The Taming of the Semi-Linear Set*

Dmitry Chistikov[†1] **and Christoph Haase**[‡2]

1 **Max Planck Institute for Software Systems (MPI-SWS), Kaiserslautern and Saarbrücken, Germany**
dch@mpi-sws.org

2 **LSV, CNRS & ENS Cachan, Université Paris-Saclay, France**
haase@lsv.ens-cachan.fr

Abstract

Semi-linear sets, which are rational subsets of the monoid $(\mathbb{Z}^d, +)$, have numerous applications in theoretical computer science. Although semi-linear sets are usually given implicitly, by formulas in Presburger arithmetic or by other means, the effect of Boolean operations on semi-linear sets in terms of the size of description has primarily been studied for explicit representations. In this paper, we develop a framework suitable for implicitly presented semi-linear sets, in which the size of a semi-linear set is characterized by its norm – the maximal magnitude of a generator.

We put together a toolbox of operations and decompositions for semi-linear sets which gives bounds in terms of the norm (as opposed to just the bit-size of the description), a unified presentation, and simplified proofs. This toolbox, in particular, provides exponentially better bounds for the complement and set-theoretic difference. We also obtain bounds on unambiguous decompositions and, as an application of the toolbox, settle the complexity of the equivalence problem for exponent-sensitive commutative grammars.

1998 ACM Subject Classification G.2 Discrete Mathematics

Keywords and phrases semi-linear sets, convex polyhedra, triangulations, integer linear programming, commutative grammars

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.128

1 Introduction

Semi-linear sets [20] are a generalisation of ultimately periodic sets of natural numbers to any dimension d. By a classic result due to Ginsburg and Spanier [6], they coincide with the sets of integers[1] definable in Presburger arithmetic (the first-order theory of the integers with addition and order), and hence enjoy closure under all Boolean operations. Their nice properties make them a versatile tool in many application domains such as formal language theory, automata theory, and database theory.

More formally, semi-linear sets are finitely represented finite and infinite subsets of $\mathbb{Z}^d$. For $d \geq 1$, a *semi-linear set* M in dimension d is a finite union of *linear sets*. The latter are presented as a base vector $\boldsymbol{b} \in \mathbb{Z}^d$ and a finite set of period vectors $P = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\} \subseteq \mathbb{Z}^d$ and have the form

$$L(\boldsymbol{b}, P) := \boldsymbol{b} + \{\lambda_1 \cdot \boldsymbol{p}_1 + \cdots + \lambda_n \cdot \boldsymbol{p}_n : \lambda_1, \ldots, \lambda_n \in \mathbb{N}\}. \tag{1}$$

[*] This research was sponsored in part by the ERC Synergy award ImPACT.

[†] Present address: Department of Computer Science, University of Oxford, UK.

[‡] Supported by Labex Digicosme, Univ. Paris-Saclay, project VERICONISS.

[1] In the literature, semi-linear sets are often defined as subsets of $\mathbb{N}^d$ instead of $\mathbb{Z}^d$ as in this paper. All of our results do, however, carry over if one wishes to restrict semi-linear sets to $\mathbb{N}^d$.

43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016).
Editors: Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi;
Article No. 128; pp. 128:1–128:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Such representations are, in fact, only rarely encountered in applications, because in many contexts semi-linear sets are defined implicitly. A semi-linear set can, for instance, be succinctly encoded by a formula in Presburger arithmetic; or a set can be just *proved* to be semi-linear with an estimation of its *norm*, $\|M\|$. The norm is the absolute value of the largest number occurring in the smallest description of M as a union of sets of the form (1). Examples of implicitly presented semi-linear sets include languages of various types of commutative grammars [10, 18] and reachability sets of reversal-bounded counter automata [13, 9].

The effect of Boolean operations is, however, not easy to track in terms of the size of vectors $\boldsymbol{b}$ and $\boldsymbol{p}_i$ if semi-linear sets are only presented implicitly. As an example, consider the set of non-negative integer solutions to a system of linear inequalities $\mathfrak{S}\colon A \cdot \boldsymbol{x} \leq \boldsymbol{c}$, which is a semi-linear set $S \subseteq \mathbb{N}^d$ encoded by $\mathfrak{S}$ with exponential succinctness. Huynh [12, 11] shows that for a given semi-linear set M, in general, whenever the complement $\overline{M} := \mathbb{N}^d \setminus M$ of M (with respect to $\mathbb{N}^d$; the same holds for $\mathbb{Z}^d$) is non-empty, then there is some $\boldsymbol{u} \in \overline{M}$ whose entries are bounded by an exponential in the explicit representation of M – which amounts to doubly exponential in the size of description of $\mathfrak{S}$. This upper bound is far from optimal: by Farkas' lemma, $\overline{M}$ contains an element $\boldsymbol{u}$ whose magnitude $\|\boldsymbol{u}\|$ is at most singly-exponential in the size of description of $\mathfrak{S}$.

Somewhat surprisingly, to the best of the authors' knowledge, there has been no unified framework for deriving bounds of this kind for implicitly presented semi-linear sets. Even if we take an explicitly given linear set as in (1) and describe it by an existential formula $\Psi(\boldsymbol{x})$ in Presburger arithmetic, the representation of the complement with a universally quantified formula $\neg\Psi(\boldsymbol{x})$ provides poor estimates on the magnitude of small elements: although upper bounds can be derived from an analysis of quantifier-elimination procedures, these bounds are only doubly exponential (see, e.g., [25]) and hence far from being optimal.

Our contribution

In this paper, we develop a framework suitable for implicitly presented semi-linear sets (explicitly presented sets are, of course, included as the simplest special case). In this framework the size of a semi-linear set $M \subseteq \mathbb{Z}^d$ is characterised by its norm, $\|M\|$, rather than the full bit-size of the description of M. We prove novel upper bounds in which, as a rule of thumb, the norm of the result of an operation is upper-bounded by $\|M\|^E$ where the exponent E behaves in a "controlled" way (say, $E = \mathsf{poly}(d)$), thus *taming* the effect of Boolean operations and decompositions. In more detail, our contributions are as follows:

- We put together a "toolbox" of operations and decompositions for semi-linear sets, with *tame* bounds, unified presentation, and simplified proofs. This toolbox includes improved bounds on the norm of the complement and, as a corollary, improved bounds on the norm of the set-theoretic difference. These bounds can give an exponential advantage over previously known techniques that upper-bound the bit-size of the result by n^E where n is the bit-size of the description of M – because n can be exponential in $\|M\|$.
- We derive from our toolbox an alternative proof of the $\mathsf{\Pi}_2^{\mathsf{P}}$ upper bound for deciding semi-linear set inclusion, shown originally by Huynh [12, 11]. As a further application, we settle the complexity of the equivalence problem for exponent-sensitive commutative grammars, which have recently been introduced by Mayr and Weihmann [18].
- We give a new proof of and provide an explicit upper bound on unambiguous decomposition of semi-linear sets. It was first asked by Ginsburg [5] whether any semi-linear set is equivalent to a semi-linear set in which every element is generated in a unique way by exactly one linear set. This question was independently positively answered by Eilenberg

and Schützenberger [3] and by Ito [14]. However, to the best of our knowledge, no bounds on this decomposition have been established so far.

We now give a brief guide to the developed techniques and to the remainder of the paper. Our starting point is the fact that the set of non-negative solutions of a system of inequalities $\mathfrak{S}$ can be obtained as $L(B,P) := \bigcup_{\boldsymbol{b} \in B} L(\boldsymbol{b}, P)$ for some finite sets $B, P \subseteq \mathbb{N}^d$. We call semi-linear sets of the form $L(B,P)$ *hybrid linear sets* and use them, instead of linear sets, as basic building blocks for general semi-linear sets. A hybrid linear set preserves more structural information about the "infinitary behaviour" of the linear sets it contains; it is, in fact, a discrete analogue of the Minkowski–Weyl representation of a convex polyhedron as the sum of a polytope and a convex cone.

