Dueling Bandits with Team Comparisons

Anonymous Author(s)
Affiliation
Address
email

Abstract

We introduce the *dueling teams problem*, a new online-learning setting in which the learner observes noisy comparisons of disjoint pairs of k-sized *teams* from a universe of n players. The goal of the learner is to minimize the number of duels required to identify, with high probability, a *Condorcet winning team*, i.e., a team which wins against any other disjoint team (with probability at least $1/2$). Noisy comparisons are linked to a total order on the teams. We formalize our model by building upon the dueling bandits setting (Yue et al., 2012) and provide several algorithms, both for stochastic and deterministic settings. For the stochastic setting, we provide a reduction to the classical dueling bandits setting, yielding an algorithm that identifies a Condorcet winning team within $\mathcal{O}((n+k\log(k))\frac{\max(\log\log n,\log k)}{\Delta^2})$ duels, where Δ is a gap parameter. For deterministic feedback, we additionally present a gap-independent algorithm that identifies a Condorcet winning team within $\mathcal{O}(nk\log(k)+k^5)$ duels.

1 Introduction

Multi-arm bandits (MAB) is a classical model of decision making under uncertainty. In spite of the simplicity of the model, it already incorporates the essential tradeoff between exploration and exploitation. In MAB, the learner performs actions and can only observe rewards of the actions performed. One of the main tasks in MAB is *best arm identification*, where the goal is to identify a near-optimal action while minimizing the number of actions executed. The MAB model has numerous practical applications, including online advertising, recommendation systems, clinical trials, and more. (See Slivkins (2019); Lattimore & Szepesvári (2020) for more background).

One weakness of the MAB model is the assumption that real-valued rewards are always available. In many applications, it is more natural to compare two actions and observe which one of them is better rather than give every single action a numerical reward. For example, recommendation systems often suggest two items and obtain only their relative preference as feedback (e.g., by a click on one of them). This leads very naturally to the well-known model of dueling bandits (Yue et al., 2012), where the learner selects a pair of actions each time and observes the binary "winner" of a duel between the two. See Busa-Fekete et al. (2018) for a survey on extensions of this model.

In this work we are interested in the case that the learner has to select two disjoint *teams* for a duel, which are k-sized subsets of the actions (which we call players). This appears naturally in sports or online games, where the goal is to pick one of the best teams from a set of players by observing the outcomes of matches (say, to be a school representative team, or to sponsor for tournaments). Examples include doubles tennis, basketball, and the online game League of Legends, where each match requires two disjoint teams of players to compete. Similar phenomena appear in working environments, where different R&D teams compete on implementing a project. Another example could be online advertisements where multiple products are bundled to a display ad and a customer can click on any of two presented bundles, e.g., some online games offer in-app bundle purchases, and the information regarding sales of different bundles can be used to improve the bundles' composition.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

Our basic model is the following. We have a universe of n players, and at each iteration the learner selects two **disjoint** teams for a duel and observes the winner. For any two different teams, there exists an unknown stationary probability that determines the winner of a duel between them. The requirement that teams need to be disjoint is in accordance with the situation in games, where a single person cannot play for both teams. The goal of the learner is to minimize the number of duels required to identify, with high probability, a *Condorcet winning team*, i.e., a team which wins against any other disjoint team (with a probability of at least $1/2$). We do assume these probabilities are linked to a strict total order on all teams. This implies the existence of a Condorcet winning team, yet it is typically not unique. We make two minimal and natural assumptions on this total order on teams, namely, that it is *consistent* to some total order among the players, and that the team probabilistic comparisons hold *Strong Stochastic Transitivity*, a common assumption in dueling bandit settings.

Clearly, given any total order among the players, the best team is the one containing the top k players, which is in particular one of the Condorcet winning teams. However, not all relations between players are deducible for the learner. In particular, even achieving accurate estimations of the latent winning probabilities between all disjoint teams might not suffice to separate the top k players from the rest. Consider for example an instance with four players $1 \succ 2 \succ 3 \succ 4$ where $k = 2$ and the total order among the teams is lexicographical, i.e., $12 \succ 13 \succ 14 \succ 23 \succ 24 \succ 34$. Then, there exist three feasible duels, each of which is won by the team containing player 1 with probability greater than $1/2$. If all three duels are won with equal probability by the team containing 1, the learner has no chance of detecting the team 12 as the top k team. However, any of the teams $12, 13$ and 14 is a Condorcet winning team.

Our main target is to present algorithms for which the number of duels is bounded by a polynomial in the number of players n and team size k, although the number of teams is exponential in k, i.e., $\Omega((\frac{n}{k})^k)$ and the number of valid duels is $\Omega(2^k(\frac{n}{2k})^{2k})$. Even if one were to accept an exponential number of arms, a direct reduction to the standard dueling bandits setting would not be feasible as not all pairs of teams are comparable in our model. In particular, duels of the form $(S \cup \{a\}, S \cup \{b\})$, which would yield a signal regarding the relation between players a and b, are forbidden. The inherent difficulty of our endeavor comes from two limitations: (1) Not all the relations between two single players are deducible, (see example above), and (2) even for pairs of players with deducible relation, having $\Omega(2^k(\frac{n}{2k})^{2k})$ valid duels and the same amount of (latent) winning probabilities makes the task of deducing their relations hard.

We start by giving a full characterization of the *deducible* pairwise relations between players, namely relations that can be detected by a learner which is allowed to perform an unlimited amount of duels. Our characterization implies that every deducible single player relation has one of two types of *witnesses*, which are constant-size sets of duels that prove their relation. We also show that, once we find a witness for one pair of players, it can often be transferred to a witness for other pairs of players.

Building upon this characterization, we introduce a parameter $\Delta_{a,b}$ which captures the distinguishability of any two players a and b and takes a value of 0 whenever the pair is not deducible. Assuming $\Delta := \Delta_{k,k+1} > 0$, where k and $k+1$ are k^{th} and $(k+1)^{\text{th}}$ best players, we give a reduction to the classic dueling bandits problem. Combining this reduction with a high-probability top-k identification algorithm for the dueling bandits setting (e.g., Mohajer et al. (2017); Ren et al. (2020)) yields a similar sample complexity upper bound, e.g., this yields a high-probability top-k identification algorithm for dueling teams with $\mathcal{O}(\Delta^{-2}(n + k\log(k))\max(\log\log n, \log k))$ duels.

