Adaptive Aggregation Methods for Infinite Horizon Dynamic Programming

DIMITRI P. BERTSEKAS, FELLOW, IEEE, AND DAVID A. CASTAÑON, MEMBER, IEEE

Abstract—**We propose a class of iterative aggregation algorithms for solving infinite horizon dynamic programming problems. The idea is to interject aggregation iterations in the course of the usual successive approximation method. An important new feature that sets our method apart from earlier proposals is that the aggregate groups of states change adaptively from one aggregation iteration to the next, depending on the progress of the computation. This allows acceleration of convergence in difficult problems involving multiple ergodic classes for which methods using fixed groups of aggregate states are ineffective. No knowledge of special problem structure is utilized by the algorithms.**

I. Introduction

CONSIDER a Markov chain with finite state-space $S = \{1, \cdots, n\}$. Let $x(t)$ denote the state of the chain at stage t. Assume that there is a finite decision space U, and that, for each state $x(t)$ and decision $u(t)$ at stage t, the state transition probabilities are given and are independent of t. Let $\alpha \in (0, 1)$ be a discount factor and $g(x(t), u(t))$ be a given cost function of state and decision. Let $\mu: S \to U$ denote a stationary control policy. The infinite horizon discounted optimal control problem consists of selecting the stationary control policy which minimizes, for all initial states i, the cost

$$J_\mu(i) = E\left\{\sum_{t=0}^{\infty} \alpha^t g(x(t), \mu(x(t))) | x(0) = i, \mu\right\}. \quad (1)$$

The optimal cost vector J^* of this problem is characterized as the unique solution of the dynamic programming equation [1]

$$J^* = \min_\mu \{g_\mu + \alpha P_\mu J^*\}. \quad (2)$$

Here the coordinates of J^* are $J^*(i) = \min_\mu J(i)$, g_μ is the vector with coordinates $g(i, \mu(i))$, P_μ is the transition probability matrix corresponding to μ, and the minimization is considered separately for each coordinate.

One of the principal methods for solving the problem is the policy iteration algorithm which iterates between a policy improvement step

$$\mu^n = \arg\min_\mu \{g_\mu + \alpha P_\mu J^{n-1}\} \quad (3)$$

yielding a new policy μ^n, and a policy evaluation step that finds the cost vector J^n corresponding to policy μ^n by solving the equation

$$J^n = g_{\mu^n} + \alpha P_{\mu^n} J^n. \quad (4)$$

Manuscript received February 1, 1987; revised January 25, 1988 and May 3, 1988. This paper is based on a prior submission of March 6, 1986. Paper recommended by Associate Editor, C. Y. Chong. This work was supported by the Office of Naval Research under Contract N00014-84-C-0577.

D. P. Bertsekas is with the Department of Electrical Engineering and Computer Science, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

D. A. Castañon is with Alphatech, Inc., Burlington, MA 01803.

IEEE Log Number 8927354.

Equation (4) is a linear $n \times n$ system which can be solved by a direct method such as Gaussian elimination. In the absence of specific structure, the solution requires $O(n^3)$ operations, and is impractical for large n. An alternative, suggested in [11] and [12] and widely regarded as the most computationally efficient approach for large problems, is to use an iterative technique for the solution of (4), such as the successive approximation method; this requires only $O(n^2)$ per iteration for dense matrices P (see the survey [2]). It appears that the most effective way to operate this type of method is not to insist on a very accurate iterative solution of (4). Two points relevant to the present paper are as follows.

1) The choice of iterative method for solving approximately (4) is open.

2) For convergence of the overall scheme, it is sufficient to terminate the iterative method at a vector J such that a norm of the residual vector

$$J - (g_{\mu^n} + \alpha P_{\mu^n} J)$$

is reduced by a certain factor over the corresponding norm of the starting residual

$$J^{n-1} - (g_{\mu^n} + \alpha P_{\mu^n} J^{n-1})$$

obtained when the policy improvement step of (3) is carried out.

This paper proposes a new iterative aggregation method for solving (4) as per 1) above. Its rate of convergence can be superior to that of other competing methods, particularly for difficult problems where there are multiple ergodic classes corresponding to the transition matrix P_{μ^n}. Its convergence is assured through the use of safeguards that enforce a guaranteed reduction of the residual vector norm as per 2) above. We have been unable to prove convergence without the use of these safeguards. On the other hand, our computational experiments indicate that the safeguards are seldom needed, and do not contribute appreciable to a deterioration of the rate of convergence of the method.

Several authors have proposed the use of aggregation–disaggregation ideas for accelerating the convergence of iterative methods for the solution of (4) (Miranker [4], Chatelin and Miranker [5], Schweitzer, Puterman, and Kindle [6], Verkhovsky [7], and Mendelshohn [8]). In [5], Chatelin and Miranker described the basic aggregation technique and derived a bound for the error reduction. However, they did not provide a specific algorithm for selecting the directions of aggregation or disaggregation. In [7], Verkhovsky proved the convergence of an aggregation method which used the current estimate of the solution J as a direction of aggregation, and a positive vector as the direction for disaggregation. This idea was extended in [6] by selecting fixed segments of the current estimate J as directions for aggregation, and certain nonnegative vectors as directions for disaggregation.

There is an important difference between the aggregation algorithms described in this paper and those developed by the previous authors. In our work, aggregation and disaggregation directions are selected adaptively based on the progress of the algorithm. In particular, *the membership of a particular state in an aggregate group changes dynamically* throughout the itera-

tions. *States with similar magnitude of residual are grouped together at each aggregation step* and, because the residual magnitudes change drastically in the course of the algorithm, the group membership of the states can also change accordingly. This is in contrast with the approach of [6], for example, where the aggregate groups are fixed through all iterations. We show via experiments and some analysis that the adaptive aggregate group formation feature of our algorithm is essential in order to achieve convergence acceleration for difficult problems involving multiple ergodic classes. For example, when P_μ is the $n \times n$ identity matrix no algorithm with fixed aggregate groups can achieve a geometric convergence rate better than α. By constrast, our algorithm converges at a rate faster than $2\alpha/m$ where m is the number of aggregate groups. We point out, however, that we have been unable to establish analytically a superior rate of convergence for the adaptive aggregation method over fixed aggregate group methods. This remains an interesting subject for investigation.

The rest of the paper is organized as follows. In Section II we provide some background material on iterative algorithms for the solution of (4), including bounds on the solution error. In Section III, we derive the equations of aggregation and disaggregation as in [5], and obtain a characterization of the error reduction produced by an aggregation step. In Section IV, we describe and motivate the adaptive procedure used to select the directions of aggregation and disaggregation. Section V analyzes in detail the error in the aggregation procedure when two aggregate groups are used. Throughout the paper we emphasize discounted problems. Our aggregation method extends, however, to average cost Markovian decision problems and in Section VI we describe the extension. In Section VII, we discuss and justify the general iterative algorithm combining adaptive aggregation steps with successive approximation steps. Section VIII presents experimental results.

