Journal of Artificial Intelligence Research 63 (2018) 265-279 Submitted 02/18; published 10/18

Revisiting the Approximation Bound for Stochastic Submodular Cover

Lisa Hellerstein LISA.HELLERSTEIN@NYU.EDU
Department of Computer Science and Engineering
NYU Tandon School of Engineering
2 Metrotech Center
Brooklyn, NY 11201

Devorah Kletenik KLETENIK@SCI.BROOKLYN.CUNY.EDU
Department of Computer and Information Science
Brooklyn College, City University of New York
2900 Bedford Avenue
Brooklyn, NY 11210

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 of a single item, and the utility function is integer-valued. This bound is similar to the $(\ln Q/\eta + 1)$ bound given by Golovin and Krause, whose analysis was recently found to have an error. Here $Q \geq R$ is the goal utility and η is the minimum gap between Q and any attainable utility $Q' < Q$. We revisit the proof of the $k(\ln R+1)$ bound of Deshpande et al., fill in the details of the proof of a key lemma, and prove two bounds for real-valued utility functions: $k(\ln R_1 + 1)$ and $(\ln R_E + 1)$. Here R_1 equals the maximum ratio between the largest increase in utility attainable from a single item, and the smallest non-zero increase attainable from that same item (in the same state). The quantity R_E equals the maximum ratio between the largest *expected* increase in utility from a single item, and the smallest non-zero *expected* increase in utility from that same item. Our bounds apply only to the stochastic setting with independent states.

1. Introduction

Golovin and Krause introduced the Stochastic Submodular Cover (StSuC) problem, a generalization of the classical NP-complete set cover problem that has applications to problems ranging from sensor placement to minimizing the cost of evaluating Boolean prediction functions (Golovin & Krause, 2011; Deshpande, Hellerstein, & Kletenik, 2016). In this problem, there are n "items". Each item can be in one of k states, and the state of each item is an independent random variable. Each item has a cost. The state of an item can only be determined by choosing it and incurring the associated cost. There is a utility function that assigns a utility value to each subset of items, and that value can depend on the states of the items. The utility function obeys certain monotonicity and submodularity properties. The problem is to sequentially choose a set of items that achieve a goal utility Q, with minimum expected cost. Formal definitions can be found in Section 2.

Golovin and Krause (2011) presented a proof showing that the Adaptive Greedy algorithm is a $(\ln Q/\eta + 1)$-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 $Q' < Q$. The quantity η can be viewed as a scaling factor.

Subsequently, Deshpande, Hellerstein, and Kletenik 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 (Deshpande et al., 2016). 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, $\eta \geq 1$, so $(\ln Q/\eta + 1) \leq (\ln Q + 1)$.

Recently, Nan and Saligrama (2017) discovered an error in Golovin and Krause's analysis of Adaptive Greedy. The error invalidates the $(\ln Q/\eta + 1)$ bound of Golovin and Krause for the StSuC problem. Golovin and Krause (2017) have since posted a new and more involved analysis, with a bound that is quadratic in $(\ln Q/\eta)$.

Given these developments, the $k(\ln R + 1)$ bound of Deshpande et al. (2016) is the only current approximation bound for the StSuC problem that is linear in $\ln Q$, for integer-valued utility functions. Because of this, we were motivated to revisit the bound.

We prove two variants of the $k(\ln R + 1)$ bound of Deshpande et al. (2016), removing the assumption that the utility function is integer valued: $k(\ln R_1 + 1)$ and $(\ln R_E + 1)$. Here R_1 equals the maximum ratio between the largest increase in utility attainable from a single item, and the smallest non-zero increase attainable from that same item (in the same state). The quantity R_E equals the maximum ratio between the largest *expected* increase in utility from a single item, and the smallest non-zero *expected* increase in utility from that same item. These bounds are similar to the $(\ln Q/\eta + 1)$ bound claimed by Golovin and Krause (2011). We obtain these bounds by tightening the analysis of Deshpande et al., and by using a different technical result of Wolsey (1982).

One of the key lemmas of Deshpande et al. (2016) 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 sufficient. We need this lemma for our new bounds and provide a complete proof of it below.

There are previous results, for other problems, that relied on the claimed bound of Golovin and Krause (2011) for the StSuC problem. Examples include a number of the results on the Stochastic Boolean Function Evaluation (SBFE) problem (e.g. Deshpande et al., 2016; Bach, Dusart, Hellerstein, & Kletenik, 2018). These previous results on the SBFE problem still hold, up to constant factors, by substituting the $k(\ln R + 1)$ bound of Deshpande et al. (2016), or the $k(\ln R_1 + 1)$ bound proved in this paper, for the claimed $(\ln Q/\eta + 1)$ bound.

