


(Near) Optimal Adaptivity Gaps for Stochastic Multi-Value Probing

Domagoj Bradac 

Department of Mathematics, Faculty of Science, University of Zagreb, Croatia
domagoj.bradac@gmail.com

Sahil Singla 

Department of Computer Science, Princeton University¹, NJ, USA
singla@cs.princeton.edu

Goran Zuzic 

Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
gzuzic@cs.cmu.edu

Abstract

Consider a kidney-exchange application where we want to find a max-matching in a random graph. To find whether an edge e exists, we need to perform an expensive test, in which case the edge e appears independently with a *known* probability p_e . Given a budget on the total cost of the tests, our goal is to find a testing strategy that maximizes the expected maximum matching size.

The above application is an example of the stochastic probing problem. In general the optimal stochastic probing strategy is difficult to find because it is *adaptive* – decides on the next edge to probe based on the outcomes of the probed edges. An alternate approach is to show the *adaptivity gap* is small, i.e., the best *non-adaptive* strategy always has a value close to the best adaptive strategy. This allows us to focus on designing non-adaptive strategies that are much simpler. Previous works, however, have focused on Bernoulli random variables that can only capture whether an edge appears or not. In this work we introduce a multi-value stochastic probing problem, which can also model situations where the weight of an edge has a probability distribution over multiple values.

Our main technical contribution is to obtain (near) optimal bounds for the (worst-case) adaptivity gaps for multi-value stochastic probing over prefix-closed constraints. For a monotone submodular function, we show the adaptivity gap is at most 2 and provide a matching lower bound. For a weighted rank function of a k -extendible system (a generalization of intersection of k matroids), we show the adaptivity gap is between $O(k \log k)$ and k . None of these results were known even in the Bernoulli case where both our upper and lower bounds also apply, thereby resolving an open question of Gupta et al. [23].

2012 ACM Subject Classification Theory of computation → Stochastic control and optimization; Theory of computation → Submodular optimization and polymatroids; Theory of computation → Approximation algorithms analysis; Theory of computation → Design and analysis of algorithms

Keywords and phrases stochastic programming, adaptivity gaps, stochastic multi-value probing, submodular functions, k -extendible systems, adaptive strategy, matroid intersection

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.49

Category RANDOM

Related Version <https://arxiv.org/abs/1902.01461>

Funding *Domagoj Bradac*: The Author thanks the Computer Science Department at Carnegie Mellon University for their support; part of his work was done with visiting in Summer 2018.

Sahil Singla: Supported in part by NSF awards CCF-1319811, CCF-1536002, and CCF-1617790 and in part by the Schmidt foundation.

Goran Zuzic: Supported in part by NSF grants CCF-1527110, CCF-1618280, CCF-1814603, CCF-1910588, NSF CAREER award CCF-1750808 and a Sloan Research Fellowship.

Acknowledgements We would like to thank Anupam Gupta for helpful discussions.

¹ Most of this work was done when the author was a graduate student at Carnegie Mellon University.



© Domagoj Bradac, Sahil Singla, and Goran Zuzic;
licensed under Creative Commons License CC-BY

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019).

Editors: Dimitris Achlioptas and László A. Végh; Article No. 49; pp. 49:1–49:21



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Consider a kidney-exchange application where we want to find a maximum matching in a random graph. To find whether an edge e exists, we need to perform an expensive test, in which case the edge e appears independently with a known probability p_e . Given a budget on the total cost of the tests, our goal is to design a testing strategy that maximizes the expected size of the found matching.

The above application can be modeled as a constrained *stochastic probing* problem [5, 21, 3, 22, 23]. In this setting, we are given a universe V of *elements* (e.g., the set of all possible edges), each with an *activation* probability p_v for $v \in V$ (e.g., the probability an edge exists). We define a random set $A \subseteq V$ of *active* elements that contains every v independently with probability p_v . A *probe* at v reveals whether $v \in A$ or $v \notin A$, and we are only allowed to probe certain *feasible subsets* $S \in \mathcal{F} \subseteq 2^V$ (e.g., subsets of edges whose tests fit in our budget). Our goal is to design a *probing strategy* to find a feasible set $S \in \mathcal{F}$ of elements to maximize $\mathbb{E}_A[f(A \cap S)]$, where f is some combinatorial function $f : 2^V \rightarrow \mathbb{R}_{\geq 0}$ (e.g., the cardinality of the maximum matching). Notice our probing strategy could be *adaptive*, i.e., we could decide which element to probe next based on the outcomes of already probed elements.

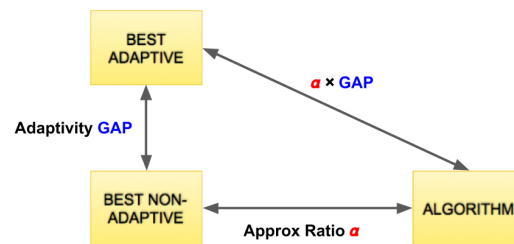
Besides matching [13, 7], stochastic probing has applications for stochastic variants of several other combinatorial problems. E.g., it can be used for Bayesian mechanism design problems [21], robot path-planning problems [22, 23], and stochastic set cover problems that arise in database applications [27, 16]. As observed in these prior works, the optimal strategy for stochastic probing can be represented as a binary *decision tree* where each node represents an element of V : You first probe the root node element, and then depending on whether it is active or inactive, you either move to the right or the left subtree. In general, such an optimal decision tree can be exponentially sized and is hard to describe. We do not even understand how to capture it for very simple functions and constraints (e.g., the max function with cardinality constraints [24]).

An alternate approach is to focus on *non-adaptive* strategies. Such a strategy commits to probing a feasible set $S \in \mathcal{F}$ in the beginning, irrespective of which of these elements turn out active. A non-adaptive strategy has several benefits: (a) it is easy to represent since we can just store the set S , (b) it is easy to find for many classes of functions and constraints (e.g., submodular functions over intersection of matroids [12]), and (c) it is parallelizable because we do not need feedback. The concern is that the expected value of the optimal non-adaptive strategy might be much smaller than that of the optimal adaptive strategy. This raises the (worst-case instance) *adaptivity gap* question: What is the maximum ratio between the expected values of the optimal adaptive and the optimal non-adaptive strategies for stochastic probing? If this ratio is small then we can focus on non-adaptive strategies and reap its benefits with only a small loss in value (see Figure 1).

Since for general combinatorial functions or constraints the adaptivity gaps can be made arbitrarily large, we need to consider special classes of functions and constraints. In a surprising result, Gupta et al. prove that for any *monotone submodular* function and any *prefix-closed constraints*², the adaptivity gap is at most 3 [23]. The best known lower bound in this setting, however, is only $\frac{e}{e-1} \approx 1.58$ due to Asadpour et al. [5]. This leaves open the following question:

For stochastic probing, what is the (worst-case) adaptivity gap for monotone submodular functions over prefix-closed constraints?

² Prefix-closed constraints stipulate that any prefix of a feasible probing sequence is also feasible. This class contains any downward-closed/packing constraint.



■ **Figure 1** An α -approximation to the best non-adaptive solution implies an $(\alpha \cdot \text{GAP})$ -approximation to the best adaptive algorithm, where GAP is the adaptivity gap.

We show that both the previously known upper bound of 3 and the lower bound of $\frac{e}{e-1}$ are not tight. Instead, the adaptivity gap is exactly 2.

One might notice that submodular functions do not capture the max-matching function used to model kidney-exchanges. This motivates us to consider more general combinatorial functions; in particular, we study the weighted rank function of a k -extendible system (defined in §2). This class generalizes intersection of k -matroids [29], e.g., a 2-extendible system captures matching in general graphs (unlike intersections of two matroids). Our goal is to bound the adaptivity gap for such functions over arbitrary prefix-closed constraints.

A major drawback of the stochastic probing model is that it only considers Bernoulli random variables. One would ideally allow for more modeling power by permitting the outcome of a probe to be a non-binary value. For example, in the kidney-exchange application, one might desire to summarize an edge probe by the risk involved in performing the match: a value of 0 describes an impossible match, a value of 1 indicates a safe match, and the possibilities in between are represented by intermediate values. Notice that the optimal adaptive strategy is still a decision tree; however, it may no longer be binary.

The main contributions of this paper are (1) a model that extends the binary stochastic probing to the multi-value setting, (2) the exact calculation of the adaptivity gap for stochastic probing of monotone submodular functions (in both the binary and multi-value setting), and (3) a nearly-tight adaptivity gap for stochastic probing of weighted rank functions over k -extendible systems.

1.1 Overview of Results

Our conceptual contribution is to present a generalization of the stochastic probing model to *stochastic multi-value probing* (SMP) described in §2. Roughly, the idea is that each element has t potential types, and a probe reveals which one of its types it takes. This trivially captures stochastic probing for $t = 2$, where the two types are active and inactive. In general these different types can be used to model different weights of an element, or to even encode different kinds of complementary relationships in the element values.

Although the SMP model is more general than the stochastic probing model, our main technical result in §3 is that for monotone submodular functions the adaptivity gap is bounded by 2. We also give a matching lower bound which proves this cannot be further reduced. This is despite the fact that the optimal decision tree for SMP may no longer be binary.

► **Theorem 1.** *The adaptivity gap for SMP where the constraints are prefix-closed and the function is monotone non-negative submodular is exactly 2.*

Since SMP is strictly more general than stochastic probing, Theorem 1 also improves the previously known upper bound of 3 for monotone submodular stochastic probing. In fact, our lower bound SMP instance in Theorem 1 is Bernoulli. Thus it resolves an open question of [23] of finding the optimal adaptivity gaps for submodular stochastic probing.

