Check for updates

Generalized Center Problems with Outliers

DEEPARNAB CHAKRABARTY and MARYAM NEGAHBANI, Dartmouth College, USA

We study the $\mathscr{F}$-center problem with outliers: Given a metric space (X, d), a general down-closed family $\mathscr{F}$ of subsets of X, and a parameter m, we need to locate a subset $S \in \mathscr{F}$ of centers such that the maximum distance among the closest m points in X to S is minimized. Our main result is a *dichotomy theorem*. Colloquially, we prove that there is an efficient 3-approximation for the $\mathscr{F}$-center problem with outliers if and only if we can efficiently optimize a *poly-bounded* linear function over $\mathscr{F}$ subject to a partition constraint. One concrete upshot of our result is a polynomial time 3-approximation for the knapsack center problem with outliers for which no (true) approximation algorithm was known.

CCS Concepts: • **Theory of computation → Facility location and clustering**;

Additional Key Words and Phrases: Approximation algorithms, clustering, k-center problem

ACM Reference format:
Deeparnab Chakrabarty and Maryam Negahbani. 2019. Generalized Center Problems with Outliers. *ACM Trans. Algorithms* 15, 3, Article 41 (July 2019), 14 pages.
https://doi.org/10.1145/3338513

41

1 INTRODUCTION

The k-center problem is a classic discrete optimization problem with numerous applications. Given a metric space (X, d) and a positive integer k, the objective is to choose a subset $S \subseteq X$ of at most k points such that $\max_{v \in X} d(v, S)$ is minimized, where $d(v, S) = \min_{u \in S} d(v, u)$. Informally, the problem is to open k centers to serve all points, minimizing the maximum distance to service. This problem has been studied for at least 50 years (Hakimi 1964, 1965), is NP-hard to approximate to a factor better than 2 (Hsu and Nemhauser 1979), and has a simple 2-approximation algorithm (Gonzalez 1985; Hochbaum and Shmoys 1985).

In many applications one is interested in a nuanced version of the problem where instead of serving all points in X, the objective is to serve at least a certain number of points. This is the so-called k-center with outliers version, or the *robust k-center* problem. This problem was first studied by Charikar et al. (2001), who give a 3-approximation for the problem. A best possible 2-approximation algorithm was recently given in Chakrabarty et al. (2016) (see also the paper by Harris et al. (2017)).

Another generalization of the k-center problem arises when the location of centers has more restrictions. For instance, if each point in X has a different weight and the constraint is that the total weight of centers opened is at most k. This problem, now called the *knapsack center* problem, was

An extended abstract appeared in the Proceedings of ICALP 2018.
Authors' addresses: D. Chakrabarty and M. Negahbani, Dartmouth College, 9 Maynard St. Hanover, NH, 03755, USA; emails: deeparnab@dartmouth.edu, maryam@cs.dartmouth.edu.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2019 Association for Computing Machinery.
1549-6325/2019/07-ART41 $15.00
https://doi.org/10.1145/3338513

studied by Hochbaum and Shmoys in Hochbaum and Shmoys (1986), who give a 3-approximation for the problem. To take another instance, X could be vectors in high dimension and the centers picked need to be linearly independent vectors. This motivates the *matroid center* problem, where the set of centers must be an independent set in a matroid. Chen et al. give a 3-approximation for this problem (Chen et al. 2013).

Naturally, the two aforementioned generalizations can be taken together. Indeed, for the *robust matroid center* problem, that is, the problem of picking centers that are an independent set and only m points need to be served, there is a 7-approximation algorithm in Chen et al. (2013). This was recently improved to a 3-approximation in Harris et al. (2017). The *robust knapsack center* problem, however, has had no non-trivial approximation algorithm till this work. Both Chen et al. (2013) and Harris et al. (2017) give *bi-criteria* 3-approximation algorithms, which violate the knapsack constraint by $(1+\varepsilon)$ (the running time of their algorithm is exponential in $1/\varepsilon$).

Our Contributions. Motivated by the state-of-affairs of the robust knapsack center problem, we study a broad generalization of the problems mentioned above. Let $\mathscr{F}$ be a general down-closed[1] family of subsets over X. In the *robust $\mathscr{F}$-center* problem, we are given a metric space (X,d), a parameter m, and the objective is to select a subset $S \in \mathscr{F}$ such that $\min_{T\subseteq X, |T|=m} \max_{v\in T} d(v,S)$ is minimized. That is, the maximum distance of service of the closest m points is minimized.

Observe that if $\mathscr{F} := \{A : w(A) \leq k\}$, then we get the robust knapsack center problem, and if $\mathscr{F}$ is the collection of independent sets of a matroid, then we get the robust matroid center problem. But this generalization captures a host of other problems. For instance, one can consider multiple (but constant) knapsack constraints. Indeed, this was studied in both Hochbaum and Shmoys (1986) and Chen et al. (2013). The former[2] only looks at the version *without* outliers and gives a polynomial time 3-approximation in the case when the weights are all polynomially bounded. The latter proves that when the weights are not polynomially bounded, there can be no approximation algorithm via a reduction to the Subset Sum problem, and gives a 3-approximation violating each knapsack constraint by at most $(1+\varepsilon)$ multiplicative factor.

Another instance is a single knapsack constraint along with a single matroid constraint. To our knowledge, this problem has not been studied earlier even in the case when outliers are not allowed. This problem seems natural: for instance, when the points are high-dimensional vectors with weights and the collection of centers needs to be a linearly independent set with total weight at most k.

The complexity of the robust $\mathscr{F}$-center problem naturally depends on the complexity of $\mathscr{F}$. To understand this, we define the following optimization problem, which depends only on the set-system $(X,\mathscr{F})$. We call it the *$\mathscr{F}$-partition constraint maximization* or simply $\mathscr{F}$-PCM. In this problem, one is given an arbitrary partition $\mathscr{P}$ of X along with $\mathscr{F}$, and a *poly-bounded* (the range is at most a polynomial in $|X|$) value $\mathsf{val}(x)$ on each $x \in X$. The objective is to find a set $S \in \mathscr{F}$ maximizing $\mathsf{val}(S) := \sum_{x\in S} \mathsf{val}(x)$, such that S contains at most one element from each part of $\mathscr{P}$. Our main result stated colloquially (and formally stated as Theorem 1 and Theorem 2 in Section 2) is the following dichotomy theorem.[3]

[1]If $A \in \mathscr{F}$ and $B \subseteq A$, then $B \in \mathscr{F}$.

[2]The complete proofs can be found in the STOC 1984 version of Hochbaum and Shmoys (1986)

[3]We are deliberately being inaccurate here. We should state the theorem for the more general *supplier* version where the set X is partitioned into $F \cup C$ and only the points in C need to be covered and only the centers in F can be opened. Being more general, the algorithmic results are therefore stronger. However, we were not able (and did not try too hard) to make our hardness go through for the center version. In the Introduction, we stick with the center version and switch to the supplier in the more formal subsequent sections.

Table 1. All the Above Results Can Be Obtained as Corollary or Simple Extensions to Our Main Result

The constraint system $\mathscr{F}$	**Without Outliers**	**Robust (With Outliers)**
Knapsack	3 (Hochbaum and Shmoys 1986)	**3** (Theorem 12)
Matroid	3 (Chen et al. 2013)	3 (Harris et al. 2017)
Multiple Knapsack (poly-bounded weights)	3 (Hochbaum and Shmoys 1986)	**3** (Theorem 13)
Knapsack and Matroid	**3** (Theorem 18)	**3** in special case (Theorem 16)
Multiple Knapsacks and Matroid	No uni-criteria approximation	**3**, $(1+\varepsilon)$ violating (Theorem 21)

The numbers in bold indicate new results.