Since the effect of operations on linear sets is primarily dominated by the magnitude and number of period vectors, reasoning in terms of hybrid linear sets lets us treat a potentially exponential number of linear sets in a uniform way. This, in turn, enables us, for instance, to obtain bounds on the representation of the intersection of two hybrid linear sets of the form $L(B,P)$ where, as one would indeed expect, the magnitude of the generators of the result does not depend on the cardinality of B (Subsection 2.3).

Our further path to the results on the complement and set-theoretic difference of semi-linear sets (Section 4) goes through another development, a *proper disjoint decomposition theorem*. It splits a hybrid linear set into a union $\bigcup_{i \in I} L(B_i, P_i)$ where each P_i is *proper* (i.e., consists of linearly independent vectors) and the *convex hulls* of $L(B_i, P_i)$ are disjoint (Section 3). For this result, we use the concept of a generalised simplex in order to construct triangulations of infinite polyhedra in $\mathbb{Q}^d$, and use the technique of half-open decompositions to ensure the disjointness of the aforementioned convex hulls.

Decomposing $\mathbb{Q}^d$ into convex polyhedra is by no means a new technique in the study of semi-linear sets. In particular, such decompositions were used by Huynh [12, 11] and recently by Kopczyński [15] in the context of semi-linear set inclusion. However, our decomposition theorem is different from theirs and gives stronger corollaries, in that we obtain a full semi-linear representation of the complement and, through intersection, of the set-theoretic difference. While the window theorem of Kopczyński in [15] gives an upper bound on the magnitude of the smallest vector in the set difference, our results upper-bound the magnitude of the largest generator.

2 Preliminaries

2.1 Basic definitions

Let $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{Q}$, and $\mathbb{Q}_{\geq 0}$ denote the set of integers, non-negative integers, rationals, and non-negative rationals, respectively. For $x \in \mathbb{Q}$, $\lfloor x \rfloor$ is the largest integer that does not exceed x. For subsets of numbers or vectors A and B, we use the Minkowski sum notation: $A + B := \{a + b : a \in A, b \in B\}$. In this and other contexts, we often omit the curly braces when referring to singletons. For sets of vectors $P = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\}, Q \subseteq \mathbb{Z}^m$, we may assume some fixed ordering on their elements, e.g., a lexicographic ordering, and thus sometimes treat P as a matrix whose column vectors are $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$. This leads to the notation $P \cdot \boldsymbol{\lambda}$ and $P \cdot Q$, for products of P with a vector $\boldsymbol{\lambda}$ and a matrix Q, respectively.

Linear, hybrid linear, and semi-linear sets

Suppose a natural $d \geq 1$ is fixed; we will call this d the *dimension*. A set $L \subseteq \mathbb{Z}^d$ is called *linear* if it is of the form

$$L = L(\boldsymbol{b}, P) := \{\boldsymbol{b} + \lambda_1 \boldsymbol{p}_1 + \cdots + \lambda_k \boldsymbol{p}_k : \lambda_1, \ldots, \lambda_k \in \mathbb{N}, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_k \in P\} \tag{2}$$

where $\boldsymbol{b} \in \mathbb{Z}^d$ and $P \subseteq \mathbb{Z}^d$ is a finite set. We call the vector $\boldsymbol{b}$ the *base vector* and vectors $\boldsymbol{p} \in P$ the *period vectors* (or simply *base* and *periods*) of L. A set $S \subseteq \mathbb{Z}^d$ is called *semi-linear* if it is a finite union of linear sets. Semi-linear sets can be represented as

$$S = \bigcup_{i \in I} L(B_i, P_i) \quad \text{where} \tag{3}$$

$$L(B_i, P_i) := \bigcup_{\boldsymbol{b}_i \in B_i} L(\boldsymbol{b}_i, P_i) \tag{4}$$

and $L(\boldsymbol{b}_i, P_i)$ is as in (2); we call sets $L(B_i, P_i)$ in (4) *hybrid linear sets.* Every linear set is also a hybrid linear set, and every hybrid linear set is semi-linear, but the converse statements are not true in general.

A hybrid linear set $L(B_i, P_i)$ is *proper* if the vectors P_i are linearly independent. Moreover, a hybrid linear set $L(B_i, P_i)$, $\#P_i = r$, is called *unambiguous* if for every $\boldsymbol{x} \in L(B_i, P_i)$ there exist a unique $\boldsymbol{b} \in B_i$ and a unique $\boldsymbol{\lambda} \in \mathbb{N}^r$ such that $\boldsymbol{x} = \boldsymbol{b} + P_i \cdot \boldsymbol{\lambda}$. A representation $\bigcup_{i \in I} L(B_i, P_i)$ is an *unambiguous decomposition* if each hybrid linear set $L(B_i, P_i)$ is unambiguous and the union is disjoint.

From the computational perspective, it is standard to represent semi-linear sets of the form (3) by listing all vectors in the sets B_i, P_i for all $i \in I$; components of the vectors are written in binary. We use the following notation to refer to various size measures for this representation. For any set A, the number of elements of A is $\#A$. For any $\boldsymbol{v} = (v_1, \ldots, v_d) \in \mathbb{Z}^d$, $\|v\| := \max_{1 \le i \le d} |v_i|$; similarly, for any $A \subseteq \mathbb{Z}^d$ we denote $\|A\| := \max_{\boldsymbol{v} \in A} \|\boldsymbol{v}\|$; observe that $\#A \le (2\|A\| + 1)^d$. Finally, for the representation (3) of a semi-linear set S we write $\|S\| := \max(\max_{i \in I} \|B_i\|, \max_{i \in I} \|P_i\|, 2)$, $\#_{\mathsf{b}} S := \max_{i \in I} \#B_i$, and $\#_{\mathsf{p}} S := \max_{i \in I} \#P_i$.

Convex polyhedra

We now introduce some terminology and notation from convex geometry (see, e.g., [22, 4, 19, 23]). For a system of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{Q}^d$, a linear combination $\lambda_1 \boldsymbol{v}_1 + \ldots + \lambda_k \boldsymbol{v}_k$ with $\lambda_1, \ldots, \lambda_k \in \mathbb{Q}$ is called: *non-negative*, or *conical*, if all $\lambda_i \ge 0$; *affine* if $\sum_{i=1}^k \lambda_i = 1$; and *convex* if it is non-negative and affine. For a possibly infinite set of vectors $A \subseteq \mathbb{Q}^d$, by $\operatorname{cone} A$, $\operatorname{aff} A$, and $\operatorname{conv} A$ we denote the *(rational) cone generated by* A, the *affine hull of* A, and the *convex hull of* A, respectively: they are the sets of all non-negative, affine, and convex combinations of finite subsets of A, respectively. We use the convention that $\boldsymbol{0} \in \operatorname{cone} A$ for any A; in particular, $\operatorname{cone} \emptyset = \{\boldsymbol{0}\}$. However, $\operatorname{conv} \emptyset = \emptyset$. Sets of the form $\boldsymbol{b} + \operatorname{cone} A$, for $\boldsymbol{b} \in \mathbb{Q}^d$, are *shifted cones*; we often refer to them simply as cones.

For any non-empty set $X \subseteq \mathbb{Q}^d$ its affine hull satisfies $\operatorname{aff} X = X_0 + \boldsymbol{v}$ for some vector $\boldsymbol{v} \in \mathbb{Q}^d$ and a uniquely determined linear subspace of $\mathbb{Q}^d$ denoted X_0. The *dimension* of X, written as $\dim X$, is the dimension of the subspace X_0.

