

# Revealing Optimal Thresholds for Generalized Secretary Problem via Continuous LP: Impacts on Online $K$ -Item Auction and Bipartite $K$ -Matching with Random Arrival Order

T-H. Hubert Chan\*      Fei Chen\*      Shaofeng H.-C. Jiang\*

## Abstract

We consider the general  $(J, K)$ -secretary problem, where  $n$  totally ordered items arrive in a random order. An algorithm observes the relative merits of arriving items and is allowed to make  $J$  selections. The objective is to maximize the expected number of items selected among the  $K$  best items.

Buchbinder, Jain and Singh proposed a finite linear program (LP) that completely characterizes the problem, but it is difficult to analyze the asymptotic behavior of its optimal solution as  $n$  tends to infinity. Instead, we prove a formal connection between the finite model and an infinite model, where there are a countably infinite number of items, each of which has arrival time drawn independently and uniformly from  $[0, 1]$ .

The finite LP extends to a continuous LP, whose complementary slackness conditions reveal an optimal algorithm which involves  $JK$  thresholds that play a similar role as the  $\frac{1}{e}$ -threshold in the optimal classical secretary algorithm. In particular, for the case  $K = 1$ , the  $J$  optimal thresholds have a nice “rational description”. Our continuous LP analysis gives a very clear perspective on the problem, and the new insights inspire us to solve two related problems.

1. We settle the open problem whether algorithms based only on relative merits can achieve optimal ratio for matroid secretary problems. We show that, for online 2-item auction with random arriving bids (the  $K$ -uniform matroid problem with  $K = 2$ ), an algorithm making decisions based only on relative merits cannot achieve the optimal ratio. This is in contrast with the folklore that, for online 1-item auction, no algorithm can have performance ratio strictly larger than  $\frac{1}{e}$ , which is achievable by an algorithm that considers only relative merits.
2. We give a general transformation technique that takes any monotone algorithm (such as threshold algorithms) for the  $(K, K)$ -secretary problem, and

constructs an algorithm for online bipartite  $K$ -matching with random arrival order that has at least the same performance guarantee.

## 1 Introduction

In computer science, guarantees are often expressed in terms of big-O notation, but occasionally the optimal performance ratio for some problems can be narrowed down to the precise number. One such example is the classical secretary problem [18, 6, 8, 7], which has been popularized in the 1950s, and since then various versions and solutions for the problem have been studied. Buchbinder, Jain and Singh [4] considered the following generalized secretary problem.

**$(J, K)$ -Secretary Problem.** There are  $n$  items (whose merits are given by a total ordering) that arrive in a random order, i.e., the arrival order is picked uniformly at random among all permutations of the  $n$  items. An *algorithm* (which knows the parameter  $n$ ) has  $J$  *quotas* for selecting (or choosing) items. It can observe the relative merits of items arrived so far, and must decide irrevocably if an item is selected when it arrives. The objective is to maximize the expected *payoff*, which is the number of items selected among the best  $K$  items, where expectation is over the random arrival permutation. The *performance ratio* is the expected payoff divided by  $\min\{J, K\}$ . Observe that no adversary is involved in the problem, and hence randomization is unnecessary to achieve optimality, although randomized algorithms can often help the analysis, in which case the performance ratio is an expectation over randomness from both the algorithm and the arrival order.

*Previous Approaches.* The simple  $(1, 1)$ -case is the classical secretary problem, where both the expected payoff and the performance ratio is the probability that the best item is selected. Buchbinder et al. [4] gave a linear programming formulation  $\text{LP}_n(J, K)$  that completely characterizes the generalized secretary problem in the sense that the optimal performance ratio can be inferred from the optimal value of the LP, which gives

\*Department of Computer Science, the University of Hong Kong. [{hubert,fchen,sfjiang}@cs.hku.hk](mailto:{hubert,fchen,sfjiang}@cs.hku.hk)

the optimal expected payoff, and every feasible solution gives a randomized algorithm with the same objective value. Indeed, for the  $(1, 1)$ -case, they showed that the optimal solution corresponds to the well-known algorithm that discards the first  $\frac{n}{e}$  items<sup>1</sup>, and after that selects the first arriving item that is the best among all already observed items. It is well-known that the asymptotic ratio approaches to  $\frac{1}{e}$  from above as  $n$  tends to infinity (indeed for finite  $n$ , the optimal ratio must be rational). For the general  $(J, K)$ -case, it is tedious to analyze the optimal LP solution and its asymptotic behavior. Moreover, the authors in [4] did not analyze the structure of the optimal LP for the general case to derive a “simple” algorithm. They could show that the  $(2, 1)$ -case has optimal asymptotic ratio  $\frac{1}{e} + \frac{1}{e^{1.5}}$  that is achievable by a similar algorithm involving the  $\frac{n}{e}$ -th and  $\frac{n}{e^{1.5}}$ -th items. However, for the  $(1, 2)$ -case, the optimal LP solution is already very complicated, and they only claimed a ratio of 0.572284, which was shown by Gilbert [9] and Gusein-Zade [10] (and also confirmed by our results) to be actually 0.573567.

**Continuous Model.** Bruss introduced the *continuous model* [3], in which there is still a totally ordered set of  $n$  items, but each item picks an arrival time independently and uniformly at random from  $[0, 1]$ . Any algorithm in the previous *step* model can still work in the continuous model by simply ignoring the arrival times, whereas any algorithm in the continuous model can be implemented as a randomized algorithm in the finite step model by first artificially generating  $n$  independent time-stamps uniformly at random from  $[0, 1]$ , and giving the  $i$ -th arriving item the  $i$ -th smallest time-stamp. Although the two models are equivalent in terms of performance ratio, the continuous model is a step towards developing a convenient tool for analyzing the asymptotic ratio, because algorithms in the continuous model might not need to know  $n$  in advance, and can rely on the current time to infer what fraction of items have already been sampled. For instance, for the  $(1, 1)$ -case, an asymptotically optimal algorithm selects the first arriving item after time  $\frac{1}{e}$  that is the best so far.

**Infinite Model.** Immorlica, Kleinberg and Mahdian [11] extended the continuous model to the *infinite model*, in which the totally ordered set of items is the set of positive integers, with a smaller integer having better merit. Each item still picks an arrival time uniformly at random from  $[0, 1]$ , and an algorithm again can only observe the relative merits of arriving items. The authors considered multiple employers competing for the best

item under the infinite model, but there was no formal treatment for the connection with the finite case.

**1.1 Our First Contribution: Optimal Thresholds for  $(J, K)$ -Secretary Problem** We give a formal treatment of the infinite model and define a special class  $\mathcal{A}$  of *piecewise continuous* infinite algorithms. Since the infinite model is ultimately just a tool to analyze the finite (continuous) model, the special class  $\mathcal{A}$  does not need to include all conceivable notions of infinite algorithms. In fact, we just need it to include the class of  *$(J, K)$ -threshold algorithms*, which we describe next.

*Quotas.* It will be helpful to imagine that there are  $J$  quotas  $Q_J, Q_{J-1}, \dots, Q_1$  available for selecting items, where a quota with larger index is used first. For instance,  $Q_J$  is used first, and  $Q_1$  last.

*Potentials.* For  $k \geq 1$ , during the execution of the secretary problem, an arriving item is a  $k$ -potential or has potential  $k$ , if it is the  $k$ -th best item among all those (including itself) that have already arrived. We say an item is a  $k_{\geq}$ -potential (pronounced as “at least  $k$ -potential”) if it is a  $k'$ -potential for some  $k' \leq k$ . Given a positive integer  $m$ , we use the notation  $[m] := \{1, 2, \dots, m\}$ .

**$(J, K)$ -Threshold Algorithm.** Such an algorithm is characterized by  $JK$  thresholds  $(\tau_{j,k})_{j \in [J], k \in [K]}$  such that (i) for all  $j \in [J]$ ,  $0 < \tau_{j,1} \leq \tau_{j,2} \leq \dots \leq \tau_{j,K} \leq 1$ ; and (ii) for all  $k \in [K]$ ,  $0 < \tau_{J,k} \leq \tau_{J-1,k} \leq \dots \leq \tau_{1,k} \leq 1$ . Quotas are used to select items as follows.

- For each  $j \in [J]$  and  $k \in [K]$ , after time  $\tau_{j,k}$ , the algorithm will select a  $k_{\geq}$ -potential, if for some  $j' \geq j$ , quota  $Q_{j'}$  is still available when it arrives, in which case we require available quota  $Q_{j'}$  with the largest  $j'$  be used. Selection is done greedily, i.e., the algorithm will select an arriving item whenever it is possible according to the above rule.

*Interpretation.* Each quota  $Q_j$  has different maturity times. For instance, at time  $\tau_{j,1}$ , the quota can only be used for selecting 1-potential. Hence, condition (i) means that there are  $K$  progressive maturity times, where after time  $\tau_{j,k}$ , quota  $Q_j$  can be used for selecting  $k_{\geq}$ -potentials. Condition (ii) means that quotas with larger indices mature to the next stage earlier.

*Finite Step Model.* Similar thresholds  $(T_{j,k})_{j \in [J], k \in [K]}$  can be defined for the finite step model, where a rule is applied after step  $T_{j,k}$  instead of after time  $\tau_{j,k}$ . In fact, any result for threshold algorithms in the continuous model can be viewed as a result for the step model in which the thresholds are randomized such that each  $T_{j,k}$  is concentrated around  $\tau_{j,k}n$  by Chernoff Bound.

**Optimality in a Nutshell.** We first show (in Proposition 3.4) that any infinite algorithm in  $\mathcal{A}$ , in particular a threshold algorithm, is feasible for some maximiza-

---

<sup>1</sup>From experiments, depending on  $n$ , the optimal threshold  $\frac{n}{e}$  can either be  $\lfloor \frac{n}{e} \rfloor$  or  $\lceil \frac{n}{e} \rceil$ .

tion primal continuous  $\text{LP}_\infty$ . To achieve this step, we use an important property of the infinite model that for any  $t \in [0, 1]$ , the sample space of arrivals in  $[0, t]$  has exactly the same structure as that in  $[0, 1]$ . Next, we describe a procedure (in Section 3.4) to construct a dual feasible solution  $q$  and pick thresholds at the same time such that the corresponding threshold algorithm gives a primal solution  $p$ , which together with  $q$  satisfies the complementary slackness conditions under the continuous primal-dual LP framework by Tyndall [19] and Levinson [17]. This establishes the optimality of some threshold algorithm in  $\text{LP}_\infty$ .

**Connection with the Finite Model.** Observe that a threshold algorithm can be applied readily to the finite (continuous) model for any  $n$ . We show that a threshold algorithm has some *monotone* property (Definition 2.2), which implies that if it has ratio  $x$  under the infinite model, then it has ratio at least  $x$  in the finite model for any  $n$ , implying that  $\text{LP}_n \geq \text{LP}_\infty$  for all  $n$  (Theorem 3.1). Coming to a full circle, we conclude from Proposition 3.5 that  $\lim_{n \rightarrow \infty} \text{LP}_n = \text{LP}_\infty$ , proving that the optimal threshold algorithm in the infinite model is also asymptotically optimal in the finite model.

**THEOREM 1.1. (OPTIMAL THRESHOLDS FOR  $(J, K)$ -SECRETARY PROBLEM)** *There is a procedure to find appropriate thresholds  $(\tau_{j,k})_{j \in [J], k \in [K]}$  such that the corresponding  $(J, K)$ -threshold algorithm is optimal under the infinite model and has a performance ratio of  $\rho_{J,K} := \frac{1}{\min\{J, K\}} \cdot (J - \sum_{j=1}^J (1 - \tau_{j,1})^K)$ . Moreover, under the finite model for any  $n$ , the same threshold algorithm achieves a performance ratio of at least  $\rho_{J,K}$ , which is asymptotically optimal. Furthermore, a result from Kleinberg [14] implies that  $\rho_{K,K} \geq 1 - O(\frac{1}{\sqrt{K}})$ .*

As a by-product, we discover that the optimal threshold algorithm for the  $(J, 1)$ -case has a very nice representation using rational numbers. Gilbert [9] gave exact values of the first four optimal thresholds using simple approximation techniques. We believe it would be too tedious to directly analyze the limiting behavior of the finite model to reach the same conclusion.

**THEOREM 1.2. (( $J, 1$ )-SECRETARY PROBLEM)** *There is a procedure to construct an increasing sequence  $\{\theta_j\}_{j \geq 1}$  of rational numbers such that for any  $J \geq 1$ , the optimal  $(J, 1)$ -threshold algorithm uses thresholds  $\{\tau_{j,1} := \frac{1}{e^{\theta_j}} | 1 \leq j \leq J\}$  (that can be computed in  $O(J^3)$  time) and has expected payoff  $\sum_{j=1}^J \tau_{j,1}$ . For instance,  $\theta_1 = 1$ ,  $\theta_2 = \frac{3}{2}$ ,  $\theta_3 = \frac{47}{24}$ ,  $\theta_4 = \frac{2761}{1152}$ ,  $\theta_5 = \frac{4162637}{1474560}$ ,  $\theta_6 = \frac{380537052235603}{117413668454400}$ .*

For general  $K \geq 2$ , the optimal solution does not have such a regular structure, but the continuous LP

still allows us to compute the exact solution. To describe the optimal solution, we use part of the principal branch of the *Lambert W function* [16]  $W : [-\frac{1}{e}, 0] \rightarrow [-1, 0]$ , where  $z = W(z)e^{W(z)}$  for all  $z \in [-\frac{1}{e}, 0]$ . Gilbert [9] and Gusein-Zade [10] obtained the thresholds  $\tau_{1,2}$  and  $\tau_{1,1}$  by using approximation calculation and a recursion, respectively; to the best of our knowledge, no previous work has computed the exact values for thresholds  $\tau_{2,1}$  and  $\tau_{2,2}$ .

**THEOREM 1.3. (( $J, 2$ )-SECRETARY PROBLEM FOR  $J = 1, 2$ )** *Define the thresholds:  $\tau_{1,2} = \frac{2}{3}$ ;  $\tau_{1,1} = -W(-\frac{2}{3e}) \approx 0.346982$ ;  $\tau_{2,2} \approx 0.517297$  is the solution of:  $x \ln x + \ln x - (2 + 3 \ln \frac{2}{3})x + 1 - \ln \frac{2}{3} = 0$ ;  $\tau_{2,1} = -W(-e^{-c/2}) \approx 0.227788$ , where  $c := -(\ln \tau_{1,1})^2 + 2 \ln \frac{2}{3} \ln \tau_{1,1} + (\ln \tau_{2,2})^2 - 2 \ln \frac{2}{3} \ln \tau_{2,2} - 2\tau_{2,2} + 4 - 2 \ln \frac{2}{3}$ .*

*Then, these thresholds can be used to achieve the following optimal performance ratios:*

- (a)  $\rho_{1,2} := 2\tau_{1,1} - \tau_{1,1}^2 \approx 0.573567$ .
- (b)  $\rho_{2,2} := \frac{1}{2} \cdot ((2\tau_{1,1} - \tau_{1,1}^2) + (2\tau_{2,1} - \tau_{2,1}^2)) \approx 0.488628$ .

Our results on the infinite model are included in our unpublished manuscript [5]. We present a simplified version in Section 3. In particular, we prove Theorem 1.2 in Section 3.3, and Theorems 1.1 and 1.3 in Section 3.4.

**Significance and Impacts.** By considering the infinite model via continuous LP, we prove that a threshold algorithm can attain the optimal performance ratio. We remark that from experiments, the finite  $\text{LP}_n$  can be solved to reveal the exact optimal thresholds for that particular  $n$  by observing precisely at which steps certain variables switch from zero to positive. However, to prove directly that the optimal solution has this threshold structure, one would need to consider messier recurrence relations instead of differential equations.

Although the idea of using complementary slackness conditions to form an optimal primal-dual pair is well-known, as seen in Section 3.4, it is non-trivial to construct the primal and the dual functions in our case. The high level idea is to start from a primal solution corresponding to some threshold algorithm, and to show that the thresholds can be chosen such that a feasible dual can be constructed to satisfy complementary slackness. Because of the two parameters  $(J, K)$ , we use double induction in the proof of Lemma 3.4 to give a construction of such a feasible dual. Careful invariants must be maintained in the induction hypothesis to complete the induction step.

To illustrate our methods, we explicitly construct the optimal thresholds for the  $(2, 2)$ -case in Section 3.4.1, which is already very technical. One could appreciate the elegance of our methods if one tries to

first solve the optimal solution of the finite LP, and then consider its asymptotic behavior.

Our continuous LP analysis gives a very clear perspective on the problem, and the new insights inspire us to answer two related questions, whose connection with our continuous LP analysis is described as follows. (1) In general secretary problems such as online  $K$ -item auction with random arriving bids [14], or in the even more general matroid secretary problems, can an algorithm making decisions based only on relative merits achieve optimal ratio?

Observe that we can use our continuous LP analysis to compute exactly the optimal ratio of the  $(2, 2)$ -secretary problem, which is the same as that of the online 2-item auction when the algorithm can only see relative merits.

Hence, to prove that observing relative merits cannot achieve optimal ratio, we construct a *filtering* algorithm that considers the actual weights, and explicitly prove that it has strictly larger ratio. To the best of our knowledge, this is the *first* result showing that algorithms based only on relative merits cannot achieve optimal ratio for matroid secretary problems. This result is in contrast with the folklore that, for online 1-item auction, no algorithm can have performance ratio strictly larger than  $\frac{1}{e}$ , which is achievable by an algorithm that considers only relative merits.

Although this result might be intuitive, our formal proof involves the computation of the optimal ratio for the  $(2, 2)$ -secretary problem, and the ratio of our filtering algorithm, both of which are technically inspired from the continuous LP analysis.

(2) Under what conditions can a secretary algorithm be modified to solve the edge-weighted online bipartite matching problem with random arrival order [15, 12]?

