The approximability and integrality gap of interval stabbing and independence problems

Shalev Ben David* Elyot Grant† Will Ma‡ Malcolm Sharpe§

May 8, 2012

Abstract

Motivated by problems such as rectangle stabbing in the plane, we study the *minimum hitting set* and *maximum independent set* problems for families of *d-intervals* and *d-union-intervals*. We obtain the following: (1) an LP-relative d-approximation for the hitting set problem on d-intervals, proven using standard techniques; (2) constructions yielding asymptotically tight lower bounds on the integrality gap of the natural linear programming relaxations of all problems; (3) a proof that the approximation ratios for independent set on families of 2-intervals and 2-union-intervals can be improved to match tight duality gap bounds obtained via topological arguments, if one has access to an oracle for PPAD-complete Borsuk-Ulam-style fixed-point problems.

1 Introduction

In this work, we examine a family of NP-hard packing and covering problems. Our study is motivated by the *minimum rectangle stabbing* problem, in which we are given a family $\mathcal{H}$ of axis-aligned rectangles in the plane, and the goal is to find a minimum-cardinality family of horizontal and vertical lines that intersect (or 'stab') each rectangle in $\mathcal{H}$ at least once. Viewing this as a geometric covering problem, we also consider the related 'dual' geometric packing problem of finding a *maximum conflict-free subset*, where the goal is to find a maximum subset of $\mathcal{H}$ containing no pair of rectangles that can be stabbed by a single horizontal or vertical line.

The rectangle stabbing and conflict-free subset problems have many applications. The rectangles themselves can be the bounding boxes of arbitrary connected objects in the plane, so applications need not be limited to problems involving rectangles. The rectangle stabbing problem can directly encode the problem of optimally subdividing the plane into a grid of axis-aligned cells so as to separate a given family of points, with applications to fault-tolerant sensor networks [3] and resource allocation in parallel processing systems [7]. The maximum conflict-free subset problem, and its higher dimensional analogues, are relevant to areas such as resource allocation, scheduling, and computational biology [2]. The properties of certain rectangle stabbing instances are also of theoretical interest in combinatorics [19].

Given a family $\mathcal{H}$ of axis-aligned rectangles, we write $\rho(\mathcal{H})$ for the minimum cardinality of a family of lines stabbing it, and $\alpha(\mathcal{H})$ for the maximum size of a conflict-free subset. It is clear that $\alpha(\mathcal{H}) \leq \rho(\mathcal{H})$. As in many geometric packing-covering dual problems, there is a bound in the other direction. In 1994, Tardos proved that $\rho(\mathcal{H}) \leq 2\alpha(\mathcal{H})$, which is easily seen to be tight [18]. However, all known proofs of Tardos's result rely on topological fixed-point theorems, and consequently do not seem to lead to polynomial-time approximations. In fact, only a 4-approximation is known for the maximum conflict-free subset problem [2],

*Comp. Sci. and A.I. Lab, MIT, `shalev@mit.edu`
†Comp. Sci. and A.I. Lab, MIT, `elyot@mit.edu`
‡Operations Research Center, MIT, `willma@mit.edu`
§Department of Combinatorics and Optimization, University of Waterloo, `sharpe.malcolm@gmail.com`

despite the fact that we can establish the optimal objective value to within a factor of 2 by solving a linear program. Improving upon this remains an important open problem.

In this paper, we obtain results for generalized versions of the stabbing and conflict-free subset problems. We examine the standard linear programming relaxations for hitting set and independent set problems involving *d-intervals*, and establish asymptotically tight upper and lower bounds on their integrality gaps. These bounds imply that no LP-relative approximation algorithm can obtain a factor below 2 for either the rectangle stabbing or the maximum conflict-free subset problems. Additionally, we establish some interesting theoretical consequences of topological methods such as Tardos's. For example, we show that the maximum conflict-free subset problem admits a 2-approximation if one has access to an oracle for a PPAD-complete problem.

This article proceeds follows: in the remainder of the current section, we define the generalized problems that we study, explain the current state of the art, and describe our contribution. Section 2 contains our integrality gap upper and lower bounds, and Section 3 contains algorithmic results that depend on PPAD oracles.

