E-Companion

This e-companion provides supplementary materials to the main paper. In Section EC.1, we give all the technical proofs omitted from the main paper. In Section EC.2, we test the robustness of GOLiQ with respect to key algorithmic hyperparameters. In Section EC.3, we compare GOLiQ to the online learning method in Huh et al. (2009). In Section EC.4, we report additional numerical studies. To facilitate readability, we formally summarize all notations in Table EC.1 including all model parameters and functions, algorithmic hyperparameters, and constants in the regret analysis.

EC.1. Proofs

EC.1.1. Proof of Lemma 1

Let Q_n^k be the queue length when customer $n-1$ in cycle k leaves the system. Then $Q_k = Q_{D_{k-1}}^{k-1} + 1$. The proof follows a stochastic ordering argument for $GI/GI/1$ models. Let $\hat{W}_n^k$, $\hat{X}_n^k$ and $\hat{Q}_n^k$ be the waiting times, observed busy periods, and queue length process in a $GI/GI/1$ queue with stationary control parameter $\mu_k \equiv \underline{\mu}$ and $p_k \equiv \underline{p}$, and with steady-state initial state, i.e., $\hat{W}_0^1 \stackrel{d}{=} W_\infty(\underline{\mu}, \underline{p})$, $\hat{X}_0^1 \stackrel{d}{=} X_\infty(\underline{\mu}, \underline{p})$ and $\hat{Q}_0^1 \stackrel{d}{=} Q_\infty(\underline{\mu}, \underline{p})$. Let's call this system the dominating system. Then, for all k,

$$\frac{U_n^k}{\lambda_n^k} \geq \frac{U_n^k}{\lambda(\underline{p})}, \text{ for } n = 1, 2, ..., Q_k, \text{ and } \frac{U_n^k}{\lambda_k} \geq \frac{U_n^k}{\lambda(\underline{p})}, \text{ for } n = Q_k + 1, 2, ..., D_k,$$

i.e., the arrival process in the dominating queue is the *upper envelope process* (UEP) for all possible arrival processes corresponding to any control sequence (μ_k, p_k). Similarly, the service process in the dominating queue is the *lower envelope process* (LEP) for all possible service processes corresponding to any control sequence. As a consequence, since $W_0^1 = 0$ and $Q_0^1 = 0$,

$$W_n^k \leq_{st} \hat{W}_n^k, \; X_n^k \leq_{st} \frac{\lambda(\underline{p})}{\lambda(\bar{p})} \cdot \hat{X}_n^k, \; Q_n^k \leq_{st} \hat{Q}_n^k.$$

Under Assumption 2, the moment generating function of the random variable $V_n/\underline{\mu} - U_n/\lambda(\underline{p})$ exists around the origin. Following Blanchet and Chen (2015), under Assumption 1, this condition can further imply that there exists a constant $\bar{\eta} > 0$ such that $\mathbb{E}[\exp\left(\bar{\eta}(V_n/\underline{\mu} - U_n/\lambda(\underline{p}))\right)] = 1$. (See the Remark on p.3222 in Blanchet and Chen (2015)) Then, following Theorem 1 of Abate et al. (1995), there exists a constant $\alpha > 0$ such that $\mathbb{P}(\hat{W}_n^k > x) \leq \alpha \exp(-\bar{\eta}x)$, for all $x > 0$. As a consequence, $\mathbb{E}[\exp(\eta \hat{W}_n^k)]$ is finite for $\eta < \bar{\eta}$,

and so are $\mathbb{E}[(\hat{W}_n^k)^m]$ for all $m \geq 1$. Given that the moments of waiting times are finite, we can conclude that $\mathbb{E}[(\hat{Q}_n^k)^m]$ and $\mathbb{E}[\exp(\eta \hat{Q}_n^k)]$ are finite for all $m \geq 1$, applying Theorem 10.4.3 in Asmussen (2003). Finally, the moments of the observed busy period $\mathbb{E}[(\hat{X}_n^k)^m]$ are finite following Proposition 4.2 in Nakayama et al. (2004). Therefore, we choose

$$M = \max_{1 \leq m \leq 4} \left\{ \mathbb{E}[(\hat{W}_n^k)^m],\ \frac{\lambda(\underline{p})^m}{\lambda(\bar{p})^m} \mathbb{E}[(\hat{X}_n^k)^m],\ \mathbb{E}[(\hat{Q}_n^k + 1)^m],\ \mathbb{E}[\exp(\eta \hat{W}_n^k)],\ \mathbb{E}[\exp(\eta(Q_n^k + 1))] \right\},$$

and this closes our proof. □

EC.1.2. Proof of Lemma 2

For $i \in \{1, 2\}$, define stopping times $\Gamma_i = \min\{n : W_n^i = 0\}$. For a fixed pair of inter-arrival and service time sequences, the consequent waiting time sequence W_k in a single-server queue is monotone in its initial state W_0. Without loss of generality, assume $W_0^1 \geq W_0^2$. Then, $W_n^1 \geq W_n^2$ for all $n \geq 1$ and therefore, $W_{\Gamma_1}^1 = W_{\Gamma_1}^2 = 0$. As the two queues are coupled with the same arrival and service time sequences, we will have $W_n^1 = W_n^2$ for all $n \geq \Gamma_1$. Therefore, we can conclude $W_n^1 = W_n^2$ for all $n \geq \max(\Gamma_1, \Gamma_2)$. For $n \leq \max(\Gamma_1, \Gamma_2)$, we have $|W_n^1 - W_n^2| \leq |W_0^1 - W_0^2|$ following Kella and Ramasubramanian (2012).

For simplicity of notation, we write $\lambda = \lambda(p)$. For $i \in \{1, 2\}$, define a random walk $R_{n+1}^i = R_n^i + S_n - \tau_n$ with $R_0^i = W_0^i$. (Recall that S_n and τ_n are the sequences of service and inter-arrival times.) By Lindley recursion, $\Gamma_i = \min\{n : R_n^i \leq 0\}$. Then, for any $n \geq 1$,

$$\begin{aligned}
\mathbb{P}(\Gamma_i \leq n) &\geq \mathbb{P}\left(\sum_{k=1}^{n} (S_k - \tau_k) < -W_0^i \right) \\
&\geq \mathbb{P}\left(\lambda \sum_{k=1}^{n} \tau_k \geq n(1-a), \mu \sum_{k=1}^{n} S_k \leq n(1+a) - \mu W_0^i \right),
\end{aligned}$$

where the second inequality holds as $(1-a)/\lambda > (1+a)/\mu$ given that $0 < a < (\underline{\mu} - \lambda(\underline{p}))/(\underline{\mu} + \lambda(\underline{p}))$ and that $\lambda/\mu \leq \lambda(\underline{p})/\underline{\mu}$. Recall that $\tau_k = U_k/\lambda$ and $S_k = V_k/\mu$. Therefore,

$$\mathbb{P}(\Gamma_i > n) \leq \mathbb{P}\left(\sum_{k=1}^{n} U_k < n(1-a) \right) + \mathbb{P}\left(\sum_{k=1}^{n} V_k > n(1+a) - \mu W_0^i \right).$$

Following Chebyshev's Inequality, we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{k=1}^{n} V_k > n(1+a) - \mu W_0^i \right) &\leq \frac{\mathbb{E}[\exp(\theta \sum_{k=1}^{n} V_k)]}{\exp(n\theta(1+a) - \mu\theta W_0^i)} = \exp(n(\phi_V(\theta) - (1+a)\theta)) \exp(\mu\theta W_0^i) \\
&\leq \exp(-n\gamma) \exp(\mu\theta W_0^i),
\end{aligned}$$

where the last inequality follows from Assumption 2. On the other hand, let Q be an exponentially tilted probability measure with respect to U, such that the likelihood ratio $\frac{dQ}{dP}(U) = \exp(-\theta U - \phi_U(-\theta))$. Then,

$$\mathbb{P}\left(\sum_{k=1}^{n} U_k < n(1-a)\right) = \mathbb{E}^Q\left[\exp\left(\theta\sum_{k=1}^{n} U_k + n\phi_U(-\theta)\right)\mathbf{1}_{\left\{\sum_{k=1}^{n} U_k < n(1-a)\right\}}\right]$$
$$\leq \exp(n(1-a)\theta + n\phi_U(-\theta)) = \exp(n((1-a)\theta + \phi_U(-\theta))) \leq \exp(-n\gamma).$$

In summary, we have $\mathbb{P}(\Gamma_i > n) \leq \exp(-n\gamma)\,(1+\exp(\mu\theta W_0^i))$, $i = 1, 2$. So, we can conclude

$$\mathbb{E}[|W_n^1 - W_n^2|^m] \leq \mathbb{P}(\max(\Gamma_1, \Gamma_2) > n)|W_0^1 - W_0^2|^m$$
$$\leq e^{-\gamma n}\left(2 + e^{\mu\theta W_0^1} + e^{\mu\theta W_0^2}\right)|W_0^1 - W_0^2|^m. \qquad \square$$

EC.1.3. Proof of Lemma 3

Define two auxiliary random walks:

$$Y_n = W_0 + \sum_{i=1}^{n}\left(\frac{V_i}{\mu_i} - \frac{U_i}{\lambda_i}\right), \tilde{Y}_n = \tilde{W}_0 + \sum_{i=1}^{n}\left(\frac{V_i}{\tilde{\mu}_i} - \frac{U_i}{\tilde{\lambda}_i}\right).$$

Then, for any $n \geq 1$, we could express W_n and $\tilde{W}_n$ as

$$W_n = Y_n - \min_{1\leq m\leq n} Y_m \wedge 0, \quad \tilde{W}_n = \tilde{Y}_n - \min_{1\leq m\leq n} \tilde{Y}_m \wedge 0.$$

Let $\tau = \arg\min_{1\leq m\leq n} Y_m$ and $\tilde{\tau} = \arg\min_{1\leq m\leq n} \tilde{Y}_m$. Note that following the above notation, for each n, W_n is the waiting time of customer n and as a consequence, $\frac{U_n}{\lambda_n}$ should be understood as the inter-arrival time between customers $n-1$ and n, and $\frac{V_n}{\mu_n}$ as the service time of customer $n-1$.

Case 1: If $Y_\tau \leq 0$ and $\tilde{Y}_{\tilde{\tau}} \leq 0$, i.e., both W_t and $\tilde{W}_t$ hit zero before n, we have

$$Y_n - Y_{\tilde{\tau}} - (\tilde{Y}_n - \tilde{Y}_{\tilde{\tau}}) \leq W_n - \tilde{W}_n = Y_n - Y_\tau - (\tilde{Y}_n - \tilde{Y}_{\tilde{\tau}}) \leq Y_n - Y_\tau - (\tilde{Y}_n - \tilde{Y}_\tau).$$

So, in this case

$$|W_n - \tilde{W}_n| \leq \sum_{i=\tau\wedge\tilde{\tau}+1}^{n}\left|\frac{1}{\mu_i} - \frac{1}{\tilde{\mu}_i}\right|V_i + \sum_{i=\tau\wedge\tilde{\tau}+1}^{n}\left|\frac{1}{\lambda_i} - \frac{1}{\tilde{\lambda}_i}\right|U_i.$$

Recall that X_n (and $\tilde{X}_n$) is the age of the server's busy time observed by customer n upon arrival. By definition, $W_\tau = 0$ and therefore,

$$X_n = \sum_{i=\tau+1}^{n}\frac{U_i}{\lambda_i}, \quad X_n + W_n = \sum_{i=\tau+1}^{n}\frac{V_i}{\mu_i}.$$

The second equation holds as the server has just served $n-\tau$ customers (indexed from τ to $n-1$) in the current busy cycle when customer n enters service. Then,

$$\sum_{i=\tau+1}^{n}\left|\frac{1}{\mu_i}-\frac{1}{\tilde{\mu}_i}\right|V_i+\sum_{i=\tau+1}^{n}\left|\frac{1}{\lambda_i}-\frac{1}{\tilde{\lambda}_i}\right|U_i\le\frac{c_\mu}{\underline{\mu}}(X_n+W_n)+\frac{c_\lambda}{\underline{\lambda}}X_n.$$

Following a similar argument, we have

$$\sum_{i=\tilde{\tau}+1}^{n}\left|\frac{1}{\mu_i}-\frac{1}{\tilde{\mu}_i}\right|V_i+\sum_{i=\tilde{\tau}+1}^{n}\left|\frac{1}{\lambda_i}-\frac{1}{\tilde{\lambda}_i}\right|U_i\le\frac{c_\mu}{\underline{\mu}}(\tilde{X}_n+\tilde{W}_n)+\frac{c_\lambda}{\underline{\lambda}}\tilde{X}_n.$$

Therefore, in this case, we have

$$|W_n-\tilde{W}_n|\le\left(\frac{c_\mu}{\underline{\mu}}+\frac{c_\lambda}{\underline{\lambda}}\right)\max(X_n,\tilde{X}_n)+\frac{c_\mu}{\underline{\mu}}\max(W_n,\tilde{W}_n).$$

Case 2: If $Y_\tau>0$ or $\tilde{Y}_{\tilde{\tau}}>0$, we can inductively derive that

$$|W_n-\tilde{W}_n|\le|W_0-\tilde{W}_0|+\sum_{i=1}^{n}\left|\frac{1}{\mu_i}-\frac{1}{\tilde{\mu}_i}\right|V_i+\sum_{i=1}^{n}\left|\frac{1}{\lambda_i}-\frac{1}{\tilde{\lambda}_i}\right|U_i.$$

In detail, it suffices to show that, for all $1\le m\le n$,

$$|W_m-\tilde{W}_m|\le|W_{m-1}-\tilde{W}_{m-1}|+\left|\frac{1}{\mu_m}-\frac{1}{\tilde{\mu}_m}\right|V_m+\left|\frac{1}{\lambda_m}-\frac{1}{\tilde{\lambda}_m}\right|U_m. \tag{EC.1}$$

Without loss of generality, we assume $Y_\tau>0$. By definition, $Y_\tau=\min_{1\le l\le n}Y_l$ and hence $W_l=Y_l>0$ for all $1\le l\le n$. Then,

$$|W_m-\tilde{W}_m|=\left|W_{m-1}-\frac{U_m}{\lambda_m}+\frac{V_m}{\mu_m}-\left(\tilde{W}_{m-1}-\frac{U_m}{\tilde{\lambda}_m}+\frac{V_m}{\tilde{\mu}_m}\right)^+\right|.$$

If $\tilde{W}_m>0$, we have

$$\begin{aligned}|W_m-\tilde{W}_m|&=\left|W_{m-1}-\frac{U_m}{\lambda_m}+\frac{V_m}{\mu_m}-\left(\tilde{W}_{m-1}-\frac{U_m}{\tilde{\lambda}_m}+\frac{V_m}{\tilde{\mu}_m}\right)\right|\\&\le|W_{m-1}-\tilde{W}_{m-1}|+\left|\frac{1}{\mu_m}-\frac{1}{\tilde{\mu}_m}\right|V_m+\left|\frac{1}{\lambda_m}-\frac{1}{\tilde{\lambda}_m}\right|U_m.\end{aligned}$$

On the other hand, if $\tilde{W}_m=0$, we have $\tilde{W}_{m-1}-\frac{U_m}{\tilde{\lambda}_m}+\frac{V_m}{\tilde{\mu}_m}\le 0$. So,

$$\begin{aligned}|W_m-\tilde{W}_m|&=W_m-0\le W_m-\left(\tilde{W}_{m-1}-\frac{U_m}{\tilde{\lambda}_m}+\frac{V_m}{\tilde{\mu}_m}\right)\\&=\left|W_{m-1}-\frac{U_m}{\lambda_m}+\frac{V_m}{\mu_m}-\left(\tilde{W}_{m-1}-\frac{U_m}{\tilde{\lambda}_m}+\frac{V_m}{\tilde{\mu}_m}\right)\right|\\&\le|W_{m-1}-\tilde{W}_{m-1}|+\left|\frac{1}{\mu_m}-\frac{1}{\tilde{\mu}_m}\right|V_m+\left|\frac{1}{\lambda_m}-\frac{1}{\tilde{\lambda}_m}\right|U_m.\end{aligned}$$

This closes the proof of (EC.1).

