

Appendix A: Additional Analysis on General Position Gap

In this section, we provide some additional analysis to complement the discussion of GPG in Section 3.1. The first lemma shows that whenever (7) has a unique optimal solution, then λ must have a positive GPG.

Lemma 9. *Suppose that Assumptions 1 and 2 hold and (7) has a unique optimal solution U^* . Then there exists some $\epsilon > 0$ such that λ has a GPG of ϵ .*

Proof. By Lemma 1, if the dual problem (7) has a unique optimal solution U^* , then so does (8). Recall that (8) is formulated as

$$\min_{U \in \mathbb{R}^n} U^\top \lambda, \quad \text{s.t.} \quad \sum_{i \in \mathcal{N}(m)} U_i \geq r_m, \quad \forall m \in \mathcal{M}, \quad (8)$$

and is the dual of

$$\max_x r^\top x, \quad \text{s.t.} \quad \sum_{m: i \in \mathcal{N}(m)} x_m = \lambda_i, \quad \forall i \in \mathcal{N}, \quad x \in \mathbb{R}_{\geq 0}^d \quad (37)$$

By applying strict complementary slackness and the uniqueness of U^* , see e.g., Exercise 4.20 of [Bertsimas and Tsitsiklis \(1997\)](#), there exists a solution x^* of (37) such that $x_m^* > 0$ for all $m \in \mathcal{M}$ where $\sum_{i \in \mathcal{N}(m)} U_i^* = r_m$. This implies that the strictly positive indices of x^* contain at least one basis. Note that while x^* is not necessarily a basic optimal solution, there exists a finite number of basic optimal solutions $x^1, x^2, \dots, x^k$ of (37), with $k \geq 1$, such that

$$x^* = \alpha_1 x^1 + \dots + \alpha_k x^k,$$

with $\alpha_1, \dots, \alpha_k$ being strictly positive and sum to one. Also, because Assumptions 1 and 2 do not allow an optimal solution to consist entirely of self-matches, we have that $\sum_m x_m^* < 1$.

Let $B(1), B(2), \dots, B(k)$ be the corresponding basis of $x^1, x^2, \dots, x^k$. Thus, we have $x^j = M_{B(j)}^{-1} \lambda$ for $j = 1, \dots, k$. Define

$$\hat{x} = \alpha_1 M_{B(1)}^{-1} \hat{\lambda} + \dots + \alpha_k M_{B(k)}^{-1} \hat{\lambda}.$$

By definition, $\hat{x} = x^*$ when $\hat{\lambda} = \lambda$. Because $\sum_m x_m^* < 1$, there exists some $\epsilon > 0$ such that for any $\hat{\lambda} \in \mathcal{B}_\epsilon(\lambda)$, we have $\sum_m \hat{x}_m < 1$, and

$$\hat{x}_m > 0 \text{ if and only if } \sum_{i \in \mathcal{N}(m)} U_i^* = r_m.$$

As a result, when λ is replaced by $\hat{\lambda} \in \mathcal{B}_\epsilon(\lambda)$, we have $\hat{x}$ and U^* forms an optimal primal and dual pair for (7). This completes the proof.

Q.E.D.

In [Kerimov et al. \(2024\)](#), the authors defined another notion of general position gap (referred to as K-GPG hereafter) using the optimal solution x^* of (3), under the assumption that the optimal solution is unique and non-degenerate.⁴ More specifically, assuming x^* has basis B , then K-GPG is defined as $\min_{m \in B} x_m^*$. The next lemma shows that the GPG in this paper has a simple characterization when x^* is unique and non-degenerate under suitable conditions. A consequence of the lemma is that our GPG is within a factor of K-GPG.

Lemma 10. *When the optimal solution x^* of (3) is unique and non-degenerate, define the optimal basis B as $B = \{m \mid x_m^* > 0\}$. Let M_B and x_B^* be the submatrix of M and x correspond to B , such that $M_B x_B^* = \lambda$. For each $m \in B$, let $v_m^\top$ be the row vector in M_B^{-1} such that $v_m^\top \lambda = x_m^*$, then the GPG is equal to*

$$\epsilon' = \min_{m \in B} \frac{x_m^*}{\|v_m\|},$$

assuming that $1 - \sum_m x_m^* > \|\sum_{m \in B} v_m\| \epsilon'$.

⁴ The formulation of [Kerimov et al. \(2024\)](#) is slightly different with constraint set $Mx \leq \lambda$. This is equivalent to formulation (3) with the appropriate self-matches having zero rewards.

Proof. By definition of B and the fact that x^* is nondegenerate, $M_B^{-1}\lambda > 0$, i.e., $x_m^* > 0$ for all $m \in B$. Also, note that when x^* is unique and nondegenerate, we also have that the optimal solution for the dual, U^* , is unique and nondegenerate.

Let the GPG equal to ϵ . We first show that $\epsilon \geq \epsilon'$. Observe that for any $\Delta \in \mathbb{R}^n$ with $\|\Delta\|_2 \leq \epsilon'$, we have

$$\begin{aligned} v_m^\top(\lambda + \Delta) &\geq v_m^\top(\lambda - \epsilon' \cdot \frac{v_m}{\|v_m\|}) \\ &= x_m^* - \|v_m\| \min_{m \in B} \frac{x_m^*}{\|v_m\|} \\ &\geq 0. \end{aligned}$$

Moreover, from our assumption that $1 - \sum_m x_m^* > \|\sum_{m \in B} v_m\| \epsilon'$, we have

$$\begin{aligned} \sum_{m \in B} v_m^\top(\lambda + \Delta) &\leq \sum_{m \in B} v_m^\top \left(\lambda + \epsilon' \cdot \frac{\sum_{m \in B} v_m}{\|\sum_{m \in B} v_m\|} \right) \\ &= \sum_{m \in B} x_m^* + \epsilon' \left\| \sum_{m \in B} v_m \right\| \\ &< 1. \end{aligned}$$

Define $\tilde{x} \in \mathbb{R}^d$, where $\tilde{x}_m = v_m^\top(\lambda + \Delta)$ for $m \in B$, and $\tilde{x}_m = 0$ for $m \in [d] \setminus B$. From the analysis above, when λ is changed to $\lambda + \Delta$, we have that $\tilde{x}$ is feasible. Furthermore, because $\tilde{x}$ and U^* satisfies the complementary slackness, it implies that U^* remains an optimal dual solution for any Δ such that $\|\Delta\|_2 \leq \epsilon'$. Therefore, $\epsilon \geq \epsilon'$.

Next, we show $\epsilon \leq \frac{x_m^*}{\|v_m\|}$ for each $m \in B$. Suppose that $\epsilon > \frac{x_m^*}{\|v_m\|}$. As $x_m^* = v_m^\top \lambda$ we have

$$v_m^\top \left(\lambda - \epsilon \frac{v_m}{\|v_m\|} \right) = x_m^* - \epsilon \|v_m\| < 0,$$

which implies that the primal basis would change. Because the dual solution U^* is non-degenerate, this implies that the optimal dual solution will change when the optimal primal basis change, which contradicts to the definition of the GPG. Therefore, we have $\epsilon \leq \frac{v_m^\top \lambda}{\|v_m\|}$ for each $m \in B$, implying

$$\epsilon \leq \min_{m \in B} \frac{v_m^\top \lambda}{\|v_m\|} = \epsilon'.$$

Therefore, we have that $\epsilon = \epsilon'$. Q.E.D.

We remark that the assumption $1 - \sum_{m \in B} x_m^* > \|\sum_{m \in B} v_m\| \epsilon'$ in Lemma 10 is mild. In most of the non-trivial systems, $\sum_{m \in B} x_m^*$ is typically much smaller than 1, while ϵ' is often significantly smaller than $1 - \sum_{m \in B} x_m^*$. Indeed, this condition is satisfied by all three numerical settings we examined in Section 5. Moreover, when the assumption $1 - \sum_{m \in B} x_m^* > \|\sum_{m \in B} v_m\| \epsilon'$ does not hold, there is a straightforward way to modify Algorithm 1 to relax this condition. Specifically, we can allow the primal-dual algorithm to schedule up to two matches per time period instead of just one, i.e., set $x(t)$ to two times the m^* -th standard basis vector. Under this modification, Lemma 10 would hold with the adjusted assumption $2 - \sum_m x_m > \|\sum_{m \in B} v_m\| \epsilon'$.

In addition, when x^* is nondegenerate (but not unique), our proof shows that $\epsilon \geq \epsilon'$ but not necessarily $\epsilon \leq \epsilon'$, as U^* may stay optimal even when $\epsilon > \frac{v_m^\top \lambda}{\|v_m\|}$ for some $m \in B$. Thus, without the uniqueness of the primal optimal solution, the computational procedure only yields a lower bound of ϵ .

With Lemma 10, we can directly compare GPG and K-GPG. Specifically, we have

$$\min_{m \in B} \|v_m\|^{-1} \cdot \left(\min_{m \in B} x_m^* \right) \leq \epsilon = \min_{m \in B} \frac{x_m^*}{\|v_m\|} \leq \max_{m \in B} \|v_m\|^{-1} \cdot \left(\min_{m \in B} x_m^* \right).$$

Appendix B: Additional Proofs

Here, we provide the proofs for some of our technical results which are fairly standard given the literature.

B.1. Proof for the Properties of the Dual Solution

Here, we provide the proof of Lemma 1 by showing that any optimal solution for the dual problem without the constraint $\sum_{m \in \mathcal{M}} x_m \leq 1$ is also an optimal solution for the dual problem with it.

Lemma 11 (Restatement of Lemma 1). *Let λ be any non-negative vector in which $\sum_{i=1}^n \lambda_i \leq 1$. If $\hat{U}$ is an optimal solution of (8), it is also an optimal solution of*

$$\min_{U \in \mathbb{R}^n} L_\lambda(U) = \min_{U \in \mathbb{R}^n} \max_{x \in \mathcal{X}} (r^\top - U^\top M) x + U^\top \lambda,$$

where we recall that $\mathcal{X} = \{x \mid x \in \mathbb{R}_{\geq 0}^d, \sum_{m \in \mathcal{M}} x_m \leq 1\}$. Furthermore, we have

$$\|\hat{U}\|_\infty \leq r_{\max} \quad \text{and} \quad \|\hat{U}\|_2 \leq \sqrt{n} r_{\max}.$$

Proof. First, we show that $\hat{U}$ is an optimal solution of $\min_{U \in \mathbb{R}^n} L_\lambda(U)$. It suffices to show that

$$\min_{U \in \mathbb{R}^n} L_\lambda(U) = \min_{U \in \mathbb{R}^n: r - M^\top U \leq 0} U^\top \lambda.$$

For any $U \in \mathbb{R}^n$ such that $\max_m (r - M^\top U)_m > 0$, define $U'_i = U_i + v$ for all $i \in [n]$, where $v = \max_m (r - M^\top U)_m$. Then $r - M^\top U' \leq 0$ and hence $L_\lambda(U') = U'^\top \lambda$. Also, note that $L_\lambda(U) = v + U^\top \lambda$. Therefore,

$$L_\lambda(U') - L_\lambda(U) = (U' - U)^\top \lambda - v = v \sum_{i=1}^n \lambda_i - v \leq 0.$$

Hence,

$$\min_{U \in \mathbb{R}^n} L_\lambda(U) = \min_{U \in \mathbb{R}^n: r - M^\top U \leq 0} L_\lambda(U) = \min_{U \in \mathbb{R}^n: r - M^\top U \leq 0} U^\top \lambda.$$

It remains to prove that $\|\hat{U}\|_\infty \leq r_{\max}$ and $\|\hat{U}\|_2 \leq \sqrt{n} r_{\max}$. By Assumption 2 and feasibility of $\hat{U}_i$, we have $\hat{U}_i \geq r_i \geq 0$. Also, by optimality of $\hat{U}$, we have that for all $i \in [n]$, $\hat{U}_i \leq r_{\max}$ as otherwise, we can decrease $\hat{U}_i$ and still obtain a feasible solution. Therefore, we have

$$\|\hat{U}\|_\infty \leq r_{\max}.$$

For $\|\hat{U}\|_2$, we have that $\|\hat{U}\|_2^2 \leq n \|\hat{U}\|_\infty^2 \leq n r_{\max}^2$. Q.E.D.

B.2. Proof of Proposition 1

Proof. By the definition given in (6),

$$L_{\hat{\lambda}}(U) = \max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle U, \hat{\lambda} \rangle = L_\lambda(U) + \langle U, \hat{\lambda} - \lambda \rangle.$$

Therefore,

$$L_{\hat{\lambda}}(U) - L_{\hat{\lambda}}(U^*) = L_\lambda(U) - L_\lambda(U^*) + \langle U - U^*, \hat{\lambda} - \lambda \rangle. \quad (38)$$

Cond.1 $\Rightarrow$ Cond.2: By the duality between $\|\cdot\|$ and $\|\cdot\|_*$, we can choose $\hat{\lambda}$ such that $\|\hat{\lambda} - \lambda\| = \epsilon$ and $\langle U - U^*, \hat{\lambda} - \lambda \rangle = -\epsilon \|U - U^*\|_*$. Thus, in view of $L_{\hat{\lambda}}(U) \geq L_{\hat{\lambda}}(U^*)$, we deduce from (38) that

$$L_\lambda(U) - L_\lambda(U^*) \geq \epsilon \|U - U^*\|_*.$$

Cond.2 $\Rightarrow$ Cond.1: Combining (38) with Cond.2 yields that for all $\|\hat{\lambda} - \lambda\| \leq \epsilon$,

$$L_{\hat{\lambda}}(U) - L_{\hat{\lambda}}(U^*) \geq \epsilon \|U - U^*\| + \langle U - U^*, \hat{\lambda} - \lambda \rangle \geq 0,$$

where the last inequality holds by the duality between $\|\cdot\|$ and $\|\cdot\|_*$. Therefore, U^* is an optimal solution to $L_{\hat{\lambda}}(U)$ for all $\|\hat{\lambda} - \lambda\| \leq \epsilon$. Q.E.D.

B.3. Proof of Corollary 2

In this subsection, we provide the proof for Corollary 2. Throughout the subsection, we will analyze the primal-dual policy considered in Corollary 2, which sets $V_t = \sqrt{t}$ and $\hat{U}(t) = \mathbf{0}$ for all t , and provide that it achieves $O(\sqrt{T})$ regret. First, we introduce the following lemma.

Lemma 12. *For any $t \in \mathbb{N}$, entries of $U(t)$ cannot be all strictly negative.*

Proof. For any $t \in \mathbb{N}$, we have $U(t) = \frac{\delta(t)}{V_t}$. Then, it suffices to show that all entries of $\delta(t)$ cannot be all strictly negative for any $t \in \mathbb{N}$, given $V_t > 0$. Next, we claim that $\sum_{i \in [n]} \delta_i(t) \geq -1$ for all $t \in \mathbb{N}$. Note that this claim, combined with the facts that $\delta(t)$ always has integer entries and any non-trivial system satisfying Assumption 2 has at least two types of agents, immediately implies that all entries of $\delta(t)$ cannot be all strictly negative.

Lastly, we are left to prove our claim. By definition, $\sum_{i \in [n]} \delta_i(t) \in \mathbb{Z}$. Note that $\sum_{i \in [n]} \delta_i(0) = 0$. By induction, it suffices to show that for any $t \in \mathbb{N}$, if $\sum_{i \in [n]} \delta_i(t) \geq -1$, then $\sum_{i \in [n]} \delta_i(t+1) \geq -1$. Fix some $t \in \mathbb{N}$. By Algorithm 1,

$$\delta(t+1) = \delta(t) + Mx(t) - A(t+1),$$

then we have

$$\begin{aligned} \sum_{i \in [n]} \delta_i(t+1) &= \sum_{i \in [n]} \delta_i(t) + \sum_{i \in [n]} (Mx(t))_i - \sum_{i \in [n]} A_i(t+1) \\ &= \sum_{i \in [n]} \delta_i(t) + \sum_{i \in [n]} (Mx(t))_i - 1, \end{aligned} \tag{39}$$

where the second equality holds because only one agent arrives at each time slot, i.e., $\sum_{i \in [n]} A_i(t+1) = 1$. Then, we consider the following two cases.

- Suppose $\sum_{i \in [n]} \delta_i(t) \geq 0$. By (39),

$$\begin{aligned} \sum_{i \in [n]} \delta_i(t+1) &= \sum_{i \in [n]} \delta_i(t) + \sum_{i \in [n]} (Mx(t))_i - 1 \\ &\geq \sum_{i \in [n]} \delta_i(t) - 1 \geq -1, \end{aligned}$$

where the first inequality holds because $\sum_{i \in [n]} (Mx(t))_i \geq 0$, and the second inequality holds by our assumption that $\sum_{i \in [n]} \delta_i(t) \geq 0$. Hence, it follows that $\sum_{i \in [n]} \delta_i(t+1) \geq -1$.

- Suppose that $\sum_{i \in [n]} \delta_i(t) = -1$. Since $\sum_{i \in [n]} \delta_i(t) = -1$, there must exist an agent $j \in [n]$ with $\delta_j(t) \leq -1$. Then, let $m' \in [d]$ denote the self-match of agent j . Then,

$$r_{m'} - U_j(t) = r_{m'} - \frac{\delta_j(t)}{V_t} > 0,$$

given that $r_{m'} \geq 0$. Recall that

$$x(t) \in \arg \max_{x \in \mathcal{X}} \left(r^\top - U(t)^\top M \right) x,$$

then $x(t)$ must be a vector with one entry equals to 1 and all other entries equal to 0, i.e., the scheduled match must be non-empty. By (39),

$$\begin{aligned} \sum_{i \in [n]} \delta_i(t+1) &= \sum_{i \in [n]} \delta_i(t) + \sum_{i \in [n]} (Mx(t))_i - 1 \\ &\geq \sum_{i \in [n]} \delta_i(t) = -1, \end{aligned}$$

where the inequality holds because $x(t)$ is a vector with one entry equals to 1 and all other entries equal to 0. Hence, it follows that $\sum_{i \in [n]} \delta_i(t+1) \geq -1$.

