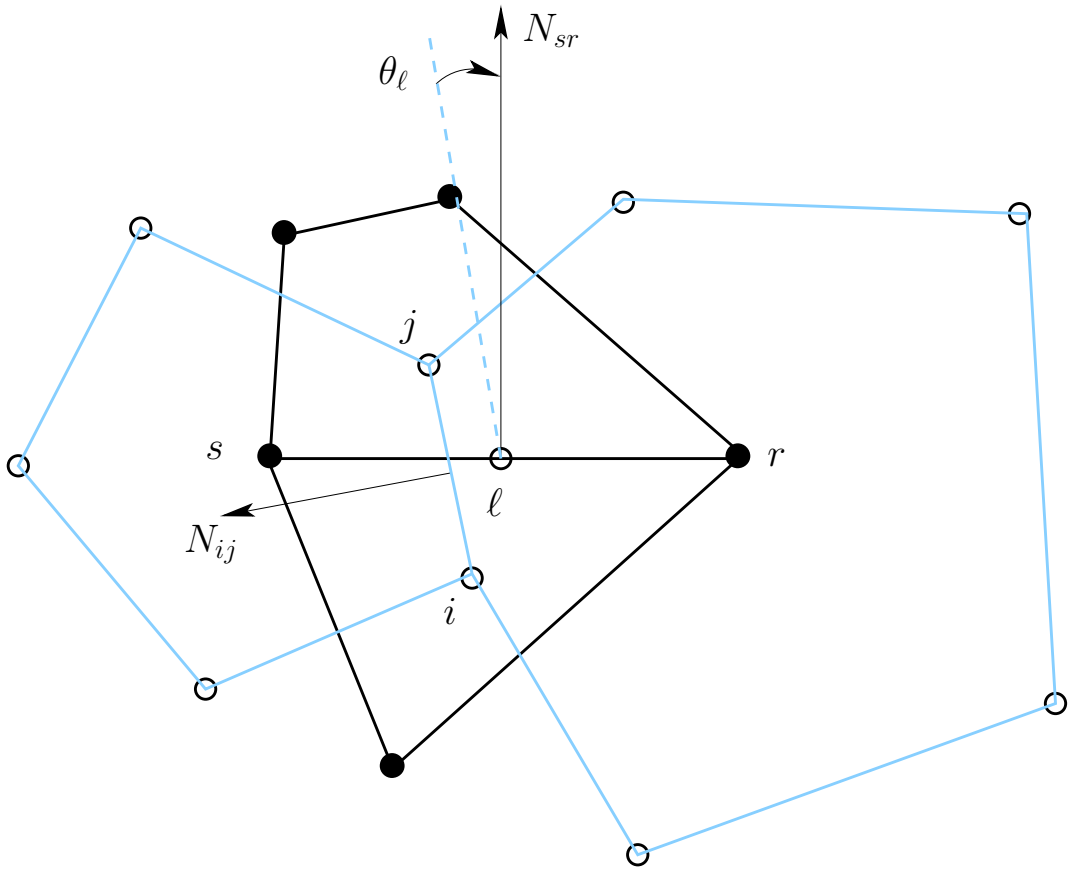


Figure 13: Two primal cells P_i, P_j (black lines) such that $P_i \cap P_j = F_\ell = \mathbf{x}_r\mathbf{x}_s$ and two dual cells D_r, D_s (blue lines) such that $D_r \cap D_s = G_\ell = \mathbf{x}_i\mathbf{x}_j$.

We define

$$h = \max_\ell(|F_\ell|, |G_\ell|).$$

Applying the method used in Sections 3 and 4.2, we have

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell) \cdot \mathbf{n}_{sr} = \frac{1}{\cos(\theta_\ell)} \frac{\bar{u}(\mathbf{x}_j) - \bar{u}(\mathbf{x}_i)}{|G_\ell|} + \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)} \frac{\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r)}{|F_\ell|} + \mathcal{O}(h),$$

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell) \cdot \mathbf{n}_{ij} = \frac{1}{\cos(\theta_\ell)} \frac{\bar{u}(\mathbf{x}_s) - \bar{u}(\mathbf{x}_r)}{|F_\ell|} + \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)} \frac{\bar{u}(\mathbf{x}_j) - \bar{u}(\mathbf{x}_i)}{|G_\ell|} + \mathcal{O}(h).$$

This is equivalent to say that $\boldsymbol{\nabla}\bar{u}$ is approximated in the diamond cell I_ℓ using the Green-Gauss formula

$$\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell) = \frac{1}{|I_\ell|}\int_{I_\ell}\boldsymbol{\nabla}\bar{u} + \mathcal{O}(h) = \frac{1}{2}\frac{1}{|I_\ell|}\left(\mathbf{N}_{sr}(u_j-u_i)+\mathbf{N}_{ij}(u_s-u_r)\right)+\mathcal{O}(h).$$

The discretization of (33) with the DDFV scheme then writes

$$\begin{cases} -\dfrac{1}{2}\displaystyle\sum_{\ell\in i,\ell\notin\partial\Omega}\frac{1}{|I_\ell|}\left(\mathbf{N}_{sr}(u_j-u_i)+\mathbf{N}_{ij}(u_s-u_r)\right)\cdot\mathbf{N}_{sr} \\ -\dfrac{1}{2}\displaystyle\sum_{\ell\in i,\ell\in\partial\Omega}\frac{1}{|I_\ell|}\left(\mathbf{N}_{sr}(u_\ell-u_i)+\mathbf{N}_{i\ell}(u_s-u_r)\right)\cdot\mathbf{N}_{sr} = |P_i|f_i, \\ -\dfrac{1}{2}\displaystyle\sum_{\ell\in r,\ell\notin\partial\Omega}\frac{1}{|I_\ell|}\left(\mathbf{N}_{sr}(u_j-u_i)+\mathbf{N}_{ij}(u_s-u_r)\right)\cdot\mathbf{N}_{ij} \\ -\dfrac{1}{2}\displaystyle\sum_{\ell\in r,\ell\in\partial\Omega}\frac{1}{|I_\ell|}\left(\mathbf{N}_{sr}(u_\ell-u_i)+\mathbf{N}_{i\ell}(u_s-u_r)\right)\cdot\mathbf{N}_{ij} = |D_r|f_r+|D_r\cap\partial\Omega|g(\mathbf{x}_r), \\ u_\ell = g(\mathbf{x}_\ell) \qquad \mathbf{x}_\ell\in\partial\Omega, \\ u_r = g(\mathbf{x}_r) \qquad \mathbf{x}_r\in\partial\Omega. \end{cases} \tag{34}$$

The following proofs are inspired from the arguments of [16] for *admissible* meshes and from [2] for general meshes (see also [12], [47]). In the sequel we will assume that the exact solution $\bar{u}$ satisfies $\bar{u}\in W^{2,\infty}(\Omega)$.

