TRANSFORM METHODS FOR HEAVY-TRAFFIC ANALYSIS

Daniela Hurtado Lange, Siva Theja Maguluri

Industrial and Systems Engineering

Georgia Institute of Technology

Georgia Institute of Technology

Overall

- #8 Top Public Universities
- #4 Most Innovative Schools
- #35 National Universities
- #71 (out of 1,000) Best Global University

As of September 2018

Undergraduate

- **#4 Best Undergraduate Engineering Program**
  - #1 Industrial / Manufacturing Engineering
  - #2 Aerospace / Aeronautical / Astronautical Engineering
  - #2 Chemical Engineering
  - #2 Civil Engineering
  - #2 Mechanical Engineering
  - #3 Biomedical Engineering
  - #3 Materials Science & Engineering
  - #4 Electrical / Electronic / Communications Engineering
  - #4 Environmental / Environmental Health Engineering
  - #5 Computer Engineering
- **#21 Best Undergraduate Business Program**
  - #6 Management Information Systems
  - #7 Quantitative Analysis
  - #7 Production / Operations Management
  - #8 Supply Chain Management / Logistics

As of September 2018

Graduate

- **#8 Best Computer Science Graduate Program**
  - #7 Artificial Intelligence
  - #9 Theory
  - #10 Systems
- **#8 Best Graduate Engineering School**
  - #1 Industrial Engineering
  - #2 Civil Engineering
  - #2 Biomedical Engineering
  - #4 Aerospace Engineering
  - #4 Computer Engineering
  - #5 Electrical Engineering
  - #5 Environmental Engineering
  - #5 Mechanical Engineering
  - #6 Chemical Engineering
  - #7 Materials Science & Engineering
  - #9 Nuclear Engineering
- **#28 Best Business School**
  - #7 Production / Operations

As of March 2018

Industrial and Systems Engineering

ISyE Rankings

#1 Undergraduate Program

- The 24th consecutive No. 1 ranking for ISyE. (USNWR's America's Best Colleges)

#1 Graduate Program

- The 28th consecutive No. 1 ranking for ISyE in the industrial/manufacturing/systems specialty. (USNWR's Best Graduate Programs)

#7 Statistics & Operational Research

- (2018 QS World University Rankings)

MBA
Master
PhD

Outline

- Heavy-Traffic analysis in the past
- MGF method
- Future work

Single server queue: Kingman, 1962

Diffusion limits

- Write a recursion for the waiting time
- Approximate the waiting time by a Reflected Brownian Motion (RBM)
- Compute steady-state distribution of the RBM
- Widely used in a variety of systems

Kingman's bound

- Write a recursion for the waiting time:
$$W_{n+1} = f(W_n)$$
- In steady-state, $E[W_{n+1}] = E[W_n]$
  - Algebra ⇒ Upper bound for $E[W_n]$
- Drift Method [Eryilmaz, Srikant 2013]

Heavy-Traffic: Diffusion Limits

- Kingman (1962):
  - Heavy-traffic parameter: $\epsilon = \mu - \lambda, n = \frac{1}{\epsilon^2}$

$$\frac{q(nk)}{\sqrt{n}} = \epsilon q\left(\frac{k}{\epsilon^2}\right) \Longrightarrow \widehat{W} \text{ as } \epsilon \to 0$$

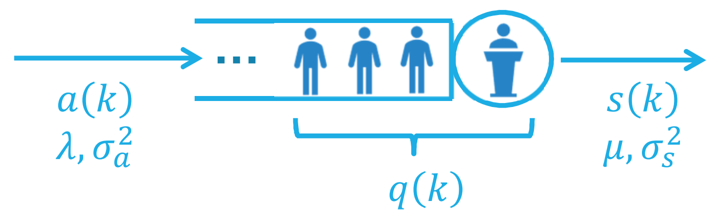
$$\downarrow$$

RBM on a line

$$\downarrow$$

Steady-state = $Expo\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)$

- Interchange of limits: $\epsilon q \Longrightarrow Expo\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)$



$$q(k+1) = [q(k) + a(k) - s(k)]^+$$

$$\begin{tikzcd}
\epsilon X^{(\epsilon)}(k) \arrow[r, "k \to \infty"] \arrow[d, "\epsilon \to 0"'] & \epsilon X^{(\epsilon)}(\infty) \arrow[d, "\epsilon \to 0"] \\
\tilde{X}(t) \arrow[r, "t \to \infty"'] & X^*
\end{tikzcd}$$

