DIMACS Technical Report 2002-23
May 2002

Generating Dual-Bounded Hypergraphs[1]

by

Endre Boros,[2] Khaled Elbassioni[3] Vladimir Gurvich[4]
Leonid Khachiyan[5]

[1]The research was supported in part by the National Science Foundation Grant IIS-0118635. The research of the first and third authors was also supported in part by the Office of Naval Research Grant N00014-92-J-1375. The second and third authors are also grateful for the partial support by DIMACS, the National Science Foundation's Center for Discrete Mathematics and Theoretical Computer Science.
[2]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; (boros@rutcor.rutgers.edu).
[3]Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway NJ 08854-8003; (elbassio@paul.rutgers.edu).
[4]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; (gurvich@rutcor.rutgers.edu).
[5]Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway NJ 08854-8003; (leonid@cs.rutgers.edu).

ABSTRACT

This paper surveys some recent results on the generation of implicitly given hypergraphs and their applications in Boolean and integer programming, data mining, reliability theory, and combinatorics.

Given a monotone property π over the subsets of a finite set V, we consider the problem of incrementally generating the family $\mathcal{F}_\pi$ of all minimal subsets satisfying property π, when π is given by a polynomial-time satisfiability oracle. For a number of interesting monotone properties, the family $\mathcal{F}_\pi$ turns out to be *uniformly dual-bounded*, allowing for the incrementally efficient enumeration of the members of $\mathcal{F}_\pi$.

Important applications include the efficient generation of minimal infrequent sets of a database (data mining), minimal connectivity ensuring collections of subgraphs from a given list (reliability theory), minimal feasible solutions to a system of monotone inequalities in integer variables (integer programming), minimal spanning collections of subspaces from a given list (linear algebra) and maximal independent sets in the intersection of matroids (combinatorial optimization).

In contrast to these results, the analogous problem of generating the family of all maximal subsets not having property π is NP-hard for most of the monotone properties π considered in the paper.

1 Introduction

Let V be a finite set of cardinality $|V| = n$. For a hypergraph (set family) $\mathcal{H} \subseteq 2^V$, let us denote by $\mathcal{I}(\mathcal{H})$ the family of its *maximal independent* subsets, i.e. maximal subsets of V not containing any edges of $\mathcal{H}$. The complement of a maximal independent subset is a *minimal transversal* of $\mathcal{H}$, i.e., a minimal subset of V intersecting all elements of $\mathcal{H}$. (Minimal transversals are also called minimal *hitting sets*.) The collection $\mathcal{H}^d$ of minimal transversals is called the *dual* or *transversal* hypergraph of $\mathcal{H}$. It is easy to see that $\mathcal{H}^d$ is a Sperner hypergraph, i.e., no edge of $\mathcal{H}^d$ contains another edge of $\mathcal{H}^d$. If $\mathcal{H}$ is also Sperner, then $\mathcal{H} = (\mathcal{H}^d)^d$.

1.1 Hypergraph dualization

Given a Sperner hypergraph $\mathcal{H}$, a frequently arising task is the generation of the transversal hypergraph $\mathcal{H}^d$. This problem, known as *dualization*, can be stated as follows:

DUAL$(\mathcal{H}, \mathcal{X})$**:** *Given a complete list of all edges of* $\mathcal{H}$*, and a set of minimal transversals* $\mathcal{X} \subseteq \mathcal{H}^d$*, either prove that* $\mathcal{X} = \mathcal{H}^d$*, or find a new transversal* $X \in \mathcal{H}^d \setminus \mathcal{X}$*.*

Clearly, we can generate all of the minimal transversals in $\mathcal{H}^d$ (equivalently, all the maximal independent sets for $\mathcal{H}$) by initializing $\mathcal{X} = \emptyset$ and iteratively solving the above problem $|\mathcal{H}^d| + 1$ times. Since $|\mathcal{H}^d|$ can be exponentially large in both $|\mathcal{H}|$ and $|V|$, the complexity of generating $\mathcal{H}^d$ is customarily measured in the input and output sizes. In particular, we say that $\mathcal{H}^d$ can be generated in *incremental polynomial time* if the above problem DUAL$(\mathcal{H}, \mathcal{X})$ can be solved in time polynomial in $|V|, |\mathcal{H}|$, and $|\mathcal{X}|$.

The dualization problem can be solved efficiently for many classes of hypergraphs. For example, for hypergraphs of bounded dimension, i.e., when the sizes of all the edges of $\mathcal{H}$ are limited by a constant

$$dim(\mathcal{H}) = \max\{|H| \mid H \in \mathcal{H}\} \leq k,$$

the dualization problem can be executed in incremental polynomial time (see e.g. [8, 26]), and even stronger, all minimal transversals of $\mathcal{H}$ can be enumerated in lexicographic order [27]. For the quadratic case ($k = 2$) there are even more efficient algorithms that enumerate all edges of $\mathcal{H}^d$ in lexicographic order with polynomial delay, i.e., in $poly(|V|, |\mathcal{H}|)$ time per each generated minimal transversal(see e.g. [41, 43, 68]). In addition, if the dimension of $\mathcal{H}$ is bounded, then problem DUAL$(\mathcal{H}, \mathcal{X})$ can also be solved efficiently in parallel:

Proposition 1 ([9]) *Problem DUAL*$(\mathcal{H}, \mathcal{X})$ *can be solved by an NC algorithm for* $dim(\mathcal{H}) \leq 3$ *and by a randomized NC algorithm for* $dim(\mathcal{H}) = 4, 5 \ldots$

Efficient algorithms also exist for the dualization of 2-monotonic, threshold, matroid, read-bounded, acyclic and some other classes of hypergraphs of unbounded dimension (see e.g. [5, 19, 22, 24, 26, 27, 46, 49, 50, 58, 59]).

Even though no incremental polynomial-time algorithm for the dualization of arbitrary hypergraphs is known, the dualization problem for hypergraphs of unbounded dimension is very unlikely to be NP-hard since it can be solved in incremental *quasi-polynomial time* [1] (see [30], or [36] for more detail):

Proposition 2 ([30]) *Problem DUAL*$(\mathcal{H}, \mathcal{X})$ *can be solved in* $O(n) + m^{o(\log m)}$ *time, where* $n = |V|$ *and* $m = |\mathcal{H}| + |\mathcal{X}|$.

The quasi-polynomial time algorithm of [30] also indicates that the duality of two hypergraphs can be disproved with limited nondeterminism [27]. As shown in [67], the quasi-polynomial term in the bound of Proposition 2 can be replaced by $(n|H|)^{O(\log(n|H|))}$.

Various applications of the hypergraph dualization problem include combinatorics [62], graph theory [41, 43, 63, 68], artificial intelligence [26], game theory [32, 33, 34, 61], reliability theory [23, 61], database theory [1, 51, 71], and learning theory [3, 24, 31].

1.2 Monotone and independence systems, membership oracles and enumeration problems

Let us consider an arbitrary monotone property π over the subsets of V, and let $\mathcal{S}_\pi$ denote the corresponding collection of all subsets having property π:

$$Y \supseteq X \in \mathcal{S}_\pi \quad \Rightarrow Y \in \mathcal{S}_\pi.$$

We shall assume that π is represented by a (polynomial-time) *membership* or *satisfiability oracle* $\mathcal{O}_\pi$, i.e., an algorithm which, given some input description $\mathcal{D}_\pi$ of π of size $|\mathcal{D}_\pi|$ and a subset $X \subseteq V$, can decide whether or not $X \in S_\pi$ in time polynomial in n and $|\mathcal{D}_\pi|$. Note that π may also be characterized by the hypergraph $\mathcal{F}_\pi \subseteq \mathcal{S}_\pi$ of all *minimal* subsets having property π. For this reason, the oracle $\mathcal{O}_\pi$ could equivalently be called a *superset oracle* for the family $\mathcal{F}_\pi$. (A set $X \subseteq V$ has property π, i.e., $X \in \mathcal{S}_\pi$, if and only if X is a superset for some $Y \subseteq X$ such that $Y \in \mathcal{F}_\pi$.) Let us also note that the complementary family $\mathcal{I}_\pi = 2^V \setminus \mathcal{S}_\pi$, i.e, the family of all sets X that does not satisfy the monotone property π, is an *independence system*:

$$Y \subseteq X \in \mathcal{I}_\pi \;\Rightarrow\; Y \in \mathcal{I}_\pi.$$

It is easy to see that the hypergraph $\mathcal{G}_\pi$ of all *maximal* subsets of $\mathcal{I}_\pi$ consists of all maximal independent sets for $\mathcal{F}_\pi$, i.e., $\mathcal{G}_\pi = \mathcal{I}(\mathcal{F}_\pi)$, where $\mathcal{I}(\cdot)$ denotes the family of all maximal independent sets for $(\cdot)$. In particular, the complementarity hypergraph $\{X \mid X = V \setminus Y,\, Y \in \mathcal{G}_\pi\}$ is the transversal hypergraph for $\mathcal{F}_\pi$.

Given a monotone property π, described by a satisfiability oracle $\mathcal{O}_\pi : \mathcal{D}_\pi \times 2^V \mapsto \{yes, no\}$, we consider the problem of incrementally enumerating all elements of $\mathcal{F}_\pi$:

[1] A function $f(t)$ is quasi-polynomial bounded if $f(t) \leq 2^{\text{polylog}\, t}$.

GEN$(\mathcal{F}_\pi, \mathcal{X})$**:** *Given a subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$, either find a new minimal satisfying set $X \in \mathcal{F}_\pi \setminus \mathcal{X}$, or prove that the given partial list is complete, $\mathcal{X} = \mathcal{F}_\pi$.*

Since the entire family $\mathcal{F}_\pi$ can be generated by initializing $\mathcal{X} = \emptyset$ and iteratively solving the above problem $|\mathcal{F}_\pi| + 1$ times, and since, as before, the size of $\mathcal{F}_\pi$ may be exponentially large both in n and the size of the input description $\mathcal{D}_\pi$, we are interested in *incrementally efficient* algorithms, i.e. algorithms which solve problem GEN$(\mathcal{F}_\pi, \mathcal{X})$ efficiently in terms of n, $|\mathcal{D}_\pi|$, and the size of the partial list $|\mathcal{X}|$.

