Numerical approximation for a superreplication problem under gamma constraints

Benjamin Bruder[*], Olivier Bokanowski[†], Stefania Maroso[‡], Hasnaa Zidani[§]

May 8, 2008

Abstract

We study a superreplication problem of European options with gamma constraints, in mathematical finance. The initially unbounded control problem is set back to a problem involving a viscosity PDE solution with a set of bounded controls. Then a numerical approach is introduced, inconditionnally stable with respect to the mesh steps. A generalized finite difference scheme is used since basic finite differences cannot work in our case. Numerical tests illustrate the validity of our approach.

Keywords. Super-replication problem, viscosity solution, numerical approximation, generalized finite difference scheme, monotone scheme, Howard's algorithm

1 Introduction

In a financial market, consisting in a non-risky asset and some risky assets, people are interested to study the minimal initial capital needed in order to superreplicate a given contingent claim, under gamma constraints. Many authors have studied this problem in different cases and with different constraints. For instance, see [13, 20], for problems in dimension 1, [9] for problems in dimension 2, and [21, 11] for problems in a general dimension d. In all these papers, authors characterize the superreplication price as the viscosity solution of an Hamilton-Jaboci-Bellman (HJB) equation with terminal and boundary conditions. In a particular case, the dual formulation of the superreplication problem leads to a standard form of optimal stochastic control problem of [9].

In this paper we study numerically an HJB equation coming from the superreplication problem in dimension 2 introduced in [9]. We discretize the HJB equation using the Generalized Finite Differences scheme [7, 8], then we study existence and uniqueness of the discrete solution. Finally we prove the convergence of the numerical solution to the viscosity solution.

[*]Lab. de Probabilités et Modèles Aléatoires, CNRS UMR 7599, Université Paris 7, also at Société Générale Asset Management. E-mail: BRUDER@MATH.JUSSIEU.FR

[†]Lab. Jacques Louis Lions, Univrsités Paris 6 & 7. E-mail: BOKA@MATH.JUSSIEU.FR

[‡]Projet TOSCA, INRIA Sophia-Antipolis. E-mailSTEFANIA.MAROSO@INRIA.FR

[§]ENSTA, 32 Bd Victor 75015 Paris. Also at INRIA Saclay Ile de France (Projet COMMANDS). E-mail: HASNAA.ZIDANI@ENSTA.FR

More precisely, we are interested by the HJB equation which comes from the two dimensional dual problem introduced in [9]:

$$\vartheta(t,x,y) = \sup_{(\rho,\xi)\in\mathcal{U}} \mathbb{E}\left[g\left(X^{\rho,\xi}_{t,x,y}(T)\right)\right], \tag{1.1}$$

where (ρ,ξ) are valued in $[-1,1]\times(0,\infty)$, the process $(X^{\rho,\xi}_{t,x,y}, Y^{\rho,\xi}_{t,y})$ is a 2-dimensional positive process which evolves according to the stochastic dynamics (2.1), and g is a payoff function. The main difficulty of the above problem is due to the non-boundedness of the control set. This implies that the Hamiltonian associated to (1.1) is not bounded, and then, numerical approximation and theoretical analysis for such a problem become more complicate.

In the literature, problems with unbounded control have been studied by many authors (for instance [1, 10]). In these papers, the authors decide to truncate the set of controls to make it bounded. This truncation simplifies the numerical analysis of the problem. However, there is no theoretical result justifying this truncation.

In this paper we do not truncate the set of controls, because we find a particular form of our HJB equation which leads us to avoid the difficulty of unbounded control. In fact, our HJB equation can be reformulated in the following way

$$\Lambda^-(J(t,x,y,D\vartheta(t,x,y),D^2\vartheta(t,x,y))) = 0,$$

where J is a symetric matrix differential operator associated to the Hamiltonian, and where $\Lambda^-(J)$ means the smallest eigenvalue of the matrix operator J. We rewrite the smallest eigenvalue as follows:

$$\Lambda^-(J) = \min_{\|\alpha\|=1} \alpha^T J\alpha,$$

where $\alpha\in\mathbb{R}^2$, and $\|\cdot\|$ denotes the Euclidean norm.

The structure of the paper is the following: in Section 2 we present the problem and the associated HJB equation. We prove boundary conditions satisfied by the value function, then the existence, uniqueness and Lipschitz property of the viscosity solution. In Section 3 we consider the discretization of the HJB equation, and recall the main properties of the Generalized Finite Differences scheme and we prove the consistency of this scheme. In Section 4, we prove existence and uniqueness of a bounded discrete numerical approximation. In Section 5 we prove the convergence of the numerical approximation. Finaly Section 6 is devoted to numerical tests and validation of the proposed algorithm.

2 Problem formulation and PDE

Let $(\Omega,\mathcal{F}_t,\mathbb{P})$ be a probability space, and $T>0$ be a fixed finite time horizon. Let $\mathcal{U}$ denotes the set of all $\mathcal{F}_t$-measurable processes $(\rho,\zeta) := \{(\rho(t),\zeta(t)); 0\le t\le T\}$ with values in $[-1,1]\times\mathbb{R}_+$:

$$\mathcal{U} := \Big\{(\rho,\zeta) \text{ valued in } [-1,1]\times(0,+\infty) \text{ and } \mathcal{F}_t\text{-measurable} \mid \int_0^T \zeta_t^2\,dt < +\infty\Big\}.$$

For a given control process (ρ,ζ), and an initial data $(t,x,y)\in(0,T)\times\mathbb{R}^+\times\mathbb{R}^+$, we consider the controlled 2-dimensional positive process $(X,Y) = (X^{\rho,\zeta}_{t,x,y}, Y^{\rho,\zeta}_{t,y})$ evolving according

to the stochastic dynamics:

$$dX(s) = \sigma(s, Y(s))X dW_s^1, s \in (t, T) \tag{2.1a}$$
$$dY^{\rho,\zeta}(s) = -\mu(s, Y(s))ds + \zeta(s)Y(s)dW_s^2, \quad s \in (t, T) \tag{2.1b}$$
$$\langle dW_s^1, dW_s^2 \rangle = \rho(s), \qquad \text{a.e } s \in (t, T) \tag{2.1c}$$
$$X(t) = x \ , \ Y(t) = y, \tag{2.1d}$$

where W_s^1 and W_s^2 denote the standard Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The volatility σ and the cash flow μ satisfy the following assumptions:

(A1) $\sigma : [0, T] \times \mathbb{R} \to \mathbb{R}^+$ is a positive function, such that σ^2 is Lipschitz. For every $t \in [0, T]$, $\sigma(t, 0) = 0$ (typically $\sigma(t, y) = \sqrt{y}$).

(A2) $\mu : (0, T) \times \mathbb{R}^+ \to \mathbb{R}^+$ is a positive Lipschitz function, with $\mu(t, 0) = 0$ for every $t \in [0, T]$.

Assumptions (A1) and (A2) ensure that the stochastic dynamic system (2.1) has a unique strong solution.

The variables $X_{t,x,y}^{\rho,\zeta}$ and $Y_{t,y}^{\rho,\zeta}$ describe two different assets from a financial market. The first asset $X_{t,x,y}^{\rho,\zeta}$ is risky, while the second one $Y_{t,y}^{\rho,\zeta}$ distributes an instaneous cash flow $\mu(s, Y_{t,y}^{\rho,\zeta}(s))$, and its price is linked to the asset $X_{t,x,y}^{\rho,\zeta}$ by the means of volatility $\sigma(s, Y_{t,y}^{\rho,\zeta}(s))$.

Remark 2.1. *It is important to remark that the evolution of the variable $Y_{t,y}^{\rho,\zeta}$ does not depend on $X_{t,x,y}^{\rho,\zeta}$.*

Now consider a function $g : \mathbb{R}^+ \to \mathbb{R}$. Different assumptions will be made on g:

(A3) g is a bounded Lipschitz function (hereafter we denote $C_0 := \|g\|_\infty$).

We shall also possibly assume

(A4) The function $f : z \to g\left(e^z\right)$ is Lipschitz continuous.

or

(A5) $g \in \mathcal{C}^2(\mathbb{R}^+)$, and the function $x \to -x^2 g''(x)$ is bounded from below on $\mathbb{R}_+$.

Consider the following stochastic control problem $(\mathcal{P}_{t,x,y})$ with its associated value function ϑ defined by:

$$\vartheta(t, x, y) := \sup_{(\rho,\zeta)\in\mathcal{U}} \mathbb{E} \to \left[g\left(X_{t,x,y}^{\rho,\zeta}(T) \right) \right]. \tag{2.2}$$

Assumption (A3) leads us to obtain a bounded value function ϑ of (2.2). Assumption (A4) is usefull to prove more regularity on the function ϑ and some boundary conditions (see section 2.2). Assumption (A5) will be needed for the scheme convergence proof.

This control problem can be interpreted in the following sense (see [9]): A trader wants to sell an European option of terminal payoff $g(X_T)$ without taking any risk. Hence we use

a superreplication framework. The underlying X of the option is a risky asset, for axample a stock, an index or a mutual fund. Unfortunately, in several cases, the volatility σ of the underlying X exhibits large random changes across time. Therefore, the Black-Scholes model fails to capture the risks of the trader. One must then use a model that features stochastic volatility. It is known that in this framework, the superreplication problem has a trivial solution (see [13]). For example, if the volatility has no a priori bound, the superreplication price is the concave envelope of the payoff $g(X(T))$, and the hedging strategy is static. To obtain more accurate prices, we introduce another financial asset Y whose price is linked to the volatility of the underlying X. For example, we can consider a variance swap which continuously pays the instantaneous variance of X (hence $\mu(t,Y) = \sigma^2$). For the sake of simplicity we assume that the price of Y and the volatility of X are driven by a single common factor (hence $\sigma = \sigma(t,Y)$). If the parameters ζ and ρ of the dynamics of the price Y were known, and if there were no transaction costs for Y, the superreplication price would simply be $\mathbb{E}\left[g\left(X^{\rho,\zeta}_{t,x,y}(T)\right)\right]$. But we face two problems:

- The parameters (ζ,ρ) of the dynamics of Y are likely to be random and difficult to measure. As there is no a priori bound to these parameters, the superreplication price is given be the supremum of $\mathbb{E}\left[g\left(X^{\rho,\zeta}_{t,x,y}(T)\right)\right]$ over all adapted processes ζ,ρ (see [15]).
- The asset Y is likely to introduce transaction costs, and hence the trader cannot buy and sell an infinite amount of asset Y during the period $[0,T]$. It is proved in [9] that the superreplication price of $g(X(T))$ under the constraint of a finite amount of transactions involving Y during $[0,T]$ is given by the value function of problem (2.2). See also [20, 21] for a similar approach.

2.1 HJB equation

Denote by $\mathcal{M}_2$ the set of symmetric 2×2 matrices. The Hamiltonian function is defined by: for $t\in(0,T)$, $x,y\in\mathbb{R}^+$, $p=(p_1,p_2)^\mathsf{T}\in\mathbb{R}^2$, and $Q\in\mathcal{M}_2$:

$$H(t,x,y,p,Q) := \inf_{(\rho,\zeta)\in[-1,1]\times\mathbb{R}_+}\left\{\mu(t,y)p_2 - \frac{1}{2}\mathrm{tr}\left(a(t,x,y,\zeta,\rho)\cdot Q\right)\right\}, \tag{2.3}$$

and the covariance matrix a is given by:

$$a(t,x,y,\zeta,\rho) := \begin{pmatrix}\sigma^2(t,y)x^2 & \rho\zeta\sigma(t,y)x\\ \rho\zeta\sigma(t,y)x & \zeta^2\end{pmatrix}.$$

Now we look for a characterization of ϑ as a viscosity solution of an HJB equation. If the minimum in (2.3) is finite, then ϑ satisfies (in the viscosity sense) the following PDE:

$$-\frac{\partial\vartheta}{\partial t} + H\left(t,x,y,D\vartheta,D^2\vartheta\right) = 0 \quad (t,x,y)\in(0,T)\times(0,+\infty)\times(0,+\infty). \tag{2.4}$$

However, the infimum in (2.3) could be infinite, and we will prove in Theorem 2.5 that the precise HJB equation satisfied by ϑ in the viscosity sense is

$$\Lambda^-\begin{pmatrix}-\frac{\partial\vartheta}{\partial t}+\mu(t,y)\frac{\partial\vartheta}{\partial y}-\frac{1}{2}\sigma^2(t,y)x^2\frac{\partial^2\vartheta}{\partial x^2} & -\frac{1}{2}\sigma(t,y)x\frac{\partial^2\vartheta}{\partial x\partial y}\\ -\frac{1}{2}\sigma(t,y)x\frac{\partial^2\vartheta}{\partial x\partial y} & -\frac{1}{2}\frac{\partial^2\vartheta}{\partial y^2}\end{pmatrix} = 0, \tag{2.5}$$

where $\Lambda^-(A)$ denotes the smallest eigenvalue of a given symmetric matrix A. We first prove that ϑ is a discontinuous viscosity solution of (2.5). We will see later on that, under (A1), ϑ is continuous thanks to a comparison principle, and even Lipschitz continuous when assumptions (A3)-(A5) hold.

First, it is easy to see that the infimum in (2.3) can only be achieved for $\rho = \pm 1$. Hence denoting ζ as $\rho\zeta$, one can see that the Hamiltonian can be rewritten as:

$$H(t,x,y,p,Q) = \inf_{\zeta\in\mathbb{R}} \left\{ \mu(t,y)p_2 - \frac{1}{2}\mathrm{tr}\left(a(t,x,y,\zeta)\cdot Q\right) \right\}, \tag{2.6}$$

where, this time, there is only one control variable ζ taking values on the whole real line, and the covariance matrix a is defined by:

$$a(t,x,y,\zeta) = \begin{pmatrix} \sigma^2(t,y)x^2 & \zeta\sigma(t,y)x \\ \zeta\sigma(t,y)x & \zeta^2 \end{pmatrix}.$$

By elementary computations, the minimization over ζ, in (2.6) gives:

$$H(t,x,y,p,Q) = -\infty \qquad \text{if } Q_{22} > 0, \tag{2.7a}$$

$$\text{or} \quad Q_{22} = 0 \text{ and } \sigma(t,y)xQ_{12} \neq 0, \tag{2.7b}$$

$$H(t,x,y,p,Q) \in \mathbb{R}, \qquad \text{otherwise.} \tag{2.7c}$$

Remark 2.2. *For this particular problem, it is not possible to find a continuous function $G : [0,T]\times\mathbb{R}^2\times\mathbb{R}^2_+\times\mathcal{M}_2 \to \mathbb{R}$ such that*

$$H(t,x,y,p,Q) > -\infty \Leftrightarrow G(t,x,y,p,Q) \geq 0.$$

Hence we can not use arguments introduced in [19] to deal with the HJB equation (2.5).

