Approximation Algorithms for Robust Covering Problems with Chance Constraints

Vineet Goyal* R. Ravi*

Abstract

Two common approaches to model uncertainty in optimization problems are to either explicitly enumerate all the realizations of the uncertain parameters or to specify them implicitly by a probability distribution over the uncertain parameters. We study robust covering problems where the demand or the right hand side of the covering constraints is uncertain, in both models of uncertainty and introduce *chance constraints* in the formulation. We consider problems with only one stage of decision making as well as a two-stage problem where there is a second recourse stage to augment the first stage decisions.

In a chance-constrained model, we are given a reliability parameter ρ (in addition to the other inputs as in the robust problem) and the goal is to construct a solution that satisfies at least a ρ fraction of the scenarios; the robust objective implies that the worst case cost over these scenarios is to be minimized. Here the parameter ρ achieves a trade-off between the reliability and the cost of the solution; when $\rho = 1$, we arrive back at the classical robust model. This model can thus be thought of as a method for automatic detection of outlier scenarios. As an example, given $\rho = \frac{k}{n}$ in an n-node graph, the one-stage chance constrained shortest path problem from a root with every node as a singleton scenario occurring with probability $\frac{1}{n}$ is to find a rooted k-MST.

While it is easy to obtain bi-criteria approximation algorithms for the chance-constrained problems that violate the chance-constraint by a small factor, we consider the problem of satisfying the chance-constraint strictly. We show that in the explicit scenario model (with more than one element in all the scenarios), both one-stage and two-stage problems are at least as hard to approximate as the dense k-subgraph (DkS) problem. The hardness arises particularly due to the requirement that the chance-constraint be satisfied strictly and does not hold if we relax that condition. For the special case when each scenario has a single element, while the one-stage problem directly reduces to a weighted partial covering problem, we show that many two-stage problems (including set cover, facility location etc) reduce to a weighted partial covering problem via a guess and prune method. We consider the two-stage shortest path problem which can not be reduced to a partial covering version and is closely related to the weighted k-MST problem where the weight function is submodular. We give an $O(\log k)$-approximation for this problem using a charging argument.

We also consider the model of uncertainty where scenarios (possibly an exponential number) are specified implicitly by a probability distribution. In particular, we consider an independent distribution where each demand occurs with a given probability independently of others. While it is not even clear if we can succinctly specify the ρ fraction of selected scenarios for the two-stage problem, we show that the one-stage problem in this model can be reduced to a weighted partial covering problem. We also extend these results to obtain a logarithmic approximation for the one-stage problem when the demand uncertainty is specified by a general probability distribution such that the cumulative probability of any demand-scenario can be computed efficiently and is *strictly-monotone* with respect to set inclusion.

*Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Supported in part NSF grant CCF-0430751.

1 Introduction

Optimization models incorporating data and demand uncertainty have long been studied in the literature due to their vast applicability in real world scenarios. In a classical optimization problem, all parameters such as demands and costs are assumed to be known precisely. However, this assumption does not hold for most real world applications. In such applications, the classical deterministic optimization approach is not useful because the solution found through such an approach is possibly sensitive to even slight changes in the problem parameters. Both one-stage and multi-stage stochastic and robust optimization models have been extensively studied to address this problem of uncertainty [14, 25, 5, 21, 20, 3, 4], with uncertainty being modeled as explicit scenarios or implicitly specified via a probability distribution over the uncertain parameters. In a one-stage model, a feasible solution for all the future scenarios needs to be constructed in a single stage, while in a multi-stage model, the first stage solution can be augmented with the recourse solution after the uncertainty has been realized (possibly in multiple stages). Stochastic optimization approaches optimize the expected costs over all scenarios while the robust optimization approaches optimize over the worst case scenario. However, both approaches are plagued by the presence of unlikely outlier scenarios which distort the optimization goals and the resulting solution.

A natural idea to overcome this problem is to prune away the outlier scenarios and solve the problem on remaining scenarios. This approach, referred to as *chance-constrained* optimization (see [10, 5]), has been studied in literature. A chance-constrained model incorporates probabilistic constraints in the traditional stochastic or robust optimization model. Thus, the problem of finding a minimum cost solution which is feasible for ρ fraction of the scenarios for a given reliability $\rho > 0$, can be modeled using chance constraints. This model is best introduced through an example: consider a one-stage shortest path problem on an undirected graph $G = (V, E)$ where we are required to construct a path between root r and an uncertain destination. Each vertex $v \in V$ occurs with probability $\frac{1}{n}$ as the destination and we are required to choose a minimum cost set of edges E_s such that with probability ρ (where ρ is given) there is a path between r and the destination. Note that the problem in this example reduces to finding a minimum cost rooted k-MST in G where $\frac{k}{n} \leq \rho < \frac{k+1}{n}$.

In a chance-constrained optimization approach, the parameter ρ captures the risk aversion of the optimizer. When $\rho = 1$, we return to the classical robust model, while at $\rho = 0$, the empty solution is feasible. In this paper, we extend the chance constrained framework to robust covering problems with demand-uncertainty (such as considered in Dhamdhere et al. [15]) in both one-stage and two-stage models where the demand-uncertainty is either given as an explicit list of scenarios or specified implicitly. (Our methods also apply directly to the stochastic versions defined e.g., in [24] but we leave the details out).

