The Gauss-Seidel Numerical Procedure for Markov Stochastic Games

Harold J. Kushner
Applied Mathematics Department
Lefschetz Center for Dynamical Systems
Brown University, Providence, RI 02912, USA
hjk@dam.brown.edu, 401-863-1400

Abstract— **Consider the problem of value iteration for solving Markov stochastic games. One simply iterates backwards, via a Jacobi-like procedure. The convergence of the Gauss-Seidel form of this procedure is shown for both the discounted and ergodic cost problems, under appropriate conditions, with extensions to problems where one stops when a boundary is hit or if any one of the players chooses to stop, with associated costs. Generally, the Gauss-Seidel procedure accelerates convergence.**

Key words. Stochastic games, Markov games, Gauss-Seidel procedure, numerical algorithms

I. Introduction

We consider two-player, zero-sum, finite-state, Markov stochastic games. There are N states and, unless noted otherwise, we suppose that the controls are feedback and not randomized. In state i, player 1's (the minimizing player) control is denoted by u_i and that of player 2 (the maximizing player) is denoted by v_i. The convergence of the value iteration procedure (see (2.2) below) for Markov stochastic games for a discounted cost function (or where there is an absorbing boundary) was established in [9], [13]. The convergence of the Gauss-Seidel procedure was first established for the control problem in [12]. It is widely used and is no less fast and is generally faster than the Jacobi procedure; see, for example, [7], [12] and [11, Chapter 6]. It will be seen that the Gauss-Seidel procedure can be viewed as an iteration with a modified transition matrix Q. The references discuss the nature of the transition probability that Q represents and show why it is faster. In particular, it has a smaller spectral radius than the transition matrix of the original problem. The ordering of the states in the iteration plays an important role in getting the best acceleration of convergence. If there is an absorbing set, then it is best to order the states so that the mean flow is toward that set. In practice, where there is no absorbing set, the ordering is often changed from cycle to cycle, say "reversing direction," to provide a greater mixing, which also accelerates convergence. See [7], [12] and [11, Chapter 6] for a discussion of preferred orderings, a point that we do not have the space to deal with here, but is the same for the control and the game problem.

This work was partially supported by Contract DAAD-19-02-1-0425 from the Army Research Office and National Science Foundation Grant ECS 0097447.

The convergence of the Gauss-Seidel form has not yet been established for the game problem. Under appropriate conditions, the convergence will be established for the discounted and ergodic cost functions, and for related problems such as where there is an absorbing boundary or optional stopping.

The u_i, v_i take values in compact sets that might depend on i. Define the control vectors $u = \{u_i, i \le N\}$, $v = \{v_i, i \le N\}$. Let $\{\tilde{p}_{ij}(u_i, v_i); i, j \le N\}$ denote the transition probabilities under controls u, v, and define $p_{ij}(u_i, v_i) = \rho\tilde{p}_{ij}(u_i, v_i)$, where $\rho \in (0, 1)$ is the discount factor. Define the degenerate transition matrix $P(u, v) = \{\rho p_{ij}(u_i, v_i); i, j \le N\}$. Hence the row sums of $P(u, v)$ are $1 - \rho$. The cost rate when in state i and under u_i, v_i is the function $k_i(u_i, v_i)$. Let $\{X_n\}$ denote the random variables of chain. Then the discounted cost under u, v is

$$C_i(u, v) = E_i^{u,v} \sum_{n=1}^{\infty} \rho^n k_{X_n}(u_{X_n}, v_{X_n}),$$

where $E_i^{u,v}$ denotes the expectation under u, v and with initial state i. It is always supposed that the $p_{ij}(u_i, v_i)$ and $k_i(u_i, v_i)$ are continuous in the u_i, v_i. Define the vector $K(u, v) = \{k_i(u_i, v_i); i \le N\}$.

