Technical Proofs

EC.1. Proofs in the Fluid Analysis

Proof of Theorem 1 The function Ψ in ODE (18) is continuous with respect to t, but not locally Lipschitz continuous with respect to (z,q). This is also reflected in the numerical solution (see Figure 3) where there are "sharp" turning points. So we cannot directly apply classical ODE theorem (e.g. Theorem VI in § 10 of Walter (1998)) which requires the function Ψ to be locally Lipschitz continuous with respect to (z,q). The idea is to divide the space $\mathbb{S}$ into several regions, and prove the existence and uniqueness in each region. Once the solution enters another region at time τ, we "restart" the ODE assuming that $(z(\tau),q(\tau))$ is the initial condition.

Note that in the space $\mathbb{S}_+$ the ODE is relatively easy to study, since only q evolves with time t according to (24). Suppose $(z(0),q(0))\in\mathbb{S}_+$, then the solution to the ODE is $z(t)=(0,\ldots,0,N)$ and $q(t)=q(0)+\int_0^t\lambda(s)ds-\gamma_K Nt$ for $t\in[0,\tau_q]$ where τ_q is the point at which $q(\cdot)$ hits 0 for the first time.

Consider next the ODE in the space $\mathbb{S}_0$, which can be divided into $\mathbb{S}_0=\bigcup_{k=0}^K\mathbb{S}_{0,k}$, where

$$\mathbb{S}_{0,k}=\left\{(z,q)\in\mathbb{S}_0 : I(z)=k\right\}. \tag{EC.1}$$

For each $k=0,1,\ldots,K$, if $(z,q)\in\mathbb{S}_{0,k}$, then

$$f_i(z,\lambda)=\begin{cases}0, & i<k-1\\ \gamma_k z_k/\lambda\wedge 1, & i=k-1\\ (1-\gamma_k z_k/\lambda)^+, & i=k\\ 0, & i>k.\end{cases} \tag{EC.2}$$

It is clear that $f(z,\lambda)$ is locally Lipschitz continuous in z on $\mathbb{S}_{0,k}$ for each $k=0,\ldots,K-1$. First, assume that the initial point $(z(0),q(0))\in\mathbb{S}_{0,0}$. This implies that $z_0(0)>0$, so there exists $\delta>0$ such that $z_0(t)>0$ for all $t\in[0,\delta)$. Plugging (EC.2) into (19)–(24) yields that for all $t\in[0,\delta)$

$$\begin{aligned}
z_0'(t)&=-\lambda(t)+\gamma_1 z_1(t),\\
z_1'(t)&=\lambda(t)-\gamma_1 z_1(t)+\gamma_2 z_2(t),\\
z_k'(t)&=-\gamma_k z_k(t)+\gamma_{k+1}z_{k+1}(t),\quad 1<k<K,\\
z_K'(t)&=-\gamma_K z_K(t),\\
q'(t)&=0.
\end{aligned}$$

These ODEs can be written in the form $(z'(t),q'(t))=\Psi_0(t,z(t),q(t))$, then Ψ_0 is locally Lipschitz continuous in (z,q) on $\mathbb{S}_{0,0}$. According to Theorem VI in § 10 of Walter (1998), there exists a unique solution in $\mathbb{S}_{0,0}$. Moreover, the solution can be extended to $[0,\tau_0]$ where $\tau_0=\inf\{t>0 : z_0(t)=0\}$.

Next, assume in general that the initial point $(z(0),q(0)) \in \mathbb{S}_{0,k}$ for $0<k<K$. According to (EC.2), there are two cases depending on the relation between $\gamma_k z_k(\cdot)$ and $\lambda(\cdot)$. If there exits $\delta>0$ such that $\gamma_k z_k(t) \le \lambda(t)$ for all $t\in[0,\delta)$, then plugging (EC.2) into the set of ODEs (19)–(24) yields that for all $t\in[0,\delta)$

$$
\begin{aligned}
z_i'(t) &= 0, \quad 0\le i<k,\\
z_k'(t) &= -\lambda(t)+\gamma_k z_k(t)+\gamma_{k+1}z_{k+1}(t),\\
z_{k+1}'(t) &= \lambda(t)-\gamma_k z_k(t)-\gamma_{k+1}z_{k+1}(t)+\gamma_{k+2}z_{k+2}(t),\\
z_i'(t) &= -\gamma_i z_i(t)+\gamma_{i+1}z_{i+1}(t), \quad k+1<i<K,\\
z_K'(t) &= -\gamma_K z_K(t),\\
q'(t) &= 0,
\end{aligned}
$$

which we write as $(z'(t),q'(t))=\Psi_{k,\le}(t,z(t),q(t))$. Again, the function $\Psi_{k,\le}$ is locally Lipschitz in (z,q) on $\mathbb{S}_{0,k}$, and the existence and uniqueness of the solution to the ODE follow from Theorem VI in § 10 of Walter (1998). Moreover, the solution can be extended to the time $\tau_1=\inf\{t>0: z_k(t)=0 \text{ or } \gamma_k z_k(t)>\lambda(t)\}$. If there does not exist such positive δ, then for any $\epsilon>0$, there exists $t_\epsilon\in(0,\epsilon)$ such that $\gamma_k z_k(t_\epsilon)>\lambda(t_\epsilon)$ (so the inequality holds on a small neighbourhood around t_ϵ). We show in this case, any solution to the ODE transits from $\mathbb{S}_{0,k}$ to $\mathbb{S}_{0,k-1}$ immediately after time 0. If not, there exists a small δ such that on $I(z(t))=k$ for all $t\in[0,\delta]$. Then for any small $\epsilon\in(0,\delta)$,