A *(rational) convex polyhedron* in $\mathbb{Q}^d$ is a set of the form $\{\boldsymbol{x} \in \mathbb{Q}^d : A \cdot \boldsymbol{x} \le \boldsymbol{c}\}$ where $A \in \mathbb{Z}^{m \times d}$, $\boldsymbol{c} \in \mathbb{Z}^m$ for some m, and $\le$ is interpreted compontent-wise. A *face* of a convex polyhedron $W \subseteq \mathbb{Q}^d$ is a set of points where some linear function $\eta \colon \mathbb{Q}^d \to \mathbb{Q}$ achieves its maximum η^* over W; if η is non-constant, the hyperplane $h = \{\boldsymbol{x} \in \mathbb{Q}^d : \eta(\boldsymbol{x}) = \eta^*\}$ is a *supporting hyperplane* of W. A face of a convex polyhedron is always a convex polyhedron itself. Faces of dimension 0, 1, and $\dim W - 1$ are *vertices*, *edges*, and *facets* respectively. All faces of W form a partial order with respect to set inclusion, the largest element being the set W itself (it is always a face).

For a hybrid linear set $L(B, P)$, we denote $K(B, P) := \operatorname{conv} L(B, P) = \operatorname{conv} B + \operatorname{cone} P$. Note that if B is a singleton, i.e., if $L(B, P)$ is a linear set, then $K(B, P)$ is a rational cone; in general, though, $K(B, P)$ is a convex polyhedron.

Given a set S, we call its representation (3) a *proper disjoint decomposition* if each hybrid linear set $L(B_i, P_i)$ is proper and $K(B_i, P_i) \cap K(B_j, P_j) = \emptyset$ for $i \neq j$.

2.2 Auxiliary tools: Systems of linear inequalities

Let $A \in \mathbb{Z}^{m \times n}$ be an integer $m \times n$ matrix and $\boldsymbol{c} \in \mathbb{Z}^m$. We call $\mathfrak{S} \colon A \cdot \boldsymbol{x} \le \boldsymbol{c}$ a *system of linear inequalities* and $\mathfrak{T} \colon A \cdot \boldsymbol{x} = \boldsymbol{c}$ a *system of linear equations*. By $[\![\mathfrak{S}]\!], [\![\mathfrak{T}]\!] \subseteq \mathbb{Z}^n$ we denote the *solution set* of $\mathfrak{S}$ and $\mathfrak{T}$, i.e, the set of all $\boldsymbol{v} \in \mathbb{Z}^n$ such that $A \cdot \boldsymbol{v} \le \boldsymbol{c}$ and $A \cdot \boldsymbol{v} = \boldsymbol{c}$, respectively. We use $[\![\mathfrak{T}]\!]_{\ge 0}$ as a shorthand for $[\![\mathfrak{T}]\!] \cap \mathbb{N}^n$, and write $(\!|\mathfrak{S}|\!)$ for the set of rational solutions from $\mathbb{Q}^n$ of $\mathfrak{S}$. Moreover, we define $\|\mathfrak{S}\|, \|\mathfrak{T}\| := \max\{\|A\|, \|\boldsymbol{c}\|\}$.

The following two propositions connect two representations of polyhedra in $\mathbb{Q}^d$.

▶ **Proposition 1** ([23]). *Let $\mathfrak{S} \colon A \cdot \boldsymbol{x} \le \boldsymbol{c}$ be a convex polyhedron in $\mathbb{Q}^d$. Then there are $B \subseteq \mathbb{Q}^d$ and $P \subseteq \mathbb{Z}^d$ such that $(\!|\mathfrak{S}|\!) = \operatorname{conv} B + \operatorname{cone} P$, $\|P\| \le 2^{O(d \log d)} \cdot \|\mathfrak{S}\|^d$, and all numerators and denominators in B are bounded by $2^{O(d \log d)} \cdot \|\mathfrak{S}\|^d$.*

Proof. By the Minkowski–Weyl theorem, there exist $C, Q \subseteq \mathbb{Q}^d$ such that $(\!|\mathfrak{S}|\!) = \operatorname{conv} C + \operatorname{cone} Q$. In fact, it is possible (cf. [23, Theorem 10.2]) to find C and Q in which all vectors have numerators and denominators of all entries bounded by $d! \cdot \|\mathfrak{S}\|^d$: they can essentially be chosen as solutions to linear systems defined by square submatrices of the matrix $\left[\begin{array}{c|c} A & \boldsymbol{c} \end{array}\right]$. ◀

▶ **Proposition 2** ([23]). *Let $M = L(\boldsymbol{b}, P) \subseteq \mathbb{Z}^d$ be a proper linear set with $r = \#P$. Then there exists a system of linear inequalities $\mathfrak{S} \colon A \cdot \boldsymbol{x} \le \boldsymbol{c}$ such that*

- *A is a $(2d - r) \times d$ matrix that does not depend on $\boldsymbol{b}$;*
- *$\|A\| \le 2^{O(r \log r)} \cdot \max(\|P\|, 1)^r$; $\|\boldsymbol{c}\| \le d \cdot \|A\| \cdot \|\boldsymbol{b}\|$; and*
- *$\operatorname{conv} L(\boldsymbol{b}, P) = (\!|\mathfrak{S}|\!)$.*

Proof (sketch). Since M is proper, $\operatorname{conv} L(\boldsymbol{b}, P)$ has exactly r facets; r inequalities in $\mathfrak{S}$ define them, with another $2(d - r)$ for $\operatorname{aff} M$. It can be shown (cf. [23, Theorem 10.2]) that there exists an appropriate A such that $(\!| A \cdot \boldsymbol{x} \le \boldsymbol{0} |\!) = \operatorname{conv} L(\boldsymbol{0}, P)$; and then $\boldsymbol{c} = A \cdot \boldsymbol{b}$. ◀

We will also need a result of von zur Gathen and Sieveking on the sets of all integer solutions of systems of linear inequalities [24].

▶ **Proposition 3.** *Let $\mathfrak{S} \colon A \cdot \boldsymbol{x} \le \boldsymbol{c}$ be a system of inequalities such that $A \in \mathbb{Z}^{m \times n}$. Then $[\![\mathfrak{S}]\!] = \bigcup_{i \in I} L(B_i, P_i)$ such that*

- *$K(B_i, P_i) \cap K(B_j, P_j) = \emptyset$ for all $i \neq j$,*
- *$\max_{i \in I} \|B_i\|, \max_{i \in I} \|P_i\| \le 2^{O(n \log n)} \cdot \|A\|^{n-1} \cdot \|\mathfrak{S}\|$, and*
- *$\#I \le 2^n$.*

Next, we additionally recall a result on the sets of integer solutions of linear equalities that follows from results of Pottier [21].

▶ **Proposition 4.** *Let $\mathfrak{S}_0 \colon A \cdot \boldsymbol{x} = \boldsymbol{0}$ and $\mathfrak{S} \colon A \cdot \boldsymbol{x} = \boldsymbol{c}$ be systems of linear Diophantine equations, where $A \in \mathbb{Z}^{m \times n}$. Then $[\![\mathfrak{S}_0]\!]_{\ge 0} = L(\boldsymbol{0}, P)$ and $[\![\mathfrak{S}]\!]_{\ge 0} = L(B, P)$ such that*

- *$\|B\| \le ((n + 1) \cdot \|A\| + \|\boldsymbol{c}\| + 1)^m$, $\|P\| \le (n \cdot \|A\| + 1)^m$.*

Finally, we will need a discrete version of Carathéodory's theorem:

▶ **Proposition 5.** *Let $M = L(C, Q) \subseteq \mathbb{Z}^d$ be a hybrid linear set. Then $M = \bigcup_{i \in I} L(B_i, P_i)$ such that*

- *$\max_{i \in I} \|B_i\| \le \|C\| + (\#Q \cdot \|Q\|)^{O(d)}$,*

- $\max_{i \in I} \#P_i \leq d$, $P_i \subseteq Q$ *and each* P_i *is linearly independent, and*
- $\#I \leq (\#Q)^d$.

The statement of Proposition 5 can essentially be shown by a combination of Lemmas 2.7 and 2.8 in [10], which, however, do not establish any concrete bounds. In our proof, we use the result on the intersection of hybrid linear sets from the following subsection 2.3.

