An Intersection Inequality for Discrete Distributions and Related Generation Problems*

Endre Boros[1], Khaled Elbassioni[1], Vladimir Gurvich[1], Leonid Khachiyan[2], and Kazuhisa Makino[3]

[1] RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; {boros,elbassio,gurvich}@rutcor.rutgers.edu

[2] Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway NJ 08854-8003; leonid@cs.rutgers.edu

[3] Division of Systems Science, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka, 560-8531, Japan; makino@sys.es.osaka-u.ac.jp

Abstract. Given two finite sets of points $\mathcal{X}, \mathcal{Y}$ in $\mathbb{R}^n$ which can be separated by a nonnegative linear function, and such that the componentwise minimum of any two distinct points in $\mathcal{X}$ is dominated by some point in $\mathcal{Y}$, we show that $|\mathcal{X}| \le n|\mathcal{Y}|$. As a consequence of this result, we obtain quasi-polynomial time algorithms for generating all maximal integer feasible solutions for a given monotone system of separable inequalities, for generating all p-inefficient points of a given discrete probability distribution, and for generating all maximal empty hyper-rectangles for a given set of points in $\mathbb{R}^n$. This provides a substantial improvement over previously known exponential algorithms for these generation problems related to Integer and Stochastic Programming, and Data Mining. Furthermore, we give an incremental polynomial time generation algorithm for monotone systems with fixed number of separable inequalities, which, for the very special case of one inequality, implies that for discrete probability distributions with independent coordinates, both p-efficient and p-inefficient points can be separately generated in incremental polynomial time.

1 Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets of points in $\mathbb{R}^n$ such that

(P1) $\mathcal{X}$ and $\mathcal{Y}$ can be separated by a nonnegative linear function: $w(x) > t \ge w(y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, where $w(x) = \sum_{i=1}^n w_i x_i$, $w_1, \ldots, w_n \in \mathbb{R}_+$ are nonnegative weights, and $t \in \mathbb{R}$ is a real threshold,

* The research of the first four authors 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. The fifth author was supported in part by the Scientific Grant in Aid of the Ministry of Education, Science, Sports, Culture and Technology of Japan.

J.C.M. Baeten et al. (Eds.): ICALP 2003, LNCS 2719, pp. 543–555, 2003.

(P2) For any two distinct points $x, x' \in \mathcal{X}$, their componentwise minimum $x \wedge x'$ is dominated by some $y \in \mathcal{Y}$, i.e. $x \wedge x' \leq y$.

Given $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$ satisfying properties (P1) and (P2), one may ask the question of how large the size of $\mathcal{X}$ can be in terms of the size of $\mathcal{Y}$. For instance, if $\mathcal{X}$ is the set of the n-dimensional unit vectors, and $\mathcal{Y} = \{\mathbf{0}\}$ is the set containing only the origin, then $\mathcal{X}$ and $\mathcal{Y}$ satisfy properties (P1), (P2), and the ratio between their cardinalities is n. We shall show that this is actually an extremal case:

Lemma 1 (Intersection Lemma). *If $\mathcal{X}$ and $\mathcal{Y} \neq \emptyset$ are two finite sets of points in $\mathbb{R}^n$ satisfying properties* (P1) *and* (P2) *above, then*

$$|\mathcal{X}| \leq n|\mathcal{Y}|. \tag{1}$$

An analogous statement for binary sets $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$ was shown in [6]. Let us also recall from [6] that condition (P1) is important, since without that $|\mathcal{X}|$ could be exponentially larger than $|\mathcal{Y}|$, already in the binary case.

Let us also remark that the nonnegativity of the weight vector w is also important. Consider for instance $\mathcal{Y} = \{(1, 1, \ldots, 1)\}$ and an arbitrary number of points in the set $\mathcal{X}$ such that $0 \leq x_i < 1$ for all $x \in \mathcal{X}$ and $i = 1, \ldots, n$. Then clearly (P2) holds, and (P1) is satisfied with $w = (-1, 0, \ldots, 0)$ and $t = -1$. However, it is impossible to bound the cardinality of $\mathcal{X}$ in terms of n and $|\mathcal{Y}| = 1$.

Let us further note that, due to the strict separation in (P1), we may assume without loss of generality that all weights are positive $w > 0$. In fact, it would be even enough to prove the lemma with $w = (1, 1, \ldots, 1)$, since scaling the ith coordinates of all points in $\mathcal{X} \cup \mathcal{Y}$ by $w_i \geq 0$ for $i = 1, \ldots, n$ always transforms the input into one satisfying (P1) with $w = (1, 1, \ldots, 1)$. Clearly, such scaling preserves the relative order of the ith coordinates of the points, and scales properly their componentwise minimum, thus the transformed point sets will satisfy (P2) as well.

As a consequence of the above lemma, we obtain new results on the complexity of several generation problems, including:

Monotone systems of separable inequalities: Given a system of inequalities on sums of single-variable monotone functions, generate all maximal feasible integer solutions of the system.

p-Efficient and p-inefficient points of discrete probability distributions: Given a random variable $\xi \in \mathbb{Z}^n$, generate all p-inefficient points, i.e. maximal vectors $x \in \mathbb{Z}^n$ whose cumulative probability $\Pr[\xi \leq x]$ does not exceed a certain threshold p, and/or generate all p-efficient points, i.e. minimal vectors $x \in \mathbb{Z}^n$ for which $\Pr[\xi \leq x] \geq p$. This problem has applications in Stochastic Programming [10,22].

Maximal k-boxes: Given a set of points in $\mathbb{R}^n$ and a nonnegative integer k, generate all maximal n-dimensional intervals (*boxes*), which contain at most k of the given points in their interior. Such intervals are called empty boxes or empty rectangles, when $k = 0$. This problem has applications in computational geometry, data mining and machine learning [1,2,8,11,16,17,20, 21].

