Fixed-Budget Best-Arm Identification in Structured Bandits

MohammadJavad Azizi[1] , **Branislav Kveton**[2*] , **Mohammad Ghavamzadeh**[3]
[1]University of Southern California
[2]Amazon
[3]Google Research
azizim@usc.edu, bkveton@amazon.com, ghavamza@google.com

Abstract

Best-arm identification (BAI) in a fixed-budget setting is a bandit problem where the learning agent maximizes the probability of identifying the optimal (best) arm after a fixed number of observations. Most works on this topic study unstructured problems with a small number of arms, which limits their applicability. We propose a general tractable algorithm that incorporates the structure, by successively eliminating suboptimal arms based on their mean reward estimates from a joint generalization model. We analyze our algorithm in linear and generalized linear models (GLMs), and propose a practical implementation based on a G-optimal design. In linear models, our algorithm has competitive error guarantees to prior works and performs at least as well empirically. In GLMs, this is the first practical algorithm with analysis for fixed-budget BAI.

1 Introduction

Best-arm identification (BAI) is a *pure exploration* bandit problem where the goal is to identify the optimal arm. It has many applications, such as online advertising, recommender systems, and vaccine tests [Hoffman *et al.*, 2014; Lattimore and Szepesvári, 2020]. In *fixed-budget (FB)* BAI [Bubeck *et al.*, 2009; Audibert *et al.*, 2010], the goal is to accurately identify the optimal arm within a fixed budget of observations (arm pulls). This setting is common in applications where the observations are costly. However, it is more complex to analyze than the *fixed-confidence (FC)* setting, due to complications in budget allocation [Lattimore and Szepesvári, 2020, Section 33.3]. In FC BAI, the goal is to find the optimal arm with a guaranteed level of confidence, while minimizing the sample complexity.

Structured bandits are bandit problems in which the arms share a common structure, e.g., *linear* or *generalized linear* models [Filippi *et al.*, 2010; Soare *et al.*, 2014]. BAI in structured bandits has been mainly studied in the FC setting with the linear model [Soare *et al.*, 2014; Xu *et al.*, 2018; Degenne *et al.*, 2020]. The literature of FB BAI for linear bandits was limited to `BayesGap` [Hoffman *et al.*, 2014] for a long time. This algorithm does not explore sufficiently, and thus, performs poorly [Xu *et al.*, 2018]. [Katz-Samuels *et al.*, 2020] recently proposed `Peace` for FB BAI in linear bandits. Although this algorithm has desirable theoretical guarantees, it is computationally intractable, and its approximation loses the desired properties of the exact form. `OD-LinBAI` [Yang and Tan, 2021] is a concurrent work for FB BAI in linear bandits. It is a sequential halving algorithm with a special first stage, in which most arms are eliminated. This makes the algorithm inaccurate when the number of arms is much larger than the number of features, a common setting in structured problems. We discuss these three FB BAI algorithms in detail in Section 7 and empirically evaluate them in Section 8.

In this paper, we address the shortcomings of prior work by developing a general successive elimination algorithm that can be applied to several FB BAI settings (Section 3). The key idea is to divide the budget into multiple stages and allocate it *adaptively* for exploration in each stage. As the allocation is updated in each stage, our algorithm adaptively eliminates suboptimal arms, and thus, properly addresses the important trade-off between *adaptive* and *static* allocation in structured BAI [Soare *et al.*, 2014; Xu *et al.*, 2018]. We analyze our algorithm in *linear* bandits in Section 4. In Section 5, we extend our algorithm and analysis to *generalized linear models* (GLMs) and present the first BAI algorithm for these models. Our error bounds in Sections 4 and 5 motivate the use of a G-optimal allocation in each stage, for which we derive an efficient algorithm in Section 6. Using extensive experiments in Section 8, we show that our algorithm performs at least as well as a number of baselines, including `BayesGap`, `Peace`, and `OD-LinBAI`.

2 Problem Formulation

We consider a general stochastic bandit with K arms. The reward distribution of each arm $i \in \mathcal{A}$ (the set of K arms) has mean μ_i. Without loss of generality, we assume that $\mu_1 > \mu_2 \geq \cdots \geq \mu_K$; thus arm 1 is optimal. Let $x_i \in \mathbb{R}^d$ be the feature vector of arm i, such that $\sup_{i \in \mathcal{A}} ||x_i|| \leq L$ holds, where $||\cdot||$ is the ℓ_2-norm in $\mathbb{R}^d$. We denote the observed rewards of arms by $y \in \mathbb{R}$. Formally, the reward of arm i is $y = f(x_i) + \epsilon$, where ϵ is a σ^2-sub-Gaussian noise and $f(x_i)$ is any function of x_i, such that $\mu_i = f(x_i)$. In this paper, we

*This work started prior to joining Amazon.

focus on two instances of f: linear (Eq. (1)) and generalized linear (Eq. (4)).

We denote by B the fixed budget of arm pulls and by ζ the arm returned by the BAI algorithm. In the FB setting, the goal is to minimize the probability of error, i.e., $\delta = \Pr(\zeta \neq 1)$ [Bubeck *et al.*, 2009]. This is in contrast to the FC setting, where the goal is to minimize the sample complexity of the algorithm for a given upper bound on δ.

