

A semidefinite optimization approach to the steady-state analysis of queueing systems

Dimitris Bertsimas · Karthik Natarajan

Received: 1 November 2005 / Revised: 28 November 2006 / Published online: 23 May 2007
© Springer Science+Business Media, LLC 2007

Abstract Computing the steady-state distribution in Markov chains for general distributions and general state space is a computationally challenging problem. In this paper, we consider the steady-state stochastic model $\mathbf{W} \stackrel{d}{=} g(\mathbf{W}, \mathbf{X})$ where the equality is in distribution. Given partial distributional information on the random variables \mathbf{X} , we want to estimate information on the distribution of the steady-state vector \mathbf{W} . Such models naturally occur in queueing systems, where the goal is to find bounds on moments of the waiting time under moment information on the service and interarrival times. In this paper, we propose an approach based on semidefinite optimization to find such bounds. We show that the classical Kingman's and Daley's bounds for the expected waiting time in a GI/GI/1 queue are special cases of the proposed approach. We also report computational results in the queueing context that indicate the method is promising.

Keywords Steady-state distribution · Waiting time · Semidefinite optimization

Research of D. Bertsimas supported in part by Singapore-MIT Alliance.

Research of K. Natarajan is supported in part by NUS Academic Research Grant R146-050-070-133 and Singapore-MIT Alliance.

D. Bertsimas
Sloan School of Management and Operations Research Center,
Massachusetts Institute of Technology, E53-363, Cambridge,
MA 02139, USA
e-mail: dbertsim@mit.edu

K. Natarajan (✉)
Department of Mathematics, National University of Singapore,
Singapore 117543, Singapore
e-mail: matkbn@nus.edu.sg

Mathematics Subject Classification (2000) 60K25 · 90C22

1 Introduction

In this paper, we study stochastic iterative models of the form:

$$\mathbf{W}_{n+1} = g(\mathbf{W}_n, \mathbf{X}_n) \quad \text{for } n = 0, 1, 2, \dots, \quad (1)$$

where \mathbf{W}_n is the state of the system at instance t_n with $t_n < t_{n+1}$ for $n = 0, 1, 2, \dots$. The random vector \mathbf{X}_n changes the state of the system from \mathbf{W}_n to \mathbf{W}_{n+1} under the mapping g . It is well known that any homogeneous Markov chain can be represented as the iterative model in (1) for i.i.d \mathbf{X}_n (see Borovkov and Foss [6] and Müller and Stoyan [15]). When steady-state exists, the model becomes:

$$\mathbf{W} \stackrel{d}{=} g(\mathbf{W}, \mathbf{X}), \quad (2)$$

where the equality is in distribution. Our goal in this paper is to obtain estimates on the distribution of the steady-state vector \mathbf{W} under (partial) distributional information on \mathbf{X} . In particular, given a finite set of moments on the distribution of \mathbf{X} , we obtain bounds on parameters of the distribution of the steady-state vector \mathbf{W} using (2). To motivate this problem, we focus on applications of the class of iterative models in a queueing context.

1.1 Application in queueing systems

Since the pioneering work of Erlang, queueing systems have been extensively studied with a focus on trying to analytically estimate, approximate or bound performance measures

such as waiting time or number of customers in the queue under varying assumptions on the inter-arrival and service time distributions. Consider a queueing system where $n = 0, 1, 2, \dots$ denotes the customer number entering the queue. Assume that the first customer arriving finds the queueing system empty. Let T_n denote the inter-arrival time between n th and $(n + 1)$ th customer and S_n denote the service time for n th customer.

(a) For a single-server GI/GI/1 queue, the waiting time of the n th customer in the queue (denoted by W_n) satisfies the relationship:

$$W_{n+1} = (W_n + S_n - T_n)_+ \quad \text{for } n = 0, 1, 2, \dots,$$

where $z_+ = \max(z, 0)$. Assume that S_n are i.i.d random variables ($S_n \sim S, E[S] = 1/\mu$) and T_n are i.i.d random variables ($T_n \sim T, E[T] = 1/\lambda$). Under the assumption of independence of S_n and T_n for each $n, \rho = \lambda/\mu < 1$, the steady-state waiting time distribution is known to be the unique solution to the recursive equation (cf. Lindley [14]):

$$W \stackrel{d}{=} (W + S - T)_+, \tag{3}$$

where the equality is in distribution and W, S and T are independent random variables.

(b) For a multi-server GI/GI/ c queue with c servers, the workload is defined as a c -dimensional vector process $\mathbf{W}_n := (W_{n1}, W_{n2}, \dots, W_{nc})$ where $W_{n1} \leq W_{n2} \leq \dots \leq W_{nc}$. This workload process satisfies the relationship:

$$\mathbf{W}_{n+1} = \mathcal{R}((\mathbf{W}_n + S_n \mathbf{e}_1 - T_n \mathbf{e})_+) \quad \text{for } n = 0, 1, 2, \dots,$$

where $\mathbf{e}_1 = (1, 0, \dots, 0), \mathbf{e} = (1, 1, \dots, 1)$ and \mathcal{R} rearranges the c components in ascending order. With $\rho_c = \lambda/(c\mu) < 1$, the steady-state distribution is the unique solution to the recursive equation (cf. Kiefer and Wolfowitz [9]):

$$\mathbf{W} \stackrel{d}{=} \mathcal{R}((\mathbf{W} + S\mathbf{e}_1 - T\mathbf{e})_+), \tag{4}$$

where the equality is in distribution and S and T are independent of \mathbf{W} . Here the first component of \mathbf{W} , i.e., W_1 is the steady-state waiting time for the customer in the multi-server queue.

It is clear from Table 1, the queueing systems in (a) and (b) fall under the category of steady-state iterative models defined by (2).

Table 1 GI/GI/1 and GI/GI/ c queueing systems

Queueing system	\mathbf{W}	\mathbf{X}	$g(\mathbf{W}, \mathbf{X})$
GI/GI/1	$W \in \mathfrak{R}_+$	$(S, T) \in \mathfrak{R}_+^2$	$(W + S - T)_+$
GI/GI/ c	$\mathbf{W} \in \mathfrak{R}_+^c$	$(S, T) \in \mathfrak{R}_+^2$	$\mathcal{R}((\mathbf{W} + S\mathbf{e}_1 - T\mathbf{e})_+)$

In each of these queueing systems, given moment information on the random interarrival and service times, we are interested in obtaining bounds on the moments or the tail probability of the steady-state waiting time. One of the earliest results in this context was derived by Kingman [10] for GI/GI/1 queue. For the single server queue, given the mean and variance of the inter-arrival and service times, he derived an upper bound on the expected steady-state waiting time using (3). We outline the proof next since it motivates our approach. As in the original proof, we work with the random variable $X := S - T$.

Proposition 1 (Kingman [10]) *Let $X := S - T$. For a GI/GI/1 queue with $E[X] < 0$, we have:*

$$E[W] \leq \frac{\text{var}(X)}{2|E[X]|}. \tag{5}$$

Proof Equating the first two moments of the random variables W and $(W + X)_+$ in (3) gives:

$$\begin{aligned} E[W] &= E[(W + X)_+], \\ E[W^2] &= E[(W + X)_+^2]. \end{aligned}$$

Defining $z_- = \max(-z, 0)$, we have:

$$z_- = z_+ - z \quad \text{and} \quad z_-^2 = z^2 - z_+^2.$$

Setting $z = W + X$ and taking expectations, we obtain:

$$\begin{aligned} E[(W + X)_-] &= E[(W + X)_+] - E[W + X] \\ &= E[W] - E[W + X] = -E[X], \\ E[(W + X)_-^2] &= 2E[W]E[X] + E[X^2] \\ &\quad (\text{since } E[WX] = E[W]E[X]). \end{aligned}$$

But $E[(W + X)_-^2] \geq E[(W + X)_-]^2$, which implies

$$2E[W]E[X] + E[X^2] - E[X]^2 \geq 0,$$

or equivalently

$$E[W] \leq \frac{\text{var}(X)}{2|E[X]|}. \quad \square$$

The two-moment bound in (5) is based on relaxing the equality of distributions to the equality of the first two moments. Kingman [11] and Daley [7, 8] used extensions of this approach to develop bounds in both single and multiple server queues. While these moment bounds are very popular due to their simplicity, there exist some shortcomings:

- (a) The upper bound on the expected waiting time in (5) is known to be tight only for special instances (for example in a D/D/1 queue). Furthermore, the bounds have been observed to be very weak under light traffic conditions (cf. Kingman [10]).

- (b) Few results are known that incorporate third and higher order moment information in finding tighter bounds in both single and multi-server queues (cf. Wolff and Wang [17]).
- (c) Few moment bounds are known on other parameters of interest, such as the probability that the steady-state waiting time exceeds a given value or the second moment of the waiting time.

