Tight Approximation Guarantees for Concave Coverage Problems

Siddharth Barman ✉
Indian Institute of Science, Bangalore, India

Omar Fawzi ✉
Univ. Lyon, ENS Lyon, UCBL, CNRS, Inria, LIP, F-69342, Lyon Cedex 07, France

Paul Fermé ✉
Univ. Lyon, ENS Lyon, UCBL, CNRS, Inria, LIP, F-69342, Lyon Cedex 07, France

Abstract

In the maximum coverage problem, we are given subsets $T_1, \ldots, T_m$ of a universe $[n]$ along with an integer k and the objective is to find a subset $S \subseteq [m]$ of size k that maximizes $C(S) := \left|\bigcup_{i \in S} T_i\right|$. It is a classic result that the greedy algorithm for this problem achieves an optimal approximation ratio of $1 - e^{-1}$.

In this work we consider a generalization of this problem wherein an element a can contribute by an amount that depends on the number of times it is covered. Given a concave, nondecreasing function φ, we define $C^{\varphi}(S) := \sum_{a \in [n]} w_a \varphi(|S|_a)$, where $|S|_a = |\{i \in S : a \in T_i\}|$. The standard maximum coverage problem corresponds to taking $\varphi(j) = \min\{j, 1\}$. For any such φ, we provide an efficient algorithm that achieves an approximation ratio equal to the *Poisson concavity ratio* of φ, defined by $\alpha_{\varphi} := \min_{x \in \mathbb{N}^*} \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(\mathbb{E}[\mathrm{Poi}(x)])}$. Complementing this approximation guarantee, we establish a matching NP-hardness result when φ grows in a sublinear way.

As special cases, we improve the result of [4] about maximum multi-coverage, that was based on the unique games conjecture, and we recover the result of [11] on multi-winner approval-based voting for geometrically dominant rules. Our result goes beyond these special cases and we illustrate it with applications to distributed resource allocation problems, welfare maximization problems and approval-based voting for general rules.

2012 ACM Subject Classification Theory of computation → Packing and covering problems; Theory of computation → Rounding techniques; Theory of computation → Algorithmic game theory

Keywords and phrases Approximation Algorithms, Coverage Problems, Concave Function

Digital Object Identifier 10.4230/LIPIcs.STACS.2021.9

Related Version *Full Version*: `https://arxiv.org/abs/2010.00970` [3]

Funding This research is supported by the French ANR project ANR-18-CE47-0011 (ACOM). SB gratefully acknowledges the support of a Ramanujan Fellowship (SERB - SB/S2/RJN-128/2015) and a Pratiksha Trust Young Investigator Award.

1 Introduction

Coverage functions are central objects of study in combinatorial optimization. Problems related to optimizing such functions arise in multiple fields, such as operations research [10], machine learning [14], algorithmic game theory [12], and information theory [2]. The most basic covering problem is the *maximum coverage* one. In this problem, we are given subsets $T_1, \ldots, T_m$ of a universe $[n]$, along with a positive integer k, and the objective is to find a size-k subset $S \subseteq [m]$ that maximizes the coverage function $C(S) := \left|\bigcup_{i \in S} T_i\right|$. A fundamental result in the field of approximation algorithms establishes that an approximation ratio of $1 - e^{-1}$ can be achieved for this problem in polynomial-time [15] and, in fact, this approximation guarantee is tight, under the assumption that $\mathrm{P} \neq \mathrm{NP}$ [13].

38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021).
Editors: Markus Bläser and Benjamin Monmege; Article No. 9; pp. 9:1–9:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Note that in the maximum coverage problem, an element $a \in [n]$ is counted at most once in the objective, even if a appears in several selected sets. However, if we think of elements $a \in [n]$ as goods or resources, there are many settings wherein the utility indeed increases with the number of copies of a that get accumulated. Motivated, in part, by such settings, we consider a generalization of the maximum coverage problem where an element a can contribute by an amount that depends on the number of times it is covered.

Given a function $\varphi : \mathbb{N} \to \mathbb{R}_+$, an integer $k \in \mathbb{N}$, a universe of elements $[n]$, positive weights w_a for each $a \in [n]$, and subsets $T_1, \ldots, T_m \subseteq [n]$, the φ-MaxCoverage problem entails maximizing $C^\varphi(S) := \sum_{a\in[n]} w_a \varphi(|S|_a)$ over subsets $S \subseteq [m]$ of cardinality k; here $|S|_a = |\{i \in S : a \in T_i\}|$.

This work focuses on functions φ that are nondecreasing and concave (i.e., $\varphi(i+2) - \varphi(i+1) \le \varphi(i+1) - \varphi(i)$ for $i \in \mathbb{N}$). We will also assume that the function φ is normalized in the sense that $\varphi(0) = 0$ and $\varphi(1) = 1$.[1] Our approximation guarantees are in terms of the *Poisson concavity ratio* of φ, which we define as follows

$$\alpha_\varphi := \inf_{x\in\mathbb{N}^*} \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(\mathbb{E}[\mathrm{Poi}(x)])} = \inf_{x\in\mathbb{N}^*} \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(x)} . \tag{1}$$

Here $\mathrm{Poi}(x)$ denotes a Poisson-distributed random variable with parameter x. We will write $\alpha_\varphi(x) := \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(x)}$, with $\alpha_\varphi(0) = 1$. One can show that $\alpha_\varphi = \min_{x\in\mathbb{N}^*} \alpha_\varphi(x) = \inf_{x\in\mathbb{R}_+} \alpha_\varphi(x)$.[2] We refer to the full version [3] for more details on the proof of this statement.

Our main result is that φ-MaxCoverage admits an efficient α_φ-approximation algorithm, when φ is normalized nondecreasing concave, and this approximation guarantee is tight when φ grows sublinearly. Formally,

▶ **Theorem 1.** For any normalized nondecreasing concave function φ, there exists a polynomial-time α_φ-approximation algorithm for the φ-MaxCoverage problem. Furthermore, for $\varphi(n) = o(n)$, it is NP-hard to approximate the φ-MaxCoverage problem within a factor better than $\alpha_\varphi + \varepsilon$, for any constant $\varepsilon > 0$.

Before detailing the proof of the theorem, we provide a few remarks and connections to related work.

1.1 Applications and related work

We can directly reduce the standard maximum coverage problem to φ-MaxCoverage by setting $\varphi(j) = \min\{j, 1\}$. In this case $\alpha_\varphi = 1 - e^{-1}$. One can also encapsulate, within our framework, the ℓ-MultiCoverage problem studied in [4] by instantiating $\varphi(j) = \min\{j, \ell\}$. In this setting, we recover the approximation ratio $\alpha_\varphi = 1 - \frac{\ell^\ell e^{-\ell}}{\ell!}$ by a simple calculation, which matches the approximation guarantee obtained in [4]. Note that the hardness result in [4] was based on the Unique Games Conjecture, whereas the current work proves that this guarantee is tight under $\mathrm{P} \neq \mathrm{NP}$.

Another application of φ-MaxCoverage is in the context of multiwinner elections that entail selecting k (out of m) candidates with the objective of maximizing the cumulative utility of n voters; here, the utility of each voter $a \in [n]$ increases as more and more approved (by a) candidates get selected. One can reduce multiwinner elections to a coverage problem

[1] One can always replace a generic φ to a normalized one without changing the optimal solutions through a simple affine transformation.

[2] We require φ to be defined for nonnegative integers and will extend it over $\mathbb{R}_+$ by considering its piecewise linear extension.

by considering subset $T_i \subseteq [n]$ as the set of voters that approve of candidate $i \in [m]$ and $\varphi(j)$ as the utility that an agent achieves from j approved selections.[3] Addressing multiwinner elections in this standard utilitarian model, Dudycz et al. [11] obtain tight approximation guarantees for some well-studied classes of utilities. Specifically, the result in [11] applies to the classic *proportional approval voting rule*, which assigns a utility of $\sum_{i=1}^{j} \frac{1}{i}$ for j approved selections. This voting rule corresponds to the coverage problem with $\varphi(j) = \sum_{i=1}^{j} \frac{1}{i}$. Section 4.1 shows that Theorem 1 holds for all the settings considered in [11] and, in fact, applies more generally. In particular, the voting version of ℓ-MultiCoverage (studied in [21]) can be addressed by Theorem 1, but not by the result in [11]. Such a separation also arises when one truncates the proportional approval voting rule to, say, ℓ candidates, i.e., upon setting $\varphi(j) = \sum_{i=1}^{\min\{j,\ell\}} \frac{1}{i}$. Given that multiwinner elections model multiple real-world settings (e.g., committee selection [21] and parliamentary proceedings [6]), instantiations of φ-MaxCoverage in such social-choice contexts substantiate the applicability of our algorithmic result.

