

A TWO-PHASE MAP FITTING METHOD WITH APH INTERARRIVAL TIME DISTRIBUTION

András Mészáros
Miklós Telek

Budapest University of Technology and Economics
Magyar Tudósok Körútja 2.
1117 Budapest, HUNGARY

ABSTRACT

Markov arrival processes (MAPs) are used extensively in traffic modeling. Consequently a wide variety of fitting procedures have been developed. Most of these however are computationally demanding or not general enough. To resolve this problem, two-step procedures of a specific type have been made, which fit a phase-type distribution (PH) to static parameters in the first step, and extend it to a MAP in the second while fitting dynamic parameters. Their general weakness is that the first step often restricts the attainable range of dynamic parameters. In our paper we present a method, that aims at providing a good starting point for the second step, by optimizing the representation of the PH that was produced by the first step.

1 INTRODUCTION

Stochastic modeling of communication and computer systems is usually based on computationally tractable flexible analytical models. With these respects, Markov arrival processes (MAPs) are one of the most attractive candidates to describe traffic processes. They are known to approximate a wide range of processes from renewal ones to long range dependent ones (Andersen and Nielsen 1998; Horváth and Telek 2002a), and they allow the use of the computationally effective matrix analytic methods (Latouche and Ramaswami 1999).

Because of the structure of the standard description of MAPs, it is a reasonable approach, to apply a fitting procedure that is composed of two steps (Buchholz and Kriege 2009; Horváth, Telek, and Buchholz 2005). The first one generates a phase type distribution for capturing traits of the inter-arrival time distribution of the process (pdf, moments, etc.), while the second one approximates parameters that characterize dynamic behavior (joint moments, lag correlations, etc.).

A phase type distribution has infinite different representations. The chosen representation affects the obtainable range of the dynamic parameters of the second step, therefore it is important to find an appropriate PH representation in the first step. In this work we propose a procedure that can be effectively used to find such a representation.

The rest of this paper is organized as follows: Section 2 presents some basic properties of PH distributions and MAPs, and Section 3 the above mentioned fitting procedure. The proposed transformation method and its variants are discussed in Section 4. Section 5 provides numerical examples, comparing the different versions. Section 6 concludes the paper.

2 PHASE TYPE DISTRIBUTIONS AND MARKOV ARRIVAL PROCESSES

2.1 Phase Type Distributions

Let us take a continuous time Markov chain (CTMC) with n transient states and one absorbing state. The time to absorption from an initial state probability vector $\hat{\pi} = (\underline{\pi}, \pi_{n+1})$ defines a phase type distribution (PH). The D generator matrix of such CTMCs has a specific structure:

$$D = \begin{bmatrix} D_0 & \underline{d}_1 \\ \underline{0} & 0 \end{bmatrix}.$$

The row sums of a CTMC generator matrix are 0, which means that $D_0 \underline{1} = -\underline{d}_1$, where $\underline{1}$ is the column vector of ones. In addition, it is usually assumed, that $\pi_{n+1} = 0$ (i.e. the process cannot start in the absorbing state). Consequently, the PH is completely characterized by $\underline{\pi}$ and D_0 .

The distribution of the time to absorption is

$$P(X < t) = 1 - \underline{\pi} e^{D_0 t} \underline{1}, \quad (1)$$

the associated density function is

$$f(t) = -\underline{\pi} e^{D_0 t} D_0 \underline{1}, \quad (2)$$

its k th moment is

$$E(X^k) = k! \underline{\pi} (-D_0)^{-k} \underline{1} \quad (3)$$

and the arrival intensity is

$$\lambda = \frac{1}{E(X)} = \frac{1}{\underline{\pi} (-D_0)^{-1} \underline{1}}.$$

The same PH can be described with different $(\underline{\pi}, D_0)$ matrix-vector pairs, which we call different representations of the distribution. We can get a new representation of the PH using the following theorem:

Theorem 1 Let $\underline{\pi}$ be the initial probability vector and D_0 be the generator matrix of a PH. By applying the following similarity transformation:

$$D'_0 = T^{-1} D_0 T, \quad \underline{\pi}' = \underline{\pi} T,$$

where T is a non-singular matrix and $T \underline{1} = \underline{1}$, we will get a different representation of the same distribution.

Proof. The theorem can be proven using:

$$\begin{aligned} f(t) &= -\underline{\pi} e^{D_0 t} D_0 \underline{1} = -\underline{\pi} \sum_{i=0}^{\infty} \frac{t^i}{i!} D_0^i D_0 \underline{1} = -\underline{\pi} \sum_{i=0}^{\infty} \frac{t^i}{i!} D_0^i D_0 T \underline{1} = \\ &= -\underline{\pi} \sum_{i=0}^{\infty} \frac{t^i}{i!} D_0^i T D'_0 \underline{1} = -\underline{\pi} T \sum_{i=0}^{\infty} \frac{t^i}{i!} D_0'^i D'_0 \underline{1} = -\underline{\pi}' e^{D'_0 t} D'_0 \underline{1}. \end{aligned}$$

□

It is important to note, that the new representation is not necessarily Markovian. This means that $\underline{\pi}'$ and D'_0 defines the same density function according to (2), but $\underline{\pi}'$ might contain negative elements, and D'_0 might contain negative non-diagonal elements, which is not possible for the initial vector and the generator of a CTMC.

