International Journal of Control

ISSN: 0020-7179 (Print) 1366-5820 (Online) Journal homepage: http://www.tandfonline.com/loi/tcon20

Phase-type distributions and representations: Some results and open problems for system theory

Christian Commault & Stéphane Mocanu

To cite this article: Christian Commault & Stéphane Mocanu (2003) Phase-type distributions and representations: Some results and open problems for system theory, International Journal of Control, 76:6, 566-580, DOI: 10.1080/0020717031000114986

To link to this article: http://dx.doi.org/10.1080/0020717031000114986

Published online: 08 Nov 2010.

INT. J. CONTROL, 2003, VOL. 76, NO. 6, 566–580

Phase-type distributions and representations: some results and open problems for system theory

CHRISTIAN COMMAULT†* and STÉPHANE MOCANU†

In this paper we consider phase-type distributions. These distributions correspond to the random hitting time of an absorbing Markov chain. They are used for modelling various random times, in particular, those which appear in manufacturing systems as processing times, times to failure, repair times, etc. The Markovian nature of these distributions allows the use of very efficient matrix based computer methods for performance evaluation. In this paper we give a system theory oriented introduction to phase-type distributions. We concentrate mainly on the representation problem which consists of finding a Markov chain associated with some phase-type distribution. This is a realization problem in the sense of system theory with a lot of links with the classical linear system theory but also with a number of constraints which make the problem harder but more interesting. Indeed this problem has strong connections with the positive realization problem in control theory. The paper recalls known results, gives some new results, and points out the main remaining problems.

1. Introduction

In this paper we deal with phase-type distributions and their representations. A phase-type distribution (PH-distribution) is the distribution of the time to absorption in a finite state absorbing Markov chain. The underlying Markov chain is called a representation of the distribution. Phase-type distributions appeared as a generalization of the exponential distribution. They allow a better modelling of positive random variables, and in particular random times. Moreover they preserve the Markovian nature of the model which is crucial for tractable computations for performance evaluation. The interest in using representations with several stages has been recognized for a long time (see Erlang 1917, Jensen 1949), and a lot of special structures have been proposed until the general definition which was given in Neuts (1975). In queuing theory specific algorithms have been developed for queues in which interarrival times and/or service times are PH distributed (Neuts 1981, Latouche and Ramaswami 1993).

The aim of this paper is to motivate the interest of control and system theory communities towards PH-distributions and representations. This interest should result, on one hand, from the fact that their domain of potential applications in modelling, performance evaluation, control of production and communication systems is now considered as intersecting our field of expertise. On the other hand, the relationship between the Laplace transform of the probability density function of a PH-distribution and a corresponding representation is very similar to the relationship between a transfer function and a corresponding state space representation. The PH-representations are then a special form of state space representations but with particular constraints on the Markov generators (Berman and Plemmons 1994). This induces a realization problem which is far from being a trivial extension of the classical linear realization problem. In particular the problem of finding the minimal order of a representation for a given PH distribution is still open. The PH-representation problem is indeed strongly connected with the positive realization problem which received a great deal of attention in the last decades in control theory, see Farina and Rinaldi (2000) for an up-to-date survey on this field. The power of modelling with canonical PH-representations and the equivalence between representations also needs a lot of further investigation. Even for the low order representations only specific cases have been completely studied. It would be of interest, for example, to fully understand the order three representations since they probably have the same importance for PH-distributions as second order systems do for classical linear systems. In summary, this is a problem with applications we are interested in, which uses tools which are familiar to us, and with a lot of interesting open questions. Why abandon it exclusively to probabilists?

This paper is not a survey of PH-representation theory. It intends to introduce the main definitions and concepts which are useful to understand this theory. It recalls a number of known results concerning mainly the representation problem in two directions: first, general properties of representations which may help in finding a representation for a given PH-distribution, and second, the particular cases which are completely solved. We point out the open problems and state a set of problems and conjectures concerning partial results. We particularly insist on the structural aspects

Received 6 December 2002.
* Author for correspondence. e-mail: commault@lag.ensieg.inpg.fr
† Laboratoire d'Automatique de Grenoble UMR 5528, INPG-CNRS-UJF, ENSIEG-BP 46, 38402 Saint Martin d'Hères Cedex, France. e-mail: mocanu@lag.ensieg.inpg.fr

International Journal of Control ISSN 0020–7179 print/ISSN 1366–5820 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0020717031000114986

of the problem and on the properties of the graph of the representation. The notion of cyclicity is particularly emphasized. We also present a result which states that the PH-representation problem and the positive realization problem are essentially the same. This opens the way for a fruitful cross-fertilization between these two fields in the future. This paper is complementary with O'Cinneide (1999) which is written with a more probabilistic point of view and also contains a number of conjectures on PH-distributions.

The paper is organized as follows. In § 2 we recall the definitions of PH-distributions and representations, we illustrate them by a simple reliability example and give two fundamental characterization theorems for PH-distributions. Section 3 contains new results which show that the PH-representation problem and the positive realization problem in control theory are essentially the same problem. In § 4 we recall some general properties of representations which concern the eigenvalue locations, the probabilistic properties of the associated distribution, and some structural aspects. Section 5 reviews some canonical forms which were introduced as progressive extensions of the exponential law. In § 6 we give some representation theorems and focus on the cases where there exists a representation whose order is the degree of the denominator of the Laplace transform of the distribution. In § 7 we deal with the important role which is played by cyclicity in the representation problem. Some concluding remarks and perspectives end the paper.

2. Definitions and characterizations

In this section we present the main definitions concerning phase-type distributions and phase-type representations. We illustrate this material with a simple example. We then give two characterization theorems for phase-type distributions.

2.1. *Definitions and an example*

The first systematic approach to the study of PH-type distributions was presented by Neuts (1981). The definition presented here follows Neuts definition.

Definition 1: A phase-type distribution is the distribution of the absorption time in a finite state absorbing Markov chain.

In the following we will only consider continuous time Markov chains, but all the definitions and results have their natural discrete time counterparts. Consider now a Markov chain with $(n+1)$ states, the first n states being transient and the last one absorbing. This Markov chain is governed by a state equation as

$$\left.\begin{aligned} \dot{\bar{x}}(t) &= \bar{x}(t)Q \\ \bar{x}(0) &= \overline{\alpha} \end{aligned}\right\} \qquad (1)$$

where the ith component of the row vector $\bar{x}(t)$ is the probability of being in state i at time t. The vector $\overline{\alpha}$ represents the initial probability distribution and is a stochastic vector. The $(n+1)\times(n+1)$ matrix Q represents the infinitesimal generator of the Markov process and can be partitioned as

$$Q = \begin{pmatrix} T & v \\ 0 & 0 \end{pmatrix} \qquad (2)$$

$T_{ij} \geq 0$ for $i \neq j, i = 1, \ldots, n; j = 1, \ldots, n$, is the transition rate from state i to state j. v_i for $i = 1, \ldots, n$, represents the transition rate from state i to the absorbing state. Due to the special structure of a Markov generator we must have $v = -T\vec{1}$ where $\vec{1}$ is the n dimensional column vector whose entries are all equal to one. The assumption that the first n states are transient implies that the matrix T is non-singular. This matrix is called a PH-generator. From now on we will assume that $\overline{\alpha} = (\alpha, 0)$ where α is n dimensional. This means that we assume that the probability of being in the absorbing state at time 0 is 0. This is for the sake of simplicity but it implies no real loss of generality. The pair (α, T) is called a representation of the corresponding distribution which is denoted by $PH(\alpha, T)$. We will also denote $\bar{x}(t) = (x(t), \bar{x}_{n+1}(t))$. It is easy to see that the distribution function of the random time to absorption, is in fact the function $\bar{x}_{n+1}(t)$. The probability density function of the absorption time, $f(t)$, is then its derivative $f(t) = \dot{\bar{x}}_{n+1}(t)$ and can be read from the state equation

$$\left.\begin{aligned} \dot{x}(t) &= x(t)T \\ f(t) &= \dot{\bar{x}}_{n+1}(t) = x(t)v \end{aligned}\right\} \qquad (3)$$

Up to a transposition, these equations are very similar to our classical linear state space equations. The solution is then $f(t) = \alpha\, e^{tT} v$. We will also frequently make use of the Laplace transform of $f(t)$ that we will denote by $\tilde{f}(s)$. $\tilde{f}(s) = \alpha(sI - T)^{-1}v$, where I denotes the order n identity matrix. In the sequel, we will for short say, Laplace transform of a distribution to mean the Laplace transform of the probability density function of the distribution. Again this function has the familiar aspect of the transfer function of a single input–single output (SISO) linear system. The *order* of a representation is the number of transient states of the Markov chain which is the dimension of the matrix T. Let $\tilde{f}(s)$ be the Laplace transform of a PH-distribution and $\tilde{f}(s) = p(s)/q(s)$, where $p(s)$ and $q(s)$ are coprime polynomials. We call the *degree* of the distribution the polynomial degree of $q(s)$. Obviously, for any PH representation the order is greater than or equal to the degree of the corresponding distribution. But, contrary to what happens for standard linear systems, given a PH-distribution there generally does not exist a representation of order equal to the degree of this distribution. Let us recall the constraints of the problem at hand compared to the usual linear system model used in control.