2.3 Intersection of semi-linear sets

▶ **Theorem 6.** *Let* M *and* N *be semi-linear sets with representations* $M = \bigcup_{j \in J} L(C_j, Q_j)$, $N = \bigcup_{k \in K} L(D_k, R_k)$. *Then the set* $L := M \cap N$ *is a semi-linear set with representation* $L = \bigcup_{i \in I} L(B_i, P_i)$ *such that* $I = J \times K$,

- $\max_{i \in I} \|B_i\|, \max_{i \in I} \|P_i\| \leq ((\#_{\mathsf{p}} M + \#_{\mathsf{p}} N) \cdot \max(\|M\|, \|N\|))^{O(d)}$, *and*
- $\#I \leq \#J \cdot \#K$.

Moreover, if $Q_j \subseteq R_k$ *and* $i = (j, k)$ *then* $P_i = Q_j$.

Proof (sketch). We have $M \cap N = \bigcup_{j \in J} L(C_j, Q_j) \cap \bigcup_{k \in K} L(D_k, R_k) = \bigcup_{j \in J, k \in K} L(C_j, Q_j) \cap L(D_k, R_k)$. Hence it suffices to show that every $L(C_j, Q_j) \cap L(D_k, R_k)$ is some $L(B_{j,k}, P_{j,k})$ with the desired properties. To this end, one can obtain the set of elements in the intersection as the set of solutions to a suitable system of linear equations and then apply the bounds from Proposition 4. Finally, the fact that if $Q_j \subseteq R_k$ then $P_i = Q_j$ follows from Theorem 5.6.1 in [5, p. 180]. ◀

3 Hybrid linear sets

In the sequel, we develop a close connection between hybrid linear sets and convex polyhedra viewed as generalized convex hulls. Convex polyhedra in $\mathbb{Q}^d$ are sets of the form $\operatorname{conv} C + \operatorname{cone} Q$ for $C, Q \subseteq \mathbb{Q}^d$; they can be viewed as a convex hulls of a set of points C and *directions* Q. Suppose $C = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r\}$ and $Q = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m\}$. The connection builds upon on the similarity of the following sets:

$$\operatorname{conv} C + \operatorname{cone} Q = \left\{ \sum_{i=1}^{r} \lambda_i \boldsymbol{b}_i + \sum_{j=1}^{m} \mu_j \boldsymbol{p}_j : \lambda_i \in \mathbb{Q}_{\geq 0},\ \sum_{i=1}^{r} \lambda_i = 1,\ \mu_j \in \mathbb{Q}_{\geq 0} \right\} \text{ and}$$

$$L(C, Q) = \left\{ \sum_{i=1}^{r} \lambda_i \boldsymbol{b}_i + \sum_{j=1}^{m} \mu_j \boldsymbol{p}_j : \lambda_i \in \mathbb{N},\ \sum_{i=1}^{r} \lambda_i = 1,\ \mu_j \in \mathbb{N} \right\}.$$

As mentioned above, $\operatorname{conv} L(C, Q) = K(C, Q) = \operatorname{conv} C + \operatorname{cone} Q$.

3.1 Proper disjoint decompositions (PDD)

Recall that $S = \bigcup_{i \in I} L(B_i, P_i)$ is a proper disjoint decomposition if vectors in each P_i are linearly independent and the convex hulls $K(B_i, P_i) = \operatorname{conv} L(B_i, P_i)$ are pairwise disjoint.

▶ **Theorem 7** (PDD for hybrid linear sets). *Every hybrid linear set* $M = L(C, Q)$ *has a proper disjoint decomposition* $\bigcup_{i \in I} L(B_i, P_i)$ *where each* P_i *is a subset of* Q *and the following inequalities hold:*

- $\|B_i\| \leq (\#Q + \|C\| + \|Q\| + d)^{O(d)} \leq \|M\|^{O(d^2)}$ *and*

- $\#I \leq (\#Q)^{d+1}$.

The idea of the proof of Theorem 7 is to rely on the connection between hybrid linear sets and convex polyhedra. We will use the observation that each set of the form $\operatorname{conv} C + \operatorname{cone} Q$ has a triangulation. While this term usually refers to the basic construction that splits a convex polygon in a plane into a number of non-overlapping triangles, we will use a construction that extends this concept in two ways: first, instead of $\mathbb{Q}^2$ the sets are in $\mathbb{Q}^d$, so triangles become simplices; second, the sets can be infinite, i.e., with $Q \neq \emptyset$.

The strategy of the proof of Theorem 7 is depicted in the following diagram:

$$\begin{tikzcd}
L(C,Q) \arrow[r, "3"] \arrow[d, "1"] & \Pi, \text{ a proper disjoint decomposition of } L(C,Q) \\
K(C,Q) \arrow[r, "2"] & \mathcal{T}, \text{ a triangulation of } K(C,Q) \arrow[u, "3"]
\end{tikzcd}$$

Step 1 is taking the convex hull, step 2 is triangulation in $\mathbb{Q}^d$, and step 3 constructs a proper disjoint decomposition given the original set $L(C,Q)$ and the triangulation of $K(C,Q)$.

A *generalized δ-dimensional simplex* T is a set of the form $T = \operatorname{conv} V + \operatorname{cone} D \subseteq \mathbb{Q}^d$ where $\#V + \#D = \delta + 1$, $V \neq \emptyset$, and the dimension of the affine hull of T is exactly δ (cf. [22, pp. 153f]). Elements of V are ordinary vertices of T, and elements of D are vertices at infinity and can be understood as directions. (The set D is, in fact, the set of *extreme directions* of the set T; see [22, p. 162].) Faces of generalized simplices $\operatorname{conv} V + \operatorname{cone} D$ are also generalized simplices and have the form $\operatorname{conv} V' + \operatorname{cone} D'$ where $V' \subseteq V$ and $D' \subseteq D$.

A *triangulation* of a set $W \subseteq \mathbb{Q}^d$ is a collection $\mathcal{T}$ of generalized simplices that satisfies the following properties:

1. $\bigcup_{F \in \mathcal{T}} F = W$.
2. For every $F \in \mathcal{T}$ and every face F' of F, it holds that $F' \in \mathcal{T}$.
3. The intersection of any two $F_1, F_2 \in \mathcal{T}$ is either empty or is a face of both F_1 and F_2.
4. All (generalized) simplices in the set of maxima of $\mathcal{T}$, denoted $\operatorname{Max} \mathcal{T} := \{F' \in \mathcal{T} : \nexists F \in \mathcal{T}.\ F' \text{ is a face of } F \text{ and } F \neq F'\}$, have the same dimension δ, denoted $\dim \mathcal{T}$.

In other words, a triangulation of W is a pure polyhedral complex (see, e.g., [4, Chapter 6]) that consists of generalized simplices and covers exactly W.

To simplify notation, we write $\mathcal{T} = (T_1, \ldots, T_m)$ whenever $\operatorname{Max} \mathcal{T} = \{T_1, \ldots, T_m\}$; of course, the set $\{T_1, \ldots, T_m\}$ is a subset of the set $\mathcal{T}$. It is straightforward that $W = T_1 \cup \ldots \cup T_m$ if $\mathcal{T} = (T_1, \ldots, T_m)$ is a triangulation of W. Conversely, if $T_1, \ldots, T_m$ are (generalized) simplices of equal dimension such that the collection $\mathcal{T}$ of all their faces satisfies Condition 3 in the definition of triangulation, then this collection $\mathcal{T}$ is a triangulation of $T_1 \cup \ldots \cup T_m$. Lemma 8 triangulates possibly unbounded convex polyhedra (for non-empty Q, it treats its elements as vertices at infinity) without introducing new vertices or directions.

