
Optimization for Approximate Submodularity

Avinatan Hassidim
Bar Ilan University and Google
avinatan@cs.biu.ac.il

Yaron Singer
Harvard University
yaron@seas.harvard.edu

Abstract

We consider the problem of maximizing a submodular function when given access to its approximate version. Submodular functions are heavily studied in a wide variety of disciplines since they are used to model many real world phenomena and are amenable to optimization. There are many cases however in which the phenomena we observe is only *approximately* submodular and the optimization guarantees cease to hold. In this paper we describe a technique that yields strong guarantees for maximization of monotone submodular functions from approximate surrogates under cardinality and intersection of matroid constraints. In particular, we show tight guarantees for maximization under a cardinality constraint and $1/(1 + P)$ approximation under intersection of P matroids.

1 Introduction

In this paper we study maximization of *approximately* submodular functions. For nearly half a century submodular functions have been extensively studied since they are amenable to optimization and are recognized as an effective modeling instrument. In machine learning, submodular functions capture a variety of objectives that include entropy, diversity, image segmentation, and clustering.

Although submodular functions are used to model real world phenomena, in many cases the functions we encounter are only *approximately* submodular. This is either due to the fact that the objectives are not exactly submodular, or alternatively they are, but we only observe their approximate version.

In the literature, approximate utility functions are modeled as surrogates of the original function that have been corrupted by random variables drawn i.i.d from some distribution. Some examples include:

- **Revealed preference theory.** Luce’s famous model assumes that an agent’s revealed utility $\tilde{f} : 2^N \rightarrow \mathbb{R}$ can be approximated by a well-behaved utility function $f : 2^N \rightarrow \mathbb{R}$ s.t. $\tilde{f}(S) = f(S) + \xi_S$ for every $S \subseteq N$ where ξ_S is drawn i.i.d from a distribution [CE16]. \tilde{f} is a utility function that approximates f , and multiple queries to \tilde{f} return the same response;
- **Statistics and learning theory.** The assumption in learning is that the data we observe is generated by $\tilde{f}(\mathbf{x}) = f(\mathbf{x}) + \xi_{\mathbf{x}}$ where f is in some well-behaved hypothesis class and $\xi_{\mathbf{x}}$ is drawn i.i.d from some distribution. The use of \tilde{f} is not to model corruption by noise but rather the fact that data is not *exactly* manufactured by a function in the hypothesis class;
- **Active learning.** There is a long line of work on noise-robust learning where one has access to a noisy membership oracle $\tilde{f}(\mathbf{x}) = f(\mathbf{x}) + \xi_{\mathbf{x}}$ and for every \mathbf{x} we have that $\xi_{\mathbf{x}}$ is drawn i.i.d from a distribution [Ang88, GKS90, Jac94, SS95, BF02, Fel09]. In this model as well, the oracle is consistent and multiple queries return the same response. For set functions, one can consider active learning in experimental design applications where the objective function is often submodular and the goal would be to optimize $f : 2^N \rightarrow \mathbb{R}$ given \tilde{f} .

Similar to the above examples, we say that a function $\tilde{f} : 2^N \rightarrow \mathbb{R}_+$ is *approximately submodular* if there is a submodular function $f : 2^N \rightarrow \mathbb{R}_+$ and a distribution \mathcal{D} s.t. for each set $S \subseteq N$ we have that $\tilde{f}(S) = \xi_S f(S)$ where ξ_S is drawn i.i.d from \mathcal{D} .¹ The modeling assumption that ξ_S is not adversarially chosen but drawn i.i.d is crucial. Without this assumption for $\tilde{f}(S) = \xi_S f(S)$ where $\xi_S \in [1 - \epsilon, 1 + \epsilon]$ even for subconstant $\epsilon > 0$ no algorithm can obtain an approximation strictly better than $n^{-1/2}$ to maximizing either \tilde{f} or f under a cardinality constraint when $n = |N|$ [HS17]. This hardness stems from the fact that approximate submodularity implies that the function is close to submodular, but its marginals (or gradients) are not well approximated by those of the submodular function. When given α approximation to the marginals, the greedy algorithm produces a $1 - 1/e^\alpha$ approximations. Furthermore, even for *continuous* submodular functions, a recent line of work shows that gradient methods produce strong guarantees with approximate gradient information of the function [HSK17, CHHK18, MHK18].

In contrast to the vast literature on submodular optimization, optimization of approximate submodularity is nascent. For distributions bounded in $[1 - \epsilon/k, 1 + \epsilon/k]$ the function is sufficiently close to submodular for the approximation guarantees of the greedy algorithm to go through [HS, LCSZ17, QSY⁺17]. Without this assumptions, the greedy algorithm performs arbitrarily poorly. In [HS17] the authors give an algorithm that obtains an approximation arbitrarily close to $1 - 1/e$ under a cardinality constraint that is sufficiently large. For arbitrary cardinality and general matroid constraints there are no known approximation guarantees.

1.1 Our Contribution

In this paper we consider the problem $\max_{S \in \mathcal{F}} f(S)$ when $f : 2^N \rightarrow \mathbb{R}$ is non-negative monotone submodular defined of a ground set N of size n and the algorithm is only given access to an approximate surrogate \tilde{f} and \mathcal{F} is a uniform matroid (cardinality) or intersection of matroids constraint. We introduce a powerful technique which we call the *sampled mean approximation* and show:

- **Optimal guarantees for maximization under a cardinality constraint.** In [HS17] the result gives an approximation arbitrarily close to $1 - 1/e$ for $k \in \Omega(\log \log n)$. This is a fundamental limitation in their technique that initializes the solution with a set of size $\Omega(\log \log n)$ used for “smoothing” the approximately submodular function (see Appendix G.1 for more details). The technique in this paper is novel and yields an approximation of $1 - 1/e$ for any $k \geq 2$, and $1/2$ for $k = 1$, which is information theoretically tight, as we later show;
- **$1/(1 + P)$ approximation for intersection of P matroids.** We utilize the sampled mean approximation method to produce the first results for the more challenging case of maximization under general matroid constraints. Our approximation guarantees are comparable with those achievable with a greedy algorithm for monotone submodular functions;
- **Information theoretic lower bounds.** We show that no randomized algorithm can obtain an approximation strictly better than $1/2 + \mathcal{O}(n^{-1/2})$ for $\max_{a \in N} f(a)$ given an approximately submodular oracle \tilde{f} , and that no randomized algorithm can obtain an approximation strictly better than $(2k - 1)/2k + \mathcal{O}(n^{-1/2})$ for maximization under cardinality constraint k ;
- **Bounds in Extreme Value Theory.** As we later discuss, some of our results may be of independent interest to Extreme Value Theory (EVT) which studies the bounds on the maximum sample (or the top samples) from some distribution. To achieve our main result we prove subtle properties about extreme values of random variables where not all samples are created equal and the distributions generalize those typically studied in EVT.

