

Non-stochastic Best Arm Identification and Hyperparameter Optimization

Kevin Jamieson

Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706

KGJAMIESON@WISC.EDU

Ameet Talwalkar

Computer Science Department, UCLA, Boelter Hall 4732, Los Angeles, CA 90095

AMEET@CS.UCLA.EDU

Abstract

Motivated by the task of hyperparameter optimization, we introduce the *non-stochastic best-arm identification problem*. Within the multi-armed bandit literature, the cumulative regret objective enjoys algorithms and analyses for both the non-stochastic and stochastic settings while to the best of our knowledge, the best-arm identification framework has only been considered in the stochastic setting. We introduce the non-stochastic setting under this framework, identify a known algorithm that is well-suited for this setting, and analyze its behavior. Next, by leveraging the iterative nature of standard machine learning algorithms, we cast hyperparameter optimization as an instance of non-stochastic best-arm identification, and empirically evaluate our proposed algorithm on this task. Our empirical results show that, by allocating more resources to promising hyperparameter settings, we typically achieve comparable test accuracies an order of magnitude faster than baseline methods.

1. Introduction

As supervised learning methods are becoming more widely adopted, hyperparameter optimization has become increasingly important to simplify and speed up the development of data processing pipelines while simultaneously yielding more accurate models. In hyperparameter optimization for supervised learning, we are given labeled training data, a set of hyperparameters associated with our supervised learning methods of interest, and a search space over these hyperparameters. We aim to find a particular configuration of hyperparameters that optimizes some evaluation criterion, e.g., loss on a validation dataset.

Since many machine learning algorithms are iterative in nature, particularly when working at scale, we can evaluate the quality of intermediate results, i.e., partially trained

learning models, resulting in a sequence of losses that eventually converges to the final loss value at convergence. For example, Figure 1 shows the sequence of validation losses for various hyperparameter settings for kernel SVM models trained via stochastic gradient descent. The figure shows high variability in model quality across hyperparameter settings. It thus seems natural to ask the question: *Can we terminate these poor-performing hyperparameter settings early in a principled online fashion to speed up hyperparameter optimization?*

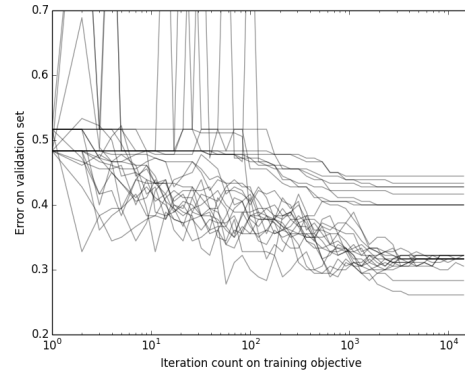


Figure 1. Validation error for different hyperparameter choices for a classification task trained using stochastic gradient descent.

Although several hyperparameter optimization methods have been proposed recently, e.g., Snoek et al. (2012; 2014); Hutter et al. (2011); Bergstra et al. (2011); Bergstra & Bengio (2012), the vast majority of them consider the training of machine learning models to be black-box procedures, and only evaluate models after they are fully trained to convergence. A few recent works have made attempts to exploit intermediate results. However, these works either require explicit forms for the convergence rate behavior of the iterates which is difficult to accurately characterize for all but the simplest cases (Agarwal et al., 2012; Swersky et al., 2014), or focus on heuristics lacking theoretical underpinnings (Sparks et al., 2015). We build upon these previous works, and in particular study the multi-armed bandit

formulation proposed in Agarwal et al. (2012) and Sparks et al. (2015), where each arm corresponds to a fixed hyperparameter setting, pulling an arm corresponds to a fixed number of training iterations, and the loss corresponds to an intermediate loss on some hold-out set.

We aim to provide a robust, general-purpose, and widely applicable bandit-based solution to hyperparameter optimization. Remarkably, however, the existing multi-armed bandits literature fails to address this natural problem setting: a *non-stochastic best-arm identification* problem. While multi-armed bandits is a thriving area of research, we believe that the existing work fails to adequately address the two main challenges in this setting:

1. We know each arm’s sequence of losses eventually converges, but we have no information about the rate of convergence, and the sequence of losses, like those in Figure 1, may exhibit a high degree of non-monotonicity and non-smoothness.
2. The cost of obtaining the loss of an arm can be disproportionately more costly than pulling it. For example, in the case of hyperparameter optimization, computing the validation loss is often drastically more expensive than performing a single training iteration.

We thus study this novel bandit setting, which encompasses the hyperparameter optimization problem, and analyze an algorithm we identify as being particularly well-suited for this setting. Moreover, we confirm our theory with empirical studies that demonstrate an order of magnitude speedups relative to standard baselines on a number of real-world supervised learning problems and datasets.

We note that this bandit setting is quite generally applicable. While the problem of hyperparameter optimization inspired this work, the setting itself encompasses the stochastic best-arm identification problem (Bubeck et al., 2009), less-well-behaved stochastic sources like max-bandits (Cirello & Smith, 2005), exhaustive subset selection for feature extraction, and many optimization problems that “feel” like stochastic best-arm problems but lack the i.i.d. assumptions necessary in that setting.

The remainder of the paper is organized as follows: In Section 2 we present the setting of interest, provide a survey of related work, and explain why most existing algorithms and analyses are not well-suited or applicable for our setting. We then study our proposed algorithm in Section 3 in our setting of interest, and analyze its performance relative to a natural baseline. We then relate these results to the problem of hyperparameter optimization in Section 4, and present our experimental results in Section 5.

2. Non-stochastic best arm identification

Objective functions for multi-armed bandits problems tend to take on one of two flavors: 1) best arm identification (or pure exploration) in which one is interested in identifying the arm with the highest average payoff, and 2) exploration-versus-exploitation in which we are trying to maximize the cumulative payoff over time (Bubeck & Cesa-Bianchi, 2012). While the latter has been analyzed in both the stochastic and non-stochastic settings, we are unaware of any work that addresses the best arm objective in the non-stochastic setting, which is our setting of interest. Moreover, while related, a strategy that is well-suited for maximizing cumulative payoff is not necessarily well-suited for the best-arm identification task, even in the stochastic setting (Bubeck et al., 2009).

Best Arm Problem for Multi-armed Bandits

input: n arms where $\ell_{i,k}$ denotes the loss observed on the k th pull of the i th arm
initialize: $T_i = 1$ for all $i \in [n]$
for $t = 1, 2, 3, \dots$
 Algorithm chooses an index $I_t \in [n]$
 Loss $\ell_{I_t, T_{I_t}}$ is revealed, $T_{I_t} = T_{I_t} + 1$
 Algorithm outputs a recommendation $J_t \in [n]$
 Receive external stopping signal, otherwise continue

Figure 2. A generalization of the best arm problem for multi-armed bandits (Bubeck et al., 2009) that applies to both the stochastic and non-stochastic settings.

The algorithm of Figure 2 presents a general form of the best arm problem for multi-armed bandits. Intuitively, at each time t the goal is to choose J_t such that the arm associated with J_t has the lowest loss in some sense. Note that while the algorithm gets to observe the value for an arbitrary arm I_t , the algorithm is only evaluated on its recommendation J_t , that it also chooses arbitrarily. This is in contrast to the exploration-versus-exploitation game where the arm that is played is also the arm that the algorithm is evaluated on, namely, I_t .

