双対制限された列挙問題：離散分布に対する交差不等式とその応用

Endre Boros[†], Khaled Elbassioni[†], Vladimir Gurvich[†],
Leonid Khachiyan[‡], 牧野和久[§]

[†]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854.
{boros,elbassio,gurvich}@rutcor.rutgers.edu
[‡]Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey, 08854-8004. leonid@cs.rutgers.edu
[§]560-8531 大阪府豊中市待兼山町 1-3 大阪大学大学院基礎工学研究科社会システム数理領域
makino@sys.es.osaka-u.ac.jp

あらまし 2つの有限集合 $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$ が非負の線形関数で分離可能であり，かつ，$\mathcal{X}$ 中の任意の相異なる2つベクトルの成分ごとに最小値をとって得られるベクトルが $\mathcal{Y}$ 中のあるベクトル以下であるとき，我々は，$|\mathcal{X}| \le n|\mathcal{Y}|$ が成立することを示す．この結果として，与えられた分離可能な不等式からなる単調システムの極大整数解列挙，与えられた離散確率分布に対する p-非効率点の列挙，$\mathbb{R}^n$ 中に与えられた点集合を一定割合以下で含む超直方体列挙に対する擬多項式時間アルゴリズムを得る．これは，整数計画，確率計画，データマイニングの分野に関連するこれらの問題に対して過去に提案された指数時間アルゴリズムを本質的に改良する．さらに，我々は，定数個の分離可能な不等式をもつ単調システムに対する逐次多項式時間列挙アルゴリズムを与える．これは，成分ごとに独立した離散確率分布における p-効率点，p-非効率点がそれぞれ逐次多項式時間で列挙できることを意味する．

Dual-Bounded Generating Problems: An Intersection Inequality for Discrete Distributions and Its Applications

Endre Boros[†], Khaled Elbassioni[†], Vladimir Gurvich[†],
Leonid Khachiyan[‡] and Kazuhisa Makino[§]

[†]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854.
{boros,elbassio,gurvich}@rutcor.rutgers.edu
[‡]Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey, 08854-8004. leonid@cs.rutgers.edu
[§]Division of Mathematical Science for Social Systems, 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 hyper-rectangles which contain a specified fraction of points of a given set 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 $t \in \mathbb{R}$ is a real threshold, and $w(x) = \sum_{i=1}^n w_i x_i$, for some nonnegative weights $w_1, \ldots, w_n \in \mathbb{R}_+$.

(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., $x \wedge x' \le y$.

Given $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$ satisfying properties (P1) and (P2), one may ask 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 actually is 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 [7]. Let us also recall from [7] 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 $w_i > 0$ 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 with respect to each coordinate of the points, and scales properly their componentwise minimum, so that the transformed point sets will satisfy (P2) as well.

We sketch Lemma 1 in Section 5. As a consequence of the 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 [9, 19].

Maximal k-boxes: Given a set of points in $\mathbb{R}^n$ and a nonnegative integer k, generate all maximal n-dimensional intervals (*boxes*), each of which contains at most k of the given points in its 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, 10, 15, 16, 17, 18].

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 that of incrementally generating the family $\mathcal{F}_\pi$ of all *maximal* elements of $\mathcal{C}$ satisfying 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 $a \in \mathcal{A}$. Let $\mathcal{I}^{-1}(\mathcal{A})$ be 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 [7], let us call $\mathcal{F}_\pi$ *uniformly dual-bounded*, if for every non-empty 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 [5] and also [4, 13, 14]):

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 [5, 12]). However, it is still open whether DUAL($\mathcal{C}, \mathcal{A}, \mathcal{B}$) has a polynomial time algorithm (e.g., [4, 11, 12]).

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 $\mathcal{E} \subseteq \mathcal{F}_\pi$ can be related to two sets of points $\mathcal{X}, \mathcal{Y}$ satisfying the conditions of Lemma 1. Then the Lemma will imply (2), which in its turn is sufficient for the efficient generation of the family $\mathcal{F}_\pi$ (see [5]).