1.1 Preliminaries

We begin by defining generalized versions of the rectangle stabbing and conflict-free subset problems. For $d \in \mathbb{N}$, a *d-interval* I is a union of d non-empty compact intervals $I^1, \ldots, I^d \subset \mathbb{R}$. The input to all problems we consider will be a finite collection $\mathcal{H}$ of d-intervals, represented explicitly. We say that a subset of $\mathcal{H}$ is *independent* if its members are pairwise disjoint, and a set $X \subseteq \mathbb{R}$ is a *hitting set* for $\mathcal{H}$ if it intersects every member of $\mathcal{H}$. We define the hypergraph $G_{\mathcal{H}} = (V, E)$ where $V = \mathcal{H}$ and E consists of all subsets $\mathcal{I} \subseteq \mathcal{H}$ such that there is a point $p \in \mathbb{R}$ that hits *exactly* the intervals in $\mathcal{I}$. Such a point p shall be called a *representative* of the hyperedge $\mathcal{I} \in E$, and we let $P(\mathcal{H})$ denote a set containing an arbitrary representative for each distinct edge in $G_{\mathcal{H}}$. Note that $|P(\mathcal{H})| \leq 2d|\mathcal{H}|$ as there are at most $2d|\mathcal{H}|$ interval endpoints. We call $G_{\mathcal{H}}$ a *d-interval hypergraph*, and observe that d-interval hypergraphs generalize d-regular hypergraphs (which are obtained when each d-interval in $\mathcal{H}$ is simply d points). We denote by $\alpha(\mathcal{H})$ the maximum size of an independent set in $\mathcal{H}$, and denote by $\rho(\mathcal{H})$ the minimum size of a hitting set for $\mathcal{H}$, in analogy with the usual notation of $\alpha(G)$ and $\rho(G)$ for the maximum independent set size and minimum edge cover size of a hypergraph G.

Special cases such as the rectangle stabbing problem arise when we impose structural restrictions on $\mathcal{H}$. If $\{J^i\}_{i=1}^d$ is a family of disjoint intervals and each d-interval $I = \cup_{i=1}^d I^i$ in $\mathcal{H}$ satisfies $I^i \subseteq J^i$ for all i, then $\mathcal{H}$ is known as a collection of *d-union-intervals*, and $G_{\mathcal{H}}$ is known as a *d-union hypergraph*. A set of d-union-intervals may also be called *d-track-intervals*, with the idea that each d-interval contains a piece from one of d different 'tracks', i.e. disjoint copies of $\mathbb{R}$ [9]. The rectangle stabbing and minimum conflict-free subset problems correspond precisely to the minimum edge cover and maximum indepedent set problems for 2-union hypergraphs, as each 'track' can be mapped onto a separate Euclidean dimension. In general, one can think of the hitting set problem for d-union-intervals as the problem of attempting to hit a family of d-dimensional 'boxes' using a minimum number of 'walls', each orthogonal to one of the coordinate axes.

For 1-interval hypergraphs (which are the same as 1-union hypergraphs), the independent set and edge cover problems can both be solved in polynomial time via simple greedy algorithms that perform a left-to-right sweep across the intervals. However, even for 2-union hypergraphs, the independent set and edge cover problems are both APX-hard. Nagashima and Yamazaki, and independently Bar-Yehuda et al., have shown the conflict-free subset problem to be APX-hard [2, 14], even when the rectangles are all unit squares with integer vertices. Kovaleva and Spieksma show that the rectangle stabbing problem is APX-hard even when each rectangle is of the form $[x, x+1] \times [y, y]$ for integers x and y [11].

For a hypergraph G, the relations $\alpha(G) \leq \rho(G)$ and $\rho(G) \leq O(\log |V|) \cdot \alpha(G)$ are well known, with the latter being tight for general hypergraphs. However, Kaiser proves that $\frac{\rho(G_{\mathcal{H}})}{\alpha(G_{\mathcal{H}})}$ is upper bounded by $d^2 - d + 1$ for d-interval hypergraphs and $d^2 - d$ for d-union hypergraphs (for $d \geq 2$), regardless of $|V|$ [10]. His approach generalizes the topological method of Tardos, which originally established a tight upper bound for the $d = 2$ case [18]. In a one-page paper, Alon shows that an upper bound of $2d^2$ can be established without topological methods by applying Turán's theorem [1]. The best known lower bounds for large d are $\Omega(\frac{d^2}{\log d})$ and $\Omega(\frac{d^2}{\log^2 d})$ for d-interval and d-union hypergraphs respectively [13].

1.2 Overview of Results

We use the term *duality gap* to denote a quantity $\sup_{\mathcal{H}} \frac{\rho(\mathcal{H})}{\alpha(\mathcal{H})}$, where $\mathcal{H}$ ranges over all collections of d-intervals or d-union-intervals. In an effort to study the duality gap, we examine standard linear programming relaxations for the hitting set and independent set problems. The standard LP relaxation for the maximum independent set problem corresponds to the *maximum fractional independent set problem*, and can be written as follows:

$$\begin{aligned} \max \quad & \sum_{I \in \mathcal{I}} x_I \\ \text{s.t.} \quad & \sum_{I \ni p} x_I \leq 1 \quad \forall\, p \in P(\mathcal{H}) \\ & x_I \geq 0 \qquad \forall\, I \in \mathcal{H}. \end{aligned} \tag{1}$$