For $t\in(0,T)$, $x,y\in\mathbb{R}^+$, $r\in\mathbb{R}$, $p=(p_1,p_2)^T\in\mathbb{R}^2$ and $Q\in\mathcal{M}_2$, introduce the notation:

$$J(t,x,y,r,p,Q) = \begin{pmatrix} -r+\mu(t,y)p_2 - \frac{1}{2}\sigma^2(t,y)x^2Q_{11} & -\frac{1}{2}\sigma(t,y)xQ_{12} \\ -\frac{1}{2}\sigma(t,y)xQ_{12} & -\frac{1}{2}Q_{22} \end{pmatrix}.$$

With straightforward computations we obtain the following result.

Lemma 2.3. *For $t\in(0,T)$, $x,y\in\mathbb{R}^+$, $r\in\mathbb{R}$, $p=(p_1,p_2)^T\in\mathbb{R}^2$ and $Q\in\mathcal{M}_2$, the following assertions hold:*

(i) $-r+H(t,x,y,p,Q)\geq 0 \Leftrightarrow \Lambda^-(J(t,x,y,r,p,Q))\geq 0.$

(ii) $-r+H(t,x,y,p,Q)\geq 0 \Rightarrow -Q_{22}\geq 0.$

(iii) $-r+H(t,x,y,p,Q)= 0 \Rightarrow \Lambda^-(J(t,x,y,r,p,Q))= 0.$

(iv) $\Lambda^-(J(t,x,y,r,p,Q))> 0 \Rightarrow -r+H(t,x,y,p,Q)> 0.$

Now, for a function $u : [0,T] \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$, we define the upper (resp. lower) semicontinuous envelope u^* (resp. u_*) of u by : for $t \in [0,T), x, y \in (0,+\infty)$,

$$u^*(t,x,y) = \limsup_{\substack{(s,w,z)\to(t,x,y)\\ s\geq 0, w,z\in(0,+\infty)}} u(s,w,z),$$
$$u_*(t,x,y) = \liminf_{\substack{(s,w,z)\to(t,x,y)\\ s\geq 0, w,z\in(0,+\infty)}} u(s,w,z).$$

With these definitions, we can give the sense of viscosity solution of (2.5), according to [2, 3, 12].

Definition 2.4. *(i) u is a discontinuous viscosity subsolution of (2.4) if for any $(\hat{t},\hat{x},\hat{y}) \in [0,T) \times (0,+\infty)^2$, and any $\phi \in C^2\left([0,T) \times (0,+\infty)^2\right)$, such that $(\hat{t},\hat{x},\hat{y})$ is a local maximum of $u^* - \phi$:*

$$\Lambda^-(J(\hat{t},\hat{x},\hat{y}), \partial_t\phi(\hat{t},\hat{x},\hat{y}), D\phi(\hat{t},\hat{x},\hat{y}), D^2\phi(\hat{t},\hat{x},\hat{y}))) \leq 0.$$

(ii) u is a discontinuous viscosity supersolution of (2.4) if for any $(\hat{t},\hat{x},\hat{y}) \in [0,T) \times (0,+\infty)^2$, and any $\phi \in C^2\left([0,T) \times (0,+\infty)^2\right)$, such that $(\hat{t},\hat{x},\hat{y})$ is a local minimum of $u_* - \phi$:

$$\Lambda^-(J(\hat{t},\hat{x},\hat{y}), \partial_t\phi(\hat{t},\hat{x},\hat{y}), D\phi(\hat{t},\hat{x},\hat{y}), D^2\phi(\hat{t},\hat{x},\hat{y}))) \geq 0.$$

(iii) u is a discontinuous viscosity solution of (2.4) if it is both sub and a super solution.

Theorem 2.5. *Under assumptions (A1)-(A2), the value function ϑ is a viscosity discontinuous solution of (2.5):*

$$\Lambda^-\begin{pmatrix} -\frac{\partial\vartheta}{\partial t} + \mu(t,y)\frac{\partial\vartheta}{\partial y} - \frac{1}{2}\sigma^2(t,y)x^2\frac{\partial^2\vartheta}{\partial x^2} & -\frac{1}{2}\sigma(t,y)x\frac{\partial^2\vartheta}{\partial x\partial y} \\ -\frac{1}{2}\sigma(t,y)x\frac{\partial^2\vartheta}{\partial x\partial y} & -\frac{1}{2}\frac{\partial^2\vartheta}{\partial y^2} \end{pmatrix} = 0.$$

Moreover ϑ is a discontinuous viscosity supersolution of

$$-\frac{\partial^2\vartheta}{\partial y^2} \geq 0. \tag{2.8}$$

Proof. The proof is splitted into two parts: the supersolution property and the subsolution property.

(a) Supersolution property. By a classical application of the Dynamic Programming Principle, as done in [18], we obtain that $\vartheta(t,x,y)$ is a viscosity supersolution of

$$-\frac{\partial\vartheta}{\partial t} + H(t,x,y,D\vartheta,D^2\vartheta) \geq 0.$$

Then, Lemma 2.3(i) implies that also

$$\Lambda^-(J(t,x,y,\partial_t\vartheta,D\vartheta,D^2\vartheta)) \geq 0,$$

and then ϑ is also a viscosity supersolution of (2.5).

Moreover, this last inequality implies that $-\frac{1}{2}\frac{\partial^2\vartheta}{\partial y^2} \geq 0$, and hence (2.8) is verified.

(b) Subsolution property. Let φ be a smooth function, and let $(\bar{t},\bar{x},\bar{y})$ be a strict maximizer of $\vartheta^* - \varphi$, such that

$$0 = (\vartheta^* - \varphi)(\bar{t},\bar{x},\bar{y}).$$

Suppose that $(\bar{t},\bar{x},\bar{y})$ belongs to the set $\mathcal{M}(\varphi)$ defined by:

$$\mathcal{M}(\varphi)=\{(t,x,y)\in[0,T)\times(0,+\infty)^2 : \Lambda^-(J(t,x,y,\partial_t\varphi(t,x,y),D\varphi(t,x,y),D^2\varphi(t,x,y)))>0\}$$

Since $\mathcal{M}(\varphi)$ is an open set, then there exists $\eta>0$ such that

$$[0\wedge(\bar{t}-\eta),\bar{t}+\eta]\times\overline{B}_\eta(\bar{x},\bar{y})\subset\mathcal{M}(\varphi),$$

where $\overline{B}_\eta(\bar{x},\bar{y})$ denotes the closed ball centered in $(\bar{x},\bar{y})$ and with radius η. From Lemma 2.3(iii), if $(t,x,y)\in\mathcal{M}(\varphi)$, then

$$-\frac{\partial\varphi}{\partial t}(t,x,y)+H(t,x,y,D\varphi(t,x,y),D^2\varphi(t,x,y))>0.$$

Using the Dynamic Programming Principle and the same arguments as in [19, Lemma 3.1], we get:

$$\sup_{\partial_p([0\wedge(\bar{t}-\eta),\bar{t}+\eta]\times\overline{B}_\eta(\bar{x},\bar{y}))}(\vartheta-\varphi)=\max_{[0\wedge(\bar{t}-\eta),\bar{t}+\eta]\times\overline{B}_\eta(\bar{x},\bar{y})}(\vartheta^*-\varphi),\tag{2.9}$$

where $\partial_p([t_1,t_2]\times\overline{B}_\eta(\bar{x},\bar{y}))$ is the forward parabolic boundary of $[t_1,t_2]\times\bar{B}_\eta(\bar{x},\bar{y})$, i.e. $\partial_p([t_1,t_2]\times\overline{B}_\eta(\bar{x},\bar{y}))=[t_1,t_2]\times\partial\overline{B}_\eta(\bar{x},\bar{y})\cup\{t_2\}\times\overline{B}_\eta(\bar{x},\bar{y})$. However, since $(\bar{t},\bar{x},\bar{y})$ is a strict maximizer of $\vartheta^*-\varphi$, equality (2.9) leads to a contradiction. Therefore, $(\bar{t},\bar{x},\bar{y})\notin\mathcal{M}(\varphi)$, and the result follows. □

In this work, we are interested by the numerical computation of the value function ϑ. Although equation (2.5) has a rigorous meaning, the formulation with the smallest eigenvalue makes difficult to deal with its numerical discretization. Of course, one can be tempted to modify the hamiltonian in the following way: for $\zeta_{\max}>0$,

$$H(t,x,y,p,Q)\cong\min_{\zeta\in[-\zeta_{\max},\zeta_{\max}]}\left\{\mu(t,y)p_2-\frac{1}{2}\mathrm{tr}(a(t,x,y,\zeta)\cdot Q)\right\}.$$

However, the choice of $\zeta_{\max}$, guaranteeing a good approximation of H, does not appear obvious to us. To avoid these difficulties, we first give an equivalent HJB equation satisfied by ϑ and which is formulated with bounded controls. More precisely, we have:

Corollary 2.6. *Under assumptions (A1)-(A3), the value function ϑ is a viscosity solution of the HJB equation:*

$$\inf_{\alpha_1^2+\alpha_2^2=1}\left\{\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}^T\begin{pmatrix}-\frac{\partial\vartheta}{\partial t}+\mu(t,y)\frac{\partial\vartheta}{\partial y}-\frac{1}{2}\sigma^2(t,y)x^2\frac{\partial^2\vartheta}{\partial x^2} & -\frac{1}{2}\sigma(t,y)x\frac{\partial^2\vartheta}{\partial x\partial y}\\ -\frac{1}{2}\sigma(t,y)x\frac{\partial^2\vartheta}{\partial x\partial y} & -\frac{1}{2}\frac{\partial^2\vartheta}{\partial y^2}\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}\right\}=0.\tag{2.10}$$

Proposition 2.7. *Let $\eta(t,x,y)$ be a continuous function such that $\eta(t,x,y)>0$ for all $(t,x,y)\in[0,T)\times(\mathbb{R}_+^*)^2$. Then equation (2.5) can be reformulated as follows,*

$$\Lambda^-\begin{pmatrix}-\frac{\partial\vartheta}{\partial t}+\mu(t,y)\frac{\partial\vartheta}{\partial y}-\frac{1}{2}\sigma^2(t,y)x^2\frac{\partial^2\vartheta}{\partial x^2} & -\frac{1}{2}\sigma(t,y)x\eta(t,x,y)\frac{\partial^2\vartheta}{\partial x\partial y}\\ -\frac{1}{2}\sigma(t,y)x\eta(t,x,y)\frac{\partial^2\vartheta}{\partial x\partial y} & -\frac{1}{2}\eta^2(t,x,y)\frac{\partial^2\vartheta}{\partial y^2}\end{pmatrix}=0.\tag{2.11}$$

In other words, Theorem 2.5 remains valid with (2.11) instead of (2.5).

Proof. This is obtained directly from (2.5) (adding the positive function $\eta(t,x,y)$ does not change the sign of the operator in (2.5) for fixed $(t,x,y,D\vartheta,D^2\vartheta)$). □

We shall use $\eta(t,x,y)=1$ unless otherwise specified. For numerical purposes it may be useful to choose $\eta\neq1$ (see for instance the uniform consistency result in Proposition 3.8).

2.2 Boundary conditions and unicity result

Unlike in most similar parabolic problems, here we do not only need a terminal condition to obtain the uniqueness, but also a border conditions when y tends to zero. Another boundary condition is hidden by the fact that we consider bounded solutions, which is, intuitively, equivalent to Neumann conditions near infinity.

Lemma 2.8. *Under assumptions (A1)-(A3), the value function ϑ is bounded and satisfies the following conditions on the boundaries $x = 0$ and $y = 0$:*

$$\lim_{(t',x',y')\to(t,x,0)} \vartheta(t',x',y') = \vartheta(t,x,0) = g(x), \forall\,(t,x) \in [0,T]\times\mathbb{R}_+^* \tag{2.12a}$$

$$\lim_{(t',x',y')\to(t,0,y)} \vartheta(t',x',y') = \vartheta(t,0,y) = g(0), \forall\,(t,y) \in [0,T]\times\mathbb{R}_+^* \tag{2.12b}$$

and the terminal condition of the equation for $t = T$ is:

$$\lim_{(t',x',y')\to(T,x,y)} \vartheta(t',x',y') = \vartheta(T,x,y) = g(x) \text{ for all } (x,y) \in (\mathbb{R}_+^*)^2. \tag{2.12c}$$

Proof. The statements (2.12a)-(2.12c) are proved in Lemma 5.6 in [9]. The proof is based on the assumptions (A1) and (A2) on σ and μ, and on the continuity and boundedness of g (see (A3)).

Now to prove statement (2.12b), we first give a representation of $\vartheta(t,x,y)$ using Doleans integral. Indeed, for every (t,x,y), we have:

$$X_{t,x,y}^{\rho,\zeta} = xZ_y^{\zeta,\rho}, \quad \text{where} \quad Z_y^{\zeta,\rho} := e^{\int_t^T \sigma(s,Y_{t,y}^{\rho,\zeta}(s))dW_s^1 + \frac{1}{2}\int_t^T (\sigma(s,Y_{t,y}^{\rho,\zeta}(s)))^2 ds}.$$

Therefore,

$$\vartheta(t,x,y) = \mathbb{E}\Big[g(X_{t,x,y}^{\rho,\zeta})(T)\Big] = \mathbb{E}\Big[g\Big(xZ_y^{\zeta,\rho},\Big)\Big]. \tag{2.13}$$

We conclude that statement (2.12b) holds. □

We recall here the uniqueness result, proved in Lemma 4.3, Proposition 4.4, and Proposition 4.6 of [9].

Theorem 2.9. *(Proposition 4.4 of [9]) Assume (A1)-(A3). Suppose that u is an upper semi-continuous viscosity subsolution of (2.5) bounded from above, and w a lower semi-continuous viscosity supersolution of (2.5) bounded from below. If, furthermore,*

$$\begin{array}{rcccl} u(T,x,y) & \leq & g(x) & \leq & w(T,x,y), \\ u(t,x,0) & \leq & g(x) & \leq & w(t,x,0), \end{array} \tag{2.14}$$

then $u(t,x,y) \leq w(t,x,y)$, for all $(t,x,y) \in [0,T]\times\mathbb{R}_+^2$. In particular, the solution of (2.5) in the viscosity sense with boundary conditions (2.12a) and (2.12c) is unique.

We give here the main ideas of the proof.

Proof. Suppose that u and w are respectively subsolution and supersolution of (2.5), and that they both satisfy the limit conditions (2.12a) and (2.12c). A classical argument (see [4]) to prove uniqueness for equation as (2.5), consists in building a strict viscosity supersolution of (2.5) w_ε, depending on the supersolution and on a parameter ε. Moreover w_ε must to be

such that, when the parameter ε goes to zero, w_ε tends to w. Then with classical arguments [12], a comparison principle between the strict supersolution and the subsolution can be obtained, and sending ε to zero we have the desired comparison principle.

