ESAIM: M2AN
Vol. 38, N° 4, 2004, pp. 723–735
DOI: 10.1051/m2an:2004034

ESAIM: Mathematical Modelling and Numerical Analysis

A FAST ALGORITHM FOR THE TWO DIMENSIONAL HJB EQUATION OF STOCHASTIC CONTROL

J. Frédéric Bonnans[1], Élisabeth Ottenwaelter[2] and Housnaa Zidani[3]

Abstract. This paper analyses the implementation of the generalized finite differences method for the HJB equation of stochastic control, introduced by two of the authors in [Bonnans and Zidani, *SIAM J. Numer. Anal.* **41** (2003) 1008–1021]. The computation of coefficients needs to solve at each point of the grid (and for each control) a linear programming problem. We show here that, for two dimensional problems, this linear programming problem can be solved in $O(p_{max})$ operations, where p_{max} is the size of the stencil. The method is based on a walk on the Stern-Brocot tree, and on the related filling of the set of positive semidefinite matrices of size two.

Mathematics Subject Classification. 49L99, 93E20.

Received: January 13, 2004. Revised: June 8, 2004.

1. Introduction

In this paper we discuss numerical schemes for the HJB equation of stochastic control. The model problem we are considering is

$$(P_{\tau,x}) \qquad \text{Min}\, I\!\!E \int_\tau^T \ell(t,y(t),u(t))\mathrm{d}t + \ell_F(y(T)); \quad \left\{ \begin{array}{l} \mathrm{d}y(t) = f(t,y(t),u(t))\mathrm{d}t + \sigma(t,y(t),u(t))\mathrm{d}w(t), \\ y(\tau) = x; \quad u(t) \in U, \;\; t \in [\tau,T], \;\; \tau \in [0,T]. \end{array} \right.$$

Here $T > 0$ is the (given) final time, $y(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ are the state and control variable, the latter subject to the constraint $u(t) \in U$ *a.e.*, where U is a compact subset of $\mathbb{R}^m$, $\ell : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\ell_F : \mathbb{R}^n \to \mathbb{R}$ are the running and final cost, $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is the drift (deterministic part of dynamics), σ is a mapping from $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$ into the space of $n \times r$ matrices, and w is a standard r dimensional Brownian motion. The control variable u has to be a function of past events, *i.e.*, is progressively measurable w.r.t. the filtration $\mathcal{F}_t$ associated with the Brownian motion. Let $\mathcal{U}$ be the set of feasible policies, *i.e.*, progressively measurable controls with values in U. We assume for the sake of simplicity that f, σ, ℓ and ℓ_F, are Lipschitz and bounded. Then (*e.g.* Fleming and Soner [7]) the stochastic differential equation is, for each policy $u \in \mathcal{U}$,

Keywords and phrases. Stochastic control, finite differences, viscosity solutions, consistency, HJB equation, Stern-Brocot tree.

[1] Projet Sydoco, Inria-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay, France.
e-mail: Frederic.Bonnans@inria.fr

[2] IUT de Paris and Projet Sydoco, Inria-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay, France.
e-mail: Elisabeth.Ottenwaelter@inria.fr

[3] Projet Sydoco, Inria-Rocquencourt and Unité de Mathématiques Appliquées, ENSTA, 32 Boulevard Victor, 75739 Paris Cedex 15, France. e-mail: zidani@ensta.fr

well posed and the corresponding expectation $W(\tau, x, u)$ is well-defined. Denote the transposition operator by $\top$. Let $a(t, x, u) := \frac{1}{2}\sigma(t, x, u)\sigma(t, x, u)^\top$, for all $(t, x, u) \in [0, T] \times \mathbb{R}^n \times U$, be the diffusion matrix. The value function V of problem $(P_{\tau,x})$, defined by $V(\tau, x) := \inf_u W(\tau, x, u)$, is (Lions [13]) the unique bounded viscosity solution of the Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} -v_t(t,x) &= \inf_{u\in U} \{\ell(t,x,u) + f(t,x,u)\cdot v_x(t,x) + a(t,x,u)\circ v_{xx}(t,x)\}\,, \\ &\qquad\qquad \text{for all } \ t,x \in [0,T]\times\mathbb{R}^n. \\ v(T,x) &= \ell_F(x), \text{ for all } \ x\in\mathbb{R}^n, \end{aligned} \tag{\textit{HJB}}$$

where v_{xx} denotes the $n \times n$ matrix of second derivatives of v with respect to x, and given two symmetric matrices A, B, of size n, $A \circ B := \sum_{i,j=1}^n A_{ij}B_{ij}$ is the scalar product associated with the Frobenius norm $\|A\| := (\sum_{i,j=1}^n A_{ij}^2)^{1/2}$ (since we do not use other norms on matrices the notation is non ambiguous). Various numerical methods have been proposed for solving this problem. Classical finite difference methods were discussed in Lions and Mercier [14], see also Menaldi [15]. Markov chain approximation were introduced in Kushner [11], see Kushner and Dupuis [12]. Camilli and Falcone [6] discuss methods based on *a priori* time discretization (and the related dynamic programming principle for discrete time problems). Krylov [10] gives an error estimate of a large class of discretization schemes. Recent improvements of the error estimates were obtained in Barles and Jakobsen [1, 2].

2. Generalized finite differences

Let us recall the generalized finite differences (GFD) method of [4] in the setting of finite horizon problems. The space discretization steps are positive real numbers $h_1, \ldots, h_n$. With a point $k \in \mathbb{Z}^n$ is associated $x_k := \sum_{i=1}^n k_i e_i$ of the state space, where e_i is the ith standard basis vector. Let $Q \in \mathbb{N}$, $Q > 1$ be the number of time steps; set $h_0 := T/Q$ and $t_q := qh_0$, for $q = 0, \ldots, Q$. Denote by v_k^q the approximation of the value function V at $(t, x) = (t_q, x_k)$.