Interestingly, it turns out that the deterministic case, i.e., when winning probabilities are in $\{0, 1\}$, constitutes a challenging special case of our problem where Δ can be particularly small, or even 0. To overcome this issue we design delicate algorithms which are independent of Δ. On a high level, a preprocessing procedure first excludes as many bad players as possible. To do so, it runs a method for identifying pairwise relations between players which performs only a small number of duels, but has little control over the pair for which the relation is uncovered. For general total orders this implies an algorithm requiring $\mathcal{O}(nk\log(k) + 2^{\mathcal{O}(k)})$ duels. For the natural case of *additive linear* orders, we present a more elaborated approach for detecting a Condorcet winning team within the reduced instance, resulting in an algorithm that performs $\mathcal{O}(nk\log(k) + k^5)$ duels.

We introduce our problem in Section 2, give a characterization of deducible relations in Section 3, discuss the stochastic setting in Section 4, and the deterministic setting in Section 5. For brevity, algorithms and (full) proofs are relegated to Section A, B, and C of the supplementary material. Section D contains a discussion and Section E a characterization of additive linear total orders.

1.1 Related Work

MAB best arm or subset identification: single arm identification was initiated in Even-Dar et al. (2006) and later studied in many works including Bubeck et al. (2011); Kaufmann et al. (2016); Chen et al. (2017). This setting was extended by Kalyanakrishnan & Stone (2010) for multiple arms identification (i.e., top k arms), using a single arm samples. Other works that address the objective of top$-k$ identification include Chen et al. (2014); Zhou et al. (2014); Bubeck et al. (2013).

Dueling bandits The work of Yue et al. (2012) lay down the framework of non-parametric bandit feedback under total order among arms, strong stochastic transitivity, and stochastic triangle inequality assumptions and were followed by many subsequent works (For more, see a survey, Busa-Fekete et al. (2018).) In particular, some subsequent works target the task of identifying the top k players in this setting Mohajer et al. (2017); Ren et al. (2020).

Dueling bandits with sets of actions One line of dueling bandits extension consider the case where the learner selects a subset of actions and observes the outcomes of all duels between all pairs of actions in the subset (Brost et al., 2016; Sui et al., 2017), or the winner of the subset Saha & Gopalan (2018); Ren et al. (2018). As a consequence, these settings give the learner strictly more information than the dueling bandits setting. In contrast, feedback in our setting reveals less information.

MAB with multiple actions selection : There are works in which the learner selects a (sometimes fixed-sized) subset of actions at each iteration, and observes either all of the individual selected arms rewards (semi-bandit feedback) or an aggregated form of the rewards (full-bandit feedback), and the task is to detect to best arm or the top k. These include *combinatorial bandits* Cesa-Bianchi & Lugosi (2012), *top-k* Rejwan & Mansour (2020), *linear bandit and routing* Awerbuch & Kleinberg (2008), and more. The main difference between combinatorial bandits and our setting is the feedback.

Comparison models: Noisy pairwise comparison models, especially for sorting and ranking, have a long history which dates backs to the 1950's (For more, see a survey, Pelc (2002).). Specifically, the mathematical problem Counterfeit coin was introduced in the form of a puzzle (Grossman, 1945): given a pile of 12 coins, determine which coins has a different weight (and therefore counterfeit) using balance scales while minimizing the number of measurements. The problem was followed by numerous generalizations (see Guy & Nowakowski (1995)). While this problem is restricted to coins with two different weights, our setting can be seen as a variant with multiple weights.

2 The Dueling Teams Problem

We formalize our problem as follows. Let $n, k \in \mathbb{N}$ with $1 \leq k \leq \frac{n}{2}$. We denote the set of players by $[n] := \{1, \ldots, n\}$ and call any set of k distinct players a *team*. Moreover, we assume the existence of an underlying strict total order among all teams, and denote it by $\succ$. We also refer to $\succ$ as the ground truth order. In particular, for any two teams A and B either $A \succ B$ holds, in which case we say that A is *better* than B, or vice versa, and this relation is transitive. Additionally, we require the total order among the teams to be consistent with a total order among players and formalize this in the *consistency* assumption at the end of this section.

In each round, the learner selects an ordered pair of two disjoint teams, A and B to perform a *duel*, and receives a noisy binary feedback about which team is better. Note that in contrast to the usual dueling bandits setting, our setting does not allow duels of the form (A, A), as selecting teams with mutual players for a duel is not an option. We denote the *observable* part of $\succ$ by $\succ_{obs}$, i.e., $A \succ_{obs} B$ iff A and B are disjoint teams and $A \succ B$. Note, $\succ_{obs}$ is not transitive, thus not even a partial order.

The outcome of a duel (A, B) is sampled independently from a Bernoulli random variable with success probability $P_{A,B}$, so in particular it holds that $P_{B,A} = 1 - P_{A,B}$. We assume that the probabilistic comparisons are linked to the total order among the teams, i.e., $A \succ B$ implies $P_{A,B} > 1/2$, and that $P_{A,B}$ exists for every pair of teams (not only disjoint ones). In the stochastic setting, we denote $A > B$ if team A is the random winner of duel (A, B). In the deterministic setting, it holds that $P_{A,B} \in \{0, 1\}$ for any teams $A \neq B$. In other words, for two disjoint teams A and B the learner can observe whether $A \succ_{obs} B$ or $B \succ_{obs} A$ by performing a single duel.

A team A is a *Condorcet winning team*[1] if $A \succ_{obs} B$ for all teams B such that $A \cap B \neq \emptyset$. From our assumption on $\succ$, there always exists a Condorcet winning team, but it is not necessarily unique. The learner's goal is to minimize the number of duels required to identify, with high probability in the stochastic setting and with probability 1 in the deterministic case, a Condorcet winning team.

[1]The name is motivated by the fact that such a team is a weak Condorcet winner for the relation $\succ_{obs}$.

In the following we formalize two more assumptions we impose on our model, the former affects the linking of the probabilities to the strict total order $\succ$, the latter restricts the total order $\succ$ itself.

Strong stochastic transitivity (SST): Similarly to the dueling bandits settings in Yue et al. (2012), we assume *strong stochastic transitivity*. Namely, for every triplet of different teams $A \succ B \succ C$ it holds that $P_{A,C} \geq \max\{P_{A,B}, P_{B,C}\}$.

Consistency: We assume that the total order $\succ$ is consistent to a total order among single players. More precisely, we say that $\succ$ satisfies *consistency* if for every two players $a, b \in [n]$ either of the following holds true:

(i) $S \cup \{a\} \succ S \cup \{b\}$ for all $S \subseteq [n] \setminus \{a, b\}, |S| = k - 1$.

(ii) $S \cup \{b\} \succ S \cup \{a\}$ for all $S \subseteq [n] \setminus \{a, b\}, |S| = k - 1$.