As a result of (EC.1), if $Y_\tau > 0$, we can conclude the system (associated with (μ_n, λ_n)) was kept busy from time 0 until customer n enters service. As a consequence, as $X_0 \geq 0$, we have

$$X_n \geq \sum_{i=1}^{n} \frac{U_i}{\lambda_i}, \quad X_n + W_n \geq \sum_{i=1}^{n} \frac{V_i}{\mu_i}.$$

Therefore,

$$\sum_{i=1}^{n} \left| \frac{1}{\mu_i} - \frac{1}{\tilde{\mu}_i} \right| V_i + \sum_{i=1}^{n} \left| \frac{1}{\lambda_i} - \frac{1}{\tilde{\lambda}_i} \right| U_i \leq \frac{c_\mu}{\underline{\mu}} (\max(X_n, \tilde{X}_n) + \max(W_n, \tilde{W}_n)) + \frac{c_\lambda}{\underline{\lambda}} \max(X_n, \tilde{X}_n),$$

and hence we can also conclude

$$|W_n - \tilde{W}_n| \leq |W_0 - \tilde{W}_0| + \left(\frac{c_\mu}{\underline{\mu}} + \frac{c_\lambda}{\underline{\lambda}} \right) \max(X_n, \tilde{X}_n) + \frac{c_\mu}{\underline{\mu}} \max(W_n, \tilde{W}_n).$$

□

EC.1.4. Proof of Lemma 4

By the inequality that $(a+b)^m \leq 2^{m-1}(a^m + b^m)$ for $m \geq 1$, we have

$$\begin{aligned}
&\mathbb{E}[|W_\infty(\mu_1, p_1) - W_\infty(\mu_2, p_2)|^m] \\
&\leq 2^{m-1} \left(\mathbb{E}[|W_\infty(\mu_1, p_1) - W_\infty(\mu_2, p_1)|^m + |W_\infty(\mu_2, p_1) - W_\infty(\mu_2, p_2)|^m] \right).
\end{aligned}$$

It suffices to prove that there exist two constant $B_1, B_2 > 0$ such that for $1 \leq m \leq 4$,

$$\begin{aligned}
\mathbb{E}[|W_\infty(\mu_1, p_1) - W_\infty(\mu_2, p_1)|^m] &\leq B_1 |\mu_1 - \mu_2|^m, \\
\mathbb{E}[|W_\infty(\mu_2, p_1) - W_\infty(\mu_2, p_2)|^m] &\leq B_2 |p_1 - p_2|^m.
\end{aligned}$$

Without loss of generality, assume $\mu_1 < \mu_2$. We now construct two stationary sequences $\{(W_n^{\mu_i} : n \leq 0), i = 1, 2\}$ that are coupled "from the past". Let V_j and U_j be two i.i.d sequences corresponding to the service and inter-arrival times. For each i, we define a random walk:

$$Y_0^{\mu_i} = 0, \; Y_n^{\mu_i} = \sum_{j=1}^{n} \left(\frac{V_j}{\mu_i} - \frac{U_j}{\lambda(p_1)} \right), \; \forall n \geq 1.$$

It is clear that $Y_n^{\mu_i}$ is a random walk with negative drift for $i = 1, 2$. Define

$$W_{-n}^{\mu_i} = \max_{j \geq n} Y_j^{\mu_i} - Y_n^{\mu_i}, n \geq 0.$$

It is known in literature (see, for example, Blanchet and Chen (2015)) that $W_{-n}^{\mu_i}$ is a stationary waiting time process of a $GI/GI/1$ queue, starting from $-\infty$, with parameter (μ_i, p_1). In particular, the dynamics of $W_{-n}^{\mu_i}$ satisfies that

$$W_{-n+1}^{\mu_i} = \left(W_{-n}^{\mu_i} + \frac{V_n}{\mu_i} - \frac{U_n}{\lambda(p_1)}\right)^+, \text{ for } n \geq 1,$$

with V_n/μ_i being the service time of customer $-n$ and $U_n/\lambda(p_1)$ being the inter-arrival time between customer $-n$ and $-n+1$. For a fixed sequence of (V_n, U_n), we have

$$W_0^{\mu_1} = \max_{j\geq 0} Y_j^{\mu_1}, \quad \text{and } W_0^{\mu_2} = \max_{j\geq 0} Y_j^{\mu_2}.$$

As $Y_j^{\mu_1} \geq Y_j^{\mu_2}$, we have $W_0^{\mu_1} \geq W_0^{\mu_2}$. Besides, let $\tau = \arg\max_{j\geq 0} Y_j^{\mu_1}$, we have

$$W_0^{\mu_1} - W_0^{\mu_2} = \max_{j\geq 0} Y_j^{\mu_1} - \max_{j\geq 0} Y_j^{\mu_2} = Y_\tau^{\mu_1} - \max_{j\geq 0} Y_j^{\mu_2} \leq Y_\tau^{\mu_1} - Y_\tau^{\mu_2}.$$

As a consequence, we have

$$|W_0^{\mu_1} - W_0^{\mu_2}| \leq \sum_{n=1}^{\tau} \left(\frac{V_n}{\mu_1} - \frac{V_n}{\mu_2}\right) \leq \frac{\mu_2 - \mu_1}{\mu_1} \sum_{n=1}^{\tau} \frac{V_n}{\mu_1}, \qquad \text{with } \tau = \inf\{n : W_{-n} = 0\}.$$

Note that V_n/μ_1 is the service time of customer $-n$ in the system with parameter (p_1, μ_1). By the definition of τ, customer $-\tau$ enters service immediately upon the arrival and the queue remains busy by arrival of customer 0. Therefore, the summation of service times on the right hand side equals to the time between the arrival of customer $-\tau$ and the departure of customer -1, which equals to the observed busy period at the arrival of customer 0 plus its waiting time, i.e.,

$$|W_0^{\mu_1} - W_0^{\mu_2}| \leq \frac{\mu_2 - \mu_1}{\mu_1} \sum_{n=1}^{\tau} \frac{V_n}{\mu_1} = \frac{\mu_2 - \mu_1}{\mu_1}(X_0^{\mu_1} + W_0^{\mu_1}).$$

Therefore, for each n,

$$\mathbb{E}[|W_0^{\mu_1} - W_0^{\mu_2}|^m] \leq \frac{(\mu_2 - \mu_1)^m}{\mu_1^m} \mathbb{E}[(X_0^{\mu_1} + W_0^{\mu_1})^m] \leq \frac{(\mu_2 - \mu_1)^m}{\underline{\mu}^m} \mathbb{E}[(X_0^{\mu_1} + W_0^{\mu_1})^m].$$

Following Lemma 1, $\mathbb{E}[(X_0^{\mu_1} + W_0^{\mu_1})^m] \leq 2^m M$. Let $B_1 = \max_{1\leq m\leq 4} 2^m M/\underline{\mu}^m$ and we conclude, for $1 \leq m \leq 4$,

$$\mathbb{E}[|W_0^{\mu_1} - W_0^{\mu_2}|^m] \leq B_1 |\mu_1 - \mu_2|^m.$$

The bound for $\mathbb{E}[|W_\infty(\mu_2, p_1) - W_\infty(\mu_2, p_2)|^m]$ follows a similar argument and therefore we only provide a sketch of the proof. Without loss of generality, we assume $p_1 < p_2$ and

consider two stationary waiting time process $\{(W_n^{p_i}: n \le 0), \lambda_i = \lambda(p_i), i = 1, 2\}$ that are coupled from past with the same sequence (V_n, U_n) in a similar way as we introduced previously. Then, we have $|W_0^{p_1} - W_0^{p_2}| \le (\lambda_1 - \lambda_2) X_0^{p_1}/\lambda_2$, and therefore,

$$\mathbb{E}[|W_0^{p_1} - W_0^{p_2}|^m] \le B_2 |p_1 - p_2|^m, \text{ with } B_2 = \max_{1 \le m \le 4, \underline{p} \le p \le \bar{p}} (M|\lambda'(p)|^m / \lambda(\bar{p})^m).$$

As a consequence, we can take

$$B = 8 \cdot \max_{1 \le m \le 4} (2^m M / \underline{\mu}^m) \vee \max_{1 \le m \le 4, \underline{p} \le p \le \bar{p}} (M|\lambda'(p)|^m / \lambda(\bar{p})^m). \tag{EC.2}$$

EC.1.5. Full Proof of Theorem 1

We first give the proofs of Corollaries 1–3.

Proof of Corollary 1 For any $n \ge d_k$,

$$\mathbb{E}[|W_n^k - \bar{W}_n^k|] = \mathbb{E}[|W_n^k - \bar{W}_n^k| 1(Q_k < d_k)] + \mathbb{E}[|W_n^k - \bar{W}_n^k| 1(Q_k \ge d_k)].$$

Given that $Q_k < d_k$, by definition, W_n^k is synchronously coupled with $\bar{W}_n^k$ for $n \ge d_k + 1$. Note that given $Q_k < d_k$, U_n^k and V_n^k are independent of Q_k for $n \ge d_k + 1$. As a consequence, by Lemma 2, the conditional expectation

$$\mathbb{E}\left[|W_n^k - \bar{W}_n^k| \;\middle|\; Q_k < d_k, W_{d_k}^k, \bar{W}_{d_k}^k \right] \le e^{-\gamma(n-d_k)} (2 + e^{\bar{\mu}\theta W_{d_k}^k} + e^{\bar{\mu}\theta \bar{W}_{d_k}^k}) \left| W_{d_k}^k - \bar{W}_{d_k}^k \right|.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[|W_n^k - \bar{W}_n^k| 1(Q_k < d_k)] &\le e^{-\gamma(n-d_k)} \mathbb{E}\left[(2 + e^{\bar{\mu}\theta W_{d_k}^k} + e^{\bar{\mu}\theta \bar{W}_{d_k}^k}) \left| W_{d_k}^k - \bar{W}_{d_k}^k \right| 1(Q_k < d_k) \right] \\
&\le e^{-\gamma(n-d_k)} \mathbb{E}\left[(2 + e^{\bar{\mu}\theta W_{d_k}^k} + e^{\bar{\mu}\theta \bar{W}_{d_k}^k}) \left| W_{d_k}^k - \bar{W}_{d_k}^k \right| \right] \\
&\le e^{-\gamma(n-d_k)} \left(2 + \mathbb{E}\left[\left(e^{\bar{\mu}\theta W_{d_k}^k} + e^{\bar{\mu}\theta \bar{W}_{d_k}^k} \right)^2 \right]^{1/2} \right) \mathbb{E}\left[\left| W_{d_k}^k - \bar{W}_{d_k}^k \right|^2 \right]^{1/2}.
\end{aligned}$$

By Lemma 1 and Assumption 2, we have

$$\begin{aligned}
\mathbb{E}\left[\left(e^{\bar{\mu}\theta W_{d_k}^k} + e^{\bar{\mu}\theta \bar{W}_{d_k}^k} \right)^2 \right] &\le 2 \left(\mathbb{E}[e^{2\bar{\mu}\theta W_{d_k}^k}] + \mathbb{E}[e^{2\bar{\mu}\theta \bar{W}_{d_k}^k}] \right) \le 4M, \\
\mathbb{E}\left[|W_{d_k}^k - \bar{W}_{d_k}^k|^2 \right] &\le 2 \left(\mathbb{E}[(W_{d_k}^k)^2] + \mathbb{E}[(\bar{W}_{d_k}^k)^2] \right) \le 4M.
\end{aligned}$$

As a consequence, we have

$$\mathbb{E}[|W_n^k - \bar{W}_n^k| 1(Q_k < d_k)] \le e^{-\gamma(n-d_k)} A, \text{with } A = 4\sqrt{M} + 4M.$$

On the other hand,

$$\mathbb{E}[|W_n^k - \bar{W}_n^k| 1(Q_k \geq d_k)] \leq \mathbb{E}\left[|W_n^k - \bar{W}_n^k|^2\right]^{1/2} \mathbb{P}(Q_k \geq d_k)^{1/2}.$$

Again, by Lemma 1, $\mathbb{E}\left[|W_n^k - \bar{W}_n^k|^2\right] \leq 4M$. As $d_k = \lceil 4\log(k)/\min(\gamma,\eta)\rceil$,

$$\mathbb{P}(Q_k \geq d_k) \leq e^{-\eta d_k}\mathbb{E}\left[e^{\eta Q_k}\right] \leq k^{-4}M.$$

In summary, we have, for $n \geq d_k + 1$,

$$\mathbb{E}[|W_n^k - \bar{W}_n^k|] \leq e^{-\gamma(n-d_k)}A + 2Mk^{-2}.$$

As a direct consequence,

$$\begin{aligned}|I_1| &= \left|\mathbb{E}\left[\sum_{n=\tilde{d}_k+1}^{D_k} W_n^k - w(\mu_k, p_k)\right]\right| \leq \sum_{n=\tilde{d}_k+1}^{D_k} \mathbb{E}[|W_n^k - \bar{W}_n^k|] \\ &\leq \sum_{n=\tilde{d}_k+1}^{\infty} e^{-\gamma(n-d_k)}A + 2Mk^{-2}D_k \leq \frac{A}{1-e^{-\gamma}}k^{-1} + 2MK_2k^{-\alpha} = O(k^{-\alpha}).\end{aligned}$$

Proof of Corollary 2 Recall that by (6), for each cycle k,

$$W_n^k = \begin{cases} \left(W_{n-1}^k + \frac{V_n^k}{\mu_k} - \frac{U_n^k}{\lambda_n^k}\right)^+ & \text{for } 1 \leq n \leq Q_k \wedge D_k; \\ \left(W_{n-1}^k + \frac{V_n^k}{\mu_k} - \frac{U_n^k}{\lambda_k}\right)^+ & \text{for } (Q_k+1)\wedge(D_k+1) \leq n \leq D_k. \end{cases}, \quad W_0^k = W_{D_{k-1}}^{k-1}.$$

Define

$$\tilde{W}_n^k = \begin{cases} \left(\tilde{W}_{n-1}^k + \frac{V_n^k}{\mu_k} - \frac{U_n^k}{\lambda_{k-1}}\right)^+ & \text{for } 1 \leq n \leq Q_k \wedge D_k; \\ \left(\tilde{W}_{n-1}^k + \frac{V_n^k}{\mu_k} - \frac{U_n^k}{\lambda_k}\right)^+ & \text{for } (Q_k+1)\wedge(D_k+1) \leq n \leq D_k. \end{cases}, \quad \tilde{W}_0^k = W_{D_{k-1}}^{k-1}.$$