Heavy-Traffic: Drift Method

- Kingman, 1962:
  - In steady state:

$$E[q^2(k+1)] = E[q^2(k)]$$

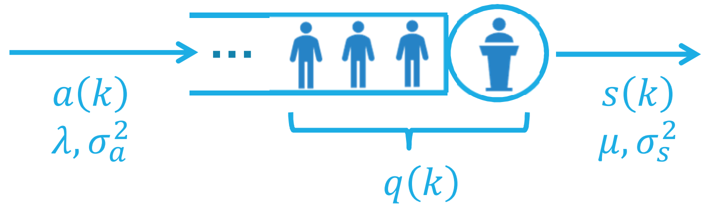
  - Yields:

$$E[q] = \frac{E[(a-s)^2]}{\underbrace{2(\mu-\lambda)}_{\epsilon}} - \frac{E[u^2]}{2(\mu-\lambda)}$$

It can be shown that this term is at most a constant and so small compared to the first term when $\lambda \to \mu$

  - Therefore,

$$\lim_{\epsilon \to 0} E[\epsilon q] = \frac{\sigma_a^2 + \sigma_s^2}{2}$$

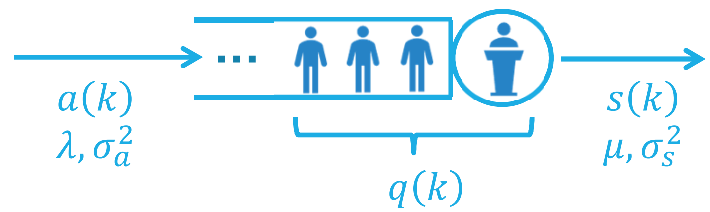


$$\begin{aligned} q(k+1) &= [q(k) + a(k) - s(k)]^+ \\ &= q(k) + a(k) - s(k) + u(k) \end{aligned}$$

$$\boxed{q(k+1)u(k) = 0}$$

Drift Method

- Drift method [Eryilmaz, Srikant 13]:

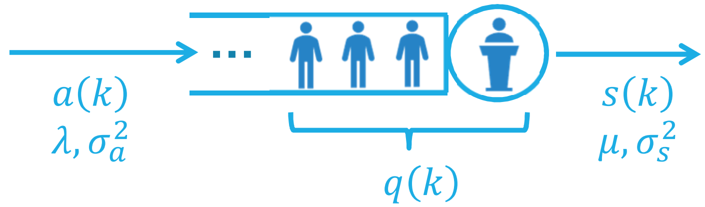


$V(q)$:	q^2	q^3	$\cdots$	q^{m+1}
Obtain:	$E[\epsilon q]$	$E[\epsilon^2 q^2]$	$\cdots$	$E[\epsilon^m q^m]$
	$\downarrow$	$\downarrow$		$\downarrow$
	$\frac{\sigma_a^2+\sigma_s^2}{2}$	$2\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)^2$		$m!\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)^m$

$$\boxed{\epsilon q \Longrightarrow Expo\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)}$$

Drift Method

- Drift method [Eryilmaz, Srikant 13]:



$$V(q): \quad 1 + \epsilon q + \frac{1}{2}\epsilon^2 \boldsymbol{q^2} + \frac{1}{3!}\epsilon^3 \boldsymbol{q^3} + \cdots + \frac{1}{(m+1)!}\epsilon^{m+1} \boldsymbol{q^{m+1}} + \cdots \; = e^{\epsilon q}$$

Obtain:

$$1 + \boldsymbol{E[\epsilon q]} + \frac{1}{2}\boldsymbol{E[\epsilon^2 q^2]} + \cdots \; + \frac{1}{m!}E[\epsilon^m q^m] + \cdots \; = E[e^{\epsilon q}]$$

