The delay of open Markovian queueing networks: Uniform functional bounds, heavy traffic pole multiplicities, and stability*†

C. Humes, Jr.,‡ J. Ou§ and P. R. Kumar¶

Abstract

For open Markovian queueing networks, we study the functional dependence of the mean number in the system (and thus also the mean delay since it is proportional to it by Little's Theorem) on the arrival rate or load factor. We obtain linear programs (LPs) which provide bounds on the pole multiplicity M of the mean number in the system, and automatically obtain lower and upper bounds on the coefficients $\{C_i\}$ of the expansion $\frac{\rho C_M}{(1-\rho)^M} + \frac{\rho C_{M-1}}{(1-\rho)^{M-1}} + \cdots + \frac{\rho C_1}{(1-\rho)} + \rho C_0$, where ρ is the load factor, which are valid for all $\rho \in [0,1)$. Our LPs can thus establish the stability of open networks for *all* arrival rates within capacity, while providing uniformly bounding functional expansions for the mean delay, valid for all arrival rates in the capacity region. The coefficients $\{C_i\}$ can be optimized to provide the best bound at any desired value of the load factor, while still maintaining its validity for all $\rho \in [0,1)$. While the above LPs feature $\frac{L(L+1)(M+1)}{2}$ variables where L is the number of buffers in the network, for balanced systems we further provide a lower dimensional LP featuring just $\frac{S(S+1)}{2}$ variables, where S is the number of stations in the network. This bound asymptotically dominates in heavy traffic a bound obtainable from the Pollaczek-Khintchine formula, and can capture interactions between multiple bottleneck stations in heavy traffic. We

*Please address all correspondence to the third author.

†The research reported here has been supported by the U.S. Army Research Office under Contract No. DAAH-04-95-1-0090, the National Science Foundation under Grant No. ECS-94-03571, the Joint Services Electronics Program under Contract No. N00014-90-J-1270, FAPESP grants 93/0603-1 and 94/1376-1, and by CNPq and IBM-Brasil.

‡Departamento de Ciencia da Computacao, Instituto de Mathematica e Estatistica, University of Sao Paulo, Brazil. This work was done while the author was visiting the University of Illinois.

§Department of Decision Sciences, Faculty of Business Administration, National University of Singapore, Singapore 0511. This work was done while the author was visiting the University of Illinois.

¶Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois, 1308 West Main St., Urbana, IL 61801.

also provide an explicit upper bound for all scheduling policies in acyclic networks, and for the FBFS policy in open re-entrant lines.

Key Words: Queueing networks, open networks, performance evaluation, scheduling policies, delay, stability, heavy traffic behavior.

1 Introduction

We address the problem of determining the delay of open Markovian queueing networks as a function of the arrival rate. A queueing network is *open* if customers arrive to the system from the outside and eventually depart. It is said to be Markovian if it has Bernoulli routing, and service and interarrival times are exponentially distributed.

For such networks, the mean delay $D^u(\lambda)$ under a scheduling policy u depends on the arrival rate λ of customers from the outside. If the mean delay $D^u(\lambda)$ is finite, we say that the system is *stable* under the scheduling policy u at the arrival rate λ. Let λ^* denote the supremal arrival rate that the system is stable, if properly scheduled. Given a particular value of $\lambda < \lambda^*$, we have provided in Kumar and Meyn (1995) a linear program (LP) that can establish stability at that specific value λ of the arrival rate. Also, LPs have been provided in Kumar and Kumar (1994) and Bertsimas, Paschalidis and Tsitsiklis (1994) that provide upper and lower bounds on the mean delay, again for specific values of λ. In Kumar and Meyn (1996) we have shown that the LPs for stability and performance are the duals of each other.

The above results however only provide bounds on the performance, pointwise for each specific value λ of the arrival rate, or stability pointwise at each specific value of λ. In this paper, our objective is to establish the behavior of networks over *all* λ's. We study the functional dependence of $D^u(\lambda)$ on λ for all arrival rates $0 \leq \lambda < \lambda^*$. If a system is stable for all λ in $[0, \lambda^*)$ for a particular scheduling policy u, we say that u is *stable throughout the capacity region*. After stability, we are interested in how the delay $D^u(\lambda)$ varies with λ in

$[0,\lambda^*)$. If $f(\lambda) \le D^u(\lambda) \le g(\lambda)$ for all $0 \le \lambda < \lambda^*$, then we call f and g *lower and upper functional bounds* on D^u, respectively. Of particular interest is the behavior of the delay in heavy traffic as $\lambda \nearrow \lambda^*$. If $D^u(\lambda) = \frac{C_M + o(1)}{(\lambda^* - \lambda)^M}$, where $C_M > 0$, then we say that M is the *pole multiplicity* of the delay in heavy traffic, and C_M is the growth constant.

We obtain LPs which provide bounds on the pole multiplicity M, and automatically obtain lower and upper bounds on the coefficients $\{C_i\}$ of the *functional expansion* $\frac{\rho C_M}{(1-\rho)^M} + \frac{\rho C_{M-1}}{(1-\rho)^{M-1}} + \cdots + \frac{\rho C_1}{(1-\rho)} + \rho C_0$ for the mean number in the system. Above, $\rho := \frac{\lambda}{\lambda^*}$ is the nominal load on the network. By this approach, we obtain an LP test that establishes stability throughout the capacity region $0 \le \rho < 1$, and simultaneously obtains uniformly bounding functional expansions valid throughout the capacity region. The functional bound can be optimized to provide the best bound at any particular value of ρ, while still maintaining its uniform validity for all $\rho \in [0,1)$. All results in this paper will be furnished in terms of the mean number in the system, and are easily translated by Little's Theorem into corresponding results concerning the mean delay by simply multiplying by λ^{-1}.

Our results develop lower bounds valid either for the class of all scheduling policies, or for a particular buffer priority policy. Our upper bounds are developed either for the class of all non-idling scheduling policies (since allowing idling allows infinite delay), or for a particular buffer priority policy.

For balanced networks where all stations are equally loaded, we further obtain a reduced dimensional LP, for a functional lower bound on the delay, consisting of just $\frac{S(S+1)}{2}$ variables rather than the $\frac{L(L+1)(M+1)}{2}$ variables for the LPs mentioned above. Here S is the number of stations in the networks. In many applications it is far smaller than L, the number of buffers in the network. We show that for re-entrant lines this functional bound asymptotically dominates in heavy traffic a lower bound that can be derived from the Pollaczek-Khintchine formula, and that it can capture nontrivial interactions between *multiple* bottleneck stations in heavy traffic.

We also study the possibility of obtaining upper bounds through our LP approach for all systems and scheduling policies known to be stable throughout the capacity region. We obtain explicit upper bounds on the mean delay of all non-idling scheduling policies in acyclic networks, and the mean delay of the First Buffer First Serve (FBFS) buffer priority policy for open re-entrant lines. We also show that our approach cannot establish the stability of the Last Buffer First Serve (LBFS) policy.

By providing uniformly bounding functional expansions of the form $\frac{\rho C_M}{(1-\rho)^M} + \frac{\rho C_{M-1}}{(1-\rho)^{M-1}} + \cdots + \frac{\rho C_1}{(1-\rho)} + \rho C_0$, valid for *all* arrival rates, which are obtained merely by solving LPs for the coefficients, these procedures can provide useful information with little computational effort for non-product form networks previously considered intractable.

Another value of our results is that they allow us to better comprehend what the linear programming approach to stability and performance is really providing us. In all examples studied by us, the functional bounds yield the same heavy traffic growth constant C_M as the pointwise bounds do numerically. Thus, through studying the closed form expressions provided by the functional bounds, we are able to assess the quality of the original pointwise bounds themselves.

The issue of bounding functional expansions seems to be relatively unexplored, though it is quite interesting for applications. We are not aware of prior work in this area. For single server queues, Gong and Hu (1992) [5] have studied Maclaurin series solutions of Lindley's equations. Girish and Hu (1995) have investigated tandem queues by this procedure. Gong and Yang (1995) have also investigated the problem of extrapolating the performance curve from a known set of explicit points on it via Pade approximations.

Our specific results are as follows:

(i) We obtain LPs which establish stability throughout the capacity region, provide lower and upper bounds on pole multiplicity, and provide uniformly bounding functional expansions for the mean delay that can be tailored for best fit at any particular loading level (Theorems 3, 4 and 5).

(ii) We provide a reduced dimensional reduced LP which gives a functional lower bound on the delay of all scheduling policies for balanced systems (Theorem 6). We show that for re-entrant lines this bound asymptotically dominates in heavy traffic a bound that can be derived from the Pollaczek-Khintchine formula (Theorem 7), and that it can capture multiple bottleneck interactions in heavy traffic (Example 4).

(iii) We provide an explicit upper bound on the delay of the FBFS policy (Theorems 8 and 9).

(iv) We provide an explicit upper bound on the delay for all acyclic networks (Theorem 10).

The rest of this paper is organized as follows. In Section 2 we begin with a review of the pointwise bounds for each fixed value of λ. In Section 3, we sow the seed of the idea for obtaining functional expansions. In Section 4, we investigate the feasible solutions of the dual LPs, exhibiting a key fundamental identity and a key inequality. In Section 5 we provide the uniformly bounding functional expansions and LPs for their coefficients. In Section 6, we provide a reduced dimensional LP for the functional lower bound for balanced systems, and in Section 7 we show that for re-entrant lines this bound asymptotically dominates in heavy traffic a bound that can be derived from the Pollaczek-Khintchine formula, and that it can capture nontrivial interactions between multiple bottleneck stations in heavy traffic. In Section 8 we turn to upper bounds, and in Sections 9 and 10 we obtain the upper bounds for the FBFS policy, and for all acyclic networks, respectively.

2 Pointwise bounds for a fixed arrival rate

We shall carry out our development on the following model of a Markovian queueing network; see Figure 1. There are S stations labeled $\{1, 2, \ldots, S\}$, and L buffers labeled $\{b_1, b_2, \ldots, b_L\}$. Buffer b_i is served by station $\sigma(i) \in \{1, 2, \ldots, S\}$. Customers arrive to the system as a Poisson process of rate λ. Upon arrival they join buffer b_i with probability p_{*i}. Customers in buffer b_i require an exponentially distributed service time with mean $1/\mu_i$ from station $\sigma(i)$. After completing service at b_i, they move to buffer b_j with probability p_{ij}, or leave the system with probability p_{i*}. We call such a system an *open Markovian queueing network.*

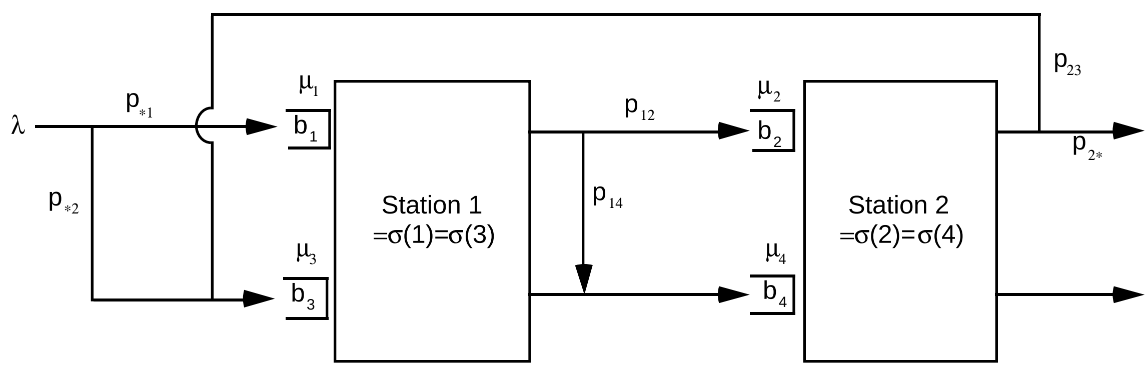

Figure 1: An open Markovian queueing network.

Let $w_i(t) := 1$ if station $\sigma(i)$ is busy serving a customer in b_i at time t; and $:= 0$ otherwise. We denote the number of customers in buffer b_i at time t by $x_i(t)$, and by $x(t) = (x_1(t), x_2(t), \ldots, x_L(t))^T$ the corresponding vector. We assume that all stochastic processes are right continuous with left limits.

We convert this system to discrete time by the method of uniformization, see Lippman (1975). We normalize time so that $\lambda + \sum_i \mu_i = 1$. Then we suppose that every buffer has either a real or virtual customer in service, and sample the system at the sequence $\{\tau_n\}$

of random times which consists of all arrival times, real service completion times, and virtual service completion times. We denote $x(\tau_n)$ and $w(\tau_n)$ by $x(n)$ and $w(n)$, for brevity.

