A Heuristic Approach for Fitting MAPs to Moments and Joint Moments

Peter Buchholz and Jan Kriege
Dep. of Computer Science, TU Dortmund, D-44221 Dortmund, Germany
{peter.buchholz, jan.kriege}@tu-dortmund.de

Author prepared version of a paper published in

Proc. of the 6th International Conference on Quantitative Evaluation of SysTems (QEST 2009), IEEE Computer Society, 2009.

http://doi.ieeecomputersociety.org/10.1109/QEST.2009.36

A Heuristic Approach for Fitting MAPs to Moments and Joint Moments

Peter Buchholz and Jan Kriege
Dep. of Computer Science, TU Dortmund, D-44221 Dortmund, Germany
{peter.buchholz, jan.kriege}@tu-dortmund.de

Abstract

Fitting of the parameters of a Phase Type (PH) Distribution or a Markovian Arrival Process (MAP) according to some quantities of measured data streams is still a challenge. This paper presents a new approach which computes in two steps for a set of moments and joint moments for an Acyclic PH distribution that is expanded into a MAP.

In contrast to other known approaches, parameters are computed to minimize the weighted squared difference between the measured moments and the moments of the resulting PH Distribution or MAP. The proposed approach is very flexible and allows one to generate a MAP of a predefined order to approximate a given set of moments and joint moments. It is shown that the approximation is often sufficiently accurate even with MAPs of a moderate size. However, we also show that the practical applicability of the approach is limited since the exact determination of higher order moments from traces requires an extremely high effort.

Keywords: Acyclic Phase Type Distributions, Markovian Arrival Processes, Moments, Fitting Procedures

1 Introduction

To apply numerical or analytical solution approaches for the analysis of queueing networks or other stochastic models, the models usually have to describe a Markov chain which implies that distributions have to be modeled as phase type distributions (PH distributions) or Markovian arrival processes (MAPs). PH distributions and MAPs have a long history in stochastic modeling [19, 20] and are powerful model classes which in theory allow the representation of almost all relevant stochastic behaviors that are observed. However, the building of a PH distribution or a MAP according to a set of characteristics of stochastic behavior that has been observed in practice is still an open and hard problem. In particular, one is often interested in PH distributions or MAPs with a small number of phases that mimic the behavior of some real system for which some field data is available. Although some progress has been made in the analysis and fitting of PH distributions and MAPs, the practical applicability of many approaches is limited [4].

In this paper we propose an approach for fitting the parameters of a MAP according to the moments or joint moments. The technique may be easily extended to other measures like values of the distribution function or joint density. The approach applies the idea of a separate fitting of the distribution by a PH distribution and the expansion of the PH distribution to a MAP in a second step [16]. For the first step, the distribution fitting, we present a heuristic approach to fit an acyclic PH distribution in canonical form to a set of moments, then this APH distribution is transformed by some equivalence transformations into

a representation that is more convenient for MAP fitting. The second step then uses some joint moments to expand the distribution into a MAP.

The paper is structured as follows. In the next section we introduce our notation, present some known results and give a brief overview of related work. Afterwards, in Section 3, the fitting approach for APH distributions and the equivalence transformations for APH distributions are presented. Then a method to fit joint moments by expanding an APH distribution into a MAP is introduced. Section 5 summarizes the two preceding steps and puts them into an algorithmic frame. Afterwards some examples are presented. The paper ends with the conclusions and some hints about possible extensions of the proposed approach.

2 Definitions, Notations and Related Work

2.1 Basic Definitions

We use a coherent notation for the representation of PH distributions and MAPs and begin with the introduction of MAPs (see [16, 27] for more details). A MAP of order n is characterized by two $n \times n$ matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ such that

- $\mathbf{D}_0(i,j) \geq 0$ for $i \neq j$ and
  $\mathbf{D}_0(i,i) \leq -\sum_{j=1,i\neq j}^{n} \mathbf{D}_0(i,j)$,
- $\mathbf{D}_1(i,j) \geq 0$ and $\mathbf{D}_0\mathbf{e}^T = -\mathbf{D}_1\mathbf{e}^T$ where $\mathbf{e}$ is the unit row vector and $\mathbf{e}^T$ is its transposed,
- $\mathbf{D}_0$ is nonsingular and $\mathbf{D}_0 + \mathbf{D}_1$ is an irreducible generator matrix.

The MAP characterizes a stochastic process where events are generated whenever a transition from $\mathbf{D}_1$ occurs, transitions in $\mathbf{D}_0$ are silent.

Matrix $\mathbf{P} = -\mathbf{D}_0^{-1}\mathbf{D}_1$ is the stochastic transition matrix of the embedded phase process at arrival instants. The stationary distribution at arrival instants is denoted by π and is given by $\pi\mathbf{P} = \pi$ and $\pi\mathbf{e}^T = 1.0$.

A PH distribution is characterized by a matrix $\mathbf{D}_0$ with the same properties as above and an initial distribution π. The PH distribution is acyclic (an APH distribution) if $\mathbf{D}_0$ is an upper (or lower) triangular matrix.

Every MAP describes an embedded PH distribution characterized by the pair $(\pi, \mathbf{D}_0)$ and a PH distribution can be expanded into an equivalent MAP by defining $\mathbf{D}_1 = (-\mathbf{D}_0\mathbf{e}^T)\pi$.

A state i of a PH distribution or a MAP is an entry state, if $\pi(i) > 0$ and it is an exit state if the ith row of $\mathbf{D}_1$ is nonzero or, equivalently, if the row sum of the ith row of $\mathbf{D}_0$ is negative.

2.2 Analysis of PH and MAPs

In this paper we introduce an approach for fitting the parameters of a PH distribution or MAP to approximate the moments or joint moments of a process. We first introduce the computation of some quantities, in particular, of the moments and joint moments for PH distributions and MAPs. Let X be a random variable with phase type distribution and $X(t)$ a stochastic process characterized by a MAP, respectively.

