Fast Pure Exploration via Frank-Wolfe

Po-An Wang
KTH Royal Institute of Technology
Stockholm, Sweden
wang9@kth.se

Ruo-Chun Tzeng
KTH Royal Institute of Technology
Stockholm, Sweden
rctzeng@kth.se

Alexandre Proutiere
EECS and Digital Futures
KTH, Stockholm, Sweden
alepro@kth.se

Abstract

We study the problem of active pure exploration with fixed confidence in generic stochastic bandit environments. The goal of the learner is to answer a query about the environment with a given level of certainty while minimizing her sampling budget. For this problem, instance-specific lower bounds on the expected sample complexity reveal the optimal proportions of arm draws an Oracle algorithm would apply. These proportions solve an optimization problem whose tractability strongly depends on the structural properties of the environment, but may be instrumental in the design of efficient learning algorithms. We devise Frank-Wolfe-based Sampling (FWS), a simple algorithm whose sample complexity matches the lower bounds for a wide class of pure exploration problems. The algorithm is computationally efficient as, to learn and track the optimal proportion of arm draws, it relies on a single iteration of Frank-Wolfe algorithm applied to the lower-bound optimization problem. We apply FWS to various pure exploration tasks, including best arm identification in unstructured, thresholded, linear, and Lipschitz bandits. Despite its simplicity, FWS is competitive compared to state-of-art algorithms.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

Contents

1 Introduction

Pure exploration in stochastic bandits [33] refers to the task of answering a given question about the reward distributions of the different arms, using as few arm pulls (or samples) as possible. The task may correspond to identifying the best arm [20], the top-m arms [49], all ϵ-good arms [36], a set of arms whose expected rewards exceed a given threshold [35], etc. To reduce the sample complexity of such a task, the learner needs to leverage as much as possible the information available about reward distributions, which typically comes as known structural properties of the set of their expected rewards. Exploiting particular structures (e.g., unimodal, Lipschitz, convex, linear) has been thoroughly studied in the regret minimization setting (see [10], and references therein), but less in the pure exploration framework, where most efforts have focused on linear structures [46, 27, 51, 47, 17, 25, 13].

In this paper, we investigate a generic learning problem proposed in [12] and covering the aforementioned pure exploration tasks with or without structure. Consider K arms whose reward distributions $(\nu_1, \dots, \nu_K)$ come from a one-dimensional exponential family and are of unknown means $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$. The parameter $\boldsymbol{\mu}$ is known to belong to $\Lambda \subset \mathbb{R}^K$, the set of possible instances. For each $\boldsymbol{\mu} \in \Lambda$, we assume that there is a unique true answer $i^\star(\boldsymbol{\mu})$ that belongs to the finite set $\mathcal{I}$ of possible answers[1] (e.g., for the best arm identification task, $i^\star(\boldsymbol{\mu}) = \arg\max_k \mu_k$). We consider pure exploration tasks in the *fixed confidence* setting where the learner wishes, for any possible $\boldsymbol{\mu} \in \Lambda$, to discover $i^\star(\boldsymbol{\mu})$ with a certain level of confidence $1-\delta$, for some $\delta \in (0,1)$. The learner's strategy is defined by (i) an adaptive sampling rule dictating the sequence of arm pulls, (ii) a stopping rule defining τ, the round where, based on the data gathered so far, the learner decides to stop pulling arms, and (iii) a decision rule specifying her answer. The goal is to devise a δ-PAC (it outputs the right answer with probability at least $1-\delta$ for any $\boldsymbol{\mu} \in \Lambda$) strategy minimizing the expected sample complexity $\mathbb{E}_{\boldsymbol{\mu}}[\tau]$.

Using the same arguments as those used in [20] for classical MAB problems, we may derive a lower bound of the expected sample complexity satisfied by any δ-PAC strategy. This lower bound, whose proof can be found in Appendix B for completeness, is given by $T^\star(\boldsymbol{\mu})\mathrm{kl}(\delta, 1-\delta)$, where the characteristic time $T^\star(\boldsymbol{\mu})$ is defined through the following optimization problem:

$$T^\star(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{\omega} \in \Sigma} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k), \tag{1}$$

where Σ is the $(K-1)$-dimensional simplex, $\mathrm{Alt}(\boldsymbol{\mu})$ is the set of *confusing* parameters $\boldsymbol{\lambda} \in \Lambda$ such that $i^\star(\boldsymbol{\mu}) \neq i^\star(\boldsymbol{\lambda})$, $\mathrm{kl}(a,b)$ is the KL divergence between two Bernoulli distributions of means a and b, and $d(\mu_k, \lambda_k)$ denotes the KL divergence of arm-k reward distributions under parameters $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$. A solution $\boldsymbol{\omega}^\star(\boldsymbol{\mu})$ of (1) can be interpreted as an optimal *allocation*, in the sense that pulling each arm i a proportion of round equal to $\boldsymbol{\omega}_i^\star(\boldsymbol{\mu})$ (in expectation) constitutes an optimal sampling rule.

Most existing algorithms achieving an asymptotically (when δ goes to 0) minimal sample complexity leverage a Track-and-Stop (TaS) framework [20]. In each round t, they plug $\hat{\boldsymbol{\mu}}(t)$ the estimated expected arm rewards in the lower bound optimization problem (1), and track the allocation $w^\star(\hat{\boldsymbol{\mu}}(t))$. As already noticed in [38], the main drawback of the Track-and-Stop framework is that it requires a recurrent access to an Oracle able to solve (1) (actually existing analyses usually assume that the Oracle outputs the exact solution for any $\boldsymbol{\mu}$). (1) is a concave program but can become difficult to solve depending the underlying structure Λ. Indeed, for complex structures, identifying the most confusing parameters leading to the objective function $\inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k)$ can be hard.

Contributions. 1) Instead of solving (1) in each round as in the TaS framework, we propose an online iterative method to approach the optimal allocation of arm pulls. Specifically, we devise Frank-Wolfe-based Sampling (`FWS`), a computationally efficient algorithm that just relies, in each round, on a single iteration Frank-Wolfe (FW) algorithm applied to (1) instantiated at $\hat{\boldsymbol{\mu}}(t)$.
2) For a wide class of pure exploration problems with or without structure, we derive an upper bound of the expected sample complexity of `FWS` for any certainty level δ, and show that this bound matches the lower bound $T^\star(\boldsymbol{\mu})\mathrm{kl}(\delta, 1-\delta)$ asymptotically as δ goes to 0.
3) We illustrate the performance of `FWS` on various pure exploration problems, including best arm identification in unstructured, linear, and Lipschitz bandits. In all tested scenarios, and despite its simplicity, `FWS` matches the performance of the best existing algorithms.

[1] Scenarios with several correct answers require a more involved analysis, see [11].

The use of the FW algorithm has been suggested in [20] in the case of best arm identification problem in unstructured bandits. In this case, FW iterations take a very simple and intuitive form (see Example 1 introduced in §3). The corresponding sampling rule is referred to as Best Challenger in [20], and leads to algorithms with remarkably low sample complexity empirically – sometimes lower than that of TaS algorithms solving (1) in each round. So far however, as discussed in [38], the analysis of FW-type sampling rules, and even their convergence, have eluded researchers. Towards the design of FWS algorithm, we devise a simple variant of the FW algorithm that yields a sampling rule whose sample complexity can be analyzed. We confirm the asymptotic optimality of as well as its empirical superiority, not only for the case of best arm identification in unstructured bandits as predicted by [20], but also for a wide class of pure exploration problems. We believe that our analysis also brings interesting solutions to the three important obstacles we needed to tackle to devise and analyze a FW-type sampling rule: (i) the objective function in (1) is not smooth; (ii) its curvature becomes infinite in general close to the boundary of Σ; and (iii) the estimate $\hat{\boldsymbol{\mu}}(t)$ is evolving and might be far from $\boldsymbol{\mu}$.

2 Related Work

Best Arm Identification (BAI) has recently received a lot of attention, either in unstructured bandit problems, see [20, 44], or in problems with various kinds of structure, e.g., linear [46, 27, 51, 47, 17, 25, 13, 42], combinatorial [32, 26, 43], spectral [30], monotone [21], cascading [54]. For BAI in unstructured bandits with fixed confidence, [20] developed the celebrated Track-and-Stop framework leading to algorithms able to asymptotically converge towards the optimal allocation of arm draws, and in turn, to achieve the lowest sample complexity possible in the high confidence regime (as δ goes to 0). It is possible to apply the TaS framework to specific structures, as this was proposed in [25] for linear bandits. However, for more involved structures, this might become computationally too difficult. Indeed TaS requires the learner to repeatedly solve the optimization problem (1).

The authors of [12] propose and exploit an interpretation of the lower bound optimization problem (1) as the solution of a 2-players game – the $\boldsymbol{\omega}$-player playing the 'sup' and the $\boldsymbol{\lambda}$-player playing the 'inf'. The algorithm presented in [12] combines two zero-regret algorithms applied sequentially by the two players, and converge to an optimal allocation. Interestingly, the algorithm uses the optimism in face of uncertainty principle to remove the need of forced exploration (the $\boldsymbol{\omega}$-player is fed with upper-confidence bounds on her rewards). As shown later, the algorithm does not perform as well as FWS. The applicability of the framework used in [12] remains unclear to us: in [13] and in [26], the authors claim that the framework cannot be applied to linear and combinatorial bandits, respectively.

In [38], the author proposes a solution close to ours. His algorithm, LMA (Lazy Miror Ascent), just runs in each round one iteration of a sub-gradient ascent algorithm applied to (1). Fortunately, the projection step usually involved in such algorithm is simple. Numerically, as illustrated later in the paper, we found that LMA may not be as efficient as TaS or FWS. We could try to explain this by remarking that LMA has similarities with the Exponential Weights algorithm (see Appendix F in [38]), an algorithm designed for adversarial online optimization problem, and may be too conservative in a stochastic setting.

As already mentioned in the introduction, FW-based algorithms for BAI in unstructured bandits have been mentioned first in [20] for their simplicity and good performance. Applying FW as if the objective function was smooth may fail at converging [38] experimentally. We believe that we manage to make, in our algorithm, the minimal modification of the FW-based algorithm so that convergence and asymptotic optimality are guaranteed. Finally note that [5] uses FW in a regret minimization problem but with a smooth objective function.

We conclude this section by mentioning existing works on the FW algorithm when applied to optimizing non-smooth functions. The proposed solutions consist by either smoothing objective function or enlarging the set of differential (this is the second approach we chose). [18, 22] apply FW on the randomly smoothed surrogate instead of the original non-smooth objective. However, computing the gradient at each iteration requires to query many time on the objective function, which may not be practical. [1, 40] use a proximal operator to replace the objective function, but as pointed out in [8], the smoothing parameters of the proximal operator are not trivial to tune. Our solution is close to those developed in [41, 8]. There, inspired by the approximate subdifferential [50], the authors propose to collect the set of the gradients in the neighborhood at each round. They show that

these collection is continuous even when the objective functions is non-smooth, which allows for the use of FW. The way we deal with the non-smoothness issue is similar but simplified by the fact that the specific form of our objective function.

3 Preliminaries

We consider the pure exploration task described in the introduction. This section presents the additional assumptions made towards the design and analysis of our algorithm. These assumptions are here illustrated for the classical Best Arm Identification (BAI) task in unstructured bandits (see Example 1); they will be verified for all other examples of pure exploration problems presented in Section 5. This section also provides useful properties of the lower bound optimization problem (1), and finally describes our choice of stopping and decision rules.

3.1 Assumptions and properties of the lower bound optimization problem

The answer map $i^\star : \Lambda \to \mathcal{I}$ allows us to decompose Λ into a union of non-overlapping sets: $\Lambda = \cup_{i\in\mathcal{I}}\mathcal{S}_i$, where $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : i^\star(\boldsymbol{\mu}) = i\}$ for all $i \in \mathcal{I}$. The answer map is known (i.e., knowing $\boldsymbol{\mu}$ is enough to output the right answer), and hence without loss of generality, we can assume that $\mathcal{S}_i \neq \emptyset$ for all $i \in \mathcal{I}$. Using this notation, the set of confusing parameters can be written as $\mathrm{Alt}(\boldsymbol{\mu}) = \cup_{i\neq i^\star(\boldsymbol{\mu})}\mathcal{S}_i$.

Assumption 1. *For each $i \in \mathcal{I}$, $\mathcal{S}_i$ is an open set and the complementary of $\mathcal{S}_i$ is a finite union of convex sets. Namely, there exists a finite collection $\mathcal{J}_i$ of convex sets $\mathcal{C}_j^i$ s.t. $\Lambda \setminus \mathcal{S}_i = \cup_{j\in\mathcal{J}_i}\mathcal{C}_j^i$.*

Example 1. The BAI task in unstructured bandits with Bernoulli rewards. For this task, we have $\Lambda = (0,1)^K, \mathcal{I} = \{1,\dots,K\}$, and for all arm i, the set of parameters for which arm i is the best arm is $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : \mu_i > \mu_k, \forall k \neq i\}$. We have: $\Lambda \setminus \mathcal{S}_i = \cup_{j\in\mathcal{J}_i}\mathcal{C}_j^i$ where $\mathcal{J}_i = \mathcal{I} \setminus \{i\}$ is the set of arms different than i and $\mathcal{C}_j^i = \{\boldsymbol{\mu} \in \Lambda : \mu_j > \mu_i\}$ is the convex set of parameters for which arm j is better than arm i. □

Now under Assumption 1, we can decompose the lower bound optimization problem as follows: $T^\star(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{\omega}\in\Sigma} F_{\boldsymbol{\mu}}(\boldsymbol{\omega})$ where $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \min_{j\in\mathcal{J}_{i^\star(\boldsymbol{\mu})}} f_j(\boldsymbol{\omega},\boldsymbol{\mu})$ and for all $j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$,

$$f_j(\boldsymbol{\omega},\boldsymbol{\mu}) = \inf_{\boldsymbol{\lambda}\in\mathcal{C}_j^{i^\star(\boldsymbol{\mu})}} \sum_{k=1}^K \omega_k d(\mu_k,\lambda_k). \tag{2}$$

Note that (2) is convex program (by convexity of the KL divergence), and that f_j is a concave function in $\boldsymbol{\omega}$ (as the minimum of concave functions). As a consequence, the objective function $F_{\boldsymbol{\mu}}$ is also concave, but not smooth. The following proposition summarizes insightful properties of the functions $f_j, j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$. It is a consequence of the envelope theorem and proved in Appendix K.2.

Proposition 1. *Let $i \in \mathcal{I}$, $j \in \mathcal{J}_i$. Define for all $(\boldsymbol{\omega},\boldsymbol{\mu}) \in \Sigma \times \mathcal{S}_i$,*

$$\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})} = \arg\min_{\boldsymbol{\lambda}\in\mathrm{cl}(\mathcal{C}_j^i)} \sum_{k=1}^K \omega_k d(\mu_k,\lambda_k), \tag{3}$$

where $\mathrm{cl}(\mathcal{C}_j^i)$ is the closure of $\mathcal{C}_j^i$. Then under Assumption 1, $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}$ is unique for all $(\boldsymbol{\omega},\boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$, where $\mathring{\Sigma}$ is the interior of Σ. In addition, f_j is continuously differentiable on $\mathring{\Sigma} \times \mathcal{S}_i$, and $\forall(\boldsymbol{\omega},\boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$,

$$\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega},\boldsymbol{\mu}) = \sum_{k=1}^K d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}_k)\boldsymbol{e}_k, \tag{4}$$

where $\boldsymbol{e}_k$ denotes the K-dimensional vector whose k-th coordinate is 1 and whose other coordinates are 0.

A key insight from the above result is that the objective function $F_{\boldsymbol{\mu}}$ is the minimum of a finite number of continuously differentiable functions. This observation will make the use of a slightly modifed FW algorithm possible (remember that the FW algorithm is known to converge for smooth

functions only). We use an additional assumption on the gradient and curvature of f_j. A controlled curvature is an essential ingredient when analyzing the convergence of FW-based algorithms, see e.g. [24]. Define $\Sigma_\gamma = \{\boldsymbol{\omega} \in \Sigma : \min_k \omega_k \geq \gamma\}$ for any $\gamma \in (0, 1/K)$. Following [24], we define $C_\psi(\mathcal{K})$, the curvature constant of the concave differentiable function $\psi : \mathcal{K} \to \mathbb{R}$ with respect to the compact set $\mathcal{K}$, as

$$C_\psi(\mathcal{K}) = \sup_{\substack{\boldsymbol{x},\boldsymbol{z} \in \mathcal{K} \\ \alpha \in (0,1] \\ \boldsymbol{y} = \boldsymbol{x} + \alpha(\boldsymbol{z} - \boldsymbol{x})}} \frac{1}{\alpha^2} \left[\psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) + \langle \boldsymbol{y} - \boldsymbol{x}, \nabla\psi(\boldsymbol{x}) \rangle\right]. \tag{5}$$

Refer to [24], for the intuition behind this defintion and examples.

Assumption 2. *For all $\boldsymbol{\mu} \in \Lambda$,*
(i) there exists $L > 0$ such that $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}, \omega \in \Sigma$, $\|\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\|_\infty \leq L$;
(ii) there exists $D > 0$ such that $\forall \gamma \in (0, 1/K)$ and $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, $C_{f_j(\cdot,\boldsymbol{\mu})}(\Sigma_\gamma) \leq \frac{D}{\gamma}$.

There is a simple way to verify whether a pure exploration problem satisfies Assumption 2, by looking at the second derivative of the function $y \mapsto d(x, y)$ at the points $(\mu_k, (\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})})_k)$ for all k. Refer to Appendix C for details.

Example 1 (cont'd). For unstructured bandits with Bernoulli rewards, we can easily compute f_j and its gradient [20, 38]: for all $j \neq i^\star(\boldsymbol{\mu})$ and all $\boldsymbol{\omega} \in \mathring{\Sigma}$, define $m_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \frac{\omega_{i^\star(\boldsymbol{\mu})}\mu_{i^\star(\boldsymbol{\mu})} + \omega_j \mu_j}{\omega_{i^\star(\boldsymbol{\mu})} + \omega_j}$. Then $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k = \mu_k$ if $k \notin \{i^\star(\boldsymbol{\mu}), j\}$ and $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k = m_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ otherwise. As a consequence:

$$\begin{cases} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \omega_{i^\star(\boldsymbol{\mu})} d(\mu_{i^\star(\boldsymbol{\mu})}, m_j(\boldsymbol{\omega}, \boldsymbol{\mu})) + \omega_j d(\mu_j, m_j(\boldsymbol{\omega}, \boldsymbol{\mu})), \\ \nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = d(\mu_{i^\star(\boldsymbol{\mu})}, m_j(\boldsymbol{\omega}, \boldsymbol{\mu}))\boldsymbol{e}_{i^\star(\boldsymbol{\mu})} + d(\mu_j, m_j(\boldsymbol{\omega}, \boldsymbol{\mu}))\boldsymbol{e}_j. \end{cases} \tag{6}$$

For this example, we can verify that Assumption 2 holds, either directly or using the tool described in Appendix C. □

3.2 Stopping and decision rules

Next we present the two last components of the FWS algorithm, namely the stopping and decision rules. These components are standard and borrowed from the existing literature. We need a few notations. For any $t \geq 1$, let A_t denote the arm selected in round t. Define $N_k(t) = \sum_{s=1}^t \mathbb{1}\{A_s = k\}$ the number of times arm k has been selected up to round t, and by $\omega_k(t) = N_k(t)/t$ the corresponding empirical proportion of draw. When $N_k(t) > 0$, the empirical average reward of arm k up to round t is denoted by $\hat{\mu}_k(t) = \sum_{s=1}^t X_k(s)\mathbb{1}\{A_s = k\}/N_k(t)$, where $X_k(s)$ is the random reward received from pulling arm k in round s.

Let us denote by τ, the stopping time defining when the algorithm stops exploring and has to output a decision. Our decision rule is obviously to output the best empirical answer: $\hat{\imath}_\tau = i^\star(\hat{\boldsymbol{\mu}}(\tau))$.

For the stopping rule, as in other existing algorithms, we leverage a Generalized Likelihood Ratio Test (GLRT). Our test boils down to comparing $tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t))$ to a threshold $\beta(t, \delta)$ (recall that $F_{\boldsymbol{\mu}}$ is the objective function of the lower bound optimization problem):

$$\tau = \inf\{t \geq 1 : tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq \beta(t, \delta)\}. \tag{7}$$

Many thresholds $\beta(t, \delta)$ have been proposed in the literature [29, 20, 25, 38]. For FWS and its analysis, we just need that the threshold statisfies the two following properties:

$$\forall t \geq 1, \quad \left(tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq \beta(t, \delta)\right) \Longrightarrow \left(\mathbb{P}_{\boldsymbol{\mu}}\left[i^\star(\hat{\boldsymbol{\mu}}(t)) \neq i^\star(\boldsymbol{\mu})\right] \leq \delta\right), \tag{8}$$

$$\exists c_1(\Lambda), c_2(\Lambda) > 0 \ : \ \forall t \geq c_1(\Lambda), \ \ \beta(t, \delta) \leq \log\left(\frac{c_2(\Lambda)t}{\delta}\right). \tag{9}$$

The first of the above properties will naturally imply that FWS returns the true answer with probability at least $1-\delta$ when stopping, whereas the second will be instrumental in the sample complexity analysis (there, $c_1(\Lambda), c_2(\Lambda)$ may depend on the set of possible instances, and on the reward distributions). In [29], the authors manage to provide, for any generic pure exploration task, a single threshold satisfying (8)-(9)). Unless otherwise mentioned, we will use the stopping rule implementing this threshold.

4 The FWS Algorithm and its Sample Complexity

In the FWS algorithm, we use the FW algorithm to learn an optimal allocation $\boldsymbol{\omega}^\star(\boldsymbol{\mu})$. In each round, an iteration of FW updates the allocation that the FWS algorithm aims at approaching using some tracking procedure. We describe this learning and tracking procedure below.

4.1 Adapting Frank-Wolfe to the non-smooth function $F_{\boldsymbol{\mu}}$

The FW algorithm [19] solves smooth convex programs by linearizing, in each iteration, the objective function and moving towards a minimizer of this linear function. Compared to the projected gradient and proximal methods, FW is computationally more efficient (e.g. it avoids the projection step), and is particularly well-suited when optimizing over polyhedra [7] (which is our case here). For a contemporary treatment of FW, refer to [24]. FW was suggested in [20] for BAI in unstructured bandits to update the allocation to be tracked. For this BAI problem, an iteration of the FW algorithm takes an intuitive form (see also Appendix A2 in [38]):

Example 1 (cont'd). For BAI in unstructured bandits, the optimal allocation $\boldsymbol{\omega}^\star(\boldsymbol{\mu})$ is the maximizer of the function $\boldsymbol{\omega} \mapsto F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \min_j f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$. $F_{\boldsymbol{\mu}}$ is smooth at points when the minimum is realized at a single arm $j^\star = \arg\min_j f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$, and there, in view of (6), its gradient is $\nabla F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = d(\mu_{i^\star(\boldsymbol{\mu})}, m_{j^\star}(\boldsymbol{\omega}, \boldsymbol{\mu}))\boldsymbol{e}_{i^\star(\boldsymbol{\mu})} + d(\mu_{j^\star}, m_{j^\star}(\boldsymbol{\omega}, \boldsymbol{\mu}))\boldsymbol{e}_{j^\star}$. Now in an iteration of the FW algorithm, one would follow the direction given by $\arg\max_{\boldsymbol{\omega}' \in \Sigma} \boldsymbol{\omega}'^\top \nabla F_{\boldsymbol{\mu}}(\boldsymbol{\omega})$. This direction is $\boldsymbol{e}_{j^\star}$ if $d(\mu_{j^\star}, m_{j^\star}(\boldsymbol{\omega}, \boldsymbol{\mu})) > d(\mu_{i^\star(\boldsymbol{\mu})}, m_{j^\star}(\boldsymbol{\omega}, \boldsymbol{\mu}))$, and $\boldsymbol{e}_{i^\star(\boldsymbol{\mu})}$ otherwise. This is precisely what the FW-type sampling rule suggested in [20] is doing: in round $(t+1)$, the best challenger is defined as $j^\star = \arg\min_j f_j(\boldsymbol{\omega}(t), \hat{\boldsymbol{\mu}}(t))$, and the arm selected corresponds to the direction given by $\arg\max_{\boldsymbol{\omega}' \in \Sigma} \boldsymbol{\omega}'^\top \nabla F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t))$, i.e., it is either the best challenger $j^\star$ or the best empirical arm $i^\star(\hat{\boldsymbol{\mu}}(t))$. □

The convergence analysis of FW usually requires that the objective function is smooth, and that its curvature can be controlled. When applying FW-type algorithms to design an optimal sampling rule (a rule that converges to the allocation $\boldsymbol{\omega}^\star(\boldsymbol{\mu})$ maximizing $F_{\boldsymbol{\mu}}$), we face three issues: (i) $F_{\boldsymbol{\mu}}$ is not smooth; (ii) $F_{\boldsymbol{\mu}}$ has an unbounded curvature close to the boundary of Σ; (iii) $\boldsymbol{\mu}$ is unknown initially, so the FW iteration in round t can be applied to $F_{\hat{\boldsymbol{\mu}}(t)}$ only. We discuss below how we circumvent these issues in the design of our algorithm.

(i) Non-smoothness of $F_{\boldsymbol{\mu}}$. In view of Proposition 1, $F_{\boldsymbol{\mu}}$ is the minimum of a finite number of smooth concave functions f_j. Hence at points where two of these functions are equal in $\boldsymbol{\omega}$, $F_{\boldsymbol{\mu}}$ is not differentiable in $\boldsymbol{\omega}$. The FW algorithm has been adapted to cope with non-smooth functions, see e.g. [41]. Typically, one constructs continuous approximations of the gradient close to non-smooth points of the functions. This construction often involves the r-subdifferential [23][2], which would be too costly to compute for $F_{\boldsymbol{\mu}}$. Instead, we can leverage the fact $F_{\boldsymbol{\mu}}$ is the minimum of concave functions, and construct the called *r-subdifferential subspace*: for $r \in (0, 1)$,

$$H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r) = \text{cov}\left\{\nabla f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) : j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}, f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) < F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) + r\right\}, \tag{10}$$

where $\text{cov}\{S\}$ denotes the convex hull of the set S. This choice greatly simplifies because it does not require to compute the gradient of f_j in a neighborhood of $\boldsymbol{\omega}$. Since the f_j are continuously differentiable, we can prove that $\boldsymbol{\omega} \mapsto H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r)$ is a continuous (i.e. upper- and lower-hemicontinuous). Using the r-subdifferential subspace, the modified FW update is given as follows. Let $\boldsymbol{x}(t)$ be the estimated optimizer of $F_{\boldsymbol{\mu}}$ in round t. In round $(t+1)$, it is updated as:

$$\begin{cases} \boldsymbol{z}(t+1) = \text{argmax}_{\boldsymbol{z} \in \Sigma} \min_{h \in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}(t), r_t)} \langle \boldsymbol{z} - \boldsymbol{x}(t), h \rangle \quad \textit{(ties broken arbitrarily)}, \\ \boldsymbol{x}(t+1) = \frac{t}{t+1}\boldsymbol{x}(t) + \frac{1}{t+1}\boldsymbol{z}(t+1). \end{cases} \tag{11}$$

Of course in the FWS algorithm, $\boldsymbol{\mu}$ is unknown, and will be simply replaced by $\hat{\boldsymbol{\mu}}(t)$ in the above update. The way we choose the sequence of parameters $\{r_t\}_{t \geq 1}$ will be discussed later. Computing $\boldsymbol{z}(t)$ is equivalent to solving a zero-sum game, which can be further formulated as a LP [52] (Chapter 20). Refer to Appendix H for a detailed description of this LP.

(ii) Unbounded curvature of $F_{\boldsymbol{\mu}}$ and (iii) unknown $\boldsymbol{\mu}$. These two issues are solved by a single trick. We impose that in the FW iterations, the update directions $\boldsymbol{z}(t)$ cover all $\boldsymbol{e}_k, k = 1, \ldots, K$

[2] For $r \in (0, 1)$, the r-subdifferential of $\psi : \mathcal{K} \to \mathbb{R}$ (where $\mathcal{K} \subset \mathbb{R}^K$ is compact and convex) is defined as $\partial_r \psi(\boldsymbol{x}) = \{h \in \mathbb{R}^K : \psi(\boldsymbol{y}) < \psi(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, h \rangle + r \text{ for all } \boldsymbol{y} \in \mathcal{K}\}$.

sufficiently often. This ensures that the target allocation $\boldsymbol{x}(t)$ stays away from the boundary of Σ, which in turn allows us to control the curvature of $F_{\hat{\boldsymbol{\mu}}(t)}$ thanks to Assumption 2. This imposed constraint can be seen as a sort of forced exploration, and further implies (thanks to our tracking procedure) that each arm is played often enough. Now, with this kind of forced exploration, $\hat{\boldsymbol{\mu}}(t)$ will concentrate around the true $\boldsymbol{\mu}$.