II. Successive Approximation and Error Bounds

For the sake of simplicity, we will drop the argument μ from (4), thereby focusing on obtaining an iterative solution to the equation

$$J = T(J) \tag{5a}$$

where the mapping $T: R^n \to R^n$ is defined by

$$T(J) \triangleq g + \alpha P J. \tag{5b}$$

A *successive approximation iteration* on a vector J simply replaces J with $T(J)$. The successive approximation method for the solution of (5) starts with an arbitrary vector J, and sequentially computes $T(J)$, $T^2(J)$, $\cdots$. Since P is a stochastic matrix (and hence has spectral radius of 1) and $\alpha \in (0, 1)$, it follows that T is a sup-norm contraction mapping with modulus α. Hence, we have

$$\lim_{k \to \infty} T^k(J) = J^* \tag{6}$$

where J^* is the solution of (5) and T^k is the composition of the mapping T with itself k times. The rate of convergence in (6) is geometric at a rate α, which is quite slow when α is close to 1.

The rate of convergence can often be substantially improved using some error bounds due to McQueen [9] and Porteus [3] (see [1] for a derivation). These bounds are based on the residual difference of $T(J)$ and J. Let $J(i)$ denote the ith component of a vector J. Let γ and β be defined as

$$\gamma = \min_i [T(J)(i) - J(i)] \tag{7a}$$

$$\beta = \max_i [T(J)(i) - J(i)]. \tag{7b}$$

Then, the solution J^* of (1) satisfies

$$T(J)(i) + \frac{\alpha\gamma}{1-\alpha} \le J^*(i) \le T(J)(i) + \frac{\alpha\beta}{1-\alpha} \tag{8}$$

for all states i. Furthermore, the bounds of (8) are monotonic and approach each other at a rate equal to the complex norm of the subdominant eigenvalue of αP, as discussed in [2] and shown in Section IV of this paper. Hence, the iterations can be stopped when the difference between the lower and upper bounds in (8) is below a specified tolerance for all states i. The value of J^* in this case is approximated by selecting a value between the two bounds.

There are also several variations of the successive approximation method such as Gauss–Seidel iteration, successive overrelaxation [10], and Jacobi iteration [2]. Depending on the problem at hand these schemes may converge faster than the successive approximation method. However, their rate of geometric convergence is often close to α when α is large and P has more than one ergodic class, in which case the subdominant eigenvalue of P has a norm of unity.

III. Aggregation Error Estimates

The basic principle of aggregation–disaggregation is to approximate the solution of (5a) by solving a smaller system of equations obtained by lumping together the states of the original system into a smaller set of aggregate states. We have a vector J and we want to make an additive correction to J of the form Wy, where y is an m-dimensional vector and W is an $n \times m$ matrix, so that

$$J + Wy \approx J^*. \tag{9}$$

In addition to W, our method makes use of another matrix Q. We will later assume that $Q = (W^{\mathrm{T}} W)^{-1} W^{\mathrm{T}}$ (superscript T denotes transpose), but it is worthwhile to postpone this assumption for later so as to develop the following error equations in generality. We thus assume the following.

Assumption 1: Q is an $m \times n$ matrix, and W is an $n \times m$ matrix, chosen so that $Q(I - \alpha P)W$ is nonsingular, and $QW = I$ where I is the m-dimensional identity.

From (5), we get

$$T(J) - J = (I - \alpha P)(J^* - J). \tag{10}$$

Multiplying this equation on the left by Q yields

$$Q(T(J) - J) = Q(I - \alpha P)(J^* - J). \tag{11}$$

We want to choose y so that $J^* - J$ is approximately equal to Wy as in (9). On the basis of (11), we see that a reasonable choice of y is the unique solution of the following $m \times m$ system obtained by replacing $J^* - J$ with Wy in (11):

$$Q(T(J) - J) = Q(I - \alpha P)Wy \tag{12}$$

or, using the fact $QW = I$,

$$Q(T(J) - J) = (I - \alpha QPW)y.$$

Thus, we define

$$y = (I - \alpha QPW)^{-1} Q(T(J) - J)$$

and consider the vector J_1 defined by [cf. (9)]

$$J_1 = J + Wy = J + W(I - \alpha QPW)^{-1} Q(T(J) - J). \tag{13}$$

The conversion of (10) to the lower dimensional (12) is known as the *aggregation step*. The *disaggregation step* is the use of (13) to approximate the solution J^*. Note that there is no claim or guarantee that J_1 approximates well J^*; this depends on the choice of the subspace W which is the key for the success of aggregation

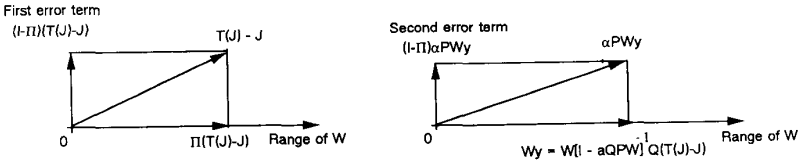


Fig. 1. Geometric illustration of the two error terms of (18). The matrix Π projects orthogonally on the range space of W. Note that if the range of W is invariant under P, the second error term is zero.

methods. If $J - J^*$ lies on the range space of W, then $J_1 = J^*$. Generally, J_1 will be close to J^* if $(J - J^*)$ nearly lies on the range space of W.

After obtaining J_1 using (13), the aggregation method performs a successive approximation iteration on it yielding

$$T(J_1) = T(J) + \alpha PWy \tag{14}$$

[this improves the quality of the solution and is also a necessary first step for the subsequent aggregation step as seen from (13)]. In some cases it is desirable to perform several successive approximation iterations between aggregation steps (see the discussion of Section IV). We thus define the *iterative aggregation method* as a sequence of iterations of the form of (13) with each pair of consecutive iterations possibly separated by one or more successive approximation iterations. Thus, an iteration of the iterative aggregation method replaces J by $T^k(J_1)$, where J_1 is given by (13), and k is some nonnegative integer. The method for choosing W will be discussed in the next section; methods for choosing k will be discussed in Section VII.

To understand the properties of the iterative aggregation method it is important to characterize the error $T(J_1) - J^*$ in terms of the error $J - J^*$. From (14) we get

$$T(J_1) - J^* = (T(J) - J) + (J - J^*) + \alpha PWy \tag{15}$$

which, using (10) and (12), yields

$$T(J_1) - J^* = \alpha P\{I - W(I - \alpha QPW)^{-1} Q(I - \alpha P)\}(J - J^*). \tag{16}$$

Equation (16) is in effect the equation obtained by Chatelin and Miranker [5] to characterize the error obtained by additive corrections based on Galerkin approximations. It applies to general linear equations where the matrix P is not necessarily stochastic. In order to better understand this equation, we will derive an expression for the residual obtained after an aggregation-disaggregation step. Define the matrix

$$\Pi = WQ \tag{17}$$

which is a projection on the range space of W. Generally, Π is not an orthogonal projection but with the choice $Q = (W^T W)^{-1} W^T$ that will be used later in this paper, Π becomes the orthogonal projection matrix on the range of W. From (16) and (10) we get

$$\begin{aligned} T(J_1) - J_1 &= (I - \alpha P)\{I - W[Q(I - \alpha P)W]^{-1} Q(I - \alpha P)\}(J^* - J) \\ &= \{I - (I - \alpha P)W[Q(I - \alpha P)W]^{-1} Q\}(I - \alpha P)(J^* - J) \\ &= (I - \Pi)(T(J) - J) + \{W[I - \alpha QPW] - (I - \alpha P)W\} \\ &\quad \cdot [I - \alpha QPW]^{-1} Q(T(J) - J) \\ &= (I - \Pi)(T(J) - J) + \alpha(I - \Pi)PW[I - \alpha QPW]^{-1} Q \\ &\quad \cdot (T(J) - J) \\ &= (I - \Pi)(T(J) - J) + \alpha(I - \Pi)PWy. \end{aligned} \tag{18}$$

