Bi-Stability Analysis of the Modified Erlang-A Model in the Quality-Driven Regime

Jing Dong

This note is a short extension of the bi-stability analysis in “Service Systems with Slowdowns: Potential Failures and Proposed Solutions” from the Quality-and-Efficiency Driven regime to Quality Driven (QD) regime.

1 Fluid analysis in QD regime

We denote the queue length process by $Q \equiv \{Q(t) : t \geq 0\}$, where $Q(t)$ counts the number of customers in the system (waiting and in service) at time $t$.

**Assumption 1.** $\mu \in C^2$ with $\mu'(x) \leq 0$ and $\mu''(x) \geq 0$ for all $x \geq 0$. $\lim_{x \to \infty} \mu(x) = \mu(\infty) > 0$.

To conduct the heavy-traffic analysis, we consider a sequence of systems indexed by $n$, where both the arrival rate and the number of servers grows with $n$. For the $n$-th system, we denote $Q_n \equiv \{Q_n(t) : t \geq 0\}$ as the queue length process (number of people in the system). We denote the arrival rate as $\lambda_n$ and the number of servers is $n$. The abandonment rate does not scale with $n$ and the service rate function takes the same form when applied to the scaled queue length process, $(Q_n - n)^+/n$. We consider the QD asymptotic regime. Without loss of generality, we assume $\mu(0) = 1$ and $\lambda_n = \rho_n$ for $\rho < 1$.

Let $A \equiv \{A(t) : t \geq 0\}$, $S \equiv \{S(t) : t \geq 0\}$ and $R \equiv \{R(t) : t \geq 0\}$ be three independent Poisson processes, each with unit rate. $A$, $S$ and $R$ generate the arrival, service completion and abandonment processes, respectively. Then, the pathwise construction of $Q_n$ is:

$$Q_n(t) = Q_n(0) + A(\lambda_n t) - S \left( \int_0^t \mu \left( \frac{(Q_n(u) - n)^+}{n} \right) (Q_n(u) \wedge n) \, du \right) - R \left( \int_0^t \theta(Q_n(u) - n)^+ \, du \right),$$

where $(x)^+ = \max(0, x)$ and $(x \wedge y) = \min(x, y)$.

We define the fluid-scaled process

$$\bar{Q}_n(t) = \frac{Q_n(t)}{n}$$

1
Theorem 1. If $\bar{Q}_n(0) \Rightarrow q(0)$ in $\mathbb{R}$, then $\bar{Q}_n \Rightarrow q$ in $\mathcal{D}$ as $n \to \infty$. The limit process $q$ is the unique solution satisfying the following integral equation

$$q(t) = q(0) + \rho t - \int_0^t \mu ((q(u) - 1)^+) (q(u) \wedge 1) \, du - \int_0^t \theta (q(u) - 1)^+ \, du.$$  

Let $f(q) = \rho - \mu (q - 1)^+ (q(u) \wedge 1) - \theta (q - 1)^+$, be the flow rate function of the fluid system at state $q$. Then we can write $q(t)$ as the solution to the following autonomous differential equation with initial value $q(0)$,

$$\dot{q} = f(q)$$

Let $\nu(x) = \mu(x) + \theta x$ for $x \geq 0$ and $\hat{x} = \arg \max_x \{\nu(x)\}$ on $[0, \infty)$. To enforce bi-stability, we impose the following assumptions on the service rate function in addition to Assumption 1.

Assumption 2. $-\mu'(0) > \theta$ and $\rho > \mu(\hat{x}) + \theta \hat{x}$.

Under Assumption 2, $\hat{x}$ is the root of $\nu'(x) = 0$ on $(0, \infty)$. Let $\bar{q} = \hat{x} + 1$. $\bar{q}$ is the point where $f(q)$ attains its maximum on $[1, \infty)$.

Lemma 1. Under Assumption 1 and 2, the fluid model has three equilibrium points, denoted as $\bar{q}_1$, $\bar{q}_2$, and $\bar{q}_3$, with $\bar{q}_1 = \rho, 1 < \bar{q}_2 < \bar{q} < \bar{q}_3 > \bar{q}$. $\bar{q}_1$ and $\bar{q}_2$ are asymptotically stable, while $\bar{q}_2$ is unstable.