A corresponding dual linear program for the *minimum fractional hitting set problem* is as follows:

$$\begin{aligned} \min \quad & \sum_{p \in P(\mathcal{H})} y_p \\ \text{s.t.} \quad & \sum_{p \in I} y_p \geq 1 \quad \forall\, I \in \mathcal{H} \\ & y_p \geq 0 \qquad \forall\, p \in P(\mathcal{H}) \end{aligned} \tag{2}$$

If $\alpha^*(\mathcal{H})$ is the optimal objective value for (1) and $\rho^*(\mathcal{H})$ is the optimal objective value for (2), then we have $\alpha(\mathcal{H}) \leq \alpha^*(\mathcal{H}) = \rho^*(\mathcal{H}) \leq \rho(\mathcal{H})$ and we can write

$$\sup_{\mathcal{H}} \frac{\rho(\mathcal{H})}{\alpha(\mathcal{H})} \leq \sup_{\mathcal{H}} \frac{\rho(\mathcal{H})}{\rho^*(\mathcal{H})} \cdot \sup_{\mathcal{H}} \frac{\alpha^*(\mathcal{H})}{\alpha(\mathcal{H})}$$

The quantity $\sup_{\mathcal{H}} \frac{\rho(\mathcal{H})}{\rho^*(\mathcal{H})}$ is called the *integrality gap* of the minimum hitting set problem (for d-intervals or d-union-intervals). Similarly, $\sup_{\mathcal{H}} \frac{\alpha^*(\mathcal{H})}{\alpha(\mathcal{H})}$ is the integrality gap of the maximum independent set problem. Since $\frac{\rho(\mathcal{H})}{\rho^*(\mathcal{H})}$ and $\frac{\alpha^*(\mathcal{H})}{\alpha(\mathcal{H})}$ are always at least 1, both integrality gaps are a lower bound on the duality gap. For the case of 1-intervals, we actually have $\alpha(\mathcal{H}) = \rho(\mathcal{H})$; both linear programs have an integrality gap of 1 because the incidence matrix of $G_{\mathcal{H}}$ exhibits the consecutive ones property and is thus totally unimodular.

Often, upper bounds on integrality gaps for packing and covering problems come alongside LP-relative approximation algorithms. Bar-Yehuda et al. employ the *local ratio* technique to obtain a polynomial-time LP-relative $2d$-approximation algorithm for the d-interval maximum independent set problem, proving that the integrality gap of maximum independent set for d-intervals is at most $2d$. Their result carries over to the version in which each element in $\mathcal{H}$ has a positive weight an a maximum weight independent set is desired. In Section 2, we show that their bound is tight up to an additive constant by establishing the following:

Theorem 1 *For any $\epsilon > 0$, there exists a collection $\mathcal{H}$ of d-intervals (respectively, d-union-intervals) for which $\frac{\alpha^*(\mathcal{H})}{\alpha(\mathcal{H})} \geq 2d - 1 - \epsilon$ (respectively, $2d - 2 - \epsilon$).*

Our constructions generalize examples from [5] and [8] but employ upon a novel amplification trick.

For both the d-interval and d-union-interval hitting set problems, we are able to prove that the integrality gap is exactly d. We show the following:

Theorem 2 *There exists a polynomial-time LP-relative d-approximation for the d-interval hitting set problem. Accordingly, for any collection $\mathcal{H}$ of d-intervals, $\frac{\rho(\mathcal{H})}{\rho^*(\mathcal{H})} \leq d$.*

Theorem 3 *For any $\epsilon > 0$, there exists a collection $\mathcal{H}$ of d-union-intervals for which $\frac{\rho(\mathcal{H})}{\rho*(\mathcal{H})} \geq d - \epsilon$.*

Theorem 2 uses standard techniques to generalize a 2-approximation algorithm for rectangle stabbing due to [7], but Theorem 3 employs a novel construction.

The table below summarizes the known integrality and duality gap bounds for $d \geq 2$:

d-Interval	Lower Bound	Upper Bound
Duality Gap	$\Omega(\frac{d^2}{\log d})$ [13]	$d^2 - d + 1$ [10]
Max-IS Integ. Gap	$2d - 1$	$2d$ [2]
Min-HS Integ. Gap	d	d
d-Union		
Duality Gap	$\Omega(\frac{d^2}{\log^2 d})$ [13]	$d^2 - d$ [10]
Max-IS Integ. Gap	$2d - 2$	$2d$ [2]
Min-HS Integ. Gap	d	d

