A queueing system with on-demand servers

A. Stolyar (UIUC)

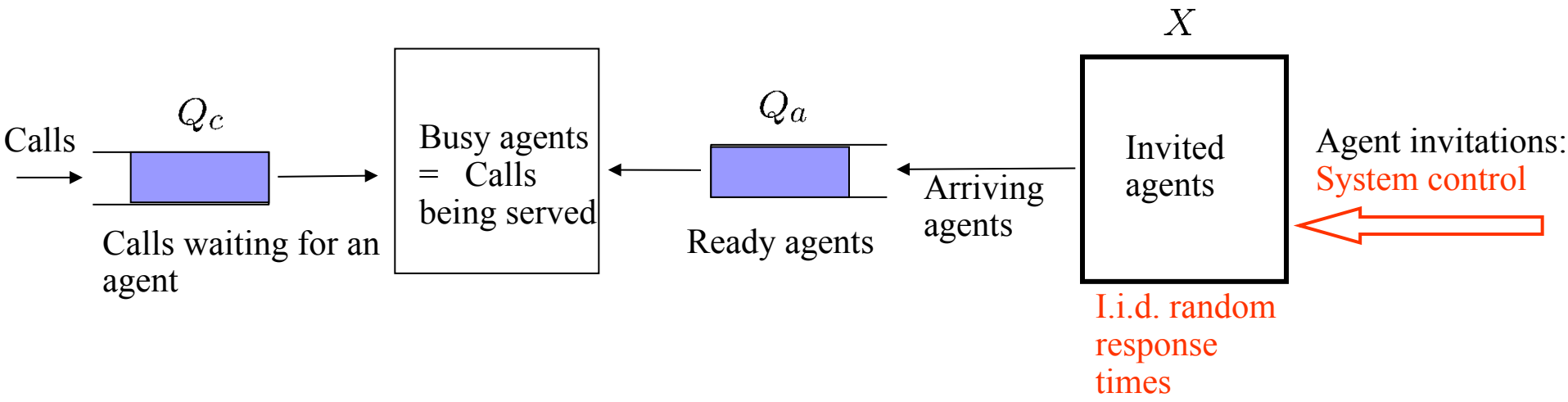
Joint work(s) with G.Pang (Pennstate), Q.Wang (UIUC), L.Nguyen (Lehigh Univ.)

June 26, 2018

Outline

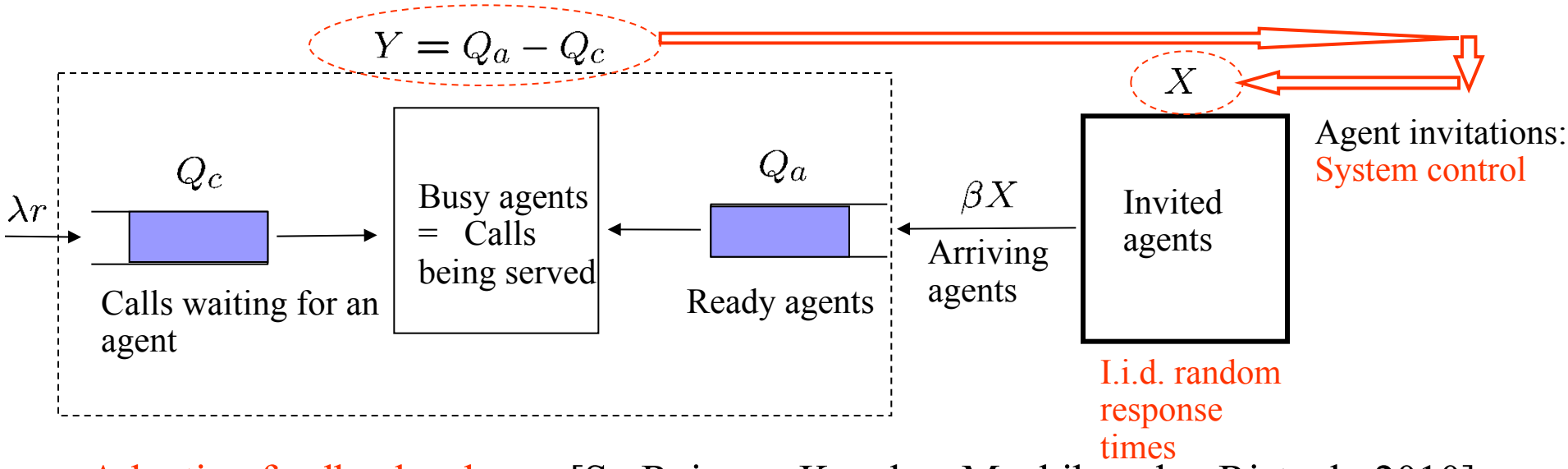
- Basic model
  - Adaptive algorithm
  - Generalized base stock (GBS) algorithm
- More general model under Adaptive algorithm

Motivation



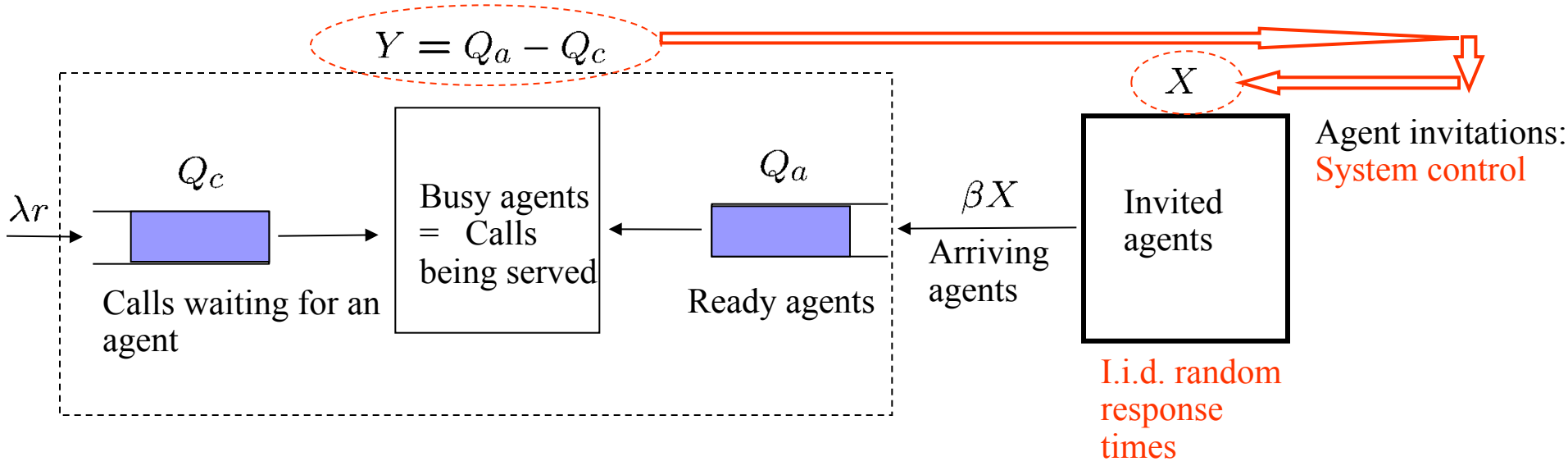
- Objective: Keep delays of both calls and agents low
- Applications:
  - Call/contact centers
  - Classical single-item inventory system with order crossovers
  - Telemedicine
  - Uber, etc.
  - ...