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

In the last section, we consider some generalizations of the intersection lemma. Namely, we show that an analogous lemma holds for families of vectors in the product of arbitrary meet semi-lattices. As an application, we obtain quasi-polynomial time algorithms for generating maximal feasible solutions for systems of monotone inequalities on sums of separable functions with bounded number of variables, and for generating maximal k-boxes whose diameter does not exceed a given threshold, for a given set of points.

Due to the space limitation, we skip most proofs of the results, which can be found in [6].

2 Systems of Monotone Separable Inequalities

For $i = 1, 2, \ldots, n$, let l_i and u_i be given integers with $l_i \leq u_i$, and let $\mathcal{C}_i \overset{\text{def}}{=} \{l_i, l_i + 1, \ldots, u_i\}$. 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 polynomial-time 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 all maximal feasible solutions for (3). Then $\mathcal{I}^{-1}(\mathcal{F}_t)$ represents the set of all minimal infeasible vectors for (3).

Generalizing results on monotone systems of *linear* inequalities from [5], we will now use Lemma 1 to prove the following:

Theorem 1 *If $\mathcal{F}_t$ is the family of all 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, $|\mathcal{I}^{-1}(\mathcal{F}_t)| \leq rn|\mathcal{F}_t|$.

Since by (4) the family $\mathcal{F}_t$ is uniformly dual-bounded, the results of [5], 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 $k = \max\{n, r, |\mathcal{X}|\}$ and $poly(k) \log(\|u - l\|_\infty + 1)$ feasibility tests for (3).*

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 (see e.g. [5]). 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*

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

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

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 time, polynomial in n, r and $\log(\|u - l\|_\infty + 1)$.*

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 $\mathcal{F}_{\mathcal{S},p}$ and $\mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$ the sets of all p-inefficient and p-efficient points for ξ. Clearly, these sets are finite since, in each dimension $i \in [n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$, 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 $k \stackrel{\text{def}}{=} \max\{n, |\mathcal{S}|, |\mathcal{E}|\}$.*

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, since the 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 $\mathcal{E} \neq \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},p})$.*

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 $\mathcal{C}_i = \{z \in \mathbb{Z} \mid l_i \leq z \leq u_i\}$, where we regard $\log 0$ as $-\infty$. 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 $k \leq |\mathcal{S}|$ be a given integer. A *maximal k-box* is a closed n-dimensional interval which contains at most 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, 10, 17, 18]), and is motivated by the discovery of missing associations or "holes" in data mining applications (see [1, 15, 16]). All known algorithms that solve this problem have running time complexity 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 $m \stackrel{\text{def}}{=} \max\{n, |\mathcal{S}|, |\mathcal{E}|\}$.*

Theorem 5 should be contrasted with the following negative result:

Proposition 2 *Given a set of points $\mathcal{S} \subseteq \mathbb{R}^n$, an integer $k \leq |\mathcal{S}|$, and a subfamily $\mathcal{X} \subseteq \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},k})$ of minimal boxes each of which contains at least k points of $\mathcal{S}$ in its interior, it is NP-complete to decide if $\mathcal{X} \neq \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},k})$.*

5 Sketch 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 $|\mathcal{X}| \geq 1$ and that $\mathcal{Y}$ is an *inclusion-wise* minimal family, each vector of which is a *component-wise* minimal 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 $|\mathcal{X}| \leq \sum_{y \in \mathcal{Y}} q(y)$, where $q(y)$ is the number of components y_i such that $y_i < u_i$.

