Advances in Mathematics **145**, 159–188 (1999)
Article ID aima.1998.1813, available online at http://www.idealibrary.com on IDEAL®

The State Reduction and Related Algorithms and Their Applications to the Study of Markov Chains, Graph Theory, and the Optimal Stopping Problem

Isaac Sonin

Department of Mathematics, University of North Carolina at Charlotte, Charlotte, North Carolina 28223

Received September 19, 1997; accepted December 14, 1998

We discuss the State Reduction/GTH (Grassmann, Taksar, Heyman) algorithm for recursively finding invariant measure. We demonstrate the relationship between this algorithm and the Freidlin–Wentzell "tree decomposition" approach to study the characteristics of Markov chains. The structure of the State Reduction/GTH algorithm suggests the natural idea for finding the distribution of a Markov chain at the moment of first visit to a given set, and some similar characteristics. We study the possible range of such algorithms. We also present a new algorithm for solving the classical problem of optimal stopping of a Markov chain based on a similar idea of sequential elimination of some states. We give shorter and more transparent proofs of some previously known results, and improve the bounds of Freidlin–Wentzell in the perturbation theory of Markov chains. Some applications to graph theory are also discussed.

1. INTRODUCTION

1.1. *Goals*

The main goals of this paper are threefold. First, we review the well-known State Reduction/GTH algorithm for the sequential calculation of invariant distribution of a Markov chain, proposed in 1985 independently by Sheskin in [1] and by Grassmann, Taksar and Heyman in [2]. We analyze this algorithm and its extensions from a more general and unified point of view, and we describe the approach that naturally stems from this algorithm. We also give shorter and more transparent proofs of some statements related to this algorithm, and present some new results.

Our second goal is to demonstrate a rather surprising relationship between this algorithm and the Wentzell–Freidlin approach to the study of Markov chains based on the "tree decomposition" of a corresponding state space. We also give a substantially shorter and simpler proof and improved bounds for perturbation type estimates.

0001-8708/99 $30.00

Our third goal is to present a new algorithm, based on the State Reduction approach, which we call the Elimination algorithm, that provides a new way to solve the classical problem of stochastic optimization–optimal stopping of a Markov chain.

1.2. *Six Problems*

Let X be a finite or countable state space, $P = \{p(x, y), x, y \in X\}$ be a stochastic (transition) matrix. Let (ξ_n), $n = 0, 1, ...$, be a Markov chain from the family of homogeneous Markov chains specified by X, P and all initial distributions. Let us denote by P_x a probabilistic measure for the Markov chain with initial point x and by E_x the expectation with respect to P_x.

We are studying the following problems related to the behavior of Markov chains (ξ_n).

PROBLEM 1. Given a set $G \subset X$ and an initial state x, calculate the distribution $u^G(x, \cdot)$ of a Markov chain at the moment τ_G of first *visit* (entry) to G ($\tau_G = \min\{n \geqslant 0,\ \xi \in G\}$), $u^G(x, y) = P_x(\xi_{\tau_G} = y)$, or the distribution $u'^G(x, \cdot)$ at the moment θ_G of the first *return* to G (the length of a sojourn from G), $\theta_G = \min\{n \geqslant 1,\ \xi_n \in G\}$. Let us denote corresponding matrices $U^G = \{u^G(x, y)\}$, $x \in X \backslash G$, $y \in G$, $U'^G = \{u'^G(x, y)\}$, $x, y \in G$. (For other values of variables x and y matrices U and U' are either trivial or coincide.)

PROBLEM 2. Given a set $G \subset X$ and an initial state x, calculate the mean time $n^G(x, y)$ spent by a Markov chain at state y until first visit to G, $n^G(x, y) = E_x \sum_{m=0}^{\tau_G - 1} I_y(\xi_m)$, I_B is the characteristic function of the set B (correspondingly calculate the mean time $n'^G(x, y)$ until first return to G). In other words the problem is to calculate matrices $N^G = \{n^G(x, y)\}$, $x, y \in X \backslash G$, and $N'^G = \{n'^G(x, y)\}$, $x \in G$, $y \in X \backslash G$.

In the sequel as a rule we will omit the superscript G in $u^G(x, \cdot)$, $n^G(x, y)$, and other matrices that depend on G when G is clear from the context.

PROBLEM 3. For the ergodic Markov chain, calculate the invariant distribution π and the fundamental matrix Z, i.e. the inverse to $(I - P + A)$, where $A = \lim_n P^n$.

PROBLEM 4 (Freidlin–Wentzell (entry-wise) perturbation type estimates). Obtain estimates for the ratios $u^G(x, s)/\tilde{u}^G(x, s)$, $n^G(x, y)/\tilde{n}^G(x, y)$, and $\pi(x)/\tilde{\pi}(x)$ for two stochastic matrices $P = \{p(x, y)\}$ and $\tilde{P} = \{\tilde{p}(x, y)\}$ such that

$$\lambda^{-1}\tilde{p}(x, y) \leqslant p(x, y) \leqslant \lambda\tilde{p}(x, y) \qquad \text{for all} \quad x, y \in X, \quad \lambda > 1. \tag{1}$$

As it happens Problems 3 and 4 are also related to some problems in graph theory. We will discuss these relationships later but now we can loosely formulate them as

PROBLEM 5. Given a graph (X, E) with a set of vertices X and a set of edges E, count (list) the spanning trees with particular properties using Markov chains specified by X and specially selected stochastic matrices.

PROBLEM 6 (The optimal stopping of a Markov chain). Given the triplet (X, P, g), where g is a *reward function*, find the *value function* $v(x)$ and the *optimal stopping time* τ_*. The value function defined as $v(x) = \sup E_x g(\xi_\tau)$, where the sup is taken over all stopping times τ ($\tau < \infty$) and τ_* is a stopping time such that for any $x \in X$ the equality $E_x g(\xi_{\tau_*}) = v(x)$ holds.

1.3. *The Classical Results* (*Closed-Form Solutions*)

Given a set $G \subset X$ let us introduce the blocks of stochastic matrix P, matrices $Q = \{p(i, j),\ i, j \in X \backslash G\}$, $R = \{p(i, j),\ i \in X \backslash G,\ j \in G\}$, $R' = \{p(i, j),\ i \in G,\ j \in X \backslash G\}$, and $Q' \{p(i, j),\ i, j \in G\}$.

The matrices Q, Q', R, R', U, U', N and N' introduced above in Problems 1 and 2 have the following dimensions, if we assume that $|X \backslash G| = k < \infty$ and $|G| = m \leqslant \infty$. The matrices Q and N are $k \times k$ matrices, R and U are $k \times m$ matrices, R' and N' are $m \times k$ matrices, and Q' and U' are $m \times m$ matrices.

The traditional answers to Problems 1–3 are well known (see, e.g., Kemeny *et al.* [3, 4] or Isaacson and Madsen [5]) and are the following. The matrix N is given by

$$N = \sum_{n=0}^{\infty} Q^n = (I - Q)^{-1}, \tag{2}$$

where I is the $k \times k$ identity matrix. The first equality in (2) follows from the definition of N; the second can be proved in the same way as the equality for geometric series $\sum_{n=0}^{\infty} x^n = (1 - x)^{-1}$. Formula (2) implies obviously

$$N = I + QN = I + NQ. \tag{3}$$

It is easy to show that similarly to (3) for N' we have the equalities

$$N' = R'N = R' + N'Q. \tag{4}$$

(Note that in the former equality we have matrix N, not N'.) Both pairs of equalities, (3) and (4), have simple probabilistic interpretations.

The distribution of a Markov chain at the moment of first visit to G is given by the matrix

$$U = NR. \tag{5}$$

This equality follows from the simple probabilistic equality

$$U = R + QU, \tag{6}$$

which means that from any state x a Markov chain can either jump to G in one step or can make one step inside of $(X \backslash G)$ and, after that, exit to G. (Similar equalities hold for U'.)

The invariant (steady-state) distribution π is the solution of the system of linear equations $\pi^T = \pi^T P$, where ^{T} denotes transposition and all vectors are assumed to be column vectors. In the ergodic case, π is the limit distribution for any initial point; and the matrix $A = \lim_n P^n$ has all rows equal to the vector π^T. The fundamental matrix Z, which is an analog of N, when ergodicity replaces absorption, is given by

$$Z = I + \sum_{n=1}^{\infty} (P - A)^n. \tag{7}$$

The elements $v(x, y)$ of the matrix $V = Z - A$, the so-called *group inverse* ($V \equiv A^{\#}$), have the probabilistic interpretation

$$v(x, y) = \lim_n \left[E_x \sum_{s=0}^{n-1} I_y(\xi_s) - n\pi(y) \right]. \tag{8}$$

Thus they measure the expected deviation in the number of visits to state y due to starting in state x instead of starting randomly according to the invariant distribution π.

1.4. *State Reduction/GTH Algorithm*

There is a vast literature on methods for computing various characteristics of Markov chains. We refer the reader to the volume "Linear Algebra, Markov Chains, and Queueing Models" [6], which represents the modern state of the art in this area.

In two pioneering papers, that appear in 1985, Sheskin [1] and Grassmann, Taksar, and Heyman [2] independently proposed practically the same algorithm to calculate invariant distribution. According to Heyman [7] (1995), "It is a variant of Gaussian elimination that accurately computes the stationary vector of a regular stochastic matrix." But both the algorithm and the broader approach for studying Markov

chains that can be developed from it mean much more. (Indeed a substantial part of that further development has been furnished by Heyman himself.)

The algorithm constructs a sequence of stochastic matrices, each having dimension one less than that of the previous matrix, and has a simple and transparent probabilistic interpretation (given in [1]): namely that each later matrix of the sequence is the transition matrix for a Markov chain whose state space is smaller by one state than that of its predecessor. This new Markov chain is specified by the initial Markov chain at the moments of its visits (entrances) to the reduced state space. The probabilistic interpretation of this algorithm given in [2] was different and was based on two rather complicated statements (Theorem 2 in Section 3 of this paper), though these statements, grounded in regenerative theory, are of significant interest on their own. In subsequent papers Grassmann, Heyman, and other authors mainly used the "visits" interpretation.

The generalization of this algorithm to the case of semi-Markov processes was studied by Kohlas in [8]. Numerous papers studied the computational properties of this algorithm and different generalizations and particular cases (see cited volume [6] and [9–14] and references therein). It was shown, among other things, that the algorithm has serious advantages over the traditional methods of calculation of π.

