Characterization of Moments and Autocorrelation in MAPs*

Giuliano Casale
College of William & Mary
Computer Science Dept.
Williamsburg, 23187, VA, US
casale@cs.wm.edu

Eddy Z. Zhang
College of William & Mary
Computer Science Dept.
Williamsburg, 23187, VA, US
eddy@cs.wm.edu

Evgenia Smirni
College of William & Mary
Computer Science Dept.
Williamsburg, 23187, VA, US
esmirni@cs.wm.edu

1. INTRODUCTION

Markovian Arrival Processes (MAPs) [9] are a general class of point processes which admits, hyper-exponential, Erlang, and Markov Modulated Poisson Processes (MMPPs) as special cases. MAPs can be easily integrated within queueing models. This makes MAPs useful for evaluating the impact of non-Poisson workloads in networking and for quantifying the performance of multi-tiered e-commerce applications and disk drives [8, 10].

In this work, we provide scalar closed-form formulas for moments and autocorrelation coefficients of general MAPs (see [7] for matrix-form computation of Markov chain descriptors). These closed-form formulas are used to define new MAPs with predefined stochastic properties.

1.1 Markovian Arrival Processes

A MAP(n) is specified by two $n \times n$ matrices, a stable matrix $\boldsymbol{D}_0$ and a nonnegative matrix $\boldsymbol{D}_1$, that describe transition rates between n states. Each transition in $\boldsymbol{D}_1$ produces a job arrival; $\boldsymbol{D}_0$ instead describes background transitions not associated with arrivals. $\boldsymbol{Q} = \boldsymbol{D}_0 + \boldsymbol{D}_1$ is the infinitesimal generator of the underlying continuous-time Markov chain. We focus on the inter-arrival (or equivalently service) time description of arrival processes. For a MAP(n), inter-arrival time moments and autocorrelations are computed using the probability vector $\boldsymbol{\pi}_e$ of the embedded process with irreducible stochastic matrix $\boldsymbol{P} = (-\boldsymbol{D}_0)^{-1}\boldsymbol{D}_1$, where $\boldsymbol{\pi}_e \boldsymbol{e} = 1$ and $\boldsymbol{e}$ is a column vector of 1's of the appropriate dimension. The MAP inter-arrival times are phase-type distributed with k-th moment

$$E[X^k] = k!\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-k}\boldsymbol{e}, \quad k \geq 0, \tag{1}$$

and squared coefficient of variation

$$\mathrm{CV}^2 = 2E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-2}\boldsymbol{e} - 1.$$

The lag-k autocorrelation of inter-arrivals is

$$\rho_k = \frac{E[X]^{-2}\boldsymbol{\pi}_e(-\boldsymbol{D}_0)^{-1}\boldsymbol{P}^k(-\boldsymbol{D}_0)^{-1}\boldsymbol{e} - 1}{\mathrm{CV}^2}, \quad k \geq 1. \tag{2}$$

Throughout the paper we refer to the $(\boldsymbol{D}_0, \boldsymbol{D}_1)$ representation as the Markovian representation of a MAP.

2. SPECTRAL CHARACTERIZATION

We obtain a spectral representation of MAPs, i.e., a simple scalar representation of (1)-(2) based on spectral properties of $(-\boldsymbol{D}_0)^{-1}$ and $\boldsymbol{P}$. This allows to represent the MAP moments and autocorrelations in terms of few scalar parameters, rather than by formulas using matrices.

*This work was supported by the National Science Foundation under grants ITR-0428330 and CNS-0720699.

2.1 Characterization of Moments

We begin by describing the moments (1) in terms of the spectrum of $(-\boldsymbol{D}_0)^{-1}$. Recall that the characteristic polynomial $\phi_{\mathbf{A}} \equiv \phi_{\mathbf{A}}(s)$ of a $n \times n$ matrix $\mathbf{A}$ is

$$\phi_{\mathbf{A}} = s^n + \alpha_1 s^{n-1} + \ldots + \alpha_{n-1}s + \alpha_n,$$

which is a polynomial in s with roots s_i equal to the eigenvalues of $\mathbf{A}$. We consider the Cayley-Hamilton theorem [4], by which the powers of $\mathbf{A}$ satisfy $\mathbf{A}^k = -\sum_{j=1}^{n}\alpha_j \mathbf{A}^{k-j}$, for $k \geq n$, i.e., matrix powers are linearly dependent according to the coefficients[1] of $\phi_{\mathbf{A}}$. Because MAP moments are computed in (1) from matrix powers of $(-\boldsymbol{D}_0)^{-1}$, they are linearly dependent.