For $|\mathcal{X}| = 1$ 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 $\mathcal{X}(i,z) = \{x \in \mathcal{X} \mid x_i \geq z\}$, $\mathcal{Y}(i,z) = \{y \in \mathcal{Y} \mid y_i \geq z\}$. Clearly, these families satisfy conditions (P1) and (P2). 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 $|\mathcal{X}(i,z)| \leq \sum_{y \in \mathcal{Y}(i,z)} q(y)$ whenever $|\mathcal{X}(i,z)| < |\mathcal{X}|$. Since the latter condition is satisfied for $z > l_i$, we can sum up the inequalities, for all values $z > l_i$, and for all indices $i \in [n]$, to obtain $\sum_{i=1}^n \int_{z>l_i} |\mathcal{X}(i,z)| dz \leq \sum_{i=1}^n \int_{z>l_i} \sum_{y \in \mathcal{Y}(i,z)} q(y) dz$. It is easily seen that the left hand side is equal to $L = \sum_{x \in \mathcal{X}} \sum_{i=1}^n (x_i - l_i)$, 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) that $(t - \sum_{i=1}^n l_i)|\mathcal{X}| < L \leq R \leq (t - \sum_{i=1}^n l_i) \sum_{y \in \mathcal{Y}} q(y)$. Since $t - \sum_{i=1}^n l_i > 0$ can be assumed, the proof is completed. □

6 Generalizations

In this section, we give some generalizations of the intersection lemma and discuss some further applications.

6.1 Intersection Lemma for Meet Semi-lattices

Let $\mathcal{P}_i$, $i = 1, \ldots, n$ be given finite partial orders such that for any index i and any two elements $x, y \in \mathcal{P}_i$, elements x and y have a *unique minimum*, i.e., the *meet* $x \wedge y \stackrel{\text{def}}{=} \min(x,y) \in \mathcal{P}_i$ exists and is well defined. Denote by "$\preceq$" the precedence relation on $\mathcal{P}$, and for $\mathcal{E} \subseteq \mathcal{P}$, let $\mathcal{E}^+ = \{y \in \mathcal{P} \mid y \succeq x \text{ for some } x \in \mathcal{E}\}$ and $\mathcal{E}^- = \{y \in \mathcal{P} \mid y \preceq x \text{ for some } x \in \mathcal{E}\}$. For simplicity, we write x^+ and x^- instead of $\{x\}^+$ and $\{x\}^-$, respectively. For $i \in [n]$ and $x \in \mathcal{P}_i$, define

$$q_i(x) = |\{z \in \mathcal{P}_i \ : \ z \notin x^- \text{ and } z \text{ has an immediate predecessor } z' \preceq x\}|,$$

and let $q(y) \stackrel{\text{def}}{=} \sum_{i=1}^n q_i(y_i)$ for $y \in \mathcal{P} \stackrel{\text{def}}{=} \mathcal{P}_1 \times \cdots \times \mathcal{P}_n$.

Lemma 2 *Let $\mathcal{P}_i$, $i = 1, \ldots, n$, be given finite meet-semi lattices, let $w : \cup_{i=1}^n \mathcal{P}_i \mapsto \mathbb{R}_+$ be a function assigning a non-negative weight to each element in $\cup_{i=1}^n \mathcal{P}_i$, and let $t \in \mathbb{R}_+$ be a given positive threshold. Assume that $\mathcal{X}$ and $\mathcal{Y} \neq \emptyset$ are subsets of $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_n$ such that*

(i) for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ we have $\sum_{i=1}^n w(x_i^-) > t \geq \sum_{i=1}^n w(y_i^-)$, where $w(\mathcal{Q}) \stackrel{\text{def}}{=} \sum_{z \in \mathcal{Q}} w(z)$, for $\mathcal{Q} \subseteq \mathcal{P}_i$ and $i \in [n]$;

(ii) For every $x' \neq x'' \in \mathcal{X}$ there exists a $y \in \mathcal{Y}$ such that $y \succeq x' \wedge x''$.