Then, in the case $Q_{k-1} < D_{k-1}$, we have $W_n^k = \tilde{W}_n^k$ for all $1 \leq n \leq D_k$. As a consequence, we have

$$|W_n^k - \bar{W}_{D_{k-1}+n}^{k-1}| \leq |\tilde{W}_n^k - \bar{W}_{D_{k-1}+n}^{k-1}| + |W_n^k - \bar{W}_{D_{k-1}+n}^{k-1}| \cdot 1(Q_{k-1} \geq D_{k-1}).$$

For the second term, by Lemma 1, we have, for $k \geq 2$,

$$\begin{aligned}\mathbb{E}[|W_n^k - \bar{W}_{D_{k-1}+n}^{k-1}| \cdot 1(Q_{k-1} \geq D_{k-1})] &\leq \mathbb{E}[(W_n^k - \bar{W}_{D_{k-1}+n}^{k-1})^2]^{1/2}\mathbb{P}(Q_{k-1} \geq D_{k-1})^{1/2} \\ &\leq \left(2\mathbb{E}[(W_n^k)^2] + 2\mathbb{E}[(\bar{W}_{D_{k-1}+n}^{k-1})^2]\right)^{1/2} \left(\exp(-\eta D_{k-1})\mathbb{E}[\exp(\eta Q_{k-1}]\right)^{1/2} \\ &\leq 2M(k-1)^{-3} \leq 16Mk^{-3}\end{aligned}$$

For the first term, by definition, $\bar{W}^{k-1}_{D_{k-1}+n}$ is a waiting time sequence with service and arrival rates $(\mu_{k-1}, \lambda(p_{k-1}))$ and $\tilde{W}^k_n$ is a sequence with rates $(\mu_k, \lambda(p_k))$ or $(\mu_k, \lambda(p_{k-1}))$. As a consequence, by applying Lemma 3, we have

$$\begin{aligned}|\tilde{W}^k_n - \bar{W}^{k-1}_{D_{k-1}+n}| \le & |\tilde{W}^k_0 - \bar{W}^{k-1}_{D_{k-1}}| + \left(\frac{|\mu_k - \mu_{k-1}|}{\underline{\mu}} + \frac{|\lambda(p_k) - \lambda(p_{k-1})|}{\lambda(\bar{p})}\right) \max(\tilde{X}^k_n, \bar{X}^{k-1}_{D_{k-1}+n}) \\ & + \frac{|\mu_k - \mu_{k-1}|}{\underline{\mu}} \max(\tilde{W}^k_n, \bar{W}^{k-1}_{D_{k-1}+n}).\end{aligned}$$

By Lemma 1, we have that $\max(\tilde{X}^k_n, \bar{X}^{k-1}_{D_{k-1}+n}) \le \frac{\lambda(p)}{\lambda(\bar{p})}\hat{X}^k_n$ and $\max(\tilde{W}^k_n, \bar{W}^{k-1}_{D_{k-1}+n}) \le \hat{W}^k_n$, where $\hat{X}^k_n$ and $\hat{W}^k_n$ are the observed busy period and waiting time in a stationary $GI/GI/1$ queue with rate $(\underline{\mu}, \underline{p})$ as defined in Lemma 1. On the other hand, under Condition (b) of Theorem 1,

$$\begin{aligned}\mathbb{E}[|\mu_k - \mu_{k-1}|^2] &\le \mathbb{E}[\|x_k - x_{k+1}\|^2] \le K_2 k^{-2\alpha} \\ \mathbb{E}[|\lambda_k - \lambda_{k-1}|^2] &\le K_2 \left(\max_p \lambda'(p)\right)^2 k^{-2\alpha} \equiv K_6 k^{-2\alpha}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[|\mu_k - \mu_{k-1}| \max(\tilde{X}^k_n, \bar{X}^{k-1}_{D_{k-1}+n})\right] &\le \mathbb{E}[(\mu_k - \mu_{k-1})^2]^{1/2} \frac{\lambda(p)}{\lambda(\bar{p})} \mathbb{E}[(\hat{X}^k_n)^2]^{1/2} \le \sqrt{K_2}\sqrt{M} k^{-\alpha}; \\ \mathbb{E}[|\lambda(p_k) - \lambda(p_{k-1})| \max(\tilde{X}^k_n, \bar{X}^{k-1}_{D_{k-1}+n})] &\le \mathbb{E}[(\lambda_k - \lambda_{k-1})^2]^{1/2} \frac{\lambda(p)}{\lambda(\bar{p})} \mathbb{E}[(\hat{X}^k_n)^2]^{1/2} \le \sqrt{K_6}\sqrt{M} k^{-\alpha}; \\ \mathbb{E}\left[|\mu_k - \mu_{k-1}| \max(\tilde{W}^k_n, \bar{W}^{k-1}_{D_{k-1}+n})\right] &\le \mathbb{E}[(\mu_k - \mu_{k-1})^2]^{1/2} \mathbb{E}[(\hat{W}^k_n)^2]^{1/2} \le \sqrt{K_2}\sqrt{M} k^{-\alpha}.\end{aligned}$$

Finally, by Corollary 1, we have

$$\mathbb{E}[|\tilde{W}^k_0 - \bar{W}^{k-1}_{D_{k-1}}|] = \mathbb{E}[|\bar{W}^{k-1}_{D_{k-1}} - W^k_0|] = \mathbb{E}[|\bar{W}^{k-1}_{D_{k-1}} - W^{k-1}_{D_{k-1}}|] \le (A+2M)(k-1)^{-2} \le (4A+8M)k^{-2}.$$

In summary, we can conclude

$$\begin{aligned}|\mathbb{E}[W^k_n - w(\mu_k, p_k)]| &\le \mathbb{E}[|w(\mu_{k-1}, p_{k-1}) - w(\mu_k, p_k)|] + \mathbb{E}[|W^k_n - \bar{W}^{k-1}_{D_{k-1}+n}|] \\ &\le B\mathbb{E}[|\mu_k - \mu_{k-1}| + |\lambda(p_k) - \lambda(p_{k-1})|] + \left(\frac{2\sqrt{K_2}}{\underline{\mu}} + \frac{\sqrt{K_6}}{\lambda(\bar{p})}\right)\sqrt{M}k^{-\alpha} + O(k^{-2}) \\ &\le B(\sqrt{K_2} + \sqrt{K_6})k^{-\alpha} + \left(\frac{2\sqrt{K_2}}{\underline{\mu}} + \frac{\sqrt{K_6}}{\lambda(\bar{p})}\right)\sqrt{M}k^{-\alpha} + O(k^{-2}) \\ &= O(k^{-\alpha}),\end{aligned}$$

where the second inequality follows from Lemma 4. As a direct consequence, $|I_2| = O(k^{-\alpha}\log(k))$ as $\tilde{d}_k = O(\log(k))$. □

Proof of Corollary 3 Note that by Lemma 1,

$$\left|h_0\mathbb{E}[W_\infty(\mu_k,p_k)]+\frac{h_0}{\mu_k}-p_k\right|\le h_0M+h_0\underline{\mu}^{-1}+\bar{p}=O(1).$$

So it suffices to show that

$$\mathbb{E}[|D_k-\lambda(p_k)T_k|]=O(k^{-\alpha}),\quad \mathbb{E}\left[\sum_{n=1}^{Q_k\wedge D_k}|p_k-p_n^k|\right]=O(k^{-\alpha}).$$

Given μ_k and p_k, T_k is the time for the D_k-th customer to enter service. Let F_n^k be the inter-service time between the $(n-1)$-th and the n-th customers in cycle k. Then, $T_k=\sum_{n=1}^{D_k}F_n^k$ and for each n,

$$F_n^k=\begin{cases}\frac{U_n^k}{\lambda_n^k}+W_n^k-W_{n-1}^k & \text{for } 1\le n\le Q_k\\ \frac{U_n^k}{\lambda_k}+W_n^k-W_{n-1}^k & Q_k+1\le n\le D_k.\end{cases}$$

Therefore,

$$\begin{aligned}T_k&=\sum_{k=1}^{D_k}F_n^k=\sum_{n=1}^{Q_k\wedge D_k}\frac{U_n^k}{\lambda_n^k}+\frac{1}{\lambda_k}\sum_{n=Q_k+1}^{D_k}U_n^k+W_{D_k}^k-W_0^k\\&=\frac{1}{\lambda_k}\sum_{n=1}^{D_k}U_n^k+W_{D_k}^k-W_0^k+\sum_{n=1}^{Q_k\wedge D_k}U_n^k\left(\frac{1}{\lambda_n^k}-\frac{1}{\lambda_k}\right).\end{aligned}$$

As a consequence,

$$|\mathbb{E}\left[(D_k-\lambda_kT_k)\right]|\le\lambda_k|\mathbb{E}[W_{D_k}^k]-\mathbb{E}[W_0^k]|+\mathbb{E}\left[\sum_{k=1}^{Q_k\wedge D_k}U_n^k\left|\frac{\lambda_k}{\lambda_n^k}-1\right|\right].$$

Following Corollary 1 and Lemma 4, for $k\ge2$, the first term

$$\begin{aligned}|\mathbb{E}[W_{D_k}^k]-\mathbb{E}[W_0^k]|&\le\mathbb{E}|W_{D_k}^k-\bar{W}_{D_k}^k|+\mathbb{E}|W_{D_{k-1}}^{k-1}-\bar{W}_{D_{k-1}}^{k-1}|+|\mathbb{E}[\bar{W}_{D_k}^k]-\mathbb{E}[\bar{W}_{D_{k-1}}^{k-1}]|\\&=(A+2M)\left(k^{-2}+(k-1)^{-2}\right)+B\sqrt{K_2}k^{-\alpha}=O(k^{-\alpha}).\end{aligned}$$

As to the second term, by definition, the customers 1 to Q_k-1 arrive to the system while customer 0 is waiting in the system, and therefore,

$$0\le\sum_{i=1}^{(Q_k-1)\wedge D_k}\frac{U_i^k}{\bar{\lambda}}\le\sum_{i=1}^{(Q_k-1)\wedge D_k}\frac{U_i^k}{\lambda_i^k}\le W_0^k\quad\Rightarrow\quad\mathbb{E}\left[\left(\sum_{i=1}^{Q_k\wedge D_k}U_i^k\right)^2\right]\le\mathbb{E}\left[(\bar{\lambda}W_0^k+U_{Q_k}^k)^2\right]\le4\bar{\lambda}^2M.$$

Here, $\mathbb{E}[(U^k_{Q_k})^2]$ is bounded since we assume that U is light-tailed (Assumption 2). For the simplicity of notation, we just assume that $\mathbb{E}\left[\left(\frac{U_i^2}{\lambda}\right)^2\right] < M$ for the same M in Lemma 1. Then,

$$\begin{aligned}\mathbb{E}\left[\sum_{k=1}^{Q_k\wedge D_k} U_n^k\Big|\frac{\lambda_k}{\lambda_n^k}-1\Big|\right] &\le \mathbb{E}\left[\sum_{k=1}^{Q_k\wedge D_k} U_n^k\Big|\frac{\lambda_k}{\lambda_{k-1}}-1\Big|\right]+\mathbb{E}\left[\sum_{k=1}^{Q_k\wedge D_k} U_n^k\cdot\frac{\bar\lambda}{\underline\lambda}\cdot 1(Q_{k-1}\ge D_{k-1})\right]\\ &\le 2\bar\lambda\sqrt{M}\mathbb{E}\left[\left|\frac{\lambda_k}{\lambda_{k-1}}-1\right|^2\right]^{1/2}+\frac{2\bar\lambda^2}{\underline\lambda}\sqrt{M}\mathbb{P}(Q_{k-1}\ge D_{k-1})^{1/2}\\ &\le \frac{2\bar\lambda\sqrt{M}K_6^{1/2}}{\underline\lambda}k^{-\alpha}+\frac{16\bar\lambda^2}{\underline\lambda}Mk^{-3}=O(k^{-\alpha}).\end{aligned}$$

Finally,

$$\begin{aligned}\mathbb{E}\left[\sum_{n=1}^{Q_k\wedge D_k}|p_k-p_n^k|\right] &\le \mathbb{E}\left[\sum_{n=1}^{Q_k\wedge D_k}|p_k-p_n^k|\cdot 1(Q_{k-1}<D_{k-1})\right]\\ &\quad+\mathbb{E}\left[\sum_{n=1}^{Q_k\wedge D_k}|p_k-p_n^k|\cdot 1(Q_{k-1}\ge D_{k-1})\right]\\ &\le \mathbb{E}\left[Q_{k-1}^2\right]^{1/2}\mathbb{E}\left[|p_k-p_{k-1}|^2\right]^{1/2}+\mathbb{E}\left[\bar p^2 Q_k^2\right]^{1/2}\mathbb{P}(Q_{k-1}\ge D_{k-1})^{1/2}\\ &\le \sqrt{MK_2}k^{-\alpha}+8\bar pMk^{-3}=O(k^{-\alpha})\end{aligned}$$

Therefore, $I_3=O(k^{-\alpha})$. □

Finishing the proof of Theorem 1 First, by Corollary 1, we have

$$|I_1|\le\frac{A}{1-e^{-\gamma}}k^{-1}+2MK_2k^{-\alpha}=O(k^{-\alpha}).$$

By Corollary 2,

$$|I_2|\le\frac{5}{\min(\gamma,\eta)}\left(B(\sqrt{K_2}+\sqrt{K_6})+\left(\frac{2\sqrt{K_2}}{\underline\mu}+\frac{\sqrt{K_6}}{\underline\lambda}\right)\sqrt{M}+(4A+8M)\right)k^{-\alpha}\log(k)=O(k^{-\alpha}\log(k)).$$

Following the proof of Corollary 3, we have

$$I_3\le(h_0M+h_0\underline\mu^{-1}+\bar p)\left(\frac{2\bar\lambda\sqrt{M}K_6^{1/2}}{\underline\lambda}+\frac{16\bar\lambda^2}{\underline\lambda}M\right)k^{-\alpha}+(\sqrt{MK_2}+8\bar pM)k^{-\alpha}=O(k^{-\alpha}).$$

Therefore, we can conclude that $\forall k\ge 2$, $R_{1,k}\le K'\cdot k^{-\alpha}\log(k)$ with

$$\begin{aligned}K'=\ &\frac{Ah_0}{1-e^{-\gamma}}+2h_0MK_2+\frac{5h_0}{\min(\gamma,\eta)}\left(B(\sqrt{K_2}+\sqrt{K_6})+\left(\frac{2\sqrt{K_2}}{\underline\mu}+\frac{\sqrt{K_6}}{\underline\lambda}\right)\sqrt{M}+(4A+8M)\right)\\ &+(h_0M+h_0\underline\mu^{-1}+\bar p)\left(\frac{2\bar\lambda\sqrt{M}K_6^{1/2}}{\underline\lambda}+\frac{16\bar\lambda^2}{\underline\lambda}M\right)+\sqrt{MK_2}+8\bar pM.\end{aligned}\tag{EC.3}$$

Let $M_0 > 0$ be the upper bound of the regret in the first cycle. Here the constant $M_0 < \infty$ since the decision region $\mathcal{B}$ is bounded and by condition (a), $D_1 \le K_2$ is also bounded. Finally, we conclude that

$$R_1(L) \le M_0 + K' \sum_{k=2}^{L} k^{-\alpha} \log(k) \le K \sum_{k=1}^{L} k^{-\alpha} \log(k).$$

with $K = K' + \frac{2M_0}{\log(2)}$. □

EC.1.6. Convergence Rate of Observed Busy Period

As an analogue of Lemma 2, we prove a uniform convergence rate for the observed busy period X_n, which will be used to bound B_k and $\mathcal{V}_k$ of the gradient estimator (18) that involves terms of X_n^k.