In most papers this algorithm is referred to as the State Reduction or GTH algorithm. But, when the short but very precise paper of Sheskin [1] is taken into account, there is no doubt that either the algorithm should be referred to as State Reduction or the letter S (after Sheskin) must somehow be added to the abbreviation GTH. Unable to solve the latter problem, in this paper we will call this algorithm the SR (for State Reduction) algorithm.

Note also that the roots of the SR algorithm can be traced to earlier papers of other authors. Lal and Bhat in [12] and their earlier works give very similar and even more general algorithms; but they stopped short of the precise form of the SR algorithm. An important contribution was made in Heyman [7], where, on the basis of SR, another algorithm, FUND was introduced to calculate the fundamental matrix Z.

In our subsequent presentation an important role is played by the transformations of state spaces and transition matrices. So in the sequel, instead of the term "Markov chain," we prefer to use the term "Markov model." A *Markov model* M is a pair (X, P), where X is a finite or countable state space, and P is a stochastic matrix.

We now present, with minor changes in notation, the SR algorithm as described in [1, 2]. In particular we have *reversed the order* in which states are eliminated. This minor change in notation facilitates our presentation of the Elimination algorithm, where the number of states to be eliminated is not known in advance (see Section 4).

Let Markov model $M=(X, P)$, where $X=\{1, 2, ..., k\}$ be given.

SR Algorithm

1. (State reduction). For $n=1, 2, ..., k-1$ do the following:

(a) Let $s_n=\sum_{y=n+1}^{k} p(n, y) \equiv 1-p(n, n)$.

(b) Let

$$p(x, y) \leftarrow p(x, y)+\frac{p(x, n)\, p(n, y)}{s_n}, \qquad x, y>n, \tag{9}$$

The calculations in Step 1 overwrite the elements of P, so let $\hat{P}=\{\hat{p}(x, y)\}$ be the contents of the array when the algorithm terminates.

2. (Back substitution). Initialize $TOT=1$ and $h(k)=1$. For $y=k-1$, $k-2, ..., 2, 1$ do the following:

(a) Let

$$h(y)=\frac{\sum_{x=y+1}^{k} h(x)\, \hat{p}(x, y)}{s_y}. \tag{10}$$

(b) Let $TOT \leftarrow TOT+h(y)$.

3. (Normalization). Let

$$\pi(y)=h(y)/TOT, \qquad y=1, 2, ..., k. \tag{11}$$

Note that at the nth, step of the first stage of this algorithm a new *stochastic* matrix $P_{n+1}=\{p_n(x, y)\}$, $x, y \in X_{n+1}=\{n+1, ..., k\}$, $n=1, 2, ..., k-1$ (of dimension $(k-n)\times(k-n)$) is calculated, $P_k=\{1\}$. It is easy to see that $\hat{p}(x, y)=p_n(x, n)$, $x \geqslant n$, $n=1, 2, ..., k-1$ and that only the elements of matrix $\hat{P}$ on the main diagonal and below (i.e., only the first columns of each of the matrices P_n) are used in the second stage of SR.

The only difference between the above presentation of the algorithm and that found in [2] is that there, before the matrix P_{n+1} is calculated, the first column of the previous matrix P_n ($P_1=P$) to the left of this matrix is divided by s_n, and therefore $h(y)$ in (10) is calculated without s_y.

Denote by h_n the sequence of vectors calculated during the backward stage of the SR algorithm, $h_n=[h(n), h(n+1), ..., h(k)]$, $n=k, k-1, ..., 1$. For the subsequent presentation we need the sequence of vectors $\hat{q}_n$ of the same dimensions, defined by

$$\hat{q}_n=s_n s_{n+1} \cdots s_{k-1} h_n, \qquad n=k, k-1, ..., 1. \tag{12}$$

In contrast to the vectors h_n, which simply include an additional coordinate at each step, vectors $\hat{q}_n$ also change proportionally the "old" coordinates. It is easy to see that the sequence of vectors $\hat{q}_n$, $n = k, ..., 1$, can be obtained by deleting s_y in (10). In Section 2, after the simple transparent probabilistic interpretation of the SR algorithm has been given, we will see that the normalized vector h_1, (as well as vector $\hat{q}_1$) is the invariant distribution π.

1.5. *Wentzell–Freidlin approach and Its relation to the SR Algorithm. Theorem* 1

In 1979 Freidlin and Wentzell in [15] developed an original approach to calculating π (as well as U and N) on the basis of tree decomposition. This well-known book is one of the main sources on the theory of large deviations, but its section on Markov chains plays only an auxiliary role in the book and seems almost unknown to specialists on Markov chains.

In this book it was proved that

$$\pi(x) = \frac{q(x)}{\sum_{y \in X} q(y)}, \tag{13}$$

where $q(y)$ is defined as follows. Let X be a finite set and P be a stochastic matrix. Let T be a *spanning tree directed* to y. This means that T is a connected graph without cycles (tree), contains all the vertices of X (spanning), and that a vertex y is designated as a *root*. In any rooted tree with a root y there is a unique path between any vertex v and y "directed" to y; and this direction makes the tree *a tree directed to y*. (In a more detailed form all necessary definitions of graph theory are given in Section 5.) Denote $G(y) = \{\text{spanning trees on } X \text{ directed to } y\}$. Then

$$q(y) = \sum_{T \in G(y)} r(T), \qquad \text{where} \quad r(T) = \prod_{(u, v) \in T} p(u,). \tag{14}$$

For example, for $X = \{1, 2\}$, $G(1) = \{T_1\}$, $T_1 = \{(2, 1)\}$, and $q(1) = r(T_1) = p(2, 1)$. For $X = \{1, 2, 3\}$, $G(1) = \{T_1, T_2, T_3\}$, $T_1 = \{(3, 2), (2, 1)\}$, $T_2 = \{(2, 3), (3, 1)\}$, $T_3 = \{(3, 1), (2, 1)\}$, and $q(1) = p(3, 2)\, p(2, 1) + p(2, 3)\, p(3, 1) + p(3, 1)\, p(2, 1)$. For $X = \{1\}$ we assume that $q(1) = r(\varnothing) = 1$. It is clear that each spanning tree on a set with k vertices (states) contains exactly $(k-1)$ edges and hence each $r(T)$ is a product of $(k-1)$ transition probabilities.

Observe that from a purely probabilistic point of view the Freidlin–Wentzell formula (13) is not only intuitively not obvious, but even strange: it uses the products of transition probabilities of length $(k-1)$, while the invariant distribution reflects the limit behavior of trajectories over large time intervals.

Note also that the Freidlin–Wentzell formula bears a strong resemblance to (and may have its roots in) Mason's formulas, developed for signal flow analysis in the 1950s and well-known in the theory of electrical circuits (see [25]). The relationship between these two fields definitely deserves to be studied, but we postpone this analysis pending more complete results.

A possible probabilistic interpretation of the quantities $r(T)$ and $q(y)$ is the following. Let each point u in $X\setminus y$ be randomly directed to some other point $v\neq u$ in X with probability $p(u, v)$, with all such directed connections independent. Then $r(T)$ is the probability that the resulting random configuration is the tree T and $q(y)$ is the probability that the resulting random configuration is one of the trees directed to y, i.e., the probability of $G(y)$.

Although formula (13) is inconvenient from a computational point of view (since the number of spanning trees in a complete graph with k vertices is k^{k-2}), it is very useful for answering questions of the type posed in Problem 4. For example, it immediately implies that if P and $\tilde{P}$ are two stochastic matrices satisfying (1), then the ratio of corresponding invariant distributions π and $\tilde{\pi}$ will be bounded by $\lambda^{2(k-1)}$. This statement (point (a) of Theorem 3 in Section 3) was proved again in [16, Theorem 1].

Using (13) and a similar but much more cumbersome representation for $u^G(x, \cdot)$ and $n^G(x, y)$, Freidlin and Wentzell obtained in [15] the important bounds for Problem 4 which they used in the theory of large deviations. We will present their results and our improved bounds in Section 3.

Regarding the second goal of this paper, to describe the relation between the SR algorithm and the results of Freidlin and Wentzell, we will prove the following surprising statement.

THEOREM 1. *Let $M=(X, P)$ be a finite ergodic Markov model, $\hat{q}_1(x)$, $x\in X$, be a vector calculated by formulas* (10) *and* (12), *and $q(x)$, $x\in X$, be a vector calculated by formula* (14). *Then $\hat{q}_1(x)=q(x)$, $x\in X$.*

The equality in Theorem 1 opens the way to using results from Markov chain theory to obtain some results in graph theory. For example, if in the formula (14) we set $p(u, v)=\text{const}>0$ for all edges of a given connected graph with vertex set X, then $q(y)$ gives the number of the spanning trees of this graph directed to y, and the SR algorithm provides sequential formulas for its calculation. The calculation of this and similar numbers is a traditional problem of graph theory. These formulas also can be helpful in other problems related to stochastic graphs or Markov chains defined on graphs, and in problems of stochastic formation of chemical or biochemical configurations. The rather complicated proof of Theorem 1 is given in Section 5.

The remainder of this paper consists of four more sections. Section 2 explains the SR approach, and then demonstrates it with solutions to Problems 1–3. Part of our statements in this section are not new but are intended to simplify and clarify the results previously known. Section 3 presents a Freidlin–Wentzell theorem giving perturbation bounds, and improves on those bounds with a short argument. Section 4 applies the SR approach to the optimal stopping problem. Section 5 is concerned with the connection of Markov chains to spanning trees and the proof of Theorem 1.

2. THE STATE REDUCTION APPROACH

2.1. *The Initial and Reduced Markov Models*

Most works on the SR algorithm refer, for basic statements, to the books [3] and [4], but it is worth noting that the basic idea underlying the elementary step of the SR algorithm appeared in the pioneering works, of Kolmogorov and Döeblin. This idea, described below in Proposition A, has been used since then in probability theory in several contexts on numerous occasions.

