A Unified Framework of Robust Submodular Optimization

Rishabh Iyer
AI & Research Division
Microsoft Corporation
Redmond, WA 98052
rishi@microsoft.com

Abstract

In this paper, we shall study a unified framework of robust submodular optimization. We study this problem both from a minimization and maximization perspective (previous work has only focused on variants of robust submodular maximization). We do this under a broad range of combinatorial constraints including cardinality, knapsack, matroid as well as graph based constraints such as cuts, paths, matchings and trees. Furthermore, we also study robust submodular minimization and maximization under multiple submodular upper and lower bound constraints. We show that all these problems are motivated by important machine learning applications including robust data subset selection, robust co-operative cuts and robust co-operative matchings. In each case, we provide scalable approximation algorithms and also study hardness bounds. Finally, we empirically demonstrate the utility of our algorithms on real world applications.

1 Introduction

Submodular functions provide a rich class of expressible models for a variety of machine learning problems. Submodular functions occur naturally in two flavors. In minimization problems, they model notions of cooperation, attractive potentials, and economies of scale, while in maximization problems, they model aspects of coverage, diversity, and information. A set function \( f: 2^V \to \mathbb{R} \) over a finite set \( V = \{1, 2, \ldots, n\} \) is submodular \([11]\) if for all subsets \( S, T \subseteq V \), it holds that \( f(S) + f(T) \geq f(S \cup T) + f(S \cap T) \). Given a set \( S \subseteq V \), we define the gain of an element \( j \notin S \) in the context \( S \) as \( f(j \mid S) = f(S \cup j) - f(S) \). A perhaps more intuitive characterization of submodularity is as follows: a function \( f \) is submodular if it satisfies diminishing marginal returns, namely \( f(j \mid S) \geq f(j \mid T) \) for all \( S \subseteq T, j \notin T \), and is monotone if \( f(j \mid S) \geq 0 \) for all \( j \notin S, S \subseteq V \).

Two central optimization problems involving submodular functions are submodular minimization \([11, 23]\) and submodular maximization \([45, 5]\). Moreover, it is often natural to want to optimize these functions subject to combinatorial constraints \([45, 23, 25, 14]\). Moreover, several combinatorial optimization problems involve minimizing one submodular function, while maximizing another function. Furthermore, in applications, these occur as maximizing or minimizing a submodular function subject to upper bound or lower bound constraints. These problems are called Submodular Cost Submodular Cover (SCSC) and Submodular Cost Submodular Knapsack (SCSK) introduced in in \([20]\).

In this paper, we shall study a framework of robust submodular optimization. Often times in applications we want to optimize several objectives (or criteria) together. There are two natural formulations of this. One is the average case, where we can optimize the (weighted) sum of the submodular functions. Examples of this have been studied in data summarization applications \([40, 51, 16]\). The other is robust or worst case, where we want to maximize (or minimize) the minimum (equivalently maximum) among the functions. Examples of this have been proposed for sensor placement and observation selection \([35]\). Robust or worst case optimization is becoming increasingly important since solutions achieved by minimization and maximization can be unstable to perturbations in data. Often times submodular functions in applications are instantiated from various properties of the data (features, similarity functions etc.) and obtaining results which are robust to perturbations and variations in this data is critical.
Given monotone submodular functions $f_1, f_2, \cdots, f_l$ to minimize and $g_1, g_2, \cdots, g_k$ to maximize, consider the following problems:

Problem 1: $\min_{X \in C} \max_{i=1:l} f_i(X)$, \hspace{1cm} Problem 2: $\max_{X \in C} \min_{i=1:k} g_i(X)$

$C$ stands for combinatorial constraints, which include cardinality, matroid, spanning trees, cuts, s-t paths etc. We shall call these problems ROBUST-SUBMIN and ROBUST-SUBMAX. Note that when $k = 1, l = 1$, we get back constrained submodular minimization and constrained submodular maximization. We will also study special cases of ROBUST-SUBMIN and ROBUST-SUBMAX when the constraints are defined via another submodular function. We study two problems: a) minimize the functions $f_i$’s while having lower bound constraints on the $g_i$’s (ROBUST-SCSK), and b) maximize the functions $g_i$’s subject to upper bound constraints on $f_i$’s (ROBUST-SCSK).

Problem 3: $\min_{X \subseteq V} \max_{i=1:l} f_i(X) \mid g_i(X) \geq c_i, i = 1, \cdots k$

Problem 4: $\max_{X \subseteq V} \min_{i=1:k} g_i(X) \mid f_i(X) \leq b_i, i = 1, \cdots l$

Problems 3 and 4 attempt to simultaneously minimize the functions $f_i$ while maximizing $g_i$. Finally, a natural extension of these problems is to have a joint average/worst case objective where we optimize $(1 - \lambda) \min_{i=1:k} g_i(X) + \lambda/k \sum_{i=1}^k f_i(X)$ and $(1 - \lambda) \max_{i=1:l} f_i(X) + \lambda/l \sum_{i=1}^l f_i(X)$ in Problems 1 - 4. We shall call these problems the MIXED versions (MIXED-SUBMAX, MIXED-SUBMIN etc.) However we point out that the MIXED case for Problems 1-4 is a special case of ROBUST optimization. It’s easy to see that we can convert this into a ROBUST formulation by defining $f'_i(X) = (1 - \lambda) f_i(X) + \lambda/l \sum_{i=1}^l f_i(X)$ and $g'_i(X) = (1 - \lambda) g_i(X) + \lambda/k \sum_{i=1}^k g_i(X)$ and then defining the ROBUST optimization on $f'_i$’s and $g'_i$.

### 1.1 Motivating Applications

This section provides an overview of two specific applications which motivate Problems 1-4. We also list down a few more motivating applications in the extended version.

**Robust Co-operative Cuts and Matchings:** Co-operative cuts have proven to be a very rich class of models for image segmentation where one can model co-operation between edges to solve the shrinking bias problem (elongated edges not getting segmented properly) [25][26]. Another application is co-operative matching where one can use co-operation among spatially similar groups of pixels to be matched together [21]. Both these cases have been formulated as submodular minimization under combinatorial constraints such as s-t cuts and matchings. The way we can model this is by finding a clustering among the pixels and defining clustered concave over modular functions [26][21]. Instead of taking a single clustering or an average among clusterings, one can pose this as a robust case problem where we minimize the worst among the clusterings to achieve more robust segmentations (equivalently assignments). In this case, ROBUST-SUBMIN is a natural formulation.

**Robust Data Subset Selection:** Submodular functions have successfully been used for several data subset selection in domains such as image classification [31], speech recognition [53] and machine translation [33][52] prove that the problem of selecting the maximum likelihood subset of a training dataset is a submodular optimization problem for several classifiers including nearest neighbor, naive Bayes etc. Another approach, which we shall study in this paper, is selecting data-sets which are robust to several data subset selection models $g_1, \cdots, g_k$. These models can be defined via different submodular functions, different choice of features, perturbations in the feature space and different target distributions where we want to use these models. In this case, we can pose this as an instance of ROBUST-SUBMAX where we want to maximize the minimum among the utility functions. Furthermore, it is also natural to select subsets of data which minimize the complexity of the dataset [41][39]. In this case, the functions $f$ captures the complexity of the data-sets $X$ (for example, vocabulary size in speech recognition and number of objects in object detection). This is naturally an instance of ROBUST-SCSK with $g_i$ being the data selection models, while $f$ is the complexity of the selected subset ($l = 1$). We can also define multiple complexity functions $f_i$, where each function is defined via different perturbations in the vocabulary (obtained by say randomly deleting a certain fraction of words from the vocabulary function).

**Robust Observation Selection:** [36] study the problem of robust submodular maximization with multiple submodular objectives. They argue how for several applications including robust experimental design and sensor placement, it is important to select observations which are robust to several objectives. This is an instance of ROBUST-SUBMAX. However, often we want to select observations with more general and with multiple co-operative cost constraints (the constraints here would be submodular) and we have an instance of ROBUST-SCSK. Another natural model here is to MIXED-SUBMAX and MIXED-SCSK.

**Submodular Models for Summarization:** Submodular Maximization is a natural formulation for various summarization problems including Document summarization [38][40], image summarization [31] and video summarization [16][32]. Most models today are average case (i.e. sums of
submodular components). A better objective is MIXED-SUBMAX where we weight both the worst case and average case objectives.

