Submodular Stochastic Probing on Matroids

Marek Adamczyk[1], Maxim Sviridenko[2], and Justin Ward[3]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, 00185 Rome, Italy, adamczyk@dis.uniroma1.it
[2]Yahoo! Labs, New York, NY 10036-3982, sviri@yahoo-inc.com
[3]Department of Computer Science, University of Warwick, Coventry CV4 7AL, United Kingdom, J.D.Ward@warwick.ac.uk

August 17, 2017

Abstract

In a *stochastic probing* problem we are given a universe E, and a probability p_e that each element $e \in E$ is active. We determine if an element is active by *probing* it, and whenever a probed element is *active*, we must permanently include it in our solution. Moreover, throughout the process we need to obey *inner* constraints on the set of elements taken into the solution, and *outer* constraints on the set of all probed elements. All previous algorithmic results in this framework have considered only the problem of maximizing a linear function of the active elements. In this paper we generalize the stochastic probing problem by providing the first constant factor approximation for maximizing a monotone submodular objective function.

For any $T \in (0,1]$, we give a $(1 - e^{-T} - o(1))/(T(k^{in} + k^{out}) + 1)$-approximation for the case in which we are given $k^{in} \geq 0$ matroids as inner constraints and $k^{out} \geq 1$ matroids as outer constraints. For $k = k^{in} + k^{out} \geq 1$, we show that the optimal value for T is given by $T = -1 - \frac{1}{k} - W_{-1}(-e^{-1-\frac{1}{k}}) \approx \sqrt{2/k} - 1/(3k)$, where W is the Lambert W function. We also obtain an improved $1/(k^{in} + k^{out})$-approximation for linear objective functions.

1 Introduction

Uncertainty in input data is a common feature of most practical problems, and research in finding good solutions (both experimental and theoretical) for such problems has a long history dating back to 1950 [8, 19]. We consider adaptive stochastic optimization problems in the framework of Dean et al. [21]. Here the solution is in fact a process, and the optimal one might even require larger than polynomial space to describe. Since the work of Dean et al. a number of such problems were introduced [15, 25, 28, 29, 6, 30, 20].

Recently, Gupta and Nagarajan [31] have given an abstract *stochastic probing* framework, which captures several adaptive stochastic problems and gives a unified view for Stochastic Matching [15] and Sequential Posted Pricing [11]. In their framework, we are given a universe E, where each element $e \in E$ is *active* with probability $p_e \in [0,1]$ independently. The only way to find out if an element is active, is to *probe* it. We call a probe *successful* if an element turns out to be active. On universe E, we execute an algorithm that probes the elements one-by-one. If an element is active, the algorithm must add it to the current solution. In this way, the algorithm gradually constructs a solution consisting of active elements. A key feature of the stochastic probing framework is the presence of two distinct types of constraints: a set of *inner* constraints, which restrict the elements that may be added to the current solution, after a successful probe, and a set of *outer* constraints, which restrict the elements that may be probed. As noted by Gupta and Nagarajan [31], these outer constraints, which must be enforced regardless of the outcome of a probe, are responsible for the richness of the framework.

1.1 Our results

Formally, we consider the problem in which we are given two independence systems of downward-closed sets: an *outer* independence system $(E, \mathcal{I}^{out})$ restricting the set of elements probed by the algorithm, and an *inner* independence system $(E, \mathcal{I}^{in})$, restricting the set of elements taken by the algorithm. Gupta and Nagarajan [31] considered many types of systems $\mathcal{I}^{in}$ and $\mathcal{I}^{out}$. Here, we focus on matroid intersections, i.e. on the special case in which $\mathcal{I}^{in}$ is an intersection of k^{in} matroids $\mathcal{M}_1^{in}, \ldots, \mathcal{M}_{k^{in}}^{in}$, and $\mathcal{I}^{out}$ is an intersection of k^{out} matroids $\mathcal{M}_1^{out}, \ldots, \mathcal{M}_{k^{out}}^{out}$. We always assume that $k^{out} \geq 1$ and $k^{in} \geq 0$.

Gupta and Nagarajan considered the case of weighted maximization, in which the goal is to obtain a probing policy that respects the inner and outer constraints and produces a solution that maximizes some linear function of the selected elements. Here we consider the generalized problem of optimizing a *submodular* function in the stochastic probing framework, subject to inner and outer matroid constraints. We provide the first constant-factor guarantees for this general problem.

Our main result is a new algorithm for the stochastic probing problem based on iterative randomized rounding of linear programs and the *continuous greedy algorithm* introduced by Vondrák [41] and further analyzed by Călinescu et al. [17]. We show that a novel, iterative, randomized rounding algorithm combined with the continuous greedy algorithm gives a $\frac{1-e^{-T}-o(1)}{T(k^{out}+k^{in})+1}$-approximation for maximizing a monotone submodular objective function in a stochastic probing problem of the sort described above, where $T \in (0,1]$ is the time at which the continuous greedy algorithm is stopped. By running the continuous greedy algorithm until $T = 1$ (as is standard) we obtain a $\frac{1-e^{-1}-o(1)}{k^{out}+k^{in}+1}$-approximation. However, we show that for $k = k^{out} + k^{in} > 1$, the optimal value of the stopping time is given by $T = -1 - \frac{1}{k} - W_{-1}(-e^{-1-\frac{1}{k}}) \approx \sqrt{\frac{2}{k}} - \frac{1}{3k}$, where W is the Lambert W function.

Additionally, we improve the bound of $\frac{1}{4(k^{in}+k^{out})}$ given by Gupta and Nagarajan [31] in the case of a linear objective. Specifically, we show that our iterative randomized rounding algorithm is a $\frac{1}{k^{in}+k^{out}}$-approximation for the stochastic probing problem with a linear objective function.

Finally, we show how our results may be generalized to include matchoid constraints, which generalize matchings in non-bipartite graphs.

1.2 Related work

The (offline) problem of maximizing a monotone submodular function subject to a cardinality constraint was studied by Nemhauser, Fisher, and Wolsey [37], who showed that the standard greedy algorithm is a $1-1/e$. Later Nemhauser and Wolsey [36] showed that any algorithm guaranteeing a $1-1/e+\epsilon$-approximate solution must evaluate the submodular function on a super-polynomial number of sets. Feige [22] considered a particular class of monotone submodular functions, given explicitly as coverage functions. He showed that even in this case, it is impossible to obtain a $1-1/e+\epsilon$-approximation algorithm for maximization subject to a single cardinality constraint, unless $P=NP$.

Fisher, Nemhauser, and Wolsey [24] showed that the standard greedy algorithm is only a $1/2$-approximation for the problem of maximizing a monotone submodular function subject to a single matroid constraint. Later, Călinescu et al. [17] gave a $1-1/e$-approximation algorithm for this problem, called the *continuous greedy algorithm*, originally developed by Vondrák [41] for the submodular welfare problem. The algorithm first approximately solves a continuous relaxation of the problem and then applies pipage rounding [2] to obtain an integral solution. Later, more sophisticated rounding approaches were developed, which handled various other constraints [12, 13], culminating in the development of *contention resolution schemes* [14], which provide a general rounding framework for combining various constraints. Here, we follow a similar direction, employing the continuous greedy algorithm to obtain a fractional solution and then iteratively rounding it to obtain an integral solution. Although our iterative rounding procedure is based on combinatorial properties of the underlying matroid constraints, as in [12, 13], the probing framework significantly complicates matters. In contrast to the offline setting, our rounding procedure corresponds to the iterative construction of a probing *policy*. Thus, it must cope with the outer constraints, the uncertainty in the input data, and the constraint that an element must be chosen if it is successfully probed. We show how to take all of these considerations into account to obtain an *adaptive* iterative rounding procedure, which in each iteration selects the next element to probe, given the outcomes of all previous iterations.

Recently, there has been increased interest in submodular maximization in the presence of uncertainty. Radlinski, Kleinberg, and Joachims [38] considered online algorithms for learning rankings, which is a special case of maximizing a coverage function subject to a cardinality constraint. More generally, Streeter and Golovin [40] have given an online algorithm for a resource allocation problem that generalizes the problem of maximizing a monotone submodular function subject to a knapsack constraint. They also give lower bounds on the regret incurred by any online algorithm for this problem, and show that their algorithm's expected regret matches these bounds up to a logarithmic factor. Golovin and Krause [27] introduced the notion of adaptive submodularity, which generalizes the standard definition of submodularity to adaptive planning problems, in which a feasible solution is a policy. They show that if a cardinality constrained problem is adaptive submodular, then a variant of the standard greedy algorithm is guaranteed to be constant-factor competitive with the optimal policy. More recently [26], they have generalized their approach to the problem of maximizing an adaptive submodular function subject to several matroid constraints. Asadpour, Nazerzadeh, and Saberi [6] considered submodular maximization in a stochastic setting in which each item is a random variable taking a non-negative real value. The goal is to choose a set of items that maximizes a submodular value function of the variables, subject to a single matroid constraint. They show that the *adaptivity gap*, which is the worst-case ratio between the expected values of an optimal adaptive and non-adaptive policy, is equal to $\frac{e}{e-1}$, and show how to obtain a non-adaptive policy whose value is at most $(1-1/e)^2$ times that of the optimal adaptive policy. In recent work [5] parallel to our own they have shown how to obtain in polynomial time, a non-adaptive policy that has value at least $(1-1/e-\epsilon)$ times the best adaptive policy, by utilizing the same improved bounds for the continuous greedy algorithm that we present in Appendix A. Their general approach is similar to ours in that they use a continuous relaxation f^+ (see Section 2.2 for details) to bound the value of the optimal adaptive policy. Their results then follow from the improved bound relating the value of the fractional solution produced by continuous greedy algorithm to the optimal value of f^+. Their general setting allows for a more general class of value functions but does not incorporate constraints on the elements that are probed. Thus, it may be viewed as a generalization of ours in the special case that $k^{out}=0$ (i.e. that there are no outer matroid constraints) and $k^{in}=1$. As our results hold only for $k^{out}\geq 1$, their work is complementary to that presented here.

Agrawal et al. [3, 4] considered stochastic optimization problems in which unknown demands may be correlated. They defined the correlation gap as the worst-case (over all possible distributions and marginals values) of the ratio of the expected cost when a realization is drawn from a distribution with some marginal values, to that of a random realization drawn from an independent distribution with the same marginal values. They showed that the correlation gap is bounded for a variety of functions. In particular, they show that for submodular functions, the correlation gap is bounded is at most $\frac{e}{e-1}$. Yan [44] considered the correlation gap in the context of submodular maximization problems, in particular the design of sequential posted pricing mechanisms. Here, as in our setting, each item may or may not be present with some marginal probability and the correlation gap of f with respect to a set of constraints measures the worst case ratio between the value of f on a random set drawn from any distribution with some marginal values to the value of f on a random set drawn from the independent distribution with these marginal values. In our setting, items will always be active independently. However, we are interested in obtaining a near-optimal *policy* for choosing items, and the set of choices made by the optimal policy will necessarily be highly correlated. Thus, we must address similar issues to those raised in [3, 44]. We discuss these issues in detail in Subsection 2.2.

1.3 Applications

We now discuss two concrete applications of the stochastic probing framework, based on stochastic matching and sequential posted pricing mechanisms, respectively.

On-line dating and related exchange problems [15] Consider an online dating service. For each pair of users, machine learning algorithms estimate the probability that they will form a happy couple. However, only after a pair meets do we know for sure if they were successfully matched (and together leave the dating service). Users have individual patience numbers that bound how many unsuccessful dates they are willing to go on until they will leave the dating service forever. The objective of the service is to maximize the number of successfully matched couples.