4.2 Algorithm

The FWS algorithm proceeds as follows. FWS maintains a target allocation, denoted by $\boldsymbol{x}(t)$, its empirical allocation $\boldsymbol{\omega}(t)$, and the empirical average rewards $\hat{\boldsymbol{\mu}}(t)$ after round t. After an initialization phase (K rounds where each arm is selected), FWS alternates between forced exploration and FW updates. More precisely:
Forced exploration occurs at rounds t where $\sqrt{\lfloor t/K \rfloor}$ is an integer and at those where $\hat{\boldsymbol{\mu}}(t-1) \notin \Lambda$ (in this case, we cannot compute the objective function). In forced exploration round t, the target allocation is updated towards the center of the simplex: $\boldsymbol{x}(t) = \frac{t-1}{t}\boldsymbol{x}(t-1) + \frac{1}{t}(1/K, \dots, 1/K)$.
FW updates happen in other rounds. There, the target allocation is updated according to our adapted version of FW (11), where in round t the unknown $\boldsymbol{\mu}$ is replaced by $\hat{\boldsymbol{\mu}}(t-1)$. In the successive FW updates, we use r-subdifferential subspaces with varying parameter r. For the analysis of FWS, we will select a sequence of parameters $\{r_t\}_{t\geq 1}$ with an appropriate decay rate.

After the target allocation is updated in round t, the algorithm tracks this allocation by selecting the arm maximizing over k the ratio $x_k(t)/\omega_k(t-1)$. Finally, FWS, whose pseudo-code is presented below, uses the stopping and decision rules described in §3.2.

Algorithm 1: FWS algorithm

Input: Confidence level δ, sequence $\{r_t\}_{t\geq 1}$
Initialization: Sample each arm once and update $\boldsymbol{\omega}(K)$, $\boldsymbol{x}(K) = (\frac{1}{K}, \dots, \frac{1}{K})$, and $\hat{\boldsymbol{\mu}}(K)$
$t \leftarrow K$
While $(tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) < \beta(\delta, t)$ or $\hat{\boldsymbol{\mu}}(t-1) \notin \Lambda)$
  $t \leftarrow t+1$
  If $(\sqrt{\lfloor t/K \rfloor} \in \mathbb{N}$ or $\hat{\boldsymbol{\mu}}(t-1) \notin \Lambda)$ (*forced exploration*) $\boldsymbol{z}(t) \leftarrow (\frac{1}{K}, \dots, \frac{1}{K})$
  Else (*FW update*)

$$\boldsymbol{z}(t) \leftarrow \operatorname*{argmax}_{\boldsymbol{z}\in\Sigma} \min_{h \in H_{F_{\hat{\boldsymbol{\mu}}(t-1)}}(\boldsymbol{x}(t-1), r_t)} \langle \boldsymbol{z} - \boldsymbol{x}(t-1), h \rangle \quad \textit{(ties broken arbitrarily)}$$

  Update $\boldsymbol{x}(t) \leftarrow \frac{t-1}{t}\boldsymbol{x}(t-1) + \frac{1}{t}\boldsymbol{z}(t)$
  Sample the arm $A_t \leftarrow \operatorname{argmax}_k x_k(t)/\omega_k(t-1)$ (*ties broken arbitrarily*)
  Update $\boldsymbol{\omega}(t)$ and $\hat{\boldsymbol{\mu}}(t)$
Output: $i^\star(\hat{\boldsymbol{\mu}}(t))$

4.3 Sample complexity

In the following theorem, we establish the asymptotic optimality of FWS.

Theorem 1. *Consider the* FWS *algorithm with a sequence* $\{r_t\}_{t\geq 1}$ *of strictly positive reals satisfying (i)* $\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} r_s = 0$, *and (ii)* $\lim_{t\to\infty} t r_t = \infty$. *Under Assumptions 1, 2, the algorithm terminates in finite time almost surely and is* δ*-PAC. Its sample complexity* τ *satisfies:*

$$\forall \boldsymbol{\mu} \in \Lambda, \ \mathbb{P}_{\boldsymbol{\mu}}\left[\limsup_{\delta\to 0} \frac{\tau}{\log(1/\delta)} \leq T^\star(\boldsymbol{\mu})\right] = 1, \ \textit{and} \ \limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau]}{\log(1/\delta)} \leq T^\star(\boldsymbol{\mu}).$$

The proof is given in Appendix I. We sketch the proof of the guarantees in expectation. The proof relies on classical concentration results, but more critically combines continuity arguments (developed in Appendix K) to account for the varying $\hat{\boldsymbol{\mu}}(t)$, and tools to analyze the convergence of the modified FW algorithm (reported in Appendix L).
1. First using concentration inequalities and the fact that FWS includes forced exploration rounds, we can define, for round t, a "good" event $\mathcal{E}_t$ under which $\hat{\boldsymbol{\mu}}(t)$ is very close to $\boldsymbol{\mu}$ and such that

$\sum_{t=1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_t^c] < \infty$. Then, several continuity arguments have to be made. In Lemma 6 (Appendix K) we show that $\boldsymbol{\mu} \mapsto F_{\boldsymbol{\mu}}$ is continuous (w.r.t. the uniform convergence norm). In Theorem 3 (Appendix K) we also prove that the solution $\boldsymbol{z}(t+1)$ of the FW update (11) is continuous in $\boldsymbol{\mu}$. The arguments above allow us to analyze the convergence of the FW updates almost as if $\hat{\boldsymbol{\mu}}(t)$ was replaced by $\boldsymbol{\mu}$ provided that the event $\mathcal{E}_t$ occurs.
2. Now we can study under the event $\mathcal{E}_t$, the impact of the FW update on the target allocation. The main step of our proof is Theorem 6 (Appendix L) characterizing how $F_{\boldsymbol{\mu}}(\boldsymbol{x}(t))$ get closer to $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))$ in each FW update. We then deduce that after a time T_1, $F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{x}(t))$ is a good approximation of $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))$.
3. We conclude the proof using similar arguments as those in [20]. According to our stopping rule, $t > \tau$ if and only if $tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) > \beta(t,\delta)$. Hence $\mathbb{E}_{\boldsymbol{\mu}}[\tau] = \sum_{t=1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[\tau > t] = \sum_{t=1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}\left[tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \leq \beta(t,\delta)\right]$ which can be approximately upper bounded by $T_1 + \sum_{t=T_1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}_t^c] + \sum_{t=1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[tF_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) \leq \beta(t,\delta)]$. The proof is concluded by remarking that in view of the property (9) of our stopping threshold, the last sum is close to $T^\star(\boldsymbol{\mu})\log(1/\delta)$ as $\delta \to 0$.
Note that our proof of Theorem 1 accounts for the possibility in certain structures (e.g. linear) of having multiple optimal allocations (these allocations form a convex set). We just reason in terms of the objective function (as in [25] for linear bandits).

Under the following additional assumption, we can derive non-asymptotic sample complexity upper bound for `FWS`. The proof of the following theorem is presented in Appendix N.

Assumption 3. *For any $\boldsymbol{\mu} \in \Lambda$, there exist constants $\kappa, E > 0$, s.t. if $\|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty \leq \kappa$, then $\boldsymbol{\pi} \in \mathcal{S}_{i^\star(\boldsymbol{\mu})}$, $\forall \boldsymbol{\omega} \in \mathring{\Sigma}$, $j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, $\nabla_{\boldsymbol{\pi}} d(\pi_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\pi})}_k)$ is continuous and $\left\|\nabla_{\boldsymbol{\pi}} d(\pi_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\pi})}_k)\right\|_1 \leq E$, $\forall k = 1,\dots,K$.*

Theorem 2. *Consider the `FWS` algorithm with a sequence $\{r_t\}_{t\geq 1}$ as in Theorem 1. Under Assumptions 1, 2, and 3, the sample complexity τ of the algorithm satisfies: for any $\boldsymbol{\mu} \in \Lambda, \delta \in (0,1)$, and any $\epsilon < \min\{\kappa E/2, 1\}$, $\tilde{\epsilon} < 1$,*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \leq \frac{1+\tilde{\epsilon}}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon}\left[\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)e}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon)}\right) + \log\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon)}\right)\right] + \Psi(K,D,E,L,c_1(\Lambda),\epsilon) + T_{\epsilon,L}^{\frac{5}{4}},$$

where $T_{\epsilon,L}$ is a constant such if $t \geq T_{\epsilon,L}$, then $\sum_{s=1}^{t} r_s < t\epsilon$ and $tr_t > L$. The constant Ψ is polynomial in $(D,E,L,c_1(\Lambda),1/\epsilon)$ and exponential in K. The precise definition of Ψ is given in Appendix N.

5 Examples and Experiments for Linear Bandits

5.1 Examples

Our framework can be applied to many pure exploration problems, including BAI in unstructured (see Example 1), linear, Lipschitz bandits. It further covers threshold bandits (the problem of identifying all arms with rewards greater than a threshold), linear threshold bandits, top-m bandits (where we wish to identify the best m arms), and dueling bandits. All these examples are presented in Appendix. Using numerical experiments, we show that `FWS` is competitive with state-of-the-art algorithms for BAI in unstructured, linear, and Lipschitz bandits, see Appendices D-E-F, respectively. To the best of our knowledge, we report the first results for BAI in Lipschitz bandits. We quote some of our results for BAI in linear bandits below.

When facing a new pure exploration problem, one can check whether it falls into our framework, by first directly verifying Assumption 1. In Appendix C, we provide a simple sufficient condition ensuring that Assumption 2 holds, and explain why all the aforementioned pure exploration problems satisfy this condition.

5.2 BAI in linear bandits

Linear bandits constitute arguably the most popular and important bandit problems with structure, and have found many applications [34, 9]. BAI in linear bandits has received a lot of attention recently,

see §2. To model linear bandits, we slightly modify our framework. The reason for this modification is that the linear structure is so strong that using our initial framework, the set Λ would be small, and we would have problems ensuring that $\hat{\boldsymbol{\mu}}(t) \in \Lambda$ after some reasonable time t. Alternatively (rather than modifying the framework), we could modify the FWS algorithm so that $\hat{\boldsymbol{\mu}}(t)$ is projected onto Λ.

Consider a set of K arms. Arm k is attached a d-dimensional feature vector $\boldsymbol{a}_k$ and its average reward $\langle \boldsymbol{a}_k, \boldsymbol{\mu} \rangle$, where $\boldsymbol{\mu} \in \mathbb{R}^d$ is unknown. Without loss of generality, we assume that $\{\boldsymbol{a}_k\}_{k \in [K]}$ spans $\mathbb{R}^d$. We modify the definition of Λ as follows: $\Lambda = \left\{ \boldsymbol{\mu} \in \mathbb{R}^d : \exists k \in [K] \text{ s.t.} \langle \boldsymbol{a}_k - \boldsymbol{a}_i, \boldsymbol{\mu} \rangle > 0, \forall i \neq k \right\}$. Hence $\boldsymbol{\mu}$ parametrizes the average rewards of the arms, but μ_k is not the average reward of arm k. The true answer is $i^\star(\boldsymbol{\mu}) = \operatorname{argmax}_k \langle \boldsymbol{a}_k, \boldsymbol{\mu} \rangle$. The lower bound optimization problem (1) becomes: $\sup_{\boldsymbol{\omega} \in \Sigma} F_{\boldsymbol{\mu}}(\boldsymbol{\omega})$ where $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\lambda})^\top \sum_k \omega_k \boldsymbol{a}_k \boldsymbol{a}_k^\top (\boldsymbol{\mu} - \boldsymbol{\lambda})$ and $\mathrm{Alt}(\boldsymbol{\mu}) = \{ \boldsymbol{\lambda} \in \Lambda : \exists k \neq i^\star(\boldsymbol{\mu}) \text{ s.t.} \langle \boldsymbol{a}_k - \boldsymbol{a}_{i^\star(\boldsymbol{\mu})}, \boldsymbol{\lambda} \rangle > 0 \}$, see e.g. [25]. From there, we can reproduce our framework: for Assumption 1, for all $j \neq i^\star(\boldsymbol{\mu})$, $\mathcal{C}_j^{i^\star(\boldsymbol{\mu})} = \{ \boldsymbol{\lambda} \in \Lambda : \langle \boldsymbol{a}_j - \boldsymbol{a}_{i^\star(\boldsymbol{\mu})}, \boldsymbol{\lambda} \rangle > 0 \}$; as for the functions f_j, they are defined through:

$$\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} = \boldsymbol{\mu} + \left(\frac{\langle \boldsymbol{a}_{i^\star(\boldsymbol{\mu})} - \boldsymbol{a}_j, \boldsymbol{\mu} \rangle}{\left\| \boldsymbol{a}_{i^\star(\boldsymbol{\mu})} - \boldsymbol{a}_j \right\|^2_{V_{\boldsymbol{\omega}}^{-1}}} V_{\boldsymbol{\omega}}^{-1} \right) \left(\boldsymbol{a}_j - \boldsymbol{a}_{i^\star(\boldsymbol{\mu})} \right), \tag{12}$$

where $V_{\boldsymbol{\omega}} = \sum_k \omega_k \boldsymbol{a}_k \boldsymbol{a}_k^\top$. In the FWS algorithm for linear bandits, we use the Least-Squares Estimator (LSE) $\hat{\boldsymbol{\mu}}(t)$ given past observations, see [25] or Appendix E for an explicit expression. It can be readily seen that this slight modification of our framework does not affect the validity of Theorem 1. We just need to use the concentration inequalities derived in [25] for $\hat{\boldsymbol{\mu}}(t)$ in the first step of its proof.

Numerical experiments. We consider the example proposed by [46]. The unknown parameter $\boldsymbol{\mu} = \boldsymbol{e}_1$ and there are $d+1$ arms, $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d, \cos(\phi)\boldsymbol{e}_1 + \sin(\phi)\boldsymbol{e}_2$ in $\mathbb{R}^d$, where $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d)$ form the standard orthonormal basis. We set $d = 6$ and $\phi = 0.1$. To assess the performance of the FWS algorithm, we compare with the following algorithms: the Lazy Track and Stop algorithm (LT) from [25]; LineGame-C (CG-C) and LineGame (Lk-C) from [13] and implemented by [45]; the XY-Adaptive algorithm (XY-A) from [46]. For information, we also run the Round Robin algorithm RR selecting each equally. For comparison, we finally compute the sample complexity lower bound $\mathrm{LB}_{\mathtt{lin}}(\delta)$ (equal to $T^\star(\boldsymbol{\mu})\mathrm{kl}(\delta, 1-\delta)$).

Except for XY-A, all algorithms implement the same stopping rule defined in (7) with threshold $\beta(t, \delta) = \log((\log(t) + 1)/\delta)$ (this threshold was initially suggested in [20], and is also used in [45] for CG-C and Lk-C). For XY-A, we use the stopping rule advocated in the corresponding papers. Refer to Appendix E for the detailed implementations.

In Table 1, we present the sample complexity (the number of samples gathered before the algorithm stops) averaged over 1000 runs for the various algorithms and for different confidence levels $\delta \in \{0.1, 0.01, 0.001, 0.0001\}$. In Appendix E, we provide detailed results, e.g. including box-plots (to show how confident we are about the values displayed in Table 1), as well as the empirical allocations achieved under the various algorithms.

Table 1: Sample complexity for the linear bandit benchmark example of [46], averaged over 1000 runs. Refer to Appendix E for details, including box-plots.

	FWS	LT	CG-C	Lk-C	XY-A	RR	$\mathrm{LB}_{\mathtt{lin}}(\delta)$
$\delta = 0.1$	1 030	919	2 498	2 319	7 016	5 451	359
$\delta = 0.01$	1 614	1 464	3 501	3 431	7 779	8 814	920
$\delta = 0.001$	2 229	1 982	4 324	4 326	9 090	12 101	1 408
$\delta = 0.0001$	2 839	2 518	5 118	5 120	9 723	15 314	1 881

6 Conclusion

We have developed FWS, a computationally and statistically efficient algorithm for active pure exploration in bandit problems with fixed confidence. In each round, FWS performs a single iteration of a modified FW algorithm to approach an optimal allocation of arm draws predicted by the

asymptotic lower bound. In the FWS algorithm, the FW iterations aim at maximizing a non-smooth function. Our main contribution is here to adapt the design of FW so that its convergence can be analyzed even for this non-smooth function. FW-based pure exploration algorithms have been discussed in the literature, with the belief that they would perform well. We confirm this belief, and even establish the asymptotic optimality of FWS in wide class of pure exploration problems.

Many interesting research directions could be investigated. Our analysis of the sample complexity in the moderate confidence regime has the advantage of being applicable to generic pure exploration problems, but may not be always tight. For bandits with specific structures, we may refine the analysis in this regime to get better upper bounds. We are also interested in investigating whether the iterative approach used in the FWS algorithm can be extended to more complex problems such as learning an optimal policy in MDPs, as well as to regret minimization problems. There, instance-specific regret lower bounds and the corresponding optimal exploration process are characterized by the solution of an optimization problem, just as in pure exploration problems.

Acknowledgments and Disclosure of Funding

The authors would like to thank the anonymous reviewers whose comments helped us to improve the manuscript. R.-C Tzeng is supported by ERC Advanced Grant REBOUND (834862). A. Proutiere's research is supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. This work was also in part financially supported by Digital Futures.

References

[1] Andreas Argyriou, Marco Signoretto, and Johan Suykens. Hybrid conditional gradient-smoothing algorithms with applications to sparse and low rank regularization. *Regularization, Optimization, Kernels, and Support Vector Machines*, 2014.

[2] Shane Barratt. On the differentiability of the solution to convex optimization problems. *arXiv preprint arXiv:1804.05098*, 2018.

[3] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

[4] Claude Berge. *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1997.

[5] Quentin Berthet and Vianney Perchet. Fast rates for bandit optimization with upper-confidence frank-wolfe. In *Proc. of NeurIPS*, 2017.

[6] KC Border. Miscellaneous notes on optimization theory and related topics. 2009.

[7] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky. The convex geometry of linear inverse problems. 2012.

[8] Edward Cheung and Yuying Li. Solving separable nonsmooth problems using frank-wolfe with uniform affine approximations. In *Proc. of IJCAI*, 2018.

[9] Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proc. of AISTATS*, 2011.

[10] Richard Combes, Stefan Magureanu, and Alexandre Proutiere. Minimal exploration in structured stochastic bandits. In *Proc. of NeurIPS*, 2017.

[11] Rémy Degenne and Wouter M Koolen. Pure exploration with multiple correct answers. In *Proc. of NeurIPS*, 2019.

[12] Rémy Degenne, Wouter M Koolen, and Pierre Ménard. Non-asymptotic pure exploration by solving games. In *Proc. of NeurIPS*, 2019.

[13] Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *Proc. of ICML*, 2020.

[14] Asen L Dontchev and R Tyrrell Rockafellar. *Implicit functions and solution mappings*. Springer, 2009.

[15] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*. SIAM, 1999.

[16] Eugene A Feinberg, Pavlo O Kasyanov, and Mark Voorneveld. Berge's maximum theorem for noncompact image sets. *Journal of Mathematical Analysis and Applications*, 2014.

[17] Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. In *Proc. of NeurIPS*, 2019.

[18] Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proc. of SODA*, 2005.

[19] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 1956.

[20] Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proc. of COLT*, 2016.

[21] Aurélien Garivier, Pierre Ménard, Laurent Rossi, and Pierre Menard. Thresholding bandit for dose-ranging: The impact of monotonicity. *arXiv preprint arXiv:1711.04454*, 2017.

[22] Elad Hazan and Satyen Kale. Projection-free online learning. In *Proc. of ICML*, 2012.

[23] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*. Springer science & business media, 2013.

[24] Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *Proc. of ICML*, 2013.

[25] Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. In *Proc. of NeurIPS*, 2020.

[26] Marc Jourdan, Mojmír Mutnỳ, Johannes Kirschner, and Andreas Krause. Efficient pure exploration for combinatorial bandits with semi-bandit feedback. In *Proc. of ALT*, 2021.

[27] Zohar S Karnin. Verification based solution for structured mab problems. In *Proc. of NeurIPS*, 2016.

[28] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *JMLR*, 2016.

[29] Emilie Kaufmann and Wouter Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *arXiv preprint arXiv:1811.11419*, 2018.

[30] Tomáš Kocák and Aurélien Garivier. Best arm identification in spectral bandits. In *Proc. of IJCAI*, 2020.

[31] Wouter Koolen. tidnabbil: Julia library for structured bandit models. https://bitbucket.org/wmkoolen/tidnabbil/src/master/, 2021. [Online; accessed 09-May-2021].

[32] Yuko Kuroki, Junya Honda, and Masashi Sugiyama. Combinatorial pure exploration with full-bandit feedback and beyond: Solving combinatorial optimization under uncertainty with limited observation. *arXiv preprint arXiv:2012.15584*, 2020.

[33] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 1985.

[34] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proc. of WWW*, 2010.

[35] Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *Proc. of ICML*, 2016.

[36] Blake Mason, Lalit Jain, Ardhendu Tripathy, and Robert Nowak. Finding all ϵ-good arms in stochastic bandits. In *Proc. of NeurIPS*, 2020.

[37] Jiri Matousek and Bernd Gärtner. *Understanding and using linear programming*. Springer Science & Business Media, 2007.

[38] Pierre Ménard. Gradient ascent for active exploration in bandit problems. *arXiv*, 2019.

[39] Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 2002.

[40] Federico Pierucci, Zaid Harchaoui, and Jérôme Malick. *A smoothing approach for composite conditional gradient with nonsmooth loss*. PhD thesis, INRIA Grenoble, 2014.

[41] Sathya N Ravi, Maxwell D Collins, and Vikas Singh. A deterministic nonsmooth frank wolfe algorithm with coreset guarantees. *Informs Journal on Optimization*, 2019.

[42] Clémence Réda, Emilie Kaufmann, and Andrée Delahaye-Duriez. Top-m identification for linear bandits. In *Proc. of AISTATS*, 2021.

[43] Idan Rejwan and Yishay Mansour. Top-k combinatorial bandits with full-bandit feedback. In *Proc. of ALT*, 2020.

[44] Daniel Russo. Simple bayesian algorithms for best arm identification. In *Annual Conference on Learning Theory*. PMLR, 2016.

[45] Xuedong Shang. Linbai: Gamification of pure exploration for linear bandits. https://github.com/xuedong/LinBAI.jl, 2021. [Online; accessed 09-May-2021].

[46] Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. In *Proc. of NeurIPS*, 2014.

[47] Chao Tao, Saúl Blanco, and Yuan Zhou. Best arm identification in linear bandits with linear dimension dependency. In *Proc. of ICML*, 2018.

[48] Robert J Vanderbei et al. *Linear programming*. Springer, 2015.

[49] Tengyao Wang, Nitin Viswanathan, and Sébastien Bubeck. Multiple identifications in multi-armed bandits. In *Proc. of ICML*, 2013.

[50] DJ White. Extension of the frank-wolfe algorithm to concave nondifferentiable objective functions. *Journal of optimization theory and applications*, 1993.

[51] Liyuan Xu, Junya Honda, and Masashi Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *Proc. of AISTATS*, 2018.

[52] Petyon Young and Shmuel Zamir. *Handbook of game theory*. Elsevier, 2014.

[53] Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

[54] Zixin Zhong, Wang Chi Cheung, and Vincent Tan. Best arm identification for cascading bandits in the fixed confidence setting. In *Proc. of ICML*, 2020.

A Table of Notations

Setting: pure exploration task	
K	Number of arms
$[m]$ for any $m \in \mathbb{N}$	The set $\{1, 2 \dots, m\}$
ν_k	Reward distribution for arm k
$X_k(t)$	Random reward received from pulling arm k in round t
$\boldsymbol{\mu} \in \mathbb{R}^K$	Vector of the expected rewards of the various arms
Λ	Set of all possible parameters $\boldsymbol{\mu}$
$\mathcal{I}$	Set of the answers
$i^\star(\boldsymbol{\mu})$	Correct answer for parameter $\boldsymbol{\mu}$
$\mathcal{S}_i$	Set of parameters for which i is the correct answer
δ	Targeted confidence level
Lower bound properties	
$\boldsymbol{\omega}$	Vector of the proportions of arm draws
Σ	Simplex
$\mathring{\Sigma}$	Interior of Σ
Σ_γ	$\{\boldsymbol{\omega} \in \Sigma : \min_k \omega_k \geq \gamma\}$
$\boldsymbol{e}_k$	The K-dimensional vector with a 1 in the k-th coordinate and 0's elsewhere.
$\mathbb{E}_{\boldsymbol{\mu}}$ and $\mathbb{P}_{\boldsymbol{\mu}}$	The expectation and probability measure corresponding to the parameter $\boldsymbol{\mu}$
$\text{Alt}(\boldsymbol{\mu})$	Set of confusing parameters for $\boldsymbol{\mu}$
$\boldsymbol{\omega}^\star(\boldsymbol{\mu})$	Optimal allocation for parameter $\boldsymbol{\mu}$
$T^\star(\boldsymbol{\mu})$	Characteristic time for parameter $\boldsymbol{\mu}$
$d(\mu, \mu')$	KL divergence between the distributions parametrized by μ and μ'
$\text{kl}(a, b)$	KL divergent between two Bernoulli distributions of means a and b
Assumptions on the objective function	
$\mathcal{J}_i$	Finite set of indexes associated with answer $i \in \mathcal{I}$
$\mathcal{C}_j^i$ where $j \in \mathcal{J}_i$	A convex set in $\Lambda \setminus \mathcal{S}_i$
$f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$	$\inf_{\boldsymbol{\lambda} \in \mathcal{C}_j^{i^\star(\boldsymbol{\mu})}} \sum_{k=1}^K \omega_k d(\mu_k, \lambda_k)$
$\text{cl}(\mathcal{K})$	The closure of $\mathcal{K}$
$\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$	$\arg\min_{\boldsymbol{\lambda} \in \text{cl}(\mathcal{C}_j^{i^\star(\boldsymbol{\mu})})} \sum_{k=1}^K \omega_k d(\mu_k, \lambda_k)$
$F_{\boldsymbol{\mu}}(\boldsymbol{\omega})$	$\min_{j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$
$C_\psi(\mathcal{K})$	$\sup_{\substack{\boldsymbol{x}, \boldsymbol{z} \in \mathcal{K} \\ \alpha \in (0,1] \\ \boldsymbol{y} = \boldsymbol{x} + \alpha(\boldsymbol{z} - \boldsymbol{x})}} \frac{1}{\alpha^2} [\psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) + \langle \boldsymbol{y} - \boldsymbol{x}, \nabla\psi(\boldsymbol{x}) \rangle]$
L	Upper bound of $\|\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\|_\infty$
D	Upper bound of $\gamma C_{f_j(\cdot, \boldsymbol{\mu})(\Sigma_\gamma)}$
τ	Stopping rule
$\hat{\imath}_\tau$	Decision rule
$\beta(t, \delta)$	Stopping threshold
$c_1(\Lambda), c_2(\Lambda)$	The constants needed for property of $\beta(t, \delta)$ (see (8))
Notations for `FWS`	
$N_k(t)$	Number of pulls of arm k up to t
$\omega_k(t)$	$N_k(t)/t$
A_t	The arm pulled in time t
$\hat{\mu}_k(t)$	$\sum_{s=1}^t X_k(s) \mathbb{1}\{A_s = k\}/N_k(t)$
$H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r)$	r-subdifferential subspace
$\boldsymbol{x}(t)$	The allocation tracked at time t
$\boldsymbol{z}(t)$	The solution for FW update at time t
$\{r_t\}_{t \geq 1}$	A sequence of positive numbers for `FWS`
$T_{\epsilon, L}$	Constant needed for the assumption on $\{r_t\}_{t \geq 1}$
κ, E	Constants needed for Assumption 3

B Proof of the Lower Bound of $\mathbb{E}_{\boldsymbol{\mu}}[\tau]$

Definition 1. *A δ-PAC strategy with stopping rule τ and decision rule $\hat{\imath}_\tau$ is a strategy such that for any $\boldsymbol{\mu} \in \Lambda$, $\mathbb{P}_{\boldsymbol{\mu}}(\tau < \infty) = 1$ and $\mathbb{P}_{\boldsymbol{\mu}}(\hat{\imath}_\tau \neq i^\star(\boldsymbol{\mu})) \leq \delta$.*

Proposition 2. *Let $\delta \in (0,1)$ and $\boldsymbol{\mu} \in \Lambda$. For any δ-PAC strategy,*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \geq T^\star(\boldsymbol{\mu}) \mathrm{kl}(\delta, 1-\delta), \tag{13}$$

where

$$T^\star(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{\omega} \in \Sigma} \inf_{\lambda \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k). \tag{14}$$

Note that $\mathrm{kl}(\delta, 1-\delta) \approx \log(1/\delta)$ as $\delta \to 0$. Hence (13) yields that

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau]}{\log(\frac{1}{\delta})} \geq T^\star(\boldsymbol{\mu}). \tag{15}$$

Proof. Consider a δ-PAC strategy. Let $\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})$. Let $\mathbb{P}_{\boldsymbol{\mu}}$ and $\mathbb{P}_{\boldsymbol{\lambda}}$ denote the probability measures generated by the parameter $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, respectively. τ is a stopping time w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 1}$ where $\mathcal{F}_t = \sigma(A_1, X_{A_1}(1), \ldots, A_t, X_{A_t}(t))$, and where A_t is the arm selected under the algorithm in round t and $X_{A_t}(t)$ is the corresponding reward. According to Definition 1, τ is almost surely finite, and Lemma 19 in [28] directly implies that

$$\sum_{k=1}^{K} \mathbb{E}_{\boldsymbol{\mu}}[N_k(\tau)]\, d(\mu_k, \lambda_k) \geq \mathrm{kl}(\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{E}), \mathbb{P}_{\boldsymbol{\lambda}}(\mathcal{E})), \tag{16}$$

where $\mathcal{E}$ can be any $\mathcal{F}_\tau$-measurable event. With the choice, $\mathcal{E} = \{\hat{\imath}_\tau = i^\star(\boldsymbol{\lambda})\}$, the definition of δ-PAC strategy and $\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})$ imply that the right-hand side of inequality (16) is $\mathrm{kl}(\mathbb{P}_{\boldsymbol{\mu}}(\mathcal{E}), \mathbb{P}_{\boldsymbol{\lambda}}(\mathcal{E})) \geq \mathrm{kl}(\delta, 1-\delta)$. (16) holds for any $\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})$. Thus,

$$\begin{aligned}
\mathrm{kl}(\delta, 1-\delta) &\leq \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \mathbb{E}_{\boldsymbol{\mu}}[\tau] \sum_{k=1}^{K} \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_k(\tau)]}{\mathbb{E}_{\boldsymbol{\mu}}[\tau]} d(\mu_k, \lambda_k) \\
&\leq \mathbb{E}_{\boldsymbol{\mu}}[\tau] \sup_{\boldsymbol{\omega} \in \Sigma} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k).
\end{aligned} \tag{17}$$

This completes the proof. □

C A Generic Method to Verify Assumptions 2

Recall that Assumption 2 is:

Assumption 2. For all $\boldsymbol{\mu} \in \Lambda$,
(i) there exists $L > 0$ such that $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, $\|\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\|_\infty \leq L$;
(ii) there exists $D > 0$ such that $\forall \gamma \in (0, 1/K)$ and $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, $C_{f_j(\cdot,\boldsymbol{\mu})}(\Sigma_\gamma) \leq \frac{D}{\gamma}$.