LEMMA EC.1. *Let X_n^1 and X_n^2 be the observed busy period of the two queueing systems coupled as in Lemma 2, with $X_0^1, X_0^2 \le_{st} \frac{\lambda(p)}{\lambda(\bar{p})} \hat{X}_0$ and $W_0^1, W_0^2 \le_{st} \hat{W}_0$.*

1. $|X_n^1 - X_n^2| \le \mathbf{1}_{\{\max(\Gamma_1,\Gamma_2)>n\}} \left(\sum_{k=1}^{n} \tau_k + X_0^1 + X_0^2\right)$.
2. *There exists a constant $K_4 > 0$ such that $|\mathbb{E}[X_n^1 - X_n^2]|^m \le K_4 e^{-0.5\gamma n} n^2$ for all $n \ge 1$ and $m \le 2$.*

Proof of Lemma EC.1 1. Following the argument in Lemma 2, if $W_0^1 \ge W_0^2$, we will have $W_{\Gamma_1}^1 = W_{\Gamma_1}^2 = 0$ and hence $X_{\Gamma_1}^1 = X_{\Gamma_1}^2 = 0$. Since the two systems share the same sequence of arrivals and service times, $X_n^1 = X_n^2$ for all $n \ge \Gamma_1$. Therefore,

$$|X_n^1 - X_n^2| \le \mathbf{1}_{\{\max(\Gamma_1,\Gamma_2)>n\}} |X_n^1 - X_n^2| \le \mathbf{1}_{\{\max(\Gamma_1,\Gamma_2)>n\}} \left(\sum_{k=1}^{n} \tau_k + X_0^1 + X_0^2\right).$$

The last inequality follows from $0 \le X_n^i \le X_0^i + \sum_{k=1}^{n} \tau_k$ for $i = 1, 2$.

2. Following 1 and part 2 of Lemma 2, for $m = 1, 2$,

$$\begin{aligned}\mathbb{E}[|X_n^1 - X_n^2|^m] &\le \mathbb{E}\left[\mathbf{1}_{\{\max(\Gamma_1,\Gamma_2)>n\}} \left(\sum_{k=1}^{n} \tau_k + X_0^1 + X_0^2\right)^m\right] \\ &\le \mathbb{P}(\max(\Gamma_1,\Gamma_2) > n)^{1/2} \mathbb{E}\left[\left(\sum_{k=1}^{n} \tau_k + X_0^1 + X_0^2\right)^{2m}\right]^{1/2}\end{aligned}$$

where

$$\mathbb{P}(\max(\Gamma_1,\Gamma_2) > n) \le e^{-n\gamma} \mathbb{E}[2 + e^{\mu\theta W_0^1} + e^{\mu\theta W_0^2}] \le e^{-n\gamma}(2 + 2M),$$

and

$$\mathbb{E}\left[\left(\sum_{k=1}^{n}\tau_k + X_0^1 + X_0^2\right)^{2m}\right] \le 3^{2m-1}\left(n^{2m}\mathbb{E}\left[\frac{U_1^{2m}}{\lambda(p)^{2m}}\right] + \mathbb{E}[(X_0^1)^{2m}] + \mathbb{E}[(X_0^2)^{2m}]\right).$$

Therefore,

$$\mathbb{E}[|X_n^1 - X_n^2|^m] \le K_4 e^{-0.5n\gamma} n^2,$$

with $K_4 = 3^m\left(\max_{1\le m\le 2}\mathbb{E}[U_1^{2m}]/\lambda(\bar{p})^{2m} + 2\frac{\lambda(p)^{2m}}{\lambda(\bar{p})^{2m}}M\right)^{1/2}(2+2M)^{1/2}$. □

EC.1.7. Proof of Theorem 2

The proof follows an induction-based approach similar to Broadie et al. (2011). For simplicity of notation, we write $\Delta_k = k^{-\beta}$. Let $\mathcal{F}_k$ be the filtration up to cycle k, i.e. including all events in the first $k-1$ cycles. Since $x_{k+1} = \pi_{\mathcal{B}}(x_k - \eta_k H_k)$,

$$\begin{aligned}
&\mathbb{E}\left[\|x_{k+1} - x^*\|^2\right] = \mathbb{E}[\|x_k - x^* - \eta_k H_k\|^2] \\
=\ &\mathbb{E}\left[\|x_k - x^*\|^2 - 2\eta_k H_k\cdot(x_k - x^*) + \eta_k^2 H_k^2\right] \\
=\ &\mathbb{E}\left[\|x_k - x^*\|^2 - 2\eta_k \nabla f(x_k)\cdot(x_k - x^*)\right] - \mathbb{E}[2\eta_k(H_k - f(x_k))\cdot(x_k - x^*)] + \mathbb{E}[\eta_k^2 H_k^2] \\
=\ &(1 - 2\eta_k K_0)\mathbb{E}\left[\|x_k - x^*\|^2\right] + \mathbb{E}[2\eta_k(H_k - f(x_k))\cdot(x^* - x_k)] + \eta_k^2\mathbb{E}[H_k^2].
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{E}[2\eta_k(H_k - \nabla f(x_k))\cdot(x^* - x_k)] = \mathbb{E}[\mathbb{E}[2\eta_k(H_k - \nabla f(x_k))\cdot(x^* - x_k)|\mathcal{F}_k]] \\
=\ &2\eta_k\mathbb{E}[\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\cdot(x^* - x_k)] \le 2\eta_k\mathbb{E}[\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2]^{1/2}\mathbb{E}[\|x^* - x_k\|^2]^{1/2} \\
\le\ &\eta_k\mathbb{E}[\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2]^{1/2}(1 + \mathbb{E}[\|x_k - x^*\|^2]).
\end{aligned}$$

The second last inequality follows from $ab + cd \le \sqrt{a^2+c^2}\sqrt{b^2+d^2}$ and the Holder Inequality, the last inequality follows from $2a \le 1 + a^2$.

Let $b_k = \mathbb{E}[\|x_k - x^*\|^2]$ and recall that $B_k = \mathbb{E}[\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2]^{1/2}$, $\mathcal{V}_k = \mathbb{E}[H_k^2]$. Then, we obtain the recursion

$$b_{k+1} \le (1 - 2K_0\eta_k + \eta_k B_k)b_k + \eta_k B_k + \eta_k^2\mathcal{V}_k.$$

By Condition (b) and (c), we have

$$b_{k+1} \le (1 - 2K_0\eta_k + \eta_k B_k)b_k + \eta_k B_k + \eta_k^2\mathcal{V}_k \le \left(1 - 2K_0\eta_k + \frac{K_0}{8}\eta_k\Delta_k\right)b_k + \frac{K_0}{8}\eta_k\Delta_k + K_3\eta_k\Delta_k.$$

Because step size $\eta_k \to 0$, for k large enough, $\eta_k K_0 \le 1/2$. Let $k_0 = \max\{k : 2\eta_k K_0 > 1\}$. Then, for $k > k_0$, $1 - 2K_0\eta_k + \frac{K_0}{8}\eta_k\Delta_k > 0$. By Condition (a), $\Delta_k/\Delta_{k+1} = (1+\frac{1}{k})^\beta \le 1 + \frac{1}{k} \le 1 + \frac{K_0}{2}\eta_k$, and by the induction assumption $b_k \le C\Delta_k$, for $k > k_0$, we have

$$\begin{aligned} b_{k+1} &\le \left(1 - 2K_0\eta_k + \frac{K_0}{8}\eta_k\Delta_k\right)\left(1 + \frac{K_0\eta_k}{2}\right)C\Delta_{k+1} + \frac{K_0}{8}\eta_k\Delta_k + K_3\eta_k\Delta_k \\ &\le C\Delta_{k+1} - \eta_k\Delta_k\left(\frac{3K_0C}{2} - \frac{K_0C}{8}\Delta_k - \frac{K_0^2C}{16}\eta_k\Delta_k - \frac{K_0}{8} - K_3\right) \end{aligned}$$

Then, we have $b_{k+1} \le C\Delta_{k+1}$ as long as

$$\frac{3K_0C}{2} - \frac{K_0C}{8}\Delta_k - \frac{K_0^2C}{16}\eta_k\Delta_k - \frac{K_0}{8} - K_3 \ge 0. \tag{EC.4}$$

To check (EC.4), note that, $\Delta_k, K_0 \le 1$ and $C \ge 8K_3/K_0$. Besides, $\eta_k K_0 \le 1/2 < 1$ for $k > k_0$. Then, for $k \ge k_0$,

$$\frac{3K_0C}{2} - \frac{K_0C}{8}\Delta_k - \frac{K_0^2C}{16}\eta_k\Delta_k - \frac{K_0}{8} - K_3 \ge \frac{3K_0C}{2} - \frac{K_0C}{8} - \frac{K_0C}{16} - \frac{K_0C}{8} - \frac{K_0C}{8} = \frac{17K_0C}{16} > 0.$$

Let

$$C = \max\left(k_0^\beta(|\bar{\mu} - \underline{\mu}|^2 + |\bar{p} - \underline{p}|^2), 8K_3/K_0\right). \tag{EC.5}$$

Then we have $\|x_k - x^*\|^2 \le C\Delta_k$ for all $1 \le k \le k_0$, and we can conclude by induction, for all $k \ge k_0$,

$$\mathbb{E}[\|x_k - x^*\|^2] \le Ck^{-\beta}.$$

By Assumption 3, there exists $\theta_0 \in [0,1]$ such that

$$|f(x_k) - f(x^*)| = |\nabla f(\theta_0(x^k - x^*) + x^*)^T(x_k - x^*)| \le K_1\|x_k - x^*\|^2.$$

As a consequence,

$$R_2(L) \le \sum_{k=1}^{L} \mathbb{E}[T_k]K_1Ck^{-\beta}.$$

Note that T_k equals to the arrival time of customer D_k plus its waiting time. Therefore,

$$\mathbb{E}[T_k] \le \mathbb{E}\left[\frac{D_k}{\lambda_k}\right] + \mathbb{E}[W^k_{D_k}] \le \frac{D_k}{\lambda(\bar{p})} + M = O(D_k),$$

and we can conclude

$$R_2(L) = O\left(\sum_{k=1}^{L} D_k k^{-\beta}\right).$$

□

EC.1.8. Proof of Theorem 3

(i) For each k, note that $x_k \in \mathcal{F}_k$, let's denote by

$$h_k^1 = -\lambda(p_k) - p_k\lambda'(p_k) + h_0\lambda'(p_k)\left[\frac{1}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(\mathbb{E}[X_n^k|\mathcal{F}_k] + \mathbb{E}[W_n^k|\mathcal{F}_k]\right) + \frac{1}{\mu}\right],$$

$$h_k^2 = c'(\mu_k) - h_0\frac{\lambda(p_k)}{\mu_k}\left[\frac{1}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(\mathbb{E}[X_n^k|\mathcal{F}_k] + \mathbb{E}[W_n^k|\mathcal{F}_k]\right) + \frac{1}{\mu}\right].$$

Then,

$$\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2 = \left|h_k^1 - \frac{\partial}{\partial p}f(\mu_k, p_k)\right|^2 + \left|h_k^2 - \frac{\partial}{\partial \mu}f(\mu_k, p_k)\right|^2.$$

Following (18),

$$|h_k^1 - \frac{\partial}{\partial p}f(\mu_k, p_k)|^2 \le \frac{h_0^2\lambda'(p_k)^2}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(|\mathbb{E}[X_n^k - x_k|\mathcal{F}_k]| + |\mathbb{E}[W_n^k - w_k|\mathcal{F}_k]|\right)^2,$$

$$|h_k^2 - \frac{\partial}{\partial \mu}f(\mu_k, p_k)|^2 \le \frac{h_0^2\lambda(p_k)^2}{\mu_k^2\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(|\mathbb{E}[X_n^k - x_k|\mathcal{F}_k]| + |\mathbb{E}[W_n^k - w_k|\mathcal{F}_k]|\right)^2,$$

where $w_k = \mathbb{E}[W_\infty(\mu_k, p_k)]$ and $x_k = \mathbb{E}[X_\infty(\mu_k, p_k)]$. Note that $\lambda(p)$, $\lambda'(p)$ and μ are bounded. Let $C_0 = \max_{(\mu,p)\in\mathcal{B}}\{h_0\lambda'(p_k), h_0\lambda(p)/\mu\}$, then

$$\begin{aligned}
\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2 &\le \frac{2C_0^2}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(|\mathbb{E}[X_n^k - x_k|\mathcal{F}_k]| + |\mathbb{E}[W_n^k - w_k|\mathcal{F}_k]|\right)^2 \\
&\le \frac{4C_0^2}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(|\mathbb{E}[X_n^k - x_k|\mathcal{F}_k]|^2 + |\mathbb{E}[W_n^k - w_k|\mathcal{F}_k]|^2\right) \\
&= \frac{4C_0^2}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(|\mathbb{E}[X_n^k - \bar{X}_k^n|\mathcal{F}_k]|^2 + |\mathbb{E}[W_n^k - \bar{W}_k^n|\mathcal{F}_k]|^2\right)
\end{aligned}$$

where the last equality follows from $\mathbb{E}[\bar{W}_k^n|\mathcal{F}_k] = w_k$ and $\mathbb{E}[\bar{X}_k^n|\mathcal{F}_k] = x_k$ and $\bar{W}_k^n$ and $\bar{X}_k^n$ are stationary versions of the waiting times and observed busy periods that are synchronously coupled with W_k^n and X_k^n respectively. Therefore, the bias

$$\begin{aligned}
B_k^2 &= \mathbb{E}[\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2] \\
&\le \mathbb{E}\left[\frac{4C_0^2}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(|\mathbb{E}[X_n^k - \bar{X}_k^n|\mathcal{F}_k]|^2 + |\mathbb{E}[X_n^k - \bar{X}_k^n|\mathcal{F}_k]|^2\right)\right] \\
&\le \frac{4C_0^2}{\lceil D_k(1-\xi)\rceil}\sum_{n>\xi D_k}^{D_k}\left(|\mathbb{E}[(X_n^k - \bar{X}_k^n)^2] + \mathbb{E}[(W_n^k - \bar{W}_k^n)^2]\right)
\end{aligned}$$

Following a similar argument as in the proof of Corollary 1, we have, for $n \geq \lceil 0.5\xi D_k \rceil$,

$$\begin{aligned}\mathbb{E}[(W_n^k - \bar{W}_k^n)^2] &\leq \mathbb{E}[(W_n^k - \bar{W}_k^n)^2 \cdot 1(Q_k < 0.5\xi D_k)] + \mathbb{E}[(W_n^k - \bar{W}_k^n)^2 \cdot 1(Q_k \geq 0.5\xi D_k)] \\ &\leq A\exp(-\gamma \cdot (n - 0.5\xi D_k)) + 2M\exp(-\eta \cdot 0.25\xi D_k)) \\ &\leq (A + 2M)\exp(-\min(\gamma,\eta) \cdot 0.25\xi D_k).\end{aligned}$$

