Finite volume method for 2D linear and nonlinear elliptic problems with discontinuities

Franck Boyer, Florence Hubert

FINITE VOLUME METHOD FOR 2D LINEAR AND NONLINEAR ELLIPTIC PROBLEMS WITH DISCONTINUITIES

FRANCK BOYER AND FLORENCE HUBERT*

Abstract. In this paper we study the approximation of solutions to linear and nonlinear elliptic problems with discontinuous coefficients in the Discrete Duality Finite Volume framework. This family of schemes allows very general meshes and inherits the main properties of the continuous problem.

In order to take into account the discontinuities and to prevent consistency defect in the scheme, we propose to modify the definition of the numerical fluxes on the edges of the mesh where the discontinuity occurs. We first illustrate our approach by the study of the 1D situation. Then, we show how to design our new scheme, called m-DDFV, and we propose its analysis. We also describe an iterative solver, whose convergence is proved, which can be used to solve the nonlinear discrete equations defining the finite volume scheme.

Finally, we provide numerical results which confirm that the m-DDFV scheme significantly improves the convergence rate of the usual DDFV method for both linear and nonlinear problems.

Key words. Finite Volume schemes, Discontinuous coefficients, Nonlinear elliptic problems.

AMS subject classifications. 35J65 - 65N15 - 74S10

1. Introduction. In this paper, we are concerned with the finite volume approximation of solutions to the following nonlinear diffusion problem with homogeneous Dirichlet boundary conditions:

$$\begin{cases} -\mathrm{div}(\varphi(z,\nabla u_e(z))) = f(z), & \text{in } \Omega, \\ u_e = 0, & \text{on } \partial\Omega, \end{cases} \tag{1.1}$$

where Ω is a given bounded polygonal domain in $\mathbb{R}^2$. We first recall the classical functional framework ensuring that the problem above is well-posed (see [16]). Let $p \in]1,\infty[$ and $p' = \frac{p}{p-1}$. The flux $\varphi : \Omega \times \mathbb{R}^2 \to \mathbb{R}^2$ in equation (1.1) is supposed to be a Caratheodory function which is strictly monotonic with respect to $\xi \in \mathbb{R}^2$:

$$(\varphi(z,\xi) - \varphi(z,\eta), \xi - \eta) > 0, \text{ for all } \xi \neq \eta, \text{ for } a.e.\ z \in \Omega\,. \tag{$\mathcal{H}_1$}$$

We also assume that there exist $C_\varphi > 0$ such that

$$(\varphi(z,\xi),\xi) \geq \frac{1}{C_\varphi}|\xi|^p - C_\varphi, \text{ for all } \xi \in \mathbb{R}^2, \text{ for } a.e.\ z \in \Omega, \tag{$\mathcal{H}_2$}$$

$$|\varphi(z,\xi)| \leq C_\varphi(1 + |\xi|^{p-1}), \text{ for all } \xi \in \mathbb{R}^2, \text{ for } a.e.\ z \in \Omega. \tag{$\mathcal{H}_3$}$$

These assumptions ensure that $u \mapsto -\mathrm{div}(\varphi(\,\cdot\,,\nabla u))$ is a Leray-Lions operator, and hence that Problem (1.1) has a unique solution in $W_0^{1,p}(\Omega)$ for any $f \in W^{-1,p'}(\Omega)$. Nevertheless, since we are particularly interested in proving error estimates for (piecewise) smooth enough solutions, we restrict our attention here to source terms f in $L^{p'}(\Omega)$.

In the present work, we concentrate on the case where the flux φ defining the equation admits discontinuities with respect to the space variable z. This kind of

*LATP, Université de Provence, 39 rue F. Joliot Curie, 13453 Marseille Cedex 13, FRANCE. [fboyer,fhubert]@cmi.univ-mrs.fr

transmission (or bimaterial) problems were, for instance, studied in the finite element framework in [11, 20] for $p = 2$ and in [17, 18] for $p \neq 2$.

Finite volume approximation of such nonlinear elliptic problems is a current research topic. We refer for instance to [3, 4, 8] for the description and the analysis of the main available schemes up to now. More precisely, we proposed in [3] to approach the solution to (1.1) by using a Discrete Duality Finite Volume method (DDFV for short). This method (previously studied in [7, 14, 15]), can be applied to a wide class of 2D meshes (note that 3D cases can also be treated, see [5, 19]) and inherits the main qualitative properties of the continuous problem. Hence, we succeeded in showing the convergence of such schemes and error estimates in the case where the flux φ and the exact solution u_e are assumed to be smooth enough. In the case where φ has discontinuous coefficients, our results in [3] show that the scheme is still convergent but the error analysis is no more valid.

Actually, it is known (even for a 1D linear equation) that such discontinuities in the coefficients imply a consistency defect in the numerical fluxes of usual finite volume schemes. In the linear case, this leads to a $\frac{1}{2}$ convergence rate in the discrete H^1 norm instead of the first order we may expect. The situation is the same for DDFV schemes and it is needed to modify the scheme in order to take into account the jumps of the coefficients of the problem and then to recover a better convergence rate.

The aim of this work it is to present a modified DDFV scheme in this framework - that we called m-DDFV- which enjoys a better convergence rate than the usual DDFV method. Then we provide the error analysis of this scheme. In particular, in the linear case, we prove the first order convergence of the m-DDFV scheme. Hence, our analysis provides a theoretical confirmation of the behavior numerically observed in a particular case in [15].

Outline. In Section 2, we propose to study a simple 1D problem where the flux φ has only one point of discontinuity. This section will let us introduce the main ideas of the method we propose and illustrate the way one can obtain the consistency estimate for the scheme under study.

In Section 3, we recall the DDFV framework for the finite volume approximation of nonlinear elliptic problems on unstructured grids. We also recall the scheme introduced and analyzed in [3]. In Section 4, we describe the m-DDFV scheme and its first properties.

Section 5 is devoted to the error analysis of this method in the case where the exact solution is assumed to be piecewise smooth enough. The main new difficulty in the analysis, compared to the ones already encountered in [3], is contained in the consistency estimate of the new discrete gradient operator introduced in Section 4. As an illustration, we give some explicit examples of the schemes under study in Section 6. Nevertheless, in general, the method is not explicit and then seems to be difficult to solve. That is the reason why in Section 7 we propose an iterative explicit algorithm to compute the approximate solution for any given data and we prove its convergence.

Notice that we also introduce a so-called *hybrid* DDFV scheme, called h-DDFV, for which a better error estimate can be obtained in the very common case where the flux is in fact smooth enough on a finite number of subdomains covering the whole domain Ω.

We finally conclude this paper by showing, in Section 8, some numerical results illustrating both the efficiency of the finite volume scheme and of the iterative solver.

2. A 1D finite volume method for a model problem.

2.1. The toy system. Let us consider in this section a model problem in 1D of the form (1.1) in order to illustrate the main steps we will follow in the sequel of this paper for 2D problems. We take $\Omega =]-1,1[$ (denoting here by x the space variable) and we define $\varphi(x,\xi) = \varphi_-(\xi)$ for $x < 0$ and $\varphi(x,\xi) = \varphi_+(\xi)$ for $x > 0$. We suppose that φ_- and φ_+ are two strictly monotonic maps from $\mathbb{R}$ to $\mathbb{R}$ such that $(\mathcal{H}_2)$ and $(\mathcal{H}_3)$ hold.

Problem (1.1) reads in this setting

$$\begin{cases} -\partial_x\big(\varphi_-(\partial_x u_e)\big) = f(x), \text{ for } x < 0, \\ -\partial_x\big(\varphi_+(\partial_x u_e)\big) = f(x), \text{ for } x > 0, \\ u_e(-1) = u_e(1) = 0, \\ \varphi_-(\partial_x u_e(0^-)) = \varphi_+(\partial_x u_e(0^+)). \end{cases} \tag{2.1}$$

2.2. The 1D finite volume scheme. Suppose now that we are given a finite volume mesh $\mathcal{T}$ of the domain Ω compatible with the discontinuity point. More precisely, let $x_0 = -1 < \ldots < x_N = 0 < \ldots < x_{N+M} = 1$ a subdivision of $[-1,1]$. We denote by $\kappa_{i+\frac{1}{2}} = [x_i, x_{i+1}]$, $i \in \{0, N+M-1\}$ the control volumes of this discretization and by $x_{i+\frac{1}{2}}$ their centers. The finite volume method associates to each center $x_{i+\frac{1}{2}}$ an unknown value $u_{i+\frac{1}{2}}$. We denote by $u^{\mathcal{T}} = (u_{i+\frac{1}{2}})_{0\le i\le N+M-1}$ the whole approximate solution and we define the usual difference quotients

$$\nabla_i u^{\mathcal{T}} = \frac{u_{i+\frac{1}{2}} - u_{i-\frac{1}{2}}}{x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}}, \ i \in \{0, N+M\},$$

where, conventionally, we set $x_{-\frac{1}{2}} = x_0 = -1$, $x_{N+M+\frac{1}{2}} = x_{N+M} = 1$ and $u_{-\frac{1}{2}} = u_{N+M+\frac{1}{2}} = 0$. To obtain the finite volume scheme, we integrate the problem (2.1) on each control volume

$$-\int_{\kappa_{i+\frac{1}{2}}} \partial_x\big(\varphi(x,\partial_x u_e)\big)\,dx = \int_{\kappa_{i+\frac{1}{2}}} f(x)\,dx, \quad \forall i \in \{1,\ldots,N+M-1\}.$$

Integrating the first term by parts, the scheme reads

$$-F_{i+1} + F_i = \int_{\kappa_{i+\frac{1}{2}}} f(x)\,dx, \quad \forall i \in \{0, N+M-1\}, \tag{2.2}$$

where F_i, $i \in \{0, N+M\}$ is an approximation of the flux $\varphi(x_i, \partial_x u_e(x_i))$. This approximation can easily be obtained away from the discontinuity in the usual way:

$$\begin{cases} F_i = \varphi_-\left(\nabla_i u^{\mathcal{T}}\right), \quad \forall i \in \{0, N-1\}, \\ F_i = \varphi_+\left(\nabla_i u^{\mathcal{T}}\right), \quad \forall i \in \{N+1, N+M\}. \end{cases} \tag{2.3}$$

The problem is: how do we choose the numerical flux F_N at the point $x_N = 0$ where the discontinuity occurs? One may imagine many naive ways to treat this problem. For instance one can define F_N by:

$$F_N\varphi_-\left(\nabla_N u^{\mathcal{T}}\right), \text{ or } F_N = \varphi_+\left(\nabla_N u^{\mathcal{T}}\right), \text{ or } F_N = \frac{1}{2}(\varphi_-\left(\nabla_N u^{\mathcal{T}}\right) + \varphi_+\left(\nabla_N u^{\mathcal{T}}\right)).$$

In fact, it can be shown that all these choices lead, in general, to a non consistent approximation of the flux at x_N. This fact is well known even in the linear case

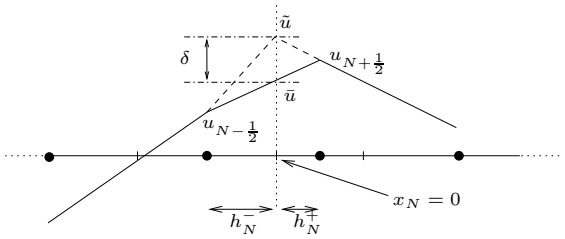


FIG. 2.1. *Illustration of the 1D case*

(see e.g. [9]). The good way to find out a consistent approximation of the flux is to introduce a new artificial unknown $\tilde{u}$ at the point of the discontinuity x_N so that we can define two different approximate gradients on both sides of the discontinuity

$$\nabla_N^+ u^{\mathcal{T}} = \frac{u_{N+\frac{1}{2}} - \tilde{u}}{h_N^+}, \quad \text{and } \nabla_N^- u^{\mathcal{T}} = \frac{\tilde{u} - u_{N-\frac{1}{2}}}{h_N^-}, \tag{2.4}$$

where we set $h_N^+ = x_{N+\frac{1}{2}} - x_N$ and $h_N^- = x_N - x_{N-\frac{1}{2}}$. In fact, it is convenient (see Figure 2.1 and the discussion below) to look for $\tilde{u}$ under the form

$$\tilde{u} = \bar{u} + \delta, \quad \text{with } \bar{u} = \frac{h_N^- u_{N+\frac{1}{2}} + h_N^+ u_{N-\frac{1}{2}}}{h_N^- + h_N^+}.$$

The value $\bar{u}$ is the value at the point 0 of the affine interpolation between $(x_{N-\frac{1}{2}}, u_{N-\frac{1}{2}})$ and $(x_{N+\frac{1}{2}}, u_{N+\frac{1}{2}})$. From now on, δ is the new artificial unknown to be determined. It follows that

$$\nabla_N^+ u^{\mathcal{T}} = \nabla_N u^{\mathcal{T}} - \frac{\delta}{h_N^+}, \quad \text{and } \nabla_N^- u^{\mathcal{T}} = \nabla_N u^{\mathcal{T}} + \frac{\delta}{h_N^-}.$$

Notice that we have

$$\nabla_N u^{\mathcal{T}} = \frac{1}{h_N^- + h_N^+} \left(h_N^- \nabla_N^- u^{\mathcal{T}} + h_N^+ \nabla_N^+ u^{\mathcal{T}} \right). \tag{2.5}$$

It is now necessary to eliminate the new unknown δ. This is done by imposing a discrete equivalent of the transmission condition in (2.1) which reads

$$\varphi_-(\nabla_N^- u^{\mathcal{T}}) = \varphi_+(\nabla_N^+ u^{\mathcal{T}}). \tag{2.6}$$

This equation uniquely defines δ as a function $\delta = \delta_N(\nabla_N u^{\mathcal{T}})$ of the usual difference quotient $\nabla_N u^{\mathcal{T}}$ since the map $\delta \mapsto \varphi_-(\nabla_N^- u^{\mathcal{T}}) - \varphi_+(\nabla_N^+ u^{\mathcal{T}})$ is strictly monotonic and tends to infinity at infinity. Notice that $\delta_N(0)$ is always 0. In the particular case where $\varphi_- = \varphi_+$ then δ_N is then identically zero and then $\nabla_N^+ u^{\mathcal{T}} = \nabla_N^- u^{\mathcal{T}} = \nabla_N u^{\mathcal{T}}$. Hence, we recover the generic situation without discontinuities in the coefficients of the equation. We can eventually define the approximate flux at the discontinuity by

$$F_N = \varphi_- \left(\nabla_N u^{\mathcal{T}} + \frac{\delta_N(\nabla_N u^{\mathcal{T}})}{h_N^-} \right) = \varphi_+ \left(\nabla_N u^{\mathcal{T}} - \frac{\delta_N(\nabla_N u^{\mathcal{T}})}{h_N^+} \right),$$

the last equality being true by definition of $\delta_N(\nabla_N u^{\mathcal{T}})$. In a more symmetric way we also have

$$F_N = \frac{h_N^- \varphi_- \left(\nabla_N u^{\mathcal{T}} + \frac{\delta_N(\nabla_N u^{\mathcal{T}})}{h_N^-} \right) + h_N^+ \varphi_+ \left(\nabla_N u^{\mathcal{T}} - \frac{\delta_N(\nabla_N u^{\mathcal{T}})}{h_N^+} \right)}{h_N^- + h_N^+}. \tag{2.7}$$

EXAMPLE 2.1. *Let us consider the case where φ_- and φ_+ are two p-laplacian like fluxes given by*

$$\varphi_-(\xi) = k_-|\xi + G_-|^{p-2}(\xi + G_-),\ \forall \xi \in \mathbb{R},$$
$$\varphi_+(\xi) = k_+|\xi + G_+|^{p-2}(\xi + G_+),\ \forall \xi \in \mathbb{R},$$

where $k_-, k_+ \in \mathbb{R}^+$ and $G_-, G_+ \in \mathbb{R}^2$. In this situation, all the computations can be made by hand. In particular, equation (2.6) can be solved and finally, the numerical flux at the discontinuity is found to be

$$F_N = \left(\frac{k_-^{\frac{1}{p-1}} k_+^{\frac{1}{p-1}} (h_N^- + h_N^+)}{h_N^+ k_-^{\frac{1}{p-1}} + h_N^- k_+^{\frac{1}{p-1}}} \right)^{p-1} \left| \nabla_N u^{\mathcal{T}} + \overline{G} \right|^{p-2} \left(\nabla_N u^{\mathcal{T}} + \overline{G} \right),$$

where $\overline{G}$ is the arithmetic mean-value between G_- and G_+ defined by

$$\overline{G} = \frac{h_N^- G_- + h_N^+ G_+}{h_N^- + h_N^+}.$$

Notice that the map $\nabla_N u^{\mathcal{T}} \mapsto F_N$ is monotonic and coercive. In the linear case (i.e. $p = 2$) we recover the well-known harmonic mean-value formula between the two diffusion coefficients k_- and k_+ (see for instance [9]):

$$F_N = \frac{k_- k_+ (h_N^- + h_N^+)}{h_N^+ k_- + h_N^- k_+} \left(\nabla_N u^{\mathcal{T}} + \overline{G} \right).$$

Let us sum up the previous study: we defined a monotonic map $\nabla_N u^{\mathcal{T}} \mapsto \delta_N(\nabla_N u^{\mathcal{T}})$ and a numerical flux F_N at the discontinuity which is also a monotonic map with respect to $\nabla_N u^{\mathcal{T}}$. The finite volume scheme is then given by (2.2) with (2.3) and (2.7).

A very important remark is that the map δ_N is defined through the implicit relation (2.6) and hence, in general, can not be computed explicitly like in Example 2.1. At a first sight, it can be considered as a major drawback of our approach. Nevertheless, we will propose in Section 7 a fully practical solver for this nonlinear scheme whose convergence is proved and whose computational cost is of the same order as in the case of continuous coefficient equations.

2.3. Consistency analysis. Let us analyse the consistency property of the flux F_N defined above. The following computations give the main ideas used in the analysis of the 2D scheme presented in the sequel of this paper. For simplicity we assume in this section that $p > 2$ and we suppose that φ_- and φ_+ satisfy the strong monotonicity assumption ($\mathcal{H}_{1'b}$) and the Hölder regularity assumption ($\mathcal{H}_{4b}$) described in Section 4.1.

Finally we suppose that the exact solution u_e of (2.1) is continuous on Ω and smooth on the two sides of the discontinuity point $x_N = 0$. In order to simplify the notations, assume that $h_N^- = h_N^+$ and denote this common value by h. Our goal is to estimate the consistency error of the flux F_N which amounts (by ($\mathcal{H}_{4b}$)) to estimate quantities like

$$R = \frac{1}{h} \int_{-h}^{0} |\partial_x u_e - \nabla_N^+ \mathbb{P}^{\mathcal{T}} u_e|^p \, dx,$$

where $\mathbb{P}^{\mathcal{T}} u_e = (u_e(x_{i+\frac{1}{2}}))_{0\leq i\leq N+M-1}$. Since u_e is smooth on $[-h,0]$ we have

$$\begin{aligned} R &\leq C\|\partial_x^2 u_e\|_\infty h^p + \frac{1}{h}\int_{-h}^0 \left| \frac{u_e(x_N) - u_e(x_{N-\frac{1}{2}})}{h} - \frac{\bar{u} - u_e(x_{N-\frac{1}{2}})}{h} \right|^p dx \\ &= C\|\partial_x^2 u_e\|_\infty h^p + \left| \frac{u_e(x_N) - \bar{u}}{h} \right|^p, \end{aligned} \tag{2.8}$$

where $\bar{u}$ is the artifical unknown defined in (2.6), that is through the equation

$$\varphi_-\left(\frac{\bar{u} - u_e(x_{N-\frac{1}{2}})}{h}\right) = \varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - \bar{u}}{h}\right). \tag{2.9}$$

Furthermore, since u_e is piecewise smooth, Taylor expansions yield

$$\frac{u_e(x_N) - u_e(x_{N-\frac{1}{2}})}{h} = \partial_x u_e(0^-) + T_1 h,$$

$$\frac{u_e(x_{N+\frac{1}{2}}) - u_e(x_N)}{h} = \partial_x u_e(0^+) + T_2 h,$$

where T_1 and T_2 are bounded with respect to h. Then, we use the transmission condition in (2.1) which gives

$$\varphi_-\left(\frac{u_e(x_N) - u_e(x_{N-\frac{1}{2}})}{h} - T_1 h\right) = \varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - u_e(x_N)}{h} - T_2 h\right). \tag{2.10}$$

Finally, we estimate the second term in the right-hand side of (2.8) by using (2.9) and (2.10). To this end, we subtract (2.10) from (2.9) and we multiply by $\frac{u_e(x_N)-\bar{u}}{h}$. It follows

$$\begin{aligned} &\left(\varphi_-\left(\frac{u_e(x_N) - u_e(x_{N-\frac{1}{2}})}{h} - T_1 h\right) - \varphi_-\left(\frac{\bar{u} - u_e(x_{N-\frac{1}{2}})}{h}\right)\right)\frac{u_e(x_N) - \bar{u}}{h} \\ &\quad + \left(\varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - \bar{u}}{h}\right) - \varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - u_e(x_N)}{h} - T_2 h\right)\right)\frac{u_e(x_N) - \bar{u}}{h} = 0. \end{aligned}$$

We add and subtract now the terms $T_1 h$ and $T_2 h$ in order to make appear expressions under the form $(\varphi_\pm(\xi) - \varphi_\pm(\eta))(\xi - \eta)$. We get

$$\begin{aligned} &\left(\varphi_-\left(\frac{u_e(x_N) - u_e(x_{N-\frac{1}{2}})}{h} - T_1 h\right) - \varphi_-\left(\frac{\bar{u} - u_e(x_{N-\frac{1}{2}})}{h}\right)\right)\left(\frac{u_e(x_N) - \bar{u}}{h} - T_1 h\right) \\ &+ \left(\varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - \bar{u}}{h}\right) - \varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - u_e(x_N)}{h} - T_2 h\right)\right)\left(\frac{u_e(x_N) - \bar{u}}{h} + T_2 h\right) \\ &\quad = -T_1 h\left(\varphi_-\left(\frac{u_e(x_N) - u_e(x_{N-\frac{1}{2}})}{h} - T_1 h\right) - \varphi_-\left(\frac{\bar{u} - u_e(x_{N-\frac{1}{2}})}{h}\right)\right) \\ &\qquad + T_2 h\left(\varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - \bar{u}}{h}\right) - \varphi_+\left(\frac{u_e(x_{N+\frac{1}{2}}) - u_e(x_N)}{h} - T_2 h\right)\right). \end{aligned}$$

Hence, using assumptions $(\mathcal{H}_{1'b})$ and $(\mathcal{H}_{4b})$ we deduce that

$$\left|\frac{u_e(x_N)-\bar{u}}{h}-T_1h\right|^p+\left|\frac{u_e(x_N)-\bar{u}}{h}+T_2h\right|^p \leq Ch\left(1+\left|\frac{u_e(x_N)-\bar{u}}{h}\right|^{p-2}\right)\left(Ch+\left|\frac{u_e(x_N)-\bar{u}}{h}\right|\right),$$

and finally we have

$$\left|\frac{u_e(x_N)-\bar{u}}{h}\right|^p \leq Ch^{\frac{p}{p-1}},$$

so that the consistency term R is finally bounded by

$$R \leq Ch^{\frac{p}{p-1}}.$$

When $p \to 2$ we recover the usual first order estimate (that is $R = O(h^2)$) whereas when p increases, this consistency order decreases.