The related *stochastic matching problem* was introduced by Chen et al. [15], who showed that the greedy strategy gives a 1/4-approximation for the unweighted case. The authors also show that the simple greedy approach gives no constant approximation in the weighted case. Their bound for the unweighted case was later improved to 1/2 by Adamczyk [1]. Bansal et al. [7] gave 1/3 and 1/4-approximations for weighted stochastic matching in bipartite and general graphs, respectively.

To model this as a stochastic probing problem, users are represented as vertices V of a graph $G = (V, E)$, where edges represent couples of users. Set E of edges is our universe on which we make probes, with p_e being the probability that a couple $e = (u_1, u_2)$ forms a happy couple after a date. The inner constraints are matching constraints — a user can be in at most one couple —, and outer constraints are b–matching — we can probe at most $t(u)$ edges adjacent to user u, where $t(u)$ denotes the patience of u. Both inner and outer constraints are intersections of two matroids for bipartite graphs. In Section 5 we discuss a generalization of our approach that allows us model matching constraints in non-bipartite graphs, as well. In the weighted bipartite case, we obtain a 1/4-approximation.

Bayesian mechanism design [31] Consider the following mechanism design problem. There are n agents and a single seller providing a certain service. Agent i's value for receiving service is v_i, drawn independently from a distribution D_i over set $\{0, 1, \ldots, B\}$. The valuation v_i is private, but the distribution D_i is known. The seller can provide service only for a subset of agents that belongs to system $\mathcal{I} \in 2^{[n]}$, which specifies feasibility constraints. A mechanism accepts bids of agents, decides on a subset of agents to serve, and sets individual prices for the service. A mechanism is called truthful if agents bid their true valuations. Myerson's theory of virtual valuations yields *truthful* mechanisms that maximize the expected revenue of a seller, although they sometimes might be impractical. On the other hand, practical mechanisms are often non-truthful.

The Sequential Posted Pricing Mechanism (SPM) introduced by Chawla et al. [11] and subsequently studied by Yan [44], and Kleinberg and Weinberg [33] gives a nice trade-off — it is truthful, simple to implement, and gives near-optimal revenue. An SPM offers each agent a "take-it-or-leave-it" price for the service. Since after a refusal a service won't be provided, it is easy to see that an SPM is a truthful mechanism.

To see an SPM as a stochastic probing problem, we consider a universe $E = [n] \times \{0, 1, \dots, B\}$, where element (i, c) represents an offer of price c to agent i. The probability that i accepts the offer is $\mathbb{P}[v_i \geq c]$, and the seller earns c then. Obviously, we can make only one offer to an agent, so outer constraints are given by a partition matroid; making at most one probe per agent also overcomes the problem that probes of $(i, 1), ..., (i, B)$ are not independent. The inner constraints on universe $[n] \times \{0, 1, \dots, B\}$ are simply induced by constraints $\mathcal{I}$ on $[n]$.

Using this reduction, Gupta and Nagarajan [31] gave an LP relaxation for any single-seller Bayesian mechanism design problem. Provided that we can optimize over $\mathcal{P}(\mathcal{I})$, the LP can be used to construct an efficient SPM. Moreover, the approximation guarantee of the constructed SPM is with respect to the optimal mechanism, which needs not be an SPM.

In the case where constraints $\mathcal{I}$ are an intersection of k matroids the resulting SPM is a $\frac{1}{4(k+1)}$-approximation [31]. Here, we give an improved approximation algorithm with a factor-$\frac{1}{k+1}$ guarantee. In particular, when $k = 1$ we obtain a 1/2-approximating, matching previous results [11, 33].

2 Preliminaries

For set $S \subseteq E$ and element $e \in E$ we use $S + e$ to denote $S \cup \{e\}$, and $S - e$ to denote $S \setminus \{e\}$. For set $S \subseteq E$ we shall denote by $\mathbf{1}_S$ a characteristic vector of set S, and for a single element e we shall write $\mathbf{1}_e$ instead of $\mathbf{1}_{\{e\}}$.

2.1 Matroids and polytopes

Let $\mathcal{M} = (E, \mathcal{I})$ be a matroid, where E is the universe of elements and $\mathcal{I} \subseteq 2^E$ is a family of independent sets. For element $e \in E$, we shall denote the matroid $\mathcal{M}$ with e contracted by $\mathcal{M}/e$, i.e. $\mathcal{M}/e = (E - e, \{S \subseteq E - e \mid S + e \in \mathcal{I}\})$.

The following lemma is a slightly modified[1] basis exchange lemma, which can be found in [39].

Lemma 2.1. *Let $A, B \in \mathcal{I}$ and $|A| = |B|$. There exists a bijection $\phi : A \to B$ such that: 1) $\phi(e) = e$ for every $e \in A \cap B$, 2) $B - \phi(e) + e \in \mathcal{I}$.*

We shall use the following corollary, where we consider independent sets of possibly different sizes.

Corollary 2.2. *Let $A, B \in \mathcal{I}$. We can find assignment $\phi_{A,B} : A \mapsto B \cup \{\perp\}$ such that:*

1. *$\phi_{A,B}(e) = e$ for every $e \in A \cap B$,*
2. *for each $f \in B$ there exists at most one $e \in A$ for which $\phi_{A,B}(e) = f$,*
3. *for $e \in A \setminus B$, if $\phi_{A,B}(e) = \perp$ then $B + e \in \mathcal{I}$, otherwise $B - \phi_{A,B}(e) + e \in \mathcal{I}$.*

[1] The difference is that we do not assume that A, B are bases, but independent sets of the same size.

Proof. Suppose $|A| \geq |B|$. We use the matroid augmentation property a sufficient number of times to extend B into $B' \in \mathcal{I}$ with $|B'| = |A|$ using elements of $A \setminus B$. From Lemma 2.1 we get a bijection $\phi : A \mapsto B'$. For each $e \in A$, we set $\phi_{A,B}(e) = \perp$, if $\phi(e) \in B' \setminus B$, and $\phi_{A,B}(e) = \phi(e)$ otherwise. Case $|A| \leq |B|$ is similar. $\square$

We consider optimization over *matroid polytopes* which have the general form:

$$\mathcal{P}(\mathcal{M}) = \left\{ x \in \mathbb{R}^E_{\geq 0} \;\middle|\; \forall_{A \in \mathcal{I}} \sum_{e \in A} x_e \leq r_{\mathcal{M}}(A) \right\},$$

where $r_{\mathcal{M}}$ is the rank function of $\mathcal{M}$. We know (see, e.g. [39]) that the matroid polytope $\mathcal{P}(\mathcal{M})$ is equivalent to the convex hull of $\{\mathbf{1}_A \mid A \in \mathcal{I}\}$, i.e. characteristic vectors of all independent sets of $\mathcal{M}$. Thus, we can represent any $x \in \mathcal{P}(\mathcal{M})$ as $x = \sum_{i=1}^m \beta_i \cdot \mathbf{1}_{B_i}$, where $B_1, \ldots, B_m \in \mathcal{I}$ and $\beta_1, \ldots, \beta_m$ are non-negative weights such that $\sum_{i=1}^m \beta_i = 1$. We shall call sets $B_1, \ldots, B_m$ a *support* of x in $\mathcal{P}(\mathcal{M})$. Cunningham [18] showed that membership in a matroid polytope can be decided in strongly-polynomial time. Moreover, for any $x \in \mathcal{P}(\mathcal{M})$, Cunningham's algorithm returns the representation $\sum_{i=1}^m \beta_i \cdot \mathbf{1}_{B_i}$ in strongly polynomial time.

2.2 Submodular functions

A set function $f : 2^E \mapsto \mathbb{R}_{\geq 0}$ is *submodular*, if for any two subsets $S, T \subseteq E$ we have $f(S \cup T) + f(S \cap T) \leq f(S) + f(T)$. We call function f *monotone*, if for any two subsets $S \subseteq T \subseteq E : f(S) \leq f(T)$. Without loss of generality, we shall assume that $f(\emptyset) = 0$. We let $f_S(e) = f(S + e) - f(S)$. Then, f is submodular if and only if it has the property of *diminishing returns*: $f_T(e) \leq f_S(e)$ for all $S \subseteq T$, and $e \notin T$.

2.2.1 Multilinear extension

We consider the *multilinear extension* $F : [0,1]^E \mapsto \mathbb{R}_{\geq 0}$ of f, whose value at a point $y \in [0,1]^E$ is given by

$$F(y) = \sum_{A \subseteq E} f(A) \prod_{e \in A} y_e \prod_{e \notin A} (1 - y_e).$$

Note that $F(\mathbf{1}_A) = f(A)$ for any set $A \subseteq E$, so F is an extension of f from the discrete domain 2^E into a real domain $[0,1]^E$. For $y \in [0,1]^E$ let $R(y)$ denote a random subset $A \subseteq E$ that is constructed by taking each element $e \in E$ independently with probability y_e. Then, $F(y) = \mathbb{E}[f(R(y))]$. Following this interpretation, Călinescu et al. [17] show that $F(y)$ can be estimated to any desired accuracy in polynomial time by using a sampling procedure.

Additionally, they show that F has the following properties, which we shall make use of in our analysis:

Lemma 2.3. *The multilinear extension F is linear along the coordinates, i.e. for any point $x \in [0,1]^E$, any element $e \in E$, and any $\xi \in [-1,1]$ such that $x + \xi \cdot \mathbf{1}_e \in [0,1]^E$, it holds that $F(x + \xi \cdot \mathbf{1}_e) - F(x) = \xi \cdot \frac{\partial F}{\partial y_e}(x)$, where $\frac{\partial F}{\partial y_e}(x)$ is the partial derivative of F in direction y_e at point x.*

Lemma 2.4. *If $F : [0,1]^E \mapsto \mathbb{R}$ is a multilinear extension of monotone submodular function $f : 2^E \mapsto \mathbb{R}$, then 1) function F has second partial derivatives everywhere; 2) for each $e \in E$, $\frac{\partial F}{\partial y_e} \geq 0$ everywhere; 3) for any $e_1, e_2 \in E$ (possibly equal), $\frac{\partial^2 F}{\partial y_{e_1} \partial y_{e_2}} \leq 0$, which means that $\frac{\partial F}{\partial y_{e_2}}$ is non-increasing with respect to y_{e_1}.*

2.2.2 Continuous greedy algorithm

In [17] the authors utilized the multilinear extension in order to maximize a submodular monotone function over a matroid constraint. They showed that the continuous greedy algorithm finds a $(1 - 1/e)$-approximate maximum of the above extension F over any solvable, downward closed polytope. In the special case of the matroid polytope, they show how to apply pipage rounding [2] to the fractional solution to obtain an integral solution. Later, Feldman et al. [23] developed the *measured* continuous greedy algorithm, which gives improved approximations in a variety of cases. In the case of monotone submodular functions, Feldman et al. show that stopping the measured continuous greedy algorithm at time $T \in [0,1]$ yields a solution x satisfying $F(x) \geq (1 - e^{-T})OPT$ and $x/T \in \mathcal{P}$, where OPT is the optimal value attained by f on integral solutions in $\mathcal{P}$. They note that these particular guarantees hold for the standard continuous greedy algorithm, as well. Because these guarantees are sufficient for our purposes, we shall focus on the standard continuous greedy algorithm.