If D_0 is an upper (or lower) triangular matrix, we speak of an acyclic phase type distribution (APH). Every APH has infinitely many different representations, including the one called CF-1 canonical form (shortly canonical form hereafter), whose D_0 matrix has the following structure:

$$D_0 = \begin{bmatrix} -\lambda_1 & \lambda_1 & & & \\ & -\lambda_2 & \lambda_2 & & \\ & & \ddots & \ddots & \\ & & & -\lambda_{n-1} & \lambda_{n-1} \\ & & & & -\lambda_n \end{bmatrix},$$

where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. The transformation of an APH representation to the distribution's canonical form is shown in Cumani (1982), He and Zhang (2006).

2.2 Markov Arrival Processes

A Markov arrival process (MAP) is usually defined by two matrices, D_0 and D_1 , such that $D = D_0 + D_1$ is the generator of the background continuous time Markov chain (CTMC), D_0 contains the transitions of the background CTMC without arrival and D_1 describes the arrivals and the associated state transitions. The row sum of matrices D_0 and D_1 satisfies $D_0\mathbf{1} = -D_1\mathbf{1}$, since the row sum of the generator of a CTMC satisfies $D\mathbf{1} = 0$, and we have $0 = D\mathbf{1} = D_0\mathbf{1} + D_1\mathbf{1}$.

The steady state probability vector of the background CTMC, $\underline{\alpha}$, is the solution of the linear system $\underline{\alpha}D = 0$, $\underline{\alpha}\mathbf{1} = 1$.

In case of MAPs, the discrete time process embedded at arrival instants plays an important role. The state transition probability matrix of the embedded process is $P = (-D_0)^{-1}D_1$. The steady state probability vector of the embedded process, $\underline{\pi}$, is the solution of the linear system $\underline{\pi}P = \underline{\pi}$, $\underline{\pi}\mathbf{1} = 1$. The steady state distributions of the original and the embedded processes are related as

$$\underline{\alpha} = \frac{\underline{\pi}(-D_0)^{-1}}{\underline{\pi}(-D_0)^{-1}\mathbf{1}} = \lambda \underline{\pi}(-D_0)^{-1}.$$

In steady state, the inter-arrival time is phase type distributed with initial probability vector $\underline{\pi}$, and generator D_0 . Thus, the inter-arrival time distribution and the moments can be calculated using (1) and (3). In addition the ρ_k lag- k correlation can be computed, using

$$\rho_k = \frac{\lambda^2 \underline{\pi}(-D_0)^{-1} P^k (-D_0)^{-1} \mathbf{1} - 1}{2\lambda^2 \underline{\pi}(-D_0)^{-1} (-D_0)^{-1} \mathbf{1} - 1}, \quad (4)$$

and $\mu_{i,j}$, the order (i, j) lag- k joint moment using

$$\mu_{i,j} = E(X_n^i X_{n+k}^j) = i!j! \underline{\pi}(-D_0)^{-i} P^k (-D_0)^{-j} \mathbf{1}.$$

3 THE TWO-STEP MAP FITTING PROCEDURE

As the previous section showed, the static characteristics of a MAP only depend on $\underline{\pi}$ and D_0 , but not on D_1 . This trait is the base of the two-step methods (Buchholz and Kriege 2009; Horváth, Telek, and Buchholz 2005), which have the following general structure:

- In the first step, the inter-arrival time distribution is fitted by a phase type distribution, which determines the D_0 matrix (the generator of the PH distribution) and the $\underline{\pi}$ vector (the initial probability vector of the PH distribution)
- Then the D_1 matrix is constructed, using a method that fits to dynamic parameters, while leaving the inter-arrival time distribution (i.e. $\underline{\pi}$ and D_0) of the resulting MAP unchanged.

3.1 Constructing $\underline{\pi}$ and D_0

The first step of the procedure is a phase type fitting problem for which we refer to (Asmussen and Nerman 1991; Bobbio and Cumani 1992; Horváth and Telek 2002b; El Abdouni Khayari, Sadre, and Haverkort 2001; Thümmel, Buchholz, and Telek 2005; Buchholz and Kriege 2009). Here we only recall that the various PH fitting methods can handle different input data. The inter-arrival time distribution of the original process can be given with its pdf or cdf, samples or by a given number of moments. The methods in (Asmussen and Nerman 1991; El Abdouni Khayari, Sadre, and Haverkort 2001; Thümmel, Buchholz, and Telek 2005) fit a phase type distribution to a set of samples. The methods in (Bobbio and Cumani 1992; Horváth and Telek 2002b) allow to fit to both, pdf/cdf and a set of samples. Exact phase type fitting of moments is available up to 3 moments (which is used in (Heindl 2004)). A heuristic approach is used to fit moments in (Buchholz and Kriege 2009).

All fitting methods in (Bobbio and Cumani 1992; Horváth and Telek 2002b; El Abdouni Khayari, Sadre, and Haverkort 2001; Thümmel, Buchholz, and Telek 2005); and (Buchholz and Kriege 2009) provide acyclic phase type distributions.

3.2 Constructing D_1

The second phase of the procedure is the fitting of D_1 to dynamic parameters. This is a non-linear optimization with linear constraints. The constraints are:

$$\text{C1: } D_{1i,j} \geq 0, \forall 1 \leq i, j \leq n,$$

$$\text{C2: } D_1 \mathbf{1} = -D_0 \mathbf{1},$$

$$\text{C3: } \pi(-D_0)^{-1} D_1 = \pi,$$

where n is the number of states of the background CTMC.

The first two are general requirements for any MAP, while the last one ensures that the inter-arrival time distribution remains the same as it was after the first step.

The goal function of the optimization is typically a weighted sum function of the normalized and squared differences of the MAP's and the original process' dynamic parameters.

