Adv. Appl. Probab. **55,** 1075–1083 (2023)
*doi:*10.1017/apr.2022.59

AN ASYMPTOTICALLY OPTIMAL HEURISTIC FOR GENERAL NONSTATIONARY FINITE-HORIZON RESTLESS MULTI-ARMED, MULTI-ACTION BANDITS: CORRIGENDUM

GABRIEL ZAYAS-CABÁN,* *University of Wisconsin-Madison*
JIAXIN LIANG,** *University of Michigan*
STEFANUS JASIN,*** *University of Michigan*
GUIHUA WANG,**** *University of Texas at Dallas*

2020 Mathematics Subject Classification: Primary 90C40
Secondary 68M20; 90B36

1. Correction to Lemma 2 in Zayas-Cabán *et al.* (2019) [1]

In our original submission (Zayas-Cabán *et al.*, 2019) [1], we have the following lemma.

Lemma 2 in [1]. *There exists a constant $M > 0$, independent of T, and a vector $\boldsymbol{\epsilon} \geq \mathbf{0}$ satisfying $\epsilon_t \leq b_t$ for all t, such that*

$$V^D(\boldsymbol{\epsilon}) - V^D(\mathbf{0}) \leq M \cdot \left[\sum_{t=1}^{T} \epsilon_t\right] = O\left(\sum_{t=1}^{T} \epsilon_t\right). \tag{1}$$

The above lemma is used to prove Theorems 1–2 and Propositions 1–3 in Sections 4 and 6 of [1]. It has been graciously pointed out to us that the bound in the lemma may not be correct in general. The original proof of this lemma uses a combination of linear program (LP) duality and sensitivity analysis results. The mistake is in the application of a known sensitivity analysis result under a certain assumption that happens to be not necessarily satisfied by our LP. Fortunately, it is possible to correct the bound in the above lemma. The new bound that we will prove in this correction note is as follows:

$$V^D(\boldsymbol{\epsilon}) - V^D(\mathbf{0}) = O\left(T \cdot \max_t \epsilon_t\right). \tag{2}$$

Received 11 December 2020; revision received 18 August 2022.
* Postal address: Mechanical Engineering Building, 1513 University Ave., Room 3011, Madison, WI 53706-1691. Email address: zayascaban@wisc.edu
** Postal address: Stephen M. Ross School of Business, University of Michigan, 701 Tappan St., Ann Arbor, MI 48109. Email address: jiaxinl@umich.edu
*** Postal address: Stephen M. Ross School of Business, University of Michigan, 701 Tappan St., Ann Arbor, MI 48109. Email address: sjasin@umich.edu
**** Postal address: Naveen Jindal School of Management, University of Texas, 800 W. Campbell Rd., Richardson, TX 75080. Email address: Guihua.Wang@utdallas.edu

In what follows, we will provide enough discussion so that the correctness of (2) can be easily verified. In Section 2, we recall the complete definition of the LP with both discounting factor and bandit arrivals that was used in [1]. In Section 3, we provide a formal statement of the new lemma and its proof. In Section 4, we discuss how this new bound affects the results in subsequent theorems and propositions in [1].

2. The linear program

Recall that the original bound (1) was used to prove the results in Sections 4 and 6 of [1]. In [1, Section 4] we analyzed a 'fixed population' model, while in [1, Section 6] we analyzed the more general 'dynamic population' model where bandits can arrive in, or depart from, the system. Since the LP used in [1, Section 6] is a generalization of the LP used in [1, Section 4], we only present our analysis for the general LP used in [1, Section 6]. The definition of the LP for any discount factor $\delta \in [0, 1]$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_T) \geq \mathbf{0}$ is given by

$$\text{LP}(\boldsymbol{\epsilon}): V^D(\boldsymbol{\epsilon}) = \min_{x,z} \sum_{t=1}^{T} \sum_{a=0}^{A} \sum_{j=1}^{J} \delta^{t-1} c_j^a \cdot x_j^a(t, \boldsymbol{\epsilon}) + \delta^T \phi \cdot z(\boldsymbol{\epsilon}) \tag{3}$$

