OPT2025: 17th Annual Workshop on Optimization for Machine Learning

Efficient Algorithms for Combinatorial-Bandits with Monotonicity

Aniket Anil Wagde AWAGDE2@UIC.EDU
University of Illinois Chicago

Aadirupa Saha AADIRUPA@UIC.EDU
University of Illinois Chicago

Abstract

In this paper, we study the problem of combinatorial bandits with stochastic full-bandit feedback, where the feedback is aggregated using an unknown operator. Unlike traditional combinatorial bandits, where the feedback is often assumed to be linear, we consider a setting with non-linear reward functions determined by any monotonic operator, which generalize various aggregation methods, including maximum, minimum, and k-order statistics. The challenge arises from the need to identify the top-k arms efficiently without explicit knowledge of the underlying aggregation operator. We propose novel algorithms that leverage sub-Gaussian noise assumptions and gap-based analysis to provide strong theoretical guarantees on the sample complexity for identifying the top-k arms with high confidence. We apply a modified version of an algorithm designed for pairwise preferences. Our results extend to various scenarios, including any monotonic function, assuming that the noise can be assumed to be sub-Gaussian, showing that our methods achieve strong performance in terms of both time and sample efficiency.

1. Introduction

Problem Motivation Multi-armed bandits (MAB) is a well-studied machine learning problem [8, 17, 23] where an agent makes sequential decisions by choosing from a finite set of options called "arms" (referencing casino gambling machines). Each selection provides feedback, typically in the form of numerical rewards, determined by the arm's mechanism. Initially knowing nothing about these mechanisms, the learner develops strategies over time. Two key factors shape the learning strategy: assumptions about the nature of feedback mechanisms and the characteristics of the learning task itself.

Traditional feedback mechanisms follow probabilistic laws [16, 22], with learning strategies acquiring knowledge about distributional properties to define optimal arms, i.e., those with the highest means. Non-stochastic settings avoid probabilistic assumptions [4, 6, 13], so are more generally applicable. Many applications focus on finding (approximately) correct answers within reasonable samples. This problem is called pure exploration, and there are two variants: fixed confidence (minimize rounds while guaranteeing confidence) and fixed budget (maximize confidence with limited rounds) [11]. The approach in this paper optimizes for fixed confidence.

Limitations of Classical MAB It falls short for complex practical situations, leading to extensions like side information [1, 3] and infinite arms [7, 18]. Another generalization allows arm sets: combinatorial bandits for numerical feedback [9] and preference-based bandits for relative feedback [5]

Combinatorial bandits are further distinguished with respect to the type of feedback between semi-bandit feedback, where feedback of each single arm in the selected set is observed (see [14] for a recent overview), and full bandit feedback, where only some aggregated value of the individual numerical feedback (rewards) is observed.[19]

In this paper, we consider the combinatorial full-bandits setting where the reward is drawn from an unknown aggregation function. This approach draws an interesting parallel between combinatorial bandits and preference-based bandits. We adapt an algorithm designed for preference-based bandits and use it to solve combinatorial bandits without affecting the time complexity of the solution. We also do not require the assumption of linear reward functions.

Applications Choosing top-k subsets with nonlinear, unclear payoffs impacts many real-world decisions. Applications include personalized recommendation systems for e-commerce platforms (Amazon, eBay) and streaming services (Netflix, Spotify), where selecting optimal products, movies, or songs involves nonlinear preference aggregation. Healthcare applications require selecting treatment combinations, medications, or diagnostic tests, where strict protocols assign varying importance to the results. General combinatorial bandit models can learn optimal intervention subsets without knowing the exact payoffs. This research advances the theoretical understanding of multi-armed bandits while providing versatile optimization tools for industries where ranking and aggregation are as important as selection.

Related Work Here, we describe some prior attempts to solve top-k combinatorial multi-armed bandits. We also have a table comparing our algorithm to others in Appendix A.

The "CSAR" algorithm [19] solves our exact problem; it assumes a linear reward structure and approximates the individual arm rewards to find the top-k set. It claims a strong sample complexity; however, its structure is highly limiting, and our approach matches its sample complexity.

The "DART"[2] algorithm also solves the problem; however, it requires the reward function to be bi-Lipschitz continuous, and we show later in Appendix A.2 that it has an infinite worst-case regret bound, rendering its theoretical guarantees trivial. Our approach doesn't need these limiting assumptions and has strong theoretical guarantees.

The "SQAM"[15] algorithm approaches the problem in two stages where it first gathers data, then attempts to compute approximations of the best arms, this a static algorithm. This suffers the same problem of assuming linear aggregation of rewards from individual arms, which makes it not as general as ours despite its strong time complexity.

Our approach of comparing pairs of arms through randomizing across the rest of the sets is what allowed us to use pairwise preference algorithms to solve combinatorial bandit problems. This alternate approach to multi-armed bandit problems might yield other benefits on further exploration, including algorithms with strong regret bounds as well as sample complexities without strong assumptions.

Our Contributions

- We introduce the problem setting `MonCMAB` and it is one of the first attempts to handle combinatorial multi-armed bandits with general subset rewards in Section 2.
- We created a method to reduce the problem of combinatorial full-feedback multi-armed bandits into relative feedback multi-armed bandits described in Section 3 and analyzed in Appendix A.6.

- Created an algorithm to tackle combinatorial unified-feedback multi-armed bandits with very few assumptions on the reward aggregation function, explained further in Section 4 and Section 5.
- Our solutions' efficiency is comparable to less general solutions, as further analyzed Section 5 and discussed in Remark 6.

2. Problem Formulation

Consider a set of n arms, denoted $[n] = \{1, \ldots, n\}$. We assume that each arm i is associated with a value μ_i that is the expected reward of a single arm i, $\mu_i \in [0, 1]$ $\forall i \in [n]$. Note that a single arm can't be sampled individually. In each round $t \in [T]$, the learner selects a subset $S_t \subseteq [n]$ of size $|S_t| = k$ and observes a reward $r_t(S_t) := F(S_t) + \eta_t$, where η_t is zero-mean sub-Gaussian noise, with scale parameter σ and $F(S_t)$ is a non-stochastic function. Note that $\mathbb{E}_{\eta_t}[r(S_t)] = F(S_t)$.

Assumption 1 *We assume that the reward $F()$ has the property of Combinatorial Monotonicity. A function has Combinatorial Monotonicity, if whenever $\mu_i > \mu_j$, then $F(S \cup i) > F(S \cup j)$ where $i, j \in [n]$ and , the elements in $\forall S \subseteq ([n] \backslash \{i, j\}), |S| = k - 1$.*

Assumption 2 (sub-Gaussian noise) *The reward $r_t(S)$ is sub-Gaussian with mean $F(S)$ and scale parameter $\sigma = 1$ for all $S \subseteq [n], |S| = k$. A random variable X with mean value $\mu = \mathbb{E}[X]$ is defined to be sub-Gaussian if $\exists$ σ such that $\mathbb{E}[\exp(\lambda(X - \mu))] \leq \exp(\sigma^2\lambda^2/2)\, \forall \lambda \in \mathbb{R}$ [24].*

Definition 1 (Optimal Set of k-arms) *The optimal set of k-arms denoted by S^*, where $S^* \subseteq [n], |S^*| = k$, is the set with the maximum expected reward, defined by $S^* = \underset{S', S' \subseteq [n], |S'|=k}{\arg\max} F(S')$.*

Theorem 2 (Optimality of Top-k Arms) *The Optimal Set of k-arms are the k-arms with the highest individual μ_i values, where $i \in [n]$.*

The proof of theorem 2 is available in the Supplementary material, Appendix A.7.

Problem Objective: The objective of the problem is to find the subset $\hat{S}$ of k-arms such that:

$$\mathbb{P}\big(F(S^*) - F(\hat{S}) > \epsilon\big) \leq \delta,$$

for any $\epsilon \in [0, \inf), \delta \in (0, 1]$, where S^* is the Optimal Set of k-arms as from Definition 1.

We will refer to the problem as the `MonCMAB`(k) problem henceforth. The algorithm we use to solve this problem is adapted from a paper by Ren et al. [20]. We are able to solve this in the special case of $\epsilon = 0$. The original algorithm takes five assumptions that our problem setting inherently satisfies, as we have shown in Appendix A.1.

3. Connecting Learning from Combinatorial Feedback to Learning from Relative Feedback

Here, we describe another framework of the multi-armed bandit, where a learning algorithm can query two arms, and feedback is given as a noisy preference of one arm over another. This setting is known as relative feedback. The problem of finding the top-k arms in this setting has been shown to be efficiently solved [20].

The main novelty of our research paper is that we have shown that the problem of top-k in monotonic combinatorial bandits with non-individualized feedback can be reduced to solving for top-k using relative feedback. This result is also highly general, requiring no assumptions outside Combinatorial Monotonicity and sub-Gaussian noise.

3.1. Preliminaries

For $i, j \in [n]$ and $S \subset [n] \setminus \{i, j\}$ sampled uniformly at random, with $|S| = k - 1$, we define the Relative-Strength on S as

$$\Delta(i, j) := \mathbb{E}_{S,\eta}[r(S \cup \{i\}) - r(S \cup \{j\})] = \mathbb{E}_S[F(S \cup \{i\}) - F(S \cup \{j\})] \tag{1}$$

Note that $\Delta(i, j) = -\Delta(j, i)$ and $\forall i \in [n]$, $\Delta(i, i) = 0$. Relative-Strength behaves analogously to relative feedback. If $\Delta(i, j) > 0$, then arm i is preferred to arm j. However, it is defined in terms of combinatorial feedback.

We also define a notion of distance from the top-k decision boundary, taken from [20]:

$$\Delta_i = \max(\Delta(i, a_{k+1}), \Delta(a_k, i)) \tag{2}$$

Note that $\Delta_{a_k} = \Delta_{a_{k+1}}$.

Definition 3 (Individualized (ϵ,k)-optimal subset [20]) *For a set S , given $k \leq |S|$, and $\epsilon \in \mathbb{R}_+$, a set $\hat{S} \subset S$ is said to be an (ϵ, k)-optimal subset of S if $|\hat{S}| = k$ and $\Delta(i, j) \geq -\epsilon$, $\forall i, j$, such that, $i \in \hat{S}, j \notin \hat{S}$.*

Note that our approach is not addressing the problem of finding an Individualized (ϵ,k)-optimal subset . Instead, we aim to find the top-k arms from Definition 1.