We prove that a secretary algorithm (such as a threshold algorithm) can be transformed with at least the same performance ratio if it satisfies some *monotone property* (Definition 2.2).

Kleinberg [14] gave an algorithm with performance ratio  $1 - \Theta(\frac{1}{\sqrt{K}})$  for online  $K$ -item auction with random arriving bids. Since Kleinberg's algorithm can be viewed as considering only relative merits, it is also a  $(K, K)$ -secretary algorithm with the same performance guarantee.

However, Kleinberg's algorithm is not monotone (Theorem 2.1). Hence, to prove that the online bipartite  $K$ -matching problem can be solved with performance ratio  $1 - \Theta(\frac{1}{\sqrt{K}})$ , we can first use our continuous LP analysis to show that the optimal asymptotic ratio can be achieved by a threshold algorithm, which is monotone and hence can be transformed to an algorithm for online bipartite matching.

**1.2 Impacts on Online  $K$ -Item Auction with Random Arriving Bids: Can Decisions Based Only On Relative Merits Be Optimal?** The online  $K$ -item auction with random arriving bids [14] is exactly the online bipartite  $K$ -matching problem with one offline node having capacity  $K$ , and hence, as observed by Buchbinder et al. [4], can be solved by any monotone secretary algorithm with at least the same guarantee.

Kleinberg [14] gave a randomized algorithm that essentially makes decisions based only on relative merits and has performance ratio  $1 - O(\frac{1}{\sqrt{K}})$ . It follows that the algorithm can be applied to the  $(K, K)$ -secretary problem with the same performance guarantee, which implies that the optimal  $(K, K)$ -threshold algorithm has ratio  $\rho_{K, K} \geq 1 - O(\frac{1}{\sqrt{K}})$ .

**Is Observing Relative Merits Sufficient?** As described in the survey by Ferguson [7], the original secretary problem was posed such that an adversary chooses  $n$  weights as in the online auction, but the performance ratio of the algorithm is still the probability that the best item is selected. The survey describes how the adversary can use a distribution  $\mathcal{D}$  to sample the  $n$  weights such that any algorithm selects the best item with probability arbitrarily close to  $\frac{1}{e}$ , which means the optimal performance ratio can be achieved by an algorithm that considers only relative merits. We do not find such a similar proof concerning online 1-item auction, but folklore [2] seems to believe that its performance ratio also cannot be larger than  $\frac{1}{e}$ . We give a simple intuition here. One can find a large enough number  $\Lambda > 0$  (depending only on  $n$ ) and apply the mapping  $w \rightarrow \Lambda^w$  to the weights generated from the same distribution  $\mathcal{D}$  such that with high probability the ratio of second largest weight to that of the largest weight is arbitrarily small. Hence, the performance ratio can be made arbitrarily close to the probability that the algorithm selects the best item, but since the algorithm can recover  $w$  from  $\Lambda^w$ , the algorithm receives the same information as before.

Perhaps one might wonder if the online  $K$ -item auction also has the same optimal performance ratio as the  $(K, K)$ -secretary algorithm. On the contrary, we show that in general considering relative weights is strictly better than just observing relative merits, even for the case  $K = 2$ .

**$\alpha$ -Filtering Threshold Algorithm.** Given a parameter  $\alpha \in [0, 1]$ , the  $\alpha$ -*filtering threshold algorithm*  $\mathcal{A}_\alpha$  is a modification of the optimal  $(2, 2)$ -threshold algorithm  $\mathcal{A}_0$ , where there is one extra condition: the algorithm selects an item only if its weight is at least  $\alpha$  times the largest weight seen so far. (Indeed, when  $\alpha = 0$ ,  $\mathcal{A}_\alpha$  is equivalent to  $\mathcal{A}_0$ .) This rule has no effect if the arriv-

ing item is a 1-potential, but it might filter out some 2-potentials.

**THEOREM 1.4. (CONSIDERING WEIGHTS DOES HELP IN ONLINE 2-ITEM AUCTION)** *For  $\alpha = 0.25289$ , the filtering threshold algorithm  $\mathcal{A}_\alpha$  has a performance ratio of at least 0.492006, which is strictly larger than the optimal ratio  $\rho_{2,2} \approx 0.488628$  of an algorithm that makes decisions based only on relative merits.*

We prove Theorem 1.4 in Section 4 under the continuous model, which can be interpreted as randomized algorithms in the finite step model. We remark that obtaining such a result would have been very difficult, if one had not known the thresholds to obtain the optimal performance ratio for the  $(2,2)$ -secretary problem.

**1.3 Impacts on Edge-weighted Online Bipartite  $K$ -Matching with Random Arrival Order: General Transformation Technique from Secretary Problem** We show that a  $(K,K)$ -threshold algorithm can be applied to the edge-weighted online bipartite  $K$ -matching problem, which was first proposed by Körula and Pál [15] and generalizes the weighted secretary problem [1].

**Edge-weighted Online Bipartite  $K$ -Matching.** An adversary fixes a bipartite graph  $G = (L \cup R, E)$  with non-negative edge weights, where  $L$  is the set of *online* nodes and  $R$  is the set of *offline* nodes. Each offline node has *capacity*  $K$ , which is the maximum number of online nodes it can match to. The algorithm has knowledge of only  $R$  and the number  $|L|$  of online nodes initially. The online nodes arrive in a random order. When an online node  $v$  arrives, all the weights  $w_{vr}$ 's of edges between  $v$  and the offline nodes  $r$  in  $R$  are revealed to the algorithm. The algorithm either keeps  $v$  unmatched or matches it irrevocably to one of the offline nodes  $r$  with non-zero remaining capacity. The objective is to maximize the performance ratio, which is the expected sum of weights of edges matched by the algorithm divided by that of the maximum weight  $K$ -matching in the graph fixed by the adversary.

Kesselheim et al. [12] essentially showed that the  $(1,1)$ -threshold algorithm can be applied to the edge-weighted online 1-matching problem to get a performance ratio of  $\frac{1}{e}$ . Independent of our work, in a recent paper Kesselheim et al. [13] considered a general class of online packing LPs. Their results imply that there is an alternative rounding-based algorithm that achieves performance ratio  $1 - O(\frac{1}{\sqrt{K}})$  for online  $K$ -matching when  $K$  is large. For small  $K$ , their results imply that there is an algorithm using one threshold that achieves performance ratio 8.1, which is not optimal.

**General Transformation.** We show that any  $(K,K)$ -

secretary algorithm satisfying certain *monotone* property (Definition 2.2) can be transformed into an algorithm for the edge-weighted online bipartite  $K$ -matching problem with performance ratio at least the same as before. In particular, threshold algorithms satisfy the monotone property.

**THEOREM 1.5. (APPLYING SECRETARY TO MATCHING)** *Any monotone  $(K,K)$ -secretary algorithm can be transformed into an algorithm for edge-weighted online bipartite  $K$ -matching with at least the same performance guarantee. In particular, the optimal  $(K,K)$ -threshold algorithm induces a deterministic algorithm with performance ratio at least  $\rho_{K,K} \geq 1 - O(\frac{1}{\sqrt{K}})$ .*

**Our Techniques.** We use a similar algorithm as in [12], in which at every step, a  $K$ -matching is formed between all arrived online nodes and offline nodes. However, we need a sophisticated objective function to form this  $K$ -matching in hindsight with the expected payoff of  $(K,K)$ -secretary problem. Next, we also consider a backward randomized process, and use a coupling argument to show that the probability that a certain offline node is available according to certain rule is at least the probability that a quota can be used according to the corresponding rule in secretary problem. Our transformation can even handle the case when offline nodes have *non-uniform capacities*; the performance ratio is limited by the minimum capacity. The details are in Section 5.

## 2 Preliminaries

**2.1 Monotonicity of Threshold Algorithms** We discuss some properties of the threshold algorithms defined in Section 1.1. Since we work with both discrete steps in  $[n]$  and continuous time in  $[0, 1]$ , we use  $\mathcal{I}$  to denote either  $[n]$  or  $[0, 1]$ , and an instant  $i \in \mathcal{I}$  can either mean a step  $i$  or a time  $t$ .

Observe that in the  $(J,K)$ -secretary problem, it does not make sense for an algorithm to select an item that is not a  $K_{\geq}$ -potential. We use  $[K^*] := \{1, \dots, K\} \cup \{\perp\}$  to denote the potential of an arriving item, where  $\perp$  means the item is not within the best  $K$  items arrived so far and should definitely be discarded.

**DEFINITION 2.1. (ECONOMICAL)** *An algorithm  $\mathcal{A}$  is economical if whether to select an item depends only on (1) at which instant  $i$  it arrives, (2) the number  $j$  of remaining quotas, (3) the item's potential  $k$ . Formally, algorithm  $\mathcal{A}$  can be described by an ensemble of functions  $\{\mathcal{A}^i : (\{0\} \cup [J]) \times [K^*] \rightarrow \{0, 1\} | i \in \mathcal{I}\}$ , where  $\mathcal{A}^i(j, k) = 1$  means that a  $k$ -potential arriving at instant  $i$  when there are  $j$  remaining quotas will be selected, and*

0 otherwise. We assume  $\mathcal{A}^\iota(0, k) = 0$  and  $\mathcal{A}^\iota(j, \perp) = 0$ , for all  $\iota \in \mathcal{I}$ ,  $j \in [J]$  and  $k \in [K]$ .

**DEFINITION 2.2. (MONOTONE)** An algorithm  $\mathcal{A}$  is monotone if it is economical and the following happens for each  $\iota \in \mathcal{I}$ .

- (1) For all  $j \in [J]$ ,  $k' \leq k \in [K]$ ,  $\mathcal{A}^\iota(j, k') \geq \mathcal{A}^\iota(j, k)$ , i.e., an item with better merit is more likely to be selected.
- (2) For all  $k \in [K]$ ,  $j \leq j' \in [J]$ ,  $\mathcal{A}^\iota(j', k) \geq \mathcal{A}^\iota(j, k)$ , i.e., a selection is more likely if there are more remaining quotas.

**PROPOSITION 2.1.** Any threshold algorithm is monotone.

**THEOREM 2.1.** Kleinberg's algorithm [14] can be implemented as an economical  $(K, K)$ -secretary algorithm, but is not monotone.

*Proof.* We first paraphrase Kleinberg's algorithm [14] for the model in which each item arrives with a random time in  $[0, 1]$ . For simplicity, we assume that  $K$  is a power of 2. For  $K = 1$ , simply run the  $\frac{1}{e}$ -threshold algorithm. For simplicity, we assume that any quotas left unused are discarded in the recursive algorithm, and this does not affect the analysis of the performance ratio. For  $K \geq 2$ , in the interval  $[0, \frac{1}{2}]$ , recursively run the  $(\frac{K}{2}, \frac{K}{2})$ -algorithm (rescaling  $[0, \frac{1}{2}]$  to  $[0, 1]$ ) to select at most  $\frac{K}{2}$  items. Suppose  $\hat{y}$  is the  $\frac{K}{2}$ -th best item among items arriving in  $[0, \frac{1}{2}]$ . Then, after time  $\frac{1}{2}$ , an item better than  $\hat{y}$  is taken until at most  $\frac{K}{2}$  items are taken in the interval  $[\frac{1}{2}, 1]$ .

We prove that the algorithm is economical by induction. For  $K = 1$ , the  $\frac{1}{e}$ -threshold algorithm is economical. For the inductive case  $K \geq 2$ , consider the  $(K, K)$ -algorithm. In the interval  $[0, \frac{1}{2}]$ , the  $(\frac{K}{2}, \frac{K}{2})$ -algorithm is economical by the induction hypothesis. Recall that we assume that if less than  $\frac{K}{2}$  quotas are used in  $[0, \frac{1}{2}]$ , then the extra quotas are discarded at time  $\frac{1}{2}$ . After time  $\frac{1}{2}$ , observe that if there are exactly  $\frac{K}{2} - j$  quotas left, then exactly  $j$  items arriving after time  $\frac{1}{2}$  have been selected that is better than  $\hat{y}$ , which means that an item arriving at this point is better than  $\hat{y}$  iff it is a  $(\frac{K}{2} + j)_{\geq}$ -potential. Hence, the algorithm is economical.

However, observe that the algorithm is not monotone. Suppose in the above analysis at some time after  $\frac{1}{2}$ , there are  $\frac{K}{2} - 1$  remaining quotas, which means that a  $(\frac{K}{2} + 1)$ -potential is better than  $\hat{y}$  and will be selected. However, if there is one more quota left, i.e., there are  $\frac{K}{2}$  remaining quotas, then only  $K_{\geq}$ -potentials will be selected. In particular, a  $(\frac{K}{2} + 1)$ -potential will not be selected. ■

## 2.2 Analyzing Monotone Algorithms with $\text{LP}_n(J, K)$

When we later transform a monotone secretary algorithm  $\mathcal{A}$  to an online bipartite matching algorithm, it will be convenient to consider the finite step model. We rewrite the linear program characterization  $\text{LP}_n(J, K)$  given in [4] as follows.

$$\text{LP}_n(J, K)$$

$$\begin{aligned} \max \quad & \sum_{j=1}^J \sum_{k=1}^K \sum_{i=1}^n \frac{1}{n} \sum_{\ell=k}^K \frac{\binom{n-i}{\ell-k} \binom{i-1}{k-1}}{\binom{n-1}{\ell-1}} z_{j|k}(i) \\ \text{s.t. } & z_{j|k}(i) \leq \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K [z_{(j+1)|\ell}(m) - z_{j|\ell}(m)], \\ & z_{J|k}(i) \leq 1 - \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K z_{J|\ell}(m), \\ & z_{j|k}(i) \geq 0, \quad \forall i \in [n], k \in [K], j \in [J]. \end{aligned}$$

**Interpretation.** Consider any  $(J, K)$ -secretary algorithm  $\mathcal{A}$  (which could be randomized). For each  $i \in [n]$ ,  $j \in [J]$  and  $k \in [K]$ , let  $z_{j|k}(i)$  be the probability that the item arriving at step  $i$  is selected using quota  $Q_j$  given that it is a  $k$ -potential, where the randomness comes from the random order and the algorithm if any. We implicitly require that  $z_{j|k}(i) = 0$  when  $i < k$ . We note that our analysis will go through with these constraints. Throughout the analysis, we will change the order of summation in the objective function for different interpretation. The following lemmas were implied by results from [4].

**LEMMA 2.1.** The probability that there are exactly  $j$  remaining quotas at the beginning of step  $i$  is

- $\sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K [z_{(j+1)|\ell}(m) - z_{j|\ell}(m)]$  for  $1 \leq j < J$ , and
- $1 - \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K z_{J|\ell}(m)$  for  $j = J$ .

*Proof.* Let  $Z_i^j$  be the event that quota  $Q_j$  has been used before step  $i$ . For each  $1 \leq m \leq i-1$ , the probability that a selection is made at step  $m$  using  $Q_j$  is  $\sum_{\ell=1}^K \frac{1}{m} z_{j|\ell}(m)$ , where  $\frac{1}{m}$  is the probability that an  $\ell$ -potential arrives at step  $m$ . Then we have

$$\Pr[Z_i^j] = \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K z_{j|\ell}(m).$$

Observe that the event of having exactly  $j$  remaining quotas at the beginning of step  $i$  is the event  $Z_i^{j+1} \wedge \overline{Z_i^j}$ , where  $Z_i^{j+1}$  is an always true event. Therefore, the probability that there are exactly  $j$  remaining quotas at the beginning of step  $i$  is

$$\begin{aligned} \Pr[Z_i^{j+1} \wedge \overline{Z_i^j}] &= \Pr[Z_i^{j+1}] - \Pr[Z_i^j] \\ &= \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K [z_{(j+1)|\ell}(m) - z_{j|\ell}(m)] \end{aligned}$$

when  $1 \leq j < J$ , and

$$\Pr[\overline{Z_i^J}] = 1 - \Pr[Z_i^J] = 1 - \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K z_{J|\ell}(m)$$

when  $j = J$ . ■

**LEMMA 2.2. (INTERPRETATIONS OF LP<sub>n</sub>)** *Given an algorithm  $\mathcal{A}$ , the corresponding  $z$  values are feasible in LP<sub>n</sub>(J, K). Moreover, the quantities in the objective function has the following interpretations.*

- $\delta_{k|l}(i) := \frac{\binom{n-i}{\ell-k} \binom{i-1}{k-1}}{\binom{n-1}{\ell-1}}$  is the conditional probability that the item arriving at step  $i$  is a  $k$ -potential, given that it is the  $\ell$ -th best item overall.
- $\gamma_k(i) := \sum_{j=1}^J z_{j|k}(i)$  is the conditional probability that an item is selected at step  $i$ , given that it is a  $k$ -potential.
- $\beta_\ell(i) := \sum_{k=1}^\ell \delta_{k|l}(i) \cdot \gamma_k(i)$  is the conditional probability that the  $\ell$ -th best item overall is selected, given that it arrives at step  $i$ .
- The expected payoff is given by the objective function, which can be rewritten as  $\sum_{\ell=1}^K \frac{1}{n} \sum_{i=1}^n \beta_\ell(i)$ .

*Proof.* Given that a  $k$ -potential arrives at step  $i$ , a selection is made using quota  $Q_j$  implies that there are exactly  $j$  remaining quotas at the beginning of step  $i$ . The fact that a  $k$ -potential arrives at step  $i$  does not affect the relative ordering of items arrived before step  $i$ , and hence does not affect the decisions of  $\mathcal{A}$  before step  $i$ . The feasibility of  $z$  immediately follows from Lemma 2.1.

Now we consider the interpretations of quantities in the objective function. Let  $a_1, \dots, a_K$  be the  $K$  best items overall in order, where  $a_1$  is the best item.