Example 1 (Hypergraph Dualization) Let $\mathcal{H} \subseteq 2^V$ be a Sperner hypergraph given by an explicit list of its edges. For $X \subseteq V$, let $\pi(X)$ be the property that X is a transversal of $\mathcal{H}$, that is, X intersects all edges of $\mathcal{H}$. Then $\mathcal{F}_\pi = \mathcal{H}^d$ is the transversal hypergraph of $\mathcal{H}$, the input list $\mathcal{D}_\pi$ of all edges of $\mathcal{H}$ provides a polynomial-time superset oracle for $\mathcal{F}_\pi$, and problem GEN$(\mathcal{F}_\pi, \mathcal{X})$ turns into the hypergraph dualization problem DUAL$(\mathcal{H}, \mathcal{X})$.

Example 2 (Knapsack inequalities) Given n non-negative weights $a_1, \ldots, a_n$ and a threshold b, define $\mathcal{S}_\pi$ to be the family of all subsets X of V whose total weight is at least b: $X \in \mathcal{S}_\pi \quad \Leftrightarrow \quad \sum_{i \in X} a_i \geq b$. Equivalently, $\mathcal{S}_\pi$ is the family of (the support sets of) all binary solutions to the knapsack inequality $ax \geq b$, where $a = (a_1, \ldots, a_n)$ and $x \in \{0,1\}^n$. This problem is defined by $\mathcal{D}_\pi = \{a_1, \ldots, a_n, b\}$, and it is clear that $\mathcal{S}_\pi$ has a polynomial-time membership oracle. As shown in [43], all elements of $\mathcal{F}_\pi$, i.e., all minimal feasible binary solutions of the knapsack inequality $ax \geq b$, can be generated in incremental polynomial time with an amortized complexity of $O(n^2)$ per generated element, i.e., problem GEN$(\mathcal{F}_\pi, \mathcal{X})$ can be solved in $O(n^2|\mathcal{X}|)$ time. Due to the obvious symmetry, all elements of $\mathcal{I}(\mathcal{F}_\pi)$, i.e., all maximal infeasible binary solutions for $ax \geq b$, can also be generated in incremental polynomial time. However, the equivalence of the enumeration problems for $\mathcal{F}_\pi$ and $\mathcal{I}(\mathcal{F}_\pi)$ does not carry over to monotone properties described by *systems* of knapsack inequalities $Ax \geq b, \quad x \in \{0,1\}^n$, where A is a given $m \times n$ non-negative matrix and b is a given m-vector. Specifically, for $\mathcal{D}_\pi = (A, b)$ we still have a trivial polynomial-time satisfiability oracle, but as shown in [11], the enumeration problem for $\mathcal{I}(\mathcal{F}_\pi)$ becomes NP-hard: Given a family $\mathcal{X} \subseteq \mathcal{I}(\mathcal{F}_\pi)$ of maximal infeasible binary vectors for $Ax \geq b$, it is NP-complete to decide whether the given collection of maximal infeasible vectors can be extended, i.e. $\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{X} \neq \emptyset$. This clearly indicates that all elements of $\mathcal{I}(\mathcal{F}_\pi)$ cannot be enumerated in incremental (or output) polynomial time, unless P=NP. It was conjectured in [43] that there are no efficient algorithms for enumerating all elements of $\mathcal{F}_\pi$ either. However, as we discuss in Section 3.6, problem GEN$(\mathcal{F}_\pi, \mathcal{X})$ reduces in polynomial time to dualization and hence can be solved in quasi-polynomial time [14]. Accordingly, all minimal binary solutions to $Ax \geq b$, i.e., all elements of $\mathcal{F}_\pi$ can be generated in incremental quasi-polynomial time, which is also true for non-negative systems of linear inequalities $Ax \geq b$ in *integer* variables (see [10],[11]). This example will be discussed in more detail in Section 2.2.

Example 3 (Cuts, Spanning Trees and Graph Connectivity). Let $G = (U, E)$ be a connected (multi)graph with $n = |E|$ edges. For a subset of edges $X \subseteq E$, let π be the property that the removal of X disconnects some vertices in U, i.e., the (multi)graph $(U, E \setminus X)$ is not connected. Then $\mathcal{D}_\pi = G$, the connectivity of $(U, E \setminus X)$ can be checked in polynomial time by depth-first search, and $\mathcal{F}_\pi$ is the collection of all minimal cuts in G. The complementarity hypergraph of $\mathcal{I}(\mathcal{F}_\pi)$, i.e., the family of all minimal transversals to $\mathcal{F}_\pi$, is the family of all spanning trees for G. It is well-known that both $\mathcal{F}_\pi$ and $\mathcal{I}(\mathcal{F}_\pi)$ can be generated efficiently (see, e.g., [63] and [60]). Suppose now that we are given two connected (multi)graphs $G_1 = (U_1, E_1)$ and $G_2 = (U_2, E_2)$, and let V be a set of n "passes", where each "pass" $v \in V$ is valid for an edge $e_1(v) \in E_1$ and another edge $e_2(v) \in E_2$. Given a subset $X \subseteq V$, let π be the property that $X \subseteq V$ disconnects at least one of the two (multi)graphs, i.e.,

$$X \in \mathcal{S}_\pi \;\Leftrightarrow\; (U_1, E_1 \setminus e_1(X)) \text{ or } (U_2, E_2 \setminus e_2(X)) \quad \text{are not connected.}$$

As before, π can be checked in polynomial time, but as we show in Section 3.2, problem GEN$(\mathcal{F}_\pi, \mathcal{X})$ becomes NP-hard: given a subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$, it is NP-complete to tell whether $\mathcal{F}_\pi \setminus \mathcal{X} \neq \emptyset$. On the other hand, enumerating all minimal transversals to $\mathcal{F}_\pi$ is equivalent with the enumeration of all minimal solutions to the system of two matroid inequalities $f_i(X) = |U_i| - c_i(X) \geq |U_i| - 1$, $i = 1, 2$, where $c_i(X)$ is the number of connected components of the graph $(U_i, e_i(X))$. As we shall see in Section 2.2, the latter problem is a special case of the so-called *matroid intersection problem* which, as shown in [43], can be solved with polynomial delay for any fixed number of matroids. We thus conclude that unlike the preceding example, all elements of $\mathcal{I}(\mathcal{F}_\pi)$ can be generated efficiently, while enumerating all elements of $\mathcal{F}_\pi$ is NP-hard. In fact, as we shall see in Section 3.2, the matroid intersection problem can be solved in incremental quasi-polynomial time even for unbounded number of matroids [13].

Example 4 (Carathéodory's Families of Cones and Vertex Enumeration). Let $\mathcal{K}$ be a family of n polyhedral cones $K_1, \ldots, K_n$ in $\Re^d$ and let $b \in \Re^d$ be a given rational vector such that $b \in \sum_{i=1}^n K_i$. Assume that each cone K_i is defined by a set of rational generators, and let π be the property that $b \in \sum_{i \in X} K_i$ for a given subsets of cones $X \subseteq V = \{1, \ldots, n\}$. Then, given $\mathcal{D}_\pi = \{K_1, \ldots, K_m, b\}$, any membership query "$X \in \mathcal{S}_\pi$?" can be answered in polynomial time by linear programming, and $\mathcal{F}_\pi$ consists of all minimal collections of the cones $K_1, \ldots, K_n$ sufficient for representing b. In particular, $|X| \leq d$ for all $X \in \mathcal{F}_\pi$ by the Carathéodory theorem. If $K_1, \ldots, K_n$ are all rays, i.e., $K_i = \mathit{cone}\{a_i\}$ for given vectors $a_1, \ldots, a_n$, and the cone generated by $a_1, \ldots, a_n$ is pointed, then each element of $\mathcal{F}_\pi$ can be identified with a vertex of the polytope $P = \{x \in \Re^n \mid \sum_{i=1}^n a_i x_i = b,\ (x_1, \ldots, x_n) \geq 0\}$. The question as to whether there exists an efficient vertex enumeration algorithm for an arbitrary polytope P is a well-known open problem in linear programming. However, it is known [42] that for dihedral cones $K_i = \mathit{cone}\{a_i, a_i'\}$ defined by pairs of generators $a_i, a_i'' \in \Re^d$, the enumeration problems for $\mathcal{F}_\pi$ and $\mathcal{I}(\mathcal{F}_\pi)$ are both NP-hard. Other examples for which both enumeration problems for $\mathcal{F}_\pi$ and $\mathcal{I}(\mathcal{F}_\pi)$ are NP-hard can be found in [35].

1.3 Joint generation and dualization

Although, in general, one or both of the enumeration problems for $\mathcal{F}_\pi$ and $\mathcal{I}(\mathcal{F}_\pi)$ may be NP-hard, the following *joint generation* problem is most likely not:

GEN$(\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y})$**:** *Given two explicitly listed set families* $\mathcal{X} \subseteq \mathcal{F}_\pi$ *and* $\mathcal{Y} \subseteq \mathcal{I}(\mathcal{F}_\pi)$, *either find a new set in* $(\mathcal{F}_\pi \setminus \mathcal{X}) \cup (\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{Y})$, *or prove that these families are complete:* $(\mathcal{X}, \mathcal{Y}) = (\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi))$.

It was observed in [6, 35] that for any polynomial-time satisfiability oracle O_π problem GEN$(\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y})$ can be reduced in polynomial time to dualization:

Proposition 3 ([6, 35]) *GEN*$(\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y})$ *can be solved in time* $n(poly(|\mathcal{X}|, |\mathcal{Y}|) + T(|\mathcal{O}_\pi|)) + T_{dual}$, *where* $T(|\mathcal{O}_\pi|)$ *is the worst-case running time of the oracle on any* $X \subseteq V$, *and* T_{dual} *denotes the time required to solve the dualization problem DUAL*$(\mathcal{X}, \mathcal{Y}^c)$, *where* $\mathcal{Y}^c = \{V \setminus Y \mid Y \in \mathcal{Y}\}$ *is the complementarity hypergraphs for* $\mathcal{Y}$.