3 Generalized Successive Elimination

Successive elimination [Karnin *et al.*, 2013] is a popular BAI algorithm in multi-armed bandits (MABs). Our algorithm, which we refer to as Generalized Successive Elimination (`GSE`), generalizes it to structured reward models f. We provide the pseudo-code of `GSE` in Algorithm 1.

`GSE` operates in $s = \lceil \log_\eta K \rceil$ stages, where η is a tunable elimination parameter, usually set to be 2. The budget B is split evenly over s stages, and thus, each stage has budget $n = \lfloor B/s \rfloor$. In each stage $t \in [s]$, `GSE` pulls arms for n times and eliminates $1 - 1/\eta$ fraction of them. We denote the set of the remaining arms at the beginning of stage t by $\mathcal{A}_t$. By construction, only a single arm remains after s stages. Thus, $\mathcal{A}_1 = \mathcal{A}$ and $\mathcal{A}_{s+1} = \{\zeta\}$. In stage t, `GSE` performs the following steps:

Projection (Line 2): To avoid singularity issues, we project the remaining arms into their spanned subspace with $d_t \leq d$ dimensions. We discuss this more after Eq. (1).

Exploration (Line 3): The arms in $\mathcal{A}_t$ are sampled according to an allocation vector $\Pi_t \in \mathbb{N}^{\mathcal{A}_t}$, i.e., $\Pi_t(i)$ is the number of times that arm i is pulled in stage t. In Sections 4 and 5, we first report our results for general Π_t and then show how they can be improved if Π_t is an *adaptive* allocation based on the G-optimal design, described in Section 6.

Estimation (Line 4): Let $X_t = (X_{1,t}, \dots, X_{n,t})$ and $Y_t = (Y_{1,t}, \dots, Y_{n,t})$ be the feature vectors and rewards of the arms sampled in stage t, respectively. Given the reward model f, X_t, and Y_t, we estimate the mean reward of each arm i in stage t, and denote it by $\hat{\mu}_{i,t}$. For instance, if f is a linear function, $\hat{\mu}_{i,t}$ is estimated using linear regression, as in Eq. (1).

Elimination (Line 5): The arms in $\mathcal{A}_t$ are sorted in descending order of $\hat{\mu}_{i,t}$, their top $1/\eta$ fraction is kept, and the remaining arms are eliminated.

At the end of stage s, only one arm remains, which is *returned* as the optimal arm. While this algorithmic design is standard in MABs, it is not obvious that it would be near-optimal in structured problems, as this paper shows.

4 Linear Model

We start with the linear reward model, where $\mu_i = f(x_i) = x_i^\top \theta_*$, for an unknown reward parameter $\theta_* \in \mathbb{R}^d$. The estimate $\hat{\theta}_t$ of θ_* in stage t is computed using least-squares regression as $\hat{\theta}_t = V_t^{-1} b_t$, where $V_t = \sum_{j=1}^n X_{j,t} X_{j,t}^\top$ is the sample covariance matrix, and $b_t = \sum_{j=1}^n X_{j,t} Y_{j,t}$. This gives us the following mean estimate for each arm $i \in \mathcal{A}_t$,

$$\hat{\mu}_{i,t} = x_i^\top \hat{\theta}_t \,. \tag{1}$$

Algorithm 1 `GSE`: Generalized Successive Elimination

Input: Elimination hyper-parameter η, budget B
Initialization: $\mathcal{A}_1 \leftarrow \mathcal{A},\ t \leftarrow 1,\ s \leftarrow \lceil \log_\eta K \rceil$

1: **while** $t \leq s$ **do**
2: **Projection**: Project $\mathcal{A}_t$ to d_t dimensions, such that $\mathcal{A}_t$ spans $\mathbb{R}^{d_t}$
3: **Exploration**: Explore $\mathcal{A}_t$ using the allocation Π_t
4: **Estimation**: Calculate $(\hat{\mu}_{i,t})_{i \in \mathcal{A}_t}$ based on observed X_t and Y_t, using Eqs. (1) or (4)
5: **Elimination**: $\mathcal{A}_{t+1} = \underset{\mathcal{A} \subset \mathcal{A}_t : |\mathcal{A}| = \lceil \frac{|\mathcal{A}_t|}{\eta} \rceil}{\arg\max} \sum_{i \in \mathcal{A}} \hat{\mu}_{i,t}$
6: $t \leftarrow t + 1$
7: **end while**
8: **Output**: ζ such that $\mathcal{A}_{s+1} = \{\zeta\}$

The matrix V_t^{-1} is well-defined as long as X_t spans $\mathbb{R}^d$. However, since `GSE` eliminates arms, it may happen that the arms in later stages do not span $\mathbb{R}^d$. Thus, V_t could be singular and V_t^{-1} would not be well-defined. We alleviate this problem by projecting[1] the arms in $\mathcal{A}_t$ into their spanned subspace. We denote the dimension of this subspace by d_t. Alternatively, we can address the singularity issue by using the pseudo-inverse of matrices [Huang *et al.*, 2021]. In this case, we remove the projection step, and replace V_t^{-1} with its pseudo-inverse.