$$
z_{k-1}(\epsilon)=\int_0^\epsilon -(\lambda(s)\wedge\gamma_k z_k(s))+\gamma_k z_k(s)ds>0,
$$

which contradicts to that $I(z(\epsilon))=k$. So in this case, we study the ODE on the region $\mathbb{S}_{0,k}\cup\mathbb{S}_{0,k-1}$. Plugging (EC.2) into the set of ODEs (19)–(24) yields that for all $t\in[0,\delta)$

$$
\begin{aligned}
z_i'(t) &= 0, \quad 0\le i<k-1,\\
z_{k-1}'(t) &= -\lambda(t)+\gamma_{k-1}z_{k-1}(t)+\gamma_k z_k(t),\\
z_k'(t) &= \lambda(t)-\gamma_{k-1}z_{k-1}(t)-\gamma_k z_k(t)+\gamma_{k+1}z_{k+1}(t),\\
z_i'(t) &= -\gamma_i z_i(t)+\gamma_{i+1}z_{i+1}(t), \quad k+1\le i<K,\\
z_K'(t) &= -\gamma_K z_K(t),\\
q'(t) &= 0,
\end{aligned}
$$

which we write as $(z'(t),q'(t))=\Psi_{k,>}(t,z(t),q(t))$. Note that $\Psi_{k,>}$ is still locally Lipschitz continuous on $\mathbb{S}_{0,k}\cup\mathbb{S}_{0,k-1}$. The existence and uniqueness of the solution follow again from Theorem VI in § 10 of Walter (1998). Moreover, the solution can be extended to the time $\tau_2=\inf\{t>0: z_{k-1}(t)=0 \text{ or } \gamma_{k-1}z_{k-1}(t)>\lambda(t)\}$.

Finally, we discuss the case where the initial point is in $\mathbb{S}_{0,K}$. Note that $\mathbb{S}_{0,K}$ is a single point $(0,\ldots,N,0)$, at which the evolution of the ODE depends on whether the relation between $\lambda(\cdot)$ and $\gamma_K z_K(\cdot)$. If there exists $\delta>0$ such that $\lambda(t)\geq\gamma_K z_K(t)=\gamma_K N$ for all $t\in[0,\delta)$, then by (EC.2) the set of ODEs becomes for all $t\in[0,\delta)$

$$z_i'(t)=0,\quad 0\leq j\leq K,$$
$$q'(t)=\lambda(t)-\gamma_K z_K(t).$$

In this case, the solution to the ODE will stay in $\mathbb{S}_{0,K}\cup\mathbb{S}_+$ till $\tau_3=\inf\{t>0:q(t)=0 \text{ and } \lambda(t)<\gamma_K N\}$. If there does not exist such a positive δ, following the same argument in the previous case, any solution to the ODE transit from $\mathbb{S}_{0,K}$ to $\mathbb{S}_{0,K-1}$ immediately after 0. So we study the ODE on the region $\mathbb{S}_{0,K-1}\cup\mathbb{S}_{0,K}$. Plugging (EC.2) into the set of ODEs yields that for all $t\in[0,\delta)$

$$z_i'(t)=0,\quad 0\leq j\leq K-2,$$
$$z_{K-1}'(t)=-\lambda(t)-\gamma_{K-1}z_{K-1}(t)+\gamma_K z_K(t), \tag{EC.3}$$
$$z_K'(t)=\lambda(t)+\gamma_{K-1}z_{K-1}(t)-\gamma_K z_K(t), \tag{EC.4}$$
$$q'(t)=0.$$

Denote the above ODE by $(z'(t),q'(t))=\Psi_{K,>}(t,z(t),q(t))$. It is clear that $\Psi_{K,>}$ is locally Lipschitz continuous on $\mathbb{S}_{0,K-1}\cup\mathbb{S}_{0,K}$. Similar to the previous analysis, the existence and uniqueness of the solution to the ODE again follow from Theorem VI in § 10 of Walter (1998). Moreover, the solution can be extended to time $\tau_4=\inf\{t>0:z_{K-1}(t)=0 \text{ or } \gamma_{K-1}z_{K-1}(t)>\lambda(t)\}$.

Proof of Proposition 1 If the initial state $(z(0),q(0))\in\mathbb{S}_+$, then by (24) and (55),

$$q'(t)=\lambda-\gamma_K N<0.$$

Thus, q will decrease to zero and the solution to the ODE will enter into $\mathbb{S}_0$. So we can assume without loss of generality that the initial state $(z(0),q(0))\in\mathbb{S}_0$. Suppose at time $s\geq 0$, $(z(s),q(s))\in\mathbb{S}_{0,i}$, which is defined in (EC.1). In other words, $i=I(z(s))$. The idea is to show that $(z(\cdot),q(\cdot))$ will eventually enter $\mathbb{S}_{0,k'}$, and then construct a Lyapunov function to show its convergence to the invariant point $(\tilde{z}(N),0)$. To study the evolution of the solution (z,q) from time s onwards, we divide the discussion into two scenarios.