Proof. Given $\mathcal{X}$ and $\mathcal{Y}$, we first check whether $\mathcal{Y} \subseteq \mathcal{I}(\mathcal{X})$, i.e., whether each set $Y \in Y$ is a *maximal* independent set for $\mathcal{X}$. (The fact that Y is an independent set for $\mathcal{X}$ follows from the monotonicity of π.) Suppose there is a set $Y \in \mathcal{Y}$ and a singleton $v \in V$ such that $Y' = Y \cup \{v\}$ is still independent of $\mathcal{X}$, i.e., $X \not\subseteq Y'$ for all $X \in \mathcal{X}$. Then $\mathcal{Y}'$ satisfies π and moreover, for any minimal subset $Z \subseteq Y'$ satisfying π we have $Z \in \mathcal{F}_\pi \setminus \mathcal{X}$. This gives a new set in $\mathcal{X}$, which can be found by querying $\mathcal{O}_\pi$ at most $|Y| \leq n$ times. Similarly, we may assume without loss of generality that each $X \in \mathcal{X}$ is a *minimal* transversal for $\mathcal{Y}^c$, for otherwise we can obtain a new set in $\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{Y}$ by querying oracle $\mathcal{O}_\pi$ at most n times. Now it is easy to see that both families $\mathcal{X}$ and $\mathcal{Y}$ are complete, i.e., $(\mathcal{X}, \mathcal{Y}) = (\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi))$ if and only if $\mathcal{X}$ and $\mathcal{Y}^c$ are mutually dual: $\mathcal{Y}^c = \mathcal{X}^d$. The latter condition can be checked by solving problem DUAL$(\mathcal{X}, \mathcal{Y}^c)$. $\square$

1.4 Dual-bounded hypergraphs and monotone properties

It is clear that for a given monotone property π, represented by a satisfiability oracle $\mathcal{O}_\pi : \mathcal{D}_\pi \times 2^V \mapsto \{yes, no\}$, we can generate both $\mathcal{F}_\pi$ and $\mathcal{I}(\mathcal{F}_\pi)$ simultaneously by starting with $\mathcal{X} = \mathcal{Y} = \emptyset$ and solving $|\mathcal{F}_\pi| + |\mathcal{I}(\mathcal{F}_\pi)| + 1$ instances of problem GEN$(\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y})$, incrementing in each iteration either $\mathcal{X}$ or $\mathcal{Y}$ by the newly found subset $S \in (\mathcal{F}_\pi \setminus \mathcal{X}) \cup (\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{Y})$, according to the answer of the oracle $\mathcal{O}_\pi$, until we have $(\mathcal{X}, \mathcal{Y}) = (\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi))$.

Though it may sound terribly redundant to use the above joint generation to generate $\mathcal{F}_\pi$ alone (and simply discard at the end all members of $\mathcal{I}(\mathcal{F}_\pi)$), it was in fact shown (see [31]) that no satisfiability oracle based algorithm can generate $\mathcal{F}_\pi$ in fewer than $|\mathcal{F}_\pi| + |\mathcal{I}(\mathcal{F}_\pi)|$ steps, in general.

Unfortunately, the above joint generation still may not be an efficient algorithm for solving either of GEN$(\mathcal{F}_\pi, \mathcal{X})$ or GEN$(\mathcal{I}(\mathcal{F}_\pi), \mathcal{Y})$ separately for the simple reason that we do not control which of the families $\mathcal{F}_\pi \setminus \mathcal{X}$ and $\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{Y}$ contains each new edge produced by the algorithm. Suppose we want to generate $\mathcal{F}_\pi$ and the family $\mathcal{I}(\mathcal{F}_\pi)$ is exponentially larger than $\mathcal{F}_\pi$. Then, if we are unlucky, we may get edges of $\mathcal{F}_\pi$ with exponential delay, while getting large subfamilies of $\mathcal{I}(\mathcal{F}_\pi)$ (which are not needed at all) in between.

For the above reasons, we shall study in this paper Sperner hypergraphs (or equivalently, monotone properties) for which joint generation in fact provides an incrementally efficient framework for generating all edges. Let us call a Sperner hypergraph $\mathcal{H} = \mathcal{F}_\pi$ (corresponding to a monotone property π) *dual-bounded* [14] if the size of $\mathcal{I}(\mathcal{F}_\pi)$ is (quasi-)polynomially limited in the size of $\mathcal{F}_\pi$, the size of V and the size $|\mathcal{D}_\pi|$ of the input description of π. Let us further call $\mathcal{H} = \mathcal{F}_\pi$ *uniformly dual-bounded*, if

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F}_\pi)| \leq (quasi\text{-})poly(|\mathcal{X}|, |V|, |\mathcal{D}_\pi|) \tag{1}$$

holds for any non-empty subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$.

¿From Proposition 3 it follows that joint generation is an incrementally efficient way of generating $\mathcal{F}_\pi$, whenever this hypergraph is uniformly dual bounded:

Proposition 4 ([14]) *If the hypergraph $\mathcal{F}_\pi$ is uniformly dual-bounded, then problem GEN($\mathcal{F}_\pi, \mathcal{X}$) is solvable in quasi-polynomial time for every subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$.*

Proof. Use the joint generation algorithm described in the proof of Proposition 3, starting with the given pair of hypegraphs $\mathcal{X}$ and cY. Suppose that after t iterations the algorithm generates t sets $Z_1, \ldots, Z_t$ all of which belong to $\mathcal{I}(\mathcal{F}_\pi) \setminus Y$. Then $Z_1, \ldots, Z_t$ also belong to $\mathcal{I}(\mathcal{X})$, and hence $t \leq |\mathcal{I}(\mathcal{F}_\pi) \cap \mathcal{I}(\mathcal{X})|$. Now Proposition 4 follows from (1). □

Perhaps surprisingly, many classes of hypergraphs (equivalently, monotone properties) arising in various application turn out to be uniformly dual-bounded. In the next section we survey several inequalities proving uniformly dual-boundedness for a number of hypergraphs (see [10, 12, 14, 17]). Finally, in Section 3 we discuss several applications of the cited results and as well as some generalizations (see [10, 12, 13, 15, 16, 17, 28, 29]).

2 Classes of Uniformly Dual-Bounded Hypergraphs

All monotone properties considered in this paper, can be described in terms a system of monotone inequalities

$$f_i(X) \geq t_i, \quad i = 1, \ldots, r, \tag{2}$$

over the subsets $X \subseteq V$ (or more generally, over elements of products of partially ordered sets), where $t_i \in \mathbb{R}$, $i = 1, \ldots, r$, are given thresholds, and $f_i : 2^V \mapsto \mathbb{R}$ are given functions. We assume that each set-function f_i is *monotone*, i.e. $f(X) \leq f(Y)$ whenever $X \subseteq Y$, and is defined via a (quasi-)polynomial-time evaluation oracle. For a subset $X \subseteq V$, let $\pi(X)$ be the property that X satisfies (2), then we have a membership oracle for $\pi(X)$ – this requires only checking the feasibility of X for all inequalities in (2). Below, we assume that $|\mathcal{D}_\pi| \geq \max\{n, r\}$, and consider four different (but possibly overlapping) classes of functions for which the hypergraph $\mathcal{F}_\pi$ is uniformly dual-bounded. Since the monotone property $\pi(X)$ will be fixed and understood to be the feasibility of X for (2), we will write simply $\mathcal{F}$ instead of $\mathcal{F}_\pi$.

2.1 Polymatroid functions

A monotone function $f : 2^V \mapsto \mathbb{Z}_+$ is called *polymatroid* if it is submodular, i.e.,

$$f(X \cup Y) + f(X \cap Y) \leq f(X) + f(Y)$$

holds for all subsets $X, Y \subseteq V$, and if $f(\emptyset) = 0$. Here we assume that f is integer-valued.

Theorem 1 ([12]) *Let $\mathcal{F}$ be the set of all minimal feasible solutions for a system of polymatroid inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r \max(n, |\mathcal{X}|^{(\log t)/c(n,|\mathcal{X}|)}), \tag{3}$$

where $t = \max\{t_1, \ldots, t_r\}$, and $c = c(n, \beta)$ is the unique positive root of the equation

$$2^c(n^{c/\log\beta} - 1) = 1. \tag{4}$$

In particular, for $\mathcal{X} = \mathcal{F}$ we get $|\mathcal{I}(\mathcal{F})| \leq r \max(n, |\mathcal{F}|^{(\log t)/c(n,|\mathcal{F}|)})$.

Let us remark that by (4), $1 = n^{-c/\log\beta} + (n\beta)^{-c/\log\beta} \geq 2(n\beta)^{-c/\log\beta}$, and hence $\beta^{1/c(n,\beta)} \leq n\beta$. Consequently, we can replace (3) by the simpler but weaker inequality

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r(n|\mathcal{X}|)^{\log t}. \tag{5}$$

On the other hand, for large $|\mathcal{X}|$ the bound of Theorem 1 becomes increasingly stronger than (3). For instance, $c(n, n) = \log(1 + \sqrt{5}) - 1 > .694$, $c(n, n^2) > 1.102$, and $c(n, n^p) \sim \log p$ for large p.

Let us remark next that the bound of Theorem 1 is reasonably sharp. As shown in [12], for any positive integer k there exists a polymatroid function f over $n = k2^k$ elements such that the inequality $f(X) \geq t$ with $t = 2^k$ has $|\mathcal{F}| = 2^{k^2}$ minimal feasible sets and and $|\mathcal{I}(\mathcal{F})| = k\binom{2^k}{2}$ maximal infeasible sets. This gives an infinite family of polymatroid inequalities for which

$$|\mathcal{I}(\mathcal{F})| \geq |\mathcal{F}|^{(.551 \log t)/c(n,|\mathcal{F}|)} \quad \text{and} \quad |\mathcal{I}(\mathcal{F})| \geq (n|\mathcal{F}|)^{(1/3)\log t}, \tag{6}$$

as $k \to \infty$; see section 3.3 for more detail.

It is also worth mentioning that for many classes of polymatroid functions, $|\mathcal{F}|$ cannot be bounded by a quasi-polynomial estimate of the form $(n|\mathcal{I}(\mathcal{F})|)^{\mathrm{poly}\log t}$. Let us consider for instance, a graph $G = l \times K_2$ consisting of l disjoint edges, and let $f(X)$ be the number of edges X intersects, for $X \subseteq V(G)$. Then f is a polymatroid function with $n = 2l$. If we let $t = l$ be the given threshold, then $|\mathcal{I}(\mathcal{F})| = l$ and $|\mathcal{F}| = 2^l$, i.e., $|\mathcal{F}| = 2^{|\mathcal{I}(\mathcal{F})|}$.

Theorem 1 and Proposition 4 imply the following result:

Proposition 5 ([12]) *Consider a system of polymatroid inequalities (2) in which the right hand sides are bounded by a quasi-polynomial in the dimension of the system:*

$$t = \max\{t_1, \dots, t_r\} \leq 2^{\mathrm{polylog}(rn)}. \tag{7}$$

Suppose further that (2) has a quasi-polynomial-time feasibility oracle. Then problem $GEN(\mathcal{F}, \mathcal{X})$ can be solved in quasi-polynomial time, and hence all minimal feasible sets for (2) can be enumerated in incremental quasi-polynomial time.

In general, (7) is a necessary requirement for the validity of Proposition 5. First of all, it is easy to show that any monotone property π can be described by a single polymatroid inequality $f(X) \geq t$ with possibly exponential threshold t (see [14] for more detail). Furthermore, the generation of $\mathcal{F}$ may be NP-hard even for polymatroid functions f defined by a polynomial-time evaluation oracle.

