CONVERGENCE OF NUMERICAL METHOD FOR MULTISTATE STOCHASTIC DYNAMIC PROGRAMMING *

F. B. HANSON** and K. NAIMIPOUR***

***Laboratory for Advanced Computing (M/C 249), University Illinois at Chicago, Chicago, IL 60680-4348*

****Dept. Computer Science, Northeastern Illinois University, 5500 N. St. Louis Avenue, Chicago, IL 60625-4699*

Abstract. Convergence of corrections is examined for a predictor-corrector method to solve Bellman equations of multi-state stochastic optimal control in continuous time. Quadratic costs and constrained control are assumed. A heuristically linearized comparison equation makes the nonlinear, discontinuous Bellman equation amenable to linear convergence analysis. Convergence is studied using the Fourier stability method. A uniform mesh ratio type condition for the convergence is results. The results are valid for both Gaussian and Poisson type stochastic noise. The convergence criteria has been extremely useful for solving the larger multi-state problems on vector supercomputers and massively parallel processors.

Key Words. Convergence of numerical methods; computational methods; dynamic programming; Markov processes; multivariable control systems; partial differential equations; feedback control; stochastic control; predictor-corrector methods

1 INTRODUCTION

For stochastic dynamic programming with both uncorrelated Gaussian and discrete Poisson noise, the Bellman partial differential equation in the quadratic cost case has the following form (Hanson, 1991; Gihman and Skorohod, 1979),

$$
\begin{aligned}
0 = {} & \frac{\partial V^*}{\partial t} + \mathcal{L}(\mathbf{x}, t) \qquad (1)\\
\equiv {} & \frac{\partial V^*}{\partial t} + \tfrac{1}{2}\text{Trace}\left[GG^T(\mathbf{x},t)\nabla_x \nabla_x^T V^*\right] \\
+ {} & \sum_{l=1}^{q} \lambda_l \cdot \left[V^*(\mathbf{x}+\mathbf{H}_l(\mathbf{x},t),t) - V^*(\mathbf{x},t)\right] \\
+ {} & C_0 + \mathbf{F}_0^T \nabla_x V^* + \tfrac{1}{2}(\mathbf{U}^* - 2\mathbf{U}_R)^T C_2 \mathbf{U}^*
\end{aligned}
$$

where V^* is the optimal value, t is time, $\mathbf{x}$ is an m-dimensional state vector, $\mathbf{u}(\mathbf{x},t)$ is the n-dimensional optimal feedback control vector, GG^T is the the $n \times r$ Gaussian noise amplitude coefficient, $\mathbf{H}_l$ is the lth column vector of the $n \times q$ Poisson noise amplitude with rate λ_l, $\{\mathbf{F}_0, F_1\}(\mathbf{x},t)$ are the coefficients of the control-linearized nonlinearity function,

$$\mathbf{F}(\mathbf{x},t,\mathbf{u}) = \mathbf{F}_0(\mathbf{x},t) + F_1(\mathbf{x},t)\mathbf{u}, \qquad (2)$$

$\{C_0, \mathbf{C}_1, C_2\}(\mathbf{x},t)$ are the coefficients of the quadratic cost term,

$$
\begin{aligned}
C(\mathbf{x},t,\mathbf{u}) &= C_0(\mathbf{x},t) + \mathbf{C}_1^T(\mathbf{x},t)\mathbf{u} \\
&+ \tfrac{1}{2}\mathbf{u}^T C_2(\mathbf{x},t)\mathbf{u}, \qquad (3)
\end{aligned}
$$

$\mathbf{U}_R(\mathbf{x},t)$ is the regular (unconstrained) control,

$$\mathbf{U}_R(\mathbf{x},t) = -C_2^{-1} \cdot (\mathbf{C}_1 + F_1^T \nabla_x V^*), \qquad (4)$$

and $\mathbf{U}^*(\mathbf{x},t)$ is the optimal (constrained) control, subject to the control constraints.

In the case of rectangular or hypercube constraints,

$$U_{min,i} \le u_i(\mathbf{x},t) \le U_{max,i}, \qquad (5)$$

for $i = 1$ to n, the optimal control has the components,

$$U_i^* = \min\left[U_{max,i}, \max[U_{min,i}, U_{R,i}]\right]. \qquad (6)$$

It should be noted that the Bellman equation (1) with quadratic costs (3) has corresponding quadratic nonlinearities in the shadow price vector $\nabla_x V^*$ since the regular control (4) is linear in the shadow price, the optimal control is at most linear in $\mathbf{U}_R$, and there are terms of (1) that are quadratic in the optimal control.

the backward nature of the PDE (1) with respect to time means that a final condition, such as $V(\mathbf{x},t_f) = V_f(\mathbf{x})$ in the case of a given salvage value, must be satisfied. The PDE is a genuine final, rather than initial, value problem and integration is in the direction of increasing time-to-go $(t_f - t)$.