A.1 Consistency of the fluxes

Let us denote by

1. $\bar{\mathcal{F}}_\ell$, $\bar{\mathcal{G}}_\ell$ the *exact* primal and dual fluxes

$$\bar{\mathcal{F}}_\ell = \int_{F_\ell}\boldsymbol{\nabla}\bar{u}\cdot\mathbf{n}_{sr}, \quad \bar{\mathcal{G}}_\ell = \int_{G_\ell}\boldsymbol{\nabla}\bar{u}\cdot\mathbf{n}_{ij},$$

2. $\mathcal{F}_\ell(\mathbf{u})$, $\mathcal{G}_\ell(\mathbf{u})$ the *approximated* primal and dual fluxes

$$\mathcal{F}_\ell(\mathbf{u}) = \frac{1}{2}\frac{1}{|I_\ell|}\left((u_j-u_i)\mathbf{N}_{sr}+(u_s-u_r)\mathbf{N}_{ij}\right)\cdot\mathbf{N}_{sr},$$
$$\mathcal{G}_\ell(\mathbf{u}) = \frac{1}{2}\frac{1}{|I_\ell|}\left((u_j-u_i)\mathbf{N}_{sr}+(u_s-u_r)\mathbf{N}_{ij}\right)\cdot\mathbf{N}_{ij},$$

3. $\mathcal{F}_\ell(\bar{\mathbf{u}})$, $\mathcal{G}_\ell(\bar{\mathbf{u}})$ what we can call the *semi-approximated* primal and dual fluxes

$$\mathcal{F}_\ell(\bar{\mathbf{u}}) = \frac{1}{2}\frac{1}{|I_\ell|}\left((\bar{u}(\mathbf{x}_j)-\bar{u}(\mathbf{x}_i))\mathbf{N}_{sr}+(\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r))\mathbf{N}_{ij}\right)\cdot\mathbf{N}_{sr},$$
$$\mathcal{G}_\ell(\bar{\mathbf{u}}) = \frac{1}{2}\frac{1}{|I_\ell|}\left((\bar{u}(\mathbf{x}_j)-\bar{u}(\mathbf{x}_i))\mathbf{N}_{sr}+(\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r))\mathbf{N}_{ij}\right)\cdot\mathbf{N}_{ij}.$$

Proposition A.1 (Consistency of the fluxes for the DDFV scheme)**.** *Assume that* **H1** *is satisfied. Then we have*

$$\begin{cases} \left|\bar{\mathcal{F}}_\ell-\mathcal{F}_\ell(\bar{\mathbf{u}})\right| \le \dfrac{C_\ell}{\cos(\theta_\ell)}|F_\ell|\left((1+|\sin(\theta_\ell)|)|F_\ell|+|G_\ell|\right) \le \dfrac{2C}{\cos(\theta_0)}h^2, \\ \left|\bar{\mathcal{G}}_\ell-\mathcal{G}_\ell(\bar{\mathbf{u}})\right| \le \dfrac{C_\ell}{\cos(\theta_\ell)}|G_\ell|\left((1+|\sin(\theta_\ell)|)|G_\ell|+|F_\ell|\right) \le \dfrac{2C}{\cos(\theta_0)}h^2, \end{cases} \tag{35}$$

where $C_\ell \le C_0\|D^2\bar{u}\|_{L^\infty}$*, where* C_0 *is a universal constant, and* $C=\max_\ell C_\ell$*.*

Proof. Using the midpoint integration formula we have

$$\bar{\mathcal{F}}_\ell = \int_{F_\ell} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{n}_{sr} + \mathcal{O}(|F_\ell|^2), \quad \bar{\mathcal{G}}_\ell = \int_{G_\ell} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{n}_{ij} + \mathcal{O}(|G_\ell|^2),$$

hence

$$\bar{\mathcal{F}}_\ell - \mathcal{F}_\ell(\bar{\mathbf{u}}) = \int_{F_\ell} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{n}_{sr} - \frac{1}{2}\frac{1}{|I_\ell|}\left((\bar{u}(\mathbf{x}_j)-\bar{u}(\mathbf{x}_i))\mathbf{N}_{sr} + (\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r))\mathbf{N}_{ij}\right)\cdot\mathbf{N}_{sr} + \mathcal{O}(|F_\ell|^2),$$

$$\bar{\mathcal{G}}_\ell - \mathcal{G}_\ell(\bar{\mathbf{u}}) = \int_{G_\ell} \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{n}_{ij} - \frac{1}{2}\frac{1}{|I_\ell|}\left((\bar{u}(\mathbf{x}_j)-\bar{u}(\mathbf{x}_i))\mathbf{N}_{sr} + (\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r))\mathbf{N}_{ij}\right)\cdot\mathbf{N}_{ij} + \mathcal{O}(|G_\ell|^2).$$