It is easy to show that $R_E \leq R_1$ if the following property holds: whenever an item j yields non-zero utility in one state, it yields non-zero utility in its other states. However, there are natural StSuC problems where this property does not hold. For example, the Stochastic Set Coverage problem is a special case of the StSuC problem with an integer-valued utility function. It is motivated by covering locations with sensors, and a sensor might cover something or nothing depending on whether its state is "working" or "broken". Adaptive Greedy applied to the Stochastic Set Coverage problem yields an approximation bound of $(\ln Q + 1)$ (Parthasarathy, 2018).[1] This bound is not implied by the bounds in this paper.

1. Golovin and Krause (2011) cited an earlier paper of Liu et al. as having proved this $(\ln Q + 1)$ bound (Liu, Parthasarathy, Ranganathan, & Yang, 2008). However, the proof in that paper had an error.

In general, the quantity Q/η in the bound of Golovin and Krause (2011) is incomparable to R_1 and R_E. However, in many cases η is equal to the smallest non-zero increase in utility from a single item, and in these cases, $R_1 \leq Q/\eta$.

The bounds we prove in this paper do not extend beyond the StSuC problem to the more general class of adaptive submodular cover problems originally considered by Golovin and Krause. The proofs of the bounds in this paper require that item states be independent, and that the utility function be "pointwise" submodular. As a result, our bounds do not apply to previous work where Adaptive Greedy was used to solve the following problems: Equivalence Class Determination, Decision Region Identification, and Scenario (sample-based) Submodular Cover (e.g., Bellala, Bhavnani, & Scott, 2012; Chen, Javdani, Karbasi, Bagnell, Srinivasa, & Krause, 2015; Grammel, Hellerstein, Kletenik, & Lin, 2017). However, we note that there are other algorithms for solving these particular problems that achieve good approximation bounds. Grammel et al. (2017) presented two algorithms for solving Scenario Submodular Cover. While one of them used Adaptive Greedy, the other did not. Subsequently, Kambadur et al. (2017) presented an algorithm that uses just one simple greedy rule, and solves an even more general class of problems.

It remains an open question whether Adaptive Greedy achieves an approximation bound of $(\ln {}^{Q}\!/\!_{\eta} + 1)$ or $(\ln R_1 + 1)$ for the StSuC problem.

Finally, we note that Deshpande et al. (2016) presented another approximation algorithm for the StSuC problem which they called Dual Adaptive Greedy. That algorithm is an extension of Fujito's algorithm for the deterministic submodular cover problem, which was based on Hochbaum's dual greedy algorithm for the classical set cover problem (Fujito, 2000). These dual greedy algorithms yield approximation bounds of an entirely different form.

2. Definitions and Background

Let $\mathcal{N} = \{1, \ldots, n\}$ be a set of *items*. Let $\mathcal{O}$ be a finite set 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 item 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 $(*, \ldots, *)$.

We define a (state dependent) utility function to be a function $g : \{0, 1, *\}^n \to \mathbb{R}^{\geq 0}$. 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 $\{j \mid b_j \neq *\}$ when they are in the states indicated by b.

We also overload $g()$ to denote the utility of a partial assignment that is given by a subset of items and a state vector. In particular, for $S \subseteq \mathcal{N}$ and $x \in \{0, 1, *\}^n$, we define $g(S, x) = g(b)$ for the $b \in \{0, 1\}^n$ satisfying $b_j = x_j$ for $j \in S$, and $b_j = *$ 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 $j \in \mathcal{N}$ where $b_j \neq *$, let $\partial g_j(S, b) = g(S \cup \{j\}, b) - g(S, b)$. We adopt the partial derivative notation here to indicate that we are measuring the change in utility produced by "increasing" the value of b_j from $*$ to a value in $\{0,1\}$. In the notation $\partial g_j(S,b)$, the value of b_j indicates whether the $*$ is being set to 0 or 1. Note that if j is already in S, then $\partial g_j(S,b) = g(S \cup \{j\}, b) - g(S,b) = 0$.

For $S \subseteq \mathcal{N}$, $b \in \{0,1,*\}^n$, $j \in \mathcal{N}$ where $b_j = *$, and $\ell \in \{0,1\}$, let $\partial g_j(S,b,\ell) = g(S \cup \{j\}, b_{j\leftarrow \ell}) - g(S,b)$. Here the third argument ℓ indicates the changed value of b_j.

Similarly, for $b \in \{0,1,*\}^n$, $j \in \mathcal{N}$ such that $b_j = *$, and $\ell \in \{0,1\}$, let $\partial g_j(b,\ell) = g(b_{j\leftarrow\ell}) - g(b)$. So, $\partial g_j(b,\ell) = \partial g_j(\mathcal{N}, b, \ell)$.

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 $\partial g_j(a,\ell) \leq \partial g_j(b,\ell)$. (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 must 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 $j \in \mathcal{N}$, p_j and c_j satisfy $0 < p_j < 1$ and $c_j \in \mathbb{R}^+$. Further, 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 $\partial g_j(\star,\ell) > 0$. Otherwise, by submodularity, choosing j can never increase utility.