Notation. In this appendix, we often use the function $d : (\mu, \pi) \mapsto d(\mu, \pi)$, defined as the KL divergence between reward distributions parametrized μ and π. π will denote its second argument. For example, $\frac{\partial d}{\partial \pi}(\mu, \lambda)$ is the partial derivate of d w.r.t. its second argument evaluated at the point (μ, λ).

C.1 Preliminaries: BAI in unstructured bandits

Before we introduce a generic way to check the assumption, we discuss the insightful case of the BAI problem in unstructured bandits with Bernoulli rewards. In this case, the gradients of f_j's are:

$$\forall j \neq i^\star(\boldsymbol{\mu}),\ \nabla f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = d(\mu_{i^\star(\boldsymbol{\mu})}, m_j(\boldsymbol{\omega}, \boldsymbol{\mu}))\boldsymbol{e}_{i^\star(\boldsymbol{\mu})} + d(\mu_j, m_j(\boldsymbol{\omega}, \boldsymbol{\mu}))\boldsymbol{e}_j,$$

where

$$m_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \frac{\omega_{i^\star(\boldsymbol{\mu})}\mu_{i^\star(\boldsymbol{\mu})} + \omega_j\mu_j}{\omega_{i^\star(\boldsymbol{\mu})} + \omega_j}.$$

We deduce that $\|\nabla f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\|_\infty \leq L = \max_{k \neq i^\star(\boldsymbol{\mu})} d(\mu_{i^\star(\boldsymbol{\mu})}, \mu_k)$ for any $\boldsymbol{\omega} \in \Sigma$. Hence, this constant L satisfies Assumption 2 (i) and depends $\boldsymbol{\mu}$ only. As for the Assumption 2 (ii), the Hessian $\nabla^2_{\boldsymbol{\omega},\boldsymbol{\omega}} f_j(\boldsymbol{\omega})$ has elements almost all equal to 0 except for those corresponding to the basis $\boldsymbol{e}_{i^\star(\boldsymbol{\mu})}, \boldsymbol{e}_j$. Extracting the non-zero elements of the Hessian, we get:

$$\begin{bmatrix} \frac{m_j(\boldsymbol{\omega},\boldsymbol{\mu})-\mu_{i^\star(\boldsymbol{\mu})}}{m_j(\boldsymbol{\omega},\boldsymbol{\mu})(1-m_j(\boldsymbol{\omega},\boldsymbol{\mu}))}\frac{(\mu_{i^\star(\boldsymbol{\mu})}-\mu_j)\omega_j}{(\omega_{i^\star(\boldsymbol{\mu})}+\omega_j)^2} & \frac{m_j(\boldsymbol{\omega},\boldsymbol{\mu})-\mu_{i^\star(\boldsymbol{\mu})}}{m_j(\boldsymbol{\omega},\boldsymbol{\mu})(1-m_j(\boldsymbol{\omega},\boldsymbol{\mu}))}\frac{(\mu_j-\mu_{i^\star(\boldsymbol{\mu})})\omega_{i^\star(\boldsymbol{\mu})}}{(\omega_{i^\star(\boldsymbol{\mu})}+\omega_j)^2} \\ \frac{m_j(\boldsymbol{\omega},\boldsymbol{\mu})-\mu_j}{m_j(\boldsymbol{\omega},\boldsymbol{\mu})(1-m_j(\boldsymbol{\omega},\boldsymbol{\mu}))}\frac{(\mu_{i^\star(\boldsymbol{\mu})}-\mu_j)\omega_j}{(\omega_{i^\star(\boldsymbol{\mu})}+\omega_j)^2} & \frac{m_j(\boldsymbol{\omega},\boldsymbol{\mu})-\mu_j}{m_j(\boldsymbol{\omega},\boldsymbol{\mu})(1-m_j(\boldsymbol{\omega},\boldsymbol{\mu}))}\frac{(\mu_j-\mu_{i^\star(\boldsymbol{\mu})})\omega_{i^\star(\boldsymbol{\mu})}}{(\omega_{i^\star(\boldsymbol{\mu})}+\omega_j)^2} \end{bmatrix}.$$

Notice that $\left|\frac{m_j(\boldsymbol{\omega},\boldsymbol{\mu})-\mu_{i^\star(\boldsymbol{\mu})}}{m_j(\boldsymbol{\omega},\boldsymbol{\mu})(1-m_j(\boldsymbol{\omega},\boldsymbol{\mu}))}\right| < 4\mu_{i^\star(\boldsymbol{\mu})}$. Thus $\left\|\nabla^2_{\boldsymbol{\omega},\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\right\|_\infty$ (defined as the maximum over rows of the L_1-norm of a row), is smaller than $\frac{4L\mu_{i^\star(\boldsymbol{\mu})}}{\gamma}$ when $\boldsymbol{\omega} \in \Sigma_\gamma$. Invoking Lemma 1.2.2 in [3] (more precisely, in its proof), one can immediately deduce that ∇f_j is $\frac{D}{\gamma}$-Lipschitz, where $D = 4L\mu_{i^\star(\boldsymbol{\mu})}$. Finally, Lemma 7 in [24] implies that a function with gradient $\frac{D}{\gamma}$-Lipschitz satisfies Assumption 2 (ii).

From the above observations, we note that the value of $m_j(\boldsymbol{\omega}, \boldsymbol{\mu})$, or equivalently the most confusing parameter $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$, plays an essential role in our assumptions. In view of Proposition 1, $\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_k d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k)\boldsymbol{e}_k$. First, if $d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k)$ is bounded for any $k \in [K]$, $\boldsymbol{\omega} \in \mathring{\Sigma}$ and $j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, then Assumption 2 (i) holds because the k-th component of $\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ is exactly $d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k)$. Then, the chain rule yields:

$$\left(\nabla^2_{\boldsymbol{\omega},\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\right)_{k,k'} = \left(\frac{\partial d}{\partial \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k)\right) \frac{\partial}{\partial \omega_{k'}} \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k. \tag{18}$$

For the BAI in unstructured bandits, we can derive Assumption 2 (ii) if $\frac{\partial d}{\partial \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k)$ is bounded and $\left\|\nabla_{\boldsymbol{\omega}} \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}\right\|_\infty$ is shown to scale as $\mathcal{O}(\frac{1}{\min_k \omega_k})$. Sometimes, however, $\nabla_{\boldsymbol{\omega}} \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ is not easy to compute. Next we provide a sufficient condition for Assumption 2 which is easier to check.

C.2 Constraint function and a sufficient condition for Assumption 2

Constraint function. To state our sufficient condition, we introduce the constraint function c^i_j to describe the set $\mathcal{C}^i_j$. Let us fix $i \in \mathcal{I}$ and $j \in \mathcal{J}_i$. The constraint function $c^i_j : \Lambda \setminus \mathcal{S}_i \to \mathbb{R}$ is a

mapping such that:

$$\mathcal{C}_j^i = \left\{ \boldsymbol{\mu} \in \Lambda \setminus \mathcal{S}_i : c_j^i(\boldsymbol{\mu}) > 0 \right\}.$$

Namely, we can define $\mathcal{C}_j^i$ by using c_j^i. For concreteness, we list below examples in which there is a constraint function.

Example 1 – BAI in unstructured bandits with Bernoulli rewards. For this task, we have $\Lambda = (0,1)^K$, $\mathcal{I} = \{1, \dots, K\}$, and for all arm i, the set of parameters for which arm i is the best arm is $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : \mu_i > \mu_k, \forall k \neq i\}$. We have: $\Lambda \setminus \mathcal{S}_i = \cup_{j \in \mathcal{J}_i} \mathcal{C}_j^i$ where $\mathcal{J}_i = \mathcal{I} \setminus \{i\}$ is the set of arms different than i and $\mathcal{C}_j^i = \{\boldsymbol{\mu} \in \Lambda : \mu_j > \mu_i\}$. Thus a constraint function is $c_j^i(\boldsymbol{\mu}) = \mu_j - \mu_i$.

Example 2 – Threshold bandits. In this task, the objective is to identify all arms whose average rewards are above a threshold $\mathfrak{I}$. With Bernoulli rewards, we have $\Lambda = (0,1)^K$ and the set of possible answers is $\mathcal{I} = 2^{[K]}$. We can decompose the set $\Lambda \setminus \mathcal{S}_\mathcal{A} = \cup_{k \in [K]} \mathcal{C}_k^\mathcal{A}$, where

$$\mathcal{C}_k^\mathcal{A} = \begin{cases} \{\boldsymbol{\mu} \in \Lambda \setminus \mathcal{S}_\mathcal{A} : \mu_k > \mathfrak{I}\} & \text{if } k \notin \mathcal{A}, \\ \{\boldsymbol{\mu} \in \Lambda \setminus \mathcal{S}_\mathcal{A} : \mu_k < \mathfrak{I}\} & \text{if } k \in \mathcal{A}. \end{cases}$$

Then a constraint function is: $c_k^\mathcal{A}(\boldsymbol{\mu}) = (\mathbb{1}\{k \notin \mathcal{A}\} - \mathbb{1}\{k \in \mathcal{A}\})(\mu_k - \mathfrak{I})$.

Example 3 – Top-m bandits. The task is to identify the best m arms. Assuming Bernoulli rewards, we have $\Lambda = \{\boldsymbol{\mu} \in (0,1) : \mu_{[1]} \geq \dots \mu_{[m]} > \mu_{[m+1]}\}$, where $\mu_{[k]}$ denotes the average reward of the arm with the k-th highest reward. The set of possible answers is $\mathcal{I} = \{\mathcal{A} \in [K] : |\mathcal{A}| = m\}$. Define $\mathcal{J}_\mathcal{A} = \{j \notin \mathcal{A}\}$ and $\mathcal{C}_j^\mathcal{A} = \{\boldsymbol{\mu} \in \Lambda \setminus \mathcal{S}_\mathcal{A} : \mu_j > \min_{k \in \mathcal{A}} \mu_k\}$. Then, we have: $\Lambda \setminus \mathcal{S}_\mathcal{A} = \cup_{j \notin \mathcal{J}_\mathcal{A}} \mathcal{C}_j^\mathcal{A}$ and a constraint function can be $c_j^\mathcal{A}(\boldsymbol{\mu}) = \mu_j - \min_{k \in \mathcal{A}} \mu_k$.

As illustrated in the above examples, the constraint functions c_j^i depend on the pure exploration task, but are simple and usually differentiable. The following lemma provides a sufficient condition for 2, involving the constraint functions only. In all the examples considered in this paper, this lemma can be applied. Its proof, provided at the end of this appendix, combines the Lagrange multiplier theorem and the implicit function theorem, and leverages similar techniques as those developed in [14, 2].

Lemma 1. *Let $\boldsymbol{\mu} \in \Lambda$. Assume that, for any $j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$,*
(a) $c_j^{i^\star(\boldsymbol{\mu})}$ is twice differentiable at the point $(\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})})$ and $\nabla^2 c_j^{i^\star(\boldsymbol{\mu})}(\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}) = 0, \forall \boldsymbol{\omega} \in \mathring{\Sigma}$,
(b) the reward distributions are Gaussian or Bernoulli,
(c) there is a constant $M > 0$ such that

$$\max \left\{ \left| d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})_k}) \right|, \left| \frac{\partial d}{\partial \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})_k}) \right| \right\} \leq M, \ \forall k \in [K], \boldsymbol{\omega} \in \mathring{\Sigma}.$$

Then Assumption 2 holds.

C.3 Applications of Lemma 1

We apply Lemma 1 to verify Assumption 2 for the pure exploration tasks presented Appendix D, E, F, and G.

Conditions (a) and (b). These conditions hold trivially because in all examples, the constraint functions are linear, and we consider only Bernoulli or Gaussian rewards.

Condition (c). First observe that:
for Bernoulli rewards,

$$d(\mu, \pi) = \mu \log \frac{\mu}{\pi} + (1-\mu) \log \frac{1-\mu}{1-\pi} \text{ and } \frac{\partial d}{\partial \pi}(\mu, \pi) = \frac{-\mu}{\pi} + \frac{1-\mu}{1-\pi}, \ \forall \mu, \pi \in (0,1),$$

and for Gaussian rewards,

$$d(\mu, \pi) = \frac{1}{2}(\mu - \pi)^2 \text{ and } \frac{\partial d}{\partial \pi}(\mu, \pi) = \pi - \mu, \ \forall \mu, \pi \in \mathbb{R}.$$

In the case of BAI for unstructured bandits (Appendix D), $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}_k = \mu_k$ if $k \notin \{i^\star(\boldsymbol{\mu}), j\}$ and $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}_k = m_j(\boldsymbol{\omega},\boldsymbol{\mu})$ otherwise, where $m_j(\boldsymbol{\omega},\boldsymbol{\mu}) = \frac{\omega_{i^\star(\boldsymbol{\mu})}\mu_{i^\star(\boldsymbol{\mu})}+\omega_j\mu_j}{\omega_{i^\star(\boldsymbol{\mu})}+\omega_j}, \forall \boldsymbol{\omega} \in \mathring{\Sigma}$. Hence, (c) holds for $m_j(\boldsymbol{\omega},\boldsymbol{\mu})$ is bounded in the interval $[\mu_j, \mu_{i^\star(\boldsymbol{\mu})}]$.

For BAI in linear bandits (Appendix E), according to (12), we have that for any $\boldsymbol{\omega} \in \mathring{\Sigma}$,

$$\begin{aligned}\left\|\overline{\boldsymbol{\lambda}_j^{\mathrm{BAI}}(\boldsymbol{\omega},\boldsymbol{\mu})}\right\|_\infty &\leq \|\boldsymbol{\mu}\|_\infty + \left\|\left(\frac{\langle \boldsymbol{a}_{i^\star} - \boldsymbol{a}_j, \boldsymbol{\mu}\rangle}{\|\boldsymbol{a}_{i^\star} - \boldsymbol{a}_j\|^2_{V_{\boldsymbol{\omega}}^{-1}}} V_{\boldsymbol{\omega}}^{-1}\right)(\boldsymbol{a}_j - \boldsymbol{a}_{i^\star})\right\|_\infty \\ &\leq \|\boldsymbol{\mu}\|_\infty + \|\boldsymbol{a}_{i^\star} - \boldsymbol{a}_j\|_\infty \|\boldsymbol{\mu}\|_\infty < \infty.\end{aligned}$$

Thus, (c) holds. The condition can be checked similarly for the threshold linear bandits. As for BAI in Lipschitz bandits (Appendix F), each component of the most confusing parameter (see (31)) is bounded in the interval of $[\min_k \mu_k, \max_k \mu_k]$, and thus, (c) holds. Finally, for the threshold bandit problem with monotone structure and the top-m arm problem in dueling bandits (Appendix G), (c) directly holds as the most confusing parameter $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}$ and $\overline{\boldsymbol{\lambda}_\sigma(\boldsymbol{\omega},\boldsymbol{\mu})}$ are fixed for any $\boldsymbol{\omega} \in \mathring{\Sigma}$.

C.4 Proof of Lemma 1

We first prove that Assumption 2 (i) holds. Let $L = M$. Observe that $\|\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega},\boldsymbol{\mu})\|_\infty = \max_k \left| d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}_k)\right| < M$, then (i) holds directly.

Let us verify Assumption 2 (ii). Let $\gamma \in (0, \frac{1}{K})$. We will prove that for $\boldsymbol{\omega} \in \Sigma_\gamma$, $\left\|\nabla^2_{\boldsymbol{\omega},\boldsymbol{\omega}} f_j(\boldsymbol{\omega},\boldsymbol{\mu})\right\|_\infty$ is bounded by $\frac{D}{\gamma}$ by some constant $D > 0$. This will imply that Assumption 2 (ii) holds, see e.g.,[3, 24].

We have:

$$\begin{cases} \frac{\partial^2 d}{\partial \pi^2}(\mu,\pi) = \frac{\mu}{\pi^2} + \frac{1-\mu}{(1-\pi)^2} \geq 1, \forall \mu, \pi \in (0,1), & \text{for Bernoulli rewards,} \\ \frac{\partial^2 d}{\partial \pi^2}(\mu,\pi) = 1, \ \forall \mu, \pi \in \mathbb{R}, & \text{for Gaussian rewards.} \end{cases} \tag{19}$$

We deduce that $\frac{\partial^2 d}{\partial \pi^2}(\mu,\pi) \geq 1$. Now, recall that $\overline{\boldsymbol{\lambda}_j(\omega,\boldsymbol{\mu})}$ is the solution of the following optimization problem:

$$\min_{\boldsymbol{\pi} \in \Lambda, c_j^{i^\star(\boldsymbol{\mu})}(\boldsymbol{\pi}) \geq 0} \sum_k \omega_k d(\mu_k, \pi_k).$$

Let $\mathcal{L} : \mathbb{R}^K \times \mathbb{R}^K \times \mathbb{R} \times \mathbb{R}^K \mapsto \mathbb{R}^K$ be the Lagrangian defined as

$$\mathcal{L}(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\boldsymbol{\pi}) = \sum_k \omega_k d(\mu_k,\pi_k) - \alpha c_j^{i^\star(\boldsymbol{\mu})}(\boldsymbol{\pi}).$$

The solution $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}$ can be identified by solving $\nabla_\alpha \mathcal{L}(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\overline{\boldsymbol{\lambda}_j(\boldsymbol{x},\boldsymbol{\mu})}) = -c_j^{i^\star(\boldsymbol{\mu})}(\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})})$ and

$$\nabla_{\boldsymbol{\pi}} \mathcal{L}(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}) = \sum_k \omega_k \frac{\partial d}{\partial \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}_k)\boldsymbol{e}_k - \alpha \nabla c_j^{i^\star(\boldsymbol{\mu})}(\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}). \tag{20}$$

By differentiating (20) with respect to $\boldsymbol{\pi}$, we get the Hessian of $\nabla^2_{\boldsymbol{\pi},\boldsymbol{\pi}}\mathcal{L}(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})})$: it is a diagonal matrix and for any $k \in [K]$, its (k,k)-th entry is

$$\begin{aligned}\nabla^2_{\boldsymbol{\pi},\boldsymbol{\pi}}\mathcal{L}(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})})_{k,k} &= \omega_k \frac{\partial^2 d}{\partial^2 \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}_k) - \alpha \nabla^2 c_j^{i^\star(\boldsymbol{\mu})}(\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}) \\ &= \omega_k \frac{\partial^2 d}{\partial^2 \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}_k) \\ &\geq \omega_k. \end{aligned} \tag{21}$$

Since $\boldsymbol{\omega} \in \Sigma_\gamma$, we deduce from (21) that $\nabla^2_{\boldsymbol{\pi},\boldsymbol{\pi}} L(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\overline{\boldsymbol{\lambda}_j(\boldsymbol{x},\boldsymbol{\mu})})$ is invertible, and that we can apply the implicit function theorem:

$$\nabla_{\boldsymbol{\omega}} \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})} = -\left(\nabla^2_{\boldsymbol{\pi},\boldsymbol{\pi}}\mathcal{L}(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})})\right)^{-1}\left(\nabla_{\boldsymbol{\omega}}\nabla_{\boldsymbol{\pi}}\mathcal{L}(\boldsymbol{\omega},\boldsymbol{\mu},\alpha,\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})})\right). \tag{22}$$

In addition, we have:

$$\left\| \left(\nabla^2_{\boldsymbol{\pi},\boldsymbol{\pi}} \mathcal{L}(\boldsymbol{\omega}, \boldsymbol{\mu}, \alpha, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}) \right)^{-1} \right\|_\infty \leq \frac{1}{\gamma}. \tag{23}$$

To derive an upper bound of $\left\| \nabla_{\boldsymbol{\omega}} \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} \right\|_\infty$, we compute the second factor in the r.h.s. of (22) by differentiating (20) with respect to $\boldsymbol{\omega}$. We can see that $\nabla_{\boldsymbol{\omega}} \nabla_{\boldsymbol{\pi}} \mathcal{L}(\boldsymbol{\omega}, \boldsymbol{\mu}, \alpha, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ is a diagonal matrix whose (k,k)-th entry is $\frac{\partial d}{\partial \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})_k})$ for any $k \in [K]$. Combining this observation with (22)-(23), we deduce that

$$\left\| \nabla_{\boldsymbol{\omega}} \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} \right\|_\infty \leq \frac{\max_k \left| \frac{\partial d}{\partial \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})_k}) \right|}{\gamma} \leq \frac{M}{\gamma}, \tag{24}$$

where the last inequality stems from (c). Finally, using (24), (c), and (18), we can upper bound $\left\| \nabla^2_{\boldsymbol{\omega},\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) \right\|_\infty$ as:

$$\left\| \nabla^2_{\boldsymbol{\omega},\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) \right\|_\infty = \max_k \sum_{k'=1}^{K} \left| \left(\frac{\partial d}{\partial \pi}(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})_k}) \right) \frac{\partial}{\partial \omega_{k'}} \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})_k} \right| \leq \frac{M^2 K}{\gamma}.$$

We have proved Assumption 2 (ii) with $D = M^2 K$. □

D BAI in Unstructured Bandits

About all our experiments. All the experiments are executed on a machine with Intel Core i5 at 1.8 GHz with 8 GB RAM. We implemented all the algorithms[3] in `Julia 1.5.4` and part of the baselines are taken from the implementation by [31, 45]. Throughout the experiments, we fix the parameters of our Frank-Wolfe-based sampling (`FWS`): $r_t = t^{-0.9}/K$, where K is the number of arms.

D.1 Preliminaries and competing algorithms

The BAI in unstructured bandits with Bernoulli rewards has been treated in Example 1 (in the main document). It is obtained by assuming $\Lambda = \{\boldsymbol{\mu} \in (0,1)^K : \exists i \in [K] \text{ s.t } \mu_i > \mu_k, \forall k \neq i\}$. The set of answers is $\mathcal{I} = [K]$ and $i^\star(\boldsymbol{\mu}) = \operatorname{argmax}_{k\in[K]} \mu_k$ and hence $\mathcal{S}_i = \{\boldsymbol{\mu} \in (0,1)^K : \mu_i > \mu_k, \forall k \neq i\}$. For this BAI, we set $\mathcal{J}_i = [K] \setminus i$ and $\mathcal{C}_j^i = \{\boldsymbol{\mu} \in \Lambda : \mu_j > \mu_i\}$. Obviously, $\{\mathcal{S}_i\}_{i\in[K]}$ are open sets and $\{\mathcal{C}_j^i\}_{j\neq i}$ are convex sets, so Assumption 1 holds. As already mentioned, we have: $\forall (\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_{i^\star(\boldsymbol{\mu})}, \forall j \neq i^\star(\boldsymbol{\mu})$,

$$\begin{aligned}
\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} &= m_j(\boldsymbol{\omega}, \boldsymbol{\mu}) \boldsymbol{e}_{i^\star(\boldsymbol{\mu})} + m_j(\boldsymbol{\omega}, \boldsymbol{\mu}) \boldsymbol{e}_j + \sum_{k\neq j, i^\star(\boldsymbol{\mu})} \mu_k \boldsymbol{e}_k, \\
f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) &= \omega_{i^\star(\boldsymbol{\mu})} d(\mu_{i^\star(\boldsymbol{\mu})}, m_j(\boldsymbol{\omega}, \boldsymbol{\mu})) + \omega_j d(\mu_j, m_j(\boldsymbol{\omega}, \boldsymbol{\mu})), \\
\nabla f_j(\omega, \boldsymbol{\mu}) &= d(\mu_{i^\star(\boldsymbol{\mu})}, m_j(\boldsymbol{\omega}, \boldsymbol{\mu})) \boldsymbol{e}_{i^\star(\boldsymbol{\mu})} + d(\mu_j, m_j(\boldsymbol{\omega}, \boldsymbol{\mu})) \boldsymbol{e}_j,
\end{aligned}$$

where

$$m_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \frac{\omega_{i^\star(\boldsymbol{\mu})} \mu_{i^\star(\boldsymbol{\mu})} + \omega_j \mu_j}{\omega_{i^\star(\boldsymbol{\mu})} + \omega_j}.$$

Assumption 2 is verified in Appendix C.

`FWS` algorithm and the FW update. To illustrate the implementation of `FWS`, we provide an example on how the FW update (11) is implemented. This update is translated into a zero-sum game that can be solved using any LP solver. Refer to Appendix H for a discussion on how to get this game in general pure exploration problems.

Let $K = 3$. Assume that we are in round t, and that we wish to apply the FW update:

$$\boldsymbol{z}(t) \leftarrow \operatorname*{argmax}_{\boldsymbol{z}\in\Sigma} \min_{h \in H_{F_{\hat{\boldsymbol{\mu}}(t-1)}(\boldsymbol{x}(t-1), r_t)}} \langle \boldsymbol{z} - \boldsymbol{x}(t-1), h \rangle \quad \textit{(ties broken arbitrarily)} .$$

Further assume that $i^\star(\hat{\boldsymbol{\mu}}(t-1)) = 3$ and that $f_1(\boldsymbol{x}(t-1), \hat{\boldsymbol{\mu}}(t-1)) \vee f_2(\boldsymbol{x}(t-1), \hat{\boldsymbol{\mu}}(t-1)) < F_t(\boldsymbol{x}(t-1)) + r_t$. We then create a 3×2 payoff matrix M, whose (k,j)-th entry is $M_{k,j} = \langle \boldsymbol{e}_k - \boldsymbol{x}(t-1), \nabla f_j(\boldsymbol{x}(t-1), \hat{\boldsymbol{\mu}}(t-1)) \rangle$ for all $k = 1,2,3$ and $j = 1,2$. In this example, the update can be formulated as

$$\begin{aligned}
&\max_{\boldsymbol{z}\in\Sigma} \min_{\boldsymbol{y}\in\mathbb{R}^2} \boldsymbol{z}^\top M \boldsymbol{y} \\
&\text{s.t. } y_1, y_2 \geq 0 \text{ and } y_1 + y_2 = 1.
\end{aligned} \tag{25}$$

Let $(\boldsymbol{z}^\star, \boldsymbol{y}^\star)$ denote the solution of (25). Then we have

$$\boldsymbol{z}(t) = \boldsymbol{x}(t-1) + \sum_{k=1}^{3} z_k^\star (\boldsymbol{e}_k - \boldsymbol{x}(t-1)) = \boldsymbol{z}^\star.$$

A standard method to solve the zero-sum game (25) is to apply any LP solver to the following problem [37, 48, 52]:

$$\begin{aligned}
&\max_{\boldsymbol{z}\in\Sigma, u\in\mathbb{R}} u \\
&\text{s.t. } (\boldsymbol{z}^\top M)_1, (\boldsymbol{z}^\top M)_2 \geq u.
\end{aligned} \tag{26}$$

The solution of (26) provides $\boldsymbol{z}^\star$ and the value of (25). Appendix H explains why and gives a short introduction for the transformation of a zero-sum game to an LP.

Competing algorithms. The list of algorithms used for comparison is provided below.

[3] https://github.com/rctzeng/NeurIPS2021-Fast-Pure-Exploration-via-Frank-Wolfe

- FWS: Our algorithm with parameters $r_t = t^{-0.9}/K$, where K is the number of arms.
- T-D: Track-and-Stop [28] with D-Tracking implemented by [31].
- D-C: AdaHedge as the $\boldsymbol{\lambda}$-player and Best-Response as the $\boldsymbol{\omega}$-player described in Section 3.1 in [12] implemented by [31].
- M-C: Lazy Mirror Ascent by [38] implemented by [31]. This method is very sensitive to the learning rate $\eta_t = 1/(L\sqrt{t})$ ($L > 0$ is a hyperparameter). Note that the implementation [31] chooses L assuming the knowledge of $\boldsymbol{\mu}$. This choice is for experimental comparison only and cannot be used in real-world scenarios.
- O-C: Optimistic Track and Stop [12] implemented by [31].
- RR: Sample arms in a round-robin manner.
- LB(δ): $T^\star(\boldsymbol{\mu})\mathrm{kl}(\delta, 1-\delta)$.