Our main technical result in §4 is that the adaptivity gap for weighted rank function of a k -extendible system is $\tilde{\Theta}(k)$.

► **Theorem 2.** *The adaptivity gap for SMP where the constraints are prefix-closed and the function is a weighted rank function of a k -extendible system is between k and $O(k \log k)$. Moreover, for unweighted rank functions, the adaptivity gap is between k and $2k$.*

Since the weighted rank of function of intersection of k -matroids is a k -extendible system, Theorem 2 implies as a corollary that the adaptivity gaps for this class is at most $\tilde{\Theta}(k)$. This improves the previously best known upper bound for intersection of k matroids of $O(k^4 \cdot \log n)$ due to Gupta et al. [22]³. We also give an $\Omega(\sqrt{k})$ -lower bound in this setting.

1.2 Techniques and Challenges

In this section we outline our main techniques and challenges for SMP adaptivity gaps.

Submodular Functions. To prove a small adaptivity gap, we need to show existence of a “good” non-adaptive solution. A priori it is not clear how to construct such a solution, e.g., LP based approaches do not extend beyond matroid constraints because of large integrality gaps. Since we only need to show *existence*, we can assume the optimal (exponential sized) decision tree is known. A crucial idea of [22] is to perform a random walk on this optimal decision tree (with probabilities given by the tree) and probing elements on the sampled root-leaf path. In other words, consider a non-adaptive strategy that randomly chooses a root-leaf path in the decision tree with the same probability as the optimal adaptive strategy. While this idea is natural in hindsight, its analysis for the non-adaptive strategy has been challenging.

In [22], the authors use Freedman’s inequality – linear functions are “well-concentrated” for a martingale – to argue that *simple* submodular functions are well-concentrated. This step requires massive union bounds over a polynomial number of linear functions, which loses logarithmic factors. To overcome this super-constant loss, in [23] the authors use an inductive approach and induct over subtrees where in each step a *stem* – the all-no path – is observed. A “stem lemma” allows them to argue that for every stem the expected value of the non-adaptive algorithm is within a factor 2 to the expected adaptive strategy. Finally, they “stitch” back the stem for induction by using submodularity, overall losing a factor of 3.

In this work, to prove the improved adaptivity gap of 2 in Theorem 1, our insight is to modify the above induction to observe a *single node* at each step (instead of a *stem* as in [23]). While we still induct over subtrees, this allows us to avoid any additional loss due to the stitching step. This induction turns out to be nontrivial because the adaptive and non-adaptive strategies can observe different types of the root element. In other words, although the non-adaptive random walk strategy follows the distribution of root-leaf paths of the adaptive strategy, it has to independently re-sample (re-probe) all the nodes on the chosen path. This hinders a direct application of induction as the marginal values in the subtrees change between the two strategies. We remedy this issue using two main ideas. First,

³ We remark that although not explicitly stated in Gupta et al. [23], their techniques can be used to remove the dependency on n , but it still only gives $\Omega(k^2)$ adaptivity gaps.

we compare the non-adaptive strategy to a “super-strategy” that can choose from both the elements chosen by the adaptive and the non-adaptive strategies. (This is also the intuition for the gap of 2 since the “super-strategy” has two chances to sample an element.) Second, the non-adaptive strategy forfeits any potential future value that the adaptive strategy gained at the root but the non-adaptive missed due to re-sampling. (This can be done by contracting the element sampled by the adaptive strategy without receiving its value.) Notice that both these steps are pessimistic and hence give a valid upper bound on the adaptivity gap. Together these ideas suffice to match the marginal values in the subtrees and apply induction without the stitching step, yielding an adaptivity gap of 2. Our lower bounds in §3.2 show examples where the super-strategy does not have any advantage over the adaptive strategy. Thus the adaptivity gap of 2 is optimal.

Rank Functions. A technical challenge in extending the above *inductive* approach to k -extendible system rank functions is that their marginal values do not belong to the same class. Namely, after contracting an element, the marginal value of a submodular function is submodular but the marginal value of a k -extendible system rank function may not even be subadditive. To overcome this, we first focus on *unweighted* rank functions. Instead of directly comparing the non-adaptive strategy to the adaptive strategy, our insight is to compare it to a *greedy procedure*. We show that this greedy procedure is a k -approximation to the adaptive strategy. Moreover, we show it has a notion of a marginal value. This allows us to compare the non-adaptive strategy to the greedy procedure in a similar way as for submodular functions, by losing another factor of 2. Our lower bound in §4.3 shows that the factor k loss in comparing to a greedy procedure is unavoidable, thereby making our analysis tight up to constants.

Finally, the challenge in proving Theorem 2 for *weighted* k -extendible system rank functions is that the greedy procedure only guarantees a k -approximation if we go in the order of decreasing weights. Instead, our adaptivity gap proofs only work when we are greedy in the root-to-leaf path order. One way around this is to partition the elements into $O(\log n)$ exponentially weighted *classes* (e.g., $1, 2, 2^2, \dots$) and apply the unweighted argument to the most valuable class. Unfortunately, this loses an $\Omega(\log n)$ factor. To obtain bounds independent of the universe size n , our insight is that picking an element in a class “removes” at most k elements from a lower weight class. We can therefore improve the $\log n$ factor loss to a $\log k$ by increasing the gap between successive classes to $\Omega(k)$. To achieve this we further combine $O(\log k)$ consecutive classes into a “super-class” (bucket). It is an interesting open question to find if this $\log k$ loss is essential in going from unweighted to weighted k -extendible system rank functions.

1.3 Further Related Work

The adaptivity gap of stochastic packing problems has seen much interest; see, e.g., for knapsack [14, 10, 28], packing integer programs [15, 13, 7], budgeted multi-armed bandits [17, 19, 26, 28], and orienteering [18, 20, 8]. All except the orienteering results rely on having relaxations that capture the constraints of the problem via linear constraints. For stochastic monotone submodular functions where the probing constraints are given by matroids, Asadpour et al. [4] bounded the adaptivity gap by $\frac{e}{e-1}$; Hellerstein et al. [25] bound it by $\frac{1}{\tau}$, where τ is the smallest probability of some set being materialized. Other relevant papers are [27, 16].

The work of Chen et al. [13] (see also [1, 7, 9, 2]) sought to maximize the size of a matching subject to b -matching constraints; this was motivated by applications to online dating and kidney exchange. See also [30, 6] for pointers to other work on kidney exchange

problems. The work of [21] abstracted out the general problem of maximizing a function (in their case, the rank function of the intersection of matroids or knapsacks) subject to probing constraints (again, intersection of matroids and knapsacks). This was improved and generalized by Adamczyk et al. [3] to submodular objectives. All these results use LPs or geometric relaxations, and do not extend to arbitrary packing constraints due to large integrality gaps of the relaxations.

2 Stochastic Multi-Value Probing Model

In this section we formally define our *stochastic multi-value probing* (SMP) model using the idea of combinatorial valuation over independent elements. We also discuss some preliminaries.

2.1 Combinatorial Valuation over Independent Elements

The multi-value paradigm is based on the notion of *type*, which represents different “values” an element can take. This leads to combinatorial valuations over independent elements where each element *independently* takes its type. Similar notions have been defined before; e.g., see [31] and references therein.

► **Definition 3** (Combinatorial valuation $\text{val}_{\mathbf{X}}$ over independent elements). *Consider a finite universe V of elements and size $n = |V|$. Each element $e \in V$ obtains exactly one type from a finite set T_e according to a given probability distribution \mathcal{D}_e over T_e . These types are assigned independently across different elements, i.e., the random vector of types $\mathbf{X} \in (T_e)_{e \in V}$ is drawn from the product distribution $\prod_{e \in V} \mathcal{D}_e$. Given a combinatorial function $f : 2^T \rightarrow \mathbb{R}_{\geq 0}$ for $T \stackrel{\text{def}}{=} \bigcup_{e \in V} T_e$, the valuation of a set $S \subseteq V$ is*

$$\text{val}_{\mathbf{X}}(S) \stackrel{\text{def}}{=} f(\{\mathbf{X}_e \mid e \in S\}) = f(\mathbf{X}_S),$$

where we define $\mathbf{X}_S \stackrel{\text{def}}{=} \{\mathbf{X}_e \mid e \in S\}$ to simplify notation.

For example, in the Bernoulli case studied in the stochastic probing literature, each element has two types: active and inactive, the distributions \mathcal{D}_e are Bernoulli, and the valuation function $\text{val}_{\mathbf{X}}(S) = f(\{e \in S \mid e \text{ is active}\})$. Another example is the multi-value max-weight matching problem described in the introduction. Here different types of an element (edge) correspond to its different weights and $\text{val}_{\mathbf{X}}(S)$ is the max-weight matching in the induced subgraph on S .

In this work we always assume the combinatorial function $f : 2^T \rightarrow \mathbb{R}_{\geq 0}$ satisfies $f(\emptyset) = 0$ and is *monotone*, i.e., $f(A) \leq f(B)$ for all $A \subseteq B$. We also assume it belongs to one of the following classes.

- *subadditive* if $f(A \cup B) \leq f(A) + f(B)$ for all $A, B \subseteq T$.
- *submodular* if $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$ for all $A, B \subseteq T$. For $S \subseteq T$, the contraction

$$f_S(A) \stackrel{\text{def}}{=} f(S \cup A) - f(S) \tag{1}$$

of a monotone submodular function is also monotone submodular.