Informal Theorem. *For any down-closed family $(X, \mathscr{F})$, the robust $\mathscr{F}$-center problem has an efficient 3-approximation algorithm if the $\mathscr{F}$-PCM problem can be solved in polynomial time. Otherwise, there is no efficient* non-trivial *approximation algorithm for the robust $\mathscr{F}$-center problem.*

Note that, in general, we are not concerned about how $\mathscr{F}$ is represented, because the only place the algorithm checks if a set S is in $\mathscr{F}$ is perhaps for solving the $\mathscr{F}$-PCM problem. So one can choose a representation that works best for the $\mathscr{F}$-PCM solver.

A series of corollaries follow from the above theorem. These are summarized in Table 1.

- When $\mathscr{F} = \{A : w(A) \leq k\}$, the $\mathscr{F}$-PCM problem can be solved in polynomial time via dynamic programming. This crucially uses that the val is poly-bounded. Therefore, we get a 3-approximation for the robust knapsack center problem (Theorem 12).
- When $\mathscr{F}$ is the independent set of a matroid, then the $\mathscr{F}$-PCM problem is a matroid intersection problem. Therefore, we get a 3-approximation for the robust matroid center problem recovering the result from Harris et al. (2017) (Theorem 14).
- When $\mathscr{F} = \{A : w_1(A) \leq k_1, w_2(A) \leq k_2, \ldots, w_d(A) \leq k_d\}$ is defined by d weight functions *and* each weight function w_i is *poly-bounded*, then $\mathscr{F}$-PCM can be solved efficiently using dynamic programming. Therefore, we get a 3-approximation algorithm for the robust multi-knapsack center problem, extending the result in Hochbaum and Shmoys (1986) to the case with outliers (Theorem 13).
- When $\mathscr{F}$ is given by the intersection of a single knapsack and a single matroid constraint, then we do not know the complexity. However, when the weight function $w(\cdot)$ is poly-bounded and the matroid is representable, then we can give a *randomized* algorithm for the $\mathscr{F}$-PCM problem via a reduction to the exact matroid intersection problem. Therefore, we get a randomized 3-approximation for this special case of robust knapsack-and-matroid center problem (Theorem 16).

Remark 1: The Zero Outlier Case. At this juncture, the reader may wonder about the complexity of the $\mathscr{F}$-center problem, which does not allow any outliers. This is related to the following decision problem. Given $(X, \mathscr{F})$ and an arbitrary sub-partition $\mathscr{P}$ of X, the problem asks whether there is a set $S \in \mathscr{F}$ such that S contains *exactly* one element from each part of $\mathscr{P}$. We call this the *$\mathscr{F}$-partition constraint feasibility* or simply the $\mathscr{F}$-PCF problem. Analogous to the informal theorem from earlier, the $\mathscr{F}$-center problem (without outliers) has an efficient 3-approximation algorithm if the $\mathscr{F}$-PCF problem can be solved efficiently; otherwise, the $\mathscr{F}$-center problem has no non-trivial approximation algorithm. Indeed, this theorem is much simpler to prove and arguably the roots of this lie in Hochbaum and Shmoys (1986).

This raises the main open question from our article: *What is the relation between the $\mathscr{F}$-PCF and the $\mathscr{F}$-PCM problem?* Clearly, the $\mathscr{F}$-PCF problem is as easy as the $\mathscr{F}$-PCM problem (set all values equal to one in the latter). But is there an $\mathscr{F}$ such that $\mathscr{F}$-PCM is "hard" while $\mathscr{F}$-PCF is "easy"? One concrete example is the corollary discussed in the last bullet point above. When $\mathscr{F}$ is a single knapsack constraint and a single matroid constraint, then the $\mathscr{F}$-PCF problem is solvable in polynomial time by minimizing a linear function over a matroid polytope and another partition matroid *base* polytope. As noted above, we do not know the complexity of the $\mathscr{F}$-PCM problem in this case.

Remark 2: Handling Approximations. If the $\mathscr{F}$-PCM problem is NP-hard, then the robust $\mathscr{F}$-center has no non-trivial approximation algorithm. However, approximation algorithms for $\mathscr{F}$-PCM translate to bi-criteria approximation algorithms for the robust $\mathscr{F}$-center problem. More precisely, if we have a ρ-approximation for the $\mathscr{F}$-PCM problem ($\rho \leq 1$), then we get a $(3, \rho)$-*bi-criteria* approximation algorithm for the robust $\mathscr{F}$-center problem. That is, we return a solution $S \in \mathscr{F}$ such that the maximum distance among the closest $\rho \cdot m$ points is at most three times the optimum value.

There could be a different notion of approximation possible for the $\mathscr{F}$-PCM problem. Given an instance, there may be an algorithm that returns a set S whose value is at least the optimum value but $S \in \mathscr{F}^R$ for some $\mathscr{F}^R \supseteq \mathscr{F}$, which is a "relaxation" of $\mathscr{F}$. For instance, if $\mathscr{F}$ is the intersection of multiple (constant) knapsack constraints that are not poly-bounded, then for any constant $\varepsilon > 0$ the $\mathscr{F}$-PCM problem can be solved (Chekuri et al. 2011; Grandoni et al. 2014), returning a set with value at least the optimum but violating each constraint by multiplicative $(1 + \varepsilon)$. We can use the same to get a polynomial time 3-approximation for the robust multiple knapsack-center problem if we are allowed to violate the knapsack constraints by $(1 + \varepsilon)$.

Our Technique. Although our theorem statement is quite general, the proof is quite easy. Let us begin with the $\mathscr{F}$-center problem without outliers. For this, we follow the algorithmic "partitioning" idea outlined by Hochbaum and Shmoys (1986). As is standard, we guess the optimum distance, which we assume to be 1 by scaling. Initially, all points are marked uncovered. Subsequently, we pick *any* uncovered point x and consider a subset B_x of points within distance 1 from it. Note that the optimum solution *must* pick at least one point from each B_x to serve x. Next, we call x "responsible" for all uncovered points within a distance 2 from it, and mark all these points covered. Observe that all the newly covered points are within distance 3 from *any* point in B_x. We continue the above procedure till all points are marked covered. Also observe that the B_x's form a sub-partition $\mathscr{P}$ of the universe where each part has a responsible point. By the above two observations, we see that the $\mathscr{F}$-PCF problem must have a feasible solution with respect to $\mathscr{P}$, and any solution to the $\mathscr{F}$-PCF problem gives a 3-approximation to the $\mathscr{F}$-center problem.

Handling outliers is a bit trickier. The above argument does not work since the "responsible" point may be an outlier in the optimal solution, and we can no longer assert that the optimal solution must contain a point from each part. Indeed, the nub of the problem seems to be figuring out which points should be outliers. The 3-approximation algorithm in Charikar et al. (2001) (see also Aggarwal et al. (2010)) cleverly chooses the partitioning via a greedy procedure, but their argument seems hard to generalize to other constraints.

A different attack used in the algorithm by Chakrabarty et al. (2016) and that by Harris et al. (2017) is by writing an LP relaxation and using the solution of the LP to recognize the outliers. At a high level, the LP assigns each point x a variable (in this article, we call it $\mathrm{cov}(x)$) that indicates the extent to which x is served. Subsequently, the partitioning procedure described in the first paragraph is run, except the responsible points are considered in decreasing order of $\mathrm{cov}(x)$. The hope is that points assigned higher $\mathrm{cov}(x)$ in the LP solution are less likely to be outliers, and therefore the partition returned by the procedure can be used to recover a 3-approximate solution. This idea does work for the natural LP relaxation of the robust matroid center problem but fails for the natural LP relaxation of the robust knapsack center problem. Indeed, the latter has unbounded integrality gap.

