Consistency of a simple multidimensional scheme for Hamilton-Jacobi-Bellman equations

Rémi Munos [a], Hasnaa Zidani [b]

[a] *Centre de mathématiques appliquées, École polytechnique, 91128 Palaiseau cedex, France*

[b] *Laboratoire de mathématiques appliquées, ENSTA, 32, boulevard Victor, 75739 Paris cedex 15, France*

Received 16 September 2004; accepted after revision 3 February 2005

Presented by Pierre-Louis Lions

Abstract

This Note presents an approximation scheme for second-order Hamilton-Jacobi-Bellman equations arising in stochastic optimal control. The scheme is based on a Markov chain approximation method. It is easy to implement in any dimension. The consistency of the scheme is proved, which guarantees its convergence. *To cite this article: R. Munos, H. Zidani, C. R. Acad. Sci. Paris, Ser. I 340 (2005).*

Résumé

Consistance d'un schéma multidimensionnel simple pour les équations de Hamilton-Jacobi-Bellman. Cette Note présente un schéma d'approximation pour les équations de Hamilton-Jacobi-Bellman qui apparaissent en contrôle optimal stochastique. Le schéma est construit selon une méthode d'approximation par chaîne de Markov. Il s'implémente facilement en n'importe quelle dimension. La consistance du schéma est prouvée, ce qui garantit sa convergence. *Pour citer cet article : R. Munos, H. Zidani, C. R. Acad. Sci. Paris, Ser. I 340 (2005).*

1. Introduction

We consider a multidimensional controlled Markov diffusion on $X = \mathbb{R}^n$ $(n \geq 1)$

$$x(t) = x + \int_0^t f(x(s), u(s))\, ds + \int_0^t \sigma(x(s), u(s))\, dW_s, \tag{1}$$

where $(W_t)_{0 \leq t \leq \infty}$ is a standard Brownian motion in $\mathbb{R}^n$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq \infty}, \mathbb{P})$, with the usual assumptions on the filtration $(\mathcal{F}_t)_{0 \leq t \leq \infty}$. The control $u(t)$ is a Lebesgue measurable function with values in a compact set $U \subset \mathbb{R}^m$, $m \geq 1$. We consider an infinite horizon discounted payoff

Email addresses: `remi.munos@polytechnique.fr` (Rémi Munos), `hasnaa.zidani@ensta.fr` (Hasnaa Zidani).

$J(x;u(\cdot)) = \mathbb{E}\left[\int_0^\infty e^{-\beta t} l(x(t),u(t))dt\right]$, with l the cost function and $\beta > 0$ the discount factor. The value function is the minimum value of the payoff

$$V(x) = \inf_{u(\cdot)} J(x;u(\cdot)). \tag{2}$$

The Hamilton-Jacobi-Bellman (HJB) equation associated with the optimal control problem is

$$-\beta v(x) + \min_{u\in U}\Big[\sum_{i=1}^{n} f_i(x,u)\partial_{x_i} v(x) + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x,u)\partial^2_{x_i,x_j} v(x) + l(x,u)\Big] = 0, \quad \text{for } x \in X$$

where $a(x,u) = \sigma(x,u)\sigma(x,u)'$ is the covariance matrix.

In this Note, we present a simple numerical scheme for the HJB equation on a grid X_h (h being the grid resolution) based on the method of Markov chain approximation [4] and prove the consistency property. This guarantees the convergence of the discrete solutions V^h of the dynamic programming equation (5) to the value function V defined by (2). It is known [4,5] that when the covariance matrix a is diagonal dominant, one can use the classical finite difference scheme which has the advantage to be convergent and easy to implement. When a is not diagonal dominant, the generalized finite difference (GFD) scheme [1] provides a consistent scheme, but requires (for each grid point) the computation of the appropriate stencil of grid points entering into the scheme.