- Given that  $a_\ell$  arrives at step  $i$ , for each  $1 \leq k \leq \ell$ , the item  $a_\ell$  is a  $k$ -potential if and only if exactly  $k-1$  items of  $\{a_1, \dots, a_{\ell-1}\}$  arrive before step  $i$ , which happens with probability  $\delta_{k|l}(i) = \frac{\binom{n-i}{\ell-k} \binom{i-1}{k-1}}{\binom{n-1}{\ell-1}}$ .
- Given that an item at step  $i$  is a  $k$ -potential, the  $J$  events that it is selected using quota  $Q_j$  for  $j \in [J]$  are disjoint; hence the item is selected with probability  $\gamma_k(i) = \sum_{j=1}^J z_{j|k}(i)$ .
- Given that  $a_\ell$  arrives at step  $i$ , for each  $1 \leq k \leq \ell$ , the probability that  $a_\ell$  is a  $k$ -potential is  $\delta_{k|l}(i)$ ; hence the conditional probability that  $a_\ell$  is selected is  $\beta_\ell(i) = \sum_{k=1}^\ell \delta_{k|l}(i) \cdot \gamma_k(i)$ .
- For  $\ell \in [K]$  and  $i \in [n]$ , item  $a_\ell$  arrives at step  $i$  with probability  $\frac{1}{n}$ . Then,  $a_\ell$  is selected by  $\mathcal{A}$  with probability  $\frac{1}{n} \sum_{i=1}^n \beta_\ell(i)$ . It follows that the expected payoff of  $\mathcal{A}$  is  $\sum_{\ell=1}^K \frac{1}{n} \sum_{i=1}^n \beta_\ell(i)$ , from which we can obtain the objective function of LP<sub>n</sub>(J, K) by substituting the functions  $\beta$ ,  $\delta$  and  $\gamma$ . ■

**LEMMA 2.3.** *Suppose  $\mathcal{A}$  is economical. Then, if  $\mathcal{A}^i(j, k) = 0$ , we have  $z_{j|k}(i) = 0$ ; if  $\mathcal{A}^i(j, k) = 1$ , then we have*

$$\begin{aligned} & z_{j|k}(i) \\ &= \begin{cases} \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K [z_{(j+1)|\ell}(m) - z_{j|\ell}(m)], & 1 \leq j < J \\ 1 - \sum_{m=1}^{i-1} \frac{1}{m} \sum_{\ell=1}^K z_{j|\ell}(m), & j = J. \end{cases} \end{aligned}$$

*Proof.* Since  $\mathcal{A}$  is economical, if  $\mathcal{A}^i(j, k) = 0$ , then clearly we have  $z_{j|k}(i) = 0$ . Now suppose  $\mathcal{A}^i(j, k) = 1$ . Then given that a  $k$ -potential arrives at step  $i$ , a selection is made using quota  $Q_j$  if and only if there are exactly  $j$  remaining quotas at the beginning of step  $i$ . Recall that a  $k$ -potential arriving at step  $i$  does not affect the decisions of  $\mathcal{A}$  before step  $i$ . Therefore,  $z_{j|k}(i)$  is the probability that there are exactly  $j$  remaining quotas at the beginning of step  $i$ . Hence, the equations follow from Lemma 2.1. ■

**LEMMA 2.4. (BETTER ITEMS MORE LIKELY SELECTED)** *Given a monotone algorithm  $\mathcal{A}$ , let  $\beta_\ell(i)$  be as defined in Lemma 2.2. For each  $i \in [n]$  and  $1 \leq k < l \leq K$ , we have  $\beta_k(i) \geq \beta_l(i)$ .*

*Proof.* Let  $\Omega_k$  be the set of permutations such that the  $k$ -th best item overall  $a_k$  arrives at step  $i$ , and define  $\Omega_l$  and  $a_l$  similarly. We define a bijection  $f : \Omega_\ell \rightarrow \Omega_k$  by swapping the positions of  $a_l$  and  $a_k$ . It suffices to prove that for all  $\omega \in \Omega_l$ , if  $a_l$  is selected in  $\omega$ , then  $a_k$  is selected in  $f(\omega)$ .

If  $a_l$  has potential  $q$  in  $\omega$ , then after swapping,  $a_k$  has potential  $p \leq q$  in  $f(\omega)$ , since  $k < l$ .

By monotonicity of  $\mathcal{A}$ , one can do a case analysis to see that if there are  $j$  remaining quotas when  $a_l$  arrives in  $\omega$ , then there are at least  $j$  remaining quotas when  $a_k$  arrives in  $f(\omega)$ . (The only tricky case is when  $a_k$  arrives at step  $i' < i$  in  $\omega$  and is selected, but after the swap,  $a_l$  is not selected in  $f(\omega)$ , which means there is one extra quota (compared to  $\omega$ ) at step  $i'$ . If this extra quota is not used up before step  $i$ , then there is one extra quota when  $a_k$  arrives; otherwise, when the extra quota is used up at some step before  $i$ , the algorithm behaves exactly like before again until step  $i$ .)

Hence, by the monotonicity of  $\mathcal{A}$  again,  $a_k$  must also be selected in  $f(\omega)$ . ■

### 3 Optimal Thresholds for (J, K)-Secretary Problem

We give a procedure to construct optimal asymptotic thresholds for (J, K)-secretary problem. Numerical values associated with the (2, 2)-case are computed in Section 3.4.1 and used in Section 4. Apart from these values, Sections 4 and 5 can be read independently.

**3.1 The Infinite Model** We use the infinite model as a tool to analyze the secretary problem when the number  $n$  of items is large. We shall describe the properties of our “infinite” algorithms, which can still be applied to finite instances to obtain conventional algorithms. We consider countably infinite number of items, whose ranks are indexed by the set  $\mathbb{N}$  of positive integers, where lower rank means better merit. Hence, the item with rank 1 is the best item. The arrival time of each item is a real number drawn independently and uniformly at random from  $[0, 1]$  (where the probability that two items arrive at the same time is 0); the (random) function  $\rho : \mathbb{N} \rightarrow [0, 1]$  gives the arrival time of each item, where  $\rho(i)$  is the arrival time of the item with rank  $i$ .

**Input Sample Space.** An algorithm can observe the arrival times  $\Sigma$  of items and their relative merit, which can be given by a total ordering  $\prec$  on  $\Sigma$ . Given  $\rho : \mathbb{N} \rightarrow [0, 1]$ , we have the set  $\Sigma_\rho := \{\rho(i) | i \in \mathbb{N}\}$ , and a total ordering  $\prec_\rho$  on  $\Sigma_\rho$  defined by  $\rho(i) \prec_\rho \rho(j)$  if and only if  $i < j$ . The sample space is  $\Omega := \{(\Sigma_\rho, \prec_\rho) | \rho : \mathbb{N} \rightarrow [0, 1]\}$ , with a probability distribution induced by the randomness of  $\rho$ ; we say each  $\omega = (\Sigma, \prec) \in \Omega$  is an *arrival sequence*. We sometimes use time  $t \in \Sigma$  to mean the item arriving at time  $t$ , for instance we might say “the algorithm selects  $t \in \Sigma$ .”

**FACT 3.1. (EVERY NON-ZERO INTERVAL CONTAINS INFINITE NUMBER OF ITEMS)** *For every interval  $\mathcal{I} \subseteq [0, 1]$  with non-zero length, the probability that there exist infinitely many items arriving in  $\mathcal{I}$  is 1.*

**Infinite Algorithm.** When an item arrives, an algorithm must decide immediately whether to select that item. Having an infinite number of items can make the notion of an infinite algorithm tricky. However, the infinite model is ultimately used as a tool to analyze threshold algorithms for the finite model. Hence, without loss of generality, we only consider monotone (and economical) algorithms in the sense defined in Definition 2.2. Observe that for any item arriving at time  $t$ , the number of items arriving before it and having better merits is finite, and hence the potential of every arriving item is well-defined. Therefore, an algorithm  $\mathcal{A}$  can still be considered as an ensemble of functions  $\{\mathcal{A}^t : [J] \times [K^*] \rightarrow \{0, 1\} | t \in [0, 1]\}$  as before.

**DEFINITION 3.1. (OUTCOME AND PAYOFF)** *Let  $\mathcal{A}$  be an (infinite) algorithm. For  $\omega \in \Omega$ , the outcome  $\mathcal{A}(\omega)$  is the number of items selected among the  $K$  best items. The expected payoff of  $\mathcal{A}$  is defined as  $P(\mathcal{A}) := \mathbf{E}_\omega [\mathcal{A}(\omega)]$ .*

The reason we consider the infinite model is that for any  $0 < t \leq 1$ , the sample space  $\Omega^{(t)}$  observed before

$t$  has the same structure as  $\Omega$  in the sense described in the following Proposition 3.1. This allows us to analyze the recursive behavior of any infinite algorithm.

**PROPOSITION 3.1. (ISOMORPHISM BETWEEN  $\Omega^{(t)}$  AND  $\Omega$ )** *For any  $0 < t \leq 1$ , the sample space  $\Omega^{(t)}$  (with distribution inherited from  $\Omega$ ) rescaled to  $[0, 1]$  (by dividing each arrival time by  $t$ ) has the same distribution as  $\Omega$ .*

*Proof.* Recall that the probability distribution over  $\Omega$  is induced by the randomness of all the infinite arrival times, each of which is a random number drawn independently and uniformly from  $[0, 1]$ . Similarly, the probability distribution over  $\Omega^{(t)}$  is induced by the randomness of arrival times before  $t$ , the number of which is infinite by Fact 3.1. Moreover, each arrival time in  $[0, t)$  is drawn independently and uniformly from  $[0, t)$ , which after rescaling is independently and uniformly distributed in  $[0, 1]$ . ■

Recall that for  $k \geq 1$ , an arriving item is a  $k$ -potential if it is the  $k$ -th best item among all those that have already arrived, and a  $k_>$ -potential if it is a  $k'$ -potential for some  $k' \leq k$ .

**PROPOSITION 3.2. (DISTRIBUTION OF POTENTIALS)** *For every  $k \geq 1$  and  $t > 0$ , with probability 1, the following conditions hold.*

1. *There exists a 1-potential in  $[0, t)$ .*
2. *There are finitely many  $k$ -potentials in  $[t, 1]$ .*

*Proof.* From Fact 3.1, with probability 1, there exists an item arriving in  $[0, t)$ . This implies that there exists  $i \in \mathbb{N}$  such that the item with rank  $i$  arrives at  $\rho(i) \in [0, t)$ . If  $\rho(i)$  is not a 1-potential, then a non-empty subset  $S$  of items with ranks in  $\{1, \dots, i-1\}$  must have arrived before  $\rho(i)$ . Since  $S$  is finite, the item among them with smallest arrival time is a 1-potential. Thus, there is a 1-potential in  $[0, t)$  with probability 1.

Similarly, from Fact 3.1, there exist  $k$  items in  $[0, t)$  and let  $r$  be the maximum rank among those  $k$ . Every  $k$ -potential in  $[t, 1]$  must have a rank in  $\{1, \dots, r-1\}$ ; that is, there are finite number of  $k$ -potentials in  $[t, 1]$  with probability 1. ■

**REMARK 3.1. (Validity of Threshold Algorithms)** *Consider a strange scenario when no items arrive at  $\tau_{j,1}$ , but, for every  $n$ , the item with rank  $n$  arrives at time  $\tau_{1,1} + \frac{1}{n}$ . It follows that every item is a 1-potential, but there is no “first” item arriving after  $\tau_{j,1}$ . One way to resolve this situation is that quota  $Q_j$  is simply “lost”. On the other hand, by Proposition 3.2, with probability 1, there are only a finite number of  $K_>$ -potentials arriving after  $\tau_{j,1}$ , and hence this strange scenario we describe happens with probability 0.*

**Piecewise Continuous.** Given an algorithm  $\mathcal{A}$ ,  $j \in [J]$  and  $k \in [K]$ , define the function  $p_{j|k}^{\mathcal{A}} : [0, 1] \rightarrow [0, 1]$  such that  $p_{j|k}^{\mathcal{A}}(x)$  is the probability that  $\mathcal{A}$  selects time  $x$  using quota  $Q_j$  given that  $x$  is a  $k$ -potential. Let  $p^{\mathcal{A}} = (p_{j|k}^{\mathcal{A}})_{j \in [J], k \in [K]}$  be the collection of functions for  $\mathcal{A}$ . We say  $\mathcal{A}$  is *piecewise continuous* if every  $p_{j|k}^{\mathcal{A}}$  is piecewise continuous. We denote by  $\mathcal{A}$  the class of piecewise continuous algorithms; as we shall see, this class of algorithms is general enough to capture the asymptotic behavior for finite models with large  $n$  number of items.

**PROPOSITION 3.3.** *Any threshold algorithm is piecewise continuous.*

*Proof.* The argument is straightforward but the full proof is tedious. As a special case, consider the first threshold  $\tau_{J,1}$  and the function  $p_{J|1}(x)$  giving the probability that a 1-potential arriving at time  $x$  is selected by using quota  $Q_J$ , which is 0 for  $x < \tau_{J,1}$ , and there is a discontinuity at  $\tau_{J,1}$ . At time  $x > \tau_{J,1}$ , the probability  $p_{J|1}(x)$  is the same as that for the event of the best item before time  $x$  arriving before  $\tau_{J,1}$ , and so  $p_{J|1}(x) = \frac{\tau_{J,1}}{x}$  for  $x > \tau_{J,1}$ . Other  $p_{j|k}$ 's can be analyzed similarly. ■

As discussed in the introduction, there is a linear programming  $\text{LP}_n$  for the finite step model such that the optimal value of the LP gives the optimal performance ratio, and every feasible solution gives a randomized algorithm with the same objective value. As we will see in Section 3.2, for the infinite model, any infinite algorithm in  $\mathcal{A}$ , in particular a threshold algorithm, is feasible for some maximization primal continuous  $\text{LP}_{\infty}$ . With slight abuse of notation, we let  $\text{LP}_n$  and  $\text{LP}_{\infty}$  denote the optimal values of the corresponding LPs.

**THEOREM 3.1. (FINITE OPTIMAL IS AT LEAST INFINITE OPTIMAL)** *Let  $\text{LP}_n$  and  $\text{LP}_{\infty}$  be optimal values of the LPs defined above. Then,  $\text{LP}_n \geq \text{LP}_{\infty}$ .*

*Proof.* Note that an infinite algorithm can be applied when the number  $n$  of items is finite, which can be further implemented as a randomized algorithm in the finite step model. As we will show in Section 3.4, there is a threshold algorithm that achieves payoff  $\text{LP}_{\infty}$ , which is optimal. It remains to prove that if a threshold algorithm has payoff  $\rho$  in the infinite model, then for any  $n$ , the payoff of the algorithm applied to the finite continuous model with  $n$  items is at least  $\rho$ .

We suppose a threshold algorithm  $\mathcal{A}$  has payoff  $\rho$  in the infinite model and fix  $n$ . For an item  $t \in [0, 1]$ , let  $\omega = (\Sigma, \prec) \in \Omega^{[t]}$  be the arrival sequence up to time  $t$ . Note that if item  $t$  has a rank larger than  $n$ ,

then  $\mathcal{A}^{[t]}(\omega) = 0$ , i.e., item  $t$  cannot be selected by  $\mathcal{A}$ . Suppose item  $t$  has a rank at most  $n$ . Let  $\omega_n = (\Sigma_n, \prec)$ , where  $\Sigma_n$  is obtained from  $\Sigma$  by removing all items with ranks larger than  $n$ . Then item  $t$  is a  $k$ -potential in  $\omega_n$  if and only if it is a  $k$ -potential in  $\omega$ . Then we have  $\mathcal{A}^{[t]}(\omega_n) = \mathcal{A}^{[t]}(\omega)$ . Therefore,  $\mathcal{A}^{[t]}(\omega) = 1$  implies  $\mathcal{A}^{[t]}(\omega_n) = 1$  for all  $t \in [0, 1]$ . It follows that the payoff of applying  $\mathcal{A}$  to the finite continuous model with  $n$  items is at least  $\rho$ . ■

**Main Approach.** Note that  $\text{LP}_n$  is the optimal payoff for the finite model with  $n$  items. In Section 3.2, we consider the continuous  $\text{LP}_{\infty}$ , which gives a connection to the (finite) LP. From Proposition 3.5 in Section 3.2, we can conclude that  $\limsup_{n \rightarrow \infty} \text{LP}_n \leq \text{LP}_{\infty}$ . In Sections 3.3 and 3.4, we show that the optimal payoff for the infinite algorithm can be achieved by a threshold algorithm, establishing Theorem 3.1. In particular, we show that the optimal payoff for the infinite model is  $\text{LP}_{\infty}$ . Hence, it follows that  $\lim_{n \rightarrow \infty} \text{LP}_n = \text{LP}_{\infty}$ , and there exist thresholds such that the threshold algorithm achieves the asymptotic optimal payoff for the finite model for large  $n$ , as stated in Theorem 1.1.

**3.2 The Continuous Linear Programming** For the finite model with random permutation, Buchbinder et al. [4] showed that there exists a linear programming  $\text{LP}_n(J, K)$  such that there is a one-to-one correspondence between an algorithm for the  $(J, K)$ -secretary problem with  $n$  items and a feasible solution of the LP; the payoff of the algorithm is exactly the objective of  $\text{LP}_n(J, K)$ . Therefore, the optimal value of the  $\text{LP}_n(J, K)$  gives the maximum payoff of the  $(J, K)$ -secretary problem with  $n$  items. We recall that the quotas are used in the order  $Q_J, Q_{J-1}, \dots, Q_1$ , and that the variable  $z_{j|k}(i)$  represents that the probability that the  $i$ -th item is selected using quota  $Q_j$  given that it is a  $k$ -potential.

For the  $(J, K)$ -secretary problem in the infinite model, we construct a continuous linear programming such that every piecewise continuous algorithm corresponds to a feasible solution, whose objective value is the payoff of the algorithm. Hence, the optimal LP gives an upper bound for the maximum payoff; we later show that there exist thresholds such that the threshold algorithm can achieve the optimal LP value.

