SIAM J. on Numerical Analysis, to appear
Version of December 10, 2002

Consistency of generalized finite difference schemes for the stochastic HJB equation

J.F. Bonnans[1] and H. Zidani[2]
December 10, 2002

Abstract

We analyze a class of numerical schemes for solving the HJB equation for stochastic control problems, which enters in the framework of Markov chain approximations, and generalizes the usual finite difference method. The latter is known to be monotonic, and hence valid, only if the scaled covariance matrix is dominant diagonal. We generalize this result by, given the set of neighboring points allowed to enter in the scheme, showing how to compute effectively the class of covariance matrices that is consistent with this set of points. We perform this computation for several cases in dimension 2 to 4.
AMS subject classifications. 93E20, 49L99

1 Motivation

This paper is devoted to the discussion of numerical algorithms for solving the stochastic optimal control problems. In order to simplify the presentation of the main ideas, consider the following model problem (Fleming and Rishel [4], Lions and Bensoussan [7])

$$(P_x) \qquad \left\{ \begin{array}{l} \mathrm{Min}\, W(x,u) = \mathbb{E}\displaystyle\int_0^\infty \ell(y_{x,u}(t),u(t))e^{-\lambda t}\mathrm{d}t; \\ \\ \left\{ \begin{array}{l} \mathrm{d}y_{x,u}(t) = f(y_{x,u}(t),u(t))\mathrm{d}t + \sigma(y_{x,u}(t),u(t))\mathrm{d}w(t), \\ y_{x,u}(0) = x, \end{array} \right. \\ \\ u(t) \in U, \quad t \in [0,\infty[. \end{array} \right.$$

[1]Inria-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay, France. Email: Frederic.Bonnans@inria.fr
[2]Unité de Mathématiques Appliquées, ENSTA, 32 Boulevard Victor, 75739 Paris Cedex 15, France. Email: zidani@ensta.fr

Here $y_{x,u}(t) \in \mathbb{R}^n$ is the state variable, $u(t) \in \mathbb{R}^m$ is the control variable, that for almost all t must belong to the set $U \subset \mathbb{R}^m$, $\lambda > 0$ is the discounting factor, $\ell : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a distributed cost, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is a deterministic dynamics, $\sigma(\cdot, \cdot)$ is a mapping from $\mathbb{R}^n \times \mathbb{R}^m$ into the space of $n \times r$ matrices, and w is a standard r dimensional Brownian motion. We are assuming full observation of the state, and are looking for a control in the class of feedback controls. In the sequel we assume f, σ, and ℓ, to be Lipschitz and bounded. Then the solution to the stochastic differential equation and associated cost are well defined (e.g. Fleming and Soner [5]). The covariance matrix is defined as is

$$a(x, u) := \sigma(x, u)\sigma(x, u)^\top, \quad \forall \ (x, u) \in \mathbb{R}^n \times \mathbb{R}^m, \tag{1.1}$$

where by $\top$ we denote the transposition operator. It is known (see P.L. Lions [8], and also Fleming and Soner [5]) that the value function V of problem (P_x), defined by $V(x) = \inf_u W(x, u)$, is the unique bounded viscosity solution of the Hamilton-Jacobi-Bellman (HJB) equation

$$\lambda V(x) \ = \ \mathcal{H}(x, V_x(x), V_{xx}(x)), \quad \text{for all } \ x \in \mathbb{R}^n; \tag{1.2}$$

the Hamiltonian $\mathcal{H}$ being defined as

$$\mathcal{H}(x, p, Q) := \inf_{u \in U} \left\{ \ell(x, u) + f(x, u) \cdot p + \tfrac{1}{2} \sum_{i,j=1}^{n} a_{ij}(x, u)\, Q_{ij} \right\}, \tag{1.3}$$

where $x \in \mathbb{R}^n$, $p \in \mathbb{R}^n$, and Q is a $n \times n$ symmetric matrix. A basic idea for discretizing this problem is as follows (an up to date synthesis of this approach is given in Kushner and Dupuis [6]). Consider a regular grid G^h of discretization of the state space $\mathbb{R}^n$, with discretization steps $h = (h_1, \ldots, h_n)$. With the coordinate $k = (k_1, \ldots, k_n)$ in $\mathbb{Z}^n$ is associated the point $x_k \in \mathbb{R}^n$ of the form

$$x_k := (k_1 h_1, \ldots, k_n h_n). \tag{1.4}$$

Of course the real computations should be performed on a finite grid. However, we will not discuss this point, and rather analyze the result of computations on this infinite grid. Let us consider an optimal control problem for a Markov chain on the grid G^h. Let $\{X_q^h, q \geq 0\}$ be the states of the Markov chain at time q, with transition probabilities denoted by $p^h(x, y \mid u)$, where

$u \in U$ is the canonical control value. Let Δt^h be an"*interpolation interval*", satisfying $\Delta t^h \to 0$ as $h \to 0$ and let $\mathbb{E}_{k,q}^{h,u}$ be the conditional expectation of X_{q+1}^h, given that $\{X_q^h = x_k\}$, and the control value u. Suppose that the chain obeys the following "*local consistency*" conditions:

$$\mathbb{E}_{k,q}^{h,u}\left[X_{q+1}^h - x_k\right] = \Delta t^h f(x_k, u) + o(\Delta t^h), \tag{1.5a}$$

$$\mathrm{Cov}_{k,q}^{h,u}\left[X_{q+1}^h - x_k\right] = \Delta t^h a(x_k, u) + o(\Delta t^h), \tag{1.5b}$$

$$\sup_q \left|X_{q+1}^h - X_q^h\right| \to 0. \tag{1.5c}$$

A possible adaptation for the cost function to this Markov chain is the following:

$$W^h(x, u^h) = \Delta t^h \mathbb{E}\left[\sum_{q\geq 0} \ell(X_q^h, u_q^h)(1 + \lambda \Delta t^h)^{-q-1}\right] \tag{1.6}$$

where $u^h = (u_q^h)$, and $u_q^h \in U$ denote the random variable which represents the control action for the chain at discrete time q. Then the dynamic programming equation for the controlled chain $\{X_q^h, q \geq 0\}$ and the cost (1.6) is:

$$V^h(x_k) = (1 + \lambda \Delta t^h)^{-1} \operatorname*{Min}_{u\in U}\left[\Delta t^h \ell(x_k, u) + \sum_{y\in G^h} p(x_k, y \mid u) V^h(y)\right], \tag{1.7}$$

for $x_k \in G^h$. It is known that the function V^h converges uniformly over compact sets to the value function V, for the original problem, as $h \to 0$, whenever the "local consistency" conditions (1.5) are satisfied, the interpolation interval possibly depending on (x, u); see Kushner and Dupuis [6].