### 1.2 Related Work and Our Contributions

**Submodular Minimization, Maximization and SCSC/SCSK:** Problems 1 and 2 are direct generalizations of constrained submodular minimization and maximization. Both these problems are NP hard under constraints even when \( f \) is monotone \([45, 55, 14, 25]\). The greedy algorithm achieves a \( 1 - 1/e \) approximation for cardinality constrained maximization and a \( 1/2 \) approximation under matroid constraints \([10, 7]\). Achieved a tight \( 1 - 1/e \) approximation for matroid constraints using the continuous greedy. \([37]\) later provided a similar \( 1 - 1/e \) approximation under multiple knapsack constraints. Constrained submodular minimization is much harder – even with simple constraints such as a cardinality lower bound constraints, the problems are not approximable better than a polynomial factor of \( \Omega(\sqrt{n}) \). \([50]\). Similarly the problems of minimizing a submodular function under covering constraints \([17]\), spanning trees, perfect matchings, paths \([13]\) and cuts \([25]\) have similar polynomial hardness factors. In all of these cases, matching upper bounds (i.e approximation algorithms) have been provided. \([50, 17, 13, 25]\). In \([23, 24]\), authors provide a scalable semi-gradient based framework which improve upon the worst case poly approximation for functions with bounded curvature \( \kappa \) (which several submodular functions occurring in real world applications have). Problems 3 and 4 generalize SCSC and SCSK studied in \([20]\). The authors provide tight approximation guarantees in this setting using the the semi-gradient framework \([23]\) and the Ellipsoidal approximations \([15]\). Similar to \([22]\) the authors also provide curvature based guarantees.

**Robust Submodular Maximization:** One of the first papers to study robust submodular maximization (problem 2) was \([35]\), where the authors study ROBUST-SUBMAX with cardinality constraints. The authors reduce this problem to a submodular set cover problem using the saturate trick to provide a bi-criteria approximation guarantee. \([3]\) extend this work and study ROBUST-SUBMAX subject to matroid constraint. They provide bi-criteria algorithms by creating a union of \( O(l/e) \) independent sets, with the union set having a guarantee of \( 1 - \epsilon \). They also discuss extensions to knapsack and multiple matroid constraints and provide bicriteria approximation of \((1 - \epsilon, O(l/e))\). \([48]\) also study the same problem. However, they take a different approach by presenting a bi-criteria algorithm that outputs a feasible set that is good only for a fraction of the \( k \) submodular functions \( g_i \). \([6, 24]\) study a slightly general problem of robust non-convex optimization (of which robust submodular optimization is a special case), but they provide weaker guarantees compared to \([6, 24]\). Another related problem, which is also called Robust Submodular Maximization \([46]\) attempts to select subsets which maximize a submodular function and are robust to up to \( \tau \) deletions. Mathematically, this problem can be cast as \( \max_{|A| \leq k} \min_{|Z \subseteq A, |Z| \leq \tau} f(A - Z) \). Note that this is a special case of the robust submodular maximization studied in \([35]\) except that \( k \) here is potentially exponential in \( \tau \). \([46]\) provide constant factor approximation algorithms for this problem. The results were later improved by \([4]\). Another version of this problem was considered by \([49]\) where they study distributionally robust submodular optimization.

**Robust Min-Max Combinatorial Optimization:** From a minimization perspective, several researchers have studied robust min-max combinatorial optimization (a special case of ROBUST-SUBMIN with modular functions) under different combinatorial constraints (see \([1, 29]\) for a survey). Unfortunately these problems are NP hard even for constraints such as knapsack, s-t cuts, s-t paths, assignments and spanning trees where the standard linear cost problems are poly-time solvable \([1, 29]\). Moreover, the lower bounds on hardness of approximation is \( \Omega((log^{1-\epsilon} l)) \) (\( l \) is the number of functions) for s-t cuts, paths and assignments \([27]\) and \( \Omega((log^{1-\epsilon} n)) \) for spanning trees \([28]\) for any \( \epsilon > 0 \). For the case when \( l \) is bounded (a constant), fully polynomial time approximation schemes have been proposed for a large class of constraints including s-t paths, knapsack, assignments and spanning trees \([2, 1, 29]\). From an approximation algorithm perspective, the best known general result is an approximation factor of \( l \) for constraints where the linear function can be exactly optimized in poly-time. For special cases such as spanning trees and shortest paths, one can achieve improved approximation factors of \( O(log n) \) \([28, 30]\) and \( \tilde{O}(\sqrt{n}) \). \([30]\).

**Our Contributions:** Among Problems 1-4, past work has mainly focused on ROBUST-SUBMAX. In this paper, we close this gap by providing approximation algorithms and hardness results for the rest of the three problems. For ROBUST-SUBMIN, we provide several approximation algorithms, which are either combinatorial or rely on continuous relaxations. We show approximation factors for a large class of constraints including cardinality, matroid span, spanning trees, cuts, paths and matchings and their corresponding hardness bounds. Our approximation bounds depend on \( n \) (the ground set) and curvature of the submodular functions \( f \). \([22]\). For ROBUST-SUBMAX, we complement previous work by providing an approximation guarantee for multiple knapsack constraints. In the

---

1. Ignores log factors
case of Problems ROBUST-SCSC and ROBUST-SCSK, we provide bicriteria approximation factors which depend on the curvature of the $f_i$’s and $n$. We also provide lower bounds for both problems. Furthermore, We show that Problems 3 and 4 are closely related to each other and are duals in that a bi-criteria approximation algorithm for one of them can provide a bicriteria approximation of the other one. Finally, we demonstrate the performance of our algorithms on real world applications.

2 Main Ideas and Techniques

In this section, we will review some of the constructs and techniques used in this paper to provide approximation algorithms for Problems 1 - 4.

The Submodular Polyhedron and Lovász extension: For a submodular function $f$, the submodular polyhedron $P_f$ and the corresponding base polytope $B_f$ are respectively defined as $P_f = \{x : x(S) \leq f(S), \forall S \subseteq V\}$, $B_f = P_f \cap \{x : x(V) = f(V)\}$. For a vector $x \in \mathbb{R}^V$ and a set $X \subseteq V$, we write $x(X) = \sum_{j \in X} x(j)$. Though $P_f$ is defined via $2^n$ inequalities, its extreme point can be easily characterized. Given any permutation $\sigma$ of the ground set $\{1, 2, \ldots, n\}$, and an associated chain $\emptyset = S^0_0 \subseteq S^1_1 \subseteq \cdots \subseteq S^n_n = V$ with $S^i_i = \{\sigma(1), \sigma(2), \ldots, \sigma(i)\}$, a vector $h^{\sigma}_f$ satisfying, $h^{\sigma}_f(\sigma(i)) = f(S^i_i) - f(S^i_{i-1}) = f(\sigma(i)|S^i_{i-1}), \forall i = 1, \cdots, n$ forms an extreme point of $P_f$. Moreover, a natural convex extension of a submodular function, called the Lovász extension, is closely related to the submodular polyhedron, and is defined as $\hat{f}(x) = \max_{h \in P_f} \langle h, x \rangle$.

Thanks to the properties of the polyhedron, $\hat{f}(x)$ can be efficiently computed: Denote $\sigma_x$ as an ordering induced by $x$, such that $x(\sigma_x(1)) \geq x(\sigma_x(2)) \geq \cdots \geq x(\sigma_x(n))$. Then the Lovász extension is $\hat{f}(x) = (h^{\sigma_x}_f, x)$ [42, 9]. The gradient of the Lovász extension $\nabla \hat{f}(x) = \sigma_x$.

Modular lower bounds (Sub-grads) and Modular upper bounds (Super-grads): Akin to convex functions, submodular functions have tight modular lower bounds. These bounds are related to the sub-differential $\partial f(Y)$ of the submodular set function $f$ at a set $Y \subseteq V$, which is defined as $\partial f(Y) = \{y \in \mathbb{R}^V : f(X) - y(X) \geq f(Y) - y(Y), \forall X \subseteq V\}$. Denote a sub-gradient at $Y$ by $h^f_Y \in \partial f(Y)$. Define $h^f_Y = h^{f_Y}_f$ (see the definition of $h^f_f$ from the previous paragraph) forms a lower bound of $f$, tight at $Y$ — i.e., $h^f_Y(X) = \sum_{j \in X} h^f_Y(j) \leq f(X), \forall X \subseteq V$ and $h^f_Y(V) = f(Y)$. Notice that the extreme points of a sub-differential are a subset of the extreme points of the submodular polyhedron. We can also define super-differentials $\partial f(Y)$ of a submodular function [26, 19] at $Y$: $\partial f(Y) = \{y \in \mathbb{R}^n : f(X) - y(X) \leq f(Y) - y(Y), \forall X \subseteq V\}$. It is possible, moreover, to provide specific supe-grads [19, 23] that define the following two modular upper bounds: $m^{f}_{X,1}(Y) \triangleq f(X) - \sum_{j \in X \setminus Y} f(j|X \setminus j) + \sum_{j \in Y \setminus X} f(j|X)$ and $m^{f}_{X,2}(Y) \triangleq f(X) - \sum_{j \in X \setminus Y} f(j|V \setminus j) + \sum_{j \in Y \setminus X} f(j|X)$. Then $m^{f}_{X,1}(Y) \geq f(Y)$ and $m^{f}_{X,2}(Y) \geq f(Y), \forall Y \subseteq V$ and $m^{f}_{X,1}(X) = m^{f}_{X,2}(X) = f(X)$.