The results above are for the problem $\max_{S \in \mathcal{F}} f(S)$ when the algorithm is given access to \tilde{f} . In some applications, however, \tilde{f} is the function that we actually wish to optimize, i.e. our goal is to solve $\max_{S \in \mathcal{F}} \tilde{f}(S)$. If $\tilde{f}(S)$ approximates $f(S)$ well on all sets S , we can use the solution for $\max_{S \in \mathcal{F}} f(S)$ as a solution for $\max_{S \in \mathcal{F}} \tilde{f}(S)$. In general, however, a solution that is good for $\max_{S \in \mathcal{F}} f(S)$ can be arbitrarily bad for $\max_{S \in \mathcal{F}} \tilde{f}(S)$. In Appendix E we give a black-box reduction showing that these problems are essentially equivalent. Specifically, we show that given a solution

¹Describing \tilde{f} as a multiplicative approximation of f is more convenient for analyzing multiplicative approximation guarantees. This is w.l.o.g as all our results apply to additive approximations as well.

to $\max_{S \in \mathcal{F}} f(S)$ one can produce a solution that is of arbitrarily close quality to $\max_{S \in \mathcal{F}} \tilde{f}(S)$ when \mathcal{F} is any uniform matroid, an intersection of matroids of rank $\Omega(\log n)$, and an intersection of matroids of any rank when the distribution \mathcal{D} has bounded support.

1.2 Technical Overview

The approximately submodular functions we consider approximate a submodular function f using samples from a distribution \mathcal{D} of the class of *generalized exponential tail* distributions defined as:

Definition. A noise distribution \mathcal{D} has a **generalized exponential tail** if there exists some x_0 such that for $x > x_0$ the probability density function $\rho(x) = e^{-g(x)}$, where $g(x) = \sum_i a_i x^{\alpha_i}$ for some (not necessarily integers) $\alpha_0 \geq \alpha_1 \geq \dots$, s.t. $\alpha_0 \geq 1$. If \mathcal{D} has bounded support we only require that either it has an atom at its supremum, or that ρ is continuous and non-zero at the supremum.

This class of distributions is carefully defined. On the one hand it is general enough to contain Gaussian and Exponential distributions, as well as any distribution with bounded support. On the other hand it has enough structure one can leverage. Note that optimization in this setting always requires that the support is independent of n and that n is sufficiently large². Throughout the paper we assume that \mathcal{D} has a generalized exponential tail and that n is sufficiently large.

Theorem. For any non-negative monotone submodular function there is a deterministic polynomial-time algorithm which optimizes the function under a cardinality constraint $k \geq 3$ and obtains an approximation ratio that is arbitrarily close to $1 - 1/e$ with probability $1 - o(1)$ using access to an approximate oracle. For $k \geq 2$ there is a randomized algorithm whose approximation ratio is arbitrarily close to $1 - 1/e$, in expectation over the randomization of the algorithm. For $k = 1$ the algorithm achieves a $1/2$ approximation in expectation, and no randomized algorithm can achieve an approximation better than $1/2 + o(1)$, in expectation.

The main part of the proof involves analysis of the following greedy algorithm. The algorithm iteratively chooses *bundles* of elements of size $\mathcal{O}(1/\epsilon)$. In each iteration, the algorithm first identifies a bundle \mathbf{x} whose addition to the current solution approximately maximizes the *approximate mean value* \tilde{F} . Informally, $\tilde{F}(\mathbf{x})$ is the average value of f evaluated on all bundles at Hamming distance one from \mathbf{x} . Then, the algorithm does not choose \mathbf{x} but rather the bundle at Hamming distance one from \mathbf{x} whose addition to the current solution maximizes the approximate submodular value \tilde{f} .

The major technical challenge is in analyzing the regime in which $k \in \Omega(1/\epsilon^2) \cap \mathcal{O}(\sqrt{\log n})$. At a high level, in this regime the analysis relies on showing that the marginal contribution of the bundle of elements selected in every iteration is approximately largest. Doing so requires proving subtle properties about extreme values of random variables drawn from the generalized exponential tail distribution, and the analysis fully leverages the properties of the distribution and the fact that $k \in \mathcal{O}(\sqrt{\log n})$. This is of independent interest to Extreme Value Theory (EVT) which tries to bound the maximum sample (or the top samples) from some distribution. If we would consider the constant function $f(S) = 1$ for $S \neq \emptyset$, and try to maximize an approximate version of f with respect to some distribution, this would be a classical EVT setting. One can view the bounds we develop as bounds on a generalization of EVT, where not all samples are created equal.

For general matroid constraints we apply the sampled-mean technique and obtain an approximation comparable to that of applying the greedy algorithm on a monotone submodular function.

Theorem. For any non-negative monotone submodular function there is a deterministic polynomial-time algorithm which optimizes the function under an intersection of P matroids constraint and obtains an approximation ratio arbitrarily close to $1/(1 + P)$ given an approximate oracle.

Paper organization. The main contribution of the paper is the definition of sampled mean approximation in Section 2 and the subsequent analysis of the algorithm for cardinality constraint in Section 3 and matroid constraints in Section 4. The techniques are novel, and the major technical crux of the paper is in analyzing the algorithm in Section 3. Optimization for small rank and lower bounds are in Appendix B.1. In Appendix G we further discuss related work, and in Appendix F we discuss extensions of the algorithms to related models.

²For example, if for every S the noise is s.t. $\xi_S = 2^{100}$ w.p. $1/2^{100}$ and 0 otherwise, but $n = 50$, it is likely that the oracle will always return 0, in which case we cannot do better than selecting an element at random.

2 The Sampled Mean Approximation

We begin by defining the sampled mean approximation of a function. This approach considers *bundles* \mathbf{x} of size $c \in \mathcal{O}(1/\epsilon)$. We can then run a variant of a greedy algorithm which adds a bundle of size c in every iteration. For a given bundle \mathbf{x} and set S , we define a *ball* to be all bundles obtained by swapping a single element from \mathbf{x} with another not in $S \cup \mathbf{x}$. We denote $\mathbf{x}_{ij} = (\mathbf{x} \setminus \{x_i\}) \cup \{x_j\}$.