The first scenario is where $i<k'$. The objective for this scenario is to show that the solution to the ODE will go from $\mathbb{S}_{0,i}$ to $\mathbb{S}_{0,i+1}$ and will never come back again. So the solution eventually enters $\mathbb{S}_{0,k'}$ and will never come back to any $\mathbb{S}_{0,i}$, $i<k'$. If $i=0$, i.e., the initial point $(z(s),q(s))\in\mathbb{S}_{0,0}$, then according to (19),

$$z_0'(t)=-\lambda+\gamma_1 z_1(t)\leq-\lambda+\gamma_1 N\overset{(a)}{<}0,$$

for $t \in [s, s_0]$, where s_0 is the first time after s when z_0 hits 0. In the above, (a) is due to (54) and (56). This means that z_0 will decrease to 0 making the solution to the ODE enter into $\mathbb{S}_{0,1}$, i.e., the smallest index i changes from 0 to 1 at time s_0. If $0 < i < k'$, i.e., $(z(s), q(s)) \in \mathbb{S}_{0,i}$, then according to (19)–(21),

$$
\begin{aligned}
z_j'(t) &= 0, \quad 0 \le j < i-1, \\
z_{i-1}'(t) &= 0 - \frac{\gamma_i z_i(t)}{\lambda}\lambda - \gamma_{i-1} z_{i-1}(t) + \gamma_i z_i(t) \\
&= -\gamma_{i-1} z_{i-1}(t) = 0, \\
z_i'(t) &= \frac{\gamma_i z_i(t)}{\lambda}\lambda - \left(1 - \frac{\gamma_i z_i(t)}{\lambda}\right)\lambda - \gamma_i z_i(t) + \gamma_{i+1} z_{i+1}(t) \\
&= -\lambda + \gamma_i z_i(t) + \gamma_{i+1} z_{i+1}(t) \\
&\overset{(b)}{<} -\lambda + \gamma_{i+1} N \overset{(c)}{<} 0,
\end{aligned}
$$

for $t \in [s, s_i]$, where s_i is the first time after s when z_i hits 0. In the above, (b) follows from (54) and (c) follows from (56). So z_i will decrease to 0 and making the solution to the ODE enter $\mathbb{S}_{0,i+1}$. Continuing the argument, the solution to the ODE enters into $\mathbb{S}_{0,k'}$ after a finite time.

For the rest of this proof, we devote to the second scenario where $i \ge k'$. The dynamics of the ODE are much more complicated in this situation. We also need to take the control threshold K into account. To better understand the evolution of the ODE in this scenario, let's start from the easy case where $K = k' + 1$. Note that by (56), we always have

$$\gamma_{K-1} z_{K-1}(t) \le \lambda < \gamma_K N. \tag{EC.5}$$

If $i = K$ (which equals $k' + 1$), i.e., $(z(s), q(s)) \in \mathbb{S}_{0,K}$, then (19)–(20) yield

$$z_j'(t) = 0, \quad j < K-2, \tag{EC.6}$$

$$z_{K-1}'(t) = -(\lambda \wedge \gamma_K N) + \gamma_K N. \tag{EC.7}$$

By (EC.5), $z_{K-1}'(t) = \gamma_K N - \lambda > 0$. This implies that the solution to the ODE will immediately enter $\mathbb{S}_{0,K-1}$. So without loss of generality, we can assume that $i = K-1$, i.e., $(z(s), q(s)) \in \mathbb{S}_{0,K-1}$. According to (19)–(22),

$$z_j'(t) = 0, \quad j < K-2, \tag{EC.8}$$

$$z_{K-2}'(t) = -(\lambda \wedge \gamma_{K-1} z_{K-1}(t)) + \gamma_{K-1} z_{K-1}(t), \tag{EC.9}$$

$$z_{K-1}'(t) = \lambda \wedge \gamma_{K-1} z_{K-1}(t) - (\lambda - \gamma_{K-1} z_{K-1}(t))^+ - \gamma_{K-1} z_{K-1}(t) + \gamma_K z_K(t), \tag{EC.10}$$

$$z_K'(t) = (\lambda - \gamma_K z_{K-1}(t))^+ - \gamma_K z_K(t), \tag{EC.11}$$

for all $t \in [s,\infty)$. This is valid for all $t \geq s$ is because $z'_{K-2}(t) = 0$ due to (EC.5), implying that the solution to the ODE will never enter $\mathbb{S}_{0,K-2}$. The inequality (EC.5) also implies that

$$\begin{aligned} z'_{K-2}(t) &= 0, \\ z'_{K-1}(t) &= -\lambda + \gamma_{K-1} z_{K-1}(t) + \gamma_K z_K(t), \\ z'_K(t) &= \lambda - \gamma_{K-1} z_{K-1}(t) - \gamma_K z_K(t). \end{aligned}$$

For this situation, we define the Lyapunov function

$$\mathcal{L}(t) = \frac{1}{2} \sum_{j=K-1}^{K} (z_j(t) - \tilde{z}_j)^2.$$

Note that $z_{K-1}(t) + z_K(t) = N$. Plugging (57) into the above yields

$$\begin{aligned} \frac{d}{dt}\mathcal{L}(t) &= [-\lambda + \gamma_{K-1} z_{K-1}(t) + \gamma_K z_K(t)]\Big\{[z_{K-1}(t) - \tilde{z}_{K-1}] - [z_K(t) - \tilde{z}_K]\Big\} \\ &= [-\lambda + \gamma_{K-1} N + (\gamma_K - \gamma_{K-1}) z_K(t)]\Big\{(N - 2z_K(t)) + \frac{2\lambda - \gamma_K N - \gamma_{K-1} N}{\gamma_K - \gamma_{K-1}}\Big\} \\ &= \frac{-2}{\gamma_K - \gamma_{K-1}}[-\lambda + \gamma_{K-1} N + (\gamma_K - \gamma_{K-1}) z_K(t)]^2 \leq 0. \end{aligned}$$

