ESAIM: M2AN
Vol. 40, N° 2, 2006, pp. 311–330
DOI: 10.1051/m2an:2006011

ESAIM: Mathematical Modelling and Numerical Analysis
www.edpsciences.org/m2an

FINITE VOLUME METHODS FOR THE VALUATION OF AMERICAN OPTIONS

Julien Berton[1] and Robert Eymard[1]

Abstract. We consider the use of finite volume methods for the approximation of a parabolic variational inequality arising in financial mathematics. We show, under some regularity conditions, the convergence of the upwind implicit finite volume scheme to a weak solution of the variational inequality in a bounded domain. Some results, obtained in comparison with other methods on two dimensional cases, show that finite volume schemes can be accurate and efficient.

Mathematics Subject Classification. 65M12.

Received: December 8, 2004. Revised: October 5, 2005.

1. Introduction

This paper concerns the use of some numerical schemes to obtain an approximate solution to a problem arising in financial mathematics. Many contracts that are traded in modern financial markets involve American options on several underlying assets. Contrary to European options, closed form or analytical solution are not available to price the American options, so numerical approximation methods are required.

Indeed, since the pioneering works on variational inequalities and stochastic control of Bensoussan and Lions [5], a large class of numerical methods have been developed to obtain an approximation of the price of an American option. Hence Cox *et al.* [13] have introduced the binomial method based on time and space discretizations, the convergence of which is proven by Amin and Khanna in [1]. Generalizations of the binomial approach are given by Boyle *et al.* [8], and Kamrad and Ritchken [20]. Carr *et al.* [12] used an integral representation of the option price, where integral formulas express the value of the American option as the value of the corresponding European option augmented by the present value of the gains from early exercise. These gains, in turn, parametrically depend on the optimal exercise boundary, which is the solution of a nonlinear integral equation subject to a boundary condition. While the option price has an explicit representation, the exercise boundary is implicitly defined by the integral equation, so that a sequential numerical procedure is required. Brennan and Schwartz [9] introduced finite difference methods related to the discretization of the variational inequality, the convergence of which is proven in Jaillet *et al.* [19]. An extension of such a method is to use a finite difference method in time combined with a finite element in space [14, 24]. Nevertheless, using these methods seem to be complex in high space dimension since a major difficulty is to obtain a spatial discretization at the same time accurate and admissible with regard to the number of unknowns.

Keywords and phrases. American option, variational inequality, finite volume method, convergence of numerical scheme.

[1] Université de Marne-la-Vallée, Champs-sur-Marne, 77454 Marne-la-Vallée, France. eymard@math.univ-mlv.fr

In order to develop some new ways to overcome this difficulty, we study in this paper finite volume schemes for pricing American options, focusing on such schemes for the following reasons:

(1) A finite volume scheme has already been successfully used by Zvan *et al.* in [26] on several numerical sample pricing problems.
(2) Following an idea proposed in [2], the use of Voronoï meshes seems to yield several advantages: the size of the control volumes can easily be reduced at the location of the moving boundary, and, for a given space step, it is possible to get a much smaller number of grid blocks in high space dimension than that obtained using hypercubic meshes. It is then easy to use finite volume schemes on Voronoï meshes for the approximation of nonlinear parabolic problems [15], whereas it does not seem to be straightforward to use finite element or finite difference schemes on such meshes.
(3) Finite volume schemes can also apply on many types of grids, including classical rectangular meshes and simplicial meshes (triangles in 2D).
(4) There exists a strong relation between the problem of pricing an American option and a Stefan problem, which describes the energy conservation within a material which changes of thermodynamical state [6]. Since finite volume methods respect the satisfaction of local balances, they lead to an accurate location of the resulting moving boundaries.

Therefore we state in Section 2 the mathematical formulation of the pricing of an American option, recalling the variational inequality issued from a stochastic differential equation, the solution of which is linked to a Stefan problem. We then define in Section 3 four variations of a finite volume scheme for this problem, crossing time explicit or implicit schemes with centered or upwind convection operators. We then proceed to the mathematical study of the implicit upwind scheme in Section 4. To our knowledge, the convergence proof of a finite volume method for a variational inequality is new. Section 5 proposes a comparison of these finite volume schemes with other methods on some numerical examples in two space dimensions, showing that the finite volume schemes can provide accurate and cheap results in some situations.

2. The continuous problem

We briefly recall the context of this problem. An American option is a contract which gives the right to receive the payoff $h(t)$ at some time t chosen between 0 and a maturity T. This payoff $h(t)$ is then given as a function of the prices $(P_t^{(i)})_{i=1,\dots,d}$ at the time t of d financial products constituting the underlying asset. Since these prices are strictly positive, we set $X_t^{(i)} = \log(P_t^{(i)})$ for $i = 1, \dots, d$, and we express $h(t)$ under the form $h(t) = \psi(X_t)$, where $\psi : \mathbb{R}^d \to \mathbb{R}$ is a given regular function. We assume that the following stochastic differential equation is satisfied by the logarithmic transformation of the prices

$$\mathrm{d}X_t^{(i)} = (r - \lambda_i - \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2)\mathrm{d}t + \sum_{j=1}^{d}\sigma_{ij}\mathrm{d}W_t^{(j)},\ i = 1, \dots, d, \tag{1}$$

where $r \geq 0$ is the interest rate, $(\lambda_i)_{i=1,\dots,d}$ are the dividend rates, $(\sigma_{ij})_{i,j=1,\dots,d}$ is invertible matrix called the volatility matrix and $(W_t)_{t\in[0,T]}$ is a standard d-dimensional Brownian motion.

Under some assumptions on financial markets (no-arbitrage principle) [19, 22], one can take for the price of such a contract a function $U(X_t, t)$ such that

$$\frac{\partial U}{\partial t}(x,t) + AU(x,t) - rU(x,t) \leq 0, \text{ for } (x,t) \in \mathbb{R}^d \times (0,T), \tag{2}$$

$$U(x,t) \geq \psi(x), \text{ for } (x,t) \in \mathbb{R}^d \times (0,T), \tag{3}$$

$$\left(\frac{\partial U}{\partial t}(x,t) + AU(x,t) - rU(x,t))(\psi(x) - U(x,t)\right) = 0, \text{ for } (x,t) \in \mathbb{R}^d \times (0,T), \tag{4}$$

$$U(x,T) = \psi(x), \text{ for } x \in \mathbb{R}^d, \tag{5}$$

where A is the second order parabolic operator defined by

$$A \ : \ U \mapsto \sum_{i=1}^{d} \left(r - \lambda_i - \frac{1}{2} \sum_{j=1}^{d} \sigma_{ij}^2 \right) \frac{\partial U}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 U}{\partial x_i \partial x_j} \sum_{k=1}^{d} \sigma_{ik} \sigma_{jk}.$$

The problem is to obtain an approximation of $U(\bar{x}, 0)$, where $\exp(\bar{x})$ denotes the initial prices, with $\bar{x} \in \mathbb{R}^2$. Setting $u(x,t) = U(x, T-t)$, the condition (5) leads to a more usual initial condition defined at $t = 0$. Under the hypothesis that A does not depend on the space variable, it is possible to make a change of variable such that the operator A is written under the form $Au = -\mathbf{V} \cdot \nabla u + \Delta u$, with $\mathbf{V} \in \mathbb{R}^d$. For discretization purposes, we only consider the above problem on a bounded domain $\Omega \subset \mathbb{R}^d$, since it is possible to control the error thus committed (see Rem. 1 below), and we now say that $u \ : \ \Omega \times (0,T) \to \mathbb{R}$ is a weak solution of the problem on the domain $\Omega \subset \mathbb{R}^d$, if it fulfills the following conditions:

$$\frac{\partial u}{\partial t}(x,t) + \operatorname{div}(u(x,t)\mathbf{V}) - \Delta u(x,t) + r u(x,t) \geq 0, \text{ for } (x,t) \in \Omega \times (0,T), \tag{6}$$

with

$$u(x,t) \geq \psi(x), \text{ for } (x,t) \in \Omega \times (0,T). \tag{7}$$

This function u must verify

$$\left(\frac{\partial u}{\partial t}(x,t) + \operatorname{div}(u(x,t)\mathbf{V}) - \Delta u(x,t) + r u(x,t) \right) (\psi(x) - u(x,t)) = 0, \text{ for } (x,t) \in \Omega \times (0,T). \tag{8}$$

We consider the initial condition

$$u(x,0) = \psi(x), \text{ for } x \in \Omega. \tag{9}$$

and the following boundary condition

$$u(x,t) = \psi(x), \text{ for } (x,t) \in \partial\Omega \times (0,T). \tag{10}$$

Remark 1 (error committed by localization)**.** An evaluation of the error committed by considering Problem (6)–(9) on a bounded domain Ω instead of $\mathbb{R}^d$ can be done. Indeed, let us consider the case where $d = 2$, and where $\kappa > 0$ is such that $0 \leq \psi(x) \leq \kappa$ for all $x \in \mathbb{R}^2$. Let us denote by $(\exp(\bar{x}_1), \exp(\bar{x}_2))$ the initial stock prices, and let us set $\bar{x} = (\bar{x}_1, \bar{x}_2)$. We denote, for some $R > 0$, by $\Omega = (\bar{x}_1 - R, \bar{x}_1 + R) \times (\bar{x}_2 - R, \bar{x}_2 + R)$. Let u be the solution (in the sense given below) of (6)–(10), and let $\widehat{u}$ be the solution (in an appropriate weak sense) of (6)–(9) with $\Omega = \mathbb{R}^2$. Then the following inequality holds:

$$|\widehat{u}(\bar{x},t) - u(\bar{x},t)| \leq 4\kappa \left[2 - \mathcal{N}\left(\frac{R - T|V_1|}{\sqrt{2T}} \right) - \mathcal{N}\left(\frac{R - T|V_2|}{\sqrt{2T}} \right) \right], \ \forall t \in [0,T],$$

where $\mathcal{N}$ is the repartition function of a standard normal distribution. The detailed proof of the above inequality is given in [7], as well as the results which can be established for other financial mathematics problems.

The following assumptions are done in this paper.

Assumption 1.

(1) $d \in \mathbb{N}_*$;
(2) $\Omega \subset \mathbb{R}^d$ *is a bounded open polygonal domain*;
(3) $T > 0$;
(4) $\psi \in C^2(\overline{\Omega}) \cap H_0^1(\Omega)$ *is such that* $\psi \geq 0$ *on* Ω;
(5) $\mathbf{V} \in \mathbb{R}^d$ *and* $r \in \mathbb{R}_+$.