To be more specific, in Buchholz and Kriege (2009), the goal function is

$$\theta = \sum_{(i,j) \in \mathcal{J}} \beta_{i,j} \left(1 - \frac{\mu_{i,j}}{v_{i,j}}\right)^2, \quad (5)$$

where $\mu_{i,j}$ is the MAP's, and $v_{i,j}$ is original process' order (i, j) lag-1 joint moment, $\beta_{i,j}$ is the corresponding weight and \mathcal{J} is the set of (i, j) pairs for which the lag-1 joint moments have to be fitted. This expression results in a least-squares problem with linear constraints.

In Horváth, Telek, and Buchholz (2005), the goal function is $\sum_{k=2}^K w_k (\rho_k - \hat{\rho}_k)^2$, where ρ_k and $\hat{\rho}_k$ are the lag- k correlation for the MAP and the original process respectively, and w_k is the corresponding weight.

4 OPTIMIZING THE REPRESENTATION OF THE PH DISTRIBUTION

The introduced fitting method has an important issue. The representation of the resulting PH distribution in the first step greatly affects the range of achievable lag correlation values. In the case of the canonical form for example (or any other representation, with one entry or exit state), the lag- k correlation will be zero, irrespectively of the actual D_1 . This makes clear, that finding a proper representation is crucial.

4.1 The Representation Optimization Method of Buchholz and Kriege

In (Buchholz and Kriege 2009) a representation optimization method is introduced, which has the canonical form of an APH as input (the procedure can be used for any APH representation), and a representation with all its states being exit states as output.

Representation Optimization Method 1

Execute the following steps for every (i, j) pair of states, where $i < j$, in lexicographical order.

1. Define $\delta \leq \delta_{max}$, where

$$\delta_{max} = \min \left(\pi(j), \frac{\pi(i) D_0(i, j)}{\lambda_j - \lambda_i}, \min_{k < i, D_0(k, i) > 0} \left(\pi(i) \frac{D_0(k, j)}{D_0(k, i)} \right) \right) \quad (6)$$

If $\lambda_i = \lambda_j$, then the second term is considered to be infinite. If $\delta_{max} > 0$, then δ can be chosen freely between δ_{max} and 0.

2. Compute

$$\pi'(k) = \begin{cases} \pi(i) + \delta, & \text{for } k = i, \\ \pi(j) - \delta, & \text{for } k = j, \\ \pi(k), & \text{otherwise,} \end{cases} \quad (7)$$

and

$$D_0(k, l) = \begin{cases} D_0(i, j) \frac{\pi(i)}{\pi(i) + \delta} - \frac{(\lambda_j - \lambda_i)\delta}{\pi(i) + \delta}, & \text{for } k = i \text{ and } l = j, \\ D_0(i, l) \frac{\pi(i)}{\pi(i) + \delta} + D_0(j, l) \frac{\delta}{\pi(i) + \delta}, & \text{for } k = i \text{ and } l \neq j, \\ D_0(k, i) \frac{\pi(i) + \delta}{\pi(i)}, & \text{for } k < j \text{ and } l = i, \\ D_0(k, j) + D_0(k, i) \frac{\delta}{\pi(i)}, & \text{for } k < j \text{ and } l = j, \\ D_0(k, l), & \text{otherwise.} \end{cases} \quad (8)$$

4.2 The Proposed Representation Optimization Method

The above described method is computationally advantageous, however it has some shortcomings:

- The connection between the attainable range of dynamic parameters and the method is not obvious, consequently the resulting representation might be even worse than the original one (if that is not the canonical form).
- The output representation depends greatly on the actual value of δ , for which $\delta = 0.9 * \delta_{max}$ is proposed by the authors, however, this proposal is based solely on experimental results.
- The input and the output of the procedure have structural constraints as they can only be APHs.
- The elements of the π_{out} output probability vector (which contains the probabilities that state i is the last state before absorption) are barely changing, which can be a limiting factor.

The procedure introduced in this section addresses these problems, and offers a more flexible alternative. To this end, we introduce matrix $T_{i,j}(x)$ ($i \neq j$) of size $n \times n$ whose k, l element is

$$[T_{i,j}(x)]_{k,l} = \begin{cases} 1 - x, & \text{if } k = l = i, \\ 1, & \text{if } k = l \neq i, \\ x, & \text{if } k = i, l = j, \\ 0, & \text{otherwise.} \end{cases}$$

Indeed only two elements of matrix $T_{i,j}(x)$ differ from the identity matrix: element i, i and i, j . The core of this method is Theorem 1 and it is composed of the following steps.

Representation Optimization Method 2

For a given goal function \mathcal{G} (a few options will be discussed later)

1. Set a to 0.5
2. Maximize $\mathcal{G}(\text{PH}(\pi T_{i,j}(x), T_{i,j}^{-1}(x) D_0 T_{i,j}(x)))$,
for $i, j \in \{1, \dots, n\}$, $i \neq j$ and $x \in \{a, -a\}$,
such that $\pi T_{i,j}(x)$ and $T_{i,j}^{-1}(x) D_0 T_{i,j}(x)$ are Markovian.
3. If there is no such representation, or its goal function value is lower than the original representation's, then halve a ,
else set

$$\pi = \pi T_{i^*, j^*}(x^*) \quad \text{and} \quad D_0 = T_{i^*, j^*}^{-1}(x^*) D_0 T_{i^*, j^*}(x^*),$$

where i^*, j^*, x^* are the parameters of the best new representation.