Coverage functions arise in numerous resource-allocation settings, such as sensor allocation [16], job scheduling, and plant location [10]. The goal, broadly, in such setups is to select k subsets of resources (out of m pre-specified ones) such that the *welfare* generated by the selected resources is maximized–each resource's contribution to the welfare increases with the number of times it is selected. This problem can be cast as φ-MaxCoverage by setting n to be the number of resources, $\{T_i\}_{i\in[m]}$ as the given collection of subsets, and $\varphi(j)$ to be the welfare contribution of a resource when it is covered j times.[4] Here, we mention a specific allocation problem to highlight the relevance of studying φ beyond the standard coverage and ℓ-coverage formulations (see Section 4.3 for details): in the Vehicle-Target Assignment problem [17, 19] the resources are n targets and covering a target j times contributes $\varphi^p(j) = \frac{1-(1-p)^j}{p}$ to the welfare; here, $p \in (0,1)$ is a given parameter. Interestingly, we find that for this problem, the approximation ratio α_φ we obtain can outperform the *Price of Anarchy* (PoA), which corresponds to the approximation ratio of any algorithm where the agents selfishly maximize their utilities (see Section 4.3 for further discussion of this point). This is to be contrasted with the resource allocation problem with $\varphi(j) = \min\{j, \ell\}$ for which it was shown in [8] that the Price of Anarchy matches with α_φ.

Theorem 1 gives us a tight approximation bound of α_φ for all the above-mentioned applications of φ-MaxCoverage. The values of α_φ for these instantiations are listed in Table 1.

It is relevant to compare the approximation guarantee, α_φ, obtained in the current work with the approximation ratio based on the notion of curvature of submodular functions. Note that if φ is nondecreasing and concave, then C^φ is submodular. One can show, via a direct calculation, that for such a submodular C^φ the curvature (as defined in [9]) is given by $c = 1 - (\varphi(m) - \varphi(m-1))$ for instances with at most m cover sets. Therefore, the algorithm of Sviridenko et al. [22] provides an approximation ratio of $1 - ce^{-1}$ for the φ-MaxCoverage problem. We note that the Poisson concavity ratio α_φ is always greater than or equal to this curvature-dependent ratio (see full version [3]). Specifically, for p-Vehicle-Target Assignment, it is strictly better for all $p \notin \{0,1\}$ and for ℓ-MultiCoverage, it is strictly better for all $\ell \geq 2$ as remarked in [4]. Therefore, for the setting at hand, the current work improves the approximation guarantee obtained in [22].

[3] Indeed, for a subset of candidates $S \subseteq [m]$, the utility of a voter $a \in [n]$ is equal to $\varphi(|S|_a)$, with $|S|_a = |\{i \in S : a \in T_i\}|$.

[4] Formally, to capture specific welfare-maximization problems in their entirety we have to a consider φ-MaxCoverage with a matroid constraint, and not just bound the number of selected subsets by k. Details pertaining to matroid constraints and the reduction appear in Section 2.2 and 4.2, respectively.

Table 1 Tight approximation ratios for particular choices of φ in the φ-MaxCoverage problem. See full version [3] for derivations of these values.

φ-MaxCoverage	$\varphi(j)$	α_φ
MaxCoverage	$\min\{j,1\}$	$1-e^{-1}$
ℓ-MultiCoverage	$\min\{j,\ell\}$	$1-\frac{\ell^\ell e^{-\ell}}{\ell!}$
Proportional Approval Voting	$\sum_{i=1}^{j}\frac{1}{i}$	$\alpha_\varphi(1)\simeq 0.7965\dots$
PAV capped at 3	$\sum_{i=1}^{\min\{j,3\}}\frac{1}{i}$	$\alpha_\varphi(1)\simeq 0.7910\dots$
p-Vehicle-Target Assignment	$\frac{1-(1-p)^j}{p}$	$\frac{1-e^{-p}}{p}$
0.1-Vehicle-Target Assignment	$\frac{1-(1-0.1)^j}{0.1}$	$\frac{1-e^{-0.1}}{0.1}\simeq 0.9516\dots$
0.1-VTA capped at 5	$\frac{1-(1-0.1)^{\min\{j,5\}}}{0.1}$	$\alpha_\varphi(5)\simeq 0.8470\dots$

1.2 Remarks on the Poisson concavity ratio α_φ

By Jensen's inequality along with the nonnegativity and concavity of φ, we have that $\alpha_\varphi \in [0,1]$. We show that α_φ can be computed numerically up to any precision $\varepsilon > 0$, in time that is polynomial in $\frac{1}{\varepsilon}$. In fact, one can show that $\alpha_\varphi(x) \geq 1-\varepsilon$ for all $x \geq N_\varepsilon := \lceil\left(\frac{6}{\varepsilon}\right)^4\rceil$ (see full version [3]). Thus, we can iterate over all $x \in \{1,2,\dots,N_\varepsilon\}$ and find $\min_{x\in[N_\varepsilon]}\alpha_\varphi(x)$ up to ε precision (under reasonable assumptions on φ). This gives us a method to overall compute α_φ, up to an absolute error of 2ε: if $\alpha_\varphi \leq 1-\varepsilon$, then computing $\min_{x\in[N_\varepsilon]}\alpha_\varphi(x)$ (up to ε precision) suffices. Otherwise, if $\alpha_\varphi \geq 1-\varepsilon$, then $\alpha_\varphi(1) \leq 1$ provides the desired bound. Furthermore, we note that even if we consider $\alpha_\varphi(x)$ over all $x \in \mathbb{R}_+$, an infimum (i.e., the value of α_φ) is achieved at an integer.

1.3 Proof techniques and organization

In Section 2, we present our approximation algorithm for the φ-MaxCoverage. The algorithm is an application of *pipage rounding*, a technique introduced in [1], on a linear programming relaxation of φ-MaxCoverage. We show that the multilinear extension F^φ of C^φ is efficiently computable and thus, we can compute an integer solution x^{int} from the optimal fractional one x^* satisfying $C^\varphi(x^{\text{int}}) \geq F^\varphi(x^*)$. Using the notion of *convex order* between distributions, we show that $F^\varphi(x^*) \geq \sum_{a\in[n]} w_a\mathbb{E}[\varphi(\mathrm{Poi}(|x^*|_a))]$, where $|x|_a = \sum_{i\in[m]:a\in T_i} x_i$. Comparing this to the value $\sum_{a\in[n]} w_a\varphi(|x^*|_a)$ taken by the linear program, we get a ratio given by the *Poisson concavity ratio* α_φ. The concavity of φ is crucial at several steps of the proof: it guarantees that the natural relaxation can be written as a linear program, it is used to relate between sums of Bernouilli random variables and a Poisson random variable via the convex order, as well as for the fact that we can restrict the infimum in the definition of α_φ to integer values of x. The generalization to matroid constraints follows in a standard way and is presented in Section 2.2.

In Section 3, we present the hardness result for φ-MaxCoverage. For this, we define a generalization of the partitioning gadget of Feige [13], extending also [4]. Roughly speaking, for an integer $x_\varphi \in \mathbb{N}$, it is a collection of x_φ-covers of the set $[n]$ (an x-cover is a collection of subsets such that each element $a \in [n]$ is covered x times, or in other words, its φ-coverage is $\varphi(x)n$) that are incompatible in the sense that if we take an element from each one of these x_φ-covers, then the φ-coverage is bounded approximately by $\mathbb{E}[\varphi(\mathrm{Poi}(x_\varphi))]n$. Then, we construct an instance of φ-MaxCoverage from an instance of the NP-hard problem *Label Cover* (as in [11]) using such a gadget with $x_\varphi \in \operatorname{argmin}_{x\in\mathbb{N}^*}\alpha_\varphi(x)$. Having set up the partitioning gadget, the analysis of the reduction can be obtained by carefully generalizing the reductions of [4] and [11].

In Section 4, we present different domains of application of our result.

2 Approximation Algorithm for φ-MaxCoverage

Fix a function $\varphi : \mathbb{N} \to \mathbb{R}_+$ that is normalized, nondecreasing and concave. The φ-MaxCoverage problem is defined as follows. The input to the problem is given by positive integers n, m, t and m subsets $T_1, \dots, T_m$ of the set $[n]$ (described as characteristic vectors), the weights $w_a \in \mathbb{Q}_+^*$ for $a \in [n]$ (described as a couple of bitstring of length t), as well as an integer $k \in \{1, \dots, m\}$. The output is a subset $S \subseteq [m]$ of size exactly k that maximizes $C^\varphi(S) = \sum_{a\in[n]} w_a \varphi(|S|_a)$, where $|S|_a = |\{i \in S : a \in T_i\}|$.

