Revisiting the Approximation Bound for Stochastic Submodular Cover

Lisa Hellerstein
Department of Computer Science and Engineering
NYU Tandon School of Engineering
lisa.hellerstein@nyu.edu

Devorah Kletenik
Brooklyn College
City University of New York
kletenik@sci.brooklyn.cuny.edu

Abstract

Deshpande et al. presented a $k(\ln R + 1)$ approximation bound for Stochastic Submodular Cover, where k is the state set size, R is the maximum utility value of a single item, and the utility function is integer-valued. Since $R \leq Q$, this bound is similar to the $(\ln Q/\eta + 1)$ bound of Golovin and Krause, whose analysis was recently found to have an error. We revisit the proof of the $k(\ln R + 1)$ bound, fill in crucial missing details in the proof of a key lemma, and prove two bounds whcih apply even when the utility function is not integer-valued: $k(\ln R/\eta_1 + 1)$ where η_1 is the smallest increase in utility of a single item, and $(\ln R/\eta_E + 1)$ where η_E is the smallest *expected* increase in utility of a single item. Our bounds apply only to the stochastic setting with independent states.

1 Introduction

Golovin and Krause introduced the Stochastic Submodular Cover (StSuC) problem [7]. They presented a proof showing that the Adaptive Greedy algorithm is a $(\ln \frac{Q}{\eta} + 1)$-approximation algorithm, for a class of adaptive submodular cover problems that includes the StSuC problem. Here Q is the "goal value", and η is the minimum gap between Q and any attainable utility value $Q' < Q$.

Subsequently, Deshpande et al. used an LP-based analysis to show that Adaptive Greedy is a $k(\ln R + 1)$-approximation algorithm for the StSuC problem, assuming an integer-valued utility function [5]. Here k is the (constant) size of the state set, and R is the maximum utility of a single item, so $R \leq Q$. For integer-valued utility functions, $\eta = 1$, so $(\ln \frac{Q}{\eta} + 1) \leq (\ln Q + 1)$.

Recently, Nan and Saligrama discovered an error in Golovin and Krause's analysis of Adaptive Greedy [11]. The error invalidates the $(\ln \frac{Q}{\eta} + 1)$ bound for the StSuC problem. The $k(\ln R + 1)$ bound of Deshpande et al. is now important as a replacement. Because of this, we were motivated to revisit this bound. To our knowledge, no similar replacement bound exists for the StSuC problem. Deshpande et al. proved a different bound on the StSuC problem using their Adaptive Dual Greedy algorithm, but it is incomparable. Bounds for other state-dependent submodular cover problems are not useful in bounding the StSuC problem (e.g. [9]).

By tightening the analysis and using a different technical result of Wolsey, we prove two variants of the bound of Deshpande et al., removing the assumption that the utility function is integer valued: $k(\ln \frac{R}{\eta_1} + 1)$ and $(\ln \frac{R}{\eta_E} + 1)$. Here η_1 is the smallest non-zero increase in utility attainable from a single item and η_E is the smallest non-zero *expected* increase in utility attainable from a single item. These bounds are similar to the $(\ln \frac{Q}{\eta} + 1)$ bound claimed by Golovin and Krause.

One of the key lemmas of Deshpande et al. lacked a convincing proof. The proof said the lemma "follows directly by linearity of expectation" from a previous result, but linearity of expectation is not

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

sufficient. We need this lemma for our improved bound and provide a complete proof in the extended version of this paper.

There are previous results, for other problems, that relied on the claimed bound of Golovin and Krause for the StSuC problem. Examples include many of the results on the Stochastic Boolean Function Evaluation (SBFE) problem (e.g. [5, 1]) and the $(\ln n + 1)$ approximation bound on the Stochastic Set Coverage problem [7]. These results still hold, up to constant factors, by substituting the $k(\ln \frac{R}{\eta_1} + 1)$ bound here (or the bound of Deshpande et al.) for the claimed $(\ln \frac{Q}{\eta} + 1)$ bound.

In fact, we note that $\eta_E \geq \eta_1$ when the following property holds: whenever an item j yields non-zero utility in one state, it yields non-zero utility in its other states. This property holds in Stochastic Boolean Function evaluation, because whenever an input variable x_i of Boolean function f is relevant to that function, discovering that the variable is either 0 or 1 must contribute non-zero utility (by submodularity). It also holds in the pipelined filter ordering problem considered in Liu et al. [10], which is a special case of Stochastic Set Coverage. Therefore, for those problems, the bound of $(\frac{\ln R}{\eta_1} + 1)$ follows from the $(\frac{\ln R}{\eta_E} + 1)$ bound. However, there are problems for which the quantity η_E can be less than η_1: in the general Stochastic Set Coverage problem, it is possible for an item to contribute zero utility in one state and non-zero utility in another. For example, a sensor may cover something or nothing, depending on whether its state is working or broken.

The bounds shown in this paper do not extend beyond the StSuC problem to the more general class of adaptive submodular cover problems. The proof of these bounds requires that item states be independent, and that the utility function be "pointwise" submodular. As a result, the bounds do not apply to previous work where Adaptive Greedy was used to solve Equivalence Class Determination, Decision Region Identification, or Scenario (sample-based) Submodular Cover (e.g., [2, 3, 8]). We note that there are now other algorithms for solving these particular problems, achieving better approximation bounds[1] The most recent is the algorithm of Kambadur et al. which uses just one simple greedy rule, and solves a more general class of problems [9].

It remains an interesting question whether an alternative proof can be found to show that Adaptive Greedy achieves something like the claimed bound of $(\ln \frac{Q}{\eta} + 1)$ for the general class of adaptive submodular cover problems considered by Golovin and Krause.