$$\begin{aligned}
\text{s.t.} \sum_{a=0}^{A} x_j^a(t, \boldsymbol{\epsilon}) &= \sum_{a=0}^{A} \sum_{i=1}^{J} x_i^a(t-1, \boldsymbol{\epsilon}) \cdot p_{ij}^a + \lambda_{jt} && \forall j \geq 1,\ t \geq 2, \\
\sum_{a=0}^{A} x_j^a(1, \boldsymbol{\epsilon}) &= n_j + \lambda_{j1} && \forall j \geq 1, \\
\sum_{a=1}^{A} \sum_{j=1}^{J} x_j^a(t, \boldsymbol{\epsilon}) &\leq b_t - \epsilon_t && \forall t \geq 1, \\
z(\boldsymbol{\epsilon}) &\geq \sum_{j \in \mathbb{U}} \sum_{a=0}^{A} \sum_{i=1}^{J} x_i^a(T, \boldsymbol{\epsilon}) \cdot p_{ij}^a - m, \\
z(\boldsymbol{\epsilon}),\ x_j^a(t, \boldsymbol{\epsilon}) &\geq 0 && \forall a \geq 0,\ j \geq 1,\ t \geq 1.
\end{aligned}$$

The decision variables in the above LP are the x's and z. It is not difficult to see that the optimal solution will satisfy $z(\boldsymbol{\epsilon}) = \big(\sum_{j \in \mathbb{U}} \sum_{a=0}^{A} \sum_{i=1}^{J} x_i^a(T, \boldsymbol{\epsilon}) \cdot p_{ij}^a - m \big)^+$. All parameters in the above LP are non-negative. In particular, p_{ij}^a is the probability of transitioning from state i to state j under action a $\big($i.e., $\sum_j p_{ij}^a = 1$ for all a and $i\big)$, λ_{jt} is the arrival rate (or expected number) of new bandits in state j at time t, and b_t is the activation budget at time t. The value of ϵ_t is assumed to be small enough so that $b_t - \epsilon_t \geq 0$; otherwise, the LP is not feasible. In [1, Section 3], we used $\lambda_{jt} = 0$ for all j and t, and the bound in (1) was originally proved for this case. We did not provide the proof for the more general case where λ_{jt} could be positive, as the proof for this case was originally deemed to be a straightforward extension of the proof for the simpler case. To avoid confusion, below we will prove the new bound (2) for the general case where λ_{jt} can also be positive.

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press

3. The new lemma

We state our new lemma.

Lemma 1. *Let* $c_{\max} = \max_{a,j} c_j^a$, $b_{\max} = \max_t b_t$, $b_{\min} = \min_{t|b_t \neq 0} b_t$, *and* $\epsilon_{\max} = \max_t \epsilon_t$. *Let* $\mathbb{1}_{\delta \neq 1}$ *and* $\mathbb{1}_{\delta = 1}$ *be indicators for the cases* $\delta \neq 1$ *and* $\delta = 1$, *respectively. If* $\epsilon_{\max} \leq b_{\min}$, *then we have the following bound:*

$$V^D(\boldsymbol{\epsilon}) - V^D(\mathbf{0})$$
$$\leq c_{\max} \cdot \epsilon_{\max} \cdot \frac{b_{\max}}{b_{\min}} \cdot \left[\left(\frac{1-\delta^T}{1-\delta} \right) \cdot \mathbb{1}_{\delta \neq 1} + T \cdot \mathbb{1}_{\delta = 1} \right] + 2\delta^T \phi \cdot \epsilon_{\max} \cdot \frac{b_{\max}}{b_{\min}}.$$