- *weighted rank function* of a family $\mathcal{F} \subseteq 2^T$ if $f(A) = \max_{B \in \mathcal{F}} w(A \cap B)$ where $w : 2^T \rightarrow \mathbb{R}_{\geq 0}$ is a linear function with non-negative weights. When w is the all ones vector (i.e., $w(A) = |A|$), we call it the *unweighted rank function* of \mathcal{F} .

In particular, we work with rank functions of two special families $\mathcal{F} \subseteq 2^V$. Subsets in the family are called *independent* subsets. A family $\mathcal{F} \ni \emptyset$ forms a

- *matroid* if for every $A, B \in \mathcal{F}$ with $|A| > |B|$ there exists $x \in A \setminus B$ such that $B \cup \{x\} \in \mathcal{F}$.
- *k-extendible system* if for every $A \subseteq B \in \mathcal{F}$ and $e \in T$ where $A \cup \{e\} \in \mathcal{F}$, we have that there is a set $Z \subseteq B \setminus A$ such that $|Z| \leq k$ and $B \setminus Z \cup \{e\} \in \mathcal{F}$.

This latter family is important because it generalizes the family of intersection of k matroids, e.g., a 2-extendible systems captures general graph matchings (see [11] for further discussion).

2.2 Adaptive Strategies and SMP

Roughly, the goal of an SMP problem is to maximize a combinatorial function over independent elements under some “feasibility constraints”. We define a *probe* of an element $e \in V$ to be an operation that reveals its random type $X_e \in T_e$. A *probing sequence* is an ordered sequence of probes on some elements.

The SMP problem only allows a family of probing sequences \mathcal{C} , which are called *feasible*. We assume minimal properties from this family. Specifically, it is *prefix-closed*, i.e., for every sequence in \mathcal{C} , each of its prefix is also in \mathcal{C} . This prefix-closed family is powerful because it generalizes any *downward-closed* family \mathcal{F} (i.e., for all $A \in \mathcal{F}$ and $B \subseteq A$ we have $B \in \mathcal{F}$) and can also capture precedence constraints.

We now define an *adaptive strategy* which constitutes a feasible solution for SMP. The nodes in this tree correspond to probes of elements

► **Definition 4** (Adaptive strategy \mathcal{T}). *It is a rooted decision tree where each non-leaf node is labeled with an element $e \in V$ and has $|T_e|$ arcs to child nodes. Each arc is uniquely labeled with a type $t \in T_e$. Whenever we encounter a node labeled e , the adaptive strategy probes e and proceeds to the subtree corresponding to the arc labeled $X_e \sim \mathcal{D}_e$. The strategy terminates on reaching a leaf and receives a value of $\text{val}_{\mathbf{X}}(S(\mathbf{X}))$, where $S(\mathbf{X}) \subseteq V$ is the set of probed elements by strategy \mathcal{T} for type vector \mathbf{X} . The objective is the expected valuation, which we denote by*

$$\text{adap}(\mathcal{T}, f) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{X}}[\text{val}_{\mathbf{X}}(S(\mathbf{X}))]. \quad (2)$$

Notice, since f is monotone, a strategy never gains value by removing a probed element. We say a strategy \mathcal{T} is *feasible* for \mathcal{C} if every root-leaf path belongs to \mathcal{C} . We now formally define an SMP problem.

► **Definition 5** (SMP problem $(\mathcal{C}, \text{val}_{\mathbf{X}})$). *Given a prefix-closed family of probing constraints \mathcal{C} and a combinatorial valuation $\text{val}_{\mathbf{X}}$ over independent elements, an SMP problem is to find a feasible adaptive strategy \mathcal{T} to maximize the expected valuation $\text{adap}(\mathcal{T}, f)$.*

2.3 Non-Adaptive Strategies and Adaptivity Gaps

A strategy to solve an SMP problem can benefit from adjusting its probing sequence based on the outcomes of the already probed elements. For instance, in the kidney-exchange example if one finds an edge incident to a vertex u , one may choose not to probe any other edges incident to u . On the other hand, a strategy that always decides the next probe independent of the outcomes of the probed elements is called *non-adaptive*. Our goal is to study the largest ratio between adaptive and non-adaptive strategies.

► **Definition 6** (Adaptivity gap for \mathcal{P}). Let \mathcal{P} be a class of SMP problems (e.g., monotone submodular functions over prefix-closed constraints). Define the adaptivity gap as the largest (worst-case instance) ratio of the optimal adaptive and optimal non-adaptive strategies for a problem $(\mathcal{C}, \text{val}_{\mathbf{X}}) \in \mathcal{P}$, i.e.,

$$\sup_{(\mathcal{C}, \text{val}_{\mathbf{X}}) \in \mathcal{P}} \frac{\sup_{\mathcal{T} \text{ is feasible in } \mathcal{P}} \text{adap}(\mathcal{T}, f)}{\sup_{S \in \mathcal{C}} \mathbb{E}_{\mathbf{X}}[\text{val}_{\mathbf{X}}(S)]}.$$

Notice that in the denominator S does not depend on \mathbf{X} .

The adaptivity gap for a general combinatorial function f is unbounded [22]. In this work we focus on monotone submodular functions and (weighted) rank functions of a k -extendible system. We bound adaptivity gaps by analyzing the following natural random walk non-adaptive strategy.

► **Definition 7** (Random walk non-adaptive strategy). For any given adaptive strategy \mathcal{T} , there is a corresponding non-adaptive strategy that (virtually) draws a sample $\mathbf{X} \sim \prod_{e \in V} \mathcal{D}_e$ from the product distribution and traverses \mathcal{T} along the root-leaf path for \mathbf{X} (i.e., when at a node labeled e , traverse the unique arc labeled X_e). Let $S(\mathbf{X})$ be the random set of elements probed by such a root-leaf path. The true (non-virtual) types of elements correspond to the vector of outcomes $\mathbf{X}' \sim \prod_{e \in V} \mathcal{D}_e$. Here \mathbf{X} and \mathbf{X}' are i.i.d. r.v.s. The random walk non-adaptive strategy probes S according to the above distribution and receives the valuation

$$\text{alg}(\mathcal{T}, f) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{X}, \mathbf{X}'}[\text{val}_{\mathbf{X}'}(S(\mathbf{X}))]. \quad (3)$$

3 Adaptivity Gaps for a Monotone Submodular Function

In this section we prove our first main result, the optimal adaptivity gap for submodular functions. In §3.1 we prove the upper bound and in §3.2 we prove the lower bound of Theorem 1.

► **Theorem 1.** *The adaptivity gap for SMP where the constraints are prefix-closed and the function is monotone non-negative submodular is exactly 2.*

3.1 Upper Bound of 2

Our non-adaptive strategy samples a random root-leaf path using the optimal adaptive strategy tree \mathcal{T} (Definition 7). In other words, it performs a “dry-run” of a random walk along the tree without probing anything. In the end it queries all the elements on this random root-leaf path. We argue that its expected value is at least half of the adaptive strategy. We encourage the reader to follow the proof idea outlined in §1.2 since algebra can conceal the main ideas.

Proof of the upper bound in Theorem 1. We induct over the depth of the tree \mathcal{T} , i.e., for any monotone submodular function f and tree \mathcal{T} of depth at most d , we have

$$\text{alg}(\mathcal{T}, f) \geq \frac{1}{2} \text{adap}(\mathcal{T}, f).$$

The base case for $d = 1$ is trivially true because the tree is a single node. For induction, let e be the root node of the optimal decision tree \mathcal{T} . Denote by $I \stackrel{\text{def}}{=} X_e$ the (random) type of element e when probed by the adaptive strategy (and also the virtual type of the

non-adaptive strategy), while $R \stackrel{\text{def}}{=} X'_e$ be the (random) true type when probed by the non-adaptive strategy. Also, let \mathcal{T}_I denote the subtree the adaptive strategy goes to when the root element is in type I and let f_I be the contraction from Eq. (1). This implies

$$\text{adap}(\mathcal{T}, f) = \mathbb{E}_I[f(I) + \text{adap}(\mathcal{T}_I, f_I)] \quad \text{and} \quad \text{alg}(\mathcal{T}, f) = \mathbb{E}_{I,R}[f(R) + \text{alg}(\mathcal{T}_I, f_R)]. \quad (4)$$

Now using submodularity and monotonicity of f on every root-leaf path of the adaptive strategy,

$$\begin{aligned} \text{adap}(\mathcal{T}, f) &\leq \mathbb{E}_{I,R}[f(I \cup R) + \text{adap}(\mathcal{T}_I, f_{I \cup R})] \\ &\leq \mathbb{E}_{I,R}[f(I) + f(R) + \text{adap}(\mathcal{T}_I, f_{I \cup R})], \end{aligned}$$

where the last inequality uses that every monotone submodular function is subadditive. Notice that I and R are i.i.d. variables. This along with linearity of expectation implies

$$\text{adap}(\mathcal{T}, f) \leq \mathbb{E}_{I,R}[2 \cdot f(R) + \text{adap}(\mathcal{T}_I, f_{I \cup R})]. \quad (5)$$

Next, we lower bound the expected value of the non-adaptive strategy from Eq. (4). We use monotonicity of f to get

$$\text{alg}(\mathcal{T}, f) = \mathbb{E}_{I,R}[f(R) + \text{alg}(\mathcal{T}_I, f_R)] \geq \mathbb{E}_{I,R}[f(R) + \text{alg}(\mathcal{T}_I, f_{I \cup R})]. \quad (6)$$

Since $f_{I \cup R}$ is also a monotone submodular function over independent elements and \mathcal{T}_I is an adaptive strategy tree of depth at most $d - 1$, by induction hypothesis

$$\text{alg}(\mathcal{T}_I, f_{I \cup R}) \geq \frac{1}{2} \text{adap}(\mathcal{T}_I, f_{I \cup R}).$$

Combining this with Eq. (5) and Eq. (6), we get

$$\text{alg}(\mathcal{T}, f) \geq \frac{1}{2} \text{adap}(\mathcal{T}, f),$$

which finishes the proof of the upper bound by induction. \blacktriangleleft

3.2 Lower Bound of 2

In this section we show a monotone non-negative submodular function and a prefix-closed set of constraints where the adaptivity gap for stochastic probing is arbitrarily close to 2. Combined with §3.1, this proves Theorem 1 that the optimal adaptivity gap is exactly 2.