Let us assume that a Markov model $M_1 = (X_1, P_1)$ is given and let (ξ_n), $n = 1, 2, ...$, be a Markov chain specified by the model M_1. Let $D \subset X$, and let $\tau_1, \tau_2, ..., \tau_n, ...$ be the sequence of Markov times of first, second, and so on, visits of (ξ_n) to the set $X_2 \equiv X_1 \setminus D$, so that $\tau_1 = \min\{k \geqslant 0 : \xi_k \in X_2\}$, $\tau_{n+1} = \min\{k : \tau_n < k, \xi_k \in X_2\}$, $0 \leqslant \tau_1 < \tau_2 < \cdots$. Let $u_1^{X_2}(x, \cdot)$ be the distribution of Markov chain (ξ_n) for the initial model M_1 at the moment τ_1 of first visit to set X_2 (first exit from D) starting at z, $z \in D$. Let us consider the random sequence $\eta_n = \xi_{\tau_n}$, $n = 1, 2, ...$

PROPOSITION A. (a) *The random sequence* (η_n) *is a Markov chain in a model* $M_2 = (X_2, P_2)$, *where*

(b) *the transition matrix* $P_2 = \{p_2(x, u)\}$ *is given by the formula*

$$p_2(x, y) = p_1(x, y) + \sum_{z \in D} p_1(x, z)\, u_1^{X_2}(z, y), \qquad x, y \in X_2. \tag{15}$$

Proof. Part (a) is immediately implied by the strong Markov property for (ξ_n) while the proof of (b) is straightforward. ∎

Since the distribution of the Markov chain at the moment of first visit to a set can be represented by formula (5), formula (15) can be represented in matrix form. This representation is proved, for example, in [3, pp. 114–116].

To simplify the presentation we will assume that $\tau_{X_2} < \infty$ with probability one. Otherwise, state space X_2 must be complemented by an additional absorbing point z_* with corresponding transition probabilities $p_2(x, z_*)$.

For the sake of brevity, we call model M_2 the *D-reduced model* (of M_1) and denote it as $M_2 = M_1(D)$.

An important case is that in which the set D consists of one point z. In this case formula (15) obviously takes the form

$$p_2(x, y) = p_1(x, y) + \frac{p_1(x, z)\, p_2(z, y)}{(1 - p_1(z, z))} \qquad (x, y \in X_2). \tag{16}$$

According to this formula, each row-vector of the new stochastic matrix P_2 is a linear combination of two rows of P_1 (with the z-column deleted). For a given row of P_2, these two rows are the corresponding row of P_1 and the z-row of P_1.

Comparing (16) with (9), we see that the above formula is the backbone of the first stage of the SR algorithm.

A few other simple statements about the transition from the initial to the reduced model can be useful.

PROPOSITION B. *If a Markov model $M_1 = (X_1, P_1)$ is irreducible aperiodic positive recurrent, then for any $D \subset X_1$ the D-reduced model $M_2 = M_1(D)$ has the same properties.*

PROPOSITION C. *If $M_1 = (X_1, P_1)$ is a Markov model and D_1, D_2 are subsets of X_1, $D_1 \cap D_2 = \varnothing$, then $M_1(D_1 \cup D_2) = (M_1(D_1))(D_2)$; i.e., the reduction in two subsequent steps coincides with the reduction in one combined step.*

PROPOSITION D. *If $M_1 = (X_1, P_1)$ is a (countable) Markov model and $D_1 \subset D_2 \subset \cdots \subset X_1$, $D = \bigcup_n D_n$ and Markov model $M_n = (X_n, P_n)$ is a D_n-reduced model of M_1, $n = 1, 2, \ldots$, $M = (X \backslash D, P)$ is a D-reduced model of M_1, then $p_n(x, y) \nearrow p(x, y)$ when $n \to \infty$ for all $x, y \in X \backslash D$.*

The proofs of B, C, D are standard and we omit them.

PROPOSITION E. *If $M_1 = (X_1, P_1)$ is a time reversible Markov model, then for any $D \subset X$, the D-reduced model $M_2 = M_1(D)$ is also time reversible.*

Proof. For the case $D = \{z\}$ this statement follows immediately from equalities $\pi_1(x)\, p_1(x, y) = \pi_1(y)\, p_1(y, x)$, $x, y \in X_1$ (time reversibility of M_1), formula (16), and proportionality of invariant distributions π_1 and π_2 for all points in smaller state space X_2. The latter property is shown below in Lemma 1. After that, Proposition C implies that Proposition E holds for any finite set D and finally, using Proposition D, we obtain the general statement. ▌

Remark 1. The interesting notion of "partition reversibility," which generalizes that of time reversibility, was introduced in [17]. Unlike time reversibility, partition reversibility is preserved, by state reduction, only for a special class of "appropriate" D, but not in general.

2.2. *The Invariant Distribution in the Initial and Reduced Models*

It is easy to understand that, although the initial Markov model $M_1 = (X_1, P_1)$ and the D-reduced model $M_2 = M_1(D)$ with the transition function $p_2(x, y)$ defined by the formula (15) are two different Markov models, having two different space sets, some of their characteristics will coincide or will have some simple relation.

For the case where the set D consists of one point z, the lemma below shows the relation between the invariant distributions in the initial and reduced models. The formulas that the lemma provides give rise in modified form to formulas (10) and (12).

LEMMA 1. *Let* $M_1 = (X_1, P_1)$ *be a Markov model,* $X_1 = X \cup \{z\}$, *and* $M_2 = (X_2, P_2)$ *be a* z*-reduced Markov model with* $X_2 = X$ *and* $p_2(x, y)$ *defined according to* (16). *Let set* X *and state* z *communicate in the model* M_1; *i.e., there are states* $i, j \in X$ *such that* $p_1(i, z) > 0$, $p_1(z, j) > 0$. *Then*

(a) *if the invariant distribution* $\pi_2(\cdot)$ *exists in model* M_2, *the invariant distribution* $\pi_1(\cdot)$ *also exists* (*in* M_1) *and can be calculated by the formulas* ($\sum_{y \in X} = \sum_y$)

$$\pi_1(y) = \alpha_1 \pi_2(y), \qquad y \in X, \tag{17}$$

$$\pi_1(z) = \beta_1 \sum_y \pi_2(y)\, p_1(y, z) = \beta_1 R_1, \tag{18}$$

where $\alpha_1 = 1 - \pi_1(z) = s_1/(R_1 + s_1)$, $\beta_1 = 1/(R_1 + s_1)$, $R_1 \equiv R_1(z)$ *is the sum in* (18), $s_1 \equiv s_1(z) = 1 - p_1(z, z)$;

(b) *if the invariant distribution* $\pi_1(\cdot)$ *exists in model* M_1, *the invariant distribution* $\pi_2(\cdot)$ *also exists* (*in* M_2) *and is given by the formula*

$$\pi_2(y) = \pi_1(y)\, \alpha_1^{-1}, \qquad y \in X, \quad \alpha_1 = 1 - \pi_1(z).$$

The relations (17) and (18) have a transparent probabilistic meaning and we omit the formal proof of Lemma 1. The invariant distribution is equal to the proportion of time spent at a state. Therefore, the invariant distributions must be proportional on X_2; i.e., the equality (17) holds. This equality is a well-known statement. Formula (18) can be easily derived from (17) and a balance equation for distribution π_1 at point z.

Note also that a similar lemma holds for the case $X_1 = X \cup D$. Then $\alpha_1 = 1 - \sum_{z \in D} \pi_1(z)$ and (18) is replaced by a system of linear equations.

Lemma 1 easily explains why the SR algorithm results in the calculation of the invariant distribution and how formula (10) for h and formula (12) for $\hat{q}_n$ appear.

Consider the sequence of models $M_n = (X_n, P_n)$, $n = 1, 2, ..., k$, where states are eliminated one by one, and the invariant distribution π_k in the last model is known. For example, if M_1 is a finite model, $X_n = \{n, n+1, ..., k\}$, and the states are eliminated in increasing order, then $\pi_k(k) = 1$. Repeated application of Lemma 1 allows us to proceed backward through this sequence, calculating each time the invariant distribution π_n for the model M_n, eventually recovering the invariant distribution π_1 of the initial model.

In the case of a finite model, if the goal is only to calculate π_1, as in the SR algorithm, we can save operations by applying at each step formulas similar to (17) and (18), but *replacing the coefficient* α_1 *by* 1 *and correspondingly* β_1 by $1/s_1$, and normalizing the result only at the final stage. In this case we have the formula for h given by (10) (the second stage of the SR algorithm).

The second way to reduce calculations is to *replace the coefficient* β_1 *by* 1 *and correspondingly* α_1 *by* s_1, and again normalizing the result only at the final stage. More precisely, slightly abusing notation, let us introduce recursively a sequence of vectors $\hat{q}_n = [\hat{q}_n(x), x \in X_n = \{n, n+k\}]$, $n = k, k-1, ..., 1$, as follows. Using generic notation y for points in X_{n+1}, and z for the additional point in X_n (i.e., $z = n$), we write

$$\begin{aligned} &\hat{q}_k(k) = 1, && \hat{q}_n(y) = s_n \hat{q}_{n+1}(y), \\ &\hat{q}_n(z) = \sum_{y \in X_{n+1}} \hat{q}_{n+1}(y)\, p_n(y, z), && n = k-1, ..., 1, \end{aligned} \tag{19}$$

where $s_n = s_n(z) = 1 - p_n(z, z)$. Then it is easy to check that vectors $\hat{q}_n$ defined in (19) coincide with vectors $\hat{q}_n$ defined in (12) and that π_n, h_n, and $\hat{q}_n$ are related by the equalities

$$\pi_n(\cdot) = h_n(\cdot) \prod_{i=n}^{k-1} \alpha_i = \hat{q}_n(\cdot) \prod_{i=n}^{k-1} \beta_i, \qquad n = k, ..., 2, 1, \tag{20}$$

where α_i and β_i are defined as in Lemma 1 for the model M_i instead of M_1.

Formula (20) implies that both normalized $\hat{q}_n$ and h_n coincide with π_n and in particular normalized h_1 gives π_1. This proves that the SR algorithm does in fact calculate the invariant distribution π_1.

Formula (19) (or (10)) makes it possible to calculate the invariant distribution for any finite and some countable Markov chains, but especially simple and fast calculations can be provided for Markov chains defined on trees or graphs with relatively small numbers of cycles or having multihierarchical structure. In many cases the calculations can be implemented in parallel.

Theorem 1 asserts that the vectors $\hat{q}_n$, and not h_n, coincide with the vectors q_n defined in (14) ($q_n = q$ for model M_n). In other words, one way of normalization is linked to "tree decomposition" and the other is not. We do not have a heuristic explanation for this.

2.3. *A Sequence of Models* (M_n). *SR Approach*

The SR algorithm consists of two parts, state reduction and backward iteration. Obviously, only the second is specifically related to the calculation of π, while the first produces a sequence of reduced models of the type described above. A natural idea, partly realized in papers [6–14], is to extend the algorithm to a more general approach, which we shall call the *state reduction* (*SR*) *approach*, and which can be described as follows.