In two dimensions, the implementation of the GFD scheme is inexpensive (see [2] for a fast algorithm). However, in higher dimensions ($n \geq 3$), this scheme is not easy to implement for a general covariance matrix. On the other hand, several works deal with the nonlocal but consistent Semi-Lagrangian scheme (see for example [3]). In such a scheme, the state is approximated by a discrete process:

$$y_{t+\tau} = y_t + \tau f(y_t,u) + \sqrt{\tau}\sigma_{j_t}\xi_t, \tag{3}$$

where τ is a time step, σ_j denotes the jth-column of σ, $j_t \in \{1,\dots,n\}$ and $\xi_t \in \{-\frac{1}{2}, +\frac{1}{2}\}$ are sequences of i.i.d. uniform random variables. Here, we consider rather the following approximation:

$$y_{t+\tau} = y_t + \tau f(y_t,u) + \eta_{j_t} v_{j_t}\xi_t,$$

where j_t and ξ_t are *non-uniform* random variables, and v_j are the eigenvectors of the covariance matrix. The diffusion steps η_j are chosen in order to obtain a consistent scheme, yet they are not necessarily equal. This scheme is also nonlocal: the diffusion steps being, for a typical implementation, of order h^α, with $\alpha < 1$. The scheme is easy to implement and provides an interesting alternative to Finite-Differences approaches. Moreover, it is built on unstructured grids.

Another appealing feature is the possible utilization of the balance conditions (6) and (7) between the weights and the diffusion steps to design boundary conditions while preserving local consistency. This point is currently under investigation and numerical experiments for comparison with the scheme (3) will be the object of future work.

2. Markov Decision Process approximation

Consider the grid $X_h \subset X$, where h is the *resolution of the grid* (i.e. for all $x \in X$, there exists $x_i \in X_h$ such that $||x - x_i|| \leq h$). For each grid point $x_i \in X_h$ and control $u \in U$, define a *discretization time-step* $\tau(x_i,u)$ and the *drifted point* $y(x_i,u) = x_i + \tau(x_i,u)f(x_i,u)$. Write $\{\alpha_j(x_i,u)\}_{1\leq j\leq q}$ (with $q \leq n$) the eigenvalues of $a(x_i,u)$ that are (strictly) positive and $\{v_j(x_i,u)\}_{1\leq j\leq q}$ the corresponding normalized eigenvectors. Since $a(x_i,u)$ is a positive semi-definite matrix, one may write

$$a(x_i,u) = \sum_{j=1}^{q} \alpha_j(x_i,u)v_j(x_i,u)v_j'(x_i,u). \tag{4}$$

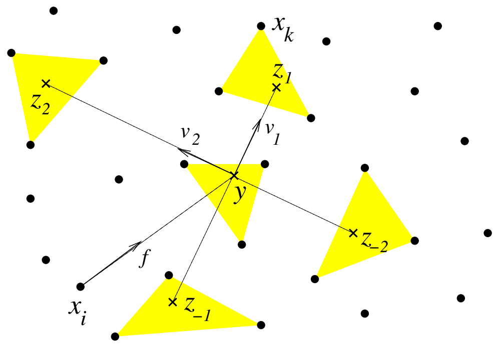

Figure 1. Construction of the MDP approximation scheme. Here $q = n = 2$.

Let us introduce $2q$ positive values, called *diffusion steps*, $\{\eta_j(x_i, u)\}_{1\leq j\leq q}$ and $_{-q\leq j\leq -1}$ which define the *diffused points* $\{z_j(x_i, u)\}_{-q\leq j\leq q}$ (see Fig. 1):

$$z_j(x_i,u) = \begin{cases} y(x_i,u) + \eta_j(x_i,u)v_j(x_i,u) & \text{for} \quad 1 \leq j \leq q \\ y(x_i,u) & \text{for} \qquad j = 0 \\ y(x_i,u) - \eta_j(x_i,u)v_{-j}(x_i,u) & \text{for} \ -q \leq j \leq -1 \end{cases}$$

Now, consider a grid-based *local linear interpolation process*: for any state x, there exists a finite set of positive coefficients $\{\lambda(x_k|x)\}$ (where $x_k \in X_h$) such that $\sum_k \lambda(x_k|x) = 1$, $\sum_k \lambda(x_k|x)x_k = x$, and such that the grid points x_k whose coefficient $\lambda(x_k|x)$ are strictly positive are at a distance $O(h)$ from x.