Parameter R in an approximation bound denotes $\max_{j\in\mathcal{N},\ell\in\{0,1\}} \partial g_j(\star,\ell)$. For $j \in \mathcal{N}$ and $\ell \in \{0,1\}$, let $\eta_1(j,\ell)$ be the minimum non-zero value of $\partial g_j(b,\ell)$ for any $b \in \{0,1,*\}^n$ where $b_j = *$. Let $r(j,\ell) = \partial g_j(\star,\ell)$. Since g is submodular, $r(j,\ell)$ is the largest increase in utility attainable from item j when it is in state ℓ. We use R_1 to denote the maximum value of the ratio ${r(j,\ell)}/{\eta_1(j,\ell)}$, over all $j \in \mathcal{N}$ and all $\ell \in \{0,1\}$.

Let $\eta_E(j)$ be the minimum non-zero value of $p_j(\partial g_j(b,1)) + q_j(\partial g_j(b,0))$, for any $b \in \{0,1,*\}^n$ where $b_j = *$. Let $r_E(j)$ be the maximum value of $p_j(\partial g_j(b,1)) + q_j(\partial g_j(b,0))$, for any $b \in \{0,1,*\}^n$ where $b_j = *$. By the submodularity of g, $r_E(j) = p_j(\partial g_j(\star,1)) + q_j(\partial g_j(\star,0))$. We use R_E to denote the maximum value of the ratio ${r_e(j)}/{\eta_E(j)}$ over all $j \in \mathcal{N}$.

A (feasible) solution to the StSuC problem is an adaptive strategy for choosing items from $\mathcal{N}$, so that utility Q is achieved, as measured by g. An adaptive strategy corresponds to a decision tree τ, where each internal node is labeled with an item j, and has a child for each of the possible states of j. An item j can appear at most once on a root-leaf path. Thus each root-leaf path in τ corresponds to a partial assignment b where for each j on the path, b_j equals the state of item j leading to the next node in the path. For each j not on the path, $b_j = *$. Partial assignment b must satisfy $g(b) = Q$.

Each $x \in \{0,1\}^n$ results in following a particular root-leaf path in the tree τ. We call the set of items on that path the *cover* constructed by τ on x. We define $cost(\tau, x)$ to be the sum of the costs of the items in the cover constructed by τ on x. The *expected cost of tree* τ is $\sum_{x\in\{0,1\}^n} P(x)\ cost(\tau, x)$,

The StSuC problem is to find an adaptive strategy satisfying the properties specified above, such that the expected cost of the corresponding decision tree τ is minimized. Thus an optimal solution minimizes $\sum_{x\in\{0,1\}^n} P(x)\ cost(\tau, x)$. However, it is not necessary to output decision tree τ explicitly (it may have exponential size). It is sufficient to find a procedure that can be used to determine, in polynomial time, which item should be chosen next in the sequence.

The Adaptive Greedy algorithm of Golovin and Krause (2011) solves the StSuC problem using the greedy rule that chooses the item that will yield the largest *expected* increase in utility, per unit cost. (Equivalently, the item minimizes the ratio between cost and expected increase in utility.) We give pseudocode in Figure 1, where we use x_j to denote the random state of item j. Thus $p_j = P(x_j = 1)$ and $q_j = P(x_j = 0)$. We treat $c_j/\Delta(j)$ as being equal to positive infinity whenever denominator $\Delta(j)$ equals 0. Lines in the pseudocode that are enclosed in square brackets assign values to variables that are not necessary for the processing of the algorithm, but that are used in the analysis.

$b \leftarrow (*, * \ldots, *)$
$F \leftarrow \emptyset$ $//F$ is set of items j chosen so far
[$t \leftarrow 0$]
while $g(b) < Q$ **do**
  [$t \leftarrow t + 1$]
  for $j \notin F$ **do**
    $\Delta(j) \leftarrow \sum_{\ell\in\{0,1\}} P(x_j = \ell)\ \partial g_j(b, \ell)$ // expected increase in utility if j is chosen
  end for
  $j^* \leftarrow \arg\min_{j\notin F} \frac{c_j}{\Delta(j)}$
  [$\theta \leftarrow \frac{c_{j^*}}{\Delta(j^*)}$]
  $\ell \leftarrow$ 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
[$T \leftarrow t$]
return b

Algorithm 1: Adaptive Greedy

3. The $(\ln R_E + 1)$ Bound

We start by reviewing some of the necessary background from the analysis of Deshpande et al. (2016). We then present our two key lemmas and complete the proof of the $(\ln R_E + 1)$ bound.

3.1 Background from Deshpande et al.

The starting point of the analysis of Deshpande et al. (2016) is the definition of a special LP whose optimal value lower bounds the optimal expected cost for the StSuC problem. This LP and a modified form of its dual are used only in the analysis of Adaptive Greedy. They do not play any role in the Adaptive Greedy algorithm itself.