In our particular case, for any $\varepsilon > 0$, we build

$$w_\varepsilon = w + \varepsilon((T-t) + \ln(1+y)).$$

From Lemma 4.3 of [9], w_ε is a strict viscosity supersolution of (2.5), bounded from below and such that conditions (2.14) are satisfied. Then we can apply Proposition 4.6 of [9] which is a comparison principle between a strict viscosity supersolution and a viscosity subsolution, and we obtain

$$w_\varepsilon \geq u,$$

for all $(t,x,y) \in [0,T] \times \mathbb{R}_+^2$. Sending ε to zero, we have the result. □

Since the boundedness property of ϑ would be tricky to manipulate numerically, in the following proposition we give some growth properties of the value function which are a sort of Neumann conditions at infinity. These conditions will guide us toward an implementable scheme.

Proposition 2.10. *Assume that (A1)-(A4) are satisfied. Then the following holds:*
(i) For any $a > 0$, the function:

$$h^1_{t,y} : x \to \vartheta(t, x+a, y) - \vartheta(t,x,y)$$

converges to zero, uniformly in (t,y), when $x \to +\infty$.
(ii) The function

$$h^2_{t,x} : y \to \vartheta(t, x, y+a) - \vartheta(t,x,y)$$

converges to zero, uniformly in (t,x), when $y \to +\infty$.

Proof. (i) Let $(t,x,y) \in (0,T) \times \mathbb{R}^+ \times \mathbb{R}^+$. As in (2.13), we have:

$$\vartheta(t,x,y) = \sup_{\zeta,\rho} \mathbb{E}\left[g\left(X^{\rho,\zeta}_{t,x,y}(T)\right)\right] = \sup_{\zeta,\rho} \mathbb{E}\left[g\left(xZ^{\zeta,\rho}_y\right)\right]. \tag{2.15}$$

By assumption (A3), the function $f : z \to g(e^z)$ is Lipschitz continuous on $\mathbb{R}$. Then, for $x' \in \mathbb{R}^+$, we get:

$$\begin{aligned}
\vartheta(t,x,y) - \vartheta(t,x',y) &= \sup_{\zeta,\rho} \mathbb{E}\left(g\left(xZ^{\zeta,\rho}_y\right)\right) - \sup_{\zeta,\rho} \mathbb{E}\left(g\left(x'Z^{\zeta,\rho}_y\right)\right) \\
&\leq \sup_{\zeta,\rho} \left\{\mathbb{E}\left(g\left(xZ^{\zeta,\rho}_y\right)\right) - \mathbb{E}\left(g\left(x'Z^{\zeta,\rho}_y\right)\right)\right\} \\
&\leq \sup_{\zeta,\rho} \left\{\mathbb{E}\left[f\left(\ln(x) + \ln\left(Z^{\zeta,\rho}_y\right)\right) - f\left(\ln(x') + \ln\left(Z^{\zeta,\rho}_y\right)\right)\right]\right\},
\end{aligned}$$

and using the Lipschitz property of f, it yields to:

$$\vartheta(t,x,y) - \vartheta(t,x',y) \leq K\left|\ln(x) - \ln(x')\right|.$$

Therefore we get that

$$\left|h^1_{t,y}(x)\right| \leq K\left|\ln\left(\frac{x+a}{x}\right)\right| \to 0 \text{ as } x \to +\infty \text{ uniformly in } (t,y). \tag{2.16}$$

To prove assertion (ii), using (2.8), we see that ϑ is a supersolution of

$$-\frac{\partial^2 v}{\partial y^2} = 0.$$

Then, from [13], we deduce that the function ϑ is concave w.r.t. y. That is, for each (t,x), $\vartheta(t,x,\cdot)$ is a concave function. Moreover, from (A3), ϑ is bounded and $\|\vartheta\|_\infty \leq M_0$ (where the constant $M_0 > 0$ is the same as in (A3)). Therefore, for any λ, the function

$$h^2_{t,x} : y \to \vartheta(t,x,y+\lambda) - \vartheta(t,x,y)$$

is decreasing. Considering that $\vartheta(t,x,n\lambda+y_0) = \vartheta(t,x,y_0) + \sum_{i=1}^n h^2_{t,x}(i\lambda+y_0)$. Hence, it follows that:

$$\vartheta(t,x,n\lambda+y_0) \geq \vartheta(t,x,y_0) + \sum_{i=1}^{n} h^2_{t,x}(n\lambda+y_0)$$

which gives:

$$h^2_{t,x}(n\lambda+y_0) \leq \frac{2M}{n}$$

and we get convergence of $h^2_{t,x}(y)$ to 0, which is uniform in (t,x). □

2.3 Lipschitz property

Here we establish the Lipschitz property of the value function ϑ.

Proposition 2.11. *Under assumptions (A1)-(A4), we have:*

(i) The value function ϑ is Lipschitz w.r.t. x.

(ii) ϑ is Lipschitz w.r.t. y.

Proof. (i) As in the proof of proposition 2.10, we consider the representation of ϑ using Doleans exponential:

$$\vartheta(t,x,y) = \sup_{\zeta,\rho} \mathbb{E}\left(g(X^{\zeta,\rho}_{t,x,y})\right) = \sup_{\zeta,\rho} \mathbb{E}\left[g\left(xZ^{\zeta,\rho}_y\right)\right] \qquad \forall t \in (0,T), x, y \in \mathbb{R}^+, \tag{2.17}$$

where $Z^{\zeta,\rho}_y = e^{\int_t^T \sigma(s,Y^{\rho,\zeta}_{t,y}(s)dW^1_s + \frac{1}{2}\int_t^T (\sigma(s,Y^{\rho,\zeta}_{t,y}(s)))^2 ds}$.

Then, for $t \in (0,T), x, x', y \in \mathbb{R}^+$ we have:

$$\left|\vartheta(t,x,y) - \vartheta(t,x',y)\right| \leq \sup_{\zeta,\rho} \mathbb{E}\left[g\left(xZ^{\zeta,\rho}_y\right) - g\left(x'Z^{\zeta,\rho}_y\right)\right].$$

As g is Lipschitz (assumption (A3)), there exists a constant $K \geq 0$ such that:

$$\left|\vartheta(t,x,y) - \vartheta(t,x',y)\right| \leq \sup_{\zeta,\rho} \mathbb{E}\left|K(x-x')Z^{\zeta,\rho}_y\right| \leq K|x-x'| \sup_{\zeta,\rho} \mathbb{E}\left(Z^{\zeta,\rho}_y\right).$$

Therefore, using the fact that the Doleans exponential is a positive local martingale, and hence a supermartingale, which implies that for any control $(\zeta,\rho) \in \mathcal{U}$:

$$\mathbb{E}\left(e^{\int_t^T \sigma^{\zeta,\rho}_u dW_u + \frac{1}{2}\int_t^T (\sigma_u)^{\zeta,\rho})^2 du}\right) \leq 1,$$

and then taking the supremum leads to:

$$\left|\vartheta(t,x,y)-\vartheta(t,x',y)\right|\leq K|x-x'|$$

which proves that $\vartheta(t,.,y)$ is a K-Lipschitz function.

(ii) Now we treat the Lipschitz property of ϑ w.r.t. y.

First,we recall that ϑ is concave w.r.t. y. Furthermore, as g is bounded, we immediately get that ϑ shares the same bound. Hence, it is sufficient to prove that ϑ is Lipschitz near the boundary $y=0$.

Recall that by (2.12a), we know that $\vartheta(t,x,0)=g(x)$ for all $(t,x)\in(0,T)\times(0,+\infty)$.

Let $(t,x,y)\in[0,T]\times(0,+\infty)^2$, with $y>0$. For any control $(\zeta,\rho)\in\mathcal{U}$, we have:

$$Y_{t,y}^{\rho,\zeta}(s)=y+\int_t^s -\mu(\tau,Y_{t,y}^{\rho,\zeta}(\tau))d\tau+\int_t^s \zeta(\tau)Y_{t,y}^{\rho,\zeta}\tau\, dW_\tau^2.$$

Furthermore, by a comparison argument for SDEs, we get, for any $\tau\in[t,T]$:

$$Y_{t,y}^{\rho,\zeta}(\tau)\geq 0.$$

Using the positivity of μ, we get:

$$0\leq Y_{t,y}^{\rho,\zeta}(s)\leq y+\int_t^s Y_{t,y}^{\rho,\zeta}(\tau)dW_\tau^2.$$

Hence, the quantity above is a super-martingale and we get:

$$\mathbb{E}\left[Y_{t,y}^{\rho,\zeta}(s)\right]\leq y. \tag{2.18}$$

Now, applying Itô's formula on $g(X_{t,x,y}^{\rho,\zeta})$:

$$\begin{aligned}
g(X_{t,x,y}^{\rho,\zeta}(s)) &= g(x)+\int_t^s g'(X_{t,x,y}^{\rho,\zeta}(\tau))dX_{t,x,y}^{\rho,\zeta}(\tau)+\\
&\quad \frac{1}{2}\int_t^s g''(X_{t,x,y}^{\rho,\zeta}(\tau))\left\langle dX_{t,x,y}^{\rho,\zeta}(\tau),dX_{t,x,y}^{\rho,\zeta}(\tau)\right\rangle\\
&= g(x)+\int_t^s g'(X_{t,x,y}^{\rho,\zeta}(\tau))dX_{t,x,y}^{\rho,\zeta}(\tau)+\\
&\quad \frac{1}{2}\int_t^s \left(X_{t,x,y}^{\rho,\zeta}(\tau)\right)^2 g''(X_{t,x,y}^{\rho,\zeta}(\tau))\sigma^2(Y_{t,y}^{\rho,\zeta}(\tau))d\tau.
\end{aligned}$$

Since $X_{t,x,y}^{\rho,\zeta}$ is a locale martingale, there exists a sequence $(s_n)_n$, with $s_n\to\infty$ such that:

$$\mathbb{E}\left(\int_t^{s_n\wedge T} g'(X_{t,x,y}^{\rho,\zeta}(u))dX_{t,x,y}^{\rho,\zeta}(u)\right)=0.$$

Using (2.18), the Lipschitz property of σ^2, and the boundedness of $x\longmapsto x^2g''(x)$, it yields: there exists a constant $C>0$, such that:

$$\left|\mathbb{E}\left(g(X_{t,x,y}^{\rho,\zeta}(s_n\wedge T))-g(x)\right)\right|\leq\int_t^{s_n\wedge T} Cy\,d\tau.$$

Finally, as g is bounded, we conclude with Fatou's lemma that:

$$\mathbb{E}\left(g(X^{\rho,\zeta}_{t,x,y}(T)) - g(x)\right) \leq C(T-t)y,$$

and since the constant C is independent of ρ, ζ, we obtain:

$$\begin{aligned}|\vartheta(t,x,y) - \vartheta(t,x,0)| &\leq \sup_{(\rho,\zeta)\in\mathcal{U}} \left\{|\mathbb{E}\left(g(X^{\rho,\zeta}_{t,x,y}(T)) - g(x)\right)|\right\} \\ &\leq CTy.\end{aligned}$$

Hence, as ϑ is concave w.r.t. y and bounded, it is Lipschitz with respect to y. □

3 Approximation Scheme

In this section we want to approximate the bounded solution of Equation (2.11) formulated as:

$$\min_{\substack{\alpha=(\alpha_1,\alpha_2)\\ \alpha_1^2+\alpha_2^2=1}} \left\{ -\alpha_1^2 \frac{\partial\vartheta}{\partial t}(t,x,y) + \mu(t,y)\alpha_1^2 \frac{\partial\vartheta}{\partial y}(t,x,y) - \frac{1}{2}\mathrm{Tr}[a(\alpha,t,x,y)D^2\vartheta(t,x,y)]\right\} = 0, \tag{3.1}$$

with boundary conditions (2.12a) and (2.12c), where μ is a positive Lipschitz function, the diffusion matrix a beeing now defined as follows:

$$\begin{aligned} a(\alpha,t,x,y) &:= \begin{pmatrix} \alpha_1^2\sigma^2(t,y)x^2 & \alpha_1\alpha_2\sigma(t,y)x\eta(t,x,y) \\ \alpha_1\alpha_2\sigma(t,y)x\eta(t,x,y) & \alpha_2^2\eta^2(t,x,y)\end{pmatrix} \\ &= \begin{pmatrix} \alpha_1\sigma(t,y)x \\ \alpha_2\eta(t,x,y)\end{pmatrix}\begin{pmatrix} \alpha_1\sigma(t,y)x \\ \alpha_2\eta(t,x,y)\end{pmatrix}^{\mathsf{T}}. \end{aligned} \tag{3.2}$$

From now on a, μ, σ and η will stand for $a(\alpha,t,x,y)$, $\mu(t,y)$, $\sigma(t,y)$ and $\eta(t,x,y)$ if there is no ambiguity.

We remark that a is *not* a diagonal dominant matrix[1], because we cannot ensure that

$$\alpha_2\eta \geq \alpha_1\sigma x, \quad \forall\,(t,x,y)\in[0,T)\times[0,+\infty)^2, \text{ and } \forall(\alpha_1,\alpha_2) \text{ s.t. } \alpha_1^2+\alpha_2^2 = 1.$$

This fact implies that we cannot choose the classical Finite Differences scheme to approximate equation (3.1). Here we shall use the Generalized Finite Differences scheme introduced in [7].

3.1 Generalized finite differences scheme

Consider a regular grid G_h on $\mathbb{R}^2_+$, with discretization space step $h > 0$:

$$G_h := \left\{(x_i,y_j),\ x_i := ih,\ y_j := jh,\ i,j\in\mathbb{N}\times\mathbb{N}^*\right\} = (h\mathbb{N})\times(h\mathbb{N}^*),$$

[1] We recall that a matrix X of dimension $N\times N$ is diagonal dominant if

$$X_{ii} \geq \sum_{i\neq j}|X_{ij}|, \quad \forall\, i = 1,\dots,N.$$

(the nodes $(x_i, 0)$ with $i \in \mathbb{N}$, are excluded from the grid), and let $\Delta t > 0$ be a time step.