3. The discrete duality finite volume framework.

3.1. Definition of the mesh. We recall the notations used in [3]. Let $\mathcal{T}$ be a triple $(\mathfrak{M}, \mathfrak{M}^*, \mathfrak{D})$ of meshes on Ω as follows. The set $\mathfrak{M}$ is a set of disjoint open

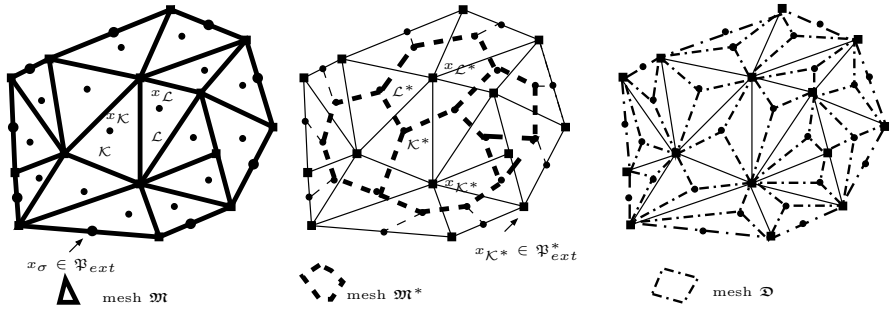


FIG. 3.1. *Example of a DDFV mesh*

polygonal convex control volumes $\kappa \in \Omega$ such that $\cup\bar{\kappa} = \bar{\Omega}$. For all adjacent volume κ and $\mathcal{L}$, we assume that $\partial\kappa \cap \partial\mathcal{L}$ is a segment that we call *an edge of the mesh* and that we denote by $\sigma = \kappa|\mathcal{L}$. Let $\mathcal{E}_{int}$ denotes the set of such edges. The set $\mathcal{E}_{ext}$ denotes the set of edges $\sigma = \partial\kappa \cap \partial\Omega$ and $\mathcal{E} = \mathcal{E}_{int} \cup \mathcal{E}_{ext}$. We associate to $\mathfrak{M}$ a family $\mathfrak{P}_{int}$ of points x_κ such that $x_\kappa \in \kappa$ and to the set $\mathcal{E}_{ext}$ a family $\mathfrak{P}_{ext}$ of points x_σ where x_σ is a point of $\sigma \in \mathcal{E}_{ext}$. Let $\mathfrak{P}^*$ be the set of vertices of the mesh $\mathfrak{M}$. The set $\mathfrak{P}^*$ can be decomposed into $\mathfrak{P}^* = \mathfrak{P}^*_{int} \cup \mathfrak{P}^*_{ext}$, where $\mathfrak{P}^*_{ext} \subset \partial\Omega$ and $\mathfrak{P}^*_{int} \cap \partial\Omega = \emptyset$. To any point $x_{\kappa^*} \in \mathfrak{P}^*_{int}$, we associate a polygon $\kappa^* \in \mathfrak{M}^*$ whose vertices are $\{x_\kappa \in \mathfrak{P} / x_{\kappa^*} \in \bar{\kappa},\ \kappa \in \mathfrak{M}\}$ sorted with respect to the clockwise order of the corresponding primal control volumes. The set $\mathcal{E}^*$ denotes the set the edges of the mesh $\mathfrak{M}^*$.

For each $\sigma = \kappa|\mathcal{L} \in \mathcal{E}_{int}$, we can associate a diamond cell $\mathcal{D}$ where $\mathcal{D}$ is the quadrangle whose diagonals are $\sigma = (x_{\kappa^*}, x_{\mathcal{L}^*})$ and $\sigma^* = \kappa^*|\mathcal{L}^* = (x_\kappa, x_\mathcal{L})$ if $\sigma \in \mathcal{E}_{int}$ and if $\sigma = (x_{\kappa^*}, x_{\mathcal{L}^*}) \in \mathcal{E}_{ext} \cap \partial\bar{\kappa}$, $\mathcal{D}$ is the triangle defined by the points $x_\kappa, x_{\kappa^*}, x_{\mathcal{L}^*}$.

The set of all diamond cell is noted $\mathfrak{D} = \mathfrak{D}_{\rm int} \cup \mathfrak{D}_{\rm ext}$. Remark that $\mathfrak{D}$ form a partition of Ω.

In this work, we assume that the diamond cells are all convex. Notice that this assumption is not necessary, in general, in order to define and analyse DDFV methods (see [3],[7]).

3.2. Notations. For any one (resp. two) dimensional set $\mathcal{V}$, we denote by $|\mathcal{V}|$ its one-dimensional (resp. two-dimensional) Lebesgue measure.

For any control volume $\kappa \in \mathfrak{M}$, we define

- $\mathfrak{D}_\kappa = \{\mathcal{D} \in \mathfrak{D} / \kappa \cap \mathcal{D} \neq \emptyset\}$.
- $\boldsymbol{\nu}_\kappa$, the outward unit normal vector to $\partial\kappa$.
- d_κ, the diameter of κ.

In the same way, for a dual control volume $\kappa^* \in \mathfrak{M}^*$, we set

- $\mathfrak{D}_{\kappa^*} = \{\mathcal{D} \in \mathfrak{D} / \kappa^* \cap \mathcal{D} \neq \emptyset\}$.
- $\boldsymbol{\nu}_{\kappa^*}$, the outward unit normal vector to $\partial\kappa^*$.
- d_{κ^*}, the diameter of κ^*.

For a diamond cell $\mathcal{D} \in \mathfrak{D}_{\rm int}$ (resp. $\mathcal{D} \in \mathfrak{D}_{\rm ext}$), recall that $(x_\mathcal{K}, x_{\mathcal{K}^*}, x_\mathcal{L}, x_{\mathcal{L}^*})$ are the vertices of $\mathcal{D}$ (resp. $(x_\mathcal{K}, x_{\mathcal{K}^*}, x_{\mathcal{L}^*})$ are vertices of $\mathcal{D}$ and $x_\sigma \in \partial\mathcal{D}$) and note :

- $\boldsymbol{\tau}$, the unit vector parallel to σ, oriented from $x_{\mathcal{K}^*}$ to $x_{\mathcal{L}^*}$.
- $\boldsymbol{\nu}$, the unit vector normal to σ, oriented from $x_\mathcal{K}$ to $x_\mathcal{L}$ (resp. from $x_\mathcal{K}$ to x_σ).
- $\boldsymbol{\tau}^*$, the unit vector parallel to σ^*, oriented from $x_\mathcal{K}$ to $x_\mathcal{L}$ (resp. from $x_\mathcal{K}$ to x_σ).
- $\boldsymbol{\nu}^*$, the unit vector normal to σ^*, oriented from $x_{\mathcal{K}^*}$ to $x_{\mathcal{L}^*}$.
- $\alpha_\mathcal{D}$, the angle between $\boldsymbol{\tau}$ and $\boldsymbol{\tau}^*$.
- $\mathrm{d}_\mathcal{D}$, the diameter of $\mathcal{D}$.
- $x_\mathcal{D}$ the intersection of $(x_\mathcal{K}, x_\mathcal{L})$ and $(x_{\mathcal{K}^*}, x_{\mathcal{L}^*})$ (resp. $x_\mathcal{D} = x_\sigma$).

Each diamond cell $\mathcal{D} \in \mathfrak{D}_{\rm int}$ (resp. $\mathcal{D} \in \mathfrak{D}_{\rm ext}$) can naturally be split into four triangles (resp. two triangles) $\mathcal{Q} \in \mathfrak{Q}_\mathcal{D}$ as shown in Figure 3.2

$$\overline{\mathcal{D}} = \overline{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} \cup \overline{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}} \cup \overline{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}} \cup \overline{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}, \text{ if } \mathcal{D} \in \mathfrak{D}_{\rm int}, \ \overline{\mathcal{D}} = \overline{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} \cup \overline{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}, \text{ if } \mathcal{D} \in \mathfrak{D}_{\rm ext}.$$

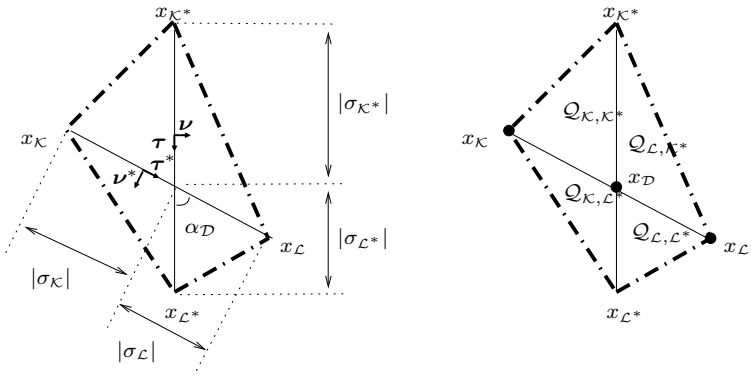


FIG. 3.2. *Notations in a diamond cell; quarter diamonds*

We denote by $\sigma_\mathcal{K}, \sigma_\mathcal{L}, \sigma_{\mathcal{K}^*}, \sigma_{\mathcal{L}^*}$ the segments $(x_\mathcal{K}, x_\mathcal{D})$, $(x_\mathcal{L}, x_\mathcal{D})$, $(x_{\mathcal{K}^*}, x_\mathcal{D})$, $(x_{\mathcal{L}^*}, x_\mathcal{D})$, so that $\sigma = \sigma_{\mathcal{K}^*} \cup \sigma_{\mathcal{L}^*}$ and $\sigma^* = \sigma_\mathcal{K} \cup \sigma_\mathcal{L}$ for $\mathcal{D} \in \mathfrak{D}_{\rm int}$. For $\mathcal{D} \in \mathfrak{D}_{\rm ext}$ we note abusively $\sigma^* = \sigma_\mathcal{K}$.We note $\mathcal{E}_\mathcal{Q}$ the set of such segments included in $\partial\mathcal{Q}$.

3.3. Regularity assumptions for the meshes. We note size($\mathcal{T}$) the maximum of the diameters of the diamond cells in $\mathfrak{D}$. The following bounds follow:

$$|\sigma| \leq \text{size}(\mathcal{T}),\ \forall \sigma \in \mathcal{E}; \quad |\sigma^*| \leq \text{size}(\mathcal{T}),\ \forall \sigma^* \in \mathcal{E}^*;$$
$$|\kappa| \leq \pi\, \text{size}(\mathcal{T})^2,\ \forall \kappa \in \mathfrak{M}; \quad |\kappa^*| \leq \pi\, \text{size}(\mathcal{T})^2,\ \forall \kappa^* \in \mathfrak{M}^*;$$
$$|\mathcal{Q}| \leq \frac{1}{2}\text{size}(\mathcal{T})^2,\ \forall \mathcal{Q} \in \mathfrak{Q}.$$

To measure how flat the diamond cells are, we introduce $\alpha_{\mathcal{T}}$ the unique real in $]0, \frac{\pi}{2}]$ such that $\sin\alpha_{\mathcal{T}} \stackrel{\text{def}}{=} \min_{\mathcal{D}\in\mathfrak{D}} |\sin\alpha_{\mathcal{D}}|$. We also need to control the ratio between the sizes of the quarter diamond cells inside each diamond $\mathcal{D}$. As a consequence, we will measure the regularity of the DDFV mesh by the following quantity

$$\text{reg}(\mathcal{T}) \stackrel{\text{def}}{=} \max\left(\frac{1}{\alpha_{\mathcal{T}}}, \max_{\mathcal{D}\in\mathfrak{D}} \max_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} \frac{\text{d}_{\mathcal{D}}}{\sqrt{|\mathcal{Q}|}} \max_{\substack{\mathcal{K}\in\mathfrak{M}\\ \mathcal{D}\in\mathfrak{D}_{\mathcal{K}}}} \frac{\text{d}_{\mathcal{K}}}{\text{d}_{\mathcal{D}}}, \max_{\substack{\mathcal{K}^*\in\mathfrak{M}^*\\ \mathcal{D}\in\mathfrak{D}_{\mathcal{K}^*}}} \frac{\text{d}_{\mathcal{K}^*}}{\text{d}_{\mathcal{D}}}\right).$$

In particular, there exists two constants C_1 and C_2 depending on reg($\mathcal{T}$) such that for any $\kappa \in \mathfrak{M}$, $\kappa^* \in \mathfrak{M}^*$ and $\mathcal{D} \in \mathfrak{D}$ such that $\mathcal{D} \cap \kappa \neq \emptyset$ and $\mathcal{D} \cap \kappa^* \neq \emptyset$ we have

$$C_1|\kappa| \leq |\mathcal{D}| \leq C_2|\kappa|, \quad C_1|\kappa^*| \leq |\mathcal{D}| \leq C_2|\kappa^*|.$$

3.4. Original DDFV approach for linear problems. The DDFV finite volume method associates to all primal control volume $\kappa \in \mathfrak{M}$ an unknown value u_κ and to all dual control volume $\kappa^* \in \mathfrak{M}^*$ an unknown value u_{κ^*}. The approximate solution $u^{\mathcal{T}}$ is denoted by

$$u^{\mathcal{T}} = \left((u_\kappa)_{\kappa\in\mathfrak{M}}, (u_{\kappa^*})_{\kappa^*\in\mathfrak{M}^*}\right).$$

The set of such unknowns $u^{\mathcal{T}}$ is denoted by $\mathbb{R}^{\mathcal{T}}$.

The method consists in introducing a discrete gradient operator $\nabla^{\mathcal{T}}$ defined to be constant on each diamond cell

$$\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}} = \frac{1}{\sin\alpha_{\mathcal{D}}}\left(\frac{u_{\mathcal{L}} - u_{\mathcal{K}}}{|\sigma^*|}\boldsymbol{\nu} + \frac{u_{\mathcal{L}^*} - u_{\mathcal{K}^*}}{|\sigma|}\boldsymbol{\nu}^*\right)$$

where $u_{\mathcal{K}^*} = 0$ (resp. $u_{\mathcal{L}^*} = 0$) if $x_{\mathcal{K}^*} \in \mathfrak{P}^*_{ext}$ (resp. if $x_{\mathcal{L}^*} \in \mathfrak{P}^*_{ext}$) and $u_{\mathcal{L}} = 0$ if $\mathcal{D} \in \mathfrak{D}_{\text{ext}}$. Then the discrete divergence operator is defined to be the adjoint of $\nabla^{\mathcal{T}}$, so that for linear equations (in particular the Laplace equation), we obtain a well-posed finite volume scheme which is for instance studied in [7, 14]. Notice that the discrete gradient operator $\nabla^{\mathcal{T}}$ was already used *e.g.* in [1, 2, 6], but in these references the values $u_{\mathcal{K}^*}$ and $u_{\mathcal{L}^*}$ on the dual mesh were not considered as unknowns of the problem but were built *via* interpolation formulas between the values of the solution on the primal mesh. In this last approach, there are less unknowns but the structure of the original equation (in particular the symmetry) is lost. The main advantage of the DDFV method is that the discrete equations inherits from the properties of the continuous one, which is crucial in particular in the nonlinear setting.

3.5. The DDFV method for nonlinear elliptic problems. In [3], we have studied the generalization of the DDFV method to the nonlinear equation (1.1). We proved that all the tools used in the study of this equation (monotonicity, compactness, etc...) can be translated to the discrete level. The scheme we proposed consists in integrating the equation (1.1) on each $\kappa \in \mathfrak{M}$ and each $\kappa^* \in \mathfrak{M}^*$ and then to

approximate fluxes $\int_\sigma (\varphi(s,\nabla u_e(s)),\boldsymbol{\nu})\,ds$ or $\int_{\sigma^*} (\varphi(s,\nabla u_e(s)),\boldsymbol{\nu}^*)\,ds$ by using the discrete gradient $\nabla^{\mathcal{T}}$ operator defined above. The scheme now writes:

$$\left\{\begin{array}{l} -\displaystyle\sum_{\mathcal{D}\in\mathfrak{D}_\mathcal{K}} |\sigma|\,(\varphi_\mathcal{D}(\nabla^{\mathcal{T}}_\mathcal{D} u^{\mathcal{T}}),\boldsymbol{\nu}_\mathcal{K}) = |\mathcal{K}| f_\mathcal{K},\ \forall \mathcal{K}\in\mathfrak{M}, \\ -\displaystyle\sum_{\mathcal{D}\in\mathfrak{D}_{\mathcal{K}^*}} |\sigma^*|\,(\varphi_\mathcal{D}(\nabla^{\mathcal{T}}_\mathcal{D} u^{\mathcal{T}}),\boldsymbol{\nu}_{\mathcal{K}^*}) = |\mathcal{K}^*| f_{\mathcal{K}^*},\ \forall \mathcal{K}^*\in\mathfrak{M}^*, \end{array}\right. \tag{3.1}$$

where $f_\mathcal{K}$ and $f_{\mathcal{K}^*}$ denotes the mean value of f over $\mathcal{K}$ and $\mathcal{K}^*$ respectively, and $\varphi_\mathcal{D}$ is the mean-value of φ over $\mathcal{D}$, that is

$$\varphi_\mathcal{D}(\xi) = \frac{1}{|\mathcal{D}|}\int_\mathcal{D} \varphi(z,\xi)\,dz. \tag{3.2}$$

We proved in [3] that this scheme is convergent for any φ satisfying assumptions $(\mathcal{H}_1)$-$(\mathcal{H}_3)$ and any source term $f\in L^{p'}(\Omega)$, and that we can adapt the scheme for source terms in $W^{-1,p'}(\Omega)$ which is the natural space in which source terms can be taken.

We also proved error estimates for the scheme above in the case where the flux φ is assumed to be smooth enough with respect to ξ and to z on the whole domain Ω, and u_e is assumed to belong to $W^{2,p}(\Omega)$.

4. Taking into account discontinuities in the DDFV framework. The point we are concerned with in this paper is that the scheme (3.1) (even though we know that it is convergent) suffers from a lost of consistency in the case where φ presents discontinuities in the space variable z. This behavior is illustrated in Section 8 in comparison with the one of the new scheme we propose in the present section. More precisely, we present a way to recover the consistency of the fluxes even when φ presents jumps. The method essentially follows the line described for the very simple toy 1D problem studied in Section 2.

4.1. Assumptions on the flux φ. We first give the precise assumptions we need on the flux φ. First of all, we reinforce the monotonicity assumption $(\mathcal{H}_1)$ in the following way:

- If $1<p\le 2$: for all $(\xi,\eta)\in\mathbb{R}^2\times\mathbb{R}^2$ and almost every $z\in\Omega$,

$$(\varphi(z,\xi)-\varphi(z,\eta),\xi-\eta) \ge \frac{1}{C_\varphi}|\xi-\eta|^2(1+|\xi|^p+|\eta|^p)^{\frac{p-2}{p}}. \tag{$\mathcal{H}_{1'a}$}$$

- If $p>2$: for all $(\xi,\eta)\in\mathbb{R}^2\times\mathbb{R}^2$ and almost every $z\in\Omega$,

$$(\varphi(z,\xi)-\varphi(z,\eta),\xi-\eta) \ge \frac{1}{C_\varphi}|\xi-\eta|^p. \tag{$\mathcal{H}_{1'b}$}$$

We also assume that the flux φ is Hölder continuous with respect to ξ:

- If $1<p\le 2$: for all $(\xi,\eta)\in\mathbb{R}^2\times\mathbb{R}^2$ and almost every $z\in\Omega$,

$$|\varphi(z,\xi)-\varphi(z,\eta)| \le C_\varphi|\xi-\eta|^{p-1}. \tag{$\mathcal{H}_{4a}$}$$

- If $p>2$: for all $(\xi,\eta)\in\mathbb{R}^2\times\mathbb{R}^2$, and almost every $z\in\Omega$,

$$|\varphi(z,\xi)-\varphi(z,\eta)| \le C_\varphi\left(1+|\xi|^{p-2}+|\eta|^{p-2}\right)|\xi-\eta|. \tag{$\mathcal{H}_{4b}$}$$

The four assumptions above are classical in the error analysis of numerical methods for nonlinear problems and are satisfied by many usual nonlinear operators. We can think for instance to p-laplacian-like operators $\varphi(z,\xi) = k(z)(A(z)\xi,\xi)^{\frac{p-2}{2}}A(z)\xi$, where k (resp. A) is a real-valued (resp. symmetric matrix-valued) bounded map satisfying a uniform coercivity assumption. We also refer to [3] for other examples.

Finally, as we have seen above we want to consider a flux φ which is piecewise smooth with respect to the space variable. The precise meaning of this statement is the following:

- If $1 < p \leq 2$: for all $\xi \in \mathbb{R}^2$, for all $\mathcal{Q} \in \mathfrak{Q}$ and almost every $(z,z') \in \mathcal{Q}^2$,

$$|\varphi(z,\xi) - \varphi(z',\xi)| \leq C_\varphi(1+|\xi|^{p-1})|z-z'|^{p-1}. \qquad (\mathcal{H}_{5a})$$

- If $p > 2$: φ is Lipschitz on any $\mathcal{Q} \in \mathfrak{Q}$, and for all $\xi \in \mathbb{R}^2$ and almost every $z \in \mathcal{Q}$ we have

$$\left|\frac{\partial\varphi}{\partial z}(z,\xi)\right| \leq C_\varphi\left(1+|\xi|^{p-1}\right). \qquad (\mathcal{H}_{5b})$$

Contrarily to the assumptions we considered in [3], the above hypothesis are localized on each quarter diamond. From a practical point of view, this means that the mesh is built is such a way that the discontinuities with respect to the space variable z of the flux φ are only allowed across edges of the primal mesh and edges of the dual mesh.

4.2. Approximate fluxes on the quarter diamond. From now on, we assume that φ is a given flux satisfying $(\mathcal{H}_2)$,$(\mathcal{H}_3)$ and either $(\mathcal{H}_{1'a})$, $(\mathcal{H}_{4a})$, $(\mathcal{H}_{5a})$ if $p \leq 2$, either $(\mathcal{H}_{1'b})$, $(\mathcal{H}_{4b})$, $(\mathcal{H}_{5b})$ if $p > 2$.