The proof below uses a stochastic probing instance on an infinite universe. Since submodular functions are defined only on finite sets, the proof below is informal. We do this to explain our main ideas and defer the formal proof to Appendix A.

Informal proof of the lower bound in Theorem 1. Our example is on a universe $V := \{e_{(k,l)} \mid k, l \in \mathbb{Z}_{\geq 0}\}$ where every element is independently active with probability ϵ for some $0 < \epsilon < 1$.

Example. We define our submodular objective f to be the weighted rank function of a partition matroid that selects at most one element from each part. The elements are partitioned according to their first label – for every $k \in \mathbb{Z}_{\geq 0}$ the set $\{e_{(k,l)} \mid l \in \mathbb{Z}_{\geq 0}\}$ is a

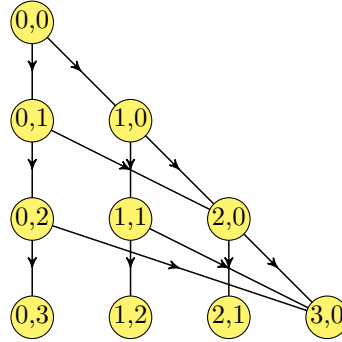
49:10 (Near) Optimal Adaptivity Gaps for Stochastic Multi-Value Probing

part of the partition matroid with weight $(1 - \epsilon)^k$. In other words, for any set $S \subseteq V$ let $K(S) := \{k \mid e_{(k,l)} \in S\}$ be the (unique) set of first labels, then

$$f(S) \stackrel{\text{def}}{=} \sum_{k \in K(S)} (1 - \epsilon)^k.$$

Note that this series always converges so f is well defined.

To define our prefix-closed constraints, we consider an infinite directed acyclic graph where every element is identified with a single node in the graph. Every node/element $e_{(k,l)}$ has exactly two outgoing edges: towards $e_{(k,l+1)}$ and towards $e_{(k+l+1,0)}$. We denote $\{e_{(k,0)}, e_{(k,1)}, \dots\}$ as the elements on *column* k . The probing constraint is that a sequence of elements can be probed if and only if it corresponds to a directed path starting at $e_{(0,0)}$. See Figure 2 for an illustration.



■ **Figure 2** Adaptivity gap lower bound example for monotone submodular functions.

Analysis. We first give an adaptive strategy with value $2 - \epsilon$ (in Eq. (7)) and later argue that every non-adaptive strategy has value at most 1 (in Eq. (8)); thereby, proving this theorem. Although, the probing constraint allows for infinite strategies, and in a different setting it would not be clear how to define their expected values, since f is monotone we include every active element in the solution. So the expected value of an infinite strategy can be defined as the limit of strategies that only probe a finite number of elements. The finite lower bound example in Appendix A is constructed by reducing V so that the resulting strategies are close to this limit.

Our adaptive strategy **adap** starts with probing element $e_{(0,0)}$. It is defined recursively: after probing $e_{(k,l)}$, the next element to probe is either $e_{(k+l+1,0)}$ if $e_{(k,l)}$ is found active, or $e_{(k,l+1)}$ otherwise. In other words, it probes elements on a column until it finds one active, and then probes another column.

Let $\text{adap}(k)$ denote the expected additional value our above adaptive strategy if the next probed element is $e_{(k,0)}$ and let $\text{adap} \stackrel{\text{def}}{=} \text{adap}(0)$ denote the expected value of the entire strategy. Note that $\text{adap}(k)$ does not depend on the set of elements found active before probing $e_{(k,0)}$ (i.e., the elements $e_{(k',l')}$ where $k' < k$). Furthermore, the subgraph reachable from $e_{(k,0)}$ is similar to the entire graph on V in the sense that one can relabel the elements in the subgraph to match the entire graph exactly, the only difference being that the value of any subset is multiplied by a factor of $(1 - \epsilon)^k$. Therefore, we have

$$\text{adap}(k) = (1 - \epsilon)^k \cdot \text{adap}(0).$$

Now, summing over the number of inactive elements on column 0, we get

$$\begin{aligned} \text{adap}(0) &= \sum_{k=0}^{\infty} (1 - \epsilon)^k \cdot \epsilon \cdot \left(1 + \text{adap}(k + 1)\right) \\ &= \sum_{k=0}^{\infty} (1 - \epsilon)^k \cdot \epsilon \left(1 + (1 - \epsilon)^{k+1} \cdot \text{adap}(0)\right), \end{aligned}$$

which uses $\text{adap}(k) = (1 - \epsilon)^k \cdot \text{adap}(0)$. Solving this equation yields the result:

$$\text{adap} = \text{adap}(0) = 2 - \epsilon. \tag{7}$$

Similarly, let $\text{alg}(k)$ denote the expected additional value of the optimal non-adaptive strategy if the next probed element is $e_{(k,0)}$, and let $\text{alg} = \text{alg}(0)$ denote the expected value of the optimal non-adaptive strategy. By the same argument as $\text{adap}(k)$, we have

$$\text{alg}(k) = (1 - \epsilon)^k \cdot \text{alg}(0).$$

Let k denote the number of elements the optimal non-adaptive strategy probes on column 0. We get

$$\text{alg}(0) = \sup_{k \geq 1} \left\{ 1 - (1 - \epsilon)^k + \text{alg}(k) \right\} = \sup_{k \geq 1} \left\{ 1 - (1 - \epsilon)^k + (1 - \epsilon)^k \cdot \text{alg}(0) \right\},$$

which uses $\text{alg}(k) = (1 - \epsilon)^k \cdot \text{alg}(0)$. This implies

$$\text{alg} = \text{alg}(0) = 1. \tag{8}$$

Combining Eq. (7) and Eq. (8), we get an adaptivity gap arbitrarily close to 2 for $\epsilon \rightarrow 0$. ◀

4 Adaptivity Gaps for a Weighted Rank Function of a k -Extendible System

For a downward-closed family \mathcal{F} , recollect that we define its rank function $f_{\mathcal{F}} : 2^V \rightarrow \mathbb{R}_{\geq 0}$ to be the largest cardinality subset in \mathcal{F} , i.e., $f_{\mathcal{F}}(S) \stackrel{\text{def}}{=} \max_{T \subseteq S \ \& \ T \in \mathcal{F}} |T| = \max_{T \in \mathcal{F}} |S \cap T|$. In this section we prove our results on the adaptivity gaps of a weighted rank function of a k -extendible system.

► **Theorem 2.** *The adaptivity gap for SMP where the constraints are prefix-closed and the function is a weighted rank function of a k -extendible system is between k and $O(k \log k)$. Moreover, for unweighted rank functions, the adaptivity gap is between k and $2k$.*

In §4.1 we prove the upper bound for unweighted k -extendible systems, and in §4.2 we give a reduction from weighted to unweighted k -extendible systems that loses a factor $O(\log k)$ in the adaptivity gap. Our lower bound is presented in §4.3.

To simplify our proofs, we define an element $e \in T$ as a *loop* in $\mathcal{F} \subseteq 2^T$ if $\{e\} \notin \mathcal{F}$. Furthermore, given a non-loop element $e \in T$, we define the *contraction* \mathcal{F}/e as $\{F \setminus \{e\} \mid F \in \mathcal{F}, e \in F\}$, i.e., the family of subsets that contain e but with e removed. We also need the following property of k -extendible systems, which intuitively means a set $E \in \mathcal{F}$ hurts at most $k \cdot |E|$ from another set $B \in \mathcal{F}$. We include the proof for completeness in Appendix B.

Let $\mathcal{F} \subseteq 2^T$ be a k -extendible system. For every $A \subseteq B \in \mathcal{F}$ and $E \subseteq T$ where $A \cup E \in \mathcal{F}$, there exists a set $Z \subseteq B \setminus A$ such that $|Z| \leq k \cdot |E|$ and $B \setminus Z \cup E \in \mathcal{F}$.

4.1 Upper Bound of $2k$ for an Unweighted k -Extendible System

Let \mathcal{T} denote the optimal adaptive strategy for maximizing the rank function f of a given k -extendible system \mathcal{F} . We prove the following unweighted upper bound of Theorem 2.

► **Theorem 8.** *The adaptivity gap for SMP where the constraints are prefix-closed and the function is an unweighted rank function of a k -extendible system is at most $2k$.*

We use the random walk strategy to convert the adaptive strategy \mathcal{T} into a non-adaptive strategy. To analyze our algorithm, we define a natural *greedy procedure* to select a subset of $A \subseteq T$ that is also in $\mathcal{F} \subseteq 2^T$. First, consider elements of A in an arbitrary order (which can even be determined on the fly). If the currently considered element is a non-loop, it gets contracted in \mathcal{F} ; otherwise it gets ignored. Any such computed set is in \mathcal{F} and the final output, the number of contracted elements, is denoted by $\text{greedy}(A)$. We first show that for k -extendible systems such a greedy procedure produces a k -approximation to the largest subset in \mathcal{F} . A similar statement has been proven by Mestre [29].