Majorization-Minimization (MMin) and Minorization-Maximization (MMax) Framework: This is a general framework of in [23, 26] which has been used for several problems including submodular minimization [23, 26, 44, 18], SCSC/SCSK [20, 9] and many others. The basic idea of this framework is quite simple. The super-grads and sub-grads defined above allow one to define upper and lower bounds of a submodular function which can be iteratively optimized. Our framework iteratively minimizes these upper bounds (equivalently maximizes these lower bounds) at every iteration.

Continuous Relaxation Framework: Another framework for submodular optimization relies on optimizing a suitable relaxation of a submodular function. In the case of minimization problems, a suitable relaxation is a Lovász extension [42, 9] while for maximization, the multilinear extension [42, 9] is based on approximating the submodular polyhedron by an ellipsoid. The main result states that any polymatroid (monotone submodular) function $f$, can be approximated by a function of the form $\sqrt{w^{\tau}(X)}$ for a certain modular weight vector $w^{\tau} \in \mathbb{R}^V$, such that $\sqrt{w^{\tau}(X)} \leq f(X) \leq O(\sqrt{n \log n}) \sqrt{w^{\tau}(X)}, \forall X \subseteq V$. One can then optimize the submodular function by optimizing the ellipsoidal approximation.

A bicriteria approximation $(\sigma, \rho)$ for a problem $\min \{f(X) | g(X) \geq c\}$ finds a set $\hat{X}$ such that $f(\hat{X}) \leq \sigma f(X^*)$ and $g(\hat{X}) \geq \rho c$ for $\sigma \geq 1, \rho \leq 1$. 


In this section, we shall go over the hardness and approximation algorithms for ROBUST-SUBMIN.

### 3 Robust Submodular Minimization and Maximization (Problems 1 and 2)

#### 3.1 Robust Submodular Minimization

We shall consider two cases, one where $f$ is bounded (i.e., its a constant), and the other where $f$ is unbounded. We start by defining the quantity $K(f, \kappa) = f/(1 + (1 - \kappa)(f - 1))$, where $\kappa$ is the curvature of the submodular function for which the bound is obtained. Note that $K(f, \kappa)$ interpolates the bound between $\max_i f_i$ and $\lambda \sum_i f_i(X)$, which is the exact minimizer of $\mathcal{I}_2$-

**Hardness:** Since ROBUST-SUBMIN generalizes robust min-max combinatorial optimization (when the functions are modular), we have the hardness bounds from [27, 28]. For the modular case, the lower bounds are $\Omega((\log^{1-\epsilon} n) (l)$ is the number of functions) for $s$-t-cuts, paths and assignments [27] and $\Omega((\log^{1-\epsilon} n) \log n)$ for spanning trees [28]. We shall consider these hardness results in the first column of Table 1. The curvature $\kappa$ corresponds to the worst curvature among the functions $f_i$ (i.e., $\kappa = \max_i \kappa_i$).

**Average Approximation:** A simple observation is that $f_{avg}(X) = 1/l \sum_{i=1}^l f_i(X)$ approximates $\max_{i=1:l} f_i(X)$. As a result, minimizing $f_{avg}(X)$ implies an approximation for ROBUST-SUBMIN.

**Theorem 1.** Given a non-negative set function $f$, define $f_{avg}(X) = \frac{1}{l} \sum_{i=1}^l f_i(X)$. Then $f_{avg}(X) \leq \max_{i=1:l} f_i(X) \leq \lambda f_{avg}(X)$. Denote $\hat{X}$ as $\beta$-approximate optimizer of $f_{avg}$. Then $\max_{i=1:l} f_i(\hat{X}) \leq \beta \max_{i=1:l} f_i(X^*)$ where $X^*$ is the exact minimizer of ROBUST-SUBMIN.

**Proof.** To prove the first part, notice that $f_i(X) \leq \max_{i=1:l} f_i(X)$, and hence $1/l \sum_{i=1}^l f_i(X) \leq \max_{i=1:l} f_i(X)$. The other inequality also directly follows since the $f_i$’s are non-negative and hence $\max_{i=1:l} f_i(X) \leq \sum_{i=1}^l f_i(X) = f_{avg}(X)$. To prove the second part, observe that $\max_{i=1:l} f_i(\hat{X}) \leq \lambda f_{avg}(\hat{X}) \leq f_{avg}(X^*) \leq \beta \max_{i=1:l} f_i(X^*)$. The first inequality holds from the first part of this result, the second inequality holds since $\hat{X}$ is a $\beta$-approximate optimizer of $f_{avg}$ and the third part of the theorem holds again from the first part of this result.

Since $f_{avg}$ is a submodular function, we can use the majorization-minimization (which we call MMin-AA) and ellipsoidal approximation (EA-AA) for constrained submodular minimization [23, 22, 14, 25].

**Corollary 2.** Using the majorization minimization (MMin) scheme with the average approximation achieves an approximation guarantee of $O(K(|X^*|, \kappa_{avg}))$ where $X^*$ is the optimal solution of ROBUST-SUBMIN and $\kappa_{avg}$ is the curvature of $f_{avg}$. Using the curvature-normalized ellipsoidal approximation algorithm from [22, 23] achieves a guarantee of $O(lK(|V| \log |V|, \kappa_{avg}))$.

This corollary directly follows by combining the approximation guarantee of MMin and EA [22, 23] with Theorem 1. Substituting the values of $|V|$ and $|X^*|$ for various constraints, we get the results in
For spanning trees and shortest path constraints, MMin achieves a
result. Next, observe that which maximizes \( w \), which uses upper bounds of the submodular functions defined via supergradients. Starting with \( X^0 = 0 \), the algorithm proceeds as follows. At iteration \( t \), it constructs modular upper bounds for each function \( f_i, m^f_i \), which is tight at \( X^t \). The set \( X^{t+1} = \text{argmin}_{X \in C} \max_i m^f_i(X) \). This is a min-max robust optimization problem. The following theorem provides the approximation guarantee for MMin.

**Theorem 3.** If \( l \) is a constant, MMin achieves an approximation guarantee of \( (1 + \epsilon)K(\|X^*\|, \kappa_{wL}) \) for the knapsack, spanning trees, matching and s-t path problems. The complexity of this algorithm is exponential in \( l \). When \( l \) is unbounded, MMin achieves an approximation guarantee of \( O(\min(\log n, l)K(n, \kappa_{wL})) \) and a \( O(\min(\sqrt{n}, l)K(n, \kappa_{wL})) \) approximation. Under cardinality and partition matroid constraints, MMin achieves a \( O(\log K(n, \kappa_{wL})/\log \log l) \) approximation.

Substituting the appropriate bounds on \( |X^*| \) for the various constraints, we get the results in Tables 1 and 2. \( \kappa_{wL} \) corresponds to the worst case curvature \( \max_i \kappa_f \). When \( l \) is bounded, the corresponding min-max robust optimization problem in every iteration of MMin can be solved via an FPTAS. In particular, we can obtain a \( 1 + \epsilon \) approximation for shortest paths in \( O(n^l+1/\epsilon^l) \), trees in \( O(n(l+1)/\epsilon \log n/\epsilon) \), matchings in \( O(n(l+1)/\epsilon \log n/\epsilon) \) and knapsack in \( O(n^{l+1}/\epsilon) \). The results for constant \( l \) is shown in Table 2 (column corresponding to MMin). When \( l \) is not constant, we cannot use the FPTAS since they are all exponential in \( l \). In the case of spanning trees and shortest path constraints, the min-max robust optimization problem can be approximated up to a factor of \( O(\log n) \) and \( O(\sqrt{n}) \) via a randomized LP relaxation \([28, 30]\). Similarly in the case of cardinality and partition matroid constraints, we can achieve a \( O(\log l/\log \log l) \) approximation using LP relaxations \([30]\).

For the other constraints, we use the modular cost \( m_r = \max_i m^f_i[r], r \in V \). This provides a \( l \) approximation. The approximation guarantees of MMin for unbounded \( l \) is shown in Table 2.

Before proving the result, we shall make a simple observation. Given a min-max objective function \( \min_{X \in C} w_{\text{max}}(X) \) where \( w_{\text{max}} = \max_j w_j(X) \), define a modular approximation \( w_{\text{modmax}}(X) = \sum_{i \in j} \max(w_{j}(i)) \). Also define \( w_{\text{avg}}(X) = 1/l \sum_i w_i(j) \).

**Lemma 4.** Given \( w_i(j) \geq 0 \), it holds that \( w_{\text{modmax}}(X) \geq w_{\text{max}}(X) \geq \frac{1}{l} w_{\text{modmax}}(X) \). Furthermore, \( w_{\text{avg}}(X) \leq w_{\text{max}}(X) \leq l w_{\text{avg}}(X) \).

