SIAM J. DISC. MATH.
Vol. 5, No. 3, pp. 386–397, August 1992

008

MEAN PASSAGE TIMES AND NEARLY UNCOUPLED MARKOV CHAINS*

REFAEL HASSIN† AND MOSHE HAVIV‡

Abstract. Let $P(0) \in R^{n\times n}$ be a stochastic matrix representing transition probabilities in a Markov chain. Also, for a matrix $A \in R^{n\times n}$ whose row-sums are zero, let $P(\varepsilon) \equiv P(0) + \varepsilon A$ be stochastic and irreducible for all $0 < \varepsilon \leq \varepsilon_{\max}$, for some $\varepsilon_{\max}$. Finally, let $M(\varepsilon)$ be a matrix whose (i,j) entry is the mean passage time from state i to state j when transitions are governed by $P(\varepsilon)$. When the Markov chain associated with $P(0)$ is decomposable into a number of independent chains plus a set of transient states, some of the entries of $M(\varepsilon)$ have singularities at zero. The orders of these poles define timescales associated with the process when ε is small. An algorithm is developed for computing these orders. The only input required is the supports of $P(0)$ and A, making the problem a combinatorial one. Finally, it is shown how the orders of the poles of $M(\varepsilon)$ at zero play a role in developing series expansions for $\pi(\varepsilon)$, the stationary distribution of $P(\varepsilon)$.

Key words. Markov chains, mean passage time, nearly uncoupled

AMS(MOS) subject classifications. 60J10, 68C05

1. Introduction. Let $P(0) \in R^{n\times n}$ be a stochastic matrix representing transition probabilities in a Markov chain. Let $A \in R^{n\times n}$ have zero row-sums. For all real ε, $0 < \varepsilon \leq \varepsilon_{\max}$, assume that $P(\varepsilon) \equiv P(0) + \varepsilon A$ are stochastic matrices representing transition probabilities in irreducible Markov chains. (The irreducibility assumption of $P(\varepsilon)$ is without loss of generality. However, we still require that the structure of the chains is the same for all ε, $0 < \varepsilon \leq \varepsilon_{\max}$. Of course, such an $\varepsilon_{\max}$ exists.) Note that we do not assume irreducibility of $P(0)$, and, in fact, we are interested in the case where the Markov chain associated with it is decomposable into a number of independent recurrent classes plus a set of transient states. In that case and for small values of ε, the matrices $P(\varepsilon)$ and the associated Markov chains are called *nearly uncoupled* or *nearly completely decomposable.*

For $0 < \varepsilon \leq \varepsilon_{\max}$, let $M(\varepsilon) \in R^{n\times n}$ be such that $M_{ij}(\varepsilon)$ is the mean passage time from state i to state j when transitions are governed by $P(\varepsilon)$. It is clear that, if $P(0)$ is decomposable, then there exist pairs (i,j) such that $\lim_{\varepsilon\to 0} M_{ij}(\varepsilon) = \infty$. As $M(\varepsilon)$ admits a Laurent series expansion (see §7), this implies that some entries of $M(\varepsilon)$ have singularities at zero; namely, they have poles there. The orders of these poles represent various timescales in the nearly uncoupled Markov chain. For example, if the order of the pole of $M_{ij}(\varepsilon)$ at zero is two, then, for a process that initiates at i, the event of hitting state j for the first time occurs after an expected time, which is of the same order of magnitude as $1/\varepsilon^2$.

The main purpose of this paper is to show that the problem of finding the orders of the above-mentioned poles is combinatorial and to develop an algorithm for computing them. Specifically, we suggest an algorithm whose input is the two binary matrices describing the supports of $P(0)$ and A. The output of the algorithm is the orders of the poles.

Finally, we discuss an application of the orders of the poles. Let $\pi(\varepsilon)$ be the

* Received by the editors September 6, 1990; accepted for publication (in revised form) July 25, 1991.

† Department of Statistics, Tel Aviv University, Tel Aviv 69978, Israel.

‡ Department of Statistics, Hebrew University of Jerusalem. Present address, Department of Econometrics, University of Sydney, N.S.W. 2006, Australia.

stationary distribution of $P(\varepsilon)$. If $\pi(\varepsilon) = \sum_{i=0}^{\infty} \pi^{(i)}\varepsilon^i$, then the series $\{\pi^{(i)}\}_{i=0}^{\infty}$ solves

$$\pi^{(0)}(I - P(0)) = 0,$$

and for $j = 0, 1, \ldots,$

$$\pi^{(j+1)}(I - P(0)) = \pi^{(j)}A.$$

If $P(0)$ is decomposable, then the first system of above equations does not determine $\pi^{(0)}$ uniquely. We show that the minimal number of consecutive sets of equations that must be considered to determine $\pi^{(0)}$ uniquely is $\max_{i,j}(u_{ij} - u_{jj}) + 1$, where u_{ij} is the order of the pole of $M_{ij}(\varepsilon)$ at zero.

Nearly uncoupled systems were introduced by Simon and Ando [18]. These were applied to some Markovian models by Courtois [3]. Also, they were considered in the Russian literature, probably begining with Gaitsgory and Pervozvansky [6]. The question of first passage times is dealt with in [10] and [11], but only for the case where $P(0)$ does not possess transient states. For this case, Latouch and Louchard (see [10] and [11]) show that the orders of the poles of $M(\varepsilon)$ at zero are zero or one. The more general case, where two independent recurrent chains at $P(0)$ are coupled at $P(\varepsilon)$ through states that are transient at $P(0)$, is considered in Delebecque [4], Coderch et al. [1], [2], and Rohlicek and Willsky [14], who developed algorithms for approximating the transient and the long-run behavior of the Markov chains associated with $P(\varepsilon)$ for $0 < \varepsilon \leq \varepsilon_{\max}$. The analysis in the above-mentioned papers is based on Kato's [8] classical perturbation results. The combinatorial version of the algorithm in [14] is given in Rohlicek and Willsky [13]. They solved a problem different from that addressed here. Specifically, let $F_{ij}^{T}(\varepsilon)$ be the probability that a process governed by $P(\varepsilon)$ and initiating at state i hits state j for the first time prior to time T.[1] Then let

$$d_{ij} = \arg\max\{d \mid \lim_{\varepsilon \to 0} F_{ij}^{T/\varepsilon^d}(\varepsilon) > 0 \text{ for some finite } T\}.$$

Next, we present an example showing that u_{ij} and d_{ij} do not necessarily coincide. This example was communicated to us by a referee.

