

# Set Covering with Ordered Replacement: Additive and Multiplicative Gaps

Friedrich Eisenbrand<sup>1</sup>, Naonori Kakimura<sup>\*2</sup>, Thomas Rothvoß<sup>\*\*1</sup>, and  
Laura Sanità<sup>\*\*\*1</sup>

<sup>1</sup> EPFL, Lausanne, Switzerland

<sup>2</sup> University of Tokyo, Japan

**Abstract** We consider set covering problems where the underlying set system satisfies a particular replacement property w.r.t. a given partial order on the elements: Whenever a set is in the set system then a set stemming from it via the replacement of an element by a smaller element is also in the set system.

Many variants of BIN PACKING that have appeared in the literature are such set covering problems with ordered replacement. We provide a rigorous account on the additive and multiplicative integrality gap and approximability of set covering with replacement. In particular we provide a polylogarithmic upper bound on the additive integrality gap that also yields a polynomial time additive approximation algorithm if the linear programming relaxation can be efficiently solved.

We furthermore present an extensive list of covering problems that fall into our framework and consequently have polylogarithmic additive gaps as well.

## 1 Introduction

SET COVER is a prominent combinatorial optimization problem that is very well understood from the viewpoint of multiplicative approximation. There exists a polynomial time factor  $O(\log n)$  approximation for SET COVER [2] and a corresponding hardness result [9]. Also the (multiplicative) integrality gap of the standard linear programming relaxation for SET COVER is known to be  $\Theta(\log n)$  [14].

Let  $\mathcal{S}$  be a family of subsets of  $[n] = \{1, \dots, n\}$ ,  $w : \mathcal{S} \rightarrow \mathbb{R}_+$  be a cost function and let  $\chi(S) \in \{0, 1\}^n$  denote characteristic vector of a set  $S \in \mathcal{S}$ . The SET COVER integer program

$$\min \left\{ \sum_{S \in \mathcal{S}} w(S)x_S \mid \sum_{S \in \mathcal{S}} x_S \cdot \chi(S) \geq \mathbf{1}, x \geq \mathbf{0}, x \text{ integral} \right\} \quad (1)$$

---

\* Supported in part by Grant-in-Aid for Scientific Research and by Global COE Program “The research and training center for new development in mathematics”, MEXT, Japan.

\*\* Supported by the Alexander von Humboldt Foundation within the Feodor Lynen program.

\*\*\* Supported by Swiss National Science Foundation within the project “Robust Network Design”.

and its linear programming relaxation is also in the focus of this paper. However, we are interested in the *additive gap* of a certain class of set covering problems. This additive gap is the *difference* between the optimum value of the integer program (1) and its linear programming relaxation. While there exists an extensive amount of literature on the (multiplicative) gap and (multiplicative) approximation algorithms, the additive gap and algorithms to construct integer solutions that are within the corresponding additive range have received less attention.

Why is it interesting to study the additive integrality gap of set covering problems? Suppose, for example that we know of a certain class of set covering problems that the additive gap is polylogarithmic,  $\log n$  say. If we then, at the same time, know that the optimum solution is at least  $\sqrt{n}$ , then the linear programming relaxation of (1) asymptotically approaches the optimum solution of the integer program yielding a  $(1 + \log n/\sqrt{n})$ -factor approximation algorithm if an integer solution respecting the gap can be efficiently computed.

Two prominent covering problems whose additive gap has been studied are MULTI-EDGE COLORING [12,16] and BIN PACKING.<sup>3</sup> For BIN PACKING, Karmarkar and Karp [13] showed that the additive gap is bounded by  $O(\log^2 n)$  and they also provide a polynomial time algorithm that constructs a solution within this range. There is an extensive amount of literature on variants of BIN PACKING (see e.g. [7,6,5,8,7,3,19,1]). The question whether the SET COVER linear programming relaxations of such variants also exhibit small additive gaps is in the focus of our paper.

It is easy to see that the additive gap of general SET COVER is  $\Theta(n)$ . For example, the VERTEX COVER problem on a disjoint union of triangles exhibits this additive gap. What makes BIN PACKING so special that polylogarithmic additive gaps can be shown to hold? It turns out that it is essentially the fact that in a feasible packing of a bin, we can replace any item by a smaller item and still remain feasible. In the setting of SET COVER this is reflected by the following. There is a partial order  $\preceq$  of the elements that we term *replacement order*. The order is *respected* by  $\mathcal{S}$  if

$$S \in \mathcal{S}, i \in S, j \notin S, j \preceq i \Rightarrow ((S \setminus \{i\}) \cup \{j\}) \in \mathcal{S}$$

We will also consider costs  $w(S)$  of sets in the family  $\mathcal{S}$ . These costs are normalized in the sense that  $w(S) \in [0, 1]$  for each  $S \in \mathcal{S}$ . The costs *respect* the replacement order if  $w(S) \geq w(S')$  whenever  $S'$  is obtained from  $S$  by replacing one element  $i \in S$  with an element  $j \preceq i$  and if  $w(S') \leq w(S)$  for any  $S' \subseteq S$ . Given a family  $\mathcal{S}$  and costs respecting the replacement order  $\preceq$ , the SET COVER WITH ORDERED REPLACEMENT problem is to solve the integer program (1). We denote the optimum value of (1) and its relaxation by  $OPT(\mathcal{S})$  and  $OPT_f(\mathcal{S})$ , respectively. The additive gap of  $\mathcal{S}$  is thus  $OPT(\mathcal{S}) - OPT_f(\mathcal{S})$ .

**Contributions.** We provide a rigorous account on additive and multiplicative integrality gaps and approximability of SET COVER WITH ORDERED REPLACEMENT if  $\preceq$  is a total order. Our main results are as follows.

<sup>3</sup> Even though coined bin “packing”, it is a covering problem.