**Proof.** The second result follows from Theorem 1, if we define \( f_i(X) = w_i(X) \). To prove the first result, we start with proving \( w_{\text{max}}(X) \leq w_{\text{modmax}}(X) \). For a given set \( X \), let \( i_X \) be the index which maximizes \( w_{\text{max}}(X) \). Then \( w_{i_X}(j) \leq \max_i w_i(j) \) from which we get the result. Next, observe that \( w_{\text{modmax}}(X) \leq \sum_{i_X} w_i(j) = lw_{\text{avg}}(X) \) which proves this part.

This shows that \( w_{\text{avg}} \) and \( w_{\text{modmax}} \) form upper and lower bounds of \( w_{\text{max}} \) and moreover, are both \( l \)-approximations of \( w_{\text{max}} \).

We now elaborate on the Majorization-Minimization algorithm and prove the result. At every round of the majorization-minimization algorithm we need to solve

\[
X^{t+1} = \text{argmin}_{X \in C} \max_i m^f_i(X).
\]

We consider three cases. The first is when \( l \) is a constant. In that case, we can use an FPTAS to solve Eq. 1. We can obtain a \( 1 + \epsilon \) approximation for shortest paths in \( O(n^{l+1}/\epsilon^l) \), trees in \( O(n(l+1)/\epsilon \log n/\epsilon) \), matchings in \( O(n(l+1)/\epsilon \log n/\epsilon) \) and knapsack in \( O(n^{l+1}/\epsilon) \). The second case is a generic algorithm when \( l \) is not constant. In this case, at every iteration of MMin use these two bounds (the \( \text{avg} \) and \( \text{modmax} \) bounds) on the function \( \max_i m^f_i(X) \) and choose the solution with a better solution. The bound of MMin directly follows from the observation that both the bounds (the \( \text{avg} \) and \( \text{modmax} \) of \( m_i(X) \)) are \( l \)-approximations of \( m_i(X) \). Finally, for the special cases of spanning trees, shortest paths, cardinality and partition matroid constraints, there exist LP relaxation based algorithms which achieve approximation factors of \( O(\log n) \), \( O(\sqrt{n} \log l/\log \log l) \), \( O(\log l/\log \log l) \) and \( O(\log l/\log \log l) \) respectively.

We now prove Theorem 5.
Proof. Assume we have an $\alpha$ approximation algorithm for solving problem $[1]$. We start MMin with $X^0 = \emptyset$. We prove the bound for MMin for the first iteration. Observe that $m^*_{\emptyset}(X)$ approximate the submodular functions $f_i(X)$ up to a factor of $K(|X|, \kappa_i)$ [22]. If $\kappa_{\text{Max}}$ is the maximum curvature among the functions $f_i$, this means that $m^*_{\emptyset}(X)$ approximate the submodular functions $f_i(X)$ up to a factor of $K(|X|, \kappa_{\text{Max}})$ as well. Hence $\max_i m^*_{\emptyset}(X)$ approximates $\max_i f_i(X)$ with a factor of $K(|X|, \kappa_{\text{Max}})$. In other words, $\max_i f_i(X) \leq \max_i m^*_{\emptyset}(X) \leq K(|X|, \kappa_{\text{Max}}) \max_i f_i(X)$. Let $\hat{X}_1$ be the solution obtained by optimizing $m^*_{\emptyset}$ (using an $\alpha$-approximation algorithms for the three cases described above). It holds that $\max_i m^*_{\emptyset}(\hat{X}_1) \leq \alpha \max_i m^*_{\emptyset}(X^*)$ where $X^*$ is the optimal solution of $\max_i m^*_{\emptyset}(X)$ over the constraint $C$. Furthermore, denote $X^*$ as the optimal solution of $\max_i f_i(X)$ over $C$. Then $\max_i m^*_{\emptyset}(X^*) \leq \max_i m^*_{\emptyset}(X^*) \leq K(|X^*|, \kappa_{\text{Max}}) \max_i f_i(X^*)$. Combining both, we see that $\max_i f(X_1) \leq \max_i m^*_{\emptyset}(X_1) \leq \alpha K(|X^*|, \kappa_{\text{Max}}) \max_i f_i(X^*)$. We then run MMin for more iterations and only continue if the objective value increases in the next round. Using the values of $\alpha$ for the different cases above, we get the results. \qed

Ellipsoidal Approximation: Next, we use the Ellipsoidal Approximation to approximate the submodular function $f_i$. To account for the curvature of the individual functions $f_i$, we use the curve-normalized Ellipsoidal Approximation [22]. We then obtain the functions $f_i(X)$ which are of the form $(1 - \kappa f_i)\sqrt{w_i(X)} + \kappa f_i \sum_{j \in X} f_i(j)$, and the problem is then to optimize $\max_i f_i(X)$ subject to the constraints $C$. This is no longer a min-max optimization problem. The following result shows that we can still achieve approximation guarantees in this case.

Theorem 5. For the case when $l$ is a constant, EA achieves an approximation guarantee of $O(K(\sqrt{|V|} \log |V|, \kappa_{\text{Max}}))$ for the knapsack, spanning trees, matching and s-t path problems. The complexity of this algorithm is exponential in $l$. When $l$ is unbounded, the EA algorithm achieves an approximation guarantee of $O(\sqrt{l} \sqrt{|V|} \log |V|)$ for all constraints. In the case of spanning trees, shortest paths, the EA achieves approximation factors of $O(\min(\log n, \sqrt{l}) \sqrt{n \log n})$, and $O(\min(n^{0.25}, \sqrt{l}) \sqrt{n \log n})$. Under cardinality and partition matroid constraints, EA achieves a $O(\sqrt{l} \sqrt{n \log l} \sqrt{n \log n})$ approximation.

For the case when $l$ is bounded, we reduce the optimization problem after the Ellipsoidal Approximation into a multi-objective optimization problem, which provides an FPTAS for knapsacks, spanning trees, matching and s-t path problems [47]. When $l$ is unbounded, we further reduce the EA approximation objective into a linear objective which then provides the approximation guarantees similar to MMin above. However, as a result, we lose the curvature based guarantee.

Proof. First we start with the case when $l$ is a constant. Observe that the optimization problem is

$$\min_{X \in C} \max_i f_i(X) = \min_{X \in C} \max_i (1 - \kappa f_i)\sqrt{w_i(X)} + \kappa f_i \sum_{j \in X} f_i(j)$$

(2)

This is of the form $\min_{X \in C} \max_i \sqrt{w_i(X)} + w_i(X)$. Define $h(y_1, y_2, y_2, \ldots, y_1, y_2) = \max_i \sqrt{y_1} + y_2$. Note that the optimization problem is $\min_{X \in C} h(w_1(X), w_2(X), \ldots, w_1(X), w_2(X))$. Observe that $h(y) \leq h(y')$ if $y \leq y'$. Furthermore, note that $y \geq 0$. Then given a $\lambda > 1$, $h(\lambda y) = \max_i \sqrt{\lambda y_1} + \lambda y_2 \leq \lambda \sqrt{y_1} + \lambda y_2 \leq \lambda h(y)$. As a result, we can use Theorem 3.3 from [43] which provides an FPTAS as long as the following exact problem can be solved on $C$: Given a constant $C$ and a vector $c \in \mathbb{R}^n$, does there exist a $x$ such that $c(x) = C$? A number of constraints including matchings, knapsacks, s-t paths and spanning trees satisfy this [47]. For these constraints, we can obtain a $1 + \epsilon$ approximation algorithm in complexity exponential in $l$.

When $l$ is unbounded, we directly use the Ellipsoidal Approximation and the problem then is to optimize $\min_{X \in C} \max_i \sqrt{w_i(X)}$. We then transform this to the following optimization problem: $\min_{X \in C} \max_i w_i(X)$. Assume we can obtain an $\alpha$ approximation to the problem $\min_{X \in C} \max_i w_i(X)$. This means we can achieve a solution $X$ such that $\max_i w_i(X) \leq \alpha \max_i w_i(X)$ where $X^*$ is the optimal solution for the problem $\min_{X \in C} \max_i w_i(X)$. Then observe that $\max_i \sqrt{w_i(X)} \leq \max_i \sqrt{w_i(X^*)} \leq \max_i f_i(X^*)$. Combining all the inequalities and also using the bound of the Ellipsoidal Approximation, we have $\max_i f_i(X) \leq \max_i \sqrt{w_i(X)} \leq \max_i \sqrt{w_i(X^*)} \leq \max_i f_i(X)$. Combining this with the approximation guarantee of $\alpha$, we get $\alpha \max_i f_i(X) \leq \alpha K(|X^*|, \kappa_{\text{Max}}) \max_i f_i(X^*)$.
\[ \beta \max_i \sqrt{w_{f_i}(X)} \leq \beta \sqrt{\alpha} \max_i \sqrt{w_{f_i}(X^{\alpha})} \leq \beta \sqrt{\alpha} \max_i \sqrt{w_{f_i}(X^*)} \leq \beta \sqrt{\alpha} \max_i f_i(X^*) \]

where \( \beta \) is the approximation of the Ellipsoidal Approximation.