Then we have

$$|\mathcal{X}| \leq \sum_{y \in \mathcal{Y}} q(y). \tag{6}$$

In particular, $|\mathcal{X}| \leq (\sum_{i=1}^n |\mathcal{P}_i| - n)|\mathcal{Y}|$.

Note that the bound of Lemma 2 is best possible, and Lemma 1 can be derived as a special case of Lemma 2.

6.2 r-Intersection Lemma

Lemma 2 can be further generalized as follows. Given two finite sets of elements $\mathcal{X}$ and $\mathcal{Y}$ in the product $\mathcal{P} \stackrel{\text{def}}{=} \mathcal{P}_1 \times \cdots \times \mathcal{P}_n$ of n meet semi-lattices, and an integer $r \geq 2$, consider the following property :

(ii′) For any r distinct elements $x^1, x^2, \ldots, x^r \in \mathcal{X}$, their componentwise meet $x^1 \wedge x^2 \wedge \ldots \wedge x^r$ is dominated by some $y \in \mathcal{Y}$, i.e., $x^1 \wedge x^2 \wedge \ldots \wedge x^r \preceq y$.

Lemma 3 *If $\mathcal{X}$ and $\mathcal{Y} \neq \emptyset$ are two finite sets of points in $\mathcal{P}$ satisfying properties (i) of Lemma 2 and (ii′) above, then*

$$|\mathcal{X}| \leq (r-1) \sum_{y \in \mathcal{Y}} q(y).$$

6.3 Systems of Monotone Inequalities on Sums of Separable Functions with Bounded Number of Variables

We can generalize Theorem 1 as follows. Let $\mathcal{H}_1, \ldots, \mathcal{H}_r \subseteq 2^{[n]}$ be r multi-hypergraphs on n vertices, and let $\mathcal{C} = \mathcal{C}_1 \times \cdots \times \mathcal{C}_n = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$, where $l, u \in \mathbb{R}^n$. For $j = 1, \ldots, r$, $H \in \mathcal{H}_j$, and $i \in H$, let $f_{H,i,j} : \mathcal{C}_i \mapsto \mathbb{R}_+$ be a single-variable (polynomial-time computable) monotone function. Consider a system of r inequalities

$$\sum_{H \in \mathcal{H}_j} \prod_{i \in H} f_{H,i,j}(x_i) \leq t_j, \quad j = 1, \ldots, r, \tag{7}$$

over $x \in \mathcal{C}$, where $t_1, \ldots, t_r$ are given real thresholds. For a hypergraph $\mathcal{H}$, define $\dim(\mathcal{H}) = \max\{|H| : H \in \mathcal{H}\}$. Function $f(x_1, x_2, x_3) = x_1^3 x_2 + 2x_1x_2 + x_3^5$ is an example with $r = 1$ and $\dim(\mathcal{H}) = 2$.

Theorem 6 *If $\dim(\mathcal{H}_j) \leq$ const for all $j = 1, \ldots, r$, then all maximal feasible solutions of a system (7) can be generated in incremental quasi-polynomial time.*

Theorem 6 is a consequence of the following.

Theorem 7 *Let $\mathcal{H}_1, \ldots, \mathcal{H}_r \subseteq 2^{[n]}$ be r multi-hypergraphs on n vertices. For $j = 1, \ldots, r$, $H \in \mathcal{H}_j$, and $i \in H$, let $f_{H,i,j} : \mathcal{C}_i \mapsto \mathbb{R}_+$ be a single-variable monotone function. If $\mathcal{F}$ is the family of all maximal feasible solutions of (7), and $\mathcal{E} \subseteq \mathcal{F}$ is non-empty, then*

$$\left|\mathcal{I}^{-1}(\mathcal{E}) \cap \mathcal{I}^{-1}(\mathcal{F})\right| \leq \left(\sum_{j=1}^{r} \sum_{H \in \mathcal{H}_j} |H|(2|\mathcal{E}|+1)^{|H|-1} \right) |\mathcal{E}|.$$

In particular, $\left|\mathcal{I}^{-1}(\mathcal{Y}) \cap \mathcal{I}^{-1}(\mathcal{F})\right| \leq d(\sum_{j=1}^{r} |\mathcal{H}_j|)(2|\mathcal{Y}|+1)^{d-1}|\mathcal{Y}|$, *where* $d = \max\{\dim(\mathcal{H}_1), \ldots, \dim(\mathcal{H}_r)\}$.