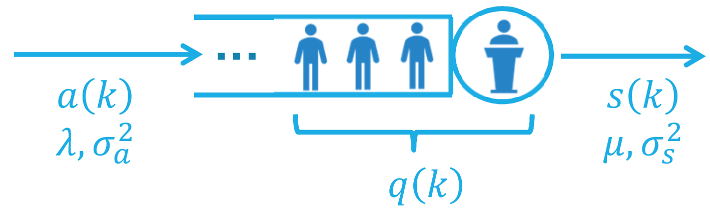
$$\boldsymbol{E[\epsilon q]} \downarrow \frac{\sigma_a^2 + \sigma_s^2}{2} \qquad \boldsymbol{E[\epsilon^2 q^2]} \downarrow 2\left(\frac{\sigma_a^2 + \sigma_s^2}{2}\right)^2 \qquad E[\epsilon^m q^m] \downarrow m!\left(\frac{\sigma_a^2 + \sigma_s^2}{2}\right)^m$$

$$\boxed{\epsilon q \Longrightarrow Expo\left(\frac{\sigma_a^2 + \sigma_s^2}{2}\right)}$$

Use $e^{\theta \epsilon q}$ as test function

MGF Method

- In steady-state: $E\left[e^{\theta\epsilon q(k+1)}\right] = E\left[e^{\theta\epsilon q(k)}\right]$



- **Key Lemma**:

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

$$q(k+1) = q(k) + a(k) - s(k) + u(k)$$

$$\boxed{q(k+1)u(k) = 0}$$

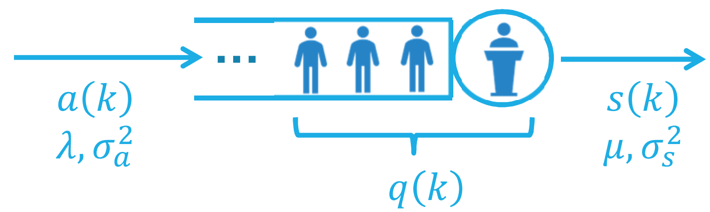
- Proof: $e^x - 1 = 0$ iff $x = 0$

- From the lemma:

$$E\left[e^{\theta\epsilon q(k+1)}\right] = E\left[e^{\theta\epsilon\left(q(k)+a(k)-s(k)\right)} - e^{-\theta\epsilon u(k)} + 1\right]$$

$$= E\left[e^{\theta\epsilon q(k)}\right]$$

MGF Method



$$\Longrightarrow E\left[e^{\theta\epsilon q}\right] = \frac{1-E\left[e^{-\theta\epsilon u}\right]}{1-E\left[e^{\theta\epsilon(a-s)}\right]}$$

$$= \frac{\theta\epsilon E[u]+o(\epsilon^2)}{-\theta\epsilon E[a-s]-\frac{\theta^2\epsilon^2}{2}E[(a-s)^2]+O(\epsilon^3)}$$

$$= \frac{\overset{1}{\cancel{\theta\epsilon^2}}+o(\epsilon^2)}{\underset{1}{\cancel{\theta\epsilon^2}}-\underset{\theta/2}{\cancel{\frac{\theta^2\epsilon^2}{2}}}(\sigma_a^2+\sigma_s^2+\epsilon^2)+O(\epsilon^3)}$$

$$\xrightarrow[\epsilon\to 0]{} \frac{1}{1-\theta\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)}$$

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

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

$$E\left[e^{\theta\epsilon q(k)}\right]$$
$$= E\left[e^{\theta\epsilon\left(q(k)+a(k)-s(k)\right)}-e^{-\theta\epsilon u(k)}+1\right]$$