The consistency assumption lets us derive a relation among the single players, by defining $a \succ b$ iff $S \cup \{a\} \succ S \cup \{b\}$ holds for some S. By team relation transitivity, $\succ$ implies a total order on $[n]$. Whenever we write $a \succ b$ for some players $a, b \in [n]$ this is short-hand notation for $S \cup \{a\} \succ S \cup \{b\}$ for all subsets $S \subseteq [n] \setminus \{a, b\}$ of size $k - 1$. For notational convenience, we assume without loss of generality that $1 \succ 2 \succ \cdots \succ n$ and write A^*_m for the set of players containing the top m players, i.e., $A^*_m = [m]$. In particular, the consistency assumption yields that A^*_k is a Condorcet winning team.

Though the ground truth ranking induces a total order among the players, the learner might not be able to deduce the entire order. In the following we give a characterization of the *deducible* part of $\succ$.

3 Witnesses: A Characterization of Deducible Relations

In this section we provide a high level description of the complete characterization of all the pairwise relations between single players that can be deduced via team duels. Though single players cannot be observed via team duels directly, we show a sufficient and necessary condition for deducible relations in the form of a constant number of winning probabilities of observable (feasible) duels. We refer to a set of players participating in such duels as *witnesses*. For completeness, we point out that a similar characterization can be done for any same-sized subsets of size less than k.

We denote by $\mathcal{C}_{obs}$ the set of strict total orders which are compatible with the winning probabilities of observable duels, i.e., $\{P_{A,B} \mid A \text{ and } B \text{ are disjoint teams}\}$, and satisfy consistency. More precisely, $\succ' \in \mathcal{C}_{obs}$ if $\succ'$ is a total order on all teams that satisfies consistency and there exist probabilities $P'_{A,B}$ for all pair of teams (A, B) such that $A \succ' B$ iff $P'_{A,B} > 1/2$, and $P'_{A,B} = P_{A,B}$ for all disjoint teams A and B. Lastly, we define $A \succ^* B$ if and only if $A \succ' B$ for all $\succ' \in \mathcal{C}_{obs}$, where A and B are not necessarily disjoint. We refer to $\succ^*$ as the *deducible* relation. For single player relations, we define $a \succ^* b$ if and only if there exists $S \subseteq [n] \setminus \{a, b\}$ such that $S \cup \{a\} \succ' S \cup \{b\}$ for all $\succ' \in \mathcal{C}_{obs}$.

Next, we define two sets of *potential witnesses* that have a simple structure and, in some cases, allow us to deduce single players relation: (1) A *potential subsets witnesses* set, denoted by $\mathcal{S}_{a,b}$, that contains all pairs (S, S') such that S and S' are disjoint subsets of $[n] \setminus \{a, b\}$ and both are of size $k - 1$, and (2) A *potential subset-team witnesses* set, denoted by $\mathcal{T}_{a,b}$, that contains all pairs (S, T) where S and T are disjoint subsets of $[n] \setminus \{a, b\}$, such that S is of size $k - 1$ and T is of size k (and is therefore a team). Below, we define under which conditions a potential witnesses is a *witness*.

Definition 3.1. *An element $(S, S') \in \mathcal{S}_{a,b}$ is a* subsets witness *for $a \succ b$ if $P_{S\cup\{a\},S'\cup\{b\}} > P_{S\cup\{b\},S'\cup\{a\}}$. An element $(S, T) \in \mathcal{T}_{a,b}$ is a* subset-team witness *for $a \succ b$ if $P_{S\cup\{a\},T} > P_{S\cup\{b\},T}$.*

We capture the set of the elements of $\mathcal{S}_{a,b}$ that are subsets witnesses for $a \succ b$ by $\mathcal{S}^*_{a,b}$ and analogously, $\mathcal{T}^*_{a,b} = \{(S, T) \in \mathcal{T}_{a,b} \mid (S, T) \text{ is a subset-team witness for } a \succ b\}$. It might be the case that $\mathcal{S}^*_{a,b} \cup \mathcal{T}^*_{a,b}$ is empty, in particular this holds when $b \succ a$. It is also possible that both $\mathcal{S}^*_{a,b} \cup \mathcal{T}^*_{a,b}$ and $\mathcal{S}^*_{b,a} \cup \mathcal{T}^*_{b,a}$ are empty, in which case we will show that the relation between players in a and b cannot be deduced. The following theorem implies that the other direction is also true.

Theorem 3.2. *Let $a, b \in [n]$. Then, $a \succ^* b$ if and only if $\mathcal{S}^*_{a,b} \cup \mathcal{T}^*_{a,b} \neq \emptyset$.*

Proof sketch. Assume that $\mathcal{S}^*_{a,b} \cup \mathcal{T}^*_{a,b} \neq \emptyset$. We show that $a \succ^* b$ by using SST, the fact that $\succ$ is a consistent strict total order, and an exhaustive case analysis. For the sake of illustration we present only one case here, namely, that $(S, S') \in \mathcal{S}^*_{a,b}$ and that both (1) $S \cup \{a\} \succ S' \cup \{b\}$ and (2) $S \cup \{b\} \succ S' \cup \{a\}$ hold. Assume for contradiction that $a \succ^* b$ does not hold. It thus

follows that there exists an order, $\succ' \in \mathcal{C}_{obs}$ for which $b \succ' a$ holds. Let $P'_{A,B}$ be the corresponding winning probabilities. Then, using consistency of $\succ'$ and (1) respectively, we get $S \cup \{b\} \succ' S \cup \{a\} \succ' S' \cup \{b\}$ and from SST $P'_{S\cup\{b\},S'\cup\{b\}} \geq P'_{S\cup\{a\},S'\cup\{b\}} > 1/2$. In addition, applying consistency again, it follows that $S \cup \{b\} \succ S' \cup \{b\} \succ S' \cup \{a\}$. Applying SST once more we get $P'_{S\cup\{b\},S'\cup\{a\}} \geq P'_{S\cup\{b\},S'\cup\{b\}} \geq P'_{S\cup\{a\},S'\cup\{b\}}$, a contradiction to $(S, S') \in \mathcal{S}^*_{a,b}$ (since this implies $P_{S\cup\{a\},S'\cup\{b\}} > P_{S\cup\{b\},S'\cup\{a\}}$).

