Fitting Matrix Geometric Distributions by Model Reduction

Nail Akar

Abstract

A novel algorithmic method is proposed to fit matrix geometric distributions of desired order to empirical data or arbitrary discrete distributions. The proposed method effectively combines two existing approaches from two different disciplines: well-established model reduction methods used in system theory, and moment matching methods of applied probability that employ second order discrete phase-type distributions. The proposed approach is validated with exhaustive numerical examples including well-known statistical data.

I. INTRODUCTION

Phase-type (PH) distributions, continuous or discrete, form a very general class of distributions that have been successfully used in performance modeling and queuing systems analysis in a wide variety of disciplines for the last few decades. Continuous Phase-type (CPH) distributions are described in detail in Neuts [1] and Latouche and Ramaswami [2]. Examples that use CPH distributions for stochastic modeling purposes but in different fields can be found in Nielsen [3], Neuts and Meier [4], and Drekic et al. [5], for models of multiple access communication systems, reliability of systems, and deficit distributions at ruin, respectively. CPH distributions have attracted more attention than their discrete counterparts, namely Discrete Phase-type (DPH) distributions, the latter first described in Neuts [6] along with their properties and their use in queuing system modeling. Both CPH and DPH distributions are closed under a number of operations. For example, convolution, mixture, minimum, or maximum, of finite independent PH-type distributions are also of PH-type; see for example Assaf and Levikson [7], O’Cinneide [8]. Other appealing features are their rational characteristic functions and representability by a pair of matrices. The most popular approach is the matrix-analytical technique to solve queuing systems involving PH-type distributions Neuts [1], Asmussen [9], Latouche and Ramaswami [2].

Matrix Exponential (ME) distributions inherit most of the features, including closure properties, of CPH distributions but they are more general than CPH; see Asmussen and O’Cinneide [10], Fackrell [11], and He and Zhang [12]. ME distributions have rational Laplace transforms similar to CPH distributions. However, they do not necessarily possess the probabilistic interpretation that CPH distributions have. Still, it has been shown that queues with ME-type arrival or service processes can also be analyzed using matrix-analytical techniques as in Asmussen and Bladt [13], Akar [14], Buchholz and Telek [15]. Similar to the generalization of ME over CPH, Matrix Geometric (MG) distributions generalize those of DPH-type; see Turin [16] and Greeuw [17]. A non-negative integer-valued discrete random variable possesses an MG distribution that is characterized with a Probability Mass Function (PMF) of the form \( p(k) = v^T h \), \( k \geq 1 \), \( p(0) = d \) for a row vector \( v \), column vector \( h \), and a finite square matrix \( T \) whose size gives the order of the MG distribution. Greeuw [17] uses the characterization of discrete phase-type distributions by O’Cinneide [8] to provide distributions that are in MG but not in DPH. Maier [18] presents an algorithm that constructs, from a given rational function \( G(z) \), a discrete-time Markov chain whose absorption-time distribution has \( G(z) \) as its probability generating function. MG distributions have rational z-transforms which makes it possible to employ matrix geometric techniques to solve queues involving matrix geometric distributions. As an example, Akar [19] provides a numerical method to solve a discrete
queuing system offered with MG-type arrival and service processes. Esparza [20] shows that the factorial moment distributions of MG-type distributions are also of MG-type.

It is well-known that PH-type distributions are dense in the set of non-negative distributions; any non-negative distribution can be arbitrarily well approximated by a PH-type distribution. However, constructing PH distributions for this purpose with a given order is not straightforward and various methods have been proposed in the literature. EMpht is a program for fitting phase-type distributions to data or parametric distributions. The Expected Maximization (EM)-based approach of EMpht is described in Asmussen et al. [21]. PhFit is a phase-type fitting tool described in Horváth and Telek [22]. PhFit fits arbitrary distributions or data not only by continuous but also discrete PH-type distributions. Thümmler et al. [23] present a novel approach for PH-type fitting with the EM algorithm and demonstrate improved accuracy compared with existing approaches. HyperStar is recently developed and described in Reinecke et al. [24] for the purpose of making PH-type fitting simpler and user-friendly. A method to approximate matrix-exponential distributions by Coxian distributions is proposed by He and Zhang [25]. In addition, Moment Matching (MM) techniques are available to match the moments of the approximating PH-type distribution to empirical moments or those of given distributions. Johnson and Taafe [26] present a nonlinear programming (NLP) approach to the problem of matching three moments to PH distributions by searching over two families of PH-type distributions: mixtures of two Erlang distributions and continuous Coxian distributions with real parameters. Second order acyclic CPH or DPH distributions are constructed to match the first three moments when possible, to the empirical moments in Telek and Heindl [27]. Canonical representations for discrete phase type distributions of order 2 and 3 are given by Tapp and Telek in [28] but this work does not focus on moment matching aspects of such distributions. Horváth and Telek [29] present an iterative approach to match an arbitrary number of moments with acyclic continuous PH-type (APH) distributions for which the computational complexity increases exponentially with the order of the approximating APH. Therefore, relatively high orders for the approximative APH may not be possible due to numerical issues. Phase-type approximations often require higher orders than their ME counterparts which has led to the work of Fackrell [30] in the context of ME fitting.

A similar problem to PH fitting exists in the field of system theory in the context of Model Reduction (MR). A linear shift-invariant multiple-input multiple-output (MIMO) discrete-time system has the unit sample response (or impulse response) matrix $H(k) = CA^k B$, $k \geq 1$, $H(0) = D$ for suitable matrices $A, B, C, D$ and the size of the square matrix $A$ gives the order of the discrete-time system; see Kailath [31] for the general theory of linear systems. In the case of a single input and single output (SISO), then we have the scalar unit sample response $h(k) = cA^{k-1}b$, $k \geq 1$, $h(0) = d$, for a row vector $c$, column vector $b$, and a finite square matrix $A$ whose size $n$ gives the order of the underlying system which is characterized with the quadruple $(c, A, b, d)$. On the other hand, a SISO linear time-invariant continuous-time system has the scalar impulse response $f(x) = ce^{Ax}b + d\delta(x)$, $x \geq 0$, where $\delta(\cdot)$ stands for the Dirac-delta function. This system is again characterized with the quadruple $(c, A, b, d)$ with order $n$ being the size of the matrix $A$. The transfer functions of these discrete-time and the continuous-time systems are $c(zI - A)^{-1}b + d$ and $c(sI - A)^{-1}b + d$, and their stabilities are indicated by all eigenvalues of $A$ being inside the unit circle, and the open left half of the complex plane, respectively, Kailath [31]. MR theory deals with finding reduced order system models of order $n_r < n$ if the order $n$ of the given SISO or MIMO system (discrete- or continuous-time) at hand is large; see Moore [32] and Antoulas et al. [33] for an overview of existing model reduction techniques. Model reduction methods are mainly classified into three classes of methods Gugercin and Antoulas [34]:

- **Singular Value Decomposition (SVD)-based methods:** Stability preserving, global bounds, relatively complex
- **Moment matching-based (or Krylov) methods:**
- **SVD-Krylov methods:**

SVD-based methods preserve stability when applied to originally stable systems and they provide global error bounds Moore [32]. However, they are computationally more intensive ($O(n^3)$) than those of Krylov methods ($O(n^2 n_r)$) of Bai [35] which match the so-called moments of the original transfer function.
at various selected frequencies, i.e., the $k$th moment at $\sigma \in \mathbb{C}$ is given by the $k$th derivative of the transfer function at $\sigma$. However, unlike the SVD-based methods, Krylov methods do not guarantee stability and error bounds are not provided. Recently, there has been increased interest in SVD-Krylov methods that benefit both from the stability and global error bound features of SVD-based methods and efficient numerical implementation and moment matching properties of Krylov methods; see Gugercin and Antoulas [34] and Gugercin [36]. Stability preservation and moment matching are both crucial for the purpose of fitting MG distributions and therefore we focus on SVD-Krylov methods in this paper. Moreover, not all linear systems satisfy the external positivity constraint, i.e., $h(k) \geq 0$ in discrete-time or $f(x) \geq 0$ in continuous-time, as would be in the case of MG or ME distributions; see Grussler [37]. Model reduction techniques generally fail to produce an externally positive reduced order model despite a start from an externally positive high order original system. For positivity preserving model reduction techniques for continuous-time systems, we refer the reader to Li et al. [38] and Grussler [37].

In this paper, we aim to fit an MG distribution of given order to data or we approximate an arbitrary discrete distribution by MG. While most existing fitting procedures involve CPH, DPH, and ME, the work on MG modeling is not as mature to the best of our knowledge. As opposed to EM or MM-based prevailing methods, we propose to use MR methods, in particular the SVD-Krylov method proposed by Gugercin and Antoulas [34]. The reasons for this choice are described below:

(a) The method of Gugercin and Antoulas [34] is specifically developed for discrete-time systems as opposed to the majority of the studies that concentrates on continuous-time systems,

(b) Being an SVD-Krylov method, the work in [34] matches $n_r$ desired moments with a reduced model of order $n_r$ and preserve stability, both of which are very critical for fitting MG-distributions. This is in contrast with pure Krylov methods that can match $2n_r$ moments but do not guarantee stability.

However, the method [34] alone does not guarantee external positivity which is also very critical in the MG-fitting setting. For the purpose of satisfying external positivity, we propose to use a mixture of the original distribution with a low-order model obtained by a particular MM-based method by Telek and Heindl [27] that obtains a second order DPH while matching the first three moments when feasible; note that the first moment is always matchable in [27]. Then, the MR-method is allowed to apply to this mixture rather than the original distribution. Subsequently, an iterative numerical algorithm is presented to find the best such mixture (in terms of the $l_2$ distance between the original distribution and that of its reduced order model) that produces an externally positive reduced order model of order $n_r$. Also note that since the moments of a mixture is obtained through the mixture of moments, any mixture (including the best one) matches the first three moments provided that [27] can match them. The reason for choosing the model produced by [27] in this mixture lies in its simplicity and the explicit moment matching algorithm provided in it. Other low-order DPHs (of order 3 for example) obtained by existing methods can also be used in the proposed mixture, which we leave outside the scope of this paper. We also propose to apply pre-smoothing on the original distribution which is shown to be beneficial when the original distribution possesses sharp edges. The method we introduce in this paper can also be viewed as an external positivity preserving model reduction technique for discrete time systems and may potentially have applications beyond the field of applied probability. The proposed MG fitting algorithm is tested for various scenarios and promising results have been obtained.

The paper is organized as follows. Section 2 describes the MG distribution in detail. Section 3 addresses the model reduction problem and the numerical algorithm we propose. We present the numerical results in Section 4. In the final section, conclusions are given.

II. PRELIMINARIES ON DISCRETE PHASE TYPE AND MATRIX GEOMETRIC DISTRIBUTIONS

The following is based on Neuts [6], Akar [19], and Greeuw [17]. A discrete phase-type (DPH) distribution is the distribution of time until absorption in a discrete-state discrete-time Markov chain (DTMC) with $n$ transient states and one absorbing state. Let the transient states be numbered as $1, 2, \ldots, n$ and the absorbing state as $n + 1$. The one step probability transition matrix of this DTMC can then be
partitioned as
\[ P = \begin{bmatrix} T & h \\ 0 & 1 \end{bmatrix}, \]  
(1)
for an \( n \times n \) sub-stochastic matrix \( T \) and an \( n \times 1 \) vector \( h = (I - T)1 \), where \( 1 \) is a column vector of ones of appropriate size. The initial probability vector can also be partitioned as \((v, d)\) for a \( 1 \times n \) row vector \( v \) and a scalar \( d \). We say \( X \sim DPH(v, T) \) and \( X \) is known to have a Probability Mass Function (PMF) \( p_X(k), k \geq 0 \) of the form
\[ p_X(k) = P(X = k) = \begin{cases} vT^{k-1}h, & k \geq 1, \\ d, & k = 0, \end{cases} \]  
(2)
and a Probability Generating Function (PGF) \( g_X(z) \) of the form
\[ g_X(z) = E[z^{-X}] = v(zI - T)^{-1}h + d. \]  
(3)
A random variable (rv) \( X \) is said to possess a matrix geometric (MG) distribution if the PMF is of the same form (2) but its parameters \( v, T, \) and \( h \), do not necessarily have the same probabilistic interpretation [20]. In this case, we say \( X \sim MG(v, T, h, d) \) and the size of the matrix \( T \) is called the order of the MG distribution. This quadruple representation is said to be irreducible if one cannot find another quadruple with lesser order satisfying (2). Actually, an MG distribution has infinitely many quadruple representations using a similarity transformation [17]. Clearly, DPH distributions form a subset of MG distributions. The \( i \)th factorial moment \( f_X^{(i)}, i \geq 1 \) of an MG-distributed rv \( X \) is given by [20]:
\[ f_X^{(i)} = E[X(X - 1) \cdots (X - i + 1)] = i!v(I - T)^{-i-1}T^{i-1}h. \]  
(4)
By definition, the zeroth factorial moment \( f_X^{(0)} = v(I - T)^{-1}h + d = 1 \). The ordinary \( i \)th moment \( m_X^{(i)} = E[X^i], i \geq 1 \) can be derived from the factorial moments \( f_X^{(i)}, 1 \leq j \leq i \) and the zeroth moment \( m_X^{(0)} = f_X^{(0)} = 1 \).