► **Lemma 9.** *Let f be a rank function of a k -extendible system $\mathcal{F} \subseteq 2^T$. Fix any subset $A \subseteq T$ and consider the output of the greedy procedure $\text{greedy}(A)$ with an arbitrary ordering of A . We have that $f(A) \leq k \cdot \text{greedy}(A)$. Even more, for any $A \subseteq B \subseteq T$ we have that $f(A) \leq k \cdot \text{greedy}(B)$.*

Proof. Let $G \subseteq B$ be the set picked by $\text{greedy}(B)$. Notice that G is a maximal set in \mathcal{F} (need not be maximum). On the other hand, let $\text{OPT} \subseteq A$ be the set picked by $f(A)$, i.e., the maximum set in \mathcal{F} on A . Our goal is to prove $|\text{OPT}| \leq k \cdot |G|$.

Let $C \stackrel{\text{def}}{=} \text{OPT} \cap G$, note that $G = C \cup (G \setminus C) \in \mathcal{F}$ and $C \subseteq \text{OPT}$, hence by Section 4 there is a $Z \subseteq \text{OPT} \setminus C$ with $|Z| \leq k \cdot |G \setminus C| = k \cdot |G| - k \cdot |C|$ such that $\text{OPT} \setminus Z \cup (G \setminus C) = (\text{OPT} \setminus C) \setminus Z \cup G \in \mathcal{F}$. However, since G is a maximal set and $(\text{OPT} \setminus C) \cap G = \emptyset$ we know that $\text{OPT} \setminus C \setminus Z = \emptyset$ and hence $|\text{OPT}| \leq |Z| + |C| \leq k \cdot |G| - k \cdot |C| + |C| = k \cdot |G| - (k-1)|C| \leq k \cdot |G|$. ◀

Given the above properties of a k -extendible system, we can now prove Theorem 8.

Proof of Theorem 8. Let \mathbf{X} and \mathbf{X}' denote the element types for the adaptive and the non-adaptive algorithms, respectively. The adaptive strategy on the optimal decision tree \mathcal{T} gets value $f(\mathbf{X}_S)$, where $S \subseteq V$ is the set of probed elements by strategy \mathcal{T} for type vector \mathbf{X} . We compare this value to a greedy strategy $\text{greedy}(\mathbf{X}_S \cup \mathbf{X}'_S)$ in which

1. we consider the elements of S in root-to-leaf order in which they appear on the tree and
2. for any $e \in S$ we first consider \mathbf{X}'_e (the true type) before \mathbf{X}_e (the virtual type) in the greedy order.

Note by Lemma 9 we have

$$\text{adap}(\mathcal{T}, f) = \mathbb{E}_{\mathbf{X}}[f(\mathbf{X}_S)] \leq k \cdot \mathbb{E}_{\mathbf{X}, \mathbf{X}'}[\text{greedy}(\mathbf{X}_S \cup \mathbf{X}'_S)].$$

By induction on the subtrees, below we prove

$$\mathbb{E}_{\mathbf{X}, \mathbf{X}'}[\text{greedy}(\mathbf{X}_S \cup \mathbf{X}'_S)] \leq 2 \cdot \text{alg}(\mathcal{T}, f). \quad (9)$$

This finishes the proof of Theorem 8 because the optimal non-adaptive algorithm has value at least

$$\text{alg}(\mathcal{T}, f) \geq \frac{1}{2} \cdot \mathbb{E}_{\mathbf{X}, \mathbf{X}'}[\text{greedy}(\mathbf{X}_S \cup \mathbf{X}'_S)] \geq \frac{1}{2k} \cdot \text{adap}(\mathcal{T}, f).$$

To prove the missing Eq. (9), we induct on the height of the tree and \mathcal{F} being any downward-closed family. For consistency, we define the notation of $\text{greedy}(\mathcal{T}, f)$ to denote the value of the above greedy strategy when run on \mathcal{T} with a rank function f . Thus, $\text{greedy}(\mathcal{T}, f) = \mathbb{E}_{\mathbf{X}, \mathbf{X}'}[\text{greedy}(\mathbf{X}_S \cup \mathbf{X}'_S)]$. Suppose $e \in V$ is the label of the root of \mathcal{T} . Denote by $I \stackrel{\text{def}}{=} X_e$ the (random) type of element e when probed by the adaptive strategy (which is also the virtual type of the non-adaptive strategy), and denote $R \stackrel{\text{def}}{=} X'_e$ the (random) true type when probed by the non-adaptive strategy. Also, let \mathcal{T}_I denote the subtree the adaptive strategy goes to when the root e is in state I . We have

$$\text{greedy}(\mathcal{T}, f) \leq \mathbb{E}_{I,R}[f(I \cup R) + \text{greedy}(\mathcal{T}_I, (f/R)/I)],$$

where by $(f/R)/I$ we mean the rank function of \mathcal{F} after we first contract R if it is a non-loop, and then contract I if it is still a non-loop. Now subadditivity of f gives

$$\begin{aligned} \text{greedy}(\mathcal{T}, f) &\leq \mathbb{E}_{I,R}[f(I) + f(R) + \text{greedy}(\mathcal{T}_I, (f/R)/I)] \\ &= \mathbb{E}_{I,R}[2 \cdot f(R) + \text{greedy}(\mathcal{T}_I, (f/R)/I)], \end{aligned} \tag{10}$$

where the last equality uses linearity of expectation as I and R are identically distributed.

Next, we lower bound the value of our non-adaptive algorithm. Although it takes a random root-leaf path and decides the set of elements to retain in the end, we lower bound its value by an online algorithm that greedily selects R (unless it is a loop), however, always also contracts I if it is a non-loop. This gives,

$$\text{alg}(\mathcal{T}, f) \geq \mathbb{E}_{I,R}[f(R) + \text{alg}(\mathcal{T}_I, (f/R)/I)]. \tag{11}$$

Since $(f/R)/I$ is also a rank function of a downward-closed system and \mathcal{T}_I is an adaptive strategy, by induction hypothesis we have

$$\text{alg}(\mathcal{T}_I, (f/R)/I) \geq \frac{1}{2} \text{greedy}(\mathcal{T}_I, (f/R)/I).$$

Combining this with Eq. (10) and Eq. (11), we get

$$\text{greedy}(\mathcal{T}, f) \leq 2 \cdot \text{alg}(\mathcal{T}, f),$$

which proves Eq. (9) by induction. ◀

4.2 Reducing Weighted to Unweighted k -Extendible System by Losing $O(\log k)$

We show how to extend the adaptivity gap result for an unweighted k -extendible system to a weighted k -extendible system by losing an $O(\log k)$ factor.

► **Theorem 10.** *For SMP over prefix-closed constraints, the adaptivity gap for a weighted rank function of a k -extendible system is at most $32k \log_2 k$.*

Proof. Given a weighted rank function f of a k -extendible system $\mathcal{F} \subseteq 2^T$ over a set of types T , we define f_j for $j \in \mathbb{Z}$ to be an unweighted rank function of the k -extendible system \mathcal{F} ; however, the new weights are changed such that only the types with original weights in

49:14 (Near) Optimal Adaptivity Gaps for Stochastic Multi-Value Probing

$(2^{j-1}, 2^j]$ participate with new weight of 1, while the other elements have a new weight of 0. Note that this partitions the set of types T into pairwise disjoint *classes*. Notice, we have

$$\text{adap}(\mathcal{T}, f) \leq \sum_j 2^j \cdot \text{adap}(\mathcal{T}, f_j), \quad (12)$$

where $\text{adap}(\mathcal{T}, f_j)$ denotes the expected value of an adaptive strategy given by the common decision tree \mathcal{T} with respect to the rank function f_j .

Now, since $\text{adap}(\mathcal{T}, f_j)$ is an unweighted k -extendible system problem, we know that a random root-leaf path returns a solution with expected value

$$\text{alg}(\mathcal{T}, f_j) \geq \frac{1}{2k} \cdot \text{adap}(\mathcal{T}, f_j). \quad (13)$$

In the following lemma, we show that these non-adaptive solutions for f_j can be combined to obtain a feasible and “high-value” non-adaptive solution for f .

► **Lemma 11.** *The random-walk non-adaptive algorithm alg has expected value*

$$\text{alg}(\mathcal{T}, f) \geq \frac{1}{16 \cdot \log k} \sum_j 2^j \cdot \text{alg}(\mathcal{T}, f_j).$$

Before proving Lemma 11, we finish the proof of Theorem 10 by combining it with Eq. (13) and Eq. (12):

$$\begin{aligned} \text{alg}(\mathcal{T}, f) &\geq \frac{1}{16 \cdot \log k} \sum_j 2^j \cdot \text{alg}(\mathcal{T}, f_j) \geq \frac{1}{32k \log k} \sum_j 2^j \cdot \text{adap}(\mathcal{T}, f_j) \\ &\geq \frac{1}{32k \log k} \cdot \text{adap}(\mathcal{T}, f). \quad \blacktriangleleft \end{aligned}$$

Informally, in the proof of Lemma 11 we combine the unweighted solutions of $\text{alg}(\mathcal{T}, f_i)$ by running a “greedy-optimal” algorithm from the higher weight to the smaller weight classes and fixing the types chosen in earlier classes. Unfortunately, in general such an approach loses an extra factor k in the approximation. To fix this, our second idea is to increase the weight gap between successive classes. We achieve this by combining $O(\log k)$ consecutive classes into a *bucket*, where in each bucket we focus on the class with the largest non-adaptive value. Because of boundary issues, we only take either odd or even buckets.