These shortcomings arise due to the inherent difficulty in solving the recursive (2) even for simple distributions of X such as the normal distribution. In this paper, we address this problem, by proposing a general algorithmic approach based on the theory of moments and semidefinite optimization to find bounds for a class of iterative stochastic models.

Our paper builds on the work of Bertsimas and Popescu [5] and Lasserre [12], who present applications of moments in the context of probability theory, Bertsimas and Popescu [4] who present applications in option pricing and Bertsimas, Natarajan and Teo [2, 3] who present applications in combinatorial optimization under uncertainty. Related to this paper is a model proposed in Chap. 12 in a recent book by Lasserre and Hernández-Lerma [13]. Therein, the authors introduce a moments approach for a special class of Markov chains and propose bounds based on semidefinite optimization. While the spirit of our method is similar, we crucially show the potential of this approach in queueing analysis.

1.2 Structure and contributions of the paper

In Sect. 2, we outline the moments based approach to compute bounds on parameters of the steady-state distribution for a class of stochastic models. Particularly, by relaxing the equality of distributions to the equality of moments and using the theory of moment cones, we formulate semidefinite relaxations for this problem.

In Sect. 3, we develop semidefinite relaxations for the GI/GI/1 queue under moment information on the inter-arrival and service times. The first order semidefinite relaxation simply reduces to Kingman’s upper bound. Interestingly, we can add higher order moment information and improve significantly on Kingman’s bound clearly indicating the potential of the approach.

In Sect. 4, we outline the extension of this approach to the GI/GI/c queue and provide preliminary numerical results. The first order semidefinite relaxation in this case is tighter than Kingman’s bound.

2 Steady-state analysis

In this section, we formulate a general moments based approach to compute performance bounds on functions of the steady-state vector W under partial distribution information

on X . Let ψ_w and ψ_x denote the probability measures of W and X supported on $\mathcal{S}_w \subseteq \mathfrak{R}^m$ and $\mathcal{S}_x \subseteq \mathfrak{R}^n$ respectively. The joint probability measure for (W, X) (under independence) is $\psi = \psi_w \times \psi_x$ with support $\mathcal{S} = \mathcal{S}_w \times \mathcal{S}_x$. The steady-state recursive equation $W \stackrel{d}{=} g(W, X)$ can be represented as:

$$\psi_w = \psi g^{-1}, \tag{6}$$

where ψ_w is the probability measure for W and ψg^{-1} is the probability measure for $g(W, X)$. To formulate the problem, we let $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m$ and $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{N}^n$ denote the multi-indices in the basis of space of real-valued polynomials in m and n variables of degree at most r respectively:

$$\begin{aligned} \{W^\alpha\}_{|\alpha| \leq r} &:= (1, W_1, \dots, W_m, W_1^2, W_1 W_2, \dots, \\ &W_m^2, \dots, W_{m-1} W_m^{r-1}, W_m^r), \\ \{X^\beta\}_{|\beta| \leq r} &:= (1, X_1, \dots, X_n, X_1^2, X_1 X_2, \dots, X_n^2, \dots, \\ &X_{n-1} X_n^{r-1}, X_n^r). \end{aligned}$$

The notation W^α stands for $W_1^{\alpha_1} \dots W_m^{\alpha_m}$ and $|\alpha| = \sum_{i=1}^m \alpha_i$. Likewise for X and β . For a given set of known moments of X , our central problem is the following steady-state model:

- Assume that we are given a finite set of moments for X up to degree $2r$. Given Borel measurable sets $\mathcal{S}_w \subseteq \mathfrak{R}^m$, $\mathcal{S}_x \subseteq \mathfrak{R}^n$ and Borel measurable functions $f : \mathcal{S}_w \rightarrow \mathfrak{R}$, $g : \mathcal{S} \rightarrow \mathcal{S}_w$, solve:

$$(\mathbb{P}) \quad \sup_{\psi_w, \psi_x} / \inf_{\psi_w, \psi_x} E_{\psi_w}[f(W)] \tag{7}$$

$$\text{s.t.} \quad \psi_w = \psi g^{-1}, \tag{8}$$

$$\psi = \psi_w \times \psi_x, \tag{9}$$

$$E_{\psi_x}[X^\beta] = m^\beta, \quad \forall |\beta| \leq 2r, \tag{10}$$

$$E_{\psi_x}[1] = E_{\psi_w}[1] = 1, \tag{11}$$

$$\psi_w \in \mathbb{M}(\mathcal{S}_w), \quad \psi_x \in \mathbb{M}(\mathcal{S}_x), \tag{12}$$

where $\mathbb{M}(\mathcal{S}_w)$, $\mathbb{M}(\mathcal{S}_x)$ is the set of finite positive Borel measures supported by \mathcal{S}_w and \mathcal{S}_x respectively.

We explicitly add the constraints $E_{\psi_x}[1] = E_{\psi_w}[1] = 1$ in (P) to ensure that the measures are probability measures. Since the probability measure of W is not uniquely determined in (P), solving (P) provides bounds on moments of functions of W . We assume that the problem is well-posed, namely there is sufficient information on X to guarantee a finite value for $E_{\psi_w}[f(W)]$.

2.1 A moments approach

To formulate the moments approach to solve (P), we make the following key assumption about the iterative function g and the objective function f .

Assumption 1 *The function $g : \mathcal{S} \rightarrow \mathcal{S}_w$ is piecewise linear and function $f : \mathcal{S} \rightarrow \mathfrak{R}$ is piecewise polynomial¹ over \mathcal{S} . Namely, there exist disjoint sets \mathcal{S}_k such that the functions $g_k(\mathbf{W}, \mathbf{X})$ are linear and functions $f_k(\mathbf{W})$ are polynomial of degree at most $2r$ on each \mathcal{S}_k and $\mathcal{S} = \bigcup_{k=1}^K \mathcal{S}_k$:*

$$\begin{aligned} g(\mathbf{W}, \mathbf{X}) &= g_k(\mathbf{W}, \mathbf{X}), \\ \text{if } (\mathbf{W}, \mathbf{X}) \in \mathcal{S}_k, \quad k &= 1, \dots, K, \\ f(\mathbf{W}) &= f_k(\mathbf{W}), \\ \text{if } (\mathbf{W}, \mathbf{X}) \in \mathcal{S}_k, \quad k &= 1, \dots, K. \end{aligned} \tag{13}$$

Furthermore, each set \mathcal{S}_k is characterized by inequality constraints among polynomials of degree at most $2r$. Such sets are known as semi-algebraic sets.²

For the GI/GI/1 and GI/GI/c queues in Table 1, these sets \mathcal{S}_k are simply polyhedral sets.

Consider the joint probability measure ψ for (\mathbf{W}, \mathbf{X}) . Let $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{N}^{m+n}$ denote the multi-indices in the basis of space of real-valued polynomials in $m + n$ variables of degree at most r :

$$\begin{aligned} \{(\mathbf{W}, \mathbf{X})^\boldsymbol{\gamma}\}_{|\boldsymbol{\gamma}| \leq r} &:= \{(\mathbf{W}^\alpha, \mathbf{X}^\beta)\}_{|\alpha|+|\beta| \leq r} \\ &= (1, W_1, \dots, X_n, \dots, W_1^r, \dots, W_m^{r-1} X_n, \dots, X_n^r). \end{aligned}$$

2.1.1 Decision variables and objective

We define the decision variables as:

$$x_k^{\alpha\beta} := E[\mathbf{W}^\alpha \mathbf{X}^\beta | \mathcal{S}_k] P[\mathcal{S}_k],$$

and let $\mathbf{x}_k := \{x_k^{\alpha\beta}\}_{|\alpha|+|\beta| \leq 2r}$ be the vector of decision variables in this basis. With this definition, we can rewrite the objective (7) as a linear function of the decision variables:

$$\begin{aligned} E[f(\mathbf{W})] &= \sum_{k=1}^K E[f_k(\mathbf{W}) | \mathcal{S}_k] P[\mathcal{S}_k] \\ &= \sum_{k=1}^K \sum_{|\alpha| \leq 2r} f_k^{\alpha\mathbf{0}} x_k^{\alpha\mathbf{0}}, \end{aligned}$$

for appropriately identified coefficients $f_k^{\alpha\mathbf{0}}$. The existence of the coefficients is guaranteed from Assumption 1.

¹Without loss of generality, we expand the definition of $f(\mathbf{W})$ from $\mathcal{S}_w \rightarrow \mathfrak{R}$ to $f(\mathbf{W}, \mathbf{X})$ from $\mathcal{S} \rightarrow \mathfrak{R}$.

²For clarity, we work directly with the set \mathcal{S}_k instead of the closure of the set. This is a minor issue since we deal with continuous functions and the results directly extend to the closure of the sets.