▶ **Lemma 8.** *Every polyhedron of the form $W = \operatorname{conv} C + \operatorname{cone} Q \subseteq \mathbb{Q}^d$ has a triangulation $\mathcal{T} = (T_1, \ldots, T_m)$ where $m \leq (\#C + \#Q)^{d+1}$ and $T_i = \operatorname{conv} C_i + \operatorname{cone} Q_i$ with $C_i \subseteq C$ and $Q_i \subseteq Q$ for all i.*

Note that adjacent simplices T_i and T_j in a triangulation can share points in common lower-dimensional faces. However, for our purposes they should be made disjoint. Suppose U is a polyhedron of the form $X = \{\boldsymbol{x} \in \mathbb{Q}^d : \boldsymbol{a}_i \cdot \boldsymbol{x} \leq c_i, 1 \leq i \leq m\}$ where $\boldsymbol{a}_i \in \mathbb{Z}^d$ and $c_i \in \mathbb{Z}$ for all i. For any $A \subseteq \{1, \ldots, m\}$, we call the set

$$X_A = \{\boldsymbol{x} \in \mathbb{Q}^d : \boldsymbol{a}_i \cdot \boldsymbol{x} < c_i, i \in A, \text{ and } \boldsymbol{a}_i \cdot \boldsymbol{x} \leq c_i, i \in \{1, \ldots, m\} \setminus A\}$$

a *half-opening* of U obtained by cutting off the hyperplanes $\boldsymbol{a}_i \cdot \boldsymbol{x} = c_i$, $i \in A$.

▶ **Lemma 9.** *Let W be a δ-dimensional polyhedron in $\mathbb{Q}^d$. For each triangulation $\mathcal{T} = (T_1, \dots, T_m)$ of W there exists a collection of sets $\mathcal{T}^0 = (T_1^0, \dots, T_m^0) \subseteq \mathbb{Q}^d$ that satisfies the following conditions:*

1. $T_1^0 \cup \ldots \cup T_m^0 = W$.
2. *For every i, T_i^0 is a half-opening of T_i.*
3. *T_i and T_j are disjoint for $i \neq j$.*

Lemma 9 is the *half-open decomposition*, originally from [1] and [16]. Our formulation is a direct corollary of Theorem 3 in the latter paper; see also [8, Section 3.2].

▶ **Lemma 10.** *Suppose $T = \operatorname{conv} V + \operatorname{cone} D$ is a generalized δ-dimensional simplex in $\mathbb{Q}^d$ where $V, D \subseteq \mathbb{Z}^d$ and $\#V + \#D = \delta + 1$. Then for any half-opening T^0 of T it holds that $T^0 \cap \mathbb{Z}^d = L(E, D)$ where $\|E\| \leq \|V\| + (d+1) \cdot \|D\|$.*

Lemma 10 makes it possible to use half-open decomposition in the proof of Theorem 7.

Proof of Theorem 7 (sketch). Take a triangulation of $W = K(C, Q) = \operatorname{conv} C + \operatorname{cone} Q$, which exists by Lemma 8, and apply Lemma 9 to this triangulation. The result is a collection $\mathcal{T}^0 = (T_1^0, \dots, T_m^0)$ where each T_i^0 is a half-opening of some generalized simplex $\operatorname{conv} C_i + \operatorname{cone} Q_i$ such that $C_i \subseteq C$ and $Q_i \subseteq Q$. By Lemma 10, $T_i^0 \cap \mathbb{Z}^d = L(D_i, Q_i)$. We now apply Theorem 6: since $Q_i \subseteq Q$, we have $L(D_i, Q_i) \cap L(C, Q) = L(B_i, P_i)$ where $P_i = Q_i$. Vectors in each set $P_i = Q_i$ are, in fact, linearly independent, because $\operatorname{conv} C_i + \operatorname{cone} Q_i$ is a generalized simplex. Moreover, $K(B_i, P_i) \subseteq \operatorname{conv} L(D_i, Q_i) \subseteq T_i^0$ for each i; since the sets $T_1^0, \dots, T_m^0$ are pairwise disjoint, so are the sets $K(B_i, P_i)$. Finally,

$$\begin{aligned} \bigcup_{i=1}^{m} L(B_i, P_i) &= \bigcup_{i=1}^{m} T_i^0 \cap \mathbb{Z}^d \cap L(C, Q) = L(C, Q) \cap \bigcup_{i=1}^{m} T_i^0 \\ &= L(C, Q) \cap W = L(C, Q) \cap \operatorname{conv} L(C, Q) = L(C, Q). \end{aligned}$$ ◀

3.2 Unambiguous decompositions (UD)

The main results of this subsection are the following theorems:

▶ **Theorem 11** (UD for proper hybrid linear sets)**.** *Every proper hybrid linear set $M = L(C, Q)$ has an unambiguous decomposition $\bigcup_{i \in I} L(B_i, P_i)$ where each P_i is a subset of Q and the following conditions are satisfied:*

- $\|B_i\| \leq \|C\|$ *and*
- $\#I \leq (2 \cdot \#C)^{\#Q}$.

▶ **Theorem 12** (UD for hybrid linear sets)**.** *Every hybrid linear set $M = L(C, Q)$ has an unambiguous decomposition $\bigcup_{i \in I} L(B_i, P_i)$ where each P_i is a subset of Q and the following inequalities hold:*

- $\|B_i\| \leq (\#Q + \|C\| + \|Q\| + d)^{O(d)} \leq \|M\|^{O(d^2)}$ *and*
- $\#I \leq ((\|C\| + \|Q\| + d)^{O(d)} + \#C)^d \cdot (d + \#Q)^{O(d^2)} \leq \|M\|^{O(d^3)}$.

We now show how to prove Theorem 11. The idea is to reduce the disambiguation of a proper hybrid linear set to disambiguation of an ideal in a finitely generated commutative monoid, captured by the following lemma. Here and below, by $\boldsymbol{e}_1, \dots, \boldsymbol{e}_r$ we denote coordinate vectors in $\mathbb{N}^r$.

▶ **Lemma 13.** *Every set of the form $U = L(F, \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_r\}) = F + \mathbb{N}^r$ with a finite $F \subseteq \mathbb{N}^r$ has a representation $U = \bigcup_{k \in K} L(G_k, E_k)$ such that the following conditions are satisfied:*

- *each set* $L(G_k, E_k)$ *is unambiguous,*
- *the polyhedra* $\operatorname{conv} L(G_k, E_k)$ *and* $\operatorname{conv} L(G_{k'}, E_{k'})$ *are disjoint for* $k \neq k'$*,*
- $\|G_k\| \leq \|F\|$*,*
- $E_k \subseteq \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r\}$*, and*
- $\#K \leq (\#F + 1)^r$*.*

Proof (sketch). The condition that a vector $\boldsymbol{x}$ belongs to $F + \mathbb{N}^r$ can be specified by a logical formula Φ over predicates of the form $x_j \geq c$. These predicates break up $\mathbb{N}^r$ into at most $(\#F + 1)^r$ disjoint regions, and each region is described by a unambiguous hybrid linear set in a straightforward way. ◀

Proof of Theorem 11 (sketch). Take $M = L(C, Q) \subseteq \mathbb{Z}^d$ where $Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r\} \subseteq \mathbb{Z}^d$ and vectors in Q are linearly independent, $r \leq d$. Consider the *point lattice* $\mathcal{L} = Q \cdot \mathbb{Z}^r = \{Q \cdot \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{Z}^r\}$; see, e.g., [17, Chapter 2]. Vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^r$ are *congruent* modulo $\mathcal{L}$, written $\boldsymbol{x} \equiv \boldsymbol{y} \pmod{\mathcal{L}}$, if and only if $\boldsymbol{x} - \boldsymbol{y} \in \mathcal{L}$. This congruence splits the set C into a disjoint union $C = C_1 \cup \ldots \cup C_s$ where $\boldsymbol{x} \in C_i$ and $\boldsymbol{y} \in C_j$ are congruent if and only if $i = j$. It is easy to see that $M = \bigcup_{1 \leq j \leq s} L(C_j, Q)$ is a disjoint union, and disambiguating each $L(C_j, Q)$ separately will disambiguate M.