Equation (18) is the basic error equation which we will be working with. There are two error terms on the right side of (18) (see Fig. 1). *Our subsequent choice of W and Q will be based on trying to minimize an estimate of the first error term on the right above.* We generally estimate errors using the pseudonorm

$$F(J) = \text{Max}_i\,(J(i)) - \text{Min}_i\,(J(i)). \tag{19}$$

Since the scalar $F(T(J) - J)$ is proportional to the difference between the upper and lower bounds in (8), we see that reducing $F(T(J) - J)$ to 0 is equivalent to having the upper and lower bounds converge to each other, thereby obtaining J^*. The second error term in (18) is a measure of how well the action of the stochastic matrix P is represented by the aggregation-disaggregation projections based on W. Note that if P maps the range of W into itself, the second term is zero since, from (17) and the condition $QW = I$ of Assumption 1, we have $(I - \Pi)W = 0$. Hence, the second term is small when the range of W is closely aligned with an invariant subspace of P. When this is not the case, the inverse in this second term introduces a tendency for instability. Despite this fact, it will be seen that the effect of this term can be adequately dealt with.

IV. Adaptive Choice of the Aggregation Matrices Based on Residual Size

We introduce a specific choice of Q and W. Partition the state-space $S = \{1, 2, \cdots, n\}$ into m disjoint sets G_j, $j = 1, \cdots, m$ (also called aggregate groups). Define the vectors w_j with ith coordinates given by

$$\begin{aligned} w_j(i) &= 1 \qquad \text{if } i \in G_j \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{20}$$

Let the matrices W and Q be defined by

$$W = [w_1, \cdots, w_m] \tag{21}$$

$$Q = (W^T W)^{-1} W^T. \tag{22}$$

Note that $W^T W$ is a diagonal matrix with $i - i$th entry equal to the number of elements in group G_i. If one of the groups is empty, then we can view the inverse above as a pseudoinverse.

Lemma 1: Assume Q and W are defined by (20)–(22). Then:

a) $QW = I$;

b) $P_a \triangleq QPW$ is a stochastic matrix;

c) Q and W satisfy Assumption 1.

Proof:

a) Immediate from the definition of (22).

b) By straightforward calculation we can verify that the (i, j)th element of P_a is

$$[P_a]_{ij} = \frac{1}{|G_i|} \sum_{k \in G_i} \sum_{m \in G_j} p_{km}$$

where $|G_i|$ is the number of states in G_i. It follows that $[P_a]_{ij} \geq 0$ for all i, j, and

$$\sum_{j=1}^{m} [P_a]_{ij} = 1, \qquad \text{for all } i = 1, \cdots, m.$$

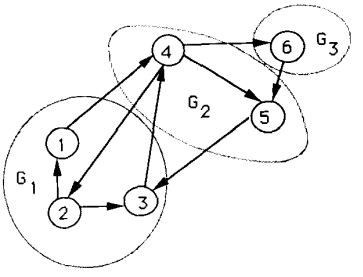


Fig. 2. Illustration of the aggregated Markov chain associated with the transition matrix $P_a = QPW$. The aggregate groups are $G_1 = \{1, 2, 3\}$, $G_2 = \{4, 5\}$, $G_3 = \{6\}$ and they correspond to states of the aggregated Markov chain. The transition probability from state G_i to state G_j equals the sum of all transition probabilities from states in G_i to states in G_j. An aggregation step can be interpreted as a policy evaluation step involving the aggregated Markov chain.

Therefore, P_a is a stochastic matrix.

c) The eigenvalues of P_a lie within the unit disk, so, in view of $\alpha < 1$, the matrix $I - \alpha P_a$ cannot have a zero eigenvalue and must therefore be invertible. This combined with part a) shows that Assumption 1 is satisfied. Q.E.D.

Fig. 2 illustrates the "aggregated Markov chain" corresponding to the stochastic matrix P_a and identifies its states with aggregate groups. This chain provides an insightful interpretation of the aggregated system of (12). By writing this system as

$$Q(T(J)-J)=(I-\alpha P_a)y$$

and by comparing it to the system of (10) we see that y is the cost vector corresponding to the aggregated Markov chain, and to a cost per stage equal to $Q(T(J) - J)$ the ith component of which is the average residual

$$\frac{1}{|G_i|}\sum_{k\in G_i}[T(J)(k)-J(k)]$$

over the ith aggregate group of states. Thus, the aggregation iteration solves in effect a (lower dimensional) dynamic programming equation corresponding to the aggregated Markov chain.

We now describe the method for selecting the aggregate groups. We write (18) as

$$T(J_1)+J_1=R_1(J)+R_2(J) \tag{23}$$

where

$$R_1(J)=(I-\Pi)(T(J)-J) \tag{24a}$$

$$R_2(J)=\alpha(I-\Pi)PW(I-\alpha QPW)^{-1}Q(T(J)-J). \tag{24b}$$

We want to select the partition G_j, $j = 1, \cdots, m$ so that $F[R_1(J)]$ is minimized. For a given value of $F(T(J) - J)$, and number of aggregate groups m, the following procedure, based on residual size, is minimax optimal against the worst possible choices of P and J. The idea is to select G_j so that the variation of residuals within each group is relatively small.

Consider

$$\gamma=\min_i[T(J)(i)-J(i)];\ \beta=\max_i[T(J)(i)-J(i)].$$

Divide the interval $[\gamma, \beta]$ into m equal length intervals, of length L, where

$$L=(\beta-\gamma)/m=(F(T(J)-J))/m. \tag{25}$$

Then, for $j < m$, we select

$$G_j=\{i\,|\,\gamma+(j-1)L\le(T(J)-J)(i)<\gamma+jL\},\qquad j<m \tag{26a}$$

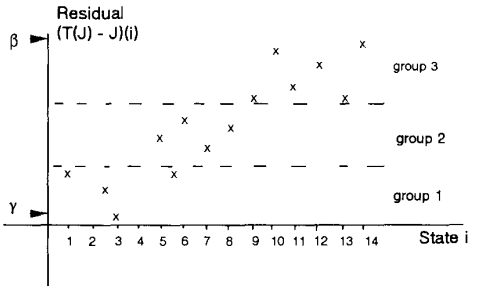


Fig. 3. Formation of aggregate groups is based on magnitude of the residuals. Here the three aggregate groups are obtained by dividing the residual range into three equal portions and grouping together the states with residuals in the same portion.

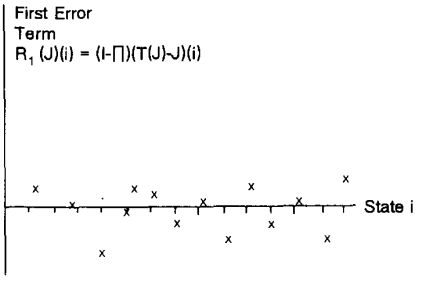


Fig. 4. Illustration of the first error term $R_1(J)$ for the case of the residuals of Fig. 3. $R_1(J)$ is obtained from $(T(J) - J)$ by subtracting the average residual over the group that contains state i.

and we select

$$G_m=\{i\,|\,\gamma+(m-1)L\le(T(J)-J)(i)\le\beta\}. \tag{26b}$$

To understand the idea behind this choice, note that if $j(i)$ is the index of the group containing state i and $|G_{j(i)}|$ is the number of states in $G_{j(i)}$, the ith coordinate of a vector $\Pi x = W(W^TW)^{-1}W^Tx$ [cf. (15) and (22)] can be calculated to be

$$(\Pi x)(i)=\sum_{k\in G_{j(i)}}\frac{x(k)}{|G_{j(i)}|}, \tag{27}$$

i.e., the average value of Πx over the group $G_{j(i)}$. Therefore, the ith coordinate of $R_1(J) = (I - \Pi)(T(J) - J)$ is the difference of the residual of state i and the average residual of the group containing state i. As a result of the choice of (25) and (26), the coordinates of $R_1(J)$ are also relatively small.