Definition. For $S \subset N$ and bundle $\mathbf{x} \subset N$, the ball around \mathbf{x} is $\mathcal{B}_S(\mathbf{x}) := \{\mathbf{x}_{ij} : i \in \mathbf{x}, j \notin S \cup \mathbf{x}\}$.

We illustrate a ball in Figure 1. Notice that as long as $|S| \leq (1 - \delta)|N|$ for some fixed $\delta > 0$, we have that $|\mathcal{B}_S(\mathbf{x})| \in \Omega(n)$. This will allow us to derive weak (but sufficient) concentration bounds.

Definition. Let $f : 2^N \rightarrow \mathbb{R}$. For a set $S \subseteq N$ and bundle $\mathbf{x} \subseteq N$, the *mean value*, *noisy mean value*, and *mean marginal contribution* of \mathbf{x} given S are, respectively:

$$\begin{aligned} (1) \quad F(S \cup \mathbf{x}) &:= \mathbb{E}_{\mathbf{z} \sim \mathcal{B}_S(\mathbf{x})} [f(S \cup \mathbf{z})] \\ (2) \quad \tilde{F}(S \cup \mathbf{x}) &:= \mathbb{E}_{\mathbf{z} \sim \mathcal{B}_S(\mathbf{x})} [\tilde{f}(S \cup \mathbf{z})] \\ (3) \quad F_S(\mathbf{x}) &:= \mathbb{E}_{\mathbf{z} \sim \mathcal{B}_S(\mathbf{x})} [f_S(\mathbf{z})] \end{aligned}$$

The following lemma implies that under the right conditions, the bundle that maximizes the noisy mean value is a good approximation of the bundle whose (non-noisy) mean value is largest.

Lemma 2.1. Let $\mathbf{x} \in \operatorname{argmax}_{\mathbf{z}: |\mathbf{z}|=c} \tilde{F}(S \cup \mathbf{z})$ where $c = 2/\epsilon$. Then, w.p. $\geq 1 - \exp(-\Omega(n^{1/4}))$:

$$F_S(\mathbf{x}) \geq (1 - \epsilon) \max_{\mathbf{z}: |\mathbf{z}|=c} F_S(\mathbf{z}).$$

The above lemma gives us a *weak* concentration bound in the following sense. While it is generally not true that $\tilde{F}(S \cup \mathbf{z}) \approx F(S \cup \mathbf{z})$, we can upper bound $\tilde{F}(S \cup \mathbf{z})$ in a meaningful way and show that $\tilde{F}(S \cup \mathbf{x}^*) \approx F(S \cup \mathbf{x}^*)$ for $\mathbf{x}^* \in \operatorname{argmax}_{\mathbf{z}} f_S(\mathbf{z})$ by using submodularity. This allows us to show that the mean marginal contribution of \mathbf{x} that maximizes the *noisy* mean value is an ϵ -approximation to the maximal (non-noisy) mean marginal contribution. Details and proofs are in Appendix A.

In addition to approximating the mean marginal contribution given a noisy oracle, an important property of the sampled-mean approach is that it well-approximates its true marginal contribution.

Lemma 2.2. For any $\epsilon > 0$ and any set $S \subset N$, let \mathbf{x} be a bundle of size $1/\epsilon$, then:

$$F_S(\mathbf{x}) \geq (1 - \epsilon)f_S(\mathbf{x}),$$

The proof is in Appendix A and exploits a natural property of submodular functions: the removal of a random element from a sufficiently large set does not significantly affect its value, in expectation.

Let $\mathbf{x}^* \in \operatorname{argmax}_{\mathbf{b}: |\mathbf{b}|=c} f_S(\mathbf{b})$. Lemma 2.1 and Lemma 2.2 together imply the following corollary:

Corollary 2.3. For a fixed $\epsilon > 0$ let $c = 3/\epsilon$, and $\mathbf{x} \in \operatorname{argmax}_{\mathbf{b}: |\mathbf{b}|=c} \tilde{F}(S \cup \mathbf{b})$. Then, w.p. at least $1 - \exp(-\Omega(n^{1/4}))$ we have that: $F_S(\mathbf{x}) \geq (1 - \epsilon)f_S(\mathbf{x}^*)$.

At a first glance, it may seem as if running the greedy algorithm with F instead of f suffices. The problem, however, is that the *mean* marginal contribution F_S may be an unbounded overestimate of the *true* marginal contribution f_S . Consider for example an instance where $S = \emptyset$ and there is a bundle \mathbf{x} of size c s.t. for every $\emptyset \neq \mathbf{z} \subseteq \mathbf{x}$ we have $f(\mathbf{z}) = \delta$ for some arbitrarily small δ , while every other subset $T \subseteq N \setminus \mathbf{x}$ is complementary to \mathbf{x} and has some arbitrarily large value M . In this case, $\mathbf{x} = \operatorname{argmax}_{\mathbf{z}: |\mathbf{z}|=c} F(\mathbf{z})$ and $F(\mathbf{x}) = M + \delta$ while $f(\mathbf{x}) = \delta$.

3 The Sampled Mean Greedy for Cardinality Constraints

The SM-GREEDY begins with the empty set S and at every iteration considers all bundles (sets) of size $c \in \mathcal{O}(1/\epsilon)$ to add to S . At every iteration, the algorithm first identifies the bundle \mathbf{x} which maximizes the noisy mean value. After identifying \mathbf{x} , it then considers all possible bundles $\mathbf{z} \in \mathcal{B}_S(\mathbf{x})$ and takes the one whose noisy value is largest. We include a formal description below.

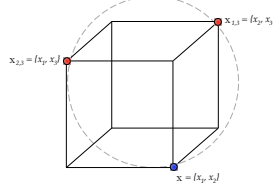


Figure 1: Illustration of a ball around $\mathbf{x} = \{x_1, x_2\}$ where $N = \{x_1, x_2, x_3\}$, and $S = \emptyset$. We think of \mathbf{x} as a point in $[0, 1]^3$ and $\mathcal{B}_S(\mathbf{x}) = \{\mathbf{x}_{12}, \mathbf{x}_{23}\} = \{\{x_2, x_3\}, \{x_1, x_3\}\} = \{(0, 1, 1), (1, 0, 1)\}$.