These problems are described in more details in the following sections. What they have in common is that each can be modelled by a property π over a set of vectors $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2 \times \cdots \times \mathcal{C}_n$, where $\mathcal{C}_i$, $i = 1, \ldots, n$ are finite subsets of the reals, and π is anti-monotone, i.e. if $x, y \in \mathcal{C}$, $x \geq y$, and x satisfies property π, then y also satisfies π. Each problem in turn can be stated as of incrementally generating the family $\mathcal{F}_\pi$ of all *maximal* elements of $\mathcal{C}$ satisfying an anti-monotone property π:

GEN$(\mathcal{F}_\pi, \mathcal{E})$**:** *Given an anti-monotone property π, and a subfamily $\mathcal{E} \subseteq \mathcal{F}_\pi$ of the maximal elements satisfying π, either find a new maximal element $x \in \mathcal{F}_\pi \setminus \mathcal{E}$, or prove that $\mathcal{E} = \mathcal{F}_\pi$.*

Clearly, the entire family $\mathcal{F}_\pi$ can be generated by initializing $\mathcal{E} = \emptyset$ and iteratively solving the above problem $|\mathcal{F}_\pi| + 1$ times.

For a subset $\mathcal{A} \subseteq \mathcal{C}$, denote by $\mathcal{I}(\mathcal{A})$ the set of maximal independent elements of $\mathcal{A}$, i.e. the set of those elements $x \in \mathcal{C}$ that are maximal with respect to the property that $x \not\geq a$ for all $a \in \mathcal{A}$. Then $\mathcal{I}^{-1}(\mathcal{A})$ is the set of elements $x \in \mathcal{C}$ that are minimal with the property that $x \not\leq a$ for all $a \in \mathcal{A}$. In particular, $\mathcal{I}^{-1}(\mathcal{F}_\pi)$ denotes the family of *minimal* elements of $\mathcal{C}$ which *do not* satisfy property π.

Following [6], let us call $\mathcal{F}_\pi$ *uniformly dual-bounded*, if for every subfamily $\mathcal{E} \subseteq \mathcal{F}_\pi$ we have

$$|\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_\pi)| \leq p(|\pi|, n, |\mathcal{E}|) \tag{2}$$

for some polynomial $p(\cdot)$, where $|\pi|$ denotes the length of the description of property π. It is known that for uniformly dual-bounded families $\mathcal{F}_\pi$ of subsets of a discrete box $\mathcal{C}$ problem GEN$(\mathcal{F}_\pi, \mathcal{E})$ can be reduced in polynomial time to the following *dualization* problem on boxes (see [4] and also [3,13]):

DUAL$(\mathcal{C}, \mathcal{A}, \mathcal{B})$**:** *Given an integer box $\mathcal{C}$, a family of vectors $\mathcal{A} \subseteq \mathcal{C}$ and a subset $\mathcal{B} \subseteq \mathcal{I}(\mathcal{A})$ of its maximal independent vectors, either find a new maximal independent vector $x \in \mathcal{I}(\mathcal{A}) \setminus \mathcal{B}$, or prove that no such vector exists, i.e., $\mathcal{B} = \mathcal{I}(\mathcal{A})$.*

It is furthermore known that problem DUAL$(\mathcal{C}, \mathcal{A}, \mathcal{B})$ can be solved in *poly*(n) + $m^{o(\log m)}$ time, where $m = |\mathcal{A}| + |\mathcal{B}|$ (see [4,12]). However, it is still open whether DUAL$(\mathcal{C}, \mathcal{A}, \mathcal{B})$ has a polynomial time algorithm (e.g., [3,12,19]).

For each of the problems described above, it will be shown that the families $\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_\pi)$ and $\mathcal{E} \subseteq \mathcal{F}_\pi$ are, respectively, in one to one correspondence with two sets of points $\mathcal{X}, \mathcal{Y}$ satisfying the conditions of Lemma 1. Thus, by Lemma 1 we can derive (2), which in its turn is sufficient for the efficient generation of the family $\mathcal{F}_\pi$ (see [4]).

In particular, it will follow that each of the above generation problems can be solved incrementally in *quasi-polynomial time*. Furthermore, we give incremental *polynomial-time* algorithms for generating

- all maximal feasible, and separately, all minimal infeasible integer vectors for systems with fixed number of monotone separable inequalities, and
- all p-efficient, and separately, all p-inefficient points of discrete probability distributions with independent coordinates

2 Systems of Monotone Separable Inequalities

For $i = 1, 2, \ldots, n$, let l_i, u_i be given integers, $l_i \leq u_i$, and let $\mathcal{C}_i \stackrel{\text{def}}{=} \{l_i, l_i + 1, \ldots, u_i\}$. A function $f : \mathcal{C}_i \mapsto \mathbb{R}$ is called *monotone* if, for $x, y \in \mathcal{C}_i$, $f(x) \geq f(y)$ whenever $x \geq y$. Let $f_{ij} : \mathcal{C}_i \mapsto \mathbb{R}$, $i = 1, 2, \ldots, n$, $j = 1, \ldots, r$ be polynomially-computable monotone functions, and consider the system of inequalities

$$\sum_{i=1}^{n} f_{ij}(x_i) \leq t_j, \quad j = 1, \ldots, r, \tag{3}$$

over the elements $x \in \mathcal{C} = \{x \in \mathbb{Z}^n \mid l \leq x \leq u\}$, where $l = (l_1, \ldots, l_n)$, $u = (u_1, \ldots, u_n)$, and $t = (t_1, \ldots, t_r)$ is a given r-dimensional real vector.

Let us denote by $\mathcal{F}_t$ the set of maximal feasible solutions for (3), and thus $\mathcal{I}^{-1}(\mathcal{F}_t)$ represents the set of minimal infeasible vectors for (3).