We note that for $d = 2$, Kaiser's topology-based upper bounds on the duality gap are actually tighter than the constructive integrality gap upper bounds of Bar-Yehuda et al. (instead, they match our integrality gap lower bounds.) Hence, despite knowing that $\frac{\rho(\mathcal{H})}{\alpha(\mathcal{H})}$ is bounded above by 3 and 2 for families of 2-intervals and 2-union-intervals respectively, no polynomial-time approximation factor below 4 is known for the maximum independent set problems. We observe, however, that Kaiser's proof can be turned into an algorithm if one has access to an oracle to solve the topological subproblems that arise in the proof. The particular topological problems in question are related to finding Borsuk-Ulam fixed-points. Unfortunately, finding Borsuk-Ulam fixed-points is a PPAD-complete problem [15] and thus seems unlikely to admit polynomial algorithms unless a major breakthrough occurs. Nevertheless, PPAD includes several problems—such as locating Brouwer fixed-points and Nash equilibria—that are sometimes amenable to pivoting algorithms that, while not guaranteed to terminate in polynomial time, may perform well in practice. Moreover, we find it independently interesting that the 2-dimensional maximum conflict-free subset problem is a natural geometric optimization problem whose best known approximability appears to improve in the presence of a PPAD oracle. In Section 3, we outline a proof of the following:

Theorem 4 *There exists an algorithm for the maximum independent set problem on 2-intervals (respectively, 2-union-intervals) returning a solution of size at least* $\frac{\alpha(\mathcal{H})}{3}$ *(respectively* $\frac{\alpha(\mathcal{H})}{2}$*) requiring* $O(\log(\alpha(\mathcal{H})))$ *calls to an oracle for a PPAD-complete fixed-point problem, and polynomial time for all other computations.*

To achieve only a logarithmic number of oracle calls, we rely on a galloping binary search.

1.3 Related Work

Many variations and special cases of d-interval stabbing and independence problems have been studied in various contexts. Kovalena and Spieksma have examined the special case of the rectangle stabbing problem in which each rectangle is a horizontal line segment [11, 12]. In their most recent work, they obtain an LP-relative $\frac{e}{e-1}$-approximation for this case, alongside an example showing that the integrality gap is precisely $\frac{e}{e-1}$.

Even et al. explore weighted and capacitated variations of d-union-interval hitting set [6]. Their results include a $3d$-approximation for a variant in which each point may only be used to hit a specified number of d-intervals, but may be purchased multiple times.

Spieksma considers the version of maximum d-interval independent set where the goal is to select a *single* interval from each d-interval such that none intersect [17]. It is shown that a straightforward greedy procedure yields a 2-approximation.

Some additional hardness results are also known. Even et al. show that there is a constant $c > 0$ such that it is NP-hard to approximate the d-interval hitting set problem to within $c \log d$ [6]. Dom et al. show that rectangle stabbing is $W[1]$-hard, even when the input consists of squares of the same size, implying that the problem is unlikely to be fixed-parameter tractable in the optimal objective value $\rho(\mathcal{H})$ [4].

2 Integrality Gap Bounds

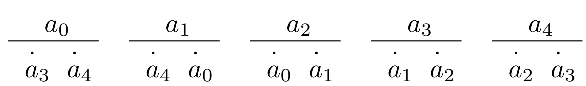
To prove Theorem 1 and establish tight lower bounds on the integrality gap of the independent set problem, we rely on an amplification lemma. We shall refer to the d individual intervals composing a d-interval as its *pieces*, and call a piece *inert* if it is a point. We shall call $\mathcal{H}$ a *clique* if $\alpha(\mathcal{H}) = 1$, and write $r(\mathcal{H})$ for the *rank* of $G_{\mathcal{H}}$—the maximum number of d-intervals intersected by any point in $\mathbb{R}$. We observe that if $\mathcal{H}$ is a clique and $r(\mathcal{H}) = p$, then a fractional independent set of value $\frac{|\mathcal{H}|}{p}$ can be obtained by simply putting weight $\frac{1}{p}$ on each d-interval in $\mathcal{H}$. In some situations, we can do better:

Lemma 5 *Suppose that $\mathcal{H}$ is a clique, and that $\mathcal{H}$ contains no two inert pieces that intersect and no d-intervals consisting entirely of inert pieces. Furthermore, suppose that $r(\mathcal{H}^*) = q$, where $\mathcal{H}^*$ is a modified version of $\mathcal{H}^*$ obtained by deleting all inert pieces. Then for any $N \in \mathbb{N}$, there is a clique $\mathcal{H}'$ of d-intervals admitting a fractional independent set of value $\frac{N|\mathcal{H}|}{Nq+1}$.*