Proof of Lemma 11. Let $a \leq b \in \mathbb{Z}$ denote the indices of the smallest and the highest weight classes. We define buckets consisting of $2 \log k$ consecutive classes, where bucket B_i consists of classes $\{b - 2i \log k, b - 2i \log k - 1, \dots, b - 2(i - 1) \log k\}$. For each B_i , let

$$j(i) \stackrel{\text{def}}{=} \operatorname{argmax}_{j \in B_i} \{2^j \cdot \text{alg}(\mathcal{T}, f_j)\}.$$

Since each bucket has size $2 \log k$, this implies

$$\sum_i 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)}) \geq \frac{1}{2 \cdot \log k} \sum_j 2^j \cdot \text{alg}(\mathcal{T}, f_j).$$

Without loss of generality we can assume the odd indices satisfy

$$\sum_{i \text{ is odd}} 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)}) \geq \frac{1}{2} \sum_i 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)}).$$

Otherwise, use the same argument for even indices. Combining the last two equations, we get

$$\sum_{i \text{ is odd}} 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)}) \geq \frac{1}{4 \cdot \log k} \sum_j 2^j \cdot \text{alg}(\mathcal{T}, f_j). \quad (14)$$

We now claim that a *greedy-optimal* algorithm has a large value: It goes over classes $j(i)$ in decreasing order of (odd) buckets, but it always selects the maximum independent set (instead of selecting a maximal greedy set) in the current class $j(i)$ given its choices in the previous. This algorithm is, therefore, a combination of greedy and optimal algorithms. The proof of the following is deferred to Appendix C.

▷ **Claim 12.** Consider an algorithm that goes over the odd numbered buckets in decreasing order of weights and selects the *maximum* set from class $j(i)$ in bucket i such that the resulting set is still feasible in \mathcal{F} . (After a set in a class is selected, it gets fixed for all smaller choices.) The finally chosen set has value at least $\frac{1}{4} \sum_{i \text{ is odd}} 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)})$.

Using Claim 12, we have

$$\text{alg}(\mathcal{T}, f) \geq \frac{1}{4} \sum_{i \text{ is odd}} 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)}),$$

which combined when with Eq. (14) proves Lemma 11. ◀

4.3 Lower Bounds

We present two very similar lower bound examples: one where the adaptivity gap is $k - o(1)$ for a rank function of an unweighted k -extendible system and another where the adaptivity gap is $\Omega(\sqrt{k})$ for a rank function of an intersection of k matroids. A related example was also shown in [23].

Example. For generality we work in the Bernoulli setting where each element in V is either active or inactive. Consider a perfect w -ary tree of depth k whose edges correspond to the ground set V . Each edge is active with probability $p > 0$. For any leaf ℓ , let P_ℓ denote the unique path from the root to ℓ . The objective value on any set is the maximum number of edges in the set on the *same* root-leaf path, i.e., for any $S \subseteq V$,

$$f(S) \stackrel{\text{def}}{=} \max_{\text{leaf } \ell} |P_\ell \cap S|.$$

The feasibility constraints are such that a set of edges can be probed if and only if there exists some root-leaf path P_ℓ such that every probed edge has at least one endpoint on P_ℓ . Note that this implies that a maximum of $w \cdot k$ edges can be probed.

Analysis. Let the adaptive strategy be the following: probe all w edges incident to the root. If any of them is active, start probing the edges directly below the active edge, otherwise below the first edge. Continue recursively until a leaf is reached. On every level, the adaptive strategy has $1 - (1 - p)^w$ probability of finding an active edge. Therefore, the expected value of the adaptive strategy is $k \cdot (1 - (1 - p)^w)$.

For any non-adaptive strategy, the feasibility constraints imply there exists a root-leaf path P_ℓ such that all probed edges have an endpoint on it. Suppose all $w \cdot k$ edges incident to P_ℓ are probed. The non-adaptive strategy can get value at most 1 from the edges not on P_ℓ and in expectation at most $k \cdot p$ from the edges on P_ℓ . So, the non-adaptive strategy has an expected value of at most $1 + k \cdot p$.

Lower Bound of k for an unweighted k -extendible system

Consider the example described above and set $w \stackrel{\text{def}}{=} k^4$ and $p \stackrel{\text{def}}{=} \frac{1}{k^3}$. The function f is trivially a rank function of a k -extendible system because the rank of the system is k , i.e., $f(V) = k$. The adaptive strategy has an expected value

$$k \cdot \left(1 - \left(1 - \frac{1}{k^3}\right)^{k^4}\right) \geq k \cdot \left(1 - \frac{1}{e^k}\right) = k - o(1),$$

whereas any non-adaptive strategy has an expected value at most $1 + \frac{1}{k^2}$. This gives an adaptivity gap of $k - o(1)$.

Lower Bound of $\Omega(\sqrt{k})$ for an unweighted intersection of k matroids

In this section we show how to model the above example as an intersection of $t = k^2$ matroids, yielding an adaptivity gap of $\Omega(\sqrt{t})$ for an intersection of t matroids. Consider the example described above and set $w \stackrel{\text{def}}{=} k$ and $p \stackrel{\text{def}}{=} \frac{1}{k}$. The adaptive strategy has an expected value of

$$k \cdot \left(1 - \left(1 - \frac{1}{k}\right)^k\right) \geq k \cdot \left(1 - \frac{1}{e}\right) = \Omega(k)$$

and the non-adaptive strategy gets at most 2 in expectation; so the adaptivity gap is $\Omega(k)$.

All that remains to show is that f can be represented as an intersection of k^2 simple partition matroids. We use the term simple partition matroid for a matroid that partitions the V into multiple parts and a set is independent if it contains at most one element in every part.

Suppose that k is prime and label each node v with a list L_v as follows: the root's label is an empty list $()$. Let $L(i)$ denote the i^{th} element of the list L and $L + x$ a list equal to L with x appended to it. All the other nodes are labeled recursively: let v be a node with children $\{v_0, v_1, \dots, v_{k-1}\}$. Define $L_{v_i} \stackrel{\text{def}}{=} L_v + i$. Hence, u is an ancestor of v if and only if L_u is a prefix of L_v , and otherwise $L_u(i) \neq L_v(i)$ for some i .

Let e_v denote the edge/element between v and its parent. We define k^2 partition matroids $M_{i,j}$ for $i \in \{1, 2, \dots, k\}$ and $j \in \{0, 1, \dots, k-1\}$. Each $M_{i,j}$ consists of k big partitions indexed from 0 to $k-1$, and all other partitions contain only a single element. Let

$$I_v(i, j) \stackrel{\text{def}}{=} L_v(i)j + d_v \pmod{k}.$$

For a node v on depth $d_v \geq i$, element e_v is in the $I_v(i, j)^{\text{th}}$ big partition of $M_{i,j}$. For a node v on depth $d_v < i$, e_v is the only element in its partition in $M_{i,j}$.

We claim that f is the rank function of $\mathcal{F} \stackrel{\text{def}}{=} \bigcap_{i=1}^k \bigcap_{j=0}^{k-1} M_{i,j}$, which is an intersection of k^2 matroids. Since \mathcal{F} is an intersection of simple partition matroids, $S \in \mathcal{F}$ if and only if $\{a, b\} \in \mathcal{F}$ for every $a, b \in S$. Now consider two nodes u, v such that $\{e_u, e_v\} \notin \mathcal{F}$. This means $I_u(i, j) = I_v(i, j)$ for some $i \leq d_u, d_v$ and $j \in \{0, 1, \dots, k-1\}$, which is equivalent to

$$L_u(i) \cdot j + d_u \equiv L_v(i) \cdot j + d_v \pmod{k}.$$

Since k is prime, this holds for some i, j if and only if $d_u = d_v$ (for $j = 0, i = 1$) or $L_u(i) \neq L_v(i)$ for any i . That is, $\{e_u, e_v\} \notin \mathcal{F}$ if and only if u and v are not ancestors of one another, which completes the proof.