3.2. Bounding Error of Estimates of the Relative-Strength between Arms

With perfect knowledge of the Relative-Strength between all the arms, it would be trivial to choose the top-k subset of arms. However, we will need to rely on estimates of the Relative-Strength between arms calculated from samples denoted by $\hat{\Delta}_t(i, j)$, specific to timestep t for any two arms i and j, defined below. Here $S_t \subseteq [n] \setminus \{i, j\}$ is sampled uniformly at random, with $|S_t| = k - 1$. Also, $\bar{\Delta}_t(i, j)$ denotes the average of the samples obtained and serves as our estimate of Relative-Strength between i and j. In Theorem 4 we bound the errors of our estimated Relative-Strength $\bar{\Delta}_t(i, j)$.

$$\hat{\Delta}_t(i, j) := r_t(S_t \cup \{i\}) - r_t(S_t \cup \{j\}) \tag{3}$$

$$\bar{\Delta}_t(i, j) = \sum_{s=1}^{t} \frac{\hat{\Delta}_s(i, j)}{t} \tag{4}$$

Theorem 4 (Error Bounds on Estimates of Relative-Strength) *A bound on the error of Relative-Strength estimates can be placed as:*

$$\mathbb{P}(|\bar{\Delta}_t(i, j) - \Delta(i, j)| \geq \epsilon) \leq 2 \exp(-\epsilon^2 t/8),$$

where $\epsilon \in \mathbb{R}_+$.

Proof Sketch: For $S_t \subseteq [n] \setminus \{i, j\}$ sampled uniformly at random, with $|S_t| = k - 1$. It can shown that $\hat{\Delta}_t(i, j) \sim$ subGaussian($\sigma = 2$) and is centered around $\Delta(i, j)$ as proved in appendix A.3. We can then use Hoeffding's inequality for sub-gaussian variables [24] as described in appendix A.4 to create bounds on error between the estimates and the true Relative-Strength between the arms.

The detailed proof can be found in appendix A.6. Theorem 4 allows us to gather and quantify preference between two arms despite only having access to combinatorial rewards.

In the next section, we use these insights to design a new highly general and computationally efficient approach to solving the `MonCMAB` problem using an algorithm intended for pairwise preference feedback.

4. Algorithm

In this section, we describe "Sequential-Elimination-Exact-k-Selection-with-Isolated-Compare" (Algorithm 4), the algorithm made by adapting the "SEEKS" algorithm from Ren et al [20]. We start with an overview of the subroutines and their purposes, and we explain the intuition of the algorithm.

Algorithm 1 "Isolated-Compare (IC) $(i, j, \epsilon, s_u, s_d, \delta, S_{up}, S_{mid}, S_{down})$" in appendix A.5 is inspired by the subroutine "Distribute-Item (DI)" from [20]. This algorithm is assigned two arms i and j, and classifies arm i based on ground truth $\Delta(i, j)$. Arm i is placed into the bins $S_{up}, S_{mid}, S_{down}$ based on its relative value to arm j. The threshold values for the bins, s_u and s_d, are for up and down, respectively. IC is also allowed an ϵ margin of error. IC compares rewards from sampling subsets that include i to rewards from subsets that include j. Effectively, this algorithm takes the exact same inputs and produces the same output as DI from [20].

Algorithm 2 "Epsilon-Quick-Select-with-Isolated-Compare (EQS-IC)(S, k, ϵ, δ)" described in appendix A.5, is highly similar to the subroutine "Epsilon-Quick-Select" from [20]. The difference being that EQS-IC uses IC instead of DI. EQS-IC operates very similarly to quickselect [12]. EQS-IC recursively pivots on a randomly chosen element until it is able to find k elements that are within ϵ error margin of the top k^{th} element, and then returns those elements. As a result EQS-IC returns an Individualized (ϵ,k)-optimal subset, it's performance is described in Appendix A.6.

Algorithm 3 "Tournament-k-Selection-with-Isolated-Compare (TKS-IC) $([n], k, \epsilon, \delta)$" is similar to the subroutine "Tournament-k-Selection" from [20]. TKS-IC operates over many rounds. In each round, it splits all the arms into batches of size $2k$ and uses EQS-IC on each batch to find an Individualized (ϵ,k)-optimal subset. It repeats this process till only k arms are left. As a result TKS-IC returns an Individualized (ϵ,k)-optimal subset with a significantly better time complexity described in Appendix A.6. Note that TKS-IC2, is nearly the same algorithm, except that it finds the worst arms. TKS-IC2 is used as a subroutine by Algorithm 4.

Algorithm 4 is similar to the algorithm "Sequential-Elimination-Exact-k-Selection" from [20], appropriately named "Sequential-Elimination-Exact-k-Selection-with-Isolated-Compare (SEEKS-IC) $([n], k, \delta)$". SEEKS-IC uses TKS-IC andTKS-IC2 to find an arm that is very close to the top k^{th} arm. It then prunes all arms below this and forces an error margin so it doesn't prune any of the top-k arms. It then repeats this process with smaller error margins till only k arms are left. SEEKS-IC is the algorithm we use since it picks the top k arms with an error of zero. Its performance is given in Theorem 5 and is proven in Appendix A.6.

Algorithm 4 is the final algorithm that we propose to find the top-k arms with combinatorial feedback. The proofs for these algorithms can be found in Appendix A.6, many of which are identical to the work in [20].

5. Theoretical Analysis

In this section, we discuss the performance of Algorithm 4 and the other subroutines that comprise of the SEEKS-IC [20] algorithm. IC, performs as a blackbox identically to DI from [20]. Also, the rest of Algorithms 2 to 4 work nearly identically to their counterparts in [20], with some small differences that don't affect the computational or the time complexity. In conclusion the "SEEKS-IC" algorithm will terminate after $O\left(\sum_{i\in[n]}\left[\Delta_i^{-2}\left(\log(n/\delta)+\log\log\Delta_i^{-1}\right)\right]\right)$ samples have been drawn and return the top-k subset of arms in our `MonCMAB` problem with a δ probability of failure.

We have included the algorithms and their corresponding proofs in Appendix A.5 and Appendix A.6 respectively.

5.1. SEEKS-IC Sample Complexity Analysis

Here, we describe the sample and computational complexity that our method requires.

Theorem 5 (Theoretical Performance of SEEKS-IC (Algorithm 4)) *With probability at least $1-\delta$, SEEKS-IC terminates after $O\left(\sum_{i\in[n]}\left[\Delta_i^{-2}\left(\log(n/\delta)+\log\log\Delta_i^{-1}\right)\right]\right)$ number of comparisons in expectation, and returns the best-k items.*

Remark 6 *Our sample complexity is particularly notable as it matches other state of the art combinatorial multi-armed bandit problems with aggregated feedback while taking far fewer assumptions. Here are some examples:*

- *"CSAR" algorithm, [19] shares our sample complexity however, requires the reward aggregation function F to be linear, significantly reducing it's generality.*
- *"DART" algorithm, [2] addresses non-linear aggregation functions and states a regret bound, however, it takes a bi-Lipschitz continuity assumption and has an infinite worst case time complexity as elaborated upon in Appendix A.2.*

6. Conclusion

We have introduced a method for tackling the problem of top-k arm identification in combinatorial bandit settings with stochastic full-bandit feedback where reward is aggregated through an unknown combinatorial monotonic function, with added sub-Gaussian noise. This works through imitating pairwise feedback by comparing rewards from actions while selecting certain arms and randomizing over other arms. This idea was incorporated into the algorithm from [20] to find the top-k arm subset. This allows us to find the top-k arms with a time complexity of $O\left(\sum_{i\in[n]}\left[\Delta_i^{-2}\left(\log(n/\delta)+\log\log\Delta_i^{-1}\right)\right]\right)$.

Future Work: We plan to find ways to extend this to a more general solution that can allow for varying tolerances of error while taking as few assumptions as possible. We also want to explore algorithms that also have strong regret bounds. We plan to find applications that could benefit from a significantly more efficient top-k search. We also plan to extend this work to contextual combinatorial multi-armed bandits and see if a similar approach is viable.

References

[1] Naoki Abe and Philip M Long. Associative reinforcement learning using linear probabilistic concepts. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 3–11, 1999.

[2] Mridul Agarwal, Vaneet Aggarwal, Abhishek Kumar Umrawal, and Chris Quinn. Dart: Adaptive accept reject algorithm for non-linear combinatorial bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6557–6565, 2021.

[3] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

[4] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 322–331. IEEE, 1995.

[5] Viktor Bengs, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based online learning with dueling bandits: A survey. *Journal of Machine Learning Research*, 22:1–108, 2021.

[6] Jasmin Brandt, Viktor Bengs, Björn Haddenhorst, and Eyke Hüllermeier. Finding Optimal Arms in Non-stochastic Combinatorial Bandits with Semi-bandit Feedback and Finite Budget. In *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=h37KyWDDC6B.

[7] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(5), 2011.

[8] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

[9] Nicolò Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

[10] Yihan Du, Yuko Kuroki, and Wei Chen. Combinatorial pure exploration with full-bandit or partial linear feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7262–7270, 2021.

[11] Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pages 3221–3229, 2012. URL https://proceedings.neurips.cc/paper/2012/hash/8b0d268963dd0cfb808aac48a549829f-Abstract.html.

[12] C. A. R. Hoare. Algorithm 65: find. *Communications of the ACM*, 4(7):321–322, July 1961. ISSN 0001-0782, 1557-7317. doi: 10.1145/366622.366647. URL https://dl.acm.org/doi/10.1145/366622.366647.

[13] Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 240–248, 2016.

[14] Marc Jourdan, Mojmír Mutnỳ, Johannes Kirschner, and Andreas Krause. Efficient pure exploration for combinatorial bandits with semi-bandit feedback. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, pages 805–849, 2021.

[15] Yuko Kuroki, Liyuan Xu, Atsushi Miyauchi, Junya Honda, and Masashi Sugiyama. Polynomial-time algorithms for multiple-arm identification with full-bandit feedback. *Neural Computation*, 32(9):1733–1773, 2020.

[16] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

[17] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[18] Rémi Munos. From bandits to Monte-Carlo tree search: The optimistic principle applied to optimization and planning. *Foundations and Trends® in Machine Learning*, 7(1):1–129, 2014.

[19] Idan Rejwan and Yishay Mansour. Top-k combinatorial bandits with full-bandit feedback. In *Proc. ALT' 20*, pages 752–776, 2020.