For the other direction we start by defining $\mathcal{D}_a$ as the set of duels (A, B) such that $a \in A$. Moreover, we define a permutation π on the set of teams, which simply exchanges the players a and b when present. We then show that $a \succ b$ implies $P_{A,B} \geq P_{\pi(A),\pi(B)}$ for all $(A, B) \in \mathcal{D}_a$. Moreover, we show that $a \succ^* b$ implies that there exists $(A, B) \in \mathcal{D}_a$ with $P_{A,B} > P_{\pi(A),\pi(B)}$ as follows. Assume not. Then we show that the relation $\succ'$ defined by $A \succ' B$ iff $\pi(A) \succ' \pi(B)$ is included in C_{obs}. However, $a \succ^* b$ implies that for any $S \subseteq [n] \setminus \{a, b\}$ of size $k-1$ it holds that $S \cup \{a\} \succ^* S \cup \{b\}$ which implies (i) $S \cup \{a\} \succ S \cup \{b\}$ as well as (ii) $S \cup \{a\} \succ' S \cup \{b\}$. Applying the definitions of $\succ'$ and π, statement (ii) implies $S \cup \{b\} = \pi(S \cup \{a\}) \succ \pi(S \cup \{b\}) = S \cup \{a\}$ and hence yields a contradiction to (i). Finally, take some $(A, B) \in \mathcal{D}_a$ with $P_{A,B} > P_{\pi(A),\pi(B)}$. If $b \in B$, then $(A \setminus \{a\}, B \setminus \{b\}) \in \mathcal{S}^*_{a,b}$, otherwise $(A \setminus \{a\}, B) \in \mathcal{T}^*_{a,b}$. $\square$

For the sake of brevity, we introduce the set $\mathcal{X}_{a,b}$ which combines the pairs from $\mathcal{S}_{a,b}$ and $\mathcal{T}_{a,b}$ into a set of triples. Formally, $\mathcal{X}_{a,b} = \{(S, S', T) \mid (S, S') \in \mathcal{S}_{a,b}, (S, T) \in \mathcal{T}_{a,b}\}$. We say that (S, S', T) is a witness for $a \succ b$ if $(S, S') \in \mathcal{S}^*_{a,b}$ or $(S, T) \in \mathcal{T}^*_{a,b}$.

4 Stochastic Setting

In this section we focus on algorithms identifying, with high probability, the top-k team, which is in particular a Condorcet winning team. The main idea is to reduce the dueling teams setting to the classic dueling bandits setting, by which we refer to Yue et al. (2012). To this end we introduce our *gap parameter*, Δ, which intuitively captures how easy it is to prove the relationship between the top-k and the top-$(k+1)$ player. We start by defining, for any element of $\mathcal{X}_{a,b}$, a random variable $X_{a,b}(S, S', T)$ combining the outcomes of the four duels which help determining whether (S, S', T) is a witness for $a \succ^* b$. Formally,

$$X_{a,b}(S, S', T) = \big(\mathbb{1}[S \cup \{a\} > S' \cup \{b\}] - \mathbb{1}[S \cup \{b\} > S' \cup \{a\}] + \mathbb{1}[S \cup \{a\} > T] - \mathbb{1}[S \cup \{b\} > T]\big)/2.$$

Observe that, for every $(S, S', T) \in \mathcal{X}_{a,b}$, we have that $\mathbb{E}[X_{a,b}(S, S', T)] > 0$ if and only if (S, S', T) is a witness for $a \succ b$. Moreover, if $\mathbb{E}[X_{a,b}(S, S', T)] = 0$ for every $(S, S', T) \in \mathcal{X}_{a,b}$ then Theorem 3.2 implies that the pairwise relation between players a, b cannot be deduced. Building upon these random variables for fixed elements in $\mathcal{X}_{a,b}$, we define the random variable $X_{a,b}$ by picking a random triplet $(S, S', T) \in \mathcal{X}_{a,b}$ and returning a realization of $X_{a,b}(S, S', T)$. Combining with the probabilistic method, we obtain the following theorem, which then bring us to the definition of a gap parameter for this problem.

Theorem 4.1. *For every two players $a, b \in [n]$ it holds that $a \succ^* b$ if and only if $\mathbb{E}[X_{a,b}] > 0$.*

Gap parameter: We define our gap parameter by $\Delta := \mathbb{E}[X_{k,k+1}]$. In the following we show that our gap parameter does not just help us to distinguish between the top k and the top $k+1$ player, but also between other players in A^*_k and players from $[n] \setminus A^*_k$. To this end, we show in Lemma 4.2 that strong stochastic transitivity holds for $\mathbb{E}[X_{a,b}]$. For most elements $(S, S', T) \in \mathcal{X}_{a,b}$ it holds that $\mathbb{E}[X_{a,c}(\pi(S), \pi(S'), \pi(T))] \geq \mathbb{E}[X_{a,b}(S, S', T)]$ (and analogously for $X_{b,c}$), where π is a permutation exchanging b and c, but, surprisingly, this is not true in general. By constructing a charging scheme, we can still show that this holds in expectation over all elements of $\mathcal{X}_{a,b}$, and derive a strong stochastic transitivity for distinguishabilities w.r.t. the total order $\succ$ on the players.

Lemma 4.2. *For a triplet of players $a \succ b \succ c$ it holds that*

$$\mathbb{E}[X_{a,c}] \geq \max\{\mathbb{E}[X_{a,b}], \mathbb{E}[X_{b,c}]\}.$$

This also yields the following theorem, which paves the way for our reduction in what follows.

Theorem 4.3. *For any $a, b \in [n]$ such that $a \in A^*_k, b \notin A^*_k$ it holds that $\mathbb{E}[X_{a,b}] \geq \mathbb{E}[X_{k,k+1}] = \Delta$. Thus, if $\Delta > 0$ and a team A it holds that $\mathbb{E}[X_{a,b}] \geq \Delta$ for every $a \in A, b \in [n] \setminus A$, then $A = A^*$.*

The reduction: We now outline the gap-dependent algorithm for the stochastic setting. The results we have derived in Section 3 will allow us to deduce, with high probability, whether a distinguishability of a given pair of players is at least Δ, and if so determine which player is the better one. Intuitively, this is done by performing $\mathcal{O}(\frac{1}{\Delta^2})$ team duels.

We will use $\mathbb{E}[X_{a,b}]$ as a proxy for the distinguishability between two single players, a, b, taking advantage of the fact that if their relation is deducible, then $\mathbb{E}[X_{a,b}] \neq 0$ and in this case $\mathbb{E}[X_{a,b}] > 0$ iff $a \succ b$. Similar the to dueling bandits setting, even though $|\mathbb{E}[X_{a,b}]| < \Delta$ for some pairs of players, identifying A_k^* with high probability is possible.