Examples of local linear interpolation processes:

- *Piecewise linear interpolation on a triangulation.* The $\{\lambda(x_k|x)\}$ are the barycentric coordinates of x in the simplex $\mathcal{T} \ni x$. Thus, there are at most $n+1$ non-zero coefficients $\{\lambda(x_k|x)\}$ corresponding to the vertices $\{x_k\}$ of $\mathcal{T}$.
- *Piecewise Multi-linear interpolation on a grid.* The $\{\lambda(x_k|x)\}$ are the multi-linear interpolation coefficients of x in the n-dimensional rectangle $\mathcal{R} \ni x$. There are at most 2^n non-zero such coefficients.

An example of an efficient local linear interpolation process in any dimension, consisting of a Coxeter-Freudenthal-Kuhn triangulation embedded in a *kd-tree*, is described in [6].

Now, consider some positive *weights* $\{\rho_j(x_i, u)\}_{-q\leq j\leq q}$ which sum to one (i.e. $\sum_{j=-q}^{q} \rho_j(x_i, u) = 1$).

The *Markov Decision Process approximation* is then defined by the state space X_h, the control space U, the cost function $c(x_i, u) = l(x_i, u)\tau(x_i, u)$, and the transition probabilities

$$p(x_k|x_i,u) = \sum_{j=-q}^{q} \rho_j(x_i,u)\lambda(x_k|z_j(x_i,u)).$$

The corresponding dynamic programming equation is:

$$V^h(x_i) = \min_{u\in U}[e^{-\beta\tau(x_i,u)} \sum_k p(x_k|x_i,u)V^h(x_k) + c(x_i,u)] \tag{5}$$

Proposition 2.1 *Define $\tau_{h,\min}$ and $\tau_{h,\max}$ such that for all h, for all $x_i \in X_h, u \in U$, $\tau_{h,\min} \leq \tau(x_i,u) \leq \tau_{h,\max}$. Assume that the weights and diffusion steps satisfy the balance conditions: for all $x_i \in X_h, u \in U$, for all $1 \leq j \leq q$,*

$$\rho_j(x_i,u)\eta_j(x_i,u) = \rho_{-j}(x_i,u)\eta_{-j}(x_i,u), \tag{6}$$

$$\rho_j(x_i,u)\eta_j^2(x_i,u) + \rho_{-j}(x_i,u)\eta_{-j}^2(x_i,u) = \alpha_j(x_i,u)\tau(x_i,u). \tag{7}$$

Then, under the CFL condition $h^2 = o(\tau_{h,\min})$ and assuming $\tau_{h,\max} = O(1)$, this approximation scheme is consistent in the sense of Kushner and Dupuis ([4], p. 71). More precisely,

$$\mathbb{E}_{x_i}[x_k - x_i] = \tau(x_i,u)f(x_i,u) \tag{8}$$

$$\mathbb{E}_{x_i}\left[x_k - \mathbb{E}_{x_i}[x_k]\right]\left[x_k - \mathbb{E}_{x_i}[x_k]\right]' = \tau(x_i,u)a(x_i,u) + o(\tau_{h,\min}) \tag{9}$$

Proof. In what follows we use simplified notation, writing f, a, τ, y, z_j, η_j, α, v_j, and ρ_j instead of $f(x_i,u)$, $a(x_i,u)$, $\tau(x_i,u)$, $y(x_i,u)$, $z_j(x_i,u)$, $\eta_j(x_i,u)$, $\alpha(x_i,u)$, $v_j(x_i,u)$, and $\rho_j(x_i,u)$.