Fig. 3 illustrates the choice of G_j for a typical $T(J) - J$ using three aggregate groups. In Fig. 4, we display the vector $R_1(J)$. Note that the spread between the maximum element and the minimum element has been reduced significantly. We have the following estimate.

Lemma 2: Let G_j be defined by (25) and (26). Then, for $m > 1$,

$$\frac{F[R_1(J)]}{F[T(J)-J]}\le\frac{2}{m}. \tag{28}$$

Proof: From (27), $\Pi(T(J) - J)$ is the vector of average values of residuals within each group G_j. The operation $(I - \Pi)(T(J) - J)$, as shown in Fig. 4, subtracts the average value of the residuals in each group from the value of the residuals in each group. Since all of the residuals in each group belong to the same interval in $[\gamma, \beta]$, so does the average value, which establishes that each coordinate of $(I - \Pi)(T(J) - J)$ lies between $-L$ and L.

Therefore, using (25), we have

$$F[(I-\Pi)(T(J)-J)] \leq 2L = 2F(T(J)-J)/m \tag{29}$$

which proves the result. Q.E.D.

We note that the argument in the proof above can be refined to give the improved estimate

$$\frac{F[R_1(J)]}{F[T(J)-J]} \leq \frac{2\lfloor 0.5n \rfloor}{m(\lfloor 0.5n \rfloor + 1)} \tag{30}$$

where $\lfloor x \rfloor$ denotes the largest integer less than x. For large n, the improvement is small. Also, the bound above is a worst-case estimate. In practice, one usually gets a reduction factor better than $1/m$ (as opposed to $2/m$). This has been verified computationally and can also be deduced from the proof of Lemma 2.

Lemma 2 establishes that with our choice of W and Q we get a substantial reduction in the error term $R_1(J)$. Hence, the aggregation step will work best in problems where the second term $R_2(J)$ is small. To illustrate this, consider the following examples.

Example 1: $P = I$, the $n \times n$ identity.

In this case, $R_2(J) = 0$ because $PW = W$. Hence, the aggregation–disaggregation step reduces the spread between the upper and lower bounds in (7) and (8) as

$$F[T(J_1)-J_1] \leq \frac{2F[T(J)-J]}{m}. \tag{31}$$

In this case, the geometric rate of convergence is accelerated by a minimum factor of $2/m$.

Example 2: $m = 1$, $W = e$ where e is the unit vector $e^T = [1, 1, \cdots, 1]$.

In this case, we obtain a scheme known as the error sum extrapolation [2]. Starting from J, a successive approximation step is used to compute $T(J)$. Then, an aggregation step is used to compute $T(J_1)$ directly as

$$T(J_1)(i) = T(J)(i) + \frac{\alpha}{n(1-\alpha)} \sum_{i=1}^{n} (T(J)-J)(i).$$

This aggregation step is followed by a sequence of successive approximation steps and aggregation steps. The rate of convergence of this method can be established using (18). The residual produced by the second successive approximation step is given by

$$\begin{aligned} T(T(J_1)-J_1) &= \alpha P(R_1(J)+R_2(J)) \\ &= \alpha P(I-\Pi)(T(J)-J) \end{aligned}$$

since $R_2(J)$ vanishes (P is a stochastic matrix and $Pe = e$). After n repetitions of successive approximation and aggregation steps, the residual r_n will be

$$\begin{aligned} r_n &= \alpha^n [P(I-\Pi)]^n (T(J)-J) \\ &= \alpha^n P(I-\Pi)P^{n-1}(T(J)-J) \end{aligned} \tag{32}$$

because from (27), $P\Pi = \Pi$ which implies that $(I - \Pi)P(I - \Pi) = (I - \Pi)P$. Consider a decomposition of $P^{n-1}(T(J) - J)$ along the invariant subspaces of P. There is a subspace corresponding to a unity eigenvalue that is spanned by e, and the component of $P^{n-1}(T(J) - J)$ along that subspace is annihilated by $(I - \Pi)$ [cf. (27)]. Therefore, r_n will converge to 0 geometrically at a rate determined by the largest complex norm of eigenvalues of αP in a direction other than e (the subdominant eigenvalue norm).

Example 3: P is block-diagonal and the aggregate groups are aligned with the ergodic classes. In this case we assume that P has multiple ergodic classes and no transient states. By reordering states if necessary, we can assume that P has the form

$$P = \text{diag}\,\{P^1, P^2, \cdots, P^r\}. \tag{33}$$

We assume also that each aggregate group G_j, $j = 1, \cdots, m$ consists of ergodic classes of states (no two states of the same ergodic class can belong to different groups). The matrix W then has the form

$$W = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & & \cdots & & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & & 0 \\ & & & & \vdots & & & & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \cdots & 1 \end{bmatrix}^T$$

and it is easily seen that $PW = W$. Therefore, the second error term $R_2(J)$ vanishes and the favorable rate estimate of (31) again holds. Note that it is not necessary that each aggregate group contain a single ergodic class. This restriction would be needed for fast convergence if the aggregate groups were to remain fixed throughout the computation.

The case of a block-diagonal matrix P is important for several reasons. First, block-diagonal matrices P present the most difficulties for the successive approximation method, regardless of whether the McQueen–Porteus error bounds are employed. Second, we can expect that algorithmic behavior on block-diagonal matrices will be replicated to a great extent on matrices with weakly coupled or sparsely coupled blocks. This conjecture is substantiated analytically in the next section and experimentally in Section VII.

The favorable rate of convergence described above is predicated on the alignment of the ergodic classes and the aggregate groups. The issue of effecting this alignment is therefore important. We first remark that even if this alignment is not achieved perfectly, we have observed experimentally that much of the favorable convergence rate can still be salvaged, particularly if an aggregation step is followed by several successive approximation steps. We provide some related substantiation in the next section, but hasten to add that we do not fully understand the mechanism of this phenomenon. We next observe that for a block-diagonal P, the eigenvectors corresponding to the dominant unity eigenvalues are of the form

$$e_j = [0 \cdots 0\ 1 \cdots 1\ 0 \cdots 0]^T \qquad j = 1, \cdots, r$$

where the unit entries correspond to the states in the jth ergodic class. Suppose that we start with some vector J and apply k successive approximation steps. The residual thus obtained will be

$$T^k(J) - T^{k-1}(J) = (\alpha P)^{k-1}(T(J)-J) \tag{34}$$

and for large k, it will be nearly a linear combination of the dominant eigenvectors. This means that $T^k(J) - T^{k-1}(J)$ is nearly constant over each ergodic class. As a result, if aggregate groups are formed on the basis of the residual $T^k(J) - T^{k-1}(J)$ and (25) and (26), they will very likely be aligned with the ergodic classes of P. This fact suggests that several successive approximation steps should be used between aggregation steps, and provides the motivation for the algorithm to be given in Section VII.