Proof. The proof is by construction. Let $\left\{x_j^a(t, \mathbf{0})\right\}$ denote an optimal solution of LP($\mathbf{0}$). We will use $\left\{x_j^a(t, \mathbf{0})\right\}$ to construct a feasible solution $\left\{\tilde{x}_j^a(t, \boldsymbol{\epsilon})\right\}$ for LP($\boldsymbol{\epsilon}$), under which we let $\tilde{z}(\boldsymbol{\epsilon}) = \left(\sum_{j \in \mathbb{U}} \sum_{a=0}^{A} \sum_{i=1}^{J} \tilde{x}_i^a(T, \boldsymbol{\epsilon}) \cdot p_{ij}^a - m \right)^+$ be the feasible z variable, and show that the gap between the objective value of LP($\boldsymbol{\epsilon}$) under $\left\{\tilde{x}_j^a(t, \boldsymbol{\epsilon})\right\}$ and $V^D(\mathbf{0})$ satisfies the bound in Lemma 1.

For ease of exposition, we will write $x_j^a(t, \mathbf{0})$ as $x_j^a(t)$ and $\tilde{x}_j^a(t, \boldsymbol{\epsilon})$ as $\tilde{x}_j^a(t)$. Define $t^* = \arg\max_{\{t|b_t \neq 0\}} \frac{\epsilon_t}{b_t}$ and let $\beta \in [0, 1]$ be such that

$$(1-\beta) \cdot b_t \geq \epsilon_t \qquad \forall t \in [T]. \tag{4}$$

In our construction of $\left\{\tilde{x}_j^a(t)\right\}$ shortly, we will see that the term βb_t can be interpreted as an upper bound of total budget consumption at time t $\left(\text{i.e., } \sum_{a=1}^{A} \sum_{j=1}^{J} \tilde{x}_j^a(t)\right)$. In particular, we use the following value of β:

$$\beta = 1 - \frac{\epsilon_{t^*}}{b_{t^*}} \geq 1 - \frac{\epsilon_{\max}}{b_{\min}}. \tag{5}$$

Let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2 \ldots \gamma_T)$, where $\gamma_t = (1-\beta) \cdot b_t$. Note that, by definition of γ_t, we have

$$\gamma_t \leq \frac{\epsilon_{\max}}{b_{\min}} \cdot b_t \leq \epsilon_{\max} \cdot \frac{b_{\max}}{b_{\min}}$$

for any t.

We now discuss the construction of $\left\{\tilde{x}_j^a(t)\right\}$. We first describe the construction for $t = 1$ and then complete the construction for $t \geq 2$ by induction. For $t = 1$, define $\left\{\tilde{x}_j^a(1)\right\}$ as follows:

$$\tilde{x}_j^a(1) = \beta x_j^a(1) \qquad \forall\, a \geq 1,$$
$$\tilde{x}_j^0(1) = \beta x_j^0(1) + (1-\beta) \cdot \left[n_j(1) + \lambda_{j1}\right]$$
$$:= \beta x_j^0(1) + \Delta_j(1),$$

where $\Delta_j(1) = (1-\beta) \cdot \left[n_j(1) + \lambda_{j1}\right]$. Clearly, $\tilde{x}_j^a(1) \geq 0$ and so $\left\{\tilde{x}_j^a(1)\right\}$ satisfies the non-negativity constraint in LP($\boldsymbol{\epsilon}$). It is also not difficult to see that $\sum_{a \geq 0} \tilde{x}_j^a(1) = n_j(1) + \lambda_{j1}$ (because $\sum_{a \geq 0} x_j^a(1) = n_j(1) + \lambda_{j1}$, as $\left\{x_j^a(t)\right\}$ is feasible for LP($\mathbf{0}$)), and so $\left\{\tilde{x}_j^a(1)\right\}$ satisfies the second constraint in LP($\boldsymbol{\epsilon}$). Moreover, by definition of β and γ_1, we have

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press

$$\sum_{a\geq 1}\sum_{i}\tilde{x}_i^a(1)=\beta\cdot\left[\sum_{a\geq 1}\sum_{i}x_i^a(1)\right]\leq\beta b_1=b_1-\gamma_1\leq b_1-\epsilon_1,$$

so that $\{\tilde{x}_j^a(1)\}$ satisfies the 'budget constraint' (i.e., the third constraint) in LP($\boldsymbol{\epsilon}$).

