Appendix A: Missing Proofs from Section 2

LEMMA 1. *For every instance G of OSOW-SO (and OSW-SO) we have* $\mathrm{OPT}^c(G) \geq \mathrm{OPT}(G)$.

Proof. To show that $\mathrm{OPT}^c(G) \geq \mathrm{OPT}(G)$, it suffices to show that there exists a vector $Y = (y_a)_{a\in\mathcal{A}} \in \Delta(\mathcal{A})$, and a distribution α over subsets of $\mathcal{X}$, such that,

$$\sum_{X\in\mathcal{X}} \alpha(X) f(X) = \mathrm{OPT}(G), \quad \sum_{X\in\mathcal{X}} \alpha(X) = 1, \quad \text{and} \sum_{X\in\mathcal{X} \mid X\ni e} \alpha(X) = p_e\, y_a \ \forall e \in N(a),\, a\in\mathcal{A}.$$

Let S denote the (random) set of actions chosen by OPT and let $Z \in \mathcal{X}$ denote the set of realized outcomes of S. Note that $Z \subseteq N(S)$. We define y_a as the probability that action $a \in S$. Then, for each $e \in N(a)$, the probability that $e \in Z$ is given by $y_a\, p_e$. This follows from the fact that OPT is non-anticipatory, meaning it chooses action a with probability (w.p.) y_a without knowing the outcome, and the outcome e occurs w.p. p_e independent of other events.

Now, define $\alpha(X)$ as the probability that $Z = X$. With this definition, we have,

$$\sum_{X\in\mathcal{X}} \alpha(X) = 1 \quad \text{and} \quad \sum_{X\in\mathcal{X}\mid X\ni e} \alpha(X) = p_e y_a \ \forall e\in N(a),\, a\in\mathcal{A},$$

which satisfies the desired conditions. Finally, since OPT has total reward $f(X)$ on sample paths where $Z = X$, we conclude that, $\mathrm{OPT}(G) = \sum_{X\in\mathcal{X}} \alpha(X)\, f(X)$.

□

LEMMA 8. *If f is monotone and has an arrival-consistent submodular order then, for every $P \in \mathcal{N}$ and non-negative real value λ, the function $\lambda\, f(N(\cdot)\cap P)$ on ground set $\mathcal{A}$ is monotone with an arrival-consistent submodular order.*

Proof. By Remark 1 in Udwani (2025b), multiplying a monotone function with a submodular-order by a non-negative scalar preserves both monotonicity and the submodular-order property. Therefore, it suffices to focus on the function $f(N(\cdot)\cap P)$. Let π denote the arrival-consistent submodular order of f. Fix an arbitrary $P \in \mathcal{N}$ and let $\psi(S) = N(S)\cap P$. Since ψ is an injective mapping from $\mathcal{A}$ to N, there is a unique arrival-consistent order $\pi_{\mathcal{A}}$ over $\mathcal{A}$ such that for any two actions a, a', we have $a \succ_{\pi_{\mathcal{A}}} a'$ if and only if $N(a)\cap P \succ_{\pi} N(a')\cap P$. More strongly, for any two action sets B and C, we have $C \succ_{\pi_{\mathcal{A}}} B$ if and only if $\psi(C) \succ_{\pi} \psi(B)$.

Now, $f(\psi(\cdot))$ is monotone because $\psi(B) \subseteq \psi(B^+)$ for all $B \subseteq B^+ \subseteq \mathcal{A}$. And $\pi_{\mathcal{A}}$ is an arrival-consistent submodular order for $f(\psi(\cdot))$ because for all $\pi_{\mathcal{A}}$-nested sets $B \subseteq B^+$ and all sets $C \succ_{\pi_{\mathcal{A}}} B^+$, we have that, $\psi(B)$ and $\psi(B^+)$ are π-nested and $\psi(C) \succ_{\pi} \psi(B^+)$. Thus,

$$f(\psi(B^+\cup C)) - f(\psi(B^+)) = f(\psi(C)\mid\psi(B^+)) \,\leq\, f(\psi(C)\mid\psi(B)) = f(\psi(B\cup C)) - f(\psi(B)),$$

here the inequality follows from the submodular order property of f.

□

Appendix B: Special Cases of OSOW-SO

We begin by proving that the OSW-SO formulation captures all of the settings we discussed in Section 3, with the exception of OBM with reusable resources, which, as shown in Appendix B.1, is a special case of OSOW.

To establish this, we define an equivalent instance of OSW-SO for each setting. We start by describing the common features shared by these instances. Recall that all the settings we are considering involve a set of resources I. We focus on instances of OSW-SO where the objective function f is a sum of monotone submodular functions $\{f_i\}_{i\in I}$. Each function f_i is defined using a subset $N_i \subseteq N$ of outcomes, with only elements in N_i contributing non-zero values to f_i. The sets N_i may overlap for different resources. We now define the specific instance for each setting.

Online matching with stochastic rewards: For each arrival $t \in [T]$, the action set A_t is the set of all edges incident to t. Action (i,t) has two possible outcomes: success or failure. The outcome of success is represented by $e_{i,t,s}$, which is included in the set N_i. Given a subset of outcomes $X \subseteq N$, the reward from resource i is captured by the monotone submodular function $f_i(X) = r_i \min\{c_i, |X \cap N_i|\}$. The overall objective function, $f = \sum_{i\in I} f_i$, is also a monotone submodular function.

Stochastic rewards with patience: In this setting, a feasible action at arrival t is an ordered sequence of resources to attempt matching with arrival t. Let A_t denote the set of all feasible actions. Each action $a \in A_t$ has a vector outcome $e_{a,t} = (e_{a,i,t})_{i\in I}$, where $e_{a,i,t} = 1$ if resource i is successfully matched to t and $e_{a,i,t} = 0$ otherwise. If $e_{a,i,t} = 1$, then we include outcome $e_{a,t}$ in the set N_i. The reward functions f_i and objective function f are the same as in the previous case.

Online assortment optimization: The action set A_t consists of the assortments that can be shown to arrival $t \in [T]$. Similar to the previous case, each action $a \in A_t$ has a vector outcome $e_{a,t} = (e_{a,i,t})_{i\in I}$, where $e_{a,i,t} = 1$ if arrival t chooses resource i and $e_{a,i,t} = 0$ otherwise. If $e_{a,i,t} = 1$, then we include outcome $e_{a,t}$ in the set N_i. The reward functions f_i and objective function f are the same as in the previous two cases.

Two-sided assortment optimization: The set of actions and outcomes is the same as online assortment optimization. However, the reward function $f_i(X)$ now represents the probability that resource i chooses at least one element from the set X. As noted by Aouad and Saban (2023), this is a monotone submodular function for a large variety of discrete choice models.

REMARK 11 (SCALAR OUTCOMES DO NOT SUFFICE). Recall that one way to represent each outcome in assortment optimization is using a binary vector $e = (e_i)_{i\in I}$ where $e_i = 1$ if i is selected by the customer and $e_i = 0$ otherwise. In Asadpour and Nazerzadeh (2016), every outcome is a scalar and we sketch an example to show that a monotone function f on scalar outcomes cannot capture the assortment problem. Consider two items i and j, with unit capacity and per unit rewards $r_i = 1$ and $r_j = 2$. Consider two arrivals with feasible assortments (actions) $U_1 = \{j\}$ at arrival 1 and $U_2 = \{i, j\}$ at arrival 2. The assortment (action)

U_2 has three possible outcomes and we can capture them using a random variable $u_2 \in [0,1]$ such that, $u_2 = \delta_{2,i}$ when item i is chosen, $u_2 = \delta_{2,j}$ when item j is chosen, and $u_2 = 0$ when the outside option is selected. Similarly, let $u_1 = \delta_{1,j}$ when item j is selected and $u_1 = 0$ otherwise. Consider a monotone and DR submodular objective function f on $[0,1]^2$. We have that $f(\{0,\delta_{2,i}\}) = r_i$ and $f(\{0,\delta_{2,j}\}) = r_j$. Since $r_j > r_i$, from the monotonicity of f we have that $\delta_{2,j} > \delta_{2,i}$. Further, $f(\{\delta_{1,j},\delta_{2,j}\}) = r_j$ (item j has unit capacity) and $f(\{\delta_{1,j},\delta_{2,i}\}) = r_j + r_i$. But this violates monotonicity of f because $f(\{\delta_{1,j},\delta_{2,i}\}) > f(\{\delta_{1,j},\delta_{2,j}\})$ even though $\delta_{2,j} > \delta_{2,i}$.

B.1. Online Matching with Reusable Resources

As mentioned in Section 3, a reusable resource is used/rented by an arrival for some time and then returned back to the system. A reward is generated each time the resource is rented to a new arrival. Specifically, suppose that arrival $t \in [T]$ arrives at *time* $a(t) \in [0,1]$. At each moment in $[0,1]$, every resource is either *available* or *unavailable*, with all resources initially available at time 0. If resource i is available at time $a(t)$, matching it to arrival t generates a reward r_i and i becomes unavailable for a fixed duration d_i. Thus, i is unavailable during the interval $(a(t), a(t)+d_i)$. Matching an arrival to an unavailable resource has no effect and generates no reward. The objective is to maximize the total reward. Notably, the classic online bipartite matching is a special case of this setting where $d_i \to +\infty\ \forall i \in I$.

Perhaps surprisingly, we find that OSW cannot capture this problem, even for instances with a *single* reusable resource.

EXAMPLE 4 (OSW FAILS TO CAPTURE REUSABILITY). Consider an instance with a single reusable resource that is used for a duration of 1.5 time units after each match and during this time the resource is unavailable. Arrivals occur at times $1, 2$, and 3. If arrival 1 is matched, the resource is used from time 1 to 2.5 and returns prior to arrival 3's arrival. Arrivals 1 and 3 can both be matched to the resource and arrival 2 can be matched only if the other two arrivals are unmatched. Now, consider a function f such that $f(\{1,2\}) = f(\{2,3\}) = f(1) = f(2) = f(3) = 1$ and $f(\{1,3\}) = f(\{1,2,3\}) = 2$. This function captures the total number of matches in every possible allocation of arrivals to the resource. Clearly, f is not a submodular function ($f(1 \mid \{2,3\}) > f(1 \mid \{2\})$).

We show that OSOW captures OBM with reusable resources (OBMR). Recall that OBMR, a unit of resource i that is matched to arrival t at time $a(t)$ is rented/used for a fixed duration d_i and returned at time $a(t)+d_i$. Given a set of arrivals S that have an edge to i, consider the process where we start with the first arrival in S and match i to every arrival in S where i is available. We refer to this as the *matching process* on S [13]. Let $n_i(S)$ denote the total number of arrivals matched to i in this process. Observe that $\sum_{i\in I} n_i(S_i)$ is the total reward of an allocation $\{S_i\}_{i\in I}$ of arrivals to resources in OBMR. It is not hard to see that n_i is

[13] This is the deterministic counterpart of the (F,σ)-random process introduced in Goyal et al. (2025).

a monotone function. In fact, this is a direct corollary of Lemma 5 in Goyal et al. (2025). We show that the arrival order is a submodular order for n_i.

LEMMA 9. *In OBMR, for every resource $i \in I$, the arrival order is a submodular order for n_i.*

Proof. Let π denote the arrival order over the ground set of arrivals. Consider π-nested sets $B \subseteq A \subseteq [T]$ and a set C that succeeds A in the arrival order. For $S \in \{A, B\}$, let t_S be the last arrival in S that is matched to i in the matching process on $S \cup C$. Let $t_{C|S}$ denote the first arrival in C that is matched in the process on $S \cup C$ and let $t_{C|S} = T + 1$ if no such arrival exists. Observe that $n_i(C \mid S) = n_i(S \cup C) - n_i(S)$ is the number of arrivals in C that are matched to i in the matching process on $S \cup C$. Due to the time nesting of A and B, we have, $t_A \geq t_B$. Thus, $t_{C|A} \geq t_{C|B}$. Now, $n_i(C \mid B) \geq n_i(C \mid A)$ follows from the monotonicity of n_i. $\square$

Beyond Deterministic Reusability. Prior work also considers a more general version of reusability where the usage durations are stochastic. Specifically, when a resource i is rented to any arrival t, it is used for a random duration d_t sampled independently from a distribution D_i. Gong et al. (2022) showed that an adaptive Greedy algorithm is 0.5-competitive in this setting.

It is possible to extend Lemma 9 to show that the arrival order is a submodular order for the expected reward function $E_{D_i}[n_i(\cdot)]$, where the expectation is taken over the random usage durations. Then, using our result for OSOW-SO, we can show that Greedy is 0.5-competitive for settings with stochastic usage durations against an offline benchmark that is *oblivious* to the realizations of usage durations. Extending our techniques to compare against the duration adaptive offline benchmark appears challenging because the arrival order is not a submodular order for every sample path of usage durations.

Nonetheless, OSOW can capture generalizations of OBMR. For example, given a monotone submodular function h_i, the function $f_i = n_i + h_i$ is monotone and the arrival order is a submodular order. The literature on online allocation of reusable resources also includes several variations of the model that we consider here (Dickerson et al. 2018, Rusmevichientong et al. 2020, Owen and Simchi-Levi 2018, Levi and Radovanović 2010, Feng et al. 2024, Manshadi and Rodilitz 2022, Baek and Ma 2022, Baek and Wang 2023, Papadigenopoulos and Caramanis 2021, Simchi-Levi et al. 2022, Huo and Cheung 2022, Ekbatani et al. 2024, Sumida 2024). It would be interesting to see if OSOW captures these settings.

B.2. Stronger Benchmark for Stochastic Rewards with Patience

In the setting of stochastic rewards with patience, the offline benchmark described in Section 2.3 may visit arrivals in any (adaptive) order, but it is required to complete all match attempts for a given arrival before moving to the next one. Borodin et al. (2022) consider a stronger benchmark for this problem, which can probe edges in any adaptive order. For example, their benchmark may attempt to match one arrival, move to attempt matching another arrival, and then return to the first arrival to try a different resource. We show that the benchmark OPT^c also serves as an upper bound for this stronger benchmark. Consequently, all our competitive ratio results for the stochastic rewards with patience setting remain valid against the strongest benchmark in the existing literature. Let $\mathrm{OPT}^+(G)$ denote the value of the benchmark of Borodin et al. (2022) on instance G.

LEMMA 10. *For every instance G of stochastic rewards with patience we have* $\mathrm{OPT}^c(G) \geq \mathrm{OPT}^+(G)$.

Proof. Consider an arbitrary arrival t, and let ω_{-t} denote the random realizations of all edges not incident to t. Conditioned on ω_{-t}, let $a(\omega_{-t})$ denote the action selected by OPT^+ at arrival t in the event that all edges incident to t fail. This action is *maximal* in the sense that it fully specifies the sequence of attempts made to match arrival t. More precisely, since OPT^+ is non-anticipatory, conditioned on ω_{-t} it attempts to match arrival t—possibly interleaving these attempts with attempts for other arrivals—according to the sequence of edges encoded by $a(\omega_{-t})$, until either a successful match occurs or arrival t exhausts its patience.

Let $\sigma_{\omega_{-t}}$ denote the induced ordering of resources attempted for arrival t under action $a(\omega_{-t})$. Conditioned on ω_{-t} and on arrival t having sufficient patience, the probability that the realized outcome of $a(\omega_{-t})$ is a successful match to resource i (i.e., $e_{a(\omega_{-t}),i,t} = 1$) is

$$p_{i,t} \prod_{j \prec_{\sigma_{\omega_{-t}}} i} (1 - p_{j,t}). \tag{7}$$

An analogous characterization applies to any online algorithm. Consequently, without loss of generality, we interpret each action in $\mathcal{A}_t$ as a *maximal* action at arrival t. Under this interpretation, if OPT^+ selects an action $a \in \mathcal{A}_t$ with probability y_a, then the probability that a particular outcome $e \in N(a)$ is realized is $y_a p_e$. Here, p_e denotes the conditional probability that action a results in outcome e, given by the product of the probability that arrival t has sufficient patience and the corresponding expression in (7). Importantly, the coefficient y_a captures the randomness induced by ω_{-t}, i.e., the realizations of all edges not incident to t.

The remainder of the proof follows the same structure as the proof of Lemma 1, which we reproduce here for completeness. It suffices to show that there exist a vector $Y = (y_a)_{a \in \mathcal{A}} \in \Delta(\mathcal{A})$ and a distribution α over subsets of $\mathcal{X}$ such that

$$\sum_{X \in \mathcal{X}} \alpha(X) f(X) = \mathrm{OPT}^+(G), \qquad \sum_{X \in \mathcal{X}} \alpha(X) = 1, \qquad \sum_{X \in \mathcal{X} : e \in X} \alpha(X) = p_e y_a \quad \forall\, e \in N(a),\, a \in \mathcal{A}.$$

Let S denote the (random) set of actions chosen by OPT^+, and let $Z \in \mathcal{X}$ denote the corresponding set of realized outcomes; note that $Z \subseteq N(S)$. Define y_a as the probability that action a is selected, i.e., $y_a = \Pr[a \in S]$. As observed above, for each outcome $e \in N(a)$, the probability that $e \in Z$ is $y_a p_e$.

Finally, define $\alpha(X)$ as the probability that $Z=X$. By construction, we have

$$\sum_{X\in\mathcal{X}}\alpha(X)=1 \quad \text{and} \quad \sum_{X\in\mathcal{X}:e\in X}\alpha(X)=p_e y_a \quad \forall\, e\in N(a),\, a\in\mathcal{A},$$

which satisfies the desired conditions. Since OPT^+ receives total reward $f(X)$ on sample paths where $Z=X$, we conclude that

$$\mathrm{OPT}^+(G)=\sum_{X\in\mathcal{X}}\alpha(X)\,f(X).$$

□

Appendix C: Missing Proofs from Section 4.1

LEMMA 2. *Consider a collection of monotone functions $\{F_1,\cdots,F_u\}$ defined on the ground set $\mathcal{A}$. Suppose that each function admits an arrival-consistent submodular order (not necessarily the same order across functions). Then, the function $F:=\sum_{j\in[u]}F_j$ satisfies the following inequality:*

$$F(\mathrm{OPT}_{T+1}\cup\mathrm{R}_{T+1}) \le F(\mathrm{R}_{T+1}) + \sum_{t\in[T]} F\big(\{o_t\}\setminus\{r_t\}\,\big|\,\cup_{\tau\in[t-1]}\{r_\tau\}\big)\,.$$

Proof. The main ingredient in this proof is Lemma 11 (shown later), which essentially follows from Corollary 2 in Udwani (2025b). For completeness, the lemma is stated and proved after the conclusion of this proof. Observe that for all $A,B\subseteq\mathcal{A}$,

$$F(A\mid B)=\sum_{j\in[u]}F_j(A\mid B).$$

Applying Lemma 11 to each function F_j, we obtain,

$$\begin{aligned}
F(\mathrm{OPT}_{T+1}\cup\mathrm{R}_{T+1}) &= \sum_{j\in[u]}F_j(\mathrm{OPT}_{T+1}\cup\mathrm{R}_{T+1}),\\
&\le \sum_{j\in[u]}\left[F_j(\mathrm{R}_{T+1}) + \sum_{t\in[T]}F_j\big(\{o_t\}\setminus\{r_t\}\,\big|\,\cup_{\tau\in[t-1]}\{r_\tau\}\big)\right],\\
&= F(\mathrm{R}_{T+1}) + \sum_{t\in[T]}F\big(\{o_t\}\setminus\{r_t\}\,\big|\,\cup_{\tau\in[t-1]}\{r_\tau\}\big)\,.
\end{aligned}$$

□

LEMMA 11. *Given a monotone submodular function F with an arrival-consistent submodular $\pi_{\mathcal{A}}$ on ground set $\mathcal{A}$, and subsets $\mathrm{OPT}_{T+1}=\{o_1,\ldots,o_T\}$ and $\mathrm{R}_{T+1}=\{r_1,\ldots,r_T\}$ of $\mathcal{A}$ such that $\{o_t,r_t\}\subseteq A_t\ \forall t\in[T]$, we have,*

$$F(\mathrm{OPT}_{T+1}\cup\mathrm{R}_{T+1}) \le F(\mathrm{R}_{T+1}) + \sum_{t\in[T]} F\big(\{o_t\}\setminus\{r_t\}\,\big|\,\cup_{\tau\in[t-1]}\{r_\tau\}\big)\,.$$

Proof. Let $\mathrm{OPT}_{t+1} = \cup_{\tau\in[t]}\{o_\tau\}$ and $\mathrm{R}_{t+1} = \cup_{\tau\in[t]}\{r_\tau\}$. To prove the lemma we first establish the following key inequality for all $t \in [T]$.

$$\begin{aligned}
&F(\{o_t, r_t, \cdots, o_T, r_T\} \mid \mathrm{R}_t) \\
&\leq F(\{r_t\} \mid \mathrm{R}_t) + F(\{o_t\}\backslash\{r_t\} \mid \mathrm{R}_t) + F(\{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\} \mid \mathrm{R}_{t+1}).
\end{aligned} \tag{8}$$

Then, using induction we establish that,

$$\begin{aligned}
&F(\mathrm{OPT}_{T+1} \cup \mathrm{R}_{T+1}) \\
&\leq F(\mathrm{R}_{t+1}) + F(\{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\} \mid \mathrm{R}_{t+1}) + \sum_{\tau\in[t]} F(\{o_\tau\} \setminus \{r_\tau\} \mid \mathrm{R}_\tau) \quad \forall t \in [T].
\end{aligned} \tag{9}$$

The main claim follows from (9) when $t = T$.