Q.E.D.

Next, we introduce the following lemma that provides an upper bound on $\langle U, Mx(t) - \lambda \rangle$ for any $U \in \mathbb{R}^n$ with at least one non-negative entry.

Lemma 13. *For any $U \in \mathbb{R}^n$ whose entries are not all strictly negative, and x^* such that*

$$x^* \in \arg \max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle,$$

we have

$$\langle U, Mx^* - \lambda \rangle \leq r_{\max} - \frac{\lambda_{\min}}{\sqrt{n}} \|U\|_2.$$

Proof. First, we prove that

$$\max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle \lambda, U \rangle \geq \lambda_{\min} \|U\|_\infty. \quad (40)$$

To prove (40), let $i^* \in \arg \max |U_i|$. We then consider the following three cases.

- If $U_{i^*} = 0$, (40) trivially holds.
- If $U_{i^*} > 0$. Let $\tilde{x}$ denote one of the feasible solutions of $Mx = \tilde{\lambda}$, where $\tilde{\lambda}_i = \lambda_i$ for $i \neq i^*$ and $\tilde{\lambda}_{i^*} = 0$. Then, it follows that $\tilde{x} \in \mathcal{X}$, where $\mathcal{X} = \{x \mid x \in \mathbb{R}_{\geq 0}^d, \sum_{m \in \mathcal{M}} x_m \leq 1\}$. Then, we have

$$\max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle \tilde{\lambda}, U \rangle \geq \langle r - M^\top U, \tilde{x} \rangle + \langle \tilde{\lambda}, U \rangle = \langle r, \tilde{x} \rangle + \langle \tilde{\lambda} - M\tilde{x}, U \rangle \geq 0. \quad (41)$$

Hence, we get

$$\begin{aligned} \max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle \lambda, U \rangle &= \max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle \tilde{\lambda}, U \rangle + \langle \lambda - \tilde{\lambda}, U \rangle \\ &\geq \langle \lambda - \tilde{\lambda}, U \rangle \\ &= \lambda_{i^*} U_{i^*} \\ &\geq \lambda_{\min} \|U\|_\infty, \end{aligned}$$

where the first inequality holds by (41), and the second equality holds because $\tilde{\lambda}_i = \lambda_i$ for $i \neq i^*$ and $\tilde{\lambda}_{i^*} = 0$.

- suppose $U_{i^*} < 0$. Since entries of U are not all strictly negative, there must exist some i' such that $U_{i'} \geq 0$, and we have that

$$\begin{aligned} \max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle \lambda, U \rangle &\geq r_{i^*} - U_{i^*} + \langle \lambda, U \rangle \\ &\geq r_{i^*} - U_{i^*} + \sum_{i \neq i'} \lambda_i U_i \\ &\geq r_{i^*} - U_{i^*} + \sum_{i \neq i'} \lambda_i U_{i^*} \\ &\geq -\lambda_{i'} U_{i^*} \\ &\geq \lambda_{\min} \|U\|_\infty, \end{aligned}$$

where the first inequality because the indicator vector at self match i^* is a feasible solution in $\mathcal{X}$; the second inequality holds because $U_{i'} \geq 0$; the third inequality holds because $U_i \geq U_{i^*}$ for any $i \in [n]$; and the fourth inequality holds because $r_{i^*} \geq 0$.

Next, applying (40), we get

$$\begin{aligned}
-\langle U, Mx^* - \lambda \rangle &= \langle r - M^\top U, x^* \rangle + \langle \lambda, U \rangle - \langle r, x^* \rangle \\
&\geq \max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle \lambda, U \rangle - r_{\max} \\
&\geq \lambda_{\min} \|U\|_\infty - r_{\max} \\
&\geq \frac{\lambda_{\min}}{\sqrt{n}} \|U\|_2 - r_{\max}.
\end{aligned}$$

where the first inequality holds by the definition of x^* and $\langle r, x^* \rangle \leq r_{\max}$ for any $x \in \mathcal{X}$; the second inequality holds by our claim; and the last inequality holds because $\|U\|_\infty \geq \frac{1}{\sqrt{n}} \|U\|_2$. Hence, our desired result follows.

Q.E.D.

With Lemma 12 and Lemma 13, we are now ready to complete the proof of Corollary 2.

Proof of Corollary 2. By (20), when $\hat{U}(t) = \mathbf{0}$,

$$\mathbb{E} \left[\|\delta(t+1)\|_2^2 \mid \mathcal{F}_t \right] \leq \|\delta(t)\|_2^2 + B + 2V_t \langle U(t) - \mathbf{0}, Mx(t) - \lambda \rangle. \quad (42)$$

Applying (42) together with Lemma 12 and Lemma 13, for any $t \in \mathbb{N}$, we get

$$\begin{aligned}
\mathbb{E} \left[\|\delta(t+1)\|_2^2 \mid \mathcal{F}_t \right] &\leq \|\delta(t)\|_2^2 + B + 2V_t r_{\max} - 2V_t \frac{\lambda_{\min}}{\sqrt{n}} \|U(t)\|_2 \\
&\leq \|\delta(t)\|_2^2 + B + 2V_t r_{\max} - 2 \frac{\lambda_{\min}}{\sqrt{n}} \|\delta(t)\|_2,
\end{aligned}$$

where the second inequality holds by $U(t) = \frac{\delta(t)}{V_t}$. Then, for any $t \in \mathbb{N}$, if

$$\|\delta(t)\|_2 \geq \frac{B + 2V_t r_{\max}}{\frac{\lambda_{\min}}{\sqrt{n}}},$$

we obtain

$$\mathbb{E} \left[\|\delta(t+1)\|_2^2 \mid \mathcal{F}_t \right] \leq \left(\|\delta(t)\|_2 - \frac{\lambda_{\min}}{2\sqrt{n}} \right)^2.$$

By Jensen's inequality, we get

$$\mathbb{E} [\|\delta(t+1)\|_2 \mid \mathcal{F}_t] \leq \sqrt{\mathbb{E} [\|\delta(t+1)\|_2^2 \mid \mathcal{F}_t]} \leq \|\delta(t)\|_2 - \frac{\lambda_{\min}}{2\sqrt{n}}.$$

By choosing $V_t = \sqrt{t}$, and invoking Lemma 6, we get that for any $t \leq T$,

$$\begin{aligned}
\mathbb{E} [\|\delta(t)\|_2] &\leq B + \frac{B + 2V_t r_{\max}}{\frac{\lambda_{\min}}{\sqrt{n}}} + B \left(\frac{B - \frac{\lambda_{\min}}{\sqrt{n}}}{2 \frac{\lambda_{\min}}{\sqrt{n}}} \right) \\
&\leq O \left(\frac{\sqrt{n}}{\lambda_{\min}} \left(B^2 + r_{\max} \sqrt{T} \right) \right).
\end{aligned}$$

It then follows from Proposition 2 that

$$t \langle r, x^* \rangle - \sum_{s=1}^t \mathbb{E} [\langle r, x(s) \rangle] \leq O(\sqrt{T}), \quad \forall t \leq T.$$

B.4. Proof of Lemma 6

Proof. Following Gupta (2024), let $\Gamma(t)$ denote a reflected random walk with step size K , reflected boundary $\Gamma_{\min} = (1 + \lceil \frac{D}{K} \rceil)K$, initial condition $\Gamma(0) = \Gamma_{\min}$, and

$$\Gamma(t+1) = \max\{\Gamma(t) + \xi(t)K, \Gamma_{\min}\} \quad \text{for } t \geq 0,$$

where $\xi(t)$ are *i.i.d.* random variables taking the value $+1$ with probability $\frac{1}{2} - \frac{\eta}{2K}$, and -1 with probability $\frac{1}{2} + \frac{\eta}{2K}$.

By (Gupta, 2024, Lemma 3), $\Psi(t) \leq_{\text{icx}} \Gamma(t)$ for all $t \geq 0$, where for any $X, Y \in \mathbb{R}^n$, $X \leq_{\text{icx}} Y$ means that X is dominated by Y in the increasing convex order, that is, $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ for any increasing convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ with valid expectation. It follows that that for any $t \geq 0$,

$$\mathbb{E}[\Psi(t)] \leq \mathbb{E}[\Gamma(t)]. \quad (43)$$

Define $\rho \triangleq \frac{\frac{1}{2} - \frac{\eta}{2K}}{\frac{1}{2} + \frac{\eta}{2K}} = \frac{K-\eta}{K+\eta}$. By (Cox and Miller, 2017, Section 2.2, Equation (52)), $\{\Gamma(t)\}$ has a unique steady state distribution under which

$$\mathbb{E}[\Gamma_{\infty}] = \Gamma_{\min} + K \frac{\rho}{1-\rho} = \Gamma_{\min} + K \left(\frac{K-\eta}{2\eta} \right). \quad (44)$$

Next, we construct $\Gamma'(t)$ such that $\Gamma'(0) \stackrel{d}{=} \Gamma_{\infty}$ follows a steady state distribution, and for any $t \geq 0$, $\Gamma'(t+1) = \max\{\Gamma'(t) + \xi(t)K, \Gamma_{\min}\}$. By induction, we have $\Gamma(t) \leq \Gamma'(t)$ for all $t \geq 0$. Therefore,

$$\mathbb{E}[\Gamma(t)] \leq \mathbb{E}[\Gamma'(t)] = \mathbb{E}[\Gamma_{\infty}], \quad (45)$$

where the last inequality holds because $\Gamma'(t) \stackrel{d}{=} \Gamma_{\infty}$ by construction. Combining (43), (44) and (45) yields the desired result (31). Q.E.D.

B.5. Proof of Lemma 7

Proof. By Definition 1,

$$\mathbb{P}\{\widehat{U}(t) \neq U^*\} \leq \mathbb{P}\left\{\left\|\widehat{\lambda}(t) - \lambda\right\|_2 > \epsilon\right\}. \quad (46)$$

It remains to establish the concentration inequality for $\left\|\widehat{\lambda}(t) - \lambda\right\|_2$. First,

$$\mathbb{E}\left[\left\|\widehat{\lambda}(t) - \lambda\right\|_2^2\right] = \mathbb{E}\left[\sum_{i=1}^n \left(\frac{1}{t} \sum_{s=1}^t A_i(s) - \lambda_i\right)^2\right] = \frac{1}{t} \sum_{i=1}^n \lambda_i(1 - \lambda_i) \leq \frac{1}{t}. \quad (47)$$

Thus, by Jensen's inequality,

$$\mathbb{E}\left[\left\|\widehat{\lambda}(t) - \lambda\right\|_2\right] \leq \sqrt{\mathbb{E}\left[\left\|\widehat{\lambda}(t) - \lambda\right\|_2^2\right]} \leq \sqrt{\frac{1}{t}}.$$

Note that the function $f: (A(1), \dots, A(t)) \rightarrow \left\|\widehat{\lambda} - \lambda\right\|_2^2$ satisfies the bounded difference property with parameter $2/t$, that is, for any $s \in [t]$ and any $A(1), \dots, A(t), A'(s)$,

$$\begin{aligned} & |f(A(1), \dots, A(s-1), A(s), A(s+1), \dots, A(t)) - f(A(1), \dots, A(s-1), A'(s), A(s+1), \dots, A(t))| \\ & \leq \frac{1}{t} \|A(s) - A'(s)\|_2 \leq \frac{2}{t}. \end{aligned}$$

Thus, by McDiarmid's inequality, for any $\Delta > 0$,

$$\mathbb{P}\left\{\left\|\widehat{\lambda}(t) - \lambda\right\|_2 \geq \mathbb{E}\left[\left\|\widehat{\lambda}(t) - \lambda\right\|_2\right] + \sqrt{\Delta/t}\right\} \leq \exp\left(-\frac{2\Delta/t}{t(2/t)^2}\right) = \exp(-\Delta/2).$$

Combining the last two displayed equations gives that

$$\mathbb{P} \left\{ \left\| \hat{\lambda}(t) - \lambda \right\|_2 \geq \sqrt{\frac{1}{t}} + \sqrt{\frac{\Delta}{t}} \right\} \leq \exp(-\Delta/2).$$

If $\Delta \geq 1$, then this further implies that

$$\mathbb{P} \left\{ \left\| \hat{\lambda}(t) - \lambda \right\|_2 \geq 2\sqrt{\frac{\Delta}{t}} \right\} \leq 2\exp(-\Delta/2);$$

If $\Delta \leq 1$, then the above is true trivially, as $2\exp(-\Delta/2) \geq 2\exp(-1/2) \geq 1$. Therefore, by letting $\epsilon = 2\sqrt{\Delta/t}$, we deduce that

$$\mathbb{P} \left\{ \left\| \hat{\lambda}(t) - \lambda \right\|_2 \geq \epsilon \right\} \leq 2\exp(-t\epsilon^2/8). \quad (48)$$

Finally, plugging the last displayed equation into (46) and summing over all $t \geq 1$, we get that

$$\sum_{t=1}^{\infty} \mathbb{P} \left\{ \hat{U}(t) \neq U^* \right\} \leq \sum_{t=1}^{\infty} 2\exp(-t\epsilon^2/8) \leq \frac{2}{\exp(\epsilon^2/8) - 1} \leq \frac{16}{\epsilon^2}, \quad (49)$$

where the last equality holds because $\exp(x) \geq 1 + x$. Q.E.D.

Appendix C: Extensions

C.1. System without Self-Matches

In this section, we discuss the primal-dual policy for a system with Assumption 1 but not Assumption 2, that is, self-matches are not necessarily allowed for all agent types. Recall that the fluid relaxation can be formulated as:

$$\begin{bmatrix} \max_x & r^\top x \\ \text{s.t.} & Mx \leq \lambda \\ & x \in \mathbb{R}_{\geq 0}^d \end{bmatrix}. \quad (50)$$

Without Assumption 2, (50) can no longer be reformulated with constraints $Mx = \lambda$. Instead, we consider the following reformulation, which adds slack variables z :

$$\begin{bmatrix} \max_{x,z} & r^\top x \\ \text{s.t.} & Mx + z = \lambda \\ & (x, z) \in \mathcal{X} \end{bmatrix}, \quad (51)$$

where $\mathcal{X} = \{(x, z) \mid (x, z) \in \mathbb{R}_{\geq 0}^{d+n}, \sum_{m \in \mathcal{M}} x_m + \sum_{i \in \mathcal{N}} z_i \leq 1\}$. Similar to Section 3, we add the redundant constraint $\sum_{m \in \mathcal{M}} x_m + \sum_{i \in \mathcal{N}} z_i \leq 1$ to ensure the optimal solution of the Lagrangian relaxed problem is bounded.

We next formally introduce the Lagrangian relaxation and our primal-dual policy in a similar fashion as Section 3, except that all of the definitions are basic on the fluid formulation (51) instead of (3). Specifically, we define the Lagrangian relaxation of (51) as

$$L_\lambda(U) = \max_{(x,z) \in \mathcal{X}} \left\{ r^\top x - U^\top (Mx + z - \lambda) \right\} = \max_{(x,z) \in \mathcal{X}} \left(r^\top - U^\top M \right) x - U^\top z + U^\top \lambda.$$

Also, we say that λ has a GPG of (at least) $\epsilon > 0$, if there exists U^* that is the unique optimal solution of

$$\min_{U \in \mathbb{R}^n} L_\lambda(U) \quad (52)$$

and U^* is optimal for

$$\min_{U \in \mathbb{R}^n} L_{\hat{\lambda}}(U), \text{ for any } \hat{\lambda} \in \mathcal{B}_\epsilon(\lambda).$$

The primal-dual policy built based on the fluid formulation (51) (hence without Assumption 2) is presented as Algorithm 2. The intuition behind Algorithm 2 is mostly the same as the intuition behind Algorithm 1. The

Algorithm 2 Primal-Dual Policy for System without Assumption 2.

-
- 1: **Input:** $\{V_t\}_{t=1}^\infty$, set $\delta(0)$ and $x(0)$ to the zero vectors.
 - 2: **for** each $t = 1, 2, \dots, T$ **do**
 - 3: Observe arrival $A(t)$
 - 4: THE SCHEDULING PROCESS:
 - 5: Update $\delta(t) = \delta(t-1) + Mx(t-1) + z(t-1) - A(t)$
 - 6: Let $\hat{\lambda}(t) = \sum_{s=1}^t A(s)/t$, and find $\hat{U}(t)$, an optimal solution of

$$\min_{U \in \mathbb{R}^n} U^\top \hat{\lambda}(t), \text{ s.t. } \sum_{i \in \mathcal{N}(m)} U_i \geq r_m, \forall m, \text{ and } U \in \mathbb{R}_{\geq 0}^n,$$

where $\hat{\lambda}(t)$ denotes the empirical arrival rate based on the arrivals from periods 1 to t .

- 7: Update $U(t) = \hat{U}(t) + \frac{\delta(t)}{V_t}$
- 8: Let $(x(t), z(t))$ to be the standard basis vector such that it is also the solution of

$$\max_{(x,z) \in \mathcal{X}} (r^\top - U(t)^\top M) x - U(t)^\top z.$$

- 9: THE REALIZATION PROCESS:
 - 10: Initialize $y(t) = \mathbf{0}_{|\mathcal{M}|+|\mathcal{N}|}$.
 - 11: Set $W_m(t) = \sum_{s=1}^t (x_m(s) - y_m(s))$, for each $m \in \mathcal{M}$
 - 12: Set $W_i(t) = \sum_{s=1}^t (z_i(s) - y_i(s))$, for each $i \in \mathcal{N}$
 - 13: **for** each $m \in \mathcal{M} \cup \mathcal{N}$ with $W_m(t) > 0$ **do**
 - 14: Let $I_m(t) = \min_{i \in \mathcal{N}(m)} \sum_{s=1}^t (A_i(s) - M_i y(s))$
 - 15: Update $y_m(t) = y_m(t) + \min\{W_m(t), I_m(t)\}$. (Note that if $m \in \mathcal{M}$, then $\min\{W_m(t), I_m(t)\}$ type m matches are realized; and if $m \notin \mathcal{M}$, then m must correspond to a self-match for some agent of type i , implying that $\min\{W_m(t), I_m(t)\}$ of type i agents are set aside, never to be used in the future)
 - 16: **end for**
 - 17: **end for**
-

main difference is that in addition to vector $x(t)$ which schedules the match types in $\mathcal{M}$, there is also $z(t)$ which can be interpreted as scheduling the virtual self-matches in $\mathcal{N}$. After the matches are scheduled, vector $y(t)$ is determined, which indicates which match types to realize. Whenever a virtual self-match is selected by $y(t)$, instead of realizing it, the policy will set the agent(s) of the corresponding type aside, and never use them in any future matches.