Throughout we assume that the scheduling policy employed is *non-idling*, i.e., a station cannot stay idle if there is work for it. (The lower bound for the class of all non-idling policies also applies to the class of all scheduling policies, as noted in the proof of Theorem 5.) Quantitatively, the property of non-idling can be expressed as $x_i(n) \geq 1 \Rightarrow \sum_{j \in \sigma(i)} w_j(n) = 1$, where, by "$j \in \sigma$" we mean $\{j : \sigma(j) = \sigma\}$. A consequence of this is,

$$x_i(n) = \sum_{j \in \sigma(i)} w_j(n) x_i(n). \tag{1}$$

In addition, one always has

$$x_i(n) \geq \sum_{j \in \sigma} w_j(n) x_i(n) \text{ for all } \sigma. \tag{2}$$

On occasion, in the sequel, we will consider *buffer priority policies*. These are defined by a buffer priority ordering $\theta = (\theta(1), \ldots, \theta(L))$ which is a permutation of $\{1, 2, \ldots, L\}$. If buffers b_i and b_j share the same station, i.e., $\sigma(i) = \sigma(j)$, then priority is given to b_j over b_i if $\theta(j) < \theta(i)$. The priority discipline is preemptive resume. As a consequence,

$$w_i(n) x_j(n) = 0 \text{ if } \sigma(i) = \sigma(j) \text{ and } \theta(j) < \theta(i). \tag{3}$$

Consider now a scheduling policy that is *stationary*, i.e., $w_i(t)$ depends only on $x(t)$, and non-idling. If the resulting system is stable with a finite second moment for $x(n)$ in steady-state, then

$$E[E[x^T(n+1) Q x(n+1) | \mathcal{F}_n] - x^T(n) Q x(n)] = 0 \text{ in steady state,}$$

where $\mathcal{F}_n$ is the past σ-algebra. Define the steady state expectations,

$$z_{ij} := E[w_i(n) x_j(n)], \tag{4}$$

$$\bar{x}_i := E(x_i(n)). \tag{5}$$

It has been shown in Kumar and Kumar (1994) and Bertsimas, Paschalidis and Tsitsiklis (1994) that by taking $Q = e_i e_j^T + e_j e_i^T$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ with a 1 in the ith place, one obtains the following set of equalities:

$$2\lambda p_{*i}\bar{x}_i - 2\mu_i z_{ii}(1-p_{ii}) + 2\sum_{j\neq i}\mu_j p_{ji} z_{ji} + 2\mu_i\rho_i(1-p_{ii}) = 0 \text{ for } i = 1, 2, \ldots, L, \quad (6)$$

$$\lambda(p_{*i}\bar{x}_j + p_{*j}\bar{x}_i) - \mu_i(p_{ij}\rho_i + z_{ij}) - \mu_j(p_{ji}\rho_j + z_{ji})$$
$$+\sum_k \mu_k(p_{ki}z_{kj} + p_{kj}z_{ki}) = 0 \text{ for all } 1 \leq i < j \leq L. \quad (7)$$

Above, the ρ_i's are the unique solution of the following *traffic equations*,

$$\lambda p_{*i} + \sum_{k=1}^{L}\mu_k p_{ki}\rho_k = \mu_i\rho_i, \text{ for } i = 1, 2, \ldots, L. \quad (8)$$

Note that ρ_i is the *nominal load* on station $\sigma(i)$ due to buffer b_i. It varies linearly with λ. Using (4,5), one can rewrite (1, 2, 3) as,

$$\bar{x}_i = \sum_{j\in\sigma(i)} z_{ji} \text{ for } 1 \leq i \leq L, \quad (9)$$

$$\bar{x}_i \geq \sum_{j\in\sigma} z_{ji} \text{ for } 1 \leq i \leq L \text{ and all } \sigma, \quad (10)$$

$$z_{ij} = 0 \text{ if } \sigma(i) = \sigma(j) \text{ and } \theta(j) < \theta(i). \quad (11)$$

It is shown in Kumar and Meyn (1996) that (6,7) are valid even if one only assumes finiteness of the *first* moment. There, by using some properties of the class of all non-idling policies, we have established the following Linear Program (LP) bounds, without even assuming finiteness of the first moment. They provide bounds, *pointwise* for each specific value of λ, and so we call them the *pointwise bound LPs.*

Theorem 1. The pointwise bound LPs. *Under any stationary non-idling scheduling policy, the mean number of customers in the system is bounded above by*

$$Max \sum_{i=1}^{L}\bar{x}_i, \quad (12)$$

subject to the constraints (6, 7, 9, 10), and

$$\bar{x}_i \geq 0, z_{ij} \geq 0 \text{ for } 1 \leq i,j \leq L. \tag{13}$$

The mean number of customers is bounded below by

$$Min \sum_{i=1}^{L} \bar{x}_i, \tag{14}$$

subject to the same constraints (6, 7, 9, 10, 13).

For buffer priority policies, the upper bound is valid only under a finite first moment assumption.

Theorem 2. Pointwise bound LPs for buffer priority policies. *Consider a buffer priority policy θ.*

(i) *The mean number of customers in steady state is bounded below by (14), subject to the constraints (6, 7, 9, 10, 13) and the buffer priority constraints (11).*

(ii) *If the mean number of customers in steady-state is finite, then it is bounded above by (12), subject to the same constraints (6, 7, 9, 10, 13, 11).*

3 Primal expansions

The above results provide LPs for obtaining *pointwise* bounds on the mean number in the system (or equivalently mean delay, through Little's Theorem) for each *fixed* value of the arrival rate λ. Our goal however is to study the *functional* dependence of the mean delay on the arrival rate λ.

Denote the nominal load on station σ by

$$\rho_\sigma := \sum_{i \in \sigma} \rho_i,$$

and the *nominal load on the system*, or more precisely on a *bottleneck station*, by

$$\rho \;\; := \;\; \mathrm{Max}_\sigma \rho_\sigma .$$

Both depend linearly on λ. The traffic *capacity* of the system is

$$\lambda^* \;\; := \;\; \sup\{\lambda : \mathrm{Max}_\sigma \rho_\sigma < 1\}.$$

Let ρ_i^* and ρ_σ^* denote the limiting values of ρ_i and ρ_σ when the arrival rate $\lambda \nearrow \lambda^*$.

To make explicit the dependence on the loading ρ (or equivalently the arrival rate λ), we will denote by $z_{ij}(\rho)$ and $\bar{x}_i(\rho)$ the expectations defined in $(4,5)$ when the arrival rate is $\lambda = \rho\lambda^*$.

We commence the development of functional dependence by examining the consequences of a truncated Laurent series type expansion for $z_{ij}(\rho)$. While this expansion itself may not be valid, we will later establish the implied results rigorously.

Truncated Laurent Series Condition (TLS)

For some $M \geq 1$, and constants $z_{ij}^{(0)}, z_{ij}^{(1)}, \ldots, z_{ij}^{(M)}$,

$$z_{ij}(\rho) \;\; = \;\; \frac{z_{ij}^{(M)}}{(1-\rho)^M} + \frac{z_{ij}^{(M-1)}}{(1-\rho)^{M-1}} + \cdots + + z_{ij}^{(0)} + O(1-\rho) \text{ for } 0 \leq \rho < 1. \tag{15}$$

Trivially, from (9), it follows that we must then have

$$\bar{x}_i(\rho) \;\; = \;\; \frac{\bar{x}_i^{(M)}}{(1-\rho)^M} + \frac{\bar{x}_i^{(M-1)}}{(1-\rho)^{M-1}} + \cdots + \bar{x}_i^0 + O(1-\rho), \text{ where} \tag{16}$$

$$\bar{x}_i^{(m)} \;\; = \;\; \sum_{j \in \sigma(i)} z_{ji}^{(m)}. \tag{17}$$

The following Lemma shows how to obtain the equality constraints satisfied by the coefficients $\{z_{ij}^{(m)}, \bar{x}_i^{(m)}\}$ of the above expansions.

Lemma 1. *Suppose a vector* $y(\rho)$ *satisfies*

(i) $y(\rho) = \sum_{m=0}^{M} \frac{y^{(m)}}{(1-\rho)} + O(1-\rho)$ *in* $0 \leq \rho < 1$,

(ii) $\rho A y(\rho) + B y(\rho) = c$.

Then the $\{y^{(m)}\}$ *satisfy the following equations:*

$$(A+B)y^{(0)} - Ay^{(1)} = c, \tag{18}$$

$$(A+B)y^{(m)} - Ay^{(m+1)} = 0 \text{ for } 1 \leq m \leq M-1, \tag{19}$$

$$(A+B)y^{(M)} = 0. \tag{20}$$

Proof. The results follow by writing $\rho = 1-(1-\rho)$, and matching coefficients of powers of $(1-\rho)$. □

From this one obtains the following constraints for $\{z_{ij}^{(m)}\}$.

Lemma 2. *Under Condition (TLS),* $\{z_{ij}^{(m)}, \bar{x}_i^{(m)}\}$ *satisfy the following for* $0 \leq m \leq M$*:*

$$2\lambda^* p_{*i}\bar{x}_i^{(m)} - 1(m \leq M-1)2\lambda^* p_{*i}\bar{x}_i^{(m+1)} - 2\mu_i z_{ii}^{(m)}(1-p_{ii}) + 2\sum_{j\neq i}\mu_j p_{ji} z_{ji}^{(m)} + 1(m=0)2\mu_i\rho_i^*(1-p_{ii}) = 0, \tag{21}$$

$$\begin{aligned}\lambda^* p_{*i}\bar{x}_j^{(m)} &- 1(m \leq M-1)\lambda^* p_{*i}\bar{x}_j^{(m+1)} + \lambda^* p_{*j}\bar{x}_j^{(m)} - 1(m \leq M-1)\lambda^* p_{*j}\bar{x}_j^{(m+1)} \\ &-1(m=0)\mu_i p_{ij}\rho_i^* - \mu_i z_{ij}^{(m)} - 1(m=0)\mu_j p_{ji}\rho_j^* - \mu_j z_{ji}^{(m)} \\ &+\sum_k \mu_k(p_{ki}z_{kj}^{(m)} + p_{kj}z_{ki}^{(m)}) = 0 \text{ for } j \geq i+1.\end{aligned} \tag{22}$$

Proof. The result is obtained by applying Lemma 1 to the constraints of the LP of Theorem 1. Noting $\lambda = \rho\lambda^*$ and $\rho_i = \rho\rho_i^*$, equation (6) can be written as

$$\rho[2\lambda^*\rho_{*i}x_i(\rho) + 2\mu_i\rho_i^*(1-p_{ii})] + [-2\mu_i(1-p_{ii})z_{ii}(\rho) + 2\sum_{j\neq i}\mu_j p_{ji}z_{ji}(\rho)] = 0. \tag{23}$$

Define the vectors $y(\rho) := (z_{1i}, z_{2i}, \ldots, z_{Li}, x_i, 1)^T$, $A := (0, \ldots, 0, 2\lambda^* p_{*i}, 2\mu_i \rho_i^*(1 - p_{ii}))$, and $B := (B_1, \ldots, B_L, 0, 0)$ with $B_j := 2\mu_j p_{ji}$ for $j \neq i$, $B_i := -2\mu_i(1 - p_{ii})$. Then (23) can be written as $\rho A y(\rho) + B y(\rho) = 0$. The equation (21) then follows from Lemma 1. Similarly (22) can be obtained. □

This suggests the possibility of directly bounding the coefficients $\{z_{ij}^{(m)}, \bar{x}_i^{(m)}\}$, and thus obtaining uniformly bounding functional expansions for the performance. One is thus led to consider the following LP, which we shall call the LP for the uniformly upper bounding functional expansion:

$$\text{Max} \sum_{i=1}^{L} \bar{x}_i^{(M)}$$

subject to the constraints (21, 22, 17) and

$$\bar{x}_i^{(m)} \geq \sum_{j \in \sigma} z_{ji}^{(m)}, \tag{24}$$

$$\bar{x}_i^{(m)} \geq 0. \tag{25}$$

(For $m \neq M$, the constraints (24,25) are not forced by the Condition (TLS)).

One hopes that if $\{\bar{x}_i^{(m)}\}$ is an optimal solution, then an uniformly upper bounding functional expansion of the following form is valid for all $0 \leq \lambda < \lambda^*$,

$$E|x(n)| \leq \sum_{i=1}^{L} \frac{\bar{x}_i^{(M)}}{(1-\rho)^M} + \sum_{i=1}^{L} \frac{\bar{x}_i^{(M-1)}}{(1-\rho)^{M-1}} + \ldots + \sum_{i=1}^{L} \bar{x}_i^{(0)} + O(1-\rho), \text{ where } |x| := \sum_{i=1}^{L} x_i.$$

However, all this is heavily contingent on the unjustified use of Condition (TLS). In fact, we have not even established stability of the system for all $\lambda \in [0, \lambda^*)$. Moreover, there is the issue of what is the appropriate value of M to choose. To resolve these issues and thus rigorously obtain functional bounds, we turn to the *duals* of the above LPs.

4 The fundamental identity and inequality

In Kumar and Meyn (1996) we have shown how to obtain Lyapunov functions that possess negative drift and thus establish stability, by taking the dual of the pointwise upper bound LP. Here we will employ duality in a different way to actually give a uniformly bounding functional *expansion* for the performance.

Associating the dual variables $(-\frac{1}{2}q_{ii}^{(m)})$ with (21), and $(-q_{ij}^{(m)})$ with (22), yields the following duals of the pointwise upper and lower bound LPs:

Dual of the LP for the uniformly upper bounding functional expansion

$$\text{Min } \sum_{i=1}^{L} \mu_i \rho_i^* (q_{ii}^{(0)} - \sum_j p_{ij} q_{ij}^{(0)}) \tag{26}$$

subject to:

$$\lambda^* \textstyle\sum_{i=1}^{L} p_{*i} \left(q_{ij}^{(m)} - 1(m \geq 1) q_{ij}^{(m-1)}\right) + \text{Max}_{i \in \sigma(j)} \mu_i \left(\displaystyle\sum_k p_{ik} q_{kj}^{(m)} - q_{ij}^{(m)}\right)$$
$$+ \textstyle\sum_{\sigma \neq \sigma(j)} \text{Max}_{i \in \sigma} \mu_i \left(\displaystyle\sum_k p_{ik} q_{kj}^{(m)} - q_{ij}^{(m)}\right)^+ \leq -1(m = M) \text{ for } 1 \leq j \leq L \text{ and } m \leq M. \tag{27}$$

Dual of the LP for the uniformly lower bounding functional expansion

$$\text{Max } \sum_{i=1}^{L} \mu_i \rho_i^* (-q_{ii}^{(0)} + \sum_j p_{ij} q_{ij}^{(0)}) \tag{28}$$

subject to:

$$\lambda^* \textstyle\sum_{i=1}^{L} p_{*i} \left(q_{ij}^{(m)} - 1(m \geq 1) q_{ij}^{(m-1)}\right) + \text{Max}_{i \in \sigma(j)} \mu_i \left(\displaystyle\sum_k p_{ik} q_{kj}^{(m)} - q_{ij}^{(m)}\right)$$
$$+ \textstyle\sum_{\sigma \neq \sigma(j)} \text{Max}_{i \in \sigma} \mu_i \left(\displaystyle\sum_k p_{ik} q_{kj}^{(m)} - q_{ij}^{(m)}\right)^+ \leq 1(m = M) \text{ for } 1 \leq j \leq L \text{ and } m \leq M. \tag{29}$$

The key to exploiting the properties of feasible solutions of these LPs is the following *fundamental identity*.