- We show that the additive gap is bounded by  $O(\log^3 n)$ . This is achieved by the use of suitable *pseudo sizes* and grouping. The pseudo sizes are responsible for the additional  $\log n$  factor compared BIN PACKING. If *natural sizes* are available, our bound matches the  $O(\log^2 n)$  bound for BIN PACKING. The grouping technique itself is not novel albeit appears to be simpler as the one in [13].
- We provide a  $\Omega(\log n)$  lower bound on the additive gap which is in contrast to BIN PACKING, where such a lower bound is not known.
- We show that SET COVER WITH ORDERED REPLACEMENT does not allow an *asymptotic polynomial time approximation scheme (APTAS)*. Also this distinguishes SET COVER WITH ORDERED REPLACEMENT from BIN PACKING.
- We show that the multiplicative gap of SET COVER WITH ORDERED REPLACEMENT is  $\Theta(\log \log n)$ . Also this is in sharp contrast to BIN PACKING, where the multiplicative gap is constant.
- Finally we provide a quasi-polynomial (running time  $n^{O(\log n)}$ ) factor 2 approximation algorithm for SET COVER WITH ORDERED REPLACEMENT.

We also bound the additive integrality gap in the case where the replacement order is not a total order. Recall that the *Dilworth number* of a partial order is the smallest number of disjoint chains that cover all elements. Let us denote the additive integrality gap as:

$$\text{gap}(n, d, k) = \max_{\mathcal{S}, w} \{OPT(\mathcal{S}) - OPT_f(\mathcal{S})\},$$

where the  $\max$  ranges over all set systems over ground set  $[n]$  (and proper cost function  $w : \mathcal{S} \rightarrow [0, 1]$ ), that respect a partial order with Dilworth number  $d$  and contain sets of size at most  $k$ .<sup>4</sup> We show that  $\text{gap}(n, d, k) = O(d^2 \log k \log^2 n)$ . Our result is an *algorithmic* result in the following sense. If the linear programming relaxation of (1) can be efficiently solved, then a solution of (1) respecting the additive gap can be efficiently computed.

We furthermore demonstrate the applicability of our bounds on the additive gap by providing an extensive list of problems from the literature that can be modeled as SET COVER WITH ORDERED REPLACEMENT.

**Related work.** We discuss many BIN PACKING variants that are covered by our framework in Section 5. For many of these problems, there exist *asymptotic polynomial time approximation schemes (APTAS)* [10,5,3,1] or *asymptotic fully polynomial time approximation schemes (AFPTAS)* [13,7,6,8]. An AFPTAS for problem (1) is a polynomial time algorithm (in  $n$  and  $1/\varepsilon$ ) that, given an  $\varepsilon > 0$  computes a solution  $APX(\mathcal{S})$  with  $APX(\mathcal{S}) \leq (1 + \varepsilon)OPT(\mathcal{S}) + f(\varepsilon)$ , where  $f$  is a fixed function. This function  $f$  can be even exponential in  $1/\varepsilon$ , see, e.g. [7,8].

<sup>4</sup> We sometimes abbreviate  $\text{gap}(n, d) := \text{gap}(n, d, n)$  for systems without cardinality restrictions and  $\text{gap}(n) := \text{gap}(n, 1)$  if the partial order is total.

While our additive gap result is incomparable with the quality achieved by an AFPTAS it however sheds some light on how large  $n$  has to be in order for such an AFPTAS to be competitive with our result in combination with a simple constant factor approximation algorithm.

We have an additive  $O(\log^3 n)$  bound for SET COVER WITH ORDERED REPLACEMENT. If we are considering a particular family of instances with  $OPT(\mathcal{S}) \geq \log^4(n)$  for each instance  $\mathcal{S}$ , then this yields an AFPTAS with  $f(\varepsilon) = O((1/\varepsilon)^3)$ . Suppose now that  $OPT(\mathcal{S}) \leq \log^4(n)$  and suppose that there exists an AFPTAS with an exponential  $f$  say  $f(\varepsilon) = 2^{1/\varepsilon}$ . Then the dimension  $n$  has to be *doubly exponential* in  $1/\varepsilon$  before the AFPTAS starts to beat the quality of a factor 2 approximation algorithm.

We would also like to mention recent related work on the additive gap of BIN PACKING. The paper [4] relates a prominent conjecture of Beck from the field of *discrepancy theory* to the question whether the additive gap of BIN PACKING is constant. If Beck's conjecture holds true, then the BIN PACKING gap is constant for 3-partitioning instances. While essentially all results on integrality gaps for BIN PACKING variants in the literature use the sparse support of basic solutions, [17] provides bounds based on probabilistic techniques. The work in [17] provides better additive gaps for example in case of BIN PACKING WITH REJECTION.

## 2 Bounding the additive gap

In this section we provide upper and lower bounds for the additive integrality gap of SET COVER WITH ORDERED REPLACEMENT. The upper bound in case of a total order is  $O(\log^3 n)$  while the lower bound is of order  $\Omega(\log n)$ . This result shows that in order to have a polylogarithmic additive integrality gap, it is sufficient to have a total ordering on the elements, like for classical BIN PACKING.

### 2.1 The upper bound

We first deal with an upper bound on the additive integrality gap. We recall that  $d$  is the Dilworth number of the partial order,  $n$  denotes the number of elements and  $k$  is an upper bound on the cardinality of the sets in the family  $\mathcal{S}$ . We show the following general theorem.

**Theorem 1.** *One has  $gap(n, d, k) = O(d^2 \log k \log^2 n)$ .*

As in [13], we will construct an integer solution to (1) from a fractional one with the claimed additive bound, by doing a sequence of iterations. At each iteration, we will cover part of our elements by rounding down an optimal fractional solution of a proper updated linear program, modify the residual instance, and re-optimize.