We now use this result to prove the theorem. Consider two cases. First, we optimize the \( \text{avg} \) and \( \text{modmax} \) versions of \( w_{f_i}(X) \) which provide \( \alpha = l \) approximation. Secondly, for the special cases of spanning trees, shortest paths, cardinality and partition matroid constraints, there exist LP relaxation based algorithms which achieve approximation factors \( \alpha \) being \( O(\log n) \), \( O(\sqrt{n} \log l / \log \log l) \), \( O(\log l / \log \log l) \) and \( O(\log l / \log \log l) \) respectively. Substitute these values of \( \alpha \) and using the fact that \( \beta = O(\sqrt{|V| \log |V|}) \), we get the approximation bound.

**Continuous Relaxation:** Here, we use the continuous relaxation of a submodular function. In particular, we use the relaxation \( \max_i f_i(x), x \in [0, 1]^{|V|} \) as the continuous relaxation of the original function \( \max_i f_i(X) \) (here \( f \) is the Lovász extension). Its easy to see that this is a continuous relaxation. Since the Lovász extension is convex, the function \( \max_i f_i(x) \) is also a convex function. This means that we can exactly optimize the continuous relaxation over a convex polytope. The remaining question is about the rounding and the resulting approximation guarantee due to the rounding. Given a constraint \( C \), define the up-monotone polytope similar to \( \mathcal{P}_C = \{ x \in [0, 1]^{|V|} : \sum_{i \in W} x_i \geq b_W \text{ for all } W \in \mathcal{W} \} \) for a family of sets \( \mathcal{W} = \{ W_1, \ldots \} \). We then round the solution using the following rounding scheme. Given a continuous vector \( \hat{x} \) (which is the optimizer of \( \max_i f_i(x), x \in \mathcal{P}_C \)), order the elements based on \( \sigma^2 \). Denote \( X_i = [\sigma_1^2, \ldots, \sigma_i^2] \) so we obtain a chain of sets \( \emptyset \subseteq X_1 \subseteq X_2 \subseteq \cdots \subseteq X_n \). Our rounding scheme picks the smallest \( k \) such that \( X_k \in \mathcal{C} \). Another way of checking this is if there exists a set \( X \in X_k \) such that \( X \in \mathcal{C} \). Since \( \mathcal{C} \) is up-monotone, such a set must exist. The following result shows the approximation guarantee.

**Theorem 6.** Given submodular functions \( f_i \) and constraints \( C \) which can be expressed as \( \{ x \in [0, 1]^{|V|} : \sum_{i \in W} x_i \geq b_W \text{ for all } W \in \mathcal{W} \} \) for a family of sets \( \mathcal{W} = \{ W_1, \ldots \} \), the continuous relaxation scheme achieves an approximation guarantee of \( \max_{W \in \mathcal{W}} |W| - b_W + 1 \). If we assume the sets in \( \mathcal{W} \) are disjoint, the integrality bounds matches the approximation bounds.

**Proof.** The proof of this theorem is closely in line with Lemma 2 and Theorem 1 from \([24]\). We first show the following result. Given monotone submodular functions \( f_i, i \in 1, \ldots, l \), and an optimizer \( \hat{x} \) of \( \max_i f_i(x) \), define \( \hat{X}_0 = \{ i : \hat{x}_i \geq \tilde{\theta} \} \). We choose \( \tilde{\theta} \) such that \( \hat{X}_0 \in C \). Then \( \max_i f_i(\hat{X}_0) \leq 1/\tilde{\theta} \max_i f_i(X^*) \) where \( X^* \) is the optimizer of \( \min_{X \in C} \max_i f_i(X) \). To prove this, observe that, by definition \( \emptyset \in \hat{X}_0 \leq \hat{X}_0 \). As a result, \( \forall i, f_i(\hat{X}_0) = \tilde{\theta} f_i(\hat{X}_0) \leq f_i(\hat{x}) \) (this follows because of the positive homogeneity of the Lovász extension. This implies that \( \tilde{\theta} \max_i f_i(\hat{X}_0) \leq f_i(\hat{x}) \leq \min_{X \in C} f_i(\hat{x}) \leq \max_{X \in C} f_i(X) \). The last inequality holds from the fact that the discrete solution is greater than the continuous one since the continuous one is a relaxation. This proves this part of the theorem.

Next, we show that the approximation guarantee holds for the class of constraints defined as \( \{ x \in [0, 1]^{|V|} : \sum_{i \in W} x_i \geq b_W \text{ for all } W \in \mathcal{W} \} \). This follows directly from the Proof of Theorem 1 in \([24]\). We can then obtain the approximation guarantees for different constraints including cardinality, spanning trees, matroids, set covers, edge covers and vertex covers, matchings, cuts and paths by appropriately defining the polytopes \( \mathcal{P}_C \) and appropriately setting the values of \( \max_{W \in \mathcal{W}} |W| - b_W + 1 \). We refer to the reader to Section 3 in \([24]\).

**Tightness of the Bounds:** Given the bounds in Tables 1 and 2, we discuss the tightness of these bounds viz-a-via the hardness. In the case when \( l \) is a constant, MMin achieves tight bounds for Trees, Matchings and Paths while the EA achieves tight bounds up to \( \log \) factors for knapsack constraints. In the case when \( l \) is not a constant, MMin achieves a tight bound up to \( \log \) factors for spanning tree constraints. The continuous relaxation scheme obtains tight bounds in the case of vertex covers. In the case when the functions \( f_i \) have curvature \( \kappa = 1 \) CR also obtains tight bounds for edge-covers and matchings. We also point out that the bounds of average approximation (AA) depend on the average case curvature as opposed to the worst case curvature. However, in practice, the functions \( f_i \) often belong to the same class of functions in which case all the functions \( f_i \) have the same curvature.

\(^11_A \) is the indicator vector of set \( A \) such that \( 1_A[i] = 1 \) iff \( i \in A \).
3.2 Robust Submodular Maximization

Krause et al [35] show that \textsc{Robust-SubMax} is inapproximable up to any polynomial factor unless P = NP even with cardinality constraints. [3] extend this to matroid and knapsack constraints. [35, 3] provide a bi-criteria approximation factor of \((1 - \epsilon, O(\log l/\epsilon))\) for cardinality constraints. For Matroid constraints, they provide bi-criteria algorithms by creating a union of \(O(\log l/\epsilon)\) independent sets, with the union set having a guarantee of \(1 - \epsilon\). This result can be extended to multiple matroid constraints [3]. Anari et al [3] also obtained a \((1 - \epsilon, O(\log l/\epsilon))\) approximation for a single knapsack constraint. In the following theorem, we provide a bi-criteria approximation for \textsc{Robust-SubMax} under multiple knapsack constraints.

**Theorem 7.** Using a modified greedy algorithm, we achieve a \((1 - \epsilon, O(l/\log l))\) bi-criteria approximation for \textsc{Robust-SubMax} under \(l\) knapsack constraints. Using the continuous greedy algorithm, we can achieve an improved factor of \((1 - \epsilon, O(\ln l/\epsilon))\) for the same problem.

Since the greedy algorithm does not work directly for multiple knapsack constraints, we convert the problem first into a single knapsack constraint which provides the bi-criteria approximation. The approximation is worse by a factor \(l\). Using the continuous greedy algorithm similar to [3] and the rounding scheme of [37] we achieve the tight approximation guarantee.

**Proof.** We start with the optimization problem:

\[
\max \min_{X \subseteq V} g_i(X) \mid w_i(X) \leq b_i, i = 1, \ldots, l. \tag{3}
\]

Note that the constraints can equivalently be written as \(\max_i w_i(X)/b_i \leq 1\). We can then define two approximations of this. One is the \text{modmax} approximation which is \(\sum_{i \in X} \max_j w_{ij}/b_i\) and the other is the \text{avg} approximation: \(\sum_{i \in X} \sum_j w_{ij}/b_i\). Both these are \(l\)-approximations of \(\max_i w_i(X)/b_i \leq 1\). Denote these approximations as \(\tilde{w}(X)\) and w.l.o.g assume that \(\tilde{w}(X) \leq \max_i w_i(X)/b_i \leq l\tilde{w}(X)\). We can then convert this to an instance of \textsc{Robust-SubMax} with a single knapsack constraint:

\[
\max \min_{X \subseteq V} g_i(X) \mid \tilde{w}(X) \leq 1 \tag{4}
\]

From [3], we know that we can achieve a \((1 - \epsilon, l/\log l/\epsilon)\) bi-criteria approximation. In other words, we can achieve a solution \(\tilde{X}\) such that \(\min_i g_i(\tilde{X}) \geq (1 - \epsilon) \min_i g_i(X^*)\) and \(\tilde{w}(\tilde{X}) \leq l\tilde{w}(X)\). Since we have that \(\max_i w_i(X)/b_i \leq l\tilde{w}(X)\), this implies \(\max_i w_i(X)/b_i \leq l\ln l/\epsilon\). This completes the first part. In practice, we can use both the approximations (i.e. the \text{modmax} and the \text{avg} approximation and choose the better among the solutions).