We say that u is a weak solution of the problem (6)–(10) if it meets the following variational inequality:

$$\begin{cases} u \in L^2(0,T;H_0^1(\Omega)), \frac{\partial u}{\partial t} \in L^2(\Omega\times(0,T)), u(x,0)=\psi(x), \text{ for } a.e.\ x\in\Omega, \\ u(x,t)\geq \psi(x), \text{ for } a.e.\ (x,t)\in\Omega\times(0,T), \\ \displaystyle\int_0^T\int_\Omega \left[\begin{array}{l} \left(\frac{\partial u}{\partial t}(x,t) + ru(x,t) + \mathrm{div}(u(x,t)\mathbf{V})\right)(v(x,t)-u(x,t)) \\ +\nabla u(x,t)\cdot\nabla(v(x,t)-u(x,t)) \end{array} \right] \mathrm{d}x\mathrm{d}t \geq 0, \\ \forall v\in L^2(0,T;H_0^1(\Omega)) \text{ such that } v(x,t)\geq\psi(x) \text{ for } \ a.e.\ (x,t)\in\Omega\times(0,T). \end{cases} \tag{11}$$

Under Assumption 1, it can be proven that there exists a unique solution u to (11) which is a classical obstacle problem (see [19, 21]). This solution u is in fact closely related to that of a Stefan problem, which is the nonlinear degenerate parabolic problem describing the energy conservation in a material which changes of thermodynamical state (see [6]). Indeed, let us assume for simplicity that $\mathbf{V}=0$ and $r=0$, and let us set $w_0=\Delta\psi$ (thus $w_0\in L^2(\Omega)$). Let us consider the function w solution of the one-phase Stefan problem $w_t-\Delta\max(w,0)=0$ in $\Omega\times(0,T)$, $\max(w,0)=0$ on $\partial\Omega\times(0,T)$, $w(\cdot,0)=w_0$ on Ω in the following weak sense: $w\in L^2(\Omega\times(0,T))$ is such that $\max(w,0)\in L^2(0,T;H_0^1(\Omega))$ and, for all $\varphi\in C_c^\infty(\Omega\times(-\infty,T))$ ($C_c^\infty(\Omega\times(-\infty,T))$ is the set of indefinitely differentiable functions with a compact support in $\Omega\times(-\infty,T)$),

$$\int_0^T\int_\Omega [w(x,t)\varphi_t(x,t) - \nabla\max(w,0)(x,t)\cdot\nabla\varphi(x,t)]\,\mathrm{d}x\mathrm{d}t + \int_\Omega w_0(x)\varphi(x,0)\,\mathrm{d}x = 0.$$

It is then proven in [6] that the solution u of (11) is such that, for all $(x,t)\in\Omega\times(0,T)$, $u(x,t)=\int_0^t\max(w(x,s),0)\,\mathrm{d}s+\psi(x)$, and $w(x,t)=\Delta u(x,t)$ for $a.e.$ $(x,t)\in\Omega\times(0,T)$. Since the Stefan problem is based on a conservation equation, it has been shown that the use of the finite volume method, which satisfies a local conservation property, accurately respects the location of the moving boundary (see [4, 15]).

3. The finite volume schemes

In order to obtain a numerical approximation of the solution of (11), let us first summarize the definition given in [15] of an admissible space and time discretization of $\Omega\times(0,T)$.

Definition 1 (space-time discretization of $\Omega\times(0,T)$)**.** An admissible finite volume discretization $\mathcal{D}$ of $\Omega\times(0,T)$ is a family $\mathcal{D}=(\mathcal{T},\mathcal{E},(x_K)_{K\in\mathcal{T}},N)$, where $\mathcal{T}$ (the family of the control volumes), $\mathcal{E}$ (the family of the edges) and $(x_K)_{K\in\mathcal{T}}$ (the family of the centers of the control volumes) are such that $(\mathcal{T},\mathcal{E},(x_K)_{K\in\mathcal{T}})$ is an admissible mesh of Ω (see [15]); in the following, we shall improperly denote only by $\mathcal{T}$ the admissible mesh $(\mathcal{T},\mathcal{E},(x_K)_{K\in\mathcal{T}})$. The integer $N\in\mathbb{N}$ is given, and the value of the time step is defined by $\delta t=\frac{T}{N+1}$. The main property of an admissible mesh is that, for two neighboring control volumes K and L, the line (x_K,x_L) is orthogonal to the common edge of K and L, denoted by $K|L$, and we set $d_{K|L}=\mathrm{d}(x_K,x_L)$. For all $K\in\mathcal{T}$, we denote by m_K the measure of K, $N_K\subset\mathcal{T}$ is the set of neighbors of K, $\mathcal{E}_K\subset\mathcal{E}$ is the set of the edges of K, $\mathcal{E}_{\mathrm{int}}$ (resp. $\mathcal{E}_{\mathrm{ext}}$) is the subset of $\mathcal{E}$ constituted by the interior edges (resp. exterior). The measure of $\sigma\in\mathcal{E}$ is denoted by m_σ. For $K\in\mathcal{T}$ and $\sigma\in\mathcal{E}_K$, $\mathbf{n}_{K\sigma}$ is the unit vector normal to σ and outward to K; we denote $\mathcal{E}_{K,\mathrm{int}}=\mathcal{E}_K\cap\mathcal{E}_{\mathrm{int}}$ and $\mathcal{E}_{K,\mathrm{ext}}=\mathcal{E}_K\cap\mathcal{E}_{\mathrm{ext}}$, and for $\sigma\in\mathcal{E}_{K,\mathrm{ext}}$, we set $d_\sigma=\mathrm{d}(x_K,\sigma)$. For a given admissible discretization $\mathcal{D}$ of $\Omega\times(0,T)$, one defines:

$$\begin{aligned} &\mathrm{size}(\mathcal{D}) = \max(\mathrm{size}(\mathcal{T}),\delta t), \\ &\mathrm{reg}(\mathcal{D}) = \mathrm{reg}(\mathcal{T}) = \max\{\tfrac{d_{K|L}}{\mathrm{d}(x_K,K|L)}, \tfrac{\mathrm{diam}(K)}{d_\sigma},\ K\in\mathcal{T},\ L\in N_K,\ \sigma\in\mathcal{E}_K\}, \end{aligned}$$

with $\mathrm{size}(\mathcal{T})=\max\{\mathrm{diam}(K),\ K\in\mathcal{T}\}$.

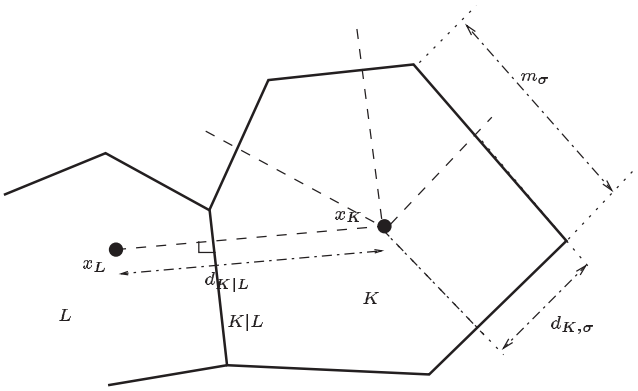

FIGURE 1. Notations for an admissible mesh.

An example of two neighboring control volumes K and L of $\mathcal{T}$ is depicted in Figure 1.

Remark 2 (variable time steps)**.** We could as well consider variable time steps. The hypothesis of a constant time step is only done here for the sake of simplicity.

Let us now introduce the space of piecewise constant functions associated with an admissible mesh and some "discrete $H_0^1(\Omega)$" norm for this space. This discrete norm will be used in the estimates on the approximate solution given by a finite volume scheme.

Definition 2 (discrete approximation function space)**.** Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$, and let $(\mathcal{T}, \mathcal{E}, (x_K)_{K\in\mathcal{T}})$ be an admissible mesh of Ω. We define $X(\mathcal{T})$ as the set of functions from Ω to $\mathbb{R}$ which are constant over each control volume of the mesh. For all $u \in X(\mathcal{T})$ and $K \in \mathcal{T}$, we denote by u_K the constant value of $u(x)$ for a.e. $x \in K$, and for all $\varphi \in C(\Omega)$, we define $P_\mathcal{T}\varphi \in X(\mathcal{T})$ by $P_\mathcal{T}\varphi_K = \varphi(x_K)$.

Definition 3 (discrete scalar product and norm)**.** Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$, and $\mathcal{T}$ an admissible mesh. We define the following scalar product on $X(\mathcal{T})$ by

$$[u,v]_{1,\mathcal{T}} = \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \frac{m_{K|L}}{d_{K|L}}(u_L-u_K)(v_L-v_K) + \sum_{\substack{K\in\mathcal{T}\\ \sigma\in\mathcal{E}_{K,\rm ext}}} \frac{m_\sigma}{d_\sigma}u_K v_K,\ \forall u,v\in X(\mathcal{T}), \tag{12}$$

and we denote by $\|\cdot\|_{1,\mathcal{T}}$ the discrete associated norm.

We recall a lemma proven in [17], which is useful to get convergence properties for approximate solutions given by a finite volume scheme.

Lemma 1 (relative compactness in $L^2(\Omega)$)**.** *Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$, with $d \in \mathbb{N}_*$. We consider a sequence $(\mathcal{T}_m, u_m)_{m\in\mathbb{N}}$ such that, for all $m \in \mathbb{N}$, $\mathcal{T}_m$ is an admissible finite volume discretization of Ω and $u_m \in X(\mathcal{T}_m)$. Let us assume that*

$$\lim_{m\longrightarrow\infty} \mathrm{size}(\mathcal{T}_m) = 0,$$

and that there exists $C > 0$ such that, for all $m \in \mathbb{N}$, $\|u_m\|_{1,\mathcal{T}_m} \le C$. Then there exist a subsequence of $(\mathcal{T}_m, u_m)_{m\in\mathbb{N}}$, again denoted $(\mathcal{T}_m, u_m)_{m\in\mathbb{N}}$, and $u \in H_0^1(\Omega)$ such that u_m tends to u in $L^2(\Omega)$ as $m \longrightarrow \infty$,

$$\lim_{m\longrightarrow\infty} [u_m, P_{\mathcal{T}_m}\varphi]_{1,\mathcal{T}_m} = \int_\Omega \nabla u(x)\cdot\nabla\varphi(x)\mathrm{d}x,\ \forall\varphi\in C^2(\overline{\Omega})\cap H_0^1(\Omega), \tag{13}$$

and

$$\int_\Omega (\nabla u(x))^2 \mathrm{d}x \leq \liminf_{m\longrightarrow\infty} \|u_m\|_{1,\mathcal{T}_m}. \tag{14}$$

Let us now turn to the approximation of the convective term.