Then, we suppose given for each quarter-diamond $\mathcal{Q} \in \mathfrak{Q}$ a probability measure $d\mu_{\overline{\mathcal{Q}}}$ on $\overline{\mathcal{Q}}$, so that we can define an approximation $\varphi_{\mathcal{Q}}(\,\cdot\,)$ of φ on $\mathcal{Q}$ by

$$\varphi_{\mathcal{Q}}(\,\cdot\,) = \int_{\overline{\mathcal{Q}}} \varphi(z,\,\cdot\,)\, d\mu_{\overline{\mathcal{Q}}}(z). \qquad (4.1)$$

This makes sense since φ is supposed to be Hölder continuous on $\mathcal{Q}$ (see assumptions $(\mathcal{H}_{5a})$-$(\mathcal{H}_{5b})$ above) and hence can be extended to a continuous map on $\overline{\mathcal{Q}}$. This quite general framework includes the case where $\varphi_{\mathcal{Q}}$ is the usual mean-value of φ on $\mathcal{Q}$ for the Lebesgue measure but also the case where $\varphi_{\mathcal{Q}}$ is chosen to be the value of φ at a given point in $\overline{\mathcal{Q}}$ or more generally an approximation of the mean-value of φ through a quadrature formula. These situations are the usual ones that we may use in practice.

Remark now that $\varphi_{\mathcal{Q}}$ inherits the monotonicity, coercivity and regularity properties of the initial flux φ, that is for any $\mathcal{Q} \in \mathfrak{Q}$:

$$(\varphi_{\mathcal{Q}}(\xi),\xi) \geq \frac{1}{C_\varphi}|\xi|^p - C_\varphi, \quad \forall \xi \in \mathbb{R}^2, \qquad (\mathcal{H}_2^\tau)$$

$$|\varphi_{\mathcal{Q}}(\xi)| \leq C_\varphi(1+|\xi|^{p-1}), \quad \forall \xi \in \mathbb{R}^2. \qquad (\mathcal{H}_3^\tau)$$

- If $1 < p \leq 2$:

$$(\varphi_{\mathcal{Q}}(\xi) - \varphi_{\mathcal{Q}}(\eta), \xi-\eta) \geq \frac{1}{C_\varphi}|\xi-\eta|^2(1+|\xi|^p+|\eta|^p)^{\frac{p-2}{p}}, \quad \forall \xi,\eta \in \mathbb{R}^2, \ (\mathcal{H}_{1'a}^\tau)$$

$$|\varphi_{\mathcal{Q}}(\xi) - \varphi_{\mathcal{Q}}(\eta)| \leq C_\varphi|\xi-\eta|^{p-1}, \quad \forall \xi,\eta \in \mathbb{R}^2. \qquad (\mathcal{H}_{4a}^\tau)$$

- If $p > 2$:

$$(\varphi_{\mathcal{Q}}(\xi) - \varphi_{\mathcal{Q}}(\eta), \xi - \eta) \geq \frac{1}{C_\varphi} |\xi - \eta|^p, \quad \forall \xi, \eta \in \mathbb{R}^2, \tag{$\mathcal{H}^{\mathcal{T}}_{1'b}$}$$

$$|\varphi_{\mathcal{Q}}(\xi) - \varphi_{\mathcal{Q}}(\eta)| \leq C_\varphi \left(1 + |\xi|^{p-2} + |\eta|^{p-2}\right) |\xi - \eta|, \quad \forall \xi, \eta \in \mathbb{R}^2. \tag{$\mathcal{H}^{\mathcal{T}}_{4b}$}$$

4.3. Local modification of the discrete gradient operator. As we saw in the 1D case (see Section 2), we need to introduce new gradient operators near the discontinuities of the flux and finally define a new approximate flux $\varphi^{\mathcal{N}}_{\mathcal{D}}$ on each diamond cell.

The new gradient operator $\nabla^{\mathcal{N}}$ we propose to consider is built upon the usual DDFV gradient $\nabla^{\mathcal{T}}$. It is chosen to be constant on all the quarter diamonds $\mathcal{Q} \in \mathfrak{Q}$. This new operator has to be thought as the 2D generalization of the definitions of ∇^+_N and ∇^-_N in (2.4). In this 1D situation, the place where the artificial unknown $\tilde{u}$ (or $\delta = \tilde{u} - \bar{u}$) must be chosen is clear: it is the point of the mesh where the discontinuity takes place, that is $x_N = 0$ in the framework of Section 2.

In the 2D setting the situation is less straightforward. In order to make the good choice it is useful to remember that the usual DDFV gradient $\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}}$ on a diamond cell $\mathcal{D}$ can be understood (it is an easy computation) as the gradient of the unique affine function $\Pi_{\mathcal{D}} u^{\mathcal{T}}$ whose value at the middle of each side of the diamond $\mathcal{D}$ is the mean-value between the two unknowns associated to the two extremities of this segment. This situation is summed up in Figure 4.1 for a given interior diamond cell $\mathcal{D} \in \mathfrak{D}_{\text{int}}$. In this figure, we introduce x_σ to be the middle of the segment σ for each $\sigma \in \{\sigma_{\mathcal{K}}, \sigma_{\mathcal{L}}, \sigma_{\mathcal{K}^*}, \sigma_{\mathcal{L}^*}\}$.

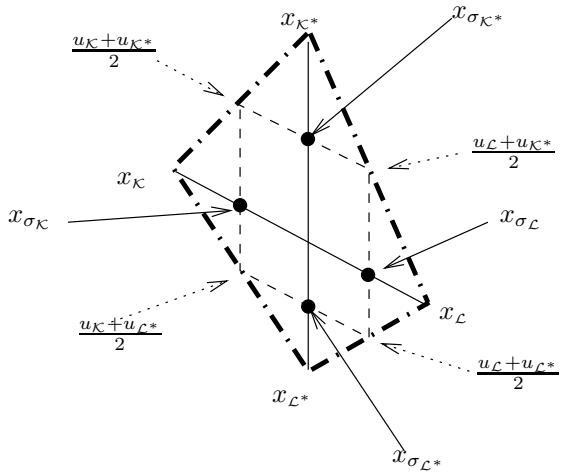


FIG. 4.1. *Affine function whose gradient is* $\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}}$

It seems now natural to define the new discrete gradient operator $\nabla^{\mathcal{N}}_{\mathcal{Q}} u^{\mathcal{T}}$ on each quarter diamond as the gradient of a function $\widetilde{\Pi}_{\mathcal{D}} u^{\mathcal{T}}$ which coincides with $\Pi_{\mathcal{D}} u^{\mathcal{T}}$ in the middle of each side of $\mathcal{D}$ and which is continuous at each point $x_{\sigma_{\mathcal{K}}}$, $x_{\sigma_{\mathcal{L}}}$, $x_{\sigma_{\mathcal{K}^*}}$ and $x_{\sigma_{\mathcal{L}^*}}$ but which is not necessarily continuous on the whole diamond $\mathcal{D}$.

This new function $\widetilde{\Pi}_{\mathcal{D}} u^{\mathcal{T}}$ is then entirely defined, for a given $u^{\mathcal{T}}$, by its four values $\widetilde{\Pi}_{\mathcal{D}} u^{\mathcal{T}}(x_\sigma)$ at each of these four points x_σ. These four values are the artificial unknowns

in our problem. Like in the 1D case, it is equivalent and more suitable to work with the new unknowns $\delta_{\mathcal{K}}^{\mathcal{D}}, \delta_{\mathcal{L}}^{\mathcal{D}}, \delta_{\mathcal{K}^*}^{\mathcal{D}}, \delta_{\mathcal{L}^*}^{\mathcal{D}}$ defined to be the differences $\widetilde{\Pi}_{\mathcal{D}} u^{\mathcal{T}}(x_\sigma) - \Pi_{\mathcal{D}}(x_\sigma)$ for each $\sigma \in \{\sigma_{\mathcal{K}}, \sigma_{\mathcal{L}}, \sigma_{\mathcal{K}^*}, \sigma_{\mathcal{L}^*}\}$. Notice that each $\Pi_{\mathcal{D}}(x_\sigma)$ can be computed as an explicit function of $u_{\mathcal{K}}, u_{\mathcal{L}}, u_{\mathcal{K}^*}$ and $u_{\mathcal{L}^*}$.

The situation is simpler in the case of exterior diamond cells $\mathcal{D} \in \mathfrak{D}_{\text{ext}}$, in which case we only need one artificial unknown, that is $\delta_{\mathcal{K}}^{\mathcal{D}}$. Hence, we define $n_{\mathcal{D}}$ to be the number of artificial unknowns needed on the diamond cell $\mathcal{D}$. From the discussion above we have $n_{\mathcal{D}} = 4$ if $\mathcal{D} \in \mathfrak{D}_{\text{int}}$ and $n_{\mathcal{D}} = 1$ if $\mathcal{D} \in \mathfrak{D}_{\text{ext}}$.

By straightforward computations, the above discussion can be summed up as follows: we define the new discrete gradient operator on each quarter diamond $\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}$, to be the gradient of $\widetilde{\Pi}_{\mathcal{D}} u^{\mathcal{T}}$, which reads

$$\nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}} = \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}} + B_{\mathcal{Q}} \delta^{\mathcal{D}},$$

where $\delta^{\mathcal{D}} \in \mathbb{R}^{n_{\mathcal{D}}}$ is an artificial set of unknowns introduced above and $(B_{\mathcal{Q}})_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}}$ is a set of $2 \times n_{\mathcal{D}}$ matrices defined as follows:

In the case where $\mathcal{D} \in \mathfrak{D}_{\text{int}}$, we take the four matrices $B_{\mathcal{Q}}$:

$$B_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} = \frac{2}{\sin\alpha_{\mathcal{D}}} \left(\frac{\boldsymbol{\nu}^*}{|\sigma_{\mathcal{K}^*}|}, 0, \frac{\boldsymbol{\nu}}{|\sigma_{\mathcal{K}}|}, 0 \right) = \frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}|} \left(|\sigma_{\mathcal{K}}| \boldsymbol{\nu}^*, 0, |\sigma_{\mathcal{K}^*}| \boldsymbol{\nu}, 0 \right), \tag{4.2}$$

$$B_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}} = \frac{2}{\sin\alpha_{\mathcal{D}}} \left(-\frac{\boldsymbol{\nu}^*}{|\sigma_{\mathcal{L}^*}|}, 0, 0, \frac{\boldsymbol{\nu}}{|\sigma_{\mathcal{K}}|} \right) = \frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}|} \left(-|\sigma_{\mathcal{K}}| \boldsymbol{\nu}^*, 0, 0, |\sigma_{\mathcal{L}^*}| \boldsymbol{\nu} \right), \tag{4.3}$$

$$B_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}} = \frac{2}{\sin\alpha_{\mathcal{D}}} \left(0, -\frac{\boldsymbol{\nu}^*}{|\sigma_{\mathcal{L}^*}|}, 0, -\frac{\boldsymbol{\nu}}{|\sigma_{\mathcal{L}}|} \right) = \frac{1}{|\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}|} \left(0, -|\sigma_{\mathcal{L}}| \boldsymbol{\nu}^*, 0, -|\sigma_{\mathcal{L}^*}| \boldsymbol{\nu} \right), \tag{4.4}$$

$$B_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}} = \frac{2}{\sin\alpha_{\mathcal{D}}} \left(0, \frac{\boldsymbol{\nu}^*}{|\sigma_{\mathcal{K}^*}|}, -\frac{\boldsymbol{\nu}}{|\sigma_{\mathcal{L}}|}, 0 \right) = \frac{1}{|\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}|} \left(0, |\sigma_{\mathcal{L}}| \boldsymbol{\nu}^*, -|\sigma_{\mathcal{K}^*}| \boldsymbol{\nu}, 0 \right). \tag{4.5}$$

In the case where $\mathcal{D} \in \mathfrak{D}_{\text{ext}}$, there is only two non-degenerate quarter-diamonds in $\mathcal{Q}$ and the two corresponding matrices $B_{\mathcal{Q}}$ are given by

$$B_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} = \frac{2}{\sin\alpha_{\mathcal{D}}} \left(\frac{\boldsymbol{\nu}^*}{|\sigma_{\mathcal{K}^*}|} \right) = \frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}|} \left(|\sigma_{\mathcal{K}}| \boldsymbol{\nu}^* \right), \tag{4.6}$$

$$B_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}} = \frac{2}{\sin\alpha_{\mathcal{D}}} \left(-\frac{\boldsymbol{\nu}^*}{|\sigma_{\mathcal{L}^*}|} \right) = \frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}|} \left(-|\sigma_{\mathcal{K}}| \boldsymbol{\nu}^* \right). \tag{4.7}$$

Notice that these matrices only depend on the geometry of the diamond cell $\mathcal{D}$. Furthermore we easily see from the formulas above that $\sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| B_{\mathcal{Q}} = 0$ for any diamond cell $\mathcal{D}$. Hence, the following straightforward result holds

LEMMA 4.1. *For all $\xi \in \mathbb{R}^2$, for all $\mathcal{D} \in \mathfrak{D}$, for all $\delta \in \mathbb{R}^{n_{\mathcal{D}}}$, we have*

$$\xi = \frac{1}{|\mathcal{D}|} \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \left(\xi + B_{\mathcal{Q}} \delta \right). \tag{4.8}$$

This Lemma implies that the new gradient has a mean value over $\mathcal{D}$ which equals the usual DDFV gradient $\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}$, that is

$$\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}} = \frac{1}{|\mathcal{D}|} \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}},$$

which is the 2D equivalent to formula (2.5).

Like in the monodimensional case presented in Section 2, we want to eliminate the additional unknowns $\delta^{\mathcal{D}}$ on each $\mathcal{D}$ in such a way that the conservativity of the numerical fluxes on all edges $\sigma \in \mathcal{E}_{\mathcal{D}}$ is ensured. More precisely, we want to choose $\delta^{\mathcal{D}}$ such that, setting $\xi = \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}$, we have

$$\left\{\begin{array}{l} \left(\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}^*\right) = \left(\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}^*\right), \\ \left(\varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}^*\right) = \left(\varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}^*\right), \\ \left(\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}\right) = \left(\varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}\right), \\ \left(\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}\right) = \left(\varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}\right), \end{array}\right. \tag{4.9}$$

in the case where $\mathcal{D} \in \mathfrak{D}_{\rm int}$. If $\mathcal{D} \in \mathfrak{D}_{\rm ext}$, the conditions on $\delta^{\mathcal{D}}$ takes the simpler form

$$\left(\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}^*\right) = \left(\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}(\xi + B_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}\delta^{\mathcal{D}}), \boldsymbol{\nu}^*\right). \tag{4.10}$$

We are now going to show that the equations (4.9) or (4.10) uniquely defines $\delta^{\mathcal{D}} \in \mathbb{R}^{\mathcal{D}}$ as a function of ξ.

PROPOSITION 4.2. *For all $\mathcal{D} \in \mathfrak{D}$ and all $\xi \in \mathbb{R}^2$, there exists a unique $\delta^{\mathcal{D}}(\xi) \in \mathbb{R}^{n_{\mathcal{D}}}$ such that* (4.9) *(resp.* (4.10)*) holds if $\mathcal{D} \in \mathfrak{D}_{\rm int}$ (resp. if $\mathcal{D} \in \mathfrak{D}_{\rm ext}$).*

Proof. We only give the proof for $\mathcal{D} \in \mathfrak{D}_{\rm int}$, since the case of boundary diamond cells can be treated in the same way. Let $\xi \in \mathbb{R}^2$ given and define $F_\xi : \mathbb{R}^4 \mapsto \mathbb{R}^4$ by

$$F_\xi(\delta) = \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}}.\varphi_{\mathcal{Q}}(\xi + B_{\mathcal{Q}}\delta).$$

By using (4.2)-(4.5), we easily see that the conditions (4.9) are equivalent to the equation $F_\xi(\delta^{\mathcal{D}}) = 0$. Hence, the claim will be proved if we show that this nonlinear equation has a unique solution. To this end, we remark that for any $\widetilde{\delta} \in \mathbb{R}^4$, we have

$$\left(F_\xi(\delta), \tilde{\delta}\right) = \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \left(\varphi_{\mathcal{Q}}(\xi + B_{\mathcal{Q}}\delta), B_{\mathcal{Q}}\widetilde{\delta}\right). \tag{4.11}$$

Hence, we deduce using assumptions $(\mathcal{H}_2^{\mathcal{T}})$ and $(\mathcal{H}_3^{\mathcal{T}})$ that there exists C depending only on p and C_φ such that

$$(F_\xi(\delta), \delta) \geq \frac{1}{C}\left(\sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|\,|\xi + B_{\mathcal{Q}}\delta|^p\right) - C|\mathcal{D}|(1 + |\xi|^p). \tag{4.12}$$

Finally, we deduce

$$(F_\xi(\delta), \delta) \geq \frac{1}{C}\left(\sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|\,|B_{\mathcal{Q}}\delta|^p\right) - C|\mathcal{D}|(1 + |\xi|^p),$$

for another constant C. Since, $\sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}||B_{\mathcal{Q}}\delta|^p \underset{|\delta| \to \infty}{\longrightarrow} \infty$, we deduce that F_ξ is coercive. By the Brouwer theorem (F_ξ is continuous since each $\varphi_{\mathcal{Q}}$ is continuous) we obtain the existence of at least one solution to the problem $F_\xi(\delta) = 0$.

Notice now that if $B_{\mathcal{Q}}\delta = B_{\mathcal{Q}}\widetilde{\delta}, \forall \mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}$ then $\delta = \widetilde{\delta}$. Hence, we deduce from (4.11) that for all $\delta \neq \widetilde{\delta}$

$$\left(F_\xi(\delta) - F_\xi(\widetilde{\delta}), \delta - \widetilde{\delta}\right) = \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \left(\varphi_{\mathcal{Q}}(\xi + B_{\mathcal{Q}}\delta) - \varphi_{\mathcal{Q}}(\xi + B_{\mathcal{Q}}\widetilde{\delta}), B_{\mathcal{Q}}\delta - B_{\mathcal{Q}}\widetilde{\delta}\right) > 0,$$

using assumption $(\mathcal{H}^{\mathcal{T}}_{1'a})$ or $(\mathcal{H}^{\mathcal{T}}_{1'b})$. This gives the uniqueness of the solution to $F_\xi(\delta) = 0$ and the claim is proved. □

EXAMPLE 4.3. *In many situations, it can happen that φ is smooth for instance in each primal control volume. In that case, it is possible to choose the approximations $\varphi_{\mathcal{Q}}$ in such a way that for each $\mathcal{D} \in \mathfrak{D}_{\rm int}$ we have*

$$\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} = \varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}, \quad \text{and } \varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}} = \varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}.$$

In that case, one can easily show that the solution $\delta^{\mathcal{D}}(\xi)$ of the equations (4.9) *appears to satisfy*

$$\delta^{\mathcal{D}}_{\mathcal{K}} = 0, \quad \delta^{\mathcal{D}}_{\mathcal{L}} = 0, \quad \text{and } \delta^{\mathcal{D}}_{\mathcal{K}^*} = \delta^{\mathcal{D}}_{\mathcal{L}^*}.$$

Hence, everything happens like in the 1D case and there is in fact only one artificial unknown ($\delta^{\mathcal{D}}_{\mathcal{K}^*}$ for instance) to determine.

A symmetric situation holds if we assume that φ is smooth in each dual control volume and that $\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} = \varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}$, $\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}} = \varphi_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}$.

Finally, notice that if we only assume that $\varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} = \varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}$ for instance, then in the general nonlinear case there is no reason why $\delta^{\mathcal{D}}_{\mathcal{K}}$ should be 0.

From now on, the new discrete gradient operator $\nabla^{\mathcal{N}}$ is completely determined by

$$\nabla^{\mathcal{N}}_{\mathcal{Q}} u^{\mathcal{T}} = \nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}} + B_{\mathcal{Q}} \delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}}), \text{ for any } \mathcal{D} \in \mathfrak{D} \text{ and } \mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}, \tag{4.13}$$

where the map $\xi \mapsto \delta^{\mathcal{D}}(\xi)$ is defined in Proposition 4.2. Notice that, in general, this new gradient operator $\nabla^{\mathcal{N}}_{\mathcal{Q}}$ is *nonlinear* contrarily to the operator $\nabla^{\mathcal{T}}_{\mathcal{D}}$ and depends on the flux φ defining the equation (and more precisely to its approximations $\varphi_{\mathcal{Q}}$).

Furthermore, let us emphasize the fact that the nonlinear map $\xi \mapsto \delta^{\mathcal{D}}(\xi)$ is only defined implicitely through the equations (4.9) or (4.10) (see also the 1D discussion in Section 2), which seems to make the new discrete gradient $\nabla^{\mathcal{N}}$ quite difficult to compute. We postpone to Section 7 the discussion on the practical way to solve this finite volume scheme.

4.4. Some useful inequalities. The usual DDFV discrete gradient and the modified one can be compared as follows.

LEMMA 4.4. *There exists a constant C that depends only on C_φ and p such that for all $\mathcal{D} \in \mathfrak{D}$, and all $u^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$, we have*

$$\int_{\mathcal{D}} |\nabla^{\mathcal{T}} u^{\mathcal{T}}(z)|^p \, dz \le \int_{\mathcal{D}} |\nabla^{\mathcal{N}} u^{\mathcal{T}}(z)|^p \, dz \le C \int_{\mathcal{D}} (1 + |\nabla^{\mathcal{T}} u^{\mathcal{T}}(z)|^p) \, dz.$$

Proof. Thanks to Lemma 4.1, we have

$$|\mathcal{D}||\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}}|^p = |\mathcal{D}| \left| \frac{1}{|\mathcal{D}|} \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \nabla^{\mathcal{N}}_{\mathcal{Q}} u^{\mathcal{T}} \right|^p = |\mathcal{D}| \left| \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} \nabla^{\mathcal{N}} u^{\mathcal{T}}(z) \, dz \right|^p.$$

Using the Jensen inequality, we deduce the first inequality.

The second one is a consequence of (4.12) applied to $\xi = \nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}}$, $\delta = \delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}})$ using that $F_{\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}}}(\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}} u^{\mathcal{T}})) = 0$ by definition of $\delta^{\mathcal{D}}$. □

Finally we can state the following discrete Poincaré inequality. Its proof is sketched in [3] and uses an argument given in [4] (see also [9]).

PROPOSITION 4.5 (Discrete Poincaré inequality). *Let $\mathcal{T}$ be a DDFV mesh on Ω. There exists a constant C depending on p, Ω and* reg($\mathcal{T}$) *such that*

$$\|u^{\mathfrak{m}}\|_{L^p} + \|u^{\mathfrak{m}^*}\|_{L^p} \le C \|\nabla^{\mathcal{T}} u^{\mathcal{T}}\|_{L^p} \le C \|\nabla^{\mathcal{N}} u^{\mathcal{T}}\|_{L^p}, \quad \forall u^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}.$$

In particular, if $\nabla^{\mathcal{N}}_{\mathcal{Q}} u^{\mathcal{T}} = 0$ for all $\mathcal{Q} \in \mathfrak{Q}$ then $u^{\mathcal{T}} = 0$.