2.1.2 Constraints

We relax the equality of the distributions of \mathbf{W} and $g(\mathbf{W}, \mathbf{X})$ in constraint (8) to the equality of moments:

$$\begin{aligned} \mathbf{W} &\stackrel{d}{=} g(\mathbf{W}, \mathbf{X}) \\ \implies E[\mathbf{W}^\alpha] &= E[g(\mathbf{W}, \mathbf{X})^\alpha] \quad \forall |\alpha| \leq 2r. \end{aligned}$$

In terms of conditional expectations, this condition is expressed as:

$$\begin{aligned} \sum_{k=1}^K E[\mathbf{W}^\alpha | \mathcal{S}_k] P[\mathcal{S}_k] - \sum_{k=1}^K E[g_k(\mathbf{W}, \mathbf{X})^\alpha | \mathcal{S}_k] P[\mathcal{S}_k] &= 0 \\ \forall |\alpha| &\leq 2r. \end{aligned}$$

By appropriately identifying the coefficients for the linear functions g_k , it is possible to re-express this condition as linear constraints in the decision variables:

$$\sum_{k=1}^K x_k^{\alpha\mathbf{0}} - \sum_{k=1}^K \sum_{|\boldsymbol{\gamma}| \leq |\alpha|} g_{k\alpha}^\boldsymbol{\gamma} x_k^\boldsymbol{\gamma} = 0 \quad \forall |\alpha| \leq 2r,$$

where $\{g_{k\alpha}^\boldsymbol{\gamma}\}$ is the corresponding set of coefficients found by taking the α th power of g_k .

We relax the independence condition of \mathbf{W} and \mathbf{X} in constraint (9) by equating the moments of the product of the variables to the product of the moments of the variables:

$$\begin{aligned} \mathbf{W}, \mathbf{X} \text{ independent} &\implies E[\mathbf{W}^\alpha \mathbf{X}^\beta] = E[\mathbf{W}^\alpha] E[\mathbf{X}^\beta] \\ \forall |\alpha| + |\beta| &\leq 2r. \end{aligned}$$

With $\mathbf{m}^\beta = E[\mathbf{X}^\beta]$, this condition is expressed as:

$$\begin{aligned} \sum_{k=1}^K E[\mathbf{W}^\alpha \mathbf{X}^\beta | \mathcal{S}_k] P[\mathcal{S}_k] - \mathbf{m}^\beta \sum_{k=1}^K E[\mathbf{W}^\alpha | \mathcal{S}_k] P[\mathcal{S}_k] &= 0 \\ \forall |\alpha| + |\beta| &\leq 2r, \end{aligned}$$

which reduces to linear constraints on the decision variables:

$$\sum_{k=1}^K x_k^{\alpha\beta} - \mathbf{m}^\beta \sum_{k=1}^K x_k^{\alpha\mathbf{0}} = 0 \quad \forall |\alpha| + |\beta| \leq 2r.$$

For $\alpha = \mathbf{0}$, the condition reduces to specifying the moments of \mathbf{X} in constraint (10).

To ensure the probabilities sum up to one in constraint (11), we have:

$$\sum_{k=1}^K P[\mathcal{S}_k] = \sum_{k=1}^K x_k^{\mathbf{0}\mathbf{0}} = 1.$$

Lastly, from constraint (12), we need to ensure that \mathbf{x}_k represents a valid moment sequence of measures over the support \mathcal{S}_k . Define the cone of moments supported on \mathcal{S}_k as:

$$\mathcal{M}_{2r}(\mathcal{S}_k) = \{ \mathbf{x}_k \mid x_k^{\alpha\beta} = E_{\psi_k}[\mathbf{W}^\alpha \mathbf{X}^\beta] \forall |\alpha| + |\beta| \leq 2r \text{ for some } \psi_k \in \mathbb{M}(\mathcal{S}_k) \},$$

Let $\overline{\mathcal{M}_{2r}(\mathcal{S}_k)}$ denotes the closure of this cone. Then, we must have

$$\mathbf{x}_k \in \overline{\mathcal{M}_{2r}(\mathcal{S}_k)}, \quad k = 1, \dots, K.$$

Proposition 2 *An upper/lower bound to the optimal objective value in problem (P) is obtained by solving the conic optimization problem:*

$$(\mathbb{P}_r) \quad \sup/\inf_{\mathbf{x}_k} \sum_{k=1}^K \sum_{|\alpha| \leq 2r} f_k^{\alpha 0} x_k^{\alpha 0}$$

$$\text{s.t.} \quad \sum_{k=1}^K x_k^{\alpha 0} - \sum_{k=1}^K \sum_{|\gamma| \leq |\alpha|} g_{k\alpha}^\gamma x_k^\gamma = 0, \quad \forall |\alpha| \leq 2r, \tag{14a}$$

$$\sum_{k=1}^K x_k^{\alpha\beta} - m^\beta \sum_{k=1}^K x_k^{\alpha 0} = 0, \quad \forall |\alpha| + |\beta| \leq 2r, \tag{14b}$$

$$\sum_{k=1}^K x_k^{00} = 1, \tag{14c}$$

$$\mathbf{x}_k \in \overline{\mathcal{M}_{2r}(\mathcal{S}_k)}, \quad k = 1, \dots, K. \tag{14d}$$

Remarks

(a) Suppose, we are interested in finding bounds on the p th moment of \mathbf{W} for $|\mathbf{p}| \leq 2r$. In this case, the conic optimization problem reduces to:

$$(\mathbb{P}_r) \quad \sup/\inf_{\mathbf{x}_k} \sum_{k=1}^K x_k^{p0}$$

$$\text{s.t.} \quad (14a-14d).$$

(b) Suppose, we are interested in computing bounds on $P[\mathbf{W} \in \mathcal{W}]$ where \mathcal{W} is a semi-algebraic set. In this case, we simply expand the partition to $2K$ sets using:

$$\mathcal{S}_k \cap \mathcal{W} \quad \text{and} \quad \mathcal{S}_k \cap \mathcal{W}^c \quad \text{for } k = 1, \dots, K, \tag{15}$$

where \mathcal{W}^c is the complement of \mathcal{W} .

Clearly, (\mathbb{P}_r) is a moments-based relaxation for the original problem (P) based on joint moments up to degree $2r$. The conic optimization problem (\mathbb{P}_r) provides upper and lower bounds that become tighter as r is increased. For a

given distribution of \mathbf{X} , this bound would converge to the exact solution in the limit as $r \uparrow \infty$ if:

- (1) All the moments of \mathbf{X} are known
- (2) The moments of \mathbf{X} completely determine the distribution of \mathbf{X} and
- (3) The distribution of \mathbf{X} determines the distribution of \mathbf{W} uniquely

These conditions are however fairly strong typically.

2.2 Semidefinite relaxations

We now propose using semidefinite relaxations to solve the conic optimization problem (\mathbb{P}_r) . We use positive semidefinite matrices to characterize the moment cone as in Lasserre [12] and Zuluaga and Pena [18].

Given a sequence $\mathbf{x} = \{x^{\alpha\beta}\}_{|\alpha|+|\beta| \leq 2r}$, let $\mathbf{M}_r(\mathbf{x})$ denote the moment matrix with rows and columns indexed in the basis of polynomials of degree at most r . The entries of the moment matrix are given as follows:

$$\mathbf{M}_r(\mathbf{x})(1, j) = x^{\alpha_1 \beta_1} \quad \text{and} \quad \mathbf{M}_r(\mathbf{x})(i, 1) = x^{\alpha_2 \beta_2}$$

$$\implies \mathbf{M}_r(\mathbf{x})(i, j) = x^{\alpha_1 + \alpha_2, \beta_1 + \beta_2}.$$

For instance, in the 2-dimensional case with $m = 1, n = 1$ the moment matrix for $r = 2$ is:

$$\mathbf{M}_2(\mathbf{x}) = \begin{pmatrix} x^{00} & x^{10} & x^{01} & x^{20} & x^{11} & x^{02} \\ x^{10} & x^{20} & x^{11} & x^{30} & x^{21} & x^{12} \\ x^{01} & x^{11} & x^{02} & x^{21} & x^{12} & x^{03} \\ x^{20} & x^{30} & x^{21} & x^{40} & x^{31} & x^{22} \\ x^{11} & x^{21} & x^{12} & x^{31} & x^{22} & x^{13} \\ x^{02} & x^{12} & x^{03} & x^{22} & x^{13} & x^{04} \end{pmatrix}.$$

A necessary condition for \mathbf{x} to be a valid truncated moment sequence is the positive semi-definiteness of the moment matrix, namely $\mathbf{M}_r(\mathbf{x}) \geq \mathbf{0}$.