2 Definitions and Background

Let $\mathcal{N} = \{1, \ldots, n\}$ be a set of *items*. Let $\mathcal{O}$ be a finite collection of *states*. For simplicity we assume $\mathcal{O} = \{0, 1\}$, but our proof extends easily to state spaces of constant size k. A state vector $x \in \{0, 1\}^n$ is an assignment of states to items, where x_i is the state of item i. A partial assignment $b \in \{0, 1, *\}^n$ represents partial information about a state assignment, with $b_i = *$ if the state of i is unknown. For $b \in \{0, 1, *\}^n$, $i \in \mathcal{N}$, and $\ell \in \{0, 1\}$, $b_{i \leftarrow \ell}$ is the assignment produced from b by setting b_i to ℓ. For $a, b \in \{0, 1, *\}^n$, we say a is an extension of b, written $a \succeq b$, if $a_i = b_i$ for all i with $b_i \neq *$. We use $\star$ to denote the assignment $(*, \ldots, *)$.

Following Deshpande et al., we define a (state dependent) utility function to be a function $g : \{0, 1, *\}^n \rightarrow \mathcal{R}^{\geq 0}$. For $b \in \{0, 1, *\}^n$, $g(b)$ is the "utility" of the information in b; in other words, it is the utility of the items in $\{j | b_j \neq *\}$ when they are in the states indicated by b. For fixed g, we define the function $G : 2^{\mathcal{N}} \times \{0, 1, *\}^n \rightarrow \mathcal{R}^{\geq 0}$, where $G(S, x) = g(b)$ for the $b \in \{0, 1, *\}^n$ satisfying $b_j = x_j$ for $j \in S$, and $b = *$ otherwise.

Let p_i be the probability that item i is in state 1 and $q_i = (1 - p_i)$ be the probability it is in state 0. Let D_p denote the product distribution defined by the p_i. Let $P(b)$ be the joint probability of the known states in b, so $P(b) = (\prod_{i:b_i=1} p_i)(\prod_{i:b_i=0} q_i)$.

For $S \subseteq \mathcal{N}, b \in \{0, 1, *\}^n$, and $j \in \mathcal{N}$ such that $b_j \neq *$, let $G_{S,b}(j) = G(S \cup \{j\}, b) - G(S, b)$. For $\ell \in \{0, 1\}$ and j such that $b_j = *$, let $G_{S,b}(j, \ell) = G(S \cup \{j\}, b_{j \leftarrow \ell}) - G(S, b)$. Similarly, for $b \in \{0, 1, *\}^n$ and j such that $b_j = *$, let $g_b(j, \ell) = g(b_{j \leftarrow \ell}) - g(b)$.

[1]The paper of Grammel et al. had 2 algorithms for solving Scenario Submodular Cover. The analysis of the second algorithm relied on the problematic result of Golovin and Krause, and is no longer valid. The first algorithm was independent of that work and its bound still holds.

Function g is *monotone* if for all $a, b \in \{0, 1, *\}$ with $a \succeq b$, we have $g(a) \geq g(b)$. Function g is *submodular* if for all $a, b \in \{0, 1, *\}^n$ where $a \succeq b$, $j \in \mathcal{N}$ such that $a_j = b_j = *$, and $\ell \in \{0, 1\}$, we have $g(a_{i \leftarrow \ell}) - g(a) \leq g(b_{i \leftarrow \ell}) - g(b)$. (Golovin and Krause call this "pointwise" submodularity.)

In the StSuC problem, we need to choose items sequentially from $\mathcal{N}$. Each item has an initially unknown state, which is a value ℓ in $\mathcal{O} = \{0, 1\}$. We continue choosing items from $\mathcal{N}$ until the chosen items achieve a certain *goal utility* Q, as measured by a given monotone submodular function $g : \{0, 1, *\}^n \to \mathbb{R}^{\geq 0}$. Choosing item j incurs a known cost c_j. We cannot see the state of an item j until after we choose it, and incur its cost. Each item can be chosen only once.

The state of each item j is an independent random variable. We are given the distribution of states for each item j. The problem is to determine the order in which to choose items, so as to minimize expected cost. The choice of the next item can depend on the states of the previously chosen items.

Formally, the inputs to the StSuC problem are as follows: itemset $\mathcal{N}$, the probabilities p_j, the costs c_j, and a monotone submodular utility function $g : \{0, 1, *\}^n \to \mathbb{R}^{\geq 0}$ (given by an oracle). For $j \in \mathcal{N}$, $0 < p_j < 1$ and $c_j \in \mathbb{R}^+$, g has the following property: there exists a value $Q \in \mathbb{R}^{\geq 0}$ such that for all full assignments $x \in \{0, 1\}^n$, $g(x) = Q$. This ensures that utility value Q can always be attained. We call Q the *goal value* of g. For $x \in \{0, 1\}^n$, we say that $S \subseteq \mathcal{N}$ is a *cover* for x if $G(S, x) = Q$.

We assume without loss of generality that for each $j \in \mathcal{N}$, there exists $\ell \in \{0, 1\}$ such that $g_\star(j, \ell) > 0$. Otherwise, by submodularity, choosing j can never increase utility.