The above derivative equals 0 only when $z(t) = \tilde{z}$. It is clear that $\mathcal{L}(t) \geq 0$ with equality holds only when $z(t) = \tilde{z}$. Note that the ODE in this proof is autonomous since $\lambda(s) \equiv \lambda$. We can view $\tilde{z}(\cdot) \equiv \tilde{z}$ as the "zero" solution to the ODE, by Theorem II in § 30 of Walter (1998), $z(t) \to \tilde{z}$ as $t \to 0$. Similar application of Lyapunov functions is also used by Perry and Whitt (2011c) to study the global asymptotic stability of the solution to the ODE for a different service model. For the rest of this proof, we will use the same argument repeatedly. For simplicity, we omit repeating the above logic, and only focus on constructing a Lyapunov function $\mathcal{L}(t)$ such that $\mathcal{L}(t) \geq 0$ with equality holds only when $z(t) = \tilde{z}$ and the derivative is strictly negative when $\mathcal{L}(t) > 0$. Consider now the general and also more difficult case where $K > k' + 1$. In this case, the smallest index i has the freedom to range between k' and K, which are more than two levels apart. Unlike the first scenario, the smallest index i may not be monotonic. It is possible that i goes up and down, calling for a different approach. Note that by (56), the following always holds

$$\gamma_{k'} z_{k'}(t) \leq \lambda < \gamma_{k'+1} N. \tag{EC.12}$$

If $i = K$, i.e., $(z(s), q(s)) \in \mathbb{S}_{0,K}$, then the ODE takes the same form as (EC.6) and (EC.7). It is clear that (EC.12) implies $z'_{K-1}(t) > 0$. Thus, the solution to the ODE will immediately enter $\mathbb{S}_{0,K-1}$. If $i = K-1$, i.e., $(z(s), q(s)) \in \mathbb{S}_{0,K-1}$, then the ODE takes the same form as (EC.8)–(EC.11). Then (EC.12) implies that

$$\begin{aligned} z'_{K-2}(t) &= 0, \\ z'_{K-1}(t) &= -\lambda + \gamma_{K-1} z_{K-1}(t) + \gamma_K z_K(t) \geq -\lambda + \gamma_{K-1} N > 0, \\ z'_K(t) &= \lambda - \gamma_{K-1} z_{K-1}(t) - \gamma_K z_K(t) < 0. \end{aligned}$$

So on $\mathbb{S}_{0,K-1}$, z_{K-1} increases and z_K decreases until some point s_1 where $\gamma_{K-1}z_{K-1}(s_1) > \lambda$. At that time, by (EC.9), we have $z'_{K-2}(s_1) > 0$. So the solution to the ODE will transit from $\mathbb{S}_{0,K-1}$ to $\mathbb{S}_{0,K-2}$. Now, assume that $k' \le i \le K-2$. On $\mathbb{S}_{0,i}$, according to (19)–(24),

$$z'_j(t) = 0, \quad j < i-1,$$

$$z'_{i-1}(t) = -(\lambda \wedge \gamma_i z_i(t)) + \gamma_i z_i(t),$$

$$z'_i(t) = \lambda \wedge \gamma_i z_i(t) - (\lambda - \gamma_i z_i(t))^+ - \gamma_i z_i(t) + \gamma_{i+1} z_{i+1}(t), \tag{EC.13}$$

$$z'_{i+1}(t) = (\lambda - \gamma_i z_i(t))^+ - \gamma_{i+1} z_{i+1}(t) + \gamma_{i+2} z_{i+2}(t), \tag{EC.14}$$

$$z'_j(t) = -\gamma_j z_j(t) + \gamma_{j+1} z_{k+1}(t), \quad i+1 < j < K, \tag{EC.15}$$

$$z'_K(t) = -\gamma_K z_K(t). \tag{EC.16}$$

Define the Lyapunov function

$$\mathcal{L}_{K-1}(t) = (z_{K-1}(t) + z_K(t)) + z_K(t).$$

We then have

$$\frac{d}{dt}\mathcal{L}_{K-1}(t) = -\gamma_{K-1} z_{K-1}(t) - \gamma_K z_K(t) \le 0,$$

and $\frac{d}{dt}\mathcal{L}_{K-1}(t) = 0$ if and only if $(z_{K-1}(t), z_K(t)) = (0,0)$. So for any

$$0 < \delta \le \frac{\gamma_{k'+1}N - \lambda}{1 + \gamma_{k'+1}\sum_{l=1}^{K}\gamma_l^{-1}},$$

there exists a time s_{K-1} such that $z_K(t) < \delta/\gamma_K$ and $z_{K-1}(t) < \delta/\gamma_{K-1}$ for all $t \ge s_{K-1}$. Now, we apply an induction argument for $j = K-2, \ldots, k'+2$. Suppose there exists s_{j+1} such that $z_{j+1}(t) < \delta/\gamma_{j+1}$ for all $t \ge s_{j+1}$. We can then show that there exists $s_j > s_{j+1}$ such that $z_j(t) < \delta/\gamma_j$ for all $t \ge s_j$. Construct the Lyapunov function

$$\mathcal{L}_j(t) = \frac{1}{2}\Big(\sum_{l=j}^{K} z_l(t)\Big)^2 + \sum_{l=j+1}^{K} z_l(t) + \ldots + \sum_{l=K}^{K} z_l(t).$$

Then

$$\frac{d}{dt}\mathcal{L}_j(t) = \Big(\sum_{l=j}^{K} z_l(t)\Big)\Big(\sum_{l=j}^{K} z'_l(t)\Big) + \sum_{l=j+1}^{K} z'_l(t) + \ldots + \sum_{l=K}^{K} z'_l(t).$$