Before proceeding with the proof we observe that the action o_t could precede, succeed, or equal r_t. To avoid a case-wise approach we define sets $\{o_t^-, r_t, o_t^+\}$ where

$$o_t^- = \begin{cases} o_t & \text{if } o_t \prec_{\pi_\mathcal{A}} r_t, \\ \emptyset & \text{otherwise}, \end{cases} \quad \text{and} \quad o_t^+ = \begin{cases} o_t & \text{if } r_t \prec_{\pi_\mathcal{A}} o_t, \\ \emptyset & \text{otherwise}. \end{cases}$$

Observe that when $o_t^- = o_t$, we have $o_t^- \prec_{\pi_\mathcal{A}} r_t \prec_{\pi_\mathcal{A}} \{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\}$. Similarly, when $o_t^+ = o_t$, we have $r_t \prec_{\pi_\mathcal{A}} o_t^+ \prec_{\pi_\mathcal{A}} \{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\}$. Note that $o_t^- = o_t^+ = \emptyset$ when $o_t = r_t$. Recall that, $F(\{\emptyset\} \mid S) = 0$ for all $S \subseteq \mathcal{A}$.

We now show inequality (8) for all $t \in [T]$.

$$\begin{aligned}
F(\{o_t, r_t, \cdots, o_T, r_T\} \mid \mathrm{R}_t) &= F\big(\{o_t^-\} \mid \mathrm{R}_t\big) + F\left(\{r_t, o_t^+\} \cup \{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\} \,\middle|\, \mathrm{R}_t \cup \{o_t^-\}\right), \\
&\leq F\big(\{o_t^-\} \mid \mathrm{R}_t\big) + F\left(\{r_t, o_t^+\} \cup \{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\} \,\middle|\, \mathrm{R}_t\right), \\
&= F\big(\{o_t^-\} \mid \mathrm{R}_t\big) + F\left(\{r_t\} \mid \mathrm{R}_t\right) + F\left(\{o_t^+\} \,\middle|\, \mathrm{R}_t \cup \{r_t\}\right) \\
&\quad + F\left(\{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\} \,\middle|\, \mathrm{R}_t \cup \{r_t, o_t^+\}\right), \\
&\leq F\big(\{o_t^-\} \mid \mathrm{R}_t\big) + F\left(\{r_t\} \mid \mathrm{R}_t\right) + F\left(\{o_t^+\} \,\middle|\, \mathrm{R}_t\right) \\
&\quad + F\left(\{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\} \mid \mathrm{R}_{t+1}\right), \\
&= F\left(\{r_t\} \mid \mathrm{R}_t\right) + F\left(\{o_{t+1}, r_{t+1}, \cdots, o_T, r_T\} \mid \mathrm{R}_{t+1}\right) + F\left(\{o_t\}\backslash\{r_t\} \mid \mathrm{R}_t\right),
\end{aligned}$$

here the two inequalities follow from the submodular order property of F applied to the $\pi_\mathcal{A}$-nested pairs $(\mathrm{R}_t, \mathrm{R}_t \cup \{o_t^-\})$, $(\mathrm{R}_t, \mathrm{R}_t \cup \{r_t\})$, and $(\mathrm{R}_{t+1}, \mathrm{R}_t \cup \{r_t, o_t^+\})$. The final identity follows from the observation that $F(\{o_t\}\backslash\{r_t\} \mid \mathrm{R}_t) = F(\{o_t^-\} \mid \mathrm{R}_t) + F(\{o_t^+\} \mid \mathrm{R}_t)$.

Now, using (8) for $t = 1$ we have,

$$F(\mathrm{OPT}_{T+1} \cup \mathrm{R}_{T+1}) \leq F(\{r_1\}) + F(\{o_1\}\backslash\{r_1\}) + F(\{o_2, r_2, \cdots, o_T, r_T\} \mid \{r_1\}),$$

which is exactly inequality (9) for $t=1$. Suppose that (9) is true for all $t \le t_0 - 1$. Then, by induction,

$$
\begin{aligned}
F(\mathrm{OPT}_{T+1} \cup \mathrm{R}_{T+1}) &\le F(\mathrm{R}_{t_0}) + F(\{o_{t_0}, r_{t_0}, \cdots, o_T, r_T\} \mid \mathrm{R}_{t_0}) + \sum_{\tau \in [t_0-1]} F(\{o_\tau\} \setminus \{r_\tau\} \mid \mathrm{R}_\tau), \\
&\le F(\mathrm{R}_{t_0+1}) + F(\{o_{t_0+1}, r_{t_0+1}, \cdots, o_T, r_T\} \mid \mathrm{R}_{t_0+1}) + \sum_{\tau \in [t_0]} F(\{o_\tau\} \setminus \{r_\tau\} \mid \mathrm{R}_\tau),
\end{aligned}
$$

here the first inequality corresponds to inequality (9) for $t = t_0 - 1$ and the second inequality follows from inequality (8) for $t = t_0$.

□

Appendix D: Proof of Lemma 3

Lemma 3 states that: (i) $\widehat{G}$ is an instance of OSOW, (ii) The dominance property holds, and (iii) The invariance property holds. The dominance property follows directly from the definition (see (5)), so we focus on proving (i) and (iii).

D.1. Proof of Claim (i)

LEMMA 4. *If G is an instance of OSOW-SO then $\widehat{G}$ is an instance of OSOW. Similarly, if G is an instance of OSW-SO then $\widehat{G}$ is an instance of OSW.*

Proof. The lemma above strengthens claim (i). Recall that $F(S_1 \cup S_2)$, as defined in (3), is a non-negative linear combination of functions $f((N(S_1) \cap P_1) \cup (N(S_2) \cap P_2))$. To establish that $\widehat{G}$ is an instance of OSOW, it suffices to show that $f((N(S_1) \cap P_1) \cup (N(S_2) \cap P_2))$ is monotone with an arrival-consistent submodular order on $\mathcal{A}$. By Remark 1 in Udwani (2025b), multiplying a monotone function with a submodular-order by a non-negative scalar preserves both monotonicity and the submodular-order property.

When G is an instance of OSW-SO, i.e., f is submodular, it suffices to show that $f((N(S_1) \cap P_1) \cup (N(S_2) \cap P_2))$ is also submodular. This implies that $\widehat{G}$ is an instance of OSW using the fact that non-negative linear combination of monotone submodular functions is also monotone submodular.

Fix arbitrary sets $P_1 \in \mathcal{N}$ and $P_2 \in \mathcal{X}$ and define $\psi : 2^{\mathscr{A}} \to 2^N$ as follows,

$$
\psi(S) = (N(S \cap \mathcal{A}) \cap P_1) \cup (N(S \cap \widehat{\mathcal{A}}) \cap P_2). \tag{10}
$$

By definition of $\mathcal{N}$ and $\mathcal{X}$, $\psi(\{a\}) = N(a) \cap P_1$ (for $a \in \mathcal{A}$) and $\psi(\{\widehat{a}_t\}) = N_t \cap P_2$ are both singleton sets that represent the realized outcomes of a and $\hat{a}$ respectively. We will show that $f(\psi(\cdot)) : 2^{\mathscr{A}} \to \mathbb{R}_+$ is a monotone submodular (order) function when f is a monotone submodular (order) function. We first show that $f(\psi(\cdot))$ is monotone. Then, we show that $f(\psi(\cdot))$ is submodular when f is submodular. Finally, we show that if f has an arrival-consistent submodular order on N then $f(\psi(\cdot))$ has an arrival-consistent submodular order on $\mathcal{A}$.

Monotonicity: Suppose that f is monotone. Observe that,

$$\psi(S) = \cup_{a\in S}\psi(a) \quad \forall S \subseteq \mathscr{A}. \tag{11}$$

Thus, $f(\psi(\cdot))$ is monotone because f is monotone and $\psi(A) \supseteq \psi(B)$ for all $B \subseteq A$.

Submodularity: Suppose that f is a monotone submodular function. Consider subsets $B \subseteq A \subseteq \mathscr{A}$, and set $C \subseteq \mathscr{A}\backslash A$. We have,

$$\begin{aligned}
f(\psi(A\cup C)) - f(\psi(A)) &= f(\psi(A\cup C)\backslash\psi(A) \mid \psi(A)),\\
&\leq f(\psi(A\cup C)\backslash\psi(A) \mid \psi(B)),\\
&= f(\psi(C)\backslash\psi(A) \mid \psi(B)),\\
&\leq f(\psi(C)\backslash\psi(B) \mid \psi(B)),\\
&= f(\psi(B\cup C)\backslash\psi(B) \mid \psi(B)),\\
&= f(\psi(B\cup C)) - f(\psi(B)).
\end{aligned}$$

The first inequality follows from submodularity of f. The second inequality follows from the monotonicity of f. The second and third equalities follow from (11). Overall, this proves that $f(\psi(\cdot))$ is submodular.

Note that we use the submodularity of f only in the first inequality. In the next part, we replace submodularity with the weaker submodular order property.

Submodular Order: Suppose that f is a monotone function with an arrival-consistent submodular order π over N. To show that $f(\psi(\cdot))$ is also a submodular order function we first define a candidate order $\pi_{\mathscr{A}}$ over $\mathscr{A}$.

Recall that, $\psi(\{a\})$ is a singleton that represents the realized outcome of action $a \in \mathscr{A}$. At a high level, we define $\pi_{\mathscr{A}}$ so that it is consistent with the order induced by π over the set $(N\cap P_1)\cup(N\cap P_2)$ of realized outcomes. Specifically, distinct actions $a_1, a_2 \in \mathscr{A}$ are ordered as follows:

(i) If $a_1, a_2 \in \mathcal{A}$, then $a_1 \succ_{\pi_{\mathscr{A}}} a_2$ if and only if $N(a_1)\cap P_1 \succ_\pi N(a_2)\cap P_1$.

(ii) If $\hat{a}_t, \hat{a}_\tau \in \hat{\mathcal{A}}$, then $\hat{a}_t \succ_{\pi_{\mathscr{A}}} \hat{a}_\tau$ if and only if $t > \tau$ (which coincides with $N_t\cap P_2 \succ_\pi N_\tau \cap P_2$).

(iii) If $a \in \mathcal{A}$ and $\hat{a}_t \in \hat{\mathcal{A}}$, then $\hat{a}_t \succ_{\pi_{\mathscr{A}}} a$ if either $N_t\cap P_2 \succ_\pi N(a)\cap P_1$ or $N_t \cap P_2 = N(a)\cap P_1$.

By definition, order $\pi_{\mathscr{A}}$ is arrival-consistent. Note that the realized outcomes of a and $\hat{a}_t$ coincide when $N_t\cap P_2 = N(a)\cap P_1$, and in this case we can set any order for a and $\hat{a}_t$.

In the earlier proof of submodularity of $f(\psi(\cdot))$ (when f is submodular), we use the submodularity of f only once to obtain the following inequality,

$$f(\psi(C)\backslash\psi(A) \mid \psi(A)) \leq f(\psi(C)\backslash\psi(A) \mid \psi(B)), \tag{12}$$

for all $B \subseteq A$ and $C \subseteq \mathscr{A} \backslash A$. To show that $\pi_{\mathscr{A}}$ is a submodular order for $f(\psi(\cdot))$, we only need to establish inequality (12) for $\pi_{\mathscr{A}}$-nested sets $B \subseteq A$ and $C \succ_{\pi_{\mathscr{A}}} A$. This will be the focus of the rest of the proof. Note that Lemma 12 is the key ingredient in the following part of the proof and is shown separately later.

From Lemma 12, we have that either $\psi(C) \backslash \psi(A) = \emptyset$ or $\psi(C) \backslash \psi(A) \succ_\pi \psi(A)$. If $\psi(C) \backslash \psi(A) = \emptyset$, the left hand side in (12) equals 0 and the inequality holds trivially by the monotonicity of f. Now, assume $\psi(C) \backslash \psi(A) \neq \emptyset$, so we have

$$\psi(C) \backslash \psi(A) \succ_\pi \psi(A).$$

Then, inequality (12) follows directly from the submodular order property of f provided that,

$$\psi(B) \text{ and } \psi(A) \text{ are } \pi\text{-nested.}$$

Using Lemma 12 again, we have that either $\psi(A \backslash B) \backslash \psi(B) = \emptyset$ or $\psi(A \backslash B) \backslash \psi(B) \succ_\pi \psi(B)$. Rewriting this using (11), we have that either $\psi(A) = \psi(B)$ or $\psi(A) \backslash \psi(B) \succ_\pi \psi(B)$. If $\psi(A) = \psi(B)$ then the two sides in (12) are equal. Otherwise, assume $\psi(A) \backslash \psi(B) \succ_\pi \psi(B)$. We conclude the proof by observing that, $\psi(A) \backslash \psi(B) \succ_\pi \psi(B)$ implies that $\psi(B)$ and $\psi(A)$ are π-nested sets.

□

LEMMA 12. *Given sets $X, Z \subseteq \mathscr{A}$ such that $Z \succ_{\pi_{\mathscr{A}}} X$, and the function ψ defined in* (10)*, it holds that either $\psi(Z) \backslash \psi(X) = \emptyset$ or $\psi(Z) \backslash \psi(X) \succ_\pi \psi(X)$.*

Proof. We assume that $\psi(Z) \backslash \psi(X) \neq \emptyset$ and show that this implies $\psi(Z) \backslash \psi(X) \succ_\pi \psi(X)$. We give a proof by contradiction.

Let

$$Z_1 = \{a \in Z \mid \exists a' \in X, \psi(a') = \psi(a)\},$$

be the set of all actions in Z that have the same realized outcome as some action in X. From (11), we have that

$$\psi(Z) \backslash \psi(X) = \psi(Z \backslash Z_1),$$

which is the set of all realized outcomes that are not common between Z and X. Now, suppose that $\psi(Z \backslash Z_1)$ does not succeed $\psi(X)$ in the order π. Then there exists distinct outcomes $e \in \psi(Z \backslash Z_1)$ and $e' \in \psi(X)$ such that $e' \succ_\pi e$. Since ψ maps every action to a single outcome, there exist distinct actions $a \in Z \backslash Z_1$ and $a' \in X$ that correspond to the outcomes e and e' respectively. By definition of the order $\pi_{\mathscr{A}}$, we have that, $a' \succ_{\pi_{\mathscr{A}}} a$. This contradicts the fact that $Z \succ_{\pi_{\mathscr{A}}} X$. Therefore, $\psi(Z \backslash Z_1) \succ_\pi \psi(X)$. □

D.2. Proof of Claim (iii): The Invariance Property

To prove the invariance property, we need to show that Greedy does not choose any action from the set $\widehat{\mathcal{A}}$. Using Lemma 5 (restated below), we show that if Greedy does not pick any of the new actions $\{\widehat{a}_1, \cdots, \widehat{a}_{t-1}\}$ prior to t then it will not pick action $\widehat{a}_t$ at t. Let $\widehat{\mathrm{R}}_t = \{\widehat{r}_1, \cdots, \widehat{r}_{t-1}\}$ denote the set of actions chosen prior to arrival t by Greedy on instance $\widehat{G}$. Suppose that $\widehat{\mathrm{R}}_t \subseteq \mathcal{A}$ and note that this is true for $t = 1$ because $\widehat{\mathrm{R}}_1 = \{\emptyset\}$. Using Lemma 5 with $S = \widehat{\mathrm{R}}_t$, we have,

$$F(\widehat{a}_t \mid \widehat{\mathrm{R}}_t) \leq \max_{a \in A_t} F(a \mid \widehat{\mathrm{R}}_t).$$

Thus, Greedy chooses an action from the original set of actions (A_t) at arrival, i.e., $\widehat{r}_t \in A_t$ and $\widehat{\mathrm{R}}_{t+1} \subseteq \mathcal{A}$. Now, our claim follows by induction over t.

LEMMA 5. *For any set function f (not necessarily monotone or submodular order), the function F defined in* (3) *satisfies,*

$$F(\widehat{a}_t \mid S) \leq \max_{a \in A_t} F(a \mid S) \qquad \forall t \in [T], S \subseteq \mathcal{A} \backslash A_t.$$

Proof. We being by restating the two main lemmas that we use to prove Lemma 5. For all $a \in \mathcal{A}$, $S \subseteq \mathcal{A} \backslash \{a\}$, and $e \in N(a)$, let

$$w_{e,S} = \sum_{P \in \mathcal{N}(S)} \gamma(P)\, f(e \mid P),$$

denote the marginal reward of outcome $e \in N(a)$ when action set S has been selected.

LEMMA 6. *For any set function f (not necessarily monotone or submodular), every arrival $t \in [T]$, action $a \in A_t$, and set $S \subseteq \mathcal{A} \backslash \{a\}$, the function F defined in* (3) *satisfies,*

$$F(a \mid S) = \sum_{e \in N(a)} p_e\, w_{e,S}.$$

LEMMA 7. *For any set function f (not necessarily monotone or submodular), every arrival $t \in [T]$ and set $S \subseteq \mathcal{A} \backslash A_t$, the function F defined in* (3) *satisfies,*

$$F(\widehat{a}_t \mid S) = \sum_{a \in A_t} y_a^c \left(\sum_{e \in N(a)} p_e\, w_{e,S} \right).$$

We are now ready to prove Lemma 5. First, we have from Lemma 7,

$$\begin{aligned} F(\widehat{a}_t \mid S) &= \sum_{a \in A_t} y_a^c \sum_{e \in N(a)} p_e\, w_{e,S}, \\ &= \sum_{a \in A_t} y_a^c\, F(a \mid S), \\ &\leq \max_{a \in A_t} F(a \mid S). \end{aligned}$$

The second equality follows from Lemma 6. The final equality follows from the fact that $\sum_{a \in A_t} y_a^c = 1$.