1.1 Previous Work

Chance constrained programming was introduced in the work of Charnes and Cooper [10]. Even with a long history, chance constraint models do not find wide applicability because of the inherent difficulty in solving these problems optimally; namely, the feasible region for a chance constrained problem depends on the underlying uncertainty and is generally non-convex. A detailed discussion of chance constrained programs, and more generally, stochastic programs can be found in [5]. Robust optimization and chance constrained optimization are very closely related (see [8, 12, 16]). Nemirovski and Shapiro [23] show how robust optimization framework provides an approximation of chance constrained programming while Chen et al. [12] propose robust optimization as a technique to obtain feasible solutions for chance constrained programs in [26].

For simple probability distributions, such as a uniform distribution (where each scenario occurs with the same probability), the chance-constrained problem reduces to a more familiar partial covering problem, where we are required to cover some k out of l scenarios with a minimum cost solution. Recall that the shortest path problem described above reduces to finding a minimum cost tree containing the root r that spans at least k vertices. This problem is a partial covering version of the spanning tree problem that has been studied extensively [6, 7, 13, 1] and for which a 2-approximation is known [19]. In general, in a partial set covering problem we are given a set family $\mathcal{F}$, set of elements U and a target $k \leq |U|$ and the goal is to select a minimum cost collection of sets from $\mathcal{F}$ that cover at least k elements. Partial covering versions of several combinatorial problems have been considered such as vertex cover [2, 18, 22], facility location, k-center [9]. However, to the best of our knowledge, there has not been any prior work in designing approximation algorithms for combinatorial problems in the general chance-constrained framework.

1.2 Our Contributions

We consider chance constraints in both one-stage as well as two-stage robust covering problems with demand-uncertainty where uncertainty is specified either as an explicit list of demand-scenarios or implicitly as a probability distribution over the demand elements that require coverage. While it is easy to obtain bi-criteria approximation algorithms for the chance-constrained problems that violate the chance constraint by a small factor (see Appendix A), we consider the problem of satisfying the chance constraint strictly.

1. We show that in the explicit scenario model (with more than one element in all the scenarios), both one-stage and two-stage problems are at least as hard to approximate as the dense k-subgraph (DkS) problem. The Dense k-Subgraph problem is conjectured to be $\Omega(n^{\delta})$-hard to approximate for some $\delta > 0$ [17].

2. For the special case when each scenario has a single element, while the one-stage problem directly reduces to a weighted partial covering problem, we show that many two-stage problems (including set cover, facility location etc) reduce to a weighted partial covering problem via a guess-and-prune method.

3. The two-stage shortest path problem does not reduce to a partial covering version but can be reduced to the weighted k-MST problem where the weight function is submodular. We give an $O(\log k)$-approximation for this problem.

<table>
<tr><th rowspan="2"></th><th colspan="3">Explicit Scenarios</th></tr>
<tr><th colspan="2">Scenarios with 1-elt</th><th>Scenarios with > 1 elts</th></tr>
<tr><td>One stage</td><td colspan="2">Reduces to partial covering</td><td>DkS-hard</td></tr>
<tr><td rowspan="2">Two stage</td><td>Set Cover, Vertex Cover, Facility Location</td><td>Reduce to partial covering</td><td rowspan="2">DkS-hard</td></tr>
<tr><td>Shortest Path</td><td>$O(\log k)$</td></tr>
</table>

Table 1: Main results for the explicit scenario uncertainty model

4. We also consider the model of uncertainty where scenarios (possibly an exponential number) are specified implicitly by a probability distribution. In particular, we consider a model where each demand occurs with a given probability independently of others referred from hereon as the *independent-scenarios* model. While it is not even clear if the two-stage problem in the independent-scenarios model is in NP, we show that the one-stage problem in this model can be reduced to a weighted partial covering problem. We also extend these results for the one-stage problem where the demand uncertainty is specified by a general probability distribution such that the *cumulative probability* of any demand-scenario can be computed efficiently and is *strictly-monotone* with respect to set inclusion.

Outline. The rest of the paper is organized as follows. In Section 2, we present the hardness of approximation of problems with more than one element per explicit scenario. In Section 3, we consider the explicit scenario model with only one element per scenario and present the reduction of the chance-constrained versions of many robust covering problems to weighted partial covering problems. Finally, in Section 4, we consider implicit models of uncertainty and show that the one-stage problems in the independent-scenario model reduce to weighted partial covering problems and also discuss extensions to the general distribution model.

2 Hardness of Approximation

We show that the one-stage chance constrained set cover problem in the explicit scenario model is at least as hard to approximate as Dense k-Subgraph even when every scenario has only two elements.

Problem Definition A one stage chance constrained set covering problem in the explicit scenario model (Explicit 1-CCSCP) is as follows: we are given a universe of elements U, a family of subsets $\mathcal{S}$, a cost function c on the subsets in $\mathcal{S}$, a list of l scenarios where scenario i is specified by a subset $S_i \subset U$ and its probability p_i, and a reliability factor $0 < \rho < 1$. The problem is to find a minimum cost partial set cover for elements in a subset of scenarios (say $\mathcal{I}$) such that $\sum_{i \in \mathcal{I}} p_i \geq \rho$.

We prove the following theorem.