[20] Wenbo Ren, Jia Liu, and Ness Shroff. The Sample Complexity of Best-k Items Selection from Pairwise Comparisons. In *Proceedings of the 37th International Conference on Machine Learning*, pages 8051–8072. PMLR, November 2020. URL https://proceedings.mlr.press/v119/ren20a.html. ISSN: 2640-3498.

[21] Omar Rivasplata. (PDF) Subgaussian random variables: An expository note. Technical report. URL https://www.researchgate.net/publication/318888443_Subgaussian_random_variables_An_expository_note.

[22] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[23] William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *BIOMETRIKA*, 25(3/4):285–294, 1933.

[24] Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2019. ISBN 978-1-108-49802-9. doi: 10.1017/9781108627771. URL https://www.cambridge.org/core/books/highdimensional-statistics/8A91ECEEC38F46DAB53E9FF8757C7A4E.

Appendix A. Supplementary Material

Algorithm	Sample complexity	Case	Reward function	Time
SEEKS-IC (ours, Thm. 5)	$O\left(\sum_{i\in[n]}\left[\Delta_i^{-2}\left(\log(n/\delta)+\log\log\Delta_i^{-1}\right)\right]\right)$	Top-k	non-linear	Poly(d)
GCB-PE [10]	$O\left(\frac{\lvert\sigma\rvert\beta_\sigma^2 L_p^2}{\Delta_{\min}^2}\log\frac{\beta_\sigma^2 L_p^2}{\Delta_{\min}^2\delta}\right)$	General	non-linear	Poly(d)
PolyALBA [10]	$\tilde{O}\left(\sum_{i=2}^{\lfloor\frac{d}{2}\rfloor}\frac{1}{\Delta_i^2}\log\frac{\lvert\mathcal{X}\rvert}{\delta}+\frac{d^2 m\xi_{\max}(\widetilde{M}(\lambda)^{-1})}{\Delta_{d+1}^2}\log\frac{\lvert\mathcal{X}\rvert}{\delta}\right)$	General	non-linear	Poly(d)
ICB [15]	$\tilde{O}\left(\frac{d\xi_{\max}(M(\lambda)^{-1})\rho(\lambda)}{\Delta_{\min}^2}\log\frac{d\xi_{\max}(M(\lambda)^{-1})\rho(\lambda)}{\Delta_{\min}^2\delta}\right)$	General	linear	Poly(d)
SAQM [15]	$\tilde{O}\left(\frac{d^{1/4}k\xi_{\max}(M(\lambda)^{-1})\rho(\lambda)}{\Delta_{\min}^2}\log\frac{d^{1/4}k\xi_{\max}(M(\lambda)^{-1})\rho(\lambda)}{\Delta_{\min}^2\delta}\right)$	Top-k	linear	Poly(d)
CSAR [19]	$\tilde{O}\left(\sum_{i=2}^{d}\frac{1}{\Delta_i^2}\log\frac{d}{\delta}\right)$	Top-k	linear	Poly(d)

Table 1: Comparison between our results and existing results for pure exploration tasks for combinatorial bandits with full-bandit feedback. "General" represents that the algorithm works for any combinatorial structure. $\tilde{O}(\cdot)$ only omits $\log\log$ factors. Main notations are defined in Section 2.

A.1. Proof of Assumptions

This is a list of the assumptions that are required by the algorithm as described in [20]:

1. **Time-Invariance.** The comparisons over two arms i and j follow a distribution that is time invariant. This is true from our definition of Relative-Strength in eq. (1).
2. **Tie-breaking.** Ties will naturally not occur since the outputs of r_t in line 8 of algorithm 1 come from continuous distributions and the probability of them producing the same value is 0.
3. **Independence.** This requires that all the samples are independent of time, items and sets. This naturally holds since the outcomes of r_t in line 8 of algorithm 1 are only affected by S_t which is uniformly sampled.
4. **Strong stochastic transitivity.** Here $i \succ j$ denotes that i is preferred to j. This requires three properties:
   (a) There is a strict order over the n items.
   (b) If $i \succ j$ then $\Delta(i,j) > 0$.
   (c) For any three items i, j, l if $i \succ j \succ l$ then $\Delta(i,l) \geq max(\Delta(i,j), \Delta(j,l))$

   Proof

   (a) This is by definition the descending order of arms from their values of μ_i.

   (b) This condition is satisfied by our assumption of Combinatorial Monotonicity.

   (c) From $i \succ j \succ l$ we know that $\mu_i > \mu_j > \mu_l$. Therefore $F(S_{t,i}) > F(S_{t,j}) > F(S_{t,l})$.
   $\Delta(i,l) = F(S_{t,i}) - F(S_{t,l}) = F(S_{t,i}) - F(S_{t,j}) + F(S_{t,j}) - F(S_{t,l})$
   $= \Delta(i,j) + \Delta(j,l)$.

So we get, $\Delta(i,l) = \Delta(i,j) + \Delta(j,l)$.

Hence Strong stochastic transitivity is satisfied.

■

5. **Stochastic triangle inequality.** For any three arms i, j, l, the inequality is $\Delta(i,l) \leq \Delta(i,j) + \Delta(j,l)$. This can be demonstrated almost identically to the proof for Strong stochastic transitivity.

A.2. Theoretical Analysis Elaborated

The "DART" algorithm [2] doesn't state a general sample complexity, however it does state a regret bound of $\mathcal{O}\left((U^3+U)K\sqrt{NKT\log 2NT}\right)$. Here the U is derived from it's Assumption 3 where it's stated that :
there exists a $U \leq \infty$ such that,

$$\frac{1}{U}\left\|\mathbb{E}[\mathbf{d}_{a_1}] - \pi\big(\mathbb{E}[\mathbf{d}_{a_2}]\big)\right\|_1 \leq |\mu_{a_1} - \mu_{a_2}| \leq U\left\|\mathbb{E}[\mathbf{d}_{a_1}] - \pi\big(\mathbb{E}[\mathbf{d}_{a_2}]\big)\right\|_1$$

for any pair of actions $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{N}$ and for any permutation π of $\mathbf{d}$.

Note that here $\mathbf{a}$ is a set of arms, chosen together to be performed as an action. We have μ_a is the expected reward of performing action $\mathbf{a}$. Also, $\mathbf{d}_a$ is a vector of rewards of the individual arm rewards of action $\mathbf{a}$.

In the case of two actions having identical μ_i, then the term $|\mu_{a_1} - \mu_{a_2}|$ can evaluate to be 0. The value of U needs to be infinity to satisfy the first inequality, since there almost certainly at least one permutation π that would cause $\left\|\mathbb{E}[\mathbf{d}_{a_1}] - \pi\big(\mathbb{E}[\mathbf{d}_{a_2}]\big)\right\|_1 > 0$. This implies that the value of U can approach $+\infty$, and as this term appears in the regret bound, in many problem settings, it has a worst-case time complexity of infinity.

A.3. Proof of sub-Gaussian sampling of $\hat{\Delta}_t(i,j)$, with $\sigma_\Delta = 2$

Proof From definition in eq. (3) $\hat{\Delta}_t(i,j) := r_t(S_t \cup \{i\}) - r_t(S_t \cup \{j\})$. From the definition of $r(S)$ in section 2, because η_t is sub-Gaussian noise, with scale parameter $\sigma = 1$, $r(S)$ is also sub-Gaussian with scale parameter $\sigma = 1$, although it is not centered at 0.
As the difference of two sub-Gaussian variables is also sub-Gaussian, and the scale parameter can be calculated to be $\sigma_\Delta = 2$ from Theorem 2.7 of [21]. ■

A.4. Hoeffding's inequality for sub-Gaussian Variables

The following is Hoeffding's inequality for sub-Gaussian variables [24].

$$\mathbb{P}(\sum_{i=1}^{N}(X_i - \mathbb{E}[X_i]) \geq s) \leq exp\left(\frac{-s^2}{2\sum_{i=1}^{N}\sigma_i^2}\right)$$

Here, $X_i\ \forall i = 1, 2...N$ are independent, and have a scale parameter of σ_i. This inequality is true for all $s \geq 0$. In section 3.2 and appendix A.6 we use this inequality on $\hat{\Delta}_t(i,j)$ which has been shown to be sub-Gaussian in Appendix A.3.

A.5. Algorithms

This section details the exact algorithms proposed. Note that Algorithms 2 to 4 are nearly identical to their corresponding algorithms in [20].

Algorithm 1: Isolated-Compare (IC) $(i, j, \epsilon, s_u, s_d, \delta, S_{up}, S_{mid}, S_{down})$

Set $t_{max} := \left\lceil \frac{32}{\epsilon^2} \ln\left(\frac{4}{\delta}\right) \right\rceil$, $\forall t \in \mathbb{Z}$, $b_t := \sqrt{\frac{8}{t} \ln \frac{2\pi^2 t^2}{3\delta}}$;
$t \leftarrow 0$;
repeat
  $t \leftarrow t + 1$;
  Sample $S_t \subseteq [n] \setminus \{i, j\}$ with $|S_t| = k - 1$ uniformly;
  Play $S_t \cup \{i\}$ and observe reward $r_t(S_t \cup \{i\})$;
  Play $S_t \cup \{j\}$ and observe reward $r_t(S_t \cup \{j\})$;
  $\hat{\Delta}_t(i, j) \leftarrow r_t(S_t \cup \{i\}) - r_t(S_t \cup \{j\})$;
  $\bar{\Delta}_t(i, j) \leftarrow \sum_{c=1}^{t} \frac{\hat{\Delta}_c(i,j)}{t}$;
  if $\bar{\Delta}_t(i, j) - b_t > s_u$ **then**
    Add i to S_{up} and **return**;
  else if $\bar{\Delta}_t(i, j) + b_t < -s_d$ **then**
    Add i to S_{down} and **return**;
  end if
until $t = t_{max}$;
if $\bar{\Delta}_t(i, j) > \frac{1}{2}\epsilon + s_u$ **then**
  Add i to S_{up};
else if $\bar{\Delta}_t(i, j) < -\frac{1}{2}\epsilon - s_d$ **then**
  Add i to S_{down};
else
  Add i to S_{mid};
end if

A.6. Proofs of theorems and performance guarantees

Note that the proofs in this section are almost identical to the proofs from [20].