The best-arm identification problems defined below require that the losses be generated by an oblivious adversary, which essentially means that the loss sequences are independent of the algorithm’s actions. Contrast this with an adaptive adversary that can adapt future losses based on all the arms that the algorithm has played up to the current time. If the losses are chosen by an oblivious adversary then without loss of generality we may assume that all the losses were generated before the start of the game. See (Bubeck & Cesa-Bianchi, 2012) for more info. We now compare the stochastic and the proposed non-stochastic best-arm identification problems.

Stochastic : For all $i \in [n], k \geq 1$, let $\ell_{i,k}$ be an i.i.d. sample from a probability distribution supported on $[0, 1]$. For each i , $\mathbb{E}[\ell_{i,k}]$ exists and is equal to some constant μ_i for all $k \geq 1$. The goal is to identify $\arg \min_i \mu_i$ while minimizing $\sum_{i=1}^n T_i$.

Non-stochastic (proposed in this work) : For all $i \in [n], k \geq 1$, let $\ell_{i,k} \in \mathbb{R}$ be generated by an oblivious adversary and assume $\nu_i = \lim_{\tau \rightarrow \infty} \ell_{i,\tau}$ exists. The goal is to identify $\arg \min_i \nu_i$ while minimizing $\sum_{i=1}^n T_i$.

These two settings are related in that we can always turn the stochastic setting into the non-stochastic setting by defining $\ell_{i,T_i} = \frac{1}{T_i} \sum_{k=1}^{T_i} \ell'_{i,k}$ where $\ell'_{i,k}$ are the losses from the stochastic problem; by the law of large numbers $\lim_{\tau \rightarrow \infty} \ell_{i,\tau} = \mathbb{E}[\ell'_{i,1}]$. In fact, we could do something similar with other less-well-behaved statistics like the minimum (or maximum) of the stochastic returns of an arm. As described in Cicirello & Smith (2005), we can define $\ell_{i,T_i} = \min\{\ell'_{i,1}, \ell'_{i,2}, \dots, \ell'_{i,T_i}\}$, which has a limit since $\ell_{i,t}$ is a bounded, monotonically decreasing sequence.

However, the generality of the non-stochastic setting introduces novel challenges. In the stochastic setting, if we set $\hat{\mu}_{i,T_i} = \frac{1}{T_i} \sum_{k=1}^{T_i} \ell_{i,k}$ then $|\hat{\mu}_{i,T_i} - \mu_i| \leq \sqrt{\frac{\log(4nT_i^2)}{2T_i}}$ for all $i \in [n]$ and $T_i > 0$ by applying Hoeffding's inequality and a union bound. In contrast, the non-stochastic setting's assumption that $\lim_{\tau \rightarrow \infty} \ell_{i,\tau}$ exists implies that there exists a non-increasing function γ_i such that $|\ell_{i,t} - \lim_{\tau \rightarrow \infty} \ell_{i,\tau}| \leq \gamma_i(t)$ and that $\lim_{t \rightarrow \infty} \gamma_i(t) = 0$. However, the existence of this limit tells us nothing about how *quickly* $\gamma_i(t)$ approaches 0. The lack of an explicit convergence rate as a function of t presents a problem as even the tightest $\gamma_i(t)$ could decay arbitrarily slowly and we would never know it.

This observation has two consequences. First, we can never *reject* the possibility that an arm is the “best” arm. Second, we can never *verify* that an arm is the “best” arm or even attain a value within ϵ of the best arm. Despite these challenges, in Section 3 we identify an effective algorithm under natural measures of performance, using ideas inspired by the fixed budget setting of the stochastic best arm problem (Karnin et al., 2013; Audibert & Bubeck, 2010; Gabilon et al., 2012).

2.1. Related work

Despite dating to back to the late 1950's, the best-arm identification problem for the stochastic setting has experienced a surge of activity in the last decade. The work has two major branches: the fixed budget setting and the fixed confidence setting. In the fixed budget setting, the algorithm is given a set of arms and a budget B and is tasked with maximizing the probability of identifying the

Exploration algorithm	# observed losses
Uniform (baseline) (B)	n
Successive Halving* (B)	$2n + 1$
Successive Rejects (B)	$(n + 1)n/2$
Successive Elimination (C)	$n \log_2(2B)$
LUCB (C), lil'UCB (C), EXP3 (R)	B

Table 1. The number of times an algorithm observes a loss in terms of budget B and number of arms n , where B is known to the algorithm. (B), (C), or (R) indicate whether the algorithm is of the fixed budget, fixed confidence, or cumulative regret variety, respectfully. (*) indicates the algorithm we propose for use in the non-stochastic best arm setting.

best arm by pulling arms without exceeding the total budget. While these algorithms were developed for and analyzed in the stochastic setting, they exhibit attributes that are very amenable to the non-stochastic setting. In fact, the algorithm we propose to use in this paper is exactly the Successive Halving algorithm of Karnin et al. (2013), though the non-stochastic setting requires its own novel analysis that we present in Section 3. Successive Rejects (Audibert & Bubeck, 2010) is another fixed budget algorithm that we compare to in our experiments.

The best-arm identification problem in the fixed confidence setting takes an input $\delta \in (0, 1)$ and guarantees to output the best arm with probability at least $1 - \delta$ while attempting to minimize the number of total arm pulls. These algorithms rely on probability theory to determine how many times each arm must be pulled in order to decide if the arm is suboptimal and should no longer be pulled, either by explicitly discarding it, e.g., Successive Elimination (Even-Dar et al., 2006) and Exponential Gap Elimination (Karnin et al., 2013), or implicitly by other methods, e.g., LUCB (Kalyanakrishnan et al., 2012) and Lil'UCB (Jamieson et al., 2014). Algorithms from the fixed confidence setting are ill-suited for the non-stochastic best-arm identification problem because they rely on statistical bounds that are generally not applicable in the non-stochastic case. These algorithms also exhibit some undesirable behavior with respect to how many losses they observe, which we explore next.

In addition to just the total number of arm pulls, this work also considers the required number of *observed* losses. This is a natural cost to consider when ℓ_{i,T_i} for any i is the result of doing some computation like evaluating a partially trained classifier on a hold-out validation set or releasing a product to the market to probe for demand. In some cases the cost, be it time, effort, or dollars, of an evaluation of the loss of an arm after some number of pulls can dwarf the cost of pulling the arm. Assuming a known time horizon (or budget), Table 1 describes the total number of times various algorithms observe a loss as a function of the budget B

and the number of arms n . We include in our comparison the EXP3 algorithm (Auer et al., 2002), a popular approach for minimizing cumulative regret in the non-stochastic setting. In practice $B \gg n$, and thus Successive Halving is a particular attractive option, as along with the baseline, it is the only algorithm that observes losses proportional to the number of arms and independent of the budget. As we will see in Section 5, the performance of these algorithms is quite dependent on the number of observed losses.

3. Proposed algorithm and analysis

The proposed Successive Halving algorithm of Figure 3 was originally proposed for the stochastic best arm identification problem in the fixed budget setting by (Karnin et al., 2013). However, our novel analysis in this work shows that it is also effective in the non-stochastic setting. The idea behind the algorithm is simple: given an input budget, uniformly allocate the budget to a set of arms for a predefined amount of iterations, evaluate their performance, throw out the worst half, and repeat until just one arm remains.