Theorem 2.1 Explicit 1-CCSCP *is at least as hard to approximate as* Dense k-Subgraph *even when each scenario has only two elements.*

Proof: In a Dense k-Subgraph instance $\mathcal{I}$, we are given a graph $G = (V, E)$ and a number k, and the objective is to find a minimum size subset of vertices $V' \subseteq V$ that induces at least k edges, i.e $|E[V'] \geq k|$.

The reduction is as follows: we construct an instance $\mathcal{I}'$ of Explicit 1-CCSCP. The element set $U = \{v_i | v_i \in V\}$. For each vertex $v_i \in V$, we have a set $S_i = \{v_i\}$ in the set family $\mathcal{F}$. For each edge $e = (v_i, v_j) \in E$, we have a scenario containing two elements $\{v_i, v_j\}$. Now, in the instance $\mathcal{I}'$ of Explicit 1-CCSCP we are required to find a minimum cardinality subset $\mathcal{S}$ of sets from $\mathcal{F}$ such that the sets in $\mathcal{S}$ satisfy at least k scenarios. Note that a scenario, $\{v_i, v_j\}$ is satisfied by $\mathcal{S}$ if both v_i and v_j are contained in some sets in $\mathcal{S}$.

Suppose there is a solution $\mathcal{S}$ for instance I'. Consider $V' = \{v_i | S_i \in \mathcal{S}\}$. Consider any scenario $\{v_i, v_j\}$ that is satisfied by $\mathcal{S}$. Note that $(v_i, v_j) \in E(G)$ and $v_i, v_j \in V'$. Thus, (v_i, v_j) is an induced edge in V' which implies $|E[V']| \geq k$. Thus, $OPT(I) \leq OPT(I')$.

Conversely, consider a solution V' of I that induces at least k edges. Consider $\mathcal{S} = \{S_i | v_i \in V'\}$. It is easy to note that for each edge $(v_i, v_j) \in E[V']$, the corresponding scenario is satisfied by the solution $\mathcal{S}$. Thus, $OPT(I') \leq OPT(I)$. ∎

A direct reduction from the above gives the following.

Theorem 2.2 Explicit 2-CCSCP *is at least as hard to approximate as* Dense k-Subgraph *even when each scenario has only two elements.*

3 Explicit Scenario Models

Note that even one stage versions of covering problems with more than one element per scenario are hard. For instance, we have the following corollary of Theorem2.1.

Corollary 3.1 *The one-stage (and hence, two-stage) chance-constrained versions of the following covering problems in the explicit scenario model are at least as hard to approximate as* Dense k-Subgraph *even when each scenario has only two elements.*

1. *Vertex Cover (scenario is described by a subset of edges)*

2. *Facility Location (scenario is described by a subset of demand points)*

3. *K-median (scenario is described by a subset of demand points)*

4. *K-center (scenario is described by a subset of demand points)*

5. *Steiner Tree (scenario is described by a subset of vertices)*

Hence in this section, we consider the case when each scenario has exactly one element.

One-stage versions. The one stage versions of the above problems with exactly one element per scenario are directly reducible to the respective partial covering variants. These variants have been well approximated in the literature (2-approximation for partial vertex cover [22], 3-approximation for partial facility location, partial k-center [9], constant-factor for partial k-median [11], and 2-approximation for partial shortest paths that reduce to k-MST [19]), hence we focus on the two-stage version henceforth.

We first give a logarithmic approximation for the two-stage chance-constrained robust version for the general set covering problem. Then, we show how the two stage versions of the above problems (in particular, Vertex Cover, Facility Location and Steiner tree) with one element per scenario can be approximated.

3.1 Two-stage Chance-constrained Set Covering Problem

Theorem 3.2 *Consider the* Explicit 2-CCSCP *where you are given a family m subsets $S_1, \ldots, S_m$ with cost function c, and l scenarios such that scenario i contains element e_i, has inflation factor σ_i and occurs with probability p_i and required reliability of the solution is ρ. This problem can be reduced to a weighted partial covering solution and thus, admits an $O(\log(\rho l))$-approximation.*

Proof: Fix an optimum solution and suppose the worst case second stage cost is B in this optimum solution. There are only l choices for B; one corresponding to the second-stage minimum cost solution for each of the l scenarios. Let c_i denote the cost of the minimum-cost set that contains e_i, $T = \{i \in [l] | \sigma_i \cdot c_i \leq B\}$. We can cover all scenarios in T in the second-stage with cost at most B. Let $\tau = \sum_{i \in T} p_i$. We need to cover a subset of scenarios from $[l] \setminus T$ whose total probability is at least $\rho - \tau$ in the first stage. Therefore, for a particular choice of B the problem reduces to a weighted partial set covering problem which admits an $O(\log k)$-approximation if you require to cover k elements. ■

The reduction in the above theorem applies to the two-stage covering problems that satisfy the following property: *If a scenario i that is covered in an optimal solution can not be independently covered in the second-stage within the worst case second-stage cost, then it must be completely covered in the first stage.*

Two-stage Chance-constrained Vertex Cover. This problem when each scenario consists of a single edge satisfies the above property. Thus, it can be reduced to a weighted partial vertex cover problem, which implies a 2-approximation using the results of [22]. Corresponding versions of the facility location problem and the shortest path problem do not satisfy this property and thus, do not directly reduce to a partial covering problem.