Proposition 6 *There exit polymatroid inequalities $f(X) \geq t$ with polynomial-time computable left-had side, for which problem GEN($\mathcal{F}, \mathcal{X}$) is NP-hard for exponentially large t.*

Proof. The result follows by a polynomial-time reduction from the following enumeration problem for relay graphs with two terminals [35]. Let $G = (U, E)$ be a graph with two distinguished vertices $p, q \in U$. To each edge $e \in E$ is assigned a relay $v \in V = \{1, \dots, n\}$ (two or more distinct edges may be assigned identical relays). Let $\mathcal{F}$ be all the minimal p-q cuts, i.e., all minimal subsets of relays that disconnect p and q. It is known that the problem of incrementally generating $\mathcal{F}$ is NP-hard [35]. For $X \subseteq V$, define $\phi(X)$ to be the number of (not necessarily simple) p-q paths of length $k = |U|$ when only the relays in X are on. Then it is not difficult to see that $\phi(X)$ is a super-modular function, i.e., $\phi(X \cup Y) + \phi(X \cap Y) \geq \phi(X) + \phi(Y)$ for all $X, Y \subseteq V$. Consequently, $f(X) = \phi(V) - \phi(V \setminus X)$ is a polymatroid function. Given $X \subseteq V$, we can compute $\phi(X)$ in polynomial time by computing the k-th power of the adjacency matrix of the graph obtained from G by deleting all edges whose relays $v \notin X$. This gives a polynomial-time evaluation oracle for $f(X)$. Finally, the p-q cuts are exactly the complements of the maximal feasible solutions for the inequality $\phi(X) \leq 0$, or equivalently, all minimal feasible solutions to $f(X) \geq t = \phi(V)$. The latter inequality is polymatroid. □

As we shall see in the next section, enumerating all elements of $\mathcal{I}(\mathcal{F})$ may be NP-hard already for polymatroid inequalities with small right-hand sides.

2.2 Smooth polymatroid functions

Let $f : 2^V \mapsto \mathbb{Z}_+$ be an integer-valued monotone function. For an integer $k \in \mathbb{Z}_+$, let us say that the function f is *k-smooth* if for any $v \in V$ and any $X \subseteq V$, we have

$$f(X \cup \{v\}) - f(X) \leq k.$$

Every 1-smooth polymatroid function is the rank function of some matroid. Note that if $f(X)$ is a k-smooth polymatroid function, then $0 \leq f(X) \leq k|X|$. Hence for systems (2) of k-smooth polymatroid inequalities we may assume without loss of generality that that $\max\{t_1, \ldots, t_r\} \leq kn$.

Theorem 2 *Let $\mathcal{F}$ be the set of all minimal feasible solutions for a system of k-smooth polymatroid inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$\sum_{Y \in \mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})} |V \setminus Y| \leq rn \sum_{X \in \mathcal{X}} |V \setminus X|^{k-1}. \tag{8}$$

In particular, we have $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq rn^k |\mathcal{X}|$, which for $\mathcal{X} = \mathcal{F}$ gives $|\mathcal{I}(\mathcal{F})| \leq rn^k |\mathcal{F}|$

Proof. We first prove (8) for $r = 1$. Given a set $Z \subseteq V$ and a singleton $v \in V \setminus Z$, call v *redundant* for Z if $f(Z \cup \{v\}) = f(Z)$. Denote by $cl(Z)$ the set obtained by adding all redundant elements to Z, then the submodularity of f implies that $f(Z) = f(cl(Z))$. Let $\mathcal{S}$ be the family of all sets of the form $cl(Y \cup \{v\})$, where $Y \in \mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})$ and $v \in V \setminus Y$. Then the left-hand side of (8) does not exceed $n|\mathcal{S}|$. On the other hand, for each set $S \in \mathcal{S}$ we have $t \leq f(S) \leq t - 1 + k$. In particular, S contains some set $X \in \mathcal{X}$ and furthermore, S can be obtained by adding at most $k - 1$ elements to X and then taking the closure of the resulting set. This gives (8) for $r = 1$. Since each maximal infeasible set for a system of monotone inequalities is a maximal infeasible set for some inequality in the system, we also obtain (8) for arbitrary r. □

The following result, implicit in the proof of Theorem 2, strengthens Proposition 5 for systems of k-smooth polymatroid inequalities.

Proposition 7 *All minimal solutions of a system of k-smooth polymatroid inequalities (2) can be generated with polynomial delay if r and k are constants. Furthermore, when r and k are bounded, then problem GEN$(\mathcal{F}, \mathcal{X})$ can be solved efficiently in parallel, i.e., GEN$(\mathcal{F}, \mathcal{X}) \in$ NC.*

There are trivial examples of 1-smooth polymatroid inequalities $f(X) \geq t$ for which the size of $\mathcal{F}$ is exponentially large in the size of $\mathcal{I}(\mathcal{F})$. For instance, let V be set of columns of the $n \times 2n$ matrix $A = [I, I]$, where I is the identity matrix. Given a set $X \subseteq V$, denote by $f(X)$ the dimension of the space spanned by the columns in X. Then the minimal solutions of the matroid inequality $f(X) \geq n$ are the column bases of A, and hence $|\mathcal{F}| = 2^n$. On the other hand, $\mathcal{I}(\mathcal{F})$ is the family of all hyperplanes spanned by the columns of A, and consequently $|\mathcal{I}(\mathcal{F})| = n$. Despite the fact that $\mathcal{I}(\mathcal{F})$ is not dual-bounded, it can be shown that for $r = k = 1$ the set $\mathcal{I}(\mathcal{F})$ can be generated efficiently: all hyperplanes (equivalently, cycles) of a matroid can be generated in incremental polynomial time [12]. Proposition 8 below shows that in general, this result cannot be improved.

Proposition 8 *There are 2-smooth polymatroid inequalities and systems of two 1-smooth polymatroid inequalities for which the generation of $\mathcal{I}(\mathcal{F})$ is NP-hard.*

Proof. Consider the following well-known NP-hard problem: Determine whether a given disjunctive normal form of N binary variables is tautological, i.e., $\phi(x_1,\dots,x_N) = C_1 \vee C_2 \vee \dots \vee C_m \equiv 1$, where each C_j is a conjunction of some literals $x_1, \bar{x}_1, \dots, x_N, \bar{x}_N$. We assume without loss of generality $m \geq 2$ and that for each $i = 1,\dots,N$, both x_i and $\bar{x}_i$ belong to some conjunctions of ϕ, i.e, that ϕ is not monotone or anti-monotone with respect to some variable. We also assume that none of the conjunctions C_j implies another conjunctions $C_{j'}$, i.e., ϕ is irredundant. Naturally, this also means that no conjunction C_j contains a variable and its negation.

We start by replacing each variable x_i in conjunction C_j by a new variable Y_{ij}, and we also replace each negation $\bar{x}_k$ in C_j by another new variable Z_{kj}. After these changes of variables, we obtain a monotone disjunctive normal form $\phi' = C'_1 \vee \dots \vee C'_m$ in $n = |C_1| + \dots + |C_m|$ literals Y_{ij} and Z_{ij}. Next we construct two parallel-series multigraphs G_1 and G_2. The first of these multigraphs is $\bullet\ C''_1\ \bullet\ C''_2\ \bullet \dots \bullet\ C''_m\ \bullet$, where each $\bullet$ is a vertex, and each C''_j consists of as many parallel edges as there are literals in conjunction C'_j. These parallel edges are marked by the corresponding literals Y_{ij} or Z_{ij}. To construct the second multigraph we first compute, for each $i = 1,\dots,N$, two sets of literals $Y_i = \{Y_{i1},\dots,Y_{im}\} \cap V$ and $Z_i = \{Z_{i1},\dots,Z_{im}\} \cap V$, where V is the set of all literals in the monotone disjunctive form ϕ'. Note that since each variable x_i and its negation $\bar{x}_i$ occur in ϕ, none of the $2N$ sets $Y_1, Z_1, \dots, Y_N, Z_N$ is empty. The second multigraph G_2 is obtained by connecting two distinguished vertices $\bullet$ and **o** by N parallel disjoint "paths" of the form $\bullet\ Y'_i\ p_i\ Z'_i\ \mathbf{o}$, where p_i is an intermediate vertex, Y'_i consists of $|Y_i|$ parallel edges labeled by all the literals in Y_i, and similarly, Z'_i consists of $|Z_i|$ parallel edges labeled by the literals of Z_i. Note that each literal in V is assigned to exactly one edge in G_1 and one edge in G_2, so that both G_1 and G_2 have $n = |C_1| + \dots + |C_m|$ edges. Also, denoting by U_1 and U_2 the vertex sets of G_1 and G_2, we have $|U_1| = m + 1$ and $|U_2| = N + 2$.

Consider the system of two inequalities $f_i(X) = |U_i| - c_i(X) \geq |U_i| - 1, \quad i = 1, 2$, where $X \subseteq V$, and where $c_i(X)$ denotes the number of connected components in G_i when only the edges labeled by the literals in X are present. It is easy to see that $f_1(X)$ and $f_2(X)$ are 1-smooth polymatroid functions, and that each set in $\mathcal{I}(\mathcal{F})$ is the complement of a minimal V-*cut*, i.e, a minimal set of literals $X \subseteq V$ that disconnect at least one of the multigraphs G_1 and G_2. (These are exactly the cuts from Example 3 of Section 1.2.) It remains to show that enumerating all such cuts is NP-hard. First of all, there are m trivial cuts $C''_1,\dots,C''_m$ in G_1, all of which are minimal in G_1 because ϕ is irredundant. None of these m cuts disconnects G_2, because $m \geq 2$ and no conjunction in ϕ contains a variable and its negation. Hence $C''_1,\dots,C''_m$ are all minimal V-cuts that disconnect G_1. It remains to enumerate all minimal V-cuts disconnecting G_2. There are N trivial minimal V-cuts $Y_1 \cup Z_1, \dots, Y_N \cup Z_N$, each of which isolates one intermediate vertex $p_i, \quad i = 1,\dots,N$ in G_2. Any additional V-cut in G_2 must disconnect the two distinguished vertices $\bullet$ and **o** by selecting, for each $i = 1,\dots,N$, either all edges in Y_i or all edges Z_i, but not both. Hence such a cut X naturally defines a truth assignment for the variables of the original disjunctive normal form: $x_i = 1$ if and only if all the variables in Y_i are in X. Now it is easily seen that in order for X to keep G_1 connected we must have $\phi(x) = 0$. Hence finding such an additional non-trivial minimal

V-cut is as hard as proving that the given disjunctive normal form ϕ is not tautological.