V. Adaptive Aggregation with Two Groups

The preceding section showed that the contribution of the second error term $R_2(J)$ of (18) is crucial for the success of our aggregation method. The analysis of this contribution seems very difficult in general, but the case where $m = 2$ is tractable and is given in this section. Experiment and some analysis show that the qualitative conclusions drawn from this case carry over to the more general case where $m > 2$. Assume that W, Q have been selected according to (20)–(22). By appropriate renumbering of

the states, assume that W is of the form

$$W=\begin{bmatrix}1 & \cdots & 1 & 0 & \cdots & 0\\ 0 & \cdots & 0 & 1 & \cdots & 1\end{bmatrix}^{T}.$$

Let k be the number of elements in the first group. Then a straightforward calculation shows that

$$P_a=\begin{bmatrix}1-b & b\\ a & 1-a\end{bmatrix} \tag{35}$$

where

$$b=\frac{1}{k}\sum_{i=1}^{k} b_i \tag{36a}$$

$$c=\frac{1}{n-k}\sum_{i=k+1}^{n} c_i \tag{36b}$$

$$b_i=\sum_{j=k+1}^{n} P_{ij}, \qquad i=1, \cdots, k \tag{37a}$$

$$c_i=\sum_{j=1}^{k} P_{ij}, \qquad i=k+1, \cdots, n. \tag{37b}$$

The right eigenvectors and eigenvalues of P_a are

$$v_1=[1\ \ 1]^{T};\ v_2=[1\ \ -c/b]^{T} \tag{38}$$

$$\lambda_1=1;\ \lambda_2=1-b-c \tag{39}$$

assuming $b \neq 0$. If $b = 0$, then v_2 can be chosen as

$$v_2=[0\ \ 1]^{T} \tag{40}$$

and $\lambda_1 = 1$, $\lambda_2 = 1 - c$. From (22) and the form of W we obtain

$$Q=\begin{bmatrix}1/k & 0\\ 0 & 1/(n-k)\end{bmatrix} W^{T}. \tag{41}$$

We can decompose the term $Q(T(J) - J)$ of (18) into its components along the eigenvectors v_1, v_2, as

$$Q(T(J)-J)=a_1v_1+a_2v_2. \tag{42}$$

We have $(I - \alpha P_a)v_1 = (1 - \alpha)v_1$ from which we obtain

$$W(I-\alpha P_a)^{-1}v_1=(1-\alpha)^{-1}Wv_1. \tag{43}$$

Hence,

$$\alpha(I-\Pi)PW(I-\alpha P_a)^{-1}v_1=\alpha(1-\alpha)^{-1}(I-\Pi)PWv_1=0$$

and it follows that the only contribution to $R_2(J)$ comes from the term a_2v_2 in (42). Using (35), (38), and (39) we obtain

$$(I-\alpha P_a)^{-1}v_2=[1-\alpha+\alpha(b+c)]^{-1}v_2. \tag{44}$$

Thus, using (24b), we obtain

$$\begin{aligned}R_2(J)&=\alpha(I-\Pi)PW(I-\alpha P_a)^{-1}a_2v_2\\ &=\alpha a_2(PW-WP_a)[1-\alpha+\alpha(b+c)]^{-1}v_2.\end{aligned} \tag{45}$$

From (34)–(37), we can calculate the $(i, 1)$ element of the matrix $PW - WP_a$ to be

$$\begin{aligned}(PW-WP_a)(i,1)&=b-b_i \qquad \text{if } i\leq k\\ &=-c+c_i \qquad \text{if } i>k.\end{aligned} \tag{46}$$

Similarly,

$$(PW-WP_a)(i,2)=-(PW-WP_a)(i,1).$$

Thus, from (45)

$$R_2(J)=\alpha a_2F(v_2)h \tag{47}$$

where h is the vector with coordinates

$$\begin{aligned}h(i)&=\frac{b-b_i}{1-\alpha+\alpha(b+c)} \qquad \text{if } i\leq k\\ &=\frac{c-c_i}{1-\alpha+\alpha(b+c)} \qquad \text{if } i>k\end{aligned} \tag{48}$$

and $F(v_2) = 1 + c/b$ [cf. (19) and (38)]. From (36), (37), and (48) we see that in order for the coordinates of h to be small, the probabilities b_i and c_i should be uniformly close to their averages b and c. If this is not so, then at least some coordinates of $R_2(J)$ will be substantial, and it is interesting to see what happens after a successive approximation step is applied to $R_2(J)$. The corresponding residual term is the vector

$$q=\alpha PR_2(J).$$

From (47) and (48) we see that the ith coordinate of q is

$$q(i)=\frac{\alpha^2a_2F(v_2)}{1-\alpha+\alpha(b+c)}\left[\sum_{j=1}^{k} p_{ij}(b-b_j)+\sum_{j=k+1}^{n} p_{ij}(c_j-c)\right]. \tag{49}$$

Since b and c are the averages of b_j and c_j, respectively, we see that the coordinates of q can be small even if the coordinates of h are large. For example, if P has a totally random structure (e.g., all elements are drawn independently from a uniform distribution), then for large n the coordinates of q will be very small by the central limit theorem. There are several other cases where either h or q (or both) are small depending on the structure of P. Several such examples will now be discussed. All of these examples involve P matrices with subdominant eigenvalues close to unity for which standard iterative methods will converge very slowly.

Case 1: P has uniformly weakly coupled classes of states which are aligned with the aggregate groups.

The matrix P in this case has the form

$$P=\begin{bmatrix}P^1 & P^2\\ P^3 & P^4\end{bmatrix} \tag{50}$$

where P^1 is $k \times k$ and the elements of P^2 and P^3 are small relative to the elements of P^1 and P^4. From (36), (37), (47), and (48) we see that if b and c are considerably smaller than $(1 - \alpha)$, then $R_2(J) \approx 0$. This will also happen if the terms b_i and c_i of (37) are all nearly equal to their averages b and c, respectively. Even if $R_2(J)$ is not near zero, from (49) we see that $q \approx 0$ if the size of the elements within each row of P^1, P^2, P^3, and P^4 is nearly uniform.

What happens when the groups identified by the adaptive aggregation process are not perfectly aligned with the block structure of P? We examine this case next.

Case 2: P block diagonal with the upper $k \times k$ submatrix not corresponding to the block structure of P.

Without loss of generality, assume that $i = 1, \cdots, m_1 \leq k$ are all elements of one group of ergodic classes of P, while $i = m_2 + 1, \cdots, n$, $m_2 \geq k$, are elements of the complementary group of ergodic classes. Note that the states $m_1 \leq i \leq m_2$ are not aligned with their ergodic classes in the adaptive aggregation process.