Therefore, $\epsilon q \Longrightarrow X$

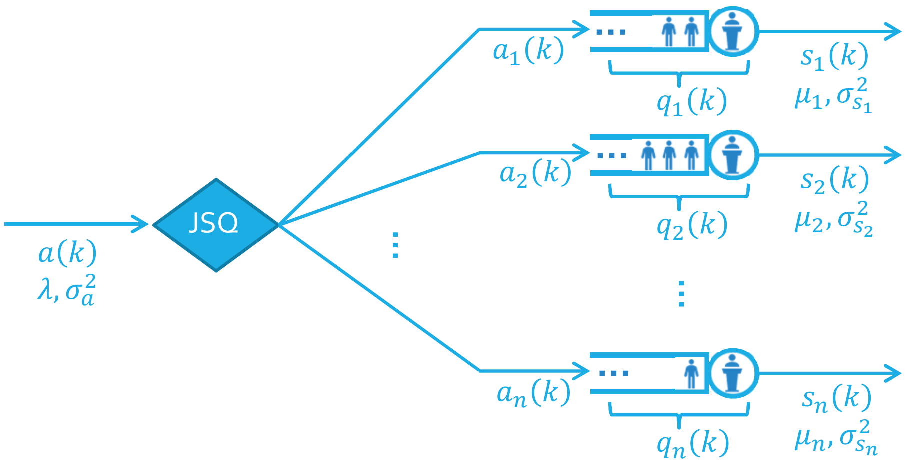
$$X \sim Expo\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right)$$

Load Balancing System

- Discrete time model
- One stream of i.i.d. arrivals routed using Join the Shortest Queue (JSQ) policy
- Dynamics of the queues:

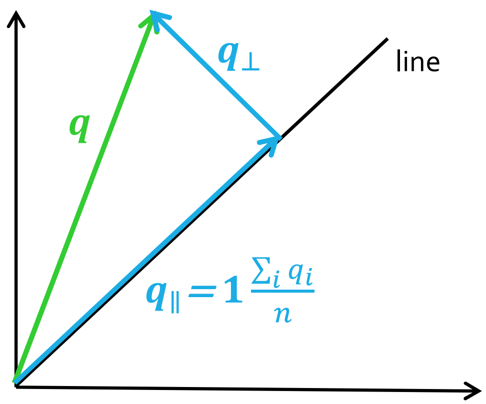
$$q_i(k+1) = [q_i(k) + a_i(k) - s_i(k)]^+$$
$$= q_i(k) + a_i(k) - s_i(k) + u_i(k)$$

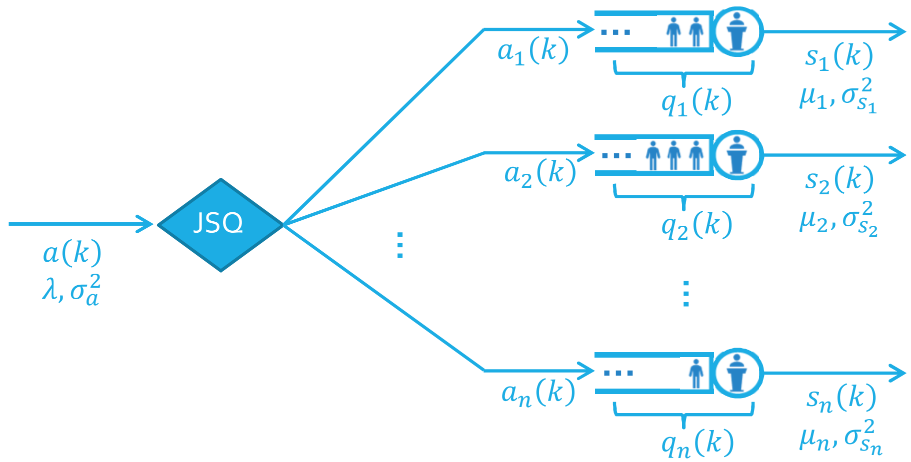
$\boxed{q_i(k+1)u_i(k) = 0 \text{ but } q_i(k+1)u_j(k) \neq 0 \text{ if } i \neq j}$