A similar proof for the multigraph $G_3 = G_1 \bullet G_2$ can be used to show that enumerating all maximal infeasible solutions to a single 2-smooth polymatroid inequality is NP-hard. Specifically, for $X \subseteq V$ define $f_3(X)$ to be $|U_3|$, the number of vertices in G_3, minus the number of connected components in the multigraph $G_3(X)$ obtained by retaining only those edges of G_3 which are labeled by the literals of X. Since there are exactly 2 edges assigned to each literal in V, we conclude that $f_3(X)$ is a 2-smooth polymatroid function. As before, enumerating all maximal infeasible sets for the inequality $f_3(X) \geq |U_3| - 1$ is equivalent to the enumeration of minimal V-cuts for G_1 and G_2. $\square$

2.3 2-Monotonic functions

A monotone function $f : 2^V \mapsto \mathbb{R}$ is called *2-monotonic*, if there exists a permutation $\sigma \in \mathbb{S}_V$ of the ground set V such that $f(X \cup \{v\} \setminus \{u\}) \geq f(X)$ whenever $u \in X \subseteq V$, $v \notin X$ and v precedes u in their σ-order. For instance, any non-negative modular function $f(X) = \sum_{v \in X} w(v)$ is 2-monotonic with respect to the permutation that puts the non-negative singleton weights $w(v)$ in non-increasing order (but not all 2-monotonic functions are modular).

Theorem 3 ([10] (see also [22])) *Let $\mathcal{F}$ be the set of all minimal feasible solutions for a system of 2-monotonic inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r \sum_{X \in \mathcal{X}} |X|.$$

In particular, $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq rn|\mathcal{X}|$, and for $\mathcal{X} = \mathcal{F}$ we get $|\mathcal{I}(\mathcal{F})| \leq rn|\mathcal{F}|$.

Let us remark that the above inequalities are reasonably tight already for systems of non-negative linear inequalities, see Section 3.6. It is also interesting to note that if the the number of inequalities r is bounded by a constant, then the size of $\mathcal{F}$ can also be polynomially bounded by the size of $\mathcal{I}(\mathcal{F})$, and both $\mathcal{F}$ and $\mathcal{I}(\mathcal{F})$ can be incrementally generated in polynomial time [10]. On the other hand, if r is not a constant, problem GEN$(\mathcal{I}(\mathcal{F}), \mathcal{Y})$ is known to be NP-hard [11] (see also [47]).

2.4 Transversal functions

Given a (not necessarily) Sperner hypergraph $\mathcal{H} \subseteq 2^V$ and a non-negative real weight $w(H) \in \mathbb{R}_+$ associated with each edge $H \in \mathcal{H}$, let us call the function $f : 2^V \mapsto \mathbb{R}_+$, defined by

$$f(X) = \sum \{w(H) \mid X \cap H \neq \emptyset,\ H \in \mathcal{H}\}$$

a *(weighted) transversal function.* Note that such a function is submodular, but not generally polymatroid since the weights are not necessarily integral. In contrast to the bound of Theorem 1, the following stronger bound is known for transversal functions.

Theorem 4 ([17]) *Let $f_1, \dots, f_r : 2^V \mapsto \mathbb{R}_+$ be r transversal functions defined by r hypergraphs $\mathcal{H}_1, \dots, \mathcal{H}_r$. Let $\mathcal{F}$ be the set of all minimal feasible solutions for the system of inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq \sum_{i=1}^{r} \sum_{X \in \mathcal{X}} |\{H \in \mathcal{H}_i \mid H \cap X \neq \emptyset\}|.$$

In particular, it follows that $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r|\mathcal{H}_{max}||\mathcal{X}|$, which for $\mathcal{X} = \mathcal{F}$ gives $|\mathcal{I}(\mathcal{F})| \leq r|\mathcal{H}_{max}||\mathcal{F}|$, where $\mathcal{H}_{max}$ is the hypergraph with largest size.

As it was shown in [17], the bound of Theorem 4 is sharp within a polylogarithmic factor. A similar reverse inequality does not hold, in general, and the incremental generation of $\mathcal{I}(\mathcal{F})$ is NP-hard [16].

3 Applications

In this section we give some specific examples of the four classes of functions described in the previous section. We then conclude by an application in data mining, and a generalization of it over products of partially ordered sets.

3.1 Matroids

Several examples of polymatroid functions can be found, for instance, in [45] and [69]. The most general example considered in this paper is the following. Let M be a matroid with rank function $\rho : 2^M \mapsto \{0, 1, \dots, |M|\}$, and let $E_1, \dots, E_n$ be some subsets of M. For each $X \subseteq V \stackrel{\text{def}}{=} \{1, \dots, n\}$, let $f(X) = \rho(\bigcup_{i \in X} E_i)$. Then f is a polymatroid function. In fact, every polymatroid function arises by this construction from some matroid, see [40, 55], and also [45]. Thus Proposition 5 implies that the family of all minimal subsets of $\{E_1, \dots, E_n\}$, the rank of whose union is at least t, can be listed incrementally in quasi-polynomial time (because t is bounded by $|M|$). Since this result holds for systems of polymatroid inequalities, we can also combine any number of matroid examples. Some examples of this construction are considered in Sections 3.3 and 3.4.

3.2 Matroid intersections

Let $M_1, \dots, M_r$ be r matroids on the ground set V of cardinality $|V| = n$. In [43] the question of generating the family $\mathcal{I}$ of all maximal sets independent in all the matroids $M_1, \dots, M_r$ was asked, and an $O(n^{r+2}|\mathcal{I}| \sum_{i=1}^{r} T_i)$ algorithm was given, where T_i is the time required for independence testing in matroid M_i. In contrast to this exponential in r bound, Proposition 5 implies that the matroid intersection problem can be solved in $k^{o(\log k)}$ oracle time, where $k = \max\{r, n, |\mathcal{I}|\}$. Indeed, let $\rho_i : 2^V \mapsto \{0, 1, \dots, n\}$ be the rank function of matroid M_i, for $i = 1, \ldots, r$. Then the functions $f_i : 2^V \mapsto \{0, 1, \ldots, n\}$, defined by $f_i(X) = \rho_i(V \setminus X) + |X| - \rho_i(V)$, $i = 1, \ldots, r$, are 1-smooth polymatroid, and a set $X \subseteq V$ is independent in M_i if and only if $\rho_i(X) \geq |X|$, i.e., $f_i(V \setminus X) \geq |V| - \rho_i(V)$. Thus letting $\mathcal{F} \stackrel{\text{def}}{=} \{V \setminus X \mid X \in \mathcal{I}\}$ to be the complementarity family for $\mathcal{I}$, we conclude that $\mathcal{F}$ is the family of minimal solutions for the system of 1-smooth polymatroid inequalities

$$f_i(X) \geq |V| - \rho_i(V), \quad i = 1, \ldots, r.$$

Therefore we get by Theorem 2,

$$\sum_{Y \in \mathcal{I}(\mathcal{F}) \cap \mathcal{I}(\mathcal{X})} |V \setminus Y| \leq rn|\mathcal{X}|$$

for each $\mathcal{X} \subseteq \mathcal{F}$. This implies the above quasi-polynomial bound on the complexity of generating $\mathcal{F}$.

3.3 Spanning a linear space by linear subspaces

The transversal hypergraph problem is equivalent to the following set covering problem: Given an s-element ground set $\mathcal{R}$ and a family $\mathcal{V}$ of n subsets of $\mathcal{R}$, enumerate all minimal subfamilies of $\mathcal{V}$ which cover the entire set $\mathcal{R}$. Replacing $\mathcal{R}$ by the vector space $\mathbf{F}^s$ over some field $\mathbf{F}$, and replacing each given subset of $\mathcal{V}$ by a linear subspace of $\mathbf{F}^s$, we arrive at the following *space covering* problem: *Given a collection* $\mathcal{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$ *of* n *linear subspaces of* $\mathbf{F}^s$, *enumerate all minimal subsets* X *of* $V = \{1, \ldots, n\}$ *such that* $\mathrm{Span}\langle\bigcup_{i \in X} \mathcal{V}_i\rangle = \mathbf{F}^s$. Generalizing further, consider the polymatroid inequality

$$f(X) = \dim(\bigcup_{i \in X} \mathcal{V}_i) \geq t, \tag{9}$$

where $t \in \{1, \ldots, s\}$ is a given threshold. Then the set $\mathcal{F}$ of minimal solutions to (9) is the collection of all minimal subsets of $\mathcal{V}$ the dimension of whose union is at least t. Theorem 1 then states that for all $t \in \{1, \ldots, s\}$, the size of $\mathcal{I}(\mathcal{F})$ can be bounded by a $\log t$-degree polynomial in n and $|\mathcal{F}|$, and thus all sets in $\mathcal{F}$ can be enumerated in incremental quasi-polynomial time.

We mention here two special cases of the above example. First, when each subspace $\mathcal{V}_i$ is spanned by a subset R_i of vectors from some fixed basis of $\mathbf{F}^s$, the value of $f(X)$ is just the size of $\bigcup_{i \in X} R_i$, which is a transversal function. Hence, by Theorem 4, we get the stronger inequality $|\mathcal{I}(\mathcal{F})| \leq s|\mathcal{F}|$.

Second, when the dimension of each input subspace $\mathcal{V}_i$, $i = 1, \ldots, n$, is bounded by some constant k, the function $f(X)$ is k-smooth. We can thus enumerate all sets in $\mathcal{F}$ with polynomial delay and the size of $\mathcal{I}(\mathcal{F})$ can be bounded by a k-degree polynomial in n and $|\mathcal{F}|$.

It is shown in [12] that the lower bounds (6) can be achieved within the subclass of rank functions defined on the subspaces of a linear space. Namely, there are kl subspaces $V_{ij} \subseteq \mathbb{R}^{2^k}$, $i = 1, \ldots, k$, $j = 0, \ldots, l-1$ of dimension 2^{k-1} each, such that for every i and $j \neq j'$ we have $\dim(V_{ij} \cup V_{ij'}) = 2^k$, while for every $(j_1, j_2, ..., j_k) \in \{0, 1, ..., l-1\}^k$ the inequality $\dim(\bigcup_{i=1}^{k} V_{ij_i}) < 2^k$ holds.

