Reducing Acyclic Phase-Type Representations

Reza Pulungan and Holger Hermanns
Department of Computer Science, Saarland University
D-66123 Saarbrücken, Germany
Email: {pulungan, hermanns}@cs.uni-sb.de

Abstract

An acyclic phase-type distribution is a phase-type distribution with a triangular matrix representation. This class of phase-type distribution exhibits special properties and characteristics, which usually result in some ease in analysis. In this paper, we propose a procedure for reducing the size of the state space of any given acyclic phase-type representations.

I. Introduction

In this paper, we focus ourselves on the class of acyclic phase-type (PH) distributions, namely PH distributions with upper triangular matrix representations.

He and Zhang [5] provided an algorithm for computing minimal representations of acyclic PH distributions. Their algorithm starts by transforming the given acyclic PH distribution to a representation that only contains the poles of the distribution. This representation is not necessarily a PH distribution, but a matrix exponential distribution. If this is the case, another state and its outgoing rate are determined and appended to the representation. This is performed one by one until a PH representation is obtained. The first PH representation found is minimal.

The algorithm of He and Zhang involves solving systems of nonlinear equations when additional states and their rates are to be determined. Since nonlinear programming is difficult, the practicality of this method for large models is not obvious, and has not been investigated so far.

The algorithm developed in this paper, on the other hand, is cubic in the size of the state space, and does only involve standard numerical steps. The goal is to reduce the size of the original representation one by one. However, unlike the algorithm of He and Zhang, the result is not guaranteed to be minimal.

We consider our algorithm worthwhile for it is easy to implement and straightforward to parallelize. It only consists of vector-matrix multiplications and the solution of well-conditioned systems of linear equations. Furthermore, because we are dealing with bidiagonal representations, these operations can be carried out even more efficiently.

II. Phase-Type Distributions

This section provides the preliminaries of PH distributions. The discussion is focused on acyclic PH distributions, their canonical forms and the transformation from arbitrary acyclic PH distribution to the canonical forms.

This work is supported by the German Research Council (DFG) as part of the Transregional Collaborative Research Center "Automatic Verification and Analysis of Complex Systems" (SFB/TR 14 AVACS). See `www.avacs.org` **for more information.**

A. Parametrization Definition

Let $\{X(t) \in \mathcal{I} | t \in \mathbb{R}^+\}$ be a homogeneous Markov process defined on a discrete and finite state space $\mathcal{I} = \{1, 2, \cdots, n, n+1\}$ and with time parameter $t \in \mathbb{R}^+ := [0, \infty)$. The Markov process is a finite continuous-time Markov chain (CTMC). If the state $n+1$ is absorbing and all other states are transient, the infinitesimal generator matrix of the Markov chain can be written as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \underline{A} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{1}$$

Matrix $\mathbf{A}$ is non-singular because the first n states in the Markov chain are transient. Vector $\underline{A}$ is a column vector where its component $\underline{A}_i$ for $i = 1, \cdots, n$ represents the transition rate from state i to the absorbing state. The Markov chain is fully specified by the generator matrix $\mathbf{Q}$ and the initial probability vector $(\underline{\alpha}, \alpha_{n+1})$, where $\underline{\alpha}$ is an n-dimensional row vector corresponding to the initial probabilities of the transient states and α_{n+1} is the initial probability to be immediately in the absorbing state. Therefore $\underline{\alpha e} + \alpha_{n+1} = 1$, where $\underline{e}$ is an n-dimensional column vector whose components are all equal to 1.

Definition 1 (Phase-Type Distribution [6]): A probability distribution on $\mathbb{R}^+$ is a **phase-type** distribution if and only if it is the distribution of the time until absorption in a Markov process of the type described above.

The pair $(\underline{\alpha}, \mathbf{A})$ is called the representation of the PH distribution and $PH(\underline{\alpha}, \mathbf{A})$ is used to denote the PH distribution with representation $(\underline{\alpha}, \mathbf{A})$.