Another extension of f studied in [9] is given by:

$$f^+(y) = \max\left\{ \sum_{A \subseteq E} \alpha_A f(A) \,\middle|\, \sum_{A \subseteq E} \alpha_A \leq 1,\ \forall A \subseteq E : \alpha_A \geq 0,\ \forall j \in E : \sum_{A : j \in A} \alpha_A \leq y_j \right\} \quad (1)$$

Intuitively, the solution $(\alpha_A)_{A \subseteq E}$ above represents the distribution over 2^E that maximizes the value $\mathbb{E}[f(A)]$ subject to the constraint that its marginal values satisfy $\mathbb{P}[i \in A] \leq y_i$. The value $f^+(y)$ is then the value of $\mathbb{E}[f(A)]$ under this distribution, while the value of $F(y)$ is the value of $\mathbb{E}[f(A)]$ under the particular distribution that places each element i in A independently. However, the following allows us to relate the value of F on the solution of the continuous greedy algorithm to the optimal value of the relaxation f^+.

Lemma 2.5. *Let f be a submodular function with multilinear extension F, and let $\mathcal{P}$ be any downward closed polytope. Let x be solution produced by the continuous greedy algorithm on F and $\mathcal{P}$ until time $T \in (0,1]$. Then:*

1. *$x/T \in \mathcal{P}$.*
2. *$F(x) \geq (1 - e^{-T} - o(1)) \max_{y \in \mathcal{P}} f^+(y)$.*

This follows from a simple modification of the continuous greedy analyses of [17], provided by Vondrák [42], together with the observations from [23].

2.3 Overview of the iterative randomized rounding approach

We now give a description of the general rounding approach that we employ in both the linear and submodular case. We consider an instance of a stochastic probing problem, with objective function f, outer matroid constraints $\mathcal{M}_j^{out}$, where $1 \leq j \leq k^{out}$ and inner matroid constraints $\mathcal{M}_j^{in}$, where $1 \leq j \leq k^{in}$. Our rounding procedure is guided by the solution of the following mathematical programming relaxation, where f^+ is the relaxation given in (1):

$$\begin{aligned} \text{maximize } & f^+(p \cdot x) \\ \text{subject to: } & x \in \mathcal{P}(\mathcal{M}_j^{out}), \quad && 1 \leq j \leq k^{out} \\ & p \cdot x \in \mathcal{P}(\mathcal{M}_j^{in}), && 1 \leq j \leq k^{in} \\ & x \in [0,1]^E \end{aligned} \tag{2}$$

We now show that the solution of the relaxation (2) is an upper bound on the expected value of the optimal feasible strategy for the related stochastic probing problem. Henceforth, we let x^+ denote the optimal solution to (2).

Lemma 2.6. *Let OPT be the optimal feasible strategy for the stochastic probing problem in our general setting, then, $\mathbb{E}[f(OPT)] \leq f^+(p \cdot x^+)$.*

Proof. We construct a feasible solution x of (2) by setting $x_e = \mathbb{P}[OPT \text{ probes } e]$. First, we show that this is indeed a feasible solution of (2). Since OPT is a feasible strategy, the set of elements Q probed by any execution of OPT is always an independent set of each outer matroid $\mathcal{M} = (E, \mathcal{I}_j^{out})$, i.e. $\forall_{j \in [k^{out}]} Q \in \mathcal{I}_j^{out}$. Thus, for any $j \in [k^{out}]$, the vector $\mathbb{E}[\mathbf{1}_Q] = x$ may be represented as a convex combination of vectors from $\{\mathbf{1}_A \mid A \in \mathcal{I}_j^{out}\}$, and so $x \in \mathcal{P}(\mathcal{M}_j^{out})$. Analogously, the set of elements S that were successfully probed by OPT satisfy $\forall_{j \in [k^{in}]} S \in \mathcal{I}_j^{in}$ for every possible execution of OPT. Hence, for any $j \in [k^{in}]$ the vector $\mathbb{E}[\mathbf{1}_S] = p \cdot x$ may be represented as a convex combination of vectors from $\{\mathbf{1}_A \mid A \in \mathcal{I}_j^{in}\}$ and so $p \cdot x \in \mathcal{P}(\mathcal{M}_j^{in})$.

The value $f^+(p \cdot x)$ gives the maximum value of $\mathbb{E}_{S \sim \mathcal{D}}[f(S)]$ over all distributions $\mathcal{D}$ satisfying $\mathbb{P}_{S \sim \mathcal{D}}[e \in S] = x_e p_e$. The solution S returned by OPT satisfies $\mathbb{P}[e \in S] = \mathbb{P}[\text{OPT probes } e] p_e = x_e p_e$. Thus, OPT defines one such distribution, and so we have $\mathbb{E}[f(OPT)] \leq f^+(p \cdot x) \leq f^+(p \cdot x^+)$. $\square$

In the case of a submodular objective function f, it is NP-hard to solve the relaxation (2) exactly. Thus, we shall use Lemma 2.5 to obtain an approximate solution; we discuss the details of this approach in Section 4.

We now describe our general rounding procedure. We are given an instance of a stochastic probing problem over universe E, specified by a set of k^{in} inner matroids, k^{out} outer matroids, an objective function f, and a probability p_e for each $e \in E$. We first obtain a feasible solution x^0 to a relaxation (2), using either linear programming or the continuous greedy algorithm. Next, we iteratively round the solution, carrying out a single probe in each iteration. In each iteration we randomly select a single element $\bar{e}$ in the support of x to probe, choosing $\bar{e}$ with probability proportional to $x_{\bar{e}}$. We probe $\bar{e}$ and update S accordingly, then update the inner and outer constraints to obtain a new relaxation (of the form given in (2)) representing the remaining problem. Finally, we update x to obtain a feasible solution for this new relaxation. The algorithm terminates when there are no elements remaining in the support of x.

Let us now describe in more detail how the algorithm carries out the updates for a single step. Suppose that at some step of the algorithm, we select $\bar{e}$ to probe. We carry out the probe, adding $\bar{e}$ to S if we are successful. Next, we replace each outer matroid $\mathcal{M}_j^{out}$ with the contracted matroid $\mathcal{M}_j^{out}/\bar{e}$, to reflect the fact that $\bar{e}$ has been probed. If the probe succeeds, we must similarly update each inner matroid constraint, replacing $\mathcal{M}_j^{in}$ by $\mathcal{M}_j^{in}/\bar{e}$, to reflect the fact that $\bar{e}$ was taken by the algorithm. If the probe fails, we do not need to update the inner constraints. Finally, we remove $\bar{e}$ from E. This gives us a new relaxation of the form given in (2).

Next, we need to further update the solution x to obtain a feasible solution for this new relaxation. Let us describe how to perform a single update corresponding to an outer or inner matroid constraint, respectively.

Suppose that $\bar{e}$ was the element probed, and consider some outer matroid $\mathcal{M}_j^{out}$, where $1 \le j \le k^{out}$. Before selecting $\bar{e}$, we have $x \in \mathcal{P}(\mathcal{M}_j^{out})$. We can thus represent x as a convex combination: $x = \sum_{i=1}^{m} \beta_i^{out} \mathbf{1}_{B_i^{out}}$, where $B_1^{out}, \ldots, B_m^{out}$ are independent sets in $\mathcal{M}_j^{out}$. We modify x to obtain a solution x' such that $x' \in \mathcal{P}(\mathcal{M}_j^{out})$ with $x'_{\bar{e}} = 1$, as follows. First, we pick one set B_a^{out} with $\bar{e} \in B_a^{out}$ to guide the update process; we choose a set $B_a^{out} \ni \bar{e}$ at random with probability $\beta_a^{out}/x_{\bar{e}}$ (note that for any element e, $\sum_{a: e \in B_a^{out}} \beta_a^{out} = x_e$). Then, for any set B_b^{out}, let $i = \phi_{a,b}(\bar{e})$, where $\phi_{a,b}$ is the mapping from B_a^{out} into B_b^{out} given by Corollary 2.2. If $i = \bar{e}$, then $\bar{e} \in B_b^{out}$ and we do nothing. If $i = \perp$, then $B_b^{out} + \bar{e} \in \mathcal{M}_j^{out}$, and so we replace B_b^{out} by $B_b^{out} + \bar{e}$. Otherwise, we substitute B_b^{out} with $B_b^{out} - i + \bar{e}$ in the support of x. Each such substitution decreases the value of coordinate i by β_b. After performing all such substitutions, we obtain a vector $x' \in \mathcal{P}(\mathcal{M}_j^{out})$. The vector x' is a convex combination of independent sets all containing $\bar{e}$. Thus, we have $x'_{\bar{e}} = 1$, while for all $i \ne \bar{e}$, we have $x'_i = x_i - \delta_i$ for some $0 \le \delta_i \le x_i$.

Similarly, if $\bar{e}$ is successfully probed we must perform a support update for each inner matroid $\mathcal{M}_j^{in}$. Here, we proceed as in the case of the outer matroids, except now we have $p \cdot x \in \mathcal{M}_j^{in}$, and we choose a set $B_a^{in} \ni \bar{e}$ to guide the support update with probability $\beta_a^{in}/(x_{\bar{e}} p_{\bar{e}})$. We consider then the vector $y = p \cdot x$ instead of x. We obtain a vector $y' \in \mathcal{M}_j^{in}$ with $y'_{\bar{e}} = 1$ and for all $i \ne \bar{e}$, $y'_i = y_i - \delta_i p_i = (x_i - \delta_i) p_i$ for some $0 \le \delta_i \le x_i$.

We now show how to combine these individual matroid updates to obtain a feasible solution for the updated relaxation. Consider some element $i \ne \bar{e}$. Each matroid update requires decreasing x_i by some value $0 \le \delta_i \le x_i$. We decrease each such x_i by the *maximum* such δ_i required by any of the $k^{out} + k^{in}$ updates, and call the resulting solution x'. Then, we have both $\{x'_i\}_{i \ne \bar{e}} \in \mathcal{P}(\mathcal{M}_j^{out}/\bar{e})$ for each $1 \le j \le k^{out}$ and $\{x'_i p_i\}_{i \ne \bar{e}} \in \mathcal{P}(\mathcal{M}_j^{in}/\bar{e})$ for each $1 \le j \le k^{in}$.

It remains to remove $\bar{e}$ from E. Note that once our algorithm sets some coordinate x_i to 0, i will never be probed, and so x_i will remain 0 for the remainder of the algorithm. Thus, in order to simplify our discussion, we do not explicitly remove $\bar{e}$ from E in each iteration. Rather, we just set $x_{\bar{e}}$ to 0. Thus, all solutions we consider will be vectors in $[0,1]^E$. Note that the coordinates of our current fractional solution x are always decreasing throughout the algorithm, either due to a matroid update step, or because we set $x_{\bar{e}}$ to 0 after probing $\bar{e}$.

We now turn to the general analysis of our rounding procedure. In order to analyze the approximation performance of our algorithm, we shall keep track of a *potential* value z, depending on the current solution x, which intuitively represents the expected value of the remaining fractional solution x, given the choices that have been made so far. Initially, our potential z will be at least some constant fraction of the optimal value of (2), and in the final step z will be equal to 0. Let x^t, S^t, and z^t be the current value of x, z, and S at the end of the tth iteration, and let x^+ be the optimal solution of (2). Our analysis proceeds by first showing that $z^0 \geq \beta \cdot f^+(x^+)$. for some constant $\beta \in (0, 1]$. Then, we consider an arbitrary step $t+1$ and analyze the expected decrease $\mathbb{E}[z^t - z^{t+1}]$ in the potential due to this step. We bound this decrease in terms of the expected increase $\mathbb{E}[S^{t+1} - S^t]$ in the probed solution S at this step, showing that:

$$\alpha \cdot \mathbb{E}[z^t - z^{t+1}] \leq \mathbb{E}[f(S^{t+1}) - f(S^t)],$$

for some $\alpha < 1$. Then, we employ the following Lemma to conclude that the algorithm is an $\alpha\beta$-approximation in expectation. The proof is based on Doob's optional stopping theorem for martingales. Hence, we need to employ language from martingale theory, such as stopping time and filtration. We provide the necessary definitions, together with a statement of Doob's theorem in Appendix B. See [43] for extended background on martingale theory.