To obtain the desired conclusion, we further analyze the derivative according to the following three subcases, depending on the relation between level j and smallest index i. It is possible that the value of i may change under each condition, but this change will not cause any trouble when we study the evolution of z_j. Subcase (1): If $j > i+1$, then the evolution of z_l, $l = j, \ldots, K-2$, follows (EC.15), thus

$$\frac{d}{dt}\mathcal{L}_j(t) = -\Big(\sum_{l=j}^{K} z_l(t)\Big)\gamma_j z_j(t) - \gamma_{j+1} z_{j+1}(t) - \ldots - \gamma_K z_K(t) \le 0.$$

The derivative equals 0 only when $z_l(t) = 0$ for all $l = j, \ldots, K$. Subcase (2): If $j = i+1$, then the evolution of z_l, $l = j+1, \ldots, K-1$, follows (EC.15), but that of z_j follows (EC.14). Thus

$$\begin{aligned} z_j'(t) &= (\lambda - \gamma_{j-1} z_{j-1}(t))^+ - \gamma_j z_j(t) + \gamma_{j+1} z_{j+1}(t), \\ &= \begin{cases} -\gamma_j z_j(t) + \gamma_{j+1} z_{j+1}(t) & \text{if } \lambda < \gamma_{j-1} z_{j-1}(t), \\ \lambda - \gamma_{j-1} z_{j-1}(t) - \gamma_j z_j(t) + \gamma_{j+1} z_{j+1}(t) & \text{if } \lambda \ge \gamma_{j-1} z_{j-1}(t). \end{cases} \end{aligned}$$

When $\lambda < \gamma_{j-1} z_{j-1}(t)$, the analysis reduces to Subcase (1). When $\lambda \ge \gamma_{j-1} z_{j-1}(t)$,

$$\frac{d}{dt}\mathcal{L}_j(t) = \Big(\sum_{l=j}^{K} z_l(t)\Big)[\lambda - \gamma_{j-1} z_{j-1}(t) - \gamma_j z_j(t)] - \gamma_{j+1} z_{j+1}(t) - \ldots - \gamma_K z_K(t).$$

Since $\gamma_{k'} \le \gamma_{j-1} < \gamma_j$ and the smallest non-zero level $i = j-1$, then

$$\begin{aligned} &\lambda - \gamma_{j-1} z_{j-1}(t) - \gamma_j z_j(t) \\ \le\, &\lambda - \gamma_{j-1}(z_{j-1}(t) + z_j(t)) \\ \le\, &\lambda - \gamma_{k'}(N - \sum_{l=j+1}^{K} z_l(t)). \end{aligned}$$

By the induction assumption that $z_l(t) \le \delta/\gamma_l$, $l = j+1, \ldots, K$,

$$\begin{aligned} &\lambda - \gamma_{k'}(N - \sum_{l=j+1}^{K} z_l(t)) \\ \le\, &\lambda - \gamma_{k'}(N - \delta \sum_{l=j+1}^{K} \gamma_l^{-1}) < 0, \end{aligned}$$

where the last inequality is due to the choice of δ. This implies that $\frac{d}{dt}\mathcal{L}_j(t) \le 0$ with the equality holding only when $z_l(t) = 0$ for all $l = j, \ldots, K$. Subcase (3): If $j = i$, then according to (EC.13)–(EC.16)

$$\begin{aligned} \frac{d}{dt}\mathcal{L}_j(t) = &\Big(\sum_{l=j}^{K-1} z_l(t)\Big)[\lambda \wedge \gamma_j z_j(t) - \gamma_j z_j(t)] \\ &+ (\lambda - \gamma_j z_j(t))^+ - \gamma_{j+1} z_{j+1}(t) \\ &- \gamma_{j+2} z_{j+2}(t) - \ldots - \gamma_K z_K(t). \end{aligned} \tag{EC.17}$$

Since the smallest non-zero level $i = j$ and $z_l(t) < \delta/\gamma_l$ for $l > j$ by the induction, we have

$$z_j(t) \ge N - \sum_{l=j+1}^{K} z_l(t) \ge N - \delta \sum_{l=j+1}^{K} \gamma_l^{-1}.$$

By the fact that $\gamma_{k'} < \gamma_{k'+1} < \gamma_j$ and the choice of δ, we have $\lambda < \gamma_j z_j(t)$. Plugging this inequality into (EC.17) reveals that $\frac{d}{dt}\mathcal{L}_j(t) \le 0$ with the equality holding only when $z_l(t) = 0$ for all $l = j, \ldots, K$. This property of the Lyapunov function $\mathcal{L}_j$ implies that there exists an s_j such that $z_j(t) < \delta/\gamma_j$ for all $t > s_j$. This completes the induction argument.

Note that the induction only goes down to $j = k'+2$. What we have now is that $z_l(t) < \delta/\gamma_l$, $l = K, K-1, \ldots, k'+2$, for all $t \ge s_{k'+2}$. This implies that $(z(t), q(t)) \in \mathbb{S}_{0,i}$ where either $i = k'$ or $i = k'+1$. If $i = k'+1$, then by (EC.12) the solution to the ODE immediately makes the transition from $\mathbb{S}_{0,k'+1}$ to $\mathbb{S}_{0,k'}$. Then, we only need to focus on the sub-region $\mathbb{S}_{0,k'}$. We now construct a final Lyapunov function

$$\mathcal{L}_{k'}(t) = \frac{1}{2}\big(z_{k'+1}(t) - \tilde{z}_{k'+1}\big)^2 + \sum_{l=k'+2}^{K} z_l(t) + \sum_{l=k'+3}^{K} z_l(t) + \ldots + \sum_{l=K}^{K} z_l(t).$$