The probability distribution of the time until absorption in the Markov chain (hence of PH distribution) is given by

$$F(t) = 1 - \underline{\alpha} \exp(\mathbf{A}t)\underline{e}, \qquad \text{for } t \geq 0. \tag{2}$$

The Laplace-Stieltjes transform (LST) of the PH distribution is given by

$$\tilde{f}(s) = \int_{-\infty}^{\infty} \exp(-st) dF(t) = \underline{\alpha}(s\mathbf{I} - \mathbf{A})^{-1}\underline{A} + \alpha_{n+1}, \tag{3}$$

where $s \in \mathbb{R}^+$ and $\mathbf{I}$ is the n-dimensional identity matrix.

Consider the LST of the PH distribution in (3). This transform is a rational function, namely

$$\tilde{f}(s) = \underline{\alpha}(s\mathbf{I} - \mathbf{A})^{-1}\underline{A} + \alpha_{n+1} = \frac{P(s)}{Q(s)},$$

for some polynomials $P(s)$ and $Q(s) \neq 0$.

For a PH distribution with an irreducible representation $(\underline{\alpha}, \mathbf{A})$, the **order of the representation** is defined as the dimension of matrix $\mathbf{A}$. The degree of the denominator polynomial of its LST expressed in irreducible ratio is called the **algebraic degree of the distribution**.

It is known [6], [7] that a given PH distribution has more than one irreducible representation. The order of a minimal irreducible representation, namely a representation with the least possible number of states, is referred to as the **order of the PH distribution**. O'Cinneide in [7] showed that the order of the PH distribution may be different from, but at least as great as, its algebraic degree.

B. Acyclic Phase-Type Distributions

An interesting subset of the family of PH distributions is the family of **acyclic PH distributions**. The family can be identified by the fact that they have triangular representations. A **triangular representation** is a representation $(\underline{\alpha}, \mathbf{A})$ where matrix $\mathbf{A}$, under some permutation of its components, is an upper triangular matrix.

In [8], O'Cinneide proved the following theorem, which characterizes acyclic PH distributions in terms of the properties of their density functions and their LSTs.

Theorem 1 (Characterization [8]): A probability distribution defined on $\mathbb{R}^+$, which is not the point mass at zero, is an **acyclic PH distribution** if and only if (1) its density function is strictly positive on $(0, \infty)$, and (2) its LST is rational and has only real poles.

Thus, any general PH representation – possibly containing cycles – represents an acyclic PH distribution (and hence has an acyclic representation) whenever the poles of its LST are all real numbers.

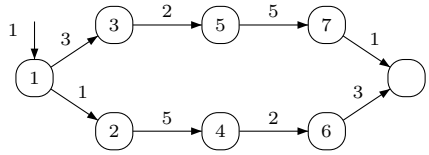

Fig. 1. An Acyclic PH Representations

Example 1 (An Acyclic PH Distribution): The graph in Fig. 1 depicts a triangular representation. Thus, it represents an acyclic PH distribution.

C. Acyclic Canonical Forms

Let an infinitesimal generator matrix of the form

$$\begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 \\ 0 & 0 & -\lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_n \end{bmatrix},$$

be denoted by $\mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. A representation $(\underline{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$ is referred to as a **bidiagonal representation**.

Cumani in [2], presented canonical forms of acyclic PH representations. In particular, he proved that every acyclic PH representation has a bidiagonal representation of the same or less order. A similar theorem was proved by O'Cinneide in [8].

Theorem 2 ([2]): Let $PH(\underline{\alpha}, \mathbf{A})$ be an acyclic PH representation and $\lambda_n \geq \lambda_{n-1} \geq \cdots \geq \lambda_1$ be the diagonal components of $-\mathbf{A}$, so ordered. Then there is an ordered bidiagonal representation such that $PH(\underline{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n)) = PH(\underline{\alpha}, \mathbf{A})$.