Model. Adaptive algorithm



- Adaptive feedback scheme [S., Reiman, Korolev, Mezhibovsky, Ristock, 2010]:
  - X is incremented by $[-\gamma \Delta Y]$ each time Y changes by ΔY *(=+1 or -1)*, $\gamma>0$ is parameter
  - Independently, X is incremented by $[-\text{sign}(Y)]$ at the instantaneous rate $|\varepsilon Y|$, where $\varepsilon>0$ is parameter
- Adaptive = Does not require knowledge of call rate or any system parameters
- We analyze it when $r \to \infty$, and prove, in particular, that steady-state delays vanish
- "Small subtlety": Assume for now that invited agents can be uninvited if necessary

Adaptive algorithm basic dynamics



INFORMALLY: $(d/dt)X = -\gamma(d/dt)Y - \epsilon Y$

$$(d/dt)Y = \beta X - \lambda r$$

Process. Fluid and diffusion scaling

- The average X to match the arrival rate is $X_* = \lambda r/\beta$
- Fluid scaled process

$$(x^r, y^r) = (X - \lambda r/\beta, Y)/r = (X/r - \lambda/\beta, Y/r)$$

Boundary: $x^r \geq -\lambda/\beta$

- Diffusion scaled process

$$(\hat{X}^r, \hat{Y}^r) = (X - \lambda r/\beta, Y)/\sqrt{r}$$

Fluid and diffusion limit dynamics
when away from boundary

- Fluid limit

$$x' = -\gamma y' - \epsilon y \qquad x' = -\gamma\beta x - \epsilon y$$
$$y' = \beta x \qquad y' = \beta x$$

Stable for any positive β, γ, ε.

When $\varepsilon < \gamma^2\beta/4$, there are two distinct eigenvectors; we assume that

- Diffusion limit

$$d\hat{X} = -\gamma\beta\hat{X}\,dt - \epsilon\hat{Y}\,dt + \gamma\sqrt{2\lambda}dW$$
$$d\hat{Y} = \beta\hat{X}\,dt - \sqrt{2\lambda}dW$$

Gaussian stationary distribution, zero mean, covariance matrix:

$$\begin{bmatrix} \frac{\lambda(\beta\gamma^2+\epsilon)}{\beta^2\gamma} & -\frac{\lambda}{\beta} \\ -\frac{\lambda}{\beta} & \frac{\lambda}{\beta\gamma} \end{bmatrix}$$

Main results for the Adaptive algorithm

Theorem 1. *For all large r the system is stable. In stationary regime:* $(x^r(\infty), y^r(\infty)) \Rightarrow (0,0)$.

Theorem 2. *The sequence of stationary distributions of* $(\hat{X}^r(\infty), \hat{Y}^r(\infty))$ *is tight. Consequently, the limit-interchange holds:*

$$(\hat{X}^r(\infty), \hat{Y}^r(\infty)) \Rightarrow (\hat{X}(\infty), \hat{Y}(\infty))$$

Related work

- Inventory models (esp., with order crossovers): Karlin-Scarf'58, Zipkin'00, Disney et al.'16, ...
- Double-ended queues and matching systems: Kashyap'66, Caldentey et al.'09, Adan-Weiss'12, Adan et al.'15, Buke-Chen'15, Bušic et al.'10, Mairesse-Moyal'14, Gurvich-Ward'14, Bušic-Meyn'14, ...
- Diffusion-scale limit interchange in many-servers regime: Halfin-Whitt'81, Jelenkovic-Mandelbaum-Momcilovic'04, Mandelbaum-Momcilovic'05, Gamarnik-Momcilovic'08, Reed'09, Gamarnik-Goldberg'11, Gamarnik-S.'12, Dai-Dieker-Gao'14, Gurvich-Whitt'09, S.-Yudovina'13, S.-Yudovina'12, S.'13, ..., see Jim Dai's talk yesterday

Actual fluid limit dynamics

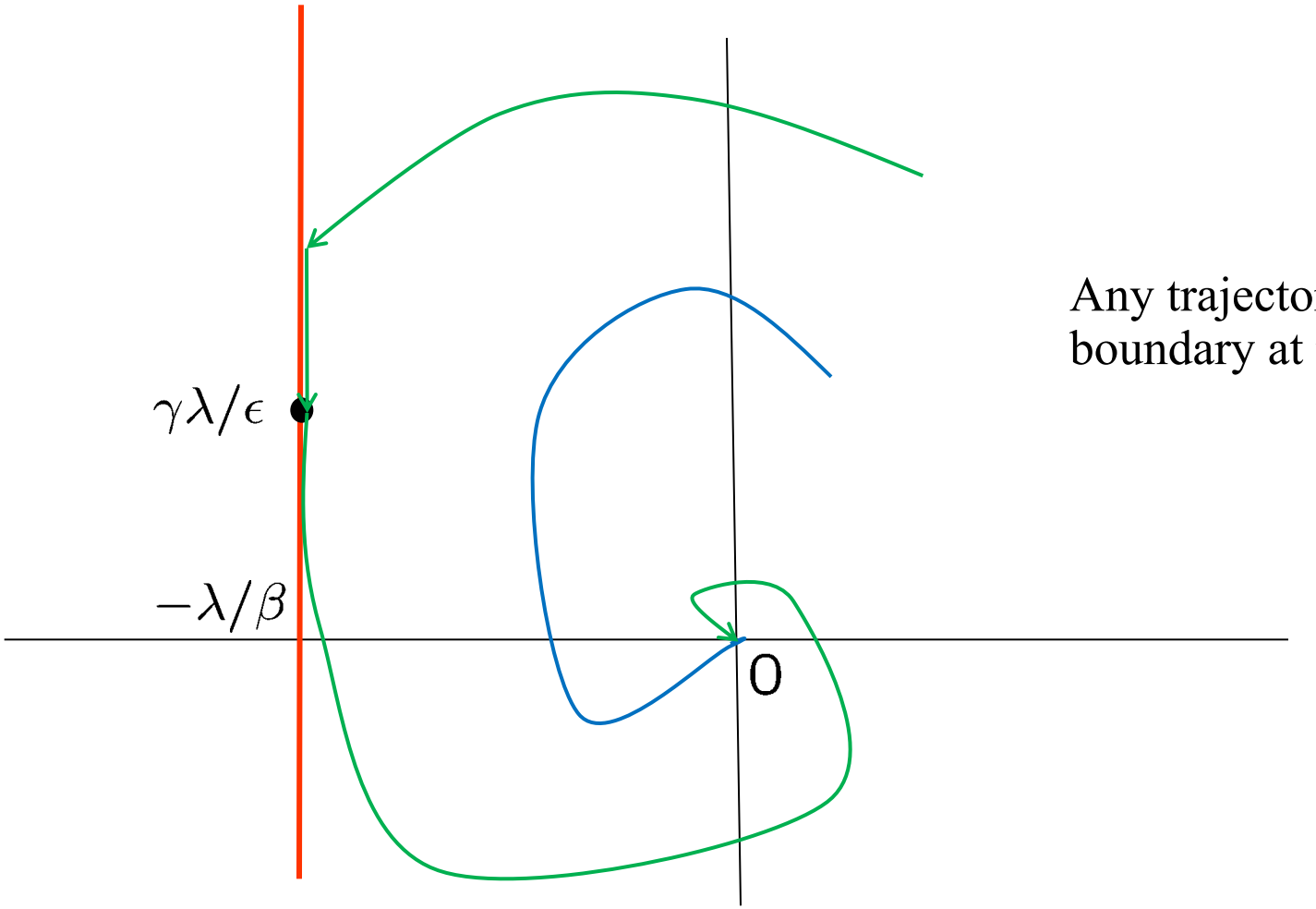
- Away from boundary

$$x' = -\gamma\beta x - \epsilon y$$
$$y' = \beta x$$

- On the boundary ($x = -\lambda/\beta$)

