A positive scheme for diffusion problems on deformed meshes

Xavier Blanc, Emmanuel Labourasse

A positive scheme for diffusion problems on deformed meshes

Xavier Blanc[1,*] **and Emmanuel Labourasse**[2]

[1] Laboratoire Jacques-Louis Lions, Université Paris Diderot, Bâtiment Sophie Germain, 5 rue Thomas Mann, 75205 Paris Cedex 13, FRANCE
[2] CEA, DAM, DIF, 91297 Arpajon Cedex, FRANCE

Received XXXX, revised XXXX, accepted XXXX
Published online XXXX

Key words deformed mesh, finite volume, diffusion, maximum principle, numerical analysis
MSC (2010) 65M08,65M12,04A25

We present in this article a positive finite volume method for diffusion equation on deformed meshes. This method is mainly inspired from [50, 52], and uses auxiliary unknowns at the nodes of the mesh. The flux is computed so as to be a two-point nonlinear flux, giving rise to a matrix which is the transpose of an M-matrix, which ensures that the scheme is positive. A particular attention is given to the computation of the auxiliary unknowns. We propose a new strategy, which aims at providing a scheme easy to implement in a parallel domain decomposition setting. An analysis of the scheme is provided: existence of a solution for the nonlinear system is proved, and the convergence of a fixed-point strategy is studied.

1 Introduction

We consider the unsteady diffusion equation

$$\begin{cases} \partial_t u - \operatorname{div}(\kappa \nabla u) = f, & \text{in } \Omega \times (0,T), \\ \gamma(\kappa \nabla u) \cdot n + \delta u = g & \text{on } \partial\Omega \times (0,T), \end{cases} \tag{1}$$

where Ω is a bounded open domain of $\mathbb{R}^2$, n is the outgoing normal to Ω. The data are $f \in L^2(\Omega)$, $g \in H^{1/2}(\partial\Omega)$, and $\kappa \in L^\infty(\Omega)$ satisfies the ellipticity condition

$$\forall x \in \Omega, \quad \kappa(x) \geq \kappa_0 > 0.$$

The functions γ and δ are smooth functions such that

$$\forall x \in \partial\Omega, \quad \delta(x) \geq 0, \quad \text{and} \quad \gamma(x) \geq \gamma_0 > 0.$$

The system (1) is completed by an initial data

$$u(t=0) = u_0 \in H^1(\Omega). \tag{2}$$

For the sake of simplicity, we restrict our attention to the case of an isotropic diffusion, that is, $\kappa(x) \in \mathbb{R}$. However, a simple adaptation allows to treat any bounded matrix $\kappa(x)$ satisfying ellipticity conditions. We refer to [50] for the details.

Under the above conditions, the system (1)-(2) has a unique solution in $C^0\left([0,T], H^1(\Omega)\right) \cap C^1\left([0,T], L^2(\Omega)\right)$. See for instance [18] for the proof. Further, we have a maximum principle: if $f \geq 0, g \geq 0$ and $u_0 \geq 0$, then $u \geq 0$ in $\Omega \times [0,T]$. The equation being linear, it also implies that, if $m \leq f \leq M$, $m\delta \leq g \leq M\delta$ and $m \leq u_0 \leq M$, then $m \leq u \leq M$.

In the case when system (1)-(2) is coupled with hydrodynamics, and if the hydrodynamic equations are solved using a Lagrangian or ALE method, the mesh may be highly distorted. Further, the unknown u is naturally discretized using a piece-wise constant approximation on this mesh. This is why we are interested in finite-volume approximations of (1)-(2) on deformed meshes.

The scheme needs to be consistent, stable, and reproduce on the discrete level qualitative properties of the system (1)-(2), such as conservativity and maximum principle. Moreover, in the case when the unknown u is a temperature or a

* Corresponding author, e-mail: blanc@ann.jussieu.fr, Phone: +33 157 279 118, Fax: +33 144 277 223

concentration, it is mandatory that u remains positive. Finally, it is desirable that the scheme produces a symmetric matrix, and is linear. However, as we will see below, these last two conditions will be dropped to the advantage of the previous ones. Further, we need the stencil of the scheme to be small, in order to make the parallelization of the scheme amenable. In summary, the ideal scheme should satisfy the following:

1. consistency;
2. stability;
3. conservativity;
4. small stencil (in link with parallelization);
5. maximum principle;
6. symmetry;
7. linearity.

The scheme we present here satisfies all these properties, except 5, 6 and 7. A weaker version of item 5 is proved, that is, the scheme is positive (see below). It should be noted that, in practice, we did not find any example in which the maximum principle is violated.

Many works have been devoted to the present subject. Let us mention those we know of, which might not be exhaustive (see the review article [12] for more details):

- To our knowledge, the first attempt to derive a consistent finite volume scheme for diffusion equations on a deformed mesh was the work of Kershaw [28] (see also [43] for a related scheme). The basic idea is to consider a transformation from a reference quadrilateral mesh to the actual deformed mesh, and write down a standard finite volume scheme in this reference configuration, then transform it in order to obtain a finite volume scheme on the deformed mesh. This scheme gives a symmetric matrix, but was proved to be consistent only on meshes consisting of parallelograms, and does not satisfy the maximum principle.

- Another approach is to use a diamond scheme, which was analyzed in [9]. The idea is to introduce auxiliary unknowns at the nodes of the mesh. These additional unknowns are computed using an interpolation method as functions of the cell unknowns. The node unknowns are then used to compute a second-order approximation of the fluxes. This method converges at order two if the interpolation method is sufficiently precise. However, this scheme is not positive, and the associated matrix is in general not symmetric.

- The method of mixed finite element [45] may be recast in a finite volume formulation. Here, we use additional degrees of freedom at the edges of the mesh. A hybrid formulation of (1) then allows to eliminate the cell unknowns, so that the system we need to solve gives the edge unknowns, and the cell unknowns are then computed accordingly. This scheme is convergent of order two, the corresponding matrix is symmetric, but it is not positive. A numerical analysis of such schemes was given in [2].

- More recently, the discrete duality finite volume method (DDFV) was proposed by F. Hermeline [21–26]. In this method, one also uses additional unknowns localized at the nodes of the mesh. These unknowns are not related to the cell unknowns by an interpolation procedure, but are computed by writing down, on a dual mesh, a diffusion scheme approximating (1). Doing so, one solves two diffusion problems instead of one. This scheme is convergent of order two, even if the mesh contains non-convex cells, and the associated matrices are symmetric, but not positive.

- The mimetic finite difference method is described in the papers [6, 7, 29, 38] (see also the review paper [36]). In this method, the fluxes are considered as additional unknowns. Moreover, the discrete system is designed so that some properties of the continuous system are reproduced, such as, for instance, the Green formula. The scheme is convergent of order two, and the corresponding matrix is symmetric. However, it does not satisfy the maximum principle (see [37]), and the number of degrees of freedom is much larger than in other schemes.

- The scheme using stabilization and harmonic interfaces (SUSHI) was proposed by R. Eymard, T. Gallouët and R. Herbin [19]. It is based on the same ideas as the standard diamond scheme. However, a stabilization term is added to the discrete gradient, which preserves consistency, while improving robustness. This scheme is convergent of order two, the corresponding matrix being symmetric, but the scheme is not positive. It is proved in [13] that, actually, SUSHI and mimetic finite difference methods are part of the same family of methods.

- The multi-point flux approximation (MPFA) [1,5,15] uses additional unknowns at the edges of the mesh, with possibly several of them on each face. These additional unknowns are used to compute a consistent approximation of the flux, and are then eliminated by imposing that the flux is continuous across each edge. These scheme are convergent, except on random meshes, and they give rise to non-symmetric matrices. They do not satisfy the maximum principle (see [16, 17, 20]).

- In [30], a finite difference type scheme has been proposed by C. Le Potier. This scheme is convergent of order one, and satisfies the maximum principle. However, in order to manage this, the stencil is enlarged to a larger distance. Thus, this scheme is difficult to use in a decomposition domain strategy.

- In [51], V. Siess studies a scheme which is convergent of order one, and satisfies the maximum principle. In order to do so, the mesh is replaced by the Voronoï mesh associated with the cell centers of the original mesh. On such a mesh, it is possible (and simple) to write down a second order consistent approximation of the fluxes using the neighbouring cell unknowns. Therefore, this scheme is consistent and satisfies the maximum principle. However, the change of mesh induces an error at order one, so that convergence is only of order one. Further, the stencil may be larger than for standard finite volume methods, and extension to anisotropic diffusion is not obvious.

- In [14], a nonlinear scheme is designed with additional unknowns which are not at the nodes nor at the faces of the mesh. These additional unknowns are computed by interpolation. The scheme is convergent of order two, satisfies the maximum principle, and gives a non-symmetric matrix in general. Its main drawback is that it is nonlinear.

- A scheme which is very similar to the preceding one was proposed in [49], in which the additional unknowns are located at the edges of the mesh. The implementation of this scheme is slightly different from [14], but the properties are very similar: it converges with the same rate, satisfies the maximum principle, produces an non-symmetric matrix, and is nonlinear.

- The scheme proposed in [50,52,53] is based on the same idea, that is, defining a nonlinear scheme so that it is positive. However, our implementation is simpler and more natural, since it uses only two-point fluxes, in the spirit of [3, 31]. This scheme is convergent of order two, and positive, that is, if $f \geq 0$, $g \geq 0$ and $u_0 \geq 0$, then $u \geq 0$. However, since the scheme is nonlinear, this does not imply the maximum principle.

- Finally, monotone nonlinear schemes based on the idea of building consistent two-point flux approximations, but without interpolation of additional unknowns, was proposed in [10,32,33,35,41]. It avoids the problem of interpolation of additional unknowns, satisfies the maximum principle, and is precise at order two. However, in the case of highly deformed meshes, it may be necessary to enlarge the stencil in order to preserve this properties.

The scheme described in [50, 52, 53] is the one we have tested. It satisfies all the points mentioned above, except items 5, 6 and 7. The maximum principle is replaced by positivity of the scheme, which is weaker since the scheme is nonlinear. Since it uses only two-point fluxes, its stencil is reduced to edge-neighbouring cells. This implies that it is very simple to use it in a decomposition domain strategy. Note that we are not able to prove that the scheme is coercive (uniformly with respect to the mesh size). However, the numerical results indicate that it is the case. Apart from studying test cases which are different from [50], our contribution is that we give proof of the existence of a solution to the scheme. We also provide a convergence analysis of the fixed point iteration needed to deal with the nonlinearity of the scheme.

Let us end this introduction by a practical remark: in order to have a positive scheme, one needs either to enlarge the stencil, or to use a nonlinear scheme (see [27, 28]). We use here the second strategy, which leads to a nonlinear scheme even if the problem (1)-(2) is linear. In principle, this induces a higher computational cost, since the nonlinear equation is in general solved using a fixed point strategy. However, for the applications we have in mind, the original problem is nonlinear. Therefore, the model problem (1)-(2) is only an intermediate step to solve a nonlinear problem. In practice, the numerical tests show that it is not necessary to reach convergence of the fixed point strategy. Only one step of this algorithm already gives satisfactory results. The outer loop (for instance a Newton algorithm) ensures consistency when it converges. Hence, the additional cost related to the nonlinearity of the scheme is relatively small, as far as the mesh distortion is not too important. For highly distorted mesh, the convergence of the Newton algorithm may be relatively slow, thereby inducing a more important additional computational cost. However, when applied to a linear problem, the present scheme is much more costly than linear schemes.

The article is organized as follows: in Section 2, we describe the method in detail, and prove that the method is positive and consistent. In Section 3, we give the implementation details. We prove in particular that, if the time step is small enough, the fixed point algorithm converges. Finally, Section 4 gives numerical tests asserting the accuracy and positivity of the scheme.

2 The method of Sheng, Yuan and Yue

2.1 Notation

We give here the notation used throughout the article: we assume that a mesh is given, and denote by

- $\mathcal{K}$ the set of all cells of the mesh;
- $\mathcal{E}$ the set of edges of the mesh;
- $\mathcal{N}$ the set of nodes of the mesh;

For any $K \in \mathcal{K}$, we still denote by K the center of this cell. For any $L \in \mathcal{K}$ sharing an edge with K, we denote by $e = K|L \in \mathcal{E}$ this common edge.

