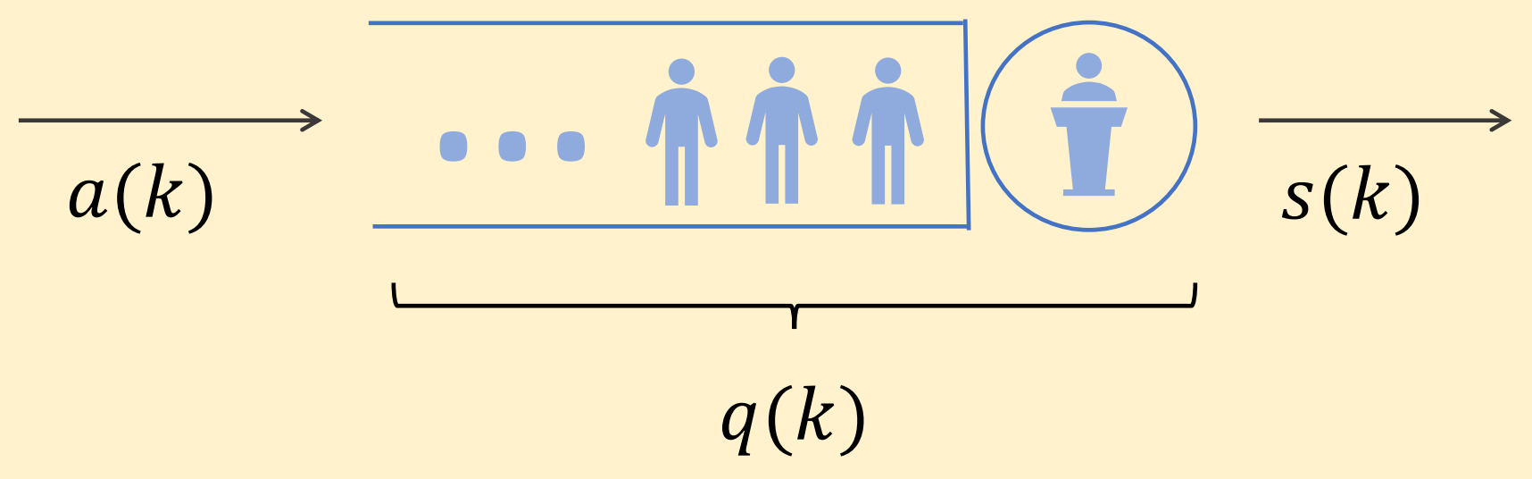
A unified view of the drift method and the MGF method for heavy-traffic analysis

Daniela Hurtado, Siva Theja Maguluri
Industrial and Systems Engineering, Georgia Institute of Technology

Single server queue: Moment Bounds



$q(k)$: Number of customers in system at the beginning of time slot k

$a(k)$: Arrivals in time slot k (i.i.d. process)

$s(k)$: Potential service in time slot k (i.i.d. process)

$u(k)$: Unused service in time slot k

$$\begin{aligned} q(k+1) &= [q(k)+a(k)-s(k)]^+ \\ &= q(k)+a(k)-s(k)+u(k) \end{aligned}$$

$$\Downarrow$$

$$q(t+1)\cdot u(t) = 0$$

Heavy traffic: Let $\epsilon = E[s(1)] - E[a(1)]$. The Heavy Traffic limit is when $\epsilon \to 0$

Kingman bound:
In steady-state, $E[q(k+1)^2] = E[q(k)^2]$

$$\Rightarrow E[q] = \frac{E[(s-a)^2]}{2\epsilon} - \frac{E[u^2]}{2\epsilon} \quad \Rightarrow \quad \lim_{\epsilon\to 0} \epsilon E[q] = \frac{\sigma_a^2+\sigma_s^2}{2}$$

Higher moments:

$V(q) = q^m$ + Steady-state $E[\Delta V(q)] = 0$ ⇒ Induction ⇒ Moment bounds

+ Heavy-traffic ⇒ Scaled moments

$V(q)$:	q^2	q^3	…	q^{m+1}	…
Obtain:	$\epsilon\, E[q]$	$\epsilon^2\, E[q^2]$	…	$\epsilon^m\, E[q^m]$	…

Add all these to get exponential

Idea!
Use e^q as the test function

Single server queue: MGF

Theorem: The scaled queue length ϵq converges in distribution to an exponential r.v. with mean $\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)$

Idea of our proof:

Test function:
$V(q) = e^{\epsilon\theta q}$

$$\left(e^{\epsilon\theta q(k+1)}-1\right)\left(e^{-\epsilon\theta u(k)}-1\right) = 0$$

$$E\left[e^{\epsilon\theta q}\right] = \frac{1-E\left[e^{-\epsilon\theta u}\right]}{1-E\left[e^{-\epsilon\theta(s-a)}\right]} \qquad \lim_{\epsilon\to 0} \approx \frac{0}{0}$$

⇓ Taylor approximation

$$E\left[e^{\epsilon\theta q}\right] = \frac{\theta\epsilon^2 + O(\epsilon^3)}{\theta\epsilon^2 - \frac{\theta^2\epsilon^2}{2}(\sigma_a^2+\sigma_s^2) + O(\epsilon^3)}$$

⇓

$$\lim_{\epsilon\to 0} E\left[e^{\epsilon\theta q}\right] = \frac{1}{1-\theta\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)}$$

MGF of expo r.v. with mean
$\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)$

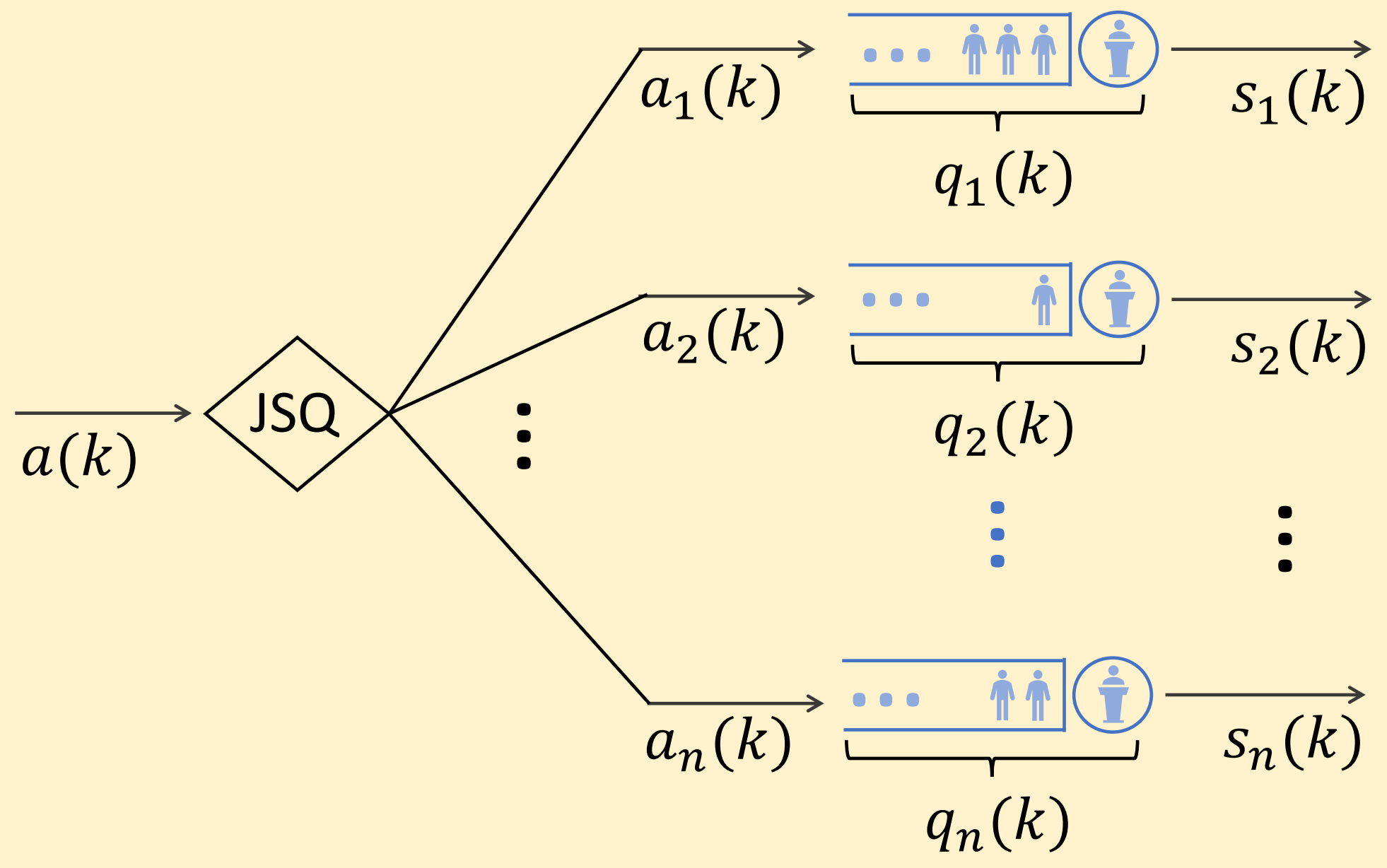
MGF method [1]: Use Moment Generating Function to prove tightness

References:

[1] Braverman, A., Dai, J. G., & Miyazawa, M. (2017). Heavy traffic approximation for the stationary distribution of a generalized Jackson network: The BAR approach. Stochastic Systems, 7(1), 143-196.