For the observed busy period X_n^k, following a similar analysis and Lemma EC.1, we have

$$\begin{aligned}&\mathbb{E}[(X_n^k - \bar{X}_n^k)^2] \\ \leq\ & K_4 e^{-0.5\gamma\xi D_k} D_k^2 + (2\mathbb{E}[(X_n^k)^4] + 2\mathbb{E}[(\bar{X}_k^n)^4])^{1/2}\mathbb{P}(Q_k \geq 0.5\xi D_k)^{1/2} \\ \leq\ & \exp(-\min(\gamma,\eta) \cdot 0.25\xi D_k)(2M + K_4 D_k^2) \leq \exp(-\min(\gamma,\eta) \cdot 0.125\xi D_k)(2M + K_4 K_5),\end{aligned}$$

where

$$K_5 = \max_{D>0} \exp(-\min(\gamma,\eta) \cdot 0.125\xi D) D^2 = \left(\frac{16}{\min(\gamma,\eta) \cdot \xi}\right)^2 e^{-2}. \tag{EC.6}$$

If we choose

$$D_k = a_D + b_D \log(k), \text{ for } a_D \geq \frac{C_D}{\min(\gamma,\eta)\xi} \text{ and } b_D \geq \frac{8}{\min(\gamma,\eta)\xi},$$

with

$$C_D = \max(8(\log((16A + 32M)C_0/K_0), 16\log((32M + 16K_4K_5)C_0/K_0)), \tag{EC.7}$$

then

$$\mathbb{E}[(W_n^k - \bar{W}_n^k)^2] \leq \frac{K_0^2}{256C_0^2k^2}, \quad \mathbb{E}[(X_n^k - \bar{X}_n^k)^2] \leq \frac{K_0^2}{256C_0k^2}.$$

As a consequence,

$$\mathbb{E}[\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2] \leq \frac{4C_0^2}{\lceil D_k(1-\xi)\rceil} \sum_{n>\xi D_k}^{D_k} \left(|\mathbb{E}[(X_n^k - \bar{X}_k^n)^2] + \mathbb{E}[(W_n^k - \bar{W}_k^n)^2]\right) \leq \frac{K_0^2}{64k^2}.$$

Therefore, we can conclude that

$$B_k = \mathbb{E}[\|\mathbb{E}[H_k - \nabla f(x_k)|\mathcal{F}_k]\|^2]^{1/2} \leq \frac{K_0}{8k}.$$

On the other hand, as $\lambda(p)$, $\lambda'(p)$ and μ are bounded, $C_1 \triangleq \max_{\mu,p\in\mathcal{B}}\{|\lambda(p) + p\lambda'(p)|, |c'(\mu)|\} < \infty$. Recall that $C_0 = \max_{(\mu,p)\in\mathcal{B}}\{h_0\lambda'(p_k), h_0\lambda(p)/\mu\}$. Then,

$$\mathbb{E}[\|H_k\|^2] \leq 8(C_1 + C_0/\underline{\mu})^2 + 8C_0^2\mathbb{E}\left[\frac{1}{\lceil(1-\xi)D_k\rceil^2}\left(\sum_{n>\xi D_k}^{D_k}\left(X_n^k + W_n^k\right)\right)^2\right].$$

By Lemma 1, we have

$$\mathbb{E}\left[\frac{1}{\lceil(1-\xi)D_k\rceil^2}\left(\sum_{n>\xi D_k}^{D_k}\left(X_n^k+W_n^k\right)\right)^2\right] \le \mathbb{E}\left[\frac{1}{\lceil(1-\xi)D_k\rceil^2}\left(\sum_{n>\xi D_k}^{D_k}\left(\frac{\lambda(\underline{p})}{\lambda(\bar{p})}\hat{X}_n^k+\hat{W}_n^k\right)\right)^2\right],$$

where $\hat{W}_n^k$ and $\hat{X}_n^k$ are defined as in Lemma 1. Note that by definition, $\hat{W}_n^k$ and $\hat{X}_n^k$ are stationary, we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\lceil(1-\xi)D_k\rceil^2}\left(\sum_{n>\xi D_k}^{D_k}\left(\frac{\lambda(\underline{p})}{\lambda(\bar{p})}\hat{X}_n^k+\hat{W}_n^k\right)\right)^2\right]\\
\le&\frac{2}{\lceil(1-\xi)D_k\rceil^2}\mathbb{E}\left[\left(\sum_{n>\xi D_k}^{D_k}\frac{\lambda(\underline{p})}{\lambda(\bar{p})}\hat{X}_n^k\right)^2\right]+\frac{2}{\lceil(1-\xi)D_k\rceil^2}\mathbb{E}\left[\left(\sum_{n>\xi D_k}^{D_k}\hat{W}_n^k\right)^2\right]\\
\le&2(1-\xi)^{-2}\mathbb{E}\left[\left(\frac{\lambda(\underline{p})}{\lambda(\bar{p})}\hat{X}_0^k\right)^2\right]+2(1-\xi)^{-2}\mathbb{E}[(\hat{W}_0^k)^2]\le 4(1-\xi)^{-2}M.
\end{aligned}$$

Therefore, $\mathcal{V}_k$ is uniformly bounded. Given that $\eta_k = c_\eta k^{-1}$, we have $\eta_k\mathcal{V}_k \le \frac{K_3}{k}$ with

$$K_3 = (8(C_1 + C_0/\underline{\mu})^2 + 32C_0^2(1-\xi)^{-2}M)c_\eta. \tag{EC.8}$$

(ii) According to the update rule, we immediately got

$$\mathbb{E}[\|x_k - x_{k+1}\|^2] \le \eta_k^2\mathbb{E}[\|H_k\|^2] \le 2k^{-2}K_3/K_0 \equiv K_2k^{-2}, \text{ with } K_2 = 2K_3/K_0.$$

(iii) We have just proved that the conditions of Theorem 1 are satisfied with $\alpha = 1$. Therefore, $R_1(L) \le K\sum_{k=1}^{L}k^{-1}\log(k) \le K\log(L)^2$ with the expression of K given in (EC.3). Besides, conditions of Theorem 2 are satisfied with $\beta = 1$ and $D_k = O(\log(k))$. In particular, $\Delta_k/\Delta_{k+1} = 1 + \frac{1}{k} \le 1 + \frac{K_0}{2\eta_k}$ given that $\eta_k = c_\eta k^{-1}$ with $c_\eta \ge 2/K_0$. Therefore,

$$R_2(L) \le CK_1\sum_{k=1}^{L}\left(\frac{D_k}{\lambda(\bar{p})} + M\right)k^{-1} = O(\log(L)^2).$$

As a consequence, the total regret

$$R(L) = R_1(L) + R_2(L) \le K_{alg}\log(L)^2 \le K_{alg}\log(M_L)^2, \text{with } M_L = \sum_{k=1}^{L}D_k.$$

The last inequality uses $\log(L)^2 \le \log(M_L)^2$. Since $M_L = O(L\log(L))$, the relaxation from L to M_L will not change the order of the regret bound. In addition, we can find a closed-expression for K_{alg} as

$$K_{alg} = K + CK_1\cdot\left(\frac{C_D+8}{\lambda(\bar{p})\min(\gamma,\eta)\xi} + M\right), \tag{EC.9}$$

where K is defined by (EC.3), C by (EC.5) and C_D by (EC.7). □

EC.1.9. Details in the Proof of Lemma 5

We first give a rigorous proof of (19) in derivation of the partial derivation $\frac{\partial}{\partial p}\mathbb{E}[W_\infty(p,\mu)]$. To better explain the proof, we adopt the notions in Glasserman (1992). We will take derivative with respective to the parameter $\theta = r = 1/\lambda(p)$. With a slight abuse of notation, we redefine $W_n(\theta) = W_n(\mu, p)$ and $\tilde{U}_n(\theta) = \frac{V_n}{\mu} - \theta U_n$ so that $\tilde{U}_n'(\theta) = -U_n$. And then, the Lindley recursion becomes

$$W_{n+1}(\theta) = \phi(W_n(\theta), \tilde{U}_n(\theta)), \quad \text{with } \phi(w,u) = (w+u)^+.$$

Note that the function ϕ is increasing and convex in w and u. In addition, the derivative process is denote as $V_n(\theta) = Z_n$. Define $\psi_w(w,u) = \psi_u(w,u) = 1(w+u>0)$, such that

$$V_{n+1}(\theta) = \psi_w(W_n(\theta), \tilde{U}_n(\theta))V_n(\theta) + \psi_u(W_n(\theta), \tilde{U}_n(\theta))\tilde{U}_n'(\theta).$$

The stationary versions of the waiting time and derivative process are denoted as $\tilde{W}_0(\theta)$ and $\tilde{V}_0(\theta)$. Then we can check Conditions (B1) to (B3) on page 377 of Glasserman (1992):

(B1) For each $\theta \in [1/\lambda(\underline{p}), 1/\lambda(\bar{p})]$, the sequence

$$\{(\tilde{U}_n(\theta), \tilde{U}_n'(\theta)), -\infty < n < \infty\} = \underbrace{\left\{\left(\frac{V_n}{\mu} - rU_n, -U_n\right), -\infty < n < \infty\right\}}_{\text{in our notation}}$$

is stationary and ergodic, as we can extend the i.i.d. sequences V_n and U_n to $-\infty < n \le 0$.

(B2) For each $\theta \in [1/\lambda(\underline{p}), 1/\lambda(\bar{p})]$, the Lindley recursion has a stationary solution $\tilde{W}_0(\theta)$, which is guaranteed by Assumption 1. Besides, following Lemma 2, for any initial state $W_0(\theta)$, the transient process $W_n(\theta)$ will converge to the stationary version in finite time almost surely.

(B3) For all $\theta \in [1/\lambda(\underline{p}), 1/\lambda(\bar{p})]$,

$$\mathbb{P}(\psi_w(\tilde{W}_0(\theta), \tilde{U}_0(\theta)) = 0) = \underbrace{\mathbb{P}\left(\left(W_\infty(\mu,p) + \frac{V_0}{\mu} - rU_0\right)^+ = 0\right)}_{\text{(in our notation)}} = \mathbb{P}(W_\infty(\mu,p) = 0) > 0.$$

According to the discussion on p.379 of Glasserman (1992), Condition (B3) holds for $GI/GI/1$ queues under the usual stability condition that $\mu > \lambda(p)$. Below, we give a detailed verification of this condition under our model setting.

Recall that $\tilde{U}_0(r) = \frac{V_0}{\mu} - rU_0$ and by Assumption 1, $\mathbb{E}[\tilde{U}_0(r)] < 0,\ \forall r \in [1/\lambda(\underline{p}), 1/\lambda(\bar{p})]$. So there exists a constant $b > 0$, such that $\mathbb{P}(\tilde{U}_0(r) < -b) > 0$ for all $r \in [1/\lambda(\underline{p}), 1/\lambda(\bar{p})]$. Let S denote the support of $W_\infty(\mu, p)$ and let $A = \inf S \geq 0$. We first show by contradiction that $A = 0$. Since A is the infimum of the support,

$$\mathbb{P}\left(W_\infty(\mu,p) \in [A, A+\varepsilon)\right) > 0, \text{ for any } \varepsilon > 0.$$

Besides, if $A > 0$,

$$\begin{aligned}\mathbb{P}(W_\infty(\mu,p) \geq A) &= \mathbb{P}\left(\left(W_\infty(\mu,p) + \tilde{U}_0(r)\right)^+ \geq A\right) \\ &= \mathbb{P}\left(W_\infty(\mu,p) + \tilde{U}_0(r) \geq A\right) = 1,\end{aligned}$$

On the other hand, we have

$$\mathbb{P}\left(W_\infty(\mu,p) + \tilde{U}_0(r) < A\right) \geq \mathbb{P}\left(W_\infty(\mu,p) \in \left[A, A+\frac{b}{2}\right),\ \tilde{U}_0(r) < -b\right) > 0,$$

where the last inequality follows from the fact that $W_\infty(\mu,p)$ and $\tilde{U}_0(r)$ are independent in the $GI/GI/1$ queue. This is a contradiction, so we can conclude that $A = 0$. Next, we show that $\mathbb{P}(W_\infty(\mu,p) = 0) > 0$. Following a similar derivation, we can conclude

$$\mathbb{P}(W_\infty(\mu,p) = 0) = \mathbb{P}\left(\left(W_\infty(\mu,p) + \tilde{U}_0(r)\right)^+ = 0\right) \geq \mathbb{P}\left(W_\infty(\mu,p) \in \left[0, \frac{b}{2}\right),\ \tilde{U}_0(r) < -b\right) > 0.$$

In addition, we have $\mathbb{E}[\tilde{W}_0(\theta)] \leq M$ and $\mathbb{E}[\tilde{V}_0(\theta)] = \mathbb{E}[\tilde{Z}_\infty] = \mathbb{E}[X_\infty(\mu,p)] \leq M$ following Lemma 1. As a consequence, we can prove (19) using the following Corollary 5.3 in Glasserman (1992):

LEMMA EC.2 **(Corollary 5.3 in Glasserman (1992))**. *Suppose that ϕ is increasing and (jointly) convex, and that W_0 and U_0 are almost surely convex. Suppose (B1)-(B3) hold, $\mathbb{E}[\tilde{W}_0(\theta)], \mathbb{E}[\tilde{V}_0(\theta)] < \infty$ for all θ in its range. Then, $\mathbb{E}[\tilde{V}_0(\theta)] = \mathbb{E}[\tilde{W}_0(\theta)]'$ and*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} V_i(\theta) = \mathbb{E}[\tilde{W}(\theta)]', a.s.$$

almost everywhere in the range of θ.

The derivation of $\frac{\partial}{\partial \mu}\mathbb{E}[W_\infty(p,\mu)]$ follows a similar argument with $\tilde{U}(\theta) \equiv V_n - \theta U_n/\lambda(p)$.

EC.2. Relaxing Theoretical Bounds of Hyperparameters

In this section, we conduct numerical experiments to investigate the robustness of GOLiQ's performance to the two main hyperparameters: (i) cycle length D_k, and (ii) step size η_k. We follow two steps:

- First, we calculate the theoretical bounds of these hyperparameters according to (20) and (21).
- Next, we test the algorithm's performance while varying these hyperparameters; we especially consider values that violate their corresponding theoretical bounds.

EC.2.1. Theoretical bounds for η_k and D_k

We follow Section 6.2 by considering the $M/M/1$ example having the objective function in (26) and demand function in (24), with $a = 4.1$, size $n = 10$ and $c_0 = 0.1$. In order to obtain the theoretic bounds for hyper-parameters, we set the region $\mathcal{B} = [6.7, 10] \times [3.7, 5]$ so that $f(\mu, p)$ is strongly convex on $\mathcal{B}$.

Theoretical bound for η_k**.** According to the conditions in Assumption 3, we note that the Hessian matrix of the objective $f(\mu, p)$ has a smallest eigenvalue 0.1231 in the specific region $\mathcal{B}$, which implies that $K_0 = 0.1231$ (and the strong convexity of the objective function on $\mathcal{B}$). Hence, following from (20), the theoretical lower bound for η_k is $c_\eta \geq \tilde{c}_\eta = 2/K_0 = 16.24$.

Theoretical bound for D_k**.** To calculate the lower bounds of a_D and b_D specified in (21), we first estimate C and (γ, η). We set $\xi = 1$. First, according to the expression (EC.7) and $K_0 = 0.1231$, we see that $C \geq 8$. Next, following (3), we select $\min(\gamma, \eta) = 0.011$ which gives the smallest theoretical lower bound.

Hence, (21) requires that $a_D \geq \tilde{a}_D = 8/0.011 = 727$ and $b_D \geq \tilde{b}_D = 8/0.011 = 727$, which leads to a bound for the cycle length $D_k \geq 727 + 727 \log(k)$.