Since

$$|I_\ell| = \frac{1}{2}\cos(\theta_\ell)|F_\ell||G_\ell|, \quad \mathbf{N}_{sr} = -\frac{1}{\cos(\theta_\ell)}\frac{|F_\ell|}{|G_\ell|}\mathbf{N}_{ij}^\perp + \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)}\mathbf{N}_{sr}^\perp, \quad \mathbf{N}_{ij} = \frac{1}{\cos(\theta_\ell)}\frac{|G_\ell|}{|F_\ell|}\mathbf{N}_{sr}^\perp - \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)}\mathbf{N}_{ij}^\perp,$$

we obtain

$$\begin{aligned}
\bar{\mathcal{F}}_\ell - \mathcal{F}_\ell(\bar{\mathbf{u}}) &= -\frac{1}{\cos(\theta_\ell)}\frac{|F_\ell|}{|G_\ell|}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{ij}^\perp + \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{sr}^\perp \\
&\quad - \frac{1}{\cos(\theta_\ell)}\frac{|F_\ell|}{|G_\ell|}(\bar{u}(\mathbf{x}_j)-\bar{u}(\mathbf{x}_i)) - \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)}(\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r)) + \mathcal{O}(|F_\ell|^2), \\
\bar{\mathcal{G}}_\ell - \mathcal{G}_\ell(\bar{\mathbf{u}}) &= \frac{1}{\cos(\theta_\ell)}\frac{|G_\ell|}{|F_\ell|}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{sr}^\perp - \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)}\boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{ij}^\perp \\
&\quad - \frac{\sin(\theta_\ell)}{\cos(\theta_\ell)}(\bar{u}(\mathbf{x}_j)-\bar{u}(\mathbf{x}_i)) - \frac{1}{\cos(\theta_\ell)}\frac{|G_\ell|}{|F_\ell|}(\bar{u}(\mathbf{x}_s)-\bar{u}(\mathbf{x}_r)) + \mathcal{O}(|G_\ell|^2).
\end{aligned}$$

Using Taylor expansions in the neighborhood of $\mathbf{x}_\ell$

$$\bar{u}(\mathbf{x}_j) = \bar{u}(\mathbf{x}_i) - \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{ij}^\perp + \mathcal{O}(|G_\ell|^2), \quad \bar{u}(\mathbf{x}_s) = \bar{u}(\mathbf{x}_r) + \boldsymbol{\nabla}\bar{u}(\mathbf{x}_\ell)\cdot\mathbf{N}_{sr}^\perp + \mathcal{O}(|F_\ell|^2),$$

we deduce (35). □

A.2 Discrete Poincaré inequality

Lemma A.2 (Discrete Poincaré inequality)**.** *Assume that* **H2** *and* **H3** *are satisfied. Consider* $\mathbf{e} = \left(\mathbf{e}^{primal}, \mathbf{e}^{dual}\right) \in \mathbb{R}^{n+m}$*, where* $\mathbf{e}^{primal} = (e_i)_{1\le i\le n}$ *and* $\mathbf{e}^{dual} = (e_r)_{1\le r\le m}$*. Assume moreover that*

$$\forall r \in \partial\Omega, \quad e_r = 0. \tag{36}$$

Then we have

$$\left(\sum_i |P_i| e_i^2 + \sum_r |D_r| e_r^2\right)^{1/2} \le 2\sqrt{2}\,\mathrm{diam}(\Omega)\frac{\sqrt{N_{\max}\xi}}{\cos(\theta_0)}\left(\sum_\ell |I_\ell|\left(\left(\frac{e_j-e_i}{|G_\ell|}\right)^2 + \left(\frac{e_s-e_r}{|F_\ell|}\right)^2\right)\right)^{1/2},$$

where we use the convention that, if $\ell \subset \partial\Omega$*, then* $e_i - e_j = e_i$ *and the constants* $N_{\max}$*,* ξ*,* θ_0 *are definied by* **H2** *and* **H3***.*

Proof. Given a point $\mathbf{x} \in \Omega$, let $\mathbf{y}(\mathbf{x})$ be the (first) point of intersection between the horizontal half line (for example) passing through $\mathbf{x}$ and the boundary $\partial\Omega$ (see Fig. 14). For any primal face F_ℓ, let $\chi_\ell : \Omega \longrightarrow \{0,1\}$ be defined by

$$\chi_\ell(\mathbf{x}) = \begin{cases} 1 & \text{if } F_\ell \cap [\mathbf{x}, \mathbf{y}(\mathbf{x})] \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

We note that

$$\int_\Omega \chi_\ell \leq \mathrm{diam}(\Omega)|F_\ell|, \tag{37}$$

where $\mathrm{diam}(\Omega) = \max_{\mathbf{x},\mathbf{y}\in\Omega} \|\mathbf{x}-\mathbf{y}\|$ is the diameter of Ω.

Fixing $\mathbf{x} \in P_i$, we write e_i^2 as a telescopic sum along the segment $[\mathbf{x}, \mathbf{y}(\mathbf{x})]$, that is,

$$e_i^2 = e_i^2 - e_j^2 + ... + e_k^2 - e_\ell^2,$$

where the difference $e_\ell = 0$, hence

$$|e_i^2| \leq |e_i^2 - e_j^2 + ... + e_k^2 - e_\ell^2| \leq \sum_\ell |e_i^2 - e_j^2|,$$

with the convention that, in the right hand side, if $\ell \subset \partial\Omega$, then $e_j = e_\ell = 0$.

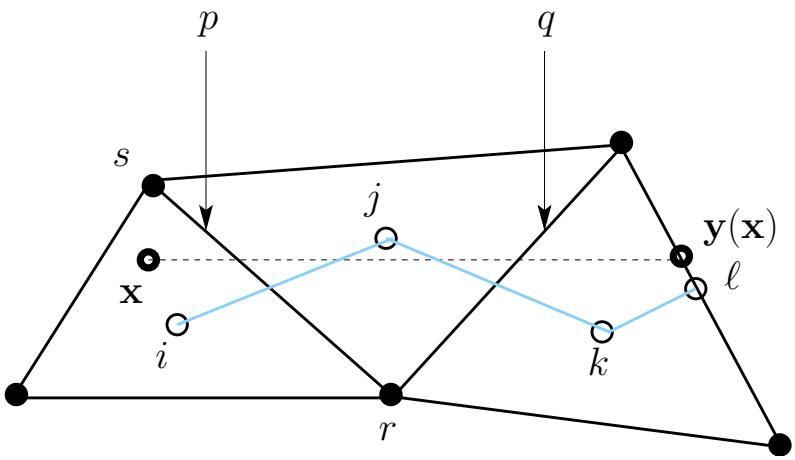


Figure 14: An example of a sequel of three adjacent primal cells P_i, P_j, P_k and a horizontal (dashed) half line coming from the point $\mathbf{x} \in P_i$ and intersecting the interior faces F_p, F_q and the boundary face F_ℓ at point $\mathbf{y}(\mathbf{x})$.