State Space Collapse

- Heavy-traffic: $\epsilon = \sum_i \mu_i - \lambda > 0$





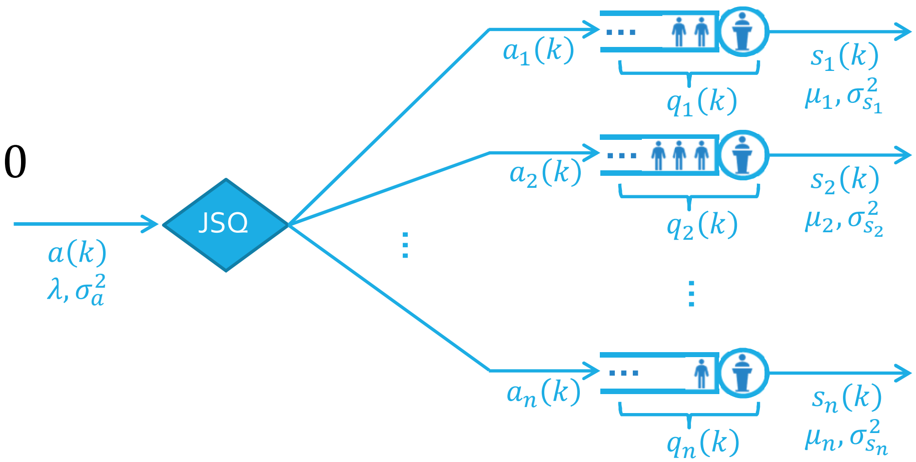
$$\boldsymbol{q}(k+1) = \boldsymbol{q}(k) + \boldsymbol{a}(k) - \boldsymbol{s}(k) + \boldsymbol{u}(k)$$

- State Space Collapse (SSC) [Eryilmaz, Srikant 2013]:
  All queues are (approximately) equal

$$E[\|\boldsymbol{q}_\perp\|^r] \le M_r$$

Unused Service

- We know: $q_i(k+1)u_i(k) = 0$
- Under SSC: $q_i(k+1)u_j(k) \approx q_j(k+1)u_j(k) = 0$



- Therefore, under SSC:

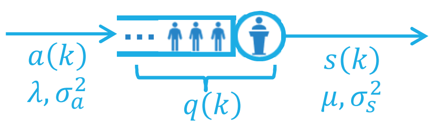
$$\left(\sum_i q_i(k+1)\right)\left(\sum_i u_i\right) \approx 0$$

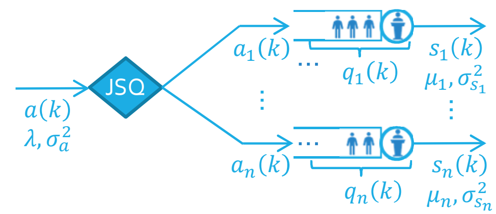
$$\boldsymbol{q}(k+1) = \boldsymbol{q}(k) + \boldsymbol{a}(k) - \boldsymbol{s}(k) + \boldsymbol{u}(k)$$

MGF method: Set to zero the drift of $V(\boldsymbol{q}) = e^{\theta\epsilon \sum_i q_i}$ &

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

MGF Method





$$\boldsymbol{q}_{\parallel} = \mathbf{1}\frac{\sum_i q_i}{n}$$
$$\& \ \boldsymbol{q}_{\perp} = \boldsymbol{q} - \boldsymbol{q}_{\parallel}$$