Definition 4 (upwind convection operator)**.** Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$, and $\mathcal{T}$ an admissible mesh. Let $\mathbf{V} \in \mathbb{R}^d$. We define the upwind convection operator $\nabla_\mathcal{T}^\mathbf{V} \ : \ X(\mathcal{T}) \to X(\mathcal{T})$ by

$$(\nabla_\mathcal{T}^\mathbf{V} u)_K = \frac{1}{m_K} \sum_{\sigma\in\mathcal{E}_K} V_{K\sigma} u_\sigma, \ \forall u \in X(\mathcal{T}), \ \forall K \in \mathcal{T}, \tag{15}$$

with

$$\begin{array}{l} V_{K\sigma} = \int_\sigma \mathbf{V}.\mathbf{n}_{K\sigma} d\gamma(x) = \mathbf{V}.\mathbf{n}_{K\sigma} m_\sigma, \ \forall K \in \mathcal{T}, \ \forall \sigma \in \mathcal{E}_K \\ V_{KL} = V_{K\ K|L}, \ \forall K \in \mathcal{T}, \ \forall L \in N_K, \end{array}$$

and denoting, for all $u \in X(\mathcal{T})$,

$$u_\sigma = \begin{cases} u_K & \text{if } \sigma \in \mathcal{E}_{\rm int}, \sigma = K|L, V_{K\sigma} \geq 0, \\ u_L & \text{if } \sigma \in \mathcal{E}_{\rm int}, \sigma = K|L, V_{K\sigma} < 0, \\ u_K & \text{if } \sigma \in \mathcal{E}_{K,\rm ext}, V_{K\sigma} \geq 0, \\ 0 & \text{if } \sigma \in \mathcal{E}_{K,\rm ext}, V_{K\sigma} < 0. \end{cases} \tag{16}$$

Remark 3. According to the definition of an admissible mesh and of $V_{K\sigma}$, for $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$, one gets:

$$\sum_{\sigma\in\ \mathcal{E}_K} V_{K\sigma} \sum_{\sigma\in\ \mathcal{E}_K} \left((V_{K\sigma})^+ - (V_{K\sigma})^-\right) \sum_{\sigma\in\ \mathcal{E}_{K,\rm int}} V_{K\sigma} + \sum_{\sigma\in\ \mathcal{E}_{K,\rm ext}} V_{K\sigma} = 0. \tag{17}$$

We set $\|\mathbf{V}\|_{L^2(\Omega)} = V$.

The above definition meets the following lemma, which depicts the convergence properties of the convection operator in presence of usual estimates for parabolic problems.

Lemma 2 (properties of the discrete convection operator $\nabla_\mathcal{T}^\mathbf{V}$)**.** *Under Assumptions 1, let $\mathcal{T}$ be an admissible mesh of Ω. Then the following relations hold*

$$\int_\Omega u(x) \nabla_\mathcal{T}^\mathbf{V} u(x) \geq 0, \ \forall u \in X(\mathcal{T}) \tag{18}$$

and

$$\|\nabla_\mathcal{T}^\mathbf{V} u\|_{L^2(\Omega)} \leq \sqrt{2d \ \mathrm{reg}(\mathcal{T})} V \|u\|_{1,\mathcal{T}}, \ \forall u \in X(\mathcal{T}). \tag{19}$$

Moreover, for a sequence $(\mathcal{T}_m, u_m)_{m\in\mathbb{N}}$ such that, for all $m \in \mathbb{N}$, $\mathcal{T}_m$ is an admissible mesh of Ω such that $\lim_{m\longrightarrow\infty} \mathrm{size}(\mathcal{T}_m) = 0$, and, for all $m \in \mathbb{N}$, $u_m \in X(\mathcal{T}_m)$ is such that there exists $C > 0$ and $u \in H_0^1(\Omega)$ with, for all $m \in \mathbb{N}$, $\|u_m\|_{1,\mathcal{T}_m} \leq C$ and u_m tends to u in $L^2(\Omega)$ as $m \longrightarrow \infty$.

Then

$$\lim_{m\longrightarrow\infty} \int_\Omega P_{\mathcal{T}_m} \varphi(x) \nabla_{\mathcal{T}_m}^\mathbf{V} u_m(x) \mathrm{d}x = \int_\Omega \varphi(x) \mathbf{V} \cdot \nabla u(x) \mathrm{d}x, \ \forall \varphi \in C^1(\overline{\Omega}). \tag{20}$$

Proof. We have, for all $u \in X(\mathcal{T})$, $\int_\Omega u(x) \nabla_\mathcal{T}^\mathbf{V} u(x) = \sum_{K\in\mathcal{T}} u_K \sum_{\sigma\in\mathcal{E}_K} V_{K\sigma} u_\sigma T_1 + T_2$, where, using Remark 3,

$$T_1 = \sum_{K\in\mathcal{T}} u_K \sum_{L\in N_K} V_{KL}(u_{K|L} - u_K)$$

and

$$T_2 = \sum_{K\in\mathcal{T}} u_K \sum_{\sigma\in\mathcal{E}_{K,\text{ext}}} V_{K\sigma}(u_\sigma - u_K).$$

We then have

$$T_1 = \sum_{K\in\mathcal{T}} u_K \sum_{L\in N_K} \left(V_{KL}^-(u_K - u_L)\right) = \frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{L\in N_K} \left(V_{KL}^-(u_K - u_L)^2\right) + \frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{L\in N_K} \left(V_{KL}^-(u_K^2 - u_L^2)\right),$$

which gives

$$T_1 \geq \frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{L\in N_K} V_{KL}(u_{K|L}^2 - u_K^2).$$

On the other hand, we have

$$T_2 = \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_{K,\text{ext}}} V_{K\sigma}^- u_K^2 \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_{K,\text{ext}}} V_{K\sigma}(u_\sigma^2 - u_K^2).$$

Gathering the above results, we get

$$\int_\Omega u(x)\nabla_{\mathcal{T}}^{\mathbf{V}} u(x) \geq \frac{1}{2}\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} V_{K\sigma}u_\sigma^2 = 0,$$

which concludes the proof of (18). Let us now turn to the proof of (19). Let $u \in X(\mathcal{T})$ and $K \in \mathcal{T}$. We have

$$m_K(\nabla_{\mathcal{T}}^{\mathbf{V}} u)_K^2 \frac{1}{m_K}\left(\sum_{L\in N_K}(V_{KL})^-(u_K - u_L) + \sum_{\sigma\in\mathcal{E}_{K,\text{ext}}}(V_{K\sigma})^- u_K\right)^2.$$

Thanks to the Cauchy-Schwarz inequality, we get

$$\begin{aligned} m_K(\nabla_{\mathcal{T}}^{\mathbf{V}} u)_K^2 \leq \quad & \frac{1}{m_K}\left(\sum_{L\in N_K}(V_{KL})^- d_{K|L} + \sum_{\sigma\in\mathcal{E}_{K,\text{ext}}}(V_{K\sigma})^- d_\sigma\right) \\ & \times\left(\sum_{L\in N_K}\frac{(V_{KL})^-}{d_{K|L}}(u_K - u_L)^2 + \sum_{\sigma\in\mathcal{E}_{K,\text{ext}}}\frac{(V_{K\sigma})^-}{d_\sigma}u_K^2\right). \end{aligned}$$

On the one hand, we have $\frac{(V_{KL})^-}{d_{K|L}} \leq V\frac{m_{K|L}}{d_{K|L}}$ and $\frac{(V_{K\sigma})^-}{d_\sigma} \leq V\frac{m_\sigma}{d_\sigma}$. On the other hand, we have

$$\sum_{L\in N_K}(V_{KL})^- d_{K|L} + \sum_{\sigma\in\mathcal{E}_{K,\text{ext}}}(V_{K\sigma})^- d_\sigma \leq V\left(\sum_{L\in N_K} m_{K|L}d_{K|L} + \sum_{\sigma\in\mathcal{E}_{K,\text{ext}}} m_\sigma d_\sigma\right) \leq Vd\ \text{reg}(\mathcal{T})m_K.$$

Thanks to the above inequalities, summing on $K \in \mathcal{T}$, we thus conclude (19). The proof of (20) is classical (see [15]) for the upwind convection operator, and this property holds in fact under the sufficient hypothesis

$$\sum_{K\in\mathcal{T}}\sum_{L\in N_K} |V_{KL}|\ |u_L - u_K| \leq C_1\,\text{size}(\mathcal{T})^{-\alpha},$$

where C_1 is a real which does not depend on $\mathcal{T}$ and $\alpha \in [0,1)$. This property is proven for the finite volume solution of a scalar hyperbolic equation with $\alpha = 1/2$, and it holds under the hypothesis $\|u_m\|_{1,\mathcal{T}_m} \leq C$ with $\alpha = 0$ (thanks to the Cauchy-Schwarz inequality). □

We can as well consider the following centered approximation of the convective term.

Definition 5 (centered convection operator)**.** We define the centered finite convection operator $\bar{\nabla}_{\mathcal{T}}^{\mathbf{V}} : X(\mathcal{T}) \to X(\mathcal{T})$ by replacing u_σ in (15) by

$$\bar{u}_\sigma = \begin{cases} \frac{1}{2}(u_K + u_L) & \text{if } \sigma \in \mathcal{E}_{\rm int}, \sigma = K|L, \\ \frac{1}{2}u_K & \text{if } \sigma \in \mathcal{E}_{K,\rm ext}. \end{cases}$$

Let $\mathcal{D}$ be a finite volume discretization of $\Omega \times (0,T)$ in the sense of Definition 1. Let us now define four finite volume schemes to discretize the problem (6)–(10), the unknown of which is $u^{n+1} \in X(\mathcal{T})$ (the discrete unknowns are then $(u_K^{n+1})_{K\in\mathcal{T},n\in\{0,\ldots,N\}}$).