It is clear that $\mathcal{L}_{k'}(t) \ge 0$ and that the derivative of the Lyapunov function is

$$\frac{d}{dt}\mathcal{L}_{k'}(t) = \big(z_{k'+1}(t) - \tilde{z}_{k'+1}\big) z'_{k'+1}(t) - \gamma_{k'+2} z_{k'+2}(t) - \gamma_{k'+3} z_{k'+3}(t) - \ldots - \gamma_K z_K(t). \tag{EC.18}$$

According to (EC.12) and (EC.13)–(EC.16),

$$z'_{k'+1}(t) = \lambda - \gamma_{k'} z_{k'}(t) - \gamma_{k'+1} z_{k'+1}(t) + \gamma_{k'+2} z_{k'+2}(t).$$

Applying the definition of $\tilde{z}$ in (57) and some algebra yields

$$\begin{aligned} z'_{k'+1}(t) &= -\gamma_{k'}\left(z_{k'}(t) - \tilde{z}_{k'}\right) - \gamma_{k'+1}\left(z_{k'+1}(t) - \tilde{z}_{k'+1}\right) + \gamma_{k'+2} z_{k'+2}(t) \\ &= \left(\gamma_{k'} - \gamma_{k'+1}\right)\left(z_{k'+1}(t) - \tilde{z}_{k'+1}\right) + \gamma_{k'} \sum_{l=k'+2}^{K} z_l(t) + \gamma_{k'+2} z_{k'+2}(t) \\ &= \left(\gamma_{k'} - \gamma_{k'+1}\right)\left(z_{k'+1}(t) - \tilde{z}_{k'+1}\right) + \gamma_{k'} \sum_{l=k'+3}^{K} z_l(t) + (\gamma_{k'} + \gamma_{k'+2}) z_{k'+2}(t). \end{aligned} \tag{EC.19}$$

When $z_{k'+1}(t) < \tilde{z}_{k'+1}$, it is clear by (EC.19) that $z'_{k'+1}(t) > 0$. So $\frac{d}{dt}\mathcal{L}_{k'}(t) < 0$. When $z_{k'+1}(t) > \tilde{z}_{k'+1}$, recall that

$$z_l(t) \le \delta/\gamma_l \text{ for all } l \ge k'+2. \tag{EC.20}$$

So δ can be chosen small enough such that

$$\kappa := \frac{\delta + \gamma_{k'} \sum_{l=k'+2}^{K} \delta/\gamma_l}{\gamma_{k'+1} - \gamma_{k'}} \le \frac{\gamma_{k'+2}}{\gamma_{k'} + \gamma_{k'+2}}.$$

If $z_{k'+1}(t) - \tilde{z}_{k'+1} > \kappa$, then according to (EC.19) and (EC.20), $z'_{k'+1}(t) < 0$. Thus $\frac{d}{dt}\mathcal{L}_{k'}(t) < 0$. If $0 < z_{k'+1}(t) - \tilde{z}_{k'+1} \le \kappa$, then plugging (EC.19) into (EC.18) yields

$$\begin{aligned} \frac{d}{dt}\mathcal{L}_{k'}(t) = &-(\gamma_{k'+1} - \gamma_{k'})\big(z_{k'+1}(t) - \tilde{z}_{k'+1}\big)^2 \\ &- \big[\gamma_{k'+2} - (\gamma_{k'} + \gamma_{k'+2})(z_{k'+1}(t) - \tilde{z}_{k'+1})\big] z_{k'+2}(t) \\ &- \sum_{l=k'+3}^{K} \big[\gamma_l - \gamma_{k'}(z_{k'+1}(t) - \tilde{z}_{k'+1})\big] z_l(t). \end{aligned}$$

The choice of κ implies that

$$\gamma_{k'+2} - (\gamma_{k'} + \gamma_{k'+2})(z_{k'+1}(t) - \tilde{z}_{k'+1}) \geq 0,$$
$$\gamma_l - \gamma_{k'}(z_{k'+1}(t) - \tilde{z}_{k'+1}) \geq 0.$$

Thus $\frac{d}{dt}\mathcal{L}_{k'}(t) < 0$. Moreover, $\frac{d}{dt}\mathcal{L}_{k'}(t) = 0$ only when $z_{k'+1}(t) = \tilde{z}_{k'+1}$ and $z_l(t) = 0$ for all $l \geq k'+2$. These properties of the Lyapunov function imply that $z(t) \to \tilde{z}$ as $t \to \infty$.

EC.2. Proofs in the Stochastic Analysis

Proof of Corollary 2 By (10), we need only to show convergence for the expectation of the holding cost,

$$\mathbb{E}\left[\frac{1}{T}\int_0^T h\left(\bar{Z}^n(s), \bar{Q}^n(s)\right) ds\right] \to \frac{1}{T}\int_0^T h(z(s), q(s))ds \quad \text{as } n \to \infty. \tag{EC.21}$$

By Theorem 2 and the continuous mapping theorem,

$$\frac{1}{T}\int_0^T h\left(\bar{Z}^n(s), \bar{Q}^n(s)\right) ds \Rightarrow \frac{1}{T}\int_0^T h(z(s), q(s))ds.$$