[2] Eryilmaz, A., & Srikant, R. (2012). Asymptotically tight steady-state queue length bounds implied by drift conditions. Queueing Systems, 72(3-4), 311-359.

Load balancing problem



For $i \in \{1,\dots,n\}$, let:

$q_i(k)$: Number of customers in system i at the beginning of time slot k

$a_i(k)$: Arrivals to system i in time slot k

$s_i(k)$: Potential service in system i in time slot k

$u_i(k)$: Unused service in system i in time slot k

$a(k)$: Arrivals to the system in time slot k

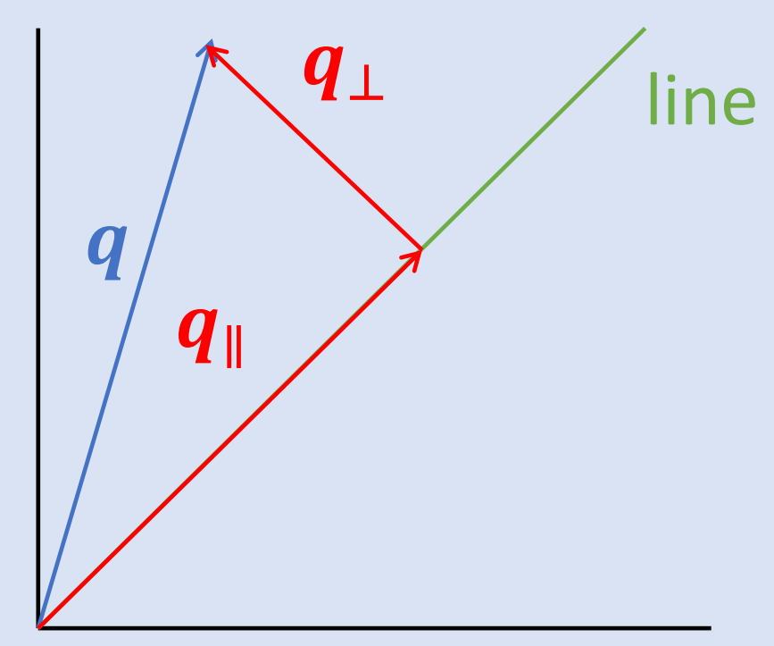
$\boldsymbol{q}(t) = \left(q_1(t),\dots,q_n(t)\right)$

$\boldsymbol{s}(t) = \left(s_1(t),\dots,s_n(t)\right)$

$\boldsymbol{u}(t) = \left(u_1(t),\dots,u_n(t)\right)$

Heavy traffic: Let $\epsilon = \sum E[s_i(1)] - E[a(1)]$.
The heavy-traffic limit is when $\epsilon \to 0$

State space collapse: In HT limit, the n-dimensional $\boldsymbol{q}$ collapses to a line. Then, $E[\|\boldsymbol{q}_\perp\|^r] \le N_r$



Theorem: The scaled vector of queue lengths $\epsilon\boldsymbol{q} \Rightarrow X\mathbf{1}$, where X is an exponential r.v. with mean $\left(\frac{\sigma_a^2+\sum\sigma_{s_i}^2}{2}\right)$

Idea of our proof:

Test function:
$V(\boldsymbol{q}) = e^{\epsilon\theta\sum q_i}$

$$E\left[\left(e^{\epsilon\theta\sum q_i(k+1)}-1\right)\left(e^{-\epsilon\theta\sum u_i(k)}-1\right)\right]$$
is $o(\epsilon^2)$

$$E\left[e^{\epsilon\theta\sum q_i}\right] = \frac{1-E\left[e^{-\epsilon\theta\sum u_i}\right]}{1-E\left[e^{-\epsilon\theta(\sum s_i - a)}\right]}$$

⇒

$$\epsilon\sum_i q_i = \epsilon q_{\parallel j} \Rightarrow X$$
$$X \sim Expo\left(\left(\frac{\sigma_a^2+\sum\sigma_{s_i}^2}{2}\right)^{-1}\right)$$

+ $\lim_{\epsilon\to 0}\epsilon^2 E[\|\boldsymbol{q}_\perp\|^2] = 0$ & $\boldsymbol{q} = \boldsymbol{q}_\parallel + \boldsymbol{q}_\perp$ ⇒ $\epsilon\boldsymbol{q} \Rightarrow X\mathbf{1}$