A (feasible) solution to the StSuC problem is an adaptive strategy for choosing a sequence of items, until they have utility Q, as measured by g. The strategy corresponds to a decision tree τ, although we do not require the tree to be output explicitly (it may have exponential size). Each internal node of τ is labeled with an item j, and has a child for each of the possible states of j. Each $x \in \{0, 1\}^n$ results in following a particular root-leaf path in the tree τ. We call the items on that path the *cover* constructed by τ on x. The expected cost incurred by τ is $\sum_{x \in \{0,1\}^n} P(x)\ cost(\tau, x)$, where $cost(\tau, x)$ is the sum of the costs of the items in the cover constructed by τ on x. The strategy corresponding to τ is an optimal solution if it incurs the minimum possible expected cost.

The Adaptive Greedy algorithm of Golovin and Krause solves the StSuC problem using the greedy rule that chooses the item with the smallest cost per expected unit of utility to be gained. We give pseudocode in Figure 1, where we use x_j to denote the random state of item j.

$b \leftarrow (*, * \ldots, *)$
$F \leftarrow \emptyset$ //F is set of items j chosen so far
while $g(b) < Q$ **do**
  for $j \notin F$ **do**
    $\Delta(j) \leftarrow \sum_{\ell \in \{0,1\}} (Prob[x_j = \ell])\ g_b(j, \ell)$ // expected increase in utility if j is chosen next
  end for
  $j^* \leftarrow \arg\min_{j \notin F} \frac{c_j}{\Delta(j)}$
  $\ell \leftarrow$ the state of j^* // observe state of j^*
  $F \leftarrow F \cup \{j^*\}$
  $b_{j^*} \leftarrow \ell$ // update b to include state of j^*
end while
return b

Algorithm 1: Adaptive Greedy

We use R to denote $\max_{j \in \mathcal{N}, \ell \in \{0,1\}} g_\star(j, \ell)$. Parameter η_1 equals the minimum value of $g_b(j, \ell)$, for any $b \in \{0, 1, *\}^n$ where $b_j = *$, and $\ell \in \{0, 1\}$. Parameter η_E is the minimum value of $p_j g_b(j, \ell) + q_j g_b(j, \ell)$, for any $b \in \{0, 1, *\}^n$ with $b_j = *$.

3 Proof of the $(\ln \frac{R}{\eta_E} + 1)$ bound

The starting point of the analysis is the definition of a special LP whose optimal value lower bounds the optimal expected cost for the StSuC problem. The LP is based on the *Neighbor Property*. This LP and its dual are used only in the analysis of Adaptive Greedy. They do not play any role in the Adaptive Greedy algorithm itself.

Let $W \subseteq \{0,1,*\}^n$ be the set of partial assignments that have exactly one $*$. For $w \in W$, let $J(w)$ denote the j such that $w_j = *$. For $\ell \in \{0,1\}$, let $w^{(\ell)} = w_{J(w) \leftarrow \ell}$. For $j \in \mathcal{N}$ and $a \in \{0,1\}^n$, let $a^j = a_{j \leftarrow *}$. Given $w \in W$, consider its two extensions $w^{(0)}$ and $w^{(1)}$. Let τ be a strategy (decision tree) solving the StSuC instance. Consider the paths taken in τ on $w^{(0)}$ and $w^{(1)}$. Either they are identical, meaning no node on them was labeled with $J(w)$, or they diverge at a node labeled with $J(w)$. This proves the Neighbor Property, which states that for each $w \in W$, the covers constructed by τ for $w^{(0)}$ and $w^{(1)}$ either both contain $J(w)$, or neither does.

We present the LP in Figure 1. It is related to an IP used by Wolsey to obtain an approximation bound for deterministic Submodular Cover [12]. It has $n2^{n-1}$ variables z_w, one for each $w \in W$.

Intuitively, a 0,1 assignment Z to the variables z_w associates a subset $F(a)$ with each $a \in \{0,1\}^n$ as follows: $F(a) = \{j \in \mathcal{N} \mid \text{ for } w = a^j, z_w = 1\}$. If Z satisfies the LP constraints, then for each a, $F(a)$ is a cover for a (this follows directly from Wolsey's analysis). Further, the value of the objective function on Z equals the expected cost of cover $F(a)$ for $a \sim D_p$.

If we further constrain the variables of the LP so that each z_w must be in $\{0,1\}$, the resulting IP asks for a cover for each a, such that the covers satisfy the Neighbor Property, and the expected cost of the cover on a random $a \sim D_p$ is minimized. These observations imply the following lemma.

Lemma 1. *[4] The optimal value of the LP in Figure 1 lower bounds the expected cost of the optimal strategy solving the associated StSuC instance.*

Take each constraint of the LP that is associated with a pair S, a, and multiply both sides of that constraint by $P(a)$. This does not change the optimal value of the LP. Taking the dual of the resulting LP, we get the Dual LP of Figure 2. By strong duality, its optimal value is equal to the optimal value of the primal, and thus also lower bounds the expected cost of the optimal strategy.

	Minimize $\sum_{w \in W} c_{J(w)} P(w)\, z_w$	
s.t.	$\sum_{j \in \mathcal{N}} G_{S,a}(j)\, z_{a^j} \geq Q - G(S,a)$	$\forall a \in \{0,1\}^n, S \subseteq N$
	$z_w \geq 0$	$\forall w \in W$

Figure 1: Primal LP

	Maximize $\sum_{a \in \{0,1\}^n} \sum_{S \subseteq \mathcal{N}} P(a)\, (Q - G(S,a))\, y_{S,a}$	
s.t.	$\sum_{S \subseteq \mathcal{N}} \sum_{\ell \in \{0,1\}} (Prob[x_j = \ell])\, G_{S,w^{(\ell)}}(j)\, y_{S,w^{(\ell)}} \leq c_j$	$\forall w, j \text{ s.t. } w \in W, j = J(w)$
	$y_{S,a} \geq 0$	$\forall S \subseteq N, a \in \{0,1\}^n$

Figure 2: Dual LP

The basic idea of the analysis is to describe an assignment Y to the dual variables $y_{S,a}$ that corresponds to information about the running of Adaptive Greedy on the different possible state vectors a. For any fixed a, the variables $y_{S,a}$ are associated with the results of running Adaptive Greedy on state vector a (i.e., when each item j is in state a_j). Our analysis diverges from that of Deshpande et al. in the proofs of two key lemmas, Lemma 2 and Lemma 4, below.