4.5. The m-DDFV scheme for discontinuous fluxes. We can finally introduce the new approximate flux $\varphi_{\mathcal{D}}^{\mathcal{N}}$ on each diamond cell to be used in the finite volume scheme instead of (3.2). For any diamond cell $\mathcal{D} \in \mathfrak{D}$, we set

$$\varphi_{\mathcal{D}}^{\mathcal{N}}(\xi) = \frac{1}{|\mathcal{D}|} \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \varphi_{\mathcal{Q}}(\xi + B_{\mathcal{Q}} \delta^{\mathcal{D}}(\xi)), \quad \forall \xi \in \mathbb{R}^2. \tag{4.14}$$

This definition is nothing but the adaptation to the 2D case of the corresponding formula in the 1D case (see (2.7)).

Applying definition (4.14) to $\xi = \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}$ for a given $u^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$, we find that

$$\varphi_{\mathcal{D}}^{\mathcal{N}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}) = \frac{1}{|\mathcal{D}|} \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \varphi_{\mathcal{Q}}(\nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}}). \tag{4.15}$$

The m-DDFV scheme that we will study in the sequel of the paper can now be defined by the set of equations

$$\left\{ \begin{array}{l} \mathfrak{a}_{\mathcal{K}}^{\mathcal{N}}(u^{\mathcal{T}}) \stackrel{\text{def}}{=} - \displaystyle\sum_{\mathcal{D} \in \mathfrak{D}_{\mathcal{K}}} |\sigma| \left(\varphi_{\mathcal{D}}^{\mathcal{N}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}), \boldsymbol{\nu}_{\mathcal{K}} \right) = |\mathcal{K}| f_{\mathcal{K}}, \ \forall \mathcal{K} \in \mathfrak{M}, \\ \mathfrak{a}_{\mathcal{K}^*}^{\mathcal{N}}(u^{\mathcal{T}}) \stackrel{\text{def}}{=} - \displaystyle\sum_{\mathcal{D} \in \mathfrak{D}_{\mathcal{K}^*}} |\sigma^*| \left(\varphi_{\mathcal{D}}^{\mathcal{N}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}), \boldsymbol{\nu}_{\mathcal{K}^*} \right) = |\mathcal{K}^*| f_{\mathcal{K}^*}, \ \forall \mathcal{K}^* \in \mathfrak{M}^*, \end{array} \right. \tag{4.16}$$

that we write under the short form $\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}) = ((|\mathcal{K}| f_{\mathcal{K}})_{\mathcal{K}}, (|\mathcal{K}^*| f_{\mathcal{K}^*})_{\mathcal{K}^*})$, with $\mathfrak{a}^{\mathcal{N}}(\,\cdot\,) \stackrel{\text{def}}{=} ((\mathfrak{a}_{\mathcal{K}}^{\mathcal{N}}(\,\cdot\,))_{\mathcal{K}}, (\mathfrak{a}_{\mathcal{K}^*}^{\mathcal{N}}(\,\cdot\,))_{\mathcal{K}^*})$. Note that the only difference between this scheme (4.16) and the previous one (3.1) is the fact that we replaced the previous mean-value approximation $\varphi_{\mathcal{D}}$ of φ over $\mathcal{D}$ by the map $\varphi_{\mathcal{D}}^{\mathcal{N}}$ defined by (4.14).

4.6. Basic properties of the scheme. Before proving existence and uniqueness of the solution for the nonlinear system (4.16) we first give some properties of the map $\mathfrak{a}^{\mathcal{N}}$. The first result is a *summation by parts* result showing that the m-DDFV scheme we propose enjoys the same discrete duality property than the original DDFV scheme.

LEMMA 4.6. *For any $u^{\mathcal{T}}$, $v^{\mathcal{T}}$ in $\mathbb{R}^{\mathcal{T}}$ we have*

$$\begin{aligned} (\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}), v^{\mathcal{T}}) &= 2 \sum_{\mathcal{D} \in \mathfrak{D}} |\mathcal{D}| \left(\varphi_{\mathcal{D}}^{\mathcal{N}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}), \nabla_{\mathcal{D}}^{\mathcal{T}} v^{\mathcal{T}} \right) \\ &= 2 \sum_{\mathcal{Q} \in \mathfrak{Q}} |\mathcal{Q}| \left(\varphi_{\mathcal{Q}}(\nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}}), \nabla_{\mathcal{Q}}^{\mathcal{N}} v^{\mathcal{T}} \right). \end{aligned}$$

Proof. The first equality can be proved in the same way than in [3] by reordering the summation on the primal and dual control volumes as a summation over the

diamond set

$$\begin{aligned}
&\sum_{\mathcal{K}\in\mathfrak{M}} \mathfrak{a}^{\mathcal{N}}_{\mathcal{K}}(u^{\mathcal{T}})v_{\mathcal{K}} + \sum_{\mathcal{K}^*\in\mathfrak{M}^*} \mathfrak{a}^{\mathcal{N}}_{\mathcal{K}^*}(u^{\mathcal{T}})v_{\mathcal{K}^*} \\
&\quad = -\sum_{\mathcal{K}\in\mathfrak{M}}\sum_{\mathcal{D}\in\mathfrak{D}_{\mathcal{K}}} |\sigma|(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}),\boldsymbol{\nu})v_{\mathcal{K}} - \sum_{\mathcal{K}^*\in\mathfrak{M}^*}\sum_{\mathcal{D}\in\mathfrak{D}_{\mathcal{K}^*}} |\sigma^*|(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}),\boldsymbol{\nu}^*)v_{\mathcal{K}^*} \\
&\quad = \sum_{\mathcal{D}\in\mathfrak{D}_{\rm int}} |\mathcal{D}| \left(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}), \frac{2}{\sin\alpha_{\mathcal{D}}}\left(\frac{v_{\mathcal{K}}-v_{\mathcal{L}}}{|\sigma^*|}\boldsymbol{\nu} + \frac{v_{\mathcal{K}^*}-v_{\mathcal{L}^*}}{|\sigma|}\boldsymbol{\nu}^*\right)\right) \\
&\qquad + \sum_{\mathcal{D}\in\mathfrak{D}_{\rm ext}} |\mathcal{D}| \left(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}), \frac{2}{\sin\alpha_{\mathcal{D}}}\frac{v_{\mathcal{K}}}{|\sigma^*|}\boldsymbol{\nu}\right) \\
&\quad = 2\sum_{\mathcal{D}\in\mathfrak{D}} |\mathcal{D}|\left(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}), \nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}}\right).
\end{aligned}$$

To prove the second equality, we use (4.15) and (4.13) to write on each diamond cell $\mathcal{D}\in\mathfrak{D}$

$$\begin{aligned}
|\mathcal{D}|\left(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}), \nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}}\right) &= \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|\left(\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}}), \nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}}\right) \\
&= \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|\left(\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}}), \nabla^{\mathcal{N}}_{\mathcal{Q}}v^{\mathcal{T}} - B_{\mathcal{Q}}\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}})\right) \\
&= \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|\left(\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}}), \nabla^{\mathcal{N}}_{\mathcal{Q}}v^{\mathcal{T}}\right) \\
&\qquad - \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}}\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}})\cdot\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}}),
\end{aligned}$$

and this last term vanishes since, by definition of the map $\delta^{\mathcal{D}}$ (see proposition 4.2), we have

$$\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}}\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}})\cdot\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}}) = (F_{\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}}(\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}})), \delta^{\mathcal{D}}(v^{\mathcal{T}})) = 0.$$

□

LEMMA 4.7. *For any $u^{\mathcal{T}}\in\mathbb{R}^{\mathcal{T}}$ we have*

$$(\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}), u^{\mathcal{T}}) \geq \frac{2}{C_\varphi}\|\nabla^{\mathcal{N}}u^{\mathcal{T}}\|^p_{L^p} - 2C_\varphi|\Omega| \geq \frac{2}{C_\varphi}\|\nabla^{\mathcal{T}}u^{\mathcal{T}}\|^p_{L^p} - 2C_\varphi|\Omega|.$$

Proof. We derive from Lemma 4.6 and assumption $(\mathcal{H}^{\mathcal{T}}_2)$ that

$$\begin{aligned}
(\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}), u^{\mathcal{T}}) &= 2\sum_{\mathcal{D}\in\mathfrak{D}}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|\left(\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}}), \nabla^{\mathcal{N}}_{\mathcal{D}}u^{\mathcal{T}}\right) \\
&\geq 2\sum_{\mathcal{D}\in\mathfrak{D}}\left(\int_{\mathcal{D}} \frac{1}{C_\varphi}|\nabla^{\mathcal{N}}u^{\mathcal{T}}(z)|^p\,dz - C_\varphi|\mathcal{D}|\right).
\end{aligned}$$

We conclude using Lemma 4.4. □

LEMMA 4.8. *We have*

$$(\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}) - \mathfrak{a}^{\mathcal{N}}(v^{\mathcal{T}}), u^{\mathcal{T}} - v^{\mathcal{T}}) > 0, \quad \forall u^{\mathcal{T}}, v^{\mathcal{T}}\in\mathbb{R}^{\mathcal{T}},\ u^{\mathcal{T}}\neq v^{\mathcal{T}}.$$

Proof. From Lemma 4.6, we have

$$(\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}) - \mathfrak{a}^{\mathcal{N}}(v^{\mathcal{T}}), u^{\mathcal{T}} - v^{\mathcal{T}}) \\ = 2 \sum_{\mathcal{Q} \in \mathfrak{Q}} |\mathcal{Q}| \left(\varphi_{\mathcal{Q}}(\nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}}) - \varphi_{\mathcal{Q}}(\nabla_{\mathcal{Q}}^{\mathcal{N}} v^{\mathcal{T}}), \nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}} - \nabla_{\mathcal{Q}}^{\mathcal{N}} v^{\mathcal{T}}\right). \quad (4.17)$$

By using the monotonicity properties of the nonlinearity $\varphi_{\mathcal{Q}}$, that is assumption $(\mathcal{H}_{1'a}^{\mathcal{T}})$ and $(\mathcal{H}_{1'b}^{\mathcal{T}})$, we deduce that the left hand side of (4.17) is non negative and vanishes if and only if $u^{\mathcal{T}} = v^{\mathcal{T}}$ (by Proposition 4.5).
□

THEOREM 4.9. *The scheme* (4.16) *admits a unique solution* $u^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$. *Furthermore, there exists a constant* C *depending only on* C_φ, $\mathrm{reg}(\mathcal{T})$ *and* p *such that this solution* $u^{\mathcal{T}}$ *satisfies*

$$\|\nabla^{\mathcal{T}} u^{\mathcal{T}}\|_{L^p} + \|\nabla^{\mathcal{N}} u^{\mathcal{T}}\|_{L^p} \leq C \left(1 + \|f\|_{L^{p'}}^{\frac{1}{p-1}}\right). \quad (4.18)$$

Proof. The map $u^{\mathcal{T}} \mapsto \mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}) - f^{\mathcal{T}}$ is continuous and coercive thanks to Lemma 4.7. We deduce from the Brouwer theorem the existence of a solution to (4.16). Uniqueness is a consequence of the monotonicity Lemma 4.8.

Estimate (4.18) comes directly from Lemmas 4.4 and 4.7. □

We finally show that the numerical solution of the scheme (4.16) depends continuously on the source term $f^{\mathcal{T}}$.

THEOREM 4.10 (Stability). *There exist a constant* $C > 0$ *depending only on* C_φ, $\mathrm{reg}(\mathcal{T})$ *and* p *such that for any* $f^{\mathcal{T}}$ *and* $g^{\mathcal{T}}$ *in* $\mathbb{R}^{\mathcal{T}}$, *we have*

$$\|\nabla^{\mathcal{N}} u^{\mathcal{T}} - \nabla^{\mathcal{N}} v^{\mathcal{T}}\|_{L^p} \leq \begin{cases} C \left(1 + \|f^{\mathcal{T}}\|_{L^{p'}} + \|g^{\mathcal{T}}\|_{L^{p'}}\right)^{\frac{2-p}{p-1}} \|f^{\mathcal{T}} - g^{\mathcal{T}}\|_{L^{p'}}, & \text{if } 1 < p \leq 2 \\ C \|f^{\mathcal{T}} - g^{\mathcal{T}}\|_{L^{p'}}^{\frac{1}{p-1}}, & \text{if } p > 2, \end{cases}$$

where $u^{\mathcal{T}}$ *(resp.* $v^{\mathcal{T}}$*) is the solution of the m-DDFV scheme* (4.16) *associated to the data* $f^{\mathcal{T}}$ *(resp.* $g^{\mathcal{T}}$*).*

Proof. We apply estimate (4.17) and obtain thanks to assumption $(\mathcal{H}_{1'b}^{\mathcal{T}})$ for $p > 2$

$$(\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}) - \mathfrak{a}^{\mathcal{N}}(v^{\mathcal{T}}), u^{\mathcal{T}} - v^{\mathcal{T}}) \geq \frac{1}{C} \sum_{\mathcal{D} \in \mathfrak{D}} \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| |\nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}} - \nabla_{\mathcal{Q}}^{\mathcal{N}} v^{\mathcal{T}}|^p.$$

For $1 < p \leq 2$ assumption $(\mathcal{H}_{1'a}^{\mathcal{T}})$ implies

$$(\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}) - \mathfrak{a}^{\mathcal{N}}(v^{\mathcal{T}}), u^{\mathcal{T}} - v^{\mathcal{T}}) \geq \\ \frac{1}{C} \sum_{\mathcal{D} \in \mathfrak{D}} \sum_{\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}| \left(1 + |\nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}}|^p + |\nabla_{\mathcal{Q}}^{\mathcal{N}} v^{\mathcal{T}}|^p\right)^{\frac{p-2}{p}} |\nabla_{\mathcal{Q}}^{\mathcal{N}} u^{\mathcal{T}} - \nabla_{\mathcal{Q}}^{\mathcal{N}} v^{\mathcal{T}}|^2.$$

Hence, if $1 < p \leq 2$ we have

$$\|\nabla^{\mathcal{N}} u^{\mathcal{T}} - \nabla^{\mathcal{N}} v^{\mathcal{T}}\|_{L^p}^2 \leq C \left(1 + \|\nabla^{\mathcal{N}} u^{\mathcal{T}}\|^{2-p} + \|\nabla^{\mathcal{N}} v^{\mathcal{T}}\|^{2-p}\right) \\ \times (\mathfrak{a}^{\mathcal{N}}(u^{\mathcal{T}}) - \mathfrak{a}^{\mathcal{N}}(v^{\mathcal{T}}), u^{\mathcal{T}} - v^{\mathcal{T}}).$$

We conclude using the definition of $f^{\mathcal{T}}$ and $g^{\mathcal{T}}$ and the discrete Poincaré inequality (Proposition 4.5). □

5. Error estimates. We first give some error estimates for the scheme (4.16) in the case where the exact solution u_e of the problem (1.1) is piecewise smooth. More precisely let us introduce for any $q \in [1, +\infty]$ the space

$$W^{2,q}(\mathfrak{D}) = \left\{ u \in W_0^{1,p}(\Omega), \quad u_{|\mathcal{Q}} \in W^{2,q}(\mathcal{Q}), \ \forall \mathcal{Q} \in \mathfrak{D} \right\},$$

endowed with the norm

$$\|u\|_{W^{2,q}(\mathfrak{D})} = \|u\|_{W_0^{1,p}(\Omega)} + \left(\sum_{\mathcal{Q} \in \mathfrak{D}} \|D^2 u\|_{L^q(\mathcal{Q})}^q \right)^{\frac{1}{q}}.$$

Our first result gives an error estimate for the solution of the m-DDFV scheme. In this result the flux φ is allowed to have discontinuities across all the edges of the primal and the dual meshes.

THEOREM 5.1. *Let $\mathcal{T}$ be a mesh on Ω. Let $f \in L^{p'}(\Omega)$ and assume that the solution u_e to (1.1) belongs to $W^{2,p}(\mathfrak{D})$.*

There exists $C > 0$ depending on $\|u_e\|_{W^{2,p}(\mathfrak{D})}$, $\mathrm{reg}(\mathcal{T})$, $\|f\|_{L^{p'}}$, C_φ and p such that the solution $u^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$ of the m-DDFV scheme (4.16) satisfies

$$\|u_e - u^{\mathfrak{M}}\|_{L^p} + \|u_e - u^{\mathfrak{M}^*}\|_{L^p} + \|\nabla u_e - \nabla^{\mathcal{N}} u^{\mathcal{T}}\|_{L^p} \leq \begin{cases} C\,\mathrm{size}(\mathcal{T})^{(p-1)^2}, & \text{if } 1 < p \leq 2, \\ C\,\mathrm{size}(\mathcal{T})^{\frac{1}{(p-1)^2}}, & \text{if } p > 2. \end{cases}$$

A typical case for which our method can be applied is the one where the domain Ω can be divided into N disjoint subdomains $(\Omega_i)_{1 \leq i \leq N}$ such that

$$\varphi \text{ is smooth over each subdomain } \Omega_i. \tag{5.1}$$

We assume that each domain Ω_i is polygonal and that the mesh is compatible with the subdomains in the sense that, for any i there exists a subset $\mathcal{E}_i$ of $\mathcal{E}$ such that $\partial\Omega_i = \cup_{\sigma \in \mathcal{E}_i} \sigma$. More generally, we may assume that the discontinuities of the flux φ only occur along a finite number of curves in Ω.

In that situation, the diamond cells naturally divide into two subsets defined by

$$\mathfrak{D}_{\mathrm{cont}} = \{\mathcal{D} \in \mathfrak{D}, \quad \exists i \in \{1, \dots, N\}, \ \mathcal{D} \subset \Omega_i\}, \quad \text{and} \quad \mathfrak{D}_{\mathrm{disc}} = \mathfrak{D} \setminus \mathfrak{D}_{\mathrm{cont}}.$$

We propose to use in that case an hybrid DDFV scheme defined as follows:

DEFINITION 5.2. *Under assumption (5.1), we call h-DDFV scheme for the problem (1.1) the DDFV scheme still under the form (4.16), but where $\varphi_{\mathcal{D}}^{\mathcal{N}}$ is defined as follows:*

- *For the diamond cells $\mathcal{D} \in \mathfrak{D}_{\mathrm{disc}}$, that is the ones where the discontinuities of the flux occur we take $\varphi_{\mathcal{D}}^{\mathcal{N}}$ as defined in (4.15).*
- *For the diamond cells $\mathcal{D} \in \mathfrak{D}_{\mathrm{cont}}$, that is away from the discontinuities, we take $\varphi_{\mathcal{D}}^{\mathcal{N}}$ to be the usual mean-value $\varphi_{\mathcal{D}}$ of φ over $\mathcal{D}$ defined in (3.2).*

Assuming that u_e is slightly more regular than in the previous result, this approach let us recover the same convergence rate than in the usual continuous flux case (see [3]).

THEOREM 5.3. *Consider the same assumptions than in Theorem 5.1 with the additionnal assumption (5.1). Assume furthermore that the solution u_e of (1.1) lies in $W^{2,q}(\mathfrak{D})$ for $q = p(p-1)^2$ if $p \geq 2$ and $q = \frac{p}{(p-1)^2}$ if $p < 2$. Then, there exists a*

constant $C > 0$ like in Theorem 5.1 such that the solution $u^{\mathcal{T}}$ to the h-DDFV scheme satisfies

$$\|u_e - u^{\mathfrak{M}}\|_{L^p} + \|u_e - u^{\mathfrak{M}^*}\|_{L^p} + \|\nabla u_e - \nabla^{\mathcal{N}} u^{\mathcal{T}}\|_{L^p} \leq \begin{cases} C\,\mathrm{size}(\mathcal{T})^{p-1}, & \text{if } 1 < p \leq 2, \\ C\,\mathrm{size}(\mathcal{T})^{\frac{1}{p-1}}, & \text{if } p > 2. \end{cases}$$

We are now going to prove these two results. The key ingredients in this analysis are the consistency error estimates on the numerical fluxes across edges of the primal and the dual control volumes.

5.1. Consistency error. In order to evaluate the error between u_e and $u^{\mathcal{T}}$ we need to introduce a projection of the exact solution u_e onto the space of discrete functions $\mathbb{R}^{\mathcal{T}}$. Notice that any function v in $W^{2,p}(\mathfrak{\Omega})$ is continuous over $\overline{\Omega}$. Hence, it makes sense to consider the center-value projection $\mathbb{P}^{\mathcal{T}}$ defined as follows:

Definition 5.4. *For any $v \in \mathcal{C}^0(\overline{\Omega})$, we define its center-value projection $\mathbb{P}^{\mathcal{T}} v \in \mathbb{R}^{\mathcal{T}}$ as the vector*

$$\mathbb{P}^{\mathcal{T}} v = \Big((v(x_{\mathcal{K}}))_{\mathcal{K} \in \mathfrak{M}}, (v(x_{\mathcal{K}^*}))_{\mathcal{K}^* \in \mathfrak{M}^*} \Big).$$

We refer to [3] for the proofs of the main properties of this projection operator.

As usual in finite volume methods, the error analysis is mainly based on estimates of consistency errors for the fluxes as defined below.

Definition 5.5. *Assume that $u_e \in W^{2,p}(\mathfrak{\Omega})$. For any $\mathcal{Q} \in \mathfrak{\Omega}$, $z \in \overline{\mathcal{Q}}$ we define*

$$\begin{aligned} R_{\mathcal{Q}}(z) &= \varphi_{|\overline{\mathcal{Q}}}(z, \nabla u_{e|\overline{\mathcal{Q}}}(z)) - \varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}} \mathbb{P}^{\mathcal{T}} u_e) \\ &= R^{\varphi}_{\mathcal{Q}}(z) + R^{\mathrm{grad}}_{\mathcal{Q}} + R^{\mathrm{z}}_{\mathcal{Q}}, \end{aligned}$$

with

$$\begin{aligned} R^{\varphi}_{\mathcal{Q}}(z) &\stackrel{\text{def}}{=} \varphi_{|\overline{\mathcal{Q}}}(z, \nabla u_{e|\overline{\mathcal{Q}}}(z)) - \frac{1}{|\mathcal{Q}|} \int_{\mathcal{Q}} \varphi(z', \nabla u_e(z'))\, dz' \\ R^{\mathrm{grad}}_{\mathcal{Q}} &\stackrel{\text{def}}{=} \frac{1}{|\mathcal{Q}|} \int_{\mathcal{Q}} \left(\varphi(z', \nabla u_e(z')) - \varphi(z', \nabla^{\mathcal{N}}_{\mathcal{Q}} \mathbb{P}^{\mathcal{T}} u_e) \right) dz', \\ R^{\mathrm{z}}_{\mathcal{Q}} &\stackrel{\text{def}}{=} \frac{1}{|\mathcal{Q}|} \int_{\mathcal{Q}} \varphi(z', \nabla^{\mathcal{N}}_{\mathcal{Q}} \mathbb{P}^{\mathcal{T}} u_e)\, dz' - \varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}} \mathbb{P}^{\mathcal{T}} u_e). \end{aligned}$$

Finally, for any $\mathcal{Q} \in \mathfrak{\Omega}$ and $\sigma \in \mathcal{E}_{\mathcal{Q}}$, we note

$$R^{\varphi}_{\mathcal{Q},\sigma} = \frac{1}{|\sigma|} \int_{\sigma} (R^{\varphi}_{\mathcal{Q}}(z), \boldsymbol{\nu}_{\mathcal{Q},\sigma})\, dz,\ R^{\mathrm{grad}}_{\mathcal{Q},\sigma} = \left(R^{\mathrm{grad}}_{\mathcal{Q}}, \boldsymbol{\nu}_{\mathcal{Q},\sigma} \right),\ R^{\mathrm{z}}_{\mathcal{Q},\sigma} = \left(R^{\mathrm{z}}_{\mathcal{Q}}, \boldsymbol{\nu}_{\mathcal{Q},\sigma} \right),$$

and $R_{\mathcal{Q},\sigma} = R^{\mathrm{grad}}_{\mathcal{Q},\sigma} + R^{\varphi}_{\mathcal{Q},\sigma} + R^{\mathrm{z}}_{\mathcal{Q},\sigma}$, where $\boldsymbol{\nu}_{\mathcal{Q},\sigma}$ is the unit normal to σ pointing outward $\mathcal{Q}$.