The definition of χ_ℓ allows to write this as follow

$$e_i^2 \leq \sum_\ell |e_j^2 - e_i^2| \chi_\ell(\mathbf{x}),$$

where the sum runs over all faces F_ℓ such that $F_\ell \cap [\mathbf{x}, \mathbf{y}(\mathbf{x})]$. Integrating this inequality over P_i with respect to $\mathbf{x}$, we have

$$\int_{P_i} e_i^2 = |P_i| e_i^2 \leq \sum_\ell |e_j^2 - e_i^2| \int_{P_i} \chi_\ell.$$

Using (37), we deduce that

$$\sum_i |P_i| e_i^2 \leq \sum_i \left(\sum_\ell |e_j^2 - e_i^2| \int_{P_i} \chi_\ell \right) \leq \sum_\ell |e_j^2 - e_i^2| \int_\Omega \chi_\ell \leq \mathrm{diam}(\Omega) \sum_\ell |F_\ell| |e_j^2 - e_i^2|,$$

that is to say,

$$\sum_i |P_i| e_i^2 \leq \mathrm{diam}(\Omega) \sum_\ell |F_\ell| |e_j^2 - e_i^2|. \tag{38}$$

Noting that

$$\sum_\ell |F_\ell||e_j^2 - e_i^2| = \sum_\ell \frac{1}{\cos(\theta_\ell)} \left(\cos(\theta_\ell)|F_\ell||G_\ell|\right)^{1/2} \frac{|e_j - e_i|}{|G_\ell|} \left(\cos(\theta_\ell)|F_\ell||G_\ell|\right)^{1/2} |e_j + e_i|,$$

and using assumption **H1**, we obtain

$$\sum_\ell |F_\ell||e_j^2 - e_i^2| \leq \sum_\ell \frac{1}{\cos(\theta_0)} \left(\cos(\theta_\ell)|F_\ell||G_\ell|\right)^{1/2} \frac{|e_j - e_i|}{|G_\ell|} \left(\cos(\theta_\ell)|F_\ell||G_\ell|\right)^{1/2} (|e_j| + |e_i|).$$

Hence, using the Cauchy-Schwarz inequality and recalling that

$$|I_\ell| = \frac{1}{2} \cos(\theta_\ell) |F_\ell||G_\ell|,$$

we infer

$$\sum_\ell |F_\ell||e_j^2 - e_i^2| \leq \frac{2}{\cos(\theta_0)} \left(\sum_\ell |I_\ell| \left(\frac{e_j - e_i}{|G_\ell|} \right)^2 \right)^{1/2} \left(\sum_\ell |I_\ell| (|e_j| + |e_i|)^2 \right)^{1/2}.$$

Since

$$(|e_i| + |e_j|)^2 \leq 2(|e_i|^2 + |e_j|^2),$$

this gives

$$\sum_\ell |F_\ell||e_j^2 - e_i^2| \leq \frac{2\sqrt{2}}{\cos(\theta_0)} \left(\sum_\ell |I_\ell| \left(\frac{e_j - e_i}{|G_\ell|} \right)^2 \right)^{1/2} \left(\sum_\ell |I_\ell| (|e_j|^2 + |e_i|^2) \right)^{1/2}. \tag{39}$$

Taking into account assumptions **H2** and **H3** we have

$$\sum_\ell |I_\ell|(e_i^2 + e_j^2) \leq \xi \sum_\ell (|P_i|e_i^2 + |P_j|e_j^2) \leq N_{\max} \xi \sum_i |P_i| e_i^2.$$

Inserting this estimate into (39), we deduce that

$$\sum_\ell |F_\ell||e_j^2 - e_i^2| \leq 2\sqrt{2} \frac{\sqrt{N_{\max}\xi}}{\cos(\theta_0)} \left(\sum_\ell |I_\ell| \left(\frac{e_j - e_i}{|G_\ell|} \right)^2 \right)^{1/2} \left(\sum_i |P_i| e_i^2 \right)^{1/2}.$$

Using Equation (38) gives

$$\left(\sum_i |P_i| e_i^2 \right)^{1/2} \leq 2\sqrt{2}\, \mathrm{diam}(\Omega) \frac{\sqrt{N_{\max}\xi}}{\cos(\theta_0)} \left(\sum_\ell |I_\ell| \left(\frac{e_j - e_i}{|G_\ell|} \right)^2 \right)^{1/2}. \tag{40}$$

Applying the same argument to the dual mesh, we also have

$$\left(\sum_r |D_r| e_r^2 \right)^{1/2} \leq 2\sqrt{2}\, \mathrm{diam}(\Omega) \frac{\sqrt{N_{\max}\xi}}{\cos(\theta_0)} \left(\sum_\ell |I_\ell| \left(\frac{e_s - e_r}{|F_\ell|} \right)^2 \right)^{1/2}. \tag{41}$$

Collecting (40) and (41), we obtain

$$\left(\sum_i |P_i| e_i^2 + \sum_r |D_r| e_r^2 \right)^{1/2} \leq 2\sqrt{2}\, \mathrm{diam}(\Omega) \frac{\sqrt{N_{\max}\xi}}{\cos(\theta_0)} \left(\sum_\ell |I_\ell| \left(\left(\frac{e_j - e_i}{|G_\ell|} \right)^2 + \left(\frac{e_s - e_r}{|F_\ell|} \right)^2 \right) \right)^{1/2},$$

which concludes the proof. □

A.3 Convergence

Proposition A.3 (Convergence of the DDFV scheme)**.** *Let $e_i = \bar{u}(\mathbf{x}_i) - u_i$ $(1 \le i \le n)$ and $e_r = \bar{u}(\mathbf{x}_r) - u_r$ $(1 \le r \le m)$, where $\mathbf{u}$ is the solution of System* (34)*. Assume that* **H1**, **H2**, **H3** *are satisfied. Then we have*

$$\left(\sum_i |P_i| e_i^2 + \sum_r |D_r| e_r^2 \right)^{1/2} \le C_1 h,$$

where C_1 is a constant independent of h.