Our solution is to use the round-or-cut framework that has recently been a powerful tool in designing many approximation algorithms (see An et al. (2014); Carr et al. (2000); Chakrabarty et al. (2017); Li (2015, 2016)). We consider the following "coverage polytope" for the robust $\mathcal{F}$-center problem: the variables are $\mathrm{cov}(x)$ denoting the extent to which x is covered by a convex combination of sets $S \in \mathcal{F}$. Of course, we cannot hope to efficiently check whether a particular cov lies in this polytope. Nevertheless, we show that for any cov in the coverage polytope, the partitioning procedure when run in the decreasing order of cov has the property that there *exists* a solution $S \in \mathcal{F}$ intersecting each part at most once, which covers at least m points. We can then use the algorithm for $\mathcal{F}$-PCM to find this set. Furthermore, and more crucially, if the partitioning procedure does not have this property, then we can efficiently find a *hyperplane separating* cov from the the coverage polytope. Therefore, we can run the ellipsoid algorithm on the coverage polytope each time either obtaining a separating hyperplane, or obtaining a cov that leads to a desired partition, and therefore a 3-approximation.

2 PRELIMINARIES

In this section, we give formal definitions and statements of our results. As mentioned in a footnote in the Introduction, we focus on the supplier version of the problem.

Definition 1 ($\mathcal{F}$-Supplier Problem). The input is a metric space (X, d) on a set of points $X = F \cup C$ with distance function $d : X \times X \longrightarrow \mathbb{R}_{\geq 0}$ and $\mathcal{F} \subseteq 2^F$ a down-closed family of subsets of F. The objective is to find $S \in \mathcal{F}$ such that $\max_{v \in C} d(v, S)$ is minimized.

Definition 2 (Robust $\mathcal{F}$-Supplier Problem). The input is an instance of the $\mathcal{F}$-supplier problem along with an integer parameter $m \in \{0, 1, \ldots, |C|\}$. The objective is to find $S \in \mathcal{F}$ and $T \subseteq C$ for which $|T| \geq m$, and $\max_{u \in T} d(u, S)$ is minimized.

Thus, an instance $\mathcal{I}$ of the robust $\mathcal{F}$-supplier problem is defined by the tuple $(F, C, d, m, \mathcal{F})$. In the definitions above, F and C are often called the set of *facilities* and *customers*, respectively.

Given the set system $\mathcal{F}$ defined over F, we define the following optimization problem.

Definition 3 ($\mathcal{F}$-PCM problem). The input is $\mathcal{J} = (F, \mathcal{F}, \mathcal{P}, \mathsf{val})$ where F is a finite set and $\mathcal{F} \subseteq 2^F$ is a down-closed family, $\mathcal{P} \subseteq 2^F$ is a sub-partition of F, and $\mathsf{val} : F \longrightarrow \{0, 1, 2, \ldots\}$ is an integer-valued function with maximum range $|\mathsf{val}|$ satisfying: $\forall f_1, f_2 \in A \in \mathcal{P}$, $\mathsf{val}(f_1) = \mathsf{val}(f_2)$. The objective is to find

$$\mathrm{opt}(\mathcal{J}) = \max_{S \in \mathcal{F}} \ \mathsf{val}(S) : \ \ |S \cap A| \leq 1, \ \forall A \in \mathcal{P}.$$

The next theorem is the main result of the article.

Theorem 1. *Given a Robust $\mathcal{F}$-Supplier instance $\mathcal{I} = (F, C, d, m, \mathcal{F})$, Let $\mathcal{A}$ be an algorithm that solves any $\mathcal{F}$-PCM instance $\mathcal{J} = (F, \mathcal{F}, \mathcal{P}, \mathsf{val})$, with $|\mathsf{val}| \leq |C|$, in time bounded by $T_{\mathcal{A}}(\mathcal{J})$. Then, there is a 3-approximation algorithm for the Robust $\mathcal{F}$-Supplier instance that runs in time $\mathrm{poly}(|\mathcal{I}|)T_{\mathcal{A}}(\mathcal{J})$.*

The next theorem is the (easier) second part of the dichotomy theorem. We show that if $\mathcal{F}$-PCM cannot be solved, then the corresponding Robust $\mathcal{F}$-Supplier cannot be approximated.

Theorem 2. *Given any non-trivial approximation algorithm $\mathcal{B}$ for the Robust $\mathscr{F}$-Supplier problem that runs in time $T_{\mathcal{B}}(|\mathscr{I}|)$ on instance $\mathscr{I}$, any $\mathscr{F}$-PCM instance $\mathscr{J} = (F, \mathscr{F}, \mathscr{P}, \mathsf{val})$ can be solved in time* $\mathrm{poly}(|\mathscr{J}|)T_{\mathcal{B}}(|\mathscr{I}|)$*, where* $|\mathscr{I}| = \mathrm{poly}(|\mathscr{J}|)$.

Proof. Given $\mathscr{J}$, we construct an instance $\mathscr{I}$ of the Robust $\mathscr{F}$-Supplier problem. The set of facilities is F. We describe the set of customers C next. Extend $\mathscr{P}$ to a partition of F denoted by $\mathcal{Q} = \mathcal{P} \cup \{\{f\} : f \in F, \nexists A \in \mathcal{P} : f \in A\}$. By definition of the $\mathscr{F}$-PCM problem, for any $A \in \mathcal{Q}$, there exists a number $n_A \in \{0, 1, 2, \ldots\}$ such that $\mathsf{val}(f) = n_A$, for all $f \in A$. For each $A \in \mathcal{Q}$, we add n_A customers to C and call this set $\phi(A)$.

We now describe the distance function. For each $A \in \mathcal{Q}$, for each pair $u, v \in A$ and $u, v \in \phi(A)$, we have $d(u, v) = 0$. For each $u \in A$ and $v \in \phi(A)$, we have $d(u, v) = 1$. All other distances are ∞. Observe that d satisfies the triangle inequality.

Finally, we let m be our guess of the value of $\mathrm{opt}(\mathscr{J})$. This completes the description of $\mathscr{I} = (F, C, d, m, \mathscr{F})$.

Suppose algorithm $\mathcal{B}$ finds $S \in \mathscr{F}$ and $T \subseteq C$ such that $|T| \geq m$ and $\max_{v \in T} d(v, S) \leq \alpha \mathrm{opt}(\mathscr{I}) = \alpha$. Without loss of generality, we can assume $|S \cap A| \leq 1$ for all $A \in \mathscr{P}$, which implies that S is a feasible solution for $\mathscr{J}$. The reason is, if there exists $f_1, f_2 \in S$ for which $f_1, f_2 \in A \in \mathscr{P}$, then $S \backslash f_2$ is still an α-approximate solution for $\mathscr{I}$. To see why this is true, recall that $\mathscr{F}$ is down-closed so $S \backslash f_2 \in \mathscr{F}$, and since $d(f_1, f_2) = 0$ then $S \backslash f_2$ covers all the customers that S covers. Next, we assert that $\mathsf{val}(S) \geq m = \mathrm{opt}(\mathscr{J})$ since $m \leq |T| \leq |\{v \in C : d(v, S) \leq \alpha\}| = \sum_{A \in \mathcal{Q} : |S \cap A| = 1} |\phi(A)| = \sum_{f \in S} \mathsf{val}(f)$, where the first equality uses the fact that for $v \in C$ and $f \in A \in \mathcal{Q}$, $d(v, f) \leq \alpha$ only if $v \in \phi(A)$.

Finally, since val is poly-bounded, which makes the value of $\mathrm{opt}(\mathscr{J})$ to be bounded by $\mathrm{poly}(|\mathscr{J}|)$, one can iterate over all the possible values for $\mathrm{opt}(\mathscr{J})$ to guess m. □

We end this section by setting a few notations used in the remainder of the article. For any $u \in F \cup C$, we let $B_C(u, r)$ be the customers in a ball of radius r around u, i.e., $B_C(u, r) = \{v \in C : d(u, v) \leq r\}$. Similarly, define $B_F(u, r)$ as the facilities in a ball of radius r around u, i.e., for $u \in F \cup C$, $B_F(u, r) = \{f \in F : d(u, f) \leq r\}$.