$\square$

Proof of Lemma 6. Since $S \cup a \subseteq \mathcal{A}$, we use the original definition of F from equation (1) to express $F(a \mid S)$ as follows,

$$\begin{aligned}
F(a \mid S) &= \sum_{P \in \mathcal{N}} \gamma(P)\left(f(N(S \cup a) \cap P) - f(N(S) \cap P)\right), \\
&= \sum_{P' \in \mathcal{N}(S)} \sum_{e \in N(a)} \gamma(P')\, p_e\, \left(f(P' \cup e) - f(P')\right), \\
&= \sum_{e \in N(a)} \sum_{P' \in \mathcal{N}(S)} \gamma(P')\, p_e\, f(e \mid P'), \\
&= \sum_{e \in N(a)} p_e \left(\sum_{P' \in \mathcal{N}(S)} \gamma(P')\, f(e \mid P') \right).
\end{aligned}$$

The third and fourth identities follow from rearranging the terms. In the second identity, we use the fact for evaluating $F(a \mid S)$, it suffices to focus only on the set of partial mappings $\mathcal{N}(S)$. Formally,

$$\begin{aligned}
&\sum_{P \in \mathcal{N}} \gamma(P)\left(f(N(S \cup a) \cap P) - f(N(S) \cap P)\right) \\
&= \sum_{P' \in \mathcal{N}(S)} \sum_{e \in N(a)} \sum_{P'' \in \mathcal{N} \setminus \mathcal{N}(S \cup a)} \gamma(P')\, p_e\, \gamma(P'')\, \left(f(P' \cup e) - f(P')\right), \\
&= \left(\sum_{P'' \in \mathcal{N} \setminus \mathcal{N}(S \cup a)} \gamma(P'') \right) \sum_{P' \in \mathcal{N}(S)} \sum_{e \in N(a)} \gamma(P')\, p_e\, f(e \mid P' \cup e),
\end{aligned}$$

here, we split $P \in \mathcal{N}$ into $P' = P \cap N(S)$ (the realized outcomes of S), e (the realized outcome of a), and $P'' = P \setminus N(S \cup a)$ (all other realized outcomes). The probability that $P' \cup e$ is the set of realized outcomes of actions in $S \cup a$ is given by $\gamma(P')\, p_e$. Observe that the marginal value $f(e \mid P')$ is independent of P'' and

$$\sum_{P'' \in \mathcal{N} \setminus \mathcal{N}(S \cup a)} \gamma(P'') = \sum_{P'' \in \mathcal{N}(\mathcal{A} \setminus (S \cup a))} \gamma(P'') = 1.$$

□

Proof of Lemma 7. Since $S \cap \widehat{\mathcal{A}} = \emptyset$, we use (4) to obtain,

$$\begin{aligned}
F(\widehat{a}_t \mid S) &= \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \gamma(P_1)\, \alpha^c(P_2)\, f(N_t \cap P_2 \mid N(S) \cap P_1), \\
&\overset{(a)}{=} \sum_{P_2 \in \mathcal{X}} \alpha^c(P_2) \left(\sum_{P_1' \in \mathcal{N}(S)} \gamma(P_1')\, f(N_t \cap P_2 \mid P_1') \right), \\
&\overset{(b)}{=} \sum_{e \in N_t} \left(\sum_{P_2 \in \mathcal{X} \mid P_2 \ni e} \alpha^c(P_2) \right) w_{e,S}, \\
&\overset{(c)}{=} \sum_{a \in A_t} \sum_{e \in N(a)} y_a^c\, p_e\, w_{e,S}. \\
&= \sum_{a \in A_t} y_a^c \sum_{e \in N(a)} p_e\, w_{e,S}.
\end{aligned}$$

Equality (a) follows from a rearrangement of terms and a change of variables (similar to the proof of Lemma 6) that is captured in the following identity,

$$\sum_{P_1 \in \mathcal{N}} \gamma(P_1) f(N_t \cap P_2 \mid N(S) \cap P_1) = \sum_{P_1' \in \mathcal{N}(S)} \gamma(P_1') f(N_t \cap P_2 \mid P_1').$$

To get equality (b) we use the fact that $N_t \cap P_2$ is a singleton (since $P_2 \in \mathcal{X}$) and also use the definition of $w_{e,S}$. Equality (c) follows from the fact that $\sum_{P_2 \in \mathcal{X} \mid P_2 \ni e} \alpha^c(P_2) = y_a^c p_e$. □

Appendix E: Proofs for Results in the RO Model

THEOREM 3. *Greedy is α-competitive ($\alpha \geq 0.5096$) for OSW-SO in the RO model.*

We will prove this theorem using the reduction technique. To use the technique in the RO model, we first show that permuting the arrivals of G and then constructing the corresponding instance without stochastic outcomes is the same as starting with the instance $\widehat{G}$ and then permuting its arrivals.

LEMMA 13. *Given an instance G of OSW-SO, let $H = G_\sigma$ for some permutation σ and let $\widehat{G}_\sigma$ denote the permuted version of $\widehat{G}$. Let $\widehat{H}$ denote the instance of OSW corresponding to H. The instances $\widehat{H}$ and $\widehat{G}_\sigma$ are equivalent.*

Proof. Recall that $\widehat{G}$ is constructed by adding an extra action $\widehat{a}_t$ for every $t \in [T]$, and defining the objective function F using the optimal solution $\alpha^c(\cdot)$ of $\mathrm{OPT}^c(G)$. $\widehat{G}_\sigma$ and $\widehat{G}$ have the same objective function but $\widehat{G}_\sigma$ has a reordered sequence of arrivals. Specifically, action set at the t-th arrival in $\widehat{G}_\sigma$ is given by $A_{\sigma(t)} \cup \widehat{a}_{\sigma(t)}$.

The instance $\widehat{H}$ is constructed by augmenting the action sets $\{A_{\sigma(t)}\}_{t \in [T]}$ of H and defining the objective function $F_{\widehat{H}}$ using the optimal solution $\alpha^c(\cdot)$ of $\mathrm{OPT}^c(H)$. Clearly, $\widehat{H}$ and $\widehat{G}_\sigma$ have an identical sequence of actions sets. Furthermore, since $\mathrm{OPT}^c(H) = \mathrm{OPT}^c(G)$, it follows that $\widehat{H}$ and $\widehat{G}_\sigma$ also have the same objective function, i.e., $F_{\widehat{H}} = F$. □

Proof of Theorem 3. Consider an arbitrary instance G of OSW-SO with a fixed arrival sequence and the corresponding instance $\widehat{G}$ of OSW. Let σ denote a random permutation of the arrival sequence and let $H = G_\sigma$ denote a randomly permuted version of G. From Lemma 13, we have that $\widehat{G}_\sigma$, the randomly permuted version of $\widehat{G}$, is identical to $\widehat{H}$.

From Lemma 3, we have

$$\mathrm{ALG}(\widehat{H}) = \mathrm{ALG}(H) \quad \text{and} \quad \mathrm{OPT}(\widehat{H}) \geq \mathrm{OPT}^c(H).$$

Using the equivalence between $\widehat{H}$ and $\widehat{G}_\sigma$ and the permutation invariance of OPT and OPT^c, we get,

$$\mathrm{R}(\widehat{G}_\sigma) = \mathrm{R}(G_\sigma) \quad \text{and} \quad \mathrm{OPT}(\widehat{G}) = \mathrm{OPT}(\widehat{G}_\sigma) \geq \mathrm{OPT}^c(G_\sigma) = \mathrm{OPT}^c(G).$$

Now, taking expectation over σ,

$$\begin{aligned}E_\sigma[\mathrm{R}(G_\sigma)] &= E_\sigma[\mathrm{R}(\widehat{G}_\sigma)],\\ &\geq \alpha\,\mathrm{OPT}(\widehat{G}_\sigma),\\ &\geq \alpha\,\mathrm{OPT}^c(G),\\ &\geq \alpha\,\mathrm{OPT}(G).\end{aligned}$$

The first inequality follows from the fact that $\widehat{G}_\sigma$ is an instance of OSW and Greedy is α-competitive for OSW in the RO model. The final inequality follows from Lemma 1.

□

Appendix F: Proofs for Results in the UIID Model

F.1. From RO to UIID

The competitive ratio of an online algorithm in the RO model is a lower bound on its competitive ratio in the UIID model. This is because the RO model "subsumes" the UIID model in the following sense: Suppose that the adversary in the RO model generates a random instance using a UIID model. Permuting the arrival sequence uniformly randomly does not change the final distribution over instances. Thus, one can generate instances from a UIID model in the RO model.

Formally, let $\mathcal{H}$ denote the set of realizable instance in a UIID model and let $\boldsymbol{H}$ denote a random instance from $\mathcal{H}$. Let H_σ represent the instance generated by randomly reordering the arrival sequence of instance $H \in \mathcal{H}$. Let β_{RO} denote the competitive ratio of an online algorithm ALG in the RO model. In the RO model, the adversary can pick the worst instance from $\mathcal{H}$. Therefore,

$$\begin{aligned}E_\sigma[\mathrm{ALG}(H_\sigma)] &\geq \beta_{RO}\,\mathrm{OPT}(H) \quad \forall H \in \mathcal{H},\\ E_D[E_\sigma[\mathrm{ALG}(\boldsymbol{H}_\sigma)]] &\geq \beta_{RO}\,E_D[\mathrm{OPT}(\boldsymbol{H})],\\ E_D[\mathrm{ALG}(\boldsymbol{H})] &\geq \beta_{RO}\,E_D[\mathrm{OPT}(\boldsymbol{H})].\end{aligned}$$

Since the last inequality holds for all sets $\mathcal{H}$ and distributions D, ALG is at least β_{RO}-competitive in the UIID model.

F.2. Deterministic Outcomes

Recall that OSW with UIID arrivals is defined by a set of T types with actions sets $\{A_t\}_{t\in[T]}$, a monotone DR submodular objective function $F:\mathbb{Z}_+^{|\mathcal{A}|}\to\mathbb{R}_+$, and a distribution D over $[T]$. In OSW-SO, in addition to the action set $\mathcal{A}$ we have a set N of possible outcomes, a monotone DR objective function $f:\mathbb{Z}_+^{|N|}\to\mathbb{R}_+$, and probabilities $\{p_e\}_{e\in N}$. A random instance $\boldsymbol{H}$ in the UIID model is generated by sampling a sequence of T IID types from D. We use H to denote a realization of $\boldsymbol{H}$ and $\mathcal{H}$ to denote the set of all possible realizations of $\boldsymbol{H}$.

We perform a change of variables to switch from the monotone DR submodular function F to a monotone submodular function F_T over an expanded ground set $\mathcal{A}_T$ with $T\,|\mathcal{A}|$ elements. As we discuss later, this will be helpful for proving Theorem 4. To familiarize the reader we introduce this change here and use the new notation in the proof of Theorem 8. In the following, we use t to index arrival types and τ to index the arrivals in $\boldsymbol{H}$. Let

$$\mathcal{A}_T = \{a_\tau \mid a \in \mathcal{A},\, \tau \in [T]\} \quad \text{and} \quad A_{\tau,t} = \{a_\tau \mid a \in A_t\}\ \forall \tau, t \in [T],$$

where a_τ is a distinct "copy" of action a. Note that $A_{\tau,t}$ is the set of feasible actions at arrival τ of type t. Before proceeding, we illustrate this change with an example.

EXAMPLE 5. Consider an instance with $T = 2$ and two types of arrivals $A_1 = \{a, b\}$ and $A_2 = \{c, d\}$. The new ground set $\mathcal{A}_T = \{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2\}$. For $\tau \in [2]$, the new sets of feasible actions are as follows: $A_{\tau,1} = \{a_\tau, b_\tau\}$ and $A_{\tau,2} = \{c_\tau, d_\tau\}$.

For $S \subseteq \mathcal{A}_T$, let $x_{a,S} = |\{\tau \mid a_\tau \in S\}|$ denote the number of copies of a in S. Let $\boldsymbol{x}_S = (x_{a,S})_{a\in\mathcal{A}} \in \mathbb{Z}_+^{|\mathcal{A}|}$. Define

$$F_T(S) = F(\boldsymbol{x}_S) \quad \forall S \subseteq \mathcal{A}_T.$$

It is not hard to see that F_T is monotone submodular. We illustrate the change of variables with a simple example.

THEOREM 8. *Greedy is $(1-1/e)$-competitive for OSW in the UIID model.*

Proof. Suppose that we add an extra layer of randomness to the UIID instance generation process: Given a random sequence $\boldsymbol{H}$, we permute the arrival sequence uniformly randomly. Let $\boldsymbol{H}_\sigma$ denote the final (permuted) sequence, where σ is a random permutation of $[T]$. Observe that $\boldsymbol{H}_\sigma$ has the same distribution as $\boldsymbol{H}$. Every sequence $\boldsymbol{H}$ describes a unique instance with T arrivals. With a slight abuse of notation, we use $\boldsymbol{H}$ to denote the sequence as well as corresponding instance. Let $O(\boldsymbol{H})$ denote the optimal offline solution (set of actions) on the instance $\boldsymbol{H}$. The optimal offline solution is permutation invariant, i.e., $O(\boldsymbol{H}) = O(\boldsymbol{H}_\sigma)$. Let $\mathrm{R}_\tau(\boldsymbol{H})$ denote the action set chosen by Greedy prior to arrival τ. Let $\mathrm{ALG}(\boldsymbol{H})$ and $\mathrm{OPT}(\boldsymbol{H})$ denote the total reward of the solution output by Greedy and OPT on $\boldsymbol{H}$. Let $E_\sigma[\cdot]$ denote expectation w.r.t. the randomness in σ and let $E_D[\cdot]$ denote expectation w.r.t. the randomness in $\boldsymbol{H}$.

Now, consider two independently drawn UIID sequences, H^1 and H^2. Let H^2_σ denote a randomly permuted version of H^2. Fix an arbitrary index $\tau \in [T-1]$ and let H^3 denote the hybrid instance with $T+\tau-1$ arrivals, constructed by taking the first $\tau - 1$ types of sequence H^1 followed by the T types of H^2_σ. For $\tau' \geq \tau$, let $H(\tau')$ denote the type of arrival τ' in H^3, which is of the same type as the $\tau' - \tau + 1$-th arrival in H^2_σ. Let $A_{\tau,H(\tau')}$ denote the set of feasible actions for arrival τ' (arrival $\tau' - \tau + 1$) in H^3 (in H^2_σ).

Let $\mathrm{R}_\tau(H^1)$ denote the set of actions chosen by Greedy prior to arrival τ on sequence H^1. Greedy also chooses the action set $\mathrm{R}_\tau(H^1)$ prior to τ on sequence H^3. At arrival τ in H^3, Greedy chooses,

$$r_\tau(H^3) = \underset{a_\tau \in A_{\tau, H(\tau)}}{\arg\max}\ F_T(a_\tau \mid \mathrm{R}_\tau).$$

Taking expectation over the randomness in σ, we have,

$$\begin{aligned}
T\, E_\sigma[F_T(r_\tau(H^3) \mid \mathrm{R}_\tau(H^1))] &= T\left[\sum_{\tau' \geq \tau}^{T+\tau-1} \frac{1}{T} \max_{a_{\tau'} \in A_{\tau', H(\tau')}} F_T(a_{\tau'} \mid \mathrm{R}_\tau(H^1))\right], \\
&\geq \sum_{a_{\tau'} \in O(H^2)} F_T(a_{\tau'} \mid \mathrm{R}_\tau(H^1)), \\
&\geq F_T(O(H^2) \mid \mathrm{R}_\tau(H^1)), \\
&= F_T(O(H^2) \cup \mathrm{R}_\tau) - F_T(\mathrm{R}_\tau(H^1)), \\
&\geq F_T(O(H^2)) - F_T(\mathrm{R}_\tau(H^1)),
\end{aligned}$$

here the first inequality follows from the fact that the τ'-th arrival in H^3 has the same type as arrival $\tau' - \tau + 1$ in H^2_σ and

$$\max_{a_{\tau'} \in A_{\tau', H(\tau')}} F_T(a_{\tau'} \mid \mathrm{R}_\tau(H^1)) \geq F_T\left(O(H^2) \cap A_{\tau'-\tau+1,\, H(\tau')} \mid \mathrm{R}_\tau(H^1)\right).$$

The second inequality follows from the submodularity of F_T and the last inequality by monotonicity of F.

Taking expectation over the randomness in H^1 and H^2 (H^3) and using the fact that the τ-th types in H^3 and H^1 have the same distribution, we get,

$$\begin{aligned}
E_D[F_T(r_\tau(\boldsymbol{H}) \mid \mathrm{R}_\tau(\boldsymbol{H}))] &\geq \frac{1}{T}\left[E_D[F_T(O(\boldsymbol{H}))] - E_D[F_T(\mathrm{R}_\tau(\boldsymbol{H}))]\right] \quad \forall \tau \in [T], \\
&= \frac{1}{T}\left[E_D[\mathrm{OPT}(\boldsymbol{H})] - E_D[F_T(\mathrm{R}_\tau(\boldsymbol{H}))]\right] \quad \forall \tau \in [T].
\end{aligned}$$

Note that, $E_D[F_T(r_\tau(\boldsymbol{H}) \mid \mathrm{R}_\tau(\boldsymbol{H}))]$ equals the expected marginal value of the Greedy solution at arrival τ in $\boldsymbol{H}$. Unrolling the recursion, we get

$$E_D[\mathrm{ALG}(\boldsymbol{H})] \geq \left(1 - \left(1 - \frac{1}{T}\right)\right)^T E_D[\mathrm{OPT}(\boldsymbol{H})].$$

$\square$

F.3. Proof of Theorem 4

Recall that an instance of OSW-SO with UIID arrivals is defined by a set of T types with actions sets $\{A_t\}_{t\in[T]}$, outcome sets $\{N_t\}_{t\in[T]}$, probabilities $\{p_e\}_{e\in N}$, monotone DR submodular objective functions $f : \mathbb{Z}_+^{|N|} \to \mathbb{R}_+$ and $F : \mathbb{Z}_+^{|\mathcal{A}|} \to \mathbb{R}_+$, and a distribution D over $[T]$. A random sequence of types $\boldsymbol{H}$ in the UIID model is generated by sampling a sequence of T IID types from D. Let H to denote a realization of $\boldsymbol{H}$ and $\mathcal{H}$ to denote the set of all (T^T) possible realizations of $\boldsymbol{H}$.

Expanding the Ground Sets: When two arrivals in realization H have the same type, their action (outcome) sets are not disjoint. However, we use the disjointedness of action (outcome) sets in several lemmas (such as Lemma 1 and Lemma 5) that are used in the proof of Theorem 2 (for adversarial arrivals). To address this issue, we create T distinct copies of every action and outcome and redefine f (and F) as set functions f_T (and F_T) on expanded ground set N_T (and $\mathcal{A}_T$). This allows us to use the lemmas that we showed previously, although it increases the complexity of our notation.

Recall the definition of the expanded set of actions (introduced in Appendix F.2),

$$\mathcal{A}_T = \{a_\tau \mid a \in \mathcal{A},\, \tau \in [T]\} \quad \text{and} \quad A_{\tau,t} = \{a_\tau \mid a \in A_t\}\ \forall t, \tau \in [T],$$

where we use τ to index arrivals and t to index arrivals types. Recall that $A_{\tau,t}$ is the set of feasible actions at arrival τ of type t. Similarly, we expand the set of outcomes by creating T copies of every outcome. Specifically, let

$$N(a_\tau) = \{e_\tau \mid e \in N(a)\},$$

be the outcome set of action $a_\tau \in \mathcal{A}_T$ with probability

$$p_{e_\tau} = p_e \quad \forall\, e_\tau \in N(a_\tau).$$

The outcome e_τ is a distinct "copy" of outcome e that can only be realized at arrival τ. Let

$$N_T = \cup_{a_\tau \in \mathcal{A}_T} N(a_\tau), \quad N(S) = \cup_{a_\tau \in S} N(a_\tau)\ \forall S \subseteq \mathcal{A}_T, \quad \text{and} \quad N_{\tau,t} = N(A_{\tau,t}). \tag{13}$$

Finally, let

$$\mathcal{N}_T = \{P \subseteq N_T \mid |P \cap N(a_\tau)| = 1\ \forall a_\tau \in \mathcal{A}_T\},$$

denote the set of all possible action to outcome mappings in the expanded ground sets. The outcome sets of distinct actions are now disjoint. Next, we define the objective functions on the expanded ground sets.