$$x' = [\gamma\lambda - \epsilon y] \vee 0$$
$$y' = \beta x$$

Actual fluid limit dynamics

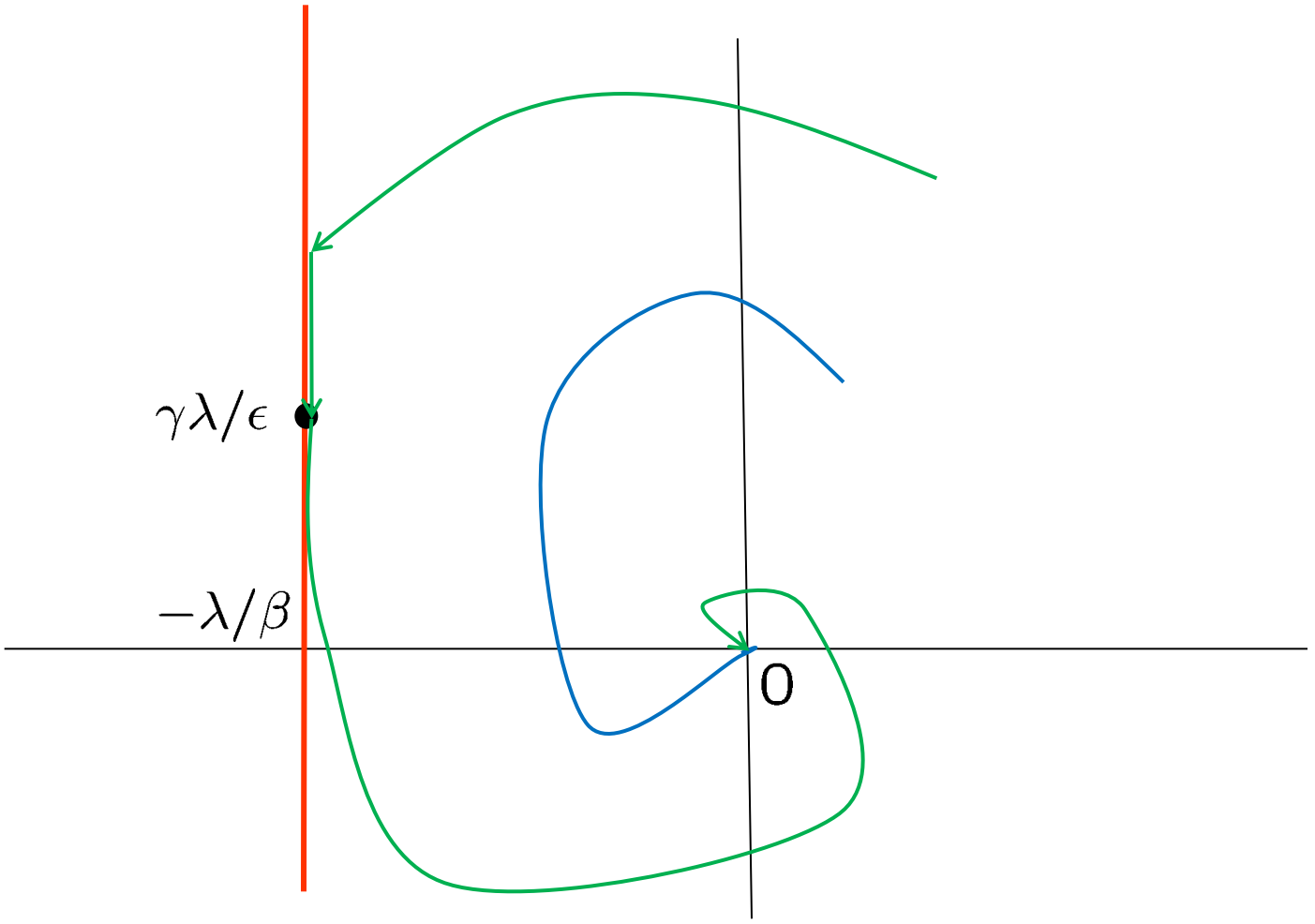


Any trajectory can hit boundary at most once

Proposition 3. *Fluid limit trajectories converge to zero, uniformly on initial states (from a bounded set).*

This does not directly imply stability of the process!

Actual fluid limit dynamics



Lemma 4. $\exists C > 0, \eta > 0$ *s.t.* $\|(x(t), y(t))\| \geq C$ *implies* $(d/dt)\|(x(t), y(t))\|_* \leq -\eta$.

$\|\cdot\|_*$ is Euclidean norm w.r.t. eigenvectors

Stability. Fluid scale tightness of stationary distributions

- Use Lyapunov drift condition for an imbedded chain, sampled at stopping times
- $c>0$ and $\delta>0$ are fixed constants
- $s^r=(x^r,y^r)$. For a given $s^r(0)$, the stopping time

$$\tau = c \wedge \inf\{t \geq 0 : \quad |\|s^r(t)\|_* - \|s^r(0)\|_*| \geq \delta\}$$

Lemma 5. *Uniformly in large r and large $\|s^r(0)\|_*$:*

$$E[\|s^r(\tau)\|_*^2 - \|s^r(0)\|_*^2 \mid s^r(0)] \leq -C_1\|s^r(0)\|_* + C_2.$$

- From here, for the sampled chain

$$E\|\widehat{s}^r(\infty)\|_* \leq C_2/C_1$$

- Then, using lower bounds on expected sampling intervals (τ 's),

$$E\|s^r(\infty)\|_* \leq C_3$$

- This and Proposition 3 implies Theorem 1

Theorem 2 proof idea

- Theorem 1 is a starting point. Uses approach in S.-Yudovina'12, S.'13

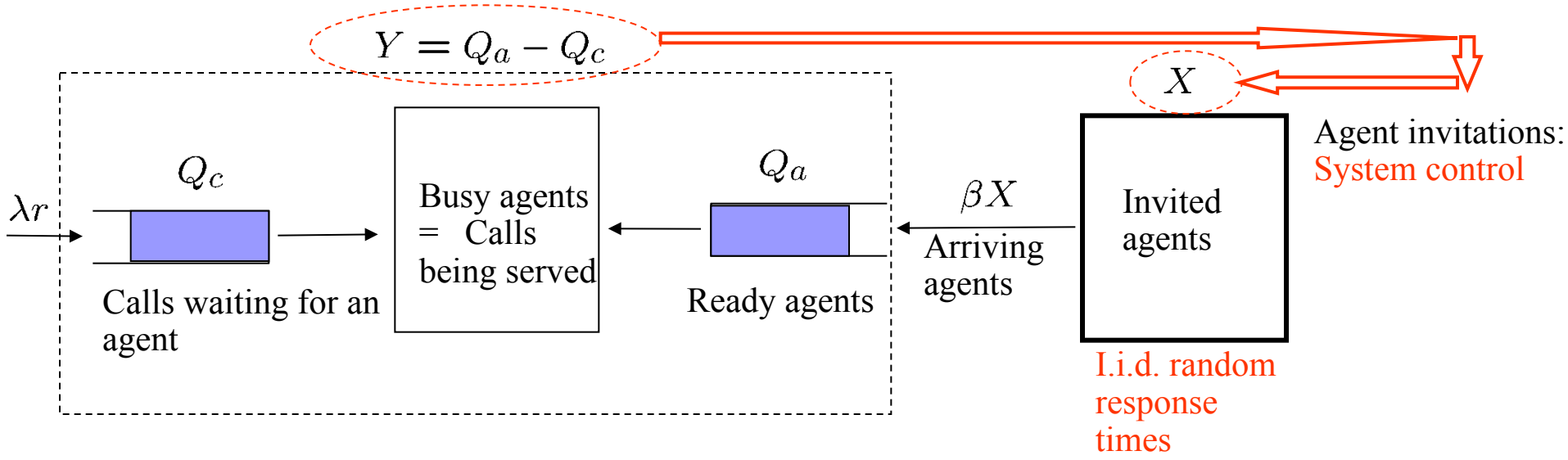
  Strengthening stationary distribution tightness in two steps:

  $o(r)$ scale => $O(r^{1/2+\alpha})$ scale => Diffusion, $O(r^{1/2})$, scale

Generalized Base Stock (GBS) algorithm

- In an inventory system cancelling an order in progress is typically not feasible
- Need to analyze algorithm without this feature, to provide apple-to-apple comparison to existing algorithms, and get limits of the improvement over existing algorithms
- We consider a different algorithm. It is less adaptive (needs to know the model parameters), but as adaptive as the classical Constant Base Stock algorithm

Generalized Base Stock (GBS) algorithm



- GBS algorithm [also in that 2010 patent]:
  - $X_* = \lambda r/\beta$; $f=f(r)$, where $f(r)/\sqrt{r} \to \infty$ and $f(r)/r \to 0$
  - $T = \min\{\max\{X_* - \gamma Y, 0\}, f\}$,
  - When T changes, $X := \max\{X, T\}$.
- Essentially, upon a call arrival $T := T + \gamma$, upon agent arrival $T:=T-\gamma$
- This scheme is does require knowledge of call rate and expected response time
- Same asymptotic regime, with $r \to \infty$. Want not only weak limits, but also limits of expectations

Basic dynamics. Relation to Constant Base Stock

- Essentially, GBS tries to keep

$$X + \gamma Y = X_*$$

- Classical Constant Base Stock (CBS):

$$X + Y = X_*$$

- So, CBS is a special case of GBS, with $\gamma = 1$
- CBS is known to be optimal for constant response times (lead times)
- Under random response times, what is the advantage of GBS over CBS?