Since all of the diagonal components of matrix $-\mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ are of a particular order (namely ascending), the first canonical form is also called the **ordered bidiagonal representation**.

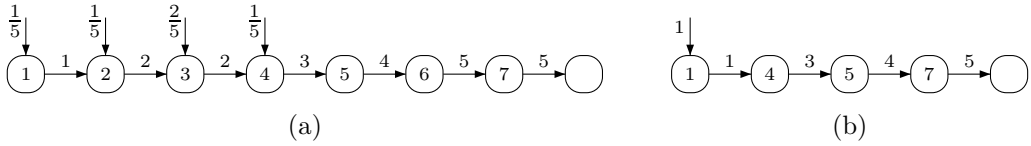

Fig. 2. (a) The Ordered Bidiagonal Representation of Fig. 1 and (b) Their Reduced Representation

Example 2 (An Ordered Bidiagonal Representation): Fig. 2 (a) depicts the graph of the ordered bidiagonal representation (first canonical form) of the triangular representation whose graph is shown in Fig. 1.

D. Transformation Algorithms

He and Zhang in [4] provided an algorithm, called the **spectral polynomial algorithm**, to obtain the bidiagonal representation of a given acyclic PH distribution.

Let $PH(\underline{\alpha}, \mathbf{A})$ be an acyclic PH distribution and let $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ be the eigenvalues of $-\mathbf{A}$. Denote the i-th column of matrix $\mathbf{P}$ by $\mathbf{P}_{*,i}$, then $\mathbf{AP} = \mathbf{PBi}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ can be expressed by

$$\mathbf{AP}_{*,i} = -\lambda_i \mathbf{P}_{*,i} + \lambda_{i-1}\mathbf{P}_{*,i-1},\ i = 1, \cdots, n,$$

with $\mathbf{P}_{*,0}=\underline{0}$. The system of equations can be rewritten as

$$\mathbf{P}_{*,i} = \frac{1}{\lambda_i}(\mathbf{A} + \lambda_{i+1}\mathbf{I})\mathbf{P}_{*,i+1},\ i = 1, \cdots, n-1.$$

For the matrix $\mathbf{P}$ to be of unit row-sums, we have to set $\mathbf{P}_{*,n} = -\mathbf{A}\underline{e}/\lambda_n$ (consult [4] for more details). If there exists a substochastic vector such that $\underline{\beta} = \underline{\alpha}\mathbf{P}$, then $PH(\underline{\alpha}, \mathbf{A}) = PH(\underline{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$. When the bidiagonal generator matrix is in canonical form, namely if $\lambda_n \geq \lambda_{n-1} \geq \cdots \geq \lambda_1$, then such substochastic vector always exists.

The spectral polynomial algorithm has complexity $\mathcal{O}(n^3)$ where n is the order of the given representation.

III. Reducing the Order of Representations

In this section, we propose a procedure to reduce the order of acyclic representations. Roughly the procedure is as follows: (1) Given an acyclic PH distribution, we transform its representation into an ordered bidiagonal representation by using the spectral polynomial algorithm of He and Zhang. (2) In the rest of the section, we will show that a smaller representation can be obtained by removing some unnecessary states from the ordered bidiagonal representation. A method for identifying and removing those unnecessary states will be provided.

A. The L-terms

Let $L(\lambda) = \frac{s+\lambda}{\lambda}$ be the reciprocal of the LST of an exponential distribution with rate $-\lambda$. We refer to a single expression of $L(\cdot)$ as an L-**term**. The LST of an ordered bidiagonal representation $(\underline{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$ can be written as

$$\begin{aligned}\tilde{f}(s) &= \frac{\underline{\beta}_1}{L(\lambda_1)\cdots L(\lambda_n)} + \frac{\underline{\beta}_2}{L(\lambda_2)\cdots L(\lambda_n)} + \cdots + \frac{\underline{\beta}_n}{L(\lambda_n)}, \\ &= \frac{\underline{\beta}_1 + \underline{\beta}_2 L(\lambda_1) + \cdots + \underline{\beta}_n L(\lambda_1)L(\lambda_2)\cdots L(\lambda_{n-1})}{L(\lambda_1)L(\lambda_2)\cdots L(\lambda_n)}. \end{aligned} \tag{4}$$

Note that the LST expression in (4) may not be in irreducible ratio form. The LST is produced in such a way that the denominator polynomial corresponds exactly to the sequence of the transition rates of the ordered bidiagonal representation. Hence, the degree of the denominator polynomial is the same as the order of the ordered bidiagonal representation.