Lemma 2.7. *Suppose our algorithm runs for τ iterations and that the potential function z satisfies $z^0 \geq \beta \cdot f^+(p \cdot x^+)$ and $z^\tau = 0$. Let $(\mathcal{F}_t)_{t \geq 0}$ be the filtration associated with our iterative algorithm, where $\mathcal{F}_i$ represents all information available after the ith iteration. Finally, suppose that in each step $t+1$ of our iterative rounding procedure:*

$$\mathbb{E}\left[f\left(S^{t+1}\right) - f\left(S^t\right) \middle| \mathcal{F}_t \right] \geq \alpha \cdot \mathbb{E}\left[z^t - z^{t+1} \middle| \mathcal{F}_t \right].$$

Then, the final solution S^τ produced by the algorithm satisfies $\mathbb{E}[f(S^\tau)] \geq \alpha\beta \cdot \mathbb{E}[f(OPT)]$.

Proof. Define the random variable $G_t = f(S^t) - f(S^{t-1})$, representing the gain in f in iteration t, and similarly let $L_t = z^{t-1} - z^t$ represent the loss in z in iteration t. Additionally, define $G_0 = L_0 = 0$. For each $0 \leq t \leq \tau$, define $D_t = G_t - \alpha \cdot L_t$. The sequence of random variables $X_t = (D_0 + D_1 + ... + D_t)$, $t \geq 0$, forms a sub-martingale, i.e.

$$\mathbb{E}\left[X_{t+1} \middle| \mathcal{F}_t \right] = \sum_{i=0}^{t} D_i + \mathbb{E}\left[G_{t+1} - \alpha \cdot L_{t+1} \middle| \mathcal{F}_t \right] \geq \sum_{i=0}^{t} D_i.$$

Let τ be the step in which the algorithm terminates, i.e. $\tau = \min\{t \mid x^t = 0^E\}$. Then, the event $\tau = t$ depends only on $\mathcal{F}_0, \ldots, \mathcal{F}_t$, so τ is a stopping time. Also, by the definition of the algorithm $x^\tau = 0^E$. It is easy to verify that all the assumptions of Doob's optional stopping theorem are satisfied, and from this theorem we get that $\mathbb{E}[\sum_{i=0}^{\tau} D_i] \geq \mathbb{E}[D_0]$. Since $D_0 = 0$, we have

$$0 \leq \mathbb{E}\left[\sum_{i=0}^{\tau} D_i\right] = \mathbb{E}\left[\sum_{i=0}^{\tau} G_i - \alpha \cdot \sum_{i=0}^{\tau} L_i\right] = \mathbb{E}\left[\sum_{i=0}^{\tau} G_i\right] - \alpha \cdot \mathbb{E}\left[\sum_{i=0}^{\tau} L_i\right].$$

Finally, we note that

$$\sum_{i=0}^{\tau} G_i = \sum_{i=1}^{\tau} \left[f(S^i) - f(S^{i-1})\right] = f(S^\tau) - f(\emptyset) = f(S^\tau)$$

and so $\mathbb{E}[\sum_{i=0}^{\tau} G_i] = \mathbb{E}[f(S^\tau)]$, and similarly,

$$\sum_{i=0}^{\tau} L_i = \sum_{i=1}^{\tau} \left[z^{i-1} - z^i\right] = z^0 - z^\tau \geq \beta \cdot f^+(p \cdot x^+) - 0 = \beta \cdot f^+(p \cdot x^+),$$

and so from Lemma 2.6, $\mathbb{E}[f(S^\tau)] \geq \alpha \cdot \mathbb{E}[\sum_{i=0}^{\tau} L_i] \geq \alpha\beta \cdot f^+(p \cdot x^+) \geq \alpha\beta \cdot \mathbb{E}[f(OPT)]$. $\square$

Henceforth, we will implicitly condition on all information $\mathcal{F}_t$ available to the algorithm just before it makes step $t+1$. That is, when discussing step $t+1$ of the algorithm, we write simply $\mathbb{E}[\cdot]$ in place of $\mathbb{E}[\cdot | \mathcal{F}_t]$.

3 Linear stochastic probing

We now consider the linear setting, in which we are given a weight w_e and a probability p_e for each element $e \in E$ and our objective $f(S)$ is simply $\sum_{e \in S} w_e$. We note that because f is linear, we in fact have

$$f^+(p \cdot x) = \sum_{e \in E} w_e p_e x_e.$$

Thus, (2) is a linear maximization problem. We can solve this problem exactly via standard linear programming techniques, using the results of Cunningham [18] and the ellipsoid algorithm. We then obtain an initial solution x^0 satisfying $f^+(p \cdot x^0) = f^+(p \cdot x^+)$.

At each step t, our algorithm randomly selects an element $\bar{e}$ to probe. Let $\Sigma^t = \sum_{e \in E} x_e^t$. Then, our algorithm chooses $\bar{e} = e$ with probability x_e^t / Σ^t. As discussed in the previous overview, the algorithm then probes $\bar{e}$ and updates the matroid constraints to reflect both the choice of $\bar{e}$ and the outcome of the probe. Finally, it updates x^t to obtain a new fractional solution x^{t+1} for the new set of constraints and removes $\bar{e}$ from the support of x.

We now turn to the analysis of the algorithm. Our potential z^t at step t will be given by:

$$z^t = \sum_{e \in E} w_e p_e x_e^t.$$

In particular, we have $z^0 = f^+(p \cdot x^0) = f^+(p \cdot x^+)$. Suppose that the algorithm terminates after τ steps. Then, $z^\tau = \sum_{e \in E} w_e p_e \cdot 0 = 0$. Hence, the conditions of Lemma 2.7 are satisfied with $\beta = 1$.

We now bound the expected loss $\mathbb{E}[z^t - z^{t+1}]$ in step $t+1$. In order to do this, we consider the value $\delta_i = p_i\left(x_i^t - x_i^{t+1}\right)$ for each $i \in E$. The decrease δ_i may be caused either by selecting i to probe, in case which we set x_i^{t+1} to 0, or by the matroid update step, in which we decrease several other coordinates of x^t to obtain x^{t+1}. Let us first consider the losses due to each matroid update.

Lemma 3.1. *Consider the update step performed for a given outer matroid $\mathcal{M}_j^{out}$ in step $t+1$, and let δ_i^{out} be the amount that x_i^t is decreased by this step. Then, $\mathbb{E}[\delta_i^{out}] \leq \frac{1}{\Sigma^t}\left(1 - x_i^t\right) x_i^t$.*

Proof. The expectation $\mathbb{E}[\delta_i^{out}]$ is over the random choice of an element $\bar{e}$ to probe and the random choice of an independent set to guide the update. Let $\mathcal{E}_a^{out}$ denote the event that the set B_a^{out} is chosen to guide a support update for $\mathcal{M}_j^{out}$.

In a given step, the probability that the set B_a^{out} is chosen to guide the support update is given by

$$\mathbb{P}\left[\mathcal{E}_a^{out}\right] = \sum_{e \in B_a^{out}} \frac{x_e^t}{\Sigma^t} \frac{\beta_a^{out}}{x_e^t} = \sum_{e \in B_a^{out}} \frac{\beta_a^{out}}{\Sigma^t} = \left|B_a^{out}\right| \frac{\beta_a^{out}}{\Sigma^t}.$$

Moreover, conditioned on the fact B_a^{out} was chosen, the probability that a particular element $e \in B_a^{out}$ was probed is uniform over the elements of B_a^{out}:

$$\mathbb{P}\left[e \text{ probed} \mid \mathcal{E}_a^{out}\right] = \mathbb{P}\left[e \text{ probed} \wedge \mathcal{E}_a^{out}\right] / \mathbb{P}\left[\mathcal{E}_a^{out}\right] = \frac{x_e^t}{\Sigma^t} \frac{\beta_a^{out}}{x_e^t} \Big/ \left|B_a^{out}\right| \frac{\beta_a^{out}}{\Sigma^t} = \frac{1}{\left|B_a^{out}\right|}. \quad (3)$$

We can write the expected decrease as $\mathbb{E}[\delta_i^{out}] = \sum_{a=1}^m \mathbb{P}\left[\mathcal{E}_a^{out}\right] \cdot \mathbb{E}\left[\delta_i^{out} \middle| \mathcal{E}_a^{out}\right]$. Note that for all $i \in B_a^{out}$, we have $\phi_{a,b}(i) = i$ for *every* set B_b^{out} such that $i \in B_b^{out}$. Thus, the support update will not change the current value of x_i^t for any $i \in B_a^{out}$, and so in fact

$$\mathbb{E}[\delta_i^{out}] = \sum_{a=1}^m \mathbb{P}\left[\mathcal{E}_a^{out}\right] \cdot \mathbb{E}\left[\delta_i^{out} \middle| \mathcal{E}_a^{out}\right] = \sum_{a : i \notin B_a^{out}} \mathbb{P}\left[\mathcal{E}_a^{out}\right] \cdot \mathbb{E}\left[\delta_i^{out} \middle| \mathcal{E}_a^{out}\right].$$

Now let us condition on taking B_a^{out} to guide the support update. Consider a set B_b^{out} containing i. If we remove i from B_b^{out}, and hence decrease x_i^t by β_b^{out}, it must be the case that we chose to probe the single element $\phi_{a,b}^{-1}(i) \in B_a^{out}$. As shown in (3), the probability that we probe this element is $\frac{1}{|B_a^{out}|}$. Hence

$$
\begin{aligned}
&\sum_{a:i\notin B_a^{out}} \mathbb{P}\left[\mathcal{E}_a^{out}\right]\cdot \mathbb{E}\left[\delta_i^{out}\middle|\,\mathcal{E}_a^{out}\right]\\
&= \sum_{a:i\notin B_a^{out}} \mathbb{P}\left[\mathcal{E}_a^{out}\right]\cdot \left(\sum_{b:i\in B_b^{out}} \beta_b^{out}\cdot \mathbb{P}\left[\phi_{a,b}^{-1}(i) \text{ is probed}\,\middle|\mathcal{E}_a^{out}\right]\right)\\
&\le \sum_{a:i\notin B_a^{out}} \mathbb{P}\left[\mathcal{E}_a^{out}\right]\cdot \left(\sum_{b:i\in B_b^{out}} \beta_b^{out}\cdot \frac{1}{|B_a^{out}|}\right)\\
&= \sum_{a:i\notin B_a^{out}} \mathbb{P}\left[\mathcal{E}_a^{out}\right]\cdot \frac{x_i^t}{|B_a^{out}|}\\
&= \sum_{a:i\notin B_a^{out}} \left|B_a^{out}\right|\frac{\beta_a^{out}}{\Sigma^t}\cdot\frac{x_i^t}{|B_a^{out}|} = \frac{1}{\Sigma^t}\sum_{a:i\notin B_a^{out}}\beta_a^{out}x_i^t = \frac{1}{\Sigma^t}\left(1-x_i^t\right)x_i^t. \qquad \square
\end{aligned}
$$

Lemma 3.2. *Consider the update step performed for a given inner matroid $\mathcal{M}_j^{in}$ in step $t+1$, and let δ_i^{in} be the amount that x_i^t is decreased by this step. Then, $\mathbb{E}[\delta_i^{in}] \le \frac{1}{\Sigma^t}\left(1-p_i x_i^t\right)x_i^t$.*