Next, we show that a policy described in Algorithm 2 achieves bounded regret under suitable choice of $\{V_t\}_{t=1}^T$. The proof of the result is via constructing a virtual system with $\mathcal{M}' = \mathcal{M} \cup \mathcal{N}$, which adds all self-matches with zero rewards. Then, we claim that Algorithm 2 in the real system is effectively the same as

Algorithm 1 for the virtual system. Thus, the regret bound from Theorem 1 for Algorithm 1 would also apply to Algorithm 2.

Corollary 3. *Consider the primal-dual policy π described in Algorithm 2, with monotonically increasing $\{V_t\}_{t=1}^T$, and $\sum_{t=1}^T 1/V_t < \infty$. Suppose that λ has a GPG of ϵ , then we have*

$$\sup_{0 \leq t \leq T} \mathbb{E} [R^{h,t} - R^{\pi,t}] \leq O \left(B + \frac{\sqrt{n} B r_{\max}}{\epsilon^2} \right), \quad (53)$$

where we recall that $B = \max_{m \in \mathcal{M}} |A(m)| + 1$, and $r_{\max} = \max_{m \in \mathcal{M}} r_m$.

Proof. Consider a virtual system with $\mathcal{M}'$, where $\mathcal{M}'$ includes $\mathcal{M}$ and all self-matches not already in $\mathcal{M}$. Additionally, let $r'_m = r_m$ for each $m \in \mathcal{M}$, and $r'_i = 0$ for each agent type $i \in \mathcal{M}' \setminus \mathcal{M}$. Clearly, the virtual system satisfies Assumption 2. In addition, observe that the fluid formulation for the virtual system is equivalent to the fluid formulation of the real system. Thus, both fluid formulations have the same optimal Lagrangian dual solution and the same GPG.

Now, assuming that the real and virtual systems are observing the exact same sequence of arrivals. We can observe that at period t , if Algorithm 2 chooses $(x(t), z(t))$ such that $x_m(t) = 1$ for some $m \in \mathcal{M}$, then a match of type m is scheduled by Algorithm 1 in the virtual system; and if Algorithm 2 chooses $(x(t), z(t))$ such that $z_i(t) = 1$ for some $i \in \mathcal{M}$, then a zero reward self-match of agent type i is scheduled by Algorithm 1 in the virtual system. As a result, both Algorithm 2 and Algorithm 1 schedule the same set of matches.

Similarly, we can see that Algorithm 2 also coincides with Algorithm 1 during the realization process. Specifically, when Algorithm 1 realizes a match of type $m \in \mathcal{M}$, so does Algorithm 2; and when Algorithm 1 realizes a zero rewards self-match that is not in $\mathcal{M}$, Algorithm 2 would set an agent of the same type aside, never using the agent to realize any matches in the future. Therefore, the only difference in realized matches between Algorithm 2 and Algorithm 1 is the number of self-matches with zero rewards. Because both the real and virtual systems face the same arrival process, we can couple the arrival sequences for both systems and apply Theorem 1, and thus obtain (53). Q.E.D.

We also note that under Algorithm 2, the expected number of agents waiting in the real system may not be finite, as it is entirely possible that $O(T)$ number of agents are set aside over T periods. Interestingly, we next illustrate that the expected queue length of the critical agent types, those types where the optimal dual solution has a positive value, is finite.

Proposition 5. *Let $q_i(t)$ represent the queue length vector for agent type i under Algorithm 2, including both agents waiting to be matched and agents set aside. Let U^* be the optimal solution of the Lagrangian dual in (52), then for any i with $U_i^* > 0$, we have*

$$\mathbb{E} [q_i(t)] \leq O \left(\frac{B}{U_i^*} + \frac{\sqrt{n} B r_{\max}}{U_i^* \epsilon^2} \right).$$

Proof. Let $y(t)$ be the first d entries of $y(t)$ from Algorithm 2. Observe that for any positive integer $s > 0$,

$$\langle U^*, My(s) - \lambda \rangle \geq \langle r, y(s) \rangle - \langle U^*, \lambda \rangle = \langle r, y(s) \rangle - \langle r, x^* \rangle,$$

where the inequality follows from the fact that $M^\top U^* \geq r$ (Lemma 1), and the equality follows from strong duality. Summing the inequalities for $s = 1, \dots, t$, and taking expectations on both sides, we obtain

$$\left\langle U^*, \mathbb{E} \left[\sum_{s=1}^t My(s) \right] - t\lambda \right\rangle \geq \sum_{s=1}^t \mathbb{E} [\langle r, y(s) \rangle] - t \langle r, x^* \rangle.$$

Combining the above equation with Corollary 3 yields

$$\left\langle U^*, t\lambda - \mathbb{E} \left[\sum_{s=1}^t My(s) \right] \right\rangle \leq t \langle r, x^* \rangle - \sum_{s=1}^t \mathbb{E} [\langle r, y(s) \rangle] \leq O \left(B + \frac{\sqrt{n} B r_{\max}}{\epsilon^2} \right).$$

For each agent type i , note that $U_i^* \geq 0$ (this is because $M^\top U^* \geq r$ so that $U_i^* \geq r_i \geq 0$) and $t\lambda_i - \mathbb{E} \left[\sum_{s=1}^t My_i(s) \right] \geq 0$. Therefore, for each i where $U_i^* > 0$, we must have

$$t\lambda_i - \mathbb{E} \left[\sum_{s=1}^t My_i(s) \right] \leq O \left(\frac{B}{U_i^*} + \frac{\sqrt{n}Br_{\max}}{U_i^* \epsilon^2} \right). \quad (54)$$

Recall that $q_i(t)$ denotes the queue length vector for the real system for agent type i . Note that

$$\mathbb{E}[q_i(t)] = \mathbb{E} \left[\sum_{s=1}^t (A_i(s) - My_i(s)) \right] = t\lambda_i - \sum_{s=1}^t \mathbb{E}[My_i(s)].$$

Therefore, for any i where $U_i^* > 0$, we have

$$\mathbb{E}[q_i(t)] \leq O \left(\frac{B}{U_i^*} + \frac{\sqrt{n}Br_{\max}}{U_i^* \epsilon^2} \right).$$

Q.E.D.

Proposition 5 implies that for any i with $U_i^* > 0$, the queue length of agents of type i remains stable. We note that the proof for Proposition 5 can be extended to any policy where an admissible matching achieves a bounded regret in the virtual system. Consequently, the stability of agent types with positive values in the optimal dual solution is a necessary condition for achieving bounded regret.

C.2. Multi-Stage Learning

In Algorithm 1, the empirical dual fluid formulation is solved during every discrete period. In this section, we consider an alternative class of policies that instead of solving the fluid formulation every period, divide the entire time horizon T into multiple time intervals $\{[T_\ell, T_{\ell+1})\}_{\ell=0}^{k-1}$, where $1 = T_0 < T_1 < \dots < T_k = T + 1$, and update $\hat{U}(t)$, the empirical shadow price, only at the beginning of each time interval. Specifically, we consider the policy that follows Algorithm 1, except that $\hat{U}(t)$ is solved from (10) only in periods $T_0, \dots, T_{k-1}$, and for any $t \in [T_\ell, T_{\ell+1})$, the empirical shadow price, $\hat{U}(t)$, is set to the plug-in estimator of the optimal dual variable at time T_ℓ . We call this class of policies *multi-stage learning* primal-dual policies as the horizon is divided into k stages, and in each stage, the empirical shadow price is computed using the empirical arrival rates from the previous stages.

By Propositions 2 and 3, for any $t \in [T_\ell, T_{\ell+1})$,

$$\begin{aligned} t \langle r, x^* \rangle - \sum_{s=1}^t \mathbb{E}[\langle r, y(s) \rangle] &\leq \sum_{s=1}^t \frac{B+1}{2V_s} + r_{\max}B + (\sqrt{n}+1)r_{\max}\mathbb{E}[\|\delta(t)\|_2] + 2r_{\max}B \sum_{s=1}^t \mathbb{P}\{\hat{U}(s) \neq U^*\} \\ &\leq \sum_{s=1}^t \frac{B+1}{2V_s} + r_{\max}B + (\sqrt{n}+1)r_{\max}\mathbb{E}[\|\delta(t)\|_2] \\ &\quad + 2r_{\max}B \left(\sum_{\ell'=1}^{\ell-1} (T_{\ell'} - T_{\ell'-1}) \mathbb{P}\{\hat{U}(T_{\ell'-1}) \neq U^*\} + (t - T_{\ell-1} + 1) \mathbb{P}\{\hat{U}(T_{\ell-1}) \neq U^*\} \right). \end{aligned}$$

By Proposition 4, for any $1 \leq t \leq T$, we have

$$\mathbb{E}[\|\delta(t)\|_2] \leq 2\sqrt{B} \left(\sum_{\ell'=1}^{\ell-1} (T_{\ell'} - T_{\ell'-1}) \mathbb{P}\{\hat{U}(T_{\ell'-1}) \neq U^*\} + (t - T_{\ell-1} + 1) \mathbb{P}\{\hat{U}(T_{\ell-1}) \neq U^*\} \right) + \frac{12B}{\epsilon} + 6\sqrt{B}.$$

Then, by combining the above two displayed inequalities, we get

$$\begin{aligned} \mathbb{E}[\|\delta(t)\|_2] &\leq \sum_{s=1}^t \frac{B+1}{2V_s} + r_{\max} \left(\left((\sqrt{n}+1) \frac{12}{\epsilon} + 1 \right) B + 6(\sqrt{n}+1)\sqrt{B} \right) \\ &\quad + 2 \left((\sqrt{n}+1)\sqrt{B} + B \right) r_{\max} \sum_{\ell'=1}^{\ell-1} (T_{\ell'} - T_{\ell'-1}) \mathbb{P}\{\hat{U}(T_{\ell'-1}) \neq U^*\} \\ &\quad + 2 \left((\sqrt{n}+1)\sqrt{B} + B \right) r_{\max} (t - T_{\ell-1} + 1) \mathbb{P}\{\hat{U}(T_{\ell-1}) \neq U^*\}, \end{aligned}$$

Hence, we have

$$\begin{aligned}
\max_{1 \leq t \leq T} \left(t \langle r, x^* \rangle - \sum_{s=1}^t \mathbb{E}[\langle r, y(s) \rangle] \right) &\leq \sum_{s=1}^T \frac{B+1}{2V_s} + r_{\max} \left(\left((\sqrt{n}+1) \frac{12}{\epsilon} + 1 \right) B + 6(\sqrt{n}+1) \sqrt{B} \right) \\
&\quad + 2 \left((\sqrt{n}+1) \sqrt{B} + B \right) r_{\max} \sum_{\ell=1}^k (T_\ell - T_{\ell-1}) \mathbb{P} \left\{ \widehat{U}(T_{\ell-1}) \neq U^* \right\} \\
&\leq \sum_{s=1}^T \frac{B+1}{2V_s} + r_{\max} \left(\left((\sqrt{n}+1) \frac{12}{\epsilon} + 1 \right) B + 6(\sqrt{n}+1) \sqrt{B} \right) \\
&\quad + 4 \left((\sqrt{n}+1) \sqrt{B} + B \right) r_{\max} \sum_{\ell=1}^k (T_\ell - T_{\ell-1}) \exp(-T_{\ell-1} \epsilon^2 / 8) \\
&\leq \sum_{s=1}^T \frac{B+1}{2V_s} + r_{\max} \left(\left((\sqrt{n}+1) \frac{12}{\epsilon} + 1 \right) B + 6(\sqrt{n}+1) \sqrt{B} \right) \\
&\quad + 4 \left((\sqrt{n}+1) \sqrt{B} + B \right) r_{\max} \cdot f(\{T_\ell\}_{\ell=0}^k), \tag{55}
\end{aligned}$$

where the second inequality holds because for any $1 \leq \ell \leq k$, by (48),

$$\mathbb{P} \left\{ \widehat{U}(T_\ell) \neq U^* \right\} \leq \mathbb{P} \left\{ \left\| \widehat{\lambda}(T_\ell) - \lambda \right\|_2 \geq \epsilon \right\} \leq 2 \exp(-T_\ell \epsilon^2 / 8),$$

and the last inequality holds by defining

$$f(\{T_\ell\}_{\ell=0}^k) \triangleq \sum_{\ell=1}^k (T_\ell - T_{\ell-1}) \exp(-T_{\ell-1} \epsilon^2 / 8). \tag{56}$$

There is an inherent tradeoff between computational efficiency and regret in this multi-stage learning algorithm. The computational cost is mainly due to resolving the linear program in (10) with the current empirical arrival rates. As we reduce the number of times the empirical shadow price are updated (i.e., decrease k), we perform fewer LP resolving steps, which improves computational efficiency. However, this comes at the cost of increased regret because the algorithm is using less frequently updated information about the arrival rates. To illustrate this trade-off, let us consider the following two extreme cases.

When $k = 2$, the algorithm reduces to a two-stage learning method, requiring only one LP resolving step, and dividing the time horizon into two stages. The first stage $[1, T_1 - 1]$ collects information about arrival rates, and the second stage $[T_1, T]$ optimizes the algorithm based on the empirical arrival rates from the first stage. If $T_1 = O(1)$, then $f(\{T_\ell\}_{\ell=0}^2) = O(T)$, and the regret at any time is upper bounded by $O(T)$ according to (55). Note that any such two-stage learning algorithm with $T_1 = O(1)$ has regret at least $\Omega(T)$, because if $T_1 = O(1)$, the probability of failing to learn the optimal basis is $\Omega(1)$, leading to regret of $\Omega(T)$. If we know a priori the time horizon T and the GPG ϵ , we can minimize the regret bound (56) for the two-stage learning method with an optimal choice of $T_1 = \Theta \left(\frac{\log(\epsilon^2 T) \vee 1}{\epsilon^2} \right)$, resulting in regret of $O \left(\frac{\log(\epsilon^2 T) \vee 1}{\epsilon^2} \right)$.

In the other extreme case, where we update the arrival rates at every time step, i.e., $T_\ell = T_{\ell-1} + 1$, we have $k = T$, recovering our previous result. This yields the lowest value for the regret bound (56) but requires resolving the LP at each time step, incurring the highest computational cost.

Another approach that lies between the two extreme cases is to apply a variant of the doubling trick. Based on (56), our main idea is to set $T_\ell = \exp(T_{\ell-1} \epsilon^2 / 16)$ so that T_ℓ grows rapidly while keeping each summand in (56) small. A minor complication arises because $\exp(x \epsilon^2 / 16) > x$ only holds when $x > \frac{32}{\epsilon^2} \log(\frac{32}{\epsilon^2})$. Therefore, we initially increase T_ℓ linearly until it exceeds the threshold $\frac{32}{\epsilon^2} \log(\frac{32}{\epsilon^2})$. Specifically, the sequence for T_ℓ is constructed as follows. Without loss of generality, by Definition 1 we assume that $0 < \epsilon \leq 1$. Let $\bar{\ell} = \lceil \log \frac{32}{\epsilon^2} \rceil$. Define $T_\ell = \frac{32}{\epsilon^2} \ell \wedge T$ for $1 \leq \ell \leq \bar{\ell}$, and $T_\ell = \exp(T_{\ell-1} \epsilon^2 / 16) \wedge T$ for any $\ell > \bar{\ell}$. Let k denote the smallest integer

with $T_k = T + 1$. Then, we have $k = O(\log_a^* T) + O(\log \frac{1}{\epsilon})$, where $\log_a^*(\cdot)$ is the iterative logarithm with base $a \triangleq \exp(\epsilon^2/16)$. Moreover, $\{T_\ell\}_{\ell=1}^k$ is a monotone increasing sequence, which follows from the fact that $x \leq a^x$ for any $x \geq \frac{32}{\epsilon^2} \log \frac{32}{\epsilon^2}$. In this case, we have

$$\begin{aligned} f(\{T_\ell\}_{\ell=0}^k) &= \sum_{\ell=1}^k (T_\ell - T_{\ell-1}) \exp(-T_{\ell-1} \epsilon^2/8) \\ &\leq \frac{32}{\epsilon^2} \sum_{\ell=1}^{\bar{\ell}} \exp(-4(\ell-1)) + \sum_{\ell \geq \bar{\ell}+1} \exp(-T_{\ell-1} \epsilon^2/16) \\ &\leq O\left(\frac{1}{\epsilon^2}\right). \end{aligned}$$

Then, by (55), the regret is $O\left(B + \frac{\sqrt{n} B r_{\max}}{\epsilon^2}\right)$. This method provides a balance between computational efficiency and regret, with the number of LP resolving steps being $O(\log_a^* T) + O(\log \frac{1}{\epsilon})$. Note that this doubling trick requires the knowledge of the magnitude of ϵ .

In summary, the tradeoff between computational efficiency and regret in this multi-stage learning algorithm is primarily driven by the number of times the LP is resolved with updated arrival rate estimates. Increasing the frequency of updates reduces regret but increases computational cost, while decreasing the frequency of updates has the opposite effect.

Appendix D: Tighter regret bounds for bipartite matching with multiple agent types

In this section, we derive a more refined regret bound for our policy with unknown arrival rates for a class of dynamic bipartite matching systems. Throughout, we focus on a symmetric bipartite matching model with $2n$ distinct agent types, partitioned into a set of “demand” agents indexed by $i \in [n]$ and a set of “supply” agents indexed by $j \in [2n] \setminus [n]$. We represent the bilateral matches between the demand and supply agents using a bipartite graph $G = (n, n, E)$, where the edge set E represents the feasible match types.