Let

$$P(\varepsilon) = \begin{pmatrix} 0 & \varepsilon & 1-\varepsilon \\ 0 & 1-\varepsilon^2 & \varepsilon^2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here $u_{13} = 1$: although $P_{13}(\varepsilon) = \Theta(1)$, the transition probability ε from state 1 into state 2, and then the additional expected time of $1/\varepsilon^2$ until hitting state 3, leads to this value of u_{13}. On the other hand, $d_{13} = 0$ as for any $T \geq 1$, $F_{13}^{T}(\varepsilon) \geq 1 - \varepsilon$.[2]

The question of series expansions for nearly uncoupled Markov chains is analyzed by Schweitzer [15]–[17] and Haviv and Ritov [7]. Schweitzer showed that the deviation matrix of $P(\varepsilon)$ has a Laurent series expansion around zero, and he gave some implicit forms for it. He also solved for the series expansion of $\pi(\varepsilon)$ and related it to the

[1] In order to ease the exposition, here we allow T to be any nonnegative number and not necessarily an integer.

[2] Note that $P(\varepsilon)$ here was not defined by a linear perturbation. However, the same phenomenon can be seen with a corresponding linear perturbation, but in a Markov chain, where state 2, above, is replaced with three other states.

expansion for the deviation matrix. His method, or, alternatively, the method given in [7], can be used to obtain the order of the poles as a by-product. Since these methods are based on solving n^2, equations they have a larger complexity $O(n^6)$, and they do not explore the combinatorial aspects of the problem. Finally, Langenhop [9] suggested a general method for inverting near singular matrices. As $M(\varepsilon)$ uniquely solves a system of equations, this method can be utilized for the problem discussed here, and then the orders of the poles are obtained as a by-product. However, for computing the orders of the poles, our direct approach is much simpler.

After introducing some terminology in §2, we present the algorithm in §3. Section 4 is devoted to proving the correctness of the algorithm. In §5 we mention a possible generalization of our algorithm to nonlinear perturbations. In §6 we state some background on Markov chains, and we use it in §7 for developing a series expansion for $\pi(\varepsilon)$ around $\varepsilon = 0$. Section 8 concludes the paper with a numerical example.

2. Terminology. A function $f : R_+ \to R_+$ is called $\Theta(\varepsilon^k)$ for some integer k (positive, zero, or negative) if there exist two positive real numbers M_1 and M_2 such that, for all $\varepsilon > 0$ small enough,

$$M_1\varepsilon^k \leq f(\varepsilon) \leq M_2\varepsilon^k.$$

If $f(\varepsilon) = \Theta(\varepsilon^{-k})$, then we say that $f(\varepsilon)$ is of *order of magnitude* k.

The following are immediate:

$$\Theta(\varepsilon^{k_1}) + \Theta(\varepsilon^{k_2}) = \Theta(\varepsilon^{min\{k_1,k_2\}}), \qquad \Theta(\varepsilon^{k_1}) * \Theta(\varepsilon^{k_2}) = \Theta(\varepsilon^{k_1+k_2}).$$

Consider $G = (V, E_e, E_r)$, a finite directed graph with a vertex set V, an edge set $E = E_e \cup E_r$ that may contain loops, and a distinguished vertex $s \in V$. G is *strongly connected*; that is, it contains, for every $i, j \in V$ a directed $i - j$ path. We call edges of E_e "*e-edges*" (epsilon edges) and those of E_r "*r-edges*" (regular edges). These sets are disjoint. A path in G is called an "*r-path*" if it consists of r-edges only (similarly for an "*r-cycle*"). An "*r-component*" is defined to be a maximal strongly connected subgraph of (V, E_r). Note that an r-component may also be a single vertex. For a subset $C \subset V$, we let $\delta(C) = \{(i,j) \in E | i \in C, j \notin C\}$ be the set of its (outward-oriented) *boundary edges.*

In terms of $P(\varepsilon)$, which was defined in the Introduction, we define the following graph G. Each vertex of V corresponds to a state. Each edge $(i,j) \in E$ is associated with a *transition probability* $p_{ij}(\varepsilon)$. If $p_{ij}(\varepsilon) = \Theta(1)$, then $(i,j) \in E_r$, while, if $p_{ij}(\varepsilon) = \Theta(\varepsilon)$, then $(i,j) \in E_e$. If $P_{ij}(\varepsilon) = 0$, then $(i,j) \notin E$. Obviously, every $i \in V$ satisfies $\sum_{j|(i,j)\in E} p_{ij}(\varepsilon) = 1$.

Fix a state $s \in S$. Let $m_i(\varepsilon)$ denote the expected time until s is first reached when the initial state is i ($m_s(\varepsilon)$ is the expected return time to s, given that it is also the initial state). It is shown that $m_i(\varepsilon) = \Theta(\varepsilon^{-u(i)})$ for some integer $u(i)$ that is zero or positive. Our problem is to compute $u(i)$ for all $i \in V$.