Let $\varphi = \{\varphi_k\}$ be a real valued function over $\mathbb{Z}^n$. The upwind finite difference operator $D_{q,k}^u$ associated with $f(t_q, x_k, u)$ at point (t_q, x_k) is

$$\left(D_{q,k}^u \varphi_k\right)_i = \frac{\varphi_{k+e_i} - \varphi_k}{h_i} \quad \text{if} \quad f(t_q, x_k, u)_i \geq 0, \quad \frac{\varphi_k - \varphi_{k-e_i}}{h_i} \quad \text{if not.} \tag{1}$$

With $\xi \in \mathbb{Z}^n$, associate the second order finite difference operator

$$\Delta_\xi \varphi_k := \varphi_{k+\xi} + \varphi_{k-\xi} - 2\varphi_k = \varphi_{k+\xi} - \varphi_k - (\varphi_k - \varphi_{k-\xi}). \tag{2}$$

The (second-order) stencil is a finite set of $\mathbb{Z}^n \setminus \{0\}$ containing at least the closest points, *i.e.* the canonical basis of $\mathbb{R}^n$: $\{e_1, \cdots, e_n\}$. For each $k \in \mathbb{Z}^n$, we perform an approximation of the second-order term in the HJB equation by a linear combination of second order finite difference operators associated with elements of the stencil, *i.e.*, the expression $\sum_{\xi\in\mathcal{S}} \alpha_{q,k,\xi}^u \Delta_\xi v_k^q$ where $\alpha_{q,k,\xi}^u$ are to be set. Let $a^h := \{a_{ij}/h_i h_j\}$ denote the scaled diffusion matrix. Following [4] we say that the operator $\sum_{\xi\in\mathcal{S}} \alpha_{q,k,\xi}^u \Delta_\xi$ is a *strongly consistent* approximation of $a(t, x, u) \circ D_{xx}^2$ if

$$\sum_{\xi\in\mathcal{S}} \alpha_{q,k,\xi}^u \xi\xi^\top = a^h(t_q, x_k, u), \qquad \text{for all} \quad k \in \mathbb{Z}^n. \tag{3}$$

From the above, we deduce the following explicit (backwards) scheme

$$\begin{aligned} \frac{v_k^{q-1} - v_k^q}{h_0} &= \inf_{u \in U} \left\{ \ell(t_q, x_k, u) + f(t_q, x_k, u) \cdot D_{q,k}^u v_k^q + \sum_{\xi \in \mathcal{S}} \alpha_{q,k,\xi}^u \Delta_\xi v_k^q \right\} \\ v_k^Q &= \ell_F, \end{aligned} \tag{4}$$

for all $q = 1, \dots, Q$ and $k \in \mathbb{Z}^n$. The scheme is monotone (*i.e.*, v_k^{q-1} is a non decreasing function of v^q) if all coefficients of $v_{k'}^q$ in (4) appear with nonnegative coefficients. This holds if the coefficients $\alpha_{q,k,\xi}^u$ are nonnegative and, in addition,

$$\sum_{i=1}^n \frac{|f_i(t_q, x_k, u)|}{h_i} + 2 \sum_{\xi \in \mathcal{S}} \alpha_{q,k,\xi}^u \leq \frac{1}{h_0}, \ \ \forall (q, k, u) \in \{1, \dots, Q\} \times \mathbb{Z}^n \times U. \tag{5}$$

This last condition ensures the non decrease w.r.t. v_k^q. Since strong consistency implies $\sum_{\xi \in \mathcal{S}} \alpha_{q,k,\xi}^u \leq$ trace $a^h(t_q, x_k, u)$ by ([4], Lemma 2.1), condition (5) is satisfied whenever

$$\sum_{i=1}^n \frac{\|f_i\|_\infty}{h_i} + 2 \| \text{ trace } a^h \|_\infty \leq \frac{1}{h_0} \cdot \tag{6}$$

Consequently, when $\min_i h_i \downarrow 0$ we may take $h_0 = C \min_i \left(h_i^2\right)$, for $C > 0$ small enough (depending on f and a), as expected.

If the strong consistency and monotonicity properties holds, then GFD are a particular case of consistent chain Markov approximations, and therefore are convergent in view of Kushner and Dupuis ([12], Chap. 10). Since these schemes are monotone, stable and consistent, convergence of these schemes is also a consequence of Barles and Souganidis ([3], Th. 2.1). It is not difficult to see that this scheme satisfies the hypotheses of Krylov [10], Barles and Jacobsen [1, 2], and hence, the error estimates of these authors apply (for the corresponding adaptation to infinite horizon problems of GDF if necessary).

The interest of GFD is that it makes easier the analyzis of consistency properties. For instance, [4] provides characterizations of the class of diffusion matrices for which the scheme is consistent with the most common stencils, for dimensions $n = 2$ to 4. Since coefficients $\alpha_{q,k,\xi}^u$ have to be nonnegative and solution of (3), they are solution of a linear program (with zero cost); their computation may be expensive if the stencil is large. Remember that this has to be done at each point of the spatial grid, for each time step (and each control if diffusions depend on the control). Define the size of a stencil $\mathcal{S}$ as

$$\text{size}(\mathcal{S}) := \max\{\|\xi\|_\infty; \ \ \xi \in \mathcal{S}\}.$$

The main result of this paper is, for two dimensional problems, an algorithm for computing the coefficients in $O(\text{size}(\mathcal{S}))$ operations. More generally, for nonconsistent problems the algorithm computes the closest consistent matrix (in the Frobenius norm) in $O(\text{size}(\mathcal{S}))$ operations. In addition, it has a recursive property: the closest consistent matrix for stencil of size p_{max} is computed in $O(1)$ operations after having obtained the closest consistent matrix for stencil of size $p_{max} - 1$.

The main result is strongly related to some geometric properties of the set of PSD (symmetric, positive semidefinite) matrices on $\mathbb{R}^2$, that are the subject of the next section.