We assume the arrival probability for each demand agents of type i satisfy $\lambda_i = \frac{1-\epsilon}{2n}$, for $i \in [n]$, and for each supply agents of type j satisfy $\lambda_j = \frac{1+\epsilon}{2n}$ for $j \in [2n] \setminus [n]$, for some small $\epsilon \in (0, 1)$. In vector notation, this is expressed as $\lambda = (\frac{1-\epsilon}{2n} \cdot \mathbf{1}_n, \frac{1+\epsilon}{2n} \cdot \mathbf{1}_n)$, where $\mathbf{1}_n$ denotes the n -dimensional vector of all ones. The set of all allowable matches, denoted by $\mathcal{M}$, consists of the edge set E alongside a self-match for every agent type. Consequently, $|\mathcal{M}| = |E| + 2n$. The reward vector is defined as

$$r_{ij} = 1 \quad \forall \{i, j\} \in E, \quad r_i = 0.1 \quad \forall i \in [2n]. \quad (57)$$

We note that our analysis extends to more general symmetric bipartite systems allowing for unequal partition sizes, provided that the agent types within each partition share identical arrival rates and self-match rewards, and that all bilateral matches yield the same reward. However, to simplify the exposition, we restrict our attention to the setting with $2n$ agent types and this specific choice of r .

Throughout the section, for any $S \subset [2n]$, we let $\mathcal{V}(S) = \{i \in [2n] : j \in S, \{i, j\} \in E\}$ denote its set of neighbors under graph G . We assume G satisfies the following connectivity condition.

Assumption 3. Consider a bipartite matching system with $2n$ agent types, where the bipartite graph $G = ([n], [2n] \setminus [n], E)$ represents the set of feasible bilateral matches. For any proper subset of demand agent types $S \subsetneq [n]$, we have

$$|\mathcal{V}(S)| > |S|. \quad (58)$$

Similarly, for any proper subset of supply agent types $T \subsetneq [2n] \setminus [n]$, we have

$$|\mathcal{V}(T)| > |T|. \quad (59)$$

Assumption 3 is satisfied by a wide class of graphs G , including complete bipartite graphs and those containing a Hamiltonian cycle of length $2n$. Intuitively, Assumption 3 ensures that the global demand-supply constraint is significantly tighter than the local structural constraints. This property is formally quantified in the following remark.

Remark 1. When Assumption 3 holds, we have $\sum_{j \in \mathcal{V}(S)} \lambda_j - \sum_{i \in S} \lambda_i \geq \frac{1}{2n}$ for any $S \subsetneq [n]$, and $\sum_{j \in [2n] \setminus [n]} \lambda_j - \sum_{i \in [n]} \lambda_i = \epsilon$.

We note that conditions similar to Assumption 3 are common in the classical literature on complete resource pooling and flexible servers (see, e.g., [Harrison, 1998](#); [Gurvich and Whitt, 2009](#); [Shi et al., 2019](#)). In essence, when ϵ is small, the assumption allows us to approximate the stochastic arrivals, originally a multi-dimensional vector, as a two-dimensional vector recording the aggregate arrivals on either partition, thereby facilitating our analysis.

We now proceed to formulate the fluid relaxation of the stochastic system. Recall that for a general matching system, the fluid relaxation is given by the linear program $\max_{x \geq 0: Mx = \lambda} \{r^\top x\}$, where M denotes the matching matrix and r the reward vector. In the context of our bipartite system, we define the variables as follows: for each $i \in [2n]$, let x_i denote the fluid rate of self-matches for agents of type i . Similarly, for each edge $\{i, j\} \in E$ with $i < j$, let x_{ij} denote the fluid rate of bilateral matches of type $\{i, j\} \in E$ between type $i \in [n]$ and type $j \in [2n] \setminus [n]$. The fluid relaxation can then be expressed as:

$$\begin{aligned} \max_x \quad & \sum_{\{i,j\} \in E} x_{ij} + 0.1 \cdot \sum_{i \in [2n]} x_i \\ \text{s.t.} \quad & \sum_{j \in \mathcal{V}(i)} x_{ij} + x_i = \lambda_i, \forall i \in [n], \\ & \sum_{i \in \mathcal{V}(j)} x_{ij} + x_j = \lambda_j, \forall j \in [2n] \setminus [n], \\ & x \geq 0. \end{aligned}$$

We say the bipartite matching system with graph G and arrival rate vector λ' as *oversupplied* if $\sum_{j \in \mathcal{V}(S)} \lambda'_j - \sum_{i \in S} \lambda'_i \geq 0$ for any subset $S \subseteq [n]$. Conversely, the system is *undersupplied* if $\sum_{j \in \mathcal{V}(T)} \lambda'_j - \sum_{i \in T} \lambda'_i \geq 0$ for any non-empty subset $T \subseteq [2n] \setminus [n]$. We define the system as *balanced* if it is both oversupplied and undersupplied.

By Remark 1, our bipartite matching system with arrival rates λ is oversupplied. Note that while the system with a general arrival rate vector λ' might be neither oversupplied nor undersupplied, the classification is straightforward when rates are symmetric within each partition. Specifically, if $\lambda' = (\alpha \cdot \mathbf{1}_n, \beta \cdot \mathbf{1}_n)$ for positive scalars α and β , then G is oversupplied if $\alpha \leq \beta$, and undersupplied if $\alpha \geq \beta$.

Applying duality theory, we can explicitly derive the optimal solution for the fluid relaxation when the system is in either of these regimes. Specifically, given that the current system is oversupplied, the optimal dual solution is $U^* = (0.9 \cdot \mathbf{1}_n, 0.1 \cdot \mathbf{1}_n)$. If the system were undersupplied with some arrival rate λ' , the optimal dual solution would be $\tilde{U} = (0.1 \cdot \mathbf{1}_n, 0.9 \cdot \mathbf{1}_n)$. We assume that the decision maker does not know the exact arrival rates, but possesses the prior knowledge that the arrival rates are symmetric within each partition.

At period t , let $\bar{A}(t)$ denote the cumulative arrival vector and $\hat{U}(t)$ denote the empirical estimate of the optimal dual solution. Given that the decision maker knows the arrival rates are symmetric within each partition, it estimates the optimal dual by identifying which side appears to be the bottleneck (the side with fewer arrivals). Specifically, it sets $\hat{U}(t)$ to assign the higher shadow price to the partition with the lower cumulative arrivals as follows.

$$\hat{U}(t) = \begin{cases} \tilde{U} = (0.1 \cdot \mathbf{1}_n, 0.9 \cdot \mathbf{1}_n) & \text{if } \sum_{i=1}^n \bar{A}_i(t) > \sum_{i=n+1}^{2n} \bar{A}_i(t) \\ U^* = (0.9 \cdot \mathbf{1}_n, 0.1 \cdot \mathbf{1}_n) & \text{if } \sum_{i=1}^n \bar{A}_i(t) < \sum_{i=n+1}^{2n} \bar{A}_i(t) \\ \hat{U}(t-1) & \text{if } \sum_{i=1}^n \bar{A}_i(t) = \sum_{i=n+1}^{2n} \bar{A}_i(t) \end{cases}. \quad (60)$$

Next, we define the aggregate imbalance process $R(t)$ as

$$R(t) = \sum_{i=1}^n \bar{A}_i(t) - \sum_{i=n+1}^{2n} \bar{A}_i(t).$$

Note that $R(t)$ is a biased random walk, where $R(t)$ increases by 1 with probability $\frac{1-\epsilon}{2}$ and decreases by 1 with probability $\frac{1+\epsilon}{2}$. We say that the process $R(t)$ undergoes a *positive complete cycle* during the interval $[t_s, t_e]$ if it performs a simple positive excursion starting at t_s and concluding at t_e . Formally, this requires $R(t_s - 1) = 0$, $R(t_s) = 1$, $R(t_e) = 0$, and $R(t) > 0$ for all intermediate steps $t \in \{t_s, \dots, t_e - 1\}$. Analogously, we define a *negative complete cycle* over $[t_s, t_e]$ if the process $-R(t)$ undergoes a positive cycle; that is, $R(t_s - 1) = 0$, $R(t_s) = -1$, $R(t_e) = 0$, and $R(t) < 0$ for all $t_s \leq t < t_e$. By definition of $\hat{U}(t)$, it is equal to $\tilde{U}$ during each positive cycle, and U^* during each negative cycle. Thus, whenever $\sum_{i=1}^n \bar{A}_i(t) = \sum_{i=n+1}^{2n} \bar{A}_i(t)$ (which implies $R(t) = 0$), a new cycle will start in period $t + 1$, while $\hat{U}(t)$ is the same as $\hat{U}(t - 1)$. Finally, we say that a cycle is *incomplete* if it starts at time $t = t_s$ and satisfies $R(t) \neq 0$ for all $t \geq t_s$. Note that, since $R(t)$ is a biased random walk, classical results from the theory of gambler's ruin (see e.g. (Hajek, 2015)) imply that, with probability one, exactly one negative incomplete cycle occurs as we run t from one to infinity.

Building on this structure, we formally present Algorithm 3, a primal-dual policy designed for the symmetric bipartite matching system with unknown arrival rates. This algorithm follows the general framework of Algorithm 1 established in Section 3, with one specific structural modification to aid our analysis. In Algorithm 3, we enforce a system regeneration whenever the aggregate imbalance vanishes (i.e., whenever $\sum_{i=1}^n \bar{A}_i(t) = \sum_{i=n+1}^{2n} \bar{A}_i(t)$). At such time steps, rather than carrying the state vectors forward, the algorithm resets $\delta(t)$ (and other vectors for checking the states) to the zero vector at the end of period t , effectively ignoring all of the agents and unrealized scheduled matches in the system going forward. This effectively resets the system state, ensuring that the process in period $t + 1$ begins effectively from time 0. While this forced renewal does not improve the regret numerically, it is analytically useful. It decouples the dynamics, allowing us to analyze the regret incurred within each cycle independently and therefore deriving a sharper regret bound than the $O(1/\epsilon^2)$ bound presented in Section 3.

Throughout our analysis, we assume that the tuning parameters $\{V_t\}_{t=1, \dots, T}$ satisfies $V_t > 2T$. This restricts the set of permissible self-matches based on the current cycle type, as formally stated in the following claim.

Claim 1. If $V_t > 2T$, Algorithm 3 will never schedule a self-match for agent types $i \in [n]$ during a negative cycle, nor for agent types $j \in [2n] \setminus [n]$ during a positive cycle.

Proof. Consider a negative cycle. In this regime, the dual estimate is given by $\hat{U}(t) = U^* = (0.9 \cdot \mathbf{1}_n, 0.1 \cdot \mathbf{1}_n)$. For any agent $i \in [n]$, the reduced cost of a self-match is

$$r_i - U_i(t) = r_i - \left(\hat{U}_i(t) + \frac{\delta_i(t)}{V_t} \right) = 0.1 - 0.9 - \frac{\delta_i(t)}{V_t} = -0.8 - \frac{\delta_i(t)}{V_t}.$$

Since the shortage, $\delta(t)$, are bounded by the time horizon (i.e., $-\delta_i(t) \leq T$) and we assume $V_t > 2T$, the term $-\frac{\delta_i(t)}{V_t}$ is strictly less than 0.5. Consequently, the reduced cost is strictly negative

$$r_i - U_i(t) < -0.8 + 0.5 = -0.3 < 0.$$

Since Algorithm 3 only schedules matches with non-negative adjusted rewards, it will never schedule a self-match for $i \in [n]$.

By symmetric arguments, during a positive cycle (where $\hat{U}(t) = \tilde{U}$), it will never schedule a self-match for agent $i \in [2n] \setminus [n]$. Q.E.D.

Algorithm 3 Primal-Dual Matching Policy for the Symmetric Bipartite System.

```

1: Input:  $\{V_t\}_{t=1}^\infty$ , set  $\bar{A}(0), \delta(0), q(0) \in \mathbb{R}^{2n}, x(0), W(0) \in \mathbb{R}^{|\mathcal{M}|}$  to the zero vectors.
2: for each  $t = 1, 2, \dots, T$  do
3:   Observe arrival vector  $A(t)$ , set  $\bar{A}(t) = \bar{A}(t-1) + A(t)$ .
4:   THE SCHEDULING PROCESS:
5:   Set  $\hat{U}(t)$  following (60).
6:   Set  $\delta(t) = (\delta(t-1) + Mx(t-1)) \cdot \mathbb{I}(R(t-1) \neq 0) - A(t)$ 
7:   Set  $U(t) = \hat{U}(t) + \frac{\delta(t)}{V_t}$ .
8:   Set  $r_0 = 0$  and  $M_{i0} = 0$  for all  $i \in [2n]$ . Select  $m^* \in \arg \max_{m \in \mathcal{M} \cup \{0\}} \{r_m - M^\top U(t)\}$ , breaking
      ties in favor of self-matches over zero, and zero over two-way matches.
9:   Set  $x(t)$  to the  $m^*$ -th standard basis vector in  $\mathbb{R}^{|\mathcal{M}|}$  if  $m^* > 0$ , and to  $\mathbf{0}_{|\mathcal{M}|}$  otherwise.
10:  THE REALIZATION PROCESS:
11:  Set  $q(t) = q(t-1)\mathbb{I}(R(t-1) \neq 0) + A(t)$ ,  $W(t) = W(t-1)\mathbb{I}(R(t-1) \neq 0) + x(t)$ .
12:  Initialize  $y(t) = \mathbf{0}_{|\mathcal{M}|}$ .
13:  for each  $m \in \mathcal{M}$  with  $W_m(t) > 0$  do
14:    Let  $I_m(t) = \min_{i \in \mathcal{N}(m)} q_i(t)$ .
15:    Compute  $y_m(t) = y_m(t) + \min\{W_m(t), I_m(t)\}$ .
16:    Update  $q(t) = q(t) - My(t)$ , and  $W(t) = W(t) - y(t)$ .
17:  end for
18: end for

```

In the rest of the section, we will analyze Algorithm 3 with $V_t > 2T$ and refer to it as π . We present our main result for the aforementioned bipartite matching system under policy π , which states that the regret of our policy π is $\Theta(\log(1/\epsilon)/\epsilon)$.

Theorem 2. *Consider a bipartite matching system satisfying Assumption 3, with arrival probabilities $\lambda = (\frac{1-\epsilon}{2n} \cdot \mathbf{1}_n, \frac{1+\epsilon}{2n} \cdot \mathbf{1}_n)$ and reward vector r defined in (57). Let π denote the policy specified in Algorithm 3. Then, for any $T \in \mathbb{N}$,*

$$\max_{0 \leq t \leq T} \left(\mathbb{E} \left[R^{h,t} \right] - \mathbb{E} \left[R^{\pi,t} \right] \right) = O \left(\frac{1}{\epsilon} \log \frac{1}{\epsilon} \right).$$

The $\log(1/\epsilon)/\epsilon$ upper bound in Theorem 2 is tight for primal–dual policies. Next, we show that for a bipartite dynamic matching system with $n = 2$, the regret of primal–dual policies, both with resetting (Algorithm 3) and without resetting (Algorithm 1), is at least $\Omega(\log(1/\epsilon)/\epsilon)$ whenever $T \geq 1/\epsilon^2$.

Theorem 3. *Consider a bipartite matching system with $n = 2$, and let π denote the policy defined in Algorithm 3. Then, for all $T > 0$,*

$$\max_{0 \leq t \leq T} \left(\mathbb{E} \left[R^{h,t} \right] - \mathbb{E} \left[R^{\pi,t} \right] \right) = \Omega \left(\min \left\{ \frac{1}{\epsilon}, \sqrt{T} \right\} \log \frac{1}{\epsilon} \right).$$

As we will see, the intuition underlying both our upper and lower bounds is as follows. The biased random walk $R(t)$ undergoes $\Theta(1/\epsilon)$ positive and negative cycles before diverging to $-\infty$. During each cycle, the expected

maximum imbalance is $\Theta(\log(1/\epsilon))$, and in each such cycle we schedule at least $\Theta(\log(1/\epsilon))$ self-matches. Consequently, the total number of self-matches is $\Theta(\log(1/\epsilon)/\epsilon)$, which leads to the stated regret bounds.

For regret upper bounds in the multi-type bipartite system with $n > 1$, additional complications arise because the imbalance of the arrivals no longer coincides, in general, with the shortage vector $\delta(t)$. In particular, the shortage vector $\delta(t)$ has two components: $\delta(t)_\parallel$, which captures the demand-supply imbalance of shortage, and $\delta(t)_\perp$, the orthogonal component (For $n = 1$, the orthogonal component is always zero). As a result, we must refine the regret characterization to depend jointly on the supply-demand arrival imbalance and on $\delta(t)_\perp$. We then show, via a novel drift argument, that $\|\delta(t)_\perp\|_2 = O(\log(1/\epsilon))$ uniformly over time.

Next, we focus on the proving Theorem 2, and then provide the proof for Theorem 3 in Section D.2.

D.1. Proof of Theorem 2

In this section, we provide the proof of Theorem 2. To highlight the high-level ideas, the proofs of several technical intermediate results, namely, Proposition 6, Lemma 14, and Proposition 7, are deferred to later subsections.

Suppose the imbalance process $R(t)$ has K complete cycles in total, with the k -th cycle spanning time $[t_s^{(k)}, t_e^{(k)}]$. By definition, $R(t)$ does not return to zero after time $t_e^{(K)}$. We decompose the total regret according to cycles. For any $1 \leq t_1 \leq t_2$, let $R^{h, [t_1, t_2]}$ denote the cumulative reward collected by the hindsight optimal policy during the time interval $[t_1, t_2]$ and define $R^{\pi, [t_1, t_2]}$ accordingly. Let $K(t)$ denote the total number of complete cycles up to time t , that is, $K(t) = \max\{k : t_e^{(k)} \leq t\}$. Then we have

$$R^{h, [1, t]} - R^{\pi, [1, t]} = \sum_{k=1}^{K(t)} \left(R^{h, [t_s^{(k)}, t_e^{(k)}]} - R^{\pi, [t_s^{(k)}, t_e^{(k)}]} \right) + R^{h, [t_e^{(K(t))} + 1, t]} - R^{\pi, [t_e^{(K(t))} + 1, t]}. \quad (61)$$

It suffices to bound the regret over each cycle. To facilitate the analysis, we define the post-decision state of shortage

$$\delta^+(\ell) = \delta(\ell) + Mx(\ell), \quad (62)$$

where M is the matching matrix and $x(\ell)$ represents the matching decision at time t . Intuitively, $\delta^+(\ell)$ represents the state of the system immediately after the scheduling decision but prior to the arrival at $\ell + 1$.