The *implicit upwind* finite volume scheme is given by

$$u^0 = P_{\mathcal{T}}\psi, \tag{21}$$

$$u^{n+1} = \max\left(\tilde{u}^{n+1}, P_{\mathcal{T}}\psi\right), \forall n \in \{0,\ldots,N\}, \tag{22}$$

$$\begin{aligned} &\int_\Omega \left(\tilde{u}^{n+1}(x) - u^n(x) + \delta t \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x) + r\delta t u^{n+1}(x)\right) v(x)\mathrm{d}x + \delta t [u^{n+1}, v]_{1,\mathcal{T}} = 0, \\ &\forall v \in X(\mathcal{T}), \forall n \in \{0,\ldots,N\}. \end{aligned} \tag{23}$$

Setting $v = 1_K$, for all $K \in \mathcal{T}$ (1_K denotes the characteristic function of K), in (23), we get as many equations as unknowns:

$$\begin{aligned} &m_K\left(\tilde{u}_K^{n+1} - u_K^n\right) + \delta t \sum_{\sigma\in\mathcal{E}_K} V_{K\sigma} u_\sigma^{n+1} + \delta t [u^{n+1}, 1_K]_{1,\mathcal{T}} + r\delta t m_K u_K^{n+1} = 0, \\ &\forall K \in \mathcal{T}, \forall n \in \{0,\ldots,N\}. \end{aligned}$$

The *explicit upwind* finite volume scheme is given by (21), (22) and

$$\begin{aligned} &\int_\Omega \left(\tilde{u}^{n+1}(x) - u^n(x) + \delta t \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n}(x) + r\delta t u^{n}(x)\right) v(x)\mathrm{d}x + \delta t [u^{n}, v]_{1,\mathcal{T}} = 0, \\ &\forall v \in X(\mathcal{T}), \forall n \in \{0,\ldots,N\}. \end{aligned} \tag{24}$$

The *implicit centered* finite volume scheme is given by (21), (22) and

$$\begin{aligned} &\int_\Omega \left(\tilde{u}^{n+1}(x) - u^n(x) + \delta t \bar{\nabla}_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x) + r\delta t u^{n+1}(x)\right) v(x)\mathrm{d}x + \delta t [u^{n+1}, v]_{1,\mathcal{T}} = 0, \\ &\forall v \in X(\mathcal{T}), \forall n \in \{0,\ldots,N\}. \end{aligned} \tag{25}$$

The *explicit centered* finite volume scheme is given by (21), (22) and

$$\begin{aligned} &\int_\Omega \left(\tilde{u}^{n+1}(x) - u^n(x) + \delta t \bar{\nabla}_{\mathcal{T}}^{\mathbf{V}} u^{n}(x) + r\delta t u^{n}(x)\right) v(x)\mathrm{d}x + \delta t [u^{n}, v]_{1,\mathcal{T}} = 0, \\ &\forall v \in X(\mathcal{T}), \forall n \in \{0,\ldots,N\}. \end{aligned} \tag{26}$$

The mathematical analysis of each of the four versions for the finite volume scheme given above can be done. All of these versions lead to L^∞ stability results, which hold under a Péclet condition on the mesh in the case of a centered scheme, namely

$$\frac{1}{2}V_{K\sigma} \le \frac{m_\sigma}{d_{K\sigma}}, \ \forall K \in \mathcal{T}, \ \forall \sigma \in \mathcal{E}_K,$$

(this inequality is satisfied under the sufficient condition $V\text{size}(\mathcal{T}) \leq 2$) and under a CFL condition, which is given in the case of the explicit centered version by

$$\delta t \leq \frac{m_K}{m_K r + \sum_{\sigma\in\mathcal{E}}(\frac{m_\sigma}{d_\sigma} + \frac{1}{2}V_{K\sigma})}, \ \forall K \in \mathcal{T},$$

and in the case of the explicit upwind version by

$$\delta t \leq \frac{m_K}{m_K r + \sum_{\sigma\in\mathcal{E}}(\frac{m_\sigma}{d_\sigma} + V^+_{K\sigma})}, \ \forall K \in \mathcal{T}.$$

These CFL conditions are then satisfied under sufficient conditions under the form $\delta t \leq C\text{size}(\mathcal{T})^2$. We prove in the next section the L^∞ stability of the upwind implicit scheme, together with the existence and uniqueness of a discrete solution, followed by the proof of its convergence. Such a proof can be done for the four versions (see [7]), and we restrict here the mathematical study to the simpler case, which is nevertheless sufficient to point out the difficulties due to the nonlinearity of the problem. We thus define the approximate solution in the particular case of the upwind implicit scheme.

Definition 6 (approximate solution given by the upwind implicit scheme)**.** Let $\mathcal{D}$ be an admissible finite volume discretization of $\Omega \times (0,T)$ in the sense of Definition 1. The approximate solution (continuous with respect to the time on $\Omega \times (0,T)$) of (6)–(10) associated to the discretization $\mathcal{D}$ is defined almost everywhere in $\Omega \times (0,T)$ by:

$$u_\mathcal{D}(x,t) = \frac{t - n\delta t}{\delta t}u^{n+1}(x) + \frac{(n+1)\delta t - t}{\delta t}u^n(x), \text{ for } a.e. \ (x,t) \in \Omega \times [n\delta t, (n+1)\delta t], \forall n \in \{0,\dots,N\},$$

where $(u^{n+1})_{n\in\{0,\dots,N\}}$ is the unique solution to (21)–(23) (thanks to Lem. 3 below).

Thanks to this Definition, one gets almost everywhere in $\Omega \times (0,T)$:

$$\frac{\partial u_\mathcal{D}}{\partial t}(x,t) = \frac{u^{n+1}(x) - u^n(x)}{\delta t}, \text{ for } a.e. \ t \in (n\delta t, (n+1)\delta t), \text{ for } a.e. \ x \in \Omega, \forall n \in \{0,\dots,N\}.$$

4. Mathematical study of the upwind implicit scheme

We now state the existence and the uniqueness of a discrete solution to the upwind implicit scheme (21)–(23), the proof of which gives at the same time the L^∞ stability of the scheme.

Lemma 3 (existence and uniqueness and L^∞ stability)**.** *Under Assumptions 1, let $\mathcal{D}$ be an admissible finite volume discretization of $\Omega \times (0,T)$ in the sense of Definition 1. Then the system of equations (21)–(23) has one and only one solution $\left(u^{n+1}\right)_{n\in\{0,\dots,N\}}$, such that*

$$P_\mathcal{T}\psi(x) \leq u^n(x) \leq \max_{x\in\overline{\Omega}}\psi(x), \ \textit{for a.e. } x \in \Omega, \ \forall n \in \{0,\dots,N+1\}. \tag{27}$$

Proof. Let us deal with the proof of the existence of a solution. We first consider, for a given $K \in \mathcal{T}$, the application $\tilde{f}_K : \mathbb{R} \to \mathbb{R}$ defined by

$$\tilde{f}_K \ : \ s \mapsto s + \max(s, P_\mathcal{T}\psi_K)\left(r\delta t + \frac{\delta t}{m_K}\sum_{\sigma\in\mathcal{E}_K}\left(\frac{m_\sigma}{d_\sigma} + (V_{K\sigma})^+\right)\right).$$

Since this function is strictly increasing and continuous and verifies $\lim_{x\to+\infty}\tilde{f}_K(x) = +\infty$, $\lim_{x\to-\infty}\tilde{f}_K(x) = -\infty$, it is therefore invertible and its reciprocal function $\tilde{f}_K^{(-1)}$ is strictly increasing and continuous as well.

Let us denote $M = \max_{x\in\overline{\Omega}} \psi(x)$. We now assume that, for a given $n \in \mathbb{N}$, a family of values $u_K^n \in [0, M]$ is given. We define the fix point application $F : X(\mathcal{T}) \to X(\mathcal{T})$, $F : u \mapsto \hat{u}$ such that

$$\hat{u}_K = \max\left\{P_{\mathcal{T}}\psi_K, \tilde{f}_K^{(-1)}\left(u_K^n + \frac{\delta t}{m_K}\sum_{L\in N_K} u_L\left((V_{KL})^- + \frac{m_{K|L}}{d_{K|L}}\right)\right)\right\},\ \forall K \in \mathcal{T}.$$

Let us assume that, for all $K \in \mathcal{T}$, $u_K \in [0, M]$. By construction, we have $\hat{u}_K \geq 0$ for all $K \in \mathcal{T}$. Let $K \in \mathcal{T}$ be such that $\hat{M} = \hat{u}_K = \max\{\hat{u}_L,\ L \in \mathcal{T}\}$. Let us assume that $\hat{M} > M$. Then this value, which is greater than $P_{\mathcal{T}}\psi_K$, must satisfy

$$\hat{M} = \tilde{f}_K^{(-1)}\left(u_K^n + \frac{\delta t}{m_K}\sum_{L\in N_K} u_L\left((V_{KL})^- + \frac{m_{K|L}}{d_{K|L}}\right)\right)$$

and therefore

$$\tilde{f}_K(\hat{M}) \leq M + \frac{\delta t}{m_K}\sum_{L\in N_K} M\left((V_{KL})^- + \frac{m_{K|L}}{d_{K|L}}\right).$$

But, using the definition of $\tilde{f}_K$, we get

$$\tilde{f}_K(\hat{M}) = \hat{M}\left(1 + r\delta t + \frac{\delta t}{m_K}\sum_{\sigma\in\mathcal{E}_K}\left(\frac{m_\sigma}{d_\sigma} + (V_{K\sigma})^+\right)\right),$$

which is impossible since $r \geq 0$ and

$$\sum_{\sigma\in\mathcal{E}_K}\left(\frac{m_\sigma}{d_\sigma} + (V_{K\sigma})^+\right) \geq \sum_{L\in N_K}\left((V_{KL})^- + \frac{m_{K|L}}{d_{K|L}}\right).$$

Therefore $\max\{\hat{u}_L,\ L \in \mathcal{T}\} \leq M$. We can then apply the Brouwer fixed point theorem, since the continuous function F is such that the image by F of the set $\{u \in X(\mathcal{T}), 0 \leq u \leq M \text{ a.e. in } \Omega\})$ is included in the same set. This proves the existence in this set of at least one $u \in X(\mathcal{T})$ such that $u = F(u)$ and therefore u satisfies $P_{\mathcal{T}}\psi(x) \leq u(x) \leq M$ for a.e. $x \in \Omega$. We then remark that, defining $u^{n+1} = u$, this function u^{n+1} is a solution to the system of equations (21)–(23), since, setting

$$\tilde{u}_K^{n+1} = \tilde{f}_K^{(-1)}\left(u_K^n + \frac{\delta t}{m_K}\sum_{L\in N_K} u_L^{n+1}\left((V_{KL})^- + \frac{m_{K|L}}{d_{K|L}}\right)\right),$$

we get

$$\tilde{f}_K(\tilde{u}_K^{n+1}) = \tilde{u}_K^{n+1} + u_K^{n+1}\left(r\delta t + \frac{\delta t}{m_K}\sum_{\sigma\in\mathcal{E}_K}\left(\frac{m_\sigma}{d_\sigma} + (V_{K\sigma})^+\right)\right).$$

Let us now prove the uniqueness of the discrete solution to the upwind implicit scheme, by adapting the technique used for the proof of Lemma 3.2 in [15]. Assume that, for $n \in \{0, \ldots, N\}$, $u^n \in X(\mathcal{T})$ is given and that $u, \tilde{u} \in X(\mathcal{T})$ and $v, \tilde{v} \in X(\mathcal{T})$ are two solutions of (22)–(23). We subtract (23) checked by $v, \tilde{v}$ from (23) checked by $u, \tilde{u}$, taking $w = u - v$ as test function. This yields, setting $\tilde{w} = \tilde{u} - \tilde{v}$:

$$\int_\Omega \left(\tilde{w}(x) + r\delta t w(x) + \delta t \nabla_{\mathcal{T}}^{\mathbf{V}} w(x)\right) w(x)\mathrm{d}x + \delta t[w, w]_{1,\mathcal{T}} = 0. \tag{28}$$

Hence, thanks to Lemma 2, we get $\int_\Omega \tilde{w}(x)w(x)\mathrm{d}x \leq 0$. Since, for all $a, b, c \in \mathbb{R}$, we have $(a-b)(\max(a,c) - \max(b,c)) \geq (\max(a,c) - \max(b,c))^2$, we then get $\int_\Omega w(x)^2\mathrm{d}x \leq 0$, which proves that the solution u^{n+1} of (22)–(23) is unique.