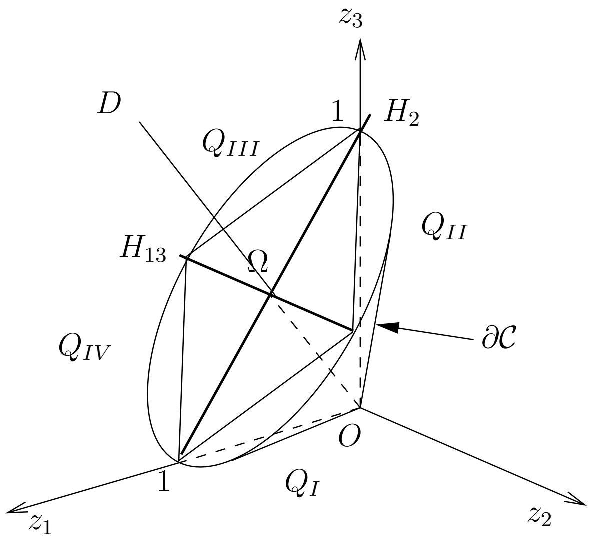

Figure 1. Cone of positive semidefinite matrices.

3. Structure of 2D diffusion matrices

We may view 2×2 symmetric matrices as elements of $\mathbb{R}^3$. The mapping

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \to (a_{11},\ \sqrt{2}a_{12},\ a_{22})^\top \tag{7}$$

is norm preserving from the space of 2×2 symmetric matrices, endowed with the Frobenius norm, onto the three dimensional Euclidean space. The image of the PSD cone by the mapping (7) is the set

$$\left\{ z \in \mathbb{R}^3; z_1 \geq 0;\ z_3 \geq 0;\ \tfrac{1}{2}(z_2)^2 \leq z_1 z_3 \right\}. \tag{8}$$

Since the set of PSD matrices is a cone, it is natural to represent directions of this cone by drawing their intersection with the hyperplane $z_1 + z_3 = 1$ (image of the set of matrices with unit trace), see Figure 1. In order to compute this intersection, consider the norm preserving mapping

$$w_1 = (z_1 - z_3)/\sqrt{2};\ w_2 = z_2;\ w_3 = (z_1 + z_3)/\sqrt{2}.$$

Taking $(w_1, w_2)/w_3$ as projective coordinates, which for matrix a means $(a_{11} - a_{22}, 2a_{12})/(a_{11} + a_{22})$, and calling the latter the *view* of matrix a, we see that the set of views of PSD matrices is the unit ball of $\mathbb{R}^2$ for the Euclidean norm. For example, the view of the identity, denoted as Ω, is the zero vector, and the view of $\eta\eta^\top$, where $\eta := (1\ 0)^\top$, is $(1\ 0)$. The lemma below eases the computation of the view of any rank one symmetric nonnegative matrix, and is illustrated in Figures 2 and 3.

Lemma 3.1. *Let $\eta = (\cos\theta, \sin\theta)$. Then the view of $\eta\eta^\top$ makes an angle of 2θ with the view of $(1,0)(1,0)^\top$.*

Proof. With η is associated $w = (\cos^2\theta - \sin^2\theta, 2\cos\theta\sin\theta) = (\cos 2\theta, \sin 2\theta)$. The result follows. □

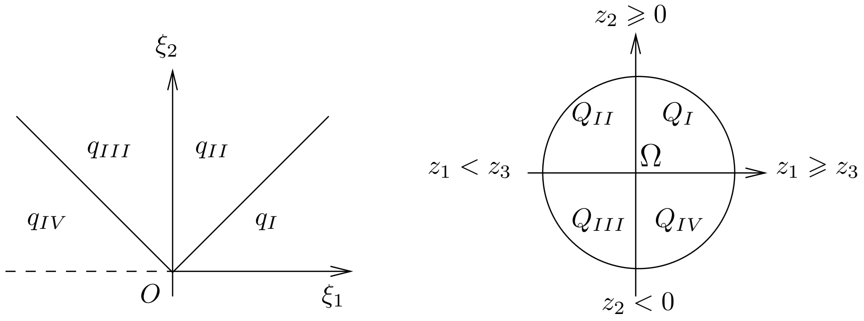

FIGURE 2. Correspondence of regions.

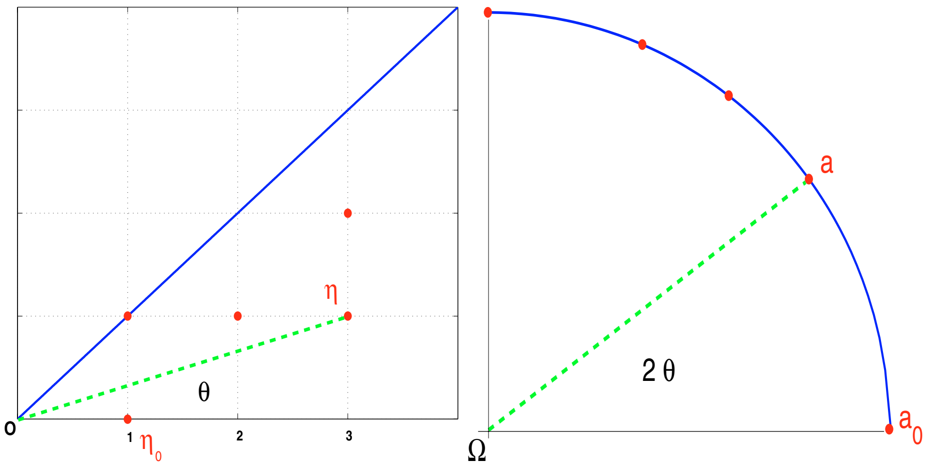

FIGURE 3. Correspondence of angles.