3 ALGORITHM AND ANALYSIS : PROOF OF THEOREM 1

We fix $\mathscr{I} = (F, C, d, \mathscr{F}, m)$ as the instance of the Robust $\mathscr{F}$-Supplier problem. We use $\widehat{\mathrm{opt}}$ to denote *our guess* of the value of the optimal solution. Without loss of generality, we can always assume $\widehat{\mathrm{opt}} = 1$, because if not, then we could scale d to meet this criteria. Our objective henceforth is to either find a set $S \in \mathscr{F}$ such that $|\{v \in C : d(v, S) \leq 1\}| \geq m$, or prove that $\mathrm{opt}(\mathscr{I}) > 1$.

There are two parts to our proof. The first part is a partitioning procedure, which given an assignment $\mathrm{cov}(v) \in \mathbb{R}_{\geq 0}$ for every customer $v \in C$, constructs an instance $\mathscr{J}$ of $\mathscr{F}$-PCM. We call cov *valuable* if $\mathscr{J}$ has optimum value $\geq m$. Our procedure ensures that if cov is valuable, then we get a 3-approximate solution for $\mathscr{I}$. This is described in Section 3.1. The second part contains the proof of Theorem 1. In particular, we show how using the round-and-cut methodology using polynomially many calls to $\mathcal{A}$ (recall this is the algorithm for $\mathscr{F}$-PCM), we can either prove $\mathrm{opt}(\mathscr{I}) > 1$, or find a valuable cov. This is described in Section 3.2.

3.1 Reduction to $\mathscr{F}$-PCM

Algorithm 1 inputs an assignment $\{\mathrm{cov}(v) \in \mathbb{R}_{\geq 0} : v \in C\}$. It returns a sub-partition $\mathscr{P}$ of F and assigns $\mathsf{val} : F \to \{0, 1, \ldots, |C|\}$ such that all the facilities in the same part of $\mathscr{P}$ get the same val. That is, it returns an $\mathscr{F}$-PCM instance $\mathscr{J} = (F, \mathscr{F}, \mathscr{P}, \mathsf{val})$ with $|\mathsf{val}| \leq |C|$.

The algorithm maintains a set of *uncovered* customers $U \subseteq C$ initialized to C (Line 1). In each iteration, it picks the customer $v \in U$ with maximum cov (Line 5) and adds it to set $\mathrm{Reps}_{\mathrm{cov}}$ (Line 6).

We add the set of facilities $B_F(v,1)$ at distance 1 from v to $\mathscr{P}$ (Lines 7 and 8). For each such v, we eke out the subset $\mathsf{Chld}(v) = B_C(v,2) \cap U$ of currently uncovered customers "represented" by v (Line 9). For every facility $f \in B_F(v,1)$, we define its *value* to be: $\mathsf{val}(f) = |\mathsf{Chld}(v)|$ (Line 10). At the end of the iteration, $\mathsf{Chld}(v)$ is removed from U (Line 11) and the loop continues till U becomes $\emptyset$. This way, the algorithm partitions C into $\{\mathsf{Chld}(v) : v \in \mathsf{Reps}_{\mathsf{cov}}\}$ (see fact(3)). Claim 5 shows that $\mathscr{P}$ is a sub-partition of F.

ALGORITHM 1: $\mathscr{F}$-PCM instance construction

Input: Robust $\mathscr{F}$-Supplier instance $(F, C, d, m, \mathscr{F})$ and assignment $\{\mathsf{cov}(v) \in \mathbb{R}_{\geq 0} : v \in C\}$
Output: $\mathscr{F}$-PCM instance $(F, \mathscr{F}, \mathcal{P}, \mathsf{val})$

1 $U \leftarrow C$ ▷ The set of uncovered customers
2 $\mathsf{Reps}_{\mathsf{cov}} \leftarrow \emptyset$ ▷ The set of representatives
3 $\mathcal{P} \leftarrow \emptyset$ ▷ The sub-partition of F that will be returned
4 **while** $U \neq \emptyset$ **do**
5 $\quad v \leftarrow \arg\max_{v \in U} \mathsf{cov}(v)$ ▷ The first customer in U in non-increasing cov order
6 $\quad \mathsf{Reps}_{\mathsf{cov}} \leftarrow \mathsf{Reps}_{\mathsf{cov}} \cup v$
7 $\quad B_F(v,1) \leftarrow \{f \in F : d(f,v) \leq 1\}$ ▷ Facilities that can cover v with a ball of radius 1
8 $\quad \mathcal{P} \leftarrow \mathcal{P} \cup B_F(v,1)$
9 $\quad \mathsf{Chld}(v) \leftarrow \{u \in U : d(u,v) \leq 2\}$ ▷ Equals to $B_C(v,2) \cap U$
10 $\quad \mathsf{val}(f) \leftarrow |\mathsf{Chld}(v)| \;\; \forall f \in B_F(v,1)$
11 $\quad U \leftarrow U \backslash \mathsf{Chld}(v)$
12 **end**

FACT 3. $\{\mathsf{Chld}(v) : v \in \mathsf{Reps}_{\mathsf{cov}}\}$ *is a partition of* C.

FACT 4. *For a* $v \in \mathsf{Reps}_{\mathsf{cov}}$ *and any* $u \in \mathsf{Chld}(v)$ *line 6 of the algorithm implies* $\mathsf{cov}(v) \geq \mathsf{cov}(u)$.

CLAIM 5. $\mathcal{P}$ *constructed by Algorithm 1 is a sub-partition of* F.

PROOF. By Line 11 of the algorithm, for each $u, v \in \mathsf{Reps}_{\mathsf{cov}}$, we have $d(u,v) > 2$ hence $B_F(u,1) \cap B_F(v,1) = \emptyset$ implying $\mathcal{P}$ is a sub-partition of F. □

CLAIM 6. *For each* $v \in \mathsf{Reps}_{\mathsf{cov}}$ *and* $f \in B_F(v,1)$, $\mathsf{Chld}(v) \subseteq B_C(f,3)$.

PROOF. For any $u \in \mathsf{Chld}(v)$, we have $d(u,v) \leq 2$ and since $d(f,v) \leq 1$, the fact that d is metric implies $d(f,u) \leq 3$. □

Definition 4. For $S \subseteq F$ let $R(S) = \{v \in \mathsf{Reps}_{\mathsf{cov}} : B_F(v,1) \cap S \neq \emptyset\}$, be the set of representative customers in $\mathsf{Reps}_{\mathsf{cov}}$ that are covered by balls of radius 1 around the facilities in S.

CLAIM 7. *Let* $S \in \mathscr{F}$ *be any feasible solution of the* $\mathscr{F}$*-PCM instance constructed by Algorithm 1. Then,* $\sum_{f \in S} \mathsf{val}(f) = \sum_{v \in R(S)} |\mathsf{Chld}(v)|$.

PROOF. For an $f \in S$, according to Line 10 of the algorithm, $\mathsf{val}(f) > 0$ only if $f \in B_F(v,1)$ for some $v \in \mathsf{Reps}_{\mathsf{cov}}$. Also, by definition of the $\mathscr{F}$-PCM problem, $|B_F(v,1) \cap S| \leq 1$ for any $v \in \mathsf{Reps}_{\mathsf{cov}}$. That is, there is exactly one $f \in B_F(v,1) \cap S$ for each $v \in R(S)$ and again by line 10, $\mathsf{val}(f) = |\mathsf{Chld}(v)|$. Summing this equality over all $v \in R(S)$ and the corresponding $f \in B_F(v,1) \cap S$ proves the claim. □

CLAIM 8. *Let* $\mathscr{I} = (F, C, d, m, \mathscr{F})$ *be a Robust* $\mathscr{F}$*-Supplier instance and let* $\mathsf{cov} : C \to \mathbb{R}_{\geq 0}$ *be a coverage function. Let* $\mathscr{J} = (F, \mathscr{F}, \mathscr{P}, \mathsf{val})$ *be the* $\mathscr{F}$*-PCM instance returned by Algorithm 1 on input*

$\mathscr{I}$ and cov*. Given any feasible solution S to $\mathscr{J}$, we can cover at least* val(S) *customers of C by opening radius 3-balls around each facility in S.*

Proof. By considering $R(S)$ from Definition 4, Claim 7 gives: $\sum_{v \in R(S)} |\mathsf{Chld}(v)| = \sum_{f \in S} \mathsf{val}(f)$. From Fact 3, we get that for all $u, v \in \mathsf{Reps}_{\mathsf{cov}}$, $\mathsf{Chld}(u) \cap \mathsf{Chld}(v) = \emptyset$. Thus, $|\bigcup_{v \in R(S)} \mathsf{Chld}(v)| = \sum_{v \in R(S)} |\mathsf{Chld}(v)| = \mathsf{val}(S)$. Furthermore, by Claim 6, $\{v \in C : d(v,S) \le 3\} \supseteq \bigcup_{u \in R(S)} \mathsf{Chld}(u)$ implying the size of the former is at least $\mathsf{val}(S)$, thus proving the lemma. □

The above claim motivates the following definition of *valuable* cov assignments, and the subsequent lemma.