4. Stop if a gets smaller than a predefined a_{min} , or the cycle limit is reached,
otherwise go to step 2.

We examined two major types of goal functions for the representation optimization. The first group consists of various functions of the entropy of the input and output probability vectors. Former is $H_{in} = \sum_i \pi(i) \log(\pi(i))$, the latter is $H_{out} = \sum_i \pi_{out}(i) \log(\pi_{out}(i))$, where π_{out} is the previously defined output probability vector which can be calculated using $\pi_{out} = (\pi D_0^{-1}) \cdot (-D_0 \mathbf{1})^T$, where \cdot denotes elementwise multiplication operation of the vectors. We experimented with the sum and product of the input and output vectors' entropies. Higher entropy means better spread of probability in the probability vector. If the probability concentrates into one state, it results in a PH with one significant initial or exit state, leading to poor dynamic behavior. (Close to zero lag-1 autocorrelation for example.)

The second group of goal functions we examined for the representation optimization is related to the achievable range of lag-1 autocorrelation (ρ_1). We inspected the maximal and minimal ρ_1 that can be attained from a representation, as well as their difference as possible goal functions. From (4) it is clear, that ρ_1 is a linear function of the elements of D_1 . As such, constraint C2 and C3 (as constraints) and (4) (as the base of the goal function) form a linear programming problem with n^2 variables and $2n$ constraints, where n is the order of the MAP, and can be solved with little cost.

Other goal functions can be used as well, the main issue to be considered is that they have to be calculated many times. (The number of calculations increases with the square of the order of the MAP in the third step, furthermore the number of iterations grows too.)

The introduced algorithm has a crucial issue. The initial representation has a significant effect on the result of the optimization procedure. This is due to two main reasons. The first one is that the method uses a greedy algorithm, thus the optimization will get stuck at local optimums. The second one is that the algorithm might not find any other valid PH representation at all. This is usually the case when one tries to optimize a canonical representation, which is the result of many PH fitting algorithms.

Taking into account the flaws of both the method of Buchholz and Kriege and the just introduced, transformation matrix based method, it is a reasonable approach to combine them: we use the former one, to leave the canonical representation, then the latter one to eliminate the weaknesses of the Buchholz-Kriege method.

5 NUMERICAL EXPERIMENTS

In this section we will demonstrate and discuss a few specialities of the method presented in the previous chapter through a number of examples. We will also compare it with the method of Buchholz and Kriege and make a few remarks regarding that procedure.

Hereinafter we will use the notation $\langle group \rangle_{\langle type \rangle}^{\langle method \rangle}$ in the figures and in the text, where $\langle group \rangle$ shows the group of the goal function, and can be ρ for lag-1 correlation and θ for the goal function in (5), $\langle type \rangle$ selects the specific goal function of the group, and $\langle method \rangle$ specifies the optimization method.

The possible categories for type are:

- *min* and *max* - the minimal and maximal achievable correlation
- *range* - the distance between the minimal and maximal achievable correlations

The possible categories for method are:

- *Buchh* - result of Buchholz-Kriege method
- *sum, prod* - result of the entropy sum and product based optimization method
- *min, max, range* - the first two are for the result of the minimal correlation, maximal correlation based optimization method, the third is for the sum absolute value of the maximal and minimal achievable correlation

For example $\rho_{\max}^{(sum)}$ denotes the maximal achievable correlation by the entropy sum based representation optimization method. In other words the maximal achievable lag-1 autocorrelation after using $H_{(sum)} = H_{(in)} + H_{(out)} = \sum_i \pi(i) + \sum_i \pi_{out}(i)$ as goal function to be maximized in the representation optimization.

In the examples we used PhFit to fit a PH distribution to data samples. The tool was introduced in Horváth and Telek (2002b) and has many different features. What is important from our work's point of view is that it fits a PH of given order in canonical representation.

In the second step of the MAP fitting we used (5) as a goal function i.e. we fitted to a weighted sum function using the normalized and squared differences of the lag-1 joint moments of the MAP and the original process. Our main reason for this choice is that it results in a least squares optimization, while in the case of the other presented option a polynomial optimization problem has to be solved.

Our decision does not conflict with the usage of correlation based goal functions in the representation optimization, as, having a fixed inter-arrival time distribution, the higher flexibility of ρ_1 is equivalent to the higher flexibility of the $\mu_{1,1}$ joint moment, as can be easily seen from the definition of the two measures.

It is also worth mentioning that, while reaching $\theta = 0$ would mean errorless fitting to the chosen lag-1 joint moments, the value set might be out of the bounds of what can be fitted with MAPs of the chosen order, therefore perfect fitting can be theoretically impossible in many cases.

5.1 Fitting MAP(5) to MAP(5)

In our first example we approximated an order 5 MAP with an other order 5 one through its inter-arrival time distribution and dynamic parameters. The original MAP is defined by

$$D_0 = \begin{bmatrix} -8 & 7.6 & 0 & 0 & 0 \\ 0 & -9 & 6.7 & 0 & 0 \\ 0 & 0 & -12.9 & 9.5 & 0 \\ 0 & 0 & 0 & -16.3 & 10.6 \\ 0 & 0 & 0 & 0.3 & -11.7 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0.1 & 0 & 0.3 \\ 0 & 1.2 & 1.1 & 0 & 0 \\ 0 & 3.4 & 0 & 0 & 0 \\ 2.7 & 3 & 0 & 0 & 0 \\ 11.4 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

First we generated a trace with 100,000 arrivals and fitted an order 5 PH to it. The fitting resulted in $\pi = [0.52298, 0.132574, 0.153669, 0.102099, 0.088678]$, and

$$D_0 = \begin{bmatrix} -6.20831 & 6.20831 & 0 & 0 & 0 \\ 0 & -14.1202 & 14.1202 & 0 & 0 \\ 0 & 0 & -17.9148 & 17.9148 & 0 \\ 0 & 0 & 0 & -18.5773 & 18.5773 \\ 0 & 0 & 0 & 0 & -19.6338 \end{bmatrix}.$$