- The non-singular matrix T represents an infinitesimal generator, then

$$\left.\begin{aligned} &T_{ij} \geq 0 \quad \text{for } i \neq j, \quad i = 1, \dots, n, \quad \text{and} \quad j = 1, \dots, n \\ &\sum_{j=1}^{n} T_{ij} \leq 0 \quad \text{for } i = 1, \dots, n \end{aligned}\right\} \tag{4}$$

- The v vector is defined as

$$v = -T\vec{1} \tag{5}$$

- The initial probability distribution is a stochastic vector, that is : $\alpha_i \geq 0$ for $i = 1, \dots, n$ and

$$\sum_{i=1}^{n} \alpha_i = 1 \tag{6}$$

It is known that a lot of qualitative properties of Markov chains can be obtained from the study of their graphs. The graph representation will thus be a very important tool for our study of phase-type representations. To a phase-type representation (α, T) we associate a directed graph $G(\alpha, T)$, where the set of vertices is composed of $(n+2)$ elements. The vertices $1, \dots, n+1$ correspond to the states of the representation. The vertex 0 is a dummy vertex corresponding to the initial state. There exists an edge (i, j), $i \neq j$, if $T_{ij} \neq 0$ for $i = 1, \dots, n$; $j = 1, \dots, n$. There exists an edge $(i, n+1)$, if $v_i \neq 0$ for $i = 1, \dots, n$. There exists an edge $(0, j)$, if $\alpha_j \neq 0$ for $j = 1, \dots, n$. To each vertex i for $i = 1, \dots, n$, we associate its outgoing rate $(-T_{ii})$. To each edge $(0, i)$ for $i = 1, \dots, n$, we associate the probability α_i. To each edge (i, j) for $i \neq j$, and for $i = 1, \dots, n$, $j = 1, \dots, n$, we associate the routing probability $p_{ij} = -T_{ij}/T_{ii}$. To each edge $(j, n+1)$, $j = 1, \dots, n$, we associate the absorption probability $p_{j,n+1} = -v_j/T_{jj}$.

A representation in which graph all the state vertices are connected to the initial vertex and to the absorbing vertex is called *irreducible* (Neuts 1981). It is obvious that, when it is not the case, the representation may be simplified and the corresponding states discarded without altering the distribution.

Example 1: Let us now present an example to illustrate the previous definitions. A highly reliable system is composed of three identical machines. Only one machine works at a time. When a machine fails another one begins to work, if there is a machine in working order. Each machine has a Time to Failure which is exponentially distributed with rate λ. When the repairer is working on it, a machine has a Repair Time which is exponentially distributed with rate μ. There is a single repairer, moreover he is rather lazy and begins repairs only if there are at least two machines down. When he is working, the repairer works until all the down machines are repaired. The system is in total catastrophic breakdown when all machines are down.

We are interested in the time between the initial situation when all the machines are in working order, and the total breakdown.

To model this problem recall that a situation which has a duration which is exponentially distributed corresponds to a markovian state. We are then able to list the states of the system:

- State x_1: the three machines are in working order, one is operating.
- State x_2: two machines are in working order, with one operating.
- State x_3: one machine is in working order, with one operating. The repairer is repairing one of the down machines.
- State x_4: two machines are in working order, with one operating. The repairer is repairing the down machine.
- State x_5: the three machines are down. This is our absorbing state.

The time we are interested in is therefore phase-type with representation (α, T) where

$$T = \begin{pmatrix} -\lambda & \lambda & 0 & 0 \\ 0 & -\lambda & \lambda & 0 \\ 0 & 0 & -(\lambda+\mu) & \mu \\ \mu & 0 & \lambda & -(\lambda+\mu) \end{pmatrix} \tag{7}$$

$$\alpha = (1, 0, 0, 0)$$

If we assume that $\lambda = 1$ and $\mu = 10$, the probability density function is

$$f(t) = 0.0028\,\mathrm{e}^{-14.251\,t} + 0.0133\,\mathrm{e}^{-7.481\,t} - 0.0638\,\mathrm{e}^{-2.221\,t} + 0.0477\,\mathrm{e}^{-0.0465\,t}$$

and its Laplace transform is

$$\tilde{f}(s) = \frac{s+11}{s^4 + 24\,s^3 + 156\,s^2 + 244\,s + 11}$$

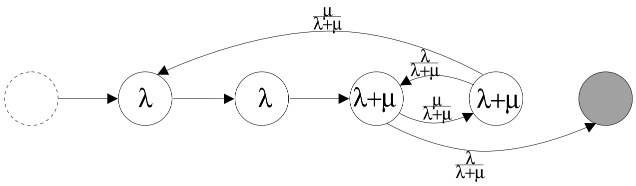

Figure 1. Graph of the Markov chain in Example 1.

Figure 1 shows the graph associated with this Markov chain. The 'dotted' node in the graph corresponds to the initial state of the chain. To the final state of the chain is associated the 'shaded' node in the graph. If (i, j) is the unique outgoing edge from the state i, then obviously $p_{ij} = 1$ and the weight is omitted in the graph.

2.2. *Characterization theorems*

We recall in this subsection some characterization theorems for phase-type distributions and comment on their consequences. The fundamental characterization theorem was first given by O'Cinneide (1990). Although the result is simple in its expression, the proof is long and technical and mainly based on convex analysis. An alternate proof for the discrete time case was given by Maier (1991). It is interesting to note that the technique of proof was completely different and used ideas of automata theory. The fundamental theorem is the following.

Theorem 1 (O'Cinneide 1990): *A probability distribution on* $[0, \infty)$ *which is not the point mass at zero is of phase-type if and only if:*

1. *it has a rational Laplace transform with a unique pole of maximal real part, and*
2. *its probability density function is positive on* $[0, \infty)$.

Let us leave aside the point mass at 0 which is directly translated into an initial probability of being in the absorbing state, and concentrate on the continuous part of the probability density function. The probability density function is a non-negative function of t and therefore its Laplace transform cannot have a dominating oscillating pole. Thus, the theorem says that all the probability distribution functions having a rational Laplace transform are *PH* except for some boundary cases. It has been shown that if the p.d.f. approaches the time axis for a finite t_0 we are lead to consider representations of increasing order. The same observation can be made concerning condition 1 of the theorem, which will be made more precise in a following section. The characterization given by the theorem is simple but one may regret that it is expressed partly in the t-domain and partly in the s-domain. Then we can set the following problem, even if it seems rather difficult since its solution would imply that we could characterize transfer functions with positive impulse response.

Problem 1: Find a characterization of PH-distributions only in terms of their Laplace transform.

The phase-type distributions have a number of natural applications in reliability theory (see our example). It was proved in (Neuts 1981) that the usual operations encountered in reliability applications preserve the phase-type nature of the distributions. Conversely the phase-type distributions can be characterized by a closure property (Maier and O'Cinneide 1992). Let μ be a distribution on R_+, p be a probability, and consider the new distribution defined as: $\mu^{(p)} = (1-p)[\mu + p\mu \circ \mu + p^2 \mu \circ \mu \circ \mu + \cdots]$, where $\circ$ is the convolution operation. This is in fact nothing more than feedback. The closure theorem can then be formulated as follows.

Theorem 2 (Maier and O'Cinneide 1992): *The family of phase-type distribution is the smallest family of distributions which:*

1. *contains the Dirac impulse in* 0 *and the exponential distributions,*
2. *is closed under the operations of convolution and convex combination,*
3. *is closed under the operation* $\mu \to \mu^{(p)}$, *where* $0 \le p < 1$.