Theorem 7 (Performance of Isolated-Compare (Algorithm 1)) *IC terminates after at most $O(\epsilon^{-2} \log \delta^{-1})$ comparisons, and with probability at least $1 - \delta$, one the following five events happens: (i) $\Delta(i, j) \geq \epsilon + s_u$ and item i is added to S_{up}; (ii) $\Delta(i, j) \in (s_u, \epsilon + s_u)$ and item i is not added to S_{down}; (iii) $\Delta(i, j) \in [-s_d, s_u]$ and item i in added to S_{mid}; (iv) $\Delta(i, j) \in (-\epsilon - s_d, -s_d)$ and item i is not added to S_{up}; and (v) $\Delta(i, j) \leq -\epsilon - s_d$ and item i is added to S_{down}.*

Proof IC terminates after at most $t_{\max} = \lceil 32\epsilon^{-2} \log(4/\delta) \rceil$ comparisons, and the sample complexity follows from the choice of $t_{\max}$. Now, we focus on proving the correctness, which is to prove that with probability at least $1 - \delta$, one of the five stated events occurs.

For any $t \in \mathbb{Z}^+$, we define a bad event that we do not want to happen,

$$\mathcal{E}_t := \left\{ \left| \bar{\Delta}_t(i, j) - \Delta(i, j) \right| \geq b_t \right\}.$$

Algorithm 2: Epsilon-Quick-Select-with-Isolated-Compare (EQS-IC)(S, k, ϵ, δ)

Randomly pick an item from S and denote it by v;
$S_{up}, S_{down} \leftarrow \emptyset$; $S_{mid} \leftarrow \{v\}$; $\delta_1 \leftarrow \frac{\delta}{|S|(|S|-1)}$;
for item i **in** S **and** $i \neq j$ **do**
  IC$(i, v, \frac{\epsilon}{2}, 0, 0, \delta_1, S_{up}, S_{mid}, S_{down})$;
end for;
if $|S_{up}| > k$ **then**
  return EQS-IC$(S_{up}, k, \epsilon, \frac{(n-1)\delta}{n})$; $\quad \#n = |S|$.
else if $|S_{up}| + |S_{mid}| \geq k$ **then**
  return $S_{up} \cup (k - |S_{up}|)$ random items of S_{mid};
else
  $k' \leftarrow k - |S_{up}| - |S_{mid}|$;
  return $S_{up} \cup S_{mid} \cup$ EQS-IC$(S_{down}, k', \epsilon, \frac{(n-1)\delta}{n})$;
end if

Algorithm 3: Tournament-k-Selection-with-Isolated-Compare (TKS-IC) $([n], k, \epsilon, \delta)$

For any $t \in \mathbb{Z}^+$, set $\epsilon_t := \frac{\epsilon}{4}\left(\frac{4}{5}\right)^t$ and $\delta_t := \frac{6\delta}{\pi^2 t^2}$;
Initialize $t \leftarrow 0$, $R_1 \leftarrow [n]$;
repeat
  $t \leftarrow t + 1$;
  Split R_t into $m_t = \left\lceil \frac{|R_t|}{2k} \right\rceil$ sets $(S_{t,i}, i \in [m_t])$, where $\forall i \in [m_t]$, $|S_{t,i}| < 2k$;
  for $i \in [m_t]$ **do**
    $A_{t,i} \leftarrow$ EQS-IC$(S_{t,i}, \min\{k, |S_{t,i}|\}, \epsilon_t, \frac{\delta_t}{k})$;
  end for;
  $R_{t+1} \leftarrow A_{t,1} \cup A_{t,2} \cup \cdots \cup A_{t,m_t}$;
until $|R_{t+1}| = k$;
return R_{t+1}

Algorithm 4: Sequential-Elimination-Exact-k-Selection-with-Isolated-Compare (SEEKS-IC) $([n], k, \delta)$

For all $t \in \mathbb{Z}^+$, set $\alpha_t := 2^{-t}$ and $\delta_t := \frac{6\delta}{\pi^2 t^2}$;
Initialize $t \leftarrow 1$, $R_1 \leftarrow [n]$, $S_1 \leftarrow \emptyset$, $k_1 \leftarrow k$;
repeat
  $A_t \leftarrow$TKS-IC$(R_t, k_t, \frac{\alpha_t}{3}, \frac{\delta_t}{3})$;
  $\{v_t\} \leftarrow$TKS-IC2$(A_t, 1, \frac{\alpha_t}{3}, \frac{\delta_t}{3})$;
  $S_{up} \leftarrow \emptyset$, $S_{mid} \leftarrow \{v_t\}$, $S_{down} \leftarrow \emptyset$;
  for items i in $R_t - v_t$ **do**
    IC$(i, v_t, \frac{\alpha_t}{3}, \frac{\alpha_t}{3}, \frac{\alpha_t}{3}, \frac{\delta_t}{3(|R_t|-1)}, S_{up}, S_{mid}, S_{down})$;
  end for;
  $S_{t+1} \leftarrow S_t \cup S_{up}$;
  $R_{t+1} \leftarrow R_t - S_{up} - S_{down}$;
  $k_{t+1} \leftarrow k_t - |S_{up}|$;
  $t \leftarrow t + 1$;
until $|S_t| \geq k$ *or* $|S_t \cup R_t| \leq k$;
return $S_t \cup \{k - |S_t|$ items in $R_t\}$;

By Theorem 4, we have that for all t in $\mathbb{Z}^+$, and we obtain the second inequality by expanding our definition of $b_t = \sqrt{\frac{8}{t}\ln(\frac{2\pi^2 t^2}{3\delta})}$

$$\mathbb{P}\{\mathcal{E}_t\} \leq 2\exp\{-\frac{tb_t^2}{8}\} \leq \frac{3\delta}{\pi^2 t^2}.$$

We define another bad event as

$$\mathcal{E}_{\text{out}} := \left\{ \left|\bar{\Delta}_t(i,j) - \Delta(i,j)\right| \geq \frac{\epsilon}{2};\ \text{where}\ \ t = t_{\max} \right\}.$$

This event's probability, by Chernoff-Hoeffding inequality, is upper bounded, and the final inequality comes from our definition of $t_{max} := \left\lceil \frac{32}{\epsilon^2} \ln\left(\frac{4}{\delta}\right) \right\rceil$

$$\mathbb{P}\{\mathcal{E}_{\text{out}}\} \leq 2\exp\left\{ \frac{-t_{max}}{8} \left(\frac{\epsilon}{2}\right)^2 \right\} \leq \delta/2.$$

Thus, by the union bound, the probability that some bad event happens is bounded by:

$$\mathbb{P}\left\{ \mathcal{E}_{\text{out}} \cup \left(\bigcup_{t=1}^{\infty} \mathcal{E}_t \right) \right\} \leq \frac{\delta}{2} + \sum_{t=1}^{\infty} \frac{3\delta}{\pi^2 t^2} = \delta.$$

In the remainder of the proof, we assume that no bad event occurs, which has a probability of at least $1 - \delta$. We split the rest of our proof into five cases, each for a specific event.

Case 1: $\Delta(i,j) \geq \epsilon + s_u$. Since none of $\mathcal{E}_t$ happens, for any round t, we have $\bar{\Delta}_t(i,j) > \Delta(i,j) - b_t \geq -b_t - s_d$, which implies that item i will not be added to S_{down} by Line 13. If IC proceeds

to Line 16, since $\mathcal{E}_{\text{out}}$ does not happen, we will have $\bar{\Delta}_t(i,j) > \Delta(i,j) - \epsilon/2 \geq \epsilon/2 + s_u$, which implies that item i will be added to S_{up} by Line 17.

Case 2: $\Delta(i,j) \in (s_u, \epsilon + s_u)$. Since none of $\mathcal{E}_t$ happens, for any round t, we have $\bar{\Delta}_t(i,j) > \Delta(i,j) - b_t > -b_t - s_d$, which implies that item i will not be added to S_{down} by Line 13. If IC proceeds to Line 15, since $\mathcal{E}_{\text{out}}$ does not happen, we will have $\bar{\Delta}_t(i,j) > \Delta(i,j) - \epsilon/2 > -\epsilon/2 - s_d$, which implies that item i will not be added to S_{down} by Line 19.

Case 3: $\Delta(i,j) \in [-s_d, +s_u]$. Since none of $\mathcal{E}_t$ happens, for any round t, we have $|\bar{\Delta}_t(i,j) - \Delta(i,j)| \leq b_t$, which implies that $\bar{\Delta}_t(i,j) + b_t > \Delta(i,j) > -s_d$ and $\bar{\Delta}_t(i,j) - b_t < \Delta(i,j) < s_u$. Thus, item i will not be added to S_{up} or S_{down} by Lines 11 or 13. IC must proceed to Line 16, since $\mathcal{E}_{\text{out}}$ does not happen, we will have $\bar{\Delta}_t(i,j) < \Delta(i,j) + \epsilon/2 < \epsilon/2 + s_u$ and $\bar{\Delta}_t(i,j) > \Delta(i,j) - \epsilon/2 \geq -\epsilon/2 - s_d$. Thus, item i will not be added to S_{up} or S_{down} by Lines 17 or 19. Therefore, item i will be added to S_{mid}.

Case 4: $\Delta(i,j) \in (-\epsilon - s_d, -s_d)$. Since none of $\mathcal{E}_t$ happens, for any round t, we have $\bar{\Delta}_t(i,j) < \Delta(i,j) + b_t < +b_t + s_u$, which implies that item i will not be added to S_{up} by Line 11. If IC proceeds to Line 16, since $\mathcal{E}_{\text{out}}$ does not happen, we will have $\bar{\Delta}_t(i,j) < \Delta(i,j) + \epsilon/2 < \epsilon/2 + s_u$, which implies that item i will not be added to S_{up} by Line 17.

Case 5: $\Delta(i,j) \leq -\epsilon - s_d$. Since none of $\mathcal{E}_t$ happens, for any round t, we have $\bar{\Delta}_t(i,j) < \Delta(i,j) + b_t \leq b_t + s_u$, which implies that item i will not be added to S_{up} by Line 11. If IC proceeds to Line 16, since $\mathcal{E}_{\text{out}}$ does not happen, we will have $\bar{\Delta}_t(i,j) < \Delta(i,j) + \epsilon/2 \leq -\epsilon/2 - s_d$, which implies that item i will be added to S_{down} by Line 15.