We present the LP in Figure 2, and call it LPI. It is based on an integer program (IP) used by Wolsey (1982) to obtain an approximation bound for the *deterministic* Submodular Cover problem. The deterministic problem can be viewed as a version of the StSuC problem where the state of each item is known in advance and the problem is simply to choose the min-cost subset of items achieving goal utility Q. In this case the utility function g is a standard set function $g : 2^{\mathcal{N}} \rightarrow \mathbb{R}^{\geq 0}$ that is monotone and submodular as per the standard definitions (for all $S', S \subseteq \mathcal{N}$ with $S' \subseteq S$, and all $j \in \mathcal{N}$, $g(S') \leq g(S)$ and $g(S' \cup \{j\}) - g(S') \geq g(S \cup \{j\}) - g(S)$). In Wolsey's IP, there is a variable $z_j \in \{0,1\}$ associated with each item j, where $z_j = 1$ means item j is included in the chosen subset. Wolsey's IP is shown in Figure 1.

$$\text{Minimize } \sum_{j \in \mathcal{N}} c_j z_j$$

s.t.

$$\sum_{j \in \mathcal{N}} [g(S \cup \{j\}) - g(S)] z_j \geq Q - g(S) \quad \forall S \subseteq \mathcal{N}$$

$$z_j \in \{0,1\} \quad \forall j \in \mathcal{N}$$

Figure 1: Wolsey's IP

The number of constraints of the IP is exponential in n. Using the monotonicity and submodularity of g, Wolsey showed that an assignment $z \in \{0,1\}^n$ satisfies these constraints iff the subset represented by z has utility equal to $Q = g(\mathcal{N})$, that is $g(\{j \mid z_j = 1\}) = Q$.

A solution to the StSuC problem is a decision tree that constructs a cover for each possible assignment of states to items. Thus it constructs feasible solutions (covers) to 2^n instances of the deterministic Submodular Cover problem, one for each of the 2^n assignments $a \in \{0,1\}^n$ of states to items. Further, these covers are related to each other, because they correspond to overlapping paths in a single decision tree.

LPI can be seen as a relaxation of the StSuC problem. The idea is to find covers for the 2^n assignments $x \in \{0,1\}^n$ which do not have to come from a decision tree, but instead must satisfy a weaker property, which we describe now. 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 unique j such that $w_j = *$. For $\ell \in \{0,1\}$, let $w^{(\ell)}$ be the full assignment produced by taking the * bit of w and setting it to ℓ; that is $w^{(\ell)} = w_{J(w) \leftarrow \ell}$. 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.

LPI has a single variable z_w for each $w \in W$. Setting $z_w = 1$ corresponds to including item $J(w)$ in the covers for both $w^{(0)}$ and $w^{(1)}$, and setting it to 0 corresponds to excluding it from both covers. Thus a 0/1 assignment Z to the variables z_w associates a subset $F(a)$ with each $a \in \{0,1\}^n$ as follows: $F(a) = \{j \in \mathcal{N} \mid \text{ for } w = a_{j\leftarrow *}, z_w = 1\}$. If Z satisfies the LP constraints, then for each a, $F(a)$ is a cover for a (this follows directly from the correctness of Wolsey's IP). Further, the value of the objective function on Z equals the expected cost of cover $F(a)$ for $a \sim D_p$. To illustrate the definitions, we present the following example:

Example 1. *For $n = 3$, $W = \{*00, *01, *10, *11, 0{*}0, 0{*}1, 1{*}0, 1{*}1, 00*, 01*, 10*, 11*\}$. Let $w = *00$. Then $w^{(0)} = 000$ and $w^{(1)} = 100$. If $z_{*00} = 1, z_{0*0} = 0$, and $z_{00*} = 1$, then $F(000) = \{1, 3\}$.*

If we further constrain the variables of LPI 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.

We now make the simplifying assumption that each p_i is strictly between 0 and 1, so that $P(a) \neq 0$ for all $a \in \{0,1\}^n$, and $P(w) \neq 0$ for all $w \in W$. This assumption was implicit in the proof of Deshpande et al. (2014). It is relatively straightforward to extend their proof, and the proof below, to remove this assumption. The idea is modify the definition of LPI, by omitting all constraints associated with $a \in \{0,1\}^n$ where $P(a) = 0$, and all terms of the objective function containing variables z_w where $P(w) = 0$. This results in a modified version of LPII, which omits the constraints corresponding to w, j where $P(w) = 0$ and the terms of the objective function with $y_{S,a}$ where $P(a) = 0$. The rest of the proof requires only small changes to deal with realizations that have no "neighbor". We omit the details here.

These observations imply the following lemma.

Lemma 1. *(Deshpande et al., 2014) The optimal value of the LP in Figure 2 lower bounds the expected cost of the optimal strategy solving the associated StSuC instance.*

Take each constraint of LPI 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 LPII in Figure 3. Thus LPII can be viewed as a non-standard version of the dual of LPI. For readability, in LPII we have designated each constraint as corresponding to a pair w, j. We could have designated the constraint as corresponding to just w, since we require $j = J(w)$.

