periodica polytechnica

Electrical Engineering
54/1-2 (2010) 11–19
doi: 10.3311/pp.ee.2010-1-2.02
web: http://www.pp.bme.hu/ee

RESEARCH ARTICLE

Spectral conditions for Phase-type representations

Tamás Demián

Received 2010-05-17

Abstract

Current paper tries to find appropriate similarity transformation that could convert a given ME (Matrix Exponential) representation to a more favorable PH (Phase-type) representation. As the main result of this paper, we give necessary conditions for the existence of such a representation. We also give methods for the search and provide conjectures on necessary and sufficient conditions too. PH distribution is the distribution of the time until absorption into the absorbent state in a Markov chain. If the arrival and service time distributions are PH distributions in a queuing system, we can use simple linear algebraic methods to derive the most important features or to perform simulation. Robust methods exist that can approximate any distribution with a ME distribution (with respect to a given measure and matrix order), but the PH transformation have not been sufficiently examined yet. This transformation is the object of the current presentation.

Keywords

Markov chain · PH distribution · ME distribution · traffic models · fitting · gradient method · cyclic matrix

Acknowledgement

The author thanks the help of Miklós Telek, Katalin Friedl and Pál Rózsa.

Tamás Demián

Department of Telecommunications, BME, H-1521 Budapest, Magyar tudósok krt. 2, Hungary

e-mail: demiantamas@vnet.hu

1 Introduction

If we want to model telecommunication networks, where the global behavior is very complex, first we should look for flexible and tractable mathematical tools. It is crucial in the network dimensioning, in the design of routing protocols, and during the operation too. A queue is usually defined by the distribution of the differences between consecutive arrival times of the resources, and by the distribution of the serving time differences (see the formalism of Kendall). The simplest resource flows follow a memoryless Poisson process, where the time distributions between the consecutive arrivals are exponentially distributed.

In this case the behavior of the queues can be characterized by the help of algebraic methods. This process allows us to use the methods of the Markov chains. We can derive the most important parameters (loss, distribution of the delay, etc.) of the system directly by solving linear equations without any simulation or statistical measurements. It can be shown how this advantageous property survives if the arrival/service processes follow the so called PH or ME processes (see [2] for PH Renewal Process, Markov Arrival Process (MAP), Matrix Exponential Arrival Process (MEP) and Quasi Birth Death Process (QBD)). As we will see, the set of ME distributions includes the set of PH distributions. The difference is that PH distributions have an expressive stochastic interpretation based on Markov-chains unlike ME distributions. However, the major part of the formulas used in the PH world is applicable in the class of ME distributions too. An important exception is e.g. the so called randomization, which gives already a practical reason for studying PH representations.

PH and ME distributions have been studied for a long time. Some important or classic results can be found here: [14][15]... The class of PH distributions is dense and hence any distribution on $[0, \infty]$ can be approximated arbitrarily close by a PH distribution. Some main studies of this approximation problem are Bux & Herzog [8], Lazowska & Addison [9] and Thummler & Buchholz & Telek [10].

Bux and Herzog have implemented a non-linear estimation approach based on the matching of the first two moments coupled with the minimization of a distance measure with respect of

the empirical distribution. Lazowska and Addison [9] provide a technique for determining a method that matches the mean and an arbitrary number of percentiles of an arbitrary distribution. Thummler, Buchholz and Telek [10] provide an efficient and numerically stable fitting method that fits a restricted class of phase-type distributions, namely mixtures of Erlang distributions, to trace data. The ME class has been proposed in [11]. The ME class includes the PH class while preserving the most of its useful properties.

A robust methods exist that can approximate any distribution with a ME distribution with respect to a given number of moments [3], but methods transforming ME representations to PH representations have not been sufficiently studied yet. This step is the object of the current paper. Since PH $\subset$ ME the existence of such a transformation is not guaranteed. As the main result of this paper, we give necessary conditions for the existence of such a representation. Stefanita Mocanu also gave a spectral condition for PH representations [12], but the strengths have not been compared yet. The major part of the current paper focuses on the so called relaxed problem, where we ignore the initial probability vector and examine only the form of the generator matrix.

In Section 2 we discuss the basic concepts and properties of ME and PH distributions. It does not contain any new results.

In Section 3 we derive the the main result of this paper: Theorem 3.5. It provides $n-1$ necessary constraints for the existence of PH representations. The constraints include only the spectral invariants of the initial representation. We do not know whether these conditions are sufficient. Theorem 3.6 is a modified version regarding to the so called phase process. This condition implicitly includes the initial vector so we take an analytic step forward to the complete (non-relaxed) PH-fitting problem too. The efficiency of the theorems is pointed out in the last sections.

In Section 4 we introduce the well known cyclic matrices. As we will show, these matrices play an important role in the PH world but they have not been examined yet in this context. These matrices have some advantageous properties beside serious drawbacks. In Theorem 4.3 we show that it is easy to find a cyclic form for a given spectrum and we point out that in the case of cyclic matrices the sufficient constraints for the existence of PH representations get simpler. The number of constraints of the relaxed problem reduced from n^2 to $n-1$. These properties inspired some conjectures on these matrices. The conjectures gave clear necessary and sufficient conditions for the relaxed problem. On the other side, cyclic matrices can not be diagonalized by a valid similarity transformation and can generate only simple exponential distributions. However, small perturbations can solve these problems. It is also an open problem at the moment, whether the $n-1$ sufficient conditions for cyclic representations and the $n-1$ necessary conditions for general representations in the previous section are equivalent.