Suppose $C_1 = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\} \subseteq \boldsymbol{x}_1 + \mathcal{L}$. Since the vectors in $Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r\}$ are linearly independent, each vector from the set $\boldsymbol{x}_1 + \mathcal{L}$ has a unique expansion of the form $\boldsymbol{x}_1 + \sum_{j=1}^{r} a_j \boldsymbol{q}_j$. Consider the mapping $\psi \colon \boldsymbol{x}_1 + \mathcal{L} \to \mathbb{Z}^r$ taking each vector $\boldsymbol{x}_1 + \sum_{j=1}^{r} a_j \boldsymbol{q}_j$ to the vector $(a_1, \ldots, a_r) \in \mathbb{Z}^r$. For each j, let a_j^0 be the smallest of the numbers $\psi(\boldsymbol{x}_t)[j]$ over $1 \leq t \leq m$; here $[j]$ refers to the jth component of an r-dimensional vector. Denote $\boldsymbol{a}^0 = (a_1^0, \ldots, a_r^0)$ and let $\psi' \colon \boldsymbol{x}_1 + \mathcal{L} \to \mathbb{Z}^r$ be given by $\psi'(\boldsymbol{x}) = \psi(\boldsymbol{x}) - \boldsymbol{a}^0$. Observe that the mapping ψ' is injective and maps C_1 to some finite set $F \subseteq \mathbb{N}^r$; in fact, $\psi'(L(C_1, Q_1)) = F + \mathbb{N}^r$. After this, it remains to apply Lemma 13. ◀

4 Semi-linear sets

In this section, we derive our main results on the complement, set-theoretic difference, and decompositions of semi-linear sets. We will rely on Theorems 7 and 11 from Section 3.

4.1 Geometric ingredients: Splitting into atomic polyhedra

Consider a semi-linear set given by $M = \bigcup_{j \in J} L(C_j, Q_j)$. Take the proper disjoint decomposition of each $L(C_j, Q_j)$ according to Theorem 7; this decomposes M as

$$M = \bigcup_{j \in J} \bigcup_{t \in T_j} L(C_{jt}, Q_{jt}) \tag{5}$$

where hybrid linear sets $L(C_{jt}, Q_{jt})$ are proper, $Q_{jt} \subseteq Q_j$, and, moreover, for any fixed j the polyhedra $K(C_{jt}, Q_{jt})$ are pairwise disjoint.

Denote by $\mathcal{H}$ the collection of principal supporting hyperplanes for shifted cones $K(\boldsymbol{b}, Q_{jt})$, $\boldsymbol{b} \in C_{jt}$, $t \in T_j$, and $j \in J$: for each cone, take its d principal supporting hyperplanes, i.e., the hyperplanes obtained in Proposition 2, each of the form $h \colon \boldsymbol{a} \cdot \boldsymbol{x} = c$ (with fixed $\boldsymbol{a} \in \mathbb{Z}^d$ and $c \in \mathbb{Z}$), and put them into $\mathcal{H}$. Note that each hyperplane h' is associated with half-spaces $h^- \colon \boldsymbol{a} \cdot \boldsymbol{x} \leq c$ and $h^+ \colon \boldsymbol{a} \cdot \boldsymbol{x} \geq c + 1$; moreover, we can pick the signs so that $K(\boldsymbol{b}, Q_{jt}) \subseteq (\!| h^- |\!)$. An *atomic polyhedron* with respect to $\mathcal{H}$ is a non-empty set of the form

$$A(H) = \bigcap_{h \in H} (\!| h^- |\!) \ \cap \bigcap_{h \in \mathcal{H} \setminus H} (\!| h^+ |\!),$$

where $H \subseteq \mathcal{H}$. Clearly, $\mathbb{Z}^d \subseteq \bigcup_{H \subseteq \mathcal{H}} A(H)$, and $A(H) \cap A(H') = \emptyset$ whenever $H \neq H'$.

▶ **Lemma 14.** *For every $L(\boldsymbol{b}, Q_{jt})$ with $\boldsymbol{b} \in C_{jt}$ and every $A = A(H)$, either $A \subseteq \operatorname{conv} L(\boldsymbol{b}, Q_{jt})$ or $A \cap \operatorname{conv} L(\boldsymbol{b}, Q_{jt}) = \emptyset$.*

Take a hybrid linear set $L(C_{jt}, Q_{jt})$ and let $\boldsymbol{b} \in C_{jt}$. We say that the linear set $L(\boldsymbol{b}, Q_{jt})$ *shares* an atomic polyhedron A iff $A \subseteq \operatorname{conv} L(\boldsymbol{b}, Q_{jt})$; otherwise we say that it *avoids* A.

▶ **Lemma 15.** *Every atomic polyhedron $A(H)$ is the set of rational solutions to a system of at most $O(d \cdot \sum_{j \in J} (\#Q_j)^{d+1})$ linear inequalities with entries bounded by $\|M\|^{O(d^2)}$.*

▶ **Lemma 16.** *The number of atomic polyhedra is at most $\left(d \cdot \sum_{j \in J} \#C_j \cdot (\#Q_j)^{d+1}\right)^{d+1}$.*

Consider an atomic polyhedron A; we will assume in the remainder of this subsection that A is shared by at least one linear set of the form $L(\boldsymbol{b}, Q_{jt})$. Even though the total number of linear sets of this form that share A can be large, the following property holds.

▶ **Lemma 17.** *If linear sets $L(\boldsymbol{b}, Q_{jt})$ and $L(\boldsymbol{b}', Q_{jt'})$ share A, then $t = t'$. In particular, the number of pairs (j, t) such that some linear set $L(\boldsymbol{b}, Q_{jt})$ shares A does not exceed $\#J$.*

▶ **Lemma 18.** *For every A there exist finite sets $E \subseteq \mathbb{Q}^d$ and $G \subseteq \mathbb{Z}^d$ that satisfy the following conditions:*

1. $A = \operatorname{conv} E + \operatorname{cone} G$.
2. *For every linear set $L(\boldsymbol{b}, Q_{jt})$ that shares A, the set G is a subset of $L(\boldsymbol{0}, Q_{jt})$.*
3. *Numerators and denominators of all entries in all $\boldsymbol{e} \in E$ are bounded by $\|M\|^{O(d^3)}$.*
4. $\|G\| \leq \|M\|^{\#J \cdot O(d^4)}$.
5. $\#G \leq \|M\|^{O(d^4)}$.

The proof of Lemma 18 first applies Proposition 1 to the representation of A from Lemma 15. The upper bound on $\|G\|$ then relies on the fact that, for every $j \in J$, our decomposition (5) ensures disjointness of $K(C_{jt}, Q_{jt})$ among $t \in T_j$; the proof uses this property via Lemma 17.

4.2 Decompositions, complement, and difference

We first state the results on decompositions of semi-linear sets and on the semi-linear representation of the complement.

▶ **Theorem 19** (PDD for semi-linear sets). *Every semi-linear set $M = \bigcup_{j \in J} L(C_j, Q_j)$ has a proper disjoint decomposition $\bigcup_{i \in I} L(B_i, P_i)$ where*

- $\|B_i\| \leq \|M\|^{\#J \cdot O(d^6)}$,
- $\|P_i\| \leq \|M\|^{\#J \cdot O(d^4)}$, *and*
- $\#I \leq \|M\|^{O(d^5)}$.

▶ **Corollary 20** (UD for semi-linear sets). *Every semi-linear set $M = \bigcup_{j \in J} L(C_j, Q_j)$ has an unambiguous decomposition $\bigcup_{i \in I} L(B_i, P_i)$ where*

- $\|B_i\| \leq \|M\|^{\#J \cdot O(d^6)}$ *and*
- $\|P_i\| \leq \|M\|^{\#J \cdot O(d^4)}$.