Note that the input to this problem can be specified using $n(m + 2t) + O(\log nmt)$ bits. To reduce the number of parameters, we will assume that t is polynomial in n and m, so that a polynomial time algorithm for this problem means an algorithm that runs in time polynomial in n and m. The counting function φ is fixed and does not depend on the instance of the problem, but for a given instance the problem only depends on the values $\varphi(0), \varphi(1), \dots, \varphi(m)$. We assume that we have black box access to φ and to ensure that all the algorithms run in polynomial time, we assume that $\varphi(j)$ can be described with a number of bits that is polynomial in j and that this description can be computed in polynomial time.

We now describe the approximation algorithm for φ-MaxCoverage that we analyze. As described above, we follow the standard relax and round strategy, as in [4]. First, we define a natural convex relaxation.

▶ **Definition 2.1** (Relaxed program).

$$\begin{aligned} \text{maximize} \quad & \sum_{a\in[n]} w_a c_a \\ \text{subject to} \quad & c_a \le \varphi(|x|_a), \forall a \in [n], \text{ with } |x|_a := \sum_{i\in[m]:a\in T_i} x_i \\ & 0 \le x_i \le 1, \forall i \in [m] \\ & \sum_{i=1}^{m} x_i = k\ . \end{aligned} \tag{2}$$

As previously mentioned, φ is defined on $\mathbb{R}_+$ by extending it in a piecewise linear fashion on non-integral points. As such, the constraint $c_a \le \varphi(|x|_a)$ is equivalent to m linear constraints. In fact, we can define φ_j to be the linear function $\varphi_j(t) = (\varphi(j) - \varphi(j-1))t - (j-1)\varphi(j) + j\varphi(j-1)$ for $j \in [m]$. Since φ is concave, we have that for all $t \in [0, m]$, $\varphi(t) = \min_{j\in[m]} \varphi_j(t)$. As such, the constraint $c_a \le \varphi(|x|_a)$ is equivalent to $c_a \le \varphi_j(|x|_a)$ for all $j \in [m]$ and so the program from Definition 2.1 is a linear program. Overall there are $n + m$ variables and $(n + 1)m + 1$ linear constraints, and by assumptions all the coefficients can be described using a number of bits that is polynomial in n and m. Hence an optimal solution of this linear program can be found in polynomial time.

Also observe that the program from Definition 2.1 is a relaxation of the φ-MaxCoverage problem. To see this, given a set S of size k, consider the characteristic vector $x \in \{0, 1\}^m$ defined by $x_i = 1$ if and only if $i \in S$. Then for all $a \in [n]$, we can set $c_a = \varphi(|x|_a) = \varphi(|S|_a)$, and we get an objective value of $\sum_{a\in[n]} w_a\varphi(|S|_a)$ which is exactly $C^\varphi(S)$. When solving the program from Definition 2.1, we get an optimal $x^* \in [0, 1]^m$ which is in general not integral. Next, we describe a method to round it to an integral vector $x^{\text{int}} \in \{0, 1\}^m$.

2.1 Rounding

For a submodular function $f : \{0, 1\}^m \to \mathbb{R}$, one can use pipage rounding [1, 23, 7] to transform, in polynomial time, any fractional solution $x \in [0, 1]^m$ satisfying $\sum_{i=1}^m x_i = k$ into an integral vector $x^{\text{int}} \in \{0, 1\}^m$ such that $\sum_{i=1}^m x_i^{\text{int}} = k$ and $F(x^{\text{int}}) \ge F(x)$, where F

corresponds to the multilinear extension of f, provided that $F(x)$ is computable in polynomial time for a given x; see e.g., [23, Lemma 3.4]. The multilinear extension $F : [0,1]^m \to \mathbb{R}$ of f is defined by $F(x_1, \ldots, x_m) := \mathbb{E}[f(X_1, \ldots, X_m)]$, where X_i are independent random variables with $X_i \sim \mathrm{Ber}(x_i)$, i.e., $X_i \in \{0,1\}$ with $\mathbb{P}(X_i = 1) = x_i$. Note that $F(x) = f(x)$ for an integral vector $x \in \{0,1\}^m$.

We apply this strategy to C^φ, which is submodular, and the solution x^* of the LP relaxation from Definition 2.1. Note that overall the algorithm is polynomial time, since here $F(x)$ is computable in polynomial time for a given x:

▶ **Proposition 2.2** ([3]). Let $F(x) := \mathbb{E}_{X\sim x}[C^\varphi(X)]$ for $x \in \{0,1\}^m$. We have an explicit formula for F:

$$F(x) = \sum_{a=1}^{n}\sum_{k=0}^{m}\Big[\frac{1}{m+1}\sum_{\ell=0}^{m}\omega_{m+1}^{-\ell k}\prod_{j\in[m]:a\in T_j}(1+(\omega_{m+1}^{\ell}-1)x_j)\Big]\varphi(k) \text{ with } \omega_{m+1} := e^{\frac{2i\pi}{m+1}}$$

Thus, F is computable in polynomial time in n and m.

We now analyze the value returned by the algorithm. Using the property of pipage rounding, with the notation $X = (X_1, \ldots, X_m)$ and $\mathrm{Ber}(x) = (\mathrm{Ber}(x_1), \ldots, \mathrm{Ber}(x_m))$, we get

$$C^\varphi(x^{\mathrm{int}}) = \mathbb{E}_{X\sim\mathrm{Ber}(x^{\mathrm{int}})}[C^\varphi(X)] \geq \mathbb{E}_{X\sim\mathrm{Ber}(x^*)}[C^\varphi(X)] \ .$$

Then it suffices to relate $\mathbb{E}_{X\sim\mathrm{Ber}(x^*)}[C^\varphi(X)]$ to the optimal value of the LP relaxation 2.1, which can only be larger than the optimal value of the φ-MaxCoverage problem.

▶ **Theorem 2.** Let x, c be a feasible solution of the program from Definition 2.1 and $X \sim \mathrm{Ber}(x)$. Recalling the definition of α_φ and $\alpha_\varphi(j)$ from (1), we have

$$\mathbb{E}_{X\sim\mathrm{Ber}(x)}[C^\varphi(X)] \geq \left(\min_{j\in[m]}\alpha_\varphi(j)\right)\sum_{a\in[n]} w_a c_a$$

In particular, this implies that the described polynomial time algorithm has an approximation ratio of α_φ:

$$C^\varphi(x^{\mathrm{int}}) \geq \alpha_\varphi \sum_{a\in[n]} w_a c_a^* \geq \alpha_\varphi \max_{S\subseteq[m]:|S|=k} C^\varphi(S) \ .$$

In order to prove this theorem, we need the following lemma:

▶ **Lemma 2.3.** For φ concave, and $p \in [0,1]^m$, we have:

$$\mathbb{E}\Big[\varphi\Big(\sum_{i=1}^{m}\mathrm{Ber}(p_i)\Big)\Big] \geq \mathbb{E}\Big[\varphi\Big(\mathrm{Poi}\Big(\sum_{i=1}^{m}p_i\Big)\Big)\Big]$$

Proof. The notion of *convex order* discussed in [20] allows us to prove this result. We say that $X \leq_{\mathrm{cx}} Y \iff \mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ for any convex f. Thanks to Lemma 2.3 of [4], we have that for $p \in [0,1]$:

$$\mathrm{Ber}(p) \leq_{\mathrm{cx}} \mathrm{Poi}(p)$$

Since this order is preserved through convolution (Theorem 3.A.12 of [20]), and the fact that $\sum_{i=1}^m \mathrm{Poi}(p_i) \sim \mathrm{Poi}\left(\sum_{i=1}^m p_i\right)$, we have:

$$\sum_{i=1}^m \mathrm{Ber}(p_i) \leq_{\mathrm{cx}} \mathrm{Poi}\left(\sum_{i=1}^m p_i\right)$$

Applying this result to $-\varphi$, which is convex, concludes the proof. ◀

We will also use the following property on $\alpha_\varphi(x)$:

▶ **Proposition 2.4** ([3]). For all $x \in \mathbb{R}_+$, we have $\alpha_\varphi(x) \geq \min\{\alpha_\varphi(\lfloor x \rfloor), \alpha_\varphi(\lceil x \rceil)\}$; here, $\alpha_\varphi(0) := \lim_{x\to 0} \alpha_\varphi(x) = 1$.