EC.2.2. Robustness to the Theoretical Bounds

Recall that the theoretical bounds in (20) and (21) require that $a_D \geq \tilde{a}_D$, $b_D \geq \tilde{b}_D$ and $c_\eta \geq \tilde{c}_\eta$. We hereby test the criticality of these lower bounds $\tilde{a}_D$, $\tilde{b}_D$ and $\tilde{c}_\eta$ by implementing GOLiQ with $(a_D, b_D, c_\eta) < (\tilde{a}_D, \tilde{b}_D, \tilde{c}_\eta)$. Specifically, in our first experiments, we consider $c_\eta = \{2, 1, 0.5, 0.1\}\tilde{c}_\eta$ for the step-size η_k, with $D_k = 10 + 10 \log(k)$ (see left-hand panels of Figure EC.1); in our second experiment, we consider $(a_D, b_D) =$

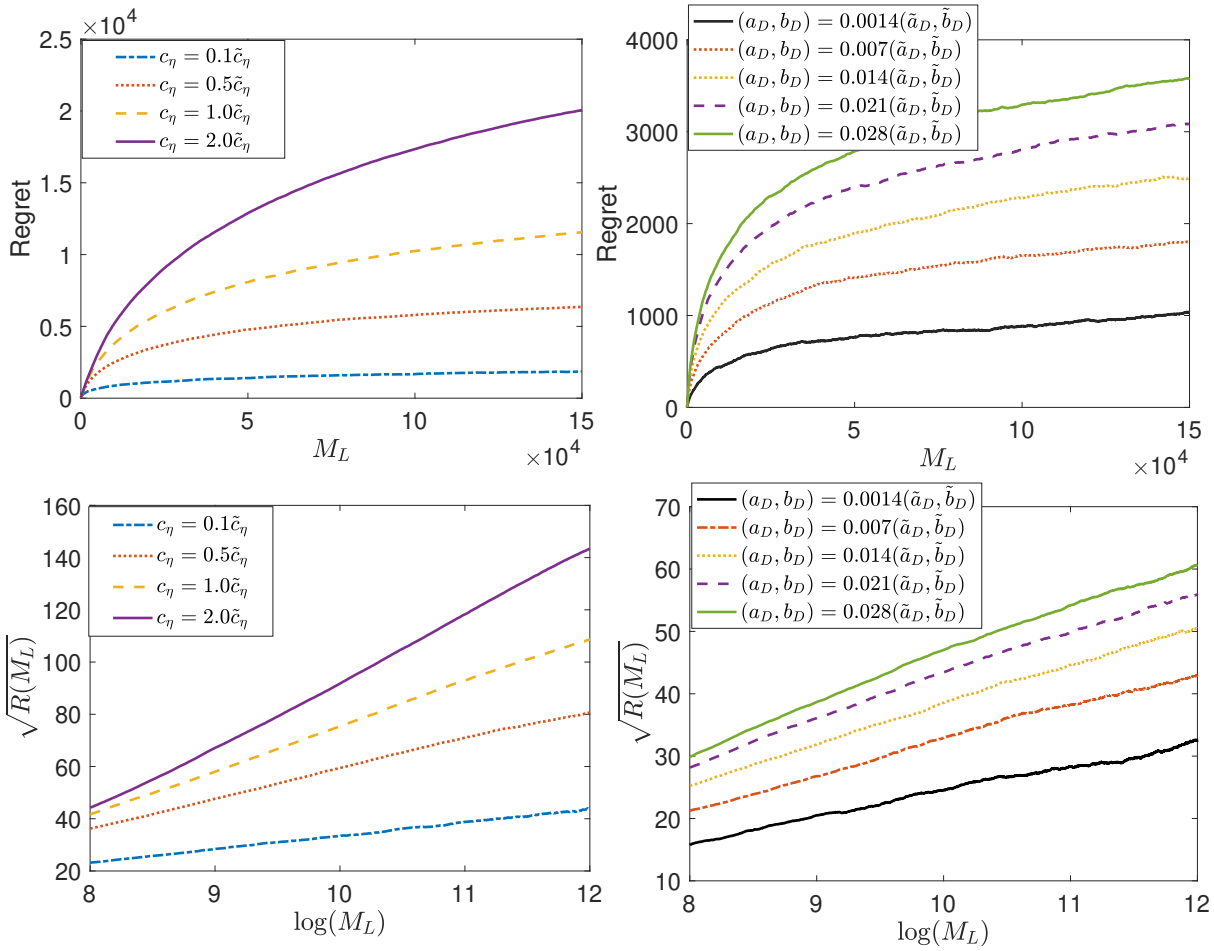

Figure EC.1 Simulated regret curves with relaxed bounds on (i) step size η_k robustness (top left panel), (ii) cycle length D_k (top right panel), and their logarithmic scales in sample size M_L (two bottom panels). All regret curves are estimated by averaging 500 independent simulation runs.

$\{0.028, 0.021, 0.014, 0.007, 0.0014\}(\tilde{a}_D, \tilde{b}_D)$ for the sample-size D_k, with $\eta_k = 3/k$ (see right-hand panels of Figure EC.1). In both experiments, we plot the average regret curves estimated by 500 independent runs.

Figure EC.1 reveals that GOLiQ continues to perform effectively even when the hyperparameters are chosen to be much smaller than their corresponding theoretical lower bounds. For η_k, our algorithm generates a logarithm regret even when $c_\eta = 0.1\tilde{c}_\eta$. (However, we discover that GOLiQ will fail to converge and yield a linear regret if we keep reducing c_η (e.g., to $0.01\tilde{c}_\eta$). For D_k, all regret curves exhibit a logarithmic order (even when $(a_D, b_D) = 0.0014(\tilde{a}_D, \tilde{b}_D)$). In summary, our numerical experiments show that the theoretical bounds for our hyperparameters do not seem to be too restrictive. In addition, the experiment in Section 6.3 serves as another piece of evidence supporting the robustness of GOLiQ. In Section 6.3, we apply GOLiQ with the same hyperparameters $\eta_k = 5k^{-1}$ and

$D_k = 10 + 10\log(k)$ for different settings with various c, c_s^2 and n (see Figure 7), and GOLiQ exhibits stable performance with similar logarithm regrets. Of course, we acknowledge that the specific selection of these hyperparameters in a practical setting will require further tuning in order to make the most efficient use of GOLiQ.

REMARK EC.1 (REQUIREMENT OF INFORMATION: ONLINE LEARNING VS. HEAVY TRAFFIC). We provide our view on how online learning relies on the system information, and we treat heavy-traffic methods as a benchmark. First, online learning in general requires less prior information of the distributions than heavy-traffic methods do. For example, to solve the problem in the present study, the diffusion limit in Lee and Ward (2014) requires the knowledge of the exact values of the second moments of arrival and service times. On the other hand, even though the efficiency of GOLiQ is subject to constraints in terms of certain model parameters, the bounds of these constraints may be relaxed without needing to sacrifice much of the algorithm's performance. Second, the required information (e.g., moments) serves as crucial input parameters for the heavy-traffic models, whereas the design and implementation of online learning algorithms do not immediately require the aforementioned information (even though it is still relevant to the tuning of hyperparameters). All that we require is that the constants in (20) and (21) are **not too small**. So as long as we follow the structure specified in (20)–(21), it will not be too difficult to find reasonably sound hyperparameters (e.g., by a trial-and-error search) even without precise information of parameters η and γ as in Assumption 2. However, trial-and-error will be ineffective for heavy-traffic methods because precise information is needed (e.g., σ^2). In this sense, online learning depends on the system information to a lesser extent.

EC.3. Selecting Hyperparameters according to Huh et al. (2009)

Huh et al. (2009) develops an online learning algorithm with the objective of finding the optimal base-stock policy for an inventory system with a non-zero replenishment lead time. At a glance, Huh et al. (2009) does not seem to be relevant to the present paper at all. Indeed, results in Huh et al. (2009) are by no means directly comparable to GOLiQ, because the two articles consider two different systems. Nevertheless, the fundamental idea in the regret analysis by Huh et al. (2009) may be used as a basis to devise a queueing-version algorithm. We give some specific reasons: First, Huh et al. (2009) analyzes the transient

regret bound of an inventory system operated under a stationary base-stock policy, of which the main framework is analogous to that in the present work. Second, the heart of the online learning algorithm in Huh et al. (2009) is an SGD method. Last, the regret in Huh et al. (2009) is also defined using the steady-state performance as the benchmark.

According to their regret analysis, Huh et al. (2009) propose to choose the hyperparameters $\eta_k = O(k^{-1/2})$ and $D_k = O(\sqrt{k})$ which yield a regret bound in the order $O(M_L^{2/3})$. However, we point out that the objective function in Huh et al. (2009) is convex while GOLiQ in the present paper is designed assuming a strongly convex objective function (Assumption 3). Therefore, to make a fair comparison between GOLiQ and the online learning algorithm proposed in Huh et al. (2009), we need to redo the regret analysis in Huh et al. (2009) under the *strong convexity*. This change, as we will show below, will yield a different set of hyperparameters.

Suppose we select $D_k = O(k^\alpha)$ and $\eta_k = O(k^{-\beta})$. Then, following Lemma 11 of Huh et al. (2009), $R_1(L) = O(L)$ (compared to $R_1(L) = o(L)$ in our analysis). Given that the objective function is strongly convex, Theorem 2 yields that $R_2(L) = O\left(\sum_{k=1}^{L} k^\alpha \cdot k^{-\beta}\right) = O\left(L^{\alpha-\beta+1}\right)$ for $\beta \in (0,1]$. As a result, the overall regret is

$$R(L) = R_1(L) + R_2(L) = O(L^{(\alpha-\beta+1)\vee 1}), \quad \text{and} \quad R(M_L) = O\left(M_L^{\frac{(\alpha-\beta+1)\vee 1}{\alpha+1}}\right),$$

with $M_L = O(L^{\alpha+1})$. Consequently, the order of $R(M_L)$ is minimized by setting

$$\eta_k = O(k^{-1}), \quad \text{and} \quad D_k = O(k), \tag{EC.10}$$

which yields an improved regret bound $O(M_L^{\frac{1}{2}})$ (as opposed to the previous regret $O(M_L^{\frac{2}{3}})$ under regular convexity).

We refer to Algorithm 1 with η_k and D_k selected according to (EC.10) as GOLiQ-H. To compare GOLiQ with GOLiQ-H, we follow the setting in Section 6.2 by considering an $M/M/1$ queue, with $c(\mu) = 0.1\mu^2$ and $\lambda(p) = 10\lambda_0(p)$.

In Figure EC.2, we plot the average regret curves (estimated by averaging 500 independent paths) for both GOLiQ and GOLiQ-H. The hyperparameters are $\eta_k = 2k^{-1}$ and $D_k = 10 + 10\log(k)$ for GOLiQ, and $\eta_k = \{2,4\}k^{-1}$ and $D_k = 10 + k$ for GOLiQ-H. Unsurprisingly, Figure EC.2 confirms that GOLiQ is more effective than GOLiQ-H. This is consistent with our theoretical analysis because GOLiQ yields a logarithmic regret while GOLi-H has a regret bound of $O(M_L^{\frac{1}{2}})$.

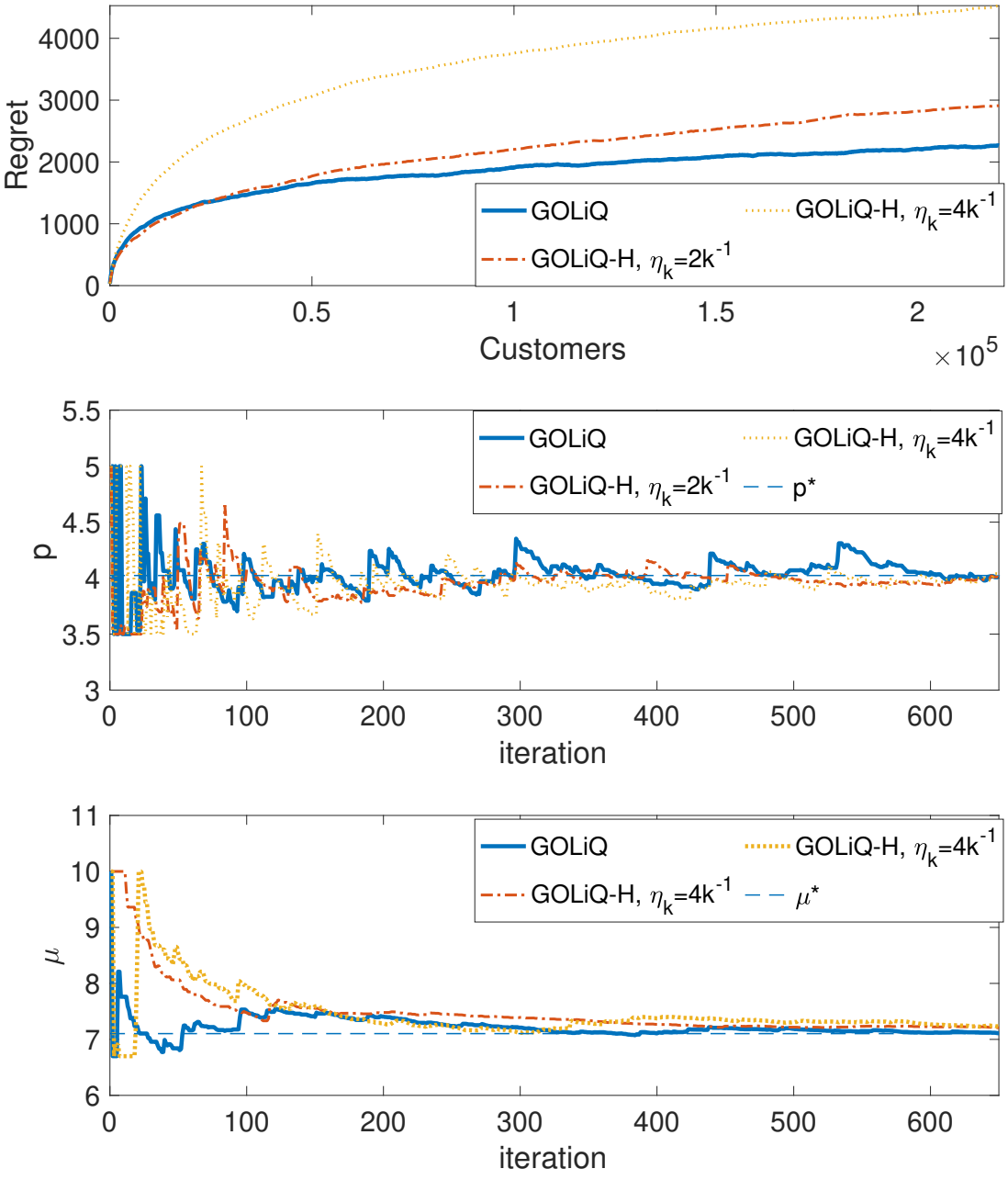

Figure EC.2 Comparing GOLiQ and GOLiQ-H: (i) regret curves (top panel), (ii) trajectory of price p_k (middle panel), and (iii) trajectory of service rate μ_k (bottom panel). Hyperparameters are $\eta_k = \{2, 4\}k^{-1}$ and $D_k = 10 + k$ for GOLiQ-H, and are $\eta_k = 2k^{-1}$ and $D_k = 10 + 10\log(k)$ for GOLiQ. All regret curves are estimated by averaging 500 independent simulation runs.