Algorithm 1 SM-GREEDY

Input: budget k , precision $\epsilon > 0$, $c \leftarrow \frac{56}{\epsilon}$

- 1: $S \leftarrow \emptyset$
- 2: **while** $|S| < c \cdot \lfloor \frac{k}{c} \rfloor$ **do**
- 3: $\mathbf{x} \leftarrow \operatorname{argmax}_{\mathbf{b}: |\mathbf{b}|=c} \tilde{F}(S \cup \mathbf{b})$
- 4: $\hat{\mathbf{x}} \leftarrow \operatorname{argmax}_{\mathbf{z} \in \mathcal{B}_S(\mathbf{x})} \tilde{f}(S \cup \mathbf{z})$
- 5: $S \leftarrow S \cup \hat{\mathbf{x}}$
- 6: **end while**
- 7: **return** S

A key step in analyzing greedy algorithms like the one above is showing that in every iteration the marginal contribution of the element selected by the algorithm is arbitrarily close to maximal. This can then be used in a standard inductive argument to show that the algorithm obtains an approximation arbitrarily close to $1 - 1/e$. The main crux of the analysis is showing that this property indeed holds in SM-GREEDY w.h.p. when $k \in \Omega(1/\epsilon^2) \cap \mathcal{O}(\sqrt{\log n})$. For $k > \sqrt{\log n}$ the analysis become substantially simpler since it suffices to argue that the algorithm chooses an element whose marginal contribution is approximately optimal in expectation (details are in the proof of Theorem 3.5).

3.1 Analysis for $k \in \Omega(1/\epsilon^2) \cap \mathcal{O}(\sqrt{\log n})$

Throughout the rest of this section we will analyze a single iteration of SM-GREEDY in which a set $S \subset N$ was selected in previous iterations and the algorithm adds a bundle $\hat{\mathbf{x}}$ of size c . Specifically, $\hat{\mathbf{x}} \in \operatorname{argmax}_{\mathbf{z} \in \mathcal{B}_S(\mathbf{x})} \tilde{f}(S \cup \mathbf{z})$ where $\mathbf{x} \in \operatorname{argmax}_{\mathbf{b}: |\mathbf{b}|=c} \tilde{F}(S \cup \mathbf{b})$.

As discussed above, for $\mathbf{x}^* \in \operatorname{argmax}_{\mathbf{b}: |\mathbf{b}|=c} f_S(\mathbf{b})$ we want to show that when $c \in \mathcal{O}(1/\epsilon)$:

$$f_S(\hat{\mathbf{x}}) \geq (1 - \epsilon)f_S(\mathbf{x}^*).$$

To do so, we will define two kinds of bundles in $\mathcal{B}_S(\mathbf{x})$, called **good** and **bad**. A good bundle is a bundle \mathbf{z} for which $f_S(\mathbf{z}) \geq (1 - \frac{2}{3}\epsilon)f_S(\mathbf{x}^*)$ and a bad bundle \mathbf{z} is one s.t. $f_S(\mathbf{z}) \leq (1 - \epsilon)f_S(\mathbf{x}^*)$. Our goal is to prove that the bundle $\hat{\mathbf{x}}$ added by the algorithm is w.h.p. not bad. Since according to the definition of good and bad the true marginal contribution of good bundles has a fixed advantage over the true marginal contribution of bad bundles, and $\hat{\mathbf{x}}$ is the bundle with largest *noisy* value, essentially what we need to show is that the largest noise multiplier of a good bundle is sufficiently close to the largest noise multiplier of a bad bundle, with sufficiently high probability.

As a first step we quantify the fraction of good bundles in a ball, which will then allow us to bound the values of the noise multipliers of good and bad bundles. The following claim implies at least half of the bundles in $\mathcal{B}_S(\mathbf{x})$ are good and at most half are bad (the proof is in Appendix B).

Claim 3.1. *Suppose $F_S(\mathbf{x}) \geq (1 - \frac{\epsilon}{3})f_S(\mathbf{x}^*)$. Then at least half of the bundles in $\mathcal{B}_S(\mathbf{x})$ are good.*

We next define two thresholds ξ_{\inf} and ξ_{\sup} . The threshold ξ_{\inf} is used as a lower bound on the largest value obtained when sampling *at least* $\frac{|\mathcal{B}_S(\mathbf{x})|}{2}$ random variables from the noise distribution. Since at least half of the bundles in the ball are good, ξ_{\inf} is a lower bound on the largest noise multiplier of a good bundle. The threshold ξ_{\sup} is used as an upper bound on the largest value obtained when sampling *at most* $\frac{|\mathcal{B}_S(\mathbf{x})|}{2}$ random variables from the noise distribution. Since at most half of the bundles in the ball are bad, ξ_{\sup} upper bounds the value of a noise multiplier of a bad bundle. Throughout the rest of this section \mathcal{D} will denote a generalized exponential tail distribution.

Definition. Let $m = \frac{|\mathcal{B}_S(\mathbf{x})|}{2}$. For probability density function $\rho(x)$ of \mathcal{D} we define:

- ξ_{sup} is the value for which: $\int_{\xi_{\text{sup}}}^{\infty} \rho(x) dx = \frac{2}{m \log n}$;
- ξ_{inf} is the value for which: $\int_{\xi_{\text{inf}}}^{\infty} \rho(x) dx = \frac{2 \log n}{m}$.

Claim 3.2. Let $m = \frac{|\mathcal{B}_S(\mathbf{x})|}{2}$, ξ_1, \dots, ξ_m be i.i.d samples of \mathcal{D} and $\xi^* = \max\{\xi_1, \dots, \xi_m\}$. Then:

$$\Pr \left[\xi_{\text{inf}} \leq \xi^* \leq \xi_{\text{sup}} \right] \geq 1 - \frac{3}{\log n}$$

The proof is in Appendix B. Since at least half of the bundles in the ball are good and at most half are bad, the above claim implies that with probability $1 - 3/\log n$ the maximal value of noise multipliers of good and bad bundles fall within the range $[\xi_{\text{inf}}, \xi_{\text{sup}}]$. Since this holds in every iteration, when $k \in \mathcal{O}(\sqrt{\log n})$ by a union bound we get that this holds in all iterations w.p. $1 - o(1)$.