Proof of Theorem 2. By linearity of expectation and the fact that the weights w_a are positive, it is sufficient to show that for all $a \in [n]$:

$$\mathbb{E}[C_a^\varphi(X)] \geq \left(\min_{j\in[m]} \alpha_\varphi(j)\right) c_a \ ,$$

where $C_a^\varphi(S) := \varphi(|S|_a)$. Note that $|X|_a = \sum_{i\in[m]:a\in T_i} X_i$, and thus:

$$\begin{aligned}
\mathbb{E}[C_a^\varphi(X)] =\ & \mathbb{E}\Big[\varphi\Big(\sum_{i\in[m]:a\in T_i} X_i\Big)\Big] = \mathbb{E}\Big[\varphi\Big(\sum_{i\in[m]:a\in T_i} \mathrm{Ber}(x_i)\Big)\Big] \\
\geq\ & \mathbb{E}\Big[\varphi\Big(\mathrm{Poi}\Big(\sum_{i\in[m]:a\in T_i} x_i\Big)\Big)\Big] \text{ thanks to Lemma 2.3} \\
=\ & \mathbb{E}[\varphi(\mathrm{Poi}(|x|_a))] \geq \min\{\alpha_\varphi(\lfloor |x|_a \rfloor), \alpha_\varphi(\lceil |x|_a \rceil)\}\varphi(|x|_a) \quad \text{by Proposition 2.4} \\
\geq\ & \left(\min_{j\in[m]} \alpha_\varphi(j)\right)\varphi(|x|_a) \geq \left(\min_{j\in[m]} \alpha_\varphi(j)\right) c_a \ .
\end{aligned} \tag{3}$$

◀

2.2 Generalization to Matroid Constraints

Instead of taking a cardinality constraint k on the size of the subset S, we look now at general matroid constraints on S. Specifically, as input, instead of k, we take a matroid $\mathcal{M}$ defined on $[m]$ and given by a set of linear constraints describing its base polytope $B(\mathcal{M})$. The output is a set $S \in \mathcal{M}$ that maximizes $C^\varphi(S)$. Note that the cardinality constraint considered above is the special case where $\mathcal{M}$ is the uniform matroid of all subsets of size at most k and the base polytope $B(\mathcal{M}) = \{x \in [0,1]^m : \sum_{i=1}^m x_i = k\}$.

We first note that in the order to establish Theorem 2, the cardinality constraint $\sum_{i=1}^m x_i = k$ is not used. Thus, since the pipage rounding strategy applies to matroid constraints $\mathcal{M}$ (see [23, Lemma 3.4]), the strategy and the analysis of its efficiency generalize immediately when applied to the following linear program:

▶ **Definition 2.5** (Relaxed program for matroid constraints).

$$\begin{aligned}
\text{maximize} \quad & \sum_{a\in[n]} w_a c_a \\
\text{subject to} \quad & c_a \leq \varphi(|x|_a), \forall a \in [n] \\
& 0 \leq x_i \leq 1, \forall i \in [m] \\
& x \in B(\mathcal{M}) \quad \text{the base polytope of } \mathcal{M}\ .
\end{aligned} \tag{4}$$

▶ **Theorem 3.** Let x, c a feasible solution of the program from Definition 2.5 and $X \sim \mathrm{Ber}(x)$. Then:

$$\mathbb{E}_{X\sim \mathrm{Ber}(x)}[C^{\varphi}(X)] \geq \left(\min_{j\in[m]} \alpha_{\varphi}(j)\right) \sum_{a\in[n]} w_a c_a \ .$$

In particular, this implies that the described polynomial time algorithm has an approximation ratio of α_{φ}:

$$C^{\varphi}(x^{\mathrm{int}}) \geq \alpha_{\varphi} \sum_{a\in[n]} w_a c_a^* \geq \alpha_{\varphi} \max_{S\in\mathcal{M}} C^{\varphi}(S) \ .$$

3 Hardness of Approximation for φ-MaxCoverage

In this section, we establish an inapproximability bound for the φ-MaxCoverage problem with weights 1 under cardinality constraints. Throughout this section we use Γ to denote the universe of elements and, hence, an instance of the φ-MaxCoverage problem consists of Γ, along with a collection of subsets $\mathcal{F} = \{F_i \subseteq \Gamma\}_{i=1}^m$ and an integer k. Recall that the objective of this problem is to find a size-k subset $S \subseteq [m]$ that maximizes $C^{\varphi}(S) = \sum_{a\in\Gamma} \varphi(|S|_a)$.

We establish the following theorem in this section:

▶ **Theorem 4.** It is NP-hard to approximate the φ-MaxCoverage problem for $\varphi(n) = o(n)$ within a factor greater that $\alpha_{\varphi} + \varepsilon$ for any $\varepsilon > 0$.

Our reduction is based on a problem called h-AryLabelCover, which is equivalent to the more standard GapLabelCover problem.

▶ **Definition 3.1** (h-AryLabelCover)**.** An instance $\mathcal{G} = (V, E, [L], [R], \{\pi_{e,v}\}_{e\in E, v\in e})$ of h-AryLabelCover is characterized by an h-uniform regular hypergraph (V, E) and bijection constraints $\pi_{e,v} : [L] \to [R]$. Here, each h-uniform hyperedge represents a h-ary constraint. Additionally, for any labeling $\sigma : V \to [L]$, we have the following notions of strongly and weakly satisfied constraints:

- An edge $e = (v_1, \dots, v_h) \in E$ is *strongly satisfied* by σ if:
$$\forall x, y \in [h], \pi_{e,v_x}(\sigma(v_x)) = \pi_{e,v_y}(\sigma(v_y))$$
- An edge $e = (v_1, \dots, v_h) \in E$ is *weakly satisfied* by σ if:
$$\exists x \neq y \in [h], \pi_{e,v_x}(\sigma(v_x)) = \pi_{e,v_y}(\sigma(v_y))$$

▶ **Proposition 3.2** (δ, h-AryGapLabelCover - [3])**.** For any fixed integer $h \geq 2$ and fixed $\delta > 0$, there exists an R_0 such that for any integer $R \geq R_0$, it is NP-hard for instances $\mathcal{G} = (V, E, [L], [R], \{\pi_{e,v}\}_{e\in E, v\in e})$ of h-AryLabelCover with right alphabet $[R]$ to distinguish between:

- **YES:** There exists a labeling σ that *strongly satisfies* all the edges.
- **NO:** No labeling *weakly satisfies* more than δ fraction of the edges.

3.1 Partitioning System

The key ingredient to prove Theorem 4 is a constant size combinatorial object called partitioning system, generalizing the work of Feige [13] and [4]. For any set $[n]$, $\mathcal{Q} \subseteq 2^{[n]}$, we overload the definition $C^{\varphi}(\mathcal{Q}) := \sum_{a\in[n]} \varphi(|\mathcal{Q}|_a)$ with $|\mathcal{Q}|_a := |\{P \in \mathcal{Q} : a \in P\}|$ and $C_a^{\varphi}(\mathcal{Q}) := \varphi(|\mathcal{Q}|_a)$. Let us take $x_{\varphi} \in \mathrm{argmin}_{x\in\mathbb{N}^*} \alpha_{\varphi}(x)$, thus $\alpha_{\varphi} = \alpha_{\varphi}(x_{\varphi})$.

We say that $\mathcal{Q}$ is an *x-cover* of $x \in \mathbb{N}$ if every element of $[n]$ is covered x times, so $C^{\varphi}(\mathcal{Q}) = n\varphi(x)$.

▶ **Definition 3.3.** An $([n], h, R, \varphi, \eta)$-*partitioning system* consists of R collections of subsets of $[n]$, $\mathcal{P}_1, \ldots, \mathcal{P}_R \subseteq 2^{[n]}$, that satisfy $\frac{x_\varphi n}{h} \in \mathbb{N}$, $x_\varphi \geq h$ and:

1. For every $i \in [R], \mathcal{P}_i$ is a collection of h subsets $P_{i,1}, \ldots, P_{i,h} \subseteq [n]$ each of size $\frac{x_\varphi n}{h}$ which is an x_φ-cover.
2. For any $T \subseteq [R]$ and $\mathcal{Q} = \{P_{i,j(i)} : i \in T\}$ for some function $j : T \to [h]$, we have $\left|C^\varphi(\mathcal{Q}) - \psi^\varphi_{|T|,h} n\right| \leq \eta n$ where:

$$\psi^\varphi_{k,h} := \mathbb{E}\Big[\varphi\Big(\operatorname{Bin}\Big(k, \frac{x_\varphi}{h}\Big)\Big)\Big] . \tag{5}$$

▶ Remark. In particular, for any $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ with Q_i of size $\frac{x_\varphi n}{h}$, we have that $C^\varphi(\mathcal{Q}) \leq n\varphi(k\frac{x_\varphi}{h})$. Indeed $C^\varphi(\mathcal{Q}) = \sum_{a\in[n]} \varphi(|\mathcal{Q}|_a)$ with $\sum_{a\in[n]} |\mathcal{Q}|_a = \sum_{i\in[k]} |Q_i| = k \cdot \frac{x_\varphi n}{h}$. By concavity of φ and Jensen's inequality, this function is maximized when all $|\mathcal{Q}|_a$ are equals, where we get $n\varphi(k\frac{x_\varphi}{h})$.