Additionally suppose the measure is supported on a polynomial $s(\mathbf{W}, \mathbf{X}) \geq 0$ of degree either $2d - 1$ (if odd) or $2d$ (if even) with $r \geq d$ and the coefficients of the polynomial given as $\mathbf{s} = \{s^\gamma\}_{|\gamma| \leq 2d}$. A localizing matrix $\mathbf{M}_{r-d}(\mathbf{s}, \mathbf{x})$ is defined as:

$$\mathbf{M}_{r-d}(\mathbf{s}, \mathbf{x})(i, j) = \sum_{|\gamma| \leq 2d} s_\gamma x^{\eta(i, j) + \gamma},$$

where $\eta(i, j)$ is the subscript of the (i, j) entry in matrix $\mathbf{M}_{r-d}(\mathbf{x})$. Suppose we have $s(\mathbf{W}, \mathbf{X}) := \mathbf{W} \geq 0$ in the previous example, then $d = 1, \mathbf{s} = (0, 1, 0, 0, 0, 0)$ and

$$\mathbf{M}_1(\mathbf{s}, \mathbf{x}) = \begin{pmatrix} x^{10} & x^{20} & x^{11} \\ x^{20} & x^{30} & x^{21} \\ x^{11} & x^{21} & x^{12} \end{pmatrix}.$$

In this case, an additional necessary condition for the sequence to be valid moments over this support is $\mathbf{M}_{r-d}(\mathbf{s}, \mathbf{x}) \geq \mathbf{0}$. While these conditions are necessary, they are not in general sufficient to ensure validity of moment sequences.

Let the set \mathcal{S}_k in (13) be defined by the intersection of a finite set of polynomial inequalities:

$$\mathcal{S}_k := \bigcap_{s \in \mathcal{S}_k} \{s(\mathbf{W}, \mathbf{X}) \geq 0\},$$

where $2d_s - 1$ or $2d_s$ represents the degree of the polynomial s and $r \geq \max_k \max_{s \in \mathcal{S}_k} d_s$.

Proposition 3 *A upper/lower bound to the optimal objective value in problem (\mathbb{P}_r) is obtained by solving the semidefinite optimization problem:*

$$\begin{aligned}
 (\mathbb{P}_r^s) \quad & \sup_{\mathbf{x}_k} / \inf \sum_{k=1}^K \sum_{|\alpha| \leq 2r} f_k^{\alpha \mathbf{0}} x_k^{\alpha \mathbf{0}} \\
 \text{s.t.} \quad & \sum_{k=1}^K x_k^{\alpha \mathbf{0}} - \sum_{k=1}^K \sum_{|\gamma| \leq |\alpha|} g_{k\alpha}^{\gamma} x_k^{\gamma} = 0, \\
 & \forall |\alpha| \leq 2r, \tag{16a} \\
 & \sum_{k=1}^K x_k^{\alpha \beta} - m^{\beta} \sum_{k=1}^K x_k^{\alpha \mathbf{0}} = 0, \\
 & \forall |\alpha| + |\beta| \leq 2r, \tag{16b} \\
 & \sum_{k=1}^K x_k^{\mathbf{0} \mathbf{0}} = 1, \tag{16c} \\
 & \mathbf{M}_r(\mathbf{x}_k) \geq \mathbf{0}, \quad k = 1, \dots, K, \tag{16d} \\
 & \mathbf{M}_{r-d_s}(\mathbf{s}, \mathbf{x}_k) \geq \mathbf{0}, \\
 & \forall s \in \mathcal{S}_k, k = 1, \dots, K. \tag{16e}
 \end{aligned}$$

Clearly, the formulations are related as:

$$\inf(\mathbb{P}_r^s) \leq \inf(\mathbb{P}_r) \leq \inf(\mathbb{P}) \leq \sup(\mathbb{P}) \leq \sup(\mathbb{P}_r) \leq \sup(\mathbb{P}_r^s).$$

The tightness of the semidefinite relaxations for a class of problems as $r \uparrow \infty$ has been shown in [13].

3 Lindley processes and the GI/GI/1 queue

We now study an application of the semidefinite optimization approach in the analysis of a Lindley process [14]. While such a process was introduced in the context of queueing, it is of independent interest for the theory of random walks. Consider a discrete time process of the form:

$$W_{n+1} = (W_n + X_n)_+ \quad \text{for } n = 0, 1, \dots,$$

where $W_0 = 0$ and X_0, X_1, \dots are i.i.d. random variables. Under the condition $E[X] < 0$, the steady-state version of the Lindley process is:

$$W \stackrel{d}{=} (W + X)_+. \tag{17}$$

In this setting, $\mathcal{S}_w = \mathfrak{R}_+$, $\mathcal{S}_x = \mathfrak{R}$, $m = 1$, $n = 1$ and $g(W, X) = (W + X)_+$. To formulate the semidefinite relaxations, we use the following result from the theory of random walks.

Proposition 4 *Consider a Lindley process (17) with $EX < 0$. Then for $p > 0$, $E[W^p] < \infty$ provided that $E[X_+^{p+1}] < \infty$. Furthermore, if $E|X|^{p+1} < \infty$ for some $p = 1, 2, \dots$, then*

$$\sum_{q=0}^p \binom{p+1}{q} E[W^q] E[X^{p+1-q}] = E[-(W + X)_-^{p+1}]. \tag{18}$$

Proof Refer to page 270 in [1]. □

Here (18) is obtained from the moment equality condition:

$$E[W^{p+1}] = E[(W + X)_+^{p+1}], \tag{19}$$

and the relationship $W + X = (W + X)_+ - (W + X)_-$. For clarity, we work directly with the condition (19) in the semidefinite formulations.

3.1 Bounds on moments of W

Let

$$s_1(W, X) := W \geq 0 \quad \text{and} \quad s_2(W, X) := W + X \geq 0.$$

We then write the order r relaxation as follows.

Proposition 5 *Given $E[X^\beta] = m^\beta < \infty$ for $\beta = 1, \dots, 2r$, an upper/lower bound on the p th moment of W in a Lindley process for $p < 2r$ is obtained by solving the semidefinite optimization problem:*

$$\begin{aligned}
 (\mathbb{P}_r^s) \quad & \sup / \inf x_1^{p0} + x_2^{p0} \\
 \text{s.t.} \quad & x_1^{\alpha 0} - \sum_{\beta=0}^{\alpha-1} \binom{\alpha}{\beta} x_2^{\beta, \alpha-\beta} = 0, \quad \forall \alpha = 1, \dots, 2r, \\
 & x_1^{\alpha \beta} + x_2^{\alpha \beta} - m^\beta (x_1^{\alpha 0} + x_2^{\alpha 0}) = 0, \\
 & \forall \alpha + \beta = 1, \dots, 2r, \\
 & x_1^{\mathbf{0} \mathbf{0}} + x_2^{\mathbf{0} \mathbf{0}} = 1, \\
 & \mathbf{M}_r(\mathbf{x}_1), \mathbf{M}_r(\mathbf{x}_2) \geq \mathbf{0}, \\
 & \mathbf{M}_{r-1}(s_1, \mathbf{x}_1), \mathbf{M}_{r-1}(s_1, \mathbf{x}_2), \\
 & \mathbf{M}_{r-1}(-s_2, \mathbf{x}_1), \mathbf{M}_{r-1}(s_2, \mathbf{x}_2) \geq \mathbf{0}.
 \end{aligned}$$

Proof For the Lindley process in (17), the piecewise linear function $g(W, X) = (W + X)_+$ can be expressed as:

$$g(W, X) = \begin{cases} 0 & \text{for } \mathcal{S}_1 = \{(W, X) \mid W \geq 0, \\ & W + X \leq 0\}, \\ W + X & \text{for } \mathcal{S}_2 = \{(W, X) \mid W \geq 0, \\ & W + X \geq 0\}. \end{cases} \quad (20)$$

Define the two sequences of decision variables as:

$$\mathbf{x}_1 = \{x_1^{\alpha\beta}\}_{\alpha+\beta \leq 2r} := \{E[W^\alpha X^\beta | \mathcal{S}_1] P[\mathcal{S}_1]\}_{\alpha+\beta \leq 2r},$$

$$\mathbf{x}_2 = \{x_2^{\alpha\beta}\}_{\alpha+\beta \leq 2r} := \{E[W^\alpha X^\beta | \mathcal{S}_2] P[\mathcal{S}_2]\}_{\alpha+\beta \leq 2r}.$$

The p th moment of W is then expressed as:

$$\begin{aligned} E[W^p] &= E[W^p | \mathcal{S}_1] P[\mathcal{S}_1] + E[W^p | \mathcal{S}_2] P[\mathcal{S}_2] \\ &= x_1^{p0} + x_2^{p0}. \end{aligned}$$

Constraint (16a) can be expressed as

$$\begin{aligned} 0 &= E[W^\alpha] - E[(W + X)_+^\alpha] \\ &= E[W^\alpha | \mathcal{S}_1] P[\mathcal{S}_1] + E[W^\alpha | \mathcal{S}_2] P[\mathcal{S}_2] \\ &\quad - E[(W + X)^\alpha | \mathcal{S}_2] P[\mathcal{S}_2] \\ &= x_1^{\alpha0} + x_2^{\alpha0} - E\left[\sum_{\beta=0}^{\alpha} \binom{\alpha}{\beta} W^\beta X^{\alpha-\beta} \mid \mathcal{S}_2\right] P[\mathcal{S}_2] \\ &= x_1^{\alpha0} + x_2^{\alpha0} - \sum_{\beta=0}^{\alpha} \binom{\alpha}{\beta} x_2^{\beta, \alpha-\beta} \\ &= x_1^{\alpha0} - \sum_{\beta=0}^{\alpha-1} \binom{\alpha}{\beta} x_2^{\beta, \alpha-\beta}. \end{aligned}$$