▶ **Theorem 21** (complement of semi-linear sets). *The complement of every semi-linear set $M = \bigcup_{j \in J} L(C_j, Q_j)$ has a representation of the form $\bigcup_{i \in I} L(B_i, P_i)$ where*

- $\|B_i\| \leq \|M\|^{\#J \cdot O(d^4)}$ *and*
- $\|P_i\| \leq \|M\|^{\#J \cdot O(d^4)}$.

We will state the results on set difference at the end of this subsection, and now we focus our attention on Theorems 19 and 21. Corollary 20 follows from Theorems 19 and 11.

Recall that in Subsection 4.1 we decomposed the space into disjoint atomic polyhedra A. Each $A = \operatorname{conv} E + \operatorname{cone} G$ by Lemma 18, with $E \subseteq \mathbb{Q}^d$ and $G \subseteq \mathbb{Z}^d$. By Carathéodory's theorem, for every vector $\boldsymbol{x} \in A$ there are $\nu_{\boldsymbol{e}} \in \mathbb{N}$, $\boldsymbol{e} \in E$, and $\mu_{\boldsymbol{g}} \in \mathbb{N}$, $\boldsymbol{g} \in G$, such that $\boldsymbol{x}$ has an expansion of the form

$$\boldsymbol{x} = \sum_{\boldsymbol{e} \in E} \nu_{\boldsymbol{e}} \cdot \boldsymbol{e} + \sum_{\boldsymbol{g} \in G'} \mu_{\boldsymbol{g}} \cdot \boldsymbol{g} = \tau(\boldsymbol{x}) + \pi(\boldsymbol{x}), \tag{6}$$

where $\tau(\boldsymbol{x}) = \sum_{\boldsymbol{e} \in E} \nu_{\boldsymbol{e}} \cdot \boldsymbol{e} + \sum_{\boldsymbol{g} \in G'} (\mu_{\boldsymbol{g}} - \lfloor \mu_{\boldsymbol{g}} \rfloor) \cdot \boldsymbol{g}$ denotes the *truncation* of $\boldsymbol{x}$, $\pi(\boldsymbol{x}) = \sum_{\boldsymbol{g} \in G'} \lfloor \mu_{\boldsymbol{g}} \rfloor \cdot \boldsymbol{g}$ denotes the *periodic part* of $\boldsymbol{x}$, and $G' \subseteq G$ is some subset of linearly independent vectors in G. We will consider sets $X = A \cap \mathbb{Z}^d \setminus M$ and $Y = A \cap \mathbb{Z}^d \cap M = A \cap M$.

It is not difficult to show that $\|\tau(X)\|$ and $\|\tau(Y)\|$ are bounded from above by $\|M\|^{\#J \cdot \mathsf{poly}(d)}$; these estimations are relevant, as we prove that the equalities $X = L(\tau(X), G)$ and $Y = L(\tau(Y), G)$ hold. While the latter equality requires no sophisticated arguments, a proof of the former turns out to be somewhat delicate. As an auxiliary statement, we show that $\tau(X) \subseteq X$; with this fact at hand, the proof of the inclusion $L(\tau(X), G) \subseteq X$ goes as follows. Suppose, for the sake of contradiction, that there exists a vector $\boldsymbol{z} \in L(\tau(X), G) \cap M$, say with $\boldsymbol{z} \in L(\boldsymbol{b}, Q_{jt})$ such that $L(\boldsymbol{b}, Q_{jt})$ shares A. This implies the existence of another vector $\boldsymbol{x} \in X$ with $\tau(\boldsymbol{x}) \in \boldsymbol{b} + Q_{jt} \cdot \mathbb{Z}^\delta$ where δ is the cardinality of Q_{jt}. At the same time, this $\tau(\boldsymbol{x})$ also belongs to X and thus to A and to the cone $K(\boldsymbol{b}, Q_{jt}) = \boldsymbol{b} + Q_{jt} \cdot \mathbb{Q}^\delta_{\geq 0}$. Since the vectors in Q_{jt} are linearly independent (recall that sets Q_{jt} come from a *proper* disjoint decomposition of $L(C_j, Q_j)$), it follows that $\tau(\boldsymbol{x}) \in \boldsymbol{b} + Q_{jt} \cdot \mathbb{N}^\delta = L(\boldsymbol{b}, Q_{jt})$, which contradicts the fact that $\tau(X) \subseteq X$, because X excludes M.

As seen from this sketch, our ability to construct the hybrid linear representation of X (which corresponds to the complement of M) relies on the fact that our decomposition of M in (5) uses linear sets with linearly independent periods only.

Proofs of Theorems 19 and 21 (sketch). Use equalities

$$M = \bigcup_{H \subseteq \mathcal{H}} A(H) \cap M \qquad \text{and} \qquad \mathbb{Z}^d \setminus M = \bigcup_{H \subseteq \mathcal{H}} A(H) \cap \mathbb{Z}^d \setminus M$$

where it suffices to consider only (non-empty) atomic polyhedra $A = A(H)$. Whenever all linear sets $L(\boldsymbol{b}, Q_{jt})$, $\boldsymbol{b} \in C_{jt}$ (see (5)) avoid a polyhedron A, we have $Y = A \cap M = \emptyset$ and $X = A \cap \mathbb{Z}^d \setminus M = A \cap \mathbb{Z}^d$. Here the case of Y is trivial, and the case of X sends us to Proposition 3. Otherwise, if at least one linear set shares A, we use the representations $X = L(\tau(X), G)$ and $Y = L(\tau(Y), G)$ as discussed above. For the purposes of proper disjoint decomposition (Theorem 19), we need to invoke Theorem 7 on $L(\tau(Y), G)$. Upper bounds on $\|B_i\|$, $\|P_i\|$, and $\#I$ follow from Lemmas 18 and 16 and from Theorem 7. ◀

▶ **Corollary 22** (difference of semi-linear sets)**.** *The set-theoretic difference $M \setminus N$ of semi-linear sets $M = \bigcup_{j \in J} L(C_j, Q_j)$ and $N = \bigcup_{k \in K} L(D_k, R_k)$ has a representation of the form $L = \bigcup_{i \in I} L(B_i, P_i)$, where*

- $\max_{i \in I} \|B_i\|, \max_{i \in I} \|P_i\| \leq \left(\#_{\mathsf{p}} M + \|M\| + \|N\|^{\#K \cdot d^5}\right)^{O(d)}$.

The following result combines Corollary 22 with Proposition 5.

▶ **Corollary 23** (small vector in set difference)**.** *Let M, N be semi-linear sets such that $\|M\|, \|N\| \leq m$ and $M \setminus N \neq \emptyset$. Then there is $\boldsymbol{v} \in M \setminus N$ such that $\|\boldsymbol{v}\| \leq 2^{m^{O(d^2)}}$.*

5 An application: Exponent-sensitive commutative grammars

In this section, we show that our bounds on the difference of semi-linear sets yield a novel and tight upper bound for the equivalence problem for a class of commutative grammars.

Let $\Sigma = \{a_1, \ldots, a_m\}$ be a finite alphabet. The free commutative monoid generated by Σ is denoted by $\Sigma^{\odot}$, and we treat elements of $\Sigma^{\odot}$ as vectors in $\mathbb{N}^d$ where $d = |\Sigma|$. By $\Sigma^{\oplus} := \Sigma^{\odot} \setminus \{\mathbf{0}\}$ we denote the free commutative semi-group generated by Σ. An *exponent-sensitive commutative grammar* (ESCG) is a tuple $G = (N, \Sigma, S, \Pi)$, where N is a finite set of non-terminal symbols; Σ is a finite alphabet, the set of terminal symbols such that $N \cap \Sigma = \emptyset$; $S \in N$ is the axiom; and $\Pi \subseteq (\bigcup_{U \in N} \{U\}^{\oplus}) \times (N \cup \Sigma)^{\odot}$ is a finite set of productions.