Let us now state a technical lemma concerning the interpolation $P_{\mathcal{T}}\psi$ of ψ, which is necessary in the course of the convergence proof. □

Lemma 4.
Under Assumptions 1, let $\mathcal{D}$ be an admissible finite volume discretization of $\Omega \times (0,T)$ in the sense of Definition 1. Then,

$$(\|P_{\mathcal{T}}\psi\|_{1,\mathcal{T}})^2 \leq \left(\|\nabla\psi\|_{L^\infty(\Omega)}\right)^2 d\ m(\Omega).$$

Proof. We have

$$\begin{aligned}
[P_{\mathcal{T}}\psi, P_{\mathcal{T}}\psi]_{1,\mathcal{T}} &= \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \frac{m_{K|L}}{d_{K|L}} \left(\psi(x_L)-\psi(x_K)\right)^2 + \sum_{\substack{K\in\mathcal{T}\\ \sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm ext}}} \frac{m_\sigma}{d_\sigma}\psi(x_{K\sigma})^2 \\
&\leq \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \frac{m_{K|L}}{d_{K|L}} \left(\|\nabla\psi\|_{L^\infty(\Omega)}\right)^2 (d_{KL})^2 + \sum_{\substack{K\in\mathcal{T}\\ \sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm ext}}} \frac{m_\sigma}{d_\sigma} \left(\|\nabla\psi\|_{L^\infty(\Omega)}\right)^2 (d_\sigma)^2.
\end{aligned}$$

We conclude, noting that $\sum_{\sigma\in\mathcal{E}} m_\sigma d_\sigma = d\ m(\Omega)$. □

We can now state a discrete $L^2(0,T;H_0^1(\Omega))$ estimate, classical in the case of parabolic problems.

Lemma 5 (discrete $L^2(0,T;H_0^1(\Omega))$ estimate)**.** *Under Assumptions 1, let $\mathcal{D}$ be a discretization of $\Omega\times(0,T)$ in the sense of Definition 1, and let $(u^n)_{n\in\{0,\dots,N+1\}}$ be the unique solution of the upwind implicit scheme (21)–(23). Then there exists C_2 only depending on Ω, T, ψ,d and V such that*

$$\sum_{n=0}^{N} \delta t \left(\|u^{n+1}\|_{1,\mathcal{T}}\right)^2 \leq C_2\,. \tag{29}$$

Proof. We take $u^{n+1}-u^0$ as test function in (23) and, summing on $n=0,\dots,N$, we get $T_3+T_4+T_5+T_6=0$, with

$$T_3 = \sum_{n=0}^{N} \int_\Omega (\tilde{u}^{n+1}(x)-u^n(x))(u^{n+1}(x)-u^0(x))\mathrm{d}x,$$

$$T_4 = r\delta t \sum_{n=0}^{N} \int_\Omega u^{n+1}(x)(u^{n+1}(x)-u^0(x))\mathrm{d}x,$$

$$T_5 = \delta t \sum_{n=0}^{N} \int_\Omega \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x)(u^{n+1}(x)-u^0(x))\mathrm{d}x,$$

and

$$T_6 = \sum_{n=0}^{N} \delta t [u^{n+1}, u^{n+1}-u^0]_{1,\mathcal{T}}.$$

We first notice that, since $u^0 = P_{\mathcal{T}}\psi$, we have

$$\left(\tilde{u}^{n+1}(x)-u^n(x)\right)\left(u^{n+1}(x)-u^0(x)\right)\left(u^{n+1}(x)-u^n(x)\right)\left(u^{n+1}(x)-u^0(x)\right),\ \text{for a.e. } x\in\Omega,\ \forall n\in\{0,\dots,N\}.$$

This leads to

$$T_3 = \sum_{n=0}^{N} \int_\Omega (u^{n+1}(x)-u^n(x))(u^{n+1}(x)-u^0(x))\mathrm{d}x,$$

and therefore

$$\begin{aligned} T_3 &= \sum_{n=0}^{N} \int_\Omega (u^{n+1}(x) - u^0(x) + u^0(x) - u^n(x))(u^{n+1}(x) - u^0(x)) \mathrm{d}x \\ &= \frac{1}{2} \int_\Omega \left((u^{N+1}(x) - u^0(x))^2 + \sum_{n=0}^{N} (u^{n+1}(x) - u^n(x))^2 \right) \mathrm{d}x, \end{aligned}$$

providing

$$T_3 \geq 0.$$

Since we have $u^{n+1}(x) \geq u^0(x) \geq 0$ for *a.e.* $x \in \Omega$, we get

$$T_4 \geq 0.$$

We have, using Lemma 2,

$$T_5 \geq -\delta t \sum_{n=0}^{N} \int_\Omega \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x) u^0(x) \mathrm{d}x.$$

We then remark that

$$\int_\Omega \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x) u^0(x) \mathrm{d}x \sum_{\substack{\sigma \in \mathcal{E}_{\mathrm{int}} \\ \sigma = K|L}} V_{KL} u_{K|L}^{n+1} (u_K^0 - u_L^0) + \sum_{\substack{K \in \mathcal{T} \\ \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}}} V_{K\sigma} u_\sigma^{n+1} u_K^0.$$

Since, for $\sigma \in \mathcal{E}_{\mathrm{int}}$ with $\sigma = K|L$, we have $|V_{KL}| \leq V m_{K|L}$ and $|u_K^0 - u_L^0| \leq \|\nabla \psi\|_{L^\infty(\Omega)} d_{K|L}$ and for $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}$, we have $|V_{K\sigma}| \leq V m_\sigma$ and $|u_K^0| \leq \|\nabla \psi\|_{L^\infty(\Omega)} d_\sigma$, we then get, using $|u_\sigma^{n+1}| \leq \|\psi\|_{L^\infty(\Omega)}$,

$$T_5 \geq -T \|\psi\|_{L^\infty(\Omega)} \|\nabla \psi\|_{L^\infty(\Omega)} d \; m(\Omega) V.$$

We can write

$$T_6 \frac{1}{2} \sum_{n=0}^{N} \delta t \left([u^{n+1}, u^{n+1}]_{1,\mathcal{T}} + [u^{n+1} - u^0, u^{n+1} - u^0]_{1,\mathcal{T}} - [u^0, u^0]_{1,\mathcal{T}} \right),$$

which leads to

$$T_6 \geq \frac{1}{2} \sum_{n=0}^{N} \delta t [u^{n+1}, u^{n+1}]_{1,\mathcal{T}} - \frac{1}{2} T [u^0, u^0]_{1,\mathcal{T}}.$$

Gathering the above results, we get

$$\frac{1}{2} \sum_{n=0}^{N} \delta t [u^{n+1}, u^{n+1}]_{1,\mathcal{T}} \leq T \|\psi\|_{L^\infty(\Omega)} \|\nabla \psi\|_{L^\infty(\Omega)} d \; m(\Omega) V + \frac{1}{2} T [u^0, u^0]_{1,\mathcal{T}},$$

which gives (29), thanks to Lemma 4. □

The following lemma concerns an estimate of the partial derivative with respect to t of the discrete solution.

Lemma 6 ($L^2(\Omega \times (0,T))$ estimate on the time derivative)**.** *Under Assumptions 1, let $\mathcal{D}$ be an admissible finite volume discretization of $\Omega \times (0,T)$ in the sense of Definition 1, let $\theta \geq \mathrm{reg}(\mathcal{D})$ and let $(u^n)_{n \in \{0,\ldots,N+1\}}$ be the unique solution of the upwind implicit scheme (21)–(23). Then there exists C_3 only depending on d, Ω, T, ψ, r, $\mathbf{V}$ and θ such that*

$$\sum_{n=0}^{N} \delta t \sum_{K \in \mathcal{T}} \left(\left\| \frac{u^{n+1} - u^n}{\delta t} \right\|_{L^2(\Omega)} \right)^2 \leq C_3 \,. \tag{30}$$

Proof. We take as test function, in (23), the function $\frac{u^{n+1}-u^n}{\delta t}$, and sum over $n \in \{0,\dots,N\}$. We get $T_7+T_8+T_9+T_{10}=0$, with

$$T_7 = \sum_{n=0}^{N} \int_\Omega (\tilde{u}^{n+1}(x) - u^n(x)) \frac{u^{n+1}(x)-u^n(x)}{\delta t} \mathrm{d}x,$$

$$T_8 = r\delta t \sum_{n=0}^{N} \int_\Omega u^{n+1}(x) \frac{u^{n+1}(x)-u^n(x)}{\delta t} \mathrm{d}x,$$

$$T_9 = \delta t \sum_{n=0}^{N} \int_\Omega \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x) (\frac{u^{n+1}(x)-u^n(x)}{\delta t}) \mathrm{d}x,$$

and

$$T_{10} = \sum_{n=0}^{N} \delta t [u^{n+1}, \frac{u^{n+1}-u^n}{\delta t}]_{1,\mathcal{T}}.$$

Since, for *a.e.* $x \in \Omega$, we have either $\tilde{u}^{n+1}(x) < P_{\mathcal{T}}\psi(x) = u^{n+1}(x) \leq u^n(x)$, or $\tilde{u}^{n+1}(x) = u^{n+1}(x)$, we get

$$\begin{array}{l} \left(\tilde{u}^{n+1}(x) - u^n(x)\right)\left(u^{n+1}(x)-u^n(x)\right) \geq \left(u^{n+1}(x)-u^n(x)\right)^2, \\ \text{for a.e. } x \in \Omega,\ \forall n \in \{0,\dots,N\}. \end{array}$$

This leads to

$$T_7 \geq \sum_{n=0}^{N} \delta t \left(\left\| \frac{u^{n+1}-u^n}{\delta t} \right\|_{L^2(\Omega)} \right)^2.$$

We now write

$$T_8 = \frac{1}{2} r \int_\Omega \left(u^{N+1}(x)^2 - u^0(x)^2 + \sum_{n=0}^{N} (u^{n+1}(x)-u^n(x))^2 \right) \mathrm{d}x,$$

which gives

$$T_8 \geq -\frac{1}{2} r \left(\|P_{\mathcal{T}}\psi\|_{L^2(\Omega)} \right)^2.$$

We then turn to the study of T_9. Applying the Young inequality and (19), we get that

$$T_9 \geq -\frac{1}{2}\delta t \sum_{n=0}^{N} \left(\left\| \frac{u^{n+1}-u^n}{\delta t} \right\|_{L^2(\Omega)} \right)^2 - \delta t d\ \theta V^2 \sum_{n=0}^{N} \|u^{n+1}\|_{1,\mathcal{T}}^2.$$

Thanks to Lemma 5, we then get

$$T_9 \geq -\frac{1}{2}\delta t \sum_{n=0}^{N} \left(\left\| \frac{u^{n+1}-u^n}{\delta t} \right\|_{L^2(\Omega)} \right)^2 - d\ \theta V^2 C_2\,.$$

We now remark that

$$\begin{aligned} T_{10} &= \frac{1}{2} \sum_{n=0}^{N} \left([u^{n+1},u^{n+1}]_{1,\mathcal{T}} + [u^{n+1}-u^n, u^{n+1}-u^n]_{1,\mathcal{T}} - [u^n,u^n]_{1,\mathcal{T}} \right) \\ &= \frac{1}{2} \left([u^{N+1},u^{N+1}]_{1,\mathcal{T}} + \sum_{n=0}^{N} [u^{n+1}-u^n, u^{n+1}-u^n]_{1,\mathcal{T}} - [u^0,u^0]_{1,\mathcal{T}} \right) \\ &\geq -\frac{1}{2} [u^0,u^0]_{1,\mathcal{T}}. \end{aligned}$$

Gathering the above results yields (30). □

According to Lemma 3.3 in [15], we deduce the following Corollary, providing sufficient conditions for applying Kolmogorov's relative compactness criterium.