Constraint (16b) can be expressed as:

$$\begin{aligned} 0 &= E[W^\alpha X^\beta] - m^\beta E[W^\alpha] \\ &= E[W^\alpha X^\beta | \mathcal{S}_1] P[\mathcal{S}_1] + E[W^\alpha X^\beta | \mathcal{S}_2] P[\mathcal{S}_2] \\ &\quad - m^\beta (E[W^\alpha | \mathcal{S}_1] P[\mathcal{S}_1] + E[W^\alpha | \mathcal{S}_2] P[\mathcal{S}_2]) \\ &= x_1^{\alpha\beta} + x_2^{\alpha\beta} - m^\beta (x_1^{\alpha0} + x_2^{\alpha0}). \end{aligned}$$

The probability of these events sum up to implying that

$$\begin{aligned} 1 &= P[\mathcal{S}_1] + P[\mathcal{S}_2] \\ &= x_1^{00} + x_2^{00}. \end{aligned}$$

Lastly, the semidefinite restrictions $\mathbf{M}_r(\mathbf{x}_1), \mathbf{M}_r(\mathbf{x}_2) \geq \mathbf{0}$ come from necessary conditions for the moment matrix. The entries of the positive semidefinite localizing matrices are

defined as:

$$\begin{aligned} \mathbf{M}_{r-1}(\mathbf{s}_1, \mathbf{x}_k)(i, j) &= x_k^{\eta(i,j)+(1,0)} \quad \text{for } k = 1, 2, \\ \mathbf{M}_{r-1}(-\mathbf{s}_2, \mathbf{x}_1)(i, j) &= -x_1^{\eta(i,j)+(1,0)} - x_1^{\eta(i,j)+(0,1)}, \\ \mathbf{M}_{r-1}(\mathbf{s}_2, \mathbf{x}_2)(i, j) &= x_2^{\eta(i,j)+(1,0)} + x_2^{\eta(i,j)+(0,1)} \end{aligned}$$

where $\eta(i, j)$ is the subscript of the (i, j) entry in matrix $\mathbf{M}_{r-1}(\mathbf{x})$. This comes from the defining inequalities $s_1(W, X) := W \geq 0$ and $s_2(W, X) := W + X \geq 0$. \square

Note that in Proposition 5, the sets \mathcal{S}_1 and \mathcal{S}_2 overlap at $W + X = 0$. This is not a restriction since the objective vanishes at $W + X = 0$.

Now consider the special case, where only the first two moments for X are known and we are interested in finding an upper bound on $E[W]$. This is the classical setting in which Kingman [10] developed his bound. Let $E[X] < 0$ and $\text{var}[X] = \sigma^2 < \infty$.

Proposition 6 *The optimal objective value to the semidefinite relaxation (\mathbb{P}_1^s) equals Kingman’s bound:*

$$\sup \mathbb{P}_1^s = \frac{\sigma^2}{2|E[X]|}.$$

Proof The first order semidefinite relaxation for computing an upper bound on $E[W]$ reduces to:

$$\begin{aligned} (\mathbb{P}_1^s) \quad &\sup x_1^{10} + x_2^{10} \\ \text{s.t.} \quad &x_1^{10} - x_2^{01} = 0, \end{aligned} \quad (21)$$

$$x_1^{20} - x_2^{02} - 2x_2^{11} = 0, \quad (22)$$

$$x_1^{01} + x_2^{01} = E[X], \quad (23)$$

$$x_1^{02} + x_2^{02} = E[X]^2 + \sigma^2, \quad (24)$$

$$x_1^{11} + x_2^{11} - (x_1^{10} + x_2^{10})E[X] = 0, \quad (25)$$

$$x_1^{00} + x_2^{00} = 1, \quad (26)$$

$$\begin{pmatrix} x_1^{00} & x_1^{10} & x_1^{01} \\ x_1^{10} & x_1^{20} & x_1^{11} \\ x_1^{01} & x_1^{11} & x_1^{02} \end{pmatrix}, \begin{pmatrix} x_2^{00} & x_2^{10} & x_2^{01} \\ x_2^{10} & x_2^{20} & x_2^{11} \\ x_2^{01} & x_2^{11} & x_2^{02} \end{pmatrix} \succeq \mathbf{0}, \quad (27)$$

$$x_1^{10} + x_1^{01} \leq 0, \quad x_2^{10} + x_2^{01} \geq 0, \quad (28)$$

$$x_1^{10}, x_2^{10} \geq 0. \quad (29)$$

We construct a feasible solution to (\mathbb{P}_1^s) that attains Kingman’s bound. This is sufficient since all the conditions used to compute Kingman’s bound are included as constraints

in (\mathbb{P}_1^s) . A feasible solution to the semidefinite relaxation is constructed below. Let $0 < \epsilon \leq \min(\sigma^2/(2E[X]^2), 1)$. Consider:

$$\begin{pmatrix} x_1^{00} & x_1^{10} & x_1^{01} \\ x_1^{10} & x_1^{20} & x_1^{11} \\ x_1^{01} & x_1^{11} & x_1^{02} \end{pmatrix} = \begin{pmatrix} 1 - \epsilon & 0 & E[X](1 - \epsilon) \\ 0 & 0 & 0 \\ E[X](1 - \epsilon) & 0 & E[X]^2(1 - \epsilon) \end{pmatrix},$$

$$\begin{pmatrix} x_2^{00} & x_2^{10} & x_2^{01} \\ x_2^{10} & x_2^{20} & x_2^{11} \\ x_2^{01} & x_2^{11} & x_2^{02} \end{pmatrix} = \begin{pmatrix} \epsilon & \frac{\sigma^2}{2|E[X]|} & E[X]\epsilon \\ \frac{\sigma^2}{2|E[X]|} & \frac{\sigma^4}{4E[X]^2\epsilon} & -\frac{\sigma^2}{2} \\ E[X]\epsilon & -\frac{\sigma^2}{2} & E[X]^2\epsilon + \sigma^2 \end{pmatrix}.$$

It can be easily verified that these two matrices are positive semidefinite thus satisfying condition (27). Furthermore for $0 < \epsilon \leq \min(\sigma^2/(2E[X]^2), 1)$, conditions (28–29) are satisfied. Adding up the two matrices:

$$\begin{pmatrix} x_1^{00} & x_1^{10} & x_1^{01} \\ x_1^{10} & x_1^{20} & x_1^{11} \\ x_1^{01} & x_1^{11} & x_1^{02} \end{pmatrix} + \begin{pmatrix} x_2^{00} & x_2^{10} & x_2^{01} \\ x_2^{10} & x_2^{20} & x_2^{11} \\ x_2^{01} & x_2^{11} & x_2^{02} \end{pmatrix} = \begin{pmatrix} 1 & \frac{\sigma^2}{2|E[X]|} & E[X] \\ \frac{\sigma^2}{2|E[X]|} & \frac{\sigma^4}{4E[X]^2\epsilon} & -\frac{\sigma^2}{2} \\ E[X] & -\frac{\sigma^2}{2} & E[X]^2 + \sigma^2 \end{pmatrix},$$

it can be verified that constraints (23–26) are satisfied. To check the feasibility of constraints (21, 22) in (\mathbb{P}_1^s) , observe that:

$$x_1^{10} - x_2^{01} = -E[X]\epsilon,$$

$$x_1^{20} - x_2^{02} - 2x_2^{11} = -E[X]^2\epsilon.$$

As $\epsilon \downarrow 0$, both these values decrease to zero. Hence in the limit with $\epsilon \downarrow 0$, the optimal objective value to (\mathbb{P}_1^s) is $x_1^{10} + x_2^{10} = \sigma^2/(2|E[X]|)$. \square

Thus with only two moment information, the semidefinite relaxation (\mathbb{P}_1^s) gives Kingman’s bound. As we will show in the computational results, the proposed approach can easily handle higher order information for X and thus improve the quality of the bounds significantly.