This suggests that one could see the set of finite dimensional linear time invariant systems as the smallest set which contains constants and integrators and which is closed under series, parallel, and feedback connections.

3. Positive realizations and PH-representations

In this section we will introduce the positive realization problem and recall some results on this control theoretic problem, see Farina and Rinaldi (2000) for a fairly complete introduction to this field. We will then prove that the PH-representation problem is in fact a particular positive realization problem.

3.1. *The positive realization problem*

A positive state space representation for a SISO system is said to be positive if it has the form

$$\left.\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned}\right\} \qquad (8)$$

where A, B and C are real matrices of respective dimensions $n \times n$, $n \times 1$, $1 \times n$ with

$$a_{ij} \geq 0 \quad \text{for } i \neq j; \quad i = 1, \ldots, \quad n; \quad j = 1, \ldots, n \tag{9}$$

$$b_i \geq 0 \quad \text{for } i = 1, \ldots, n \tag{10}$$

$$c_i \geq 0 \quad \text{for } i = 1, \ldots, n \tag{11}$$

This state space representation is such that for any initial state $x(0) \geq 0$ and any input function $u(t) \geq 0$ for $t \geq 0$, the state and output functions remain non-negative. Positive representations appear naturally in modelling of a number of physical, economical and ecological systems (Luenberger 1979).

A linear system defined by its rational transfer function $\tilde{g}(s)$ is said to be (externally) positive if its impulse response $g(t)$, the inverse Laplace transform of $\tilde{g}(s)$, is such that $g(t) \geq 0$ for $t \geq 0$. We assume that the transfer function is strictly proper, that is $g(0) = 0$.

The positive realization problem is the following: given a positive system defined by its transfer function $\tilde{g}(s)$, find, if possible, a positive state space representation (A, B, C) such that $\tilde{g}(s) = C(sI - A)^{-1}B$. The problem of existence of such a positive realization was solved in Farina (1996).

Theorem 3: *A linear system with transfer function $\tilde{g}(s)$ and impulse response $g(t)$ has a positive realization if and only if:*

1. *$\tilde{g}(s)$ has a unique pole with maximal real part, and*
2. *$g(t) > 0$ for $t > 0$.*

Since it is clear that a PH-representation is a particular positive realization with specific and seemingly hard constraints, it is surprising that the conditions of Theorems 1 and 3 are exactly the same. This raises an interesting question: is not the PH-representation simply a canonical form which can always be achieved, provided of course that the impulse response has the same properties as a p.d.f.? This question will be studied and answered affirmatively in the next subsection.

3.2. *From a positive realization to a PH-representation*

To a positive state space representation one can associate a graph as we did for PH-representations. The vertices are associated with input, state and output variables, therefore we have one vertex u, one vertex y and n vertices $(x_1, \ldots, x_n)$. There is an edge (x_i, x_j) (resp. (u, x_i), (x_i, y)) if $a_{ji} > 0$ (resp. $b_i > 0$, $c_i > 0$). This graph is called the influence graph in Farina and Rinaldi (2000). We say that the representation is irreducible if any state vertex belongs to an input–output path in the associated graph. Irreducibility is equivalent to both excitability and transparency in the terminology of Farina and Rinaldi (2000). It is clear that when a representation is not irreducible it can be simplified by discarding some states, therefore we may restrict our attention, without loss of generality, to irreducible representations. The following result which is specific to positive realizations is of importance.

Proposition 1: *Consider a linear system with transfer function $\tilde{g}(s)$ which admits a positive realization. For any irreducible positive realization (A, B, C) the eigenvalue of A with maximal real part is the pole of maximal real part of $\tilde{g}(s)$.*

Proof: The result is proved in Anderson *et al.* (1996) for discrete-time systems but the same arguments can be used for continuous-time systems. Also the notion of irreducibility is not explicit in the formulation of Anderson *et al.* (1996) □

Although irreducibility is weaker than the assumption of joint controllability and observability, the proposition says that for positive irreducible representations, the dominant eigenvalue of A can never be simplified. The same result was proved for PH-representations in Neuts (1984).

From the basic properties of a p.d.f. it follows that the impulse response $g(t)$ of a linear positive system may be interpreted as a p.d.f. if the corresponding transfer function $\tilde{g}(s)$ is asymptotically stable and if the normalization condition $\tilde{g}(0) = 1$ is satisfied. This allows us to give the correspondence between positive realizations and PH-representations.

Theorem 4: *Consider a linear system with transfer function $\tilde{g}(s)$ and impulse response $g(t)$ which admits a positive irreducible realization of order n, (F, G, H):*

1. *if $\tilde{g}(s)$ is asymptotically stable there exists an order n positive irreducible realization of $\tilde{g}(s)$, (A, B, C), which satisfies*

$$\sum_{j=1}^{n} a_{ij} + b_i = 0 \quad \text{for } i = 1, \ldots, n \tag{12}$$

   and the graph of (A, B, C) is the same as the graph of (F, G, H),

2. *if moreover $\tilde{g}(0) = 1$ then*

$$\sum_{i=1}^{n} c_i = 1 \tag{13}$$

Before proving the theorem let us state a technical lemma.

Lemma 1 *Consider an irreducible positive representation (F, G, H), where F is asymptotically stable, the equation*

$$F\gamma = -G \tag{14}$$

has a unique solution $\gamma = (\gamma_1, \ldots, \gamma_n)^{\mathrm{T}}$ *with* $\gamma_i > 0$ *for* $i = 1, \ldots, n$.

Proof: The solution of the equation exists and is unique since F is asymptotically stable and then non-singular. From Berman *et al.* (1989, Theorem 4.12), this solution is non-negative. Let us now prove that it has indeed only positive entries. For i such that $g_i \neq 0$ the ith equation implies that $\gamma_i > 0$. Let $j \neq i$ be such that $f_{ji} > 0$, then we must have $\gamma_j > 0$. From the irreducibility assumption, each state is connected to the input and we can repeat the reasoning for all the entries of γ. □

Proof of Theorem 4:

1. Define the $n \times n$ diagonal non-singular matrix $M = \mathrm{Diag}(\gamma_1, \ldots, \gamma_n)$ where γ is the solution of $F\gamma = -G$. Define the new triplet (A, B, C) where

$$A = M^{-1}FM, \quad B = M^{-1}G, \quad C = HM \tag{15}$$

(A, B, C) is a new realization of $\tilde{g}(s)$ and it is clearly positive. The change of basis corresponds here to a simple rescaling of the state variables. It can easily be checked that the relation (12) follows from the definition of γ.

The change of basis amounts to a multiplication of rows and columns (resp. rows, columns) of F (resp. G, H) by positive numbers. Then an entry of a matrix which is positive remains positive after the change of basis, an entry of a matrix which is 0 remains 0 after the change of basis. Therefore the graph of (A, B, C) is the same as the graph of (F, G, H).

2. Since C is a non-zero matrix with non-negative entries we can write $C = \delta\hat{C}$ where

$$\delta = \sum_{i=1}^{n} c_i \tag{16}$$

the entries of $\hat{C}$ are non-negative and sum up to one. Therefore the triplet $(A, B, \hat{C})$ corresponds to a PH-representation and the corresponding distribution has a Laplace transform which satisfies $\hat{g}(0) = \hat{C}A^{-1}B = 1$. If we start with $\tilde{g}(0) = 1$ then $\tilde{g}(0) = HF^{-1}G = CA^{-1}B = \delta\hat{C}A^{-1}B = \delta = 1$ which implies

$$\sum_{i=1}^{n} c_i = 1 \tag{17}$$

□

Remark 1: Theorem 4 is a generalization of a result in Maeda *et al.* (1981) in which it is proved that any asymptotically stable positive system which has a positive realization may be seen as a compartmental system.

Remark 2: A very important consequence of Theorem 4 is that most of the results in the realization theory of positive systems may be used for PH-representations and vice versa. As an example Theorem 1 follows immediately from Theorems 3 and 4.

With this last remark in mind we will continue the paper with the point of view of PH-representations.

4. Canonical forms for PH-representations

The search for and the study of particular representations have been subject of important work in the theory of phase-type representations. The particularities of a representation can be of a structural nature, that is its graph has some special features, or of a parametric nature. The interest of these studies is twofold. First, in fitting a given data set, it is much easier to identify the parameters of a representation which is constrained than those of a general representation which is overparameterized. Second, given some canonical form it is important to characterize the set of representations which are equivalent in the sense that they lead to the same PH-distribution. We will not develop in this paper the identification aspects although this is a very promising subject (Johnson and Taaffe 1990 a, b, 1991, Ryden 1996).