Define the unit direction vector $\mathbf{d} = (\mathbf{1}_n, -\mathbf{1}_n)/\sqrt{2n}$. For any vector $v \in \mathbb{R}^{2n}$, we define its projection onto the subspace spanned by $\mathbf{d}$ as $v_\parallel = \langle v, \mathbf{d} \rangle \cdot \mathbf{d}$, and the orthogonal residual as $v_\perp = v - v_\parallel$. Next, we present the following proposition, which provides an upper bound for the regret on each cycle with respect to the maximum imbalance, the norm of the shortage, $\|\delta(\ell)\|_2$, and its orthogonal residual $\|\delta^+(\ell)_\perp\|_2$.

Proposition 6. *For every cycle $1 \leq k \leq K$,*

$$\max_{t_s^{(k)} \leq \ell \leq t_e^{(k)}} \left(R^{h, [t_s^{(k)}, \ell]} - R^{\pi, [t_s^{(k)}, \ell]} \right) \leq 2 \max_{t_s^{(k)} \leq \ell \leq t_e^{(k)}} |R(\ell)| + 8n \max_{t_s^{(k)} \leq \ell \leq t_e^{(k)}} \|\delta^+(\ell)_\perp\|_2 + 2B \triangleq G_k. \quad (63)$$

Moreover, for any $t > t_e^{(K)}$,

$$R^{h, [t_e^{(K)} + 1, t]} - R^{\pi, [t_e^{(K)} + 1, t]} \leq 2\|\delta(t)\|_2 + 2B. \quad (64)$$

Combining (61) and Proposition 6, we deduce that

$$R^{h, t} - R^{\pi, t} \leq \sum_{k=1}^K G_k + 2\|\delta(t)\|_2 \mathbb{I}(t > t_e^{(K)}) + 2B,$$

Taking expectation over both sides yields

$$\begin{aligned} \mathbb{E} [R^{h, t} - R^{\pi, t}] &\leq \mathbb{E} \left[\sum_{k=1}^K G_k \right] + 2\mathbb{E} [\mathbb{I}(t > t_e^{(K)}) \|\delta(t)\|_2] + 2B \\ &= \mathbb{E} [K] \cdot \mathbb{E} [G_1] + 2\mathbb{E} [\mathbb{I}(t > t_e^{(K)}) \|\delta(t)\|_2] + 2B, \end{aligned} \quad (65)$$

where the equality holds by Wald's equation, because each completed cycle constitutes an i.i.d. renewal process, so that $G_1, G_2, \dots$ are i.i.d. and independent from K , the number of completed cycles.

The following lemma bounds the expected number of completed (positive or negative) cycles and the maximum imbalance over a completed cycle, and shows that $R(t)$ diverges to $-\infty$ after the last completed cycle.

Lemma 14. *We have $\mathbb{E}[K] = \Theta(1/\epsilon)$ and*

$$\mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \mid R(t_s^{(1)}) = 1 \right] = \mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \right] = \Theta \left(\log \frac{1}{\epsilon} \right).$$

Moreover, with probability one, $R(t)$ diverges to $-\infty$ before returning to zero after $t_e^{(K)}$.

In addition to (65) and Lemma 14, our proof relies on a key bound on $\|\delta^+(\ell)_\perp\|_2$ during a cycle, which we state next.

Proposition 7. *It holds that*

$$\mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} \|\delta^+(\ell)_\perp\|_2 \right] \leq O \left(\log \frac{1}{\epsilon} \right).$$

Now, we are ready to finish the proof of Theorem 2.

Proof of Theorem 2. Combining Lemma 14 and Proposition 7, we get that

$$\mathbb{E}[G_1] \leq O \left(\log \frac{1}{\epsilon} \right).$$

Apply the inequality above to (65) with the fact that $\mathbb{E}[K] = O(\frac{1}{\epsilon})$ from Lemma 14, we have

$$\mathbb{E} \left[R^{h,t} - R^{\pi,t} \right] \leq O \left(\frac{1}{\epsilon} \log \frac{1}{\epsilon} \right) + 2\mathbb{E} \left[\|\delta(t)\|_2 \mathbb{I}(t > t_e^{(K)}) \right]. \quad (66)$$

It remains to bound $\mathbb{E} \left[\|\delta(t)\|_2 \mathbb{I}(t > t_e^{(K)}) \right]$. Note that conditional on $t > t_e^{(K)}$, $R(t)$ no longer has iid increments. To deal with this technical issue, we couple $\delta(t)$ with an auxiliary inventory process $\delta'(t)$. In particular, let $\delta'(t)$ be the shortage in Algorithm 3 where $\widehat{U}(t)$ is always fixed to be U^* for all t . Note that $\delta(t)$ and $\delta'(t)$ are driven by the same arrival process. Since we reset the inventory to be 0 when $R(t) = 0$, we have $\delta(t) = \delta'(t) = 0$ for $t = t_e^{(K)}$. Moreover, by Lemma 14, $R(t)$ diverges to $-\infty$ before returning to zero for $t_e^{(K)}$. Therefore $\widehat{U}(t) = U^*$ for all $t > t_e^{(K)}$. It follows that with probability one, $\delta(t) = \delta'(t)$ for all $t \geq t_e^{(K)}$ and hence

$$\mathbb{E} \left[\|\delta(t)\|_2 \mathbb{I}(t > t_e^{(K)}) \right] = \mathbb{E} \left[\|\delta'(t)\|_2 \mathbb{I}(t > t_e^{(K)}) \right] \leq \mathbb{E} \left[\|\delta'(t)\|_2 \right]. \quad (67)$$

Finally, we can bound $\mathbb{E} \left[\|\delta'(t)\|_2 \right]$ via the drift analysis exactly as Proposition 4, except for minor differences due to resetting. Specifically, fix any constants η and D satisfying

$$0 < \eta < \epsilon, D \geq \eta + 1, \text{ and } B - 2(\epsilon - \eta)D \leq \eta^2.$$

For any $\ell \geq 1$, whenever $\|\delta(\ell)\|_2 \geq D$ and $\widehat{U}(\ell) = U^*$, we claim that

$$\mathbb{E} \left[\|\delta'(\ell+1)\|_2 \mid \mathcal{F}_\ell \right] \leq \|\delta'(\ell)\|_2 - \eta.$$

If $R(\ell) = 0$, we have $\delta'(\ell+1) = -A(\ell+1)$ by Algorithm 3 implying that $\|\delta'(\ell+1)\|_2 \leq 1 \leq \|\delta'(\ell)\|_2 - \eta$, as $\|\delta'(\ell)\|_2 \geq D \geq \eta + 1$; and if $R(\ell) \neq 0$, it follows from Lemma 5. Then, analagous to Proposition 4, we have $\mathbb{E} \left[\|\delta'(t)\|_2 \right] = O(\frac{1}{\epsilon})$. Combining it with (66) and (67) completes the proof.

Q.E.D.

The proof of Proposition 6 is presented in Section D.1.1, while the proof of Lemma 14 appears in Section D.1.2. Finally, the proof of Proposition 7—arguably the most technical, as it relies on a delicate drift analysis—is provided in Section D.1.3.

D.1.1. Proof of Proposition 6: Analysis of cycle regret Without loss of generality, we focus on $k = 1$. For ease of notation, we write t_e as $t_e^{(1)}$. Let $\bar{x}_s(\ell)$, $\bar{x}_b(\ell)$ denote the cumulative number of self-matches and bilateral matches scheduled up to period ℓ , respectively. The following lemma connects the accumulation of self-matches $\bar{x}_s(\ell)$ to the inventory $\|\delta^+(\ell)\|_1$ and the imbalance $|R(\ell)|$, assuming that $[1, \ell]$ is contained within the first complete or incomplete cycle. Throughout the section, we again adopt the convention that the first cycle spans the interval $[1, t_e]$, where $t_e < \infty$ denotes a complete cycle and $t_e = \infty$ denotes an incomplete one.

Lemma 15. *During the first cycle from time periods 1 through t_e , recall that $\bar{x}_s(\ell)$ denotes the number of self-matches scheduled up to period ℓ . For any $1 \leq \ell \leq t_e$,*

$$\bar{x}_s(\ell) = \|\delta^+(\ell)\|_1 + |R(\ell)|,$$

The next lemma establishes an upper bound on $\bar{x}_s(t')$ in terms of maximum imbalance $R(\ell)$ and $\|\delta^+(\ell)_\perp\|_2$ over time $1 \leq \ell \leq t'$.

Lemma 16. *For any $t' \in \{1, 2, \dots, t_e\}$, the cumulative number of self-matches at t' satisfies*

$$\bar{x}_s(t') \leq \max_{1 \leq \ell \leq t'} \{-R(\ell) + 4n \cdot \|\delta^+(\ell)_\perp\|_2\} \quad (68)$$

Now, we are ready to finish the proof of Proposition 6.

Proof of Proposition 6. Since there are in total $(\ell + R(\ell))/2$ agents of types in $[n]$ and $(\ell - R(\ell))/2$ agents of types in $[2n] \setminus [n]$ arriving up to time ℓ , and the cumulative reward collected by the hindsight optimal policy is maximized by using bilateral matches, it follows that

$$R^{h,[1,\ell]} \leq \frac{\ell - |R(\ell)|}{2} + 0.1|R(\ell)| = \frac{\ell}{2} - 0.4|R(\ell)|. \quad (69)$$

On the other hand, by Proposition 3 from Section 4, we have

$$R^{\pi,[1,\ell]} = \sum_{\kappa=1}^{\ell} r^\top y(\kappa) \geq \sum_{\kappa=1}^{\ell} r^\top x(\kappa) - \|\delta(\ell)\|_1 - B = 0.1\bar{x}_s(\ell) + \bar{x}_b(\ell) - \|\delta(\ell)\|_1 - B.$$

Plugging in the identity $\bar{x}_s(\ell) + 2\bar{x}_b(\ell) = \sum_{i \in [2n]} \bar{A}_i(\ell) + \sum_{i \in [2n]} \delta_i^+(\ell) = \ell + \sum_{i \in [2n]} \delta_i^+(\ell)$, the RHS of the last display can be rewritten as

$$\begin{aligned} R^{\pi,[1,\ell]} &= 0.1\bar{x}_s(\ell) + \frac{\ell - \bar{x}_s(\ell) + \sum_{i \in [2n]} \delta_i^+(\ell)}{2} - \|\delta(\ell)\|_1 - B \\ &= \frac{\ell}{2} - 0.4\bar{x}_s(\ell) + \frac{1}{2} \sum_{i \in [2n]} \delta_i^+(\ell) - \|\delta(\ell)\|_1 - B. \end{aligned} \quad (70)$$

Combining (69) and (70) yields that

$$R^{h,[1,\ell]} - R^{\pi,[1,\ell]} \leq 0.4\bar{x}_s(\ell) - 0.4|R(\ell)| - \frac{1}{2} \sum_{i \in [2n]} \delta_i^+(\ell) + \|\delta(\ell)\|_1 + B. \quad (71)$$

It follows that

$$\begin{aligned} R^{h,[1,\ell]} - R^{\pi,[1,\ell]} &\leq 0.4\bar{x}_s(\ell) - 0.4|R(\ell)| + 1.5 \|\delta^+(\ell)\|_1 + 2B \\ &\leq 1.9\bar{x}_s(\ell) + 2B \\ &\leq 2 \max_{1 \leq \kappa \leq \ell} \{-R(\kappa) + 4n \cdot \|\delta^+(\kappa)_\perp\|_2\} + 2B \\ &\leq G_1, \end{aligned}$$

where the first inequality holds due to $\|\delta(\ell)\|_1 \leq \|\delta^+(\ell)\|_1 + B$; the second inequality holds due to Lemma 15; and the third inequality follows from Lemma 16.

To prove (63), we assume $k = 1$ without loss of generality. Taking the maximum over $1 \leq \ell \leq t_e^{(1)}$ completes the proof of (63) for $k = 1$.

Next, we prove (64). Again, without loss of generality, let us shift the time index so that $t_e^{(K)} = 0$ (i.e., $K = 0$). Combining (71) with the identity $\bar{x}_s(\ell) = \|\delta^+(\ell)\|_1 + |R(\ell)|$ from Lemma 15, we get

$$\begin{aligned} R^{h,[1,\ell]} - R^{\pi,[1,\ell]} &\leq 0.4 (\|\delta^+(\ell)\|_1 + |R(\ell)|) - 0.4 |R(\ell)| - \frac{1}{2} \sum_{i \in [2n]} \delta_i^+(t) + \|\delta(\ell)\|_1 + B \\ &\leq 2 \|\delta(\ell)\|_1 + 2B, \end{aligned}$$

where the last inequality holds due to $\|\delta^+(\ell)\|_1 \leq \|\delta(\ell)\|_1 + B$.

Q.E.D.

It remains to prove Lemma 15 and Lemma 16. We focus on the negative cycle, as the proof for the positive cycle is completely analogous.

Properties of a negative cycle. We begin by deriving several basic properties of π during either a complete or incomplete negative cycle starting from $t = 1$. Throughout the section, we again adopt the convention that the cycle spans the interval $[1, t_e]$, where $t_e < \infty$ denotes a complete cycle and $t_e = \infty$ denotes an incomplete one.

Recall that for any $\ell \in [1, t_e]$, we have $\widehat{U}(\ell) = U^* = (0.9 \cdot \mathbf{1}_n, 0.1 \cdot \mathbf{1}_n)$. The following claim characterizes the scheduling decisions made by π upon the arrival of an agent in period ℓ during a negative cycle.

Claim 2. Consider a negative cycle spanning $[1, t_e]$, where t_e may be infinite. For each $\ell \in \{1, \dots, t_e\}$, the policy π schedules matches as follows:

- If agent $j \in [2n] \setminus [n]$ arrives:
 - If $\delta_j(\ell) \leq 0$ and $\delta_i(\ell) \geq 0$ for all neighbors $i \in \mathcal{V}(j)$, π schedules a self-match for j .
 - If $\delta_j(\ell) \leq 0$ and it has a neighbor with $\delta_i(\ell) < 0$, π schedules a bilateral match $\{i, j\}$ with the neighbor $i \in \arg \min_{i' \in \mathcal{V}(j)} \delta_{i'}(\ell)$.
 - If $\delta_j(\ell) > 0$ and there exists a neighbor i such that $\delta_i(\ell) + \delta_j(\ell) < 0$, π schedules a bilateral match $\{i, j\}$ with $i \in \arg \min_{i' \in \mathcal{V}(j)} \delta_{i'}(\ell)$.
 - If $\delta_j(\ell) > 0$ and $\delta_i(\ell) + \delta_j(\ell) \geq 0$ for all neighbors $i \in \mathcal{V}(j)$, π schedules no match.
- If agent $i \in [n]$ arrives:
 - If there exists a neighbor j such that $\delta_i(\ell) + \delta_j(\ell) < 0$, π schedules a bilateral match $\{i, j\}$ with the neighbor $j \in \arg \min_{j' \in \mathcal{V}(i)} \delta_{j'}(\ell)$.
 - Otherwise, π schedules no match.

Proof of Claim 2. Immediate from the description of Algorithm 3 and Claim 1. Q.E.D.

The following lemma establishes key properties for $\delta(\ell)$ (pre-decision) and $\delta^+(\ell)$ (post-decision) during a negative cycle.

Lemma 17. During a negative cycle spanning 1 to t_e , the following properties hold for all $1 \leq \ell \leq t_e$.

$$\delta_i(\ell) + \delta_j(\ell) \geq -1, \forall \{i, j\} \in E \quad (72)$$

$$\delta_j(\ell) \geq -1, \forall j \in [2n] \setminus [n] \quad (73)$$

$$\delta_i(\ell) \leq 0, \forall i \in [n], \quad (74)$$

$$\text{and } \delta_i^+(\ell) + \delta_j^+(\ell) \geq 0, \forall \{i, j\} \in E \quad (75)$$

$$\delta_j^+(\ell) \geq 0, \forall j \in [2n] \setminus [n] \quad (76)$$

$$\delta_i^+(\ell) \leq 0, \forall i \in [n]. \quad (77)$$

Proof of Lemma 17. We prove by induction. For the base case $\ell = 1$, (72)-(74) hold directly given $\delta(\ell) = -A(\ell)$. Moreover (75)-(77) hold because if agent of type $i \in [n]$ arrives, then we schedule a match $\{i, j\}$ for some $j \in \mathcal{V}(i)$, implying $\delta_{i'}^+(\ell) = 0$ and $\delta_{j'}^+(\ell) \geq 0$ for any $i' \in [n]$, $j' \in [2n] \setminus [n]$; if agent of type $j \in [2n] \setminus [n]$ arrives, then we schedule a self-match for j , implying $\delta_{i'}^+(\ell) = 0$ and $\delta_{j'}^+(\ell) = 0$ for any $i' \in [n]$, $j' \in [2n] \setminus [n]$.

For the inductive step, assume that conditions (75)-(77) hold for time t . Because $\delta^+(\ell)$ satisfies (75)-(77), and $\delta(\ell + 1) = \delta^+(\ell) - A(\ell + 1)$, where $A(\ell + 1)$ is a standard basis vector, the bounds (72)-(74) for $\delta(\ell + 1)$ follow immediately. We now focus on verifying (75)-(77) for $\delta^+(\ell + 1)$.

Case 1: an agent of type $j \in [2n] \setminus [n]$ arrives at time $\ell + 1$, i.e., $A(\ell + 1) = \mathbf{e}_j$. By Claim 2, and the fact $\delta^+(\ell + 1) = \delta(\ell + 1) + Mx(\ell + 1)$, the following holds.