3. The algorithm. The algorithm computes $u(i)$ for $i \neq s$ by first assigning and revising temporary values and, finally, changing them to permanent. The set of vertices with temporary values is denoted by T. In a final step, $u(s)$ is computed from the other values. In the course of the algorithm, a multigraph $G' = (V', E'_r, E'_e)$ is maintained. Then sets $C \subset V'$ are *condensed* into a single vertex c as follows: Edges $(i,j) \in \delta(C)$ are replaced by (possibly parallel) edges (c,j), and, similarly, edges in $\delta(V \setminus C)$ are replaced by (possibly parallel) edges (i,c). Edges (i,j) such that $i,j \in C$ are simply deleted. The correspondence of each edge in G' to the original edge in G is

maintained. The algorithm is stated next. A numerical example is given in §8. The reader may wish to consider this example prior to reading the rest of the paper.

Input: $G = (V, E_e, E_r)$.

Output: $u(i)$, $i \in V$.

Step 1 (Initialization).

Construct a graph $G' = (V', E'_e, E'_r)$ from G by first setting $G' \leftarrow G$, and then deleting all loops $(i,i) \in E_e$ and all edges going out of s. Set $u(i) \leftarrow 0$ and $S(i) \leftarrow \{i\}$ for all $i \in V$.

Step 2 (Elimination of loops).[3]

If G' contains no loops, go to Step 3. Otherwise, let $(i,i) \in E'_r$. Set $E'_r \leftarrow E'_r \setminus \{(i,i)\}$. If $\delta(\{i\}) \cap E'_r = \phi$, set $E'_r \leftarrow E'_r \cup \delta(\{i\})$, $E'_e \leftarrow E'_e \setminus \delta(\{i\})$ and $u(i) \leftarrow 1$. Repeat Step 2.

Step 3 (Condensation of cycles).

If G' contains no directed r-cycles, go to Step 4.

Let C be (the vertex set in V' of) such a cycle.[4] Condense C into a single vertex c.

Case (i). $\delta(C) \cap E'_r \neq \phi$.

Set $u(c) \leftarrow \max\{u(i) | i \in C\}$.

Case (ii). $\delta(C) \subseteq E'_e$.

Set $u(c) \leftarrow 1 + \max\{u(i) | i \in C\}$.

Set $E'_r \leftarrow E'_r \cup \delta(C)$, $E'_e \leftarrow E'_e \setminus \delta(C)$.

Set $S(c) \leftarrow \cup_{i \in C} S(i)$.

Repeat Step 3.

Step 4 (Solution of the problem for r-acyclic graphs).

Set $T \leftarrow V'$. Let $u(j) = \max\{u(i) | i \in T\}$. (Break ties arbitrarily.) Delete j from T (thus turning $u(j)$ into permanent for j). For r-edges (i,j) where $i \in T$, set $u(i) \leftarrow u(j)$. For e-edges (i,j) where $i \in T$, set $u(i) \leftarrow \max\{u(i), u(j) - 1\}$. If $T = \phi$, go to Step 5. Else, repeat Step 4.

Step 5 (Computation of $u(i)$ $i \in V \setminus \{s\}$).

$\{S(v') | v' \in V'\}$ is a partition of V. For each $v \in V$ find $v' \in V'$ such that $v \in S(v')$ and set $u(v) \leftarrow u(v')$.

Step 6 (Computation of $u(s)$).

Set $u(s) \leftarrow \max\{\max\{u(i) | (s,i) \in E_r\}, \max\{u(i) - 1 | (s,i) \in E_e\}\}$.

4. Validation of the algorithm. In this section, we use the following notation; we distinguish between *states*, which correspond to the initial Markov process, and *vertices* $v \in V'$, which correspond to the sets of states $S(v)$; we also distinguish between *transitions*, which correspond to the initial Markov process, and *edges* $(i,j) \in E'_r \cup E'_e$, which correspond to moves of the process between vertices. It should be emphasized that each edge $(i,j) \in E'$ is associated with a transition from a state in $S(i)$ to a state in $S(j)$. It is possible, however, that an r-edge in V' is associated with a transition of a $\Theta(\varepsilon)$ probability, since such changes are performed at Step 3, Case (ii) in the above algorithm.

Let C denote an r-cycle (or an r-component) of G'; We say for a state j that $j \in C$ whenever $j \in S(v)$ for some $v \in C$. Also, let $P_i^*(\varepsilon) = \sum_{j | (i,j) \in \delta(C)} p_{ij}(\varepsilon)$ be the probability of an immediate exit from C, given that the current state i is in C. For

[3] This step can be deleted; instead, loops will be considered as directed cycles in Step 3. It is included here only to simplify the proofs below and, in particular, to guarantee that the inductive assumption in Theorem 4.5 (i), below, holds.

[4] Instead of r-cycles, we may choose r-components.

a vertex $v \in V'$ and for states i and j in $S(v)$, let $Q_{ij}(\varepsilon)$ be the expected number of visits in state $j \in S(v)$ before first exiting $S(v)$, given the current state is $i \in S(v)$. A similar notation is used for cycles instead of vertices. Note that, for simplicity, we omit from the notation the cycle or the vertex. However, from the context, it is clear which is the cycle or the vertex under consideration.

The following two lemmas give the orders of magnitude of the expected number of visits in a state prior to an exit from a cycle. Their proofs are similar, and thus we supply a detailed proof only to one of them.