Let \( X_A \sim MG(v_A, T_A, h_A, d_A) \) and \( X_B \sim MG(v_B, T_B, h_B, d_B) \) be independent and let \( X \) be an \( \alpha \)-mixture of the ordered pair \((X_A, X_B)\), i.e., \( p_X(k) = \alpha p_{X_A}(k) + (1 - \alpha)p_{X_B}(k), k \geq 0, 0 < \alpha < 1 \). In this case, we have
\[ X \sim MG \left( \begin{bmatrix} \alpha v_A & (1 - \alpha)v_B \end{bmatrix}, \begin{bmatrix} T_A & 0 \\ 0 & T_B \end{bmatrix}, \begin{bmatrix} h_A \\ h_B \end{bmatrix}, \alpha d_A + (1 - \alpha)d_B \right). \]

The parameter \( \alpha \) is called the mixing coefficient. Moreover, the factorial moments of \( X \) can be written as
\[ f_X^{(i)} = \alpha f_{X_A}^{(i)} + (1 - \alpha)f_{X_B}^{(i)}, i \geq 0. \]
Also, let \( Y = X_A + X_B \). Consequently, \( Y \) has a matrix geometric representation characterized with the quadruple:
\[ Y \sim MG \left( \begin{bmatrix} d_Bv_A & v_B \\ h_Bv_A & T_B \end{bmatrix}, \begin{bmatrix} h_A \\ h_Bd_A \end{bmatrix}, d_Ad_B \right). \]

Let us now try to quantify if the PMFs \( p_{X_A} \) and \( p_{X_B} \) are close to each other. For the purpose of quantifying proximity (or distance) between two distributions, we propose to use the \( l_2 \) (Euclidian) norm of \( (p_{X_A} - p_{X_B}) \):
\[ dist(X_A, X_B) = \|p_{X_A} - p_{X_B}\|_2 = \left( \sum_{k=0}^{\infty} (p_{X_A}(k) - p_{X_B}(k))^2 \right)^{1/2} \]
as the \( l_2 \) distance measure between the two PMFs \( X_A \) and \( X_B \). Let \( X \) be the original rv and \( X_r \) denote the rv corresponding to its reduced order model. In this case, the distance \( dist(X, X_r) \) represents the
(modeling) error. For other distance measures between two PMFs used in the literature, we refer the reader to a comprehensive survey on distance measures [39].

Several well-known discrete distributions will be used in the numerical examples throughout the paper. The rv $X$ is said to have a discrete uniform distribution characterized with parameter pair $(u, w)$, i.e., $X \sim Unif(u, w)$ when

$$p_X(k) = \frac{1}{(w - u + 1)}; u \leq k \leq w, 0 \leq u \leq w,$$

and zero otherwise. The rv $X$ is said to possess a discrete triangular distribution (see Kokonendji et al. [40]) characterized with the parameter pair $(u, w)$, i.e., $X \sim Tri(u, w)$ when

$$p_X(k) = \frac{w + 1 - |k - u|}{(w + 1)^2}, u - w \leq k \leq u + w, u \geq w \geq 0,$$

and zero otherwise. In this case, $u$ is the called the center parameter and $w$ is the arm parameter of the corresponding distribution. The rv $X$ is Binomial distributed with parameter pair $(N, p)$, i.e., $X \sim Bin(N, p)$, where $g_X(z) = (1 - p + pz^{-1})^N, N \geq 1, 0 < p < 1$. The rv $X$ is geometrically distributed with the parameter $p$, i.e., $X \sim Geom(p)$, when

$$p_X(k) = (1 - p)^{k-1}p, k \geq 1,$$

with $E[X] = 1/p$.

III. FITTING MATRIX GEOMETRIC DISTRIBUTIONS

The problem studied in this paper is to find an integer-valued non-negative rv $X_r \sim MG(v_r, T_r, h_r, d_r)$ of reduced order $n_r$ that mimics, in distribution, an original rv $X \sim MG(v, T, h, d)$ with order $n$ such that $n_r < n$. Typically, $n_r << n$ and we refer to this problem as model reduction. This particular problem arises in the following representative scenarios:

a) The original MG-distribution order is undesirably high and order reduction is necessary,

b) The original distribution is obtained from observed data. For this purpose, let us assume $p_X(k), 0 \leq k \leq K$ is available for some $K$ which is the largest observed data. Let $u_X(k) = 1 - \sum_{i=1}^{k} \frac{p_X(i)}{1 - p_X(0)}, 1 \leq k \leq K$ and $u_X(0) = 1$. Also let $v_X(k) = u_X(k)/u_X(k - 1), 1 \leq k \leq K$. Based on the results presented in Alfa [41], one can show that $X \sim DPH(v, T)$ with order $K$ where

$$v(1) = 1 - p_X(0), v(j) = 0, j \neq 1, T_{j,j+1} = v_X(j), 1 \leq j < K, T_{i,j} = 0, j \neq i + 1.$$

c) The original distribution is not of MG-type but can well be approximated by an MG distribution with potentially large orders. For this purpose, let us assume the existence of an integer $K$ such that $P(X > K) = \varepsilon$ for negligibly small $\varepsilon$. One can then approximate the rv $X$ by $\tilde{X}$ whose PMF can be given by

$$p_{\tilde{X}}(k) = \begin{cases} 
  p_X(k) + \varepsilon & \text{if } k = K, \\
  p_X(k) & \text{if } 0 \leq k < K, \\
  0 & \text{if } k > K.
\end{cases} \quad (5)$$

This method is called truncation. Since the order $K$ will generally be large, there may be a need for model reduction.

For the purpose of model reduction, the following goals are generally set:

i) It should hold that $m_X^{(0)} = m_{\tilde{X}}^{(0)} = 1$.

ii) The $I \geq 1$ moments of the original distribution and those of the reduced model are the same, i.e.,

$$m_X^{(i)} = m_{\tilde{X}}^{(i)} = 1 \leq i \leq I.$$

iii) We attempt to reduce the distance between the original PMF and that of its reduced order model where a distance measure given in [39] can be used for this purpose. In this paper, we use the $l_2$ distance between two PMFs in the numerical examples.

We first describe two approaches from the existing literature to partially attain these goals and then propose two new algorithms for further improvement.
A. Moment Matching Using a DPH of Second Order

In Telek and Heindl [27], a second order DPH is constructed that attempts to fit the first \( I = 3 \) moments of an original distribution. The DPH is assumed to be acyclic, hence named ADPH. In this case, the sub-stochastic matrix component is upper triangular, and moreover there is no probability mass at zero. Telek and Heindl [27] provide permissible ranges for the factorial moments within which it is feasible to construct a second order DPH that exactly matches the first three moments. In case this problem is not feasible, an algorithm (referred to as ADPH2 in the current article) is provided that suitably adjusts the original second and third factorial moments to be matched so that the problem becomes feasible. With the adjusted moments, one can construct an ADPH of second order that always matches the first moment, and approximately matches the other two. For details on ADPH2, the reader is referred to [27].