In addition, unless noted otherwise, we assume that the Isaacs condition holds; namely, that for any N-vector $H = \{h_i, i \le N\}$,

$$\sup_v \inf_u [P(u, v)H + K(u, v)] = \inf_u \sup_v [P(u, v)H + K(u, v)]. \tag{1.1}$$

In vector forms such as (1.1), it is always supposed that the inf and sup are taken line by line, so that the ith line is $\sup_{v_i} \inf_{u_i} [\sum_j p_{ij}(u_i, v_i)h_i + k_i(u_i, v_i)]$ and involves the inf and sup over u_i and v_i only. The condition (1.1) is used for notational simplicity. Otherwise, one must randomize the controls. Then, when the number of control

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Gauss-Seidel Numerical Procedure for Markov Stochastic Games	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Brown University,Division of Applied Mathematics,182 George Street,Providence,RI,02912**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**6**	

values is finite, the control is replaced by the vector of probabilities, and the analog of (1.1) holds. If the controls take values in a continuum and (1.1) does not hold, then the randomization is more complicated, but the results can be readily extended. The condition (1.1) commonly holds for the games arising as numerical approximations to stochastic differential games under the conditions of [6], [8].

Section 2 concerns the discounted cost problem and also remarks on cases where there is forced stopping on hitting a boundary or with optional stopping. The ergodic cost problem is dealt with in Section 3 and the cost under u, v is

$$\gamma(u,v) = \lim_n \frac{1}{n} E_i^{u,v} \sum_{l=1}^{n} k_{X_l}(u_{X_l}, v_{X_l}). \tag{1.2}$$

In Section 3, it is first shown that the game version of a classical value iteration method converges. Then this is adapted to the Gauss-Seidel procedure. If controls $\hat{u}^n, \hat{v}^n$ are used at time n, then write $p_{ij}^{(n)}(\hat{u}^n, \hat{v}^n; \hat{u}^{n-1}, \hat{v}^{n-1}; \cdots; \hat{u}^1, \hat{v}^1)$ for the n-step transition probabilities. The ergodic cost problem uses the additional assumption that there is an $\epsilon > 0$, a state j_0, and an integer $m \leq N$, such that

$$p_{ij_0}^{(m)}(\hat{u}^m, \hat{v}^m; \hat{u}^{m-1}, \hat{v}^{m-1}; \cdots; \hat{u}^1, \hat{v}^1) \geq \epsilon \tag{1.3}$$

for all possible controls. This is a standard condition for the ergodic cost problem in the control literature [15]. See also [1, Vol 2] and [7, pp156–158]. Of particular interest are Markov chain games that arise as numerical approximations of games with diffusion models as in [6], [8], where (1,3) will commonly hold under the assumptions on the nondegeneracy of the diffusion in [6].

To date, there have not been proofs of the convergence of the Gauss-Seidel method for either the game or the control problem with ergodic cost criteria. Indeed, it does not always converge, even under (1.3) for ergodic models. But, it will converge if (1.3) holds for a modified transition probability. This will be discussed further in Section 3. The modified condition holds for the chains obtained as approximations in [6] under the nondegeneracy conditions used there. These chains are obtained via the Markov chain approximation methods of [11]. The book [4] discusses other numerical procedures, based on nonlinear programming methods and (under some smoothness conditions) develops a convergent modified Newton procedure that has the rate of the policy iteration procedure whenever that converges. The paper [5] discusses what might be called a type of combined value iteration and approximation in policy space method. The paper [3] numerically compares a variety of approaches and shows that the policy iteration algorithm performs best when it converges (which is not always the case). The papers [2], [14] discuss a variety of modifications of value and policy iteration.

II. The Discounted Cost Problem and Extensions

Until further notice, we consider the discounted cost case. Let $\bar{C}_i$ denote the value of the game when starting in state i and define $\bar{C} = \{\bar{C}_i; i \leq N\}$. In vector form, the equation for the value is

$$\bar{C} = \sup_v \inf_u \left[P(u,v)\bar{C} + K(u,v)\right] = \inf_u \sup_v \left[P(u,v)\bar{C} + K(u,v)\right]. \tag{2.1}$$

In all such vector equations, The inf sup is taken by line; the ith line is over u_i, v_i. Recall that the discounting is incorporated into the $P(u,v)$, Hence, for any integer $m \geq 1$, $P^m(\hat{u}^m, \hat{v}^m; \cdots; \hat{u}^1, \hat{v}^1)$ is a contraction (in the Euclidean norm sense) uniformly in the choices of the controls $\{\hat{u}^n, \hat{v}^n\}$. A unique solution $\bar{C}$ exists and is the value [4, Theorem 3.1.1]. Let $\bar{u}, \bar{v}$ denote any controls that realize (2.1).