Lemma 3. The fundamental identity. *Let $Q = Q^T = [q_{ij}]$. Under any non-idling stationary scheduling policy the following identity holds:*

$$
\begin{aligned}
E\left[\frac{1}{2}x^T(t+1)Qx(t+1)\right] &= E\left[\frac{1}{2}x^T(t)Qx(t)\right] \\
&+E\sum_{j=1}^{L} x_j(t)\left[\lambda\sum_i p_{*i}q_{ij} + \sum_i \mu_i w_i(t)\left(\sum_k p_{ik}q_{kj} - q_{ij}\right)\right] \quad (30)\\
&+\frac{1}{2}\sum_i \lambda p_{*i}q_{ii} + \frac{1}{2}\sum_{i=1}^{L}\mu_i w_i(t)\left(q_{ii} - 2\sum_j p_{ij}q_{ij} + \sum_j p_{ij}q_{jj}\right).
\end{aligned}
$$

Proof. This is based on direct calculation, using

$$
\begin{aligned}
E\left[x^T(t+1)Qx(t+1)|\mathcal{F}_t\right] &= \sum_i \lambda p_{*i}(x(t)+e_i)^T Q(x(t)+e_i) \\
&+\sum_i \mu_i w_i(t)\sum_j p_{ij}(x(t)-e_i+e_j)^T Q(x(t)-e_i+e_j) \\
&+\sum_i \mu_i w_i(t) p_{i*}(x(t)-e_i)^T Q(x(t)-e_i). \quad \square
\end{aligned}
$$

From the fundamental identity we obtain the following *fundamental inequality*.

Lemma 4. The fundamental inequality.

(i) *Consider any non-idling scheduling policy. Let $Q^{(m)} = [q_{ij}^{(m)}]$, $0 \le m \le M$, be symmetric matrices which satisfy the constraints (27) and*

$$q_{ii}^{(m)} - 2\sum_j p_{ij}q_{ij}^{(m)} + \sum_j p_{ij}q_{jj}^{(m)} \ge 0 \text{ for } 1 \le i \le L. \quad (31)$$

Then

$$
\begin{aligned}
&E\left[\tfrac{1}{n}\textstyle\sum_{t=0}^{n-1}|x(t)|\right] + \lambda^*(1-\rho)E\sum_j \tfrac{1}{n}\sum_{t=0}^{n-1} x_j(t)\sum_i p_{*i}q_{ij}^{(M)} \\
&\le \textstyle\sum_{m=0}^{M}\sum_i \frac{\mu_i\rho_i}{(1-\rho)^m}\left(q_{ii}^{(M-m)} - \sum_j p_{ij}q_{ij}^{(M-m)}\right) \quad (32)\\
&+E\textstyle\sum_{m=0}^{M}\frac{x^T(0)Q^{(M-m)}x(0) - x^T(n)Q^{(M-m)}x(n)}{2n(1-\rho)^m} + o(1) \text{ (where } \lim_n o(1) = 0).
\end{aligned}
$$

(ii) *Consider any non-idling scheduling policy with* $\lim_n \frac{1}{n}\sum_{t=0}^{n-1} Ew_i(t) = \rho_i$ *for all* i. *Let* $Q^{(m)} = [q_{ij}^{(m)}]$, $0 \le m \le M$ *be symmetric matrices which satisfy (29). Then*

$$E\left[\frac{1}{n}\sum_{t=0}^{n-1}|x(t)|\right] - \lambda^*(1-\rho)E\sum_j \frac{1}{n}\sum_{t=0}^{n-1} x_j(t)\sum_i p_{*i}q_{ij}^{(M)}$$
$$\ge -\sum_{m=0}^{M}\sum_i \frac{\mu_i\rho_i}{(1-\rho)^m}\left(q_{ii}^{(M-m)} - \sum_j p_{ij}q_{ij}^{(M-m)}\right) \tag{33}$$
$$-E\sum_{m=0}^{M}\frac{x^T(0)Q^{(M-m)}x(0) - x^T(n)Q^{(M-m)}x(n)}{2n(1-\rho)^m} + o(1).$$

Proof. (i) With $Q = Q^{(M)}$, the term in (30) can be rewritten as

$$\lambda\sum_i p_{*i}q_{ij}^{(M)} = \lambda^*\sum_i p_{*i}q_{ij}^{(M)} - \lambda^*\sum_i p_{*i}q_{ij}^{(M-1)} + (\lambda-\lambda^*)\sum_i p_{*i}q_{ij}^{(M)} + \lambda^*\sum_i p_{*i}q_{ij}^{(M-1)}. \tag{34}$$

Applying the fundamental identity with $Q^{(M)}$ in place of Q thus gives,

$$E\left[\tfrac{1}{2}x^T(t+1)Q^{(M)}x(t+1)\right] = E\left[\tfrac{1}{2}x^T(t)Q^{(M)}x(t)\right]$$
$$+E\sum_{j=1}^{L} x_j(t)\left[\lambda^*\sum_i p_{*i}(q_{ij}^{(M)} - q_{ij}^{(M-1)}) + \sum_i \mu_i w_i(t)\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right)\right]$$
$$+(\lambda-\lambda^*)E\left[\sum_{j=1}^{L} x_j(t)\sum_{i=1}^{L} p_{*i}q_{ij}^{(M)}\right] + \lambda^*E\left[\sum_{j=1}^{L} x_j(t)\sum_{i=1}^{L} p_{*i}q_{ij}^{(M-1)}\right]$$
$$+\tfrac{1}{2}\sum_i \lambda p_{*i}q_{ii}^{(M)} + \tfrac{1}{2}E\sum_{i=1}^{L}\mu_i w_i(t)\left(q_{ii}^{(M)} - 2\sum_j p_{ij}q_{ij}^{(M)} + \sum_j p_{ij}q_{jj}^{(M)}\right). \tag{35}$$

Now note that whenever $x_j(t) > 0$ one of the quantities in $\{w_i(t) : i \in \sigma(j)\}$ is 1, while the rest are zero, due to the assumption that the scheduling policy is non-idling. Hence,

$$x_j(t)\sum_{i\in\sigma(j)}\mu_i w_i(t)\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right) \le x_j(t)\left[\operatorname*{Max}_{i\in\sigma(j)}\mu_i\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right)\right].$$

Also, if $\sigma \ne \sigma(j)$, then the $w_i(t)$'s with $i \in \sigma$ may be zero or one. Hence,

$$x_j(t)\sum_{i\in\sigma}\mu_i w_i(t)\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right) \le x_j(t)\left[\operatorname*{Max}_{i\in\sigma}\mu_i\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right)^+\right] \quad \text{for } \sigma \ne \sigma(j)$$

where $a^+ := \text{Max}(a,0)$. Hence

$$\begin{aligned} x_j(t)\sum_i \mu_i w_i(t)\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right) &= x_j(t)\left[\sum_{i\in\sigma(j)} \mu_i w_i(t)\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right)\right. \\ &\quad \left. + \sum_{\sigma\neq\sigma(j)}\sum_{i\in\sigma} \mu_i w_i(t)\left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right)\right] \\ &\leq x_j(t)\left[\underset{i\in\sigma(j)}{\text{Max}}\, \mu_i \left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right)\right. \\ &\quad \left. + \sum_{\sigma\neq\sigma(j)} \underset{i\in\sigma}{\text{Max}}\, \mu_i \left(\sum_k p_{ik}q_{kj}^{(M)} - q_{ij}^{(M)}\right)^+\right]. \end{aligned}$$

Substituting this in (35) and using (27) yields,

$$\begin{aligned} &E\left[\tfrac{1}{2}x^T(t+1)Q^{(M)}x(t+1)\right] \leq E\left[\tfrac{1}{2}x^T(t)Q^{(M)}x(t)\right] - E|x(t)| \\ &+(\lambda-\lambda^*)E\left[\textstyle\sum_{j=1}^L x_j(t)\sum_{i=1}^L p_{*i}q_{ij}^{(M)}\right] + \lambda^* E\left[\textstyle\sum_{j=1}^L x_j(t)\sum_{i=1}^L p_{*i}q_{ij}^{(M-1)}\right] \\ &+E\left[\tfrac{1}{2}\textstyle\sum_i \lambda p_{*i}q_{ii}^{(M)} + \tfrac{1}{2}E\sum_{i=1}^L \mu_i w_i(t)\left(q_{ii}^{(M)} - 2\sum_j p_{ij}q_{ij}^{(M)} + \sum_j p_{ij}q_{jj}^{(M)}\right)\right]. \end{aligned} \tag{36}$$

Now note that since ρ_i is the rate at which work arrives for b_i,

$$\frac{1}{n}\sum_{t=0}^{n-1} Ew_i(t) \leq \rho_i + o(1), \text{ where } \lim_{n\to\infty} o(1) = 0.$$

Hence, using (31) and (8),

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{t=0}^{n-1}\left[\tfrac{1}{2}\sum_i \lambda p_{*i}q_{ii}^{(M)} + \tfrac{1}{2}\sum_{i=1}^L \mu_i E(w_i(t))\left(q_{ii}^{(M)} - 2\sum_j p_{ij}q_{ij}^{(M)} + \sum_j p_{ij}q_{jj}^{(M)}\right)\right] \\ &\qquad \leq \textstyle\sum_{i=1}^L \mu_i\rho_i\left(q_{ii}^{(M)} - \sum_j p_{ij}q_{ij}^{(M)}\right) + o(1). \end{aligned}$$

Define $\overline{X}_j(n) := \frac{1}{n}\sum_{t=0}^{n-1} x_j(t)$. Summing (36), and taking expectations we obtain

$$\begin{aligned} &E|\overline{X}(n)| - (\lambda-\lambda^*)E\textstyle\sum_{j=1}^L \overline{X}_j(n)\sum_{i=1}^L p_{*i}q_{ij}^{(M)} - \lambda^* E\sum_{j=1}^L \overline{X}_j(n)\sum_{i=1}^L p_{*i}q_{ij}^{(M-1)} \\ &\leq \textstyle\sum_{i=1}^L \mu_i\rho_i\left(q_{ii}^{(M)} - \sum_j p_{ij}q_{ij}^{(M)}\right) + \frac{1}{2}\frac{x^T(0)Q^{(M)}x(0)}{n} - \frac{1}{2}E\frac{x^T(n)Q^{(M)}x(n)}{n} + o(1). \end{aligned} \tag{37}$$

Similarly, for $Q^{(m)}$, $0 \leq m \leq M-1$, we obtain

$$\begin{aligned} &-(\lambda-\lambda^*)E\textstyle\sum_{j=1}^L \overline{X}_j(n)\sum_{i=1}^L p_{*i}q_{ij}^{(m)} - 1(m\geq 1)\lambda^* E\sum_{j=1}^L \overline{X}_j(n)\sum_{i=1}^L p_{*i}q_{ij}^{(m-1)} \\ &\leq \textstyle\sum_{i=1}^L \mu_i\rho_i\left(q_{ii}^{(m)} - \sum_j p_{ij}q_{ij}^{(m)}\right) + \frac{1}{2}\frac{x^T(0)Q^{(m)}x(0)}{n} - \frac{1}{2}E\frac{x^T(n)Q^{(m)}x(n)}{n} + o(1). \end{aligned} \tag{38}$$

Recursively substituting for $E\sum_{j=1}^{L}\overline{X}_j(n)\sum_{i=1}^{L}p_{*i}q_{ij}^{(m)}$ from (38) into (37) gives the result.

(ii) The proof is similar. □

5 Uniformly bounding functional expansions

In this section we obtain the LPs for the uniformly bounding functional expansions.

To obtain uniformly *upper* bounding functional expansions for the mean number in the system as a function of ρ, we would like to neglect the term $x^T(n)Q^{(M-m)}x(n)$ in (32).

Definition. *A symmetric matrix Q is said to be* copositive *if $x^TQx \geq 0$ for all $x \geq 0$.*

We note that copositive matrices have been extensively studied (see Cottle, Habetler and Lemke (1970), Murty and Kabadi (1987), and Andersson, Chang and Elfving (1993)), e.g., in linear complementarity theory. They are characterized by the signs of certain determinants. However, testing for copositivity is NP-Complete. Clearly all positive semidefinite matrices (i.e., $x^TQx \geq 0$ for all x), and all non-negative matrices (i.e., $q_{ij} \geq 0$ for all i,j), are copositive.

Theorem 3. Uniformly upper bounding functional expansions. *Suppose $Q^{(m)}$, $0 \leq m \leq M$, is a family of symmetric matrices satisfying (27,31) such that*

$$\sum_{m=0}^{M}\frac{1}{(1-\rho)^m}Q^{(M-m)} \text{ is copositive, and} \tag{39}$$

$$\sum_i p_{*i}q_{ij}^{(M)} \geq 0 \text{ for all } j. \tag{40}$$

Then, for every non-idling scheduling policy, for every arrival rate λ in $[0,\lambda^*)$, one has

$$\limsup_n E\left[\frac{1}{n}\sum_{t=0}^{n-1}|x(t)|\right] \leq \frac{\rho C_M}{(1-\rho)^M}+\frac{\rho C_{M-1}}{(1-\rho)^{M-1}}+\cdots+\frac{\rho C_1}{(1-\rho)}+\rho C_0, \tag{41}$$

where

$$C_m := \sum_i \mu_i \rho_i^* \left(q_{ii}^{(M-m)} - \sum_j p_{ij} q_{ij}^{(M-m)} \right). \tag{42}$$

If the scheduling policy is stationary, then the above bound also applies to the steady state value of $E|x(n)|$.