The ADPH2 algorithm assumes \( p_X(0) = 0 \). Let \( p_X(k), k \geq 0 \) be the PMF of the rv \( X \) with \( p_X(0) \neq 0 \). Let \( \tilde{X} \) with PMF \( \tilde{p}_X(k) \) be such that \( \tilde{p}_X(0) = 0 \) and \( \tilde{p}_X(k) = p_X(k)/(1 - p_X(0)), k \geq 1 \). Since \( \tilde{X} \) does not have a probability mass at zero, one can employ ADPH2 on \( \tilde{X} \) to construct a second order reduced model \( \tilde{X}_r \sim DPH(\tilde{v}_r, \tilde{T}_r) \) for \( \tilde{X} \). Consequently, \( X_r \sim DPH((1 - p_X(0))\tilde{v}_r, \tilde{T}_r) \) yields a second order model for the original rv \( X \) while fitting the non-zero probability mass at zero. This is the approach we will take in the current paper while still referring to it as ADPH2.

B. Moment Matching with Least Squares (MMLS)

The model reduction techniques of system theory attempt to reduce the model order of a given linear shift-invariant dynamical system as opposed to a distribution; see Moore [32] and Antoulas et al. [33]. A finite-dimensional shift-invariant discrete-time dynamical system is characterized with the quadruple \((v, T, h, d)\) with order being the size of the matrix \( T \) if the unit sample response of the system is in the form (2) but does not necessarily satisfy the conditions for being a PMF. The reader is referred to Kailath [31] and Chen [42] for linear dynamical systems and their state-space representations, i.e., quadruple representations.

We now describe the particular model reduction method for discrete-time systems by least squares based on the work of Gugercin and Antoulas [34] which will also be the basic building block of our proposed approach. Although, any set of complex numbers can be chosen for transfer function moment matching, we limit ourselves to the two values \( \sigma = 0 \) and \( \sigma \rightarrow \infty \) only. Let the original discrete-time rv \( X \sim MG(v, T, h, d) \) with order \( n \). We are interested in finding \( X_r \sim MG(v_r, T_r, h_r, d_r) \) with order \( n_r < n \). Let \( n_r^{(1)} \) and \( n_r^{(2)} \) be such that \( n_r = n_r^{(1)} + n_r^{(2)}, n_r^{(2)} \geq 1 \). We then define the \( n \times n_r \) matrix \( Q \) whose column space denoted by \( C(Q) \) is given by

\[
C(Q) = C([h, Th, T^2h, \ldots, T^{n_r^{(1)}-1}h, (I - T)^{-1}h, (I - T)^{-2}h, \ldots, (I - T)^{-n_r^{(2)}h}]). \tag{6}
\]

Moreover, let \( O \) be the solution to the Stein equation, also referred to as a discrete Lyapunov equation in Benner at al. [43]:

\[
T^T O T + v^T v = O. \tag{7}
\]

Actually, \( O \) is the observability gramian of the dynamical system characterized with \((v, T, h, d)\). We also define

\[
Z^T = (Q^T O Q)^{-1} Q^T O. \tag{8}
\]

Based on Gugercin and Antoulas [34], the representation \( X_r \sim MG(v_r, T_r, h_r, d_r) = MG(vQ, Z^T TQ, Z^T h, d) \) provides a reduced order matrix geometric representation with order \( n_r \). However, the proposed representation is not guaranteed to correspond to an actual PMF. Actually, it is quite possible to have an integer \( j > 0 \) such that

\[
v_r T_r^{j-1} h_r < 0.
\]

The reduced order model is known to have the following properties; see [34]:
Algorithm 1 The MMLLS method

1: function MMLLS($X \sim MG(v, T, h, d), n^{(1)}_r, n^{(2)}_r$)
2: Define the matrix $H$ as
   \[ H = \begin{bmatrix} h & Th & T^2h & \cdots & T^{n^{(1)}_r-1}h & (I - T)^{-1}h & (I - T)^{-2}h & \cdots & (I - T)^{-n^{(2)}_r}h \end{bmatrix} \]
3: Find the QR decomposition (see [44]) of the matrix $H$, i.e., $H = QR$, where $Q$ is orthogonal and $R$ is upper-triangular. In Matlab, we propose to use $[Q, R] = qr(H)$.
4: Solve the discrete Lyapunov equation given in (7) for the matrix $O$. In Matlab, we propose to use $O = dlyap(T^T, v^Tv)$.
5: Define the matrix $Z$ as in identity (8).
6: $v_r \leftarrow vQ, T_r \leftarrow Z^TQ, h_r \leftarrow Z^Th, d_r \leftarrow d$
7: Return $X_r \sim MG(v_r, T_r, h_r, d_r)$ and the error $e \leftarrow ||p_X - p_{X_r}||_2$
8: end function

i) The first $n^{(1)}_r + 1$ PMF values (starting from zero) are matched, i.e.,
   \[ p_{X_r}(i) = p_X(i), i = 0, 1, \ldots, n^{(1)}_r. \]

ii) The first $n^{(2)}_r$ moments (starting from zero) are matched, i.e.,
   \[ m^{(i)}_{X_r} = m^{(i)}_X, i = 0, 1, \ldots, n^{(2)}_r - 1, \]
   or equivalently the first $n^{(2)}_r$ factorial moments (starting from zero) are matched, i.e.,
   \[ f^{(i)}_{X_r} = f^{(i)}_X, i = 0, 1, \ldots, n^{(2)}_r - 1. \]

iii) The matrix $T_r$ has all eigenvalues inside the unit circle and $p_{X_r}(k) = v_r T_r^{k-1}h_r$ does not grow without bounds as $k \to \infty$.

iv) This model minimizes a certain $l_2$ error between the original and reduced order systems when $n_r = n^{(1)}_r$ [34]. In this case, the model has been shown in [34] to minimize $||\hat{a}_k \ast (p_X(k) - p_{X_r}(k))||_2$ where $\ast$ denotes the convolution operator and $\hat{a}_k = 0$ for $k > n_r$ and the coefficients $\hat{a}_j, j \leq n_r$ correspond to the coefficients of the denominator of $A(z) = v_r(zI - T_r)^{-1}h_r + d_r$. For a proof of this statement and for details, the reader is referred to Gugercin and Antoulas [34].

v) The procedure does not require the explicit computation of any of the moments.

vi) There is no guarantee that the reduced order model corresponds to a distribution even when the original model does. This arises especially when the original PMF vanishes at a certain number of points and there are sharp edges around these points, as will be shown in the numerical examples.