Since a matrix is diagonal dominant iff $|a_{11} - a_{22}| + 2|a_{12}| \leq a_{11} + a_{22}$, the view of such matrices is the unit ball in the norm ℓ^1 of $\mathbb{R}^2$. Diagonal dominant matrix have the well-known decomposition

$$a = (a_{11} - |a_{12}|)\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix} + (a_{22} - |a_{12}|)\begin{pmatrix}0\\1\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix} + \max(a_{12}, 0)\begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}1 & 1\end{pmatrix} + \max(-a_{12}, 0)\begin{pmatrix}-1\\1\end{pmatrix}\begin{pmatrix}-1 & 1\end{pmatrix}. \tag{9}$$

Let us call “inner region” of the PSD cone, the set of diagonal dominant matrices. There are four outer regions corresponding to the violation of one of the four constraints $\pm a_{12} \leq a_{ii}$, for $i = 1, 2$. They are numbered from I to IV according to Figure 2. The outer region I is the set of PSD and non diagonal dominant matrices such that $a_{22} < a_{12} < a_{11}$. It is easy to reduce any diffusion matrix to this case by permutation of variables and change of sign of one state variable. Therefore in the sequel we will discuss essentially the fast decomposition of such matrices. Note that for PSD and diagonal dominant matrices in region I an alternative decomposition, involving the identity matrix, and referred to in Section 5, is

$$a = (a_{11} - a_{22})\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} + (a_{22} - a_{12})\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} + a_{12}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}. \tag{10}$$

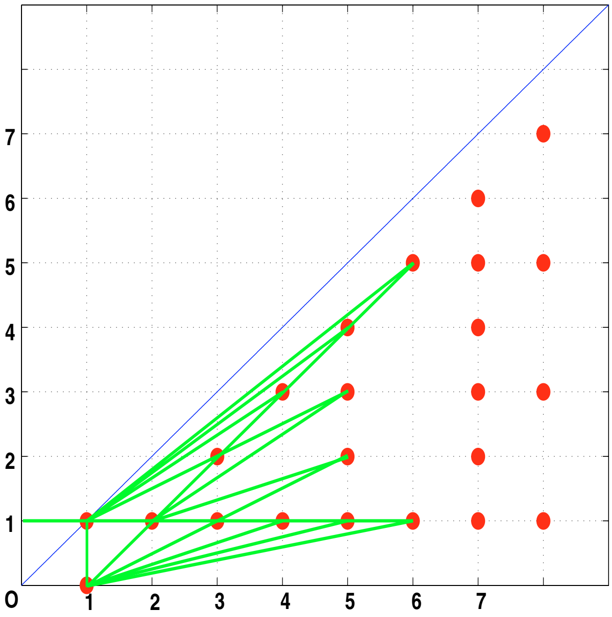

Figure 4. Family relations in regular grid.

4. The Stern-Brocot tree

If the function φ of Section 2, defined over $\mathbb{Z}^n$, is the value at grid points of a smooth function $\Phi : \mathbb{R}^n \to \mathbb{R}$, *i.e.*, $\varphi_k = \Phi(x_k)$, where $x_k := \sum_i k_i h_i$, then the operator Δ_ξ defined in (2) allows, as it can be seen by a Taylor expansion of Φ around x_k, to obtain a consistent approximation of $\Phi''(x_k)(x_\xi, x_\xi)$, the curvature of Φ at x_k along direction x_ξ. The consistency condition (3) expresses the fact that a nonnegative combination of such curvatures equals the second order term of the HJB equation. Two elements of the stencil generate the same direction if they are not linearly independent. Since the algorithm should use points in the stencil as close to x_k as possible, it suffices to take such ξ with relatively prime components.

For two dimensional problems on which we focus now, such points have a specific structure. Assume for simplicity that $k = 0$. For reason of symmetries, we have displayed in Figure 4 one eighth of the neighbouring points, namely the points ξ in $\mathbb{Z}^2_+$, such that $\xi_2 \leq \xi_1$. Those with an irreducible associated (symbolic) fraction ξ_2/ξ_1, that we will call irreducible points, are in red (boldface in black and white). These points are connected by segments that represent the arcs of a tree that we introduce now.

A very effective way for generating direction with irreducible components is to use the *Stern-Brocot tree*, see *e.g.* [8], (which, by the way, is not a tree in the classical sense), displayed in Figure 5. In the sequel, when we write q/p this should be understood as the pair (p, q), so that $p = 0$ makes no problem.

The tree starts with two roots $0/1$ and $1/0$. At any stage of the construction, between two adjacent nodes q/p and q'/p', called the parents, insert the child node $(q + q')/(p + p')$. The two roots are adjacent, and hence, the first child is $1/1$. Then each child is made adjacent with each of his two parents, and we can repeat the process of generating children (in any order).

Figure 4 shows the links between parents and child for the first nodes of the Stern-Brocot tree. One finds the two parents of a child-node following the two segments starting from this point and going to the left.

For convenience we give a short proof of classical properties of the Stern-Brocot tree (see also [8], Sect. 4.5).

Lemma 4.1. *Let q/p and q'/p' be adjacent nodes such that $q/p < q'/p'$, with child q''/p'', where $p'' = p + p'$, $q'' = q + q'$. Then*

(i) $q/p < q''/p'' < q'/p'$;

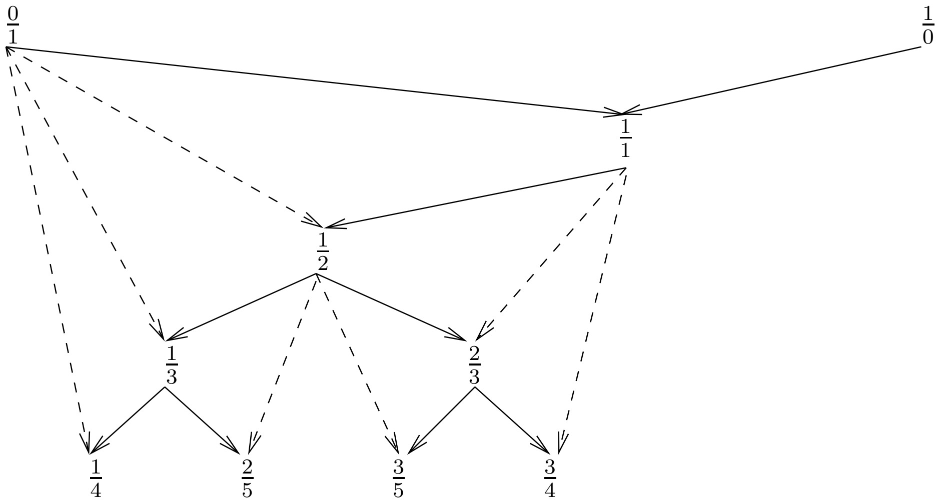

FIGURE 5. Stern-Brocot tree.