Proof. Under (39), the term in the last summation in the RHS in (32), involving $x(n)$, can be dropped. Similarly, under (40), the second term on the LHS of (32) can be dropped. □

This result provides a sufficient condition for stability for *all* arrival rates in the capacity region $0 \le \lambda < \lambda^*$. It also provides a bound on the order of the growth to infinity in heavy traffic, i.e., *pole multiplicity*, as $\rho \nearrow 1$, $E|x(n)| = O\left(\frac{1}{(1-\rho)^M}\right)$.

Any selection of the constants $(C_1, C_2, \ldots, C_m)$ which satisfies the constraints of Theorem 3 furnishes a functional upper bound on the mean number in the system. One can exploit the freedom that exists in the choice of the constants $(C_1, C_2, \ldots, C_m)$ while meeting these constraints to choose a functional upper bound which has the lowest value at a *particular* value ρ_0 of the nominal load that may be of special interest.

For example, suppose one is interested in heavy traffic performance, i.e., $\rho_0 = 1$, or more precisely $\rho \nearrow 1$. Then one first minimizes the coefficient $C_M = \sum_i \mu_i \rho_i^* \left(q_{ii}^0 - \sum_j p_{ij} q_{ij}^0 \right)$. After minimizing C_M, then one can still exploit the residual freedom by minimizing C_{M-1}. Then one can minimize $C_{M-2}, C_{M-3}, \ldots, C_0$ recursively as above.

On the other hand, if one is particularly interested in a nominal value of the load $0 < \rho_0 < 1$, then one minimizes

$$\sum_{m=0}^{M} \frac{C_m}{(1-\rho_0)^m} = \sum_{m=0}^{M} \sum_i \frac{\mu_i \rho_i^*}{(1-\rho_0)^m} \left(q_{ii}^{(M-m)} - \sum_j p_{ij} q_{ij}^{(M-m)} \right).$$

After doing this there may still be residual freedom in choosing the coefficients. This can be exploited as one chooses. For example, one could minimize $C_M, C_{M-1}, \ldots, C_0$ as above if there is secondary interest in heavy traffic behavior.

No matter how one exploits the freedom in choosing the coefficients $\{C_i\}$ which satisfy the constraints of Theorem 3, one always obtains a functional bound that is uniformly valid for all ρ in $[0,1)$ (provided of course that there is a feasible solution). Thus one can construct several such functional bounds and take their functional minimum which will yield another functional upper bound.

The constraints of Theorem 3 can be written as linear constraints on the variables $\{q_{ij}, C_m\}$, except for the copositivity constraint (39). As noted earlier, the copositivity of a matrix is characterized by the signs of certain determinants, though testing for copositivity is NP-complete.

Instead of directly incorporating the copositivity constraints (39) in any optimization procedure for selecting the $\{C_m\}$, one has two options. For the first option, one can simply ignore the copositivity constraint (39), and optimize the value of the C_m's as desired. Then one can check whether the resulting optimal $Q^{(m)}$'s are copositive. This yields a linear programming procedure followed by copositivity test. However it may be computationally complex for large problems due to the NP-Completeness of the copositivity test. The second option is to replace the copositivity condition (39) by the stronger *non-negativity* condition

$$Q^{(m)} \geq 0, \text{ for } 0 \leq m \leq M,$$

where the matrix inequality is required to hold componentwise. Since these constraints are linear, the entire procedure can then be performed by linear programming.

The algorithms above can be implemented beginning with $M = 1$, and increasing M by one at each step until one finds the smallest value of M for which one has a feasible solution to the constraints.

Other search procedures for the smallest value of M can also be devised. One can easily note that if there is a feasible solution for a particular value of $\overline{M}$ for M, then there is a feasible solution for all larger values of $M \geq \overline{M}$. In particular, if $\{Q^{(0)}, Q^{(1)}, \ldots, Q^{(\overline{M})}\}$ is a feasible solution for $\overline{M}$, then $\{0, 0, \ldots, 0, Q^{(0)}, Q^{(1)}, \ldots, Q^{(\overline{M})}\}$ is a feasible solution for M. Thus if one can find a large value of $\overline{M}$ for which there is a feasible solution, then one can proceed to find the smallest value of M for which there is a feasible solution by a bisection search of $\{0, 1, 2, \ldots, \overline{M}\}$.

Example 1

Consider the system shown in Figure 2. The LP for the uniformly upper bounding functional

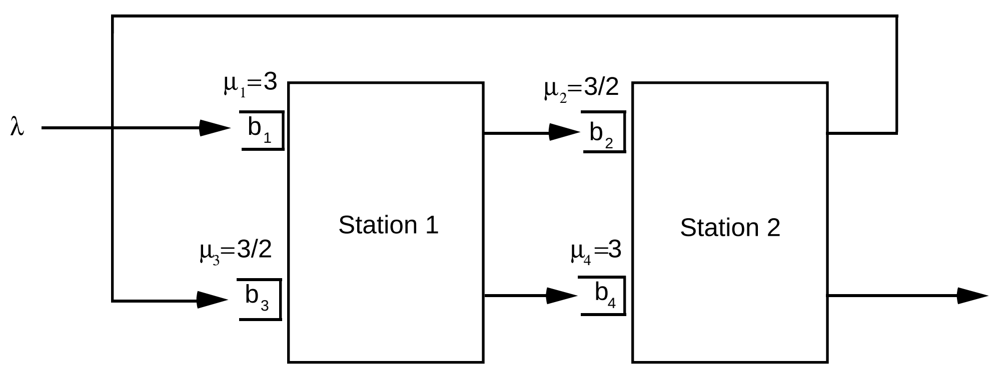

Figure 2: System of Examples 1, 2 and 3.

expansion is infeasible for $M = 1$. So we turn to $M = 2$. First we obtain the best constants for the heavy traffic upper bound. We obtain a feasible solution, with a minimal value for $C_2 = 17/9$. Recursively minimizing C_1 (after fixing C_2), and then C_0 (after fixing C_2 and C_1), the uniformly upper bounding functional expansion obtained is,

$$E|x(n)| \leq \frac{17\rho}{9(1-\rho)^2} + \frac{11\rho}{9(1-\rho)} \text{ for all } \rho \in [0, 1). \tag{43}$$

The solution is

$$Q^{(0)} = \begin{bmatrix} 1 & 2/3 & 2/3 & 0 \\ 2/3 & 2/3 & 2/9 & 2/9 \\ 2/3 & 2/9 & 2/3 & -2/9 \\ 0 & 2/9 & -2/9 & 2/9 \end{bmatrix}, Q^{(1)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 11/9 & 5/9 & 1 \\ 1 & 5/9 & 5/9 & 1/3 \\ 1 & 1 & 1/3 & 1/3 \end{bmatrix}, Q^{(2)} = 0.$$

The matrix $Q^{(0)}$ is copositive but *not* non-negative. Like all the other upper bounds in this paper, it is obtained by solving the LP without any copositivity constraint, and merely verifying that the final answer is copositive. It may be noted that there is *no* non-negative feasible solution for $Q^{(0)}$. In fact, at the nominal loading of 0.8 and larger, the original Stability LP of Kumar and Meyn (1995) does not possess a feasible non-negative solution, though it does possess a feasible copositive solution. This resolves in the negative (see Kumar and Meyn (1996)) an open problem concerning whether it is sufficient to restrict attention to non-negative solutions for Q.

In Figure 3 we compare this bound (43) with the results of the pointwise bound LPs from Kumar and Kumar (1994). Computing $\lim_{\rho \nearrow 1}$ (Upper Bound) $(1-\rho)^2$ numerically gives a value of 1.8889, which matches (43). The difference between the bounds, imperceptible in the leftmost graph of Figure 3, is about 1.11ρ in light traffic, and appears to be *exactly* $\frac{16\rho}{9(1-\rho)}$ in heavy traffic.

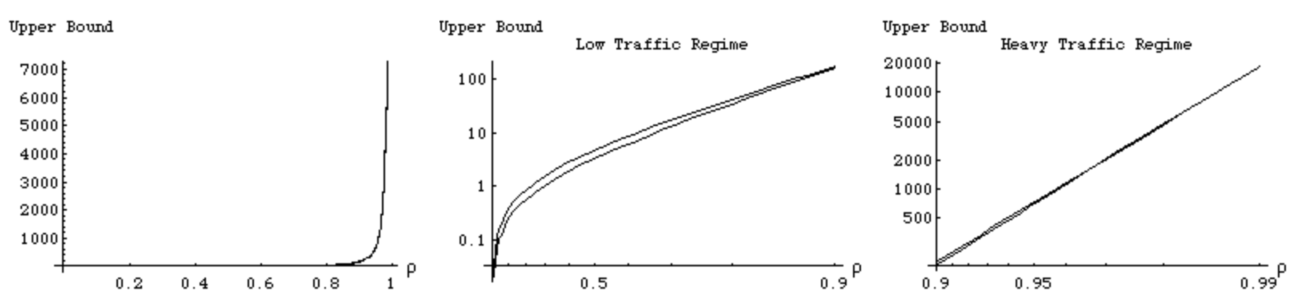

Figure 3: Comparison of functional upper bound (43) with pointwise upper bounds in Example 1. The difference between them, imperceptible in the graph on the left, is more perceptible in the other two graphs where the scales are linear in Log($|x|$) and Log($1-\rho$).

Next, as an illustration, we compute the functional bounds optimized for light traffic, which corresponds to ρ in the neighborhood of $\rho_0 = 0$. First we minimize $C_2 + C_1 + C_0$. Then we fix $C_2 + C_1 + C_0$ at its minimum value, and recursively minimize C_2, then C_1, and finally C_0, as above. This gives the bound

$$E|x(n)| \leq \frac{272\rho}{117(1-\rho)^2} + \frac{86\rho}{117(1-\rho)} \text{ for all } \rho \in [0,1), \tag{44}$$

which comes from the solution,

$$Q^{(0)} = \begin{bmatrix} 16/13 & 32/39 & 32/39 & 0 \\ 32/39 & 32/39 & 32/117 & 32/117 \\ 32/39 & 32/117 & 32/39 & -32/117 \\ 0 & 32/117 & -32/117 & 32/117 \end{bmatrix}, Q^{(1)} = \begin{bmatrix} 1 & 14/3 & 1 & 1 \\ 14/3 & 137/117 & 95/117 & 1 \\ 1 & 95/117 & 53/117 & 1/3 \\ 1 & 1 & 1/3 & 1/3 \end{bmatrix}, Q^{(2)} = 0.$$

Note that since both (43,44) are valid, one can take the minimum of them and obtain,

$$E|x(n)| \leq \text{Min}\{\frac{17\rho}{9(1-\rho)^2} + \frac{11\rho}{9(1-\rho)}, \frac{272\rho}{117(1-\rho)^2} + \frac{86\rho}{117(1-\rho)}\} \text{ for all } \rho \in [0,1). \quad \square$$

The following are the uniformly *lower* bounding functional expansion counterparts of these results. There is *no* need to require the copositivity condition (39), the positivity condition (31), or to restrict attention to non-idling policies.

Theorem 4. Uniformly lower bounding functional expansions. *Suppose $Q^{(m)}$, $0 \leq m \leq M$, is a family of symmetric matrices satisfying (29,40). Then for every scheduling policy, for every arrival rate λ in $[0, \lambda^*)$,*

$$\liminf_n E\left[\frac{1}{n}\sum_{t=0}^{n-1}|x(t)|\right] \geq \frac{\rho C_M}{(1-\rho)^M} + \frac{\rho C_{M-1}}{(1-\rho)^{M-1}} + \cdots + \frac{\rho C_1}{(1-\rho)} + \rho C_0 \text{ for } 0 \leq \rho < 1, \tag{45}$$

where

$$C_m := \sum_i \mu_i \rho_i^* \left(\sum_j p_{ij} q_{ij}^{(M-m)} - q_{ii}^{(M-m)}\right) \tag{46}$$

If the scheduling policy is stationary, this lower bound also applies to the steady state value of $E|x(n)|$.

Proof. For a stationary scheduling policy with a finite first moment it is shown in Kumar and Meyn (1996) that $\frac{1}{n}E[x^T(n+1)Qx(n+1) - x^T(n)Qx(n)] \to 0$ and so the second to last term in (33) vanishes as $n \to +\infty$. (It is for this reason that unlike Theorem 3 no copositivity condition is needed on Q). Also, for such a policy $\frac{1}{n}\sum_{t=0}^{n-1} E(w_i(t)) \to \rho_i$. (It is for this reason that unlike Theorem 3 there is no need for the positivity condition (31)). Hence the bound (45) follows from Lemma 4.ii for stationary non-idling scheduling policies with a finite first moment. If a stationary policy does not have a finite first moment, then the lower bound (45) holds trivially. Thus the bound holds for all stationary non-idling policies.

From Borkar (1983) it follows that given an initial condition for the Markovian network there exists a stationary non-idling scheduling policy that is optimal in the class of all non-anticipative non-idling scheduling policies, for the problem of minimizing the long term average of the mean number in the system. Hence the lower bound (45) applies to all non-idling policies. Finally, given any policy π that is not non-idling, one can construct a new non-idling policy $\tilde{\pi}$ under which every station works on a customer at a time that π would have worked on it, provided that under $\tilde{\pi}$ that same customer is present at the station at that time, but $\tilde{\pi}$ is non-idling since it works on an available customer in First Come First Serve order at other times. Under such a policy $\tilde{\pi}$ every customer leaves the system no later than it would have under π. Thus, for every policy, there is a non-idling policy that is at least as good. Hence the lower bound (45) applies to all policies. $\square$

We thus obtain a lower bound (using Knuth's notation) on the growth rate in heavy traffic, $E|x(n)| = \Omega\left(\frac{1}{(1-\rho)^M}\right)$, and a bound on pole multiplicity. The best constants C_m for a particular nominal load ρ_0 are found as in the case of the upper bounds. Since there is no copositivity condition to check, these are purely linear programming procedures.