B. Reducing the Order of Bidiagonal Representations

Observing (4), if we are to remove a state from the ordered bidiagonal representation, we will have to find a common L-term in both of the numerator and denominator polynomials of (4). Removing such a common L-term from the numerator and denominator polynomials means removing the corresponding state from the representation.

However, such removal of a state can only be carried out if after the removal, the initial probability distribution $\underline{\beta}$ is redistributed in a correct manner, namely the new initial probability distribution, say $\underline{\gamma}$, is a substochastic vector. This procedure of identifying and

properly removing a state from an ordered bidiagonal representation is described formally in the following lemma.

Lemma 1: Let $(\underline{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$ be an ordered bidiagonal representation of order n. If for some $1 \leq i \leq n$,

$$\underline{\beta}_1 + \underline{\beta}_2 L(\lambda_1) + \cdots + \underline{\beta}_i L(\lambda_1)L(\lambda_2)\cdots L(\lambda_{i-1})$$

is divisible by $L(\lambda_i)$ then $PH(\underline{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n)) = PH(\underline{\gamma}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n))$ if $\underline{\gamma}$ is a substochastic vector. $(\underline{\gamma}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n))$ is an ordered bidiagonal representation of order $n-1$.

Proof: Assume that the conditions are true. Then we can find a common L-term in both the numerator and denominator polynomials of the LST of $(\underline{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$ because the rest of the numerator polynomial is divisible by $L(\lambda_i)$. Removing this common L-term reduces the order by 1 and we have a new representation $(\underline{\gamma}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n))$.

Since the new representation is constructed from the same LST (albeit simplified), it constitutes the same PH distribution. ■

The first observation we can gain from inspecting the form of (4) and Lemma 1 is that $L(\lambda_1)$ cannot divide the numerator polynomial if $\underline{\beta}_1 \neq 0$. Hence, Erlang distributions and the convex combination of Erlang distributions with the same parameter are always irreducible.

To reduce the order of an ordered bidiagonal representation, we need to check two conditions in Lemma 1, namely the divisibility of the numerator polynomial and the substochasticity of the resulting initial probability vector. For this, let

$$R(s) = \underline{\beta}_1 + \underline{\beta}_2 L(\lambda_1) + \cdots + \underline{\beta}_i L(\lambda_1)L(\lambda_2)\cdots L(\lambda_{i-1}). \tag{5}$$

By ordinary algebra, we can check whether $R(s)$ is divisible by $L(\lambda_i)$ simply by checking whether $R(-\lambda_i) = 0$.

The substochasticity of the resulting initial probability vector, on the other hand, can be checked while computing it. Referring back to Lemma 1, let us rename the generator matrices $\mathbf{Bi}_1(\cdot) := \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ and $\mathbf{Bi}_2(\cdot) := \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n)$. The two ordered bidiagonal representations represent the same PH distribution, hence

$$\begin{aligned} PH(\underline{\beta}, \mathbf{Bi}_1(\cdot)) &= PH(\underline{\gamma}, \mathbf{Bi}_2(\cdot)), \\ 1 - \underline{\beta}\exp(\mathbf{Bi}_1(\cdot)t)\underline{e_n} &= 1 - \underline{\gamma}\exp(\mathbf{Bi}_2(\cdot)t)\underline{e_{n-1}}, \end{aligned} \tag{6}$$

where $\underline{e_n}$ is a column vector of dimension n whose components are all equal to 1. To compute vector $\underline{\gamma}$ from vector $\underline{\beta}$, we need $n-1$ equations relating their components. We can evaluate (6) at $n-1$ different t values to obtain such required system of equations. However, such function evaluations can be costly. To avoid this, we proceed differently.