Proof. We construct $\mathcal{H}'$ by making N copies of each d-interval in $\mathcal{H}$, and then perturbing all inert pieces in the resulting family of d-intervals such that no two inert pieces intersect, while preserving intersections of inert pieces with non-inert pieces. It is immediate that $\mathcal{H}'$ is still a clique; note that copies of the same d-interval in $\mathcal{H}$ must intersect in $\mathcal{H}'$ because no d-interval consists entirely of inert pieces. Moreover, $r(\mathcal{H}') = Nq + 1$, so $\mathcal{H}'$ admits a fractional independent set of value $\frac{N|\mathcal{H}|}{Nq+1}$ by placing weight $\frac{1}{Nq+1}$ on each d-interval in $\mathcal{H}'$. □

By taking the limit as $N \to \infty$, Lemma 5 yields an integrality gap lower bound of $\frac{|\mathcal{H}|}{q}$ given a d-interval graph satisfying the necessary requirements. We note that the amplification in Lemma 5 also works for d-union-intervals. We proceed with the proof of Theorem 1:

Proof of Theorem 1. For the case of d-interval graphs, we exhibit a clique $\mathcal{H}$ satisfying the conditions of Lemma 5 with $|\mathcal{H}| = 2d-1$ and $q = 1$. We label the d-intervals $\{a_0, a_1, \ldots, a_{2d-2}\}$. Each d-interval will have exactly one non-inert piece (a closed interval in $\mathbb{R}$) and $d-1$ inert pieces. We position the non-inert pieces such that no two intersect, which ensures that $q = 1$. Then for all $0 \le i \le 2d-2$, we position the remaining $d-1$ inert pieces of a_i (each of which is a point) on the non-inert pieces of d-intervals $\{a_{i+1}, \ldots, a_{i+(d-1)}\}$, where the addition is modulo $2d-1$. This ensures that an inert piece of a_i intersects non-inert pieces in $\{a_{i+1}, \ldots, a_{i+(d-1)}\}$, hence ensuring that inert pieces of $\{a_{i-1}, \ldots, a_{i-(d-1)}\}$ all intersect the non-inert piece of interval a_i. This proves that the construction yields a clique, from which it follows that the independent set problem in d-interval graphs has an integrality gap of $\frac{|\mathcal{H}|}{q} = 2d-1$. An example of the construction for $d = 3$ is shown (intervals are vertically separated for clarity):



For the case of d-union-intervals, we exhibit a clique $\mathcal{H}$ of size $4d-4$ satisfying the conditions of Lemma 5 with $q = 2$. This yields an integrality gap $\frac{|\mathcal{H}|}{q} = 2d-2$.

We label the $4d-4$ d-union-intervals by $\{a_1^i, a_2^i, a_3^i, a_4^i\}_{i=1}^{d-1}$. We shall say that each interval has its k^{th} piece in the k^{th} *track*, where each track is a copy of $\mathbb{R}$. We first explain what happens in tracks 1 through $d-1$, and then explain what happens in the final track, which is treated differently. For $1 \le i \le d-1$, all of the pieces in track i are inert (single points) except for the i^{th} pieces of a_1^i, a_2^i, a_3^i, and a_4^i, which are arranged as follows:

Now, for all $j \ne i$, the i^{th} pieces of $\{a_1^j, a_2^j, a_3^j, a_4^j\}$ are positioned according to the following rules:

- If $j < i$, put a_1^j, a_2^j in $a_1^i \cap a_2^i$; put a_3^j, a_4^j in $a_3^i \cap a_4^i$
- If $j > i$, put a_3^j, a_4^j in $a_1^i \cap a_2^i$; put a_1^j, a_2^j in $a_3^i \cap a_4^i$

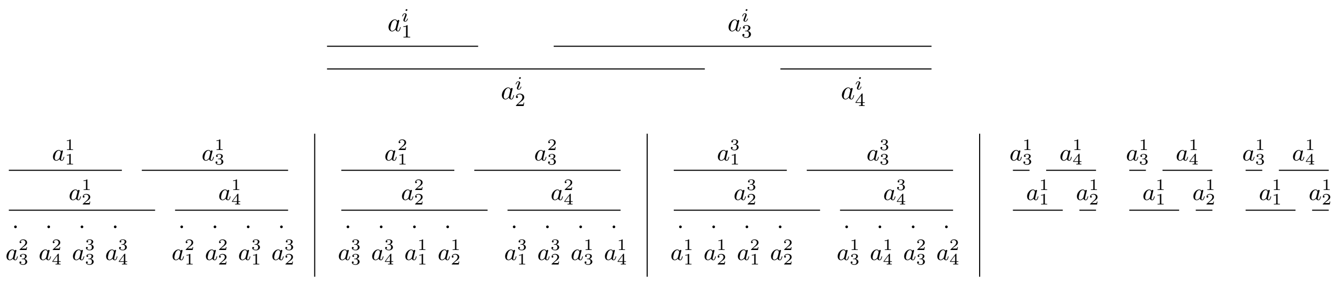


Figure 1: A clique of 12 4-union-intervals satisfying the conditions of Lemma 5 with $q = 2$

In the last track d, none of the intervals need to be inert. For all $1 \leq i \leq d-1$, the d^{th} pieces of $\{a_1^i, a_2^i, a_3^i, a_4^i\}$ are positioned similarly to the diagram above, but are permuted to induce the remaining three dependencies among the d-intervals. Figure 2 illustrates this and provides an example of the entire construction for $d = 4$.