- Set drift of $V(q) = e^{\theta\epsilon q}$ to zero

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

- Yields

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

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

- Set drift of $V(\boldsymbol{q}) = e^{\theta\epsilon\sum_i q_i}$ to zero

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

- Yields

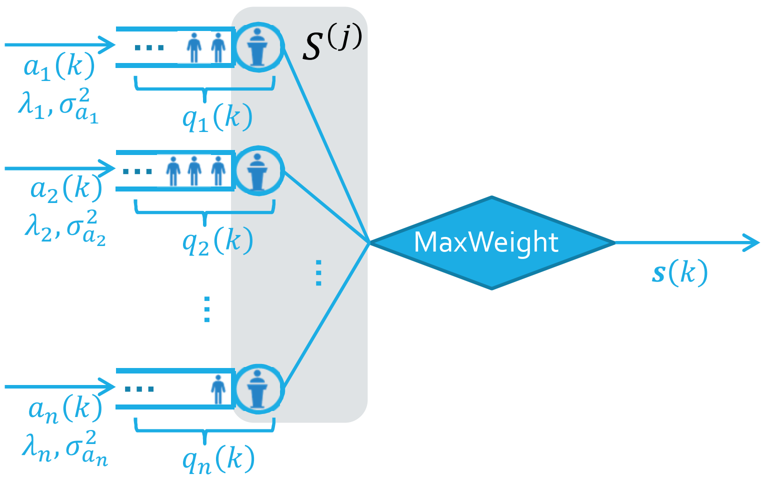
$$\lim_{\epsilon\to 0} E\left[e^{\theta\epsilon\sum_i q_i}\right] = \frac{1}{1 - \theta\left(\frac{\sigma_a^2 + \sum_i \sigma_{s_i}^2}{2}\right)}$$

- Then, $\epsilon\boldsymbol{q}_{\parallel} \Longrightarrow \frac{X}{n}\mathbf{1}$, where $X \sim Expo\left(\frac{\sigma_a^2+\sum_i \sigma_{s_i}^2}{2}\right)$
- SSC implies $\epsilon\boldsymbol{q}_{\perp} \Longrightarrow \mathbf{0}$
- Also, $\epsilon\boldsymbol{q} = \epsilon\boldsymbol{q}_{\parallel} + \epsilon\boldsymbol{q}_{\perp}$

$$\boxed{\epsilon\boldsymbol{q} \Rightarrow \frac{X}{n}\mathbf{1}}$$

Generalized Switch

- Discrete time model
- n queues with i.i.d. arrival process and a server
  - Arrival to different queues are independent
- Interference constraints
  - S is the set of feasible service rate vectors
- Scheduling: MaxWeight
- Fading
  - Channel state is random
  - If the channel state is j, the set of feasible service rates is $S^{(j)}$
  - ψ_j is the probability that the channel is in state j