Before we proceed with the construction of $\{\tilde{x}_j^a(t)\}$ for $t\geq 2$, we define $n_j(t)$ and $\tilde{n}_j(t)$ as follows:

$$n_j(t)=\sum_{a\geq 0}\sum_{i}x_i^a(t-1)p_{ij}^a\quad\text{and}\quad\tilde{n}_j(t)=\sum_{a\geq 0}\sum_{i}\tilde{x}_i^a(t-1)p_{ij}^a.$$

For $t\geq 2$, we define $\Delta_j(t)$ and $\{\tilde{x}_j^a(t)\}$ recursively as follows:

$$\Delta_j(t)=\sum_{i}\Delta_i(t-1)\cdot p_{ij}^0+(1-\beta)\lambda_{jt},$$

$$\tilde{x}_j^a(t)=\beta x_j^a(t)\qquad\forall\, a\geq 1,$$

$$\tilde{x}_j^0(t)=\beta x_j^0(t)+\Delta_j(t).$$

We prove the following identities by induction:

$$\tilde{n}_j(t)=\beta n_j(t)+\sum_{i}\Delta_i(t-1)p_{ij}^0, \tag{6}$$

$$\tilde{n}_j(t)+\lambda_{jt}=\beta\sum_{a\geq 0}x_j^a(t)+\Delta_j(t), \tag{7}$$

$$\sum_{j}\Delta_j(t)=(1-\beta)\sum_{j}\sum_{a\geq 0}x_j^a(t), \tag{8}$$

$$\sum_{a\geq 0}\tilde{x}_j^a(t)=\tilde{n}_j(t)+\lambda_{jt}, \tag{9}$$

$$\sum_{a\geq 1}\sum_{i}\tilde{x}_i^a(t)=\beta\cdot\left[\sum_{a\geq 1}\sum_{i}x_i^a(t)\right]\leq\beta\cdot b_t=b_t-\gamma_t\leq b_t-\epsilon_t, \tag{10}$$

$$\sum_{j}\left[\tilde{x}_j^0(t)-x_j^0(t)\right]=(1-\beta)\sum_{j}\sum_{a\geq 1}x_j^a(t)\leq(1-\beta)b_t=\gamma_t. \tag{11}$$

First note that Equation (6) follows directly from the definition of $\tilde{n}_j(t)$ and $\{\tilde{x}_j^a(t)\}$:

$$\tilde{n}_j(t)=\sum_{a\geq 0}\sum_{i}\tilde{x}_i^a(t-1)p_{ij}^a=\beta n_j(t)+\sum_{i}\Delta_i(t-1)p_{ij}^0.$$

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press

Next, Equation (7) follows from (6) and the fact that $\sum_{a\geq 0} x_j^a(t) = n_j(t) + \lambda_{jt}$ (because $\{x_j^a(t)\}$ is feasible for LP($\mathbf{0}$)):

$$\begin{aligned}\tilde{n}_j(t) + \lambda_{jt} &= \beta n_j(t) + \sum_i \Delta_i(t-1)p_{ij}^0 + \lambda_{jt} \\ &= \beta[n_j(t) + \lambda_{jt}] + \left[\sum_i \Delta_i(t-1)p_{ij}^0 + (1-\beta)\lambda_{jt}\right] \\ &= \beta[n_j(t) + \lambda_{jt}] + \Delta_j(t) \\ &= \beta \sum_{a\geq 0} x_j^a(t) + \Delta_j(t).\end{aligned}$$