Successive Halving Algorithm

input: Budget B , n arms where $\ell_{i,k}$ denotes the k th loss from the i th arm

Initialize: $S_0 = [n]$.

For $k = 0, 1, \dots, \lceil \log_2(n) \rceil - 1$

 Pull each arm in S_k for $r_k = \lfloor \frac{B}{|S_k| \lceil \log_2(n) \rceil} \rfloor$ additional times and set $R_k = \sum_{j=0}^k r_j$.

 Let σ_k be a bijection on S_k such that $\ell_{\sigma_k(1), R_k} \leq \ell_{\sigma_k(2), R_k} \leq \dots \leq \ell_{\sigma_k(|S_k|), R_k}$

$S_{k+1} = \{i \in S_k : \ell_{\sigma_k(i), R_k} \leq \ell_{\sigma_k(\lfloor |S_k|/2 \rfloor), R_k}\}$.

output : Singleton element of $S_{\lceil \log_2(n) \rceil}$

Figure 3. Successive Halving was originally proposed for the stochastic best arm identification problem in Karnin et al. (2013) but is also applicable to the non-stochastic setting.

The budget as an input is easily removed by the “doubling trick” that attempts $B \leftarrow n$, then $B \leftarrow 2B$, and so on. This method can reuse existing progress from iteration to iteration and effectively makes the algorithm parameter free. But its most notable quality is that if a budget of B' is necessary to succeed in finding the best arm, by performing the doubling trick one will have only had to use a budget of $2B'$ in the worst case without ever having to know B' in the first place. Thus, for the remainder of this section we consider a fixed budget.

3.1. Analysis of Successive Halving

We first show that the algorithm never takes a total number of samples that exceeds the budget B :

$$\sum_{k=0}^{\lceil \log_2(n) \rceil - 1} |S_k| \left\lfloor \frac{B}{|S_k| \lceil \log_2(n) \rceil} \right\rfloor \leq \sum_{k=0}^{\lceil \log_2(n) \rceil - 1} \frac{B}{\lceil \log_2(n) \rceil} \leq B.$$

Next we consider how the algorithm performs in terms of identifying the best arm. First, for $i = 1, \dots, n$ define $\nu_i = \lim_{\tau \rightarrow \infty} \ell_{i,\tau}$ which exists by assumption. Without loss of generality, assume that

$$\nu_1 < \nu_2 \leq \dots \leq \nu_n.$$

We next introduce functions that bound the approximation error of $\ell_{i,t}$ with respect to ν_i as a function of t . For each $i = 1, 2, \dots, n$ let $\gamma_i(t)$ be the point-wise smallest, non-increasing function of t such that

$$|\ell_{i,t} - \nu_i| \leq \gamma_i(t) \quad \forall t.$$

In addition, define $\gamma_i^{-1}(\alpha) = \min\{t \in \mathbb{N} : \gamma_i(t) \leq \alpha\}$ for all $i \in [n]$. With this definition, if $t_i > \gamma_i^{-1}(\frac{\nu_i - \nu_1}{2})$ and $t_1 > \gamma_1^{-1}(\frac{\nu_i - \nu_1}{2})$ then

$$\begin{aligned} \ell_{i,t_i} - \ell_{1,t_1} &= (\ell_{i,t_i} - \nu_i) + (\nu_1 - \ell_{1,t_1}) + 2\left(\frac{\nu_i - \nu_1}{2}\right) \\ &\geq -\gamma_i(t_i) - \gamma_1(t_1) + 2\left(\frac{\nu_i - \nu_1}{2}\right) > 0. \end{aligned}$$

Indeed, if $\min\{t_i, t_1\} > \max\{\gamma_i^{-1}(\frac{\nu_i - \nu_1}{2}), \gamma_1^{-1}(\frac{\nu_i - \nu_1}{2})\}$ then we are guaranteed to have that $\ell_{i,t_i} > \ell_{1,t_1}$. That is, comparing the intermediate values at t_i and t_1 suffices to determine the ordering of the final values ν_i and ν_1 . Intuitively, this condition holds because the envelopes at the given times, namely $\gamma_i(t_i)$ and $\gamma_1(t_1)$, are small relative to the gap between ν_i and ν_1 . This line of reasoning is at the heart of the proof of our main result, and the theorem is stated in terms of these quantities. All proofs can be found in the appendix.

Theorem 1 Let $\nu_i = \lim_{\tau \rightarrow \infty} \ell_{i,\tau}$, $\bar{\gamma}(t) = \max_{i=1,\dots,n} \gamma_i(t)$ and

$$\begin{aligned} z &= 2\lceil \log_2(n) \rceil \max_{i=2,\dots,n} i \left(1 + \bar{\gamma}^{-1}\left(\frac{\nu_i - \nu_1}{2}\right)\right) \\ &\leq 2\lceil \log_2(n) \rceil \left(n + \sum_{i=2,\dots,n} \bar{\gamma}^{-1}\left(\frac{\nu_i - \nu_1}{2}\right)\right). \end{aligned}$$

If the budget $B > z$ then the best arm is returned from the algorithm.

The representation of z on the right-hand-side of the inequality is very intuitive: if $\bar{\gamma}(t) = \gamma_i(t) \quad \forall i$ and an oracle gave us an explicit form for $\bar{\gamma}(t)$, then to merely verify that the i th arm’s final value is higher than the best arm’s, one must pull each of the two arms at least a number of

times equal to the i th term in the sum (this becomes clear by inspecting the proof of Theorem 3). Repeating this argument for all $i = 2, \dots, n$ explains the sum over all $n - 1$ arms. While clearly not a proof, this argument along with known lower bounds for the stochastic setting (Audibert & Bubeck, 2010; Kaufmann et al., 2014), a subset of the non-stochastic setting, suggest that the above result may be nearly tight in a minimax sense up to log factors.

Example 1 Consider a feature-selection problem where you are given a dataset $\{(x_i, y_i)\}_{i=1}^n$ where each $x_i \in \mathbb{R}^D$ and you are tasked with identifying the best subset of features of size d that linearly predicts y_i in terms of the least-squares metric. In our framework, each d -subset is an arm and there are $n = \binom{D}{d}$ arms. Least squares is a convex quadratic optimization problem that can be efficiently solved with stochastic gradient descent. Using known bounds for the rates of convergence (Nemirovski et al., 2009) one can show that $\gamma_a(t) \leq \frac{\sigma_a \log(nt/\delta)}{t}$ for all $a = 1, \dots, n$ arms and all $t \geq 1$ with probability at least $1 - \delta$ where σ_a is a constant that depends on the condition number of the quadratic defined by the d -subset. Then in Theorem 1, $\bar{\gamma}(t) = \frac{\sigma_{\max} \log(nt/\delta)}{t}$ with $\sigma_{\max} = \max_{a=1, \dots, n} \sigma_a$ so after inverting $\bar{\gamma}$ we find that $z = 2\lceil \log_2(n) \rceil \max_{a=2, \dots, n} a \frac{4\sigma_{\max} \log\left(\frac{2n\sigma_{\max}}{\delta(\nu_a - \nu_1)}\right)}{\nu_a - \nu_1}$ is a sufficient budget to identify the best arm. Later we put this result in context by comparing to a baseline strategy.