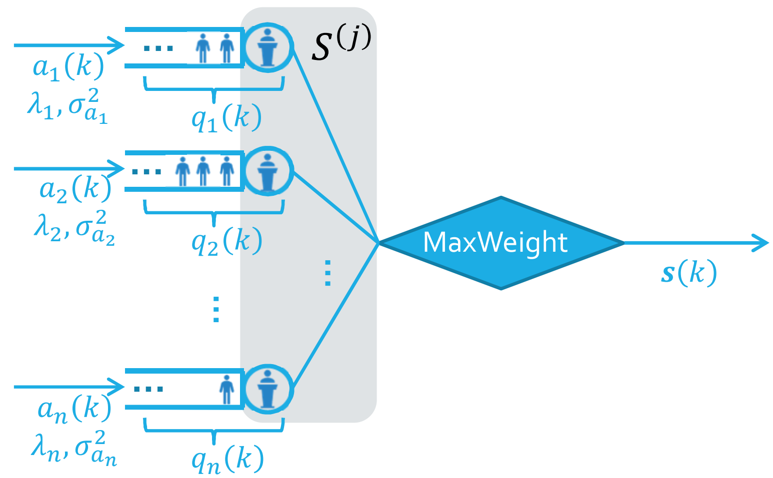
Generalized Switch

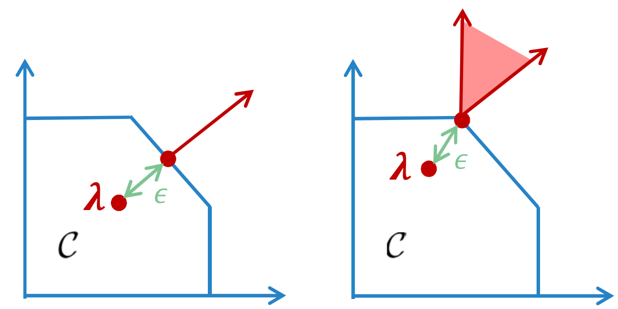
- Capacity region

$$C = \sum_j \psi_j \, ConvexHull\left(S^{(j)}\right)$$

- Heavy-traffic: ϵ is the distance between λ and the boundary of C

- State Space Collapse
  - One dimensional
  - Multidimensional





Overview

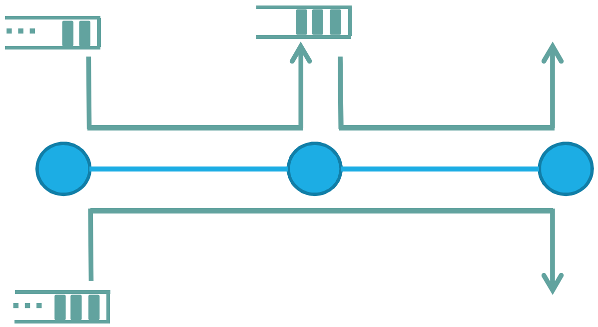
- Transform (MGF) Method for Heavy-Traffic
  - Use exponential test function in the drift method
  - Further simplifies the Drift Method (which simplified the Diffusion Method)
  - Unifies the Drift Method and BAR method [Braverman, Miyazawa, Dai 15] for Heavy-Traffic
- Complete Resource Pooling (One dimensional SSC)
  - Single Server Queue
  - Load Balancing
  - Generalized Switch
- Ongoing Work
  - When CRP is not satisfied (Multi dimensional SSC)
  - Bandwidth Sharing Network
  - Input Queued Switch

$$\begin{tikzcd}
\epsilon q^{(\epsilon)}(k) \arrow[r, "k \to \infty"] \arrow[d, "\epsilon \to 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \to 0"] \\
\tilde{q}(t) \arrow[r, "t \to \infty"'] & q^{*}
\end{tikzcd}$$

Direct Methods

Bandwidth Sharing Network

- Continuous time model
- Network with L links and R flows
- Flow r:
  - Arrivals are Poisson(λ_r)
  - Job sizes are Exponential(μ_r)
- Each link has capacity C_l
- Rates allocation: Proportional scheduling



$L = 2$
$R = 3$

[Massoulié et. al 2007; Paganini et al 2012]

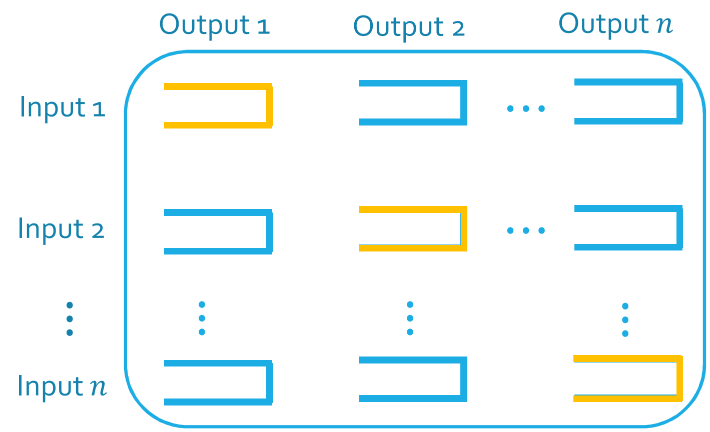
$$\begin{tikzcd}
\epsilon q^{(\epsilon)}(k) \arrow[r, "k \to \infty"] \arrow[d, "\epsilon \to 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \to 0"] \\
\tilde{q}(t) \arrow[r, "t \to \infty"'] & q^*
\end{tikzcd}$$

[Wang, et.al, 2018]

[Kang et.al, 2009 ; Ye and Yao, 2012; Vlasiou et.al, 2014]

Input-queued Switch

- nxn input queued switch in discrete time
  - Packets arrive to input ports
  - Service takes exactly 1 time slot
  - At most one queue from each row, and one from each column can be served in each time slot
  - Scheduling: MaxWeight



- Drift method [Maguluri, Srikant 2016]

$$\lim_{\epsilon\to 0} \epsilon\, E\left[\sum_{i,j} q_{ij}\right] = \left(1-\frac{1}{2n}\right)\|\sigma^2\|$$

Thank you!

Questions?