Our aim is the computation of $\bar{C}$, hence of optimizing controls as well. A variety of computational methods are available. In [9], [10], [13] it was shown that, for any C^0, the C^n in the iteration in value space algorithm

$$C^{n+1} = \sup_v \inf_u [P(u,v)C^n + K(u,v)] \tag{2.2}$$

converge to $\bar{C}$.

The Gauss-Seidel procedure for the game problem is the iteration in value space with successive substitutions, taken in the order $i = 1, 2, \ldots,$

$$C_i^{n+1} = \sup_{v_i} \inf_{u_i} \left[\sum_{j=1}^{i-1} p_{ij}(u_i, v_i) C_j^{n+1} + \sum_{j=i}^{N} p_{ij}(u_i, v_i) C_j^n + k_i(u_i, v_i)\right]. \tag{2.3}$$

Taking the sup inf in (2.3) is equivalent to solving a matrix game. Except for this sup inf, it is just the standard Gauss-Seidel method for iteratively solving linear equations. The convergence proof in Theorem 2.1 adapts the method of [7], [12]. The ordering of the states can vary with n.

Before proceeding, it is convenient to define a transition probability $Q(u,v)$ and cost vector $\hat{K}(u,v)$ that play a crucial role in the analysis. This Q will be the effective transition probability that determines the behavior of the Gauss-Seidel procedure. Consider the set of linear equations in an unknown $D = \{D_{ij}\}$, where the vector C is given, solved by successive substitution in the order $i = 1, 2, \ldots, N$:

$$D_i = \left[\sum_{j=1}^{i-1} p_{ij}(u_i, v_i) D_j + \sum_{j=i}^{N} p_{ij}(u_i, v_i) C_j + k_i(u_i, v_i)\right]. \tag{2.4}$$

This uniquely defines a matrix $Q(u,v) = \{q_{ij}(u,v); i,j \le N\}$ and vector $\hat{K}(u,v) = \{\hat{k}_i(u,v); i \le N\}$ such that $D = Q(u,v)C + \hat{K}(u,v)$. In detail, by successive substitutions in (2.4), we find that

$$\begin{aligned} q_{1j}(u,v) &= p_{1j}(u_1,v_1), \quad 1 \le j \le N, \\ q_{21}(u,v) &= p_{21}(u_2,v_2)q_{11}(u,v), \\ q_{2j}(u,v) &= p_{2j}(u_2,v_2) + p_{21}(u_2,v_2)q_{1j}(u,v), \quad 2 \le j \le N. \end{aligned}$$

In general,

$$\begin{aligned} q_{ij}(u,v) &= p_{ij}(u_i,v_i) + \sum_{k=1}^{i-1} p_{ik}(u_i,v_i)q_{kj}(u,v), \quad j \ge i, \\ q_{ij}(u,v) &= \sum_{k=1}^{i-1} p_{ik}(u_i,v_i)q_{kj}(u,v), \quad 1 \le j < i. \end{aligned} \tag{2.5}$$

$Q(u,v)$ can also be defined from (2.4) in terms of the upper and lower triangular matrices formed from $P(u,v)$, but we prefer to write the details. Also,

$$\begin{aligned} \hat{k}_1(u,v) &= k_1(u_1,v_1), \\ \hat{k}_2(u,v) &= p_{21}(u_2,v_2)\hat{k}_1(u,v) + k_2(u_2,v_2), \end{aligned}$$

and, in general,

$$\hat{k}_i(u,v) = \sum_{k=1}^{i-1} p_{ij}(u_i,v_i)\hat{k}_j(u,v) + k_i(u_i,v_i). \tag{2.6}$$

Note that, for our discounted cost problem where the discount factor is incorporated into the $p_{ij}(u_i,v_i)$, $Q(u,v)$ is a degenerate transition matrix since the row sums satisfy $\sum_j q_{ij}(u,v) \le \rho$ for all i and controls. If there is no discounting (i.e., $\rho = 1$), then the row sums are always unity. These facts are easily proved by induction, starting with $i = 1$.

Theorem 2.1. *For any C^0, the C^n in* (2.3) *converges to* $\bar{C}$.