- If we schedule a self-match, it must be of agent type j . This implies $\delta^+(\ell) = \delta^+(\ell + 1)$ and (75)-(77) holds for $\ell + 1$.
- If we schedule match $\{i, j\}$, then we must have $i \in \arg \min_{i' \in \mathcal{V}(j)} \delta_{i'}(\ell + 1)$ and $\delta_i(\ell + 1) < 0$. Because $\delta^+(\ell + 1) = \delta^+(\ell) + \mathbf{e}_i$, where $\mathbf{e}_i$ is the i -th standard basis vector. By induction hypothesis, adding the non-negative vector $\mathbf{e}_i$ preserves conditions (75) and (76). For condition (77), since $\delta_i(\ell + 1)$ must be strictly negative, we have $\delta_i^+(\ell + 1) = \delta_i(\ell + 1) + 1 \leq 0$.
- If we schedule no match, then we must have $\delta_j(\ell + 1) > 0$ and $\delta_i(\ell + 1) + \delta_j(\ell + 1) \geq 0$ for any $i \in \mathcal{V}(j)$. Because $\delta^+(\ell + 1) = \delta^+(\ell) - A(\ell + 1)$, (77) holds. Also, $\delta_j^+(\ell + 1) = \delta_j(\ell + 1) \geq 0$ and $\delta_{j'}^+(\ell + 1) = \delta_{j'}^+(\ell) \geq 0$ for all $j' \in [2n] \setminus [n]$, $j' \neq j$, so (76) holds. Moreover, since no match is scheduled, and we must have $\delta_i(\ell + 1) + \delta_j(\ell + 1) \geq 0$ for any $i \in \mathcal{V}(j)$, we have $\delta_i^+(\ell + 1) + \delta_j^+(\ell + 1) = \delta_i(\ell + 1) + \delta_j(\ell + 1) \geq 0$ for all $i \in \mathcal{V}(j)$. Finally, for any $j' \in [2n] \setminus [n]$ and $j' \neq j$, and $i' \in \mathcal{V}(j')$, $\delta_{i'}^+(\ell + 1) + \delta_{j'}^+(\ell + 1) = \delta_{i'}^+(\ell) + \delta_{j'}^+(\ell) \geq 0$ by induction hypothesis, verifying (75).

Case 2: Suppose that agent of type $i \in [n]$ arrives at time $\ell + 1$. By Claim 2, the following holds.

- If we schedule match $\{i, j\}$, then we must have $\delta_i(\ell + 1) + \delta_j(\ell + 1) < 0$ with $j \in \arg \min_{j' \in \mathcal{V}(i)} \delta_{j'}(\ell + 1)$. Similar to Case 1, adding $Mx(\ell + 1) - A(\ell + 1) = \mathbf{e}_j$ preserves lower bounds (75) and (76). For (77), we note $\delta_{i'}^+(\ell + 1) = \delta_{i'}^+(\ell) \leq 0$, for all $i' \in [n]$.
- If we schedule no match, then we must have $\delta_i(\ell + 1) + \delta_j(\ell + 1) \geq 0$ for all $j \in \mathcal{V}(i)$. Because there is no match at $\ell + 1$, $\delta_{i'}^+(\ell + 1) = \delta_{i'}^+(\ell)$ for all $i' \neq i \in [n]$ and $\delta_j^+(\ell + 1) = \delta_j^+(\ell)$ for all $j \in [2n] \setminus [n]$. Thus, we only need to check condition (75) and (77) associated with agent type i . Because $\delta^+(\ell + 1) = \delta(\ell + 1)$, condition (75) holds with agent type i . Finally, Condition (77) holds as $\delta_i^+(\ell) \leq 0$ and $\delta_i^+(\ell + 1) = \delta_i^+(\ell) - 1$.

This completes the induction. Q.E.D.

With Lemma 17, we next prove Lemma 15 and Lemma 16 when the first cycle is negative; we skip the proof for the positive cycle as it is completely analogous.

Proof of Lemma 15. By definition,

$$\delta^+(\ell) = \sum_{\kappa=1}^{\ell} Mx(\kappa) - \bar{A}(\ell).$$

For negative cycles, we have

$$-\|\delta^+(\ell)\|_1 = \langle \delta^+(\ell), (\mathbf{1}_n, -\mathbf{1}_n) \rangle = \left\langle \sum_{\kappa=1}^{\ell} Mx(\kappa), (\mathbf{1}_n, -\mathbf{1}_n) \right\rangle - \langle \bar{A}(\ell), (\mathbf{1}_n, -\mathbf{1}_n) \rangle = -\bar{x}_s(\ell) - R(\ell),$$

where the first equality holds because $\delta_j^+(\ell) \geq 0$ for $j \in [2n] \setminus [n]$ and $\delta_i^+(\ell) \leq 0$ for $i \in [n]$ in view of (76) and (77) in Lemma 17; the last equality holds by the fact that all matches are either bilateral matches or self-matches $j \in [2n] \setminus [n]$ due to Claim 2, and by the definition of $R(\ell)$.

Q.E.D.

Proof of Lemma 16. For each $j \in [2n] \setminus [n]$, define auxiliary vector $\mathbf{f}^j$ as: $\mathbf{f}^j = -\sqrt{2n} \cdot \mathbf{d} - 2n \cdot \mathbf{e}_j$. We first establish that for any $t' \in \{1, 2, \dots, t_e\}$,

$$\bar{x}_s(t') \leq \max_{1 \leq \ell \leq t'} \left\{ -R(\ell) + \max_{n+1 \leq j \leq 2n} \langle \mathbf{f}^j, \delta^+(\ell) \rangle \right\} \quad (78)$$

We now establish (78) by induction. The base case for $t' = 1$ is straightforward. For the inductive step, assume (78) holds for $t' < t_e$, (78) holds. Consider the following cases for $t' + 1$.

Case 1: $\bar{x}_s(t' + 1) = \bar{x}_s(t')$, since $RHS(t')$ is non-decreasing in t' , (78) holds for $t' + 1$.

Case 2: $\bar{x}_s(t' + 1) = \bar{x}_s(t') + 1$, and $RHS(t') > \bar{x}_s(t')$. Since the RHS of (78) is always an integer, (78) holds for $t' + 1$.

Case 3: $\bar{x}_s(t' + 1) = \bar{x}_s(t') + 1$, and $RHS(t') = \bar{x}_s(t')$. Then, let $j' \in [2n] \setminus [n]$ be the agent type that arrived at $t + 1$ and was self-matched. By Claim 2 and Lemma 17, we must have $\delta_{j'}^+(t') = 0$, implying $\langle \mathbf{f}^{j'}, \delta^+(t') \rangle = -\sqrt{2n} \langle \mathbf{d}, \delta^+(t') \rangle$. Thus, we have

$$\begin{aligned} -R(t') + \langle \mathbf{f}^{j'}, \delta^+(t') \rangle &= -R(t') - \sqrt{2n} \langle \mathbf{d}, \delta^+(t') \rangle \\ &= -R(t') - \sum_{i=1}^n \delta_i^+(t') + \sum_{j=n+1}^{2n} \delta_j^+(t') \\ &= \bar{x}_s(t'), \end{aligned}$$

where the last equality follows from Lemma 15. Because $j' \in [2n] \setminus [n]$ be the agent type that arrived at $t' + 1$ and was self-matched, we also have

$$\bar{x}_s(t' + 1) = -R(t' + 1) + \langle \mathbf{f}^{j'}, \delta^+(t' + 1) \rangle \leq \max_{1 \leq \ell \leq t'+1} \left\{ -R(\ell) + \max_{n+1 \leq j \leq 2n} \langle \mathbf{f}^j, \delta^+(\ell) \rangle \right\}.$$

By induction, we have (78) for any $t' \in \{1, 2, \dots, t_e\}$.

To prove (68), observe that $\mathbf{f}^j$ is orthogonal to the drift vector $\mathbf{d}$. Specifically, for any $j \in [2n] \setminus [n]$,

$$\langle \mathbf{f}^j, \mathbf{d} \rangle = -\sqrt{2n} \langle \mathbf{d}, \mathbf{d} \rangle - 2n \langle \mathbf{e}_j, \mathbf{d} \rangle = -\sqrt{2n}(1) - 2n \left(\frac{-1}{\sqrt{2n}} \right) = -\sqrt{2n} + \sqrt{2n} = 0.$$

Since $\mathbf{f}^j \perp \mathbf{d}$, we have $\langle \mathbf{f}^j, \delta^+(\ell) \rangle = \langle \mathbf{f}^j, \delta^+(\ell)_\perp \rangle$. Applying the Cauchy-Schwarz inequality:

$$\langle \mathbf{f}^j, \delta^+(\ell) \rangle \leq \|\mathbf{f}^j\|_2 \|\delta^+(\ell)_\perp\|_2.$$

Given $\|\mathbf{f}^j\|_2 = \sqrt{4n^2 - 2n} \leq 4n$, the result follows. Q.E.D.

D.1.2. Proof of Lemma 14

Proof of Lemma 14. The proof follows from the standard analysis of a biased random walk. Recall that $R(t) = \sum_{i=1}^n \bar{A}_i(t) - \sum_{i=n+1}^{2n} \bar{A}_i(t)$ is a biased random walk, where $R(t)$ increases by 1 with probability $\frac{1-\epsilon}{2}$ and decreases by 1 with probability $\frac{1+\epsilon}{2}$. Let $\rho = \frac{1+\epsilon}{1-\epsilon}$ and $m \geq 1$ be an integer. Recall the classical Gambler's ruin formula (see, e.g. (Hajek, 2015, page 114)): letting $a, b > 0$ be two integers,

$$\mathbb{P}\{R(t) \text{ starts at } 0 \text{ and reaches } a \text{ before reaching } -b\} = \frac{\rho^b - 1}{\rho^{a+b} - 1}. \quad (79)$$

Therefore,

$$\mathbb{P}\{R(t) \text{ starts at } 1 \text{ and reaches } m \text{ before reaching } 0\} = \frac{\rho - 1}{\rho^m - 1} \quad (80)$$

Analogously,

$$\mathbb{P}\{R(t) \text{ starts at } -1 \text{ and reaches } -m \text{ before reaching } 0\} = \frac{\rho^{-1} - 1}{\rho^{-m} - 1}.$$

Hence, by taking $m \rightarrow \infty$, we deduce that $R(t)$ starts at 1 and diverges to $+\infty$ before returning to 0 with zero probability, while $R(t)$ starts at -1 and diverges to $-\infty$ before returning to 0 with probability $1 - \rho^{-1} = \frac{2\epsilon}{1+\epsilon} \triangleq \epsilon'$. As a consequence, if $R(t)$ starts from 0, it will generate a positive cycle (resp. negative cycle) with probability $\frac{1-\epsilon}{2}$ (resp. $\frac{1-\epsilon}{2}$), and will diverge with probability $\epsilon' \times \frac{1+\epsilon}{2} = \epsilon$. Thus, the total number of cycles K follows a geometric distribution $K \sim \text{Geo}(\epsilon)$. Consequently, the expected number of complete cycles is $\mathbb{E}[K] = \frac{1-\epsilon}{\epsilon} = \Theta(\frac{1}{\epsilon})$.

Next, we bound the maximum imbalance over a complete cycle. By symmetry, $R(t)$ over a positive complete cycle has the same distribution as $-R(t)$ over a negative complete cycle. Therefore, it follows from (80) that

$$\mathbb{P} \left\{ \max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \geq m \mid R(t_s^{(1)}) = -1 \right\} = \mathbb{P} \left\{ \max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \geq m \mid R(t_s^{(1)}) = 1 \right\} = \frac{\rho - 1}{\rho^m - 1}. \quad (81)$$

Therefore,

$$\mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \right] = \sum_{m=1}^{\infty} \mathbb{P} \left\{ \max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \geq m \right\} = \sum_{m=1}^{\infty} \frac{\rho - 1}{\rho^m - 1}. \quad (82)$$

By letting $M = \lfloor 1/(\rho - 1) \rfloor$,

$$\begin{aligned} \mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \right] &= \sum_{m=1}^M \frac{\rho - 1}{\rho^m - 1} + \sum_{m=M+1}^{\infty} \frac{\rho - 1}{\rho^m - 1} \\ &\leq \sum_{m=1}^M \frac{1}{m} + \sum_{m=M+1}^{\infty} \frac{\rho^M (\rho - 1)}{(\rho^M - 1) \rho^m} \\ &\leq 1 + \ln M + \frac{1}{\rho^M - 1} = O(\log(1/\epsilon)), \end{aligned}$$

where the first inequality uses $\rho^m - 1 \geq m(\rho - 1)$ (Bernoulli's inequality) for the first term, and $\rho^m - 1 > \frac{\rho^M - 1}{\rho^M} \rho^m$ for the second term given $m \geq M$.

Note that for $m \leq M$, we have $\rho^m - 1 - em(\rho - 1) \leq 0$ for $1 \leq m \leq \lfloor 1/(\rho - 1) \rfloor$ which follows by the convexity of $f(x) = \rho^x - 1 - ex(\rho - 1)$ on x , and the fact that $f(x) \leq 0$ for $x = 1$ and $x = \lfloor 1/(\rho - 1) \rfloor$. Then, for the lower bound, we have

$$\mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \right] \geq \sum_{m=1}^M \frac{\rho - 1}{\rho^m - 1} \geq \sum_{m=1}^M \frac{1}{em} = \Omega(\log(1/\epsilon)).$$

Combining the above three displayed inequalities, we conclude that

$$\mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \right] = \sum_{m=1}^{\infty} \frac{\rho - 1}{\rho^m - 1} = \Theta(\log(1/\epsilon))$$

Moreover, by (81) and (82), we have

$$\mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \mid R(t_s^{(1)}) = 1 \right] = \mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \right].$$

Hence, our desired results follow.

Q.E.D.

D.1.3. Proof of Proposition 7 Without loss of generality, we focus on $k = 1$. For ease of notation, we write t_e as $t_e^{(1)}$. The proof is based on a delicate drift analysis, showing that $\|\delta^+(\ell)_\perp\|_2$ has a negative drift, independent of ϵ .

Recall the Lagrangian dual function defined as $L_\lambda(U) \triangleq \max_{x \in \mathcal{X}} \langle r - M^\top U, x \rangle + \langle U, \lambda \rangle$. The following result establishes the stability of the optimal dual U^* .

Lemma 18. Under Assumption 3, U^* is an optimal solution to $\min_U L_{\lambda'}(U)$ for all $\lambda' \in \mathbb{R}^{2n}$ such that $\|(\lambda' - \lambda)_{\parallel}\| \leq \frac{\epsilon}{\sqrt{2n}}$ and $\|(\lambda' - \lambda)_{\perp}\| \leq \frac{1}{4n\sqrt{2n}}$, for any $\epsilon \leq \frac{1}{4n}$.

Proof. By complementary slackness, U^* remains an optimal dual solution if there exists x' such that

$$\begin{aligned} \sum_{j \in N(i)} x'_{ij} &= \lambda'_i, \forall i \in [n], \\ \sum_{i \in N(j)} x'_{ij} + x'_j &= \lambda'_j, \forall j \in [2n] \setminus [n], \\ x' &\geq 0. \end{aligned}$$

The condition above is equivalent the existence of a flow of value $\sum_{i \in [n]} \lambda'_i$ where each $i \in [n]$ act as a source node with supply λ'_i and each $j \in [2n] \setminus [n]$ act as a sink node with demand of at most λ'_j . Thus, by max-flow min-cut theorem, U^* remains an optimal dual solution if and only if

$$\sum_{j \in \mathcal{V}(S)} \lambda'_j - \sum_{i \in S} \lambda'_i \geq 0, \forall S \subset [n]. \quad (83)$$

For $S \subsetneq [n]$, we have

$$\sum_{j \in \mathcal{V}(S)} \lambda'_j - \sum_{i \in S} \lambda'_i \geq \sum_{j \in \mathcal{V}(S)} \lambda_j - \sum_{i \in S} \lambda_i - \|\lambda' - \lambda\|_1 \geq \frac{1}{2n} - \|\lambda' - \lambda\|_1, \quad (84)$$

as the last inequality holds by Remark 1 and the values of λ . Next, observe that

$$\|\lambda' - \lambda\|_1 \leq \sqrt{2n} \|\lambda' - \lambda\| \leq \sqrt{2n} \|(\lambda' - \lambda)_{\parallel}\| + \sqrt{2n} \|(\lambda' - \lambda)_{\perp}\| \leq \epsilon + \frac{1}{4n}, \quad (85)$$

given that $\epsilon \leq 1$. Combining the two inequalities above, we have

$$\sum_{j \in \mathcal{V}(S)} \lambda'_j - \sum_{i \in S} \lambda'_i \geq \frac{1}{2n} - \epsilon - \frac{1}{4n} \geq 0, \quad (86)$$

where the last inequality holds because $\epsilon \leq \frac{1}{4n}$.

Finally, when $S = [n]$, we have $\mathcal{V}(S) = [2n] \setminus [n]$ and

$$\sum_{i=1}^n \lambda'_i - \sum_{i=n+1}^{2n} \lambda'_i = \left[\sum_{i=1}^n (\lambda' - \lambda)_i - \sum_{i=n+1}^{2n} (\lambda' - \lambda)_i \right] + \left[\sum_{i=1}^n \lambda_i - \sum_{i=n+1}^{2n} \lambda_i \right] \leq \sqrt{2n} \|(\lambda' - \lambda)_{\parallel}\| - \epsilon \leq 0,$$

where the second equality holds by assumption $\sum_{i=1}^n \lambda_i - \sum_{i=n+1}^{2n} \lambda_i = -\epsilon$, and $\sum_{i=1}^n (\lambda' - \lambda)_i - \sum_{i=n+1}^{2n} (\lambda' - \lambda)_i = \sqrt{2n} \langle \lambda' - \lambda, \mathbf{d} \rangle \leq \sqrt{2n} \|(\lambda' - \lambda)_{\parallel}\|$ and the last inequality holds by assumption that $\|(\lambda' - \lambda)_{\parallel}\| \leq \frac{\epsilon}{\sqrt{2n}}$.

Q.E.D.