Now the Markov chain approximation method consists in finding a chain $\{X_q^h\}$ satisfying the "local consistency" (1.5). A standard way for the construction of such approximating chain is to use the finite difference approximations. However, this works only if the matrix a has a dominant diagonal (see section 3 for details), whereas this matrix may be an arbitrary semidefinite positive matrix. In some cases it is possible to make a change of variables in the state space in order that this hypothesis is satisfied, see e.g. Kushner and Dupuis [6, Section 5.4]. However, when the control enters in the matrix σ, and hence, also in a, this is no more possible in general. By contrast, the Markov chain approximation method is, in principle, able to handle the case

when the covariance matrix is not dominant diagonal. In fact, relation (1.5) essentially gives linear relations (to be satisfied approximately) on the transition probabilities, while the latter have to be nonnegative, and of sum equal to 1. Several questions then arise. First of all, since the Markov chain represents the discretization of a partial differential equation (PDE), it is highly desirable to limitate the transitions from one point of the grid to other points that are not too far. Also, for computational complexity reasons, the number of transitions should be as small as possible.

We are led, then, to the following question. Given a point in the grid with coordinate k and a control, choose a set of other points in the grid to which transitions are allowed. For instance we may allow transitions to points for which the coordinate k' are such that $|k'_i - k_i| \leq 1$ for all i. More generally, we choose a set of neighbors defined by constraints on the difference of coordinates $k' - k$. Is it possible then to compute consistent transition probabilities ? In other words, what is the class of covariance matrices that is compatible with such a choice of possible transitions ? And then, what is the cost of computing the transition probabilities themselves ? Finally, on what basis should we choose the neighbors ?

These are several delicate questions. The paper is essentially devoted to the first of them, i.e., how to check the consistency condition. Note that our results apply also to finite horizon problems, in which the value function depends on time and space, since the analysis of consistency for these problems leads basically to questions of the same nature. Similarly, we discuss only explicit schemes, but implicit schemes (in connection to the policy iteration methods, see Kushner and Dupuis [6, Section 6.2]) also lead to the same questions and our results apply also to this case. Note that, in the case of a covariance matrix that is a smooth function of the state only, it is possible to state a consistent approximation using finite elements, see Chung, Hanson and Xu [3]. However, it is not easy to extend this idea to the case when the covariance matrix either is not differentiable, or depends also on the control.

2 Generalized finite differences

Let us present a generalization of the usual finite difference schemes; we will see later that these generalized finite differences are in fact a particular case of Markov chain approximation. Let $\varphi = \{\varphi_k\}$ be a real valued function over $\mathbb{Z}^n$. With $\xi \in \mathbb{Z}^n$, associate the *shift* operator δ_ξ defined by $\delta_\xi \varphi_k := \varphi_{\xi+k}$.

Consider the finite difference operator $\Delta_\xi = \delta_\xi + \delta_{-\xi} - 2\delta_0$, in other words,

$$\Delta_\xi \varphi_k := \varphi_{k+\xi} + \varphi_{k-\xi} - 2\varphi_k = \varphi_{k+\xi} - \varphi_k - (\varphi_k - \varphi_{k-\xi}). \tag{2.8}$$

If Φ is a C^2 (twice continuously differentiable) function over $\mathbb{R}^n$, and $\varphi_k = \Phi(x_k)$ for all k, then by a standard Taylor expansion, we have that

$$\Delta_\xi \varphi_k := \sum_{i,j=1}^{n} h_i h_j \xi_i \xi_j \Phi_{x_i x_j} + o(\|h\|^2). \tag{2.9}$$

For instance, when ξ is equal to e_i (the ith element of the natural basis of $\mathbb{R}^n$) and $e_i \pm e_j$, resp., we obtain

$$\left\{ \begin{array}{lcl} \Delta_{e_i} \varphi_k & = & (h_i)^2 \Phi_{x_i x_i} + o(\|h\|^2), \\ \Delta_{e_i \pm e_j} \varphi_k & = & (h_i)^2 \Phi_{x_i x_i} + (h_j)^2 \Phi_{x_j x_j} \pm 2 h_i h_j \Phi_{x_i x_j} + o(\|h\|^2). \end{array} \right. \tag{2.10}$$

Denote by v_k the approximation of the value function V at x_k. Let $D_k^u v_k$ be a notation for the upwind spatial finite difference,

$$(D_k^u v_k)_i = \frac{v_{k+e_i} - v_k}{h_i} \quad \text{if} \quad f(x_k, u)_i \geq 0, \quad \frac{v_k - v_{k-e_i}}{h_i} \quad \text{if not.} \tag{2.11}$$

Now let $\mathcal{S}$ be a finite set of $\mathbb{Z}^n \setminus \{0\}$ containing $\{e_1, \cdots, e_n\}$. We consider explicit schemes based on the difference operators that we just discussed, namely

$$\lambda v_k \quad = \quad \inf_{u \in U} \left\{ \ell(x_k, u) + f(x_k, u) \cdot D_k^u v_k + \sum_{\xi \in \mathcal{S}} \alpha_{k,\xi}^u \Delta_\xi v_k \right\}, \tag{2.12}$$

for all $k \in \mathbb{Z}^n$. We will see soon how to chose the coefficients $\alpha_{k,\xi}^u$ in order to have a convergent approximation. Note that, since $\Delta_\xi = \Delta_{-\xi}$, we may assume without loss of generality that either $\alpha_{k,\xi}^u$ or $\alpha_{k,-\xi}^u$ is zero, for all ξ. In particular, we may assume that $\alpha_{k,-e_i}^u$ is zero for all i. Note that there are other possibilities than (2.11) for discretizing the first-order term. For instance, it may be useful to consider centered differences in order to obtain (if the solution is smooth enough) higher orders of accuracy. However, since the difficulty for obtaining consistency lies in the discretization of the second-order term in the HJB equation, we will not elaborate on this.