LEMMA 1. *In a MAP(n), any $n+1$ consecutive moments are linearly dependent, i.e.,*

$$E[X^k] = -\textstyle\sum_{j=1\ldots n} b_j E[X^{k-j}], \quad E[X^0] = 1, \quad k \geq n, \tag{3}$$

where $b_j = m_j k!/(k-j)!$, and m_j is the coefficient of s^{n-j} in $\phi_{(-\boldsymbol{D}_0)^{-1}}$.

PROOF. Using the Cayley-Hamilton theorem,

$$E[X^k] = -k!\boldsymbol{\pi}_e(\textstyle\sum_{j=1\ldots n} m_j(-\boldsymbol{D}_0)^{-(k-j)})\boldsymbol{e} \tag{4}$$

which immediately proves the theorem by (1). □

Observing that the coefficients of a characteristic polynomial $\phi_{\mathbf{A}}$ are functions of the eigenvalues of $\mathbf{A}$, we can derive the relation between eigenvalues of $(-\boldsymbol{D}_0)^{-1}$ and moments.

THEOREM 1. *Let $(-\boldsymbol{D}_0)^{-1}$ have n eigenvalues, among which only $m \leq n$ eigenvalues $\theta_t \in \mathbb{C}$, $1 \leq t \leq m$, are distinct. Let q_t be the multiplicity of θ_t, $\sum_{t=1\ldots m} q_t = n$. Then the moments may be computed as*

$$E[X^k] = \textstyle\sum_{t=1\ldots m} k!\,\theta_t^k \sum_{j=1\ldots q_t} M_{t,j}k^{j-1}, \tag{5}$$

$$E[X^0] = \textstyle\sum_{t=1\ldots m} M_{t,1} = 1, \tag{6}$$

and the constants $M_{t,j}$'s are independent of k.

PROOF. Note that (3) is an homogeneous linear recurrence of order n in $E[X^k]/k!$ with constant coefficients m_j. Its closed-form solution has the same structure of (5) and depends on n constants $M_{t,j}$'s and on the roots of the characteristic equation of (3) that are immediately found to be the eigenvalues of $(-\boldsymbol{D}_0)^{-1}$. □

[1] The α_j's are easily computed, e.g., with the MATLAB function `poly`.

Observing that the $M_{t,j}$ and θ_t are no more than $2n$ parameters and that the $M_{t,j}$'s are linearly dependent due to the condition $E[X^0] = 1$, we have the following corollary.

COROLLARY 1. *A MAP(n) process can fit up to $2n-1$ independent moments.*

Corollary 1 shows that only $2n-1$ parameters are needed to impose the maximum number of fittable moments, and stresses the redundancy of the Markovian representation [11].

2.2 Characterization of Autocorrelation

The spectral characterization can be extended to autocorrelations by considering the spectrum of $\boldsymbol{P}$, which determines the properties of the matrix powers $\boldsymbol{P}^k$ in (2).

LEMMA 2. *In a MAP(n), any $n+1$ consecutive autocorrelations are linearly dependent, i.e.,*

$$\rho_k = -\sum_{j=1\dots n} a_j \rho_{k-j}, \;\; \rho_0 = \tfrac{1}{2}\left(1 - \tfrac{1}{\mathrm{CV}^2}\right), \;\; k \geq n, \quad (7)$$

where a_j is the coefficient of s^{n-j} in $\phi_{\boldsymbol{P}}$ and $\sum_{j=1}^{n} a_j = 0$.

PROOF. We want to prove that $\sum_{j=0\dots n} a_j \rho_{k-j} = 0$, where $a_0 = 1$. By definition of ρ_k, this is equivalent to prove

$$\sum_j a_j (\boldsymbol{\pi}_e (-\boldsymbol{D}_0)^{-1} \boldsymbol{P}^{k-j} (-\boldsymbol{D}_0)^{-1} \boldsymbol{e} - E[X]^2) = 0.$$