By strong duality, the optimal value of LPII is equal to the optimal value of LPI, and thus also lower bounds the expected cost of the optimal strategy. The remainder of our analysis will use LPII.

The basic idea of the analysis of Adaptive Greedy is to define an assignment Y to the variables $y_{S,a}$ of LPII 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). The analysis we give below is the same as that of Deshpande et al., except in the proofs of two key lemmas, Lemma 2 and Lemma 3, below.

In Lemma 2, we show that the value of the objective function of LPII, on assignment Y, equals the expected cost incurred by Adaptive Greedy. This is the lemma presented by Deshpande et al. without adequate proof.

$$
\begin{array}{ll}
\text{Minimize } \sum_{w \in W} c_{J(w)}\ P(w)\ z_w & \\
\text{s.t.} & \\
\sum_{j \in \mathcal{N}} \partial g_j(S,a)\ z_{a_{j \leftarrow *}} \geq Q - g(S,a) & \forall a \in \{0,1\}^n, S \subseteq N \\
z_w \geq 0 & \forall w \in W
\end{array}
$$

Figure 2: LPI

$$
\begin{array}{ll}
\text{Maximize } \sum_{a \in \{0,1\}^n} \sum_{S \subseteq \mathcal{N}} P(a)\ (Q - g(S,a))\ y_{S,a} & \\
\text{s.t.} & \\
\sum_{S \subseteq \mathcal{N}} \sum_{\ell \in \{0,1\}} P(x_j = \ell)\ \partial g_j(S, w^{(\ell)})\ y_{S,w^{(\ell)}} \leq c_j & \forall w, j \text{ s.t. } w \in W,\ j = J(w) \\
y_{S,a} \geq 0 & \forall S \subseteq N, a \in \{0,1\}^n
\end{array}
$$

Figure 3: LPII

In Lemma 3, we show that Y exceeds the right hand side of the constraints of LPII by a factor of at most $c_j(\ln R_E + 1)$. (An analogous lemma of Deshpande et al. shows that for integer valued utility functions, Y exceeds the right hand side of each constraint of LPII by a factor of at most $k(\ln R + 1)$.)

Let Y' be the result of dividing each entry in Y by $(\ln R_E + 1)$. By Lemma 3, Y' is a feasible solution to LPII.

The analysis of Adaptive Greedy can then be completed as follows. Let OPTDT be the expected cost of the optimal strategy and let OPTII be the optimal value of LPII. Let AGCOST be the expected cost of Adaptive Greedy and let $q(y)$ denote the objective function of LPII.

Then:

$$
\begin{aligned}
q(Y') &\leq \mathsf{OPTII} && \text{because } Y' \text{ is a feasible solution to LPII} \\
\Rightarrow q(Y') &\leq \mathsf{OPTDT} && \text{since } \mathsf{OPTII} \leq \mathsf{OPTDT} \text{ by Lemma 1} \\
\Rightarrow q(Y) &\leq (\ln R_E + 1)\mathsf{OPTDT} && \text{by definition of } Y' \text{ and the linearity of } q() \\
\Rightarrow \mathsf{AGCOST} &\leq (\ln R_E + 1)\mathsf{OPTDT} && \text{since } q(Y) = \mathsf{AGCOST} \text{ by Lemma 2}
\end{aligned}
$$

3.2 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^t_x and F^t_x be the values of b and F at the end of iteration t. So $|F^t_x| = t$, and b^t_x represents the states of items in F^t_x. If $j \notin F^{t-1}_x$ then let $\Delta^t_x(j)$ be the value of $\Delta(j)$ at the end of iteration t, else let $\Delta^t_x(j) = 0$. Let j^t_x be the item j^* chosen in iteration t. Let θ^t_x be the value of $c_j/\Delta^t_x(j)$ for $j = j^t_x$. We refer to θ^t_x as the *rate* in iteration t.

Define Y to be the assignment to the variables in LPII such that for all $x \in \{0,1\}^n$:

$$y_{S,x} = \begin{cases} \theta^1_x \text{ if } S = F^0 \\ (\theta^{t+1}_x - \theta^t_x) \text{ if } S = F^t \text{ and } t \in \{1 \dots T^x - 1\} \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

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

Deshpande et al. (2014) claimed that Lemma 2 below followed directly from a result of Wolsey (1982) by linearity of expectation. 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 $c_j/\partial g_j(F,x)$, 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.