Example 2

Consider the system shown in Figure 1. First we obtain the best coefficients for the heavy traffic region. The LP for the uniformly lower bounding functional expansion in (i) above yields $Q^{(0)} \equiv 0$ when $M = 2$. So we turn to $M = 1$. This gives

$$E|x(n)| \geq \frac{7\rho}{9(1-\rho)} + \frac{7\rho}{9} \text{ for all } \rho \in [0,1). \tag{47}$$

The solution is

$$Q^{(0)} = \begin{bmatrix} -1 & -2/3 & -2/3 & 0 \\ -2/3 & -4/9 & -4/9 & 0 \\ -2/3 & -4/9 & -4/9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } Q^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1/9 & 1/9 & 0 \\ 0 & 1/9 & 1/9 & 1/3 \\ 0 & 0 & 1/3 & -1/3 \end{bmatrix}.$$

The pointwise lower bounds in Kumar and Kumar (1994) numerically give $\lim_{\rho \nearrow 1}(1-\rho)$ (Lower Bound) $= 0.7778$. Thus the uniformly lower bounding functional expansion (47) directly obtains the same constant. The graphs in Figure 4 compare the uniformly lower bounding functional expansion (47) with the pointwise bounds. The difference between the two bounds is about 0.44ρ is light traffic, and about 1.78ρ in heavy traffic.

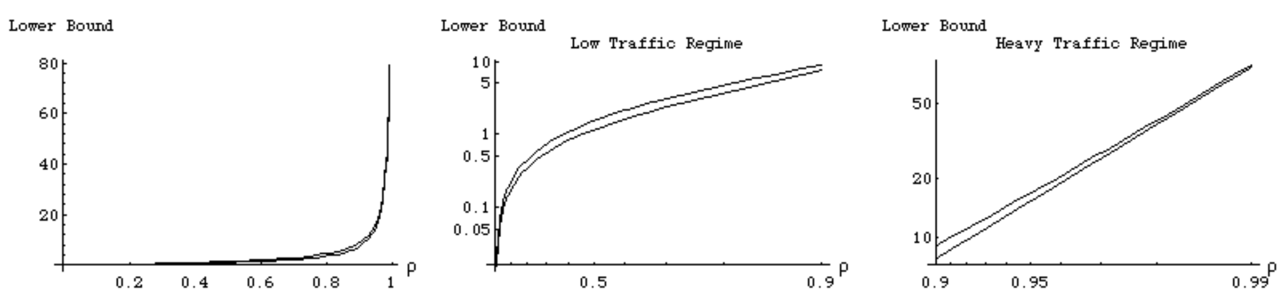

Figure 4: Comparison of functional lower bound (47) with pointwise lower bounds for system of Example 2.

Next we fit a bound optimized for the low traffic region $\rho_0 = 0$. First we maximize $C_2 + C_1 + C_0$. Then fixing $C_2 + C_1 + C_0$ at its maximum value, we recursively maximize C_2, C_1, and C_0. This yields

$$E|x(n)| \geq \frac{16\rho}{9} \text{ for all } \rho \in [0,1), \tag{48}$$

which comes from the solution,

$$Q^{(0)} = \begin{bmatrix} 0 & 1/3 & 0 & 0 \\ 1/3 & -4/9 & 0 & 0 \\ 0 & 0 & -2/3 & 0 \\ 0 & 0 & 0 & -1/3 \end{bmatrix} \text{ and } Q^{(1)} = 0.$$

Since both bounds are uniformly valid, we can take their functional maximum, and obtain

$$E|x(n)| \geq \text{Max}\{\frac{7\rho}{9(1-\rho)} + \frac{7\rho}{9}, \frac{16\rho}{9}\} \text{ for all } \rho \in [0,1). \qquad \Box$$

Consider now a buffer priority policy θ. We have seen earlier that the additional constraint (11) is satisfied in the primal. The dual LP correspondingly restricts the "$\text{Max}_{i\in\sigma(j)}$" in (27) and (29) to "$\text{Max}_{\{i:i\in\sigma(j) \text{ and } \theta(i)\leq\theta(j)\}}$. The uniformly lower and upper bounding functional expansions carry over with this change.

Theorem 5. Uniformly bounding functional expansions for buffer priority policies. *Consider a buffer priority policy θ. The bounds of Theorems 3 and 4 hold with the modification that "$\text{Max}_{i\in\sigma(j)}$" in (27) and (29) is replaced by "$\text{Max}_{\{i:i\in\sigma(j) \text{ and } \theta(i)\leq\theta(j)\}}$."*

Example 3

Consider the system shown in Figure 1. There are four buffer priority policies $\theta_{LBFS} = (4,3,2,1)$, $\theta_{FBFS} = (1,2,3,4)$, θ_{3214} and θ_{1432}. The uniformly lower and upper bounding

	Functional Lower Bound	Functional Upper Bound
θ_{LBFS}	$\frac{11\rho}{9(1-\rho)} + \frac{7\rho}{9}$	$\frac{17\rho}{6(1-\rho)} + \frac{2\rho}{3}$
θ_{FBFS}	$\text{Max}\{\frac{7\rho}{3(1-\rho)} - 7\rho, \frac{16\rho}{9}\}$	$\frac{11\rho}{2(1-\rho)}$
θ_{3214}	$\frac{7\rho}{9(1-\rho)} + \frac{7\rho}{9}$	$\frac{17\rho}{9(1-\rho)^2} + \frac{11\rho}{9(1-\rho)}$
θ_{1432}	$\frac{29\rho}{24(1-\rho)} + \frac{25\rho}{48}$	$\frac{17\rho}{6(1-\rho)} + \frac{2\rho}{3}$

Figure 5: Uniformly lower and upper bounding functional expansions for all buffer priority policies in system of Example 3.

functional expansions, both optimized for the heavy traffic region, are shown in Figure 5. For θ_{FBFS} we take the maximum of the heavy and light traffic optimized bounds. In all cases, the coefficient C_M of the highest power is the same as that obtained numerically in the limit from the pointwise bounds. It is worth noting that for θ_{LBFS} the functional lower bound appears to coincide *exactly* with the pointwise lower bounds for all values of ρ. $\square$

In all the Examples 1, 2 and 3, the heavy traffic growth constant C_M produced by the functional bound LPs has coincided with the value computed numerically from the pointwise bound LPs. Hence the closed form solution provided by the functional bound LPs allows us to comprehend the quality of the results provided by the original pointwise bound LPs.

6 Reduced dimensional LP for the functional lower bound for balanced systems

We now explicitly describe a family of feasible solutions of the LP for the uniformly lower bounding functional expansions for balanced systems where $\rho_\sigma \equiv \rho$ for all σ. These explicit solutions provide functional lower bounds of the form $E|x(n)| \geq \frac{C_1 \rho}{1-\rho}$. They are described by parameters that can be optimized through considerably lower dimensional LPs. While the original LP for $M = 1$ features $L(L+1)$ variables $\left(\frac{L(L+1)}{2}\right.$ variables for each of $Q^{(0)}$ and

$Q^{(1)}\Big)$, our explicit solution only features $\frac{S(S+1)}{2}$ variables, where S is the number of stations, which is in many applications of interest far smaller than the number L of buffers.

The feasible solutions are expressed in terms of $W_{\sigma i}$, the *mean work remaining* to be done on a customer in buffer b_i by station σ, prior to that customer's exit from the system. They are obtained as the unique solution of the equations,

$$W_{\sigma i} = \frac{1}{\mu_i} 1(i \in \sigma) + \sum_j p_{ij} W_{\sigma j} \text{ for all } i. \tag{49}$$

The key idea is to restrict attention to quadratics in the mean remaining work $w_\sigma := \sum_i W_{\sigma i} x_i$ for stations, rather than more general quadratics in the system state consisting the number customers in each buffer. Specifically, instead of considering quadratics such as $\sum_{ij} q_{ij} x_i x_j$, one restricts attention to quadratics such as $-\sum_{\sigma\sigma} a_{\sigma\sigma'} w_\sigma w_{\sigma'}$. Thus we look for feasible solutions of the form $Q^{(m)} = -W^T A^{(m)} W$ where $W := [W_{\sigma i}]$. In what follows we detail a construction for $Q^{(0)}$ which with $M = 1$ and $Q^{(1)} = 0$ yields a feasible solution. Our construction works only for balanced systems with $\rho_\sigma \equiv \rho$, i.e., those systems for which all stations are equally loaded. The result is a lower bound with a pole multiplicity of order 1. In our numerical studies we have been unable to find any queueing network for which a lower bound with pole multiplicity greater than or equal to 2 can be established.

The following lemma provides the details of our construction.

Lemma 5. *Consider* a balanced *system, i.e.,* $\rho_\sigma \equiv \rho$ *for all* σ. *Let* $A = A^T = [a_{\sigma\sigma'}]$ *be an* $S \times S$ *matrix which satisfies,*

$$\sum_{\sigma\sigma'} a_{\sigma'\sigma} W_{\sigma j} \leq 1 \text{ for all } j, \tag{50}$$

$$\sum_{\sigma} a_{\sigma'\sigma} W_{\sigma j} \geq 0 \text{ for all } \sigma' \text{ and } j, \tag{51}$$

$$a_{\sigma\sigma'} = a_{\sigma'\sigma} \text{ for all } \sigma, \sigma'. \tag{52}$$

(i) *The constraints (29,40) are satisfied with* $M := 1$*,*

$$\begin{aligned} Q^{(0)} &:= -W^T AW, \text{ where } W := [W_{\sigma i}], \text{ and} \\ Q^{(1)} &:= 0. \end{aligned}$$

(ii) $\sum_i \mu_i \rho_i^* \left(\sum_j p_{ij} q_{ij}^{(0)} - q_{ii}^{(0)} \right) = \sum_i \rho_i^* \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma i}$.

Proof. Since $M = 1$ and $Q^{(1)} = 0$, it suffices for (29) to verify that the following inequalities hold:

$$-\lambda^* \sum_i p_{*i} q_{ij}^{(0)} \leq 1 \text{ for all } j, \text{ and} \tag{53}$$

$$\begin{aligned} &\lambda^* \sum_i p_{*i} q_{ij}^{(0)} + \mathrm{Max}_{i \in \sigma(j)} \mu_i \left(\sum_k p_{ik} q_{kj}^{(0)} - q_{ij}^{(0)} \right) \\ &+ \sum_{\sigma \neq \sigma(j)} \mathrm{Max}_{i \in \sigma} \mu_i \left(\sum_k p_{ik} q_{kj}^{(0)} - q_{ij}^{(0)} \right)^+ \leq 0. \end{aligned} \tag{54}$$

For (53), note that since the system is balanced, the capacity λ^* satisfies

$$\lambda^* \sum_i p_{*i} W_{\sigma i} = 1 \text{ for all } \sigma. \tag{55}$$

Now $q_{ij}^{(0)} = -\sum_{\sigma\sigma'} a_{\sigma'\sigma} W_{\sigma' i} W_{\sigma j}$. Hence (53) is verified since,

$$\begin{aligned} -\lambda^* \sum_i p_{*i} q_{ij}^{(0)} &= \lambda^* \sum_i p_{*i} \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma' i} W_{\sigma j} \\ &= \sum_{\sigma'} \left(\sum_\sigma a_{\sigma\sigma'} W_{\sigma j} \right) \sum_i (\lambda^* p_{*i} W_{\sigma' i}) \\ &= \sum_{\sigma'} \sum_\sigma a_{\sigma\sigma'} W_{\sigma j} \text{ (due to (55))} \\ &\leq 1. \end{aligned} \tag{56}$$

Now we turn to showing (54). First note that

$$\sum_k p_{ik} q_{kj}^{(0)} - q_{ij}^{(0)} = -\sum_k p_{ik} \sum_{\sigma\sigma'} a_{\sigma'\sigma} W_{\sigma' k} W_{\sigma j} + \sum_{\sigma\sigma'} a_{\sigma'\sigma} W_{\sigma' i} W_{\sigma j}$$

$$
\begin{aligned}
&= -\sum_{\sigma\sigma'} \left[W_{\sigma' i} - \frac{1}{\mu_i} 1(i \in \sigma') \right] a_{\sigma'\sigma} W_{\sigma j} + \sum_{\sigma\sigma'} a_{\sigma'\sigma} W_{\sigma' i} W_{\sigma j} \text{ (from (49))} \\
&= \frac{1}{\mu_i} \sum_{\sigma} a_{\sigma(i)\sigma} W_{\sigma j} \qquad (57) \\
&\geq 0 \text{ (from (51))},
\end{aligned}
$$

which also establishes (ii) as a by product of (57).

Hence

$$
\begin{aligned}
\text{LHS of (54)} &= \lambda^* \sum_i p_{*i} q_{ij}^{(0)} + \sum_{\sigma\sigma'} a_{\sigma\sigma'} W_{\sigma j} \\
&= 0 \text{ (from (56)},
\end{aligned}
$$

which proves (54), and thus (i), since (40) trivially holds.

The equality in (ii) follows from (49). □

Once we have discovered constraints on $A^{(0)}$ which guarantee that the resulting $\{Q^{(0)}, Q^{(1)} = 0\}$ are feasible, we can optimize the lower bound by linear programming. Thus we obtain the following reduced dimensional LP for the functional lower bound.

Theorem 6. Reduced dimensional LP for functional lower bound for balanced systems. *Consider a balanced system. Let C_1 be the value of the LP,*

$$
\text{Max} \sum_i \rho_i^* \sum_{\sigma} a_{\sigma(i)\sigma} W_{\sigma i}
$$

subject to (50, 51, 52). Then, for any scheduling policy,

$$
\liminf_N \frac{1}{N} \sum_{n=0}^{N-1} E|x(n)| \geq \frac{\rho C_1}{1-\rho}. \qquad (58)
$$

7 Asymptotic heavy traffic domination of the Pollaczek-Khintchine bound

We now show that the functional lower bound (58) produced by the reduced dimensional LP of Theorem 6 is always at least as good asymptotically in heavy traffic as a bound for networks obtainable from the Pollaczek-Khintchine formula which captures the behavior of any one bottleneck. More precisely, we show that as $\rho \nearrow 1$ the limiting ratio of (58) to the lower bound obtained from the Pollaczek-Khintchine formula is guaranteed to be at least one. We also provide an example where the limiting ratio is strictly larger than one. This example shows that our functional lower bound (58) (and thus also (45)) captures nontrivial interactions between multiple bottleneck stations in heavy traffic.