Observe that any two d-union-intervals with the same superscript i must be adjacent in either track i or track d. For $1 \leq i < j \leq d-1$, we check that all 16 dependencies between $a_1^i, a_2^i, a_3^i, a_4^i$ and $a_1^j, a_2^j, a_3^j, a_4^j$ are accounted for: In track i, a_3^j and a_4^j intersect $a_1^i \cap a_2^i$; a_1^j and a_2^j intersect $a_3^i \cap a_4^i$. In track j, a_1^i and a_2^i intersect $a_1^j \cap a_2^j$; a_3^i and a_4^i intersect $a_3^j \cap a_4^j$. Thus $\mathcal{H}$ is a clique. It is easy to verify that the other conditions of Lemma 5 are satisfied with $q = 2$, so the proof is complete. $\square$

Next, we establish Theorem 2 by giving a set of d-union-intervals with a hitting set integrality gap of $d - \epsilon$:

Proof of Theorem 2. Fix $\epsilon > 0$. Choose any integer $a \geq \frac{2d^2}{\epsilon}$ and any integer $n \geq \frac{2a}{\epsilon}$. Fix some small δ, say $\delta = 0.1$. Here, we regard the d tracks $\{J^1, \ldots, J^d\}$ as disjoint copies of $\mathbb{R}$. A d-union-interval I is called *aligned* if, for all $1 \leq k \leq d$, the piece of I in J^k has the form $[i_k + \delta, j_k - \delta]$ for some integers $0 \leq i_k < j_k \leq n$. In other words, a d-union-interval is aligned if the endpoints of all of its pieces each barely miss an integer point between 0 and n. Let $\mathcal{H}$ be the collection of all aligned d-union-intervals I such that the total length of all pieces in I is exactly $a - 2d\delta$. Note that $|\mathcal{H}|$ is finite.

Let P contain all points of the form $i + 0.5$ for $i \in \{0, 1, \ldots, n-1\}$ in each of the d tracks, for a total of dn points. Each d-union-interval in $\mathcal{H}$ must contain at least a points in P, so we can obtain a fractional hitting set of total weight $\frac{dn}{a}$ by placing a value of $\frac{1}{a}$ at each point in P. This shows that $\rho^*(\mathcal{H}) \leq \frac{dn}{a}$.

Let Q be any feasible integral hitting set. Let b_i be the number of points of Q in J^i. We assume that $b_i < n$; if not, then $|Q| \geq n$, which will suffice for our lower bound. We say that an open interval in J^i is *missed* if it contains no points of Q and has integral endpoints. By the pigeonhole principle, there must exist a missed subinterval of $[0, n]$ in J^i having length at least $\frac{n-b_i}{b_i+1}$. By combining these missed subintervals over all d tracks, we obtain a d-union-interval of total length $\sum_{i=1}^d \frac{n-b_i}{b_i+1}$ that is missed by Q in all tracks. Since no aligned d-union-interval in I is missed by Q, we must have

$$\sum_{i=1}^{d} \frac{n - b_i}{b_i + 1} < a.$$

By rearranging this, we obtain

$$d \left(\sum_{i=1}^{d} \frac{1}{b_i + 1} \right)^{-1} > \frac{d(n+1)}{a+d}.$$

The left side of the above equation is a harmonic mean. Since an arithmetic mean is always greater than or equal to the corresponding harmonic mean, we get

$$\frac{1}{d}\sum_{i=1}^{d}(b_i+1) > \frac{d(n+1)}{a+d}$$

and hence $\rho(\mathcal{H}) \geq |Q| = \sum_{i=1}^{d} b_i > \frac{d^2(n+1)}{a+d} - d$. Dividing by the upper bound we had for $\rho^*(\mathcal{H})$ gives

$$\frac{\rho(\mathcal{H})}{\rho^*(\mathcal{H})} > \frac{ad(n+1)}{n(a+d)} - \frac{a}{n} > \left(1 - \frac{d}{a+d}\right)d - \frac{a}{n},$$

where the last inequality is due to $\frac{n+1}{n} > 1$. Since we chose a and n such that $a \geq \frac{2d^2}{\epsilon}$ and $n \geq \frac{2a}{\epsilon}$, we get

$$\frac{\rho(\mathcal{H})}{\rho^*(\mathcal{H})} > \left(1 - \frac{d}{\frac{2d^2}{\epsilon}+d}\right)d - \frac{a}{\frac{2a}{\epsilon}} = d - \frac{\epsilon}{2+\frac{\epsilon}{d}} - \frac{\epsilon}{2} > d - \epsilon,$$

completing the proof of Theorem 2. □

Finally, we provide a polynomial algorithm yielding a matching upper bound of d for the integrality gap of the general d-interval hitting set problem:

Proof of Theorem 3. Let $\mathcal{H}$ be a collection of d-intervals, and let $\{y_p^* : p \in P(\mathcal{H})\}$ be an optimal fractional hitting set of weight $\rho^*(\mathcal{H})$ obtained by solving linear program (2). We demonstrate how to round $\{y_p^*\}$ to an integral solution of weight at most $d \cdot \rho^*(\mathcal{H})$. For a d-interval I in $\mathcal{H}$, let $I^* \subseteq I$ be any piece of I that is hit by weight at least $\frac{1}{d}$ under $\{y_p^*\}$ (one must exist by the pigeonhole principle). Then the set $\mathcal{C} = \{I^* : I \in \mathcal{H}\}$ is a set of intervals in $\mathbb{R}$ that are each hit by weight at least $\frac{1}{d}$ under $\{y_p^*\}$.

By multiplying solution $\{y_p^*\}$ by d, we obtain a new fractional hitting set $\{dy_p^*\}$ of weight $d\rho^*(\mathcal{H})$ that hits, with weight 1, all elements of $\mathcal{C}$. However, the incidence matrix for the hitting set problem on 1-intervals is totally unimodular, so there must exist an integral hitting set Q of weight at most $d\rho^*(\mathcal{H})$ that hits all of $\mathcal{C}$—one can be found by simply solving linear program (2) again for $\mathcal{C}$ instead of $\mathcal{H}$. Of course, Q is also a hitting set for $\mathcal{H}$, from which it follows that $\rho(\mathcal{H}) \leq d \cdot \rho^*(\mathcal{H})$. Moreover, by simply returning Q, we obtain a polynomial-time LP-relative d-approximation for the d-interval hitting set problem, completing the proof. □

We note that the above algorithm also works for the *weighted* variant of the minimum hitting set problem, in which each point $p \in P(\mathcal{H})$ is given a positive cost, and the goal is to compute a minimum cost hitting set.

3 Topology-based algorithms

In this section, we sketch a proof of Theorem 4, illustrating how to obtain a 3-approximation (respectively, a 2-approximation) for the independent set problem on 2-intervals (respectively, 2-union-intervals), supposing one has access to oracles for topological subproblems. We assume familiarity with topological fixed-point theorems and their connection to the complexity class PPAD; see [15].

Our approach follows Kaiser's duality gap upper bound proof [10], which we outline here. We first consider the case of 2-union-intervals. For concreteness, we consider a family $\mathcal{H}$ of axis-aligned rectangles in the plane. Let n be an arbitrary positive integer. Kaiser demonstrates how to construct a family of $2n+2$ real-valued functions $h_1^{\mathcal{H}}, \ldots, h_{2n+2}^{\mathcal{H}}$ defined on the space $S^n \times S^n$ (where S^n is the boundary of an $(n+1)$-dimensional unit ball) with the following properties:

1. Each point $x \in S^n \times S^n$ can be mapped to a set of n horizontal lines and n vertical lines in the plane.
2. If $h_i^{\mathcal{H}}(x) = 0$ for all i, then x corresponds to a set of lines that intersect all rectangles in $\mathcal{H}$.

3. If $h_i^{\mathcal{H}}(x) \neq 0$ for all i, then x corresponds to a set of n horizontal lines and n vertical lines defining a grid from which we can easily find a conflict-free set of rectangles of size $n+1$.

Kaiser then establishes that, for topological reasons, there must exist a point $x \in S^n \times S^n$ such that $h_1^{\mathcal{H}}(x) = h_2^{\mathcal{H}}(x) = \ldots = h_{2n+2}^{\mathcal{H}}(x)$ and thus a point x must exist satisfying item 2 or item 3 above. Tardos's result that $\rho(\mathcal{H}) \leq 2\alpha(\mathcal{H})$ follows immediately by setting $n = \alpha(\mathcal{H})$, since then a point where $h_i^{\mathcal{H}}(x) \neq 0$ for all i cannot exist, and thus a stabbing consisting of $\alpha(\mathcal{H})$ horizontal lines and $\alpha(\mathcal{H})$ vertical lines must exist.