Now, with Lemma 18, we can prove the following locally polyhedral conditions at U^* and $\tilde{U}$.

Lemma 19. For any $U \in \mathbb{R}^n$ and any $\eta \leq \frac{1}{4n\sqrt{2n}}$, we have

$$\begin{aligned} L_{\lambda}(U) - L_{\lambda}(U^*) &\geq \frac{\epsilon}{\sqrt{2n}} \|(U - U^*)_{\parallel}\| + \eta \|(U - U^*)_{\perp}\|, \\ L_{\lambda}(U) - L_{\lambda}(\tilde{U}) &\geq \frac{\epsilon}{\sqrt{2n}} \|(U - \tilde{U})_{\parallel}\| + \eta \|(U - \tilde{U})_{\perp}\| - \epsilon \sqrt{\frac{2}{n}} \langle U - \tilde{U}, \mathbf{d} \rangle. \end{aligned}$$

Proof. Fix some $U \in \mathbb{R}^n$. Choose λ' such that $\|(\lambda' - \lambda)_{\parallel}\| = \frac{\epsilon}{\sqrt{2n}}$, $\|(\lambda' - \lambda)_{\perp}\| = \eta \leq \frac{1}{4n\sqrt{2n}}$, and

$$\langle U - U^*, \lambda' - \lambda \rangle = \langle (U - U^*)_{\parallel}, (\lambda' - \lambda)_{\parallel} \rangle + \langle (U - U^*)_{\perp}, (\lambda' - \lambda)_{\perp} \rangle = -\frac{\epsilon}{\sqrt{2n}} \|(U - U^*)_{\parallel}\| - \eta \|(U - U^*)_{\perp}\|.$$

It follows that

$$\begin{aligned}
L_\lambda(U) &= L_{\lambda'}(U) + \langle U, \lambda - \lambda' \rangle \\
&\geq L_{\lambda'}(U^*) + \langle U, \lambda - \lambda' \rangle \\
&= L_\lambda(U^*) + \langle U^*, \lambda' - \lambda \rangle + \langle U, \lambda - \lambda' \rangle \\
&= L_\lambda(U^*) + \frac{\epsilon}{\sqrt{2n}} \|(U - U^*)_\parallel\| + \eta \|(U - U^*)_\perp\|.
\end{aligned}$$

where the second equality follows from Lemma 18.

Next, define $\tilde{\lambda}$ such that

$$\tilde{\lambda}_i = \begin{cases} \lambda_{i+n} & 1 \leq i \leq n \\ \lambda_{i-n} & n+1 \leq i \leq 2n \end{cases}.$$

Since $\tilde{U}$ is an optimal solution of $\min_U L_{\tilde{\lambda}}(U)$, analogous to the above argument, we have

$$L_{\tilde{\lambda}}(U) - L_{\tilde{\lambda}}(\tilde{U}) \geq \frac{\epsilon}{\sqrt{2n}} \|(U - \tilde{U})_\parallel\| + \eta \|(U - \tilde{U})_\perp\|.$$

Therefore,

$$L_\lambda(U) - L_\lambda(\tilde{U}) = L_{\tilde{\lambda}}(U) - L_{\tilde{\lambda}}(\tilde{U}) + \langle U - \tilde{U}, \lambda - \tilde{\lambda} \rangle \geq \frac{\epsilon}{\sqrt{2n}} \|(U - \tilde{U})_\parallel\| + \eta \|(U - \tilde{U})_\perp\| - \epsilon \sqrt{\frac{2}{n}} \langle U - \tilde{U}, \mathbf{d} \rangle,$$

where the last equality holds due to $\lambda - \tilde{\lambda} = -\epsilon \sqrt{\frac{2}{n}} \mathbf{d}$, given that $\mathbf{d} = (\mathbf{1}_n, -\mathbf{1}_n)/\sqrt{2n}$. Q.E.D.

Next, leveraging Lemma 19, we establish that $\|\delta^+(\ell)_\perp\|$ has a negative drift when $\|\delta^+(\ell)_\perp\| \geq D$ for some suitably chosen large D . Note that

$$\|\delta^+(\ell)_\perp\| - \|\delta^+(\ell-1)_\perp\| \leq \frac{\|\delta^+(\ell)_\perp\|^2 - \|\delta^+(\ell-1)_\perp\|^2}{2\|\delta^+(\ell-1)_\perp\|}. \quad (87)$$

The numerator can be rewritten as

$$\|\delta^+(\ell)_\perp\|^2 - \|\delta^+(\ell-1)_\perp\|^2 = \|\delta^+(\ell)\|^2 - \|\delta^+(\ell-1)\|^2 - \left(\|\delta^+(\ell)_\parallel\|^2 - \|\delta^+(\ell-1)_\parallel\|^2 \right). \quad (88)$$

Thus we bound the drift of $\|\delta^+(\ell-1)\|^2$ and that of $\|\delta^+(\ell-1)_\parallel\|^2$ separately. The following lemma bounds from above the drift of $\|\delta^+(\ell-1)\|^2$.

Lemma 20. *It holds that for any $\ell \geq 1$,*

$$\begin{aligned}
&\mathbb{E} \left[\|\delta^+(\ell)\|^2 - \|\delta^+(\ell-1)\|^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] \\
&\leq -2 \left(\frac{\epsilon}{\sqrt{2n}} \|\delta(\ell)_\parallel\| + \eta \|\delta(\ell)_\perp\| \right) + 2\epsilon \sqrt{\frac{2}{n}} \langle \delta(\ell), \mathbf{d} \rangle \mathbb{I}(\hat{U}(\ell) = \tilde{U}) + 4B
\end{aligned}$$

Proof. By (62), $\delta^+(\ell) = \delta^+(\ell-1) - A(\ell) + Mx(\ell)$ for any $\ell \geq 1$ where $\delta^+(0) = \mathbf{0}$. Then, we get

$$\begin{aligned}
\|\delta^+(\ell)\|_2^2 - \|\delta^+(\ell-1)\|_2^2 &= \|Mx(\ell) - A(\ell)\|_2^2 + 2\langle \delta^+(\ell-1), Mx(\ell) - A(\ell) \rangle \\
&\stackrel{(a)}{\leq} B + 2\langle \delta(\ell) - A(\ell), Mx(\ell) - A(\ell) \rangle \\
&\stackrel{(b)}{\leq} 2B + 2\langle \delta(\ell), Mx(\ell) - A(\ell) \rangle \\
&\leq 2B + 2\langle \delta(\ell), Mx(\ell) - \lambda \rangle + 2\langle \delta(\ell), \lambda - A(\ell) \rangle \\
&\stackrel{(c)}{\leq} 2B + 2\langle \delta(\ell), Mx(\ell) - \lambda \rangle + 2\langle \delta(\ell-1), \lambda - A(\ell) \rangle - 2\langle A(\ell) - Mx(\ell-1), \lambda - A(\ell) \rangle \\
&\stackrel{(d)}{\leq} 4B + 2\langle \delta(\ell), Mx(\ell) - \lambda \rangle + 2\langle \delta(\ell-1), \lambda - A(\ell) \rangle, \quad (89)
\end{aligned}$$

where (a), (b) and (d) holds by $B = \max_{m \in \mathcal{M}} |\mathcal{N}(m)| + 1$, and (c) holds because $\delta(\ell) = \delta(\ell-1) - A(\ell) + Mx(\ell-1)$, given $R(\ell-1) \neq 0$. Then, we have

$$\mathbb{E} \left[\|\delta^+(\ell)\|_2^2 - \|\delta^+(\ell-1)\|_2^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] \leq 4B + 2V_t \left\langle U(\ell) - \widehat{U}(\ell), Mx(\ell) - \lambda \right\rangle.$$

If $\widehat{U}(\ell) = U^*$, we get

$$\begin{aligned} \mathbb{E} \left[\|\delta^+(\ell)\|_2^2 - \|\delta^+(\ell-1)\|_2^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] &\leq 4B + 2V_t \left\langle U(\ell) - U^*, Mx(\ell) - \lambda \right\rangle \\ &\leq 4B + 2V_t (L_\lambda(U^*) - L_\lambda(U(\ell))) \\ &\leq 4B - 2V_t \left(\frac{\epsilon}{\sqrt{2n}} \|(U(\ell) - U^*)_\parallel\| + \eta \|(U(\ell) - U^*)_\perp\| \right) \\ &= 4B - 2 \left(\frac{\epsilon}{\sqrt{2n}} \|\delta(\ell)\|_\parallel + \eta \|\delta(\ell)\|_\perp \right), \end{aligned}$$

where the second inequality holds by (29), and the last inequality hold by Lemma 19.

Analogous to (29), we can get

$$\left\langle U(\ell) - \widetilde{U}, Mx(\ell) - \lambda \right\rangle \leq L_\lambda(U^*) - L_\lambda(\widetilde{U}),$$

which follows from the optimality of $x(\ell)$ given in (12),

$$L_\lambda(U(\ell)) = \langle r, x(\ell) \rangle + \langle U(\ell), \lambda - Mx(\ell) \rangle,$$

and by the feasibility of $x(\ell)$,

$$L_\lambda(\widetilde{U}) \geq \langle r, x(\ell) \rangle + \left\langle \widetilde{U}, \lambda - Mx(\ell) \right\rangle.$$

Then, if $U(\ell) = \widetilde{U}$, we get

$$\begin{aligned} \mathbb{E} \left[\|\delta^+(\ell)\|_2^2 - \|\delta^+(\ell-1)\|_2^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] &\leq 4B + 2V_t (L_\lambda(U^*) - L_\lambda(\widetilde{U})) \\ &\leq 4B - 2 \left(\frac{\epsilon}{\sqrt{2n}} \|\delta(\ell)\|_\parallel + \eta \|\delta(\ell)\|_\perp - \epsilon \sqrt{\frac{2}{n}} \langle \delta(\ell), \mathbf{d} \rangle \right), \end{aligned}$$

where the last inequality hold by Lemma 19. Q.E.D.

We then bound from below the drift of $\|\delta^+(\ell)\|_\parallel^2$.

Lemma 21. *It holds that*

$$\mathbb{E} \left[\|\delta^+(\ell)\|_\parallel^2 - \|\delta^+(\ell-1)\|_\parallel^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] \geq \epsilon \sqrt{\frac{2}{n}} \langle \delta(\ell), \mathbf{d} \rangle - 1.$$

Proof. Condition on $R(\ell-1) \neq 0$. Then, by (62), $\delta^+(\ell) = \delta^+(\ell-1) - A(\ell) + Mx(\ell)$. It follows that

$$\begin{aligned} \|\delta^+(\ell)\|_\parallel^2 - \|\delta^+(\ell-1)\|_\parallel^2 &= \langle \delta^+(\ell), \mathbf{d} \rangle^2 - \langle \delta^+(\ell-1), \mathbf{d} \rangle^2 \\ &= (\langle \delta^+(\ell-1), \mathbf{d} \rangle + \langle Mx(\ell) - A(\ell), \mathbf{d} \rangle)^2 - \langle \delta^+(\ell-1), \mathbf{d} \rangle^2 \\ &\geq 2 \langle \delta^+(\ell-1), \mathbf{d} \rangle \langle Mx(\ell) - A(\ell), \mathbf{d} \rangle. \end{aligned}$$

Let E_ℓ denote the event that the scheduled match at time ℓ is not a self-match. Since $\delta^+(0) = 0$, if a self-match is used, then $\delta^+(\ell) = \delta^+(\ell-1)$ because self-matches will always match the newly arriving agent. Therefore,

$$\|\delta^+(\ell)\|_\parallel^2 - \|\delta^+(\ell-1)\|_\parallel^2 = \left(\|\delta^+(\ell)\|_\parallel^2 - \|\delta^+(\ell-1)\|_\parallel^2 \right) \mathbb{I}(E_\ell).$$

It follows that

$$\begin{aligned}
& \mathbb{E} \left[\left\| \delta^+(\ell) \right\|^2 - \left\| \delta^+(\ell-1) \right\|^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] \\
&= \mathbb{E} \left[\left(\left\| \delta^+(\ell) \right\|^2 - \left\| \delta^+(\ell-1) \right\|^2 \right) \mathbb{I}(E_\ell) \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] \\
&\geq 2\mathbb{E} \left[\langle \delta^+(\ell-1), \mathbf{d} \rangle \langle Mx(\ell) - A(\ell), \mathbf{d} \rangle \mathbb{I}(E_\ell) \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] \\
&\stackrel{(a)}{\geq} -2 \langle \delta^+(\ell-1), \mathbf{d} \rangle \langle \mathbb{E}[A(\ell) \mathbb{I}(E_\ell) \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0], \mathbf{d} \rangle \\
&= -2 \langle \delta^+(\ell-1), \mathbf{d} \rangle \langle \mathbb{E}[A(\ell) \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0] - \mathbb{E}[A(\ell) \mathbb{I}(E_\ell^c) \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0], \mathbf{d} \rangle \\
&\stackrel{(b)}{\geq} -2 \langle \delta^+(\ell-1), \mathbf{d} \rangle \langle \lambda, \mathbf{d} \rangle = \epsilon \sqrt{\frac{2}{n}} \langle \delta^+(\ell-1), \mathbf{d} \rangle \\
&\stackrel{(c)}{\geq} \epsilon \sqrt{\frac{2}{n}} \langle \delta(\ell), \mathbf{d} \rangle - 1,
\end{aligned}$$

where (a) holds because on event E_ℓ , only two-way matching is conducted or no match is conducted, and then $\langle Mx(\ell), \mathbf{d} \rangle = 0$; (b) holds because on event E_ℓ^c , if in a negative cycle, $\langle A(\ell), \mathbf{d} \rangle \leq 0$, and $\langle \delta^+(\ell-1), \mathbf{d} \rangle \leq 0$ by Lemma 17, and then by symmetry, if in a positive cycle, $\langle A(\ell), \mathbf{d} \rangle \geq 0$, and $\langle \delta^+(\ell-1), \mathbf{d} \rangle \geq 0$; (c) holds because $\delta^+(\ell-1) = \delta(\ell-1) + Mx(\ell-1) = \delta(\ell) + A(\ell)$ by (62).

Q.E.D.

Combining the above two lemmas, we derive a desired upper bound to the drift of $\left\| \delta^+(\ell)_\perp \right\|^2$.

Lemma 22. *It holds that*

$$\mathbb{E} \left[\left\| \delta^+(\ell)_\perp \right\|^2 - \left\| \delta^+(\ell-1)_\perp \right\|^2 \mid \mathcal{F}_{\ell-1} \right] \leq -2\eta \left\| \delta^+(\ell-1)_\perp \right\| + 4B + 2,$$

for any $\eta \leq \frac{1}{4n\sqrt{n}}$. In particular, if $\left\| \delta^+(\ell-1)_\perp \right\| \geq (4B+2)/\eta$, we have

$$\mathbb{E} \left[\left\| \delta^+(\ell)_\perp \right\| - \left\| \delta^+(\ell-1)_\perp \right\| \mid \mathcal{F}_{\ell-1} \right] \leq -\eta/2.$$

Proof. Applying Lemma 20 and Lemma 21 to (88), we have

$$\begin{aligned}
& \mathbb{E} \left[\left\| \delta^+(\ell)_\perp \right\|^2 - \left\| \delta^+(\ell-1)_\perp \right\|^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) \neq 0 \right] \\
&\leq -2 \left(\frac{\epsilon}{\sqrt{2n}} \left\| \delta(\ell)_\perp \right\| + \eta \left\| \delta(\ell)_\perp \right\| \right) + \epsilon \sqrt{\frac{2}{n}} \langle \delta(\ell), \mathbf{d} \rangle \left(2\mathbb{I}(\widehat{U}(\ell-1) = \widetilde{U}) - 1 \right) + 4B + 1 \\
&\leq -2\eta \left\| \delta(\ell)_\perp \right\| + 4B + 1 \\
&\leq -2\eta \left\| \delta^+(\ell-1)_\perp \right\| + 4B + 2,
\end{aligned}$$

where the last inequality holds given that $\delta^+(\ell-1) = \delta(\ell) + A(\ell)$ by (62) and $\eta \leq \frac{1}{4n\sqrt{2n}} \leq \frac{1}{4}$. Moreover, we have

$$\mathbb{E} \left[\left\| \delta^+(\ell)_\perp \right\|^2 - \left\| \delta^+(\ell-1)_\perp \right\|^2 \mid \mathcal{F}_{\ell-1}, R(\ell-1) = 0 \right] \leq 1 - \left\| \delta^+(\ell-1)_\perp \right\|^2 \leq -2\eta \left\| \delta^+(\ell-1)_\perp \right\| + 4B + 2,$$

given $\eta \leq \frac{1}{4n\sqrt{2n}} \leq \frac{1}{2}$ and $B \geq 1$. Then, by combining the above two displayed inequalities, we obtain

$$\mathbb{E} \left[\left\| \delta^+(\ell)_\perp \right\|^2 - \left\| \delta^+(\ell-1)_\perp \right\|^2 \mid \mathcal{F}_{\ell-1} \right] \leq -2\eta \left\| \delta^+(\ell-1)_\perp \right\| + 4B + 2.$$

By above inequality and (87), it follows that

$$\mathbb{E} \left[\left\| \delta^+(\ell)_\perp \right\| - \left\| \delta^+(\ell-1)_\perp \right\| \mid \mathcal{F}_{\ell-1} \right] \leq -\eta + \frac{4B+2}{2\left\| \delta^+(\ell-1)_\perp \right\|} \leq -\eta/2,$$

where the last inequality holds under the assumption that $\left\| \delta^+(\ell-1)_\perp \right\| \geq (4B+2)/\eta$. Q.E.D.

Based on the negative drift of $\left\| \delta^+(\ell)_\perp \right\|$, we can derive a uniform upper bound to $\mathbb{E} \left[\max_{\ell \leq L} \left\| \delta^+(\ell)_\perp \right\| \right]$ for any fixed time L .