We have an initial Markov model $M_1 = (X_1, P_1)$ and a characteristic $\mathscr{L}_1$ of this model to be calculated. Characteristic $\mathscr{L}_1$ can be a function on X_1, as invariant distribution $\pi(\cdot)$ or a value function $v(\cdot)$, where the former is a solution of a system of linear equations, and the latter is a solution of an optimization problem (see Problem 6). Characteristic $\mathscr{L}_1$ can be a matrix, like matrices U, N, Z, V, defined in Problems 1–3 as mathematical expectations with respect to E_x, $x \in X_1$. More generally, characteristic $\mathscr{L}_1$ can be a property, as is time reversibility. Two basic questions need to be answered.

Q1. When is a characteristic of an initial model *preserved* by reduction?

By preserved, we mean that the characteristics of the initial and reduced models coincide or coincide on the shared parts of the two domains.

Q2. When can a characteristic of an initial model $M_1 = (X_1, P_1)$ be *restored* (calculated backward) given matrix P_1 and the values of this characteristic in the reduced model?

Suppose that the answer for Question 1 or Question 2 is positive for a characteristic $\mathscr{L}$, at least for some pairs of models M and M', where M' is a reduced model of M. The SR approach can be described as the construction of a sequence of models $M_n = (X_n, P_n)$, $n = 1, ..., k$, such that the model M_1 is an initial model; each subsequent model is a reduced model of the previous one; in the last model M_k the characteristic $\mathscr{L}_k$ can be found in explicit form; and characteristic $\mathscr{L}_n$ either coincides with

characteristic $\mathscr{L}_{n+1}$ or it can be calculated through matrix P_n and characteristic $\mathscr{L}_{n+1}$ for all n up to $\mathscr{L}_1$. The SR algorithm for calculating the invariant distribution is an exact example of such a situation.

If, in the initial model, set X_1 is a finite set, $X_1 = \{1, 2, ..., k\}$, then the first part of the SR approach is easy to implement in a standard way. Then $M_n = (X_n, P_n)$, $n = 1, 2, ..., k$, with $X_n = \{n, ..., k\}$, and stochastic matrices P_n are calculated recurrently, using (16), eliminating one point at a time. But generally any suitable sequence of models can be used in the first stage and in some cases even when the initial space set X_1 is countable. In this case, of course, there must be at least one step at which an infinite set D must be eliminated. This is possible if the transition probabilities or a state space X_n has some specific properties, for example, when the Markov chain is a random walk on a line. Such examples are discussed, by Sonin in [19] in the context of optimal stopping problems. Briefly, the SR approach can be described as the first (generalized) stage of the SR algorithm, followed by backward calculation of a characteristic in question.

The most difficult part is the second stage, which requires an affirmative answer either to Question 1 or to Question 2, and some knowledge of how the characteristic of interest is linked in the initial and reduced models. To construct an algorithm, obviously it is sufficient to give the formulas for one step. So here, as in other statements in this paper, to simplify notation we consider this step for models M_1 and M_2, where M_2 is a D-reduced model of M_1.

Lemma 1 in Section 2.2 (and hence the SR algorithm) gives an example of a situation in which the answer to Question 2 is positive. Below we provide sufficient conditions for the answer to Question 1 to be positive.

Heuristically, any property will be preserved if that property depends only on the behavior of trajectories when they are outside of D. To give the formal statements we need some definitions.

Let us denote by $P_{i,x}$, $E_{i,x}$ the measure and expectation for a Markov chain with an initial point x in model M_i, $i = 1, 2$. We again will suppress the x, when it is clear what the initial point is, and we will assume for simplicity that $P_{1,x}(\tau_{X_1 \setminus D} < \infty) = 1$ for all $x \in D$. Denote by $H_i = \{h = (x_0 x_1 \cdots x_s \cdots),\ x_s \in X_i\}$ the set of all (infinite) trajectories in model M_i, $i = 1, 2$, and let $\mathscr{B}_i$, $i = 1, 2$, be corresponding Borel σ-algebras. Let us introduce the projection mapping $F: H_1 \to H_2$ by $F(x_0 x_1 \cdots x_s \cdots) = (x_{\tau_1} \cdots x_{\tau_k} \cdots)$, where $\tau_1, \tau_2, ..., \tau_n, ...$ are the moments of first, second, and so on, visits to the set $X_1 \setminus D$. Thus all $x_i \in D$, $0 \leqslant i$, are deleted from the trajectory $(x_0 x_1 \cdots x_s \cdots)$ in H_1.

Let L_i, be two functionals defined on H_i, $i = 1, 2$. We call such functionals F-equivalent, if they satisfy the condition

$$L_1(h) = L_2(F(h)) \qquad \text{for all} \quad h \in H_1. \tag{21}$$

LEMMA 2. *Let* $M_1=(X_1, P_1)$ *be a Markov model,* $D\subset X_1$, $M_2=(X_2, P_2)$ *be the D-reduced model, and* L_i, $i=1, 2$, *be functionals satisfying condition* (21). *Then*

$$E_{1,x}L_1=E_{2,x}L_2, \qquad x\in X_2. \tag{22}$$

Proof. First, using formula (15) it is easy to check, that $P_2(B)=P_1(F^{-1}(B))$ for any set $B\in\mathscr{B}_2$. This implies that the statement is true for all pairs of functionals L_1, L_2 having a form $L_2(h)=I_B(h)$, $h\in H_2$, $L_1(h)=L_2(F(h))$, $h\in H_1$. After that (22) follows by a limit transition. ▌

Lemma 2 gives a formal justification to the intuitively obvious

LEMMA 3. *Let* $M_1=(X_1, P_1)$ *be a Markov model,* $G\subset X_1$, $D\subset X_1\setminus G$, *and* $M_2=(X_2, P_2)$ *be a D-reduced Markov model, i.e.,* $X_2=X_1\setminus D$, *and* $p_2(x, y)$ *is defined according to* (15). *Let* U_i, N_i, U_i', N_i', $i=1, 2$, *be matrices defined in Problems* 1 *and* 2 *for models* M_1 *and* M_2 (*for a set* G). *Then the elements of* U_1, N_1, U_1', N_1' *restricted to the model* M_2 *coincide with the corresponding elements of* U_2, N_2, U_2', N_2', *i.e.,* $u_1^G(x, y)=u_2^G(x, y)$, $x\in X_2\setminus G$, $y\in G$; $n_1'^G(x, y)=n_2'^G(x, y)$, $x\in G$, $y\in X_2\setminus G$, *and similarly for other matrices.*

Proof. It is sufficient to note that the number of visits to a state $y\in X_2$ and the position of a trajectory at the moment of first visit (or first return) to a set G are examples of functionals satisfying condition (21). ▌

Lemma 3 implies that the distribution $u_1^G(x, \cdot)$ of a Markov chain at the moment of first visit to set G and the mean time $n_1^G(x, y)$ spent at y before this moment in the initial model $M_1=(X_1, P_1)$ both remain unchanged for those states x, y which remain in the reduced model $M_2=(X_2, P_2)$. This property holds true for any finite number of repetitions of the reduction step, provided that x remains in the state space in the case of calculation of $u_1^G(x, \cdot)$, and both x and y in the case of calculation of $n_1^G(x, y)$. If $X_1=\{1, 2, ..., k\}\cup G$ and on the ith step of the reduction stage state i is eliminated, $i=1, 2, ..., k-1$, then at the final, $(k-1)$th step, in model $M_k=(X_k, P_k)$, set $(X_k\setminus G)$ consists of only one point k, and therefore

$$\begin{aligned} &u_1^G(k, \cdot)=u_2^G(k, \cdot)=\cdots=u_k^G(k, \cdot)=p_k(k, \cdot)\, s_k^{-1},\\ &n_1^G(k, k)=\cdots=n_k^G(k, k)=s_k^{-1}, \end{aligned} \tag{23}$$

where $s_k=1-p_k(k, k)$.

Thus we know values $u_i^G(k, \cdot)$ and $n_i^G(k, k)$ for all models M_i, i.e., the characteristics u and n are preserved, but we need computational formulas for sequential calculation for other points. As usual, we consider two models M_1 and M_2, defined in Lemma 1, and we assume that the values

$n_1(x, y) = n_2(x, y)$ are given for all $x \in X_2 = X$. Thus we need to fill the first row $n_1(z, y)$ and the first column $n_1(x, z)$ of matrix N_1. For this goal we will use (3). The first of these equalities gives $n_1(z, y) = p_1(z, z)\, n_1(z, y) + \sum_{x \in X} p_1(z, x)\, n_1(x, y)$, $y \neq z$, which implies the formula to fill the first row in matrix N_1 (except element $n_1(z, z)$),

$$n_1(z, y) = \sum_{x \in X} p_1(z, x)\, n_2(x, y)/s_1(z), \qquad y \neq z. \tag{24}$$

Now, similarly the second equality in (3) provides the first (full) column of matrix N_1,

$$n_1(x, z) = \left[I_z(x) + \sum_{y \in X} n_1(x, y)\, p_1(y, z) \right] \Big/ s_1(z), \qquad x \in X \cup z. \tag{25}$$

Now we can calculate $U_i = \{u_i^G(x, \cdot)\}$ either by (5) or sequentially using (6).

Remark 2. The other method for calculating matrix N is given by Sheskin in [18]. In this paper the two stages, state reduction and backward computation, are combined in one stage when a matrix of double dimension is calculated using formula (9). It is an interesting question whether similar construction is always possible if the answer to Question 2 is positive.

Remark 3. An algorithm for calculating the fundamental matrix Z, similar to FUND given in Heyman [7], also can be described as a sequential calculation of Z, on the basis of Z_{n+1} and P_n, $n = k-1, ..., 2, 1$.

2.4. *A Short Proof of the Grassmann, Taksar, Heyman Statements*

Let $M = (X, P)$ be an irreducible aperiodic positive recurrent Markov model, $X = \{1, 2, ...\}$, $G^s = \{1, 2, ..., s\}$, $n^s(x, y)$ be the expected number of visits of a Markov chain to a state y before the moment of first *return* to G^s (i.e., during a sojourn from G^s) with initial point x, i.e., $n^s(x, y) = E_x \sum_{m=0}^{\theta-1} I_y(\xi_m)$, where θ is the moment of first return to G^s, $\tau = \min\{n \geqslant 1, \xi_n \in G^s\}$. (Note that $n^s(x, y) \equiv n'^{G^s}(x, y)$ in notation of Problems 1 and 2 and Lemma 3.)