Generalizing results on monotone systems of *linear* inequalities from [4], we will now use Lemma 1 to prove the following:

Theorem 1. *If $\mathcal{F}_t$ is the family of maximal feasible solutions of* (3)*, and $\mathcal{E} \subseteq \mathcal{F}_t$ is non-empty, then*

$$\left|\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_t)\right| \leq rn|\mathcal{E}|. \tag{4}$$

In particular, $|\mathcal{I}^{-1}(\mathcal{F}_t)| \leq rn|\mathcal{F}_t|$.

Proof. Let us define a monotonic mapping $\phi : \mathcal{C} \mapsto \mathbb{R}^n$ by setting $\phi(x) = (f_{1j}(x_1), \ldots, f_{nj}(x_n))$ for $x \in \mathcal{C}$. Let $\mathcal{Y} \stackrel{\text{def}}{=} \{\phi(x) \mid x \in \mathcal{E}\}$, and let $\mathcal{X}_j \stackrel{\text{def}}{=} \{\phi(x) \mid x \in \mathcal{I}^{-1}(\mathcal{E}), \sum_{i=1}^{n} f_{ij}(x_i) > t_j\}$, for $j = 1, \ldots, r$. In other words, $\mathcal{X}_j$ is the ϕ-mapping of those minimal infeasible solutions of (3) in $\mathcal{I}^{-1}(\mathcal{E})$ which violate the jth inequality. Since the functions f_{ij} are monotone, and since we consider only maximal feasible or minimal infeasible vectors for (3), the mappings $\mathcal{E} \longrightarrow \mathcal{Y}$ and $\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_t) \longrightarrow \mathcal{X}_1 \cup \cdots \cup \mathcal{X}_r$ are one-to-one.

It is also easy to see that the sets $\mathcal{X}_j$ and $\mathcal{Y}$ satisfy the conditions of Lemma 1 with $w = (1, 1, \ldots, 1)$ and $t = t_j$, for $j = 1, \ldots, r$, and thus (4) follows readily by Lemma 1. □

Since by (4) the family $\mathcal{F}_t$ is uniformly dual-bounded, the results of [4], as we cited earlier, directly imply the following.

Corollary 1. *Problem GEN($\mathcal{F}_t, \mathcal{X}$) of incrementally generating maximal feasible solutions for (3) can be solved in $k^{o(\log k)}$ time, where $k = \max\{n, r, |\mathcal{X}|, \log(\|u - l\|_\infty + 1)\}$.*

It should be mentioned that in contrast to (4), the size of $\mathcal{F}_t$ cannot be bounded by a polynomial in n, r, and $|\mathcal{I}^{-1}(\mathcal{F}_t)|$, even for monotone systems of linear inequalities. However, for systems (3) with constant r, we shall show that such a bound exists, and further that the generation problem can be solved in polynomial time:

Theorem 2. *If $\mathcal{F}_t$ is the family of maximal feasible solutions of* (3)*, and $\mathcal{E} \subseteq \mathcal{I}^{-1}(\mathcal{F}_t)$ is non-empty, then*

$$|\mathcal{I}(\mathcal{E}) \cap \mathcal{F}_t| \leq (n|\mathcal{E}|)^r. \tag{5}$$

In particular, $|\mathcal{F}_t| \leq \Big(n|\mathcal{I}^{-1}(\mathcal{F}_t)|\Big)^r$.

Theorem 3. *If the number of inequalities in (3) is bounded, then both the maximal feasible and minimal infeasible vectors can be generated in incremental polynomial time, in n, r and $\sum_{i=1}^n |\mathcal{C}_i|$.*

The proofs of Theorem 2 and 3 will be given in Section 6. In the next section, we consider an application of Theorem 3 for the case of $r = 1$.

3 *p*-Efficient and *p*-Inefficient Points of Probability Distributions

Let ξ be an n-dimensional random variable on $\mathbb{Z}^n$, with a finite support $\mathcal{S} \subseteq \mathbb{Z}^n$, i.e., $\sum_{q \in \mathcal{S}} \Pr[\xi = q] = 1$, and $\Pr[\xi = q] > 0$ for $q \in \mathcal{S}$. Given a threshold probability $p \in [0,1]$, a point $x \in \mathbb{Z}^n$ is said to be *p-efficient* if it is minimal with the property that $\Pr[\xi \leq x] > p$. Let us conversely say that $x \in \mathbb{Z}^n$ is *p-inefficient* if it is maximal with the property that $\Pr[\xi \leq x] \leq p$. Denote respectively by $\mathcal{F}_{\mathcal{S},p}$, $\mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$ the sets of p-inefficient, and p-efficient points for ξ. Clearly, these sets are finite since, in each dimension $i \in [n] \stackrel{\text{def}}{=} \{1, \dots, n\}$, we need to consider only the projections $\mathcal{C}_i \stackrel{\text{def}}{=} \{q_i, q_i - 1 \mid q \in \mathcal{S}\} \subseteq \mathbb{Z}$. In other words, the sets $\mathcal{F}_{\mathcal{S},p}$ and $\mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$ can be regarded as subsets of a finite integral box $\mathcal{C} = \mathcal{C}_1 \times \cdots \times \mathcal{C}_n$ of size at most $2|\mathcal{S}|$ along each dimension.