Definition 5. An assignment $\{\mathsf{cov}(v) \in \mathbb{R}_{\ge 0} : v \in C\}$ is *valuable* with respect to a Robust $\mathscr{F}$-Supplier instance $\mathscr{I} = (F, C, d, m, \mathscr{F})$, iff $\mathsf{opt}(\mathscr{J}) \ge m$, where $\mathscr{J}$ is the $\mathscr{F}$-PCM instance returned by Algorithm 1 from $\mathscr{I}$ and cov.

Lemma 9. *Given an instance $\mathscr{I}$ of the Robust $\mathscr{F}$-Supplier problem with* $\mathsf{opt}(\mathscr{I}) = 1$*, and a valuable assignment* cov *with respect to it, we can obtain a 3-approximate solution in time* $\mathsf{poly}(|\mathscr{I}|) + T_{\mathcal{A}}(\mathscr{J})$ *where $\mathscr{J}$ is the instance constructed by Algorithm 1 from $\mathscr{I}$ and* cov*.*

Proof. Since cov is valuable, $\mathsf{opt}(\mathscr{J}) \ge m$. We use solver $\mathcal{A}$ to return an optimal solution $S \in \mathscr{F}$ with $\mathsf{val}(S) \ge m$. Claim 8 implies that S is a 3-approximate solution to $\mathscr{I}$. □

3.2 The Round and Cut Approach

If the guess $\widehat{\mathsf{opt}} = 1$ for $\mathscr{I} = (F, C, d, m, \mathscr{F})$ is at least $\mathsf{opt}(\mathscr{I})$, then the following polytope must be non-empty. To see this, if $S^* \in \mathscr{F}$ is the optimal solution to $\mathscr{I}$, then set $z_{S^*} := 1$ and $z_S := 0$ for $S \in \mathscr{F} \backslash S^*$:

$$
\begin{aligned}
\mathscr{P}^{\mathscr{I}}_{\mathsf{cov}} = \{(\mathsf{cov}(v) : v \in C) : \quad & \sum_{v \in C} \mathsf{cov}(v) \ge m && (\mathscr{P}^{\mathscr{I}}_{\mathsf{cov}}.1)\\
\forall v \in C,\ \mathsf{cov}(v) - & \sum_{S \in \mathscr{F} : d(v,S) \le 1} z_S = 0 && (\mathscr{P}^{\mathscr{I}}_{\mathsf{cov}}.2)\\
& \sum_{S \in \mathscr{F}} z_S = 1 && (\mathscr{P}^{\mathscr{I}}_{\mathsf{cov}}.3)\\
\forall S \in \mathscr{F}, \quad & z_S \ge 0\} && (\mathscr{P}^{\mathscr{I}}_{\mathsf{cov}}.4)
\end{aligned}
$$

The variables z_S for all $S \in \mathscr{F}$ represent a convex combination of a set of integral solutions, and for $v \in C$, $\mathsf{cov}(v)$ indicates how much v is covered by this (possibly fractional) solution. Even though $\mathscr{P}^{\mathscr{I}}_{\mathsf{cov}}$ has exponentially many auxiliary variables (z_S for all $S \in \mathscr{F}$), its dimension is still $|C|$. The following gives a family of valid inequalities for $\mathscr{P}^{\mathscr{I}}_{\mathsf{cov}}$ via Farkas lemma.

Lemma 10. *Let $\lambda(v) \in \mathbb{R}$ for every $v \in C$ be such that*

$$\sum_{\substack{v \in C : \\ d(v,S) \le 1}} \lambda(v) < m \quad \forall S \in \mathscr{F}. \tag{V1}$$

Then any $\mathsf{cov} \in \mathscr{P}^{\mathscr{I}}_{\mathsf{cov}}$ *satisfies*

$$\sum_{v \in C} \lambda(v)\mathsf{cov}(v) < m. \tag{V2}$$

Proof. Given $\text{cov} \in \mathscr{P}^{\mathscr{I}}_{\text{cov}}$, there exists $\{z_S : S \in \mathscr{F}\}$ such that together they satisfy ($\mathscr{P}^{\mathscr{I}}_{\text{cov}}$.1)–($\mathscr{P}^{\mathscr{I}}_{\text{cov}}$.4):

$$\sum_{v \in C} \lambda(v)\text{cov}(v) =_{(\mathscr{P}^{\mathscr{I}}_{\text{cov}}.2)} \sum_{v \in C} \lambda(v) \sum_{\substack{S \in \mathscr{F}: \\ d(v,S) \le 1}} z_S = \sum_{S \in \mathscr{F}} z_S \sum_{\substack{v \in C: \\ d(v,S) \le 1}} \lambda(v)$$

$$<_{(\text{V1}),(\mathscr{P}^{\mathscr{I}}_{\text{cov}}.4)} m \sum_{S \in \mathscr{F}} z_S =_{(\mathscr{P}^{\mathscr{I}}_{\text{cov}}.3)} m. \qquad \square$$

The next lemma shows that all cov's in $\mathscr{P}^{\mathscr{I}}_{\text{cov}}$ are valuable.

Lemma 11. *Suppose an assignment $\{\text{cov}(v) \in \mathbb{R}_{\ge 0} : v \in C\}$ is not valuable with respect to $\mathscr{I} = (F, C, d, m, \mathscr{F})$. Then there is a hyper-plane separating it from $\mathscr{P}^{\mathscr{I}}_{\text{cov}}$ that can be constructed in polynomial time.*

Proof. If $\sum_{v \in C} \text{cov}(v) < m$, then this inequality itself is a separating hyper-plane and we are done. So, we may assume $\sum_{v \in C} \text{cov}(v) \ge m$.

Let $\mathscr{J} = (F, \mathscr{F}, \mathscr{P}, \text{val})$ be the $\mathscr{F}$-PCM instance constructed by Algorithm 1 from $\mathscr{I}$ and cov. Fix $S \in \mathscr{F}$ and recall from Definition 4 that $R(S) = \{v \in \text{Reps}_{\text{cov}} : B_F(v, 1) \cap S \ne \emptyset\}$. Pick an arbitrary $T \subseteq S$ for which $|B_F(v, 1) \cap T| = 1$, for all $v \in R(S)$. Observe that by down-closedness of $\mathscr{F}$, we have $T \in \mathscr{F}$, which implies T is a feasible solution for $\mathscr{J}$, and since cov is not valuable $\text{val}(T) < m$. Furthermore, Claim 7 applied to T gives $\text{val}(T) = \sum_{v \in R(T)} |\text{Chld}(v)|$. Since $R(S) = R(T)$ and $|\text{Chld}(v)|$ is integer-valued, we get

$$\sum_{v \in R(S)} |\text{Chld}(v)| \le m - 1. \tag{1}$$

Define $\lambda(v)$ for $v \in C$ as

$$\lambda(v) = \begin{cases} |\text{Chld}(v)| & v \in \text{Reps}_{\text{cov}} \\ 0 & \text{for all other } v \in C \end{cases}.$$