Kaiser's proof outlines a procedure by which, given an integer n, we can find either a stabbing of size $2n$ or a conflict-free set of rectangles of size $n+1$, using only a single oracle call to a solver for a topological fixed-point problem. Using this, we can obtain a 2-approximation for the maximum conflict-free subset problem by simply finding a *cutoff point* t such that Kaiser's method returns a conflict-free set of size t when run with $n = t-1$, but returns a stabbing of size $2t$ when run with $n = t$. Although there might exist multiple cutoff points t, one must exist in the interval $[\frac{\rho(\mathcal{H})}{2}, \alpha(\mathcal{H})]$. A cutoff point can thus be found in $O(\alpha(\mathcal{H}))$ guesses by a galloping binary search that first guesses $t = 1, t = 2, t = 4, \ldots$ until a stabbing of size $2t$ is returned, and then binary searches between $\frac{t}{2}$ and t to find a cutoff point.

Kaiser's topological argument employs a result of Ramos that generalizes the Borsuk-Ulam theorem to cross products of spheres [16], and thus our algorithm requires oracles for Ramos-style fixed-points. Ramos's theorem is proven using a parity-based induction argument that implicitly shows that an appropriate computational version of the Ramos fixed-point problem lies in the complexity class PPAD (since a search for Ramos fixed-points can be completed by solving a cleverly constructed instance of the *end-of-the-line* problem [15]). We also note that, by the discreteness of our problem, the particular instances that we must solve can be efficiently represented and have rational solutions. This establishes Theorem 4 for 2-union-intervals.

For the case of general 2-intervals, Kaiser provides a related argument that can be adapted in a similar manner to yield a binary search algorithm. In this case, only oracles for the standard Borsuk-Ulam theorem are required. However, due to changes in Kaiser's analysis, we can only obtain a 3-approximation. Still, our integrality gap bounds imply that this is optimal among all LP-relative approximations.

References

[1] N. Alon. Piercing d-intervals. *Disc. Comput. Geom.*, 19(3, Special Issue):333–334, 1998.

[2] R. Bar-Yehuda, M. Halldórsson, J. Naor, H. Shachnai, and I. Shapira. Scheduling split intervals. *SIAM J. Comput.*, 36(1):1–15 (electronic), 2006.

[3] G. Călinescu, A. Dumitrescu, H. Karloff, and P. Wan. Separating points by axis-parallel lines. *Internat. J. Comput. Geom. Appl.*, 15(6):575–590, 2005.

[4] M. Dom, M. Fellows, and F. Rosamond. Parameterized complexity of stabbing rectangles and squares in the plane. In Proc. *WALCOM*, LNCS 5341:298–309, 2009.

[5] A. Dumitrescu. On two lower bound constructions. In *Proc. 11th Canadian Conf. Comput. Geom.*, 1999.

[6] G. Even, R. Levi, D. Rawitz, S. Baruch, S. Shahar, and M. Sviridenko. Algorithms for capacitated rectangle stabbing and lot sizing with joint set-up costs. *ACM Trans. Algorithms*, 4(3), 2008.

[7] D. Gaur, T. Ibaraki, and R. Krishnamurti. Constant ratio approximation algorithms for the rectangle stabbing problem and the rectilinear partitioning problem. *J. Algorithms*, 43(1): 138–152, 2002.

[8] A. Gyárfás and J. Lehel. A Helly-type problem in trees. *Comb. Theory and its Appl., II*, 571–584, 1970.

[9] M. Jiang. Recognizing d-interval graphs and d-track interval graphs. In *Proc. 4th FAW*, 160–171, 2010.

[10] T. Kaiser. Transversals of d-intervals. *Disc. Comput. Geom.*, 18(2):195–203, 1997.

[11] S. Kovaleva and F. Spieksma. Primal-dual approximation algorithms for a packing-covering pair of problems. *RAIRO Oper. Res.*, 36(1):53–71, 2002.

[12] S. Kovaleva and F. Spieksma. Approximation algorithms for rectangle stabbing and interval stabbing problems. *SIAM J. Disc. Math.*, 20(3):748-768, 2006.

[13] J. Matoušek. Lower bounds on the transversal numbers of d-intervals. *Disc. Comput. Geom.*, 26(3):283–287, 2001.

[14] H. Nagashima and K. Yamazaki. Hardness of approximation for non-overlapping local alignments. *Disc. Appl. Math.*, 137(3):293-309, 2004.

[15] C. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *J. Comput. System Sci.*, 48(3):498–532, 1994.

[16] E. Ramos. Equipartition of mass distributions by hyperplanes. *Disc. Comput. Geom.*, 15(2):147–167, 1996.

[17] F. Spieksma. On the approximability of an interval scheduling problem. *J. Sched.*, 2(5):215-227, 1999.

[18] G. Tardos. Transversals of 2-intervals, a topological approach. *Combinatorica*, 15(1):123–134, 1995.

[19] V. Vatter. Small permutation classes. In *Proc. Lond. Math. Soc. (3)*, 103(5):879-921, 2011.