Given an edge e of a cell K, we will denote by $n_{K,e}$ the outer normal to K on e. Finally, we define the global mesh size by

$$\Delta x = \max\{|e|, \quad e \in \mathcal{E}\}. \tag{3}$$

2.2 Discrete fluxes

In order to write down a finite volume scheme for (1)-(2), we integrate it over a cell K:

$$\int_K \frac{\partial u}{\partial t} dx - \int_K \nabla \cdot (\kappa \nabla u) dx = \int_K f dx,$$

and use the divergence formula:

$$\int_K \frac{\partial u}{\partial t} dx - \int_{\partial K} \kappa \nabla u \cdot n_K d\Gamma = \int_K f dx,$$

where n_K is the outer unit normal to cell K, and ∂K is the boundary of K. We define an edge e as the intersection of two neighbouring cells K and L. We note $e = \partial K \cap \partial L = K|L$. We have

$$\int_K \frac{\partial u}{\partial t} dx + \sum_{e \in \partial K} \left(- \int_e \kappa \nabla u \cdot n_{K_e} d\Gamma \right) = \int_K f dx. \tag{4}$$

We set

$$\mathcal{F}_{K,e} = - \int_e \kappa(x) \nabla u(x,t) \cdot n_{K_e} d\Gamma,$$

the flux going out of K through e. The point is to write down an approximation of $\mathcal{F}_{K,e}$ for each edge e, as a function of the cell unknowns.

For this purpose, we introduce additional unknowns at the nodes of the mesh. In a second step, we will express these additional unknowns by an interpolation method. Let us first define the notation: (recall that K and L denote both the cells and the corresponding centers.) The points M_1 and M_2 are nodes of K such that the basis $(\overrightarrow{KM_1}, \overrightarrow{KM_2})$ is direct, and such that the decomposition of $n_{K,e}$ in this basis gives non-negative coefficients. See figure 1. The points M_3 and M_4 are defined in a similar way, with the constraint that the basis $(\overrightarrow{LM_3}, \overrightarrow{LM_4})$ is direct, and that the decomposition of the normal vector n_{Le} in this basis gives non-negative coefficients. We define O_1 (respectively O_2) the intersection between the half line starting at K (respectively L) with direction $n_{K,e}$ (respectively $n_{L,e}$) and the boundary of K (respectively L). We also define the angles

$$\theta_{K_1} = (\overrightarrow{KM_1}, \overrightarrow{KO_1}), \quad \theta_{K_2} = (\overrightarrow{KO_1}, \overrightarrow{KM_2}), \quad \theta_{L_1} = (\overrightarrow{LM_3}, \overrightarrow{LO_2}), \quad \theta_{L_2} = (\overrightarrow{LO_2}, \overrightarrow{LM_4}),$$

$$\theta_K = \theta_{K_1} + \theta_{K_2}, \quad \theta_L = \theta_{L_1} + \theta_{L_2}.$$

Note that the nodes M_i may be different from the endpoints of the edge e (see Figure 2). We then write the vector $n_{K,e}$ as a linear combination of $\overrightarrow{KM_1}$ and $\overrightarrow{KM_2}$. We thus compute $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ such that

$$n_{K_e} = \alpha \frac{\overrightarrow{KM_1}}{\|\overrightarrow{KM_1}\|} + \beta \frac{\overrightarrow{KM_2}}{\|\overrightarrow{KM_2}\|}. \tag{5}$$

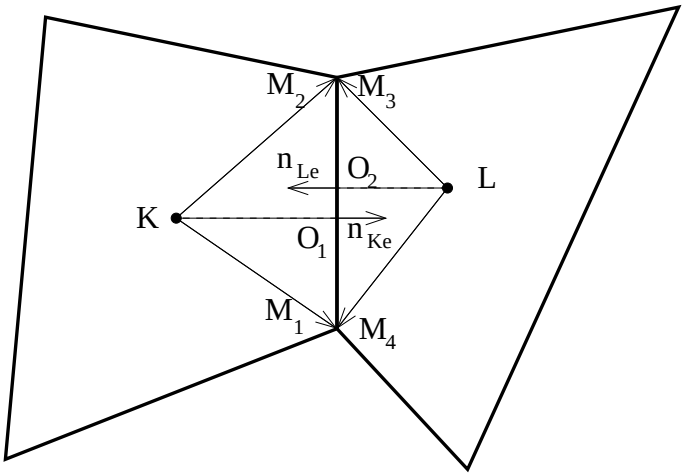

Fig. 1 The cells K and L, their centers, and the points M_i for $1 \leq i \leq 4$.

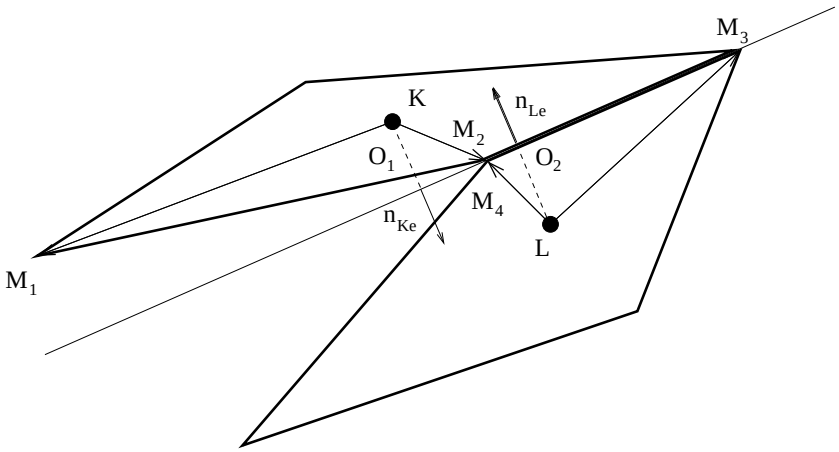

Fig. 2 The cells K, L and the points M_i for $1 \leq i \leq 4$ in the case when these points are not the endpoints of the edge $e = K|L$.

We compute the outer product of (5) with $\overrightarrow{KM_1}$, and find

$$\left.\begin{array}{ll} \|n_{K_e} \wedge \overrightarrow{KM_1}\| & = \|\overrightarrow{KM_1}\| \, |\sin\theta_{K_1}| \\ \|n_{K_e} \wedge \overrightarrow{KM_1}\| & = |\beta| \, \|\overrightarrow{KM_1}\| \, |\sin\theta_K| \end{array}\right\} \Longrightarrow |\beta| = \left|\frac{\sin\theta_{K_1}}{\sin\theta_K}\right|.$$

Since $\theta_K \in (0, \pi]$ and $\theta_{K_1} \in (0, \pi]$, we have $0 < \sin\theta_K \leq 1$ and $0 < \sin\theta_{K_1} \leq 1$ whence

$$\beta = \frac{\sin\theta_{K_1}}{\sin\theta_K}$$

Likewise, $\alpha = \dfrac{\sin\theta_{K_2}}{\sin\theta_K}$. We thus have

$$n_{K_e} = \frac{\sin\theta_{K_2}}{\sin\theta_K} \frac{\overrightarrow{KM_1}}{\|\overrightarrow{KM_1}\|} + \frac{\sin\theta_{K_1}}{\sin\theta_K} \frac{\overrightarrow{KM_2}}{\|\overrightarrow{KM_2}\|}. \tag{6}$$

A similar argument gives

$$n_{L_e} = \frac{\sin\theta_{L_2}}{\sin\theta_L} \frac{\overrightarrow{LM_3}}{\|\overrightarrow{LM_3}\|} + \frac{\sin\theta_{L_1}}{\sin\theta_L} \frac{\overrightarrow{LM_4}}{\|\overrightarrow{LM_4}\|}. \tag{7}$$

We infer

$$\mathcal{F}_{K,e} = -\int_e \left(\frac{\sin\theta_{K_2}}{\sin\theta_K} \frac{\nabla u \cdot \overrightarrow{KM_1}}{\|\overrightarrow{KM_1}\|} + \frac{\sin\theta_{K_1}}{\sin\theta_K} \frac{\nabla u \cdot \overrightarrow{KM_2}}{\|\overrightarrow{KM_2}\|} \right) \kappa(x) d\Gamma$$
$$\mathcal{F}_{L,e} = -\int_e \left(\frac{\sin\theta_{L_2}}{\sin\theta_L} \frac{\nabla u \cdot \overrightarrow{LM_3}}{\|\overrightarrow{LM_3}\|} + \frac{\sin\theta_{L_1}}{\sin\theta_L} \frac{\nabla u \cdot \overrightarrow{LM_4}}{\|\overrightarrow{LM_4}\|} \right) \kappa(x) d\Gamma$$

We use a finite difference approximation to compute the integrand in the above formula:

$$\nabla u \cdot \frac{\overrightarrow{KM_i}}{\|\overrightarrow{KM_i}\|} = \frac{u(M_i) - u(K)}{\|\overrightarrow{KM_i}\|} + O(\Delta x).$$

Thus,

$$\mathcal{F}_{K,e} = -|e| \left(\frac{\sin\theta_{K_2}}{\sin\theta_K} \frac{u_{M_1} - u_K}{\|\overrightarrow{KM_1}\|} + \frac{\sin\theta_{K_1}}{\sin\theta_K} \frac{u_{M_2} - u_K}{\|\overrightarrow{KM_2}\|} \right) \kappa_e + O(\Delta x^2),$$
$$\mathcal{F}_{L,e} = -|e| \left(\frac{\sin\theta_{L_2}}{\sin\theta_L} \frac{u_{M_3} - u_L}{\|\overrightarrow{LM_3}\|} + \frac{\sin\theta_{L_1}}{\sin\theta_L} \frac{u_{M_4} - u_L}{\|\overrightarrow{LM_4}\|} \right) \kappa_e + O(\Delta x^2),$$

where κ_e is the value of $\kappa(x)$ at the center of the edge e. We define

$$F_1 = -|e|\kappa_e \left(\frac{\sin\theta_{K_2}}{\sin\theta_K} \frac{u_{M_1} - u_K}{\|\overrightarrow{KM_1}\|} + \frac{\sin\theta_{K_1}}{\sin\theta_K} \frac{u_{M_2} - u_K}{\|\overrightarrow{KM_2}\|} \right), \tag{8}$$
$$F_2 = -|e|\kappa_e \left(\frac{\sin\theta_{L_2}}{\sin\theta_L} \frac{u_{M_3} - u_L}{\|\overrightarrow{LM_3}\|} + \frac{\sin\theta_{L_1}}{\sin\theta_L} \frac{u_{M_4} - u_L}{\|\overrightarrow{LM_4}\|} \right). \tag{9}$$

In order to use formulas (8) and (9), we need to compute the values u_{M_i} at the nodes M_i, and the value of κ on the edge e. This value is in general computed as a linear combination of κ_K and κ_L, as for instance

$$\kappa_e = \frac{\lambda_{K,e} + \lambda_{L,e}}{\frac{\lambda_{L,e}}{\kappa_L} + \frac{\lambda_{K,e}}{\kappa_K}},$$

where $\lambda_{K,e}$ and $\lambda_{L,e}$ are, respectively, the distance between the center of K (resp. L) and the edge e.

The unknowns u_{M_i} are computed via an interpolation method, for which we refer to Section 3.1 below.