Proof. Since

$$\bar{C}_i = \sup_{v_i} \inf_{u_i} \left[\sum_{j=1}^{i-1} p_{ij}(u_i,v_i)\bar{C}_j + \sum_{j=i}^{N} p_{ij}(u_i,v_i)\bar{C}_j + k_i(u_i,v_i) \right], \tag{2.7}$$

by successive substitutions, we can write (2.1) in the equivalent form

$$\bar{C} = \sup_v \inf_u \left[Q(u,v)\bar{C} + \hat{K}(u,v) \right] = Q(\bar{u},\bar{v})\bar{C} + \hat{K}(\bar{u},\bar{v}). \tag{2.8}$$

Similarly, with u^n, v^n realizing (2.3), the following is equivalent to (2.3):

$$\begin{aligned} C^{n+1} &= \sup_v \inf_u \left[Q(u,v)C^n + \hat{K}(u,v) \right] \\ &= Q(u^n,v^n)C^n + \hat{K}(u^n,v^n). \end{aligned} \tag{2.9}$$

In (2.8) and (2.9), it is understood that the inf and sup are again taken line by line, in the order $i = 1, 2 \ldots$. The inf and sup in line 1 are over u_1 and v_1, and in turn, that in line i are over u_i, v_i.

For any u, v, and $i = 1, \ldots, N$, (2.1) yields

$$\begin{aligned} &\sum_{j=1}^{i-1} p_{ij}(\bar{u}_i,v_i)\bar{C}_j + \sum_{j=i}^{N} p_{ij}(\bar{u}_i,v_i)\bar{C}_j + k_i(\bar{u}_i,v_i) \\ &\le \bar{C}_i = \sup_{v_i} \inf_{u_i} \left[\sum_{j=1}^{i-1} p_{ij}(u_i,v_i)\bar{C}_j + \sum_{j=i}^{N} p_{ij}(u_i,v_i)\bar{C}_j + k_i(u_i,v_i) \right] \\ &= \sum_{j=1}^{i-1} p_{ij}(\bar{u}_i,\bar{v}_i)\bar{C}_j + \sum_{j=i}^{N} p_{ij}(\bar{u}_i,\bar{v}_i)\bar{C}_j + k_i(\bar{u}_i,\bar{v}_i) \\ &\le \sum_{j=1}^{i-1} p_{ij}(u_i,\bar{v}_i)\bar{C}_j + \sum_{j=i}^{N} p_{ij}(u_i,\bar{v}_i)\bar{C}_j + k_i(u_i,\bar{v}_i). \end{aligned} \tag{2.10}$$

In vector notation, this can be written as

$$P(\bar{u},v)\bar{C} + K(\bar{u},v) \le \bar{C} \le P(u,\bar{v})\bar{C} + K(u,\bar{v}).$$

It can also be written as

$$Q(\bar{u},v)\bar{C} + \hat{K}(\bar{u},v) \le \bar{C} \le Q(u,\bar{v})\bar{C} + \hat{K}(u,\bar{v}). \tag{2.11}$$

For any u, v, (2.3) or, equivalently, (2.9) yields

$$\begin{aligned} &Q(u^n,v)C^n + \hat{K}(u^n,v) \\ &\le C^{n+1} = \sup_v \inf_u \left[Q(u,v)C^n + \hat{K}(u,v) \right] \le Q(u,v^n)C^n + \hat{K}(u,v^n). \end{aligned} \tag{2.12}$$

Selecting $(u,v) = (u^n,v^n)$ in (2.11) and $(u,v) = (\bar{u},\bar{v})$ in (2.12) yields

$$\begin{aligned} &Q(u^n,\bar{v})\left(C^n - \bar{C}\right) \\ &\quad = \left[Q(u^n,\bar{v})C^n + \hat{K}(u^n,\bar{v}) \right] - \left[Q(u^n,\bar{v})\bar{C} + \hat{K}(u^n,\bar{v}) \right] \\ &\quad \le C^{n+1} - \bar{C} \\ &\quad \le \left[Q(\bar{u},v^n)C^n + \hat{K}(\bar{u},v^n) \right] - \left[Q(\bar{u},v^n)\bar{C} + \hat{K}(\bar{u},v^n) \right] \\ &\quad = Q(\bar{u},v^n)\left(C^n - \bar{C}\right). \end{aligned} \tag{2.13}$$