It is also worth mentioning that even though the space covering problem can be solved in incremental quasi-polynomial time, the following close modification of the problem is NP-hard: Enumerate all minimal subsets $X \subseteq V$ such that $\mathrm{Span}\langle \bigcup_{i \in X} \mathcal{V}_i \rangle$ contains a given linear subspace $\mathcal{V}_0$. In fact, the above *subspace* covering problem is NP-hard even when $\mathcal{V}_0$ is a line and $\dim(\mathcal{V}_i) = 2$ for all $i = 1, \ldots, n$ (see [42]).

3.4 Spanning collections of graphs

Let R be a finite set of s vertices and let $E_1, \ldots, E_n \subseteq R \times R$ be a collection of n graphs on R. Given a set $X \subseteq = \{1, \ldots, n\}$ define $c(X)$ to be the number of connected components of graph $(R, \bigcup_{i \in X} E_i)$. Then $c(X)$ is an anti-monotone supermodular function and hence for any integral threshold t, the inequality

$$f(X) = s - c(X) \geq t$$

is polymatroid. In particular, if $t = s - 1$ and $c(\{1, \ldots, n\}) = 1$, then $\mathcal{F}$ is the family of all minimal collections of the input graphs $E_1, \ldots, E_n$ which interconnect all vertices in R. (If the n input graphs are just n disjoint edges, then $\mathcal{F}$ is the set of all spanning trees in the graph $(R, E_1 \cup \ldots \cup E_n)$.)

Since $c(X)$ can be evaluated at any set X in polynomial time, Proposition 2 implies that for each fixed $t \in \{1, \ldots, s\}$, all elements of $\mathcal{F}$ can be enumerated in incremental quasi-polynomial time. For example, given a collection of n *equivalence relations* (partitions) on R, we can enumerate in incremental quasi-polynomial time all minimal subsets of the given relations whose transitive closure puts all elements of R in one equivalence class, or more general, produces at most $s - t$ equivalence classes. Note that this example is a special case of the space covering problem, where $\mathbf{F} = \mathbf{GF}(2)$ and each subspace $\mathcal{V}_i$ is the span of the incidence vectors of the edges of the set E_i, for $i = 1, \ldots, n$.

Interestingly, enumerating all minimal collections of $E_1, \ldots, E_n$ which connect two distinguished vertices $p, q \in R$ turns out to be NP-hard even if the input sets $E_1, \ldots, E_n$ are disjoint and contain at most 2 edges each, see [35] and also [42]. Needless to say that as before, generating all maximal collections of $E_1, \ldots, E_n$ for which the number of connected components of $(R, \bigcup_{i \in X} E_i)$ exceeds a given threshold remains NP-hard.

3.5 Dualization of read-k CNFs

Let

$$\phi(x_1, \ldots, x_n) = \bigwedge_{H \in \mathcal{H}} \bigvee_{i \in H} x_i$$

be a monotone conjunctive form defined by a family of disjunctions $\mathcal{H} \subseteq 2^{\{1,\dots,n\}}$ of n binary variables $x_1, \dots, x_n$. Assume without loss of generality that $\mathcal{H}$ is a Sperner hypergraph, i.e., CNF $\phi(x)$ is irredundant. It is easily seen that the computation of the irredundant disjunctive normal form for $\phi(x)$ is equivalent to the dualization of $\mathcal{H}$. The CNF $\phi(x)$ is called *read-k* if each variable x_i occurs in at most k disjunctions:

$$\max_{i \in V} |\{H \in \mathcal{H} \mid i \in H\}| \leq k.$$

Let us consider *transversal* function $f(X) = |\{H \in \mathcal{H} \mid H \cap X \neq \emptyset\}|$. Obviously, this function is polymatrid, moreover, it is k-smooth whenever CNF ϕ is read-k. Hence for each fixed k, read-k CNFs can be dualized with polynomial delay; see [24] and [27] for alternative proofs of this result.

3.6 Monotone systems of Boolean and integer inequalities

Consider a system of r linear inequalities in n binary variables

$$a_i^T x \geq b_i, \quad x \in \{0,1\}^n, \quad i = 1, \dots, r \tag{10}$$

where $a_1, \dots, a_r$ are given non-negative n-vectors and $b_1, \dots, b_r$ are given scalars. For a set $X \subseteq V = \{1, \dots, n\}$, let $x \in \{0,1\}^n$ be the incidence vector of X, and define $f_i(X) = a_i^T x$. As mentioned above, f_i is 2-monotonic with respect to the permutation that puts the components of a_i in non-increasing order. Accordingly, from Theorem 3 we conclude that for any non-empty set $\mathcal{X} \subseteq \mathcal{F}$ of minimal feasible solutions to (10) we have the inequalities

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r \sum_{x \in \mathcal{X}} p(x) \leq nr|\mathcal{X}|, \tag{11}$$

where $p(x)$ is number of positive components of x. In particular, for any feasible system (10) we have

$$|\mathcal{I}(\mathcal{F})| \leq nr|\mathcal{F}|. \tag{12}$$

The above bounds are sharp when $r = 1$, for instance, for the inequality $x_1 + \ldots + x_n \geq n$. For large r, these bounds are accurate up to a factor polylogarithmic in r. For instance, for any positive integer k, the system of $r = 2^k$ inequalities in $n = 2k$ binary variables

$$x_{i_1} + \cdots + x_{i_k} \geq 1, \ i_1 \in \{1,2\}, \dots, i_k \in \{2k-1, 2k\}$$

has 2^k maximal infeasible binary vectors and only k minimal feasible binary vectors, i.e.,

$$|\mathcal{I}(\mathcal{F})| = \frac{nr}{2(\log r)^2} |\mathcal{F}|.$$

It was shown in [14] that inequalities (11) and (12) actually hold for any *monotone* system of linear inequalities (10), i.e., under the assumption that the feasibility of $x \in \{0,1\}^n$ for

(10) is a monotone property. (This is clearly true when the vectors a_i are non-negative, but in general, the non-negativity of the a_i's is not necessary.) In fact, the inequalities stated in (11) and (12) also hold for any monotone system of r linear inequalities in n *integer* variables

$$a_i^T x \geq b_i,\ x \in \mathcal{C} \stackrel{\text{def}}{=} \{x \in \mathbb{Z}^n \mid 0 \leq x \leq c\},\ i = 1, \ldots, r \tag{13}$$

where c is a given n vector some or all components of which may be infinite. Furthermore, all minimal feasible integer solutions to a given monotone system of integer inequalities (13) can be generated in incremental quasi-polynomial time [10, 11]. As mentioned in Example 2 of Section 1.2, this result should be contrasted with the conjecture of [43] that for non-negative a_i's and $c = (1, \ldots, 1)$ all minimal solutions to (10) cannot be enumerated in in incremental polynomial time unless P=NP. On the other hand, the problem of generating all maximal infeasible binary vectors for (10) is NP-hard already for binary vectors a_i, see [47] and also [14] for more detail.

3.7 Maximal frequent and minimal infrequent sets for binary matrices

The notion of frequent sets in data-mining [64] can be related naturally to the transversal functions considered in Section 2.4. Let V be a finite set of binary attributes of a database. Let $\mathcal{D} : \mathcal{R} \times V \mapsto \{0, 1\}$ be a given $s \times n$ binary matrix representing a set $\mathcal{R}$ of transactions over V. To each subset of columns $X \subseteq V$, we associate the subset $S(X) = S_{\mathcal{D}}(X) \subseteq \mathcal{R}$ of all those rows $i \in \mathcal{R}$ for which $\mathcal{D}(i, j) = 1$ in every column $j \in X$. The cardinality of $S(X)$ is called the *support* of X and it is easy to see that the function

$$f(X) = s - |S(X)|$$

is a transversal function with respect to the hypergraph $\mathcal{H}$ defined by the anti-incidence matrix $\mathcal{D}$.

A column set $X \subseteq V$ is called *t-frequent* if $|S(X)| \geq t$ and otherwise, is said to be *t-infrequent*. Thus the families $\mathcal{F}$ and $\mathcal{I}(\mathcal{F})$ of minimal feasible and maximal infeasible sets for the inequality $f(X) \geq s - t + 1$ correspond, respectively, to the minimal infrequent and maximal frequent sets for $\mathcal{D}$.

The generation of (maximal) frequent sets of a given binary matrix is an important task of knowledge discovery and data mining, e.g., it is used for mining association rules [7, 31, 52, 53, 56, 57, 70], correlations [20], sequential patterns [2], episodes [54], emerging patterns [25], and appears in many other applications. Most practical procedures to generate frequent sets are based on the anti-monotone *Apriori* heuristic (see [1]) and build frequent sets in a bottom-up way, running in time proportional to the number of frequent sets. It was demonstrated recently in [21] that these methods are inadequate in practice when there are (many) frequent sets of large size (see also [4, 39, 44]), due to the fact that the number of frequent sets can be exponentially larger than $|\mathcal{I}(\mathcal{F})|$. Thus these results show that it is perhaps more efficient to find the *boundary* of the frequent sets, i.e. the union $\mathcal{F} \cup \mathcal{I}(\mathcal{F})$

(proposed e.g. in [64]), and use it as a condensed representation of the data set, as suggested in [52]. Furthermore, no algorithm using membership queries "Is X frequent?" can generate all (maximal) frequent sets in fewer than $|\mathcal{F} \cup \mathcal{I}(\mathcal{F})|$ steps (see e.g. [31]). There were several other examples presented in [52] to show the usefulness of maximal frequent sets and minimal infrequent sets, e.g. providing error bounds for the confidence of an arbitrary Boolean rule, in terms of minimal infrequent sets.

It follows by the results of Section 2.4 that the hypergraph of minimal infrequent sets is dual bounded:

$$|\mathcal{I}(\mathcal{F})| \leq (s - t + 1)|\mathcal{F}| \tag{14}$$

for all $t \in \{1, \ldots, s\}$. Let us note that these inequalities are best possible. For instance, they are sharp when $\mathcal{D}$ is an $s \times (s - t + 1)$ matrix in which every entry is 1, except the diagonal entries in the first $s - t + 1$ rows, which are 0. In addition, (14) stays accurate, up to a factor of $\log s$, even when $s \gg n$ and $|\mathcal{F}|$ and $|\mathcal{I}(\mathcal{F})|$ are arbitrarily large. Let us consider for instance a binary matrix $\mathcal{D}$ with $s = 2^k$ rows and $n = 2k$ columns ($k \geq 1$, integer), such that each row contains exactly one 0 and one 1 in each pair of the adjacent columns $\{1, 2\}$, $\{3, 4\}$, $\ldots$, $\{2k - 1, 2k\}$, and in all 2^k possible ways in the $s = 2^k$ rows. It is not difficult to see that for $t = 1$ there are 2^k maximal 1-frequent sets (every row of the matrix is the characteristic vector of a maximal 1-frequent set), and that there are only k minimal 1-infrequent sets, namely $\{2i - 1, 2i\}$ for $i = 1, \ldots , k$. Thus for such examples we have

$$|\mathcal{I}(\mathcal{F})| = \frac{s}{\log s}|\mathcal{F}|.$$

The same examples also show that $|\mathcal{F}|$ cannot be bounded by a quasi-polynomial in $|\mathcal{I}(\mathcal{F})|$, n and s. Indeed, it was shown in [16, 48] that problem $\mathrm{GEN}(\mathcal{I}(\mathcal{F}), \mathcal{Y})$ is NP-hard even if $|\mathcal{Y}| = O(n)$ and $|\mathcal{I}(\mathcal{F})|$ is exponentially large in n whenever $\mathcal{I}(\mathcal{F}) \setminus \mathcal{Y} \neq \emptyset$. A generalization of inequalities (14) for *weighted* maximal frequent and minimal infrequent sets can be found in [17].