LEMMA 4.1. *Let C be an r-cycle in the original graph G. Suppose that $\delta(C) \cap E_r \neq \phi$. Then $Q_{ij}(\varepsilon) = \Theta(1)$ for all $i, j \in C$.*

LEMMA 4.2. *Let C be an r-cycle in the original graph G. Suppose that $\delta(C) \subseteq E_e$. Then $Q_{ij}(\varepsilon) = \Theta(\varepsilon^{-1})$ for all $i, j \in C$.*

Proof. Let $T(\varepsilon)$ be the submatrix of $P(\varepsilon)$ representing transition probabilities in an r-cycle C. It is well known that $Q(\varepsilon) = (I - T(\varepsilon))^{-1}$. Hence, $Q_{ij}(\varepsilon)$ is a ratio between two polynomials in ε and therefore it has an integer order of magnitude. Since for all $i, j \in C$ there exists an $i-j$ r-path, the orders of magnitude of all the entries of $Q(\varepsilon)$ are identical. Consequently, it is sufficient to prove that $\sum_{j \in C} Q_{ij}(\varepsilon) = \Theta(\varepsilon^{-1})$. Indeed, since the exit probabilities from all states are at most $\Theta(\varepsilon)$, the time to the first exit from C (regardless of the current state) is stochastically dominated by a geometric random variable whose expectation is $\Theta(\varepsilon^{-1})$. Hence, $\sum_{j \in C} Q_{ij}(\varepsilon)$ is at least $\Theta(\varepsilon^{-1})$. On the other hand, in order for $\sum_{j \in C} Q_{ij}(\varepsilon)$ to be $\Theta(\varepsilon^{-k})$ for $k \geq 2$ there should be at least one $j \in C$ with this property. An r-path from j to an exit state $e \in C$ exists, however, and hence the number of visits to state e is also of the order of magnitude k. This contradicts the fact that the number of visits to state e is at most $1/P_e^{\star}(\varepsilon) = \Theta(\varepsilon^{-1})$. □

The following two lemmas correspond to exit probabilities from a cycle C in the original graph G. Basically, they show that the order of exit probabilities from various states $i \in C$ and via various pairs $(i, j) \in \delta(C)$ does not depend on the entering state.

LEMMA 4.3. *For an r-cycle C in the original graph G, suppose that $\delta(C) \cap E_r \neq \phi$. Then, for every current state in C, an exit from C occurs with probability $\Theta(1)$ for every $(i, j) \in \delta(C) \cap E_r$ and with probability $\Theta(\varepsilon)$ for every $(i, j) \in \delta(C) \cap E_e$.*

Proof. For a current state i, with $(i, j) \in \delta(C) \cap E_r$, the exit probability is, of course, $\Theta(1)$. For all other states k in C, there exists an r-path from k to i, making the probability of an exit through (i, j) also $\Theta(1)$ for any current state k. For $(i, j) \in \delta(C) \cap E_e$, pick an arbitrary $(g, h) \in \delta(C) \cap E_r$. As there is an r-path from g to i, it is easy to see that, conditional on an exit from (i, j) or (g, h), the exit is from (i, j) with probability $\Theta(\varepsilon)$. Extending that to conditioning on an exit completes the proof. □

LEMMA 4.4. *For a cycle C in the original graph G, suppose that $\delta(C) \subseteq E_e$. Let $(i, j) \in \delta(C)$. Then, regardless of the current state in C, exit occurs via (i, j) with probability $\Theta(1)$.*

Proof. We prove that i is the exit state with probability $\Theta(1)$. The claim then follows immediately by considering conditional probabilities. For $i, j \in C$, let $G_{ij}(\varepsilon)$ be the probability of exiting through j, given that i is the entering (or current) state. If $i \neq j$, then

$$G_{ij}(\varepsilon) = \sum_{k \in C} T_{ik}(\varepsilon) G_{kj}(\varepsilon),$$

and otherwise

$$G_{ii}(\varepsilon) = \sum_{k \in C} T_{ik}(\varepsilon) G_{ki}(\varepsilon) + P_i^\star(\varepsilon).$$

Hence $G(\varepsilon) = Q(\varepsilon)\text{diag}(P^\star(\varepsilon))$ where $\text{diag}(P^\star(\varepsilon))$ is a diagonal matrix whose ith diagonal entry equals $P_i^\star(\varepsilon)$. Solving for $G(\varepsilon)$, we obtain that $G_{ij}(\varepsilon) = Q_{ij}(\varepsilon)P_j^\star(\varepsilon)$, which is $\Theta(\varepsilon^{-1})\Theta(\varepsilon) = \Theta(1)$ in the case where j is an exit state. □

THEOREM 4.5. *After each execution of Step 3,*

(i) *$u(c)$ is the order of the expected sojourn time at $S(c)$, regardless of the current state in $S(c)$;*

(ii) *if $\delta(C) \cap E_r' \neq \phi$ (case (i)), then the conditional exit probability from $S(c)$ is $\Theta(1)$ for each transition associated with r-edges in $\delta(C)$, and $\Theta(\varepsilon)$ for those associated with e-edges of $\delta(C)$, regardless of the current state in $S(c)$;*

(iii) *if $\delta(C) \subseteq E_e'$ (case (ii)), then the conditional exit probability from $S(c)$ is $\Theta(1)$ for each transition associated with edges in $\delta(C)$, regardless of the current state in $S(c)$.*

Proof. The proof is by induction on the number of executions of Step 3. Lemmas 4.1–4.4 establish the claims of the theorem for the first execution. Next, we consider an arbitrary execution assuming the theorem holds for the previous executions.

1. Let $v \in C$ be a vertex and let $i \in S(v)$ be a state. Also, let $u_i(v)$ be the order of the expected sojourn time at $S(v)$ given i as the current state. By the induction assumption, $u_i(v) = u(v)$ is independent of i for every $i \in S(v)$. We wish to next show that $u_i(c)$ is independent of i for all $i \in C$ and that $u_i(c) = u(c)$ (the value generated by the algorithm). Suppose that this is not the case. Following the r-edges defining C, at least one corresponds to $v, w \in C$, with $u_g(c) < u_h(c)$, where $g \in S(v)$ and $h \in S(w)$. Two possibilities exist.