The main idea of the Generalized Finite Differences scheme consists in the approximation of the diffusion term $\mathrm{Tr}(a \cdot D^2\phi)$ by a linear combination of elementary diffusions $\Delta_\xi^h \phi$ pointing towards grid points. More precisely, for a given direction $\xi = (\xi_1, \xi_2)^T \in \mathbb{Z}^2$, we define a second order finite difference operator (for $x, y \in \mathbb{R}$) as follows:

$$\Delta_\xi^h \phi(x,y) := \frac{1}{h^2} \left(\phi(x + \xi_1 h, y + \xi_2 h) + \phi(x - \xi_1 h, y - \xi_2 h) - 2\phi(x,y) \right).$$

By a Taylor expansion, we have

$$\begin{aligned} \Delta_\xi^h \phi(x,y) &= \sum_{i,j=1}^{2} \xi_i \xi_j \frac{\partial^2 \phi}{\partial x_i \partial x_j}(x,y) + ||\xi||^4 \, O(h^2) \\ &= \mathrm{Tr}[\xi \xi^T \cdot D^2\phi] + ||\xi||^4 \, O(h^2) \end{aligned} \tag{3.3}$$

(where x_1 and x_2 stand for x and y in (3.3)).

Now let p be in $\mathbb{N}^*$. The covariance matrix a defined in (3.2) is of rank one. Therefore, we have two cases.

Case 1. The direction of diffusion $\begin{pmatrix} \alpha_1 \sigma x \\ \alpha_2 \eta \end{pmatrix}$ points toward a grid point (rh, qh), with $r, q \in \mathbb{Z}^2$ and $|r|, |q| \leq p$. Then we consider the vector $\xi = \xi_{r,q} := \begin{pmatrix} r \\ q \end{pmatrix}$ and we have $a = \gamma_\xi^\alpha \xi \xi^\mathsf{T}$ with $\gamma_\xi^\alpha = \mathrm{Tr}(a)/||\xi||^2$, for a given $\alpha = (\alpha_1, \alpha_2)$. The second order diffussion term $\mathrm{Tr}(a \cdot D^2\phi)$ can be approximated by:

$$\mathrm{Tr}(a \cdot D^2\phi) = \gamma_\xi^\alpha \Delta_\xi^h \phi + ||\xi||^4 \, O(h^2)$$

Case 2. In general the direction of the diffusion $\begin{pmatrix} \alpha_1 \sigma x \\ \alpha_2 \eta \end{pmatrix}$ has a real slope, or is pointed towards a grid point (rh, qh) with $\max(|r|, |q|) > p$. In this case, we consider a set of natural integers

$$\mathcal{Q}_p := \{ \xi = (\xi_1, \xi_2)^T \in \mathbb{Z} \times \mathbb{N} \ \max(|\xi_1|, \xi_2) \leq p, \ (|\xi_1|, \xi_2) \text{ irreducible} \},$$

and the associated cone of positive symmetric matrices

$$\mathcal{C}(\mathcal{Q}_p) = \left\{ \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi \xi \xi^T, \gamma_\xi \geq 0 \right\}.$$

(This cone is also generated by the matrices $\xi\xi^T$ where $\xi = (r,q)^T$ for all $(r,q) \in \mathbb{Z}^2$ and such that $|r|, |q| \leq p$). Then we consider a particular projection a_p of a on $\mathcal{C}(\mathcal{Q}_p)$ that can be written as the sum of two terms as follows:

$$a_p = \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha \xi \xi^T = \gamma_{\xi^+}^\alpha \xi^+ \xi^{+T} + \gamma_{\xi^-}^\alpha \xi^- \xi^{-T} \tag{3.4}$$

where $\gamma_\xi \geq 0$, and $\xi^\pm$ are two elements of $\mathcal{Q}_p$ that depend on (α, t, x, y) ($\xi^\pm$ correspond to the two closest directions to $\begin{pmatrix} \alpha_1 \sigma x \\ \alpha_2 \eta \end{pmatrix}$).

Thus we approximate $\mathrm{Tr}(a \cdot D^2\phi)$ by $\mathrm{Tr}(a_p \cdot D^2\phi)$, and use that

$$\mathrm{Tr}(a_p \cdot D^2\phi) = \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha \Delta_\xi^h \phi + \left(\sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha ||\xi||^4 \right) O(h^2) \tag{3.5}$$

(by (3.3) and (3.4)).

Representation of symmetric positive matrices. We recall that, as in [7], we can represent the 2×2 symmetric matrix a by an element of $\mathbb{R}^3$ using the following coordinates:

$$z_1 = a_{11}, \quad z_2 = \sqrt{2}\, a_{12}, \quad z_3 = a_{22}. \tag{3.6}$$

The cone of positive symmetric matrices is then defined as the set $\{z \in \mathbb{Z}^3,\ z_1, z_3 \geq 0,\ z_2^2 \leq 2 z_1 z_3\}$ and represented in Figure 1 (a), together with the cone $\mathcal{C}(\mathcal{Q}_1)$ of diagonally dominant matrices. Cuts of the cones $\mathcal{C}(\mathcal{Q}_1)$, $\mathcal{C}(\mathcal{Q}_2)$, and $\mathcal{C}(\mathcal{Q}_3)$ with the plan of trace one matrices ($z_1 + z_3 = 1$) are represented in Figure 2.

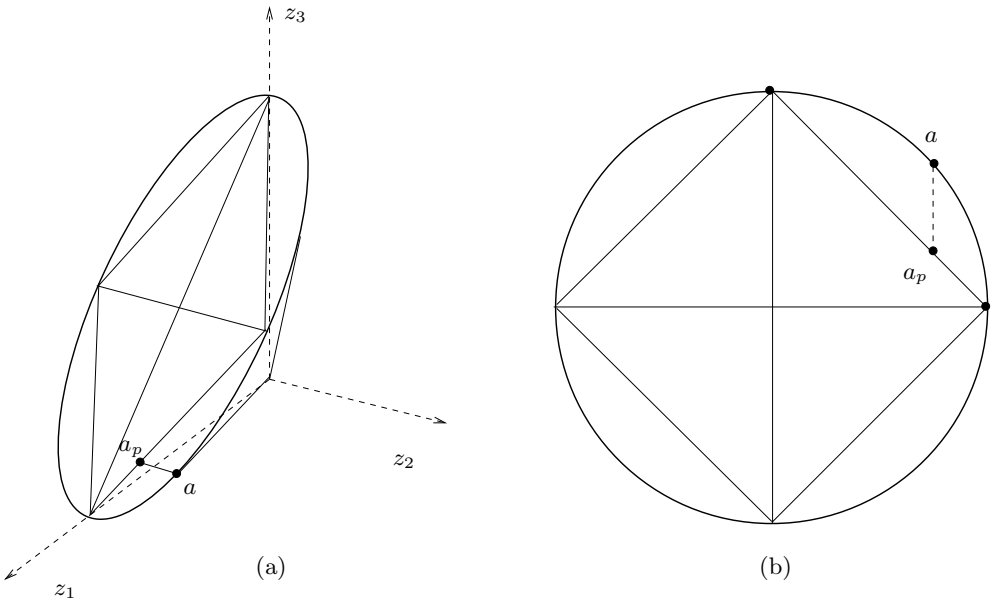

Figure 1: (a) Cone of positive definite matrices, embedding the cone of diagonally dominant matrices $\mathcal{C}(\mathcal{Q}_1)$; projection of a matrix a on $\mathcal{C}(\mathcal{Q}_1)$, in the case $\mathrm{Tr}(a) = 1$ (b) Same figure, where we draw the cut of the cone with trace one matrices (i.e, such that $z_1 + z_3 = 1$).

Definition of a_p. For any $p \geq 1$, the matrix a_p is defined as the projection of a on the cone $\mathcal{C}(\mathcal{Q}_p)$ parallely to the z_2 axis (see Fig. 1). The order p is the order of neighboring points allowed to enter in the scheme (this order may depends on where we are situated on the grid and on the direction of the diffusion). We notice that a_{11} and a_{22} are unchanged by the projection and only a_{12} is modified (since only z_2 is modified):

$$a = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \quad \rightarrow \quad a_p = \begin{pmatrix} a_{11} & a'_{12} \\ a'_{12} & a_{22} \end{pmatrix}. \tag{3.7}$$

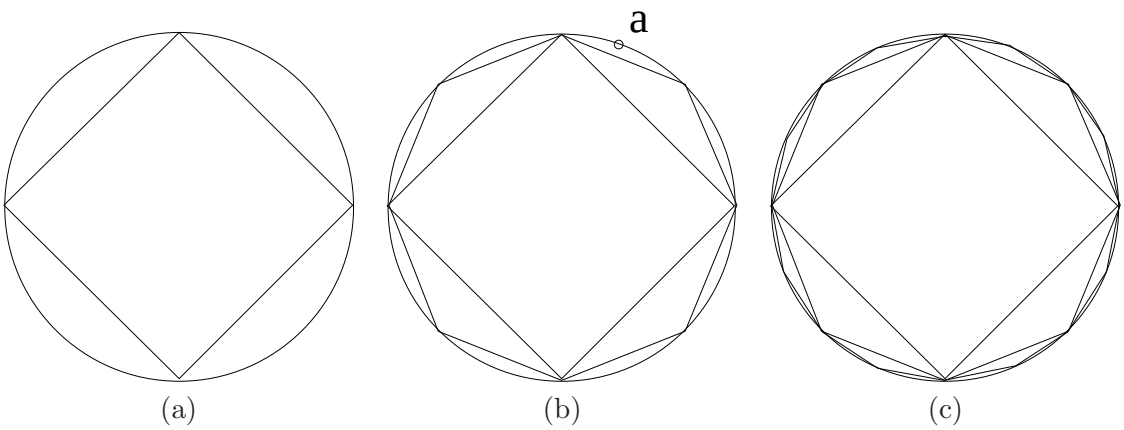

Figure 2: (a) Symmetric semi-definite positive matrix with trace equal to 1 and cone of diagonal dominant matrix. (b) Cone $\mathcal{C}(\mathcal{Q}_1)$, a is on the border of the semi-definite positive matrix. (c) Cone $\mathcal{C}(\mathcal{Q}_2)$.

In the special case of a border grid point $(x = 0, y)$, we see that a is of the form $a = \gamma_\xi^\alpha \xi\xi^T$ with $\xi = (0, 1)^T$ hence $a \in \mathcal{Q}_1$ and there is no projection error ($a_p = a$).

Remark 3.1. *Our projection is not the same as the orthogonal projection used in [7]. This modification is important to prove the global convergence of our scheme (see section 5). However the generation of the optimal directions $\xi^\pm$ are the same as for the orthogonal projection of [7], and can be performed in $O(p)$ operations by using Stern-Brocot algorithm [16]. These directions, geometrically, corresponds to the two closest directions (in angle) from the direction* $\begin{pmatrix} \alpha_1\sigma x \\ \alpha_2\eta \end{pmatrix}$ *defining the matrix a.*

Remark 3.2. *The choice of the order p depends on where we are situated on the grid. For instance, if we consider a point (x, y) in the middle of the grid, and we want to discretize $\mathrm{Tr}(a \cdot D^2\phi(t, x, y))$, we can follow the direction of diffusion and choose the biggest order of discretization p, because this will give a better approximation of the covariance matrix a. On the other hand, if we consider a point (x, y) near to the boundary, it can often happen that following the direction of the diffusion, we involve in the discretization some points which are out of the grid. In this case the choice of p will be reduce in order to stay in the grid.*

Remark 3.3. *In all the decompositions, the coefficients γ_ξ^α and also the vectors ξ are in terms of α and (t, x, y). For simplicity of notations we may omit to specify this dependence.*

For a symmetric matrix $b = (b_{ij})$ of dimension 2 we consider the Frobenius norm

$$\|b\|_F := \mathrm{Tr}(bb^T)^{1/2} = (\sum_{i,j=1,2} b_{ij}^2)^{1/2}.$$

and the following notations

$$|\partial_t^2\phi|_0 := ||\frac{\partial^2\phi}{\partial t^2}||_{L^\infty((0,T)\times\mathbb{R}^2_+)},$$
$$|\partial_y^2\phi|_0 := ||\frac{\partial^2\phi}{\partial y^2}||_{L^\infty((0,T)\times\mathbb{R}^2_+)},$$
$$|D^k\phi|_0 := \max_{i,j\geq 0,\ i+j=k} ||\frac{\partial^k\phi}{\partial x^i\partial y^j}||_{L^\infty((0,T)\times\mathbb{R}^2_+)}, \quad \text{for } k\in\mathbb{N}.$$

Then we have the following error estimates.

Lemma 3.4 (Error projection)**.** *Let $(t,x,y,\alpha=(\alpha_1,\alpha_2))$ be given in $[0,T]\times\mathbb{R}^2_+\times\mathbb{R}^2$, with $\alpha_1^2+\alpha_2^2=1$, and let a^p be the projected matrix associated to a as defined in (3.2).*
(i) For $p\geq 1$, we have

$$||a(\alpha,t,x,y)-a_p(\alpha,t,x,y)||_F \leq \frac{2}{p}\mathrm{Tr}(a(\alpha,t,x,y)).$$

(ii) For any $p\geq 1$ and any regular function ϕ, we have

$$\left|\mathrm{Tr}(a(\alpha,t,x,y)\cdot D^2\phi(t,x,y)) - \sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^\alpha\Delta_\xi^h\phi(t,x,y)\right|$$
$$\leq \quad 4|D^2\phi|_0\frac{\mathrm{Tr}(a(\alpha,t,x,y))}{p} + \frac{2}{3}|D^4\phi|_0\mathrm{Tr}(a(\alpha,t,x,y))\,p^2h^2$$

Proof. *(i)* It suffices to consider the case when $\mathrm{Tr}(a)=1$. The norm $||a-a_p||_F$ is also the Euclidean distance in $\mathbb{R}^3$ beetwen the two matrices a and a_p represented using their coordinates as in (3.6).

Let $\xi^\pm$ be the two vectors of $\mathcal{Q}_p$ associated to non-zero α_ξ. Then following the arguments of [7], we have

$$||a-a_p||_F \leq ||\frac{\xi^+}{||\xi^+||}-\frac{\xi^-}{||\xi^-||}|| \leq 2(\widehat{\xi^-\xi^+}) \leq \frac{2}{p}.$$

where $(\widehat{\xi^-\xi^+})$ denotes the angle of between the vectors ξ^+ and ξ^-.