The density of X realized by a PH distribution $(\pi, \mathbf{D}_0)$ is given by

$$f_{PH}(t) = \pi e^{\mathbf{D}_0 t}(-\mathbf{D}_0)\mathbf{e}^T = \pi \sum_{k=0}^{\infty} \frac{(\mathbf{D}_0 t)^k}{k!}(-\mathbf{D}_0)\mathbf{e}^T \tag{1}$$

and the joint density of $X(t)$ realized by a MAP $(\mathbf{D}_0, \mathbf{D}_1)$ is given by

$$f_{MAP}(t_1, t_2, \ldots, t_m) = \pi e^{\mathbf{D}_0 t_1}\mathbf{D}_1 e^{\mathbf{D}_0 t_2}\mathbf{D}_1 ... e^{\mathbf{D}_0 t_m}\mathbf{D}_1\mathbf{e}^T \ . \tag{2}$$

The moments and joint moments are derived from the moment matrix $\mathbf{M} = -\mathbf{D}_0^{-1}$. Such that we obtain for the moments

$$\mu_i = E(X^i) = i!\pi\mathbf{M}^i\mathbf{e}^T \tag{3}$$

and for the joint moments

$$\mu_{i,j} = E(X^i, X^j) = i!j!\pi\mathbf{M}^i\mathbf{P}\mathbf{M}^j\mathbf{e}^T \ . \tag{4}$$

As shown in [27], a PH distribution of order n is characterized by $2n-1$ moments and an order n MAP is characterized by n^2 joint moments, if it is non redundant. Since a PH distribution has n^2+n-1 free parameters and a MAP has $2n^2-n$ free parameters, the matrix representations are highly redundant which makes the fitting of parameters even harder. In the sequel we speak of a PH/APH distribution or a MAP, if the concrete and unique distribution/stochastic process is meant and we use PH/APH/MAP representation for the matrices and vectors related to the concrete representation. Thus, each representation describes a unique distribution and for a distribution infinitely many representations exist.

One of the problems for any fitting approach is that canonical representations of PH distributions or MAPs do not exist in general. This implies that different representations of the same distribution are available and fitting algorithms that work on the matrix representations may switch between different equivalent representations resulting in bad approximations. With the exception of $n = 2$ or 3 canonical representations exist only for APH distributions. The canonical representation and a method to transform a APH representation into the canonical form are given in [9]. We briefly present this form here since it is the base of our transformation method.

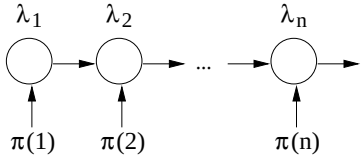

Figure 1. Canonical representation of APH distributions.

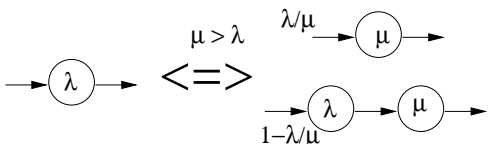

Figure 2. Equivalent representations of an exponential distribution.

The structure of the canonical representation of APH distributions with n phases is shown in Fig. 1. For the canonical representation $\lambda_n \geq \lambda_{n-1} \geq \ldots \geq \lambda_1$ holds. It has 1 exit and up to n entry states. The transformation of an APH representation into the canonical form is based on equivalent representations of the exponential distribution as shown in Fig. 2, details about an algorithm to transform any APH distribution into its canonical form can be found in [9] (see also [11, 15]). Unfortunately, the results cannot be easily extended to cyclic PH distributions and this class is larger than the class of APH distributions [4, 11].

2.3 Fitting Problems

The goal of the algorithms developed in this paper is to compute a MAP that matches or approximates the moments and joint moments of some observed process. Let ν_i and $\nu_{i,j}$ be the observed (joint) moments that should be approximated. Usually the values are extracted from some measurements and are therefore random variables which would in principle require a statistical evaluation. However, this point is usually not taken into account in fitting approaches. We assume here that ν_i and $\nu_{i,j}$ are unbiased estimates of the true values.

Even if an APH distribution of order n is completely characterized by $2n-1$ moments, $2n-1$ measured moments usually will not define a proper APH distribution of order n. Unfortunately, it is not even known which combinations of moments can be matched by an order n PH or APH distribution. Consequently, it is recommendable to go from an exact matching to a good approximation. Thus, the fitting problem becomes for distributions

$$\min_{(\pi,\mathbf{D}_0)} \left(\sum_{i \in \mathcal{M}} \left(\beta_i \frac{\mu_i}{\nu_i} - \beta_i \right)^2 \right) \tag{5}$$

where $(\pi, \mathbf{D}_0)$ is a valid PH (or APH) representation with moments μ_i, $\mathcal{M}$ is a set of moments to be approximated and β_i is a non negative weight which allows one to give different weights to the moments, e.g., to privilege lower order moments. There is no relation between the order of the (A)PH distribution and the number of moments in $\mathcal{M}$ since we are searching for an *optimal* approximation which becomes exact if the minimum in (5) becomes zero.

In the second step we start with a PH representation $(\pi, \mathbf{D}_0)$ that is expanded to a MAP representation $(\mathbf{D}_0, \mathbf{D}_1)$ to match additionally some joint moments $\nu_{i,j}$ from a set $\mathcal{J}$. The resulting minimization problem becomes

$$\min_{\mathbf{D}_1,\ \pi\mathbf{D}_0^{-1}\mathbf{D}_1=\pi} \left(\sum_{(i,j) \in \mathcal{J}} \left(\beta_{i,j} \frac{\mu_{i,j}}{\nu_{i,j}} - \beta_{i,j} \right)^2 \right) \tag{6}$$

Here $(\mathbf{D}_0, \mathbf{D}_1)$ is a valid MAP representation with a given matrix $\mathbf{D}_0$ and a given stationary arrival distribution π.

2.4 Related Work

There is an enormous amount of material on PH distributions and less but still much material on MAPs. In particular, a large number of fitting methods exists but almost all approaches have problems when applied to field data which is in the area of computer or communication systems often given as a trace with $10^5, 10^6$ or even more elements. We cannot give a comprehensive overview of related work and only highlight some major results related to our work.