2 Analysis of stationary distribution

Let $\pi_n$ denote the stationary distribution of the $n$-th system, then we have the following detailed balance equation for Birth-and-Death process.

$$\lambda_n \pi_n(q - 1) = \left( \mu \left( \frac{(q - n)^+}{n} \right) ((q) \wedge n) + \theta (q - n)^+ \right) \pi_n(q)$$
Then we have when \(\lambda_n \geq \mu ((q - n)^+/n)(q \land n) + \theta(q - n)^+\), \(\pi_n(q) \geq \pi_n(q - 1)\). Under Assumption 2, let \(\tilde{x}\) denote the root of \(\rho - \nu(x) = 0\) on \((0, \tilde{x})\) and \(\hat{x}\) denote the root of \(\rho - \nu(x) = 0\) on \((\hat{x}, \infty)\). Then we have

**Lemma 2.** Under Assumption 2, \(\pi_n(\cdot)\) has two peaks, one at \(\tilde{q}_n, 1 = \lfloor \lambda_n \rfloor\), the other at \(\tilde{q}_n, 2 = \lfloor (\hat{x} + 1)n \rfloor\). The minimum point between the two peaks (valley) is, \(\tilde{q}_n = \lfloor (\tilde{x} + 1)n \rfloor\).

**Proof.** Let \(f_n(q) = \lambda_n - \mu \left(\frac{(q - n)^+/n (q \land n)}{n} + \theta(q - n)^+\right)\). For \(q < \lambda_n\), \(f_n(q) = \lambda_n - \mu(0)q > 0\). For \(\lambda_n < q < n\), \(f_n(q) = \lambda_n - \mu(0)q < 0\). When \(q > n\), let \(x_n = (q - n)/n\).

Then \(f_n(q)/n = \rho - \nu(x_n)\). As \(\rho - \nu(x_n) < 0\) for \(0 \leq x_n < \tilde{x}\), \(\rho - \nu(x) \geq 0\) for \(\tilde{x} \leq x_n \leq \hat{x}\) and \(\rho - \mu(x) < 0\) for \(x_n > \hat{x}\), we have \(f_n(q) < 0\) for \(s_n \leq q < (\tilde{x} + 1)n\), \(f_n(q) \geq 0\) for \((\tilde{x} + 1)n \leq q \leq (\tilde{x} + 1)n\) and \(f_n(q) < 0\) for \(q > (\tilde{x} + 1)n\). \(\square\)

The following theorem characterize the relationship among the values of \(\tilde{q}_{n,1}\), \(\tilde{q}_{n,2}\) and \(\tilde{q}_n\).

**Theorem 2.**

\[
\frac{1}{n} \log \frac{\pi_n(q_{n,1})}{\pi_n(\tilde{q}_n)} = I_1
\]

and

\[
\frac{1}{n} \log \frac{\pi_n(q_{n,2})}{\pi_n(\tilde{q}_n)} = I_2
\]

where

\[
I_1 = (1 - \rho) \log \mu(0) + \int_{\rho}^{1} \log x \, dx + \int_{0}^{\tilde{x}} \log \nu(x) \, dx - (\tilde{x} + 1 - \rho) \log \rho.
\]

and

\[
I_2 = -\int_{\tilde{x}}^{\hat{x}} \log \nu(x) \, dx + (\tilde{x} - \hat{x}) \log \rho.
\]

**Proof.** As

\[
\pi_n(q_{n,1}) = \prod_{q = \tilde{q}_{n,1} + 1}^{\tilde{q}_n} \frac{\mu ((q - n)^+/n)(q \land n) + \theta(q - n)^+}{\lambda_n \pi_n(\tilde{q}_n)}
\]

\[
= \exp \left( \sum_{q = \tilde{q}_{n,1} + 1}^{n} \log \mu(0) \frac{q}{n} + \sum_{q = n+1}^{\tilde{q}_n} \log \nu \left( \frac{q - n}{n} \right) - (\tilde{q}_n - \tilde{q}_{n,1}) \log \frac{\lambda_n}{n} \right) \pi_n(\tilde{q}_n),
\]
then under our scaling parameters \((n = n, \lambda_n = \rho n)\), we have

\[
\frac{1}{n} \log \frac{\pi_n(q_{n,1})}{\pi_n(q_n)} = \frac{n - \rho n}{n} \log \mu(0) + \frac{1}{n} \sum_{q = \rho n + 1}^{n} \log \left( \frac{q}{n} \right) + \frac{1}{n} \sum_{k = 1}^{\bar{x} n} \log \left( \frac{\nu}{n} \left( \frac{k}{n} \right) \right)
\]

\[
\quad - \frac{(\bar{x} + 1)n - \rho n}{n} \log \rho
\]

\[
\rightarrow (1 - \rho) \log \mu(0) + \int_{\rho}^{1} \log(x) dx + \int_{0}^{\bar{x}} \log \nu(x) dx - (\bar{x} + 1 - \rho) \log \rho
\]

as \(n \to \infty\). Likewise, we have

\[
\pi_n(q_{n,2}) = \exp \left( - \sum_{q = \rho n + 1}^{\bar{x} n} \log \left( \frac{1}{n} + 1 \frac{k}{n} \right) \right)
\]

\[
\quad \pi_n(\tilde{q}_n) = \exp \left( - \sum_{\tilde{q}_n + 1}^{\bar{x} n} \log \left( \frac{1}{n} + 1 \frac{k}{n} \right) \right)
\]