The last equation is indeed true if $\sum_{j=0}^{n} a_j \boldsymbol{P}^{k-j} = 0$ and $\sum_{j=0}^{n} a_j = 0$, and the first relation holds true by the Cayley-Hamilton theorem; the second relation follows by the stochasticity of $\boldsymbol{P}$, as we have that its largest eigenvalue is always $\gamma_1 = 1$ and thus for $s = \gamma_1$ it is $\phi(\boldsymbol{P}) = 0 = \sum_{j=0}^{n} a_j$,which finally proves $\rho_k = -\sum_{j=1\dots n} a_j \rho_{k-j}$. The formula for ρ_0 follows by evaluating (2) for $k = 0$, i.e.,

$$\rho_0 = (E[X]^{-2} \boldsymbol{\pi}_e (-\boldsymbol{D}_0)^{-2} \boldsymbol{e} - 1)/\mathrm{CV}^2 = \left(1 - 1/\mathrm{CV}^2\right)/2.$$

since $\boldsymbol{\pi}_e (-\boldsymbol{D}_0)^{-2} \boldsymbol{e} = E[X^2]/2 = (1 + \mathrm{CV}^2) E[X]^2/2$. □

Using a proof analogous to that of Theorem 1, it is possible to relate ρ_k with the eigenvalues of $\boldsymbol{P}$ and characterize the maximum number of fittable autocorrelations.

THEOREM 2. *Let $\gamma_t \in \mathbb{C}$, $1 \leq t \leq m$, be an eigenvalue of $\boldsymbol{P}$ with multiplicity r_t, and where $\gamma_1 = 1$. Then*

$$\rho_k = \sum_{t=2\dots m} \gamma_t^k \sum_{j=1\dots r_t} A_{t,j} k^{j-1}, \quad (8)$$

$$\rho_0 = \sum_{t=2\dots m} A_{t,1}, \quad k \geq 1, \quad (9)$$

where the $A_{t,j}$'s constants are independent of k.

Observing that the distinct $A_{t,j}$ and γ_t in (8) are no more than $2n-2$, and that fixing CV^2 imposes ρ_0, we have the following corollary.

COROLLARY 2. *A MAP(n) process can fit up to $2n-2$ independent autocorrelations ρ_k, $k \geq 1$. With given CV^2, the maximum independent coefficients become $2n-3$.*

Similarly to the moments, (8) has a much simpler structure than the corresponding Markovian formula (2).

2.3 Example: MAP(2) Characterization

MAP(2)s are popular MAP models thanks to their small parametrization space of just six parameters. MAP(2) characterization results are first given in [3] using matrix exponentials. Here we illustrate the simplicity of characterizing MAP(2)s using spectral methods, by immediately deriving the structure of MAP(2) moments and autocorrelations from the results of the previous sections. From (5), for $\theta_1 \neq \theta_2$

$$E[X^k] = k! M_{1,1} \theta_1^k + k!(1 - M_{1,1}) \theta_2^k, \quad (10)$$

and up to three independent moments may be fitted. Similarly, the autocorrelations are

$$\rho_k = \gamma_2^k A_{2,1} = \gamma_2^k \rho_0 = \frac{\gamma_2^k}{2}\left(1 - \frac{1}{\mathrm{CV}^2}\right), \; k \geq 1, \gamma_2 \in \mathbb{R}, \quad (11)$$

which can fit a single autocorrelation for fixed CV^2. Eqn. (11) indicates that the MAP(2) offers limited versatility in exploring the impact of non-renewal workloads on systems or fitting trace data, since the autocorrelations are always geometrically decaying if $|\gamma_2| < 1$. Based on the spectral results, we consider a more general class of MAPs for fitting.

3. MAP PROCESS DESIGN

Defining MAPs with predefined moments and autocorrelations is necessary for fitting data traces, for model sensitivity analysis, and for the design of processes with specific theoretical properties. The general problem is difficult and several approaches have been investigated, e.g., [1,2,5], but important issues remain open. For instance whenever a trace's autocorrelation exhibits oscillations that would require complex eigenvalues, no simple way is known to impose this behavior to the inter-arrival times[2]. We address the problem by a technique based on the Kronecker product.

3.1 Kronecker Product Composition

We create a MAP(n) by composition of a set of smaller MAPs, typically MAP(2)s. Previous work has intended MAP composition as a superposition of J processes [1,9]

$$\{\boldsymbol{D}_0^{\mathrm{super}}, \boldsymbol{D}_1^{\mathrm{super}}\} = \{\boldsymbol{D}_0^1 \oplus \boldsymbol{D}_0^2 \oplus \cdots \oplus \boldsymbol{D}_0^J, \boldsymbol{D}_1^1 \oplus \boldsymbol{D}_1^2 \oplus \cdots \oplus \boldsymbol{D}_1^J\},$$

where $\oplus$ is the Kronecker sum operator. Obtaining a certain autocorrelation for the superposed process requires imposing the eigenvalues of $\boldsymbol{P}^{\mathrm{super}} = (-\boldsymbol{D}_0^{\mathrm{super}})^{-1} \boldsymbol{D}_1^{\mathrm{super}}$, but this can be prohibitive since $\boldsymbol{P}^{\mathrm{super}}$ has eigenvalues that are not in simple relation with the entries of $\boldsymbol{D}_0^j$ and $\boldsymbol{D}_1^j$, $1 \leq j \leq J$.