More precisely, we will consider the following (more general) linear program

$$\min \left\{ \sum_{S \in \mathcal{S}} w(S)x_S \mid \sum_{S \in \mathcal{S}} x_S \cdot \chi(S) \geq b, x \geq \mathbf{0} \right\}, \quad (2)$$

where  $b \in \mathbb{N}_0^n$  is a non-negative vector. The number  $b_i$  denotes the *multiplicity* of the item  $i$ , i.e., how many times it needs to be covered. The reason of having multiplicity is that during our iterations, we will reduce the number of constraints of the above linear program at the expense of increasing multiplicity for a subset of the items. Of course, when multiplicity comes into play, we allow  $S \in \mathcal{S}$  to be multiset, since e.g. if a set  $S' = \{i, j, h\} \in \mathcal{S}$  and  $i \preceq j \preceq h$ , the replacement order implies  $S = \{i, i, i\}$  to be in  $\mathcal{S}$  as well.

Let us denote the optimum of this linear program with multiplicity and corresponding integer program by  $OPT_f(\mathcal{S}, b)$  and  $OPT(\mathcal{S}, b)$  respectively. As in [13] we will consider an optimal vertex solution  $x^*$  of the linear program (2) and extract the partial covering  $\lfloor x_S^* \rfloor$ ,  $S \in \mathcal{S}$ . This partial covering will leave some elements of the multiset uncovered. This multiset of uncovered elements defines a *residual instance*, encoded by  $b' = \sum_{S \in \mathcal{S}} \{x_S^*\} \cdot \chi(S)$  where  $\{x_S^*\} = x_S^* - \lfloor x_S^* \rfloor$  denotes the fractional part of  $x^*$ . The following relation holds even for arbitrary set covering problems:

$$OPT(\mathcal{S}, b) - OPT_f(\mathcal{S}, b) \leq OPT(\mathcal{S}, b') - OPT_f(\mathcal{S}, b'). \quad (3)$$

The key point of considering a residual instance, is that we will modify it to reduce the number of constraints in the next iteration by using a grouping technique similar to the one given for BIN PACKING in [13]. However, the grouping in [13] crucially relies on the *size of an element* which is part of a BIN PACKING instance, and that is instead missing in our abstract setting. In order to still apply grouping techniques, we will define *pseudo sizes*  $s_i \in ]0, 1]$  for each element  $i$  in our ground set below. These pseudo sizes satisfy the following properties:

- (i)  $s_i \geq s_j$  if  $i \succeq j$ ;
- (ii) We can cover all the elements of any (not necessarily maximal) chain  $C$  at cost  $O(\sum_{i \in C} s_i) + O(\log \frac{1}{s_{\min}})$ , where<sup>5</sup>  $s_{\min} := \min_{i \in [n]} s_i$ .

Notice that the usual size of an element in a BIN PACKING instance is also a pseudo size<sup>6</sup>. Given a vector of pseudo sizes  $s$ , we can define the *total size* of the elements by  $\text{size}(b) = b^T s$ . Suppose now that the largest total pseudo size of a set in  $\mathcal{S}$  is  $\alpha = \max\{\sum_{i \in S} s_i \mid S \in \mathcal{S}\}$ . An important observation is that the total size of the residual instance is bounded by

$$\text{size}(b') = b'^T s = \sum_{S \in \mathcal{S}} \{x^*(S)\} \chi(S)^T s \leq \text{support}(b) \cdot \alpha, \quad (4)$$

where  $\text{support}(b)$  denotes the number of nonzero entries in  $b$ .

We are now ready to state the following Lemma, which generalizes the result of Karmakar and Karp [13].

**Lemma 1.** *Let  $\mathcal{S}$  be a set system on  $[n]$  with cost function  $w : \mathcal{S} \rightarrow [0, 1]$ ,  $\preceq$  be a partial order respected by  $\mathcal{S}$  of Dilworth number  $d$  and let  $b \in \mathbb{N}_0^n$ . Still,*

<sup>5</sup> For notational convenience, we always assume that  $s_{\min} \leq 1/2$  so that  $\log \frac{1}{s_{\min}} \geq 1$ .

<sup>6</sup> Here one can cover a subset of elements even with  $2 \sum_{i \in C} s_i + 1$  bins alone and the minimal size does not need to be considered.

let  $s$  be a vector of pseudo sizes which satisfies properties (i) and (ii). Then  $OPT(\mathcal{S}, b) - OPT_f(\mathcal{S}, b) = O(\alpha \log(1/s_{\min}) \cdot d \log n)$ , where  $\alpha = \max\{\sum_{i \in S} s_i \mid S \in \mathcal{S}\}$ .

*Proof.* Initially, we partition the elements into chains  $C_1, \dots, C_d$  w.r.t. the order  $\succeq$ . The bound follows from iterating the following procedure. First, we replace  $b$  by  $b'$  which encodes its residual instance. According to (3), this does not decrease the additive integrality gap. We then apply the following grouping procedure to each chain  $C_\mu$  with  $\mu = 1, \dots, d$ . The chain  $C_\mu$  is partitioned into *classes*:

$$U_\ell^\mu = \left\{ i \in C_\mu \mid \left(\frac{1}{2}\right)^{\ell+1} < s_i \leq \left(\frac{1}{2}\right)^\ell \right\} \text{ for } \ell = 0, \dots, \lfloor \log(1/s_{\min}) \rfloor.$$

For each such class  $U_\ell^\mu$ , build groups of  $4 \cdot 2^\ell \alpha$  consecutive elements (the elements are always counted with multiplicity), starting from the largest element (the last group could contain less elements), and discard the first and the last group. In this way, we discard at most  $8 \cdot 2^\ell \cdot \alpha$  elements in the class  $U_\ell^\mu$ . Those elements have total size at most  $8 \cdot \alpha$ , hence the total size of discarded elements in chain  $C_\mu$  is bounded by  $8 \cdot \alpha \cdot (\log(1/s_{\min}) + 1)$ . By (ii) we can cover them at cost  $O(\alpha \cdot \log(1/s_{\min}))$ . This amounts to a cost of  $O(d \cdot \alpha \cdot \log(1/s_{\min}))$  to cover all discarded elements of all chains.