Iterating yields

$$\begin{aligned} &\left[Q(u^n,\bar{v})\cdots Q(u^1,\bar{v})\right]\left(C^1 - \bar{C}\right) \\ &\le C^{n+1} - \bar{C} \le \left[Q(\bar{u},v^n)\cdots Q(\bar{u},v^1)\right]\left(C^1 - \bar{C}\right). \end{aligned} \tag{2.14}$$

For the discounted problem the row sums of $Q(u,v)$ satisfy $\sum_j q_{ij}(u,v) \le \rho$ for all i and controls. Hence $C^n \to \bar{C}$ as $n \to \infty$. ■

Stopping when hitting a boundary set. Now, we allow $\rho \in (0,1]$, so that the discounting can be dropped if desired. Suppose that the process stops when a boundary

set is hit and that the mean time to reach the boundary set is bounded, uniformly in the controls and initial condition. Thus we can suppose that the boundary set is absorbing and has zero cost. Without loss of generality, let 0 denote the boundary state. Let $P(u,v)$ still denote the matrix of transition probabilities among the states $1,\dots,N$ only. With $\bar{C} = \{\bar{C}_i; 1 \le i \le N\}$ and $C^n = \{C_i^n; 1 \le i \le N\}$ the products on either side of (2.14) go to zero. Thus $C^n \to \bar{C}$. It is preferable if state 1 is connected to the boundary and the states are ordered so that the "mean flow" is toward the boundary as one goes from the lower to the higher numbered states, where possible. The transition matrix Q associated with such an ordering has a faster absorption of the process at the boundary, which implies a faster convergence of the algorithm [7], [12] and [11, Chapter 6].

Optional stopping problems. As in the above paragraph, let $\rho \in (0,1]$. Various forms of optional stopping can be handled. There are now three ways that the process can be stopped. One is by hitting a predefined stopping set, denoted by state 0, as in the previous paragraph. Call the time τ_0. Otherwise, either player can decide to have the game stopped. The associated times are called τ_i for player i. After stopping for whatever reason, the state goes to absorbing 0, with zero holding cost there. The $P(u,v)$ represents the transition probabilities only among the states $1,\dots,N$. For given functions $g_i(\cdot)$, the cost is now

$$C_i(u,v) = E_i^{u,v} \sum_{n=0}^{\min\{\tau_0,\tau_1,\tau_2\}-1} k_{X_n}(u_{X_n}, v_{X_n}) + E_i^{u,v} g_1(X_{\tau_1}) I_{\{\tau_1 \le \tau_2, \tau_1 < \tau_0\}} + E_i^{u,v} g_2(X_{\tau_2}) I_{\{\tau_2 < \min\{\tau_1,\tau_0\}\}}. \tag{2.15}$$

The controls u_i, v_i can now take the new value *stop* as well as the original values used in Theorem 2.1. Let $k_i(u_i,v_i) \ge \epsilon > 0$ for all i, u_i, v_i values other than the value *stop*, and suppose that $g_1(\cdot) \ne g_2(\cdot)$ but $g_1(i) \ge g_2(i)$. Extend the definition of the $k_i(\cdot)$ to include the control value *stop*, by writing $k_i(\text{stop}, v_i) = g_1(i)$ and let $k_i(u_i, \text{stop}) = g_2(i)$ if $u_i \ne$ stop and let it be zero otherwise. Then the Gauss-Seidel algorithm can be written as (2.3). We have $g_2(i) \le C_i^n \le g_1(i)$. Similarly (2.1) holds and $g_2(i) \le \bar{C}_i \le g_1(i)$.

Let (u^n, v^n) satisfy (2.3) and $(\bar{u}, \bar{v})$ satisfy (2.1). Due to the positivity of $k_i(u_i,v_i)$, for u_i and/or v_i not equal to *stop*, if player 1 uses u^n and player 2 uses some $\hat{v}^n$ at time n, then $P(u^n,\tilde{v}^n)\cdots P(u^0,\hat{v}^0) \to 0$ and $Q(u^n,\tilde{v}^n)\cdots Q(u^0,\hat{v}^0) \to 0$ as $n \to \infty$ uniformly in the $\{\hat{v}^n\}$ choices. Analogously, $Q(\bar{u},\tilde{v}^n)\cdots Q(\bar{u},\hat{v}^0) \to 0$ for all $\{\hat{v}^n\}$ choices. Using these facts and following the logic of the proof of Theorem 2.1 yields the convergence $C^n \to \bar{C}$ for this problem.