Recall that the Pollaczek-Khintchine formula for the mean number of customers in an $M/G/1$ queue operated under the First Come First Serve (FCFS) policy is $\frac{2\rho-\rho^2+\lambda^2 \text{var(service time)}}{2(1-\rho)}$. We can apply this formula to calculate a lower bound on the mean number in the system shown in Figure 6 consisting of a single station revisited several times.

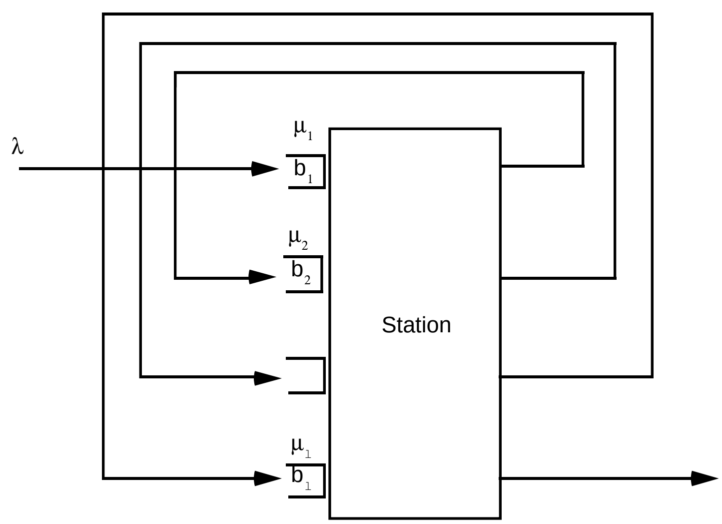

Figure 6: A single station with multiple revisits.

Under any non-idling scheduling policy the *work* in this system is conserved. It is also easy to see that the mean *number of customers* in the system is a minimum when the Last Buffer First Serve buffer priority policy is used, which gives priority to the buffers $\{b_\ell, b_{\ell-1}, \ldots, b_1\}$ in that order. The resulting queueing system is equivalent, in terms of the number of customers in the system at any given time, to an $M/G/1$ queue where each customer's service time is the sum of ℓ independent exponentially distributed random variables with means $\frac{1}{\mu_1}, \frac{1}{\mu_2}, \ldots, \frac{1}{\mu_\ell}$, when it is operated under the FCFS policy. Such a combined service time has mean $\sum_{i=1}^{\ell} \frac{1}{\mu_i}$ and variance $\sum_{i=1}^{\ell} \frac{1}{\mu_i^2}$, with $\rho = \lambda \sum_{i=1}^{\ell} \frac{1}{\mu_i}$. Thus, we obtain that the mean number of customers under any non-idling policy (and in fact under any scheduling policy) for the single station multiple revisit system of Figure 6 is lower bounded by $\frac{2\rho - \rho^2 + \lambda^2 \sum_{i=1}^{\ell} \frac{1}{\mu_i^2}}{2(1-\rho)}$.

Using this, we can derive a lower bound for any open *re-entrant line*, i.e., a network as in Section 2 with every p_{ij} either 0 or 1 (as in Figure 2). Let σ be a fixed station in such a network. Now if all customers spend zero time at all buffers in the network, *except* the buffers of station σ, then one obtains a network of the form shown in Figure 6, which has the lower bound $\frac{2\rho - \rho^2 + \lambda^2 \sum_{i \in \sigma} \frac{1}{\mu_i^2}}{2(1-\rho)}$. It is easy to show that this lower bound is also a lower bound on the number of customers in the original network where the service times at all other stations are not zero.

For every station σ, one therefore obtains such a lower bound, from which it follows by maximizing over σ that for the open re-entrant line, under any scheduling policy in steady-state,

$$E \mid x(t) \mid \ \geq \ \operatorname*{Max}_{\sigma} \frac{2\rho - \rho^2 + \lambda^2 \sum_{i \in \sigma} \frac{1}{\mu_i^2}}{2(1-\rho)}. \tag{59}$$

We call this the *Pollaczek-Khintchine lower bound.* In a sentence, it captures the worst bottleneck in the system, when all stations except one, are made transparent to customers.

We will show below in Theorem 7 that if we consider the functional lower bound of the previous section with the further restriction that A is a *non-negative matrix*, then the value of the reduced dimensional LP is

$$C_1 \quad = \quad \frac{1 + \text{Max}_\sigma \lambda^{*2} \sum_{i \in \sigma} \frac{1}{\mu_i^2}}{2}. \tag{60}$$

Once we do this in the following Theorem 7, we will then have shown that the lower bound (45) produced with this additional restriction to a non-negative A is at least as good as,

$$\frac{\rho \left(1 + \text{Max}_\sigma \lambda^{*2} \sum_{i \in \sigma} \frac{1}{\mu_i^2}\right)}{2(1-\rho)}. \tag{61}$$

Taking the heavy traffic limit of the ratio of (61) and (59), and noting that $\lambda \nearrow \lambda^*$ as $\rho \nearrow 1$, we obtain

$$\lim_{\rho \nearrow 1} \frac{\rho \left(1 + \text{Max}_\sigma \lambda^{*2} \sum_{i \in \sigma} \frac{1}{\mu_i^2}\right)}{2(1-\rho)} \Bigg/ \frac{\text{Max}_\sigma \left(2\rho - \rho^2 + \lambda^2 \sum_{i \in \sigma} \frac{1}{\mu_i^2}\right)}{2(1-\rho)} = 1.$$

Since our bound (45) does *not* restrict A to be non-negative, it is at least as high as that obtained with the additional non-negativity restriction, and we will thus have proved asymptotic dominance in heavy traffic. In fact, we will later present a example where the asymptotic dominance is strict.

It only remains to establish that (60) is indeed the value of the reduced LP when the additional constraint that A be a non-negative matrix is appended. This is done in the following theorem.

Theorem 7. Asymptotic heavy traffic domination of the Pollaczek-Khintchine Bound. *Consider a balanced re-entrant line. The value of the LP,*

$$\text{Max} \sum_i \rho_i^* \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma i}, \tag{62}$$

subject to (50,51,52) and

$$a_{\sigma\sigma'} \quad \geq \quad 0 \text{ for all } \sigma, \sigma', \tag{63}$$

is (60).

The proof of this result follows from the following two Lemmas.

Lemma 6. *The value of the LP in Theorem 7 is*

$$\frac{\lambda^{*2} Max_{\sigma,\sigma'} \left[\sum_{i\in\sigma'} \frac{1}{\mu_i} W_{\sigma i} + \sum_{i\in\sigma} \frac{1}{\mu_i} W_{\sigma' i}\right]}{2}.$$

Proof: Define $c_{\sigma'\sigma} := \sum_{i\in\sigma'} \rho_i^* W_{\sigma i}$. Using the symmetry of $a_{\sigma\sigma'}$ we can write the objective (62) as,

$$\begin{aligned}
\sum_i \rho_i^* \sum_\sigma a_{\sigma(i)\sigma} W_{\sigma i} &= \sum_{\sigma\sigma'} a_{\sigma'\sigma} \sum_{i\in\sigma'} \rho_i^* W_{\sigma i} \\
&= \sum_{\sigma\sigma'} a_{\sigma'\sigma} c_{\sigma'\sigma} \\
&= \sum_{\sigma\sigma'} a_{\sigma'\sigma} \frac{(c_{\sigma\sigma'} + c_{\sigma'\sigma})}{2}. \qquad (64)
\end{aligned}$$

Now note that since the system is a re-entrant line and is also balanced, from (55) we have $W_{\sigma 1} = \frac{1}{\lambda^*}$ for all σ. Hence the constraint (51) for $j = 1$ is,

$$\sum_{\sigma\sigma'} a_{\sigma'\sigma} \le \lambda^*.$$

Fix a choice of $\overline{\sigma}$ and $\overline{\sigma}'$ such that $c_{\overline{\sigma}\,\overline{\sigma}'} + c_{\overline{\sigma}'\overline{\sigma}} \ge c_{\sigma\sigma'} + c_{\sigma'\sigma}$ for all σ, σ'. Since $a_{\sigma'\sigma} \ge 0$, it follows that an optimal "allocation" of the $a_{\sigma'\sigma}$'s in the LP is to set $a_{\overline{\sigma}\,\overline{\sigma}'} = a_{\overline{\sigma}'\overline{\sigma}} = \frac{\lambda^*}{2}$ if $\overline{\sigma} \ne \overline{\sigma}'$, or $a_{\overline{\sigma}\,\overline{\sigma}} = \lambda^*$ if $\overline{\sigma} = \overline{\sigma}'$, and set all other $a_{\sigma\sigma'}$'s to zero. Since this is a feasible allocation, the result follows from (64) upon noting that $\rho_i^* = \frac{\lambda^*}{\mu_i}$. □

Lemma 7.

$$\sum_{i\in\sigma'} \frac{1}{\mu_i} W_{\sigma i} + \sum_{i\in\sigma} \frac{1}{\mu_i} W_{\sigma' i} = \frac{1}{\lambda^{*2}} + 1(\sigma = \sigma') \sum_{i\in\sigma} \frac{1}{\mu_i^2}. \qquad (65)$$

Proof. Let $a_i := 1(i \in \sigma)$ and $b_i := 1(i \in \sigma')$. Then

$$\sum_{i\in\sigma'} \frac{1}{\mu_i} W_{\sigma i} + \sum_{i\in\sigma} \frac{1}{\mu_i} W_{\sigma' i} = \sum_i \frac{1}{\mu_i} b_i \sum_{j\ge i} \frac{1}{\mu_j} a_j + \sum_i \frac{1}{\mu_i} a_i \sum_{j\ge i} \frac{1}{\mu_j} b_j$$

$$
\begin{aligned}
&= \sum_i \sum_{j \geq i} \frac{1}{\mu_i \mu_j} b_i a_j + \sum_i \sum_{j \geq i} \frac{1}{\mu_i \mu_j} a_i b_j \\
&= \sum_i \sum_j \frac{1}{\mu_i \mu_j} b_i a_j + \sum_i \frac{1}{\mu_i^2} a_i b_i \\
&= \left(\sum_i \frac{1}{\mu_i} b_i \right) \left(\sum_j \frac{1}{\mu_j} a_j \right) + \sum_{i \in \sigma} \frac{1}{\mu_i^2} 1(i \in \sigma) 1(i \in \sigma') \\
&= \frac{1}{\lambda^{*2}} + 1(\sigma = \sigma') \sum_{i \in \sigma} \frac{1}{\mu_i^2}. \qquad \square
\end{aligned}
$$

Now we provide an example to show that when the sign of $a_{\sigma\sigma'}$ is unrestricted as in (45), it can give a strictly better lower bound than (60). Hence the asymptotic dominance in heavy traffic of our lower bound (45) over the Pollaczek-Khintchine bound (59) is strict. Thus we see that the lower bound can capture complicated interactions between multiple bottleneck stations.

Example 4

Consider the system of Figure 7.

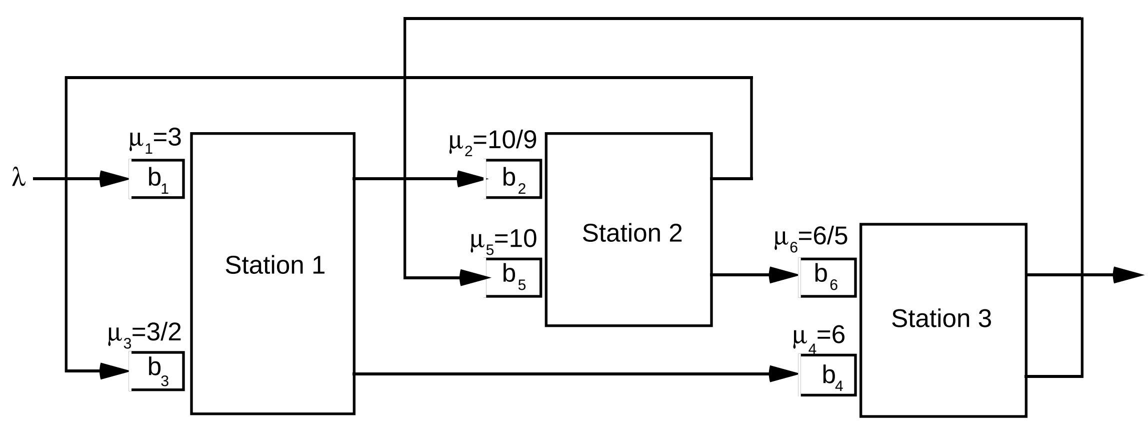

Figure 7: System of Example 4.

The optimal non-negative A gives $C_1 = 91/100$, with

$$A_{\text{Non-negative}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

It clearly captures only the behavior of the single bottleneck Station 2.

On the other hand, the optimal sign indefinite solution for A is,

$$A_{\text{Sign indefinite}} = \begin{bmatrix} 99/467 & -75/467 & 9/467 \\ -75/467 & 500/467 & 0 \\ 9/467 & 0 & 0 \end{bmatrix}.$$

It clearly captures the interactions between all the three bottleneck stations, and it gives the strictly larger value of the constant $C_1 = 466/467$. □

8 Explicit pointwise upper bounds

The only topological configurations which have been proved to be stable for all non-idling policies throughout the capacity region $[0, \lambda^*)$ are acyclic networks. The only buffer priority policies which have been proved stable throughout the capacity region $[0, \lambda^*)$ for all re-entrant lines are the First Buffer First Serve Policy (FBFS) $\theta_{FBFS} = \{1, 2, \ldots, L\}$, and the Last Buffer First Serve Policy (LBFS) $\theta_{LBFS} = \{L, L-1, \ldots, L\}$, see Lu and Kumar (1991), Dai (1995), Kumar and Kumar (1995), and Dai and Weiss (1994). It is only for such systems which are provably stable throughout $[0, \lambda^*)$ that one can hope to establish functional upper bounds.