For PH distributions different papers exist that consider the range of distributions that can be expressed with a PH distribution of fixed order and which search for canonical or unique representations (e.g., [21, 11]). However, apart from the

mentioned canonical representation for APH distributions [9] and analytical results for PH distributions of order two [26], a general characterization of the set of reachable moments is still missing for PH distributions. In [15] a recursive approach is developed that allows one to fit the parameters of an APH distribution of order n to a given set of $2n-1$ moments, if such a fitting is possible. Since the approach is exact, the moments have to be representable by an order n APH distribution, it is not possible to compute an approximation or increase the order of the distribution to fit a given set of moments.

An overview of different heuristic fitting methods for PH distributions and MAPs can be found in [13]. Most of these heuristic techniques are based on the EM algorithm [2] which maximizes the likelihood according to a trace or a set of computed values of the distribution function. Since the original algorithm is very slow, several variants have been developed subsequently [17, 10, 28, 23], which are more efficient but still might have problems if the order of the phase type distribution is large or the trace contains many values. Additionally, other heuristic nonlinear optimization approaches have been applied to compute ML estimates for the parameters of an APH distribution according to an empirical density [3, 14]. The major limitation of EM based approaches compared to moment fitting is that the former techniques have to work on the whole and possibly huge trace whereas moment fitting is independent of the length of the trace. On the other hand, the meaning of higher order moments is not really clear and their estimation is often unreliable as shown by our examples.

The fitting of MAPs is significantly more complex than the fitting of PH distributions. Moment and joint moment fitting has been developed for acyclic MAPs of order two [12]. EM algorithms can be and have been extended to fit MAPs [5, 24] but are usually rather inefficient and therefore only applicable if the order of the MAP is small and the trace is not too long. Alternative techniques like the KPC approach [7] build a MAP by the composition of several MAPs of order 2 for which moment fitting is easier. However, the price of this approach might be a MAP with a huge state space.

Most related to our work is [27] where a characterization of PH distributions and MAPs is introduced and a moment fitting method is presented. In contrast to our moment fitting method, [27] first fits a matrix exponential distribution and a matrix exponential process and then transforms these representations into a PH distribution or MAP. The transformation is the crucial step which is done via a general purpose optimization technique which may fail to find an appropriate representation. In contrast, our method always finds an APH distribution and a MAP but usually only approximates the given moments and joint moments. If one keeps in mind that moments and joint moments are only estimates derived from some sample, an approximation is often as good as an exact fitting, if the approximation error is small.

3 Approximate Fitting of APH-Distributions

In this section we propose an approach to fit the parameters of an APH distribution according to some moments which usually have been measured. The problem is treated as a general optimization problem which is solved approximately in subsection 3.1 by the repeated optimization of simpler problems. The resulting APH distribution in canonical form is not suited for a subsequent MAP fitting since it contains only a single exit state. Thus, it is transformed into an equivalent APH representation with additional exit states.

3.1 Generation of an APH Distribution in Canonical Form

We now develop an algorithm to find an APH distribution in canonical form that approximates a set of moments $\mathcal{M}$. In other words, we solve the optimization problem (5). If $\mathcal{M}$ contains $2n-1$ moments which can be exactly matched by an order n APH distribution, then an exact matching using the approach from [15] is preferable. However, since this is usually not the case for empirical moments, the use of a general optimization approach has the advantage of computing approximations independently of the number and size of the moments and the order of the APH distribution.

The matrix $\mathbf{D}_0$ of an APH distribution in canonical form has the following structure.

$$\begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & 0 & -\lambda_{n-1} & \lambda_{n-1} \\ 0 & \cdots & \cdots & 0 & -\lambda_n \end{pmatrix}$$

such that matrix $\mathbf{M}$ becomes

$$\begin{pmatrix} \frac{1}{\lambda_1} & \frac{1}{\lambda_2} & \cdots & \cdots & \frac{1}{\lambda_n} \\ 0 & \frac{1}{\lambda_2} & \cdots & \cdots & \frac{1}{\lambda_n} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & 0 & \frac{1}{\lambda_{n-1}} & \frac{1}{\lambda_n} \\ 0 & \cdots & \cdots & 0 & \frac{1}{\lambda_n} \end{pmatrix}$$

Let $\mathbf{m}_i = \mathbf{M}^i \mathbf{e}^T$ be the vector of the ith conditional moment, then $\mu_i = \pi \mathbf{m}_i$. Assume that $\Lambda = (\lambda_1, \ldots, \lambda_n)$ is known, then the minimization problem (5) according to π becomes

$$\min_{\pi : \pi \mathbf{e}^T = 1, \pi \geq 0} \left(\sum_{i \in \mathcal{M}} \left(\beta_i \frac{\pi \mathbf{m}_i}{\nu_i} - \beta_i \right)^2 \right) \tag{7}$$

This is a non negative least squares problem with a single linear constraint for which efficient solution algorithms exist [18].

Optimization according to Λ is harder since matrices $\mathbf{M}^i$ are required for the computation of the ith moment. Thus, we consider the optimization according to a single rate λ_r for $1 \leq r \leq n$. Assume that we modify λ_r such that $1/\lambda_r$ becomes $1/\lambda_r + \Delta$ (i.e. λ_r becomes $\lambda_r/(1+\lambda_r\Delta)$). Then the moments matrix becomes

$$\mathbf{M}_{\Delta,r} = \mathbf{M} + \Delta \mathbf{E}_r$$

where $\mathbf{E}_r$ is a $n \times n$ matrix with 1 in the position $(1,r), \ldots, (r,r)$ and 0 elsewhere. The ith conditional moment for a fixed initial vector π and fixed rates λ_s ($s \neq r$) is then given by

$$\mu_i(\Delta, r) = \pi \left(\mathbf{M} + \Delta \mathbf{E}_r \right)^i \mathbf{e}^T$$

such that the minimization problem

$$\min_{\Delta} \left(\sum_{i \in \mathcal{M}} \left(\beta_i \frac{\mu_i(\Delta, r)}{\nu_i} - \beta_i \right)^2 \right) \tag{8}$$

has to be solved. $\mu_i(\Delta, r)$ is a polynomial of order i and a minimum of (8) can be found with standard optimization techniques. Either Δ is chosen from $[1/\lambda_{i+1} - 1/\lambda_i, 1/\lambda_{i-1} - 1/\lambda_i]$ such that the resulting APH distribution is still in canonical form or the optimum is computed for $\Delta \in (-1/\lambda_r, \infty)$ that assures that λ_r remains non negative and the resulting APH distribution can be transformed into the canonical form using the approach from [9].