3.2 Two Stage Chance-Constrained Facility Location Problem 2-CCFLP

Problem Definition Given a metric (V, d), a set of potential facilities $\mathcal{F}$ and a set of l scenarios where scenario i is specified by a demand point $v_i \in V$ and inflation factor σ_i and occurs with probability p_i and required reliability is ρ. Opening a facility $j \in \mathcal{F}$ in the first stage costs c_j while opening it in the second stage in scenario i costs $\sigma_i \cdot c_j$. The goal is to select a ρ fraction of the scenarios (say $\mathcal{I}$) and open a set of facilities F_1 to open in the first stage and for each of the selected scenario i, connect to one of the open facilities in F_1 or open a new facility and connect to it in the second stage if that scenario materializes. Let x_j be a binary variable denoting whether $j \in \mathcal{F}$ is opened in the first stage or not and let $f_i(x)$ denote the minimum second-stage cost in scenario i given the first stage solution is x. The objective is to minimize

$$\sum_{j \in \mathcal{F}} c_j x_j + \max_{i \in \mathcal{I}} f_i(x)$$

We reduce the above problem to a weighted partial covering problem and thus, give a 3-approximation for 2-CCFLP.

For the sake of simplicity, we assume that all scenarios occur with the same probability $p = \frac{1}{l}$; essentially the same algorithm and analysis extend to the general problem.

Theorem 3.3 *There is a* 3*-approximation for the* 2-CCFLP *with* l *scenarios where each scenario has only one element and the required reliability is* $\rho = \frac{k}{l}$*.*

Proof: Fix an optimum solution and suppose the first stage facility opening cost is C_1^* and the worst case second stage cost is C_2^* in this optimum solution. There are only $2l \cdot |\mathcal{F}|$ choices for C_2^*. Let $f_i(x)$ denote the minimum-cost solution for scenario i when the first stage solution is x. Let $T = \{i \in [l] | \sigma_i \cdot f_i(0) \leq C_2^*\}$. Note that computing $f_i(0)$ is easy: consider the minimum cost of opening (in the second stage) and connecting v_i to the open facility. We can cover all scenarios in T in the second-stage with cost at most C_2^*. Therefore, the first stage problem is to open a set of facilities such that for at least $k' = k - |T|$ scenarios from $[l] \setminus T$, there is an open facility within a distance $\alpha \cdot C_2^*$ from the demand-point for some approximation factor $\alpha > 0$. Note that there is a set of facilities of cost C_1^* such that for at least k' scenarios in $[l] \setminus T$, the demand-point is within a distance C_2^* of some open facility.

The first stage problem thus reduces to a version of the partial k-center problem considered in Charikar et al. [9] who give a 3-approximation for the problem. Therefore, we can find a set of facilities of cost at most C_1^* such that at least k' demand-scenarios are within a distance $3C_2^*$ of some open facility which gives a 3-approximation for 2-CCFLP. ■

3.3 Two Stage Chance Constrained Shortest Path Problem (2-CCSPP)

Problem Definition Given a graph $G = (V, E)$ with edge costs c, a root vertex r, a reliability level ρ and a list of l scenarios. Each scenario i is specified by a terminal t_i, an inflation factor σ_i and a probability p_i. The goal is to select a ρ fraction of the scenarios, buy some edges E_f in the first stage and for each selected scenario i, augment the first stage solution in the recourse stage with edges E_s^i (bought at an inflated cost) such that $E_f \cup E_s^i$ contains a path from r to t_i. The objective minimizes the worst case cost over all scenarios.

For the sake of simplicity, we consider the case of uniform probabilities and a uniform inflation factor across all scenarios. Thus, the reliability level ρ translates to covering $k = l\rho$ out of l terminals. However, it is not difficult to extend this algorithm and the analysis to general problem with different probabilities and inflation factors for different scenarios.

Using the structural theorem in Dhamdhere et al. [15], we obtain the following lemma,

Lemma 3.4 *For the uniform robust* 2-CCSPP *that requires to cover* k *out of* l *terminals, there exists a first stage solution* E_f *and a set* I *of* k *scenarios such that* E_f *is a tree containing* r *and can be augmented by* E_r^i *to obtain a feasible solution for scenarios in* I *and* $c(E_f) + \max_{i \in I} \sigma c(E_r^i) \leq 2\mathsf{OPT}$*, where* OPT*is the optimal solution for robust* 2-CCSPP*.*

Fix an optimal solution to the robust 2-CCSPP such that the first stage solution is connected to r, say $O = (O_f, O_r^1, \ldots, O_r^l)$ (some of the recourse edge sets may be empty). From Lemma 3.4, we know that $c(O_f) + \max_i \sigma c(O_r^i) \leq 2\mathsf{OPT}$. Let σC be the maximum second stage cost for any scenario in O. We can assume that σC is known (as there are only nl choices of C). Thus, the tree O_f is within a distance C from at least k of the l terminals(say $t_1^*, \ldots, t_k^*$).