In Section 5 we propose two straightforward and useful measures for gradient algorithms which work on the space of the matrix entries. Theorem 5.2 allows us to improve the 'goodness' of representation using gradient methods with a differentiable error measure. Theorem 5.3 can be used in general, for the correction of non-valid transformation matrices. Theorem 5.4 states that cyclic matrices fulfil the necessary conditions of a local maximum. So this result also supports our conjectures.

In Section 6 case studies will demonstrate the efficiency of our spectral conditions.

2 Basic properties

PH distribution is the distribution of the time until absorption into the absorbent state in a continuous time Markov chain (we focus on continuous Phase Type Distribution in this paper).

The row of the absorbent state is redundant in the generator matrix, since in this state there is no escape to other states. The intensities of these transitions are zero. The column of the absorbent state is also redundant, since the rowsums are zero in the generator matrix (in the continuous case). The 'transient generator matrix' (A) of a PH distribution is the generator matrix of the Markov chain without the corresponding row and column of the absorbent state. The order (n) of a PH distribution is the size of A. In the followings we use the term 'initial probability vector' (π) without the corresponding probability of the absorbent state, since this value is also redundant.

The probability density function and the cumulative distribution function of the absorption time is the following:

$$f(t) = \pi e^{At} a, \tag{2.1}$$

$$F(t) = 1 - \pi e^{At} \mathbb{1}, \tag{2.2}$$

where $\mathbb{1}$ is the column vector of ones and $a = -A\mathbb{1}$ is the column vector of the absorbent transition intensities.

If $n \to \infty$ then any distributions can be approximated by PH distributions with arbitrary accuracy, but some properties of the PH distributions always survive:

- the domain of PH distributions is infinite: $[0, \infty]$
- that is why the variance can not reach zero: $\sigma \geq \frac{1}{n} > 0$

Now we summarize the constraints over the (A, π) representation:

$$\begin{aligned} \forall i, j, i \neq j: \quad & A_{ij} \geq 0 \\ \forall i: \quad & a_i \geq 0 \\ \forall i: \quad & \pi_i \geq 0 \\ & 1 - \pi \mathbb{1} \geq 0 \end{aligned}$$

The constraints over A imply that the diagonal entries are negative. These inequalities simply come from the representation of the Markov chains. Roughly speaking, these constraints make the difference between PH and ME distributions.

ME distribution is a time distribution that can be expressed by Eq. 2.2 (and by Eq. 2.1) where there are no constraints over the (A, π) representation. The only constraint is the next:

$$\forall t \geq 0 : f(t) \geq 0, F(\infty) = 1$$

namely $f(t)$ is a probability density function. These conditions hold automatically in the case of PH distributions, $PH \subset ME$. We will use terms 'valid transient generator matrix' and 'valid initial probability vector' in the case of ME representations, if the appropriate PH constraints hold for A or π. If both of them are valid then we have a PH representation. If we demand the validity of A only and neglect the form of π, we call it 'relaxed problem'.

Definition 2.1 *Two ME representations – ME(π, A) and ME(π', A') – are* similar *, if there exists such an invertible B transformation matrix that*

$$A' = B^{-1}AB, \tag{2.3}$$

$$B\mathbb{1} = \mathbb{1}, \tag{2.4}$$

$$\pi' = \pi B. \tag{2.5}$$

B is a similarity transformation over the representation. If Eq. (2.4) holds for B (the sum of all rows are one), we say B is 'valid'. The set of valid transformations is a group.

Theorem 2.1 *If two ME representations – ME(π, A) and ME(π', A') – are* similar*, then they generate the same ME distribution.*

Proof 2.1 $F'(t) = 1 - \pi' e^{A't}\mathbb{1} =$
$= 1-(\pi B)(e^{B^{-1}ABt})\mathbb{1} = 1-(\pi B)(B^{-1}e^{At}B)\mathbb{1} = 1-\pi e^{At}\mathbb{1} = F(t)$

Theorem 2.2 *If two ME representations of the same ME distribution have minimal order, then the representations are similar.*

Now we can define the subject of this article more formally: Given the initial ME representation we search for a similar PH representation (valid ME representation).

It is clear that for every ME representation there is a dominant λ_{max} eigenvalue that is real, and it has the highest real part. It means that the determinant of $-A$ is positive. If the dominant eigenvalue would not exist, the sign of the density function would alternate. If this value were not negative then $F(t)$ would not converge to 1. The uniqueness of the dominant eigenvalue does not hold in general (e.g. if $A = -I$).

3 The residual polynomial

In this section we prove the main theorem of the current paper. Theorem 3.5 is a very efficient tool for proving the non-existence of any valid representation. At first we derive an inequality over the coefficients of the characteristic polynomial and the so called diagonal polynomial defined below. Then we apply this inequality to the maximal diagonal polynomial too, which is representation independent unlike an arbitrary diagonal polynomial, so we get a stronger and representation independent statement: Theorem 3.5.