▶ **Proposition 3.4** ([3]). For $R, h \in \mathbb{N}$ with $h \geq x_\varphi$, $\eta \in (0, 1)$, $n \geq \eta^{-2} R\varphi(R)^2 \log(20(h + 1))$ such that $\frac{x_\varphi n}{h} \in \mathbb{N}$, there exists an $([n], h, R, \varphi, \eta)$-partitioning system, which can be found in time $\exp(Rn \log(n))\cdot\operatorname{poly}(h)$.

3.2 The Reduction

Proof of Theorem 4. Let $\varepsilon > 0$. Without loss of generality, we can assume that $\varepsilon < 1$. We show that it is NP-hard to reach an approximation greater than $\alpha_\varphi + \varepsilon$ for the φ-MaxCoverage problem, via a reduction from δ, h-AryGapLabelCover. Define:

- $\eta = \frac{\varphi(x_\varphi)}{4x_\varphi}\varepsilon$, so $0 < \eta \leq \varepsilon < 1$,
- $h \geq x_\varphi$ such that $\left|\psi^\varphi_{h,h} - \alpha_\varphi \varphi(x_\varphi)\right| \leq \eta$ (see (5) for the definition of ψ^φ); one can show that such a choice exists by bounding the total variation between Bernouilli and Poisson laws, together with the fact that $\varphi(x) = o(x)$ (see full version [3]),
- θ such that for all $x \geq \theta$, $\frac{\varphi(x)}{x} \leq \eta$, which exists since $\varphi(x) = o(x)$,
- $\xi = \frac{x_\varphi}{\theta}$,
- $\delta = \frac{\eta}{2}\frac{\xi^3}{h^2}$,
- $R \geq h$ large enough for Proposition 3.2 to hold.

Given an instance $\mathcal{G} = (V, E, [L], [R], \Sigma, \{\pi_{e,v}\}_{e\in E, v\in e})$ of δ, h-AryGapLabelCover, we construct an instance $(\Gamma, \mathcal{F}, k)$ of the φ-MaxCoverage problem with:

- n a large enough integer to have an $([n], h, R, \varphi, \eta)$-partitioning system (Proposition 3.4),
- $\Gamma = [n] \times E$,
- $k = |V|$,
- Consider a $([n], h, R, \varphi, \eta)$-partitioning system, and call $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_R\}$ the corresponding set of collections. Define sets $T^{e,v_j}_\beta = P_{\pi_{e,v_j}(\beta),j} \times \{e\}$ for $e = (v_1, \ldots, v_h) \in E, j \in [h], \beta \in [L]$. Then, choose as cover sets $F^v_\beta := \bigsqcup_{e\in E: v\in e} T^{e,v}_\beta$ and take $\mathcal{F} := \{F^v_\beta, v \in V, \beta \in [L]\}$.

We will now prove that if we are in a YES instance, we have that there exists $\mathcal{T}$ of size k such that $C^\varphi(\mathcal{T}) \geq \varphi(x_\varphi)|\Gamma|$ (completeness). Moreover, if we are in a NO instance, then we have that for all $\mathcal{T}$ of size $k = |V|$, $C^\varphi(\mathcal{T}) \leq (\alpha_\varphi + \varepsilon)\varphi(x_\varphi)|\Gamma|$ (soundness). Establishing these two properties will conclude the proof.

In order to achieve this, let us define $C^{\varphi,e} := \sum_{a\in[n]\times\{e\}} C^\varphi_a$. In particular, $C^\varphi = \sum_{a\in\Gamma} C^\varphi_a = \sum_{e\in E} C^{\varphi,e}$. For $\mathcal{T} \subseteq \mathcal{F}$, we define the relevant part of $\mathcal{T}$ on e by:

$$\mathcal{T}_e := \{T^{e,v}_\beta : v \in e, \beta \in [L], F^v_\beta \in \mathcal{T}\} = \{F^v_\beta \cap ([n] \times \{e\}), F^v_\beta \in \mathcal{T}\}.$$

Note that $C^{\varphi,e}(\mathcal{T}) = C^{\varphi,e}(\mathcal{T}_e)$, and in particular $C^\varphi(\mathcal{T}) = \sum_{e\in E} C^{\varphi,e}(\mathcal{T}_e)$.

3.3 Completeness

Suppose the given h-AryLabelCover instance $\mathcal{G}$ is a YES instance. Then, there exists a labeling $\sigma : V \mapsto [L]$ which strongly satisfies all edges. Consider the collection of $|V|$ subsets $\mathcal{T} := \{F^v_{\sigma(v)} : v \in V\}$. Fix $e = (v_1, \ldots, v_h) \in E$. Since e is strongly satisfied by σ, there exists $r \in [R]$ such that $\pi_{e,v_i}(\sigma(v_i)) = r$ for all $i \in [h]$. Thus, $\mathcal{T}_e = \{T^{e,v_i}_{\sigma(v_i)}\}_{i\in[h]} = \{P_{r,i} \times \{e\}\}_{i\in[h]}$ is an x_φ-cover of $[n] \times \{e\}$, and so $C^{\varphi,e}(\mathcal{T}_e) = n\varphi(x_\varphi)$. Thus $C^\varphi(\mathcal{T}) = \sum_{e\in E} C^{\varphi,e}(\mathcal{T}_e) = |E|\varphi(x_\varphi)n = \varphi(x_\varphi)|\Gamma|$.

3.4 Soundness

Suppose the given h-AryLabelCover instance $\mathcal{G}$ is a NO instance. Let us prove the contrapositive of the soundness: we suppose that there exists $\mathcal{T}$ of size $k = |V|$ such that $C^\varphi(\mathcal{T}) > (\alpha_\varphi + \varepsilon)\varphi(x_\varphi)|\Gamma|$. Let us show that there exists a labelling σ that weakly satisfies a strictly larger fraction of the edges than δ.

For every vertex $v \in V$, we define $L(v) := \{\beta \in [L] : F^v_\beta \in \mathcal{T}\}$ to be the candidate set of labels that can be associated with the vertex v. We extend this definition to hyperedges $e = (v_1, \ldots, v_h)$ where we define $L(e) := \bigcup_{i\in[h]} L(v_i)$ to be the *multiset* of all labels associated with the edge. Note that $|\mathcal{T}_e| = |L(e)|$.

We say that $e = (v_1, \ldots, v_h) \in E$ is *consistent* if and only if $\exists x \neq y \in [h], \pi_{e,v_x}(L(v_x)) \cap \pi_{e,v_y}(L(v_y)) \neq \emptyset$. We then decompose E in three parts:

- B is the set of edges $e \in E$ with $|L(e)| \geq \frac{h}{\xi}$.
- N is the set of consistent edges $e \in E$ with $|L(e)| < \frac{h}{\xi}$.
- $I = E - (B \cup N)$ is the set of inconsistent edges $e \in E$ with $|L(e)| < \frac{h}{\xi}$.

We want to show that the contribution of N is not too small, which we will use to construct a labelling weakly satisfying enough edges. This comes from the following lemmas:

▶ **Lemma 3.5.** $\sum_{e\in E} |L(e)| = |E|h$

Proof. Recall that our h-uniform hypergraph is regular; call d its regular degree. In particular, we have that $d|V| = |E|h$. Note also that $\sum_{v\in V} |L(v)| = |\mathcal{T}| = |V|$. Thus:

$$\sum_{e\in E} |L(e)| = \sum_{e\in E} \sum_{v\in V: v\in e} |L(v)| = \sum_{v\in V} \sum_{e\in E: v\in e} |L(v)| = d \sum_{v\in V} |L(v)| = d|V| = |E|h \ . \tag{6}$$

◀

Next, we bound the contribution of B:

▶ **Lemma 3.6.** $\sum_{e\in B} C^{\varphi,e}(\mathcal{T}_e) \leq \frac{\varepsilon}{4}\varphi(x_\varphi)|\Gamma|$.