Given a set $P \subseteq N_T$, let, $y_{e,P} = |\{\tau \mid e_\tau \in P\}|$ denote the number of copies of e in P. Let $\boldsymbol{y}_P = (y_{e,P})_{e \in N} \in \mathbb{Z}_+^{|N|}$. The new objective functions are

$$f_T(P) = f(\boldsymbol{y}_P) \quad \forall P \subseteq N_T$$

$$\text{and} \quad F_T(S) = \sum_{P \in \mathcal{N}_T} \gamma(P)\, f_T(N(S) \cap P) \quad \forall S \subseteq \mathcal{A}_T. \tag{14}$$

It is not hard to see that f_T and F_T are monotone submodular.

Proof of Theorem 4. Let $I = (G, D)$ denote an occurrence of OSW-SO with UIID arrivals. We claim that there exists an occurrence $\widehat{I} = (\widehat{G}, D)$ of OSW (without stochastic outcomes) with UIID arrivals such that,

$$(i)\ E_D[\mathrm{OPT}(\widehat{\boldsymbol{H}})] \geq E_D[\mathrm{OPT}(\boldsymbol{H})] \quad \text{and} \quad (ii)\ E_D[\mathrm{ALG}(\boldsymbol{H})] = E_D[\mathrm{ALG}(\widehat{\boldsymbol{H}})],$$

here $\boldsymbol{H}$ denotes a random sequence of T types that induces an instance of I as well as $\widehat{I}$. With a slight abuse of notation, we use $\boldsymbol{H}$ to denote the instance of I and $\widehat{\boldsymbol{H}}$ to denote the instance of $\widehat{I}$.

Assuming the truthfulness of (i) and (ii), we have,

$$E_D[\text{ALG}(\boldsymbol{H})] = E_D[\text{ALG}(\widehat{\boldsymbol{H}})] \geq (1-1/e)\, E_D[\text{OPT}(\widehat{\boldsymbol{H}})] \geq \, E_D[\text{OPT}(\boldsymbol{H})],$$

where the first inequality follows from Theorem 8. In the rest of the proof, we construct the occurrence $\widehat{I}$ and then prove claims (i) and (ii).

We begin by introducing some notation. Consider an arbitrary realization $H \in \mathcal{H}$ of $\boldsymbol{H}$. For each arrival $\tau \in [T]$, let $H(\tau)$ denote its type. Since only actions corresponding to the realized arrival types are available, we define

$$A_H = \cup_{\tau \in [T]} A_{\tau, H(\tau)},$$

to be the set of actions that are feasible in realization H. Let

$$\mathcal{X}_H = \{X \mid X \subseteq N(A_H),\, |X \cap N_{\tau, H(\tau)}| = 1 \; \forall \tau \in [T]\}$$

be the collection of all realizable outcome sets that can arise from feasible selections of actions in H (exactly one action from each arrival). We compare Greedy with the upper bound $\text{OPT}^c(H) = \max_{Y \in \Delta(A_H)} F_T^c(Y)$ on the offline benchmark. From Lemma 1, we have,

$$\text{OPT}^c(H) \geq \text{OPT}(H). \tag{15}$$

Thus, $E_D[\text{OPT}^c(\boldsymbol{H})] \geq E_D[\text{OPT}(\boldsymbol{H})]$. Let $Y_H^c = (y^c_{a_\tau, H})_{a_\tau \in A_H}$ denote the optimizer of $\text{OPT}^c(H)$ and let α^c denote the optimal probability distribution over $\mathcal{X}_H$ such that,

$$\sum_{X \in \mathcal{X}_H} \alpha^c_H(X) = 1, \qquad \sum_{X \in \mathcal{X}_H | X \ni e_\tau} \alpha^c_H(X) = p_{e_\tau}\, y^c_{a_\tau, H} \quad \forall e_\tau \in N(a_\tau),\, a_\tau \in A_H, \tag{16}$$

and

$$\text{OPT}^c(H) = F^c(Y_H^c) = \sum_{X \in \mathcal{X}_H} \alpha^c_H(X)\, f_T(X). \tag{17}$$

Construction of $\widehat{I}$: The overall construction is quite similar to the adversarial case, in that we augment the set of actions in G and then extend the function F_T to the new ground set. The main difference is that, in the enlarged action set, we include *multiple* new actions for every arrival type so that we can capture the optimal offline solution for every realization H. For the sake of intuition, as we describe each component of the construction, we first describe it in context of the original (non-expanded) ground set of actions before switching to the expanded ground set where we simply have T copies of all types and actions. Note that the distribution D over types remains unchanged.

New Actions: Consider arrival number $k \in [T]$ in a realization H. Similar to the adversarial case, we introduce a new action $\widehat{a}_{k,H}$ to capture the optimal offline decision at arrival k in H. Doing this for every k and every H gives a set of new actions $\{\widehat{a}_{k,H}\}_{k\in[T],H\in\mathcal{H}}$ that together represent the optimal offline decision at every arrival of every possible instance in $\mathcal{H}$.

In the expanded ground set framework, every action has T copies, resulting in the following set of new actions:

$$\widehat{\mathcal{A}}_T = \{\widehat{a}_{\tau,k,H} \mid \tau \in [T],\, k \in [T], H \in \mathcal{H}\}.$$

Here, the set $\{\widehat{a}_{\tau,k,H}\}_{\tau\in[T]}$ consists of T copies of $\widehat{a}_{k,H}$. Next, we distribute the new actions among the various types. Let $\mathscr{A}_T = \mathcal{A}_T \cup \widehat{\mathcal{A}}_T$ denote the new ground set of actions.

Enlarged Action Set for Each Arrival Type: In the original (non-expanded) ground set, we augment the action set of type t with the new set

$$\widehat{A}_t = \{\widehat{a}_{k,H} \mid H(k) = t,\, k \in [T],\, H \in \mathcal{H}\},$$

that captures the offline decision at every type-t arrival across all realizations in $\mathcal{H}$. In the expanded ground set, this augmentation is applied at the level of individual arrivals. Specifically, for each arrival τ of type t, we define

$$\widehat{A}_{\tau,t} = \{\widehat{a}_{\tau,k,H} \mid H \in \mathcal{H},\, H(k) = t\}.$$

Consequently, arrival τ of type t in instance $\widehat{H}$ of $\widehat{I}$ has the set of feasible actions $A_{\tau,t} \cup \widehat{A}_{\tau,t}$. A key feature of the construction is that feasibility depends only on the arrival type. In particular, the action $\widehat{a}_{\tau,k,H}$ is feasible at arrival τ in every realization $H' \in \mathcal{H}$ that satisfies $H'(\tau) = H(k)$, independent of the rest of the realization.

A Canonical Feasible Solution: Before extending F_T to the enlarged ground set $\mathscr{A}_T$, observe that the set

$$\widehat{A}_H = \{\widehat{a}_{\tau,\tau,H} \mid \tau \in [T]\}$$

is a feasible solution to instance $\widehat{H}$ of $\widehat{I}$, since it includes exactly one (feasible) action from $\widehat{A}_{\tau,H(\tau)}$ for each arrival $\tau \in [T]$. In the next step of the construction, we will define the extended objective function so that this canonical solution satisfies $F_T(\widehat{A}_H) = \mathrm{OPT}^c(H)$.

Objective: We now extend F_T over the ground set $\mathscr{A}_T$. Although $\widehat{G}$ does not have stochastic outcomes, for the purpose of this definition we use the expanded set of outcomes N_T defined in (13) and let

$$N(\widehat{a}_{\tau,k,H}) = N_{\tau,H(k)} \quad \forall\, \widehat{a}_{\tau,k,H} \in \mathscr{A}_T \quad \text{and} \quad N(\widehat{A}_H) = N(A_H). \tag{18}$$

For every $S \subseteq \mathscr{A}_T$ such that $S_1 = S \cap \mathcal{A}_T$ and $S_2 = S \cap \widehat{\mathcal{A}}_T$, let

$$F_T(S) := \sum_{P_1 \in \mathcal{N}_H} \sum_{P_H \in \mathcal{X}_H\ \forall H \in \mathcal{H}} \gamma(P_1) \prod_{H\in\mathcal{H}} \alpha^c_H(P_H)\, f_T\left(\cup_{H\in\mathcal{H}}(N(S_2)\cap P_H) \cup (N(S_1)\cap P_1)\right). \tag{19}$$

We have defined F_T as the expected value of the realized outcomes when we independently sample the mappings P_1 and $\{P_H\}_{H\in\mathcal{H}}$ w.p. $\gamma(P_1)$ and $\{\alpha^c_H(P_H)\}_{H\in\mathcal{H}}$ respectively. When $S_2 \subseteq \widehat{A}_H$ for some H, this definition is consistent with the function defined in (3). In particular,

$$F_T(\widehat{A}_H) = \sum_{P\in\mathcal{X}_H} \alpha^c_H(P)\; f_T\left(N(A_H)\cap P\right) = \sum_{P\in\mathcal{X}_H} \alpha^c_H(P)\; f_T(P) = \mathrm{OPT}^c(H), \tag{20}$$

here we used (18) and the fact that $\sum_{P_1\in\mathcal{N}_T}\gamma(P_1)=1$ and $\sum_{P_{H'}\in\mathcal{X}_{H'}}\alpha^c_{H'}=1$ for every $H'\in\mathcal{H}$. This completes the definition of instance $\widehat{G}$. Similar to the function F that we defined in (3), we have that F_T is monotone submodular. For completeness, we include a proof of this after the present proof concludes (see Lemma 14). Overall, $\widehat{G}$, which is given by feasible action sets $A_{\tau,t}\cup\widehat{A}_{\tau,t}$ and F_T, and the distribution D together define an occurrence $\widehat{I}$ of OSW with UIID arrivals. Next, we show properties (i) and (ii).

To see (i), let $\boldsymbol{H}$ denote a random sequence of T types (and induced instance of I) and let $\widehat{\boldsymbol{H}}$ denote the induced instance of $\widehat{I}$. Observe that

$$E_D[\mathrm{OPT}(\widehat{\boldsymbol{H}})] \geq E_D[F_T(\widehat{A}_{\boldsymbol{H}})] = E_D[\mathrm{OPT}^c(\boldsymbol{H})] \geq E_D[\mathrm{OPT}(\boldsymbol{H})].$$

The first inequality follows from the fact that $\widehat{A}_{\boldsymbol{H}}$ is a feasible solution of instance $\widehat{\boldsymbol{H}}$. The equality follows from (20) and the final inequality follows from (15).

To establish (ii), it is sufficient to show that $\mathrm{ALG}(\widehat{H})=\mathrm{ALG}(H)$. Similar to the adversarial case, we establish this by proving that Greedy does not choose any of the new actions of instance $\widehat{H}$. Specifically, we show that on $\widehat{H}$, Greedy chooses from the original action set $A_{\tau,H(\tau)}$ at every arrival $\tau\in[T]$. The key ingredient is as follows: For every arrival $\tau\in[T]$ of $\widehat{H}$ with feasible action set $A_{\tau,H(\tau)}\cup\widehat{A}_{\tau,H(\tau)}$, and every set $S\subseteq\mathcal{A}_T\backslash A_{\tau,H(\tau)}$, we have

$$F_T(\widehat{a}_{\tau,k,H'}\mid S) \leq \max_{a_\tau\in A_{\tau,H(\tau)}} F_T(a_\tau\mid S) \quad \forall \widehat{a}_{\tau,k,H'}\in\widehat{A}_{\tau,H(\tau)}.$$

Therefore, in terms of marginal value, every action in $\widehat{A}_{\tau,H(\tau)}$ is dominated by the best action in $A_{\tau,H(\tau)}$ provided that S does not contain any action from $\widehat{\mathcal{A}}_T\cup A_{\tau,H(\tau)}$. This is a generalization of Lemma 5, which is formally stated and proved after the proof of this theorem (see Lemma 15). Using this result for $\tau=1$ and $S=\emptyset$, we have that Greedy does not choose a new action at arrival 1. If Greedy does not choose any new action prior to arrival τ, then using the result above we have that Greedy chooses an original action at τ, completing the proof. $\square$

LEMMA 14. *The function F_T defined in* (19) *is monotone submodular.*

Proof. Fix arbitrary sets $P_1\in\mathcal{N}_T$ and $P_H\in\mathcal{X}_H\ \forall H\in\mathcal{H}$. Let $P_2=\cup_{H\in\mathcal{H}}P_H$ and define $\psi: 2^{\mathscr{A}_T}\to 2^{N_T}$ as follows,

$$\psi(S) = (N(S\cap\mathcal{A}_T)\cap P_1)\cup(N(S\cap\widehat{\mathcal{A}}_T)\cap P_2).$$

Observe that F_T is a linear combination of functions of the form $f_T(\psi())$. Since the family of monotone submodular functions is closed under addition, it suffices to show that $f_T(\psi(\cdot)) : 2^{\mathscr{A}_T} \to \mathbb{R}_+$ is a monotone submodular function when f_T is monotone submodular. The rest of the proof mimics the proof of Lemma 4 and we repeat it for completeness.

Monotonicity: Suppose that f_T is monotone. Observe that,

$$\psi(S) = \cup_{x\in S}\psi(x) \quad \forall S \subseteq \mathscr{A}_T. \tag{21}$$

Thus, $f(\psi(\cdot))$ is monotone because f is monotone and $\psi(A) \supseteq \psi(B)$ for all $B \subseteq A$.

Submodularity: Suppose that f_T is a monotone submodular function. Consider an sets $B \subseteq A \subseteq \mathscr{A}_T$, and set $C \subseteq \mathscr{A}_T \backslash A$, we have,

$$\begin{aligned}
f_T(\psi(A\cup C)) - f_T(\psi(A)) &= f_T(\psi(A\cup C)\backslash\psi(A) \mid \psi(A)),\\
&\le f_T(\psi(A\cup C)\backslash\psi(A) \mid \psi(B)),\\
&= f_T(\psi(C)\backslash\psi(A) \mid \psi(B)),\\
&\le f_T(\psi(C)\backslash\psi(B) \mid \psi(B)),\\
&= f_T(\psi(B\cup C)\backslash\psi(B) \mid \psi(B)),\\
&= f_T(\psi(B\cup C)) - f_T(\psi(B)).
\end{aligned}$$

The first inequality follows from submodularity of f_T. The second inequality follows from the monotonicity of f_T. The second and third equalities follow from (21). This proves that $f_T(\psi(\cdot))$ is submodular.

□

LEMMA 15. *For every arrival $\tau \in [T]$ of sequence $H \in \mathcal{H}$ and every set $S \subseteq \mathcal{A}_T \backslash A_{\tau,H(\tau)}$, we have*

$$F_T(\widehat{a}_{\tau,k,H'} \mid S) \le \max_{a_\tau \in A_{\tau,t}} F_T(a_\tau \mid S) \quad \forall \widehat{a}_{\tau,k,H'} \in \widehat{A}_{\tau,H(\tau)}.$$

Proof. Fix an arbitrary action $\widehat{a}_{\tau,k,H'} \in \widehat{A}_{\tau,t}$. For brevity, let $t = H(\tau)$ and note that $H'(k) = t$ because $\widehat{a}_{\tau,k,H'} \in \widehat{A}_{\tau,t}$. We have,

$$\begin{aligned}
F_T(\widehat{a}_{\tau,k,H'} \mid S) &= \sum_{P\in\mathcal{N}_T} \sum_{P_{H'}\in\mathcal{X}_{H'}} \gamma(P)\, \alpha^c_{H'}(P_{H'})\, f_T(N_{\tau,t}\cap P_{H'} \mid N(S)\cap P), && (22)\\
&= \sum_{P\in\mathcal{N}_T} \gamma(P) \left(\sum_{P_{H'}\in\mathcal{X}_{H'}} \sum_{e_\tau\in N_{\tau,t}\cap P_{H'}} \alpha^c_{H'}(P_{H'})\, f_T(e_\tau \mid N(S)\cap P) \right), && (23)\\
&= \sum_{P\in\mathcal{N}_T} \gamma(P) \left(\sum_{e_\tau\in N_{\tau,t}} f_T(e_\tau \mid N(S)\cap P) \left(\sum_{P_{H'}\in\mathcal{X}_{H'} \mid P_{H'}\ni e_\tau} \alpha^c_{H'}(P_{H'}) \right) \right),
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{P \in \mathcal{N}_T} \gamma(P) \left(\sum_{a \in A_{\tau,t}} \sum_{e_\tau \in N(a_\tau)} y^c_{a_\tau,H'}\, p_{e_\tau}\, f_T(e_\tau \mid N(S) \cap P) \right), && (24)\\
&= \sum_{a_\tau \in A_{\tau,t}} y^c_{a_\tau,H'} \left(\sum_{P \in \mathcal{N}_T} \gamma(P) \left(\sum_{e_\tau \in N(a_\tau)} p_{e_\tau}\, f_T(e_\tau \mid N(S) \cap P) \right) \right), \\
&= \sum_{a_\tau \in A_{\tau,t}} y^c_{a_\tau,H'}\, F(a_\tau \mid S), && (25)\\
&\le \max_{a_\tau \in A_{\tau,t}} F(a_\tau \mid S). && (26)
\end{aligned}
$$

We obtain equality (22) by ignoring the realized mapping $P_{H''}$ for every instance H'' ($\neq H'$) because the sets $S \cup \{\widehat{a}_{\tau,k,H'}\}$ and $\widehat{A}_{H''}$ are disjoint for all $H'' \in \mathcal{H}\backslash\{H'\}$. We get equality (23) by using the fact that the set $N_{\tau,t} \cap P_{H'}$ is a singleton. Equality (24) follows from (16). Inequality (26) follows from the fact that $Y^c_{H'} \in \Delta(A_{H'})$, which implies that $\sum_{a_\tau \in A_{\tau,t}} y^c_{a_\tau,H'} = 1$. It remains to show equality (25). For every $a_\tau \in A_{\tau,t}$ and $X \subseteq \mathcal{A}_T \backslash A_{\tau,t}$, we have

$$
\begin{aligned}
F_T(a_\tau \mid X) &= \sum_{P \in \mathcal{N}_T} \gamma(P) \left[f_T(N(X \cup a_\tau) \cap P) - f_T(N(X) \cap P) \right], \\
&= \sum_{P \in \mathcal{N}_T} \gamma(P)\, f_T(N(a_\tau) \cap P \mid N(X) \cap P), \\
&= \sum_{P \in \mathcal{N}_T} \sum_{e_\tau \in P \cap N(a_\tau)} \gamma(P \backslash N(a_\tau))\, p_{e_\tau}\, f_T(e_\tau \mid N(X) \cap P), \\
&\overset{(*)}{=} \sum_{P \in \mathcal{N}_T} \gamma(P \backslash N(a_\tau)) \left[\sum_{e'_\tau \in N(a_\tau)} p_{e'_\tau} \left(\sum_{e_\tau \in N(a_\tau)} p_{e_\tau}\, f_T(e_\tau \mid N(X) \cap (P \cup e'_\tau)) \right) \right], \\
&= \sum_{Q \in \mathcal{N}_T} \gamma(Q) \left(\sum_{e_\tau \in N(a_\tau)} p_{e_\tau}\, f_T(e_\tau \mid N(X) \cap Q) \right),
\end{aligned}
$$

To obtain equation $(*)$, we use the fact that for $X \subseteq \mathcal{A}_T \backslash A_{\tau,t}$, we have,

$$N(X) \cap (P \cup e'_\tau) = N(X) \cap P,$$

and the following identities,

$$\sum_{e'_\tau \in N(a_\tau)} p_{e'_\tau} = \sum_{e \in N(a)} p_e = 1.$$

□

Appendix G: The Single-Arrival Problem and Approximate-Greedy

As noted in Remark 7, the Greedy algorithm solves an instance of the SAP (restated below) at each arrival $t \in [T]$, with weights (marginal reward values) $w_{e,\mathrm{R}_t} = \sum_{P \in \mathcal{N}(\mathrm{R}_t)} \gamma(P)\, f(e \mid P)$.

$$\text{SAP at } t: \qquad \arg\max_{a \in A_t} \sum_{e \in N(a)} p_e\, w_{e,\mathrm{R}_t}.$$

Thus far, we have assumed that the SAP can be solved efficiently at each arrival using value oracles for f and F. However, the problem can be quite challenging to solve in various applications due to the following reasons:

(i) SAP *may be NP-hard:* In certain settings, the set of feasible actions A_t is defined implicitly and can be exponentially large in the size of the instance, making the SAP NP-hard. For example, in online assortment optimization, the SAP is equivalent to the problem of finding a revenue optimal assortment, which is an NP-hard problem for many choice models (Heger and Klein 2024).