In Lemma 2, we show that the value of the objective function of Dual LP, on assignment Y, equals the expected cost incurred by Adaptive Greedy. This is the lemma presented by Deshpande et al. without adequate proof. In Lemma 4, we show that Y exceeds the right hand side of the constraints of Dual LP by a factor of at most $(\ln R/\eta_E + 1)$.

As in Deshpande et al., we combine the above lemmas to complete the analysis, Let OPTDT be the expected cost of the optimal strategy, and OPTDLP be the optimal value of Dual LP. Let AGCOST be the expected cost of Adaptive Greedy and let $q(y)$ denote the objective function of Dual LP. By Lemma 1, $\mathsf{OPTDLP} \leq \mathsf{OPTDT}$. By Lemma 4, if we divide Y by $(\ln R/\eta_E + 1)$, the resulting assignment is a feasible solution to Dual LP. Call that assignment Y'. Since Y' is a feasible solution of Dual LP, $q(Y') \leq \mathsf{OPTDLP} \leq \mathsf{OPTDT}$. Since $q()$ is a linear function,

$q(Y) \leq (\ln R/\eta_E + 1)\mathsf{OPTDT}$. Finally, by Lemma 2, $q(Y)$ is the expected cost of Adaptive Greedy, which is therefore at most $(\ln R/\eta_E + 1)\mathsf{OPTDT}$.

3.1 The Two Lemmas

It remains to describe assignment Y and to prove the two key lemmas. Consider execution of Adaptive Greedy on a state vector x. Number the iterations of the while loop starting from 1. Let T^x be the total number of iterations. Let b_x^t, F_x^t, and $\Delta_x^t(j)$ be the values of b, F, and $\Delta(j)$ at the end of iteration t. So $|F_x^t| = t$, and b_x^t represents the states of items in F_x^t. Set $\theta_x^t = c_j/\Delta_x^t(j)$ where j is the value of j^* chosen in iteration t. Thus θ_x^t is the *rate* in iteration t (cost per expected unit of utility). Let j_x^t be the item chosen in iteration t.

Define Y be the assignment to the variables in the dual LP such that for all $x \in \{0,1\}^n$, $y_{F^0,x} = \theta_x^1$, $y_{F^t,x} = (\theta_x^{t+1} - \theta_x^t)$ for $t \in \{1 \ldots T^x - 1\}$, and $y_{S,x} = 0$ for all other S.

Let $q^x(Y) = \sum_{S \subset \mathcal{N}} (Q - G(S,x)) y_{S,x}$.

In the proof of Deshpande et al., Lemma 2 was claimed to follow directly from a result of Wolsey by linearity of expectation [4]. This would be the case if state vector x was given at the start of Adaptive Greedy, and item j chosen in loop iteration t was the minimizer of the quantity $c_j/G_{F,x}(j)$, whose denominator is the guaranteed increase in utility from choosing j with known x.

However, Adaptive Greedy chooses the item j that minimizes $c_j/\Delta(j)$, whose denominator is the *expected* increase in utility from choosing j. Linearity of expectation is not sufficient here. We modify Wolsey's analysis by "averaging" the expected value $\Delta(j)$ over the two different possible states of j, to obtain a full and correct proof.

Lemma 2. *The expected cost of the cover constructed by Adaptive Greedy is equal to $q(Y)$.*

Proof. For each fixed $x \in \{0,1\}^n$, we have the following (omitting the subscripts and superscript x on θ, F, q, and T for readability):

$$\begin{aligned}
&q^x(Y) && \\
&= \textstyle\sum_{S \subset \mathcal{N}} (Q - G(S,x)) y_{S,x} && \text{by definition of } q \\
&= \textstyle\sum_{t=1}^{T} (Q - G(F^{t-1},x)) y_{F^{t-1},x} && \text{since } y_{S,x} = 0 \text{ for } S \notin \{F^0, \ldots, F^{T-1}\} \\
&= (Q - G(F^0,x))\theta^1 + \textstyle\sum_{t=2}^{T} (Q - G(F^{t-1},x))(\theta^t - \theta^{t-1}) && \text{by definition of } Y \\
&= (Q - G(F^{T-1},x))\theta^T + \textstyle\sum_{t=1}^{T-1} (G(F^t,x) - G(F^{t-1},x))\theta^t && \text{grouping by multiples of } \theta^t \\
&= \textstyle\sum_{t=1}^{T} (G(F^t,x) - G(F^{t-1},x))\theta^t && \text{because } Q = G(F^T)
\end{aligned}$$

Therefore (restoring subscripts and superscript x):

$$E[q^x(Y)] = \sum_{x \in \{0,1\}^n} \sum_{t=1}^{T^x} P(x)[G(F_x^t,x) - G(F_x^{t-1},x)]\theta_x^t \tag{1}$$

Consider the decision tree τ corresponding to Adaptive Greedy. Running Adaptive Greedy on input x corresponds to following a path in τ from the root to a leaf. For each internal node v of τ, let $j(v)$ denote the item tested in that node. Let $X = \{(x,t) \mid x \in \{0,1\}^*, 1 \leq t \leq T^x\}$. Let X^v denote the set of $(x,t) \in X$ such that v is node number t on the root-leaf path that is followed in τ on state vector x (with the root as node number 1 on that path). Thus for $x \in \{0,1\}^n$ and $1 \leq t \leq T^x$, the pair (x,t) belongs to exactly one set X^v, and the X^v form a partition of X. Clearly each pair $(x,t) \in X^v$ has the same value for t, which is the number of nodes on the path from the root of τ to node v.