Main results for GBS

Theorem 6. Consider a fixed integer $\gamma \geq 1$. Suppose the response time distribution is exponential (with mean $1/\beta$). Then, under GBS policy, the process $(X^r(t), Y^r(t)),\ t \geq 0$, (which is an irreducible continuous-time countable Markov chain) is positive recurrent (stochastically stable) for any sufficiently large r. The following convergence holds:

$$\frac{Y^r(\infty)}{\sqrt{r}} \Rightarrow \mathcal{N}(0, \lambda(\beta\gamma)^{-1}).$$

Moreover, the expectation of $|Y^r(\infty)/\sqrt{r}|$ converges to that of $|\mathcal{N}(0, \lambda(\beta\gamma)^{-1})|$:

$$\mathbb{E}\left|\frac{Y^r(\infty)}{\sqrt{r}}\right| \to 2\sqrt{\frac{\lambda}{2\pi\beta\gamma}}.$$

Main results for GBS

Corollary 7. Suppose we are in the conditions of Theorem 6, except γ may depend on r. Then, the dependence $\gamma = \gamma(r)$ can be chosen in a way such that

$$\mathbb{E}\frac{|Y^r(\infty)|}{\sqrt{r}} \to 0.$$

- Linear holding cost under CBS: $\mathbb{E}|Y^r(\infty)| = O(\sqrt{r})$
- Under GBS (with optimal $\gamma = \gamma(r)$): $\mathbb{E}|Y^r(\infty)| = o(\sqrt{r})$

Proof of Theorem 6: Basic properties of (X,Y)

- Conservation law

$$\mathbb{E}X^r(\infty) = (\lambda/\beta)r$$

- Uniformly bounded gap

$$\mathbb{E}[X^r(\infty) - T^r(\infty)] \leq C, \quad \forall r$$

  - When X-T is large, essentially,
    - » upon a call arrival X-T decreases γ,
    - » upon agent arrival X-T decreases by $\gamma - 1$

  - Prob{An arrival is a call} $\geq 1/2 - \varepsilon$

Proof of Theorem 6: Artificial process

- Artificial process is same as under GBS, except invited agents can be removed at any time. The number of invited agents is always exactly "on target", i.e. it is the deterministic function of the queue length:

$$\tilde{X} = \tilde{T} = \min\{\max\{X_* - \gamma\tilde{Y}, 0\}, f\}$$

- Conservation law

$$\mathbb{E}\tilde{X}^r(\infty) = (\lambda/\beta)r$$

- $\tilde{Y}$ is simply a birth-death process. Stationary distribution is analyzed directly

Lemma 8. For the artificial process:

$$\frac{\tilde{Y}^r(\infty)}{\sqrt{r}} \Rightarrow \mathcal{N}(0, \lambda(\beta\gamma)^{-1}).$$

Moreover:

$$\mathbb{E}\left|\frac{\tilde{Y}^r(\infty)}{\sqrt{r}}\right| \to 2\sqrt{\frac{\lambda}{2\pi\beta\gamma}}.$$

Proof of Theorem 6: Actual Vs Artificial comparison

- Actual process can also be viewed as a birth-death process, but with random birth rates, which are greater than those for the artificial process. This implies:

$$\tilde{Y}^r(\infty) \leq_{st} Y^r(\infty)$$

- To prove the theorem, suffices to show

$$\frac{1}{\sqrt{r}}\left[\mathbb{E}Y^r(\infty) - \mathbb{E}\tilde{Y}^r(\infty)\right] \to 0$$

- To illustrate the proof, let us pretend that

$$T^r = X_* - \gamma Y^r, \qquad \tilde{X}^r = \tilde{T}^r = X_* - \gamma \tilde{Y}^r$$

- Then

$$\mathbb{E}Y^r(\infty) - \mathbb{E}\tilde{Y}^r(\infty) = \frac{1}{\gamma}\mathbb{E}\tilde{X}^r(\infty) - \frac{1}{\gamma}\mathbb{E}T^r(\infty) = \frac{1}{\gamma}\mathbb{E}\left[X^r(\infty) - T^r(\infty)\right]$$

- The actual bound is

$$\mathbb{E}Y^r(\infty) - \mathbb{E}\tilde{Y}^r(\infty) = o(\sqrt{r})$$

GBS Vs CBS

- Linear "holding cost" under CBS:

$$\mathbb{E}|Y^r(\infty)| = O(\sqrt{r})$$

- Under GBS (with optimal $\gamma = \gamma(r)$):

$$\mathbb{E}|Y^r(\infty)| = o(\sqrt{r})$$

  - simulations suggest: $\mathbb{E}|Y^r(\infty)| = O(r^{0.38})$
  - moreover, looks like Cost(GBS)/Cost(OPT) = *O(1)*

- GBS retains substantial advantage over CBS for non-exponential response time distributions
  - Moreover, for some distributions (Pareto), the advantage is even larger

Simulation: EXP response time

$X_* = r/\beta$	GBS Policy		CBS Policy
	γ	cost	cost
2	1.6	1.00	1.08
10	2.2	2.01	2.50
20	2.4	2.66	3.55
100	3.4	4.95	7.97
200	4.8	6.41	11.3
400	5.6	8.22	16.0
600	5.8	9.53	19.5
800	6.8	10.5	22.6
1000	6.8	11.4	25.2
1200	7.8	12.2	27.6
1400	7.8	12.9	29.9
1600	8.6	13.5	31.9
1800	8.6	14.1	33.8
2000	8.6	14.6	35.7

Table 1: A Comparison between GBS and CBS Policies: lead time exponentially distributed with mean $\beta = 1/2$; $h = \theta = 1$