It is fundamental to notice that, by definition of the new discrete gradient operator $\nabla^{\mathcal{N}}$ and since $u_e \in W^{2,p}(\mathfrak{\Omega})$, we have the conservativity property

$$R_{\mathcal{Q},\sigma} = -R_{\mathcal{Q}',\sigma}, \text{ if } \sigma = \mathcal{Q}|\mathcal{Q}'. \tag{5.2}$$

The objective is now to estimate each of the three terms involved in this consistency error. The terms $R^{\varphi}_{\mathcal{Q}}$ and $R^{\mathrm{z}}_{\mathcal{Q}}$ can be easily controlled by using the same techniques as in [3]. This is the aim of the following proposition.

PROPOSITION 5.6. *Let $\mathcal{T}$ be a mesh on Ω and assume that the solution u_e to problem (1.1) lies in $W^{2,p}(\mathfrak{Q})$. There exists a constant $C>0$ depending on p, $\mathrm{reg}(\mathcal{T})$ and C_φ such that*

$$|\mathcal{Q}|\,|R^{\varphi}_{\mathcal{Q},\sigma}|^{\frac{p}{p-1}} \le C\mathrm{d}_{\mathcal{Q}}^{p\alpha_p}\int_{\mathcal{Q}}(1+|\nabla u_e|^p+|D^2u_e|^p)\,dz,\quad \forall\mathcal{Q}\in\mathfrak{Q},\forall\sigma\in\mathcal{E}_{\mathcal{Q}},$$

$$|\mathcal{Q}|\,|R^{\mathrm{z}}_{\mathcal{Q}}|^{\frac{p}{p-1}} \le C\mathrm{d}_{\mathcal{Q}}^{p\alpha_p}|\mathcal{Q}|(1+|\nabla^{\mathcal{N}}_{\mathcal{Q}}\mathbb{P}^{\mathcal{T}}u_e|^p),\quad \forall\mathcal{Q}\in\mathfrak{Q},$$

where $\alpha_p=1$ if $1<p\le 2$ and $\alpha_p=\frac{1}{p-1}$ if $p>2$.

Proof. The proof of the first point is the same as [3, Proposition 7.6].

Note that in the case where the approximate flux $\varphi_{\mathcal{Q}}$ is chosen to be the mean-value of φ on $\mathcal{Q}$, then $R^{\mathrm{z}}_{\mathcal{Q}}$ just vanishes. For other choices of $\varphi_{\mathcal{Q}}$, we write that

$$|R^{\mathrm{z}}_{\mathcal{Q}}|\le\frac{1}{|\mathcal{Q}|}\int_{\mathcal{Q}}\int_{\overline{\mathcal{Q}}}|\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}}\mathbb{P}^{\mathcal{T}}u_e)-\varphi(z',\nabla^{\mathcal{N}}_{\mathcal{Q}}\mathbb{P}^{\mathcal{T}}u_e)|\,dz\,d\mu_{\overline{\mathcal{Q}}}(z')$$

and we use assumption $(\mathcal{H}_{5a})$ or $(\mathcal{H}_{5b})$. The claim follows using Jensen's inequality. □

We can now proceed to the study of the consistency estimate for the new gradient operator $\nabla^{\mathcal{N}}$ that we have introduced. This is the main difference between the present study and our previous works since the definition of the new discrete gradient depends on the jumps of φ in each diamond cell. Hence, the consistency estimate for this operator can not be obtained as in the usual way, that is only by applying well chosen Taylor formulas. The proof of the estimate is much more involved. We also want to point out the fact that, when $p\neq 2$, we do not obtain the usual first order consistency estimate as we obtained in [3] for the operator $\nabla^{\mathcal{T}}_{\mathcal{D}}$. Indeed, due to the degeneracy/singularity of the nonlinear operator near the origin, we only recover a consistency property of order less than one.

PROPOSITION 5.7. *Let $\mathcal{T}$ be a mesh on Ω and assume that the solution u_e to problem (1.1) lies in $W^{2,p}(\mathfrak{Q})$. There exists a constant $C>0$ depending on $p,\mathrm{reg}(\mathcal{T})$ and C_φ such that for any $\mathcal{D}\in\mathfrak{D}$ we have*

$$\int_{\mathcal{D}}|\nabla u_e(z)-\nabla^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e(z)|^p\,dz\le C\mathrm{d}_{\mathcal{D}}^{p(p-1)\alpha_p^2}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}\left(1+|\nabla u_e|^p+|D^2u_e|^p\right)\,dz.$$

Proof.

Let us give the proof in the case where $\mathcal{D}$ is an interior diamond cell. The case $\mathcal{D}\in\mathfrak{D}_{\mathrm{ext}}$ can be treated in a similar, and in fact simpler, way.

Let us define the projection $\mathbb{P}^{\mathfrak{Q}}u_e$ of u_e on the set of quarter diamonds as follows. For each quarter diamond $\mathcal{Q}\in\mathfrak{Q}$, the restriction of $\mathbb{P}^{\mathfrak{Q}}u_e$ to the triangle $\mathcal{Q}$ is the unique affine function $\mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}}u_e$ which coincides with u_e at the middle of the semi-edges $\sigma\in\mathcal{E}_{\mathcal{Q}}$ and whose value at the middle of the third side of $\mathcal{Q}$ is the mean-value of the values of u_e at the extremities of this side. Notice that this definition makes sense since $u_e\big|_{\overline{\mathcal{Q}}}\in W^{2,p}(\mathcal{Q})\subset\mathcal{C}^0(\overline{\mathcal{Q}})$. As an example, in the case of the quarter diamond $\mathcal{Q}=\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}$ (see Figure 5.1), this definition reads

$$\begin{aligned}\mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}u_e(x_{\sigma_{\mathcal{K}}})&=u_e(x_{\sigma_{\mathcal{K}}}),\\ \mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}u_e(x_{\sigma_{\mathcal{K}^*}})&=u_e(x_{\sigma_{\mathcal{K}^*}}),\\ \mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}u_e\left(\frac{x_{\mathcal{K}}+x_{\mathcal{K}^*}}{2}\right)&=\frac{u_e(x_{\mathcal{K}})+u_e(x_{\mathcal{K}^*})}{2}.\end{aligned}$$

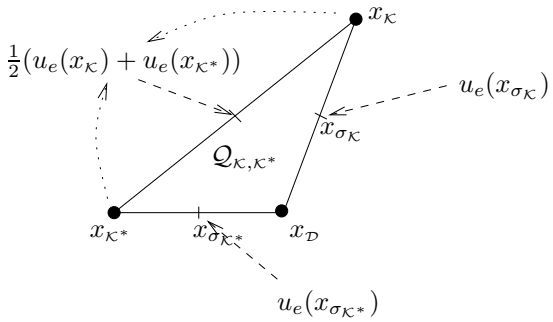


FIG. 5.1. *The affine interpolation* $\mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}$ *on the quarter diamond* $\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}$

The gradient of $\mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} u_e$ is then given by

$$\nabla \mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} u_e = \frac{2}{\sin\alpha_{\mathcal{D}}} \left(\frac{u_e(x_{\sigma_{\mathcal{K}^*}}) - \frac{u_e(x_{\mathcal{K}}) + u_e(x_{\mathcal{K}^*})}{2}}{|\sigma_{\mathcal{K}}|} \boldsymbol{\nu} + \frac{u_e(x_{\sigma_{\mathcal{K}}}) - \frac{u_e(x_{\mathcal{K}}) + u_e(x_{\mathcal{K}^*})}{2}}{|\sigma_{\mathcal{K}^*}|} \boldsymbol{\nu}^* \right).$$

Let us now define the consistency error for this projection $\mathbb{P}^{\mathfrak{Q}}$ as follows

$$T_{\overline{\mathcal{Q}}}(z) = \nabla u_e(z) - \nabla \mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}} u_e, \quad \forall z \in \mathcal{Q}, \quad \forall \mathcal{Q} \in \mathfrak{Q}. \tag{5.3}$$

By usual Taylor expansions inside each quarter diamond $\mathcal{Q}$ (see [3] for instance) we can easily show that there exists a constant $C > 0$ as in the claim on the proposition such that

$$\int_{\mathcal{Q}} |T_{\overline{\mathcal{Q}}}(z)|^p \, dz \leq C \mathrm{d}_{\mathcal{Q}}^p \int_{\mathcal{Q}} |D^2 u_e(z)|^p \, dz, \quad \forall \mathcal{Q} \in \mathfrak{Q}. \tag{5.4}$$

By the discussion of Section 4.3 we remark that, $\mathcal{D}$ being an interior diamond cell, there exists $\tilde{\delta}_{\mathcal{D}} \in \mathbb{R}^4$ such that

$$\nabla \mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}} u_e - \nabla^{\mathcal{T}}_{\mathcal{D}} \mathbb{P}^{\mathcal{T}} u_e = B_{\mathcal{Q}} \widetilde{\delta}_{\mathcal{D}}, \, \forall \mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}},$$

and then, by the definition (4.13) of $\nabla^{\mathcal{N}}_{\mathcal{Q}}$, we deduce that there exists $\overline{\delta^{\mathcal{D}}} \in \mathbb{R}^4$ such that

$$\nabla \mathbb{P}^{\mathfrak{Q}}_{\mathcal{Q}} u_e - \nabla^{\mathcal{N}}_{\mathcal{Q}} \mathbb{P}^{\mathcal{T}} u_e = B_{\mathcal{Q}} \overline{\delta^{\mathcal{D}}}, \, \forall \mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}. \tag{5.5}$$

Since u_e solves (1.1) with $f \in L^{p'}(\Omega)$, we know that the following transmission property holds

$$\int_{\sigma_{\mathcal{K}}} \varphi_{|\frac{(z, \nabla u_e|_{\frac{(s))}{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}}}{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}} \cdot \boldsymbol{\nu}^* \, ds = \int_{\sigma_{\mathcal{K}}} \varphi_{|\frac{(z, \nabla u_e|_{\frac{(s))}{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}}}{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}} \cdot \boldsymbol{\nu}^* \, ds.$$

Recall that the gradient operator $\nabla^{\mathcal{N}}$ is built to ensure that the discrete equivalent of this property, that is the first equation of (4.9), holds. It follows that

$$\left(\frac{1}{|\sigma_{\mathcal{K}}|} \int_{\sigma_{\mathcal{K}}} \varphi_{|\frac{(s, \nabla u_e|_{\frac{(s))}{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}}}{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}} \, ds - \varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}(\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} \mathbb{P}^{\mathcal{T}} u_e), \boldsymbol{\nu}^* \right)$$
$$- \left(\frac{1}{|\sigma_{\mathcal{K}}|} \int_{\sigma_{\mathcal{K}}} \varphi_{|\frac{(s, \nabla u_e|_{\frac{(s))}{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}}}{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}} \, ds - \varphi_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}(\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}} \mathbb{P}^{\mathcal{T}} u_e), \boldsymbol{\nu}^* \right) = 0.$$

By using Definition 5.5, we get

$$\left(\frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}|}\int_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}\Big(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}^*\right)$$
$$-\left(\frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}|}\int_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}\Big(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}^*\right)$$
$$=R^{\varphi}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*},\sigma_{\mathcal{K}}}-R^{z}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*},\sigma_{\mathcal{K}}}-R^{\varphi}_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*},\sigma_{\mathcal{K}}}+R^{z}_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*},\sigma_{\mathcal{K}}}.$$

Similarly we obtain for the other three semi-edges in the diamond under study the following relations

$$\left(\frac{1}{|\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}|}\int_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}\Big(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}^*\right)$$
$$-\left(\frac{1}{|\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}|}\int_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\Big(\varphi(z,\nabla u_e(z))\,dz-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}^*\right)$$
$$=R^{\varphi}_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*},\sigma_{\mathcal{L}}}-R^{z}_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*},\sigma_{\mathcal{L}}}-R^{\varphi}_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*},\sigma_{\mathcal{L}}}+R^{z}_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*},\sigma_{\mathcal{L}}}.$$

$$\left(\frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}|}\int_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}\Big(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}\right)$$
$$-\left(\frac{1}{|\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}|}\int_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}\Big(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}\right)$$
$$=R^{\varphi}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*},\sigma_{\mathcal{K}^*}}-R^{z}_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*},\sigma_{\mathcal{K}^*}}-R^{\varphi}_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*},\sigma_{\mathcal{K}^*}}+R^{z}_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*},\sigma_{\mathcal{K}^*}}.$$

$$\left(\frac{1}{|\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}|}\int_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}\Big(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}\right)$$
$$-\left(\frac{1}{|\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}|}\int_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\Big(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\mathbb{P}^{\mathcal{T}}u_e)\Big)\,dz,\boldsymbol{\nu}\right)$$
$$=R^{\varphi}_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*},\sigma_{\mathcal{L}^*}}-R^{z}_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*},\sigma_{\mathcal{L}^*}}-R^{\varphi}_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*},\sigma_{\mathcal{L}^*}}+R^{z}_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*},\sigma_{\mathcal{L}^*}}.$$

Multiplying these equations respectively by $|\sigma_{\mathcal{K}}|\overline{\delta_{\mathcal{K}}}$, $|\sigma_{\mathcal{L}}|\overline{\delta_{\mathcal{L}}}$, $|\sigma_{\mathcal{K}^*}|\overline{\delta_{\mathcal{K}^*}}$, and $|\sigma_{\mathcal{L}^*}|\overline{\delta_{\mathcal{L}^*}}$ and summing, we obtain

$$\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}\left(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla^{\mathcal{N}}_{\mathcal{Q}}\mathbb{P}^{\mathcal{T}}u_e),B_{\mathcal{Q}}\overline{\delta^{\mathcal{D}}}\right)dz$$
$$\leq\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}||B_{\mathcal{Q}}\overline{\delta^{\mathcal{D}}}|\sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}}(|R^{\varphi}_{\mathcal{Q},\sigma}|+|R^{z}_{\mathcal{Q},\sigma}|),$$

where we used the definitions (4.2)-(4.5). Using (5.3) and (5.5) we finally deduce

$$\begin{aligned}
&\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}\left(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla_{\mathcal{Q}}^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e),\nabla u_e(z)-\nabla_{\mathcal{Q}}^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e\right)\,dz\\
&\leq\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\left(\int_{\mathcal{Q}}|\nabla u_e(z)-\nabla_{\mathcal{Q}}^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e(z)|\,dz+\int_{\mathcal{Q}}|T_{\overline{\mathcal{Q}}}(z)|\,dz\right)\sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}}(|R_{\mathcal{Q},\sigma}^{\varphi}|+|R_{\mathcal{Q},\sigma}^{z}|)\\
&\qquad+\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}\left(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla_{\mathcal{Q}}^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e),T_{\overline{\mathcal{Q}}}(z)\right)\,dz. \quad (5.6)
\end{aligned}$$

In the case $p>2$, using assumptions $(\mathcal{H}_{1'b})$ and $(\mathcal{H}_{4b})$ and Young's inequality, we deduce from formula (5.6) that

$$\begin{aligned}
&\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}|\nabla u_e(z)-\nabla_{\mathcal{Q}}^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^p\,dz\\
&\qquad\leq C\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}}|\mathcal{Q}|\left(|R_{\mathcal{Q},\sigma}^{\varphi}|^{\frac{p}{p-1}}+|R_{\mathcal{Q},\sigma}^{z}|^{\frac{p}{p-1}}\right)+C\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}|T_{\overline{\mathcal{Q}}}(z)|^p\,dz\\
&\qquad\qquad+C\left(\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}|T_{\overline{\mathcal{Q}}}(z)|^p\,dz\right)^{\frac{1}{p-1}}\left(\int_{\mathcal{D}}(1+|\nabla u_e(z)|^p)\,dz\right)^{\frac{p-2}{p-1}}.
\end{aligned}$$

From (5.4) and the estimates in Proposition 5.6, it follows that

$$\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}|\nabla u_e(z)-\nabla_{\mathcal{Q}}^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^p\,dz\leq C\mathrm{d}_{\mathcal{D}}^{\frac{p}{p-1}}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}\int_{\mathcal{Q}}(1+|\nabla u_e|^p+|D^2u_e|^p)\,dz,$$

and the claim is proved.

In the case $1<p\leq 2$, using assumptions $(\mathcal{H}_{1'a})$ and $(\mathcal{H}_{4a})$, we deduce from

formula (5.6) that

$$
\begin{aligned}
&\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q |\nabla u_e(z)-\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^p\,dz\\
&\le\Bigg[\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\left(\int_Q|\nabla u_e(z)-\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e(z)|\,dz+\int_Q|T_{\overline{Q}}(z)|\,dz\right)\sum_{\sigma\in\mathcal{E}_Q}(|R^{\varphi}_{Q,\sigma}|+|R^{\mathrm{z}}_{Q,\sigma}|)\\
&\qquad\qquad+\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q(\varphi(z,\nabla u_e(z))-\varphi(z,\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e),T_{\overline{Q}}(z))\,dz\Bigg]^{\frac{p}{2}}\\
&\times\left(\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q(1+|\nabla u_e(z)|^p+|\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^p)\,dz\right)^{\frac{2-p}{2}}\\
&\le C\Bigg[\left(\left(\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q|\nabla u_e(z)-\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e(z)|^p\,dz\right)^{\frac12}+\left(\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q|T_{\overline{Q}}(z)|^p\,dz\right)^{\frac12}\right)\\
&\qquad\qquad\times\left(\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}|\sum_{\sigma\in\mathcal{E}_Q}|R^{\varphi}_{Q,\sigma}|^{\frac{p}{p-1}}+|R^{\mathrm{z}}_{Q,\sigma}|^{\frac{p}{p-1}}\right)^{\frac{p-1}{2}}\\
&+\left(\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q|\nabla u_e(z)-\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^p\right)^{\frac{p-1}{2}}\left(\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q|T_{\overline{Q}}(z)|^p\,dz\right)^{\frac12}\Bigg]\\
&\times\left(\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q(1+|\nabla u_e(z)|^p+|\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^p)\,dz\right)^{\frac{2-p}{2}}
\end{aligned}
$$

Using Young's inequality, (5.4) and the estimates in Proposition 5.6, the claim follows. □

We can now estimate the consistency error of the scheme due to the approximation of the gradient as follows.

PROPOSITION 5.8. *Let $\mathcal{T}$ be a mesh on Ω and assume that u_e lies in $W^{2,p}(\mathfrak{Q})$. There exists a constant $C>0$ depending on $p,\mathrm{reg}(\mathcal{T})$ and C_φ such that for any $\mathcal{D}\in\mathfrak{D}$ we have*

$$
\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}||R_Q^{\mathrm{grad}}|^{\frac{p}{p-1}}\le C\mathrm{d}_{\mathcal{D}}^{p(p-1)\alpha_p^3}\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q\left(1+|\nabla u_e|^p+|D^2u_e|^p\right)\,dz.
$$

Proof. In the case $1<p\le2$, using assumption $(\mathcal{H}_{4a})$ and the consistency estimate of Proposition 5.6, we deduce that

$$
\begin{aligned}
\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}||R_Q^{\mathrm{grad}}|^{\frac{p}{p-1}}&\le C_\varphi^{\frac{p}{p-1}}\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}|\left(\frac{1}{|\mathcal{Q}|}\int_Q|\nabla u_e(z)-\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^{p-1}\,dz\right)^{\frac{p}{p-1}}\\
&\le C_\varphi^{\frac{p}{p-1}}\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q|\nabla u_e(z)-\nabla_Q^{\mathcal{N}}\mathbb{P}^{\mathcal{T}}u_e|^p\,dz\\
&\le C\mathrm{d}_{\mathcal{D}}^{p(p-1)\alpha_p^2}\sum_{Q\in\mathfrak{Q}_{\mathcal{D}}}\int_Q\left(1+|\nabla u_e|^p+|D^2u_e|^p\right)\,dz,
\end{aligned}
$$

which gives the claim since $\alpha_p = 1$ as soon as $p \geq 2$.