If we again discard the Dirac impulse whose representation is only an absorbing state, the simplest representation consists of one transient state with rate λ. This representation leads to the exponential distribution which has probability density function $f(t) = \lambda\,\mathrm{e}^{-\lambda t}$, and Laplace transform

$$\tilde{f}(s) = \frac{\lambda}{s + \lambda} \tag{18}$$

If we consider the convolution of k exponentials with the same rate λ, this is obtained by the series connection of k states, we then have

$$f(t) = \frac{\lambda^k t^{k-1}}{(k-1)!}\,\mathrm{e}^{-\lambda t} \tag{19}$$

and

$$\tilde{f}(s) = \left(\frac{\lambda}{s + \lambda}\right)^k \tag{20}$$

This representation is called the Erlang representation. It has been known for a long time and has demonstrated its utility within modelling tele-traffic problems (Erlang 1917). The simple graph associated with an Erlang distribution is composed of k 'chained' identical states with each transition rate λ (figure 2).

We call a generalized Erlang representation one that has the series structure of the Erlang representation but

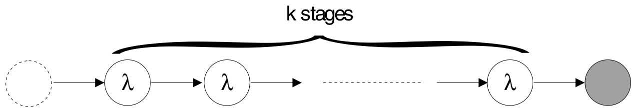

Figure 2. Graph of the Markov chain of an Erlang representation

where the rates of the states are not necessarily equal. The Laplace transform of the distribution is then

$$\tilde{f}(s) = \left(\frac{\lambda_1}{s+\lambda_1}\right) \cdot \left(\frac{\lambda_2}{s+\lambda_2}\right) \ldots \left(\frac{\lambda_k}{s+\lambda_k}\right) \quad (21)$$

The associated graph is very similar to the one in figure 2 but the corresponding transition rates are different according to (21).

A generalization of the previous form, by allowing a direct routing to the absorbing state after leaving a given state, results in the Cox representation (Cox 1955) which leads to a distribution whose Laplace transform is

$$\tilde{f}(s) = \frac{\lambda_1}{s+\lambda_1}\left(p_1 + (1-p_1)\frac{\lambda_2}{s+\lambda_2}\left(p_2 + (1-p_2)\frac{\lambda_3}{s+\lambda_3}\ldots \right.\right.$$
$$\left.\left. +(1-p_{k-1})\frac{\lambda_k}{s+\lambda_k}\right)\ldots\right) \quad (22)$$

The corresponding graph topology can be easily derived from (22), and is shown in figure 3.

We call a Cox representation such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$, an ordered Cox representation.

The hyperexponential distribution is obtained as a convex combination of several exponentials, which in terms of graphs is simply a parallel connection (see figure 4). If the initial probability distribution on the k states is $\alpha = (\alpha_1, \ldots, \alpha_k)$ the Laplace transform of the distribution is

$$\tilde{f}(s) = \alpha_1\left(\frac{\lambda_1}{s+\lambda_1}\right) + \alpha_2\left(\frac{\lambda_2}{s+\lambda_2}\right) + \cdots + \alpha_k\left(\frac{\lambda_k}{s+\lambda_k}\right) \quad (23)$$

The Mixture of Generalized Erlang (MGE) distribution is defined as the convex combination of a finite set of generalized Erlang distributions.

All the previously introduced representations have the common feature that their graph is acyclic. In general we call a representation whose graph is acyclic, triangular. The term triangular comes from the fact that, in an acyclic graph, there is a natural order among the vertices and renumbering the states according to

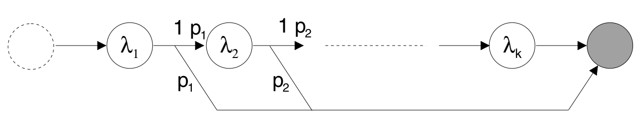

Figure 3. Graph of the Markov chain of a Cox representation.

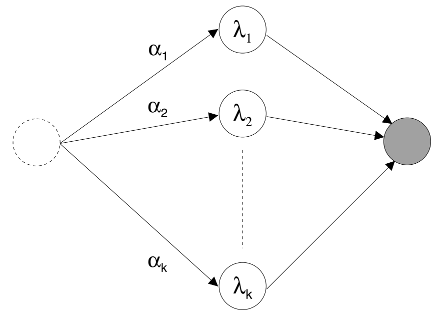

Figure 4. Graph of the Markov chain of a hyperexponential distribution.

this order induces a PH generator T which is upper triangular.

We will now state an important result on equivalence between representations. We say that two representations are *equivalent* if they give rise to the same PH-distribution.

Theorem 5 (Cumani 1982): *A triangular PH-representation is equivalent to an ordered Cox representation of at most the same order.*

Proof (sketch): We will not give a complete and formal proof of this result but instead present the main ideas and illustrate them on a simple example. This will be representative of some simple techniques which can be used to simplify representations.

1. First, we note that we can view a PH-representation as the movement of an entity in the graph. At time 0 the entity routes towards state i with probability α_i. In state i it spends an exponentially distributed time with rate $-T_{ii}$, the outgoing rate of state i. Then it moves to state j with probability $p_{ij} = -(T_{ij}/T_{ii})$ and so on, until it reaches the absorbing state. One can then get another representation which is the finite convex combination of all the possible paths from the initial state to the absorbing state, in fact a MGE. Let us perform this operation on the example in figure 5 whose representation is (24). We get the representation of figure 6.

$$T = \begin{pmatrix} -3 & 1 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & -2 \end{pmatrix} \quad \alpha = (0.5, 0.2, 0.3) \quad (24)$$

2. In each path we reorder the states in decreasing order of magnitude of their outgoing rate.
3. At this point we use the following simple trick. It is easy to check that an exponential stage of rate μ is equivalent to the two stage Cox representation

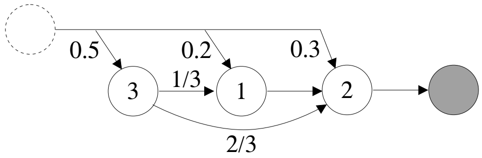

Figure 5. The acyclic graph corresponding to the upper-triangular phase-type representation (24).

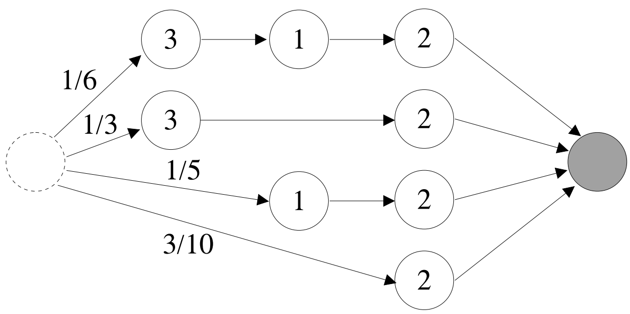

Figure 6. The convex combination of all the paths in the graph of figure 5.

of figure 7 for any λ such that $\mu < \lambda$, and $p = \mu/\lambda$. We can then put ahead in each path a stage with rate equal to the maximal rate. This stage can be 'factorized' on all paths. We have then a stage with maximal rate, followed by a convex combination of paths in which this maximal rate has an order reduced by one. Continue the procedure until a Cox representation is obtained. □

From the procedure it is clear that we will end up with a Cox representation in which the outgoing rates are the outgoing rates of the initial representation appearing in decreasing order of magnitude. Note that this procedure may result in a strict decrease of the number of states (think of a hyperexponential with two states of the same rate which reduces to a unique exponential). The ordered Cox representation of the distribution considered in (24) and figure 5 is given in figure 8.

Since, up to a reordering of states, the T matrix of an acyclic representation can always be considered as triangular, the poles of the Laplace transform of the corresponding distribution are real. If we want to take into account the possibility of non-real poles we must introduce representations with a cyclic graph. The simplest cyclic representation is obtained by adding a feedback on an Erlang representation. The result is called a *feedback Erlang representation*. For example, adding a feedback of weight z on the Erlang representation of figure 2 leads to the representation in figure 9.

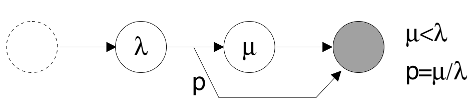

Figure 7. A second order Cox representation that reduces to the exponential of rate μ.