Lemma 3.1 [19]*(page 59.) If A is valid, then A^{-1} consists of only non positive entries.*

Definition 3.1 *$P(\lambda)$: Characteristic matrix: $\lambda I - A$*
$P(\lambda) = \sum_{i=0}^{n} P_i \lambda^i$: Characteristic polynomial: $|\lambda I - A|$
$D(\lambda) = \sum_{i=0}^{n} D_i \lambda^i$: Diagonal polynomial: $|\lambda I - A \circ I|$
$R(\lambda) = \sum_{i=0}^{n} R_i \lambda^i$: Residual polynomial: $R(\lambda) = P(\lambda) - D(\lambda)$

where $\circ$ denotes the element wise matrix product.

Lemma 3.2 *If A is valid, then the product of the positive diagonal entries of $-A$ is greater or equal than the product of the positive eigenvalues (the determinant) of $-A$:*

$$D_0 \geq P_0.$$

Proof 3.1

$$\frac{\partial \det(-A)}{\partial(-A)_{ij}} = [\mathrm{adj}(-A)]_{ji} =$$
$$[(-A)^{-1}]_{ji} \det(-A) = -[A^{-1}]_{ji} \det(-A) \geq 0, \tag{3.1}$$

if A is valid. We also know that $-A_{ij} \leq 0$ if $i \neq j$. So if we increase the non diagonal, negative entries of $-A$, then the determinant of the resulted matrix will not decrease. Moreover, the result is also the additive inverse of a valid transient generator matrix, so the properties above survive. It means that while we increase all the non diagonal (negative) entries of $-A$ up to zero, the determinant will not decrease. If all the non diagonal entries reach zero, the determinant will be the product of the remaining diagonal entries which is the 0-th coefficient of the diagonal polynomial.

Lemma 3.3 *If A is a valid transient generator matrix of order n then*

$$\forall i : D_i \geq P_i \text{ or } R_i \leq 0. \tag{3.2}$$

Proof 3.2 *We get the value of P_i if we get all the combinations of $(n-i)$ pieces of diagonal entries from the $[-A]$ matrix, calculate the determinant of the corresponding submatrices separately and finally sum these values. (It means $\binom{n}{n-i}$ submatrices.) We can apply the previous lemma on these submatrices and we can give an upper bound for the determinant of these submatrices by the product of the appropriate diagonal entries. The sum of these upper bounds is D_i.*

Some coefficients of $R(\lambda)$ have no information content. It is easy to see that $R_n = R_{n-1} = 0$.

We have to emphasize that the $D(\lambda)$ diagonal polynomial is representation-dependent. We can achieve arbitrary diagonal patterns using appropriate similarity transformations, but there is one constraint. The sum of the diagonal entries is independent from the representation.

Let us denote the ith diagonal entry of A by d_i: $d_i = A_{ii}$ and the average of the negative diagonal entries of A by $\bar{d}$.

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{-P_{n-1}}{n}$$

.

Lemma 3.4 *All coefficients of $D(\lambda)$ are maximal if the diagonal entries are equal (under the constraint of the next inequalities $\forall i : d_i \leq 0$ and $\sum_i d_i$ is constant).*

Proof 3.3 *Consider the case that there exists such a pair of roots (d_i, d_j) of $D(d)$ that fulfills the next inequality:*

$$d_i \leq \bar{d} \leq d_j.$$

We can always find such a pair, if the roots are not equal. Let us separate these roots from the polynomial:

$$D(d) = (d - d_i)(d - d_j)D_{ij}(d) = d^2 D_{ij}(d) - d(d_i + d_j)D_{ij}(d) + d_i d_j D_{ij}(d). \quad (3.3)$$

Since all the roots of $D(d)$ are negative, the coefficients of $D_{ij}(d)$ are also positive. Let c denote $\bar{d} - d_i$. Let us change the value of the root pair as follows:

$$\acute{d}_i = d_i + c = \bar{d}, \qquad \acute{d}_j = d_j - c.$$

It can be achieved by an appropriate similarity transformation. In this case the sum of the altered roots remains the same but the product increases:

$$\acute{d}_i\acute{d}_j = \bar{d}(d_j - \bar{d} + d_i) = -\bar{d}^2 + \bar{d}(d_i + d_j) - d_i d_j + d_i d_j = \\ = -(\bar{d} - d_i)(\bar{d} - d_j) + d_i d_j \geq d_i d_j,$$

so the altered roots result increased coefficients in the $D(d)$ polynomial (see Eq. (3.3)). If we repeat this root alteration (at most $(n-1)$ times) we arrive at such a $D(d)$ polynomial, where all roots are equal to $\bar{d}$. During this procedure the coefficients do not decrease. Since it works for arbitrary initial diagonal entries, the resulted polynomial provides the highest coefficient values for $D(d)$:

$$D_{max}(d) = (d - \bar{d})^n = \sum_{i=0}^{n} \binom{n}{i} (-\bar{d})^{n-i} d^i = \sum_{i=0}^{n} D_{max\ i} d^i$$

where

$$\forall i : D_i \leq D_{max\ i} = \binom{n}{i} \bar{d}^{n-i} \quad (3.4)$$

Definition 3.2 *$R_{min}(\lambda)$: Minimal residual polynomial: $R_{min}(\lambda) = P(\lambda) - D_{max}(\lambda) = P(\lambda) - (\lambda + \frac{P_{n-1}}{n})^n$ so $R_{min}(\lambda)$ is also independent from the representation.*

Now we can rewrite lemma 3.3 to a representation-independent form without loss of generality:

Theorem 3.5 *If*

$$\forall i : R_{min\ i} \leq 0 \quad (3.5)$$

does not hold then there is no valid representation.