Before continuing with the joint moment fitting, let us examine how the achievable correlations change with $\delta_{rate} = \delta/\delta_{\max}$. In Figure 1 the results of the entropy based and the correlation based optimizations can be compared to that of the Buchholz-Kriege method. The correlation of the trace is $\rho_1 = 0.1197$. Perhaps the most striking observation is that the Buchholz-Kriege method reaches its maximum around $\delta_{rate} = 0.63$, with $\rho_{\max}^{(Buchh)} = 0.106$ and by $\delta_{rate} = 0.9$ it decreases to 0.02. This behavior can be explained by taking a look at the initial probability vector of the PH (i.e. the embedded probability vector of the MAP) in the $\delta_{rate} = 0.9$ point: $\pi = [0.943, 0.021, 0.025, 0.007, 0.004]$. The probabilities of the latter states get near zero, in other words the representation gets close to one with a single entry state, which cannot be complemented to a MAP that has any dynamic behavior. This drift to the initial states is a general characteristic of the method, as can be easily seen from (7). (If δ_{\max} is small, or the original values of the initial probabilities are high enough, they might remain sufficiently high after using the Buchholz-Kriege method even for bigger δ_{rate} values, in which case this is not an issue.)

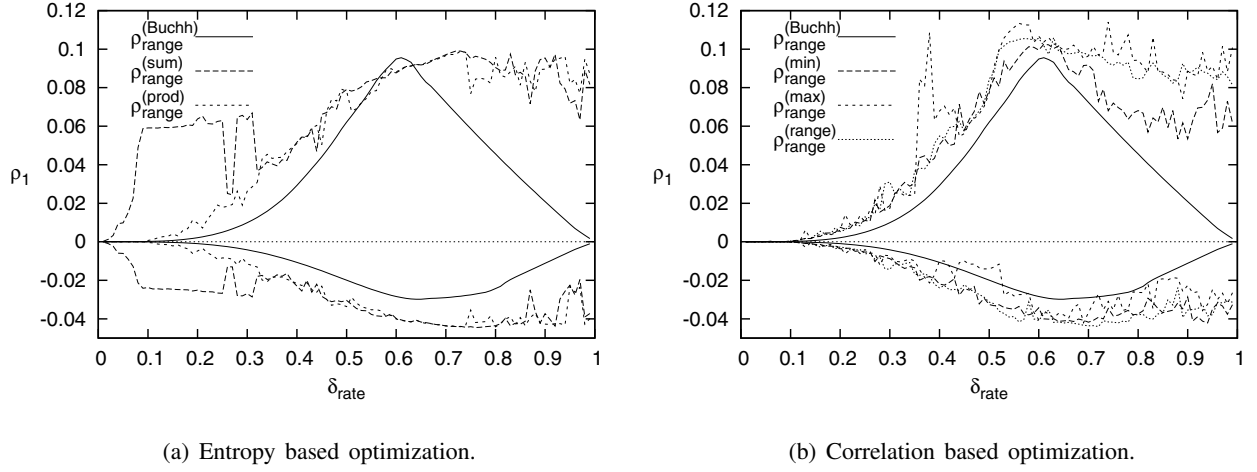


Figure 1: Example 1: Range of achievable correlations after optimizing the representation of the PH.

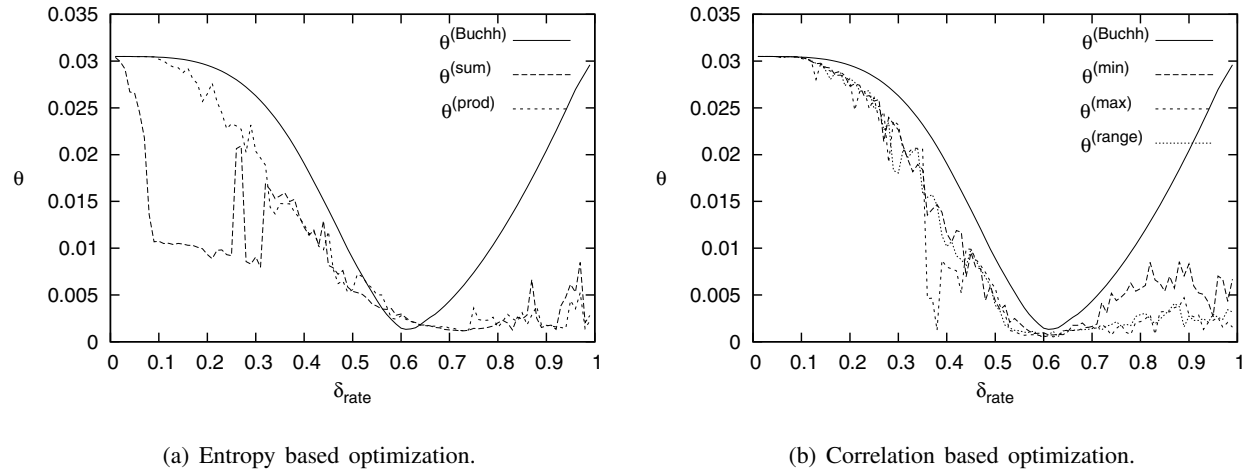


Figure 2: Example 1: Result of the joint moment fitting.

The entropy based optimizations address exactly this problem. In Figure 1(a) we can see that the usage of the two goal functions have similar outcome for a large scale of δ_{rate} . The differences at smaller δ_{rate} values are mainly due to the small local optimums in which the entropy product based optimization gets stuck, and not because of the significant difference in the goal functions themselves. In general, when the maximal correlations are similar for the two cases, the optimized representations are similar as well.