The two steps (7) and (8) can be iterated by optimization according to π and according to λ_r ($r = 1, \ldots, n$) until the parameters remain almost constant. This approach results in a sequence of APH distributions with a decreasing error in (5). However, the approach converges to local minima such that it should be restarted several times at random points to obtain a good approximation. We come back to concrete algorithms in section 5.

3.2 Transformation of the Canonical Form

A transformation from the canonical form to a general APH distribution with more than one exit state has been proposed in [16]. We present here another transformation which basically inverts the transformation process into the canonical form as given in [9]. However, this inversion is not unique and influences the range of joint moments that can be reached in the steps where the APH representation is expanded to a MAP. The transformation performs modification on $\mathbf{D}_0$ and π that do not alter the distribution. Of course, the transformation steps can be applied to any APH representation, it need not be in canonical form. Unfortunately, it is not clear yet how to find the most flexible APH representation for a distribution where most flexible means the one with widest range of joint moments that can be represented by expanding the representation into a MAP. Thus, several transformation may be checked.

The transformation consists of a sequence of steps and in each step two states i and j ($i < j$) that are connected by a transition are handled. Each transformation step can be done locally by considering only incoming and outgoing transitions of i, incoming transitions of j, and the probabilities $\pi(i)$ and $\pi(j)$. Consequently, the representation remains acyclic. Furthermore, our transformations keep the transition rates $\lambda_1, \ldots, \lambda_n$ unchanged such that $\lambda_i \leq \lambda_j$ holds for $i < j$.

In the following description of a transformation step, we choose two states i and j with $i < j$, $q(i,j) > 0$ and compute new transition rates $q'(.,.)$ and initial probabilities $\pi'(.)$ for a different representation of the same distribution.

Define $\delta \leq \delta^*$ where

$$\begin{aligned} \delta^* = \quad & \min \Big(\pi(j), \tfrac{\pi(i)q(i,j)}{\lambda_j - \lambda_i}, \\ & \min_{k<i, q(k,i)>0} \Big(\pi(i) \tfrac{q(k,j)}{q(k,i)} \Big) \Big) \ . \end{aligned} \tag{9}$$

for $\lambda_j > \lambda_i$. If $\lambda_i = \lambda_j$, then the second term in the minimum (9) becomes ∞ and does not count such that the minimum is computed according to the remaining two conditions. If $\delta^* > 0$, we can choose $\delta^* > \delta > 0$ and perform the following transformations for π

$$\pi'(k) = \begin{cases} \pi(i) + \delta & \text{for } k = i \\ \pi(j) - \delta & \text{for } k = j \\ \pi(k) & \text{otherwise} \end{cases} \tag{10}$$

and $q(k,l)$.

$$q'(k,l) = \begin{cases} q(i,j)\frac{\pi(i)}{\pi(i)+\delta} - \\ \quad \frac{(\lambda_j-\lambda_i)\delta}{\pi(i)+\delta} & \text{for } k=i \text{ and } l=j \\ q(i,l)\frac{\pi(i)}{\pi(i)+\delta} + \\ \quad q(j,l)\frac{\delta}{\pi(i)+\delta} & \text{for } k=i \text{ and } l\neq j \\ q(k,i)\frac{\pi(i)+\delta}{\pi(i)} & \text{for } k<i \text{ and } l=i \\ q(k,j) - q(k,i)\frac{\delta}{\pi(i)} & \text{for } k<j \text{ and } l=j \\ q(k,l) & \text{otherwise} \end{cases} \tag{11}$$

The following theorems summarize some properties of the distribution resulting from the transformation. Proofs can be found in a technical report [6].

Theorem 1. *If (10) and (11) are applied to an APH representation, then the resulting representation is still an APH representation and describes the same distribution.*

Theorem 2. *If (10) and (11) are applied to an APH representation, then* $\sum_{l=k+1}^{n} q'(k,l) = \sum_{l=k+1}^{n} q(k,l)$ *for* $k \neq i$ *and*

$$\sum_{l=i+1}^{n} q'(i,l) = \sum_{l=i+1}^{n} q(i,l) + \frac{\delta}{\pi(i)+\delta} (q(i,n+1) - q(j,n+1))$$

where $q(k,n+1) = \lambda_k - \sum_{l=k+1}^{n} q(k,l)$ *(*$k = i, j$*).*

Theorem 2 implies that if j is an exit state, then also i becomes an exit state. The following theorem shows that the repeated application of the transformation results in an APH representation with n exit states under certain conditions.

Theorem 3. *If we apply the transformation rules (10) and (11) with* $\delta^* > \delta > 0$ *consecutively to states* $i = 1, 2, \ldots, n$ *and* $j = i+1, \ldots, n$ *of an APH representation in canonical form with* $\pi(i) > 0$ *for all* $i = 1, \ldots, n$*, we obtain an APH representation of the same distribution where all states are exit states.*

The resulting APH representation of the distribution depends on the choice of δ in every step and defines the range of the joint moments that can be reached in the subsequent fitting step. We obtained good results by choosing $\delta = 0.9\delta^*$.

Theorem 3 implies that in the canonical representation all states are entry states. If this is not the case, then the transformation may generate a representation with fewer exit states. E.g., for an Erlang n distribution no transformation is possible. By a slight modification of (7), it can be assured that $\pi(i) > \epsilon > 0$ for some $\epsilon < 1/n$. In the modified problem the minimum is computed according to vector ϕ with the constraint $\phi \mathbf{e}^T = 1 - n\epsilon$ and $\phi \geq 0$. For $\epsilon > 0$ we obtain a transformed APH representation where every state is entry and exit state.

4 Fitting of Joint Moments for MAPs

Starting from some APH or PH representation we now present an approach to fit additionally some joint moments by expanding the distribution into a MAP. The following approach can be applied to any PH representation. Thus, it need not be

an APH representation and it can be generated with any of the available methods. Alternatively to the approach presented in the previous section, exact moments fitting as in [15] may be applied, if possible, or a PH representation is generated using one of the approaches presented in [2, 5, 17, 14, 28]. Of course, the range of joint moments that can be fitted depends on the structure of the PH representation. Representations with only one entry or exit state have no flexibility and cannot be used to fit joint moments.