Proof. We have:

$$\begin{aligned}
\sum_{e\in B} C^{\varphi,e}(\mathcal{T}_e) \leq & \sum_{e\in B} n\varphi\Big(|L(e)|\frac{x_\varphi}{h}\Big) && \text{by the remark on Definition 3.3 and } |\mathcal{T}_e| = |L(e)| \\
\leq & \ |B| \cdot n\varphi\Big(\frac{\sum_{e\in B} |L(e)|}{|B|}\frac{x_\varphi}{h}\Big) && \text{by Jensen's inequality on concave } \varphi \\
\leq & \ |B| \cdot n\varphi\Big(\frac{|E|h}{|B|}\frac{x_\varphi}{h}\Big) && \text{since } \varphi \text{ nondecreasing and } \sum_{e\in B} |L(e)| \leq |E|h \\
= & \ \frac{\varphi\big(\frac{|E|x_\varphi}{|B|}\big)}{\frac{|E|x_\varphi}{|B|}} x_\varphi |\Gamma| \ .
\end{aligned}$$

(7)

We have seen that $\sum_{e\in B}|L(e)| \leq |E|h$, but $\sum_{e\in B}|L(e)| \geq |B|\frac{h}{\xi}$ by definition of B, so we have that $\frac{|B|}{|E|} \leq \xi$. Thus $\frac{|E|x_\varphi}{|B|} \geq \frac{x_\varphi}{\xi} = \theta$. By definition of θ, we get that $\sum_{e\in B} C^{\varphi,e}(\mathcal{T}_e) \leq \eta x_\varphi|\Gamma| = \frac{\varepsilon}{4}\varphi(x_\varphi)|\Gamma|$. ◀

In order to bound the contribution of I, we will prove a property on inconsistent edges:

▶ **Proposition 3.7.** Let $e = (v_1, \dots, v_h) \in E$ be an inconsistent hyperedge with respect to $\mathcal{T}$. Then we have that $\left|C^{\varphi,e}(\mathcal{T}_e) - \psi^{\varphi}_{|L(e)|,h}n\right| \leq \eta n$.

Proof. Since e is inconsistent, $\forall x \neq y \in [h], \pi_{e,v_x}(L(v_x)) \cap \pi_{e,v_y}(L(v_y)) = \emptyset$. Therefore, for every $i \in [R]$, there is at most one $v \in e$ such that $i \in \pi_{e,v}(L(v))$, i.e., $\mathcal{T}_e$ intersects with $\mathcal{P}_i \times \{e\}$ in at most one subset. This gives us a subset $T \subseteq [R]$ and a function $j : T \to [h]$ such that $\mathcal{T}_e = \{P_{i,j(i)} \times \{e\} : i \in T\}$. As a consequence, $|T| = |\mathcal{T}_e| = |L(e)|$ and by the second condition of the partitioning system, we get the expected result. ◀

Now, we can bound the contribution of I:

▶ **Lemma 3.8.** $\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \leq (\alpha_\varphi + \frac{\varepsilon}{2})\varphi(x_\varphi)|\Gamma|$.

Proof. Thanks to Proposition 3.7, we have:

$$\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \leq \sum_{e\in I}(\psi^{\varphi}_{|L(e)|,h} + \eta)n \leq \sum_{e\in E}(\psi^{\varphi}_{|L(e)|,h} + \eta)n \ , \tag{8}$$

since $I \subseteq E$ and $\psi^{\varphi}_{|L(e)|,h} \geq 0$. But $\sum_{e\in E}|L(e)| = |E|h$ by Lemma 3.5, and one can show that $x \mapsto \psi^{\varphi}_{x,h}$ is concave (see full version [3]), so we can use Jensen's inequality to get $\sum_{e\in E}\psi^{\varphi}_{|L(e)|,h} \leq |E|\psi^{\varphi}_{\frac{\sum_{e\in E}|L(e)|}{|E|},h} = |E|\psi^{\varphi}_{h,h}$ and thus:

$$\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \leq (\psi^{\varphi}_{h,h} + \eta)n|E| \leq (\alpha_\varphi\varphi(x_\varphi) + 2\eta)|\Gamma| \ , \tag{9}$$

by definition of h. This implies that the total contribution of inconsistent edges I is at most $\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \leq (\alpha_\varphi\varphi(x_\varphi) + 2\eta)|\Gamma| \leq (\alpha_\varphi + \frac{\varepsilon}{2})\varphi(x_\varphi)|\Gamma|$ by definition of η. ◀

▶ **Lemma 3.9.** $\sum_{e\in N} C^{\varphi,e}(\mathcal{T}_e) > \frac{\varepsilon}{4}\varphi(x_\varphi)|\Gamma|$ and thus $\frac{|N|}{|E|} \geq \xi\eta$.

Proof. Since we have supposed that $\sum_{e\in E} C^{\varphi,e}(\mathcal{T}_e) = C^{\varphi}(\mathcal{T}) > (\alpha_\varphi + \varepsilon)\varphi(x_\varphi)|\Gamma|$, and with the help of Lemmas 3.6 and 3.8, we have that the contribution of N is:

$$\sum_{e\in N} C^{\varphi,e}(\mathcal{T}_e) > \frac{\varepsilon}{4}\varphi(x_\varphi)|\Gamma| \ .$$

However, we have that for $e \in N$ that $C^{\varphi,e}(\mathcal{T}_e) \leq n\varphi\left(|\mathcal{T}_e|\frac{x_\varphi}{h}\right) = n\varphi\left(|L(e)|\frac{x_\varphi}{h}\right) \leq n\varphi\left(\frac{x_\varphi}{\xi}\right) \leq \frac{nx_\varphi}{\xi}$ thanks to the remark on Definition 3.3 and the bound $|L(e)| < \frac{h}{\xi}$. This implies that:

$$\frac{|N|}{|E|} \geq \frac{\xi}{x_\varphi}\frac{\varepsilon\varphi(x_\varphi)}{4} = \xi\eta \ .$$ ◀

From this, we construct a randomized labeling $\sigma : V \mapsto [L]$ as follows: for $v \in V$, if $L(v) \neq \emptyset$, set $\sigma(v)$ uniformly from $L(v)$, otherwise set it arbitrarily. We claim that in expectation, this labeling must weakly satisfy δ fraction of the hyperedges.

To see this, fix any $e = (v_1, \ldots, v_h) \in N$. Thus $\exists x \neq y \in [h], \pi_{e,v_x}(L(v_x)) \cap \pi_{e,v_y}(L(v_y)) \neq \emptyset$. Furthermore $|L(v_x)|, |L(v_y)| \leq \frac{h}{\xi}$. Thus, we have that $\pi_{e,v_x}(L(v_x)) = \pi_{e,v_y}(L(v_y))$ with probability at least $\frac{1}{|L(v_x)||L(v_y)|} \geq \left(\frac{\xi}{h}\right)^2$.

Therefore:

$$\begin{aligned} & \mathbb{E}_\sigma \mathbb{E}_{e \sim E}[\sigma \text{ weakly satisfies } e] \\ \geq \quad & \xi\eta \mathbb{E}_\sigma \mathbb{E}_{e \sim E}[\sigma \text{ weakly satisfies } e | e \in N] \quad \text{by Lemma 3.9} \\ > \quad & \frac{\eta}{2} \frac{\xi^3}{h^2} = \delta \ . \end{aligned} \tag{10}$$

In particular there exists some labeling σ such that $\mathbb{E}_{e \sim E}[\sigma \text{ weakly satisfies } e] > \delta$, and thus the soundness is also proved. ◀

4 Applications

This section show that instantiations of φ-MaxCoverage encapsulate and generalize multiple problems from fields such as computational social choice [5] and algorithmic game theory [18].

4.1 Multiwinner Elections

As mentioned previously, multiwinner elections (with a utilitarian model for the voters) entail selection of k (out of m) candidates that maximize the utility across n voters. Here, the utility of each voter $a \in [n]$ increases with the number of approved (by a) selections. The work of Dudycz et al. [11] study the computational complexity of such elections and, in particular, address classic voting rules in which – for a specified sequence of nonnegative weights $(w_1, w_2, \ldots)$ – voter a's utility is equal to $\sum_{i=1}^{j} w_i$, when she approves of j candidates among the selected ones. One can view this election exercise as a coverage problem by considering subset $T_i \subseteq [n]$ as the set of voters that approve of candidate $i \in [m]$ and $\varphi(j) = \sum_{i=1}^{j} w_i$. Indeed, for a subset of candidates $S \subseteq [m]$, the utility of a voter $a \in [n]$ is equal to $\varphi(|S|_a)$, with $|S|_a = |\{i \in S : a \in T_i\}|$.

Dudycz et al. [11] show that if the weights satisfy $w_1 \geq w_2 \geq \ldots$ (i.e., bear a diminishing returns property) along with geometric dominance ($w_i \cdot w_{i+2} \geq w_{i+1}^2$ for all $i \in \mathbb{N}^*$) and $\lim_{i \to \infty} w_i = 0$, then a tight approximation guarantee can be obtained for the election problem at hand. Note that the diminishing returns property implies that $\varphi(j) = \sum_{i=1}^{j} w_i$ is concave and $\lim_{i \to \infty} w_i = 0$ ensures that φ is sublinear. Furthermore, one can show that:

▶ **Proposition 4.1** ([3]). If $w_i := \varphi(i) - \varphi(i-1)$ is *geometrically dominant*, ie. $\forall i \in \mathbb{N}^*$, $\frac{w_i}{w_{i+1}} \geq \frac{w_{i+1}}{w_{i+2}}$, then $\alpha_\varphi = \alpha_\varphi(1)$.