Theorem 4. *Given a partial list $\mathcal{E} \subseteq \mathcal{F}_{\mathcal{S},p}$ of p-inefficient points, problem GEN($\mathcal{F}_{\mathcal{S},p}, \mathcal{E}$) can be solved in $k^{o(\log k)}$ time, where $k \stackrel{\text{def}}{=} \max\{n, |\mathcal{S}|, |\mathcal{E}|\}$.*

Proof. This statement is again a consequence of the fact that the set $\mathcal{F}_{\mathcal{S},p}$ is uniformly dual-bounded, i.e. that

$$\left|\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})\right| \leq |\mathcal{S}||\mathcal{E}|, \tag{6}$$

for any non-empty subset $\mathcal{E} \subseteq \mathcal{F}_{\mathcal{S},p}$. To see (6), let $\mathcal{X} = \{\phi(x) \mid x \in \mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})\}$ and $\mathcal{Y} = \{\phi(y) \mid y \in \mathcal{E}\}$, where $\phi : \mathbb{Z}^n \mapsto \mathbb{R}^{|\mathcal{S}|}$ is the mapping defined by: $\phi(x) = (\Pr[\xi = q] \ : \ q \in \mathcal{S},\ q \leq x)$ for $x \in \mathbb{Z}^n$. One can easily check that the mapping ϕ is one-to-one between $\mathcal{X}$ and $\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$, and that the families $\mathcal{X}$ and $\mathcal{Y}$ satisfy properties (P1) and (P2) with $w = (1, 1, \dots, 1)$ and $t = p$. Therefore, (6) follows from the intersection lemma. □

In particular, all p-inefficient points of a discrete probability distribution can be enumerated incrementally in quasi-polynomial time. In general, a result analogous to that for p-efficient points is highly unlikely to hold, as there exist examples for which the corresponding problem is NP-hard:

Proposition 1. *Given a discrete random variable ξ on a finite support set $\mathcal{S} \subseteq \mathbb{R}^n$, a threshold probability $p \in [0,1]$, and a partial list $\mathcal{E} \subseteq \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$ of p-efficient points for ξ, it is* NP*-complete to decide if $\mathcal{E} \neq \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$.*

Proof. Consider the well-known NP-complete problem of deciding whether a given graph $G = (V, E)$ contains an independent set of size t, where $t \geq 2$ is a given threshold. Let $\mathcal{S} \subseteq \{0,1\}^V$ be the set of points consisting of the $|V|$ incidence vectors of the vertices of G, and $t-2$ copies of the $|E|$ incidence vectors of the edges. Let ξ be an n-dimensional integer-valued random variable having uniform distribution on $\mathcal{S}$, i.e., $\Pr[\xi = q] = 1/|\mathcal{S}|$ if and only if $q \in \mathcal{S}$. Then, for $p = t/|\mathcal{S}|$, the incidence vector of each edge is a p-efficient point for ξ, and it is easy to see that there is another p-efficient point if and only if there is an independent set of G of size t. □

Finally we observe that if ξ is an integer-valued finite random variable with *independent* coordinates $\xi_1, \ldots, \xi_n$, then the generation of both $\mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$ and $\mathcal{F}_{\mathcal{S},p}$ can be done in polynomial time, even if the number of points $\mathcal{S}$, defining the distribution of ξ, is *exponential* in n (but provided that the distribution function for each component ξ_i is computable in polynomial-time). Indeed, by independence we have $\Pr[\xi \leq x] = \prod_{j=1}^n \Pr[\xi_j \leq x_j]$. Defining $f(x) = \log \Pr[\xi \leq x] = \sum_{j=1}^n \log \Pr[\xi_j \leq x_j]$, we can write $f(x)$ as the sum of single-variable monotone functions $f_1, \ldots, f_n$, where $f_i = \log \Pr[\xi_i \leq x_i]$, for $i = 1, \ldots, n$. Let $l_i = \min\{x_i \in \mathbb{Z} \mid \Pr[\xi_i \leq x_i] > 0\} - 1$, $u_i = \min\{x_i \in \mathbb{Z} \mid \Pr[\xi_i \leq x_i] = 1\}$, and $\mathcal{C}_i = \{z \in \mathbb{Z} \mid l_i \leq z \leq u_i\}$. Then the p-inefficient (p-efficient) points are the maximal feasible (respectively, minimal infeasible) solutions of the *monotone separable* inequality $\sum_{i=1}^n f_i(x_i) \leq t \stackrel{\text{def}}{=} \log p$ over the product $\mathcal{C} \stackrel{\text{def}}{=} \mathcal{C}_1 \times \cdots \times \mathcal{C}_n$. Consequently, Theorem 3 immediately gives the following:

Corollary 2. *If the coordinates of a random variable ξ over $\mathbb{Z}^n$ are independent, then both the p-efficient and the p-inefficient points for ξ can be enumerated in incremental polynomial time.*

4 Maximal *k*-Boxes

Let $\mathcal{S}$ be a set of points in $\mathbb{R}^n$, and k be a given integer, $k \leq |\mathcal{S}|$. A *maximal k-box* is an n-dimensional interval which does not contain more than k points of $\mathcal{S}$ in its interior, and which is maximal with respect to this property (i.e. cannot be extended in any direction without strictly enclosing more points of $\mathcal{S}$). Let $\mathcal{F}_{\mathcal{S},k}$ be the set of all maximal k-boxes. The problem of generating all elements of $\mathcal{F}_{\mathcal{S},0}$ has been studied in the machine learning and computational geometry literatures (see [2,8,11,20,21]), and is motivated by the discovery of missing associations or "holes" in data mining applications (see [1,16,17]). All known algorithms that solve this problem have running time complexity which is exponential in the dimension n of the given point set. In contrast, we show in this paper that the problem can be solved in quasi-polynomial time:

Theorem 5. *Given a point set $\mathcal{S} \subseteq \mathbb{R}^n$, an integer k, and a partial list of maximal empty boxes $\mathcal{E} \subseteq \mathcal{F}_{\mathcal{S},k}$, problem GEN$(\mathcal{F}_{\mathcal{S},k}, \mathcal{E})$ can be solved in $m^{o(\log m)}$ time, where $m \stackrel{\text{def}}{=} \max\{n, |\mathcal{S}|, |\mathcal{E}|\}$.*