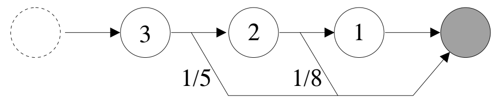

Figure 8. The ordered Cox representation of the distribution in (24)

The Laplace transform of a feedback Erlang distribution is obtained by a formula similar to the closed loop transfer function expression. If one thinks of equation (20) as the open loop transfer function of a system, adding a feedback of constant value z leads to the expression of the Laplace transform of the feedback Erlang distribution

$$\tilde{f}(s) = \frac{(1-z)\lambda^k}{(s+\lambda)^k - z\lambda^k} \tag{25}$$

One can easily provide analytical expressions for the poles of the Laplace transform. They can be obtained by elementary manipulations of the kth order roots of unity

$$\lambda_j = \lambda\left(1 - z^{(1/k)}\, \mathrm{e}^{(2j\pi/k)i}\right) \qquad j = 0 \ldots k-1 \quad \forall k > 3$$

One can construct canonical cyclic representations as extensions of the existing acyclic representations. The idea is to 'replace' one or more states in an acyclic representation by a feedback Erlang block. The representations obtained in such a way are called *monocyclic* representations (Mocanu and Commault 1999). If in (21) we replace some of the exponential stages by distributions with Laplace transform of the form (25) we obtain a series structure analogous to the generalized Erlang representation but obtained as convolutions of exponentials and feedback Erlang blocks (figure 10). We call such a representation a *generalized monocyclic Erlang representation*. As exponentials and Erlang distributions can be seen as feedback Erlangs of adequate order and zero feedback weight we can express the Laplace transform of a monocyclic Erlang representation as

$$\tilde{f}(s) = \prod_{i=1}^{k} \frac{(1-z_i)\lambda^{n_i}}{(s+\lambda)^{n_i} - z_i\lambda^{n_i}} \tag{26}$$

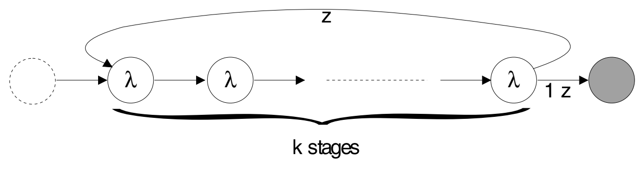

Figure 9. Feedback Erlang representation.

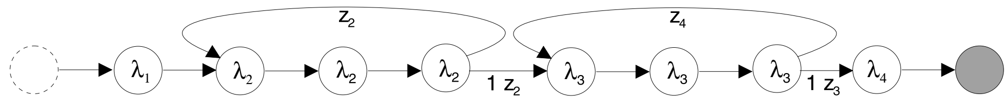

Figure 10. Monocyclic Erlang representation.

where n_i, $i = 1, \dots, k$ are the orders of the Erlang blocks in the convolution.

There are two main characteristics of monocyclic representations.

- a state vertex of the graph may belong to at most one cycle of the graph;
- inside a cycle the transition rates of all the states are equal.

Representations such as the generalized monocyclic Erlang can be extended easily to other canonical cyclic representations analogous to Cox or hyperexponential representations. A class of particular interest in the following is obtained by considering convex combinations of generalized monocyclic Erlang distributions. We obtain then *Mixtures of generalized Monocyclic Erlang representations* (MME).

A new type of representation can also be obtained by adding feedbacks to the Cox representation. It is called the *feedback Cox*. A particular form of such a representation (which contains only the outer feedback) was called a *unicyclic representation* in O'Cinneide (1999). This feedback Cox representation, whose graph is depicted in figure 11, recalls us the canonical observable form and therefore allows hope for a large power of representation. Although several authors have suggested this representation, no evidence was given of the power of representation via a generalization of Theorem 5 for the cyclic case. We conclude this section with the question.

Problem 2: Investigate the power of representation of the feedback Cox representations.

5. Some properties of phase-type representations

In this section we give some general properties of phase-type representations which will be useful later in the search for a representation of a given phase-type distribution. They will in particular induce lower bounds on the order of possible representations. The first theorem comes from the numerous studies which have been done on the eigenvalue location of stochastic matrices following, for example, the Perron–Frobenius theorem. The result was obtained independently in Dmitriev and Dynkin (1945) and O'Cinneide (1991).

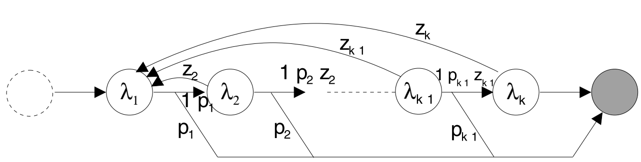

Figure 11. Graph of the Markov chain of a feedback Cox representation.

Theorem 6 (Dmitriev and Dynkin 1945, O'Cinneide 1991)**:** *Let T be a phase-type generator of order n. Let $-\lambda_1$, $\lambda_1 > 0$, be its eigenvalue with maximal real part and $-\lambda_2 \pm i\theta$, $\lambda_2 > 0$ and $\theta > 0$, be any pair of its complex eigenvalues. The following relation is satisfied*

$$\frac{\theta}{\lambda_2 - \lambda_1} \leq \cot \frac{\pi}{n} \tag{27}$$

The previous relation may be seen as

$$n \geq \frac{\pi}{\arctan\left(\dfrac{\lambda_2 - \lambda_1}{\theta}\right)} \tag{28}$$

Since the poles of the Laplace transform of the distribution are eigenvalues of the T matrix of any representation, it follows that the order of the representation increases when the angle between the position of complex poles and the vertical line passing through the real dominating pole decreases. This makes part 2 of Theorem 1 more precise and gives a first bound on the order of possible representations of a given distribution. This result definitely kills the hope that one could find a representation whose order is the degree of the distribution (recall that the degree of a distribution is the polynomial degree of the denominator of its Laplace transform). The previous theorem gives us interesting and general information about the pole location of the Laplace transform of a PH distribution, but nothing general is known about the zeros, so we can ask the question, which is a relaxed version of Problem 1.

Problem 3: Find general properties about the zero location, or the relative pole/zero location, of the Laplace transform of a PH-distribution.

For contributions to this problem see Zemanian (1959, 1961) and Sumita and Masuda (1987). But these results concern sufficient conditions which moreover are explicit only for small degree distributions.

Up to now we have not mentioned the parameters which are of main interest for probabilists. Consider a probability distribution μ. We denote by $m(\mu)$ its mean value and $\sigma(\mu)$ its standard deviation. We call coefficient of variation the value

$$Cv(\mu) = \frac{\sigma(\mu)}{m(\mu)} \tag{29}$$

The coefficient of variation is a normalized measure of the variability of the distribution. It is well known that the exponential distribution of rate λ has a mean $m = 1/\lambda$ and a standard deviation $\sigma = 1/\lambda$, therefore its coefficient of variation is $Cv = 1$. It is of interest to know how phase-type distributions fit in with given pre-specified probabilistic features, in particular a given mean and a given coefficient of variation. The first answer is given in the following theorem.

Theorem 7 (Aldous and Shepp 1987): *Consider an order* n *representation and let* μ *be the corresponding distribution, we have*

$$Cv(\mu) \geq \frac{1}{\sqrt{n}}. \tag{30}$$

Moreover, the equality holds only in the case of the n state Erlang representation.

From this theorem it follows that getting a low coefficient of variation implies a high order representation. Also the best way to approximate a constant time is to use an Erlang representation. On the other hand, it is easy to check that with a two state hyperexponential representation we can fit any positive mean value and any coefficient of variation greater than one. In fact, the previous theorem shows that the coefficient of variation reduction implies series connection of states. This can be made more precise concerning the structure of acyclic representations in the following corollary.

Corollary 1: *Consider an acyclic PH-representation and let m be the maximal number of states which are visited before absorption. The coefficient of variation of the corresponding distribution* μ *satisfies*

$$Cv(\mu) \geq \frac{1}{\sqrt{m}} \tag{31}$$

The proof is postponed to the Appendix. The previous corollary gives structural information on the possible representations of a PH-distribution. Further structural information appears in the following theorem.