Let $\Delta t^h > 0$ denote a *fictitious* time step (fictitious in the sense that the discrete scheme involves space, but not time, so that this time step has no

influence on the solution). Multiplying (2.10) by Δt^h and adding v_k on both sides, get

$$v_k := (1+\lambda\Delta t^h)^{-1} \inf_{u\in U}\left\{ v_k + \Delta t^h\, \ell(x_k,u) + \Delta t^h\, f(x_k,u)\cdot D_k^u v_k + \Delta t^h \sum_{\xi\in\mathcal{S}} \alpha_{k,\xi}^u \Delta_\xi v_k \right\}. \tag{2.13}$$

With straightforward calculations, we can remark that the approximation (2.13) can be written in the form of (1.7), with the transition probabilities:

$$\begin{aligned}
&p^h(x_k, x_k \mid u) = 1 - \Delta t^h \sum_{i=1}^n \left(\frac{|f_i(x_k,u)|}{h_i} + 2\sum_{\xi\in\mathcal{S}} \alpha_{k,\xi}^u \right),\\
&p^h(x_k, x_{k\pm e_i} \mid u) = \Delta t^h \left(\frac{f_i^{\pm}(x_k,u)}{h_i} + \alpha_{k,e_i}^u \right),\\
&p^h(x_k, x_{k\pm\xi} \mid u) = \Delta t^h\, \alpha_{k,\xi}^u \qquad \text{for } \xi\in\mathcal{S}, \xi\neq e_i,\\
&p^h(x_k, y) = 0 \qquad \text{for } y \notin x_{k+\mathcal{S}},
\end{aligned}$$

where $f_i^+(x_k,u) = \max(f_i(x_k,u),0)$ and $f_i^-(x_k,u) = -\min(f_i(x_k,u),0)$.

Note that the sum of transition probabilities is, whatever the choice of coefficients $\alpha_{k,\xi}^u$ is, equal to one. However, that these transition probabilities are nonnegative adds the following condition on $\alpha_{k,\xi}^u$:

$$\alpha_{k,\xi}^u \geq 0, \quad \forall\, (\xi,k,u) \in \mathcal{S}\times\mathbb{Z}^n\times U, \tag{2.14a}$$

$$\sum_{i=1}^n \frac{|f_i(x_k,u)|}{h_i} + 2\sum_{\xi\in\mathcal{S}} \alpha_{k,\xi}^u \leq \left(\Delta t^h\right)^{-1}, \quad \forall\, (k,u)\in\mathbb{Z}^n\times U. \tag{2.14b}$$

The second condition (2.14b) is, always satisfied when Δt^h is small enough, if the left-hand-side of (2.14b) is uniformly bounded. We obtain in (2.23) such a bound. Here again we could take the more general point of view of having a time step depending on (x,u). Again, we prefer not to be general in order to concentrate on the main difficulties.

Assume (2.14) to be satisfied (we will see that (2.14b) is satisfied as soon as the time step is small enough), so that the scheme is a Markov chain approximation (of a specific type since transition probabilities to points of the form $x_{k\pm\xi}$ are equal if $\xi \neq e_i$ for some i. We therefore concentrate on

the local consistency condition (1.5). Since the terms multiplied by each coefficient $\alpha^u_{k,\xi}$ have a mean equal to x_k, we have that:

$$\mathbb{E}^{h,u}_{k,q}\Big[X^h_{q+1} - x_k\Big] = \Delta t^h \sum_i f_i^+(x_k,u)e_i - \Delta t^h \sum_i f_i^-(x_k,u)e_i = \Delta t^h f(x_k,u) \tag{2.15}$$

so that condition (1.5a) is always satisfied. This in turn implies, denoting $\hat{x} := x_k + \Delta t^h f(x_k,u)$,

$$\mathrm{Cov}^{h,u}_{k,q}\Big[X^h_{q+1} - x_k\Big] = \mathbb{E}^{h,u}_{k,q}\left[(X^h_{q+1} - \hat{x}_k)(X^h_{q+1} - \hat{x}_k)^\top\right] + o(\Delta t^h), \tag{2.16}$$

and, therefore,

$$\mathrm{Cov}^{h,u}_{k,q}\Big[X^h_{q+1} - x_k\Big] = \Delta t^h \sum_{\xi\in\mathcal{S}}\sum_{i,j} h_i h_j \xi_i \xi_j \alpha^u_{k,\xi} e_i e_j^\top + o(\Delta t^h). \tag{2.17}$$

In view of (2.17), and since $\Delta t^h \to 0$ as $h \to 0$, local consistency holds iff we have

$$\sum_{\xi\in\mathcal{S}}\sum_{i,j} h_i h_j \xi_i \xi_j \alpha^u_{k,\xi} e_i e_j^\top = a(x_k,u) + o(1). \tag{2.18}$$

In the sequel, we discuss the *strong consistency* property

$$\sum_{i,j} h_i h_j \xi_i \xi_j \alpha^u_{k,\xi} e_i e_j^\top = a(x_k,u) \quad \text{for all} \quad k \in \mathbb{Z}^n. \tag{2.19}$$

Let a^h denote the scaled covariance matrix $\{a_{ij}/h_i h_j\}$. Then a condition equivalent to (2.19) is

$$\sum_{\xi\in\mathcal{S}} \alpha^u_{k,\xi} \xi\xi^\top = a^h(x_k,u), \quad \text{for all} \quad k \in \mathbb{Z}^n. \tag{2.20}$$

Since every $\alpha^u_{k,\xi}$ is nonnegative, strong consistency means that the symmetric matrix $a^h(x_k,u)$ belongs, for all k and u, to the cone generated by the set $\{\xi\xi^\top; \xi \in \mathcal{S}\}$ that we denote

$$\mathcal{C}(\mathcal{S}) := \left\{ \sum_{\xi\in\mathcal{S}} \alpha_\xi \xi\xi^\top; \alpha \in \mathbb{R}_+^{|\mathcal{S}|} \right\}. \tag{2.21}$$

Note that strong consistency implies a bound on the coefficients $\alpha^u_{k,\xi}$, which in turn allows to obtain an estimate of the fictitious time step.