Bellman equations such as (1) arise in multistate resource management problems in disastrous environments

* This is a preprint of the paper published in "Proceedings of Int. Federation of Automatic Control 12th World Congress," vol. 9, pp. 501-504, July 1993.

considered by Hanson (1987) and Pratico et al. (1992). However, other possible applications are aerospace dynamics in disastrous weather, financial market crashes and repair downtime in manufacturing.

2 NUMERICAL METHOD

Due to the mentioned nonlinear properties of the Bellman equation (1), a predictor-corrector, Crank-Nicholson method (Douglas and Dupont, 1970) is used to approximate (1). The continuous, multistate vector $\mathbf{x}$ is replaced by its discretization,

$$\mathbf{X_j} = [X_{i,j_i}]_{m\times 1} = [X_{i,1} + (j_i - 1)DX_i]_{m\times 1}, \qquad (7)$$

where $\mathbf{DX}$ is the vector stepsize and $\mathbf{j} = [j_i]_{m\times 1}$ is the multistate vector index while $j_i = 1$ to M_i nodes per state and $i = 1$ to m states. The discretized time-to-go $t_f - t$ is

$$T_k = t_f - (k-1)\cdot DT, \qquad (8)$$

for $k = 1$ to K, with $T_1 = t_f$. Consequently, the optimal costs V^* has the discrete representation:

$$V_{\mathbf{j}k} \simeq V^*(\mathbf{X_j}, T_k), \qquad (9)$$

and the Bellman equation corrector step has the form:

$$V^{(\gamma+1)}_{\mathbf{j},k+1} = V_{\mathbf{j},k} + DT\cdot \mathcal{L}^{(\gamma)}_{\mathbf{j},k+\frac{1}{2}}, \qquad (10)$$

where the spatial operator $\mathcal{L}$ is evaluated at the γth correction for $\gamma = 0, 1, 2, \cdots$, and at the Crank-Nicolson temporal midpoint $k + \frac{1}{2}$ for $O(DT^2)$ accuracy, which in turn is approximated by

$$V^{(\gamma)}_{\mathbf{j},k+\frac{1}{2}} = \tfrac{1}{2}\left(V^{(\gamma)}_{\mathbf{j},k+1} + V_{\mathbf{j}k}\right), \qquad (11)$$

the usual average. The zeroth correction $\gamma = 0$ is defined as the prediction. The result from the prior time step $V_{\mathbf{j},k}$ is the final asymptotic corrected result $V^{(\infty)}_{\mathbf{j},k}$ in the theoretical case or $V^{(\gamma_{\max})}_{\mathbf{j},k}$ in the realistic, finite correction case. Extrapolation,

$$V^{(-1)}_{\mathbf{j},k+\frac{1}{2}} = \tfrac{1}{2}\left(3\cdot V^{(-1)}_{\mathbf{j},k} - V_{\mathbf{j},k-1}\right), \qquad (12)$$

i.e., $\gamma = -1$, is used to accelerate convergence prior to each predictor step except the first, once answers are available for two time steps. In (10), the spatial derivatives are replaced by central finite differences, while the Poisson induced delay terms are replaced by linear interpolants of compatible accuracy, i.e., $O(DX_l^2)$.

3 APPROXIMATE LINEAR COMPARISON EQUATION

Linearization and localization of the Bellman equation using diffusion approximation arguments, along with the use of *worst case* local estimates of variable coefficients, formally leads to a simpler, linearized, constant coefficient, comparison equation:

$$0 = \frac{\partial V^+}{\partial t} + \text{Trace}\left[\text{Diag}[A]\nabla_x\nabla_x^T V^+\right] + \mathbf{B}\cdot\nabla_x V^+. \qquad (13)$$

Here, Diag$[A]$ is an $m\times m$ diagonal constant matrix and $\mathbf{B}$ is a constant m-vector. The linearized, constant coefficient, comparison equation makes convergence and asymptotic stability amenable to analysis.