Theorem 8 (Commault and Chemla 1996): *Consider a PH-representation and let* μ *be the corresponding distribution. Denote by* $\tilde{f}(s) = p(s)/q(s)$ *the Laplace transform of this distribution, where* $p(s)$ *and* $q(s)$ *are coprime polynomials. Then the difference* $\deg(q(s)) - \deg(p(s))$ *is equal to the minimal number of states which are visited before absorption in the representation.*

This theorem establishes a relation between the difference of degrees in the denominator and numerator, and the length of a shortest path in the graph of any representation. Recall that this difference of degrees is, up to one, the number of null derivatives at $t = 0$ of the probability density function. This number, which is known as the infinite zero order for a usual transfer function, is related to the minimal number of integrators we must cross before reaching the output in a state space realization. However, the equality holds only generically (Commault *et al.* 1991, van der Woude 1991). It is the set of constraints on the matrices α, T which insures that the result always holds true for PH-representations.

The results of this section will be guides in the search for a representation of a given PH-distribution. Let us now tackle this problem that is the realization problem in control theory.

6. Getting representations from PH-distributions

We now attack the problem: given a PH-distribution, that is a probability distribution defined on $[0, \infty)$ which satisfies the conditions of Theorem 1, find a representation for this distribution. We are interested, if possible, in getting sympathic representations. Sympathic representations may be representations of minimal order or with a special structure (something like our reachable or observable canonical forms). These representations with a special structure are desired because they generally avoid an overparameterization. The minimal representation problem is still open and, in general, we are far from having a complete realization theory as in the usual linear systems context. In the first subsection we present the general realization theorems.

6.1. *General representation results*

As noted before, since in an acyclic representation the states may be numbered so that the T matrix is triangular, the poles of the corresponding PH-distribution are real. It happens that the converse result is true, that is given a PH-distribution with real poles one can find a triangular or acyclic representation. As we noted previously that acyclic representations are equivalent to ordered Cox representations, we finally have the theorem.

Theorem 9 (O'Cinneide 1991): *Let* μ *be a PH-distribution whose Laplace transform has only real poles. Then* μ *has an ordered Cox representation of some order.*

The proof of the sufficiency part of the theorem is constructive and provides us with such a representation. A similar result was obtained in the general case using a suitable generalization of Cox representations for complex poles. This generalization relies on the notion of monocyclic Erlang representations which were introduced in § 4.

Theorem 10 (Mocanu and Commault 1999): *Let μ be a PH-distribution, μ has a representation which is a mixture of monocyclic Erlang (MME).*

The proofs of both theorems are based on convex analysis and may be summarized as follows:

- Find a generator T whose eigenvalue set contains the poles of the Laplace transform of the distribution.
- Construct a pseudo-representation (α', T) where α' is not necessarily stochastic.
- Construct Euler approximants in a suitable space of distributions. Each step in the approximation corresponds to the addition of a new state to the representation.
- Obtain a representation of the distribution.

Although some tricks may be used to limit the number of additional states, there is no guarantee that the obtained representation will be minimal. For a practical implementation of this method see http://www.multimania.com/mocanu/

6.2. *When the order of the representation equals the degree of the distribution*

Given a distribution, there generally does not exist a representation whose order is the degree of the distribution. But it is of interest to characterize the representations whose corresponding distribution has a degree equal to the order. This is for us clearly related to the notions of reachability and observability of a realization.

Definition 2 (O'Cinneide 1989): A PH generator T is simple if any two initial probability distributions $\alpha_1 \neq \alpha_2$, give rise to two different PH-distributions. That is $PH(\alpha_1, T) \neq PH(\alpha_2, T)$.

A representation defined by the matrix T is simple if and only if the matrix

$$R = [v, Tv, \ldots, T^{n-1}v] = -T[\vec{1}, T\vec{1}, \ldots, T^{n-1}\vec{1}] \quad (32)$$

has rank n. This notion is very close to our observability notion. Up to now, a direct counterpart of reachability is not known. This raises a new question.

Problem 4: Give a natural interpretation of the counterpart of reachability for PH-representations

However, we can get such a counterpart through the duality notion which is related to the reverse time point of view (Ramaswami 1990). Consider a PH-representation (α, T), and denote by τ the absorption instant. We call the *dual or reverse time representation* the Markov process in which we are in state i at time t if we are in state i of the original process at time $(\tau - t)$. The matrices of this representation are obtained as

$$\alpha* = v^{\mathrm{T}}M, \quad v* = M^{-1}\alpha^{\mathrm{T}}, \quad T* = M^{-1}T^{\mathrm{T}}M \quad (33)$$

where the matrix M is just a scaling diagonal matrix.

$$M = \mathrm{diag}(m_1, \ldots, m_n) \quad (34)$$

The row vector $m = (m_1, \ldots, m_n)$ is given by

$$m = -\alpha T^{-1} \quad (35)$$

Some interesting properties of the reverse time representation can be derived (Commault and Chemla 1993).

- The representation and its reverse time representation give rise to the same PH-distribution
- The two representations have the same number of states and there is a one-to-one correspondence between these states. The corresponding states have the same outgoing rate.
- m_i is the average time which is spent in state i before absorption.
- The graph of the reverse time representation is the reverse graph of the initial one. More precisely, it is obtained by reversing the direction of all edges. The initial vertex becomes the absorbing one and conversely.

From our first courses in control theory the following result is readily obtained.

Theorem 11 (Commault and Chemla 1993): *An order n representation (α, T) which is simple and whose reverse time representation is simple gives rise to a distribution $PH(\alpha, T)$ which has degree n.*

This theorem characterizes representations which cannot be reduced but it is not very helpful in finding representations.

Let us now give some examples of distributions of a given degree n for which we can provide an order n representation. First note that a distribution whose Laplace transform is

$$\tilde{f}(s) = \left(\frac{\lambda_1}{s + \lambda_1}\right) \cdot \left(\frac{\lambda_2}{s + \lambda_2}\right) \cdots \left(\frac{\lambda_n}{s + \lambda_n}\right) \quad (36)$$

can be represented by a convolution of n exponential states. This result can be generalized as follows.

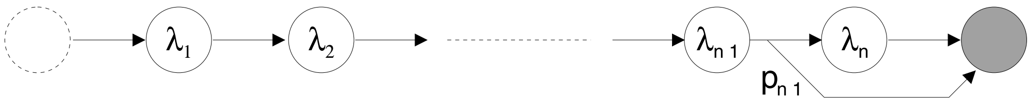

Figure 12. The Cox representation of the distribution (37).

Theorem 12 (Commault and Chemla 1996)**:** *Consider a PH-distribution with Laplace transform $\tilde{f}(s) = p(s)/q(s)$, where $p(s)$ and $q(s)$ are coprime polynomials, such that $q(s)$ has degree n with n real roots and $p(s)$ has degree less than or equal to one. This distribution has an order n representation.*

Denote $p(s) = (1 + s/\mu)$, with $\mu \geq 0$, and

$$q(s) = \prod_{i=1}^{n}\left(1 + \frac{s}{\lambda_i}\right), \qquad \lambda_1 \geq \lambda_2, \ldots \geq \lambda_n > 0 \quad (37)$$

It can be shown that the inverse Laplace transform of $\tilde{f}(s)$ is positive only if $\mu \geq \lambda_n$. Under this assumption it is easy to check that the ordered Cox representation of figure 12 with $p_{n-1} = \lambda_n/\mu$ corresponds to this distribution. A generalization of the previous theorem can be obtained but only as of a sufficient condition.

Theorem 13: *Consider a PH-distribution with Laplace transform $\tilde{f}(s) = p(s)/q(s)$, where $p(s)$ and $q(s)$ are coprime polynomials with real roots*

$$p(s) = \prod_{i=1}^{m}\left(1 + \frac{s}{\mu_i}\right), \qquad \mu_1 \geq \mu_2, \ldots \geq \mu_m > 0 \quad (38)$$

and

$$q(s) = \prod_{i=1}^{n}\left(1 + \frac{s}{\lambda_i}\right), \qquad \lambda_1 \geq \lambda_2, \ldots \geq \lambda_n > 0, \quad n > m \quad (39)$$

If

$$\mu_m \geq \lambda_n, \ldots, \mu_1 \geq \lambda_{n-m+1} \quad (40)$$

this distribution has an order n representation.

Proof: The Laplace transform can be written

$$\tilde{f}(s) = \prod_{i=1}^{n-m}\frac{1}{(1 + (s/\lambda_i))}\prod_{j=n-m+1}^{n}\frac{(1 + (s/\mu_{j-n+m}))}{(1 + (s/\lambda_j))} \quad (41)$$

A representation is obtained through a convolution of $(n - m)$ exponential states and m stages which have the form of the rightmost part of figure 12. □

In the case of complex poles of the Laplace transform, to our knowledge, the only general result is the following.