Let v be a node in τ, and let $i = j(v)$ be the item labeling v. We define $p_v = p_i$ and $q_v = q_i$. We define $X_1^v = \{(x,t) \in X^v | x_i = 1\}$, and $X_0^v = \{(x,t) \in X^v | x_i = 0\}$. Each $(x,t) \in X_1^v$ has a corresponding "neighbor" $(x',t) \in X_0^v$, where x differs from x' only in position i. Therefore, $X^v = X_1^v \cup X_0^v$ and there is a bijection between X_1^v and X_0^v mapping each $(x,t) \in X_1^v$ to $(x',t) \in X_0^v$.

Let S^v denote the set of items labeling the nodes on the path from the root down to node v, not including the item labeling node v. Let b^v denote the partial assignment indicating the outcomes of the tests in S^v, corresponding to the path down to (but not including) node v.

For $(x,t) \in X^v$, $F_x^{t-1} = S^v$ and $F_x^t = S^v \cup \{j(v)\}$. Also,

$$G(F_x^t, x) - G(F_x^{t-1}, x) = G(S^v \cup \{j(v)\}, x) - G(S^v, x) \tag{2}$$

For $\ell \in \{0,1\}$, let $g_v(\ell)$ be the increase in utility obtained at node v, if the element in that node is in state ℓ. That is, $g_v(\ell) = g_{b^v}(j(v), \ell)$.

We define $\Delta(v) = \sum_{\ell \in \{0,1\}} (Prob[x_i = \ell])\, g_v(\ell)$. Thus $\Delta(v)$ is the expected increase in utility at node v. We have

$$\Delta(v) = p_v g_v(1) + q_v g_v(0) \tag{3}$$

Clearly, $P(b^v) = \sum_{(x,t)\in X^v} P(x)$ and $p_v P(b^v) = \sum_{(x,t)\in X_1^v} P(x)$. Recall that for $x \in \{0,1\}^n$ and $i \in \mathcal{N}$, x^i denotes the partial assignment that is produced from x by setting bit i to $*$. Since for $(x,t) \in X_1^v$, $P(x^{j(v)}) = \frac{1}{p_v} P(x)$, it follows that $\sum_{(x,t)\in X_1^v} P(x^{(j)}) = P(b^v)$ and hence

$$\sum_{(x,t)\in X_1^v} P(x^{j(v)}) = \sum_{(x,t)\in X^v} P(x) \tag{4}$$

Let $\theta^v = c_{j(v)}/\Delta(v)$. Then for each $(x,t) \in X^v$, $\theta_x^t = \theta^v$.

Thus,

$$\begin{aligned}
&E[q^x(Y^x)] \\
&= \sum_{x\in\{0,1\}^n} \sum_{t=1}^{T^x} P(x)[G(F_x^t,x) - G(F_x^{t-1},x)]\theta_x^t \qquad \text{by (1)} \\
&= \sum_v \sum_{(x,t)\in X^v} P(x)[G(S^v\cup\{j(v)\},x) - G(S^v,x)]\theta^v \qquad \text{by (2) since the } X^v \text{ partition the } (x,t) \\
&= \sum_v \theta^v \left(\sum_{(x,t)\in X^v} P(x)[G(S^v\cup\{j(v)\},x) - G(S^v,x)]\right) \qquad \text{moving } \theta^v \text{ forward} \\
&= \sum_v \theta^v \left(\sum_{(x,t)\in X^v} P(x)[g_v(x_{j(v)})]\right) \qquad \text{by definition of } g_v(\ell) \\
&= \sum_v \theta^v \left(\sum_{(x,t)\in X_1^v} P(x)[g_v(1)] + \sum_{(x,t)\in X_0^v} P(x)[g_v(0)]\right) \qquad \text{separating } X^v \text{ into } X_0^v \text{ and } X_1^v \\
&= \sum_v \theta^v \left(\sum_{(x,t)\in X_1^v} P(x^{j(v)})p_v[g_v(1)] + \sum_{(x,t)\in X_0^v} P(x^{j(v)})q_v[g_v(0)]\right) \\
&\qquad \text{since for } i\in\mathcal{N},\ P(x) = P(x^i)p_i \text{ if } x\in X_1^v \text{ and } P(x) = P(x^i)q_i \text{ if } x \in X_0^v \\
&= \sum_v \theta^v \left(\sum_{(x,t)\in X_1^v} P(x^{j(v)})[p_v g_v(1) + q_v g_v(0)]\right) \qquad \text{pairing } (x,t)\in X_1^v \text{ with } (x',t)\in X_0^v \\
&= \sum_v \theta^v \left(\sum_{(x,t)\in X_1^v} P(x^{j(v)})\Delta(v)\right) \qquad \text{by (3)} \\
&= \sum_v c_{j(v)} \left(\sum_{(x,t)\in X_1^v} P(x^{j(v)})\right) \qquad \text{by definition of } \theta(v) \\
&= \sum_v \sum_{(x,t)\in X^v} c_{j(v)} P(x) \qquad \text{by (4)} \\
&= \sum_{x\in\{0,1\}^n} \sum_{t=1}^{T^x} c_{j_x^t} P(x) \qquad \text{since the } X^v \text{ partition the } (x,t)
\end{aligned}$$

The final expression is equal to the expected cost of the cover constructed by Adaptive Greedy. □

For $w \in W$, let $h'_w(Y)$ denote the function of the variables $y_{S,a}$, computed in the left hand side of the constraint for w, in Dual LP. We will bound $h'_w(Y)$.