(ii) First using (3.3) (more precisely, $|\Delta_\xi^h\phi - \mathrm{Tr}[\xi\xi^T\cdot D^2\phi]| \leq 2\frac{h^2}{4!}(\sum_{i+j=4}C_4^i|\xi_1^i\xi_2^j|)|D^4\phi|_0$), $\sum_{i+j=4}C_4^i|\xi_1|^i|\xi_2|^j = (|\xi_1|+|\xi_2|)^4| \leq 4||\xi||^4$, and the fact that $||\xi||^2\leq 2p^2$, we obtain

$$\left|\mathrm{Tr}(a_p\cdot D^2\phi) - \sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^\alpha\Delta_\xi^h\phi\right| \leq \frac{2}{3}|D^4\phi|_0(\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^\alpha||\xi||^2)\,p^2h^2 = \frac{2}{3}|D^4\phi|_0\mathrm{Tr}(a_p)\,p^2h^2.$$

Also $\mathrm{Tr}(a_p)=\mathrm{Tr}(a)$ by using (3.7). Then we have

$$\left|\mathrm{Tr}(a\cdot D^2\phi) - \sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^\alpha\Delta_\xi^h\phi\right| \leq \left|\mathrm{Tr}(a\cdot D^2\phi) - \mathrm{Tr}(a_p\cdot D^2\phi)\right| + \frac{2}{3}|D^4\phi|_0\mathrm{Tr}(a)\,p^2h^2. \quad (3.8)$$

To conclude the proof we use (3.8) together with the inequalities $\mathrm{Tr}(AB)\leq||A||_F||B||_F$ for any symmetric matrices A,B, $||A||_F\leq\mathrm{Tr}(A)$ for any symetric and positive matrix A, and $||D^2\phi||_F\leq 2|D^2\phi|_0$. □

Remark 3.5. *Using an orthogonal projection as in [7], we would obtain $||a-a_p||_F\leq\frac{1}{4p^2}\mathrm{Tr}(a)$.*

3.2 The discrete equation

From now on, $\lceil r \rceil$ will denote the smallest integer greater than r, $p_{\max} \in \mathbb{N}$ the maximal order of grid points allowed to enter in the scheme ($p_{\max}$ will typically depend of h), and Δt the time step. Set $\rho = (\Delta t, h, p_{\max})$, $r \in \mathbb{R}$ and $\phi : [0,T] \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$, and define

$$p(x,y) := \max(1, \min(p_{max}, \lceil x/h \rceil, \lceil y/h \rceil)).$$

In particular, we have $p = p_{max}$ if $x - p_{max}h \geq 0$ and $y - p_{max}h \geq 0$ (points in the interior of the domain), or $p = \min(\lceil x/h \rceil, \lceil y/h \rceil)$ for the points near to the boundary and such that $x \neq 0$, and $p = 1$ in the case $x = 0$.

Now we define the function $\mathcal{S}^\rho$ as follows:

$$\begin{aligned}\mathcal{S}^\rho(t,x,y,r,\phi) \quad := \quad & \min_{\substack{\alpha_1^2+\alpha_2^2=1 \\ \alpha=(\alpha_1,\alpha_2)}} \left\{ -\alpha_1^2 \frac{\phi(t+\Delta t,x,y) - r}{\Delta t} + \alpha_1^2 \mu \frac{r - \phi(t,x,y-h)}{h} \right. \\ & \left. -\frac{1}{2} \sum_{\xi \in \mathcal{Q}_{p(x,y)}} \gamma_\xi^\alpha(t,x,y) \frac{\phi(t,x-\xi_1 h, y-\xi_2 h) - 2r + \phi(t,x+\xi_1 h, y+\xi_2 h)}{h^2} \right\}, \end{aligned} \tag{3.9}$$

for $(t,x,y) \in [0,T) \times (0,\infty)^2$. In (3.9) we have used the decomposition of the projected matrix associated to a: $a_p(\alpha,t,x,y) = \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha \Delta_\xi^h \phi$, with $p = p(x,y)$. (we recall that there are only two non vanishing terms γ_ξ^α in the sum).

Now let $t_n := n\Delta t$. The discrete scheme for (3.1) is defined as the bounded solution $v_h(t_n,.)$ (if it exists) of

$$\mathcal{S}^\rho(t_n,x,y,v_h(t_n,x,y),v_h) = 0, \quad (x,y) \in G_h, \tag{3.10a}$$

for $n = N-1, \ldots, 1, 0$ and with the boundary conditions:

$$v_h(T,x,y) = g(x), \quad \forall (x,y) \in G_h, \tag{3.10b}$$

$$v_h(t_n,x,0) = g(x), \quad n = 0, \ldots, N, \ x \in h\mathbb{N}. \tag{3.10c}$$

For $t \in [t_n, t_{n+1}]$, $v_h(t,.)$ is defined by a P1 interpolation of $v_h(t_n,.,.)$ and $v_h(t_{n+1},.)$. The solution v_h will stand for an approximation of the value function ϑ.

We also define a continuous fonction $\mathcal{F}$, in view of the left hand side of (2.11), for $u \in \mathbb{R}$, $p = (p_1,p_2) \in \mathbb{R}^2$ and $Q = (Q_{ij})$ any 2×2 symmetric matrix, as follows:

$$\mathcal{F}(t,x,y,u,p,Q) := \inf_{\substack{\alpha=(\alpha_1,\alpha_2) \\ \alpha_1^2+\alpha_2^2=1}} \left\{ -\alpha_1^2 u + \alpha_1^2 \mu(t,y) p_2 - \frac{1}{2} \mathrm{Tr}[a(\alpha,t,x,y) \cdot Q] \right\}$$

Then our equation (2.11) is now equivalent to

$$\mathcal{F}(t,x,y,\partial_t \vartheta(t,x,y), D\vartheta(t,x,y), D^2\vartheta(t,x,y)) = 0$$

We remark that $\mathcal{F}$ is continous and backward parabolic (in the sense that it is decreasing in the variable u, and decreasing with respect to symetric matrices Q[2]). In view of the general abstract convergence result of [5], we shall use monotonity, stability and consistency results for the scheme (3.10) in order to obtain its convergence.

First we have the following result, easily deduced from Lemma 3.4(ii) and by standard first order consistency estimates.

[2] If $Q_2 - Q_1$ is positive then $\mathcal{F}(t,x,y,u,p,Q_1) \geq \mathcal{F}(t,x,y,u,p,Q_2)$

Lemma 3.6. *Let $\rho = (\Delta t, h, p_{\max})$, $(x,y) \in \mathbb{R}_+^2$, $t \in [0,T[$, $\phi \in C^4([0,T]\times[0,\infty)^2)$. We have*

$$\left|S^\rho(t,x,y,\phi(t,x,y),\phi) - \mathcal{F}(t,x,y,\partial_t\phi(t,x,y),D\phi(t,x,y),D^2\phi(t,x,y))\right| \\ \leq \frac{1}{2}|\partial_t^2\phi|_0\,\Delta t + \frac{1}{2}\mu|\partial_y^2\phi|_0\,h + 4|D^2\phi|_0\frac{\mathrm{Tr}(a)}{p} + \frac{2}{3}|D^4\phi|_0\mathrm{Tr}(a)\,p^2h^2, \tag{3.11}$$

where a_p is the projection of a on $\mathcal{C}(\mathcal{Q}_p)$ and $p := p(x,y)$. Furthermore the term $4|D^2\phi|_0\frac{\mathrm{Tr}(a)}{p}$ vanishes in the case $x = 0$.

Proposition 3.7. *The scheme (3.10) satisfies the following properties:*
(i) ***Monotonicity:*** *for all $r \in \mathbb{R}$, $x,y \in \mathbb{R}_+$, $u,v \in C([0,T]\times[0,\infty)^2)$ such that $u \leq v$, we have*

$$\mathcal{S}^\rho(t,x,y,r,u) \geq \mathcal{S}^\rho(t,x,y,r,v).$$

(ii) ***Stability:*** *For all $\rho = (h,\Delta t) \in (\mathbb{R}_+^*)^2$ and $p_{\max} \in \mathbb{N}^*$, there exists a bounded solution v_h of (3.10) such that $||v_h||_{L^\infty((0,T)\times\mathbb{R}_+^2)} \leq C_0 := ||g||_{L^\infty(\mathbb{R}_+)}$.*
(iii) ***Consistency:*** *Let $p_{\max}$ be such that $hp_{\max} \to 0$ as $h \to 0$. Then $\forall (x,y) \in (\mathbb{R}_+^*)2$, $\forall t \in [0,T[$, for every $\phi \in C^4([0,T]\times[0,\infty)^2)$,*

$$\lim_{\substack{(t_n,x_i,y_j)\to(t,x,y)\\ \Delta t\to 0,\ h\to 0,\ p_{\max}\to\infty}} S^\rho(t_n,x_i,y_j,\phi(t_n,x_i,y_j),\phi) = \mathcal{F}(t,x,y,\partial_t\phi(t,x,y),D\phi(t,x,y),D^2\phi(t,x,y))$$

The monotonicity property (i) is immediate, and the consistency property (iii) is deduced from the previous Lemma 3.6 and the continuity of $\mathcal{S}^\rho$. Hence there remains to prove the well-posedness of the scheme and the stability (ii) This will be done in Section 4.

We deduce also from the Lemma 3.6 and the fact that $\mathrm{argmin}_p(\frac{1}{p} + p^2h^2)$ is obtained for $p = 2^{-\frac{1}{3}}h^{-\frac{2}{3}}$, the following consistency error result.

Proposition 3.8. *Suppose that $\eta(t,x,y) := x\sqrt{y}$, and $p_{\max}\sim Ch^{-\frac{2}{3}}$ as $h \to 0$, for some constant $C > 0$. Let (t,x,y) be in $[0,T]\times\mathbb{R}_+\times\mathbb{R}_+$. Then as $\rho = (\Delta t, h) \to 0$ and $(t_n,x_i,y_j) \to (t,x,y)$, we have*

$$\left|S^\rho(t_n,x_i,y_j,\phi(t_n,x_i,y_j),\phi) - \mathcal{F}(t,x,y,\partial_t\phi(t,x,y),D\phi(t,x,y),D^2\phi(t,x,y))\right| \\ = \quad O(h^{\frac{2}{3}}) + O(\Delta t). \tag{3.12}$$

Furthermore the bound is uniform on compact sets with respect to (t,x,y).

Proof. The fact that $p_{\max} = o(\frac{1}{h})$ ensures that for any fixed point (x,y) with $x,y > 0$ and for h sufficiently small, $p = p_{\max}$ neighbor grid points can be used for the approximation of a and in (3.11), and the proof is immediate.

We only have to check that for points close to the boundary ($x = 0$ or $y = 0$) we can establish a uniform bound. The case of $p < p_{\max}$ happens when $p = p(x,y) = \min(\lceil x/h\rceil, \lceil y/h\rceil)$, so either $p = \lceil x/h\rceil$, or $p = \lceil y/h\rceil$.

In the case when $x = 0$, we have $p = 1$, there is no projection error, and the error term is bounded by $O(\Delta t + h)$.

In the case when $x \neq 0$, we first have the bound $\mathrm{Tr}(a) \leq Cx^2y + \eta(t,x,y)^2 \leq C'x^2y$ for some constant C'. If $p = \lceil \frac{x}{h} \rceil$, we have $p \geq \frac{x}{h}$ and $\frac{\mathrm{Tr}(a)}{p} \leq C'xyh = O(h)$. If $p = \lceil \frac{y}{h} \rceil$, we have $p \geq \frac{y}{h}$ and $\frac{\mathrm{Tr}(a)}{p} \leq C'x^2h = O(h)$. □

Remark 3.9. *In the case when the direction of the diffusion points toward a node of the grid, the consistency remains the same, except for the term* $4|D^2\phi|_0 \frac{\mathrm{Tr}(a)}{p}$ *which vanishes.*

4 Numerical solution: existence and stability

In this section we prove the well-posedness of the implicit scheme (3.10), and prove the stability property stated in Proposition 3.7(*ii*).

We first initialize the scheme by

$$v_h(T,x,y) := g(x), \qquad (x,y) \in G_h.$$

Then, given $v_h(t+\Delta t,.)$ for some time $t = t_n$, we need to find a bounded $v_h(t,.)$ such that

$$\min_{\substack{\alpha=(\alpha_1,\alpha_2)\\ \alpha_1^2+\alpha_2^2=1}} \left\{ \alpha_1^2 \frac{v_h(t,x,y) - v_h(t+\Delta t,x,y)}{\Delta t} + \alpha_1^2 \mu(t,y) \frac{v_h(t,x,y) - v_h(t,x,y-h_2)}{h_2} \right.$$
$$\left. -\frac{1}{2} \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha (\Delta_\xi^h v_h)(t,x,y) \right\} = 0, \quad \forall (x,y) \in G_h, \tag{4.1}$$

and with the following boundary conditions:

$$v_h(t,x,0) = g(x), \quad \forall x \in h\mathbb{N}. \tag{4.2}$$

Scheme in abstract form. Since for all $(x,y) \in G_h$ with $y > 0$, an optimal control (α_1,α_2) must be found, we introduce $S^1 := \{\alpha = (\alpha_1,\alpha_2), \alpha_1^2 + \alpha_2^2 = 1\}$ and

$$\mathcal{A} := (S^1)^{\mathbb{N}\times\mathbb{N}^*}$$

the set of controls associated to the grid mesh $G_h = h(\mathbb{N}\times\mathbb{N}^*)$.

The scheme can then be expressed in the following abstract form: find $X := v_h(t,.,.) \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}^*}$, bounded, such that

$$\min_{w\in\mathcal{A}} \Big(A(w)X - b(w) \Big) = 0, \tag{4.3}$$

where $A(w)$ is a linear operator on $\mathbb{R}^{\mathbb{N}\times\mathbb{N}^*}$, and $b(w)$ is a vector of $\mathbb{R}^{\mathbb{N}\times\mathbb{N}^*}$, and are made precise below.

Definition of the matrix $A(w)$ and vector $b(w)$: We denote by $X = (X_{ij})_{i\geq 0,\, j\geq 1}$, (resp. $w = (\alpha_{ij})_{i\geq 0,\, j\geq 1}$, with $\alpha_{ij} = (\alpha_{ij,1}, \alpha_{ij,2})$) values (resp. controls) corresponding to the mesh points (x_i, y_j) of G_h. Then

- $A(w)$ is an infinite matrix determined by $\forall X$, $\forall i \geq 0$, $\forall j \geq 1$,

$$\begin{aligned}(A(w)X)_{ij} &:= \frac{\alpha_{ij,1}^2}{\Delta t}X_{ij} + \alpha_{ij,1}^2\mu(t,y_j)\frac{1}{h}(X_{ij} - (1-\kappa_{j-1})X_{i,j-1}) \\ &\quad + \frac{1}{2}\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^{\alpha_{ij}}(-(1-\kappa_{j-\xi_2})X_{i-\xi_1,j-\xi_2} + 2X_{ij} - X_{i+\xi_1,j+\xi_2})\end{aligned}$$

  where $\kappa_k := 1$ if $k = 0$ and $\kappa_k := 0$ if $k \neq 0$.