Proof. Let us define $\mathcal{C}_i \stackrel{\text{def}}{=} \{p_i - \epsilon, p_i, p_i + \epsilon \mid p \in \mathcal{S}\}$ for $i = 1, \ldots, n$, where $\epsilon > 0$ is small enough, and let us consider the family of boxes $\mathcal{B} = \{[a, b] \subseteq \mathbb{R}^n \mid a, b \in \mathcal{C}_1 \times \cdots \times \mathcal{C}_n,\ a \leq b\}$. Then $\mathcal{F}_{\mathcal{S},k} \subseteq \mathcal{B}$, and $\mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},k})$ corresponds to minimal boxes of $\mathcal{B}$ containing at least $k+1$ points of $\mathcal{S}$ in their interior. Then, to prove the theorem it is enough to show that, for any non-empty subset $\emptyset \neq \mathcal{E} \subseteq \mathcal{F}_{\mathcal{S},k}$, we have

$$|\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},k})| \leq |\mathcal{S}||\mathcal{E}|. \tag{7}$$

Let us note first that for $k = 0$ we have $|\mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},0})| = |\mathcal{S}|$, implying (7) readily, thus we assume $k > 0$ in the sequel. Let $u = (u_1, \ldots, u_n)$ where $u_i = \max \mathcal{C}_i$ for $i = 1, \ldots, n$, let $\mathcal{C}_i^* \stackrel{\text{def}}{=} \{u_i - p \mid p \in \mathcal{C}_i\}$ for $i = 1, \ldots, n$, and let us consider the $2n$-dimensional box $\mathcal{C} = \mathcal{C}_1^* \times \cdots \times \mathcal{C}_n^* \times \mathcal{C}_1 \times \cdots \times \mathcal{C}_n$. Let us further represent every n-dimensional interval $[a, b]$ in $\mathcal{F}_{\mathcal{S},k} \cup \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},k})$ as a $2n$-dimensional vector $(u - a, b) \in \mathcal{C}$. It is now easy to see that if $x, y \in \mathcal{C}$ are two boxes, $x \leq y$ (componentwise, as usual), and x defines a box, then indeed y also defines a box which contains x (though not all elements of $\mathcal{C}$ define a box, since $a_i > b_i$ is possible for some $(u - a, b) \in \mathcal{C}$).

Let us now define the anti-monotone property π to be satisfied by an $x \in \mathcal{C}$ if and only if it contains at most k points in its interior, where x contains no point in its interior if it does not define a box. Clearly, $\mathcal{F}_\pi$ for this property and $\mathcal{F}_{\mathcal{S},k}$ differ by at most $\sum_{i=1}^n |\mathcal{C}_i| - 1$ elements, in which $a_i > b_i$ for exactly one of the indices i, and the values a_i and b_i are consecutive in $\mathcal{C}_i$.

Finally, consider the sets $\mathcal{X} = \{\phi(x) \mid x \in \mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},k})\}$ and $\mathcal{Y} = \{\phi(y) \mid y \in \mathcal{E}\}$, where $\phi(x) \in \{0, 1\}^{\mathcal{S}}$ is the characteristic vector of the subset of $\mathcal{S}$ contained in the interior of box $x \in \mathcal{C}$. It is easy to see now that the mapping ϕ is one-to-one between $\mathcal{X}$ and $\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},k})$, and that the sets $\mathcal{X}$ and $\mathcal{Y}$ satisfy properties (P1) and (P2) with $w = (1, 1, \ldots, 1)$ and $t = k$. Thus, inequality (7) follows by applying the intersection lemma. □

5 Proof of the Intersection Lemma

As mentioned in the introduction, we may assume without loss of generality that all the weights are 1's. We can further assume that $\mathcal{Y}$ is a minimal family for properties (P1) and (P2). For $i = 1, \ldots, n$, let $l_i \stackrel{\text{def}}{=} \min\{x_i \mid x \in \mathcal{X}\}$, and $u_i \stackrel{\text{def}}{=} \max\{x_i \mid x \in \mathcal{X}\}$.

To prove the lemma, we shall show by induction on $|\mathcal{X}|$ that

$$|\mathcal{X}| \leq \sum_{y\in\mathcal{Y}} q(y), \tag{8}$$

where $q(y)$ is the number of components y_i such that $y_i < u_i$.

Clearly, for $|\mathcal{X}| \leq 1$ the statement is true since $\mathcal{Y}$ is non-empty and $q(y) = 0$ for $y \in \mathcal{Y}$ implies by (P1) that $\mathcal{X} = \emptyset$. Let us assume therefore that $|\mathcal{X}| \geq 2$, and define for every $i = 1, \ldots, n$ and $z \in \mathbb{R}$ the families

$$\mathcal{X}(i,z) = \{x \in \mathcal{X} \mid x_i \geq z\}, \quad \mathcal{Y}(i,z) = \{y \in \mathcal{Y} \mid y_i \geq z\}.$$

Clearly, these families satisfy conditions (P1) and (P2) and therefore satisfy the conclusion of the lemma whenever $\mathcal{Y}(i,z) \neq \emptyset$. Furthermore, we may assume without loss of generality that $\mathcal{Y}(i,z) = \emptyset$ implies $\mathcal{X}(i,z) = \emptyset$ for all $i \in [n]$ and $z \in \mathbb{R}$. Indeed, by (P2), if $|\mathcal{Y}(i,z)| = 0$ then $|\mathcal{X}(i,z)| \in \{0,1\}$. If there is an $i \in [n]$ and $z \in \mathbb{R}$, such that $\mathcal{X}(i,z) = \{x\}$ and $\mathcal{Y}(i,z) = \emptyset$, then deleting the element x from $\mathcal{X}$ reduces $|\mathcal{X}|$ by 1 and reduces the sum $\sum_{y\in\mathcal{Y}} q(y)$ by at least 1.