Proof. Because we only perform a support update when the probe of a chosen element is successful, the expectation $\mathbb{E}[\delta_i^{in}]$ is over the random result of the probe, as well as the random choice of element $\bar{e}$ to probe and the random choice of a base to guide the update. We proceed as in the case of Lemma 3.1, now letting $\mathcal{E}_a^{in}$ denote the event that the probe was successful and B_a^{in} is chosen to guide the support update. We have:

$$
\begin{aligned}
&\mathbb{P}\left[\mathcal{E}_a^{in}\right] = \sum_{e\in B_a^{in}} p_e\frac{x_e^t}{\Sigma^t}\frac{\beta_a^{in}}{p_e x_e^t} = \sum_{e\in B_a^{in}}\frac{\beta_a^{in}}{\Sigma^t} = \left|B_a^{in}\right|\frac{\beta_a^{in}}{\Sigma^t},\\
&\mathbb{P}\left[e \text{ probed}\mid \mathcal{E}_a^{in}\right] = \mathbb{P}\left[e\text{ probed}\wedge\mathcal{E}_a^{in}\right]/\mathbb{P}\left[\mathcal{E}_a^{in}\right] = p_e\frac{x_e^t}{\Sigma^t}\frac{\beta_a^{in}}{p_e x_e^t}\bigg/\left|B_a^{in}\right|\frac{\beta_a^{in}}{\Sigma^t} = \frac{1}{|B_a^{in}|}.
\end{aligned}
$$

By a similar argument as in Lemma 3.1 we then have that $\mathbb{E}[\delta_i^{in}]$ is at most:

$$
\begin{aligned}
\sum_{a:i\notin B_a^{in}} \mathbb{P}\left[\mathcal{E}_a^{in}\right]\cdot\left(\sum_{b:i\in B_b^{in}}\beta_b^{in}\cdot\frac{1}{|B_a^{in}|}\right) &= \sum_{a:i\notin B_a^{in}}\mathbb{P}\left[\mathcal{E}_a^{in}\right]\cdot\frac{x_i^t}{|B_a^{in}|}\\
&= \sum_{a:i\notin B_a^{out}}\left|B_a^{in}\right|\frac{1}{\Sigma^t}\beta_a^{in}\cdot\frac{x_i^t}{|B_a^{in}|} = \frac{1}{\Sigma^t}\sum_{a:i\notin B_a^{in}}\beta_a^{in}x_i^t = \frac{1}{\Sigma^t}(1-p_i x_i^t)x_i^t.
\end{aligned}
$$

$\square$

We perform the matroid updates sequentially for each of the k^{in} and k^{out} matroids to obtain a new solution x^{t+1}. Now, we consider the expected decrease of a single coordinate of x^t due to both the initial probing step, in which we decrease the probed element's coordinate to 0, and the matroid updates.

Lemma 3.3. *For each step $t+1$ in the iterative rounding procedure,*

$$\mathbb{E}[p_i\delta_i] = \mathbb{E}[p_i(x_i^t - x_i^{t+1})] \leq \frac{k^{in}+k^{out}}{\Sigma^t} p_i x_i^t$$

for all $i \in E$.

Proof. We must decrease x_i^t either by x_i^t, in the case that i is probed, or by the maximum value δ_i^{out} or δ_i^{in} required by any matroid update. Then, the total decrease in x_i^t is less than or equal to the *sum* of all the decreases required by the probing step and each individual update step. Thus, we have

$$\begin{aligned}
\mathbb{E}[p_i\delta_i] &\leq \mathbb{P}\left[i \text{ probed}\right] p_i x_i^t + k^{out} p_i \mathbb{E}[\delta_i^{out}] + k^{in} p_i \mathbb{E}[\delta_i^{in}] \\
&\leq \frac{x_i^t}{\Sigma^t} p_i x_i^t + k^{out} \frac{1}{\Sigma^t}(1-x_i^t) p_i x_i^t + k^{in} \frac{1}{\Sigma^t}(1-p_i x_i^t) p_i x_i^t \\
&\leq \frac{1}{\Sigma^t}\left(k^{out} p_i x_i^t + k^{in}(1-p_i x_i^t) p_i x_i^t\right) \\
&\leq \frac{1}{\Sigma^t}(k^{out}+k^{in}) p_i x_i^t,
\end{aligned}$$

where the second inequality follows from Lemmas 3.1 and 3.2 and the third one uses the fact that $k^{out} \geq 1$. □

Using Lemma 3.3 we can now prove our main result for linear stochastic probing.

Theorem 3.4. *For a linear objective function f, the solution S produced by our randomized rounding algorithm satisfies $\mathbb{E}[f(S)] \geq \frac{1}{k^{out}+k^{in}} \mathbb{E}[f(OPT)]$, where OPT is the solution produced by the optimal policy.*

Proof. Because z^t is a linear function of x^t, the expected total decrease of z in step $t+1$ is given by:

$$\mathbb{E}[z^t - z^{t+1}] = \sum_i w_i \mathbb{E}[p_i(x_i^t - x_i^{t+1})] \leq \frac{k^{out}+k^{in}}{\Sigma^t} \sum_i w_i p_i x_i^t,$$

where the inequality follows from Lemma 3.3.

On the other hand, the expected gain in $f(S)$ is

$$\sum_{e \in E} w_e p_e \mathbb{P}\left[e \text{ probed}\right] = \frac{1}{\Sigma^t} \sum_{e\in E} w_e p_e x_e^t \geq \frac{1}{k^{out}+k^{in}} \mathbb{E}[z^t - z^{t+1}].$$

Applying Lemma 2.7, with $\beta = 1$ and $\alpha = \frac{1}{k^{out}+k^{in}}$, the final solution S^τ produced by the algorithm satisfies $\mathbb{E}[f(S^\tau)] \geq \frac{1}{k^{out}+k^{in}} \mathbb{E}[f(OPT)]$. □

4 Submodular stochastic probing

We now consider the case in which the objective function $f : 2^E \to \mathbb{R}_{\geq 0}$ is a monotone submodular function. Obtaining an optimal solution to the relaxation (2) is NP-hard in this case [9], but we can obtain a constant-factor approximation using the continuous greedy algorithm. That is, we run the continuous greedy algorithm on the multilinear relaxation F of f and the polytope $\mathcal{P} = \bigcap_{j=1}^{k^{out}} \mathcal{P}(\mathcal{M}_j^{out}) \cap \bigcap_{j=1}^{k^{in}} \mathcal{P}(\mathcal{M}_j^{in})$. We consider the solution $\hat{x}$ produced by the algorithm when it is terminated at some time $T \in (0, 1]$. According to Lemma 2.5, we have $\hat{x}/T \in \mathcal{P}$ and $F(p \cdot \hat{x}) \geq (1 - e^{-T} - o(1)) f^+(p \cdot x^+)$. We then start our iterative rounding procedure with initial solution $x^0 = \hat{x}/T$, and define the potential z^t by

$$z^t = F(\mathbf{1}_{S^t} + T \cdot (p \cdot x^t)) - F(\mathbf{1}_{S^t}),$$

for all $0 \leq t \leq \tau$. Then, we have

$$z^0 = F(\mathbf{1}_{S^0} + T \cdot (p \cdot x^0)) - F(\mathbf{1}_{S^0}) = F(T \cdot (p \cdot x^0)) = F(p \cdot \hat{x}) \geq (1 - e^{-T} - o(1)) f^+(p \cdot x^+).$$

If the probing algorithm stops after τ steps, then we have $z^\tau = F(\mathbf{1}_{S^\tau} + T \cdot p \cdot 0^E) - F(\mathbf{1}_{S^\tau}) = F(\mathbf{1}_{S^\tau}) - F(\mathbf{1}_{S^\tau}) = 0$. Thus, the potential z satisfies the conditions of Lemma 2.7 with $\beta = 1 - e^{-T} - o(1)$.

Given the initial value x^0, our iterative rounding algorithm proceeds exactly as in the linear case. We now apply Lemma 2.7 with the potential z to analyze the expected performance of our probing algorithm.

Theorem 4.1. *For a monotone submodular objective function f, and any stopping time $T \in (0, 1]$ for the continuous greedy phase, the solution S produced by our randomized rounding algorithm satisfies $\mathbb{E}[f(S)] \geq \frac{(1-e^{-T}-o(1))}{T(k^{out}+k^{in})+1} \mathbb{E}[f(OPT)]$, where OPT is the solution produced by the optimal policy.*

Proof. We analyze the expected decrease $\mathbb{E}[z^t - z^{t+1}]$ due to step $t + 1$ of the algorithm. Suppose that the algorithm selects element $\bar{e}$ to probe. Then, we have $S^{t+1} = S^t + \bar{e}$ with probability $p_{\bar{e}}$, and $S^{t+1} = S^t$ otherwise. In general, we have

$$\begin{aligned}
&\mathbb{E}[z^t - z^{t+1}] \\
&= \mathbb{E}[F(\mathbf{1}_{S^t} + T \cdot (p \cdot x^t)) - F(\mathbf{1}_{S^t})] - \mathbb{E}[F(\mathbf{1}_{S^{t+1}} + T \cdot (p \cdot x^{t+1})) - F(\mathbf{1}_{S^{t+1}})] \\
&= \mathbb{E}[F(\mathbf{1}_{S^{t+1}}) - F(\mathbf{1}_{S^t})] + \mathbb{E}[F(\mathbf{1}_{S^t} + T \cdot (p \cdot x^t)) - F(\mathbf{1}_{S^{t+1}} + T \cdot (p \cdot x^{t+1}))] \\
&\leq \mathbb{E}[F(\mathbf{1}_{S^{t+1}}) - F(\mathbf{1}_{S^t})] + \mathbb{E}[F(\mathbf{1}_{S^t} + T \cdot (p \cdot x^t)) - F(\mathbf{1}_{S^t} + T \cdot (p \cdot x^{t+1}))],
\end{aligned} \tag{4}$$

where in the last line, we have used the fact that $S^{t+1} \supseteq S^t$ and F is increasing in all directions (Lemma 2.4).

We shall first bound the second expectation in (4). For each $i \in E$, we define

$$w_i = \frac{\partial F}{\partial x_i}(\mathbf{1}_{S^t}) = F(\mathbf{1}_{S^t + i}) - F(\mathbf{1}_{S^t}) = f(S^t + i) - f(S^t).$$

As in the linear case, let $\delta = x^t - x^{t+1}$, and note that $\delta_i = 0$ for all $i \in S^t$. Let $y = \mathbf{1}_{S^t} + T \cdot (p \cdot x^t)$ and suppose that we decrease the coordinates of y sequentially in some arbitrary order to obtain the vector $\mathbf{1}_{S^t} + T \cdot (p \cdot x^{t+1}) = y - T \cdot (p \cdot \delta)$. We consider the total change in F when an arbitrary coordinate i is decreased. Let y' be the vector obtained from y after all coordinates preceding i have been decreased. Lemma 2.3 states that F behaves as a linear function when only one coordinate changes, and so the total change in F from decreasing coordinate i is given by:

$$F(y') - F(y' - T \cdot ((p \cdot \delta) \cdot \mathbf{1}_i)) = T \cdot p_i \delta_i \frac{\partial F}{\partial x_i}(y') \leq T \cdot p_i \delta_i \frac{\partial F}{\partial x_i}(\mathbf{1}_{S^t}) = T \cdot w_i p_i \delta_i,$$

where the inequality follows from $y' \geq \mathbf{1}_{S^t}$ and Lemma 2.4, which states that the partial derivatives of F are coordinate-wise non-increasing. Thus, we have:

$$\begin{aligned} \mathbb{E}[F(\mathbf{1}_{S^t} + T \cdot (p \cdot x^t)) - F(\mathbf{1}_{S^t} + T \cdot (p \cdot x^{t+1}))] &= \mathbb{E}[F(y) - F(y - T \cdot (p \cdot \delta))] \\ \leq \mathbb{E}\left[\sum_{i \in E} T \cdot w_i p_i \delta_i\right] &= \sum_{i \in E} T \cdot w_i p_i \cdot \mathbb{E}[\delta_i] \leq \frac{1}{\Sigma^t}(k^{out} + k^{in}) \sum_{i \in E} T \cdot w_i p_i x_i^t, \end{aligned} \tag{5}$$

where the last inequality follows from Lemma 3.3, since our algorithm alters the solution x^t exactly as in the linear case.