Simulation: EXP response time

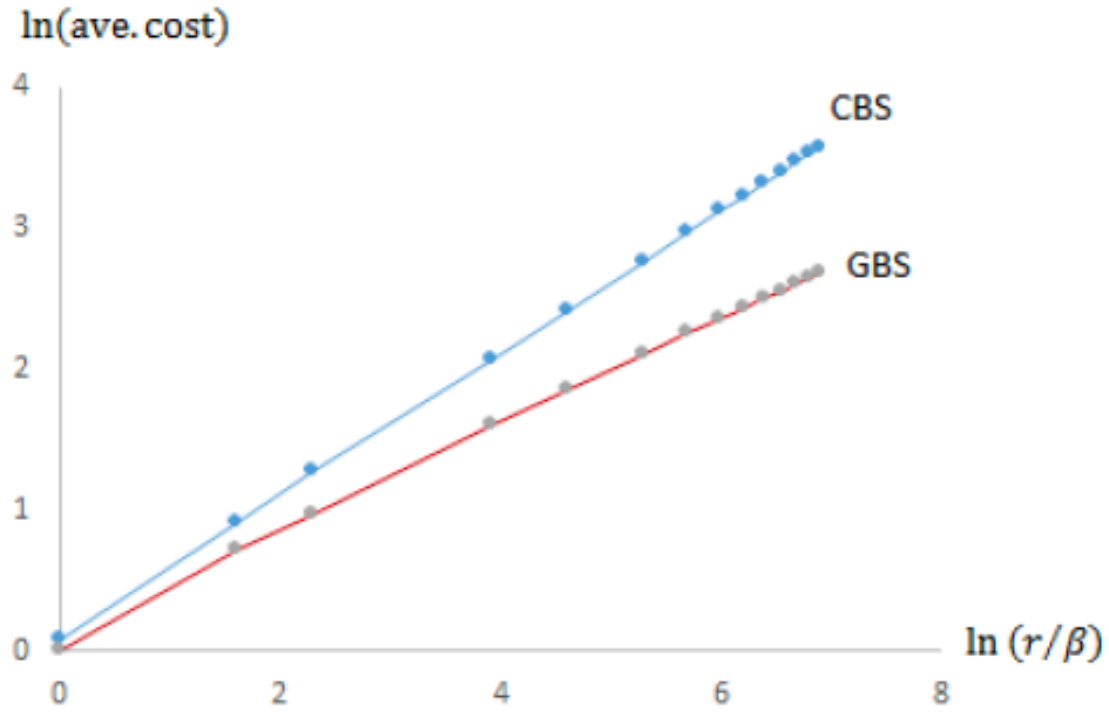


Figure 1: Changes of Costs with Mean Lead Time Demand on Log Scale

Simulation: Response time = d + EXP

ave. lead time demand $X_* = r/\beta$	GBS Policy		CBS Policy
	γ	cost	cost
2	1.4	1.02	1.08
10	1.8	2.12	2.51
20	2.2	2.84	3.56
100	2.8	5.64	7.93
200	3.2	7.53	11.3
400	3.8	10.1	15.9
600	4.4	12.0	19.5
800	4.4	13.6	22.6
1000	4.8	15.0	25.2
1200	5.2	16.2	27.6
1400	5.4	17.4	30.0
1600	5	18.4	31.9
1800	5.6	19.3	33.8
2000	5.2	20.3	35.8

Table 3: Comparison between GBS and CBS Policies: mean lead time=2, deterministic component $d = 0.2$

Simulation: Response time = d + EXP

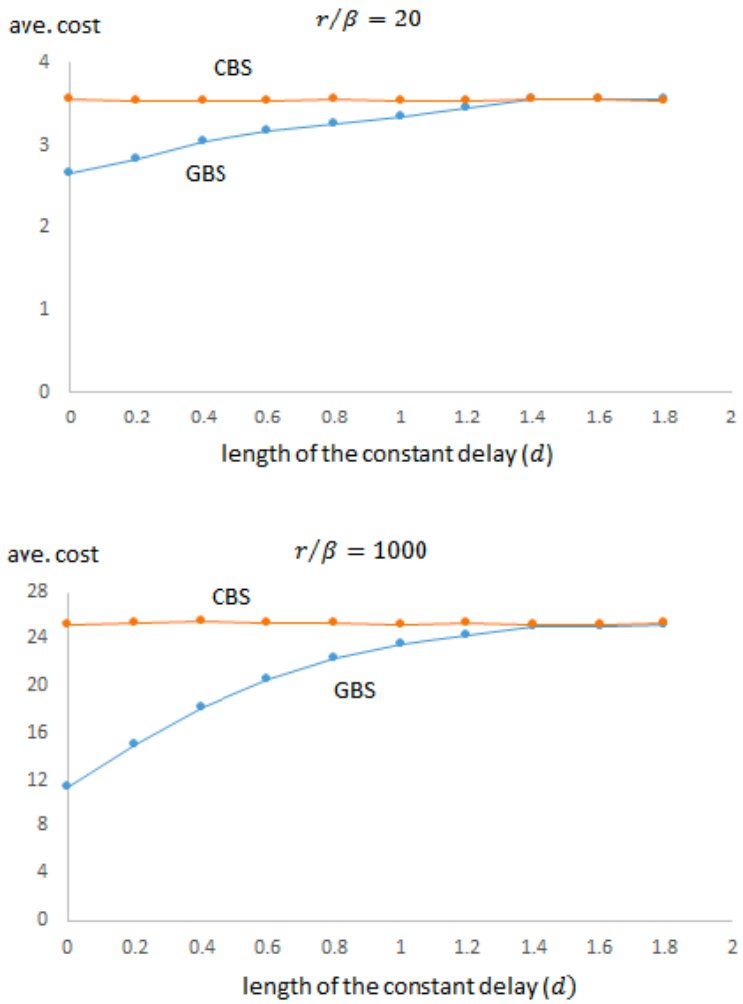


Figure 2: Changes of inventory cost under the GBS and CBS policies with d, the deterministic component of the lead time

Simulation: Pareto response time

$$1 - F(x) = \frac{1}{(1+\tau x)^q}$$

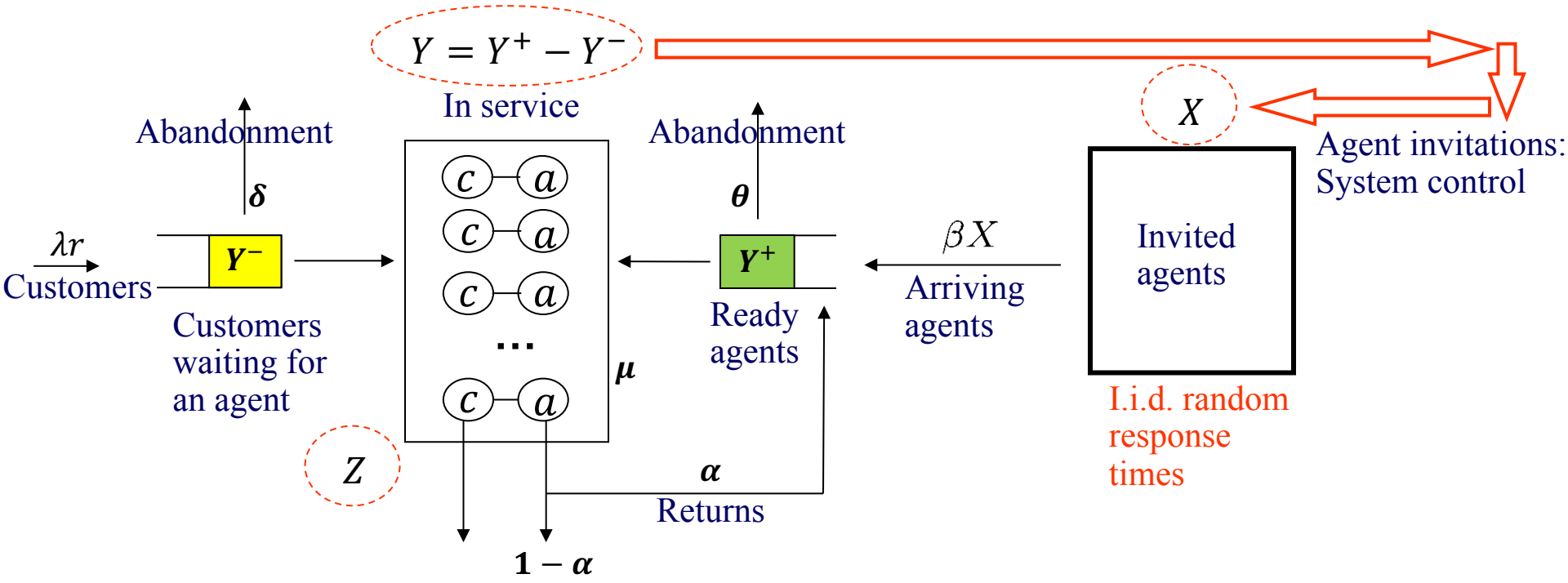
ave. lead time demand	GBS		CBS
$X_* = r/\beta$	γ	cost	cost
2	1.6	0.96	1.08
10	1.8	1.93	2.51
20	2.4	2.47	3.57
100	3.8	4.52	8.00
200	4.6	5.80	11.2
400	5.6	7.42	15.9
600	5.8	8.55	19.6
800	6.4	9.47	22.6
1000	6.8	10.2	25.1
1200	7.6	10.9	27.5
1400	8.0	11.5	29.7
1600	8.0	12.0	31.6
1800	8.2	12.5	33.9
2000	8.4	13.0	35.5