Note that $\bar{Z}^n(s) \leq \bar{N}^n$ and $\bar{Q}^n(s) \leq \bar{Q}^n(0) + \bar{\Lambda}^n(s)$ for any $s \geq 0$. By monotonicity of h and (28),

$$\begin{aligned}\frac{1}{T}\int_0^T h\left(\bar{Z}^n(s), \bar{Q}^n(s)\right) ds &\leq h\left(\bar{N}^n e, \bar{Q}^n(0) + \bar{\Lambda}^n(T)\right)\\ &\leq h\left(2Ne, \bar{Q}^n(0) + \bar{\Lambda}^n(T)\right)\end{aligned}$$

for all large enough n. Pick a positive $\epsilon < 1$, for the constant C specified in Assumption 2,

$$\begin{aligned}&\mathbb{E}\left[h^{1+\epsilon}\left(2Ne, \bar{Q}^n(0) + \bar{\Lambda}^n(T)\right)\right]\\ &\leq \mathbb{E}\left[h^{1+\epsilon}\left(2Ne, \bar{Q}^n(0) + \bar{\Lambda}^n(T)\right) | \bar{Q}^n(0) + \bar{\Lambda}^n(T) \leq C\right] \mathbb{P}(\bar{Q}^n(0) + \bar{\Lambda}^n(t) \leq C)\\ &\quad + \mathbb{E}\left[h^{1+\epsilon}\left(2Ne, \bar{Q}^n(0) + \bar{\Lambda}^n(T)\right) | \bar{Q}^n(0) + \bar{\Lambda}^n(t) > C\right] \mathbb{P}(\bar{Q}^n(0) + \bar{\Lambda}^n(t) > C)\\ &\leq h^{1+\epsilon}(2Ne, C) + \mathbb{E}[A^{1+\epsilon}\exp(\alpha(1+\epsilon)/2\bar{Q}^n(0))\exp(\alpha(1+\epsilon)/2\bar{\Lambda}^n(T))]\\ &\leq h^{1+\epsilon}(2Ne, C) + \mathbb{E}[A^{1+\epsilon}\exp(\alpha\bar{Q}^n(0))]\exp\left(\left(\exp\left(\frac{\alpha}{n}\right) - 1\right) n\int_0^T \bar{\lambda}^n(s)ds\right).\end{aligned}$$

By Assumption 1, $\exp\left(\left(\exp\left(\frac{\alpha}{n}\right) - 1\right) n\int_0^T \bar{\lambda}^n(s)ds\right) < \infty$. This combined with (29) yields that $\frac{1}{T}\int_0^T h\left(\bar{Z}^n(s), \bar{Q}^n(s)\right)$ is uniformly integrable. According to Theorem 5.5.2 in Durrett (2010), we have (EC.21).

Proof of Lemma 1 It suffices to prove the relative compactness of $\{(\bar{Z}^n, \bar{Q}^n)\}$ and $\{\nu^n\}$ separately. We have already shown the compactness of the space $\mathbb{M}$ because of the compactness of $[0,T] \times \bar{\mathbb{Z}}^{K+1}$. So by Prohorov's theorem (cf. Theorem 11.6.1 in Whitt (2002)), $\{\nu^n\}$ is relatively compact in $\mathbb{M}$. It remains to verify that the sequence $\{(\bar{Z}^n, \bar{Q}^n)\}_{n\in\mathbb{N}}$ is relatively compact.

By the assumption of initial condition (25), for any $\epsilon > 0$ there exists a $C > 0$ and $n_0 > 0$ such that

$$\mathbb{P}\left(|\bar{Q}^n(0)| > C\right) < \epsilon, \text{ for all } n > n_0. \tag{EC.22}$$

Note that $\bar{Z}_k^n(\cdot) \leq N^n/n \leq 2N$ for sufficiently large n. For any $\delta > 0$, define the modulus of continuity $\mathbf{w}_T(y(\cdot),\delta)$ for a function x on $[0,T]$ as

$$\mathbf{w}_T(x(\cdot),\delta) = \sup_{|t-s|\leq\delta,\ s,t\in[0,T]} |x(t) - y(s)|.$$

Now, we study the modulus of continuity of $\bar{Q}^n$ and $\bar{Z}_k^n$, $k = 0,\ldots,K$. According to (38), we have

$$|\bar{Q}^n(t) - \bar{Q}^n(s)| \leq |\bar{M}_a^n(t) - \bar{M}_a^n(s)| + |\bar{M}_K^n(t) - \bar{M}_K^n(s)| \\ + \int_s^t \bar{\lambda}^n(\tau)d\tau + \gamma_k \bar{N}^n|t-s|.$$

Note that the last two terms are deterministic, so for any $\delta < \epsilon/(3\sup_n(\sup_{\tau\in[0,T]}\bar{\lambda}^n(\tau) + \bar{N}^n \max_k \gamma_k))$,

$$\mathbb{P}^n\Big(\sup_{|t-s|\leq\delta,\ s,t\in[0,T]} |\bar{Q}^n(t) - \bar{Q}^n(s)| > \epsilon\Big) \\ \leq \mathbb{P}^n\Big(\sup_{|t-s|\leq\delta,\ s,t\in[0,T]} |\bar{M}_a^n(t) - \bar{M}_a^n(s)| > \epsilon/3\Big) + \mathbb{P}^n\Big(\sup_{|t-s|\leq\delta,\ s,t\in[0,T]} |\bar{M}_k^n(t) - \bar{M}_k^n(s)| > \epsilon/3\Big)$$

Note that both $\bar{M}_a^n$ and $\bar{M}_k^n$ are square-integrable martingales. Thus, Doob's inequality (cf. Proposition 2.2.16 in Ethier and Kurtz (1986)) implies that $\bar{M}_a^n \Rightarrow 0$ and $\bar{M}_k^n \Rightarrow 0$ as $n \to \infty$. So

$$\mathbb{P}^n\Big(\mathbf{w}_T(\bar{Q}^n(\cdot),\delta) > \epsilon\Big) < \epsilon, \quad \text{for all large } n. \tag{EC.23}$$