Property (8) derives from the definition of the interpolation coefficients:

$$\begin{aligned}\mathbb{E}_{x_i}[x_k - x_i] &= \textstyle\sum_k p(x_k|x_i,u)(x_k - x_i) = \sum_{j=-q}^{q} \rho_j \sum_k \lambda(x_k|z_j)(x_k - x_i) \\ &= \textstyle\sum_{j=-q}^{q} \rho_j \Big[\sum_k \lambda(x_k|z_j)(x_k - z_j) + z_j - y + y - x_i \Big] = y - x_i = \tau f\end{aligned}$$

since $\sum_k \lambda(x_k|z_j)x_k = z_j$ and $\sum_{j=-q}^{q} \rho_j z_j = y + \sum_{j=1}^{q}(\rho_j\eta_j - \rho_{-j}\eta_{-j}) = y$ from (6).

Property (9) follows from the decomposition

$$\begin{aligned}\mathbb{E}_{x_i}\left[x_k - \mathbb{E}_{x_i}[x_k]\right]\left[x_k - \mathbb{E}_{x_i}[x_k]\right]' &= \textstyle\sum_k p(x_k|x_i,u)(x_k - y)(x_k - y)' \\ &= \textstyle\sum_{j=-q}^{q} \rho_j \left[\sum_k \lambda(x_k|z_j)(x_k - z_j)(x_k - z_j)' + (z_j - y)(z_j - y)' \right]\end{aligned}$$

because $\sum_{j=-q}^{q} \rho_j \sum_k \lambda(x_k|z_j)(x_k - z_j)(z_j - y)' = 0$.

Since the interpolation process is local, i.e. for all k such that $\lambda(x_k|z_j) > 0$, $x_k - z_j = O(h)$, we deduce that $\sum_{j=-q}^{q} \rho_j \sum_k \lambda(x_k|z_j)(x_k - z_j)(x_k - z_j)' = O(h^2)$.

Now, by noticing that $(z_j - y)(z_j - y)' = \eta_j^2 v_j v_j'$, from (4) and (7) we deduce that

$$\textstyle\sum_{j=-q}^{q} \rho_j (z_j - y)(z_j - y)' = \sum_{j=1}^{q}(\rho_j\eta_j^2 + \rho_{-j}\eta_{-j}^2)v_j v_j' = \sum_{j=1}^{q} \alpha_j \tau v_j v_j' = \tau a.$$

Thus, from the CFL condition, we have

$$\mathbb{E}_{x_i}\left[x_k - \mathbb{E}_{x_i}[x_k]\right]\left[x_k - \mathbb{E}_{x_i}[x_k]\right]' = a(x_i,u)\tau + o(\tau_{h,\min}). \quad \square$$

Remark 1 Here are some specific cases for which the conditions (6) and (7) hold :

- *The weights are constant $\rho_j = \rho = 1/(2q)$ for $j \neq 0$ and $\rho_0 = 0$. Thus the steps $\eta_j = \sqrt{\alpha_j \tau q}$.*
- *The steps η_j are constant $\eta_j = \eta$. If we choose $\rho_0 = 0$ then for $j \geq 1$ the weights $\rho_j = \alpha_j \frac{\tau}{2\eta^2}$. Since they sum to one, we deduce that ρ_j is proportional to α_j: for $j \geq 1$, $\rho_j = \alpha_j/(2\sum_{i=1}^{q}\alpha_i)$.*

References

[1] F. Bonnans, H. Zidani, Consistency of generalized finite difference schemes for the stochastic HJB equation, SIAM J. Numer. Anal. 41 (3) (2003) 1008–1021.

[2] F. Bonnans, E. Ottenwaelter, H. Zidani, A fast algorithm for the two dimensional HJB equation of stochastic control, Math. Model. Numer. Anal. 38 (4) (2004) 723–735.

[3] F. Camilli, M. Falcone, An approximation scheme for the optimal control of diffusion processes, Math. Model. Numer. Anal. 29 (1) (1995) 97–122.

[4] H.J. Kushner, P.G. Dupuis, Numerical methods for stochastic control problems in continuous time, 2nd ed., Appl. math. 24, Springer-Verlag, New York, 2001.

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

[6] R. Munos, A. Moore, Variable Resolution Discretization in Optimal Control, Mach. Learn. 49 (2-3) (2002) 291–323.