Proof. The fluxes $\bar{\mathcal{F}}_\ell$, $\mathcal{F}_\ell(\mathbf{u})$, $\bar{\mathcal{G}}_\ell$, $\mathcal{G}_\ell(\mathbf{u})$ are such that

$$-\sum_{\ell \in i} \bar{\mathcal{F}}_\ell = -\sum_{\ell \in i} \mathcal{F}_\ell(\mathbf{u}) = \int_{P_i} f \quad \text{and} \quad -\sum_{\ell \in r} \bar{\mathcal{G}}_\ell = -\sum_{\ell \in r} \mathcal{G}_\ell(\mathbf{u}) = \int_{D_r} f.$$

Therefore,

$$\sum_{\ell \in i} \bar{\mathcal{F}}_\ell = \sum_{\ell \in i} \mathcal{F}_\ell(\mathbf{u}), \quad \sum_{\ell \in r} \bar{\mathcal{G}}_\ell = \sum_{\ell \in r} \mathcal{G}_\ell(\mathbf{u}).$$

Given $e_i = \bar{u}(\mathbf{x}_i) - \mathbf{u}_i$ and $e_r = \bar{u}(\mathbf{x}_r) - \mathbf{u}_r$ we deduce that

$$\sum_{\ell \in i} \left(\mathcal{F}_\ell(\bar{\mathbf{u}}) - \mathcal{F}_\ell(\mathbf{u}) \right) = \sum_{\ell \in i} \left(\mathcal{F}_\ell(\bar{\mathbf{u}}) - \bar{\mathcal{F}}_\ell \right) = \frac{1}{2} \sum_{\ell \in i} \left(\frac{1}{|I_\ell|} \left((e_j - e_i)\mathbf{N}_{sr} + (e_s - e_r)\mathbf{N}_{ij} \right) \cdot \mathbf{N}_{sr} \right),$$
$$\sum_{\ell \in r} \left(\mathcal{G}_\ell(\bar{\mathbf{u}}) - \mathcal{G}_\ell(\mathbf{u}) \right) = \sum_{\ell \in r} \left(\mathcal{G}_\ell(\bar{\mathbf{u}}) - \bar{\mathcal{G}}_\ell \right) = \frac{1}{2} \sum_{\ell \in r} \left(\frac{1}{|I_\ell|} \left((e_j - e_i)\mathbf{N}_{sr} + (e_s - e_r)\mathbf{N}_{ij} \right) \cdot \mathbf{N}_{ij} \right).$$

Multiplying these relations respectively by e_i and e_r and summing over the primal cells P_i and dual cells D_r, we obtain

$$\sum_i e_i \sum_{\ell \in i} (\mathcal{F}_\ell(\bar{\mathbf{u}}) - \bar{\mathcal{F}}_\ell) + \sum_r e_r \sum_{\ell \in r} (\mathcal{G}_\ell(\bar{\mathbf{u}}) - \bar{\mathcal{G}}_\ell) = \frac{1}{2} \sum_i \sum_{\ell \in i} \left(e_i \frac{1}{|I_\ell|} \left((e_j - e_i)\mathbf{N}_{sr} + (e_s - e_r)\mathbf{N}_{ij} \right) \cdot \mathbf{N}_{sr} \right)$$
$$+ \frac{1}{2} \sum_r \sum_{\ell \in r} \left(e_r \frac{1}{|I_\ell|} \left((e_j - e_i)\mathbf{N}_{sr} + (e_s - e_r)\mathbf{N}_{ij} \right) \cdot \mathbf{N}_{ij} \right).$$

Exchanging the sums, this reads as

$$\sum_\ell \left(\left(\mathcal{F}_\ell(\bar{\mathbf{u}}) - \bar{\mathcal{F}}_\ell \right)(e_j - e_i) + \left(\mathcal{G}_\ell(\bar{\mathbf{u}}) - \bar{\mathcal{G}}_\ell \right)(e_s - e_r) \right)$$
$$= \frac{1}{2} \sum_\ell \frac{1}{|I_\ell|} \left((e_j - e_i)^2 \mathbf{N}_{sr} \cdot \mathbf{N}_{sr} + (e_s - e_r)^2 \mathbf{N}_{ij} \cdot \mathbf{N}_{ij} + 2(e_j - e_i)(e_s - e_r)\mathbf{N}_{sr} \cdot \mathbf{N}_{ij} \right)$$
$$= 2 \sum_\ell \frac{1}{\cos(\theta_\ell)^2} |I_\ell| \left(\left(\frac{e_j - e_i}{|G_\ell|} \right)^2 + \left(\frac{e_s - e_r}{|F_\ell|} \right)^2 + 2\sin(\theta_\ell) \frac{e_j - e_i}{|G_\ell|} \frac{e_s - e_r}{|F_\ell|} \right). \quad (42)$$

This expression is nonnegative owing to the following inequality, which holds for all $X, Y \in \mathbb{R}^n$

$$X^2 + Y^2 \le \frac{1}{1 - |\sin(\theta_\ell)|} \left(X^2 + Y^2 + 2\sin(\theta_\ell) XY \right) = \frac{1 + |\sin(\theta_\ell)|}{\cos(\theta_\ell)^2} \left(X^2 + Y^2 + 2\sin(\theta_\ell) XY \right). \quad (43)$$

Estimate (43) and equality (42) imply

$$\sum_\ell |I_\ell| \left(\left(\frac{e_j - e_i}{|G_\ell|} \right)^2 + \left(\frac{e_s - e_r}{|F_\ell|} \right)^2 \right)$$

$$\leq \sum_\ell \frac{1+|\sin(\theta_\ell)|}{\cos(\theta_\ell)^2}|I_\ell|\left(\left(\frac{e_j-e_i}{|G_\ell|}\right)^2+\left(\frac{e_s-e_r}{|F_\ell|}\right)^2+2\sin(\theta_\ell)\frac{e_j-e_i}{|G_\ell|}\frac{e_s-e_r}{|F_\ell|}\right)$$

$$\leq 2\sum_\ell \frac{1}{\cos(\theta_\ell)^2}|I_\ell|\left(\left(\frac{e_j-e_i}{|G_\ell|}\right)^2+\left(\frac{e_s-e_r}{|F_\ell|}\right)^2+2\sin(\theta_\ell)\frac{e_j-e_i}{|G_\ell|}\frac{e_s-e_r}{|F_\ell|}\right)$$

$$= \sum_\ell \left(\left(\mathcal{F}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{F}}_\ell\right)(e_j-e_i)+\left(\mathcal{G}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{G}}_\ell\right)(e_s-e_r)\right).$$