Equation (9) follows directly from the definition of $\{\tilde{x}_j^a(t)\}$ and (7), whereas Equation (10) follows from the definition of $\{\tilde{x}_j^a(t)\}$ and the fact that $\sum_j \sum_{a\geq 1} x_j^a(t) \leq b_t$ (because $\{x_j^a(t)\}$ is feasible for LP($\mathbf{0}$)). Equation (11) follows from the definition of $\{\tilde{x}_j^a(t)\}$ together with (8) and the fact that $\sum_j \sum_{a\geq 1} x_j^a(t) \leq b_t$. Thus, among the six identities (6)–(11), we only need to show (8) by induction. Note that Equations (9) and (10) imply that the constructed $\{\tilde{x}_j^a(t)\}$ for $t \geq 2$ satisfies the first and third constraints in LP($\boldsymbol{\epsilon}$). Since $\tilde{x}_j^a(t)$ is obviously non-negative by construction, it also satisfies the non-negative constraint. As a result, the constructed $\{\tilde{x}_j^a(t)\}$ is feasible for LP($\boldsymbol{\epsilon}$).

We prove (8) by induction starting with $t = 2$. By definition of $\Delta_j(2)$,

$$\begin{aligned}\sum_j \Delta_j(2) &= \sum_j \left[\sum_i \Delta_i(1)p_{ij}^0 + (1-\beta)\lambda_{j,2}\right] \\ &= \sum_i \Delta_i(1) + (1-\beta)\sum_j \lambda_{j,2} \\ &= (1-\beta)\left[\sum_i \sum_{a\geq 0} x_i^a(1) + \sum_j \lambda_{j,2}\right] \\ &= (1-\beta)\left[\sum_j n_j(2) + \sum_j \lambda_{j,2}\right] \\ &= (1-\beta)\sum_j \left[n_j(2) + \lambda_{j,2}\right] \\ &= (1-\beta)\sum_j \sum_{a\geq 0} x_j^a(2),\end{aligned}$$

where the third equality follows since, by definition, $\Delta_i(1) = (1-\beta)[n_i(1) + \lambda_{i,1}] = (1-\beta)\sum_{a\geq 0} x_i^a(1)$ (from the second constraint in LP($\mathbf{0}$)); the fourth equality follows by the definition of $n_j(2)$; and the last equality follows by the first constraint in LP($\mathbf{0}$).

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press

Now, suppose that (6)–(11) hold for all times $s \leq t$. Then

$$\begin{aligned}
\sum_j \Delta_j(t+1) &= \sum_j \left[\sum_i \Delta_i(t) p_{ij}^0 + (1-\beta)\lambda_{j,t+1} \right] \\
&= \sum_i \Delta_i(t) + (1-\beta) \sum_j \lambda_{j,t+1} \\
&= (1-\beta)\left[\sum_i \sum_{a\geq 0} x_i^a(t) + \sum_j \lambda_{j,t+1} \right] \\
&= (1-\beta)\left[\sum_j n_j(t+1) + \sum_j \lambda_{j,t+1} \right] \\
&= (1-\beta) \sum_j \left[n_j(t+1) + \lambda_{j,t+1} \right] \\
&= (1-\beta) \sum_j \sum_{a\geq 0} x_j^a(t+1),
\end{aligned}$$

where the third equality follows by the induction hypothesis, the fourth equality follows by the definition of $n_j(t)$, and the last equality follows by the first constraint in LP($\mathbf{0}$). This completes our inductive step and thus the proof by induction.

We have so far shown that the constructed $\left\{\tilde{x}_j^a(t)\right\}$ is feasible for LP($\boldsymbol{\epsilon}$). We now compute a bound for $V^D(\boldsymbol{\epsilon}) - V^D(\mathbf{0})$. Let $V^D(\boldsymbol{\epsilon}, \tilde{x})$ denote the objective value of LP($\boldsymbol{\epsilon}$) under $\left\{\tilde{x}_j^a(t)\right\}$. Then $V^D(\boldsymbol{\epsilon}) - V^D(\mathbf{0}) \leq V^D(\boldsymbol{\epsilon}, \tilde{x}) - V^D(\mathbf{0})$. Now,