The correctness follows from the above five cases, and the proof is complete. ■

Theorem 4 (Error Bounds on Estimates of Relative-Strength) *A bound on the error of Relative-Strength estimates can be placed as:*

$$\mathbb{P}(|\bar{\Delta}_t(i,j) - \Delta(i,j)| \geq \epsilon) \leq 2\exp(-\epsilon^2 t/8),$$

where $\epsilon \in \mathbb{R}_+$.

Proof The rewards are drawn from subgaussian distributions with $\sigma = 1$ and so, using definitions from eq. (1), eq. (4) and appendix A.3, we know that $\hat{\Delta}(i,j)$ follows a subGaussian distribution with mean $\Delta(i,j)$ and $\sigma_\Delta = 2$. Hence, we can apply Hoeffding's inequality in appendix A.4 to obtain:

$$\begin{aligned}
&\mathbb{P}\left(\sum_{x=1}^{t}(\hat{\Delta}_x(i,j) - \Delta(i,j)) \geq s\right) \leq \exp\left(\frac{-s^2}{2\sum_{x=1}^{t}\sigma_\Delta^2}\right), \\
\xrightarrow{(a)}\ &\mathbb{P}\left(\left(\frac{\sum_{x=1}^{t}\hat{\Delta}_x(i,j)}{t}\right) - \Delta(i,j) \geq \frac{s}{t}\right) \leq \exp\left(\frac{-s^2}{2\sum_{x=1}^{t}\sigma_\Delta^2}\right), \\
\xrightarrow{(b)}\ &\mathbb{P}\left(\left|\left(\frac{\sum_{x=1}^{t}\hat{\Delta}_x(i,j)}{t}\right) - \Delta(i,j)\right| \geq \frac{s}{t}\right) \leq 2\exp\left(\frac{-s^2}{2\sum_{x=1}^{t}\sigma_\Delta^2}\right), \\
\xrightarrow{(c)}\ &\mathbb{P}\left(\left|\left(\frac{\sum_{x=1}^{t}\hat{\Delta}_x(i,j)}{t}\right) - \Delta(i,j)\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-\epsilon^2 t}{8}\right),
\end{aligned}$$

where step (a) came from dividing the left side by t, step (b) came from the fact that subGaussian distributions are symmetric, and step (c) came from the fact that we know $\sigma_\Delta = 2$ and letting $\epsilon = s/t$. As it's necessary for $s > 0$, so $\epsilon > 0$ as well. ■

Theorem 8 (Theoretical Performance of EQS-IC (Algorithm 2)) *Given an input set S with $|S| = n$, $1 \leq k \leq n/2$, and $\epsilon > 0, \delta \in (0,1)$, EQS-IC (S, k, ϵ, δ) terminates after $O(n\epsilon^{-2}\log(n/\delta))$ number of comparisons in expectation, and with probability at least $1-\delta$, returns an Individualized (ϵ,k)-optimal subset of S.*

Proof The proof consists of two parts: the proof of the correctness and the proof of the sample complexity. To avoid ambiguity, we use EQS-IC to denote the algorithm and subEQS-IC to denote the EQS-IC function called by the algorithm.

Let $\mathcal{E}$ be the event that all calls of IC return correct results, i.e., for each call of IC, one of the five events stated in theorem 7 happens. By theorem 7 and the union bound, $\mathcal{E}$ happens with probability at least $1 - \delta/n$.

Proof of the correctness

We prove the correctness by induction. First, let $n = 1$. In this case, k must be one. Since the only item is chosen as the pivot, and the pivot is added to S_{mid}, EQS-IC simply returns $\{1\}$ as the answer, which is correct with probability 1. Thus, when $n = 1$, EQS-IC returns an Individualized $(\epsilon, 1)$-optimal subset of S with probability 1.

Now we consider the case where $n > 1$. We make the following hypothesis to prove the correctness by induction.

Hypothesis 1. For all sets S' with size less than n, $k' \in \{1, 2, \ldots, |S'|\}$, and $\delta' \in (0, \delta]$, EQS-IC $(S', k', \epsilon, \delta')$ returns an Individualized (ϵ,k)-optimal subset of S' with probability at least $1 - \delta'$.

We note that when $n = 1$, EQS-IC returns an Individualized $(\epsilon, 1)$-optimal subset of S with probability 1, and thus, Hypothesis 1 holds for $n = 2$.

From now on till the end of the proof of the correctness, we assume that $\mathcal{E}$ happens and subEQS-IC (i.e., the EQS-IC called by the algorithm) also returns a correct result. We have shown that $\mathbb{P}\{\mathcal{E}\} \geq 1 - \delta/n$, and Hypothesis 1 claims that subEQS-IC returns a correct result with probability at least $1 - (n-1)\delta/n$. Thus, this assumption holds with probability at least $1 - \delta$.

First, we show a property about the sets S_{up}, S_{mid}, and S_{down}. Since $\mathcal{E}$ happens, according to theorem 7, all items i added to S_{up} have $i \succ v$, all items i added to S_{mid} have $\Delta(i, v) \in (-\epsilon/2, +\epsilon/2)$, and all items i added to S_{down} have $v \succ i$. Here we note that v is a pivot randomly picked from S.

Now, let item i in $S_{\text{up}} \cup S_{\text{mid}}$ and item j in $S_{\text{mid}} \cup S_{\text{down}}$ be given. There are four cases about items i and j.

Case 1: item i is in S_{up} and item j is in S_{mid}. Since $i \succ v$, we have $\Delta(i, j) \geq \Delta(v, j) \geq -\Delta(v, j) \geq -\epsilon/2 > -\epsilon$.

Case 2: item i is in S_{up} and item j is in S_{down}. In this case, we have $i \succ v \succ j$, which implies that $\Delta(i,j) > 0 > -\epsilon$.

Case 3: item i is in S_{mid} and item j is in S_{mid}. By the definition of STI, we have $\Delta(i,j) \leq \Delta(i,v) + \Delta(j,v) \leq \epsilon$, which implies that $\Delta(i,j) \geq -\Delta(i,j) \geq -\epsilon$.

Case 4: item i is in S_{mid} and item j is in S_{down}. Since $v \succ j$, we have $\Delta(i,j) \geq \Delta(i,v) \geq -\Delta(i,v) \geq -\epsilon/2 > -\epsilon$.

Thus, from the above four cases, we conclude that for any item i in $S_{\text{up}} \cup S_{\text{mid}}$ and j in $S_{\text{mid}} \cup S_{\text{down}}$, $\Delta(i,j) \geq -\epsilon$.

Next, we finish the proof of the correctness by analyzing the following three cases. Let R be the returned set of EQS-IC . Let i be an item in R and j be an item not in R.

Case 1: $|S_{\text{up}}| > k$. In this case, item i is in S_{up}. If j is in S_{up}, by Hypothesis 1, the set returned by subEQS-IC is an Individualized (ϵ,k)-optimal subset of S_{up}, and thus, $\Delta(i,j) \geq -\epsilon$. For the case where j is in $S_{\text{mid}} \cup S_{\text{down}}$, we have shown that $\Delta(i,j) \geq -\epsilon$.

Case 2: $|S_{\text{up}}| \leq k$ and $|S_{\text{up}}| + |S_{\text{mid}}| \geq k$. In this case, we have that i is in $S_{\text{up}} \cup S_{\text{mid}}$ and j is in $S_{\text{mid}} \cup S_{\text{down}}$. We have shown that $\Delta(i,j) \geq -\epsilon$.

Case 3: $|S_{\text{up}}| + |S_{\text{mid}}| < k$. In this case, j is in S_{down}. For the case where i is in $S_{\text{up}} \cup S_{\text{mid}}$, we have shown that $\Delta(i,j) \geq -\epsilon$. If i is in S_{down}, then i is in the returned set of subEQS-IC , which by Hypothesis 1 implies that $\Delta(i,j) \geq -\epsilon$.

Therefore, if Hypothesis 1 holds for n, EQS-IC returns a correct Individualized (ϵ,k)-optimal subset of S with probability at least $1-\delta$. Since $k \leq n$ and $\delta <$ are arbitrary, Hypothesis 1 holds for $n+1$. Also, since Hypothesis 1 holds for $n = 2$, Hypothesis 1 holds for all $n \geq 2$. This completes the proof of the correctness.

Proof of the sample complexity

We prove the sample complexity by induction. Let $c_1 > 0$ be the hidden constant of the sample complexity of IC stated in theorem 7. For any positive integer n_1, we use $T(n_1, k_1, \epsilon, \delta_1)$ to denote the upper bound of the expected number of comparisons conducted by the call of EQS-IC $([n_1], k_1, \epsilon, \delta_1)$, where $[n_1]$ denotes an arbitrary set consisting of n_1 items, k_1 is a positive integer with $k_1 \leq \min\{n_1, k\}$, and δ_1 is in $(0, \delta]$.

When there is only one item, we have $T(1, k_1, \epsilon, \delta_1) = 0$, as we do not need to conduct any comparison. When there are two items, since we only need to compare the two items in the call of IC, we have $T(2, k_1, \epsilon, \delta_1) \leq c_1 \epsilon^{-2} \log \delta^{-1}$ for any $k_1 \leq \min\{2, k\}$ and $\delta_1 \in (0, \delta]$.

Now we let $n_1 > 2$, $k_1 \leq \min\{n_1, k\}$, and $\delta_1 \in (0, \delta]$ be given, we make the following hypothesis. Note that we have shown that when $n_1 = 3$, Hypothesis 2 holds.

Hypothesis 2. For all $n_2 < n_1$, $k_2 \leq \min\{n_2, k_1\}$, and $\delta_2 \in (0, \delta_1]$, $T(n_2, k_2, \epsilon, \delta_2) \leq c_2 n_2 \epsilon^{-2} \log(n_2/\delta_2)$, where $c_2 > 0$ is a sufficiently large constant.

For the call of EQS-IC $([n_1], k_1, \epsilon, \delta_1)$, we use v to denote its pivot, and use l to denote the rank of item v in $[n_1]$, i.e., item v ranks the l-th best in $[n_1]$. Since the pivot v is picked at random, l is uniformly distributed on $[n_1]$.

We recall that $\mathcal{E}$ is the event that all ICs called by EQS-IC $([n_1], k_1, \epsilon, \delta_1)$ return correct results, i.e., for each call of IC, one of the five events stated in theorem 7 happens. By theorem 7, $\mathcal{E}$ happens with probability at least $1 - \delta/n_1$.