Next, we prove the second part of the theorem. This uses the continuous greedy algorithm and we use the proof technique from [3]. In particular, first truncate \(g_i\)’s to \(c\), so we can define \(g_i^c(X) = \min(g_i(X), c)\). Define the multi-linear extension of \(g_i^c\) as \(G_i^c\). First we argue that we can obtain a solution \(y(\tau)\) at time \(\tau\) such that \(G_i^c(y(\tau)) \geq (1 - c - \epsilon)c\), \(\forall i\) where \(y(\tau) \in \tau\mathcal{P}_c\). In other words, \(y(\tau)\) satisfies \(\langle y, w_i \rangle / b_i \leq \tau, \forall i\). This follows from Claim 1 in [3] since the result holds for any down-monotone polytope [7]. Next, we use the rounding technique from [37]. In order to do this, we first set \(\tau = \ln l/\epsilon\) so we can achieve a solution so \(G_i^c(y) \geq (1 - \epsilon/l)c\). Notice that this implies \(G(X) = \sum_i G_i^c(y) \geq (1 - \epsilon/l)c\). Since \(G^c\) is a single submodular function, we can round the obtained fractional solution \(y\) using the rounding scheme from [37]. This will achieve a discrete solution \(X\) such that \(g_i^c(X) \geq (1 - \epsilon/l - \epsilon)c\) and satisfies \(\max_i w_i(X)/b_i \leq l\ln l/\epsilon\). The question is what we can say about the original objective. We can show that for all \(i = 1: l, g_i^c(X) \geq (1 - \epsilon - \epsilon/l)c\) since suppose this were not the case then there would exist atleast one \(i\) such that \(g_i^c(X) < (1 - \epsilon - \epsilon/l)c\).

Since \(g_i^c \leq c, \forall j = 1: l\), this implies that \(\sum_{j=1}^l g_j^c/l < c(l - 1)/l + (1 - \epsilon - \epsilon/l)c < c(1 - \epsilon/l - \epsilon)c\) which refutes the fact that \(g_i^c(X) = \sum_{j=1}^l g_j^c/l \geq (1 - \epsilon/l - \epsilon)c\). Define a new \(epsilon\) as \(\epsilon + \epsilon/l\) and this shows that we can achieve a discrete solution \(X\) such that \(g_i^c(X) < (1 - \epsilon)c\) and \(\max_i w_i(X)/b_i \leq l\ln l/\epsilon\). We can then the binary search over the values of \(c\) (similar to [3]) and obtain a \((1 - \epsilon, O(\ln l/\epsilon))\) bicriteria approximation. \(\square\)

4 Robust SCSC and Robust SCSK (Problems 3 and 4)

We first start by showing that \textsc{Robust-SCSC} and \textsc{Robust-SCSK} are closely related and an approximation algorithm for one of the problems provides an approximation algorithm for the other. We shall call these two problems duals of each other. Proofs of all results are in the extended version.
Theorem 8. We can obtain a \(((1 - \epsilon)\rho, \sigma)\) approximation for ROBUST-SCSK using 
\[
\log_{\min_i g_i(V) / \min_j g_j(j)} \text{ calls to an } (\sigma, \rho) \text{ bicriteria approximation algorithm for ROBUST-SCSC.}
\]
Conversely, we can achieve a \(((1 + \epsilon)\sigma, \rho)\) bicriteria approximation for ROBUST-SCSC with 
\[
\log_{\max_i f_i(V) / \max_j \min_i f_i(j)} \text{ calls to a } (\rho, \sigma) \text{ bicriteria approximation for ROBUST-SCSC.}
\]

Proof. To prove this result, we transform Problems 3 and 4 into the following problems:
\[
\min \max_{X \subseteq V} f_i(X) \mid \min_{i=1:k} g_i(X) / c_i \geq 1,
\]
\[
\max \min_{X \subseteq V} g_i(X) \mid \max_{i=1:k} f_i(X) / b_i \leq 1,
\]
Immediately notice that this is similar to SCSC and SCSK except that the objectives \(\max_i f_i(X)\) and \(\min_i g_i(X)\) are not submodular. They are however monotone and we can use algorithms 2 and 3 (for a linear search) and algorithms 4 and 5 (for a binary search version) from [20] to convert the problems from one form into another. This result follows from Theorem 3.1 and Theorem 3.2 from [20].

Next we discuss the hardness of Problems 3 and 4. Since Problem 3 generalizes ROBUST-SUBMIN with cardinality, knapsack and spanning tree constraints \((g)\) is the rank of the spanning tree matroid). This provides a hardness of \(\Omega(\max\{K(\sqrt{n}, \kappa), \log^{1-\epsilon} n\})\) for any \(\epsilon > 0\). The hardness of \(K(\sqrt{n}, \kappa)\) since it generalized SCSC and \(\log^{1-\epsilon} n\) from generalizing the case when \(f_i\)'s are modular and \(g\) is the rank of a spanning tree matroid [28]. Problem 4 generalizes ROBUST-SUBMAX which means one cannot achieve any polynomial approximation factor unless \(P = NP\). Since Problems 3 and 4 can be converted into one another, this means Problem 4 also has the same bi-criteria hardness of \(\Omega(\max\{K(\sqrt{n}, \kappa), \log^{1-\epsilon} n\})\) for any \(\epsilon > 0\).

Average Approximation: We start with the average case approximation of \(f_{\max}(X) = \max_i f_i(X)\) as \(f_{\text{avg}} = 1/l \sum_{i=1}^l f_i(X)\) and use \(f_{\text{avg}}\) instead of the \(f_{\max}(X)\) in Problems 3 and 4. We start with Problem 3. Using the saturate trick [25], we notice that the constraint is equivalent to \(\sum_{i=1}^k \min(g_i(X), c_i) \geq \sum_i c_i\), which immediately converts ROBUST-SCSC into an instance of SCSC. Next, we can use the Majorization-Minimization algorithm on SCSC (which we call MMin-AA) or the Ellipsoidal Approximation (which we call EA-AA) to approximate \(f_{\text{avg}}\). We can similarly solve ROBUST-AA. In the case of ROBUST-SCSK, we consider two cases. If we use the majorization-minimization scheme to approximate \(f_{\text{avg}}\), at every iteration of the majorization-minimization, we get an instance of ROBUST-SUBMAX subject to a single knapsack constraint. In the case of the Ellipsoidal Approximation, we use Theorem 8 and use the approximation algorithm for ROBUST-SCSC to provide the result.

Theorem 9. MMin-AA achieves an approximation factor of \(1/l \sum_{i=1}^l g_i(j)\) for Problem 3 and a bicriteria factor of \((1 - \epsilon) H(\max_i g_i(j))\) for Problem 4. Similarly, EA-AA achieves an approximation factor of \(O(1/l \sum_{i=1}^l g_i(j))\) for Problem 3 and a bicriteria factor of \((1 - \epsilon) O(1/l \sum_{i=1}^l g_i(j))\) for Problem 4.

Proof. In this proof, we elaborate a little bit on the summary above. We rewrite the optimization problem 3 and 4 using the average approximation:

Problem 3': \[
\min_{X \subseteq V} 1/l \sum_{i=1}^l f_i(X) \mid g_i(\mathcal{X}) \geq c_i, i = 1, \ldots, k
\]

Problem 4': \[
\max_{X \subseteq V} \min_{i=1:k} g_i(X) \mid 1/l \sum_{i=1}^l f_i(X) / b_i \leq 1
\]

Note that for Problem 3, we approximate \(\max_i f_i(X)\) with \(\frac{1}{l} \sum_{i=1}^l f_i(X)\), while for Problem 4 we approximate \(\max_i f_i(X) / b_i\) with \(\frac{1}{l} \sum_{i=1}^l f_i(X) / b_i\). Now, observe that the constraints in Problem 3' can be written as \(\sum_{i=1}^k \min(g_i(X), c_i) = \sum_{i=1}^k c_i\). Problem 3 then becomes an instance of SCSC. We can use the majorization-minimization algorithm (called ISSC in [20]) which

\[\text{Note that Problem 3 does admit approximation guarantees while Problem 4 can only admit bicriteria guarantees due to this result.}\]

\[\text{Note that the harmonic function}\]
achieves a guarantee of $K(n, \kappa_{avg})H(\max_j \sum_i g_i(j))$ for Problem 3’. Finally, we note that $f_{avg}$ is an $l$-approximation of $f_{max}$ which provides us with the guarantee for MMin. Next, we can also use EA for Problem 3’ and use the algorithm EASSC from [20]. This provides us with a tighter guarantee of $O(K(\sqrt{n}, \kappa_{avg})H(\max_j \sum_i g_i(j)))$ and again using the fact that $f_{avg}$ is a $l$-approximation of $f_{max}$ achieves the guarantee. Next, we consider Problem 4’. For Problem 4’, we can use the majorization-minimization (super-gradient) based upper bounds for the function $\frac{1}{l} \sum_{i=1}^{l} f_i(X)/b_i$. This then converts Problem 4’ into an instance of ROBUST-SUBMAX subject to a knapsack constraints. This achieves a $(1-\epsilon, O(l(\log l(\epsilon))))$ approximation [3]. Furthermore, we know that $\frac{1}{l} \sum_{i=1}^{l} f_i(X)/b_i$ is a $l$-approximation to $\max_{i=1:l} f_i(X)/b_i$ and we use the fact that $\max_{i=1:l} f_i(X)/b_i \leq l(\frac{1}{l} \sum_{i=1}^{l} f_i(X)/b_i) \leq O(l(\log l(\epsilon)))$. This proves the result. Finally, for the EA approximation for Problem 4, we use the fact that the EA-AA algorithm achieves a $O(K(\sqrt{n}, \kappa_{avg})H(\max_j \sum_i g_i(j)))$-approximation for Problem 3 and using Theorem 8 we obtain a bound for Problem 4.