References

- 1 Marek Adamczyk. Improved analysis of the greedy algorithm for stochastic matching. *Inf. Process. Lett.*, 111(15):731–737, 2011.
- 2 Marek Adamczyk, Fabrizio Grandoni, and Joydeep Mukherjee. Improved approximation algorithms for stochastic matching. In *Algorithms-ESA 2015*, pages 1–12. Springer, 2015.
- 3 Marek Adamczyk, Maxim Sviridenko, and Justin Ward. Submodular Stochastic Probing on Matroids. In *STACS*, pages 29–40, 2014.
- 4 Arash Asadpour and Hamid Nazerzadeh. Maximizing stochastic monotone submodular functions. *Management Science*, 62(8):2374–2391, 2016.
- 5 Arash Asadpour, Hamid Nazerzadeh, and Amin Saberi. Stochastic submodular maximization. In *International Workshop on Internet and Network Economics*, pages 477–489. Springer, 2008. Full version appears as [4].
- 6 Itai Ashlagi and Alvin E. Roth. New Challenges in Multihospital Kidney Exchange. *American Economic Review*, 102(3):354–59, 2012.
- 7 Nikhil Bansal, Anupam Gupta, Jian Li, Julián Mestre, Viswanath Nagarajan, and Atri Rudra. When LP Is the Cure for Your Matching Woes: Improved Bounds for Stochastic Matchings. *Algorithmica*, 63(4):733–762, 2012.
- 8 Nikhil Bansal and Viswanath Nagarajan. On the Adaptivity Gap of Stochastic Orienteering. In *IPCO*, pages 114–125, 2014.
- 9 Alok Baveja, Amit Chavan, Andrei Nikiforov, Aravind Srinivasan, and Pan Xu. Improved Bounds in Stochastic Matching and Optimization. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA*, pages 124–134, 2015.
- 10 Anand Bhargat, Ashish Goel, and Sanjeev Khanna. Improved Approximation Results for Stochastic Knapsack Problems. In *SODA*, pages 1647–1665, 2011.
- 11 Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a Monotone Submodular Function Subject to a Matroid Constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011.
- 12 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Submodular Function Maximization via the Multilinear Relaxation and Contention Resolution Schemes. *SIAM J. Comput.*, 43(6):1831–1879, 2014. doi:10.1137/110839655.
- 13 Ning Chen, Nicole Immorlica, Anna R. Karlin, Mohammad Mahdian, and Atri Rudra. Approximating Matches Made in Heaven. In *Automata, Languages and Programming, 36th International Colloquium, ICALP 2009, Rhodes, Greece, July 5-12, 2009, Proceedings, Part I*, pages 266–278, 2009.
- 14 Brian C. Dean, Michel X. Goemans, and Jan Vondrák. Approximating the stochastic knapsack problem: The benefit of adaptivity. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 208–217. IEEE, 2004.
- 15 Brian C. Dean, Michel X. Goemans, and Jan Vondrák. Adaptivity and approximation for stochastic packing problems. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 395–404, 2005.
- 16 Amol Deshpande, Lisa Hellerstein, and Devorah Kletenik. Approximation Algorithms for Stochastic Boolean Function Evaluation and Stochastic Submodular Set Cover. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1453–1466, 2014. doi:10.1137/1.9781611973402.107.
- 17 Sudipto Guha and Kamesh Munagala. Approximation algorithms for budgeted learning problems. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 104–113. ACM, 2007.

- 18 Sudipto Guha and Kamesh Munagala. Multi-armed Bandits with Metric Switching Costs. In *Automata, Languages and Programming, 36th International Colloquium, ICALP 2009, Rhodes, Greece, July 5-12, 2009, Proceedings, Part II*, pages 496–507, 2009.
- 19 Anupam Gupta, Ravishankar Krishnaswamy, Marco Molinaro, and R. Ravi. Approximation Algorithms for Correlated Knapsacks and Non-martingale Bandits. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 827–836, 2011.
- 20 Anupam Gupta, Ravishankar Krishnaswamy, Viswanath Nagarajan, and R. Ravi. Approximation algorithms for stochastic orienteering. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 1522–1538, 2012.
- 21 Anupam Gupta and Viswanath Nagarajan. A Stochastic Probing Problem with Applications. In *Integer Programming and Combinatorial Optimization - 16th International Conference, IPCO 2013, Valparaíso, Chile, March 18-20, 2013. Proceedings*, pages 205–216, 2013.
- 22 Anupam Gupta, Viswanath Nagarajan, and Sahil Singla. Algorithms and adaptivity gaps for stochastic probing. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1731–1747. SIAM, 2016.
- 23 Anupam Gupta, Viswanath Nagarajan, and Sahil Singla. Adaptivity Gaps for Stochastic Probing: Submodular and XOS Functions. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1688–1702. SIAM, 2017.
- 24 Jian Li Hao Fu and Pan Xu. A PTAS for a Class of Stochastic Dynamic Programs. In *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2018*, 2018.
- 25 Lisa Hellerstein, Devorah Kletenik, and Patrick Lin. Discrete Stochastic Submodular Maximization: Adaptive vs. Non-adaptive vs. Offline. In *Algorithms and Complexity - 9th International Conference, CIAC 2015, Paris, France, May 20-22, 2015. Proceedings*, pages 235–248, 2015. doi:10.1007/978-3-319-18173-8_17.
- 26 Jian Li and Wen Yuan. Stochastic combinatorial optimization via poisson approximation. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 971–980, 2013. doi:10.1145/2488608.2488731.
- 27 Zhen Liu, Srinivasan Parthasarathy, Anand Ranganathan, and Hao Yang. Near-optimal algorithms for shared filter evaluation in data stream systems. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 133–146, 2008. doi:10.1145/1376616.1376633.
- 28 Will Ma. Improvements and Generalizations of Stochastic Knapsack and Multi-Armed Bandit Approximation Algorithms: Extended Abstract. In *SODA*, pages 1154–1163, 2014.
- 29 Julián Mestre. Greedy in approximation algorithms. In *European Symposium on Algorithms*, pages 528–539. Springer, 2006.
- 30 Alvin E. Roth, Tayfun Sönmez, and M.Ütku Ünver. Pairwise kidney exchange. *J. Econom. Theory*, 125(2):151–188, 2005. doi:10.1016/j.jet.2005.04.004.
- 31 Aviad Rubinfeld and Sahil Singla. Combinatorial prophet inequalities. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1671–1687. SIAM, 2017.

A Adaptivity Gap Lower Bound of 2 for Submodular Functions

Proof. As mentioned, the finite lower bound example is constructed by reducing the infinite example given in Section 3.2. However, this reduction loses the nice similarity properties of the graph so much more calculation is required in order to bound the strategies.

Let $0 < \epsilon < 1/2$ and D be the smallest integer such that $(1 - \epsilon)^D < \epsilon^2$. The ground set is the result of removing elements $e_{(k,l)}$ where $k + l > D$, that is $V \stackrel{\text{def}}{=} \{e_{(k,l)} : k, l \in \mathbb{Z}_{\geq 0}, k + l \leq D\}$ where each node is active with probability ϵ . The probing constraint

and the objective function f are naturally reduced to this set: a sequence of elements can be probed if they correspond to a (finite) path starting at $e_{(0,0)}$ in the given graph, and $f(S) \stackrel{\text{def}}{=} \sum_{k \in K(S)} (1 - \epsilon)^k$ where $K(S)$ is the set of (unique) first labels which now finite. Similarly as before, we will denote $\{e_{(k,0)}, e_{(k,1)}, \dots, e_{(k,D-k)}\}$ as the vertices on the *column* k .

We first show that any non-adaptive strategy has expectation at most 1. Let $\text{alg}(k)$ denote the additional expected value of the optimal non-adaptive strategy if the next probed element is $e_{(k,0)}$. We will inductively prove $\text{alg}(k) < (1 - \epsilon)^k$, which is sufficient for our claim. For the base case $k = D$, the inequality clearly holds since $\text{alg}(D) = \epsilon(1 - \epsilon)^D < (1 - \epsilon)^D$. For $0 \leq k < D$ let i be the second label of the last vertex probed on the column k .

$$\begin{aligned} \text{alg}(k) &= \max_{i=0}^{D-k} \left[(1 - \epsilon)^k \Pr[k \in K(\text{active})] + \text{alg}(k + i + 1) \right] \\ &= \max_{i=0}^{D-k} \left[(1 - \epsilon)^k (1 - (1 - \epsilon)^{i+1}) + \text{alg}(k + i + 1) \right] \\ &< \max_{i=0}^{D-k} \left[(1 - \epsilon)^k (1 - (1 - \epsilon)^{i+1}) + (1 - \epsilon)^{k+i+1} \right] = (1 - \epsilon)^k. \end{aligned}$$

This completes the induction and proves that non-adaptive strategies get at most 1.

Finally, we show that there exists an adaptive strategy with expected value at least $2 - O(\epsilon)$ for sufficiently small $\epsilon > 0$. This finalizes the proof since it implies a gap of 2 by taking $\epsilon \rightarrow 0$. The strategy is naturally reduced: first probe $e_{(0,0)}$ and after probing some $e_{(k,l)}$ terminate if $k + l = D$, otherwise probe $e_{(k+l+1,0)}$ if $e_{(k,l)}$ is active and $e_{(k,l+1)}$ if not. Let $\text{adap}(k)$ denote the expected value this strategy gets when the next probed element is $e_{(k,0)}$, for $0 \leq k \leq D$. For convenience, define $\text{adap}(D + i) \stackrel{\text{def}}{=} 0$ for all $i \geq 1$.

We prove by induction that $\text{adap}(k) > \frac{4-6\epsilon}{2-\epsilon}(1 - \epsilon)^k - 8\epsilon$, which is sufficient to finalize the proof since then $\text{adap}(0) > 2 - O(\epsilon)$. For k large enough that $\frac{4-6\epsilon}{2-\epsilon}(1 - \epsilon)^k < 8\epsilon$, the inequality clearly holds and presents our base case. Otherwise, $(1 - \epsilon)^k \geq 8\frac{2-\epsilon}{4-6\epsilon}\epsilon > 4\epsilon$. Let i be the second label of the last vertex probed on the column k and let A denote the set of active elements.

$$\begin{aligned} \text{adap}(k) &= \sum_{i=0}^{D-k} \Pr \left[v_{(k,i)} \in A, v_{(k,0)} \notin A, \dots, v_{(k,i-1)} \notin A \right] \left[(1 - \epsilon)^k + \text{adap}(k + i + 1) \right] \\ &= \sum_{i=0}^{D-k} (1 - \epsilon)^i \epsilon \left[(1 - \epsilon)^k + \text{adap}(k + i + 1) \right] \\ &= \epsilon \cdot \sum_{i=0}^{D-k} (1 - \epsilon)^{k+i} + \epsilon \cdot \sum_{i=0}^{D-k} (1 - \epsilon)^i \text{adap}(k + i + 1) \\ &= \epsilon \cdot \frac{1}{\epsilon} (1 - \epsilon)^k \left(1 - (1 - \epsilon)^{D-k+1} \right) + \epsilon \cdot \sum_{i=0}^{D-k} (1 - \epsilon)^i \cdot \text{adap}(k + i + 1). \end{aligned}$$