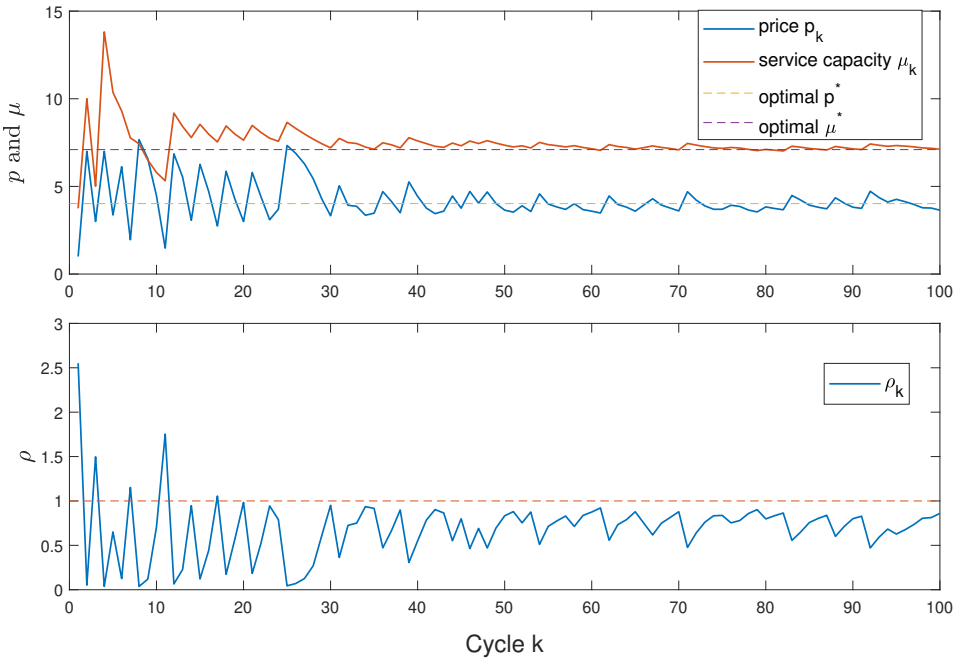

Figure EC.3 Joint pricing and staffing for the $M/M/1$ model in Section 6.2 without uniform stability.

EC.4. Additional Numerical Examples

In this section we conduct additional numerical experiments to confirm the practical effectiveness of our algorithm. In what follows, we first test the case where the uniform stability condition is relaxed; we next report the algorithm performance for $GI/GI/1$ queueing models with phase-type and lognormal distributions.

EC.4.1. Violation of Uniform Stability

We extend the $M/M/1$ example considered in Section 6.2 with the uniform stability condition relaxed. Specifically, we begin with an initial setting of (p_0, μ_0) such that $\rho_0 \equiv \lambda(p_0)/\mu_0 = 2.55$, which violates the stability condition. As shown in Figure EC.3, the pricing and staffing policies (p_k, μ_k) remain convergent to (p^*, μ^*). Consistently, the resulting traffic intensity $\rho_k \equiv \lambda(p_k)/\mu_k$ is quickly controlled to fall below 1; that is, the workload is kept in check despite of the unstable performance in the initial cycle.

EC.4.2. $M/G/1$ with Phase-Type Service

To test the performance of our online learning algorithm for queues with non-exponential service times, we consider phase-type distributions: hyperexponential with n phases (H_n) and Erlang with n phases (E_n). In Figure EC.4 we report the convergent sequence (p_k, μ_k)

with H_2 service with service-time SCV $c_s^2 = 8$ (top panel), M service with $c_s^2 = 1$ (middle panel), and E_8 service with $c_s^2 = 1/8$ (bottom panel). Other parameters include the step size $\eta_k = 4/k$, cycle length $D_k = 20 + 10\log(k)$ and initial condition $p_0 = 4$ and $\mu_0 = 12$ ($\lambda_0 = 5.249$).

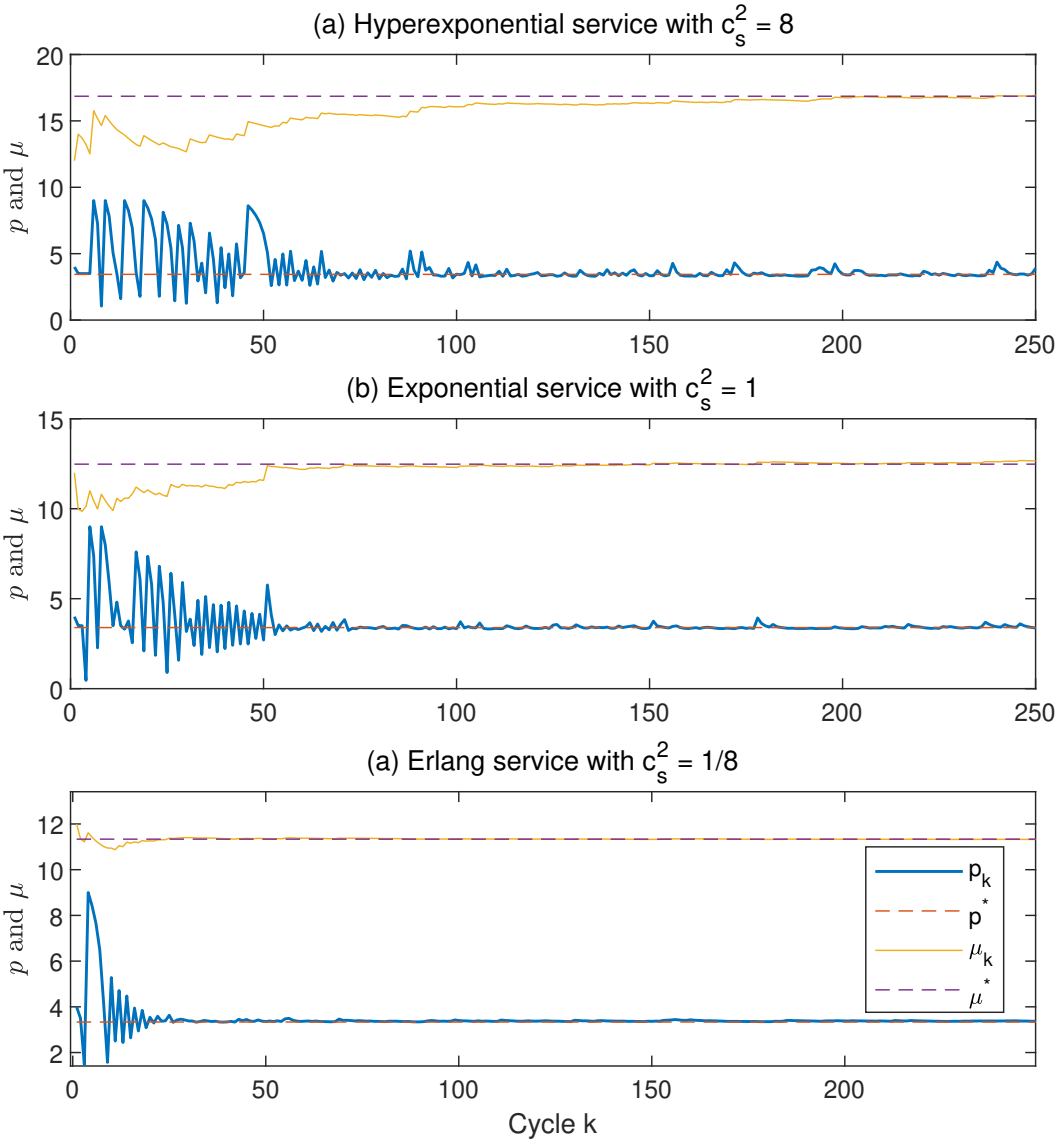

Figure EC.4 Joint pricing and staffing for an $M/G/1$ queue having (a) H_2 service with $c_s^2 = 8$ (top panel), (b) M service (middle panel), and (c) Erlang service with $c_s^2 = 1/8$ (bottom panel). Other parameters are step length $\eta_k = 4/k$, cycle length $D_k = 20 + 10\log(k)$, initial condition $p_0 = 4$, $\mu_0 = 12$. The optimal pricing and staffing solutions are: (i) $(p^*, \mu^*) = (3.44, 16.86)$; (ii) $(p^*, \mu^*) = 3.40, 12.48)$; (iii) $(p^*, \mu^*) = 3.38, 11.34)$.

Figure EC.4 confirms that our algorithm remains effective. In addition, the convergence is faster as the CSV c_s^2 decreases. This is intuitive because a less variable service-time distribution yields a smaller $\mathcal{V}_k$ for the gradient estimator.

EC.4.3. $GI/GI/1$ Examples

We consider an $LN/LN/1$ queue with service and interarrival times following lognormal (LN) distributions. Our consideration here follows from the recent empirical confirmations of LN distributed service times in real service systems.

We let $c_s^2 = c_a^2 = 2$ with c_a^2 being the SCV of the LN-distributed interarrival times. The other parameters remain the same as in Section EC.4.2. Because the exact optimal solutions (p^*, μ^*) are unavailable for this model, we are unable to provide an estimate of the regret as done in Figure 6, nor can we confirm the convex structure of the problem. Nevertheless, Figure EC.5 shows that our online algorithm continues to work well, despite the fact that LN is no longer a light-tail distribution (Assumption 2 does not hold in this case).

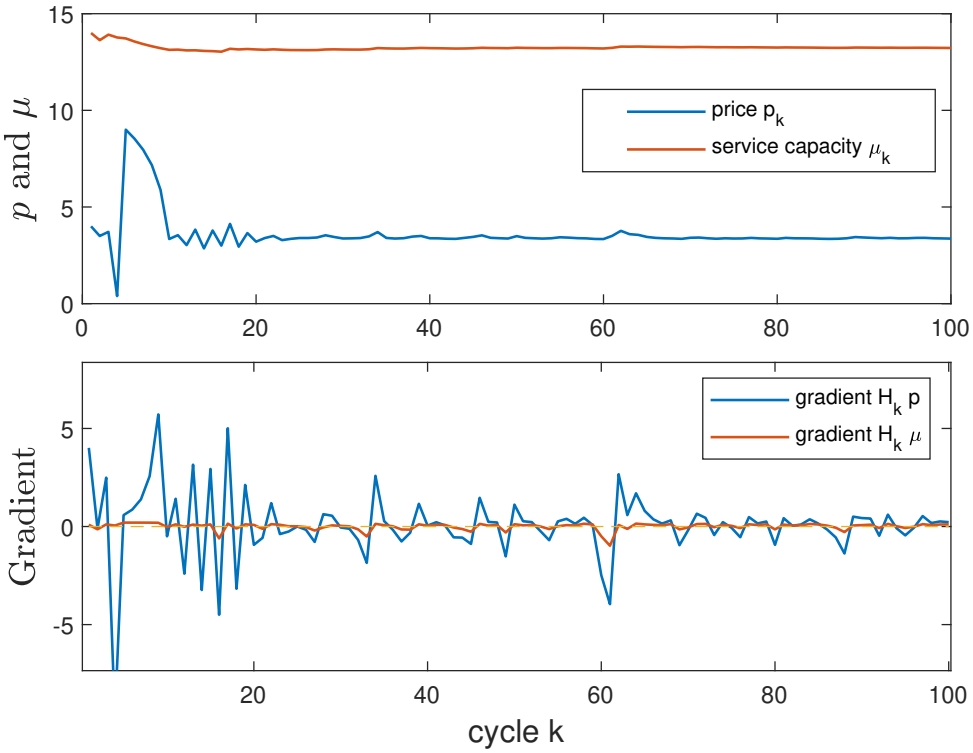

Figure EC.5 Joint online pricing and staffing for an $LN/LN/1$ having lognormal service and interarrival times with CSVs $c_s^2 = c_a^2 = 2$. Other parameters are $\eta_k = 4/k$, $D_k = 20 + 10\log(k)$, $p_0 = 4, \mu_0 = 14$.

EC.4.4. Extended Comparison of GOLiQ and Heavy-traffic Methods

Supplementing our investigations in Section 6.3, we provide additional numerical results. Recall that the heavy-traffic results in Lee and Ward (2014) are obtained by constructing a sequence of $GI/GI/1$ models indexed by n, where the n^{th} model has scaled arrival rate $\lambda_n(p) = n\lambda(p)$ and service rate $\mu_n = n\mu$, so that both λ_n and μ_n grow to ∞ as n increases. Lee and Ward (2014) develop asymptotic staffing and pricing solutions for the $GI/GI/1$ queue; they show that, as the scaling factor $n \to \infty$, the optimal price $p_n^* \to p_\infty$ and service capacity $\mu_n^*/n \to \mu_\infty$, with $\rho_\infty \equiv \lambda(p_\infty)/\mu_\infty = 1$.

We repeat our experiment in Section 6.2 with the scaling parameter $n \in \{10, 50, 100, 500, 1000, 2000\}$ for the arrival rate function (24). But we now focus on the optimal traffic intensity as n varies. In Figure EC.6 we plot the optimal price and service rate as n increases. In each experiment, we compute the optimal p_n and μ_n using their average value in cycles 300–500 of Algorithm 1. Consistent with Lee and Ward (2014), Figure EC.6 shows that p_n, μ_n/n and ρ_n approach p_∞, μ_∞ and $\rho_\infty = 1$. On the other hand, when the scale n is not very large, the heavy-traffic solutions can become inaccurate. For instance, when $n = 100$ the optimal traffic ρ_{100} is around 0.8, which is not close to 1.

EC.4.5. Alternative Definition of Regret

In this subsection, we attempt to rationalize our regret definition in (9). We consider a potential alternative to (9) which benchmarks the system revenue under GOLiQ with the nonstationary revenue under (μ^*, p^*). Because the nonstationary queue length is intractable, we conduct additional numerical experiments to estimate the expected nonstationary regret via Monte-Carlo simulations.

Specifically, we simulate the regret in (10) under (p^*, μ^*) with the queueing system starting empty (of which the dynamics is nonstationary). We use the $M/M/1$ model in Section 6 having a logit demand function (24) with $n = 10$ and a quadratic staffing cost in (25) with $c = 0.1$.

In Figure EC.7 we graph both versions of the regret under GOLiQ under the same experimental setting, with hyperparameters $\eta_k = 3k^{-1}$ and $D_k = 10 + 10\log(k)$. Figure EC.7 confirms that the these two versions of regret appear to be nearly indistinguishable. This is due to the geometric ergodicity of $G/G/1$ queue.