To tackle this problem, we propose a different approach which we call Kronecker product composition (kpc), i.e.,

$$\{\boldsymbol{D}_0^{\mathrm{kpc}}, \boldsymbol{D}_1^{\mathrm{kpc}}\} = \{(-1)^{J-1} \boldsymbol{D}_0^1 \otimes \cdots \otimes \boldsymbol{D}_0^J, \boldsymbol{D}_1^1 \otimes \cdots \otimes \boldsymbol{D}_1^J\}$$

where $\otimes$ is the Kronecker product. We assume that all eigenvalues of $\boldsymbol{D}_0^j$ and $\boldsymbol{P}^j = (-\boldsymbol{D}_0^j)^{-1} \boldsymbol{D}_1^j$ are distinct, and require that all J processes, except at most one, have $\boldsymbol{D}_0^j$ diagonal. Recall that when a MAP has diagonal $\boldsymbol{D}_0$, its coefficient of variation cannot be less than one. In general, $\boldsymbol{D}_0^{\mathrm{kpc}}$ is not diagonal since one of the J processes does not have restrictions on its $\boldsymbol{D}_0^j$ matrix, hence the composed process may also have $\mathrm{CV}^2 < 1$. The properties of $\{\boldsymbol{D}_0^{\mathrm{kpc}}, \boldsymbol{D}_1^{\mathrm{kpc}}\}$ follow recursively from those of $\{-\boldsymbol{D}_0^1 \otimes \boldsymbol{D}_0^2, \boldsymbol{D}_1^1 \otimes \boldsymbol{D}_1^2\}$ given below.

THEOREM 3. *The moments of $\{-\boldsymbol{D}_0^1 \otimes \boldsymbol{D}_0^2, \boldsymbol{D}_1^1 \otimes \boldsymbol{D}_1^2\}$ are $E[X^k] = E[X_1^k] E[X_2^k]/k!$, where $E[X_j^k]$ is the k-th moment*

[2]Circulant matrices have been used for assigning the dependence structure of the counting process, e.g., in [6]. However, the technique does not immediately extend to the joint problem of fitting moments and autocorrelations of inter-arrivals.

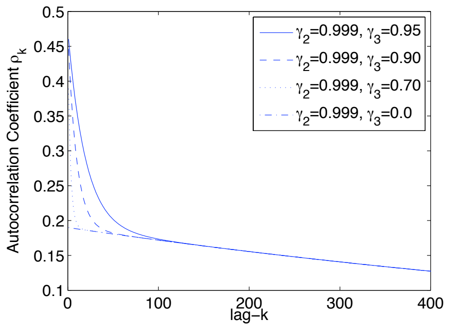

Figure 1: MAP(4) processes with varying low-lag autocorrelations

of the process $\{D_0^j, D_1^j\}$. *In particular, the squared coefficient of variation satisfies* $1+CV^2 = (1+CV_1^2)(1+CV_2^2)/2$, *and* CV_j^2 *is the squared coefficient of variation of* $\{D_0^j, D_1^j\}$.

THEOREM 4. *The autocorrelation of* $\{-D_0^1 \otimes D_0^2, D_1^1 \otimes D_1^2\}$ *satisfies*

$$CV^2\rho_k = (CV_1^2)\rho_k^1 + (CV_2^2)\rho_k^2 + (CV_1^2 CV_2^2)\rho_k^1\rho_k^2,$$

where ρ_k^j *is the lag-k autocorrelation of* $\{D_0^j, D_1^j\}$. *In particular, the eigenvalues* $1, \gamma_2, \gamma_3, \ldots$ *of* P^{kpc}, *which shape the autocorrelation, are obtained by the Kronecker product of the eigenvalues of* $P^1 = -(D_0^1)^{-1}D_1^1$ *and* $P^2 = -(D_0^2)^{-1}D_1^2$.

The above relations simplify the generation of processes with predefined properties of the moments and autocorrelations, and can be proved by properties of the Kronecker product.