In this case, we have

$$b_i = \sum_{j=k+1}^{m_2} P_{ij} \qquad \text{if } i \le m_1$$

$$= \sum_{j=m_2+1}^{n} P_{ij} \qquad \text{if } k \ge i > m_1 \tag{51}$$

$$c_i = \sum_{j=1}^{m_1} P_{ij} \qquad \text{if } m_2 \ge i > k$$

$$= \sum_{j=m_1+1}^{k} P_{ij} \qquad \text{if } m_2 < i \le n. \tag{52}$$

Suppose

$$k - m_1 \approx m_2 - k;\ k \approx n/2;\ k - m_1 \ll k \tag{53}$$

so that the aggregate groups are nearly aligned with the block structure of P. The ergodic classes corresponding to group 1 consist of the set of states $i = 1, \cdots, m_1$ and $i = k+1, \cdots, m_2$, while the remaining states correspond to the ergodic classes in group 2. From (51) we see that b_i will tend to be small for $i = 1, \cdots, m_1$ and large for $i = m_1 + 1, \cdots, k$. Similarly, c_i will tend to be small for $i = m_2 + 1, \cdots, n$ and large for $i = k+1, \cdots, m_2$. It follows from (48) that

$$h(i) > 0 \qquad \text{if } i = 1, \cdots, m_1 \text{ or } i = k+1, \cdots, m_2$$

$$h(i) < 0 \qquad \text{otherwise.} \tag{54}$$

Hence, $R_2(J)$ is contributing terms of opposite sign to the ergodic classes in groups 1 and 2. By following the aggregation step with repeated successive approximation iterations, this contribution will be smoothed throughout the ergodic classes. Thus, the next aggregation step will be able to identify groups which are aligned with the block structure of P, thereby reducing the error as in Case 1.

Case 3: P has sparsely-coupled classes of states.

In this case, P has the general form

$$P = \begin{bmatrix} P^1 & P^2 \\ P^3 & P^4 \end{bmatrix} \tag{55}$$

where elements of P^1, P^4, P^2, P^3 are of the same order, and P^1, P^4 are dense while P^2, P^3 are very sparse. Assume that the groups are aligned with the block structure of P. Then we have

$$b_i = \sum_{j=1}^{n-k} P^2_{ij} \qquad \text{if } i \le k \tag{56}$$

$$c_i = \sum_{j=1}^{k} P^3_{ij} \qquad \text{if } i > k. \tag{57}$$

As in Case 1, if b_i and c_i are small [of the order of $(1 - a)$], or vary little from the corresponding averages b and c, then $R_2(J) \approx 0$. If the size of the elements within P^1 and P^4 is nearly uniform, then from (49) we see that $q \approx 0$. Furthermore, the behavior observed in Case 2 is replicated in this case and, when the aggregate groups are not aligned with the block structure of the P matrix, the term $R_2(J)$ forces the next aggregation step to be better aligned with the block structure of P.

In conclusion, the cases studied in this section indicate that, for classes of problems where there are multiple eigenvalues with norm near unity, a combination of several successive approximation steps, followed by an aggregation step, will minimize the contribution of $R_2(J)$ to the error, and thereby accelerate the convergence of the iterative process as in Lemma 2. In Section VII, we formalize these ideas in terms of an overall iterative algorithm.

VI. Extension to the Average Cost Problem

The aggregation procedure described in Section III can also be used in the policy evaluation step of the policy iteration algorithm in the average cost case. Here the cost vector for a stationary policy μ is given by

$$J_\mu = \lim_{T \to \infty} (1/T) E \left\{ \sum_{t=0}^{T} g(x(t), \mu(x(t))) | \mu \right\}. \tag{58}$$

As in the discounted cost case, the average cost incurred by policy μ satisfies the linear equation (see [1] for a detailed derivation)

$$J_\mu + h_\mu = g_\mu + P_\mu h_\mu. \tag{59}$$

The vector h_μ is the differential cost incurred by policy μ. In what follows we drop the subscript μ.

The solution of (59) can be computed under certain conditions using the successive approximation method [1]. We fix a state which for concreteness is taken to be State 1. Starting with an initial guess h^0 for the differential cost, the successive approximation method computes h^{n+1} as

$$h^{n+1} = T(h^n) - e e_1^{\mathrm{T}} T(h^n) \tag{60}$$

where $T(h)$ is defined by

$$T(h) = g + Ph,$$

$e = [1, 1, \cdots, 1]^{\mathrm{T}}$ and $e_1 = [1, 0, \cdots, 0]^{\mathrm{T}}$ is the coordinate vector corresponding to the fixed state 1. Equation (60) can be written as

$$h^{n+1} = g_A + P_A h^n \tag{61}$$

where

$$g_A = (I - e e_1^{\mathrm{T}}) g$$

$$P_A = (I - e e_1^{\mathrm{T}}) P.$$

We assume that all eigenvalues of P except for a single unity eigenvalue lie strictly within the unit circle (see [1] for a method that works under the weaker assumption that P has a single ergodic class). A straightforward calculation shows that $P_A^2 = P_A P$ from which we obtain $P_A^k = P_A P^{k-1}$ for all $k > 0$. Since P_A annihilates the eigenvector e corresponding to the unit eigenvalue of P, it follows that the eigenvalues of P_A all lie strictly inside the unit circle, guaranteeing the convergence of the iteration of (61). Furthermore, the rate of convergence is specified by the subdominant eigenvalue of P.

Note that the iteration in (61) is identical to the discounted cost iteration

$$h^{n+1} = g + \alpha P h^n$$

except that g_A replaces g and P_A replaces αP. Thus, the aggregation and error equations of Section III can be extended to the average cost problem using the above substitutions. The following lemma establishes that the choice of the matrices Q and W used in Section IV result in a well-posed aggregate problem provided the fixed State 1 forms an aggregate group by itself.

Lemma 3: Assume Q and W are defined by (20)–(22) with the set G_1 consisting of just State 1, and that all eigenvalues of P except for a single unity eigenvalue lie strictly within the unit circle. Then the aggregate matrix QP_AW has spectral radius less than unity.

Proof: It is straightforward to verify that

$$QP_A W = (I - e_m e_{1,m}^{\mathrm{T}}) P_a \tag{62}$$

where $P_a = QPW$ is the aggregate stochastic matrix defined in Lemma 1b, e_m is the m-dimensional vector of all 1's, and $e_{1,m}$ is the m-dimensional vector with first coordinate 1, and all other coordinates 0. Therefore, as earlier, we obtain $(QP_A W)^2 = (QP_A W)P_a$ from which

$$(QP_A W)^k = (QP_A W)P_a^{k-1} = (I - e_m e_{1,m}^{\mathrm{T}})P_a^k, \qquad \text{for all } k > 0. \tag{63}$$

We have $P_a^k = (QPW)^k = QP^k W$ for all $k > 0$, and from this we obtain that P_a has all its eigenvalues strictly within the unit circle except for a single unity eigenvalue. Using this fact, (63), and the fact that $(I - e_m e_{1,m}^{\mathrm{T}})$ annihilates the eigenvector e_m corresponding to the single unity eigenvalue of P_a, we see that $QP_A W$ must have all its eigenvalues strictly within the unit circle. Q.E.D.

Equation (62) illustrates that the solution to the aggregate linear equation is the solution of an aggregate average-cost problem with transition probabilities P_a. The equations for the aggregation step are as follows.

$$h_1 = h + W(I - QP_A W)^{-1} Q(g_A + P_A h - h).$$

Using this equation we obtain error equations similar to (23) and (24), indicating that the same choice of Q and W will result in similar acceleration as in the discounted case. This has been verified by the experiments of Section VIII.

VII. Iterative Aggregation Algorithms

The method for imbedding our aggregation ideas into an algorithm is straightforward. Each iteration consists of one or more successive approximation steps, followed by an aggregation step. The number of successive approximation steps in each iteration may depend on the progress of the computation.