(ii) *every node of the Brocot tree is irreducible;*

(iii) *every irreducible fraction b/a belongs to the Brocot tree.*

Furthermore, if q/p and q'/p' are adjacent nodes of the tree such that $q/p < b/a < q'/p'$, then

$$a \geq p + p'; \quad b \geq q + q'. \tag{11}$$

Proof.

(i) It is easily checked that $q/p < (q+q')/(p+p') < q'/p'$. (This property explains why generation of sons may be made in any order.)

(ii) We prove by induction that, if q/p and q'/p' are adjacent nodes of the tree, then

$$q'p - qp' = 1. \tag{12}$$

The relation is obviously true for the root nodes 0/1 and 1/0. Assume that it is satisfied for adjacent nodes q/p and q'/p'. It follows from (12) that $q'(p+p') - p'(q+q') = 1$ and $p(q+q') - q(p+p') = 1$, proving the induction. Combining (12) and Bézout's theorem, we obtain (ii).

(iii) Let b/a be an irreducible fraction, with $0 < b/a < 1$, and q/p, q'/p' be adjacent nodes of the tree such that $q/p < b/a < q'/p'$. Then $bp - aq \geq 1$ and $aq' - bp' \geq 1$. Multiply the first (second) inequality by p' (by p) and add them; multiply the first (second) inequality by q' (by q) and add them; using (12), relation (11) follows. Since $p'' \geq \max(p, p') + 1$, this relation implies that there is a finite number of couple of adjacent nodes $(q/p, q'/p')$ in the tree such that $q/p < b/a < q'/p'$ holds. This is the case for the two root nodes. Assume now that b/a does not belong to the Stern-Brocot tree. If $q/p < b/a < q'/p'$, setting $q'' = q + q'$ and $p'' = p + p'$, we see that either $q/p < b/a < q''/p''$, or $q''/p'' < b/a < q'/p'$. In this way we generate an infinite sequence of adjacent nodes such that $q/p < b/a < q'/p'$. The desired contradiction follows. □

5. DECOMPOSITION OF THE SCALED DIFFUSION MATRIX

In the sequel of this paper we will present a fast algorithm for computing the decomposition of diffusion matrices, when the stencil is the set of directions with integer irreducible components, with bound p_{max} on the

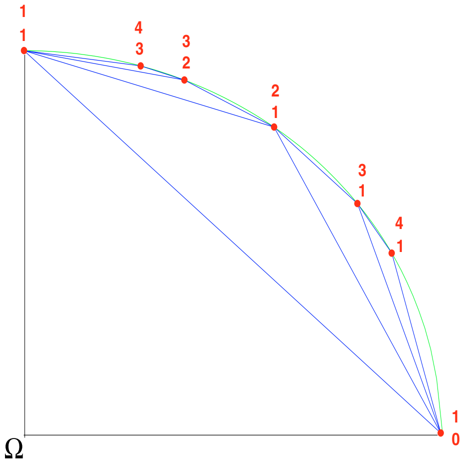

Figure 6. Correspondence of directions.

absolute value of components:

$$\mathcal{S}_{p_{max}} := \{(\xi_1,\xi_2) \in \mathbb{Z}\times\mathbb{N};\ \max(|\xi_1|,\xi_2) \le p_{max};\ (|\xi_1|,\xi_2) \text{ irreducible}\}\,.$$

(The point $(0,0)$ is considered as not irreducible here.) The polyhedral cone generated by these directions is $\mathcal{C}(\mathcal{S}_{p_{max}}) = \{\sum_{\xi\in\mathcal{S}_{p_{max}}} \alpha_\xi \xi\xi^\top;\ \alpha_\xi \ge 0\}$.

As discussed at the end of Section 3, it suffices to discuss the case when the matrix a^h is in outer region I; *i.e.*, when it is PSD and non diagonal dominant, and $a_{22} < a_{12} < a_{11}$. In Figure 6, this means that the view of a^h belongs to the quarter of ball in the upper right side, and is not in the triangle with vertices of coordinates $(0,0)$, $(1,0)$ and $(0,1)$. The latter correspond to the identity matrix, and to degenerate diffusions with horizontal and angle of $\pi/4$ diffusions. (The cone generated by these three points is a set of diagonal dominant matrices.)

With every node q/p of the Stern-Brocot tree, $q \le p$, associate directions $\xi_{p,q} := (p\ q)^\top$ and $X_{p,q} := \xi_{p,q}\xi_{p,q}^\top$. With two adjacent nodes is associated the plane $H(q/p, q'/p')$ generated by $X_{p,q}$ and $X_{p',q'}$, and two half spaces, the inner one (containing the identity matrix) and the outer one. Denote by $P_H(q/p, q'/p')$ the orthogonal projection onto this plane (since the mapping (7) onto $\mathbb{R}^3$ is norm invariant, projection w.r.t. Frobenius norm is equivalent to the Euclidean projection in the image space $\mathbb{R}^3$).

Beginning the search of a decomposition, we are in the following situation: the matrix a^h belongs to the outer half space of $H(0/1, 1/1)$. So, let us assume more generally that a^h belongs to the outer half space of $H(q/p, q'/p')$, where q/p and q'/p' are adjacent nodes. In Figure 6 we have drawn the views of the first nodes of the Stern-Brocot tree; the segments are the views of the segment between two neighbouring nodes of this tree.

We see, using Lemma 3.1 that we have to use another element of stencil of the form $\hat{q}/\hat{p}$, with $\hat{q}$ and $\hat{p}$ nonnegative, such that $q/p < \hat{q}/\hat{p} < q'/p'$, and as small as possible. In view of (11), the optimal choice is to take the child $q''/p'' = (q+q')/(p+p')$. Then (see Fig. 6) there are two possibilities:

- The matrix a^h belongs to both inner half spaces of $H(q/p, q''/p'')$ and $H(q''/p'', q'/p')$. Then a^h belongs to the cone generated by $X_{p,q}$, $X_{p',q'}$ and $X_{p'',q''}$. Since these three matrices are linearly independent, the corresponding coefficients are unique nonnegative solution of the invertible (three dimensional) system
$$\alpha_{p,q}\ X_{p,q} + \alpha_{p',q'}\ X_{p',q'} + \alpha_{p'',q''}\ X_{p'',q''} = a^h. \tag{13}$$
- The matrix a^h belongs to at least one outer half space. Since $X_{p'',q''}$ belongs to the boundary of the cone of PSD matrices, a^h cannot belong to both outer half spaces (see Fig. 6). We are therefore lead to the situation at the beginning, setting either q/p or q'/p' to q''/p''.