(ii) *Computing the weights may be difficult:* To solve the SAP at arrival t, we may need to compute the weight w_{e,R_t} for each $e \in N_t$. In some settings, such as online assortment optimization, the weights can be computed in polynomial-time (see Appendix G.1). However, in general, w_{e,R_t} is a sum of exponentially many terms, making it computationally infeasible to compute the exact value.

These challenges are common in combinatorial optimization, both in online and offline settings (Golrezaei et al. 2014, Asadpour and Nazerzadeh 2016). The standard approach to addressing them is to find an approximately optimal solution to the SAP, which we will discuss in more detail below. Although this may reduce the competitive ratio of the online algorithm, we show that by extending the reduction technique, the impact on the competitive ratio remains the same in both settings with and without stochastic outcomes.

In the following discussion, let $w_e(P) = f(e \mid P)$ for all $P \in \mathcal{N}(\mathrm{R}_t)$. We assume that $w_e(P)$ is easy to compute given P. For brevity, we omit the subscript R_t and use the shorthand $w_e = \sum_{P \in \mathcal{N}(\mathrm{R}_t)} \gamma(P)\, f(e \mid P)$. Note that $w_e = E[w_e(P)]$, where the expectation is taken w.r.t. the stochastic outcomes.

Approximation Oracle for SAP. When SAP is NP-hard, we assume the availability of an oracle that outputs an η-approximate solution for the SAP with $\eta \in (0,1]$. Specifically, for any instance of the SAP, the oracle returns a solution a_t such that:

$$\eta\text{-approximation for SAP:} \qquad \sum_{e \in N(a_t)} p_e\, w_e \geq \eta \left(\max_{a \in A_t} \sum_{e \in N(a)} p_e\, w_e \right).$$

There exists a large body of literature on approximation algorithms for the SAP in various settings, including assortment optimization and stochastic rewards with patience. For further references, we refer to Golrezaei et al. (2014) and Brubach et al. (2025).

Sample Average Approximation of w_e. When w_e cannot be computed exactly, we use a sample average approximation (SAA) of w_e. Specifically, we sample J (partial) mappings $P^j \in \mathcal{N}(\mathrm{R}_t)$ for $j \in [J]$ and compute,

$$\text{SAA of } w_e: \qquad w_e^{apx} = \frac{1}{J} \sum_{j \in [J]} w_e(P^j) \qquad \forall e \in N_t,$$

Note that each partial mapping P^j only includes the realized outcomes of the $t-1$ actions in the set R_t. Using a standard Chernoff bound, we can derive the following result.

LEMMA 16. *For $\delta_t \in (0,1)$, given $J_t = \frac{4\log\frac{|N_t|}{\delta_t}}{\delta_t^2}$ independent samples from $\mathcal{N}(\mathrm{ALG}_t)$, we have*

$$P\Big(|w_e^{apx} - w_e| < \delta_t f(e)\ \forall e \in N_t\Big) \geq 1 - \delta_t.$$

We include the proof in Appendix G.2. Later, we discuss the appropriate values of $\{J_t\}_{t\in[T]}$ to ensure that $\sum_{t\in[T]} \delta_t$ is as small as desired, without requiring any knowledge of the number of arrivals T.

The Approximate-Greedy Algorithm. Let Apx-Greedy denote the modified Greedy algorithm where we select an approximately optimal solution to the SAP instance with approximate weights at each arrival. Specifically, at arrival $t \in [T]$, we compute the approximate weights w_e^{apx} using J_t samples and choose the action determined by the η-approximation oracle for the SAP with weights $\{w_e^{apx}\}_{e\in N_t}$. Let $\beta(\eta, \sum_{t\in T}\delta_t)$ denote the competitive ratio of Apx-ALG. Note that Apx-ALG is a randomized algorithm and the competitive ratio compares the expected performance of the algorithm with the optimal offline. We expect the performance of Apx-ALG to degrade as the quality of the SAP solution deteriorates, i.e., as η decreases and $\sum_{t\in[T]}\delta_t$ increases. We establish the following competitive ratio guarantees with deterministic and stochastic outcomes.

THEOREM 9. *Given an η-approximate oracle for the* SAP *and $J_t = \frac{4\log\frac{N_t}{\delta_t}}{\delta_t^2}$ independent samples of the partial mapping at every arrival $t \in [T]$, the Apx-Greedy algorithm is at least $\left(\frac{\eta}{1+\eta} - \frac{3}{2}\sum_{t\in[T]}\delta_t\right)$-competitive for OSOW in the adversarial model.*

THEOREM 10. *In the adversarial, RO, and UIID models, Apx-Greedy algorithm has the same competitive ratio both with and without stochastic outcomes.*

The proofs of Theorem 9 and Theorem 10 can be found in Appendix G.3 and Appendix G.4, respectively. Note that $\frac{\eta}{1+\eta} > \frac{\eta}{2}$ for all $\eta \in (0,1)$. For deterministic outcomes, we believe that results similar to the one in Theorem 9 should hold in the other arrival models as well as for the Greedy-like algorithms discussed in Section 5. Formally deriving these individual results is beyond the scope of this paper.

Choosing J_t (and δ_t). Ideally, we want $\sum_{t\in[T]}\delta_t = O(\epsilon)$, where $\epsilon \in (0,1)$ is a tunable parameter and the big-O suppresses constant factors. This condition holds when $\delta_t = \frac{\epsilon}{T}\ \forall t \in [T]$, i.e., when we draw $\frac{4}{\epsilon^2}T^2\log\frac{T|N_t|}{\epsilon}$ samples at each arrival. When T is unknown, it suffices to set $\delta_t = \frac{\epsilon}{t^2}$, yielding

$$J_t = \frac{4}{\epsilon^2}t^4\log\frac{t^2|N_t|}{\epsilon} \qquad \forall t \in [T],$$

since $\sum_{t=1}^{+\infty}\frac{\epsilon}{t^2} = \frac{\epsilon\pi^2}{6}$. We note that our goal was not to determine the minimum number of required samples; a smaller number may suffice with further refinements.

For illustration, in the context of online two-sided assortment optimization, where each action corresponds to a subset of I (the set of resources) and each action has at most $|I|+1$ outcomes since an arrival chooses at most one resource, we have $|N_t| = O(|I|\,2^I)$. In the case of stochastic rewards with patience, we get $|N_t| = O(2^I \times |I|!)$, as there are $O(|I|!)$ possible actions, each corresponding to a permutation of the resources. In both cases, $\log|N_t| = O(|I|\log|I|)$, which remains polynomial in the problem size.

G.1. The Single Arrival Problem in Special Cases of OSW-SO

As mentioned earlier, the Single Arrival Problem, SAP, given by

$$\arg\max_{a \in A_t} F(a \mid S),$$

can be a computationally challenging optimization problem. The marginal value $F(a \mid S)$ is the expected increase in total reward due to action a. In the following, we examine the complexity of computing $F(a \mid S)$ in two settings.

Online assortment optimization: In this setting, each action corresponds to a subset of the set of resources I. For the sake of simplicity, consider the unit capacity setting, i.e., $c_i = 1$ for all $i \in I$. This is without loss of generality because we can always split a resource with more than 1 unit of capacity into several identical resources each with unit capacity. Now, when assortment $a \subseteq I$ is shown to arrival t, the arrival chooses resource $i \in a$ with probability $\phi_t(i,a)$. This generates a reward of r_i if no arrival before t has chosen resource i. Let $p_i(t)$ denote the probability of the event that no arrival before t chose resource i. Then,

$$\begin{aligned}
F(a \mid S) &= \sum_{e \in N(a)} p_e \left(\sum_{P \in \mathcal{N}(S)} \gamma(P) \sum_{i \in I} f_i(e \mid P) \right), \\
&= \sum_{e \in N(a)} p_e \left(\sum_{P \in \mathcal{N}(S)} \gamma(P) \sum_{i \in I} r_i \mathbb{1}(i \text{ unmatched in } P \text{ and } e_i = 1) \right), \\
&= \sum_{i \in I} r_i \sum_{e \in N(a) | e_i = 1} p_e \left(\sum_{P \in \mathcal{N}(S)} \gamma(P) \mathbb{1}(i \text{ unmatched in } P) \right), \\
&= \sum_{i \in I} r_i \, p_i(t) \sum_{e \in N(a) | e_i = 1} p_e, \\
&= \sum_{i \in I} r_i \, p_i(t) \, \phi_t(i,a), \\
&= \sum_{i \in a} r_i \, \phi_t(i,a) \, p_i(t).
\end{aligned}$$

Here, we use the fact that $p_i(t) = \sum_{P \in \mathcal{N}(S)} \gamma(P) \mathbb{1}(i \text{ unmatched in } P)$ and $\phi_t(i,a) = 0 \ \forall i \notin a$. Thus, the SAP problem in assortment optimization is equivalent to finding the revenue optimal assortment over ground set I, with item prices given by $\{r_i \, p_i(t)\}_{i \in I}$. The probabilities $p_i(t)$ are easy to compute. Notably, if assortment a_t is shown to arrival t, the probabilities for arrival $t+1$ are given by,

$$p_i(t+1) = p_i(t)(1 - \phi_t(i,a_t)) \quad \forall i \in I.$$

Clearly, $p_i(1) = 1$ for all $i \in I$ and it is straightforward to update the probabilities after each arrival. Therefore, the SAP problem for assortment optimization is as hard or as easy as the problem of finding a revenue optimal assortment.

Two-sided assortment optimization: As we will see, computing the SAP weights exactly can be much more challenging in two-sided assortment optimization. At a high level, this is because the SAP objective aims to maximize the expected total gain in the probability that some resource selects arrival t, and this probability may depend on the choices made by all previous arrivals. Formally, let $T_{i,t}$ denote the set of arrivals up to and including t that choose resource $i \in I$. Let $\phi_i(T_{i,t})$ denote the probability that resource i chooses an arrival from the set $T_{i,t}$. Let a_τ denote the (non-adaptive) assortment shown to arrival $\tau \in [t-1]$. Then the marginal increase in objective from showing assortment a_t to arrival t is as follows:

$$F(a_t \mid \{a_1, \cdots, a_{t-1}\}) = \sum_{i \in I} \left[\sum_{T_{i,t} \subseteq [t]} \left(\prod_{\tau \in T_{i,t}} \phi_\tau(i, a_\tau) \prod_{\tau \in [t] \setminus T_{i,t}} (1 - \phi_\tau(i, a_\tau)) \right) \left(\phi_i(T_{i,t}) - \phi_i(T_{i,t} \setminus \{t\}) \right) \right].$$

Here, $\prod_{\tau \in T_{i,t}} \phi_\tau(i, a_\tau) \prod_{\tau \in [t] \setminus T_{i,t}} (1 - \phi_\tau(i, a_\tau))$ is the probability that $T_{i,t} \subseteq [t]$ is the set of arrivals that choose i and $\phi_i(T_{i,t}) - \phi_i(T_{i,t} \setminus \{t\})$ is the increase in the probability that resource i chooses an arrival, conditioned on the set $T_{i,t}$. Note that $\phi_i(T_{i,t}) - \phi_i(T_{i,t} \setminus \{t\}) = 0$ when $t \notin T_{i,t}$.

G.2. Proof of Lemma 16

LEMMA 16. *Given $J_t = \frac{4 \log \frac{|N_t|}{\delta_t}}{\delta_t^2}$ independent samples from $\mathcal{N}(\mathrm{ALG}_t)$, we have,*

$$P\left(|w_e^{apx} - w_e| < \delta_t f(e) \; \forall e \in N_t\right) \geq 1 - \delta_t.$$

Proof. We first note that all weights w_e and their approximations w_e^{apx} lie in the interval $[0, f(e)]$. This holds because

$$0 \leq f(e \mid P) \leq f(e)$$

for any monotone arrival-consistent submodular-order (and submodular) function f, for every $e \in N_t$ and every $P \subseteq \bigcup_{\tau < t} N_\tau$. We use the following version of the Chernoff-Hoeffding bound (Lemma 9 in Vondrák (2010)):

(Chernoff Bound) Let $X_1, X_2, \cdots, X_n$ be n independent random variables such that $X_i \in [0, b]$ for some $b \in (0,1)$ and for all i. Let $X = \frac{1}{n} \sum_{i \in [n]} X_i$. Then, for all $\delta > 0$,

$$P\left(\left| \frac{1}{n} \sum_{i \in [n]} X_i - X \right| \geq \delta b \right) < e^{-\frac{n\delta^2}{4}}. \tag{27}$$

Fix an arbitrary outcome $e \in N_t$. Using inequality (27) with $\delta = \delta_T$ and $n = \frac{4 \log \frac{|N_t|}{\delta_t}}{\delta_t^2}$, we have,

$$P\left(|w_e^{apx} - w_e| > \delta_t f(e)\right) \leq e^{-\frac{4 \log \frac{|N_t|}{\delta_t} \delta_t^2}{4 \delta_t^2}} = \frac{\delta_t}{|N_t|}.$$

Applying the union bound, we have,

$$P\left(|w_e^{apx} - w_e| \geq \delta_t f(e) \; \forall e \in N_t\right) \leq \delta_t,$$

as desired.

□

G.3. Apx-Greedy in the Adversarial Model

THEOREM 9. *Given an η-approximate oracle for the* SAP *and $J_t = \frac{4 \log \frac{N_t}{\delta_t}}{\delta_t^2}$ independent samples of the partial mapping at every arrival $t \in [T]$, the Apx-Greedy algorithm is at least $\left(\frac{\eta}{1+\eta} - \frac{3}{2} \sum_{t \in [T]} \delta_t\right)$-competitive for OSOW in the adversarial model.*

Proof. Consider an arbitrary instance G of OSOW. Recall that NA-OPT is the offline benchmark for OSOW. Let OPT_{T+1} denote the output of NA-OPT on instance G. Let $o_t = \mathrm{OPT}_{T+1} \cap A_t$ denote the action selected at t by NA-OPT and let $\mathrm{OPT}_t = \{o_1, \cdots, o_{t-1}\}$ for every $t \in [T]$. Note that Apx-Greedy is a randomized algorithm due to the randomness in estimating the weights at each arrival. Let $\boldsymbol{r}_t$ denote the (random) action chosen by Apx-Greedy at arrival t and let $\boldsymbol{R}_t = \{\boldsymbol{r}_1, \cdots, \boldsymbol{r}_{t-1}\}$ denote the set of actions chosen prior to arrival t. Let us fix an arbitrary sample path (by fixing all the samples used for SAA). Let $\boldsymbol{r}_t = r_t$ and let $\boldsymbol{R}_t = \mathrm{R}_t$ on this sample path. We have,

$$
\begin{aligned}
F(\mathrm{OPT}_{T+1}) &\leq F\left(\mathrm{R}_{T+1} \cup \mathrm{OPT}_{T+1}\right), \\
&\leq F\left(\cup_{t \in [T]} \{r_t\}\right) + \sum_{t \in [T]} F\left(\{o_t\} \backslash \{r_t\} \mid \cup_{\tau \in [t-1]} \{r_\tau\}\right), \\
&\leq F(\mathrm{R}_{T+1}) + \sum_{t \in [T]} F(o_t \mid \mathrm{R}_t),
\end{aligned} \tag{28}
$$

here the first and the last inequalities follow from monotonicity of F. The second inequality follows from the submodular order property and Lemma 2. Next, we will consider the expectation over randomness in Apx-R on the RHS of inequality (28). Let $E[\cdot]$ denote the expectation w.r.t. randomness in Apx-Greedy. Using Lemma 17 (which is stated subsequently), we will show that,

$$
E[F(\boldsymbol{r}_t \mid \boldsymbol{R}_t)] \geq \eta\, E[F(o_t \mid \boldsymbol{R}_t)] - (1 + 2\eta)\, \delta_t \max_{a \in A_t} F(a). \tag{29}
$$

First, we prove the main claim under the assumption that inequality (29) holds for all $t \in [T]$. We have,

$$
\begin{aligned}
F(\mathrm{OPT}_{T+1}) &\leq E[F(\boldsymbol{R}_{T+1})] + \sum_{t \in [T]} E[F(o_t \mid \boldsymbol{R}_t)], \\
&\leq E[F(\boldsymbol{R}_{T+1})] + \sum_{t \in [T]} \frac{1}{\eta} \left(E[F(\boldsymbol{r}_t \mid \boldsymbol{R}_t)] + (1+2\eta)\delta_t \max_{a \in A_t} F(a) \right), \\
&= \left(1 + \frac{1}{\eta}\right) E[F(\boldsymbol{R}_{T+1})] + \left(2 + \frac{1}{\eta}\right) \sum_{t \in [T]} \delta_t \left(\max_{a \in A_t} F(a) \right), \\
&\leq \left(1 + \frac{1}{\eta}\right) E[F(\boldsymbol{R}_{T+1})] + \left(2 + \frac{1}{\eta}\right) F(\mathrm{OPT}_{T+1}) \sum_{t \in [T]} \delta_t.
\end{aligned}
$$

The first inequality follows by taking expectation over randomness in Apx-Greedy on both sides of inequality (28). The second inequality follows from (29). The (first) equality follows from the fact that $F(\boldsymbol{R}_{T+1}) = \sum_{t \in [T]} F(\boldsymbol{r}_t \mid \boldsymbol{R}_t)$. To see the final inequality, observe that $F(\mathrm{OPT}_{T+1}) \geq \max_{t \in [T],\, a \in A_t} F(a)$ because the singleton action set $\{a\}$ is a feasible solution for the offline problem. Rearranging the terms on both sides of the final inequality, we obtain,

$$\begin{aligned}\left(1+\frac{1}{\eta}\right) E[F(\boldsymbol{R}_{T+1})] &\geq \left(1-\left(2+\frac{1}{\eta}\right)\sum_{t\in[T]}\delta_t\right) F(\mathrm{OPT}_{T+1}),\\ E[F(\boldsymbol{R}_{T+1})] &\geq \left(\frac{\eta}{1+\eta}-\left(\frac{2\eta+1}{\eta+1}\right)\sum_{t\in[T]}\delta_t\right) F(\mathrm{OPT}_{T+1}),\\ &\geq \left(\frac{\eta}{1+\eta}-\frac{3}{2}\sum_{t\in[T]}\delta_t\right) F(\mathrm{OPT}_{T+1}).\end{aligned}$$

Here, we used the fact that $\frac{2\eta+1}{\eta+1}\leq\frac{3}{2}$ for all $\eta\in(0,1]$.

To complete the proof, we need to establish inequality (29). Fix an arbitrary arrival $t\in[T]$. Since Apx-A-Greedy draws independent samples for the SAA at every arrival, it suffices to show that conditioned $\boldsymbol{R}_t = \mathrm{R}_t$, we have

$$E_t[F(\boldsymbol{r_t}\mid \mathrm{R}_t)] \geq \eta \max_{a'\in A_t} F(a'\mid \mathrm{R}_t) - (1+2\eta)\,\delta_t \max_{a\in A_t} F(a), \tag{30}$$

here $E_t[\cdot]$ denotes the conditional expectation w.r.t. randomness in Apx-Greedy at arrival t, given a fixed sample path prior to t. Observe that $\max_{a'\in A_t} F(a'\mid \mathrm{R}_t)\geq F(o_t\mid \mathrm{R}_t)$. Now, the inequality (30) follows directly from Lemma 17 (stated subsequently), thereby completing the proof.