Thus, we can assume by induction on the number of elements in $\mathcal{X}$ that

$$|\mathcal{X}(i,z)| \leq \sum_{y\in\mathcal{Y}(i,z)} q(y) \tag{9}$$

whenever

$$|\mathcal{X}(i,z)| < |\mathcal{X}|.$$

Let us now sum up inequalities (9), for all indices $i \in [n]$ and for all values $z > l_i$ (for which $|\mathcal{X}(i,z)| \neq |\mathcal{X}|$), yielding

$$\sum_{i=1}^{n} \int_{z>l_i} |\mathcal{X}(i,z)| dz \leq \sum_{i=1}^{n} \int_{z>l_i} \sum_{y\in\mathcal{Y}(i,z)} q(y) dz. \tag{10}$$

It is easily seen that the left hand side of (10) is equal to

$$L = \sum_{x\in\mathcal{X}} \sum_{i=1}^{n} (x_i - l_i)|\mathcal{X}|,$$

while the right hand side is equal to

$$R = \sum_{y\in\mathcal{Y}} q(y) \sum_{i=1}^{n} (y_i - l_i).$$

Thus, we get by (P1) and (10) that

$$(t - \sum_{i=1}^{n} l_i)|\mathcal{X}| < L \leq R \leq (t - \sum_{i=1}^{n} l_i) \sum_{y\in\mathcal{Y}} q(y). \tag{11}$$

Note that $t - \sum_{i=1}^n l_i > 0$ can be assumed without loss of generality. Indeed, if $t \leq \sum_{i=1}^n l_i$ then for an arbitrary $y \in \mathcal{Y}$ (and $\mathcal{Y} \neq \emptyset$) we have $\sum_{i=1}^n y_i \leq t \leq \sum_{i=1}^n l_i$ by (P1). By the minimality of $\mathcal{Y}$, we must have $y_i \geq l_i$, for all $i = 1, \ldots, n$, implying that $t = \sum_{i=1}^n l_i$. But then we can replace t by $t + \epsilon$, for a sufficiently small $\epsilon > 0$, and still satisfy property (P1). Thus inequality (8) follows from (11). □

Remark 1. Lemma 1 can be generalized as follows. Given two finite sets of points $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$ and an integer $r \geq 2$, such that $\mathcal{X}$ and $\mathcal{Y}$ can be separated by a nonnegative linear function and for any r distinct points $x^1, x^2, \ldots, x^r \in \mathcal{X}$, their componentwise minimum $x^1 \wedge x^2 \wedge \ldots \wedge x^r$ is dominated by some $y \in \mathcal{Y}$ (i.e. $x^1 \wedge x^2 \wedge \ldots \wedge x^r \leq y$), then

$$|\mathcal{X}| \leq n(r-1)|\mathcal{Y}|.$$

6 Proof of Theorems 2 and 3

For $j = 1, 2, \ldots, r$, let $f_j(x) = \sum_{i=1}^n f_{ij}(x_i)$, where $x \in \mathcal{C} = \{x \in \mathbb{Z}^n \mid l_i \leq x_i \leq u_i, i = 1, 2, \ldots, n\}$. For a given real vector $t = (t_1, \ldots, t_r)$, let $\mathcal{F}_t$ be the set of maximal feasible solutions of system (3).

For each $i \in [n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$, let $\Delta_{ij} : \{l_i - 1, l_i, \ldots, u_i\} \to \mathbb{R}$ be the *difference* of f_{ij} defined by

$$\Delta_{ij}(x_i) = \begin{cases} f_{ij}(x_i+1) - f_{ij}(x_i) & \text{if } x_i \in \{l_i, l_i+1, \ldots, u_i - 1\} \\ +\infty & \text{if } x_i \in \{l_i - 1, u_i\}. \end{cases} \tag{12}$$

Let us now define, for each $j \in [r]$, a mapping μ^j from pairs of a vector $x \in \mathcal{C}$ and a component $i \in [n]$ with $x_i > l_i$ to vectors $y \in \mathcal{C}$ by

$$\mu^j(x,i)_k = \begin{cases} x_k - 1 & \text{if } k = i \\ x_k + \alpha_k & \text{otherwise,} \end{cases} \tag{13}$$

where $\alpha_k = \alpha_k(x, k, j)$ is a non-negative integer such that $\Delta_{kj}(x_k + \alpha_k) \geq \Delta_{ij}(x_i - 1)$ and $\Delta_{kj}(x_k + s) < \Delta_{ij}(x_i - 1)$ for all $s = 0, 1, \ldots, \alpha_k - 1$. Note that such α_k always exists by our definition (12).

Given any $x \in \mathcal{I}^{-1}(\mathcal{F}_t)$, there exists an index $j = \rho(x) \in [r]$ such that x violates the jth inequality of the system, i.e. $f_j(x) > t_j$. For $\mathcal{E} \subseteq \mathcal{I}^{-1}(\mathcal{F}_t)$ and $j \in [r]$, let $\rho_{\mathcal{E}}^{-1}(j) \stackrel{\text{def}}{=} \{x \in \mathcal{E} \mid \rho(x) = j\}$.

Proof of Theorem 2 Let us consider an arbitrary non-empty subset $\mathcal{E} \subseteq \mathcal{I}^{-1}(\mathcal{F}_t)$. Consider a vector $y \in \mathcal{I}(\mathcal{E}) \cap \mathcal{F}_t$ and let y_i be a component of y such that $y_i < u_i$ (such a component always exists since $\mathcal{E}$ is non-empty). Then, by the maximality of y, there exists a vector $x = x^i \in \mathcal{E}$ such that $x \leq y + \mathbf{e}^i$, where $\mathbf{e}^i$ is the ith unit vector. Let $j = \rho(x) \in [r]$ be an index such that x violates the jth inequality of the system.

Claim 1. $y \leq \mu^j(x,i)$.