Since we cannot directly sample $X_{a,b}$, we will instead sample uniformly at random a triplet of sets, (S, S', T) from $\mathcal{X}_{a,b}$. Using $(S, S')(\in \mathcal{S}_{a,b})$ and $(S, T)(\in \mathcal{T}_{a,b})$, we can then perform all the duels required for an unbiased sample of $X_{a,b}(S, S', T)$, which is by itself a sampling of $X_{a,b}$. Given any dueling teams instance, we define a dueling bandits instance as follows: for every two players $a, b \in [n]$,

$$P_{a,b} = 1/2 + \mathbb{E}[X_{a,b}]. \tag{1}$$

Where $P_{a,b}$ is the probability that a wins in a (singles) duel against b. Clearly, $1 - P_{a,b} = P_{b,a}$. In addition, Theorem 4.1 implies that a is better than b in this dueling bandits instance iff $a \succ^* b$. So whenever a dueling bandits algorithm is asking for a duel query, (a, b), we can make an independent unbiased sample of $X_{a,b}$ by returning a random sampling of $X_{a,b}(S, S', T) + 1/2$ (i.e., $X_{a,b}$ is a Bernoulli random variable with bias $\mathbb{E}[X_{a,b}(S, S', T)] + 1/2$.). In cases where the realization of $X_{a,b}$ is 0 and the algorithm does not consider ties as a valid duel feedback, we can randomly assign a duel winner. We formalize this idea in the sub-procedure *singlesDuel* (in the supplementary material), that simulates a duel for classical dueling bandits settings using team duels. Notice that, by Lemma 4.2, the probabilities $P_{a,b}$ defined in (1) satisfy SST with respect to the total order among the players induced by the ground truth order $\succ$. In addition, the feedback of each single player duel we perform is time-invariant, thus all the non-parametric assumptions for dueling bandits settings apply here. The reduction allows us to identify the top k players using any dueling bandit algorithm with the same goal that works for total order on arms that satisfy SST, and a gap between the top k and $k+1$ arms as assumptions. Formally,

Theorem 4.4. *Given any dueling teams instance with n and k (namely, $P_{A,B}$ for every two teams that hold strict total order, SST, and consistency), we have that the dueling bandit instance defined by (1) satisfies SST with respect to the ground truth order among players $\succ$ and for any two players $a \succ b$ it holds that $P_{a,b} \geq 1/2$. Moreover, $P_{k,k+1} = 1/2 + \Delta$.*

Using the above theorem we can use any dueling bandit algorithm for top k identification to solve our problem. Mohajer et al. (2017) provide an algorithm that returns the top k players with probability exceeding $1 - (\log n)^{-c_0}$ with sample complexity at most $c_1(n + k \log k)\frac{\max(\log\log n, \log k)}{\Delta^2_{k,k+1}}$ in expectation, where c_0 and c_1 are universal positive constants and $\Delta_{k,k+1}$ is the distinguishability between the k and the $k+1$ best players (see Algorithm 2 and Theorem 1 in Mohajer et al. (2017)).

Ren et al. (2020) show an algorithm that returns the top k players with probability at least $1 - \delta$ with sample complexity $\mathcal{O}(\sum_{i\in[n]}(\Delta_i^{-2}(\log(n/\delta) + \log\log \Delta_i^{-1}))$, where $\Delta_i = \mathbb{1}_{i \succ k+1} \cdot \Delta_{i,k+1} + \mathbb{1}_{k \succ i} \cdot \Delta_{k,i}$ and $k, k+1$ are the top k and the top $k+1$ players, respectively (see Algorithm 5 and Theorem 8 in Ren et al. (2020))[2].

These algorithms, together with Theorem 4.1 allow us to derive the following theorem.

Theorem 4.5. *There exists an algorithm that returns A_k^* with probability exceeding $1 - (\log n)^{-c_0}$ with sample complexity at most $c_1(n + k \log k)\frac{\max(\log\log n, \log k)}{\Delta^2}$ in expectation, where c_0 and c_1 are universal positive constants.*

In addition, there exists an algorithm that returns A_k^* with probability at least $1 - \delta$ with sample complexity $\mathcal{O}(\sum_{i\in[n]}(\Delta_i^{-2}(\log(n/\delta) + \log\log \Delta_i^{-1}))$, where $\Delta_i = \mathbb{1}_{i \succ k+1} \cdot \mathbb{E}[X]_{i,k+1} + \mathbb{1}_{k \succ i} \cdot \mathbb{E}[X]_{k,i}$ and i denotes the top i players, thus $\Delta_i \geq \Delta$ for every $i \in [n]$.

[2] We remark that Ren et al. (2020) also assume Stochastic triangle inequality which we do not, however it is only used to derive a lower bound.

5 Deterministic Setting

In the previous section we showed the existence of algorithms that identify the top k team with a number of duels that depends on Δ. But what if Δ is very small or even 0? One reason for that can be that all relevant probabilities are close to $1/2$. More precisely, $P_{\{k\}\cup S,\{k+1\}\cup S'}$, $P_{\{k+1\}\cup S,\{k\}\cup S'}$, $P_{\{k+1\}\cup S,T}$, and $P_{\{k\}\cup S,T}$ are very close to $1/2$ for all $(S,S',T) \in \mathcal{X}^*_{k,k+1}$. This might also occur in classic dueling bandits settings, when the target is to separate the top k players from the rest (e.g., Mohajer et al. (2017); Ren et al. (2020)). As a result, a gap between the top k and $k+1$ players is often a parameter of the sample complexity in such settings. For these cases, our approach presented in the stochastic section very much resembles the current literature.

The other, more interesting reason for Δ to be small is when there exist only a small number of witnesses. This is in particular the case when the probability matrix contains only few distinct values, as for example when feedback is deterministic, i.e., $P_{A,B} \in \{0,1\}$. Note that in this setting, $(S,T) \in \mathcal{T}_{a,b}$ is a witness if and only if $S \cup \{a\} \succ_{obs} T \succ_{obs} S \cup \{b\}$, and $(S,S') \in \mathcal{S}_{a,b}$ is a witness if and only if $S \cup \{a\} \succ_{obs} S' \cup \{b\}$ and $S' \cup \{a\} \succ_{obs} S \cup \{b\}$. This follows as for any other potential witness $(S,S',T) \in \mathcal{X}_{a,b}$ it holds that $\mathbb{E}[X_{a,b}(S,S',T)] = 0$. It is possible to come up with deterministic instances where up to $(2k-1)^2$ pairs do not have any witness to distinguish them. To overcome this issue, we design algorithms for the deterministic case that are independent of Δ within this section. In the appendix we show that these results can be extended to a slightly stochastic environment.