For each  $j \in [J]$  and  $k \in [K]$ , let  $p_{j|k}(x)$  be a function of  $x$  that is piecewise continuous in  $[0, 1]$ . In the rest of this paper, we use “ $\forall x$ ” to denote “for almost all  $x$ ”, which means for all but a measure zero set. Define

$\text{LP}_\infty(J, K)$  as follows.

$$\begin{aligned} & \text{LP}_\infty(J, K) \\ \max \quad w(p) = & \sum_{j=1}^J \sum_{k=1}^K \int_0^1 \left( \sum_{\ell=k}^K \binom{\ell-1}{k-1} (1-x)^{\ell-k} \right) x^{k-1} p_{j|k}(x) dx \\ \text{s.t.} \quad p_{j|k}(x) \leq & \int_0^x \frac{1}{y} \sum_{\ell=1}^K [p_{(j+1)|\ell}(y) - p_{j|\ell}(y)] dy, \\ & \forall x \in [0, 1], k \in [K], 1 \leq j < J \\ p_{J|k}(x) \leq & 1 - \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{J|\ell}(y) dy, \quad \forall x \in [0, 1], k \in [K] \\ p_{j|k}(x) \geq 0, \quad & \forall x \in [0, 1], k \in [K], j \in [J]. \end{aligned}$$

Fix an algorithm  $\mathcal{A} \in \mathcal{A}$ . For each  $x$  and  $j$  and  $k$ , the events  $E_x^j$ ,  $Z_x^j$ ,  $V_x^k$  and  $W_x^k$  are defined as follows. Let  $E_x^j$  be the event that time  $x$  is selected using quota  $Q_j$ . Let  $Z_x^j$  be the event that quota  $Q_j$  has already been used before time  $x$ , i.e., all quotas  $Q_{j'}$  for  $j' \geq j$  have been used. Let  $V_x^k$  be the event that time  $x$  is a  $k$ -potential. Let  $W_x^k$  be the event that time  $x$  is the  $k$ -th best item overall. Note that  $Z_x^j$  implies  $Z_x^{j+1}$ , and  $Z_x^{j+1} \wedge \overline{Z_x^j}$  is the event that quota  $Q_j$  is the next quota available to be used at time  $x$ , for  $1 \leq j < J$ . Also observe that  $E_x^j$  implies  $Z_x^{j+1} \wedge \overline{Z_x^j}$ .

**LEMMA 3.1. (INDEPENDENCE BETWEEN POTENTIAL AND PAST HISTORY)** *For  $0 < x \leq 1$ , and positive integer  $k$ , the event  $V_x^k$  that  $x$  is a  $k$ -potential is independent of the arrival sequence observed before time  $x$ . In particular, this implies that for any  $K > 1$ , the event that  $x$  is a  $K_{\geq}$ -potential is also independent of the arrival sequence observed before time  $x$ .*

*Proof.* By Proposition 3.1 the arrival sequence observed before time  $x$  can be generated by sampling a random arrival time for each integer in  $\mathbb{N}$  independently and uniformly in  $[0, x]$ . We distinguish two cases: (1) without knowledge of  $x$ , this sequence is generated for all integers in  $\mathbb{N}$ ; (2) given that  $x$  is a  $k$ -potential for some  $k \in [K]$ , this sequence is generated for all integers in  $\mathbb{N} \setminus \{k\}$ . Since the total ordering on a sequence observed before  $x$  is inherited from  $\mathbb{N}$  and there is a bijection between  $\mathbb{N}$  and  $\mathbb{N} \setminus \{k\}$ , the sequences generated in the two cases have the same distribution. Hence, the event  $V_x^k$  is independent of  $\Omega^{(x)}$ . Since the  $V_x^k$ 's for  $k \in [K]$  are disjoint, the event that  $x$  is a  $K_{\geq}$ -potential and  $\Omega^{(x)}$  are independent. ■

**LEMMA 3.2.** *For all  $j \in [J]$  and  $x \in [0, 1]$ , we have  $\Pr(Z_x^j) = \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{j|\ell}(y) dy$ .*

*Proof.* For  $\ell \in [K]$ , let  $y_\ell$  be the arrival time of the  $\ell$ -th best item in  $[0, x]$ . Define  $Y := \max_{\ell \in [K]} \{y_\ell\}$ . Then for each  $y \in [0, x]$  we have  $\Pr(Y \leq y) = \frac{y^K}{x^K}$ . It follows that

the probability density function of  $Y$  is  $f(y) = \frac{Ky^{K-1}}{x^K}$ . Also note that given  $Y = y$ , we have  $\Pr(y_\ell = y) = \frac{1}{K}$  for all  $\ell \in [K]$ . It follows that  $\Pr(E_y^j | Y = y) = \sum_{\ell=1}^K \Pr(E_y^j | V_y^\ell) \Pr(y_\ell = y) = \frac{1}{K} \sum_{\ell=1}^K p_{j|\ell}(y)$ .

There is no  $K_{\geq}$ -potential in  $(y, x]$  and hence no item is selected. Thus  $Z_x^j$  happens if and only if either  $Z_y^j$  or  $E_y^j$  (i.e.  $Z_y^{j+1} \wedge \overline{Z_y^j} \wedge E_y^j$ ) happens. By using arguments similar to the proof of Proposition 3.1, we can show that, whether the event  $Y = y$  happens or not, the distribution of sample space  $\Omega^{(y)}$  of arrival time observed before time  $y$  remains the same; in particular, the events  $Z_y^j$  and  $Y = y$  are independent. Moreover the events  $Z_y^j$  and  $E_y^j$  are disjoint. By the law of total probability we have

$$\begin{aligned} \Pr(Z_x^j) &= \int_0^x \Pr(Z_x^j | Y = y) f(y) dy \\ &= \int_0^x [\Pr(Z_y^j | Y = y) + \Pr(E_y^j | Y = y)] \frac{Ky^{K-1}}{x^K} dy \\ &= \frac{K}{x^K} \int_0^x [\Pr(Z_y^j) + \frac{1}{K} \sum_{\ell=1}^K p_{j|\ell}(y)] y^{K-1} dy. \end{aligned}$$

Fix  $j$  and let  $g(x) := \Pr(Z_x^j)$  be a function with respect to  $x$ . Taking derivatives on both sides of  $x^K g(x) = K \int_0^x [g(y) + \frac{1}{K} \sum_{\ell=1}^K p_{j|\ell}(y)] y^{K-1} dy$  and using piecewise continuity, we have  $g'(x) = \frac{1}{x} \sum_{\ell=1}^K p_{j|\ell}(y) dy$  for almost all  $x$ . Then  $g(x) = \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{j|\ell}(y) dy + c$  for some constant  $c$ . By definition  $g(0) = 0$  and thus  $c = 0$ . Therefore we have  $\Pr(Z_x^j) = \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{j|\ell}(y) dy$ . ■

**PROPOSITION 3.4. (OPTIMAL PAYOFF AT MOST OPTIMAL  $\text{LP}_\infty(J, K)$ )** *Let  $\mathcal{A} \in \mathcal{A}$  be an algorithm for the  $(J, K)$ -secretary problem. Let  $p = (p_{j|k})_{j \in [J], k \in [K]}$  be the functions such that for  $j \in [J]$  and  $k \in [K]$  and  $x \in [0, 1]$ , the probability that time  $x$  is selected by  $\mathcal{A}$  using quota  $Q_j$  given that time  $x$  is a  $k$ -potential is  $p_{j|k}(x)$ . Then  $p$  is a feasible solution of  $\text{LP}_\infty(J, K)$ . Moreover, the payoff of  $\mathcal{A}$  is exactly the objective  $w(p) = \sum_{j=1}^J \sum_{k=1}^K \int_0^1 \left( \sum_{\ell=k}^K \binom{\ell-1}{k-1} (1-x)^{\ell-k} \right) x^{k-1} p_{j|k}(x) dx$ .*

*Proof.* We first show that the payoff  $P(\mathcal{A}) = w(p)$ . Consider the relation between  $\Pr(E_x^j | V_x^k)$  and  $\Pr(E_x^j | W_x^k)$ . If time  $x$  is the  $\ell$ -th best item overall, then it must be a  $k$ -potential for some  $k \leq \ell$ . Moreover, we have  $\Pr(V_x^k | W_x^\ell) = \binom{\ell-1}{k-1} x^{k-1} (1-x)^{\ell-k}$  (by convention  $0^0 = 1$ ). Then

$$\begin{aligned} \Pr(E_x^j | W_x^\ell) &= \sum_{k=1}^{\ell} \Pr(E_x^j | V_x^k \wedge W_x^\ell) \Pr(V_x^k | W_x^\ell) \\ &= \sum_{k=1}^{\ell} \Pr(E_x^j | V_x^k) \Pr(V_x^k | W_x^\ell) \\ &= \sum_{k=1}^{\ell} \binom{\ell-1}{k-1} x^{k-1} (1-x)^{\ell-k} p_{j|k}(x). \end{aligned}$$

Let  $\mathbb{I}_\ell$  be the indicator that the  $\ell$ -th best item is selected. Since the probability density function of

each arrival time is uniform in  $[0, 1]$ , the payoff of the algorithm is

$$\begin{aligned}
P(\mathcal{A})\mathbf{E}\left[\sum_{\ell=1}^K \mathbb{I}_\ell\right] &= \sum_{\ell=1}^K \mathbf{E}[\mathbb{I}_\ell] \\
&= \sum_{\ell=1}^K \Pr(\ell\text{-th best item selected}) \\
&= \sum_{\ell=1}^K \sum_{j=1}^J \int_0^1 1 \cdot \Pr(E_x^j | W_x^\ell) dx \\
&= \sum_{j=1}^J \int_0^1 \sum_{\ell=1}^K \sum_{k=1}^{\ell} \binom{\ell-1}{k-1} x^{k-1} (1-x)^{\ell-k} p_{j|k}(x) dx \\
&= w(p).
\end{aligned}$$

For the constraints, by Lemma 3.2 we have  $p_{J|k}(x) = \Pr(E_x^J | V_x^k) \leq \Pr(\overline{Z_x^J}) = 1 - \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{J|\ell}(y) dy$ , and  $p_{j|k}(x) = \Pr(E_x^j | V_x^k) \leq \Pr(Z_x^{j+1} \wedge Z_x^j) = \Pr(Z_x^{j+1}) - \Pr(Z_x^j) = \int_0^x \frac{1}{y} \sum_{\ell=1}^K [p_{(j+1)|\ell}(y) - p_{j|\ell}(y)] dy$  for  $1 \leq j < J$ , where the second last equality follows since  $Z_x^j$  implies  $Z_x^{j+1}$ . ■

The following proposition relates the optimal values of  $\text{LP}_n(J, K)$  and  $\text{LP}_\infty(J, K)$ . The proof is omitted here and can be found in Proposition 3.2 of [5].

**PROPOSITION 3.5. (RELATION BETWEEN  $\text{LP}_n(J, K)$  AND  $\text{LP}_\infty(J, K)$ )** Let  $\text{LP}_n$  and  $\text{LP}_\infty$  be the optimal values of  $\text{LP}_n(J, K)$  and  $\text{LP}_\infty(J, K)$ , respectively. Then, for every  $\epsilon > 0$ , there exists  $N \in \mathbb{N}$  such that  $\text{LP}_\infty \geq \text{LP}_n - \epsilon$  for all  $n \geq N$ .

**3.3 A Primal-Dual Method for Finding Thresholds** We give a primal-dual procedure that finds appropriate thresholds for which the corresponding  $(J, K)$ -threshold algorithm corresponds to an optimal solution in the continuous linear program  $\text{LP}_\infty(J, K)$ . To illustrate our primal-dual method, we first consider the special case  $K = 1$  as described in Theorem 1.2; the general case is given in Section 3.4.

A  $J$ -threshold algorithm is a special case with  $t_j := \tau_{j,1}$ , and recall that any algorithm in the class  $\mathcal{A}$  corresponds to a feasible solution in the following primal continuous LP:

$$\begin{aligned}
&\text{LP}_\infty(J) \\
\max \quad &w(p) = \sum_{j=1}^J \int_0^1 p_j(x) dx \\
\text{s.t.} \quad &p_j(x) \leq \int_0^x \frac{1}{y} [p_{j+1}(y) - p_j(y)] dy, \\
&\quad \forall x \in [0, 1], 1 \leq j < J \\
&p_J(x) \leq 1 - \int_0^x \frac{p_J(y)}{y} dy, \quad \forall x \in [0, 1] \\
&p_j(x) \geq 0, \quad \forall x \in [0, 1], j \in [J].
\end{aligned}$$

The dual LP for  $\text{LP}_\infty(J)$  is as follows (see [17] for details on primal-dual continuous LP):

$$\begin{aligned}
&\text{D}_\infty(J) \\
\min \quad &\int_0^1 q_J(x) dx \\
\text{s.t.} \quad &q_1(x) + \frac{1}{x} \int_x^1 q_1(y) dy \geq 1, \quad \forall x \in [0, 1] \\
&q_j(x) + \frac{1}{x} \int_x^1 [q_j(y) - q_{j-1}(y)] dy \geq 1, \\
&\quad \forall x \in [0, 1], 1 < j \leq J \\
&q_j(x) \geq 0, \quad \forall x \in [0, 1], j \in [J].
\end{aligned}$$

**Weak Duality.** Similar to normal LP, for any feasible primal  $p$  and dual  $q$ , the value of the primal objective is at most that of the dual objective. Moreover, if their objective values are equal, then both are optimal. We also have the following complementary slackness conditions.

**FACT 3.2. (COMPLEMENTARY SLACKNESS CONDITIONS)** Let  $p = (p_1, \dots, p_J)$  and  $q = (q_1, \dots, q_J)$  be feasible solutions of  $\text{LP}_\infty(J)$  and  $\text{D}_\infty(J)$ , respectively. Then,  $p$  and  $q$  are primal and dual optimal, respectively, if they satisfy the following conditions  $\forall x \in [0, 1]$ :

$$\begin{aligned}
&(p_J(x) + \int_0^x \frac{1}{y} p_J(y) dy - 1) q_J(x) = 0 \\
&(p_j(x) + \int_0^x \frac{1}{y} [p_j(y) - p_{j+1}(y)] dy) q_j(x) = 0, \quad 1 \leq j < J \\
&(q_1(x) + \frac{1}{x} \int_x^1 q_1(y) dy - 1) p_1(x) = 0 \\
&(q_j(x) + \frac{1}{x} \int_x^1 [q_j(y) - q_{j-1}(y)] dy - 1) p_j(x) = 0, \\
&\quad 1 < j \leq J.
\end{aligned}$$

**Primal-Dual Method.** We start from a primal feasible solution  $p$  corresponding to a  $J$ -threshold algorithm, whose thresholds are to be determined. We can determine the values of the thresholds one by one in order to construct a dual  $q$  such that complementary slackness conditions hold, which implies that with those found thresholds the corresponding  $J$ -threshold algorithm is optimal.

**(1) Forming Feasible Primal Solution  $p$ .** Suppose  $p$  is a (feasible) primal corresponding to a  $J$ -threshold algorithm with thresholds  $0 < t_J \leq t_{J-1} \leq \dots \leq t_1 \leq 1$ . We denote  $E_x^j$  to be the event that item at  $x$  is selected by using quota  $Q_j$  (where quotas with larger  $j$ 's are used first),  $V_x$  to be the event that  $x$  is a 1-potential, and  $Z_x^j$  to be the event that at time  $x$ , quota  $Q_j$  has already been used (and so have the quotas with indices larger than  $j$ ). For notational convenience,  $Z_x^{J+1}$  is the whole sample space, i.e., an always true event.

For each  $j \in [J]$ , consider the conditional probability  $\Pr(E_x^j | Z_x^{j+1} \wedge Z_x^j \wedge V_x)$  of the event that item at  $x$  is selected by using quota  $Q_j$ , given that  $x$  is a 1-potential and quota  $Q_j$  is the next available quota at time  $x$ . By definition of threshold algorithms, this conditional probability is 0 if  $x < t_j$  and is 1 if  $x \geq t_j$ . Hence, we have the following.

$$\begin{aligned} \Pr(E_x^j | Z_x^{j+1} \wedge \overline{Z_x^j} \wedge V_x) &= \frac{\Pr(E_x^j | V_x)}{\Pr(Z_x^{j+1} \wedge \overline{Z_x^j} | V_x)} \\ &= \frac{p_j(x)}{\Pr(Z_x^{j+1} \wedge \overline{Z_x^j} | V_x)} = \begin{cases} 0, & 0 \leq x < t_j \\ 1, & t_j \leq x \leq 1, \end{cases} \end{aligned}$$

where from independence of  $V_x$  and  $Z_x^j$ , and Lemma 3.2, we have:

$$\begin{aligned} \Pr(Z_x^{j+1} \wedge \overline{Z_x^j} | V_x) &= \Pr(Z_x^{j+1} \wedge \overline{Z_x^j}) \\ &= \begin{cases} \int_0^x \frac{1}{y} [p_{j+1}(y) - p_j(y)] dy, & 1 \leq j < J \\ 1 - \int_0^x \frac{1}{y} p_j(y) dy, & j = J. \end{cases} \end{aligned}$$

This implies that in the primal  $\text{LP}_\infty(J)$ , the  $j$ -th constraint is equality in the range  $[t_j, 1]$ , but might be strict inequality in the range  $[0, t_j]$  (hence forcing  $q_j$  to 0); the function  $p_j$  is zero in the range  $[0, t_j]$ , but might be strictly positive in the range  $[t_j, 1]$  (hence forcing equality for the  $j$ -th constraint in dual).

**(2) Finding Feasible Dual  $q$  to Satisfy Complementary Slackness.** To ensure that a dual solution  $q$  satisfies complementary slackness together with the above primal  $p$ , we require the following for each  $j \in [J]$ , where for notational convenience we write  $q_0 \equiv 0$ .