Typical estimates of the coefficients in (13) are appropriate bounds on the infinitesimal moments of the diffusion approximation:

$$\begin{aligned} B_l &= \max_{(\mathbf{x},T,\mathbf{u})}\left[\text{mean}[dX_l(t)|X_l(T) = x_l]\right]/DT \\ &= \max_{(\mathbf{x},t,\mathbf{u})}[F_l(\mathbf{x},t,\mathbf{u}) + \sum_j H_{lj}(\mathbf{x},t)\lambda_j], \end{aligned} \qquad (14)$$

and

$$\begin{aligned} A_{ll} &= \tfrac{1}{2}\cdot\max_{(\mathbf{x},T,\mathbf{u})}\left[\text{var}\left[dX_l(t)|X_l(T) = x_l\right]\right]/DT \\ &= \tfrac{1}{2}\cdot\max_{(\mathbf{x},t)}\left[\sum_j\left(G^2_{lj}(\mathbf{x},t) + H^2_{lj}(\mathbf{x},t)\cdot\lambda_j\right)\right] \end{aligned} \qquad (15)$$

Naimipour et al. (1993) give linear coefficient bounds for the single state, quadratic costs case of the Bellman equation.

4 CORRECTOR CONVERGENCE

The von Neumann's Fourier method is applied to analyze the convergence properties of the predictor-corrector method. Discrete Fourier representation of the final optimal feedback value is assumed:

$$V_{\mathbf{j},1} = C\cdot\exp\left(i\cdot\sum_{l=1}^{m} j_l\cdot\beta_l\cdot DX_l\right). \qquad (16)$$

Only an arbitrary, single vector mode $\boldsymbol{\beta} = [\beta_l]_{m\times 1}$ need be examined, instead of a infinite Fourier representation, since the comparison equation is linear.

The predictor-corrector Bellman equation corresponding to the linear comparison equation (13),

$$V^{(\gamma+1)}_{\mathbf{j},k+1} = V_{\mathbf{j},k} + DT\cdot(\mathbf{A}\cdot\mathbf{DDV} + \mathbf{B}\cdot\mathbf{DV})^{(\gamma)}_{\mathbf{j},k+\frac{1}{2}}, \qquad (17)$$

where **DV** is the finite difference vector for the gradient, **DDV** is the finite difference vector for the vector of second derivatives and $\mathbf{A} = [A_{ll}]_{m\times 1}$ is the vector of diffusion coefficients.

The beginning *extrapolation evaluation* step is taken as the *final condition*,

$$V^{(-1)}_{\mathbf{j},\frac{3}{2}} = V_{\mathbf{j},1}\,, \tag{18}$$

due to the backward nature of the problem. Upon substitution of (18), with Fourier assumption (16), the prediction becomes

$$V^{(0)}_{\mathbf{j},2} = (1 + 2\cdot\theta)\cdot V_{\mathbf{j},1}\,, \tag{19}$$

where the *corrector convergence parameter* is defined as

$$\begin{aligned}\theta \;&\equiv\; DT\cdot\sum_{l=1}^{m}\left(\frac{i\cdot B_l}{2\cdot DX_l}\cdot\sin(\beta_l\cdot DX_l)\right.\\ &-\;\left.\frac{2\cdot A_{ll}}{DX_l^2}\cdot\sin^2(\beta_l\cdot DX_l/2)\right).\end{aligned} \tag{20}$$

The *predictor evaluation* is

$$V^{(0)}_{\mathbf{j},\frac{3}{2}} = \tfrac{1}{2}(V^{(0)}_{\mathbf{j},2} + V_{\mathbf{j},1}) = (1+\theta)\cdot V_{\mathbf{j},1}. \tag{21}$$

Using induction, the general correction for the second time step is shown to be

$$V^{(\gamma)}_{\mathbf{j},2} = \left[\rho - \frac{2\cdot\theta^{\gamma+2}}{1-\theta}\right]\cdot V_{\mathbf{j},1}, \tag{22}$$

after some algebra, where the *temporal stability parameter* is defined as

$$\rho \equiv \frac{1+\theta}{1-\theta}, \tag{23}$$

which will be used later.