The situation is a bit more complicated in the correlation based methods' case. The representations might differ considerably despite the close correlation values. This seems to suggest that the representation could not be improved further significantly. This assumption is also supported by the fact that the achievable correlation does not change substantially for larger values of δ_{rate} . In accordance with the expectations, the representation with maximal attainable correlation ($\rho_1 = 0.114$) can be obtained using the maximal correlation as goal function.

Now let us move on to the construction of D_1 . We made the fitting based on the $v_{i,j}$, $1 \leq i, j \leq 4$ joint moments of the trace, with $\beta_{i,j} = 0.25^{i+j-2}$. The results can be seen in Figure 2. By comparing Figure 1 and 2, it is obvious, that the θ goal function shows close connection with the ρ_1 values. This is partly

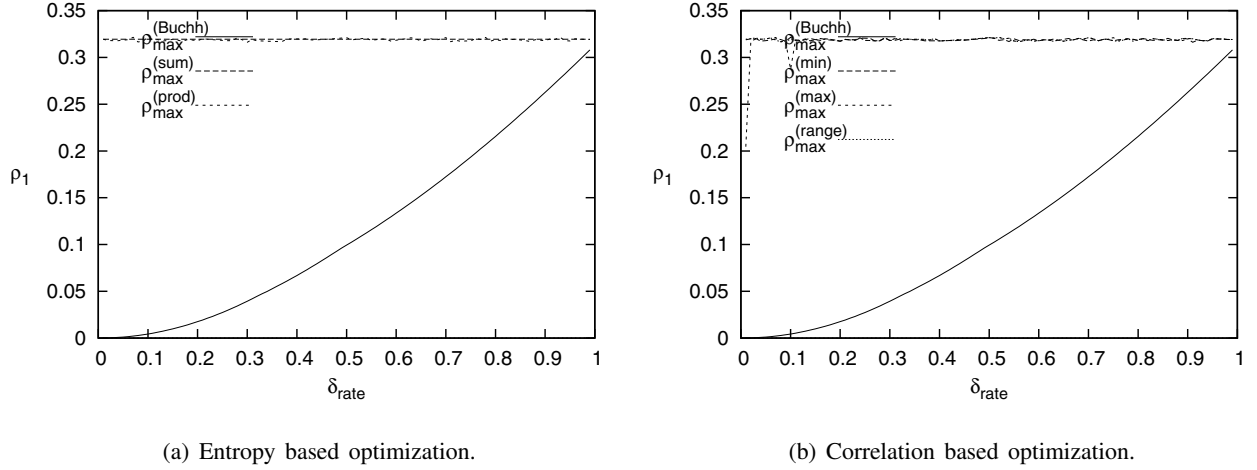


Figure 3: Example 2: Range of achievable correlations after optimizing the representation of the PH.

because we approximated a MAP with a MAP of the same order, but the other reason is that the most important $\mu_{1,1}$ joint moment is in direct connection with ρ_1 .

5.2 Fitting a Real Trace

In these examples we fitted to real data, the first 100,000 arrivals of the pAug89 traffic trace (The Internet Traffic Archive 1989), which contains arrival times of Ethernet packages. The correlation of the data is $\rho_1 = 0.167$.

First we fitted with an order 3 MAP. The PH fitting set resulted in

$$\underline{\pi} = [0.0382422, 0.70902, 0.252738] \text{ and } D_0 = \begin{bmatrix} -89.7068 & 89.7068 & 0 \\ 0 & -863.648 & 863.648 \\ 0 & 0 & -881.203 \end{bmatrix},$$

Figure 3 shows the maximum achievable correlations after the different optimizations of this representation.

We can see that the different optimizations give almost identical results for the whole range of δ_{rate} (except for the Buchholz-Kriege method). Furthermore, the actual representations are very similar for all of the cases. This suggests that the PH distribution is inherently flexible. Also it affirms that the entropy based optimization can give good results in many cases, while having lower computational cost than the correlation based ones.

Figure 4 shows the outcome of the joint moment fitting. The curves are in accordance with our earlier observations: the higher the maximal correlation is, the better the fitting can be. In this case, however, the maximal achievable correlation is significantly bigger, than what the trace has. The reason why they are still that closely connected is that the variance of the original trace is higher than the MAP's. ($\sigma_{trace}^2 = 1.95 \times 10^{-5}$, $\sigma_{MAP}^2 = 1.21 \times 10^{-5}$) The Buchholz-Kriege method gives approximately equal result ($\theta = 0.22$) to the other methods at its optimum, which is just below $\delta_{rate} = 1$, while at $\delta_{rate} = 0.9$, $\theta = 0.34$.

In our third example we fitted the pAug89 trace with an order 5 MAP. The PH fitting resulted in $\underline{\pi} = [0.042174, 0.359523, 0.454311, 0.000272131, 0.14372]$ and

$$D_0 = \begin{bmatrix} -96.5378 & 96.5378 & 0 & 0 & 0 \\ 0 & -1036.68 & 1036.68 & 0 & 0 \\ 0 & 0 & -1257.18 & 1257.18 & 0 \\ 0 & 0 & 0 & -1474.71 & 1474.71 \\ 0 & 0 & 0 & 0.3 & -2951.7 \end{bmatrix}.$$