Returning to the first expectation in (4), we have

$$\mathbb{E}[F(\mathbf{1}_{S^{t+1}}) - F(\mathbf{1}_{S^t})] = \sum_i \mathbb{P}\left[i \text{ probed}\right] p_i (F(S^t + i) - F(S^t)) = \frac{1}{\Sigma^t} \sum_{i \in E} x_i^t p_i w_i. \tag{6}$$

Combining inequalities (4), (5), and (6) we obtain:

$$\begin{aligned} \mathbb{E}[z^t - z^{t+1}] &\leq \frac{1}{\Sigma^t} \sum_{i \in E} w_i p_i x_i^t + \frac{1}{\Sigma^t}(k^{out} + k^{in}) \sum_{i \in E} T \cdot w_i p_i x_i^t \\ &= (T \cdot (k^{out} + k^{in}) + 1) \frac{1}{\Sigma^t} \sum_i w_i p_i x_i^t. \end{aligned}$$

On the other hand, the expected increase of $f(S^{t+1}) - f(S^t)$ in this step is:

$$\frac{1}{\Sigma^t} \sum_{i \in E} x_i^t p_i (f(S^t + i) - f(S^t)) = \frac{1}{\Sigma^t} \sum_{i \in E} x_i^t p_i w_i \geq \frac{1}{T \cdot (k^{out} + k^{in}) + 1} \mathbb{E}[z^t - z^{t+1}].$$

Applying Lemma 2.7, with $\beta = 1 - e^{-T} - o(1)$ and $\alpha = \frac{1}{T \cdot (k^{out} + k^{in}) + 1}$, the final solution S^τ produced by the algorithm satisfies

$$\mathbb{E}[f(S^\tau)] \geq \frac{1 - e^{-T} - o(1)}{T \cdot (k^{out} + k^{in}) + 1} \mathbb{E}[f(OPT)]. \qquad \square$$

Let $k = k^{out} + k^{in}$. Then, Theorem 4.1 shows that our probing algorithm returns a solution that has expected value at least $\rho(T) - o(1)$ times that of the optimal policy, where $\rho(T) \triangleq \frac{1-e^{-T}}{Tk+1}$. Setting just $T = 1$ gives us $\rho(T) = \left(1 - \frac{1}{e}\right) / (k+1)$, but in Appendix C, we derive the optimal value T_{opt} for T, showing that for $k = 1$, $T_{opt} = 1$ and for all $k > 1$,

$$T_{opt} = -1 - \frac{1}{k} - W_{-1}(-e^{-1-\frac{1}{k}}),$$

where W_{-1} is the lower, real-valued branch of the Lambert W function. We refer the reader to [16] for a thorough introduction to the Lambert W function. Chatzigeorgiou [10] gives the following bounds on $W_{-1}(e^{-u-1})$ that hold for all $u \in (0, 1)$:

$$-1 - \sqrt{2u} - \frac{3}{4}u \leq W_{-1}(-e^{-u-1}) \leq -1 - \sqrt{2u} - \frac{2}{3}u \leq W_{-1}(e^{-u-1}).$$

Using $u = \frac{1}{k}$, in this bound, we obtain

$$-\sqrt{\frac{2}{k}} + \frac{1}{4k} \leq 1 + \frac{1}{k} + W_{-1}(-e^{-1-\frac{1}{k}}) \leq -\sqrt{\frac{2}{k}} + \frac{1}{3k},$$

and hence

$$T_{opt} = \sqrt{\frac{2}{k}} - \frac{1}{\gamma_k k},$$

for $\gamma_k \in [3, 4]$. Using the fact that $e^{-x} = 1 - x + \Theta(x^2)$, we find that the optimal approximation ratio satisfies

$$\rho\left(T_{opt}\right) = \frac{1 - e^{-\sqrt{\frac{2}{k}} + \frac{1}{\gamma_k k}}}{\left(\sqrt{\frac{2}{k}} - \frac{1}{\gamma_k k}\right) k + 1} = \frac{\sqrt{\frac{2}{k}} - \Theta\left(\frac{1}{k}\right)}{\sqrt{2k} + 1 - \frac{1}{\gamma_k}} = \frac{1 - \Theta\left(\frac{1}{\sqrt{k}}\right)}{k + \sqrt{\frac{k}{2}}\left(1 - \frac{1}{\gamma_k}\right)}.$$

5 Non-Bipartite Matching and Matchoid Constraints

We now briefly consider a generalization of the matroid intersection setting studied in the previous sections. We can formulate the maximum-weight matching problem in a general graph $G = (V, E)$ as follows: for each $v \in V$ let $E(v) \subseteq E$ be the set of edges incident to v, and let $\mathcal{M}'_v = (\mathcal{I}'_v, E(v))$ be a uniform matroid of rank 1 defined on $E(v)$. This gives $|V|$ different matroids, each defined on some subset E. For convenience, we can easily extend each matroid $\mathcal{M}'_v$ to a matroid $\mathcal{M}_v$ on all of E, by letting $\mathcal{M}_v = (\mathcal{I}_v, E)$ be the union of $\mathcal{M}'_v$ and a free matroid on $E \setminus E(v)$. The maximum-weight matching problem in G is then equivalent to finding a maximum weight set S such that $S \in \mathcal{I}_v$ for every $v \in V$.

Note that a set $S \subseteq E$ is independent in $\mathcal{M}_v$ if and only if $S \cap E(v)$ is independent. We say that an element $e \in E$ *participates* in the matroid constraint $\mathcal{M}_v$ if and only if $e \in E(v)$. In our formulation of the maximum weighted matching problem, a given element $e \in E$ can participate in only 2 matroid constraints, since each edge e has 2 endpoints.

We can formulate the maximum weighted matching problem in a k-uniform hypergraph in the same fashion; each $e \in E$ is now a set of k vertices, and will hence participate in k matroid constraints. Additionally, we can naturally extend this construction to b-matchings by letting each $\mathcal{M}'_v$ be a uniform matroids of rank b. Even more generally, we can consider the *k-matchoid problem* in which we have a k-uniform hypergraph and each matroid $\mathcal{M}'_v$ is some *arbitrary* matroid. Matchoids generalize matroid intersections, since we can represent k simultaneous matroid constraints on a ground set E as a hypergraph with k vertices (one corresponding to each matroid constraint) and $|E|$ parallel hyperedges, each incident to all k vertices.

We now describe a simple adaptation of our probing algorithm to this generalized setting. Suppose that we are given a k^{out}-matchoid constraining which elements may be probed, and a k^{in}-matchoid constraining the elements that may be selected. Specifically, we are given a ground set E, a k^{out}-uniform hypergraph $H^{out} = (V^{out}, E^{out})$, and a k^{in}-uniform hypergraph $H^{in} = (V^{in}, E^{in})$ where $|E| = |E^{out}| = |E^{in}|$. We suppose that for each element $e \in E$ there is a unique corresponding edge $e^{out} \in E^{out}$ and $e^{in} \in E^{in}$. Additionally, we are given two sets of matroids $\{\mathcal{M}_v^{out}\}_{v \in V}$ and $\{\mathcal{M}_v^{in}\}_{v \in V}$ on E, where only the elements of E corresponding to those edges of $E^{out}(v)$ (respectively, $E^{in}(v)$) participate in the constraint $\mathcal{M}_v^{out}$ (respectively, $\mathcal{M}_v^{in}$). As in the case of matroid intersections, our goal is to find and probe a subset of elements $Q \subseteq E$ that is independent in each outer matroid, so that the subset of active elements $S \subseteq Q$ is independent in each inner matroid and has maximum value under a given function $f : 2^E \to \mathbb{R}_{\geq 0}$.

We first obtain a solution to the following mathematical optimization program using either linear programming or the continuous greedy algorithm:

$$\begin{aligned} \text{maximize } & f^+(p \cdot x) \\ \text{subject to: } & x \in \mathcal{P}(\mathcal{M}_v^{out}), \qquad v \in V \\ & p \cdot x \in \mathcal{P}(\mathcal{M}_v^{in}), \quad v \in V \\ & x \in [0,1]^E \end{aligned} \tag{7}$$

Note that now our program may have significantly more than $k^{in} + k^{out}$ matroid constraints. However, since H^{out} is k^{out}-uniform and H^{in} is k^{in}-uniform, we may assume that $|V| \leq (k^{in} + k^{out})|E|$. Thus, the total number of matroid constraints is still polynomial, and we can use the same techniques that we applied to (2) to solve (7).

Given a solution x of (7), our rounding procedure proceeds exactly as in the case of matroid intersection, randomly selecting an element $\bar{e} \in E$ to probe at each time step t with probability proportional to $x^{\bar{e}}t$, probing $\bar{e}$, then performing the necessary updates in each outer and inner matroid to obtain a new solution of (7). As in the intersection case, each of our outer and inner matroid updates satisfy Lemmas 3.1 and 3.2, respectively. Finally, we claim that Lemma 3.3 holds in our general setting. Indeed, we observe that, although there are now potentially many more than k outer and inner matroids, we must only update the constraints in which e participates. Each element $e \in E$ participates in only k^{out} outer matroid constraints and k^{in} inner matroid constraints, and so the maximum number of updates is exactly the same as in the proof of Lemma 3.3. It then follows that Lemma 3.3, and hence also our main technical results, Theorems 3.4 and 4.1, are valid even in the more general setting of matchoid constraints.

6 Conclusion

Here, we have considered the problem of maximizing a monotone submodular set function in the general stochastic probing model with matroid constraints on both the elements probed and the elements taken. Our general approach makes use of a new iterative rounding algorithm for linear set functions, together with a new observation regarding the performance of the continuous greedy algorithm for submodular maximization. We believe that this latter ingredient (given as Lemma 2.5 and described in detail in Appendix A) may be of use for the design and analysis of other stochastic algorithms in other contexts, as well.

One potential generalization in this direction would be to determine whether our results may be generalized to non-monotone functions using, for example, the measured continuous greedy algorithm of Feldman et al. [23]. Additionally, we ask whether a similar approach may be employed for other forms of constraints, such as knapsack constraints. For example, although b–matchings in general (non-bipartite) graphs are not intersections of two matroids, it is possible to exploit the matching structure to give a factor-$1/4$ approximation using our techniques.

Our analysis relies heavily on items being active with independent probabilities. Another direction for further work involves considering the case in which these probabilities may be correlated. While existing work on the correlation gap [44] implies that we can obtain a constant factor by simply ignoring correlations, it may be possible to obtain a better factor. In particular, we ask whether it might be possible to employ techniques similar to those that we employ to deal with correlations between choices in the probing policy to deal with correlations between items.

A final open question is whether it is possible to show that our results are the best possible. We note that even for offline, unweighted k-matroid intersection, the best known inapproximability result is $O(\ln k/k)$, which follows from inapproximability of k-set packing [32]. In contrast, the best known algorithms for matroid intersection problems have approximation guarantees of only $2/(k+\epsilon)$ in the unweighted case [34], and $1/(k-1+\epsilon)$ in the weighted case [35]. It would be interesting to investigate whether stronger lower bounds may be obtained in the stochastic model for multiple inner or outer matroid constraints.

Acknowledgments.