If $p'' > p_{max}$, we replace a^h by its projection onto the cone generated by matrices of the form X_{p_i,q_i}, with either $q_i/p_i < q/p$ or $q'/p' < q_i/p_i$. Note that this projection belongs to the cone generated by $X_{p,q}$ and $X_{p',q'}$. As above, since these two matrices are linearly independent, the corresponding coefficients are unique nonnegative solution of the system

$$\alpha_{p,q}\ X_{p,q} + \alpha_{p',q'}\ X_{p',q'} = P_H(q/p, q'/p')\ a^h. \tag{14}$$

This leads to an effective algorithm, that will stop either if the exact decomposition is obtained, or if either $p'' > p_{max}$, or if the projection of a^h onto $H(q/p, q'/p')$ is close enough to a^h. The precise algorithm is as follows; ε is a threshold for the distance to the projection of a^h onto the class of consistent matrices, and p_{max} is the size of stencil:

Algorithm DECOMP

```
INITIAL PHASE: Data  $\varepsilon \geq 0,\ p_{max} > 0$. Set $k := 0$.
  • If $a^h$ is diagonal dominant: set $\alpha$ using (10) and stop.
  • Reduction to region I, i.e. $a^h_{22} < a^h_{12} < a^h_{11}$.
    Set $q_0/p_0 := 0/1$, $q'_0/p'_0 := 1/1$.
REPEAT
  • Compute $a' := P_H(q/p, q'/p')a^h$.
  • If $\|a' - a^h\| \leq \varepsilon\|a^h\|$ or $p + p' > p_{max}$: compute $\alpha$, decomposition of $a'$ using (14) and stop.
  • Set $q''/p'' := (q + q')/(p + p')$.
  • If $a^h$ in inner half spaces of $H(q/p, q''/p'')$ and $H(q/p, q''/p'')$: compute $\alpha$ using (13) and stop.
  • If $a$ is in outer half space of $H(q/p, q''/p'')$: $q'/p' := q''/p''$.
    Otherwise $q/p := q''/p''$.
  • $k := k + 1$.
END REPEAT
```

From the above discussion we have the following result.

Theorem 5.1. *Algorithm* **DECOMP** *provides a decomposition of a^h with a relative error lower than ε, and stops after at most p_{max} iterations. The cost of each iteration is $O(1)$ operations, and hence, its total cost is no more than $O(p_{max})$.*

Obviously it is useful to compute the largest distance between a^h and its projection (as a function of p_{max}) and to evaluate the resulting approximation error. This is the subject of the next section.

6. Projection errors for scaled diffusion matrices

Let us compute the expression of the maximal relative distance of a diffusion matrix to the polyhedral cone $\mathcal{C}(\mathcal{S}_{p_{max}})$:

$$\varepsilon_{p_{max}} := \max_a \operatorname{dist}\left(\frac{a}{\|a\|},\ \mathcal{C}(\mathcal{S}_{p_{max}})\right),$$

where the maximum is over all possible nonzero diffusion matrices, *i.e.*, nonzero PSD matrices. By $\lceil r \rceil$ we denote the smallest integer greater than r.

Lemma 6.1. *The distance from a PSD matrix a to $\mathcal{C}(\mathcal{S}_{p_{max}})$ is at most $\varepsilon_{p_{max}}\|a\|$, and*

$$\varepsilon_{p_{max}} = \frac{\sqrt{p_{max}^2 + 1} - p_{max}}{\sqrt{2}\ \sqrt{2\ p_{max}^2 + 1}} \leq \frac{1}{4}\ p_{max}^{-2}. \tag{15}$$

Conversely, given $\varepsilon > 0$, the distance from a to $\mathcal{C}(\mathcal{S}_{p_{max}})$ is at most ε when $p_{max} \geq p_\varepsilon$, with

$$p_\varepsilon := \left\lceil \frac{\sqrt{1 - \varepsilon^2} - \varepsilon}{2\sqrt{\varepsilon\sqrt{1 - \varepsilon^2}}} \right\rceil. \tag{16}$$

TABLE 1. First values of $\varepsilon_{p_{max}}$ and p_ε.

p_{max}	1	2	3	4	5	15
$\varepsilon_{p_{max}}$	0.169102	0.055642	0.026325	0.015153	0.009804	0.001109

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-7}
p_ε	2	5	16	20	159	1 582

Proof. The first sentence is a consequence of the fact that the projection onto a cone with vertex at 0 is positively homogeneous. We may assume that $\|a\| = 1$. Let a' be the projection of a onto $\mathcal{C}(\mathcal{S}_{p_{max}})$. Let us prove first that, if a' is the projection on the hyperplane spanned by $\xi\xi^\top$ and $\xi'(\xi')^\top$, then

$$\|a - a'\| \leq \frac{\left(1 - \cos(\widehat{\xi, \xi'})\right)}{\sqrt{2} \cdot \sqrt{1 + \cos^2(\widehat{\xi, \xi'})}} \, \|a\| \tag{17}$$

the bound being sharp. Indeed, we may assume that $\xi = (\cos\theta \;\; \sin\theta)^\top$ and $\xi' = (\cos\theta' \;\; \sin\theta')^\top$. Set $\theta'' := \frac{1}{2}(\theta + \theta')$ and $\xi'' = (\cos\theta'' \;\; \sin\theta'')^\top$. By reasons of symmetry, the maximal error is reached for $a = \xi''(\xi'')^\top$, with $\xi = (\cos\theta'' \;\; \sin\theta'')$, and its projection is of the form $a' = \alpha b$, where $b := (\xi\xi^\top + \xi'(\xi')^\top)$, for some $\alpha \in \mathbb{R}_+$. The minimum w.r.t. α of $\|a - \alpha b\|^2$ is