The Track-and-Stop algorithm has two versions, one with D-tracking (directly tracking the optimal allocation) and another one with C-tracking (tracking the cumulative optimal allocation). We found that D-tracking always performs better than C-tracking numerically. Hence, we report the performance of Track-and-Stop with D-tracking only.

Stopping rule. In all the algorithms, we use the same stopping rule (7) for unstructured bandits, with the same threshold $\beta(t, \delta) = \log((\log(t) + 1)/\delta)$, suggested in [20].

D.2 Numerical experiments

Bernoulli rewards. In the first experiment, we consider Bernoulli rewards with $\boldsymbol{\mu} = [0.3, 0.21, 0.2, 0.19, 0.18]$ used in [28]. We average our results over 3000 runs. In Table 2, we provide the sample complexity for various confidence levels $\delta \in \{0.1, 0.01, 0.001, 0.0001\}$. To provide a more detailed comparison, at the confidence level $\delta = 0.01$, we show the sample complexity in box-plot in Figure 3a and compare the allocation of arm draws achieved under the various algorithms in Table 3.

In Figure 1, we plot the number of rounds (the median over all runs) FWS is in force exploration or the r-subdifferential subspace used in FWS contains the gradient of only one function (in this round, our FW update coincides with the traditional FW update as if the objective function was smooth). In Figure 2, we provide the distribution of the number of functions involved in the FW updates. It is interesting to note that 60% of the time our update differs from the usual FW update.

Table 2: Sample complexity in unstructured bandits with Bernoulli rewards with $\boldsymbol{\mu} = [0.3, 0.21, 0.2, 0.19, 0.18]$ and $\delta = 0.01$, averaged over 3000 runs.

	FWS	T-D	D-C	M-C	O-C	RR	LB(δ)
$\delta = 0.1$	1 365	1 337	1 859	1 668	1 818	2 326	574
$\delta = 0.01$	2 125	2 066	2 674	2 509	2 706	3 460	1 471
$\delta = 0.001$	2 899	2 823	3 465	3 362	3 584	4 555	2 252
$\delta = 0.0001$	3 645	3 589	4 279	4 231	4 457	5 621	3 008

Table 3: Allocation of arm draws in unstructured bandits with Bernoulli rewards with $\boldsymbol{\mu} = [0.3, 0.21, 0.2, 0.19, 0.18]$ and $\delta = 0.01$, averaged over 3000 runs.

	FWS	T-D	D-C	M-C	O-C	RR	$\boldsymbol{\omega}^\star(\boldsymbol{\mu})$
$\boldsymbol{a}_1$	34.08	35.60	29.40	31.72	31.31	20.00	32.59
$\boldsymbol{a}_2$	24.82	23.47	21.46	22.37	20.94	20.00	25.15
$\boldsymbol{a}_3$	17.45	17.30	17.94	17.81	17.99	20.00	17.66
$\boldsymbol{a}_4$	13.38	13.22	16.11	15.06	15.79	20.00	13.24
$\boldsymbol{a}_5$	10.27	10.42	15.08	13.04	13.97	20.00	10.36

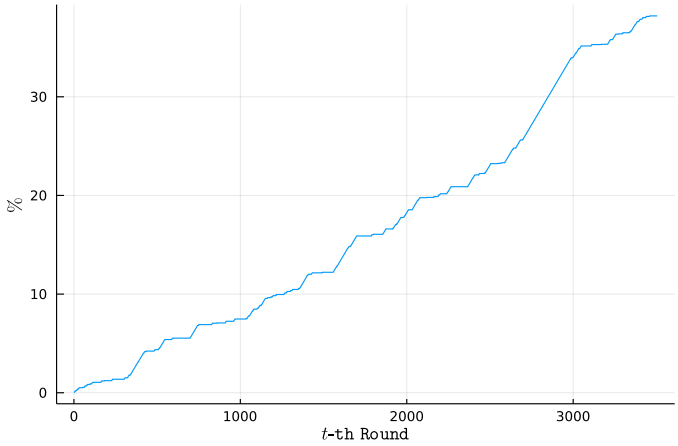


Figure 1: Number of rounds where FWS is in forced exploration or the FW update in FWS corresponds to the usual FW update. BAI in unstructured bandits with Bernoulli rewards and $\boldsymbol{\mu} = [0.3, 0.21, 0.2, 0.19, 0.18]$, $\delta = 0.0001$.

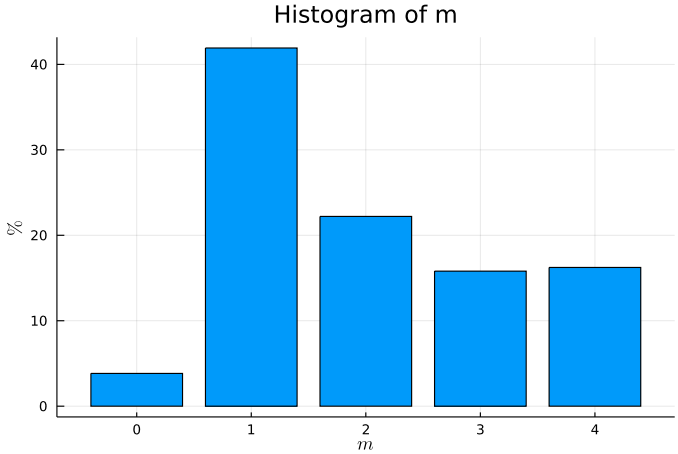
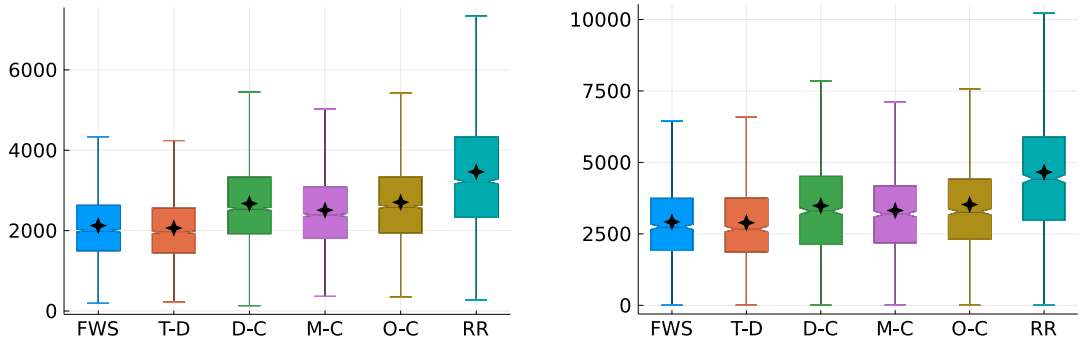


Figure 2: Proportions of rounds where we have m functions in the linear program involved in the FW updates (i.e. $m = \left|\{j : f_j(\boldsymbol{x}(t), \hat{\boldsymbol{\mu}}(t)) < F_{\hat{\boldsymbol{\mu}}(t-1)}(\boldsymbol{x}(t)) + r_t\}\right|$, and $m = 0$ in forced exploration). BAI in unstructured bandits with Bernoulli rewards and $\boldsymbol{\mu} = [0.3, 0.21, 0.2, 0.19, 0.18]$, $\delta = 0.0001$.



(a) Bernoulli rewards: $\boldsymbol{\mu} = [0.3, 0.21, 0.2, 0.19, 0.18]$. (b) Gaussian rewards: $\boldsymbol{\mu} = [1, 0.85, 0.8, 0.7]$.

Figure 3: Sample complexity for the unstructured best-arm identification problem at $\delta = 0.01$, plotted in boxplots where the stars represent the averaged sample complexity and the outliers are hidden.

Gaussian rewards. In the second experiment, we consider Gaussian rewards with means $\boldsymbol{\mu} = [1, 0.85, 0.8, 0.7]$ and unit variance as proposed in [38]. The results are averaged over 1000 runs. In Table 4, we compare the sample complexity for $\delta \in \{0.1, 0.01, 0.001, 0.0001\}$. At the confidence level $\delta = 0.01$, we show the sample complexity in box-plot in Figure 3b and compare the allocation of arm draws achieved under the various algorithms in Table 5.

Table 4: Sample complexity in unstructured bandits with Gaussian rewards and $\boldsymbol{\mu} = [1, 0.85, 0.8, 0.7]$ and $\delta = 0.01$, averaged over 1000 runs.

	FWS	T-D	D-C	M-C	O-C	RR	LB(δ)
$\delta = 0.1$	1 857	1 874	2 286	2 160	2 272	2 994	791
$\delta = 0.01$	2 919	2 891	3 487	3 313	3 528	4 659	2 026
$\delta = 0.001$	3 990	4 000	4 640	4 449	4 732	6 219	3 101
$\delta = 0.0001$	5 056	5 038	5 739	5 575	5 896	7 855	4 142

Table 5: Allocation of arm draws in unstructured bandits with Gaussian rewards $\boldsymbol{\mu} = [1, 0.85, 0.8, 0.7]$ and $\delta = 0.01$, averaged over 1000 runs.

	FWS	T-D	D-C	M-C	O-C	RR	$\boldsymbol{\omega}^\star(\boldsymbol{\mu})$
$\boldsymbol{a}_1$	41.05	42.00	34.37	37.46	39.70	25.00	41.25
$\boldsymbol{a}_2$	36.01	36.04	30.94	32.47	31.60	25.00	37.93
$\boldsymbol{a}_3$	16.94	16.11	19.56	18.51	18.92	25.00	15.21
$\boldsymbol{a}_4$	6.00	5.85	15.13	11.56	9.78	25.00	5.61

In all these results, we observe that FWS and T-D exhibit very close performance. FWS is as efficient as T-D. The two algorithms outperform other algorithms. We further observe that the allocations achieved under FWS and T-D are closer to the optimal allocation $\boldsymbol{\omega}^\star(\boldsymbol{\mu})$ than those of other algorithms.

E Linear Bandits

E.1 Preliminaries and competing algorithms

Linear bandits have been extensively applied in online advertisement, [34, 9], and have become the most relevant of bandit problems with structure. BAI in linear bandits has been investigated for example in [25, 46, 13].

Consider a bandit problem with K arms and Gaussian rewards. Each arm k is associated with a d-dimensional vector $\boldsymbol{a}_k$. Without loss of generality, we assume that $\{\boldsymbol{a}_k\}_{k\in[K]}$ spans the space $\mathbb{R}^d$. We study two learning tasks: BAI and the so-called threshold bandit task where the objective is to identify the set of arms whose expected rewards are above a given threshold.

We slightly modify our framework as described in Section 5. We define for the two tasks:

$$\Lambda_{\text{BAI}} = \left\{\boldsymbol{\mu}\in\mathbb{R}^d : \exists k\in[K] \text{ s.t.} \langle \boldsymbol{a}_k - \boldsymbol{a}_i, \boldsymbol{\mu}\rangle > 0, \forall i\neq k\right\}, \qquad i^\star_{\text{BAI}}(\boldsymbol{\mu}) = \operatorname*{argmax}_k \langle \boldsymbol{a}_k, \boldsymbol{\mu}\rangle;$$

$$\Lambda_{\mathfrak{I}} = \left\{\boldsymbol{\mu}\in\mathbb{R}^d : \langle \boldsymbol{a}_k, \boldsymbol{\mu}\rangle \neq \mathfrak{I}, \forall k\in[K]\right\}, \qquad i^\star_{\mathfrak{I}}(\boldsymbol{\mu}) = \{k\in[K] : \langle \boldsymbol{a}_k, \boldsymbol{\mu}\rangle > \mathfrak{I}\}.$$

For BAI, $\mathcal{S}_i = \{\boldsymbol{\mu}\in\mathbb{R}^d : \langle \boldsymbol{a}_k - \boldsymbol{a}_i, \boldsymbol{\mu}\rangle > 0, \forall i\neq k\}$ is clearly an open set. For the threshold problem, $\mathcal{S}_A = \{\boldsymbol{\mu}\in\mathbb{R}^d : \langle \boldsymbol{a}_k, \boldsymbol{\mu}\rangle > \mathfrak{I}, \forall k\in A\}$, where A is a subset of $[K]$, is an open set too. To implement our algorithm, we introduce $V_{\boldsymbol{\omega}} = \sum_k \omega_k \boldsymbol{a}_k\boldsymbol{a}_k^\top$, and build the Least-Squares Estimator of $\boldsymbol{\mu}$:

$$\hat{\boldsymbol{\mu}}(t) = V^\dagger_{\boldsymbol{\omega}(t)} \sum_{s=1}^{t} X_{A_t}(t)\boldsymbol{a}_{A_t}.$$

For this LSE, we use in the analysis the concentration results derived in Lemmas 3 and 4 in [25]. The objective function that has to be maximized to get an optimal allocation is:

$$F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \inf_{\boldsymbol{\lambda}\in \text{Alt}(\boldsymbol{\mu})} \frac{\|\boldsymbol{\mu}-\boldsymbol{\lambda}\|^2_{V_{\boldsymbol{\omega}}}}{2},$$

where $\text{Alt}(\boldsymbol{\mu})$ depends on the task. Let us describe the most confusing parameter $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega},\boldsymbol{\mu})}$ for both tasks.

BAI. Let $\boldsymbol{\mu}\in\Lambda_{\text{BAI}}$ and $j\neq i^\star_{\text{BAI}}(\boldsymbol{\mu})$, let $\mathcal{C}_j^{i^\star_{\text{BAI}}(\boldsymbol{\mu})} = \{\boldsymbol{\lambda}\in\Lambda^{\text{BAI}} : \langle\boldsymbol{\lambda}, \boldsymbol{a}_j\rangle > \langle\boldsymbol{\lambda}, \boldsymbol{a}_{i^\star_{\text{BAI}}(\boldsymbol{\mu})}\rangle\}$, which is a convex set (as any convex combination of two points in $\mathcal{C}_j^{i^\star_{\text{BAI}}(\boldsymbol{\mu})}$ is still in $\mathcal{C}_j^{i^\star_{\text{BAI}}(\boldsymbol{\mu})}$). Applying the Lagrange multiplier theorem (see Appendix of [13]), we get that:

$$\overline{\boldsymbol{\lambda}^{\text{BAI}}_j(\boldsymbol{\omega},\boldsymbol{\mu})} = \boldsymbol{\mu} + \left(\frac{\langle \boldsymbol{a}_{i^\star_{\text{BAI}}(\boldsymbol{\mu})} - \boldsymbol{a}_j, \boldsymbol{\mu}\rangle}{\left\|\boldsymbol{a}_{i^\star_{\text{BAI}}(\boldsymbol{\mu})} - \boldsymbol{a}_j\right\|^2_{V_{\boldsymbol{\omega}}^{-1}}} V_{\boldsymbol{\omega}}^{-1}\right)\left(\boldsymbol{a}_j - \boldsymbol{a}_{i^\star_{\text{BAI}}(\boldsymbol{\mu})}\right), \ \forall\boldsymbol{\omega}\in\mathring{\Sigma}. \tag{27}$$

Threshold bandit. For all $j\in[K]$, we let $\mathcal{C}_j^{i^\star_{\mathfrak{I}}(\boldsymbol{\mu})} = \{\boldsymbol{\lambda}\in\Lambda_{\mathfrak{I}} : \text{sign}(\mathfrak{I} - \langle\boldsymbol{a}_j,\boldsymbol{\lambda}\rangle) \neq \text{sign}(\mathfrak{I} - \langle\boldsymbol{a}_j,\boldsymbol{\mu}\rangle)\}$, which is a convex set (as again any convex combination of two points in $\mathcal{C}_j^{i^\star_{\mathfrak{I}}(\boldsymbol{\mu})}$ is still in $\mathcal{C}_j^{i^\star_{\mathfrak{I}}(\boldsymbol{\mu})}$). Likewise, the Lagrange multiplier theorem yields that

$$\overline{\boldsymbol{\lambda}^{\mathfrak{I}}_j(\boldsymbol{\omega},\boldsymbol{\mu})} = \boldsymbol{\mu} + \text{sign}(\mathfrak{I} - \langle\boldsymbol{a}_j,\boldsymbol{\mu}\rangle)\left(\frac{(\mathfrak{I} - \langle\boldsymbol{a}_j,\boldsymbol{\mu}\rangle)}{\|\boldsymbol{a}_j\|^2_{V_{\boldsymbol{\omega}}^{-1}}} V_{\boldsymbol{\omega}}^{-1}\right)\boldsymbol{a}_j, \ \forall\boldsymbol{\omega}\in\mathring{\Sigma}. \tag{28}$$

$\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega},\boldsymbol{\mu})$ can be obtained by directly plugging (27) or (28) into (4) and Proposition 1 shows that $f_j(\boldsymbol{\omega},\boldsymbol{\mu}) = \langle\boldsymbol{\omega}, \nabla f_j(\boldsymbol{\omega},\boldsymbol{\mu})\rangle$. We have checked Assumption 2 in Appendix C.3, using Lemma 1.

Competing algorithms. For BAI in linear structure, we compare the performance of the following algorithms.

- `FWS`: Our algorithm with parameters $r_t = t^{-0.9}/K$.

- LT: Lazy Track and Stop (LT) by [25].
- CG-C and Lk-C: LineGame-C (CG-C) and LineGame (Lk-C) from [13] implemented by [45].
- XY-A: XY-Adaptive [46]. The hyperparameter α is set equal to 0.1 as done by [46].
- RR: Round Robin
- $\mathtt{LB}_{\mathtt{lin}}(\delta)$: $T^\star(\boldsymbol{\mu})\mathrm{kl}(\delta, 1-\delta)$ exploiting the linear structure

To our best knowledge, the linear threshold bandit problem was only studied in [13]. Hence for this problem, we only compare our algorithm with CG-C, Lk-C and RR.

Stopping rules. For BAI in linear bandits, except for XY-A, all algorithms use the same stopping rule (7), with the same threshold $\beta(t,\delta) = \log((\log(t)+1)/\delta)$ (Note that the implementations in [45] make this choice as well).
For linear threshold bandits, we also use the stopping rule (7) with the same threshold $\beta(t,\delta) = \log((\log(t)+1)/\delta)$ for all algorithms.

E.2 Numerical experiments

BAI in linear bandits. We consider the example proposed by [46]. The unknown parameter is $\boldsymbol{\mu} = \boldsymbol{e}_1$ and there are $d+1$ arms, $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d, \cos(\phi)\boldsymbol{e}_1 + \sin(\phi)\boldsymbol{e}_2$ in $\mathbb{R}^d$, where $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d$ form the standard orthonormal basis. We set $d = 6$ and $\phi = 0.1$.

In Table 6, we provide the sample complexity of the various algorithms averaged over 1000 runs for various confidence levels $\delta \in \{0.1, 0.01, 0.001, 0.0001\}$. To provide a more detailed comparison, at the confidence level $\delta = 0.01$, we show the sample complexity in box-plot in Figure 6a and compare the allocation of arm draws achieved under the various algorithms in Table 7.

Table 6: Sample complexity for BAI in linear bandits for the benchmark example of [46], averaged over 1000 runs.

	FWS	LT	CG-C	Lk-C	XY-A	RR	$\mathtt{LB}_{\mathtt{lin}}(\delta)$
$\delta = 0.1$	1 030	919	2 498	2 319	7 016	5 451	359
$\delta = 0.01$	1 614	1 464	3 501	3 431	7 779	8 814	920
$\delta = 0.001$	2 229	1 982	4 324	4 326	9 090	12 101	1 408
$\delta = 0.0001$	2 839	2 518	5 118	5 120	9 723	15 314	1 881

Table 7: Allocation of arm draws for the benchmark example of [46] at $\delta = 0.01$, averaged over 1000 runs.

	FWS	LT	CG-C	Lk-C	XY-A	RR	$\boldsymbol{\omega}^\star(\boldsymbol{\mu})$
$\boldsymbol{a}_1$	1.02	4.4	13.64	13.10	9.35	14.29	0.38
$\boldsymbol{a}_2$	94.22	91.11	35.75	36.66	69.80	14.28	97.72
$\boldsymbol{a}_3$	0.93	1.02	12.46	12.25	4.35	14.28	0.38
$\boldsymbol{a}_4$	0.93	1.01	12.43	12.22	3.63	14.28	0.38
$\boldsymbol{a}_5$	0.93	1	12.38	12.25	4.45	14.28	0.38
$\boldsymbol{a}_6$	0.93	1	12.44	12.23	2.64	14.28	0.38
$\boldsymbol{a}_7$	1.01	0.46	0.89	1.29	5.78	14.28	0.38

All the results above suggest that FWS is really competitive with the state-of-art algorithm, LT, and it achieves an allocation closer to the optimal allocation than its competitors. In Figure 4, we plot the number of rounds (the median over all runs) FWS is in force exploration or the r-subdifferential subspace used in FWS contains the gradient of only one function (in this round, our FW update coincides with the traditional FW update as if the objective function was smooth). In Figure 5, we provide the distribution of the number of functions involved in the FW updates. Most of the time, the update used in FWS coincides with the usual FW update (which contrasts with the case of BAI in unstructured bandits).

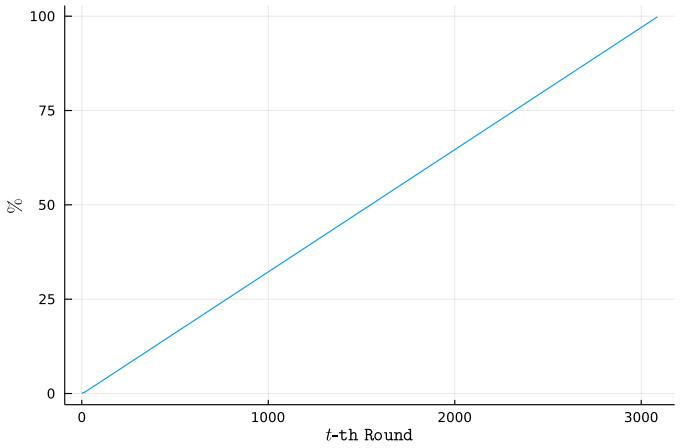


Figure 4: Number of rounds where the FW update in FWS corresponds to the usual FW update, and the force exploration. BAI in linear bandits for the benchmark example of [46] with $\delta = 0.0001$.

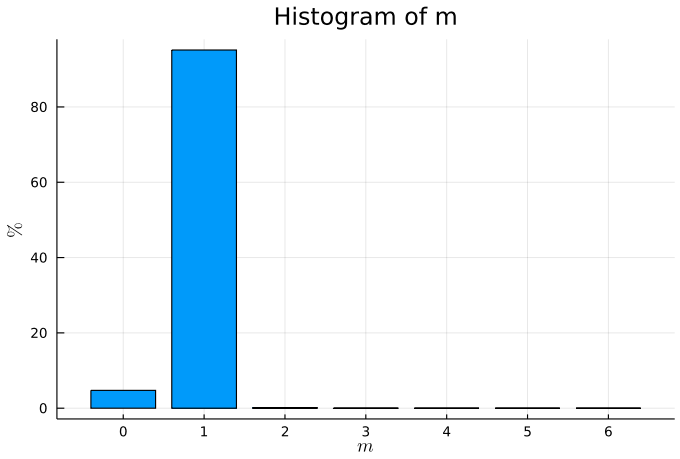
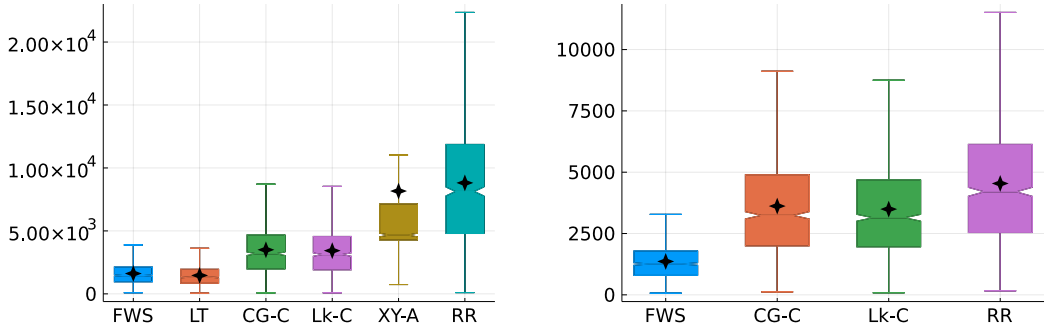


Figure 5: Proportions of rounds where we have m functions in the linear program involved in the FW updates (i.e. $m = \left|\{j : f_j(\boldsymbol{x}(t), \hat{\boldsymbol{\mu}}(t)) < F_{\hat{\boldsymbol{\mu}}(t-1)}(\boldsymbol{x}(t)) + r_t\}\right|$, and $m = 0$ in forced exploration). BAI in linear bandits for the benchmark example of [46] with $\delta = 0.0001$.



(a) Linear best-arm identification problem.

(b) Linear threshold problem.

Figure 6: Sample complexity for linear bandits at $\delta = 0.01$, plotted in boxplots where the stars represent the averaged sample complexity and the outliers are hidden.

Threshold bandit. Consider the linear threshold bandit problem, obtained by modifying the example of [46]: $\boldsymbol{\mu} = \boldsymbol{e}_1$ and there are $d + 1$ actions associated with $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d, \cos(\phi)\boldsymbol{e}_1 + \sin(\phi)\boldsymbol{e}_2$ in $\mathbb{R}^d$, where $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d)$ form the standard orthonormal basis.

Table 8: Sample complexity for the linear threshold bandit problem, averaged over 1000 runs.

	FWS	CG-C	Lk-C	RR	$\mathtt{LB}_{\mathtt{lin}}(\delta)$
$\delta = 0.1$	874	2 673	2 511	2 904	374
$\delta = 0.01$	1 362	3 618	3 494	4 540	957
$\delta = 0.001$	1 865	4 437	4 372	6 213	1465
$\delta = 0.0001$	2 398	5 163	5 132	7 807	1957

Table 9: Allocation of arm draws for the linear threshold bandit problem at $\delta = 0.01$, averaged over 1000 runs.

	FWS	CG-C	Lk-C	RR	$\boldsymbol{\omega}^\star(\boldsymbol{\mu})$
$\boldsymbol{a}_1$	19.62	19.08	19.19	14.30	0.38
$\boldsymbol{a}_2$	1.26	12.94	12.72	14.28	1.13
$\boldsymbol{a}_3$	1.34	12.75	12.45	14.28	1.17
$\boldsymbol{a}_4$	1.35	12.71	12.43	14.28	1.17
$\boldsymbol{a}_5$	1.30	12.73	12.45	14.28	1.17
$\boldsymbol{a}_6$	1.28	12.73	12.46	14.28	1.17
$\boldsymbol{a}_7$	73.84	17.07	18.31	14.28	93.80

We set $d = 6$, $\phi = 0.01$, and the threshold $\mathfrak{I} = 0.9$. The goal is to identify all arms whose mean is larger than the $\mathfrak{I}$.

In Table 8, we provide the sample complexity of the various algorithms for $\delta \in \{0.1, 0.01, 0.001, 0.0001\}$ and averaged over 1000 runs. To provide a more detailed comparison, at the confidence level $\delta = 0.01$, we show the sample complexity in box-plot in Figure 6b and compare the allocation of arm draws achieved under the various algorithms in Table 9.

We have the similar observations as those made for the BAI task. FWS clearly outperforms its competitors.

F BAI in Lipschitz Bandits

F.1 Preliminaries and competing algorithms

We consider the BAI task in the following Lipschitz bandit. There is a finite number K of arms. Each arm k is associated with a feature vector or *position* $\boldsymbol{a}_k \in \mathbb{R}^d$ for some $d \in \mathbb{N}$. The reward distributions are assumed to be Gaussian with a fixed and known variance. The mapping from the arm feature vector to the corresponding average reward is known to be Lipschitz, which means that:

$$\Lambda = \left\{\boldsymbol{\mu} \in \mathbb{R}^K : \exists i \in [K] \text{ s.t. } \mu_i > \mu_k, \forall k \neq i \text{ and } |\mu_k - \mu_{k'}| < \ell \left\|\boldsymbol{a}_k - \boldsymbol{a}_{k'}\right\|_\infty, \ \forall k, k' \in [K]\right\},$$

where the constant $\ell > 0$ is known in advance. The answer map is $i^\star(\boldsymbol{\mu}) = \operatorname{argmax}_i \mu_i$, since we consider a BAI task. Let us fix some $i \in [K]$ and $\boldsymbol{\mu} \in \mathcal{S}_i$, $\mathrm{Alt}(\boldsymbol{\mu})$ can be divided into a union of sets $\cup_{j \neq i}\{\boldsymbol{\lambda} \in \Lambda : \lambda_j > \lambda_i\}$. Hence $\mathcal{J}_i = [K] \setminus \{i\}$. Assumption 1 holds as the set

$$\mathcal{S}_i = \left\{\boldsymbol{\mu} \in \mathbb{R}^K : \mu_i > \mu_k, \forall k \neq i \text{ and } |\mu_k - \mu_{k'}| < \ell \left\|\boldsymbol{a}_k - \boldsymbol{a}_{k'}\right\|_\infty, \ \forall k, k' \in [K]\right\}$$

is open set for all $i \in [K]$ and

$$\mathcal{C}_j^i = \{\boldsymbol{\lambda} \in \Lambda : \lambda_j > \lambda_i, \text{ and } |\lambda_k - \lambda_{k'}| < \ell \left\|\boldsymbol{a}_k - \boldsymbol{a}_{k'}\right\|_\infty, \ \forall k, k' \in [K]\}$$

is a convex set, for all $j \neq i$ (one can readily check that any convex combination of any two points in $\mathcal{C}_j^i$ is still in $\mathcal{C}_j^i$).