Proof 3.4 *For any representation we can write $\forall i : R_{min\ i} \leq R_i$ using the previous lemma. If there is a valid representation, we can write $\forall i : R_i \leq 0$ using lemma 3.3. So in this case we can also write: $\forall i : R_{min\ i} \leq R_i \leq 0$.*

This theorem gives us $n - 1$ inequalities. We do not believe that this necessary condition is sufficient too, but we have not found any counterexamples yet.

We can also derive the so called phase process if we initiate the Markov chain every time we reach the absorbent state. (Phase processes occur e.g. at Quasi Birth-Death (QBD) Processes.) It is easy to see that the generator matrix of the phase process is the next: $G = A + a\pi$.

Since G have to be valid, we can repeat the procedure discussed above. We can also define the minimal residual polynomial for G: R^G_{min}, and we can rewrite theorem 3.5 for the phase process too:

Theorem 3.6 *If*

$$\forall i : R^G_{min\ i} \leq 0. \quad (3.6)$$

does not hold then there is no valid representation.

In this paper we examine only the relaxed problem except the last theorem, where π took part implicitly. An other exception occurs in Section 5, where the gradient methods use functions of the complete ME representations.

Finally we invoke an old result from Stefanita Mocanu [12] who also gives a spectral condition for a PH representation:

Theorem 3.7 *If there is such a complex conjugate pair of eigenvectors – $\lambda_{m_r} \pm \lambda_{m_i} i$ –, that fulfils:*

$$\left| \frac{\lambda_{m_r} - \lambda_{max}}{\lambda_{m_i}} \right| \leq \cot \frac{\pi}{n}, \quad (3.7)$$

then there is no similar valid transient generator matrix.

This theorem gives only one inequality unlike Theorem 3.5, but we have not performed a complete comparison yet.

4 Cyclic matrices

In this section we introduce the cyclic matrices because these matrices play an important role in the world of valid transient generator matrices.

Definition 4.1 *An A quadratic matrix is cyclic, if the A_{ij} values depend only from the $(j - i) \bmod n$ value. This kind of matrix can be characterized by the entries of the first row (c).*

E.g.:

$$\mathbf{A(c)} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{pmatrix}, \textit{ where } c = (1,2,3).$$

Definition 4.2 *Let us denote the next matrix with U:*

$$\mathbf{U} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_1 & \omega_2 & \dots & \omega_{n-1} \\ 1 & \omega_1^2 & \omega_2^2 & \dots & \omega_{n-1}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_1^{n-1} & \omega_2^{n-1} & \dots & \omega_{n-1}^{n-1} \end{pmatrix},$$

where $\omega_k = e^{\frac{2k\pi i}{n}}$.

$\frac{1}{\sqrt{n}}U$ is unitary and symmetric (the Discrete Fourier Transform), so $U^{-1} = \bar{U}/n$. (Here $\bar{U}$ is the complex conjugate of U.)

Definition 4.3 *If MAM^{-1} is diagonal then M is a left modal matrix of A.*

Each row of the left modal matrix is a left eigenvector of the corresponding eigenvalue in the diagonal form, so if we multiply these rows separately by arbitrary constants, the resulted matrix remains also a left modal matrix. This degree of freedom allows us to ensure the validity of the left modal matrix in general.

Theorem 4.1 [4]*(page 255.) U is a left modal matrix of any cyclic matrix.*

Unfortunately all the rowsums of U (left modal matrix) are zero except the first one, which is n, so we can not diagonalize cyclic matrices with valid similarity transformations. It is not good since the ordered diagonal form could be the intermediate step between the cyclic and the initial representations[1]. Moreover, it can be shown easily that cyclic matrices can generate only geometric distributions. Using Theorem 2.1 in this case, we also get that there is no valid transformation between general PH representations and cyclic forms. By all means, small perturbations can eliminate the zero rowsums of the left modal matrix. This elimination process has not been examined yet.

Theorem 4.2 [4]*(page 255.) If A is a cyclic matrix then there is a simple linear transformation between the λ eigenvalues and c:*

$$\lambda = cU, \qquad or \qquad c = \frac{\lambda \bar{U}}{n}. \tag{4.1}$$

In the next theorem we show that it is easy to find a real-valued cyclic form for a given spectrum. We just have to order the λ eigenvalues properly.