3.1.1 Computational results for Lindley process

We consider a Lindley process for a normally distributed random variable X . Even in this simple case, there is no simple form for the distribution of W (see the discussion in [1], p. 243). We use parameters $E[X] = -0.25$ and $\sigma^2 = 0.25$

Table 2 Bounds on $E[W^p]$ for $p = 1, \dots, 4$ in Lindley process (– indicates not defined)

Moments	$E[W]$	$E[W^2]$	$E[W^3]$	$E[W^4]$
\mathbb{P}_1^s -UB	0.500	∞	–	–
\mathbb{P}_2^s -UB	0.376	0.341	0.477	∞
\mathbb{P}_3^s -UB	0.323	0.315	0.456	0.929
\mathbb{P}_4^s -UB	0.310	0.304	0.449	0.907
\mathbb{P}_5^s -UB	0.287	0.286	0.433	0.886
Simulation	0.265	0.275	0.415	0.830
\mathbb{P}_5^s -LB	0.232	0.258	0.390	0.774
\mathbb{P}_4^s -LB	0.199	0.241	0.372	0.734
\mathbb{P}_3^s -LB	0.179	0.228	0.362	0.700
\mathbb{P}_2^s -LB	0.146	0.217	0.200	0.167
\mathbb{P}_1^s -LB	0	0	–	–

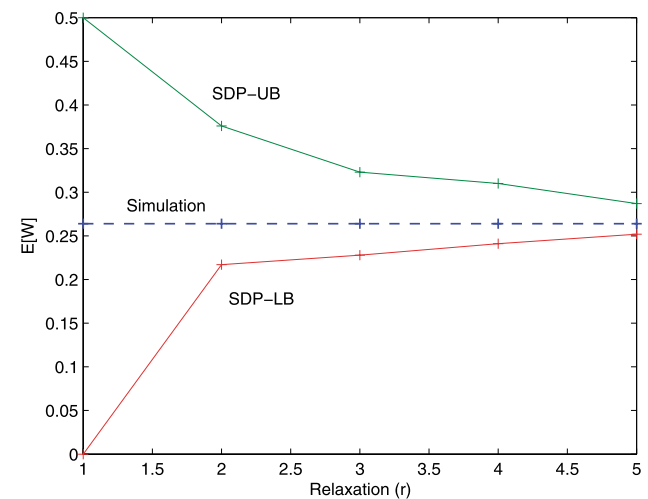


Fig. 1 Bounds on $E[W]$ in the Lindley process

for X .³ The Lindley process was simulated over 500000 replications of X and the results were averaged to estimate the first four moments of W . We compare this with the upper and lower bounds on $E[W^p]$ for $p = 1, \dots, 4$ by solving higher order semidefinite relaxations, thereby incorporating more moment information on X .⁴ The lower and upper bounds from the semidefinite relaxations and the simulated values are displayed in Table 2. The bounds on the first moment $E[W]$ are also plotted in Fig. 1.

³The first ten moments are $(-0.25, 0.3125, -0.2031, 0.2852, -0.2744, 0.425, -0.5179, 0.8733, -1.2541, 2.2785)$.

⁴The experiments were run by a PC with a Intel Pentium 4-M 1.72 GHz CPX, 256 MB of RAM and Microsoft Windows XP Professional operation system. It was coded in MATLAB 7.04 and SeDuMi 1.1 [16] was used as the solver for the semidefinite optimization problems.

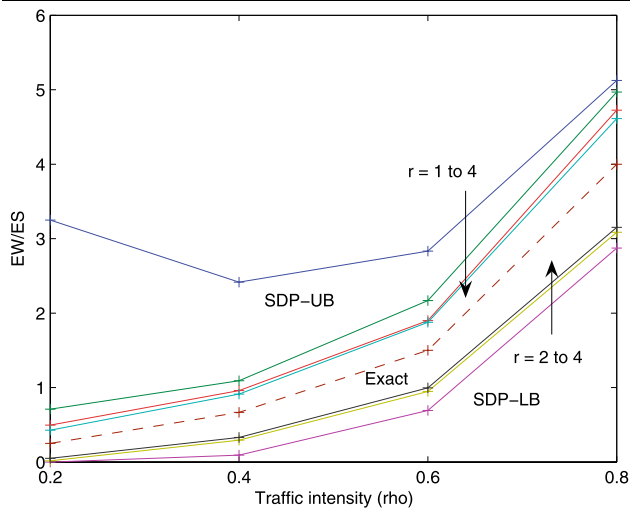


Fig. 2 Bounds on $E[W]/E[S]$ in the M/M/1 queue for $r = 1$ to $r = 4$

From the results, we observe that:

- With only two known moments ($r = 1$), we get exactly Kingman’s upper bound for $E[W]$. Furthermore, the two moment lower bound on $E[W]$ is zero (cf. [11]).
- By increasing the order of the relaxation, we clearly get tighter lower and upper bounds on the four moments of W . The relative error on the upper bound for the first moment decreases from around 90% ($r = 1$) to around 9% ($r = 5$). The results are clearly indicative of the potential of the approach.

3.1.2 Computational results for M/M/1 queue

For a GI/GI/1 queue, it is possible to tighten the bounds by directly working with the equality:

$$W \stackrel{d}{=} (W + S - T)_+,$$

and using the additional information that $S, T \geq 0$. We compare the results from the first four semidefinite relaxations with the exact expected waiting time for a M/M/1 queue. The results are plotted in Fig. 2 for four different values of $\rho = 0.2, 0.4, 0.6, 0.8$ ranging from light to heavy traffic conditions. It is clear that by increasing the order of the semidefinite relaxation, we improve on the bound. Also for the semidefinite relaxation of order $r > 1$, the expected waiting time is clearly increasing with ρ which is not true for Kingman’s bound ($r = 1$).

In fact, Kingman’s bound in the GI/GI/1 queue has been tightened by Daley [7] by using the additional inequality:

$$E[(T - S - W)_+^2] \geq E[T^2](1 - \rho)^2.$$

Such a moment inequality, simply reduces to an additional linear constraint in the semidefinite optimization problem.

We can similarly recover Daley’s bound using the proposed approach.

3.2 Bounds on tail probability of W

In this section, we derive moment based bounds on the probability that W exceeds a given positive value w i.e., $P(W \geq w)$. Let

$$s_1(W, X) := W \geq 0, \quad s_2(W, X) := W + X \geq 0 \quad \text{and}$$

$$s_3(W, X) := W - w \geq 0.$$

Define:

$$\mathcal{S}_1 = \{(W, X) \mid W + X \leq 0, W \geq w\},$$

$$\mathcal{S}_2 = \{(W, X) \mid W + X \geq 0, W \geq w\},$$

$$\mathcal{S}_3 = \{(W, X) \mid W \geq 0, W + X \leq 0, W < w\},$$

$$\mathcal{S}_4 = \{(W, X) \mid W \geq 0, W + X \geq 0, W < w\},$$

where $g(W, X) = 0$ over \mathcal{S}_1 and \mathcal{S}_3 and $g(W, X) = W + X$ over \mathcal{S}_2 and \mathcal{S}_4 . We then write the order r relaxation as follows.

Proposition 7 Given $E[X^\beta] = m^\beta < \infty$ for $\beta = 1, \dots, 2r$, an upper/lower bound on $P(W \geq w)$ in a Lindley process is obtained by solving the semidefinite optimization problem:

$$(\mathbb{P}_r^s) \quad \sup/\inf \quad x_1^{00} + x_2^{00}$$

$$\text{s.t.} \quad x_1^{\alpha 0} + x_3^{\alpha 0} - \sum_{\beta=0}^{\alpha-1} \binom{\alpha}{\beta} (x_2^{\beta, \alpha-\beta} + x_4^{\beta, \alpha-\beta}) = 0,$$

$$\forall \alpha = 1, \dots, 2r,$$

$$\sum_{k=1}^4 x_k^{\alpha \beta} - m^\beta \left(\sum_{k=1}^4 x_k^{\alpha 0} \right) = 0,$$

$$\forall \alpha + \beta = 1, \dots, 2r,$$

$$\sum_{k=1}^4 x_k^{00} = 1,$$

$$\mathbf{M}_r(\mathbf{x}_k) \geq \mathbf{0}, \quad k = 1, \dots, 4,$$

$$\mathbf{M}_{r-1}(s_1, \mathbf{x}_3), \mathbf{M}_{r-1}(s_1, \mathbf{x}_4) \geq \mathbf{0},$$

$$\mathbf{M}_{r-1}(-s_2, \mathbf{x}_1), \mathbf{M}_{r-1}(s_2, \mathbf{x}_2),$$

$$\mathbf{M}_{r-1}(-s_2, \mathbf{x}_3), \mathbf{M}_{r-1}(s_2, \mathbf{x}_4) \geq \mathbf{0},$$

$$\mathbf{M}_{r-1}(s_3, \mathbf{x}_1), \mathbf{M}_{r-1}(s_3, \mathbf{x}_2),$$

$$\mathbf{M}_{r-1}(-s_3, \mathbf{x}_3), \mathbf{M}_{r-1}(-s_3, \mathbf{x}_4) \geq \mathbf{0}.$$