The goal is to find a matrix $\mathbf{D}_1$ for given π and $\mathbf{D}_0$ such that $\mathbf{D}_1 \geq \mathbf{0}$, $-\mathbf{D}_0\mathbf{e}^T = \mathbf{D}_1\mathbf{e}^T$ and $\pi\mathbf{M}\mathbf{D}_1 = \pi$. The first condition is a non negativity condition for all elements in the matrix, the second and third condition define together $2n$ linear constraints for the elements of $\mathbf{D}_1$. Define

$$\mathbf{v}^i = \pi\mathbf{M}^{i+1} \text{ and } \mathbf{w}^i = \mathbf{M}^i\mathbf{e}^T \ ,$$

then

$$\mu_{i,j} = i!j! \sum_{r=1}^{n}\sum_{s=1}^{n} \mathbf{v}^i(r)\mathbf{D}_1(r,s)\mathbf{w}^j(s) \ . \tag{12}$$

Now consider again the general minimization problem (6) where (12) is plugged in for $\mu_{i,j}$. The resulting problem is a non negative least squares problem with n^2 variables and $2n$ linear constraints. This problem can be solved with standard algorithms for non negative least squares problems [18].

In summary, for a given PH representation the expansion into a MAP to approximate a set of joint moments is fairly easy and can be done efficiently as long as n is moderate (i.e. in the range of 20 to 50). If for an APH representation where the moments are adequately fitted, the fitting of joint moments is not good enough, one may either repeat moment fitting with additional states or one may expand the given APH representation by substituting one of its phases with rate λ_i by two phases with rates λ_i and μ_i ($> \lambda_i$) according to the equivalence in Fig. 2. This transformation keeps the distribution unmodified but enlarges the flexibility of fitting joint moments.

5 An Algorithmic Approach for Fitting

The two steps presented in the previous section will now be combined and described in an algorithmic framework resulting in a complete fitting algorithm. The first step is the determination of an APH distribution in canonical representation. As mentioned, a single optimization step converges towards a local minimum such that the algorithm should be restarted with random values several times. The following algorithm performs the fitting.

Algorithm 1.

1. *repeat*
2. *generate n positive random numbers ;*
3. *order the generated numbers and initialize λ_i ;*
4. *generate randomly an initial distribution π ;*
5. *repeat*

6. $\lambda'_i = \lambda_i$ *and* $\pi'(i) = \pi(i)$ *(*$i = 1, \ldots, n$*) ;*
7. *for (*$i = 1, \ldots, n$*) do*
8. *compute* $\Delta \in (-1/\lambda_i, \infty)$ *that minimizes (8) ;*
9. *recompute the canonical representation using the approach from [9] ;*
10. *recompute* π *by computing the minimum of (7) ;*
11. *until* $|\lambda'_i - \lambda_i| < \epsilon$ *and* $|\pi'(i) - \pi(i)| < \epsilon$ *(*$i = 1, \ldots, n$*) ;*
12. *until maximal number of restarts is reached or the moment approximation is good enough ;*

If the approximation of the moments is not appropriate, then n is set to $n + 1$ and the algorithm is restarted or the number of restarts is increased. In the next step, the canonical representation is transformed to increase the number of exit states using the following algorithm.

Algorithm 2.

1. *Initialize* $\mathbf{Q}(i, i) = -\lambda_i$ *(*$i = 1, \ldots, n$*) ;*
2. *Initialize* $\mathbf{Q}(j, j + 1) = \lambda_j$ *(*$j = 1, \ldots, n - 1$*) ;*
3. *for (*$i = 1, \ldots, n$*)*
4. *for (*$j = i + 1, \ldots, n$*)*
5. *compute* π' *from* π *using (10) ;*
6. *compute* $\mathbf{Q}'$ *from* $\mathbf{Q}$ *using (11) ;*
7. $\pi = \pi'$ *and* $\mathbf{Q} = \mathbf{Q}'$ *;*
8. $\mathbf{D}_0 = \mathbf{Q}$ *;*

If the representation is appropriate (i.e., moments are appropriately fitted and several entry and exit states exist which can be always assured by defining a minimal non zero value for the elements in π), the third algorithm expands the distribution into a MAP.

Algorithm 3.

1. *compute* $\mathbf{v}^i$ *and* $\mathbf{w}^j$ *for all* $(i, j) \in \mathcal{J}$ *;*
2. *solve (6) representing the joint moments by (12) and considering the constraints* $\pi\mathbf{M}\mathbf{D}_1 = \pi$ *and* $-\mathbf{D}_0\mathbf{e}^T = \mathbf{D}_1\mathbf{e}^T$ *;*

If the joint moments are not adequately approximated, then n is set to $n+1$ and the whole approach starts again or the representation is expanded by cloning some of its phases which results in a larger state space and more flexibility for the fitting of joint moments.

For the solution of the non negative least square problems in (6) and (7) the procedure [25] is used. Constraints are integrated in the goal function using Lagrange multipliers. The most costly step of the approach is the moment fitting with the APH distribution since it is usually recommended to do some $(10-100)$ restarts for a good fitting. However, even this step requires only a few second, if n is not too large. If other methods for distribution fitting are available, these may as well be used. We made good experience with the *gfit* approach from [28] which uses an EM algorithm to fit hyper Erlang distribution. For n equal to 10 or below, the whole fitting requires only a few seconds. In particular the fitting of joint moments using non negative least squares is much faster than EM algorithms or similar approaches working on a complete trace.

6 Experiments and Results

We present now some examples to show the potential and limitations of the proposed fitting algorithm. In practice we are interested in fitting MAPs to real traffic traces. However, in the first subsection we consider some results generated from a MAP which are fitted by MAPs of the same size. In this case, exact values for moments and joint moments are known. Afterwards we consider some real traces.