Algorithm Consider ball B_i of radius $2C$ around terminal t_i. We select a maximal independent set $\mathcal{I}$ on balls $B_1, \ldots, B_l$ as follows:

1. Initialize $\mathcal{I} \leftarrow \phi$, $\mathcal{T} \leftarrow \{t_1, \ldots, t_l\}$.

2. while ($\mathcal{T} \neq \phi$), do

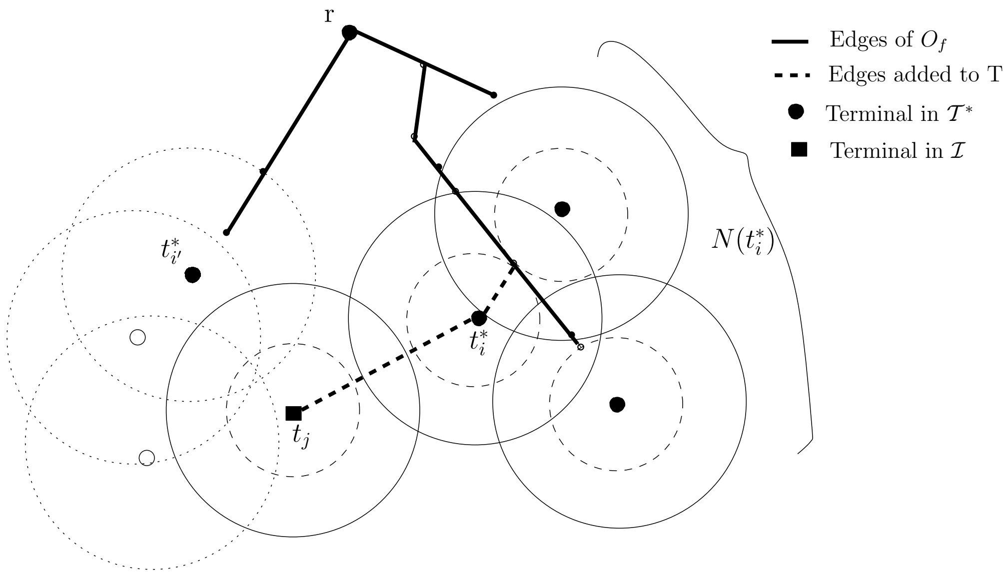

Figure 1: Constructing k-MST of cost $O(\mathsf{OPT})$ from O_f under our weight function

(a) Consider the terminal $t_i \in \mathcal{T}$ such that the number of terminals in $\mathcal{T}$ within a distance of at most $4C$ from t_i is maximum (resolve ties arbitrarily).

(b) Let $N(t_i)$=set of terminals in $\mathcal{T}$ that are within distance $4C$ from t_i (including itself) and let $w(t_i) = |N(t_i)|$.

(c) Add t_i to $\mathcal{I}$ and remove all terminals that are within distance $4C$ of t_i from $\mathcal{T}$.

Now, construct a minimum cost spanning tree, T_A containing r that spans terminals of weight at least k. Note that only terminals in the independent set $\mathcal{I}$ have a non-zero weight.

Lemma 3.5 $c(T_A) = O(\log k)c(O_f)$

Proof: Suppose $\mathcal{I}$ has terminals $t_1, \ldots, t_q$ (listed in the order they were added to $\mathcal{I}$). Note that $w(t_1) \geq w(t_2) \geq \ldots \geq w(t_q)$. Let $B(t_i^*)$ denote the ball of radius $2C$ around terminal t_i^* and $\mathcal{T}^* = \{t_1^*, \ldots, t_k^*\}$. Recall that t_i^* is a terminal within distance of at most C from O_f; therefore, $B(t_i^*)$ intersects with O_f. For the sake of argument, whenever an edge $e \in O_f$ crosses the ball $B(t_i^*)$ for any $t_i^* \in \mathcal{T}^*$, we introduce a new vertex at the point of intersection and edge e is subdivided into two. It is easy to note that $\sum_{e \in B(t_i^*) \cap O_f} c(e) \geq C$ for any $t_i^* \in \mathcal{T}^*$.

We will now construct a tree T from O_f that contains r and spans terminals in $\mathcal{I}$ of weight at least k. Initialize $T \leftarrow O_f$.

Consider the terminal $t_i^* \in \mathcal{T}^*$ such that the ball $B(t_i^*)$ intersects the ball $\hat{B}_j, j = 1, \ldots, q$ of highest weight. Let $N(t_i^*)$ denote the terminals in $\mathcal{T}^*$ that intersect with $B(t_i^*)$ (including itself) and let $n(t_i^*) = |N(t_i^*)|$. We claim that $n(t_i^*) \leq w(t_j)$. At the time t_j was added to $\mathcal{I}$, t_i^* was also a candidate. Furthermore, all the terminals in $N(t_i^*)$ were also candidates; otherwise one of them would intersect with $\hat{B}_j, j = 1, \ldots, q$ of higher weight contradicting our choice of t_i^*. Since, t_j was chosen in $\mathcal{I}$, $w(t_j) \geq n(t_i^*)$. Thus, the tree T can be extended to reach t_j by charging to the cost of edges in $O_f \cap B(t_i^*)$ since, we know $\sum_{e \in B(t_i^*) \cap O_f} c(e) \geq C$.

In doing so, we have updated the weight of T under our weight function by $w(t_j) \geq n(t_i^*)$. We update $\mathcal{T}^* \leftarrow \mathcal{T}^* \setminus (N(t_i^*) \cup N(t_j))$ and continue. Note that by updating the set terminals $\mathcal{T}^*$ by removing t_i^* and all other terminals in $N(t_i^*)$ we ensure that we do not charge to the same cost of $\mathsf{OPT}(O_f)$ in some other iteration.

Note that we might have removed $w(t_j)+n(t_i^*)$ terminals from $\mathcal{T}^*$ and added only $w(t_j) \geq 1/2(w(t_j+n(t_i^*))$ weight in T. Thus, we would obtain a tree T that has cost $O(c(O_f))$ and spans terminals of weight at least $k/2$.