(3.1)

$$\begin{cases} q_j(x) = 0, & x \in [0, t_j] \\ q_j(x) + \frac{1}{x} \int_x^1 [q_j(y) - q_{j-1}(y)] dy = 1, & x \in [t_j, 1]. \end{cases}$$

The astute reader might notice that we have imposed an extra condition  $q_j(t_j) = 0$ . This will ensure that as long as (3.1) is satisfied by some non-negative  $q_j$ , the  $j$ -th constraint in  $\text{D}_\infty(J)$  is also automatically satisfied. For  $x \in [t_j, 1]$ , the constraint is clearly satisfied with equality; for  $x \in [0, t_j]$ , observing that both  $q_j$  and  $q_{j-1}$  vanishes below  $t_j$  the left hand side reduces to  $\frac{1}{x} \int_{t_j}^1 [q_j(y) - q_{j-1}(y)] dy$ , which is larger than  $q_j(t_j) + \frac{1}{t_j} \int_{t_j}^1 [q_j(y) - q_{j-1}(y)] dy = 1$ .

As we shall see soon, in the recursive equations (3.1), the function  $q_1$  and the threshold  $t_1$  does not depend on  $J$ . In particular, the thresholds  $t_j$ 's and functions  $q_j$ 's found for  $\text{D}_\infty(J)$  can be used to extend to the solution for  $\text{D}_\infty(J+1)$ . This explains the nice structure of the solution that appears in Theorem 1.2.

**Objective Value.** The objective value of  $\text{D}_\infty(J)$  is  $\int_0^1 q_J(y) dy = \int_{t_j}^1 q_J(y) dy$ . From the second equation of (3.1) evaluating at  $x = t_j$ , we have the recursive definition  $\int_{t_j}^1 q_j(y) dy = \int_{t_{j-1}}^1 q_{j-1}(y) dy + t_j$ , which immediately implies that the objective value for  $\text{D}_\infty(J)$  is  $\sum_{j=1}^J t_j$ , as stated in Theorem 1.2. Hence, it suffices to show the existence of dual functions as required in (3.1).

**LEMMA 3.3. (EXISTENCE OF FEASIBLE DUAL SATISFYING COMPLEMENTARY SLACKNESS)** *There is a pro-*

*cEDURE to generate an increasing sequence  $\{\theta_j\}_{j \geq 1}$  of rational numbers producing  $t_j := \frac{1}{e^{\theta_j}}$ , and a sequence  $\{q_j : [0, 1] \rightarrow \mathbb{R}^+\}_{j \geq 1}$  of non-negative functions that satisfy (3.1).*

*Proof.* We show the existence result by induction; our induction proof actually gives a method to generate such  $t_j$ 's and  $q_j$ 's. We explicitly describe the method in Section 3.3.1, and it can be seen that the time to generate the first  $J$  thresholds is  $O(J^3)$ .

For convenience, we denote  $q_0(x) \equiv 0$  and set  $\theta_0 := 0$  and  $t_0 := 1$ . Suppose for some  $j \geq 1$  we have constructed the function  $q_{j-1}$  which is continuous and can be positive only in  $[t_{j-1}, 1]$ . We next wish to find continuous function  $q_j$  and threshold  $t_j < t_{j-1}$  satisfying (3.1). If such  $q_j$  and  $t_j$  exist, then we must have the following for  $x \in [t_j, 1]$ :

$$\begin{aligned} q_j(x) + \frac{1}{x} \int_x^1 [q_j(y) - q_{j-1}(y)] dy &= 1 \\ xq_j(x) + \int_x^1 [q_j(y) - q_{j-1}(y)] dy &= x \\ (xq'_j(x) + q_j(x)) - q_j(x) + q_{j-1}(x) &= 1 \\ q'_j(x) &= \frac{1}{x}(1 - q_{j-1}(x)). \end{aligned}$$

Since  $q_j(1) = 1$ , we must have

$$(3.2) \quad q_j(x) = 1 + \ln x + \int_x^1 \frac{q_{j-1}(y)}{y} dy, \quad \forall x \in [t_j, 1]$$

To show that both  $q_j$  and  $t_j$  exist, we need a stronger induction hypothesis. Hence, we first explicitly solve for  $q_1$  and  $t_1$ , and state what properties we can assume. Since  $q_0(x) \equiv 0$ , we have  $q_1(x) = 1 + \ln x$  on  $[t_1, 1]$ . In order to have  $q_1(t_1) = 0$ , we must have  $t_1 := \frac{1}{e}$  and  $\theta_1 := 1$ . We give our induction hypothesis, which is true for  $j = 1$ .

**Induction Hypothesis.** Suppose for some  $j \geq 1$ , there exist functions  $\{q_i\}_{i=0}^j$  and thresholds  $\{t_i\}_{i=0}^j$  satisfying (3.1) such that the following holds.

1. The function  $q_j$  is non-negative and continuous.
2. There exists an increasing sequence  $\{\theta_i\}_{i=0}^j$  of rational numbers that defines the thresholds  $t_i := \exp(-\theta_i)$  such that  $q_j$  is 0 on  $[0, t_j]$  and between successive thresholds,  $q_j(x)$  is given by a polynomial in  $\ln x$  with rational coefficients.
3. For  $x \in (t_j, 1)$ ,  $q_j(x) > q_{j-1}(x)$ .

We next show the existence of  $q_{j+1}$  and  $t_{j+1}$ .

**Finding  $q_{j+1}$ .** From (3.2),  $q_{j+1}(x)$  must agree on  $[t_{j+1}, 1]$  with the function  $q(x)$  given by  $q(x) = 1 + \ln x + \int_x^1 \frac{q_j(y)}{y} dy$ , which is continuous.

We first check that we can set  $q_{j+1}(x) := q(x)$  for  $x \in [t_j, 1]$ . Since from the induction hypothesis we have  $q_j > q_{j-1}$  on  $(t_j, 1)$ , we immediately have  $\forall x \in [t_j, 1]$ ,  $q(x) = 1 + \ln x + \int_x^1 \frac{q_j(y)}{y} dy > 1 + \ln x + \int_x^1 \frac{q_{j-1}(y)}{y} dy =$

$q_j(x)$ . In particular, we have  $q(t_j) > q_j(t_j) = 0$ , and also  $q(x) \geq q_j(x) \geq 0$  for  $x \in [t_j, 1]$ .

From the induction hypothesis on  $q_j$ , we can conclude that between successive thresholds in  $[t_j, 1]$ ,  $q_{j+1}(x)$  can also be represented by a polynomial in  $\ln x$  with rational coefficients. Hence, it follows that  $d_j := \int_{t_j}^1 \frac{q_j(y)}{y} dy$  is rational.

**Finding  $t_{j+1}$ .** We next consider the behavior of  $q$  for  $x \leq t_j$ . Observe that in this range,  $q(x) = 1 + \ln x + d_j$ , which is a polynomial in  $\ln x$  with rational coefficients, and strictly increasing in  $x$ . Moreover, we have  $q(t_j) > 0$  and as  $x$  tends to 0,  $q(x)$  tends to negative infinity. Hence, there is a unique  $t_{j+1} \in (0, t_j)$  such that  $q(t_{j+1}) = 0$ ; we set  $t_{j+1} := \exp(-\theta_{j+1})$ , where  $\theta_{j+1} := 1 + d_j$ , which is rational.

Hence, we can set  $q_{j+1}(x) := q(x)$  for  $x \in [t_{j+1}, 1]$  and 0 for  $x \in [0, t_{j+1}]$ . We can check that the conditions in the induction hypothesis hold for  $q_{j+1}$  and  $t_{j+1}$  as well. This completes the induction proof. ■

**3.3.1 Explicit Methods for the  $(J, 1)$ -Case** For  $K = 1$  with thresholds denoted by  $t_j = \tau_{j,1}$  for  $j \in [J]$ , the proof of Lemma 3.3 gives a method to generate the dual variables  $q_j$ 's and thresholds  $t_j$ 's, which we describe below.

#### $J$ -THRESHOLDS GENERATOR

Set  $\theta_1 := 1$  and  $t_1 := e^{-\theta_1}$ . Let  $q_1$  be a function defined on  $[0, 1]$  such that  $q_1(x) = 0$  for  $x \in [0, t_1)$  and  $q_1(x) = 1 + \ln x$  for  $x \in [t_1, 1]$ .

For  $j = 1, 2, \dots, J - 1$

Set  $\theta_{j+1} := 1 + \int_{t_j}^1 \frac{q_j(y)}{y} dy$  and  $t_{j+1} := e^{-\theta_{j+1}} \in (0, t_j)$ .

Let  $q_{j+1}$  be a function defined on  $[0, 1]$  such that

$$q_{j+1}(x) = \begin{cases} 0, & x \in [0, t_{j+1}) \\ 1 + \ln x + \int_{t_j}^1 \frac{q_j(y)}{y} dy, & x \in [t_{j+1}, t_j) \\ 1 + \ln x + \int_x^1 \frac{q_j(y)}{y} dy, & x \in [t_j, 1]. \end{cases}$$

One can see that the  $q_j(x)$ 's are polynomials in  $\ln x$  with rational coefficients. Hence we can describe the algorithm by maintaining the rational coefficients of the polynomials. The modified method is described as follows, and it can be seen that generating the first  $J$  thresholds takes  $O(J^3)$  time.

**3.4 Primal-Dual Method for General  $(J, K)$ -case** In Section 3.3 we have shown that there exist thresholds such that the corresponding  $J$ -threshold algorithm is optimal. We apply our primal-dual method to the general  $(J, K)$ -case following a similar framework.

#### $\theta$ -GENERATOR

For integer  $n \geq 1$ , let  $0^n$  be the zero vector with  $n$  coordinates. Let  $\theta_0 := 0$ .

For positive integers  $j$  and  $k$ , let  $c_{j,k} \in \mathbb{R}^{J+1}$  be a vector (corresponding to  $q_j$  in interval  $[t_k, t_{k-1}]$  where  $t_0 := 1$ ). Denote by  $c_{j,k}(i)$  the  $i$ -th coordinate of  $c_{j,k}$  for  $i \in [J+1]$ .

Set  $c_{1,1} := (1, 1, 0^{J-1})$  and  $\theta_1 := 1$ .

For  $j = 1, 2, 3, \dots$

Let  $\alpha \in \mathbb{R}$  be an auxiliary variable with initial value  $\alpha := 0$ .

For  $k = 1, 2, \dots, j$

Let  $d \in \mathbb{R}^J$  be such that  $d(i) = -\frac{c_{j,k}(i)}{i}$  for each  $i \in [J]$ .

If  $k > 1$ , then set  $\alpha := \alpha +$

$$\sum_{i=1}^{J+1} \frac{c_{j,k-1}(i)}{i} [(-\theta_{k-2})^i - (-\theta_{k-1})^i].$$

Set  $c_{j+1,k} := (1, 1, 0^{J-1}) +$

$$(\sum_{i=1}^{J+1} \frac{c_{j,k}(i)}{i} (-\theta_{k-1})^i, d) + (\alpha, 0^J).$$

Set  $\alpha := \alpha + \sum_{i=1}^{J+1} \frac{c_{j,j}(i)}{i} [(-\theta_{j-1})^i - (-\theta_j)^i]$ .

Set  $c_{j+1,j+1} := (1, 1, 0^{J-1}) + (\alpha, 0^J)$ .

Set  $\theta_{j+1} := c_{j+1,j+1}(1)$ .

After proving Theorem 1.1, we apply the construction procedure to the  $(2, 2)$ -case as an illustration, and hence prove Theorem 1.3, in Section 3.4.1.

For  $k \in [K]$  and  $x \in [0, 1]$ , define  $\alpha_k(x) := \sum_{\ell=k}^K \binom{\ell-1}{k-1} (1-x)^{\ell-k} x^{k-1}$ . The dual continuous LP for  $\text{LP}_\infty(J, K)$  is as follows.

$$\begin{aligned} & \text{D}_\infty(J, K) \\ \min \quad & \sum_{k=1}^K \int_0^1 q_{J|k}(x) dx \\ \text{s.t.} \quad & q_{1|k}(x) + \frac{1}{x} \int_x^1 \sum_{\ell=1}^K q_{1|\ell}(y) dy \geq \alpha_k(x), \\ & \forall x \in [0, 1], k \in [K] \\ & q_{j|k}(x) + \frac{1}{x} \int_x^1 \sum_{\ell=1}^K [q_{(j)|\ell}(y) - q_{(j-1)|\ell}(y)] dy \geq \alpha_k(x) \\ & \forall x \in [0, 1], k \in [K], 1 < j \leq J \\ & q_{j|k}(x) \geq 0, \quad \forall x \in [0, 1], k \in [K], j \in [J]. \end{aligned}$$

For  $\text{LP}_\infty(J, K)$  and  $\text{D}_\infty(J, K)$ , we say a constraint is the  $(j, k)$ -th constraint if  $p_{j|k}$  or  $q_{j|k}$  is the concerned function in the constraint. For instance, the  $(j, k)$ -th constraint in the dual with  $1 < j \leq J$  is  $q_{j|k}(x) + \frac{1}{x} \int_x^1 \sum_{\ell=1}^K [q_{(j)|\ell}(y) - q_{(j-1)|\ell}(y)] dy \geq \alpha_k(x)$ ,  $\forall x \in [0, 1]$ . We still have weak duality and the following complementary slackness conditions.

**FACT 3.3. (COMPLEMENTARY SLACKNESS CONDITIONS)** Let  $p = (p_{j|k})_{j \in [J], k \in [K]}$  and  $q = (q_{j|k})_{j \in [J], k \in [K]}$  be feasible solutions of  $\text{LP}_\infty(J, K)$  and  $\text{D}_\infty(J, K)$ , respectively. Then  $p$  and  $q$  are primal and dual optimal, respectively, if they satisfy the following conditions

$\forall x \in [0, 1], k \in [K]$ :

$$\begin{aligned} (p_{j|k}(x) + \int_0^x \frac{1}{y} \sum_{\ell=1}^K [p_{j|\ell}(y) - p_{(j+1)|\ell}(y)] dy) q_{j|k}(x) &= 0 \\ 1 \leq j < J \\ (p_{J|k}(x) + \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{J|\ell}(y) dy - 1) q_{J|k}(x) &= 0 \\ (q_{j|k}(x) + \frac{1}{x} \int_x^1 \sum_{\ell=1}^K [q_{(j)|\ell}(y) - q_{(j-1)|\ell}(y)] dy - \alpha_k(x)) \\ \cdot p_{j|k}(x) &= 0, \quad 1 < j \leq J \\ (q_{1|k}(x) + \frac{1}{x} \int_x^1 \sum_{\ell=1}^K q_{1|\ell}(y) dy - \alpha^k(x)) p_{1|k}(x) &= 0. \end{aligned}$$

**Primal-Dual Method.** We start from a primal feasible solution  $p$  corresponding to a  $(J, K)$ -threshold algorithm, whose thresholds are to be determined. We can determine the values of the thresholds one by one in order to construct a dual  $q$  such that complementary slackness conditions hold, which implies that with those found thresholds the corresponding  $(J, K)$ -threshold algorithm is optimal.

**(1) Forming Feasible Primal Solution  $p$ .** Suppose  $p$  is a (feasible) primal corresponding to a  $(J, K)$ -threshold algorithm with  $JK$  thresholds  $\tau_{j,k}$  such that  $0 < \tau_{J,k} \leq \tau_{J-1,k} \leq \dots \leq \tau_{1,k} \leq 1$  for  $k \in [K]$  and  $0 < \tau_{j,1} \leq \tau_{j,2} \leq \dots \leq \tau_{j,K} \leq 1$  for  $j \in [J]$ . Suppose  $E_x^j$  is the event that the item at  $x$  is selected by using quota  $Q_j$  (where quotas with larger  $j$ 's are used first),  $V_x^k$  is the event that  $x$  is a  $k$ -potential, and  $Z_x^j$  is the event that at time  $x$ , quota  $Q_j$  has already been used (and so have the quotas with indices larger than  $j$ ). For notational convenience,  $Z_x^{J+1}$  is the whole sample space, i.e., an always true event.

For each  $j \in [J]$ , consider the conditional probability  $\Pr(E_x^j | Z_x^{j+1} \wedge Z_x^j \wedge V_x^k)$  of the event that item at  $x$  is selected by using quota  $Q_j$ , given that  $x$  is a  $k$ -potential and quota  $Q_j$  is the next available quota at time  $x$ . By definition of threshold algorithms, this conditional probability is 0 if  $x < \tau_{j,k}$  and is 1 if  $x \geq \tau_{j,k}$ . Hence, we have the following.

$$\begin{aligned} \Pr(E_x^j | Z_x^{j+1} \wedge \overline{Z_x^j} \wedge V_x^k) &= \frac{\Pr(E_x^j | V_x^k)}{\Pr(Z_x^{j+1} \wedge \overline{Z_x^j} | V_x^k)} \\ &= \frac{p_{j|k}(x)}{\Pr(Z_x^{j+1} \wedge \overline{Z_x^j} | V_x^k)} = \begin{cases} 0, & 0 \leq x < \tau_{j,k} \\ 1, & \tau_{j,k} \leq x \leq 1, \end{cases} \end{aligned}$$

where from independence of  $V_x^k$  and  $Z_x^j$  (Lemma 3.1), and Lemma 3.2, we have:

$$\begin{aligned} \Pr(Z_x^{j+1} \wedge \overline{Z_x^j} | V_x^k) \\ = \begin{cases} \int_0^x \frac{1}{y} \sum_{\ell=1}^K [p_{(j+1)|\ell}(y) - p_{j|\ell}(y)] dy, & 1 \leq j < J \\ 1 - \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{J|\ell}(y) dy, & j = J. \end{cases} \end{aligned}$$

This implies that in the primal  $\text{LP}_\infty(J, K)$ , the  $(j, k)$ -th constraint is equality in the range  $[\tau_{j,k}, 1]$ , but might be strict inequality in the range  $[0, \tau_{j,k})$  (hence

forcing  $q_{j|k}$  to be 0); the function  $p_{j|k}$  is zero in the range  $[0, \tau_{j,k})$ , but might be strictly positive in the range  $[\tau_{j,k}, 1]$  (hence forcing equality for the  $(j, k)$ -th constraint in dual).