Then, we “round-up” the elements in each group to the largest element in this group. In other words, for each group we now consider to have one item type (the largest one, according to the chain) with multiplicity  $4 \cdot 2^\ell \alpha$ . This way, we obtain a new “rounded” instance that is represented by a vector  $b'' \in \mathbb{N}_0^n$ . Since the discarded groups compensate the round-up operation within each group, one has  $OPT_f(\mathcal{S}, b'') \leq OPT_f(\mathcal{S}, b')$ . Also,  $OPT(\mathcal{S}, b') \leq OPT(\mathcal{S}, b'') + O(d \cdot \alpha \cdot \log(1/s_{\min}))$  and thus

$$OPT(\mathcal{S}, b') - OPT_f(\mathcal{S}, b') \leq OPT(\mathcal{S}, b'') - OPT_f(\mathcal{S}, b'') + O(d \cdot \alpha \cdot \log(1/s_{\min})).$$

We will next show that  $\text{support}(b'') \leq \text{support}(b)/2$ . The assertion then follows, since the support of the rounded instance (and hence the corresponding additive integrality gap) will be 1 after  $O(\log n)$  iterations of the above described procedure.

The support of  $b''$  is the number of non-discarded groups. Each non-discarded group  $U_\ell^\mu$  contains a number of elements equal to  $4 \cdot 2^\ell \alpha$ , each of size at least  $\frac{1}{2^{\ell+1}}$ . Then the total size of the group is at least  $2 \cdot \alpha$ . Thus  $2 \cdot \alpha \cdot \text{support}(b'') \leq \text{size}(b')$ . But since  $b'$  encodes a residual instance, one has  $\text{size}(b') \leq \text{support}(b) \cdot \alpha$  from (4). That is,  $\text{support}(b'') \leq \text{support}(b)/2$  follows.  $\square$

Notice that the  $O(\log^2 n)$  upper bound of Karmarkar and Karp [13] for BIN PACKING also follows from this lemma by considering the original sizes given in the instances as pseudo sizes, and setting  $d$ ,  $\alpha$  and all the initial  $b_i$ 's to one. Items of size less than  $1/n$  can be removed and distributed on top of the solution later. If one needs an additional bin, the gap is even bounded by a constant.

We now show how to define good pseudo sizes of the items for *any* given set system  $\mathcal{S}$ . Let

$$\text{size}(i) := \min \left\{ \frac{w(S)}{|S|} \mid S \in \mathcal{S} \text{ contains only elements } j \succeq i \right\}$$

Note that  $\text{size}(i) \geq \text{size}(j)$  holds if  $i \succeq j$ , that is, property (i) of Lemma 1 holds. The next Lemma shows that also (ii) holds as well.

**Lemma 2.** *Let  $\mathcal{S}$  be a set system with replacement property. We can cover all the elements of a (not necessarily maximal) chain  $C$  at cost at most  $2 \sum_{i \in C} \text{size}(i) + O(\log \frac{1}{s_{\min}})$ .*

*Proof.* Again let  $U_\ell = \{i \in C \mid (\frac{1}{2})^{\ell+1} < \text{size}(i) \leq (\frac{1}{2})^\ell\}$  be the  $\ell$ th size class of chain  $C$ . We construct a solution for each size class separately. For class  $\ell$ , consider iteratively the largest uncovered element  $i$  and let  $S(i)$  be a set, defining the quantity  $\text{size}(i)$ , i.e.  $S(i)$  contains only elements that are at least as large as  $i$  and  $\frac{w(S(i))}{|S(i)|} \leq \text{size}(i)$ . By the replacement property, this set  $S(i)$  can be used to cover the largest  $|S(i)|$  many uncovered elements.

Let  $S_1, \dots, S_p$  be the sets selected in this procedure. Note that all the  $S_j$ , but possibly  $S_p$ , cover exactly  $|S_j|$  elements. Then, since  $\frac{w(S_j)}{|S_j|} \leq 2 \cdot \text{size}(i)$  for every  $i \in S_j$ , we have

$$\sum_{j=1}^p w(S_j) = \sum_{j=1}^{p-1} \sum_{i \in S_j} \frac{w(S_j)}{|S_j|} + w(S_p) \leq \sum_{i \in U_\ell} 2 \cdot \text{size}(i) + 1.$$

Summing over the  $(\log \frac{1}{s_{\min}} + 1)$  many size classes then gives the claim.  $\square$

We are ready to prove Theorem 1.

*Proof (Proof of Theorem 1).* Let  $\mathcal{S}$  be a set system on  $[n]$ , and  $\preceq$  be a partial order respected by  $\mathcal{S}$ . We may assume that  $\preceq$  consists of  $d$  incomparable chains  $C_1, \dots, C_d$ . We define  $s_i = \text{size}(i)$  for any element  $i$ .

We first claim that we can discard all items  $i$  with  $s_i < \frac{1}{n^2}$ : in fact, by the definition of the pseudo sizes any such element is contained in a set of cost at most  $\frac{1}{n}$ , hence all such tiny elements can be covered at a total cost of 1.

Let  $S \in \mathcal{S}$  and consider an element  $i$  which is the  $j$ th largest in  $S \cap C_\mu$ . By the definition of the pseudo sizes,  $\text{size}(i) \leq \frac{w(S)}{j}$ . Thus

$$\sum_{i \in S} \text{size}(i) \leq d \sum_{j=1}^k \frac{w(S)}{j} \leq 2d \log k.$$

Therefore, it follows from Lemma 1 (by setting  $b = \mathbf{1}$ ) that  $\text{gap}(n, d, k) = O(d^2 \log k \log^2 n)$ , since  $\alpha = 2d \log k$  and  $s_{\min} \geq 1/n^2$ .  $\square$

We want to remark that the above result is constructive in the sense that once a fractional solution  $x$  and the partial order are given, a feasible integer solution matching the bound of Theorem 1 can be computed in polynomial time.

If all costs are one, i.e.,  $w(S) = 1$  for each  $S \in \mathcal{S}$ , then we have  $\text{size}(i) \geq \frac{1}{k}$ . Thus we have the following corollary.