4.1 Analysis

In this section, we prove an error bound for `GSE` with the linear model. Although this error bound is a special case of that for GLMs (see Theorem 2), we still present it because more readers are familiar with linear bandit analysis than GLMs. To reduce clutter, we assume that all logarithms have base η. We denote by $\Delta_i = \mu_1 - \mu_i$, the sub-optimality gap of arm i, and by $\Delta_{\min} = \min_{i>1} \Delta_i$, the minimum gap, which by the assumption in Section 2 is just Δ_2.

Theorem 1. `GSE` *with the linear model (Eq.* (1)*) and any valid*[2] *allocation strategy* Π_t *identifies the optimal arm with probability at least* $1 - \delta$ *for*

$$\delta \leq 2\eta \log(K) \exp\left(\frac{-\Delta_{\min}^2 \sigma^{-2}}{4 \max\limits_{i \in \mathcal{A}, t \in [s]} ||x_i - x_1||_{V_t^{-1}}^2} \right). \tag{2}$$

where $||x||_V = \sqrt{x^\top V x}$ *for any* $x \in \mathbb{R}^d$ *and matrix* $V \in \mathbb{R}^{d \times d}$*. If we use the G-optimal design (Algorithm 2) for* Π_t*, then*

$$\delta \leq 2\eta \log(K) \exp\left(\frac{-B\Delta_{\min}^2}{4\sigma^2 d \log(K)} \right). \tag{3}$$

We sketch the proof in Section 4.2 and defer the detailed proof to Appendix A.

The error bound in (3) scales as expected. Specifically, it is tighter for a larger budget B, which increases the statistical power of `GSE`; and a larger gap $\Delta_{\min}$, which makes the

[1] The projection can be done by multiplying the arm features with the matrix whose columns are the orthonormal basis of the subspace spanned by the arms [Yang and Tan, 2021].

[2] Allocation strategy Π_t is valid if V_t is invertible.

optimal arm easier to identify. The bound is looser for larger K and d, which increase with the instance size; and larger reward noise σ, which increases uncertainty and makes the problem instance harder to identify. We compare this bound to the related works in Section 7.

There is no lower bound for FB BAI in structured bandits. Nevertheless, in the special case of MABs, our bound ((3)) matches the FB BAI lower bound $\exp\left(\frac{-B}{\sum_{i\in\mathcal{A}}\Delta_i^{-2}}\right)$ in Kaufmann *et al.* [2016], up to a factor of $\log K$. It also roughly matches the tight lower bound of Carpentier and Locatelli [2016], which is $\exp\left(\frac{-B}{\log(K)\sum_{i\in\mathcal{A}}\Delta_i^{-2}}\right)$. To see this, note that $\sum_{i\in\mathcal{A}}\Delta_i^{-2}\approx K\Delta_{\min}^{-2}$ and $d=K$, when we apply GSE to a K-armed bandit problem.

4.2 Proof Sketch

The key idea in analyzing GSE is to control the probability of eliminating the optimal arm in each stage. Our analysis is modular and easy to extend to other elimination algorithms. Let E_t be the event that the optimal arm is eliminated in stage t. Then, $\delta=\Pr(\cup_{t=1}^s E_t)\le\sum_{t=1}^s\Pr(E_t|\bar{E}_1,\dots,\bar{E}_{t-1})$, where $\bar{E}_t$ is the complement of event E_t. In Lemma 1, we bound the probability that a suboptimal arm has a higher estimated mean reward than the optimal arm. This is a novel concentration result for linear bandits in successive elimination algorithms.

Lemma 1. *In* GSE *with the linear model of Eq.* (1)*, the probability that any suboptimal arm i has a higher estimated mean reward than the optimal arm in stage t satisfies* $\Pr(\hat{\mu}_{i,t}>\hat{\mu}_{1,t})\le 2\exp\left(\frac{-\Delta_i^2\sigma^{-2}}{2||x_i-x_1||^2_{V_t^{-1}}}\right)$.

This lemma is proved using an argument mainly driven from a concentration bound. Next, we use it in Lemma 2 to bound the probability that the optimal arm is eliminated in stage t.

Lemma 2. *In* GSE *with the linear model (Eq.* (1)*), the probability that the optimal arm is eliminated in stage t satisfies* $\Pr(\tilde{E}_t)\le 2\eta\exp\left(\frac{-\Delta_{\min,t}^2\sigma^{-2}}{2\max_{i\in\mathcal{A}_t}||x_i-x_1||^2_{V_t^{-1}}}\right)$, *where* $\Delta_{\min,t}=\min_{i\in\mathcal{A}_t\setminus\{1\}}\Delta_i$ *and* $\tilde{E}_t$ *is a shorthand for event* $E_t|\bar{E}_1,\dots,\bar{E}_{t-1}$.

This lemma is proved by examining how another arm can dominate the optimal arm and using Markov's inequality. Finally, we bound δ in Theorem 1 using a union bound. We obtain the second bound in Theorem 1 by the Kiefer-Wolfowitz Theorem [Kiefer and Wolfowitz, 1960] for the G-optimal design described in Section 6.