First, we consider the case where $\mathcal{E}$ does not happen. In this case, since v is added to S_{mid}, we have $|S_{\text{up}}| \leq n_1 - 1$ and $|S_{\text{down}}| \leq n_1 - 1$, and subEQS-IC (if existing) will only be executed on one of S_{up} and S_{down}. Hence, in this case, the expected number of comparisons conducted by EQS-IC $([n_1], k_1, \epsilon, \delta_1)$ is

$$\begin{aligned} T_1 &\leq \max_{k' \in [k_1]} T(n_1 - 1, k', \epsilon, \delta_1) + \frac{c_1(n_1 - 1)}{\epsilon^2} \log \frac{n_1}{\delta_1} \\ &\leq (c_2 + c_1) \cdot \frac{n_1}{\epsilon^2} \log \frac{n_1}{\delta_1}. \end{aligned}$$

Next, we consider the case where $\mathcal{E}$ happens. In this case, since theorem 7 states that no item less preferred than the pivot v will be added to S_{up} and no item more preferred than the pivot v will be added to S_{down}, we have $|S_{\text{up}}| \leq l - 1$ and $|S_{\text{down}}| \leq n_1 - l$. If $l > k_1$, then we have $|S_{\text{up}}| + |S_{\text{mid}}| = n_1 - |S_{\text{down}}| > l > k_1$, which implies that subEQS-IC (if existing) will only be executed on the set S_{up}, and the size of S_{up} is no more than $(l - 1)$. If $l \leq k_1$, then we have $|S_{\text{up}}| \leq l \leq k_1$, which implies that subEQS-IC (if existing) will only be executed on S_{down}, and the size of S_{down} is at most $(n_1 - l)$. Hence, when $\mathcal{E}$ happens, the expected number of comparisons conducted by the call of EQS-IC $([n_1], k_1, \epsilon, \delta_1)$ is

$$\begin{aligned} T_2 \leq\ & \frac{1}{n_1} \sum_{l=1}^{k_1} \left[T\left(n_1 - l, k_1 - l, \epsilon, \frac{n_1 - 1}{n_1} \cdot \delta_1\right) \right] \\ & + \frac{1}{n_1} \sum_{l=k_1+1}^{n_1} \left[T\left(l - 1, k_1, \epsilon, \frac{n_1 - 1}{n_1} \cdot \delta_1\right) \right] \\ & + \frac{c_1(n_1 - 1)}{\epsilon^2} \log \frac{n_1}{\delta} \\ \leq\ & \frac{c_2}{n_1} \sum_{l=1}^{k_1} \left[\frac{n_1 - l}{\epsilon^2} \log \frac{(n_1 - l)n_1}{(n_1 - 1)\delta_1} \right] \\ & + \frac{c_2}{n_1} \sum_{l=k_1+1}^{n_1} \left[\frac{l - 1}{\epsilon^2} \log \frac{(l - 1)n_1}{(n_1 - 1)\delta_1} \right] + \frac{c_1 n_1}{\epsilon^2} \log \frac{n_1}{\delta_1} \\ \leq\ & \frac{c_2}{n_1} \left\{ \sum_{l=1}^{k_1} \left[\frac{n_1 - l}{\epsilon^2} \log \frac{n_1}{\delta_1} \right] + \sum_{l=k_1+1}^{n_1} \left[\frac{l - 1}{\epsilon^2} \log \frac{n_1}{\delta_1} \right] \right\} \\ & + \frac{c_1 n_1}{\epsilon^2} \log \frac{n_1}{\delta_1} \end{aligned}$$

$$
\begin{aligned}
&= \frac{c_2}{n_1\epsilon^2}\log\frac{n_1}{\delta_1}\left\{\frac{(2n_1-1-k_1)k_1}{2}+\frac{(n_1+k_1-1)(n_1-k_1)}{2}\right\}+\frac{c_1n_1}{\epsilon^2}\log\frac{n_1}{\delta_1}\\
&= \frac{c_2}{n_1\epsilon^2}\log\frac{n_1}{\delta_1}\cdot\left[k_1(n_1-k_1)+\frac{1}{2}n_1(n_1-1)\right]+\frac{c_1n_1}{\epsilon^2}\log\frac{n_1}{\delta_1}\\
&\le \frac{c_2}{n_1\epsilon^2}\log\frac{n_1}{\delta_1}\cdot\frac{3}{4}n_1^2+\frac{c_1n_1}{\epsilon^2}\log\frac{n_1}{\delta_1}\\
&= \left(\frac{3}{4}c_2+c_1\right)\frac{n_1}{\epsilon^2}\log\frac{n_1}{\delta_1}.
\end{aligned}
$$

Summarizing the numbers of comparisons in these two cases, by $\mathbb{P}\{\mathcal{E}\}\ge 1-\delta_1/n_1$, $n_1>2$, and $\delta_1<$, we get

$$
\begin{aligned}
T(n_1,k_1,\epsilon,\delta_1) &\le \left(1-\frac{\delta_1}{n_1}\right)\cdot T_2+\frac{\delta_1}{n_1}\cdot T_1\\
&\le \left(\left(\frac{3}{4}+\frac{\delta_1}{4n_1}\right)c_2+c_1\right)\frac{n_1}{\epsilon^2}\log\frac{n_1}{\delta_1}\\
&\le \left(\frac{19}{24}c_2+c_1\right)\frac{n_1}{\epsilon^2}\log\frac{n_1}{\delta_1}.
\end{aligned}
$$

Choose $c_2\ge 4.8c_1$, and then we have $T(n_1,k_1,\epsilon,\delta_1)\le c_2n_1\epsilon^{-2}\log(n_1/\delta_1)$. Thus, if Hypothesis 2 holds for n_1, it will hold for n_1+1. We also recall that when $n_1\le 3$, Hypothesis 2 holds. Therefore, by induction, Hypothesis 2 holds for all values of n_1. Hence, EQS-IC $([n],k,\epsilon,\delta)$ terminates after at most $c_2n\epsilon^{-2}\log(n/\delta)$ number of comparisons in expectation. This completes the proof of the sample complexity, and the proof of theorem 8 is complete. ■

Theorem 9 (Theoretical Performance of TKS-IC (Algorithm 3)) *Given input $1\le k\le n/2$, and $\epsilon,\delta\in(0,1/2)$, TKS-IC terminates after $O(n\epsilon^{-2}\log(k/\delta))$ number of comparisons in expectation, and with probability at least $1-\delta$, returns an Individualized (ϵ,k)-optimal subset.*

Proof We first prove the correctness of TKS-IC and then prove its sample complexity. Here, we let T be the number of rounds, and thus, the returned set is R_{T+1}.

Proof of the correctness

Step 1 is to prove that for any round t, R_{t+1} contains an Individualized (ϵ,k)-optimal subset of R_t. Let $b_1,b_2,\ldots,b_k$ be the best-k items of R_t, and denote $B=\{b_1,b_2,\ldots,b_k\}$. Also, for all $l\in[k]$, we use S_{t,s_l} to denote the split set that contains b_l.

We let $\mathcal{E}^t_{\text{good}}$ be the event that for all $l\in[k]$, the calls of EQS-IC on S_{t,s_l} return correct results. By theorem 8 and the union bound, we have $\mathbb{P}\{\mathcal{E}^t_{\text{good}}\}\ge 1-\delta_t$. During the proof of the correctness, we assume that $\mathcal{E}^t_{\text{good}}$ happens for all t, and by the union bound, we have that

$$
\mathbb{P}\left\{\bigcap_{t=1}^{T}\mathcal{E}^t_{\text{good}}\right\}\ge 1-\sum_{t=1}^{T}\mathbb{P}\left\{(\mathcal{E}^t_{\text{good}})^c\right\}
$$

$$\geq 1 - \sum_{t=1}^{T} \delta_t$$
$$\geq 1 - \sum_{t=1}^{\infty} \frac{6\delta}{\pi^2 t^2} = 1 - \delta. \tag{5}$$

We complete Step 1 by constructing a subset $U \subset R_{t+1}$ that is an Individualized (ϵ,k)-optimal subset of R_t. The construction consists of stages. We note that we only need to prove the existence of such a set, and thus, in the construction, we have the oracle knowledge about the values of $b_1, b_2, \ldots, b_k$.

Stage 0: Let U be the empty set.

Stage 1: If b_1 is not in A_{t,s_1}, then by theorem 8, all items i in A_{t,s_1} have $\Delta(i, b_1) \geq -\epsilon_t$. By the definition of SST, this implies that $\Delta(i, j) \geq -\epsilon_t$ for all items j in R_t. In this case, we let $U = A_{t,s_1}$, which is an Individualized (ϵ,k)-optimal subset of R_t, and the construction of U is complete. If b_1 is in A_{t,s_1}, then we add b_1 to U and the construction proceeds to Stage 2.

Stage l for any l in $\{2, 3, \ldots, k\}$: We hypothesize that either (i) the construction has ended at an earlier stage, or (ii) $b_1 \in A_{t,s_1}, b_2 \in A_{t,s_2}, \ldots, b_{l-1} \in A_{t,s_{l-1}}$, and $U = \{b_1, b_2, \ldots, b_{l-1}\}$. Now we assume that the construction has not ended, otherwise we skip this stage. If b_l is not in A_{t,s_l}, then by the property of EQS-IC stated in Theorem 4 and the definition of SST, in $A_{t,s_l} - \{b_1, b_2, \ldots, b_{l-1}\}$, there are at least $|A_{t,s_l}| - l + 1 = k - l + 1$ items i such that for all items j in $R_t - \{b_1, b_2, \ldots, b_{l-1}\}$, we have $\Delta(i, j) \geq \Delta(i, b_l) \geq -\epsilon_t$. In this case, we add these $(k - l + 1)$ items to U. Then U is an Individualized (ϵ,k)-optimal subset of R_t, and the construction of U is complete. If b_l is not in A_{t,s_l}, then we add b_l to U, and the construction proceeds to Stage $(l + 1)$.

Stage 1 does not require any hypothesis, and after each Stage l for l in $[k-1]$, the hypothesis required by Stage $(l + 1)$ is satisfied. Also, each stage adds at least one item to U. Hence, the construction completes after at most k stages. From the above induction, we have that U is an Individualized (ϵ,k)-optimal subset of R_t. Thus, for any t, given that $\mathcal{E}_t$ happens, R_{t+1} contains an Individualized (ϵ,k)-optimal subset of R_t.