**(2) Finding Feasible Dual  $q$  to Satisfy Complementary Slackness.** To ensure that a dual solution  $q$  satisfies complementary slackness together with the above primal  $p$ , we require the following for each  $j \in [J]$  and  $k \in [K]$ , where for notational convenience we write  $q_{0|k} \equiv 0$  for all  $k \in [K]$ .

$$(3.3) \quad \begin{cases} q_{j|k}(x) = 0, & x \in [0, \tau_{j,k}]; \\ q_{j|k}(x) + \frac{1}{x} \int_x^1 \sum_{\ell=1}^K [q_{(j)|\ell}(y) \\ - q_{(j-1)|\ell}(y)] dy = \alpha_k(x), & x \in [\tau_{j,k}, 1]. \end{cases}$$

Here the extra condition  $q_{j|k}(\tau_{j,k}) = 0$  ensures that as long as (3.3) is satisfied by some non-negative  $q_{j|k}$ , the  $(j, k)$ -th constraint in  $\text{D}_\infty(J)$  is also automatically satisfied. Proof for this indication is not straightforward and requires stronger conditions for the dual functions, which we provide along the way we prove Theorem 1.1. From the recursive equations (3.3), the thresholds  $\tau_{j,k}$ 's and functions  $q_{j|k}$ 's found for  $\text{D}_\infty(J, K)$  can be used to extend to the solution for  $\text{D}_\infty(J+1, K)$ .

**Objective Value.** The objective value of  $\text{D}_\infty(J, K)$  is  $\int_0^1 \sum_{k=1}^K q_{J|k}(x) dx = \int_{\tau_{J,1}}^1 \sum_{k=1}^K q_{J|k}(x) dx$ . From equations (3.3) we have for  $j \in [J]$ ,

$$\begin{aligned} \int_{\tau_{j,1}}^1 \sum_{k=1}^K q_{j|k}(x) dx - \int_{\tau_{j-1,1}}^1 \sum_{k=1}^K q_{(j-1)|k}(x) dx \\ = \tau_{j,1} \alpha_1(\tau_{j,1}), \end{aligned}$$

where  $\tau_{0,1} = 1$ . This together with  $\alpha_1(x) = \frac{1-(1-x)^K}{x}$  implies that the objective value of  $\text{D}_\infty(J, K)$  is

$$\begin{aligned} \int_{\tau_{J,1}}^1 \sum_{k=1}^K q_{J|k}(x) dx &= \sum_{j=1}^J \tau_{j,1} \alpha_1(\tau_{j,1}) \\ &= J - \sum_{j=1}^J (1 - \tau_{j,1})^K. \end{aligned}$$

To prove Theorem 1.1, it suffices to show the existence of dual functions as required in (3.3). The proof of the following lemma is omitted here and can be found in Lemmas 5.1, 5.2 and 5.3 of [5].

**LEMMA 3.4. (EXISTENCE OF FEASIBLE DUAL SATISFYING COMPLEMENTARY SLACKNESS)** *There is a procedure to find appropriate thresholds  $(\tau_{j,k})_{j \in [J], k \in [K]}$  and a collection  $(q_{j|k})_{j \in [J], k \in [K]}$  of non-negative functions that satisfy (3.3).*

**3.4.1 An Illustration: the  $(2, 2)$ -case** As an example, we solve the  $(2, 2)$ -secretary problem using the primal-dual method. In particular, we settle thresholds for an optimal  $(2, 2)$ -threshold algorithm. Then we calculate the primal variables and expected payoff that will be useful in Section 4.

**Finding Optimal Thresholds.** We follow the primal-dual method described above to find optimal thresholds  $\tau_{j,k}$  for  $j, k \in \{1, 2\}$  for the  $(2, 2)$ -case. In particular, we use the equations (3.3) to set the dual  $q_{j|k}$ 's. Let  $r_1(x) := q_{1|1}(x) + q_{1|2}(x)$  and  $r_2(x) := q_{2|1}(x) + q_{2|2}(x)$ . Observe that  $\alpha_1(x) = 2 - x$  and  $\alpha_2(x) = x$ .

**Finding  $\tau_{1,2}$ .** Note that in  $[\tau_{1,2}, 1]$ ,  $r_1(x)$  must agree with the function  $r(x)$  such that

$$r(x) + \frac{2}{x} \int_x^1 r(y) dy = 2.$$

Solving the equation we get  $r(x) = 4x - 2$ . Also note that  $q_{1,2}(x)$  in  $[\tau_{1,2}, 1]$  must agree with  $q(x)$  such that

$$q(x) + \frac{1}{x} \int_x^1 r(y) dy = x.$$

Solving the equation we get  $q(x) = 3x - 2$ . Setting  $q(x) = 0$  we get  $x = \frac{2}{3} \in (0, 1)$ . Hence  $\tau_{1,2} = \frac{2}{3}$ . Also, for  $x \in [\tau_{1,2}, 1]$ , we have

$$\int_x^1 r_1(y) dy = 2x - 2x^2.$$

In the following calculation, we will use the substitution  $\tau_{1,2} = \frac{2}{3}$  when necessary.

**Finding  $\tau_{1,1}$ .** Note that  $q_{1|2}(x) = 0$  in  $[0, \tau_{1,2}]$ , and  $q_{1|1}(x) = r_1(x)$  in  $[\tau_{1,1}, \tau_{1,2}]$  must agree with  $q(x)$  such that

$$q(x) + \frac{1}{x} \int_x^{\tau_{1,2}} q(y) dy + \frac{1}{x} \int_{\tau_{1,2}}^1 r_1(y) dy = 2 - x.$$

Solving the equation we get  $q(x) = 2 \ln \frac{x}{\tau_{1,2}} - 2x + 3\tau_{1,2}$ . Setting  $q(x) = 0$  we get  $x = -W(-\frac{2}{3e}) \approx 0.346982 \in (0, \tau_{1,2})$ . Hence  $\tau_{1,1} = -W(-\frac{2}{3e})$ . Also, for  $x \in [\tau_{1,1}, \tau_{1,2}]$ , we have

$$\int_x^{\tau_{1,2}} r_1(y) dy = x^2 - 2x \ln \frac{x}{\tau_{1,2}} - \frac{4}{9}$$

**Finding  $\tau_{2,2}$ .** For  $x \in [\tau_{1,2}, 1]$ , let  $q(x)$  and  $r(x)$  be functions such that

$$\begin{aligned} r(x) + \frac{2}{x} \int_x^1 r(y) dy - \frac{2}{x} \int_x^1 r_1(y) dy &= 2 \\ q(x) + \frac{1}{x} \int_x^1 r(y) dy - \frac{1}{x} \int_x^1 r_1(y) dy &= x. \end{aligned}$$

Solving the equations we get  $r(x) = 8x - 8x \ln x - 6$  and  $q(x) = 5x - 4x \ln x - 4$ . Since  $q(\tau_{1,2}) = -\frac{8}{3} \ln \frac{2}{3} - \frac{2}{3} \approx 0.414573 > 0$ , the function  $r_2(x)$  agrees with  $r(x)$  in  $[\tau_{1,2}, 1]$ . Also, we have

$$\int_{\tau_{1,2}}^1 r_2(y) dy = \frac{4}{3} + \frac{16}{9} \ln \tau_{1,2}.$$

Now consider the case  $x \in [\tau_{1,1}, \tau_{1,2}]$ . Let  $q(x)$  and  $r(x)$  be functions such that

$$\begin{aligned} r(x) + \frac{2}{x} \int_x^{\tau_{1,2}} r(y) dy + \frac{2}{x} \int_{\tau_{1,2}}^1 r_2(y) dy \\ - \frac{2}{x} \int_x^1 r_1(y) dy &= 2 \\ q(x) + \frac{1}{x} \int_x^{\tau_{1,2}} r(y) dy + \frac{1}{x} \int_{\tau_{1,2}}^1 r_2(y) dy \\ - \frac{1}{x} \int_x^1 r_1(y) dy &= x. \end{aligned}$$

Solving the equations we get  $r(x) = 4x \ln x + 4 \ln \frac{x}{\tau_{1,2}} - (10 + 12 \ln \tau_{1,2})x + 6$  and  $q(x) = 2x \ln x + 2 \ln \frac{x}{\tau_{1,2}} - (4 + 6 \ln \tau_{1,2})x + 2$ . Setting  $q(x) = 0$  we get  $x \approx 0.517297 \in [\tau_{1,1}, \tau_{1,2}]$ . Hence  $\tau_{2,2} \approx 0.517297$  is the solution of  $x \ln x + \ln x - (2 + 3 \ln \frac{2}{3})x + 1 - \ln \frac{2}{3} = 0$ . Then, in  $[\tau_{2,2}, \tau_{1,2}]$ , the functions  $r_2(x)$  agree with  $r(x)$ . Also, we have

$$\begin{aligned} \int_{\tau_{2,2}}^{\tau_{1,2}} r_2(y) dy &= -\frac{4}{3} - \frac{16}{9} \ln \tau_{1,2} + 6\tau_{2,2}^2 \ln \tau_{1,2} \\ - 2\tau_{2,2}^2 \ln \tau_{2,2} + 6\tau_{2,2}^2 - 4\tau_{2,2} \ln \frac{\tau_{2,2}}{\tau_{1,2}} - 2\tau_{2,2}. \end{aligned}$$

**Finding  $\tau_{2,1}$ .** Note that  $q_{2|2}(x) = 0$  in  $[0, \tau_{2,2}]$ . For  $x \in [\tau_{1,1}, \tau_{2,2}]$ , let  $q(x)$  be a function such that

$$\begin{aligned} q(x) + \frac{1}{x} \int_x^{\tau_{2,2}} q(y) dy + \frac{1}{x} \int_{\tau_{2,2}}^1 r_2(y) dy \\ - \frac{1}{x} \int_x^1 r_1(y) dy = 2 - x. \end{aligned}$$

Solving the equation we get  $q(x) = -(\ln \frac{x}{\tau_{1,2}})^2 + (\ln \frac{\tau_{2,2}}{\tau_{1,2}})^2 + 2 \ln \frac{\tau_{2,2}}{\tau_{1,2}} - 6\tau_{2,2} \ln \tau_{1,2} + 2\tau_{2,2} \ln \tau_{2,2} - 6\tau_{2,2} + 4$ . Since  $q(\tau_{1,1}) > 0$ , the function  $q_{2|1}(x) = r_2(x)$  agrees with  $q(x)$  in  $[\tau_{1,1}, \tau_{2,2}]$ . Also, we have

$$\begin{aligned} \int_{\tau_{1,1}}^{\tau_{2,2}} r_2(y) dy &= 4\tau_{2,2} \ln \frac{\tau_{2,2}}{\tau_{1,2}} - 6\tau_{2,2}^2 \ln \tau_{1,2} \\ + 2\tau_{2,2}^2 \ln \tau_{2,2} - 6\tau_{2,2}^2 + 2\tau_{2,2} - \tau_{1,1}(\ln \frac{\tau_{2,2}}{\tau_{1,2}})^2 \\ - 2\tau_{1,1} \ln \frac{\tau_{2,2}}{\tau_{1,2}} + 6\tau_{1,1} \tau_{2,2} \ln \tau_{1,2} - 2\tau_{1,1} \tau_{2,2} \ln \tau_{2,2} \\ + 6\tau_{1,1} \tau_{2,2} - 2\tau_{1,1} + \tau_{1,1}(\ln \frac{\tau_{1,1}}{\tau_{1,2}})^2 - 2\tau_{1,1} \ln \frac{\tau_{1,1}}{\tau_{1,2}}. \end{aligned}$$

Now consider the case  $x \in [0, \tau_{1,1}]$ , where  $r_1(x) = 0$ . Let  $q(x)$  be a function such that

$$\begin{aligned} q(x) + \frac{1}{x} \int_x^{\tau_{1,1}} q(y) dy + \frac{1}{x} \int_{\tau_{1,1}}^1 r_2(y) dy \\ - \frac{1}{x} \int_{\tau_{1,1}}^1 r_1(y) dy = 2 - x. \end{aligned}$$

Solving the equation we get  $q(x) = 2 \ln x - 2x + c$ , where  $c$  is a constant defined as

$$\begin{aligned} c &= -(\ln \frac{\tau_{1,1}}{\tau_{1,2}}) + (\ln \frac{\tau_{2,2}}{\tau_{1,2}})^2 + 2 \ln \frac{\tau_{2,2}}{\tau_{1,2}} - 6\tau_{2,2} \ln \tau_{1,2} \\ &\quad + 2\tau_{2,2} \ln \tau_{2,2} - 6\tau_{2,2} + 2\tau_{1,1} - 2 \ln \tau_{1,1} + 4 \\ &= -(\ln \tau_{1,1})^2 + 2 \ln \frac{2}{3} \ln \tau_{1,1} + (\ln \tau_{2,2})^2 \\ &\quad - 2 \ln \frac{2}{3} \ln \tau_{2,2} - 2\tau_{2,2} + 4 - 2 \ln \frac{2}{3}, \end{aligned}$$

where the last equation follows from the definitions of  $\tau_{1,2}$ ,  $\tau_{1,1}$  and  $\tau_{2,2}$ . Setting  $q(x) = 0$  we get  $x = -W(-e^{-c/2}) \approx 0.227788 \in (0, \tau_{1,1})$ . Hence  $\tau_{2,1} = -W(-e^{-c/2})$ . The function  $q_{2,1}(x)$  agrees with  $q(x)$  in  $[\tau_{2,1}, \tau_{1,1}]$ .

**Calculating Primal Variables and Expected Pay-off.** Given thresholds  $\tau_{j,k}$ 's, the primal variables can be calculated using the following recursion. For  $0 \leq x <$

$\tau_{j,k}$ ,  $p_{j|k}(x) = 0$ ; for  $\tau_{j,k} \leq x \leq 1$ ,

$$p_{j|k}(x) = \begin{cases} \int_0^x \frac{1}{y} \sum_{\ell=1}^K [p_{(j+1)|\ell}(y) - p_{j|\ell}(y)] dy, & 1 \leq j < J \\ 1 - \int_0^x \frac{1}{y} \sum_{\ell=1}^K p_{J|\ell}(y) dy, & j = J. \end{cases}$$

For the  $(2, 2)$ -case, setting  $J = K = 2$ , we obtain the following:

$$\begin{aligned} p_{2|1}(x) &= \begin{cases} \frac{\tau_{2,1}}{x}, & x \in [\tau_{2,1}, \tau_{1,1}] \\ \frac{\tau_{2,1}}{x}, & x \in [\tau_{1,1}, \tau_{2,2}] \\ \frac{\tau_{2,1}\tau_{2,2}}{x^2}, & x \in [\tau_{2,2}, \tau_{1,2}] \\ \frac{\tau_{2,1}\tau_{2,2}}{x^2}, & x \in [\tau_{1,2}, 1] \end{cases} \\ p_{1|1}(x) &= \begin{cases} \frac{\tau_{1,1}-\tau_{2,1}}{x} + \frac{\tau_{2,1}}{x} \ln \frac{x}{\tau_{1,1}}, & x \in [\tau_{1,1}, \tau_{2,2}] \\ \frac{\tau_{1,1}+\tau_{2,1}}{x} + \frac{\tau_{2,1}}{x} \ln \frac{\tau_{2,2}}{\tau_{1,1}} - \frac{2\tau_{2,1}\tau_{2,2}}{x^2}, & x \in [\tau_{2,2}, \tau_{1,2}] \\ \frac{2\tau_{2,1}\tau_{2,2}}{x^2} \ln \frac{x}{\tau_{1,2}} + \frac{\tau_{1,2}(\tau_{2,1}+\tau_{1,1})-2\tau_{2,1}\tau_{2,2}}{x^2} \\ + \frac{\tau_{2,1}\tau_{1,2}}{x^2} \ln \frac{\tau_{2,2}}{\tau_{1,1}}, & x \in [\tau_{1,2}, 1] \end{cases} \\ p_{2|2}(x) &= \begin{cases} \frac{\tau_{2,1}\tau_{2,2}}{x^2}, & x \in [\tau_{2,2}, \tau_{1,2}] \\ \frac{\tau_{2,1}\tau_{2,2}}{x^2}, & x \in [\tau_{1,2}, 1] \end{cases} \\ p_{1|2}(x) &= \frac{2\tau_{2,1}\tau_{2,2}}{x^2} \ln \frac{x}{\tau_{1,2}} + \frac{\tau_{1,2}(\tau_{2,1}+\tau_{1,1})-2\tau_{2,1}\tau_{2,2}}{x^2} + \frac{\tau_{2,1}\tau_{1,2}}{x^2} \ln \frac{\tau_{2,2}}{\tau_{1,1}}, \\ &x \in [\tau_{1,2}, 1] \end{aligned}$$

The objective function of  $\text{LP}_\infty$ , which is the expected payoff of the threshold algorithm, can be rewritten as  $\sum_{\ell=1}^K \int_0^1 \sum_{k=1}^\ell \left( \binom{\ell-1}{k-1} (1-x)^{\ell-k} \right) x^{k-1} \sum_{j=1}^J p_{j|k}(x) dx$ , where each integral for  $\ell \in [K]$  is the probability that the  $\ell$ -th best item overall is selected. For the  $(2, 2)$ -case, the probabilities that the best and second best item are selected are

$$\begin{aligned} p_1 &:= \int_0^1 (p_{2|1}(x) + p_{1|1}(x)) dx \\ p_2 &:= \\ &\int_0^1 [(1-x)(p_{2|1}(x) + p_{1|1}(x)) + x(p_{2|2}(x) + p_{1|2}(x))] dx, \end{aligned}$$

respectively. In particular, for  $p_2$ , we consider  $p_2^{(1)} := \int_0^1 (1-x)(p_{2|1}(x) + p_{1|1}(x)) dx$  and  $p_2^{(2)} := \int_0^1 x(p_{2|2}(x) + p_{1|2}(x)) dx$ , which are the probabilities that the second best item is selected as a 1-potential and a 2-potential, respectively. Substituting the functions  $p_{j|k}$ 's and the thresholds  $\tau_{j,k}$ 's calculated above, we obtain the following results:

$$\begin{aligned} p_1 &\approx 0.555711, & p_2^{(1)} &\approx 0.240104, \\ p_2^{(2)} &\approx 0.181441, & p_2 &\approx 0.421545. \end{aligned}$$

#### 4 Online 2-Item Auction with Random Arriving Bids: Decisions Based Only On Relative Merits Are Not Optimal

As mentioned in the introduction, the optimal threshold algorithm for the classical  $(1, 1)$ -secretary problem also achieves optimal ratio for the online 1-item auction with random arriving bids. We investigate if observing relative merits alone is enough to achieve optimal ratio in general for  $K$ -item auction. We focus on  $K = 2$ , and consider the continuous time model, where each item arrives at an independent time in  $[0, 1]$ .