5 Generalized Linear Model

We now study FB BAI in *generalized linear models (GLMs)* [McCullagh and Nelder, 1989], where $\mu_i=f(x_i)=h(x_i^\top\theta_*)$, where h is a monotone function known as the *mean function*. As an example, $h(x)=(1+\exp(-x))^{-1}$ in logistic regression. We assume that the derivative of the mean function, h', is bounded from below, i.e., $c_{\min}\le h'(x_i^\top\tilde{\theta}_t)$, for some $c_{\min}\in\mathbb{R}^+$ and all $i\in\mathcal{A}$. Here $\tilde{\theta}_t$ can be any convex combination of θ_* and its *maximum likelihood estimate* $\hat{\theta}_t$ in stage t. This assumption is standard in GLM bandits [Filippi *et al.*, 2010; Li *et al.*, 2017]. The existence of $c_{\min}$ can be guaranteed by performing forced exploration at the beginning of each stage with the sampling cost of $O(d)$ [Kveton *et al.*, 2020]. As $\hat{\theta}_t$ satisfies $\sum_{j=1}^n\left(Y_{j,t}-h(X_{j,t}^\top\hat{\theta}_t)\right)X_{j,t}=0$, it can be computed efficiently by *iteratively reweighted least squares* [Wolke and Schwetlick, 1988]. This gives us the following mean estimate for each arm $i\in\mathcal{A}_t$,

$$\hat{\mu}_{i,t}=h(x_i^\top\hat{\theta}_t)\,. \tag{4}$$

5.1 Analysis

In Theorem 2, we prove similar bounds to the linear model. The proof and its sketch are presented in Appendix B. These are the first BAI error bounds for GLM bandits.

Theorem 2. GSE *with the GLM (Eq.* (4)*) and any valid Π_t identifies the optimal arm with probability at least $1-\delta$ for*

$$\delta\le 2\eta\log(K)\exp\left(\frac{-\Delta_{\min}^2\sigma^{-2}c_{\min}^2}{8\max\limits_{i\in\mathcal{A},t\in[s]}||x_i||^2_{V_t^{-1}}}\right). \tag{5}$$

If we use the G-optimal design (Algorithm 2) for Π_t, then

$$\delta\le 2\eta\log(K)\exp\left(\frac{-B\Delta_{\min}^2c_{\min}^2}{8\sigma^2 d\log(K)}\right). \tag{6}$$

The error bounds in Theorem 2 are similar to those in the linear model (Section 4.1), since $\max_{i\in\mathcal{A},t\in[s]}||x_i-x_1||_{V_t^{-1}}\le 2\max_{i\in\mathcal{A},t\in[s]}||x_i||_{V_t^{-1}}$.The only major difference is in factor $c_{\min}^2$, which is 1 in the linear case. This factor arises because GLM is a linear model transformed through some non-linear mean function h. When $c_{\min}$ is small, h can have flat regions, which makes the optimal arm harder to identify. Therefore, our GLM bounds become looser as $c_{\min}$ decreases. Note that the bounds in Theorem 2 depend on all other quantities same as the bounds in Theorem 1 do.

The novelty in our GLM analysis is in how we control the estimation error of θ_* using our assumptions on the existence of $c_{\min}$. The rest of the proof follows similar steps to those in Section 4.2 and are postponed to Appendix B.

6 G-Optimal Allocation

The stochastic error bounds in (2) and (5) can be optimized by minimizing $2\max_{i\in\mathcal{A},t\in[s]}||x_i||_{V_t^{-1}}$ with respect to V_t, in particular, with respect to X_t. In each stage t, let $g_t(\pi,x_i)=||x_i||^2_{V_t^{-1}}$, where $V_t=n\sum_{i\in\mathcal{A}_t}\pi_i x_i x_i^\top$ and $\sum_{i\in\mathcal{A}_t}\pi_i=1$. Then, optimization of V_t is equivalent to solving $\min_\pi\max_{i\in\mathcal{A}_t}g_t(\pi,x_i)$. This leads us to the G-optimal design [Kiefer and Wolfowitz, 1960], which minimizes the maximum variance along all x_i.

We develop an algorithm based on the Frank-Wolfe (FW) method [Jaggi, 2013] to find the G-optimal design. Algorithm 2 contains the pseudo-code of it, which we refer to as FWG. The G-optimal design is a convex relaxation of

Algorithm 2 Frank-Wolfe G-optimal allocation (FWG)