Hence, Theorem 1, together with Proposition 4.1, can be invoked to recover the result in [11] where we get $\alpha_\varphi = \alpha_\varphi(1)$. In fact, Theorem 1 does not require geometric dominance among the weights and, hence, applies to a broader class of voting rules. For instance, the geometric dominance property does not hold if one considers the voting weights induced by ℓ-MultiCoverage, i.e., $w_i = 1$, for $1 \leq i \leq \ell$, and $w_j = 0$ for $j > \ell$. However, using Theorem 1, we get that for this voting rule we can approximate the optimal utility within a factor of $\alpha_\varphi = 1 - \frac{\ell^\ell e^{-\ell}}{\ell!}$. Another example of such a separation arises if one truncates the proportional approval voting. The standard proportional approval voting corresponds to $w_i = \frac{1}{i}$, for all $i \in \mathbb{N}$ (equivalently, $\varphi(j) = \sum_{i=1}^{j} \frac{1}{i}$) and falls within the purview of [11]. While the truncated version with $\varphi(j) = \sum_{i=1}^{\min\{j,\ell\}} \frac{1}{i}$, for a given threshold ℓ, does not satisfy geometric dominance, Theorem 1 continues to hold and provide a tight approximation ratio that can be computed numerically (see Table 1 for examples):

▶ **Proposition 4.2** ([3])**.** If $\forall x \geq \ell, \varphi(x) = \varphi(\ell) > 0$, then $\alpha_\varphi(x)$ is nondecreasing from ℓ to $+\infty$ and $\alpha_\varphi(x) = \frac{\varphi(\ell) - e^{-x} \sum_{k=0}^{\ell-1} (\varphi(\ell) - \varphi(k)) \frac{x^k}{k!}}{\varphi(x)}$. In particular, $\alpha_\varphi = \min_{x \in [\ell]} \alpha_\varphi(x)$, and the argmin can be computed numerically.

4.2 Resource Allocation in Multiagent Systems

A significant body of prior work in algorithmic game theory has addressed game-theoretic aspects of maximizing welfare among multiple (strategic) agents; see, e.g., [19]. Complementing such results, this section shows that the optimization problem underlying multiple welfare-maximization games can be expressed in terms of φ-MaxCoverage.

Specifically, consider a setting with n resources, k agents, and a (counting) function $\varphi : \mathbb{N} \mapsto \mathbb{R}_+$. Every agent i is endowed with a collection of resource subsets $\mathcal{A}_i = \{T_1^i, \ldots, T_{m_i}^i\} \subseteq 2^{[n]}$ (i.e., each $T_j^i \subseteq [n]$). The objective is to select a subset $A_i \in \mathcal{A}_i$, for all $i \in [k]$, so as to maximize $W^\varphi(A_1, A_2, \ldots, A_k) \coloneqq \sum_{a \in [n]} w_a \, \varphi(|A|_a)$. Here, $w_a \in \mathbb{R}_+$ is a weight associated with $a \in [n]$ and $|A|_a \coloneqq |\{i \in [k] : a \in A_i\}|$. We will refer to this problem as the φ-Resource Allocation problem.

While φ-Resource Allocation does not directly reduce to φ-MaxCoverage, the next theorem shows that it corresponds to maximizing φ-coverage functions subject to a matroid constraint. Hence, invoking our result from Section 2.2, we obtain a tight α_φ-approximation for φ-Resource Allocation:

▶ **Theorem 5** ([3])**.** For any normalized nondecreasing concave function φ, there exists a polynomial-time α_φ-approximation algorithm for φ-Resource Allocation. Furthermore, for $\varphi(n) = o(n)$, it is NP-hard to approximate φ-Resource Allocation within a factor better than $\alpha_\varphi + \varepsilon$, for any constant $\varepsilon > 0$.

4.3 Vehicle-Target Assignment

Vehicle-Target Assignment [17, 19] is another problem which highlights the applicability of coverage problems, with a concave φ. In particular, Vehicle-Target Assignment can be directly expressed as φ-Resource Allocation: the $[n]$ resources correspond to targets, the agents correspond to vehicles $i \in [k]$, each with a collection of covering choices $\mathcal{A}_i \subseteq 2^{[n]}$, and $\varphi^p(j) = \frac{1-(1-p)^j}{p}$, for a given parameter $p \in (0,1)$. As limit cases, we define $\varphi^0(j) \coloneqq \lim_{p \to 0} \varphi^p(j) = j$ and $\varphi^1(j) \coloneqq 1$. Since $\varphi^p(j)$ is concave, by a simple calculation and Theorem 5, we obtain a novel tight approximation ratio of $\alpha_{\varphi^p} = \frac{1-e^{-p}}{p}$ for this problem. Also, one can look at the capped version of this problem, $\varphi_\ell^p(j) \coloneqq \varphi^p(\min\{j,\ell\})$. In particular, we recover the ℓ-MultiCoverage function when $p = 0$. In Figure 1, we have plotted several cases of the tight approximations $\alpha_{\varphi_\ell^p}$ in function of ℓ for several values of ℓ:

Paccagnan and Marden [19] study the game-theoretic aspects of Vehicle-target assignment. A key goal in [19] is to bound the welfare loss incurred due to strategic selection by the k vehicles, i.e., the selection of each $A_i \in \mathcal{A}_i$ by a self-interested vehicle/agent $i \in [k]$. The loss is quantified in terms of the *Price of Anarchy* (PoA). Formally, this performance metric is defined as ratio between the welfare of the worst-possible equilibria and

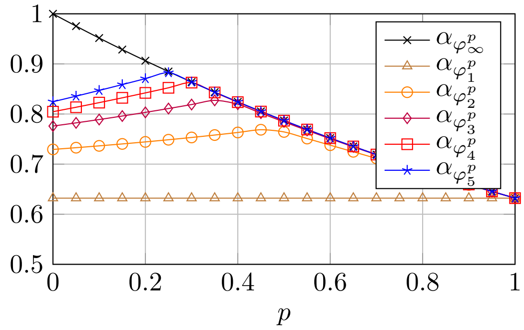

Figure 1 Tight approximation ratios $\alpha_{\varphi^p_\ell}$, where ℓ is the rank of the capped version of the p-VEHICLE-TARGET ASSIGNMENT problem. When $p = 0$, we recover the ℓ-coverage problem.

the optimal welfare. Paccagnan and Marden [19] show that, for computationally tractable equilibrium concepts (in particular, for coarse correlated equilibria), tight price of anarchy bounds can be obtained via linear programs.

Note that our hardness result (Theorem 1) provides upper bounds on PoA of tractable equilibrium concepts–this follows from the observation that computing an equilibrium provides a specific method for finding a coverage solution. In [8] and in the particular case of the ℓ-MULTICOVERAGE problem, it is shown that this in fact an equality, i.e., PoA $= \alpha_\varphi$ if $\varphi(j) = \min\{j, \ell\}$ for all values of ℓ. However, numerically comparing the approximation ratio for VEHICLE-TARGET ASSIGNMENT, $\alpha_{\varphi^p} = \frac{1-e^{-p}}{p}$, with the optimal PoA bound, we note that α_{φ^p} can in fact be strictly greater than the PoA guarantee; see Figure 2.

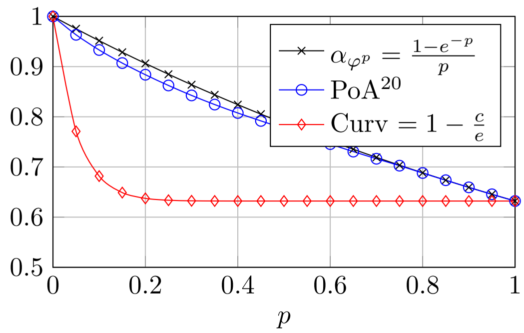

Figure 2 Comparison between the PoA and α_φ for the VEHICLE-TARGET ASSIGNMENT problem. Using the linear program found in [19], we were able to compute the blue curve PoA^{20}, the *Price of Anarchy* of this problem for $m = 20$ players. Since the PoA only decreases when the number of players grows, this means that PoA $< \alpha_\varphi$ in that case. As a comparison, the red curve Curv depicts the general approximation ratio (see [22]) obtained for submodular function with curvature c, with $c = 1 - \varphi^p(m) + \varphi^p(m-1)$ here.