Formulas (8) and (9) give a consistent approximation of the flux through the edge e (up to a change of sign). We are going to combine them in order to obtain a positive scheme. For this purpose, we write

$$F_{K,e} = \mu_1(u)F_1 - \mu_2(u)F_2, \tag{10}$$
$$F_{L,e} = \mu_2(u)F_2 - \mu_1(u)F_1, \tag{11}$$

where the coefficients μ_1 and μ_2 are chosen below. We need these formulas to be consistent, so we impose $\mu_1 + \mu_2 = 1$. The idea of [50] is to compute μ_1 and μ_2 in such a way that (10) and (11) are two-point approximations of the flux. Going back to (8) and (9), we insert them into (10) and (11):

$$\begin{aligned} F_{K,e} &= \mu_1 F_1 - \mu_2 F_2 \\ &= -\mu_1|e|\kappa_e \left(\frac{\sin\theta_{K_2}}{\sin\theta_K} \frac{u_{M_1} - u_K}{\|\overrightarrow{KM_1}\|} + \frac{\sin\theta_{K_1}}{\sin\theta_K} \frac{u_{M_2} - u_K}{\|\overrightarrow{KM_2}\|} \right) \\ &\quad +\mu_2|e|\kappa_e \left(\frac{\sin\theta_{L_2}}{\sin\theta_L} \frac{u_{M_3} - u_L}{\|\overrightarrow{LM_3}\|} + \frac{\sin\theta_{L_1}}{\sin\theta_L} \frac{u_{M_4} - u_L}{\|\overrightarrow{LM_4}\|} \right). \end{aligned}$$

Thus,

$$\begin{aligned} F_{K,e} &= \mu_1|e|\kappa_e\left(\frac{\sin\theta_{K_1}}{\sin\theta_K}\frac{1}{\|\overrightarrow{KM_2}\|}+\frac{\sin\theta_{K_2}}{\sin\theta_K}\frac{1}{\|\overrightarrow{KM_1}\|}\right)u_K \\ &\quad -\mu_2|e|\kappa_e\left(\frac{\sin\theta_{L_1}}{\sin\theta_L}\frac{1}{\|\overrightarrow{LM_4}\|}+\frac{\sin\theta_{L_2}}{\sin\theta_L}\frac{1}{\|\overrightarrow{LM_3}\|}\right)u_L \\ &\quad \underbrace{-\mu_1|e|\kappa_e\left(\frac{\sin\theta_{K_1}}{\sin\theta_K}\frac{u_{M_2}}{\|\overrightarrow{KM_2}\|}+\frac{\sin\theta_{K_2}}{\sin\theta_K}\frac{u_{M_1}}{\|\overrightarrow{KM_1}\|}\right)}_{=a_1} \\ &\quad \underbrace{+\mu_2|e|\kappa_e\left(\frac{\sin\theta_{L_1}}{\sin\theta_L}\frac{u_{M_4}}{\|\overrightarrow{LM_4}\|}+\frac{\sin\theta_{L_2}}{\sin\theta_L}\frac{u_{M_3}}{\|\overrightarrow{LM_3}\|}\right)}_{=a_2}. \end{aligned} \tag{12}$$

Note that, due to our choice or M_1, M_2, M_3, M_4, (see figures 1 and 2), we have $a_1 \geq 0$ and $a_2 \geq 0$ (provided $u \geq 0$, which is the case, as we will see below). We are going to impose that $a_1\mu_1 = a_2\mu_2$ in order to have two-point fluxes. We thus need to solve the following system:

$$\left\{\begin{array}{lcl} \mu_1+\mu_2 & = & 1, \\ a_1\mu_1 - a_2\mu_2 & = & 0. \end{array}\right.$$

As far as $a_1 + a_2 \neq 0$, this system has a unique solution, which reads

$$\mu_1 = \frac{a_2}{a_1+a_2}, \quad \mu_2 = \frac{a_1}{a_1+a_2}.$$

When $a_1 + a_2 = 0$, in [50, 53], it is proposed to arbitrarily choose $\mu_1 = \mu_2 = \frac{1}{2}$. However, this may cause the flux to be a discontinuous function of u. This is not desirable, so we use slightly different values for a_1 and a_2, namely

$$\tilde{a}_1 = a_1 + \Delta x^2, \quad \tilde{a}_2 = a_2 + \Delta x^2,$$

where Δx is defined by (3). Then we solve the system defining μ_1 and μ_2 with these values

$$\tilde{\mu}_1 = \frac{a_2+\Delta x^2}{a_1+a_2+2\Delta x^2}, \quad \tilde{\mu}_2 = \frac{a_1+\Delta x^2}{a_1+a_2+2\Delta x^2}.$$

Note that, if the coefficient a_i are non-negative, which is the case in the interpolation procedures mentioned below (see Section 3.1), this definition of $\tilde{\mu}_1$ and $\tilde{\mu}_2$ is a continuous function of u (see the proof of Proposition 3.1 below). Finally, we re-define the fluxes as

$$\begin{aligned} F_{K,e} = -F_{L,e} = \tilde{\mu}_1|e|\kappa_e\left(\frac{\sin\theta_{K_1}}{\sin\theta_K}\frac{1}{\|\overrightarrow{KM_2}\|}+\frac{\sin\theta_{K_2}}{\sin\theta_K}\frac{1}{\|\overrightarrow{KM_1}\|}\right)u_K \\ -\tilde{\mu}_2|e|\kappa_e\left(\frac{\sin\theta_{L_1}}{\sin\theta_L}\frac{1}{\|\overrightarrow{LM_4}\|}+\frac{\sin\theta_{L_2}}{\sin\theta_L}\frac{1}{\|\overrightarrow{LM_3}\|}\right)u_L. \end{aligned} \tag{13}$$

Since the definition (12) of the fluxes is consistent with the exact fluxes of order 2 in Δx, so is (13).

Remark 1 *In the definition of $\tilde{a}_1$ and $\tilde{a}_2$, we have chosen to use a global mesh size Δx, but it is also possible to use a local one, for instance*

$$\Delta x_{LM} = \max\left\{|e|, e \in \mathcal{E} \cap (K \cup L)\right\}.$$

This would allow smaller values of the regularizing terms where the mesh is fine, which numerically is relevant.

Since we have imposed $F_{L,e} = -F_{K,e}$, the scheme is conservative (see Proposition 2.4 below). Moreover, if the values u_{M_i} are non-negative, which is the case for the interpolation procedures we use, we have $\tilde{\mu}_1 > 0$ and $\tilde{\mu}_2 > 0$, hence

$$F_{K,e} = -F_{L,e} = A_{K,e}u_K - A_{L,e}u_L,$$

with $A_{K,e} \geq 0$ and $A_{L,e} \geq 0$. This implies that the matrix is the transpose of an M-matrix (see Definition 2.1 below). This in turn implies that the scheme is well-defined (and positive), as we will see below in Proposition 2.3.

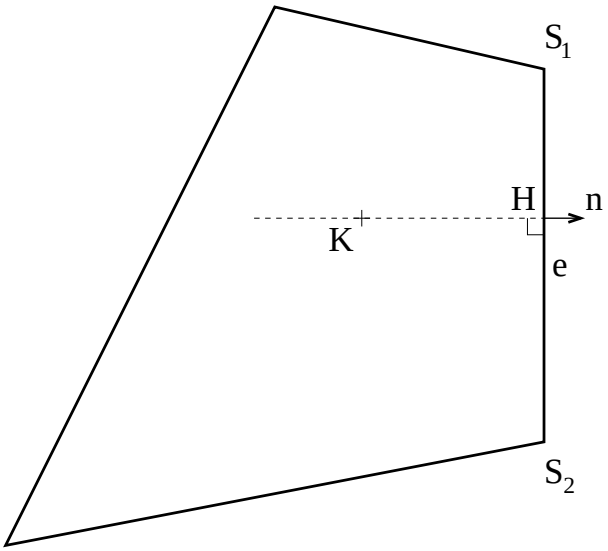

Fig. 3 The boundary cell K and its boundary edge e

2.3 Boundary conditions

The boundary conditions of (1) are taken into account as follows: let K be a boundary cell, with an edge e on the boundary (see figure 3). We denote by S_1 and S_2 the nodes of e, and we impose the value

$$-\gamma_e F_{K,e} + \delta_e u_e = g_e, \tag{14}$$

where g_e is the value of g on e, and similarly for γ_e and δ_e. The value u_e is an approximation of the unknown u on e, yet to be determined, together with the flux $F_{K,e}$ through e. Now, let us define the point H, which is the orthogonal projection of K onto the line (S_1S_2). Then, a consistent approximation of the flux is given by $F_{K,e} = \kappa_e \dfrac{u_K - u_H}{\|\overrightarrow{KH}\|}$, hence,

$$u_H = u_K - \frac{\|\overrightarrow{KH}\|}{\kappa_e} F_{K,e} \tag{15}$$

Moreover, given the values of u at the nodes S_1 and S_2, the following approximation of u_e is of second order:

$$u_e = \frac{1}{2}\left(u_{S_1} + u_{S_2}\right).$$

It is then possible to express the value u_H by an affine approximation:

$$u_e - u_H = \left(\frac{1}{2} + \frac{\overrightarrow{S_2K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2}\right) u_{S_1} + \left(\frac{1}{2} - \frac{\overrightarrow{S_1K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2}\right) u_{S_2}. \tag{16}$$

We now re-write (14) as

$$-\gamma_e F_{K,e} + \delta_e u_H + \delta_e \left(u_e - u_H\right) = g_e.$$

Inserting the value of $u_e - u_H$ given by (16) into this equation, and then using the value of u_H given by (15), we infer

$$-\left(\gamma_e + \delta_e \frac{\|\overrightarrow{KH}\|}{\kappa_e}\right) F_{K,e} + \delta_e u_K = g_e - \delta_e \left[\left(\frac{1}{2} + \frac{\overrightarrow{S_2K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2}\right) u_{S_1} + \left(\frac{1}{2} - \frac{\overrightarrow{S_1K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2}\right) u_{S_2}\right].$$

Hence, we have the following value for the flux:

$$F_{K,e} = \frac{\delta_e \left[\left(\frac{1}{2} + \frac{\overrightarrow{S_2K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2}\right) u_{S_1} + \left(\frac{1}{2} - \frac{\overrightarrow{S_1K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2}\right) u_{S_2}\right] - g_e + \delta_e u_K}{\gamma_e + \delta_e \frac{\|\overrightarrow{KH}\|}{\kappa_e}} \tag{17}$$

The values u_{S_1} and u_{S_2} being computed by the interpolation method described below (Section 3.1), we use this value of the flux in the finite volume formulation.

The advantages of this method (comparing with the one described in [50]) are that it avoids the introduction of additional degrees of freedom on the boundary. Further, in the case of a Cartesian mesh, we recover the approximation by standard finite difference methods.

Note that the node values u_{S_1} and u_{S_2} are necessary to compute $F_{K,e}$. As we will see below, in practice, we use an explicit value for these unknowns: u_{S_i} are the values computed at the preceding step of the fixed point algorithm.

2.4 A summary of the scheme

Let us summarize the scheme we have just derived: a consistent approximation of the flux from cell K to cell L through edge $e = K|L$ is given by

$$F_{K,e} = A_{K,e}u_K - A_{L,e}u_L, \quad \text{and} \quad F_{L,e} = -F_{K,e},$$

with

$$A_{K,e} = \tilde{\mu}_1 |e| \kappa_e \left(\frac{\sin\theta_{K_1}}{\sin\theta_K} \frac{1}{\|\overrightarrow{KM_2}\|} + \frac{\sin\theta_{K_2}}{\sin\theta_K} \frac{1}{\|\overrightarrow{KM_1}\|} \right), \tag{18}$$

$$A_{L,e} = \tilde{\mu}_2 |e| \kappa_e \left(\frac{\sin\theta_{L_1}}{\sin\theta_L} \frac{1}{\|\overrightarrow{LM_4}\|} + \frac{\sin\theta_{L_2}}{\sin\theta_L} \frac{1}{\|\overrightarrow{LM_3}\|} \right). \tag{19}$$

The coefficients $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are given by

$$\tilde{\mu}_1 = \frac{a_2 + \Delta x^2}{a_1 + a_2 + 2\Delta x^2}, \quad \tilde{\mu}_2 = \frac{a_1 + \Delta x^2}{a_1 + a_2 + 2\Delta x^2},$$

with

$$a_1 = |e| \kappa_e \left(\frac{\sin\theta_{K_1}}{\sin\theta_K} \frac{u_{M_2}}{\|\overrightarrow{KM_2}\|} + \frac{\sin\theta_{K_2}}{\sin\theta_K} \frac{u_{M_1}}{\|\overrightarrow{KM_1}\|} \right),$$

$$a_2 = |e| \kappa_e \left(\frac{\sin\theta_{L_1}}{\sin\theta_L} \frac{u_{M_4}}{\|\overrightarrow{LM_4}\|} + \frac{\sin\theta_{L_2}}{\sin\theta_L} \frac{u_{M_3}}{\|\overrightarrow{LM_3}\|} \right).$$