Input: Stage budget n, N number of iterations
Initialization: $\pi_0 \leftarrow (1, \dots, 1)/|\mathcal{A}_t| \in \mathbb{R}^{|\mathcal{A}_t|}$, $i \leftarrow 0$
while $i < N$ **do**
  $\pi'_i \leftarrow \arg\min_{\pi' : ||\pi'||_1 = 1} \nabla_\pi g_t(\pi_i)^\top \pi'$ {Surrogate}
  $\gamma_i \leftarrow \arg\min_{\gamma \in [0,1]} g_t(\pi_i + \gamma(\pi'_i - \pi_i))$ {Line search}
  $\pi_{i+1} \leftarrow \pi_i + \gamma_i(\pi'_i - \pi_i)$ {Gradient step}
  $i \leftarrow i + 1$
end while
Output: $\Pi_t = \text{ROUND}(n, \pi_N)$ {Rounding}

the G-optimal allocation; an allocation is the (integer) number of samples per arm while a design is the proportion of n for each arm. Defining $g_t(\pi) = \max_{i \in \mathcal{A}_t} g_t(\pi, x_i)$, by Danskin's theorem [Danskin, 1966], we know $\nabla_{\pi_j} g_t(\pi) = -n(x_j^\top V_t^{-1} x_{\max})^2$, where $x_{\max} = \arg\max_{i \in \mathcal{A}_t} g_t(\pi, x_i)$. This gives us the derivative of the objective function so we can use it in a FW algorithm. In each iteration, FWG first minimizes the 1st-order surrogate of the objective, and then uses line search to find the best step-size and takes a gradient step. After N iterations, it extracts an allocation (integral solution) from π_N using an efficient rounding procedure from Allen-Zhu *et al.* [2017], which we call it $\text{ROUND}(n, \pi)$. This procedure takes budget n, design π_N, and returns an allocation Π_t.

In Appendix C, we show that the error bounds of Theorems 1 and 2 still hold for large enough N, if we use Algorithm 2 to obtain the allocation strategy Π_t at the exploration step (Line 3 of Algorithm 1). This results in the deterministic bounds in (3) and (6) in these theorems.

7 Related Work

To the best of our knowledge, there is no prior work on FB BAI for GLMs and our results are the first in this setting. However, there are three related algorithms for FB BAI in linear bandits that we discuss them in detail here. Before we start, note that there is no matching upper and lower bound for FB BAI in any setting [Carpentier and Locatelli, 2016]. However, in MABs, it is known that *successive elimination* is near-optimal [Carpentier and Locatelli, 2016].

BayesGap [Hoffman *et al.*, 2014] is a Bayesian version of the gap-based exploration algorithm in Gabillon *et al.* [2012]. This algorithm models correlations of rewards using a Gaussian process. As pointed out by Xu *et al.* [2018], BayesGap does not explore enough and thus performs poorly. In Appendix D.1, we show under few simplifying assumptions that the error probability of BayesGap is at most $KB \exp\left(\frac{-B\Delta_{\min}^2}{32K}\right)$. Our error bound in Eq. (3) is at most $2\eta \log(K) \exp\left(\frac{-B\Delta_{\min}^2}{4d\log(K)}\right)$. Thus, it improves upon BayesGap by reducing dependence on the number of arms K, from linear to logarithmic; and on budget B, from linear to constant. We provide a more detailed comparison of these bounds in Appendix D.1. Our experimental results in Section 8 support these observations and show that our algorithm always outperforms BayesGap in the linear setting.

Peace [Katz-Samuels *et al.*, 2020] is mainly a FC BAI algorithm based on a transductive design, which is modified to be used in the FB setting. It minimizes the *Gaussian-width* of the remaining arms with a progressively finer level of granularity. However, Peace cannot be implemented exactly because the Gaussian width does not have a closed form and is computationally expensive to minimize. To address this, Katz-Samuels *et al.* [2020] proposed an approximation to Peace, which still has some computational issues (see Remark D.2 and Section 8.1). The error bound for Peace, although is competitive, only holds for a relatively large budget (Theorem 7 in [Katz-Samuels *et al.*, 2020]). We discuss this further in Remark D.1. Although the comparison of their bound to ours is not straightforward, we show in Appendix D.2 that each bound can be superior in certain regimes that depend mainly on the relation of d and K. In particular, we show two cases: (i) Based on few claims in Katz-Samuels *et al.* [2020] that are not rigorously proved (see (i) in Appendix D.2 for more details), their error bound is at most $2\lceil\log(d)\rceil \exp\left(\frac{-B\Delta_{\min}^2}{\max_{i\in\mathcal{A}}||x_i - x_1||_{V^{-1}} \log(d)}\right)$ which is better than our bound (Eq. (2)) only if $K > \exp(\exp(\log(d)\log\log(d)))$. (ii) We can also show that their bound is at most $2\lceil\log(d)\rceil \exp\left(\frac{-B\Delta_{\min}^2}{d\log(K)\log(d)}\right)$ under the G-optimal design, which is worse than our error bound (Eq. (3)).

In our experiments with Peace in Section 8, we implemented its approximation and it never performed better than our algorithm. We also show in Section 8.1 that approximate Peace is much more computationally expensive compared to our algorithm.