Most confusing parameter. Unlike in the previous examples, there is no close form for the $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$. However, there is a simple strategy to compute it efficiently. Fix $\boldsymbol{\mu} \in \Lambda$, a suboptimal arm $j \neq i^\star(\boldsymbol{\mu})$, and $\boldsymbol{\omega} \in \mathring{\Sigma}$. The most confusing parameter solves (2), which in the case of Gaussian rewards translates to:

$$\min_{\boldsymbol{\lambda} \in \mathcal{C}_j^{i^\star(\boldsymbol{\mu})}} \sum_{k=1}^K \frac{\omega_k (\lambda_k - \mu_k)^2}{2}. \tag{29}$$

Observe that for the solution of (29), we should have $\lambda_j = \lambda_{i^\star(\boldsymbol{\mu})}$. Hence, the problem (29) can be simplified by setting $\lambda_j = \lambda_{i^\star(\boldsymbol{\mu})} = \theta \in [\mu_j, \mu_{i^\star(\boldsymbol{\mu})}]$ and by remarking that other values are decided by exploiting Lipschitz structure and minimizing the distance to $\boldsymbol{\mu}$. After simplification, we get a single-parameter (here θ) optimization problem:

$$\min_{\theta \in [\mu_j, \mu_{i^\star(\boldsymbol{\mu})}]} \sum_k \frac{\omega_k}{2} \left\{[(\theta - \ell \left\|\boldsymbol{a}_k - \boldsymbol{a}_j\right\|_\infty - \mu_k)^+]^2 + [(\mu_k - \theta - \ell \left\|\boldsymbol{a}_k - \boldsymbol{a}_{i^\star(\boldsymbol{\mu})}\right\|_\infty)^+]^2\right\}. \tag{30}$$

Figure 7 provides a simple example to explain this transformation.

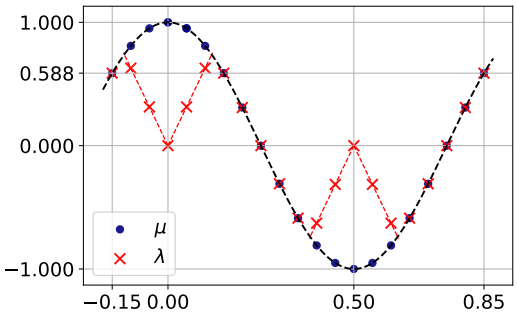


Figure 7: An example of most confusing parameters $\boldsymbol{\lambda} \in \mathcal{C}_j^{i^\star(\boldsymbol{\mu})}$. Along the x-axis, we have arm positions, and on the y-axis, the average rewards. Dots represent $\boldsymbol{\mu}$ and crosses $\boldsymbol{\lambda}$. Note that $\lambda_{i^\star(\boldsymbol{\mu})} = \lambda_j = \theta$ and that other components of $\boldsymbol{\lambda}$ are selected to get a minimal modification of $\boldsymbol{\mu}$ to satisfy the Lipschitz constraint. $\mu_k = \cos(2\pi(-0.15 + 0.05k)), \forall k = 0, 1, \ldots, 20$, $\ell = 2\pi$ and $\theta = 0$.

The minimal value of the above problem (30) is exactly $f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ and for any k, the k-th component of $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ is

$$\min\{\max\{\theta_j^\star - \ell \left\|\boldsymbol{a}_k - \boldsymbol{a}_j\right\|_\infty, \mu_k\}, \theta_j^\star + \ell \left\|\boldsymbol{a}_k - \boldsymbol{a}_{i^\star(\boldsymbol{\mu})}\right\|_\infty\}, \tag{31}$$

where $\theta_j^\star$ is the solution of the problem (30). $\theta_j^\star$ can be found by using simple binary search. $\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ can be obtained by directly plugging $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ into the equation (4) and Proposition 1 shows that $f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \langle \boldsymbol{\omega}, \nabla f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) \rangle$. To check Assumption 2, we can use Lemma 1 as shown in Appendix C.

Implementing `FWS`**.** When the Lipschitz constant ℓ is tight, (i.e. $\max_{k \neq k'} \frac{|\mu_k - \mu_{k'}|}{\|\boldsymbol{a}_k - \boldsymbol{a}_{k'}\|_\infty} \approx \ell$), we need accurate estimate of $\boldsymbol{\mu}$ so that $\hat{\boldsymbol{\mu}}(t)$ satisfies the Lipschitz assumption, and belongs to Λ. In this case, the Lipschitz structure is too strong and in its initial design, `FWS` may use numerous rounds of forced exploration so that finally $\hat{\boldsymbol{\mu}}(t) \in \Lambda$. To circumvent this issue, we could project $\hat{\boldsymbol{\mu}}(t)$ to Λ. We use another solution that consists in artificially enlarging Λ. To this aim, we pick a Lipschitz constant ℓ' larger than ℓ, and define

$$\Lambda_{\ell'} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^K : \exists i \in [K] \text{ s.t. } \mu_i > \mu_k, \forall k \neq i \text{ and } |\mu_k - \mu_{k'}| < \ell' \|\boldsymbol{a}_k - \boldsymbol{a}_{k'}\|_\infty, \ \forall k, k' \in [K] \right\}.$$

Note that $\Lambda \subset \Lambda_{\ell'}$. Now when $\hat{\boldsymbol{\mu}}(t) \notin \Lambda$ (although there exists an unique empirical best arm under $\hat{\boldsymbol{\mu}}(t)$), we can find $\ell' > \ell$ s.t. $\hat{\boldsymbol{\mu}}(t) \in \Lambda_{\ell'}$. Inspired by this observation, we just replace in `FWS` the condition $\hat{\boldsymbol{\mu}}(t) \notin \Lambda$ by $\hat{\boldsymbol{\mu}}(t) \notin \Lambda_{\ell_{\texttt{pseudo}}(t)}$, where

$$\ell_{\texttt{pseudo}}(t) = \max \left\{ \ell, \max_{k \neq k'} \frac{|\hat{\mu}_k(t) - \hat{\mu}_{k'}(t)|}{\|\boldsymbol{a}_k - \boldsymbol{a}_{k'}\|_\infty} \right\}.$$

Note that $\ell_{\texttt{pseudo}}(t) \geq \ell$ and $\hat{\boldsymbol{\mu}}(t) \in \Lambda$ if and only if $\ell_{\texttt{pseudo}}(t) = \ell$. After this change, `FWS` does not have a forced exploration round each time $\hat{\boldsymbol{\mu}}(t) \notin \Lambda$. Removing this forced exploration condition does not affect our analysis.

Competing algorithms. As far as we know, this paper is the first to consider BAI in Lipschitz bandits. Hence, for this task, we just investigate the following algorithms and baselines:

- `FWS`: Our algorithm with parameters $r = t^{-0.9}/K$, where K is the number of arms.
- `T-D`: Track and Stop [28] with D-Tracking. `T-D` is the strongest baseline without prior knowledge of the structure, and we include it to estimate the gains achieved when exploiting the Lipschitz structure.
- `M-C`: Here we use $\ell_{\texttt{pseudo}}(t)$, introduced above, and Proposition 1 to construct the subdifferential for LMA [38]. The learning rate η_t is chosen with the knowledge of $\boldsymbol{\mu}$ (as discussed previously). Note that there is not known theoretical guarantees for LMA in Lipschitz bandits.
- $\texttt{LB}_{\texttt{Lip}}(\delta)$: $T^\star(\boldsymbol{\mu})\text{kl}(\delta, 1-\delta)$ with Lipschitz structure.

Stopping rules. `FWS` and `M-C` use the stopping rule (7) for Lipschitz bandits while `T-D` uses the same stopping rule for the unstructured bandits. The threshold is set equal to $\beta(t, \delta) = \log((\log(t) + 1)/\delta)$ for both algorithms.

F.2 Numerical experiments

We consider two experiments.

Experiment L1. In this experiment, the average rewards are given by the ℓ-Lipschitz function $f(x) = 9\cos(x)/(x^2 + 10)$, where $\ell = 0.9$. We have 20 arms with mean $\boldsymbol{\mu} = [f(x_1), \cdots, f(x_{20})]$, where $x_i = 1.25 + 0.25(i-1), \forall i = 1, \cdots, 20$, as shown in Figure 8.

In Table 10, we provide the sample complexity of the various algorithms averaged over 100 runs for confidence levels $\delta \in \{0.1, 0.01\}$. To provide a more detailed comparison, at the confidence level $\delta = 0.01$, we show the sample complexity in box-plot in Figure 11a and compare the allocation of arm draws achieved under the various algorithms in Table 11. Observe that `FWS` outperforms `M-C`, and manages to almost halve the sample complexity compared to `T-D`. Exploiting the structure yields critical improvements.

In Figure 9, we plot the number of rounds (the median over all runs) `FWS` is in force exploration or the r-subdifferential subspace used in `FWS` contains the gradient of only one function (in this round, our FW update coincides with the traditional FW update as if the objective function was smooth). In Figure 10, we provide the allocation that number of functions are involved in FW update.

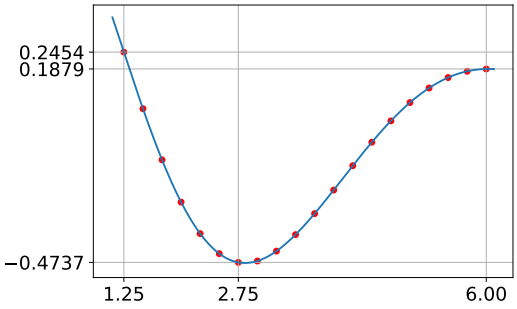


Figure 8: Experiment L1. The positions of the arms (x-axis) and their expected rewards (y-axis).

Table 10: Sample complexity for Experiment L1 averaged over 100 runs.

	FWS	M-C	T-D	$\mathrm{LB}_{\mathrm{Lip}}(\delta)$
$\delta = 0.1$	21 791	30 999	41 182	6 798
$\delta = 0.01$	30 051	41 481	56 810	17 415

Table 11: The average rewards and the allocation of arm draws (%) in Experiment L1 with $\delta = 0.01$ averaged over 100 runs.

		$\boldsymbol{a}_1$	$\boldsymbol{a}_2$	$\boldsymbol{a}_3$	$\boldsymbol{a}_4$	$\boldsymbol{a}_5$	$\boldsymbol{a}_6$	$\boldsymbol{a}_7$	$\boldsymbol{a}_8$	$\boldsymbol{a}_9$	$\boldsymbol{a}_{10}$
	Positions	1.25	1.5	1.75	2.0	2.25	2.5	2.75	3.0	3.25	3.5
	μ	0.25	0.05	-0.12	-0.27	-0.38	-0.44	-0.47	-0.47	-0.44	-0.38
Allocation (%)	FWS	29.15	1.14	0.34	0.39	0.15	0.15	0.14	0.14	0.14	0.15
	M-C	24.56	2.88	2.79	1.93	1.67	1.59	1.57	1.57	1.58	1.65
	T-D	35.08	0.85	0.39	0.38	0.38	0.38	0.38	0.38	0.38	0.38
		$\boldsymbol{a}_{11}$	$\boldsymbol{a}_{12}$	$\boldsymbol{a}_{13}$	$\boldsymbol{a}_{14}$	$\boldsymbol{a}_{15}$	$\boldsymbol{a}_{16}$	$\boldsymbol{a}_{17}$	$\boldsymbol{a}_{18}$	$\boldsymbol{a}_{19}$	$\boldsymbol{a}_{20}$
	Positions	3.75	4.0	4.25	4.5	4.75	5.0	5.25	5.5	5.75	6.0
	μ	-0.31	-0.23	-0.14	-0.06	0.01	0.07	0.12	0.16	0.18	0.19
Allocation (%)	FWS	0.17	0.21	0.29	0.43	0.79	1.56	3.04	7.25	23.00	31.57
	M-C	1.92	2.17	2.68	2.90	3.09	3.41	4.35	6.90	13.00	17.78
	T-D	0.38	0.38	0.38	0.41	0.52	1.09	2.59	7.53	19.44	28.30

Experiment L2. In the second experiment, we consider the arms with positions and average rewards presented in the second and third rows of Table 12, respectively. The reward function is Lipschitz, and the learner is informed that this function has a Lipschitz constant $\ell = 0.01$. This example is chosen because identifying the best arm $\boldsymbol{a}_1$ is hard without leveraging the Lipschitz structure. Indeed, to identify $\boldsymbol{a}_1$, the learner will need to select $\boldsymbol{a}_1$ and $\boldsymbol{a}_6$ (the second best arm) often. Imagine now that the average rewards of these two arms are well known. If the learner is not aware of the Lipschitz structure, she will need to further explore all other arms. However, if she is aware that the reward function is 0.01-Lipschitz, knowing that the average reward of $\boldsymbol{a}_6$ is roughly 1, she will deduce that the average rewards of all other arms (except $\boldsymbol{a}_1$) must be in the interval $[0.96, 1.04]$ ($\boldsymbol{a}_2$ and $\boldsymbol{a}_{10}$ are at a distance 4 from $\boldsymbol{a}_6$). These arms are then worse than $\boldsymbol{a}_1$, and an informed learner does not really need to explore them. In summary, we expect that exploiting the structure in L2 will bring significant improvement in the sample complexity.

In Table 13, we report the sample complexity of the various algorithms averaged over 100 runs for confidence levels $\delta \in \{0.1, 0.01\}$. To provide a more detailed comparison, at the confidence level $\delta = 0.01$, we show the sample complexity in box-plot in Figure 11b and compare the allocation of arm draws achieved under the various algorithms in Table 12. Observe that again, FWS outperforms M-C, and in this experiment, it manages to almost divide the sample complexity by factor 3 compared to T-D. As expected, exploiting the structure yields an even greater improvement than in Experiment L1.

Table 12: Average rewards and allocation of arm draws (%) at $\delta = 0.01$ averaged over 100 runs.

		$\boldsymbol{a}_1$	$\boldsymbol{a}_2$	$\boldsymbol{a}_3$	$\boldsymbol{a}_4$	$\boldsymbol{a}_5$	$\boldsymbol{a}_6$	$\boldsymbol{a}_7$	$\boldsymbol{a}_8$	$\boldsymbol{a}_9$	$\boldsymbol{a}_{10}$
	Positions	0	96	97	98	99	100	101	102	103	104
	μ	1.06	0.99	0.99	0.99	0.99	1	0.99	0.99	0.99	0.99
Allocations (%)	FWS	26.63	7.30	8.45	6.47	7.69	12.39	7.26	7.93	7.95	7.92
	M-C	27.85	7.76	8.05	8.16	7.99	8.81	8.21	7.86	8.09	7.22
	T-D	25.37	7.84	10.06	8.10	5.50	15.20	7.06	9.16	5.59	6.11

Table 13: Sample complexity for Experiment L2 averaged over 100 runs.

	FWS	M-C	T-D	$\mathrm{LB}_{\mathrm{Lip}}(\delta)$
$\delta = 0.1$	29 308	35 582	75 154	6 046
$\delta = 0.01$	41 909	47 759	98 188	15 490

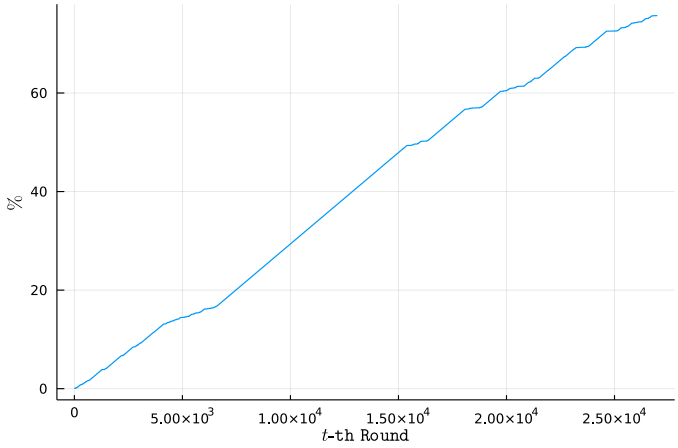


Figure 9: Number of rounds where FWS is in forced exploration or the FW update in FWS corresponds to the usual FW update. Experiment L1.

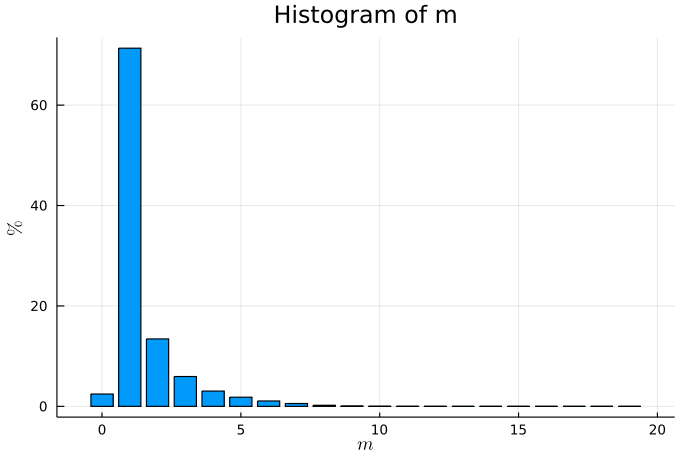
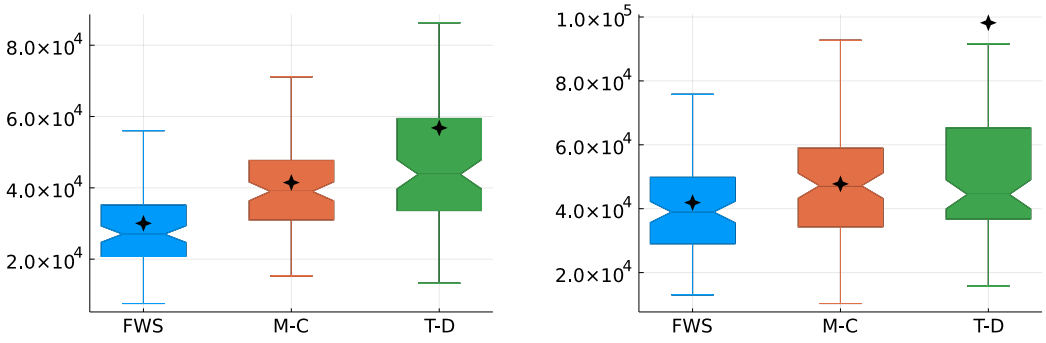


Figure 10: Proportions of rounds where we have m functions in the linear program involved in the FW updates (i.e. $m = \left|\{j : f_j(\boldsymbol{x}(t), \hat{\boldsymbol{\mu}}(t)) < F_{\hat{\boldsymbol{\mu}}(t-1)}(\boldsymbol{x}(t)) + r_t\}\right|$, and $m = 0$ in forced exploration). Experiment L1.



(a) Experiment L1

(b) Experiment L2

Figure 11: Sample complexity averaged over 100 runs at $\delta = 0.01$. The stars in the boxplots represent the averaged sample complexity and the outliers are hidden.

G Additional Examples

In this section, we present two additional examples to illustrate the applicability of our framework. We do not report any numerical experiments on these.

G.1 Threshold problem in monotone bandits

This task has applications in clinical trials. The learner aims at determining the maximum tolerable dose (MTD) (the maximum amount of the drug that can be given to a person without any potential danger). Arms represent the increasing doses, and the risk the potential adverse effects is drawn from a Gaussian distribution whose average increases with the dose. The learner is given a threshold of tolerance risk $\mathfrak{I} \in \mathbb{R}$, and wish to identify the first arm with risk that exceeds this threshold. Refer to [21] for details.

This pure exploration task can be investigated using our framework. We have: $\Lambda = \{\boldsymbol{\mu} \in \mathbb{R}^K : \mu_1 < \mu_2 < \ldots < \mu_K, \text{ and } \mu_k \neq \mathfrak{I},\ \forall k \in [K]\}$. the set of possible answer is $\mathcal{I} = [K] \cup \emptyset$: if the true answer is k, arm k is the last arm below the threshold, and $\emptyset$ refers to the case where all arms have a risk above the threshold. The set of parameters $\boldsymbol{\mu}$ for which k is the correct answer is $\mathcal{S}_k = \{\boldsymbol{\mu} \in \Lambda : \mu_k < \mathfrak{I} < \mu_{k+1}\}$, which is an open set. Observe that it is also convex. Similarly $\mathcal{S}_\emptyset = \{\boldsymbol{\mu} \in \Lambda : \mu_1 > \mathfrak{I}\}$ is also open and convex.

Let us now identify the set of confusing parameters. To simplify the presentation, we assume that $\boldsymbol{\mu}$ is such that $1 < i^\star(\boldsymbol{\mu}) < K$. The set of confusing parameters can be decomposed as $\mathrm{Alt}(\boldsymbol{\mu}) = \cup_{u \neq i^\star(\boldsymbol{\mu})} \mathcal{C}_u^{i^\star(\boldsymbol{\mu})}$, where $\mathcal{C}_u^{i^\star(\boldsymbol{\mu})} = \mathcal{S}_u$ is convex. Assumption 1 is hence verified. Let $u \neq i^\star(\boldsymbol{\mu})$ and $\boldsymbol{\omega} \in \mathring{\Sigma}$. Elementary calculus yields that

$$\overline{\boldsymbol{\lambda}_u(\boldsymbol{\omega}, \boldsymbol{\mu})} = \begin{cases} \boldsymbol{\mu} + \sum_{s=u}^{i^\star(\boldsymbol{\mu})} (\mathfrak{I} - \mu_s) \boldsymbol{e}_s & \text{if } u < i^\star(\boldsymbol{\mu}), \\ \boldsymbol{\mu} + \sum_{s=i^\star(\boldsymbol{\mu})}^{u} (\mathfrak{I} - \mu_s) \boldsymbol{e}_s, & \text{otherwise.} \end{cases}$$

This implies that

$$\nabla_{\boldsymbol{\omega}} f_u(\boldsymbol{\omega}, \boldsymbol{\mu}) = \begin{cases} \sum_{s=u}^{i^\star(\boldsymbol{\mu})} \frac{(\mathfrak{I}-\mu_s)^2}{2} \boldsymbol{e}_s & \text{if } u < i^\star(\boldsymbol{\mu}), \\ \sum_{s=i^\star(\boldsymbol{\mu})}^{u} \frac{(\mathfrak{I}-\mu_s)^2}{2} \boldsymbol{e}_s, & \text{otherwise.} \end{cases} \tag{32}$$

As shown in Proposition 1, $\langle \boldsymbol{\omega}, \nabla f_u(\boldsymbol{\omega}, \boldsymbol{\mu}) \rangle = f_u(\boldsymbol{\omega}, \boldsymbol{\mu})$, and thus

$$f_\ell(\boldsymbol{\omega}, \boldsymbol{\mu}) = \begin{cases} \sum_{s=u}^{i^\star(\boldsymbol{\mu})} \omega_s \frac{(\mathfrak{I}-\mu_s)^2}{2} & \text{if } u < i^\star(\boldsymbol{\mu}), \\ \sum_{s=i^\star(\boldsymbol{\mu})}^{u} \omega_s \frac{(\mathfrak{I}-\mu_s)^2}{2}, & \text{otherwise.} \end{cases} \tag{33}$$

In view of (32), Assumptions 2 (i) holds (as $\nabla_{\boldsymbol{\omega}} f_j$ is bounded). Assumption 2 (ii) can be easily verified by differentiating (32) with respect to $\boldsymbol{\omega}$ or using the sufficient condition provided in Appendix C.

G.2 Top-m arms indentification in dueling bandits

The top-m arms identification task consists in identifying the m best arms. To solve this task in dueling bandits [53], the learner is allowed to sequentially pick pairs of arms. If the pair (i, j) is selected, the learner observes the realization of a Bernoulli r.v. with mean $\mu_{i,j}$. If $\mu_{i.j} > 1/2$, we say that arm i is better than arm j. The preference matrix $\boldsymbol{\mu} = (\mu_{i,j})$ is assumed to satisfy:

(a) $\mu_{i,j} = 1 - \mu_{j,i}, \quad \forall (i,j) \in [K]^2.$

(b) $\mu_{i,i} = \dfrac{1}{2}.$

(c) if $\min(\mu_{i,j}, \mu_{j,k}) \geq \dfrac{1}{2}$, then $\mu_{i,k} \geq \dfrac{1}{2}.$

Under this assumption, $\boldsymbol{\mu}$ induces a total order $\succ_{\boldsymbol{\mu}}$, defined by $i \succ_{\boldsymbol{\mu}} j$ if and only if $\mu_{i,j} \geq 1/2$. Also note that under this assumption, $\boldsymbol{\mu}$ is defined only through its entries above the diagonal, hence by $\frac{K(K-1)}{2}$ parameters. We denote by $\sigma_{\boldsymbol{\mu}}$ a permutation of $[K]$, such that $\sigma_{\boldsymbol{\mu}}(1) \succ_{\boldsymbol{\mu}} \sigma_{\boldsymbol{\mu}}(2) \succ_{\boldsymbol{\mu}} \ldots \succ_{\boldsymbol{\mu}} \sigma_{\boldsymbol{\mu}}(m) \succ_{\boldsymbol{\mu}} \ldots \succ_{\boldsymbol{\mu}} \sigma_{\boldsymbol{\mu}}(K)$. We are ready to define

$$\Lambda = \left\{ \boldsymbol{\mu} \in (0,1)^{\frac{K(K-1)}{2}} : \boldsymbol{\mu} \text{ satisfies (c) and (d)} \right\},$$

where (d) ensures that the set of m best arms is unique:

$$\text{(d)} \quad \text{we can select } \sigma_{\boldsymbol{\mu}} \text{ such that } \mu_{\sigma_{\boldsymbol{\mu}}(m),\sigma_{\boldsymbol{\mu}}(m+1)} > \frac{1}{2}.$$

In our framework, the set of answers is $\mathcal{I} = \{\mathcal{A} \subset [K] : |\mathcal{A}| = m\}$ and for any $\mathcal{A} \in \mathcal{I}$, $\mathcal{S}_{\mathcal{A}} = \{\boldsymbol{\mu} \in \Lambda : \mu_{i,j} > 1/2 \text{ if } i \in \mathcal{A} \text{ but } j \notin \mathcal{A}\}$. We can readily check that $\mathcal{S}_{\mathcal{A}}$ is open.

Now let $\boldsymbol{\mu} \in \Lambda$. Assume w.l.o.g. that $\sigma_{\boldsymbol{\mu}} = Id$ (identity permutation); in particular $1 \succ_{\boldsymbol{\mu}} 2 \succ_{\boldsymbol{\mu}} \ldots \succ_{\boldsymbol{\mu}} m \succ_{\boldsymbol{\mu}} \ldots \succ_{\boldsymbol{\mu}} K$. The true answer is $[m]$. If we define:

$$\mathcal{J}_{[m]} = \{\sigma \in \Theta : \exists k > m \text{ s.t } \sigma(k) \leq m\} \quad \text{and} \quad \forall \sigma \in \mathcal{J}_{[m]},\ \mathcal{C}_{\sigma}^{[m]} = \{\boldsymbol{\lambda} \in \Lambda : \sigma_{\boldsymbol{\lambda}} = \sigma\},$$

where Θ is the set of all the permutations of $[K]$, then $\text{Alt}(\boldsymbol{\mu}) = \cup_{\sigma \in \mathcal{J}_{[m]}} \mathcal{C}_{\sigma}^{[m]}$ and $\mathcal{C}_{\sigma}^{[m]}$ is a convex set (for any $\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}} \in \mathcal{C}_{\sigma}^{[m]}$, for any of their convex combinations $\boldsymbol{\lambda}'$, we have $\sigma_{\boldsymbol{\lambda}'} = \sigma$). Assumption 1 is hence verified. For each $\sigma \in \mathcal{J}_{[m]}$, we discuss the most confusing parameter in the set $\mathcal{C}_{\sigma}^{[m]}$ against $\boldsymbol{\mu}$ at the point $\boldsymbol{\omega}$. Namely, we solve

$$\min_{\boldsymbol{\lambda} \in \mathcal{C}_{\sigma}^{[m]}} \sum_{k<\ell} \omega_{k,\ell} d(\mu_{k,\ell}, \lambda_{k,\ell}),$$

where $\omega_{k,\ell}$ is the proportion of times that (k,ℓ) is pulled (in dueling bandits, pulling (k,ℓ) is equivalent to pulling (ℓ,k), hence we only count for $k < \ell$). For any $k < \ell$, we can readily show that

$$\overline{\boldsymbol{\lambda}_{\sigma}(\boldsymbol{\omega}, \boldsymbol{\mu})}_{k,\ell} = \begin{cases} \frac{1}{2} & \text{if } \sigma(\ell) < \sigma(k), \\ \mu_{k,\ell}, & \text{otherwise.} \end{cases}$$

This implies that

$$\nabla_{\boldsymbol{\omega}} f_{\sigma}(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_{\substack{k<\ell \\ \sigma(\ell)<\sigma(k)}} d(\mu_{k,\ell}, \frac{1}{2}). \tag{34}$$

As shown in Proposition 1, $\langle \boldsymbol{\omega}, \nabla f_{\sigma}(\boldsymbol{\omega}, \boldsymbol{\mu}) \rangle = f_{\sigma}(\boldsymbol{\omega}, \boldsymbol{\mu})$, and thus

$$f_{\sigma}(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_{\substack{k<\ell \\ \sigma(\ell)<\sigma(k)}} \omega_{k,\ell} d(\mu_{k,\ell}, \frac{1}{2}). \tag{35}$$

In view of (34), Assumptions 2 (i) holds (as $\nabla_{\boldsymbol{\omega}} f_{\sigma}$ is bounded). Assumption 2 (ii) can be easily verified by differentiating (34) with respect to $\boldsymbol{\omega}$ or using the sufficient condition provided in Appendix C.