At this point, our lower bound on the largest noisy value of a good bundle is:

$$\max_{\mathbf{z} \in \text{good}} \tilde{f}(S \cup \mathbf{z}) \geq \xi_{\text{inf}} \times \left(f(S) + \left(1 - \frac{2}{3}\epsilon\right) f_S(\mathbf{x}^*) \right)$$

and the upper bound on the noisy value of any bad bundle is:

$$\max_{\mathbf{z} \in \text{bad}} \tilde{f}(S \cup \mathbf{z}) \leq \xi_{\text{sup}} \times (f(S) + (1 - \epsilon) f_S(\mathbf{x}^*)).$$

Let $\hat{\mathbf{x}} = \arg\max_{\mathbf{z} \in \mathcal{B}_S(\mathbf{x})} \tilde{f}(S \cup \mathbf{z})$. We show that given the right bound on ξ_{inf} against ξ_{sup} , then $\hat{\mathbf{x}}$ is not a bad bundle (it does not have to be a good bundle). Lemma 3.3 below gives us such a bound.

Lemma 3.3. For any generalized exponential tail distribution, fixed $\gamma > 0$ and sufficiently large n :

$$\xi_{\text{inf}} \geq \left(1 - \frac{\gamma}{\sqrt{\log n}}\right) \xi_{\text{sup}}$$

The proof is quite technical and the fully leverages the fact that $k \in \mathcal{O}(\sqrt{\log n})$ and the properties of generalized exponential tail distributions. The main challenge is that the tail of the distribution is not necessarily monotone (see Appendix B for further discussion). We defer the proof to Appendix B.

Proving the Main Lemma. We can now prove our main lemma which shows that for any $k \in \omega(1/\epsilon) \cap \mathcal{O}(\sqrt{\log n})$ taking a bundle according to the sample mean approach is guaranteed to be close to the optimal bundle. We state it for a cardinality constraint, but this fact holds more generally for any matroid of rank k . We include a short proof sketch and defer the full proof to Appendix B.

Lemma 3.4. Let $\epsilon > 0$, and assume that $k \in \omega(1/\epsilon) \cap \mathcal{O}(\sqrt{\log n})$ and that $f_S(\mathbf{x}^*) \in \Omega\left(\frac{f(S)}{k}\right)$. Then, in every iteration of SM-GREEDY we have that with probability at least $1 - \frac{4}{\log n}$:

$$f_S(\hat{\mathbf{x}}) \geq (1 - \epsilon) f_S(\mathbf{x}^*).$$

Proof Sketch. From Corollary 2.3 we know that in every iteration with overwhelming high probability $F_S(\mathbf{x}) \geq (1 - \delta) f(\mathbf{x}^*)$ for $\delta = \epsilon/3$. Since Lemma 3.3 applies for any fixed $\gamma > 0$ we know that for sufficiently large n we can lower bound the value of the maximal good bundle in the ball:

$$\max_{\mathbf{z} \in \text{good}} \tilde{f}(S \cup \mathbf{z}) \geq \left(1 - \frac{\delta^2}{3\sqrt{\log n}}\right) \xi_{\text{sup}} \times [f(S) + (1 - 2\delta) f_S(\mathbf{x}^*)]$$

Let \mathbf{b} be the bad bundle with the maximal noisy value for f_S . To bound this:

$$\tilde{f}(S \cup \mathbf{b}) = \max_{\mathbf{z} \in \text{bad}} \xi_{S \cup \mathbf{z}} f(S \cup \mathbf{z}) \leq \xi_{\text{sup}} \times [f(S) + (1 - 3\delta) f_S(\mathbf{x}^*)]$$

The difference between the two bounds is positive, implying that a bad bundle is not selected. \square

Theorem 3.5. For any fixed $\epsilon > 0$ and $k \geq 4/\epsilon^2$ SM-GREEDY returns a set S s.t. w.p. $1 - o(1)$:

$$f(S) \geq (1 - 1/e - \epsilon) \text{OPT}.$$

Proof. Let $\kappa = \lfloor \frac{k}{c} \rfloor$ and use O to denote the set of κ bundles of size c whose total value is largest. To simplify notation we will treat sets of bundles as sets of elements. We will show that $f(S) \geq (1 - 1/e - \delta)f(O)$ where $\delta = \epsilon/2$. Notice that this implies $f(S) \geq (1 - 1/e - \epsilon)$ since $\kappa > \frac{k}{c} - c$ and by submodularity $f(O) \geq (1 - \delta)\text{OPT}$ when $k \geq \frac{1}{\delta^2} = \frac{4}{\epsilon^2}$.

We introduce some notation: we use $\hat{\mathbf{x}}_i$ to denote the bundle selected at iteration $i \in [\kappa]$, $S_i = \cup_{j \leq i} \hat{\mathbf{x}}_j$, $\mathbf{x}_i \in \arg\max_{\mathbf{z}} \tilde{F}(S_{i-1} \cup \mathbf{z})$, $\mathbf{x}_i^* \in \arg\max_{\mathbf{z}} f_{S_{i-1}}(\mathbf{z})$, and O is the set of κ bundles with maximal value. Define $\Delta_i = f(O) - f(S_{i-1})$, and $S_0 = \emptyset$. From submodularity and monotonicity:

$$f_{S_{i-1}}(\mathbf{x}_i^*) \geq \frac{1}{\kappa} \Delta_i$$

Consider now a set of bundles $\{\mathbf{z}_1, \dots, \mathbf{z}_\kappa\}$ where for every $i \in [\kappa]$ we have that \mathbf{z}_i is drawn u.a.r. from $\mathcal{B}_{S_{i-1}}(\mathbf{x}_i)$. For each such bundle we can assign a random variable ζ_i for which $f_{S_{i-1}}(\mathbf{z}_i) = \zeta_i f_{S_{i-1}}(\mathbf{x}_i^*)$. Since in every iteration $i \in [\kappa]$ we choose the set whose value is maximal in $\mathcal{B}_{S_{i-1}}(\mathbf{x}_i)$, by stochastic dominance we know that $f_{S_{i-1}}(\hat{\mathbf{x}}_i) \geq f(\mathbf{z}_i)$ and therefore:

$$f(S_i) - f(S_{i-1}) = f_{S_{i-1}}(\hat{\mathbf{x}}_i) \geq \zeta_i f_{S_{i-1}}(\mathbf{x}_i^*) \geq \frac{\zeta_i}{\kappa} \Delta_i$$

We will now show by induction that for all $i \in [\kappa]$ we have that $\Delta_i \leq \prod_{j=1}^i \left(1 - \frac{\zeta_j}{\kappa}\right) f(O)$. This is clearly the case for $i = 1$ when $S_0 = \emptyset$ and in general applying the inductive hypothesis we get:

$$\Delta_{i+1} = f(O) - f(S_i) = \Delta_i - (f(S_i) - f(S_{i-1})) \leq \Delta_i \left(1 - \frac{\zeta_{i+1}}{\kappa}\right) \leq \prod_{j=1}^{i+1} \left(1 - \frac{\zeta_j}{\kappa}\right) f(O)$$