Table 5: Comparison between GBS and CBS Policies when the lead time has Pareto distribution with $q = 3$ and $\tau = 0.25$

Current/future work on the basic model

- Adaptive algorithm
  - Analyze without the simplifying assumption
- GBS
  - Explain the cost growth rate
  - Prove Cost(GBS)/Cost(OPT) = *O(1)*
  - Fundamental cost limits for non-exponential response time distributions

More general model. Adaptive algorithm



- Same Adaptive scheme (including the assumption that agents may be uninvited):
  - X is incremented by $[-\gamma \Delta Y]$ each time Y changes by ΔY *(=+1 or -1)*, $\gamma > 0$ is parameter
  - Independently, X is incremented by $[-\text{sign}(Y)]$ at the instantaneous rate $|\varepsilon Y|$, where $\varepsilon > 0$ is parameter

INFORMALLY: $(d/dt)X = -\gamma(d/dt)Y - \epsilon Y$

$$(d/dt)Y = \beta X - \lambda r + \alpha\mu Z + \delta Y^- - \theta Y^+$$

Process. Fluid scale analysis

- Consider the system process (X^r, Y^r, Z^r) with parameter $r \to \infty$, while $\alpha, \beta, \mu, \delta, \theta, \epsilon, \gamma$ do not depend on r

- Conservation laws:

$$\beta \mathrm{E}X^r + \alpha\mu \mathrm{E}Z^r = \lambda r, \qquad \mathrm{E}Z^r = \lambda r/\mu$$

- Centering:
  - Z^r : $\lambda r/\mu$
  - Y^r : 0
  - X^r : $\lambda r(1-\alpha)/\beta$

- Convenient to consider the process (X^r, Y^r, V^r) where $V^r = (Y^r)^+ + Z^r$

- Fluid-scaled processes with centering

$$(x^r, y^r, v^r) = r^{-1}(X^r - \lambda r(1-\alpha)/\beta, Y^r, V^r - \lambda r/\beta)$$

Fluid limit

- Fluid-scaled processes with centering

$$(x^r, yr, vr) = r^{-1}(X^r - \lambda r(1-\alpha)/\beta, Y^r, V^r - \lambda r/\beta)$$

- Fluid limit

$$(x(\cdot), y(\cdot), v(\cdot)) = \lim_{r\to\infty}(x^r(\cdot), y^r(\cdot), v^r(\cdot))$$

satisfies conditions

$$\begin{cases} x' = \begin{cases} -\gamma y' - \epsilon y, if\ x > -\dfrac{\lambda(1-\alpha)}{\beta} \\ [-\gamma y' - \epsilon y] \vee 0, if\ x = -\dfrac{\lambda(1-\alpha)}{\beta} \end{cases} \\ y' = \beta x + \alpha\mu(v - y^+) + \delta y^- - \theta y^+ \\ v' = \beta x - (1-\alpha)\mu(v - y^+) - \theta y^+ \end{cases} \tag{10}$$

Fluid limit

- Fluid-scaled processes with centering

$$(x^r, yr, vr) = r^{-1}(X^r - \lambda r(1-\alpha)/\beta, Y^r, V^r - \lambda r/\beta)$$

- Fluid limit

$$(x(\cdot), y(\cdot), v(\cdot)) = \lim_{r\to\infty}(x^r(\cdot), y^r(\cdot), v^r(\cdot))$$

satisfies conditions

$$\begin{cases} x' = \begin{cases} -\gamma y' - \epsilon y, if\ x > -\dfrac{\lambda(1-\alpha)}{\beta} \\ [-\gamma y' - \epsilon y] \vee 0, if\ x = -\dfrac{\lambda(1-\alpha)}{\beta} \end{cases} \\ y' = \beta x + \alpha\mu(v - y^+) + \delta y^- - \theta y^+ \\ v' = \beta x - (1-\alpha)\mu(v - y^+) - \theta y^+ \end{cases} \tag{10}$$

Is system (10) stable?

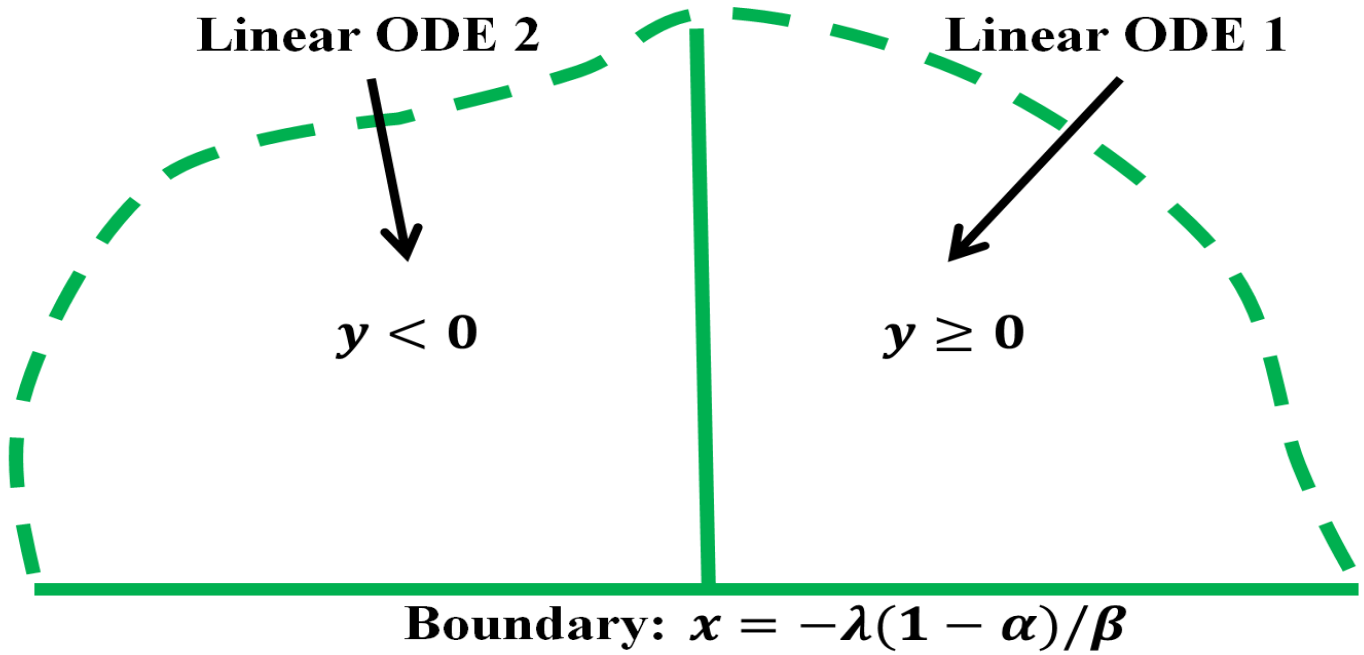
Stability: $(x, y, v)(t) \to (0,0,0)$ as $t \to \infty$

Behavior of fluid limit trajectories

- It's complicated:

  - A "reflecting" boundary

  - Two domains where the trajectories follow different ODEs (but the RHS of the ODE is continuous everywhere)



Global vs. local stability

- Consider a dynamic system in $\mathbb{R}^3$ described by

$$\begin{cases} x' = -\gamma y' - \epsilon y \\ y' = \beta x + \alpha\mu(v - y^+) + \delta y^- - \theta y^+ \\ v' = \beta x - (1-\alpha)\mu(v - y^+) - \theta y^+ \end{cases} \tag{11}$$

- Fluid limit is *globally stable* if every fluid limit trajectory converges to the equilibrium point (0,0,0).

- Fluid limit is *locally stable* if every solution of the dynamic system (11) converges to the equilibrium point (0,0,0).