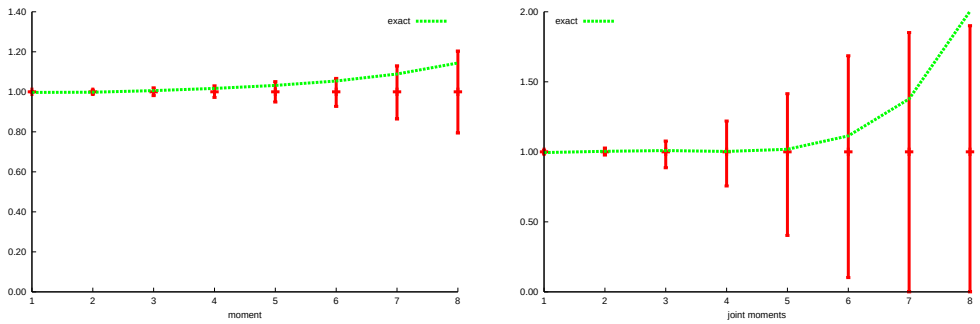

Figure 3. Confidence intervals for the moments $E(X^i)$ and joint moments $E(X^i, X^i)$ of the example MAP.

6.1 Fitting MAPs

We present results for one of the example MAPs from [16]. Of course, theoretically, if $2n-1$ moments and $(n-1)^2$ joint moments are known, then a MAP can be recreated. However, in practice the following problems may come up:

- We use APH distributions rather than PH distributions and restrict for $n > 2$ the class of distributions that can be represented.
- The moment fitting approach for APH is only a heuristic that may fail to find the optimal representation.
- The transformation of the canonical form to a general APH is non unique and the resulting distribution may not be adequate for fitting the required joint moments.

- In practice, moments are estimated from some trace since the exact moments are not known such that the quality of the MAP depends on the quality of the estimates.
- In practice $2n-1$ moments and $(n-1)^2$ joint moments often will not define a proper MAP of order n and even if such a MAP exists it is not clear whether it adequately approximates additional quantities of the trace like additional moments or lag k autocorrelations.

The first three points are related to the fitting method, the last two are more general. We begin with the analysis of the following MAP with 3 states that has been used in [16] as example one.

$$\mathbf{D}_0 = \begin{pmatrix} -3.721 & 0.5 & 0.02 \\ 0.1 & -1.206 & 0.005 \\ 0.001 & 0.002 & -0.031 \end{pmatrix}$$
$$\mathbf{D}_1 = \begin{pmatrix} 0.2 & 3.0 & 0.001 \\ 1.0 & 0.1 & 0.001 \\ 0.005 & 0.003 & 0.02 \end{pmatrix}$$

The MAP is cyclic but can be adequately represented by a MAP with acyclic matrix $\mathbf{D}_0$. The algorithm computes the following MAP.

$$\mathbf{D}_0 = \begin{pmatrix} -0.031 & 0.00244 & 0.00267 \\ 0.0 & -0.115 & 0.0107 \\ 0.0 & 0.0 & -1.637 \end{pmatrix}$$
$$\mathbf{D}_1 = \begin{pmatrix} 0.02 & 0.000432 & 0.00538 \\ 0.0 & 0.104 & 0.0 \\ 0.00951 & 0.00235 & 1.625 \end{pmatrix}$$

The whole computation of the MAP requires about 2 seconds which is much faster than with other known methods. Although the original and the fitted MAP look fairly different, their moments and joint moments are almost the same.

$$(\mu_i) = (1.14986, 35.7530, 3375.66, 435509, 70277728)$$
$$(\nu_i) = (1.14986, 35.7398, 3378.09, 435482, 70253190)$$
$$(\mu_{i,j}) = \begin{pmatrix} 1.18e+1 & 7.37e+2 & 7.12e+4 \\ 7.36e+2 & 4.70e+4 & 4.55e+6 \\ 7.12e+4 & 4.55e+6 & 4.41e+8 \end{pmatrix}$$
$$(\nu_{i,j}) = \begin{pmatrix} 1.19e+1 & 7.36e+2 & 7.12e+4 \\ 7.35e+2 & 4.70e+4 & 4.55e+6 \\ 7.11e+4 & 4.55e+6 & 4.41e+8 \end{pmatrix}$$

The lag k autocorrelations are also almost identically for the original and the fitted MAP, whereas the density functions differ slightly at some points. In summary, for the example MAP the fitting was efficient and adequate. However, in practice we do not have the exact moments and joint moments for a MAP, they have to be extracted from a trace. To show the potential

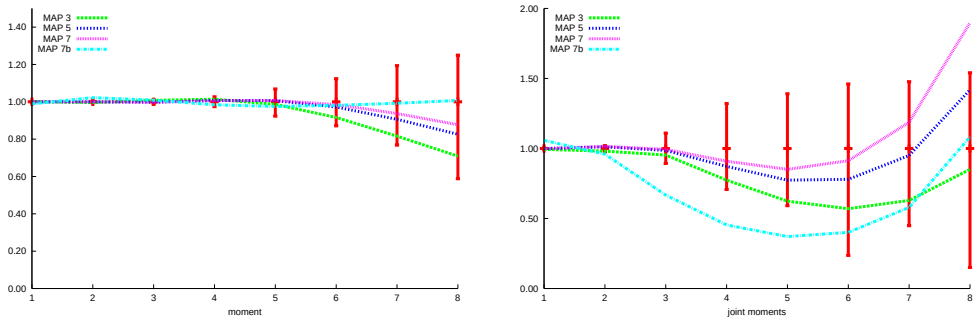

Figure 4. Confidence intervals for the moments $E(X^i)$ and joint moments $E(X^i, X^i)$ of the LBL-trace and the fitted moments and joint moments.

problems which may come up in such a case, we generate a trace from the example MAP and estimate the moments and joint moments from the trace. The accuracy of the estimates is shown by computing confidence intervals for the moments and joint moments. Since the values in the trace are dependent, we use bootstrapping to compute confidence intervals by applying the approach for mean values from [8].