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):

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

Therefore (restoring subscripts and superscript x):

$$E[q^x(Y)] = \sum_{x \in \{0,1\}^n} \sum_{t=1}^{T^x} P(x)[g(F_x^t, x) - g(F_x^{t-1}, x)]\theta_x^t \tag{2}$$

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. Let $X = \{(x,t) \mid x \in \{0,1\}^n, 1 \le t \le T^x\}$. 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 $1 \le t \le T^x$, the pair (x,t) belongs to exactly one set X^v, and the X^v form a partition of X. 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 $j(v)$ be the item labeling v. We define $p_v = p_{j(v)}$ and $q_v = q_{j(v)}$. We define $X_1^v = \{(x,t) \in X^v | j(v) = 1\}$, and $X_0^v = \{(x,t) \in X^v | j(v) = 0\}$. 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 $j(v)$. 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 F^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 F^v, corresponding to the path down to (but not including) node v.

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

$$g(F_x^t, x) - g(F_x^{t-1}, x) = g(F^v \cup \{j(v)\}, x) - g(F^v, x) \tag{3}$$

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

We define $\Delta(v)$ to be the expected increase in utility at node v, that is

$$\Delta(v) = p_v \partial g_v(1) + q_v \partial g_v(0) \tag{4}$$

Clearly, $P(b^v) = \sum_{(x,t) \in X^v} P(x)$ and $p_v P(b^v) = \sum_{(x,t) \in X_1^v} P(x)$. Since for $(x,t) \in X_1^v$, $P(x_{j(v) \leftarrow *}) = (1/p_v) P(x)$, it follows that $\sum_{(x,t) \in X_1^v} P(x_{j(v) \leftarrow *}) = P(b^v)$ and hence

$$\sum_{(x,t) \in X_1^v} P(x_{j(v) \leftarrow *}) = \sum_{(x,t) \in X^v} P(x) \tag{5}$$

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

Thus,

$E[q^x(Y)]$

$$= \sum_{x \in \{0,1\}^n} \sum_{t=1}^{T^x} P(x)[g(F_x^t, x) - g(F_x^{t-1}, x)]\theta_x^t \qquad \text{by (2)}$$

$$= \sum_{v} \sum_{(x,t) \in X^v} P(x)[g(F^v \cup \{j(v)\}, x) - g(F^v, x)]\theta^v$$

by (3) since the X^v partition the (x,t)

$$
\begin{aligned}
&= \sum_v \theta^v \Big(\sum_{(x,t)\in X^v} P(x)[g(F^v \cup \{j(v)\}, x) - g(F^v, x)] \Big) && \text{moving } \theta^v \text{ forward} \\
&= \sum_v \theta^v \Big(\sum_{(x,t)\in X^v} P(x)[\partial g_v(x_{j(v)})] \Big) && \text{by definition of } \partial g_v(\ell) \\
&= \sum_v \theta^v \Big(\sum_{(x,t)\in X_1^v} P(x)[\partial g_v(1)] + \sum_{(x,t)\in X_0^v} P(x)[\partial g_v(0)] \Big) && \text{separating } X^v \text{ into } X_0^v \text{ and } X_1^v \\
&= \sum_v \theta^v \Big(\sum_{(x,t)\in X_1^v} P(x_{j(v)\leftarrow *}) p_v[\partial g_v(1)] + \sum_{(x,t)\in X_0^v} P(x_{j(v)\leftarrow *}) q_v[\partial g_v(0)] \Big) \\
& \qquad \text{since for } i \in \mathcal{N},\ P(x) = P(x_{i\leftarrow *})p_i \text{ if } x_i = 1 \text{ and } P(x) = P(x_{i\leftarrow *})q_i \text{ if } x_i = 0 \\
&= \sum_v \theta^v \Big(\sum_{(x,t)\in X_1^v} P(x_{j(v)\leftarrow *})[p_v \partial g_v(1) + q_v \partial g_v(0)] \Big) && \text{pairing } (x,t) \in X_1^v \text{ with } (x',t) \in X_0^v \\
&= \sum_v \theta^v \Big(\sum_{(x,t)\in X_1^v} P(x_{j(v)\leftarrow *}) \Delta(v) \Big) && \text{by (4)} \\
&= \sum_v c_{j(v)} \Big(\sum_{(x,t)\in X_1^v} P(x_{j(v)\leftarrow *}) \Big) && \text{by definition of } \theta(v) \\
&= \sum_v \sum_{(x,t)\in X^v} c_{j(v)} P(x) && \text{by (5)} \\
&= \sum_{x\in\{0,1\}^n} \sum_{t=1}^{T^x} c_{j_x^t} P(x) && \text{since the } X^v \text{ partition the } (x,t)
\end{aligned}
$$

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 (and the associated $j = J(w)$), in LPII. We will bound $h'_w(Y)$.

The analysis of Deshpande et al. relied on a bound given in a technical lemma of Wolsey (1982), as quoted by Fujito (2000). We use a different bound of Wolsey, which we present here. It comes from the same technical lemma.

Wolsey's Bound. *(Wolsey, 1982) Given two sequences of real numbers,* $0 < \alpha^{(1)} \le \alpha^{(2)} \le \ldots \le \alpha^{(T)}$ *and* $\beta^{(1)} \ge \beta^{(2)} \ge \ldots \ge \beta^{(T)} > 0$*, the following holds:*

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

The next lemma, Lemma 3, upper bounds the left hand side of the constraints for the $w \in W$, when evaluated at Y.