4.4 Welfare Maximization for φ-Coverage

Maximizing (social) welfare by partitioning items among agents is a key problem in algorithmic game theory; see, e.g., the extensive work on combinatorial auctions [18]. The goal here is to partition t items among a set of k agents such that the sum of values achieved by the agents – referred to as the social welfare – is maximized. That is, one needs to partition $[t]$ into k pairwise disjoint subsets $A_1, A_2, \ldots, A_k$ with the objective of maximizing $\sum_{i=1}^k v_i(A_i)$. Here, $v_i(S)$ denotes the valuation that agent i has for a subset of items $S \subseteq [t]$.

When each agent's valuation v_i is submodular, a tight $(1 - e^{-1})$-approximation ratio is known for social welfare maximization [23]. This section shows that improved approximation guarantees can be achieved if, in particular, the agents' valuations are φ-coverage functions. Towards a stylized application of such valuations, consider a setting in which each "item" $b \in [t]$ represents a bundle (subset) of goods $T_b \subseteq [n]$ and the value of an agent increases with the number of copies of any good $a \in [n]$ that get accumulated. Indeed, if each agent's value for j copies of a good is $\varphi(j)$, then we have a φ-coverage function and the overall optimization problem is find a k-partition, $A_1, A_2, \ldots, A_k$, of $[t]$ that maximizes $\sum_{i=1}^{k} \left(\sum_{a \in [n]} \varphi\left(|A_i|_a\right) \right)$, where $|A_i|_a \coloneqq \{b \in A_i : a \in T_b\}$.

In the current setup, one can obtain an α_φ approximation ratio for social-welfare maximization by reducing this problem to φ-coverage with a matroid constraint, and applying the result from Section 2.2. Specifically, we can consider a partition matroid over the universe $[t] \times [k]$: for a bundle/item $b \in [t]$ and an agent $i \in [k]$, the element (b, i) in the universe represents that bundle b is assigned to agent i, i.e., $b \in A_i$. The partition-matroid constraint is imposed to ensure that each bundle b is assigned to at most one agent. Furthermore, we can create k copies of the underlying set of goods $[n]$ and set $T_{(b,i)} \coloneqq \{(a, i) : a \in T_b\}$ to map the φ-coverage over the universe to the social-welfare objective. This, overall, gives us the desired α_φ approximation guarantee.

Conclusion

We have introduced the φ-MaxCoverage problem where having c copies of element a gives a value $\varphi(c)$. We have shown that when φ is normalized, nondecreasing and concave, we can obtain an approximation guarantee given by the *Poisson concavity ratio* $\alpha_\varphi \coloneqq \min_{x \in \mathbb{N}^*} \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(\mathbb{E}[\mathrm{Poi}(x)])}$ and we showed it is tight for sublinear functions φ. The Poisson concavity ratio strictly beats the bound one gets when using the notion of curvature submodular functions, except in very special cases such as MaxCoverage where the two bounds are equal.

An interesting open question is whether there exists combinatorial algorithms that achieve this approximation ratio. As mentioned in [4], for the ℓ-MultiCoverage with $\ell \geq 2$, which is the special case where $\varphi(x) = \min\{x, \ell\}$, the simple greedy algorithm only gives a $1 - e^{-1}$ approximation ratio, which is strictly less than the ratio $\alpha_\varphi = 1 - \frac{\ell^\ell e^{-\ell}}{\ell!}$ in that case. Also, for any geometrically dominant vector $w = (\varphi(i+1) - \varphi(i))_{i \in \mathbb{N}}$ which is not p-geometric, such as Proportional Approval Voting, the greedy algorithm achieves an approximation ratio which is strictly less than α_φ (see Theorem 18 of [11]).

Another open question is whether the hardness result remains true even when $\varphi(n) \neq o(n)$. A good example is given by $\varphi(0) = 0$ and $\varphi(1 + t) = 1 + (1 - c)t$ with $c \in (0, 1)$. We know that the problem is hard for $c = 1$ but easy for $c = 0$. One can show that the approximation ratio achieved by our algorithm is $\alpha_\varphi = 1 - \frac{c}{e}$ in that case (which is the same approximation ratio obtained from the curvature in [22]), but the tightness of this approximation ratio remains open.

References

1 Alexander A. Ageev and Maxim Sviridenko. Pipage rounding: A new method of constructing algorithms with proven performance guarantee. *J. Comb. Optim.*, 8(3):307–328, 2004. `doi:10.1023/B:JOCO.0000038913.96607.c2`.

2 Siddharth Barman and Omar Fawzi. Algorithmic aspects of optimal channel coding. *IEEE Trans. Inf. Theory*, 64(2):1038–1045, 2018. `doi:10.1109/TIT.2017.2696963`.

3 Siddharth Barman, Omar Fawzi, and Paul Fermé. Tight approximation guarantees for concave coverage problems. *CoRR*, abs/2010.00970, 2020. arXiv:2010.00970.

4 Siddharth Barman, Omar Fawzi, Suprovat Ghoshal, and Emirhan Gürpinar. Tight approximation bounds for maximum multi-coverage. In Daniel Bienstock and Giacomo Zambelli, editors, *Integer Programming and Combinatorial Optimization - 21st International Conference, IPCO 2020, London, UK, June 8-10, 2020, Proceedings*, volume 12125 of *Lecture Notes in Computer Science*, pages 66–77. Springer, 2020. doi:10.1007/978-3-030-45771-6_6.

5 Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016. doi:10.1017/CBO9781107446984.

6 Markus Brill, Jean-François Laslier, and Piotr Skowron. Multiwinner approval rules as apportionment methods. In Satinder P. Singh and Shaul Markovitch, editors, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 414–420. AAAI Press, 2017. doi:10.1177/0951629818775518.

7 Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011. doi:10.1137/080733991.

8 Rahul Chandan, Dario Paccagnan, and Jason R. Marden. Optimal mechanisms for distributed resource-allocation. *CoRR*, abs/1911.07823, 2019. arXiv:1911.07823.

9 Michele Conforti and Gérard Cornuéjols. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the Rado-Edmonds theorem. *Discret. Appl. Math.*, 7(3):251–274, 1984. doi:10.1016/0166-218X(84)90003-9.

10 Gérard Cornuéjols, Marshall L Fisher, and George L Nemhauser. Exceptional paper—location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management science*, 23(8):789–810, 1977. doi:10.1287/mnsc.23.8.789.

11 Szymon Dudycz, Pasin Manurangsi, Jan Marcinkowski, and Krzysztof Sornat. Tight approximation for proportional approval voting. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 276–282. ijcai.org, 2020. doi:10.24963/ijcai.2020/39.

12 Shaddin Dughmi and Jan Vondrák. Limitations of randomized mechanisms for combinatorial auctions. *Games Econ. Behav.*, 92:370–400, 2015. doi:10.1016/j.geb.2014.01.007.

13 Uriel Feige. A threshold of ln n for approximating set cover. *J. ACM*, 45(4):634–652, 1998. doi:10.1145/285055.285059.

14 Vitaly Feldman and Pravesh Kothari. Learning coverage functions and private release of marginals. In Maria-Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 679–702. JMLR.org, 2014. URL: http://proceedings.mlr.press/v35/feldman14a.html.

15 Dorit S. Hochbaum. Approximation algorithms for NP-hard problems. *SIGACT News*, 28(2):40–52, 1997. doi:10.1145/261342.571216.

16 Jason R. Marden and Adam Wierman. Distributed welfare games with applications to sensor coverage. In *Proceedings of the 47th IEEE Conference on Decision and Control, CDC 2008, December 9-11, 2008, Cancún, Mexico*, pages 1708–1713. IEEE, 2008. doi:10.1109/CDC.2008.4738800.

17 Robert A Murphey. Target-based weapon target assignment problems. In *Nonlinear assignment problems*, pages 39–53. Springer, 2000. doi:10.1007/978-1-4757-3155-2_3.

18 Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007. doi:10.1017/CBO9780511800481.

19 Dario Paccagnan and Jason R Marden. Utility design for distributed resource allocation–part II: Applications to submodular, covering, and supermodular problems. *CoRR*, abs/1807.01343, 2018. arXiv:1807.01343.

20 Moshe Shaked and J George Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007. `doi:10.1007/978-0-387-34675-5`.

21 Piotr Skowron, Piotr Faliszewski, and Jérôme Lang. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artif. Intell.*, 241:191–216, 2016. `doi:10.1016/j.artint.2016.09.003`.

22 Maxim Sviridenko, Jan Vondrák, and Justin Ward. Optimal approximation for submodular and supermodular optimization with bounded curvature. *Math. Oper. Res.*, 42(4):1197–1218, 2017. `doi:10.1287/moor.2016.0842`.

23 Jan Vondrák. Submodularity in Combinatorial Optimization. *Univerzita Karlova, Matematicko-Fyzikální Fakulta*, 2007. URL: `https://dspace.cuni.cz/bitstream/handle/20.500.11956/13738/140038775.pdf`.