Proof. Let us first note that $x_i = y_i + 1$, since $x_i \leq y_i + 1$ and we have $f_j(x) \leq t_j$ if $x_i \leq y_i$, contradicting the fact that $x \in \mathcal{I}^{-1}(\mathcal{F}_t)$. This means $y_i = \mu^j(x,i)_i$. Moreover, if $x_k < y_k - \alpha_k$ for some $k \neq i$, then we have

$$\begin{aligned} f_j(y) - f_j(x) = \sum_{h \neq i,k} (f_{hj}(y_h) - f_{hj}(x_h)) \\ +(f_{kj}(y_k) - f_{kj}(x_k)) - (f_{ij}(x_i) - f_{ij}(y_i)) \\ \geq \Delta_{kj}(x_k + \alpha_k) - \Delta_{ij}(x_i - 1), \end{aligned} \tag{14}$$

where the last inequality follows from the monotonicity of the functions f_{ij}, and the facts that $x_k \leq y_k$ for all $k \neq i$, $y_i = x_i - 1$, and $y_k \geq x_k + \alpha_k + 1$. Since $\Delta_{kj}(x_k + \alpha_k) - \Delta_{ij}(x_i - 1) \geq 0$ by the definition of $\alpha_k = \alpha_k(x,k,j)$, we get $f_j(y) \geq f_j(x) > t_j$, a contradiction to the fact that $y \in \mathcal{F}_t$. Therefore, $y_k \leq x_k + \alpha_k$ must hold for all components $k \neq i$, proving the calim. □

Claim 2. $y_k = \mu^j(x,i)_k$ for all components $k \in [n]$ for which

$$\Delta_{kj}(y_k) \geq \Delta_{ij}(y_i). \tag{15}$$

Proof. Let $k \neq i$ satisfy (15), then for $s = 0, 1, \ldots, \alpha_j - 1$, we have

$$\Delta_{kj}(y_k) \geq \Delta_{ij}(y_i) = \Delta_{ij}(x_i - 1) > \Delta_{kj}(x_k + s), \tag{16}$$

by definition of $\alpha_k = \alpha_k(x,k,j)$. Since $x_k \leq y_k$, it follows from (16) that $y_k \geq x_k + \alpha_k = \mu^j(x,i)_k$, and therefore the result follows from Claim 1. □

Claims 1 and 2 imply that

$$y = \bigwedge_{i \in [n] \colon y_i < u_i} \mu^j(x^i, i), \tag{17}$$

where for vectors $v, u \in \mathcal{C}$ we let, as before, $v \wedge u$ denote the component-wise minimum of v and u.

Not all of the vectors $\mu^j(x^i, i)$ are necessary for this representation. Suppose that there exist two vectors $x^i, x^k \in \mathcal{E}$ such that $x^i_i > l_i$, $x^k_k > l_k$, $x^i \leq y + \mathbf{e}^i$, $x^k \leq y + \mathbf{e}^k$, and $\rho(x^i) = \rho(x^k) = j$. Suppose further that $\Delta_{kj}(x^k_k - 1) \leq \Delta_{ij}(x^i_i - 1)$. Then Claim 2 implies that (17) remains valid even if we drop x^k. In other words, we can identify, for each $j \in [r]$, a single vector $x^{i_j} \in \rho_{\mathcal{E}}^{-1}(j)$, and obtain consequently at most r vectors $\mu^j(x^{i_j}, i_j)$ such that

$$y = \bigwedge_{j \in [r]} \mu^j(x^{i_j}, i_j), \tag{18}$$

where we have $\mu^j(x^{i_j}, i_j) = u$ if there exists no vector x^i in $\rho_{\mathcal{E}}^{-1}(j)$ The latter representation readily implies (5). □

Proof of Theorem 3 Note that for constant r, the sizes of $\mathcal{F}_t$ and $\mathcal{I}^{-1}(\mathcal{F}_t)$ are polynomially related by inequalities (4) and (5). Hence, the theorem follows from the following lemma which gives an algorithm for generating all minimal true points and/or all maximal false points of a monotone separable system (3), with bounded number of inequalities r, in incremental polynomial time.

For $\mathcal{E} \subseteq \mathcal{C}$, denote by $\mathcal{E}^+ = \{y \in \mathcal{C} \mid y \geq x, \text{ for some } x \in \mathcal{E}\}$ and $\mathcal{E}^- = \{y \in \mathcal{C} \mid y \leq x, \text{ for some } x \in \mathcal{E}\}$.

Lemma 2. *Let $\mathcal{F}_t$ be the set of maximal feasible solutions for (3), and let $\mathcal{Y} \subseteq \mathcal{F}_t$ and $\mathcal{X} \subseteq \mathcal{I}^{-1}(\mathcal{F}_t)$, such that $\mathcal{X} \neq \emptyset$. Then $\mathcal{Y} = \mathcal{F}_t$ and $\mathcal{X} = \mathcal{I}^{-1}(\mathcal{F}_t)$ if and only if*

(i) *For all $x \in \mathcal{X}$ and $i \in [n]$ such that $x_i > l_i$, and for all $k \neq i$ such that $\mu^j(x,i)_k < u_k$, where $j = \rho(x)$, the vector $\overline{x} = \overline{x}(x,i,k)$ given by*

$$\overline{x}_h = \begin{cases} x_h - 1 & \text{if } h = i \\ \mu^j(x,i)_h + 1 & \text{if } h = k \\ x_h & \text{otherwise,} \end{cases} \tag{19}$$

is in $\mathcal{X}^+$.

(ii) *For every collection $(x^j \in \rho_{\mathcal{X}}^{-1}(j) \mid j \in [r])$, and for every selection of indices $(k_1, \ldots, k_r)$ such that $x^j_{k_j} > l_{k_j}$, the vector $y = \wedge_{j\in[r]}\nu^j$ is in $\mathcal{X}^+ \cup \mathcal{Y}^-$, where ν_j is either $\mu^j(x^j, k_j)$ or u.*

Proof. Note that if $x \in \mathcal{X}$, $i, k \in [n]$ and $j \in [r]$ satisfy the conditions specified in (i), and $\overline{x} = \overline{x}(x,i,k)$ is given by (19), then $f_j(\overline{x}) - f_j(x) \geq 0$ follows, implying that both (i) and (ii) are indeed necessary conditions for duality (i.e. for $\mathcal{Y} = \mathcal{F}_t$ and $\mathcal{X} = \mathcal{I}^{-1}(\mathcal{F}_t)$).