One reason why we want to control the number of successive approximation steps per iteration is to guarantee convergence. In contrast to a successive approximation step, the aggregation step need not improve any measure of convergence. We may wish therefore to ensure that sufficient progress has been made via successive approximation between aggregation steps to counteract any divergence tendencies that may be introduced by aggregation. Indeed, we have observed experimentally that the error $F(T(J) - J)$ often tends to deteriorate immediately following an aggregation step due to the contribution of $R_2(J)$, while unusually large improvements are made in the next few successive approximation steps. This is consistent with some of the analytical conclusions of the previous section. An apparently effective scheme is to continue with successive approximation steps as long as $F(T(J) - J)$ keeps decreasing by a "substantial" factor.

One implementation of the algorithm will now be formally described.

Step 0 (Initialization): Choose initially a vector J, and scalars $\epsilon > 0$, β_1, β_2 in (0, 1), $\omega_1 \approx \infty$ and $\omega_2 \approx \infty$.

Step 1 (Successive Approximation Step): Compute $T(J)$.

Step 2 (Termination Test): If $F(T(J) - J) < \epsilon$, stop and accept

$$T(J) + (1/2)\alpha(1-\alpha)^{-1}\,[\max_i (T(J) - J)(i) - \min_i (T(J) - J)(i)]$$

as the solution [cf. the bounds in (8)]. Else go to Step 3.

Step 3 (Test for an Aggregation Step): If

$$F(T(J) - J) \le \omega_1 \tag{64}$$

and

$$F(T(J) - J) \ge \omega_2 \tag{65}$$

set $\omega_1 := \beta_1 F(T(J) - J)$ and go to Step 4. Else, set $\omega_2 := \beta_2 F(T(J) - J)$, $J := T(J)$ and go to Step 1.

Step 4 (Aggregation Step): Form the aggregate groups of states G_j, $j = 1, \cdots, m$ based on $T(J) - J$ as in (26). Compute $T(J_1)$ using (13) and (14). Set $J := T(J_1)$, $\omega_2 \approx \infty$, and go to Step 1.

The purpose of the test of (65) is to allow the aggregation step only when the progress made by the successive approximation step is relatively small (a factor no greater than β_2). The test of (64) guarantees convergence of the overall scheme. To see this, note that the test of (64) ensures that, before Step 4 is entered, $F(T(J) - J)$ is reduced to a level below the target ω_1, and ω_1 converges to zero when an infinite number of aggregation steps are performed. If only a finite number of aggregation steps are performed, the algorithm reduces eventually to the convergent successive approximation method.

An alternative implementation is to eliminate the test of (65) and perform an aggregation step if (64) is satisfied and the number of consecutive iterations during which an aggregation step was not performed exceeds a certain threshold.

VIII. Computational Results

A large number of randomly generated problems with 100 states or less were solved using the adaptive aggregation methods of this paper. The conclusion in summary is that problems that are easy for the successive approximation method (single ergodic class, dense matrix P) are also easy for the aggregation method; but problems that are hard for successive approximation (several weakly coupled blocks, sparse structure) are generally easier for aggregation and often dramatically so.

Tables I and II summarize representative results relating to problems with 75 states grouped in three blocks of 25 each. The elements of P are either zero or randomly drawn from a uniform distribution. The probability of an element being zero was controlled thereby allowing the generation of matrices with approximately prescribed degree of density. Table I compares various methods on block diagonal problems with and without additional transient states, which are full (100 percent) dense, and 25 percent dense within each block. Table II considers the case where the blocks are weakly coupled with 2 percent coupling (size of elements outside the blocks is on the average 0.02 times the average size of the elements inside the blocks), and the case where the blocks are 100 percent coupled (all nonzero elements of P have nearly the same size). Each entry in the tables is the number of steps for the corresponding method to reach a prescribed difference (10^{-6}) between the upper and lower bounds of Section II. Our accounting assumes that an aggregation step requires roughly twice as much computation as a successive approximation step which is quite realistic for most problems. Thus, the entries for the aggregation methods represent the sum of the number of successive approximation and twice the number of aggregation steps. In all cases the starting vector was zero, and the components of the cost vector g were randomly chosen on the basis of a uniform distribution over [0, 1].

The methods are successive approximation [with the error bounds of (8)], and six aggregation methods corresponding to all combinations of three and six aggregate groups, and three, five, and ten successive approximation steps between aggregation steps. Naturally these methods do not utilize any knowledge about the block structure of the problem.

Table I shows the dramatic improvement offered by adaptive aggregation as predicted by Example 3 in Section IV. The improvement is substantial (although less pronounced) even when there are transient states. Generally speaking, the presence of transient states has a detrimental effect on the performance of the

TABLE I
DISCOUNT FACTOR 0.99, BLOCK DIAGONAL P, 3 BLOCKS, 25 STATES EACH, TOLERANCE FOR STOPPING: 1.0 E-6

	Successive Approx. (SA)	3 SA Steps per aggr. step 3 aggr. groups	3 SA Steps per aggr. step 6 aggr. groups	5 SA Steps per aggr. step 3 aggr. groups	5 SA Steps per aggr. step 6 aggr. groups	10 SA Steps per aggr. step 3 aggr. groups	10 SA Steps per aggr. step 6 aggr. groups
100% density 0 trans. states	1195	11	11	15	15	25	25
100% density 20 trans. states	1225	31	16	58	17	170	27
25% density 0 trans. states	1212	23	26	29	23	27	27
25% density 20 trans. states	1197	186	105	177	72	194	50

TABLE II
DISCOUNT FACTOR 0.99, COUPLED P, 3 BLOCKS, 25 STATES EACH, TOLERANCE FOR STOPPING: 1.0 E-6

	Successive Approx. (SA)	3 SA Steps per aggr. step 3 aggr. groups	3 SA Steps per aggr. step 6 aggr. groups	5 SA Steps per aggr. step 3 aggr. groups	5 SA Steps per aggr. step 6 aggr. groups	10 SA Steps per aggr. step 3 aggr. groups	10 SA Steps per aggr. step 6 aggr. groups
100% density 2% coupling	170	17	17	22	22	37	37
25% density 2% coupling	167	38	33	36	32	40	40
100% density 100% coupling	6	7	7	8	7	7	7
3% density 100% coupling	66	56	66	60	64	64	66

aggregation method when there are multiple ergodic classes. Repeated successive approximation steps have the effect of making the residuals nearly equal across the ergodic classes; however, the residuals of transient states tend to drift at levels which are intermediate between the corresponding levels for the ergodic classes. As a result, even if the alignment of aggregate groups and ergodic classes is perfectly achieved, the aggregate groups typically contain a mixture of ergodic classes and transient states. This has an adverse effect on both error terms of (18). As the results of Table I show, it appears advisable to increase the number of aggregate groups m when there are transient states. It can be seen also from Table I that the number of successive approximation steps performed between aggregation steps influences the rate of convergence. Generally speaking, there seems to be a problem-dependent optimal value for this number which increases as the problem structure deviates from the ideal block diagonal structure. For this reason it is probably better to use an adaptive scheme to control this number in a general purpose code as discussed in Section VII.

Table II shows that as the coupling between blocks increases (and consequently, the modulus of the subdominant eigenvalue of P decreases), the performance of both successive approximation and adaptive aggregation improves. When there is full coupling between the blocks the methods become competitive, but when the coupling is weak the aggregation methods hold a substantial edge as predicted by our analysis.