THEOREM 2. (a) [2, Theorem 1]. *Let π be the steady-state distribution in the model M. Then for any d and $y \in X$*

$$\pi(y) = \sum_{x=1}^{d} \pi(x)\, n^d(x, y). \tag{26}$$

(b) [2, Proposition 1]. *For any* $x, d, y \in X$, $1 \leqslant x \leqslant d \leqslant y$,

$$n^d(x, y+1) = n^y(x, y+1) + \sum_{u=d+1}^{y} n^d(x, u)\, n^y(u, y+1). \tag{27}$$

Proof. Let M_1 denote the model in Theorem 2. Accordingly all quantities in (26) and (27) will receive the subscript 1. Thus $X_1 = \{1, 2, ...\}$, $\pi = \pi_1$, $n^d = n_1^d$, and so on.

(a) The nontrivial case is the case where $y > d$. Let $D = \{i: i > d, i \neq y\}$ and $M_2 = M_1(D)$ be the D-reduced model, i.e., $X_2 = (\{1, 2, ..., d\} \cup \{y\})$. By definition of invariant distribution ($\pi_2^T = \pi_2^T P_2$) we have $\pi_2(y)(1 - p_2(y, y)) = \sum_{x=1}^{d} \pi_2(x)\, p_2(x, y)$. It is easy to see that in the model M_2, where the only possibility of a sojourn from G^d is a jump to y, cycling a few times in y and returning to G^d, we have $n_2^d(y, y) = 1/(1 - p_2(y, y))$ and $n_2^d(x, y) = p_2(x, y)\, n_2^d(y, y) = p_2(x, y)/(1 - p_2(y, y))$. Therefore $\pi_2(y) = \sum_{x=1}^{d} \pi_2(x)\, n_2^d(x, y)$.

By Lemma 3 applied to set $G = G^d$ ($n_i'^{G^d}(x, y) \equiv n_i^d(x, y)$) we have $n_1^d(x, y) = n_2^d(x, y)$. Now, using this equality and the proportionality of π_1 and π_2 (point (a) of Lemma 1) we receive (26).

(b) Let $D = \{y+2, y+3, ...\}$ and $M_2 = M_1(D)$ be the D-reduced model, i.e., $X_2 = \{1, 2, ..., d, d+1, y, y+1\}$. Applying Lemma 3 correspondingly for $G = G^d$ and $G = G^y$, we obtain $n_1^d(x, u) = n_2^d(x, u)$ for $x \in X_2$, $u \in \{d+1, ..., y, y+1\}$, and $n_1^y(u, y+1) = n_2^y(u, y+1)$ for $u \in X_2$. As is easy to see (we used a similar equality in the proof of point (a)), in model M_2 we have also $n_2^y(u, y+1) = p_2(u, y+1)/c_2$, where $c_2 = 1 - p_2(y+1, y+1)$ for $u \in X_2$, $u \neq y+1$. Hence (27) can be rewritten as

$$n_2^d(x, y+1) = p_2(x, y+1)/c_2 + \sum_{u=d+1}^{y} n_2^d(x, u)\, p_2(u, y+1)/c_2.$$

This formula is obviously equivalent to

$$n_2^d(x, y+1) = p_2(x, y+1) + \sum_{u=d+1}^{y+1} n_2^d(x, u)\, p_2(u, y+1), \tag{28}$$

where the sum has one more term. Now it is sufficient to notice that (28) for $1 \leqslant x \leqslant d \leqslant y$ is simply the second equality in formula (4) for states x and $y+1$ with set $G = \{1, 2, ..., d\}$ and $X \backslash G = \{d+1, ..., y+1\}$ in model M_2. ∎

3. IMPROVED FREIDLIN–WENTZELL PERTURBATION BOUNDS FOR FINITE MARKOV CHAINS

For matrices P and $\tilde{P}$, satisfying inequalities (1), in the sequel we will use notation $P/\tilde{P} = O(\lambda)$ and we will call them *λ-comparable*. We will apply the same convention to two functions (numbers) $f(x)$, $g(x)$ satisfying inequalities similar to (1) for all values of their arguments.

THEOREM 3 [15, Lemmas 3.2–3.4]. *Let $M = (X, P)$ be a Markov model, $G \subset X$, $|X \backslash G| = k < \infty$, and P and $\tilde{P}$ be two stochastic matrices, $P/\tilde{P} = O(\lambda)$.*

(a) *If $\pi(x)$, $\tilde{\pi}(x)$ are the invariant distributions for Markov models $M = (X, P)$ and $\tilde{M} = (X, \tilde{P})$, $(G = \varnothing)$ then*

$$\pi(x)/\tilde{\pi}(x) = O(\lambda^{2(k-1)}), \qquad x \in X, \tag{29}$$

(b) *If $u(x, \cdot)$, $\tilde{u}(x, \cdot)$, $n(x, y)$, $\tilde{n}(x, y)$ are the distributions at the moment of first visit to G and the mean time spent at y until such moment for Markov models M and $\tilde{M}$, then*

$$u(x, \cdot)/\tilde{u}(x, \cdot) = O(\lambda^{4^k}), \qquad x \in X \backslash G, \tag{30}$$

$$n(x, y)/\tilde{n}(x, y) = O(\lambda^{4^k}), \qquad x, y \in X \backslash G. \tag{31}$$

Remark 4. In Theorem 3 we presented the Freidlin–Wentzell results only in the context of Problem 4. They also studied the relation between aggregated and disaggregated models. The second part of their results follows from the above estimates relatively easily.

The basis for Theorem 3 is provided by formula (13). The proof of this formula in [15, Lemma 3.1] is relatively simple and short. For the sake of completeness it is given as a part of the proof of Theorem 1 (Lemma 6 in Section 5). Statement (a) of Theorem 3 follows immediately from this formula, since both numerator and denominator contain products of transition probabilities of length $(k-1)$. In contrast to (13), the proofs of estimates (30) and (31) in [15] require a few pages and are very complicated. Using the idea of State Reduction, we can easily prove

THEOREM 4. *Let the conditions of Theorem* 3 *be satisfied. Then the bounds in* (30) *and* (31) *are true with* 4 *replaced by* 3.

Proof. To prove Theorem 4 we need only two simple statements.

PROPOSITION 1. *Let a, b, c, a', b', c' be positive numbers such that $a/a' = O(s_1)$, $b/b' = O(s_1)$, and $c/c' = O(s_2)$. Then $a'/a = O(s_1)$, $(a+b)/(a'+b) = O(s_1)$, $ac/a'c' = O(s_1 s_2)$.*

Proof. Obvious. ∎

LEMMA 4. *Let $M_1 = (X_1, P_1)$ be a Markov model, P_1 and $\tilde{P}_1$ be two stochastic matrices, $P_1/\tilde{P}_1 = O(s)$. Then after elimination of one point from X_1 the corresponding matrices P_2 and $\tilde{P}_2$ are s^3-comparable, i.e., $P_2/\tilde{P}_2 = O(s^3)$.*

Proof. By the definition of s-comparability we have $p_1(x, y) \leqslant s\tilde{p}_1(x, y)$, $p_1(z, y) \leqslant s\tilde{p}_1(z, y)$, $1 - p(z, z) = \sum_{y \neq z} p_1(z, y) \geqslant s^{-1} \sum_{y \neq z} \tilde{p}_1(z, y) = s^{-1}(1 - \tilde{p}_1(z, z))$, and hence by (16) and Proposition 1 we obtain the statement of Lemma 4. ∎

Applying Lemma 4 sequentially to models M_n, $n = 1, 2, ..., k$ we obtain that $P_2/\tilde{P}_2 = O(\lambda^3)$, $P_3/\tilde{P}_3 = O(\lambda^{3^2})$, ..., $P_{k-1}/\tilde{P}_{k-1} = O(\lambda^{3^{k-2}})$, and $P_k/\tilde{P}_k = O(\lambda^{3^{k-1}})$. Now using formula (23), a similar formula for $n_{k-1}(x, y)$, $x \neq y$, and Lemma 4 we get that $u_1/\tilde{u}_1 = u_k/\tilde{u}_k = O(\lambda^{2 \cdot 3^{k-1}})$, $n_1/\tilde{n}_1 = n_{k-1}/\tilde{n}_{k-1} = O(\lambda^{3^{k-1}})$, which implies Theorem 4. ∎

Remark 5. Using (16) it is not difficult to construct examples that show that the constant 3 in the formulation of Theorem 4 can not be decreased.

4. THE OPTIMAL STOPPING PROBLEM AND THE ELIMINATION ALGORITHM

There are two different, though essentially equivalent, approaches to OSP, usually called "the martingale theory of OSP of general stochastic sequences (processes)" and "the OSP of Markov chains," represented by the two classical monographs [21, 22]. In this paper we use the terminology from the latter approach.

The OSP of a Markov chain is specified by the triplet $M = (X, P, g)$, where X is a state space, which is here assumed countable or finite; $P = \{p(x, y)\}$ is a transition matrix; and g is a *reward function*. Accordingly, let v be a *value function* for this OSP, i.e., $v(x) = \sup E_x g(\xi_\tau)$, where the sup is taken over all stopping times τ ($\tau < \infty$). Let $Tf(x) = \sum_y p(x, y) f(y)$ be the *averaging operator*. It is well known that the value function v satisfies the Bellman (optimality) equation

$$v(x) = \max(g(x), Tv(x)), \tag{32}$$

and that v is the minimal excessive function which majorizes function g, i.e., the minimal function satisfying $v(x) \geqslant g(x)$, $v(x) \geqslant Tv(x)$ for all $x \in X$ (in the terminology of the martingale approach $v(\xi_n)$ is the Snell's envelope).

Let $G \subseteq X$ and τ_G be the moment of first visit to G. We call a set S an *optimal stopping set* if $S = \{x: v(x) = g(x)\}$ and $P_x\{\tau_S < \infty\} = 1$ for all $x \in X$. It is known that if such a set S exists and g is a bounded function then $\tau \equiv \tau_S$ is an optimal stopping time and $v(x) = E_x g(\xi_\tau)$. To simplify the presentation we will assume that the optimal stopping sets do exist though this assumption can be relaxed. This assumption always holds, for example, in the case of finite state space X.