Then

\[
\frac{1}{n} \log \frac{\pi_n(q_{n,2})}{\pi_n(q_n)} = - \frac{1}{n} \sum_{\tilde{q}_n + 1}^{\bar{x} n} \log \left( \frac{1}{n} + 1 \frac{k}{n} \right) + \frac{(\bar{x} + 1)n - (\bar{x} + 1)n}{n} \log \rho
\]

\[
\quad \rightarrow - \int_{\bar{x}}^{\tilde{x}} \log \nu(x) dx + (\bar{x} - \tilde{x}) \log \rho.
\]

\[\square\]

**Lemma 3.** For any fixed \(0 < y_1 < y_2 < \infty\)

\[
\lim_{n \to \infty} \log \frac{\pi_n([\rho n + \sqrt{n} y_2])}{\pi_n([\rho n + \sqrt{n} y_1])} = - \int_{y_1}^{y_2} \frac{1}{\rho} y \, dy
\]

and

\[
\lim_{n \to \infty} \log \frac{\pi_n([\rho n - \sqrt{n} y_1])}{\pi_n([\rho n - \sqrt{n} y_2])} = - \int_{-y_2}^{-y_1} \frac{1}{\rho} y \, dy
\]

**Proof.** We prove the first equation only, as the proof of the second equation follows exactly the same line of analysis.

\[
\log \frac{\pi_n([\rho n + \sqrt{n} y_2])}{\pi_n([\rho n + \sqrt{n} y_1])} = - \sum_{k = [\sqrt{n} y_1] + 1}^{[\sqrt{n} y_2]} \log \left( 1 + \frac{k}{\rho n} \right)
\]

\[
= - \sum_{k = [\sqrt{n} y_1] + 1}^{[\sqrt{n} y_2]} \frac{1}{\rho} \frac{k}{\sqrt{n} \sqrt{n}} + O\left( \frac{1}{\sqrt{n}} \right)
\]

\[
\rightarrow - \int_{y_1}^{y_2} \frac{1}{\rho} y \, dy.
\]

\[\square\]
Lemma 4. Recall that \( \bar{x} > 0 \) is the strictly positive root of \( \rho - \mu(x) - \theta x = 0 \). For any fixed \( 0 < y_1 < y_2 < \infty \),

\[
\lim_{n \to \infty} \log \frac{\pi_n([\bar{x} + 1)n + \sqrt{n}y_2])}{\pi_n([\bar{x} + 1)n + \sqrt{n}y_1])} = -\int_{y_1}^{y_2} \frac{\mu'(\bar{x}) + \theta}{\rho} x \, dx
\]

and

\[
\lim_{n \to \infty} \log \frac{\pi_n([\bar{x} + 1)n - \sqrt{n}y_1])}{\pi_n([\bar{x} + 1)n - \sqrt{n}y_2])} = -\int_{-y_1}^{-y_2} \frac{\mu'(\bar{x}) + \theta}{\rho} x \, dx
\]

Proof. We prove the first equation only, as the proof of the second equation follows exactly the same line of analysis.

\[
\log \frac{\pi_n([\bar{x} + 1)n + \sqrt{n}y_2])}{\pi_n([\bar{x} + 1)n + \sqrt{n}y_1])} = -\sum_{k=[\sqrt{n}y_1]+1}^{[\sqrt{n}y_2]} \log \left( 1 + \frac{1}{\rho} \left( \mu \left( \bar{x} + \frac{k}{n} \right) - \mu(\bar{x}) + \theta \frac{k}{n} \right) \right)
\]

\[
= -\sum_{k=[\sqrt{n}y_1]+1}^{[\sqrt{n}y_2]} \frac{\mu'(\bar{x}) + \theta}{\rho} \frac{k}{\sqrt{n}} \frac{1}{\sqrt{n}} + O\left( \frac{1}{\sqrt{n}} \right)
\]

\[
\to -\int_{y_1}^{y_2} \frac{\mu'(\bar{x}) + \theta}{\rho} y \, dy.
\]

\[ \square \]