In the following Sections 9 and 10, we are able to provide upper bounds for all acyclic topologies for all non-idling policies, and the FBFS policy for all re-entrant lines, but first we demonstrate by a counterexample that the method of using quadratic functions cannot give the proof of stability for LBFS throughout the capacity region.

Example 5: Unbounded pointwise upper bound LP for LBFS

Consider the system shown in Figure 8, operating under the LBFS policy.

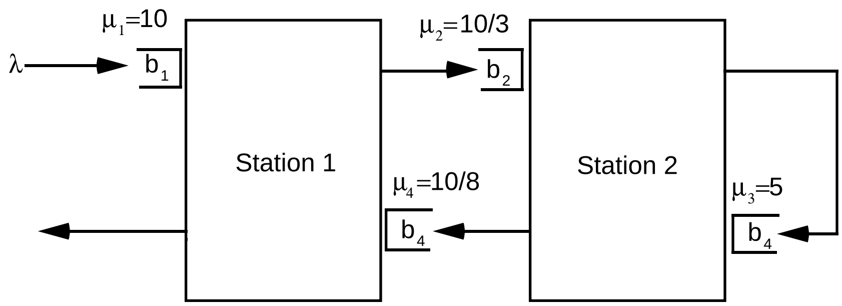

Figure 8: System of Example 5 for which pointwise upper bound LP for the LBFS policy is unbounded.

The pointwise upper bound LP is,

$$\text{Min } \lambda \left(\sum_{i=1}^{L} q_{ii} - \sum_{i=1}^{L-1} q_{i,i+1} \right)$$

subject to:

$$\lambda q_{1j} + \text{Max}_{\{i:i\in\sigma(j) \text{ and } i\geq j\}} \mu_i (q_{i+1,j} - q_{ij}) + \text{Max}_{i\notin\sigma(j)} \mu_i (q_{i+1,j} - q_{ij})^+ \leq -1.$$

It is the dual of the pointwise upper bound LP in Theorem 2. If the mean number in the system is finite in steady state, then the value of this LP is an upper bound on it. As shown in Figure 9, the value is unbounded for $0.8892 < \lambda < \frac{10}{9} = \lambda^*$. As a consequence, the LP for the uniformly upper bounding functional expansion is infeasible. □

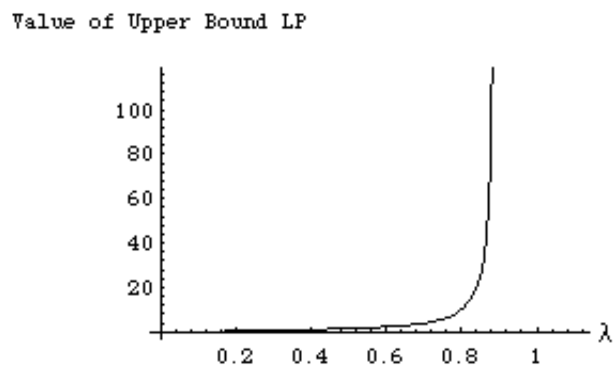

Figure 9: Plot of value of pointwise upper bound LP for LBFS policy in Example 5. It is unbounded for $\lambda > 0.8892$.

9 Explicit upper bound for the FBFS policy

In this section we determine an upper bound on the pole multiplicity for re-entrant lines operating under the FBFS policy. We will do so by directly determining an explicit feasible solution for the original pointwise upper bound LP itself. Our method will provide a functional form for the Q matrix for the pointwise upper bound LP, and we establish the pole multiplicity upper bound by studying this. With this result, the FBFS policy becomes the only one known, to the best of our knowledge, for which a pole multiplicity upper bound is available for all re-entrant lines.

If the mean number in the system is finite, a pointwise upper bound on the FBFS policy, for each arrival rate λ, is given by the value of the pointwise upper bound LP of Theorem 2. Its dual is,

$$\text{Min } \lambda \left(\sum_{i=1}^{L} q_{ii} - \sum_{i=1}^{L-1} q_{i,i+1} \right)$$

subject to,

$$\lambda q_{1j} + \text{Max}_{\{i:i\in\sigma(j) \text{ and } i\leq j\}} \mu_i (q_{i+1,j} - q_{ij}) + \sum_{\sigma\neq\sigma(j)} \text{Max}_{i\in\sigma} \mu_i (q_{i+1,j} - q_{ij})^+ \leq -1. \qquad (66)$$

Let us append the following non-negativity condition and call it the *Stability LP*,

$$q_{ij} = q_{ji} \geq 0. \tag{67}$$

Then, it follows that if there is a feasible solution, it is automatically copositive, and so the mean number in the system is indeed finite, and hence the value of the Stability LP is a pointwise upper bound, see Kumar and Meyn (1996).

Theorem 8. Upper bound for FBFS. *Consider a re-entrant line operating under the FBFS policy.*

(i) *Let*

$$b_\sigma := \mathrm{Max}_{i\in\sigma} W_{\sigma,i} \quad (= W_{\sigma,1})\,. \tag{68}$$

For a fixed $\epsilon_1 > 0$, recursively define,

$$\epsilon_{i+1} = \epsilon_i \ \mathrm{Min}\left\{ Min_{1\leq j\leq i-1} \frac{W_{\sigma(i)j}}{W_{\sigma(i+1)j}}, \frac{\mu_i W_{\sigma(i)i} - \lambda b_{\sigma(i)}}{2\mu_i W_{\sigma(i+1),i}} \right\} \ \textit{for } 1 \leq i \leq L-1.$$

(For $i = 1$, ϵ_2 is taken equal to the second term). Now scale ϵ_1, thus rescaling all the other ϵ_js, so that

$$\epsilon_j \ \mathrm{Max}\left\{ \lambda b_{\sigma(j)} - 1, \frac{\left(\lambda b_{\sigma(j)} - \mu_j W_{\sigma(j)j}\right)}{2} \right\} \leq -1 \ \textit{for all } j. \tag{69}$$

Then using the notation $i \vee j := \max(i,j)$ and $i \wedge j := \min(i,j)$,

$$q_{ij} := \epsilon_{i\vee j} W_{\sigma(i\vee j), i\wedge j}$$

satisfies the constraints (66,67) of the Stability LP.

(ii) *Hence $E(x_n) \leq \lambda(\sum_{i=1}^{L} q_{ii} - \sum_{i=1}^{L-1} q_{i,i+1})$ in steady state.*

Proof. Note first that since $\epsilon_1 > 0$, it follows that $\epsilon_i > 0$ for all i, since $\mu_i W_{\sigma(i)i} - \lambda b_{\sigma(i)} \geq 1 - \lambda b_{\sigma(i)} > 0$. Moreover each $\epsilon_i < +\infty$ since $W_{\sigma(i+1)1} > 0$ for $1 \leq i \leq L-1$.

We will now bound $\mu_i(q_{i+1,j} - q_{ij})$ by considering several cases.

Case i: $i < j$

$$\mu_i(q_{i+1,j} - q_{ij}) = \mu_i(\epsilon_j W_{\sigma(j),i+1} - \epsilon_j W_{\sigma(j)i}) = -1(i \in \sigma(j))\epsilon_j \leq 0.$$

Case ii: $i > j$

$$\mu_i(q_{i+1,j} - q_{ij}) = \mu_i(\epsilon_{i+1} W_{\sigma(i+1)j} - \epsilon_i W_{\sigma(i)j}) \leq 0.$$

Case iii: $i = j$

$$\begin{aligned} \mu_j(q_{j+1,j} - q_{jj}) &= \mu_j(\epsilon_{j+1} W_{\sigma(j+1)j} - \epsilon_j W_{\sigma(j)j}) \\ &\leq \mu_j \left(\epsilon_j \frac{\mu_j W_{\sigma(j)j} - \lambda b_{\sigma(j)}}{2\mu_j W_{\sigma(j+1)j}} W_{\sigma(j+1)j} - \epsilon_j W_{\sigma(j)j} \right) \\ &= -\epsilon_j \frac{\mu_j W_{\sigma(j)} + \lambda b_{\sigma(j)}}{2} \\ &< 0. \end{aligned}$$

Hence

$$\begin{aligned} \text{LHS of (66)} &= \lambda \epsilon_j W_{\sigma(j)1} + \text{Max}_{\{i: i \in \sigma(j) \text{ and } i \leq j\}} \left\{ -\epsilon_j, \frac{-\epsilon_j \left(\mu_j W_{\sigma(j)j} + \lambda b_{\sigma(j)} \right)}{2} \right\} \\ &= \epsilon_j \, \text{Max} \left\{ (\lambda b_{\sigma(j)} - 1), \frac{(\lambda b_{\sigma(j)} - \mu_j W_{\sigma(j)j})}{2} \right\} \qquad (70) \\ &< 0. \end{aligned}$$

After scaling as in (69), ϵ_1 is chosen to make

$$\text{LHS of (66)} \leq -1 \text{ for } 1 \leq j \leq L,$$

thus satisfying (66). □

As a corollary, from Kumar and Meyn (1995) it follows that FBFS is stable in the following very strong sense.

Corollary. Exponential stability of FBFS. *Under FBFS, for any $0 \leq \lambda < \lambda^*$, the system is $e^{\epsilon x}$-uniformly ergodic for some small $\epsilon > 0$. In particular it has moments of all polynomial orders and even an exponential moment, and they all converge geometrically fast to their steady state values.*

Next we obtain the upper bound on the pole multiplicity of the growth rate of FBFS. We study the behavior when λ is close to λ^* by employing a slightly different construction which separates the bottleneck stations and their last buffers, from the others.

Theorem 9. Upper bound on pole multiplicity of FBFS. *Under the FBFS policy,*

$$E|x_n| = O\left(\frac{1}{(1-\rho)^B}\right), \tag{71}$$

where B is the number of bottleneck stations.

Proof. The proof is again by constructing an explicit solution for Q. We separate the last buffers at the bottleneck stations from the others. With $\mathcal{B} := \{\sigma : \lambda^* W_{\sigma 1} = 1\}$ = set of bottleneck stations, let

$$\begin{aligned} \mathcal{L} &:= \{i : i = \max\{j : j \in \sigma\} \text{ for some } \sigma \in \mathcal{B}\} \\ &= \text{ set of last buffers at the bottleneck stations,} \end{aligned}$$

let

$$\begin{aligned} \beta_i &:= \text{Min}_{1 \leq j \leq i-1} \frac{W_{\sigma(i)j}}{W_{\sigma(i+1)j}} \text{ for } 2 \leq i \leq L-1, \\ \delta_i^* &:= \frac{\mu_i W_{\sigma(i)i} - \lambda^* b_{\sigma(i)}}{2\mu_i W_{\sigma(i+1)i}} \text{ for } i \notin \mathcal{L}, 1 \leq i \leq L-1, \\ \delta_i(\lambda) &:= \frac{\mu_i W_{\sigma(i)i} - \lambda b_{\sigma(i)}}{2\mu_i W_{\sigma(i+1)i}} = \frac{1-\rho_{\sigma(i)}}{2\mu_i W_{\sigma(i+1)i}} \text{ for } i \in \mathcal{L}, 1 \leq i \leq L-1. \end{aligned} \tag{72}$$

Let

$$\gamma_i(\lambda) := \begin{cases} \gamma_i^* := \text{Min}\{\beta_i, \delta_i^*\} \text{ for } i \notin \mathcal{L}, 1 \leq i \leq L-1, \\ \text{Min}\{\beta_i, \delta_i(\lambda)\} \text{ for } i \in \mathcal{L}1, 1 \leq i \leq L-1. \end{cases}$$

With $\epsilon_1 := 1$, we recursively set

$$\epsilon_{i+1} := \epsilon_i \gamma_i(\lambda) \text{ for } i = 1, 2, \ldots, L-1.$$

We note that since $\mu_i W_{\sigma(i)i} = 1 = \lambda^* b_{\sigma(i)}$ for $i \in \mathcal{L}$, $\delta_i(\lambda^*) = 0$, and so

$$\gamma_i(\lambda) = \delta_i(\lambda) \text{ for all } i \in \mathcal{L}, \text{ for } \lambda \text{ in some open interval } (\lambda^* - \epsilon, \lambda^*). \tag{73}$$

Consequently,

$$\epsilon_{i+1} = \prod_{\substack{1 \le k \le i \\ k \notin \mathcal{L}}} \gamma_k^* \prod_{\substack{1 \le k \le i \\ k \in \mathcal{L}}} \gamma_k(\lambda) \text{ for } 1 \le i \le L-1, \text{ and } \lambda \text{ in } (\lambda^* - \epsilon, \lambda^*). \tag{74}$$

Define $q_{ij} := \epsilon_{i \vee j} W_{\sigma(i \vee j) i \wedge j}$.

We will henceforth restrict attention to λ in $(\lambda^* - \epsilon, \lambda^*)$. Exactly as in Theorem 8, we can verify that the bounds of Cases i, ii, and iii hold, and that (70) holds.

Now we note that with $\epsilon_1 := 1$, the objective function satisfies

$$\begin{aligned}
\lambda \left(\sum_{i=1}^{L} q_{ii} - \sum_{i=1}^{L-1} q_{i,i+1} \right) &= \lambda \left(\sum_{i=1}^{L} \epsilon_i W_{\sigma(i)i} - \sum_{i=1}^{L-1} \epsilon_{i+1} W_{\sigma(i+1)i} \right) \\
&\le \lambda \sum_{i=1}^{L} \epsilon_i W_{\sigma(i)i} \\
&\le c, \text{ where } c \text{ does not depend on } \lambda.
\end{aligned} \tag{75}$$

Now we investigate (70) since we want to scale ϵ_1, which will result in the scaling of (75), so that the LHS of (66) is bounded above by (-1).