If the number of non-zero entries in each row of the matrix $\mathcal{D}$ is bounded by some constant k, then the number of frequent sets is bounded by n^k and therefore, they can be even listed efficiently in parallel. It is interesting to note that if the number of non-zero entries in each row of $\mathcal{D}$ is at least $n - k$ for some constant k, then for any constant t both maximal t-frequent and minimal t-infrequent sets can be also generated efficiently in parallel, see [9] for more details.

3.8 Maximal frequent and minimal infrequent elements in products of partially ordered sets

Let us finally consider a natural generalization of frequent sets to databases defined over products of partially ordered sets (posets). Specifically, let $\mathcal{P} \stackrel{\text{def}}{=} \mathcal{P}_1 \times \ldots \times \mathcal{P}_n$ be the product of n posets. Let us use $\preceq$ to denote the precedence relation in $\mathcal{P}$ and also in $\mathcal{P}_1, \ldots, \mathcal{P}_n$, i.e., if $p = (p_1, \ldots, p_n) \in \mathcal{P}$ and $q = (q_1, \ldots, q_n) \in \mathcal{P}$, then $p \preceq q$ in $\mathcal{P}$ if and only if $p_i \preceq q_i$ in $\mathcal{P}_i$, $i = 1, \ldots, n$. For $\mathcal{A} \subseteq \mathcal{P}$, denote by $\mathcal{A}^+ = \{x \in \mathcal{P} \mid x \succeq a, \text{ for some } a \in \mathcal{A}\}$ and $\mathcal{A}^- = \{x \in \mathcal{P} \mid x \preceq a, \text{ for some } a \in \mathcal{A}\}$, the *filter* and *ideal* generated by $\mathcal{A}$, respectively. Any element in $\mathcal{P} \setminus \mathcal{A}^+$ is called *independent of* $\mathcal{A}$. Let $\mathcal{I}(\mathcal{A})$ be the set of all maximal independent elements for $\mathcal{A}$:

$$\begin{aligned} \mathcal{I}(\mathcal{A}) \;\stackrel{\text{def}}{=}\; & \{p \in \mathcal{P} \mid p \notin \mathcal{A}^+ \text{ and} \\ & (q \in \mathcal{P}, q \succeq p, q \neq p \;\Rightarrow\; q \in \mathcal{A}^+)\}. \end{aligned}$$

Consider a database $\mathcal{D} \subseteq \mathcal{P}$ of transactions, each of which is an n-dimensional vector of attribute values over $\mathcal{P}$. For an element $p \in \mathcal{P}$, let us denote by

$$S(p) = S_{\mathcal{D}}(p) \stackrel{\text{def}}{=} \{q \in \mathcal{D} \mid q \succeq p\},$$

the set of transactions in $\mathcal{D}$ that *support* $p \in \mathcal{P}$. Note that, by this definition, the function $|S(\cdot)| : \mathcal{P} \mapsto \{0, 1, \ldots, |\mathcal{D}|\}$ is an anti-monotone function, i.e., $|S(p)| \leq |S(q)|$, whenever $p \succeq q$.

Given $\mathcal{D} \subseteq \mathcal{P}$ and an integer threshold t, let us say that an element $p \in \mathcal{P}$ is t-frequent if it is supported by at least t transactions in the database, i.e., if $|S_{\mathcal{D}}(p)| \geq t$. Conversely, $p \in \mathcal{P}$ is said to be t-infrequent if $|S_{\mathcal{D}}(p)| < t$. Since the function $|S_{\mathcal{D}}(\cdot)|$ is anti-monotone, we may restrict our attention only to *maximal frequent* and *minimal infrequent* elements. Denote by $\mathcal{F} = \mathcal{F}_{\mathcal{D},t}$ the set of all minimal t-infrequent elements of $\mathcal{P}$ with respect to the database $\mathcal{D}$. Then $\mathcal{I}(\mathcal{F}_{\mathcal{D},t})$ is the set of all maximal t-frequent elements.

As in the Boolean case, the separate and joint generation of maximal frequent and minimal infrequent elements of a database can be used for finding association rules in data mining applications. If the database $\mathcal{D}$ contains categorical (e.g., zip code, make of car), or quantitative (e.g., age, income) attributes, and the corresponding posets $\mathcal{P}_i$ are total orders, then the above generation problems can be used to mine the so called *quantitative* association rules [66]. More generally, each attribute a_i in the database can assume values belonging to some partially ordered set $\mathcal{P}_i$. For example, [65] describes applications where items in the database belong to sets of *taxonomies* (or *is-a hierarchies*), and proposes several algorithms for mining association rules among these hierarchical data (see also [37, 38]). Furthermore, many data analysis applications assume data values ranging over lattices of small size, see e.g. [18].

Extending the notion of uniformly dual-bounded hypergraphs to subsets of products of posets in the obvious way, it can be derived from Theorem 4 that if each poset $\mathcal{P}_i$ is a join semi-lattice (i.e. for every two elements $x, y \in \mathcal{P}_i$, there is a unique upper bound $x \vee y$), then for any integer threshold t, the set of minimal infrequent elements $\mathcal{F}_{\mathcal{D},t}$ of a database $\mathcal{D}$ is uniformly dual-bounded:

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F}_{\mathcal{D},t})| \leq (|\mathcal{D}| - t + 1)|\mathcal{X}|$$

for any nonempty $\mathcal{X} \subseteq \mathcal{F}_{\mathcal{D},t}$. It is also easily verified that for such uniformly dual-bounded subsets of a poset $\mathcal{P}$, the incremental generation of these subsets reduces in polynomial time to the following natural generalization of the hypergraph transversal problem on products of posets:

DUAL$(\mathcal{P}, \mathcal{A}, \mathcal{B})$**:** *Given a subset* $\mathcal{A} \subseteq \mathcal{P}$ *in a poset* $\mathcal{P}$ *and a collection of maximal independent elements* $\mathcal{B} \subseteq \mathcal{I}(\mathcal{A})$*, either find a new maximal independent element* $p \in \mathcal{I}(\mathcal{A}) \setminus \mathcal{B}$*, or prove that* $\mathcal{A}$ *and* $\mathcal{B}$ *form a dual pair:* $\mathcal{B} = \mathcal{I}(\mathcal{A})$.

The complexity of the dualization problem for products of general posets is an open question. Two special cases are known to be quasi-polynomially solvable [28, 29], namely when the underlying precedence graph of each poset $\mathcal{P}_i$ is a rooted tree (is-a hierarchy), and when each poset $\mathcal{P}_i$ a join semi-lattice of bounded width (where the width of $\mathcal{P}_i$ is the maximum size of an antichain in $\mathcal{P}_i$). It follows then that generating minimal infrequent elements in these classes of posets can be done in incremental quasi-polynomial time.

References

[1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo, Fast discovery of association rules, in U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy eds., *Advances in Knowledge Discovery and Data Mining*, 307-328, AAAI Press, Menlo Park, California, 1996.

[2] R. Agrawal and R. Srikant, Mining sequential patterns, *Proc. 11th International Conference on Data Engineering, 1995*, pp.3-14.

[3] M. Anthony and N. Biggs, *Computational Learning Theory. An Introduction*, Cambridge University Press, Cambridge, UK, 1992.

[4] R.J. Bayardo, Efficiently mining long patterns from databases, *Proc. the 1998 ACM-SIGMOD International Conference on Management of Data*, pp. 85-93.

[5] P. Bertolazzi and A. Sassano, An $O(mn)$ time algorithm for regular set-covering problems, *Theoretical Computer Science* 54 (1987) 237-247.

[6] J. C. Bioch and T. Ibaraki, Complexity of identification and dualization of positive Boolean functions, *Information and Computation* 123 (1995) 50-63.

[7] M. M. Bongard, *Pattern Recognition*, Rochelle Park, NJ Hayden Book Co, Spartan Books, 1970.

[8] E. Boros, V. Gurvich, and P.L. Hammer, Dual subimplicants of positive Boolean functions, *Optimization Methods and Software*, 10 (1998) 147-156. (RUTCOR Research Report 11-1993).

[9] E. Boros, K. Elbassioni, V. Gurvich and L. Khachiyan, An efficient incremental algorithm for generating all maximal independent sets in hypergraphs of bounded dimension, *Parallel Processing Letters*, 10 (2000) 253-266.

[10] E. Boros, K. Elbassioni, V. Gurvich, L. Khachiyan and K. Makino, Generating all minimal integer solutions to a monotone system of linear inequalities, *Automata, Languages and Programming, 28-th International Colloquium, ICALP 2001* (F. Orejas, P.G. Spirakis and Jan van Leeuwen, eds.), *Lecture Notes in Computer Science* 2076, pp. 92-103, 2001.

[11] E. Boros, K. Elbassioni, V. Gurvich, L. Khachiyan and K. Makino, Dual-bounded generating problems: all minimal integer solutions of a monotone systems of linear inequalities, to appear in *SIAM J. Comput* (DIMACS Technical Report 2001-12, Rutgers University.)

[12] E. Boros, K. Elbassioni, V. Gurvich and L. Khachiyan, An inequality for polymatroid functions and its applications, to appear in *Discrete Applied Mathematics*, 2002. (DIMACS Technical Report 2001-14, Rutgers University.)

[13] E. Boros, K. Elbassioni, V. Gurvich and L. Khachiyan, Matroid intersections, polymatroid inequalities, and related problems, to appear in: *Proc. 27th International Symposium on Mathematical Foundations of Computer Science, MFCS 2002.*

[14] E. Boros, V. Gurvich, L. Khachiyan and K. Makino, Dual-bounded generating problems: partial and multiple transversals of a hypergraph, *SIAM J. Comput.*, 30 (2001) 2036-2050.