Theorem 14 (Commault and Chemla 1996)**:** *Consider a PH-distribution with Laplace transform $\tilde{f}(s) = 1/q(s)$*

$$q(s) = \left(1 + \frac{s}{c}\right)\left(1 + \frac{s}{a + ib}\right)\left(1 + \frac{s}{a - ib}\right) \quad (42)$$

where $a, b, c \geq 0$. This distribution has an order 3 representation if and only if

$$b \leq \frac{a - c}{\sqrt{3}} \quad (43)$$

It is clear from Theorem 6 that condition (43) is necessary for obtaining an order 3 representation. In fact, using Theorem 8 and some simplifying tricks we can always obtain a representation which has the graph given in figure 13.

For PH-distributions with Laplace transform $\tilde{f}(s) = 1/q(s)$ where $q(s)$ is as in (42), Theorem 6 gives a minimal order n_{min} for a representation. An interesting question is the following.

Conjecture 1: *There exists a representation of order $n_{\min}$ for such a distribution.*

In Mocanu (1999) it is only proved that we can obtain a representation of order $(n_{\min} + 1)$.

A more interesting problem, both from the theoretical and practical points of view, would be the characterization of order 3 PH-distributions. Since it can be easily proved that order 2 PH-representations must have real eigenvalues, the order 3 representations appear to be the smallest ones with a sufficient power of representation (or approximation). In that sense they probably could have the same importance for PH-distributions as order 2 linear systems for classical control applications. We can state our question.

Problem 5: Consider a PH-distribution whose Laplace transform is $\tilde{f}(s) = p(s)/q(s)$, where $p(s)$ has degree two and $q(s)$ degree three. Under which conditions (besides condition (43)) has this distribution an order 3 representation?

The question was solved in Anderson *et al.* (1996) for discrete-time positive systems, but the solvability condition is not easy to check.

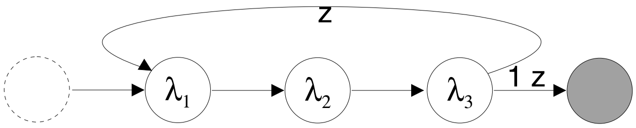

Figure 13. An order 3 representation for the PH-type distribution (42) under the condition (43).

7. On cyclicity of representations

It appears that the cyclicity issue plays a crucial role in the search for representations. In some cases the cyclicity is compulsory, that is when the Laplace transform of the distribution has non-real poles. From Theorem 9, when all the poles are real it follows that acyclic representations exist. But in some cases there may exist cyclic representations which may be of a smaller order than any acyclic one. To make more precise this questions let us go back to our example.

7.1. *Example*

Consider again Example 1. From the physical modelling of the problem we obtained a cyclic representation. It is also clear that the more likely behaviour of the system consists in the cycling around states x_1, x_2, x_3, x_4, that is, the cyclicity here appears to be somewhat intrinsic. But if we compute the poles of $\tilde{f}(s)$, these poles are real and from Theorem 12 we know that there exists a Cox representation of order 4 for this distribution. If we change the input data to reinforce the cyclicity (in an intuitive sense) by increasing for example the repair rate to $\mu = 20$, we obtain

$$\tilde{f}(s) = \frac{s + 21}{s^4 + 44\,s^3 + 506\,s^2 + 884\,s + 21} \tag{44}$$

but the poles remain real. Even more paradoxical, the non-real poles appear for small values of μ. It can be proved that non-real poles occur for $\mu \in (0,\ 5.729)$. For example, with $\mu = 5$ we get

$$\tilde{f}(s) = \frac{s + 6}{s^4 + 14\,s^3 + 56\,s^2 + 74\,s + 6} \tag{45}$$

which has a pair of non-real poles. See also the behaviour of the feedback Erlang where the poles are non-real for any value of the feedback probability.

From this discussion it appears that the relation between the cyclicity of the representation and the existence of non-real poles is not yet fully understood.

7.2. *The triangular order*

Since distributions with real poles have acyclic representations, it is natural to define the *triangular order* as the minimal order of an acyclic representation of this distribution. The definition and some properties of the triangular order were given by O'Cinneide in (1999). The paper also contains a characterization of this order. This characterization allows the author to compute the triangular order for a distribution whose Laplace transform is $\tilde{f}(s) = p(s)/q(s)$, where $p(s)$ is a polynomial of degree 2 and $q(s) = (1 + s/\lambda)^3$. Unfortunately this is not a general practical tool to compute the triangular order for a given distribution. Moreover, it is easy to build examples where the triangular order is strictly greater than the order. An example with order 3 and triangular order 4 is given in Chemla (1993) (see also Botta *et al.* 1987, Harris *et al.* 1992).

It turns out that we can get a modest generalization of Theorem 12 for a case where the triangular order is equal to the degree of the distribution.

Theorem 15 (Commault and Chemla 1996)**:** *Consider a PH-distribution with Laplace transform* $\tilde{f}(s) = p(s)/q(s)$*, where* $p(s)$ *and* $q(s)$ *are coprime polynomials with real roots, such that*

$$p(s) = (1 + s/\mu_1)(1 + s/\mu_2), \qquad \mu_1 \geq \mu_2 > 0 \tag{46}$$

$$q(s) = \prod_{i=1}^{n}\left(1 + \frac{s}{\lambda_i}\right), \qquad \lambda_1 \geq \lambda_2, \ldots \geq \lambda_n > 0 \tag{47}$$

The distribution has triangular order n if and only if

$$\mu_2 > \lambda_n \tag{48}$$

and

$$(\mu_1 + \mu_2) \geq (\lambda_{n-1} + \lambda_n). \tag{49}$$

7.3. *What is a good representation?*

Since the generator T is to be used in matrix computations for performance evaluation, it is of interest that this matrix has the least possible order and has a nice structure. In particular the numerical inversion of T or the formal inversion of $(sI - T)$ would highly benefit from a triangular form. In linear system theory we are used to the comfortable situation where we can achieve at the same time the minimal order and various canonical forms. This is no longer the case for the representation problem and in particular, as seen in the previous subsection, when triangular representations exist they may be of an order strictly larger than a cyclic one. Comparing the respective benefits of the incompatible goals of low order and triangular form raises the question: what is really a good practical representation? As far as cyclic representations are concerned, it has been proved (Maier 1993) that representations with embedded cycles are equivalent to representations with independent cycles (the 'no cycles within cycles' condition). Therefore, even if cyclicity is obligatory, this cyclicity can be made as less intricate as possible.

Although the minimal representation problem remains a wonderful theoretical problem our feeling is that in practice canonical sparse forms should be used. For example, the constructive procedure in the proof of Theorems 9 and 10 gives rise to representations with a very interesting structure and of a reasonable order when they are implemented with some care.

8. Conclusion

In this paper we have tried to make the system theory community aware of the problem of phase-type representations. This problem which is of a realization type has a lot of things in common with the classical linear system realization problem. It has also a lot of proper features which make it rather difficult and interesting. The general problem of finding a minimal representation for a given PH-distribution is still open and appears to be so difficult that it seems hopeless to attack it in its generality. Instead, we suggest that a set of partial problems could be solved to better understand the properties of the representations and incidentally solve some practical problems. In particular we feel that a better understanding of the properties of low order PH-distributions together with a complete characterization of their power of representation and approximation would certainly enhance considerably the appeal of this field of research.

We have also in this paper put a bridge between the PH-representation problem and the positive realization theory. These two fields had their own development starting with a common root which was the Perron–Frobenius theory. Our contribution, in particular through Theorem 4, is to make clear that the problems are essentially the same and that most of the results obtained in one of these fields could be translated in the other one. For example the structural results and canonical forms obtained for PH-representations have their natural counterparts for positive systems. Finally, we think that the communities of control and probability theory should work together on the exciting remaining open problems.