Using the Cauchy-Schwarz inequality we obtain

$$\sum_\ell \left((\mathcal{F}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{F}}_\ell)(e_j-e_i)+(\mathcal{G}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{G}}_\ell)(e_s-e_r)\right)$$

$$= \sum_\ell \left(\frac{|G_\ell|}{|I_\ell|^{1/2}}\left(\mathcal{F}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{F}}_\ell\right)\frac{|I_\ell|^{1/2}}{|G_\ell|}(e_j-e_i)+\frac{|F_\ell|}{|I_\ell|^{1/2}}\left(\mathcal{G}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{G}}_\ell\right)\frac{|I_\ell|^{1/2}}{|F_\ell|}(e_s-e_r)\right)$$

$$\leq \left(\sum_\ell \frac{|G_\ell|^2}{|I_\ell|}\left(\mathcal{F}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{F}}_\ell\right)^2+\frac{|F_\ell|^2}{|I_\ell|}\left(\mathcal{G}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{G}}_\ell\right)^2\right)^{1/2}\left(\sum_\ell |I_\ell|\left(\left(\frac{e_j-e_i}{|G_\ell|}\right)^2+\left(\frac{e_s-e_r}{|F_\ell|}\right)^2\right)\right)^{1/2},$$

hence

$$\left(\sum_\ell |I_\ell|\left(\left(\frac{e_j-e_i}{|G_\ell|}\right)^2+\left(\frac{e_s-e_r}{|F_\ell|}\right)^2\right)\right)^{1/2} \leq \left(\sum_\ell \frac{|G_\ell|^2}{|I_\ell|}\left(\mathcal{F}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{F}}_\ell\right)^2+\frac{|F_\ell|^2}{|I_\ell|}\left(\mathcal{G}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{G}}_\ell\right)^2\right)^{1/2}. \quad (44)$$

Applying the consistency of fluxes (35) we have

$$\sum_\ell \frac{|G_\ell|^2}{|I_\ell|}\left(\mathcal{F}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{F}}_\ell\right)^2+\frac{|F_\ell|^2}{|I_\ell|}\left(\mathcal{G}_\ell(\bar{\mathbf{u}})-\bar{\mathcal{G}}_\ell\right)^2$$

$$\leq 4\sum_\ell \frac{C_\ell^2}{\cos(\theta_\ell)^4}\left(((1+|\sin(\theta_\ell)|)|F_\ell|+|G_\ell|)^2+((1+|\sin(\theta_\ell)|)|G_\ell|+|F_\ell|)^2\right)$$

$$\leq 8\frac{C^2}{\cos(\theta_0)^4}|\Omega|(2+\sigma)^2h^2, \quad (45)$$

with

$$C=\max_\ell C_\ell, \quad \sigma=\max_\ell |\sin(\theta_\ell)|.$$

Inserting (45) into (44), we deduce

$$\left(\sum_\ell |I_\ell|\left(\left(\frac{e_j-e_i}{|G_\ell|}\right)^2+\left(\frac{e_s-e_r}{|F_\ell|}\right)^2\right)\right)^{1/2} \leq \sqrt{8}\frac{C}{\cos(\theta_0)^2}|\Omega|^{\frac{1}{2}}(2+\sigma)h. \quad (46)$$

Applying Lemma A.2 to the left-hand side of Equation (46), we conclude that

$$\left(\sum_i |P_i|e_i^2+\sum_r |D_r|e_r^2\right)^{1/2} \leq C_1 h,$$

with

$$C_1 = 8\,\mathrm{diam}(\Omega)|\Omega|^{1/2}\frac{C}{\cos(\theta_0)^3}(2+\sigma)\sqrt{N_{\max}\xi},$$

hence the method is (at least) first-order convergent. □

A.4 Coercivity

Lemma A.4 (Coercivity)**.** *Let* $\mathbf{A}$ *be the matrix associated with the DDFV discretization* (34) *of equation* (33)*. There exists a constant* C_2 *independent of* h *such that*

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \|\mathbf{u}\|_2^2 \leq C_2 \mathbf{u}^t \mathbf{A} \mathbf{u}.$$

Proof. Owing to the identity

$$|I_\ell| = \frac{1}{2}\cos(\theta_\ell)|F_\ell||G_\ell|,$$

we have

$$\begin{aligned}
\mathbf{u}^t\mathbf{A}\mathbf{u} &= \frac{1}{2}\sum_{\ell\notin\partial\Omega}\frac{1}{|I_\ell|}\|\mathbf{N}_{sr}(u_j-u_i)+\mathbf{N}_{ij}(u_s-u_r)\|^2 + \frac{1}{2}\sum_{\ell\in\partial\Omega}\left(\frac{1}{|I_\ell|}\|\mathbf{N}_{sr}(u_\ell-u_i)+\mathbf{N}_{i\ell}(u_s-u_r)\|^2\right)\\
&= 2\sum_{\ell\notin\partial\Omega}\frac{1}{\cos(\theta_\ell)^2}|I_\ell|\left(\left(\frac{u_j-u_i}{|G_\ell|}\right)^2+\left(\frac{u_s-u_r}{|F_\ell|}\right)^2+2\sin(\theta_\ell)\frac{u_j-u_i}{|G_\ell|}\frac{u_s-u_r}{|F_\ell|}\right)\\
&+2\sum_{\ell\in\partial\Omega}\frac{1}{\cos(\theta_\ell)^2}|I_\ell|\left(\left(\frac{u_\ell-u_i}{|G_\ell|}\right)^2+\left(\frac{u_s-u_r}{|F_\ell|}\right)^2+2\sin(\theta_\ell)\frac{u_\ell-u_i}{|G_\ell|}\frac{u_s-u_r}{|F_\ell|}\right).
\end{aligned}$$