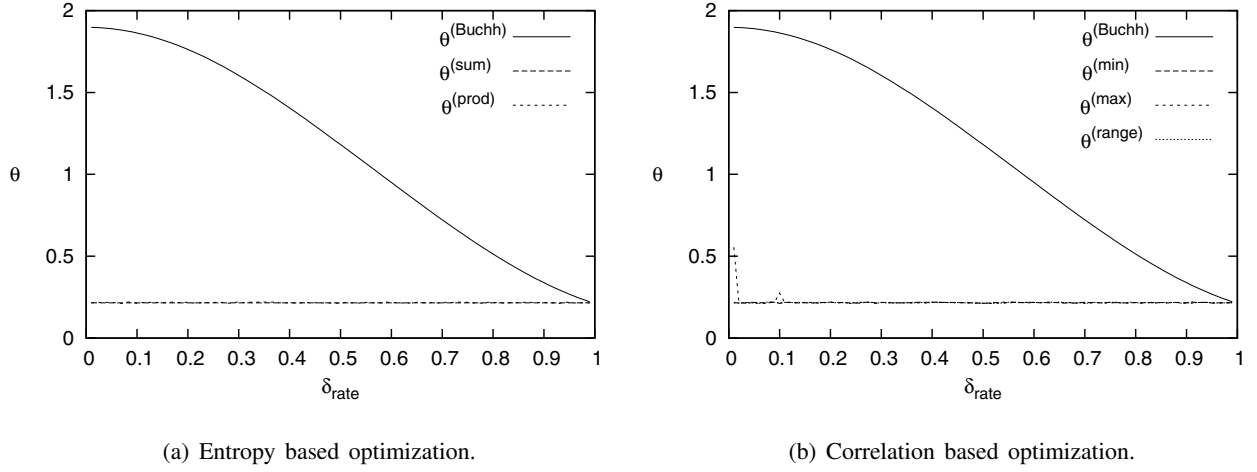


Figure 4: Example 2: Result of the joint moment fitting.

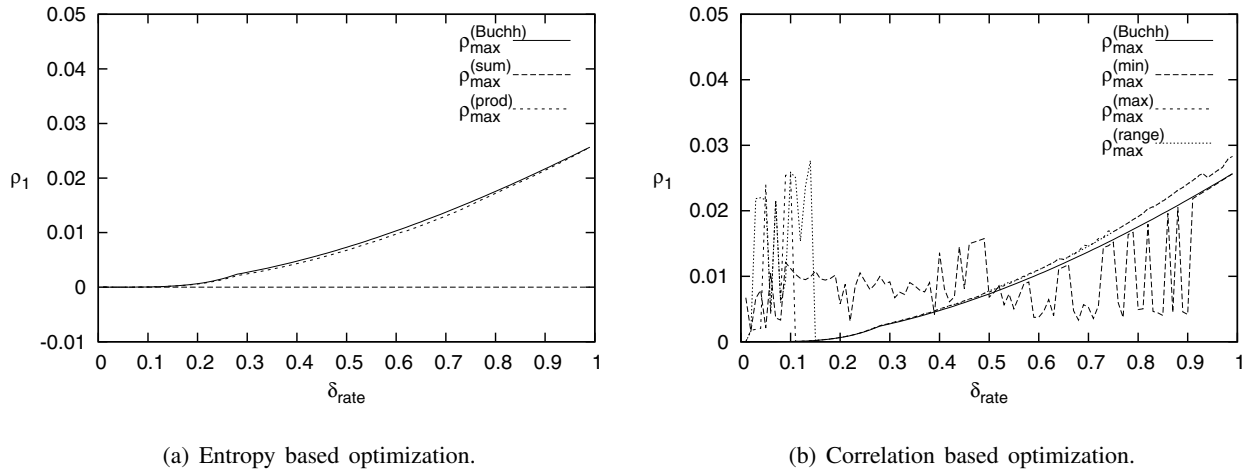


Figure 5: Example 3: Range of achievable correlations after optimizing the representation of the PH.

The maximum achievable correlations of the different optimizations can be seen in Figure 5. The results show that the applied methods could not improve the representation significantly. The entropy sum based optimization even worsens the attainable correlation, which decreases to zero regardless of the value of δ_{rate} . This reveals that the entropy based goal functions are not completely reliable.

The results suggest that the PH itself is inflexible. The Buchholz-Kriege method can barely move the representation from the canonical form (this is because the δ_{max} value is far smaller than in previous cases), and the first and fourth elements of $\underline{\pi}$ reach zero rapidly when the optimizations increase the maximal achievable correlation.

The results of the joint moment fitting are shown in Figure 6, and are in accordance with the expectations based on the previous examples. The θ goal function's value is in close connection with the maximal achievable correlation.

The second and third example show, that the fitting does not necessarily improve with the increase of the order of the MAP, on the contrary, our experiments show that higher order PHs are often less flexible than smaller ones.