(a) We have that $P_{gh}(\varepsilon) = \Theta(1)$. Considering this transition probability, we obtain that $\Theta(\varepsilon^{-u_g(c)}) \geq \Theta(1) * \Theta(\varepsilon^{-u_h(c)})$ so $u_g(c) \geq u_h(c)$, a contradiction.

(b) We have that $P_{gh}(\varepsilon) = \Theta(\varepsilon)$. Since now the transition $g \to h$ is associated with an r-edge, g belonged to some cycle C' that was condensed at some earlier execution of Step 3, Case (ii). By the induction hypothesis on part (iii) of Theorem 4.5, the probability of an exit from C' by an $g \to h$ conditional on an exit from C', is $\Theta(1)$. Again, it follows that $\Theta(\varepsilon^{-u_g(c)})$ is at least $\Theta(1) * \Theta(\varepsilon^{-u_h(c)}) = \Theta(\varepsilon^{-u_h(c)})$, so that $u_g(c) \geq u_h(c)$, a contradiction.

2. Suppose that the Markov process initiates at C, and consider the clock that runs only while transitions between vertices in C or outward of C occur (and thus transitions between states belonging to the same vertex do not progress the clock). This is not a Markov process, as the transition out of a vertex may depend on the entering state. However, by the induction assumptions on parts (ii) and (iii) of the theorem, the orders of these transition probabilities are independent of the entering state. Thus, if C satisfies the condition of Case (i) (respectively, Case (ii)) in Step 3, then, as in Lemma 4.1 (respectively, Lemma 4.2), the expected number of visits at each of the vertices of C (in the new clock) until an exit from C is $\Theta(1)$ (respectively, $\Theta(\varepsilon^{-1})$). Thus, returning to the original clock, the expected time until an exit from C (regardless of the current state in C) is $\Theta(1) * \Theta(\varepsilon^{-\max_{v \in C} u(v)})$ (respectively, $\Theta(\varepsilon^{-1}) * \Theta(\varepsilon^{-\max_{v \in C} u(v)})$), which is of the order of magnitude $\max_{v \in C} u(v)$ (respectively, $1 + \max_{v \in C} u(v)$). Indeed, this is the value computed by algorithm.

3. Parts (ii) and (iii) of the theorem are the counterparts of Lemmas 4.3 and 4.4., respectively, when considering the above-mentioned new clock. For a formal proof, the

argument of these lemmas may be repeated, in addition to the need to condition on the initial state in C. However, by the induction hypothesis, the orders of magnitude are not a function of the entering state and hence the result follows. □

LEMMA 4.6. *Step 4 of the algorithm determines the order of magnitude of the time until absorption at s for all vertices in the final graph.*

Proof. Step 4 of the algorithm is reached when G' contains no r-cycles. Also, note that s can be reached from all vertices via an r-path. Since G' has no r-cycles, then regardless of the initial state, the expected number of moves between vertices until the process reaches s is $\Theta(1)$. To see this, note that the vertices can be numbered in such a way that an r-edge from vertex i to vertex j exists only if $i < j$. In particular, s obtains the maximum index. Hence the probability of moving backward is either zero or $\Theta(\varepsilon)$.

Let $m = \max\{u(i) | i \in V'\}$. Starting at a vertex j with $u(j) = m$, each move to a new vertex contributes expected time until absorption of the order of magnitude of at most m, with this value being achieved at least once. Since, in expectation, there are $\Theta(1)$ such moves, we conclude by a simple renewal argument that the expected time until absorption in state s is $\Theta(\varepsilon^{-m})$. Hence the algorithm determines the correct values for those vertices that were first to be deleted from T. To see that, the order m applies for those vertices from which j can be reached by an r-edge is now immediate.

Consider the case when an e-edge exists from vertex i to j (but no r-path from i to j exists). There is an $\Theta(\varepsilon)$ probability to enter j, and this contributes to the expected time until absorption at state s, while initiating at state i a value of $\Theta(\varepsilon) * \Theta(\varepsilon^{-m}) = \Theta(\varepsilon^{-(m-1)})$. As a larger order cannot be achieved, this is the order of time until absorption at s that such a vertex has. These arguments extend inductively to other vertices to prove the lemma. □

COROLLARY 4.7. *Steps 5 and 6 correctly determine $u(i)$ for all states $i \in V$.*

Proof. The proof for $i \neq s$ is immediate from Lemma 4.6 and part (i) of Theorem 4.5. Then the claim follows for $i = s$ by a simple expectation argument. □

Complexity of the algorithm. The complexity of the algorithm is dominated by Step 3. This step can be executed in $O(n^2)$ time (cf. Fox and Landi [5]), so that an $O(n^3)$-time complexity obtains for a fixed target state and $O(n^4)$ when all target states are required. We suspect, however, that sophisticated dynamic data structures can be used to reduce this bound.

5. Generalization to a nonlinear perturbation. The algorithm can be extended to solve a more general problem. Let $P(0) \in R^{n \times n}$ be a stochastic matrix representing transition probabilities in a Markov chain. Let $A(x) \in R^{n \times n}$ have zero row-sums for $x \in X$. X is a set of positive real numbers that are not necessarily integers. For all real ε, $0 < \varepsilon \leq \varepsilon_{\max}$, assume that $P(\varepsilon) \equiv P(0) + \sum_{x \in X} \varepsilon^x A(x)$ are stochastic matrices representing transition probabilities in irreducible Markov chains. As before, we are interested in computing $u(i)$, the order of magnitude of the expected time to reach a given state s given an initial state i.