In the above example we computed upper bounds on the γ_i functions in terms of problem dependent parameters to provide us with a sample complexity by plugging these values into our theorem. However, we stress that constructing tight bounds for the γ_i functions is very difficult outside of very simple problems like the one described above, and even then we have unspecified constants. Fortunately, because our algorithm is agnostic to these γ_i functions, it is also in some sense *adaptive* to them: the faster the arms' losses converge, the faster the best arm is discovered, without ever changing the algorithm. This behavior is in stark contrast to the hyperparameter tuning work of Swersky et al. (2014) and Agarwal et al. (2012), in which the algorithms explicitly take upper bounds on these γ_i functions as input, meaning the performance of the algorithm is only as good as the tightness of these difficult to calculate bounds.

3.2. Comparison to a uniform allocation strategy

We can also derive a result for the naive uniform budget allocation strategy. For simplicity, let B be a multiple of n so that at the end of the budget we have $T_i = B/n$ for all $i \in [n]$ and the output arm is equal to $\hat{i} = \arg \min_i \ell_{i, B/n}$.

Theorem 2 (Uniform strategy – sufficiency) Let $\nu_i =$

$$\lim_{\tau \rightarrow \infty} \ell_{i, \tau}, \bar{\gamma}(t) = \max_{i=1, \dots, n} \gamma_i(t) \text{ and}$$

$$z = \max_{i=2, \dots, n} n \bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right).$$

If $B > z$ then the uniform strategy returns the best arm.

Theorem 2 is just a sufficiency statement so it is unclear how the performance of the method actually compares to the Successive Halving result of Theorem 1. The next theorem says that the above result is tight in a worst-case sense, exposing the real gap between the algorithm of Figure 3 and the naive uniform allocation strategy.

Theorem 3 (Uniform strategy – necessity) For any given budget B and final values $\nu_1 < \nu_2 \leq \dots \leq \nu_n$ there exists a sequence of losses $\{\ell_{i, t}\}_{t=1}^\infty$, $i = 1, 2, \dots, n$ such that if

$$B < \max_{i=2, \dots, n} n \bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right)$$

then the uniform budget allocation strategy will not return the best arm.

If we consider the second, looser representation of z on the right-hand-side of the inequality in Theorem 1 and multiply this quantity by $\frac{n-1}{n-1}$ we see that the sufficient number of pulls for the Successive Halving algorithm essentially behaves like $(n-1) \log_2(n)$ times the average $\frac{1}{n-1} \sum_{i=2, \dots, n} \bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right)$ whereas the necessary result of the uniform allocation strategy of Theorem 3 behaves like n times the maximum $\max_{i=2, \dots, n} \bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right)$. The next example shows that the difference between this average and max can be very significant.

Example 2 Recall Example 1 and now assume that $\sigma_a = \sigma_{\max}$ for all $a = 1, \dots, n$. Then Theorem 3 says that the uniform allocation budget must be at least $n \frac{4\sigma_{\max} \log\left(\frac{2n\sigma_{\max}}{\delta(\nu_2 - \nu_1)}\right)}{\nu_2 - \nu_1}$ to identify the best arm. To see how this result compares with that of Successive Halving, let us parameterize the ν_a limiting values such that $\nu_a = a/n$ for $a = 1, \dots, n$. Then a sufficient budget for the Successive Halving algorithm to identify the best arm is just $8n\lceil \log_2(n) \rceil \sigma_{\max} \log\left(\frac{n^2 \sigma_{\max}}{\delta}\right)$ while the uniform allocation strategy would require a budget of at least $2n^2 \sigma_{\max} \log\left(\frac{n^2 \sigma_{\max}}{\delta}\right)$. This is a difference of essentially $4n \log_2(n)$ versus n^2 .

3.3. A pretty good arm

Up to this point we have been concerned with identifying the *best* arm: $\nu_1 = \arg \min_i \nu_i$ where we recall that $\nu_i = \lim_{\tau \rightarrow \infty} \ell_{i, \tau}$. But in practice one may be satisfied with merely an ϵ -good arm i_ϵ in the sense that $\nu_{i_\epsilon} - \nu_1 \leq \epsilon$. However, with our minimal assumptions, such a statement

is impossible to make since we have no knowledge of the γ_i functions to determine that an arm's final value is within ϵ of any value, much less the unknown final converged value of the best arm. However, as we show in Theorem 4, the Successive Halving algorithm cannot do much worse than the uniform allocation strategy.

Theorem 4 For a budget B and set of n arms, define \hat{i}_{SH} as the output of the Successive Halving algorithm. Then

$$\nu_{\hat{i}_{SH}} - \nu_1 \leq \lceil \log_2(n) \rceil 2\bar{\gamma} \left(\left\lfloor \frac{B}{n \lceil \log_2(n) \rceil} \right\rfloor \right).$$

Moreover, \hat{i}_U , the output of the uniform strategy, satisfies

$$\nu_{\hat{i}_U} - \nu_1 \leq \ell_{\hat{i}_U, B/n} - \ell_{1, B/n} + 2\bar{\gamma}(B/n) \leq 2\bar{\gamma}(B/n).$$

Example 3 Recall Example 1. Both the Successive Halving algorithm and the uniform allocation strategy satisfy $\hat{\nu}_i - \nu_1 \leq \tilde{O}(n/B)$ where \hat{i} is the output of either algorithm and \tilde{O} suppresses poly log factors.

We stress that this result is merely a fall-back guarantee, ensuring that we can never do much worse than uniform. However, it does not rule out the possibility of the Successive Halving algorithm far outperforming the uniform allocation strategy in practice. Indeed, we observe order of magnitude speed ups in our experimental results.

4. Hyperparameter optimization for supervised learning

In supervised learning we are given a dataset that is composed of pairs $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ for $i = 1, \dots, n$ sampled i.i.d. from some unknown joint distribution $P_{X,Y}$, and we are tasked with finding a map (or model) $f : \mathcal{X} \rightarrow \mathcal{Y}$ that minimizes $\mathbb{E}_{(X,Y) \sim P_{X,Y}} [\text{loss}(f(X), Y)]$ for some known loss function $\text{loss} : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$. Since $P_{X,Y}$ is unknown, we cannot compute $\mathbb{E}_{(X,Y) \sim P_{X,Y}} [\text{loss}(f(X), Y)]$ directly, but given m additional samples drawn i.i.d. from $P_{X,Y}$ we can approximate it with an empirical estimate, that is, $\frac{1}{m} \sum_{i=1}^m \text{loss}(f(x_i), y_i)$. We do not consider arbitrary mappings $\mathcal{X} \rightarrow \mathcal{Y}$ but only those that are the output of running a fixed, possibly randomized, algorithm \mathcal{A} that takes a dataset $\{(x_i, y_i)\}_{i=1}^n$ and algorithm-specific parameters $\theta \in \Theta$ as input so that for any θ we have $f_\theta = \mathcal{A}(\{(x_i, y_i)\}_{i=1}^n, \theta)$ where $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$. For a fixed dataset $\{(x_i, y_i)\}_{i=1}^n$ the parameters $\theta \in \Theta$ index the different functions f_θ , and will henceforth be referred to as *hyperparameters*. We adopt the train-validate-test framework for choosing hyperparameters (Hastie et al., 2005):

1. Partition the total dataset into TRAIN, VAL, and TEST sets with $\text{TRAIN} \cup \text{VAL} \cup \text{TEST} = \{(x_i, y_i)\}_{i=1}^m$.
2. Use TRAIN to train a model $f_\theta = \mathcal{A}(\{(x_i, y_i)\}_{i \in \text{TRAIN}}, \theta)$ for each $\theta \in \Theta$,

3. Choose the hyperparameters that minimize the empirical loss on the examples in VAL: $\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{|\text{VAL}|} \sum_{i \in \text{VAL}} \text{loss}(f_\theta(x_i), y_i)$
4. Report the empirical loss of $\hat{\theta}$ on the test error: $\frac{1}{|\text{TEST}|} \sum_{i \in \text{TEST}} \text{loss}(f_{\hat{\theta}}(x_i), y_i)$.