We thank Jan Vondrák for sharing a sketch of the analysis presented in Appendix A. Marek Adamczyk was supported by the ERC StG project PAAl no. 259515. Justin Ward was supported by EPSRC grant EP/J021814/1.

References

[1] Marek Adamczyk. Improved analysis of the greedy algorithm for stochastic matching. *Inf. Processing Letters*, 111:731–737, August 2011.

[2] Alexander A. Ageev and Maxim Sviridenko. Pipage rounding: A new method of constructing algorithms with proven performance guarantee. *J. Comb. Optim.*, 8(3):307–328, 2004.

[3] Shipra Agrawal, Yichuan Ding, Amin Saberi, and Yinyu Ye. Correlation robust stochastic optimization. In *Proc. 21st Annual ACM-SIAM Sympos. on Discrete Algorithms, SODA '10*, pages 1087–1096, Philadelphia, 2010. SIAM.

[4] Shipra Agrawal, Yichuan Ding, Amin Saberi, and Yinyu Ye. Price of correlations in stochastic optimization. *Oper. Res.*, 60(1):150–162, 2012.

[5] Arash Asadpour and Hamid Nazerzadeh. Maximizing stochastic monotone submodular functions, 2014. arXiv preprint. Available at: `http://arxiv.org/abs/0908.2788`.

[6] Arash Asadpour, Hamid Nazerzadeh, and Amin Saberi. Stochastic submodular maximization. In *Internet and Network Economics, 4th International Workshop*, pages 477–489, Berlin, 2008. Springer.

[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] E. M. L. Beale. Linear programming under uncertainty. *J. of the Royal Statistical Soc., Ser. B*, 17:173–184, 1955.

[9] Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a submodular set function subject to a matroid constraint (extended abstract). In *Integer Programming and Combinatorial Optimization, 12th International Conference*, pages 182–196, Berlin, 2007. Springer.

[10] Ioannis Chatzigeorgiou. Bounds on the Lambert function and their application to the outage analysis of user cooperation. *IEEE Comm. Letters*, 17(8):1505–1508, 2013.

[11] Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proc. 42nd ACM Sympos. on the Theory of Comput., STOC '10*, pages 311–320, New York, 2010. ACM.

[12] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Dependent randomized rounding via exchange properties of combinatorial structures. In *Proc. 51st IEEE Sympos. on Foundations of Comput. Sci., FOCS '10*, pages 575–584, Washington, DC, 2010. IEEE Computer Society.

[13] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Multi-budgeted matchings and matroid intersection via dependent rounding. In *Proc. 22nd Annual ACM-SIAM Sympos. on Discrete Algorithms, SODA '11*, pages 1080–1097, Philadelphia, 2011. SIAM.

[14] 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.

[15] 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*, pages 266–278, Berlin, 2009. Springer.

[16] Robert M. Corless, Gaston H. Gonnet, D. E. G. Hare, David J. Jeffrey, and Donald E. Knuth. On the Lambert W function. *Advances in Computational Math.*, 5(1):329–359, 1996.

[17] Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a Submodular Set Function Subject to a Matroid Constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011.

[18] William H Cunningham. Testing membership in matroid polyhedra. *J. of Comb. Theory, Ser. B*, 36(2):161–188, 1984.

[19] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.

[20] Brian C. Dean, Michel X. Goemans, and Jan Vondrák. Adaptivity and approximation for stochastic packing problems. In *Proc. 15th Annual ACM-SIAM Sympos. on Discrete Algorithms, SODA '05*, pages 395–404, Philadelphia, 2005. SIAM.

[21] Brian C. Dean, Michel X. Goemans, and Jan Vondrák. Approximating the stochastic knapsack problem: The benefit of adaptivity. *Math. Oper. Res.*, 33(4):945–964, 2008.

[22] Uriel Feige. A threshold of ln n for approximating set cover. *J. ACM*, 45:634–652, 1998.

[23] Moran Feldman, Joseph Naor, and Roy Schwartz. A unified continuous greedy algorithm for submodular maximization. In *Proc. 52nd IEEE Sympos. on Foundations of Comput. Sci., FOCS '11*, pages 570–579, Washington, DC, 2011. IEEE Computer Society.

[24] M L Fisher, G L Nemhauser, and L A Wolsey. An analysis of approximations for maximizing submodular set functions—ii. *Math. Programming Studies*, 8:73–87, 1978.

[25] Michel X. Goemans and Jan Vondrák. Stochastic covering and adaptivity. In *Theoretical Informatics, 7th Latin American Sympos.*, pages 532–543, Berlin, 2006. Springer.

[26] Daniel Golovin and Andreas Krause. Adaptive submodular optimization under matroid constraints, 2011. arXiv preprint. Available at: `http://arxiv.org/abs/1101.4450`.

[27] Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *J. of Artificial Intelligence Res.*, 42:427–486, 2011.

[28] Sudipto Guha and Kamesh Munagala. Approximation algorithms for budgeted learning problems. In *Proc. 39th ACM Sympos. on the Theory of Comput., STOC '07*, pages 104–113, New York, 2007. ACM.

[29] Sudipto Guha and Kamesh Munagala. Model-driven optimization using adaptive probes. In *Proc. 18th Annual ACM-SIAM Sympos. on Discrete Algorithms, SODA '07*, pages 308–317, Philadelphia, 2007. SIAM.

[30] Anupam Gupta, Ravishankar Krishnaswamy, Marco Molinaro, and R. Ravi. Approximation algorithms for correlated knapsacks and non-martingale bandits. In *Proc. 52nd IEEE Sympos. on Foundations of Comput. Sci., FOCS' 11*, pages 827–836, Washington, DC, 2011. IEEE Computer Society.

[31] Anupam Gupta and Viswanath Nagarajan. A stochastic probing problem with applications. In *Integer Programming and Combinatorial Optimization, 16th International Conference*, pages 205–216, Berlin, 2013. Springer.

[32] Elad Hazan, Shmuel Safra, and Oded Schwartz. On the complexity of approximating k-set packing. *Computational Complexity*, 15:20–39, 2006.

[33] Robert Kleinberg and S. Matthew Weinberg. Matroid prophet inequalities. In *Proc. 44th ACM Sympos. on the Theory of Comput., STOC '12*, pages 123–136, New York, 2012. ACM.

[34] Jon Lee, Maxim Sviridenko, and Jan Vondrák. Matroid matching: the power of local search. In *Proc. 42nd ACM Sympos. on the Theory of Comput., STOC '10*, pages 369–378, New York, 2010. ACM.

[35] Jon Lee, Maxim Sviridenko, and Jan Vondrák. Submodular maximization over multiple matroids via generalized exchange properties. *Math. Oper. Res.*, 35(4):795–806, 2010.

[36] G L Nemhauser and L A Wolsey. Best Algorithms for Approximating the Maximum of a Submodular Set Function. *Math. Oper. Res.*, 3(3):177–188, 1978.

[37] G L Nemhauser, L A Wolsey, and M L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Math. Programming*, 14(1):265–294, 1978.

[38] Filip Radlinski, Robert Kleinberg, and Thorsten Joachims. Learning diverse rankings with multi-armed bandits. In *Machine Learning, Proc. of the Twenty-Fifth International Conference*, pages 784–791, New York, 2008. ACM.

[39] A. Schrijver. *Combinatorial Optimization - Polyhedra and Efficiency.* Springer, 2003.

[40] Matthew Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Inf. Processing Systems 21*, pages 1577–1584. Curran Associates, 2008.

[41] Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *Proc. 40th ACM Sympos. on the Theory of Comput., STOC '08*, pages 67–74, New York, 2008. ACM.

[42] Jan Vondrák. Personal correspondence, 2013.

[43] David Williams. *Probability with Martingales*. Cambridge mathematical textbooks. Cambridge University Press, 1991.

[44] Qiqi Yan. Mechanism design via correlation gap. In *Proc. 22nd Annual ACM-SIAM Sympos. on Discrete Algorithms, SODA '11*, pages 710–719, Philadelphia, 2011. SIAM.

A The Continuous Greedy Algorithm

Let $\mathcal{I}$ be an independence system with a corresponding downward-closed solvable[2] polytope $\mathcal{P}$ and let $x^+ = \arg\max_{y \in \mathcal{P}} f^+(y)$, where f^+ is the relaxation given by (1) for some monotone submodular function f. Suppose that we run the continuous greedy algorithm until time $T \in (0, 1]$, obtaining a solution x.

We show how to modify the analysis of the continuous greedy algorithm to obtain the following the following stronger guarantees.

1. $x/T \in \mathcal{P}$
2. $F(x) \geq (1 - e^{-T} - o(1)) \max_{y \in \mathcal{P}} f^+(y)$

Our analysis is based on a proof due to Vondrák [42].

We consider the extension:

$$f^*(y) = \min_{S \subseteq E} \left[f(S) + \sum_{j \in E} y_j f_S(j) \right].$$

Călinescu et al. [9] show that for any value $y \in [0, 1]^E$, $f^+(y) \leq f^*(y)$.

Let $y(t)$ be the value of the fractional solution at time t of the continuous greedy algorithm. First, we note that at each step the continuous greedy algorithm adds $\delta \cdot v(t)$ to the current solution $y(t)$, where $v(t)$ is some vector in the polytope $\mathcal{P}$. Then, $x = y(T)$ can be written as T times a convex combination of the points $v(t) \in \mathcal{P}$:

$$\sum_{t=1}^{T/\delta} \delta v(t) = T \sum_{t=1}^{T/\delta} \frac{\delta}{T} v(t).$$

[2] A polytope is *solvable* if the linear optimization problem $\max_{v \in \mathcal{P}} \sum_{j \in E} v_j w_j$ can be solved for any set of weights $w \in \mathbb{R}^E_{\geq 0}$. All of the constraints considered in this paper have solvable polytopes.

Hence, we have $x/T \in \mathcal{P}$ as required.

We now show how to obtain our approximation guarantee with respect to $f^+(x^+)$. Our proof closely follows that of [17]. Suppose throughout that we discretize time into increments of size $\delta = \frac{1}{9n^2}$.