**Corollary 1.** *If all costs of feasible sets are one, one has  $\text{gap}(n, d, k) = O(d^2 \log^2 k \log n)$ .*

When  $d = 1$ , this theorem says  $\text{gap}(n, 1, k) = O(\log^2 k \log n)$ , which is better than the result of [4]:  $\text{gap}(n, 1, k) = O(k \log n)$ .

## 2.2 The lower bound

In this section, we give a lower bound for the additive integrality gap of set systems with replacement property. For simplicity, we first assume that  $\preceq$  consists of only one chain.

**Lemma 3.** *One has  $\text{gap}(n) \geq \Omega(\log n)$ .*

*Proof.* Let  $m$  be a parameter, that we determine later. Define a unit cost set system  $\mathcal{S}$  as follows. For any  $\ell \in \{1, \dots, m\}$ , introduce  $3 \cdot 100^\ell$  many  $\ell$ -level elements  $U_\ell$ . Hence  $U := \bigcup_{\ell=1}^m U_\ell$  with  $|U_\ell| = 3 \cdot 100^\ell$  is the set of all elements. We choose  $U_1 \succeq U_2 \succeq \dots \succeq U_m$ <sup>7</sup> and an arbitrary order within every  $U_\ell$ . Any set  $S$  of at most  $2 \cdot 100^\ell$  many  $\ell$ -level elements forms an  $\ell$ -level set, e.g.  $\mathcal{S} = \bigcup_{\ell=1}^m \{S \subseteq U_\ell \cup \dots \cup U_m \mid |S| \leq 2 \cdot 100^\ell\}$ . By taking 3  $\ell$ -level sets to the fractional extend of  $\frac{1}{2}$ , we can cover all  $\ell$ -level elements, hence  $\text{OPT}_f(\mathcal{S}) \leq \frac{3}{2} \cdot m$ . It remains to lower bound  $\text{OPT}(\mathcal{S})$ . Let  $n_\ell$  be the number of  $\ell$ -level sets in any integer solution. To cover all level  $\ell$ -level elements we must either have  $n_\ell \geq 2$ , or  $\sum_{j < \ell} n_j \cdot 2 \cdot 100^j \geq 100^\ell$ , i.e.  $4 \sum_{j < \ell} n_j \cdot 100^{j-\ell} \geq 2$ . In any case, the sum of the left hand sides must be at least 2:  $n_\ell + 4 \sum_{j < \ell} n_j \cdot 100^{j-\ell} \geq 2$ . We add up this inequality for  $\ell = 1, \dots, m$  and obtain (\*)

$$\sum_{\ell=1}^m n_\ell \cdot \underbrace{\left(1 + 4 \sum_{i \geq 1} \frac{1}{100^i}\right)}_{=\frac{103}{99}} \geq \sum_{\ell=1}^m \left(n_\ell + 4 \sum_{j < \ell} n_j \cdot 100^{j-\ell}\right) \stackrel{(*)}{\geq} 2m$$

This gives  $\text{OPT}(\mathcal{S}) - \text{OPT}_f(\mathcal{S}) \geq \sum_{\ell=1}^m n_\ell - \frac{3}{2}m \geq 2 \frac{99}{103}m - \frac{3}{2}m > 0.4 \cdot m$ . The number of elements in the instance is  $n = \sum_{\ell=1}^m 3 \cdot 100^\ell$ , hence  $m = \Omega(\log n)$ .  $\square$

More generally, the following holds:

**Theorem 2.**  *$\text{gap}(n, d) \geq \Omega(d \cdot \log(n/d))$ .*

*Proof.* Apply the construction from Lemma 3 to obtain families  $\mathcal{S}_1, \dots, \mathcal{S}_d$ , each on a disjoint set of  $n/d$  elements. Then the union  $\bigcup_{j=1}^d \mathcal{S}_j$  has Dilworth number  $d$  and the claimed additive gap.  $\square$

<sup>7</sup> We abbreviate  $U \succeq U' \Leftrightarrow \forall i \in U, j \in U' : i \succeq j$ .



### 3 Approximation & intractability

In Section 5, we mention many variants of BIN PACKING that admit an APTAS (or even an AFPTAS) and fall into our general framework of SET COVER WITH ORDERED REPLACEMENT. It is thus natural to ask whether the SET COVER WITH ORDERED REPLACEMENT itself has an APTAS. This cannot be expected for arbitrary partial orders. In this section we show that an APTAS does not exist, even if the replacement order is a total order. On the positive side however, we show that there exists a quasi-polynomial time factor 2 approximation algorithm in this case. This is obtained by first rounding the instance to  $O(\log n)$  different item types and then solving the rounded instance exactly by dynamic programming. However, due to lack of space, we postpone the proof to the full version of this paper.

From now on, we restrict our view exclusively on set systems, respecting a total order  $\preceq$ . To define such a set system with unit-costs, it will be convenient to consider just the set of *generators*. These are the sets that are maximal with respect to the order. More formally, if there is an injective map  $\varphi : S \rightarrow S'$  with  $i \preceq \varphi(i)$  for all  $i \in S$  (i.e. we can obtain a set  $S$  from  $S'$  by applying the replacement rule), then we say that  $S'$  *dominates*  $S$ . Hence a set family  $\mathcal{S}$  is called the set of *generators* for the set system

$$g(\mathcal{S}) = \{S \subseteq [n] \mid \exists S' \in \mathcal{S} : S' \text{ dominates } S\}$$

Hence, if  $\mathcal{S}$  is an arbitrary set family, by definition  $g(\mathcal{S})$  respects the replacement rule. For a proof of the next proposition we refer to the full version of this paper.