For the following results a trace with 10^7 elements has been generated from the MAP. From this trace the moments and joint moments are computed. Figure 3 shows the width of the confidence intervals for the moments and joint moments relative to the estimated mean values for significance level $1 - \alpha = 90\%$. Furthermore, the exact values of the moments and joint moments which have been computed numerically from the MAP are drawn in the figure relative to the estimated moments. It can be seen that for lower order moments and joint moments confidence intervals are small and the estimated and exact values are very similar. However, for higher order moments and, in particular joint moments, the situation is problematic. The width of the confidence interval of the fifth moment is already 10% of the estimated mean and for the ninth moment the confidence interval is 60% of the estimated mean. It can also be seen that higher order moments and joint moments of the trace and the MAP differ significantly. To show the practical problems which arise, we consider some of the values resulting from three traces with 10^7 elements which are generated from the original MAP with different seeds of the random number generator. For the MAP μ_1 equals 1.149 from the traces we estimate $(1.152, 1.150, 1.150)$, μ_5 equals $7.028e+7$ from the traces we estimate $(7.192e+7, 6.758e+7, 6.808e+7)$ and for $\mu_{3,3}$ the exact value equals $4.406e+8$ whereas the traces result in $(4.031e+8, 4.856e+8, 4.368e+8)$. Similarly, the lag 10 autocorrelation of the MAP equals $5.69e-2$ and the traces result in values between $5.45e-2$ and $6.19e-2$.

If we fit a MAP according to the estimated values of one of the traces, we obtain a process that significantly differs in its behavior from the original MAP. We analyzed several other MAPs (e.g., all the examples presented in [16]). In all cases the fitting results are very good. Thus, the fitting procedure is much more accurate than the information we obtain from a trace even with 10 million entries. We will come back to this point after considering some real traces.

6.2 Fitting Real Traces

As real examples we use two traces from the internet traffic archive [22, 1] which often have been used as benchmarks for MAP or distribution fitting. The first traces is the *LBL-TCP-3* trace which describes two hours of wide-area TCP traffic

capturing about 1.8 million packets. The second trace is the *pAug89* trace which includes the interarrival times of one million Ethernet packets.

For both traces we estimated ν_i and $\nu_{i,j}$ for $i, j = 1, \ldots, 8$ from the traces and computed 90% confidence intervals via bootstrapping. Then MAPs with 3, 5 and 7 states are fitted according to the moments and joint moments. We use the weights $\beta_i = 2^{i+1}$ and $\beta_{i,j} = 2^{-(i-1)(j-1)}$. Of course, this choice is somehow arbitrary but captures the observation that lower order moments are more important and can be estimated more reliable. However, an appropriate choice of weights requires further investigations.

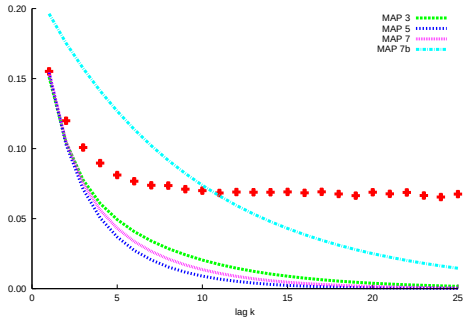

Figure 5. Lag k autocorrelation of the LBL-trace and the fitted MAPs.

Fig. 4 shows the results of moment and joint moment fitting for the LBL-trace. In the figures we plotted the width of the confidence intervals for the moments and joint moments relative to the absolute values. For the joint moments we show only the values for $\nu_{i,i}$, however, results for $\nu_{i,j}$ ($i \neq j$) are similar. Like in the previous examples, confidence intervals for the first three moments and joint moments are acceptable, afterwards the confidence intervals become too large. The 1.7 million elements in the trace are not sufficient for an accurate estimation of the higher order moments and joint moments. The figures also include curves for μ_i/ν_i and $\mu_{i,i}/\nu_{i,i}$ where $\mu_\cdot$ are the values of the fitted MAP and $\nu_\cdot$ are the estimated values. *MAPx* stands for a MAP with x states fitted according to the weighted moments, whereas *MAPxb* stands for a MAP of order x fitted according to the raw moments and joint moments. It can be seen that fitting of the weighted moments results in an almost exact fitting of the first five moments and the first three joint moments in all cases whereas higher order moments are underestimated and higher order joint moments are overestimated. An increased number of states results in a better fitting, e.g. going from 3 to 7 states reduces the squared error in (5) by a factor of 10 and the squared error of (6) by a factor of 2. Fitting without weights results in a better approximation of higher order moments but the price are small errors in lower order moments. E.g., with *MAP7b* the first moment is underestimated by 2% and the second moment is overestimated by 2% whereas the MAPs fitted according to the weighted moments yield exact results.

Fig. 5 shows the autocorrelation of the original trace and the fitted MAPs. It should be remarked that the autocorrelation plays no role in the fitting algorithm. It can be seen that the confidence intervals for the lag k autocorrelation are small and that the MAPs do not adequately represent the autocorrelation structure. Only the first two lags are described by the MAPs fitted according to the weighted traces whereas the unweighted fitting first over- and then underestimates the lag k

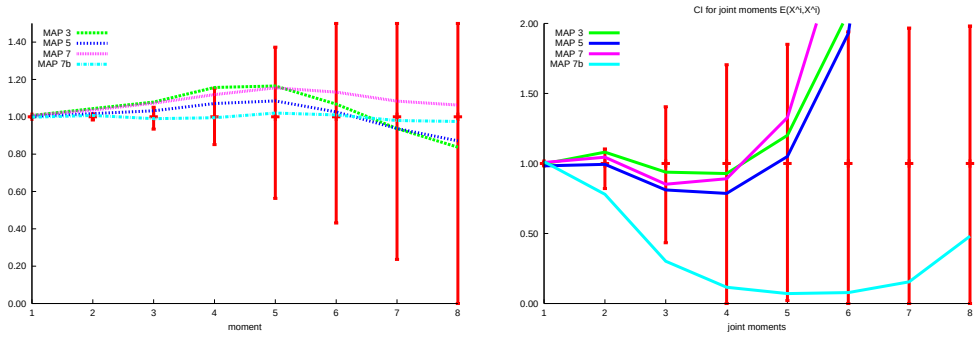

Figure 6. Confidence intervals for the moments $E(X^i)$ and joint moments $E(X^i, X^i)$ of the pAug-trace and the fitted moments and joint moments.

autocorrelation. The results indicates that moment and joint moment fitting without fitting the autocorrelation for larger lags can be questionable. However, it is also the other way round, fitting according to the autocorrelation without the moments usually results in a bad approximation of joint moments other than the first one which is determined by the lag 1 autocorrelation coefficient.