III. The Ergodic Cost Problem

Now $\rho = 1$ and $P(u,v)$ is the transition matrix for a controlled Markov chain which is ergodic under any u, v. We adapt the procedure of [7, pp156–158], originally due to White [15]. Let $\mathbf{e}$ denote the N-vector, all of whose components are unity.

A Jacobi procedure. We first consider the analog of the simple backwards iteration (Jacobi) procedure (2.2), whose convergence for the game with ergodic payoffs has not been proved to date in the literature. For arbitrary W^0, define the vectors W^n, w^n recursively by

$$\begin{aligned} W^n &= \sup_v \inf_u \left[P(u,v)w^{n-1} + K(u,v)\right] \\ w^n &= W^n - W^n_{j_0}\mathbf{e}, \end{aligned} \tag{3.1}$$

where j_0 is defined above (1.3). There is a value for the game [4, Section 5.2]. The value $\bar{\gamma}$ is given by

$$\bar{W} + \bar{\gamma}\mathbf{e} = \sup_v \inf_u \left[P(u,v)\bar{W} + K(u,v)\right]. \tag{3.2}$$

As for the control problem, the value of $\bar{W}$ is unique, up to the addition of a vector with constant components, and the value of $\bar{\gamma}$ is unique. An alternative way of writing (3.2) is as

$$\begin{aligned} \bar{W} &= \sup_v \inf_u \left[P(u,v)\bar{w} + K(u,v)\right] \\ \bar{w} &= \bar{W} - \bar{W}_{j_0}\mathbf{e}. \end{aligned} \tag{3.2a}$$

Theorem 3.1. *w^n converges to the value $\bar{\gamma}$ of the game.*

Proof. Recall the condition (1.3) and the definitions of m and j_0 there. Let u^n, v^n be the selected controls in (3.1). Define $c_n = W^n_{j_0}\mathbf{e}$. Then, for any u, v,

$$\begin{aligned} &P(u^n,v)w^{n-1} + K(u^n,v) \\ &\quad \le W^n = P(u^n,v^n)w^{n-1} + K(u^n,v^n) \\ &\quad \le P(u,v^n)w^{n-1} + K(u,v^n). \end{aligned} \tag{3.3}$$

Let $(u,v) = (u^{n-1}, v^{n-1})$ in (3.3) and use the definition of w^n in (3.1) to get

$$\begin{aligned} &P(u^n,v^{n-1})w^{n-1} + K(u^n,v^{n-1}) - c_n \\ &\quad \le w^n \le P(u^{n-1},v^n)w^{n-1} + K(u^{n-1},v^n) - c_n. \end{aligned}$$

Replacing n with $n-1$ in (3.3) and letting $(u,v) = (u^n,v^n)$ yields, for $i \le N$,

$$\begin{aligned} &P(u^{n-1},v^n)w^{n-2} + K(u^{n-1},v^n) - c_{n-1} \\ &\quad \le w^{n-1} \le P(u^n,v^{n-1})w^{n-2} + K(u^n,v^{n-1}) - c_{n-1}. \end{aligned}$$

The last two inequalities yield

$$\begin{aligned} &P(u^n,v^{n-1})\left(w^{n-1} - w^{n-2}\right) - (c_n - c_{n-1}) \le w^n - w^{n-1} \\ &\quad \le P(u^{n-1},v^n)\left(w^{n-1} - w^{n-2}\right) - (c_n - c_{n-1}). \end{aligned} \tag{3.4}$$

Iterating (3.4) $m-1$ times leads to

$$\begin{aligned} &P(u^n, v^{n-1}) \cdots P(u^{n-m+1}, v^{n-m})(w^{n-m} - w^{n-m-1}) \\ &\quad -(c_n - c_{n-m}) \quad \le w^n - w^{n-1} \\ &\le P(u^{n-1}, v^n) \cdots P(u^{n-m}, v^{n-m+1})(w^{n-m} - w^{n-m-1}) \\ &\quad -(c_n - c_{n-m}). \end{aligned} \tag{3.5}$$