For a PH representation $(\underline{\alpha}, \mathbf{A})$, the i-th derivative, for $i > 0$, of its distribution functions is given by

$$F^{(i)}(t) = -\underline{\alpha}\mathbf{A}^i \exp(\mathbf{A}t)\underline{e}.$$

Evaluating these derivatives at $t = 0$ allows us to avoid computing the exponential of matrices. Hence, the components of vector $\underline{\gamma}$ can be computed by solving the following system of equations

$$\underline{\beta}\mathbf{Bi}_1(\cdot)^i\underline{e_n} = \underline{\gamma}\mathbf{Bi}_2(\cdot)^i\underline{e_{n-1}}, \quad i = 0, \cdots, n-2. \tag{7}$$

Once (7) is solved, the substochasticity of $\underline{\gamma}$ can be determined simply by checking that all of its components are nonnegative real numbers.

Example 3 (A Reduced Acyclic PH Representation): Fig. 2 (b) depicts the graph of the reduced acyclic PH representation obtained by applying the reduction method to representation whose graph is shown in Fig. 1.

C. The Algorithm

Let such a state which corresponds to the L-term $L(\lambda_i)$ in Lemma 1 be called a **removable** state. Thus, a state is removable in an ordered bidiagonal representation, if the numerator polynomial of the LST is divisible by the L-term corresponding to the state and if removing the state results in a valid initial probability distribution.

Lemma 1 can be turned into an algorithm that reduces the order of a given acyclic PH representation.

Algorithm 1 (Reducing Acyclic Representation):

```
 1: function RedAcycRep(α, A)
 2:     (β, Bi) ← Spa(α, A)
 3:     n ← OrderOf(β, Bi)
 4:     i ← 2
 5:     while i ≤ n do
 6:         if Removable(i, (β, Bi)) then
 7:             (β', Bi') ← Remove(i, (β, Bi))
 8:             n ← n − 1
 9:         else
10:             i ← i + 1
11:         end if
12:         (β, Bi) ← (β', Bi')
13:     end while
14:     return (β, Bi)
15: end function
```

Algorithm 1 inputs an acyclic PH representation and outputs its reduced ordered bidiagonal representation. Function Spa$(\cdot)$, in the algorithm transforms an acyclic representation to its ordered bidiagonal representation using the spectral polynomial algorithm of He and Zhang ([4]). Function OrderOf$(\cdot)$ returns the order of the given representation. Function Removable$(\cdot)$ returns True if state i is removable from the given representation. Removing it means redistributing the initial probability distribution, whose computation was shown in (7). This is basically what function Remove$(\cdot)$ does, namely removing state i from the given representation.

The algorithm proceeds by checking each state whether it is removable or not. If it is removable then it is eliminated. Further, the algorithm terminates once all states have been checked and the removable ones have been eliminated. Hence, the algorithm does what it is supposed to do and terminates.

Let n be the order of the given acyclic PH representation. The spectral polynomial algorithm of He and Zhang entails n matrix-vector multiplications, thus amounts to $\mathcal{O}(n^3)$. Checking whether state i is removable needs an evaluation of (5), which requires $\left(i^2 - i\right)$ multiplications and $\left(\frac{i^2+i}{2} - 1\right)$ additions. At worst this costs $\mathcal{O}(n^2)$.