□

LEMMA 17. *Let a^* denote an optimal solution (action) for the* SAP *with weights w_e and let a^{apx} denote the (random) output of an η-approximation oracle for* SAP *for the (random) instance with weights w_e^{apx}. Then, we have the following inequality,*

$$E_t\left[\sum_{e\in N(a^{apx})} p_e\, w_e^{apx}\right] \geq \eta \sum_{e\in N(a^*)} p_e\, w_e - (1+2\eta)\delta_t \max_{a\in A_t} F(a).$$

Here, the expectation is taken over the randomness in w_e^{apx} at arrival t.

Proof. Consider the event that,

$$|w_e^{apx}-w_e|\leq \delta_t\, f(e) \quad \forall e\in N_t.$$

We use GOOD to denote this event. From Lemma 16, the probability of this event is at least $1-\delta_t$. Conditioned on GOOD, we have w.p. 1,

$$\begin{aligned}\sum_{e\in N(a^*)} p_e\, w_e &\leq \sum_{e\in N(a^*)} p_e\,(w_e^{apx}+\delta_t\, f(e)),\\ &= \sum_{e\in N(a^*)} p_e\, w_e^{apx} + \delta_t \sum_{e\in N(a^*)} p_e\, f(e),\end{aligned}$$

$$\begin{aligned}
&\leq \max_{a' \in A_t} \sum_{e \in N(a')} p_e\, w_e^{apx} + \delta_t \max_{a \in A_t} \left(\sum_{e \in N(a)} p_e\, f(e) \right), \\
&\leq \frac{1}{\eta} \sum_{e \in N(a^{apx})} p_e\, w_e^{apx} + \delta_t \max_{a \in A_t} F(a), \\
&\leq \frac{1}{\eta} \sum_{e \in N(a^{apx})} p_e\, (w_e + \delta_t\, f(e)) + \delta_t \max_{a \in A_t} F(a), \\
&= \frac{1}{\eta} \sum_{e \in N(a^{apx})} p_e\, w_e + \left(1 + \frac{1}{\eta}\right) \delta_t \max_{a \in A_t} F(a). \qquad (31)
\end{aligned}$$

Observe that,

$$\begin{aligned}
&E\left[\sum_{e \in N(a^{apx})} p_e\, w_e^{apx} \right] \\
&= P(\textsc{Good})\; E\left[\sum_{e \in N(a^{apx})} p_e\, w_e^{apx} \mid \textsc{Good} \right] + (1 - P(\textsc{Good}))\; E\left[\sum_{e \in N(a^{apx})} p_e\, w_e^{apx} \mid \neg\textsc{Good} \right], \\
&\geq (1 - \delta_t) \left[\eta \sum_{e \in N(a^*)} p_e\, w_e - (1 + \eta)\; \delta_t \max_{a \in A_t} F(a) \right],
\end{aligned}$$

here we used inequality (31) and the fact that $P(\textsc{Good}) \geq 1 - \delta_t$ to lower bound the first term in the summation and we used the non-negativity of w_e^{apx} to lower bound the second term by 0. Finally, note that,

$$\begin{aligned}
(1 - \delta_t) \left[\eta \sum_{e \in N(a^*)} p_e\, w_e - (1 + \eta)\; \delta_t \max_{a \in A_t} F(a) \right] &\geq (1 - \delta_t)\, \eta \sum_{e \in N(a^*)} p_e\, w_e - (1 + \eta)\; \delta_t \max_{a \in A_t} F(a), \\
&\geq \eta \sum_{e \in N(a^*)} p_e\, w_e - (1 + 2\eta)\; \delta_t \max_{a \in A_t} F(a).
\end{aligned}$$

The second inequality follows from the fact that $\sum_{e \in N(a)} p_e\, w_e \leq \sum_{e \in N(a)} p_e\, f(e) \leq F(a)$ for all $a \in A_t$.

□

G.4. Applying the Reduction Technique to Apx-Greedy

THEOREM 10. *In the adversarial, RO, and UIID models, Apx-Greedy algorithm has the same competitive ratio both with and without stochastic outcomes.*

Proof. As mentioned earlier in Remark 6, to use the reduction technique for an algorithm ALG, it suffices to establish the invariance property for ALG. Specifically, for Apx-Greedy, we need to show that $E[\text{Apx-Greedy}(G)] = E[\text{Apx-Greedy}(\widehat{G})]$, where the expectation is w.r.t. the randomness in Apx-Greedy. At first glance, this may seem challenging because the approximation oracle for the SAP and the sample average approximation of weights may lead to the selection of actions from $\widehat{\mathcal{A}}$; potentially violating the invariance property. We show that this issue has a simple fix. The key observation is that Apx-Greedy is at least $\beta(\eta,\epsilon)$-competitive for *every possible* η-approximation oracle for the SAP. Specifically, it suffices to show that there exists a choice of η-feasible solutions for the SAP instances arising at each arrival of $\widehat{G}$ such that Apx-Greedy$(\widehat{G}) =$ Apx-Greedy(G).

We compare Apx-Greedy(G) and Apx-Greedy$(\widehat{G})$ at the sample-path level. To this end, we define a natural coupling of the executions of Apx-Greedy on instances G and $\widehat{G}$. For each $t \in [T]$, let $\mathcal{P}_t$ denote a sufficiently large collection of i.i.d. random samples from the set $\mathcal{N}\left(\bigcup_{\tau<t} A_\tau\right)$ of partial mappings. Each mapping in $\mathcal{N}\left(\bigcup_{\tau<t} A_\tau\right)$ specifies the realized outcome of every action in $\bigcup_{\tau<t} A_\tau$. We couple the two executions by using the same collection $\mathcal{P}_t$ to compute the SAA weights w_e^{apx} for all outcomes $e \in N_t$, for each $t \in [T]$, in both instances G and $\widehat{G}$. Although the mappings in $\mathcal{P}_t$ may not realize every possible outcome prior to arrival t, this is not required. It suffices that the resulting estimates w_e^{apx} are identical across the two instances and provide sufficiently accurate approximations of w_e, which follows from Lemma 16.

Now, fix an arbitrary sample path of the algorithm. For instance G, let SAP_t denote the SAP instance at arrival t and let r_t denote the action selected at t. Similarly, let $\widehat{\text{SAP}}_t$ and $\widehat{r}_t$ denote the SAP instance and the action selected at arrival t of $\widehat{G}$. We will use induction to show that for all $t \in [T]$, there exists an η-approximate solution to $\widehat{\text{SAP}}_t$ such that $\widehat{r}_t = r_t$.

Base case ($t = 1$): The instances $\widehat{\text{SAP}}_1$ and SAP_1 have identical weights but $\widehat{\text{SAP}}_1$ has additional feasible solutions. In the adversarial and RO models, using Lemma 5, we know that $\widehat{a}_1$ is not an optimal solution of $\widehat{\text{SAP}}_1$. Therefore, action r_1 is an η-approximate solution for $\widehat{\text{SAP}}_1$, and we set $\widehat{r}_1 = r_1$. In the UIID model, we have several new actions at arrival 1. However, by applying Lemma 15 – the UIID specific extension of Lemma 5 – we reach the same conclusion.

Inductive step: Suppose that for some τ, we have $\widehat{r}_t = r_t$ for all $t < \tau$. Now consider the instance $\widehat{\text{SAP}}_\tau$ and SAP_τ, which have the same weights. Using Lemma 5 with $S = \{r_t, \cdots, r_{\tau-1}\}$, we have that $\widehat{a}_\tau$ is not an optimal solution of $\widehat{\text{SAP}}_\tau$. Therefore, action r_τ is an η-approximate solution of $\widehat{\text{SAP}}_\tau$. By the induction hypothesis, we have $\widehat{r}_t = r_t$ for all t. Similar to the base case, we use Lemma 15 to reach this conclusion in the UIID model. □

Appendix H: Missing Details for Greedy-like Algorithms

H.1. Extending the Reduction Technique to RAVO

Recall that Balance is a family of deterministic algorithms for RAVO-DO that favors actions generating the highest *perturbed rewards*. Formally, the algorithm selects the following action at arrival t:

$$\text{Balance:} \quad \arg\max_{a \in A_t} \sum_{i \in I | a_i = 1} r_i \, u\left(c_i, y_i(t)\right),$$

here $y_i(t) = \sum_{a \in \mathrm{R}_t} a_i$ is the total capacity of resource i allocated prior to t. In the presence of stochastic outcomes, $y_{P,i}(t)$ represents the total capacity of i allocated prior to t under a given (partial) action to outcome mapping $P_t \in \mathcal{N}(\mathrm{R}_t)$, and is given by

$$y_{P_t,i}(t) = \sum_{a \in \mathrm{R}_t} \sum_{e \in N(a) \cap P_t} e_i.$$

The non-adaptive verion of Balance for RAVO selects the following action t,

$$\text{Non-adaptive Balance:} \quad \arg\max_{a \in A_t} \sum_{P_t \in \mathcal{N}(\mathrm{R}_t)} \gamma(P_t) \sum_{e \in N(a)} p_e \sum_{i \in I | e_i = 1} r_i \, u\left(c_i, y_{P_t,i}(t)\right).$$

Devanur et al. (2016) showed that an instance of Balance (Algorithm 1 in Devanur et al. (2016)) is $(1-1/e)$-competitive for RAVO-DO when $c_{\min} \to +\infty$. Using the reduction technique, we show that non-adaptive Balance has the same competitive ratio for RAVO and RAVO-DO. Crucially, this equivalence holds in the large capacity regime because the RAVO-DO instance $\widehat{G}$ constructed by the reduction has the same minimum resource capacity $c_{\min}$ as the original RAVO instance G. As a result, non-adaptive Balance is asymptotically $(1-1/e)$-competitive for RAVO in the large capacity regime. ‘ We restate our result for (non-adaptive) Balance.

THEOREM 5. *Any (non-adaptive) Balance algorithm has the same asymptotic competitive ratio for* RAVO *and* RAVO-DO.

Proof. Let ALG denote an instance of Balance with competitive ratio β for RAVO-DO. Consider an arbitrary instance G of RAVO. Since RAVO-DO is a special case of RAVO, it suffices to show that,

$$\frac{\mathrm{ALG}(G)}{\mathrm{OPT}(G)} \geq \beta.$$

We prove this result by extending the reduction technique. Specifically, we show that there exists an instance $\widehat{G}$ of RAVO-DO with the same minimum resource capacity such that $\mathrm{OPT}(\widehat{G}) \geq \mathrm{OPT}^c(G)$ (dominance) and $\mathrm{ALG}(\widehat{G}) = \mathrm{ALG}(G)$ (invariance). To prove invariance, we show that ALG solves an instance of the SAP at each arrival, relying on the fact that Lemmas 6 and 7 hold for arbitrary functions, without requiring monotonicity or submodularity.

The key difference is that we must now ensure that $\widehat{G}$ is an instance of RAVO-DO, rather than OSW or OSOW. As in the original construction, $\widehat{G}$ is defined using an augmented action set $\mathscr{A} = \mathcal{A} \cup \hat{\mathcal{A}}$. However, the objective function requires a slight modification: we introduce a customized objective function $\hat{F}$ tailored to the RAVO-DO setting. We begin by defining this new objective function and then specify the resulting instance $\widehat{G}$ of RAVO-DO.

Objective Function $\widehat{F}$: Let $r_{P_1,P_2,i} = \gamma(P_1)\,\alpha^c(P_2)\,r_i$ for all $P_1 \in \mathcal{N}$, $P_2 \in \mathcal{X}$, and $i \in I$. Consider the objective function,

$$\widehat{F}(S_1 \cup S_2) = \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \sum_{i \in I} r_{P_1,P_2,i} \min\left\{c_i, \sum_{e \in N(S_1) \cap P_1} e_i + \sum_{e \in N(S_2) \cap P_2} e_i\right\} \quad \forall S_1 \subseteq \mathcal{A},\ S_2 \subseteq \widehat{\mathcal{A}}, \quad (32)$$

where $N(S_2) = \cup_{t | \widehat{a}_t \in S_2} N_t$. We examine the differences between $\widehat{F}$ and the function F defined in (3) in Remark 12, which follows this proof. Observe that,

$$\widehat{F}(\widehat{\mathcal{A}}) = \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \sum_{i \in I} r_{P_1,P_2,i} \min\left\{c_i, \sum_{e \in P_2} e_i\right\} = \sum_{P_2 \in \mathcal{X}} \sum_{i \in I} \alpha^c(P_2)\, f_i(P_2) = \mathrm{OPT}^c(G), \quad (33)$$

$$\widehat{F}(S) = \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \sum_{i \in I} r_{P_1,P_2,i} \min\left\{c_i, \sum_{e \in N(S) \cap P_1} e_i\right\} = \sum_{P \in \mathcal{N}(S)} \gamma(P)\, f(P) = F(S) \quad \forall S \subseteq \mathcal{A}. \quad (34)$$

It remains to show that $\widehat{F}$ is the objective for an instance of RAVO-DO with action set $\mathscr{A}$.

Instance of RAVO-DO*:* Consider an instance of RAVO-DO with the set of resources $\widehat{I} = \{(P_1, P_2, i) \mid P_1 \in \mathcal{N},\ P_2 \in \mathcal{X},\ i \in I\}$, where the resource $(P_1, P_2, i) \in \widehat{I}$ has capacity c_i and per unit reward $r_{P_1,P_2,i}$. The (deterministic) outcome of an action $a \in \mathscr{A}$ in this instance is a binary vector $\widehat{e}_a = (\widehat{e}_{a,P_1,P_2,i})_{(P_1,P_2,i) \in \widehat{I}}$ with one component per resource in $\widehat{I}$. Fix P_1, P_2, and i, then for all $a \in \mathcal{A}$, the component $\widehat{e}_{a,P_1,P_2,i} = e_i$, where e_i is the resource i component of the outcome $e \in N(a) \cap P_1$ in instance G. Here, $N(a) \cap P_1$ is a singleton. Similarly, for every action $\widehat{a}_t \in \widehat{\mathcal{A}}$, we have $\widehat{e}_{\widehat{a}_t,P_1,P_2,i} = e_i$ where e_i is the resource i component of the outcome $e \in N_t \cap P_2$. Given these definitions, the reward function of resource (P_1, P_2, i) in this instance of RAVO-DO is given by,

$$f_{P_1,P_2,i}(X) = r_{P_1,P_2,i} \min\left\{c_i, \sum_{a \in \mathscr{A} | \widehat{e}_a \in X} \widehat{e}_{a,P_1,P_2,i}\right\}.$$

It follows that the total reward across all resources is given by the function $\widehat{F}$ defined in (32). Specifically, for all action sets $S_1 \subseteq \mathcal{A}$ and $S_2 \subseteq \widehat{\mathcal{A}}$,

$$\begin{aligned}\sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \sum_{i \in I} f_{P_1,P_2,i}(\cup_{a \in S_1 \cup S_2} \widehat{e}_a) &= \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \sum_{i \in I} r_{P_1,P_2,i} \min\left\{c_i, \sum_{e \in N(S_1) \cap P_1} e_i + \sum_{e \in N(S_2) \cap P_2} e_i\right\}, \\ &= \widehat{F}(S_1 \cup S_2)\end{aligned}$$

Now, observe that $\mathrm{OPT}(\widehat{G}) \geq \widehat{F}(\widehat{\mathcal{A}})$, because $\widehat{\mathcal{A}}$ is a feasible solution to $\widehat{G}$. From (33), we have that $\widehat{F}(\widehat{\mathcal{A}}) = \mathrm{OPT}^c(G)$. Thus, $\mathrm{OPT}(\widehat{G}) \geq \mathrm{OPT}^c(G)$ and we have the dominance property. If the invariance property is also true, then we have,

$$\frac{\mathrm{ALG}(G)}{\mathrm{OPT}(G)} = \frac{\mathrm{ALG}(\widehat{G})}{\mathrm{OPT}(G)} \geq \frac{\mathrm{ALG}(\widehat{G})}{\mathrm{OPT}(\widehat{G})} \geq \beta.$$

Proof of the Invariance Property: To show that $\mathrm{ALG}(G) = \mathrm{ALG}(\widehat{G})$, it suffices to show that ALG produces the same output on both G and $\widehat{G}$. This is sufficient because $\widehat{F}$ and F are identical when restricted to the ground set $\mathcal{A}$ (see (34)). For the randomized PG algorithm, we assume that the random values $u_i \sim D_i$ have been fixed arbitrary for all $i \in I$, and we prove invariance on every sample path.

Let $t \in [T]$ be an arbitrary arrival, and suppose that, prior to arrival t, ALG selects the same set of actions $S \subseteq \mathcal{A}$ on both G and $\widehat{G}$. We will show that ALG selects the same action at arrival t in both cases. Assuming this holds, $\mathrm{ALG}(G) = \mathrm{ALG}(\widehat{G})$ follows by induction over t.