The limitation of the set of witnesses makes the problem of identifying a Condorcet winning team in the deterministic setting surprisingly nontrivial. For general total orders, a crucial difficulty lies in efficiently proving that a given team is indeed Condorcet winning. However, we are still able to get the following result:

Theorem 5.1. *For deterministic feedback, there exists an algorithm that performs $\mathcal{O}(kn\log(k) + k^2\log(k)2^{5k})$ duels and outputs a Condorcet winning team.*

For the natural special case of *additive total orders* we obtain a stronger result. A total order $\succ$ is *additive total*, if there exist values for the players denoted by $v(a), a \in [n]$ such that $A \succ B$ iff $\sum_{a\in A} v(a) > \sum_{b\in B} v(b)$. We present an algorithm that identifies a Condorcet winning team after polynomial many duels and also outputs a proof.

Theorem 5.2. *For deterministic feedback and additive total orders, there exists an algorithm that finds a Condorcet winning team within $\mathcal{O}(kn\log(k) + k^5)$ duels.*

Both algorithms rely on the same preprocessing procedure called *ReducePlayers* which reduces the number of players from n to $\mathcal{O}(k)$. At the heart of this procedure is a subroutine called *Uncover*. After describing *Uncover* and *ReducePlayers*, we prove Theorem 5.1. Towards proving Theorem 5.2, we introduce two more subroutines, namely *NewCut* and *Compare*, which are crucial for identifying and proving a Condorcet winning team within the smaller instance. Finally, Algorithm *CondorcetWinning* combines all components and proves Theorem 5.2.

The Uncover Subroutine Given two disjoint teams $A \succ B$, the *Uncover* subroutine finds a pair of players $a \in A$ and $b \in B$ and a subsets witness for their relation, i.e., an element from $\mathcal{S}^*_{a,b}$. To understand the idea of the subroutine, consider some arbitrary ordering of the elements in A and B, respectively, i.e., $A = \{a_1,\dots,a_k\}$ and $B = \{b_1,\dots,b_k\}$. Then, iteratively exchange the elements a_1 and b_1, a_2 and b_2, resulting in sets $A_0 = A, B_0 = B, A_1 = \{b_1,a_2,\dots,a_k\}, B_1 = \{a_1,b_2,\dots,b_k\}$, $A_2 = \{b_1,b_2,a_3,\dots,a_k\}$, and so on. Since $A_0 \succ B_0$ but $A_0 = B_k \succ A_k = B_0$ holds, there needs to be some earliest point in time $i \le k$ for which $B_i \succ A_i$ is true. This implies $a_i \succ b_i$ as $(\{a_1,\dots a_{i-1},b_{i+1},\dots b_k\},\{b_1,\dots,b_{i-1},a_{i+1},\dots,a_k\})$ is a witness for this relation.

While the above sketched subroutine is simple, it performs k duels in the worst case. We refine this idea by a binary search approach, decreasing the number of duels to $\log(k)$. In the appendix we give a slightly stronger version of Lemma 5.3, which allows us to divide A and B into two subsets each. Under some requirements, we can then control from which of the subsets the pairwise relation is revealed.

Lemma 5.3. *Let A and B be two disjoint teams with $A \succ B$. After performing $\mathcal{O}(\log(k))$ duels,* Uncover *returns (a,b) with $a \in A$, $b \in B$ and $(S,S') \in \mathcal{S}^*_{a,b}$.*

Reducing the Number of Players to $\mathcal{O}(k)$ The fact that we can eliminate a subset of the players and still find (and prove) a Condorcet winning team is due to the following observation.

Observation 5.4. *Let $X \subseteq [n]$ such that $A^*_{2k} \subseteq X$. Let $\hat{A} \subseteq X$ be a team such that $\hat{A} \succ A$ for all teams $A \subseteq X \setminus \hat{A}$. Then, $\hat{A}$ is a Condorcet winning team.*

The procedure *ReducePlayers* reduces the set of players $[n]$ to some subset $X \subseteq [n]$ guaranteeing that $A^*_{2k} \subseteq X$ and $|X| < 6k$. The algorithm maintains a dominance graph $D = (V, E)$ on the set of players. More precisely, the nodes of D are the players, i.e., $V = [n]$, and there exists an arc from node a to node b if the algorithm has proven that $a \succ b$. The set $V_{<2k}$ is the subset of the players having an indegree smaller than $2k$ in D. The high level idea of the algorithm is the following: It starts with the empty dominance graph $D = ([n], \emptyset)$. The algorithm then iteratively identifies pairwise relations of the players with help of *Uncover* and adds the corresponding arcs to the graph. By adding more and more arcs to D, the set of nodes $V_{<2k}$ shrinks more and more while $A^*_{<2k} \subseteq V_{<2k}$ is always guaranteed. At some point, the algorithm cannot identify any more pairwise relations and returns $V_{<2k}$. How does the algorithm identify pairwise relations? At any point it tries to find a matching between $2k$ players, say $\{(a_1, b_1), \ldots, (a_k, b_k)\}$ with the constraint that, for all $i \in [k]$, none of the arcs (a_i, b_i) or (b_i, a_i) is present within the graph D yet. The algorithm ends when it cannot find such a matching anymore. We show that this only happens after $|V_{<2k}| < 6k$.

Lemma 5.5. *Given the set of players $[n]$,* ReducePlayers *returns $X \subseteq [n]$ with $|X| \leq 6k - 2$ and $A^*_{2k} \subseteq X$.* ReducePlayers *performs $\mathcal{O}(nk \log(k))$ duels and runs in time $\mathcal{O}(n^2 k^2)$.*

Proof Sketch (of Theorem 5.1). Let D be the dominance graph at the end of *ReducePlayers*. Then, the learner selects a k-sized subset of $V_{<2k}$, call it $\hat{A}$, with the property that there is no arc from some node in $V_{<2k} \setminus \hat{A}$ towards some node in $\hat{A}$. Then, the learner tests $\hat{A}$ against all possible teams containing players from $V_{<2k} \setminus \hat{A}$, which are $\mathcal{O}(2^{5k})$ many. If $\hat{A}$ wins all of these duels, then $\hat{A}$ is a Condorcet winning team by Observation 5.4. However, if there exists $A \succ \hat{A}$, then, by the choice of $\hat{A}$, there does not exist any arc from A towards $\hat{A}$. Hence, by calling the subroutine *Uncover* for two arbitrary orderings of A and $\hat{A}$, the learner will identify one additional arc. This procedure can be repeated $\mathcal{O}(k^2)$ many times and hence shows Theorem 5.1. □

Subroutines NewCut and Compare The *NewCut* subroutine takes as input a subset of the players $X \subseteq [n]$, a pair $a, b \in X$, and a witness proving that $a \succ b$, i.e., $(S, T) \in \mathcal{S}^*_{ab} \cup \mathcal{T}^*_{ab}$. That means, T can be either of size $k - 1$ or k, and S and T are not required to be subsets of X. The subroutine outputs a partition of X into two non-empty sets U and L with $U \triangleright L$, which is short-hand notation for $u \succ \ell$ for any $u \in U$ and $\ell \in L$. The main idea of the algorithm is to make use of the transitivity among witnesses which we established in Lemma 4.2. More precisely, whenever the algorithm has found a witness showing that some player in X is better than b, it stores this witness in a list $\mathcal{W}$. Iteratively applying the transfer function used in the proof of Lemma 4.2 to all witnesses in $\mathcal{W}$ and elements in X makes sure that after termination of the the algorithm $U \triangleright L$ holds, where U is defined by all players u for which the algorithm found a witness for $u \succ b$ and $L = X \setminus U$.