Using the induction hypothesis, we get

$$\begin{aligned} \text{adap}(k) &> (1 - \epsilon)^k - (1 - \epsilon)^{D+1} + \epsilon \sum_{i=0}^{D-k} (1 - \epsilon)^i \left(\frac{4-6\epsilon}{2-\epsilon} (1 - \epsilon)^{k+i+1} - 8\epsilon \right) \\ &= (1 - \epsilon)^k - (1 - \epsilon)^{D+1} + \epsilon \sum_{i=0}^{D-k} \frac{4-6\epsilon}{2-\epsilon} (1 - \epsilon)^{k+2i+1} - 8\epsilon^2 \sum_{i=0}^{D-k} (1 - \epsilon)^i \\ &= (1 - \epsilon)^k - (1 - \epsilon)^{D+1} + (1 - \epsilon)^{k+1} \frac{4-6\epsilon}{(2-\epsilon)^2} \left(1 - (1 - \epsilon)^{2(D-k+1)} \right) \\ &\quad - 8\epsilon \left(1 - (1 - \epsilon)^{D-k+1} \right). \end{aligned}$$

49:20 (Near) Optimal Adaptivity Gaps for Stochastic Multi-Value Probing

After dropping some positive summands and using $(1 - \epsilon)^D < \epsilon$ and $(1 - \epsilon)^k > \epsilon$, we get

$$\text{adap}(k) > (1 - \epsilon)^k - \epsilon^2 + (1 - \epsilon)^{k+1} \frac{4 - 6\epsilon}{(2 - \epsilon)^2} (1 - \epsilon^2) - 8\epsilon.$$

It is sufficient to prove

$$(1 - \epsilon)^k - \epsilon^2 - 8\epsilon + (1 - \epsilon)^{k+1} \frac{4 - 6\epsilon}{(2 - \epsilon)^2} (1 - \epsilon^2) > \frac{4 - 6\epsilon}{2 - \epsilon} (1 - \epsilon)^k - 8\epsilon.$$

Multiplying by $\frac{(2 - \epsilon)^2}{(1 - \epsilon)^k} > 0$, we get an equivalent statement to prove:

$$(2 - \epsilon)^2 - \epsilon^2 \cdot \frac{(2 - \epsilon)^2}{(1 - \epsilon)^k} + (1 - \epsilon)(4 - 6\epsilon)(1 - \epsilon^2) > (4 - 6\epsilon)(2 - \epsilon).$$

Finally, using $\epsilon^2 \frac{(2 - \epsilon)^2}{(1 - \epsilon)^k} < \epsilon^2 (2 - \epsilon)^2 \frac{1}{4\epsilon} = \epsilon + O(\epsilon^2)$ and expanding out, we note that the left-hand side is $8 - 15\epsilon + O(\epsilon^2)$, while the right-hand side is $8 - 16\epsilon + O(\epsilon^2)$. Therefore, the inequality holds for sufficiently small $\epsilon > 0$. This concludes the proof. \blacktriangleleft

B Proof of the k -Extendible Property for Set Extension

Let $\mathcal{F} \subseteq 2^T$ be a k -extendible system. For every $A \subseteq B \in \mathcal{F}$ and $E \subseteq T$ where $A \cup E \in \mathcal{F}$, there exists a set $Z \subseteq B \setminus A$ such that $|Z| \leq k \cdot |E|$ and $B \setminus Z \cup E \in \mathcal{F}$.

Proof. Enumerate the elements $E = \{e_1, \dots, e_r\}$ where $r \stackrel{\text{def}}{=} |E|$ and denote by $E_i \stackrel{\text{def}}{=} \{e_1, \dots, e_i\}$ for $0 \leq i \leq r$. Initialize $Z_0 \stackrel{\text{def}}{=} \emptyset$ and consider the following procedure to construct Z_1, Z_2, \dots, Z_r that satisfies the invariants $A \subseteq B \setminus Z_i$, $B \setminus Z_i \cup E_i \in \mathcal{F}$ and $|Z_i| \leq k \cdot i$.

In the i^{th} step we have that $A \cup E_{i-1} \cup \{e_i\} \in \mathcal{F}$ by downward-closeness and $A \cup E_{i-1} \subseteq B \setminus Z_{i-1} \cup E_{i-1}$ by the induction hypothesis. Hence by k -extendibility we can find $Z' \subseteq B \setminus (Z_{i-1} \cup A \cup E_{i-1})$ with $|Z'| \leq k$ and where $(B \setminus Z_{i-1} \cup E_{i-1}) \setminus Z' \cup \{e_i\} = B \setminus (Z_{i-1} \cup Z') \cup E_i \in \mathcal{F}$. Set $Z_i \stackrel{\text{def}}{=} Z_{i-1} \cup Z'$ and note that $|Z_i| \leq |Z_{i-1}| + |Z'| \leq (i - 1) \cdot k + k = i \cdot k$. Furthermore, already deduced that $B \setminus Z_i \cup E_i \in \mathcal{F}$ and finally $A \subseteq B \setminus Z_i = B \setminus Z_{i-1} \setminus Z'$ since $Z' \cap A = \emptyset$. We satisfied all stipulations of the induction, hence we report Z_r as the solution. \blacktriangleleft

C Proof of Claim 12

\triangleright **Claim 12.** Consider an algorithm that goes over the odd numbered buckets in decreasing order of weights and selects the *maximum* set from class $j(i)$ in bucket i such that the resulting set is still feasible in \mathcal{F} . (After a set in a class is selected, it gets fixed for all smaller choices.) The finally chosen set has value at least $\frac{1}{4} \sum_{i \text{ is odd}} 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)})$.

Proof. The intuition is that for a k -extendible system by Section 4 any selected member can “hurt” at most k members from lower buckets. Since we only consider odd numbered buckets, two types in different buckets differ in their weights by at least a factor of $2^{2 \log k} = k^2$. Thus, losing k types of lower weight should not significantly impact the value.

Let ℓ be the random variable denoting the leaf reached by the random walk on the decision tree \mathcal{T} , and let R be the random set of elements seen by the random-walk non-adaptive strategy on this path. Furthermore, let A_i denote the set of elements picked by the non-adaptive strategy with respect to $f_{j(i)}$, let $A'_i \subseteq A_i$ be the set of elements picked by our greedy-optimal non-adaptive strategy from bucket i , and let $A'_{<i}$ denote $\bigcup_{i' < i : i' \text{ is odd}} A_{i'}$.

In other words, $A'_{<i}$ is the greedy-optimal solution up to bucket number i and A'_i is the maximum subset of A_i such that $A'_i \cup A'_{<i} \in \mathcal{F}$. Note that A_i , A'_i and $A'_{<i}$ are random variables depending on ℓ and R .

Using Section 4 on the k -extendible system \mathcal{F} with the preconditions $\emptyset \cup A'_{<i} \in \mathcal{F}$ and $\emptyset \subseteq A_i$, there exists a set Z with $|Z| \leq k \cdot |A'_{<i}|$ such that $A_i \setminus Z \in \mathcal{F}$. Hence, we have

$$|A'_i| \geq |A_i \setminus Z| \geq |A_i| - k \cdot |A'_{<i}|.$$

Multiplying by $2^{j(i)}$ and summing over all odd i gives

$$\begin{aligned} \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A'_i| &\geq \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A_i| - k \cdot \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A'_{<i}| \\ &= \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A_i| - k \cdot \sum_{i \text{ is odd}} |A'_i| \sum_{i' > i : i' \text{ is odd}} 2^{j(i')}. \end{aligned} \quad (15)$$

Now, since every bucket i contains $2 \log k$ classes, where two successive class weights differ by a factor of 2, we know

$$2^{j(i+2)} \leq \frac{2^{j(i)}}{k^2}.$$

Combining this with Eq. (15) gives

$$\begin{aligned} \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A'_i| &\geq \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A_i| - k \cdot \sum_{i \text{ is odd}} |A'_i| \sum_{i' > i : i' \text{ is odd}} \frac{2^{j(i'+2)}}{k^2} \\ &\geq \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A_i| - \sum_{i \text{ is odd}} |A'_i| \cdot 2^{j(i)}, \end{aligned}$$

where the last inequality uses

$$\sum_{i' > i : i' \text{ is odd}} 2^{j(i'+2)} = \sum_{i' \geq i : i' \text{ is odd}} 2^{j(i')} \leq 2 \cdot 2^{j(i)} \leq k \cdot 2^{j(i)}.$$

After rearranging,

$$\sum_{i \text{ is odd}} 2^{j(i)} \cdot |A'_i| \geq \frac{1}{2} \cdot \sum_{i \text{ is odd}} 2^{j(i)} \cdot |A_i|.$$

Notice that by definition of a class, each type in class $j(i)$ has weight at least $2^{j(i)-1}$. Using this fact and taking expectation over ℓ and R , we get

$$\begin{aligned} \text{alg}(\mathcal{T}, f) &\geq \mathbb{E}_{\ell, R} \left[\sum_{i \text{ is odd}} 2^{j(i)-1} \cdot |A'_i| \right] \\ &\geq \frac{1}{4} \mathbb{E}_{\ell, R} \left[\sum_{i \text{ is odd}} 2^{j(i)} \cdot |A_i| \right] = \frac{1}{4} \sum_{i \text{ is odd}} 2^{j(i)} \cdot \text{alg}(\mathcal{T}, f_{j(i)}), \end{aligned}$$

which finishes the proof of Claim 12. \triangleleft