We therefore have that:

$$\Delta_\kappa \leq \prod_{j=1}^{\kappa} \left(1 - \frac{\zeta_j}{\kappa}\right) f(O) \leq e^{-\frac{1}{\kappa} \sum_{j=1}^{\kappa} \zeta_j} f(O)$$

Observe that the solution of the algorithm S respects $f(S) = f(S_\kappa) = f(O) - \Delta_\kappa$, thus:

$$f(S) \geq \left(1 - e^{-\frac{1}{\kappa} \sum_{j=1}^{\kappa} \zeta_j}\right) f(O) \quad (1)$$

From Lemma 3.4, when $k \leq \sqrt{\log n}$ for every $i \in [\kappa - 1]$ we have that w.p. $1 - 4/\log(n)$:³

$$f_{S_{i-1}}(\hat{\mathbf{x}}_i) \geq (1 - \delta) f_{S_{i-1}}(\mathbf{x}_i^*)$$

Therefore by a union bound, with probability $1 - o(1)$, we have that $\zeta_i \geq (1 - \delta)$ for all $i \in [\kappa]$. In particular, $\frac{1}{\kappa} \sum_{j=1}^{\kappa} \zeta_j \geq (1 - \delta)$. Otherwise, when $k > \sqrt{\log n}$ from Lemma 2.1:

$$\mathbb{E}_{\mathbf{z} \sim \mathcal{B}(\mathbf{x}_i)}[f_{S_{i-1}}(\mathbf{z})] = F_{S_{i-1}}(\mathbf{x}_i) \geq \left(1 - \frac{\delta}{2}\right) f_{S_{i-1}}(\mathbf{x}_i^*)$$

Thus, $\mathbb{E}[\zeta_{i+1}] \geq (1 - \delta/2)$, and by Chernoff when $\kappa > \sqrt{\log n}$ we get $\frac{1}{\kappa} \sum_{i=1}^{\kappa} \zeta_i \geq (1 - \delta)$ w.p. at least $1 - \exp(-\delta^2 \kappa/8)$. Therefore, in both cases, when $k \leq \sqrt{\log n}$ and when $k \geq \sqrt{\log n}$ we have that $\frac{1}{\kappa} \sum_{j=1}^{\kappa} \zeta_j \geq (1 - \delta)$ w.p. $1 - o(1)$. With 1 this implies $f(S) \geq (1 - 1/e - \delta)f(O)$. \square

Constant k and information theoretic lower bounds. For any constant, a single iteration of a minor modification of SM-GREEDY suffices. In Appendix B.1 we show an approximation arbitrarily close to $1 - 1/k$ w.h.p. and $1 - 1/(k+1)$ in expectation. For $k = 1$ this is arbitrarily close to $1/2$. In Appendix D we show nearly matching lower bounds, and in particular that no randomized algorithm can obtain an approximation ratio better than $1/2 + o(1)$ when $k = 1$, and that it is impossible to obtain an approximation better than $(2k - 1)/2k + O(1/\sqrt{n})$ for the optimal set of size k .

Theorem 3.6. *For any non-negative monotone submodular function there is a deterministic polynomial-time algorithm which optimizes the function under a cardinality constraint $k \geq 3$ and obtains an approximation ratio that is arbitrarily close to $1 - 1/e$ with probability $1 - o(1)$ using access to an approximate oracle. For $k \geq 2$ there is a randomized algorithm whose approximation ratio is arbitrarily close to $1 - 1/e$, in expectation over the randomization of the algorithm. For $k = 1$ the algorithm achieves a $1/2$ approximation in expectation, and no randomized algorithm can achieve an approximation better than $1/2 + o(1)$, in expectation.*

³W.l.o.g we assume that in every iteration $f_{S_{i-1}}(\mathbf{x}_i^*) \in \Omega(\frac{f(S_{i-1})}{\kappa})$ to apply Lemma 3.4. Since $\kappa \in \Omega(1/\epsilon)$, $k \in \Omega(1/\epsilon^2)$, and $f(S) \leq \text{OPT}$, ignoring iterations where this does not hold costs $\gamma \epsilon \text{OPT}$ for a small fixed γ .

4 Approximation Algorithm for Matroids

For intersection of matroids \mathcal{F} the algorithm from the Section 3 is generalized as described below.

Algorithm 2 SM-MATROID-GREEDY

Input: intersection of matroids \mathcal{F} , precision $\epsilon > 0$, $c \leftarrow \frac{56}{\epsilon}$

```

1:  $S \leftarrow \emptyset, X \leftarrow N$ 
2: while  $X \neq S$  do
3:    $X \leftarrow X \setminus \{\mathbf{x} : S \cup \mathbf{x} \notin \mathcal{F}\}$ 
4:    $\mathbf{x} \leftarrow \operatorname{argmax}_{\mathbf{b}: |\mathbf{b}|=c} \tilde{F}(S \cup \mathbf{b})$ 
5:    $\hat{\mathbf{x}} \leftarrow \operatorname{argmax}_{\mathbf{z} \in \mathcal{B}(\mathbf{x})} \tilde{f}(S \cup \mathbf{z})$ 
6:    $S \leftarrow S \cup \hat{\mathbf{x}}$ 
7: end while
8: return  $S$ 

```

The analysis of the algorithm uses the lemma below, which is a generalization of the classic result of [NWF78a]. The proof can be found in the Appendix.