**Ratio Depends on Relative Weights.** To see why considering relative weights can do better than observing just relative merits, consider the performance ratio of the optimal  $(2, 2)$ -threshold algorithm  $\mathcal{A}_0$ , which is  $\rho_{2,2} = \frac{p_1+p_2}{2}$ , where  $p_1$  and  $p_2$  are the probabilities that the best and the second best items, respectively, are selected. By monotonicity and Lemma 2.4, we have  $p_1 \geq p_2$ ; our calculation shows that the inequality is actually strict. Suppose the largest weight is  $W_1$  and the second largest is  $W_2$  (resolving ties consistently); we abuse notation and also use  $W_1$  and  $W_2$  to refer to the corresponding items. Hence, if  $\alpha = \frac{W_2}{W_1}$ , then the performance ratio (by just looking at relative merits) is at least  $\frac{p_1+p_2\alpha}{1+\alpha}$ , even if the weights of all other items are zero. Notice this bound is minimized and achieves  $\rho_{2,2}$  when  $\alpha = 1$ . This suggests that the adversary should pick the weights such that  $\alpha$  is close to 1. However, if indeed this is the worst case, then the algorithm could potentially filter out 2-potentials whose weight is not large enough. Before we analyze this cat-and-mouse game between an adversary and an algorithm, we first see how an algorithm might use relative weights to its advantage.

**$\alpha$ -Filtering Threshold Algorithm.** Given a parameter  $\alpha \in [0, 1]$ , the  $\alpha$ -filtering threshold algorithm  $\mathcal{A}_\alpha$  is a modification of the optimal threshold algorithm  $\mathcal{A}_0$ , where there is one extra condition: the algorithm selects an item only if its weight is at least  $\alpha$  times the largest weight seen so far. (Indeed, when  $\alpha = 0$ ,  $\mathcal{A}_\alpha$  is equivalent to  $\mathcal{A}_0$ .) This rule has no effect if the arriving item is a 1-potential, but it might filter out some 2-potentials.

**Analyzing Threshold Algorithm  $\mathcal{A}_0$ .** We first analyze the behavior of  $\mathcal{A}_0$  by rewriting the objective function of  $\text{LP}_\infty(2, 2)$ , which is the expected payoff of the  $(2, 2)$ -secretary problem and can be solved exactly by our techniques in Section 3.4.1:

$$\begin{aligned} &\sum_{j=1}^2 \sum_{k=1}^2 \int_0^1 \left( \sum_{\ell=k}^K \binom{\ell-1}{k-1} (1-x)^{\ell-k} \right) x^{k-1} p_{j|k}(x) dx \\ &= p_1 + p_2^{(1)} + p_2^{(2)}, \end{aligned}$$

where

$$\bullet p_1 := \int_0^1 \left( \sum_{j=1}^2 p_{j|1}(x) \right) dx \approx 0.555711 \text{ is the}$$

- probability that  $W_1$  is selected.
- $p_2^{(1)} := \int_0^1 \left( \sum_{j=1}^2 (1-x) \cdot p_{j|1}(x) \right) dx \approx 0.240104$  is the probability that  $W_2$  is selected as a 1-potential.
  - $p_2^{(2)} := \int_0^1 \left( \sum_{j=1}^2 x \cdot p_{j|2}(x) \right) dx \approx 0.181441$  is the probability that  $W_2$  is selected as a 2-potential.
  - $p_2 := p_2^{(1)} + p_2^{(2)} \approx 0.421545$  is the probability that  $W_2$  is selected.

**Analyzing  $\alpha$ -Filtering Threshold Algorithm.** We consider  $\mathcal{A}_\alpha$  for some  $\alpha \in [0, 1]$ .

**Case**  $\frac{W_2}{W_1} < \alpha$ . This is simple, because  $W_2$  will be filtered out when considered as a 2-potential; hence, the expected payoff is at least  $p_1 W_1 + p_2^{(1)} W_2$ .

**Case**  $\frac{W_2}{W_1} \geq \alpha$ .  $W_2$  will pass the filter when it is considered as a 2-potential; hence, the expected payoff is at least  $p_1 W_1 + p_2 W_2$  as before. However, in order to get an improvement, we do a more careful analysis to consider the probability  $\hat{p}_2$  that the algorithm includes an item with weight at least  $\alpha W_1$  as a 2-potential. Observe that this event includes  $W_2$  being selected as a 2-potential, and so  $\hat{p}_2 \geq p_2^{(2)}$ . Hence, there is an improvement in the expected payoff:  $p_1 W_1 + p_2 W_2 + \tilde{p}\alpha W_1$ , where  $\tilde{p} = \hat{p}_2 - p_2^{(2)} \geq 0$ .

**Why Does Filtering Help?** Intuitively, this offers improvement because, if other items have weights less than  $\alpha W_1$ , then they might be filtered out and this increases the chance of  $W_2$  being selected. On the other hand, if there are other items with weights at least  $\alpha W_1$ , then they make a significant contribution to the payoff if they are selected, as opposed to zero contribution in the secretary case. We give a lower bound for the probability  $\hat{p}_2$  in the following lemma.

**LEMMA 4.1. (POWER OF FILTERING)** Suppose  $\frac{W_2}{W_1} \geq \alpha$ . The probability that algorithm  $\mathcal{A}_\alpha$  selects an item with weight at least  $\alpha W_1$  as a 2-potential is  $\hat{p}_2 \geq 0.208159$ . In particular, this implies that  $\tilde{p} = \hat{p}_2 - p_2^{(2)} \geq 0.026718$ .

*Proof.* In the proof, we shall consider lower bounds for the conditional probabilities of the form that a certain quota  $Q_j$  is still available at some time, given some event. Since having more items can only make this probability smaller, the infinite model is the worst case.

Recall that in  $\mathcal{A}_0$  (and also in  $\mathcal{A}_\alpha$ ), quota  $Q_2$  is used before  $Q_1$ , and  $\tau_{j,k}$  is the time after which quota  $Q_j$  may be used for a  $k$ -potential. The optimal thresholds  $\tau_{2,1} \leq \tau_{1,1} \leq \tau_{2,2} \leq \tau_{1,2}$  are calculated exactly by our techniques in Section 3.4.1, and we also obtain the optimal solution  $p_{j|k}(x)$ 's.

We analyze disjoint events, each of which implies the event in question. Hence, the sum of their probabilities gives a lower bound.

- $W_1$  arrives in  $[0, \tau_{2,1}]$  and  $W_2$  arrives in  $[\tau_{2,2}, 1]$ . In this case, there must exist some item arriving in  $[\tau_{2,2}, 1]$  that has weight at least  $\alpha W_1$  and is selected as a 2-potential. The probability is  $\tau_{2,1}(1 - \tau_{2,2})$ .
- In the following events,  $W_1$  arrives after  $\tau_{2,1}$ , and hence they are disjoint from the case above. In these cases, quota  $Q_1$  is used for the 2-potential with weight at least  $\alpha W_1$ .
  - $W_1$  arrives in  $[\tau_{2,1}, \tau_{1,1}]$  and  $W_2$  arrives in  $[\tau_{1,2}, 1]$ . This happens with probability  $(\tau_{1,1} - \tau_{2,1})(1 - \tau_{1,2})$ .
  - (i)  $W_1$  arrives in  $[\tau_{1,1}, \tau_{2,2}]$ , (ii)  $Q_2$  is still available when  $W_1$  arrives and (iii)  $W_2$  arrives in  $[\tau_{1,2}, 1]$ . Events (i) and (ii) ensure that quota  $Q_1$  is still available after time  $\tau_{2,2}$ , and event (iii) ensures that there is some item (such as  $W_2$ ) that can pass the filter. Event (iii) happens with probability  $(1 - \tau_{1,2})$ , which is independent of the first two events in the infinite model.  
We compute the probability that (i) and (ii) happen. Suppose  $W_1$  arrives at some  $y \in [\tau_{1,1}, \tau_{2,2}]$ , then we require the largest-weighted item in  $[0, y)$  arrive before  $\tau_{2,1}$ , which happens with probability  $\frac{\tau_{2,1}}{y}$ . Hence, the first two events happen with probability given by the integral  $\int_{\tau_{1,1}}^{\tau_{2,2}} \frac{\tau_{2,1}}{y} dy$ .
  - (i)  $W_1$  arrives in  $[\tau_{2,2}, \tau_{1,2}]$ , (ii)  $Q_2$  is still available when  $W_1$  arrives and (iii)  $W_2$  arrives in  $[\tau_{1,2}, 1]$ . This case is similar to (b2), except that if  $W_1$  arrives at  $y \in [\tau_{2,2}, \tau_{1,2}]$ , among items arriving in  $[0, y)$ , we require the best item arrive before  $\tau_{2,1}$  and the stronger condition that the second best arrive before  $\tau_{2,2}$ , i.e, we exclude the case that a 2-potential could arrive in  $[\tau_{2,2}, y)$  but is filtered. This ensures that  $Q_1$  is still available at time  $\tau_{1,2}$ . Events (i) and (ii) happen with probability at least  $\int_{\tau_{2,2}}^{\tau_{1,2}} \frac{\tau_{2,1}\tau_{2,2}}{y^2} dy$ , and as before, event (iii) happens independently with probability  $(1 - \tau_{1,2})$ .
  - (i)  $W_1$  arrives in  $[\tau_{1,2}, 1]$ , (ii)  $Q_2$  is still available when  $W_1$  arrives, and (iii)  $W_2$  arrives after  $W_1$ . Conditioning on  $W_2$  arriving at  $x \in [\tau_{1,2}, 1]$ , and  $W_1$  arriving at  $y \in [\tau_{1,2}, x]$ , using the same analysis as in (b3), the conditional probability that  $Q_2$  is still available at time  $y$  is at least  $\frac{\tau_{2,1}\tau_{2,2}}{y^2}$ . Hence, the probability for this case is at least  $\int_{\tau_{1,2}}^1 \int_{\tau_{1,2}}^x \frac{\tau_{2,1}\tau_{2,2}}{y^2} dy dx$ .

Hence, summing the above probabilities, we have

$$\begin{aligned}\hat{p}_2 &\geq \tau_{2,1}(1 - \tau_{2,2}) + (1 - \tau_{1,2}) \left( \tau_{1,1} - \tau_{2,1} + \int_{\tau_{1,1}}^{\tau_{2,2}} \frac{\tau_{2,1}}{y} dy \right. \\ &\quad \left. + \int_{\tau_{2,2}}^{\tau_{1,2}} \frac{\tau_{2,1}\tau_{2,2}}{y^2} dy \right) + \int_{\tau_{1,2}}^1 \int_{\tau_{1,2}}^x \frac{\tau_{2,1}\tau_{2,2}}{y^2} dy dx.\end{aligned}$$

Substituting values for optimal thresholds  $\tau_{2,1}$ ,  $\tau_{1,1}$ ,  $\tau_{2,2}$  and  $\tau_{1,2}$ , we have  $\hat{p}_2 > 0.208159$  and  $\tilde{p} := \hat{p}_2 - p_2^{(2)} > 0.026718$ . ■

**Picking a Suitable Filter  $\alpha$ .** If  $\alpha$  is too large, then the adversary can choose  $W_2$  to be just below the threshold to cause the algorithm to filter  $W_2$  out as a 2-potential. On the other hand, if  $\alpha$  is too small, then the benefit of filtering would be insignificant. We show how to pick a suitable value for  $\alpha$ .

**THEOREM 4.1. (OPTIMAL  $\mathcal{A}_\alpha$  BEATS  $\mathcal{A}_0$ )** For  $\alpha = 0.25289$ , for the online 2-item auction, the filtering threshold algorithm  $\mathcal{A}_\alpha$  has a performance ratio of at least 0.492006, which is strictly larger than  $\rho_{2,2} \approx 0.488628$ , the ratio of algorithm  $\mathcal{A}_0$ , whose decisions are based on relative merits.

*Proof.* Consider a filtering threshold algorithm  $\mathcal{A}_\alpha$  for some  $\alpha \in [0, 1]$ . There are two situations for the worst case.

(1) The ratio  $\frac{W_2}{W_1}$  is below  $\alpha$ , causing  $\mathcal{A}_\alpha$  to miss  $W_2$  as a 2-potential. In this case, the performance ratio is  $\frac{p_1 W_1 + p_2^{(1)} W_2}{W_1 + W_2} \geq \frac{p_1 + p_2^{(1)} \alpha}{1 + \alpha}$ , where the worst case is that  $\frac{W_2}{W_1}$  approaches  $\alpha$  from below.

(2) The ratio  $\frac{W_2}{W_1}$  is above  $\alpha$ , in which case the performance ratio is at least  $\frac{p_1 W_1 + p_2 W_2 + \tilde{p} \alpha W_1}{W_1 + W_2} \geq \frac{p_1 + p_2 + \tilde{p} \alpha}{2}$ , where the worst case is when  $W_1 = \tilde{W}_2$ .

Hence, it suffices to choose  $\alpha$  to maximize  $\min\{\frac{p_1 + p_2^{(1)} \alpha}{1 + \alpha}, \frac{p_1 + p_2 + \tilde{p} \alpha}{2}\}$ . When  $\alpha = 0.25289$ , the minimum is at least 0.492006, as required. ■

## 5 General Transformation from Secretary to Matching

In online bipartite matching, the offline nodes can have different capacities. We call this problem edge-weighted online bipartite  $K$ -matching, where each offline node  $r$  has a capacity  $K_r$ . Theorem 1.5 is implied by the following result.

**THEOREM 5.1. (APPLYING SECRETARY TO MATCHING WITH NON-UNIFORM CAPACITIES)** For the online bipartite  $K$ -matching with offline nodes set  $R$ , any collection of monotone algorithms  $\mathcal{A}_r$  for the  $(K_r, K_r)$ -secretary problem can be transformed into an algorithm  $\widehat{\mathcal{A}}$  for the matching problem such that the performance ratio of  $\widehat{\mathcal{A}}$  is the minimum of performance ratios of

$\mathcal{A}_r$  over  $r \in R$ . In particular, the optimal  $(K_r, K_r)$ -threshold algorithms induce a deterministic algorithm with performance ratio at least  $1 - O(\frac{1}{\sqrt{K_{\min}}})$ , where  $K_{\min} = \min_{r \in R} K_r$ .

Given a collection of monotone algorithms  $\mathcal{A}_r$  for the  $(K_r, K_r)$ -secretary problem, we construct an algorithm  $\widehat{\mathcal{A}}$  for the online bipartite  $K$ -matching, where we can imagine that each offline node  $r$  runs its  $(K_r, K_r)$ -secretary algorithm  $\mathcal{A}_r$  to decide whether to select an online node. Observe that the  $z$  values are determined by the recursion in Lemma 2.3. The monotone property will be needed in the performance analysis.

**Algorithm  $\widehat{\mathcal{A}}$  for the online  $K$ -matching problem**

**Input:** Graph  $G = (L \cup R, E)$  with edge weight  $w$ , economical  $(K_r, K_r)$ -secretary algorithms  $\mathcal{A}_r$  for all  $r \in R$

Let  $M := \emptyset$ ,  $\widehat{R} := \emptyset$  and  $L_0 := \emptyset$ .

For each node  $r \in R$ , make  $K_r$  copies  $r^1, \dots, r^{K_r}$  of  $r$  and add them to  $\widehat{R}$ .

For the  $i$ -th online node  $v \in L$ :

Let  $\widehat{E} := \emptyset$ .

Compute  $z_{j|k}^r(i)$  with respect to  $\mathcal{A}_r$  for each  $r \in R$ ,  $j \in [K_r]$  and  $k \in [K_r]$  according to Lemma 2.3.

Let  $\gamma_k^r(i) := \sum_{j=1}^{K_r} z_{j|k}^r(i)$  for  $r \in R$  and  $k \in [K_r]$ .

Let  $L_i := L_{i-1} \cup \{v\}$ .

For each  $v' \in L_i$  and  $r \in R$ , if  $\{v', r\}$  is an edge in  $G$ , then for each  $k \in [K_r]$ , add an edge  $\{v', r^k\}$  with weight  $\widehat{w}_{v', r^k} := \gamma_k^r(i) \cdot w_{v', r}$  to  $\widehat{E}$ .

Find a maximum weight (1-)matching  $M_i$  on the graph  $(L_i \cup \widehat{R}, \widehat{E})$  with respect to  $\widehat{w}$ .

If  $v$  is matched to some  $r^k$  in  $M_i$ :

Let  $j$  be the remaining capacity of  $r$  after matching nodes in  $M$ .

If  $\mathcal{A}_r^i(j, k) = 1$ , then add edge  $\{v, r\}$  to  $M$ .