When $p > 2$, by Jensen's and Hölder's inequality, we have

$$\begin{aligned}
&\sum_{\mathcal{Q}\in\mathfrak{Q}_\mathcal{D}} |\mathcal{Q}||R_\mathcal{Q}^{\rm grad}|^{\frac{p}{p-1}} \\
&\quad\leq C_\varphi^{\frac{p}{p-1}} \sum_{\mathcal{Q}\in\mathfrak{Q}_\mathcal{D}} |\mathcal{Q}| \Bigg(\frac{1}{|\mathcal{Q}|} \int_\mathcal{Q} \left(1 + |\nabla u_e(z)|^{p-2} + |\nabla_\mathcal{Q}^\mathcal{N} \mathbb{P}^\mathcal{T} u_e|^{p-2}\right) \\
&\qquad\qquad \times |\nabla u_e(z) - \nabla_\mathcal{Q}^\mathcal{N} \mathbb{P}^\mathcal{T} u_e| \, dz \Bigg)^{\frac{p}{p-1}} \\
&\quad\leq C \left(\sum_{\mathcal{Q}\in\mathfrak{Q}_\mathcal{D}} \int_\mathcal{Q} \left(1 + |\nabla u_e(z)|^p + |\nabla_\mathcal{Q}^\mathcal{N} \mathbb{P}^\mathcal{T} u_e|^p\right) \, dz \right)^{\frac{p-2}{p-1}} \\
&\qquad\qquad \times \left(\sum_{\mathcal{Q}\in\mathfrak{Q}_\mathcal{D}} \int_\mathcal{Q} |\nabla u_e(z) - \nabla_\mathcal{Q}^\mathcal{N} \mathbb{P}^\mathcal{T} u_e|^p \, dz \right)^{\frac{1}{p-1}} \\
&\quad\leq C \mathrm{d}_\mathcal{D}^{p\alpha_p^2} \sum_{\mathcal{Q}\in\mathfrak{Q}_\mathcal{D}} \int_\mathcal{Q} \left(1 + |\nabla u_e|^p + |D^2 u_e|^p\right) \, dz,
\end{aligned} \tag{5.7}$$

and we conclude by noting that $p\alpha_p^2 = p(p-1)\alpha_p^3$ when $p > 2$. ∎

5.2. Proof of Theorem 5.1. We have

$$\|\nabla u_e - \nabla^\mathcal{N} u^\mathcal{T}\|_{L^p} \leq \|\nabla u_e - \nabla^\mathcal{N} \mathbb{P}^\mathcal{T} u_e\|_{L^p} + \|\nabla^\mathcal{N} \mathbb{P}^\mathcal{T} u_e - \nabla^\mathcal{N} u^\mathcal{T}\|_{L^p}.$$

Proposition 5.7 gives a bound for the first term. We come back to the proof of the stability Theorem 4.10 to evaluate the second one. That proof shows that the estimate of $\|\nabla^\mathcal{N} \mathbb{P}^\mathcal{T} u_e - \nabla^\mathcal{N} u^\mathcal{T}\|_{L^p}$ requires the control of

$$I \stackrel{\rm def}{=} \left(\mathfrak{a}^\mathcal{N}(\mathbb{P}^\mathcal{T} u_e) - \mathfrak{a}^\mathcal{N}(u^\mathcal{T}), \mathbb{P}^\mathcal{T} u_e - u^\mathcal{T}\right).$$

By classical manipulations (using the conservativity of numerical fluxes) we express I through the consistency errors thanks to

$$\begin{aligned}
\mathfrak{a}_\mathcal{K}^\mathcal{N}(u^\mathcal{T}) - \mathfrak{a}_\mathcal{K}^\mathcal{N}(\mathbb{P}^\mathcal{T} u_e) &= \sum_{\mathcal{Q}\subset\mathcal{K}} \sum_{\sigma\in\mathcal{E}_\mathcal{Q}\cap\partial\mathcal{K}} |\sigma| R_{\mathcal{Q},\sigma}, \, \forall \mathcal{K} \in \mathfrak{M} \\
\mathfrak{a}_{\mathcal{K}^*}^\mathcal{N}(u^\mathcal{T}) - \mathfrak{a}_{\mathcal{K}^*}^\mathcal{N}(\mathbb{P}^\mathcal{T} u_e) &= \sum_{\mathcal{Q}\subset\mathcal{K}^*} \sum_{\sigma\in\mathcal{E}_\mathcal{Q}\cap\partial\mathcal{K}^*} |\sigma| R_{\mathcal{Q},\sigma}, \, \forall \mathcal{K}^* \in \mathfrak{M}^*.
\end{aligned}$$

If we define the error $e^\mathcal{T} = u^\mathcal{T} - \mathbb{P}^\mathcal{T} u_e$, the formulas above yield

$$I = \sum_{\mathcal{K}\in\mathfrak{M}} \sum_{\mathcal{Q}\subset\mathcal{K}} \sum_{\sigma\in\mathcal{E}_\mathcal{Q}\cap\partial\mathcal{K}} |\sigma| R_{\mathcal{Q},\sigma} e_\mathcal{K} + \sum_{\mathcal{K}^*\in\mathfrak{M}^*} \sum_{\mathcal{Q}\subset\mathcal{K}^*} \sum_{\sigma\in\mathcal{E}_\mathcal{Q}\cap\partial\mathcal{K}^*} |\sigma| R_{\mathcal{Q},\sigma} e_{\mathcal{K}^*}.$$

Reordering the sum over the diamond cells, we find that

$$\begin{aligned}
I = \sum_{\mathcal{D}\in\mathfrak{D}} \Bigg(&|\sigma_{\mathcal{K}^*}| (R_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} e_\mathcal{K} + R_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} e_\mathcal{L}) \\
&+ |\sigma_{\mathcal{L}^*}| (R_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*,\sigma_{\mathcal{L}^*}}} e_\mathcal{K} + R_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*,\sigma_{\mathcal{L}^*}}} e_\mathcal{L}) \\
&+ |\sigma_\mathcal{K}| (R_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*,\sigma_\mathcal{K}}} e_{\mathcal{K}^*} + R_{\mathcal{Q}_{\mathcal{K},\mathcal{L}^*,\sigma_\mathcal{K}}} e_{\mathcal{L}^*}) \\
&+ |\sigma_\mathcal{L}| (R_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*,\sigma_\mathcal{K}}} e_{\mathcal{K}^*} + R_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*,\sigma_\mathcal{K}}} e_{\mathcal{L}^*}) \Bigg).
\end{aligned} \tag{5.8}$$

Using the conservativity property (5.2), the first term in the sum above reads

$$\begin{aligned}
&|\sigma_{\mathcal{K}^*}|(R_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} e_{\mathcal{K}} + R_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} e_{\mathcal{L}}) \\
&\quad = -|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}| R_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} \frac{2}{\sin\alpha_{\mathcal{D}}} \frac{e_{\mathcal{L}} - e_{\mathcal{K}}}{|\sigma_{\mathcal{K}}| + |\sigma_{\mathcal{L}}|} + |\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}| R_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} \frac{2}{\sin\alpha_{\mathcal{D}}} \frac{e_{\mathcal{L}} - e_{\mathcal{K}}}{|\sigma_{\mathcal{K}}| + |\sigma_{\mathcal{L}}|} \\
&\qquad = -|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}| R_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} (\nabla_{\mathcal{D}}^{\mathcal{T}} e^{\mathcal{T}}, \boldsymbol{\tau}^*) + |\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}| R_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} (\nabla_{\mathcal{D}}^{\mathcal{T}} e^{\mathcal{T}}, \boldsymbol{\tau}^*). \quad (5.9)
\end{aligned}$$

We remark now, by using (4.2) and (4.5), that for any $\delta \in \mathbb{R}^4$ we have

$$|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}| \left(B_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}} \delta, \boldsymbol{\tau}^*\right) = |\sigma_{\mathcal{K}^*}| \sin\alpha_{\mathcal{D}} \delta_{\mathcal{K}^*} = -|\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}| \left(B_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}} \delta, \boldsymbol{\tau}^*\right).$$

Hence by using once more the conservativity property and the definition (4.13), we can replace $\nabla_{\mathcal{D}}^{\mathcal{T}} e^{\mathcal{T}}$ by the corresponding $\nabla_{\mathcal{Q}}^{\mathcal{N}} e^{\mathcal{T}}$ in the right-hand side of (5.9). It follows

$$\begin{aligned}
&|\sigma_{\mathcal{K}^*}|(R_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} e_{\mathcal{K}} + R_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} e_{\mathcal{L}}) \\
&\qquad = -|\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}| R_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} (\nabla_{\mathcal{Q}_{\mathcal{K},\mathcal{K}^*}}^{\mathcal{N}} e^{\mathcal{T}}, \boldsymbol{\tau}^*) + |\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}| R_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*,\sigma_{\mathcal{K}^*}}} (\nabla_{\mathcal{Q}_{\mathcal{L},\mathcal{K}^*}}^{\mathcal{N}} e^{\mathcal{T}}, \boldsymbol{\tau}^*).
\end{aligned}$$

The other terms in (5.8) being treated in the same way, it follows that

$$\begin{aligned}
I &\le C \sum_{\mathcal{Q}\in\mathfrak{Q}} \sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}} |\mathcal{Q}||R_{\mathcal{Q},\sigma}||\nabla_{\mathcal{Q}}^{\mathcal{N}} e^{\mathcal{T}}| \\
&\le C \left(\sum_{\mathcal{Q}\in\mathfrak{Q}} \sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}} |\mathcal{Q}||R_{\mathcal{Q},\sigma}|^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} \|\nabla^{\mathcal{N}} e^{\mathcal{T}}\|_{L^p}.
\end{aligned}$$

Using assumptions $(\mathcal{H}_{1'a})$ and $(\mathcal{H}_{1'b})$, we derive that,

$$\begin{aligned}
\|\nabla^{\mathcal{N}} e^{\mathcal{T}}\|_{L^p}^2 \le C(1 + \|\nabla^{\mathcal{N}} u^{\mathcal{T}}\|_{L^p}^{2-p} + \|\nabla^{\mathcal{N}} \mathbb{P}^{\mathcal{T}} u_e\|_{L^p}^{2-p}) \\
\times \left(\sum_{\mathcal{Q}\in\mathfrak{Q}} \sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}} |\mathcal{Q}||R_{\mathcal{Q},\sigma}|^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} \|\nabla^{\mathcal{N}} e^{\mathcal{T}}\|_{L^p}, \quad \text{if } 1 < p \le 2,
\end{aligned}$$

and

$$\|\nabla^{\mathcal{N}} e^{\mathcal{T}}\|_{L^p}^p \le C \left(\sum_{\mathcal{Q}\in\mathfrak{Q}} \sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}} |\mathcal{Q}||R_{\mathcal{Q},\sigma}|^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} \|\nabla^{\mathcal{N}} e^{\mathcal{T}}\|_{L^p}, \quad \text{if } p > 2. \quad (5.10)$$

The claim follows by using the estimate (4.18) and Propositions 5.6, 5.7 and 5.8.

5.3. Proof of Theorem 5.3. We only give the proof in the case $p > 2$ since the other case can be treated in the same way. We come back to (5.10) which is still valid for the h-DDFV scheme. It follows

$$\|\nabla^{\mathcal{N}} e^{\mathcal{T}}\|_{L^p} \le C \left(\sum_{\mathcal{Q}\in\mathfrak{Q}} \sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}} |\mathcal{Q}||R_{\mathcal{Q},\sigma}|^{\frac{p}{p-1}} \right)^{\frac{1}{p}}.$$

The terms $R_{\mathcal{Q},\sigma}$ contain the three respective contributions of $R^{\mathrm{z}}_{\mathcal{Q}}$, $R^{\varphi}_{\mathcal{Q},\sigma}$ and $R^{\mathrm{grad}}_{\mathcal{Q}}$. As far as $R^{\mathrm{z}}_{\mathcal{Q}}$ and $R^{\varphi}_{\mathcal{Q},\sigma}$ are concerned, the estimate of Proposition 5.6 is still valid for the hybrid scheme so that

$$\left(\sum_{\mathcal{Q}\in\mathfrak{Q}}\sum_{\sigma\in\mathcal{E}_{\mathcal{Q}}}|\mathcal{Q}|\left(|R^{\mathrm{z}}_{\mathcal{Q},\sigma}|^{\frac{p}{p-1}}+|R^{\varphi}_{\mathcal{Q},\sigma}|^{\frac{p}{p-1}}\right)\right)^{\frac{1}{p}}\leq C\mathrm{size}(\mathcal{T})^{\frac{1}{p-1}}\|u_e\|_{W^{2,p}(\mathfrak{D})}.$$

We split now the contribution of $R^{\mathrm{grad}}_{\mathcal{Q}}$ in two parts : the one coming from diamond cells in $\mathfrak{D}_{\mathrm{cont}}$ where the usual DDFV approximate flux is usued and the one coming from the diamond cells in $\mathfrak{D}_{\mathrm{disc}}$ where we used our new discrete gradient and flux. It follows

$$\left(\sum_{\mathcal{Q}\in\mathfrak{Q}}|\mathcal{Q}||R^{\mathrm{grad}}_{\mathcal{Q}}|^{\frac{p}{p-1}}\right)^{\frac{1}{p}}\leq C\left(\sum_{\mathcal{D}\in\mathfrak{D}_{\mathrm{cont}}}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}||R^{\mathrm{grad}}_{\mathcal{Q}}|^{\frac{p}{p-1}}\right)^{\frac{1}{p}}+C\left(\sum_{\mathcal{D}\in\mathfrak{D}_{\mathrm{disc}}}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}||R^{\mathrm{grad}}_{\mathcal{Q}}|^{\frac{p}{p-1}}\right)^{\frac{1}{p}}. \quad (5.11)$$

Near the discontinuities of the flux, that is for each diamond cell $\mathcal{D}\in\mathfrak{D}_{\mathrm{disc}}$, we use the estimate (5.7). Away from the discontinuities, i.e. for $\mathcal{D}\in\mathfrak{D}_{\mathrm{cont}}$ we used the usual DDFV scheme (that is $\nabla^{\mathcal{N}}=\nabla^{\mathcal{T}}$), so that the gradient consistency estimate given by (5.7) reads, since $u_e\in W^{2,p}(\mathcal{D})$:

$$\int_{\mathcal{D}}|\nabla u_e(z)-\nabla^{\mathcal{T}}\mathbb{P}^{\mathcal{T}}u_e(z)|^p\,dz\leq C\mathrm{d}_{\mathcal{D}}^p\int_{\mathcal{D}}\left(1+|\nabla u_e|^p+|D^2u_e|^p\right)\,dz.$$

This estimate is proved for instance in [3, Lemma 7.5]. Hence, (5.11) now gives

$$\left(\sum_{\mathcal{Q}\in\mathfrak{Q}}|\mathcal{Q}||R^{\mathrm{grad}}_{\mathcal{Q}}|^{\frac{p}{p-1}}\right)^{\frac{1}{p}}\leq C\mathrm{size}(\mathcal{T})^{\frac{1}{p-1}}\|u_e\|_{W^{2,p}(\mathfrak{D})}+C\mathrm{size}(\mathcal{T})^{\frac{1}{(p-1)^2}}\left(\int_{\Omega_{\mathrm{disc}}}(1+|\nabla u_e|+|D^2u_e|)^p\,dz\right)^{\frac{1}{p}},$$

where we introduced $\Omega_{\mathrm{disc}}=\bigcup_{\mathcal{D}\in\mathfrak{D}_{\mathrm{disc}}}\mathcal{D}$. Since we assumed that φ is smooth on each subdomain Ω_i, we see that the set Ω_{disc} is an $\mathrm{size}(\mathcal{T})$-neighborhood of union of the boundaries of the Ω_i's. Hence, there exists $C>0$ such that $|\Omega_{\mathrm{disc}}|\leq C\mathrm{size}(\mathcal{T})$. It follows by the Hölder inequality and using the assumption $u_e\in W^{2,q}(\mathfrak{Q})$, that

$$\left(\sum_{\mathcal{Q}\in\mathfrak{Q}}|\mathcal{Q}||R^{\mathrm{grad}}_{\mathcal{Q}}|^{\frac{p}{p-1}}\right)^{\frac{1}{p}}\leq C\mathrm{size}(\mathcal{T})^{\frac{1}{p-1}}\|u_e\|_{W^{2,p}(\mathfrak{D})}+C\mathrm{size}(\mathcal{T})^{\left(\frac{1}{(p-1)^2}+\frac{q-p}{pq}\right)}\|u_e\|_{W^{2,q}(\mathfrak{Q})},$$

and the claim is proved since $\frac{1}{(p-1)^2}+\frac{q-p}{pq}\geq\frac{1}{p-1}$ as soon as $q\geq p(p-1)^2$.

6. Examples. In the case of a linear problem where $\varphi(z,\xi) = A(z)\xi$, it is easily seen that, for any $\mathcal{D} \in \mathfrak{D}$ the numerical flux $\varphi_{\mathcal{D}}^{\mathcal{N}}$ is a linear map of the DDFV gradient $\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}$. More precisely, there exists a unique definite positive matrix $\overline{A}_{\mathcal{D}}$ such that $\varphi_{\mathcal{D}}^{\mathcal{N}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}) = \overline{A}_{\mathcal{D}} \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}}$.

In general, it is difficult to give an explicit formula for the matrix $\overline{A}_{\mathcal{D}}$ but it can be evaluated by computing the map $\delta^{\mathcal{D}}$ that is, following Proposition 4.2 and its proof, by computing the inverse of the $n_{\mathcal{D}} \times n_{\mathcal{D}}$ matrix $\sum_{\mathcal{D} \in \mathfrak{D}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}} B_{\mathcal{Q}}$. This operation has a very low computational cost and has to be made only once.

In some particular cases, it is possible to find an explicit form for $\overline{A}_{\mathcal{D}}$ which is interesting in order to illustrate our approach and to compare the results with the 1D case. Let us consider a given diamond cell $\mathcal{D} \in \mathfrak{D}_{\rm int}$ whose diagonals are $\sigma = \kappa|\mathcal{L}$ and $\sigma^* = \kappa^*|\mathcal{L}^*$.

- *First example:* We assume that $A(z)$ is constant on each control volume. We denote by A_κ the value of $A(z)$ on the control volume κ. The matrix $\overline{A}_{\mathcal{D}}$ is then defined by

$$(\overline{A}_{\mathcal{D}}\boldsymbol{\nu},\boldsymbol{\nu}) = \frac{(|\sigma_\kappa| + |\sigma_\mathcal{L}|)(A_\kappa\boldsymbol{\nu},\boldsymbol{\nu})(A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu})}{|\sigma_\mathcal{L}|(A_\kappa\boldsymbol{\nu},\boldsymbol{\nu}) + |\sigma_\kappa|(A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu})}, \tag{6.1}$$

$$\begin{aligned}(\overline{A}_{\mathcal{D}}\boldsymbol{\nu}^*,\boldsymbol{\nu}^*) = {}& \frac{|\sigma_\mathcal{L}|(A_\mathcal{L}\boldsymbol{\nu}^*,\boldsymbol{\nu}^*) + |\sigma_\kappa|(A_\kappa\boldsymbol{\nu}^*,\boldsymbol{\nu}^*)}{|\sigma_\kappa| + |\sigma_\mathcal{L}|} \\ & - \frac{|\sigma_\kappa||\sigma_\mathcal{L}|}{|\sigma_\kappa| + |\sigma_\mathcal{L}|} \frac{((A_\kappa\boldsymbol{\nu},\boldsymbol{\nu}^*) - (A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu}^*))^2}{|\sigma_\mathcal{L}|(A_\kappa\boldsymbol{\nu},\boldsymbol{\nu}) + |\sigma_\kappa|(A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu})},\end{aligned} \tag{6.2}$$

$$(\overline{A}_{\mathcal{D}}\boldsymbol{\nu},\boldsymbol{\nu}^*) = \frac{|\sigma_\mathcal{L}|(A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu}^*)(A_\kappa\boldsymbol{\nu},\boldsymbol{\nu}) + |\sigma_\kappa|(A_\kappa\boldsymbol{\nu},\boldsymbol{\nu}^*)(A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu})}{|\sigma_\mathcal{L}|(A_\kappa\boldsymbol{\nu},\boldsymbol{\nu}) + |\sigma_\kappa|(A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu})}. \tag{6.3}$$

  We recognize in (6.1) the weighted harmonic mean-value of $(A_\kappa\boldsymbol{\nu},\boldsymbol{\nu})$ and $(A_\mathcal{L}\boldsymbol{\nu},\boldsymbol{\nu})$ and in the first term of (6.2) the weighted arithmetic mean-value of $(A_\kappa\boldsymbol{\nu}^*,\boldsymbol{\nu}^*)$ and $(A_\mathcal{L}\boldsymbol{\nu}^*,\boldsymbol{\nu}^*)$.
- *Second example:* We assume that $A(z) = \lambda(z)\mathrm{Id}$ is isotropic, continuous on each quarter-diamond and we assume that the mesh is orthogonal, that is $\sigma \perp \sigma^*$ or equivalently $\sin\alpha_{\mathcal{D}} = 1$. Introducing $\lambda_{\mathcal{Q}} = \int_{\overline{\mathcal{Q}}} \lambda(z)\, d\mu_{\overline{\mathcal{Q}}}(z)$, the mean-value of λ over $\overline{\mathcal{Q}}$ with respect to the measure $d\mu_{\overline{\mathcal{Q}}}$, the equivalent matrix $\overline{A}_{\mathcal{D}}$ satisfies in that case:

$$\begin{aligned}(\overline{A}_{\mathcal{D}}\boldsymbol{\nu},\boldsymbol{\nu}) = {}& \frac{|\sigma_\kappa| + |\sigma_\mathcal{L}|}{|\sigma_{\kappa^*}| + |\sigma_{\mathcal{L}^*}|}\left(\frac{|\sigma_{\kappa^*}|\lambda_{\mathcal{Q}_{\mathcal{L},\kappa^*}}\lambda_{\mathcal{Q}_{\kappa,\kappa^*}}}{|\sigma_\kappa|\lambda_{\mathcal{Q}_{\mathcal{L},\kappa^*}} + |\sigma_\mathcal{L}|\lambda_{\mathcal{Q}_{\kappa,\kappa^*}}}\right. \\ & \left. + \frac{|\sigma_{\mathcal{L}^*}|\lambda_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\lambda_{\mathcal{Q}_{\kappa,\mathcal{L}^*}}}{|\sigma_\kappa|\lambda_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}} + |\sigma_\mathcal{L}|\lambda_{\mathcal{Q}_{\kappa,\mathcal{L}^*}}}\right),\end{aligned} \tag{6.4}$$

$$\begin{aligned}(\overline{A}_{\mathcal{D}}\boldsymbol{\nu}^*,\boldsymbol{\nu}^*) = {}& \frac{|\sigma_{\kappa^*}| + |\sigma_{\mathcal{L}^*}|}{|\sigma_\kappa| + |\sigma_\mathcal{L}|}\left(\frac{|\sigma_\kappa|\lambda_{\mathcal{Q}_{\kappa,\mathcal{L}^*}}\lambda_{\mathcal{Q}_{\kappa,\kappa^*}}}{|\sigma_{\kappa^*}|\lambda_{\mathcal{Q}_{\kappa,\mathcal{L}^*}} + |\sigma_{\mathcal{L}^*}|\lambda_{\mathcal{Q}_{\kappa,\kappa^*}}}\right. \\ & \left. + \frac{|\sigma_\mathcal{L}|\lambda_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}}\lambda_{\mathcal{Q}_{\mathcal{L},\kappa^*}}}{|\sigma_{\kappa^*}|\lambda_{\mathcal{Q}_{\mathcal{L},\mathcal{L}^*}} + |\sigma_{\mathcal{L}^*}|\lambda_{\mathcal{Q}_{\mathcal{L},\kappa^*}}}\right),\end{aligned} \tag{6.5}$$

$$(\overline{A}_{\mathcal{D}}\boldsymbol{\nu},\boldsymbol{\nu}^*)=0. \tag{6.6}$$

Notice that even though $A(z)$ is isotropic, the matrix $\overline{A}_{\mathcal{D}}$ is only diagonal (in the orthogonal frame $(\boldsymbol{\nu},\boldsymbol{\nu}^*)$) and not isotropic in general. Furthermore, we see that (6.4) and (6.5) combine arithmetic mean-value of the coefficients in the transverse direction and harmonic mean-value of the coefficients along the direction we are looking at.