We repeat this procedure on the remaining terminals of OPTthat are not covered in T. This implies that after $\log k$ rounds, we will obtain a tree T spanning terminals of cumulative weight at least k and $c(T) = O(\log k)c(O_f)$.

Thus, there exists a tree T containing r that spans a subset of terminals in $\mathcal{I} = \{t_1, \ldots, t_q\}$ whose cumulative weight is at least k and $c(T) = O(\log k)c(O_f)$. ■

Theorem 3.6 *There is an $O(\log k)$-approximation to robust* 2-CCSPP.

Proof: Let T_A be the first stage tree returned by the algorithm. It is easy to note that there are at least k terminals from $t_1, \ldots, t_l$ that are within a distance of $4C$ from T_A. Lemma 3.5 implies that $c(T_A) = O(\log k)c(O_f)$. Thus, the cost of the solution returned by our algorithm is $(O(\log k)c(O_f) + 4\sigma C) = O(\log k)\mathsf{OPT}$. ■

We would like to remark here that finding an approximate first stage tree that reaches within a distance C to k of the n terminals is $\Omega(\log n)$-hard by a simple reduction from a set cover problem. In the above algorithm, we find an $O(\log n)$-approximation to the first stage tree. However, we do not obey the distance bound of C strictly and find a tree that is within a distance of $4C$ from at least k terminals. Obtaining a constant approximation for robust-2-CCSPP is an interesting open problem.

4 Implicit Scenario Models

In this section, we consider chance-constrained covering problem with implicit scenarios. we restrict our discussion to only one stage problems since it is not even clear whether the two-stage versions are in NP (the set of scenarios satisfying the chance constraint may not be described succinctly).

For the one stage problems, we consider an *independent scenarios* model where each element occurs with a given probability independent of others and extend to a class of general distributions.

4.1 Independent Scenarios Model: Reduction to Partial Weighted Covering Problem

Consider the one-stage set covering problem where we are given a set family $\mathcal{F}$ and a universe of elements U. The demand-uncertainty is specified by an independent scenarios model where each element e occurs independently with probability p_e (we refer it as Independent 1-CCSCP). The probability $p(E)$ of any subset E is $\Pi_{e \in E} p_e$. Also,

$$\sum_{E' \subset E} p(E') = \Pi_{e \notin E}(1 - p_e)$$

Theorem 4.1 Independent 1-CCSCP *can be reduced to a weighted partial set covering problem.*

Proof: Let z_e be a $0-1$ variable that denotes whether e is covered or not. Also, let x_S denote whether set $S \in \mathcal{F}$ is picked in the solution or not. Then, the probabilistic constraint can be written as,

$$\Pi_{z_e=0}(1 - p_e) \geq \rho$$

Taking logarithms on both sides, we get

$$\begin{aligned} \textstyle\sum_{z_e=0} \log(1 - p_e) &\geq \log \rho \\ \Rightarrow \textstyle\sum_{e \in U}(1 - z_e) \log(1 - p_e) &\geq \log \rho \\ \Rightarrow \textstyle\sum_{e \in U} -z_e \log(1 - p_e) &\geq \log \frac{\rho}{\Pi_{e \in U}(1 - p_e)} \end{aligned}$$

For each element $e \in U$, let $w_e = -\log(1-p_e)$. Also, let $W = \log \frac{\rho}{\Pi_{e \in U}(1-p_e)}$. Note that $w_e > 0$, for all $e \in U$. Now, the chance constrained set covering problem can be reduced to a weighted partial set covering problem where weight of element e is w_e and the goal is to select a minimum-cost family of subsets from $\mathcal{F}$ that cover elements of weight at least W. ■

For the general set-covering problem, the greedy algorithm gives an $O(\log W)$-approximation where W is the required weight target computed in the proof above.

The following corollary is immediate for the one-stage vertex cover problem and the spanning tree problem in the independent scenarios model by reducing the problems to the corresponding weighted partial covering versions.

One-stage Independent-Chance-Constrained Vertex Cover Problem (Independent 1-CCVCP)We are given a graph $G = (V, E)$ with costs on vertices and a reliability level ρ.. Each edge e occurs with probability p_e, independently of others. The objective is to find a minimum cost vertex cover C such that it covers ρ fraction of the scenarios (a scenario corresponds to a realization of the edges).

One-stage Independent-Chance-Constrained Shortest Path Problem (Independent 1-CCSPP)We are given a graph $G = (V, E)$ with costs on edges, a root vertex r and a reliability level ρ. Each vertex v occurs with probability p_v, independently of others. The objective is to find a minimum cost tree T containing the root r such that for ρ fraction of the scenarios there is a path in T from the root to each vertex in the scenario.

Corollary 4.2 *We obtain the following approximation guarantees for* Independent 1-CCVCP *and* Independent 1-CCSPP.

1. *There is a* 2*-approximation for* Independent 1-CCVCP.
2. *There is a* 5*-approximation for* Independent 1-CCSPP.

For the vertex cover problem, the algorithm of [22] gives a 2-approximation for Independent 1-CCVCP. From Theorem 4.1 we know that the Independent 1-CCSPP reduces to a weighted k-MST problem. Chudak et al. [13] give Lagrangian relaxation based 5-approximation for the unweighted k-MST problem that can be adapted to obtain a 5-approximation for the weighted version. This gives the result for Independent 1-CCSPP.