As we have assumed that $u = g = 0$ on $\partial\Omega$ we can use the Lemma A.2 to $\mathbf{u} = ((u_i)_{1\leq i\leq n}, (u_r)_{1\leq r\leq m})$ instead of $\mathbf{e} = ((e_i)_{1\leq i\leq n}, (e_r)_{1\leq r\leq m})$. Therefore there exists a constant C_2 independent of h such that

$$\left(\sum_i |P_i|u_i^2 + \sum_r |D_r|u_r^2\right)^{1/2} \leq C_2\left(\sum_\ell |I_\ell|\left(\left(\frac{u_j-u_i}{|G_\ell|}\right)^2+\left(\frac{u_s-u_r}{|F_\ell|}\right)^2\right)\right)^{1/2}.$$

Using inequality (43), we have

$$\begin{aligned}
&\sum_\ell |I_\ell|\left(\left(\frac{u_j-u_i}{|G_\ell|}\right)^2+\left(\frac{u_s-u_r}{|F_\ell|}\right)^2\right)\\
&\leq 2\sum_\ell \frac{1}{\cos(\theta_\ell)^2}|I_\ell|\left(\left(\frac{u_j-u_i}{|G_\ell|}\right)^2+\left(\frac{u_s-u_r}{|F_\ell|}\right)^2+2\sin(\theta_\ell)\frac{u_j-u_i}{|G_\ell|}\frac{u_s-u_r}{|F_\ell|}\right),
\end{aligned}$$

which allows to conclude the proof. □

A.5 Stability

Lemma A.5 (Stability)**.** *Let* $\mathbf{u}$ *be the solution to* (34)*. We have*

$$\|\mathbf{u}\|_2 \leq C_2\|\mathbf{f}\|_2,$$

where C_2 *does not depend on* $\mathbf{u}$*,* $\mathbf{f}$ *and* h*.*

Proof. We have

$$\mathbf{u}^t\mathbf{A}\mathbf{u} = \sum_i |P_i|f_iu_i + \sum_r |D_r|f_ru_r,$$

hence, owing to the Cauchy-Schwarz inequality

$$\mathbf{u}^t\mathbf{A}\mathbf{u} \leq \left(\sum_i |P_i|f_i^2 + \sum_r |D_r|f_r^2\right)^{1/2}\left(\sum_i |P_i|u_i^2 + \sum_r |D_r|u_r^2\right)^{1/2} = \|\mathbf{f}\|_2\|\mathbf{u}\|_2.$$

Now, thanks to lemma A.4, we obtain

$$\|\mathbf{u}\|_2^2 \leq C_2\mathbf{u}^t\mathbf{A}\mathbf{u} \leq C_2\|\mathbf{f}\|_2\|\mathbf{u}\|_2,$$

which allows to conclude. □

References

[1] I. Aavatsmark, G.T. Eigestad, R.A. Klausen, M.F. Wheeler, and I. Yotov. Convergence of a symmetric MPFA method on quadrilateral grids. *Comput. Geosci.*, 11(4):333–345, 2007.

[2] B. Andreianov, F. Boyer, and F. Hubert. Discrete duality finite volume schemes for Leray-Lions type elliptic problems on general 2D meshes. *Numer. Methods Partial Differ. Eq.*, 23:pp 145–195, 2007.

[3] E. Bertolazzi and G. Manzini. A second-order maximum principle preserving finite volume method for steady convection-diffusion problems. *SIAM J. Numer. Anal.*, 43(5):2172–2199, 2005.

[4] X. Blanc, F. Hermeline, E. Labourasse, and J. Patela. High-order monotone finite-volume schemes for 1D elliptic problems. *HAL ID: cea-03421015*, 2022.

[5] X. Blanc and E. Labourasse. A positive scheme for diffusion problems on deformed meshes. *Z. Angew. Math. Mech.*, 96(6):660–680, 2016.

[6] J.-S. Camier and F. Hermeline. A monotone nonlinear finite volume method for approximating diffusion operators on general meshes. *Int. J. Numer. Meth. Engng*, 107:496–519, 2016.

[7] C. Cancès and C. Guichard. Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure. *Found. Comput. Math.*, 17:1525–1584, 2017.

[8] F. Cao, Y. Yao, Y. Yu, and G. Yuan. A conservative enforcing positivity-preserving algorithm for diffusion scheme on general meshes. *Int. J. Numer. Anal. Model.*, 13(5):739–752, 2016.

[9] P. Ciarlet. Discrete maximum principle for finite-difference operators. *Aeq. Math.*, 4:338–352, 1970.

[10] P. Ciarlet and P.-A. Raviart. Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure. *Found. Comput. Math.*, 2:17–31, 1973.

[11] B. Després. Non linear schemes for the heat equation in 1D. *ESAIM: M2AN*, 48(1):107–134, 2014.

[12] K. Domelevo and P. Omnes. A finite volume method for the laplace equation on almost arbitrary two-dimensional grids. *ESAIM: M2AN*, 39(6):1203–1249, 2005.

[13] J. Droniou and C. Le Potier. Construction and convergence study of schemes preserving the elliptic local maximum principle. *SIAM J. Numer. Anal.*, 49(2):459–490, 2011.

[14] M. Dumbser, W. Boscheri, M. Semplice, and G. Russo. Central weighted eno schemes for hyperbolic conservation laws on fixed and moving unstructured meshes. *SIAM J. Sci. Comput.*, 39(6):A2564–A2591, 2017.

[15] L. Evans. Application of nonlinear semigroup theory to certain partial differential equations. In Michael G. Crandall, editor, *Nonlinear Evolution Equations*, pages 163–188. Academic Press, 1978.

[16] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In Ph. G. Ciarlet and J.-L. Lions, editors, *Handbook of numerical analysis*, volume VII. North-Holland, Amsterdam, 2000.

[17] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: A scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.*, 30(4):1009–1043, 2010.

[18] Y. Gao, G. Yuan, S. Wang, and X. Hang. A finite volume element scheme with a monotonicity correction for anisotropic diffusion problems on general quadrilateral meshes. *J. Comput. Phys.*, 407:109143, 2020.

[19] Z. Gao and J. Wu. A second-order positivity-preserving finite volume scheme for diffusion equations on general meshes. *SIAM Journal on Scientific Computing*, 37(1):A420–A438, 2015.