- $b(w)$ is defined by

$$\begin{aligned}b_{i,j}(w) &:= \frac{\alpha_{ij,1}^2}{\Delta t}v_h(t+\Delta t, x_i, y_j) + \alpha_{ij,1}^2\frac{\mu(t,y_j)}{h}\kappa_{j-1}g(x_i) \\ &\quad + \frac{1}{2}\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^{\alpha_{ij}}\kappa_{j-\xi_2}g(x_{i-\xi_1})\end{aligned} \tag{4.4}$$

  where $v_h(t+\Delta t, x, y)$ is the solution at the previous time step and is assumed to be bounded.

We shall also denote

$$\delta_{ij}(w) := \frac{\alpha_{ij,1}^2}{\Delta t} + \frac{\alpha_{ij,1}^2}{h}\mu(t,y_j)\kappa_{j-1} + \frac{1}{2}\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^{\alpha_{ij}}\kappa_{j-\xi_2}.$$

Remark 4.1. *The matrix $A(w)$ is $\delta(w)$-diagonal dominant in the following sense:*

$$A_{(i,j),(i,j)}(w) = \delta_{ij}(w) + \sum_{(k,\ell)\neq(i,j)}|A_{(i,j),(k,\ell)}(w)|$$

Remark 4.2. *In the case no border points $y = 0$ are involved (i.e. when $j > p_{max}$), we have the more simple expressions:*

$$\begin{aligned}(A(w)X)_{ij} &:= \frac{\alpha_{ij,1}^2}{\Delta t}X_{ij} + \frac{\alpha_{ij,1}^2}{h}\mu(t,y_j)(X_{ij} - X_{i,j-1}) \\ &\quad + \frac{1}{2}\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^{\alpha_{ij}}(-X_{i-\xi_1,j-\xi_2} + 2X_{ij} - X_{i+\xi_1,j+\xi_2}).\end{aligned}$$

and

$$b_{i,j}(w) := \frac{\alpha_{ij,1}^2}{\Delta t}v_h(t+\Delta t, x_i, y_j), \qquad \delta_{ij}(w) := \frac{\alpha_{ij,1}^2}{\Delta t}.$$

Remark 4.3. *On the boundary $x = 0$, if we assume that $v_h(t+\Delta t, 0, y) = g(0)$ then the scheme reads*

$$\begin{aligned}\min_{\alpha_1^2+\alpha_2^2=1}\Bigg\{&\alpha_1^2\frac{v_h(t,0,y) - g(0)}{\Delta t} + \alpha_1^2\mu(t,y)\,\frac{v_h(t,0,y) - v_h(t,0,y-h)}{h} \\ &+ \frac{1}{2}\alpha_2^2(-v_h(t,0,y-h) + 2v_h(t,0,y) - v_h(t,0,y+h))\Bigg\} = 0, \quad \forall y \in h\mathbb{N}^*\end{aligned} \tag{4.5}$$

and with $v_h(t,0,0) = g(0)$. One can show that $v_h(t,0,y) = const = g(0)$ is the only bounded solution of (4.5) (using the results of Lemma A.1, Proposition A.4 and Proposition B.1). Hence by recursion we see that $v_h(t,0,y) = g(0)$ for all t and $y \in h\mathbb{N}$. In order to simplify the presentation of $A(w)$ and $b(w)$ we have preferred not to add this knowledge in a boundary condition at $x = 0$.

Preliminary results. In order to find a solution of (4.3), we first consider the linear system

$$A(w)X = b(w),$$

for a given $w \in \mathcal{A}$. For clarity, some specific results for such systems have been postponed to Appendix A (this requires some work since the systems are of infinite dimension). We can check that $(A(w), b(w))$ satisfy all the assumptions of Proposition A.6. In particular, we obtain that $A(w)$ is a *monotone matrix*, in the sense that if $X = (X_{i,j})_{i\geq 0, j\geq 1}$ is bounded (or bounded from below) and such that

$$\forall i \geq 0,\ \forall j \geq 1, \quad \delta_{ij}(w) = 0 \ \Rightarrow (A(w)X)_{ij} = 0, \tag{4.6}$$

then

$$A(w)X \geq 0 \Rightarrow \ X \geq 0.$$

Here (4.6) is equivalent to

$$\forall i \geq 0,\ j \geq 1, \quad \alpha_{ij,1} = 0 \ \Rightarrow -X_{i,j-1} + 2X_{ij} - X_{i,j+1} = 0.$$

Since $b(w)$ satisfies $\delta_{ij}(w) = 0 \Rightarrow b_{i,j}(w) = 0$, and that

$$\max_{i,j;\delta_{ij}(w)>0} \frac{|b_{ij}(w)|}{\delta_{ij}(w)} \leq \max(||v_h(t+\Delta t,.,.)||_\infty, ||g||_\infty),$$

we also obtain by Proposition A.6 (ii) that there exists a unique bounded X such that $A(w)X = b(w)$, and satisfying furthermore

$$||X||_\infty := \max_{i\geq 0, j\geq 1} |X_{ij}| \leq \max(||v_h(t+\Delta t,.,.)||_\infty,\ ||g||_\infty).$$

Howard's algorithm We can now consider the following Howard's algorithm [17] for solving (4.3).

Let $w^0 \in \mathcal{A}$ be a given initial control value

Iterate for $k \geq 0$

- Find X^k bounded, such that $A(w^k)X^k = b(w^k)$.
- $w^{k+1} := \text{argmin}_{w\in\mathcal{A}}(A(w)X^k - b(w))$.

In the second step, the minimization is done component by component, since $(A(w)X^k - b(w))_{ij}$ depends only of the control α_{ij}; the minimum is also well defined since the control set S^1 for α_{ij} is compact.

For convergence proof of Howard's algorithm in a general setting see [6] and references therein. In our case we have the following result, whose proof is postponed to Appendix B.

Proposition 4.4. *There exists a unique bounded solution X to the problem*

$$\min_{w\in\mathcal{A}}(A(w)X - b(w)) = 0,$$

and the sequence X^k converges pointwisely towards X, i.e., $\lim_{k\to\infty} X^k_{ij} = X_{ij}$ $\forall i,j \geq 0$.

Proof of the existence of v_h and of stability property. First, the convergence of Howard's algorithm leads to the existence of a scheme solution. Also we obtain the bound $||v_h(t,.)||_\infty = ||X||_\infty \leq \max(||v_h(t+\Delta t,.)||_\infty, ||g||_\infty)$. Hence by recursion we obtain $||v_h(t,.)||_\infty \leq ||g||_\infty$. This shows the stability property, and complete the proof of Proposition 3.7(ii).

Remark 4.5. *The stability and monotonicity results are obtained inconditionnaly with respect to the mesh sizes $h > 0$ and $\Delta t > 0$.*

We have also the following stronger monotonicity result, and that will be useful for the next Section.

Proposition 4.6. *if $v_h^1(t+\Delta t)$ and $v_h^1(t+\Delta t)$ are two bounded vectors defined on the grid, and X^1 and X^2 denotes the two corresponding solutions of (4.3), then*

$$v_h^1(t+\Delta t,.) \leq v_h^2(t+\Delta t,.) \quad \Rightarrow \quad X^1 \leq X^2.$$

Proof. Let us denote $b^q(w)$, for $q = 1,2$, the vectors corresponding to $v_h^q(t+\Delta t)$ as defined in (4.5). We remark that $b^1(w) \leq b^2(w)$, $\forall w \in \mathcal{A}$. Let w^1 be an optimal control for X^1. Then

$$\begin{aligned} A(w^1)X^1 - b^1(w^1) &= 0 = \min_{w\in\mathcal{A}}(A(w)X^2 - b^2(w)) \\ &\leq A(w^1)X^2 - b^2(w^1) \\ &\leq A(w^1)X^2 - b^1(w^1), \end{aligned}$$

and thus $A(w^1)(X^2 - X^1) \geq 0$. By the monotonicity property of $A(w^1)$ and the fact that if $\delta_{ij}(w^1) = 0$ then $b^2_{ij}(w^1) - b^1_{ij}(w^1) = 0$, we conclude to $X^1 \leq X^2$. □

5 Convergence

Since the scheme is monotone, stable and consistent, the idea is to use the same arguments as in [5, Theorem 2.1] to conclude the convergence of ϑ_h toward ϑ, taking into account the comparison principle Theorem 2.9.

We first establish the following discrete comparison principle for the scheme.

Lemma 5.1. *Let $Y = Y_{h,\Delta t}(t,x,y)$ be defined on $(x,y) \in G_h$ and for $t \in T - \Delta t\mathbb{N}$. Suppose that Y is a supersolution of the scheme (resp. subsolution of the scheme), in the following sense:*
(i) $\forall t + \Delta t \leq T$, $\forall (x,y) \in G_h$, $\mathcal{S}^\rho(t,x,y,Y(t,x,y),Y) \geq 0$ (resp. ≤ 0),
(ii) $\forall (x,y) \in G_h$, $Y(T,x,y) \geq g(x)$ (resp $Y(T,x,y) \leq g(x)$),
(iii) $\forall t \leq T$, $(x,y) \in G_h$, $Y(t,x,0) \geq g(x)$ (resp $Y(t,x,0) \leq g(x)$),
(iv) $Y(t,x,y)$ is bounded from below (resp. from above).
Then $Y \geq v_h$ (resp $Y \leq v_h$), where $v_h = v_h(t,x,y)$ are the scheme values.

Proof. The proof can be obtained by recursion (using $Y(t+\Delta t,.) \geq v_h(t+\Delta t,.)$ to show that $Y(t,.) \geq v_h(t,.)$) following the same arguments as in Proposition 4.6. In order to conclude from $A(w_1)(Y(t,.) - v_h(t,.)) \geq 0$ to $Y(t,.) - v_h(t,.) \geq 0$ (for a given control w_1), we use the fact that $Y(t,.) - v_h(t,.)$ is bounded from below and Proposition A.6 1). The proof for the subsolution is similar. □

We can now give the main convergence result.

Theorem 5.2. *We assume (A1)-(A3) and (A5). Suppose that $p_{max} = o(\frac{1}{h})$ as $h \to 0$. Then the scheme converges locally uniformly to ϑ when $(\Delta t, h) \to 0$.*

Proof. Let $\bar{v}$ and $\underline{v}$ be defined by

$$\overline{v}(t,x,y) := \limsup_{h,\Delta t \to 0,(t',x',y') \to (t,x,y)} v_h(t',x',y'),$$
$$\underline{v}(t,x,y) := \liminf_{h,\Delta t \to 0,(t',x',y') \to (t,x,y)} v_h(t',x',y')$$

The function $v_h(t,x,y)$ defined for (x,y) in the grid G_h and for $t = T - n\Delta t$ can be extended to $[0,T] \times \mathbb{R}^+ \times \mathbb{R}^+$ by a P1 interpolation in time. As in [5, Theorem 2.1], using properties $(i) - (iii)$ obtained in Proposition 3.7, we can prove that $\bar{v}$ and $\underline{v}$ are respectively bounded viscosity subsolution and supersolution of (3.1). Furthermore, if the following inequalities hold:

$$\overline{v}(T,x,y) \leq g(x) \leq \underline{v}(T,x,y) \tag{5.1}$$
$$\overline{v}(t,x,0) \leq g(x) \leq \underline{v}(t,x,0) \tag{5.2}$$

then, by the comparison principle of Theorem 2.9, we obtain $\overline{v} \leq \underline{v}$. Hence $\overline{v} = \underline{v}$ and the convergence of v_h towards the unique viscosity solution ϑ of (3.1).

Step 1: $\underline{v}(T,x,y) \geq g(x)$, and $\underline{v}(t,x,0) \geq g(x)$.
Considering $Y(t,x,y) := g(x)$, we see that Y is a subsolution of the scheme (in the sense of Lemma 5.1). Hence $v_h \geq Y$ and we deduce the two inequalities $\underline{v}(T,x,y) \geq g(x)$ and $\underline{v}(t,x,0) \geq g(x)$.

Step 2: $\bar{v}(T,x,y) \leq g(x)$, and $\bar{v}(t,x,0) \leq g(x)$.
Let $B \geq 0$ and $L \geq 0$ be constants such that $-g''(x) \geq -B$ for all $x \in [0,2]$, $-x^2 g''(x) \geq -B$ for all $x \geq 0$ and $\sigma^2(t,y) \leq Ly$, for all $y \geq 0$. Let

$$Y(t,x,y) := K\,(T-t)y + g(x), \quad \text{with } K := 2BL.$$

We consider a given point $(t,x,y) \in (T - \Delta t\mathbb{N}) \times G_h$, and a minimiser $\alpha = (\alpha_1, \alpha_2) \in S^1$ associated to $\mathcal{S}^\rho$ as in (3.9).

Let us write the decomposition of $a_p := \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha \xi \xi^T$. Then we have

$$\mathcal{S}^\rho(t,x,y,Y(t,x,y),Y) \;=\; -\alpha_1^2 \frac{\partial Y}{\partial t} + \alpha_1^2 \mu \frac{\partial Y}{\partial y} - \frac{1}{2} \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha \Delta_\xi^h Y$$

Since $g \in C^2$, we notice that $(\Delta_\xi^h Y)(t,x,y) = (\Delta_\xi^h g)(x) = g''(x_\xi)\xi_1^2$ for some $x_\xi \in (x - |\xi_1|h, x + |\xi_1|h) \subset [x - ph, x + ph]$, and thus

$$\mathcal{S}^\rho(t,x,y,Y(t,x,y),Y) \;\geq\; \alpha_1^2 K y - \frac{1}{2} \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^\alpha \xi_1^2 g''(x_\xi).$$

Let $h > 0$ be such that $h \leq \frac{1}{2p_{\max}}$. In the case $x \in [0,1]$ we have $ph \leq p_{\max}h \leq \frac{1}{2}$ and thus $x_\xi \in [0,2]$, and $-g''(x_\xi) \geq -B$. Hence, for all $x \in [0,1]$,

$$\begin{aligned}\mathcal{S}^\rho(t,x,y,Y(t,x,y),Y) &\geq \alpha_1^2 Ky - \frac{1}{2}B\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^\alpha \xi_1^2 \\ &\geq \alpha_1^2 Ky - \frac{1}{2}\alpha_1^2 BLy \\ &\geq 0,\end{aligned}$$

where we have used the definition of K and the fact that

$$\begin{aligned}\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^\alpha\xi_1^2 &= (a_p)_{11} = a_{11} = \alpha_1^2 x^2\sigma^2(t,y) \\ &\leq \alpha_1^2 Ly.\end{aligned} \tag{5.3}$$

Now in the case $x \geq 1$, we have $x \leq x_\xi + p_{\max}h \leq x_\xi + \frac{1}{2}$, with $x_\xi \geq \frac{1}{2}$, and thus $-x^2 g''(x_\xi) = -\frac{x^2}{x_\xi^2}x_\xi^2 g''(x_\xi) \geq -4B$. We obtain

$$\begin{aligned}\mathcal{S}^\rho(t,x,y,Y(t,x,y),Y) &\geq \alpha_1^2 Ky - 2B\frac{1}{x^2}\sum_{\xi\in\mathcal{Q}_p}\gamma_\xi^\alpha\xi_1^2 \\ &\geq \alpha_1^2 Ky - \alpha_1^2 2BLy \\ &\geq 0.\end{aligned}$$

Hence, Y satisfies the assumptions (i)-(iv) of Lemma 5.1, and thus $Y \geq v_h$. In particular,

$$\overline{v}(T,x,y) = \limsup_{h\to 0,(t',x',y')\to(T,x,y)} v_h(t',x',y') \leq \limsup_{h\to 0,(t',x',y')\to(T,x,y)} Y(t,x,y) = g(x).$$

We obtain $\overline{v}(t,x,0) \leq g(x)$ in the same way. □

Remark 5.3. *The choice of the projection a_p of a is made such that the identity (5.3) be true. This approach do not work with the orthogonal projection of a on $\mathcal{C}(\mathcal{Q}_p)$.*

6 Numerical results

We consider the approximation scheme of Section 3, hereafter refered as the Implicit Euler scheme, or (IE) scheme, and test it on some numerical examples. In all cases we have taken

$$\mu(t,y) = 0,$$

i.e. no transport term, because this is not the main difficulty of the equation. Also we fixed

$$\sigma(t,y) = \sqrt{y}. \tag{6.1}$$

We choose $\eta(t,x,y) := y$ in (2.11), altough this do not affect much numerical results.