$$\begin{aligned}
&V^D(\boldsymbol{\epsilon}, \tilde{x}) - V^D(\mathbf{0}) \\
&= \sum_j \sum_{t=1}^T \delta^{t-1} c_j^0 \left[\tilde{x}_j^0(t) - x_j^0(t)\right] + \sum_{j,a\geq 1} \sum_{t=1}^T \delta^{t-1} c_j^a \left[\tilde{x}_j^a(t) - x_j^a(t)\right] + \delta^T \phi\left[\tilde{z}(\epsilon) - z(0)\right] \\
&= \sum_j \sum_{t=1}^T \delta^{t-1} c_j^0 \left[\tilde{x}_j^0(t) - x_j^0(t)\right] - (1-\beta) \sum_{j,a\geq 1} \sum_{t=1}^T \delta^{t-1} c_j^a x_j^a(t) + \delta^T \phi\left[\tilde{z}(\epsilon) - z(0)\right] \\
&\leq c_{\max} \cdot \sum_{t=1}^T \delta^{t-1} \gamma_t + \delta^T \phi\left[\tilde{z}(\epsilon) - z(0)\right] \\
&\leq c_{\max} \cdot \epsilon_{\max} \cdot \frac{b_{\max}}{b_{\min}} \cdot \left[\left(\frac{1-\delta^T}{1-\delta}\right) \cdot \mathbb{1}_{\delta\neq 1} + T \cdot \mathbb{1}_{\delta=1} \right] + \delta^T \phi\left[\tilde{z}(\epsilon) - z(0)\right],
\end{aligned}$$

where the first inequality follows from (11) and the last inequality follows since $\gamma_t \leq \epsilon_{\max} \cdot b_{\max}/b_{\min}$. It remains to bound $\delta^T \phi[\tilde{z}(\epsilon) - z(0)]$.

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press

To this end, recall that

$$\tilde{z}(\epsilon) = \left(\sum_{j\in\mathbb{U}} \sum_{a=0}^{A} \sum_{i=1}^{J} \tilde{x}_i^a(T, \boldsymbol{\epsilon}) \cdot p_{ij}^a - m \right)^+ \quad \text{and} \quad z(\epsilon) = \left(\sum_{j\in\mathbb{U}} \sum_{a=0}^{A} \sum_{i=1}^{J} x_i^a(T, \boldsymbol{\epsilon}) \cdot p_{ij}^a - m \right)^+ .$$

Since $\epsilon = 0$ corresponds to the optimal LP solution, $z(0) = \left(\sum_{j\in\mathbb{U}} \sum_{a=0}^{A} \sum_{i=1}^{J} x_i^a(T) \cdot p_{ij}^a - m\right)^+$. Next, recall that ϵ corresponds to perturbing the original LP and as such represents a generalization of this original LP. It follows that $\tilde{z}(\epsilon) \le \tilde{z}(0) = \left(\sum_{j\in\mathbb{U}} \sum_{a=0}^{A} \sum_{i=1}^{J} \tilde{x}_i^a(T) \cdot p_{ij}^a - m\right)^+$. From this it follows that $\tilde{z}(\epsilon) - z(0)$ is bounded above by

$$\left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} \tilde{x}_i^a(T) p_{ij}^a - m \right)^+ - \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} x_i^a(T) p_{ij}^a - m \right)^+ ,$$

and this last expression is bounded above by

$$\left| \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} \tilde{x}_i^a(T) p_{ij}^a - m \right)^+ - \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} x_i^a(T) p_{ij}^a - m \right)^+ \right| .$$

Applying the property $\max\{a, b\} = \frac{1}{2}(a + b + |a - b|)$, where a and b are arbitrary real numbers, yields

$$\begin{aligned} \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} \tilde{x}_i^a(T) p_{ij}^a - m \right)^+ &= \max \left\{ \sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} \tilde{x}_i^a(T) p_{ij}^a - m, 0 \right\} \\ &= \frac{1}{2} \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} \tilde{x}_i^a(T) p_{ij}^a - m + \left| \sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} \tilde{x}_i^a(T) p_{ij}^a - m \right| \right) \end{aligned}$$

and

$$\begin{aligned} \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} x_i^a(T) p_{ij}^a - m \right)^+ &= \max \left\{ \sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} x_i^a(T) p_{ij}^a - m, 0 \right\} \\ &= \frac{1}{2} \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} x_i^a(T) p_{ij}^a - m + \left| \sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} x_i^a(T) p_{ij}^a - m \right| \right) . \end{aligned}$$