Corollary 1 (space and time translates estimates)**.** *Under Assumptions 1, let $\mathcal{D}$ be an admissible finite volume discretization of $\Omega \times (0,T)$ in the sense of Definition 1, let $\theta \geq \mathrm{reg}(\mathcal{D})$ and let $u_{\mathcal{D}}$ be given by Definition 6 from the upwind implicit scheme (21)–(23), prolonged by zero on $\mathbb{R}^{d+1} \setminus \Omega \times (0,T)$. Then, there exists C_4 and C_5 only depending on d, Ω, T, ψ, r, $\mathbf{V}$ and θ such that*

$$\|u_{\mathcal{D}}(.+\eta,.) - u_{\mathcal{D}}(.,.)\|^2_{L^2(\mathbb{R}^{d+1})} \leq C_4 \,|\eta|\,(|\eta| + 4\mathrm{size}(\mathcal{T}))\,,\ \forall \eta \in \mathbb{R}^d$$

and

$$\|u_{\mathcal{D}}(.,.+\lambda) - u_{\mathcal{D}}(.,.)\|^2_{L^2(\mathbb{R}^{d+1})} \leq \lambda C_5\,,\ \forall \lambda \in \mathbb{R}.$$

Proof. The space translate is a classical consequence of the discrete $L^2(0,T;H^1_0(\Omega))$ estimate (see Lem. 5). The time translate estimate is, on the one hand, a consequence of Lemma 6, on the other hand, a consequence of the L^∞ estimate (27). □

Thanks to the preceding estimates, we can state a relative compactness result.

Corollary 2. *Under Assumptions 1, let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be sequence of admissible finite volume discretizations of $\Omega \times (0,T)$ in the sense of Definition 1 such that there exists $\theta > 0$ with $\mathrm{reg}(\mathcal{D}) \leq \theta$ for all $m \in \mathbb{N}$ and $\mathrm{size}(\mathcal{D}_m) \to 0$ as $m \to +\infty$. For all $m \in \mathbb{N}$, Let $u_{\mathcal{D}_m}$ be given by Definition 6 from the upwind implicit scheme (21)–(23). Then there exists a subsequence of $(u_{\mathcal{D}_m})_{m\in\mathbb{N}}$, again denoted $(u_{\mathcal{D}_m})_{m\in\mathbb{N}}$ and there exists $\bar{u} \in H^1(\Omega\times(0,T)) \cap L^2\left(0,T;H^1_0(\Omega)\right)$ such that $\{u_{\mathcal{D}_m}\}_{m\in\mathbb{N}}$ converges to $\bar{u}$ in the strong topology of $L^2(\Omega\times(0,T))$ and $\{\frac{\partial u_{\mathcal{D}_m}}{\partial t}\}_{m\in\mathbb{N}}$ converges to $\frac{\partial \bar{u}}{\partial t}$ for the weak topology of $L^2(\Omega\times(0,T))$. Moreover, we have*

$$\int_{\Omega\times(0,T)} (\nabla \bar{u}(x,t))^2 \mathrm{d}x\mathrm{d}t \leq \liminf_{m\to\infty} \int_0^T \left(\|u_{\mathcal{D}_m}(\cdot,t)\|_{1,\mathcal{T}_m}\right)^2 \mathrm{d}t. \tag{31}$$

Proof. Thanks to Lemma 6, we get that the sequence $(\frac{\partial u_{\mathcal{D}_m}}{\partial t})_{m\in\mathbb{N}}$ is bounded in $L^2(\Omega\times(0,T))$. Therefore there exist $Z \in L^2(\Omega\times(0,T))$ and a subsequence of $(\mathcal{D}_m)_{m\in\mathbb{N}}$ again denoted $(\mathcal{D}_m)_{m\in\mathbb{N}}$, such that the sequence $(\frac{\partial u_{\mathcal{D}_m}}{\partial t})_{m\in\mathbb{N}}$ converges to Z for the weak topology of $L^2(\Omega\times(0,T))$. Thanks to the L^∞ estimate (27) and to Corollary 1, we deduce from the Kolmogorov theorem that there exist $\bar{u} \in L^2(\Omega\times(0,T))$ and a subsequence of $(\mathcal{D}_m)_{m\in\mathbb{N}}$ again denoted $(\mathcal{D}_m)_{m\in\mathbb{N}}$, such that the sequence $(u_{\mathcal{D}_m})_{m\in\mathbb{N}}$ converges to $\bar{u}$ in $L^2(\Omega\times(0,T))$ (this is the generalization to time dependent functions of Lem. 1). Moreover, since $\mathrm{size}(\mathcal{T}_m) \to 0$ as $m \to +\infty$, we get from Corollary 1 that

$$\|\bar{u}(.+\eta,.) - \bar{u}(.,.)\|^2_{L^2(\mathbb{R}^{d+1})} \leq C_2\,|\eta|^2,\ \forall \eta \in \mathbb{R}^d.$$

Applying Proposition IX.3 in [10], we get that $\bar{u} \in L^2\left(0,T;H^1(\Omega)\right)$. Furthermore, thanks to Theorem 1 in [16], $\bar{u}(t,.) \in H^1_0(\Omega)$ *a.e.* $t \in (0,T)$. Finally, since $\frac{\partial u_{\mathcal{D}_m}}{\partial t}$ weakly tends to Z and $u_{\mathcal{D}_m}$ tends to $\bar{u}$ in $L^2(\Omega\times(0,T))$ as $m \to +\infty$, we deduce that $\frac{\partial \bar{u}}{\partial t} = Z$ a.e. in $(0,T)\times\Omega$. Then (31) is obtained in a similar way as (14) in the case of a steady state problem. □

We can now state the convergence result for the upwind implicit scheme.

Theorem 1. *Under Assumptions 1, let $\theta > 0$ and let $\mathcal{D}$ be an admissible finite volume discretization of $\Omega\times(0,T)$ in the sense of Definition 1, such that $\mathrm{reg}(\mathcal{D}) \leq \theta$. Let $u_{\mathcal{D}}$ be given by Definition 6 from the upwind implicit scheme (21)–(23). Then $u_{\mathcal{D}}$ converges in $L^2(\Omega\times(0,T))$ to $\bar{u}$, the unique weak solution to (11), as $\mathrm{size}(\mathcal{D})$ tends to 0. Moreover, $\frac{\partial u_{\mathcal{D}}}{\partial t}$ weakly converges to $\frac{\partial \bar{u}}{\partial t}$ in $L^2(\Omega\times(0,T))$.*

Proof. Under Assumptions 1, let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of admissible finite volume discretizations of $\Omega\times(0,T)$ in the sense of Definition 1 such that $\mathrm{reg}(\mathcal{D}) \leq \theta$ for all $m \in \mathbb{N}$ and $\mathrm{size}(\mathcal{D}_m) \to 0$ as $m \to +\infty$. From Corollary 2, we get the existence of a subsequence of $(\mathcal{D}_m)_{m\in\mathbb{N}}$ and of $\bar{u} \in H^1(\Omega\times(0,T)) \cap L^2(0,T;H^1_0(\Omega))$ such that

$\{u_{\mathcal{D}_m}\}_{m\in\mathbb{N}}$ converges to $\bar{u}$ for the strong topology of $L^2(\Omega\times(0,T))$, $\{\frac{\partial u_{\mathcal{D}_m}}{\partial t}\}_{m\in\mathbb{N}}$ converges to $\frac{\partial \bar{u}}{\partial t}$ for the weak topology of $L^2(\Omega\times(0,T))$, and (31) holds.
We obtain the initial condition $\bar{u}(x,0)=\psi(x)$ by passing to the limit on the relation

$$\int_\Omega \varphi(x) P_{\mathcal{T}_m}\psi(x)\mathrm{d}x - \frac{1}{T}\int_0^T\int_\Omega \varphi(x)\left((T-t)\frac{\partial u_{\mathcal{D}_m}}{\partial t}(x,t) - u_{\mathcal{D}_m}(x,t)\right)\mathrm{d}x\mathrm{d}t,$$

for all $\varphi\in C_c^\infty(\Omega)$. We also obtain that $\bar{u}(x,t)\geq\psi(x)$ for *a.e.* $(x,t)\in\Omega\times(0,T)$ by passing *a.e.* to the limit on $u_{\mathcal{D}_m}(x,t)\geq P_{\mathcal{T}_m}\psi(x)$. It now suffices to prove that $\bar{u}$ is solution to the last relation of (11), since the uniqueness of the solution to (11) implies that all the sequence converges. This is then sufficient to conclude the proof of Theorem 1.