All tests where done in Scilab (equivalent of Matlab), on a Pentium 4, 3Ghz computer.

6.1 Consistency test

Here we perform a verification of the consistency error of the spatial discretization. We consider the function

$$v(t,x,y) := 1 - e^{-x^2-y^2} + (T-t)^2, \tag{6.2}$$

and define f such that

$$f(t,x,y) := \tag{6.3}$$
$$\inf_{\alpha_1^2+\alpha_2^2=1} \left\{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}^T \begin{pmatrix} -\frac{\partial v}{\partial t} - \frac{1}{2}\sigma^2(t,y)x^2\frac{\partial^2 v}{\partial x^2} & -\frac{1}{2}\sigma(t,y)xy\frac{\partial^2 v}{\partial x \partial y} \\ -\frac{1}{2}\sigma(t,y)xy\frac{\partial^2 v}{\partial x \partial y} & -\frac{1}{2}y^2\frac{\partial^2 v}{\partial y^2} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right\}.$$

The function f corresponds to the first member of (2.11) with $\vartheta = v$ and $\eta(t,x,y) = y$. An exact computation gives

$$f(t,x,y) = \frac{1}{2}\Bigg[-\frac{\partial v}{\partial t} - \frac{1}{2}\sigma^2(t,x)x^2\frac{\partial^2 v}{\partial x^2} - \frac{1}{2}y^2\frac{\partial^2 v}{\partial y^2} \tag{6.4}$$
$$+\sqrt{\left(\frac{\partial v}{\partial t} + \frac{1}{2}\sigma^2(t,x)x^2\frac{\partial^2 v}{\partial x^2} - \frac{1}{2}y^2\frac{\partial^2 v}{\partial y^2}\right)^2 + \left(\sigma(t,x)xy\frac{\partial^2 v}{\partial x \partial y}\right)^2}\Bigg].$$

Following the definition of $\mathcal{S}^\rho$ in (3.10), and using the fact that $\mu = 0$, we define here $\mathcal{S}^{\rho,N_u}$ such that

$$\mathcal{S}^{\rho,N_u}(t,x,y,v(t,x,y),v) =$$
$$\min_{k=1,\dots,N_u} \left\{ \alpha_{1,k}^2 \frac{v(t,x,y) - v(t+\Delta t,x,y)}{\Delta t} - \frac{1}{2}\sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^{\alpha_k} \Delta_\xi v(t,x,y) \right\},$$

where $\alpha_k = (\alpha_{1,k}, \alpha_{2,k}) := e^{2i\pi k/(2N_u)}$ (we remark that it is sufficient to take half of the unit circle for the controls α in the definition of $\mathcal{S}^\rho$, and we do the same for $\mathcal{S}^{\rho,N_u}$).

Then we compute at time $t = T$, the value of

$$\mathcal{S}^{\rho,N_u}(t,x,y,v(t,x,y),v) - f(t,x,y). \tag{6.5}$$

The results are shown in Table 1, in L^∞ and L^2 norms. The space domain is $[0, x_{\max}] \times [0, y_{\max}]$ with $x_{\max} = y_{\max} = 3$, and we have used here Neumann boundary conditions on $x = x_{\max}$ and on $y = y_{\max}$ (following Proposition 2.10).

number of space steps	N_u	p_{max}	L^2 error	L^∞ error	CPU time (seconds)
20× 20	20	2	0.0215	0.0396	0.23
40× 40	40	3	0.0094	0.0212	1.72
80× 80	80	4	0.0046	0.0121	14.98
160× 160	160	6	0.0020	0.0058	156.09

Table 1: Consistency error

Remark 6.1. *Contrary to the the definition of the projection of the matrix a in section 3, we chose an orthogonal projection of a on $\mathcal{C}(\mathcal{S}_p)$, as in [7]. Even if we did not prove convergence in that case, it gives better numerical results (recall that $\|a - a^p\|_F \leq \frac{1}{4p_{\max}^2}\mathrm{Tr}(a)$ from [7]).*

Remark 6.2. *A key parameter for the discretization scheme is the maximum order $p_{\max}$ that we consider. From the expression of the theoretical consistency error and of Proposition 3.8, we can take $p_{\max}$ of the order of $h^{-\frac{2}{3}}$ In practice we observe that a small $p_{\max}$, as in Table 1, is numerically sufficient to obtain a consistency of order $O(h)$.*

We obtain a consistency error that converges to zero with rate h in both L^∞ and L^2 norms. To this end we also found numerically that it was sufficient to increase the number of controls as the number of space steps (as is done in Table 1).

On the other hand we have also observed that the consistency error behaves as $O(\frac{1}{N_u})$. For large N_u, and fixed space steps, the error no more diminishes, because the spatial error dominates (see Table 2).

N_u	L^2 error	L^∞ error
5	0.035	0.051
10	0.014	0.024
20	0.010	0.022
40	0.009	0.021
80	0.009	0.021

Table 2: Error with varying number of controls N_u. Space steps 40×40 here.

6.2 Convergence test

Now we consider the time-dependant equation (2.11), with unkown ϑ and with a second member f defined by (6.4) and (6.2), and with terminal data $\vartheta(T,.,.) = v(T,.,.)$. In this case we know that the value of the solution is $\vartheta = v$.

The results are given in Table 3, where we test the Implicit Euler scheme and also the Crank-Nicolson (CN) scheme (see Remark 6.3) that is second order in time [14]. We have used $T = 1$ with different time steps. We find that the (IE) scheme converges with rate $O(h) + O(\Delta t)$. The CN scheme gives better numerical results with a similar computational cost.

From Table 4 (varying time steps/space steps), we see that only few time steps are needed in order to obtain a good accuracy.

This shows the feasibility of the (EI) and (CN) scheme in reasonable time.

The CPU time could be reduced by using approximate sparse solvers for the linear systems involved in Howard's algorithm.

Remark 6.3. *The Crank-Nicolson scheme is defined here by the following implicit scheme:*

$$0 = \min_{k=1,\dots,N_u} \left\{ \alpha_{1,k}^2 \frac{v(t,x,y) - v(t+\Delta t,x,y)}{\Delta t} \right.$$
$$\left. - \frac{1}{2} \sum_{\xi \in \mathcal{Q}_p} \gamma_\xi^{\alpha_k} \frac{\Delta_\xi v(t,x,y) + \Delta_\xi v(t+\Delta t,x,y)}{2} \right\}$$

number of space steps	N_u	N	$p_{\max}$	L^2 error (EI)	L^∞ error (EI)	CPU time (seconds)	L^2 error (CN)	L^∞ error (CN)
20×20	20	20	2	0.0590	0.0822	98	0.0136	0.0333
40×40	40	40	3	0.0284	0.0367	946	0.0051	0.0117
80×80	80	80	4	0.0138	0.0178	10120	0.0023	0.0053

Table 3: Error for the Implicit Euler scheme and the Crank-Nicholson Scheme.

varying time steps

number of space steps	number of time steps	L^2 error	L^∞ error
80×80	5	0.0026	0.0062
80×80	10	0.0023	0.0052
80×80	80	0.0023	0.0053

varying space steps

number of space steps	number of time steps	L^2 error	L^∞ error
20×20	80	0.0136	0.0332
40×40	80	0.0051	0.0118
80×80	80	0.0023	0.0053

Table 4: Error with varying number of time steps (resp. space steps) for the Crank Nicholson Scheme. Only a few time steps are needed in order to obtain a good accuracy.

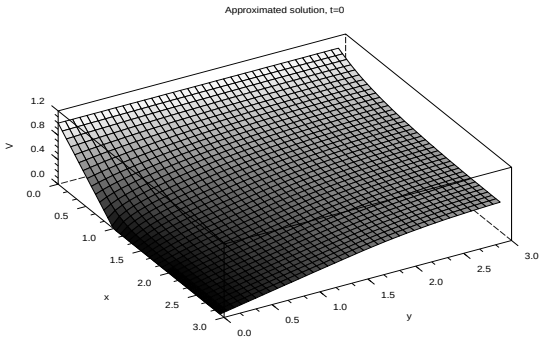

Figure 3: Surreplication price at time $t = 0$, with $T = 1$, $K = 1$ and payoff $(K - x)_+$.

(in our test, $\gamma_\xi^{\alpha_k}$ does not depend on time).

6.3 Application

We apply the method to a financial example: we compute the price of a put option of strike $K = 1$ and maturity $T = 1$. In this model, X represents the price of the underlying of the put, and Y represents the price of the forward variance swap on the underlying X. Therefore, the terminal condition is

$$v(T, x, y) = g(x) := (K - x)_+ .$$

(Note that even if g does not satisfy $(A5)$, it can be uniformly approximated by some g_ε which satisfies, for any $\varepsilon > 0$, $(A5)$ and $||g - g_\varepsilon||_{L^\infty(\mathbb{R}_+)} \leq \varepsilon$. In this way a convergence result can still be obtained).

Numerically we compute the price of the option for larges values of Y (i.e., $Y \simeq 3$), in order to use Neumann conditions for large Y. This approach is coherent with the value of interest which are typically for Y lower (or of the order of) unity. The result is shown in Fig. 3

Acknowledgments. The authors would like to thank H. Pham for useful comments.

A Properties of some infinite linear system

In this section we give some basic results for solving some specific infinite linear system that are involved in our scheme.

Notations. We say that $A = (a_{ij})_{1 \leq i,j}$, $i, j \in \mathbb{N}^*$, with $a_{ij} \in \mathbb{R}$ is an *infinite* matrix if $\{j \geq 1,\ a_{ij} \neq 0\}$ is finite $\forall i \geq 1$. If $X = (x_i)_{i \geq 1}$ then we denote $(AX)_i = \sum_{j \geq 1} a_{ij} x_j$. We also denote $X \geq 0$ if $x_i \geq 0$, $\forall i \geq 1$.

The following Lemma generalizes the monotony property of M-matrices.

Lemma A.1 (monotony)**.** *Let $A = (a_{ij})_{1 \leq i,j}$ be a real infinite matrix such that*
(i) For all $i \geq 1$, $\exists \delta_i \geq 0$, $a_{ii} = \delta_i + \sum_{j \neq i} |a_{ij}|$,
(ii) $a_{ij} \leq 0$ $\forall i \neq j$,

(iii) $\delta_1 > 0$,
(iv) $\forall i \geq 1$, $\sum_j j\ a_{ij} \geq 0$.
(v) $\forall i \geq 2$, *if* $\delta_i = 0$ *then* $\exists q_i > 0$ *such that* $(AX)_i = q_i(-x_{i-1} + 2x_i - x_{i+1})$.
Then A *is monotone in the following sense: if* $X = (x_i)_{i\geq 1}$ *is bounded from below and such that* $\forall i \geq 1$, $\delta_i = 0 \Rightarrow (AX)_i = 0$, *then*

$$AX \geq 0 \Rightarrow X \geq 0.$$

Remark A.2. *Note that from Lemma A.1 we deduce the uniqueness of bounded solutions of* $AX = b$ *for any* b *such that* $\delta_i = 0 \Rightarrow b_i = 0$.

Proof of Lemma A.1. Let $m = \min_{i\geq 1} x_i$.

Step 1. We first assume that there exists $i \geq 1$ such that $m = x_i$. Then

$$0 \leq a_{ii}x_i + \sum_{j\neq i} a_{ij}x_j = \delta_i x_i + \sum_{j\neq i} |a_{ij}|(x_i - x_j) \leq \delta_i x_i$$

If $\delta_i > 0$, then $x_i \geq 0$. In the case $\delta_i = 0$, by assumption (v) we obtain that $m = x_i = x_{i-1} = x_{i+1}$. In particular the minimum m is also reached by x_{i-1}. Since $\delta_1 > 0$, by a recursion argument we will arrive at a point j such that $\delta_j > 0$ and thus $x_j \geq 0$.

Step 2. In the general case we consider $Y = (y_i)$ with $y_i := x_i + \varepsilon\ i$ for some $\varepsilon > 0$. We note that $y_i \to +\infty$, hence $i \to y_i$ has a minimum. Also, $(AY)_i = (AX)_i + \varepsilon \sum_j j\ a_{ij} \geq 0$. Hence $AY \geq 0$ and $Y \geq 0$ by Step 1. Since this is true for any $\varepsilon > 0$, we conclude that $X \geq 0$. □

Remark A.3. *Note that in Lemma A.1 we can relax the assumption* (x_i) *bounded from below by* $\liminf_{i\to\infty} \frac{x_i}{i} \geq 0$.