Results for the second trace *pAug89* are shown in the Fig. 6. For this example moments are slightly easier to fit, whereas higher order joint moments are hard to fit by low order MAPs.

7 Conclusions

We presented an approach to fit the moments and joint moments of a MAP according to the moments and joint moments of a traffic trace. In contrast to other approaches, moment fitting is considered as a minimization problem such that the fitting of MAP parameters according to the joint moments becomes a non negative least squares problem, if the distribution has already been fitted. Since distribution fitting is usually easier than fitting of MAPs, the approach describes an easy and fast method to expand a PH distribution to a MAP. The method usually results in good fitting results, if moments and joint moments generated from some MAP are used for fitting. However, the situation becomes much more complex, if moments and joint moments result from traces. It has been shown that even for large traces the estimates for higher order moments or joint moments are unreliable such that moment fitting can only be recommended for the low order moments. In so far, the question for the best measures to be used in a fitting approach is still open. However, our results indicate that a statistical evaluation of the trace is recommended since an accurate fitting of unreliable measures is not better than a loose approximation of exact measures.

Of course, the approach presented here requires some additional tuning in particular to find appropriate values for the weights used for the goal functions and to find the best representation of an APH distribution for expansion into a MAP. These steps will be considered in future research.

Acknowledgment

We thank the anonymous reviewers for their detailed reviews indicating several flaws in the submitted version of the paper.

References

[1] The internet traffic archive. http://ita.ee.lbl.gov/html/traces.html.

[2] S. Asmussen, O. Nerman, and M. Olsson. Fitting phase-type distributions via the EM-algorithm. *Scand. J. Stat.*, 23(4):419–441, 1996.

[3] A. Bobbio and A. Cumani. ML estimation of the parameters of a PH distribution in triangular canonical form. In G. Balbo and G. Serazzi, editors, *Computer Performance Evaluation*, pages 33–46. Elsevier, 1992.

[4] L. Bodrog, A. Heindl, G. Horváth, M. Telek, and A. Horváth. Current results and open questions on PH and MAP characterization. In D. Bini, B. Meini, V. Ramaswami, M. Remiche, and P. Taylor, editors, *Numerical Methods for Structured Markov Chains*, number 07461 in Dagstuhl Seminar Proceedings, 2008.

[5] P. Buchholz. An EM-algorithm for MAP fitting from real traffic data. In P. Kemper and W. H. Sanders, editors, *Computer Performance Evaluation / TOOLS*, volume 2794 of *Lecture Notes in Computer Science*, pages 218–236. Springer, 2003.

[6] P. Buchholz and J. Kriege. Equivalence transformations for acyclic phase type distributions. Technical report, Dep. of Informatics, TU Dortmund, 2009.

[7] G. Casale, E. Z. Zhang, and E. Smirni. KPC-toolbox: Simple yet effective trace fitting using Markovian arrival processes. In *QEST*, pages 83–92. IEEE Computer Society, 2008.

[8] D. Choquet, P. L'Ecuyer, and C. Léger. Bootstrap confidence intervals for ratios of expectations. *ACM Trans. Model. Comput. Simul.*, 9(4):326–348, 1999.

[9] A. Cumani. On the canonical representation of homogeneous Markov processes modeling failure-time distributions. *Micorelectronics and Reliability*, 22(3):583–602, 1982.

[10] A. Feldmann and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31:245–279, 1998.

[11] Q. M. He and H. Zhang. A note on unicyclic representations of phase type distributions. *Stochastic Models*, 21:465–483, 2005.

[12] A. Heindl, G. Horváth, and K. Gross. Explicit inverse characterizations of acyclic MAPs of second order. In A. Horváth and M. Telek, editors, *EPEW*, volume 4054 of *Lecture Notes in Computer Science*, pages 108–122. Springer, 2006.

[13] A. Horváth and M. Telek. Markovian modeling of real data traffic: Heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In M. Calzarossa and S. Tucci, editors, *Performance*, volume 2459 of *Lecture Notes in Computer Science*, pages 405–434. Springer, 2002.

[14] A. Horvath and M. Telek. Phfit: A general purpose phase type fitting tool. In *Performance Tools 2002*, volume 2324 of *LNCS*, pages 82–91. Springer, 2002.

[15] A. Horvath and M. Telek. Matching more than three moments with acyclic phase type distributions. *Stochastic Models*, 23:167–194, 2007.

[16] G. Horvath, M. Telek, and P. Buchholz. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag-correlation. In *QEST*, pages 124–133. IEEE CS Press, 2005.

[17] R. El Abdouni Khayari, R. Sadre, and B. Haverkort. Fitting world-wide web request traces with the EM-algorithm. *Performance Evaluation*, 52:175–191, 2003.

[18] C. L. Lawson and B. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, 1974.

[19] M. F. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779, 1979.

[20] M. F. Neuts. *Matrix-geometric solutions in stochastic models*. Johns Hopkins University Press, 1981.

[21] C. A. O'Cinneide. Phase-type distributions: open problems and a few properties. *Stochastic Models*, 15(4):731–757, 1999.

[22] V. Paxson and S. Floyd. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3:226–244, 1995.

[23] A. Riska, V. Diev, and E. Smirni. An EM-based technique for approximating long-tailed data sets with PH distributions. *Performance Evaluation*, 55:147–164, 2004.

[24] T. Rydén. An EM algorithm for estimationin Markov modulated Poisson processes. *Computationals Statistics and Data Analysis*, 21:431–447, 1996.

[25] S. Suvrit. Source of the non negative least squares algorithm from [18]. http://www.kyb.tuebingen.mpg.de/bs/people/suvrit/work/progs/nnls.html.

[26] M. Telek and A. Heindl. Matching moments for acyclic discrete and continuous phase-type distributions of second order. *International Journal of Simulation Systems, Science and Technology*, 3, 2002.

[27] M. Telek and G. Horváth. A minimal representation of markov arrival processes and a moments matching method. *Perform. Eval.*, 64(9-12):1153–1168, 2007.

[28] A. Thümmler, P. Buchholz, and M. Telek. A novel approach for phase-type fitting with the EM algorithm. *IEEE Trans. Dep. Sec. Comput.*, 3(3):245–258, 2006.