3.1.1 Example: Closed Model Sensitivity

Applications of kpc span from fitting trace data to process design. We give a simple example related to the sensitivity analysis of a closed queueing network composed by two or three queues in series. The network population is set to $N = 50$ jobs. The mean service rate is identical at all queues and equal to $\mu = 1$ job/s. The service process at the first queue is a MAP(4) obtained by kpc of two MAP(2)s; the other queues have exponential service. Using kpc, we vary the two largest autocorrelation eigenvalues γ_2 and γ_3 of the MAP(4) to investigate how different dependence profiles may impact on the network throughput. Some of the generated service processes are shown in Figure 1; the network throughput computed by global balance is given in Table 1. The first row refers to the renewal case where the MAP(4) is exponential with the mean $\mu = 1$ job/s and no autocorrelation. The following rows explore the impact of different MAP(4) autocorrelations for $\text{CV}^2 = 20$, e.g., in the fifth row it is shown that $\gamma_2 = 0.999$ decreases the throughput to 0.507 job/s, a -33.4% degradation with respect to the renewal case; in the next row it is shown that adding an eigenvalue $\gamma_3 = 0.700$ has a negligible effect with respect to the previous case (-33.5% versus -33.4%).

These results stress the importance of accounting for non-renewal features, since up to 35% of the throughput can be affected by autocorrelation at a single queue, although the degradation seems to be mitigated when the number of queues is increased. Further, the last part of the table shows that the workloads in Figure 1, although quite different, perform almost identically, with the throughput degradation always between -33.4% and -34.9%. This suggests the need for further investigation on the actual impact of low-lag coefficients on the performance of closed models. Observations of this type are impossible with MMPP(2)s or MAP(2)s, and promote the kpc method to improve understanding of systems under non-renewal workloads.

Table 1: Throughput [job/s] for different MAP(4) service processes at queue 1

Queue 1 MAP	2 Queues	3 Queues
$\gamma_2 = 0.000, \gamma_3 = 0.000$	0.762 (renewal)	0.261 (renewal)
$\gamma_2 = 0.700, \gamma_3 = 0.000$	0.701 (-8.0%)	0.254 (-2.7%)
$\gamma_2 = 0.900, \gamma_3 = 0.000$	0.621 (-18.5%)	0.237 (-9.2%)
$\gamma_2 = 0.950, \gamma_3 = 0.000$	0.577 (-24.2%)	0.226 (-13.4%)
$\gamma_2 = 0.999, \gamma_3 = 0.000$	0.507 (-33.4%)	0.211 (-19.2%)
$\gamma_2 = 0.999, \gamma_3 = 0.700$	0.506 (-33.5%)	0.211 (-19.2%)
$\gamma_2 = 0.999, \gamma_3 = 0.900$	0.502 (-34.1%)	0.211 (-19.2%)
$\gamma_2 = 0.999, \gamma_3 = 0.950$	0.496 (-34.9%)	0.209 (-19.9%)

4. CONCLUSION

We have proposed a spectral characterization of moments and autocorrelation which simplifies the analysis of MAP processes. Our method applies to MAP characterization and synthesis. In particular, the kpc method allows to define easily MAPs with predefined moments and autocorrelations. Ongoing work includes the fitting of traffic traces using kpc.

5. REFERENCES

[1] A.T. Andersen, B.F. Nielsen. A Markovian approach for Modeling Packet traffic with Long-Range Dependence. *IEEE JSAC*, 16(5), 719–732, 1998.

[2] P. Buchholz, A. Panchenko. A Two-Step EM Algorithm for MAP Fitting. LNCS 3280, 217–227, 2004, Springer.

[3] A. Heindl, K. Mitchell, A. van de Liefvoort. Correlation bounds for second-order MAPs with application to queueing network decomposition. *PEVA*, 63(6), 2006.

[4] F. E. Hohn. *Elementary Linear Algebra*. Dover, 1973.

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

[6] L.A. Kulkarni, S. Li. Transient behaviour of queueing systems with correlated traffic. *PEVA* 27, 1996.

[7] R. M. M. Leao, E. de Souza e Silva, S. C. de Lucena. A Set of Tools for Traffic Modeling, Analysis and Experimentation LNCS 1786, 40–55, 2000, Springer.

[8] N. Mi, Q. Zhang, A. Riska, E. Smirni, E. Riedel Performance Impacts of Autocorrelation Flows in Systems. To appear in Performance 2007.

[9] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, 1989.

[10] A. Riska, E. Riedel. Disk Drive Level Workload Characterization. *Proc. USENIX 2006*, 97–102, 2006.

[11] M. Telek, G. Horwath A minimal representation of Markov arrival processes and a moments matching method. To appear in Performance 2007.