We begin by describing the action selected by (non-adaptive) ALG on instance G. At arrival t, the algorithm selects

$$a_G = \arg\max_{a \in A_t} \sum_{P \in \mathcal{N}(S)} \gamma(P) \sum_{e \in N(a)} p_e \sum_{i \in I | e_i = 1} r_i\, u\left(c_i, y_{P,i}(t)\right).$$

For brevity, let

$$\tilde{f}(e \mid P) = \sum_{i \in I | e_i = 1} r_i\, u\left(c_i, y_{P,i}(t)\right).$$

Using Lemma 6, which holds for any function $\tilde{f}$, we rewrite a_G as follows,

$$a_G = \arg\max_{a \in A_t} \sum_{e \in N(a)} p_e \left(\sum_{P \in \mathcal{N}(S)} \gamma(P) \tilde{f}(e \mid P) \right).$$

We now turn to the action selected by ALG on the constructed instance $\widehat{G}$. At arrival t, ALG selects

$$\begin{aligned}
a_{\widehat{G}} &= \arg\max_{a \in A_t \cup \{\widehat{a}_t\}} \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \sum_{i \in I \,|\, \widehat{e}_{a,P_1,P_2,i} = 1} r_{P_1,P_2,i}\, u\left(c_{P_1,P_2,i}, \sum_{a' \in S} \widehat{e}_{a',P_1,P_2,i} \right), \\
&= \arg\max_{a \in A_t \cup \{\widehat{a}_t\}} \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \gamma(P_1) \alpha^c(P_2) \sum_{i \in I \,|\, \widehat{e}_{a,P_1,P_2,i} = 1} r_i\, u\left(c_i, \sum_{a' \in S} \widehat{e}_{a',P_1,P_2,i} \right), \\
&= \arg\max_{a \in A_t \cup \{\widehat{a}_t\}} \sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \gamma(P_1) \alpha^c(P_2) \sum_{i \in I \,|\, \widehat{e}_{a,P_1,P_2,i} = 1} r_i\, u(c_i, y_{P_1 \cap N(S),i}(t)),
\end{aligned}$$

where the second equality follows from the construction of $\widehat{G}$. The third equality follows from the fact that $S \subseteq \mathcal{A}$, which implies

$$\sum_{a' \in S} \widehat{e}_{a',P_1,P_2,i} = \sum_{a' \in S} \sum_{e \in N(a') \cap P_1} e_i = y_{P_1 \cap N(S),i}(t).$$

To conclude that $a_{\widehat{G}} = a_G$, it suffices to show that the marginal value of the auxiliary action $\widehat{a}_t$ is a convex combination of the marginal values of actions in A_t. This essentially follows from Lemma 7 and we repeat the argument below for completeness.

$$\textstyle\sum_{P_1 \in \mathcal{N}} \sum_{P_2 \in \mathcal{X}} \gamma(P_1) \alpha^c(P_2) \displaystyle\sum_{i \in I \,|\, \widehat{e}_{\widehat{a}_t,P_1,P_2,i} = 1} r_i\, u(c_i, y_{P_1 \cap N(S),i}(t))$$

$$
\begin{aligned}
&= \sum_{P_1\in\mathcal{N}}\sum_{P_2\in\mathcal{X}}\gamma(P_1)\alpha^c(P_2)\sum_{i\in I}(\widehat{e}_{\widehat{a}_t,P_1,P_2,i})\, r_i\, u(c_i, y_{P_1\cap N(S),i}(t))\\
&= \sum_{P_1\in\mathcal{N}}\gamma(P_1)\sum_{i\in I} r_i\, u(c_i, y_{P_1\cap N(S),i}(t))\sum_{P_2\in\mathcal{X}}\alpha^c(P_2)\,\widehat{e}_{\widehat{a}_t,P_1,P_2,i}\\
&= \sum_{P_1\in\mathcal{N}}\gamma(P_1)\sum_{i\in I} r_i\, u(c_i, y_{P_1\cap N(S),i}(t))\sum_{P_2\in\mathcal{X}}\sum_{e\in N_t\cap P_2}\alpha^c(P_2) e_i\\
&= \sum_{P\in\mathcal{N}(S)}\gamma(P)\sum_{i\in I} r_i\, u(c_i, y_{P,i}(t))\sum_{e\in N_t|e_i=1}\left(\sum_{P_2\in\mathcal{X}|P_2\ni e}\alpha^c(P_2)\right)\\
&= \sum_{P\in\mathcal{N}(S)}\gamma(P)\sum_{a\in A_t,\, e\in N(a)}\sum_{i\in I|e_i=1} r_i\, u(c_i, y_{P,i}(t))\,(y_a^c p_e)\\
&= \sum_{a\in A_t} y_a^c\sum_{e\in N(a)} p_e\left(\sum_{P\in\mathcal{N}(S)}\gamma(P)\tilde{f}(e\,|\,P)\right)\\
&\leq \sum_{e\in N(a_G)} p_e\left(\sum_{P\in\mathcal{N}(S)}\gamma(P)\tilde{f}(e\,|\,P)\right)\\
&= \sum_{P\in\mathcal{N}(S)}\gamma(P)\sum_{e\in N(a_G)} p_e\sum_{i\in I|e_i=1} r_i\, u(c_i, y_{P,i}(t))\\
&= \sum_{P_1\in\mathcal{N}}\gamma(P_1)\sum_{e\in N(a_G)\cap P_1}\sum_{i\in I|e_i=1} r_i\, u(c_i, y_{P_1\cap N(S),i}(t))\\
&= \sum_{P_1\in\mathcal{N}}\sum_{P_2\in\mathcal{X}}\gamma(P_1)\alpha^c(P_2)\sum_{e\in N(a_G)\cap P_1}\sum_{i\in I|e_i=1} r_i\, u(c_i, y_{P_1\cap N(S),i}(t))\\
&= \sum_{P_1\in\mathcal{N}}\sum_{P_2\in\mathcal{X}}\gamma(P_1)\alpha^c(P_2)\sum_{i\in I\,|\,\widehat{e}_{a_G,P_1,P_2,i}=1} r_i\, u(c_i, y_{P_1\cap N(S),i}(t)).
\end{aligned}
$$

□

REMARK 12 (DIFFERENCE BETWEEN $\hat{F}$ AND F). Recall that

$$\hat{F}(S_1\cup S_2) = \sum_{P_1\in\mathcal{N}}\sum_{P_2\in\mathcal{X}}\sum_{i\in I} r_{P_1,P_2,i}\min\left\{c_i, \sum_{e\in N(S_1)\cap P_1} e_i + \sum_{e\in N(S_2)\cap P_2} e_i\right\},$$

whereas, from (3), we have

$$F(S_1\cup S_2) = \sum_{P_1\in\mathcal{N}}\sum_{P_2\in\mathcal{X}}\sum_{i\in I} r_{P_1,P_2,i}\min\left\{c_i, \sum_{e\in(N(S_1)\cap P_1)\cup(N(S_2)\cap P_2)} e_i\right\}.$$

Observe that

$$\sum_{e\in(N(S_1)\cap P_1)\cup(N(S_2)\cap P_2)} e_i \leq \sum_{e\in N(S_1)\cap P_1} e_i + \sum_{e\in N(S_2)\cap P_2} e_i.$$

Consequently,

$$F(S_1\cup S_2) \leq \hat{F}(S_1\cup S_2).$$

This distinction is crucial. While $\hat{F}$ corresponds to the objective function of a valid instance of RAVO-DO, the function F does not, in general, correspond to any feasible RAVO-DO instance.

To illustrate this, consider a single resource i with capacity $c_i = 2$, and singleton sets $S_1 = \{a\}$, $S_1' = \{a'\}$ and $S_2 = \{\hat{a}\}$. Suppose there exist mappings P_1 and P_2 such that $N(S_1) \cap P_1 = N(S_2) \cap P_2 = \{e\}$ and $N(S_1') \cap P_1 = \{e'\}$. Then,

$$\hat{F}(S_1 \cup S_2) = \hat{F}(S_1 \cup S_1') = \hat{F}(S_1' \cup S_2) = 2r_i,$$

whereas

$$F(S_1 \cup S_2) = r_i, \quad \text{and} \quad F(S_1 \cup S_1') = F(S_1' \cup S_2) = 2r_i.$$

In a valid instance of RAVO-DO, the objective behaves linearly when there is sufficient remaining resource capacity. The subadditive behavior exhibited by F in this example cannot arise from any valid instance of RAVO-DO, highlighting the difference between $\hat{F}$ and F.

H.2. Competitive Ratio Upper Bounds Using Reduction Technique

Given a set I of resources, let $F = \sum_{i \in I} F_i$ denote the objective function in an instance of OSW-SO, where $F_i(S) = \sum_{P \in \mathcal{N}(S)} \gamma(P) f_i(P)$ is the expected total reward from resource $i \in I$. We assume that $f = \sum_{i \in I} f_i$ and $F = \sum_{i \in I} F_i$ are monotone submodular functions but make no assumptions on the component functions f_i and F_i.

We are interested in GLAs for OSW that are greedy w.r.t. the perturbed objective function $\tilde{F} = \sum_{i \in I} \tilde{F}_i$. Specifically, at arrival t, the algorithm selects

$$\arg\max_{a \in \mathrm{R}_t} \sum_{i \in I} \tilde{F}_i(a \mid \mathrm{R}_t),$$

where

$$\tilde{F}_i(a \mid \mathrm{R}_t) = u_i \left(\{F_j(X)\}_{j \in I, X \subseteq \mathrm{R}_t} \right) F_i(a \mid \mathrm{R}_t),$$

and each u_i is an arbitrary (possibly randomized) function of the current state of resource i, represented by the collection $\{F_j(X)\}_{j \in I, X \subseteq \mathrm{R}_t}$.

THEOREM 6. *The family of* GLA*s defined above has competitive ratio at most 0.5 for OSW.*

Proof. To prove this theorem we leverage Theorem 5 from Mehta and Panigrahi (2012), which states that when $c_{\min} = 1$, no non-adaptive algorithm has a competitive ratio higher than 0.5 for the problem of online matching with stochastic rewards (OMSR), a special case of OSW-SO. By applying the reduction technique in the reverse direction, we show that this result imposes an upper bound of 0.5 on the competitive ratio of all GLAs for OSW.

Formally, we argue by contradiction. Suppose that there exists a GLA for OSW with competitive ratio $0.5 + \epsilon$ for some $\epsilon > 0$. In the first part of the proof, we use the reduction technique to show that this implies the existence of a non-adaptive GLA with competitive ratio $0.5 + \epsilon$ for OSW-SO against the benchmark

OPT^c. Therefore, there exists a non-adaptive algorithm with competitive ratio $0.5+\epsilon$ (against OPT^c) for OMSR. At this point, one might expect an immediate contradiction with Theorem 5 of Mehta and Panigrahi (2012), which rules out competitive ratios above 0.5 for non-adaptive OMSR algorithms when compared against an LP benchmark that upper bounds OPT. However, since the result of Mehta and Panigrahi (2012) is stated with respect to an LP benchmark, an additional step is required. In the second part of the proof, we strengthen their result by showing that the same 0.5 upper bound continues to hold even when performance is measured against OPT^c.

Part 1: In this part, we use the reduction technique to show that a non-adaptive GLA has the same competitive ratio for OSW-SO and OSW. Let G be an instance of OSW-SO with objective functions $f = \sum_{i\in I} f_i$ and $F = \sum_{i\in I} F_i$. Using the original construction from Lemma 3, we construct an instance $\widehat{G}$ of OSW with an extended objective function defined over the enlarged ground set $\mathscr{A}$.

Specifically, the extended objective function is

$$F(S) = \sum_{P_1\in\mathcal{N}}\sum_{P_2\in\mathcal{X}} \gamma(P_1)\,\alpha^c(P_2)\,f\left((N(S_1)\cap P_1)\cup(N(S_2)\cap P_2)\right),$$

for all $S\subseteq\mathscr{A}$, where we define $S_1 = S\cap\mathcal{A}$ and $S_2 = S\cap\widehat{\mathcal{A}}$. We retain the same resource set I and define corresponding extended component functions F_i so that

$$F_i(S) = \sum_{P_1\in\mathcal{N}}\sum_{P_2\in\mathcal{X}} \gamma(P_1)\,\alpha^c(P_2)\,f_i\left((N(S_1)\cap P_1)\cup(N(S_2)\cap P_2)\right) \quad \forall i\in I.$$

This construction defines a valid extension of the original component functions: for every set $S\subseteq\mathcal{A}$¡ the value of $F_i(S)$ remains unchanged, and the extended objective decomposes as $F=\sum_{i\in I} F_i$. The dominance property holds by construction of $\widehat{G}$. To complete the reduction, it remains to establish the invariance property holds for the family of GLAs. As in Lemma 3, we show that the GLA selects the same sequence of actions on both G and $\widehat{G}$.

To this end, consider the GLA applied to the instance $\widehat{G}$, which operates using the extended perturbed marginal value functions of the form

$$\tilde{F}_i(a\,|\,S) = u_i(\{F_j(X)\}_{j\in I,\,X\subseteq S})\;F_i(a\,|\,S), \qquad \forall\, S\subseteq\mathscr{A},\, i\in I.$$

When the functions u_i are randomized, we define a natural coupling between their instantiations on the two instances. Specifically, we couple the random choices underlying u_i so that, for every set $S\subseteq\mathcal{A}$, the value

$$u_i(\{F_j(X)\}_{j\in I,\,X\subseteq S})$$

is identical when evaluated on G and on $\widehat{G}$. This coupling is well defined because, for every $S\subseteq\mathcal{A}$, the collection of values $\{F_j(X)\}_{j\in I,\,X\subseteq S}$ coincide across the two instances. The restriction to subsets of S is essential: if u_i were allowed to depend on all actions revealed at (and prior to) arrival t, i.e., $\cup_{\tau\le t}(A_\tau \cup$

$\{\widehat{a}_\tau\}$), then the presence of additional actions in $\widehat{G}$ could change the perturbation values, breaking the coupling and preventing us from establishing invariance.

It suffices to show that for any set $S \subseteq \mathcal{A}\backslash A_t$,

$$\tilde{F}(\widehat{a}_t \mid S) = \sum_{a\in A_t} y_a \tilde{F}(a \mid S). \tag{35}$$

As an immediate consequence, we obtain $\tilde{F}(\widehat{a}_t \mid S) \leq \max_{a\in A_t} \tilde{F}(a \mid S)$, and the invariance property follows by induction over t, exactly as in the proof of Lemma 3. We now verify (35),

$$\begin{aligned}
\tilde{F}(\widehat{a}_t \mid S) &= \sum_{i\in I} \tilde{F}_i(\widehat{a}_t \mid S), \\
&= \sum_{i\in I} u_i(\{F_j(X)\}_{j\in I, X\subseteq S}) F_i(\widehat{a}_t \mid S), \\
&= \sum_{i\in I} u_i(\{F_j(X)\}_{j\in I, X\subseteq S}) \left(\sum_{a\in A_t} y_a F_i(a \mid S)\right), \\
&= \sum_{i\in I}\sum_{a\in A_t} y_a u_i(\{F_j(X)\}_{j\in I, X\subseteq S}) F_i(a \mid S), \\
&= \sum_{a\in A_t} y_a \sum_{i\in I} \tilde{F}_i(a \mid S).
\end{aligned}$$

All equalities follow directly from the definitions, except the third, which follows from Lemma 7 applied to the functions f_i and F_i. Importantly, Lemma 7 holds for arbitrary functions that are not necessarily non-negative, monotone, or submodular.

Part 2: In this part, we extend Theorem 5 of Mehta and Panigrahi (2012) and show that the upper bound of 0.5 holds even when we compare non-adaptive algorithms against OPT^c. Mehta and Panigrahi (2012) showed the upper bound using the following family of 'hard' instances of OMSR. Consider a bipartite graph between the set of resources, $[n]$, and arrivals $[T]$. Let E denote the set of edges. Every edge in E has success probability p and assume that $\frac{1}{p}$ is an integer. The first $\frac{1}{p}$ arrivals have an edge to every resource. The second $\frac{1}{p}$ arrivals have an edge to every resource except resource 1. In general, arrival $t \in \{\frac{i-1}{p}+1, \frac{i}{p}\}$ has an edge to every resource in $\{i, \cdots, n\}$, and $T = \frac{n}{p}$. In a simplified view, they showed that as $p \to 0$, the total expected reward of a non-adaptive algorithm cannot exceed $0.5n + o(n)$. However, they compare online algorithms for OMSR with the following LP upper bound on OPT.

$$\begin{aligned}
\max_{x_{i,t}\ \forall (i,t)\in E} \quad & \sum_{(i,t)\in E} p\, x_{i,t}, \\
& \sum_{i \mid (i,t)\in E} x_{i,t} \leq 1 \quad \forall t\in[T], \\
& \sum_{t \mid (i,t)\in E} p\, x_{i,t} \leq 1 \quad \forall i\in[n], \\
& x_{i,t} \geq 0 \quad \forall (i,t)\in E.
\end{aligned}$$

It is easy to see that, for the family of instances described above, the LP has optimal value n which is achieved by the following optimal solution: for all $i \in [n]$ we can set $x_{i,t} = 1$ if t is one of the last $\frac{1}{p}$ arrivals that have an edge to i and $x_{i,t} = 0$ for all other arrivals.

We show that the benchmark OPT^c has value at least n for these instances by finding a feasible solution with value n for the optimization problem in OPT^c. For $k \in [\frac{1}{p}]$, let X_k denote the outcome where edge $(i, \frac{i-1}{p} + k)$ succeeds for all $i \in [n]$ and all other edges fail. Consider the probability distribution α such that $\alpha(X) = p$ if $X = X_k$ for some $k \in [\frac{1}{p}]$ and $\alpha(X) = 0$ otherwise. Observe that,

$$\sum_{X \in \mathcal{X}} \alpha(X) = \sum_{k \in [\frac{1}{p}]} \alpha(X_k) = 1.$$

For all $i \in [n]$, let $y_{(i,t)} = 1$ if $t \in \{\frac{i-1}{p} + 1, \frac{i}{p}\}$ and let $y_{(i,t)} = 0$ otherwise. Clearly, $(y_{(i,t)})_{(i,t) \in E} \in \Delta(E)$ and

$$\sum_{k \in [\frac{1}{p}] | (i,t) \in X_k} \alpha(X_k) = y_{(i,t)}\, p \quad \forall (i,t) \in E.$$

We have $f(X_k) = n$ for all $k \in [\frac{1}{p}]$ because every resource is matched in X_k. Thus, $\sum_{X \in \mathcal{X}} \alpha(X) f(X) = \frac{1}{p}\, p\, n = n$, which concludes the proof.

□

H.3. Perturbed Greedy (PG) Algorithm

H.3.1. PG Algorithms Captured by GLA. Inspired by the RANKING algorithm of Karp et al. (1990), the Perturbed Greedy algorithm for vertex-weighted online bipartite matching (OBM) (Aggarwal et al. 2011) matches each arrival t to an available resource with the highest *perturbed reward*, given by

$$r_i\big(1 - e^{y_i}\big),$$

where each y_i is sampled independently and uniformly at random from $[0, 1]$. This algorithm fits naturally within the family of GLAs introduced in Section 5.2, with perturbation functions $u_i = 1 - e^{y_i}$ for all $i \in I$.

Several works have studied natural generalizations of this approach to settings with non-unit inventories, stochastic rewards, and possibly non-binary bids (Albers and Schubert 2021, Vazirani 2023, Udwani 2025a, Goyal and Udwani 2023), all of which are captured as special cases of the GLA framework. More broadly, this framework also encompasses the Perturbed Greedy algorithm of Hathcock et al. (2024), which applies to settings where the reward associated with each resource is given by a matroid rank function.

H.3.2. Perturbed Greedy (PG) Algorithm for RAVO-DO. Consider a natural extension of the family of PG algorithms for the RAVO-DO formulation. Given the set of previously selected actions R_t, a PG algorithm selects the following action at arrival t:

$$\arg\max_{a \in A_t} \sum_{i \in I | a_i = 1} u_i\, r_i\, \min\{1, c_i - y_i(t)\}. \tag{36}$$

where $\{u_i\}_{i\in I}$ are independent random perturbations drawn from a distribution U. We recover the Greedy algorithm when $u_i = u = 1$ for each resource i.

A natural question is whether there exists a PG algorithm with competitive ratio exceeding 0.5 for RAVO-DO. We answer this question in the negative by presenting a simple instance showing that the family of PG algorithms defined by (36) has competitive ratio at most 0.5.

Without loss of generality, assume $u_i \in [0,1]$ for all $i \in I$; this can always be achieved by appropriately scaling the rewards r_i. We further assume that the distribution U is regular in the sense that $E[u_i] := \mu_i > 0$. Now, consider an instance of RAVO-DO with two disjoint groups of resources, I_1 and I_2, each containing m resources, where m is sufficiently large. All resources have unit capacity. Each resource in I_1 has per-unit reward $1+\epsilon$, for some arbitrarily small $\epsilon > 0$, while each resource in I_2 has per-unit reward 1. There are two arrivals. The first arrival has two feasible actions $\{a_1, a_2\}$. Action a_1 uses a unit of every resource in I_1, and action a_2 uses a unit of every resource in I_2. The second arrival has a single feasible action a_1' that uses a unit of every resource in I_1. The optimal offline solution selects actions a_2 and a_1', achieving a total reward of $2m$.

Under a PG algorithm, the perturbed reward of action a_1 is $(1+\epsilon)\sum_{i\in I_1} u_i$. By the law of large numbers, this quantity concentrates within a factor of $(1\pm\delta)$ of $(1+\epsilon)mE[u_i]$ with high probability, where $\delta \to 0$ as $m \to \infty$. Consequently, for any fixed ϵ, there exists a sufficiently large m such that the PG algorithm selects action a_1 with high probability. As a result, the algorithm obtains total reward of $(1+\epsilon)m$, yielding a competitive ratio of at most $(1+\epsilon)\,0.5$, which can be made arbitrarily close to 0.5.

Finally, we note that applying the reduction technique for RAVO to provide upper bounds on the performance of the general family of GLAs for RAVO-DO leads to a violation of the invariance property. Specifically, the reduction for RAVO introduces new resources (see Appendix H.1), and without more restrictive assumptions on the perturbation functions u_i, a GLA may no longer select the same sequence of actions on the original instance G and the reduced instance $\widehat{G}$. As a result, the extended reduction technique developed in Appendix H.1 cannot be directly used to transfer performance guarantees for general GLAs from RAVO to RAVO-DO.

Appendix I: Missing Details for Non-monotone Functions

I.1. Non-Adaptive Algorithm: Challenges and Potential Approaches

In this section, we discuss the challenges of extending the reduction technique to analyze a non-adaptive algorithm for ONSW-SO. First, as noted in Remark 10, Cascade Sampling does not satisfy the invariance property. To address this issue, we consider a simpler algorithm called Greedy Sampling (Algorithm 2).