The edge-set of the graph G is now associated with the number k_{ij} denoting the order of the transition probability from i to j. The value of k_{ij} is set to zero if $P_{ij}(0) > 0$, and in this case where $(i, j) \in E_r$. Else, k_{ij} is set to the lowest index x for which $A_{ij}(x) > 0$, if such an index exists. Since the sum of the probabilities on the outgoing transitions from a state is 1, there is at least one r-edge (i, j) for every $i \in V$. After deleting the edges, leaving s, it is true, as was obvious in the special case above, that if G' contains no r-cycle then there is an r-path from every $i \neq s$ to s.

The following modifications are needed in the algorithm.

Replace cases (i) *and* (ii) *of Step* 3 *by the following lines.* Set $k \leftarrow \min\{k_{ij} | (i,j) \in \delta(C)\}$. Set $u(c) \leftarrow k + \max\{u(i) | i \in C\}$. Set $k_{ij} \leftarrow k_{ij} - k$ for all $(i,j) \in \delta(C)$. Modify E_r by adding to it edges $(i,j) \in \delta(C)$ that now have $k_{ij} = 0$.

Replace Step 4. Let $u(j) = \max\{u(i) | i \in T\}$. Delete j from T (thus turning $u(j)$ into permanent for j). For edges (i,j) where $i \in T$, set $u(i) \leftarrow \max\{u(i), u(j) - k_{ij}\}$. If $T = \phi$, go to Step 5. Else, repeat Step 4.

Replace Step 6. Set $u(s) \leftarrow \max\{u(i) - k_{si} | (s,i) \in E\}$.

It is obvious that the generalization of the problem is a natural one and that it leads to a simpler presentation of the algorithm. We have chosen to present the special case of linear perturbation because of its more obvious applicability and also because the proofs are somewhat simpler.

Another obvious extension of the problem is that in which the time from entrance to a state to the next transition is state-dependent. In this case, numbers $t(i)$ $i \in V$ are given, representing the order of the expected time from entrance to i till the next transition. Hardly any modification is needed in the algorithm to account for this case. The only change is at the initialization step, where, instead of starting with $u(i) \leftarrow 0$ for all $i \in V$, we start with $u(i) \leftarrow t(i)$ for all $i \in V$.

6. Background on Markov chains. Next, we briefly introduce some preliminaries concerning Markov chains (see Meyer [12]). Let $P \in R^{n \times n}$ be a stochastic matrix representing transition probabilities in the Markov chain. If P is irreducible (or ergodic), then a unique positive probability vector π with $\pi = \pi P$ exists. It is called the *stationary distribution* associated with P. The stochastic matrix P is also associated a matrix M, which is the matrix of mean passage times; namely, M_{ij} is the expected time until the process first hits state j when it initiates at i. In particular, $M_{ii} = 1/\pi_i$. Denote by Y the *deviation matrix* of P. This is the unique matrix Y satisfying the following three requirements:

$$(I - P)Y(I - P) = I - P,$$

$$Y(I - P)Y = Y,$$

$$Y(I - P) = (I - P)Y.$$

Furthermore, for each pair of states i and j,

$$\frac{M_{ij}}{M_{jj}} = \delta_{ij} + Y_{jj} - Y_{ij}, \tag{6.1}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. This relation between the matrices M and Y implies also that $Y_{jj} \geq Y_{ij}$ for all i and j. If P is aperiodic (or regular), then

$$Y = \lim_{N \to \infty} \left[\sum_{t=0}^{N} P^t - (N+1)E \right],$$

where E is matrix whose rows coincide with π. This representation of the deviation matrix gives its probabilistic interpretation. Also, $Y = [I - (P - E)]^{-1} - E$. Finally, zero is an eigenvalue of Y with $\pi Y = \underline{0}$ and with $Y\underline{1} = \underline{0}$, where $\underline{0}$ (respectively, $\underline{1}$) is a vector all its entries are 0 (respectively, 1).

7. Series expansion for the stationary distribution. We now return to nearly uncoupled stochastic matrices with linear perturbation as defined in §§1–4. Let $\pi(\varepsilon)$ be the stationary distribution of $P(\varepsilon)$. Schweitzer [16] showed that, for all $\varepsilon > 0$ small enough,

$$\pi(\varepsilon) = \sum_{i=0}^{\infty} \pi^{(i)} \varepsilon^i$$

for some sequence $\{\pi^{(i)}\}_{i=0}^{\infty}$. Also, $\pi^{(i)} = \pi^{(0)} U^i$ for some matrix U. Hence finding $\pi^{(0)}$ is a key for a complete description of the series expansion of $\pi(\varepsilon)$.

To determine the terms of such a series expansion and, in particular, to find the leading term, consider the identity $\pi(\varepsilon) = \pi(\varepsilon)(P(0) + \varepsilon A)$. It is clear then that $\{\pi^{(i)}\}_{i=0}^{\infty}$ solves the following systems of difference equations:

$$\pi^{(0)}(I - P(0)) = 0,$$

and, for $j = 0, 1, \ldots$

$$\pi^{(j+1)}(I - P(0)) - \pi^{(j)} A = 0.$$

We refer to the above as the *systems of fundamental equations.* If $P(0)$ is irreducible, then the first set of equations $\pi^{(0)}(I - P(0) = 0$ (plus the needed normalization) is sufficient to determine $\pi^{(0)}$ uniquely. However, if $P(0)$ is decomposable, then a larger number of sets of fundamental equations is needed to determine $\pi^{(0)}$ uniquely, and the question of how many sets are needed to determine $\pi^{(0)}$ uniquely arises. By having this value in advance, we can improve the naive approach of adding a set of fundamental equations one at a time and solving the resulting system until a unique (up to a normalization constant) $\pi^{(0)}$ emerges.

We argue that, for a decomposable $P(0)$, the above question is related to the question of mean passage times at $P(\varepsilon)$ for small values of ε and their orders of magnitude. Recall that $M(\varepsilon)$ and $Y(\varepsilon)$ are the mean passage time matrix and the deviation matrix, respectively, of $P(\varepsilon)$. The following result is taken from Schweitzer [16].