The proof that $\bar{u}$ is solution to (11) is performed in two steps. In the first step, we prove by passing to the limit on the scheme that

$$\int_0^T\int_\Omega \left[\begin{array}{l}\left(\dfrac{\partial\bar{u}}{\partial t}(x,t) + r\bar{u}(x,t) + \mathrm{div}(\bar{u}(x,t)\mathbf{V})\right)(\bar{u}(x,t)-\psi(x)) \\ +\nabla\bar{u}(x,t)\nabla(\bar{u}(x,t)-\psi(x))\end{array}\right]\mathrm{d}x\mathrm{d}t\leq 0. \tag{32}$$

Note that the above relation is equivalent to

$$\int_0^T\int_\Omega \left[\begin{array}{l}\left(\dfrac{\partial\bar{u}}{\partial t}(x,t) + r\bar{u}(x,t)\right)(\bar{u}(x,t)-\psi(x)) - \mathrm{div}(\bar{u}(x,t)\mathbf{V})\psi(x) \\ +\nabla\bar{u}(x,t)\nabla(\bar{u}(x,t)-\psi(x))\end{array}\right]\mathrm{d}x\mathrm{d}t\leq 0, \tag{33}$$

since, using $\bar{u}\in L^2(0,T;H_0^1(\Omega))$, we have

$$\begin{aligned}\int_0^T\int_\Omega \bar{u}(x,t)\mathrm{div}(\bar{u}(x,t)\mathbf{V})\mathrm{d}x\mathrm{d}t &= \frac{1}{2}\int_0^T\int_\Omega \mathrm{div}(\bar{u}^2(x,t)\mathbf{V})\mathrm{d}x\mathrm{d}t \\ &= \frac{1}{2}\int_0^T\int_{\partial\Omega}\bar{u}^2(x,t)\mathbf{V}.\mathbf{n}(x)\mathrm{d}\gamma(x)\mathrm{d}t = 0.\end{aligned}$$

Let $m\in\mathbb{N}$ be given; we then omit the index m in some discretized expressions, denoting $\mathcal{D}=\mathcal{D}_m$. We again introduce, in (23), the test function $u^{n+1}-P_{\mathcal{T}}\psi = u^{n+1}-u^0$ and we sum over $n\in\{0,\ldots,N\}$. We again get $T_3^{(m)}+T_4^{(m)}+T_5^{(m)}+T_6^{(m)}=0$, with

$$T_3^{(m)} = \sum_{n=0}^N\int_\Omega(\tilde{u}^{n+1}(x)-u^n(x))(u^{n+1}(x)-u^0(x))\mathrm{d}x,$$

$$T_4^{(m)} = r\delta t\sum_{n=0}^N\int_\Omega u^{n+1}(x)(u^{n+1}(x)-u^0(x))\mathrm{d}x,$$

$$T_5^{(m)} = \delta t\sum_{n=0}^N\int_\Omega \nabla_{\mathcal{T}}^{\mathbf{V}}u^{n+1}(x)(u^{n+1}(x)-u^0(x))\mathrm{d}x,$$

and

$$T_6^{(m)} = \sum_{n=0}^N\delta t[u^{n+1},u^{n+1}-u^0]_{1,\mathcal{T}}.$$

We again remark that we have

$$T_3^{(m)} = \sum_{n=0}^N\int_\Omega(u^{n+1}(x)-u^n(x))(u^{n+1}(x)-u^0(x))\mathrm{d}x.$$

Let us denote by $\hat{u}_{\mathcal{D}}$, the function defined by $\hat{u}_{\mathcal{D}}(x,t) = u^{n+1}(x)$ for *a.e.* $x \in \Omega$ and $t \in [n\delta t, (n+1)\delta t)$. We get that $\hat{u}_{\mathcal{D}_m}$ also converges to $\bar{u}$ in $L^2(\Omega \times (0,T))$, since Lemma 6 proves that $\hat{u}_{\mathcal{D}_m} - u_{\mathcal{D}_m}$ tends to 0 in $L^2(\Omega \times (0,T))$. We then get that

$$T_3^{(m)} = \int_0^T \int_\Omega \frac{\partial u_{\mathcal{D}_m}}{\partial t}(x,t)(\hat{u}_{\mathcal{D}_m}(x,t) - P_{\mathcal{T}_m}\psi(x))\mathrm{d}x\mathrm{d}t,$$

and therefore, passing to the limit on a product of weakly-strongly convergent sequences, we get

$$\lim_{m\to\infty} T_3^{(m)} = \int_0^T \int_\Omega \frac{\partial \bar{u}}{\partial t}(x,t)(\bar{u}(x,t) - \psi(x))\mathrm{d}x\mathrm{d}t.$$

We also get that

$$T_4^{(m)} = r\int_0^T \int_\Omega \hat{u}_{\mathcal{D}_m}(x,t)(\hat{u}_{\mathcal{D}_m}(x,t) - u^0(x))\mathrm{d}x\mathrm{d}t,$$

and therefore

$$\lim_{m\to\infty} T_4^{(m)} = r\int_0^T \int_\Omega \bar{u}(x,t)(\bar{u}(x,t) - \psi(x))\mathrm{d}x\mathrm{d}t.$$

Thanks to Lemma 2, we have $T_5^{(m)} \geq T_{11}^{(m)}$ with

$$T_{11}^{(m)} = -\delta t \sum_{n=0}^{N} \int_\Omega \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x) P_{\mathcal{T}_m}\psi(x)\mathrm{d}x.$$

With a straightforward adaptation of Lemma 2 to time-dependent functions, we get

$$\lim_{m\to\infty} T_{11}^{(m)} = -\int_0^T \int_\Omega \psi(x)\mathbf{V}\cdot\nabla\bar{u}(x,t)\mathrm{d}x\mathrm{d}t.$$

Thanks to (31) and to the time-dependent version of Lemma 1, we get

$$\liminf_{m\to\infty} T_6^{(m)} \geq \int_0^T \int_\Omega \nabla\bar{u}(x,t)\cdot(\nabla\bar{u}(x,t) - \nabla\psi(x))\mathrm{d}x\mathrm{d}t.$$

Gathering the above results yields (33) and therefore (32).

We now turn to the second step, which consists in proving that, for all $v \in C_c^\infty(\Omega\times(0,T))$ such that, for all (x,t) in $\Omega\times(0,T)$, $v(x,t) \geq \psi(x)$, we have

$$\int_0^T \int_\Omega \left[\begin{array}{l} \left(\dfrac{\partial \bar{u}}{\partial t}(x,t) + r\bar{u}(x,t) + \mathrm{div}(\bar{u}(x,t)\mathbf{V})\right)(v(x,t) - \psi(x)) \\ +\nabla\bar{u}(x,t)\nabla(v(x,t) - \psi(x)) \end{array} \right] \mathrm{d}x\mathrm{d}t \geq 0. \tag{34}$$

Then, the subtraction of (32) to (34) gives (11) with test functions in $C_c^\infty(\Omega\times(0,T))$ instead of $L^2(0,T;H_0^1(\Omega))$. One then concludes by density of $C_c^\infty(\Omega\times(0,T))$ in $L^2(0,T;H_0^1(\Omega))$. We therefore introduce, for all $n \in \{0,\ldots,N\}$, the test function $w^{n+1} = P_{\mathcal{T}_m} v(\cdot,(n+1)\delta t) - P_{\mathcal{T}_m}\psi$ in (23) and we sum over $n \in \{0,\ldots,N\}$.

We thus get $T_{12}^{(m)} + T_{13}^{(m)} + T_{14}^{(m)} + T_{15}^{(m)} = 0$, with

$$T_{12}^{(m)} = \sum_{n=0}^{N} \int_{\Omega} (\tilde{u}^{n+1}(x) - u^n(x)) w^{n+1}(x) \mathrm{d}x,$$

$$T_{13}^{(m)} = r\delta t \sum_{n=0}^{N} \int_{\Omega} u^{n+1}(x) w^{n+1}(x) \mathrm{d}x,$$

$$T_{14}^{(m)} = \delta t \sum_{n=0}^{N} \int_{\Omega} \nabla_{\mathcal{T}}^{\mathbf{V}} u^{n+1}(x) w^{n+1}(x) \mathrm{d}x,$$

and

$$T_{15}^{(m)} = \sum_{n=0}^{N} \delta t [u^{n+1}, w^{n+1}]_{1,\mathcal{T}}.$$

Since $w^{n+1}(x) \geq 0$ and $\tilde{u}^{n+1}(x) \leq u^{n+1}(x)$, for *a.e.* $x \in \Omega$ and for all $n \in \{0, \ldots, N\}$, we get $T_{12}^{(m)} \leq T_{16}^{(m)}$ with

$$T_{16}^{(m)} = \sum_{n=0}^{N} \int_{\Omega} (u^{n+1}(x) - u^n(x)) w^{n+1}(x) \mathrm{d}x \sum_{n=0}^{N} \int_{n\delta t}^{(n+1)\delta t} \int_{\Omega} \frac{\partial u_{\mathcal{D}_m}}{\partial t}(x,t) w^{n+1}(x) \mathrm{d}x \mathrm{d}t.$$

We then have, as previously,

$$\lim_{m\to\infty} T_{16}^{(m)} = \int_0^T \int_{\Omega} \frac{\partial \bar{u}}{\partial t}(x,t)(v(x,t) - \psi(x)) \mathrm{d}x \mathrm{d}t,$$

$$\lim_{m\to\infty} T_{13}^{(m)} = r \int_0^T \int_{\Omega} \bar{u}(x,t)(v(x,t) - \psi(x)) \mathrm{d}x \mathrm{d}t,$$

$$\lim_{m\to\infty} T_{14}^{(m)} = \int_0^T \int_{\Omega} (v(x,t) - \psi(x)) \mathbf{V} \cdot \nabla \bar{u}(x,t) \mathrm{d}x \mathrm{d}t,$$

and

$$\liminf_{m\to\infty} T_{15}^{(m)} \int_0^T \int_{\Omega} \nabla \bar{u}(x,t) \cdot (\nabla v(x,t) - \nabla \psi(x)) \mathrm{d}x \mathrm{d}t.$$

Gathering the above results produces (34) since $T_{16}^{(m)} + T_{13}^{(m)} + T_{14}^{(m)} + T_{15}^{(m)} \geq 0$. This concludes the proof of Theorem 1. □

5. Applications and open problems

We consider Problem (2)–(5) in the case of the American put option on the minimum of two underlying assets, the payoff of which is given, for all $x = (x_1, x_2) \in \mathbb{R}^2$ by $\psi(x) = \max(\kappa - \min(\exp(x_1), \exp(x_2)), 0)$ where $\kappa > 0$ is the exercise price. We assume that $(\sigma_{ij})_{i,j=1,2}$ is defined by $\sigma_{11} = \tilde{\sigma}_{11}$, $\sigma_{12} = 0$, $\sigma_{21} = \rho\tilde{\sigma}_{22}$, and $\sigma_{22} = (1-\rho^2)^{1/2}\tilde{\sigma}_{22}$ for given reals $\tilde{\sigma}_{11} > 0$, $\tilde{\sigma}_{22} > 0$ and $\rho \in (-1, 1)$. The following table provides eighteen test cases, crossing three values for the correlation ρ and three values for the exercise price κ on two different sets of data.

$\tilde{\sigma}_{11}$	$\tilde{\sigma}_{22}$	ρ	$(\bar{x}_i)_{i=1,2}$	r	$(\lambda_i)_{i=1,2}$	κ	T
0.2	0.3	–0.5, 0.0, 0.5	log(40)	log(1.05)	log(1.02)	36, 40, 44	1
0.3	0.4	–0.5, 0.0, 0.5	log(100)	log(1.05)	log(1.02)	90, 100, 110	1

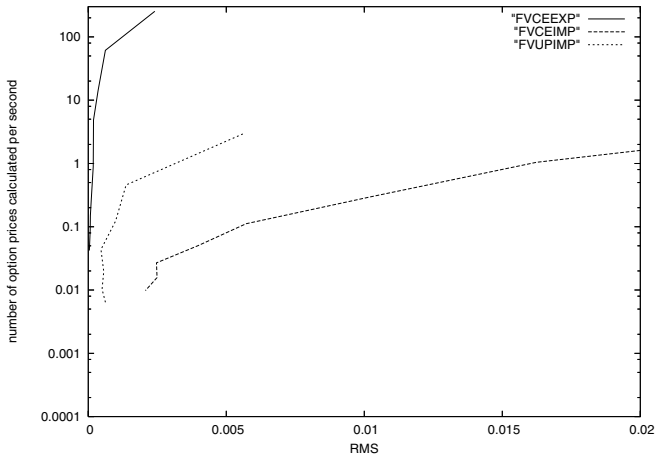

FIGURE 2. Comparison between finite volume numerical schemes.