Appendix

Proof of Corollary 1: As noted before, an acyclic representation can be transformed into a representation which is a convex combination (or mixture) of generalized exponentials (MGE). In this MGE, the longest path has the same length as the longest path in the original representation. The length (or number of edges) of this path in the graph is in fact the maximal number m of states which are visited before absorption plus one. If we consider all the distributions corresponding to paths in the MGE, the theorem says that their coefficient of variation is greater than or equal to $1/\sqrt{m}$. If we can prove that the convex combination produces a coefficient of variation which is greater than or equal to the smallest coefficient of variation of the paths, we are done. If we prove this for two distributions the general result will follow easily. Consider two distributions μ_1 and μ_2 (which need not be phase-type), with respective mean values m_1 and m_2, standard deviations σ_1 and σ_2, coefficients of variation Cv_1 and Cv_2. We assume that $Cv_1 \le Cv_2$. We consider the convex combination with probabilities p and $(1-p)$ of these distributions, called μ. We get for μ the first and second moments which are

$$E(\mu) = pm_1 + (1-p)m_2$$

and

$$E(\mu^2) = pE(\mu_1^2) + (1-p)E(\mu_2^2) = p(\sigma_1^2 + m_1^2) + (1-p)(\sigma_2^2 + m_2^2)$$

Inserting these expressions in

$$Cv^2(\mu) = \frac{E(\mu^2) - E(\mu)^2}{E(\mu)^2}$$

we get after some calculations

$$Cv^2(\mu) = \frac{p(1-p)(m_1 - m_2)^2 + p\sigma_1^2 + (1-p)\sigma_2^2}{E(\mu)^2}$$

then

$$Cv^2(\mu) \ge \frac{pm_1^2 Cv_1^2 + (1-p)m_2^2 Cv_2^2}{E(\mu)^2}$$

Since $Cv_1^2 \le Cv_2^2$ we get

$$Cv^2(\mu) \ge Cv_1^2 \frac{pm_1^2 + (1-p)m_2^2}{(pm_1 + (1-p)m_2)^2}$$

It is easy to see that

$$pm_1^2 + (1-p)m_2^2 \ge (pm_1 + (1-p)m_2)^2$$

And finally

$$Cv^2(\mu) \ge Cv_1^2$$

which ends the proof. □

References

Aldous, D., and Shepp, L., 1987, The least variable phase-type distribution is Erlang. *Communications in Statistics—Stochastic Models*, **3**, 467–473.

Anderson, B. B. O., Deistler, M., Farina, L., and Benvenuti, L., 1996, Nonnegative realization of a linear system with nonnegative impulse response. *IEEE Transactions on Circuits and Systems*, *I*, **43**, 134–142.

Berman, A., Neumann, M., and Stern, R. J., 1989, *Nonnegative Matrices in Dynamic Systems* (New York: Wiley).

Berman, A., and Plemmons, R. J., 1994, *Nonnegative Matrices in the Mathematical Sciences*, volume 9 of *Classics in Applied Mathematics* (Philadelphia: SIAM).

Botta, R. B., Harris, C. M., and Marchal, W. G., 1987, Characterization of generalized hyperexponential distribution functions. *Communications in Statistics—Stochastic Models*, **3**, 115–148.

Chemla, J.-P., 1993, Les lois de type phase: caractérisation et représentations. PhD thesis, INPG, Grenoble, France.

Commault, C., and Chemla, J.-P., 1993, On dual and minimal phase type representations. *Communications in Statistics—Stochastic Models*, **9**, 421–434.

Commault, C., and Chemla, J.-P., 1996, An invariant of representations of phase-type distributions and some applications. *Journal of Applied Probability*, **33**, 368–381.

Commault, C., Dion, J. M., and Perez, A., 1991, Disturbance rejection for structured systems. *IEEE Transactions on Automatic Control*, **36**, 884–887.

Cox, D. R., 1955, A use of complex probabilities in the theory of stochastic processes. *Proceedings of the Cambridge Philosophical Society*, **51**, 313–319.

Cumani, A., 1982, On the canonical representation of homogenous Markov processes modelling failure-time distributions. *Microelectronics and Reliability*, **22**, 583–602.

Dmitriev, N., and Dynkin, E. B., 1945, On the characteristic numbers of a stochastic matrix. *Comptes Rendus (Doklady) de l'Académie de Sciences de l'URSS (Nouvelle Série)*, **49**, 159–162.

Erlang, A. K., 1917, Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Post Office Electrical Engineers' Journal*, **10**, 189–197.

Farina, L., 1996, On the existence of a positive realization. *Systems and Control Letters*, **28**, 219–226.

Farina, L., and Rinaldi, S., 2000, *Positive Linear Systems* (New York: Wiley).

Harris, C. M., Marchal, W. G., and Botta, R. B., 1992, A note on generalized hyperexponential distributions. *Communications in Statistics—Stochastic Models*, **8**, 179–191.

Jensen, A., 1949, Distribution patterns composed of a limited number of exponential distributions. In *Den 11. Skandinaviske Matematikerkongres*, pp. 209–215.

Johnson, M. A., and Taaffe, M. R., 1990 a, Matching moments to phase distributions: Density function shapes. *Communications in Statistics—Stochastic Models*, **6**, 283–306.

Johnson, M. A., and Taaffe, M. R., 1990 b, Matching moments to phase distributions: Nonlinear programming approaches. *Communications in Statistics—Stochastic Models*, **6**, 259–281.

Johnson, M. A., and Taaffe, M. R., 1991, A graphical investigation of error bounds for moment-based queueing approximations. *Queueing Systems*, **8**, 295–312.

Latouche, G., and Ramaswami, V., 1993, A logarithmic reduction algorithm for quasi-birth-death processes. *Journal of Applied Probability*, **30**, 650–674.

Luenberger, D., 1979, *Introduction to Dynamical Systems: Theory, Models and Applications* (New York: Wiley).

Maeda, H., Kodama, S., and Kajiya, F., 1981, Compartmental systems analysis: realization of a class of linear systems with physical constraints. *IEEE Transactions on Circuits and Systems I*, **28**, 39–47.

Maier, R. S., 1991, The algebraic construction of phase-type distributions. *Commun. Stat., Stochastic Models*, **7**, 573–602.

Maier, R. S., 1993, Phase-type distributions and the structure of finite Markov chains. *Communications in Statistics—Simulation and Computation*, **46**, 449–453.

Maier, R. S., and O'Cinneide, C. A., 1992, A closure characterisation of phase-type distributions. *Journal of Computational and Applied Mathematics*, **29**, 92–103.

Mocanu, S., 1999, Construction et propriétés des représentations monocycliques des lois de type phase. PhD thesis, INPG.

Mocanu, S., and Commault, C., 1999, Sparse representations of phase-type distributions. *Communications in Statistics—Stochastic Models*, **15**, 759–778.

Neuts, M., 1975, Distributions of phase-type. In *Liber Amicorum Professor Emeritus H. Florin*, pp. 173–206.

Neuts, M., 1981, *Matrix-Geometric Solutions in Stochastic Models. An Algorithm Approach* (Baltimore and London: The John Hopkins University Press).

Neuts, M., 1984, The abscissa of convergence of the Laplace-Stieltjes transform of a PH-distribution. *Communications in Statistics—Simulation and Computation*, **13**, 367–373.

O'Cinneide, C. A., 1989, On non-uniqueness of representations of phase-type distribution. *Communications in Statistics—Stochastic Models*, **5**, 247–259.

O'Cinneide, C. A., 1990, Characterization of phase-type distributions. *Communications in Statistics–Stochastic Models*, **6**, 1–57.

O'Cinneide, C. A., 1991, Phase-type distributions and invariant polytopes. *Advances in Applied Probability*, **23**, 515–535.

O'Cinneide, C. A., 1993, Triangular order of triangular phase-type distributions. *Communications in Statistics—Stochastic Models*, **9**, 507–529.

O'Cinneide, C. A., 1999, Phase-type distributions: open problems and a few properties. *Communications in Statistics—Stochastic Models*, **15**.

Ramaswami, V., 1990, A duality theorem for matrix paradigms in queueing theory. *Communications in Statistics—Stochastic Models*, **6**, 151–161.

Ryden, T. R., 1996, On identifiability and order of continuous-time aggregated Markov chains, Markov modulated Poisson processes, and phase-type distributions. *Journal of Applied Probability*, **33**, 640–653.

Sumita, U., and Masuda, Y., 1987, Classes of probability density functions having laplace transforms with negative zeros and poles. *Advances in Applied Probability*, **19**, 632–651.

van der Woude, J. W., 1991, On the structure at infinity of structured systems. *Linear Algebra and its Applications*, **148**, 145–169.

Zemanian, A. H., 1959, On the pole and zero locations of rational Laplace transforms of non-negative functions. *Proceedings of the American Mathematical Society*, **10**, 868–872.

Zemanian, A. H., 1961, On the pole and zero locations of rational Laplace transforms of non-negative functions II. *Proceedings of the American Mathematical Society*, **12**, 870–874.