OD-LinBAI [Yang and Tan, 2021] uses a G-optimal design in a sequential elimination framework for FB BAI. In the first stage, it eliminates all the arms except $\lceil d/2 \rceil$. This makes the algorithm prone to eliminating the optimal arm in the first stage, especially when the number of arms is larger than d. It also adds a linear (in K) factor to the error bound. In Appendix D.3, we provide a detailed comparison between the error bound of OD-LinBAI and ours, and show that similar to the comparison with Peace, there are regimes where each bound is superior. However, we show that our bound is tighter in the more practically relevant setting of $K = \Omega(d^2)$. In particular, we show that their error is at most $\left(\frac{4K}{d} + 3\log(d)\right) \exp\left(\frac{(d^2-B)\Delta_{\min}^2}{32d\log(d)}\right)$. Now assuming $K = d^q$ for some $q \in \mathbb{R}$, if we divide our bound (Eq. (3)) with theirs, we obtain $O\left(\frac{q\log(d)}{d^{q-1}+\log(d)} \exp\left(\frac{-d^2\Delta_{\min}^2}{d\log(d)}\right)\right)$, which is less than 1, so in this case *our error bound is tighter*. However, for $K < d(d+1)/2$, their bound is tighter. Finally, we note that our experiments in Section 8 and Appendix E.3 support these observations.

8 Experiments

In this section, we compare GSE to several baselines including all linear FB BAI algorithms: Peace, BayesGap, and OD-LinBAI. Others are variants of cumulative regret (CR) bandits and FC BAI algorithms. For CR algorithms, the baseline

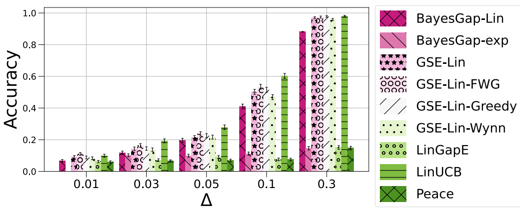

Figure 1: Static allocation.

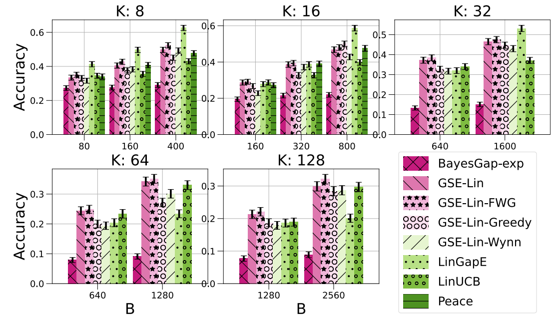

Figure 2: Adaptive instance for $d = K - 1$.

stops at the budget limit and returns the most pulled arm.[3] We use LinUCB [Li *et al.*, 2010] and UCB-GLM [Li *et al.*, 2017], which are the state-of-the-art for linear and GLM bandits, respectively. LinGapE (a FC BAI algorithm) [Xu *et al.*, 2018] is used with its stopping rule at the budget limit. We tune its δ using a grid search and only report the best result. In Appendix F, we derive proper error bounds for these baselines to further justify the variants.

The *accuracy* is an estimate of $1 - \delta$, as the fraction of 1000 Monte Carlo replications where the algorithm finds the optimal arm. We run GSE with linear model and uniform exploration (GSE-Lin), with FWG (GSE-Lin-FWG), with sequential G-optimal allocation of Soare *et al.* [2014] (GSE-Lin-Greedy), and with Wynn's G-optimal method (GSE-Lin-Wynn). For Wynn's method, see Fedorov [1972]. We set $\eta = 2$ in all experiments, as this value tends to perform well in successive elimination [Karnin *et al.*, 2013]. For LinGapE, we evaluate the Greedy version (LinGapE-Greedy) and show its results only if it outperforms LinGapE. For LinGapE-Greedy, see [Xu *et al.*, 2018]. In each experiment, we fix K, B/K, or d; depending on the experiment to show the desired trend. Similar trends can be observed if we fix the other parameters and change these. For further detail of our choices of kernels for BayesGap and also our real-world data experiments, see Appendix E.

8.1 Linear Experiment: Adaptive Allocation

We start with the example in Soare *et al.* [2014], where the arms are the canonical d-dimensional basis $e_1, e_2, \ldots, e_d$ plus a disturbing arm $x_{d+1} = (\cos(\omega), \sin(\omega), 0, \ldots, 0)^\top$ with $\omega = 1/10$. We set $\theta_* = e_1$ and $\epsilon \sim \mathcal{N}(0, 10)$. Clearly the optimal arm is e_1, however, when the angle ω is as small as 1/10, the disturbing arm is hard to distinguish from e_1. As argued in Soare *et al.* [2014], this is a setting where an adaptive strategy is optimal (see Appendix G.1 for further discussion on Adaptive vs. Static strategies).

Fig. 2 shows that GSE-Lin-FWG is the second-best algorithm for smaller K and the best for larger K. BayesGap-Lin performs poorly here, and thus, we omit it. We *conjecture* that BayesGap-Lin fails because it uses Gaussian processes and there is a very low correlation between the arms in this experiment. LinGapE wins mostly for smaller K and loses for larger K. This could be because its regret is linear in K (Appendix D). Peace has lower accuracy than several other algorithms. We could only simulate Peace for $K \leq 16$, since its computational cost is high for larger values of K. For instance, at $K = 16$, Peace completes 100 runs in 530 seconds; while it only takes 7 to 18 seconds for the other algorithms. At $K = 32$, Peace completes 100 runs in 14 hours (see Appendix E.1).

In this experiment, $K \approx d$ and both OD-LinBAI and GSE have $\log(K)$ stages and perform similarly. Therefore, we only report the results for GSE. This also happens in Section 8.2.