Table 3 Bounds on $P(W \geq w)$ in the Lindley process

w	0.25	0.5	0.75	1	1.5	2
\mathbb{P}_1^s -UB	1.000	0.736	0.555	0.438	0.306	0.235
\mathbb{P}_2^s -UB	0.901	0.626	0.445	0.302	0.127	0.052
\mathbb{P}_3^s -UB	0.690	0.505	0.359	0.236	0.103	0.049
\mathbb{P}_4^s -UB	0.617	0.454	0.307	0.201	0.090	0.040
Simulation	0.319	0.205	0.125	0.076	0.028	0.010
\mathbb{P}_4^s -LB	0.103	0.056	0.027	0.012	0.003	0.002
\mathbb{P}_3^s -LB	0.074	0.035	0.017	0.007	0.003	0.001
\mathbb{P}_2^s -LB	0.043	0.010	0.000	0.000	0.000	0.000
\mathbb{P}_1^s -LB	0.000	0.000	0.000	0.000	0.000	0.000

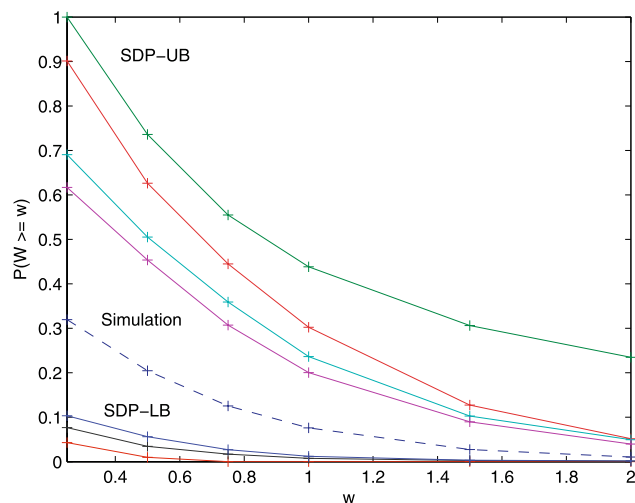


Fig. 3 Bounds on $P(W \geq w)$ in the Lindley process

3.2.1 Computational results

As before, we consider a Lindley process for a normally distributed random variable X with $E[X] = -0.25$ and $\sigma^2 = 0.25$. The simulated value obtained by over 500000 replications of X is compared with the semidefinite formulations \mathbb{P}_1^s to \mathbb{P}_4^s . The tail probabilities $P(W \geq w)$ were evaluated for $w = 0.25, 0.5, 0.75, 1, 1.5, 2$. The results are displayed in Table 3 and Fig. 3.

From the results, we observe that:

- By increasing the order of the relaxation, we obtain tighter lower and upper bounds on the probability $P(W \geq w)$. The relative errors are typically larger for smaller values of w .

4 The GI/GI/c queue

Consider a c server FCFS queueing system with renewal arrival process and i.i.d service times. The dynamics of the GI/GI/c queueing system is characterized by the workload process vector $\mathbf{W}_n := (W_{n1}, W_{n2}, \dots, W_{nc})$ with $W_{n1} \leq W_{n2} \leq \dots \leq W_{nc}$. The workload vector is defined by the recursion:

$$\mathbf{W}_{n+1} = \mathcal{R}((\mathbf{W}_n + S_n \mathbf{e}_1 - T_n \mathbf{e})_+) \quad \text{for } n = 0, 1, 2, \dots$$

The service times S_n are i.i.d random variables ($S_n \sim S$) and interarrival times T_n are i.i.d random variables ($T_n \sim T$). Under the assumption that $\rho_c = \lambda/(c\mu) < 1$, the workload process is known to be the unique solution to the recursive equation (cf. Kiefer and Wolfowitz [9]):

$$\mathbf{W} \stackrel{d}{=} \mathcal{R}((\mathbf{W} + S \mathbf{e}_1 - T \mathbf{e})_+),$$

where the equality is in distribution and S and T are independent of \mathbf{W} . The first component of \mathbf{W} , i.e., W_1 is the steady-state waiting time for the customer in the multi-server queue.

For the purpose of exposition, we focus on the GI/GI/2 queue wherein:

$$(W_1, W_2) \stackrel{d}{=} \mathcal{R}((W_1 + S - T)_+, (W_2 - T)_+). \tag{30}$$

Proposition 8 For a GI/GI/2 queue, the function $\mathcal{R}((W_1 + S - T)_+, (W_2 - T)_+)$ is a piecewise linear function over five polyhedral sets.

Proof For $c = 2$, the function is a piecewise linear function over five polyhedral sets:

$$g(\cdot) = \begin{cases} (W_2 - T, W_1 + S - T) & \text{if } \mathcal{S}_1 = \{(W_1, W_2, S, T) \in \mathfrak{R}_+^4 \mid W_1 + S - T \geq W_2 - T \geq 0\}, \\ (0, W_1 + S - T) & \text{if } \mathcal{S}_2 = \{(W_1, W_2, S, T) \in \mathfrak{R}_+^4 \mid W_1 + S - T \geq 0 \geq W_2 - T\}, \\ (W_1 + S - T, W_2 - T) & \text{if } \mathcal{S}_3 = \{(W_1, W_2, S, T) \in \mathfrak{R}_+^4 \mid W_2 - T \geq W_1 + S - T \geq 0\}, \\ (0, W_2 - T) & \text{if } \mathcal{S}_4 = \{(W_1, W_2, S, T) \in \mathfrak{R}_+^4 \mid W_2 - T \geq 0 \geq W_1 + S - T\}, \\ (0, 0) & \text{if } \mathcal{S}_5 = \{(W_1, W_2, S, T) \in \mathfrak{R}_+^4 \mid 0 \geq W_1 + S - T, 0 \geq W_2 - T\}. \end{cases}$$

□

In general for a c server queue, this is piecewise linear function over $\frac{c(c+3)}{2}$ polyhedral sets. This can be observed by noting that there are c positions for 0 relative to the ordered values $W_2 - T \leq \dots \leq W_c - T$. If we index each of these positions as $i = 1, \dots, c$, then there are $c - i + 2$ positions for the value of $W_1 + S - T$ relative to the ordered values of $0 \leq W_{i+1} - T \leq \dots \leq W_c - T$. This gives a total of $\frac{c(c+3)}{2}$ polyhedral sets.

4.1 Bounds on moments of waiting time

Let

$$s_1(\cdot) := W_1 + S - T \geq 0, \quad s_2(\cdot) := W_2 - T \geq 0 \quad \text{and}$$

$$s_3(\cdot) := W_1 - W_2 + S \geq 0.$$

Furthermore, let

$$s_4(\cdot) := W_1 \geq 0, \quad s_5(\cdot) := W_2 \geq 0, \quad s_6(\cdot) := S \geq 0 \quad \text{and} \\ s_7(\cdot) := T \geq 0.$$

We then write the order r relaxation as follows.

Proposition 9 *Given $E[S^{\beta_1} T^{\beta_2}] = m^{\beta_1 \beta_2} < \infty$ for $\beta_1 + \beta_2 = 1, \dots, 2r$, an upper/lower bound on the p th moment of the waiting time in a GI/GI/2 queue for $p < 2r$ is obtained by solving the semidefinite optimization problem:*

$$(\mathbb{P}_r^s) \quad \sup/\inf \sum_{k=1}^5 x_k^{p000} \\ \text{s.t.} \quad \sum_{k=1}^5 x_k^{\alpha_1 \alpha_2 00} - \sum_{j=0}^{\alpha_2} \sum_{k=0}^{\alpha_2-j} \sum_{r=0}^{\alpha_1} (-1)^{\alpha_2-j-k+\alpha_1-r} \binom{\alpha_2}{j} \binom{\alpha_2-j}{k} \binom{\alpha_1}{r} x_1^{j,r,k,\alpha_1-j-k+\alpha_2-r} \\ - \sum_{j=0}^{\alpha_1} \sum_{k=0}^{\alpha_1-j} \sum_{r=0}^{\alpha_2} (-1)^{\alpha_1-j-k+\alpha_2-r} \binom{\alpha_1}{j} \binom{\alpha_1-j}{k} \binom{\alpha_2}{r} x_3^{j,r,k,\alpha_1-j-k+\alpha_2-r} \\ - 0^{\alpha_1} \sum_{j=0}^{\alpha_2} \sum_{k=0}^{\alpha_2-j} (-1)^{\alpha_2-j-k} \binom{\alpha_2}{j} \binom{\alpha_2-j}{k} x_2^{j,0,k,\alpha_2-j-k} \\ - 0^{\alpha_1} \sum_{r=0}^{\alpha_2} (-1)^{\alpha_2-r} \binom{\alpha_2}{r} x_4^{0,r,0,\alpha_2-r} = 0, \quad \forall \alpha_1 + \alpha_2 = 1, \dots, 2r, \\ \sum_{k=1}^5 x_k^{\alpha_1 \alpha_2 \beta_1 \beta_2} - m^{\beta_1 \beta_2} \sum_{k=1}^5 x_k^{\alpha_1 \alpha_2 00} = 0, \quad \forall \alpha_1 + \alpha_2 + \beta_1 + \beta_2 = 1, \dots, 2r, \\ \sum_{k=1}^5 x_k^{0000} = 1, \\ \mathbf{M}_r(\mathbf{x}_k) \geq \mathbf{0}, \quad k = 1, \dots, 5, \\ \mathbf{M}_{r-1}(s_3, \mathbf{x}_1), \mathbf{M}_{r-1}(s_2, \mathbf{x}_1), \mathbf{M}_{r-1}(s_1, \mathbf{x}_2), \mathbf{M}_{r-1}(-s_2, \mathbf{x}_2) \geq \mathbf{0}, \\ \mathbf{M}_{r-1}(-s_3, \mathbf{x}_3), \mathbf{M}_{r-1}(s_1, \mathbf{x}_3), \mathbf{M}_{r-1}(s_2, \mathbf{x}_4), \mathbf{M}_{r-1}(-s_1, \mathbf{x}_4) \geq \mathbf{0}, \\ \mathbf{M}_{r-1}(-s_1, \mathbf{x}_5), \mathbf{M}_{r-1}(-s_2, \mathbf{x}_5) \geq \mathbf{0}, \\ \mathbf{M}_{r-1}(s_4, \mathbf{x}_k), \mathbf{M}_{r-1}(s_5, \mathbf{x}_k), \mathbf{M}_{r-1}(s_6, \mathbf{x}_k), \mathbf{M}_{r-1}(s_7, \mathbf{x}_k) \geq \mathbf{0}, \quad k = 1, \dots, 5.$$