The analysis of Deshpande et al. relied on a bound given in a technical lemma of Wolsey, as quoted by Fujito [6]. We use a different bound of Wolsey, given in that same technical lemma [12].

Lemma 3. *[12] Given two sequences of real numbers,* $0 < \alpha^{(1)} \leq \alpha^{(2)} \leq \ldots \leq \alpha^{(T)}$ *and* $\beta^{(1)} \geq \beta^{(2)} \geq \ldots \geq \beta^{(T)} > 0$*, the following holds:*

$$\alpha^{(1)}\beta^{(1)} + (\alpha^{(2)} - \alpha^{(1)})\beta^{(2)} + \ldots + (\alpha^{(T)} - \alpha^{(T-1)})\beta^{(T)} \leq (\max_{1\leq t\leq T} \alpha^{(t)}\beta^{(t)}) \left[\ln \frac{\beta^{(1)}}{\beta^{(T)}} + 1\right]$$

Lemma 4 upper bounds the left hand side of the constraints for the $w \in W$, when evaluated at Y. Deshpande et al. proved an upper bound of $kc_j H(R)$ for integer-valued utility functions, where $H(m)$ is the mth harmonic number. We prove a bound of $c_j(\ln R/\eta_E + 1)$.

Lemma 4. *For every $x \in \{0,1\}^n$ and $j \in \mathcal{N}$, $h'_{x^j}(Y) \leq c_j(\ln \frac{R}{\eta_E} + 1)$.*

Proof. Fix $x \in \{0,1\}^n$ and $j \in \mathcal{N}$. Let x' be obtained from x by complementing x_j. Assume $x_j = 1$ (and $x'_j = 0$). Let $w = x^j$, that is, $w = x_{j \leftarrow *}$. Thus $w^{(1)} = x$ and $w^{(0)} = x'$. Consider Y.

Let τ be the decision tree corresponding to Adaptive Greedy. Consider the root-leaf paths of τ taken on x and x'. If j does not appear on these paths, then the paths are the same. Otherwise, they diverge on a node containing j. We consider these two cases separately.

In the first case, the paths are identical and $y_{S,x} = y_{S,x'}$ for all $S \subset \mathcal{N}$, Further, $T^x = T^{x'}$, $F^t_x = F^t_{x'}$ for all t with $0 \leq t \leq T^x$, and unless $S = F^t_x$ for some t with $0 \leq t \leq T^x - 1$, $y_{S,x} = y_{S,x'} = 0$.

Therefore, we have

$$\begin{aligned} h'_w(Y) &= \textstyle\sum_{S \subseteq \mathcal{N}} [p_j\, G_{S,w^{(1)}}(j)\, y_{S,w^{(1)}} + q_j\, G_{S,w^{(0)}}(j)\, y_{S,w^{(0)}}] \\ &= \textstyle\sum_{t=0}^{T^x - 1} y_{F^t_x,x} \left[p_j\, G_{F^t_x,w}(j,1) + q_j\, G_{F^t_x,w}(j,0)\right] \\ &= \textstyle\sum_{t=0}^{T^x-1} y_{F^t_x,x} \Delta^{t+1}_x(j) \qquad \text{by the definition of } \Delta^t_x \end{aligned}$$

In the second case, the paths diverge at a node labeled j. Let v be the node. Numbering the nodes on the path from the root to v, starting at 1, let t^* be the number of v. Then for $0 \leq t \leq t^* - 1$, $F^t_x = F^t_{x'}$, and for $1 \leq t \leq t^*$, and $\theta^t_x = \theta^t_{x'}$. For $t^* \leq t \leq T^x$, $j \in F^{t^*}_x$, so $G_{F^t_x,x}(j) = 0$ and hence $G_{F^t_x,x}(j)\, y_{S,x} = 0$. If $S \notin \{F^0_x, \ldots, F^{T^x-1}_x\}$, then $y_{S,x} = 0$. Thus if $G_{S,x}(j)\, y_{S,x} \neq 0$, then $S = F^t_x$ for some t where $0 \leq t \leq t^* - 1$. Similarly, if $G_{S,x'}(j)\, y_{S,x'} \neq 0$, $S = F^{t^*}_x$ for some t where $1 \leq t \leq t^* - 1$. Therefore, analogous to the other case, we have

$$h'_w(Y) \quad = \quad \sum_{t=0}^{t^*-1} y_{F^t_x,x} \Delta^{t+1}_x(j)$$

Recall that $\Delta^t_x(j)$ is the expected increase in utility during iteration t, on input x, if j were chosen in that iteration. By the assumption in the definition of the StSuC problem, there exists $\ell \in \{0,1\}$ such that $g_\star(j,\ell) > 0$, for $\star = (*, \ldots, *)$. Therefore, $\Delta^1_x(j) > 0$.

In the first case above, let $\hat{T}$ be the the maximum value of t such that $1 \leq t \leq T^x$ and $\Delta^t_x(j) > 0$. In the second, let $\hat{T}$ be the maximum value of t such that $1 \leq t \leq t^*$ and $\Delta^t_x(j) > 0$. In both cases, the first $\hat{T}$ nodes of the paths for x and x' in τ are identical, so $\Delta^t_x(j) = \Delta^t_{x'}(j)$ for $1 \leq t \leq \hat{T}$.