Majorization-Minimization: Next, we directly apply majorization-minimization via the modular upper bounds (super-gradients) to Problems 3 and 4. We then obtain a sequential procedure where at every iteration, we have solve Problems 3 and 4 with $m_{f_{ji}}^{l_{i}}$ instead of $f_i$. We analyze Problem 4 since it is easier. At every iteration, we then have ROBUST-SUBMAX with multiple knapsack constraints. Invoking Theorem 7, we then get the corresponding approximation guarantee. To obtain an approximation guarantee for Problem 3, we invoke Theorem 8 to transfer over the bound.

**Theorem 10.** Majorization-Minimization (MMin) achieves a bicriteria approximation of $(K(n, \kappa_{wuc}) \log l/\epsilon (1+\epsilon), 1-\epsilon)$ for Problem 3 and a factor of $(1-\epsilon, K(n, \kappa_{wuc}) \log l/\epsilon)$ for Problem 4.

We get rid of the dependency on $l$ by directly applying Majorization-Minimization. However, we have a slightly weaker dependency on $n$ compared to the EA based schemes.

**Proof.** To prove this result, we start with Problem 4. We apply the majorization-minimization on the $f_i$’s by sequentially creating modular upper bounds starting at the empty set. Each sub-problem in the MMin becomes an instance of ROBUST-SUBMAX subject to multiple knapsack constraints. We can use Theorem 7 which provides a $(1-\epsilon, O(\ln l/\epsilon))$ approximation. Now the modular upper bounds of $f_i$ satisfies $f_i(X) \leq m_{f_{ji}}^{l_{i}}(X) \leq K(n, \kappa_{wuc}) f_i(X)$. At every iteration, majorization-minimization solves the following optimization problem (starting with $X^0 = \emptyset$):

$$\max_X \min_i g_i(X) \mid m_{f_{ji}}^{l_{i}}(X)/K(n, \kappa_{wuc}) \leq b_i$$

We now do the analysis for the first iteration. Observe that we can obtain a set $\hat{X}$ such that $m_{f_{ji}}^{l_{i}}(X)/K(n, \kappa_{wuc}) \leq b_i \ln l/\epsilon$. This implies that $m_{f_{ji}}^{l_{i}}(X) \leq K(n, \kappa_{wuc}) b_i \ln l/\epsilon$ and correspondingly, $f_i(X) \leq K(n, \kappa_{wuc}) b_i \ln l/\epsilon$. The first iteration of MMin achieves a bi-criteria approximation factor of $(1-\epsilon, K(n, \kappa) \log l/\epsilon)$ for Problem 4. Since subsequent iterations will only improve the objective value, the result holds. To prove the bound for Problem 3, we use Theorem 8 to transform the above bound into a bicriteria approximation for Problem 3.

Ellipsoidal Approximation: Finally, we approximate the $f_i$’s with the Ellipsoidal Approximation $\hat{f}_i(X) = \sqrt{w_{f_{ji}}^l(X)}$. Again for simplicity, we start with Problem 4. The constraints here are $\sqrt{w_{f_{ji}}^l(X)} \leq b_i$ which is equivalent to $w_{f_{ji}}^l(X) \leq b_i^2$. We then have ROBUST-SUBMAX subject to multiple knapsack constraints and using Theorem 7 we can achieve the following result.

**Theorem 11.** The Ellipsoidal Approximation Algorithm (EA) achieves a bicriteria approximation of $(O(\sqrt{n} \log n \sqrt{\log l/\epsilon}, 1-\epsilon)$ for Problem 3 and a factor of $(1-\epsilon, O(\sqrt{n} \log n \sqrt{\log l/\epsilon}))$ for Problem 4.

EA matches the hardness upto log factors when the curvature $\kappa = 0$.

**Theorem 12.** We start with Problem 4. The Ellipsoidal Approximation $\sqrt{w_{f_{ji}}^l(X)}$ approximates the functions $f_i(X)$ such that $\forall i, \sqrt{w_{f_{ji}}^l(X)} \leq f_i(X) \leq O(\sqrt{n} \log n) \sqrt{w_{f_{ji}}^l(X)}$. We then replace the functions $f_i$’s in Problem 4 with the Ellipsoidal Approximation:

$$\max_X \min_i g_i(X) \mid \sqrt{w_{f_{ji}}^l(X)} \leq b_i$$

(6)
This is equivalent to the constraints \( w_j(X) \leq b_j^2 \). From Theorem 7 we can obtain a set \( \hat{X} \) such that \( \min_i g_i(\hat{X}) \leq (1 - \epsilon) \min_i g_i(X^*) \) and \( w_j(X) \leq b_j^2 \log l/\epsilon \), which implies that \( \sqrt{w_j(X)} \leq b_j \sqrt{\log l/\epsilon} \). Finally note that \( f_i(X) \leq O(\sqrt{n} \log n) \sqrt{w_j(X)} \leq O(\sqrt{n} \log n \sqrt{\log l/\epsilon}) b_j \) which then proves the result. To achieve an approximation factor for Problem 3, we use Theorem 8 to obtain a bicriteria approximation for Problem 3 as well.

5 Experimental Results

Synthetic Experiments: The first set of experiments are synthetic experiments. We define \( f_j(X) = \sum_{i=1}^{c_j} \sqrt{w(X \cap C_{ij})} \) for a clustering \( C_j = \{C_1, C_2, \ldots, C_{c_j}\} \). We define \( l \) different random clusterings (with \( l = 3 \) and \( l = 10 \)). We choose the vector \( w \) at random with \( w \in [0, 1]^n \) and \( n = 50 \). We compare the different algorithms under cardinality constraints \( |X| \leq 10 \). The results are shown in Figure 1 (left) and are over 20 runs of random choices in \( w \) and \( C \)’s. We observe first that as expected, the AA versions of MMin and EA don’t perform as well since it optimizes the average case instead of the worst case. Directly optimizing the worst case performs much better. Next, we observe that MMin performs comparably to EA though its a simpler algorithm (a fact which has been noticed in several other scenarios as well [23, 20, 25].

Co-operative Matchings: Next, we compare ROBUST-SUBMIN in co-operative cuts. We follow the experimental setup in [21]. We run this on the House Dataset [6]. The results in Figure 1 (middle) are with respect to a simple modular (additive) baseline where the image correspondence problem becomes an assignment problem. We compare two baselines. The first is SubMod [21] which uses a single submodular function. The second is ROBUST-SUBMIN where we define several functions (over different clusterings of the pixels in the two images) and we define a robust objective. We run our experiments with \( l = 10 \) different clustering each obtained by different random initializations of k-means. The class of functions is exactly the same as in [21]. We see that the robust technique outperforms a single submodular function. In this experiment, we consider all pairs of images with the difference in image numbers being 20, 40, 60 and 80 (The \( x \)-axis in Figure 1 (center) – this is similar to the setting in [21]).

Limited Vocabulary Speech Data Selection: Here, we follow the experimental setup from [41]. The function \( f \) is the vocabulary function, \( f(X) = \gamma(X) \) [41]. Here, we define several different functions \( f_i \) (we set \( l = 10 \)). To define each function, we randomly delete 20% of the words from the vocabulary and define \( f_i(X) = \gamma_i(X) \). The function \( g \) is the diversity function and we use the feature based function from [41], except that we define \( k = 10 \) different functions \( g_i \) each defined via small perturbations of the feature values. The functions \( g_i \) are normalized so they are all between [0, 1]. The intuition of selecting \( f_i \)’s and \( g \)'s in this way is to be robust to perturbations in features and vocabulary words. We run these experiments on TIMIT [12] and we restrict ourselves to a ground-set size of \( |V| = 100 \). Again, we compare the different algorithms and arrive at similar conclusions to the synthetic experiments. Firstly, the average case algorithms MMin-AA and EA-AA don’t perform as well as MMin and EA themselves. Moreover, MMin and EA perform comparatively even though EA is orders of magnitude slower.