**Proposition 1.** *Let  $\mathcal{S} \subseteq 2^{[n]}$  be a family of sets and  $\succeq$  the total order with  $1 \succeq \dots \succeq n$ .*

- i) If  $\mathcal{S}' \subseteq g(\mathcal{S})$  is a feasible solution (i.e.  $\bigcup_{S \in \mathcal{S}'} S = [n]$ ) then  $\sum_{S \in \mathcal{S}'} |S \cap \{1, \dots, i\}| \geq i$  for  $i = 1, \dots, n$ .*
- ii) If  $\mathcal{S}' \subseteq \mathcal{S}$  are generators with  $\sum_{S \in \mathcal{S}'} |S \cap \{1, \dots, i\}| \geq i \forall i \in [n]$ , then sets  $\mathcal{S}'$  can be replaced by dominated ones which form a feasible solution of the same cardinality.*

#### 3.1 Ruling out an APTAS

We will now see that unless  $\mathbf{P} = \mathbf{NP}$ , there is no APTAS for a generic problem defined on a set system that respects a total order.

**Theorem 3.** *For every  $\varepsilon > 0$  and any  $C > 0$  there is a generic problem with unit-cost sets respecting a total order for which it is  $\mathbf{NP}$ -hard to find a solution of cost  $(\frac{3}{2} - \varepsilon)OPT + C$ .*

*Proof.* We will prove the theorem by constructing a set system such that for any fixed integer  $k > 0$ , it is  $\mathbf{NP}$ -hard to distinguish whether an optimum solution consists of at most  $2k$  or at least  $3k$  sets. Choosing  $k := k(\varepsilon, C)$  large enough then gives the claim.

To establish this hardness result, we will use a reduction from the **NP**-hard PARTITION [11] problem. An instance  $\mathcal{I}$  of PARTITION, is given by a set of  $n$  items with sizes  $a_1 \geq a_2 \geq \dots \geq a_n$ , and asks for a partition of the items into two sets  $A_1$  and  $A_2$ , such that  $\sum_{j \in A_1} a_j = \sum_{j \in A_2} a_j =: A$ . Given such an instance, we create groups  $L_1, \dots, L_k$ , where the group  $L_p$  contains  $n^{2p}$  copies of item  $i$ , i.e.  $L_p = \{v_{p,i}^j \mid i = 1, \dots, n; j = 1, \dots, n^{2p}\}$ . Note that the total number of elements is  $N := n^{O(k)}$ . We define a total order with  $L_1 \succeq \dots \succeq L_k$  and  $v_{p,i}^j \succeq v_{p,i'}^{j'}$  whenever  $i < i'$  (and  $v_{p,i}^1 \succeq \dots \succeq v_{p,i}^{n^{2p}}$  for the sake of completeness). Let  $S(I, p) := \{v_{p,i}^j \mid i \in I; j = 1, \dots, n^{2p}\}$  be the set induced by  $I \subseteq [n]$  in group  $L_p$ . We define generators

$$\mathcal{S} = \left\{ S(I, p) \mid \forall p = 1, \dots, k; \forall I \subseteq [n] : \sum_{i \in I} a_i \leq A \right\}$$

*Completeness:*  $\mathcal{I} \in \text{PARTITION} \Rightarrow \text{OPT}(g(\mathcal{S})) \leq 2k$ . Let  $I \subseteq [n]$  with  $\sum_{i \in I} a_i = A$  be a PARTITION solution. Then the  $2k$  sets of the form  $S([n] \setminus I, p), S(I, p)$  for  $p = 1, \dots, k$  cover all  $N$  elements.

*Soundness:*  $\mathcal{I} \notin \text{PARTITION} \Rightarrow \text{OPT}(g(\mathcal{S})) \geq 3k$ . We may assume that  $3k < n$ . Now suppose for the sake of contradiction, there is no PARTITION solution, but  $\mathcal{S}' \subseteq \mathcal{S}$  is a family of less than  $3k$  generating sets, dominating a feasible solution, satisfying the second condition in Prop. 1. Then there must be a group  $L_p$  such that  $\mathcal{S}'$  contains generators  $S(I_1, p), \dots, S(I_m, p)$  with  $m \leq 2$ . Then from Prop. 1 we obtain that

$$i \cdot n^{2p} \stackrel{\text{Prop. 1}}{\leq} \sum_{S \in \mathcal{S}'} |S \cap (L_1 \cup \dots \cup L_{p-1} \cup S(\{1, \dots, i\}, p))| \leq 3k \cdot n \cdot n^{2p-2} + \sum_{\ell=1}^m n^{2p} |I_\ell \cap \{1, \dots, i\}|$$

Hence:  $\sum_{\ell=1}^m |I_\ell \cap \{1, \dots, i\}| \geq \lceil i - \frac{3k}{n} \rceil = i$ , since the left hand side is integral and  $3k < n$ . Since  $\sum_{\ell=1}^m |I_\ell \cap \{1, \dots, i\}| \geq i$  for all  $i \in [n]$ , we conclude by applying again Prop. 1, that elements in  $I_1, \dots, I_m$  can be replaced by smaller ones and obtain  $I'_1, \dots, I'_m$  such that still  $\sum_{i \in I'_\ell} a_i \leq A$  but  $\bigcup_{\ell=1}^m I'_\ell = [n]$ . Since  $m \leq 2$ , this is a contradiction.  $\square$

## 4 Multiplicative integrality gaps

Next, we show a  $\Theta(\log \log n)$  multiplicative gap for the case of a total order.