[15] E. Boros, V. Gurvich, L. Khachiyan and K. Makino, Generating partial and multiple transversals of a hypergraph, *Automata, Languages and Programming, 27th International Colloquium, ICALP 2000* (U. Montanari, J.D.P. Rolim, andE. Welzl, eds.), *Lecture Notes in Computer Science* 1853, pp. 588-599, 2000.

[16] E. Boros, V. Gurvich, L. Khachiyan and K. Makino, On the complexity of generating maximal frequent and minimal infrequent sets, *Proc. 19th International Symposium on Theoretical Aspects of Computer Science, STACS 2002*, (H. Alt and A. Ferreira, eds.), *Lecture Nores in Computer Science* 2285, pp. 133-141, 2002.

[17] E. Boros, V. Gurvich, L. Khachiyan and K. Makino, Generating weighted transversals of a hypergraph, *Proc. 2nd Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications*, pp. 13-22. Hungarian Academy of Sciences, Budapest, Hungary, April 20-23, 2001. (DIMACS Technical Report 2000-17, Rutgers University.)

[18] E. Boros, P.L. Hammer and J.N. Hooker, Predicting cause-effect relationships from incomplete discrete observations, *SIAM J. Discrete Math.*, 7 (1994), 481-491.

[19] E. Boros, P. L. Hammer, T. Ibaraki and K. Kawakami, Polynomial time recognition of 2-monotonic positive Boolean functions given by an oracle, *SIAM J. Comput.*, 26 (1997) 93-109.

[20] S. Brin, R. Motwani, and C. Silverstein, Beyond market basket: Generalizing association rules to correlations, *Proc. the 1997 ACM-SIGMOD Conference on Management of Data*, pp. 265-276.

[21] S. Brin, R. Motwani, J. Ullman, and S. Tsur, Dynamic itemset counting and implication rules for market basket data, *Proc. the 1997 ACM-SIGMOD Conference on Management of Data*, pp. 255-264.

[22] Y. Crama, Dualization of regular Boolean functions, *Discrete Applied Mathematics*, 16 (1987) 79-85.

[23] C J. Colbourn, *The Combinarotics of Network Reliability*, Oxford University Press, New Your, 1987

[24] C. Domingo, N. Mishra, and L. Pitt, Efficient read-k monotone CNF/DNF dualization using a learning with membership queries approach, *Machine Learning*, 37 (1999) 89-110.

[25] G. Dong and J. Li, Efficient mining of emerging patterns, *Proc. the 1999 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 43-52.

[26] T. Eiter and G. Gottlob, Identifying the minimal transversals of a hypergraph and related problems, *SIAM J. Comput.*, 24 (1995) 1278-1304.

[27] T. Eiter, G. Gottlob, K. Makino, New results on monotone dualization and generating hypergraph transversals, to appear in *Proc. 34th Annual ACM Symposium on Theory of Computing (STOC 2002).*

[28] K. Elbassioni, On dualization in products of forests, in: *Proc. 19th International Symposium on Theoretical Aspects of Computer Science, STACS 2002*, (H. Alt and A. Ferreira, eds.), *Lecture Nores in Computer Science* 2285, pp. 143-153, 2002.

[29] K. Elbassioni, Dualization in products of bounded-width lattices, Manuscript.

[30] M. L. Fredman and L. Khachiyan, On the complexity of dualization of monotone disjunctive normal forms, *J. Algorithms*, 21 (1996) 618-628.

[31] D. Gunopulos, R. Khardon, H. Mannila, and H. Toivonen, Data mining, hypergraph transversals and machine learning, *Proc. 16th ACM-SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, (1997) pp. 12-15.

[32] V. Gurvich, To the theory of multistep games, *U.S.S.R. Comput. Math. and Math. Phys.*, 13 (1973) 1485-1500.

[33] V. Gurvich, Nash-solvability of games in pure strategies, *U.S.S.R. Comput. Math. and Math. Phys.*, 15 (1975) 357-371.

[34] V. Gurvich, Nash equilibrium in pure strategies, Soviet Math. Dokl., 38, 3 (1988) 597-602.

[35] V. Gurvich and L. Khachiyan, On generating the irredundant conjunctive and disjunctive normal forms of monotone Boolean functions, *Discrete Applied Mathematics*, 96-97 (1999) 363-373. (See also Rutcor Research Report, RRR-35-95, Rutgers University.)

[36] V. Gurvich and L. Khachiyan, On the frequency of the most frequently occurring variable in dual DNFs, *Discrete Mathematics* 169 (1997) 245-248.

[37] J. Han, Y. Cai and N. Cercone, Data driven discovery of quantitative rules in relational databases, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5 No. 1, (1993) pp. 29-40.

[38] J. Han and Y. Fu, Discovery of multiple-level association rules from large databases, *Proc. 21st International Conference on Very Large Data Bases*, pp. 420-431, 1995.

[39] J. Han, J. Pei, and Y. Yin, Mining frequent patterns without candidate generation, *Proc. the 2000 ACM-SIGMOD Conference on Management of Data*, pp. 1-12.

[40] T. Helgason, Aspects of the theory of hypermatroids, in Hypergraph Seminar (C. Berge and D.K. Ray-Chaudhuri eds.), Lecture Notes in Math. 411 (1975) Springer, 191-214.

[41] D. S. Johnson, M. Yannakakis and C. H. Papadimitriou, On generating all maximal independent sets, *Information Processing Letters*, 27 (1988) 119-123.

[42] L. Khachiyan, Transversal hypergraphs and families of polyhedral cones, in : *Advances in Convex Analysis and Global Optimization*, Honoring the memory of K. Carathéodory, (N. Hadjisavvas and P. Pardalos Eds.), 105-118, Kluwer Academic Publishers, Dordrecht/Boston/London, 2001.

[43] E. Lawler, J. K. Lenstra and A. H. G. Rinnooy Kan, Generating all maximal independent sets: NP-hardness and polynomial-time algorithms, *SIAM J. Comput.*, 9 (1980) 558-565.

[44] D. Lin and Z.M. Kedem, Pincer-search: a new algorithm for discovering the maximum frequent set, *Proc. 6th European Conference on Extending Database Technology*, to appear.

[45] L. Lovász, Submodular functions and convexity, in *Mathematical Programming: The State of the Art*, Bonn 1982, pp. 235-257, (Springer Verlag, 1983).

[46] K. Makino, Efficient dualization of $O(\log n)$-term monotone disjunctive normal forms, to appear in *Discrete Applied Mathematics.* (Technical Report 00-07, Discrete Mathematics and Systems Science, Osaka University, 2000.)

[47] K. Makino and T. Ibaraki, Interior and exterior functions of Boolean functions, *Discrete Applied Mathematics*, 69 (1996) 209-231.

[48] K. Makino and T. Ibaraki, Inner-core and outer-core functions of partially defined Boolean functions, *Discrete Applied Mathematics*, 96-97, (1999), 307-326.

[49] K. Makino and T. Ibaraki. The maximum latency and identification of positive Boolean functions. *SIAM J. Comput.*, 26 (1997) 1363–1383.

[50] K. Makino and T. Ibaraki, A fast and simple algorithm for identifying 2-monotonic positive Boolean functions, *J. Algorithms*, 26 (1998) 291-305.

[51] H. Manilla and K. J. Räihä, Design by example: An application of Armstrong relations, *Comp. System. Sci.* 22 (1986) 126-141.

[52] H. Mannila and H. Toivonen, Multiple uses of frequent sets and condensed representations, *Proc. 2nd International Conference on Knowledge Discovery and Data Mining*, (1996) pp. 189-194.

[53] H. Mannila and H. Toivonen, Levelwise search and borders of theories in knowledge discovery. Series of Publications C C-1997-8, University of Helsinki, Department of Computer Science (1997).

[54] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1 (1997), 259-289.

[55] C.G.H. McDiarmid, Rado's theorem for polymatroids, *Math. Proc. Cambridge Phil. Soc.* 78 (1975) 263-281.

[56] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, Discovering frequent closed itemsets for association rules. *Proc. 7th ICDT Conference*, Jerusalem, Israel, January 10-12, 1999; *Lecture Notes in Computer Science*, 1540, pp. 398-416, Springer Verlag, 1999.

[57] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, Closed set based discovery of small covers for association rules, *Proc. 15emes Journees Bases de Donnees Avancees, BDA*, pp. 361-381, 1999.

[58] U. N. Peled and B. Simeone, Polynomial-time algorithm for regular set-covering and threshold synthesis, *Discrete Applied Mathematics* 12 (1985) 57-69.

[59] U. N. Peled and B. Simeone, An $O(nm)$-time algorithm for computing the dual of a regular Boolean function, *Discrete Applied Mathematics* 49 (1994) 309-323.

[60] J. S. Provan and D. R. Shier, A Paradigm for Listing (s, t)-Cuts in Graphs. *Algorithmica* 15(4) (1996) 351-372

[61] K.G. Ramamurthy, *Coherent Structures and Simple Games*, Kluwer Academic Publishers, Dordrecht,the Netherlands, 1990

[62] R. c. Read, Every one is a winner, or how to avoid isomorphism when cataloging combinatorial configurations, *Ann. Discrete Math.*,2 (1978) 107-120

[63] R. C. Read and R. E. Tarjan, Bounds on backtrack algorithms for listing cycles, paths, and spanning trees, *Networks*, 5 (1975) 237-252.

[64] R. H. Sloan, K. Takata, G. Turan, On frequent sets of Boolean matrices, *Annals of Mathematics and Artificial Intelligence* 24 (1998) 1-4.

[65] R. Srikant and R. Agrawal, Mining generalized association rules, *Proc. 21st International Conference on Very Large Data Bases*, pp. 407-419, 1995.

[66] R. Srikant and R. Agrawal, Mining quantitative association rules in large relational tables, *Proc. of the ACM-SIGMOD 1996 Conference on Management of Data*, pp. 1-12, 1996.

[67] H. Tamaki. Space-efficient enumeration of minimal transversals of a hypergraph. IPSJ-AL 75 (2000) 29-36.

[68] S. Tsukiyama, M. Ide, H. Ariyoshi and I. Shirakawa, A new algorithm for generating all maximal independent sets, *SIAM J. Comput.*, 6 (1977) 505-517.

[69] D.J.A. Welsh, *Matroid Theory*, Academic Press, London, New York, San Francisco 1976.

[70] M.J. Zaki and M. Ogihara, Theoretical foundations of association rules, *3rd SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, June 1998.

[71] *Principles of Database and Knowledge Base Systems*, Vols. 1 and 2, Computer Science Press, Rockville, MD, 1988.