Main result (sufficient local stability conditions)

Theorem 8: Fluid limit is locally stable if either

$$\gamma > \max\left\{\frac{\alpha\mu-\delta}{\beta}, \sqrt{\frac{(2-\alpha)\epsilon\mu+\alpha\epsilon\delta}{\beta\mu}}\right\} \quad \text{(i)}$$

or

$$\gamma > \max\left\{\frac{\alpha\mu-\delta+\sqrt{(\alpha\mu-\delta)^2+4\alpha\mu^2}}{2\beta}, \sqrt{max\left\{\frac{\alpha\epsilon(\delta-\mu)}{\beta\mu}, 0\right\}}\right\} \quad \text{(ii)}$$

Some existing theory

- Even without the reflecting (regulating) boundary on x, we have ODE with 2 domains ($y \geq 0$ and $y < 0$). This is a switched linear system
- For local stability (stability of the system without boundary), it is sufficient that a Common Quadratic Lyapunov Function (CQLF) exists.
- There is literature on existence of CQLF for switched linear systems:
  - R. Shorten, O. Mason, F. O'Cairbre, P. Curran. A unifying framework for the siso circle criterion and other quadratic stability criteria. International Journal of Control, 77(1): 1-9, 2004.
  - R. Shorten, F. Wirth, O. Mason, K. Wulff, C. King. Stability criteria for switched and hybrid systems. Society for Industrial and Applied Mathematics, 49(4):545-592, 2007.
  - H. Lin, P. J. Antsaklis. Stability and stabilizability of switched linear systems: A survey of recent results. IEEE Transactions on Automatic Control, 54(2):308-322, 2009.

Fluid limit dynamics (when away from boundary)

- Fluid limit dynamics when away from boundary

$$\begin{cases} x' = -\gamma y' - \epsilon y \\ y' = \beta x + \alpha\mu(v - y^+) + \delta y^- - \theta y^+ \\ v' = \beta x - (1-\alpha)\mu(v - y^+) - \theta y^+ \end{cases} \tag{11}$$

- 2 domains:

$y \geq 0$

$$\begin{cases} x' = (-\gamma\beta)x + (\gamma\alpha\mu + \gamma\theta - \epsilon)y + (-\gamma\alpha\mu)v \\ y' = (\beta)x + (-\alpha\mu - \theta)y + (\alpha\mu)v \\ v' = (\beta)x + \big((1-\alpha)\mu - \theta\big)y + (-(1-\alpha)\mu)v \end{cases}$$

$y < 0$

$$\begin{cases} x' = (-\gamma\beta)x + (\gamma\delta - \epsilon)y + (-\gamma\alpha\mu)v \\ y' = (\beta)x + (-\delta)y + (\alpha\mu)v \\ v' = (\beta)x + (-(1-\alpha)\mu)v \end{cases}$$

Fluid limit dynamics (when away from boundary)

- In matrix form: $u(t) = (x(t), y(t), v(t))^T$

$$y \geq 0$$
$$u'(t) = A_1 u(t)$$

$$A_1 = \begin{pmatrix} -\gamma\beta & \gamma\alpha\mu + \gamma\theta - \epsilon & -\gamma\alpha\mu \\ \beta & -\alpha\mu - \theta & \alpha\mu \\ \beta & -(1-\alpha)\mu & -(1-\alpha)\mu \end{pmatrix}$$

$$y < 0$$
$$u'(t) = A_2 u(t)$$

$$A_2 = \begin{pmatrix} -\gamma\beta & \gamma\delta - \epsilon & -\gamma\alpha\mu \\ \beta & -\delta & \alpha\mu \\ \beta & 0 & -(1-\alpha)\mu \end{pmatrix}$$

Existence of CQLF

Necessary and sufficient condition for the existence of CQLF for switched linear systems [Shorten et al, 2007]

Proposition 9: Let A_1 and A_2 be Hurwitz matrices in $\mathbb{R}^{n\times n}$, such that $A_1 - A_2$ has rank one. Then the two systems

$$u'(t) = A_1 u(t) \quad \text{and} \quad u'(t) = A_2 u(t)$$

have a CQLF if and only if $A_1 A_2$ has no negative real eigenvalues.

Theorem 8: Proof outline

- A_1 is always Hurwitz
- A_2 is Hurwitz if $\gamma > \frac{\alpha\mu - \delta}{\beta}$
- $rank(A_1 - A_2) = 1$
- Key part: $A_1 A_2$ has no negative real eigenvalues if either (i) or (ii) holds

$$\gamma > \max\left\{\frac{\alpha\mu - \delta}{\beta}, \sqrt{\frac{(2-\alpha)\epsilon\mu + \alpha\epsilon\delta}{\beta\mu}}\right\} \quad \text{(i)}$$

$$\gamma > \max\left\{\frac{\alpha\mu - \delta + \sqrt{(\alpha\mu - \delta)^2 + 4\alpha\mu^2}}{2\beta}, \sqrt{\max\left\{\frac{\alpha\epsilon(\delta - \mu)}{\beta\mu}, 0\right\}}\right\} \quad \text{(ii)}$$

Theorem 8: Proof outline

Proposition 10 [Shorten et al, 2004]: If ${A_1}^{-1}$ is non-singular, $A_1 A_2$ has no negative real eigenvalues if and only if ${A_1}^{-1} + \tau A_2$ is non-singular for all $\tau \geq 0$.

$\det\left[{A_1}^{-1} + \tau A_2\right] < 0$ if either (i) or (ii) holds

$$\gamma > \max\left\{\frac{\alpha\mu - \delta}{\beta}, \sqrt{\frac{(2-\alpha)\epsilon\mu + \alpha\epsilon\delta}{\beta\mu}}\right\} \quad \text{(i)}$$

$$\gamma > \max\left\{\frac{\alpha\mu - \delta + \sqrt{(\alpha\mu - \delta)^2 + 4\alpha\mu^2}}{2\beta}, \sqrt{\max\left\{\frac{\alpha\epsilon(\delta - \mu)}{\beta\mu}, 0\right\}}\right\} \quad \text{(ii)}$$

Some useful corollaries (sufficient local stability conditions)

Corollary 11: Given all other parameters are fixed, fluid limit is locally stable for all sufficiently large γ

Corollary 12: If $\alpha\mu \leq \delta$, then fluid limit is locally stable for all sufficiently small ϵ

Corollary 13: If $\alpha\mu > \delta$ and $\epsilon \leq \frac{(\alpha\mu-\delta)^2\mu}{(2-\alpha)\mu\beta+\alpha\delta\beta}$, then fluid limit is locally stable under condition $\gamma > \frac{\alpha\mu-\delta}{\beta}$

Some useful corollaries (sufficient local stability conditions)

Corollary 14: If $\mu > \delta$, then fluid limit is locally stable under condition

$$\gamma > \frac{\alpha\mu - \delta + \sqrt{(\alpha\mu - \delta)^2 + 4\alpha\mu^2}}{2\beta}$$ (does not depend on ϵ)

Corollary 15: If $\alpha = 0$, then fluid limit is locally stable for all positive $\beta, \mu, \epsilon, \gamma$, and $\delta \geq 0, \theta \geq 0$

Numerical and simulation results

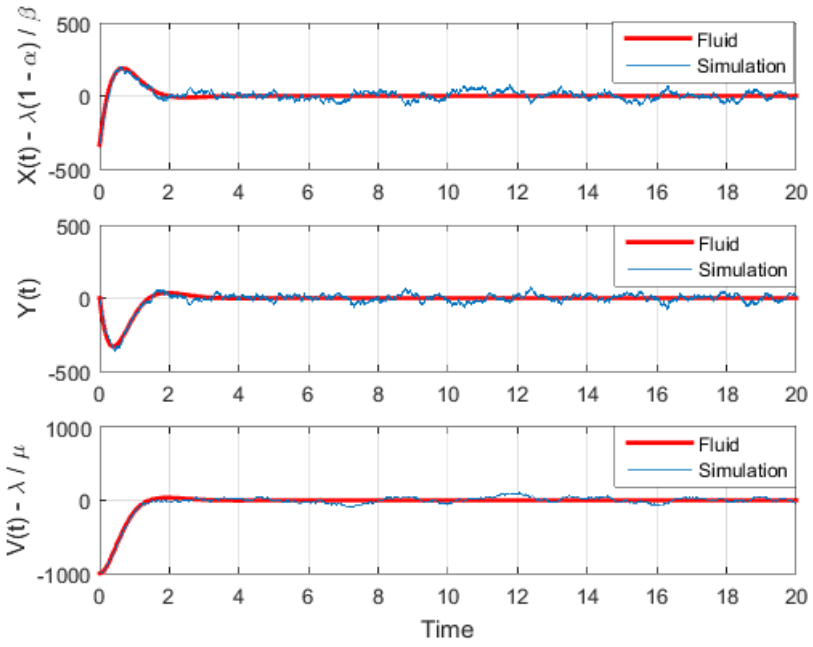
- Simulate the true system, with the boundary
- Vary parameters and initial conditions
- Is there a gap between local and global stability?