Since the method has moment matching (properties i and ii) and least squares minimization (property iv) features, we call the method MMLLS (Moment Matching with Least Squares) (referred to similarly in the original paper [34]) throughout this article and is presented in Algorithm 1. In Algorithm 1, the image of $H$ spans a Krylov subspace and it can be obtained with a computational complexity of $O(n^2 n_r)$ [34]. The observability gramian $O$ is not only computationally more intensive, i.e., $O(n^3)$, but computing a full-rank $O$ is known to be ill-conditioned in large-scale settings Penzl [45], Penzl [46], Gugercin and Antoulas [34]. Although there are low-rank Smith-type methods to produce low-rank approximations to the full rank $O$ with improved numerical stability such as Penzl [45], Gugercin et al. [47], we calculate the full rank $O$ in this paper via the discrete Lyapunov equation (7).

The method we propose that combines ADPH2 and MMLLS in order to obtain a reduced order distribution is described next.
C. Iterative Moment Matching with Least Squares (IMMLS)

Recall that we are interested in finding a rv $X_r \sim MG(v_r, T_r, h_r, d_r)$ of reduced order $n_r$ given an original rv $X \sim MG(v, T, h, d)$ with order $n$ such that $2 < n_r < n$. If $n_r = 2$, we propose to use the already existing ADPH2 method. Let $X_2 \sim (v_2, T_2, h_2, d_2)$ be the matrix geometric reduced order distribution found using ADPH2. Let $X^{(\alpha)}$ be the $\alpha$-mixture of the ordered pair $(X, X_2)$ and let $X^{(\alpha)}_r \sim MG(v^{(\alpha)}_r, T^{(\alpha)}_r, h^{(\alpha)}_r, d^{(\alpha)}_r)$ be the reduced order model of $X^{(\alpha)}$ with order $n_r$, using the MMLS algorithm for $0 < \alpha < 1$. We have two observations: Stemming from property iv, as $\alpha \to 0$, then $X^{(\alpha)}_r \overset{d}{\to} X_2$ where $\overset{d}{\to}$ denotes convergence in distribution. Therefore, we are guaranteed to have a valid reduced order model corresponding to a distribution as $\alpha \to 0$. On the other hand, as $\alpha \to 1$, we take advantage of moment matching and least squares features of the MMLS algorithm but we are not guaranteed to have a valid distribution. In this article, we propose to iteratively search for the largest feasible $\alpha$ in the interval $(0, 1)$ such that $MG(v^{(\alpha)}_r, T^{(\alpha)}_r, h^{(\alpha)}_r, d^{(\alpha)}_r)$ corresponds to a legitimate distribution unless $MG(v^{(1)}_r, T^{(1)}_r, h^{(1)}_r, d^{(1)}_r)$ produces a legitimate distribution. In the latter case, there is no need for a search. In case needed, the iterative search is done in the following manner. In the first step, we fix $\alpha$ to zero and set $\beta$ to a random number uniformly distributed in the interval $(\alpha, 1)$ and check if $MG(v^{(\beta)}_r, T^{(\beta)}_r, h^{(\beta)}_r, d^{(\beta)}_r)$ corresponds to a legitimate distribution. Then, we set $\alpha = \beta$ if a legitimate distribution is obtained with lower error. Otherwise $\alpha$ remains intact. This procedure is repeated for at most $S$ steps which is called the search depth of the procedure. Note that we can not use the binary search algorithm in place of randomized search since it is generally not true that there is a particular value of $\alpha$ below which all produced reduced order models are legitimate and others not.

To test whether a quadruple representation leads to a legitimate distribution (which is essential to the search procedure), we propose to check if the following holds for a given arbitrarily large integer $M_d$:

$$D_{X_r} = \sum_{i=0}^{M_d} p_{X_r}(i) I_{X_r}(i) \geq -\eta_d$$

(9)

for some very small $\eta_d \geq 0$ and the indicator function $I_{X_r}(i) = 1$ when $p_{X_r}(i) < 0$ and $I_{X_r}(i) = 0$ otherwise. Note that the deviation parameter $D_{X_r}$ provides a quantitative measure of how much $X_r$ deviates from being an actual rv. When condition (9) is satisfied, then we say $X_r$ has a legitimate distribution. When $\eta_d = 0$ and as $M_d \to \infty$, this test is exact. Obviously, the use of a finite $M_d$ reduces the computation time for validation but at the expense of a slight reduction in validation accuracy. On the other hand, the choice of a non-zero $\eta_d$ violates the definition for an exact distribution but this parameter can be used as an instrument to reduce modeling errors despite a slight deviation from being an actual distribution. In case such relaxing can not be tolerated, the parameter $\eta_d$ may be set to zero. We call this method IMMLS (Iterative MMLS) for which a pseudo-code is presented in Algorithm 2.

The mixing coefficient found using this iterative procedure is denoted by $\alpha$ and the distance between the original PMF and its reduced model is denoted by $e$ representing the $l_2$ error. If the resulting $\alpha$ is close to unity, then the reduced order model is dominated by the MMLS algorithm and we benefit substantially from its moment matching and least squares minimization capabilities. For example, if $\alpha = 1$, then the first $n_r^{(1)} + 1$ PMF values and the first $n_r^{(2)} - 1$ ordinary moments are matched to those of the original distribution. The least-squares property of the MMLS algorithm is additionally effective in matching the original distribution. Note that for all values of the mixing coefficient $\alpha$, if ADPH2 matches the first three moments, then the combined algorithm IMMLS using ADPH2 and MMLS also produces a reduced order distribution which also matches the first three moments but also makes an attempt to approximately match higher order moments with increased values of $\alpha$. With this matrix-analytical methodology, it is quite possible to improve ADPH2 without sacrificing its moment matching capability.