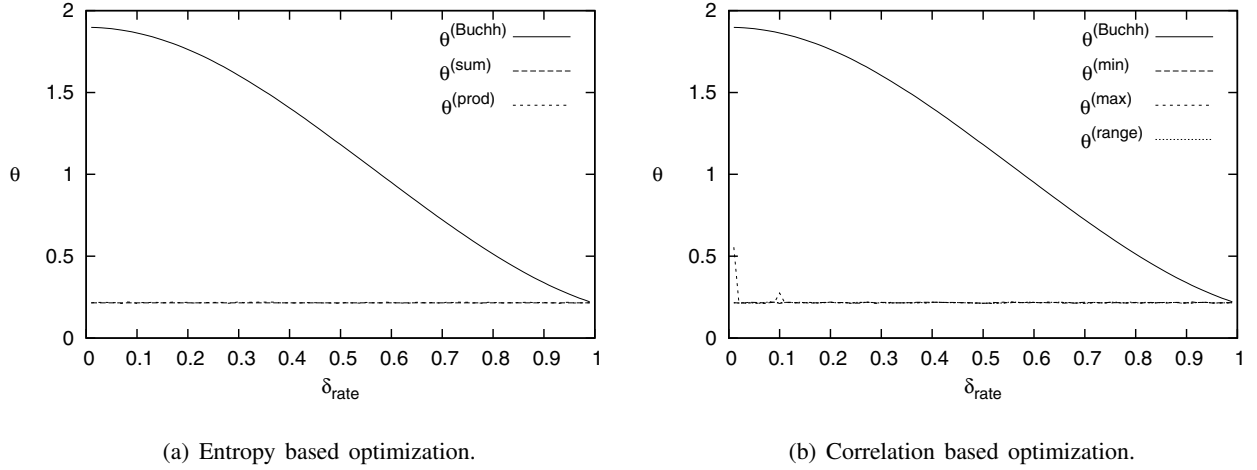


Figure 6: Example 3: Result of the joint moment fitting.

6 CONCLUSIONS

In this paper we presented a new PH transformation algorithm, which can improve the existing two-step MAP fitting methods. The algorithm optimizes the representation of the PH obtained in the first step, this way produces a better representation for the second step.

The numerical examples show that generally the new transformation algorithm gives better results than the method of Buchholz and Kriege. We have to note however, that the correlation based optimization is relatively slow (and its run time increases rapidly with the order of the PH), although the algorithm's speed can be controlled by changing the a_{min} value in the stopping condition, while the entropy based optimization gives worse results than the Buchholz-Kriege method in some cases. The experiments also affirm that the achievable lag-1 autocorrelation can be effectively used to judge the flexibility of a representation in general.

It is also worth pointing out, that the $\delta_{rate} = 0.9$ choice is generally suboptimal for the Buchholz-Kriege method, thus it may be beneficial to check the results for multiple δ_{rate} values.

The last example also made clear that the representation optimization has its limitations, as the structure of the PH may narrow down the possible range of dynamic behavior significantly. This is a serious flaw of the two step optimization procedures in general. It is among our future goals to develop an optimization method that eliminates this weakness.

REFERENCES

- Andersen, A. T., and B. F. Nielsen. 1998. "A Markovian approach for Modeling Packet Traffic with Long-Range Dependence". *IEEE Journal on Selected Areas in Communications* 16 (5): 719–732.
- Asmussen, S., and O. Nerman. 1991. "Fitting Phase-type distributions via the EM algorithm". In *Proceedings: "Symposium i Advent Statistik"*, 335–346. Copenhagen.
- Bobbio, A., and A. Cumani. 1992. "ML estimation of the parameters of a PH distribution in triangular canonical form". In *Computer Performance Evaluation*, edited by G. Balbo and G. Serazzi, 33–46: Elsevier Science Publishers.
- Buchholz, P., and J. Kriege. 2009. "A Heuristic Approach for Fitting MAPs to Moments and Joint Moments". In *Proc. QEST'09*, 53–64. Budapest, Hungary: CPS press.
- Cumani, A. 1982. "On the canonical representation of homogeneous Markov processes modelling failure-time distributions". *Microelectronics and Reliability* 22:583–602.

- El Abdouni Khayari, R., R. Sadre, and B. Haverkort. 2001. "Fitting world-wide web request traces with the EM-algorithm". In *Proc. of SPIE*, Volume 4523, 211–220. Denver, USA.
- He, Q.-M., and H. Zhang. 2006. "Spectral polynomial algorithms for computing bi-diagonal representations for matrix-exponential distributions and phase-type distributions". *Stochastic models* 22:289–317.
- Heindl, A. 2004. "Inverse characterization of hyperexponential MAP(2)s". In *11th Int. Conf. on Analytical and Stochastic Modelling Techniques and Applications (ASMTA)*, 183–189. Magdeburg, Germany.
- Horváth, A., and M. Telek. 2002a, July. "A Markovian point process exhibiting multifractal behaviour and its application to traffic modeling". In *Matrix analytic methods: Theory and applications, MAM4*, edited by G. Latouche and P. Taylor, 183–208. World Scientific.
- Horváth, A., and M. Telek. 2002b, April. "PhFit: A general purpose Phase type fitting tool". In *Tools 2002*, 82–91. London, England: Springer, LNCS 2324.
- Horváth, G., M. Telek, and P. Buchholz. 2005. "A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation". In *Proc. QEST'09*, 124–133: IEEE CS.
- Latouche, G., and V. Ramaswami. 1999. *Introduction to matrix analytic methods in stochastic modeling*. SIAM.
- The Internet Traffic Archive 1989. "BC-pAug89 traffic trace". <http://ita.ee.lbl.gov/index.html>.
- Thümmler, A., P. Buchholz, and M. Telek. 2005. "A Novel Approach for Fitting Probability Distributions to Trace Data with the EM Algorithm". In *Int. Conf. on Dependable Systems and Networks (DSN)*. Yokohama, Japan: IEEE CS.

AUTHOR BIOGRAPHIES

ANDRÁS MÉSZÁROS is a Ph.D. student at the Budapest University of Technology and Economics. His research interests are in stochastic modeling.

MIKLÓS TELEK received the M. Sc. degree in Electrical Engineering from the Technical University of Budapest in 1987. In the same year he joined the Hungarian Post Research Institute where he studied the modeling, analysis and planning aspects of communication networks. Since 1990 he has been with the Department of Telecommunications of the Technical University of Budapest, where he is a full professor now. He received the C.Sc. and D.Sc. degree from the Hungarian Academy of Science in 1995 and 2004, respectively. His current research interest includes stochastic performance modeling and analysis of computer and communication systems.