4.2 General Distribution Model

We consider an implicit model where scenarios come from a general distribution such that the *cumulative probability* of every demand-scenario can be computed efficiently and satisfies *strict-monotonicity* with respect to set inclusion. We show that a greedy algorithm gives a logarithmic approximation for the one-stage set cover problem in this model.

In the one-stage set cover problem in this model, we are given a set family $\mathcal{F}$, a universe of elements U and a reliability level ρ. The demand-uncertainty is specified by a probability distribution $P : 2^U \rightarrow [0, 1]$ (possibly a black-box) such that any subset $E \subset U$ occurs with probability $P(E)$. We further assume that P satisfies the following properties.

1. (**Efficiency**) Cumulative probability $F(E) = \sum_{E' \subset E} P(E')$ can be computed efficiently for any subset $E \subset U$.
2. (**Strict-Monotonicity**) For any $E_1, E_2 \subset U$, $E_1 \subsetneq E_2 \Rightarrow F(E_1) < F(E_2)$.

We obtain a logarithmic approximation for the set cover problem in this model using a greedy algorithm described below. Let OPT denote the cost of an optimal solution. We prune away all sets $S \in \mathcal{F}$ such that $c_S >$ OPT. Clearly, the modified instance is feasible. Also, let $S_{max} = \text{argmax}\{P(S) | S \in \mathcal{F}\}$. Since P is monotone, $p_{max} = P(S_{max}) > 0$. The algorithm is as follows:

1. Initialize $i \leftarrow 1, E_1 \leftarrow S_{max}$ and $\mathcal{C} \leftarrow \{S_{max}\}$.
2. While $(F(E_i) < \rho)$
   (a) Find a set $S \in \mathcal{F} \setminus \mathcal{C}$ that minimizes $\frac{c_S}{F(E_i \cup S) - F(E_i)}$.
   (b) Update $E_{i+1} \leftarrow E_i \cup S$ and $\mathcal{C} \leftarrow \mathcal{C} \cup S$.
   (c) Update $i \leftarrow i + 1$.

By a standard averaging argument, we obtain the following theorem.

Theorem 4.3 *The greedy algorithm gives an $O(\log \frac{\rho}{p_{max}})$-approximation for the one-stage set cover problem where uncertainty is given by a probability distribution (possibly a black-box) such that the cumulative probability F of any demand-scenario can be computed efficiently.*

Proof: In the greedy step i, we can find a set S_i such that $\frac{c_{S_i}}{F(E_i \cup S) - F(E_i)} \leq \frac{\mathsf{OPT}}{\rho - F(E_i)}$. Therefore,

$$\begin{aligned} c_{S_i} &\leq \mathsf{OPT}\frac{F(E_i \cup S_i) - F(E_i)}{\rho - F(E_i)} \\ &\leq \mathsf{OPT}\int_{x=F(E_i)}^{F(E_i \cup S_i)} \frac{1}{\rho - F(E_i)} dx \\ &\leq \mathsf{OPT}\int_{x=F(E_i)}^{F(E_i \cup S_i)} \frac{1}{\rho - x} dx \end{aligned}$$

Let the number of steps in the greedy algorithm be k. We know that due to the pruning step, $c_{S_k} \leq \mathsf{OPT}$ where S_k is set added in the k^{th} step. The total cost of the sets added in the first $(k-1)$ steps can be bounded as:

$$\begin{aligned} \sum_{i=1}^{k-1} c_{S_i} &\leq \sum_{i=1}^{k-1} \mathsf{OPT}\int_{x=F(E_i)}^{F(E_i \cup S_i)} \frac{1}{\rho - x} dx \\ &\leq \mathsf{OPT}\int_{x=p_{max}}^{\rho} \frac{1}{x} dx \\ &\leq \mathsf{OPT}\log\left(\frac{\rho}{p_{max}}\right) \end{aligned}$$

This proves the required approximation. ■

References

[1] S. Arora and G. Karakostas. A 2+ epsilon approximation algorithm for the k-MST problem. *Mathematical Programming*, 107(3):491–504, 2006.

[2] R. Bar-Yehuda. Using Homogeneous Weights for Approximating the Partial Cover Problem. *Journal of Algorithms*, 39(2):137–144, 2001.

[3] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming Series B*, 98:49–71, 2003.

[4] D. Bertsimas and M. Sim. The price of robustness. *Operation Research*, 52(2):35–53, 2004.

[5] J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, Berlin, 1997.

[6] A. Blum, P. Chalasani, and S. Vempala. A constant-factor approximation for the k-MST problem in the plane. *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 294–302, 1995.

[7] A. Blum, R. Ravi, and S. Vempala. A constant-factor approximation algorithm for the k MST problem (extended abstract). *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 442–448, 1996.

[8] G. Calafiore and MC Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[9] M. Charikar, S. Khuller, D.M. Mount, and G. Narasimhan. Algorithms for facility location problems with outliers. *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 642–651, 2001.

[10] A. Charnes and WW Cooper. Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints. *Operations Research*, 11(1):18–39, 1963.

[11] K. Chen. A constant factor approximation algorithm for k-median clustering with outliers. *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 826–835, 2008.

[12] X. Chen, M. Sim, and P. Sun. A Robust Optimization Perspective of Stochastic Programming. *Optimization Online*, 2005.