We present two numerical examples to demonstrate the operational principles of IMMLS while setting the algorithm parameters $S = 50$, $\eta_d = 10^{-6}$, and $M_d = 5000$, and the reduced order model parameters $n_r^{(1)} = 4$ and $n_r^{(2)} = 6$ in both examples. In the first example, $X \sim Tri(100, 40)$ and we run ADPH2,
Algorithm 2 The IMMLS method

1: function IMMLS$(X \sim MG(v, T, h, d), n_r^{(1)}, n_r^{(2)}, \eta_d, M_d, S)$
2:    Obtain $X_r \sim MG(v_r, T_r, h_r, d_r) \leftarrow$ MMLS$(X \sim MG(v, T, h, d), n_r^{(1)}, n_r^{(2)})$
3:    $\alpha \leftarrow 1$, $e \leftarrow \|p_X - p_{X_r}\|_2$
4:    if $D_{X_r} \geq -\eta_d$ then
5:        Goto Step 18
6:    else
7:        Obtain $X_2 \sim (v_2, T_2, h_2, d_2)$ using the ADPH2 method.
8:        $\alpha \leftarrow 0$, $X_r \leftarrow X_2$, $e \leftarrow \|p_X - p_{X_2}\|_2$
9:        for $i \leftarrow 1, S$ do
10:           Pick $\beta$ uniformly distributed in $(\alpha, 1)$
11:           Obtain $X^{(\beta)}$ as the $\beta$-mixture of the ordered pair $(X, X_2)$
12:           Find $Y \sim MG(v_0, T_0, h_0, d_0) \leftarrow$ MMLS$(X^{(\beta)}, n_r^{(1)}, n_r^{(2)})$ and calculate its error
13:           if $D_Y \geq -\eta_d$ and error $< e$ then
14:              $\alpha \leftarrow \beta$, $X_r \leftarrow Y$, $e \leftarrow \|p_X - p_Y\|_2$
15:           end if
16:        end for
17:    end if
18:    Return the mixing coefficient $\alpha$, $X_r \sim MG(v_r, T_r, h_r, d_r)$, and the error $e$.
19: end function

MMLS, and IMMLS algorithms using MATLAB. With IMMLS, we iteratively find the mixing coefficient $\alpha$ as 0.7913. The relevant performance metrics, in particular, the first $n_r^{(1)} + 1$ values of the PMF, namely $p_{X_r}(k), 0 \leq k \leq 4$, and the first $n_r^{(2)}$ factorial moments (starting from the zeroth moment) $f_{X_r}^{(i)}, 0 \leq i \leq 5$, are tabulated in Table I for all the three algorithms as well as the corresponding values for the original PMF. We repeat the same experiment in the second example with $X \sim Unif(90, 110)$ in which case IMMLS finds the mixing coefficient $\alpha$ to be 0.3821. The original PMF as well as the PMFs obtained by ADPH2, MMLS, and IMMLS algorithms are depicted in Fig. 2, whereas the associated performance metrics are presented in Table II. The machine epsilon is 2.22e-16 in MATLAB and any value less than the machine epsilon in absolute value is reported as zero in Tables I and II. We have the following observations:

(i) ADPH2 matches only the first moment but approximately matches the second and third moments in both examples. Actually, ADPH2 algorithm produced the same model for both examples due to the fact that the second and third moments could not be matched by ADPH2 and they are adjusted to the same values in both examples.

(ii) MMLS matches the first $n_r^{(1)} + 1$ PMF values as well as the the first $n_r^{(2)} - 1$ ordinary moments. However, MMLS falls short of providing a legitimate PMF in both examples.

(iii) IMMLS, as conjectured, provides a legitimate PMF (note the value of $D_{X_r}$) and its PMF and moment matching capability stand between that of ADPH2 and MMLS.

(iv) The mixing coefficient of IMMLS is smaller and consequently the error with IMMLS is relatively higher in the second numerical example that has sharp edges in the original distribution in comparison with the first numerical example. When there are sharp edges in the PMF of the original random variable, the absolute value of the deviation parameter of the MMLS solution increases since the MMLS method presents oscillatory behavior to capture the sharp edges. When sharp edges are around zero, such oscillations violate positivity. Subsequently, the IMMLS solution presents larger errors for the second example.
Fig. 1. The original PMF and the reduced order model PMFs with ADPH2, MMLS, and IMMLS when $X_r \sim Tri(100, 40)$ and $n_r^{(1)} = 4$ and $n_r^{(2)} = 6$.

<table>
<thead>
<tr>
<th>$p_{X_r}(0)$</th>
<th>ADPH2</th>
<th>MMLS</th>
<th>IMMLS</th>
<th>Original PMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{X_r}(1)$</td>
<td>0</td>
<td>4.4210e-16</td>
<td>8.0421e-16</td>
<td>0</td>
</tr>
<tr>
<td>$p_{X_r}(2)$</td>
<td>4.0000e-04</td>
<td>8.3494e-05</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$p_{X_r}(3)$</td>
<td>7.8400e-04</td>
<td>1.6365e-04</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$p_{X_r}(4)$</td>
<td>1.1525e-03</td>
<td>2.4056e-04</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

| $f_{X_r}^{(0)}$ | 1.0000e+00 | 1.0000e+00 | 1.0000e+00 | 1.0000e+00 |
| $f_{X_r}^{(1)}$ | 1.0000e+02 | 1.0000e+02 | 1.0000e+02 | 1.0000e+02 |
| $f_{X_r}^{(2)}$ | 1.4800e+04 | 1.1144e+04 | 1.0180e+04 | 1.0180e+04 |
| $f_{X_r}^{(3)}$ | 2.9106e+06 | 1.4410e+06 | 1.0534e+06 | 1.0534e+06 |
| $f_{X_r}^{(4)}$ | 7.1454e+08 | 2.3666e+08 | 1.1060e+08 | 1.1060e+08 |
| $f_{X_r}^{(5)}$ | 2.1036e+11 | 5.3215e+10 | 1.1762e+10 | 1.1762e+10 |

Table I: Several performance metrics obtained by the ADPH2, MMLS, and IMMLS algorithms when run with the test example $X \sim Tri(100, 40)$.

error $||p_X - p_{X_r}||_2$ | 1.0195e-01 | 8.9433e-03 | 2.3652e-02 | N/A |
deviation $D_{X_r}$ | N/A | -1.7687e-02 | 0 | N/A |

Fig. 2. The original PMF and the reduced order model PMFs with ADPH2, MMLS, and IMMLS when $X_r \sim Unif(90, 110)$ and $n_r^{(1)} = 4$ and $n_r^{(2)} = 6$. 
TABLE II
SEVERAL PERFORMANCE METRICS OBTAINED BY THE ADPH2, MMLS, AND IMMLS ALGORITHMS WHEN RUN WITH THE TEST EXAMPLE \(X \sim \text{Unif}(90,110)\).