Lemma 2.1 *Assume that the strong consistency condition holds. Then*

$$\sum_{\xi \in \mathcal{S}} \alpha^u_{k,\xi} \leq \operatorname{trace} a^h(x_k, u), \tag{2.22}$$

and hence, condition (2.14b) *for the fictitious time step is satisfied whenever*

$$\sum_{i=1}^{n} \frac{\|f_i\|_\infty}{h_i} + 2\| \operatorname{trace} a^h\|_\infty \leq \left(\Delta t^h\right)^{-1}. \tag{2.23}$$

Proof. Taking the trace of both sides of (2.20), and since the trace of $\xi\xi^\top$ is greater or equal 1, obtain (2.22). The second part of the lemma is immediate. ∎

It follows from this lemma that, when $h \to 0$, we may take Δt^h of order $O(\min_i h_i^2)$, as expected.

Now we can summarize the results of this section in the following theorem:

Theorem 2.1 *Let $\mathcal{S}$ be a fixed finite set of $\mathbb{Z}^n \setminus \{0\}$ containing $\{e_1, \cdots, e_n\}$, and let h be a fixed step size. Assume that, for every $k \in \mathbb{Z}^n$ and every $u \in U$, the scaled covariance matrix $a^h(x_k, u)$ belongs to the cone $\mathcal{C}(\mathcal{S})$. Then the scheme* (2.13) *is a consistent Markov chain approximation whenever the coefficients $(\alpha^u_{k,\xi})$ are nonegative and satisfy condition* (2.20)*, the time step being such that* (2.23) *is satisfied.*

As said before, condition (2.23) is not really restrictive since Δt^h is just a fictious time step. Note implicit schemes can be used, as already mentioned, in connection to the policy iteration algorithm, and in that case it is easily seen that one can take a time step of order $O(\min_i h_i)$. Similar results hold in the finite horizon case. The most important condition in the above theorem is that the scaled matrix might belong to the cone $\mathcal{C}(\mathcal{S})$. Before going on the characterization of $\mathcal{C}(\mathcal{S})$, we will first compare our scheme to the classical finite differences approximations.

Remark 2.1 The results of this section are close to the analysis in section 5.4.4 of [6], where consistency for an arbitrary set of transition to neighbors is discussed. The point of view of this paper is rather to fix the set $\mathcal{S}$ of the neighbors allowed to enter in the scheme, and then to characterize the class of covariance matrices for which consistency holds. We will see in sections 4 and 5 (and this is the main novelty of the paper) how to obtain an effective characterization.

Remark 2.2 It is possible to study consistency taking the point of view of the discretization of the HJB equation (1.2), the solution being defined in the sense of viscosity. Barles and Souganidis [2] give a systematic way of obtaining convergent approximation schemes for second order partial differential equations whose solution satisfies some strong uniqueness property. There approach applies to (1.2) and leads to the same conditions than those of theorem 2.1.

3 Classical finite differences approximations

Let us show that the generalized finite difference algorithm, given in the above section, is indeed a generalization of the classical finite differences approximations, that we recall now. Let Φ be a C^2 function over $\mathbb{R}^n$, and let $\varphi_k := \Phi(x_k)$ for all k. Given any $\xi \in \mathbb{Z}^n$, we can approximate the second order derivatives of Φ by the following finites differences:

$$\frac{\delta_{\xi+e_i+e_j} - \delta_{\xi+e_i} - \delta_{\xi+e_j} + \delta_\xi}{h_i h_j} \varphi_k = \Phi_{x_i x_j}(x_k) + o(1). \tag{3.24}$$

Denote the corresponding operators as follows:

$$d_{ij}^\xi := \frac{\delta_{\xi+e_i+e_j} - \delta_{\xi+e_i} - \delta_{\xi+e_j} + \delta_\xi}{h_i h_j}. \tag{3.25}$$

Viewing i (resp. j) as the first (second) coordinate, when $\xi = 0$, we call this operator d_{ij}^ξ the right upper approximation of $\Phi_{x_i x_j}$. We can similarly define left upper, right lower, and left lower approximations of $\Phi_{x_i x_j}$, by taking ξ equal to $-e_i$, $-e_j$, and $-e_i - e_j$, resp. By combining these amounts, we can define centered approximations; the corresponding operators are, along the main and second diagonals:

$$D_{ij}^+ := \tfrac{1}{2}(d_{ij}^0 + d_{ij}^{-e_i-e_j}), \qquad D_{ij}^- := \tfrac{1}{2}(d_{ij}^{-e_i} + d_{ij}^{-e_j}). \tag{3.26}$$

In other words,

$$\begin{aligned} D_{ij}^+ &= \frac{1}{2h_i h_j}(\delta_{e_i+e_j} + \delta_{-e_i-e_j} + 2\delta_0 - \delta_{e_i} - \delta_{-e_i} - \delta_{e_j} - \delta_{-e_j}), \\ D_{ij}^- &= \frac{1}{2h_i h_j}(\delta_{e_i} + \delta_{-e_i} + \delta_{e_j} + \delta_{-e_j} - \delta_{e_i-e_j} - \delta_{e_j-e_i} - 2\delta_0). \end{aligned} \tag{3.27}$$

In addition, for the approximation of diagonal second order derivatives we take the standard centered formula

$$D_{ii} := \frac{\delta_{e_i} + \delta_{-e_i} - 2\delta_0}{h_i h_i}. \tag{3.28}$$

The classical finite differences approximation of (1.2) is

$$\begin{array}{rcl} \lambda v_k & = & \inf\limits_{u\in U}\left(\ell(x_k,u) + f(x_k,u)\cdot D_k^u v_k + \frac{1}{2}\sum\limits_{i,j=1}^n a_{ij}(x_k,u) D_{ij}^{\pm} v_k\right) \\ & & \text{for all} \quad k\in\mathbb{Z}^n, \quad q\in\mathbb{N}, \\ y_k^0 & = & 0, \qquad \text{for all} \quad k\in\mathbb{Z}^n, \end{array} \tag{3.29}$$

where if $i\neq j$, $D_{ij}^{\pm}$ is equal either to D_{ij}^+ or D_{ij}^-, and D_k^u is the the upwind spatial finite difference defined in (2.11). The above scheme is equivalent to the following one:

$$\begin{array}{l} v_k := (1+\lambda\Delta t^h)^{-1} \\ \quad \inf\limits_{u\in U}\left\{ v_k + \Delta t^h \ell(x_k,u) + \Delta t^h f(x_k,u)\cdot D_k^u v_k + \frac{1}{2}\Delta t^h \sum\limits_{i,j=1}^n a_{ij}(x_k,u) D_{ij}^{\pm} v_k. \right\}. \end{array} \tag{3.30}$$

It is known that this scheme is a consistant Markov chain approximation under restrictive assumptions that we explicit now (this is a reformulation of known results; see e.g. [6] or [9]).