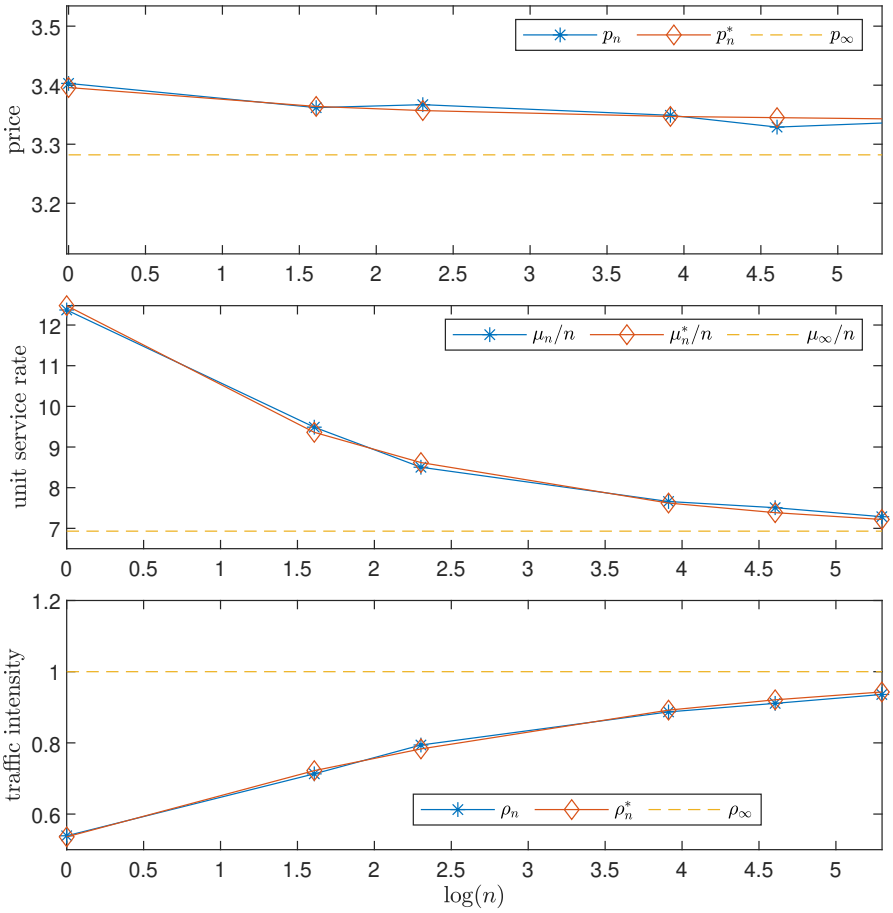

Figure EC.6 Comparison of (i) online optimization solutions (p_n, μ_n, ρ_n); (ii) exact solutions $(p_n^*, \mu_n^*, \rho_n^*)$; and (iii) heavy-traffic solutions $(p_\infty, \mu_\infty, \rho_\infty) = (3.282, 6.932, 1)$ in Lee and Ward (2014), as the system scale $n = M_0$ increases, with parameters $D_k = n(10 + 10\log(k))$ and $\eta_k = 3k^{-1}$.

EC.5. Additional discussion on Assumption 3

In this section, we first provide some sufficient conditions for strong convexity in the $M/GI/1$ case.

LEMMA EC.3. *For $M/GI/1$ queues, if $c(\mu)$ is convex, $\lambda(\underline{p})/\underline{\mu} < 1$, $\lambda''(p)$ is continuous, and in addition,*

$$\frac{\lambda'(p)^2}{2\lambda(p)} < \lambda''(p) < \frac{-2\lambda'(p)}{p}, \quad \text{and} \quad \frac{\lambda(\bar{p})}{\bar{\mu}} > 1 - 1/\sqrt{2} \approx 0.29, \tag{EC.11}$$

then $f(\mu, p)$ is strongly convex in $\mathcal{B}$.

Proof of Lemma EC.3 Recall that $f(\mu,p) = -p\lambda(p) + h_0\mathbb{E}[Q_\infty(\mu,p)] + c(\mu)$, and $(-p\lambda(p))'' = -p\lambda'' - 2\lambda'$. Under condition (EC.11), we have $-p\lambda'' - 2\lambda' > 0$. Therefore, both

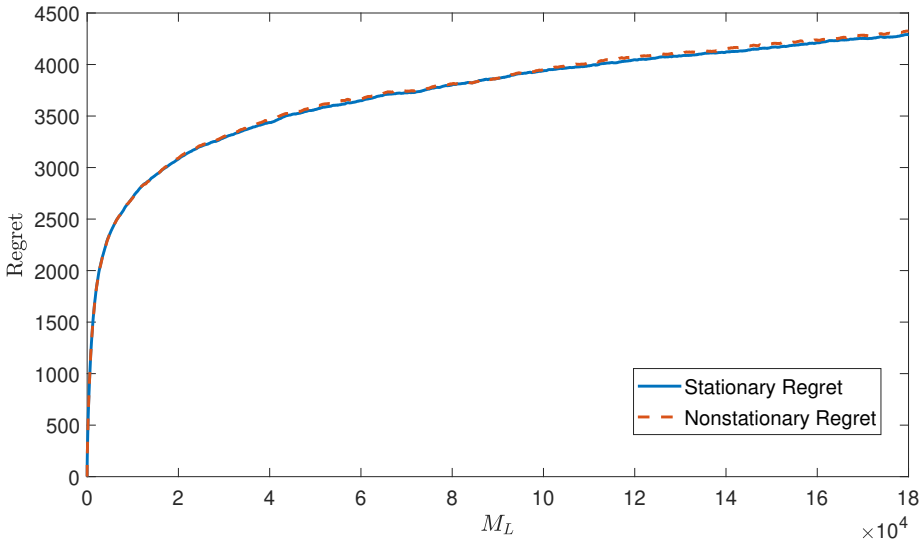

Figure EC.7 Comparing two versions of regret: (i) "stationary regret" that benchmarks with the steady-state performance under (μ^*, p^*) (defined in (9)), and (ii) "nonstationary regret" that benchmarks with the nonstationary performance under (μ^*, p^*). Hyperparameters are $\eta_k = 3k^{-1}$, $D_k = 10 + 10\log(k)$. Both regret curves are estimated by 1,000 independent runs.

$-p\lambda(p)$ and $c(\mu)$ are convex, and it suffices to show that $\mathbb{E}[Q_\infty(\mu, p)]$ is strongly convex in μ and p. For $M/GI/1$ queues, Pollaczek-Khinchine formula yields that

$$q(\mu, p) \equiv \mathbb{E}[Q_\infty(\mu, p)] = C\frac{\lambda(p)}{\mu - \lambda(p)} + (1 - C)\frac{\lambda(p)}{\mu},$$

with $C \equiv \frac{1+c_s^2}{2}$. For any given pair of (μ, p), let H_q be the Hessian matrix of $q(\mu, p)$. We next verify that H_q is positively definite. By direct calculation, we have

$$\partial_p^2 q = C\frac{\mu}{(\mu-\lambda)^3}\left(2(\lambda')^2 + (\mu-\lambda)\lambda''\right) + (1-C)\frac{\lambda''}{\mu},$$
$$\partial_\mu^2 q = C\frac{2\lambda}{(\mu-\lambda)^3} + (1-C)\frac{2\lambda}{\mu^3}, \quad \partial_p\partial_\mu q = -C\frac{\lambda+\mu}{(\mu-\lambda)^3}\lambda' - (1-C)\frac{\lambda'}{\mu^2},$$

with λ', λ'' being the first and second order derivatives of $\lambda(p)$. As a result, the determinant of Hessian matrix of H_q is

$$|H_q| = \frac{C^2}{(\mu-\lambda)^5}\left(2\mu\lambda\lambda'' - (\mu-\lambda)(\lambda')^2\right) + \frac{(1-C)^2}{\mu^4}\left(2\lambda\lambda'' - (\lambda')^2\right) + \frac{2C(1-C)}{\mu^2(\mu-\lambda)^3}\left((2\mu-\lambda)\lambda\lambda'' - (\mu-\lambda)(\lambda')^2\right).$$

To show that H_q is positively definite, it suffices to show that $\partial_\mu^2 q$, $\partial_p^2 q$ and $|H_q|$ are all positive. First, it is clear that

$$\partial_\mu^2 q = 2\lambda C\left(\frac{1}{(\mu-\lambda)^3} - \frac{1}{\mu^3}\right) + \frac{2\lambda}{\mu^3} > 0.$$

Next, we compute

$$\begin{aligned}\partial_p^2 q &= C\frac{\mu}{(\mu-\lambda)^3}\left(2(\lambda')^2+(\mu-\lambda)\lambda''\right)+(1-C)\frac{\lambda''}{\mu}\\ &= \frac{2C\mu}{(\mu-\lambda)^3}(\lambda')^2+\frac{C\mu}{(\mu-\lambda)^2}\lambda''+(1-C)\frac{\lambda''}{\mu}\\ &\overset{(a)}{>} \frac{C\mu}{(\mu-\lambda)^2}\lambda''+(1-C)\frac{\lambda''}{\mu}\\ &= C\lambda''\Big(\underbrace{\frac{\mu}{\mu-\lambda}}_{>1}\cdot\frac{1}{\mu-\lambda}-\frac{1}{\mu}\Big)+\frac{\lambda''}{\mu}\overset{(b)}{>} C\lambda''\left(\frac{1}{\mu-\lambda}-\frac{1}{\mu}\right)+\frac{\lambda''}{\mu}>0.\end{aligned}$$

Here, inequality (a) follows from that $\frac{2C\mu}{(\mu-\lambda)^3}(\lambda')^2>0$. Inequality (b) holds due to the facts that $\frac{\mu}{\mu-\lambda}>1$ and that $\lambda''>\frac{(\lambda')^2}{2\lambda}\geq 0$. The last inequality holds because $\frac{1}{\mu-\lambda}>\frac{1}{\mu}$. As a result, we have $\partial_p^2 q, \partial_\mu^2 q>0$. Next, we verify that $|H_q|>0$. Because $2\lambda\lambda''-(\lambda')^2>0$, we have

$$2\mu\lambda\lambda''>(\mu-\lambda)(\lambda')^2, \quad \text{and} \quad (2\mu-\lambda)\lambda\lambda''>(\mu-\lambda)(\lambda')^2.$$

Therefore,

$$\begin{aligned}|H_q| = &\frac{C^2}{(\mu-\lambda)^5}\left(2\mu\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)+\frac{(1-C)^2}{\mu^4}\left(2\lambda\lambda''-(\lambda')^2\right)\\ &+\frac{2C(1-C)}{\mu^2(\mu-\lambda)^3}\left((2\mu-\lambda)\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)\\ \overset{(c)}{>}&\frac{C^2}{(\mu-\lambda)^5}\left(2\mu\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)+\frac{2C(1-C)}{\mu^2(\mu-\lambda)^3}\left((2\mu-\lambda)\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)\\ >&\frac{C^2}{(\mu-\lambda)^5}\Big(\underbrace{2\mu\lambda\lambda''}_{>(2\mu-\lambda)\lambda\lambda''}-(\mu-\lambda)(\lambda')^2\Big)-\frac{2C^2}{\mu^2(\mu-\lambda)^3}\left((2\mu-\lambda)\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)\\ >&\frac{C^2}{(\mu-\lambda)^5}\left((2\mu-\lambda)\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)-\frac{2C^2}{\mu^2(\mu-\lambda)^3}\left((2\mu-\lambda)\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)\\ =&\frac{C^2\left((2\mu-\lambda)\lambda\lambda''-(\mu-\lambda)(\lambda')^2\right)}{(\mu-\lambda)^3}\left(\frac{1}{(\mu-\lambda)^2}-\frac{2}{\mu^2}\right)>0.\end{aligned}$$

Inequality (c) holds because $\frac{(1-C)^2}{\mu^4}(2\lambda\lambda''-(\lambda')^2)$ is positive by (EC.11). The last inequality follows from $\frac{\lambda(p)}{\mu}>1-1/\sqrt{2}$. Therefore, we have, for any pair $(\mu,p)\in\mathcal{B}$, $\nabla^2 f(\mu,p)$ is positive-definite. Because each component of $\nabla^2 f(\mu,p)$ is continuous with respect to (μ,p), and hence the smallest eigenvalue of $\nabla^2 f(\mu,p)$ is also continuous with respect to (μ,p) (See Appendix D of Horn and Johnson (2012)). Consequently, given that $\mathcal{B}$ is a compact region, there exists an $\varepsilon>0$ such that $\nabla^2 f(\mu,p)-\varepsilon I$ is positive-definite. This shows that $f(\mu,p)$ is strongly convex in $\mathcal{B}$ (See Appendix B.5 in Bertsekas (1997)). □

	Notation	Description
Model parameters and functions	$\mathcal{B}=[\underline{p},\bar{p}]\times[\underline{\mu},\bar{\mu}]$	Feasible action space
	$c(\mu)$	Staffing cost function
	$f(\mu,p)$	Objective (loss) function
	h_0	Customer holding cost
	$\lambda(p)$	Demand function
	μ	Service rate/capacity
	n	Market size/ System scale in Section 6
	p	Service fee
	$Q_\infty(\mu,p)$	Stationary queue length under (p,μ)
	S_n^k	Service time of the $(n-1)^{\text{th}}$ customer in cycle k
	τ_n	Interarrival time between $(n-1)^{\text{th}}$ and n^{th} customers
	θ,γ,η	Parameters of light-tail assumptions (Assumption 2)
	U_n, V_n	Unscaled random "seeds" of interarrival and service times
	$x^*=(p^*,\mu^*)$	Optimal decision fee and service rate
Algorithmic hyperparameters	D_k	Sample size (number of customers served) in cycle k
	η_k	Step size or learning rate in cycle k
	H_k	Gradient estimator in cycle k
	M_L	Cumulative number of customers served by cycle L
	Q_k	Queue content leftover from cycle $k-1$
	W_n^k	Delay of the n^{th} customer in cycle k
	ξ	Warm up rate
	X_n^k (X_n)	Server's busy time observed by customer n in cycle k
Constants and bounds in regret analysis	a_D, b_D	Constants for D_k in equation (21)
	$A=4\sqrt{M}+4M$	Constant in Corollary 1
	B	Constant of stationary waiting times in Lemma 4
	$B_k,\ \mathcal{V}_k$	Upper bounds for bias and Variance of H_k
	c_η	Constant for η_k in equation (20)
	c_μ, c_λ	Constants in Lemma 3
	$C=\max\{\|x_0-x^*\|^2, 8K_3/K_0\}$	Constant in Theorem 2
	$C_0=\max_{x\in\mathcal{B}}\{h_0\lambda'(p), h_0\lambda(p)/\mu\}$	Constant in the proof of Theorem 3
	$C_1=\max_{x\in\mathcal{B}}\{\lvert\lambda(p)+p\lambda'(p)\rvert, \lvert c'(\mu)\rvert\}$	Constant in the proof of Theorem 3
	C_D	Constant for the selection of D_k (Theorem 3)
	$d_k=\lceil 4\log(k)/\min(\theta,\gamma)\rceil$	Constant of warm-up time (Theorem 1)
	$\tilde{d}_k=\lceil 5\log(k)/\min(\theta,\gamma)\rceil$	Constant of warm-up time (Theorem 1)
	$\Gamma_i,\ i=1,2$	Stopping time of random walks (proof of Lemma 2)
	I_1, I_2, I_3	Three terms of the regret of nonstationary
	K_{alg}	Bound of the cumulative regret (Theorem 3)
	K'	Constant for regret of nonstationary (proof of Theorem 1)
	$K=K'+2M_0/\log(2)$	Constant in the proof of Theorem 1
	K_0, K_1	Constants for convexity and smoothness (Assumption 1)
	$K_2=2K_3/K_0$	Constant for D_k (Theorem 1)
	K_3	Constants for variance in Theorem 3 (EC.8)
	K_4	Constants for convergence rate of busy time in Lemma EC.1
	$K_5=32e^4/(\min(\theta,\gamma))$	Bound in the proof of Theorem 3 (EC.6)
	$K_6=K_2\max_p\lambda'(p)$	Bound in the proof of Corollary 2
	$\bar{\lambda},\underline{\lambda}$	Upper and lower bounds for $\lambda(p)$
	M	Uniform bound for queueing functions in Lemma 1
	M_0	Upper bound of the regret in the first cycle
	R_k	Total regret during cycle k
	$R_{1,k}, R_{2,k}$	Regret of nonstationary/suboptimality in cycle k
	$R(L), R_1(L), R_2(L)$	Cumulative/nonstationary/suboptimal regret by cycle L
	T_k	Length of cycle k

Table EC.1 Glossary of notation.