[13] Fabián A. Chudak, Tim Roughgarden, and David P. Williamson. Approximate k-msts and k-steiner trees via the primal-dual method and lagrangean relaxation. In *Proceedings of the 8th International IPCO Conference on Integer Programming and Combinatorial Optimization*, pages 60–70, London, UK, 2001. Springer-Verlag.

[14] G. B. Dantzig. Linear programming under uncertainty. *Management Sci.*, 1:197–206, 1955.

[15] K. Dhamdhere, V. Goyal, R. Ravi, and M. Singh. How to pay, come what may: Approximation algorithms for demand-robust covering problems. In *FOCS*, pages 367–378, 2005.

[16] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1):37–61, 2006.

[17] U. Feige, D. Peleg, and G. Kortsarz. The Dense k-Subgraph Problem. *Algorithmica*, 29(3):410–421, 2001.

[18] R. Gandhi, S. Khuller, and A. Srinivasan. Approximation algorithms for partial covering problems. *Journal of Algorithms*, 53(1):55–84, 2004.

[19] N. Garg. Saving an epsilon: a 2-approximation for the k-MST problem in graphs. *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 396–402, 2005.

[20] W.K.K. Haneveld and M.H. van der Vlerk. Stochastic Programming. *Department of Econometrics and OR, University of Groningen, Netherlands*, 2003.

[21] W.K. Klein Haneveld and M.H. van der Vlerk. Stochastic integer programming: General models and algorithms. *Annals of Operations Research*, 85:39–57, 1999.

[22] J. Mestre. A Primal-Dual Approximation Algorithm for Partial Vertex Cover: making educated guesses. *Lecture Notes in Computer Science*, 3624:182, 2005.

[23] A. Nemirovski and A. Shapiro. Convex Approximations of Chance Constrained Programs. *SIAM Journal on Optimization*, 17:969, 2006.

[24] R. Ravi and A. Sinha. Hedging uncertainty: Approximation algorithms for stochastic optimization problems. In *IPCO*, pages 101–115, 2004.

[25] M. J. Rosenblatt and H. L. Lee. A robustness approach to facilities design. *International Journal of Production Research*, 25:479–486, 1987.

[26] A. Ruszczynski and A. Shapiro. Stochastic programming models. *Handbooks in Operations Research and Management Science*, 10:1–64.

A Bicriteria Results

We show that if the chance-constraint can be violated by a constant factor, we can obtain an $O(\alpha)$-approximation when an α-approximation is known for the robust problem without the chance-constraints. For the sake of exposition, we consider the Explicit 1-CCSCP problem but essentially the same argument extends to the two-stage problems.

Let there be l scenarios $S_1, \dots, S_l$ with probabilities $p_1, \dots, p_l$ respectively. The problem is to satisfy a subset of scenarios whose probabilities sum to the reliability factor ρ.

To formulate this as an integer program (IP1), let z_i be a binary variable that denotes whether scenario i is covered or not.

$$\begin{array}{rcl} \min \sum_{S \in \mathcal{S}} c_S x_S & & \\ \sum_{S: e \in S} x_S & \geq & z_i \ \ \forall e \in S_i \ \forall i = 1, \dots, l \\ \sum_{i=1}^{l} p_i z_i & \geq & \rho \\ x_S & \in & \{0,1\} \ \ \forall S \in \mathcal{S} \\ z_i & \in & \{0,1\} \ \ \forall i = 1, \dots, l \end{array}$$

If the deterministic set covering problem has an α-approximation, then we can give an $\frac{\alpha}{\epsilon}$-approximation for any constant $\epsilon > 0$ to the chance constrained problem that violates the chance-constraint and covers only a $\rho' = \frac{\rho - \epsilon}{1 - \epsilon}$ fraction of the scenarios.

Theorem A.1 *Suppose there is an α-approximation to the deterministic set covering problem. Then for the* Explicit 1-CCSCP *with reliability ρ, there is an $\frac{\alpha}{\epsilon}$-approximation for any constant $\epsilon > 0$ that covers $\rho' = \frac{\rho - \epsilon}{1 - \epsilon}$ scenarios.*

Proof: Assume wlog that each scenario has probability $p = \frac{1}{l}$; otherwise we can consider multiple copies of the same scenario. Now, consider the optimal solution (say $(\tilde{x}, \tilde{z})$) of the LP relaxation of IP1. We know $\sum_{i=1}^{l} \frac{1}{l} \tilde{z}_i \geq \rho$ Let $H = \{i | \tilde{z}_i \geq \epsilon\}$ and $h = |H|$. Therefore,

$$h + (l - h) \cdot \epsilon \geq l\rho$$

$$\Rightarrow h \geq \frac{l(\rho - \epsilon)}{1 - \epsilon}$$

Consider the solution, $\hat{x} = \frac{1}{\epsilon}\tilde{x}$. Clearly $\hat{x}$ is a fractional solution that is feasible for all scenarios in H and can be rounded using the deterministic α approximation to an integer solution. Furthermore, the total probability of scenarios is H is $\frac{\rho - \epsilon}{1 - \epsilon}$. Therefore, we obtain an $\frac{\alpha}{\epsilon}$-approximate solution to Explicit 1-CCSCP that violates the chance constraint and covers $\rho' = \frac{\rho - \epsilon}{1 - \epsilon}$ scenarios. ■