ESCG are essentially equivalent to a generalisation of communication-free Petri nets in which incoming arcs may have multiplicity greater than one [18]. The size $|G|$ of G is the number of symbols required write it down; in particular we assume that commutative words from $\Sigma^{\odot}$ are encoded in *binary*. Subsequently, we write $V \to W$ whenever $(V, W) \in \Pi$. Let $D, E \in (N \cup \Sigma)^{\odot}$, we say D directly generates E, written $D \Rightarrow_G E$, iff there are $F \in (N \cup \Sigma)^{\odot}$ and $\pi \in \Pi$ such that $\pi = V \to W$, $D = V + F$ and $E = F + W$. We write $U \Rightarrow_G^* W$ for the reflexive transitive closure of $\Rightarrow_G$ and say that U generates W in this case. If G is clear from the context, we omit the subscript G. The language $\mathcal{L}(G)$ generated by G is defined as

$$\mathcal{L}(G) := \{W \in \Sigma^{\odot} : S \Rightarrow^* W\}.$$

Given ESCG G, H and $w \in \Sigma^{\odot}$, the word problem is to decide whether $w \in \mathcal{L}(G)$, and equivalence is to decide whether $\mathcal{L}(G) = \mathcal{L}(H)$. The word problem is PSPACE-complete; the equivalence problem was shown PSPACE-hard and decidable in 2-EXPSPACE by Mayr and Weihmann [18]. The latter result has recently been improved to coNEXP-hardness and membership in co-2NEXP in [7]. An application of Corollary 23 enables us to settle the complexity of the equivalence problem for ESCG.

▶ **Theorem 24.** *Equivalence for ESCG is* coNEXP*-complete.*

Proof (sketch). Let G, H be ESCG such that $\mathcal{L}(G) \neq \mathcal{L}(H)$, and with no loss of generality assume that there is some $w \in \mathcal{L}(G) \setminus \mathcal{L}(H)$. It is shown in [18] that $M = \mathcal{L}(G)$ and $N = \mathcal{L}(H)$ are semi-linear with $\|M\|, \|N\| \leq 2^{p(|G|+|H|)}$ for some fixed polynomial p. Consequently, by Corollary 23 we may assume that $\|w\| \leq 2^{2^{q(|G|+|H|)}}$ for some fixed polynomial q, and hence the representation size n of w is upper-bounded by $2^{q(|G|+|H|)}$. Thus, for the coNEXP upper bound it only remains to show that $w \in \mathcal{L}(G)$ and $w \notin \mathcal{L}(H)$ can be checked in time polynomial in the n. This is not completely obvious since the word problem for ESCG is PSPACE-complete. In the full version of this paper, we show how this obstacle can be avoided, bringing in a strategy that was used by Huynh [10] in order to show a coNEXP upper bound for the equivalence problem for context-free commutative grammars. ◀

References

1 M. Beck and F. Sottile. Irrational proofs for three theorems of Stanley. *Eur. J. Comb.*, 28(1):403–409, 2007. doi:10.1016/j.ejc.2005.06.003.

2 E. Domenjoud. Solving systems of linear Diophantine equations: An algebraic approach. In *Mathematical Foundations of Computer Science, MFCS*, pages 141–150, 1991. doi:10.1007/3-540-54345-7_57.

3 S. Eilenberg and M.P Schützenberger. Rational sets in commutative monoids. *J. Algebra*, 13(2):173–191, 1969. doi:10.1016/0021-8693(69)90070-2.

4 J. Gallier. Notes on convex sets, polytopes, polyhedra, combinatorial topology, Voronoi diagrams and Delaunay triangulations, 2012. Manuscript available at `http://www.cis.upenn.edu/~jean/gbooks/convexpoly.html`.

5 S. Ginsburg. *The mathematical theory of context free languages.* McGraw-Hill, 1966.

6 S. Ginsburg and E.H. Spanier. Bounded ALGOL-like languages. *T. Am. Math. Soc.*, pages 333–368, 1964. `doi:10.2307/1994067`.

7 C. Haase and P. Hofman. Tightening the complexity of equivalence problems for commutative grammars. In *Symposium on Theoretical Aspects of Computer Science, STACS*, volume 47 of *LIPIcs*, pages 41:1–41:14. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016. `doi:10.4230/LIPIcs.STACS.2016.41`.

8 C. Haase, B. Nill, and A. Paffenholz. Lecture notes on lattice polytopes. Preliminary version of December 7, 2012, available at `http://polymake.org/polytopes/paffenholz/data/preprints/ln_lattice_polytopes.pdf`.

9 M. Hague and A.W. Lin. Model checking recursive programs with numeric data types. In *Computer Aided Verification, CAV*, volume 6806 of *Lect. Notes Comp. Sci.*, pages 743–759. Springer, 2011. `doi:10.1007/978-3-642-22110-1_60`.

10 D.T. Huynh. The complexity of equivalence problems for commutative grammars. *Information and Control*, 66(1–2):103–121, 1985. `doi:10.1016/S0019-9958(85)80015-2`.

11 D.T. Huynh. A simple proof for the Σ_2^p upper bound of the inequivalence problem for semilinear sets. *Elektron. Inf.verarb. Kybern.*, 22(4):147–156, 1986.

12 T.-D. Huynh. The complexity of semilinear sets. *Elektron. Inf.verarb. Kybern.*, 18(6):291–338, 1982.

13 O.H. Ibarra. Reversal-bounded multicounter machines and their decision problems. *J. ACM*, 25(1):116–133, 1978. `doi:10.1145/322047.322058`.

14 R. Ito. Every semilinear set is a finite union of disjoint linear sets. *J. Comput. Syst. Sci.*, 3(2):221–231, 1969. `doi:10.1016/S0022-0000(69)80014-0`.

15 E. Kopczyński. Complexity of problems of commutative grammars. *Log. Meth. Comput. Sci.*, 11(1), 2015. `doi:10.2168/lmcs-11(1:9)2015`.

16 M. Köppe and S. Verdoolaege. Computing parametric rational generating functions with a primal Barvinok algorithm. *Electr. J. Comb.*, 15(1), 2008.

17 J. Matoušek. *Lectures on discrete geometry.* Graduate texts in mathematics. Springer, 2002. `doi:10.1007/978-1-4613-0039-7`.

18 E.W. Mayr and J. Weihmann. Completeness results for generalized communication-free Petri nets with arbitrary edge multiplicities. In *Reachability Problems (RP'13)*, volume 8169 of *LNCS*, pages 209–221. Springer, 2013. `doi:10.1007/978-3-642-41036-9_19`.

19 A. Paffenholz. Polyhedral geometry and linear optimization. Preliminary version of July, 2010, available at `http://www.mathematik.tu-darmstadt.de/~paffenholz/daten/preprints/ln.pdf`.

20 R. Parikh. On context-free languages. *J. ACM*, 13(4):570–581, 1966. `doi:10.1145/321356.321364`.

21 L. Pottier. Minimal solutions of linear Diophantine systems: Bounds and algorithms. In *Rewriting Techniques and Applications, RTA*, volume 488 of *Lect. Notes Comp. Sci.*, pages 162–173. Springer, 1991. `doi:10.1007/3-540-53904-2_94`.

22 R.T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

23 A. Schrijver. *Theory of linear and integer programming.* John Wiley & Sons, 1986.

24 J. von zur Gathen and M. Sieveking. A bound on solutions of linear integer equalities and inequalities. *P. Am. Math. Soc.*, 72(1):155–158, 1978. `doi:10.1090/S0002-9939-1978-0500555-0`.

25 V. Weispfenning. The complexity of almost linear Diophantine problems. *J. Symb. Comp.*, 10(5):395–403, 1990. `doi:10.1016/S0747-7171(08)80051-X`.

Revision Notice

This is a revised version. It removes incorrectly stated upper bounds on the cardinalities of the sets of generators in Proposition 4 and Theorems 6 and 7.

Dagstuhl Publishing – October 26, 2016.