<table>
<thead>
<tr>
<th></th>
<th>ADPH2</th>
<th>MMLS</th>
<th>IMMLS</th>
<th>Original PMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{X}(0))</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(p_{X}(1))</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(p_{X}(2))</td>
<td>4.0000e-04</td>
<td>0</td>
<td>2.4715e-04</td>
<td>0</td>
</tr>
<tr>
<td>(p_{X}(3))</td>
<td>7.8400e-04</td>
<td>0</td>
<td>4.8440e-04</td>
<td>0</td>
</tr>
<tr>
<td>(p_{X}(4))</td>
<td>1.1525e-03</td>
<td>0</td>
<td>7.1207e-04</td>
<td>0</td>
</tr>
<tr>
<td>(f_{X}^{(1)})</td>
<td>1.0000e+00</td>
<td>1.0000e+00</td>
<td>1.0000e+00</td>
<td>1.0000e+00</td>
</tr>
<tr>
<td>(f_{X}^{(2)})</td>
<td>1.0000e+02</td>
<td>1.0000e+02</td>
<td>1.0000e+02</td>
<td>1.0000e+02</td>
</tr>
<tr>
<td>(f_{X}^{(3)})</td>
<td>1.4800e+04</td>
<td>9.9367e+03</td>
<td>1.2942e+04</td>
<td>9.9367e+03</td>
</tr>
<tr>
<td>(f_{X}^{(4)})</td>
<td>2.9106e+06</td>
<td>9.8109e+05</td>
<td>2.1733e+06</td>
<td>9.8109e+05</td>
</tr>
<tr>
<td>(f_{X}^{(5)})</td>
<td>7.1454e+08</td>
<td>9.6246e+07</td>
<td>4.7827e+08</td>
<td>9.6246e+07</td>
</tr>
<tr>
<td>error (</td>
<td>p_X - p_{X'}</td>
<td>|_2)</td>
<td>2.0429e-01</td>
<td>1.2739e-01</td>
</tr>
<tr>
<td>deviation (D_{X'})</td>
<td>N/A</td>
<td>-2.7899e-01</td>
<td>-4.3728e-08</td>
<td>N/A</td>
</tr>
</tbody>
</table>

D. Iterative Moment Matching with Least Squares and Smoothing (IMMLSS)

Since sharp edges in the original PMF are problematic for IMMLS (as demonstrated in the second numerical example of the previous subsection), there is an apparent need for performance improvement in obtaining reduced order models when sharp edges are present. In this article, we propose to smooth edges of the original PMF in the pre-processing stage and then apply IMMLS on the smoothed model. For smoothing purposes, we use in this article the two-sided rv \(Z^{[f]}\) with \(p_{Z^{[f]}}(k) = \frac{1}{2f+1}, -f \leq k \leq f\), and then use the operation \(X^{[f]} = \max(0, X + Z^{[f]}\) which amounts to smoothing \(g_{X}(z)\) with a smoothing filter with windowing parameter \(f\). Moreover, the max operator in the above operation ensures that the rv \(X^{[f]}\) is non-negative. Other possible smoothing operations that can also be used for edge smoothing are left outside the scope of this paper. If \(X \sim MG(v, T, h, d)\), then \(X^{[f]} = \max(0, X + Z^{[f]}\) is again of MG-type. Let \(Y^{[f]} \sim \text{Unif}(0, 2f)\). Then \(X + Y^{[f]} \sim MG(c, A, b, g)\) for some quadruple \((c, A, b, g)\) since the sum of MGs is also an MG (see Section 1). Finally, \(X^{[f]} \sim MG(cA^{f}, A, b, g + \sum_{i=1}^{f} cA^{i}b)\). However, using a large smoothing parameter \(f\) also distorts the original PMF and may also lead to increased errors. For this purpose, we propose an exhaustive search algorithm that applies smoothers with different smoothing parameters on the original PMF and finds a reduced model for each smoothed PMF. One of the reduced models that is closest to (in the \(l_2\) sense) to the original PMF is then chosen. As an illustrative example, we use the second test example \((X \sim \text{Unif}(90,110))\) of the previous subsection and employ PMF pre-smoothing with varying parameter \(f\). The reduced order model’s error is depicted as a function of parameter \(f\) for four different choices of the parameter pairs \((n_r^{(1)}, n_r^{(2)})\) in Fig. 3a. Subsequently, for each of the four \((n_r^{(1)}, n_r^{(2)})\) scenarios, the value of \(f\) is chosen so as to minimize the modeling error. The PMFs obtained with this approach are depicted in Fig. 3b. We observe that smoothing is helpful in terms of the reduction of the modeling error. The optimal value of the smoothing parameter appears to depend on the model size. Moreover, as far as minimization of error is concerned, it is more desirable to match as many moments as possible as opposed to matching the PMF values in the vicinity of zero.

Let \(\mathcal{F}\) be a set of smoothing parameters and we present the pseudo-code for IMMLSS (IMMLS and Smoothing) in Algorithm 3 with the set \(\mathcal{F}\) being input to the function IMMLSS. In Algorithm 3, when IMMLSS calls IMMLS, the error of using the model \(X_{r}^{[f]}\) is always calculated by IMMLSS as \(||p_X - p_{X'}||_2\) as opposed to \(||p_X - p_{X'}||_2\).
Smoothing parameter $f$

error $||p^X - p^X_r||_2$

Fig. 3. a) The error as a function of the smoothing parameter $f$

b) The PMF of the reduced order model, both for four different pairs of values for $n_r^{(1)}$ and $n_r^{(2)}$.

Algorithm 3 The IMMLSS method

1: function IMMLSS($X \sim MG(v, T, h, d), n_r^{(1)}, n_r^{(2)}, \eta_d, M_d, S, F)$
2:    while $f \in F$ do
3:        Smooth $X$ with a smoothing filter with parameter $f$ to obtain $X^{[f]} \sim MG(v^{[f]}, T^{[f]}, h^{[f]}, d^{[f]})$
4:        Obtain $X_r^{[f]} \sim MG(v_r^{[f]}, T_r^{[f]}, h_r^{[f]}, d_r^{[f]})$, the mixing coefficient $\alpha^{[f]} \leftarrow$ IMMLS($X^{[f]}, n_r^{(1)}, n_r^{(2)}, \eta_d, M_d, \epsilon, S$)
5:        $e^{[f]} \leftarrow ||p^X - p^X_r^{[f]}||_2$
6:    end while
7:    $f^* \leftarrow \arg\min_f \{e^{[f]}, f \in F\}$
8:    Return $X_r \sim MG(v_r^{[f^*]}, T_r^{[f^*]}, h_r^{[f^*]}, d_r^{[f^*]})$, the error $e^* \leftarrow e^{[f^*]}$, and the mixing coefficient $\alpha^* \leftarrow \alpha^{[f^*]}$
9: end function

IV. Numerical Examples

We provide numerical results using IMMLSS with the purpose of constructing reduced order MG distributions of a given order $n_r$ given higher order original distributions. Let $X$ denote the original rv with PMF $p_X(k), k \geq 0$. The IMMLSS algorithm parameters we use for all numerical examples are presented in Table III. We test the IMMLSS algorithm in the following five numerical examples:

<table>
<thead>
<tr>
<th>Algorithm parameters used for IMMLSS in all numerical examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_d$</td>
</tr>
<tr>
<td>$\eta_d$</td>
</tr>
<tr>
<td>$F$</td>
</tr>
<tr>
<td>$\epsilon$</td>
</tr>
<tr>
<td>$n_r^{(1)}$</td>
</tr>
<tr>
<td>$S$</td>
</tr>
</tbody>
</table>
A. Example 1

Let $X_i, i = 1, 2, 3$ be Binomial distributed with parameter pair $(N_i, p_i)$, i.e., $X_i \sim Bin(N_i, p_i)$ and let $X_i$s be independent. Also let $X$ be a tri-mixture of $X_i$s, i.e., $p_X(k) = \sum_{i=1}^{3} \gamma_i p_{X_i}(k)$, $k \geq 0, \gamma_i \geq 0, \sum_{i=1}^{3} \gamma_i = 1$. We set $N_1 = N_2 = 1000, N_3 = 3000, \gamma_1 = 0.5, \gamma_2 = \gamma_3 = 0.25, p_1 = 0.18, p_2 = 0.24, p_3 = 0.04$. For this example, $E[X] = 180$. In this scenario, we first approximated the original distribution of $X$ by that of $\tilde{X}$ which is obtained through $X$ by truncation with truncation parameter $\varepsilon = 10^{-13}$ as in Eqn. (5) which yields $n = 341$ for the order of the MG representation for $\tilde{X}$. We then run IMMLSS operating on $\tilde{X}$. The PMFs obtained with IMMLSS are depicted in Fig. 4. We observe that IMMLSS matches the entire PMF with increased model order $n_r$ with high mixing coefficients, low smoothing parameter, and very low error.

B. Example 2

In Example 2a, $X_i \sim Unif(u_i, w_i), i = 1, 2$, $X_i$s are independent, and $X$ is a bi-mixture of $X_i$s, i.e., $p_X(k) = \sum_{i=1}^{2} \gamma_i p_{X_i}(k)$, $k \geq 0, \gamma_i \geq 0, \sum_{i=1}^{2} \gamma_i = 1$. We set $u_1 = 40, w_1 = 60, u_2 = 160, w_2 = 240, \gamma_1 = \frac{1}{2}$ which yields $E[X] = 125$ and $n = 240$. Example 2b is the same as the previous example with $u_1 = 115, w_1 = 135, u_2 = 85, w_2 = 165$. As in Example 2a, $E[X] = 125$. The PMFs obtained by IMMLSS for Examples 2a and 2b are depicted in figs. 5 and 6, respectively. Compared to the previous example, the smoothing parameter $f^*$ increased in both examples with much larger errors stemming from sharp edges of the original PMF especially around the horizontal axis. Edges that are far from the horizontal axis are captured better as in Example 2b in comparison with Example 2a; note the errors in both examples.
C. Example 3

Let $Y_i, i = 1, 2$ be geometric distributed with parameter $p_i$, i.e., $Y_i \sim Geom(p_i)$, where $p_Y(k) = (1 - p_i)^{k-1}p_i, 0 < p_i < 1, k \geq 1$. We assume $Y_i$s are independent. Let $X_i = Y_i + h_i$ where $h_i$ is fixed and non-negative. Let $X$ be a mixture of $X_i$s, i.e., $p_X(k) = \sum_{i=1}^{2} \gamma_i p_{X_i}(k), k \geq 0, \gamma_i \geq 0, \sum_{i=1}^{2} \gamma_i = 1$. We set $p_1 = 0.004, p_2 = 0.02, h_1 = 50, h_2 = 250, \gamma_1 = 5/6$ in Example 3a. Examples 3b and 3c are the same as Example 3a except that $h_1 = 0$ in Example 3b and $\gamma_1 = 1/3, \gamma = 2/3, p_1 = 0.04$ in Example 3c. The PMFs obtained by IMMLSS for Examples 3a, 3b, and 3c are depicted in figs. 7, 8, and 9, respectively.
Fig. 9. The original PMF and the PMFs of those obtained by IMML S for three different values of $n_r$ and the corresponding $f^*$, $\alpha^*$, $e^*$ values for Example 3c.

Fig. 10. The original PMF and the PMFs of those obtained by IMM LS for three different values of $n_r$ and the corresponding $f^*$, $\alpha^*$, $e^*$ values for Example 4.

D. Example 4

The original rv $X$ of order $n = 1100$ is obtained by Geyser data from Azzalini and Bowman [48] by windowing. Actually, $X$ represents the waiting time until the next eruption for which the time unit is set to 0.1 minutes. The PMFs obtained by IMMLSS for Example 4 are depicted in Fig. 10.

E. Example 5

The packet sizes in units of Bytes used in a high speed communications link are obtained by MAWI (Measurement and Analysis on the WIDE Internet) at samplepoint-F (a trans-Pacific link) [49]. The statistics we use are the ones collected on the first working day of 2013 at 14:00. The smallest (largest) packet size observed is 60 (1514) Bytes. The original rv $X$ representing the Internet packet sizes is obtained by smoothing original data with parameter $f = 32$ which then yields a finite support distribution for $X$ of order $n = 1546$. The PMFs obtained by IMMLSS for Example 5 are depicted in Fig. 11.

F. Observations

We have the following observations:

- Smoothing the original distribution appears to be necessary especially when the original PMF vanishes at certain points and there are abrupt jumps around these points; see for example figs. 5, 6, and 9. The requirement for smoothing diminishes for the case of abrupt jumps elsewhere; see Fig. 8.

- The error $e^*$ decreases with reduced model order $n_r$. Although the best mixing coefficient $\alpha^*$ generally appears to increase with reduced model order $n_r$, there are cases otherwise.

- The largest original MG order $n$ we tried is $n = 1546$ for Example 5. As explained before, the equation (7) is ill-conditioned for large-scale problems and for larger values of $n$ and for values of $n_r$ exceeding 64, we encountered Matlab warnings. Using low-rank approximations to the gramian matrix $O$ as part of IMMLSS in this numerically challenging-regime is left for future research.
Fig. 11. The original PMF and the PMFs of those obtained by IMM LS for three different values of $n_r$ and the corresponding $f^\ast$, $\alpha^\ast$, $e^\ast$ values for Example 4.

V. CONCLUSIONS

A novel algorithmic method is proposed to fit matrix geometric distributions. The proposed method effectively combines two existing approaches from two different disciplines, namely model reduction methods of system theory and moment matching methods of applied probability. Promising results have been obtained in several bi-modal and tri-modal scenarios and for some well-known statistical data. Improving the numerical algorithm so as to operate in more numerically-challenging regimes including large-scale problems and extension of the proposed method for fitting ME-distributions in continuous-time are considered as future work.

ACKNOWLEDGMENT

This work is supported in part by the Science and Research Council of Turkey (Tübitak) under project grant EEEAG-111E106.

REFERENCES