Lemma 23. *For any fixed time $L > 0$, it holds that*

$$\mathbb{E} \left[\max_{1 \leq \ell \leq L} \|\delta^+(\ell)_\perp\| \right] \leq \sqrt{B} + \frac{4B+2}{\eta} + \frac{2\sqrt{B}}{\eta} \log(eL).$$

Proof. Note that

$$\left| \|\delta^+(\ell+1)_\perp\| - \|\delta^+(\ell)_\perp\| \right| \leq \|\delta^+(\ell+1)_\perp - \delta^+(\ell)_\perp\| \leq \|\delta^+(\ell+1) - \delta^+(\ell)\| \leq \sqrt{B} \triangleq K.$$

Furthermore, by Lemma 22,

$$\mathbb{E} [\|\delta^+(\ell+1)_\perp\| - \|\delta^+(\ell)_\perp\| \mid \mathcal{F}_t] \leq -\eta/2, \quad \text{when } \|\delta^+(\ell)_\perp\| \geq D \triangleq \frac{4B+2}{\eta}.$$

Thus, following the proof of Lemma 6, let $\Gamma(\ell)$ denote a reflected random walk with step size $K = \sqrt{B}$, reflected boundary $\Gamma_{\min} = (1 + \lceil \frac{D}{K} \rceil) K$, initial condition $\Gamma(0) = \Gamma_{\min}$, and

$$\Gamma(\ell+1) = \max \{ \Gamma(\ell) + \xi(\ell)K, \Gamma_{\min} \}, \quad \text{for } \ell \geq 0,$$

where $\xi(\ell)$ are i.i.d. random variables taking the value $+1$ with probability $p = \frac{1}{2} - \frac{\eta}{4K}$ and -1 with probability $q = \frac{1}{2} + \frac{\eta}{4K}$. By (Gupta, 2024, Lemma 3), $\{\|\delta^+(\ell)_\perp\|\}_{\ell \in [1, L]} \leq_{\text{icx}} \{\Gamma(\ell)\}_{\ell \in [0, L]}$. Since $\max_{1 \leq \ell \leq L} x(\ell)$ is a convex function of vector $\{x(\ell)\}_{\ell \in [0, L]}$, it follows that

$$\mathbb{E} \left[\max_{\ell \in [1, L]} \|\delta^+(\ell)_\perp\| \right] \leq \mathbb{E} \left[\max_{\ell \in [1, L]} \Gamma(\ell) \right]$$

By (Cox and Miller, 2017, Section 2.2, Equation (52)), $\{\Gamma(\ell)\}$ has a unique steady state distribution:

$$\mathbb{P} \{ \Gamma_\infty \geq \Gamma_{\min} + k \} = \rho^k, \tag{90}$$

where $\rho = p/q$. Moreover, we can construct $\Gamma'(\ell)$ such that $\Gamma'(0) \triangleq \Gamma_\infty$ follows a steady state distribution, and for any $\ell \geq 0$, $\Gamma'(\ell+1) = \max \{ \Gamma'(\ell) + \xi(\ell)K, \Gamma_{\min} \}$. Given $\Gamma(0) = \Gamma_{\min}$, by induction, we have $\Gamma(\ell) \leq \Gamma'(\ell)$ for all $\ell \geq 0$. Thus,

$$\mathbb{P} \{ \Gamma(\ell) \geq \Gamma_{\min} + k \} \leq \mathbb{P} \{ \Gamma'(\ell) \geq \Gamma_{\min} + k \} = \mathbb{P} \{ \Gamma_\infty \geq \Gamma_{\min} + k \} = (1 - \rho)\rho^k \leq \rho^k.$$

Therefore,

$$\begin{aligned} \mathbb{E} \left[\max_{\ell \in [1, L]} \Gamma(\ell) \right] - \Gamma_{\min} &= \sum_{k=1}^{\infty} \mathbb{P} \left\{ \max_{\ell \in [1, L]} \Gamma(\ell) \geq \Gamma_{\min} + k \right\} \\ &\leq \sum_{k=1}^{\infty} \min \left\{ 1, \sum_{\ell=1}^L \mathbb{P} \{ \Gamma(\ell) \geq \Gamma_{\min} + k \} \right\} \\ &\leq \sum_{k=1}^{\infty} \min \{ 1, L\rho^k \}. \end{aligned}$$

Choose $k^* = \lfloor \frac{\log(L)}{\log(1/\rho)} \rfloor$. Then

$$\sum_{k=1}^{\infty} \min \{ 1, L\rho^k \} \leq k^* + L \sum_{k \geq k^*+1}^{\infty} \rho^k \leq k^* + \frac{L\rho^{k^*+1}}{1-\rho} \leq k^* + \frac{1}{1-\rho} \leq \frac{\log(L)}{\log(1/\rho)} + \frac{1}{1-\rho}$$

Therefore,

$$\mathbb{E} \left[\max_{\ell \in [1, L]} \|\delta^+(\ell)_\perp\| \right] \leq \mathbb{E} \left[\max_{\ell \in [1, L]} \Gamma(\ell) \right] \leq \Gamma_{\min} + \frac{\log(L)}{\log(1/\rho)} + \frac{1}{1-\rho}$$

Finally, note that $\log(1/\rho) \geq \log(1 + \eta/(2K)) \geq \eta/(2K)$ and $1 - \rho \geq \eta/(2K)$ Q.E.D.

Proof of Proposition 7. Finally, we derive a uniform upper bound to $\mathbb{E} [\max_{\ell \leq t_e} \|\delta^+(\ell)_\perp\|]$, where t_e denotes the random duration time of a (either positive or negative) cycle. Let $\bar{A}(\ell)$ denote the cumulative arrivals up to time step ℓ and $R(\ell) = \sqrt{2n} \langle \bar{A}(\ell), \mathbf{d} \rangle$. Then $R(\ell)$ is a biased random walk, where $R(\ell)$ increases by 1 with probability $p = \frac{1-\epsilon}{2}$ and decreases by 1 with probability $q = \frac{1+\epsilon}{2}$. Then by the definition of cycles,

$$t_e = \inf\{\ell \geq 1 : R_\ell = 0, R_1 = 1\}$$

Note that

$$\mathbb{P}\{t_e = 2n + 2\} = C_n p^n q^{n+1} \leq \frac{1}{n} 2^{2n} p^n q^{n+1} \triangleq a_n,$$

where $C_n = \frac{1}{n+1} \binom{2n}{n}$ is the Catalan number which counts the number of string of length $2n$ with an equal number of $+1$ and -1 bits such that no initial segment of the string has more -1 than $+1$. If so, the following proof needs minor adjustment.

Then, for any fixed time L , let $M(L) = \max_{\ell \leq L} \|\delta^+(\ell)_\perp\|$, and

$$\mathbb{E}[M(t_e)] = \mathbb{E}[M(t_e)\mathbb{I}(t_e \leq L)] + \mathbb{E}[M(t_e)\mathbb{I}(t_e > L)] \leq \mathbb{E}[M(L)] + \mathbb{E}[t_e \mathbb{I}(t_e > L)].$$

where the inequality holds because $\delta^+(\ell)$ can change by 1 at most at each time step so that $\|\delta^+(\ell)_\perp\| \leq \|\delta^+(\ell)\| \leq \ell \leq L$ for all $1 \leq \ell \leq L$.

By Lemma 23, $\mathbb{E}[M(L)] \leq C \log L$, where C is a constant independent of L and ϵ . In addition, we have

$$\mathbb{E}[t_e \mathbb{I}(t_e > L)] = \sum_{n=\lceil (L-1)/2 \rceil}^{\infty} (2n+2) a_n$$

Finally,

$$\sum_{n \geq \lceil (L-1)/2 \rceil} (2n+2) a_n \leq 4 \sum_{n \geq \lceil (L-1)/2 \rceil} 4^n p^n q^{n+1} \leq 4q \frac{(4pq)^{\lceil (L-1)/2 \rceil}}{1-4pq}.$$

Note that $4pq = 1 - \epsilon^2$. Thus, we get

$$\mathbb{E}[t_e \mathbb{I}(t_e > L)] \leq \frac{2(1+\epsilon)}{\epsilon^2} \exp\left(-\epsilon^2(L-1)/2\right).$$

Finally, we set $L = 1/\epsilon^4$ and get that

$$\mathbb{E}[M(t_e)] \leq \mathbb{E}[M(L)] + \mathbb{E}[t_e \mathbb{I}(t_e > L)] \leq 4C \log \frac{1}{\epsilon} + \frac{2(1+\epsilon)}{\epsilon^2} \exp\left(\frac{\epsilon^2}{2} - \frac{1}{2\epsilon^2}\right) = 4C \log \frac{1}{\epsilon} + C',$$

where both C and C' are constants independent of ϵ .

Q.E.D.

D.2. Proof of Theorem 3

Let $\bar{x}_1(\ell)$ denote the cumulative number of self-matches of type 1 up to period ℓ for any $\ell \geq 1$. We have

$$\langle r, \bar{y}(t) \rangle \leq \langle r, \bar{x}(t) \rangle \leq 0.1 \cdot \bar{x}_1(t) + 1 \cdot (\bar{A}_1(t) - \bar{x}_1(t)) + 0.1 (\bar{A}_2(t) - \bar{A}_1(t) + \bar{x}_1(t)) = 0.9 \bar{A}_1(t) + 0.1 \bar{A}_2(t) - 0.8 \bar{x}_1(t).$$

It follows that $\langle r, tx^* - \mathbb{E}[\bar{y}(t)] \rangle \geq 0.8 \mathbb{E}[\bar{y}_1(t)]$ and hence $\max_{0 \leq t \leq T} \langle r, tx^* - \mathbb{E}[\bar{y}(t)] \rangle \geq 0.8 \mathbb{E}[\bar{x}_1(T)]$. Then, it suffices to show that $\mathbb{E}[\bar{x}_1(t)] = \Omega\left(\frac{1}{\epsilon} \log(1/\epsilon)\right)$.

Recall $\bar{x}_s(\ell)$ denotes the cumulative number of self-matches up to period ℓ for any $\ell \geq 1$. During the first cycle from time periods 1 through $t_e^{(1)}$, by Lemma 15, for any $1 \leq \ell \leq t_e^{(1)}$, we get

$$\bar{x}_s(\ell) = \|\delta^+(\ell)\|_1 + |R(\ell)| \geq |R(\ell)|.$$

Since $\bar{x}_s(\ell)$ is non-decreasing in $1 \leq \ell \leq t_e^{(1)}$,

$$\bar{x}_s(t_e^{(1)}) \geq \max_{1 \leq \ell \leq t_e^{(1)}} |R(\ell)|.$$

Recall that K denotes the total number of complete cycles. Then, by symmetry, for any $1 < k \leq K$,

$$\bar{x}_s(t_e^{(k)}) - \bar{x}_s(t_e^{(k-1)}) \geq \max_{t_s^{(k)} \leq \ell \leq t_e^{(k)}} |R(\ell)|$$

By Claim 1, we know that only self-matches of type 1 can be conducted during the positive cycles. Then, we obtain

$$\begin{aligned} \bar{x}_1(t) &= \sum_{k=1}^{K(t)} \left(\bar{x}_s(t_e^{(k)}) - \bar{x}_s(t_e^{(k-1)}) \right) \mathbb{I}(R(t_s^{(k)}) = 1) \\ &\geq \sum_{k=1}^{K(t)} \max_{t_s^{(k)} \leq \ell \leq t_e^{(k)}} |R(\ell)| \cdot \mathbb{I}(R(t_s^{(k)}) = 1), \end{aligned}$$

where $K(t)$ denotes the total number of complete cycles up to time t . Taking the expectation over both hand sides yields that

$$\begin{aligned} \mathbb{E}[\bar{x}_1(t)] &\geq \mathbb{E} \left[\sum_{k=1}^{K(t)} \max_{t_s^{(k)} \leq \ell \leq t_e^{(k)}} |R(\ell)| \cdot \mathbb{I}(R(t_s^{(k)}) = 1) \right] \\ &\stackrel{(a)}{=} \mathbb{E}[K(t)] \cdot \mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \cdot \mathbb{I}(R(t_s^{(1)}) = 1) \right] \\ &\stackrel{(b)}{=} \frac{1}{2} \mathbb{E}[K(t)] \cdot \mathbb{E} \left[\max_{t_s^{(1)} \leq \ell \leq t_e^{(1)}} |R(\ell)| \right] \\ &\stackrel{(c)}{=} \Omega \left(\min \left\{ \frac{1}{\epsilon}, \sqrt{t} \right\} \log(1/\epsilon) \right) \end{aligned}$$

where (a) holds by Wald's equation because each completed cycle constitutes an i.i.d. renewal process that is independent from $K(t)$; (b) holds by symmetry; (c) holds by Lemma 14 and our claim that

$$\mathbb{E}[K(t)] \geq \Omega \left(\min \left\{ \frac{1}{\epsilon}, \sqrt{t} \right\} \right).$$

This completes our proof.

It remains to prove our claim. By definition, $K(t) = \sum_{\ell=1}^t \mathbb{I}(R(\ell) = 0)$. Thus

$$\mathbb{E}[K(t)] = \sum_{\ell=1}^t \mathbb{P}\{R(\ell) = 0\} = \sum_{\ell=1}^{t/2} \mathbb{P}\{R(2\ell) = 0\} = \sum_{\ell=1}^{t/2} \binom{2\ell}{\ell} \left(\frac{1-\epsilon}{2} \right)^\ell \left(\frac{1+\epsilon}{2} \right)^\ell \triangleq S_t, \quad (91)$$

where the second equality holds because ℓ needs to be even so that $R(\ell) = 0$; the third equality holds $R(2\ell) = 0$ if and only if there are exactly ℓ arrivals of each type.

As $t \rightarrow \infty$, $S_t \rightarrow S_\infty$. Using the identity $\sum_{\ell=1}^{\infty} \binom{2\ell}{\ell} x^\ell = (1-4x)^{-1/2} - 1$, we get a closed form expression for S_∞ :

$$S_\infty = \sum_{\ell=1}^{\infty} \binom{2\ell}{\ell} \left(\frac{1-\epsilon^2}{4} \right)^\ell = \left(1 - (1-\epsilon)^2 \right)^{-1/2} - 1 = \frac{1}{\epsilon} - 1.$$

With a bit of extra effort, we can derive a lower bound to S_t for finite t . Specifically, using the following lower bound to the central binomial coefficient

$$\binom{2\ell}{\ell} \geq \frac{1}{2\sqrt{\ell}} 4^\ell, \quad \text{for all } \ell \geq 1,$$

and letting $t_0 = \lfloor \min\{t, 1/\epsilon^2\}/2 \rfloor$, we get that

$$S_t \geq \sum_{\ell=1}^{t_0} \frac{1}{2\sqrt{\ell}} (1-\epsilon^2)^\ell \geq \frac{\sqrt{t_0}}{2} (1-\epsilon^2)^{t_0} = \frac{\sqrt{t_0}}{2} (1-\epsilon^2)^{\epsilon^2/2} \geq \frac{\sqrt{t_0}}{2} \left(1 - \frac{1}{2(1-\epsilon^2)} \right) = \frac{\sqrt{t_0}(1-2\epsilon^2)}{4(1-\epsilon^2)}.$$

where the last inequality holds because $(1-a)^x \geq 1 - \frac{ax}{1-a}$ for $0 < a < 1$. Plugging the above lower bound on S_t into (91), our claim follows.

Appendix E: Table of notation

Notation	Definition
$\mathcal{N} = [n]$	Set of agent types
$\mathcal{N}(m)$	Set of agent types participating in the match m
$\mathcal{M} = [d]$	Set of matches
λ	Arrival rates vector
$A_i(t)$	Number of arrivals of type i at time t
M	Network matrix, where for any $i \in [n]$, let M_i denote its i th row such that for $m \in [d]$, $M_{im} = \mathbb{I}(i \in \mathcal{N}(m))$
r_m	Reward for match m
$r_{\max}$	Maximum reward across all matches
B	$\max_{m \in \mathcal{M}} \mathcal{N}(m) + 1$
$x(t)$	Vector of scheduled matches at time t
$y(t)$	Vector of realized matches at time t
$y^h(\omega)$	Vector of realized matches by the hindsight optimal policy under a random scenario ω
$\delta(t)$	Vector of shortage of agents at period t , where $-\delta_i(t)$ represents the difference between the number of agents of type i that have arrived by time t and the total number of agents of type i required for all matches scheduled before time t .
$\hat{\lambda}(t)$	Empirical arrival rates vector at time t
U^*	Optimal solution of the Lagrangian dual problem
$\hat{U}(t)$	Empirical estimator of the optimal dual variable U^* at time t under $\hat{\lambda}(t)$
$U(t)$	Dual variable (i.e., Lagrangian multipliers) at t
V_t	Some policy parameter (possibly time-varying)
$L_\lambda(U)$	Lagrangian relaxation
ϵ	General Position Gap (GPG)
$R^{h,t}$	Reward under hindsight optimal policy up to time t
$R^{\pi,t}$	Reward under policy π up to time t

Table 1 Table of Notations

References

- Bertsimas, Dimitris, John N Tsitsiklis. 1997. *Introduction to linear optimization*, vol. 6. Athena Scientific Belmont, MA.
- Cox, David Roxbee, Hilton David Miller. 2017. *The theory of stochastic processes*. Routledge.
- Gupta, Varun. 2024. Greedy algorithm for multiway matching with bounded regret. *Operations Research* **72**(3) 1139–1155.
- Gurvich, Itay, Ward Whitt. 2009. Scheduling flexible servers with convex delay costs in many-server service systems. *Manufacturing & Service Operations Management* **11**(2) 237–253.
- Hajek, Bruce. 2015. *Random processes for engineers*. Cambridge university press.
- Harrison, J Michael. 1998. Heavy traffic analysis of a system with parallel servers: asymptotic optimality of discrete-review policies. *The Annals of Applied Probability* **8**(3) 822–848.
- Kerimov, Süleyman, Itai Ashlagi, Itai Gurvich. 2024. Dynamic matching: Characterizing and achieving constant regret. *Management Science* **70**(5) 2799–2822.
- Shi, Cong, Yehua Wei, Yuan Zhong. 2019. Process flexibility for multiperiod production systems. *Operations Research* **67**(5) 1300–1320.