Consequently, in the limit of a large number of corrections, $\gamma \to +\infty$, the asymptotic value on the second time step is

$$V^{(\infty)}_{\mathbf{j},2} = \rho\cdot V_{\mathbf{j},1}, \tag{24}$$

provided the *corrector convergence condition* holds,

$$|\theta| < 1, \tag{25}$$

so that $\theta^{\gamma+2} \to 0$ as $\gamma \to +\infty$. Similarly, on earlier time steps,

$$\begin{aligned}V^{(\gamma)}_{\mathbf{j},k} \;&=\; \left[\rho^{k-1} + \frac{2\cdot\rho^{k-3}\cdot(1-\rho)\cdot\theta^{\gamma+2}}{1-\theta}\right]\cdot V_{\mathbf{j},1}\\ &\to\; V^{(\infty)}_{\mathbf{j},k} = \rho^{k-1}\cdot V_{\mathbf{j},1}\,,\end{aligned} \tag{26}$$

as $\gamma \to +\infty$ when $|\theta| < 1$.
Since the corrector convergence condition (25) is not in a helpful form, θ must be simplified using the original complex definition (20), which can be transformed to its real squared modulus,

$$\begin{aligned}|\theta|^2 \;&=\; \Re^2[\theta] + \Im^2[\theta] \\ &=\; DT^2\cdot\left[\left(\sum_{l=1}^{m}\frac{B_l}{2\cdot DX_l}\cdot\sin(\beta_l\cdot DX_l)\right)^2\right.\\ &+\;\left.\left(\sum_{l=1}^{m}\frac{2\cdot A_{ll}}{DX_l^2}\cdot\sin^2(\beta_l\cdot DX_l/2)\right)^2\right].\end{aligned} \tag{27}$$

A simpler, usable expression can be found using the inequality $0 \le \sin^2(\phi/2) \le 1$ on the diffusion terms since $A_{ll} \ge 0$ and $B_l\sin(\phi) \le |B_l|$ on the drift terms since the sign of the drift term can be anything. Finally, the condition on $|\theta|$ can be replaced by following simplified, but uniform in the A_{ll} and B_l, corrector convergence criterion:

$$\begin{aligned}\sigma \equiv DT\cdot\sqrt{\left(\sum_l\frac{|B_l|}{2\cdot DX_l}\right)^2 + \left(\sum_l\frac{2\cdot A_{ll}}{(DX_l)^2}\right)^2},\\ < 1,\end{aligned} \tag{28}$$

where DT is the time-step size and DX_l is the lth state-step size. The numerical Bellman equation is *parabolic-dominant* or *diffusion-dominant* if the diffusion coefficients A_{ll} dominate in (28), while *hyperbolic-dominant* or *convection-dominant* when the drift terms B_l dominate.

In theory, (28) is the criterion for an infinite number of corrections. However, a practical corrector stopping criterion would be to select the maximum number of corrections $\gamma_{\max}$ according to

$$\left|V^{(\gamma)}_{\mathbf{j},k+1} - V^{(\gamma-1)}_{\mathbf{j},k+1}\right| < \mathrm{tol}\cdot\left|V^{(\infty)}_{\mathbf{j},k+1}\right|, \tag{29}$$

where tol is the relative tolerance (e.g., $tol = \frac{1}{2}\cdot 10^{1-d}$ for approximately d relative digits accuracy). Using the formula (26) in (29), an estimate of maximum number of corrections corresponding to d-accuracy is given by the ceiling function

$$\gamma_{\max} = \left\lceil\frac{\log\left(\frac{|1+\theta|^2\cdot 10^{1-d}}{8\cdot|1-\theta|\cdot|\theta|^2}\right)}{\log(|\theta|)}\right\rceil, \tag{30}$$

which is used in actual computations.

5 ASYMPTOTIC STABILITY

Investigating the asymptotic stability for long times-to-go, $t_f - t$, with respect to Fourier final data noise, requires another formulation of the problem. Let

$$E^{(\gamma)}_{\mathbf{j},k} = \widetilde{V}^{(\gamma)}_{\mathbf{j},k} - V^{(\gamma)}_{\mathbf{j},k} \tag{31}$$

be the error in the numerical approximation with dirty final data $\widetilde{V}_{\mathbf{j},k}$ relative to the exact numerical approximation. Both approximations satisfy the discretized Bellman Eq. (17), so that

$$\begin{aligned} E_{\mathbf{j},k+1}^{(\gamma+1)} &= E_{\mathbf{j},k} \\ &+ DT \cdot (\mathbf{A} \cdot \mathbf{DDE} + \mathbf{B} \cdot \mathbf{DE})_{\mathbf{j},k+\frac{1}{2}}^{(\gamma)}, \end{aligned} \tag{32}$$

where $\mathbf{DE}$ and $\mathbf{DDE}$ are the respective vector of finite differences for the derivatives. The two approximations only differ in the final discrete Fourier error

$$E_{\mathbf{j},1} = C \cdot \exp\left(i \cdot \sum_{l=1}^{m} j_l \cdot \beta_l \cdot DX_l\right), \tag{33}$$

as in (16). Since the analysis is the same as in the corrector convergence case, the asymptotic error in the limit of a large number of corrections is