Lemma 5.6. *Let $X \subseteq [n]$, $a, b \in X$ and $(S, T) \in \mathcal{S}^*_{a,b} \cup \mathcal{T}^*_{a,b}$. Then,* $\mathrm{NewCut}(X, (a, b), (S, T))$ *returns a partition of X into U and L such that $U \triangleright L$, $a \in U$ and $b \in L$. The number of duels performed by* NewCut *and its running time can be bounded by $\mathcal{O}(|X|^2)$.*

From now on we assume additive linear orders. The *compare* subroutine is crucial for obtaining upper bounds for differences of values of subsets of players. It is used in the following situation. Let (a, b) be a pair of players and $(S, S') \in \mathcal{S}^*_{a,b}$ be a witness for $a \succ b$. Then, it can be easily shown that $v(a) - v(b) > |v(S) - v(S')|$. We will be interested in the question whether a similar relation holds for two subsets of S and S', namely, $C \subseteq S$ and $D \subseteq S'$ of equal size. The *compare* subroutine checks whether such a relation holds by performing two additional duels. If it returns *True*, then $v(a) - v(b) > |v(C) - v(D)|$. Otherwise, there can be found a pair $c \in C$ and $d \in D$ and a witness for their relation by one call to the *Uncover* subroutine. This observation is formalized Lemma 5.7.

Lemma 5.7. *Let $a \succ b$ be two players, $(S, S') \in \mathcal{S}^*_{a,b}$ and $C \subseteq S, D \subseteq S'$ with $|C| = |D|$. If* $\mathrm{Compare}((a, b), (S, S'), (C, D))$ *returns* True*, then $v(a) - v(b) > |v(C) - v(D)|$. Otherwise, one call to* Uncover *returns $c \in C$ and $d \in D$ together with a witness for their relation.*

Algorithm CondorcetWinning The algorithm maintains a partition of the players into a weak ordering, i.e., $\mathcal{T} = \{T_1, \ldots, T_\ell\}$ with $T_1 \triangleright T_2 \triangleright \cdots \triangleright T_\ell$. We introduce the short-hand notation

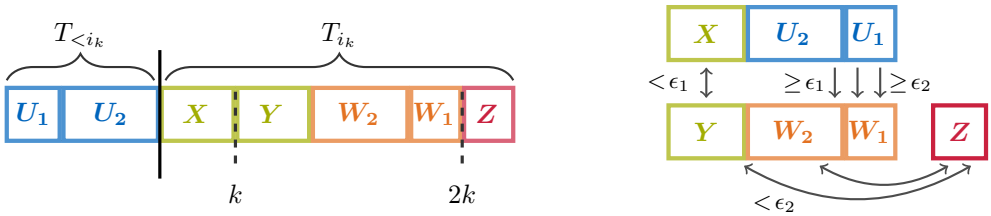

Figure 1: Illustration of the proof technique of algorithm *CondorcetWinning1*. In the left illustration, the solid black line indicates that all players left to it were proven to be better than all players right to it. The dashed line marked with "k" indicates that the sets to its left contain k players in total. However, this line does not indicate proven relations, e.g., a player from X can be better than a player from Y. The right figure illustrates the proof for $X \cup U_1 \cup U_2$ being Condorcet winning.

$T_{\leq j} = \bigcup_{m\in[j]} T_m$ and $T_{<j} = \bigcup_{m\in[j-1]} T_m$. After the application of the preprocessing procedure *ReducePlayers*, this partition consists of one set, namely $\mathcal{T} = \{T_1\}$, where $|T_1| \in \mathcal{O}(k)$ and $A^*_{2k} \subseteq T_1$. At any point in the execution of the algorithm, we are especially interested in two indices, namely $i_k \in [\ell]$ such that $|T_{<i_k}| < k < |T_{\leq i_k}|$ and similarly $i_{2k} \in [\ell]$ such that $|T_{<i_{2k}}| < 2k < |T_{\leq i_{2k}}|$.[3] Observe that all players from $T_{<i_k}$ are guaranteed to be among the top-k players. On the other hand, among the players from T_{i_k} some belong to A^*_k and others do not. The main idea of the algorithm is then the following: Take a prefix of $\mathcal{T}$ of size k, i.e., this team contains the set of players $T_{<i_k}$ and is a subset of the players in $T_{\leq i_k}$, and either prove that this prefix is a Condorcet Winning team, or refine the partition $\mathcal{T}$ and repeat the process. We provide two different algorithms, namely *CondorcetWinning1* for the case $i_k = i_{2k}$ and *CondorcetWinning2* when $i_k \neq i_{2k}$. Unsurprisingly, the latter case requires a strictly less sophisticated approach, which is why we focus on *CondorcetWinning1* in the following.

The algorithm starts by partitioning the set $T_{<i_k}$ into two sets U_1 and U_2, where U_1 is a prefix of $T_{<i_k}$ of size $|T_{\leq i_k}| - 2k$. It partitions the set T_{i_k} into five sets X, Y, W_1, W_2, and Z. In particular it is known that $(U_1 \cup U_2) \triangleright (X \cup Y \cup W_1 \cup W_2 \cup Z)$ but no relation among any pair in T_{i_k} is known. Regarding the sizes of the sets it holds that $|U_i| = |W_i|$ for $i \in \{1, 2\}$, $|X| = |Y| = k - |U_1| - |U_2|$ and $|U_1| = |Z|$. The main aim of the algorithm will be to define $0 < \epsilon_1 < \epsilon_2$ and prove that the following statements are true:

(i) $|v(X) - v(Y)| < \epsilon_1$

(ii) $|v(a) - v(b)| < \epsilon_2$ for all $a \in Y \cup W_1 \cup W_2$ and $b \in Z$, and

(iii) there exist $u_1, \ldots, u_{|Z|+1} \in U_1 \cup U_2$ as well as $w_1, \ldots, w_{|Z|+1} \in W_1 \cup W_2$ such that

  (a) $v(u_1) - v(w_1) \geq \epsilon_1$ and

  (b) $v(u_i) - v(w_i) \geq \epsilon_2$ for all $i \in \{2, \ldots, |Z| + 1\}$.