Example 4 Consider a linear classification example where $\mathcal{X} \times \mathcal{Y} = \mathbb{R}^d \times \{-1, 1\}$, $\Theta \subset \mathbb{R}_+$, $f_\theta = \mathcal{A}(\{(x_i, y_i)\}_{i \in \text{TRAIN}}, \theta)$ where $f_\theta(x) = \langle w_\theta, x \rangle$ with $w_\theta = \arg \min_w \frac{1}{|\text{TRAIN}|} \sum_{i \in \text{TRAIN}} \max(0, 1 - y_i \langle w, x_i \rangle) + \theta \|w\|_2^2$, and finally $\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{|\text{VAL}|} \sum_{i \in \text{VAL}} \mathbf{1}\{y f_\theta(x) < 0\}$.

In the simple above example involving a single hyperparameter, we emphasize that for each θ we have that f_θ can be efficiently computed using an iterative algorithm (Shalev-Shwartz et al., 2011), however, the selection of \hat{f} is the minimization of a function that is not necessarily even continuous, much less convex. This pattern is more often the rule than the exception. We next attempt to generalize and exploit this observation.

4.1. Posing as a best arm non-stochastic bandits problem

Let us assume that the algorithm \mathcal{A} is iterative so that for a given $\{(x_i, y_i)\}_{i \in \text{TRAIN}}$ and θ , the algorithm outputs a function $f_{\theta,t}$ every iteration $t > 1$ and we may compute

$$\ell_{\theta,t} = \frac{1}{|\text{VAL}|} \sum_{i \in \text{VAL}} \text{loss}(f_{\theta,t}(x_i), y_i).$$

We assume that the limit $\lim_{t \rightarrow \infty} \ell_{\theta,t}$ exists¹ and is equal to $\frac{1}{|\text{VAL}|} \sum_{i \in \text{VAL}} \text{loss}(f_\theta(x_i), y_i)$.

With this transformation we are in the position to put the hyperparameter optimization problem into the framework of Figure 2 and, namely, the non-stochastic best-arm identification formulation developed in the above sections. We generate the arms (different hyperparameter settings) uniformly at random (possibly on a log scale) from within the region of valid hyperparameters (i.e. all hyperparameters within some minimum and maximum ranges) and sample enough arms to ensure a sufficient cover of the space (Bergstra & Bengio, 2012). Alternatively, one could input a uniform grid over the parameters of interest. We note that random search and grid search remain the default choices for many open source machine learning packages

¹We note that $f_\theta = \lim_{t \rightarrow \infty} f_{\theta,t}$ is not enough to conclude that $\lim_{t \rightarrow \infty} \ell_{\theta,t}$ exists (for instance, for classification with 0/1 loss this is not necessarily true) but these technical issues can usually be surped for real datasets and losses (for instance, by replacing $\mathbf{1}\{z < 0\}$ with a very steep sigmoid). We ignore this technicality in our experiments.

such as LibSVM (Chang & Lin, 2011), scikit-learn (Pedregosa et al., 2011) and MLlib (Kraska et al., 2013). As described in Figure 2, the bandit algorithm will choose I_t , and we will use the convention that $J_t = \arg \min_{\theta} \ell_{\theta, T_{\theta}}$. The arm selected by J_t will be evaluated on the test set following the work-flow introduced above.

4.2. Related work

We aim to leverage the iterative nature of standard machine learning algorithms to speed up hyperparameter optimization in a robust and principled fashion. We now review related work in the context of our results. In Section 3.3 we show that no algorithm can provably identify a hyperparameter with a value within ϵ of the optimal without known, explicit functions γ_i , which means no algorithm can reject a hyperparameter setting with absolute confidence without making potentially unrealistic assumptions. Swersky et al. (2014) explicitly defines the γ_i functions in an ad-hoc, algorithm-specific, and data-specific fashion which leads to strong ϵ -good claims. A related line of work explicitly defines γ_i -like functions for optimizing the computational efficiency of structural risk minimization, yielding bounds (Agarwal et al., 2012). We stress that these results are only as good as the tightness and correctness of the γ_i bounds, and we view our work as an empirical, data-driven driven approach to the pursuits of Agarwal et al. (2012). Also, Sparks et al. (2015) empirically studies an early stopping heuristic for hyperparameter optimization similar in spirit to the Successive Halving algorithm.

We further note that we fix the hyperparameter settings (or arms) under consideration and adaptively allocate our budget to each arm. In contrast, Bayesian optimization advocates choosing hyperparameter settings adaptively, but with the exception of Swersky et al. (2014), allocates a fixed budget to each selected hyperparameter setting (Snoek et al., 2012; 2014; Hutter et al., 2011; Bergstra et al., 2011; Bergstra & Bengio, 2012). These Bayesian optimization methods, though heuristic in nature as they attempt to simultaneously fit and optimize a non-convex and potentially high-dimensional function, yield promising empirical results. We view our approach as complementary and orthogonal to the method used for choosing hyperparameter settings, and extending our approach in a principled fashion to adaptively choose arms, e.g., in a mini-batch setting, is an interesting avenue for future work.

5. Experiment results

In this section we compare the proposed algorithm to a number of other algorithms, including the baseline uniform allocation strategy, on a number of supervised learning hyperparameter optimization problems using the experimental setup outlined in Section 4.1. Each experiment was im-

plemented in Python and run in parallel using the multiprocessing library on an Amazon EC2 c3.8xlarge instance with 32 cores and 60 GB of memory. In all cases, full datasets were partitioned into a training-base dataset and a test (TEST) dataset with a 90/10 split. The training-base dataset was then partitioned into a training (TRAIN) and validation (VAL) datasets with an 80/20 split. All plots report loss on the test error.

To evaluate the different search algorithms’ performance, we fix a total budget of iterations and allow the search algorithms to decide how to divide it up amongst the different arms. The curves are produced by implementing the doubling trick by simply doubling the measurement budget each time. For the purpose of interpretability, we reset all iteration counters to 0 at each doubling of the budget, i.e., we do not warm start upon doubling. All datasets, aside from the collaborative filtering experiments, are normalized so that each dimension has mean 0 and variance 1.