The node values u_{M_i} are given by the interpolation procedure described in Section 3.1 below. The important fact is that it should satisfy

$$u_{M_i} \geq 0.$$

Finally, we use an implicit discretization in time, so that the scheme for system (1)-(2) is given by

$$\begin{cases} \dfrac{1}{\Delta t}\left(u^{n+1} - u^n\right) + M(u^{n+1})u^{n+1} = f + g, \\ u^{n+1} \geq 0, \end{cases} \tag{20}$$

where we have denoted by u^n the vector $(u_K^n)_{K\in\mathcal{K}}$. Here, $\mathcal{K}$ is the set of all cells of the mesh. The vector f is $(f_K)_{K\in\mathcal{K}}$, where $f_K = \int_K f$, so that u_K is an approximation of $\int_K u$. The vector g corresponds to the boundary conditions:

$$g_K = \sum_{e\in K\cap\partial\Omega} g_{K,e},$$

with

$$g_{K,e} = \frac{g_e - \delta_e \left[\left(\frac{1}{2} + \frac{\overrightarrow{S_2K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2} \right) u_{S_1}^{n+1} + \left(\frac{1}{2} - \frac{\overrightarrow{S_1K}\cdot\overrightarrow{S_1S_2}}{\|\overrightarrow{S_1S_2}\|^2} \right) u_{S_2}^{n+1} \right]}{\gamma_e + \delta_e \frac{\|\overrightarrow{KH}\|}{\kappa_e}}$$

The matrix $M(u)$ is defined as follows:

$$[M(u)]_{KK} = \sum_{e\in K} A_{K,e} + \sum_{e\in K\cap\partial\Omega} \left(\frac{\delta_e}{\gamma_e + \delta_e \frac{\|\overrightarrow{KH}\|}{\kappa_e}} \right), \tag{21}$$

$$[M(u)]_{KL} = -A_{L,e} \quad \text{if } K \neq L. \tag{22}$$

Implicitly, we set $A_{L,e} = 0$ if e is not an edge of L. Hence, if K and L do not share an edge, then $[M(u)]_{KL} = 0$.

2.5 Properties of the scheme

First, we prove that the scheme is well-posed, that is, (20) has a solution. For this purpose, we first need some definitions:

Definition 2.1 We say that a matrix $M \in \mathbb{R}^{N\times N}$ is an M-matrix if it satisfies the following inequalities:

$$\forall i \neq j, \quad m_{ij} \leq 0, \tag{23}$$

$$\forall i, \quad \sum_{j=1}^{N} m_{ij} \geq 0. \tag{24}$$

Moreover, if (24) is strict for all i, we say that M is a strict M-matrix.

Remark 2 *Definition 2.1 is not the standard definition of an M-matrix. Indeed, in the literature, an M-matrix is a matrix M such that $M = c\mathbf{Id} - A$, where A has non negative entries, and $c \geq \rho(A)$, where $\rho(A)$ is the spectral radius of A (see [44] for instance). It is an easy exercise to prove that Definition 2.1 is a special case of this latter definition.*

We have the following standard fact:

Lemma 1 *Assume that M is a strict M-matrix. Then, M satisfies the following properties:*

$$\text{if } X \in \mathbb{R}^N \text{ is such that } \forall i, (MX)_i \geq 0, \text{ then } \forall i, \quad X_i \geq 0. \tag{25}$$

$$\text{if } X \in \mathbb{R}^N \text{ is such that } \forall i, (MX)_i > 0, \text{ then } \forall i, \quad X_i > 0. \tag{26}$$

These properties are equivalent to the fact that M is invertible, with M^{-1} having non-negative coefficients.

The proof of this result can be found in many linear algebra textbooks, as for instance [44, 47].

Definition 2.2 In the sequel, for any $X \in \mathbb{R}^N$, we will denote by $X \geq 0$ the fact that all the components of X are non-negative:

$$X \geq 0 \iff \forall 1 \leq i \leq N, \quad X_i \geq 0.$$

Similarly, $X > 0$ means that all components of X are positive:

$$X > 0 \iff \forall 1 \leq i \leq N, \quad X_i > 0.$$

We are now in position to prove that the scheme is well-posed:

Proposition 2.3 *If $f + g \geq 0$ and $u^n \geq 0$, then system* (20) *has a solution u^{n+1}.*

Recall that, according to Definition 2.2, $u \geq 0$ means that each component of the vector u are non-negative.

Proof: we reproduce here the proof given in [14]. Equation (20) may be written $\phi(u^{n+1}) = u^{n+1}$, where

$$\phi(u) = \left(\mathbf{Id} + \Delta t M(u)\right)^{-1} \left(u^n + \Delta t(f + g)\right). \tag{27}$$

Hence, we need to prove that ϕ has a fixed point $u \geq 0$ in order to prove that (20) has a solution.

To this end, we first note that, in view of (21) and (22), $M(u^{n+1})^T$ is always an M-matrix. Hence, $\left(\mathbf{Id} + \Delta t M(u^{n+1})\right)^T$ is a strict M-matrix. Applying Lemma 1, we infer that the matrix $\left(\mathbf{Id} + \Delta t M(u^{n+1})\right)^{-T}$ has positive coefficients. Hence, since $u^n + \Delta t(f + g) \geq 0$, we have

$$\forall u \in \mathbb{R}^N, \quad \phi(u) \geq 0.$$

Moreover, multiplying (27) by the constant vector $(1, \ldots, 1)$ on the left, we find that

$$\sum_{K\in\mathcal{K}} \left[\phi(u)\right]_K = \sum_{K\in\mathcal{K}} \left(u_K^n + f_K + g_K\right) := C_0,$$

which is a constant independent of u. We thus define the set

$$\mathcal{C} = \left\{ u \in \mathbb{R}^N, \quad u \geq 0, \quad \sum_{K\in\mathcal{K}} u_K = C_0 \right\}. \tag{28}$$

It is a convex compact subset of $\mathbb{R}^N$, and the application ϕ maps $\mathcal{C}$ into itself. Moreover, ϕ is continuous. Indeed, each coefficient of $M(u)$ is a continuous function of u, and $M \mapsto (I + M)^{-1}$ is a continuous application from the set of M-matrices to the set of matrices. Hence, we may apply Brouwer's theorem [46], which implies that ϕ has a fixed point in $\mathcal{C}$, hence (20) has a solution. □

Remark 3 *Here again, the positivity property is a consequence of the structure of the matrix M, and does not depend on the fact that we use $M = M\left(u^{n+1}\right)$ in the scheme* (20)*. In particular, as far as $u^n \geq 0$, we still have positivity for the scheme* (30)*. However, only the exact solution to the linear system is non-negative. In particular, if the linear system is solved with a poor precision, it may still exhibit negative values.*

Proposition 2.4 *The scheme defined by* (20) *is conservative.*

Proof: the only thing we need to check is that if $f = 0$, $\delta = 0$ and $g = 0$, then the scheme satisfies a discrete version of the equality $\frac{d}{dt}\int_\Omega u = 0$, that is,

$$\forall n \geq 0, \quad \sum_{K\in\mathcal{K}} u_K^{n+1} = \sum_{K\in\mathcal{K}} u_K^n. \tag{29}$$

For this purpose, we write (20):

$$\forall K \in \mathcal{K}, \quad \frac{1}{\Delta t}\left(u_K^{n+1} - u_K^n\right) + \sum_{L\in\mathcal{K}} \left[M\left(u^{n+1}\right)\right]_{KL} u_L^{n+1} = 0,$$

and we sum up with respect to K. This gives

$$\sum_{K\in\mathcal{K}} u_K^{n+1} - \sum_{K\in\mathcal{K}} u_K^n + \Delta t \sum_{K\in\mathcal{K}}\sum_{L\in\mathcal{K}} \left[M\left(u^{n+1}\right)\right]_{KL} u_L^{n+1} = 0.$$

Using formulas (21) and (22), we find

$$\sum_{K\in\mathcal{K}} M_{KL} = M_{LL} + \sum_{K\neq L} M_{KL} = \sum_{e\in L} A_{L,e} - \sum_{e\in L} A_{L,e} = 0.$$

□

Remark 4 *Here again, the above property is a consequence of the structure of the matrix M. In particular, M need not be exactly equal to $M(u^{n+1})$ in order to have this property. In other words, we do not need to reach convergence of the fixed point strategy in order to have conservativity. In particular, if we replace* (20) *by*

$$\frac{1}{\Delta t}\left(u^{n+1} - u^n\right) + M(u^n)u^{n+1} = f + g, \tag{30}$$

the scheme is still conservative.

3 Implementation

3.1 Interpolation

Here we explain how we compute the node unknowns as functions of the cell unknowns, in order to compute the matrix $M(u)$ of the scheme. For this purpose, a natural way to proceed is the one proposed in [48] or [50]: take a node P of the mesh, and consider the neighbouring cells $(K_i)_{1\leq i\leq p}$, that is, all the cells having P as a node. Then, write down a linear approximation of u_P as a function of the values u_{K_i}:

$$u_P = \sum_{i=1}^{P} \omega_i u_{K_i}. \tag{31}$$

In order to have a second-order approximation, so that the scheme may still be second-order convergent, we require formula (31) to be exact for any affine function. This gives the following system:

$$\left\{\begin{array}{rcl} \displaystyle\sum_{i=1}^{p} \omega_i & = & 1, \\ \displaystyle\sum_{i=1}^{p} \omega_i x_{K_i} & = & x_P, \\ \displaystyle\sum_{i=1}^{p} \omega_i y_{K_i} & = & y_P. \end{array}\right. \tag{32}$$

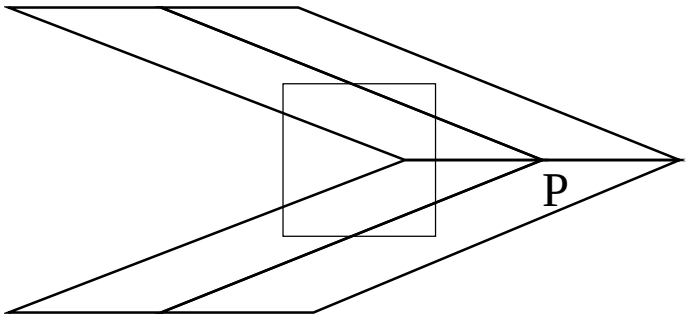

Fig. 4 An example in which the point P is not in the convex envelop of the neighbouring cell centers.

In this system, the unknowns are the weights ω_i, and for each point M, we denote its components by (x_M, y_M). Hence, we have three equations and p unknowns. In general, $p \neq 3$. We solve the system using a least squares procedure [4]. In general, $p \geq 3$ (for a structured mesh of quadrangles, $p = 4$ when P is an interior node), and we thus find a solution to this system.

However, as we have seen, it is important that $u_P \geq 0$ in order for the scheme to be well-defined (see Proposition 2.3). A way to ensure this is to have only non-negative weights in the interpolation procedure. This is always the case if P is in the convex envelop of the points K_i. When the mesh is highly deformed, this may not be the case (see Figure 4). It is possible to circumvent this difficulty by extending the interpolation to non-neighbouring cells. However, by doing so, we extend the stencil of the scheme, which we want to avoid do to parallelization concerns. At this point, we see two main possible strategies:

- first compute the orthogonal projection Q of P onto the convex hull of the points K_i. Then use this projection Q in system (32) instead of P itself. Doing so, you are sure that $\omega_i \geq 0$ for all i, so that $u_P \geq 0$. However, the projection may result in poor accuracy, since Q might be rather far from P.

- Another possibility is to first solve system (32), and compute u_P, but then truncate the result:

$$u_P = \max\left(0, \sum_{i=1}^{P} \omega_i u_{K_i}\right).$$

  Doing so, we are sure that $u_P \geq 0$, and in principle we do not affect the accuracy, since we know, *a posteriori*, that $u_P \geq 0$. Moreover, contrary to what happens when truncating cell values in the same way, the scheme is still conservative. Indeed, this is a property of the matrix which is not modified by the present truncation.

This last method is the one we have retained, since it seems to us the best compromise between robustness and precision, given the fact that we do not want to enlarge the stencil of the scheme as for the strategy proposed in [50].