8.2 Linear Experiment: Static Allocation

As in Xu *et al.* [2018], we take arms $e_1, e_2, ..., e_{16}$ and $\theta_* = (\Delta, 0, \ldots, 0)$, where $K = d = 16$ and $B = 320$. In this experiment, knowing the rewards does not change the allocation strategy. Therefore, a static allocation is optimal [Xu *et al.*, 2018]. The goal is to evaluate the ability of the algorithm to adapt to a static situation.

Our results are reported in Fig. 1. We observe that LinUCB performs the best when Δ is small (harder instances). This is expected since suboptimal arms are well away from the optimal one, and CR algorithms do well in this case (Appendix D). Our algorithms are the second-best when Δ is sufficiently large, converging to the optimal static allocation. BayesGap-exp, LinGapE, and Peace cannot take advantage of larger Δ, probably because they adapt to the rewards too early. This example demonstrates how well our algorithms adjust to a static allocation, and thus, properly address the tradeoff between static and adaptive allocation.

8.3 Linear Experiment: Randomized

In this experiment, we use the example in Tao *et al.* [2018] and [Yang and Tan, 2021]. For each bandit instance, we generate i.i.d. arms sampled from the unit sphere centered at the origin with $d = 10$. We let $\theta_* = x_i + 0.01(x_j - x_i)$, where x_i and x_j are the two closest arms. As a consequence, x_i is the optimal arm and x_j is the disturbing arm. The goal is to evaluate the expected performance of the algorithms for a random instance to avoid bias in choosing the bandit instances.

We fix B/K in Fig. 3 and compare the performance for different K. GSE-Lin-FWG has competitive performance with other algorithms. We can see that G-optimal policies have similar expected performance while FWG is slightly better. Again, LinGapE performance degrades as K increases and Peace underperforms our algorithms. Moreover, the performance of OD-LinBAI worsens as K increases, especially for $K > \frac{d(d+1)}{2}$. We report more experiments in this setting, comparing GSE to OD-LinBAI, in Appendix E.3.

[3] In Appendix D, we argue that this is a reasonable stopping rule.

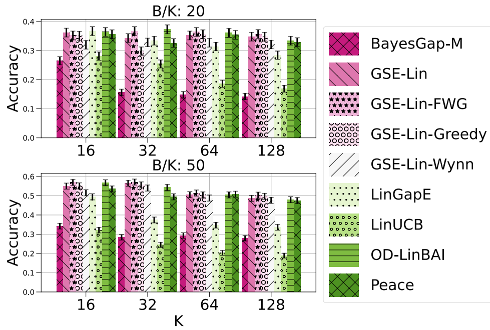

Figure 3: Randomized linear experiment.

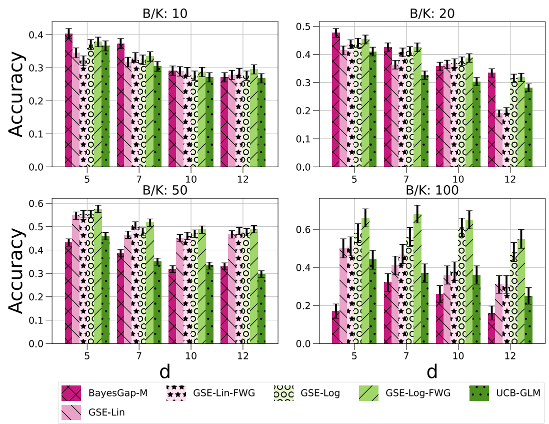

Figure 4: Logistic bandit experiment for $K = 8$.

8.4 GLM Experiment

As an instance of GLM, we study a *logistic bandit*. We generate i.i.d. arms from uniform distribution on $[-0.5, 0.5]^d$ with $d \in \{5, 7, 10, 12\}$, $K = 8$, and $\theta_* \sim N(0, \frac{3}{d} I_d)$, where I_d is a $d \times d$ identity matrix. The reward of arm i is defined as $y_i \sim \text{Bern}(h(x_i^\top \theta_*))$, where $h(z) = (1 + \exp(-z))^{-1}$ and $\text{Bern}(z)$ is a Bernoulli distribution with mean z. We use `GSE` with a logistic regression model (`GSE-Log`) and also with the linear models to evaluate the robustness of `GSE` to *model misspecification*. For exploration, we only use `FWG` (`GSE-Log-FWG`), as it performs better than the other G-optimal allocations in earlier experiments. We also use a modification of `UCB-GLM` [Li *et al.*, 2017], a state-of-the-art GLM CR algorithm, for FB BAI.

The results in Fig. 4 show `GSE` with logistic models outperforms linear models, and `FWG` improves on uniform exploration in the GLM case. These experiments also show the robustness of `GSE` to model misspecification, since the linear model only slightly underperforms the logistic model. `UCB-GLM` results confirm that CR algorithms could fail in BAI. `BayesGap-M` falls short for $B/K \geq 50$; the extra B in their error bound also suggests failure for large B. In contrast, the performance of `GSE` keeps improving as B increases.