5.1. Ridge regression

We first consider a ridge regression problem trained with stochastic gradient descent on this objective function with step size $.01/\sqrt{2 + T_{\lambda}}$. The ℓ_2 penalty hyperparameter $\lambda \in [10^{-6}, 10^0]$ was chosen **uniformly at random on a log scale per trial**, with 10 values (i.e., arms) selected per trial. We use the Million Song Dataset year prediction task (Lichman, 2013) where we have down sampled the dataset by a factor of 10 and normalized the years such that they are mean zero and variance 1 with respect to the training set. The experiment was repeated for 32 trials. Error on the VAL and TEST was calculated using mean-squared-error. In the left panel of Figure 4 we note that LUCB, lil’UCB perform the best in the sense that they achieve a small test error two to four times faster, in terms of iterations, than most other methods. However, in the right panel the same data is plotted but with respect to wall-clock time rather than iterations and we now observe that Successive Halving and Successive Rejects are the top performers. This is explainable by Table 1: EXP3, lil’UCB, and LUCB must evaluate the validation loss on every iteration requiring much greater compute time. This pattern is observed in all experiments so in the sequel we only consider the uniform allocation, Successive Halving, and Successive Rejects algorithm.

5.2. Kernel SVM

We now consider learning a kernel SVM using the RBF kernel $\kappa_{\gamma}(x, z) = e^{-\gamma||x-z||_2^2}$. The SVM is trained using Pegasos (Shalev-Shwartz et al., 2011) with ℓ_2 penalty hyperparameter $\lambda \in [10^{-6}, 10^0]$ and kernel width $\gamma \in [10^0, 10^3]$ both chosen uniformly at random on a log scale per trial. Each hyperparameter was allocated 10 samples resulting in $10^2 = 100$ total arms. The experiment was

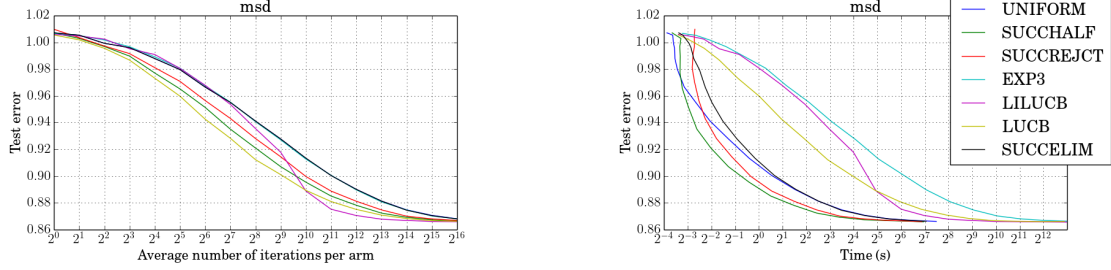


Figure 4. Ridge Regression. Test error with respect to both the number of iterations (left) and wall-clock time (right). Note that in the left plot, uniform, EXP3, and Successive Elimination are plotted on top of each other.

repeated for 64 trials. Error on the VAL and TEST was calculated using 0/1 loss. Kernel evaluations were computed online (i.e. not precomputed and stored). We observe in Figure 5 that Successive Halving obtains the same low error more than an order of magnitude faster than both uniform and Successive Rejects with respect to wall-clock time, despite Successive Halving and Success Rejects performing comparably in terms of iterations (not plotted).

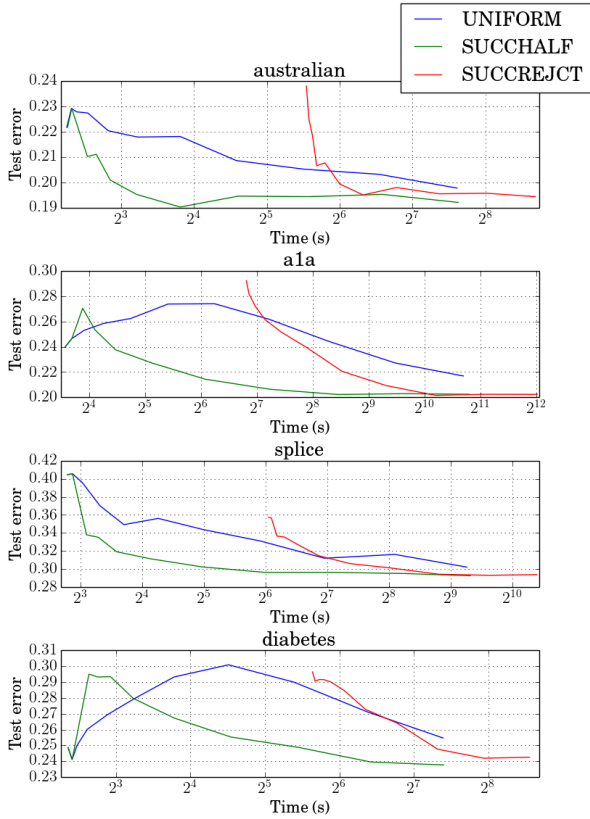


Figure 5. Kernel SVM. Successive Halving and Successive Rejects are separated by an order of magnitude in wall-clock time.

5.3. Collaborative filtering

We next consider a matrix completion problem using the Movielens 100k dataset trained using stochastic gradient

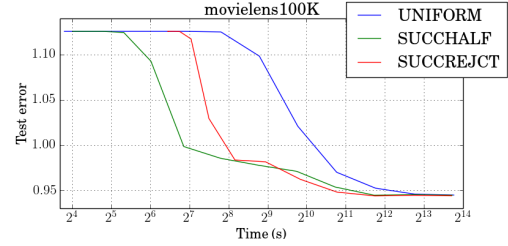


Figure 6. Matrix Completion (bi-convex formulation).

descent on the bi-convex objective with step sizes as described in Recht & Ré (2013). To account for the non-convex objective, we initialize the user and item variables with entries drawn from a normal distribution with variance σ^2/d , hence each arm has hyperparameters d (rank), λ (Frobenium norm regularization), and σ (initial conditions). $d \in [2, 50]$ and $\sigma \in [.01, 3]$ were chosen uniformly at random from a linear scale, and $\lambda \in [10^{-6}, 10^0]$ was chosen uniformly at random on a log scale. Each hyperparameter is given 4 samples resulting in $4^3 = 64$ total arms. The experiment was repeated for 32 trials. Error on the VAL and TEST was calculated using mean-squared-error. One observes in Figure 6 that the uniform allocation takes two to eight times longer to achieve a particular error rate than Successive Halving or Successive Rejects.

6. Future directions

Our theoretical results are presented in terms of $\max_i \gamma_i(t)$. An interesting future direction is to consider algorithms and analyses that take into account the specific convergence rates $\gamma_i(t)$ of each arm, analogous to considering arms with different variances in the stochastic case (Kaufmann et al., 2014). Incorporating pairwise switching costs into the framework could model the time of moving very large intermediate models in and out of memory to perform iterations, along with the degree to which resources are shared across various models (resulting in lower switching costs). Finally, balancing solution quality and time by adaptively sampling hyperparameters as is done in Bayesian methods is of considerable practical interest.