A similar argument based on (35)–(37) can show that the same inequality as the above holds for $\bar{Z}_k^n$, for all $k = 0,\ldots,K$. This implies that

$$\mathbb{P}^n\Big(\mathbf{w}_T((\bar{Z}^n,\bar{Q}^n)(\cdot),\delta) > \epsilon\Big) < \epsilon, \quad \text{for all large } n. \tag{EC.24}$$

Inequalities (EC.22) and (EC.24) have verified that conditions (i) and (ii) of Theorem 3.7.2 in Ethier and Kurtz (1986) hold for the sequence $\{(\bar{Z}^n,\bar{Q}^n)\}_{n\in\mathbb{N}}$. Thus the relative compactness has been proved.

Proof of Lemma 3 We now restrict our attention to a convergent subsequence that converges to the limit $((z,q),\nu)$. With a little abuse of the notation, we still use the superscript n to index the convergent subsequence. It is convenient to assume, using Skorokhod's representation theorem (cf. Theorem 3.1.8 in Ethier and Kurtz (1986)), that the stochastic process for all n as well as the limit are defined on the same probability space on which the convergence $((\bar{Z}^n,\bar{Q}^n),\nu^n) \to ((z,q),\nu)$ is almost surely.

According to the stochastic dynamic equations (5)–(8), for any bounded functions $g:\bar{\mathbb{Z}}_+^{K+1}\to\mathbb{R}$,

$$\begin{aligned}&g(Z^n(t))-g(Z^n(0))\\ &=\sum_{j=0}^{K-1}\int_0^t\left[g\left(Z^n(s-)-e_j+e_{j+1}\right)-g(Z^n(s-))\right]\mathbf{1}_{\{Z^n(s-)\in\mathcal{A}_j\}}d\Lambda^n(s)\\ &\quad+\sum_{j=1}^{K}\int_0^t\left[g\left(Z^n(s-)-e_j+e_{j-1}\right)-g(Z^n(s-))\right]\mathbf{1}_{\{Q^n(s-)=0\}}dD_j^n(s).\end{aligned}$$

Since $\bar{M}_a^n$ and $\bar{M}_k^n$ defined in (33) and (34) are Martingales, it follows that

$$\begin{aligned}\bar{M}_g^n(t)=&\frac{1}{n}[g(Z^n(t))-g(Z^n(0))]\\ &-\sum_{j=0}^{K-1}\int_0^t\left[g(Z^n(s-)-e_j+e_{j+1})-g(Z^n(s-))\right]\mathbf{1}_{\{Z^n(s-)\in\mathcal{A}_j\}}\bar{\lambda}^n(s)ds\\ &-\sum_{j=1}^{K}\int_0^t\left[g\left(Z^n(s-)-e_j+e_{j-1}\right)-g(Z^n(s-))\right]\mathbf{1}_{\{\bar{Q}^n(s-)=0\}}\gamma_j\bar{Z}_j^n(s)ds\end{aligned}$$

is a Martingale for all bounded function $g:\bar{Z}_+^{K+1}\to\mathbb{R}$. Since g is bounded, $\mathbb{E}[(\bar{M}_g^n(t))^2]\to 0$ as $n\to\infty$. It follows from Doob's inequality (cf. Proposition 2.2.16 in Ethier and Kurtz (1986)) that $\bar{M}_g^n\Rightarrow 0$ as $n\to\infty$. Suppose $I(z(t))=k$ for some $k=0,1,\ldots,K$. The continuity of the limit (z,q) and the convergence imply that there exists a small interval $[t,t+\delta]$ where $\bar{Z}_k^n(\cdot)>0$ for all sufficiently large n. The boundedness of the function g implies that $\frac{1}{n}[g(Z^n(t+\delta))-g(Z^n(t))]\to 0$ as $n\to\infty$. Thus, the rest of the terms in $\bar{M}_g^n(t+\delta)-\bar{M}_g^n(t)$ will converge, as $n\to\infty$, to

$$\begin{aligned}&\sum_{j=0}^{k\wedge(K-1)}\int_{[t,t+\delta]\times\bar{Z}_+^{K+1}}\left[g(y-e_j+e_{j+1})-g(y)\right]\mathbf{1}_{\{y\in\mathcal{A}_j\}}\lambda(s)\nu(ds\times dy)\\ +&\sum_{j=k}^{K}\int_{[t,t+\delta]\times\bar{Z}_+^{K+1}}\left[g\left(y-e_j+e_{j-1}\right)-g(y)\right]\mathbf{1}_{\{q(s)=0\}}\gamma_j z_j(s)\nu(ds\times dy)\end{aligned}$$

which should be 0. Letting $\delta\to 0$ gives (51).

References

Walter, Wolfgang. 1998. *Ordinary differential equations*, *Graduate Texts in Mathematics*, vol. 182. Springer-Verlag, New York.

Perry, Ohad, Ward Whitt. 2011c. An ODE for an overloaded X model involving a stochastic averaging principle. *Stochastic Systems* **1** 17–66.

Durrett, Rick. 2010. *Probability: Theory and Examples*. 4th ed. Cambridge University Press.

Whitt, Ward. 2002. *Stochastic-process limits*. Springer Series in Operations Research, Springer-Verlag, New York.

Ethier, Stewart N., Thomas G. Kurtz. 1986. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.