On the negative side, we have the following proposition.

Proposition 3 *Given a hypergraph $\mathcal{H} \subseteq 2^{[n]}$ and an integer threshold t, incrementally generating all minimal infeasible vectors for the inequality $f(x) = \sum_{H \in \mathcal{H}} \prod_{i \in H} x_i \leq t$ over $x \in \{0,1\}^{[n]}$ is NP-hard, even if* $\dim(\mathcal{H}) = 2$.

6.4 Maximal Packings/Coverings of Points into/by Boxes

Let $\mathcal{S}$ be a set of points in $\mathbb{R}^n$. Let $C : \mathcal{S} \mapsto \{1, 2, \ldots, r\}$ and $w : \mathcal{S} \mapsto \mathbb{R}_+$ be respectively a coloring and a weighting of the point set $\mathcal{S}$, i.e., mappings that assign respectively one of r colors and a non-negative real weight to each point in $\mathcal{S}$. Given a non-negative threshold vector $t = (t_1, \ldots, t_r) \in \mathbb{R}_+^r$, let us define a *packing* of the point set $\mathcal{S}$, with respect to (w, C, t), to be a box containing (in its interior) a subset of $\mathcal{S}_i \stackrel{\text{def}}{=} \{p \in \mathcal{S} \mid C(p) = i\}$ of total weight at most t_i for all $i = 1, \ldots, r$. Let us define conversely a (C, w, t)-*covering* of $\mathcal{S}$, to be any box that contains a subset of $\mathcal{S}_i$ of total weight greater than t_i for some $i = 1, \ldots, r$. Denote respectively by $\mathcal{F}_{\mathcal{S},C,w,t}$ and $\mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},C,w,t})$ the families of all maximal packings and all minimal coverings of the point set $\mathcal{S}$ with respect to (w, C, t). Clearly, if $r = 1$, $t = k$, and all weights are ones, then $\mathcal{F}_{\mathcal{S},C,w,t}$ is just the family of maximal k-boxes discussed in Section 4. Therefore, Theorem 5 is a special case of the following.

Theorem 8 *All maximal packings of a given point set $\mathcal{S} \subseteq \mathbb{R}^n$, with respect to a given coloring $C : \mathcal{S} \mapsto \{1, 2, \ldots, r\}$, a non-negative weight $w : \mathcal{S} \mapsto \mathbb{R}_+$, and a given threshold vector $t \in \mathbb{R}^r_+$, can be generated incrementally in $k^{o(\log k)}$ time, where $k \stackrel{\text{def}}{=} \max\{n, |\mathcal{S}|, |\mathcal{X}|\}$.*

This follows again from a generalization of the dual-bounding inequality, which can be proved using the intersection lemma:

Theorem 9 *Let $\mathcal{S}$ be a given set of points in $\mathbb{R}^n$, $C : \mathcal{S} \mapsto \{1, 2, \ldots, r\}$ and $w : \mathcal{S} \mapsto \mathbb{R}_+$ be respectively a coloring and a weighting of $\mathcal{S}$, and $t \in \mathbb{R}^r_+$ be a given non-negative real-threshold. If $\mathcal{F} = \mathcal{F}_{\mathcal{S},C,w,t}$ is the set of packings of the point set $\mathcal{S}$, with respect to (C, w, t), then*

$$|\mathcal{I}^{-1}(\mathcal{Y}) \cap \mathcal{I}^{-1}(\mathcal{F})| \le \sum_{i=1}^{r} \sum_{y \in \mathcal{Y}} |\{p \in \mathcal{S}_i \mid \text{point } p \notin \text{the interior of box } y\}|, \tag{8}$$

for any $\emptyset \neq \mathcal{Y} \subseteq \mathcal{F}$, where $\mathcal{S}_i = \{p \in \mathcal{S} \mid C(p) = i\}$. In particular, $|\mathcal{I}^{-1}(\mathcal{F})| \le r|\mathcal{S}||\mathcal{F}|$.