References

- Agarwal, Alekh, Bartlett, Peter, and Duchi, John. Oracle inequalities for computationally adaptive model selection. *arXiv preprint arXiv:1208.0129*, 2012.
- Audibert, Jean-Yves and Bubeck, Sébastien. Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pp. 13–p, 2010.
- Auer, Peter, Cesa-Bianchi, Nicolo, Freund, Yoav, and Schapire, Robert E. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.
- Bergstra, James and Bengio, Yoshua. Random search for hyper-parameter optimization. *JMLR*, 2012.
- Bergstra, James, Bardenet, Rémi, Bengio, Yoshua, and Kégl, Balázs. Algorithms for Hyper-Parameter Optimization. *NIPS*, 2011.
- Bubeck, Sébastien and Cesa-Bianchi, Nicolo. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.
- Bubeck, Sébastien, Munos, Rémi, and Stoltz, Gilles. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, pp. 23–37. Springer, 2009.
- Chang, Chih-Chung and Lin, Chih-Jen. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 2011.
- Cicirello, Vincent A and Smith, Stephen F. The max k-armed bandit: A new model of exploration applied to search heuristic selection. In *Proceedings of the National Conference on Artificial Intelligence*, volume 20, pp. 1355. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.
- Even-Dar, Eyal, Mannor, Shie, and Mansour, Yishay. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *The Journal of Machine Learning Research*, 7:1079–1105, 2006.
- Gabillon, Victor, Ghavamzadeh, Mohammad, and Lazaric, Alessandro. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems*, pp. 3212–3220, 2012.
- Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, and Franklin, James. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.
- Hutter, Frank, Hoos, Holger H, and Leyton-Brown, Kevin. Sequential Model-Based Optimization for General Algorithm Configuration. pp. 507–523, 2011.
- Jamieson, Kevin, Malloy, Matthew, Nowak, Robert, and Bubeck, Sébastien. lil’ucb: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of The 27th Conference on Learning Theory*, pp. 423–439, 2014.
- Kalyanakrishnan, Shivaram, Tewari, Ambuj, Auer, Peter, and Stone, Peter. Pac subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pp. 655–662, 2012.
- Karnin, Zohar, Koren, Tomer, and Somekh, Oren. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pp. 1238–1246, 2013.
- Kaufmann, Emilie, Cappé, Olivier, and Garivier, Aurélien. On the complexity of best arm identification in multi-armed bandit models. *arXiv preprint arXiv:1407.4443*, 2014.
- Kraska, Tim, Talwalkar, Ameet, Duchi, John, Griffith, Rean, Franklin, Michael, and Jordan, Michael. MLbase: A Distributed Machine-learning System. In *CIDR*, 2013.
- Lichman, M. UCI machine learning repository, 2013. URL <http://archive.ics.uci.edu/ml>.
- Nemirovski, Arkadi, Juditsky, Anatoli, Lan, Guanghui, and Shapiro, Alexander. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.
- Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830, 2011.
- Recht, Benjamin and Ré, Christopher. Parallel stochastic gradient algorithms for large-scale matrix completion. *Mathematical Programming Computation*, 5(2): 201–226, 2013.
- Shalev-Shwartz, Shai, Singer, Yoram, Srebro, Nathan, and Cotter, Andrew. Pegasos: Primal estimated sub-gradient solver for svm. *Mathematical programming*, 127(1):3–30, 2011.
- Snoek, Jasper, Larochelle, Hugo, and Adams, Ryan. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, 2012.

Snoek, Jasper, Swersky, Kevin, Zemel, Richard, and Adams, Ryan. Input warping for bayesian optimization of non-stationary functions. In *International Conference on Machine Learning*, 2014.

Sparks, Evan R, Talwalkar, Ameet, Franklin, Michael J., Jordan, Michael I., and Kraska, Tim. TuPAQ: An efficient planner for large-scale predictive analytic queries. *arXiv preprint arXiv:1502.00068*, 2015.

Swersky, Kevin, Snoek, Jasper, and Adams, Ryan Prescott. Freeze-thaw bayesian optimization. *arXiv preprint arXiv:1406.3896*, 2014.

A. Proof of Theorem 1

Proof For notational ease, define $[\cdot] = \{\{\cdot\}_{t=1}^n\}_{i=1}^n$ so that $[\ell_{i,t}] = \{\{\ell_{i,t}\}_{t=1}^\infty\}_{i=1}^n$. Without loss of generality, we may assume that the n infinitely long loss sequences $[\ell_{i,t}]$ with limits $\{\nu_i\}_{i=1}^n$ were fixed prior to the start of the game so that the $\gamma_i(t)$ envelopes are also defined for all time and are fixed. Let Ω be the set that contains all possible sets of n infinitely long sequences of real numbers with limits $\{\nu_i\}_{i=1}^n$ and envelopes $[\bar{\gamma}(t)]$, that is,

$$\Omega = \left\{ [\ell'_{i,t}] : [|\ell'_{i,t} - \nu_i| \leq \bar{\gamma}(t)] \wedge \lim_{\tau \rightarrow \infty} \ell'_{i,\tau} = \nu_i \quad \forall i \right\}$$

where we recall that \wedge is read as “and” and \vee is read as “or.” Clearly, $[\ell_{i,t}]$ is a single element of Ω .

We present a proof by contradiction. We begin by considering the singleton set containing $[\ell_{i,t}]$ under the assumption that the Successive Halving algorithm fails to identify the best arm, i.e., $S_{\lceil \log_2(n) \rceil} \neq 1$. We then consider a sequence of subsets of Ω , with each one contained in the next. The proof is completed by showing that the final subset in our sequence (and thus our original singleton set of interest) is empty when $B > z$, which contradicts our assumption and proves the statement of our theorem.

To reduce clutter in the following arguments, it is understood that S'_k for all k in the following sets is a function of $[\ell'_{i,t}]$ in the sense that it is the state of S_k in the algorithm when it is run with losses $[\ell'_{i,t}]$. We now present our argument in detail, starting with the singleton set of interest, and using the definition of S_k in Figure 3.

$$\begin{aligned} & \left\{ [\ell'_{i,t}] \in \Omega : [\ell'_{i,t} = \ell_{i,t}] \wedge S'_{\lceil \log_2(n) \rceil} \neq 1 \right\} \\ &= \left\{ [\ell'_{i,t}] \in \Omega : [\ell'_{i,t} = \ell_{i,t}] \wedge \bigvee_{k=1}^{\lceil \log_2(n) \rceil} \{1 \notin S'_k, 1 \in S'_{k-1}\} \right\} \\ &= \left\{ [\ell'_{i,t}] \in \Omega : [\ell'_{i,t} = \ell_{i,t}] \wedge \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ \sum_{i \in S'_k} \mathbf{1}\{\ell'_{i,R_k} < \ell'_{1,R_k}\} > \lfloor |S'_k|/2 \rfloor \right\} \right\} \\ &= \left\{ [\ell'_{i,t}] \in \Omega : [\ell'_{i,t} = \ell_{i,t}] \wedge \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ \sum_{i \in S'_k} \mathbf{1}\{\nu_i - \nu_1 < \ell'_{1,R_k} - \nu_1 - \ell'_{i,R_k} + \nu_i\} > \lfloor |S'_k|/2 \rfloor \right\} \right\} \\ &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ \sum_{i \in S'_k} \mathbf{1}\{\nu_i - \nu_1 < |\ell'_{1,R_k} - \nu_1| + |\ell'_{i,R_k} - \nu_i|\} > \lfloor |S'_k|/2 \rfloor \right\} \right\} \\ &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ \sum_{i \in S'_k} \mathbf{1}\{2\bar{\gamma}(R_k) > \nu_i - \nu_1\} > \lfloor |S'_k|/2 \rfloor \right\} \right\}, \end{aligned} \quad (1)$$