Numerical and simulation examples: Example 1

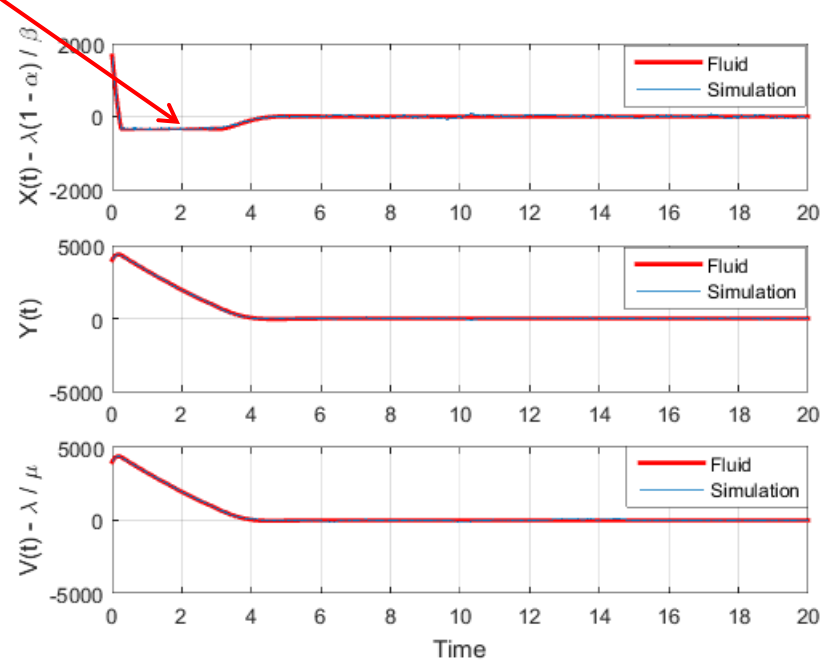
- The sufficient local stability conditions are satisfied

$$\lambda r = 2000\,, \alpha = 0.5\,, \beta = 3\,, \mu = 2\,, \gamma = 1\,, \epsilon = 1.5\,, \delta = 1\,, \theta = 0.1$$

Trajectory hits boundary on x



$(X(0), Y(0), Z(0)) = (0,0,0)$



$(X(0), Y(0), Z(0)) = (2000,4000,1000)$

Different initial conditions

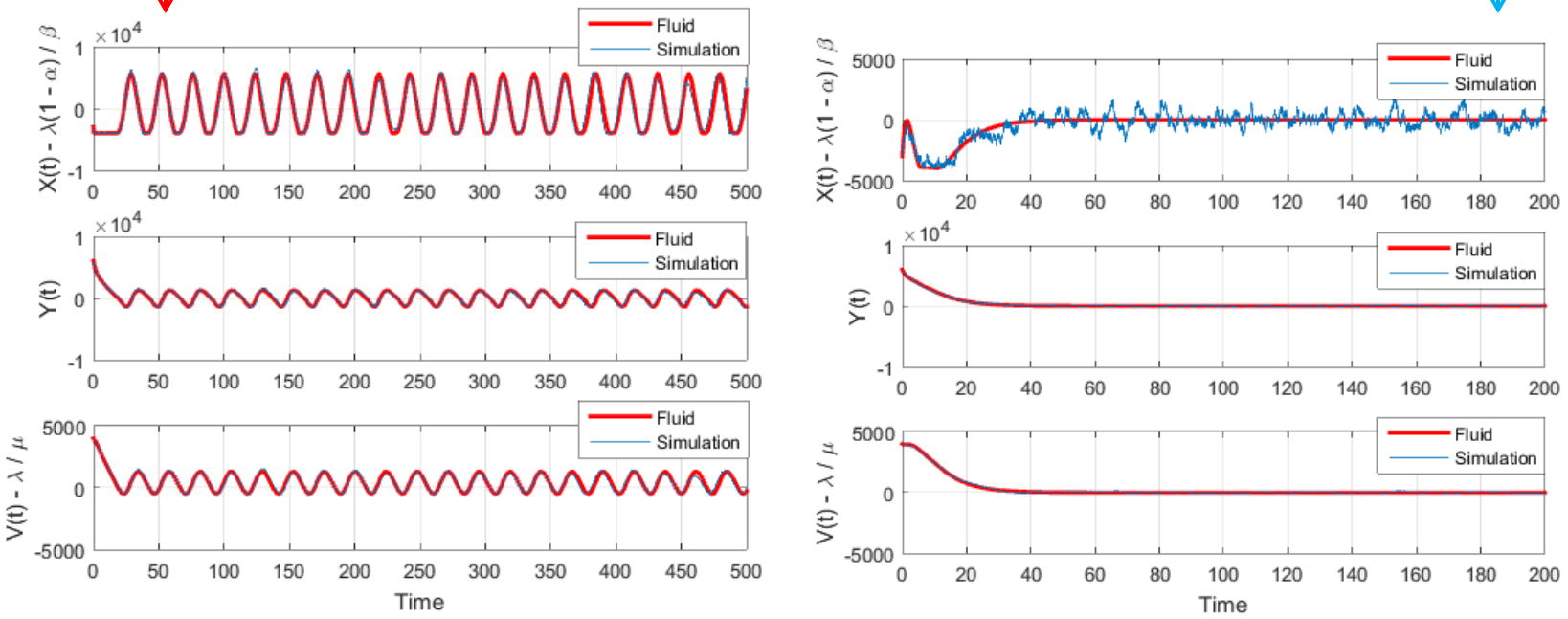
<u>Conjecture 16</u>: Our system is globally stable if it is locally stable.

Numerical and simulation examples: Example 2

- Stabilizing impact of larger γ

$\lambda r = 2000\,, \alpha = 0.9\,, \beta = 0.05\,, \mu = 0.5\,, \boldsymbol{\gamma = 1}\,, \epsilon = 1\,, \delta = 0.01\,, \theta = 0.01$

$\lambda r = 2000\,, \alpha = 0.9\,, \beta = 0.05\,, \mu = 0.5\,, \boldsymbol{\gamma = 10}\,, \epsilon = 1\,, \delta = 0.01\,, \theta = 0.01$



$(X(0), Y(0), Z(0)) = (1000,6000,2000)$

Increasing γ from 1 to 10 makes system locally stable (Corollary 1). Simulation results indicate that it also makes fluid limit globally stable => supports Conjecture 16.

Discussion of / further work on the generalized model(s)

- Adaptive algorithm
  - Relation between local and global stability: seems challenging
  - Boundary causes major difficulties
- Further model extensions
  - multi-class customers, multi-type agents, finite pools of agents, …
- Different algorithms

Papers

G. Pang, A.L. Stolyar. A service system with on-demand agent invitations. Queueing Systems, 82(3), 259-283, 2016.

A.L. Stolyar, Q. Wang, Exploiting random lead times for significant inventory cost savings, 2018, https://arxiv.org/abs/1801.02646.

L. Nguyen, A.L. Stolyar, A queueing system with on-demand servers: local stability of fluid limits, Queueing Systems, 2017, DOI 10.1007/s11134-017-9564-8.

A.L. Stolyar., M.I. Reiman, N. Korolev, V. Mezhibovsky, H. Ristock. Pacing in knowledge worker engagement. US Patent Application 20100266116-A1, 2010.