$$\Delta = \|a\|^2 - (a \circ b)^2/\|b\|^2 = 1 - (a \circ b)^2/\|b\|^2.$$

This amount being invariant w.r.t. translation of angles, we may assume that $\theta + \theta' = 2\theta'' = 0$, and hence $\theta' = -\theta$, $a = (1,\ 0,\ 0)^\top$, $b = \left(2\cos^2\theta,\ 0,\ 2\sin^2\theta\right)^\top$. We obtain $\|b\|^2 = 4(\cos^4\theta + \sin^4\theta)$ and $a \circ b = 2\cos^2\theta$. It follows that $\Delta = 1 - (a \circ b)^2/\|b\|^2 = \sin^4\theta/(\cos^4\theta + \sin^4\theta)$. Setting $\delta = |\theta' - \theta| = 2|\theta|$, and combining with

$$\begin{aligned} 2\sin^2\theta &= 1 - \cos^2\theta + \sin^2\theta = 1 - \cos\delta \\ \cos^4\theta + \sin^4\theta &= (\cos^2\theta - \sin^2\theta)^2 + 2\cos^2\theta\sin^2\theta = \cos^2\delta + \tfrac{1}{2}\sin^2\delta \end{aligned}$$

we get (17).

In the stencil of size p_{max}, the greatest angle between two consecutive vectors is the angle between $\xi_0 = (1\ 0)^\top$ and $\xi_1 = (p_{max}\ 1)^\top$. By (17) and $\cos(\xi_0, \xi_1) = p_{max}/\sqrt{p_{max}^2 + 1}$ we have that, in the p_{max}-stencil, the largest error is (15).

Relation (16) is a simple consequence of (15), see details in the prepublication [5]. □

We display in Table 1 below the first values of $\varepsilon_{p_{max}}$ and some values of p_ε. An algorithm involving only the closest neighbor can make up to 17 % of relative error on diffusion matrices, and hence, will perform poorly in general. A relative precision of 1 % needs to take $p = 5$. This motivates our effort to make a theory for arbitrary large values of p_{max}.

Remark 6.2. If consistency does not hold, and $\varepsilon = 0$, then algorithm **DECOMP** computes the decomposition of the projection $a'(t, x, u)$ of $a(t, x, u)$ onto $\mathcal{C}(\mathcal{S}_{p_{max}})$. In that case, the numerical scheme can be interpreted as a consistent approximation for the perturbed HJB equation

$$\begin{aligned} -v_t(t,x) &= \inf_{u \in U} \left\{ \ell(t,x,u) + f(t,x,u) \cdot v_x(t,x) + a'(t,x,u) \circ v_{xx}(t,x) \right\}, \\ &\quad \text{for all } \; t, x \in [0,T] \times \mathbb{R}^n. \\ v(T,x) &= \ell_F(x), \; \text{ for all } \; x \in \mathbb{R}^n. \end{aligned} \tag{HJB_p}$$

Denote by v' the (well-defined) corresponding solution. When the step sizes vanishes, the limit of error between the solution of HJB and the one of the scheme is $\|v - v'\|_\infty$. By ([9], Th. 4.1), and then combining with Lemma 6.1, we obtain estimates of the type

$$\|v - v'\|_\infty \leq C\|a - a'\|_\infty^{1/2} \leq C'\varepsilon_{p_{max}}^{1/2} \leq C''/p_{max}, \tag{18}$$

where C and C' do not depend on p_{max}. If diffusions are uniformly invertible we have that $\|v - v'\|_\infty \leq C_1\|a - a'\|_\infty$, and hence, $\|v - v'\|_\infty \leq C_1'/p_{max}^2$, where C_1 and C_1' do not depend on p_{max}. For infinite horizon problems we can obtain similar results applying ([1], Lem. 2.6), (with the same exponents if the discounting coefficient is "large enough", and otherwise with smaller exponents).

7. Numerical results

We have implemented algorithm **DECOMP** in the C programming language and tested it on two academic examples in which the value function is known. Also we integrate on a finite rectangular domain with exact values on the boundary. This allows to compute the error made by the scheme and to see if its behavior is in agreement with the theory. For points of the grid close to the boundary, the size of the stencil may be smaller than p_{max} since points out of the domain are not used. Therefore, in the vicinity of the boundary the errors of approximation of diffusion matrices are larger than far from the boundary. We use the reverse-time function $W(s,x) = V(T-s,x)$ in order to integrate t from 0 to T.

7.1. An uncontrolled problem

Our first test function is

$$\begin{cases} W(t,x_1,x_2) = (1+t)\sin x_1 \sin x_2 \\ 0 \leqslant x_1 \leqslant \pi; \quad 0 \leqslant x_2 \leqslant \pi; \quad 0 \leqslant t \leqslant 1. \end{cases} \tag{19}$$

We choose $\Delta x := h_1 = h_2$, $N_1h_1 = N_2h_2 = \pi$, and the measure of error is $e := \|W_{approx} - W_{exact}\|_1/(N_1N_2)$, where $\| W\|_1 := \sum_{i,j} |W_{i,j}|$. The following expressions for ℓ, f and σ are compatible with the HJB equation:

$$\begin{cases} \ell(t,x_1,x_2) = \sin x_1 \sin x_2[1 + (1+2\beta)(1+t)] \\ \qquad\qquad\qquad -2(1+t)\ \cos x_1 \cos x_2 \sin(x_1+x_2)\cos(x_1+x_2) \\ f(t,x_1,x_2) = 0 \\ a(t,x_1,x_2) = \begin{pmatrix} \sin^2(x_1+x_2)+\beta^2 & \sin(x_1+x_2)\cos(x_1+x_2) \\ \sin(x_1+x_2)\cos(x_1+x_2) & \cos^2(x_1+x_2)+\beta^2 \end{pmatrix} \end{cases}$$

here $\sigma(t,x_1,x_2) = \sqrt{2}\begin{pmatrix} \sin(x_1+x_2) & \beta & 0 \\ \cos(x_1+x_2) & 0 & \beta \end{pmatrix}$.

We display in Figure 7 the error as a function of discretization step, for $\beta^2 = 0.1$ and 0, and $p_{max} = 5$. The scheme is consistent only in the first case. Accordingly, the error decreases when the space step is reduced in the first case, but not in the other.