6 Conclusions

In this paper, we study four classes of robust submodular optimization problems: ROBUST-SUBMIN, ROBUST-SUBMAX, ROBUST-SCSC and ROBUST-SCSK. We study approximation algorithms and hardness results in each case. We propose a scalable majorization-minimization algorithm which has near optimal approximation bounds and also scales well. In future work, we would like to address the gap between the hardness and approximation bounds, and achieve tight curvature-based bounds in each case. We would also like to study other settings and formulations of robust optimization in future work.
References

[1] H. Aissi, C. Bazgan, and D. Vanderpooten. Min–max and min–max regret versions of combinatorial optimization problems: A survey. *European journal of operational research*, 197(2):427–438, 2009.

[2] H. Aissi, C. Bazgan, and D. Vanderpooten. General approximation schemes for min–max (regret) versions of some (pseudo-) polynomial problems. *Discrete Optimization*, 7(3):136–148, 2010.

[3] N. Anari, N. Haghtalab, S. Pokutta, M. Singh, A. Torrico, et al. Robust submodular maximization: Offline and online algorithms. *In Proc. AISTATS*, 2019.

[4] I. Bogunovic, S. Mitrović, J. Scarlett, and V. Cevher. Robust submodular maximization: A non-uniform partitioning approach. *arXiv preprint arXiv:1706.04918*, 2017.

[5] N. Buchbinder, M. Feldman, J. Naor, and R. Schwartz. A tight (1/2) linear-time approximation to unconstrained submodular maximization. *In FOCS*, 2012.

[6] T. S. Caetano, J. J. McAuley, L. Cheng, Q. V. Le, and A. J. Smola. Learning graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(6):1048–1058, 2009.

[7] C. Chekuri, J. Vondrák, and R. Zenklusen. Submodular function maximization via the multilinear relaxation and contention resolution schemes. *STOC*, 2011.

[8] R. S. Chen, B. Lucier, Y. Singer, and V. Syrgkanis. Robust optimization for non-convex objectives. In *Advances in Neural Information Processing Systems*, pages 4705–4714, 2017.

[9] J. Edmonds. Submodular functions, matroids and certain polyhedra. *Combinatorial structures and their Applications*, 1970.

[10] M. Fisher, G. Nemhauser, and L. Wolsey. An analysis of approximations for maximizing submodular set functions—ii. *Polyhedral combinatorics*, pages 73–87, 1978.

[11] S. Fujishige. *Submodular functions and optimization*, volume 58. Elsevier Science, 2005.

[12] J. Garofolo, F. Lamel, L., J. W., Fiscus, D. Pallet, and N. Dahlgren. Timit, acoustic-phonetic continuous speech corpus. In *DARPA*, 1993.

[13] G. Goel, C. Karande, P. Tripathi, and L. Wang. Approximability of combinatorial problems with multi-agent submodular cost functions. *In FOCS*, 2009.

[14] G. Goel, P. Tripathi, and L. Wang. Combinatorial problems with discounted price functions in multi-agent systems. *In FSTTCS*, 2010.

[15] M. Goemans, N. Harvey, S. Iwata, and V. Mirrokni. Approximating submodular functions everywhere. In *SODA*, pages 535–544, 2009.

[16] M. Gygli, H. Grabner, and L. Van Gool. Video summarization by learning submodular mixtures of objectives. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3090–3098, 2015.

[17] S. Iwata and K. Nagano. Submodular function minimization under covering constraints. *In FOCS*, pages 671–680. IEEE, 2009.

[18] R. Iyer and J. Bilmes. Algorithms for approximate minimization of the difference between submodular functions, with applications. *In UAI*, 2012.

[19] R. Iyer and J. Bilmes. The submodular Bregman and Lovász-Bregman divergences with applications. *In NIPS*, 2012.

[20] R. Iyer and J. Bilmes. Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints. *In NIPS*, 2013.

[21] R. Iyer and J. Bilmes. Near optimal algorithms for hard submodular programs with discounted cooperative costs. *In Proc. AISTATS*, 2019.

[22] R. Iyer, S. Jegelka, and J. Bilmes. Curvature and Optimal Algorithms for Learning and Minimizing Submodular Functions. *In Neural Information Processing Society (NIPS)*, 2013.

[23] R. Iyer, S. Jegelka, and J. Bilmes. Fast Semidifferential based submodular function optimization. *In ICML*, 2013.

[24] R. Iyer, S. Jegelka, and J. Bilmes. Monotone closure of relaxed constraints in submodular optimization: Connections between minimization and maximization: Extended version. *In UAI*, 2014.

[25] S. Jegelka and J. A. Bilmes. Approximation bounds for inference using cooperative cuts. *In ICML*, 2011.

[26] S. Jegelka and J. A. Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. *In CVPR*, 2011.
[27] A. Kasperski and P. Zieliński. On the approximability of minmax (regret) network optimization problems. *Information Processing Letters*, 109(5):262–266, 2009.

[28] A. Kasperski and P. Zieliński. On the approximability of robust spanning tree problems. *Theoretical Computer Science*, 414(4-5):365–374, 2011.

[29] A. Kasperski and P. Zieliński. Robust discrete optimization under discrete and interval uncertainty: A survey. In *Robustness analysis in decision aiding, optimization, and analytics*, pages 113–143. Springer, 2016.

[30] A. Kasperski and P. Zieliński. Approximating some network problems with scenarios. *arXiv preprint arXiv:1806.08936*, 2018.

[31] V. Kaushal, R. Iyer, S. Kothawade, R. Mahadev, K. Doctor, and G. Ramakrishnan. Learning from less data: A unified data subset selection and active learning framework for computer vision. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1289–1299. IEEE, 2019.

[32] V. Kaushal, R. Iyer, S. Kothawade, S. Subramanian, and G. Ramakrishnan. A framework towards domain specific video summarization. In *WACV*, 2019.

[33] K. Kirchhoff and J. Bilmes. Submodularity for data selection in machine translation. In *Empirical Methods in Natural Language Processing (EMNLP)*, October 2014.

[34] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos. Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management*, 134(6):516–526, 2008.

[35] A. Krause, B. McMahan, C. Guestrin, and A. Gupta. Robust submodular observation selection. *Journal of Machine Learning Research (JMLR)*, 9:2761–2801, 2008.

[36] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *JMLR*, 9:235–284, 2008.

[37] A. Kulik, H. Shachnai, and T. Tamir. Maximizing submodular set functions subject to multiple linear constraints. In *SODA*, 2009.

[38] H. Lin and J. Bilmes. A class of submodular functions for document summarization. In *ACL*, 2011.

[39] H. Lin and J. Bilmes. Optimal selection of limited vocabulary speech corpora. In *Interspeech*, 2011.

[40] H. Lin and J. Bilmes. Learning mixtures of submodular shells with application to document summarization. In *UAI*, 2012.

[41] Y. Liu, R. Iyer, K. Kirchhoff, and J. Bilmes. Switchboard-ii and fisver-i: Crafting high quality and low complexity conversational english speech corpora using submodular function optimization. *Computer Speech & Language*, 42:122–142, 2017.

[42] L. Lovász. Submodular functions and convexity. *Mathematical Programming*, 1983.

[43] S. Mittal and A. S. Schulz. A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one. *Operations Research*, 61(2):386–397, 2013.

[44] M. Narasimhan and J. Bilmes. A submodular-supermodular procedure with applications to discriminative structure learning. In *UAI*, 2005.

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

[46] J. B. Orlin, A. S. Schulz, and R. Udwani. Robust monotone submodular function maximization. *Mathematical Programming*, 172(1-2):505–537, 2018.

[47] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 86–92. IEEE, 2000.

[48] T. Powers, D. W. Krout, J. Bilmes, and L. Atlas. Constrained robust submodular sensor selection with application to multistatic sonar arrays. *IET Radar, Sonar & Navigation*, 11(12):1776–1781, 2017.

[49] M. Staib, B. Wilder, and S. Jegelka. Distributionally robust submodular maximization. *arXiv preprint arXiv:1802.05249*, 2018.

[50] Z. Svitkina and L. Fleischer. Submodular approximation: Sampling-based algorithms and lower bounds. In *FOCS*, pages 697–706, 2008.

[51] S. Tschiatschek, R. Iyer, H. Wei, and J. Bilmes. Learning mixtures of submodular functions for image collection summarization. In *Neural Information Processing Society (NIPS)*, Montreal, CA, December 2014.
[52] K. Wei, R. Iyer, and J. Bilmes. Submodularity in data subset selection and active learning. In International Conference on Machine Learning, pages 1954–1963, 2015.

[53] K. Wei, Y. Liu, K. Kirchhoff, and J. Bilmes. Using document summarization techniques for speech data subset selection. In NAACL-HLT, 2013.

[54] B. Wilder. Equilibrium computation for zero sum games with submodular structure. arXiv preprint arXiv:1710.00996, 2017.

[55] L. A. Wolsey. An analysis of the greedy algorithm for the submodular set covering problem. Combinatorica, 2(4):385–393, 1982.