Lemma 3.1 *The classical finite differences approximation scheme can be interpreted as a consistent Markov chain approximation iff the following three conditions hold:*
(i) *If $i\neq j$ is such that $a_{ij}(x_k,u)\neq 0$, then $D_{ij}^{\pm} = D_{ij}^+$ if $a_{ij}(x_k,u)>0$, and $D_{ij}^{\pm} = D_{ij}^-$ if $a_{ij}(x_k,u)<0$,*
(ii) *The matrix $a^h(x_k,u)$ is dominant diagonal, or equivalently,*

$$\frac{a_{ii}(x_k,u)}{h_i} \geq \sum_{j\neq i} \frac{|a_{ij}(x_k,u)|}{h_j} \qquad \textit{for all} \quad i=1,\ldots,n, \tag{3.31}$$

(iii) *The time step Δt^h satisfies the following condition*

$$\sum_{i=1}^n \frac{|f(x_k,u)_i|}{h_i} + \sum_{i=1}^n \left(2\frac{a_{ii}(x_k,u)}{h_i^2} - \sum_{j\neq i} \frac{|a_{ij}(x_k,u)|}{h_i h_j} \right) \leq \left(\Delta t^h\right)^{-1}. \tag{3.32}$$

We now make explicit the link between the two approaches by expressing the classical finite differences approximation scheme as a Markov chain approximation scheme. If the conditions of the above lemma are satisfied, then we can write the approximation of second order terms as

$$\begin{aligned}\sum_{i,j=1}^{n} a_{ij}(x_k,u)D_{ij}^{\pm} &= \sum_{\substack{i\neq j\\ a_{ij}>0}} \frac{a_{ij}}{h_i h_j}\Delta_{e_i+e_j} - \sum_{\substack{i\neq j\\ a_{ij}<0}} \frac{a_{ij}}{h_i h_j}\Delta_{e_i-e_j} + \\ &\quad \sum_i \left(\frac{a_{ii}}{(h_i)^2} - \sum_{j\neq i}\frac{|a_{ij}|}{h_i h_j}\right)\Delta_{e_i}\end{aligned} \tag{3.33}$$

The weights of the transitions are nonnegative iff condition (ii) of lemma 3.1 is satisfied. It follows that the classical finite difference scheme is equivalent to the generalized finite difference scheme where the set $\mathcal{S}$ is equal to

$$\hat{\mathcal{S}} := \{e_1, \cdots, e_n\} \cup \{e_i \pm e_j, 1 \leq i \neq j \leq n\}.$$

We have that $\mathcal{C}(\hat{\mathcal{S}})$ is precisely the cone of dominant diagonal matrices.

4 Characterization of finitely generated cones

Let us come back to the analysis of the generalized finite difference method. In the sequel we will concentrate on characterizations of the strong consistency condition, with special attention to the case when $\mathcal{S}$ is the set $\mathcal{S}^q$ of neighboring points of order q, defined by

$$\mathcal{S}^q := \{\xi \in \mathbb{Z}^n; |\xi_i| \leq q,\ i = 1, \ldots, n\}. \tag{4.34}$$

Characterizing a finitely generated cone happens to be a classical problem of convex analysis and polyhedral combinatorics, and can be solved using the notion of polar cone. Let us recall these classical results; an excellent reference on this subject is Pulleyblank [10].

Let $\mathcal{C}$ be a nonempty closed convex cone in $\mathbb{R}^p$. The associated (positively) polar cone is

$$\mathcal{C}^* = \{x^* \in X^*\ ;\ \langle x^*, x\rangle \geq 0,\ \forall x \in \mathcal{C}\}.$$

It is known that $(\mathcal{C}^*)^* = \mathcal{C}$. Let $\mathcal{C}$ be finitely generated, say by $g_1, \ldots, g_q$. Then $\mathcal{C}^* = \bigcap_i \{x^* \in X^*\ ;\ \langle x^*, g_i\rangle \geq 0\}$. It happens that the set $\mathcal{C}^*$ is also

finitely generated, say by $g_1^*, \ldots, g_r^*$; this dual generator can be computed by a certain recursion. Since $\mathcal{C} = (\mathcal{C}^*)^*$, it follows that

$$\mathcal{C} = \{x; \ \ \langle g_i^*, x\rangle \geq 0, \ \ i = 1, \ldots, r\}.$$

This means that the cone $\mathcal{C}$ is characterized by a finite number of linear inequalities, whose coefficients can be computed.

Let us specialize this result to the case of the cone $\mathcal{C}(\mathcal{S})$. Let $\mathcal{M}$ be the set of symmetric matrices, and $\mathcal{M}_+$ be the set of symmetric definite positive matrices. Using the Frobenius scalar product $A \cdot B = \sum_{i,j} A_{ij} B_{ij}$, for which $B \cdot \xi\xi^\top = \xi^\top B \xi$, for all square $n \times n$ symmetric matrix B and n dimensional vector ξ, we have that the polar cone is

$$\mathcal{C}(\mathcal{S}^q)^* = \left\{ B \in \mathcal{M} \ ; \ \xi^\top B \xi \geq 0, \ \forall \xi \in \mathcal{S} \right\}. \tag{4.35}$$

Consider the example when $\mathcal{S} = \mathcal{S}^q$ defined in (4.34). Using the following facts: $\mathcal{C}(\mathcal{S}^q)$ is strictly increasing with q, is a subset of the cone $\mathcal{M}_+$, and $(\mathcal{M}_+)^* = \mathcal{M}_+$, we have the infinite chain of strict inclusions

$$\mathcal{C}(\mathcal{S}^1) \subset \mathcal{C}(\mathcal{S}^2) \cdots \subset \mathcal{M}_+ \subset \cdots \mathcal{C}(\mathcal{S}^2)^* \subset \mathcal{C}(\mathcal{S}^1)^*. \tag{4.36}$$

It can be noticed that, since the cone $\mathcal{C}(\mathcal{S}^q)$ contains every nonnegative diagonal matrices, each element of its dual has a nonnegative diagonal.