$$E_{\mathbf{j},k}^{(\infty)} = \rho^{k-1} \cdot E_{\mathbf{j},1}, \tag{34}$$

as in (26). Hence, $\rho^{k-1} \to 0$ is needed as the temporal index $k \to \infty$ for asymptotic stability relative to Fourier perturbations in the final data. Thus, $|\rho| < 1$ is necessary. In order to show this, let $\theta = \Re[\theta] + i\Im[\theta]$ with complex conjugate $\overline{\theta} = \Re[\theta] - i\Im[\theta]$, where from (20),

$$\Re[\theta] = -DT \cdot \sum_{l=1}^{m} \frac{2 \cdot A_{ll}}{DX_l^2} \cdot \sin^2(\beta_l \cdot DX_l/2) \tag{35}$$

which is negative, and

$$\Im[\theta] = DT \cdot \sum_{l=1}^{m} \frac{B_l}{2 \cdot DX_l} \cdot \sin(\beta_l \cdot DX_l), \tag{36}$$

so that

$$\begin{aligned} |\rho|^2 &= \frac{(1+\theta)}{(1-\theta)} \cdot \frac{(1+\overline{\theta})}{(1-\overline{\theta})} \\ &= \frac{(1+\Re[\theta])^2 + (\Im[\theta])^2}{(1-\Re[\theta])^2 + (\Im[\theta])^2} < 1, \end{aligned} \tag{37}$$

or $|\rho| < 1$, because $\Re[\theta] = -|\Re[\theta]| < 0$ and $(1 + |\Re[\theta]|) > (1 - |\Re[\theta]|)$. Note that corrector convergence, $|\theta| < 1$, automatically implies asymptotic stability, since $|\theta| < 1$ implies that $|\Re[\theta]| < 1$ also. Thus, $E_{\mathbf{j},k}^{(\infty)} \to 0$ as $k \to +\infty$, independent of the final data error.

6 CONCLUSIONS

In this paper, a uniform mesh criterion has been derived for corrector convergence of the numerical approximation to solution of the Bellman equation for stochastic dynamic programming. The results are good for fairly general Markov noise in continuous time. The corrector convergence condition applies whether the numerical Bellman equation is parabolic-dominant or hyperbolic-dominant.

In addition, it has been shown that the numerical solution to the linearized comparison equation is asymptotic stable for long times-to-go when the corrector convergence condition is satisfied, at least in theory.

These results have been used with much success for super vector and massively parallel computations in relatively large scale stochastic control applications by Hanson and coworkers (1987,1991a,1991b,1992).

Acknowledgement. This work was supported in part by the National Science Foundation Grants DMS-8806099 and DMS-9102343, by Los Alamos National Laboratory's Advanced Computing Laboratory, and by the National Center for Supercomputing Applications.

7 REFERENCES

Chung, S.-L., F. B. Hanson, and H. H. Xu (1992). Parallel stochastic dynamic programming: Finite element methods. *Lin. Alg. Applic.*, **172**, 197-218.

Douglas, J. Jr., and T. DuPont (1970). Galerkin methods for parabolic equations. *SIAM J. Num. Anal.*, **7**, 575-626.

Gihman, I. I., and A. V. Skorohod (1979). *Controlled Stochastic Processes*. Springer-Verlag, New York.

Hanson, F. B. (1987). Bioeconomic model of the Lake Michigan alewife fishery. *Can. J. Fish. Aquat. Sci.*, **44 (Suppl. II)**, 298-305

Hanson, F. B. (1991a). Computational dynamic programming on a vector multiprocessor. *IEEE Trans. Automatic Control*, **36**, 507-511.

Kushner, H. J., and P. Dupuis (1992). *Numerical Methods for Stochastic Control in Continuous time*. Springer-Verlag, New York.

Naimipour, K., and F. B. Hanson (1993). Convergence of a numerical method for the Bellman equation of stochastic optimal control with quadratic costs and constrained control. *Dynamics and Control*, to appear, 1-27.

Pratico, C. J., F. B. Hanson, H. H. Xu, D. J. Jarvis and M. S. Vetter (1992). Visualization for the management of renewable resources in an uncertain environment. In: *Proc. Supercomputing '92*, 258-266 (color plates p. 843).

Xu, H. H., F. B. Hanson, and S.-L. Chung (1991b). Optimal data parallel methods for stochastic dynamic programming. In: *Proc. 1991 Int. Conf. on Parallel Processing, Vol. III Algorithms and Applications*, 142-146.