H Zero-sum Game: the Equivalent Linear Program

In this section, we explain how to transform the zero-sum game (11) used in our FW update to a simple Linear Program (LP). The zero-sum game is:

$$\boldsymbol{z}(t) \leftarrow \operatorname*{argmax}_{\boldsymbol{z}\in\Sigma} \min_{h\in H_{F_{\hat{\boldsymbol{\mu}}(t-1)}}(\boldsymbol{x}(t-1),r_t)} \langle \boldsymbol{z}-\boldsymbol{x}(t-1), h\rangle$$

For clarity, we use the following notations: $\boldsymbol{x} = \boldsymbol{x}(t-1) \in \mathring{\Sigma}$, $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}(t-1)$, $r = r_t$, and we assume w.l.o.g. that $j = 1, \dots, J$ are the indexes in $\mathcal{J}_{i^\star(\boldsymbol{\mu})}$ verifying $f_j(\boldsymbol{x}, \boldsymbol{\mu}) < F_{\boldsymbol{\mu}}(\boldsymbol{x}) + r$. Hence, $H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}, r) = \operatorname{cov}(\{\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{x}, \boldsymbol{\mu})\}_{j=1}^J)$.

Define the payoff matrix $M \in \mathbb{R}^{K\times J}$ with $M_{k,j} = \langle \boldsymbol{e}_k - \boldsymbol{x}, \nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{x}, \boldsymbol{\mu})\rangle$, for all $k \in [K], j \in [J]$. Then the problem (11) can be formulated as

$$\begin{aligned} &\max_{\boldsymbol{z}\in\Sigma} \min_{\boldsymbol{y}\in\mathbb{R}^J} \boldsymbol{z}^\top M \boldsymbol{y} \\ &\text{s.t. } y_j \geq 0, \forall j \in [J] \text{ and } y_1 + y_2 + \ldots + y_J = 1. \end{aligned} \tag{36}$$

Denote by $(\boldsymbol{z}^\star, \boldsymbol{y}^\star)$ the solution of the problem (36). Then the solution $\boldsymbol{z}(t)$ of (11) is

$$\boldsymbol{z}(t) = \boldsymbol{x} + \sum_k z_k^\star (\boldsymbol{e}_k - \boldsymbol{x}) = \boldsymbol{z}^\star.$$

Standard textbooks in game theory present procedures to solve (36) by transforming it into an LP [37, 48, 52]. We give below the method we used in our experiments.

If $\boldsymbol{z} \in \mathbb{R}^K$ is fixed, the best response of the $\boldsymbol{y}$-player is a pure strategy. The pay-off of this strategy is of course $\min\{(\boldsymbol{z}^\top M)_1, \dots, (\boldsymbol{z}^\top M)_J\}$. As a consequence, the optimal strategy for the $\boldsymbol{z}$-player is to solve the following problem:

$$\max_{\boldsymbol{z}\in\Sigma} \left\{\min\{(\boldsymbol{z}^\top M)_1, \dots, (\boldsymbol{z}^\top M)_J\}\right\}. \tag{37}$$

(37) is transformed to an LP by introducing an auxiliary parameter $u \in \mathbb{R}$ as a lower bound of $(\boldsymbol{z}^\top M)_j$. The problem (37) becomes

$$\begin{aligned} &\max_{\boldsymbol{z}\in\Sigma, u\in\mathbb{R}} u \\ &\text{s.t. } (\boldsymbol{z}^\top M)_j \geq u, \forall j = 1, \dots, J. \end{aligned} \tag{38}$$

I Asymptotic Sample Complexity Upper Bound

This section is devoted to the proof of Theorem 1. This theorem summarizes our analysis of FWS, and its proof heavily relies on results presented in subsequent appendices. Specifically in Appendix J, we state and prove concentration results quantifying how $\hat{\boldsymbol{\mu}}(t)$ concentrates around $\boldsymbol{\mu}$, and how the FW update in FWS differs from the same update obtained assuming that $\boldsymbol{\mu}$ is known. In turn, to establish these results, we will need continuity arguments presented in Appendix K (e.g., our FW update needs to be continuous in $\boldsymbol{\omega}$, $\boldsymbol{\mu}$ and the parameter r). In Appendix L, we provide useful results related to the convergence of our variant of the FW algorithm. The proof of Theorem 1 will finally require us to study the tracking rule, which is done in Appendix M.

Coming back to the present appendix, we start with the almost sure upper bound and then proceed with the expected upper bound.

I.1 Almost sure upper bound

The proof starts by defining the event

$$\mathcal{E} = \left\{ F_{\boldsymbol{\mu}}(\boldsymbol{\omega}(t)) \xrightarrow{t\to\infty} F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) \text{ and } \hat{\boldsymbol{\mu}}(t) \xrightarrow{t\to\infty} \boldsymbol{\mu} \right\}.$$

We know that $\mathbb{P}_{\boldsymbol{\mu}}[\mathcal{E}] = 1$ based on the Theorem 7 in Appendix L and on the law of the large number (every arm will be pulled infinite times because of forced exploration rounds). Since $F_{\boldsymbol{\mu}}(\boldsymbol{\omega})$ is continuous w.r.t. $\boldsymbol{\mu}$ (Lemma 6 in Appendix K), we also have that $F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}) \xrightarrow{t\to\infty} F_{\boldsymbol{\mu}}(\boldsymbol{\omega})$ uniformly over $\boldsymbol{\omega} \in \mathring{\Sigma}$ almost surely. This further implies that $F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \xrightarrow{t\to\infty} F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))$ a.s. (by applying triangular inequality). Let $\epsilon \in (0,1)$. Under the event $\mathcal{E}$, there exists a constant t_1 such that for $t \geq t_1$, $F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq (1-\epsilon) F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))$. Hence, denoting $\mathbb{N}^* = \mathbb{N} \cup \{\infty\}$, we get:

$$\begin{aligned}
\tau &= \inf\left\{ t \in \mathbb{N}^* : t F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq \beta(t,\delta) \right\} \\
&\leq t_1 \vee \inf\left\{ t \in \mathbb{N}^* : t(1-\epsilon) F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) \geq \beta(t,\delta) \right\} \\
&\leq t_1 \vee \inf\left\{ t \in \mathbb{N}^* : t \geq \frac{\beta(t,\delta) T^\star(\boldsymbol{\mu})}{(1-\epsilon)} \right\} \\
&\leq c_1(\Lambda) \vee t_1 \vee \inf\left\{ t \in \mathbb{N}^* : t \geq \frac{\log(c_2(\Lambda) t) T^\star(\boldsymbol{\mu})}{(1-\epsilon)\delta} \right\}.
\end{aligned}$$

where the second inequality stems from the fact that $T^\star(\boldsymbol{\mu})^{-1} = F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))$ and the final inequality stems from (7). Finally, applying Lemma 4 (presented at the end of this appendix) with $\alpha = 1$, $c_1 = \frac{(1-\epsilon)\delta}{T^\star(\boldsymbol{\mu})}$ and $c_2 = c_2(\Lambda)$ to the above inequality yields that

$$\tau \leq c_1(\Lambda) + t_1 + \frac{T^\star(\boldsymbol{\mu})}{(1-\epsilon)\delta}\left[\log\left(\frac{T^\star(\boldsymbol{\mu}) c_2(\Lambda) e}{\delta(1-\epsilon)} \right) + \log\log\left(\frac{T^\star(\boldsymbol{\mu}) c_2(\Lambda)}{\delta(1-\epsilon)} \right) \right].$$

This implies $\mathbb{P}_{\boldsymbol{\mu}}\left[\limsup_{\delta\to 0} \frac{\tau}{\log(1/\delta)} \leq T^\star(\boldsymbol{\mu}) \right] = 1$ and $\mathbb{P}_{\boldsymbol{\mu}}\left[\tau < \infty\right] = 1$, for all $\delta \in (0,1)$. The fact that the algorithm is δ-PAC directly follows from the property of $\beta(t,\delta)$, i.e. (8).

I.2 Expected upper bound

Let $\epsilon \in (0,1)$. Based on the conditions imposed on $\{r_t\}$, there exist $T_{\epsilon,L} \in \mathbb{N}$ such that

$$\sum_{s=1}^{t} r_s < t\epsilon \text{ and } t r_t > L \text{ if } t \geq T_{\epsilon,L}. \tag{39}$$

Let $M = \max\{ \left(\frac{32D+3L}{\epsilon}\right)^{11}, T_{\epsilon,L}^{\frac{11}{8}}, (4K+1)^{\frac{11}{8}} \}$ and for any $T \geq M$, define the functions

$$\begin{cases}
\underline{h}(T) = \min\left\{ t \in \mathbb{N} : t \geq T^{\frac{8}{11}} + 2, \sqrt{t/K} \in \mathbb{N} \right\}, \\
\overline{h}(T) = \min\left\{ t \in \mathbb{N} : t \geq T^{\frac{2}{11}} \underline{h}(T), \sqrt{t/K} \in \mathbb{N} \right\}.
\end{cases} \tag{40}$$

We are now ready to introduce our "good" events $\mathcal{E}_{1,\epsilon}(T) = \left(\bigcap_{t=\underline{h}(T)}^{T} \mathcal{E}_{1,\epsilon}^{(t)}\right)$ and $\mathcal{E}_{2,\epsilon}(T) = \left(\bigcap_{t=\underline{h}(T)}^{T} \mathcal{E}_{2,\epsilon}^{(t)}\right)$, where

$$\mathcal{E}_{1,\epsilon}^{(t)} = \left\{ \max_{\boldsymbol{z}\in\Sigma} \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}(t-1),r_t)} \langle \boldsymbol{z}-\boldsymbol{x}(t-1), h\rangle - \epsilon < \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}(t-1),r_t)} \langle \boldsymbol{z}(t)-\boldsymbol{x}(t-1), h\rangle \right\}$$
$$\mathcal{E}_{2,\epsilon}^{(t)} = \left\{ \hat{\boldsymbol{\mu}}(t) \in \mathcal{S}_{i^\star(\boldsymbol{\mu})} \text{ and } \left|F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega})\right| < \epsilon,\ \forall \boldsymbol{\omega} \in \mathring{\Sigma} \right\}.$$

$\mathcal{E}_{1,\epsilon}^{(t)}$ can be seen as the event that the error of solution in FW-update (11) is bounded by ϵ, which yields that $F_{\boldsymbol{\mu}}(\boldsymbol{x}(t))$ converges to $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star)$. As a consequence of the tracking rule, $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}(t))$ converges to $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star)$ as well. More precisely, as stated in Lemma 3 at the end of this appendix, under $\mathcal{E}_{1,\epsilon}(T)$, $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega}(t)) < 5\epsilon$. Now, $\mathcal{E}_{2,\epsilon}^{(t)}$ is the event that the error of objective function is bounded by ϵ uniformly, so that FWS can stop while it is close to the real maximum. Overall, under $\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T)$, we obtain that

$$\begin{aligned}
\min\{\tau, T\} &\le \overline{h}(T) + \sum_{t=\overline{h}(T)}^{T} \mathbb{1}\{\tau > t\} \\
&\le \overline{h}(T) + \sum_{m=\overline{h}(T)}^{T} \mathbb{1}\{t F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) < \beta(t,\delta)\} \\
&\le \overline{h}(T) + \sum_{t=\overline{h}(T)}^{T} \mathbb{1}\{t(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon) < \beta(t,\delta)\} \\
&\le \overline{h}(T) + \frac{\beta(T,\delta)}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon},
\end{aligned}$$

where the third inequality is due to the fact that under event $\mathcal{E}_{2,\epsilon}(T)$ and in view of Lemma 3, when $t \ge \overline{h}(T)$, we have $F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \ge F_{\boldsymbol{\mu}}(\boldsymbol{\omega}(t)) - \epsilon \ge F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon$.

Now introduce a constant

$$T_0(\delta) = \inf\{T \in \mathbb{N} : \overline{h}(T) + \frac{\beta(T,\delta)}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon} \le T\}.$$

The above inequalities show that $\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T) \subset \{\tau \le T\}$. Therefore,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}[\tau] &\le \sum_{T=1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[\tau \ge T] \\
&\le \left(\frac{32D+3L}{\epsilon}\right)^{11} + T_{\epsilon,L}^{\frac{11}{8}} + (4K+1)^{\frac{11}{8}} + T_0(\delta) + \sum_{T=M+1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{1,\epsilon}(T))^c].
\end{aligned} \tag{41}$$

The term $\sum_{T\ge 1} \mathbb{P}_{\boldsymbol{\mu}}[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T))^c]$ on the right-hand side of the inequality (41) can be upper bounded by concentration inequalities, which we summarize in Lemma 2 and prove in the next appendix. As for $T_0(\delta)$, we further introduce another small constant $\tilde{\epsilon} \in (0,1)$ and observe that

$$T - \overline{h}(T) \ge \frac{T}{1+\tilde{\epsilon}} \text{ when } T \ge \left(\frac{2}{\tilde{\epsilon}}\right)^{11}.$$

Therefore, based on the above fact, and (7),

$$
\begin{aligned}
T_0(\delta) &\leq \left(\frac{2}{\tilde{\epsilon}}\right)^{11} + \inf\left\{T \in \mathbb{N} : \frac{\beta(T,\delta)}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon} \leq \frac{T}{1+\tilde{\epsilon}}\right\} \\
&\leq \max\left\{\left(\frac{2}{\tilde{\epsilon}}\right)^{11}, c_1(\Lambda)\right\} + \inf\left\{T \in \mathbb{N} : \frac{1}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon} \log(\frac{c_2(\Lambda)T}{\delta}) \leq \frac{T}{1+\tilde{\epsilon}}\right\} \\
&\leq \max\left\{\left(\frac{2}{\tilde{\epsilon}}\right)^{11}, c_1(\Lambda)\right\} \\
&+ \frac{1+\tilde{\epsilon}}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon}\left[\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)e}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon)}\right) + \log\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon)}\right)\right],
\end{aligned} \tag{42}
$$

where the second inequality is due to (7) and the last inequality is a consequence of Lemma 4 with $\alpha = 1$, $c_1 = (F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon)/(1+\tilde{\epsilon})$ and $c_2 = c_2(\Lambda)/\delta$. Substituting the upper bounds provided by (42) and Lemma 2 into (41), we obtain that

$$
\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau]}{\log(1/\delta)} \leq \frac{1+\tilde{\epsilon}}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon}.
$$

Since ϵ and $\tilde{\epsilon}$ can be arbitrary small and $T^\star(\boldsymbol{\mu}) = \frac{1}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))}$, we get the desired result.

I.3 Additional lemmas

Lemma 2. *Under Assumptions 1, we have*

$$
\sum_{T=M}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{1,\epsilon}(T))^c] < \infty.
$$

Refer to Appendix J for a proof.

Lemma 3. *For any $T \geq M = \max\{\left(\frac{32D+3L}{\epsilon}\right)^{11}, T_{\epsilon,L}^{\frac{11}{8}}, (4K+1)^{\frac{11}{8}}\}$, under event $\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T)$ and Assumption 2, `FWS` algorithm with a sequence $\{r_t\}_{t\geq 1}$, satisfying (i), (ii) stated in Theorem 1 attains that*

$$
F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega}(t)) \leq 5\epsilon, \ \forall t = \overline{h}(T), \overline{h}(T)+1, \ldots, T.
$$

Refer to Appendix L.3 for a proof.

Lemma 4 (Lemma 18 in [20])**.** *For $\alpha \in [1, e/2]$, any two constants c_1, c_2,*

$$
x = \frac{1}{c_1}\left[\log\left(\frac{c_2 e}{c_1^\alpha}\right) + \log\log\left(\frac{c_2}{c_1^\alpha}\right)\right]
$$

is such that $c_1 x \geq \log(c_2 x^\alpha)$.

J Concentration Results

This section presents the proof of Lemma 2 and the necessary technical lemmas (see Appendix J.2). We restate the lemma:

Lemma 2. *Under Assumptions 1, we have*

$$\sum_{T=1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}}[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{1,\epsilon}(T))^c] < \infty.$$

J.1 Proof of Lemma 2

We first derive sufficient conditions for the events $\mathcal{E}_{1,\epsilon}(T)$ and $\mathcal{E}_{2,\epsilon}(T)$ to hold separately. Then, we will conclude applying the concentration inequality.

(i) The event $\mathcal{E}_{1,\epsilon}(T)$:
Let $t = \underline{h}(T), \dots, T$. Applying the second part of Theorem 3 in Appendix K with $\boldsymbol{\omega} = \boldsymbol{x}(t-1)$, $r = r_t$, and $\boldsymbol{\pi} = \hat{\boldsymbol{\mu}}(t-1)$, and hence $\boldsymbol{z}(\boldsymbol{\omega}, r, \boldsymbol{\pi}) = \boldsymbol{z}(t)$, we get that: if $\|\hat{\boldsymbol{\mu}}(t-1) - \boldsymbol{\mu}\|_\infty < \xi_{1,\epsilon}$,

$$\max_{\boldsymbol{z}\in\Sigma} \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}(t-1),r_t)} \langle \boldsymbol{z} - \boldsymbol{x}(t-1), h\rangle - \epsilon < \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}(t-1),r_t)} \langle \boldsymbol{z}(t) - \boldsymbol{x}(t-1), h\rangle.$$

From the definition of $\mathcal{E}_{1,\epsilon}^{(t)}$, we deduce that:

$$\mathcal{E}_{1,\epsilon}^{(t)} \subset \{\|\hat{\boldsymbol{\mu}}(t-1) - \boldsymbol{\mu}\|_\infty < \xi_{1,\epsilon}\}\,,\ \forall t = \underline{h}(T), \dots, T.$$

(ii) The event $\mathcal{E}_{2,\epsilon}(T)$:
From Lemma 6 in Appendix K, we directly deduce that

$$\mathcal{E}_{2,\epsilon}^{(t)} \subset \{\|\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}\|_\infty < \xi_{2,\epsilon}\}\,,\ \forall t = \underline{h}(T), \dots, T.$$

Summarizing (i), (ii), we get that

$$\mathbb{P}_{\boldsymbol{\mu}}\left[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T))^c\right] \le \sum_{t=\underline{h}(T)-1}^{T} \sum_{k=1}^{K} \mathbb{P}_{\boldsymbol{\mu}}\left[|\hat{\mu}_k(t) - \mu_k| \ge \xi(\epsilon)\right], \tag{43}$$

where $\xi(\epsilon) = \min\{\xi_{1,\epsilon}, \xi_{2,\epsilon}\}$. To ensure the distance between the $\hat{\boldsymbol{\mu}}(t)$ and $\boldsymbol{\mu}$ is small, we need to pull each arm sufficiently often up to t. From Lemma 13 in Appendix M, we have

$$\min_k t x_k(t) \ge \sqrt{\frac{t}{K}} - 1, \forall t \ge 4K.$$

Hence, Lemma 12 from Appendix M implies that

$$N_k(t) \ge \sqrt{\frac{t}{K}} - K, \forall k \in [K], t \ge 4K.$$

Applying Chernoff inequalities yields that $\forall k \in [K], t \ge 4K$,

$$\mathbb{P}_{\boldsymbol{\mu}}\left[|\hat{\mu}_k(t) - \mu_k| \ge \xi(\epsilon)\right] \le e^K \left[\exp\left(-\sqrt{t} A_k^-\right) + \exp\left(-\sqrt{t} A_k^+\right)\right], \tag{44}$$

where $A_k^- = d(\mu_k - \xi(\epsilon), \mu_k)/\sqrt{K}$ and $A_k^+ = d(\mu_k + \xi(\epsilon), \mu_k)/\sqrt{K}$. Substituting the upper bound (44) into (43), we get using a union bound for any $T \ge M$,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}}\left[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T))^c\right] &\le e^K \sum_{k=1}^{K} \sum_{t=\underline{h}(T)-1}^{T} \left[\exp\left(-\sqrt{t} A_k^-\right) + \exp\left(-\sqrt{t} A_k^+\right)\right] \\
&\le e^K \sum_{k=1}^{K} \int_{T^{\frac{8}{11}}}^{\infty} \left[\exp\left(-\sqrt{t} A_k^-\right) + \exp\left(-\sqrt{t} A_k^+\right)\right] dt,
\end{aligned}$$

where the second inequality follows from the definition (40) of $\underline{h}(T)$. We then apply Lemma 5 presented below with $\alpha = \frac{8}{11}, \beta = \frac{1}{2}$ and $A = A_k^+ (= A_k^-$ resp.) and deduce that

$$\begin{aligned}
\sum_{T=M}^{\infty} \mathbb{P}_{\boldsymbol{\mu}} \left[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T))^c \right] &\leq e^K \sum_{k=1}^{K} \int_0^{\infty} \left(\int_{T^{\frac{8}{11}}}^{\infty} \exp\left(-\sqrt{t} A_k^-\right) + \exp\left(-\sqrt{t} A_k^+\right) dt \right) dT \\
&= e^K \sum_{k=1}^{K} \frac{2\Gamma(\frac{19}{4})}{d(\mu_k - \xi(\epsilon), \mu_k)^{\frac{19}{4}}} + \frac{2\Gamma(\frac{19}{4})}{d(\mu_k + \xi(\epsilon), \mu_k)^{\frac{19}{4}}} \\
&< 34 e^K \sum_{k=1}^{K} \frac{1}{d(\mu_k - \xi(\epsilon), \mu_k)^{\frac{19}{4}}} + \frac{1}{d(\mu_k + \xi(\epsilon), \mu_k)^{\frac{19}{4}}} < \infty,
\end{aligned}$$

where the second inequality is due to $\Gamma(\frac{19}{4}) < 17$. This concludes the proof.

J.2 Technical lemmas

Lemma 5. *Let $\alpha, \beta \in (0,1)$ and $A > 0$.*

$$\int_0^{\infty} \left(\int_{T^{\alpha}}^{\infty} \exp(-At^{\beta}) dt \right) dT = \frac{\Gamma(\frac{1}{\alpha\beta} + \frac{1}{\beta})}{\beta A^{\frac{1}{\alpha\beta} + \frac{1}{\beta}}}.$$

Proof.

$$\begin{aligned}
\int_0^{\infty} \left(\int_{T^{\alpha}}^{\infty} \exp(-At^{\beta}) dt \right) dT &= \int_0^{\infty} \alpha T^{\alpha} \exp(-AT^{\alpha\beta}) dT \\
&= \frac{1}{\beta A^{\frac{1}{\alpha\beta} + \frac{1}{\beta}}} \int_0^{\infty} x^{\frac{1}{\alpha\beta} + \frac{1}{\beta} - 1} e^{-x} dx \\
&= \frac{\Gamma(\frac{1}{\alpha\beta} + \frac{1}{\beta})}{\beta A^{\frac{1}{\alpha\beta} + \frac{1}{\beta}}}.
\end{aligned}$$

□

K Continuity Arguments

The main goal of this section is to prove Proposition 1. We also state and prove Theorem 3 and Lemma 6. These results are used in Appendix J.

In the first subsection K.1, we present some of the ingredients used to establish our continuity results. The proofs of Proposition 1, Theorem 3 and Lemma 6 are presented in K.2, K.3 and K.4, respectively.

Theorem 3. *For any $\epsilon > 0$, there exist a constant $\xi_{1,\epsilon} > 0$, which depends on $\boldsymbol{\mu}$ and ϵ, such that if $\|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty < \xi_{1,\epsilon}$, then $\boldsymbol{\mu} \in \Lambda$,*

$$\left| \max_{\boldsymbol{z}\in\Sigma} \min_{h\in H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega},r)} \langle \boldsymbol{z}-\boldsymbol{\omega}, h\rangle - \max_{\boldsymbol{z}\in\Sigma} \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega},r)} \langle \boldsymbol{z}-\boldsymbol{\omega}, h\rangle \right| < \frac{\epsilon}{2},\ \forall (\boldsymbol{\omega}, r) \in \mathring{\Sigma} \times (0,1), \tag{45}$$

and

$$\left| \min_{h\in H_{F_{\boldsymbol{\pi}}(\boldsymbol{\omega},r)}} \langle \boldsymbol{z}-\boldsymbol{\omega}, h\rangle - \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega},r)} \langle \boldsymbol{z}-\boldsymbol{\omega}, h\rangle \right| < \frac{\epsilon}{2},\ \forall (\boldsymbol{z}, \boldsymbol{\omega}, r) \in \Sigma \times \mathring{\Sigma} \times (0,1). \tag{46}$$

As a consequence, if we fix some $(\boldsymbol{\omega}, r, \boldsymbol{\pi}) \in \mathring{\Sigma} \times (0,1) \times \Lambda$, where $\|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty < \xi_{1,\epsilon}$, and further select $\boldsymbol{z}(\boldsymbol{\omega}, r, \boldsymbol{\pi}) \in \operatorname{argmax}_{\boldsymbol{z}\in\Sigma} \min_{h\in H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega},r)} \langle \boldsymbol{z}-\boldsymbol{\omega}, h\rangle$, the above two inequalities yield that

$$\max_{\boldsymbol{z}\in\Sigma} \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega},r)} \langle \boldsymbol{z}-\boldsymbol{\omega}, h\rangle - \epsilon < \min_{h\in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega},r)} \langle \boldsymbol{z}(\boldsymbol{\omega}, r, \boldsymbol{\pi})-\boldsymbol{\omega}, h\rangle.$$

Lemma 6. *For any $\epsilon > 0$, there is $\xi_{2,\epsilon} > 0$, which depends on $\boldsymbol{\mu}$ and ϵ, s.t. if $\|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty < \xi_{2,\epsilon}$, then*

$$\boldsymbol{\pi} \in \mathcal{S}_{i^\star(\boldsymbol{\mu})} \text{ and } |F_{\boldsymbol{\pi}}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega})| < \epsilon, \forall \boldsymbol{\omega} \in \mathring{\Sigma}.$$

K.1 Continuity and differentiablity of value functions

We introduce some definitions and results taken from [16], and also used recently in [11, 10] in the bandit literature. [11] concerns the continuity of the optimal allocation when there are multiple correct answers for active learning. [10] applies it for the regret minimization problem, but it is restricted to the single-valued analysis.

Definition 2. *Let $f : U \to \mathbb{R}$ be a function where U is a non-empty subset of a topological space. The level sets of f is defined as for $y \in \mathbb{R}$,*

$$L_f(y, U) = \{x \in U \ :\ f(x) \leq y\},$$
$$L_f^<(y, U) = \{x \in U \ :\ f(x) < y\}.$$

*We say that f is **lower semi-continuous** on U if all the level sets $L_f(y, U)$ are closed. It is **inf-compact** on U if all these level sets are compact. And it is **upper semi-continuous** if all the strict level sets $L_f^<(y, U)$ are open.*

Suppose $\mathbb{X}$ and $\mathbb{Y}$ are Hausdorff topological spaces. Let $u : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ be a function and $\Phi : \mathbb{X} \rightrightarrows \mathbb{S}(\mathbb{Y})$ be a set-valued function, where $\mathbb{S}(\mathbb{Y})$ is the set of non-empty subsets of $\mathbb{Y}$. We are interest in a minimization problem of the form:

$$v(x) = \inf_{y\in\Phi(x)} u(x, y),$$
$$\Phi^*(x) = \{y \in \Phi(x) \ :\ u(x, y) = v(x)\}.$$

For $U \subset \mathbb{X}$, let the graph of Φ restricted to U be $Gr_U(\Phi) = \{(x, y) \in U \times \mathbb{Y} \ :\ y \in \Phi(x)\}$.

Definition 3. *A function $u : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}}$ is called $\mathbb{K}$**-inf-compact** on $Gr_{\mathbb{X}}(\Phi)$ if for all non-empty compact subset C of $\mathbb{X}$, u is inf-compact on $Gr_C(\Phi)$.*

There are two versions of Berge's theorem used in our paper. The first one asks Φ to be compact-valued. The second one relaxes this assumption but requires an additional assumption on the object function u. Besides, we introduce $\mathbb{K}(\mathbb{X}) = \{F \in \mathbb{S}(\mathbb{X}) : F \text{ is compact }\}$.

Theorem 4 ([4]). *Let $\mathbb{X}$ and $\mathbb{Y}$ be Hausdorff topological spaces. Assume that*

- $\Phi : \mathbb{X} \rightrightarrows \mathbb{K}(\mathbb{X})$ *is continuous (i.e. both lower and upper hemicontinous),*
- $u : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ *is continuous.*

Then the function $v : \mathbb{X} \to \mathbb{R}$ is continuous and the solution multifunction $\Phi^* : \mathbb{X} \to \mathbb{S}(\mathbb{Y})$ is upper hemicontinuous and compact valued.