Now observe that for any $S \in \mathscr{F}$,

$$\sum_{v \in C : d(v,S) \le 1} \lambda(v) = \sum_{v \in \text{Reps}_{\text{cov}} : d(v,S) \le 1} |\text{Chld}(v)| = \sum_{v \in R(S)} |\text{Chld}(v)| \le m - 1. \tag{2}$$

That is, $\lambda(v)$'s satisfy Equation (V1). Now, we prove Equation (V2) is not satisfied, thus it can be used to separate cov from $\mathscr{P}^{\mathscr{I}}_{\text{cov}}$:

$$\begin{aligned} \sum_{v \in C} \lambda(v)\text{cov}(v) &= \sum_{v \in \text{Reps}_{\text{cov}}} |\text{Chld}(v)|\text{cov}(v) = \sum_{v \in \text{Reps}_{\text{cov}}} \sum_{u \in \text{Chld}(v)} \text{cov}(v) \\ &\ge_{\text{Fact 4}} \sum_{v \in \text{Reps}_{\text{cov}}} \sum_{u \in \text{Chld}(v)} \text{cov}(u) =_{\text{Fact 3}} \sum_{v \in C} \text{cov}(v) \ge m. \end{aligned} \tag{3}$$

$\square$

Proof of Theorem 1. Given the guess $\widehat{\text{opt}}$, which is scaled to 1, we use the ellipsoid algorithm to check if $\mathscr{P}^{\mathscr{I}}_{\text{cov}}$ is empty or not. Whenever an ellipsoid asks if a given cov is in $\mathscr{P}^{\mathscr{I}}_{\text{cov}}$ or not, run Algorithm 1 for this given cov to construct the corresponding $\mathscr{F}$-PCM instance $\mathscr{J}$ and use algorithm $\mathscr{A}$, promised in the statement of Theorem 1, to solve it. If $\text{opt}(\mathscr{J}) \ge m$, then Lemma 9 implies that we have a 3-approximate solution. Otherwise, cov is not valuable, and we can use Lemma 11 to find a separating hyperplane. In polynomial time, either we get a $\text{cov} \in \mathscr{P}^{\mathscr{I}}_{\text{cov}}$, which by Lemma 11 has to be valuable, or we prove $\mathscr{P}^{\mathscr{I}}_{\text{cov}}$ is empty and we modify our $\widehat{\text{opt}}$ guess. For the correct guess, the latter case will not occur and we get a 3-approximate solution.

To bound the number of ellipsoid iterations, notice that we have two types of separating hyper-planes constructed in Lemma 11. One is of the form $\sum_{v \in C} \mathrm{cov}(v) \geq m$ and the other one is like $\sum_{v \in C} \lambda(v)\mathrm{cov}(v) < m$. Since $0 \leq m \leq |C|$ and for all $v \in C$, we have $0 \leq \mathrm{cov}(v) \leq 1$ and $0 \leq \lambda(v) \leq |C|$, each of our separating hyper-planes can be described in $O(|C| \log |C|)$ bits. Thus, the polyhedron described by our separating hyper-planes has facet complexity $O(|C| \log |C|)$ and according to Theorem 6.5.16 of Grötschel et al. (1993), ellipsoid terminates in $\mathrm{poly}(|C|)$ many rounds. □

4 APPLICATIONS AND EXTENSIONS

In this section, we elaborate on the applications and extensions stated in the Introduction. We begin with looking at specific instances of $\mathscr{F}$ that have been studied in the literature and some that have not.

Single and Multiple Knapsack Constraints. We look at

$$\mathscr{F}_{\mathsf{KN}} := \{S \subseteq F : \text{ for } i = 1, \ldots, d, \sum_{v \in S} w_i(v) \leq k_i\},$$

where there are d weight functions over F and k_i's are upper bounds on these weights. Of special interest is the case $d = 1$, in which we get the robust knapsack supplier problem also called the weighted k-supplier problem with outliers.

The $\mathscr{F}$-PCM problem for the above $\mathscr{F}_{\mathsf{KN}}$ has the following complexity: When $d = 1$, the problem can be solved in polynomial time. Indeed, given a partition $\mathscr{P}$, since $\mathsf{val}(u) = \mathsf{val}(v)$ for all v in the same part, any solution that picks a facility from a part $A \in \mathscr{P}$ may as well pick the one with the smallest weight in that part. Thus, the problem boils down to the usual knapsack problem, in which we have $|\mathscr{P}|$ items where the item corresponding to part $A \in \mathscr{P}$ has weight $\min_{v \in A} w(v)$ and value $\mathsf{val}(v)$. Since the values are poly-bounded, this problem is solvable in polynomial time. Thus, we get the following corollary to Theorem 1 resolving the open question raised in Chen et al. (2013) and Harris et al. (2017).

Theorem 12. *There is a polynomial time 3-approximation to the robust knapsack center problem.*

When $d > 1$, then the $\mathscr{F}$-PCM problem is NP-hard even when val is poly-bounded. However, if the w_i's are also poly-bounded (actually one of them can be general), then the $\mathscr{F}$-PCM problem can be solved in polynomial time using dynamic programming. This problem was in fact studied in Hochbaum and Shmoys (1986) (the conference version) and is called the *suitcase* problem there. Thus, we get the following corollary to Theorem 1 extending the result in Hochbaum and Shmoys (1986).

Theorem 13. *There is a polynomial time 3-approximation to the robust multiple-knapsack center problem if the number of weights is a constant and all but possibly one weight function are poly-bounded.*

Single and Multiple Matroid Constraints. We look at

$$\mathscr{F}_{\mathsf{Mat}} := \{S \subseteq F : S \in \mathscr{I}_{M_i}, \forall i = 1, \ldots, d\}.$$

When $d = 1$, we get the robust matroid center problem. The $\mathscr{F}$-PCM paper reduces to finding a maximum value set in $\mathscr{I}_M$ and a partition matroid induced by $\mathscr{P}$. This is solvable in polynomial time even when val is general and not poly-bounded, and even when $\mathscr{I}_M$ is given as an independent set oracle. Thus, we get the following corollary to Theorem 1 obtaining the result in Harris et al. (2017).

Theorem 14 [Theorem 1.1 in Harris et al. (2017)]. *There is a polynomial time 3-approximation to the robust matroid center problem even when the matroid is described as an independent set oracle.*

When there are $d > 1$ matroids, then the $\mathscr{F}$-PCM problem is NP-hard. Therefore, Theorem 2 implies that for instance, we can have *no* unicriteria approximation for the robust matroid-intersection center problem.

Single Knapsack and Single Matroid Constraint. We look at

$$\mathscr{F}_{\mathsf{KN}\cap\mathsf{Mat}} := \{S \subseteq F : \sum_{v \in S} w(v) \le k,\ S \in \mathscr{I}_M\},$$

which is the intersection of a single matroid and a single knapsack constraint. To the best of our knowledge, the resulting Robust $\mathscr{F}$-Supplier problem has not been studied before. One natural instantiation is when F is a collection of high-dimensional vectors with weights and the constraint on the centers is to pick a linearly independent set with total weight at most k.

The corresponding $\mathscr{F}$-PCM problem asks us, given a partition $\mathscr{P}$ and poly-bounded values val, to find a set $S \in \mathscr{I}_{\mathcal{M}} \cap \mathscr{I}_{\mathscr{P}}$ of maximum value such that $w(S) \le k$, where $\mathscr{I}_{\mathscr{P}}$ is the partition matroid induced by $\mathscr{P}$. We do not know if this problem can be solved in polynomial time, even in the case when M is another partition matroid.

However, the above problem is related to the *exact matroid intersection* problem. In this problem, we are given two matroids $\mathcal{M}$ and $\mathscr{P}$, and a weight function $\mathbf{w}$ on each ground element and a budget $\mathbf{W}$. The objective is to decide whether or not there is a set $S \in \mathscr{I}_{\mathcal{M}} \cap \mathscr{I}_P$ such that $\mathbf{w}(S) = \mathbf{W}$. Understanding the complexity of this problem is a long standing challenge (Camerini et al. 1992; Mulmuley et al. 1987; Papadimitriou and Yannakakis 1982). When the matroids are representable over the same field, then Camerini et al. (1992) gives a randomized pseudopolynomial time algorithm for the problem. The following claim shows the relation between $\mathscr{F}$-PCM and the exact matroid intersection problem; this claim is essentially present in Berger et al. (2011).