[20] R. Herbin and F. Hubert. Benchmark on discretization schemes for anisotropic diffusion problems on general grids. In R. Eymard and J.-M. Herard, editors, *Finite volume for complex applications, problems and perspectives V*. Wiley, 2008.

[21] F. Hermeline. A finite volume method for the approximation of diffusion operators on distorted meshes. *J. Comput. Phys.*, 160(2):481–499, 2000.

[22] F. Hermeline. Approximation of diffusion operators with discontinuous tensor coefficients on distorted meshes. *Comput. Methods Appl. Mech. Eng.*, 192(16-18):1939–1959, 2003.

[23] F. Hermeline. *Nouvelles méthodes de volumes finis pour approcher des équations aux dérivées partielles sur des maillages quelconques.* Habilitation à diriger des recherches, CEA/DAM Ile de France, 2008.

[24] J. Karátson, S. Korotov, and M. Křížek. On discrete maximum principles for nonlinear elliptic problems. *Math. Comput. Simul.*, 76(1):99–108, 2007.

[25] S. Korotov, M. Křížek, and P. Neittaanmäki. Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle. *Math. Comp.*, 70(233):107–119, 2000.

[26] M. Käser and A. Iske. Ader schemes on adaptive triangular meshes for scalar conservation laws. *J. Comput. Phys.*, 205(2):486–508, 2005.

[27] O. Larroche. An efficient explicit numerical scheme for diffusion-type equations with a highly inhomogeneous and highly anisotropic diffusion tensor. *J. Comput. Phys.*, 223:436–450, 2007.

[28] C. Le Potier. Schéma volumes finis monotone pour des opérateurs de diffusion fortement anisotropes sur des maillages de triangles non structurés. *C. R. Math.*, 341(12):787–792, 2005.

[29] C. Le Potier. Correction non linéaire et principe du maximum pour la discrétisation d'opérateurs de diffusion avec des schémas volumes finis centrés sur les mailles. *C. R. Math.*, 348(11-12):691–695, 2010.

[30] K. Lipnikov, M. Shashkov, D. Svyatskiy, and Yu. Vassilevski. Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes. *J. Comput. Phys.*, 227(1):492–512, 2007.

[31] K. Lipnikov, D. Svyatskiy, and Y. Vassilevski. Interpolation-free monotone finite volume method for diffusion equations on polygonal meshes. *J. Comput. Phys.*, 228(3):703–716, 2009.

[32] K. Lipnikov, D. Svyatskiy, and Y. Vassilevski. Minimal stencil finite volume scheme with the discrete maximum principle. *Russian J. Numer. Anal. Math. Modelling*, 27(4):369–385, 2012.

[33] R. Liska and M. Shashkov. Enforcing the discrete maximum principle for linear finite element solutions of second-order elliptic problems. *Commun. Comput. Phys.*, 3(4):852–877, 2008.

[34] R. Loubère, M. Staley, and B. Wendroff. The repair paradigm: New algorithms and applications to compressible flow. *J. Comput. Phys.*, 211(2):385–404, 2006.

[35] R.J. Plemmons. m-matrix characterizations.i – nonsingular m-matrices. *Linear Algebra and its Applications*, (2):175 – 188.

[36] M. Schneider, L. Agélas, G. Enchéry, and B. Flemisch. Convergence of nonlinear finite volume schemes for heterogeneous anisotropic diffusion on general meshes. *J. Comput. Phys.*, 351:80–107, 2017.

[37] Z. Sheng and G. Yuan. The finite volume scheme preserving extremum principle for diffusion equations on polygonal meshes. *J. Comput. Phys.*, 230(7):2588–2604, 2011.

[38] Z. Sheng and G. Yuan. A new nonlinear finite volume scheme preserving positivity for diffusion equations. *J. Comput. Phys.*, 315:182–193, 2016.

[39] Z. Sheng, J. Yue, and G. Yuan. Monotone finite volume schemes of nonequilibrium radiation diffusion equations on distorted meshes. *SIAM J. Sci. Comput.*, 31(4):2915–2934, 2009.

[40] R. S. Varga. *Matrix iterative analysis*, volume 1. Prentice Hall, 1962.

[41] T. Vejchodskỳ and P. Šolín. Discrete maximum principle for higher-order finite elements in 1D. *Math. Comp.*, 76(260):1833–1846, 2007.

[42] J. Wang, Z. Sheng, and G. Yuan. A finite volume scheme preserving maximum principle with cell-centered and vertex unknowns for diffusion equations on distorted meshes. *Appl. Math. Comput.*, 398(1):1–21, 2021.

[43] S. Wang, G. Yuan, Y. Li, and Z. Sheng. Discrete maximum principle based on repair technique for diamond type scheme of diffusion problems. *Int. J. Numer. Methods Fluids*, 70(9):1188–1205, 2012.

[44] J. Wu and Z. Gao. Interpolation-based second-order monotone finite volume schemes for anisotropic diffusion equations on general grids. *Journal of Computational Physics*, 275:569–588, 2014.

[45] Y. Yao and G. Yuan. Enforcing positivity with conservation for nine-point scheme of nonlinear diffusion equations. *Comput. Methods Appl. Mech. Eng.*, 223:161–172, 2012.

[46] Y. Yu, X. Chen, and G. Yuan. A finite volume scheme preserving maximum principle for the system of radiation diffusion equation with three temperatures. *SIAM J. Sci. Comput.*, 41(1):93–113, 2019.

[47] G. Yuan and Z. Sheng. Analysis of accuracy of a finite volume scheme for diffusion equations on distorted meshes. *J. Comput. Phys.*, 224:1170, 2007.

[48] G. Yuan and Z. Sheng. Monotone finite volume schemes for diffusion equations on polygonal meshes. *J. Comput. Phys.*, 227(12):6288–6312, 2008.

[49] X. Zhang, S. Su, and J. Wu. A vertex-centered and positivity-preserving scheme for anisotropic diffusion problems on arbitrary polygonal grids. *Journal of Computational Physics*, 344:419–436, 2017.

[50] F. Zhao, Z. Sheng, and G. Yuan. A monotone combination scheme of diffusion equations on polygonal meshes. *Z. Angew. Math. Mech.*, 100(5):1–25, 2020.