6.5 Maximal Packings with Certain Geometric Properties

We conclude with one more application of Lemma 2. Let $\mathcal{S}$ be a set of points in $\mathbb{R}^n$. For $i = 1, \ldots, n$, consider the set of projection points $\mathcal{S}_i \stackrel{\text{def}}{=} \{p_i \in \mathbb{R} \mid p \in \mathcal{S}\}$, and let $\mathcal{L}_i$ be the *lattice of intervals* whose elements are the different intervals defined by the projection points $\mathcal{S}_i$, and ordered by containment "$\succeq$". The meet of any two intervals in $\mathcal{L}_i$ is their *intersection*, and the join is their *span*, i.e., the minimum interval containing both of them. The minimum element l_i of $\mathcal{L}_i$ is the empty interval. Let $\mathcal{L} = \mathcal{L}_1 \times \cdots \times \mathcal{L}_n$, and for a box $x \in \mathcal{L}$, and $i \in [n]$, denote by $|x_i|$ the length of the interval x_i. Let $f_{ij} : \mathbb{R}_+ \mapsto \mathbb{R}_+$, $i = 1, 2, \ldots, n$, $j = 1, \ldots, r$ be monotone *supermodular* functions, i.e., $f_{ij}(x) \ge f_{ij}(y)$ for $x \succeq y$, and

$$f_{ij}(x \vee y) + f_{ij}(x \wedge y) \ge f_{ij}(x) + f_{ij}(y) \tag{9}$$

for all $x, y \in \mathcal{L}_i$. Let us also say that $f : \mathcal{L}_i \mapsto \mathbb{R}_+$ is *locally supermodular* if (9) is satisfied for all $x, y \in \mathcal{L}_i$ for which $x \vee y$ is an immediate successor of x, y. It is not hard to see that local supermodularity is a sufficient condition for the supermodularity of a monotone function on the lattice $\mathcal{L}_i$.

Consider the system of inequalities

$$\sum_{i=1}^{n} f_{ij}(|x_i|) \le t_j, \quad j = 1, \ldots, r, \tag{10}$$

over the set of n-dimensional boxes $x \in \mathcal{L}$, where $t = (t_1, \ldots, t_r)$ is a given nonnegative r-dimensional real vector. Let us denote by $\mathcal{F}_{\mathcal{S},t}$ the set of all maximal feasible solutions for (10).

Theorem 10 *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a given point set, $f_{ij} : \mathbb{R}_+ \mapsto \mathbb{R}_+$, $i = 1, 2, \ldots, n$, $j = 1, \ldots, r$ be monotone* supermodular *functions, and $t \in \mathbb{R}^r_+$ be a given threshold vector. Then for any non-empty subset $\mathcal{Y}$ of the maximal feasible solutions $\mathcal{F}_{\mathcal{S},t}$ of (10), we have*

$$|\mathcal{I}^{-1}(\mathcal{Y}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},t})| \le rn|\mathcal{S}||\mathcal{Y}|. \tag{11}$$

Corollary 3 *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a given point set, $f_{ij} : \mathbb{R}_+ \mapsto \mathbb{R}_+$, $i = 1, 2, \ldots, n$, $j = 1, \ldots, r$ be monotone* convex *functions, and $t \in \mathbb{R}^r_+$ be a given threshold vector. Then for any non-empty subset $\mathcal{Y}$ of the maximal feasible solutions $\mathcal{F}_{\mathcal{S},t}$ of the system*