Step 2 is to finish the proof of the correctness. Step 1 has shown that for each $t \in [T]$, there exists a set $U_{t+1} \subset R_{t+1}$ such that U_{t+1} is an Individualized (ϵ,k)-optimal subset of R_t. Recall that T is the last round, and the loop ends only when $|R_t|$ reaches k. Thus, $|R_{T+1}| = k$, and $U_{T+1} = R_{T+1}$.

Let $t > 1$ be given, and let u_{t+1} be an item in U_{t+1}. If u_{t+1} is in U_t, then we let $u_t = u_{t+1}$, which implies that $\Delta(u_{t+1}, u_t) = 0 \geq -\epsilon_t$. If u_{t+1} is not in U_t, then by the fact that $|U_t| = |U_{t+1}|$, $U_t - U_{t+1}$ contains at least one item, and we denote this item by u_t. By Step 1, u_t has $\Delta(u_{t+1}, u_t) \geq -\epsilon_t$. Thus, in both cases, we have $\Delta(u_{t+1}, u_t) \geq -\epsilon_t$.

Let i be an item in R_{T+1} and j be an item in $[n] - R_{T+1}$. Since $j \in [n] = R_1$ and $j \notin R_{T+1}$, there is an r such that $j \in R_r$ and $j \notin R_{r+1}$. We use u_{T+1} to denote i. By the above paragraph, there exists a sequence of items $u_T \in U_T, u_{T-1} \in U_{T-1}, \ldots, u_{r+1} \in U_{r+1}$ such that $\Delta(u_{t+1}, u_t) \geq -\epsilon_t$ for all $t \in \{T, T-1, T-2, \ldots, r+1\}$. Also, since item j is in R_r but not in R_{r+1}, we have $\Delta(u_{r+1}, j) \geq -\epsilon_r$. By this sequence, we conclude that

$$\begin{aligned}
\Delta(i,j) &= \Delta(u_{T+1}, j) \\
&\geq \Delta(u_T, j) - \epsilon_T \\
&\geq \Delta(u_{T-1}, j) - \epsilon_T - \epsilon_{T-1} \\
&\vdots \\
&\geq \Delta(u_{r+1}, j) - \sum_{s=r+1}^{T} \epsilon_s \\
&> -\sum_{s=r}^{T} \epsilon_s \\
&\geq -\sum_{s=1}^{\infty} \epsilon_s = -\epsilon.
\end{aligned}$$

Thus, when $\mathcal{E}^t_{\text{good}}$ happens for all t, the returned set of EQS-IC is an Individualized (ϵ,k)-optimal subset of $[n]$. By eq. (5), the joint event $\bigcap_{t=1}^{T} \mathcal{E}^t_{\text{good}}$ happens with probability at least $1-\delta$. This completes the proof of the correctness.

Proof of the sample complexity

At each round t, there are $\lceil |R_t|/m \rceil$ calls of EQS-IC. Each call of EQS-IC involves at most $2k$ items with parameters k (or less), ϵ_t, and δ_t, and returns at most k items. Thus, we have $|R_{t+1}| = \lceil |R_t|/(2k) \rceil k$. If $|R_t| \leq \lceil n/(2^{t-1}k) \rceil k$, then we have $|R_{t+1}| \leq \lceil \lceil n/(2^{t-1}k) \rceil /2 \rceil k \leq \lceil n/(2^t k) \rceil k$. Also, we have $|R_1| = n \leq \lceil n/k \rceil k$, and thus, by induction, for any t,

$$|R_t| \leq \lceil n/(2^{t-1}k) \rceil k \leq c_3 n \cdot 2^{-t},$$

where $c_3 > 0$ is some universal constant. By the fact that $|R_t| \geq k$, we also get that the number of EQS-IC called by round t is at most

$$\lceil |R_t|/(2k) \rceil \leq |R_t|/k \leq c_3 n \cdot 2^{-t}/k.$$

Let $c_4 > 0$ be the hidden constant factor in the sample complexity stated in theorem 8. We conclude that the expected number of comparisons conducted by TKS-IC is

$$\mathbb{E}[N] \leq \mathbb{E}\left\{ \sum_{t=1}^{T} \left\lceil \frac{|R_t|}{2k} \right\rceil \cdot c_4 \left(\frac{2k}{\epsilon_t^2} \log\left(\frac{2k^2}{\delta_t} \right) \right) \right\} \tag{6}$$

$$\leq \sum_{t=1}^{\infty} \left\lceil \frac{|R_t|}{2k} \right\rceil \cdot c_4 \left(\frac{2k}{\epsilon_t^2} \log\left(\frac{2k^2}{\delta_t} \right) \right) \tag{7}$$

$$\leq \sum_{t=1}^{\infty} \left[\frac{c_3 n}{2^t k} \cdot c_4 \left(\frac{2k}{\epsilon^2} \cdot \left(\frac{5}{4} \right)^{2t} \log\left(\frac{2\pi^2 k^2 t^2}{6\delta} \right) \right) \right] \tag{8}$$

$$= \frac{2c_3c_4n}{\epsilon^2} \sum_{t=1}^{\infty} \left[\left(\frac{25}{32}\right)^t \left(2\log t + \log\left(\frac{2\pi^2k^2}{6\delta}\right)\right)\right] \tag{9}$$

$$= O\left(\frac{n}{\epsilon^2}\log\frac{k}{\delta}\right). \tag{10}$$

This completes the proof of the sample complexity, and the proof of theorem 9 is complete. ■

Theorem 5 (Theoretical Performance of SEEKS-IC (Algorithm 4)) *With probability at least* $1-\delta$*, SEEKS-IC terminates after* $O\left(\sum_{i\in[n]}\left[\Delta_i^{-2}\left(\log(n/\delta)+\log\log\Delta_i^{-1}\right)\right]\right)$ *number of comparisons in expectation, and returns the best-*k *items.*

Proof

Notations. We use round t to denote the t-th iteration of Lines 3 to 14. For any item i, we use I_i^t to denote the index of the round when i is assured (i.e., the round when item i is added to S_{up} or S_{down} and not added to R_{t+1}) and define $T_i := \min\{T, I_i\}$ as the index of the last round when item i is involved in some comparisons. We use T to denote the index of the last round. Assume that the unknown true order of these n items is $r_1 \succ r_2 \succ \cdots \succ r_n$. Define $U := \{r_1, r_2, \ldots, r_k\}$ as the set of the best-k items, and $U_t := U \cap R_t$.

Proof of the correctness

To prove that if SEEKS-IC returns, then the returned set is U with probability at least $1-\delta$. We prove the correctness by induction.

Hypothesis 4. Let $t \le T+1$ be given. We hypothesize that $S_t \subset U \subset R_t \cup S_t$ with probability at least $1 - \sum_{r=1}^{t-1}\delta_r$.

When $t=1$, we have $R_1 = [n]$ and $S_1 = \emptyset$, which implies that $S_1 = \emptyset \subset U \subset [n] = R_1$ with probability 1. Thus, Hypothesis 4 holds for $t=1$. Now, we consider the case where $t \ge 2$.

First, we bound an event. Let $\mathcal{E}_t$ be the event that in round t, all the calls of TKS-IC, TKS-IC2, and ICs return correct results. By theorem 9, theorem 7, and the union bound, we have

$$\mathbb{P}\{\mathcal{E}_t\} \ge 1 - \frac{\delta_t}{3} - \frac{\delta_t}{3} - \frac{\delta_t}{3(|R_t|-1)} \cdot (|R_t|-1) \tag{11}$$

$$\ge 1 - \delta_t. \tag{12}$$

In the proof of the correctness, we assume $\mathcal{E}_t$ happens.

Second, we show a useful property of the pivot v_t. In each iteration, items in S_{up} are added to S_t and k_t is decreased by $|S_{\text{up}}|$, and thus, $k_t = k - |S_t|$. By Hypothesis 4, we have $S_t \subset U \subset R_t \cup S_t$, $U_t \subset R_t$, and $S_t \cap R_t = \emptyset$, and thus, $U_t = U - S_t$ and $|U_t| = |U - S_t| = k - |S_t| = k_t$. By Theorem 5, for any item i in A_t and j in $(R_t - A_t)$, we have $\Delta(i,j) \ge -\alpha_t/3$. If $U_t = A_t$, then we have $v_t \succ r_k$, which implies that $\Delta(v_t, r_k) \ge 0 > -\alpha_t/3$. If $U_t \ne A_t$, then $R_t - A_t$ contains some item in U (which implies that $v \succeq k$), and thus, $\Delta(v_t, r_k) \ge \Delta(v_t, v) \ge -\alpha_t/3$. Thus, in both cases, we have $\Delta(v_t, r_k) \ge -\alpha_t/3$.

For Line 5, we recall that TKS-IC2 is almost the same as TKS-IC with the only difference being that TKS-IC2 is used for finding the worst items. By Theorem 5, we have that for any item j in $A_t - \{v_t\}$, $\Delta(v_t, j) \leq \alpha_t/3$. Since $|A_t| = |U_t| = k_t$ and $A_t \cap U \subset R_t \cap U \subset U_t$, m_t the worst item in A_t has $r_k \succeq m_t$. Thus, $\Delta(v_t, r_k) \leq \Delta(v_t, m_t) \leq \alpha_t/3$. Therefore, we conclude

$$-\epsilon/3 \leq \Delta(v_t, r_k) \leq \alpha_t/3. \tag{13}$$

The third step is to show that in round t, $S_{\text{up}} \subset U_t$ and $S_{\text{down}} \cap U_t = \emptyset$. Let item i in U_t be given. Since $\mathcal{E}_t$ happens, the calls of IC on items i and j give correct results. Since item i is in U_t, we have $\Delta(i, r_k) \geq 0$, which by eq. (13) implies that $\Delta(i, v_t) \geq -\alpha_t/3$. By theorem 7, item i is not added to S_{down}. Hence, no item in U_t is added to S_{down}, which implies $S_{\text{down}} \cap U_t = \emptyset$.

Let item j in $R_t - U_t$ be given. Since $r_k \succ j$, we have $\Delta(r_k, j > 0$, which implies that $\Delta(j, r_k) \leq \alpha_t/3$. By theorem 7, item j is not added to S_{up}. Thus, no item in $R_t - U_t$ is added to S_{up}, which implies $S_{\text{up}} \subset U_t$.