An important observation is that $\mathcal{S}^q$, and therefore also $\mathcal{C}(\mathcal{S}^q)$, are invariant through the linear transformations in $\mathbb{R}^n$ that correspond to a permutation of coordinates, and also to the change of sign of coordinates. The permutation of coordinates i and j of $\xi \in \mathbb{R}^n$ result in the permutation of elements of $\xi\xi^\top$ of coordinates (i,k) and (j,k), and (k,i) and (k,j), for all k, while changing the sign of ξ_i results in changing the sign of elements of $\xi\xi^\top$ of coordinates (i,j), for $j \neq i$. Since these transformations are self adjoint, for each $B \in \mathcal{C}(\mathcal{S}^q)^*$, the matrices obtained by the same (adjoint) transformations (so that the scalar product with B remains invariant) also belong to $\mathcal{C}(\mathcal{S}^q)^*$. In particular, a generator of $\mathcal{C}(\mathcal{S}^q)^*$ can be partitioned into classes of equivalence corresponding to the above mentioned transformations. This allows to give a compact description of the set of generators.

5 Specific examples

We have performed the computation of generators of dual cones using the Qhull algorithm by Barbet et al. [1]. The latter computes, given a finite set

in $\mathbb{R}^m$, a minimal set of linear inequalities characterizing its convex hull. This computation is made using the floating point arithmetic of the C language. However, the risk of numerical errors due to the floating point arithmetic is limited, since we were able to compute a scaling of the data for which all coefficients are small integers, up to an absolute precision of 10^{-10}.

The link between the convex hull of a finite set and the generator of a dual cone is as follows. Consider a generator $g_1, \ldots, g_n$, and set $g_0 := 0$. Then compute a minimal characterization of the convex hull of $g_0, \ldots, g_n$, of the form $\langle g_i^*, \cdot \rangle \geq b_i$, $i = 1, \ldots, r$. A minimal generator of the dual cone is given by the homogeneous inequalities, i.e., the dual cone is

$$\{g \in \mathbb{R}^m; \ \ \langle g_i^*, g \rangle \geq b_i, \ \ \ i \in I\}$$

with $I := \{1 \leq i \leq r; \ \ b_i = 0\}$.

Our actual computations deal with spaces of symmetric matrices of size n. Each of them can be represented by its upper triangular part, and thus is viewed as an element of $\mathbb{R}^m$, $m = \frac{1}{2}n(n+1)$; in particular, $m = 3$, 6, and 10, for $n = 2$, 3, and 4, respectively.

Once a generator of the dual cone has been obtained, it remains to identify the classes of equivalence (defined in the previous section) in order to obtain compact expression. This was done by sorting the elements following the (ordered) weights of diagonal elements (the latter being, as we already know, nonnegative). It appears that this suffices for identifying the equivalence classes, as can be checked by generating them using the formulas given below and comparing both sets.

Dimension 2. In the case $n = 2$, we computed characterizations of the sets $\mathcal{C}(\mathcal{S}^q)$, $q = 1$ to 10. We display detailed results for $q = 1$ to 7. The set $\mathcal{C}(\mathcal{S}^1)$ is characterized by 4 constraints and 1 equivalence class:

$$a_{ii} \geq |a_{ij}|, \quad 1 \leq i \neq j \leq 2.$$

The set $\mathcal{C}(\mathcal{S}^2)$ is characterized by 8 constraints and 2 equivalence classes:

$$\left\{ \begin{array}{l} 2a_{ii} \geq |a_{ij}| \\ 2a_{ii} + a_{jj} \geq 3|a_{ij}| \end{array} \right.$$

for $1 \leq i \neq j \leq 2$. The set $\mathcal{C}(\mathcal{S}^3)$ is characterized by 16 constraints and 4 equivalence classes:

$$\begin{cases} 3a_{ii} \geq |a_{ij}| \\ 3a_{ii} + 2a_{jj} \geq 5|a_{ij}| \\ 6a_{ii} + a_{jj} \geq 5|a_{ij}| \\ 6a_{ii} + 2a_{jj} \geq 7|a_{ij}| \end{cases}$$

for $1 \leq i \neq j \leq 2$. The set $\mathcal{C}(\mathcal{S}^4)$ is characterized by 24 constraints and 6 equivalence classes:

$$\begin{cases} 4a_{ii} \geq |a_{ij}| \\ 4a_{ii} + 3a_{jj} \geq 7|a_{ij}| \\ 6a_{ii} + a_{jj} \geq 5|a_{ij}| \\ 6a_{ii} + 2a_{jj} \geq 7|a_{ij}| \\ 12a_{ii} + a_{jj} \geq 7|a_{ij}| \\ 12a_{ii} + 6a_{jj} \geq 17|a_{ij}| \end{cases}$$

for $1 \leq i \neq j \leq 2$. The set $\mathcal{C}(\mathcal{S}^5)$ is characterized by 40 constraints and 10 equivalence classes:

$$\begin{cases} 5a_{ii} \geq |a_{ij}| \\ 5a_{ii} + 4a_{jj} \geq 9|a_{ij}| \\ 10a_{ii} + 2a_{jj} \geq 9|a_{ij}| \\ 10a_{ii} + 3a_{jj} \geq 11|a_{ij}| \\ 12a_{ii} + a_{jj} \geq 7|a_{ij}| \\ 12a_{ii} + 6a_{jj} \geq 17|a_{ij}| \\ 15a_{ii} + 2a_{jj} \geq 11|a_{ij}| \\ 15a_{ii} + 6a_{jj} \geq 19|a_{ij}| \\ 20a_{ii} + a_{jj} \geq 9|a_{ij}| \\ 20a_{ii} + 12a_{jj} \geq 31|a_{ij}| \end{cases}$$

for $1 \leq i \neq j \leq 2$. The set $\mathcal{C}(\mathcal{S}^6)$ is characterized by 48 constraints and 12