where the last set relaxes the original equality condition to just considering the maximum envelope $\bar{\gamma}$ that is encoded in Ω . The summation in Eq. 1 only involves the ν_i , and this summand is maximized if each S'_k contains the first $|S'_k|$ arms. Hence we have,

$$\begin{aligned} (1) &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ \sum_{i=1}^{|S'_k|} \mathbf{1}\{2\bar{\gamma}(R_k) > \nu_i - \nu_1\} > \lfloor |S'_k|/2 \rfloor \right\} \right\} \\ &= \left\{ [\ell'_{i,t}] \in \Omega : \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ 2\bar{\gamma}(R_k) > \nu_{\lfloor |S'_k|/2 \rfloor + 1} - \nu_1 \right\} \right\} \\ &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ R_k < \bar{\gamma}^{-1} \left(\frac{\nu_{\lfloor |S'_k|/2 \rfloor + 1} - \nu_1}{2} \right) \right\} \right\}, \end{aligned} \quad (2)$$

where we use the definition of $\bar{\gamma}^{-1}$ in Eq. 2. Next, we recall that $R_k = \sum_{j=0}^k \lfloor \frac{B}{|S_k| \lceil \log_2(n) \rceil} \rfloor \geq \frac{B/2}{(\lfloor |S_k|/2 \rfloor + 1) \lceil \log_2(n) \rceil} - 1$ since $|S_k| \leq 2(\lfloor |S_k|/2 \rfloor + 1)$. We note that we are underestimating by almost a factor of 2 to account for integer effects in

favor of a simpler form. By plugging in this value for R_k and rearranging we have that

$$\begin{aligned}
 (2) &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : \bigvee_{k=0}^{\lceil \log_2(n) \rceil - 1} \left\{ \frac{B/2}{\lceil \log_2(n) \rceil} < (\lfloor |S'_k|/2 \rfloor + 1)(1 + \bar{\gamma}^{-1} \left(\frac{\nu_{\lfloor |S'_k|/2 \rfloor + 1} - \nu_1}{2} \right)) \right\} \right\} \\
 &= \left\{ [\ell'_{i,t}] \in \Omega : \frac{B/2}{\lceil \log_2(n) \rceil} < \max_{k=0, \dots, \lceil \log_2(n) \rceil - 1} (\lfloor |S'_k|/2 \rfloor + 1)(1 + \bar{\gamma}^{-1} \left(\frac{\nu_{\lfloor |S'_k|/2 \rfloor + 1} - \nu_1}{2} \right)) \right\} \\
 &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : B < 2 \lceil \log_2(n) \rceil \max_{i=2, \dots, n} i (\bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right) + 1) \right\} = \emptyset
 \end{aligned}$$

where the last equality holds if $B > z$.

The second, looser, but perhaps more interpretable form of z is thanks to (Audibert & Bubeck, 2010) who showed that

$$\max_{i=2, \dots, n} i \bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right) \leq \sum_{i=2, \dots, n} \bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right) \leq \log_2(2n) \max_{i=2, \dots, n} i \bar{\gamma}^{-1} \left(\frac{\nu_i - \nu_1}{2} \right)$$

where both inequalities are achievable with particular settings of the ν_i variables. ■

B. Proof of Theorem 2

Proof Recall the notation from the proof of Theorem 1 and let $\hat{i}([\ell'_{i,t}])$ be the output of the uniform allocation strategy with input losses $[\ell'_{i,t}]$.

$$\begin{aligned}
 \left\{ [\ell'_{i,t}] \in \Omega : [\ell'_{i,t}] = \ell_{i,t} \wedge \hat{i}([\ell'_{i,t}]) \neq 1 \right\} &= \left\{ [\ell'_{i,t}] \in \Omega : [\ell'_{i,t}] = \ell_{i,t} \wedge \ell'_{1,B/n} \geq \min_{i=2, \dots, n} \ell'_{i,B/n} \right\} \\
 &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : 2\bar{\gamma}(B/n) \geq \min_{i=2, \dots, n} \nu_i - \nu_1 \right\} \\
 &= \left\{ [\ell'_{i,t}] \in \Omega : 2\bar{\gamma}(B/n) \geq \nu_2 - \nu_1 \right\} \\
 &\subseteq \left\{ [\ell'_{i,t}] \in \Omega : B \leq n\bar{\gamma}^{-1} \left(\frac{\nu_2 - \nu_1}{2} \right) \right\} = \emptyset
 \end{aligned}$$

where the last equality follows from the fact that $B > z$ which implies $\hat{i}([\ell_{i,t}]) = 1$. ■

C. Proof of Theorem 3

Proof Let $\beta(t)$ be an arbitrary, monotonically decreasing function of t with $\lim_{t \rightarrow \infty} \beta(t) = 0$. Define $\ell_{1,t} = \nu_1 + \beta(t)$ and $\ell_{i,t} = \nu_i - \beta(t)$ for all i . Note that for all i , $\gamma_i(t) = \bar{\gamma}(t) = \beta(t)$ so that

$$\begin{aligned}
 \hat{i} = 1 &\iff \ell_{1,B/n} < \min_{i=2, \dots, n} \ell_{i,B/n} \\
 &\iff \nu_1 + \bar{\gamma}(B/n) < \min_{i=2, \dots, n} \nu_i - \bar{\gamma}(B/n) \\
 &\iff \nu_1 + \bar{\gamma}(B/n) < \nu_2 - \bar{\gamma}(B/n) \\
 &\iff \bar{\gamma}(B/n) < \frac{\nu_2 - \nu_1}{2} \\
 &\iff B \geq n\bar{\gamma}^{-1} \left(\frac{\nu_2 - \nu_1}{2} \right).
 \end{aligned}$$
■

D. Proof of Theorem 4

We can guarantee for the Successive Halving algorithm of Figure 3 that the output arm \hat{i} satisfies

$$\begin{aligned}
 \nu_{\hat{i}} - \nu_1 &= \min_{i \in S_{\lceil \log_2(n) \rceil}} \nu_i - \nu_1 \\
 &= \sum_{k=0}^{\lceil \log_2(n) \rceil - 1} \min_{i \in S_{k+1}} \nu_i - \min_{i \in S_k} \nu_i \\
 &\leq \sum_{k=0}^{\lceil \log_2(n) \rceil - 1} \min_{i \in S_{k+1}} \ell_{i, R_k} - \min_{i \in S_k} \ell_{i, R_k} + 2\bar{\gamma}(R_k) \\
 &= \sum_{k=0}^{\lceil \log_2(n) \rceil - 1} 2\bar{\gamma}(R_k) \leq \lceil \log_2(n) \rceil 2\bar{\gamma} \left(\lfloor \frac{B}{n^{\lceil \log_2(n) \rceil}} \rfloor \right)
 \end{aligned}$$

simply by inspecting how the algorithm eliminates arms and plugging in a trivial lower bound for R_k for all k in the last step.