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

we have

$$|\mathcal{I}^{-1}(\mathcal{Y}) \cap \mathcal{I}^{-1}(\mathcal{F}_{\mathcal{S},t})| \le rn|\mathcal{S}||\mathcal{Y}|. \tag{12}$$

Finally, we mention two applications of Corollary 3:

(i) Given a set of points $\mathcal{S} \subseteq \mathbb{R}^n$, a coloring $C : \mathcal{S} \mapsto \{1, 2, \ldots, r\}$, a weighting $w : \mathcal{S} \mapsto \mathbb{R}_+$, and a non-negative real threshold $t \in \mathbb{R}^r_+$, generate all maximal (w, C, t)-packings of $\mathcal{S}$ with diameter not exceeding a given threshold $\delta \geq 0$. If $x \in \mathcal{L}$ is such a packing, then it must further satisfy the inequality $(\sum_{i=1}^n |x_i|^p)^{1/p} \leq \delta$ which is in the form covered by Corollary 3 for any finite $p \geq 1$.

(ii) Given n sets $\mathbb{P}_1, \cdots, \mathbb{P}_n \subseteq \mathbb{R}$, and a nonnegative real threshold δ, generate all minimal boxes $[a, b] \in \mathcal{L}$ with $\{a_i, b_i\} \subseteq \mathbb{P}_i$, for $i = 1, \ldots, n$, and with volume at least δ. In fact, these boxes are the minimal feasible solutions of the inequality $\sum_{i=1}^n \log |x_i| \geq \log \delta$, over the lattice $\mathcal{L}$. If $\mathcal{F}$ is the family of all minimal feasible solutions to this inequality, then, as was done in Theorem 10 and Corollary 3, one can use Lemma 2 to prove that $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq \sum_{i=1}^n |\mathbb{P}_i||\mathcal{X}|$. for any non-empty subset $\mathcal{X} \subseteq \mathcal{F}$. Thus all minimal boxes with volume at least δ can be generated in quasi-polynomial time.

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] L. Budach, B. Graw, C. Meinel and S. Waack , *Algebraic and topological properties of finite partially ordered sets*, Teubner-texte zur Mathematik, Band 109, 1988.

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

[5] 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 J. Comput.*, 31 (2002) 1624-1643.

[6] E. Boros, K. Elbassioni, V. Gurvich, L. Khachiyan and K.Makino, An intersection inequality for discrete distributions and related generation problems, Rutcor Research Report, RRR-17-2003, Rutgers University, 2003.

[7] 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.

[8] B. Chazelle, R. L. (Scot) Drysdale III and D. T. Lee, Computing the largest empty rectangle, *SIAM J. Comput.*, 15 (1986) 550-555.

[9] D. Dentcheva, A. Prékopa and A. Ruszczynski, Concavity and efficient points of discrete distributions in Probabilistic Programming, *Mathematical Programming* 89 (2000) 55–77.

[10] 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.

[11] T. Eiter, G. Gottlob, and K. Makino. New results on monotone dualization and generating hypergraph transversals. *SIAM J. Comput.*, 32 (2003) 514–537. Preliminary paper in Proc. ACM STOC 2002.

[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] B. Liu, L.-P. Ku and W. Hsu, Discovering interesting holes in data, In *Proc. IJCAI*, pp. 930–935, Nagoya, Japan, 1997.

[16] 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.

[17] A. Namaad, W. L. Hsu and D. T. Lee, On the maximum empty rectangle problem. *Discrete Applied Mathematics*, 8 (1984) 267-277.

[18] M. Orlowski, A new algorithm for the large empty rectangle problem, *Algorithmica* 5 (1990) 65-73.

[19] A. Prékopa, *Stochastic Programming*, (Kluwer, Dordrecht, 1995).