We have computed these eighteen test cases, using the following numerical schemes (all the computations have been performed on a PC Pentium IV 2.66 GH computer with a RAM capacity equal to 512 Mb):

(1) the explicit DPEXP scheme [3];
(2) the DPADI scheme [25] using the same number of steps for the space and time discretizations;
(3) the KR scheme [20];
(4) the binomial BEG scheme [8];
(5) the centered explicit finite volume scheme FVCEEXP (defined in this paper by (21), (22), (26)), using a time step equal to 0.31 times the stability condition and rectangular control volumes;
(6) the centered implicit finite volume scheme FVCEIMP (defined in this paper by (21), (22), (25)), using a time step equal to the stability condition of the explicit scheme, a preconditioning ILU(0) GMRES method to solve linear systems (see [23]) and rectangular control volume;
(7) the upwind implicit finite volume scheme FVUPIMP (defined in this paper by (21), (22), (23)), using a time step equal to twice the stability condition of the explicit scheme, a preconditioning ILU(0) GMRES method to solve linear systems and rectangular control volumes.

We take as reference values for $U(\bar{x},0) = u(\bar{x},T)$ in the eighteen test cases, denoted by $(\hat{u}_i)_{i=1,\dots,18}$, which are computed using the multinomial tree scheme with 3000 steps [8]. From the computation of the eighteen approximations $(u_i^{(\mathrm{NS})})_{i=1,\dots,18}$ of $u(\bar{x},T)$, given by the numerical schemes NS = DPEXP, DPADI, KR, BEG, FVCEEXP, FVCEIMP and FVUPIMP, we define the root-mean-squared (RMS) relative error by (see Broadie and Detemple [11])

$$\mathrm{RMS}^{(\mathrm{NS})} = \left(\frac{1}{18} \sum_{i=1}^{18} \left(\frac{u_i^{(\mathrm{NS})} - \hat{u}_i}{\hat{u}_i} \right)^2 \right)^{1/2}.$$

In Figure 2, we compare the number of option prices calculated per second CPU with respect to $\mathrm{RMS}^{(\mathrm{FVCEEXP})}$, $\mathrm{RMS}^{(\mathrm{FVCEIMP})}$ and $\mathrm{RMS}^{(\mathrm{FVUPIMP})}$. We obtain that, for a given computing time, the centered explicit finite volume scheme gives a better precision on these eighteen cases than the implicit finite volume schemes, and that, classically, the centered scheme is more precise than the upstream weighting scheme. In Figure 3, we draw the $\mathrm{RMS}^{(\mathrm{DPADI})}$, $\mathrm{RMS}^{(\mathrm{DPEXP})}$ and $\mathrm{RMS}^{(\mathrm{FVCEEXP})}$ errors with respect to the number of space steps. This figure shows that, for a given number of space steps, the centered explicit finite volume methods gives a better approximation of the eighteen prices than the DPADI and DPEXP numerical schemes. Figure 4 gives the number of option prices calculated per second CPU with respect to the RMS error. We see that the computing

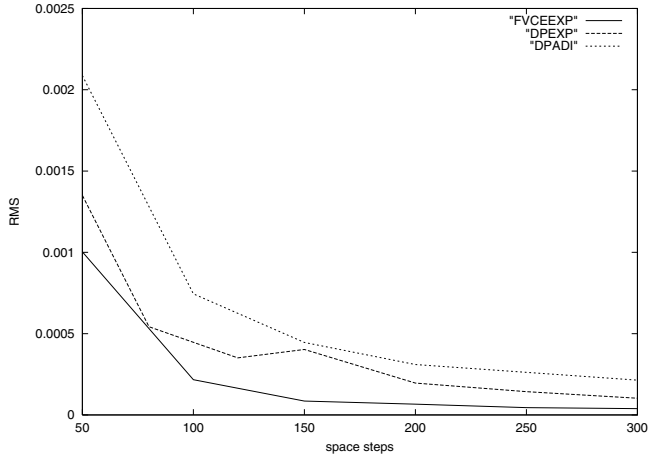

FIGURE 3. Precision with respect to the size of the mesh.

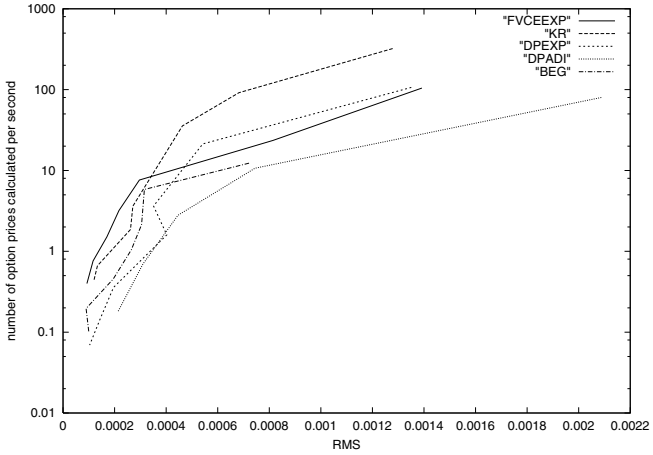

FIGURE 4. Number of option prices calculated per second CPU with respect to the relative error.

time needed for the explicit centered finite volume scheme (FVCEEXP) is similar to that of the trinomial scheme (KR), whereas the efficiency of these two schemes appears to be similar to that of the binomial scheme (BEG) and of the dynamic programming explicit scheme (DPEXP) and better than that of the dynamic programming ADI scheme (DPADI).

Hence this paper shows that the use of the finite volume scheme for solving some probabilistic problems issued from financial mathematics leads, in some cases, to accurate results for an acceptable computing time. The application of this scheme to Voronoï grids and the use of Sparse Grids adaptive meshes (see [18]) will be the main points of further works.

REFERENCES

[1] K. Amin and A. Khanna, Convergence of American option values from discrete- to continuous-time financial models. *Math. Finance* **4** (1994) 289–304.

[2] V. Bally and G. Pages, A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli* **9** (2003) 1003–1049.

[3] G. Barles, Ch. Daher and M. Romano, Convergence of numerical Schemes for problems arising in Finance theory. *Math. Mod. Meth. Appl. Sci.* **5** (1995) 125–143.
[4] J. Bénard, R. Eymard, X. Nicolas and C. Chavant, Boiling in porous media: model and simulations. *Transport Porous Med.* **60** (2005) 1–31.
[5] A. Bensoussan and J.L. Lions, *Applications des inéquations variationnelles en contrôle stochastique*, Dunod, Paris (1978). *Application of variational inequalities in stochastic control*, North Holland (1982).
[6] J. Berton and R. Eymard, *Une méthode de volumes finis pour le calcul des options américaines*, Congrès d'Analyse Numérique. La Grande Motte, France (2003). `http://www.math.univ-montp2.fr/canum03/`
[7] J. Berton, *Méthodes de volumes finis pour des problèmes de mathématiques financières*. Thèse de l'Université de Marne-la-Vallée, France (in preparation).
[8] P. Boyle, J. Evnine and S. Gibbs, Numerical evaluation of multivariate contingent claims. *Rev. Financ. Stud.* **2** (1989) 241–250.
[9] M.J. Brennan and E. Schwartz, The valuation of the American put option. *J. Financ.* **32** (1977) 449–462.
[10] H. Brézis, *Analyse fonctionnelle (Théorie et applications)*. Dunod, Paris (1999).
[11] M. Broadie and J. Detemple, American option valuation: new bounds, approximations, and a comparison of existing methods securities using simulation. *Rev. Financ. Stud.* **9** (1996) 1221–1250.
[12] P. Carr, R. Jarrow and R. Myneni, Alternative characterizations of American put options. *Math. Financ.* **2** (1992) 87–106.
[13] J.C. Cox, S.A. Ross and M. Rubinstein, Options pricing: A simplified approach. *J. Financ. Econ.* **7** (1979) 229–263.
[14] J.N. Dewynne, S.D. Howison, I. Rupf and P. Wilmott, Some mathematical results in the pricing of American options, *Eur. J. Appl. Math.* **4** (1993) 381–398.
[15] R. Eymard, T. Gallouët and R. Herbin, Finite Volume Methods, in *Handb. Numer. Anal.*, Ph. Ciarlet and J.L. Lions (Eds.) **7** (2000) 715–1022.
[16] R. Eymard, T. Gallouët and R. Herbin, Convergence of finite volume schemes for semilinear convection diffusion equations, *Numer. Math.* **82** (1999) 90–116.
[17] R. Eymard, T. Gallouët, R. Herbin and A. Michel, Convergence of a finite volume scheme for nonlinear degenerate parabolic equations, *Numer. Math.* **92** (2001) 41–82.
[18] P.W. Hemker, Sparse-grid finite-volume multigrid for 3D-problems. *Adv. Comput. Math* **4** (1995) 83–110.
[19] P. Jaillet, D. Lamberton and B. Lapeyre, Variational inequalities and the pricing of American options. *Acta Appl. Math.* **21** 3 (1990) 263–289.
[20] B. Kamrad and P. Ritchken, Multinomial approximating models for options with k-state variables. *Manage. Sci.* **37** (1991) 1640–1652.
[21] O.A. Ladyzhenskaya, V.A. Solonnikov and N.N. Ural'tseva, Linear and quasi-linear equations of parabolic type. Translated from the Russian by S. Smith. Transl. Math. Monogr. (AMS) **23** (1968) xi+648.
[22] D. Lamberton and B. Lapeyre, *Introduction au calcul stochastique appliqué à la finance*. Ellipses, Paris, New York, London (1997) 176.
[23] Y. Saad, *Iterative methods for sparse linear systems*. First edition, SIAM (1996).
[24] I. Sapariuc, M.D. Marcozzi and J.E. Flaherty, A numerical analysis of variational valuation techniques for derivative securities, *Appl. Math. Comput.* **159** (2004) 171–198.
[25] S. Villeneuve and A. Zanette, Parabolic A.D.I. methods for pricing American options on two stocks, *Math. Oper. Res.* **27** (2002) 121–149.
[26] R. Zvan, P.A. Forsyth and K.R. Vetzal, A finite volume approach for contingent claims valuation, *IMA J. Numer. Anal.* **21** (2001) 703–731.