3.2 Fixed-point iteration

As we have seen above, the system is nonlinear (the matrix M in (20) depends on u^{n+1}). Hence in order to solve (20), one needs to use a fixed point strategy. For instance, the following algorithm is the most simple and most natural one:

$$\left\{ \begin{array}{l} v^0 = u^n, \\ \forall k \geq 0, \quad \frac{1}{\Delta t}\left(v^{k+1} - u^n\right) + M(v^k)v^{k+1} = f + g, \end{array} \right. \tag{33}$$

with a stopping criterion of the form $\|v^{k+1} - v^k\| \leq \varepsilon \|v^k\|$. The value of u^{n+1} is then given by v^{k+1}.

We have the following convergence result:

Proposition 3.1 *Assume that $f + g \geq 0$ and $u^n > 0$. Then, if Δt is small enough, the sequence $(v^k)_{k\in\mathbb{N}}$ defined by* (33) *converges to u^{n+1}, which is a solution to* (20)*.*

Proof: Here again, we use the mapping ϕ defined in the proof of Proposition 2.3 :

$$\begin{array}{rccl} \phi : & \mathcal{C} & \longrightarrow & \mathcal{C} \\ & u & \longmapsto & \left(\mathbf{Id} + \Delta t M(u)\right)^{-1}\left(u^n + \Delta t(f+g)\right). \end{array}$$

The set $\mathcal{C}$ is defined by (28). We are going to prove that this map is a contraction, so that the iteration $v^{k+1} = \phi(v^k)$ converges to a fixed point of ϕ.

In order to do so, we first prove the following fact: if Δt is small enough, then we have

$$\forall u \in \mathcal{C}, \quad \forall v \in \mathcal{C}, \quad \forall x \in \mathbb{R}^N, \quad \left\| \left[\mathbf{Id} + \frac{\Delta t}{2} \left(M(u) + M(v) \right) \right] x \right\| \geq \frac{1}{2} \|x\|. \tag{34}$$

To prove our claim, we note that, in view of (18) and (19), we have

$$|A_{K,e}| \leq 2 \frac{|e| \kappa_e}{\min\left(\|\overrightarrow{KM_1}\|, \|\overrightarrow{KM_2}\| \right)},$$

where M_1 and M_2 are the endpoints of e. This upper bound, in particular, does not depend on u. Inserting this into (21) and (22), we have

$$|[M(u)]_{KK}| \leq \sum_{e \in K} \left(2 \frac{|e| \kappa_e}{\min\left(\|\overrightarrow{KM_1}\|, \|\overrightarrow{KM_2}\| \right)} \right) + \sum_{e \in K \cap \partial\Omega} \frac{\delta_e}{\gamma_e},$$

and

$$|[M(u)]_{KL}| \leq 2 \frac{|e| \kappa_e}{\min\left(\|\overrightarrow{LM_1}\|, \|\overrightarrow{LM_2}\| \right)},$$

All these bounds are independent of u, hence, we have proved that there is a constant C_1, which depends only on the mesh, and on δ, γ and κ, such that $\|M(u)\| \leq C_1$. Hence, if $2C_1 \Delta t \leq 1$, we have

$$\left\| \left[\mathbf{Id} + \frac{\Delta t}{2} \left(M(u) + M(v) \right) \right] x \right\| \geq \|x\| - \frac{\Delta t}{2} \|(M(u) + M(v))x\| \geq \|x\| - \frac{\Delta t}{2} 2C_1 \|x\| \geq \frac{1}{2} \|x\|,$$

which proves (34).

Next, we prove the following fact: there exists a constant C_2 depending only on the mesh and on δ, γ and κ, such that,

$$\forall u \in \mathcal{C}, \quad \forall v \in \mathcal{C}, \quad \|M(u) - M(v)\| \leq \frac{C_2}{\Delta x^2} \|u - v\|. \tag{35}$$

It is clear from the above argument (and from the formulas (21) and (22)), that it is sufficient to prove this inequality for $A_{K,e}$. Now, $A_{K,e}$ depends on u only through $\tilde{\mu}_1$ and is a linear function of $\tilde{\mu}_1$, so it is sufficient to prove (35) for μ_1. We thus write :

$$\tilde{\mu}_1(u) = \psi(a_1(u), a_2(u)), \quad \text{where} \quad \psi(\alpha, \beta) := \frac{\beta + \Delta x^2}{\alpha + \beta + 2\Delta x^2}.$$

The function $(\alpha, \beta) \mapsto \psi(\alpha, \beta)$ is differentiable in the set $\{\alpha \geq 0, \beta \geq 0\}$, and

$$\frac{\partial \psi}{\partial \alpha} = -\frac{\beta + \Delta x^2}{\left(\alpha + \beta + 2\Delta x^2 \right)^2}, \quad \frac{\partial \psi}{\partial \beta} = \frac{\alpha}{\left(\alpha + \beta + 2\Delta x^2 \right)^2}.$$

Hence, $|\nabla \psi| \leq \left(2\Delta x^2 \right)^{-1}$, thus

$$|\tilde{\mu}_1(u) - \tilde{\mu}_1(v)| = |\psi(a_1(u), a_1(u)) - \psi(a_1(v), a_2(v))| \leq \frac{1}{2\Delta x^2} \left(|a_1(u) - a_1(v)| + |a_2(u) - a_2(v)| \right).$$

Finally, in view of the definition of a_1 and a_2, we have $|a_i(u) - a_i(v)| \leq C_2 \|u - v\|$, where C_2 is a constant depending only on κ and on the mesh. Hence, we infer

$$|\tilde{\mu}_1(u) - \tilde{\mu}_1(v)| \leq \frac{C_2}{2\Delta x^2} \|u - v\|.$$

This proves (35) for $\tilde{\mu}_1$, hence for M.

We are now in position to prove that ϕ is a contraction: indeed, we have, from the definition of ϕ,

$$\left[\mathbf{Id} + \frac{\Delta t}{2} \left(M(u) + M(v) \right) \right] \left(\phi(u) - \phi(v) \right) = \frac{1}{2} \Delta t \left(M(v) - M(u) \right) \left(\phi(u) + \phi(v) \right).$$

Applying (34) on the one hand, and (35) on the other hand, we have

$$\frac{1}{2}\|\phi(u)-\phi(v)\| \leq C_2 \frac{\Delta t}{\Delta x^2}\|u-v\|\,\|\phi(u)+\phi(v)\|. \tag{36}$$

Due to the definition of $\mathcal{C}$, we know that $\|\phi(u)\|$ and $\|\phi(v)\|$ are bounded independently of u and v, respectively. Hence, choosing Δt even smaller if necessary, we have proved that ϕ is a contraction on $\mathcal{C}$. This proves the result. □

Remark 5 *This result is of limited practical interest for several reasons.*
First, a careful examination of the constants in the proof indicate that, at best, the condition would read

$$\Delta t \leq \min\left(\frac{1}{2C_1}, C_2\Delta x^2\right), \tag{37}$$

for some C_1 and C_2 that in principle still depend on the mesh. Hence, we end up with a condition which is the stability condition for an explicit scheme, loosing the benefit of an implicit scheme.

Second, the constant C_2 is not explicit and could in principle be large for highly deformed meshes.

However, it indicates that, if convergence of the fixed point should fail, decreasing the time step would help. Actually, it should be noted that in the numerical tests we have done, we did not find any situation in which the sequence defined by (33) *does not converge (even with rather large values of Δt). An alternative strategy would be to use acceleration techniques, such as the one presented in [34].*

Remark 6 *It is also possible to prove a similar result when assuming that $u \geq m > 0$. In such a case, roughly speaking, Δx^2 in* (36) *would be replaced by $m+\Delta x^2$, therefore making condition* (37) *nicer. However, in such a case, it is necessary that the interpolation procedure giving the node unknowns satisfies the following property:*

$$\textit{if} \quad \forall K \in \mathcal{K}, \quad u_K \geq m, \quad \textit{then} \quad \forall N \in \mathcal{N}, \quad u_N \geq m. \tag{38}$$

This property is not satisfied by the interpolation method using truncation (see section 3.1).

If the problem we need to solve is really (1)-(2), then the above fixed-point strategy is necessary to obtain good results. However, the problems we deal with in practice are nonlinear. Hence, problem (1)-(2) is only an intermediate step in solving the master equation using, for instance, a Newton algorithm. When this algorithm reaches convergence, u^n is close to u^{n+1}, so the convergence of the fixed point algorithm may not be needed. In numerical simulations, we have compared two strategies: the first one is to reach convergence in the fixed point algorithm. The second one consists in dropping the fixed points strategy, that is, use only one iteration. It happens that, in most cases, it is less costly to use the second strategy. This amounts to using the scheme (30). The precision of the two methods is comparable. Here, Remarks 4 and 3 are of particular interest: this strategy does not change the main properties of the scheme.

4 Numerical tests

4.1 Steady analytic problem

We first assess the scheme on a steady problem with an analytically solution. Since the positivity of the solution is a requirement for our scheme, we choose a solution positive everywhere:

$$u(x,y) = 1 + \cos(\pi x)\cos(\pi y) \tag{39}$$

This solution is obtained in solving the following equation in $\Omega = (0,1)^2$

$$\begin{cases} -\operatorname{div}(\nabla u) = 2\pi^2\cos(\pi x)\cos(\pi y), & \text{in } \Omega, \\ (\nabla u)\cdot n = 0 & \text{on } \partial\Omega, \end{cases} \tag{40}$$

with the constraint $\int_\Omega u = 1$. We perform this problem on several meshes, which are classical for diffusion operator. These meshes are obtained by perturbing an initial uniform mesh made of n^2 squares of size $1/n^2$:

- Random mesh – all vertex are translated by the vector $0.4(r_x/n, r_y/n)$, where r_x and r_y are two random numbers uniformly distributed on $[0,1]$. This mesh includes non-convex zones, see left part of fig. 5.

- Kershaw mesh [28] – mesh is highly skewed, see middle part of fig. 5.

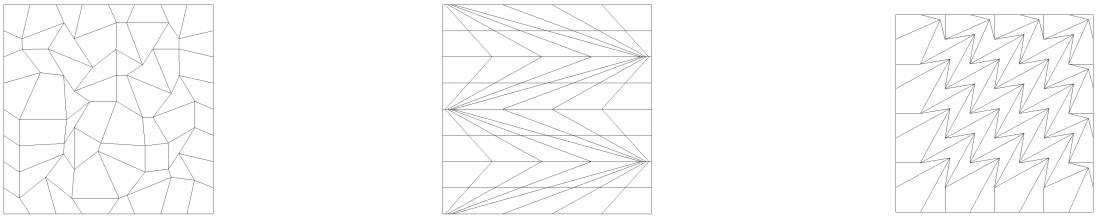

Fig. 5 8^2 zones meshes: random (left), Kershaw (middle), non-convex (right).

- Strongly non-convex mesh, constructed by moving the central vertex of every the four-cell patterns of the Cartesian mesh of a vector $(0.8\sqrt{2}/n, 0.8\sqrt{2}/n)$, see right part of fig. 5.

We perform a convergence study on this problem and for these meshes, starting with the 8^2 mesh to a 512^2 mesh. We also compute the ratio

$$c_r = \frac{\sum_{e\in\mathcal{E}} F_{K,e}\left(u_K - u_L\right)}{\sum_{e\in\mathcal{E}} \frac{\kappa_e}{d_e}|e|\left(u_K - u_L\right)^2}.$$

to measure numerically the coercivity of the scheme. Results are summarized in Table 1.

	random			Kershaw			non-convex		
n	L_2 error	order	c_r	L_2 error	order	c_r	L_2 error	order	c_r
8	1.07×10^{-2}		0.88	7.31×10^{-2}		0.36	3.90×10^{-2}		0.18
16	2.22×10^{-3}	2.27	0.81	4.12×10^{-2}	0.83	0.26	1.03×10^{-2}	1.92	0.15
32	5.18×10^{-4}	2.10	0.80	1.91×10^{-2}	1.11	0.22	2.61×10^{-3}	1.98	0.13
64	1.11×10^{-4}	2.22	0.79	7.35×10^{-3}	1.38	0.19	7.14×10^{-4}	1.87	0.12
128	2.69×10^{-5}	2.04	0.79	2.27×10^{-3}	1.70	0.18	1.99×10^{-4}	1.84	0.12
256	9.57×10^{-6}	1.49	0.79	6.18×10^{-4}	1.88	0.18	5.82×10^{-5}	1.77	0.12
512	2.69×10^{-6}	1.83	0.79	1.60×10^{-4}	1.96	0.18	1.84×10^{-5}	1.66	0.12

Table 1 Convergence table for the analytic problem

The scheme achieves almost second order accuracy even on these highly deformed meshes. In the case of non-convex mesh (last column), it seems that the interpolation procedure we use degrades the order. The coefficient c_r, which accounts for the coercivity, converges for all the grid patterns.