Unfortunately, in the nonlinear case there are very few cases where all the computations can be performed explicitly (see for instance the 1D example 2.1). That is the reason why we propose in the following section a fully practical method to solve the m-DDFV and h-DDFV schemes in any situation.

7. Numerical implementation of the scheme. In this section we present a fully explicit algorithm to solve the finite volume scheme under study and we prove its convergence. From now on, we suppose given a DDFV mesh $\mathcal{T}$ on Ω and a source term f.

7.1. Some remarks on the potential case. We assume, only in this paragraph, that φ derives from a potential Φ, that is

$$\left\{\begin{array}{lcl}\varphi(z,\xi) &=& \nabla_\xi\Phi(z,\xi), \quad \text{for all } \xi\in\mathbb{R}^2 \ \text{ and a.e. } z\in\Omega,\\ \Phi(z,0) &=& 0, \quad \text{for a.e. } z\in\Omega.\end{array}\right. \tag{7.1}$$

We can now define an approximation of Φ on each quarter-diamond by $\Phi_{\mathcal{Q}}(\cdot)=\int_{\overline{\mathcal{Q}}}\Phi(z,\cdot)d\mu_{\mathcal{Q}}(z)$, that satisfies $\nabla\Phi_{\mathcal{Q}}=\varphi_{\mathcal{Q}}$. Since φ is strictly monotonic, the function Φ is strictly convex.

PROPOSITION 7.1. *The solution $u^{\mathcal{T}}$ of the scheme* (4.16) *is the unique minimizer of the functional defined by*

$$J^{\mathcal{T}}(v^{\mathcal{T}})=2\sum_{\mathcal{D}\in\mathfrak{D}}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}|\Phi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}v^{\mathcal{T}})-\sum_{\mathcal{K}}|\mathcal{K}|f_{\mathcal{K}}v_{\mathcal{K}}-\sum_{\mathcal{K}^*}|\mathcal{K}^*|f_{\mathcal{K}^*}v_{\mathcal{K}^*},\quad\forall v^{\mathcal{T}}\in\mathbb{R}^{\mathcal{T}}.$$

Proof. By using assumptions $(\mathcal{H}_1)$, $(\mathcal{H}_2)$ and $(\mathcal{H}_3)$, the definition (7.1) and the Poincaré inequality, it is easily seen that $J^{\mathcal{T}}$ is strictly convex and coercive on $\mathbb{R}^{\mathcal{T}}$ and thus has a unique minimizer that we call $u^{\mathcal{T}}$.

Let us now write the Euler-Lagrange equation for this minimization problem. The equation corresponding to the unknown $u_{\mathcal{K}}$ reads

$$2\sum_{\mathcal{D}\in\mathfrak{D}_{\mathcal{K}}}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}|\left(\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}}),(\mathrm{Id}+\mathrm{B}_{\mathcal{Q}}.\mathrm{D}\delta^{\mathcal{D}})\frac{\partial\nabla^{\mathcal{T}}_{\mathcal{D}}\mathrm{u}^{\mathcal{T}}}{\partial\mathrm{u}_{\mathcal{K}}}\right)=|\mathcal{K}|f_{\mathcal{K}}.$$

By definition of $\delta^{\mathcal{D}}$, for any $\mathcal{D}$ we have $\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}|^tB_{\mathcal{Q}}\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}})=0$ so that the term containing the derivative $D\delta^{\mathcal{D}}$ of $\delta^{\mathcal{D}}$ vanishes. Furthermore, by definition of $\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}$ we have

$$\frac{\partial\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}}{\partial u_{\mathcal{K}}}=-\frac{1}{\sin\alpha_{\mathcal{D}}}\frac{\boldsymbol{\nu}_{\mathcal{K}}}{|\sigma^*|},$$

hence it follows, using (4.15),

$$\begin{aligned}|\mathcal{K}|f_{\mathcal{K}}&=-2\sum_{\mathcal{D}\in\mathfrak{D}_{\mathcal{K}}}\left(\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}}|\mathcal{Q}|\varphi_{\mathcal{Q}}(\nabla^{\mathcal{N}}_{\mathcal{Q}}u^{\mathcal{T}}),\frac{1}{\sin\alpha_{\mathcal{D}}}\frac{\boldsymbol{\nu}_{\mathcal{K}}}{|\sigma^*|}\right)\\&=-2\sum_{\mathcal{D}\in\mathfrak{D}_{\mathcal{K}}}|\mathcal{D}|\left(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}),\frac{1}{\sin\alpha_{\mathcal{D}}}\frac{\boldsymbol{\nu}_{\mathcal{K}}}{|\sigma^*|}\right)=-\sum_{\mathcal{D}\in\mathfrak{D}_{\mathcal{K}}}|\sigma|\left(\varphi^{\mathcal{N}}_{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}),\boldsymbol{\nu}_{\mathcal{K}}\right),\end{aligned}$$

since $|\mathcal{D}| = \frac{1}{2}|\sigma||\sigma^*|\sin\alpha_{\mathcal{D}}$, and the claim is proved. □

From now on, we denote by $\Delta = \bigoplus_{\mathcal{D}\in\mathfrak{D}} \mathbb{R}^{n_{\mathcal{D}}}$ the space in which the artificial unknowns $(\delta^{\mathcal{D}})_{\mathcal{D}}$ are lying. We introduce the following new functional defined on $\mathbb{R}^{\mathcal{T}} \times \Delta$

$$J^{\mathcal{T},\Delta}(v^{\mathcal{T}},\tilde{\delta}) = 2\sum_{\mathcal{D}\in\mathfrak{D}}\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|\Phi_{\mathcal{Q}}(\nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}} + B_{\mathcal{Q}}\tilde{\delta}^{\mathcal{D}}) - \sum_{\mathcal{K}}|\mathcal{K}|f_{\mathcal{K}}v_{\mathcal{K}} - \sum_{\mathcal{K}^*}|\mathcal{K}^*|f_{\mathcal{K}^*}v_{\mathcal{K}^*}, \quad \forall v^{\mathcal{T}}\in\mathbb{R}^{\mathcal{T}},\ \forall\tilde{\delta}\in\Delta.$$

Proposition 7.2. *The functional $J^{\mathcal{T},\Delta}$ has a unique minimizer which is given by $(u^{\mathcal{T}}, (\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}u^{\mathcal{T}}))_{\mathcal{D}})$.*

Proof. It is easily seen that, for any $v^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$ fixed, the functional $\tilde{\delta}\in\Delta \mapsto J^{\mathcal{T},\Delta}(v^{\mathcal{T}},\tilde{\delta})$ decouples into a sum over $\mathfrak{D}$ of independant maps depending only on $\tilde{\delta}^{\mathcal{D}}$ for a given $\mathcal{D}\in\mathfrak{D}$. Following the proof of Proposition 4.2, each of these maps has a unique minimum exactly given by $\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}})$. Hence we proved

$$J^{\mathcal{T}}(v^{\mathcal{T}}) = J^{\mathcal{T},\Delta}\left(v^{\mathcal{T}}, (\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}}))_{\mathcal{D}}\right) \le J^{\mathcal{T},\Delta}(v^{\mathcal{T}},\tilde{\delta}), \quad \forall\tilde{\delta}\in\Delta,$$

with equality if and only if $\tilde{\delta} = (\delta^{\mathcal{D}}(\nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}}))_{\mathcal{D}}$, which gives the claim. □

7.2. Derivation of the decomposition-coordination method. In the case of a potential flux φ, we proved in Proposition 7.2 that the solution of the scheme can be obtained by minimizing a functional which can be computed explicitly but depending on much more unknowns than the cardinality of $\mathbb{R}^{\mathcal{T}}$ in which we look for the approximate solution.

We propose, for this non-quadratic minimization problem, a saddle-point formulation (in the very spirit of [12, 13]). Let us define the set $(\mathbb{R}^2)^{\mathfrak{Q}}$ of families of vectors of $\mathbb{R}^2$ indexed by the set of quarter diamonds. We suppose given a family $\mathcal{A} = (A_{\mathcal{Q}})_{\mathcal{Q}\in\mathfrak{Q}}$ of definite positive 2×2 matrices which is aimed to play the role of heterogeneous and isotropic augmentation parameters. More precisely, we introduce now the augmented lagrangian

$$L^{\mathcal{T},\Delta}_{\mathcal{A}}(v^{\mathcal{T}},\tilde{\delta},g,\lambda) = 2\sum_{\mathcal{Q}\in\mathfrak{Q}}|\mathcal{Q}|\Phi_{\mathcal{Q}}(g_{\mathcal{Q}}) + 2\sum_{\mathcal{Q}\in\mathfrak{Q}}|\mathcal{Q}|(\lambda_{\mathcal{Q}}, g_{\mathcal{Q}} - \nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}} - B_{\mathcal{Q}}\tilde{\delta}^{\mathcal{D}}) + \sum_{\mathcal{Q}\in\mathfrak{Q}}|\mathcal{Q}|\left(A_{\mathcal{Q}}(g_{\mathcal{Q}} - \nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}} - B_{\mathcal{Q}}\tilde{\delta}^{\mathcal{D}}), (g_{\mathcal{Q}} - \nabla^{\mathcal{T}}_{\mathcal{D}}v^{\mathcal{T}} - B_{\mathcal{Q}}\tilde{\delta}^{\mathcal{D}})\right) - \sum_{\mathcal{K}}|\mathcal{K}|f_{\mathcal{K}}v_{\mathcal{K}} - \sum_{\mathcal{K}^*}|\mathcal{K}^*|f_{\mathcal{K}^*}v_{\mathcal{K}^*}, \quad \forall v^{\mathcal{T}}\in\mathbb{R}^{\mathcal{T}}, \forall\tilde{\delta}\in\Delta, \forall g,\lambda\in(\mathbb{R}^2)^{\mathfrak{Q}}.$$

If, for any $\mathcal{Q}\in\mathfrak{Q}$, we take $A_{\mathcal{Q}} = r\mathrm{Id}$ for a given parameter $r > 0$ we recover the augmented lagrangian algorithm proposed in [12, 13]. It is easily seen that this

lagrangian has a unique saddle-point $(u^{\mathcal{T}}, \delta^{\mathcal{D}}, p, \lambda)$ satisfying the equilibrium equations

$$\begin{cases} \varphi_{\mathcal{Q}}(g_{\mathcal{Q}}) + \lambda_{\mathcal{Q}} + A_{\mathcal{Q}}(g_{\mathcal{Q}} - \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}} - B_{\mathcal{Q}}\delta^{\mathcal{D}}) = 0, \quad \forall \mathcal{Q} \in \mathfrak{Q}, \\ \displaystyle\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}} A_{\mathcal{Q}}(B_{\mathcal{Q}}\delta^{\mathcal{D}} + \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}} - g_{\mathcal{Q}}) - \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}}\lambda_{\mathcal{Q}} = 0, \quad \forall \mathcal{D} \in \mathfrak{D}, \\ g_{\mathcal{Q}} - \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}} - B_{\mathcal{Q}}\delta^{\mathcal{D}} = 0, \quad \forall \mathcal{Q} \in \mathfrak{Q}, \\ \displaystyle 2\sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|\,(A_{\mathcal{Q}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}} + B_{\mathcal{Q}}\delta^{\mathcal{D}} - g_{\mathcal{Q}}), \nabla_{\mathcal{D}}^{\mathcal{T}} v^{\mathcal{T}}) = \sum_{\mathcal{K}} |\mathcal{K}| f_{\mathcal{K}} v_{\mathcal{K}} + \sum_{\mathcal{K}^*} |\mathcal{K}^*| f_{\mathcal{K}^*} v_{\mathcal{K}^*} \\ \qquad \displaystyle + 2\sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|(\lambda_{\mathcal{Q}}, \nabla_{\mathcal{D}}^{\mathcal{T}} v^{\mathcal{T}}), \quad \forall v^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}. \end{cases} \tag{7.2}$$

These equations are clearly equivalent to

$$\begin{cases} \lambda_{\mathcal{Q}} = -\varphi_{\mathcal{Q}}(g_{\mathcal{Q}}), \quad \forall \mathcal{Q} \in \mathfrak{Q}, \\ \displaystyle\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}}\lambda_{\mathcal{Q}} = 0, \quad \forall \mathcal{D} \in \mathfrak{D}, \\ g_{\mathcal{Q}} = \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}} + B_{\mathcal{Q}}\delta^{\mathcal{D}}, \quad \forall \mathcal{Q} \in \mathfrak{Q}, \\ \displaystyle -2\sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|(\lambda_{\mathcal{Q}}, \nabla_{\mathcal{D}}^{\mathcal{T}} v^{\mathcal{T}}) = \sum_{\mathcal{K}} |\mathcal{K}| f_{\mathcal{K}} v_{\mathcal{K}} + \sum_{\mathcal{K}^*} |\mathcal{K}^*| f_{\mathcal{K}^*} v_{\mathcal{K}^*}, \quad \forall v^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}. \end{cases} \tag{7.3}$$

The first three equations imply that $\delta^{\mathcal{D}} = \delta^{\mathcal{D}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T}})$, then the fourth equation is nothing but a different way to write (4.16). As a consequence, the saddle-point of $L_{\mathcal{A}}^{\mathcal{T},\Delta}$ gives the unique solution to the finite volume scheme.

From equations (7.2), we deduce an iterative method to solve our problem following the same idea than [12, ALG 2, p. 170]. In our setting, the algorithm reads as follows: we suppose given $\lambda^0 \in (\mathbb{R}^2)^{\mathfrak{Q}}$, $g^0 \in (\mathbb{R}^2)^{\mathfrak{Q}}$ then for any $n \geq 1$:

1. Find $(u^{\mathcal{T},n}, \delta_{\mathcal{D}}^n) \in \mathbb{R}^{\mathcal{T}} \times \Delta$ solution to the linear problem

$$\begin{aligned} &2\sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|\left(A_{\mathcal{Q}}(\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T},n} + B_{\mathcal{Q}}\delta_{\mathcal{D}}^n - g_{\mathcal{Q}}^{n-1}), \nabla_{\mathcal{D}}^{\mathcal{T}} v^{\mathcal{T}}\right) \\ &\qquad = \sum_{\mathcal{K}} |\mathcal{K}| f_{\mathcal{K}} v_{\mathcal{K}} + \sum_{\mathcal{K}^*} |\mathcal{K}^*| f_{\mathcal{K}^*} v_{\mathcal{K}^*} + 2\sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|(\lambda_{\mathcal{Q}}^{n-1}, \nabla_{\mathcal{D}}^{\mathcal{T}} v), \quad \forall v^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}. \\ &\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}} A_{\mathcal{Q}}(B_{\mathcal{Q}}\delta_{\mathcal{D}}^n + \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T},n} - g_{\mathcal{Q}}^{n-1}) - \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}}\lambda_{\mathcal{Q}}^{n-1} = 0, \forall \mathcal{D} \in \mathfrak{D}. \end{aligned} \tag{7.4}$$

Notice that the second equation explicitly gives, locally on each diamond cell $\mathcal{D}$, the expression of $\delta_{\mathcal{D}}^n$ as an affine function of $\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T},n}$. It is only needed to compute, one time at the beginning of the algorithm, the inverse of all the definite positive symmetric matrices $\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}}\, A_{\mathcal{Q}}\, B_{\mathcal{Q}}$ whose size is $n_{\mathcal{D}} \times n_{\mathcal{D}}$ (which is low since $n_{\mathcal{D}} = 1$ for boundary diamond cells and $n_{\mathcal{D}} = 4$ for interior diamond cells).
Finally, once the second equation in (7.4) is solved, we can introduce the expression of $\delta_{\mathcal{D}}^n$ as a function of $\nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T},n}$ in the first equation of (7.4). This first equation is now an explicit large linear system in the variables $u^{\mathcal{T},n}$. Notice that the matrix of this large sparse linear system is the same at each iteration which is very important if one wants to use direct linear solvers

for instance. Finally, notice that the size and the stencil of this system is exactly the same than the one of the DDFV matrix for the Laplace equation for instance.

2. For any $\mathcal{Q} \in \mathfrak{Q}$, find $g_{\mathcal{Q}}^n$ satisfying

$$\varphi_{\mathcal{Q}}(g_{\mathcal{Q}}^n) + \lambda_{\mathcal{Q}}^{n-1} + A_{\mathcal{Q}}(g_{\mathcal{Q}}^n - \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T},n} - B_{\mathcal{Q}}\delta_{\mathcal{D}}^n) = 0. \tag{7.5}$$

This is the unique nonlinear part of the algorithm. The equation is localized on each quarter diamond and consists in solving a nonlinear equation in $\mathbb{R}^2$ defined by the explicit map $\varphi_{\mathcal{Q}}+A_{\mathcal{Q}}$. Notice that $\varphi_{\mathcal{Q}}+A_{\mathcal{Q}}$ is strictly monotonic and coercive on $\mathbb{R}^2$ so that the solution $g_{\mathcal{Q}}^n$ to (7.5) exists and is unique.
From a practical point of view, one can use here a Newton method to solve all these equations simultaneously (this step can be massively parallelized).

3. Finally compute $\lambda_{\mathcal{Q}}^n$ through

$$\lambda_{\mathcal{Q}}^n = \lambda_{\mathcal{Q}}^{n-1} + \gamma A_{\mathcal{Q}}(g_{\mathcal{Q}}^n - \nabla_{\mathcal{D}}^{\mathcal{T}} u^{\mathcal{T},n} - B_{\mathcal{Q}}\delta_{\mathcal{D}}^n), \quad \forall \mathcal{Q} \in \mathfrak{Q}, \tag{7.6}$$

where $\gamma > 0$ is given parameter.

The choice of the *best* augmentation matrices $\mathcal{A}$ and parameter γ is a complex problem (see the discussion in [12] for instance). We will give some examples of such choices in Section 8.

7.3. General fluxes. Convergence of the iterative solver. The above algorithm is deduced from the lagrangian formulation of the scheme which is only available in the variational case (7.1). Nevertheless, the iterative algorithm (7.4)-(7.6) can be used in the general case of any monotonic flux φ. We can now prove the convergence of this algorithm in the non-variational setting.

THEOREM 7.3. *Let $\mathcal{T}$ be a DDFV mesh on Ω and $(\varphi_{\mathcal{Q}})_{\mathcal{Q}}$ a family of strictly monotonic continuous maps from $\mathbb{R}^2$ onto itself. Then for any augmentation matrices family $\mathcal{A}$ and any $\gamma \in \left]0, \frac{1+\sqrt{5}}{2}\right[$, the algorithm given by* (7.4)-(7.6) *converges, when n goes to infinity, towards the unique solution to* (7.3) *that is the unique solution to the m-DDFV scheme* (4.16).

The proof we present here is an adaptation to our framework of the arguments given in [12]. Furthermore, a similar algorithm and result is available for the hybrid h-DDFV scheme.

Proof. For any real-valued or vector-valued families $f = (f_{\mathcal{Q}})_{\mathcal{Q}\in\mathfrak{Q}}$ and $g = (g_{\mathcal{Q}})_{\mathcal{Q}\in\mathfrak{Q}}$ we introduce the following inner products and associated norms:

$$\begin{aligned}
(f,g)_0 &= \sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|(f_{\mathcal{Q}}, g_{\mathcal{Q}}), \quad \|f\|_0 = (f,f)_0^{\frac{1}{2}}, \\
(f,g)_{\mathcal{A}} &= \sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|(A_{\mathcal{Q}} f_{\mathcal{Q}}, g_{\mathcal{Q}}), \quad \|f\|_{\mathcal{A}} = (f,f)_{\mathcal{A}}^{\frac{1}{2}}, \\
(f,g)_{\mathcal{A}^{-1}} &= \sum_{\mathcal{Q}\in\mathfrak{Q}} |\mathcal{Q}|(A_{\mathcal{Q}}^{-1} f_{\mathcal{Q}}, g_{\mathcal{Q}}), \quad \|f\|_{\mathcal{A}^{-1}} = (f,f)_{\mathcal{A}^{-1}}^{\frac{1}{2}}.
\end{aligned}$$

Let us define the error terms $v^{\mathcal{T},n} = u^{\mathcal{T},n} - u^{\mathcal{T}}$, $h^n = (h_{\mathcal{Q}}^n)_{\mathcal{Q}} \in (\mathbb{R}^2)^{\mathfrak{Q}}$ with $h_{\mathcal{Q}}^n = g_{\mathcal{Q}}^n - g_{\mathcal{Q}}$, $\mu^n = (\mu_{\mathcal{Q}}^n)_{\mathcal{Q}} \in (\mathbb{R}^2)^{\mathfrak{Q}}$ with $\mu_{\mathcal{Q}}^n = \lambda_{\mathcal{Q}}^n - \lambda_{\mathcal{Q}}$ and $\beta^n = (\beta_{\mathcal{D}}^n)_{\mathcal{D}} \in \Delta$ with $\beta_{\mathcal{D}}^n = \delta_{\mathcal{D}}^n - \delta^{\mathcal{D}}$. Finally, we introduce the nonlinear map $\varphi_{\mathfrak{Q}} : (\mathbb{R}^2)^{\mathfrak{Q}} \mapsto (\mathbb{R}^2)^{\mathfrak{Q}}$ defined by $(\varphi_{\mathfrak{Q}}(g))_{\mathcal{Q}} = \varphi_{\mathcal{Q}}(g_{\mathcal{Q}})$ and the linear map $\mathcal{B} : \Delta \mapsto (\mathbb{R}^2)^{\mathfrak{Q}}$ defined by $(\mathcal{B}\beta)_{\mathcal{Q}} = B_{\mathcal{Q}}\beta_{\mathcal{D}}$, where $\mathcal{D}$ is the diamond cell such that $\mathcal{Q} \in \mathfrak{Q}_{\mathcal{D}}$.