Lemma 3. *For every* $w \in W$ *and* $j = J(w)$, $h'_w(Y) \le c_j(\ln R_E + 1)$.

Proof. Let $x = w^{(1)}$ and $x' = w^{(0)}$.

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, $T^x = T^{x'}$, $F_x^t = F_{x'}^t$ for all t with $0 \leq t \leq T^x$, and unless $S = F_x^t$ for some t with $0 \leq t \leq T^x - 1$, $y_{S,x} = y_{S,x'} = 0$.

Therefore, we have

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

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^v be the number of v. Then for $0 \leq t \leq t^v - 1$, $F_x^t = F_{x'}^t$, and for $1 \leq t \leq t^v$, and $\theta_x^t = \theta_{x'}^t$. For $t^v \leq t \leq T^x$, $j \in F_x^{t^v}$, so $\partial g_j(F_x^t, x) = 0$ and hence $\partial g_j(F_x^t, x) \ y_{S,x} = 0$. If $S \notin \{F_x^0, \ldots, F_x^{T^x - 1}\}$, then $y_{S,x} = 0$. Thus if $\partial g_j(S, x) \ y_{S,x} \neq 0$, then $S = F_x^t$ for some t where $0 \leq t \leq t^v - 1$. Similarly, if $\partial g_j(S, x') \ y_{S,x'} \neq 0$, $S = F_{x'}^t$ for some t where $1 \leq t \leq t^v - 1$. Therefore, analogous to the other case, we have

$$
h'_w(Y) = \sum_{t=0}^{t^v - 1} y_{F_x^t, x} \Delta_x^{t+1}(j)
$$

Recall that $\Delta_x^t(j)$ 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 $\partial g_j(\star, \ell) > 0$, for $\star = (*, \ldots, *)$. Therefore, $\Delta_x^1(j) > 0$.

In the first case above, let $\hat{T}$ be the the maximum value of t such that $1 \leq t \leq T^x$ and $\Delta_x^t(j) > 0$. In the second, let $\hat{T}$ be the maximum value of t such that $1 \leq t \leq t^v$ and $\Delta_x^t(j) > 0$. In both cases, the first $\hat{T}$ nodes of the paths for x and x' in τ are identical, so $\Delta_x^t(j) = \Delta_{x'}^t(j)$ 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, $0 < \theta_x^1 \leq \ldots \leq \theta_x^{\hat{T}}$. By the submodularity of g, $\Delta_x^1(j) \geq \ldots \geq \Delta_x^{\hat{T}}(j) > 0$.

Thus Wolsey's bound applies to the non-decreasing subsequence $\theta_x^1, \theta_x^2, \ldots, \theta_x^{\hat{T}}$ and the non-increasing subsequence $\Delta_x^1(j), \Delta_x^2(j), \ldots, \Delta_x^{\hat{T}}(j)$. Suppressing the subscript x on F^t, θ , and Δ for readability, and using the fact that $w = x_{j \leftarrow *}$, we have

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

□

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* $(\ln R_E + 1)$ *times the expected cost incurred by the optimal strategy.*

4. The $k(\ln R_1 + 1)$ Bound

We can also prove an approximation bound of $k(\ln R_1 + 1)$ for Adaptive Greedy, where k is the size of the state space. This generalizes the bound of Deshpande et al. (2016) 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 $h'_w(Y)$. To obtain this upper bound, we use almost the same argument as that in the proof of Lemma 8 in the work of Deshpande et al. (2016), but apply a different bound of Wolsey (1982) (the one given above). We include the full argument here for completeness. As in the previous section, we make the simplifying assumption that all the p_i are strictly between 0 and 1.

Lemma 4. *For every* $w \in W$ *and* $j = J(w)$, $h'_w(Y) \leq kc_j(\ln R_1 + 1)$, *where* k *is the size of the state space.*

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

Let $x = w^{(1)}$ and $x' = w^{(0)}$. Let $D^t_x(j) = \partial g_j(F^t_x, x)$ and $D^t_{x'}(j) = \partial g_j(F^t_{x'}, x')$. Thus $D^t_x(j)$ denotes 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 item j^x_t). $D^t_{x'}(j)$ is the analogous value for x'. Thus

$$
\Delta^t_x(j) = p_j(D^t_x(j)) + q_j(D^t_{x'}(j)). \tag{6}
$$

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, $D^1_x(j) \geq D^2_x(j) \geq \ldots \geq D^{T^x}_x(j)$.