Lastly, we show that Hypothesis 4 holds for all t. We have already proved that when Hypothesis 4 holds for t, with probability at least $1 - \delta_t$ (i.e., when $\mathcal{E}_t$ happens), $S_{\text{up}} \subset U_t$ and $S_{\text{down}} \cap U_t = \emptyset$. By $S_{\text{up}} \subset U_t$ and $S_t \subset U$, we get

$$S_{t+1} = S_t \cup S_{\text{up}} \subset U.$$

By $S_{\text{down}} \cap U_t = \emptyset$ and $U \subset R_t \cup S_t$, we get

$$U_t \cap (R_t - S_{t+1} - R_{t+1}) = U_t \cap S_{\text{down}} = \emptyset,$$

which implies that $U_t \subset S_{t+1} \cup R_{t+1}$. Hence,

$$\begin{aligned} U &= U_t \cup (U - U_t) \\ &= U_t \cup ((R_t \cup S_t) \cap U - R_t \cap U) \\ &= U_t \cup (S_t \cap U) \\ &\subset R_{t+1} \cup S_{t+1} \cup S_t \\ &= R_{t+1} \cup S_{t+1}. \end{aligned}$$

Thus, we conclude that with probability at least $1 - \sum_{r=1}^{t-1} \delta_r - \delta_t = 1 - \sum_{r=1}^{t} \delta_r$, $S_{t+1} \subset U \subset R_{t+1} \cup S_{t+1}$. This means that if Hypothesis 4 holds for t, then it holds for $t + 1$. It has also been shown that when $t = 1$, Hypothesis 4 holds. Thus, Hypothesis 4 holds for all $t \leq T + 1$.

Therefore, with probability at least

$$1 - \sum_{r=1}^{T} \delta_r \geq 1 - \sum_{r=1}^{\infty} \frac{6\delta}{\pi^2 r^2} \geq 1 - \delta, \tag{14}$$

$S_{T+1} \subset U \subset R_{T+1} \cup S_{T+1}$. Also, we have $|R_{T+1} \cup S_{T+1}| \leq k$. Thus, the returned set $S_{T+1} \cup R_{T+1}$ is exactly U. This completes the proof of the correctness.

Proof of the sample complexity

In the proof of the sample complexity, we assume that $\bigcap_{t=1}^{T} \mathcal{E}_t$ happens. By eq. (14), $\bigcap_{t=1}^{T} \mathcal{E}_t$ happens with probability at least $1-\delta$. Thus, with probability at least $1-\delta$, all the calls of TKS-IC, TKS-IC2, and IC return correct results.

Let N denote the number of comparisons conducted by SEEKS-IC. In round t, the comparisons are conducted by the calls of TKS-IC (Line 4), TKS-IC2 (Line 5), and IC (Line 8). By theorem 9, the expected number of comparisons conducted by TKS-IC is at most $O(|R_t|\alpha_t^{-2}\log(n/\delta_t))$, and that of TKS-IC2 is at most $O(k_t\alpha_t^{-2}\log(n/\delta_t))$. By theorem 7, the expected number of comparisons conducted by each call of IC is at most $O(\alpha_t^{-2}\log(|R_t|/\delta_t))$. Thus, in round t, the expected number of comparisons is at most $O(|R_t|\alpha_t^{-2}\log(|R_t|/\delta_t)) = O(|R_t|\alpha_t^{-2}\log(n/\delta_t))$. Recall that for any item i, T_i is the index of the round when item i is assured (i.e., item i is not added to R_{t+1}) or the algorithm terminates. Thus, we have

$$\begin{aligned}\mathbb{E}[N] &\le c_9\mathbb{E}\left\{\sum_{t=1}^{T}\left[|R_t|\alpha_t^2\log(n/\delta_t)\right]\right\}\\ &\le c_9\sum_{i\in[n]}\mathbb{E}\left\{\sum_{t=1}^{T_i}\left[\alpha_t^{-2}\log(n/\delta_t)\right]\right\},\end{aligned} \tag{15}$$

where $c_9 > 0$ is a universal constant.

Now let item $i \neq r_k$ be given. Define $\tau_i := \inf\{t \in \mathbb{Z}^+ : \alpha_t < |\Delta(i, r_k)|\}$, i.e., when $t > \tau_i$, we have $\alpha_t < |\Delta(i, r_k)|$. Since $\alpha_t = 2^{-t}$, we have $\tau_i \le 1 + \log_2 |\Delta(i, r_k)|^{-1}$.

Let $t \ge \tau_i$ be given. First, we consider the case where i is in $[n] - U$. When $t \ge \tau_i$, we have $\alpha_t < |\Delta(i, r_k)|$—, i.e., $\Delta(i, r_k) = -|\Delta(i, r_k)| < -\alpha_t$. By eq. (13), we have $|\Delta_{v_t,r_k}| \le \alpha_t/3$, which implies that $\Delta(i, v_t) \le 1/2 - (|\Delta_{i,r_k}| - |\Delta_{r_k,v_t}|) < -2\alpha_t/3$. Since $\mathcal{E}_t$ happens, by theorem 7, at round t, item i is added to S_{down}, i.e., item i is not added to R_{t+1}. Second, we consider the case where $i \in U - \{r_k\}$. Since $t \ge \tau_i$, we have $\alpha_t < |\Delta(i, r_k)|$, i.e., $\Delta(i, r_k) > \alpha_t$. By eq. (13), we have $|\Delta(v_t, r_k)| \le \alpha_t/3$, which implies that $\Delta(i, v_t) = |\Delta(i, v_t)| \ge (|\Delta(i, r_k)| - |\Delta(v_t, r_k)|) \ge -2\alpha_t/3$. Since $\mathcal{E}_t$ happens, by Lemma 3, at round t, item i is added to S_{up}, i.e., item i is not added to R_{t+1}. Thus, when $\bigcap_{t=1}^{T} \mathcal{E}_t$ happens,

$$T_i \le \tau_i \le 1 + \log_2 |\Delta(i, r_k)|^{-1},$$

from which it follows that

$$\begin{aligned}\mathbb{E}\left\{\sum_{t=1}^{T_i}\left[\alpha_t^{-2}\log(n/\delta_t)\right]\right\} &\le \sum_{t=1}^{\tau_i}\left[4^t\log\left(\frac{\pi^2t^2n}{6\delta}\right)\right]\\ &\le \sum_{t=1}^{\tau_i}\left[4^t\log(\tau_i^2)\right] + \sum_{t=1}^{\tau_i}\left[4^t\log\left(\frac{\pi^2n}{6\delta}\right)\right]\\ &\le c_{10}\cdot 4^{\tau_i}\left(\log\tau_i + \log\left(\frac{\pi^2n}{6\delta}\right)\right)\end{aligned}$$

$$\leq c_{11} \cdot 4^{1+\log_2 |\Delta(i,r_k)|^{-1}} \left(\log(1+\log_2 |\Delta(i,r_k)|^{-1}) + \log(n/\delta)\right)$$
$$\leq c_{12} |\Delta(i,r_k)|^{-2} \left(\log(n/\delta) + \log\log |\Delta(i,r_k)|^{-1}\right), \tag{16}$$

where $c_{10}, c_{11}, c_{12} > 0$ are three universal constants.

Also, we observe that when all items in $[n] - U$ are assured, SEEKS-IC will terminate and conduct no more comparisons. At round t with $t \geq \max_{i\in[n]-U} \tau_i = \tau_{r_{k+1}}$, since $\bigcap_{t=1}^{T} \mathcal{E}_t$ happens, all items not in U are assured. Thus, we have $T_{r_1}, T_{r_2}, \ldots, T_{r_k} \leq \tau_{r_{k+1}}$. Similar to eq. (16), we have that for any item i in U,

$$\mathbb{E}\left\{\sum_{t=1}^{T_i} \left[\alpha_t^{-2} \log(n/\delta_t)\right]\right\} \leq \sum_{t=1}^{\tau_{r_{k+1}}} \left[\alpha_t^{-2} \log(n/\delta_t)\right]$$
$$\leq c_{12} \Delta_{r_{k+1},r_k}^{-2} \left(\log(n/\delta) + \log\log |\Delta(r_k, r_{k+1})|^{-1}\right). \tag{17}$$

Note that for any item i in $U = \{r_1, r_2, \ldots, r_k\}$, $\Delta_i = \Delta_{i,r_{k+1}}$ and $\Delta_i = |\Delta(i,r_k)| + |\Delta(r_k, r_{k+1})|$, which implies that $\min\{\Delta(i,r_k)^{-1}, \Delta(r_k,r_{k+1})^{-1}\} \leq 2\Delta_i^{-1}$. Therefore, by eq. (16) and eq. (17), for any item i in U, we have

$$\mathbb{E}[N] = O\left(\Delta_i^{-2}(\log(n/\delta) + \log\log \Delta_i^{-1})\right). \tag{18}$$

Thus, by eq. (15) and eq. (18), and the definition of Δ_i's stated in eq. (2), we conclude that when $\bigcap_{t=1}^{T} \mathcal{E}_t$ happens,

$$\mathbb{E}[N] \leq c_9 \sum_{i\in[n]} \mathbb{E}\left\{\sum_{t=1}^{T_i} \left[\alpha_t^{-2} \log(n/\delta_t)\right]\right\} \tag{19}$$
$$= O\left(\sum_{i\in[n]} \left[\Delta_i^{-2}(\log(n/\delta) + \log\log \Delta_i^{-1})\right]\right). \tag{20}$$

This completes the proof of the sample complexity, and the proof of theorem 5 is complete. ■

A.7. Top-k Optimality

Theorem 2 (Optimality of Top-k Arms) *The Optimal Set of k-arms are the k-arms with the highest individual μ_i values, where $i \in [n]$.*

Proof Assume without loss of generality that $\mu_1 \geq \mu_2 \geq ...\mu_n$. Let a_i represent the i^{th} arm ordered in descending order, such that a_1 represents the best arm, a_k represents the worst arm in the top-k set, and a_{k+1} represents the best arm that wasn't good enough to be included in the top-k set. Hence the top-k set of arms is $S = \{a_i, \forall i \in [k]\}$. The expected reward of set S is $F(S)$.

If any arm $a_i \in S$, (a top-k arm) is replaced by a non top-k arm $a_j \in [n], j > k$, we know that ($\mu_i < \mu_j$). It follows from Combinatorial Monotonicity that, $F(S) > F(S_{\setminus\{i\}} \cup j)$. Where $S_{\setminus\{i\}}$ is the set S excluding the arm i.

Hence, replacing any top-k arm with a non-top-k arm will result in a lower expected reward, so the optimal set of arms must be the top-k μ_i arms. ■