Claim 15. *Given an algorithm for the exact matroid intersection problem, one can solve the $\mathscr{F}$-PCM problem in polynomial time when the weights w are poly-bounded.*

Proof. We guess V^* to be the optimum value of the $\mathscr{F}$-PCM problem; since val is poly-bounded, there are only polynomially many guesses. We also guess $k^* \le k$ to be the total w-weight of the optimum set. Again if w is poly-bounded, there are polynomially many guesses. We define a weight function $\mathbf{w}$ as follows. Let $\phi = w(F) + 1$ be a large enough upper-bound on the possible values of $w(S), S \subseteq F$. Define $\mathbf{w}(f) = \phi\mathsf{val}(f) + w(f)$ for all $f \in F$ and $\mathbf{W} = \phi\mathsf{V}^* + k^*$.

We claim that there is a set S in $\mathscr{I}_M \cap \mathscr{I}_{\mathscr{P}}$ with $\mathbf{w}(S) = \mathbf{W}$ iff $\mathsf{val}(S) = \mathsf{V}^*$ and $w(S) = k^*$. The if-direction is trivial.

However, if $\mathbf{w}(S) = \mathbf{W}$, then we get $k^* = \mathbf{w}(S) - \phi\mathsf{V}^* = \phi\mathsf{val}(S) + w(S) - \phi V^*$. Now, if $\mathsf{val}(S) \neq \mathsf{V}^*$, since val is integer-valued and since $\phi > w(S)$ for any $S \subseteq F$, then the right-hand side is either negative or $> w(F)$. In any case, it cannot be k^*. Therefore, we must have $\mathsf{val}(S) = \mathsf{V}^*$, which implies $w(S) = k^*$. □

Armed with the non-trivial result about exact matroid intersection from Camerini et al. (1992), we get the following.

Theorem 16. *Given a linear matroid $\mathcal{M}$ and a poly-bounded weight function, there is a randomized polynomial time 3-approximation to the robust knapsack-and-matroid center problem.*

4.1 The Case of No Outliers

The $\mathscr{F}$-supplier problem, that is the case of $m = |C|$, may be of special interest. In this case the problem is easier and the complexity is defined by the complexity of the following decision problem.

Definition 6 ($\mathcal{F}$-PCF Problem). The input is $\mathcal{J} = (F, \mathcal{F}, \mathcal{P})$ where F is a finite set, $\mathcal{F} \subseteq 2^F$ is a down-closed family and $\mathcal{P} \subseteq 2^F$ is an arbitrary sub-partition of F. The objective is to decide whether there *exists* a set $S \in \mathcal{F}$ such that $|S \cap A| = 1, \ \forall A \in \mathcal{P}$.

Theorem 17. *If the $\mathcal{F}$-PCF problem can be solved in polynomial time for any partition $\mathcal{P}$, then the $\mathcal{F}$-supplier problem has a polynomial time 3-approximation. Otherwise, there is no non-trivial approximation possible for the $\mathcal{F}$-supplier problem.*

Sketch. Run Algorithm 1 with an arbitrary assignment cov (and ignore the val's). Let $\mathcal{J} = (F, \mathcal{F}, \mathcal{P})$ be the resulting $\mathcal{F}$-PCF instance. If the guess $\widehat{\mathrm{opt}} = 1$ is correct, then note that the optimum solution S^* must satisfy $S^* \cap A \neq \emptyset$ for all $A \in \mathcal{P}$; if not, then the corresponding $v \in \mathrm{Reps}_{\mathrm{cov}}$ cannot be served. Conversely, any S satisfying $S \cap A \neq \emptyset$ for all $A \in \mathcal{P}$ implies a 3-approximate solution. Therefore, an algorithm for $\mathcal{F}$-PCF can either give a 3-approximate solution or prove the guess $\widehat{\mathrm{opt}}$ is too low. □

Theorems 1 and 17 raise the question: Is there any set of constraints for which the problem without outliers is significantly easier than the problem with outliers? We do not know the answer to this question, although we guess the answer is yes. For this, it suffices to design a set system for which $\mathcal{F}$-PCF is easy but $\mathcal{F}$-PCM is hard (perhaps NP-hard). To see the difference between these problems consider the $\mathcal{F}_{\mathrm{KN} \cap \mathrm{Mat}}$ family described in the previous subsection. We do not know if $\mathcal{F}$-PCM is easy or hard, but $\mathcal{F}$-PCF is easy: this amounts to minimizing $w(S)$ over $S \in \mathcal{I}_{\mathcal{M}} \cap \mathcal{B}_{\mathcal{P}}$ where $\mathcal{B}_{\mathcal{P}}$ is the base polytope induced by $\mathcal{P}$. This is the weighted matroid intersection problem, which can be solved in polynomial time. Thus, we get the following corollary.

Theorem 18. *There is a polynomial time 3-approximation to the knapsack-and-matroid center problem.*

4.2 Handling Approximation

The technique used to prove Theorem 1 is robust enough to translate approximation algorithms for the $\mathcal{F}$-PCM problem to *bi-criteria* approximation algorithms for the Robust $\mathcal{F}$-Supplier problem. There are two notions of approximation algorithms for the $\mathcal{F}$-PCM problem and they lead to two notions of bi-criteria approximation.

The first is the standard notion: a ρ-approximation (for $\rho \leq 1$) algorithm that takes instance $\mathcal{J}$ of $\mathcal{F}$-PCM, returns a solution $S \in \mathcal{F}$ of value $\mathrm{val}(S) \geq \rho \cdot \mathrm{opt}(\mathcal{J})$. The corresponding bi-criteria approximation notion for the Robust $\mathcal{F}$-Supplier problem is the following: an (α, β)-approximation algorithm for instance $\mathcal{I}$ of Robust $\mathcal{F}$-Supplier returns a solution that opens centers at $S \in \mathcal{F}$ and the distance of at least βm customers to S is $\leq \alpha \cdot \mathrm{opt}(\mathcal{I})$. The proof of Theorem 1, in fact, implies the following.

Theorem 19. *Let $\mathcal{A}$ be a polynomial time ρ-approximate algorithm for the $\mathcal{F}$-PCM problem. Then there is a polynomial time $(3, \rho)$-bi-criteria approximation algorithm for the Robust $\mathcal{F}$-Supplier problem.*

Proof. The argument is analogous to the proof of Theorem 1. The only difference is after constructing the $\mathcal{F}$-PCM instance $\mathcal{J}$. Run algorithm $\mathcal{A}$ to construct a solution S for $\mathcal{J}$. If $\mathrm{opt}(\mathcal{J}) \geq m$, then S must have value at least m/ρ and according to Claim 8, at least m/ρ customers are at distance at most 3 of this S. Thus, S is a $(3, \rho)$-approximate solution and we are done. Otherwise, it must be that $\mathrm{opt}(\mathcal{J}) < m$ so cov is not valuable and we can use Lemma 11 to get a separating hyperplane. Note that the definition of *valuable* is still the same and proof of Lemma 11 is independent of S. □

The second notion of approximation for the $\mathscr{F}$-PCM problem is one which satisfies the constraints approximately. This notion is more problem dependent and makes sense only if there is a notion of an approximate relaxation $\mathscr{F}^R$ for the set $\mathscr{F}$. For example, an $(1+\varepsilon)$-relaxation for $\mathscr{F}_{\mathsf{KN}}$ could be the subsets S with $w_i(S) \leq (1+\varepsilon) \cdot k_i$ for all i. A ρ-violating algorithm for an instance $\mathscr{J}$ of $\mathscr{F}$-PCM would then return a set S with $\mathsf{val}(S) \geq \mathsf{opt}(\mathscr{J})$ but $S \in \mathscr{F}^R$, which is an ρ-relaxation for $\mathscr{F}$. This defines a different bi-criteria approximation notion for the Robust $\mathscr{F}$-Supplier problem. An α-approximate β-violating algorithm for the Robust $\mathscr{F}$-Supplier problem takes an instance $\mathscr{I}$ and returns a solution $S \in \mathscr{F}^R$, which is a β-relaxation for $\mathscr{F}$ such that at least m customers in C are at distance at most $\alpha \cdot \mathsf{opt}(\mathscr{I})$ to S.