4.2 Planar nonlinear heat wave in a cold wall

A well-known analytic solution of the system (1) is referred in the literature as Marshak waves. This self-similar solution has been at first exhibited by Marshak [40]. It consists in considering a diffusion coefficient $\kappa = \kappa_0 u^k$, where κ_0 is a constant and $k \in \mathbb{N}^+$, a boundary condition $u(0,t) = u_0$ and a initial condition $u(x,t) = 0, x > 0$. This generates a self-similar wave propagating into the half space $x > 0$.

Introducing dimensionless variables

$$\mu = \mu(\xi) = \frac{u}{u_0}, \quad \xi = \left(\frac{k+1}{2\kappa_0 u_0^k}\right)^{1/2} \frac{x}{\sqrt{u}}, \tag{41}$$

the first equation of system (1) reduces to an ordinary differential equation

$$\frac{d^2\left(\mu^{k+1}\right)}{d\xi^2} + \xi\frac{d\mu}{d\xi} = 0. \tag{42}$$

The boundary conditions associated with equation (42) are $\mu(0) = 1$ and $\mu = \mu^k(d\mu/d\xi) = 0$ for $\xi = \xi_0$. The position ξ_0 is the solution of the boundary value problem.

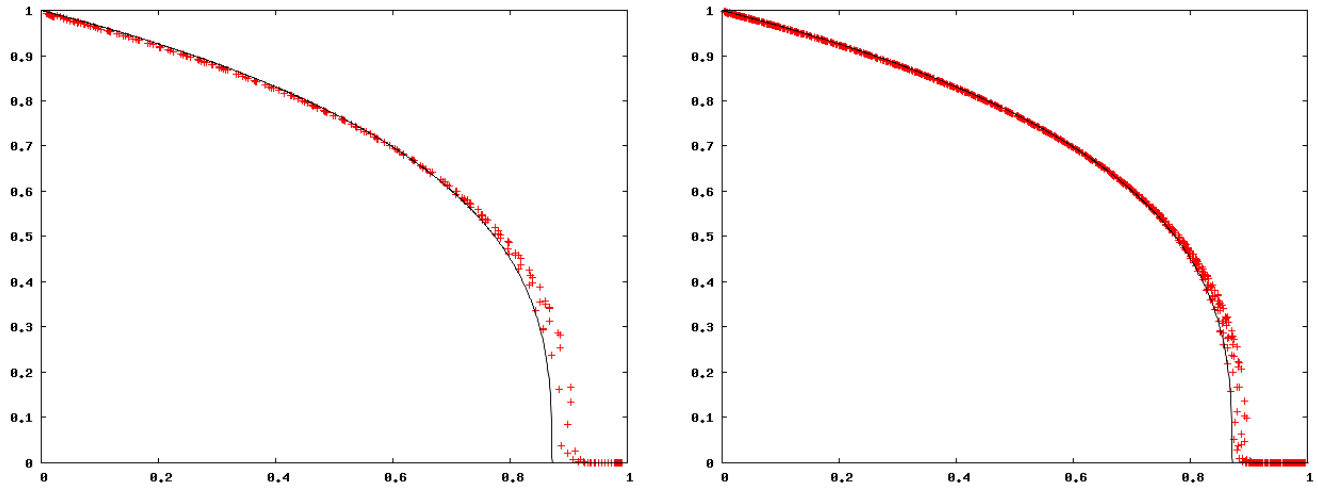

Fig. 6 Plain line: analytic solution, symbols: numerical solution. Left 32^2 Kershaw mesh, right 64^2 Kershaw mesh.

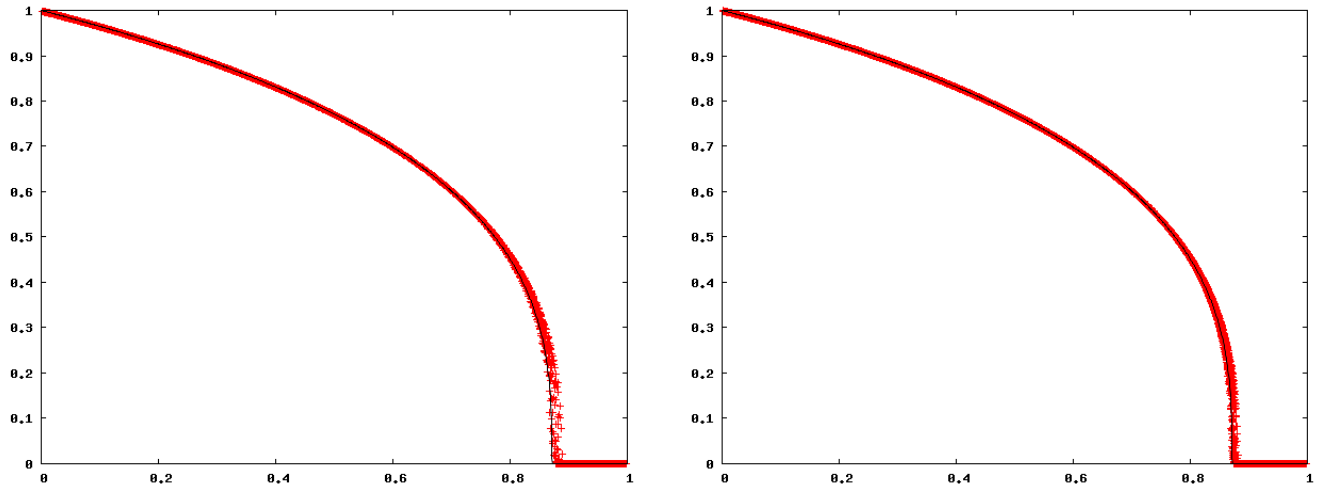

Fig. 7 Plain line: analytic solution, symbols: numerical solution. Left 128^2 Kershaw mesh, right 256^2 Kershaw mesh.

For our application, we choose $k = 3, \kappa_0 = 1, u_0 = 1$ and run the code until $t = 1$. We perform this calculation on 4 Kershaw meshes. Domain size is 1×1, and the number of zones goes from 32^2 zones for the coarsest mesh to 256^2 zones for the finest one. Numerical results are displayed on figures 6 and 7. On these figures, we have plotted the values of the temperature for all the zones of the mesh, as a function of x. The oscillatory aspect of the plot is due to the fact that the mesh is not 1D. We illustrate this point on figure 8 for the coarsest grid. We see on this plot that the solution has no oscillation.

We observe convergence to the analytic solution. Moreover, the solution remains positive all along the simulation. To have an idea of the convergence rate, we compute the L^∞ norm of the function u^4, for which the spatial derivative is bounded. We obtain a convergence rate of ≈ 1.2.

4.3 Radiative shock

Our last test problem is more challenging. The goal is to assess the robustness and accuracy of the method, in a configuration which is relevant for our applications. In theses cases, radiation is strongly coupled with the Euler equations for hydrodynamics. Since we use Arbitrary Lagrangian-Eulerian methods to discretize the Euler equations, the mesh moves at each time-step of the calculation, and the distortion of the zones is induced by the fluid motion.

In the following, we consider the following set of equations corresponding to the Euler equation for the hydrodynamics coupled with an equilibrium diffusion model for the radiation. We solve the equations in the Lagrangian frame:

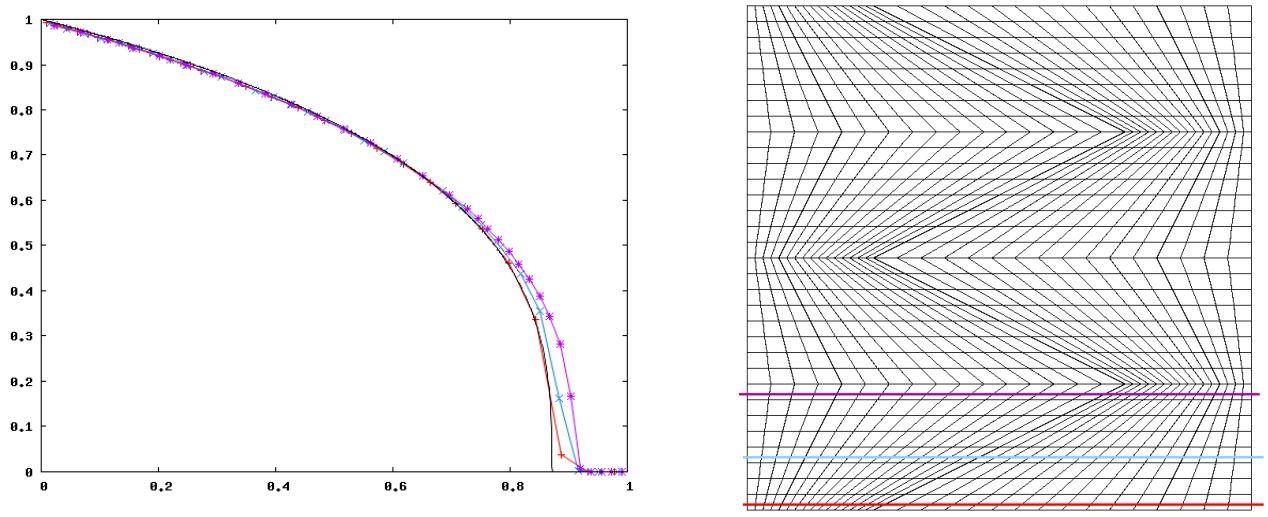

Fig. 8 Black line: analytic solution, color lines: numerical solutions on the 32^2 Kershaw mesh. Each color corresponds to the colored line on the Kershaw mesh on the right part of the figure, where the numerical solution is measured.

$$\left\{\begin{array}{rcl} \rho d_t \tau &=& \operatorname{div} \mathbf{u}, \\ \rho d_t \mathbf{u} &=& -\nabla(p+p_r), \\ \rho d_t(e+e_r) &=& -\operatorname{div}\left((p+p_r)\mathbf{u}\right) + \operatorname{div}\left(\frac{c}{3\sigma_R}\nabla e_r\right), \\ \rho d_t(\varepsilon+e_r) &=& -\operatorname{div}\left(p_r\mathbf{u}\right) + \operatorname{div}\left(\frac{c}{3\sigma_R}\nabla e_r\right), \end{array}\right. \tag{43}$$

where $d_t\varphi \equiv (\partial_t + \mathbf{u}\cdot\nabla)\,\varphi$ is the Lagrangian derivative, ρ the density, $\tau = 1/\rho$ the specific volume, $\mathbf{u}$ the fluid velocity, p the material pressure, $p_r = 4/3aT^4$ the radiative pressure, T the temperature, $e = \varepsilon + 1/2\mathbf{u}^2$ the total material energy, $e_r = aT^4$ the radiative energy, c the speed of light, $\sigma_R(T)$ the Rosseland opacity and $\varepsilon(\rho, T)$ the internal energy.

To solve system (43), we split it into a hydrodynamic part and a radiative part. The hydrodynamic part is solved using the Godunov-type Lagrangian scheme Glace [8, 11]. The radiative part consists in solving the following equation:

$$\rho\partial_t(\varepsilon + e_r) = \operatorname{div}\left(\frac{c}{3\sigma_R}\nabla e_r\right). \tag{44}$$

Note that this equation is nonlinear even if σ_R does not depend on T.

We use a Newton procedure to solve the equation (44). As for the Marshak wave, we use the procedure described in section 3.2 in order to converge at the same time the non-linearity of the system and of the scheme.

In order to properly assess the capability of our code to compute such flows, we compare our results to the semi-analytic solution calculated by Lowrie and Rauenzahn [39]. We refer to this work for all details concerning the semi-analytic solution. We choose the configuration so that the shock is super-critical. The structure of the solution consists in a radiation precursor and then an isothermal, hydrodynamic shock.