**Lemma 4.** *Let  $\mathcal{S}$  any set family on  $n$  elements respecting a total order and let  $w : \mathcal{S} \rightarrow [0, 1]$  be a cost function. Then  $\text{OPT}(\mathcal{S}) \leq O(\log \log n) \cdot \text{OPT}_f(\mathcal{S})$ .*

*Proof.* Consider consecutive groups  $L_0, \dots, L_k$  such that  $|L_i| = 2^i$  (group  $L_k$  might have fewer elements),  $k \leq \log n$  and all elements in  $L_i$  are larger than those in  $L_{i+1}$ . Let  $x \in [0, 1]^{\mathcal{S}}$  be a fractional solution. We buy set  $S$  independently with probability  $\lambda \cdot x_S$  where  $\lambda := \max\{8 \cdot \log(4 + 4 \log n), 4\}$  (in fact we may assume that  $\lambda \cdot x_S \leq 1$ , otherwise we buy  $\lfloor \lambda x_S \rfloor$  sets deterministically and then

another set with probability  $\lambda x_S - \lfloor \lambda x_S \rfloor$ ). Let  $X_S$  be the indicator variable, telling whether we bought  $S$  or not. Define  $E_i := \{\sum_S X_S \cdot |S \cap L_i| < 2 \cdot |L_i|\}$  as the event that we bought less than two times enough sets for the  $i$ th group. Recall the following Chernov bound (see<sup>8</sup> e.g. [15]):

*Let  $Y_1, \dots, Y_m$  be independent random variables with  $Y_i \in [0, 1]$ ,  $Y = \sum_{i=1}^m Y_i$  and  $0 < \delta < 1$ . Then  $\Pr[Y \leq (1 - \delta)E[Y]] \leq e^{-E[Y]\delta^2/2}$ .*

Applying Chernov bound with  $\delta := 1/2$ ,  $Y_S := X_S \frac{|S \cap L_i|}{|L_i|}$  and  $E[Y] = \lambda$ , we obtain

$$\Pr[E_i] = \Pr\left[\sum_{S \in \mathcal{S}} X_S \cdot |S \cap L_i| < 2 \cdot |L_i|\right] \leq \Pr\left[\sum_{S \in \mathcal{S}} Y_S < (1 - \delta)\lambda\right] \leq e^{-\lambda/8} \leq \frac{1}{4(1 + \log n)}.$$

By the union bound,  $\Pr[E_0 \cup \dots \cup E_k] \leq \frac{1}{4}$ . Furthermore  $\Pr[\sum_{S \in \mathcal{S}} X_S w(S) > 4\lambda \cdot OPT_f] \leq \frac{1}{4}$  by Markov's inequality. Overall, with probability at least  $1/2$ , we obtain an integer solution  $\mathcal{S}' = \{S \in \mathcal{S} \mid X_S = 1\}$  of cost at most  $O(\log \log n) \cdot OPT_f$  that reserves at least  $2|L_i|$  slots for elements in  $L_i$ . Those slots are enough to cover all elements in  $L_{i+1}$ .  $\square$

Note that the result in Lemma 4 provides a randomized polynomial time algorithm, provided that a near-optimal fractional solution  $x$  can be obtained.

**Lemma 5.** *There exists a set system  $\mathcal{S}' \subseteq 2^{[n]}$  with unit cost sets and respecting a total order such that  $OPT(\mathcal{S}') \geq \Omega(\log \log n) \cdot OPT_f(\mathcal{S}')$ .*

*Proof.* Let  $k \in \mathbb{N}$  be a parameter. To construct our instance, we use as starting point a SET COVER instance defined by a set system  $\mathcal{C}$  with  $2^k - 1$  sets  $C_1, \dots, C_{2^k - 1}$  and  $2^k - 1$  elements  $U = \{1, \dots, 2^k - 1\}$  such that one needs at least  $k$  sets to cover all elements, while  $OPT_f(\mathcal{C}) \leq 2$  (see Example 13.4 in the book of Vazirani [18]).

For every element  $i \in U$ , create groups  $L_i$  with  $|L_i| = (2k)^i$ . For any  $C_j$  in the original set system, define a set  $S_j := (\bigcup_{i \in C_j} L_i)$  with unit cost, and let  $\mathcal{S} = \{S_1, \dots, S_{2^k - 1}\}$ . In other words, we take a standard SET COVER instance and replace the  $i$ th element by  $(2k)^i$  elements. We define a total order such that all items in  $L_i$  are larger than those in  $L_{i+1}$  (and any order within the groups). The claim is that the set system  $\mathcal{S}' := g(\mathcal{S})$ , which is generated by the sets  $\mathcal{S}$ , has a covering integrality gap of at least  $k/2$ . First note that still  $OPT_f(\mathcal{S}') \leq 2$ . Now suppose for contradiction that there are generators (after reindexing)  $S_1, \dots, S_m \subseteq \mathcal{S}$  with  $m < k$  that satisfy the condition in Prop. 1. Since  $m < k$ , there must be an index  $i$  such that  $i \notin \{C_1 \cup \dots \cup C_m\}$ . Then

$$(2k)^i \leq \sum_{\ell=1}^i |L_\ell| \stackrel{\text{Prop. 1}}{\leq} \sum_{j=1}^m |S_j \cap (L_1 \cup \dots \cup L_i)| = \sum_{j=1}^m \sum_{\ell \in C_j, \ell \leq i} (2k)^\ell \leq m \cdot 2(2k)^{i-1} = \frac{m}{k}(2k)^i$$

Rearranging yields that  $m \geq k$ . Note that the number of elements in the system  $\mathcal{S}'$  is  $n = \sum_{i=1}^{2^k - 1} (2k)^i \leq 2 \cdot (2k)^{2^k}$ , hence  $k = \Omega(\log \log n)$ .  $\square$

<sup>8</sup> To be precise, the claim in [15] is for 0/1 distributed random variables, but the same proof goes through if  $0 \leq Y_i \leq 1$ .

## 5 Applications

We now demonstrate the versatility of our framework by establishing small additive integrality gaps for several BIN PACKING variants from the literature, listed in the following. We refer to the full version of this paper for the proofs of the bounds provided below.

**CARDINALITY BIN PACKING.** Here we are given a BIN PACKING instance  $s_1, \dots, s_n$  with an additional parameter  $k \in \mathbb{N}$ . A subset of items  $S$  can be assigned to a bin only if  $\sum_{i \in S} s_i \leq 1$  and  $|S| \leq k$ . There is an AFPTAS due to Epstein & Levin [7]. With our technique, we obtain:  $OPT(\mathcal{S}) - OPT_f(\mathcal{S}) = O(\log k \log n)$ .