[1] here we have to remind the reader, that the set of valid transformations is a group

Theorem 4.3 *$c = \frac{\lambda \bar{U}}{n}$ is a real vector if and only if $\forall i : \bar{\lambda_i} = \lambda_{[(1-i) \mod n]+1}$.*

Proof 4.1 *It is easy to see that for every row/column of U there is a conjugate row/column too:*

$$\forall i : \bar{U}_{i,:} = U_{[(1-i) \mod n]+1,:},$$

(Here $\bar{U}_{i,:}$ denotes the i-th row of $\bar{U}$.) The first row/column is the conjugate of itself. If n is even then that is the case with the $\left(\frac{n}{2}+1\right)$-th row/column too. The complex components fall out in the expression of c_k:

$$(\lambda_i \bar{U}_{ik} + \lambda_{[(1-i) \mod n]+1} \bar{U}_{[(1-i) \mod n]+1,k}) =$$

$$(\lambda_i \bar{U}_{ik} + \bar{\lambda}_i U_{ik}) \in \Re.$$

The other direction is straightforward.

We know that c_1 is the average of the eigenvalues because $c = \frac{\lambda \bar{U}}{n}$. So the diagonal entries are negative. Moreover, the sum of any row in the cyclic form is λ_1 because $\lambda = cU$, which is also negative. Only the positiveness of the non-diagonal entries is not trivial. As we see the number of constraints of the relaxed problem reduced from n^2 to $n-1$ in the case of cyclic matrices. Since the number of appropriate orderings is huge, we can vary them to find a valid cyclic form. Usually there are more than two real-valued eigenvalues in the initial form of A unlike in cyclic matrices. In this case we can divide the eigenvalues into more groups and construct cyclic blocks separately.

Definition 4.4 *A matrix is called block-cyclic, if it is block-diagonal, and the blocks are cyclic.*

Now we can express our main conjecture regarding to cyclic matrices.

Conjecture 4.4 *If all the similar real-valued block-cyclic representations are not valid, then there is no similar valid transient generator matrix at all.*

This statement could be a necessary and sufficient condition for the existence of a valid similar transient generator matrix. We do not know any counter-example till now. Unfortunately the number of real-valued block-cyclic representations is huge but finite and we think, there could be efficient algorithms or at least good heuristics to find valid ones. This question have not been sufficiently examined yet.

Finally we have to mention that the 'best' diagonal polynomial in the previous section came from a setup where diagonal entries were equal. It remains us to the cyclic matrices. Moreover, the number of validity constraints of Theorem 3.5 is exactly the same as in the case of cyclic forms. However, in the first case constraints are necessary conditions whilst in the second case these are sufficient ones. If the constraints were equivalent, we would find the final answer to the relaxed problem by providing a necessary and sufficient condition.

The next section also contains a conjecture regarding to such cyclic matrices, where there is at most two real-valued eigenvalues. It will also say that cyclic forms are the 'best' but uses a bit different terminology.

5 Algorithms, goodness and error functions

If a ME distribution satisfies all the known necessary conditions above for the existence of a similar PH representation, we should try to find such a representation. In the followings we will use valid infinitesimal transformations iteratively using the gradient of different goodness/error functions and choose step sizes adaptively.

Definition 5.1 *B is a valid infinitesimal transformation if $B = I + \beta$ where $\beta\mathbb{1} = 0$ and the entries are so small that β^2 is negligible according to β. That is why $B^{-1} = I - \beta$.*

Definition 5.2 *Adaptive step size (regarding to any gradient method): The gradient is β in the transformation space. The transformation can be written so: $B = I + p\beta$, where p is the adaptive step size. If the last step have just taken and the previous step size was p' then the current value is $p = 1.2p'$. We transform the matrix using this p and derive the value of the goodness or error function and compare it with the original one. If the value gets better, we take the step and calculate the new gradient. Otherwise we halve p until the value gets better or at least remains the same then we also take the step.*

First of all we have to introduce the indicator matrix and the indicator vector.

Definition 5.3

$$\breve{\mathbf{A}} = \begin{pmatrix} a_1 & A_{12} & \dots & A_{1n} \\ A_{21} & a_2 & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & a_n \end{pmatrix}$$

$$\breve{\pi} = \pi = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}$$

The role of indicators is obvious. The corresponding (π, A) pair is valid if and only if all entries of the indicators are non negative. We have to emphasize that the diagonal entries of the indicator matrix are the rowsums of $-A$. We will define the goodness/error functions with the help of the indicators.

Definition 5.4 *The R goodness functions:*

$$R_A = \min \breve{A}, \qquad R_\pi = \min \breve{\pi}$$

$$R = \min(R_A,\ c_R R_\pi), \qquad \text{where } c_R \geq 0$$

Definition 5.5 *The E error functions:*

$$E_A(c) = \sum_{x \in \breve{A}} e^{-cx}, \qquad E_\pi(d) = \sum_{x \in \breve{\pi}} e^{-dx}$$

$$E = E_A(c) + c_E E_\pi(d), \quad \text{where} c_E \geq 0, \quad c,\ d > 0$$

c_E, c_R values are the binding coefficients. We usually set them as follows: $c_E = 1, c_R = n$.

Theorem 5.1 *$A\ (\pi, A)$ pair is PH representation if and only if $R \geq 0$.*

The R goodness functions are quite informative from the viewpoint of the PH representations unlike the E function, but R is not differentiable everywhere. That is why we use the E function in the gradient method. The next theorem allows us to enhance the value of R by using the gradients of E.