For the GI/GI/2 queue, the first constraint in Proposition 9 is obtained by equating the moments in (30):

$$E[W_1^{\alpha_1} W_2^{\alpha_2}] = E[(W_2 - T)^{\alpha_1} (W_1 + S - T)^{\alpha_2} | \mathcal{S}_1] P[\mathcal{S}_1] \\ + E[0^{\alpha_1} (W_1 + S - T)^{\alpha_2} | \mathcal{S}_2] P[\mathcal{S}_2]$$

$$+ E[(W_1 + S - T)^{\alpha_1} (W_2 - T)^{\alpha_2} | \mathcal{S}_3] P[\mathcal{S}_3] \\ + E[0^{\alpha_1} (W_2 - T)^{\alpha_2} | \mathcal{S}_4] P[\mathcal{S}_4],$$

and using binomial expansions.

4.1.1 Computational results

We now evaluate the quality of the bounds by comparing the results with the M/M/2 queue. We focus on the dimensionless quantity $E[W_1]/E[S]$. For the M/M/2 queue with $\rho = \lambda/(2\mu) < 1$, the exact value for the expected waiting time is:

$$E[W_1]/E[S] = \frac{\rho^2}{1 - \rho^2}. \tag{31}$$

The two server upper bound obtained by Kingman [11] for the M/M/2 queue reduces to:

$$E[W_1]/E[S] = \frac{1 + 3\rho^2}{4\rho(1 - \rho)}. \tag{32}$$

We consider four different values of $\rho = 0.2, 0.4, 0.6, 0.8$ ranging from light to heavy traffic conditions. The upper and lower bounds on $E[W_1]/E[S]$ computed using the semidefinite formulations are displayed in Table 4 and Fig. 4.

We can observe the following from the results:

Table 4 Bounds on $E[W_1]/E[S]$ in the M/M/2 queue

ρ	0.2	0.4	0.6	0.8
Kingman-UB	1.750	1.542	2.167	4.563
\mathbb{P}_1^s -UB	1.687	1.375	1.792	3.563
\mathbb{P}_2^s -UB	0.262	0.527	1.162	2.877
\mathbb{P}_3^s -UB	0.155	0.380	0.887	2.303
Exact	0.042	0.191	0.563	1.778
\mathbb{P}_3^s -LB	0.000	0.035	0.245	1.180
\mathbb{P}_2^s -LB	0.000	0.000	0.109	1.015
\mathbb{P}_1^s -LB	0.000	0.000	0.000	0.000

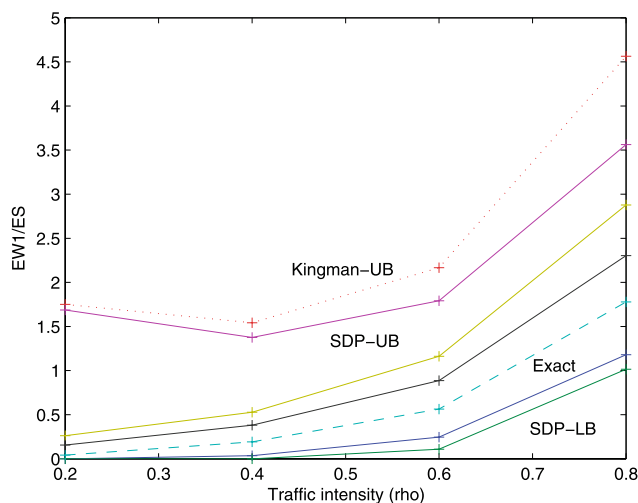


Fig. 4 Bounds on $E[W_1]/E[S]$ in the M/M/2 queue

- In the multi-server case the first order relaxation given only two moments is in fact tighter than Kingman’s upper bound [11]. While the bound in [11] is based on exploiting the symmetry of the sum and sum of squares of the components of the workload vector, our formulation in Proposition 9 uses the complete structure of the workload vector. An interesting open problem is to solve the first order semidefinite relaxation for the GI/GI/c queue in closed form, if possible.
- As in the single server case, we get tighter lower and upper bounds by increasing the order of the relaxation. The size of the semidefinite problems, however increases rapidly with number of servers and the order of the relaxation.

5 Conclusions

In this paper, we have proposed a semidefinite optimization approach to compute steady-state distributions with a focus on queueing applications. The algorithmic approach systematically generalizes some of the known two moment bounds to higher order moment bounds for the expected waiting time in queues. The numerical results for the GI/GI/1 queue and the GI/GI/2 queue are promising. It is natural to study several extensions of this approach to tandem queues, fork-join queues and queueing networks.

Acknowledgements We would like to thank two anonymous referees for their valuable and insightful comments. We would particularly like to thank the referee who pointed out the $\frac{c(c+3)}{2}$ decomposition for the GI/GI/c queue.

References

1. Asmussen, S.: Applied Probability and Queues, 2nd edn. Springer, New York (2003)
2. Bertsimas, D., Natarajan, K., Teo, C.P.: Probabilistic combinatorial optimization: moments, semidefinite programming and asymptotic bounds. *SIAM J. Optim.* **15**(1), 185–209 (2004)
3. Bertsimas, D., Natarajan, K., Teo, C.P.: Persistence in discrete optimization under data uncertainty. *Math. Program. Ser. B* **108**(2), 251–274 (2006)
4. Bertsimas, D., Popescu, I.: On the relation between option and stock prices: a convex optimization approach. *Oper. Res.* **50**, 358–374 (2002)
5. Bertsimas, D., Popescu, I.: Optimal inequalities in probability theory: a convex optimization approach. *SIAM J. Optim.* **15**(3), 780–804 (2005)
6. Borovkov, A.A., Foss, S.G.: Stochastically recursive sequences and their generalizations. *Sib. Adv. Math.* **2**, 16–81 (1992)
7. Daley, D.J.: Inequalities for moments of tails of random variables with queueing applications. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **41**, 139–143 (1977)
8. Daley, D.J.: Some results for the mean waiting-time and workload in GI/GI/k queues. In: Dshalalow, J.H. (ed.) *Frontiers in Queueing*, pp. 35–59. CRC Press, Boca Raton (1997)

9. Kiefer, J., Wolfowitz, J.: On the theory of queues with many servers. *Trans. Am. Math. Soc.* **78**, 1–18 (1955)
10. Kingman, J.F.C.: Some inequalities for the GI/G/1 queue. *Biometrika* **49**, 315–324 (1962)
11. Kingman, J.F.C.: Inequalities in the theory of queues. *J. Roy. Stat. Soc. Ser. B* **32**, 102–110 (1970)
12. Lasserre, J.B.: Bounds on measures satisfying moment conditions. *Ann. Appl. Probab.* **12**, 1114–1137 (2002)
13. Lasserre, J.B., Hernández-Lerma, O.: *Markov Chains and Invariant Probabilities*. Progress in Mathematics, vol. 211. Birkhäuser, Basel (2003)
14. Lindley, D.V.: On the theory of queues with a single server. *Proc. Camb. Philos. Soc.* **48**, 277–289 (1952)
15. Müller, A., Stoyan, D.: *Comparison Methods for Stochastic Models and Risks*. Wiley, New York (2002)
16. SeDuMi version 1.1. Available from <http://sedumi.mcmaster.ca/>
17. Wolff, R.W., Wang, C.L.: Idle period approximations and bounds for the GI/GI/1 queue. *Adv. Appl. Probab.* **35**, 773–792 (2003)
18. Zuluaga, L., Pena, J.F.: A conic programming approach to generalized Tchebycheff inequalities. *Math. Oper. Res.* **30**(2), 369–388 (2005)