Basically there are three methods of solving OSP. The first one can be conventionally called "the direct solution of the Bellman equation" and generally can be applied only when M has a specific structure. The second method is the value iteration method, in which, instead of solving Eq. (32), one considers the sequence of functions $v_n(x)$ satisfying the relations $v_{n+1}(x) = \max(g(x), Tv_n(x))$, $v_0 = g$, $n = 0, 1, \ldots$. The third well-known method is applicable when the state space X is a finite set. In this case, using the Bellman equation, $v(x)$ can be represented as a solution of a linear programming problem.

These three methods can be complemented by the algorithm which we call the Elimination algorithm. This algorithm was proposed by the author in [19] and uses the first stage of the SR approach. (At that time the author was not aware of a vast literature on the SR algorithm, and no proper references were made.)

The idea of the Elimination algorithm is based on the following simple fact. Though in OSP it may be difficult to find the states where it is optimal to stop, it is relatively easy to find a state (states) where it is optimal *not to stop*. For example, it is optimal not to stop at all states where $Tg(\cdot) > g(\cdot)$; i.e., the expected reward of doing one more step is larger than the reward from stopping. Now we can exclude these states by using state reduction. By Lemma 3 that will keep the distributions at the moments of visits to any subset of remaining states the same and the excluded states do not matter since it is not optimal to stop there. After that, in the reduced model we can repeat the first step, and so on. The formal justification of the transition from model M_1 to model M_2 is given by the following theorem.

THEOREM 5 [19, 20]. *Let $M_1 = (X_1, P_1, g)$ be an optimal stopping problem, $D \subseteq \{z \in X_1 : T_1 g(z) > g(z)\}$ and $P_{1,x}(\tau_{X_1 \setminus D} < \infty) = 1$ for all $x \in D$. Consider an optimal stopping problem $M_2 = (X_2, P_2, g)$ with $X_2 = X_1 \setminus D$, $p_2(x, y)$ defined by* (15). *Let S be the optimal stopping set in problem M_2. Then S is the optimal stopping set in M_1 also and $v_1(x) = v_2(x)$ for all $x \in X_2$.*

Let $M_1 = (X_1, P_1, g)$ be an OSP with finite $X_1 = \{x_1, ..., x_k\}$ and T_1 be the corresponding averaging operator. The implementation of the Elimination algorithm consists of the sequential application of two basic steps.

The first is to calculate the differences $g(x_i) - T_i g(xi)$, $i = 1, 2, ..., k$, until the first state occurs where this difference is negative. If there is no such state, i.e., if all differences are nonnegative, it means that function $g(x)$ is an excessive function (for stochastic matrix P_1) and therefore $g(x)$ is a minimal excessive majorant of $g(x)$, i.e., $g(x) = v(x)$ for all x and X_1 is a stopping set.

Otherwise there is a state, say z, where $g(z) < T_1 g(z)$. This implies (by (32)) that $g(z) < v(z)$ and hence z is not in the stopping set. Then we apply a basic step of the Elimination algorithm: we consider a new, "reduced" model of OSP, $M_2 = (X_2, P_2, g)$ with state set $X_2 = (X_1 \setminus \{z\})$ and transition probabilities $p_2(x, y)$, $x, y \in X_2$, recalculated by (16). Theorem 5 guarantees that the stopping set in the reduced model M_2 coincides with the optimal stopping set in the initial model M_1.

Now we repeat both steps in the model M_2, i.e., check the differences $g(x) - T_2 g(x)$ for $x \in X_2$, where T_2 is an averaging operator for stochastic matrix P_2, and so on. Obviously, in no more than k steps we shall come to the model $M_m = (X_m, P_m, g)$, where $g(x) - T_m g(x) \geqslant 0$ for all $x \in X_m$, and therefore X_m is a stopping set in this and in all previous models, including the initial model M_1.

In [20] examples in which the Elimination algorithm can be applied to a countable set are given. Using a backward stage similar to that of the SR approach, we can also sequentially calculate the values of $v(x)$ for all points eliminated previously.

5. PROOF AND APPLICATION OF THEOREM 1 (TREE COUNTING)

5.1. *Proof of Theorem* 1

Let set X in Theorem 1 contain k points, $X = \{1, 2, ..., k\}$, and M_n, $n = 1, 2, ..., k$, be a standard sequence of models for the initial model $M = M_1 : M_n = (X_n, P_n)$, $n = 1, 2, ..., k$ with $X_n = \{n, ..., k\}$, and stochastic matrices P_n are calculated recurrently, using (16), eliminating one point at a time. Let $\hat{q}_n(x)$ be defined by (10) and (12) and $q_n(x)$ be defined by Freidlin–Wentzell formula (14), both in the models M_n, $n = 1, 2, ..., k$. Thus in our notation $q(x) = q_1(x)$. We obviously will prove Theorem 1 if we show that

$$\hat{q}_n(x) = q_n(x), \qquad x \in X_n, \quad n = 1, 2, ..., k. \tag{33}$$

From Section 2 (see paragraphs after Lemma 1) we know that $\hat{q}_n(x)$ can be also recursively defined by (19). Since by definitions $\hat{q}_k(k) = q_k(k) = 1$, to prove (33) it is sufficient to prove that recursive formulas for $\hat{q}_n(x)$ and $q_n(x)$ coincide for all n, i.e., $q_n(x)$ also satisfy formulas (19). As usual it is sufficient to formulate the corresponding statement (Lemma 5, below) for one step.

Note that Lemma 5 mirrors Lemma 1 in saying that if the normalizing factor $(R_1 + s_1)$ is omitted in coefficients α_1 and β_1 in formulas (17) and (18) then π_1 and π_2 can be replaced by q_1 and q_2. In accordance with notation in Section 1.5 and notation in Lemma 1 denote $G_i(y) = \{\text{spanning trees on } X_i \text{ directed to } y\}$, $i = 1, 2$.

LEMMA 5. *Let Markov models M_1 and M_2 defined as in Lemma* 1, $s_1 = 1 - p_1(z, z)$, *and let* $q_1(x)$, $x \in X \cup \{z\}$, *and* $q_2(y)$, $y \in X$, *calculated by* (14) *in corresponding models, i.e.,*

$$q_1(x) = \sum_{T' \in G_1(x)} \prod_{(u, v) \in T'} p_1(u, v), \qquad q_2(y) = \sum_{T \in G_2(y)} \prod_{(u, v) \in T} p_2(u, v).$$

Then

$$q_1(y) = s_1 q_2(y), \qquad y \in X, \tag{34}$$

$$q_1(z) = \sum_{y \in X} q_2(y)\, p_1(y, z). \tag{35}$$

To prove Lemma 5 we will prove Lemmas 6–10 but first we need some definitions and facts from graph theory (see, e.g., [23]) in addition to those given in Section 1.5.

Graph H is a pair of sets $H = (X, E)$, where X is a set of vertices (states) and E is a set of edges, i.e., a set of unordered pairs (u, v), $u, v \in X$. If E is a set of *ordered* pairs we have a *directed graph* or *digraph*. We consider graphs and digraphs on a given finite set X. If E is not mentioned explicitly we assume that E is the set of all edges, i.e., H is a complete graph.

The number of trees in $G(y) = \{\text{spanning trees on } X \text{ directed to } y\}$ (the same as the number of spanning trees on X) is equal to k^{k-2}, where $k = |X|$. This equality follows from the well-known (in combinatorics) so-called *Enumerator by degree sequence* (see formula (49) at the end of this section).

A collection $F = \{T_u, u \in G \subset X\}$ of *disjoint* trees with vertices in X, such that each T_u is a tree directed to a state u and each state $x \in X$ belongs to some tree T_u, is a *spanning directed forest over set of roots* G (some trees T_u may have no edges, only roots u). In the sequel, for the sake of brevity,

directed forest will always stand for a spanning directed forest over some set of roots. Notice that if there is a spanning tree T_y directed to y and some subset of edges is deleted from T_y (but all vertices remain intact) then the set of remaining edges is always a directed forest over a set $G \cup \{y\}$, where $G = \{u \in X: (u, v) \in T_y, (u, v) \text{ was deleted}\}$.

Given set X, every (directed) graph, tree, forest is uniquely specified by a set of its edges, so we will use the same letter to denote the directed graph, etc., itself and the set of all edges in it. Further we will use letters T, T' to denote trees and directed trees and F to denote directed forests.

Our first step in proving Lemma 5 is Freidlin–Wentzell formula (13). To keep our paper self-contained we give its proof, which is a more detailed version of their short proof.

LEMMA 6 [15, Lemma 3.2]. *Let $M = (X, P)$ be a finite Markov model such that the invariant distribution π exists and $q(\cdot)$ is defined by formula* (14). *Then π is given by formula* (13).

Proof. To prove Lemma 6 it is sufficient to prove that q satisfies the same balance equation as invariant distribution π, i.e.,

$$q(x) \sum_{v \neq x} p(x, v) = \sum_{y \neq x} q(y)\, p(y, x). \tag{36}$$

Given tree $T \in G(x)$ and a state v, consider the set of edges $T \cup (x, v)$. This is a directed graph with one cycle containing both state v and state x. Let $y = y(T, v)$ be the last state in the path from v to x in the tree T. Denote the $S = S(T, v)$ graph resulting from deletion from the set of edges $T \cup (x, v)$ of edge (y, x). It is easy to see that S is a spanning tree directed to y. Then by (14) the left part of (36) is equal to

$$\sum_{T \in G(x)} r(T) \sum_{v \neq x} p(x, v) = \sum_{y \neq x} \left[\sum_{v \neq x} \sum_{T \in G(x):\, y(T, v) = y} r(S(T, v)) \right] p(y, x). \tag{37}$$

Now using the definition of S and the usual properties of trees it can be checked that trees $S(T, v)$ are different for different pairs (T, v) and that $\bigcup_{v \neq x,\, T \in G(x):\, y(T, v) = y} S(T, v) = G(y)$. Therefore the right side of (37) coincides with the right side of (36) and hence the equality (36) is proved. ∎

Our next step, in proving Lemma 5 is to show that formula (34) immediately implies formula (35). Actually; if $q_1(y) = s_1 q_2(y)$ for all $y \in X$, then if we take into account the fact that $s_1 \equiv 1 - p_1(z) = \sum_{x \neq z} p_1(z, x)$, formula (36) for $x = z$ and model $M = M$, becomes (35).