Theorem 5.2 *If the $E_1(c, d, c_E)$, $E_2(c, d, c_E)$, $R_1(c_R)$, $R_2(c_R)$ values are the goodness and error functions of two ME representations, then*

$$\forall c_E,\ c_R \geq 0 : \lim_{c,d \to \infty} \frac{E_1(c,d)}{E_2(c,d)} = \begin{cases} \infty & \text{if} R_2 > R_1 \\ \text{positive} & \text{if} R_2 = R_1 \\ 0 & \text{if} R_2 < R_1 \end{cases} \tag{5.1}$$

Theorem can be proved easily if we take the limit of the expression above. This means that if we want to minimize $E(c)$, we get similar results as if we would maximize R, if c is high enough.

We do not describe our algorithm in details because we could test it only on small problem instances where $n \leq 3$ (see the case studies below for details). The main idea is that we use grad E at first with small c, d values. The minimum is unique if c and d is small enough because in this case we can substitute the Taylor expansion of E with a positive definite quadratic expression. So we can reach this unique minimum from any similar representation. Then we increase the c, d values slowly while we perform some iteration steps until numerical difficulties arise: E diverges fast if R is negative. MATLAB sources are available here: [6].

Sometimes it is easier to calculate the gradient in the complete space of the infinitesimal transformations and then we derive the closest valid one. In this case the next theorem could be useful. It can be applied not only to infinitesimal transformations but to general ones too.

Theorem 5.3 *We look for such a B transformation matrix that is not far from C but valid. If we maximize the $\frac{\langle B|C\rangle}{\sqrt{\langle B|B\rangle\langle C|C\rangle}}$ value[2] beside the validity condition $B\mathbb{1} = \mathbb{1}$, then*

$$B = \frac{\mathbb{1}\mathbb{1}^T}{n} + xC\left(I - \frac{\mathbb{1}\mathbb{1}^T}{n}\right)$$

where $x = \frac{n}{\mathbf{1}^T C \mathbf{1}}$. If we minimize $\langle B - C|B - C\rangle$ beside the same condition then $x = 1$.

[2] $\langle X|Y\rangle := \sum_{ij} X_{ij} Y_{ij}$

Proof 5.1 *We show only the second case. At first we have to express the partial derivatives of the expression what we minimalize/maximalize in the space of valid transformations. I used the set of valid infinitesimal transformations which differ from the identity matrix only in one non-diagonal entry while validity is ensured by a diagonal correction. In this case* $\frac{\partial\langle B-C|B-C\rangle}{\partial b_{ij}} = 0$ *implies that* $\forall ij : 2(b_{ij} - c_{ij}) - 2(b_{ii} - c_{ii}) = 0$*. Now we have to merge these Eqs. into a matrix-equation as follows:*

$$C - \mathrm{diag}(C)\mathbb{I}^T = B - \mathrm{diag}(B)\mathbb{I}^T,$$

where diag() *gives back a column-vector including the appropriate diagonal entries of the argument. Let us multiply the equation with* $\mathbb{I}$ *on the right.*

$$C\mathbb{I} - n\mathrm{diag}(C) = \mathbb{I} - n\mathrm{diag}(B),$$

so we san write

$$\mathrm{diag}(B) = \frac{\mathbb{I} - C\mathbb{I}}{n} + \mathrm{diag}(C).$$

If we substitute it into the first matrix-equation, we get back the statement of the theorem.

Now we mention an other aspect of the cyclic matrices.

Theorem 5.4 *If a goodness/error function has the next form:* $F(A) = \sum_{x\in\check{A}} f(x)$ *where* f *is differentiable and* A *is cyclic then the gradient of* $F(A)$ *is zero (in the space of valid infinitesimal transformations).*

The proof is too long to detail here, but we have to use the same partial derivative of the infinitesimal transformations as in the previous proof. The steps are straightforward. The property of the cyclic matrices inspired the next conjecture:

Conjecture 5.5 *If* A *has similar real cyclic representations (consisting only one block) then the global maximum of* R_A *comes from one of these representations.*

Conjecture 4.4 can be considered as the generalization of this conjecture to the block-diagonal form. If there are more blocks then $R_A \leq 0$ of course.

6 Case studies

We will show the power of our theorems and methods on real data sets[5].

6.1 BC trace

The BellCore dataset (available at [5]) is a traffic data of a LAN Ethernet network recorded 1989 in the Bellcore Research Center in Morristown. The dataset was first analyzed at this reference: [1].

The empirical moments are the next:

	moments	8	35725624614.8032684326172
1	1.000000000000000000000	9	3414160455559.02880859375
2	4.22360967056936154052000	10	339725592277090.125000000
3	64.7632019524897799556100	11	34637644509814440.0000000
4	1862.57143119453121471452	12	3587161604074151936.00000
5	82474.2108922783372690901	13	375441643826279284736.000
6	5136400.25600292067974806	14	39589661071358423990272.0
7	401428800.155969798564911	15	4197680784515048876802048

$2n - 1$ moments can determine the corresponding ME function with order n, so maximal order is 8.

Now we fit ME functions where $n = 1, 2, 3, \ldots, 8$ with the method of Appie [3].