If state i is removable, eliminating it entails solving (7). We observe that for any bidiagonal generator $\mathbf{Bi}$ of dimension d, $\mathbf{Bi}(i,i) = -\mathbf{Bi}(i,i+1)$ for $0 < i < d$. Now, since both matrices in the system of equations are bidiagonal generators, this observation allows us to show – by induction on n – that (7) can be transformed into

$$\underline{b} = \mathbf{A}\underline{\gamma}^T \tag{8}$$

where $\mathbf{A}$ is an upper triangular matrix and $\underline{b}$ is a column vector which is obtained by evaluating the left hand side of (7) $n-1$ times. A detailed analysis shows that this transformation requires $(\frac{3n^2+5n}{2})$ multiplications and (n^2) additions, and is thus of order $\mathcal{O}(n^2)$

time. Equation (8), on the other hand, can be solved in $\mathcal{O}(n^2)$ time, because $\mathbf{A}$ is upper triangular, which means we only need to apply backward substitutions.

Since the procedure for checking whether a state is removable and then removing it is actually carried out at most n times, hence the overall time complexity of the algorithm is $\mathcal{O}(n^3 + n(n^2 + n^2 + n^2)) = \mathcal{O}(n^3)$.

IV. Conclusions and Future Work

This paper has introduced an algorithm to reduce acyclic representations of PH distributions (viz. acyclic absorbing CTMCs). In each iteration, the algorithm requires quadratic time (in the current number of states), to reduce the representation by at least one state, as long as no further reduction is possible. We have highlighted that the resulting CTMC is not necessarily minimal

In practice, one may wonder whether it is beneficial to run an overall cubic algorithm to reduce the matrix representation of a PH distribution. This, of course, depends on the application context. If one intends to numerically compute the absorption probability at many time points (or at a single large time point) it might very well be worthwhile to run the suggested algorithm as a preprocessing step. This step reduces the dimensions of the involved matrices and vectors, and hence speeds up the subsequent (usually uniformization-based) iterations.

Further, if the CTMC representation is used in a concurrency context, then a one-state saving in a single component – prior to exploring the crossproduct with other components – saves states in the order of the entirety of all other components. This can in general lead to an exponential saving in size of the overall system.

We have made a prototypical implementation of the algorithm using exact arithmetic. Preliminary results from experimenting with acyclic PH representations obtained from analyzing fault-tolerant parallel processor models [3] with the compositional method of [1] are encouraging. By varying the number of processors in each group, we generated several models of increasing state space size. Our reduction method can reduces the size of the state space of each model by orders of magnitude, for instance from 1728 states to 64 states. This reduction generally takes much less than the time required to obtain the model.

In the future we intend to improve the reduction method and its implementation. We also plan to gather more stochastic models with acyclic PH representations and carry out more systematic experiments for them.

References

[1] Hichem Boudali, Pepijn Crouzen, and Mariëlle Stoelinga. A compositional semantics for Dynamic Fault Trees in terms of Interactive Markov Chains. In *5th International Symposium on Automated Technology for Verification and Analysis, ATVA'07*, **2007 (accepted).**

[2] Aldo Cumani. Canonical representation of homogeneous Markov processes modelling failure time distributions. *Microelectronics and Reliability*, **2(3):583–602, 1982.**

[3] J.B. Dugan, S.J. Bavuso, and M.A. Boyd. Dynamic fault-tree models for fault-tolerant computer systems. *IEEE Transaction on Reliability*, **41(3):363–377, 1992.**

[4] Qi-Ming He and Hanqin Zhang. Spectral polynomial algorithms for computing bi-diagonal representations for phase type distributions and matrix-exponential distributions. *Stochastic Models*, **2(2):289–317, 2006.**

[5] Qi-Ming He and Hanqin Zhang. An algorithm for computing the minimal coxian representation. *INFORMS Journal of Computing*, **2007 (accepted).**

[6] Marcel F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. **Dover, 1981.**

[7] Colm A. O'Cinneide. Characterization of phase-type distributions. *Communications in Statistics - Stochastic Models*, **6(1):1–57, 1990.**

[8] Colm A. O'Cinneide. Phase-type distributions and invariant polytopes. *Advances in Applied Probability*, **23(43):515–535, 1991.**