By the submodularity of g and the greedy rule used by Adaptive Greedy, the rate paid during each iteration of Adaptive Greedy, on input x, cannot decrease in subsequent iterations. Therefore, $0 < \theta^1_x \leq \ldots \leq \theta^{\hat{t}}_x$. By the submodularity of g, $\Delta^1_x(j) \geq \ldots \geq \Delta^{\hat{T}}_x(j) > 0$.

Thus Lemma 3 applies to the non-decreasing subsequence $\theta^1_x, \theta^2_x, \ldots, \theta^{\hat{T}}_x$ and the non-increasing subsequence $\Delta^1_x(j), \Delta^2_x(j), \ldots, \Delta^{\hat{T}}_x(j)$. Suppressing the subscript x on F^t, θ , and Δ for readability, and using our assumption that $w = x^j$, we have

$$\begin{aligned} h'_w(Y) &= \textstyle\sum_{t=0}^{\hat{T}} y_{F^t,x} \Delta^{t+1}(j) \\ &= \theta^1 \Delta^1(j) + \Sigma_{t=2}^{\hat{T}} (\theta^t - \theta^{t-1}) \Delta^t(j) \\ &\leq (\max_{1 \leq t \leq \hat{T}} \theta^t \Delta^t(j)) \left[\ln \tfrac{\Delta^1(j)}{\Delta^{\hat{T}}(j)} + 1\right] && \text{by Lemma 3} \\ &\leq c_j [\ln \tfrac{\Delta^1(j)}{\Delta^{\hat{T}}(j)} + 1] && \text{since } \theta^t_x \leq c_j / \Delta^t_x(j) \text{ by the Adaptive Greedy choice} \\ &\leq c_j [\ln(R/\Delta^{\hat{T}}_x) + 1] && \text{by the definition of R} \\ &\leq c_j [\ln(R/\eta_E) + 1] && \text{by the definition of } \eta_E \end{aligned}$$

$\square$

In fact, by the above analysis, the lemma (and the theorem below) still holds if R is redefined to be the largest *expected* increase in utility attainable from a single item.

By the lemmas above, and the previous analysis, we have the following theorem.

Theorem 1. *For the Stochastic Submodular Cover problem, the expected cost incurred by Adaptive Greedy is at most* $(\ln \frac{R}{\eta_E} + 1)$ *times the expected cost incurred by the optimal strategy.*

4 Generalizing the Deshpande et al. bound to real-valued utility functions

We can also prove an approximation bound of $k(\ln R/\eta_1 + 1)$ for Adaptive Greedy, where k is the size of the state space. This generalizes the bound of Deshpande et al. to utility functions that are not necessarily integer-valued. The proof is essentially the same as the proof in the previous section, except that it relies on a different upper bound on $h'_{x^j}(y)$. To obtain this upper bound, we use the same argument as that in Deshpande et al., but apply a different bound of Wolsey (the one given in Lemma 3, above). We state the upper bound here, and give a full proof for completeness.

Lemma 5. *For every* $x \in \{0,1\}^n$ *and* $j \in \{1, \ldots, N\}$, $h'_{x^j}(Y) \leq kc_j(\ln \frac{R}{\eta_1} + 1)$*, where* k *is the size of the state space.*

Proof. As above, we give the proof for the state space $\{0,1\}$ (i.e., where $k = 2$) but the proof easily generalizes to constant $k > 2$.

Fix $x \in \{0,1\}^n$ and $j \in N$. Let x' be the assignment obtained from x by complementing x_j. Without loss of generality, assume that $x_j = 1$ (and $x'_j = 0$).

Let $q_j = 1 - p_j$. Let $D^t_x(j)$ denote the amount of additional utility that would have been attained in iteration t of Adaptive Greedy, on input x, if item j had been chosen (rather than j^x_t). Let $D^t_{x'}(j)$ be the analogous value for x'.

Therefore, $D^t_x(j) = G_{F^t_x,x}(j)$, $D^t_{x'}(j) = G_{F^t_{x'},x'}(j)$, and we have

$$\Delta^t_x(j) = p_j(D^t_x(j)) + q_j(D^t_{x'}(j)). \tag{5}$$

Let κ be the value of t that maximizes $(\theta^t_x)(D^t_x(j))$. Similarly, let κ' be the value of t that maximizes $(\theta^t_{x'})(D^t_{x'}(j))$.

We have $\theta^1_x \leq \theta^2_x \ldots \leq \theta^{T^x}_x$. By the submodularity of g, $D^1_x(j) \geq D^2_x(j) \geq \ldots \geq D^{T^x}_x(j)$.