If $n = 8$ or $n = 7$, there is no such an ME function that has the same moments as above according to Appie's method. Otherwise this method gives the next representations.

$$\mathbf{A_6} = \begin{pmatrix} -1.069 & -0.137 & 0.1479 & 0.023 & -0.002 & 0.001 \\ -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.720 & -1.779 & 0.843 & 0.133 & -0.014 & 0.003 \\ -32.083 & 32.147 & -15.352 & -2.267 & 0.240 & -0.050 \\ 34.026 & -36.225 & 17.285 & 2.661 & -0.387 & 0.080 \\ 487.515 & -505.237 & 241.161 & 36.464 & -4.422 & 0.708 \end{pmatrix}$$

$$\mathbf{A_5} = \begin{pmatrix} -1.43750 & 0.24460 & -0.03413 & -0.00413 & 0.00081 \\ -1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ -1.38050 & 0.39352 & -0.19447 & -0.02351 & 0.00463 \\ 3.69680 & -4.85760 & 2.31210 & 0.40033 & -0.07890 \\ -23.76400 & 23.54400 & -11.24500 & -1.64660 & 0.12743 \end{pmatrix}$$

$$\mathbf{A_4} = \begin{pmatrix} -1.32920 & 0.13267 & 0.01929 & 0.00394 \\ -1.00000 & 0.00000 & 0.00000 & 0.00000 \\ -0.76393 & -0.24415 & 0.10991 & 0.02245 \\ -6.80400 & 6.00260 & -2.87190 & -0.38239 \end{pmatrix}$$

$$\mathbf{A_3} = \begin{pmatrix} -1.38800 & 0.19348 & -0.00973 \\ -1.00000 & 0.00000 & 0.00000 \\ -1.09890 & 0.10230 & -0.05546 \end{pmatrix}$$

$$\mathbf{A_2} = \begin{pmatrix} -1.37070 & 0.17553 \\ -1.00000 & 0.00000 \end{pmatrix}$$

$$\mathbf{A_1} = \begin{pmatrix} -1.00000 \end{pmatrix}$$

$$\pi_k = e_{(1)} = (1,\ 0,\ 0,\ \ldots)$$

Unfortunately if $n = 5$ then the ME function is not an ME distribution because $\lim_{t\to\infty} f(t) = -\infty$ (there is a positive eigenvalue) even though the moments formally fit. In the other cases the ME functions are distributions.

If $n = 6$ then PH representation is impossible if we apply theorem 3.5:

$$R_{min}(\lambda) = 0\lambda^6 + 0\lambda^5 + 0.02175\lambda^4 - 0.06599\lambda^3 -$$

$$-0.09069\lambda^2 - 0.02258\lambda - 0.00173$$

has a positive coefficient.

If $n = 4,\ 5,\ 6$ then PH representations are also impossible if we apply theorem 3.6, since the R^G_{min} polynomials have positive coefficients.

After running our algorithm if $n = 1, 2, 3$, we get PH representations only in cases $n = 1, 2$:

$$\pi_2' = (0.98957,\ 0.01043),$$

$$\mathbf{A_2'} = \begin{pmatrix} -1.22730 & 0.02085 \\ 0.02085 & -0.14337 \end{pmatrix},$$

If $n = 3$ then resulted local extrema has the next goodness functions: $R_{A_3'} = -0.00144670998799$, $R_{\pi_3'} = 0.00173310249387$.

Using only E-steps and adjusting c and d manually, we found the next PH representation:

$$\pi_3' = (0.0000154,\ 0.0109904,\ 0.9889941),$$

$$\mathbf{A_3'} = \begin{pmatrix} -0.051950 & 0.017028 & 0.017767 \\ 0.000230 & -0.152746 & 0.079693 \\ 0.000001 & 0.021915 & -1.238799 \end{pmatrix}.$$

To get this result, we increased c and d very slowly to avoid the extreme error values and the numerical instabilities. The result appeared at $c \approx 6000,\ d \approx 8000$.

6.2 DEC trace

DEC is an other traffic dataset (available here:[5]).

	moments		
1	1.000000000000000000000000000	9	1799517979420836.500000000000
2	3.204810514595462400680000000	10	450666369931268544.0000000000
3	28.062661397070076674250000000	11	112863712680956608512.0000000
4	2048.980478830916581500790000	12	28265296203065805766656.00000
5	460828.2335480888723395764800	13	70786877101982044253061 12.000
6	114632132.1903925687074661255	14	17727682523682461877975 12192
7	28693310339.73722076416015625	15	44396749874194860592798367744

If $n = 8$, there is no such an ME function that has the same moments as above according to Appie's method. If $n = 7, 6, 5, 4$ then the ME functions are not ME distributions. In other cases the method of Appie provides the next ME representations:

$$\mathbf{A_3} = \begin{pmatrix} -1.06871 & -0.13677 & 0.14791 \\ -1.00000 & 0.00000 & 0.00000 \\ 0.72033 & -1.77922 & 0.84266 \end{pmatrix},$$

$$\mathbf{A_2} = \begin{pmatrix} -1.06871 & -0.13677 \\ -1.00000 & 0.00000 \end{pmatrix},$$

$$\mathbf{A_1} = \begin{pmatrix} -1.00000 \end{pmatrix},$$

$$\pi_k = e_{(1)} = (1,\ 0,\ 0,\ \ldots).$$

After running our algorithm if $n = 1, 2, 3$, we get PH representations only in cases $n = 1, 2$:

$$\pi_2' = (0.97983,\ 0.020168),\quad \mathbf{A_2'} = \begin{pmatrix} -1.22337 & 0.03139 \\ 0.03139 & -0.23424 \end{pmatrix}$$