To see the sufficiency, suppose that (i) and (ii) hold, and let y be a maximal element in $\mathcal{C} \setminus (\mathcal{X}^+ \cup \mathcal{Y}^-)$. Since $y \neq u$ by assumption, there is an $i \in [n]$ such that $y_i < u_i$. By maximality of y, there exists an $x \in \mathcal{X}$ such that $y + \mathbf{e}^i \geq x$. Let $j = \rho(x)$. If $y_k \geq \mu^j(x,i)_k + 1$, for some $k \neq i$, then $y \geq \overline{x}(x,i,k)$, and hence by (i), $y \in \mathcal{X}^+$, yielding a contradiction. We conclude therefore that $y \leq \mu^j(x,i)$, and consequently, as in the proof of Theorem 2, y is in the form given in (18). But then, by (ii), $y \in \mathcal{X}^+ \cup \mathcal{Y}^-$, another contradiction. □

References

1. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo, Fast discovery of association rules, in *Advances in Knowledge Discovery and Data Mining* (U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, eds.), pp. 307–328, AAAI Press, Menlo Park, California, 1996.
2. M. J. Atallah and G. N. Fredrickson, A note on finding a maximum empty rectangle, *Discrete Applied Mathematics* 13 (1986) 87–91.
3. J. C. Bioch and T. Ibaraki, Complexity of identification and dualization of positive Boolean functions, *Information and Computation* 123 (1995) 50–63.
4. E. Boros, K. Elbassioni, V. Gurvich, L. Khachiyan and K.Makino, Dual-bounded generating problems: All minimal integer solutions for a monotone system of linear inequalities, *SIAM Journal on Computing*, **31** (5) (2002) pp. 1624–1643.

5. E. Boros, K. Elbassioni, V. Gurvich and L. Khachiyan, An inequality for polymatroid functions and its applications, to appear in *Discrete Applied Mathematics*, 2003. (DIMACS Technical Report 2001–14, Rutgers University, ("http://dimacs.rutgers.edu/TechnicalReports/2001.html").
6. E. Boros, V. Gurvich, L. Khachiyan and K. Makino, Dual bounded generating problems: partial and multiple transversals of a hypergraph, *SIAM Journal on Computing* **30** (6) (2001) 2036–2050.
7. E. Boros, V. Gurvich, L. Khachiyan and K. Makino, On the complexity of generating maximal frequent and minimal infrequent sets in binary matrices. In: Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science (STACS 2002). (H. Alt and A. Ferreira, eds., Antibes Juan-les-Pins, France, March 14–16, 2002), Lecture Notes in Computer Science 2285 (2002) pp. 133–141, (Springer Verlag, Berlin, Heidelberg, New York).
8. B. Chazelle, R. L. (Scot) Drysdale III and D. T. Lee, Computing the largest empty rectangle, *SIAM Journal on Computing*, 15(1) (1986) 550–555.
9. Y. Crama, Dualization of regular Boolean functions, *Discrete Applied Mathematics* 16 (1987) 79–85.
10. D. Dentcheva, A. Prékopa and A. Ruszczynski, Concavity and efficient points of discrete distributions in Probabilistic Programming, *Mathematical Programming* **89** (2000) 55–77.
11. J. Edmonds, J. Gryz, D. Liang and R. J. Miller, Mining for empty rectangles in large data sets, in *Proc. 8th Int. Conf. on Database Theory (ICDT)*, Jan. 2001, *Lecture Notes in Computer Science* 1973, pp. 174–188.
12. M. L. Fredman and L. Khachiyan, On the complexity of dualization of monotone disjunctive normal forms, *Journal of Algorithms*, 21 (1996) 618–628.
13. 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.
14. D. Gunopulos, R. Khardon, H. Mannila, and H. Toivonen, Data mining, hypergraph transversals and machine learning, in *Proceedings of the 16th ACM-SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, (1997) pp. 12–15.
15. E. Lawler, J. K. Lenstra and A. H. G. Rinnooy Kan, Generating all maximal independent sets: NP-hardness and polynomial-time algorithms, *SIAM Journal on Computing*, 9 (1980) 558–565.
16. B. Liu, L.-P. Ku and W. Hsu, Discovering interesting holes in data, In *Proc. IJCAI*, pp. 930–935, Nagoya, Japan, 1997.
17. B. Liu, K. Wang, L.-F. Mun and X.-Z. Qi, Using decision tree induction for discovering holes in data, In *Proc. 5th Pacific Rim International Conference on Artificial Intelligence*, pp. 182–193, 1998.
18. K. Makino and T. Ibaraki, Interior and exterior functions of Boolean functions, *Discrete Applied Mathematics*, 69 (1996) 209–231.
19. K. Makino and T. Ibaraki. The maximum latency and identification of positive Boolean functions. *SIAM Journal on Computing*, 26:1363–1383, 1997.
20. A. Namaad, W. L. Hsu and D. T. Lee, On the maximum empty rectangle problem. *Discrete Applied Mathematics*, **8**(1984) 267–277.
21. M. Orlowski, A new algorithm for the large empty rectangle problem, *Algorithmica* 5(1) (1990) 65–73.
22. A. Prékopa, *Stochastic Programming*, (Kluwer, Dordrecht, 1995).
23. R. C. Read and R. E. Tarjan, Bounds on backtrack algorithms for listing cycles, paths, and spanning trees, *Networks* 5 (1975) 237–252.

24. R. Srikant and R. Agrawal, Mining generalized association rules. In *Proc. 21st International Conference on Very Large Data Bases*, pp. 407–419, 1995.
25. R. Srikant and R. Agrawal, Mining quantitative association rules in large relational tables. In *Proc. of the ACM-SIGMOD 1996 Conference on Management of Data*, pp. 1–12, 1996.