With these three statements we can show that $U_1 \cup U_2 \cup X$ is a Condorcet winning team. More precisely, one can show that $v(U_1 \cup U_2 \cup X) - v(W_1 \cup W_2 \cup Y) > |Z| \cdot \epsilon_2$ and $v(W_1 \cup W_2 \cup Y) - v(B^*) > -|Z| \cdot \epsilon_2$, where B^* is the best response towards $U_1 \cup U_2 \cup X$. See Figure 1 for an illustration of the argument.

It remains to sketch how the algorithm defines ϵ_1, ϵ_2 and proves $(i) - (iii)$. For simplicity assume $U_1 \triangleright U_2$. The algorithm then attempts to do the following steps: (1) Find a witness for players $\bar{u} \in U_2$ and $\bar{w} \in W_2$, using *Uncover*. (2) Use *Compare*, to prove that $|v(X) - v(Y)| < v(\bar{u}) - v(\bar{w})$ and $|v(a) - v(b)| < v(\bar{u}) - v(\bar{w})$ holds for all players $a \in W_1 \cup W_2 \cup Y$ and $b \in Z$. (3) Repeat step (2) by replacing $\bar{w}$ with any player of W_1. If one of the steps (1)-(3) fails, we show that the partition $\mathcal{T}$ can be refined. Otherwise, we show that $(i) - (iii)$ hold for $\epsilon_1 = v(\bar{u}) - v(w^*_1)$ and $\epsilon_2 = v(\bar{u}) - v(w^*_2)$, where w^*_1 and w^*_2 are the best and second best players from $W_1 \cup \{\bar{w}\}$, respectively. The following Lemma concludes the proof sketch of Theorem 5.2.

Lemma 5.8. *After performing $\mathcal{O}(k^5)$ many duels,* CondorcetWinning1 *has identified a Condorcet winning team.* CondorcetWinning2 *identifies a Condorcet winning team after $\mathcal{O}(k^2 \log(k))$ duels.*

[3] In case one of these indices does not exist, this implies that we have either identified the set A^*_k or A^*_{2k}. In the first case we have found a Condorcet winning team and in the second case Observation 5.4 implies that we can find one by performing one additional duel. For the sake of brevity we disregard this case from now on.

References

Awerbuch, B., & Kleinberg, R. (2008). Online linear optimization and adaptive routing. *Journal of Computer and System Sciences*, *74*(1), 97–114.

Brost, B., Seldin, Y., Cox, I. J., & Lioma, C. (2016). Multi-dueling bandits and their application to online ranker evaluation. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM)*, (pp. 2161–2166).

Bubeck, S., Munos, R., & Stoltz, G. (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, *412*(19), 1832–1852.

Bubeck, S., Wang, T., & Viswanathan, N. (2013). Multiple identifications in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*.

Busa-Fekete, R., Hüllermeier, E., & Mesaoudi-Paul, A. E. (2018). Preference-based online learning with dueling bandits: A survey. *CoRR*.

Cesa-Bianchi, N., & Lugosi, G. (2012). Combinatorial bandits. *Journal of Computer and System Sciences*, *78*(5), 1404–1422.

Chen, L., Li, J., & Qiao, M. (2017). Towards instance optimal bounds for best arm identification. In *Proceedings of the 30th Conference on Learning Theory (COLT)*, (pp. 535–592). PMLR.

Chen, S., Lin, T., King, I., Lyu, M. R., & Chen, W. (2014). Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*.

Even-Dar, E., Mannor, S., & Mansour, Y. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, *7*, 1079–1105.

Grossman, H. D. (1945). The twelve-coin problem. *Scripta Math*.

Guy, R. K., & Nowakowski, R. J. (1995). Coin-weighing problems. *The American Mathematical Monthly*, *102*(2), 164–167.

Kalyanakrishnan, S., & Stone, P. (2010). Efficient selection of multiple bandit arms: Theory and practice. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*.

Kaufmann, E., Cappé, O., & Garivier, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *Journal of MachineLearning Research*, *17*, 1:1–1:42.

Lattimore, T., & Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press.

Mohajer, S., Suh, C., & Elmahdy, A. (2017). Active learning for top-k rank aggregation from noisy comparisons. In D. Precup, & Y. W. Teh (Eds.) *Proceedings of the 34th International Conference on Machine Learning (ICML)*, vol. 70, (pp. 2488–2497).

Pelc, A. (2002). Searching games with errors - fifty years of coping with liars. *Theoretical Computer Science*, *270*(1-2), 71–109.

Rejwan, I., & Mansour, Y. (2020). Top-k combinatorial bandits with full-bandit feedback. In A. Kontorovich, & G. Neu (Eds.) *Algorithmic Learning Theory (ALT)*, (pp. 752–776). PMLR.

Ren, W., Liu, J., & Shroff, N. (2020). The sample complexity of best-k items selection from pairwise comparisons. In H. D. III, & A. Singh (Eds.) *Proceedings of the 37th International Conference on Machine Learning*, vol. 119 of *Proceedings of Machine Learning Research*, (pp. 8051–8072). PMLR.

Ren, W., Liu, J., & Shroff, N. B. (2018). PAC ranking from pairwise and listwise queries: Lower bounds and upper bounds. Tech. rep., http://arxiv.org/abs/1806.02970.

Saha, A., & Gopalan, A. (2018). Battle of bandits. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Slivkins, A. (2019). Introduction to multi-armed bandits. *Foundations and Trends in Machine Learning*, *12*(1-2), 1–286.

Sui, Y., Zhuang, V., Burdick, J., & Yue, Y. (2017). Multi-dueling bandits with dependent arms. In *Proceedings of 33rd the Conference on Uncertainty in Artificial Intelligence (UAI)*.

Yue, Y., Broder, J., Kleinberg, R., & Joachims, T. (2012). The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, *78*(5), 1538 – 1556.

Zhou, Y., Chen, X., & Li, J. (2014). Optimal pac multiple arm identification with applications to crowdsourcing. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] This is a theoretical work and we do not participate any negative social impact.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] In the main part of the paper we provide proof sketches. In the appendix (uploaded as supplementary material) we provide complete proofs for all of our technical results.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]