If $n = 3$ then resulted local extrema has the next goodness functions: $R_{\pi_3'} = 0.0009580680434057619$, $R_{A_3'} = -0.0009149475456419218$. Using only E-steps and adjusting c and d manually, we found the next PH representation:

$$\pi_3' = (0.290516,\ 0.000004,\ 0.709480),$$

$$\mathbf{A_3'} = \begin{pmatrix} -0.476016 & 0.000001 & 0.021968 \\ 0.006860 & -0.020668 & 0.006675 \\ 0.053149 & 0.000026 & -2.068253 \end{pmatrix}.$$

The method is the same as above but the parameters were a lot higher when the PH representation appeared: $c \approx 26000,\ d \approx 33000$.

7 Conclusions

We can say that our theorems are very efficient in proving the absence of any PH representation. We do not know such an A matrix that satisfies the conditions of our theorems but has no similar valid representations at the moment. Unfortunately in our case studies there were no PH distributions if n was high $(4, 5, 6, \ldots)$, so we could not test our algorithm on bigger problem instances. We do not know whether this problem comes from the particular datasets or from the fitting method. However, this negative result could be achieved hard without our spectral theorems.

We showed that cyclic matrices play an important role in the world of PH representations. We do not know any counterexample for the conjectures regarding to cyclic matrices till now. The conjectures could be used for a necessary and sufficient condition for the existence of similar valid representations of the transient generator matrix.

References

1 **Fowler H J, Leland W E**, *Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management*, IEEE Journal of Selected Areas in Communications **9** (1991), 1139–1149, DOI 10.1109/49.103559.

2 **Bini D A, Latouche G, Meini B**, *Numerical Methods for Structured Markov Chains*, Oxford University Press, 2005.

3 **van de Liefvoort A**, *The moment problem for continuous distributions*, Technical report WP-CM-1990-02, University of Missouri, Kansas City, USA, 1990.

4 **Rózsa P**, *Lineáris Algebra és alkalmazásai*, Műszaki könyvkiadó, 1974.

5 **Danzig P, Mogul J, Paxson V, Schwartz M**, *The Internet Traffic Archive*, available at `http://ita.ee.lbl.gov`.

6 **Demián T**, *Matlab sources*, available at `demiantamas.atw.hu/me2ph.zip`.

7 ———, *Markovi forgalmi modellek illesztése*, 2006, `http://demiantamas.atw.hu/diploma.pdf`. diploma thesis.

8 **Bux W, Herzog U**, *The phase concept: Approximation of measured data and performance analysis*, Computer Performance (Chandy K M, Reiser M, eds.), 1977, pp. 23–38.

9 **Lazowska Edward D., Addison Cliff**, *Selecting Parameter Values for Servers of the Phase Type*, Proceedings of the Third International Symposium on Modelling and Performance Evaluation of Computer Systems, ISBN:0-444-85332-4, 1979, pp. 407–420.

10 **Thummler A, Buchholz P, Telek M**, *A Novel Approach for PhaseType Fitting with the EM Algorithm*, IEEE Transactions on Dependable and Secure Computing 3, 2006, pp. 245–258, DOI 10.1109/TDSC.2006.27, (to appear in print).

11 **Lipsky L**, *Queuing theory – A linear algebraic approach*, Mcmillan, New York, 1992.

12 **Stefanita Mocanu**, *Construction et propriétés des représentations monocycliques des lois de type phase*, Laboratoire d'Automatique de Grenoble, France, 1999. Ph. D. dissertation.

13 **Horváth A, Rácz S, Telek M**, *Moments characterization of the class of order 3 matrix exponential distributions*, 16th Int. Conf. (ASMTA), Vol. 5513 of LNCS, Springer, Madrid, Spain, June 2009, pp. 174 –188.

14 **Neuts M F**, *Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach*, Dover Publications Inc., 1981.

15 **Latouche G, Ramaswami V**, *Introduction to Matrix Analitic Methods in Stochastic Modelling*, ASA SIAM, 1999.

16 **Bean N G, Fackrell M, Taylor P**, *Characterization of matrix-exponential distributions*, Stochastic Models **24** (2008), 339–363, DOI 10.1080/15326340802232186.

17 **Bladt M, Neuts M F**, *Matrix-exponential distributions: calculus and interpretations via flows*, Stochastic Models **19** (2003), 113–124, DOI 10.1081/STM-120018141.

18 **He Q M, Zhang H**, *On matrix exponential distributions*, Adv. in Appl. Probab. **39** (2007), no. 1, 271–292, DOI 0.1239/aap/1175266478.

19 **Stoyan Gisbert, Takó Galina**, *Numerikus Módszerek I.*, Typotex Kiadó, 2002.