Lemma 4.1. *Let O be the optimal solution, $k = |O|$, and for every iteration i of SM-MATROID-GREEDY let S_i be the set of elements selected and $\mathbf{x}_i^* \in \operatorname{argmax}_{|\mathbf{z}|=c} f_{S_{i-1}}(\mathbf{z})$. Then:*

$$f(O) \leq (P+1) \sum_{i=1}^{\frac{k}{c}} f_{S_i}(\mathbf{x}_i^*)$$

Theorem 4.2. *Let \mathcal{F} denote the intersection of $P \geq 1$ matroids on the ground set N , and $f : 2^N \rightarrow \mathbb{R}$ be a non-negative monotone submodular function. Then with probability $1 - o(1)$ the SM-MATROID-GREEDY algorithm returns a set $S \in \mathcal{F}$ s.t.:*

$$f(S) \geq \frac{1-\epsilon}{P+1} OPT$$

Proof Sketch. If the rank of the matroid is $\mathcal{O}(1/\epsilon^2)$ we can simply apply the case of small k as in the previous section. Otherwise, assume the rank is at least $\Omega(1/\epsilon^2)$. Let $\kappa = \frac{k}{c}$, $S_i = \{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_i\}$ be current solutions of bundles of size c at iteration $i \in [\kappa]$ of the algorithm, and let \mathbf{x}_i^* be the optimal bundle at iteration i , i.e. $\mathbf{x}_i^* = \operatorname{argmax}_{\mathbf{b}: |\mathbf{b}| \leq c} f_{S_{i-1}}(\mathbf{b})$. In every iteration $i \in [\kappa]$, similar to the proof of Theorem 3.5, since we choose the set whose value is maximal in $\mathcal{B}_{S_{i-1}}(\mathbf{x}_i)$ we have:

$$f_{S_{i-1}}(\hat{\mathbf{x}}_i) \geq \zeta_i f_{S_{i-1}}(\mathbf{x}_i^*)$$

where ζ_i is a random variable with mean $(1 - \frac{\delta}{2})$. Therefore:

$$f(S) = \sum_{i=1}^{\kappa} f_{S_{i-1}}(\hat{\mathbf{x}}_i) \geq \sum_{i=1}^{\kappa} \zeta_i f_{S_{i-1}}(\mathbf{x}_i^*)$$

From Lemma 3.4, when $k \leq \sqrt{\log n}$ for every $i \in [\kappa - 1]$ we have that w.p. $1 - 4/\log(n)$:

$$f_{S_{i-1}}(\hat{\mathbf{x}}_i) \geq (1 - \delta) f_{S_{i-1}}(\mathbf{x}_i^*)$$

Therefore by a union bound, with probability $1 - o(1)$, we have that $\zeta_i \geq (1 - \delta)$ for all $i \in [\kappa]$. Otherwise, when $k > \sqrt{\log n}$ we apply a Chernoff bound. We get that with probability $1 - o(1)$:

$$f(S) \geq \sum_{i=1}^{\kappa} \zeta_i f_{S_{i-1}}(\mathbf{x}_i^*) \geq (1 - \delta) \sum_{i=1}^{\kappa} f_{S_{i-1}}(\mathbf{x}_i^*)$$

From Lemma 4.1 this implies the result. \square

Acknowledgements. This research was supported by NSF grant CAREER CCF 1452961, NSF CCF 1301976, BSF grant 2014389, NSF USICCS proposal 1540428, a Google Research award, and a Facebook research award.

References

- [Ang88] Dana Angluin. Queries and concept learning. *Machine Learning*, 2(4):319–342, 1988.
- [AS04] Alexander A. Ageev and Maxim Sviridenko. Pipage rounding: A new method of constructing algorithms with proven performance guarantee. *J. Comb. Optim.*, 8(3), 2004.
- [BDF⁺10] D. Buchfuhrer, S. Dughmi, H. Fu, R. Kleinberg, E. Mossel, C. H. Papadimitriou, M. Schapira, Y. Singer, and C. Umans. Inapproximability for VCG-based combinatorial auctions. In *SIAM-ACM Symposium on Discrete Algorithms (SODA)*, pages 518–536, 2010.
- [BF02] Nader H. Bshouty and Vitaly Feldman. On using extended statistical queries to avoid membership queries. *Journal of Machine Learning Research*, 2:359–395, 2002.
- [BFNS12] Niv Buchbinder, Moran Feldman, Joseph Naor, and Roy Schwartz. A tight linear time $(1/2)$ -approximation for unconstrained submodular maximization. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 649–658, 2012.
- [BFNS14] Niv Buchbinder, Moran Feldman, Joseph Naor, and Roy Schwartz. Submodular maximization with cardinality constraints. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1433–1452, 2014.
- [BH11] Maria-Florina Balcan and Nicholas J. A. Harvey. Learning submodular functions. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 793–802, 2011.
- [BMKK14] Ashwinkumar Badanidiyuru, Baharan Mirzasoleiman, Amin Karbasi, and Andreas Krause. Streaming submodular maximization: massive data summarization on the fly. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 671–680, 2014.
- [BSS10] D. Buchfuhrer, M. Schapira, and Y. Singer. Computation and incentives in combinatorial public projects. In *EC*, pages 33–42, 2010.
- [CCPV07] Grigore Calinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a submodular set function subject to a matroid constraint. In *Integer programming and combinatorial optimization*, pages 182–196. Springer, 2007.
- [CE11] Chandra Chekuri and Alina Ene. Approximation algorithms for submodular multiway partition. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 807–816, 2011.
- [CE16] C.P. Chambers and F. Echenique. *Revealed Preference Theory*. Econometric Society Monographs. Cambridge University Press, 2016.
- [CHHK18] Lin Chen, Christopher Harshaw, Hamed Hassani, and Amin Karbasi. Projection-free online optimization with stochastic gradient: From convexity to submodularity. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 813–822, 2018.
- [CJV15] Chandra Chekuri, T. S. Jayram, and Jan Vondrák. On multiplicative weight updates for concave and submodular function maximization. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 201–210, 2015.
- [DFK11] Shahar Dobzinski, Hu Fu, and Robert D. Kleinberg. Optimal auctions with correlated bidders are easy. In *STOC*, pages 129–138, 2011.