**Output:** A  $K$ -matching  $M$ .

For the  $i$ -th online node  $v \in L$ , if  $v$  is matched to  $r^k$  in  $M_i$ , then we say  $v$  and  $r$  are *partners* of each other in  $M_i$ , and  $\{v, r^k\}$  is a *potential k* edge in  $M_i$ . The following property is useful in our analysis.

**FACT 5.1.** For any real numbers  $x_1 \geq \dots \geq x_K$  and  $y_1 \geq \dots \geq y_K$ , we have

$$\sum_{k=1}^K x_k y_k \geq \frac{1}{K} \sum_{k=1}^K x_k \cdot \sum_{k'=1}^K y_{k'}.$$

**LEMMA 5.1. (COUPLING BETWEEN MATCHING AND SECRETARY)** Suppose the  $i$ -th online node  $v_i$  is matched to some  $r^k$  in  $M_i$  for some  $r \in R$  and  $k \in [K_r]$ . Then, the edge  $\{v, r\}$  is added to  $M$  with probability at least  $\gamma_k^r(i)$ , where  $\gamma_k^r(i)$  is the probability that  $\mathcal{A}_r$  makes a selection at step  $i$ , given that a  $k$ -potential arrives.

*Proof.* We denote  $\mathcal{A} := \mathcal{A}_r$ ,  $K := K_r$  and  $\gamma := \gamma^r$ . For  $\iota \in [i]$ , let  $v_\iota$  be the online node that arrives at step  $\iota$ .

**Conditioning on  $L_i$  and  $v_i$ .** As argued in [12], we condition on the set  $L_i$  and node  $v_i$  and show that the remaining randomness of the relative order among nodes in  $L_{i-1}$  is enough to give the desired lower bound on the probability.

**Potential Sequence.** For  $\iota \in [i]$ , if node  $v_\iota$  is matched to  $r^k$  in  $M_\iota$  for some  $k \in [K]$ , then we say that  $r$  observes potential  $k$  at step  $\iota$ , otherwise  $r$  observes potential  $\perp$ . From the hypothesis,  $r$  observes potential  $k$  at step  $i$ . Suppose  $\sigma_0 \in [K^*]^{i-1}$  is the potential sequence observed by node  $r$  in the first  $i-1$  steps. It follows that the remaining capacity of  $r$  at the beginning of step  $i$  is determined from  $\sigma_0$  by algorithm  $\mathcal{A}$ .

Consider a potential sequence  $\sigma_1 \in [K^*]^{i-1}$  observed by an economical algorithm in the first  $i-1$  steps in the  $(K, K)$ -secretary problem.

**Coupling.** The crucial step in the proof is to set up a coupling between  $\sigma_0$  and  $\sigma_1$  such that for any  $\iota < i$ , if  $\sigma_0(\iota) = k$  for some  $k \in [K]$ , then  $\sigma_1(\iota) = k$ .

Observe that when an economical algorithm is run on two such potential sequences  $\sigma_0$  and  $\sigma_1$ , at the beginning every step  $\iota \leq i$ , the remaining capacity in  $\sigma_0$  is at least the number of remaining quotas in  $\sigma_1$ . This can be proved by induction on  $\iota$ , where the base case  $\iota = 1$  is trivial. For the inductive step, suppose the claim is true at the beginning of step  $\iota$ . The capacity of  $r$  can decrease at step  $\iota$  only if  $\sigma_0(\iota) \in [K]$ , in which case  $\sigma_1(\iota) = \sigma_0(\iota)$ . Hence, if the remaining capacity is the same as the number of remaining quotas, then both quantities remain the same after step  $\iota$ ; otherwise, if the remaining capacity is strictly larger, then even if it is decreased by 1, the conclusion still holds after step  $\iota$ .

Hence, it follows that at the beginning of step  $i$ , the remaining capacity at  $r$  after observing  $\sigma_0$  is at least the number of remaining quotas after observing  $\sigma_1$ . Hence, by the monotonicity of  $\mathcal{A}$ ,  $r$  is matched to  $v_i$  in  $\sigma_0$ , if a selection is made at step  $i$  after observing  $\sigma_1$ , given that a  $k$ -potential arrives. This gives the required bound on the probability. We next set up the required coupling to complete the proof.

**Backward Analysis.** As in [12], we consider a backward randomized procedure. Suppose the potential sequences  $\sigma_0$  and  $\sigma_1$  have already been determined from step  $\iota+1$  to  $i-1$ , during which the identities of nodes  $v_{\iota+1}$  to  $v_{i-1}$  have been revealed, and we now perform sampling at step  $\iota$  to determine the identity of  $v_\iota$ .

An important property of random permutation is that given the potential sequence  $\sigma_1[\iota+1..i-1]$ , the distribution of  $\sigma_1[1..\iota]$  does not change. Hence, setting  $\kappa = \min\{K, \iota\}$ , for  $k \in [\kappa]$ , it follows that  $\sigma_1(\iota) = k$  with probability  $\frac{1}{\iota}$ , and  $\sigma_1(\iota) = \perp$  with probability  $1 - \frac{\kappa}{\iota}$ .

Another consequence is that by the monotonicity of  $\mathcal{A}$ ,  $z_{j|1}(\iota) \geq z_{j|2}(\iota) \geq \dots \geq z_{j|K}(\iota)$ , which implies that  $\gamma_1(\iota) \geq \gamma_2(\iota) \geq \dots \geq \gamma_K(\iota)$ . Since  $M_\iota$  is a maximum weight matching in  $(L_\iota \cup \widehat{R}, \widehat{E})$  with respect to  $\widehat{w}$ , it follows that there exists  $\ell \leq \kappa$  such that  $r^1, r^2, \dots, r^\ell$  are matched in  $M_\iota$ , but  $r^{\ell+1}, \dots, r^K$  are not. Hence, by using the randomness to reveal the identity of  $v_\iota$ , we see that for  $k \in [\ell]$ ,  $\sigma_0(\iota) = k$  with probability  $\frac{1}{\iota}$ , and  $\sigma_0(\iota) = \perp$  with probability  $1 - \frac{\ell}{\iota}$ .

This readily gives a way to form the coupling. First, generate  $\sigma_1(\iota)$  as described above. If  $\sigma_1(\iota) \in [\ell]$ , then set  $\sigma_0(\iota) := \sigma_1(\iota)$ , otherwise set  $\sigma_0(\iota) := \perp$ . This completes the construction for the coupling argument and also the proof of Lemma 5.1. ■

**LEMMA 5.2. (LOCAL PERFORMANCE)** *Let  $\text{OPT}$  be the weight of an optimal  $K$ -matching in the graph fixed by the adversary. Let  $\rho^*$  be the minimum performance ratio of  $\mathcal{A}_r$  over  $r \in R$ . Then, the expected payoff is at least  $\rho^* \cdot \text{OPT}$ .*

*Proof.* Recall that  $M_i$  is a maximum weight 1-matching in  $G_i := (L_i \cup \widehat{R}, \widehat{E})$  with respect to  $\widehat{w}$ . (We suppress the dependence on  $i$  in  $\widehat{E}$  without risk of ambiguity.) Moreover,  $M$  is the final matching returned by the algorithm.

**Conditioning on  $L_i$ .** Conditioning on  $L_i$ , the matching  $M_i$  is determined. We consider the randomness of the node  $v_i$  among  $L_i$  to arrive at step  $i$ . Let  $e_i$  be the edge matching  $v_i$  in  $M_i$  if  $v_i$  is matched, and a dummy edge of weight 0 otherwise.

For  $r \in R$  and  $k \in [K_r]$ , let  $e_k^r$  be the edge matching  $r^k$  in  $M_i$  if  $r^k$  is matched, and a dummy edge of weight 0 otherwise. Note that for each  $e \in M_i$ ,  $\Pr[e_i = e | M_i] = \frac{1}{i}$ . From Lemma 5.1,  $\Pr[v_i \text{ matched in } M | e_i = e_k^r, M_i] \geq \gamma_k^r(i)$ . Hence, the expected payoff from  $v_i$  (conditioning on  $L_i$ ) is

$$\begin{aligned} & \sum_{r \in R} \sum_{k=1}^{K_r} w_{e_k^r} \cdot \Pr[e_i = e_k^r \wedge v_i \text{ matched in } M | M_i] \\ &= \sum_{r \in R} \sum_{k=1}^{K_r} w_{e_k^r} \cdot \Pr[e_i = e_k^r | M_i] \\ &\quad \cdot \Pr[v_i \text{ matched in } M | e_i = e_k^r, M_i] \\ &\geq \frac{1}{i} \sum_{r \in R} \sum_{k=1}^{K_r} w_{e_k^r} \cdot \gamma_k^r(i) \\ &= \frac{1}{i} \sum_{r \in R} \sum_{k=1}^{K_r} \widehat{w}_{e_k^r} = \frac{1}{i} \cdot \widehat{w}(M_i). \end{aligned}$$

**Comparing with Offline Optimal  $M^*$ .** Let  $M^*$  be an offline optimal  $K$ -matching in the adversary graph  $G$ . Given  $L_i$ , we shall construct a 1-matching  $M_i^*$  in  $G_i$ . By the choice of  $M_i$ , we have  $\widehat{w}(M_i) \geq \widehat{w}(M_i^*)$ . Hence, it suffices to define such an  $M_i^*$  and give a lower bound for the expectation of its weight with respect to  $\widehat{w}$ .

Given  $L_i$ , we first consider the induced matching  $M^*[L_i]$  obtained by keeping only online nodes in  $L_i$ . We form a 1-matching  $M_i^*$  in  $G_i$  in the following way.

For each offline node  $r$ , suppose that its degree in  $M^*[L_i]$  is  $1 \leq \kappa \leq K_r$ ; if it has degree 0, then it is not matched in  $M_i^*$ . Sort the edges in non-increasing order of weights (which were given by the adversary):  $\omega_1 \geq \omega_2 \geq \dots \geq \omega_\kappa$  (where ties are resolved consistently using the identities of nodes for instance) such that edge  $\{u^k, r\} \in M^*[L_i]$  has weight  $\omega_k$ . For  $k \in [\kappa]$ , include  $\{u^k, r^k\}$  in  $M_i^*$ , in which case we say that edge  $\{u^k, r\} \in M^*$  becomes a potential- $k$  edge  $\{u^k, r^k\}$  in  $M_i^*$ . Define  $M_i^*[r] := \{\{u^k, r^k\} : k \in [\kappa]\} \subset M_i^*$ , and  $M^*[r]$  to be the set of edges incident to  $r$  in  $M^*$ .

**Analyzing  $\hat{w}(M_i^*)$ .** Fix  $r \in R$ . It suffices to compare the expectation  $\mathbf{E}[\hat{w}(M_i^*[r])]$  with  $w(M^*[r])$ . Suppose the weights of edges in  $M^*[r]$  is  $w_1 \geq w_2 \geq \dots \geq w_{K_r}$  (where ties are resolved consistently). Suppose the degree of  $r$  in  $M^*$  is  $\kappa$ . For convenience of notation, if  $\kappa < K_r$ , then for  $l > \kappa$ , we have  $w_l := 0$ , and hence, all terms with index  $l > \kappa$  are multiplied by zero.

For  $\ell \in [\kappa]$ , observe that the edge  $e_\ell \in M^*$  with weight  $w_\ell$  is included in  $M_i^*$  with probability  $\frac{i}{n}$ . Now, it is useful to recall the quantities  $\delta, \gamma, \beta$  defined in Lemma 2.2. For  $k \in [\ell]$ ,  $\delta_{k|\ell}(i)$  is exactly the conditional probability that edge  $e_\ell$  becomes a potential- $k$  edge in  $M_i^*$ , given that  $e_\ell$  is included in  $M_i^*$ . Recall that if edge  $e_\ell$  becomes a potential- $k$  edge in  $G_i$ , its weight is multiplied by  $\gamma_k^r(i)$  in  $\hat{w}$ . It follows that  $\mathbf{E}[\hat{w}(M_i^*[r])] = \frac{i}{n} \sum_{\ell \in [K_r]} \sum_{k \in [\ell]} \delta_{k|\ell} \cdot \gamma_k^r(i) \cdot w_\ell = \frac{i}{n} \sum_{\ell \in [K_r]} \beta_\ell(i) \cdot w_\ell$ , where  $\beta_\ell(i) := \sum_{k=1}^\ell \delta_{k|\ell}(i) \cdot \gamma_k^r(i)$ .

From Lemma 2.4,  $\beta_\ell(i)$  is non-increasing in  $\ell$ , and so is  $w_\ell$ . By Fact 5.1,  $\sum_{\ell \in [K_r]} \beta_\ell(i) \cdot w_\ell \geq \frac{1}{K_r} \sum_{\ell \in [K_r]} \beta_\ell(i) \cdot \sum_{\ell \in [K_r]} w_\ell = \frac{1}{K_r} \sum_{\ell \in [K_r]} \beta_\ell(i) \cdot w(M^*[r])$ . Hence  $\mathbf{E}[\hat{w}(M_i^*[r])] \geq \frac{i}{K_r n} \sum_{\ell \in [K_r]} \beta_\ell^r(i) \cdot w(M^*[r])$ .

**Wrapping Up.** Observe that each  $\mathcal{A}_r$  has performance ratio  $\frac{1}{K_r n} \sum_{i=1}^n \sum_{\ell \in [K_r]} \beta_\ell^r(i) \geq \rho^*$ . Hence, the expected payoff of  $\hat{\mathcal{A}}$  is at least

$$\begin{aligned} \sum_{i=1}^n \frac{1}{i} \cdot \mathbf{E}[\hat{w}(M_i)] &\geq \sum_{i=1}^n \frac{1}{i} \sum_{r \in R} \mathbf{E}[\hat{w}(M_i^*[r])] \\ &\geq \sum_{i=1}^n \sum_{r \in R} \frac{1}{K_r n} \sum_{\ell \in [K_r]} \beta_\ell^r(i) \cdot w(M^*[r]) \\ &= \sum_{r \in R} w(M^*[r]) \cdot \frac{1}{K_r n} \sum_{i=1}^n \sum_{\ell \in [K_r]} \beta_\ell^r(i) \\ &\geq \sum_{r \in R} w(M^*[r]) \cdot \rho^* = \rho^* \cdot \text{OPT}. \end{aligned} \quad \blacksquare$$

### Proof of Theorem 5.1 (Performance Ratio):

Since the optimal  $(k, k)$ -threshold algorithm has ratio  $\rho_{k,k} \geq 1 - O(\frac{1}{\sqrt{k}})$ , it follows that, if each  $\mathcal{A}_r$  is the optimal  $(K_r, K_r)$ -threshold algorithm, then algorithm  $\hat{\mathcal{A}}$  has a ratio at least  $\rho^* \geq 1 - O(\frac{1}{\sqrt{K_{\min}}})$ , where  $K_{\min} = \min_{r \in R} K_r$ , completing the proof of Theorem 5.1.  $\blacksquare$

## References

- [1] Moshe Babaioff, Michael Dinitz, Anupam Gupta, Nicole Immorlica, and Kunal Talwar. Secretary problems: weights and discounts. In *SODA*, pages 1245–1254, 2009.
- [2] Moshe Babaioff, Nicole Immorlica, David Kempe, and Robert Kleinberg. Online auctions and generalized secretary problems. *SIGecom Exchanges*, 7(2), 2008.
- [3] F. Thomas Bruss. A unified approach to a class of best choice problems with an unknown number of options. *The Annals of Probability*, 12(3):pp. 882–889, 1984.
- [4] Niv Buchbinder, Kamal Jain, and Mohit Singh. Secretary problems via linear programming. In *IPCO*, pages 163–176, 2010.
- [5] T.-H. Hubert Chan and Fei Chen. A primal-dual continuous LP method on the multi-choice multi-best secretary problem. *CoRR*, abs/1307.0624, 2013.
- [6] E. B. Dynkin. The optimum choice of the instant for stopping a Markov process. *Soviet Math. Dokl*, 4, 1963.
- [7] Thomas S. Ferguson. Who solved the secretary problem? *Statistical Science*, 4(3):pp. 282–289, 1989.
- [8] P. R. Freeman. The secretary problem and its extensions: A review. *International Statistical Review*, 51(2):pp. 189–206, 1983.
- [9] John P Gilbert and Frederick Mosteller. Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, pages 35–73, 1966.
- [10] S. Gusein-Zade. The problem of choice and the optimal stopping rule for a sequence of independent trials. *Theory of Probability & Its Applications*, 11(3):472–476, 1966.
- [11] Nicole Immorlica, Robert D. Kleinberg, and Mohammad Mahdian. Secretary problems with competing employers. In *WINE*, pages 389–400, 2006.
- [12] Thomas Kesselheim, Klaus Radke, Andreas Tönnis, and Berthold Vöcking. An optimal online algorithm for weighted bipartite matching and extensions to combinatorial auctions. In *ESA*, pages 589–600, 2013.
- [13] Thomas Kesselheim, Klaus Radke, Andreas Tönnis, and Berthold Vöcking. Primal beats dual on online packing LPs in the random-order model. In *STOC*, pages 303–312, 2014.
- [14] Robert D. Kleinberg. A multiple-choice secretary algorithm with applications to online auctions. In *SODA*, pages 630–631, 2005.
- [15] Nitish Korula and Martin Pál. Algorithms for secretary problems on graphs and hypergraphs. In *ICALP* (2), pages 508–520, 2009.
- [16] Johann Heinrich Lambert. *Acta Helvetica Physico-Mathematico-Anatomico-Botanico-Medica*, volume v. 3 (1758). Basileae., 1758.
- [17] N. Levinson. A class of continuous linear programming problems. *Journal of Mathematical Analysis and Applications*, 16:73–83, 1966.
- [18] D. V. Lindley. Dynamic programming and decision theory. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 10(1):pp. 39–51, 1961.
- [19] William F. Tyndall. A duality theorem for a class of continuous linear programming problems. *Journal of the Society for Industrial and Applied Mathematics*, 13(3):pp. 644–666, 1965.