ALGORITHM 2: Greedy Sampling (Non-adaptive)

Input: Parameter $p \in [0,1]$;

Set $\mathrm{R}_1 = \emptyset$;

for *every arrival* $t \in [T]$ **do**

Find element with maximum marginal value, $a_t = \arg\max_{a \in A_t} F(a \mid \mathrm{R}_t)$;

Choose action a_t w.p. p and choose the null action w.p. $1-p$;

Let r_t denote the chosen action and set $R_{t+1} = R_t \cup \{r_t\}$;

end

At every arrival, Algorithm 2 selects the greedy action a_t w.p. p and selects the null action (0_t) w.p. $1-p$. The action a_t has non-negative marginal value because $F(a_t \mid R_t) \geq F(0_t \mid \mathrm{R}_t) = 0$. The sampling probability p is a parameter that influences the competitive ratio and can be set to any value in $[0,1]$. Choosing $p=1$ gives us the deterministic Greedy algorithm which has a competitive ratio of 0 for ONSW (Ganz et al. 2023).

REMARK 13. Algorithm 2 was introduced by Harshaw et al. (2022), who showed that it is $\frac{p(1-p)}{1+p}$-competitive for ONSW. Note that $\max_{p\in[0,1]} \frac{p(1-p)}{1+p} = 3-2\sqrt{2}$. In Appendix I.2, we provide a short and simplified proof of this result by extending the proof template for Theorem 1, which may be of independent interest.

Challenges with Extending the Reduction Technique: Given an instance G of ONSW-SO, consider the instance $\widehat{G}$ as defined in Section 4.2. It can be verified that Greedy Sampling satisfies the invariance property. Specifically, when it selects an action with a positive value, it does so greedily. By extending Lemma 5, we can show that the greedy action a_t lies in the set $\mathcal{A}$ for each arrival $t \in [T]$.

However, as mentioned earlier, another challenge in extending the reduction technique to ONSW-SO is that the objective F (as defined in (3)) may not be submodular when f is non-monotone. Specifically, in the proof of Lemma 4 (included in Appendix D.1), we rely on both the *monotonicity* and submodularity of f to establish that F is submodular.

We believe that this a fundamental issue. To pinpoint the source of the problem, consider the extreme case where G has deterministic outcomes. In this case, the optimal offline solution, OPT, is a deterministic subset of $\mathcal{A}$. Let o_t denote the action selected at arrival $t \in [T]$ in OPT. To define instance $\widehat{G}$, we extend F over $\mathcal{A} \cup \widehat{\mathcal{A}}$ such that, for each arrival t, the marginal values of the new action $\widehat{a}_t$ mimic those of action o_t. This ensures that Greedy (and Greedy-like algorithms) do not select actions in $\widehat{\mathcal{A}}$, while also guaranteeing that $F(\widehat{\mathcal{A}}) = \mathrm{OPT}(G)$.

When F is monotone and submodular on the ground set $\mathcal{A}$, we can accomplish this extension while preserving the monotonicity and submodularity properties by setting,

$$F(\widehat{a}_t \mid X) = F(o_t \mid X) \quad \forall X \subseteq \mathcal{A}.$$

In particular, we have $F(\widehat{a}_t \mid X \cup o_t) = 0$. However, when F is non-monotone, we must have $F(\widehat{a}_t \mid X \cup o_t) \le F(\widehat{a}_t \mid X)$ in order to preserve submodularity over the expanded ground set. This presents a problem: if $F(\widehat{a}_t \mid X)$ is negative and $F(X \cup o_t) = 0$, we cannot guarantee the non-negativity of F on the expanded ground set (see Example 6). In other words, for non-monotone functions, it is unclear whether we can extend F over the new ground set in a way that preserves both its non-negativity and submodularity, while also ensuing that the marginal values of the new actions mimic those of certain existing actions.

EXAMPLE 6. Consider a ground set $\mathcal{A} = \{1, 2\}$ and let $F(\emptyset) = F(1) = 0$, $F(2) = 1$, and $F(\{1, 2\}) = 0$. We add a new element $\widehat{1}$ and extend F such that, $F(\widehat{1} \mid 2) = F(1 \mid 2) = -1$. To ensure submodularity of F, we need that, $F(\widehat{1} \mid \{1, 2\}) \le F(\widehat{1} \mid 2) = -1$. Observe that the resulting function has a negative value, $F(\{1, \widehat{1}, 2\}) \le F(2) + 2F(1 \mid 2) = -1$.

I.2. Simplified Analysis of Greedy Sampling

In the analysis of many of the algorithms in this paper, we use monotonicity to argue that the (set) union of the optimal solution and the solution of an online algorithm has at least as much value as the optimal solution alone. This is not true for non-monotone functions and the following lemma is the main ingredient that fills the technical gap created in the absence of monotonicity.

LEMMA 18 **(Lemma 2.2 in Buchbinder et al. (2014)).** *Consider a non-negative submodular function w over ground set W. Let $W(p)$ denote a random subset of W where each element appears with probability at most p (not necessarily independently). Then,*

$$E[w(W(p) \cup S)] \ge (1 - p)\, w(S) \quad \forall S \subseteq W.$$

Lemma 18 does not hold for submodular order functions and finding a suitable substitute for this lemma appears to be a challenging technical problem.

THEOREM 11. *Algorithm 2 with sampling probability $p \in [0, 1]$ is $\frac{p(1-p)}{1+p}$-competitive for ONSW with adversarial arrivals.*

Proof. Let R_{T+1} denote the final (random) set of actions output by Algorithm 2. Observe that each action $a \in \mathcal{A}$ appears in R_{T+1} with probability at most p. Let OPT_{T+1} denote the optimal offline solution, i.e., $\mathrm{OPT}_{T+1} = \arg\max_{a_t \in A_t \ \forall t \in [T]} F(\cup_{t \in [T]} \{a_t\})$. Finally, for $t \in [T]$, let $\{r_t\} = \mathrm{R}_{T+1} \cap A_t$, $\mathrm{R}_t = \{r_1, \cdots, r_{t-1}\}$, and $\{o_t\} = \mathrm{OPT}_{T+1} \cap A_t$. Let $E[\cdot]$ denote expectation over the randomness in Algorithm 2.

$$\begin{aligned}
(1 - p)\, F(\mathrm{OPT}_{T+1}) &\le E[F(\mathrm{R}_{T+1} \cup \mathrm{OPT}_{T+1})] \\
&= E\left[F\left(\mathrm{R}_{T+1} \cup \left(\cup_{t \in [T]} \{o_t\} \backslash \{r_t\}\right)\right)\right], \\
&\le E[F(\mathrm{R}_{T+1})] + \sum_{t \in [T]} E[F(\{o_t\} \backslash \{r_t\} \mid \mathrm{R}_t)],
\end{aligned}$$

$$\leq E[F(\mathrm{R}_{T+1})] + \sum_{t\in[T]} \frac{1}{p} E[F(r_t \mid \mathrm{R}_t)],$$
$$= \left(1+\frac{1}{p}\right) E[F(\mathrm{R}_{T+1})].$$

Assuming correctness, this completes the proof. Now, the first inequality follows from Lemma 18 with $w = F$, $W(p) = \mathrm{R}_{T+1}$ and $S = \mathrm{OPT}_{T+1}$. The second inequality follows by linearity of expectation and submodularity of F. The third inequality follows from the following claim,

$$E[F(r_t \mid \mathrm{R}_t)] \geq p\, E[F(\{o_t\}\backslash\{r_t\} \mid \mathrm{R}_t)] \quad \forall t \in [T].$$

To show this claim, we fix $t \in [T]$ and let $a_t = \arg\max_{a\in A_t} F(a \mid \mathrm{R}_t)$. Note that $F(a \mid \mathrm{R}_t) \geq \max\{0,\, F(\{o_t\}\backslash\{r_t\} \mid \mathrm{R}_t)\}$ because $0_t \in A_t$. Let $E_t[\cdot]$ denote expectation w.r.t. the random sampling of action a_t. We have,

$$E_t[F(r_t \mid \mathrm{R}_t)] = p\, F(a_t \mid \mathrm{R}_t) + (1-p) F(0_t \mid \mathrm{R}_t) = p\, F(a_t \mid \mathrm{R}_t) \geq p\, F(\{o_t\}\backslash\{r_t\} \mid \mathrm{R}_t)\,.$$

□

I.3. Analysis of Cascade Sampling

THEOREM 7. *Cascade Sampling (Algorithm 1) is* 0.25*-competitive for ONSW-SO in the adversarial model.*

Proof. Since Cascade Sampling is both randomized and adaptive, we need substantial new notation for this proof that will be introduced at various points as needed to clarify the argument.

Let ALG denote the Cascade Sampling algorithm and let OPT denote the optimal adaptive offline benchmark. Fix an arbitrary realized mapping $P \in \mathcal{N}$ for both ALG and OPT. With P fixed, every action maps to a unique outcome and OPT outputs a deterministic set of actions. Let e_t^O and e_t^A denote the realized outcomes of the actions chosen by ALG and OPT at arrival t. Let $\mathcal{E}^A = \{e_1^A, e_2^A, \cdots, e_T^A\}$ and $\mathcal{E}^O = \{e_1^O, e_2^O, \cdots, e_T^O\}$ denote the set of all outcomes in ALG and OPT respectively. Let $\mathcal{E}_t^A = \{e_1^A, \cdots, e_{t-1}^A\}$ denote the set of realized outcomes in ALG prior to arrival t.

Let $E[\cdot]$ denote expectation w.r.t. the randomness in ALG. Note that every outcome $e \in P$ appears in $\mathcal{E}^A$ with probability at most 0.5 (not necessarily independently) because (excluding the null action) ALG does not choose any action w.p. more than 0.5. Thus,

$$\begin{aligned}
0.5\, f(\mathcal{E}^O) &\leq E[f\left(\mathcal{E}^A \cup \mathcal{E}^O\right)],\\
&= E\left[f\left(\mathcal{E}^A \cup \left(\cup_{t\in[T]}\{e_t^O\}\backslash\{e_t^A\}\right)\right)\right],\\
&\leq E[f(\mathcal{E}^A)] + \sum_{t\in[T]} E\left[f\left(\{e_t^O\}\backslash\{e_t^A\} \mid \mathcal{E}_t^A\right)\right] \qquad (37)
\end{aligned}$$

We use Lemma 18 with $w = f$, $W(0.5) = \mathcal{E}^A$, and $S = \mathcal{E}^O$ to get the first inequality. The second inequality follows by linearity of expectation and submodularity of f.

Thus far, we fixed P and followed the proof template of Theorem 1 to derive upper and lower bounds on $f(\mathcal{E}^A \cup \mathcal{E}^O)$. Now, we consider inequality (37) in expectation w.r.t. the randomness in the mapping P. We begin with some notation. Let $\boldsymbol{P}$ denote a random mapping from actions to outcomes. We decompose $\boldsymbol{P}$ into $\boldsymbol{P}_t$ and $\boldsymbol{P}_{-t}$, where $\boldsymbol{P}_t$ is the random mapping from A_t to N_t and $\boldsymbol{P}_{-t}$ is the random mapping from $\mathcal{A}\backslash A_t$ to $N\backslash N_t$. Let $E_{\boldsymbol{P}}[\cdot], E_{\boldsymbol{P}_t}[\cdot], E_{\boldsymbol{P}_{-t}}[\cdot]$ denote expectation w.r.t. the randomness in $\boldsymbol{P}$, $\boldsymbol{P}_t$, and $\boldsymbol{P}_{-t}$ respectively. Note that the sets $\mathcal{E}^A$ and $\mathcal{E}^O$ are now random sets but but we continue to use the original notation for brevity. Also, the set of outcomes $\mathcal{E}^A_t$ is independent of the mapping $\boldsymbol{P}_t$. Taking expectation w.r.t. $\boldsymbol{P}$ on both sides of inequality (37), we have,

$$
\begin{aligned}
0.5E_{\boldsymbol{P}}[f(\mathcal{E}^O)] &\leq E_{\boldsymbol{P}}[E[f(\mathcal{E}^A)]] + \sum_{t\in[T]} E_{\boldsymbol{P}}\left[E\left[f\left(\{e^O_t\}\backslash\{e^A_t\} \mid \mathcal{E}^A_t\right)\right]\right], \\
&\leq E_{\boldsymbol{P}}[E[f(\mathcal{E}^A)]] + \sum_{t\in[T]} E_{\boldsymbol{P}_{-t}}\left[E_{\boldsymbol{P}_t}\left[E\left[f\left(\{e^O_t\}\backslash\{e^A_t\} \mid \mathcal{E}^A_t\right)\right]\right] \,\middle|\, \boldsymbol{P}_{-t} = P_{-t}\right], \qquad (38)\\
&\leq E_{\boldsymbol{P}}[E[f(\mathcal{E}^A)]] + \sum_{t\in[T]} E_{\boldsymbol{P}_{-t}}\left[E_{\boldsymbol{P}_t}\left[E\left[f\left(e^A_t \mid \mathcal{E}^A_t\right)\right]\right] \,\middle|\, \boldsymbol{P}_{-t} = P_{-t}\right], \qquad (39)\\
&= E_{\boldsymbol{P}}[E[f(\mathcal{E}^A)]] + \sum_{t\in[T]} E_{\boldsymbol{P}}\left[E\left[f\left(e^A_t \mid \mathcal{E}^A_t\right)\right]\right], \qquad (40)\\
&= 2\,E_{\boldsymbol{P}}[E[f(\mathcal{E}^A)]].
\end{aligned}
$$

First, note that the final inequality gives us the desired lower bound on competitive ratio of ALG because $E_{\boldsymbol{P}}[f(\mathcal{E}^O)]$ is the objective value of the offline benchmark and $E_{\boldsymbol{P}}[E[f(\mathcal{E}^A)]]$ is the objective value of ALG. Inequality (38) and identity (40) follow from the law of total expectation (also called the tower rule). To complete the proof, we need to prove inequality (39). To this end, it suffices to show that the following inequality holds for every $t \in [T]$, conditioned on $\boldsymbol{P}_{-t} = P_{-t}$:

$$
E_{\boldsymbol{P}_t}\left[E\left[f\left(\{e^O_t\}\backslash\{e^A_t\} \mid \mathcal{E}^A_t\right)\right]\right] \leq E_{\boldsymbol{P}_t}\left[E\left[f\left(e^A_t \mid \mathcal{E}^A_t\right)\right]\right]. \qquad (41)
$$

Proof of (41)*:* Let o_t denote the action chosen by OPT at arrival t. Since OPT is non-anticipatory, the selection of o_t is independent of $\boldsymbol{P}_t$. Let $\{a_1, a_2, \cdot, a_{m_t}\}$ denote the ordered set of actions with non-negative marginal value at arrival t in ALG, sorted in descending order of marginal values. Note that inequality (41) holds trivially when $o_t \notin \{a_1, \cdots, a_{m_t}\}$, since the left-hand-side of (41) is at most zero, while the right-hand-side is non-negative. Similarly, if $m_t = 0$, i.e., there are no actions with non-negative marginal value at t, then the right-hand-side is zero whereas the left-hand-side is at most zero. So we assume that $m_t \geq 1$ and $o_t = a_\ell$ for some $\ell \in [m_t]$.

Recall that ALG chooses action a_ℓ with probability $2^{-\ell}$. Interchanging the order of expectations on the right-hand-side of inequality (41) we get,

$$
E_{\boldsymbol{P}_t}\left[E\left[f\left(e^A_t \mid \mathcal{E}^A_t\right)\right]\right] = E\left[E_{\boldsymbol{P}_t}\left[f\left(e^A_t \mid \mathcal{E}^A_t\right)\right]\right] = \sum_{j\in[m_t]} 2^{-j} \sum_{e\in N(a_j)} p_e\, f(e \mid \mathcal{E}^A_t),
$$

here we ignored the null action because it contributes no value. Now, we make two important observations. First, with probability $1-1/2^{\ell}$, ALG does not choose a_ℓ (which is the same as o_t), and the set $\{e_t^O\}\backslash\{e_t^A\} = \{e_t^O\} = \boldsymbol{P}_t \cap N(a_\ell)$. With probability $1/2^{\ell}$, ALG chooses a_ℓ and we have $\{e_t^O\}\backslash\{e_t^A\} = \emptyset$. Thus,

$$E_{\boldsymbol{P}_t}\left[E\left[f\left(\{e_t^O\}\backslash\{e_t^A\}\mid \mathcal{E}_t^A\right)\right]\right] = (1-2^{-\ell})\sum_{e\in N(a_\ell)} p_e\, f(e\mid \mathcal{E}_t^A).$$

The second observation is simply that,

$$\sum_{j\le \ell} 2^{-j} = 1-2^{-\ell}.$$

Using these two observations along with the fact that every action in $\{a_1, a_2, \cdots, a_{m_t}\}$ has non-negative marginal value we obtain,

$$\begin{aligned}
E_{\boldsymbol{P}_t}\left[E\left[f\left(\{e_t^O\}\backslash\{e_t^A\}\mid \mathcal{E}_t^A\right)\right]\right] &= (1-2^{-\ell})\sum_{e\in N(a_\ell)} p_e\, f(e\mid \mathcal{E}_t^A),\\
&= \sum_{j\in[\ell]}\left[2^{-j}\sum_{e\in N(a_\ell)} p_e\, f(e\mid \mathcal{E}_t^A)\right],\\
&\le \sum_{j\in[\ell]}\left[2^{-j}\sum_{e\in N(a_j)} p_e\, f(e\mid \mathcal{E}_t^A)\right],\\
&\le \sum_{j\in[m_t]}\left[2^{-j}\sum_{e\in N(a_j)} p_e\, f(e\mid \mathcal{E}_t^A)\right],\\
&= E_{\boldsymbol{P}_t}\left[E\left[f\left(\{e_t^A\}\mid \mathcal{E}_t^A\right)\right]\right].
\end{aligned}$$

The two inequalities follow from the fact that, $\sum_{e\in N(a_1)} p_e\, f(e\mid \mathcal{E}_t^A) \ge \sum_{e\in N(a_2)} p_e\, f(e\mid \mathcal{E}_t^A) \ge \cdots \ge \sum_{e\in N(a_\ell)} p_e\, f(e\mid \mathcal{E}_t^A) \ge \cdots \ge \sum_{e\in N(a_{m_t})} p_e\, f(e\mid \mathcal{E}_t^A) \ge 0.$ □

REMARK 14 (A-GREEDY IS 0.5-COMPETITIVE FOR OSOW-SO). For several special cases of OSOW-SO, it is known that A-Greedy achieves a competitive ratio of 0.5. However, in Example 2, we presented an instance of OSW-SO in which A-Greedy attains only half the expected reward of Greedy. In fact, a refinement of the analysis of Cascade Sampling shows that the competitive ratio of A-Greedy is exactly 0.5. Specifically, when f is monotone and admits an arrival-consistent submodular order, we obtain the following strengthened version of inequality (37):

$$f(\mathcal{E}^O) \le f(\mathcal{E}^A) + \sum_{t\in[T]} f\left(\{e_t^O\}\backslash\{e_t^A\}\mid \mathcal{E}_t^A\right). \tag{42}$$

Moreover, by the definition of A-Greedy, we have

$$E_{\boldsymbol{P}_t}\left[f\left(\{e_t^O\}\backslash\{e_t^A\}\mid \mathcal{E}_t^A\right)\right] \le E_{\boldsymbol{P}_t}\left[f\left(\{e_t^A\}\mid \mathcal{E}_t^A\right)\right].$$

Taking expectation over $\boldsymbol{P}$ on both sides of (42), we conclude that

$$E_{\boldsymbol{P}}[f(\mathcal{E}^O)] \le 2\, E_{\boldsymbol{P}}[f(\mathcal{E}^A)],$$

as desired.