Define $\delta w^n = w^n - w^{n-1} = \{\delta w_i^n; i \le N\}$. Then the right hand inequality of (3.5) yields, for $i \le N$,

$$\begin{aligned} &\delta w_i^n \le \\ &\sum_j p_{ij}(u^{n-1}, v^n; u^{n-2}, v^{n-1}; \cdots; u^{n-m}, v^{n-m+1}) \delta w_j^{n-m} \\ &\quad - \left[W_{j_0}^n - W_{j_0}^{n-m}\right]. \end{aligned} \tag{3.6}$$

Since $w_{j_0}^n = 0$ for all n, we have $\delta w_{j_0}^n = 0$. This, with (3.6) and (from (1.3))

$$p_{ij_0}^{(m)}(u^{n-1}, v^n; u^{n-2}, v^{n-1}; \cdots; u^{n-m}, v^{n-m+1}) \ge \epsilon > 0$$

for all i, n, and controls, yields for $i \le N$

$$\max_i \delta w_i^n \le (1-\epsilon) \max_j \delta w_j^{n-m} - \left[W_{j_0}^n - W_{j_0}^{n-m}\right].$$

Analogously, using the fact that $\min_i w_i^n \le 0$ and that

$$p_{ij_0}(u^n, v^{n-1}; u^{n-1}, v^{n-2}; \cdots; u^{n-m+1}, v^{n-m}) \ge \epsilon$$

for all i, n, the left hand inequality of (3.5) yields, for $i \le N$,

$$\min_i \delta w_i^n \ge (1-\epsilon) \min_j \delta w_j^{n-m} - \left[W_{j_0}^n - W_{j_0}^{n-m}\right].$$

Hence, for all i, where we define $[\max_i - \min_i] a_i = \max_i a_i - \min_i a_i$,

$$\left[\max_i - \min_i\right] \delta w_i^n \le (1-\epsilon) \left[\max_i - \min_i\right] \delta w_i^{n-m}, \tag{3.7}$$

which implies that w^n converges to, say, $\bar{w}$. Hence W^n converges to, say, $\bar{W}$, and the limits satisfy (3.2a). Hence, (3.2) holds with $\bar{\gamma} = \bar{w}_{j_0}$. ■

The Gauss-Seidel procedure. The Gauss-Seidel form of (3.1) is, in order $i = 1, 2, \ldots,$

$$\begin{aligned} W_i^n = \sup_{v_i} \inf_{u_i} \Bigg[&\sum_{j=1}^{i-1} p_{ij}(u_i, v_i) W_j^n \\ &+ \sum_{j=i}^{N} p_{ij}(u_i, v_i) \left[W_j^{n-1} - W_{j_0}^{n-1}\right] + k_i(u_i, v_i)\Bigg], \end{aligned} \tag{3.8}$$

Recall the definition of $Q(u, v)$ and $\hat{K}(u, v)$ from Section 2. Then, in matrix notation, (3.8) can be written as

$$\begin{aligned} &W^n = \sup_v \inf_u \left[Q(u, v) w^{n-1} + \hat{K}(u, v)\right], \\ &w^n = W^n - W_{j_0}^n \mathbf{e}. \end{aligned} \tag{3.9}$$

The condition (1.3) is no longer sufficient for convergence. For arbitrary controls $\{\hat{u}^n, \hat{v}^n\}$, let $q_{ij}^{(n)}(\hat{u}^n, \hat{v}^n; \cdots; \hat{u}^1, \hat{v}^1)$ denote the i, jth element of $Q(\hat{u}^n, \hat{v}^n) \cdots Q(\hat{u}^1, \hat{v}^1)$. We now require the additional condition that there are $\epsilon > 0, j_0$, and an integer m, such that for all controls $\{\hat{u}^n, \hat{v}^n\}$ and all i,

$$q_{ij_0}^{(m)}(\hat{u}^m, \hat{v}^m; \hat{u}^{m-1}, \hat{v}^{m-1}; \cdots; \hat{u}^1, \hat{v}^1) \ge \epsilon > 0. \tag{3.10}$$

The condition is discussed below the theorem.