LEMMA 7.1. *$M(\varepsilon)$ and $Y(\varepsilon)$ admit Laurent series expansion.*

Proof. It is well known (see, e.g., Meyer [12]) that $M(\varepsilon)$ is the unique matrix in $R^{n \times n}$ satisfying $(I - P(\varepsilon))M(\varepsilon) = J - P(\varepsilon)M_{dg}(\varepsilon)$, where J is a matrix, all its entries are one, and $M_{dg}(\varepsilon)$ is a diagonal matrix whose diagonal coincides with the diagonal of $M(\varepsilon)$. Solving for $M(\varepsilon)$, say by Cramer's rule, then, for each i, j, $M_{ij}(\varepsilon)$ is the ratio between two polynomials in ε and hence it admits a Laurent expansion.[5] This completes the proof for $M(\varepsilon)$. For the matrix $Y(\varepsilon)$ the result follows from the above proof coupled with (6.1). □

It is clear that $M_{ij}(\varepsilon) = \Theta(\varepsilon^{-u_{ij}})$ where u_{ij} is the order of the pole of $M_{ij}(\varepsilon)$. Thus this order can be computed via the algorithm presented in §3.

THEOREM 7.2. *Let u_{ij} be the order of the pole of $M_{ij}(\varepsilon)$ at zero. Similarly, let v_{ij} be the corresponding order of $Y_{ij}(\varepsilon)$.*[6] *Then,*

$$\max_{ij}(u_{ij} - u_{jj}) = \max_{ij} v_{ij}.$$

[5] Moreover, the order of the pole is bounded by n since the polynomial in the numerator cannot have a larger degree.

[6] Note that, in the case of an analytic function, the order is zero.

Proof. As for any irreducible P, the corresponding π and Y satisfy $\pi Y = 0$, and, as $Y_{jj} \geq Y_{ij}$, the maximal order of the poles over $Y_{jj}(\varepsilon) - Y_{ij}(\varepsilon)$ coincides with the corresponding maximization over $Y_{ij}(\varepsilon)$. The proof is now completed by relation (6.1). □

It was shown in Haviv and Ritov [7] that the maximal value over the order of the poles of the entries of $Y(\varepsilon)$ at zero equals the minimal number of sets of fundamental equations needed to determine $\pi^{(0)}$ uniquely, minus one.[7] By Theorem 7.2, this value can be computed from the orders of the poles at zero of the entries of $M(\varepsilon)$. Thus we have the following theorem.

THEOREM 7.3. *The minimal number of sets of fundamental equations needed to construct a system of equations whose corresponding solution to* $\pi^{(0)}$ *is unique (up to a normalizing constant) is* $\max_{ij}(u_{ij} - u_{jj}) + 1$.

8. An example. Let

$$P(\varepsilon) = P(0) + \varepsilon A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Recall that

$$E_r = \{(i,j) | P_{ij}(0) > 0\}, \quad \text{and} \quad E_e = \{(i,j) | A_{ij} > 0\}.$$

This information is summarized by the following graph where r-edges are represented by bold arrows and e-edges by dashed arrows. See Fig. 1.

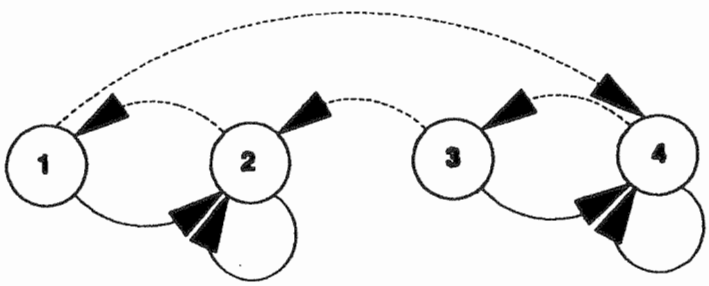

FIG. 1

The Markov chain associated with $P(0)$ is has two independent chains (states 2 and 4, which are absorbing states) and two transient states (states 1 and 3). We proceed through the steps of the algorithm with the target state $s = 4$. Edges emanating form vertex 4 are ignored by the first part of the algorithm, so that we consider Fig. 2, below.

The algorithm initiates with the values $u(1) = u(2) = u(3) = 0$. There is only one r-cycle: (2). All edges emanating from it are e-edges, so that Case (ii) occurs in Step 3. Hence $u(2)$ is set to 1, and the next multigraph to be considered is Fig. 3.

Now there is an r-cycle (2,1). It is condensed to a vertex c. As all edges emanating from it are e-edges, Case (ii) of Step 3 occurs again, and $u(c)2$ is set to 2. See Fig. 4.

[7] In Schweitzer [17] it is stated that the maximal order of the poles is an upper bound for the number of fundamental equations needed to determine $\pi^{(0)}$.