7.2. Numerical example, optimal control

We consider here an optimal control problem where $\sigma(\cdot)$ and $a(\cdot)$ do not depend on the control. Also, the drift is $f(t,x,u) = u$, with restriction $u_1^2 + u_2^2 \leq 1$. The test function is

$$\begin{cases} W(t,x_1,x_2) = (1+t)\sin x_1 \sin x_2 \\ -1 \leqslant x_1 \leqslant 1; \quad -1 \leqslant x_2 \leqslant 1; \quad 0 \leqslant t \leqslant 0.5. \end{cases} \tag{20}$$

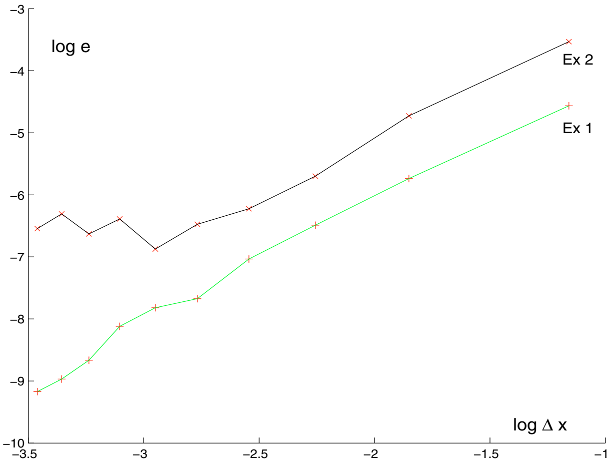

Figure 7. Error *vs.* discretization step, $p_{max} = 5$.

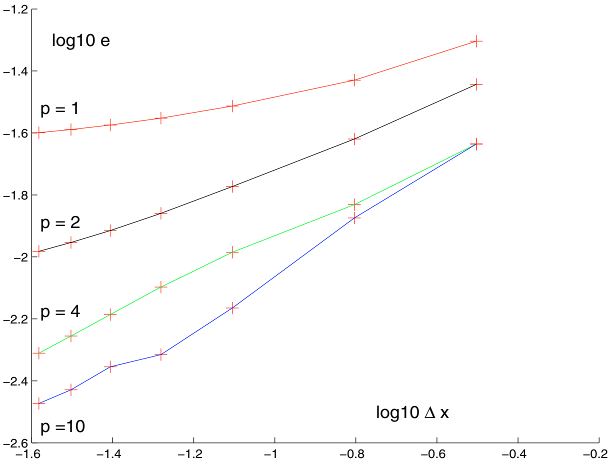

Figure 8. Error *vs.* discretization step, optimal control, $p_{max} = 1,\ 2,\ 4,\ 10$.

We have here a degenerate diffusion $a(t, x_1, x_2) = \frac{1}{2}\sigma(t, x_1, x_2)\sigma(t, x_1, x_2)^\top$ with

$$\sigma_1(t, x_1, x_2) = \sqrt{2}\ \sin(x_1 + x_2),\ \ \sigma_2(t, x_1, x_2) = \sqrt{2}\ \cos(x_1 + x_2).$$

The resulting running cost $\ell(t, x_1, x_2)$ is here

$$\sin(x_1)\sin(x_2) + (1+t)\Big[\big(\cos^2(x_1)\sin^2(x_2) + \sin^2(x_1)\cos^2(x_2)\big)^{1/2} + \sin(x_1)\sin(x_2) - 2\sin(x_1+x_2)\cos(x_1+x_2)\cos(x_1)\cos(x_2)\Big].$$

We display in Figure 8 the error, as defined in Section 7.1, *vs.* the discretization step for several values of p_{max}. Although the scheme is not consistent, it appears that the discretization errors are quite small.

Acknowledgements. The authors thank two anonymous referees for their useful remarks.

References

[1] G. Barles and E.R. Jakobsen, On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations. *ESAIM: M2AN* **36** (2002) 33–54.

[2] G. Barles and E.R. Jakobsen, Error bounds for monotone approximation schemes for Hamilton-Jacobi-Bellman equations (to appear).

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

[4] J.F. Bonnans and H. Zidani, Consistency of generalized finite difference schemes for the stochastic HJB equation. *SIAM J. Numer. Anal.* **41** (2003) 1008–1021.

[5] J.F. Bonnans, E. Ottenwaelter and H. Zidani, *A fast algorithm for the two dimensional HJB equation of stochastic control.* Technical report, INRIA (2004). Rapport de Recherche 5078.

[6] F. Camilli and M. Falcone, An approximation scheme for the optimal control of diffusion processes. *RAIRO Modél. Math. Anal. Numér.* **29** (1995) 97–122.

[7] W.H. Fleming and H.M. Soner, *Controlled Markov processes and viscosity solutions.* Springer, New York (1993).

[8] R.L. Graham, D.E. Knuth and O. Patashnik, *Concrete Mathematics, A Foundation For Computer Science.* Addison-Wesley, Reading, MA (1994). Second edition.

[9] E.R. Jakobsen and K.H. Karlsen, Continuous dependence estimates for viscosity solutions of fully nonlinear degenerate parabolic equations. *J. Differ. Equations* **183** (2002) 497–525.

[10] N.V. Krylov, On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients. *Probab. Theory Related Fields* **117** (2000) 1–16.

[11] H.J. Kushner, *Probability methods for approximations in stochastic control and for elliptic equations.* Academic Press, New York (1977). *Math. Sci. Engrg.* **129**.

[12] H.J. Kushner and P.G. Dupuis, *Numerical methods for stochastic control problems in continuous time.* Springer, New York, *Appl. Math.* **24** (2001). Second edition.

[13] P.-L. Lions, Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. Part 2: Viscosity solutions and uniqueness. *Comm. Partial Differential Equations* **8** (1983) 1229–1276.

[14] P.-L. Lions and B. Mercier, Approximation numérique des équations de Hamilton-Jacobi-Bellman. *RAIRO Anal. Numér.* **14** (1980) 369–393.

[15] J.-L. Menaldi, Some estimates for finite difference approximations. *SIAM J. Control Optim.* **27** (1989) 579–607.