Discussion of (3.10). Consider a one dimensional reflected diffusion on the finite interval $[A, B], B > A$, and let the variance be strictly positive. Approximate this by an N-dimensional Markov chain via the methods of [11]. The reflecting states are 1 and N, which correspond to A and B, resp. If the discretization interval is small enough, then each state communicates with its immediate neighbors only, with probabilities that are bounded away from zero, uniformly in the controls. Then $\inf_{u,v,i} q_{i,2}(u, v) > 0$ and we can use any $m \ge 1$ and $j_0 = 2$ in (3.10). This is a consequence of the form of the Gauss-Seidel iteration, which connects states to those that are lower in the order of the iteration. An analogous result holds for the multidimensional case, if the diffusion being approximated is non-degenerate. See [11] for details concerning the approximation, which is the same for the game problem.

Theorem 3.2. *w^n converges to the value $\bar{\gamma}$ of the game.*

Proof. The proof is just an adaptation of that of Theorem 3.1, analogously to the way that the proof of Theorem 2.1 is an adaptation of the proof of the convergence of the classical procedure (2.2) of value iteration for the discounted cost problem. Let u^n, v^n be the selected values in (3.8) or (3.9). Then the inequalities (3.3) hold with $(Q, \hat{K})$ replacing (P, K). Analogously to the development in Theorem 3.1, this and (3.10) imply (3.7) and the theorem. ■

IV. Conclusions.

For solving optimization problems for control and games for finite-state Markov chain models via value iteration, the Gauss-Seidel method is faster than the Jacobi procedure. The proof of convergence for the control problem is well known, but was not available for the game problem. For the problem of games, it is shown that the Gauss-Seidel procedure converges for the discounted, optimal stopping, and ergodic cost criteria.

References

[1] D.P. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice-Hall, Englewood Cliffs, NJ, 1987.

[2] M. Breton, J. A. Filar, A. Haurie, and T. A. Schultz. On the computation of equilibria in discounted stochastic dynamic

games. In T. Basar, editor, *Lecture Notes in Mathematical and Economic Systems, Vol 265*, pages 64–87. Springer-Verlag, New York and Berlin, 1986.

[3] M. Breton and P. L'Écuyer. Approximate solutions to continuous stochastic games. In R. P. Hämäläinen and H. K. Ehtamo, editors, *Lecture Notes in Control and Information Sciences, Vol 156*, pages 257–264. Springer-Verlag, Berlin and New York, 1991.

[4] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, Berlin and New York, 1996.

[5] A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Management Science*, 12:359–370, 1966.

[6] H.J. Kushner. Numerical approximations for stochastic differential games: The ergodic case. *SIAM J. Control Optim.*, 42:1911-1933, 2003

[7] H.J. Kushner. *Introduction to Stochastic Control Theory*. Holt, Rinehart and Winston, New York, 1972.

[8] H.J. Kushner. Numerical methods for stochastic differential games. *SIAM J. Control Optim.*, 41:457–486, 2002.

[9] H.J. Kushner and S.G. Chamberlain. Finite state stochastic games: Existence theorems and computational procedures. *IEEE Trans. on Automatic Control*, 14:248–255, 1969.

[10] H.J. Kushner and S.G. Chamberlain. On stochastic differential games: Sufficient conditions that a given strategy be a saddle point and numerical procedures for the solution of the game. *J. Math. Anal. Appl.*, 26:560–575, 1969.

[11] H.J. Kushner and P. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag, Berlin and New York, 1992. Second edition, 2001.

[12] H.J. Kushner and A.J. Kleinman. Numerical methods for the solution of degenerate nonlinear equations arising in optimal stochastic control theory. *IEEE Trans. Automat. Contr.*, 13:344–353, 1968.

[13] L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. U.S.A.*, 39:1095–1100, 1953.

[14] B. Tolwinski. Solving dynamic games via Markov chain approximations. In R. P. Hämäläinen and H. K. Ehtamo, editors, *Lecture Notes in Control and Information Sciences, Vol 156*, pages 265–274. Springer-Verlag, Berlin and New York, 1991.

[15] D.J. White. Dynamic programming, Markov chains and the method of successive approximations. *J. Math. Anal Appl.*, 6:373–376, 1963.