Lemma A.1 (Modification of Lemma 3.1 in [17]). *Consider any $y \in [0,1]^E$ and let R denote a random set in which each element $e \in E$ occurs independently with probability y_e. Then,*

$$f^+(x^+) \leq F(y) + \max_{v \in \mathcal{P}} \sum_{j \in E} v_j \mathbb{E}[f_R(j)].$$

Proof. We have the fractional solution $x^+ \in \mathcal{P}$, and so:

$$\sum_{j \in E} x_j^+ \mathbb{E}[f_R(j)] \leq \max_{v \in \mathcal{P}} \sum_{j \in E} v_j \mathbb{E}[f_R(j)]. \tag{8}$$

We note that $F(y) = \mathbb{E}[f(R)]$, and so (8) implies

$$\begin{aligned} F(y) + \max_{v \in \mathcal{P}} \sum_{j \in E} v_j \mathbb{E}[f_R(j)] &\geq \mathbb{E}[f(R)] + \sum_{j \in E} x_j^+ \mathbb{E}[f_R(j)] \\ &= \mathbb{E}\left[f(R) + \sum_{j \in E} x_j^+ f_R(j)\right] \\ &\geq \min_{S \subseteq E} \left[f(S) + \sum_{j \in E} x_j^+ f_S(j)\right] \\ &= f^*(x^+) \geq f^+(x^+). \end{aligned}$$

□

Lemma A.2 (Modification of Lemma 3.2 in [17]). *With high probability, the continuous greedy algorithm for every time step t finds a vector $v(t) \in \mathcal{P}$ such that*

$$\sum_{j \in E} v_j(t) \mathbb{E}[f_{R(t)}(j)] \geq (1 - 2n\delta) OPT - F(y(t)).$$

Proof. Let $y(t)$ be the value of y at time t, and let $R(t)$ denote the random set in which each element $j \in E$ appears independently with probability $y_j(t)$. At each time step t, the continuous greedy algorithm computes a solution to linear optimization problem $\max_{v \in \mathcal{P}} \sum_{j \in E} v_j \omega_j(t)$, where each $\omega_j(t)$ is an estimate the true expected marginal $\mathbb{E}[f_{R(t)}(j)]$. Călinescu et al. show that, given the choice of δ, the estimates satisfy $|\omega_j(t) - \mathbb{E}[f_{R(t)}(j)]| \leq \delta \cdot OPT$ with high probability. Now, let $M = \max_{v \in \mathcal{P}} \sum_{j \in E} v_j \omega_j(t)$ and let $v^* \in \mathcal{P}$ be some vector attaining this maximum. By Lemma A.1, we have $M \geq f^+(x^+) - F(y(t))$, and the solution $v(t)$ of the linear optimization problems satisfies $\sum_{j \in E} v_j(t) \omega_j(t) \geq \sum_{j \in E} v_j^* \omega_j(t) \geq$

$M - n\delta \cdot OPT$. Therefore, with high probability

$$\begin{aligned}\sum_{j \in E} v_j(t) \mathbb{E}[f_{R(t)}(j)] &\geq \sum_{j \in E} v_j(t)\omega_j(t) - n\delta \cdot OPT \\ &\geq M - 2n\delta \cdot OPT \\ &\geq f^+(x^+) - F(y(t)) - 2n\delta \cdot OPT \\ &\geq (1-2n\delta) f^+(x^+) - F(y(t)),\end{aligned}$$

where the last inequality follows from the fact that f^+ is a relaxation of f, and so $OPT \leq f^+(x^+)$. □

Lemma A.3. *With high probability the fractional solution $y(T)$ produced by the continuous greedy algorithm stopping at time $T \in (0, 1]$ satisfies:*

$$F(y(T)) \geq \big(1 - e^{-T} - o(1)\big)\, f^+(x^+).$$

Proof. As in Lemma A.2, let $R(t)$ be a random set in which each element $j \in E$ appears independently with probability $y_j(t)$. Let $D(t)$ be a random set in which each element $j \in E$ appears with probability $\delta v_j(t)$. Then, $\mathbb{E}[f(R(t+\delta))] \geq \mathbb{E}[f(R(t) \cup D(t))]$, since f is monotone and each element j appears in $R(t+\delta)$ with probability $y_j(t) + \delta v_j(t)$ but in $R(t) \cup D(t)$ with probability only $1 - (1 - y_j(t))(1 - \delta v_j(t)) \leq y_j(t) + \delta v_j(t)$. Therefore, applying Lemma A.2, with high probability we have:

$$\begin{aligned}F(y(t+\delta)) - F(y(t)) &= \mathbb{E}[f(R(t+\delta)) - f(R(t))] \\ &\geq \mathbb{E}[f(R(t) \cup D(t)) - f(R(t))] \\ &\geq \sum_{j \in E} \mathbb{P}\left[D(t) = \{j\}\right] \mathbb{E}[f_j(R(t))] \\ &= \sum_{j \in E} \delta v_j(t) \prod_{i \neq j} \big[1 - \delta v_i(t)\big] \mathbb{E}[f_j(R(t))] \\ &\geq \sum_{j \in E} \delta v_j(t)(1-\delta)^{n-1} \mathbb{E}[f_j(R(t))] \\ &\geq \delta(1 - n\delta) \sum_{j \in E} v_j(t) \mathbb{E}[f_j(R(t))] \\ &\geq \delta(1-n\delta)\big[(1-2n\delta) f^+(x^+) - F(y(t))\big] \\ &\geq \delta\big[(1-3n\delta) f^+(x^+) - F(y(t))\big].\end{aligned}$$

Let $\tilde{OPT} = (1 - 3n\delta) f^+(x^+)$. Then, we can rewrite the above inequality as:

$$\tilde{OPT} - F(y(t+\delta)) \leq (1-\delta)\left[\tilde{OPT} - F(y(t))\right].$$

Proceeding inductively, we then have:

$$\tilde{OPT} - F(y(T)) \leq (1-\delta)^{T/\delta}\left[\tilde{OPT} - F(y(0))\right] \leq (1-\delta)^{T/\delta}\tilde{OPT} \leq e^{-T} \cdot \tilde{OPT}.$$

Finally, since $\delta = \frac{1}{9n^2}$, we have:

$$F(y(1)) \geq \left(1 - e^{-T}\right) \tilde{OPT} = \left(1 - e^{-T}\right) \left(1 - \frac{1}{3n}\right) f^+(x^+) \geq \left(1 - e^{-T} - \frac{1}{3n}\right) f^+(x^+). \quad \square$$

B Martingale Theory

Here, we give a brief overview of the basic notions from martingale theory necessary for the proof of Lemma 2.7.

Definition B.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is a sample space, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. Sequence $\{\mathcal{F}_t : t \geq 0\}$ is called a *filtration* if it is an increasing family of sub-σ-algebras of $\mathcal{F}$: $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$.

Intuitively speaking, when considering a stochastic process, σ-algebra $\mathcal{F}_t$ represents all information available to us right after making step t. In our case σ-algebra $\mathcal{F}_t$ contains all information about each randomly chosen element to probe, about outcome of each probe, and about each support update for every matroid, that happened before or at step t.

Definition B.2. A process $(X_t)_{t\geq 0}$ is called a *martingale* if for every $t \geq 0$ all following conditions hold:

1. random variable X_t is $\mathcal{F}_t$-measurable,
2. $\mathbb{E}[|X_t|] < \infty$,
3. $\mathbb{E}\left[X_{t+1} \middle| \mathcal{F}_t\right] = X_t$.

In our case we actually consider a *sub-martingale* $(X_t)_{t\geq 0}$ which satisfies $\mathbb{E}\left[X_{t+1} \middle| \mathcal{F}_t\right] \geq X_t$ instead of equality in the above definition.

Definition B.3. Random variable $\tau : \Omega \mapsto \{0, 1, \ldots\}$ is called a *stopping time* if $\{\tau = t\} \in \mathcal{F}_t$ for every $t \geq 0$.

Intuitively, τ represents a moment when an event happens. We have to be able to say whether it happened at step t given only the information from steps $0, 1, 2, \ldots, t$. In our case we define τ as the first moment when the fractional solution becomes zero. It is clear that this is a stopping time according to the above definition.

Theorem B.1 (Doob's Optional-Stopping Theorem)**.** *Let τ be a stopping time. Let $(X_t)_{t\geq 0}$ be a sub-martingale. If there exists a constant N such that always $\tau < N$, then $\mathbb{E}[X_\tau] \geq \mathbb{E}[X_0]$.*

The above is not Doob's theorem in its full generality, but rather the simplest variant that still holds in our setting.

C Derivation of T_{opt}

Recall that our submodular stochastic probing algorithm's approximation performance is given by $\rho(T) - o(1)$, where:

$$\rho(T) = \frac{1 - e^{-T}}{Tk + 1},$$

where T is the stopping time of the continuous greedy algorithm and $k = k^{out} + k^{in} \geq 1$.

The first derivative of ρ with respect to T is given by:

$$\frac{d}{dT}\rho(T) = \frac{(Tk+1)e^{-T} - k\left(1 - e^{-T}\right)}{(Tk+1)^2}.$$

We have $\frac{d}{dT}\rho(T) = 0$ if and only if

$$k\left(T + \frac{1}{k} + 1\right)e^{-T} - k = 0,$$

or, equivalently (since $k \geq 1$):

$$\begin{aligned}
\left(T + 1 + \frac{1}{k}\right)e^{-T} &= 1 \\
-\left(T + 1 + \frac{1}{k}\right)e^{-T-1-\frac{1}{k}} &= -e^{-1-\frac{1}{k}} \\
-T - 1 - \frac{1}{k} &= W\left(-e^{-1-\frac{1}{k}}\right) \\
T &= -1 - \frac{1}{k} - W\left(-e^{-1-\frac{1}{k}}\right),
\end{aligned}$$

where W is the Lambert W function, defined by the equation $z = W(z)e^{W(z)}$ (for a detailed discussion of the Lambert W function, we refer the reader to [16]). The function $W(z)$ has 2 real-valued branches in the range $[-e^{-1}, 0]$. In this range, the upper branch W_0 of W takes values in $[-1, 0]$ and so $W_0(-e^{-1-\frac{1}{k}})$ yields a negative value for T, since $-1 - \frac{1}{k} < -1$. Thus, we restrict ourselves to the lower branch, W_{-1}, which takes values in $[-1, -\infty]$ over this range. The single, real-valued critical point of $\rho(T)$ satisfying $T \geq 0$ is thus given by:

$$T_{opt} = -1 - \frac{1}{k} - W_{-1}\left(-e^{-1-\frac{1}{k}}\right). \tag{9}$$

Now, we show that T_{opt} is a maximizer of ρ. Let $B(T) = Tk + 1$ and note that $B(T) > 0$ for all $T \geq 0$. Then, we have:

$$\begin{aligned}
\frac{d^2}{dT^2}\rho(T) &= \frac{B(T)^2(-B(T)e^{-T}) - 2B(T)k\left(B(T)e^{-T} - k + ke^{-T}\right)}{B(T)^4} \\
&= \frac{-B(T)^2e^{-T} - 2B(T)ke^{-T} + 2k^2 - 2k^2e^{-T}}{B(T)^3} \\
&= \frac{e^{-T}}{B(T)^3}\left[-B(T)^2 - 2B(T)k + 2k^2e^{T} - 2k^2\right].
\end{aligned}$$

We now show that $\frac{d^2}{dT^2}\rho(T_{opt}) < 0$. Because $B(T_{opt}) > 0$, it suffices to show that:

$$-B(T_{opt})^2 - 2B(T_{opt})k + 2k^2 e^{T_{opt}} - 2k^2 < 0. \tag{10}$$

Let $C = e^{-1-\frac{1}{k}}$. Then, we have $T_{opt} = -1 - \frac{1}{k} - W_{-1}(-C)$, and so

$$\begin{aligned}
T_{opt} + 1 + \frac{1}{k} &= -W_{-1}(-C) \\
&= -W_{-1}(-C)e^{W_{-1}(-C)}e^{-W_{-1}(-C)} \\
&= -(-C)e^{-W_{-1}(-C)} \\
&= e^{-1-\frac{1}{k}}e^{-W_{-1}(-C)} \\
&= e^{T_{opt}}.
\end{aligned}$$

Thus, we have

$$2k^2 e^{T_{opt}} = 2k^2\left(T_{opt} + 1 + \frac{1}{k}\right) = 2k\left(T_{opt}k + k + 1\right) = 2B(T_{opt})k + 2k^2. \tag{11}$$

Substituting (11) in (10) we obtain

$$-B(T_{opt})^2 - 2B(T_{opt})k + 2B(T_{opt})k + 2k^2 - 2k^2 = -B(T_{opt})^2 < 0,$$

which is true for all $k \geq 0$, since $B(T_{opt}) > 0$ for all such k. Thus, T_{opt} is a local maximum for ρ. Because T_{opt} is the only critical point of $\rho(T)$ for $T > 0$, it follows that T_{opt} is a global maximum of ρ. We note that for $k = 1$, we have $T_{opt} > 1$. In this case, the optimal value for $T \in [0, 1]$ is given by 1, since T must be non-decreasing on the interval $[0, T_{opt}]$ in order for T_{opt} to be a global maximum. In all other cases, we have $T_{opt} \leq 1$.