Theorem 20. *Let $\mathcal{A}$ be a polynomial time ρ-violating algorithm for the $\mathscr{F}$-PCM problem. Then there is a polynomial time 3-approximate-ρ-violating algorithm for the Robust $\mathscr{F}$-Supplier problem.*

Proof. Again, the proof is analogous to the proof of Theorem 1. Here, the only change is the argument for determining if cov is valuable or not. After running $\mathcal{A}$ on the $\mathscr{F}$-PCM instance, we get $S \in \mathscr{F}^R$. Since $\mathscr{F} \subseteq \mathscr{F}^R$, if the value of S is less than m, then it means that $\mathsf{opt}(\mathscr{J}) < m$, so cov is not valuable and we get a separating hyper-plane by Lemma 11. Otherwise, by Claim 8, at least m customers are at distance at most 3 of S and it is a 3-approximate-ρ-violating solution. Recall that the rest of the proof (including the proof of Lemma 11) and the definition of *valuable* cov is the same as before. □

When $\mathscr{F}$ is described by constant d knapsack constraints (with general weights) and a single matroid constraint, for any constant $\varepsilon > 0$, Chekuri et al. give an $(1+\varepsilon)$-approximation algorithm for the $\mathscr{F}$-PCM (Chekuri et al. 2011). Without the matroid constraint, Grandoni et al. give an $(1+\varepsilon)$-violating algorithm (Grandoni et al. 2014). Together, we get the following corollary. The latter recovers a result from Chen et al. (2013).

Theorem 21. *Fix any constant $\varepsilon > 0$. There is a polynomial time $(3, (1+\varepsilon))$-bi-criteria approximation algorithm for the robust supplier problem with constant many knapsack constraints and one matroid constraint. There is a polynomial time 3-approximate $(1+\varepsilon)$-violating algorithm for the robust supplier problem with constant many knapsack constraints.*

REFERENCES

Gagan Aggarwal, Rina Panigrahy, Tomás Feder, Dilys Thomas, Krishnaram Kenthapadi, Samir Khuller, and An Zhu. 2010. Achieving anonymity via clustering. *ACM Trans. Algor.* 6, 3 (2010), 49:1–49:19. DOI: https://doi.org/10.1145/1798596.1798602

Hyung-Chan An, Mohit Singh, and Ola Svensson. 2014. LP-based algorithms for capacitated facility location. In *Proceedings of the IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS'14).* IEEE Computer Society, Washington, DC, 256–265. DOI: https://doi.org/10.1109/FOCS.2014.35

André Berger, Vincenzo Bonifaci, Fabrizio Grandoni, and Guido Schäfer. 2011. Budgeted matching and budgeted matroid intersection via the gasoline puzzle. *Math. Program.* 128, 1 (2011), 355–372. DOI: https://doi.org/10.1007/s10107-009-0307-4

Paolo M. Camerini, Giulia Galbiati, and Francesco Maffioli. 1992. Random pseudo-polynomial algorithms for exact matroid problems. *J. Algor.* 13, 2 (1992), 258–273. DOI: https://doi.org/10.1016/0196-6774(92)90018-8

Robert D. Carr, Lisa K. Fleischer, Vitus J. Leung, and Cynthia A. Phillips. 2000. Strengthening integrality gaps for capacitated network design and covering problems. In *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'00).* Society for Industrial and Applied Mathematics, Philadelphia, PA, 106–115. Retrieved from http://dl.acm.org/citation.cfm?id=338219.338241.

Deeparnab Chakrabarty, Prachi Goyal, and Ravishankar Krishnaswamy. 2016. The non-uniform k-center problem. In *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP'16).* 67:1–67:15. DOI: https://doi.org/10.4230/LIPIcs.ICALP.2016.67

Deeparnab Chakrabarty, Ravishankar Krishnaswamy, and Amit Kumar. 2017. The heterogeneous capacitated k-center problem. In *Proceedings of the 19th International Conference on Integer Programming and Combinatorial Optimization (IPCO'17).* 123–135. DOI: https://doi.org/10.1007/978-3-319-59250-3_11

Moses Charikar, Samir Khuller, David M. Mount, and Giri Narasimhan. 2001. Algorithms for facility location problems with outliers. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'01)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 642–651.

Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. 2011. Multi-budgeted matchings and matroid intersection via dependent rounding. *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*. 1080–1097. DOI: https://doi.org/10.1137/1.9781611973082.82

Danny Z. Chen, Jian Li, Hongyu Liang, and Haitao Wang. 2013. Matroid and knapsack center problems. In *Proceedings of the Conference on Integer Programming and Combinatorial Optimization (IPCO'13)*. 110–122.

Teofilo F. Gonzalez. 1985. Clustering to minimize the maximum intercluster distance. *Theoret. Comput. Sci.* 38 (1985), 293–306. DOI: https://doi.org/10.1016/0304-3975(85)90224-5

Fabrizio Grandoni, R. Ravi, Mohit Singh, and Rico Zenklusen. 2014. New approaches to multi-objective optimization. *Math. Program.* 146, 1–2 (2014), 525–554. DOI: https://doi.org/10.1007/s10107-013-0703-7

Martin Grötschel, László Lovász, and Alexander Schriver. 1993. *Geometric Algorithms and Combinatorial Optimization.* Springer, Berlin. DOI: https://doi.org/10.1007/978-3-642-78240-4

Seifollah L. Hakimi. 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. *Ops Res. INFORMS* 12 (1964), 450–459. DOI: https://doi.org/10.1287/opre.12.3.450

Seifollah L. Hakimi. 1965. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Ops Res. INFORMS* 13 (1965), 462–475. DOI: https://doi.org/10.1287/opre.13.3.462

David G. Harris, Thomas Pensyl, Aravind Srinivasan, and Khoa Trinh. 2017. A lottery model for center-type problems with outliers. In *Proceedings of the Conference on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM'17)*. 10:1–10:19. DOI: https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2017.10

Dorit S. Hochbaum and David B. Shmoys. 1985. A best possible heuristic for the k-center problem. *Math. Oper. Res.* 10, 2 (1985), 180–184. DOI: https://doi.org/10.1287/moor.10.2.180

Dorit S. Hochbaum and David B. Shmoys. 1986. A unified approach to approximation algorithms for bottleneck problems. *J. ACM* 33, 3 (1986), 533–550. DOI: https://doi.org/10.1145/5925.5933

Wen-Lian Hsu and George L. Nemhauser. 1979. Easy and hard bottleneck location problems. *Discrete Appl. Math.* 1, 3 (1979), 209–215. DOI: https://doi.org/10.1016/0166-218X(79)90044-1

Shi Li. 2015. On uniform capacitated k-median beyond the natural LP relaxation. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'15)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 696–707.

Shi Li. 2016. Approximating capacitated k-median with $(1+\varepsilon)k$ open facilities. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 786–796. Retrieved from http://dl.acm.org/citation.cfm?id=2884435.2884491.

Ketan Mulmuley, Umesh V. Vazirani, and Vijay V. Vazirani. 1987. Matching is as easy as matrix inversion. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC'87)*. ACM, New York, NY, 345–354. DOI: https://doi.org/10.1145/28395.383347

Christos H. Papadimitriou and Mihalis Yannakakis. 1982. The complexity of restricted spanning tree problems. *J. ACM* 29, 2 (Apr. 1982), 285–309. DOI: https://doi.org/10.1145/322307.322309

Received September 2018; revised April 2019; accepted May 2019