Since $|a - b| \ge |a| - |b|$ and, similarly, $|b - a| = |a - b| \ge |b| - |a| = -(|a| - |b|)$ for any two real numbers a and b, the last calculation yields that the expression

$$\left| \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} \tilde{x}_i^a(T) p_{ij}^a - m \right)^+ - \left(\sum_{a\ge 0} \sum_{i} \sum_{j\in\mathbb{U}} x_i^a(T) p_{ij}^a - m \right)^+ \right|$$

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press

is bounded above by

$$\left|\sum_{a\geq 0}\sum_{i}\sum_{j\in\mathbb{U}}\tilde{x}_i^a(T)p_{ij}^a-\sum_{a\geq 0}\sum_{i}\sum_{j\in\mathbb{U}}x_i^a(T)p_{ij}^a\right|.$$

The triangle inequality then implies that this last expression is bounded above by

$$\left|\sum_{i,a\geq 1}\sum_{j\in\mathbb{U}}\tilde{x}_i^a(T)p_{ij}^a-\sum_{i,a\geq 1}\sum_{j\in\mathbb{U}}x_i^a(T)p_{ij}^a\right|+\left|\sum_{i}\sum_{j\in\mathbb{U}}\tilde{x}_i^0(T)p_{ij}^0-\sum_{i}\sum_{j\in\mathbb{U}}x_i^0(T)p_{ij}^a\right|.$$

The property (10) implies that the terms inside the first set of absolute values equals $(1-\beta)\sum_{a\geq 1}\sum_i\sum_{j\in U}x_i^a(T)p_{ij}^a$, so that the expression above equals

$$(1-\beta)\sum_{a\geq 1}\sum_{i}\sum_{j\in\mathbb{U}}x_i^a(T)p_{ij}^a+\left|\sum_{i}\sum_{j\in\mathbb{U}}\tilde{x}_i^0(T)p_{ij}^0-\sum_{i}\sum_{j\in\mathbb{U}}x_i^0(T)p_{ij}^0\right|,$$

which is bounded above by

$$(1-\beta)\sum_{a\geq 1}\sum_{i}\sum_{j}x_i^a(T)p_{ij}^a+\left|\sum_{i}\sum_{j\in\mathbb{U}}\tilde{x}_i^0(T)p_{ij}^0-\sum_{i}\sum_{j\in\mathbb{U}}x_i^0(T)p_{ij}^0\right|,$$

which, by (10), equals

$$(1-\beta)\sum_{a\geq 1}\sum_{i}x_i^a(T)+(1-\beta)\sum_{j\in\mathbb{U}}\sum_{i}\sum_{a\geq 1}x_i^a(T)p_{ij}^0.$$

This last expression is bounded above by $2(1-\beta)\sum_{a\geq 1}\sum_i x_i^a(T)\leq 2(1-\beta)b_T\leq 2\gamma_T$. The choice of β and definition of γ_T yield that $2\gamma_T$ is bounded above by $2\epsilon_{max}\cdot\frac{b_{max}}{b_{min}}$, as claimed. □

4. Impact on other results in Zayas-Cabán *et al.* (2019) [1]

As noted earlier, the bound in the original version of [1, Lemma 2] was used to prove Theorems 1–2 and Propositions 1–3 in Sections 4 and 6 of [1]. It turns out that the new bound in Lemma 1 does not change the results of Theorem 1, Theorem 2, or Proposition 3, but it does slightly change the bounds in Propositions 1 and 2. We discuss all of these below.