9 Conclusions

In this paper, we studied fixed-budget best-arm identification (BAI) in linear and generalized linear models. We proposed the `GSE` algorithm, which offers an adaptive framework for structured BAI. Our performance guarantees are near-optimal in MABs. In generalized linear models, our algorithm is the first practical fixed-budget BAI algorithm with analysis. Our experiments show the efficiency and robustness (to model misspecification) of our algorithm. Extending our `GSE` algorithm to more general models could be a future direction (see Appendix H).

References

[Abbasi-yadkori *et al.*, 2011] Yasin. Abbasi-yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

[Allen-Zhu *et al.*, 2017] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal design of experiments via regret minimization. In *International Conference on Machine Learning*, pages 126–135, 2017.

[Audibert *et al.*, 2010] Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best Arm Identification in Multi-Armed Bandits. In *Proceedings of the 23th Conference on Learning Theory*, 2010.

[Berthet and Perchet, 2017] Quentin Berthet and Vianney Perchet. Fast rates for bandit optimization with upper-confidence frank-wolfe. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[Bubeck *et al.*, 2009] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pages 23–37, 2009.

[Carpentier and Locatelli, 2016] Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pages 590–604, 2016.

[Damla Ahipasaoglu *et al.*, 2008] Selin Damla Ahipasaoglu, Peng Sun, and Michael J Todd. Linear convergence of a modified Frank–Wolfe algorithm for computing minimum-volume enclosing ellipsoids. *Optimization Methods and Software*, 23(1):5–19, 2008.

[Danskin, 1966] John M. Danskin. The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–664, 1966.

[Degenne *et al.*, 2019] Rémy Degenne, Wouter M Koolen, and Pierre Ménard. Non-asymptotic pure exploration by solving games. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[Degenne *et al.*, 2020] Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pages 2432–2442, 2020.

[Fedorov, 1972] Valerii Vadimovich Fedorov. *Theory of Optimal Experiments*. Probability and Mathematical Statistics. Academic Press, 1972.

[Filippi *et al.*, 2010] Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, 2010.

[Gabillon *et al.*, 2012] Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems*, pages 3212–3220, 2012.

[Hoffman *et al.*, 2014] Matthew Hoffman, Bobak Shahriari, and Nando Freitas. On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning. In *Artificial Intelligence and Statistics*, pages 365–374, 2014.

[Huang *et al.*, 2021] Ruiquan Huang, Weiqiang Wu, Jing Yang, and Cong Shen. Federated linear contextual bandits. *Advances in Neural Information Processing Systems*, 34, 2021.

[Jaggi, 2013] Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pages 427–435, 2013.

[Jedra and Proutiere, 2020] Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 10007–10017, 2020.

[Karnin *et al.*, 2013] Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*, page 1238–1246, 2013.

[Katz-Samuels *et al.*, 2020] Julian Katz-Samuels, Lalit Jain, Zohar Karnin, and Kevin Jamieson. An empirical process approach to the union bound: Practical algorithms for combinatorial and linear bandits. In *Advances in Neural Information Processing Systems*, pages 10371–10382, 2020.

[Kaufmann *et al.*, 2016] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

[Khachiyan, 1996] Leonid G Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2):307–320, 1996.

[Kiefer and Wolfowitz, 1960] Jack Kiefer and Jacob Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.

[Kumar and Yildirim, 2005] Piyush Kumar and E Alper Yildirim. Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1):1–21, 2005.

[Kveton *et al.*, 2020] Branislav Kveton, Manzil Zaheer, Csaba Szepesvari, Lihong Li, Mohammad Ghavamzadeh, and Craig Boutilier. Randomized exploration in generalized linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2066–2076, 2020.

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

[Li *et al.*, 2010] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. *International Conference on World Wide Web*, 2010.

[Li *et al.*, 2017] Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pages 2071–2080, 2017.

[McCullagh and Nelder, 1989] Peter McCullagh and John A Nelder. *Generalized Linear Models*. Chapman & Hall, 1989.

[Riquelme *et al.*, 2018] Carlos Riquelme, George Tucker, and Jasper Snoek. Deep bayesian bandits showdown: An empirical comparison of bayesian deep networks for Thompson sampling. In *International Conference on Learning Representations*, 2018.

[Soare *et al.*, 2014] Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836, 2014.

[Tao *et al.*, 2018] Chao Tao, Saúl Blanco, and Yuan Zhou. Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, pages 4877–4886, 2018.

[Tripuraneni *et al.*, 2021] Nilesh Tripuraneni, Chi Jin, and Michael Jordan. Provable meta-learning of linear representations. In *International Conference on Machine Learning*, pages 10434–10443, 2021.

[Vershynin, 2019] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics, 2019.

[Wolke and Schwetlick, 1988] Robert. Wolke and Hartmut Schwetlick. Iteratively reweighted least squares: algorithms, convergence analysis, and numerical comparisons. *SIAM Journal on Scientific and Statistical Computing*, 9(5):907–921, 1988.

[Xu *et al.*, 2018] Liyuan Xu, Junya Honda, and Masashi Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851, 2018.

[Yang and Tan, 2021] Junwen Yang and Vincent YF Tan. Towards minimax optimal best arm identification in linear bandits. *arXiv preprint arXiv:2105.13017*, 2021.