equivalence classes:

$$\begin{cases} 6_{ii} \geq |a_{ij}| \\ 6a_{ii} + 5a_{jj} \geq 11|a_{ij}| \\ 10a_{ii} + 2a_{jj} \geq 9|a_{ij}| \\ 10a_{ii} + 3a_{jj} \geq 11|a_{ij}| \\ 12a_{ii} + a_{jj} \geq 7|a_{ij}| \\ 12a_{ii} + 6a_{jj} \geq 17|a_{ij}| \\ 15a_{ii} + 2a_{jj} \geq 11|a_{ij}| \\ 15a_{ii} + 6a_{jj} \geq 19|a_{ij}| \\ 20a_{ii} + a_{jj} \geq 9|a_{ij}| \\ 20a_{ii} + 12a_{jj} \geq 31|a_{ij}| \\ 30a_{ii} + a_{jj} \geq 11|a_{ij}| \\ 30a_{ii} + 20a_{jj} \geq 49|a_{ij}| \end{cases}$$

or $1 \leq i \neq j \leq 2$. The set $\mathcal{C}(\mathcal{S}^7)$ is characterized by 72 constraints and 18 equivalence classes:

$$\begin{cases} 7a_{ii} \geq |a_{ij}| \\ 7a_{ii} + 6a_{jj} \geq 13|a_{ij}| \\ 14a_{ii} + 3a_{jj} \geq 13|a_{ij}| \\ 14a_{ii} + 4a_{jj} \geq 15|a_{ij}| \\ 15a_{ii} + 2a_{jj} \geq 11|a_{ij}| \\ 15a_{ii} + 6a_{jj} \geq 19|a_{ij}| \\ 20a_{ii} + a_{jj} \geq 9|a_{ij}| \\ 20a_{ii} + 12a_{jj} \geq 31|a_{ij}| \\ 21a_{ii} + 2a_{jj} \geq 13|a_{ij}| \\ 21a_{ii} + 10a_{jj} \geq 29|a_{ij}| \\ 28a_{ii} + 2a_{jj} \geq 15|a_{ij}| \\ 28a_{ii} + 15a_{jj} \geq 41|a_{ij}| \\ 30a_{ii} + a_{jj} \geq 11|a_{ij}| \\ 30a_{ii} + 20a_{jj} \geq 49|a_{ij}| \\ 35a_{ii} + 6a_{jj} \geq 29|a_{ij}| \\ 35a_{ii} + 12a_{jj} \geq 41|a_{ij}| \\ 42a_{ii} + a_{jj} \geq 13|a_{ij}| \\ 42a_{ii} + 30a_{jj} \geq 71|a_{ij}| \end{cases}$$

Dimension 3. When $n = 3$, we computed characterizations of the sets $\mathcal{C}(\mathcal{S}^q)$, $q = 1$ to 2. The set $\mathcal{C}(\mathcal{S}^1)$ is characterized by

$$\begin{cases} a_{ii} \geq |a_{ij}| \\ a_{ii} + a_{jj} \geq (-1)^p a_{ik} + (-1)^q a_{jk} + 2(-1)^{p+q+1} a_{ij} \end{cases}$$

for $i \neq j \neq k$ and $p, q \in \{1, 2\}$. As was expected, this cone is larger than the cone of dominant diagonal matrices. The set $\mathcal{C}(\mathcal{S}^2)$ is characterized by

$$\begin{cases} 2a_{ii} \geq |a_{ij}| \\ 2a_{ii} + a_{jj} \geq 3|a_{ij}| \\ 2a_{ii} + 2a_{jj} \geq 4(-1)^p a_{ij} + (-1)^q a_{jk} - (-1)^{p+q} a_{ik} \\ 2a_{ii} + 2a_{jj} + a_{kk} \geq 4(-1)^p a_{ij} + 3(-1)^q a_{jk} - 3(-1)^{p+q} a_{ik} \\ 3a_{ii} + 2a_{jj} + 2a_{kk} \geq 5(-1)^p a_{ij} + 4(-1)^q a_{jk} - 5(-1)^{p+q} a_{ik} \\ 6a_{ii} + a_{jj} + a_{kk} \geq 5(-1)^p a_{ij} + 2(-1)^q a_{jk} - 5(-1)^{p+q} a_{ik} \\ 6a_{ii} + 2a_{jj} + a_{kk} \geq 7(-1)^p a_{ij} + 3(-1)^q a_{jk} - 5(-1)^{p+q} a_{ik} \\ 6a_{ii} + 2a_{jj} + 2a_{kk} \geq 7(-1)^p a_{ij} + 4(-1)^q a_{jk} - 7(-1)^{p+q} a_{ik} \\ 8a_{ii} + 2a_{jj} \geq 8(-1)^p a_{ij} + (-1)^q a_{jk} - 2(-1)^{p+q} a_{ik} \\ 8a_{ii} + 2a_{jj} + a_{kk} \geq 8(-1)^p a_{ij} + 3(-1)^q a_{jk} - 6(-1)^{p+q} a_{ik} \\ 8a_{ii} + 3a_{jj} + 2a_{kk} \geq 10(-1)^p a_{ij} + 5(-1)^q a_{jk} - 8(-1)^{p+q} a_{ik} \\ 8a_{ii} + 6a_{jj} + 2a_{kk} \geq 14(-1)^p a_{ij} + 7(-1)^q a_{jk} - 8(-1)^{p+q} a_{ik} \\ 12a_{ii} + 2a_{jj} + a_{kk} \geq 10(-1)^p a_{ij} + 3(-1)^q a_{jk} - 7(-1)^{p+q} a_{ik} \\ 12a_{ii} + 4a_{jj} + a_{kk} \geq 14(-1)^p a_{ij} + 4(-1)^q a_{jk} - 7(-1)^{p+q} a_{ik} \\ 12a_{ii} + 6a_{jj} + 2a_{kk} \geq 17(-1)^p a_{ij} + 7(-1)^q a_{jk} - 10(-1)^{p+q} a_{ik} \\ 12a_{ii} + 6a_{jj} + 4a_{kk} \geq 17(-1)^p a_{ij} + 10(-1)^q a_{jk} - 14(-1)^{p+q} a_{ik} \\ 18a_{ii} + 2a_{jj} + a_{kk} \geq 12(-1)^p a_{ij} + 3(-1)^q a_{jk} - 9(-1)^{p+q} a_{ik} \\ 18a_{ii} + 8a_{jj} + a_{kk} \geq 24(-1)^p a_{ij} + 6(-1)^q a_{jk} - 9(-1)^{p+q} a_{ik} \\ 18a_{ii} + 10a_{jj} + 2a_{kk} \geq 27(-1)^p a_{ij} + 9(-1)^q a_{jk} - 12(-1)^{p+q} a_{ik} \end{cases}$$