We use the following values ahead of the shock: $\rho_0 = 1$, $\mathbf{u}_0 = (1.58551e8, 0., 0.)^T$, $T_0 = 1.53783e6$, $\sigma = 485.8$. Moreover, we use perfect gas equation of state, with $\gamma = 5/3$. This gives the following values for the main parameters of the simulation

$$\mathcal{P}_0 = \frac{aT_0^4}{\rho a_0^2} = 10^{-4}, \mathcal{M}_0 = \frac{|\mathbf{u}_0| - s_{shock}}{a_0} = 10 \text{ and } \mathcal{K} = \frac{aT_0^4 c}{3\sigma\rho_0 a_0^3} = 10^{-4},$$

with $a_0 = \sqrt{\gamma p_0/\rho_0}$ the sound speed ahead of the shock and s_{shock} the shock speed. Lowrie and Rauenzahn show that these three numbers are sufficient to determine entirely the self-similar solution of the problem. $\mathcal{M}_0$ is the shock Mach number, $\mathcal{K}$ controls the amount of thermal diffusion and consequently the extent of the radiation precursor in front of the

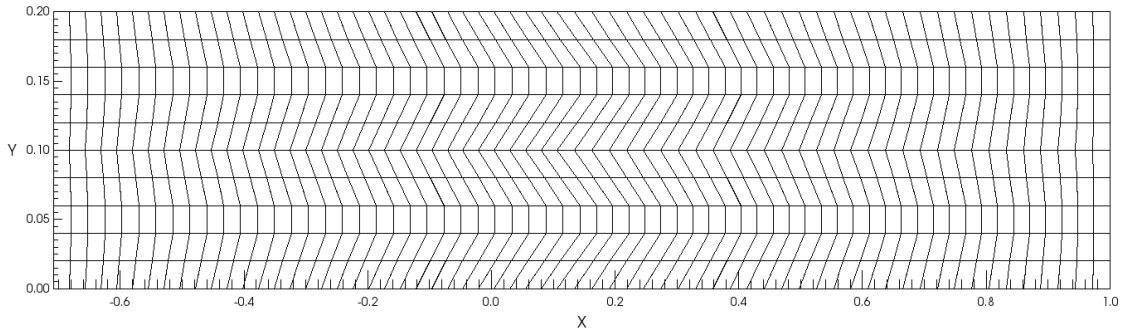

Fig. 9 Initial coarsest mesh for the radiative shock problem.

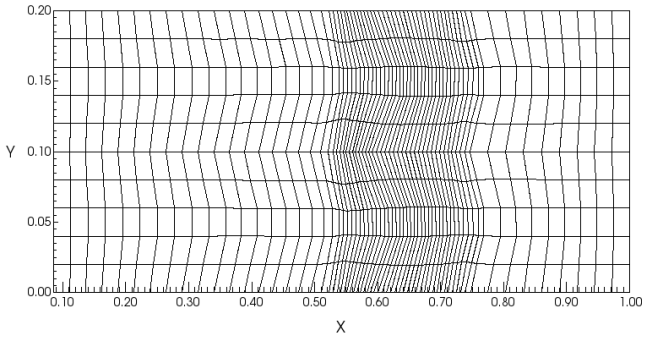

Fig. 10 Final coarsest mesh for the radiative shock problem.

shock and $\mathcal{P}_0$ measures the influence of radiation on the flow dynamics.

For the 1D problem, we use the following initial conditions: $x \in [-0.707245, 1]$, and exact solution for ρ, $\mathbf{u}$ and T for a shock position $x_{shock} = 0.764237$. Boundary conditions are $\mathbf{u} = \mathbf{u}_0$ and zero net flux in $x = -0.707245$ and $\mathbf{u} = \mathbf{0}$ and zero net flux in $x = 1$. Then we run the calculation until $t_f = 5 \times 10^{-9}$.

To assess the scheme for this kind of problem, we artificially extend the calculation in the y-direction: $y \in [0, 0.1]$, and perform a convergence study. Number of zones goes from 64×10 to 512×80 and we apply a Kershaw like pattern on the initial mesh. Resulting coarsest mesh (64×10) is depicted in figure 9.

During the calculation, the Lagrangian motion of the mesh deforms the zones giving rise of a more perturbed mesh (see figure 10).

We display on figures 11 to 14 the results of the simulations on the different meshes in term of velocity and temperature, compared to the analytic solution. All the values of the different layers are displayed on these figures.

It shows the convergence of the numerical simulation to the semi-analytic solution, assessing the global hydro-radiative scheme to capture such radiative shock configuration. Moreover, we don't observe any undershoots on the temperature profiles (undershoots on velocity are due to wall-heating effect [42]). On these meshes, the numerical cost of the non-linearity of the scheme is weak, because the non-linearity of the problem is as hard to solve than the non-linearity of the scheme.

5 Conclusion

We have presented in this article a finite-volume scheme for diffusion equation which has a two-point stencil, which is consistent, and is positive, even on deformed meshes. Therefore, assuming the solution is positive, we propose a method to ensure the positivity of the scheme, while keeping a very compact stencil. Consequently, the scheme is easy to use on a parallel computing architecture.

We have conducted numerical tests which assess convergence on highly deformed mesh for smooth solutions. If the deformation is not too strong, second-order accuracy is achieved. Further, we have observed convergence for nonlinear

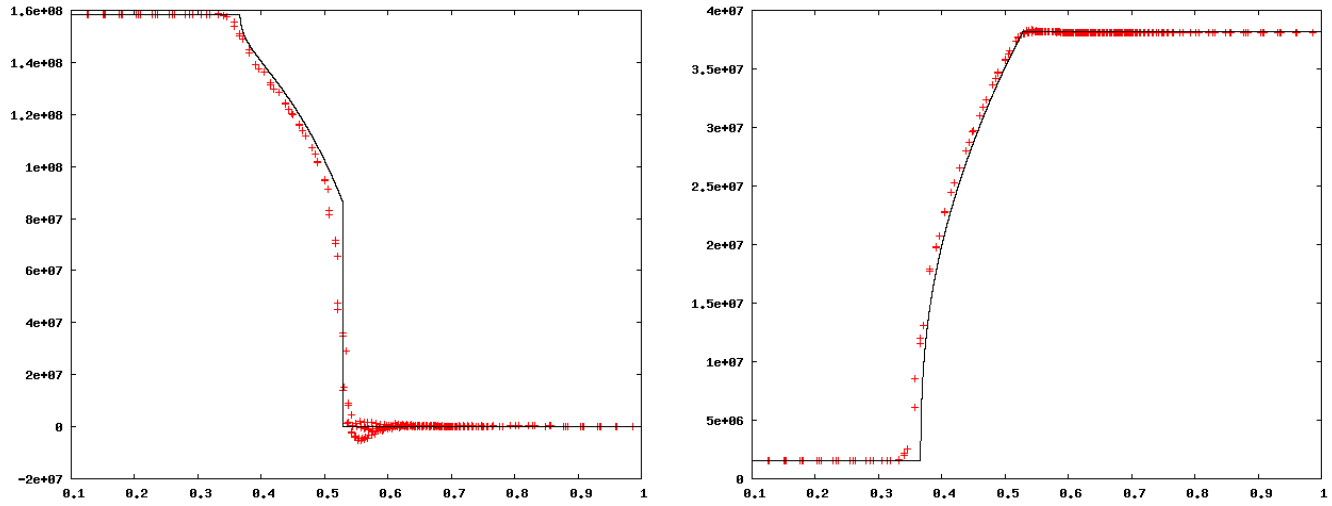

Fig. 11 1D-plot of the numerical solution (symbols) compared to analytic (plain line) for the 64×10 zones mesh. Left: velocity, right: temperature.

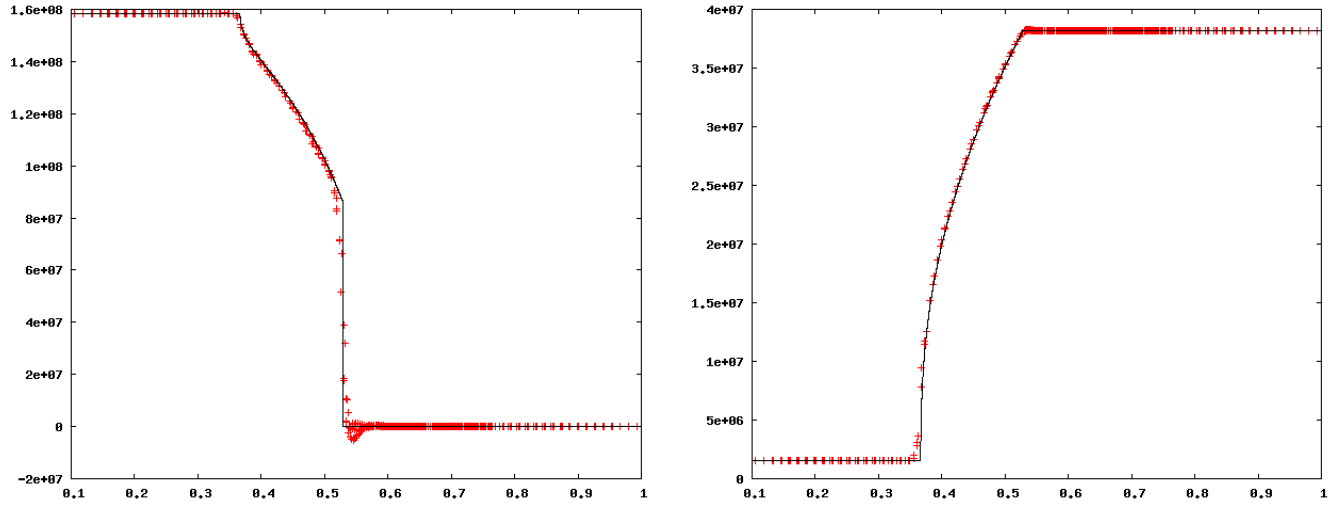

Fig. 12 1D-plot of the numerical solution (symbols) compared to analytic (plain line) for the 128×20 zones mesh. Left: velocity, right: temperature.

Marshak waves on the one hand, and for hydro-radiative shock solutions on the other hand. In the first case, analytic solutions are available, while in the second case, semi-analytic solutions are used for comparison.

The main drawback is that the scheme is non-linear. Therefore, a fixed-point loop is in principle necessary for computing the solution. We prove that if Δt is small enough, this algorithm converges.

The fact that the scheme is non-linear indicates that it is not efficient for solving simple linear problems. However, in the case of nonlinear problems, the outer loop (in our case a Newton algorithm) allows for dropping the fixed-point strategy. Indeed, at convergence of the Newton algorithm, two successive values of u, namely u^n and u^{n+1}, are very close, so only one iteration of the fixed-point strategy is necessary. Hence, for non-linear problems, the cost of the present scheme remains comparable to that of linear schemes.

Acknowledgments

The authors address many thanks to G. Samba and F. Hermeline for useful advises and fruitful discussions about this work.

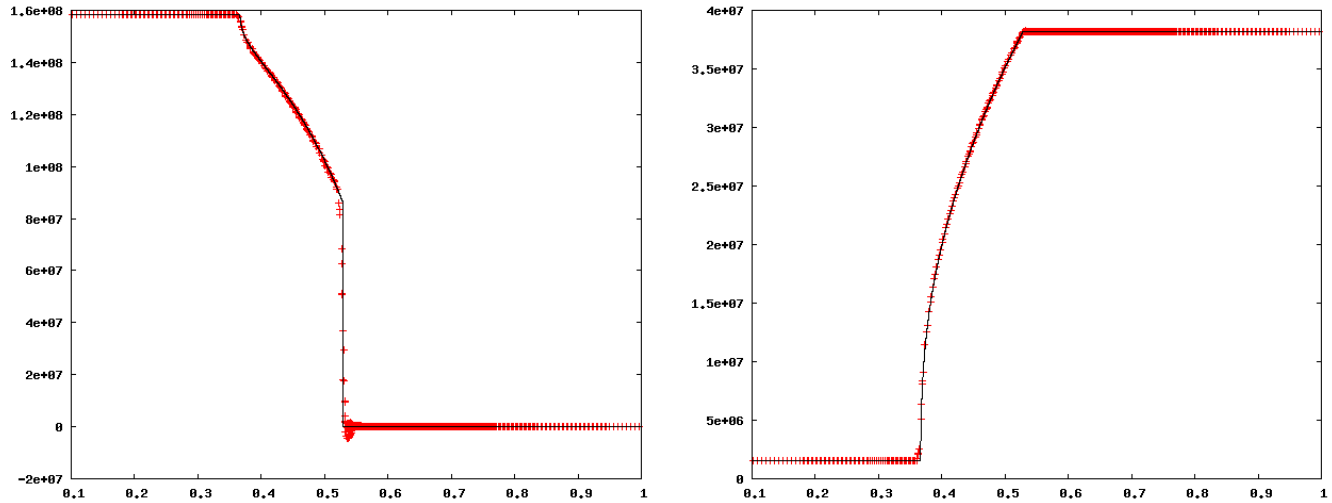

Fig. 13 1D-plot of the numerical solution (symbols) compared to analytic (plain line) for the 256×40 zones mesh. Left: velocity, right: temperature.