Since $x_j = 1$, $D^1_x(j) = \partial g_j(\star, 1)$. By the definition of $\eta_1(j)$, $D^t_x(j) \geq \eta_1(j)$ for every t such that $D^t_x(j) > 0$. Thus by the definition of R_1 and Wolsey's bound,

$$\theta^1_x(D^1_x(j)) + \Sigma_{t=2}^{T^x}(\theta^t_x - \theta^{t-1}_x)(D^t_x(j)) \leq \theta^\kappa_x(D^\kappa_x(j))(\ln R_1 + 1) \tag{7}$$

Similarly:

$$\theta^1_{x'}(D^1_{x'}(j)) + \Sigma_{t=2}^{T^{x'}}(\theta^t_{x'} - \theta^{t-1}_{x'})(D^t_{x'}(j)) \leq \theta^{\kappa'}_{x'}(D^{\kappa'}_{x'}(j))(\ln R_1 + 1) \tag{8}$$

Then:

$$\begin{aligned}
&h'_w(Y)\\
&= \sum_{S \subseteq \mathcal{N}} [p_j \partial g_j(S, x) y_{S,x} + q_j \partial g_j(S, x') y_{S,x'}]\\
&= \begin{array}{l} p_j[\theta^1_x(D^1_x(j)) + \Sigma_{t=2}^{T^x}(\theta^t_x - \theta^{t-1}_x)(D^t_x(j))] + \\ q_j[\theta^1_{x'}(D^1_{x'}(j)) + \Sigma_{t=2}^{T^{x'}}(\theta^t_{x'} - \theta^{t-1}_{x'})(D^t_{x'}(j))] \end{array}\\
&\leq p_j \theta^\kappa_x(D^\kappa_x(j))(\ln R_1 + 1) + q_j \theta^{\kappa'}_{x'}(D^{\kappa'}_{x'}(j))(\ln R_1 + 1) && \text{by (7) and (8)}\\
&\leq (\ln R_1 + 1)[p_j \theta^\kappa_x(D^\kappa_x(j)) + q_j \theta^{\kappa'}_{x'}(D^{\kappa'}_{x'}(j))]\\
&\leq \begin{array}{l} (\ln R_1 + 1)[p_j \theta^\kappa_x(D^\kappa_x(j)) + q_j \theta^\kappa_x(D^\kappa_{x'}(j)) + \\ \qquad p_j \theta^{\kappa'}_{x'}(D^{\kappa'}_x(j)) + q_j \theta^{\kappa'}_{x'}(D^{\kappa'}_{x'}(j))] \end{array}\\
& && \text{since this just adds extra non-negative terms}\\
&= (\ln R_1 + 1)(\theta^\kappa_x \Delta^\kappa_x(j) + \theta^{\kappa'}_{x'} \Delta^{\kappa'}_{x'}(j)) && \text{factoring out the } \theta\text{s and using (6)}\\
&\leq (\ln R_1 + 1)(c_j + c_j) && \text{by the Adaptive Greedy choices}\\
&\leq 2c_j(\ln R_1 + 1)
\end{aligned}$$

The factor of 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 $k(\ln R_1 + 1)$ 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

Bach, E., Dusart, J., Hellerstein, L., & Kletenik, D. (2018). Submodular goal value of boolean functions. *Discrete Applied Mathematics*, *238*, 1–13.

Bellala, G., Bhavnani, S., & Scott, C. (2012). Group-based active query selection for rapid diagnosis in time-critical situations. *IEEE Transactions on Information Theory*, *58*, 459–478.

Chen, Y., Javdani, S., Karbasi, A., Bagnell, J. A., Srinivasa, S. S., & Krause, A. (2015). Submodular surrogates for value of information. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pp. 3511–3518.

Deshpande, A., Hellerstein, L., & Kletenik, D. (2014). 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*, pp. 1453–1467. Society for Industrial and Applied Mathematics.

Deshpande, A., Hellerstein, L., & Kletenik, D. (2016). Approximation algorithms for stochastic submodular set cover with applications to boolean function evaluation and min-knapsack. *ACM Transactions on Algorithms (TALG)*, *12*(3), 42:1–42:28.

Fujito, T. (2000). Approximation algorithms for submodular set cover with applications. *IEICE Transactions on Information and Systems*, *83*(3), 480–487.

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

Golovin, D., & Krause, A. (2017). Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *CoRR*, *abs/1003.3967v5*.

Grammel, N., Hellerstein, L., Kletenik, D., & Lin, P. (2017). Scenario submodular cover. In *Approximation and Online Algorithms - 14th International Workshop, WAOA 2016, Aarhus, Denmark, August 25-26, 2016, Revised Selected Papers*, Vol. 10138 of *Lecture Notes in Computer Science*, pp. 116–128. Springer.

Kambadur, P., Nagarajan, V., & Navidi, F. (2017). Adaptive submodular ranking. In *Proceedings of the 19th International Conference on Integer Programming and Combinatorial Optimization (IPCO)*, pp. 317–329.

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

Nan, F., & Saligrama, V. (2017). 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*, *63*(11), 7612–7614.

Parthasarathy, S. (2018). An analysis of the greedy algorithm for stochastic set cover. *CoRR*, *abs/1803.07639*.

Wolsey, L. (1982). An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, *2*, 385–393.