Dimension 4. When $n = 4$, the set $\mathcal{C}(\mathcal{S}^1)$ is characterized by

$$\begin{cases} a_{ii} & \geq |a_{ij}| \\ a_{ii} + a_{jj} & \geq (-1)^p a_{ik} + (-1)^q a_{jk} - 2(-1)^{p+q} a_{ij} \\ a_{ii} + a_{jj} + a_{kk} & \geq (-1)^p a_{il} + (-1)^q a_{jl} + (-1)^r a_{kl} \\ & \quad -2(-1)^{p+q} a_{ij} - 2(-1)^{p+r} a_{ik} - 2(-1)^{q+r} a_{jk} \\ 2a_{ii} + a_{jj} + a_{kk} + a_{ll} & \geq 3(-1)^p a_{ij} + 3(-1)^q a_{ik} + 3(-1)^r a_{il} \\ & \quad -2(-1)^{p+q} a_{jk} - 2(-1)^{p+r} a_{jl} - 2(-1)^{q+r} a_{kl} \\ 4a_{ii} + a_{jj} + a_{kk} & \geq 2(-1)^p a_{il} + (-1)^q a_{jl} + (-1)^r a_{kl} \\ & \quad -4(-1)^{p+q} a_{ij} - 4(-1)^{p+r} a_{ik} - 2(-1)^{q+r} a_{jk} \\ 4a_{ii} + 2a_{jj} + a_{kk} + a_{ll} & \geq 6(-1)^p a_{ij} + 4(-1)^q a_{ik} + 4(-1)^r a_{il} \\ & \quad -3(-1)^{p+q} a_{jk} - 3(-1)^{p+r} a_{jl} - 2(-1)^{q+r} a_{kl} \end{cases}$$

for $i \neq j \neq k \neq l$ and $p, q, r \in \{1, 2\}$.

Summary of results. The following table summarizes the various steps of our calculation, and highlights the importance of reduction of constraints using the classes of equivalence.

Here $\mathcal{S}^*$ is the set of matrices in $\mathcal{S}$ of trace not greater than 1.

n	q	size of generator of primal cone	# of constraints defining $\mathcal{S}^*$	# of constraints defining $\mathcal{C}$	# of classes of equivalence
2	1	4	6	4	1
2	2	8	13	8	2
2	3	16	27	16	4
2	4	24	39	24	6
2	5	40	67	40	10
2	6	48	87	48	12
2	7	72	123	72	18
2	8	88	159	88	22
2	9	112	203	112	28
2	10	128	239	128	32
3	1	13	31	24	2
3	2	49	563	372	19
4	1	40	476	328	6

6 Discussion of results

In this paper we have worked in the framework of "Markov chain approximations" discussed in Kushner and Dupuis [6]. Our main result is a method for computing a characterization of the class of covariant matrices that are strongly consistent with an a priori choice of neighboring points to which transitions are allowed. In the computation and display of the results, we use in an essential way the property of invariance of these cones with respect to some transformations. Although we are looking for linear inequalities with integer coefficients, the computations were made in floating-point arithmetic. However, once properly scaled, the results are up to a precision of 10^{-10} equal to very small integers, as may be seen in the tables of the previous section, and hence it seems that these results are exact, despite the fact that our method is not a mathematical proof (we tried an exact approach based on computer algebra, but without success, since many singularities were encountered). So, we have performed the computations, giving explicit results, for dimensions of the state space between 2 and 4, and when only a limited number of neighboring points are allowed (which is a highly desirable feature). This said, it seems that we have performed the computations for essentially all cases for which the numerical resolution of the stochastic HJB equation is of reasonable complexity. Indeed, when the number of linear inequalities characterizing strongly consistent matrices is large, we may expect that computing the coefficient of the algorithm will be expensive.

On the other hand, our results are only a preliminary step towards an efficient numerical algorithm. There are two main difficulties. The first is to design fast algorithms for computing the coefficients $\alpha^u_{k,\xi}$. The latter are, by the definition, solution of a linear programming problem, but using a linear programming solver for each control, at each point of the grid would be inefficient. The second difficulty is to deal with the case when consistency does not hold, e.g. by approximating the matrix $a(x_k, u)$ by a consistent matrix and then performing an error analysis. We are now pursuing some research in these directions.

Acknowledgments. We thank the three anonymous referees for their useful remarks.

References

[1] C.B. Barber, D.P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions for Mathematical Software*, 22:469–483, 1996.

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

[3] S.L. Chung, F.B. Hanson, and H.H. Xu. Parallel stochastic dynamic programming: finite element methods. *Linear Algebra and its Applications*, 172, 1992.

[4] W. H. Fleming and R. Rishel. *Deterministic and stochastic optimal control*, volume 1 of *Applications of mathematics*. Springer, New York, 1975.

[5] W. H. Fleming and H.M. Soner. *Controlled Markov processes and viscosity solutions*. Springer, New York, 1992.

[6] H. J. Kushner and P. G. Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24 of *Applications of mathematics*. Springer, New York, 2001. Second edition.

[7] J. L. Lions and A. Bensoussan. *Application des inéquations variationnelles en contrôle stochastique*, volume 6 of *Méthodes mathématiques de l'informatique*. Dunod, Paris, 1978.

[8] P.L. Lions. Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. Part 2: viscosity solutions and uniqueness. *Communications in partial differential equations*, 8:1220–1276, 1983.

[9] P.L. Lions and B. Mercier. Approximation numérique des équations de Hamilton-Jacobi-Bellman. *RAIRO Analyse numérique*, 14:369–393, 1980.

[10] W.R. Pulleyblank. Polyhedral combinatorics. In G.L. Nemhauser et al., editor, *Optimization*. Elsevier, Amsterdam, 1989.