Using those notations, we see from (7.4)-(7.6) that these quantities solve the following equations:

1. Equation for $v^{\mathcal{T},n}$:

$$\left(\nabla^{\mathcal{T}} v^{\mathcal{T},n} + \mathcal{B}\beta^n - h^n, \nabla^{\mathcal{T}}_{\mathcal{D}} w^{\mathcal{T}}\right)_{\mathcal{A}} + \left(h^n - h^{n-1}, \nabla^{\mathcal{T}} w^{\mathcal{T}}\right)_{\mathcal{A}} = \left(\mu^{n-1}, \nabla^{\mathcal{T}} w^{\mathcal{T}}\right)_0, \quad \forall w^{\mathcal{T}} \in \mathbb{R}^{\mathcal{T}}. \quad (7.7)$$

2. Equation for $\beta^n_{\mathcal{D}}$:

$$\sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}} A_{\mathcal{Q}} (B_{\mathcal{Q}}\beta^n_{\mathcal{D}} + \nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} - h^n_{\mathcal{Q}}) + \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}} A_{\mathcal{Q}} (h^n_{\mathcal{Q}} - h^{n-1}_{\mathcal{Q}}) - \sum_{\mathcal{Q}\in\mathfrak{Q}_{\mathcal{D}}} |\mathcal{Q}|^t B_{\mathcal{Q}} \mu^{n-1}_{\mathcal{Q}} = 0, \quad \forall \mathcal{D} \in \mathfrak{D}. \quad (7.8)$$

3. Equation for $h^n_{\mathcal{Q}}$:

$$\varphi_{\mathcal{Q}}(g^n_{\mathcal{Q}}) - \varphi_{\mathcal{Q}}(g_{\mathcal{Q}}) + \mu^{n-1}_{\mathcal{Q}} + A_{\mathcal{Q}}(h^n_{\mathcal{Q}} - \nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} - B_{\mathcal{Q}}\beta^n_{\mathcal{D}}) = 0, \quad \forall \mathcal{Q} \in \mathfrak{Q}. \quad (7.9)$$

4. Equation for $\mu^n_{\mathcal{Q}}$:

$$\mu^n_{\mathcal{Q}} = \mu^{n-1}_{\mathcal{Q}} + \gamma A_{\mathcal{Q}}(h^n_{\mathcal{Q}} - \nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} - B_{\mathcal{Q}}\beta^n_{\mathcal{D}}), \quad \forall \mathcal{Q} \in \mathfrak{Q}. \quad (7.10)$$

We take $w^{\mathcal{T}} = v^{\mathcal{T},n}$ in (7.7), we get

$$\left(\nabla^{\mathcal{T}} v^{\mathcal{T},n} + \mathcal{B}\beta^n - h^n, \nabla^{\mathcal{T}} v^{\mathcal{T},n}\right)_{\mathcal{A}} + \left(h^n - h^{n-1}, \nabla^{\mathcal{T}} v^{\mathcal{T},n}\right)_{\mathcal{A}} = \left(\mu^{n-1}, \nabla^{\mathcal{T}} v^{\mathcal{T},n}\right)_0. \quad (7.11)$$

From (7.10), we deduce

$$\frac{1}{2}\|\mu^n\|^2_{\mathcal{A}^{-1}} - \frac{1}{2}\left\|\mu^{n-1}\right\|^2_{\mathcal{A}^{-1}} - \gamma\left(\mu^{n-1}, h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n\right)_0 - \frac{1}{2}\gamma^2 \|h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n\|^2_{\mathcal{A}} = 0. \quad (7.12)$$

Taking the $(\cdot,\cdot)_0$ inner product of (7.9) with $h^n = g^n - g$, we get

$$(\varphi_{\mathfrak{Q}}(g^n) - \varphi_{\mathfrak{Q}}(g), g^n - g)_0 + (h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n, h^n)_{\mathcal{A}} + \left(\mu^{n-1}, h^n\right)_0 = 0.$$

Multiplying (7.8) by $\beta^n_{\mathcal{D}}$, summing over $\mathfrak{D}$ and using (7.11), it follows

$$(\varphi_{\mathfrak{Q}}(g^n) - \varphi_{\mathfrak{Q}}(g), g^n - g)_0 + \|h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n\|^2_{\mathcal{A}} + \left(\mu^{n-1}, h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n\right)_0 + \left(h^n - h^{n-1}, \nabla^{\mathcal{T}} v^{\mathcal{T},n} + \mathcal{B}\beta^n\right)_{\mathcal{A}} = 0.$$

Multiplying this equation by γ and adding to (7.12), we get rid of the term containing $\mu^{n-1}_{\mathcal{Q}}$. We obtain

$$\frac{1}{2}\|\mu^n\|^2_{\mathcal{A}^{-1}} - \frac{1}{2}\left\|\mu^{n-1}\right\|^2_{\mathcal{A}^{-1}} + \gamma(\varphi_{\mathfrak{Q}}(g^n) - \varphi_{\mathfrak{Q}}(g), g^n - g)_0 + \frac{\gamma}{2}(2-\gamma)\|h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n\|^2_{\mathcal{A}} + \gamma\left(h^n - h^{n-1}, \nabla^{\mathcal{T}} v^{\mathcal{T},n} + \mathcal{B}\beta^n\right)_{\mathcal{A}} = 0. \quad (7.13)$$

From (7.5) we get, for any $\mathcal{Q} \in \mathfrak{Q}$,

$$\varphi_{\mathcal{Q}}(g^n_{\mathcal{Q}}) - \varphi_{\mathcal{Q}}(g^{n-1}_{\mathcal{Q}}) + (\mu^{n-1}_{\mathcal{Q}} - \mu^{n-2}_{\mathcal{Q}}) \\ + A_{\mathcal{Q}}\Big((h^n_{\mathcal{Q}} - h^{n-1}_{\mathcal{Q}}) - (\nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} - \nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n-1}) - B_{\mathcal{Q}}(\beta^n_{\mathcal{D}} - \beta^{n-1}_{\mathcal{D}})\Big) = 0,$$

so that, using (7.10) it follows

$$\varphi_{\mathcal{Q}}(g^n_{\mathcal{Q}}) - \varphi_{\mathcal{Q}}(g^{n-1}_{\mathcal{Q}}) + A_{\mathcal{Q}}(h^n_{\mathcal{Q}} - h^{n-1}_{\mathcal{Q}}) + A_{\mathcal{Q}} h^{n-1}_{\mathcal{Q}} \\ - A_{\mathcal{Q}}\big(\nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} + B_{\mathcal{Q}}\beta^n_{\mathcal{D}}\big) = (1-\gamma) A_{\mathcal{Q}}(h^{n-1}_{\mathcal{Q}} - \nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n-1} - B_{\mathcal{Q}}\beta^{n-1}_{\mathcal{D}}).$$

Taking the inner product $(\cdot,\cdot)_0$ of these equations by $g^n - g^{n-1} = h^n - h^{n-1}$ we get

$$\big(\varphi_{\mathfrak{Q}}(g^n) - \varphi_{\mathfrak{Q}}(g^{n-1}), g^n - g^{n-1}\big)_0 + \big\|h^n - h^{n-1}\big\|^2_{\mathcal{A}} \\ + \big(h^{n-1}, h^n - h^{n-1}\big)_{\mathcal{A}} - \big(\nabla^{\mathcal{T}} v^{\mathcal{T},n} + \mathcal{B}\beta^n, h^n - h^{n-1}\big)_{\mathcal{A}} \\ = (1-\gamma)\big(h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}, h^n - h^{n-1}\big)_{\mathcal{A}}.$$

If we add to this equation the following algebraic relation

$$\frac{1}{2}\|h^n\|^2_{\mathcal{A}} - \frac{1}{2}\big\|h^{n-1}\big\|^2_{\mathcal{A}} - \frac{1}{2}\big\|h^n - h^{n-1}\big\|^2_{\mathcal{A}} = \big(h^{n-1}, h^n - h^{n-1}\big)_{\mathcal{A}},$$

we get

$$\frac{1}{2}\|h^n\|^2_{\mathcal{A}} - \frac{1}{2}\big\|h^{n-1}\big\|^2_{\mathcal{A}} + \frac{1}{2}\big\|h^n - h^{n-1}\big\|^2_{\mathcal{A}} \\ + \big(\varphi_{\mathfrak{Q}}(g^n) - \varphi_{\mathfrak{Q}}(g^{n-1}), g^n - g^{n-1}\big)_0 - \big(\nabla^{\mathcal{T}} v^{\mathcal{T},n} + \mathcal{B}\beta^n, h^n - h^{n-1}\big)_{\mathcal{A}} \\ = (1-\gamma)\big(h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}, h^n - h^{n-1}\big)_{\mathcal{A}}.$$

Multiplying this equation by γ and summing with (7.13), it follows:

$$\frac{1}{2}\Big(\|\mu^n\|^2_{\mathcal{A}^{-1}} + \gamma\|h^n\|^2_{\mathcal{A}}\Big) - \frac{1}{2}\Big(\big\|\mu^{n-1}\big\|^2_{\mathcal{A}^{-1}} + \gamma\big\|h^{n-1}\big\|^2_{\mathcal{A}}\Big) \\ + \gamma\big(\varphi_{\mathfrak{Q}}(g^n) - \varphi_{\mathfrak{Q}}(g^{n-1}), g^n - g^{n-1}\big)_0 + \gamma\big(\varphi_{\mathfrak{Q}}(g^n) - \varphi_{\mathfrak{Q}}(g), g^n - g\big)_0 \\ + \frac{\gamma}{2}\big\|h^n - h^{n-1}\big\|^2_{\mathcal{A}} + \frac{\gamma}{2}(2-\gamma)\|h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n\|^2_{\mathcal{A}} \\ = \gamma(1-\gamma)\big(h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}, h^n - h^{n-1}\big)_{\mathcal{A}}. \quad (7.14)$$

We can now use Cauchy-Schwarz's and Young's inequalities to get

$$\gamma(1-\gamma)\big(h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}, h^n - h^{n-1}\big)_{\mathcal{A}} \\ \leq \frac{\gamma}{2}\big\|h^n - h^{n-1}\big\|^2_{\mathcal{A}} + \frac{\gamma(1-\gamma)^2}{2}\big\|h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}\big\|^2_{\mathcal{A}} \\ - \frac{\gamma}{2}\Big[\big\|h^n - h^{n-1}\big\|_{\mathcal{A}} - (1-\gamma)\big\|h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}\big\|_{\mathcal{A}}\Big]^2.$$

By the assumption on γ, we have $(1-\gamma)^2 \leq (2-\gamma)$ so that (7.14) yields

$$\begin{aligned}
&\frac{1}{2}\left(\|\mu^n\|^2_{\mathcal{A}^{-1}} + \gamma\,\|h^n\|^2_{\mathcal{A}} + \gamma(2-\gamma)\,\|h^n - \nabla^{\mathcal{T}} v^{\mathcal{T},n} - \mathcal{B}\beta^n\|^2_{\mathcal{A}}\right)\\
&\quad - \frac{1}{2}\left(\left\|\mu^{n-1}\right\|^2_{\mathcal{A}^{-1}} + \gamma\left\|h^{n-1}\right\|^2_{\mathcal{A}} + \gamma(2-\gamma)\left\|h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}\right\|^2_{\mathcal{A}}\right)\\
&\qquad + \gamma\left(\varphi_{\mathfrak{D}}(g^n) - \varphi_{\mathfrak{D}}(g^{n-1}), g^n - g^{n-1}\right)_0 + \gamma\left(\varphi_{\mathfrak{D}}(g^n) - \varphi_{\mathfrak{D}}(g), g^n - g\right)_0\\
&\qquad + \frac{\gamma}{2}\left[\left\|h^n - h^{n-1}\right\|_{\mathcal{A}} - (1-\gamma)\left\|h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}\right\|_{\mathcal{A}}\right]^2 \leq 0. \quad (7.15)
\end{aligned}$$

Since each map $\varphi_{\mathcal{Q}}$ is monotonic so is $\varphi_{\mathfrak{D}}$, we deduce that the sequence

$$\left(\|\mu^n\|^2_{\mathcal{A}^{-1}} + \gamma\,\|h^n\|^2_{\mathcal{A}} + \gamma(2-\gamma)\left\|h^{n-1} - \nabla^{\mathcal{T}} v^{\mathcal{T},n-1} - \mathcal{B}\beta^{n-1}\right\|^2_{\mathcal{A}}\right)_n$$

of non-negative numbers is non-increasing and then converges. Coming back to (7.15), we deduce that $(g^n_{\mathcal{Q}})_n$ converges towards $g_{\mathcal{Q}}$ for any $\mathcal{Q} \in \mathfrak{Q}$, that is $h^n_{\mathcal{Q}} \to 0$. If $\gamma \neq 1$ we also deduce from (7.15) that

$$h^n_{\mathcal{Q}} - \nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} - B_{\mathcal{Q}}\beta^n_{\mathcal{D}} \xrightarrow{n\to\infty} 0, \quad \forall \mathcal{Q} \in \mathfrak{Q}$$

and hence $\nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} + B_{\mathcal{Q}}\beta^n_{\mathcal{D}}$ goes to 0. Then it follows that $\mu^n_{\mathcal{Q}} \to 0$ by (7.9). Using now (4.8), we find that for any $\mathcal{D} \in \mathfrak{D}$ we have $\nabla^{\mathcal{T}}_{\mathcal{D}} v^{\mathcal{T},n} \to 0$ and finally $\beta^n_{\mathcal{D}} \to 0$.

In the case $\gamma = 1$, we can draw the same conclusions directly from (7.14) since the right-hand side is 0 in that case. □

8. Numerical results. We present here some numerical results in the following situation: we consider the domain $\Omega =]0,1[\times]0,1[$, and the flux φ defined by

$$\varphi(z,\xi) = \begin{cases} |\xi|^{p-2}\xi, & \text{if } z_1 < 0.5,\\ (A\xi,\xi)^{\frac{p-2}{2}} A\xi, & \text{if } z_1 > 0.5, \end{cases} \quad (8.1)$$

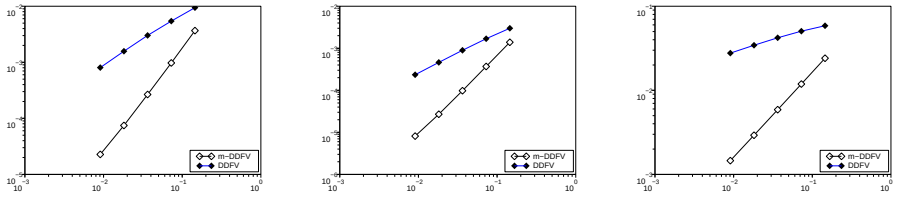
where A is the matrix $A = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$, $\alpha, \beta \in]1, +\infty[$. Then we construct the source term f, and the boundary data in such a way that the solution of (1.1) is given by

$$u_e(z) = \begin{cases} (\alpha z_1 + \gamma z_2)^2, & \text{if } z_1 < 0.5,\\ \left(z_1 + \gamma z_2 + \dfrac{\alpha - 1}{2}\right)^2, & \text{if } z_1 > 0.5, \end{cases}$$

where $\gamma = \sqrt{\alpha\frac{1-\alpha}{1-\beta}}$. It is easily seen that this function u_e satisfies the transmission condition on the line $\{z_1 = \frac{1}{2}\}$, for any value of p.

In a first test case, we choose $p = 3.0$ and $(\alpha, \beta) = (5.0, 2.0)$ so that the problem we consider has discontinuities and is anisotropic. We show in Figure 8.1 the errors in three different norms as a function of the mesh size, in a logarithmic scale, for the original DDFV scheme (3.1)-(3.2) (marked by ⧫) and for the m-DDFV scheme (4.15)-(4.16) (marked by ◊). Notice that, since the operator defined by (8.1) is piecewise constant, the h-DDFV scheme of Definition 5.2 is exactly the same than the m-DDFV scheme.

As predicted by the theory, the m-DDFV scheme provides a much better convergence rate than the original DDFV scheme. Furthermore, and it is an important

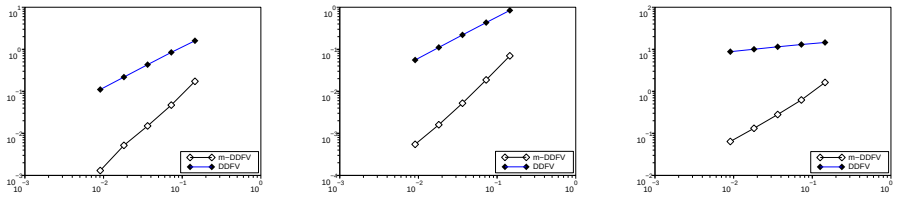


L^∞ error slopes 1.71 and 0.97 | L^p error slopes 1.73 and 0.97 | $W^{1,p}$ error slopes 1.0 and 0.31

FIG. 8.1. *Test case 1 :* $p = 3.0$, $\alpha = 5$, $\beta = 2$

point, the error (in any of the three norms we consider) obtained by the m-DDFV scheme is better even in the case of coarse meshes.

As a second test case we assume that $p = 5.0$ and $(\alpha, \beta) = (10.0, 10.0)$. In this situation the operator is isotropic but the jump of the diffusion coefficient is of order $10^{2.5}$. We observe (see Figure 8.2) the same overall behavior of the two schemes.



L^∞ error slopes 1.98 and 0.98 | L^p error slopes 1.54 and 0.99 | $W^{1,p}$ error slopes 1.04 and 0.19

FIG. 8.2. *Test case 2 :* $p = 5.0$, $\alpha = \beta = 10.0$

Finally, we want to illustrate the behavior of the decomposition-coordination algorithm proposed in Section 7. First of all, it is shown in [12] for instance that such algorithms can be applied, in suitable infinite dimensional functional spaces, directly to the continuous problem (1.1). This fact let us hope that the convergence rate of the present method may not depend too much on the size of the mesh we consider. Actually, in our numerical computations, we observed that the number of iterations needed to achieve a given residual norm was essentially the same in each level of refinement of the mesh we considered.

Let us now illustrate the fact that one can take advantage of using the heterogeneous and isotropic augmentation matrices family $\mathcal{A}$. To this end, we consider the first example given above (Figure 8.1) and we plot in Figure 8.3 the evolution during the iterations of the L^p, $W^{1,p}$ errors and the residual norm of the algorithm. The left plot is the one obtained with the classical augmentation term, that is when $A_{\mathcal{Q}} = r\mathrm{Id}$ for any $\mathcal{Q} \in \mathfrak{Q}$, with $r = 1.5$. In the right plot, we have chosen $A_{\mathcal{Q}} = r\mathrm{Id}$ if $\mathcal{Q} \subset \{z_1 < \frac{1}{2}\}$ and $A_{\mathcal{Q}} = rA$ if $\mathcal{Q} \subset \{z_1 > \frac{1}{2}\}$.

We see that the use of anisotropic and heterogeneous augmention terms let us achieve the tolerance 10^{-7} on the residual norm in ~ 130 iterations instead of ~ 180 in the isotropic case. Furthermore, we see that the error due to the scheme is achieved after ~ 40 iterations, that is when the residual norm is $\sim 5.10^{-6}$ in the first case whereas it is only achieved in ~ 150 iterations, for a residual norm of 5.10^{-7}. This

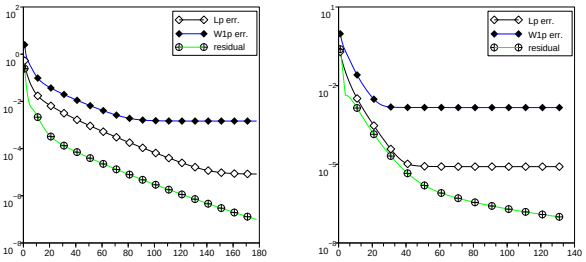


FIG. 8.3. *Convergence of the iterative solver. Isotropic (left) and anisotropic augmentation (right)*

illustrate the fact that the choice of suitable augmentation matrices $\mathcal{A}$ may let us save a significant amount of computational time to solve our scheme.

9. Conclusions. In this paper we provide a modification of the DDFV finite volume scheme for nonlinear elliptic problems on general 2D grids in order to take into account discontinuities in the coefficients. The m-DDFV scheme we obtained is proved to present a better consistency of the fluxes at the discontinuities. The performance of the scheme is illustrated by numerical results on heterogeneous and anisotropic p-laplacian equations. Furthemore, we proposed a generalisation of the decomposition-coordination method of Glowinski in order to solve our scheme. We show that the use of heterogeneous and anisotropic augmentation terms in this approach may lead to much better performance of the algorithm.

REFERENCES

[1] I. Aavatsmark, T. Barkve, O. Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. I. Derivation of the methods. *SIAM J. Sci. Comput.*, 19(5):1700–1716 (electronic), 1998.

[2] I. Aavatsmark, T. Barkve, O. Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. II. Discussion and numerical results. *SIAM J. Sci. Comput.*, 19(5):1717–1736 (electronic), 1998.

[3] B. Andreianov, F. Boyer, and F. Hubert. Discrete duality finite volume schemes for Leray-Lions type elliptic problems on general 2D-meshes. *Num. Meth. for PDEs*, 2006. `http://dx.doi.org/10.1002/num.20170`.

[4] B. Andreianov, M. Gutnic, and P. Wittbold. Convergence of finite volume approximations for a nonlinear elliptic-parabolic problem: a "continuous" approach. *SIAM J. Numer. Anal.*, 42(1):228–251, 2004.

[5] Y. Coudière, C. Pierre, and R. Turpault. Solving the fully coupled heart and torso problems of electrocardiology with a 3D discrete duality finite volume method. *submitted*, 2006. `http://hal.ccsd.cnrs.fr/ccsd-00016825`.

[6] Y. Coudière, J.-P. Vila, and P. 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.

[7] K. Domelevo and P. Omnes. A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *M2AN Math. Model. Numer. Anal.*, 39(6):1203–1249, 2005.

[8] J. Droniou. Finite volume approximations for fully non-linear elliptic equations in divergence form. *submitted*, 2005. `http://hal.ccsd.cnrs.fr/ccsd-00009614`.

[9] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000.

[10] R. Eymard, T. Gallouët, and R. Herbin. A cell-centred finite-volume approximation for

anisotropic diffusion operators on unstructured meshes in any space dimension. *IMA J. Numer. Anal.*, 26(2):326–353, 2006.

[11] M. Feistauer and V. Sobotíková. Finite element approximation of nonlinear elliptic problems with discontinuous coefficients. *RAIRO Modél. Math. Anal. Numér.*, 24(4):457–500, 1990.

[12] R. Glowinski. *Numerical methods for nonlinear variational problems*. Springer Series in Computational Physics. Springer-Verlag, New York, 1984.

[13] R. Glowinski and A. Marrocco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *Rev. Française Automat. Informat. Recherche Opérationnelle*, 9(R-2):41–76, 1975.

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

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

[16] J. Leray and J.-L. Lions. Quelques résultats de Višik sur les problèmes elliptiques nonlinéaires par les méthodes de Minty-Browder. *Bull. Soc. Math. France*, 93:97–107, 1965.

[17] W. B. Liu. Degenerate quasilinear elliptic equations arising from bimaterial problems in elastic-plastic mechanics. *Nonlinear Anal.*, 35(4, Ser. A: Theory Methods):517–529, 1999.

[18] W. B. Liu. Finite element approximation of a nonlinear elliptic equation arising from bimaterial problems in elastic-plastic mechanics. *Numer. Math.*, 86(3):491–506, 2000.

[19] C. Pierre. *Modélisation et simulation de l'activité électrique du coeur dans le thorax, analyse numérique et méthodes de volumes finis*. PhD thesis, Université de Nantes, 2005. `http://tel.ccsd.cnrs.fr/tel-00010705`.

[20] A. Ženíšek. The finite element method for nonlinear elliptic equations with discontinuous coefficients. *Numer. Math.*, 58(1):51–77, 1990.