Proposition A.4 (Existence of solutions for linear systems)**.** *We consider* A, *an infinite matrix, such that*
(i) $\forall i \geq 1$, $\exists \delta_i \geq 0$, $a_{ii} = \delta_i + \sum_{j\neq i} |a_{ij}|$,
(ii) $\delta_1 > 0$.
(iii) $\forall i \geq 2$, *if* $\delta_i = 0$ *then* $\exists q_i > 0$, *such that* $(AX)_i = q_i(-x_{i-1} + 2x_i - x_{i+1})$.
Let also $b = (b_i)_{i\geq 1}$ *be such that*

$$\forall i, \delta_i = 0 \Rightarrow b_i = 0, \qquad \text{and} \qquad \max_{k\geq 1,\ \delta_k \neq 0} \frac{|b_k|}{\delta_k} < \infty.$$

Then there exists a unique X, *in the space of bounded sequences, such that* $AX = b$, *and furthermore we have*

$$\max_{k\geq 1} |x_k| \leq \max_{k\geq 1,\ \delta_k\neq 0} \frac{|b_k|}{\delta_k}.$$

Proof. We look for solutions $x^{(n)} = (x_1^{(n)}, \ldots, x_n^{(n)})^T \in \mathbb{R}^n$ of the first n linear equations of $AX = b$, and set also $x_k^{(n)} := 0$, $\forall k > n$. (Dirichlet type boundary conditions on the right border). This leads to solve the finite dimensional system

$$A^{(n)}x^{(n)} = b^{(n)} \tag{A.1}$$

where $A^{(n)} := (a_{ij})_{1\leq i,j\leq n}$ and $b^{(n)} := (b_1, \ldots, b_n)^T$,

Lemma A.5. *There exists a unique $x^{(n)}$ solution of (A.1) and furthermore it satisfies the inequality*

$$\max_{1\le k\le n} |x_k^{(n)}| \le \max_{1\le k\le n,\ \delta_k\neq 0} \frac{|b_k|}{\delta_k}. \tag{A.2}$$

Proof of Lemma A.5. Suppose that $x^{(n)}$ exists, and let i be such that $|x_i^{(n)}| = \max_{1\le j\le n} |x_j^{(n)}|$. Note that we still have $\forall 1 \le i \le n$, $a_{ii}^{(n)} = \delta_i + \sum_{j\neq i} |a_{ij}^{(n)}|$. If $\delta_i > 0$,

$$|b_i| \ge |a_{ii}^{(n)} x_i^{(n)}| - \sum_{j\neq i} |a_{ij}^{(n)}||x_j^{(n)}| \ge \delta_i |x_i^{(n)}|$$

thus $|x_i^{(n)}| \le \frac{|b_i|}{\delta_i}$. If $\delta_i = 0$, we consider

$$i_0 := \sup\{k < i,\ \delta_k > 0\}.$$

(i_0 exists since $\delta_1 > 0$). Then $-x_{k-1}^{(n)} + 2x_k^{(n)} - x_{k+1}^{(n)} = b_k/q_k = 0$ for $k = i_0+1, \dots, i$, and $x_{k+1}^{(n)} - x_k^{(n)} = const = c_0$ for $k = i_0, \dots, i$. But $x_i^{(n)}$ is an extremum of $x_{i-1}^{(n)}$, $x_i^{(n)}$ and $x_{i+1}^{(n)}$. This implies that $x_{i-1}^{(n)} = x_i^{(n)} = x_{i+1}^{(n)}$, and thus $c_0 = 0$ and $x_{i_0}^{(n)} = x_i^{(n)}$ is also an extremum. Since $\delta_{i_0} > 0$, we can estimate $|x_{i_0}^{(n)}|$ as before. This implies the invertibility of $A^{(n)}$, and thus the uniqueness of $x^{(n)}$. □

Now we shall prove that the sequence $X^{(n)} = (x^{(n)}, 0, 0, \dots)^T$, which satisfies already $||X^{(n)}||_\infty \le C := \max_{\delta_k\neq 0} \frac{|b_k|}{\delta_k}$, converges pointwisely towards a solution X of the problem. We first suppose that $b \ge 0$. We can see that $A^{(n)}$ is still a monotone matrix (following the proof of Lemma A.1). Hence $x^{(n)} \ge 0$. Now we consider $x^{(n+1)}$ and for $i \le n$ we see that

$$\left(A^{(n)} x^{(n+1)}\right)_i = b_i - a_{i,n+1} x_{n+1}^{(n+1)} \ge b_i = \left(A^{(n)} x^{(n)}\right)_i.$$

Hence we obtain that

$$(x_1^{(n+1)}, \dots, x_n^{(n+1)})^T \ge (x_1^{(n)}, \dots, x_n^{(n)})^T,$$

and in particular $X^{(n)} \le X^{(n+1)}$. Since $||X||_\infty \le C$, we obtain the (pointwise) convergence of $X^{(n)}$ towards some vector X such that $||X||_\infty \le C$. In the general case, we can decompose $b = b^+ - b^-$ with $b^+ = \max(b, 0)$, $b^- = \max(-b, 0)$, and proceed in the same way. We obtain the pointwise convergence of $X^{(n)} = X^{(n),+} - X^{(n),-}$ towards some X, with $X^{(n),\pm} \ge 0$ and $||X^{(n),\pm}||_\infty \le C$, hence also $||X||_\infty \le C$.

Since $\{j,\ a_{ij}^{(n)} \neq 0\}$ is finite, for any given i we can pass to the limit $n \to \infty$ in $\sum_{j\ge 1} a_{ij}^{(n)} x_j^{(n)} = b_i$, and obtain $(AX)_i = b_i$. □

Case of infinite 2d matrices. We say that the set of real numbers $A = (A_{(i,j),(k,\ell)})_{1\le i,j,k,\ell}$ is an *infinite 2d* matrix if $\{(k,\ell),\ A_{(i,j),(k,l)} \neq 0\}$ is finite $\forall i, j \ge 1$ (A is also an "infinite" tensor). If $X = (X_{i,j})_{i,j\ge 1}$ then we denote $(AX)_{i,j} = \sum_{k,\ell\ge 1} A_{(i,j),(k,\ell)} X_{k,\ell}$. We also denote $X \ge 0$ if $X_{i,j} \ge 0$, $\forall i, j$.

The previous results can be easily generalized to infinite 2d matrices. We state here the results without proof.

Proposition A.6. *Let $A = (A_{(i,j),(k,\ell)})_{1\leq i,j,k,\ell}$ be an infinite 2d matrix such that*
(i) For all $i,j \geq 1$, $A_{(i,j),(i,j)} = \delta_{ij} + \sum_{(k,\ell)\neq(i,j)} |A_{(i,j),(k,\ell)}|$ with $\delta_{ij} \geq 0$,
(ii) $A_{(i,j),(k,\ell)} \leq 0$ $\forall (i,j) \neq (k,\ell)$,
(iii) $\delta_{i1} > 0$, $\forall i \geq 1$,
(iv) $\forall i,j \geq 1$, $\sum_{(k,\ell)}(k+\ell)A_{(i,j),(k,\ell)} \geq 0$,
(v) $\forall i \geq 1$, $\forall j \geq 2$, if $\delta_{ij} = 0$ then $\exists q_{ij} > 0$ such that

$$(AX)_{ij} = q_{ij}(-X_{i,j-1} + 2X_{i,j} - X_{i,j+1}).$$

1) Then A is monotone in the following sense: if $X = (X_{i,j})_{i,j\geq 1}$ is bounded from below and such that $\forall i,j \geq 1$, $\delta_{i,j} = 0 \Rightarrow (AX)_{i,j} = 0$, then

$$AX \geq 0 \Rightarrow X \geq 0.$$

2) If $b = (b_{ij})_{i,j\geq 1}$ is such that $\delta_{ij} = 0 \Rightarrow b_{i,j} = 0$, and $\max_{i,j\geq 1,\ \delta_{ij}>0} \frac{|b_{ij}|}{\delta_{ij}} < \infty$, then there is a unique bounded X such that $AX = b$, and furthermore

$$\max_{i,j\geq 1} |X_{ij}| \leq \max_{i,j\geq 1,\ \delta_{ij}>0} \frac{|b_{ij}|}{\delta_{ij}}.$$

B Convergence of the Howard algorithm

In this section we prove the following result.

Proposition B.1. *Let S be a compact set, and $\mathcal{A} := S^{\mathbb{N}}$, the set of infinite sequences of S. For all $w \in \mathcal{A}$, let $A(w) := (a_{ij}(w))_{i,j\geq 1}$ be an infinite matrix, and $b(w) := (b_i(w))_{i\geq 1}$. We assume furthermore that*
(i) If $w = (w_i)_{i\geq 1}$, $a_{ij}(w)$ depends only of w_i, and also $b_i(w)$ depends only of w_i, and this dependence is continuous.
(ii) $\forall i$, $\sup_{w\in\mathcal{A}} \big(Card\{j,\ a_{ij}(w) \neq 0\}\big) < \infty$.
(iii) (monotony) For all $w \in \mathcal{A}$ and X bounded,

$$A(w)X \geq 0 \quad \Rightarrow \quad X \geq 0.$$

(iv) $\exists C \geq 0$, $\forall w \in \mathcal{A}$, $\exists X$ solution of $A(w)X = b(w)$ and such that

$$||X||_\infty \leq C.$$

Then
(i) there exists a unique bounded solution X to the problem

$$\min_{w\in\mathcal{A}}(A(w)X - b(w)) = 0. \tag{B.1}$$

(ii) the Howard algorithm as defined in section 4 converges pointwisely towards X.

Remark B.2. *Proposition B.1 can then be adapted in order to prove Proposition 4.4. The proof is left to the reader.*

Proof. Let us first check the uniqueness. Let X and Y be two solutions, and let $\bar{w}$ be an optimal control associated to Y. Then

$$\begin{aligned} A(\bar{w})Y - b(\bar{w}) &= 0 \\ &= \min_{w\in\mathcal{A}}(A(w)X - b(w)) \\ &\leq A(\bar{w})X - b(\bar{w}). \end{aligned}$$

Hence $A(\bar{w})(Y-X) \leq 0$ and thus $Y \leq X$ using the monotony property. We can prove $Y \geq X$ in the same way, hence $X = Y$ which proves uniqueness.

The existence now is obtained by considering the sequence X^k and controls w^k as in the Howard algorithm of section 4.

We first remark that for all $k \geq 0$, $X^k \leq X^{k+1}$, because

$$\begin{aligned} A(w^{k+1})X^{k+1} - b(w^{k+1}) &= 0 \\ &= A(w^k)X^k - b(w^k) \\ &\geq \min_{w}(A(w)X^k - b(w)) \\ &\geq A(w^{k+1})X^k - b(w^{k+1}) \end{aligned}$$

and using the monotony of $A(w^{k+1})$. Also, X^k is bounded. Hence X^k converges pointwisely towards some bounded X. It remains to show that X satisfies (B.1).

Let $F_i(X)$ be the i-th component of $\min_{w\in\mathcal{A}}(A(w)X - b(w))$, i.e.

$$F_i(X) = \min_{w\in\mathcal{A}}(A(w)X - b(w))_i$$

For a given i, since $(A(w)X)_i$ involves only a finite number of matrix continuous coefficients $(a_{ij}(w))_{j\leq j_{\max}}$, we obtain that $\lim_{k\to\infty} F_i(X^k) = F_i(X)$. Also by compactness of S, by a diagonal extraction argument, there exists a subsequence of $(w^k)_{k\geq 0}$, denoted w^{ϕ_k}, that converges pointwisely towards some $w \in \mathcal{A}$.

Passing to the limit in $(A(w^{\phi_k})X^{\phi_k} - b(w^{\phi_k}))_i = 0$, we obtain $(A(w)X - b(w))_i = 0$. On the other hand,

$$\begin{aligned} F_i(X) &= \lim_{k\to\infty} F_i(X^{\phi_k - 1}) \\ &= \lim_{k\to\infty} \left(A(w^{\phi_k})X^{\phi_k} - b(w^{\phi_k})\right)_i \\ &= \big(A(w)X - b(w)\big)_i \end{aligned}$$

Hence $F_i(X) = 0$, $\forall i$, which concludes the proof. □

References

[1] M. Akian, J.L. Menaldi, and A. Sulem. On an investement-consumption model with transaction costs. *SIAM J. Control Optim.*, 34:329–364, 1996.

[2] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.* Systems and Control: Foundations and Applications. Birkhäuser, Boston, 1997.

[3] G. Barles. An approach of deterministic control problems with unbounded data. *Ann. Inst. H. Poincaré, Anal. Non Linéaire*, 7(4):235–258, 1990.

[4] G. Barles. *Solutions de viscosité des équations de Hamilton-Jacobi*, volume 17 of *Mathématiques et Applications*. Springer, Paris, 1994.

[5] G. Barles and P.E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4:271–283, 1991.

[6] O. Bokanowski, S. Maroso, and H. Zidani. Some convergence results for Howard's algorithm. *Preprint Inria http://hal.inria.fr/inria-00179549/fr/*, 2007.

[7] J.F. Bonnans, E. Ottenwaelter, and H. Zidani. Numerical schemes for the two dimensional second-order HJB equation. *ENSAIM: M2AN*, 38:723–735, 2004.

[8] J.F. Bonnans and H. Zidani. Characterization of consistency of some numerical schemes for the stochastic HJB equation. In J.L. Menaldi E. Rofman and A. Sulem, editors, *Optimal Control and PDE - Innovations et Applications*. IOS Press, 2000.

[9] B. Bruder. Super-replication of european options with a derivative asset under constrained finite variation strategies. *Preprint HAL http://hal.archives-ouvertes.fr/hal-00012183/fr/*, 2005.

[10] J.P. Chancelier, B. Øksendal, and A. Sulem. Combined stochastic control and optimal stopping, and application to numerical approximation of combined stochastic and impulse control. *Tr. Mat. Inst. Steklova*, 237(Stokhast. Finans. Mat.):149–172, 2002.

[11] P. Cheridito, H.M. Soner, and N. Touzi. The multi-dimensional super-replication problem under gamma constraints. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 22(5):633–666, 2005.

[12] M. Crandall, H. Ishii, and P.L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc.*, 27:1–67, 2000.

[13] J. Cvitanic, H. Pham, and N. Touzi. Super-replication in stochastic volatility models with portfolio constraints. *Journal of Applied Probability*, 36:523–545, 1999.

[14] R. Dautray and J.L. Lions. *Mathematical analysis and numerical methods for science and technology. Volume 5 . Evolution Problems I.* Springer, 2000.

[15] N. El Karoui and M.C. Quenez. Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM J. Control Optim.*, 33(1):29–66, 1995.

[16] R.L. Graham, D.E. Knuth, and O. Patashnik. *Concrete Mathematics, A Foundation For Computer Science*. Addison-Wesley, Paris, 1994.

[17] R.A. Howard. *Dynamic Programming and Markov Process*. MIT, 1960.

[18] P.L. Lions. Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. Part 2: viscosity solutions and uniqueness. *Communications in partial differential equations*, 8:1229–1276, 1983.

[19] H. Pham. On some recent aspects of stochastic control and their applications. *Probability Surveys*, 2:1–549, 2005.

[20] H.M. Soner and N. Touzi. Super replication under gamma constraints. *SIAM Journal on Control and Optimization*, 39(1):73–96, 2000.

[21] H.M. Soner and N. Touzi. Stochastic target problems, dynamic programming and viscosity solutions. *SIAM Journal on Control and Optimization*, 41:404–424, 2002.