Theorem 5 ([16]). *Assume that*

- $\mathbb{X}$ *is compactly generated,*
- $\Phi : \mathbb{X} \rightrightarrows \mathbb{S}(\mathbb{Y})$ *is lower hemicontinuous,*
- $u : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ *is $\mathbb{K}$-inf-compact and upper semi-continuous on $Gr_{\mathbb{X}}(\Phi)$.*

Then the function $v : \mathbb{X} \to \mathbb{R}$ is continuous and the solution multifunction $\Phi^* : \mathbb{X} \rightrightarrows \mathbb{S}(\mathbb{Y})$ is upper hemicontinuous and compact valued.

All the above theorems are about the continuity of the value function of a parameterized optimization problem. We also need an additional lemma to guarantee the differentiability. The following lemma is one of the variant of the envelope theorem, which provides an important tool in optimization and has several applications in economics.

Lemma 7 (Corollary 299 in [6]). *Let $\mathbb{X}$ be a metric space and Y is a nonempty open subset in $\mathbb{R}^K$. Let $u : \mathbb{X} \times Y \to \mathbb{R}$ and assume $\frac{\partial u}{\partial y}$ exists and is continuous in $\mathbb{X} \times Y$. For each $y \in Y$, let $x^\star(y)$ minimizes $u(x, y)$ over $x \in \mathbb{X}$. Set*

$$v(y) = u(x^\star(y), y).$$

Assume that $x^\star : Y \to \mathbb{X}$ is a continuous function. Then v is continuously differentiable and

$$\frac{d}{dy} v(y) = \frac{\partial u}{\partial y}\left(x^\star(y), y\right).$$

K.2 Proof of Proposition 1

With fixed $j \in \mathcal{J}_i$, we prove the proposition in two steps.

(i) f_j is continuous on $\Sigma \times \mathcal{S}_i$ and $\overline{\boldsymbol{\lambda}}_j$ is unique, continuous on $\mathring{\Sigma} \times \mathcal{S}_i$.

We apply Theorem 4 with the following substitutions:

- $\mathbb{X} = \Sigma \times \mathcal{S}_i$,
- $\mathbb{Y} = \text{cl}(\mathcal{C}_j^i)$,
- $\Phi(\boldsymbol{\omega}, \boldsymbol{\mu}) = \text{cl}(\mathcal{C}_j^i)$,
- $u(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{k=1}^K \omega_k d(\mu_k, \lambda_k)$.

As Φ is a constant correspondence and u is a continuous mapping, we immediate obtain that $\overline{\boldsymbol{\lambda}}_j$ is upper hemicontinuous and f_j is continuous on $\Sigma \times \mathcal{S}_i$ by Theorem 4.

Observe that $d(\mu, \cdot)$ is strictly convex on $\text{cl}(\mathcal{C}_j^i)$ when the distributions are from a one-parameter exponential family and recall $\min_k \omega_k > 0$ for all $\boldsymbol{\omega} \in \mathring{\Sigma}$, so is the weighted sum. We conclude that the uniqueness of the solution function, $\overline{\boldsymbol{\lambda}}_j$, stems from the strict convexity of the objective function. Thus, the continuity of $\overline{\boldsymbol{\lambda}}_j$ holds as the consequence of the uniqueness and its upper hemicontinuity.

(ii) f_j is differentiable on $\mathring{\Sigma} \times \mathcal{S}_i$ and $\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_{k=1}^K d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k)$ on $\mathring{\Sigma} \times \mathcal{S}_i$.

This is a consequence of Lemma 7 using the following substitutions:

- $\mathbb{X} = \text{cl}(\mathcal{C}_j^i)$,
- $Y = \mathring{\Sigma} \times \mathcal{S}_i$,
- $x^\star(\boldsymbol{\omega}, \boldsymbol{\mu}) = \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$,
- $u(\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}, \boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_{k=1}^K \omega_k d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k)$.

Under these substitutions, f_j is continuously differentiable as $x^\star$ is continuous from (i). The results follow directly.

K.3 The continuity of solution of (11) – Proof of Theorem 3

Before we prove the theorem, we state and prove some preliminary results. For the simplicity and clarity, we make a convention that $\boldsymbol{\mu} \in \mathcal{S}_i$ for some $i \in \mathcal{I}$. We also define the maps ψ_1 and ψ_2 as:

$$\psi_1 : (\boldsymbol{\omega}, r, \boldsymbol{\pi}, \boldsymbol{z}) \mapsto \min_{h \in H_{F_{\boldsymbol{\pi}}(\boldsymbol{\omega}, r)}} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle,$$
$$\psi_2 : (\boldsymbol{\omega}, r, \boldsymbol{\pi}) \mapsto \max_{\boldsymbol{z} \in \Sigma} \min_{h \in H_{F_{\boldsymbol{\pi}}(\boldsymbol{\omega}, r)}} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle.$$

Lemma 8. *$\psi_1(\boldsymbol{\omega}, r, \boldsymbol{\pi}, \boldsymbol{z})$ is continuous on $\mathring{\Sigma} \times (0,1) \times \mathcal{S}_i \times \Sigma$.*

Proof. We apply Theorem 4 with the following substitutions:

- $\mathbb{X} = \mathring{\Sigma} \times (0,1) \times \mathcal{S}_i \times \Sigma$,
- $\mathbb{Y} = \mathbb{R}^K$,
- $\Phi(\boldsymbol{\omega}, r, \boldsymbol{\pi}, \boldsymbol{z}) = H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r)$,
- $u(\boldsymbol{\omega}, r, \boldsymbol{\pi}, \boldsymbol{z}, h) = \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle$.

As u is obviously continuous, it only remains to prove that Φ is continuous.

Let $\{(\boldsymbol{\omega}_n, r_n, \boldsymbol{\pi}_n)\}_{n=1}^\infty$ be a sequence converging to $(\boldsymbol{\omega}, r, \boldsymbol{\pi}) \in \mathring{\Sigma} \times (0,1) \times \mathcal{S}_i$. Also, let $H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r) = \text{cov}\{\nabla_{\boldsymbol{\omega}} f_{j_m}(\boldsymbol{\omega}, \boldsymbol{\pi})\}_{m=1}^M$ for some $\{j_m\}_{m=1}^M \in \mathcal{J}_i$. Arbitrarily select $h \in H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r)$. Then there exists $\alpha_1, \ldots, \alpha_m \geq 0$ such that

$$\sum_{m=1}^M \alpha_{j_m} = 1 \text{ and } h = \sum_{m=1}^M \alpha_{j_m} \nabla_{\boldsymbol{\omega}} f_{j_m}(\boldsymbol{\omega}, \boldsymbol{\pi}).$$

As $(\boldsymbol{\omega}_n, r_n, \boldsymbol{\pi}_n) \xrightarrow{n \to \infty} (\boldsymbol{\omega}, r, \boldsymbol{\pi})$ and $\{f_j\}_{j \in \mathcal{J}_i}$ are continuous from Proposition 1, there is $N \in \mathbb{N}$ such that

$$\nabla_{\boldsymbol{\omega}} f_{j_m} \in H_{F_{\boldsymbol{\pi}}(\boldsymbol{\omega}_n, r_n)}, \text{ or equivalently } f_{j_m}(\boldsymbol{\omega}_n, \boldsymbol{\pi}_n) < F(\boldsymbol{\omega}_n, \boldsymbol{\pi}_n) + r_n,$$

for all $m = 1, \ldots, M,\ n \geq N$. In the following, we show lower and upper hemicontinuity for Φ separately.

Lower hemicontinuity:

For $n \geq N$, we select $h_n = \sum_{m=1}^M \alpha_{j_m} \nabla_{\boldsymbol{\omega}} f_{j_m}(\boldsymbol{\omega}_n, \boldsymbol{\pi}_n)$ then $h_n \xrightarrow{n \to \infty} h$ as $\nabla_{\boldsymbol{\omega}} f_{j_m}$'s are continuous by Proposition 1. This implies the lower hemicontinuity of Φ

Upper hemicontinuity:

Let $\mathcal{U}$ be an open set containing $H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r) = \text{cov}\{\nabla_{\boldsymbol{\omega}} f_{j_m}(\boldsymbol{\omega}, \boldsymbol{\pi})\}_{m=1}^M$. Because $\mathcal{U}$ is open, there exist $\epsilon > 0$ such that $H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r) + B(0, \epsilon) \subset \mathcal{U}$, where $+$ is a Minkowski addition and $B(0, \epsilon)$ is the K-dimensional ball with diameter ϵ. According to Proposition 1, there exists an integer $N' \geq N$ such that $\|\nabla_{\boldsymbol{\omega}} f_{j_m}(\boldsymbol{\omega}_n, \boldsymbol{\pi}_n) - \nabla_{\boldsymbol{\omega}} f_{j_m}(\boldsymbol{\omega}, \boldsymbol{\pi})\|_\infty < \epsilon$ for all $m = 1, \ldots, M,\ n \geq N'$. Thus, if $n \geq N'$,

$$H_{F_{\boldsymbol{\pi}_n}}(\boldsymbol{\omega}_n, r_n) = \text{cov}\{\nabla_{\boldsymbol{\omega}} f_{j_m}(\boldsymbol{\omega}_n, \boldsymbol{\pi}_n)\}_{m=1}^M\} \subset H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r) + B(0, \epsilon) \subset \mathcal{U},$$

and the upper hemicontinuity follows.

Summarizing, by continuity of Φ and u, we conclude that ψ_1 is also continuous by Theorem 4. □

Lemma 9. *$\psi_2(\boldsymbol{\omega}, r, \boldsymbol{\pi})$ is continuous on $\mathring{\Sigma} \times (0,1) \times \mathcal{S}_i$.*

Proof. We apply Theorem 4 with the following substitutions:

- $\mathbb{X} = \mathring{\Sigma} \times (0,1) \times \mathcal{S}_i$,
- $\mathbb{Y} = \Sigma$,
- $\Phi(\boldsymbol{\omega}, r, \boldsymbol{\pi}) = \Sigma$,
- $u(\boldsymbol{\omega}, r, \boldsymbol{\pi}, \boldsymbol{z}) = \psi_1(\boldsymbol{\omega}, r, \boldsymbol{\pi}, \boldsymbol{z})$.

From Lemma 8, ψ_1 is continuous. Notice that Φ is a constant map and hence continuous, so Theorem 4 directly implies the conclusion.

□

We are now ready to prove the theorem.

Proof of Theorem 3: We prove the inequality (45) using Lemma 9. The inequality (46) can be obtained analogously using Lemma 8. Let ϕ be a function defined on $\mathcal{S}_i$ as

$$\phi(\boldsymbol{\pi}) = \min \left\{ - |\psi_2(\boldsymbol{\omega}, r, \boldsymbol{\pi}) - \psi_2(\boldsymbol{\omega}, r, \boldsymbol{\mu})| : (\boldsymbol{\omega}, r) \in \mathring{\Sigma} \times (0,1) \right\}.$$

ϕ is a continuous function on $\mathring{\Sigma} \times \mathcal{S}_i$:

We apply Theorem 5 with the following substitutions:

- $\mathbb{X} = \mathcal{S}_i$,
- $\mathbb{Y} = \mathring{\Sigma} \times (0,1)$,
- $\Phi(\boldsymbol{\pi}) = \mathring{\Sigma} \times (0,1)$,
- $u(\boldsymbol{\lambda}, \boldsymbol{\omega}, r) = - |\psi_2(\boldsymbol{\omega}, r, \boldsymbol{\pi}) - \psi_2(\boldsymbol{\omega}, r, \boldsymbol{\mu})|$.

As $\mathbb{X} = \mathcal{S}_i$ is a metric space, it is compactly generated. Φ is continuous for it is a constant map. As for u, the upper semi-continuity follows from Lemma 9. It only remains to show that u is $\mathbb{K}$-inf compact. Let $C \subset \mathcal{S}_i$ be a compact set and let $y \in \mathbb{R}$. We show that $L_u(y, C \times \mathring{\Sigma} \times (0,1))$ is a compact by checking that it is bounded and closed. Boundness directly follows from the fact $\mathring{\Sigma} \times (0,1)$ is bounded and C is compact. As for closeness, u is a continuous function from Lemma 9, which also implies that $L_u(y, C \times \mathring{\Sigma} \times (0,1))$ is closed. Thus Theorem 5 implies that ϕ is a continuous function.

By definition of ϕ, $\phi(\boldsymbol{\mu}) = 0$. Since ϕ is continuous, there exists $\xi_{1,\epsilon}$ such that $\phi(\boldsymbol{\pi}) > -\epsilon/2$ for all $|\boldsymbol{\pi} - \boldsymbol{\mu}| < \xi_{1,\epsilon}$. In other words, the inequality (45) holds.

K.4 Proof of Lemma 6

Assume $i^\star(\boldsymbol{\mu}) = i$ for some $i \in \mathcal{I}$ for clarity. According to Assumption 1, $\mathcal{S}_i$ is open, and we know that $\boldsymbol{\pi} \in \mathcal{S}_i$ when $\boldsymbol{\pi}$ is close enough to $\boldsymbol{\mu}$. Hence, it remains to show that there exists a constant $\xi_{2,\epsilon} > 0$ such that $|F_{\boldsymbol{\pi}}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega})| < \frac{\epsilon}{2}$, for all $|\boldsymbol{\pi} - \boldsymbol{\mu}| < \xi_{2,\epsilon}$. We consider a function ϕ, which is defined below, and show its continuity.

$$\phi(\boldsymbol{\pi}) = \min_{\boldsymbol{\omega} \in \mathring{\Sigma}} - |F_{\boldsymbol{\pi}}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega})|, \ \forall \boldsymbol{\pi} \in \mathcal{S}_i.$$

ϕ is a continuous function on $\mathcal{S}_i$:

We apply Theorem 5 with the following substitutions:

- $\mathbb{X} = \mathcal{S}_i$,
- $\mathbb{Y} = \mathring{\Sigma}$,
- $\Phi(\boldsymbol{\lambda}) = \mathring{\Sigma}$,
- $u(\boldsymbol{\lambda}) = - |F_{\boldsymbol{\pi}}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega})|$.

As $\mathbb{X} = \mathcal{S}_i$ is a metric space, it is compactly generated. Φ is continuous for it is a constant map. As for u, the upper semi-continuity is followed by Proposition 1. It only remains to show that u is $\mathbb{K}$-inf compact. Let $C \subset \mathcal{S}_i$ be a compact set and let $y \in \mathbb{R}$. We show that $L_u(y, C \times \mathring{\Sigma})$ is a compact by checking that it is bounded and closed. Boundness directly follows from the fact $\mathring{\Sigma}$ is bounded and C is compact. As for closeness, u is a continuous function from Proposition 1, which also implies that $L_u(y, C \times \mathring{\Sigma})$ is closed. Theorem 5 hence implies that ϕ is a continuous function.

By the definition of ϕ, $\phi(\boldsymbol{\mu}) = 0$. Since ϕ is continuous, there exists $\xi_{2,\epsilon}$ such that $\phi(\boldsymbol{\lambda}) > -\epsilon/2$, or equivalently $|F_{\boldsymbol{\pi}}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega})| < \frac{\epsilon}{2}$, for all $|\boldsymbol{\pi} - \boldsymbol{\mu}| < \xi_{2,\epsilon}$. This completes the proof.

L Convergence of the Frank-Wolfe Algorithm

In this appendix, we study the performance of our variant of the FW algorithm. We assume that the real parameter $\boldsymbol{\mu}$ is used in the updates rather than its estimate.

Notations. In the following, for brevity, we drop the subscript $\boldsymbol{\mu}$. For instance, $F_{\boldsymbol{\mu}}$ is replaced by F; $\mathcal{J}_{i\star(\boldsymbol{\mu})}$ and $i^\star(\boldsymbol{\mu})$ become $\mathcal{J}$ and $i^\star$. We also use ∇f_j instead of $\nabla_{\boldsymbol{\omega}} f_j$ (as we will not differentiate f_j w.r.t. another argument).

L.1 Smoothness of the objective function

We state below the main properties of the objective function F. These properties will be instrumental in our convergence analysis.

L.1.1 F is Lipschitz

Proposition 3. *F is a L-Lipschitz function on Σ with respect to the infinity norm.*

Proof. Recall Assumption 2 and apply of mean value theorem. We get that f_j's are L-Lipschitz on $\mathring{\Sigma}$. As f_j's are continuous functions on Σ (see K.2 (i)), we can further extend the Lipschitzness from $\mathring{\Sigma}$ to Σ. Next we show that F is L-Lipstchitz. For any $\boldsymbol{x}, \boldsymbol{y} \in \Sigma$, we have that

$$\begin{aligned} F(\boldsymbol{x}) &= \min_{j\in\mathcal{J}} f_j(\boldsymbol{x}) \geq \min_{j\in\mathcal{J}} \left(f_j(\boldsymbol{y}) - L\left\|x-y\right\|_\infty\right) \\ &\geq \min_{j\in\mathcal{J}} f_j(\boldsymbol{y}) - L\left\|x-y\right\|_\infty = F(\boldsymbol{y}) - L\left\|x-y\right\|_\infty . \end{aligned}$$

This concludes the proof. □

L.1.2 Curvature of F

The definition (5) of the curvature and Assumption 2 allow us to bound the curvature of f_j inside Σ_γ. The following Proposition states that inside Σ_γ, the first order approximation of f_j remains controlled.

Proposition 4. *Let $\gamma \in (0, \frac{1}{K})$, $\boldsymbol{x} \in \Sigma_\gamma$ and $\boldsymbol{z} \in \Sigma$. Under Assumption 2, we have*

$$f_j(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f_j(\boldsymbol{x})\rangle - f_j(\boldsymbol{y}) \leq \frac{8D\alpha^2}{\gamma},$$

where $j \in \mathcal{J}$ and $\boldsymbol{y} = \boldsymbol{x} + \alpha(\boldsymbol{z}-\boldsymbol{x})$ for some $\alpha \in (0, \frac{1}{2}]$.

Proof. Let us drop the subscript j in f_j for clarity. Consider

$$\boldsymbol{u} = \frac{1}{2}(\boldsymbol{x}+\boldsymbol{z}). \tag{47}$$

As $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}$ are on the same line, we can re-write $\boldsymbol{y}$ as:

$$\boldsymbol{y} = \boldsymbol{x} + \alpha(\boldsymbol{z}-\boldsymbol{x}) = 2\alpha\boldsymbol{u} + (1-2\alpha)\boldsymbol{x}.$$

The definition of $\boldsymbol{u}$ implies that $\boldsymbol{x}, \boldsymbol{u} \in \Sigma_{\frac{\gamma}{2}}$. Let $\alpha' = 2\alpha \in [0,1]$, Assumption 2 (ii) and the definition (5) lead to

$$f(\boldsymbol{x}) + \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle - f(\boldsymbol{y}) \leq \frac{2D\alpha'^2}{\gamma} = \frac{8D\alpha^2}{\gamma}.$$

□

Next, consider $F = \min_{j\in\mathcal{J}} f_j$ instead of f_j.

Corollary 1. *Let $\gamma \in (0, 1/K), r \in (0,1)$, $\boldsymbol{x} \in \Sigma_\gamma$ and $\boldsymbol{z} \in \Sigma$. If α is a positive number s.t. $\alpha < \min\{\frac{1}{2}, \frac{r}{L}\}$, then*

$$F(\boldsymbol{y}) \geq F(\boldsymbol{x}) + \alpha \min_{h \in H_F(\boldsymbol{x}, r)} \langle \boldsymbol{z} - \boldsymbol{x}, h \rangle - \frac{8D\alpha^2}{\gamma},$$

where $\boldsymbol{y} = (1-\alpha)\boldsymbol{x} + \alpha \boldsymbol{z}$.

Proof. If $F(\boldsymbol{x}) = f_j(\boldsymbol{x}) = f_j(\boldsymbol{y}) = F(\boldsymbol{y})$ for some $j \in \mathcal{J}$, the result directly follows from Proposition 4 as $\nabla f_j(\boldsymbol{x}) \in H_F(\boldsymbol{x}, r)$.

Otherwise, assume that we have two distinct $j_1, j_2 \in \mathcal{J}$ such that $F(\boldsymbol{x}) = f_{j_1}(\boldsymbol{x}) < f_{j_2}(\boldsymbol{x})$ and $F(\boldsymbol{y}) = f_{j_2}(\boldsymbol{y}) < f_{j_1}(\boldsymbol{y})$. As shown in the proof of Proposition 3, f_j is L-Lipschitz and $\|\boldsymbol{x} - \boldsymbol{y}\|_\infty = \alpha \|\boldsymbol{z}\|_\infty < \frac{r}{L}$. We deduce that $f_{j_2}(\boldsymbol{x}) < F(\boldsymbol{x}) + r$, which is equivalent to $\nabla f_{j_2}(\boldsymbol{x}) \in H_F(\boldsymbol{x}, r)$. Consequently, choosing $h = \nabla f_{j_2}(\boldsymbol{x})$ yields that

$$F(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, h \rangle - F(\boldsymbol{y}) \leq f_{j_2}(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, h \rangle - f_{j_2}(\boldsymbol{y}) \leq \frac{8D\alpha^2}{\gamma},$$

where the last inequality is from Proposition 4. The corollary is proved. □

L.2 Properties of $H_\Phi(\boldsymbol{x}, r)$

Here we consider some functions, $\phi_1, \ldots, \phi_n$ on Σ and define $\Phi(\boldsymbol{x}) = \min_i \phi_i(\boldsymbol{x}), \forall \boldsymbol{x} \in \Sigma$. It is clear that $\partial\Phi(\boldsymbol{x}) \subset H_\Phi(\boldsymbol{x}, r)$. The following result relates $H_\Phi(\boldsymbol{x}, r)$ to the r-subdifferential of Φ. Recall that for $r \in (0,1)$, the r-subdifferential of Φ is defined as $\partial_r\Phi(\boldsymbol{x}) = \{h \in \mathbb{R}^K : \Phi(\boldsymbol{y}) < \Phi(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, h \rangle + r \text{ for all } \boldsymbol{y} \in \Sigma\}$.

Lemma 10. *If $\Phi = \min_{i \in [n]} \phi_i$ where $\{\phi_j\}_{j=1}^n$ are concave differentiable functions defined on $\mathring{\Sigma}$, then*

$$H_\Phi(\boldsymbol{x}, r) \subset \partial_r\Phi(\boldsymbol{x}), \forall x \in \mathring{\Sigma},\ r > 0. \tag{48}$$

Proof. Let $\boldsymbol{x} \in \mathring{\Sigma}, r > 0$ be fixed and $\mathcal{A} = \{i \in [n] : \phi_i(\boldsymbol{x}) < \Phi(\boldsymbol{x}) + r\}$. Let $h \in H_\Phi(\boldsymbol{x}, r)$. It can be written as $h = \sum_{i \in \mathcal{A}} \alpha_i \nabla\phi_i(\boldsymbol{x}) \in H_\Phi(\boldsymbol{x}, r)$, where $\alpha_i \geq 0, \forall i \in \mathcal{A}$ and $\sum_{i \in \mathcal{A}} \alpha_i = 1$. Observe that for any $\boldsymbol{y} \in \mathring{\Sigma}$, $\Phi(\boldsymbol{y}) \leq \phi_i(\boldsymbol{y}),\ \forall i \in \mathcal{A}$. Thus,

$$\Phi(\boldsymbol{y}) - \Phi(\boldsymbol{x}) - \langle \boldsymbol{y} - \boldsymbol{x}, h \rangle < \sum_{i \in \mathcal{A}} \alpha_i \left[\phi_i(\boldsymbol{y}) - \phi_i(\boldsymbol{x}) + r - \langle \boldsymbol{y} - \boldsymbol{x}, \nabla\phi_i(\boldsymbol{x}) \rangle\right] \leq \sum_{i \in \mathcal{A}} \alpha_i r = r,$$

where the last inequality stems for the concavity of the ϕ_i's. The above inequality, valid for any $h \in H_\Phi(\boldsymbol{x}, r)$, implies that $H_\Phi(\boldsymbol{x}, r) \subset \partial_r\Phi(\boldsymbol{x})$. □

The next property is sometimes called primal-dual gap, see [24]. Interestingly, this property together with Lemma 10 tell us that the maxmin value computed at each iteration (11) can serve as an estimate of the gap.

Lemma 11 ([41])**.** *Let $\Phi = \min_{i \in [n]} \phi_i$ where $\{\phi_j\}_{j=1}^n$ are concave differentiable functions defined on Σ. Then, for any $\boldsymbol{x} \in \Sigma$,*

$$\max_{\boldsymbol{z} \in \Sigma} \min_{h \in \partial_r\Phi(\boldsymbol{x})} \langle \boldsymbol{z} - \boldsymbol{x}, h \rangle \geq \max_{\boldsymbol{y} \in \Sigma} \Phi(\boldsymbol{y}) - \Phi(\boldsymbol{x}) - r.$$

L.3 The convergence of FWS under $\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T)$

Recall the definition of our "good" event (see Appendix I): $\mathcal{E}_{1,\epsilon}(T) = \left(\bigcap_{t=\underline{h}(T)}^T \mathcal{E}_{1,\epsilon}^{(t)}\right)$ and $\mathcal{E}_{2,\epsilon}(T) = \left(\bigcap_{t=\underline{h}(T)}^T \mathcal{E}_{2,\epsilon}^{(t)}\right)$ where

$$\mathcal{E}_{1,\epsilon}^{(t)} = \left\{ \max_{\boldsymbol{z} \in \Sigma} \min_{h \in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}(t-1), r_t)} \langle \boldsymbol{z} - \boldsymbol{x}(t-1), h \rangle - \epsilon < \min_{h \in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{x}(t-1), r_t)} \langle \boldsymbol{z}(t) - \boldsymbol{x}(t-1), h \rangle \right\},$$

$$\mathcal{E}_{2,\epsilon}^{(t)} = \left\{ \hat{\boldsymbol{\mu}}(t) \in \mathcal{S}_{i^\star(\boldsymbol{\mu})} \text{ and } \left| F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) \right| < \epsilon,\ \forall \boldsymbol{\omega} \in \mathring{\Sigma} \right\}.$$

In what follows, we use the notation: $\Delta_t = F(\boldsymbol{\omega}^\star) - F(\boldsymbol{x}(t))$. In our convergence analysis, we first show that Δ_t is a decreasing sequence under $\mathcal{E}_{1,\epsilon}(T)$. Then, we prove that Δ_t becomes small when $t \geq \overline{h}(T)$.

Theorem 6. *Let $t \in \mathbb{N}$ satisfying that $\lfloor \sqrt{\frac{t}{K}} \rfloor \notin \mathbb{N}$ and $t \geq 4K$. Under the event $\mathcal{E}_{1,\epsilon}^{(t)} \cap \mathcal{E}_{2,\epsilon}^{(t)}$ and iteration (11) with parameter such that $L < r_t t$, we have*

$$\Delta_t \leq \frac{t-1}{t}\Delta_{t-1} + \frac{r_t + \epsilon}{t} + \frac{16D\sqrt{K}}{t^{\frac{3}{2}}}. \tag{49}$$

Proof. To simplify our presentation, we denote $\boldsymbol{y} = \boldsymbol{x}(t)$, $\boldsymbol{x} = \boldsymbol{x}(t-1)$ and $\boldsymbol{z} = \boldsymbol{z}(t)$. Also, let α be the step size $\frac{1}{t}$ and $r = r_t$.

Lemma 13 implies that $\boldsymbol{x} \in \Sigma_{\frac{1}{2\sqrt{tK}}}$ (when $t \geq 4K$), and hence, Corollary 1 with $\gamma = \frac{1}{2\sqrt{tK}}$ yields:

$$\begin{aligned} F(\boldsymbol{y}) &\geq F(\boldsymbol{x}) + \alpha \min_{h \in H_F(\boldsymbol{x}, r)} \langle \boldsymbol{z} - \boldsymbol{x}, h \rangle - 16D\alpha^2\sqrt{tK} \\ &\geq F(\boldsymbol{x}) + \alpha \left(\max_{\boldsymbol{\omega} \in \Sigma} \min_{h \in H_F(\boldsymbol{x}, r)} \langle \boldsymbol{\omega} - \boldsymbol{x}, h \rangle - \epsilon \right) - 16D\alpha^2\sqrt{tK}, \end{aligned} \tag{50}$$

where the second inequality directly follows from the selection of $\boldsymbol{z}$ and the event $\mathcal{E}_{1,\epsilon}^{(t)}$. As $H_F(\boldsymbol{x}, r) \subset \partial_r F(\boldsymbol{x})$, shown in Lemma 10, Lemma 11 implies that the second term in the right-hand side of inequality (50) can be lower bounded as:

$$\begin{aligned} \max_{\boldsymbol{\omega} \in \Sigma} \min_{h \in H_F(\boldsymbol{x}, r)} \langle \boldsymbol{\omega} - \boldsymbol{x}, h \rangle - \epsilon &\geq \max_{\boldsymbol{\omega} \in \Sigma} \min_{h \in \partial_r F(\boldsymbol{x})} \langle \boldsymbol{\omega} - \boldsymbol{x}, h \rangle - \epsilon \\ &\geq \Delta_{t-1} - r - \epsilon. \end{aligned} \tag{51}$$

Substituting the inequalities (51) into (50), we obtain that

$$F(\boldsymbol{y}) \geq F(\boldsymbol{x}) + \alpha \left(\Delta_{t-1} - r - \epsilon \right) - 16D\alpha^2\sqrt{tK}.$$

Subtracting $F(\boldsymbol{\omega}^\star)$ on both sides of the above inequality, we get that

$$\Delta_t \leq (1-\alpha)\Delta_{t-1} + \alpha(r + \epsilon) + 16D\alpha^2\sqrt{tK}. \tag{52}$$

The result follows from the inequality (52) and $\alpha = \frac{1}{t}$.