Let us prove formula (34) of Lemma 5. Let a state $y \in X$ be fixed. By definition of $q_1(y)$ (formula (14)), we have

$$q_1(y) = \sum_{T' \in G_1(y)} r_1(T') = \sum_{T' \in G_1(y)} \prod_{(u, v) \in T'} p_1(u, v). \tag{39}$$

By definition of $q_2(y)$

$$s_1 q_2(y) = s_1 \sum_{T \in G_2(y)} r_2(T) = s_1 \sum_{T \in G_2(y)} \prod_{(u, v) \in T} p_2(u, v). \tag{39}$$

Thus our aim is to prove that the right sides of (38) and (39) coincide. The key element of this long proof is the two-stage partitioning of sets $G_1(y)$ and $G_2(y)$, first according to different subsets $D \subset X \setminus y$ and then by different directed forests F over sets $D \cup \{y\}$.

First we apply this idea to decomposing $q_1(y)$ in formula (38). Let $T' \in G_1(y)$. Denote by $D = \{u: (u, z) \in T'\}$, $D = D(T')$. Because y is a root in T', $D \subset X \setminus y$. Let us delete all edges in T' with starting points at D, an edge (z, j), $j \in X$, which is the first edge on the unique path from z to y in T', and a state z. Thus we have removed $|D| + 1$ edges and one state. The set F of remaining edges of T', $F \equiv F_{T', D} = \{(u, v) \in T',\ u \notin D,\ u \neq z\}$, specifies a directed forest $F = \{T_u, u \in D \cup y\}$. Note that state $T_y \equiv T_y(F)$, because if $j \in T_u$, $u \neq y$, we would have a cycle in T' made of the path from j to u and edges (u, z) and (z, j).

Denote $\mathscr{F}(D) = \{F: F = F_{T', D},\ T' \in G_1(y),\ D(T') = D\}$. It is easy to see that $\mathscr{F}(D)$ is the set of all forests F (in X) over set $D \cup \{y\}$.

Obviously, any $T' \in G_1(y)$ uniquely specifies a triplet (D, F, j). Vice versa, if we have a triplet (D, F, j), where D is a subset of $X \setminus y$, F is a directed forest $F = \{T_u, u \in D \cup y\}$ and a point $j \in T_y$, we can obtain a tree $T' \in G_1(y)$ adding to set X a state z, and adding to F edges (u, z) for all $u \in D$ and edge (z, j). This correspondence proves the first point of the following lemma.

LEMMA 7. (a) *There is one-to-one correspondence between trees T' in $G_1(y)$ and triplets (D, F, j) described above.*

(b) *For any tree $T' \in G_1(y)$, $T' = (D, F, j)$*

$$\begin{aligned} r_1(T') &= \prod_{u \in D} p_1(u, z) \prod_{(u, v) \in F} p_1(u, v) \cdot p_1(z, j) \\ &= A(D)\, B(F)\, p_1(z, j), \end{aligned} \tag{40}$$

where $A(D)$, $B(F)$ are the first and second products in (40).

(c) *The probability* $q_1(y)$ *is given by the formula*

$$q_1(y) = \sum_{D \subset X \setminus y} A(D) \sum_{F \in \mathscr{F}(D)} B(F) \sum_{j \in T_y(F)} p_1(z, j). \tag{41}$$

Proof of (b) follows immediately from (a) and the formula for $r_1(T')$, (38). Proof of (c) follows immediately from (38) and (40) and the usual algebraic manipulations. ▮

Now let us analyze formula (39) for $s_1 q_2(y)$. Let $T \in G_2(y)$. The transition probabilities p_2 and p_1 are related by formula (16). Therefore

$$r_2(T) = \prod_{(u, v) \in T} p_2(u, v) = \prod_{(u, v) \in T} (p_1(u, v) + p_1(u, z)\, p_2(z, v)\, s_1^{-1}). \tag{42}$$

Let us consider a monomial in the expansion of this product, i.e., a term with coefficient 1. This monomial is the product of some number of the first terms and of some number of the second terms in the brackets in (42). Let D be the set of all *states* u participating in the second terms and F be the set of all *edges* (u, v) participating as the first terms in the product for this monomial. Because y is a root in T, $D \subset X \setminus y$. Obviously set $D \subset X \setminus y$ uniquely specifies set $F = F_{T, D}$, which also can be described as a result of the deletion from tree T of all edges (u, v), $u \in D$ (but not states). Thus F is a directed forest $F = \{T_u, u \in D \cup y\}$. Given $T \in G_2(y)$ and $D \subset X \setminus y$, denote

$$f(v) \equiv f_{T, D}(v) = |\{u : (u, v) \in T, u \in D\}|, \qquad v \in X; \tag{43}$$

i.e., $f(v)$ is the number of edges in T directed to v from set D, or which is the same, the number of trees T_u directed to v in T. Let $f \equiv f_{T, D}$ be a corresponding $(k-1)$-dimensional vector $f = [f(v), v \in X]$. Obviously $\sum_v f(v) = |D|$. In this notation the monomial can be rewritten as

$$\begin{aligned} &\prod_{(u, v) \in F = F_{T, D}} p_1(u, v) \prod_{u \in D} (p_1(u, z)\, p_1(z, v)\, s_1^{-1}) \\ &\quad = \prod_{u \in D} p_1(u, z) \prod_{(u, v) \in F} p_1(u, v) \prod_v (p_1(z, v))^{f(v)}\, s_1^{-|D|} \\ &\quad = A(D)\, B(F)\, P(T, D)\, s_1^{-|D|}, \end{aligned} \tag{44}$$

where $A(D)$, $B(F)$, and $P(T, D)$ are corresponding products in (44). Notice that $A(D)$ and $B(F)$ are the same as the terms in (40).

Vice versa, given $T \in G_2(y)$, each set $D \subset X \setminus y$ uniquely specifies a monomial in the expansion of (42) and the set $F = F_{T, D}$. This one-to-one correspondence between subsets D and monomials in (42) and formula (44) implies the first point of the following lemma.

LEMMA 8. (a) *For any tree* $T \in G_2(y)$ *its probability*

$$r_2(T) = \sum_{D \subset X \setminus y} A(D)\, B(F)\, P(T, D)\, s_1^{-|D|}, \tag{45}$$

where $A(D)$, $B(F)$, $F = F_{T,D}$, *and* $P(T, D)$ *are defined in* (44).

(b) *The probability* $q_2(y)$ *is given by the formula*

$$q_2(y) = \sum_{D \subset X \setminus y} A(D) \sum_{F \in \mathscr{F}(D)} B(F) \sum_{T \in G_2(y):\, F_{T,D} = F} P(T, D)\, s_1^{-|D|}. \tag{46}$$

To prove point (b) let us notice that given set $D \subset X \setminus y$, a set $\{F: F = F_{T,D}$ for some $T \in G_2(y)\}$ coincides with the set $\mathscr{F}(D)$ of all possible forests $F = \{T_u, u \in D \cup y\}$ (in, X) over set $\{D \cup y\}$, introduced before the formulation of Lemma 7. Therefore by definition of $q_2(y)$ (see formula (39)) and formula (45) we obtain (46). ∎

Now comparing (46) and (41) we see that to prove that $q_1(y) = s_1 q_2(y)$ we need only that the following equality hold for any given set $D \subset X \setminus y$, and any forest $F \in \mathscr{F}(D)$,

$$\sum_{T \in G_2(y):\, F_{T,D} = F} P(T, D) = s_1^{|D|-1} \sum_{j \in T_y(F)} p_1(z, j), \tag{47}$$

where $P(T, D)$ are defined in (44), $s_1 = 1 - p_1(z, z) = \sum_{v \in X} p_1(z, v)$, and function $f = f_{T,D}$ was defined in (43). Notice that the sum in the left side of this equality is taken over trees in the model M_2, but the transition probabilities in both sides of this equality are taken from the model M_1.

To prove (47) we will prove a more general statement for the generating functions defined on set X, which is of interest in its own right. We receive this generalized statement if we replace all transition probabilities $p_1(z, v)$ in both sides of (47) by variables x_v. To simplify notations we will assume $X = \{1, 2, ..., k\}$.

LEMMA 9. *Let* $y \in X = \{1, 2, ..., k\}$, $D \subset X \setminus y$, $F \in \mathscr{F}(D)$ *be a directed forest* $\{T_u\}$ *over set* $D \cup \{y\}$, *with* $f = f_{T,D}$ *defined by* (43). *Then for any tuple* $(x_1, x_2, ..., x_k)$

$$\sum_{T \in G(y):\, F_{T,D} = F} \prod_v x_v^{f(v)} = \left(\sum_{v \in X} x_v \right)^{|D|-1} \sum_{j \in T_y(F)} x_j. \tag{48}$$

To prove Lemma 9 we need some additional notions and results from graph theory. Let G be the set of all *nondirected* spanning trees on a set $Z = \{1, 2, ..., n\}$ and let $T \in G$. With any tree T we can associate a monomial $M(T) = x_1^{d(1)} x_2^{d(2)} \cdots x_n^{d(n)}$, where $d(i)$ is a *degree* of a vertex i,

i.e., the number of edges in T touching i. The theorem, well known in graph theory (see, e.g., [23, Chap. 4, Theorem 2]) states that

$$\sum_{T \in G} M(T) = (x_1 + x_2 + \cdots + x_n)^{n-2} x_1 x_2 \cdots x_n. \tag{49}$$

The sum in the left side is called *the enumerator-by-degree sequence*.

We need similar statements for *directed* trees. Let $y \in Z$ and $T \in G(y)$. Similarly to $M(T)$ let us define the monomial $I(T)$ by $I(T) = x_1^{g(1)} x_2^{g(2)} \cdots x_n^{g(n)}$, where $g(i)$ is an *indegree* of a vertex i, i.e., the number of edges directed to i in a tree T.