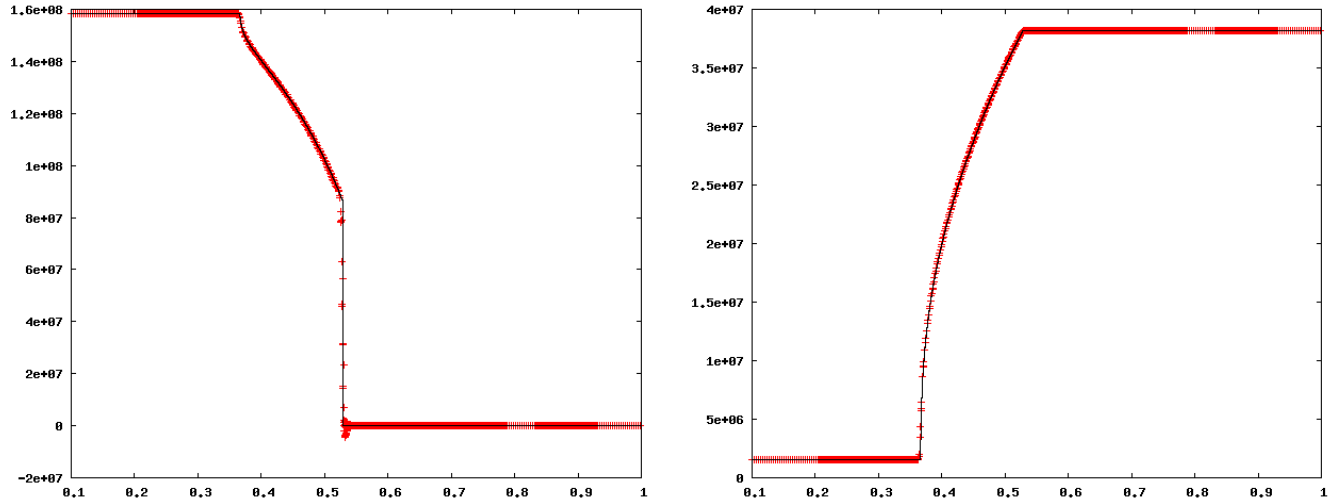

Fig. 14 1D-plot of the numerical solution (symbols) compared to analytic (plain line) for the 512×80 zones mesh. Left: velocity, right: temperature.

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] Todd Arbogast, Mary F. Wheeler, and Ivan Yotov. Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences. *SIAM J. Numer. Anal.*, 34(2):828–852, 1997.

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

[4] Enrico Bertolazzi and Gianmarco Manzini. On vertex reconstructions for cell-centered finite volume approximations of 2d anisotropic diffusion problems. *Math. Mod. Meth. Appl. Sci.*, 17:1–32, 2007.

[5] Jérôme Breil and Pierre-Henri Maire. A cell-centered diffusion scheme on two-dimensional unstructured meshes. *J. Comput. Phys.*, 224(2):785–823, 2007.

[6] Franco Brezzi, Annalisa Buffa, and Konstantin Lipnikov. Mimetic finite differences for elliptic problems. *ESAIM, Math. Model. Numer. Anal.*, 43(2):277–295, 2009.

[7] Franco Brezzi, Konstantin Lipnikov, and Mikhail Shashkov. Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes. *SIAM J. Numer. Anal.*, 43(5):1872–1896, 2005.

[8] G. Carré, S. Delpino, B. Després, and E. Labourasse. A cell-centered Lagrangian hydrodynamics scheme on general unstructured meshes in arbitrary dimension. *J. Comput. Phys.*, 228:5160–5183, 2009.

[9] Yves Coudière, Jean-Paul Vila, and Philippe Villedieu. Convergence rate of a finite volume scheme for a two dimensional convection-diffusion problem. *M2AN, Math. Model. Numer. Anal.*, 33(3):493–516, 1999.

[10] A. A. Danilov and Yu. V. Vassilevski. A monotone nonlinear finite volume method for diffusion equations on conformal polyhedral meshes. *Russian J. Numer. Anal. Math. Modelling*, 24(3):207–227, 2009.

[11] B. Després and C. Mazeran. Lagrangian gas dynamics in 2D and lagrangian systems. *Arch. Rat. Mech. anal.*, 178:327–372, 2005.

[12] Jerome Droniou. Finite volume schemes for diffusion equations: Introduction to and review of modern methods. *Math. Models Methods Appl. Sci.*, 24(8):1575–1619, 2014.

[13] Jérôme Droniou, Robert Eymard, Thierry Gallouët, and Raphaèle Herbin. A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. *Math. Models Methods Appl. Sci.*, 20(2):265–295, 2010.

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

[15] M.G. Edwards and C. F. Rogers. Finite volume discretization with imposed flux continuity for the general tensor pressure equation. *Comput. Geosci.*, 2:259–290, 1998.

[16] Michael G. Edwards and Hongwen Zheng. A quasi-positive family of continuous Darcy-flux finite-volume schemes with full pressure support. *J. Comput. Phys.*, 227(22):9333–9364, 2008.

[17] G. T. Eigestad, I. Aavatsmark, and M. Espedal. Symmetry and M-matrix issues for the O-method on an unstructured grid. *Comput. Geosci.*, 6(3-4):381–404, 2002. Locally conservative numerical methods for flow in porous media.

[18] L.C. Evans. Application of nonlinear semigroup theory to certain partial differential equations. *Nonlinear Evolution Equations*, pages 163–188, 1978.

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

[20] Helmer A. Friis and Michael G. Edwards. A family of MPFA finite-volume schemes with full pressure support for the general tensor pressure equation on cell-centered triangular grids. *J. Comput. Phys.*, 230(1):205–231, 2011.

[21] F. Hermeline. A finite volume method for second-order elliptic equations. (Une méthode de volumes finis pour les équations elliptiques du second ordre.). *C. R. Acad. Sci. Paris, Ser. I*, 1998.

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

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

[24] F. Hermeline. Approximation of 2D and 3D diffusion operators with variable full tensor coefficients on arbitrary meshes. *Comput. Methods Appl. Mech. Eng.*, 196(21-24):2497–2526, 2007.

[25] 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.

[26] F. Hermeline. A finite volume method for approximating 3D diffusion operators on general meshes. *J. Comput. Phys.*, 228(16):5763–5786, 2009.

[27] Eirik Keilegavlen, Jan M. Nordbotten, and Ivar Aavatsmark. Sufficient criteria are necessary for monotone control volume methods. *Appl. Math. Lett.*, 22(8):1178–1180, 2009.

[28] David S. Kershaw. Differencing of the diffusion equation in Lagrangian hydrodynamic codes. *J. Comput. Phys.*, 39:375–395, 1981.

[29] Y. Kuznetsov, K. Lipnikov, and M. Shashkov. The mimetic finite difference method on polygonal meshes for diffusion-type problems. *Comput. Geosci.*, 8(4):301–324, 2005.

[30] Christophe Le Potier. A linear scheme satisfying a maximum principle for anisotropic diffusion operators on distorted grids. (Un schéma linéaire vérifiant le principe du maximum pour des opérateurs de diffusion très anisotropes sur des maillages déformés.). *C. R., Math., Acad. Sci. Paris*, 347(1-2):105–110, 2009.

[31] 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.

[32] 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.

[33] K. Lipnikov, D. Svyatskiy, and Y. Vassilevski. A monotone finite volume method for advection-diffusion equations on unstructured polygon meshes. *J. Comput. Phys.*, 229(11):4017–4032, 2010.

[34] K. Lipnikov, D. Svyatskiy, and Y. Vassilevski. Anderson acceleration for nonlinear finite volume scheme for advection-diffusion problems. *SIAM J. Sci. Comput.*, 35(2):A1120–A1136, 2013.

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

[36] Konstantin Lipnikov, Gianmarco Manzini, and Mikhail Shashkov. Mimetic finite difference method. *Journal of Computational Physics*, 257, Part B(0):1163 – 1227, 2014. Physics-compatible numerical methods.

[37] Konstantin Lipnikov, Gianmarco Manzini, and Daniil Svyatskiy. Monotonicity conditions in the mimetic finite difference method. In *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, volume 4 of *Springer Proc. Math.*, pages 653–661. Springer, Heidelberg, 2011.

[38] Konstantin Lipnikov, Mikhail Shashkov, and Daniil Svyatskiy. The mimetic finite difference discretization of diffusion problem on unstructured polyhedral meshes. *J. Comput. Phys.*, 211(2):473–491, 2006.

[39] R. Lowrie and R. Rauenzahn. Radiative shock solutions in the equilibrium diffusion limit. *Shock waves*, 16:445–453, 2007.

[40] R. E. Marshak. Effect of radiation on shock wave behavior. *Phys. Fluids*, 1:24, 1958.

[41] K. Nikitin and Yu. Vassilevski. A monotone nonlinear finite volume method for advection-diffusion equations on unstructured polyhedral meshes in 3D. *Russian J. Numer. Anal. Math. Modelling*, 25(4):335–358, 2010.
[42] W. F. Noh. Errors for calculations of strong shocks using artificial viscosity and an artificial heat flux. *J. Comput. Phys.*, 72:78–120, 1987.
[43] G. J. Pert. Physical constraints in numerical calculations of diffusion. *J. Comput. Phys.*, 42(1):20–52, 1981.
[44] R.J. Plemmons. M-matrix characterizations.i – nonsingular m-matrices. *Linear Algebra and its Applications*, 18(2):175 – 188, 1977.
[45] P.-A. Raviart and J.-M. Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles*. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris, 1983.
[46] C. A. Rogers. A less strange version of Milnor's proof of Brouwer's fixed-point theorem. *Amer. Math. Monthly*, 87(7):525–527, 1980.
[47] Denis Serre. *Matrices*, volume 216 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. Theory and applications, Translated from the 2001 French original.
[48] Zhiqiang Sheng and Guangwei Yuan. A nine point scheme for the approximation of diffusion operators on distorted quadrilateral meshes. *SIAM J. Sci. Comput.*, 30:1241–1361, 2008.
[49] Zhiqiang Sheng and Guangwei Yuan. The finite volume scheme preserving extremum principle for diffusion equations on polygonal meshes. *J. Comput. Phys.*, 230(7):2588–2604, 2011.
[50] Zhiqiang Sheng, Jingyan Yue, and Guangwei Yuan. Monotone finite volume schemes of nonequilibrium radiation diffusion equations on distorted meshes. *SIAM J. Sci. Comput.*, 31(4):2915–2934, 2009.
[51] V. Siess. A linear and accurate diffusion scheme respecting the maximum principle on distorted meshes. *C. R. Acad. Sci. Paris, Ser. I*, 347:1317–1320, 2009.
[52] Shuai Wang, Guangwei Yuan, Yonghai Li, and Zhiqiang Sheng. A monotone finite volume scheme for advection-diffusion equations on distorted meshes. *Internat. J. Numer. Methods Fluids*, 69(7):1283–1298, 2012.
[53] Guangwei Yuan and Zhiqiang Sheng. Monotone finite volume schemes for diffusion equations on polygonal meshes. *Journal of Computational Physics*, 227(12):6288–6312, June 2008.