Also, by the definitions of η_1 and R, $D^{T^x}_x(j) \geq \eta_1$ and $D^1_x(j) = g_\star(j,1) \leq R$. Thus by Lemma 3,

$$\theta^1_x(D^1_x(j)) + \Sigma^{T^x}_{t=2}(\theta^i_x - \theta^{t-1}_x)(D^t_x(j)) \leq \theta^\kappa_x(D^\kappa_x(j))(\ln \frac{R}{\eta_1} + 1) \tag{6}$$

By an analogous argument:

$$\theta^1_x(D^1_0(j)) + \Sigma^{T^{x'}}_{t=2}(\theta^t_{x'} - \theta^{t-1}_{x'})(D^t_0(j)) \leq \theta^{\kappa'}_{x'}(D^\kappa_{x'}(j))(\ln \frac{R}{\eta_1} + 1) \tag{7}$$

Then:

$$
\begin{aligned}
&h'_{x^j}(Y) \\
&= \quad \textstyle\sum_{S \subseteq N} [p_j G_{S,x}(j) y_{S,x} + q_j G_{S,x'}(j) y_{S,x'}] \\
&= \quad \begin{array}{c} p_j[\theta_x^1(D_x^1(j)) + \Sigma_{t=2}^{T^x}(\theta_x^t - \theta_x^{i-1})(D_x^t(j))]+ \\ q_j[\theta_{x'}^1(D_{x'}^1(j)) + \Sigma_{t=2}^{T^{x'}}(\theta_{x'}^t - \theta_{x'}^{t-1})(D_{x'}^t(j))] \end{array} \\
&\leq \quad p_j \theta_x^\kappa(D_{x'}^\kappa(j))(\ln \tfrac{R}{\eta_1} + 1) + q_j \theta_{x'}^{\kappa'}(D_{x'}^{\kappa'}(j))(\ln \tfrac{R}{\eta_1} + 1) && \text{by (6) and (7)} \\
&\leq \quad (\ln \tfrac{R}{\eta_1} + 1)[p_j \theta_x^\kappa(D_x^\kappa(j)) + q_j \theta_{x'}^{\kappa'}(D_{x'}^{\kappa'}(j))] \\
&\leq \quad \begin{array}{r} (\ln \tfrac{R}{\eta_1} + 1)[p_j \theta_x^\kappa(D_x^\kappa(j)) + q_j \theta_x^\kappa(D_x^\kappa(j))+ \\ p_j \theta_{x'}^{\kappa'}(D_{x'}^{\kappa'}(j)) + q_j \theta_{x'}^{\kappa'}(D_{x'}^{\kappa'}(j))] \end{array} && \text{since this just adds extra non-negative terms} \\
&= \quad (\ln \tfrac{R}{\eta_1} + 1)(\theta_x^\kappa \Delta_x^\kappa(j) + \theta_{x'}^{\kappa'} \Delta_{x'}^{\kappa'}(j)) && \text{factoring out the } \theta\text{s and using (5)} \\
&\leq \quad (\ln \tfrac{R}{\eta_1} + 1)(c_j + c_j) && \text{due to the greedy choices of Algorithm 1} \\
&\leq \quad 2c_j(\ln \tfrac{R}{\eta_1} + 1) && \text{by submodularity and the definition of } R
\end{aligned}
$$

The 2 in the bound is replaced by k when there are k states. □

The bound on Adaptive Greedy then follows immediately from the arguments in the previous section.

Theorem 2. *For the Stochastic Submodular Cover problem, the expected cost incurred by Adaptive Greedy is at most $k(\ln \frac{R}{\eta_1} + 1)$ times the expected cost incurred by the optimal strategy, where k is the size of the state set.*

Acknowledgments

L. Hellerstein was partially supported by NSF Grant 1217968. D. Kletenik was partially supported by a PSC-CUNY Award, jointly funded by The Professional Staff Congress and The City University of New York.

References

[1] Eric Bach, Jeremie Dusart, Lisa Hellerstein, and Devorah Kletenik. Submodular goal value of Boolean functions. *arXiv preprint arXiv:1702.04067*, 2017.

[2] G. Bellala, S. Bhavnani, and C. Scott. Group-based active query selection for rapid diagnosis in time-critical situations. *IEEE Transactions on Information Theory*, 2012.

[3] Yuxin Chen, Shervin Javdani, Amin Karbasi, J. Andrew Bagnell, Siddhartha S. Srinivasa, and Andreas Krause. Submodular surrogates for value of information. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 3511–3518, 2015.

[4] 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*, pages 1453–1467. Society for Industrial and Applied Mathematics, 2014.

[5] Amol Deshpande, Lisa Hellerstein, and Devorah Kletenik. Approximation algorithms for stochastic submodular set cover with applications to boolean function evaluation and min-knapsack. *ACM Trans. Algorithms*, 12(3):42:1–42:28, 2016.

[6] T. Fujito. Approximation algorithms for submodular set cover with applications. *IEICE Transactions on Informations and Systems*, 83, 2000.

[7] D. Golovin and A. Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

[8] Nathaniel Grammel, Lisa Hellerstein, Devorah Kletenik, and Patrick Lin. Scenario submodular cover. In *Approximation and Online Algorithms - 14th International Workshop, WAOA 2016, Aarhus, Denmark, August 25-26, 2016, Revised Selected Papers*, volume 10138 of *Lecture Notes in Computer Science*, pages 116–128. Springer, 2017.

[9] Prabhanjan Kambadur, Viswanath Nagarajan, and Fatemeh Navidi. Adaptive submodular ranking. In *Proceedings of the 19th International Conference on Integer Programming and Combinatorial Optimization (IPCO)*, pages 317–329, 2017.

[10] Z. Liu, S. Parthasarathy, A. Ranganathan, and H. Yang. Near-optimal algorithms for shared filter evaluation in data stream systems. In *Proceedings of the 28th ACM SIGMOD International Conference on Management of Data,*, 2008.

[11] Feng Nan and Venkatesh Saligrama. Comments on the proof of adaptive stochastic set cover based on adaptive submodularity and its implications for the group identification problem in "group-based active query selection for rapid diagnosis in time-critical situations". *IEEE Transactions on Information Theory*, 2017.

[12] L. Wolsey. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2:385–393, 1982. 10.1007/BF02579435.