- [DK14] J. Djolonga and A. Krause. From MAP to marginals: Variational inference in bayesian submodular models. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.
- [DLN08] Shahar Dobzinski, Ron Lavi, and Noam Nisan. Multi-unit auctions with budget limits. In *FOCS*, 2008.
- [DNS05] Shahar Dobzinski, Noam Nisan, and Michael Schapira. Approximation algorithms for combinatorial auctions with complement-free bidders. In *STOC*, pages 610–618, 2005.
- [DRY11] Shaddin Dughmi, Tim Roughgarden, and Qiqi Yan. From convex optimization to randomized mechanisms: toward optimal combinatorial auctions. In *STOC*, pages 149–158, 2011.
- [DS06] Shahar Dobzinski and Michael Schapira. An improved approximation algorithm for combinatorial auctions with submodular bidders. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1064–1073. Society for Industrial and Applied Mathematics, 2006.
- [DV12] Shahar Dobzinski and Jan Vondrák. The computational complexity of truthfulness in combinatorial auctions. In *EC*, pages 405–422, 2012.
- [Fei98] Uriel Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.
- [Fel09] Vitaly Feldman. On the power of membership queries in agnostic learning. *Journal of Machine Learning Research*, 10:163–182, 2009.
- [FFI⁺15] Uriel Feige, Michal Feldman, Nicole Immorlica, Rani Izsak, Brendan Lucier, and Vasilis Syrgkanis. A unifying hierarchy of valuations with complements and substitutes. In *AAAI*, pages 872–878, 2015.
- [FMV11] Uriel Feige, Vahab S Mirrokni, and Jan Vondrak. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.
- [FNW78] Marshall L Fisher, George L Nemhauser, and Laurence A Wolsey. *An analysis of approximations for maximizing submodular set functions—II*. Springer, 1978.
- [FV06] Uriel Feige and Jan Vondrak. Approximation algorithms for allocation problems: Improving the factor of $1-1/e$. In *Foundations of Computer Science, 2006. FOCS’06. 47th Annual IEEE Symposium on*, pages 667–676. IEEE, 2006.
- [GFK10] D. Golovin, M. Faulkner, and A. Krause. Online distributed sensor selection. In *IPSN*, 2010.
- [GK10] R. Gomes and A. Krause. Budgeted nonparametric learning from data streams. In *Int. Conference on Machine Learning (ICML)*, 2010.
- [GK11] D. Golovin and A. Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *JAIR*, 42:427–486, 2011.
- [GKS90] Sally A. Goldman, Michael J. Kearns, and Robert E. Schapire. Exact identification of circuits using fixed points of amplification functions (abstract). In *Proceedings of the Third Annual Workshop on Computational Learning Theory, COLT 1990, University of Rochester, Rochester, NY, USA, August 6-8, 1990.*, page 388, 1990.
- [HS] Thibaut Horel and Yaron Singer. Maximizing approximately submodular functions. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*.
- [HS17] Avinatan Hassidim and Yaron Singer. Submodular optimization under noise. 2017. COLT.

- [HSK17] S. Hamed Hassani, Mahdi Soltanolkotabi, and Amin Karbasi. Gradient methods for submodular maximization. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5843–5853, 2017.
- [Jac94] Jeffrey C. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. In *35th Annual Symposium on Foundations of Computer Science, Santa Fe, New Mexico, USA, 20-22 November 1994*, pages 42–53, 1994.
- [JB11a] S. Jegelka and J. Bilmes. Approximation bounds for inference using cooperative cuts. In *Int. Conference on Machine Learning (ICML)*, 2011.
- [JB11b] S. Jegelka and J. Bilmes. Submodularity beyond submodular energies: Coupling edges in graph cuts. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1897–1904, 2011.
- [KG07] A. Krause and C. Guestrin. Nonmyopic active learning of gaussian processes. an exploration–exploitation approach. In *Int. Conference on Machine Learning (ICML)*, 2007.
- [KKT03] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2003.
- [KLMM05] Subhash Khot, Richard J Lipton, Evangelos Markakis, and Aranyak Mehta. Inapproximability results for combinatorial auctions with submodular utility functions. In *Internet and Network Economics*, pages 92–101. Springer, 2005.
- [KMVV13] Ravi Kumar, Benjamin Moseley, Sergei Vassilvitskii, and Andrea Vattani. Fast greedy algorithms in mapreduce and streaming. In *SPAA*, 2013.
- [KOJ13] P. Kohli, A. Osokin, and S. Jegelka. A principled deep random field for image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [LB11a] H. Lin and J. Bilmes. A class of submodular functions for document summarization. In *ACL/HLT*, 2011.
- [LB11b] H. Lin and J. Bilmes. Optimal selection of limited vocabulary speech corpora. In *Proc. Interspeech*, 2011.
- [LCSZ17] Qiang Li, Wei Chen, Xiaoming Sun, and Jialin Zhang. Influence maximization with ϵ -almost submodular threshold functions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 3804–3814, 2017.
- [LKG⁺07] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2007.
- [LMNS09] Jon Lee, Vahab S. Mirrokni, Viswanath Nagarajan, and Maxim Sviridenko. Non-monotone submodular maximization under matroid and knapsack constraints. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 323–332, 2009.
- [LSST13] Brendan Lucier, Yaron Singer, Vasilis Syrgkanis, and Éva Tardos. Equilibrium in combinatorial public projects. In *WINE*, pages 347–360, 2013.
- [MHK18] Aryan Mokhtari, Hamed Hassani, and Amin Karbasi. Conditional gradient method for stochastic submodular maximization: Closing the gap. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pages 1886–1895, 2018.

- [MSV08] Vahab S. Mirrokni, Michael Schapira, and Jan Vondrák. Tight information-theoretic lower bounds for welfare maximization in combinatorial auctions. In *Proceedings 9th ACM Conference on Electronic Commerce (EC-2008)*, Chicago, IL, USA, June 8-12, 2008, pages 70–77, 2008.
- [NW78] George L Nemhauser and Leonard A Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of operations research*, 3(3):177–188, 1978.
- [NWF78a] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions ii. *Math. Programming Study* 8, 1978.
- [NWF78b] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978.
- [PP11] Christos H. Papadimitriou and George Pierrakos. On optimal single-item auctions. In *STOC*, pages 119–128, 2011.
- [PSS08] Christos H. Papadimitriou, Michael Schapira, and Yaron Singer. On the hardness of being truthful. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 250–259, 2008.
- [QSY⁺17] Chao Qian, Jing-Cheng Shi, Yang Yu, Ke Tang, and Zhi-Hua Zhou. Subset selection under noise. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 3563–3573, 2017.
- [RLK11] M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. *ACM TKDD*, 5(4), 2011.
- [SGK09] M. Streeter, D. Golovin, and A. Krause. Online learning of assignments. In *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [SS95] Eli Shamir and Clara Schwartzman. Learning by extended statistical queries and its relation to PAC learning. In *Computational Learning Theory: Eurocolt '95*, pages 357–366. Springer-Verlag, 1995.
- [SS08] Michael Schapira and Yaron Singer. Inapproximability of combinatorial public projects. In *WINE*, pages 351–361, 2008.
- [VCZ11] Jan Vondrák, Chandra Chekuri, and Rico Zenklusen. Submodular function maximization via the multilinear relaxation and contention resolution schemes. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing, STOC '11*, pages 783–792, New York, NY, USA, 2011. ACM.
- [Von08] Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *STOC*, pages 67–74, 2008.