An interesting issue is the choice of the number of aggregate groups m. According to Lemma 2, the first error term $R_1(J)$ of (24) is reduced by a factor proportional to m at each aggregation step. This argues for a large value of m, and indeed we have often found that increasing m from two to something like three or four leads to a substantial improvement. On the other hand, the benefit from reduction of $R_1(J)$ is usually exhausted when m rises above four, since then the effect of the second error term $R_2(J)$ becomes dominant. Also the aggregation step involves the solution of the m-dimensional linear system of (12), so when m is large the attendant overhead can become substantial. In the extreme case where $m = n$ and each state forms by itself an aggregate group, the solution is found in a single aggregation step. The corresponding dynamic programming method is then equivalent to the policy iteration algorithm.

TABLE III
AVERAGE COST INFINITE HORIZON PROBLEMS, COUPLED P, 3 BLOCKS, 25 STATES EACH, STOPPING TOLERANCE 1.0 E-6

	Successive Approx. (SA)	Adaptive aggregation 2 aggr. groups	Adaptive aggregation 3 aggr. groups
100% density 2% coupling	184	62	13
25% density 2% coupling	164	26	26
100% density 1% coupling	338	64	13
25% density 2% coupling	307	43	27
100% density 0.1% coupling	LARGE	71	10
25% density 0.1% coupling	LARGE	50	26

Table III shows the performance of adaptive aggregation algorithms for the infinite horizon average cost case. In these algorithms, the number of successive approximation steps between aggregation steps was determined adaptively as in the algorithm of Section VII, by performing aggregation steps whenever the rate of error reduction of successive approximation steps was slower than 0.9. Table III shows that, while the rate of convergence of successive approximation methods is very sensitive to the strength of the coupling between blocks of P, the rate of

convergence of the adaptive aggregation methods remains largely unaffected. In particular, the results for the adaptive algorithms using only two aggregate groups illustrate that major reductions in computation time can be achieved even if the number of aggregate groups is smaller than the number of strongly-connected components of the stochastic matrix P.

IX. Conclusion

In this paper, we have developed aggregation techniques for the iterative solution of large-scale linear systems of equations arising in dynamic programming. The distinguishing feature of our method is its adaptive character; the aggregation directions are selected on the basis of the residual vector of the iteration, and can vary among iterations. Computational results using our method show impressive acceleration of the convergence rate over the ordinary successive approximation method, particularly for problems with weakly-coupled classes of states. This acceleration is obtained even when the number of aggregate states used by the method is much smaller than the number of weakly-coupled classes of states in the original problem. Thus, it is not necessary to know *a priori* the special structure of the problem for the method to be effective.

The intuitive reason for the improved convergence rate is as follows: based on monitoring the residuals, the adaptive aggregation iteration identifies some of the eigenspaces along which convergence is slow. Each aggregation–disaggregation step then removes most of the component of the iteration error along these eigenspaces. At each iteration, the errors along different slowly-converging directions are removed. However, because these directions change from one iteration to the next, it is sufficient to use a small number of aggregate states.

Extensions of the adaptive aggregation method to obtain iterative solutions of general linear systems of equations are straightforward; however, the specific choice of aggregation matrix W in this paper is based on the ergodic eigenstructure of stochastic matrices, and should be reconsidered for general linear equations. Other potential extensions include development of higher order adaptive aggregation schemes which use lagged values of the residual vectors in order to identify aggregation–disaggregation directions, and analysis of the convergence rates of these aggregation schemes.

References

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

[2] E. L. Porteus, "Overview of iterative methods for discounted finite Markov and semi-Markov decision chains," in *Recent Developments in Markov Decision Processes*, R. Hartley, L. C. Thomas, and D. J. White, Eds. New York: Academic, 1980.

[3] ——, "Some bounds for discounted sequential decision processes," *Management Sci.*, vol. 18, 1971.

[4] W. L. Miranker, "Hierarchical relaxation," *Comput.*, vol. 23, 1979.

[5] F. Chatelin and W. L. Miranker, "Acceleration by aggregation of successive approximation methods," *Linear Alg. Appl.*, vol. 43, 1982.

[6] P. J. Schweitzer, M. Puterman, and K. W. Kindle, "Iterative aggregation-disaggregation procedures for solving discounted semi-Markovian reward processes," *Operat. Res.*, vol. 33, pp. 589–606, 1985.

[7] B. S. Verkhovsky, "Smoothing system optimal design," IBM Research Division, Yorktown Heights, NY, Tech. Rep. 6085, 1976.

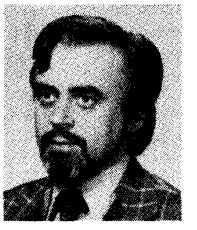
[8] R. Mendelshohn, "An iterative aggregation procedure for Markov decision processes," *Operat. Res.*, vol. 30, 1982.

[9] J. McQueen, "A modified dynamic programming method for Markovian decision problems," *J. Math. Anal. Appl.*, vol. 14, 1966.

[10] H. J. Kushner and A. J. Kleinman, "Accelerated procedures for the solution of discounted Markov control problems," *IEEE Trans. Automat. Contr.*, vol. AC-16, 1971.

[11] M. L. Puterman and M. C. Shin, "Modified policy iteration algorithms for discounted Markov decision problems," *Management Sci.*, vol. 24, 1979.

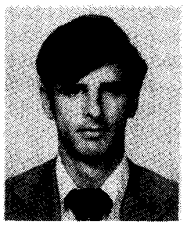
[12] ——, "Action elimination procedures for modified policy iteration algorithms," *Operat. Res.*, vol. 30, 1982.



Dimitri P. Bertsekas (S'70–SM'77–F'84) received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1971.

He has held faculty positions with the Department of Engineering-Economic Systems, Stanford University, from 1971 to 1974 and with the Department of Electrical Engineering, University of Illinois, Urbana, from 1974 to 1979. Since 1979 he has been on the faculty of the Massachusetts Institute of Technology, where he is currently Professor of Electrical Engineering and Computer Science. He also consults regularly with private industry. His research interests are in the areas of estimation and control of stochastic systems, linear, nonlinear and dynamic programming, data communication networks, and parallel and distributed computation, and has written numerous papers in each of these areas. He is the author of *Dynamic Programming and Stochastic Control* (New York: Academic, 1976), *Constrained Optimization and Lagrange Multiplier Methods* (New York: Academic, 1982), *Dynamic Programming: Deterministic and Stochastic Models* (Englewood Cliffs, NJ: Prentice-Hall, 1987); and coauthor of *Stochastic Optimal Control: The Discrete-Time Case* (New York: Academic, 1978), *Data Networks* (Englewood Cliffs, NJ: Prentice-Hall, 1987), and *Parallel and Distributed Computation: Numerical Methods* (Englewood Cliffs, NJ: Prentice-Hall, 1989).

Dr. Bertsekas has held editorial positions for several journals.



David A. Castañon (S'68–M'79) was born in La Habana, Cuba, on November 24, 1950. He received the B.S. degree in electrical engineering, with high honors, from Tulane University, New Orleans, LA, in 1971, and the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology, Cambridge, in 1976.

From 1976 to 1982 he was a Research Associate with the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology, Cambridge. He is presently a Senior Scientist at Alphatech, Inc., Burlington, MA. His research interests include stochastic control and estimation, team theory, dynamic game theory, large scale systems, and numerical optimization.