**OPEN END BIN PACKING.** We are here given a BIN PACKING instance  $s_1, \dots, s_n$ , but a set  $S \subseteq [n]$  is feasible if  $\sum_{i \in S \setminus \{j\}} s_i \leq 1$  holds for every  $j \in S$ . There is an AFPTAS for this variant by Epstein & Levin [6]. We can prove that:  $OPT(\mathcal{S}) - OPT_f(\mathcal{S}) = O(\log^2 n)$ .

**BIN PACKING WITH GENERAL COST STRUCTURE.** We are given item sizes  $s_1, \dots, s_n$ , and a cost function  $f : \{0, \dots, n\} \rightarrow [0, 1]$ , which is a monotonically nondecreasing concave function with  $f(0) = 0$ . The goal is to find sets  $S_1, \dots, S_p$  to cover the items such that  $\sum_{i \in S_j} s_i \leq 1$  for every  $j = 1, \dots, p$  and the total cost  $\sum_{i=1}^p f(|S_i|)$  is minimized. This problem admits an AFPTAS [8]. We prove:  $OPT(\mathcal{S}) - OPT_f(\mathcal{S}) = O(\log^2 n)$ .

**GENERALIZED COST VARIABLE SIZED BIN PACKING.** We are given item sizes  $s_1, \dots, s_n$ , and bin types  $B_1, \dots, B_k$  each one with a different capacity  $a_1, \dots, a_k$  and a different cost  $c_1, \dots, c_k \in [0, 1]$ . The goal is to select a minimum cost subset of bins to pack the items, where a subset of items can be packed into a bin  $B_i$  if the sum of their sizes does not exceed the bin capacity  $a_i$ . See [5] for an APTAS. Our result is:  $OPT(\mathcal{S}) - OPT_f(\mathcal{S}) = O(\log^3 n)$ .

**BIN PACKING WITH REJECTION.** We are given a BIN PACKING instance  $s_1, \dots, s_n$ , but additionally any item  $i$  has a rejection cost  $c_i$ . An item can either be packed, or discarded at cost  $c_i$ . See [7] for an AFPTAS. We have:  $OPT(\mathcal{S}) - OPT_f(\mathcal{S}) = O(\sqrt{n} \log^{3/2} n)$ .

**TRAIN DELIVERY.** We are given  $n$  items, each having a size  $s_i$  and position  $p_i \in [0, 1]$ . The goal is to transport all items to a depot, located at 0, using a unit capacity train and minimizing the total tour length. The authors in [3] give an APTAS. We can prove that:  $OPT(\mathcal{S}) - OPT_f(\mathcal{S}) = O(\sqrt{n} \log^{3/2} n)$ .

**$m$ -DIMENSIONAL VECTOR PACKING.** Let  $V$  be a set of vectors  $v_1, \dots, v_n \in [0, 1]^m$ . The goal is to partition  $V$  into bins  $B_1, \dots, B_k$ , such that  $k$  is minimized and  $\sum_{i \in B_j} v_i \leq \mathbf{1}$ . We say that  $v_i \preceq v_j$ , if  $v_j$  is componentwise not smaller than  $v_i$ . The Dilworth number of  $V$  is then the smallest  $d$  such that  $V$  can be partitioned into  $V^1, \dots, V^d$ , such that:  $\forall v_i, v_j \in V^h$ , either  $v_i \preceq v_j$  or  $v_j \preceq v_i$ . If there is no bound on  $d$ , there is no APTAS possible already in 2-dimensions [19]. Differently, for constant  $d$  there is an APTAS given in [1]. Our result is:  $OPT(\mathcal{S}) - OPT_f(\mathcal{S}) = O(d^2 \log^3 n)$ .

## References

1. A. Caprara, H. Kellerer, and U. Pferschy. Approximation schemes for ordered vector packing problems. *Naval Research Logistics*, 50:58–69, 2003.
2. V. Chvátal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4(3):233–235, 1979.
3. A. Das, C. Mathieu, and S. Mozes. The train delivery problem- vehicle routing meets bin packing, 2010.
4. F. Eisenbrand, D. Pálvölgyi, and T. Rothvoß. Bin packing via discrepancy of permutations. In *Symposium on Discrete Algorithms, SODA 2011*. SIAM, 2011.
5. L. Epstein and A. Levin. An APTAS for generalized cost variable-sized bin packing. *SIAM Journal on Computing*, 38(1), 2008.
6. L. Epstein and A. Levin. Asymptotic fully polynomial approximation schemes for variants of open-end bin packing. *Inf. Process. Lett.*, 109(1):32–37, 2008.
7. L. Epstein and A. Levin. AFPTAS results for common variants of bin packing: A new method to handle the small items. *SIAM Journal on Optimization*, 2010. to appear.
8. L. Epstein and A. Levin. Bin packing with general cost structures. *Mathematical Programming*, to appear, 2010.
9. Uriel Feige. A threshold of  $\ln n$  for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.
10. W. Fernandez de la Vega and G. S. Lueker. Bin packing can be solved within  $1 + \epsilon$  in linear time. *Combinatorica*, 1(4):349–355, 1981.
11. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, New York, 1979.
12. J. Kahn. Asymptotics of the chromatic index for multigraphs. *Journal of Combinatorial Theory. Series B*, 68(2):233–254, 1996.
13. N. Karmarkar and R. M. Karp. An efficient approximation scheme for the one-dimensional bin-packing problem. In *23rd annual symposium on foundations of computer science (Chicago, Ill., 1982)*, pages 312–320. IEEE, New York, 1982.
14. L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383–390, 1975.
15. M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, Cambridge, 2005. Randomized algorithms and probabilistic analysis.
16. M. Plantholt. A sublinear bound on the chromatic index of multigraphs. *Discrete Mathematics*, 202(1-3):201–213, 1999.
17. T. Rothvoß. The entropy rounding method in approximation algorithms. *submitted*, 2010.
18. V. V. Vazirani. *Approximation Algorithms*. Springer-Verlag, 2001.
19. G. J. Woeginger. There is no asymptotic PTAS for two-dimensional vector packing. *Information Processing Letters*, 64(6):293 – 297, 1997.