Theorem 1 in [1, Section 4]. *In this theorem, we consider the setting where* $\lambda_{jt}=0$ *for all* j *and* t*, and* $\delta=1$*. We can use exactly the same* ϵ_t *as defined in the original version of* [1, *Theorem* 1]*. By the new lemma (Lemma* 1 *of this paper), we still have* $V_\theta^D(\boldsymbol{\epsilon})-V_\theta^D(\mathbf{0})=O(T\sqrt{d\cdot\theta\ln\theta})$*. As a result, there are no changes and we still get exactly the same bound as in the original Theorem* 1. □

Proposition 1 in [1, Section 4]. *In this proposition, we consider the same setting considered in* [1, *Theorem* 1]*, with the exception that we set* $\delta\in(0,1)$*. If we use the same* ϵ_t *as defined in the original* [1, *Proposition* 1]*, by the new Lemma* 1 *we have* $V_\theta^D(\boldsymbol{\epsilon})-V_\theta^D(\mathbf{0})=O(\sqrt{d\cdot\ln T\cdot\theta\ln\theta})$

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press

(the original bound under the old Lemma 2 *of* [1] *was* $O(\sqrt{d\cdot\theta\ln\theta})$*). This implies that the new bound for* [1, *Proposition* 1] *is given by*

$$\frac{V_\theta^{RAC}-V_\theta^D(\mathbf{0})}{V_\theta^D(\mathbf{0})} = O\left(\frac{1}{\theta^d}+\sqrt{\frac{d\cdot\ln T\cdot\ln\theta}{\theta}}\right).$$

Note that if we instead apply the same ϵ_t *as defined in [*1*, Theorem* 1*] to [*1*, Proposition* 1*], it is not difficult to check that the bound becomes*

$$\frac{V_\theta^{RAC}-V_\theta^D(\mathbf{0})}{V_\theta^D(\mathbf{0})} = O\left(\frac{T^2}{\theta^d}+\sqrt{\frac{d\cdot\ln\theta}{\theta}}\right),$$

which, with a proper choice of d, essentially has the same order of magnitude as the bound in Theorem 1. □

Theorem 2 in [1, Section 6]. *In this theorem, we consider the setting where* λ_{jt} *could be positive, and* $\delta=1$*. We can use exactly the same* ϵ_t *as defined in the original version of* [1, *Theorem* 2]*. By the new Lemma* 1*, we still have* $V_\theta^D(\boldsymbol{\epsilon})-V_\theta^D(\mathbf{0})=O(T^{3/2}\sqrt{d\cdot\theta\ln\theta})$*. As a result, nothing changes and we still get exactly the same bound as in the original Theorem* 2. □

Proposition 2 in [1, Section 6]. *In this proposition, we consider the setting where* λ_{jt} *may be positive and* $\delta\in(0,1)$*. If we use the same* ϵ_t *as defined in the original version of* [1, *Proposition* 2]*, the new bound in Proposition* 2 *is given by*

$$\frac{V_\theta^{RAC}-V_\theta^D(\mathbf{0})}{V_\theta^D(\mathbf{0})} = O\left(\frac{1}{\theta^{d/2}}+\sqrt{\frac{d\cdot T\ln T\cdot\ln\theta}{\theta}}\right).$$

□

Proposition 3 in [1, Section 6]. *In this proposition, we consider the setting where bandits can complete service or abandon. Since* $\alpha\in(0,1)$*, we have* $\epsilon_{\max}=O\left(\sqrt{\frac{d\theta\ln\theta}{1-\beta}}\right)$*. So, by the new Lemma* 1, $V_\theta^D(\boldsymbol{\epsilon})-V_\theta^D(\mathbf{0})=O\left(T\sqrt{\frac{d\cdot\theta\ln\theta}{1-\beta}}\right)$*. This does not change anything in the proof of Proposition* 3*, and so the final bound in Proposition* 3 *also does not change.* □

Reference

[1] Zayas-Cabán, G., Jasin, S., and Wang, G. (2019). An asymptotically optimal heuristic for general nonstationary finite-horizon restless multi-armed, multi-action bandits. *Adv. Appl. Prob.* **51**, 745–772.

https://doi.org/10.1017/apr.2022.59 Published online by Cambridge University Press