For $j \in \sigma \notin \mathcal{B}$, $\rho_{\sigma(j)} - \mu_j W_{\sigma(j)j} = \lambda W_{\sigma(j)1} - \mu_j W_{\sigma(j)j} \le \lambda^* W_{\sigma(j)1} - 1 \le c' < 0$ for some c'. Hence for $j \in \sigma \notin \mathcal{B}$, LHS of (66) $\le -\frac{1}{2}\epsilon_j c'$. For $j \in \sigma \in \mathcal{B}$, $\rho_{\sigma(j)} - \mu_j W_{\sigma(j)j} \le \rho_{\sigma(j)} - 1 = \rho - 1$. Hence for $j \in \sigma \in \mathcal{B}$, LHS of (66) $\le -\frac{\epsilon_j}{2}(1-\rho)$. Therefore, dividing ϵ_1 by $c''(1-\rho)\text{Min}_j \epsilon_j$ yields (66). Noting from (72, 73, 74) that $\epsilon_{i+1} \ge c'''(1-\rho)^{B-1}$, the bound (71) follows. □

10 Upper bound for acyclic systems

In this section we obtain an explicit upper bound for all *acyclic* networks, i.e., systems for which

$$\sigma(i+1) \neq \sigma(i) \text{ implies that } \sigma(j) \neq \sigma(i) \text{ for all } j \geq i+1.$$

Note that immediate revisits to stations are allowed. This is again done through an explicit feasible solution for the original pointwise upper bound LP itself.

Define b_σ as in (68), and

$$a_\sigma \quad := \quad \text{Min}_i W_{\sigma i} \quad (= \frac{1}{\mu_{i^*}} \text{ where } i^* := \text{ArgMax } \{i : i \in \sigma\}).$$

Theorem 10. Upper bound for acyclic systems. *Let*

$$\begin{aligned}
\delta_1 &:= Max_\sigma \left[\frac{1}{a_\sigma \prod_{s=1}^{\sigma} \frac{(1-\lambda b_s)}{2} \prod_{s=1}^{\sigma-1} \frac{a_s^2}{b_{s+1}^2}} \right] \\
\delta_{\sigma+1} &:= \delta_\sigma (\frac{1-\lambda b_\sigma}{2}) \frac{a_\sigma^2}{b_{\sigma+1}^2} \text{ for } \sigma = 1, 2, \ldots, S-1.
\end{aligned}$$

Then

$$q_{ij} \quad := \quad \delta_{\sigma(i) \vee \sigma(j)} W_{\sigma(i) \vee \sigma(j), i} W_{\sigma(i) \vee \sigma(j), j}$$

satisfies

$$\lambda q_{1j} + \text{Max}_{\{i : i \in \sigma(j)\}} \mu_i (q_{i+1,j} - q_{ij}) + \sum_{\sigma \neq \sigma(j)} \text{Max}_{i \in \sigma} \mu_i (q_{i+1,j} - q_{ij})^+ \leq -1. \tag{76}$$

Hence, for all non-idling policies,

$$\begin{aligned}
E(x_n) &\leq \lambda \left(\sum_{i=1}^{L} \delta_{\sigma(i)} W_{\sigma(i),i}^2 - \sum_{i=1}^{L-1} \delta_{\sigma(i+1)} W_{\sigma(i+1),i} W_{\sigma(i+1),i+1} \right) \text{ for all } 0 \leq \lambda < \lambda^* \\
&= O(\frac{1}{(1-\rho)^B}).
\end{aligned}$$

Proof: Note that

$$\delta_\sigma = \delta_1 \prod_{s=1}^{\sigma-1} \frac{(1-\lambda b_s)}{2} \frac{a_s^2}{b_{s+1}^2} \geq \frac{2}{a_\sigma(1-\lambda b_\sigma)} \text{ for } \sigma \geq 2.$$

We bound $\mu_i(q_{i+1,j} - q_{ij})$.

Case i: $\sigma(i) < \sigma(j)$

$$\begin{aligned} \mu_i(q_{i+1,j} - q_{ij}) &= \mu_i(\delta_{\sigma(j)} W_{\sigma(j),j} W_{\sigma(j),i+1} - \delta_{\sigma(j)} W_{\sigma(j),j} W_{\sigma(j),i}) \\ &= 0. \end{aligned}$$

Case ii: $\sigma(i) \geq \sigma(j)$ **and** $\sigma(i+1) = \sigma(i)$

$$\begin{aligned} \mu_i(q_{i+1,j} - q_{ij}) &= \mu_i(\delta_{\sigma(i)} W_{\sigma(i),j} W_{\sigma(i),i+1} - \delta_{\sigma(i)} W_{\sigma(i),j} W_{\sigma(i),i}) \\ &= -\delta_{\sigma(i)} W_{\sigma(i),j} \\ &\leq 0. \end{aligned}$$

Case iii: $\sigma(i) \geq \sigma(j)$ **and** $\sigma(i+1) = \sigma(i) + 1$

$$\begin{aligned} \mu_i(q_{i+1,j} - q_{ij}) &= \mu_i\left(\delta_{\sigma(i)+1} W_{\sigma(i)+1,j} W_{\sigma(i)+1,i+1} - \delta_{\sigma(i)} W_{\sigma(i),j} W_{\sigma(i),i}\right) \\ &= \frac{1}{a_{\sigma(i)}} \left(\delta_{\sigma(i)} \frac{(1-\lambda b_{\sigma(i)})}{2} \frac{a_{\sigma(i)}^2}{b_{\sigma(i)+1}^2} b_{\sigma(i)+1}^2 - \delta_{\sigma(i)} W_{\sigma(i),j} a_{\sigma(i)}\right) \\ &= \delta_{\sigma(i)} \left(\frac{(1-\lambda b_{\sigma(i)})}{2} a_{\sigma(i)} - W_{\sigma(i),j}\right) \\ &\leq 0 \text{ (since } a_{\sigma(i)} \leq W_{\sigma(i),j}). \end{aligned}$$

Hence,

$$\begin{aligned} \text{LHS of (76)} &\leq \lambda \delta_{\sigma(j)} b_{\sigma(j)} W_{\sigma(j),j} + \text{Max}\left\{(-\delta_{\sigma(j)} W_{\sigma(j),j}), \delta_{\sigma(j)} \frac{1-\lambda b_{\sigma(j)}}{2} a_{\sigma(j)} - \delta_{\sigma(j)} W_{\sigma(j),j}\right\} \\ &= \lambda \delta_{\sigma(j)} b_{\sigma(j)} W_{\sigma(j),j} + \delta_{\sigma(j)} \frac{(1-\lambda b_{\sigma(j)})}{2} a_{\sigma(j)} - \delta_{\sigma(j)} W_{\sigma(j),j} \end{aligned}$$

$$
\begin{aligned}
&= (\lambda b_{\sigma(j)} - 1)\delta_{\sigma(j)} W_{\sigma(j),j} + \frac{(1-\lambda b_{\sigma(j)})}{2}\delta_{\sigma(j)} a_{\sigma(j)} \\
&\leq \frac{\lambda b_{\sigma(j)} - 1}{2}\delta_{\sigma(j)} a_{\sigma(j)} \text{ (since } a_{\sigma(j)} \leq W_{\sigma(j),j}) \\
&\leq -1.
\end{aligned}
$$
□

If the pole multiplicity bound is too loose, one may wonder whether that is due to the functional bound LPs or the original pointwise bounds themselves. The following tandem example shows that the pointwise bounds themselves are simply not cognizant of distributional results such as Burke's' Theorem.

Example 6: Upper bound for tandem system

Consider the tandem system shown in Figure 10.

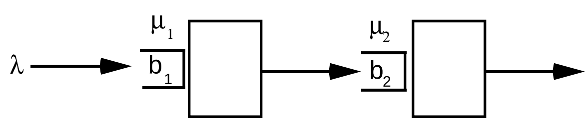

Figure 10: Tandem system of Example 6.

In the balanced case $\mu_1 = \mu_2 = \lambda^*$, the functional upper bound LP gives,

$$M = 2, C_2 = 1, C_1 = 1, C_0 = 0, Q^0 = \frac{1}{\lambda^*}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, Q^{(1)} = \frac{1}{\lambda^*}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, Q^{(2)} = 0,$$

which yields the functional upper bound,

$$E|x(n)| \leq \frac{\rho}{(1-\rho)^2} + \frac{\rho}{1-\rho}.$$

The loose pole multiplicity bound of 2 originates in the original pointwise bounds themselves, since the pointwise LP gives

$$\lim_{\rho \nearrow 1}(1-\rho)^2 \text{ Upper Bound} = 1.$$

However in the unbalanced case, one does obtain first order bounds. If $\mu_1 > \mu_2 = \lambda^*$, then

$$M = 1, Q^{(0)} = \frac{1}{\lambda^*}\begin{bmatrix} \frac{\mu_1}{\mu_1 - \lambda^*} & 1 \\ 1 & 1 \end{bmatrix}, Q^{(1)} = 0,$$

is feasible. This gives the upper bound,

$$E|x_n| \leq \frac{(\frac{\mu_1}{\mu_1 - \lambda^*})\rho}{1 - \rho}.$$

If $\mu_2 > \mu_1 = \lambda^*$, then

$$M = 1, Q^{(0)} = \frac{1}{\lambda^*}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, Q^{(1)} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\mu_2 - \lambda^*} \end{bmatrix},$$

is feasible. This gives

$$E|x_n| \leq \frac{\rho}{1 - \rho} + \frac{\rho\lambda^*}{\mu_2 - \lambda^*}. \qquad \square$$

11 Concluding remarks

For the performance analysis of open Markovian queueing networks, we have obtained LPs that provide uniformly bounding functional expansions for the performance throughout the capacity region. We have shown that our bounds can capture nontrivial interactions between multiple bottleneck stations in heavy traffic.

Several interesting open questions remain. Little appears to be known concerning the pole multiplicity in heavy traffic. This is not surprising since until recently even the issue of stability was not regarded as a major issue. However, it is important to characterize pole multiplicity as well as the growth constants if one wants to comprehend the behavior of a network as traffic increases. A pole multiplicity of M corresponds to a steady state distribution $\pi(|x| = n) = \Omega(n^{(M-1)}\rho^n)$. Currently we are not aware of any examples of high pole multiplicity, though instability for arrival rates short of capacity corresponds to a pole multiplicity of $M = +\infty$. This is a challenging area for the future, as is the whole issue of obtaining uniform functional bounds. By considering a scheduling policy which applies a stable stationary policy off of a large compact set in the state-space, but which within the compact set applies a destabilizing policy, it should be possible to obtain pole multiplicities which are arbitrarily high. Whether such large pole multiplicities can result from buffer priority policies is a more challenging problem. We conjecture that this may be possible too.

Acknowledgments The authors thank all those involved in the reviewing process for their efforts.

References

[1] P. R. Kumar and S. P. Meyn, "Stability of queueing networks and scheduling policies," *IEEE Transactions on Automatic Control*, vol. 40, pp. 251–260, February 1995.

[2] S. Kumar and P. R. Kumar, "Performance bounds for queueing networks and scheduling policies," *IEEE Transactions on Automatic Control*, vol. AC-39, pp. 1600–1611, August 1994.

[3] D. Bertsimas, I. C. Paschalidis, and J. N. Tsitsiklis, "Optimization of multiclass queueing networks: Polyhedral and nonlinear characterizations of achievable performance," *Annals of Applied Probability*, vol. 4, pp. 43–75, 1994.

[4] P. R. Kumar and S. P. Meyn, "Duality and linear programs for stability and performance analysis of queueing networks and scheduling policies," *IEEE Transactions on Automatic Control*, vol. 41, pp. 4–17, January 1996.

[5] W.-B. Gong and J.-Q. Hu, "The Maclaurin series for the GI/G/1 queue," *Journal of Applied Probability*, vol. 29, pp. 176–184, 1992.

[6] M. Girish and J.-Q. Hu, "Higher order approximations for queueing networks," tech. rep., Department of Manufacturing Engineering, Boston University, 1995.

[7] W.-B. Gong and H. Yang, "Rational approximants for some performance problems," tech. rep., University of Massachusetts, Amherst, 1995. To appear in *IEEE Transactions on Computers*.

[8] S. Lippman, "Applying a new device in the optimization of exponential queueing systems," *Operations Research*, vol. 23, pp. 687–710, 1975.

[9] R. W. Cottle, G. J. Habetler, and C. E. Lemke, "On classes of copositive matrices," *Linear Algebra and Its Applications*, vol. 3, pp. 295–310, 1970.

[10] K. G. Murty and S. N. Kabadi, "Some NP-complete problems in quadratic and nonlinear programming," *Mathematical Programming*, vol. 39, pp. 117–129, 1987.

[11] L. Andersson, G. Chang, and T. Elfving, "Criteria for copositive matrices and non-negative Bezier patches," Tech. Rep. LiTH-MAT-R-93-27, Linkoping University and University of Science and Technology, China, August 1993.

[12] V. S. Borkar, "Controlled Markov chains and stochastic networks," *SIAM Journal on Control and Optimization*, vol. 21, no. 4, pp. 652–666, 1983.

[13] S. H. Lu and P. R. Kumar, "Distributed scheduling based on due dates and buffer priorities," *IEEE Transactions on Automatic Control*, vol. AC-36, pp. 1406–1416, December 1991.

[14] J. G. Dai, "On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit model," *Annals of Applied Probability*, vol. 5, pp. 49–77, 1995.

[15] S. Kumar and P. R. Kumar, "Fluctuation smoothing policies are stable for stochastic re-entrant lines," technical report, Coordinated Science Laboratory, University of Illinois, Urbana, IL, 1995. To appear in *Journal of Discrete Event Dynamic Systems: Theory and Applications*.

[16] J. Dai and G. Weiss, "Stability and instability of fluid models for certain re-entrant lines." Preprint, February 1994. To appear in *Mathematics of Operations Research*.