□

The following theorem states the convergence of `FWS`. This convergence is obtained by repeatedly applying Theorem 6.

Theorem 7. *Let $\{r_t\}_{t \geq 1}$ be a sequence of positive numbers satisfying (i) $\lim_{t \to \infty} \frac{1}{t} \sum_{s=1}^{t} r_s = 0$, and (ii) $\lim_{t \to \infty} t r_t = \infty$. Suppose $T \geq \max\{\left(\frac{32D+3L}{\epsilon}\right)^{11}, T_{\epsilon,L}^{\frac{11}{8}}, (4K+1)^{\frac{11}{8}}\}$, where $T_{\epsilon,L}$ is defined in (39). Then, under event $\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T)$, applying `FWS` algorithm with $\{r_t\}_{t \geq 1}$, we have:*

$$F(\boldsymbol{\omega}^\star) - F(\boldsymbol{x}(t)) \leq 4\epsilon, \ \forall t = \overline{h}(T), \overline{h}(T)+1, \ldots, T.$$

Proof. We start the proof by dividing the time horizon into several blocks, where each block consists of K successive rounds. We introduce m as the index of the block, with a slight abuse of notation, we denote $\tilde{\Delta}_m = \Delta_{mK}$ as the gap at the end of the m-th block.

We provide recursive properties of $\tilde{\Delta}_m$ in two cases (a) m is a square number, (b) m is not a square number for $mK \geq \underline{h}(T)$.

Step 1. Recursive properties of $\tilde{\Delta}_m$ under (a) and (b).

For (a), $\boldsymbol{x}(mK) = \frac{1}{m}(\frac{1}{K}, \ldots, \frac{1}{K}) + \frac{m-1}{m}\boldsymbol{x}(mK - K)$. Proposition 3 directly yields that

$$\tilde{\Delta}_m \leq \tilde{\Delta}_{m-1} + \frac{L}{m}.$$

Since $\tilde{\Delta}_{m-1}$ is bounded by L, the above equation implies that

$$m\tilde{\Delta}_m \leq (m-1)\tilde{\Delta}_{m-1} + 2L. \tag{53}$$

For (b), recall that $T \geq T_{\epsilon,L}^{\frac{11}{8}}$, (39) and (40), for any $t \geq \underline{h}(T)$, we have that $t \geq \max\{T_{\epsilon,L}, 4K\}$. Thus, letting $Z = 16D\sqrt{K}$, Theorem 6 can be applied to derive a series of inequalities for $t = (m-1)K+1, \ldots, mK$:

$$\begin{aligned}
[(m-1)K+1]\Delta_{(m-1)K+1} &\leq (m-1)K\Delta_{(m-1)K} + \epsilon + \frac{Z}{[(m-1)K+1]^{\frac{1}{2}}} + r_{(m-1)K+1},\\
[(m-1)K+2]\Delta_{(m-1)K+2} &\leq [(m-1)K+1]\Delta_{(m-1)K+2} + \epsilon\\
&\quad + \frac{Z}{[(m-1)K+2]^{\frac{1}{2}}} + r_{(m-1)K+2},\\
&\vdots\\
(mK)\Delta_{mK} &\leq (mK-1)\Delta_{mK-1} + \epsilon + \frac{Z}{(mK)^{\frac{1}{2}}} + r_{mK}.
\end{aligned}$$

Summing over them and dividing by K on both sides, we obtain that:

$$m\tilde{\Delta}_m \leq (m-1)\tilde{\Delta}_{(m-1)} + \epsilon + Z(m) + r(m), \tag{54}$$

where $Z(m) = \sum_{t=(m-1)K+1}^{mK} \frac{Z}{K\sqrt{t}}$ and $r(m) = \sum_{t=(m-1)K+1}^{mK} \frac{r_t}{K}$.

Our next step consists in studying the recursion between two successive square numbers. We introduce:

$$\underline{p} = \sqrt{\frac{\underline{h}(T)}{K}},\ \overline{p} = \sqrt{\frac{\overline{h}(T)}{K}},\ \text{and } \mathcal{P}(T) = \{p \in \mathbb{N} : \underline{h}(T) \leq Kp^2 \leq T\}. \tag{55}$$

Step 2. For any $q \geq \overline{p}$, $\tilde{\Delta}_q \leq 3\epsilon$.

We first fix some $p \in \mathcal{P}(T)$, summing the inequalities (54) over $m = p^2+1, \ldots, (p+1)^2-1$ and inequality (53) with $m = (p+1)^2$ gives:

$$(p+1)^2\tilde{\Delta}_{(p+1)^2} \leq p^2\tilde{\Delta}_{p^2} + 2p\epsilon + 2L + \sum_{m=p^2+1}^{(p+1)^2} Z(m) + r(m). \tag{56}$$

Then for any $q \geq \overline{p}$, we sum the inequalities (56) from $p = \underline{p}$ to $p = q-1$ and get that

$$\begin{aligned}
q^2\tilde{\Delta}_{q^2} &\leq \underline{p}^2\tilde{\Delta}_{\underline{p}^2} + 2\epsilon\sum_{p=\underline{p}}^{q-1} p + 2L(q-p-2) + \sum_{m=\underline{p}^2}^{q^2} Z(m) + r(m)\\
&\leq \underline{p}^2 L + 2\epsilon\int_0^q t dt + 2Lq + \int_0^{q^2K} \frac{Z}{K\sqrt{t}} dt + \sum_{t=1}^{q^2K} \frac{r_t}{K}\\
&\leq \frac{\underline{h}(T)L}{K} + \epsilon q^2 + 2Lq + \frac{2Zq}{\sqrt{K}} + \sum_{t=1}^{q^2K} \frac{r_t}{K}.
\end{aligned} \tag{57}$$

Recall from (40) and (55) that $\overline{h}(T) \geq T^{\frac{2}{11}}\underline{h}(T)$ and $q^2K \geq \overline{p}^2K = \overline{h}(T) \geq \max\{T^{\frac{2}{11}}K, T_{\epsilon,L}\}$. Divide by q^2 both sides of the inequality (57). We obtain:

$$\begin{aligned}
\tilde{\Delta}_{q^2} &\leq \frac{L}{T^{\frac{2}{11}}} + \epsilon + \frac{2(Z/\sqrt{K}+L)}{T^{\frac{1}{11}}} + \frac{1}{q^2K}\sum_{t=1}^{q^2K} r_t\\
&\leq \epsilon + \frac{2Z/\sqrt{K}+3L}{T^{\frac{1}{11}}} + \frac{1}{q^2K}\sum_{t=1}^{q^2K} r_t \leq 3\epsilon,
\end{aligned}$$

where the last inequality stems from $T \geq \left(\frac{32D+3L}{\epsilon}\right)^{11} = \left(\frac{2Z/\sqrt{K}+3L}{\epsilon}\right)^{11}$ and the definition of $T_{\epsilon,L}$ (see (39)).

Step 3. For any $t \geq \overline{h}(T)$, we have $\Delta_t \leq 4\epsilon$.

Now suppose $t \in \{Kq^2+1, \ldots, K(q+1)^2-1\}$ for some $q \geq \overline{p}$. Recall that

$$\boldsymbol{x}(t) = \frac{Kq^2}{t}\boldsymbol{x}(Kq^2) + \frac{(t-Kq^2)}{t}\boldsymbol{u}, \text{ for some } \boldsymbol{u} \in \Sigma,$$

which yields that

$$\left\|\boldsymbol{x}(Kq^2) - \boldsymbol{u}\right\|_\infty \leq \frac{t-Kq^2}{t}\left\|\boldsymbol{u}\right\|_\infty \leq \frac{K(2q+1)}{Kq^2} \leq \frac{3}{q} \leq \frac{\epsilon}{L},$$

as (40) implies that $q \geq \frac{T^{\frac{2}{11}}\underline{h}(T)}{K} \geq \frac{3L}{\epsilon}$. Consequently, Proposition 3 yields

$$\left|F(\boldsymbol{x}(t)) - F(\boldsymbol{x}(Kq^2))\right| \leq \epsilon.$$

Combining this with the inequality from **Step 2**, we conclude that $F(\boldsymbol{\omega}^\star) - F(\boldsymbol{x}(t)) \leq 4\epsilon$ for all $t \geq \overline{h}(T)$.

□

A consequence of Lemma 12 about the tracking rule (presented in the next appendix) and Theorem 7 is Lemma 3.

Proof of Lemma 3. Lemma 12 implies that $\left\|\boldsymbol{\omega}(t) - \boldsymbol{x}(t)\right\|_\infty \leq \frac{K-1}{t}$ and then by Proposition 3,

$$F(\boldsymbol{\omega}(t)) \geq F(\boldsymbol{x}(t)) - \frac{(K-1)L}{t} \geq F(\boldsymbol{x}(t)) - \epsilon,$$

where the last inequality is due to the fact that $t \geq \overline{h}(T) \geq T^{\frac{2}{11}}\underline{h}(T) \geq \frac{KL}{\epsilon}$ (see definition of $\underline{h}(T)$ and $\overline{h}(T)$ (40)). Combining Theorem 7 and the above inequality leads to the desired result. □

M Tracking Rule

This section presents the analysis of the tracking rule in FWS and related results.

Lemma 12 (Lemma 7 in [12])**.** *Let $\{\boldsymbol{z}(s)\}_{s\in\mathbb{N}} \in \Sigma$ be a sequence of vectors such that $\boldsymbol{z}(1),\dots,\boldsymbol{z}(K)$ are $\boldsymbol{e}_1,\dots,\boldsymbol{e}_K$. We recursively define for $t \geq K$,*

$$\forall k \in [K], N_k(K) = 1,$$

$$\forall t \geq K+1,\ \ A_t \in \operatorname*{argmax}_{k'} \frac{\sum_{s=1}^{t} z_{k'}(s)}{N_{k'}(t-1)},\ \ \forall k \in [K],\ N_k(t) = \sum_{s=1}^{t} \mathbb{1}\{A_s = k\},$$

(where the tie-breaking rule in the argmax *is arbitrary). Then for all $t \geq K$, all $k \in [K]$,*

$$\sum_{s=1}^{t} z_k(s) - (K-1) \leq N_k(t) \leq \sum_{s=1}^{t} z_k(s) + 1.$$

Lemma 13. *At any time $t \geq 4K$, under* FWS*, we have $\boldsymbol{x}(t) \in \Sigma_{\sqrt{\frac{1}{Kt}} - \frac{1}{t}} \subset \Sigma_{\frac{1}{2\sqrt{tK}}}$.*

Proof. This lemma directly follows from the forced exploration procedure of FWS when $\lfloor t/K \rfloor$ is a square number, $\boldsymbol{x}(t)$ move to the center of Σ for K successive rounds. Hence, for all $k = 1,\dots,K$

$$\begin{aligned} t x_k(t) = \sum_{s=1}^{t} z_k(s) &\geq \frac{1}{K} \sum_{s=1}^{t} \mathbb{1}\{\boldsymbol{z}(s) = (1/K,\dots,1/K)\} \\ &= \sqrt{\lfloor \frac{t}{K} \rfloor} \geq \sqrt{\frac{t}{K}} - 1 \geq \frac{1}{2}\sqrt{\frac{t}{K}}. \end{aligned}$$

Dividing t on the both sides, we get the result. □

N Non-asymptotic Sample Complexity

Looking back at our asymptotic sample complexity analysis, we note that the reason why we could not derive results for the mild confidence regime (non-asymptotic) is that we cannot quantify the cost paid for events $\mathbb{P}\left[(\mathcal{E}_{1,\epsilon} \cap \mathcal{E}_{2,\epsilon})^c\right]$ (see (41)-(42)). Looking in more details, the probability of these events cannot be precisely controlled because our continuity arguments (see Lemma 6 and Theorem 3 in Appendix K) rely on maximal theorems, and the constants $\xi_{1,\epsilon}$ and $\xi_{2,\epsilon}$ involved there have an unknown dependence in ϵ.

To get non-asymptotic sample complexity upper bounds, we use mean value theorems instead, and obtain simple upper bounds of $\xi_{1,\epsilon}$ and $\xi_{2,\epsilon}$. This section is organized as follows: we present a stronger version of Lemma 6 and Theorem 3 in N.1 and N.2, respectively; the proof of Theorem 2 is then provided in N.3.

For convenience, we restate the additional assumption and our non-asymptotic sample complexity upper bound.

Assumption 3. *For any $\boldsymbol{\mu} \in \Lambda$, there exist constants $\kappa, E > 0$, s.t. if $\|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty \leq \kappa$, then $\boldsymbol{\pi} \in \mathcal{S}_{i^\star(\boldsymbol{\mu})}, \forall \boldsymbol{\omega} \in \mathring{\Sigma},\ j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}, \nabla_{\boldsymbol{\pi}} d(\pi_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\pi})}_k)$ is continuous and $\left\|\nabla_{\boldsymbol{\pi}} d(\pi_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\pi})}_k)\right\|_1 \leq E,\ \forall k = 1, \ldots, K$.*

Theorem 2. *Consider* FWS *algorithm with a sequence $\{r_t\}_{t \geq 1}$ as in Theorem 1. Under Assumptions 1, 2, and 3, the sample complexity τ of the algorithm satisfies: for any $\boldsymbol{\mu} \in \Lambda, \delta \in (0,1)$, and any $\epsilon < \min\{\kappa E/2, 1\}$, $\tilde{\epsilon} < 1$,*

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \leq \frac{1+\tilde{\epsilon}}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon}\left[\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)e}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon)}\right) + \log\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu})) - 6\epsilon)}\right)\right] + \Psi(K, D, E, L, c_1(\Lambda), \epsilon) + T_{\epsilon,L}^{\frac{5}{4}},$$

where $T_{\epsilon,L}$ is a constant such if $t \geq T_{\epsilon,L}$, then $\sum_{s=1}^t r_s < t\epsilon$ and $tr_t > L$. The constant Ψ is polynomial in $(D, E, L, c_1(\Lambda), 1/\epsilon)$ and exponential in K. The precise definition of Ψ is given at the end of this section.

N.1 Continuity of the primal problem

First, we present the analogue of Lemma 6 (Appendix K).

Lemma 14. *Under Assumptions 1 and 3, for any $\boldsymbol{\mu} \in \Lambda$ and $\epsilon \in (0, \kappa E)$, if $\|\boldsymbol{\mu} - \boldsymbol{\pi}\|_\infty \leq \frac{\epsilon}{E}$, then*

$$|F_{\boldsymbol{\pi}}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega})| < \epsilon, \forall \boldsymbol{\omega} \in \mathring{\Sigma}.$$

Proof. Fix $j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, and let $\boldsymbol{\omega} \in \mathring{\Sigma}$. Define the function $g : [0,1] \to \mathbb{R}$ as

$$g(t) = f_j(\boldsymbol{\omega}, t\boldsymbol{\mu} + (1-t)\boldsymbol{\pi}).$$

Since $\|\boldsymbol{\mu} - \boldsymbol{\pi}\|_\infty \leq \frac{\epsilon}{E} \leq \kappa$, Assumption 3 says that $t\boldsymbol{\mu} + (1-t)\boldsymbol{\pi} \in \mathcal{S}_{i^\star(\boldsymbol{\mu})}$, which implies g is well-defined. Based on the mean value theorem, there is $t \in (0,1)$ s.t.

$$\langle g'(t), \boldsymbol{\mu} - \boldsymbol{\pi} \rangle = f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) - f_j(\boldsymbol{\omega}, \boldsymbol{\pi}). \tag{58}$$

For clarity, we denote $t\boldsymbol{\mu} + (1-t)\boldsymbol{\pi} = \tilde{\boldsymbol{\pi}}$ and its k-th component as $\tilde{\pi}_k$. Also, $\frac{\partial f_j}{\partial \boldsymbol{\mu}}(\boldsymbol{\omega}, \boldsymbol{\mu})$ is the partial derivative of $f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ w.r.t. $\boldsymbol{\mu}$. (58) yields that

$$\begin{aligned}
|f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) - f_j(\boldsymbol{\omega}, \boldsymbol{\pi})| = |g'(t)| &= \left|\langle \frac{\partial f_j}{\partial \boldsymbol{\mu}}(\boldsymbol{\omega}, \tilde{\boldsymbol{\pi}}), \boldsymbol{\mu} - \boldsymbol{\pi} \rangle\right| \\
&= \sum_{k=1}^K \omega_k \langle \nabla_{\tilde{\boldsymbol{\pi}}} d(\tilde{\pi}_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \tilde{\boldsymbol{\pi}})}_k), \boldsymbol{\mu} - \boldsymbol{\pi} \rangle \\
&\leq \sum_{k=1}^K \omega_k \left\|\nabla_{\tilde{\boldsymbol{\pi}}} d(\tilde{\pi}_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \tilde{\boldsymbol{\pi}})}_k)\right\|_1 \|\boldsymbol{\mu} - \boldsymbol{\pi}\|_\infty \leq \epsilon,
\end{aligned} \tag{59}$$

where the last inequality is the result of Assumption 3, $\|\boldsymbol{\mu}-\boldsymbol{\pi}\|_\infty < \frac{\epsilon}{E}$ and $\boldsymbol{\omega} \in \mathring{\Sigma}$.

As for the objective function, we have

$$F_{\boldsymbol{\pi}}(\boldsymbol{\omega}) - F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \min_j f_j(\boldsymbol{\omega}, \boldsymbol{\pi}) - \min_j f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) \leq \min_j f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) + \epsilon - \min_j f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \epsilon,$$

where the inequality holds in view of (59). The other inequality follows similarly, which completes the proof. □

N.2 Envelope theorem at a saddle point

As Lemma 14 in Appendix K, we wish to apply the envelop theorem for the perturbation analysis of the equation (11). For clarity, we redefine the notations used in Appendix K.

Let $\mathbb{X}$ and $\mathbb{Y}$ be Hausdorff topological spaces and $u : \mathbb{X} \times \mathbb{Y} \times [0,1] \to \mathbb{R}$ be a function. We introduce $\mathbb{K}(\mathbb{X})$ (resp. $\mathbb{K}(\mathbb{Y})$) as the collection of all compact sets in $\mathbb{X}$ (resp. $\mathbb{Y}$). Assume that $X : [0,1] \rightrightarrows \mathbb{K}(\mathbb{X})$ and $Y : [0,1] \rightrightarrows \mathbb{K}(\mathbb{Y})$ are nonempty correspondences. We say that $(x^\star(t), y^\star(t))$ is a *saddle point* of u at some fixed $t \in [0,1]$ if it satisfies that

$$\max_{x \in X(t)} u(x, y^\star(t), t) \leq u(x^\star(t), y^\star(t), t) \leq \min_{y \in Y(t)} u(x^\star(t), y, t). \tag{60}$$

It is well-known that the existence of the above saddle point $(x^\star(t), y^\star(t))$ implies that (see e.g. [15] Chapter 6. Proposition 1.2)

$$\max_{x \in X(t)} \min_{y \in Y(t)} u(x, y, t) = \min_{y \in Y(t)} \max_{x \in X(t)} u(x, y, t) = u(x^\star(t), y^\star(t), t). \tag{61}$$

Next, introduce the value function $v(t)$ for $t \in [0,1]$ as $v(t) = \max_{x \in X(t)} \min_{y \in Y(t)} u(x, y, t)$. The existence of the saddle point $(x^\star(t), y^\star(t))$ further implies that there exist subsets $X^\star(t) \subseteq X(t)$, $Y^\star(t) \subset Y(t)$ such that ([15] Chapter 6. Proposition 1.4)

$$u(x, y, t) = u(x^\star(t), y^\star(t), t), \ \forall (x, y) \in X^\star(t) \times Y^\star(t).$$

In this case, we can derive an envelope theorem at the saddle points [39].

Theorem 8 (Theorem 5 in [39])**.** *Let u and its derivative with respect to t, u_t, be continuous functions on $X \times Y \times [0,1]$. Let X, Y be continuous correspondences such that the existence of saddle point is guaranteed for all $t \in [0,1]$. Then $v(t)$ is differentiable in $(0,1)$, and*

$$v'(t) = \max_{x \in X(t)} \min_{y \in Y(t)} u_t(x, y, t) = \min_{y \in Y(t)} \max_{x \in X(t)} u_t(x, y, t), \ \forall t \in (0,1).$$

Using Theorem 8, we are able to develop the stronger version of Theorem 3.

Theorem 9. *Let $\boldsymbol{\mu} \in \Lambda$, $\epsilon \in (0, \kappa E)$. For any $r \in (0,1)$, and $\boldsymbol{\omega} \in \mathring{\Sigma}$, if another parameter $\boldsymbol{\pi} \in \Lambda$ satisfies that $\|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty \leq \frac{\epsilon}{2E}$, then under Assumptions 1 and 3, we get:*

$$\left| \max_{z \in \Sigma} \min_{h \in H_{F_{\boldsymbol{\pi}}}} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle - \max_{z \in \Sigma} \min_{h \in H_{F_{\boldsymbol{\mu}}}} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle \right| \leq \frac{\epsilon}{2}, \ \forall (\boldsymbol{\omega}, r) \in \mathring{\Sigma} \times (0,1), \tag{62}$$

and

$$\left| \min_{h \in H_{F_{\boldsymbol{\pi}}}} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle - \min_{h \in H_{F_{\boldsymbol{\mu}}}} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle \right| \leq \frac{\epsilon}{2}, \ \forall (\boldsymbol{z}, \boldsymbol{\omega}, r) \in \Sigma \times \mathring{\Sigma} \times (0,1). \tag{63}$$

Proof. We prove (62). (63) will hold for similar reasons as discussed later. Fix $\boldsymbol{\mu} \in \Lambda$, $r \in (0,1)$ and $\boldsymbol{\omega} \in \mathring{\Sigma}$.

(i) Verifying the conditions of Theorem 8.

We apply Theorem 8 with $\mathbb{X} = \Sigma - \boldsymbol{\omega} = \{\boldsymbol{x} \in \mathbb{R}^K : \exists \boldsymbol{z} \in \Sigma \text{ s.t } \boldsymbol{x} = \boldsymbol{z} - \boldsymbol{\omega}\}$ and $\mathbb{Y} = \Sigma(\mathcal{J}_{i^\star(\boldsymbol{\mu})})$, which denotes the $\left|\mathcal{J}_{i^\star(\boldsymbol{\mu})}\right| - 1$-simplex. As $\|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty \leq \frac{\epsilon}{2E} < \kappa$, Assumption 3 holds. Thus,

$\boldsymbol{\pi} \in \mathcal{S}_{i^\star(\boldsymbol{\mu})}$, we then define u and its derivative with respect to t as:

$$u(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_k \sum_j x_k y_j d(\tilde{\pi}(t)_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \tilde{\boldsymbol{\pi}}(t))}_k),$$
$$u_t(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_k \sum_j x_k y_j \langle \nabla_{\tilde{\boldsymbol{\pi}}(t)} d(\tilde{\pi}(t)_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \tilde{\boldsymbol{\pi}}(t))}_k), \boldsymbol{\mu} - \boldsymbol{\pi} \rangle.$$

where $\tilde{\boldsymbol{\pi}}(t) = t\boldsymbol{\mu} + (1-t)\boldsymbol{\pi}$, for all $t \in [0,1]$. Observe that both u, u_t are continuous on $\mathbb{X} \times \mathbb{Y} \times [0,1]$. Further define the correspondences

$$\begin{cases} X(t) = \mathbb{X} = \Sigma - \boldsymbol{\omega}, \\ Y(t) = \left\{ \boldsymbol{y} \in \Sigma(\mathcal{J}_{i^\star(\boldsymbol{\mu})}) : y_j = 0 \text{ if } f_j(\boldsymbol{\omega}, \tilde{\boldsymbol{\pi}}(t)) \geq F_{\tilde{\boldsymbol{\pi}}(t)}(\boldsymbol{\omega}) + r \right\}. \end{cases}$$

$X(t)$ is a constant so it is continuous. As for the continuity of $Y(t)$, the argument is similar to that used to prove the hemicontinuity of $H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r)$ (see the proof of Lemma 8). Since $\max_{x \in X(t)} \min_{y \in Y(t)} u(x, y, t)$ forms a zero-sum matrix game for any $t \in [0,1]$, the saddle point always exist (von Neumann minimax theorem, see [52] chapter 20). Thus, the conditions of Theorem 8 are verified.

(ii) Applying mean value theorem.

Observe that

$$\begin{aligned} v(0) &= \max_{\boldsymbol{x} \in \Sigma - \boldsymbol{\omega}} \min_{\boldsymbol{y} \in Y(0)} \sum_k \sum_j x_k y_j d(\pi_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\pi})}_k) \\ &= \max_{\boldsymbol{x} \in \Sigma - \boldsymbol{\omega}} \min_{h \in H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r)} \langle \boldsymbol{x}, h \rangle = \max_{z \in \Sigma} \min_{h \in H_{F_{\boldsymbol{\pi}}}(\boldsymbol{\omega}, r)} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle. \end{aligned} \tag{64}$$

Likewise, we have

$$v(1) = \max_{z \in \Sigma} \min_{h \in H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r)} \langle \boldsymbol{z} - \boldsymbol{\omega}, h \rangle. \tag{65}$$

Theorem 8 implies that $v(t)$ is differentiable and the mean value theorem yields that there exists $t_0 \in (0,1)$ such that $v(1) - v(0) = \max_{\boldsymbol{x} \in X(t_0)} \min_{\boldsymbol{y} \in Y(t_0)} u_t(\boldsymbol{x}, \boldsymbol{y}, t_0)$. Therefore,

$$\begin{aligned} |v(1) - v(0)| &= \left| \max_{\boldsymbol{x} \in X(t_0)} \min_{\boldsymbol{y} \in Y(t_0)} u_t(\boldsymbol{x}, \boldsymbol{y}, t_0) \right| \\ &\leq \max_{k,j} \left\| \nabla_{\tilde{\boldsymbol{\pi}}(t_0)} d(\tilde{\pi}(t_0)_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \tilde{\boldsymbol{\pi}}(t_0))}_k) \right\|_1 \|\boldsymbol{\pi} - \boldsymbol{\mu}\|_\infty \\ &\leq (E)(\frac{\epsilon}{2E}) \leq \frac{\epsilon}{2}, \end{aligned} \tag{66}$$

where the first inequality is Hölder inequality and the second inequality stems from Assumption 3. By substituting equations (64)-(65) into the left-hand side of the inequality (66), we deduce the inequality (62) claimed in the theorem.

As for the inequality (63), the argument will hold by replacing $X(t) = \{\boldsymbol{z}\}$.

□

N.3 Completing the non-asymptotic analysis

Based on Theorem 9 and Lemma 14 in this section, we can state the new concentration result that will replace Lemma 2 (Appendix I.3).

Lemma 15. *Under Assumptions 1 and 3, for any $\boldsymbol{\mu} \in \Lambda, \epsilon \in (0, \kappa E)$, under* FWS*,*

$$\sum_{T=1}^{\infty} \mathbb{P}_{\boldsymbol{\mu}} \left[(\mathcal{E}_{1,\epsilon}(T) \cap \mathcal{E}_{2,\epsilon}(T))^c \right] \leq \sum_{k=1}^{K} \frac{34e^K}{d(\mu_k - \frac{\epsilon}{2E}, \mu_k)^{\frac{19}{4}}} + \frac{34e^K}{d(\mu_k + \frac{\epsilon}{2E}, \mu_k)^{\frac{19}{4}}}.$$

Proof. Replacing $\xi(\epsilon)$ by $\frac{\epsilon}{2E}$ in the proof of Lemma 2 (see Appendix J.1), we obtain the lemma (as in the proof of Lemma 2).

□

Proof of Theorem 2
Plugging the inequality derived in Lemma 15 and (42) in (41), we conclude that

$$\begin{aligned}\mathbb{E}[\tau] \leq & \sum_{k=1}^{K}\left(\frac{34e^K}{d(\mu_k-\frac{\epsilon}{2E},\mu_k)^{\frac{19}{4}}}+\frac{34e^K}{d(\mu_k+\frac{\epsilon}{2E},\mu_k)^{\frac{19}{4}}}\right)+\left(\frac{32D+3L}{\epsilon}\right)^{11} \\ & +(4K+1)^{11}+\max\left\{c_1(\Lambda),\left(\frac{2}{\tilde{\epsilon}}\right)^{11}\right\} \\ & +\frac{1+\tilde{\epsilon}}{F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))-6\epsilon}\left[\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)e}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))-6\epsilon)}\right)+\log\log\left(\frac{(1+\tilde{\epsilon})c_2(\Lambda)}{\delta(F_{\boldsymbol{\mu}}(\boldsymbol{\omega}^\star(\boldsymbol{\mu}))-6\epsilon)}\right)\right].\end{aligned}$$

This is the upper bound claimed in Theorem 2. □