LEMMA 10. *Let $y \in Z$, $|X| = n$, and $I(T)$ be the monomials defined above. Then*

$$\sum_{T \in G(y)} I(T) = (x_1 + x_2 + \cdots + x_n)^{n-2} x_y. \tag{50}$$

Proof. Obviously we have $d(y) = g(y)$ and $d(i) = g(i) + 1$ for any vertex $i \neq y$. Hence $M(T) = I(T) \prod_{i \neq y} x_i$ and (50) follows immediately from (49). ∎

Proof of Lemma 9. Let $F = \{T_u\}$ be a directed forest over set $\tilde{D} = D \cup \{y\}$ and a tree $T \in G(y)$ be such that $F_{T,D} = F$. We can define a unique tree $\tilde{T}$ directed to y on a set $\tilde{D}$ as follows. If $\omega, u \in \tilde{D}$ then an edge $(\omega, u) \in \tilde{T}$ iff tree T_ω is connected to tree T_u in T by some edge $(\omega, v_\omega) \in T$, $v_\omega \in T_u$. Denote $\tilde{G}(y)$ the set of all trees on set $\tilde{D}$ directed to y. Let $\tilde{I}(\tilde{T})$ be a monomial for this tree in variables t_u, $u \in \tilde{D}$, i.e., $\tilde{I}(\tilde{T}) = \prod_{u \in \tilde{D}} t_u^{\tilde{g}(u)}$, where $\tilde{g}(u)$ is an indegree of u in $\tilde{T}$, which coincide with the number of T_ω directed to T_u in T. Then by Lemma 10

$$\sum_{\tilde{T} \in \tilde{G}(y)} \tilde{I}(\tilde{T}) = \left(\sum_{u \in \tilde{D}} t_u \right)^{|D|+1-2} t_y. \tag{51}$$

Now we can notice that each tree $T \in G(y)$, such that $F_{T,D} = F$, is specified by a corresponding tree $\tilde{T}$ and by a specification of vectors $V_u = [v_\omega,\ \omega \in D : (\omega, v_\omega) \in T,\ v_\omega \in T_u]$, $u \in \tilde{D}$. It is easy to see that the dimension of vector V_u is $\tilde{g}(u)$ and that $\sum_{v \in T_u} f(v) = \tilde{g}(u)$. Given $\tilde{T}$ and hence $\tilde{g}(u)$, each vector V_u is in one-to-one correspondence with monomials of generating function $(\sum_{v \in T_u} x_v)^{\tilde{g}(u)}$. It means that each t_u in both sides of (51) must be replaced by $(\sum_{v \in T_u} x_v)$. Then (51) is transformed into (48). Thus Lemma 9 and the equality (47) are proved. This proves Lemma 5 and Theorem 1. ∎

If we set $x_1 = x_2 = \cdots = x_k = 1$ then formula (48) implies the following

COROLLARY 1 (of Lemma 9). *Let $y \in X$, $|X| = k$, $D \subset X \setminus y$, $F = \{T_u\}$ be a directed forest over set $D \cup \{y\}$. Then there are $k^{|D|-1} |T_y|$ spanning trees on X directed to y and containing given directed forest F.*

Remark 6. Lemma 9 and formula (48) are in fact a generalization of the theorem given by (49). The latter formula is a particular case of (48) when $D = X \setminus y$, $F = \varnothing$. In this case formula (48) takes the form (50), which is equivalent to (49).

5.2. *A Possible Probabilistic Interpretation of Lemma* 9

Let y, D, $F = \{F_u\} \in \mathscr{F}(D)$ be as in Lemma 9 and $(x_1, x_2, ..., x_k)$ be a *probability distribution* on X, $\sum_{i=1}^{k} x_i \leqslant 1$. Let each tree T_u be randomly connected (directed) by a point u to a point $v \in X$ with probability x_v, and let all such directed connections be independent. For any resulting random configuration C, we can define $f_C(v)$ as the number of such connections to point v. If $C = T$ for some tree $T \in G(y)$, then $f_C = f_{T,D}$ and $\prod_v x_v^{f(v)}$ is the probability that random configuration will result in a tree T. Thus, the left side of (48) is the probability of a set $\{T \in G(y): F_{T,D} = F\}$, i.e., the probability of all random configurations that are trees T, directed to point y and *having as their part given directed forest* $F = \{T_u\} \in \mathscr{F}(D)$.

Formula (47) is a special case of such an interpretation when all $x_v = p_1(z, v)$. Note that by formula (48) this probability does not depend on the structure of trees T_u, $u \neq y$. If $\sum_{i=1}^{k} x_i = 1$, then this probability depends only on set T_y and if $\sum_{i=1}^{k} x_i < 1$ then this probability depends only on the value of this sum, set T_y, and the number of components $|D|$.

5.3. *The Sequential Calculation of the Number of Spanning Trees in a Given Graph G*

Let G be a connected graph on a set $X = \{1, 2, ..., k\}$. Denote by N_G the number of spanning trees on this graph. This number coincides with the number of spanning trees directed to any point x, for example, $x = k$. Let d be the maximal value of degree in G, i.e., $d = \max_{x \in V} d(x)$. Let us define the transition probabilities for the Markov chain on this graph as

$$\begin{aligned} &p(u, v) = 1/d \qquad \text{if} \quad (u, v) \in G, \quad u \neq v, \\ &p(u, u) = 1 - \sum_{v \neq u} p(u, v) = 1 - d(u)/d. \end{aligned} \tag{52}$$

By (14) $r(T) = 1/d^{k-1}$ for every spanning tree on this graph and hence $q(k) \equiv q_1(k) = N_G/d^{k-1}$. By Theorem 1 and Lemma 5, $q_1(k) = \hat{q}_1(k)$, where

$\hat{q}_1(k)$ can be calculated sequentially by the first of formulas (19) ($y=k$) starting from $\hat{q}_k(k)=1$. Thus

$$N_G = d^{k-1} q_1(k) = d^{k-1} \prod_{n=1}^{k-1} s_n(n),$$

where $s_n(n)=1-p_n(n,n)$. This provides an alternative way to calculate the number of spanning trees of a graph G. The traditional way is to calculate the determinant of a matrix specified by the adjacency matrix and the matrix of outdegrees. Note also that, since the uniform distribution is an invariant distribution for any Markov chain with symmetrical transition probabilities, this holds for the Markov chain defined by (52). Therefore by point (b) of Lemma 1 the uniform distribution is also an invariant distribution in any model M_n. Hence all $\hat{q}_n(x)$ for fixed n coincide for different $x \in X_n$.

Similarly we can find the number of spanning trees of a graph G containing a given collection of disjoint trees $\{T_u\}$. To this goal we need to define the transition probabilities on X by two different constants. Other interesting relationships between behavior of time reversible Markov chains on graphs and the algebraic properties of their transition matrices can be found in [24].

ACKNOWLEDGMENT

The author thanks John Thornton, who carefully read a few versions of this paper and made many valuable comments.

REFERENCES

1. T. J. Sheskin, A Markov partitioning algorithm for computing steady-state probabilities, *Oper. Res.* **33** (1985), 228–535.
2. W. K. Grassmann, M. I. Taksar, and D. P. Heyman, Regenerative analysis and steady-state distributions for Markov chains, *Oper. Res.* **33** (1985), 1107–1116.
3. J. G. Kemeny and J. L. Snell, "Finite Markov Chains," Van Nostrand–Reinhold, Princeton, NJ, 1960.
4. J. G. Kemeny, J. L. Snell, and A. V. Knapp, "Denumerable Markov Chains," Springer-Verlag, New York, 1976.
5. D. L. Isaacson and R. W. Madsen, "Markov Chains: Theory and Applications," Wiley, New York, 1976.
6. "Linear Algebra, Markov Chains, and Queueing models" (C. D. Meyer and R. J. Plemmons, Eds.), IMA Vol. Math. Appl., Vol. 48, Springer-Verlag, New York, 1993.
7. D. P. Heyman, Accurate computation of the fundamental matrix of a Markov chain, *SIAM J. Matrix Anal. Appl.* **16** (1995), 954–963.

8. J. Kohlas, Numerical computation of mean passage times and absorption probabilities in Markov and semi-markov models, *Z. Operat. Res.* **30** (1986), A197–A207.
9. S. Kumar, W. K. Grassmann, and R. Billinton, A stable algorithm to calculate steady-state probability and frequency of a Markov system, *IEEE Trans. Reliability* **R36** (1987), 58–62.
10. D. P. Heyman, Further comparisons of direct methods for computing stationary distributions of Markov chains, *SIAM J. Algebraic Discrete Methods* **8**, No. 2 (1987), 226–232.
11. D. P. Heyman and A. Reeves, Numerical solution of linear equations arising in Markov chain models, *ORSA J. Comput.* **1** (1989), 52–60.
12. R. Lal and U. N. Bhat, Reduced systems algorithms for Markov chains, *Management Sci.* **34** (1988), 1202–1220.
13. W. K. Grassmann and D. P. Heyman, Equilibrium distribution of block-structured Markov chains with repeating rows, *J. Appl. Probab.* **27** (1990), 557–576.
14. D. P. Heyman, A direct algorithm for computing the stationary distribution of a p-cyclic Markov chain, *in* "Linear Algebra, Markov Chains, and Queueing Models," pp. 205–209, IMA Vol. Math. Appl., Vol. 48, Springer-Verlag, New York, 1993.
15. M. I. Freidlin and A. D. Wentzell, "Perturbations of Stochastic Dynamic Systems," Springer-Verlag, Berlin/New York, 1984. [transl. of Russian ed., Nauka, Moscow, 1979]
16. C. A. O'Cinneide, Entrywise perturbation theory and error analysis for Markov chains, *Numer. Math.* **65** (1993), 109–120.
17. C. Alexopoulos, A. El-Tannir, and R. Serfozo, Partition-reversible Markov processes, preprint, 1997.
18. T. J. Sheskin, Computing the fundamental matrix for a reducible Markov chain, *Math. Mag.* **68**, No. 5 (1995), 393–398.
19. I. M. Sonin, Two simple theorems in the problems of optimal stopping, *in* "Proceedings, INFORMS Appl. Probab. Conf., Atlanta, GA, 1995."
20. I. M. Sonin, Two simple theorems in the problems of optimal stopping and the Elimination Algorithm, preprint, 1997.
21. Y. S. Chow, H. Robbins, and D. Sigmund, "Great Expectations: The Theory of Optimal Stopping," Houghton Mifflin, New York, 1971.
22. A. N. Shiryayev, "Optimal Stopping Rules," Springer-Verlag, New York, 1978. [transl. of Russian ed., Nauka, Moscow, 1976, 1969]
23. K. Bogart, "Introductory Combinatorics," Harcourt Brace Jovanovich, New York, 1990.
24. P. Diaconis and D. Strook, Geometric bounds for eigenvalues of Markov chains, *Ann. of Appl. Probab.* **1** (1991), 36–61.
25. "The Circuits and Filters. Handbook" (Wai-Kai Chen, Ed.), CRC Press, Boca Raton, FL, 1995.