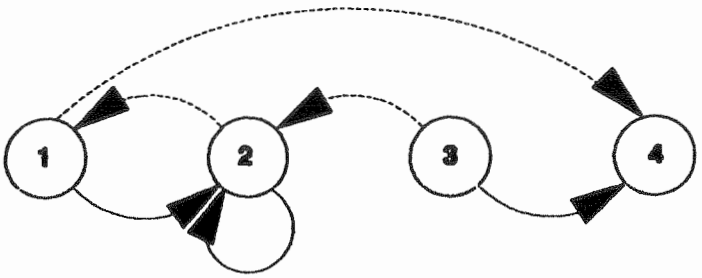

FIG. 2

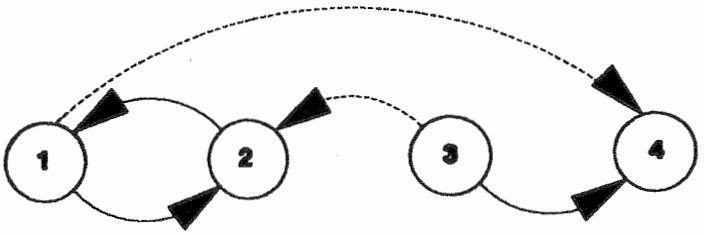

FIG. 3

At this stage, no more r-cycles exist, and the algorithm moves to Step 4. Then the algorithm seeks the set of vertices currently having the largest value and makes this order permanent. Thus $u(c) = 2$ and $T = \{3, 4\}$. Next, the algorithm seeks vertices in T from which vertex c can be reached via paths that are constructed only of r-edges. There are no such vertices in our example. Then it seeks those vertices from which c can be reached in one step via an e-edge. Vertex 3 is such a vertex. It gives it the value of 1, which becomes permanent in the next execution of Step 4. Then $T = \phi$, and the algorithm moves to Step 5. There, it sets $u(1) = u(2) = 2$ and $u(3) = 1$. Finally, in Step 6, Fig. 5, below, is considered.

Then $u(4) = 0$ is computed. Thus $u_{14} = 2$, $u_{24} = 2$, $u_{34} = 1$ and $u_{44} = 0$.

For state 3 as the target state, the algorithm obtains that $u_{13} = 2$, $u_{23} = 2$, $u_{33} = 1$, and $u_{43} = 1$. The other values can be found by symmetry. Hence $\max_{ij}(u_{ij} - u_{jj}) = 2$, and three sets of fundamental equations, $\pi^{(0)}(I - P(0)) = 0$, $\pi^{(1)}(I - P(0)) = \pi^{(0)}A$, and $\pi^{(2)}(I - P(0)) = \pi^{(1)}A$, are required to obtain a system of equations in which the $\pi^{(0)}$ component of a solution $(\pi^{(0)}, \pi^{(1)}, \pi^{(2)})$ is unique (up to a normalizing constant). As is solved in Haviv and Ritov [7], $\pi^{(0)} = (0, 1/2, 0, 1/2)$.

Acknowledgments. The authors thank Y. Ritov and P. J. Schweitzer for their helpful observations.

REFERENCES

[1] M. CODERCH, A. S. WILLSKY, S. S. SASTRY, AND D.A. CASTANON, *Hierarchical aggregation of singularity perturbed finite state Markov processes*, Stochastics, 8 (1983), pp. 259–289.

[2] ———, *Hierarchical aggregation of linear systems with multiple time scales*, IEEE Trans. Automat. Control, AC–28 (1983), pp. 1017–1030.

[3] P. J. COURTOIS, *Decomposability: Queueing and Computing Systems*, Academic Press, New York, 1977.

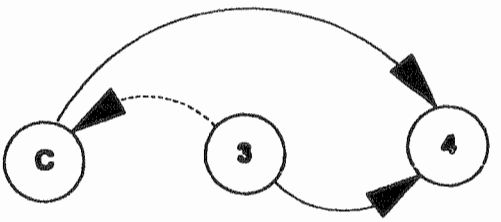

FIG. 4

FIG. 5

[4] F. DELEBECQUE, *A reduction process for perturbed Markov chains*, SIAM J. Appl. Math., 43 (1983), pp. 325–350.
[5] B. L. FOX AND D. M. LANDI, *An algorithm for identifying the ergodic subchains and transient states for stochastic matrices*, Comm. Assoc. Comput. Mach., 11 (1968), pp. 619–621
[6] V. G. GAITSGORY AND A. A. PERVOZVANSKY, *Aggregation of systems in Markov chains with weak interactions*, Kibernetika, 11 (1975), pp. 91–98. (In Russian.) (pp. 441–450 in the English translation.)
[7] M. HAVIV AND J. RITOV, *Series expansion for stochastic matrices*, unpublished.
[8] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, New York, 1966.
[9] C. E. LANGENHOP, *The Laurent expansion for a nearly singular matrix*, Linear Algebra Appl., 4 (1971), pp. 329–340.
[10] G. LATOUCH AND G. LOUCHARD, *Return times in nearly decomposable stochastic processes*, J. Appl. Prob., 15 (1978), pp. 251–267.
[11] G. LOUCHARD AND G. LATOUCH, *Random times in nearly completely decomposable transient Markov chains*, Cahiers du C.E.R.D., 24 (1982), pp. 321–352.
[12] C. D. MEYER, *The role of the group generalized inverse in the theory of finite Markov chains*, SIAM Rev., 17 (1975), pp. 443–464.
[13] J. R. ROHLICEK AND A. S. WILLSKY, *Structural decomposition of multiple time scale Markov processes*, in Proc. of the 25th Allerton Conference on Communications, Control and Computing, pp. 674–683, University of Illinois, Urbanna, IL, pp. 674–683.
[14] ———, *The reduction of Markov generators: An algorithm exposing the role of transient states*, J. Assoc. Comput. Mach., 35 (1988), pp. 675–696.
[15] P. J. SCHWEITZER, *Perturbation theory and finite Markov chains*, J. Appl. Probab., 5 (1968), pp. 401–413.
[16] ———, *Perturbation series expansions of nearly completely decomposable Markov chains*, Working Papers Series No. 8122, The Graduate School of Management, The University of Rochester, Rochester, NY, 1981.
[17] ———, *Perturbation series expansions of nearly completely decomposable Markov chains*, in Teletraffic Analysis and Computer Performance Evaluation, O. J. Boxma, J. W. Cohen and H. C. Tijm, eds., Elsevier, North–Holland, Amsterdam, 1986.
[18] H. A. SIMON AND A. ANDO, *Aggregation of variables in dynamic systems*, Econometrica, 29 (1961), pp. 111–138.